IntechOpen

Genetic Polymorphisms

New Insights

Edited by Mahmut Çalışkan

Genetic Polymorphisms - New Insights

Edited by Mahmut Çalışkan

Published in London, United Kingdom



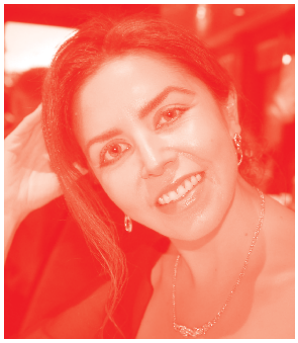




IntechOpen





Supporting open minds since 2005

Genetic Polymorphisms - New Insights
http://dx.doi.org/10.5772/intechopen.95195
Edited by Mahmut Çalışkan

Contributors
Tung Nguyen-Thanh, Thong Ba-Nguyen, Thuan Dang-Cong, Necla Benlier, Nevhiz Gundogdu, Mehtap Ozkur, Arshad A Pandith, Aabid Koul, Sheikh Mansoor, Usma Manzoor, Ina Bhat, Fozia Mohammad, Iqra Anwar, Qurat Ul Aein, Carmen Vladulescu, Shahid M. Baba, Daniel Kepple, Anthony Ford, Ebony Little, Gabrielle Kolesar, Beka R. Abagero, Ashley N. Blackwell, Swarnapali De Silva Indrasekara, Delenasaw Yewhalaw, Eugenia Lo, Nouha Bouayed Abdelmoula, Balkiss Abdelmoula, Orlex B. Baylen Yllano, Aloysius Brown, Leilani D. Dela Roca Arce, Epharaim A. Evangelista, Ferdinand A. Esplana, Lester Harris R. Catolico, Merbeth Christine L. Pedro, Subodh Kumar Jain, Shweta Yadav, Sapna Sedha, Shridhar C. Ghagane, Aimen Akbar, Sridevi I. Puranik, Surayya Siddiqui, Rakesh Choudhary, Subhash Chand, Tejveer Singh, Rajesh K. Singhal, Vinay K. Chourasiya, Indu

Assistant to the Editor: Nalan Tavşanlı

First published in London, United Kingdom, 2022 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Genetic Polymorphisms - New Insights
Edited by Mahmut Çalışkan
p. cm.
Print ISBN 978-1-83968-810-2
Online ISBN 978-1-83968-811-9
eBook (PDF) ISBN 978-1-83968-812-6

Meet the editor

Mahmut Çalışkan is a Professor of Genetics and Molecular Biology in the Department of Biology, Biotechnology Division, Istanbul University, Turkey. He obtained a BSc from Middle East Technical University, Ankara, and a Ph.D. from the University of Leeds, England. His main research areas include the role of germin gene products during early plant development, analysis of genetic variation, polymorphisms, and the characterization and biotechnological use of halophilic archaea.

Contents

Preface

Since the dawn of humankind, people have continually sought improvement in all aspects of life on Earth. Once it was established that living beings' characteristics are inherited and that genetic richness is an advantage against changing environmental conditions, people began to develop new strategies to reveal and conserve genetic diversity in both animal and human populations. The main evolutionary mechanisms of mutation, natural selection, and genetic drift have produced a wide variety of organisms, which resulted in the formation of many well-defined breeds with different levels of performance. Knowledge of the genetic polymorphisms and processes that underlie the origins and maintenance of organisms is important to providing critical insights into the structure and dynamics of living populations. Genetic polymorphisms are essential for populations to evolve and cope with environmental changes, new diseases, and pest epidemics. Genetic polymorphisms also provide the opportunity to trace the history of populations, species, and their ancestors. Therefore, the assessment of genetic variation in species and among populations is important for the conservation of genetic resources. Over the last few decades, there have been dramatic advances in molecular genetics and these developments have provided scientists involved in the conservation and improvement of genetic resources with a range of new techniques for their research.

Nowadays, techniques are available to determine changes at the DNA level in all kinds of organisms. Differences in gene sequences can be directly observed and described with a degree of precision previously impossible to achieve. Many of the techniques that have been developed have already been used to study the extent and distribution of changes in species gene pools and to investigate evolutionary and taxonomic relationships. With the development of polymerase chain reaction (PCR)-based techniques, in particular, numerous molecular technologies that can be used for the detection, characterization, and evaluation of genetic changes in populations have been, and continue to be, developed. These techniques can be chosen based on how they display genetic differences, the type of data they produce, the taxonomic levels at which they can be most appropriately applied, and their technical and financial requirements. The aim of the book is to provide current knowledge of genetic polymorphisms by presenting the works of scientists who are engaged in the production of new information that can be used to reveal genetic polymorphisms, often from very different perspectives. In the book, the genetic polymorphism studies that are carried out on different kinds of organisms at the DNA level or gene expression level are particularly important to evaluate the process of genetic polymorphisms.

Chapter 1, "Genetic Polymorphism in Animals", Jain et al, examine the recent development of polymorphisms in certain animals, including protozoa, earthworms, honeybees, silkworms, pearl oysters, fishes, and birds. In Chapter 2, "From Genes to Biomarkers: Understanding the Biology of Malaria Gametocytes and Their Detection", Kepple et al. provide information about genetic polymorphisms in gametocyte biomarkers that enable researchers to develop a more

sensitive and accurate diagnostic test for Plasmodium gametocytes. Chapter 3, "Thrombosis-Related DNA Polymorphisms", by Abdelmoula and Abdelmoula, discusses thrombotic disorders and their related diseases, particularly cardiovascular diseases, which are among the most common causes of morbidity and mortality in the world. The authors consider multiple, single, and combined genetic variations and polymorphisms, especially in the genes of coagulation and hemostasis pathway as well as the genes of inflammation and other genes interacting with lifestyle and environmental factors, such as immune and oxidative systems. In Chapter 4, "Cytokine Gene Polymorphism and Cancer Risk: A Promising Tool for Individual Susceptibility and Prognostic Implications", Pandith et al. reveal that germ line constitutional mutations or variations in polymorphic sequences are key instruments used as molecular tools to predict the susceptibility of a person to numerous cancers. They also point out that cytokines are thought to be essential molecules that have a dual face due to pro-inflammatory as well as anti-inflammatory mechanisms in carcinogenesis. Chapter 5, "Genetic Polymorphism and Prostate Cancer: An Update", by Siddiqui et al, provides current information about the occurrence and mortality rates of prostate cancer, genes associated with prostate cancer risk, single nucleotide polymorphisms (snps), genetic and environmental risk factors, and prostate cancer treatment. Chapter 6, "$HER2^{Ile655Val}$ Polymorphism and Risk of Breast Cancer", by Nguyen-Thanh et al., examines the $HER2^{Ile655Val}$ single nucleotide polymorphism and its association with risk of early-onset breast cancer. In Chapter 7, "Human Genetic Polymorphisms Associated with Susceptibility to COVID-19 Infection and Response to Treatment", Benlier et al. discuss the SARS-CoV-2 RNA virus, ACE2 carboxypeptidase, transmembrane serine protease, elastase activity, cathepsin-L peptidase activity, human alpha-1 antitrypsin glycoproteins, and HLA molecules in the scope of genetic polymorphisms. Chapter 8, "Characterization, Comparative, and Phylogenetic Analyses of Retrotransposons in Diverse Plant Genomes", by Brown et al., provides details about transposable elements, specifically the role of retrotransposons in diverse plant genomes. The authors also handle the issues of classes and types of transposable elements, characterization of retrotransposons, mechanism of action, the role of retrotransposons, multiple sequence alignment, conservation of retrotransposons, and phylogenetic analysis. Finally, in Chapter 9, "Sex Determination", Choudhary et al. consider mechanisms of sex determination, evolutionary differentiation of sex chromosomes, single gene sex determination, environmental sex determination, maternal and cytoplasmic sex determination, and mixed-sex determination systems.

I would like to express my deepest gratitude to all the authors who contributed to this book by sharing their valuable works. This book is a useful resource for students, researchers, and experts in molecular genetics.

Dr. Mahmut Çalışkan
Professor,
Faculty of Sciences,
Biology Department,
Biotechnology Division,
İstanbul University,
İstanbul, Turkey

Chapter 1

Genetic Polymorphism in Animals

Subodh Kumar Jain, Shweta Yadav and Sapna Sedha

Abstract

Biological diversity is the variability among living organisms from all sources of nature. Genetic polymorphism study support a lot when any economically important particular species is taken into consideration. The knowledge of genetic background of a species and its population structure is very essential for their successful conservation and management. Molecular techniques have been supporting in the determination of population diversity and also to determine the genetic architecture of a wide variety of closely related individuals. Molecular techniques based on DNA polymorphism are now used in population genetic studies, systematic and molecular taxonomy. This chapter will provide information on genetic diversity of various economically important species such as protozoa, worms, insects, pearl oyster, fishes and birds. The study of genetic variations in economically important species has practical significance for developing strategies to control the disease, to improve reproductive traits, yield more beneficiary products like honey, silk, pearl, manure, etc. Since there are some data gaps, most suitable and promising technology must be used to elucidate the role of every single gene involved in the pathways to be studied in order to apply for more benefit to the society.

Keywords: genetic diversity, economically important species, protozoa, worms, insects, pearl oyster, fishes, birds

1. Introduction

Biological diversity is the variability among living organisms from all sources of nature. Information on molecular structure of economically important organism is useful for optimizing identification of stock, stock enhancement, breeding program, management of sustainable yield and preservation of diversity. Genetic polymorphism study plays an important role to understand the basis of population differentiation and species diagnostics. In recent times molecular techniques have been supporting in the determination of population diversity and also to determine the genetic architecture of a wide variety of closely related individuals. The discovery of PCR has a major impact on eukaryotic genome and contributed to the development and application of various DNA markers. Within the population, organisms are identified by their morphology but there are some small invisible changes observed due to environmental effect. Genetic polymorphism study support a lot when any economically important particular species is taken into consideration.

This chapter will provide information on genetic diversity of various economically important species such as Protozoa, Worms, Insects, Pearl oyster, Fishes and Birds.

2. Protozoa

There are many factors associated with protozoan parasites genetic diversity such as: transmission and passage history in laboratory conditions; occurrence in different hosts or geographic regions; selective pressure of drugs and competitive interactions among populations. However, the number of examined isolates of parasites and genetic markers, assortment of methods, probes, primers and reagents used is also of significance. The significance of genetic variability in parasite populations is still the subject of interest and controversy. A simple interpretation of such variation is impossible because of the complexity of host–parasite interactions. The knowledge of parasite diversity at the nucleic acids level has continually increased, but a correct interpretation of this phenomenon requires at least the same knowledge of genetic variability in host populations [1].

The study of genetic variation in malaria parasites has practical significance for developing strategies to control the disease. The genome diversity of the important human pathogen *Plasmodium vivax*, however, remains essentially unknown. The data about Single Nucleotide Polymorphisms (SNPs) show that *P. vivax* has a highly diverse genome, and provide useful information for further understanding the genome diversity of the parasite [2].

All RAPD primers yield patterns that differ between species and thus could serve as species-diagnostic traits. But the extensive polymorphisms observed within each species for most RAPD primers preclude their practical use for species diagnosis. There are, however, primers that yield monomorphic patterns within a species and thus can readily be used for species diagnosis, which may be useful for epidemiological and other purposes. For example, two primers are monomorphic in *T. cruzi* and thus could be used in epidemiological practice for differentiating *T. cruzi* from *T. rangeli*, a species that infects humans [3].

3. Worms

For stem cells and regeneration study, free-living flatworms (the planarian *Schmidteamediterranea)* are extensively used as model organisms. The germline-enriched genes of the flatworm *M. lignano* have a high fraction of flatworm-specific genes. The *Mlig-sperm1* gene responsible for producing healthy spermatozoa has been identified as a member of the novel gene family conserved only in free-living flatworms [4].

TIM29 (mitochondrial inner membrane protein) was shown to stabilize the protein import complex TIM22 by interacting with it, but its biological function remains largely unknown. Till now, it was classified as one of the Domain of Unknown Function (DUF) genes, with a conserved protein domain DUF2366 of unclear function. It has been demonstrated that DUF2366/TIM29 knockdown in *Macrostomumlignano* prevents worms from adapting to a highly proliferative state required for regeneration with least effect during the normal homeostatic condition [5].

Low response to ivermectin (IVM) in patients infected with *Onchocerca volvulus* indicates that the parasite might be under a selection process toward potential resistance. In order to limit this process, the characterization of *O. volvulus genes* is very crucial. It has been observed that a deficit of heterozygous female worms leading to Hardy Weinberg disequilibrium, which might be explained by a shorter life-span of these worms compared to the homozygous worms. Also the heterozygous female worms were much less fertile than the homozygotes: more than two thirds of the homozygotes were fertile, whereas only 37% of the heterozygotes were fertile [6].

Control of onchocerciasis or river blindness by mass treatment of the population with IVM has been a great success until now, so that in certain foci its elimination has become feasible. However, after more than 20 years of repeated IVM mass treatment, the disease still persists in many endemic countries. Sub-optimal responses and genetic changes have been reported in *Onchocerca volvulus* populations under high IVM pressure but more work is needed to determine whether resistance is developing. In a study four SNPs occurring in the β-tubulin gene of these parasites were investigated and found changes in genotype frequencies in *O. volvulus* β-tubulin gene associated with IVM treatments. The SNP at position 1545 (A/G) showed a significant increase in frequency of the less common nucleotide in the female worms following treatment. After three-monthly treatments, female worm homozygotes with the less common genotype, prior to treatment, increased in frequency. The selected homozygotes, as well as heterozygotes, appeared to be less fertile than the wild-type homozygotes. These results provide additional evidence for genetic selection [7].

4. Earthworms

From a unicellular organism like protozoans, nature has used its nimble fingers to create the structural complexity inside the organisms at diverse stages, to conform themselves very correctly to the prevailing conditions. In this scheme of their polymorphism, it'd be suitable to emphasize the evolutionary significance of earthworms which cause them to masters the soil invertebrate community. In general, lumbricid earthworms have ruled of their distribution in temperate soils and the megascolecid earthworms predominate the sub-tropical and tropical soils. The depth of competition is intense in temperate areas with a slender area of interest where litter forms the primary food source [8]. On the opposite hand, niche and morphogenecity have been enlarged with greater food diversity for tropical earthworms and they display awhole lot of variant in size and behavioral patterns.

4.1 Molecular markers and earthworm genetic polymorphism

Despite being efficient in soils, their relevance and research over the past 130 years are still fragmentary therefore, the various species complex and morphs are not being resolved adequately. Molecular systematics can be an essential supply of facts to delimit species and to assign taxonomic categories in complex species. Several genes including cytochrome oxidase 1 (COI), 18S, 16S, 28S, and protein-coding histone H3 genes had been currently used for reading phylogeny and polymorphism in earthworm species. The COI gene is a 658 bp trendy marker that has been demonstrated to be powerful for earthworm identifications, molecular systematics, ecology, phylogeography, and cryptic speciation. Furthermore, the other markers specifically 18 s, 28 s, are often involved to have a look at interfamily, intrageneric, and shallow intraspecific diversity. Most of the polymorphic research are primarily based on mitochondrial markers mainly COI, COII, and different protein-coding genes as they evolve extra unexpectedly than nuclear genes thus resulting in the accretion of differences between closely related species.

The Barcode of Life Data System (BOLD, http://www.barcodinglife.org), a valuable integrative bioinformatics platform, serves as a systematic workbench helping all stages of the analytical pathway from specimen collection to validation was formulated by Ratnasingham [9]. Chang et al. [10] investigated polymorphism in *Metaphire formosae* species group, a member of the *Pheretima* complex using DNA sequences of (*COI*), 16S ribosomal (r)RNA, and NADH dehydrogenase

subunit 1 (ND1) and revealed the presence of 13 taxa of the *M. formosae* species group, including a cryptic species. A study sequenced the *COI*, 16S, tRNAs, and 28S genes in 202 Hormogastridae earthworm individuals in the Iberian Peninsula and suggested the presence of excessive genetic range with the presence of five cryptic allopatric species [11]. Another study identified two new earthworm species namely *E. nordenskioldimongol* and *E. nordenskioldionon* from Mongolia which were justified by molecular taxonomy using mitochondrial DNA barcoding [12]. Similarly, Dario et al. [13] provided the description of one new earthworm species viz., *Eiseniona gerardoi* primarily based on mitochondrial in addition to nuclear molecular markers, within the controversial genus *Eiseniona* of lumbricidae from the region of Extremadura (Spain). Dominguez et al. [14] studied the evolution of lumbricids *via* way of means of reading one hundred sixty earthworm individuals using the sequencing of two nuclear genes and seven mitochondrial tRNAs, with 22 morphological characters, discovered 84 lumbricid species under 28 genera. Also, Dmitry and Gennady [15] studied the taxonomical status of the exceptional ecological forms of *D. ghilarovi* using *COI* and 16 s genes as molecular tools. They discovered apparent differentiation among the meadow-swamp black form and the forest gray form of *D. ghilarovi*. Decaens et al. [16] sequenced 651 individuals using *COI* gene, which corresponds to 48 MOUTs, and concluded that factors like ecological processes in addition to long-time period diversification are critical in structuring and diversity of earthworm communities in tropical rainforests of French Guiana. Moreover, Hong and Csuzdi [17] revoked the phylogenetic affinities of Korean *E. nordenskioldi* specimens and compared them with Siberian *E. nordenskioldi* individuals on the basis of mitochondrial gene marker. Also, Shekhovtsov et al. [18] investigated the genetic variations of *Eisenia norden skioldipallida* Malevic from various climatic zones of Northern Asia by sequencing of *COI* and ITS2 loci and detected five cryptic genetic lineages within *E. nordenskioldipallida*. Teerapong et al. [19] explored new earthworm species, *Pontodrilus longissimus* sp. nov., from beaches of Thailand and Peninsular Malaysia primarily based on morphological investigations and partial sequencing of mitochondrial *COI* gene. The molecular phylogeny of Lumbricidae earthworms was investigated by Farnaz et al. [20] using phylogenetic evaluation of two nuclear gene regions (28S rDNA and 18S rDNA) and 11 mitochondrial genes (16S rDNA, 12S rDNA, NADH-I, *COI* and *COII* and tRNAsAsn, Asp, Val, Leu, Ala and Ser) that lead to the addition of one new genus; *Philomontanus gen*. Nov and its three new species namely, *Philomontanussarii* sp. nov., *Philomontanusmahmoudi* sp. nov. and *Philomontanusbaloutchi* sp. nov.

In India, the studies on earthworm polymorphism using molecular tools were unavailable till 2008. However, two studies used the RAPD-PCR technique for the first time to evaluate the genetic diversity of Indian earthworms [21, 22]. Various molecular markers viz., Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphism (RFLP), and Simple Sequence Repeat (SSR) [23, 24] were sequenced in addition to mitochondrial *COI* gene of six species of earthworms and determined that the *COI* exhibited a unique barcode to a particular species. One of the latest works [25] explored four new species of *Kanchuria* (Megascolecidae) from Meghalaya and updated the checklist of the Northeastern region of India [26]. Also, Thakur et al. [27] studied polymorphism in *Eutyphoeus* sp. using mitochondrial gene marker. Tiwari et al. [26] studied the earthworm polymorphism of the Sagar district of India using integrative methods. Nonetheless, earthworm molecular taxonomy based on DNA sequencing in India is at its initial stage, such that very little information is available. Recently, 801 DNA sequences of Indian earthworm (belonging to family Acanthodrilidae 4, Almidae 3, Eudrilidae 22, Hormogastridae 7, Lumbricidae 6, Megascolecidae 426, Moniligastridae 122, Octochaetidae 199, Rhinodrilidae 6) have been made available on Bold-system

under the project diversity studies of Indian earthworms using DNA barcodes. This signifies only a fraction of Indian earthworms and thus further challenges remain open for biology fanatics to research in the field of genetic polymorphism using molecular approach.

4.2 Reproductive organs polymorphism

The absence of certain reproductive organs is ordinarily essential for biparental replica in hermaphroditic oligochaetes [28]. During the evolutionary process, the organs important in biparental reproduction in hermaphroditic animals were found removed in certain Japanese species of *Pheretima* and evolved into a various intraspecific morphs viz., H (primary, from which different morphs derived and with biparental replica), Hp (secondary, reproductive machine incomplete however a number of organs stay juvenile), A (without spermathecae, storage of sperm impossible), R (without terminalia, discharge of sperm impossible), AR (without spermathecae and terminali, storage of foreign sperm and extrusion of own sperm impossible) and ARZ (testes and other male organs and spermathecae lacking). That reproductive-organ polymorphism in *Pheretima* emphasizes a normal alternate into being uniparental. Organ disorder denial of alternative among uni and biparental may also make parthenogenesis pseudo-obligatory in selected complex of earthworms.

4.3 Biochemical polymorphism in earthworms

Since, the genetic variation in earthworm species may cause mild effect on phenotype, voluminous work is focused on the effects of only one or a few polymorphic loci at any one time. There is an upcoming concern of an individual possessing numerous polymorphism (set of polymorphic loci), which may act in concert to affect phenotype. That can be assessed through correlating genetic houses of loci with physical and/or biochemical properties of the respective gene products. The variation in distribution of polymorphism among different pathways such as glycolysis, Kreb's cycle, hexose monophosphate shunt inside glucose metabolism and ancillary pathways leading to amino acid metabolism was first highlighted [29]. The polymorphic loci which persist in conserved pathways should have common variants with intermediate biochemical activities and less uncommon alleles than polymorphic loci in different pathways. Comparisons of the activities of rate limiting enzymes among different invertebrates suggest that earthworms have a fairly low metabolic rate [30–32]. The studies are beneficial particularly in resolving polytypic complexes and the genetic consequences of parthenogenesis in invasive species.

5. Insects

Insects represent a major life form on earth. So far nearly 0.9 million insect species have been discovered comprising 75% of all the recorded animal species. Some of the insect species are easy to identify and categorize while for others it is difficult due to their small size and environmental factors that resulted in to morphological variation. To overcome these problems, the advanced molecular techniques viz. PCR, RFLP, and ALFP have been a great help. RAPD markers have been used in gene mapping to characterize cultivars and species genetically, infer phylogeny and biogeography of insect population and understand modes of evolution and evolutionary trajectories [33]. RAPD-PCR analysis was used to confirm

genetic differentiation between two cockroach species *Perplaneta Americana* and *Blatellagermanicana* [34].

5.1 Honey bee

Honey bee is the colonial insect with complex social behavior. Beside its economic importance it has long been important for the production of honey, wax, behavioral study and pollination of crops. Nemobiologists and behaviorists have used the honey bee as a model organism to study the molecular basis of learning [35]. For genetic analysis honey bee genome sequencing project also proposed that it will benefit human health and medicine in diverse areas including venom toxicology, allergic diseases, mental illness, infectious diseases, parasitology and gerontology and will also improve human nutrition by enabling enhanced pollination of food plants and accelerated delivery of hymenopteran parasitoids for biological control of pests [36]. Honey bee has a higher rate of meiotic recombination than any other known metazoan [37]. The higher recombination rates effectively increase the accuracy of linkage mapping and high recombination rate and the low incidence of repetitive DNA should facilitate map based cloning of genes in the honey bee [38]. AFLP markers and microsatellites have been used in dissecting the guarding and stinging behavior in honey bee. Division of labour, expression of guarding and stinging behavior is influenced by specific quantitative trait loci [39]. By using multilocus fingerprinting super and half-sister in a colony of honey bee has been discriminated with synthetic oligonucleotide [40].

5.2 Silkworm

The silkworm *Bombyxmori* is domesticated for silk production for about 5000 The silkworm *B. mori* is domesticated for silk production for about 5000 years. The well characterized mutations of *B. mori* (a well studied lepidopteran model system) affect every aspect of organism's morphology, development, behavior and its considerable economic importance [41].

Since *B. mori* is of great economic importance to silk producing countries such as India, Russia, Japan, Korea, China, Bulgaria and Iran, a number of silkworm breeds have been collected suitable for a wide range of agroclimatic conditions. More than 4000 strains are maintained in the germplasm of *B. mori* and 46 institutes are involving silkworm genetic resources maintenance, which includes univoltine, bivoltine, and polyvoltine strains. These different genotypes possess so many differences in their quantitative and qualitative traits which control silk yield. It was estimated that silkworm genome contains about 108 bp and its genetic information volume is nearly one-sixth of human being. There are over 450 morphological, physiological, and biochemical characters recorded at present, among them 300 (including multiallele) had been located on 27 groups of the total 28 chromosomes [41]. Apart from a rich biodiversity of geographical races, there are also a large number of mutants for a variety of characters present in *B. mori* [42]. Zhang et al. [43] reported that genetic distances within Japanese strains are closer than those of Chinese strains and within a strain; the individual polymorphism is significantly higher in wild silkworm than those of domesticated silkworm. According to Liu [44] at the species level, *Antheraeapernyi* and *Bombyxmori* showed high levels of genetic diversity, whereas *Samia Cynthia ricini* showed low level of genetic diversity. However, at the strains level, *Antheraeapernyi* had relatively the highest genetic diversity and *B. mori* had the lowest genetic diversity. Yukuhiro et al. [45] analyzed PCR amplified carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD) gene fragments from 146 *Bombyxmori* native strains

and found extremely low levels of DNA polymorphism. CAD haplotype analysis of 42 samples of Japanese *B. mandarina* revealed four haplotypes. No common haplotype was shared between the two species and at least five base substitutions were detected. These results suggesting that low levels of gene flow between the two species. Further extremely low level of DNA polymorphism in *B. mori* compared to its wild relatives suggested that the CAD gene itself or its tightly linked regions are possible targets for silkworm domestication. This information clearly indicates narrow level of genetic diversity in silkworm.

The existence of genetic variation within a population is crucial for its ability to evolve in response to novel environmental challenges. In order to adapt to new environment and conditions, the genetically variable populations are thought to evolve their morphological, physiological or behavioral mechanisms [46]. This not only results in better adaptation to their local environments, but may also lead to reduction in the genetic variations thereotically. Domestication, phenotypic selection, breeding systems, genetic drift is thought to be the main reason for reduced genetic diversity in silkworms.

6. Pearl oyster

The pearl oyster *Pinctadafucata* is a commercially important marine shellfish cultured for producing saltwater pearls manly in China and Japan [47]. Researchers using genome-wide genetic data from specimens collected across the western Pacific, elucidated how pearl oyster populations vary genetically and geographically. Their analyses provide insight into how these pearl oysters have adapted to environmental changes over time. They sequenced the genomes of the specimens and analyzed 36,203 single nucleotide polymorphism (SNP) sites [48].

Genetic variability and the pattern of population structure among 9 samples of Calafia pearl oyster Pinctadamazatlanica collected from Mexico to Panama, using mtDNA RFLP analysis of two genes 12S rRNA and subunit one of Cytochrome oxydase (COI). Haplotype diversity varied from 0.000to 0.856. The Panama population appeared to be monomorphic, while the other samples exhibited a level of haplotypic variability. A test for the impact of demographic history on genetic diversity was applied on the sequence data, and the results were congruent with a recent decline of population sizes. Three significantly distinct groups could then be defined, which correspond to Northern Mexico, Southern Mexico, and Panama [49]. The pearl oyster, *Pinctadafucata*, a marine bivalve belonging to the family Pteriidae, is the primary species cultured for marine pearls in China and Japan [50]. In a study 16 SNP markers have been reported, may provide a useful tool for population genetics and evolutionary analysis [51].

7. Fish

Fishes are economically important animal consumed globally. Information on the molecular structure of fish species is useful for optimizing identification of stocks, stock enhancement, breeding programs, and management of sustainable yield and preservation of genetic diversity. The knowledge of genetic background of a species and its population structure is very essential for successful fisheries conservation and management. A number of methods have been developed to measure genetic diversity within the species [52]. Molecular techniques based on DNA polymorphism are now used in population genetic studies, systematic and molecular taxonomy. Molecular techniques played an important role to understand

the basis of polymorphism of a species, species diagnostics and population differentiation. Three fish species of fish *Labeorohita, Catlacatla, Cirrhinamrigala* (family Cyprinidae) of India have beenstudied by RAPD for molecular identification. The diverse nature of DNA bands indicated the genetic distance between fish species and presence of common bands attributed to an evolutionary relationship. Pattern of species specific unique bands are useful for identification [53]. In addition to *Labeorohita* and *Catlacatla,* another fish *Nile tilapia* of the same family has also been studied showing genetic relationship [54]. Kempter et al. used microsatellites for the authentication of torpedo scad (Megalaspiscordyla) -fish species by successfully extracting a DNA fragment of the nuclear rhodopsin gene (RH1) and amplification of nine microsatellite regions (SSRs). After analysis of the products, they found differences between the RH1 sequences and those obtained from Gene Bank. This study was used to characterize and assess the genetic diversity of the populations as well as effectively categorized the populations over the areas of the western Indian Ocean and the western Pacific [55].

8. Birds

Major factors of economic concern in the modern poultry industry are the reproductive traits i.e. age at first egg, number of eggs and weight of the egg [56–59]. These reproductive traits are of great importance as studies on these genes helps in revealing the genetic mechanisms affecting egg-laying performance and for breeding the laying hens with high productivity and quality [60–64]. RT-PCR results of a study showed that the growth differentiation factor 9 *(GDF9* gene*)* is involved in determining reproductive traits in chicken as they observed its high expression in stroma with cortical follicles (STR) and prehierarchal follicles [62].

Very low density apolipoprotein-II (apoVLDL-II) is a major polypeptide component of avian VLDL. The function of apoVLDL-II is the transport of neutral lipids (triacylglycerol) in the form of VLDL in the plasma. The apoVLDLII gene is dormant in embryos, chicks and roosters but can be activated by estrogen. Genotyping for the apoVLDL-II gene showed a mutation in 492-bp fragment located on the first intron. Polymorphism in apoVLDL-II gene was significantly associated with body weight at 6 week (BW6), carcass weight (CW), breast muscle weight (BMW), drumstick weight (DW) and wing weight (WINW). Association between single nucleotide polymorphism of apoVLDL-II gene with growth and body composition traits in Iranian commercial broiler line [65].

Genetic diversity measured at the molecular level does not always correspond to phenotypic breed diversity, because a long history of exchange, upgrading and crossbreeding has sometimes created new genotypes within old phenotypes. For example, breeders of fancy breeds are mainly concerned about the phenotype, whereas the genotype of phenotypically different breeds may be very similar [66].

In a study, it has been reported that the three structurally related orphan G protein-coupled receptors (GRP3, GPR6, GPR12) are constitutively active and thought to regulate neuronal outgrowth and oocyte meiotic arrest in mammals. But the information is scanty related to this data in case of non mammalian vertebrates therefore require further research. The cloned duck GPR3 and duck/chicken GPR6 and GPR12 are intron-less and encode receptors that show high amino acid sequence identities (66–88%) with their respective mammalian orthologs. It has been demonstrated that GPR3, GPR6, GPR12, and GPR12L are constitutively active and capable of stimulating the cAMP/PKA signaling pathway without ligand stimulation in birds (and zebrafish), indicating their conserved signaling property across vertebrates [67]. Several researches have demonstrated that RNA-Seq is the most

suitable and promising technology which gives a definitive view of the pathways involved and the role of every single gene in that process [68]. RNA-seq data/qRT-PCR assays revealed that GPR6 and GPR12L expression is mainly restricted to the chicken brain, while GPR12 is highly expressed in chicken ovarian granulosa cells (GCs) and oocytes of 6 mm growing follicles and its expression in cultured GCs is upregulated by progesterone [67].

Signal transducers and activators of transcription (STATs) represent a family of latent cytoplasmic proteins that mediate a variety of peptide hormones and cytokines in a target cell [69], controlling the action of growth hormones on target genes [70]. STAT5B regulates ovary development and sexual maturation [71]. Moreover, in another study, STAT5B knock-out mice had no breast development or lactation [72].

9. Conclusion

The study of genetic variations in economically important species has practical significance for developing strategies to control the disease, to improve reproductive traits, yield more beneficiary products like honey, silk, pearl, manure, etc. Further, the study of genetic variation will help in area wise differentiation of the species. Also, the data pertaining to G protein-coupled receptors (GRP3, GPR6, GPR12) is scanty in case of non mammalian vertebrates which needs further research. RNA-Seq -the most suitable and promising technology must be used to elucidate the role of every single gene involved in the pathways to be studied in order to apply for more benefit to the society.

Author details

Subodh Kumar Jain*, Shweta Yadav and Sapna Sedha
Department of Zoology, Harisingh Gour University, Sagar, Madhya Pradesh, India

*Address all correspondence to: subjain@gmail.com

References

[1] Majewska AC, Sulima P (1999) Source and significance of genetic polymorphism of selected parasitic protozoa. Wiad Parazytol, 45(3): 293-307.

[2] Xiaorong Feng, Jane M Carlton, Deirdre A Joy, Jianbing Mu, Tetsuya Furuya, Bernard B Suh, Yufeng Wang, John W Barnwell, Xin-Zhuan Su (2003) Single-nucleotide polymorphisms and genome diversity in Plasmodium vivax. Proc Natl Acad Sci USA, 100(14):8502-8507. DOI: 10.1073/pnas.1232502100

[3] Michel Tibayrenc, T. Katjaneubauer, Christian Barnabo, Franoiseguerrini, Douglas Skarecky, and Francisco J. Ayala (1993) Genetic characterization of six parasitic protozoa: Parity between random-primer DNA typing and multilocus enzyme electrophoresis (phylogeny/clonal theory/ Trypanosoma/Leishmanial/Plasmoum). Proc Natl Acad Sci USA, 90:1335-1339.

[4] Magda Grudniewska, Stijn Mouton, Margriet Grelling, Anouk HG, Wolters, Jeroen Kuipers, Ben NG, Giepmans, and Eugene Berezikov (2018) A novel flatworm-specific gene implicated in reproduction in *Macrostomumlignano*. Scientific reports 8, 3192. doi: 10.1038/ s41598-018-21107-4

[5] Stijn Mouton, Kirill Ustyantsev, Frank Beltman, Lisa Glazenburg, and Eugene Berezikov (2021) TIM29 is required for enhanced stem cell activity during regeneration in the flatworm *Macrostomumlignano*. Sci Rep, 11(1):1166, doi: 10.1038/s41598-020-80682-7

[6] C Bourguinat, S D S Pion, J Kamgno, J Gardon, N Gardon-Wendel, B O L Duke, R K Prichard, M Boussinesq (2006) Genetic polymorphism of the beta- tubulin gene of Onchocera volvulus in ivermectin naïve patients from Cameroon, and its relationship with fertility of the worms. Parasitology, 132 (2) 255-262, Doi10.1017/S003118200 5008899

[7] Hugues Nana-Djeunga, Catherine Bourguinat, Sebastien DS Pion, Joseph Kamgno, Jacques Gardon, Flobert Njiokou, Michel Boussinesq, Roger K Prichard (2012) Single nucleotide polymorphisms in β-tubulin selected in *Onchocerca volvulus* following repeated ivermectin treatment: possible indication of resistance selection. Mol Biochem Parasitol, 185(1): 10-18, DOI: 10.1016/j.molbiopara.2012.05.005

[8] Kale, Radha D (1998) Earthworm - Cinderella of Organic farming, Prism Books Pvt Ltd, Bangalore, India. p. 88.

[9] Ratnasingham S, Hebert, PDN (2007) Barcoding Bold: The Barcode of Life Data System (www.barcodinglife. org). Molecular Ecology Notes, 7: 355-364.

[10] Chang CH, Lin SM, Chen JH (2008) Molecular systematics and phylogeography of the gigantic earthworms of the *Metaphireformosae* species group (Clitellata, Megascolecidae). Mol. Phylogenet. Evol., 49: 958-968.

[11] Novo M, Almodovar A, Fernandez R, Trigo D, Cosın DJD (2010) Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. Molecular Phylogenetics and Evolution, 56(1):507-512.

[12] Blakemore RJ (2013) Earthworms newly from Mongolia (Oligochaeta, Lumbricidae, *Eisenia*). ZooKeys, 285:1-21.

[13] Dario J, Díaz C, Marta N, Rosa F, Daniel FM, Mónica G (2014) A new earthworm species within a controversial genus: *Eisenionagerardoi*

sp. n. (Annelida, Lumbricidae) description based on morphological and molecular data. ZooKeys, 399: 71-87

[14] Dominguez J, Aira M, Breinholt JW, Stojanovic M, James SW, Perez-Losada M (2015) Underground evolution: new roots for the old tree of lumbricid earthworms. Molecular Phylogenetics and Evolution, 83:7-19.

[15] Dmitry MA, Gennady NG (2015) Genetic differentiation of black and grey colored forms of the earthworm *Drawidaghilarovi* Gates, 1969 (Moniligastridae, Oligochaeta) on Russian Far East. European Journal of Soil Biology, 67: 12-16.

[16] Decaens T, Porco D, James SW, Brown GG, Chassany V, Dubs F, Roy V (2016) DNA barcoding reveals diversity patterns of earthworm communities in remote tropical forests of French Guiana. Soil Biology & Biochemistry, 92: 171-183.

[17] Hong Y, Csuzdi C (2016) New data to the earthworm fauna of the Korean peninsula with redescription of *Eiseniakoreana* (Zicsi) and remarks on the *Eisenianordenskioldi* species group (Oligochaeta, Lumbricidae). Zool Studies, 55:12

[18] Shekhovtsov SV, Berman DI, Bazarova NE, Bulakhova NA, Porco D, PeltekSE (2016) Cryptic genetic lineages in *Eisenianordenskioldipallida* (Oligochaeta, Lumbricidae). European Journal of Soil Biology, 75: 151-156.

[19] Teerapong S, Chirasak S, Parin J, Ratmanee C, Somsak P (2018) Morphological and molecular evidence reveal a new species of the earthworm genus Pontodrilus Perrier, 1874 (Clitellata, Megascolecidae) from Thailand and Peninsular Malaysia. Zootaxa, 4496 (1): 218-237.

[20] Farnaz B, Marjan S, Masoumeh M, Manuel A, Marcos P, Jorge D (2019) Multigene phylogeny reveals a new Iranian earthworm genus (Lumbricidae: Philomontanus) with three new species. PLoS ONE, 14(1): e0208904. doi: 10.1371/journal.pone.0208904

[21] Giraddi RS, Meenatchi R, Suresh B, Biradar MD, Biradar DP (2009) Standardization of method for genomic DNA extraction in earthworms. Karnataka Journal of Agricultural Science,22: 918-920.

[22] Meenatchi R, Giraddi RS, Biradar DP (2009) Assessment of genetic variability among strains of earthworm, *Eudriluseugeniae* (Kinberg) using PCR-RAPD technique. Karnataka Journal of Agricultural Science, 22: 942-945.

[23] Yadav S, Mullah M (2017) A Review on Molecular markers as tools to study earthworm diversity. International Journal of Pure and Applied Zoology, 5: 62-69.

[24] Jaya M, Aja M, Nair VK (2015) Biomolecular approach to Oligochaete Taxonomy. International Journal of New Technologies in Science and Engineering, 2: 74-83.

[25] Lone AR, Tiwari N, Thakur SS, Pearlson O, Pavlicek T, Yadav S (2020) Exploration of four new Kanchuria sp. of earthworms (Oligochaeta: Megascolecidae) from the North Eastern Region of India using DNA bar-coding approach. Journal of Asia-Pacific Biodiversity, 13(2): 268-281

[26] Tiwari N, Lone AR, Thakur SS, Yadav S (2020) Interrogation of earthworm (Clitellata: Haplotaxida) taxonomy and the DNA sequence database, Journal of Asia-Pacific Biodiversity, 14(1): 40-52.

[27] Thakur SS, Lone AR, Tiwari N, Yadav S (2020) Exploring new records of *Eutyphoeus* sp. (Haplotaxida: Octochaetidae) from garo hills,

Meghalaya, North Eastern state of India with use of DNA barcodes, Mitochondrial DNA Part A, 31: 265-272. DOI: 10.1080/24701394.2020.1781834.

[28] Gates G E (1956) Reproductive Organ Polymorphism in Earth-Worms of the Oriental Megascolecine Genus *PheretimaKinberg* 1867. Evolution, 10 (2): 213-227

[29] Diehl Walteri J and Williams L Donald (1992) Carbohydrate metabolism in the earthworm *Eiseniafetida* (Oligochaeta).Comp. Biochem. Physiol., 101 :83-90.

[30] Crabtree B and Newsholme E A (1972) The activities of phosphorylase, hexokinase, phosphor fructokinase, lactate dehydrogenase and the glycerol 3-phosphate dehydrogenases in muscles from vertebrates and invertebrates. Biochem.J.,126:49-58

[31] Zammit V A and Newsholme E A (1976) The maximum activities of hexokinase, phosphorylase, phosphofructo- kinase, glycerol phosphate dehydrogenases, lactate dehydrogenase, octopine dehydrogenase, phosphoenol- pyruvate carboxykinase, nucleoside diphosphatekinase, glutamate-oxaloacetate transaminase and arginine kinase in relation to carbohydrate utilization in muscles from marine invertebrates. Biochem. J., 160: 447-462.

[32] Alpe P R, Newsholme E A and Zammit VA (1976) Activities of citrate and NAD^+linked and $NADP^+$linkedisocitrate dehydrogenase in muscle from vertebrates and invertebrates. Biochemistry Journal, 154: 689-700.

[33] Jain SK, Neekhra B, Pandey D, Jain K (2010) RAPD marker system in insect study: A review Ind J Biotechnology, 9: 7-12.

[34] Neekhra B, Pandey D, Jain SK (2012) RAPD-PCR based marker approach for the genetic differentiation of two species of cockroach (Order Dictioptera). J. of Life Sc., 6:1328-1333.

[35] Neekhra B, Pandey D, Mishra M, Jain SK (2012) Molecular marker approach in honey bee: A review , Int J Pharma Bio Sci, 3(3B): 261-271.

[36] Heckel DG(2003) Genomics in pure and applied entomology. Annual review of Entomology, 48: 235-260.

[37] Beye M, Hunt GJ, Page RE, Fondrk MK, Grohmann L, Moritz RFA (1999) Unusually high recombination rate detected in the sex locus region of the honey bee *Apismellifera*. Genetics, 153: 1701-1708.

[38] Tomkins JP, Luo M, Fang GC, Main D, Goicoechea JL, Atkins M, Frisch DA, . Page RE, Guzman-novoa E, Yu Y, Hunt G and Wing RA (2002) New genomic resources for the honey bee Apismellifera L: development of a deep-coverage BAC library and a prelimnary STC database. Genet. Mol. Res., 1 (4): 306-316.

[39] Arechavaleta-Velasco ME, Hunt GJ Emore C (2003) Quatitative trait loci that influence the expression of guarding and stinging behaviors of individual honeybees. Behavioral Genetics, 33:357-364.

[40] Blanchetot A (1991) Genetic relatedness in honey bees as established by DNA fingerprinting. J. Hered., 82: 391-396.

[41] Nagaraju J and Goldsmith MR (2002) Silkworm genomics-progress and prospects. Current Science 83 (4): 415-425.

[42] BannoY, Shimada T, Kajiura Z, and Sezutsu H (2010) The silkworm—an attractivebioresource supplied by Japan. Experimental Animals, 59(2): 139-146.

[43] Zhang L, Huang Y, Miao X, Qian M and Lu C (2005) Microsatellite markers

application on domesticated silkworm and wild silkworm. Insect Science, 12(6): 413-419.

[44] Liu Y, Qin L, Li Y et al. (2010) Comparative genetic diversity and genetic structure of three chinese silkworm species Bombyxmori L. (Lepidoptera: Bombycidae), Antheraeapernyiguerin-meneville and samiacynthiaricinidonovan (Lepidoptera: Saturniidae). Neotropical Entomology, 39(6):967– 976.

[45] Yukuhiro K, Sezutsu H, Tamura T et al. (2012) Little gene flow between domestic silkmothBombyxmori and its wild relative Bombyxmandarina in Japan, and possible artificial selection on the CAD gene of B. Mori. Genes Genetics Systemics, 87: 331-340.

[46] Falconer D S and T F C (1996) Mackay, Introduction to Quantitative Genetics, Longman, Delhi, India.

[47] Yu D and Chu K H (2006) Genetic variation in wild and cultured populations of the pearl oyster Pinctadafucata from southern China. Aquaculture, 258: 220-227.

[48] Takeshi Takeuchi, TetsujiMasaoka, Hideo Aoki, Ryo Koyanagi, Manabu Fujie, Noriyuki Satoh (2020) Divergent northern and southern populations and demographic history of the pearl oyster in the western Pacific revealed with genomic SNPs. Evolutionary Applications, DOI: 10.1111/eva. 12905

[49] Arnaud, S, Monteforte, M, Galtier, N (2000) Population structure and genetic variability of pearl oyster Pinctadamazatlanica along Pacific coasts from Mexico to Panama. Conservation Genetics, 1: 299-308. doi: 10.1023/A:1011575722481

[50] Zhang L (2002) The development of international pearl industry and the counter measures taken by China to speed up Chinese pearl industry. Mar. Sci., 26: 10-13.

[51] Xiande Huang, Shanzeng WU, Yunyan Guan, Yaoguo LI, And Maoxian HE (2014) Identification of sixteen single-nucleotide polymorphism markers in the pearl oyster, Pinctadafucata, for population genetic structure analysis. Journal of Genetics, 93.

[52] Kumla S, Doolgindachbaporn S, Sudmoon R, Sattayasai N (2012) Genetic variation, population structure and identification of yellow catfish, *Mystusnemurus* (C&V) in Thailand using RAPD, ISSR and SCAR marker, Mol. Biol. Rep. 39(5):5201-5210.

[53] Neekhra B, Mansoori AA, Verma S, Koiri RK, Jain SK (2014) RAPD-PCR based biomarker study in fish species (Family Cyprinidae) of Madhya Pradesh. Austin J. Mol& Cell Biol., 1 (1): 1-6.

[54] Mansoori AA, Khan H, Jain SK (2018) RAPD-PCR based biomarker study for molecular identification and polymorphism in fish species (Cyprinidae family). Trends in fisheries Research. 7 (3): 72-78.

[55] Kempter J, Kielpinski M, Panicz R, Pruffer K, Keszka S (2017) Development of the method for identification of selected populations of torpedo scad, Megalaspiscordyla (Linnaeus, 1758), using microsatellite DNA analyses. CELFISH project - Part 4. Food Chemistry, 221:944-949. DOI: 10.1016/j. foodchem.2016.11.070.

[56] FranceschA, EstanyJ, Alfonso L, and Iglesias M (1997) Genetic parameters for egg number, egg weight, and eggshell color in three catalan poultry breeds. Poultry Science, 76 (12): 1627-1631.

[57] Akbaş Y and Takma C (2005) Canonical correlation analysis for

studying the relationship between egg production traits and body weight, egg weight and age at sexual maturity in layers. Czech Journal of Animal Science, 50(4):163-168.

[58] Liu L B, Li DY, Zhao Z L, Liu YP, Wang Y, and Zhu Q (2012) Polymorphism of Prolactin Receptor gene and its association with egg production traits in Erlang mountainous chicken. Asian Journal of Animal and Veterinary Advances, 7(11): 1183-1190.

[59] Wu N, Zhu Q, Chen B, GaoJ, XuZ, and Li D (2017) High-throughput sequencing of pituitary and hypothalamic microRNA transcriptome associated with high rate of egg production. BMC Genomics, 18(1): 255.

[60] Zhu M and Zhao S(2007) Candidate gene identification approach: Progress and challenges. International Journal of Biological Sciences, 3(7): 420-427.

[61] Xu H Y, Wang Y, Liu Y P, Wang J W, and Zhu Q(2012) Polymorphisms and expression of the chicken POU1F1 gene associated with carcass traits. Molecular Biology Reports, 39(8):8363-8371.

[62] Zhang Y, Du H,Chen J, Yang G, and Zhang X (2008) Porcine growth differentiation factor 9 gene polymorphisms and their associations with litter size. Journal of Genetics and Genomics, 35(3):163-169.

[63] Chu M X, Wu Z H, Feng T et al. (2011) Polymorphism of GDF9 gene and its association with litter size in goats. Veterinary Research Communications, 35(6):pp. 329-336.

[64] VageD I, Husdal M, Kent M P, Klemetsdal G, and Boman I A (2013) A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep, BMC Genetics, 14:1.

[65] Hamid Reza Seyedabadi, Cyrus Amirinia, Nour Amirmozafari, Rasoul Vaez Torshizi and Mohammad Chamani(2010). African Journal of Biotechnology, 9(27):4175-4178.

[66] Hoffmann (2005) World's Poultry Science Journal, 61, doi 10.1079/WPS200449

[67] Zejiao Li, Biying Jiang, Baolong Cao, Zheng Zhang, Jiannan Zhang, Juan Li, Yan Huang, and Yajun Wang (2021) Characterization of Four Orphan Receptors (GPR3, GPR6, GPR12 and GPR12L) in Chickens and Ducks and Regulation of *GPR12* Expression in Ovarian Granulosa Cells by Progesterone. Genes , 12(4):489. doi: 10.3390/genes12040489

[68] Perini F, Cendron F, Rovelli G, Castellini C, Cassandro M, &Lasagna E (2021) Emerging Genetic Tools to Investigate Molecular Pathways Related to Heat Stress in Chickens: A Review. Animals: An open access journal from MDPI, 11(1): 46. doi: 10.3390/ani11010046

[69] Darnell JE (1997) STATs and gene regulation. Science. 277(5332): 1630-1635. doi:10.1126/science.277.5332.1630

[70] Argetsinger L, and Carter-Su (1996) Growth hormone signalling mechanisms: involvement of the tyrosine kinase JAK2. Horm. Res. Paediatr. 45(Suppl. 1): 22-24. doi:10.1159/000184823

[71] Ou J, Tang S, Sun D and Zhang Y (2009) Polymorphisms of three neuroendocrine-correlated genes associated with growth and reproductive traits in the chicken. Poult. Sci. 88(4): 722-727.

[72] Udy GB, Towers RP, Snell RG, Wilkins RJ, et al. (1997) Requirement of STAT5B for sexual dimorphism of body growth rates and liver gene expression. Proc. Natl. Acad. Sci. U. S. A. 94(14): 7239-7244. doi:10.1073/pnas.94.14.7239

Chapter 2

From Genes to Biomarkers: Understanding the Biology of Malaria Gametocytes and Their Detection

Daniel Kepple, Anthony Ford, Ebony Little, Gabrielle Kolesar, Beka R. Abagero, Ashley N. Blackwell, Swarnapali De Silva Indrasekara, Delenasaw Yewhalaw and Eugenia Lo

Abstract

Each year, approximately 230 million malaria cases and 400,00 malaria deaths are reported worldwide. Malaria is a life-threatening disease caused by *Plasmodium* parasites that are transmitted from one individual to another through the bites of infected female *Anopheles* mosquitoes. Malaria parasites replicate asexually in the human host, and, in each replication cycle, a portion of the asexual stages develops into sexual gametocytes that permit transmission. The proportion of infections that carries gametocytes and the infectivity of gametocytes are indicators of human-to-mosquito transmission potential. In *P. falciparum*, gametocytes appear 10–14 days after infection, whereas in *P. vivax* gametocytes appear simultaneously with asexual schizonts. Such difference in development not only increases the length of time that an individual is infectious, but also increases the likelihood of transmission before treatment. The conversion from asexual parasites to gametocytes is also highly variable between infections. Differences in age, host immune response, parasite genetic composition, density of red blood cells, presence of co-infecting parasite strains, and antimalarial drug use could affect gametocytes production. In *P. vivax*, the unique ability to produce hypnozoites, a dormant liver stage of the parasite, may allow gametocytes to be produced periodically from relapse and contribute to transmission. In this chapter, we will provide an overview of the biology of *Plasmodium* gametocytes, existing tools for gametocyte detection, and features of gametocyte genes. The biological insights and genetic findings are essential to developing better detection biomarkers and effective strategies to reduce transmission in malaria-endemic countries.

Keywords: *Plasmodium*, gametocyte, epidemiology, biomarkers, transmission

1. Introduction

1.1 Malaria epidemiology

Malaria is a mosquito-borne disease. In humans, malaria is primarily caused by six different *Plasmodium* parasite species including *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, *P. knowlesi*, and *P. simium* [1]. To date, there is an estimated 229 million cases of malaria worldwide each year resulting in approximately 409,000 deaths [2]. This has resulted in an estimated direct cost (treatments, premature death, and loss of household income) of $12 billion USD per year, with lost economic growth being many times greater [3]. The deadliest of these is *P. falciparum*, primarily located in the African continent [4], due to the high parasitemia levels and rapid growth [4]. *P. falciparum* utilizes several host receptors, including Band 3 and Glycophorin A, B, and C for red blood cell invasion [5], which contributes to high parasite loads in infected patients and spread across the ethnically diverse continent. Contrary to *P. falciparum*, *P vivax* primarily uses the Duffy antigen receptor found on the surface of reticulocytes for invasion, leading to lower parasitemia levels. *P. vivax* is globally widespread in Asia, South America, and northern and eastern parts of Africa [6]. This parasite can cause relapse from weeks to months after the initial infection [6].

Most *P. ovale* cases were reported in the tropics of western Africa [7] and is considerably less prevalent and possibly less fatal in comparison to *P. falciparum* and *P. vivax* [8]. *P. ovale* has two different subspecies: *P. ovale curtisi* and *P. ovale wallikeri*, both of which may produce dormant stage hypnozoites similarly to *P. vivax* [9, 10]. *P. ovale* has been shown to be transmitted across international borders from Africa to China [11] as well as to Korea [11, 12], in countries that were considered malaria free [12]. *P. malariae* is mostly reported in sub-Saharan Africa and southeast Asia. Similar to *P. ovale*, *P. malariae* is less common and less fatal. In sub-Saharan Africa, infections of *P. malariae* are often found to also contain *P. falciparum* [13]. Though less common, *P. knowlesi*, a primate parasite infecting long-tailed and pig-tailed macaques [14], has recently been reported to infect humans in Southeast Asia [15, 16] and is mainly transmitted by *Anopheles leucosphyrus* and *A. hackeri* [16–18]. *P. knowlesi* utilizes two gene families: DBL and RBP, to bind the Duffy antigen/chemokine receptor (DARC) to invade human erythrocytes [19]. Last but not least, *P. simium* is a common primate parasite that has recently been detected through molecular assays in humans from Atlantic Forest region of southeastern Brazil [20]. *P. simium* remains largely uncharacterized due to its novelty and rarity in humans, but is thought to have similar invasion mechanisms to *P. vivax* due to genetic and morphological similarities [20]. The presence of these less-common malarial species emphasizes the need for better detection tools and control measures especially in countries approaching the elimination phase.

1.2 Development and transmission of *Plasmodium*

The lifecycle of *Plasmodium* is divided into two stages: the human asexual reproduction stage (also known as the erythrocytic stage) and the mosquito sexual reproduction stage (known as the sporogonic cycle) [20], both of which are necessary for population growth and genetic diversification [19]. The parasite first enters the human body in the form of sporozoites through the infected salivary gland of an *Anopheles* mosquito during a blood meal. Within minutes, sporozoites infect the liver and begin asexual replication to form schizonts (a process called exoerythrocytic schizogony) within a few days to a few weeks [20]. Once matured, schizonts bust to release hundreds of merozoites into the blood stream [20]. *P. falciparum*, *P. ovale*, *P. malariae*, and *P. simium* infect mature erythrocytes [2, 19, 21] while *P. vivax* and *P. knowlesi* infect younger reticulocytes [2, 19, 22–24] to feast on

hemoglobin and further reproduce asexually [2, 19]. At this stage, the intracellular *Plasmodium* is considered as a ring stage parasite because the young parasite takes up hemoglobin in a single, large vacuole that looks like a ring. Once the hemoglobin is engulfed, the parasite then begins feeding and hemoglobin is acquired by endocytosis of erythrocyte cytoplasm within cytostomes, known as the trophozoite stage. After the hemoglobin is consumed, the trophozoite may either undergo schizogony to asexually reproduce and start the erythrocytic cycle with new merozoites or mature into a macrogametocyte (egg) or microgametocyte (sperm). Because human body temperature is considerably warm for *Plasmodium* sexual reproduction [25], both forms of gametocytes are taken up by an *Anopheles* mosquito where sexual reproduction occurs within the gut of the mosquito producing zygotes. Once matured, the zygote will become mobile and elongated, developing into ookinete to infect the midgut wall of the mosquito [26] and develop further into oocyst. Oocysts asexually divide and eventually rupture into sporozoites that travel to the salivary gland of the mosquito to infect a new human host [2, 19].

All six human malaria parasites require gametocytes to infect female *Anopheles* mosquito to reproduce sexually and continue development into sporozoites before infecting a new human host [27]. This has led to complications in the control of malaria transmission and gametocyte detection [28]. Furthermore, *P. vivax* and *P. ovale's* unique ability to form dormant stage hypnozoites complicates parasite clearance and can reintroduce old parasite strains into transmission reservoirs [2, 19]. Gametocytes are critical for malaria transmission and possible immune evasion in both the human and mosquito hosts [29, 30], emphasizing the importance of accurate detection. This chapter aims to provide a systemic review that highlights the complexity of *Plasmodium* gametocytes at the biological and genetic levels and current methods used to track and detect *Plasmodium* gametocytes. We further examine the capability of biomarkers used across *Plasmodium* species and provide new candidate biomarkers to further enhance detection protocols. Lastly, we examine the immunogenicity of gametocytes in both humans and mosquitoes.

2. Gametocytogenesis

2.1 Gametocyte commitment and development

Gametocytogenesis is the commitment of a *Plasmodium* parasite to produce male and female gametocytes through mitotic division during the trophozoite stage that involves multiple epigenetic and transcriptional regulations [31]. Changes in temperature, pH, and host age help stimulate gametogenesis, the emergence of the gametes in the mosquito mid-gut. A drop in temperature from 38°C to 20-26°C, the exposure to gametocyte activating factors and/or mosquito exflagellation factors, as well as a rise in pH can trigger gametogenesis [32–34]. Some merozoites, for reasons not fully understood, differentiate into the sexual forms of the parasites, the gametocytes. Upon ingestion by the mosquito, the decline in pH, a drop in temperature, and other mosquito derived factors such as xanthurenic acid can together activate gametocytes to transform within 5–10 minutes to male (microgamete) and female gametes (macrogamete) within the mosquito midgut. When gametocytes are taken up during a mosquito's blood meal, a number of factors including temperature, oxygen and carbon dioxide concentration, pH and exflagellation factor contribute to the maturation of gametocytes [35]. Some species, such as *P. ovale* in humans and *P. yoelii* in mice, can develop into morphologically distinct male and female gametocytes directly from hepatic merozoites for further transmission [7]. Gametocyte commitment is largely based on stress factors including high parasitemia, anemia,

drug treatments, and host immune responses [36–39]. Although no clinical symptoms are experienced during gametocytogenesis, this developmental stage is critical for sexual replication in the mosquitoes and subsequent infection of a new human host in the form of sporozoites. There is considerable variation in the development time among the different human *Plasmodium* species, ranging from 7 to 10 days after the initial establishment of asexual parasites for *P. falciparum* [40] and 7–15 days for *P. vivax* [41]. It is yet unclear about the time for gametocyte development in *P. knowlesi*, *P. malariae*, *P. ovale*, and *P. simium*. During this prolonged maturation period, gametocytes undergo five morphologically distinct stages (I-V) [42]. The immature gametocytes developed during stages I-IV sequester mostly in host tissues, particularly in bone marrow and spleen [43]. At stage V during maturation, gametocytes become more deformable and return in the blood circulation for uptake by a new mosquito host [44]. For proper sporogonic development to occur, both male and female gametes must be taken up during a blood meal as merozoites are incapable at forming both male and female gametocytes from the same schizont [45]. Most regulation is performed through RNA binding proteins, such as acetylation lowers binding efficiency 4 (ALBA4), that increases exflagellation events in an unknown manner [46, 47].

To date, most of our knowledge on gametocyte commitment is derived from *P. falciparum* due to a lack of viable culturing methods for *P. vivax* [48] and low prevalence of other *Plasmodium* species. Molecular processes including epigenetic regulation and histone posttranslational modifications play vital roles in gametocyte commitment [49–51]. A well-studied example of histone regulation in *P. falciparum* is H3K9me3, which is normally restricted to multigene families in subtelomeric regions [52], and recruits heterochromatin protein 1 (HP1) [53, 54]. Both proteins are strongly associated with the *AP2-G* locus, the primary regulator for gametocytogenesis [55, 56]. Previous studies have shown that a majority of asexual parasites have *AP2-G* silenced [47] and this gene silencing may be induced by histone deacetylase (Hda2) [57]. Depletion of HP1 through *GDV1*, another gene that regulates gametocyte production, was shown to increase schizont development into gametocytes by 50% [47, 58]. Further, the presence of Stabilization ligand Shield 1 (Shld1) may also increase transcription of *AP2-G* by stabilizing the protein complex for transcription and enhance gametocyte production. Although the molecular mechanisms of gametocytogenesis are well established in *P. falciparum*, there are currently no sensitive and reliable methods available in clinical settings for front-line detection of *in vivo* gametocytes due to strain polymorphisms, limiting gametocyte densities, and variations in the timing of gametocyte production and development.

2.2 Polymorphisms of gametocyte genes

Recently, advances in next generation sequencing provide new insights in our understanding of genetic variation and gene expression across different stages of *Plasmodium* species [59]. The mosquito immune system is a significant barrier for some *Plasmodium* isolates to infect the mosquito hosts. For example, in *P. falciparum*, highly expressed Pfs47 protein allows the parasites to become "invisible" and escape the immune system of the mosquito. This protein is expressed on the surface of female gametocytes and ookinetes. The African isolates of *P. falciparum* that are known to express Pfs47 have been shown to escape the *Anopheles gambiae* immune system by suppressing the Jun-N-terminal kinase (JNK) signaling and avoiding the induction of epithelial nitration [59]. The Pfs47 protein gene is highly polymorphic implying that the mosquito immune system may be the driving force for diversity, and that parasites with compatible Pfs47 haplotypes and/or escape immune evasion are preferably selected [60].

The innate immune response in the mosquito vector is mediated in most part by the hemocytes, which eliminate pathogens such as bacteria, fungi, and protozoa. *Anopheles* mosquitoes are known to have a complement C3-like protein called thioester-containing proteins (TEP). TEP of *An. gambiae* (AgTEP1) has been shown to initiate immune defense against *P. berghei* [61]. TEP1 facilitates the interaction between the parasite and hemocytes with late encapsulation that kill the parasite. Knockdown of the TEP1 gene renders genetically selected refractory *Anopheles* strain susceptible to infection and increases the infectivity rates [61]. However, *P. falciparum* has been shown to bypass the TEP-based defense mechanism using its 6-cystein protein P47-like [61]. In *P. berghei*, the P47-like protein is important for female gamete fertility, whereas in *P. falciparum*, it promotes the gametocyte-to-ookinete development and protects the ookinete from complement-dependent lysis [61]. Furthermore, infection of *An. gambiae* mosquitoes by ookinetes of *P. berghei* has been shown to module the mosquito's immune system by up-regulating expression of the peptide defensin and a putative gram-negative bacteria-binding protein and a TNF-α factor-like transcription factor (LL3) [61].

2.3 Expression of gametocyte genes

During the different stages of gametocyte development, various genes express differentially. For instance, *Pfs*16 expresses the earliest and highest in stage II gametocytes, the alpha-tubulin II gene shows maximum expression levels in both stage II and III gametocytes, *Pfs*230 expresses in stage III gametocytes, PfsMR5 in stage IV gametocytes, and *Pfs*28 in stage V gametocytes [62]. Such gene expression pattern is directly related to the activities involved in each specific stage. *Pfs*16 is known as a marker for detecting sexually committed ring stages parasites and it likely plays a key role in gametocyte maturation given its continual expression during the entire gametocyte maturation process [63]. Recent immunofluorescences assay identified a small population of schizonts that expressed the *Pfs*16 gene. *Pfs*16 mRNA increases in the asexual cycle before schizonts develop into stage I gametocytes, though translation is delayed until the onset of stage I gametocytogenesis [62]. This finding suggests that both *Pfs*16 transcription and translation may begin prior to invasion of a committed merozoite and development into a stage I gametocyte.

The female gametocyte specific gene *Pfs*25, which expresses solely in the mosquito, is a glycosylphosphatidylinositol-linked protein expressed on the surface of ookinetes [64]. This gene is used for the detection of stage V female mature gametocytes as well as in the quantification of gametocytes from field studies [63]. Because *Pfs*25 is solely expressed inside the mosquito hosts with limited immune selective pressure, sequence variation between isolates is relatively low [65]. Antibodies against *Pfs*25 have been shown to reduce oocyst production in *in vitro* membrane feeding experiments. Both high sequence conservation and antibody response make *Pfs*25 a leading target for transmission-blocking vaccine design. Apart from *Pfs*25, another family of gametocyte protein with cysteine-rich domains includes *Pfs*230 and *Pfs*48/45 that have also been targets of transmission-blocking vaccine development [66, 67].

*Pfs*230 expresses in both male and female gametocytes with a prodomain that is processed during gametocytogenesis and mediates red blood cell binding, specifically during oocyst development [68]. The Pfs230 protein appears on the surface of gametes as a complex with Pfs48/45 (a glycosylphosphatidylinositol (GPI) anchored protein [69]) and appears to be critical for gamete fusion. The double domains in Pfs230 reveal a structure resembling the surface antigen 1 (SAG1) protein with a double beta-sandwich structure found in *Toxoplasma gondii*, another apicomplexan parasite. The complete gene encoding the Pfs230 protein from different isolates of *P. falciparum* showed that 27 nonsynonymous polymorphic sites [70]. Among

them, the amino acids at eight polymorphic sites map to positions that point their side chains toward the surface of the protein [70]. Five of those eight sites map to a confined region on the same side of the beta-sandwich of the protein structure. Even though the distribution of the amino acids are over much of the length of the domain IV sequence, the polymorphic sites appear to outline, although loosely, a contiguous surface region of the model [70]. This structural feature may provide relevant sites with respect to the interaction of male and female gametes [70].

In both male and female gametocytes, *Pfs*230 is expressed without a known membrane anchor and appears as a complex with the Pfs48/45 proteins on the surface of gametes, though this colocalization is not required. Pfs230 has been observed on the surface of live macrogametes in the absence of Pfs48/45 [71]. Pfs48/45 is a cysteine rich surface protein that is vital in male gamete fertility. Sequence analyses of *Pfs*48/45 revealed that polymorphisms are rare for residues involved at the binding interface [72]. Furthermore, unlike other pre-erythrocytic blood stage antigens Pfs48/45 is less polymorphic. The number of synonymous substitutions per synonymous site exceeds the number of non-synonymous substitutions per non-synonymous site [73].

Recent clustering analyses indicated that there is a clear distinction in the expression levels between male and female gametocyte genes. Male and female gametocytes have been shown to be differentiated by their gene expression levels as early as stage III. The female gametes are generally separated by stage, indicating differentiating gene expression levels with respect to mid and late-stage female gametocytes [74]. For male gametocytes, PF3D7_1325200, a putative lactate dehydrogenase gene, is highly expressed [74]. Another male gametocyte gene PF3D7_1311100 is a putative meiosis-specific nuclear structure protein 1. This gene is essential for normal assembly of the sperm flagella in mice, suggesting that it may have a role in the male gamete development and exflagellation. Other genes including Pf3D7_1114000, Pf3D7_1122900, Pfg14–748, HAP2, and MAPK2 are found to be associated with male gametocytes, though both Pf3D7_1114000 and Pf3D7_1122900 are also expressed in a few female gametocytes [74].

3. Genetic markers for detection and prevention

3.1 Conventional genes used for gametocyte detection

Previous study has indicated that approximately 10% of *P. falciparum* and 60% of *P. vivax* infections had concurrent detectable low-density gametocytemia [75]. Gametocyte densities are typically lower than asexual parasite densities (**Figure 1A**) and directly associated with total asexual parasite densities (**Figure 1B**). Given the relatively low gametocyte densities, it is conceivable that many of these infections are submicroscopic and remain undetected in communities where malaria occurs, contributing to continuous transmission.

Molecular tests for diminutive amounts of gametocytes use reverse-transcription polymerase chain reaction (RT-PCR) to amplify RNA transcripts of gametocyte-specifically expressed genes. Compared to DNA-based assay, qRT-PCR of targeted RNA transcripts revealed higher sensitivity in detecting gametocytes of considerably low densities. For example, there are more than 10^6 copies of 18S rRNA transcript per cell but only 5 copies of the 18 s rRNA gene per genome [76]. The production of high transcript copies in parasite cells allow for greater detection limits. For *P. falciparum*, several targeted genes such as *Pfs*25, *Pfg*136, *Pfg*84, and *Pfg*17 are specific to female gametocytes. The lack of genes specific to male gametocytes could underestimate the total gametocytes in an infection and thus

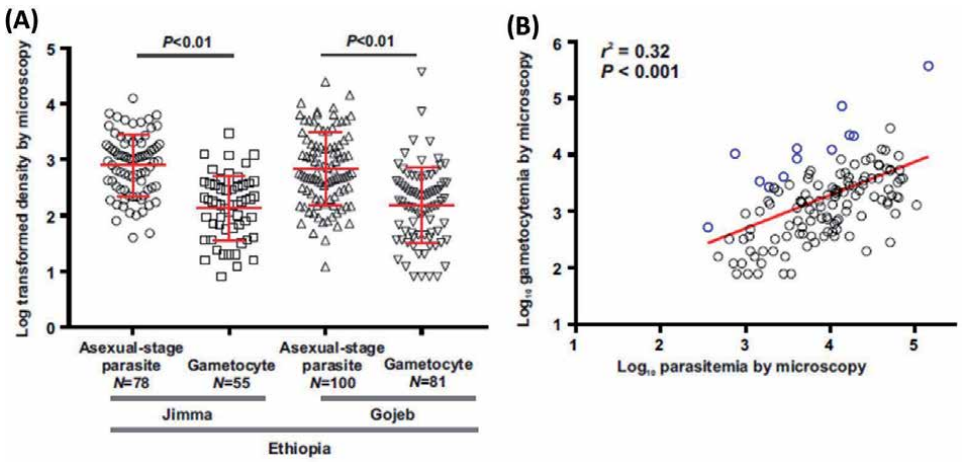


Figure 1.
(A) Comparison of asexual parasitemia and gametocytemia among P. vivax *infections from Ethiopia. (B) Significant correlation was observed between asexual parasitemia and gametocytemia.*

its transmission potential. Moreover, genes often show wide variations in detection limit based on the level of RNA transcript production, ranging from 100 gametocytes per microliter of blood by *Pfg*84, 70.7 gametocytes/μL by *Pfg*136, 25.3 gametocytes/μL by *Pfs*25 to as low as 10 gametocytes/μL by *Pfg*17 [77]. *Pfg*17 is considered as the most sensitive biomarker so far because it can detect the lowest number of gametocytes. For example, in malaria-endemic areas of Ghana, *Pfg*17 offers higher sensitivity than *Pfs*25 in detecting *P. falciparum* gametocytes in clinical samples collected from children and adults [61]. Of the 80 children, 47 were tested negative and 33 were positive for asexual blood stage *P. falciparum*. Nine out of the 47 (19%) negative children were detected positive for *P. falciparum* gametocytes by Pfg17. Among 30 adults aged from 32 to 60 years, 21 were (70%) tested positive for *P. falciparum* gametocytes by *Pfg*17, but only 7 (23.3%) were positive by *Pfs*25 [61].

For *P. vivax*, *Pvs* 25 and Pvs16 that are specific to female gametocytes are two conventional gene markers for gametocyte detection [78]. One gametocyte roughly corresponds to four *Pvs*25 transcripts per cell [79], and *Pvs*25 can detect from a mean of 0.34 gametocytes per μL blood from *P. vivax* patients in Papua New Guinea [78] to a mean of 2 gametocytes per μL blood in patients from Ethiopia [75]. Such low gametocyte densities make them extremely difficult to be detected by microscopy. The number of *Pvs*25 gene transcript copes detected by qRT-PCR directly correlates with the number of mature gametocytes as well as the overall parasite densities [41, 80] and was shown a nearly normal distribution with a mean of 1.2×10^7 copies/μL (ranging from 1.1 to 4.8×10^8 copies/μL) blood among 42 symptomatic *P. vivax* patients from northwestern Brazil [81]. Prior studies showed that age is tightly associated with gametocytemia. A lower proportion of infections with gametocytes was found with increasing age [79, 82]. Gametocytes are generally detected in ~20% of the infections among adults [41], but at much higher proportions in children under the age of 12 [82, 83]. Yet, gametocytemia in adults is up to 20-fold higher than in children [84, 85]. In areas with low levels of transmission, a large proportion of infections that are undetected by microscopy could be reservoir for parasites with high infectious gametocytes [86]. In Ethiopia, symptomatic *P. vivax* infections are nearly four times more infectious than asymptomatic ones [87]. Other factors such as host immune response, parasite strains, red blood cell density, antimalarial drug use, and relapse can also affect gametocyte production [79, 88, 89]. For *P. malariae* and *P. ovale*, there is yet no gametocyte assays due to little success in finding *Pfs*25 or *Pvs*25 orthologues in these species.

3.2 Novel gene candidates to improve detection sensitivity

Recently, the male-specific gene *pfs*13 that offers the lowest female to male expression ratio was identified as a new male gametocyte biomarker used in qRT-PCR assays from field isolates [90]. Other male-specific markers including PF3d7_1311100, PF3D7_1325200, and Pfg14–748 were also shown with abundant stage V male gametocytes in infected samples that made them as sensitive biomarker candidates [74], although their specific functions are unclear. Furthermore, *CCp*1, *CCp*3, and *P25* were identified as better biomarkers in differentiating female from male gametocytes [74, 91]. These genes are highly expressed in stages III to V female gametocytes. Though another gene *NEK*4 is female-specific, it is not sufficiently expressed in the transcriptome until stages IV and V, suggesting that it could be a good late-stage female biomarker candidate. PF3d7_1107800, a putative AP2 transcription factor, also strongly correlates as a female specific gametocyte gene similar to *NEK*4. PF3d7_1107800 is highly expressed only during stages IV and V, and thus, could potentially a biomarker for late-stage female gametocyte detection [74]. Besides, female gametocyte marker gene *CCp*4 was also identified as a new target for gametocyte detection [92]. The design of intron-spanning primers of these novel genes allows for the amplification of mRNA only without a DNA digestion step [92].

A recent study of 26 *P. vivax* samples from Cambodian patients indicated that the expression profile of 21 predicted gametocyte genes were clustered in two distinct groups [93]. One group includes *Pvs*25, *ULG*8, gametocyte developmental protein 1, guanylate kinase, *HMGB*1, and five CPW-WPC proteins that associate with intracellular trafficking and histone remodeling in the female gametocytes. The other group includes *Pvs*47, *Pvs*48/45. *Hap*2, the gamete egress and sporozoite traversal protein, s16, and three CPW-WPC proteins that associate with microtubular development in the male gametocytes. It remains to be determined if these male and female gametocyte genes show higher expression than the conventional marker *Pvs*25, offer high detectability of the total potentially transmitting gametocyte densities (both male and female gametocytes), and allow for robust gametocyte sex-ratio estimates in field studies given their stability under suboptimal storage conditions [92].

3.3 Treatment and prevention

In *P. falciparum* infections, gametocytes express surface antibodies against the surface antigens for circumsporozoite protein (PfCSP) 2A10, apical membrane antigen 1 (PfAMA1), and thrombospondin-related adhesive protein (TRAP). In *P. vivax* infections, Pvs48/45 and PfCLAG9 have been shown to elicit naturally acquired immune responses [65, 66]. Gene 1613 has been identified to be critical for the development and maturation of gametocytes, which could potentially impact the elicitation of humoral immune response and the ability of gametocytes to transmit [94]. These could be critical antigen proteins for vaccine design. As gametocytes mature from one stage to another, the production of antibodies in human hosts will cause gametocytes to develop and mature at a much slower rate than normal in the bone marrow [94–96]. Combining multiple antigens involved in different stages of gametocyte development and sexual progression can help interfere and diminish transmission [69]. Predicting the structure of the parasite antigen binding domains e.g., in AMA1 and Pfs48/45, and delineating epitopes targeted by the host antibodies could uncover key elements in blocking transmission and provide a benchmark for evaluating vaccine efficacy [70]. Vaccines that block the ability of gametocytes to transmit can reduce infections and lower the transmission potential from mosquitoes to humans in malaria-endemic areas [67]. Future studies should focus on expanding our understanding of gene interactions and their functions as well as disease control by way of transmission blocking vaccine

development [71]. Furthermore, uncharacterized chemotypes that possess activity against sexual parasites have also been recently identified to inhibit transmission to mosquito and human contact [68]. Interference by JmjC inhibitor known as ML324 causes expression to subsequently cease methylation activity followed by gametocyte lysis for *P. falciparum.* Identifying key surface proteins present in multi-parasite life rostrums could potentially be means of a preventative antimalarial against gametocytes [68].

4. Conclusions

To date, gametocytogenesis and gametocyte transmission tracking remain largely uncharacterized due to low prevalence and technological hurdles. As several countries approach the elimination phase for malaria, the need for sensitive and reliable biomarkers for gametocyte detection is more urgent than ever. Microscopy has low detection limit to overcome low parasitemia loads and existing qPCR biomarkers fail to accurately detect both sexes of the parasites. Information on genetic polymorphisms and expression levels of gametocyte biomarkers enable researchers to develop a more sensitive and accurate diagnostic test for *Plasmodium* gametocytes. Knowledge of gametocyte reservoirs and their interactions with host immune system will help develop effective treatment and preventive strategies to minimize the risk of malaria transmission. We recommend future studies focus on gene interactions and protein functions involved in gametocyte development, as well as polymorphisms in novel gametocyte genes that could be targets for developing better diagnostics or transmission blocking vaccines.

Acknowledgements

We thank our colleagues at the University of North Carolina at Charlotte and Jimma University for the inspiration and discussion that lead to the conceptualization of this paper. We also thank the reviewers for their insightful comments.

Author contributions

Conceptualization, D.K., A.F., and E.L.; resources, D.Y. and E.L.; writing—original draft preparation, D.K., A.F., E.L., G.K., D.Y., and E.L.; writing—review and editing, D.K., B.A., A.B., S.D.S.I., D.Y., and E.L.; funding acquisition, E.L. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by NIH/NIAID, grant number R15 AI138002.

Conflict of interest

The authors declare no conflict of interest.

Author details

Daniel Kepple[1*], Anthony Ford[2], Ebony Little[1], Gabrielle Kolesar[1], Beka R. Abagero[3], Ashley N. Blackwell[4], Swarnapali De Silva Indrasekara[4], Delenasaw Yewhalaw[3] and Eugenia Lo[1,5*]

1 Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

2 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

3 Tropical Infectious Disease Research Center, Jimma University, Jimma, Ethiopia

4 Department of Chemistry, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

5 School of Data Science, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

*Address all correspondence to: dkepple@uncc.edu and eugenia.lo@uncc.edu

IntechOpen

References

[1] Mayxay, M., et al., *Mixed-species malaria infections in humans.* Trends Parasitol, 2004. **20**(5): p. 233-240.

[2] Muhseen, Z.T., et al., *Natural products for treatment of Plasmodium falciparum malaria: An integrated computational approach.* Comput Biol Med, 2021. **134**: p. 104415.

[3] Organization, G.W.H., *World malaria report 2020: 20 years of global progress and challenges*. 2020.

[4] Snow, R.W., et al., *The prevalence of Plasmodium falciparum in sub-Saharan Africa since 1900.* Nature, 2017. **550**(7677): p. 515-518.

[5] Jaskiewicz, E., et al., *Erythrocyte glycophorins as receptors for Plasmodium merozoites.* Parasit Vectors, 2019. **12**(1): p. 317.

[6] Dayananda, K.K., R.N. Achur, and D.C. Gowda, *Epidemiology, drug resistance, and pathophysiology of Plasmodium vivax malaria.* J Vector Borne Dis, 2018. **55**(1): p. 1-8.

[7] Okafor, C.N. and N.A. Finnigan, *Plasmodium Ovale Malaria*, in *StatPearls*. 2021, StatPearls Publishing Copyright © 2021, StatPearls Publishing LLC.: Treasure Island (FL).

[8] Baird, J.K., Purnomo, and S. Masbar, *Plasmodium ovale in Indonesia.* Southeast Asian J Trop Med Public Health, 1990. **21**(4): p. 541-4.

[9] Groger, M., et al., *A systematic review of the clinical presentation, treatment and relapse characteristics of human Plasmodium ovale malaria.* Malar J, 2017. **16**(1): p. 112.

[10] Robinson, L.J., et al., *Strategies for understanding and reducing the Plasmodium vivax and Plasmodium ovale hypnozoite reservoir in Papua New Guinean children: a randomised placebo-controlled trial and mathematical model.* PLoS Med, 2015. **12**(10): p. e1001891.

[11] Zhou, R., et al., *Characterization of Plasmodium ovale spp. imported from Africa to Henan Province, China.* Sci Rep, 2019. **9**(1): p. 2191.

[12] Shin, H.I., et al., *Diagnosis and Molecular Analysis on Imported Plasmodium ovale curtisi and P. ovale wallikeri Malaria Cases from West and South Africa during 2013-2016.* Korean J Parasitol, 2020. **58**(1): p. 61-65.

[13] Yman, V., et al., *Persistent transmission of Plasmodium malariae and Plasmodium ovale species in an area of declining Plasmodium falciparum transmission in eastern Tanzania.* PLoS Negl Trop Dis, 2019. **13**(5): p. e0007414.

[14] Singh, B. and C. Daneshvar, *Human infections and detection of Plasmodium knowlesi.* Clin Microbiol Rev, 2013. **26**(2): p. 165-184.

[15] Barber, B.E., et al., *World Malaria Report: time to acknowledge Plasmodium knowlesi malaria.* Malar J, 2017. **16**(1): p. 135.

[16] White, N.J., *Plasmodium knowlesi: The Fifth Human Malaria Parasite.* Clinical Infectious Diseases, 2008. **46**(2): p. 172-173.

[17] Sallum, M.A., E.L. Peyton, and R.C. Wilkerson, *Six new species of the Anopheles leucosphyrus group, reinterpretation of An. elegans and vector implications.* Med Vet Entomol, 2005. **19**(2): p. 158-199.

[18] WHARTON, R.H. and D.E. EYLES, *Anopheles hackeri, a vector of Plasmodium knowlesi in Malaya.* Science, 1961. **134**(3474): p. 279-280.

[19] Milner, D.A., Jr., *Malaria Pathogenesis.* Cold Spring Harb Perspect Med, 2018. **8**(1).

[20] Kori, L.D., N. Valecha, and A.R. Anvikar, *Insights into the early liver stage biology of Plasmodium.* J Vector Borne Dis, 2018. **55**(1): p. 9-13.

[21] Sterling, C.R., et al., *Erythrocyte membrane alterations induced by Plasmodium simium infection in Saimiri sciureus: relation to Schüffner's dots.* J Parasitol, 1975. **61**(2): p. 177-188.

[22] Golassa, L., et al., *The biology of unconventional invasion of Duffy-negative reticulocytes by Plasmodium vivax and its implication in malaria epidemiology and public health.* Malar J, 2020. **19**(1): p. 299.

[23] Meyer, E.V., et al., *The reticulocyte binding-like proteins of P. knowlesi locate to the micronemes of merozoites and define two new members of this invasion ligand family.* Mol Biochem Parasitol, 2009. **165**(2): p. 111-121.

[24] Semenya, A.A., et al., *Two functional reticulocyte binding-like (RBL) invasion ligands of zoonotic Plasmodium knowlesi exhibit differential adhesion to monkey and human erythrocytes.* Malar J, 2012. **11**: p. 228.

[25] Rossati, A., et al., *Climate, environment and transmission of malaria.* Infez Med, 2016. **24**(2): p. 93-104.

[26] Beier, J.C., *Malaria parasite development in mosquitoes.* Annu Rev Entomol, 1998. **43**: p. 519-543.

[27] Ngotho, P., et al., *Revisiting gametocyte biology in malaria parasites.* FEMS Microbiol Rev, 2019. **43**(4): p. 401-414.

[28] Basu, S. and P.K. Sahi, *Malaria: An Update.* Indian J Pediatr, 2017. **84**(7): p. 521-528.

[29] Kepple, D., et al., *Alternative Invasion Mechanisms and Host Immune Response to Plasmodium vivax Malaria: Trends and Future Directions.* Microorganisms, 2020. **9**(1).

[30] Bansal, G.P. and N. Kumar, *Immune Responses in Malaria Transmission.* Current Clinical Microbiology Reports, 2018. **5**(1): p. 38-44.

[31] Ngwa, C.J., T.F.d.*A. Rosa*, and G. Pradel, *The Biology of Malaria Gametocytes, Current Topics in Malaria*, in *Current Topics in Malaria*, A.J. Rodriguez-Morales, Editor. 2016: IntechOpen.

[32] Saini, E., et al., *Photosensitized INA-Labelled protein 1 (PhIL1) is novel component of the inner membrane complex and is required for Plasmodium parasite development.* Sci Rep, 2017. **7**(1): p. 15577.

[33] Carter, R., et al., *Plasmodium falciparum: an abundant stage-specific protein expressed during early gametocyte development.* Exp Parasitol, 1989. **69**(2): p. 140-149.

[34] Drakeley, C.J., et al., *Transmission-reducing immunity is inversely related to age in Plasmodium falciparum gametocyte carriers.* Parasite Immunol, 2006. **28**(5): p. 185-190.

[35] Sinden, R.E., *Infection of mosquitoes with rodent malaria*, in *The Molecular Biology of Insect Disease Vectors*, J.M. Crampton, C.B. Beard, and C. Louis, Editors. 1997, Springer, Dordrecht. p. 67-91.

[36] Talman, A.M., et al., *Influence of chemotherapy on the Plasmodium gametocyte sex ratio of mice and humans.* Am J Trop Med Hyg, 2004. **71**(6): p. 739-744.

[37] Gautret, P., et al., *Enhanced gametocyte formation by Plasmodium chabaudi in immature erythrocytes: pattern of production, sequestration, and infectivity to mosquitoes.* J Parasitol, 1996. **82**(6): p. 900-6.

[38] Smalley, M.E. and J. Brown, *Plasmodium falciparum gametocytogenesis stimulated by lymphocytes and serum from infected Gambian children.* Trans R Soc Trop Med Hyg, 1981. **75**(2): p. 316-317.

[39] Trager, W. and G.S. Gill, *Enhanced gametocyte formation in young erythrocytes by Plasmodium falciparum in vitro.* J Protozool, 1992. **39**(3): p. 429-432.

[40] Gardiner, D.L. and K.R. Trenholme, *Plasmodium falciparum gametocytes: playing hide and seek.* Ann Transl Med, 2015. **3**(4): p. 45.

[41] Bousema, T. and C. Drakeley, *Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination.* Clin Microbiol Rev, 2011. **24**(2): p. 377-410.

[42] Hawking, F., M.E. Wilson, and K. Gammage, *Evidence for cyclic development and short-lived maturity in the gametocytes of Plasmodium falciparum.* Trans R Soc Trop Med Hyg, 1971. **65**(5): p. 549-559.

[43] Joice, R., et al., *Plasmodium falciparum transmission stages accumulate in the human bone marrow.* Sci Transl Med, 2014. **6**(244): p. 244re5.

[44] Tibúrcio, M., et al., *A switch in infected erythrocyte deformability at the maturation and blood circulation of Plasmodium falciparum transmission stages.* Blood, 2012. **119**(24): p. e172-e180.

[45] Silvestrini, F., P. Alano, and J.L. Williams, *Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum.* Parasitology, 2000. **121 Pt 5**: p. 465-471.

[46] Muñoz, E.E., et al., *ALBA4 modulates its stage-specific interactions and specific mRNA fates during Plasmodium yoelii growth and transmission.* Mol Microbiol, 2017. **106**(2): p. 266-284.

[47] Trelle, M.B., et al., *Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum.* J Proteome Res, 2009. **8**(7): p. 3439-3450.

[48] Gunalan, K., et al., *Plasmodium vivax Infections of Duffy-Negative Erythrocytes: Historically Undetected or a Recent Adaptation?* Trends in Parasitology, 2018. **34**(5): p. 420-429.

[49] Josling, G.A., K.C. Williamson, and M. Llinás, *Regulation of Sexual Commitment and Gametocytogenesis in Malaria Parasites.* Annu Rev Microbiol, 2018. **72**: p. 501-519.

[50] Coetzee, N., et al., *Quantitative chromatin proteomics reveals a dynamic histone post-translational modification landscape that defines asexual and sexual Plasmodium falciparum parasites.* Sci Rep, 2017. **7**(1): p. 607.

[51] Miao, J., et al., *The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation.* Gene, 2006. **369**: p. 53-65.

[52] Lopez-Rubio, J.J., L. Mancio-Silva, and A. Scherf, *Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites.* Cell Host Microbe, 2009. **5**(2): p. 179-190.

[53] Flueck, C., et al., *Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors.* PLoS Pathog, 2009. **5**(9): p. e1000569.

[54] Fraschka, S.A., et al., *Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites.* Cell Host Microbe, 2018. **23**(3): p. 407-420.e8.

[55] Bechtsi, D.P. and A.P. Waters, *Genomics and epigenetics of sexual commitment in Plasmodium.* International Journal for Parasitology, 2017. **47**(7): p. 425-434.

[56] Campbell, T.L., et al., *Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite.* PLoS Pathog, 2010. **6**(10): p. e1001165.

[57] Coleman, B.I., et al., *A Plasmodium falciparum histone deacetylase regulates antigenic variation and gametocyte conversion.* Cell Host Microbe, 2014. **16**(2): p. 177-186.

[58] Chawla, J., J. Oberstaller, and J.H. Adams, *Targeting Gametocytes of the Malaria Parasite Plasmodium falciparum in a Functional Genomics Era: Next Steps.* Pathogens, 2021. **10**(3).

[59] Ramphul, U.N., et al., *Plasmodium falciparum evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells.* Proc Natl Acad Sci U S A, 2015. **112**(5): p. 1273-1280.

[60] Molina-Cruz, A., et al., *Plasmodium evasion of mosquito immunity and global malaria transmission: The lock-and-key theory.* Proc Natl Acad Sci U S A, 2015. **112**(49): p. 15178-15183.

[61] Kengne-Ouafo, J.A., et al., *Immune Responses to the Sexual Stages of Plasmodium falciparum Parasites.* Front Immunol, 2019. **10**: p. 136.

[62] Eksi, S., A. Suri, and K.C. Williamson, *Sex- and stage-specific reporter gene expression in Plasmodium falciparum.* Mol Biochem Parasitol, 2008. **160**(2): p. 148-151.

[63] Jafari-Guemouri, S., et al., *Dynamics of Plasmodium falciparum gametocyte carriage in pregnant women under intermittent preventive treatment with sulfadoxine-pyrimethamine in Benin.* Malar J, 2018. **17**(1): p. 356.

[64] Scally, S.W., et al., *Molecular definition of multiple sites of antibody inhibition of malaria transmission-blocking vaccine antigen Pfs25.* Nat Commun, 2017. **8**(1): p. 1568.

[65] McLeod, B., et al., *Potent antibody lineage against malaria transmission elicited by human vaccination with Pfs25.* Nat Commun, 2019. **10**(1): p. 4328.

[66] Williamson, K.C., *Pfs230: from malaria transmission-blocking vaccine candidate toward function.* Parasite Immunol, 2003. **25**(7): p. 351-359.

[67] Theisen, M., M.M. Jore, and R. Sauerwein, *Towards clinical development of a Pfs48/45-based transmission blocking malaria vaccine.* Expert Rev Vaccines, 2017. **16**(4): p. 329-336.

[68] Eksi, S., et al., *Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production.* Mol Microbiol, 2006. **61**(4): p. 991-998.

[69] Kumar, N., *Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex.* Parasite Immunol, 1987. **9**(3): p. 321-335.

[70] Gerloff, D.L., et al., *Structural models for the protein family characterized by gamete surface protein Pfs230 of Plasmodium falciparum.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13598-13603.

[71] Singh, K., et al., *Structure and function of a malaria transmission blocking vaccine targeting Pfs230 and Pfs230-Pfs48/45 proteins.* Commun Biol, 2020. **3**(1): p. 395.

[72] Kundu, P., et al., *Structural delineation of potent transmission-blocking epitope I on malaria antigen Pfs48/45.* Nat Commun, 2018. **9**(1): p. 4458.

[73] Escalante, A.A., et al., *Polymorphism in the gene encoding the Pfs48/45 antigen of Plasmodium falciparum. XI. Asembo Bay Cohort Project.* Mol Biochem Parasitol, 2002. **119**(1): p. 17-22.

[74] Walzer, K.A., et al., *Single-Cell Analysis Reveals Distinct Gene Expression and Heterogeneity in Male and Female Plasmodium falciparum Gametocytes.* mSphere, 2018. **3**(2).

[75] Tadesse, F.G., et al., *The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia.* Malar J, 2017. **16**(1): p. 99.

[76] Nishimoto, Y., et al., *Evolution and phylogeny of the heterogeneous cytosolic SSU rRNA genes in the genus Plasmodium.* Mol Phylogenet Evol, 2008. **47**(1): p. 45-53.

[77] Essuman, E., et al., *A Novel Gametocyte Biomarker for Superior Molecular Detection of the Plasmodium falciparum Infectious Reservoirs.* J Infect Dis, 2017. **216**(10): p. 1264-1272.

[78] Wampfler, R., et al., *Strategies for detection of Plasmodium species gametocytes.* PLoS One, 2013. **8**(9): p. e76316.

[79] Koepfli, C., et al., *Blood-Stage Parasitaemia and Age Determine Plasmodium falciparum and P. vivax Gametocytaemia in Papua New Guinea.* PLoS One, 2015. **10**(5): p. e0126747.

[80] Bharti, A.R., et al., *Experimental infection of the neotropical malaria vector Anopheles darlingi by human patient-derived Plasmodium vivax in the Peruvian Amazon.* Am J Trop Med Hyg, 2006. **75**(4): p. 610-616.

[81] Lima, N.F., M.S. Bastos, and M.U. Ferreira, *Plasmodium vivax: reverse transcriptase real-time PCR for gametocyte detection and quantitation in clinical samples.* Exp Parasitol, 2012. **132**(3): p. 348-354.

[82] Nacher, M., et al., *Seasonal fluctuations in the carriage of Plasmodium vivax gametocytes in Thailand.* Ann Trop Med Parasitol, 2004. **98**(2): p. 115-120.

[83] Olliaro, P.L., et al., *Implications of Plasmodium vivax Biology for Control, Elimination, and Research.* Am J Trop Med Hyg, 2016. **95**(6 Suppl): p. 4-14.

[84] Dixon, M.W., et al., *Sex in Plasmodium: a sign of commitment.* Trends Parasitol, 2008. **24**(4): p. 168-175.

[85] Reece, S.E., R.S. Ramiro, and D.H. Nussey, *Plastic parasites: sophisticated strategies for survival and reproduction?* Evol Appl, 2009. **2**(1): p. 11-23.

[86] Hofmann, N.E., et al., *Assessment of ultra-sensitive malaria diagnosis versus standard molecular diagnostics for malaria elimination: an in-depth molecular community cross-sectional study.* Lancet Infect Dis, 2018. **18**(10): p. 1108-1116.

[87] Tadesse, F.G., et al., *The Relative Contribution of Symptomatic and Asymptomatic Plasmodium vivax and Plasmodium falciparum Infections to the Infectious Reservoir in a Low-Endemic Setting in Ethiopia.* Clin Infect Dis, 2018. **66**(12): p. 1883-1891.

[88] Baker, D.A., *Malaria gametocytogenesis.* Mol Biochem Parasitol, 2010. **172**(2): p. 57-65.

[89] Drakeley, C., et al., *The epidemiology of Plasmodium falciparum gametocytes: weapons of mass dispersion.* Trends Parasitol, 2006. **22**(9): p. 424-430.

[90] Santolamazza, F., et al., *Detection of Plasmodium falciparum male and female gametocytes and determination of parasite sex ratio in human endemic populations by novel, cheap and robust RTqPCR assays.* Malar J, 2017. **16**(1): p. 468.

[91] Bennink, S., M.J. Kiesow, and G. Pradel, *The development of malaria parasites in the mosquito midgut.* Cell Microbiol, 2016. **18**(7): p. 905-918.

[92] Meerstein-Kessel, L., et al., *A multiplex assay for the sensitive detection and quantification of male and female Plasmodium falciparum gametocytes.* Malar J, 2018. **17**(1): p. 441.

[93] Kim, A., et al., *Plasmodium vivax transcriptomes reveal stage-specific chloroquine response and differential regulation of male and female gametocytes.* Nat Commun, 2019. **10**(1): p. 371.

[94] Julien, J.P. and H. Wardemann, *Antibodies against Plasmodium falciparum malaria at the molecular level.* Nat Rev Immunol, 2019. **19**(12): p. 761-775.

[95] de Jong, R.M., et al., *Immunity against sexual stage Plasmodium falciparum and Plasmodium vivax parasites.* Immunol Rev, 2020. **293**(1): p. 190-215.

[96] Hall, C.E., et al., *Mosquito Bite-Induced Controlled Human Malaria Infection with Plasmodium vivax or P. falciparum Generates Immune Responses to Homologous and Heterologous Preerythrocytic and Erythrocytic Antigens.* Infect Immun, 2019. **87**(3).

Chapter 3

Thrombosis-Related DNA Polymorphisms

Nouha Bouayed Abdelmoula and Balkiss Abdelmoula

Abstract

Venous and arterial thrombosis are complex disorders involving several genetic inherited thrombotic and environmental risk factors as well as many mechanistic pathways including those of hemostatic, inflammatory and oxidative homeostasis. To provide an overview of genetic polymorphisms associated with thrombotic disorders, we studied related pathways and mechanisms of venous and arterial thrombosis along with their genetic polymorphisms in association with their clinical significance. We considered classical polymorphisms in the coagulation pathway factors, particularly the thrombophilia predisposition factors: Factor V, Prothrombin and MTHFR as well as PROC, PROS and antithrombin III. Other known and novel genetic polymorphisms having an impact on the pathogenesis of and the susceptibility to venous and/or arterial thrombotic disorders, in particular those involving inflammatory, immune and oxidant/antioxidant/redox signaling systems, were reviewed.

Keywords: ACE, Antioxidant systems, Arterial thrombosis, Coagulation pathway factors, Factor V, Genetic polymorphisms, Hemostatic systems, Leiden, MTHFR, Predisposition factors, Prothrombin, Venous thrombosis

1. Introduction

Thrombotic disorders and their related diseases, particularly cardiovascular diseases, are among the most common causes of morbidity and mortality in the world, causing a heavy burden on public health.

At the era of precision medicine ecosystem, omics technologies are increasingly being used to provide new molecular taxonomy of diseases and more precise approaches to assess risks, to predict and diagnose, and to monitor prognosis, therapeutic management and progression of these diseases.

During the genomic and the post-genomic phases and after the success of the Human Genome project, many genetic markers of complex common multifactorial diseases, in particular thrombosis and cardiovascular diseases have been tested using genetic association studies.

Multiple single and combined genetic variations and polymorphisms especially in the genes of coagulation and hemostasis pathways, as well as in the genes of inflammation and other genes interacting with lifestyle and environmental factors such as immune and oxidative systems were considered.

However, some of those variants remain under debate and clinical genotype–phenotype correlations continue to be uncertain.

To provide an overview of thrombosis-related DNA polymorphisms, we reviewed thrombosis related mechanisms as well as genetic variants associated with arterial and venous thrombosis and embolism and their clinical manifestations.

Classical polymorphisms in the hemostasis and coagulation pathways factors are reported in this first chapter. Besides, other types of genetic polymorphisms and variants having an impact on the susceptibility to venous and arterial thrombotic disease will be documented in a second future part.

2. Definitions and stratification of thrombotic disorders

2.1 Thrombosis vs. embolism

Thrombosis is defined as the formation, development, or presence of a blood clot, known as a thrombus, within a blood vessel that may be either a vein or an artery. The prefix "thrombo" come from the Greek "thrombos" meaning a lump or clump.

When thrombosis detaches and travels through blood vessels to another part of the body, it becomes an embolism.

Thrombosis has catastrophic complications by obstructing blood flow leading to ischemia and even infarction of the tissues supplied by the occluded blood vessels.

Embolism is often considered more dangerous than thrombosis because of its predilection to obstruct the entire blood vessel.

2.2 Arterial thrombosis vs. venous thrombosis

Usually, thrombotic diseases are classified according to their occurrence in the venous system of low flow and pressure or in the arterial system of high-flow and pressure.

Venous and arterial thrombotic disorders have long been viewed as separate pathophysiological entities partly because of the recognizable anatomical differences, despite the idea of a common pathogenesis of all thrombosis which is fundamentally the disturbance of hemostasis [1].

In fact, arterial thrombosis has long been apprehended to be a phenomenon of platelet activation, whereas venous thrombosis has been mainly held to be secondary to the activation of the clotting system [1].

Differences observed in the composition of the thrombi, which are platelet rich thrombi in arterial thrombosis and fibrin rich thrombi in venous thrombosis, and the presence of vascular wall damage in particular atheroma in arterial thrombosis reinforced this dichotomy in the concepts of arterial vs. venous thrombosis [2].

Arterial thrombosis involves the formation of white platelet-rich thrombi that occurs after the rupture of atherosclerotic plaques and the exposure of procoagulant material such as lipid-rich macrophages, collagen, tissue factor and/or endothelial breach, in a high shear environment [3].

In contrast, venous thrombosis is usually associated with plasma hypercoagulability and activation of the clotting system with expression of procoagulant factors on an intact endothelium. This activation is the result of the inflammatory process associated or not to a reduced blood flow or a stasis subsequent to prolonged immobility [1].

Actually, the distinctions are not absolute, and there are many evidences that arterial and venous thrombosis have many common underlying mechanisms involving biological factors either responsible for activating coagulation or inflammatory pathways in both the arterial and the venous systems [4].

Moreover, it was shown that patients with venous thromboembolism are at a higher risk of arterial thrombotic complications than matched control individual supporting the interplay between venous and arterial thrombosis pathogenesis [5].

2.3 Arterial vs. venous thrombotic related diseases

Arterial and venous thrombosis and embolism are associated to a variety of diseases and clinical manifestations such as systemic arterial thrombosis or embolism, ischemic strokes and acute infarction as well as superficial vein thrombosis and acute peripheral venous phlebitis, deep vein thrombosis and acute pulmonary embolism [6]. Obstetrical and placental thrombosis is another clinical presentation of thrombotic disease.

2.4 Coagulopathies vs. thrombophilia thrombotic related states

The definition of the term coagulopathies is controversial. They are various conditions in which the aptitude of blood to clot is impaired. However, the term of coagulopathies is used by many health-professionals to design thrombotic states and disorders of coagulation [7].

The definition of the term thrombophilia is more consensual and refers to inherited defects leading to enhanced coagulation, especially of the venous system. On the other hand and for many authors, thrombophilia may be, inherited or acquired, and the hypercoagulability state may arise from an excess or hyperfunction of a procoagulant or a deficiency of an anticoagulant factor [8].

2.5 Microthrombosis vs. macrothrombosis

Microvascular thrombosis is defined by the occurrence of microthrombi within the microcirculation. This microthrombogenesis process is associated to various severe clinical diseases such as thrombotic thrombocytopenic purpura, disseminated intravascular coagulation, and antiphospholipid syndrome as well as other thrombotic microangiopathies. It is also observed during systemic infections, cancer, myocardial infarction, stroke and neurodegenerative diseases [9].

Microvascular thrombosis often occurs subsequently to disordered clot formation and disordered inflammation pathway. Recently, during the coronavirus pandemic, it was shown that the novel severe acute respiratory syndrome coronavirus 2, is characterized by a dysregulated immune system and hypercoagulability recognized on the basis of profound d-dimer elevations and evidence of microthrombi and macrothrombi, both in venous and arterial systems. The complex crosstalking between the innate immune system and coagulation pathways culminates in the model of immunothrombosis, ultimately causing microthrombotic complications [9]. Microvessel thrombosis can then cause greatly differing symptoms that range from limited changes in plasma coagulation markers to severe multi-organ failure. Immunothrombosis is critically supported by neutrophil elastase and the activator molecules of blood coagulation tissue factor and factor XII. Identification of the biological driving forces of microvascular thrombosis should help to elucidate the mechanisms promoting pathological vessel occlusions in both microvessels and large vessels [10].

3. Thrombosis pathogenesis

Thrombosis is a pathologic phenomenon related to the disturbance of the dynamic balance of hemostasis. In fact, under normal circumstances, there is a fine balance between the procoagulant, anticoagulant and fibrinolytic pathways.

Hemostasis is the physiological mechanism aiming to protect the vascular system and to keep it intact after injury. The dynamic hemostatic balance comprising interactions between endothelial cells, thrombocytes, coagulation, and fibrinolysis prompts the regulation of hemostasis in order to assure the function of tissues and organs. This mechanism ensures the control of hemorrhage and thrombosis pathway activation and provide a matrix in wound healing and tissue repair. The amount of fibrin layers, at a site of injury inducing the progress of the tissue repair, is controlled by hemostasis balance [11].

When this equilibrium is disturbed under any condition, the physiologic process becomes pathologic leading to bleeding or to thrombotic troubles.

Numerous genetic, acquired and environmental factors can disturb the balance in favor of coagulation, leading to the pathologic formation of thrombi in veins, arteries, or cardiac chambers [1, 12].

The German pathologist Rudolf Virchow recognized that if this dynamic balance was altered by venous stasis, abnormal coagulability and vessel wall damages, microthrombi could propagate to form macroscopic thrombi. This understanding of thrombosis formation has been baptized as the triad theory in 1856 (**Figure 1**) [13]. Mechanisms of stasis, hypercoagulability, and endothelial dysfunction have been, subsequently elucidated as well as the different major factors involved in the hemostasis and coagulation cascade and fibrinolytic system, both at the biochemical and genetic levels [14].

Typically, arterial thrombotic disease is interrelated to atherosclerosis and thrombosis, as well as their interaction designed by the term atherothrombosis. Acute arterial thrombosis occurs at the site of a ruptured, lipid-rich atherosclerotic plaque. This event contributes to the transition of a stable atherosclerotic disease to an acute state [3]. Depending on the localization of atherosclerotic plaques, arterial thrombotic disease may be an acute infarction such as myocardial or brain infarctions, ischemic stroke, or peripheral arterial occlusion leading to ischemia. Arterial thrombosis can also be due to other pathological conditions favoring arterial clotting and turbulences such as atrial fibrillation and antiphospholipid syndrome [1, 3].

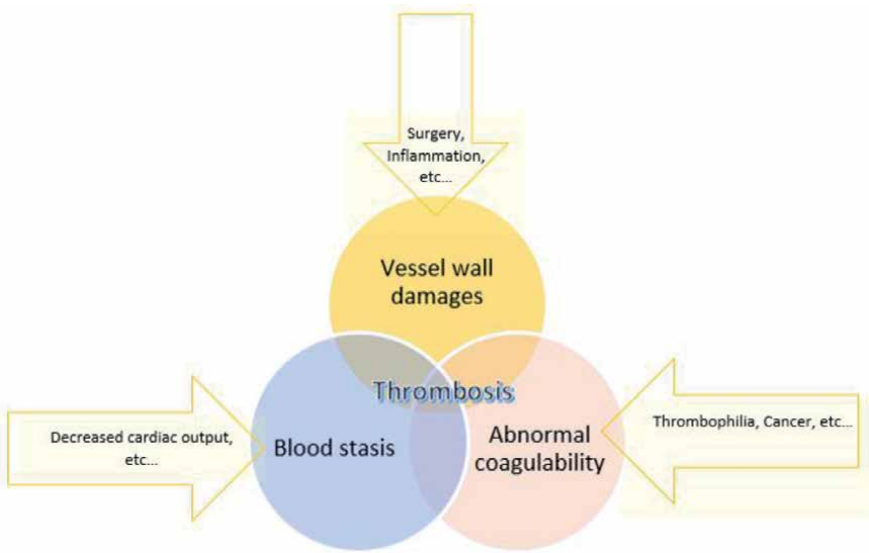


Figure 1.
The triad contributing together to venous thrombosis described by Rudolph Virchow.

In veins, Virchow's triad is traditionally invoked to explain pathophysiologic mechanisms leading to thrombosis. In fact, abnormalities in blood composition with plasma hypercoagulability, alterations in the wall components of blood vessel and changes of the blood flow with stasis, are the three components involved in the development of venous thrombosis [13]. Clinical manifestations of venous thrombosis include acute peripheral venous thrombosis (phlebitis) and deep venous thrombosis or venous thromboembolism as well as pulmonary embolism, the most serious acute complication of deep venous thrombosis. Long-term complications represented by post-thrombotic syndromes are due to damages affecting the valves in the veins. Venous thromboembolism and superficial vein thrombosis account for about 90% of venous thrombosis. Other rarer forms include retinal vein thrombosis, splanchnic vein thrombosis, cerebral venous sinus thrombosis, renal vein thrombosis, and ovarian vein thrombosis [15].

4. Thrombosis related pathways

The complex thrombosis related pathways crosstalking implies hemostasis and coagulation, inflammation and immune system as well as the contributing role of Redox homeostasis and the interplay of oxidative/nitrosative stress to both inflammation and coagulation [16, 17].

4.1 Hemostasis and coagulation pathways

Understanding, components and factors as well as steps of hemostasis and coagulation pathways, is important in defining the molecular variants related to the thrombosis pathogenesis [18].

Hemostasis encompasses the tightly regulated processes of blood clotting, platelet activation, and vascular repair. Hemostasis, which is the physiologic response to vascular endothelial injury, encompasses a series of processes to maintain blood within the vascular system through the formation of a clot. It involves three basic steps: vascular spasm, platelet clot formation, and coagulation, in which activation of the coagulation cascade clotting factors promotes the formation of a fibrin clot. Fibrinolysis is the process in which a clot is degraded [19].

Hemostasis can be divided into primary hemostasis and secondary hemostasis corresponding to the coagulation process. The fibrinolytic pathway called the tertiary hemostasis interacts to regulate fibrin deposition and removal during healing [18, 20].

Primary hemostasis consists of the formation of the platelet clot and includes the blood vessel constriction or vasoconstriction, platelet adhesion, activation and aggregation at the site of the vessel injury. Secondary hemostasis is characterized by the transformation of fibrinogen into fibrin and changes the platelet clot into a stable fibrin clot [20, 21].

During hemostasis, three distinctive pathways can be involved: intrinsic, extrinsic, and common pathways. Activation of the intrinsic pathway is promoted through exposed endothelial collagen, while activation of the extrinsic pathway is stimulated through tissue factor released by various cells in particular the endothelial cells after external damage. These pathways initiate separately at the beginning but at a specific moment with the presence of factor X, the Stuart-Prower factor, they converge, leading to common pathway with the generation of the prothrombinase complex that cleaves the prothrombin into thrombin and then fibrin activation process and platelet clot stabilization with a fibrin webbing. Coagulation cascade

involving conversion of inactive coagulation factors to their active forms following a series of enzymatic reactions including multiple cofactors and that ends with the conversion of fibrinogen to fibrin, leads to the formation of the definitive fixed blood clot with scrambled blood cells. The factors II, VII, IX, X, XI and XII circulate as zymogens that are activated into serine proteases to act as catalytic agents cleaving the next zymogen into more serine proteases, whereas factors V, VIII, XIII are not serine proteases [20–22].

To look at the multiple involved factors and actors in hemostasis pathways as well as their complex interactions; we reviewed recent literature reviews detailing the physiology of hemostasis and coagulation pathways. The biologic and molecular factors, cofactors and actors of the hemostasis pathways are summarized in **Table 1** and **Figure 2**.

At the cellular level, hemostatic reactions involves plasma, platelet, and vascular components. After a blood vessel injury the extracellular matrix and the collagen become unprotected and in contact with the blood within an area of vasoconstriction, leading to the liberation of cytokines and inflammatory markers. Consequently, platelet adhesion, activation and aggregation at that site are sequentially mediated by interactions between various receptors including tyrosine kinase receptors, glycoprotein receptors, other G-protein receptors and proteins within the platelets [23]. Platelet degranulation induces the liberation of Adenosine diphosphate, thromboxane A2, serotonin, and multiple other activation factors. The conversion of fibrinogen to fibrin and the formation of a platelet-fibrin hemostatic clot ends by a coagulation cascade involving the formation of fibrin polymer mesh catalyzed by activated factor XIII that stimulates the lysine and the glutamic acid side chains causing the cross-linking of the fibrin molecules and the formation of a stable fibrin clot [24].

The fibrinolytic pathway called the tertiary hemostasis interacts to regulate fibrin deposition and removal during healing. The activities of thrombin and other serine proteases are modulated by the serine protease inhibitors (serpins),

Clotting factors	Aliases names	Gene	Chromosome	Exons	Gene ID
Factor I	Fibrinogen	FGB	4q31.3	8	2244
Factor II	Prothrombin	FII	11p11.2	14	2147
Factor III	Thromboplastin -Tissue Factor	FIII	1p21.3	6	2152
Factor IV	Ionized calcium				
Factor V	Proaccelerin	FV	1q24.2	25	2153
Factor VI	Unassigned				
Factor VII	Proconvertin -Stable factor	FvII	13q34	10	2155
Factor VIII	Anti-hemophilic factor A	FVIII	Xq28	27	2157
Factor IX	Christmas factor	FIX	Xq27.1	8	2158
Factor X	Stuart-Prower factor	FX	13q34	8	2159
Factor XI	Plasma thromboplastin	FXI	4q35.2	15	2160
Factor XII	Hageman factor	FXII	5q35.3	15	2161
Factor XIII	Fibrin stabilizing factor	FXIIIA1	6p25.1	15	2162
		FXIIIB	1q31.3	12	2165

www.genecards.org and www.ncbi.nlm.nih.gov/gene/

Table 1.
Clotting factors and their encoding genes.

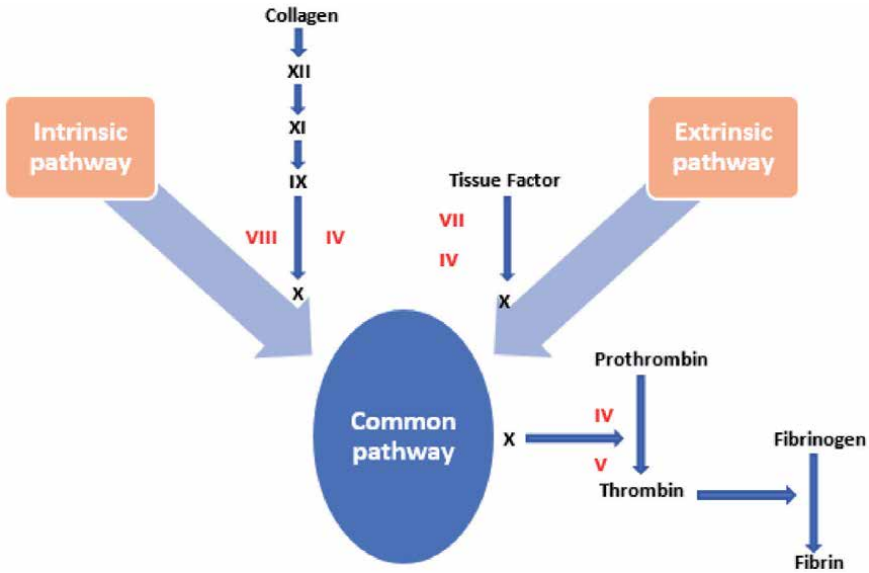


Figure 2.
The three distinctive pathways of hemostasis: intrinsic, extrinsic, and common pathways.

including antithrombin III and heparin cofactor II which are important in regulating the physiological anticoagulant action of glycosaminoglycans at the endothelium [25].

Under normal circumstances, there exists a fine balance between the procoagulant and anticoagulant pathway and hemostasis is under the inhibitory control of several inhibitors that limit clot formation, thereby avoiding thrombus propagation. This balance is disturbed whenever the procoagulant activity of the coagulation factors is increased, or the activity of naturally occurring inhibitors is decreased [14, 26]. As thrombin acts as a procoagulant, it also acts as a negative feedback by activating plasminogen (Serpine 1 Gene ID: 5054 7q22.1 with 9 exons) to plasmin and stimulating the production of antithrombin (Serpinc1 Gene ID: 462, 1q25.1 with 9 exons). Plasmin acts directly on the fibrin mesh and breaks it down. Antithrombin decreases the production of thrombin from prothrombin and decreases the amount of activated factor X. Protein C or blood coagulation factor XIV (PROC Gene ID: 5624, 2q14.3 with 8 exons) and protein S (PROS1 Gene ID: 5627 3q11.1 with 16 exons) act to prevent coagulation, mainly by inactivating factors V and VIII. The Kunitz-type protease inhibitor tissue factor pathway inhibitor (TFPI Gene ID: 7035, 2q32.1 with 13 exons) limits the diffusion of the coagulation cascade. TFPI binds to FXa or the TF-FVIIa-FXa complex to restrict coagulation function [https://www.genecards.org, https://www.ncbi.nlm.nih.gov/gene/].

Under abnormal circumstances, the formation of thrombi happens in a not breached vessel, in particular in venous thrombosis where thrombi are formed subsequently to the activation of the clotting system. However, in arterial thrombosis, thrombi are considered typically as the result of a phenomenon of atherothrombosis after rupture of atherosclerotic plaques leading to platelet activation and interactions between platelet activation, tissue factor vesicle expression from plaque macrophages, and then activation of the coagulation cascade [27].

4.2 Inflammatory and immune pathways

Understanding the role of inflammation in thrombosis disorders is important in defining the molecular variants related to the thrombosis pathogenesis.

The complex pathways of inflammation and hemostasis appear to have a common evolutionary origin and interrelated pathophysiologic processes [28].

In fact, there is an inflammation-hemostasis cycle in which each activated process promotes the other, and the two systems function in a positive feedback circle. The mechanisms responsible in the relations between thrombosis and inflammation involve all components of the hemostatic system including associated cells and plasma coagulation/fibrinolysis cascades [1, 28, 29].

The first event in thrombus formation is probably the stimulation of an inflammatory response with the activation of endothelial cells, platelets, and leukocytes. Initiation of inflammation leads to the formation of microparticles that activate the coagulation system through the induction of tissue factor. In fact, throughout the inflammatory response, various inflammatory mediators, in particular proinflammatory cytokines play a central role in ever-changing the hemostatic activity towards procoagulant state by triggering endothelial cell dysfunction, increased platelet reactivity, activation of the plasma coagulation cascade, impaired function of physiologic anticoagulants and inhibited fibrinolytic activity [30].

On the other side, coagulation cascade augments inflammation by means of thrombin-induced secretion of proinflammatory cytokines and growth factors. Platelets may also trigger inflammation, in particular by activating the dendritic cells. In abnormal circumstances, other inflammatory factors are implicated such as chemokines, adhesion molecules, platelet-derived mediators linking thrombosis and atherosclerosis and thrombosis, infection and immunity [28, 31].

Recently, there is a consensus that vascular thrombosis diseases are simultaneously, triggered by biological stimuli responsible for activating coagulation and inflammatory pathways in both the arterial and the venous systems [32]. In fact, while it is commonly recognized that the pathogenesis of arterial thrombotic disease is related to the chronic lipid-driven inflammatory disease of the arterial wall characterized by the involvement of the innate and adaptive immune systems or atherosclerosis, it is only recently that inflammation has been accepted as a common pathway of venous thrombosis formation [33, 34].

The most well described pathophysiologic process, in which there are an established relation between inflammation and hemostasis is the arterial atherothrombosis generated consequently to ruptured atherosclerotic plaque. Besides, chronic inflammation may cause endothelial damage, resulting in the loss of physiologic anticoagulant, antiaggregant and vasodilatory properties of endothelium. There are, many systemic inflammatory diseases characterized by thrombotic tendency in the absence of vessel wall damage, including chronic autoimmune diseases and vasculitis, such as Behçet disease, antineutrophilic cytoplasmic antibody-associated vasculitis, Takayasu arteritis, rheumatoid arthritis, systemic lupus erythematosus, antiphosholipid syndrome, familial Mediterranean fever, thromboangiitis obliterans and inflammatory bowel diseases [28, 35, 36].

Inflammation-induced venous thrombosis developed in the absence of vessel wall damage, in particular during malignancies, is also well demonstrated. Malignant proliferation induces prothrombotic substrates such as tissue factor and production of inflammatory cytokines, including tumor necrosis factor (TNF Gene ID: 7124, 6p21.33 with 4 exons) and interleukin-1 (IL1A Gene ID: 3552, 2q14.1 with 8 exons and IL1B Gene ID: 3553, 2q14.1 with 7 exons). This leads to the shift of the hemostatic state to procoagulant state that predisposes to the development of venous thrombosis [1, 37].

On the other hand, during the coordinated intravascular coagulation response of platelets in response to various blood pathogens and consequent tissue damage recently termed immunothrombosis, the risk of thrombosis that manifests as

arterial or venous thrombosis (and may contribute to atherosclerosis). During this process of immunothrombosis, inflammation-dependent activation of the coagulation system is part of the host response to pathogens, aiming to limit their systemic spread in the bloodstream [38, 39]. This response is achieved through an interplay between innate immune cells and platelets, triggering the activation of the coagulation system and the releasing of the complement system. Platelets and immune cells form, in fact, a physical barrier of confinement preventing dissemination of pathogens and potentially leading to activation of the innate and adaptive branches of the immune system [40]. Interestingly, platelets mediate the crosstalk between the hemostatic and the immune system utilizing similar pathways. The dysregulated and excessive activation of immunothrombosis results in thromboinflammation, causing tissue ischemia by microvascular and macrovascular thrombosis. Pulmonary immunothrombosis in severe COVID-19 correlating with a systemic prothrombotic phenotype provides clinical evidence for the partnership between inflammation and thrombosis [12, 41].

4.3 Oxidative pathways and redox homeostasis

There is now a strong evidence for the participation of reactive oxygen and nitrogen species in the pathogenesis of thrombosis as well as solid proofs of an interplay of oxidative and nitrosative stress, inflammation and thrombosis [42].

First, it is well established that reactive oxygen species (ROS) participate in vascular cell signaling and proatherogenic gene expression by modulation of oxidation–reduction (Redox) reactions pathways [43–45].

The cellular redox state, or balance between cellular oxidation and reduction reactions, serves as a vital antioxidant defense system that is linked to all important cellular activities. Redox homeostasis is thought to be achieved by careful regulation of both ROS formation and removal from the body system [46].

Recently, redox processes in cell signaling imply, beside ROS, Reactive Nitrogen Species (RNS). In fact, ROS and RNS were identified as key players in initiating, mediating, and regulating the cellular and biochemical complexity of oxidative stress either as physiological or as pathogenic processes [47].

Oxidative stress is a term associated with both enhanced production of ROS and reduced efficacy of protection by antioxidant enzymes. Nevertheless, after the discovery of NO as a biological entity and the powerful role of superoxide radicals O2- and NO as oxidants via ONOO- formation, oxidative stress has become unavoidably related to nitrosative stress and RNS. The key species associated with oxidative and nitrosative stress as well as their interactions have been reviewed and it was suggested that ROS are the initial reactants produced from an ionization event, whereas RNS are the effectors/activators of redox-dependent cellular signal transduction pathways [48, 49].

Depending on the severity of the oxidative stress, adaptive processes occur by increasing antioxidant capacities and by growing the capacity of the oxidative damages reparation. In extreme cases, metabolic processes shift away from oxidative metabolism towards glycolytic metabolism. In the case of chronic metabolic oxidative stress along with the accumulation of oxidative damage to critical biomolecules, potentially pathological conditions can develop due to the cumulative oxidative damage interacting with proteins, lipids, and DNA [48, 50].

An excessive ROS generation or a defect in the antioxidant defense system impacts a wide variety of biological molecules, lipids in the plasma and mitochondrial membranes, causing lipid peroxidation that impairs membrane selective permeability, proteins (resulting in structural instability and damage to their enzymatic function) and nucleic acids, thus inducing pathways of apoptosis [51].

In particular, oxidative stress is responsible for the disruption of the coagulation cascade at various stages leading to anomalies in blood coagulability and platelet reactivity. Nicotinamide adenine dinucleotide phosphate (NADPH) oxidase family (NOX) enzymes appears to be the most important source for ROS involved in processes related to thrombosis [52]. On the other hand, oxidative stress contributes to the development of atherosclerosis leading to atherothrombosis [53].

Oxidative pathways of thrombosis can occur throughout endothelial apoptosis, impairment of red blood cells quality and function, endothelial dysfunction and damage of endothelial cell lining, activation of platelets and leukocytes, and consequently by affecting the clotting system. The endothelial cell lining is essential in triggering the prothrombotic events as an intact endothelial cell lining prevents platelets adherence and activation. Furthermore, endothelial injury-associated oxidative stress promotes tissue factor expression having a potent procoagulant activity. In fact, ROS increases the expression of tissue factor in endothelial cells, monocytes and vascular smooth muscle cells, with essentially the contribution of NOX enzymes [54]. Whereas TFPI, which is the only physiologic regulator of tissue factor activity, can be inhibited by oxidative stress and exerts a procoagulant effect. ROS can also directly inactivate major anticoagulant proteins such as protein C and its upstream agonist thrombomodulin (THBD Gene ID: 7056, 20p11.21 with one exon). The stimulation of protease-activated receptors (PARs) may also lead to endothelial tissue factor induction via mitochondrial ROS signaling [55].

ROS stimulates platelet reactivity and during this ROS generation/platelet reactivity, platelet ROS are mostly generated by reduced NADPH oxidase. NOX2 expressed in platelets is an important regulator of platelet activation associated thrombosis. Engagement of primary platelet receptors, GPIb-IX-V and GPVI, that initiate thrombus formation, leads to a rapid increase in intracellular ROS above basal levels, and is a key step in platelet activation following exposure to physiological ligands such as von Willebrand factor (VWF)/collagen [55–57].

Thus, in contrast to endothelial injury-associated thrombus, platelet-dependent thrombus formation may also be influenced by alteration of platelet redox state or other cells or vascular redox state. Settings and pathways that influence the formation of superoxide and nitric oxide, as well as their metabolism, may specifically influence platelet function and thrombus formation [58].

Many studies emphasis that ROS influences venous thrombus formation and resolution through the modulation of the coagulation, the fibrinolysis, the proteolysis and the complement system, as well as the regulation of effector cells such as platelets, endothelial cells, erythrocytes, neutrophils, mast cells, monocytes and fibroblasts. During antiphosholipid syndrome for example, venous thrombosis occurs in patients having alterations in their redox homeostasis [42].

Reactive free radicals are defined as any chemical species capable of independent existence that contains one or more unpaired electrons. Reactive oxygen species (ROS) and Reactive nitrogen species (RNS) are free radicals that are associated with the oxygen atom (O) and other molecules, with stronger reactivity. Other biologically important free radicals exist such as lipid hydroperoxide (ROOH), lipid peroxyl radical (ROO), and lipid alkoxyl radical (RO), which are associated with membrane lipids and thiol radical (RS), which has an unpaired electron on the sulfur atom [59].

ROS consists of radical and non-radical oxygen species formed by the partial reduction of oxygen, including superoxide anion (O2-), hydroxyl radicals (OH), singlet oxygen (1O2) and hydrogen peroxide (H2O2). Generally, ROS are generated endogenously as natural by-products of aerobic metabolism during mitochondrial oxidative metabolic rate and by the means of cytochrome p450, cyclooxygenase, lipoxygenase and NOX enzymes. ROS are, also generated in response to stimulation such as by cytokines and other inflammatory mediators [59, 60].

Imbalance between oxidative stress and antioxidant status may be the result of an up-regulation of ROS-producing enzymes, such as NADPH oxidase and myeloperoxidase, along with down-regulation of antioxidant enzymes, such as superoxide dismutase (SOD) and glutathione peroxidase (GPx). Antioxidant defenses modulate the steady state balance of ROS with the implication of other several antioxidant enzymes such as catalase, glutathione peroxidases, heme oxygenase, thioredoxin system as well as small molecule antioxidants such as glutathione, vitamins A, C and E. These enzymes are produced to scavenge ROS, thereby limiting their detrimental effects [60, 61].

For example, age-dependent increased venous thrombosis is related to age-related endothelial dysfunction involving upregulation of the NADPH oxidase and cyclooxygenases (COXs)-dependent oxidative stress pathways. On the other hand, overexpression of the antioxidant enzyme GPX-1 protects from age-dependent increased venous thrombosis. Moreover, during aging, abnormal aged red blood cells may also adhere to the endothelium or extracellular matrix, activate platelets and other cells, and enhance local thrombin generation during thrombosis [52].

Second, in nitrosative stress, RNS involve various species such as nitric oxide (NO), nitrogen dioxide (NO2) and peroxynitrite (ONOO-). NO is generated in biological tissues by specific nitric oxide synthases (NOSs) and acts as an important oxidative biological signaling molecule in defense mechanisms and immune regulation. In the extracellular milieu, NO reacts with oxygen and water to form within an enzymatic cascade more reactive products. Immune cells, including macrophages and neutrophils, simultaneously release NO and superoxide into their phagocytic vacuoles. Other inflammatory cells can also produce reactive chemicals that can result in 3-NT formation, including the peroxidases in activated neutrophils and eosinophils. 3-NT is a characteristic marker of nitrosative stress and, commonly, inflammation [47].

NO was originally discovered as a vasodilator product of the endothelium and later as a factor having important antiplatelet actions, inhibitor effects of leukocyte adhesion and migration, and other inflammatory cells adhesion. By activating guanylyl cyclase, inhibiting phosphoinositide 3-kinase, impairing capacitative calcium influx, and inhibiting Cox-1, endothelial NO limits platelet activation, adhesion, and aggregation. Platelets are also an important source of NO, and this platelet-derived NO pool limits recruitment of platelets to the platelet-rich thrombus [62, 63].

A deficiency of bioactive NO is associated with arterial thrombosis in individuals with endothelial dysfunction and patients with a deficiency of the extracellular antioxidant enzyme GPx-3. Impaired NO availability also seems to be caused by inactivation of NOS and the levels of an endogenous inhibitor of NOS such as asymmetric dimethylarginine (ADMA). On the other hand, it seems that the decline in vascular NO production may be a characteristic feature of mammalian aging [47, 64].

Low NO availability is incriminated in thrombus formation. In fact, in the presence of low NO availability, endothelium-dependent vasodilation is impaired, which leads to abnormal red blood cells adhesion, and may contribute to increased platelet activation [65].

NO is considered as having a dual role as a protective or harmful molecule depending on tissue concentration levels and interaction with oxidative stress. In the endothelium, NO has antiatherogenic actions related to the inhibition of platelet function and inflammatory cell adhesion, promotion of fibrinolysis, and attenuation of smooth muscle cell proliferation. Oxidative stress and enhanced ROS production seem to be involved in the down-regulation of the protective NO pathway [48].

Finally, ROS/RNS generated during inflammation and inflammatory response and crosstalk between cellular redox state and the ROS/RNS network during inflammation constitute an emerging field. In states of inflammation, NO production by the vasculature increases considerably and, in conjunction with other ROS, contributes to oxidative stress [47].

ROS/RNS have also emerged as important modulators of intracellular transduction signaling. These radicals interact with redox-sensitive signaling molecules including protein tyrosine phosphatases, protein kinases and ion channels, regulating cellular processes like growth factor signaling, hypoxic signal transduction, autophagy, immune responses, and stem cell proliferation and differentiation. Moreover, the level of miRNAs can be modulated at the transcription and/or processing level by stress-induced factors like p53 or NF-kB as well as the presence of intracellular hydrogen peroxide levels [43, 44, 51, 66].

5. Thrombosis-related DNA polymorphisms

Among the many important insights derived from completion of the Human Genome Project was the recognition of the abundance of single nucleotide polymorphisms (SNPs) as a major source of genetic variation. Studies over the past last years have resulted in increasing recognition of the critical role of structural genetic variation in particular of copy-number variation in modulating gene expression and disease phenotype. Recently, genome-wide surveys and association studies are being widely applied to identify genetic factors that affect complex diseases or traits. However, while for SNP-association studies there are well-developed available resources, resources for structural genetic variation identified via genome-wide association studies are still in their early phases [67–69].

Arterial and venous thrombosis, with their clinical manifestations classified as complex multi-factorial diseases are related to various genetic variations that are both deleterious mutations and disease-susceptibility sequence polymorphisms, since the last century. In fact, beyond deficits in coagulation inhibitors, which have been known for a long time, the two most known thrombogenic mutations have been discovered in 1994 and 1996 [69].

Genetic studies in thrombosis started with the conception of the term thrombophilia by Jordan and Nandorff in 1956 [70]. Nine years later, antithrombin deficiency was identified as a genetic risk factor of thrombosis [71]. In the 1990s, activated PC resistance and the causal genetic variation of factor V Leiden, the first-born thrombogenic mutation, were revealed in a family setting. Factor V Leiden risk factor was also the first prothrombotic defect in a procoagulant protein. In 1996, case–control studies revealed the common prothrombin 20210 G > A mutation [72].

Using association studies and commonly PCR-RFLP tools, many genetic variations in almost all coagulation proteins were tested. However, until the beginning of the new century, only few gene loci were significantly associated with thrombotic diseases. The majority of the genetic variations was identified in the coding genes of hemostasis and coagulation factors and was shown as important risk factors for particularly venous thrombosis [73]. In contrast, in arterial thrombosis disease commonly related to atherosclerosis, many polymorphisms were identified, with significant relationships demonstrated between genotypes and plasma phenotypes. However, the exact contribution of genotypes to clinical phenotypes remains regularly uncertain. These variants associated in majority to a small risk, if any risk at all, have a limited usefulness as relevant biomarkers of the thrombotic diseases that must be prescreened to guide prevention, prognosis and treatment [74].

Nowadays, despite the success of genome-wide association studies in identifying new genetic factors determining many thrombosis-related diseases such as coronary artery disease, results for arterial and venous thrombosis have yielded little success. One of the reasons for the limited number of loci identified is most likely the lack of power due to the small sample sizes of studied cases [75].

Since the underlying pathogenic mechanisms are only partially known, regarding the complex interplay of many pathways in thrombosis-related pathogenesis as shown in the chapter 3 and the emerging role of intricate modulators of intracellular transduction signaling, there is mounting evidence indicating the challenging struggle of the mapping disease-susceptibility genes in thrombotic disorders. Genetic variations within thrombosis related pathways involving hemostasis and coagulation, inflammation, and immune system as well as Redox homeostasis and oxidative/nitrosative stress might be potential risk factors. Subsequently, many new loci need identification as risk factors as well as characterization in the future, through studies of candidate genes and genome-wide association studies.

6. Classic thrombophilia-related DNA polymorphisms

Thrombophilia traditionally refers to rare inherited defects leading to enhanced coagulation, especially of the venous system. Thrombophilia may be inherited or acquired. Acquired causes of thrombophilia include trauma, surgery, pregnancy, use of oral contraceptives, antiphospholipid syndrome, paroxysmal nocturnal hemoglobinuria and heparin induced thrombocytopenia. Inherited causes of thrombophilia are related to the hypercoagulability state that may arise from an excess or hyperfunction of a procoagulant or a deficiency of an anticoagulant factor [72].

In 1937, Nygaard and Brown first used the term thrombophilia, when they described sudden occlusion of large arteries, sometimes with coexistent venous thrombosis. Investigation of thrombophilia causes started by familial setting within families characterized by a predisposition to thromboembolic diseases and a strong tendency to venous thrombosis. This approach leaded to the early description of the deficiency of antithrombin causing venous thromboembolism at a young age related to thrombophilia entity in several members of a Norwegian family. Deficiencies of protein C and protein S were discovered after few years like the novel hereditary thrombophilia risk factors in other anticoagulant proteins. The inherited risk factors implying prothrombotic factors for thrombophilia were identified in following years like the underlying causes linked to venous thrombosis. They included FV Leiden variant linked to resistance to activated protein C (APC resistance) and FII 20210 G/A transition linked to elevated levels of Prothrombin (**Table 2**) [82, 83].

The factor V G1691A and prothrombin G20210A polymorphisms in arterial disease have been subjects of numerous reports. Many of these—some large—concerned with their association with arterial disease in young, middle-aged, and elderly populations had negative results. In contrast, some studies report positive associations, particularly when the interaction of these polymorphisms with environmental factors were formally evaluated [82, 83].

In 1988, Kang et al. [84] described a heat-labile form of MTHFR associated with mild hyperhomocysteinemia. The C677T polymorphism of the MTHFR gene has led to the identification of many more cases of thrombophilia.

Hereditary thrombophilia manifests more or less severely and early depending on the genotype, which may be heterozygous (a single affected allele), homozygous or composite heterozygous (two affected alleles) or associated with several different genetic risk factors.

Thrombophilia risk factor	Prevalence in the general population (%)	Mode of transmission	Reference
FV Leiden	2–15	Dominant	Bertina et al. [76]
FII G20210A	2–3	Dominant	Poort et al. [77]
MTHFR C677T	1–11	Recessive	Frosst et al. [78]
ATIII deficiency	0.02–2	Dominant	Egeberg [79]
PC deficiency	0.2–0.5	Dominant	Griffin et al. [80]
PS deficiency	0.1–2.1	Dominant	Comp and Esmon [81]

Table 2.
Prevalence and mode of transmission of inherited thrombophilia.

Association studies between inherited thrombophilia and venous thromboembolic disease showed dominance of factor V Leiden, and factor II G20210A variant in comparison with coagulation protein deficiencies (**Table 3**).

6.1 Factor V mutations

The factor V gene (Gene ID: 2153) is located on the long arm of chromosome 1 at q24.2. It consists of 25 exons that span a region of approximately 80. The FV gene encodes a propeptide of 2224 amino acids containing a 28-residue signal peptide, excised after translocation into the endoplasmic reticulum. This propeptide, called proaccelerin, is a monomeric protein of 330 kDa. It is characterized by a domain structure with 3 A domains (330aa), 2 C domains (150aa) and a B domain [N-A1-A2-B-A3-C1-C2--COOH]. Proaccelerin consists of two calcium-stabilized non-covalent chains: a heavy chain (110,000) and a light chain (74,000–71000). It is highly N- and O-glycosylated (about 13–25% of the mass, i.e. 37 N-glycosylation sites: 25 at the B-domain, 9 heavy chains and 3 light chains). Factor V exists in two forms V1 and V2, related to the heterogeneity of the molecular mass of the light chain and caused by the partial glycosylation at the Asn2181 residue (3 and 2 carbohydrate chains respectively for FV1 and FV2). These 2 forms differ in their functions: FV2a would have more affinity for membrane phospholipids and thrombin generation. Factor V is synthesized by hepatocytes and megakaryocytes. Approximately 80% of pro-accelerin is circulating in plasma (with a concentration of 5-10 mg/l = 21 nM) and only 20% is stored in the α-granules of platelets (i.e. between 4600 and 14000 molecules/platelet) in association with BPM (binding protein multimerin). This platelet fraction is released during platelet activation [91, 92].

Proaccelerin has a dual function: procoagulant and anticoagulant. Procoagulant action: Factor Va is part of the prothrombinase complex (Xa;Va;PL;Ca) that converts prothrombin (FII) to thrombin (FIIa), which is responsible for thrombus formation. It is initially activated by thrombin, which eliminates the B domain, and then its Xa cofactor binds to the C2 domain. At the molecular level, factor V performs its procoagulant function after proteolysis by thrombin and FXa at three arginine residues (Arg709, Arg1018, and Arg1545). This results in the elimination of the B domain. Thus FVa is formed by a heavy chain (105 KDa) [A1-A2] and a light chain (74–71 KDa) [A3-C1-C2] stabilized by calcium [92].

Once activated, FV complexed with FXa on membrane phospholipids and in the presence of Calcium forms the prothrombinase complex that converts prothrombin to thrombin. Anticoagulant action: Factor V binds to protein S and is a synergistic

	Thrombophilia n/N	Absence of thrombophilia n/N	OR (CI = 95%)
FV Leiden HETEROZYGOTE			
Dilley et al. [85]	8/9	32/107	18.75 (2.25–156.15)
Gerhardt et al. [86]	47/65	72/287	7.80 (4.26–14.28)
Martinelli et al. [87]	22/28	97/323	8.54 (3.36–21.73)
Murphy et al. [88]	3/16	29/556	4.19 (1.13–15.54)
Tormene et al. [89]	6/94	1/81	5.45 (0.64–46.29)
TOTAL (95%)	86/212	231/1354	8.94(2.32–50.79)
FII G20210A HETEROZYGOTE			
Dilley et al. [85]	4/4	36/112	18.86 (0.99–359.76)
Gerhardt et al. [86]	20/23	98/321	15.17 (4.41–52.24)
Martinelli et al. [87]	7/14	112/337	2.01 (0.69–5.87)
TOTAL (95%)	31/41	246/770	12.01(2.03–139.29)
MTHFR C677T HOMOZYGOTE			
Dilley et al. [85]	5/13	22/63	1.16 (0.34–3.99)
Murphy et al. [88]	1/57	9/223	0.42 (0.05–3.42)
Ogunyemi et al. [90]	2/12	28/48	0.14 (0.03–0.72)
TOTAL	8/82	59/334	0.57(0.42–2.71)
ATIII Deficiency			
Gerhardt et al. [86]	6/8	83/212	4.66 (0.92–23.65)
Martinelli et al. [87]	1/2	118/349	1.96 (0.12–31.58)
TOTAL	7/10	201/561	3.31(0.52–13.61)
PC Deficiency			
Gerhardt et al. [86]	15/24	91/312	4.05 (1.71–9.58)
(Ogunyemi et al. [90]	2/2	28/58	5.35 (0.25–116.31)
TOTAL	17/26	119/370	4.7(0.98–62.94)
PS Deficiency			
Gerhardt et al. [86]	13/24	92/309	2.79 (1.2–6.45)
Martinelli et al. [87]	2/3	117/348	3.95 (0.35–44.00)
TOTAL	15/27	209/657	3.37 (0.77–25.22)

Table 3.
Association studies between inherited thrombophilia and venous thromboembolic disease.

cofactor for inhibition of VIIIa by activated protein C. The anticoagulant action requires the inactivation of FVa. This is mediated by protein C which successively proteolyses it at Arg506, Arg306 and Arg 679. The first cleavage at Arg506 reduces the activity of FVa (25–40%) as well as its affinity for FXa, this partial inactivation is completed after cleavage at Arg 306, however Arg 679 is less important in this process. Thus, factor V will be fragmented into FVai (composed of the A1 domain associated with the light chain) and two fragments derived from the A2 domain (A2N and A2C respectively on the N and C terminal side). Alternatively, the inactivation of FVa is mediated by thrombin, which cleaves Arg643 in the presence

of endothelial cells, resulting in a reduction of affinity between the two heavy and light chains [91–93].

On the other hand, activated protein C can degrade intact FV and thus confers an anticoagulant property (FVac). Thus, the latter would be a cofactor of APC and protein S in the degradation of FVIIIa. This anticoagulant property requires cleavage at different sites (Arg306, Arg506, Arg679 and Lys994). However, only Arg 506 is required for the expression of the FV-APC cofactor activity [93]. This functional duality of FV in the coagulation process is dependent on the local concentration of procoagulant and anticoagulant enzymes such as thrombin, FXa and APC, which are responsible for the conversion of FV into a procoagulant or anticoagulant cofactor [94].

Various mutations affecting the FV gene have been described in association with a thrombotic phenotype. Among these missense mutations, the most prevalent is the Factor V Leiden mutation, initially described by Bertina et al. (**Table 4**) [76].

6.1.1 Factor V Leiden

Many studies have focused on the pathogenicity of VF in the occurrence of thrombophilia. Since its discovery in 1994 by Bertina et al. [76], FV Leiden represents the major anomaly in thromboembolic patients. It is a 1691G-A transition in exon 10 of the FV gene resulting in an Arg506-to-Gln substitution (R506Q) (**Figure 3**). This mutation is responsible for resistance to activated protein C, since it affects the potential site of cleavage by activated protein C, both for FVa degradation and FV-APC cofactor activity for FVIIIa inactivation [98, 99].

The thrombotic risk depends on the form of expression of the FV Leiden allele: it is 5–7 times in heterozygotes, 30 times in homozygotes and intermediate in pseudo homozygotes. The latter is a particular form, where the FVL allele is associated with another deficient or null allele [98].

Different mutations have been described in association with FV Leiden. The first mutation is a A4070G transition at exon 13 of the FV gene, resulting in a His1299Arg substitution, described in 1997 [100]. This mutation is an allelic form characterized by a moderate deficiency in Factor V not counterbalanced by FV Leiden; hence the occurrence of thrombosis.

In 1998, it was identified a null mutation that consists of a C2308T transition, at exon 13 of the FV gene affecting codon Arg 712(CGA), and producing a stop codon (TGA) resulting in a truncated protein at the level of its light chain (A3, C1, and C2 domains) unable to perform its anticoagulant function. As a result, only FV Leiden molecules are present and thus responsible for thrombosis [101].

The geographic distribution of FVL is extremely heterogeneous: it is absent in Asians, Africans, Americans, and Australians; however, it is prevalent in the Caucasian population. The existence of a single haplotype of FV Leiden worldwide suggests a single mutational event that occurred about 30,000 years ago, after the migration from Africa and the segregation of Asians from Europe. The age of factor V Leiden is estimated at 21,340 years [102].

Codon	AA	Name	Reference
ARG306GLY	1090A-G	FV HONG KONG	Chan et al. [95]
ARG306THR	1091G-C	FV CAMBRIDGE	Williamson et al. [96]
ILE359THR	1250 T-C	FV LIVERPOOL	Steen et al. [97]
Arg506Gln	1691G-A	FV LEIDEN	Bertina et al. [76]

Table 4.
Various mutations affecting the FV gene.

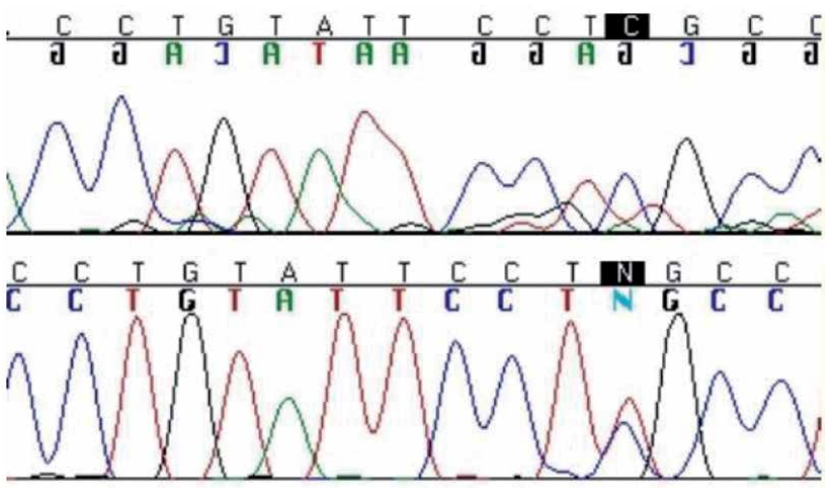


Figure 3.
Sanger sequence of the G1691A Factor V Leiden mutation in exon 10 of the FV gene.

FV Leiden is a remarkable genetic anomaly in more ways than one. It is common in the general population (2–15%) and affects 20–25% of patients with at least one episode of venous thromboembolism. Carriers of the anomaly are heterozygous in the vast majority of cases, which increases their risk of thrombosis by a factor of 5. However, homozygotes are not rare (0.05 to 0.25%) in the general population. They have a significant risk of venous thromboembolism (RR in the range of 20–30), but generally have no thrombotic events in childhood. They may remain asymptomatic, even in the homozygous state as the Factor V Leiden mutation has incomplete penetrance [103].

6.1.2 Factor V Hong Kong

In the Chinese population, it was identified a 1090A-G substitution in exon 7 of the FV gene resulting in an Arg306Gly (R306G) substitution and the mutation is named Factor V Hong Kong. It is one of the sites of cleavage by activated protein C that is affected, leading to a loss of the procoagulant activity of FVa. However, no predisposition to thrombosis was detected [95].

6.1.3 Factor V Cambridge

In 1998, at Addenbrooke's Hospital (Cambridge, England), Williamson et al. [96] identified a new mutation of the FV gene in a thrombophilic patient. It is a G to C transversion that results in an Arg306Thr (R306T) substitution. This mutation affects the APC cleavage site and is responsible for a loss of procoagulant activity of FVa. Unlike FV Hong Kong, FV Cambridge is associated with resistance to activated protein C.

6.1.4 Factor V Liverpool

In 2004, Steen et al. [97] identified a new mutation affecting the FV gene, called FV Liverpool. It is a 1250 T-C transition resulting from an Ile359Thr substitution (I359T). This mutation reduces the susceptibility of FVa to proteolysis by APC, since it alters the N- glycosylation at the asn357 residue of the A2 domain. Thus, the anticoagulant activity of FV is decreased and resistance to protein C is observed.

6.2 Factor II or prothrombin gene mutations

The factor II gene (Gene ID: 2147) is about 21 Kb long. It is located on chromosome 11, near the centromere (band 11p11.2). It has 20,241 bp, 14 exons of 25 to 315 bp and 13 introns of 84 and 9447 bp.

Factor II protein or prothrombin is one of the coagulation factors whose hepatocyte synthesis depends on vitamin K. It is a globular protein of about 72 kDa. It consists of a polypeptide chain of 579 amino acids, formed by functional domains found in several coagulation factors. The propeptide (residues −43 to −1), encoded by exons 1 and 2, is cleaved before secretion of the protein by hepatocytes. The Gla domain (residues 1 to 37), encoded by exon 2, is characteristic of vitamin K-dependent proteins: it contains 10 glutamic acid (Glu) residues that are converted to c-carboxy-glutamic acid (Gla) post-translationally by a vitamin K-dependent carboxylase present in the endoplasmic reticulum of the hepatocyte. Gla residues are involved in binding to anionic membrane phospholipids (mainly phosphatidylserine) in the presence of calcium. Two "Kringle" domains (Kringle 1: residues 65–143; Kringle 2: residues 170–248) are encoded by exons 5–6 and exon 7 respectively. Kringle 2 is involved in the binding of prothrombin to factor Va. The C-terminal part of prothrombin carries the serine protease domain, encoded by exons 8–14 [104].

Thrombin is the active form of factor II. It is composed of two polypeptide chains joined by a disulfide bridge: an A chain (residues 284–320) encoded by exons 8 and 9 and a B chain (residues 321–579) encoded by exons 9 to 14. Prothrombin synthesized in hepatocytes is found in the bloodstream. Prothrombin is activated by the prothrombinase complex into thrombin, a potential factor in the coagulation cascade. The prothrombinase complex consists of an enzyme (factor Xa, or FXa), a protein cofactor (factor Va, or FVa) and anionic phospholipids. FXa performs two cleavages at the Arg271/Thr272 and Arg320/Ile321 peptide bonds, giving rise to thrombin and activation peptides (F1, F2, F3) [104, 105].

Thrombin, a serine protease, is both multifunctional and highly selective. It has procoagulant properties, converting fibrinogen to fibrin, activating factor XIII, and amplifying its own formation (activation of platelets and factors V, VIII, XI). It is also anticoagulant, since by binding to thrombomodulin, it becomes capable of activating protein C, the negative regulator of coagulation. It activates not only platelets but almost all cell types (blood, vascular and non-vascular cells), thus intervening in many processes other than hemostasis (inflammation, angiogenesis, tissue remodeling, etc....) [106].

Various mutations have been described associated with the occurrence of thrombophilia (**Table 5**). These mutations affect both the regulatory and the splicing system and lead to thrombosis and hyperprothrombinemia respectively. Nevertheless, the G 20210A polymorphism first described by Poort et al. [77] is the most incriminated in thrombophilia.

6.2.1 G20210A In the 3'UTR region of the FII gene

The G20210A mutation affects the 3'UTR region of the FII gene. It leads to transcriptional efficiency through facilitated polyadenylation, resulting in increased prothrombin synthesis (https://www.ncbi.nlm.nih.gov/gene/2147).

The geographic distribution is due to a founder effect that dates back to 23,720 years [102]. The G20210A prothrombin mutation is quite common in the general population (2–3%). It is found in about 10% of patients with venous thrombosis, with carriers having a 3–4 fold increased risk. Homozygotes are rare and are likely to have a higher risk [103].

Localization	AA	Reference	Phenotype
3'UTR	C20209T	Warshawsky et al. [107]	Thrombosis
3'UTR	C20221T	Wylenzek et al. [108]	Thrombosis
3'UTR	G20210A	Poort et al. [77]	Thrombosis
	A19911G	Ceelie et al. [109]	Hyperprothrombinemia

Table 5.
Various mutations affecting the FII gene.

6.2.2 Other molecular polymorphisms of the FII gene

In 2001, Ceelie et al. [109] focused on the genetic causes of hyperprothrombinemia by analyzing variations in the prothrombin sequence in homozygous 20210-GG subjects. A homozygous 19911-G mutation is associated with an elevated prothrombin level. However, it does not affect thrombotic risk. On the other hand, this risk is potentially increased in association with the G20210A mutation (OR = 1.6 for 19911A versus 4.7 for 19911G). In the same year, Wylenzek et al. [108] detected a new mutational point in a Lebanese family: a C20221T substitution in the 3'UTR region of the F2 gene. In addition, in 2002, Warshawsky et al. [107] identified a mutation in four patients of African American origin presenting venous thrombosis. It is a C20209T substitution of the 3'UTR region of the FII gene [110].

6.3 Polymorphisms of the methylenetetrahydrofolate reductase MTHFR

The MTHFR gene (Gene ID: 4524) is located on chromosome 1 (1p36.22) and consists of 12 exons (https://www.ncbi.nlm.nih.gov/gene/4524). Several transcriptional start sites, alternative splicing and polyadenylation sites have been observed for MTHFR. Transcription start sites are located in two regions, and two promoters have been characterized. The latter contain multiple binding sites for transcription factors [111].

5,10-methylenetetrahydrofolate reductase (MTHFR) is a 150 kDa dimer comprising two isoforms of varying sizes: 77 kDa and 70 kDa. It consists of 656 AA and has two domains: catalytic on the N-terminal side and regulatory on the C-terminal side. The MTHFR protein is a homodimer with a βα structure. Each monomer is formed by 8 alpha helices and 8 beta sheets. It is the cofactor of the flavine adenine dinucleotide (FAD) [111].

MTHFR is a cytoplasmic enzyme found in the spleen, lymph nodes and bone marrow. Dimeric flavoprotein 5,10-methylenetetrahydrofolate reductase (MTHFR) is an NADPH-dependent enzyme that catalyzes the reduction of 5,10-MTHFR, the major carbon donor in nucleotide biosynthesis, to 5-MTHFR, which is the predominant form of folate and the donor of the methyl radical in the reaction of homocysteine remethylation to methionine [111].

Indeed, homocysteine is a sulfur-containing amino acid formed during the conversion of methionine to cysteine (demethylation). The catabolism of homocysteine follows two pathways: on one hand, transsulfuration (conversion to cystathionine and then to cysteine) involving cystathionine betasynthase (CBS) and, on the other hand, remethylation (regeneration of methionine) involving methionine synthase (MS) and methylene tetrahydrofolate reductase (MTHFR). These enzymes have as enzymatic cofactors certain vitamins of the B group (B6, folic acid, B12) [112].

The 5,10-methylene tetrahydrofolate reductase (MTHFR) catalyzes the irreversible reduction of 5,10-methylene tetrahydrofolate (CH2THF) to 5-methyltetrahydrofolate (CH3THF). MTHFR activity thus affects the availability of CH2THF,

which influences RNA and DNA synthesis. CH3THF is required for the remethylation of homocysteine to methionine (MET), which in turn is involved in protein synthesis and methylation of DNA and other compounds (CH3-X) [113].

Hyperhomocysteinemia may be both a genetic and acquired abnormality. Homocystinuria and hyperhomocysteinemia can be caused by rare inborn errors of metabolism that result in marked elevations of plasma and urine homocysteine concentrations [114].

Genetic polymorphisms result from common mutations that are usually ignored because they are often benign. However, some polymorphisms are not without health consequences. Two SNPs are described for MTHFR: C677T polymorphism and A1298C polymorphism of the MTHFR. Recently, rare variants in MTHFR have been detected by whole exome sequencing in association with the occurrence and the recurrence of pulmonary embolism [115].

6.3.1 C677T polymorphism of MTHFR

In 1995, Frosst et al. [78] identified a C to T substitution of nucleotide 677 that converts an alanine to a valine in 222. This mutation affects the catalytic domain of the MTHFR protein responsible for the generation of a thermolabile enzyme whose enzymatic activity is reduced by half at 37°C and absent at 46°C. This thermolability depends on the transmitted form: in homozygotes, the residual activity is only 18–22% while it is 56% in heterozygotes. On the other hand, the presence of the mutation in the homozygous state alters the metabolism of folates and induces a moderate increase in plasma homocysteine concentrations. The study of the biochemical characteristics of this thermolabile factor revealed a tendency to segregate into monomers as well as the dissociation of its cofactor FAD in solution [116].

The C677T mutation of MTHFR is a common polymorphism in the general population (allelic frequency is 0.38). Its frequency in the homozygous state varies between 1 and 21% with a significant heterogeneous distribution among different ethnic groups [117]. The homozygous TT genotype is particularly prevalent in northern China (18%), eastern Italy (18%) and California (21%). In addition, there is a geographical gradient in Europe (increase from north to south) and China (decrease from north to south). Furthermore, the genotypic frequency is low in African ancestors, intermediate in Europeans and prevalent in Americans [111].

6.3.2 A1298C polymorphism of the MTHFR

In 1998, Van der Put et al. [118] identified another SNP less frequent than C677T (allelic frequency is 0.33): it is a 1298A-C substitution that converts Glu429 to Ala (E429A) and thus destroys the restriction site for MboII. This mutation affects the regulatory domain of the protein and is associated with a more pronounced reduction in MTHFR enzymatic activity in homozygotes than in heterozygotes. However, the E429A protein is biochemically similar to the normal protein [116].

6.4 Protein C deficiency

Protein C (PROC), protein S (PROS1), and antithrombin (Serpinc1) have been demonstrated to play important roles in the anticoagulation process and thrombophilia [69, 72, 119].

PROC gene is located on chromosome 2 (2q14.3) and consists of 8 exons. Gene ontology annotations related to this gene include calcium ion binding and peptidase activity. An important paralog of this gene is PROZ (https://www.ncbi.nlm.nih.gov/gene/5624).

This gene encodes a vitamin K-dependent plasma glycoprotein, produced and secreted by hepatic cells as a zymogen. Along with its cofactor protein S (PS), activated protein C plays the role of an inactivator of the coagulation factors and an important factor of the regulation of the blood clotting pathway. It has a proteolytic effect on of the activated forms of coagulation factors V and VIII (Va and VIIIa). Protein C is a multi-domain glycoprotein composed of a non-catalytic light chain linked to the catalytic heavy chain by a single disulfide bond. The light chain harbors the vitamin K-dependent N-terminal g-carboxyglutamic acid (Gla) domain followed by two epidermal growth factor (EGF)-like domains. The C-terminal catalytic heavy chain with a trypsin-like substrate specificity is preceeded by an activation peptide, which is removed during the activation of protein C by the thrombin- thrombomodulin complex. The encoded protein C is cleaved to its activated form or APC by the thrombin-thrombomodulin complex. APC with the serine protease domain leads to the degradation of the coagulation factors Va and VIIIa in the presence of PS acting as a non-enzymatic cofactor, calcium ions and phospholipids. Protein C exerts also a protective effect on the endothelial cell barrier function [119–121].

Various conditions have been shown to cause acquired protein C deficiency. These conditions include vitamin K deficiency, warfarin therapy, severe liver disease, disseminated intravascular coagulation, severe bacterial infections in the young, and some chemotherapy drugs [122].

In contrast, inherited protein C deficiency is caused by genetic variations in the PROC gene. Prevalence of hereditary PROC deficiency is estimated at 0.2–0.5%. The milder form is caused by an alteration in one PROC gene and is inherited in an autosomal dominant manner. The severe form is caused by an alteration in both PROC genes (homozygous or compound heterozygotes) and is inherited in an autosomal recessive manner. In the other hand, heterozygous mutations in many adults may be asymptomatic for life but other heterozygous protein C deficiencies are characterized by recurrent venous thrombosis. Individuals with decreased amounts of protein C are classically referred to as having type I deficiency and those with normal amounts of a functionally defective protein as having type II deficiency [123].

PC deficiency is found in 3% of patients with primary venous thromboembolic disease. However, regarding the complex forms of inherited PC deficiency, studies of thrombophilic patients have shown that the prevalence of PC deficiency associated with thrombosis is between 1/16,000 and 1/36,000 [124]. A much higher prevalence of asymptomatic PC deficiency has been shown in a healthy blood donor population (1/200 to 1/700) [125].

Mutations in PROC gene have been long associated with thrombophilia with an increased tendency toward thromboembolic disease risk. In 1981, it was first described by Griffin et al. [80] that hereditary PROC deficiency was responsible of an hypercoagulability state. Hereditary PROC deficiency considered as autosomal dominant by familial studies arises from several distinct mutations in the PROC gene. PROC mutations leading to homozygous deficiency are detected during neonatal purpura fulminans [126].

Protein C database analysis suggests that there are about 380 mutations of PROC gene and that the mutations are scattered on both light and heavy chains and involve all functional domains of the protein (Gla, EGF1, EGF2 and catalytic domains), and recurrent venous thrombosis [127]. ClinVar database records that mention thrombosis and PROC showed 200 genetic variations. PROC variants occurring as the result of these genetic changes can lead to severe intracellular impairments and ineffective PROC release or non-functional PROC release.

PROC gene neighboring sequence contains several transcriptional regulatory regions. Distinct polymorphic loci were identified on promoter region of the human PROC gene. It was shown that polymorphic regions of the PROC gene: -1654C > T, -1641A > G and -1476A > T were associated with deep venous thromboembolism in some countries. Pulmonary embolism incidence in Chinese population seems to be associated with TT phenotype of -1654C > T polymorphism of PROC gene [128].

Recently, it was shown that PC as well as its PS cofactor are not only partners in the anticoagulant system, but also proteins closely involved in the mechanisms of inflammation, apoptosis, and in vascular permeability [129].

6.5 Protein S deficiency

PROS1 gene is located on chromosome 3 at 3q11.1 with 16 exons. There are two genes with 98% homology: an active gene with 15 exons spanning more than 80 kb and a non-coding pseudogene b, which is very close to the PSa gene (https://www.ncbi.nlm.nih.gov/gene/5627). This gene encodes Protein S (PS), the major cofactor of PC. It is a single-stranded, vitamin K-dependent glycoprotein of 69 kDa. PS is produced by the liver, but has also been localized in the endothelial cell, the megakaryocyte and the Leydig cell. It is synthesized as a 676 AA precursor comprising a leader sequence eliminated before secretion, a hydrophobic signal peptide, and a propeptide with the carboxylase recognition site analogous to that of other vitamin K-dependent factors. The mature form of PS (635 AA) consists of a GLA domain with 11 GLA residues, a binding peptide, a thrombin-sensitive loop (TSL), four EGF domains, and a carboxyterminal region with areas of homology to the hormone-binding globulin (SHBG) [130–133].

PS acts to prevent coagulation, mainly by inactivating factors V and VIII. PS increases the affinity of PCa for negatively charged phospholipids, forming a membrane-bound PCa-PS complex that makes factors Va and VIIIa more accessible to cleavage by PCa. PS circulates in the plasma partly under the influence of PCa. Free form (40% of circulating PS) is active in the coagulation system. Whereas, 60% is in the complexed form with C4b-binding protein (C4bBP), a protein of the complement system that binds PS at the SHBG domain. C4bBP-bound PS has no cofactor effect on PCa [131].

Other mechanisms of action independent of PC have been suggested for PS but their physiological importance is not firmly established; in particular, PS may have direct anticoagulant activity through its ability to bind and inhibit factors Xa, Va, and VIIIa and to compete with procoagulant factors for binding to phospholipids. It may also stimulate inhibition of the tissue factor pathway inhibitor or TFPI.

PS deficiency is found in 2–3% of thrombophilic patients. The prevalence in the general population may be in the range of 0.05% to 0.1% [130].

6.6 Antithrombin III deficiency

Antithrombin belongs to the serine protease inhibitor superfamily: the serpins. It exerts its physiological function by inhibiting procoagulation factors, such as thrombin, factor Xa, factor IIa, and other factors of the blood coagulation system. It contributes to the regulation of clot formation both by inhibiting thrombin activity directly and by interfering with earlier stages of the clotting cascade [134].

Antithrombin (AT) is a single-stranded plasma glycoprotein with a molecular weight of 58 kDa and 432 amino acids (AA) and four oligosaccharide side chains. AT is synthesized by the liver. It mainly inactivates thrombin and activated factor Xa, but also, in the presence of heparin, factors VIIa, XIa and XIIa [135].

Inactivation of the protease involves the formation of an irreversible bond between the active site of the enzyme and the reactive site of the inhibitor, formed by Arg 393 and Ser 394 (P1-P1'). The AT acts as a pseudosubstrate for the enzyme. Indeed, cleavage of the P1-P1' linkage induces a major conformational change in the AT, which can then form a stable complex with the target protease by incorporation of the AAs located upstream of Arg 393 into a b-sheet structure consisting of five strands in the uncleaved form and six strands in the cleaved form, with the sixth strand being the P1-P14 segment [136].

Inhibition of the enzyme by AT is catalyzed by heparin and proteoglycans of the vascular endothelium. This interaction accelerates thrombin inhibition by a factor of approximately 2000. In the presence of heparin, the AT reactive site loop is more exposed at the protein surface and more readily fits the catalytic site of certain activated factors such as factor Xa [137]. In the case of thrombin, which like AT has binding sites for heparin, a ternary complex is formed that brings the enzyme closer to its inhibitor. The heparin-binding domain of AT comprises the region of AA 41–49 on the one hand and AA 107–156 on the other. Both regions are rich in basic AAs that can interact with the sulfate groups of heparin. They are similar in the tertiary structure of the protein [138].

SERPINC1 is the gene encoding antithrombin. It is located on chromosome 1 at 1q25.1 with 9 exons. There are ten Alu sequences in the introns, representing 22% of the intronic sequences, four times more than in the entire human genome. These repetitive elements may contribute to the occurrence of many mutations and deletions in the gene (https://www.ncbi.nlm.nih.gov/gene/462).

Antithrombin deficiency, a rare autosomal dominant disorder (MIM#107300), is caused by rare genetic variations of SERPINC1 gene. There are two types of antithrombin deficiency. In type I antithrombin deficiency, functional and antigenic levels are proportionally decreased. In type II antithrombin deficiency, antigenic levels are normal while the functional activity is abnormal. In around 0.02–0.25% of a healthy population with antithrombin deficiency, there is a 5- to 50-fold increased risk of developing venous. AT deficiency is found in 1–2% of patients with primary venous thromboembolic disease. The prevalence of symptomatic AT deficiency in the general population is between 1:2000 and 1:5000 [139].

The first variation linked to antithrombin deficiency was characterized in 1983 and, to date, more than 200 variants have been reported to be associated with the risk of thrombosis. The homozygous variant (Phe229Leu) of SERPINC1 leading to spontaneous antithrombin polymerization in vivo has been shown to be associated with severe childhood thrombosis [140]. The heterozygous variant is mainly associated with a high risk of venous thrombosis [141].

However, most of SERPINC1 genetic variants (currently 399 different mutations reported) are rare, usually found in a single family baptized as private or orphan and occasionally discovered in more than one population. For example, in a Dutch population, Bezemer et al. [142] reported the 5301G > A polymorphism of SERPINC1 gene, to be associated with the risk of venous thrombosis. The frequencies of this rs2227589 polymorphism were around 0.10 and 0.329 in the East Asian population. In Spanish Caucasian population, it was shown the presence of a functional effect of the 5301G > A on antithrombin levels [143]. All of these conclusions were debated in different other studies [139].

Although thrombophilia can be identified in about half of all patients presenting with venous thrombosis, genetic testing or screening for hereditary thrombophilia is indicated only in selected cases [144, 145].

7. Conclusions

This chapter has focused on thrombosis-related genetic polymorphisms, particularly the variations of hemostatic genes involved in classical inherited thrombophilia diseases. They are the earliest and the most studied polymorphisms in the field.

Despite the increasing knowledge about thrombosis-related genetic polymorphisms, genetic testing for inherited thrombophilia remains considered, most often, not helpful to guide clinical decisions and not recognized on a routine basis.

The current knowledge of the contribution of thrombosis-related genetic polymorphisms showed an accumulation of understanding over the years for more than half a century that has led to robust results regarding their roles in thrombotic disorders and their potential clinical consideration, particularly as genetic markers of the diseases.

Despite their recognition as risk factors with well-established frequencies and sufficiently convincing associations, the implementation of the genetic testing as diagnosis/prognosis tools failed to attain an international consensus for clinical application. In fact, even though, genetic and genomic testing and screening are expected to have a greatly increased role in healthcare with a gradually likely to be ordered in routine for many diseases, genetic test reports in thrombosis miss the ability to deliver with the results, their clinical implications clearly and unambiguously. Guidelines and recommendations on thrombosis related genetic polymorphisms laboratory analysis remained limited to a narrow range of specific clinical situations and patients and are not uniform worldwide. The conditions under which genetic testing for thrombophilia have been in fact defined, were engaged, validated and published by some working groups and medical associations [144–153].

Furthermore, literature review showed that the AT, PC and, PS deficiencies, as well as FV Leiden, prothrombin mutation and MTHFR polymorphisms mentioned above, are considered as having an increased risk for venous thrombosis but have little or no effect on arterial thrombosis. In fact, the available evidence indicates that Leiden FV variant is not a major risk factor of any sort in arterial thrombosis and micro thrombosis, including myocardial infarction and strokes. The same conclusions were demonstrated for the other thrombophilia polymorphisms [154–156]. However, hyperhomocysteinaemia is still considered as a mixed risk factor for both arterial and venous thrombosis [1, 154].

Additionally, analysis of the literature revealed the description of several other polymorphisms that predisposes to the development of thrombosis, mainly those involving the hemostatic pathway factors like fibrinogen, factors VIII and VII, factor XIII, activated protein C receptor, thrombomodulin, plasminogen activator inhibitor, tissue plasminogen activator, Thrombin-activatable fibrinolysis inhibitor and platelet receptors (GPIIb-IIIa, GPIb-IX-V, GPIa-IIa, GPVI and others) etc... Two exhaustive reviews reporting these polymorphisms and especially their clinical significance were identified [146, 157].

In contrast, there is little clarity in relation to arterial thrombotic disease and the initial promise that genetic risk factors might contribute appreciably to an explanation of the development of arterial thromboses has largely been unfulfilled. As well, the expectations raised by early reports of positive associations have been tempered by inconsistent results with almost the majority of the studied hemostatic genes. In reality, the most consistent associations that have been found involve fibrinogen and the factor XIII [146]. Recently, most association studies of arterial thrombosis-related genetic polymorphisms are focused more much towards genes and factors involved in the other pathogenic pathways leading to thrombosis, i.e., inflammatory, immune and oxidative pathways.

Interrogation of the NCBI ClinVar database with an inquiry linking polymorphism and thrombosis revealed more than 3500 variations. The pathogenicity and the cause-and-effect relationship of these genetic variations was not strongly validated in the majority of cases. Most of related studies involved limited number of patients with untested statistical associations on a large scale. Additionally, some of the problems in identifying causal genetic markers are related to difficulties associated with the precise definition of the clinical phenotype under study. The sampling of patients in the studies was characterized by obvious heterogeneity both in terms of pathologies and in terms of the physiopathogenic origins of considered thrombosis. Indeed, there are many problematical uses of the terminology and other difficulties regarding the disorders stratification and the thrombosis typology. These difficulties of clinical and biological heterogeneity limit fundamentally the statistical homogeneity of the studied subgroups and the comparative effectiveness during the associative relationship approaches.

To overcome these shortcomings and to reach effective associations, several other technical and biological obstacles must be considered and defeated.

Translational medicine and research findings in the field of polymorphisms during thrombotic disorders are promising by the use of omics approaches and genome-wide association analysis, which will permit the identification of new risk loci. They will provide mechanistic insights into the genetic pathogenesis of thrombotic entities and put on view greater overlap among venous, arterial and microvascular thrombotic disorders than previously thought. The next step will be at the interactomic level to disclose binary and complex interactions between genes, factors and actors of thrombosis pathways. Indeed, it has been evident through this review that there are permanent interactions between the different pathogenic factors at the origin of thrombosis and almost permanent functional dualities for each factor under the influence of the dynamic homeostatic states of the organism and of the cells facing various situations to establish adaptive equilibriums. The importance of environmental influences and the complexity of the processes involved in vascular thrombotic disease suggest another myriad of other interacting metabolic factors. This understanding further increases the importance of lifelong risk interactions and may suggest an explanation for some of the inconsistencies in case–control studies. It is also important to consider that studies of population genetics of polygenic disorders, such as thrombotic disease, would ideally require prior knowledge of the relationship between the protein level/receptor density and disease, the degree of heritability of variance in the plasma levels of the protein/receptor density, and the genetic determinants of heritability.

It should be emphasized that despite the difficulties delaying the approval of genetic polymorphisms as reliable markers of thrombotic disorders, some of them have already been integrated into the preventive approaches of precision medicine, while others could be adopted in the near future in the predictive approaches of precision medicine.

Finally, it is important to mention that the most advantageous achievements of the studies on the associated thrombosis genetic variations are perceived through the pharmaceutical and biotechnological industries discovers in the field of thrombosis. These progresses on drug discoveries during the past, the present and the future are closely related to the deep understanding of the pathogenic mechanisms of thrombosis as well as their complex interplay.

Conflict of interest

Nouha Bouayed Abdelmoula and Balkiss Abdelmoula declare they have no conflict of interest.

Author details

Nouha Bouayed Abdelmoula* and Balkiss Abdelmoula
Genomics of Signalopathies at the Service of Medicine, Medical University of Sfax, Tunisia

*Address all correspondence to: nouha_abdelmoulabouayed@yahoo.fr

References

[1] Previtali E, Bucciarelli P, Passamonti SM, Martinelli I. Risk factors for venous and arterial thrombosis. Blood Transfus. 2011;9(2):120-138. doi:10.2450/2010.0066-10

[2] Chernysh IN, Nagaswami C, Kosolapova S, et al. The distinctive structure and composition of arterial and venous thrombi and pulmonary emboli. Sci Rep. 2020;10(1):5112. doi:10.1038/s41598-020-59526-x

[3] Badimon L, Padró T, Vilahur G. Atherosclerosis, platelets and thrombosis in acute ischaemic heart disease. Eur Heart J Acute Cardiovasc Care. 2012;1(1):60-74. doi:10.1177/2048872612441582

[4] Ashorobi D, Ameer MA, Fernandez R. Thrombosis. 2021 May 1. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 30860701.

[5] Abdol Razak NB, Jones G, Bhandari M, Berndt MC, Metharom P. Cancer-Associated Thrombosis: An Overview of Mechanisms, Risk Factors, and Treatment. Cancers (Basel). 2018;10(10):380. Published 2018 Oct 11. doi:10.3390/cancers10100380

[6] Morrone D, Morrone V. Acute Pulmonary Embolism: Focus on the Clinical Picture [published correction appears in Korean Circ J. 2018 Jul;48(7):661-663]. Korean Circ J. 2018;48(5):365-381. doi:10.4070/kcj.2017.0314

[7] Hunt BJ. Bleeding and coagulopathies in critical care. N Engl J Med. 2014 Feb 27;370(9):847-59. doi: 10.1056/NEJMra1208626. PMID: 24571757.

[8] Nascimento CMDB, Machado AMN, Guerra JCC, et al. Consensus on the investigation of thrombophilia in women and clinical management. Einstein (Sao Paulo). 2019;17(3):eAE4510. Published 2019 Aug 19. doi:10.31744/einstein_journal/2019AE4510 doi: 10.1056/NEJMra1208626. PMID: 24571757.

[9] Bray MA, Sartain SE, Gollamudi J, Rumbaut RE. Microvascular thrombosis: experimental and clinical implications. Transl Res. 2020;225:105-130. doi:10.1016/j.trsl.2020.05.006

[10] Pfeiler S, Massberg S, Engelmann B. Biological basis and pathological relevance of microvascular thrombosis. Thrombosis Research. 2014; 133 (Suppl 1): S35-S37. https://doi.org/10.1016/j.thromres.2014.03.016.

[11] Versteeg HH, Heemskerk JW, Levi M, Reitsma PH. New fundamentals in hemostasis. Physiol Rev. 2013 Jan;93(1):327-58. doi: 10.1152/physrev.00016.2011. PMID: 23303912

[12] Esmon CT. Basic mechanisms and pathogenesis of venous thrombosis. Blood Rev. 2009;23(5):225-229. doi:10.1016/j.blre.2009.07.002

[13] Kushner A, West WP, Pillarisetty LS. Virchow Triad. [Updated 2020 Sep 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK539697/

[14] Palta S, Saroa R, Palta A. Overview of the coagulation system. Indian J Anaesth. 2014;58(5):515-523. doi:10.4103/0019-5049.144643

[15] Waheed SM, Kudaravalli P, Hotwagner DT. Deep Vein Thrombosis. [Updated 2020 Aug 10]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK507708/

[16] Chang R, Mamun A, Dominic A, Le NT. SARS-CoV-2 Mediated Endothelial Dysfunction: The Potential Role of Chronic Oxidative Stress. Front Physiol. 2021;11:605908. Published 2021 Jan 15. doi:10.3389/fphys.2020.605908

[17] Gu SX, Tyagi T, Jain K, et al. Thrombocytopathy and endotheliopathy: crucial contributors to COVID-19 thromboinflammation. Nat Rev Cardiol. 2021;18(3):194-209. doi:10.1038/s41569-020-00469-1

[18] Gale AJ. Continuing education course #2: current understanding of hemostasis. Toxicol Pathol. 2011;39(1):273-280. doi:10.1177/0192623310389474

[19] Yau JW, Teoh H, Verma S. Endothelial cell control of thrombosis. BMC Cardiovasc Disord. 2015;15:130. Published 2015 Oct 19. doi:10.1186/s12872-015-0124-z

[20] LaPelusa A, Dave HD. Physiology, Hemostasis. [Updated 2021 May 9]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK5452

[21] Periayah MH, Halim AS, Mat Saad AZ. Mechanism Action of Platelets and Crucial Blood Coagulation Pathways in Hemostasis. Int J Hematol Oncol Stem Cell Res. 2017;11(4):319-327.

[22] Chaudhry R, Usama SM, Babiker HM. Physiology, Coagulation Pathways. [Updated 2020 Sep 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482253/

[23] Zgheib C, Xu J, Liechty KW. Targeting Inflammatory Cytokines and Extracellular Matrix Composition to Promote Wound Regeneration. Adv Wound Care (New Rochelle). 2014;3(4):344-355. doi:10.1089/wound.2013.0456

[24] Estevez B, Du X. New Concepts and Mechanisms of Platelet Activation Signaling. Physiology (Bethesda). 2017;32(2):162-177. doi:10.1152/physiol.00020.2016

[25] Rau JC, Beaulieu LM, Huntington JA, Church FC. Serpins in thrombosis, hemostasis and fibrinolysis. J Thromb Haemost. 2007;5 Suppl 1(Suppl 1):102-115. doi:10.1111/j.1538-7836.2007.02516.x

[26] Wolberg AS, Aleman MM, Leiderman K, Machlus KR. Procoagulant activity in hemostasis and thrombosis: Virchow's triad revisited. Anesth Analg. 2012;114(2):275-285. doi:10.1213/ANE.0b013e31823a088c

[27] Badimon L, Vilahur G. Thrombosis formation on atherosclerotic lesions and plaque rupture. J Intern Med. 2014 Dec;276(6):618-32. doi: 10.1111/joim.12296. Epub 2014 Sep 25. PMID: 25156650.

[28] Margetic S. Inflammation and haemostasis. Biochem Med (Zagreb). 2012;22(1):49-62.

[29] Patalakh I. Hemostatic Soluble Plasma Proteins During Acute-Phase Response and Chronic Inflammation. In: Francisco Veas editors. Acute Phase Proteins. IntechOpen; 2011. Chp5. DOI: 10.5772/20408

[30] Iba T, Levy JH. Inflammation and thrombosis: roles of neutrophils, platelets and endothelial cells and their interactions in thrombus formation during sepsis. J Thromb Haemost. 2018 Feb;16(2):231-241. doi: 10.1111/jth.13911. Epub 2017 Dec 21. PMID: 29193703.

[31] Oikonomopoulou K, Ricklin D, Ward PA, Lambris JD. Interactions between coagulation and complement--their role in inflammation. Semin

Immunopathol. 2012;34(1):151-165. doi:10.1007/s00281-011-0280-x

[32] Becatti M, Emmi G, Bettiol A, et al. Behçet's syndrome as a tool to dissect the mechanisms of thrombo-inflammation: clinical and pathogenetic aspects. Clin Exp Immunol. 2019;195(3):322-333. doi:10.1111/cei.13243

[33] Linton MRF, Yancey PG, Davies SS, et al. The Role of Lipids and Lipoproteins in Atherosclerosis. [Updated 2019 Jan 3]. In: Feingold KR, Anawalt B, Boyce A, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK343489/

[34] Branchford BR, Carpenter SL. The Role of Inflammation in Venous Thromboembolism. Front Pediatr. 2018;6:142. Published 2018 May 23. doi:10.3389/fped.2018.00142

[35] Rajendran P, Rengarajan T, Thangavel J, et al. The vascular endothelium and human diseases. Int J Biol Sci. 2013;9(10):1057-1069. Published 2013 Nov 9. doi:10.7150/ijbs.7502

[36] Hoang MP, Park J. Vasculitis. Hospital-Based Dermatopathology. 2020;245-296. Published 2020 Feb 29. doi:10.1007/978-3-030-35820-4_7

[37] Aksu K, Donmez A, Keser G. Inflammation-induced thrombosis: mechanisms, disease associations and management. Curr Pharm Des. 2012;18(11):1478-93. doi: 10.2174/138161212799504731. PMID: 22364132.

[38] Gaertner F, Massberg S. Blood coagulation in immunothrombosis-At the frontline of intravascular immunity. Semin Immunol. 2016 Dec;28(6):561-569. doi: 10.1016/j.smim.2016.10.010. Epub 2016 Nov 17. PMID: 27866916.

[39] Shi C, Yang L, Braun A, Anders HJ. Extracellular DNA-A Danger Signal Triggering Immunothrombosis. Front Immunol. 2020;11:568513. Published 2020 Oct 7. doi:10.3389/fimmu.2020.568513

[40] Eriksson O, Mohlin C, Nilsson B, Ekdahl KN. The Human Platelet as an Innate Immune Cell: Interactions Between Activated Platelets and the Complement System. Front Immunol. 2019;10:1590. Published 2019 Jul 10. doi:10.3389/fimmu.2019.01590

[41] Jayarangaiah A, Kariyanna PT, Chen X, Jayarangaiah A, Kumar A. COVID-19-Associated Coagulopathy: An Exacerbated Immunothrombosis Response. Clin Appl Thromb Hemost. 2020;26:1076029620943293. doi:10.1177/1076029620943293

[42] Gutmann C, Siow R, Gwozdz AM, Saha P, Smith A. Reactive Oxygen Species in Venous Thrombosis. International Journal of Molecular Sciences. 2020; 21(6):1918. https://doi.org/10.3390/ijms21061918

[43] Bae YS, Oh H, Rhee SG, Yoo YD. Regulation of reactive oxygen species generation in cell signaling. Mol Cells. 2011;32(6):491-509. doi:10.1007/s10059-011-0276-3

[44] Forrester SJ, Kikuchi DS, Hernandes MS, Xu Q, Griendling KK. Reactive Oxygen Species in Metabolic and Inflammatory Signaling. Circ Res. 2018;122(6):877-902. doi:10.1161/CIRCRESAHA.117.311401

[45] Go YM, Jones DP. Intracellular proatherogenic events and cell adhesion modulated by extracellular thiol/disulfide redox state. Circulation. 2005 Jun 7;111(22):2973-80. doi: 10.1161/CIRCULATIONAHA.104.515155. Epub 2005 May 31. PMID: 15927968.

[46] He L, He T, Farrar S, Ji L, Liu T, Ma X. Antioxidants Maintain Cellular

Redox Homeostasis by Elimination of Reactive Oxygen Species. Cell Physiol Biochem. 2017;44(2):532-553. doi: 10.1159/000485089. Epub 2017 Nov 17. PMID: 29145191.

[47] Di Meo S, Reed TT, Venditti P, Victor VM. Role of ROS and RNS Sources in Physiological and Pathological Conditions. Oxid Med Cell Longev. 2016;2016:1245049. doi:10.1155/2016/1245049

[48] Pizzino G, Irrera N, Cucinotta M, et al. Oxidative Stress: Harms and Benefits for Human Health. Oxid Med Cell Longev. 2017;2017:8416763. doi:10.1155/2017/8416763

[49] Mikkelsen RB, Wardman P. Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms. Oncogene. 2003 Sep 1;22(37):5734-54. doi: 10.1038/sj.onc.1206663. PMID: 12947383.

[50] Filomeni G, De Zio D, Cecconi F. Oxidative stress and autophagy: the clash between damage and metabolic needs. Cell Death Differ. 2015;22(3):377-388. doi:10.1038/cdd.2014.150

[51] Redza-Dutordoir M, Averill-Bates DA. Activation of apoptosis signalling pathways by reactive oxygen species. Biochim Biophys Acta. 2016 Dec;1863(12):2977-2992. doi: 10.1016/j.bbamcr.2016.09.012. Epub 2016 Sep 17. PMID: 27646922.

[52] Wang Q, Zennadi R. Oxidative Stress and Thrombosis during Aging: The Roles of Oxidative Stress in RBCs in Venous Thrombosis. Int J Mol Sci. 2020;21(12):4259. Published 2020 Jun 15. doi:10.3390/ijms21124259

[53] Heinle H, Brehme U, Kelber O, Schneider W, Weiser D. Oxidative stress in atherogenesis: Basic mechanisms and problems of therapy with antioxidants. Exp Clin Cardiol. 2001;6(2):77-80.

[54] Mittal M, Siddiqui MR, Tran K, Reddy SP, Malik AB. Reactive oxygen species in inflammation and tissue injury. Antioxid Redox Signal. 2014;20(7):1126-1167. doi:10.1089/ars.2012.5149

[55] Banfi C, Brioschi M, Barbieri SS, Eligini S, Barcella S, Tremoli E, Colli S, Mussoni L. Mitochondrial reactive oxygen species: a common pathway for PAR1- and PAR2-mediated tissue factor induction in human endothelial cells. J Thromb Haemost. 2009 Jan;7(1):206-16. doi: 10.1111/j.1538-7836.2008.03204.x. Epub 2008 Oct 25. PMID: 18983479.

[56] Masselli E, Pozzi G, Vaccarezza M, et al. ROS in Platelet Biology: Functional Aspects and Methodological Insights. Int J Mol Sci. 2020;21(14):4866. Published 2020 Jul 9. doi:10.3390/ijms21144866

[57] Qiao J, Arthur JF, Gardiner EE, Andrews RK, Zeng L, Xu K. Regulation of platelet activation and thrombus formation by reactive oxygen species. Redox Biol. 2018;14:126-130. doi:10.1016/j.redox.2017.08.021

[58] Freedman JE. Oxidative stress and platelets. Arterioscler Thromb Vasc Biol. 2008 Mar;28(3):s11-6. doi: 10.1161/ATVBAHA.107.159178. Epub 2008 Jan 3. PMID: 18174453.

[59] Phaniendra A, Jestadi DB, Periyasamy L. Free radicals: properties, sources, targets, and their implication in various diseases. Indian J Clin Biochem. 2015;30(1):11-26. doi:10.1007/s12291-014-0446-0

[60] Li R, Jia Z, Trush MA. Defining ROS in Biology and Medicine. React Oxyg Species (Apex). 2016;1(1):9-21. doi:10.20455/ros.2016.803

[61] Sena CM, Leandro A, Azul L, Seiça R, Perry G. Vascular Oxidative Stress: Impact and Therapeutic Approaches. Front Physiol. 2018 Dec

4;9:1668. doi: 10.3389/fphys.2018.01668. PMID: 30564132; PMCID: PMC6288353.

[62] Roberts W, Michno A, Aburima A, Naseem KM. Nitric oxide inhibits von Willebrand factor-mediated platelet adhesion and spreading through regulation of integrin alpha(IIb)beta(3) and myosin light chain. J Thromb Haemost. 2009 Dec;7(12):2106-15. doi: 10.1111/j.1538-7836.2009.03619.x. Epub 2009 Sep 18. PMID: 19765213.

[63] Loscalzo J. Nitric oxide insufficiency, platelet activation, and arterial thrombosis. Circ Res. 2001 Apr 27;88(8):756-62. doi: 10.1161/hh0801.089861. PMID: 11325866.

[64] Lubos E, Handy DE, Loscalzo J. Role of oxidative stress and nitric oxide in atherothrombosis. Front Biosci. 2008;13:5323-5344. Published 2008 May 1. doi:10.2741/3084

[65] Weisel JW, Litvinov RI. Red blood cells: the forgotten player in hemostasis and thrombosis. J Thromb Haemost. 2019 Feb;17(2):271-282. doi: 10.1111/jth.14360. Epub 2019 Jan 7. PMID: 30618125; PMCID: PMC6932746.9

[66] Lin YH. MicroRNA Networks Modulate Oxidative Stress in Cancer. Int J Mol Sci. 2019;20(18):4497. Published 2019 Sep 11. doi:10.3390/ijms20184497

[67] Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009 Jan;93(1):22-6. doi: 10.1016/j.ygeno.2008.08.012. Epub 2008 Oct 19. PMID: 18822366; PMCID: PMC2631358

[68] Pranavchand R, Reddy BM. Genomics era and complex disorders: Implications of GWAS with special reference to coronary artery disease, type 2 diabetes mellitus, and cancers. J Postgrad Med. 2016;62(3):188-198. doi:10.4103/0022-3859.186390

[69] Jorine S. Koenderman and Pieter H. Reitsma (November 9th 2011). Inherited Thrombophilia: Past, Present, and Future Research, Thrombophilia, Andrea Luigi Tranquilli, IntechOpen, DOI: 10.5772/26050. Available from: https://www.intechopen.com/books/thrombophilia/inherited-thrombophilia-past-present-and-future-research

[70] JORDAN FL, NANDORFF A. The familial tendency in thrombo-embolic disease. Acta Med Scand. 1956 Dec 31;156(4):267-75. doi: 10.1111/j.0954-6820.1956.tb00084.x. PMID: 13394174.

[71] Corral J, de la Morena-Barrio ME, Vicente V. The genetics of antithrombin. Thromb Res. 2018 Sep;169:23-29. doi: 10.1016/j.thromres.2018.07.008. Epub 2018 Jul 5. PMID: 30005274.

[72] Khan S, Dickerman JD. Hereditary thrombophilia. Thromb J. 2006;4:15. Published 2006 Sep 12. doi:10.1186/1477-9560-4-15

[73] Smith NL, Chen MH, Dehghan A, Strachan DP, Basu S, Soranzo N, Hayward C, Rudan I, Sabater-Lleal M, Bis JC, de Maat MP, Rumley A, Kong X, Yang Q, Williams FM, Vitart V, Campbell H, Mälarstig A, Wiggins KL, Van Duijn CM, McArdle WL, Pankow JS, Johnson AD, Silveira A, McKnight B, Uitterlinden AG; Wellcome Trust Case Control Consortium;, Aleksic N, Meigs JB, Peters A, Koenig W, Cushman M, Kathiresan S, Rotter JI, Bovill EG, Hofman A, Boerwinkle E, Tofler GH, Peden JF, Psaty BM, Leebeek F, Folsom AR, Larson MG, Spector TD, Wright AF, Wilson JF, Hamsten A, Lumley T, Witteman JC, Tang W, O'Donnell CJ. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. Circulation. 2010 Mar

30;121(12):1382-92. doi: 10.1161/CIRCULATIONAHA.109.869156. Epub 2010 Mar 15. Erratum in: Circulation.2010 Jul 20;122(3):e399. PMID: 20231535; PMCID: PMC2861278.

[74] Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet. 2017;18(6):331-344. doi:10.1038/nrg.2016.160

[75] Hinds DA, Buil A, Ziemek D, et al. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. Hum Mol Genet. 2016;25(9):1867-1874. doi:10.1093/hmg/ddw037

[76] Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature. 1994 May 5;369(6475):64-7. doi: 10.1038/369064a0. PMID: 8164741.

[77] Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood. 1996 Nov 15;88(10):3698-703. PMID: 8916933

[78] Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995 May;10(1):111-3. doi: 10.1038/ng0595-111. PMID: 7647779

[79] Egeberg O. Inherited Antithrombin Deficiency Causing Thrombophilia. Thromb Diath Haemorrh. 1965 Jun 15;13:516-30. PMID: 14347873

[80] Griffin JH, Evatt B, Zimmerman TS, Kleiss AJ, Wideman C. Deficiency of protein C in congenital thrombotic disease. J Clin Invest. 1981 Nov;68(5):1370-3. doi: 10.1172/jci110385. PMID: 6895379; PMCID: PMC370934

[81] Comp PC, Esmon CT. Recurrent venous thromboembolism in patients with a partial deficiency of protein S. N Engl J Med. 1984 Dec 13;311(24):1525-8. doi: 10.1056/NEJM198412133112401. PMID: 6239102

[82] Merriman L, Greaves M. Testing for thrombophilia: an evidence-based approach. Postgrad Med J. 2006;82(973):699-704. doi:10.1136/pgmj.2006.048090

[83] Middeldorp S. Inherited thrombophilia: a double-edged sword. Hematology Am Soc Hematol Educ Program. 2016;2016(1):1-9. doi:10.1182/asheducation-2016.1.1

[84] Kang SS, Wong PW, Zhou JM, Sora J, Lessick M, Ruggie N, Grcevich G. Thermolabile methylenetetrahydrofolate reductase in patients with coronary artery disease. Metabolism. 1988 Jul;37(7):611-3. doi: 10.1016/0026-0495(88)90076-5. PMID: 3386531.

[85] Dilley A, Austin H, El-Jamil M et al. Genetic factors associated with thrombosis in pregnancy in a United States population. Am J Obstet Gynecol. 2000 Nov;183(5):1271-7.

[86] Gerhardt A, Scharf RE, Beckmann MW et al. Prothrombin and factor V mutations in women with a history of thrombosis during pregnancy and the puerperium. N Engl J Med. 2000 Feb 10;342(6):374-80.

[87] Martinelli I, De Stefano V, Taioli E et al. Inherited thrombophilia and first venous thromboembolism during pregnancy and puerperium. Thromb Haemost. 2002 May;87(5):791-5.

[88] Murphy RP, Donoghue C, Nallen RJ et al. Prospective evaluation of the risk conferred by factor V Leiden and thermolabile methylenetetrahydrofolate reductase polymorphisms in pregnancy. Arterioscler Thromb Vasc Biol. 2000 Jan;20(1):266-70.

[89] Tormene D, Simioni P, Prandoni P et al. Factor V Leiden mutation and the risk of venous thromboembolism in pregnant women. Haematologica. 2001 Dec;86(12):1305-9

[90] Ogunyemi D, Cuellar F, Ku W, Arkel Y. Association between inherited thrombophilias, antiphospholipid antibodies, and lipoprotein A levels and venous thromboembolism in pregnancy. Am J Perinatol. 2003 Jan;20(1):17-24.

[91] Duga S, Asselta R, Tenchini ML. Coagulation factor V. Int J Biochem Cell Biol. 2004 Aug;36(8):1393-9. doi: 10.1016/j.biocel.2003.08.002. PMID: 15147718.

[92] Kalafatis M, Beck DO, Mann KG. Structural requirements for expression of factor Va activity. J Biol Chem. 2003 Aug 29;278(35):33550-61. doi: 10.1074/jbc.M303153200. Epub 2003 Jun 4. PMID: 12788947.

[93] Thorelli E, Kaufman RJ, Dahlbäck B. Cleavage of factor V at Arg 506 by activated protein C and the expression of anticoagulant activity of factor V. Blood. 1999 Apr 15;93(8):2552-8. PMID: 10194434.

[94] Dahlbäck B. Advances in understanding pathogenic mechanisms of thrombophilic disorders. Blood. 2008 Jul 1;112(1):19-27. doi: 10.1182/blood-2008-01-077909. PMID: 18574041.

[95] Chan WP, Lee CK, Kwong YL, Lam CK, Liang R. A novel mutation of Arg306 of factor V gene in Hong Kong Chinese. Blood. 1998 Feb 15;91(4):1135-9. PMID: 9454741.

[96] Williamson D, Brown K, Luddington R, Baglin C, Baglin T. Factor V Cambridge: a new mutation (Arg306->Thr) associated with resistance to activated protein C. Blood. 1998 Feb 15;91(4):1140-4. PMID: 9454742.

[97] Steen M, Norstrøm EA, Tholander AL, Bolton-Maggs PH, Mumford A, McVey JH, Tuddenham EG, Dahlbäck B. Functional characterization of factor V-Ile359Thr: a novel mutation associated with thrombosis. Blood. 2004 May 1;103(9):3381-7. doi: 10.1182/blood-2003-06-2092. Epub 2003 Dec 24. PMID: 14695241.

[98] Brugge JM, Simioni P, Bernardi F, Tormene D, Lunghi B, Tans G, Pagnan A, Rosing J, Castoldi E. Expression of the normal factor V allele modulates the APC resistance phenotype in heterozygous carriers of the factor V Leiden mutation. J Thromb Haemost. 2005 Dec;3(12):2695-702. doi: 10.1111/j.1538-7836.2005.01634.x. PMID: 16359508.

[99] Rosén SB, Sturk A. Activated protein C resistance--a major risk factor for thrombosis. Eur J Clin Chem Clin Biochem. 1997 Jul;35(7):501-16. PMID: 9263726.

[100] Castaman G, Lunghi B, Missiaglia E, Bernardi F, Rodeghiero F. Phenotypic homozygous activated protein C resistance associated with compound heterozygosity for Arg506Gln (factor V Leiden) and His1299Arg substitutions in factor V. Br J Haematol. 1997 Nov;99(2):257-61. doi: 10.1046/j.1365-2141.1997.3993213.x. PMID: 9375735.

[101] Lunghi B, Castoldi E, Mingozzi F, Bernardi F. A new factor V gene polymorphism (His 1254 Arg) present in subjects of african origin mimics the R2 polymorphism (His 1299 Arg). Blood. 1998 Jan 1;91(1):364-5. PMID: 9414311.

[102] Zivelin A, Mor-Cohen R, Kovalsky V, Kornbrot N, Conard J, Peyvandi F, Kyrle PA, Bertina R, Peyvandi F, Emmerich J, Seligsohn U. Prothrombin 20210G>A is an ancestral prothrombotic mutation that occurred in whites approximately 24,000 years ago. Blood. 2006 Jun 15;107(12):4666-8. doi: 10.1182/blood-2005-12-5158. Epub 2006 Feb 21. PMID: 16493002.

[103] Emmerich J, Aiach M. Facteurs génétiques de risque de thrombose [Genetic risk factors of thrombosis]. Ann Cardiol Angeiol (Paris). 2002 Jun;51(3):129-34. French. doi: 10.1016/s0003-3928(02)00084-7. PMID: 12471642

[104] Friedmann AP, Koutychenko A, Wu C, Fredenburgh JC, Weitz JI, Gross PL, Xu P, Ni F, Kim PY. Identification and characterization of a factor Va-binding site on human prothrombin fragment 2. Sci Rep. 2019 Feb 21;9(1):2436. doi: 10.1038/s41598-019-38857-4. PMID: 30792421; PMCID: PMC6385242

[105] Chinnaraj M, Planer W, Pozzi N. Structure of Coagulation Factor II: Molecular Mechanism of Thrombin Generation and Development of Next-Generation Anticoagulants. Front Med (Lausanne). 2018;5:281. Published 2018 Oct 2. doi:10.3389/fmed.2018.00281

[106] Narayanan S. Multifunctional roles of thrombin. Ann Clin Lab Sci. 1999 Oct-Dec;29(4):275-80. PMID: 10528826.

[107] Warshawsky I, Hren C, Sercia L, Shadrach B, Deitcher SR, Newton E, Kottke-Marchant K. Detection of a novel point mutation of the prothrombin gene at position 20209. Diagn Mol Pathol. 2002 Sep;11(3):152-6. doi: 10.1097/00019606-200209000-00005. PMID: 12218454.

[108] Wylenzek M, Geisen C, Stapenhorst L, Wielckens K, Klingler KR. A novel point mutation in the 3′ region of the prothrombin gene at position 20221 in a Lebanese/Syrian family. Thromb Haemost. 2001 May;85(5):943-4. PMID: 11372696.

[109] Ceelie H, Bertina RM, van Hylckama Vlieg A, Rosendaal FR, Vos HL. Polymorphisms in the prothrombin gene and their association with plasma prothrombin levels. Thromb Haemost. 2001 Jun;85(6):1066-70. PMID: 11434686.

[110] Ropero P, González FA, Nieto JM, Villegas A, Sevilla J, Pérez G, Alonso JM, Recasens V, Abio M, Vagace JM, Vanegas RJ, González Fernández B, Martínez R. C>A substitution in NT 46 of the 3' UTR region (the α complex protected region) of the alpha-1 globin gene: a non-deletional mutation or polymorphism? J Clin Pathol. 2020 Jan;73(1):14-16. doi: 10.1136/jclinpath-2019-206004. Epub 2019 Aug 21. PMID: 31434698

[111] Leclerc D, Sibani S, Rozen R. Molecular Biology of Methylenetetrahydrofolate Reductase (MTHFR) and Overview of Mutations/Polymorphisms. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6561/

[112] Townsend DM, Tew KD, Tapiero H. Sulfur containing amino acids and human disease. Biomed Pharmacother. 2004;58(1):47-55. doi:10.1016/j.biopha.2003.11.005

[113] Škovierová H, Vidomanová E, Mahmood S, et al. The Molecular and Cellular Effect of Homocysteine Metabolism Imbalance on Human Health. Int J Mol Sci. 2016;17(10):1733. Published 2016 Oct 20. doi:10.3390/ijms17101733

[114] Barroso M, Handy DE, Castro R. The Link Between

Hyperhomocysteinemia and Hypomethylation: Implications for Cardiovascular Disease. Journal of Inborn Errors of Metabolism and Screening. 2017; April 5. doi. org/10.1177/2326409817698994

[115] Tan JS, Yan XX, Wu Y, Gao X, Xu XQ, Jiang X, Jia L, Hu S, Hua L, Wang XJ. Rare variants in MTHFR predispose to occurrence and recurrence of pulmonary embolism. Int J Cardiol. 2021 May 15;331:236-242. doi: 10.1016/j. ijcard.2021.01.073. Epub 2021 Feb 8. PMID: 33571559.

[116] Yamada K, Chen Z, Rozen R, Matthews RG. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. Proc Natl Acad Sci U S A. 2001 Dec 18;98(26):14853-8. doi: 10.1073/ pnas.261469998. Epub 2001 Dec 11. PMID: 11742092; PMCID: PMC64948.

[117] Contreras-Cubas C, Sánchez-Hernández BE, García-Ortiz H, et al. Heterogenous Distribution of MTHFR Gene Variants among Mestizos and Diverse Amerindian Groups from Mexico. PLoS One. 2016;11(9):e0163248. Published 2016 Sep 20. doi:10.1371/ journal.pone.0163248

[118] van der Put NM, Gabreëls F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, van den Heuvel LP, Blom HJ. A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? Am J Hum Genet. 1998 May;62(5):1044-51. doi: 10.1086/301825. PMID: 9545395; PMCID: PMC1377082.

[119] Soare AM, Popa C. Deficiencies of proteins C, S and antithrombin and activated protein C resistance--their involvement in the occurrence of Arterial thromboses. J Med Life. 2010;3(4):412-415.

[120] Pilli VS, Plautz W, Majumder R. The Journey of Protein S from an Anticoagulant to a Signaling Molecule. JSM Biochem Mol Biol. 2016;3(1):1014.

[121] Rezaie AR. Regulation of the protein C anticoagulant and antiinflammatory pathways. Curr Med Chem. 2010;17(19):2059-2069. doi:10.2174/092986710791233706

[122] Dinarvand P, Moser KA. Protein C Deficiency. Arch Pathol Lab Med. 2019 Oct;143(10):1281-1285. doi: 10.5858/ arpa.2017-0403-RS. Epub 2019 Feb 1. PMID: 30702334

[123] Gupta A, Patibandla S. Protein C Deficiency. [Updated 2021 Jan 29]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm. nih.gov/books/NBK542222/

[124] Tuddenham EG, Cooper DN. Protein C and protein C inhibitor. In: Tuddenham EG, Cooper DN, editors. The molecular genetics of haemostasis and its inherited disorders. NewYork: Oxford University Press; 1994. p. 149-60.

[125] Miletich J, Sherman L, Broze G Jr. Absence of thrombosis in subjects with heterozygous protein C deficiency. N Engl J Med. 1987 Oct 15;317(16):991-6. doi: 10.1056/NEJM198710153171604. PMID: 3657866

[126] Zhang H, Bi X, Su Z, Tu X, Wang L, Shen B. A novel compound heterozygous mutations in protein C gene causing neonatal purpura fulminans. Blood Coagul Fibrinolysis. 2018 Mar;29(2):216-219. doi: 10.1097/ MBC.0000000000000687. PMID: 29356699.

[127] Lu Y, Mehta-D'souza P, Biswas I, et al. Ile73Asn mutation in protein C introduces a new N-linked glycosylation site on the first EGF-domain of protein C and causes thrombosis.

Haematologica. 2020;105(6):1712-1722. doi:10.3324/haematol.2019.227033

[128] Winther-Larsen A, Kjaergaard AD, Larsen OH, Hvas AM, Nissen PH. Protein C deficiency; PROC gene variants in a Danish population. Thromb Res. 2020 Jan;185:153-159. doi: 10.1016/j.thromres.2019.11.027. Epub 2019 Nov 30. PMID: 31821907.

[129] Danese S, Vetrano S, Zhang L, Poplis VA, Castellino FJ. The protein C pathway in tissue inflammation and injury: pathogenic role and therapeutic implications. Blood. 2010;115(6):1121-1130. doi:10.1182/blood-2009-09-201616

[130] Briët E, BroekmansAW, Engesser L. Hereditary protein S deficiency. In: Bertina RM, editor. Protein C and related proteins. Edinburgh: Churchill Livingstone; 1988. p. 203-20.

[131] Dahlbäck B. The tale of protein S and C4b-binding protein, a story of affection. Thromb Haemost. 2007 Jul;98(1):90-6. PMID: 17597997

[132] Stenflo J. Contributions of Gla and EGF-like domains to the function of vitamin K-dependent coagulation factors. Crit Rev Eukaryot Gene Expr. 1999;9(1):59-88. PMID: 10200912

[133] Collin O, Bergh A. Leydig cells secrete factors which increase vascular permeability and endothelial cell proliferation. Int J Androl. 1996 Aug;19(4):221-8. doi: 10.1111/j.1365-2605.1996.tb00466.x. PMID: 8940660

[134] Hepner M, Karlaftis V. Antithrombin. Methods Mol Biol. 2013;992:355-64. doi: 10.1007/978-1-62703-339-8_28. PMID: 23546728.

[135] Hsu E, Moosavi L. Biochemistry, Antithrombin III. [Updated 2020 Sep 11]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK545295/

[136] Olson ST, Gettins PG. Regulation of proteases by protein inhibitors of the serpin superfamily. Prog Mol Biol Transl Sci. 2011;99:185-240. doi: 10.1016/B978-0-12-385504-6.00005-1. PMID: 21238937

[137] Pike RN, Buckle AM, le Bonniec BF, Church FC. Control of the coagulation system by serpins. Getting by with a little help from glycosaminoglycans. FEBS J. 2005 Oct;272(19):4842-51. doi: 10.1111/j.1742-4658.2005.04880.x. PMID: 16176258

[138] Li Y, Sun C, Yates EA, Jiang C, Wilkinson MC, Fernig DG. Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification. Open Biol. 2016;6(3):150275. doi:10.1098/rsob.150275

[139] Yue Y, Sun Q, Xiao L, et al. Association of SERPINC1 Gene Polymorphism (rs2227589) With Pulmonary Embolism Risk in a Chinese Population. Front Genet. 2019;10:844. Published 2019 Sep 13. doi:10.3389/fgene.2019.00844

[140] Picard V, Dautzenberg MD, Villoutreix BO, Orliaguet G, Alhenc-Gelas M, Aiach M. Antithrombin Phe229Leu: a new homozygous variant leading to spontaneous antithrombin polymerization in vivo associated with severe childhood thrombosis. Blood. 2003 Aug 1;102(3):919-25. doi: 10.1182/blood-2002-11-3391. Epub 2003 Feb 20. PMID: 12595305.

[141] Toderici M, de la Morena-Barrio ME, Padilla J, Miñano A, Antón AI, Iniesta JA, Herranz MT, Fernández N, Vicente V, Corral J. Identification of Regulatory Mutations in SERPINC1 Affecting Vitamin D

Response Elements Associated with Antithrombin Deficiency. PLoS One. 2016 Mar 22;11(3):e0152159. doi: 10.1371/journal.pone.0152159. Erratum in: PLoS One. 2016;11(7):e0159987. PMID: 27003919; PMCID: PMC4803246

[142] Bezemer I. D., Bare L. A., Doggen C. J., Arellano A. R., Tong C., Rowland C. M., et al. (2008). Gene variants associated with deep vein thrombosis. *JAMA* 299 (11), 1306-1314. 10.1001/jama.299.11.1306

[143] Anton A. I., Teruel R., Corral J., Minano A., Martinez-Martinez I., Ordonez A., et al. (2009). Functional consequences of the prothrombotic SERPINC1 rs2227589 polymorphism on antithrombin levels. *Haematologica* 94 (4), 589-592. 10.3324/haematol.2008. 000604

[144] Pernod G, Biron-Andreani C, Morange PE, Boehlen F, Constans J, Couturaud F, Drouet L, Jude B, Lecompte T, Le Gal G, Trillot N, Wahl D; French group on haemostasis and thrombosis; French Society of vascular medicine. Recommendations on testing for thrombophilia in venous thromboembolic disease: a French consensus guideline. J Mal Vasc. 2009 May;34(3):156-203. English, French. doi: 10.1016/j.jmv.2009.02.005. PMID: 19645086

[145] Colucci G, Tsakiris DA. Thrombophilia screening revisited: an issue of personalized medicine. J Thromb Thrombolysis. 2020 May;49(4):618-629. doi: 10.1007/ s11239-020-02090-y. PMID: 32248336; PMCID: PMC7182628

[146] Lane DA, Grant PJ. Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. Blood. 2000 Mar 1;95(5):1517-32. PMID: 10688804.

[147] Zhang S, Taylor AK, Huang X, Luo B, Spector EB, Fang P, Richards CS; ACMG Laboratory Quality Assurance Committee. Venous thromboembolism laboratory testing (factor V Leiden and factor II c.*97G>A), 2018 update: a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2018 Dec;20(12):1489-1498. doi: 10.1038/ s41436-018-0322-z. Epub 2018 Oct 5. PMID: 30297698.

[148] Connors JM. Thrombophilia Testing and Venous Thrombosis. N Engl J Med. 2017 Sep 21;377(12):1177-1187. doi: 10.1056/NEJMra1700365. PMID: 28930509.

[149] Cooper PC, Pavlova A, Moore GW, Hickey KP, Marlar RA. Recommendations for clinical laboratory testing for protein C deficiency, for the subcommittee on plasma coagulation inhibitors of the ISTH. J Thromb Haemost. 2020 Feb;18(2):271-277. doi: 10.1111/ jth.14667. PMID: 31999059.

[150] Van Cott EM, Orlando C, Moore GW, Cooper PC, Meijer P, Marlar R; Subcommittee on Plasma Coagulation Inhibitors. Recommendations for clinical laboratory testing for antithrombin deficiency; Communication from the SSC of the ISTH. J Thromb Haemost. 2020 Jan;18(1):17-22. doi: 10.1111/ jth.14648. PMID: 31894660.

[151] Marlar RA, Gausman JN, Tsuda H, Rollins-Raval MA, Brinkman HJM. Recommendations for clinical laboratory testing for protein S deficiency: Communication from the SSC committee plasma coagulation inhibitors of the ISTH. J Thromb Haemost. 2021 Jan;19(1):68-74. doi: 10.1111/jth.15109. PMID: 33405382.

[152] Moore GW, Van Cott EM, Cutler JA, Mitchell MJ, Adcock DM; subcommittee on plasma coagulation inhibitors. Recommendations for clinical laboratory testing of activated

protein C resistance; communication from the SSC of the ISTH. J Thromb Haemost. 2019 Sep;17(9):1555-1561. doi: 10.1111/jth.14532. Epub 2019 Jul 17. PMID: 31317658.

[153] Tran HA, Gibbs H, Merriman E, Curnow JL, Young L, Bennett A, Tan CW, Chunilal SD, Ward CM, Baker R, Nandurkar H. New guidelines from the Thrombosis and Haemostasis Society of Australia and New Zealand for the diagnosis and management of venous thromboembolism. Med J Aust. 2019 Mar;210(5):227-235. doi: 10.5694/mja2.50004. Epub 2019 Feb 10. Erratum in: Med J Aust. 2019 Jul;211(2):94. Erratum in: Med J Aust. 2020 Feb;212(3):108. PMID: 30739331.

[154] Brisson RT, Arruda JFL, Silva LD, de Jesus DLJ, Zetola VF, Kauffmann Nogueira MAC. Stroke in Young Military Men With Heterozygous for MTHFR Gene Mutation or Factor V Leiden Gene Mutation Associated With Patent Foramen Ovale: Report of Two Cases and Therapeutic Strategy. Mil Med. 2021 May 16:usab192. doi: 10.1093/milmed/usab192. Epub ahead of print. PMID: 33993312.

[155] Boekholdt SM, Kramer MH. Arterial thrombosis and the role of thrombophilia. Semin Thromb Hemost. 2007 Sep;33(6):588-96. doi: 10.1055/s-2007-985755. PMID: 17768691.

[156] Mandala E, Lafaras C, Tsioni C, et al. Prevalence of thrombophilic mutations in patients with unprovoked thromboembolic disease. A comparative analysis regarding arterial and venous disease. Hippokratia. 2012;16(3): 250-255.

[157] Alhenc-Gelas M. Mutations et polymorphismes des protéines de l'hémostase prédisposant à la thrombose. Elsevier Masson SAS. 2011; http://dx.doi.org/10.1016/S1166-4568(11)52983-3

Chapter 4

Cytokine Gene Polymorphism and Cancer Risk: A Promising Tool for Individual Susceptibility and Prognostic Implications

Arshad A. Pandith, Ina Bhat, Sheikh Mansoor, Aabid Koul, Usma Manzoor, Iqra Anwar, Fozia Mohammad, Qurat Ul Aein, Shahid M. Baba and Carmen Vladulescu

Abstract

Cytokines are potent molecules produced mainly by specific activated immune cells to control inflammatory responses besides other biologic processes. Although active participation of cytokines provides defense against carcinogenesis on the other hand, deregulation at the genetic level influences their activity to promote tumor development. Among many aspects, constitutional polymorphic sequence variations are key factors that derange the cytokine expression to lead an individual's propensity to risk for different cancers. Cytokine polymorphisms are now believed to alter these critical molecules that have a dual face in carcinogenesis as, when implicated in the activation of the immune response, these molecules check the cancer development while their persistent inflammatory reaction can envisage the development of malignancy and tumor growth. We have given ample evidence of case-control studies in a range of cancers where substantial evidence, as reported in this chapter, links polymorphism of cytokine gene susceptibility with numerous cancers. Cytokine gene polymorphism is vital to be significant bimolecular genetic determinants of susceptibility and prognosis of cancer. A strong need is felt for more case-control association studies in cytokine candidate genes involved in specific pathways for particular cancer in bigger powered sample sizes involving additional variables to disclose their factual risk for cancer.

Keywords: cytokines, polymorphism, malignancy, tumor, immune response, case-control, inflammatory reaction

1. Introduction

Cancer is the world's second leading cause of death and accounts for around 1 in 6 deaths worldwide. A projected 18.1 million new cancer-related cases and 9.6 million cancer-related deaths were recorded in 2018, according to the 2018 GLOBOCAN study. Lung cancer is the most frequently diagnosed cancer for both sexes with 11.6% of the total cases of cancer and is also the leading cause of cancer death with 18.4% of the total cancer deaths. Cancer, a cellular overgrowth disorder,

involves cellular transformation, apoptotic dysregulation, uncontrolled proliferation, invasion, angiogenesis, and metastasis [1]. A significant correlation between chronic infection, inflammation, and certain forms of cancer have been suggested in clinical and epidemiological research, with inflammation frequently occurring in and around tumors [2].

Recent studies have shown that chronic inflammation that regulates the microenvironment of the tumor is involved in the initiation of the tumor and is a critical component of tumor promotion and progression [3]. The insistent inflammatory microenvironment contributes to increased promotion of the tumor, accelerated progression, invasion of the surrounding tissues, angiogenesis, and eventually metastasis [4]. The tumor microenvironment has a vast abundance of cytokines as well as other inflammatory mediators which impact immunosuppression, cancer growth, tissue remodeling, and angiogenesis [5]. Cytokines are a heterogeneous group of glycoproteins or small soluble polypeptides (secreted or membrane-bound) chiefly produced by the immune cells which under normal conditions are produced in response to specific stimuli exerting pleiotropic and redundant effects thereby altering the behavior of the identical or different cells by regulating the growth, differentiation, and activation of normal cells especially immune cells [6]. Depending on the microenvironment, cytokines may either have pro- or anti-inflammatory or may even have immunosuppressive activity [7]. In response to both antigen-specific and nonspecific stimuli, the development of multiple cytokines by immune cells plays a crucial role in the outcome of inflammatory immune responses [8]. They are either released in response to a variety of cellular stresses, including carcinogen-induced injury, infection, inflammation, and immunity that inhibits tumor growth and progression, or are host-derived cytokines that cause cancer cells to promote growth, reduce apoptosis, aid invasion and metastasis. The role of cytokines in subsequent settings is to cause a host response aimed at controlling cellular stress and reducing cell damage. Although critical insult control encourages tissue repair, failure to control the injury can lead to the insistent development of cytokines, leading to further aggravation of tissue damage. Depending on several stages of cancer development and progression, host responses to cellular stress may influence it inherently [9]. Therefore, the cytokine mixture produced in the tumor microenvironment plays an important role in the pathogenesis of cancer [10]. The molecular and cellular changes that eventually lead to cancer stimulate changes in the local cytokine environment, stimulate immune cell intrusions, and release additional cytokines that function in an autocrine or paracrine manner [11]. In addition, pro-inflammatory cytokines are associated with anorectic and cachectic disease in patients with progressive cancer, pain in the form of both algia and dynia, toxicity, and resistance to treatment. Physical exercise can, however, alter cytokine levels and decrease fatigue in cancer patients, and may also increase their prognosis [7].

Cytokines provide a key molecular link between inflammation, tumor promotion, and progression which mainly include the interferons (IFNs), the interleukins (ILs), the tumor necrosis factor super family (TNFs), the transforming growth factor super family (TGFs), and Colony Stimulating Factors (CSFs) [12].

Genetic variations in the human genome can govern the risk of cancer development, symptoms, treatment, and its outcome [7]. In recent years, within cytokine gene sequences, many single nucleotide polymorphisms (SNPs) and a small number of microsatellite polymorphisms have been reported, mainly within their promoter regions [13] Indeed the most common variations in the genome are the SNPs [14]. Some suggest that the differences in SNPs may also help explain cancer disparities among the various ethnic groups [15]. Differential levels of gene transcription, including TNF -308 and IL-10-10822 can be associated with

some of these polymorphisms [16]. Several genetic studies have attempted to associate these cytokine polymorphisms with cancers; however, the studies are still limited, but increasingly evolving. The link between TNF alpha-SNPs and specific cancers, including oral carcinoma, has been documented in several studies [17] and non-Hodgkin's lymphoma [18], but most of these associations are refuted by others. IL-1 gene polymorphisms that provide increased expression of this pro-inflammatory cytokine are associated with an increased risk of cancers, mainly of gastric origin [19]. The association between IL-1B/IL-1RN polymorphisms and the development of gastric adenocarcinoma following *Helicobacter pylori* infection is also well known [20]. In relation to cancer, IL-10 polymorphisms are of particular concern since IL-10 has immunosuppressive and anti-angiogenic properties [8]. In view of this, this chapter highlights the essential role of genetic polymorphisms with respect to a host of cancers and their susceptibility in different geographical regions.

2. Pro-inflammatory cytokine gene polymprphisms

2.1 Solid tumors

2.1.1 Breast cancer

2.1.1.1 Interleukin 6

IL-6 is a major inflammatory pleiotropic cytokine that is considered a main growth-promoting factor [21]. Several IL-6 target genes are involved in the progression of the cell cycle and apoptosis suppression, which emphasizes the significance of IL-6 in tumorigenesis [22]. Many polymorphic IL-6 studies have shown that it is associated with breast cancer risk and prognosis. A study indicates that IL-6 is a predisposing genetic factor that contributes to the prognosis of breast cancer, with a G/C polymorphism associated with high levels of IL-6 production within the promoter region of the IL-6 gene correlating with a worse prognosis. While few studies have been published to date on polymorphisms within the IL-6 gene cluster and breast cancer [23], there is still controversy about the evidence on 174 G/C IL-6 polymorphism in breast cancer. The presence of IL-6 polymorphism was associated with an improved outcome in high-risk breast cancer, reported by DeMichele et al. [24]. Iacopetta et al. [25] found that the IL-6 polymorphism was predictive of the phenotype of aggressive breast cancer

2.1.1.2 Interleukin 10

IL-10 is an immunosuppressive and anti-angiogenic multifunctional cytokine that can have both tumor-promoting and inhibiting properties [26]. The IL-10 gene promoter has been recognized to have a significant number of polymorphisms (primarily SNPs) [27]. As a consequence, promoter polymorphisms have been the most scrutinized, particularly with regard to possible gene transcription and protein production influences. Many studies on the interactions between genotypes of IL-10 and breast cancer have been published. One small breast cancer study confirmed that the low-expression 21082 AA genotype was associated with a high risk of disease [28]. In comparison, a larger case-control study found a correlation between the 2592 AA genotype and the reduced risk of breast cancers [29] but did not associate with any clinical parameters.

2.1.1.3 Interferon-γ

Interferon-γ (IFN-γ) is the cytokine that likely has a vital role in carcinogenesis. Several combined epidemiological studies have shown that IFN-γ can function as a major risk factor for breast cancer development and progression, primarily due to genetic polymorphisms of IFN-γ [30]. Interestingly, genetic polymorphisms in the IFN-γ gene are likely to affect the degree of IFN-γ expression, leading to impaired function or decreased IFN-γ activity; subsequent low IFN-γ expression levels would encourage tumor development, which can contribute to can susceptibility to breast cancer [31]. In order to be more precise, many common IFN-γ gene polymorphisms have been found to increase the risk of breast cancer, such as rs20697 (−1615C/T) and rs2430561 (+874T/A) [30, 32, 33]. Contradictory findings have also been reported recently, although several studies have also suggested that IFN-γ genetic polymorphisms can play a critical role in breast cancer pathogenesis [30, 31]. Gene polymorphisms, with low levels of IFN-γ, can alter the function and expressions of IFN-γ. The rs2430561 A>T polymorphism was revealed by Liu et al. in intron 1 of IFN-γ [34]. The gene can transform its transcription functionally and lead to susceptibility to breast cancer [35]. Another meta-analysis shows that genetic polymorphisms within the IFN-γ gene were significantly associated with an increased risk of breast cancer, especially the polymorphism of 2430561 T>A. No connection was, however, identified between rs2069705 C>T polymorphism and breast cancer susceptibility. Genetic polymorphisms and the risk of breast cancer in Asians, but not among Caucasians, suggest that ethnic differences may affect the susceptibility of individuals to breast cancer [36].

2.1.1.4 Tumor necrosis factor-alpha (TNF-α)

TNF-α is an essential pro-inflammatory cytokine for human cancer growth and progression [37] by stimulating the development of genotoxic molecules (NO, ROS) that can lead to DNA damage and mutations, TNF-α can promote tumor initiation and progression. The increased risk of many cancers, including breast cancer, is associated with genetic polymorphisms that boost TNF-α development [33].

Increasing evidence has shown that a SNP in the promoter region of the TNF-α gene (−308G>A, rs1800629) induces genetic susceptibility in many forms of tumors like BC [38]. The TNF-α-308A allele tends to have a higher constitutive and inducible expression, as the −308G>A mutation affects the AP-22020 consensus binding site [39]. The results on the association of –308 (G/A) TNF-α polymorphism with breast cancer development are rather contradictory [40]. A rise in homozygous (−308AA) TNF-α genotype frequency was significantly associated with the development of breast cancer and poor prognosis in Tunisian women [40]. No link between the homozygous (−308AA) TNF-α genotype and breast cancer risk was identified in Holland. The association of allele (−308A) TNF-α with tumor vascularization has, however, been shown to be [41]. Where the majority of studies found no link between-308 (G/A) TNF-α polymorphism and the occurrence of breast cancer [42, 43]. Two recent meta-analyses have confirmed that the genotypes TNF-α-308GA and AA were significantly associated with a reduced risk of breast cancer in Caucasians [14, 16, 44, 45]. The allele frequencies in the controls of some studies [42, 46] did not show compliance with Hardy-Weinberg (HW) in the meta-analyses. In addition, Yang et al meta-analysis included one study comparing the frequencies of various genotypes of TNF-α-308 polymorphism in patients with benign breast disease and controls, and another study that did not have frequencies for each genotype [14, 16, 44, 45]

2.1.2 Hepato-celllular carcinoma

2.1.2.1 Interleukin-1

IL-1 α and IL-1β are pro-inflammatory and potent cytokines. It has been stated that the genes of this family are highly polymorphic that modulate the expression of IL-1β. IL-1β −511T/C, −31C/T, and +3593C/T are the most studied SNPs of the IL-1 gene family, in addition to 86bp of VNTR in intron 2 of the IL-1RN gene with 5 different alleles [47]. IL-1β-31C/T is associated with increased transcription initiation factor binding that has been shown to participate in hepatic carcinogenesis [48]. In the presence of IL-1β (−511CC and −31TT) genotypes, allele 2 was known to be a risk for HBV-related hepatocellular carcinoma (HCC). The IL-1β-511C allele was found to be an IL-1RN risk with no HCCC risk [49]. On the other hand, there was no connection between the IL-1β-511C allele and HCC in several studies. In the Wang et al. analysis [50], the IL-1β-31TT genotype and the IL-1β-511/−31CT haplotype were associated with an increased risk of HCC-related HCC and the risk factor was not the IL-1RN VNTR polymorphism.

2.1.2.2 Tumor necrosis factor-α

Polymorphisms of TNF-α (−1031T/C, −863C/A, −857C/T, −308G/A and −238G/A) are found to alter TNF-α development [51]. A critical risk factor for HCCC is regarded as TNF-α −308G/A [45], but TNF-α-238G/A is engineered to play a passive role in the risk of HCC [52]. On the other hand, several other studies indicated an increased risk of HCC for the allele TNF-α-308A [51]. The association between TNF-α promoter alleles such as TNF-α-308A and TNF-α-238A and various TNF-α-238A expressions is indicated in several studies [39].

2.1.2.3 Interleukin-6

In order to counteract inflammatory reactions, IL-6 is a key factor in viral infections. In the promoter area, the most studied genetic polymorphisms in the IL-6 gene are located downstream (−597G/A, −572G/C, −174G/C, and −373A/T), and have an effect on IL-6 development levels at the transcriptional stage [52]. These are reported to be associated with chronic hepatitis, where IL-6-572G/C was found to be associated with the risk of HCC-related HBV [53].

2.1.3 Gastric cancer (GC)

2.1.3.1 Interleukin-1Beta

There are three associated genes in the IL1 cluster: IL1-A, IL-1B, and IL-1RA that encode the signal proteins IL-1, IL-1, and their receptor, IL-1RA, respectively. The association between IL-1B and IL-1RA gene polymorphisms and the development of gastric cancer (GC) has been identified in numerous studies [47]. Polymorphisms of IL-1 are located in positions-511 (CT, rs16944), -31 (TC, rs1143627), and 3954 (CT, rs1143634) that affect the expression of IL-1. The majority of studies classify the IL1β-511 polymorphism T allele as normal among Caucasian individuals with non-cardiac GC but preferably for cancer of the intestinal subtype. It is, therefore, feasible to propose this SNP as a possible GC predictive marker. Even after numerous meta-analyses have been carried out, investigations of the IL1B-31 TATA-box polymorphism appear to be controversial. Around 14 studies [54] reported a slight non-significant correlation between the C variant allele and GC risk in comparison

with TT homozygotes. Again, in contrast with Hispanic or Caucasian cultures, there was no correlation between Asian groups. A modest increase in the intestinal GC subtype among C allele carriers in Caucasian populations was suggested by histologic stratification, but this statement was not true for the diffuse GC subtype. Wang et al. also found that the +3954 gene polymorphism T allele contributes to the GC risk. A lack of interaction between the +3954 polymorphism and GC was stated by Xue et al. [55]. A limited number of studies were conducted on the IL-1β-+3954 C/T polymorphism [55].

2.1.3.2 Interleukin-2

In the differentiation of CD41-positive T cells into Th1 and Th2 effector subsets, IL-2 effectively controls the immune response and plays an important role. IL-2 leads to the activation and transmission of immune responses that are inflammatory, including *H. pylori*-induced gastric inflammation. In the promoter region, IL-2, IL-2-330, and -384 have two types of SNPs that affect IL-2 development [56, 57]. Wu et al. [57], reported that the T allele significantly reduced the risk of gastric cardiac cancer with IL-2-330 polymorphism. Another research, on failed to show a significant association with IL-2-330, while Togawa et al. [58] on the other hand, stated in 2005 that the IL-2-330 T/T genotype in Japan increased the risk of gastric atrophy associated with GC. The IL-2-330 polymorphism yields contradictory findings. In addition, for IL-2-384 and +11.4 polymorphisms and GC growth, no significant association has been seen [59].

2.1.3.3 Interleukin-6

IL-6 appears to be involved in gastric oncogenesis, as serum IL-6 levels show an increase in the gastrointestinal cells and mucosa of patients suffering from GC [60]. The data collected indicated that the polymorphism of IL6-174G/C was related to the risk of GC in the West. The suppression of tumor necrosis factor-α and IL-11 could explain the carcinogenic properties of IL-6 [61]. In 2009, Kang et al. showed a strong negative association among HP-positive cases and controls between the IL6-572G/C polymorphism GG genotype and duodenal ulcer risk. The effect of the G allele on the rate of synthesis of proteins has not been determined to date, and the role of this SNP remains unclear.

2.1.3.4 Interleukin-8

IL-8 improves the proliferation and migration of cells and serves as a chemical attractor and mediates chronic inflammatory processes [62]. In contrast, mucosal levels of IL-8 were found to be elevated in GC patients, and the prognosis was significantly lower in patients with high expression of IL-8 compared to patients with moderate levels of this protein [63]. IL-8 can induce Reg protein expression in stomach cells, which intensifies gastric mucosal cell proliferation and may indirectly promote the initiation of GC [64]. In the IL-8 gene locus, fifteen functional polymorphisms occur, and some can alter gene expression [65]. Several IL8- case-control studies on 251A/T (rs4073), IL8 +396T/G (rs2227307), and IL8+ 781C/T (rs2227306) were conducted in which IL8-251A/T, A allele was associated with increased GCC [66]. Kang et al. [67] also found that higher GC risk is correlated with the AA genotype of HP-positive individuals. Due to conflicting findings, the role of IL-8 gene polymorphisms in GC remains unclear. The −251 polymorphism of

the gene, however, seems to play a major role in the development of GC and needs further study. In this region, the +396 and +781 SNPs are almost unexplored.

2.1.3.5 Interleukin-17A and 17F

IL-17A and IL-17F, synthesized by activated T cells, are characterized as pro-inflammatory. Five variants of this cytokine are known which differ in properties and sites of expression from the founding member IL-17A. Several studies have shown that IL-17 leads to gastric carcinoma growth and progression [68]. It has recently been suggested that a low IL-17 tumor expression rate can imply a poor prognosis in GC patients. Wu et al. reported that in comparison with the mutant AA genotype, the GA and GG genotypes of +7488 SNP were correlated with an increased GC risk and also reported the absence of a link between IL17-197G/A polymorphism and GC risk [69].

2.1.3.6 Tumor necrosis factor-α

One of the pro-inflammatory cytokines strongly expressed in *H. pylori*-induced gastritis is TNF-α, a potent gastric acid secretion inhibitor [70, 71]. While the TNF-α promoter has recorded several polymorphisms, most studies have concentrated on the G/A polymorphism at position 307 since most of the other polymorphisms are functionally silent. In patients with polymorphism at position 307 with malignant tumors, multiple studies find a higher concentration of TNF-α [72]. TNF-α is a pro-inflammatory cytokine, which is a mediator of the immune response in *H. pylori* and shares many biological behaviors with IL-1 [73]. It doubles the risk of gastric cancer without any correlation with the risk of gastric cancer of the esophagus or cardium [74, 75]. Studies done on the effect of TNF-α 307 polymorphisms on the expression of mucosal cytokine and showed no substantial differences in TNF-α level between different allele carriers that indicate that this polymorphism does not affect the development of cytokine. Some race-specific associations have been proposed by another meta-analysis summarizing data on TNF-α, 308 variants, with an increased risk of gastric cancer in various ethnic populations. TNF-α 238 polymorphisms were not substantially associated with the risk of gastric cancer, consistent with previous results [76]. Another research showed that a possible risk factor for gastric cancers is TNF-α 857 T allele.

2.1.4 Prostate cancer

2.1.4.1 Interleukin-1

The relation between the risk of IL-1 family polymorphisms, including IL-1A, IL-1B, and IL-1RN and prostate Cancer (PCa), has been less studied. There was no important correlation between PCa risk and IL-1B/IL-1RN polymorphism in several studies. In a recent meta-analysis, IL-1A, IL-1RN (rs315951 and rs3087263), and IL-1B+3953 (rs1143634) polymorphisms were not significantly correlated with the risk of PCa. In homozygote and recessive models, IL-1B-511 (rs16944) polymorphism was significantly associated with PCa risk, and in the heterozygote model, the allele comparison IL-1B-31 (rs1143627) polymorphism was also marginally significantly associated with PCa risk. Therefore, the meta-analysis indicated that IL-1B-511 (rs16944) and IL-1B-31 (rs1143627) sequence variants were significantly associated with PCa risk. This result presents more new evidence that pro-inflammatory cytokines and inflammation play an important role in the etiology of PCa [77].

2.1.4.2 Interleukin-6

Interleukin 6 (IL 6) plays a crucial role in the inflammatory phase among the cytokines involved in inflammation. Mandić et al., found that IL-6-174 SNP differs between ethnicities and that single polymorphic cytokine variants most likely have little effect on the susceptibility of PCa [78]. The study by Pierce et al. showed that circulating IL- 6 and its gene polymorphism did not affect the risk of PCa [79]. Whereas Mandal et al. had a contrary view [80]. Another meta-analysis of 11 independent studies, including 10,745 cases and 13,473 controls based on several recently published studies which indicated an inconsistent conflicting and inconsistent trends of association between IL- 6 (174 G/C) and PCa. IL- 6 (174 G/C) polymorphism has been found not to be a risk factor for prostate cancer in the general population [81].

2.1.5 Esophageal cancer

2.1.5.1 Interleukin 6

IL-6 (interleukin-6) is a pro-inflammatory peptide that is actively involved in tumorigenesis [82]. The effect of IL-6 and IL-6 receptors (IL-6R) on the prognosis of esophageal cancer (EC) has been identified [83]. Systemic and/or local IL-6 therefore appears to be a central molecule in the stimulation of ESCC progression. Different studies have shown an association between various IL-6 polymorphisms and cancers. Polymorphism of IL-6-634G>C and prognosis after EC esophagectomy revealed that when EC patients were treated surgically, those with the IL-6-634G/G or G/C genotype had a 3-fold poorer prognosis than those with the C/C genotype. In comparison, IL-6R polymorphism and IL-6 tumor expression are not associated with prognosis [84]. With respect to EC, Oka et al. [85] stated that poor survival was associated with high serum IL-6 levels. Buraczyn et al., on the other hand, stated that in the presence of inflammatory stimulation, patients carrying the G allele and to an even greater extent the G/G genotype showed higher IL-6 output at position −634 in the IL-6 promoter region than patients carrying the C allele [86]. Kitamura et al. also reported that the 634 G allele is associated in vitro with increased IL-6 production in peripheral mononuclear blood cells [87]. While IL-6 levels are associated with different cancers, the relationship between IL-6 and EC polymorphisms requires further evaluation because less studies are available [88].

2.1.5.2 Interleukin 12

The pro-inflammatory cytokine family is the Interleukin 12 (IL-12) family, which is essential for host tumor resistance [89]. A few studies have determined whether polymorphisms and serum levels of the IL-12 family (IL-12A gene rs568408, IL-12Bgene rs3212227, IL-27 gene rs153109, rs17855750, rs181206) and its family receptor (IL-12Rb1, 378 C/G) gene are correlated with EC. Studies that revealed polymorphisms of IL-12 rs568408, rs3212227, and IL-12Rb1 gene 378 C/G and serum levels of IL-12p40 and IL-27p28 were significantly associated with the risk of EC [90]. IL-12Rb1 gene Codon 378 C to G causes a transition in amino acid (glycine to arginine), which may further weaken the transcript's subsequent biological activity and is linked to many malignancies such as leiomyoma [91].

2.1.5.3 Interleukin-18

Interleukin-18 (IL-18) is another cytokine that is mainly involved in the inflammatory immune response and is a potent factor triggering IFN-γ. Genetic IL-18 gene polymorphisms have recently been observed to have a significant effect on the vulnerability of a number of inflammatory diseases and various malignancies, including EC [92]. Different SNPs have been identified in the promoter region of the IL-18 gene and are likely to influence gene activity [93]. In the promoter region of the same gene, three SNPs were found at distinct positions −137, 607, and 656 relatives to the transcriptional start site. A study conducted by Ye et al. in the Chinese population investigated 137 G/C and 607 C/A polymorphisms of the IL-18 gene in EC patients. The 137 GC and CC genotypes were associated with a significantly increased risk of ESCC as compared with the −137 GG genotypes as G to C substitution at position 137 abolishes a histone4 transcription factor-1(H4TF-1) nuclear factor-binding site and hence has an impact on IL-18 activity. Nevertheless, in EC patients, the genotype and allele frequencies of IL-18 promoter 607 C/A polymorphism were not substantially different from those in healthy controls. The results, therefore, indicate that IL-18 137 G/C polymorphism could be used as an ESCCC genetic susceptibility marker [94].

2.1.5.4 Tumor necrosis factor-α

Tumor necrosis factor-alpha (TNF-α) is a pro-inflammatory cytokine that plays a major role in host defense and inflammatory responses but also induces cell death and tissue degradation in some instances [95]. There have been reports of dysregulated expression of TNF-α associated with a number of tumors, including EC, [96]. A number of SNPs of TNF-α gene have been found, which include TNF-α-238 G/A (rs361525), TNF-α-308G/A (rs1800629), TNF-α-857C/T (rs179972), TNF-α-863C/A (rs1800630), TNF-α-509C/T (rs1800469) and TNF-α-1031T/C (rs1799964) [83]. Among these, the most common TNF-α polymorphisms occur at position 308 in the promoter region and are extensively studied, showing a strong association with increased TNF-α production [97]. Many studies have focused on the association between TNF-α-308 G>A (rs1800629) and EC risk [98]. Deans et al. have found that the genotype AA TNF-α 308 is associated with an adverse prognosis of gastroesophageal cancer [99]. In contrast, the results of the Cui et al. study found that TNF-α-308G/A polymorphism was not associated with EC risk [100]. In addition, Guo et al. did not find a significant difference between EC patients and controls in the overall genotypic distribution of TNF-α-308G/A polymorphism [101]. In addition, few meta-analyzed studies report that TNF-α-308 G>A (rs1800629) is poorly associated with an increased risk of EC [102]. Conclusions on the role of TNF-α -308G/A gene polymorphism in the risk of EC have, therefore, been inconsistent.

2.1.5.5 Interferon-γ

The cytokine produced against viral and intracellular bacterial infections in the human body is Interferon-gamma (IF-γ). Many polymorphisms have been investigated mostly in IFN-gamma, which are significantly associated with many complications such as susceptibility to many infections [103]. It has been stated that a risk factor for EC may be the IFN-γ + 874AT genotype. The study by Du et al. has shown that IFN-γ + 874, T allele may predispose for EC [104]. Another research found that an important association existed between the genetic polymorphism of

INF-γ 874A>T and infectious complications following esophagectomy in a cohort of EC patients [105].

2.1.6 Pancreatic cancer

2.1.6.1 Interleukin-1 beta

Interleukin 1b (IL1b), a central IL-1b gene-encoded pro-inflammatory cytokine, has been associated with chronic inflammation and plays an important role in inflammatory pancreatic diseases, including pancreatic cancer [106]. In the production of pNETs, the functions of IL-1b −511C/T and +3954 C/T genotypes remain unclear. PNETs are a heterogeneous group with different biology and prognosis of tumors; they can occur as solitary tumors, and up to 15% of pNETS are part of hereditary syndromes. The findings of Maja et al. [107] showed a significant correlation between the IL-1b −511C/T genotype and CTCT −511/+3954 genotype combination and susceptibility to functional pNET growth, and the risk of non-functional pNET development were associated with patients with CTCC −511/+3954 genotype combination. All of these findings indicate IL-1b participation in the growth of pNET [107].

2.1.6.2 Interleukin-2

In T cell-dependent immunity, IL-2 is a potent pro-inflammatory cytokine and a central regulatory cytokine. Several SNPs) have been found to alter IL-2 production in the promoter region and to be associated with inflammation-based cancer [58]. The G-allele in the IL-2 promoter at the −330 position seems to correlate with higher IL-2 output [108]. It has been shown that during the cell cycle, dividing cancer cells release IL-2 in various concentrations and that IL-2 promotes cancer development [109]. Hofsli and collaborators found that GEP-NET carcinogenesis growth factors downregulate the IL-2 receptor, presenting proof of an additional mechanism through which neuroendocrine cancer cell growth is promoted [110]. It was found that G-allele raises the expression of IL-2 at the −330 position and that individuals homozygous for the G-allele have three times higher IL-2 values [108].

2.1.6.3 TNF-α

In the inflammatory etiology of pancreatic cancer, TNF-α plays a very crucial role. The risk factor for various forms of cancer, such as hepatocellular carcinoma, gastric cancer, and breast cancer, has been identified as TNF-α-308 polymorphism [111–113], while previous studies have reported that polymorphisms in the TNF-A-308 A/G gene are not linked to pancreatic cancer [78].

2.1.7 Bladder cancer

2.1.7.1 Interleukin-6

IL6 is an inflammatory pleiotropic cytokine released by different types of lymphoid/non-lymphoid cells [114]. Immune response, cell survival, proliferation, and apoptosis are critical for [113]. In the IL-6 promoter region, numerous polymorphic variants have been identified that are associated with IL-6 transcription activity [115] which indicates a connection to cancer [116].

IL-6 polymorphism has been shown to modulate changes in its expression to cause cancer risk, including bladder cancer. Association studies of IL-6 gene

polymorphisms conducted in India and globally confirm the risk of several CC genotype and C allele cancers, especially BCC allele cancers [117, 118]. Fishman et al. have stated that IL-6-174 G/C polymorphic variation affects transcription and IL-6 protein expression [119]. Conversely, several studies from other parts of the Indian and Caucasian population indicated that conflicting findings were present [120, 121]. The variant genotype IL-6-174 G/C was shown to be significantly associated with an increased risk of BC [122] but no BC risk association was found by other authors [123]. In addition, there was no proven association between IL-6-572 G/C, -596 A/G polymorphisms, and BC risk [118].

2.1.7.2 TNF-α gene polymorphisms

A major inflammatory cytokine is the tumor necrosis factor-alpha (TNF-α) gene, which mediates a connection to all steps involved in tumor growth and development. TNF-α gene polymorphisms and their receptors have conceptualized the understanding of the genetic effects of inflammatory consequences [124]. Polymorphic sequence differences are mainly documented in association studies of various cancers like BC, with some studies refuting and other studies showing the association for cancer risk, mostly single nucleotide changes in TNF-α promoter [125, 126]. This contentious link between TNF-α polymorphisms and UBC is due to the race, sample size, and technical aspects involved. Of the 7 variants of the TNF-α gene studied by Marsh et al., 859T and +488A polymorphisms were significantly linked to BC risk [127]. Another research investigates 3 TNF-al-1031 T>C polymorphism SNPs where high risk was identified for controls of BC cases [128]. Similarly, experiments on TNF-al-308 G>A polymorphism have also reported controversial results [129, 130]. Although TNF-α-308 A variant alleles with a major risk relationship to BC risk were observed by Lima et al. [131].

2.1.8 Gliomas

2.1.8.1 IL-4 gene polymorphism

Research by Brenner et al. found that the polymorphism of IL-4 (rs2243248, −1098T>G) was substantially correlated with the overall risk of glioma [132]. Another study showed that IL-4 induced aberrant Stat3 activation in glioblastoma cells but not in normal human astrocytes, and hypothesized that aberrant Stat3 activation induced by IL-4 could contribute to the pathogenesis of GBM cells [133].

IL4R encodes the interleukin-4 receptor alpha chain that can connect interleukin 4 and interleukin 13 to regulate the development of IgE [134]. In 2013, TianboJin et al. reported that rs1801275 in the IL-4R gene can predict the over-dominant model of Glioma susceptibility by 2.29-fold [135]. Furthermore, another article also stated that rs1801275 could increase the risk of studying glioblastoma [136]. However, an important link between mutant IL-4R alpha rs1801275 and gliomas was not found by Ruan et al. [137].

2.1.8.2 IL-12gene polymorphism

Many SNPS, like 1188A/C, in the IL-12 gene will affect its levels of expression and are associated with several tumors such as nasopharyngeal carcinoma [138] and breast cancer [139]. This SNP might influence the susceptibility of individuals to glioma. Haidar et al. stated that neither mutant (hetero or homozygous) genotypes nor mutant IL-12p40 1188A/C variant allele appears to be associated with gliomas

[140]. In line with a meta-analysis review, this outcome [141]. The stratification by cancer type showed that the variant only represents a risk factor for cervical and nasopharyngeal cancer, although these authors found a substantial association of the variant with overall cancer.

2.1.8.3 IL-13 gene polymorphism

IL-13 plays an important role in allergies and is essential for the suppression of tumor immune surveillance through an immune regulatory pathway [142]. SNPs, including rs25041, rs1800925, and rs1295686, found in the IL-13 gene, are closely linked to IL-13 expression. In human malignant glioma cell lines and in primary tumor cell cultures, IL-13 has been shown to be over-expressed [143]. The serum immunoglobulin E (IgE) level could be significantly increased by the rs20541 A allele (glutamine (Gln, Q) form) in different populations [144]. Evidence has also been established by previous research that IgE levels in patients with glioma were lower than in people without glioma [136]. It is very fair to expect the IL-13 gene rs20541 polymorphism to exert some effect on glioma susceptibility due to its important roles in immune surveillance and serum IgE level modulation. A recent meta-analysis review, however, indicates that the IL-13+rs1800925 genotype could be a risk factor for gliomas and that IL-13+rs20541 contributes to cancer [145] (**Table 1**).

2.2 Hematological malignancies

2.2.1 Lymphomas

2.2.1.1 IL-4 and IL-5 gene polymorphisms

In the proliferation of T cells, Il-4 and Il-5 play a key role. The study of the relationship between SNPs was chosen from the main immunological genes of cytokines and NHL in several studies. This represents a pathway-based approach to the investigation of common genetic variants in the cytokine network of Th1 and Th2. Common genetic variants of the Th2 cytokine genes have been observed and are associated with NHL risk. SNPS has been shown to be substantially associated with an increased risk of NHL in the TH2 genes IL-4 (−1098T>G) and IL-5 (−745C>T) [165].

2.2.1.2 IL-6 gene polymorphism

While monocytes are the main source of IL-6, many cells, including dendritic cells, lymphocytes, neutrophils, mast cells, mesenchymal cells, and tumor cells, may produce it. IL-6 rs1800797 (IL-6 rs1800795G>C, rs1800796G>C, rs1800797G>A) was the only SNP to demonstrate substantial survival outcomes, with DLBCL co-dominant model (GG/AG/AA) and recessive model (AA genotype versus combined GG/GA genotype) subjects having worse overall survival. The correlation between polymorphism of the IL-6 gene promoter and the risk of lymphomas, however, shows inconsistent results. A common case-control study conducted by the International Lymphoma Epidemiology (Inter Lymph) consortium in 2006 to examine the relationship between gene polymorphisms and the risk of lymphoma showed that there is no correlation between IL-6 promoter polymorphism (174G>C rs1800795) and the risk of NHL [166].

Activity	Cancer type		Cytokines	Polymorphisms	Association	Allele	References
Pro-inflammatory	Solid tumors	Breast	IL-6	−174 G/C	High grade tumor	−174 CC	[25]
			IL-10	21082 G>A	Worst prognosis	−21082 AA	[28]
					High risk	−21082 AA	[29]
			IFN-γ	+874T/A	Increased risk	+874 AA	[36]
				−1615C/T	No	—	[36]
			TNF-α	−308G/A	Poor prognosis	−table308GA, AA	[40]
					No	—	[43]
		HCC	TNF-α	1031T>C	No	1031T>C	[146]
				863C>A	Yes	863C>A	[146]
				857C>T	No	857C>T	[146]
				308G>A	Yes	308G>A	[146]
			IFN- γ	+874T>A	Risk factor	+874T>A	[147]
			IL-18	607A>C	Risk factor	607A>C	[148]
				137C>G	Risk factor	137C>G	[149]
				148G>C	Risk factor	148G>C	[150]
			IL-16	rs11556218T>G	Risk factor	rs11556218T>G	[151]
				rs4072111C>T	Risk factor	rs4072111C>T	[151]
				rs4778889	Risk factor	rs4778889	[152]
			IL-12A	rs3212227A>C	Risk factor	rs3212227A>C	[153]
			IL-12B	rs2243115T>G	Risk factor	rs2243115T>G	[147]

Activity	Cancer type	Cytokines	Polymorphisms	Association	Allele	References
	Gastric	TNF-α	TNF-α 307	Race sp. asso.		[150]
			TNF-α 307, 1031, 863, 857, and 238	Not favoring		[154]
						[75]
			TNF-α 307	Not favoring		[155]
		IL-1β	IL1β –31 T/C	Slight non sig	C allele	[155]
			IL1β –511C/T	Significant association	TT	[94]
				Association	T allele	[156]
				Association	T allele	[55]
			IL1β +3954 C/T	Lack of association	T allele	[55]
				Association	T allele	[156]
		IL-6	Il-6 –174 G/C	Association	GG	[157]
			IL-6 –572 G/C	Negative association	GG	
			IL-6 –597 G/A	Lack of association	G allele	[158]
		IL-8	IL8 –251A/T,	Lack of association	TA	[158]
			IL8 +396T/G,	Lack of association	GG	[158]
			IL8+781C/T	Lack of association	CT	[157]
		IL-2	IL-2 -330	Sig reduced risk	T allele	[57]
				Non-sig association	T allele	[159]
				Sig association	TT genotype	[58]
			IL-2 –384G/T, +114G/T	Non-sig association		[59]
		IL17A and IL17F	–197G/A	Lack of association	AG	[4]
			+7488A/G	Association	GG	[4]

Activity	Cancer type		Cytokines	Polymorphisms	Association	Allele	References
		Prostate	IL-6	IL-6 −174	Least association		[78]
			IL-1	IL-6	No. assoc.		[79]
				IL-6	Association		[80]
				IL-6 −174 G/C	No association		[81]
				IL-1α/IL-1RN/ILβ +3953	No association		[77]
				IL-1β-511/IL-1β-31	Association		[77]
		Esophageal	IL-6	−634	Poor prognosis	−634GG	[84]
			IL-12A	rs568408 G>A	Susceptibility	rs568408 GA,GG	[160]
			IL-12B	rs3212227A>C	Significant risk	rs3212227AC,CC	[160]
			IL-12Rb1	−378	Susceptibility	−378 GG	[160]
			IL-18	−137 G/C	Increased risk	−137 CC,	[94]
				−607 C/A	No	−607 C/A	[94]
			TNF-α	−308	Adverse prognosis	−308 A/A	[99]
			IFN-γ	+874A>T	Protective	+874AT,TT	[104]
		Pancreatic	IL-1β	−511C/T, −511/+3954	Favouring	CT genotype	[106]
			IL-2		Favouring	CT genotype	[107]
			TNF-α-308	IL-2 −330	Favouring	G allele	[108]
			A/G	TNF-α-308 A/G	Not favouring	AG	[69]
		Bladder cancer	TNF-α	1031T>C	No	1031T>C	[161]
				−857C>T	Yes	−857C>T	[161]
				−308G>A	Yes	−308G>A	[161]
			IFN-γ	IFN-γ +874	Yes	IFN-γ +874	[162]

Activity	Cancer type		Cytokines	Polymorphisms	Association	Allele	References
			IL-1β	−511C>T	No	−511C>T	[123]
			IL1-RN	VNTR	Yes	—	[163]
		Glioma	IL-4	−588C>T	Suggestive association	−588TT	[132]
				−1098T>G		−1098GT,GG	[132]
			IL-4R	+828A>G	Overall risk	+828A>G	[164]
					Overall risk	—	[137]
			IL-12	−1188A>C	No	—	[140]
			IL-13	+2044A>G	No susceptibility	+2044AA,AG	[136]

Table 1.
A meta-data of different pro-inflammatory cytokine polymorphisms and their association in various solid tumors.

2.2.1.3 IL-10 gene polymorphism

A number of studies have evaluated the association between specific polymorphisms of IL-10 genes and the risk of non-Hodgkin lymphoma but fewer studies analyzed the impact of IL-10 polymorphisms on the prognosis of patients with DLBCL. There is clear evidence that in hematologic malignancies, immune-regulatory cytokines play a major role. The relationship of seven single nucleotide polymorphisms (SNPs) in two selected cytokines (IL-6 rs1800795 G C, rs1800796 G>C, rs1800797 G>A, IL-10 rs1800871 G A, rs1800872 G>T, rs1800890 A>T, rs1800896 T>C) to risk and overall survival was examined in different studies [167].

2.2.2 Leukemias

2.2.2.1 IL-15, -18 and-1β gene polymorphism in acute lymphoblastic leukemia (ALL)

The proliferation of T, B, and NK cells is encouraged by IL-15, a pro-inflammatory cytokine. Some studies have previously shown that IL-15 can protect hematologic tumors from drug-induced in vitro apoptosis, and high expression of IL-15 is associated with childhood ALL CNS disease. Several minimal residual diseases (MRD)-related IL-15 SNPs have been shown to have a link to increased in vitro transcription/translation efficiency of IL-15. As found by the GWAS scan, there are 5 SNPs in the IL-15 locus that are significantly correlated with childhood ALL therapy response. In the most recent study, polymorphisms of the IL-15 rs10519612 CC genotype have been shown to be associated with adult ALL1 [69]. Polymorphisms of IL-1β (rs16944) and IL-18 (rs1946518) have been shown to predict the prognostic consequences and manage ALL. The impact of the rs1884444 sequence variant on relapse rate and connection of rs10889677 AA genotype with favorable prognostic factors recommend the influence of the investigated SNPs on ALL response to treatment and outcome.

2.2.2.2 Other gene polymorphisms in acute and chronic leukemia (AML and CML) and chronic lymphocytic leukemia

Polymorphisms of IL-1β (rs16944) and IL-18 (rs1946518) have been shown to predict the prognostic consequences and manage ALL care. In cytokine genes, several SNPs have been identified, indicating that certain alleles may lead to alterations in cytokine production [168]. It is, therefore, hypothesized that variants of the cytokine gene can influence the expression of genes and may be associated with leukemia pathogenesis.

The relation between the polymorphism of IFNγ +874T>A (rs2430561) and the risk of CML [169] or earlier, CLL was assessed. Various studies have suggested that the IFNγ+874T>A polymorphism leads to the susceptibility of CML and CLL [168]. No association between TGFβ1 rs1800470 polymorphism and leukemia was identified in earlier studies. Nursal et al., on the other hand, found that variants of the genes TNF-α rs361525, IL-10 (−1082G>A rs1800896, −819C>T rs1800871, −592C>A rs1800872) and TGF-β1 (codon 25) may have a significant association with AML etiopathogenesis [170]. The relation between TNF-α 308G>A, IL-10 (−592T>G, −819T>C, −1082T>C), IFN-γ +874T>A and TGF-β1 (codons 10 and 25) was not identified to confer any risk to CML [169].

2.2.3 Myelomas

2.2.3.1 IL-1gene polymorphism

IL-1 is a potent cytokine that is pro-inflammatory and functions as an endogenous pyrogen. IL-1 has two cytokines formed by two diverse genes, IL-1 alpha, and IL-1β. Cytokine plays a crucial role in the development, differentiation, and function of different immune cells in cells. An important interaction between IL-1 alpha-889C>T and IL-1β-3737C>T and the risk of MM was shown to have an association with polymorphisms of IL-1 [171].

2.2.3.2 IL-6 gene polymorphism

As a proliferative factor for multiple myeloma (MM), IL-6, a cytokine with broad inflammation and immunity functions, is identified. Some myeloma cells and bone marrow stromal cells may produce IL-6 and can restrain apoptosis in myeloma [172]. Previous studies have shown that IL-6 expression can be partly genetically modulated in the promoter region of IL-6 by polymorphisms located at the rs1800795 location. Recent studies have shown that the association of IL-6 rs1800795 polymorphism with MM risk is negligible.

2.2.3.3 TNF-α gene polymorphism

Gene polymorphisms of TNF-α may also be essential for its functional variation. Research has shown that a TNF-α allele (−308) has been expressed at lower levels in MM subjects, indicating that an allele can have a protective effect against disease [173]. However, the GG genotype of TNF-α (−238) was shown to be correlated to early progression in MM in another study [174].

2.2.3.4 IL -12 gene polymorphism

IL-12 is a cytokine that stimulates immunity that is both innate and adaptive. This induces cytotoxicity of Th1 cells and has been shown to have effective immunomodulatory and anti-tumor activities [175]. However, while IL-12 is an inflammatory cytokine, protection from neoplastic disease appears to be the prevalent activity of the cytokine in this case. There is no clear link between IL-12 (rs1801131) (A>C), polymorphism, and the risk of MM in recent studies [176] (**Table 2**).

3. Anti-inflammatory cytokine gene polymprphisms

3.1 Solid tumors

3.1.1 Breast cancer

3.1.1.1 IL-1β gene polymorphism

A variety of cell types, including monocytes, macrophages, and epithelial cells, are formed by interleukin-1β (IL-1β) belonging to the IL-1 family and have multiple biological effects [190]. IL-1β induces the expression of pro-inflammatory genes that may play a key role in the early stages of carcinogenesis, such as cyclooxygenase type 2, inducible nitric oxide synthase, and other cytokines/chemokines. A number of studies have documented the association of IL-1β polymorphisms with

Activity	Cancers type			Cytokines	Polymorphisms	Association	Allele	References
Pro-inflammatory	Hematological cancers	Lymphomas	HL	IL-1	+4345 T>G	Significant	−TG	—
				IL-6	−174 G>C	Significant	174 GC	[177]
				TNF-α	−863/308 C>A/G>A	Not significant	−863/308 CA,GA	[177]
			NHL	IL-6	rs1800795 G> C	Not significant	−174GC	[166]
					rs1800796 G>C	Significant	−598GA	
				TNF-α	rs1800629 G>C	Significant	−308GA	[157]
				TNF-β	rs909253 A>G	Significant	−252AG	[178]
				IL-4	rs2243250T>G	Significant	−1098TG	[178]
				IL-5	rs2069812 C>T	Not significant	−745CT	
		Leukemias	ALL	IL-1β	rs16944 G>A	Significant	TT	[179]
				IL-18	rs1946518	Significant	TT	[180]
				IL-15	rs10519612	Significant	CC	[181]
				IL-6	174G/C	Significant	rs1800795 GC	[182]
				IL-4	590C/T	Not significant	rs2243250 CT	[170, 183]
			AML, CML, CLL	IL-1α	889C>T-	Significant	889CT	
				IL-1β	3737C>T	Significant	3737CT	[184]
				IL-2	rs1800795 G>C	Significant	174/52GC	[185]
				IL-6	rs1801131C>G rs361525G>A	Not Significant	CG	[184]
				TNF-α	rs2430561T>A	Significant	-308/238 GA	[186]
				IFN-γ	889C>T-	Significant	+874T TA	

Activity	Cancers type	Cytokines	Polymorphisms	Association	Allele	References
	Myelomas	IL-1 α	889C>T	Highly significant	CT	[187]
		IL-1β	3737C>T	Highly significant	CT	[187]
		IL-6	174/52G>C	Significant	rs1800795 GC	[188]
		IL-12	rs1801131 A>C	Not significant	rs1801131AC	[185]
		TNF-α	−308/238 G>A	Significant	rs1800629-308/238 GG	[189]

Table 2.
A meta-data of different pro-inflammatory cytokine polymorphisms and their association in various liquid cancers.

the risk of breast cancer, establishing the important role of IL-1β in its development. The findings are contradictory due to the relatively small sample size of the study. Ito et al. [191] first recorded that rs1143627 is significantly correlated with the risk of breast cancer (CC vs. TT: OR = 1.82). The correlation in the Chinese population was followed by another case-control analysis by Liu et al. (CC vs. TT: adjusted OR=1.72). All the studies for rs1143627 included in the meta-analysis were conducted in Asian populations, suggesting that rs1143627 could contribute to the risk of breast cancer, especially in Asians [34].

3.1.1.2 TGFβ -1 gene polymorphism

A multifunctional cytokine that is important for maintaining homeostasis involving bone and muscle differentiation, immune response, and tumor suppression is the transforming growth factor beta-1(TGFβ-1). Increased TGFβ-1 development occurs in different tumor types and is associated with tumor grade severity [192]. There is evidence that when the proliferative inhibition effect of the TGFβ-1 signaling pathway has been overridden by other oncogenic mutations, TGFβ-1 functions as a suppressor of tumor initiation but as a promoter of tumor progression [193]. The 29T/C transformation that generates a Leu10Pro replacement in the TGFβ-1 precursor signal peptide has been reported to be associated with TGFβ-1 secretion of the protein and thus may have altered the risk of breast cancer. In the published clinical trials, Dunning et al. [194] concluded that in studies (20,837 patients and 22,879 controls) from European countries, TGFβ-129T/C T polymorphism was strongly correlated with breast cancer risk (C/C versus T-carrier, 1.21; 95% CI 1.05–1.37); Hishida et al. [195] found that the C/C genotype was substantially correlated with a decreased risk of breast cancer relative to the T/T genotype (OR = 0.45, 0.20–0.98); polymorphisms are not likely to be associated with major increases in the overall risk of breast cancer among Caucasian women in the promoter region of TGFβ-1. Another meta-analysis review indicated that the 29T/C polymorphism did not contribute to breast cancer risk in both recessive, dominant, and other genetic models [196].

3.1.2 Hepato-cellular carcinoma

3.1.2.1 IL-10 gene polymorphism

During chronic viral infections, IL-10, an immune suppressor cytokine, plays a vital role in the weakened host immune response, where its secretion is stated to be genetically regulated [197]. The most studied SNPs that have an influence on IL-10.1 development are IL-10 (−1082G/A, −819T/C, and −592A/C). The reports suggested that −1082 G allele and GCC haplotype are correlated with high IL-10 expression, where less expression association is shown by ATA haplotype. On the other hand, studies have shown that IL-10GCC and IL-10 haplotype AAGCC (−3575, −2763, −1082, −819, and −592) are associated with lower IL-10 haplotype concentration [198].

3.1.2.2 TGF-β1 gene polymorphism

In addition to +869T/C (Codon 10), +915G/C (Codon 25) and +788C/T (codon 263), TGF-β1 polymorphisms are identified in TGF-β1-988C/A, -800G/A, -509C/T and in insertion/deletion [199, 200]. Several SNPs are identified in TGF-β1 (+869T/C and +915G/C) [201] while TGF-β1 +869T allele was found to be correlated with high TGF-β11 expression [202].

3.1.3 Gastric cancer

3.1.3.1 IL-4 gene polymorphism

IL-4 is also recognized as a pro-inflammatory cytokine suppressor and an enhancer of anti-inflammatory cytokine synthesis [203]. The polymorphism of the promoter IL4-590C/T (rs2243250) is widely studied because the mutant T allele has been documented to increase IL-4 expansion compared to the C allele [204]. In contrast with the TT genotype, Wu et al. found an increased risk of developing diffuse form and cardiac GC for the CT/CC genotype in 2003. It is also proposed that lower IL-4 levels in the gastric mucosa favor the growth of GC. The genetic polymorphism IL4-168T/C (rs2070874) tested by Wu et al. [205] stated that a reduced GC risk among the Chinese population was correlated with mutant C allele.

3.1.3.2 IL-10 gene polymorphism

As an immune suppressor and anti-inflammatory mediator, IL-10 typically acts. From the transcriptional start site, there are 3 functional promoters SNPs at-1082 (A to G, rs1800896), -819 (C to T, rs1800871), and -592 (A to C, rs1800872) pairs in the IL10 locus. Studies of IL10-1082A/G, IL10-592A/C, and IL10-819C/T polymorphisms indicate the absence of major discrepancies in the distribution of genotypes between GC patients and healthy controls [206–208].

3.1.4 Prostate cancer (PCa)

3.1.4.1 IL-10 gene polymorphism

The most studied cytokine for PCa risk is IL-10. Controversial findings have emerged in this regard, where various studies show conflicting results and the latest meta-analysis showed that IL-10 (rs1800871) polymorphism was not correlated with PCa danger [209]. On the contrary, another meta-analysis indicated that polymorphic variants of IL-10 (rs1800871) and IL-10 (rs1800872) reinforce the idea that these may be moderately correlated with advanced PCa and thus affect disease progression [210]. The 2011 study by Shao et al found no substantial evidence suggesting variations in allele frequency or genotype distribution between PCa patients and control subjects for any of the three SNPs IL-10-1082 A>G, -819 C>T and -592 C>A.

3.1.5 Esophageal cancer

3.1.5.1 IL-1β gene polymorphism

In cell proliferation and apoptosis, interleukin 1β (IL1β) is involved which is an essential mediator of inflammatory response. IL1β induces COX-2 expression, inducible synthase nitric oxide, and other cytokines/chemokines, which can function significantly in the early stages of carcinogenesis. Many cancers such as gastric cancer and inflammatory bowel disease are associated with IL1β polymorphisms [211]. One of the polymorphisms associated with IL1β levels in the promoter area is rs16944 G>A [212]. According to Liang et al., IL1ß rs16944 GA variant heterozygotes had a substantially decreased chance of ESCC rather than IL1ß rs16944 AA homozygotes, in line with the Thai population study showing that IL1ß rs16944 GG genotype was found to be more common in patients with gastric cancer [213, 214]. However, minimal studies have been performed that can provide us with

clear proof of the functionality in the risk of ESCC of IL1ß rs16944 G>A polymorphism.

3.1.5.2 TGF-β1 gene polymorphism

Two main components of TGF-β signaling that play an important role in carcinogenesis are the transforming growth factor β1 (TGF-β1) and its TGF-β1RII receptor. Several functional polymorphisms were observed in TGF β1 and TGF β1 RII and were correlated with elevated TGF-β1 serum or plasma levels and increased TGFβ1RIII transcription activity. Epidemiological evidence has shown that two transforming growth factor-beta 1 (TGF-β1) gene polymorphisms (namely rs1800468G>A and rs1800471G>C) could be involved in the production of cancer. Their function in the carcinogenic process of esophageal squamous cell carcinoma (ESCC) has, however, been less well described. In recent studies, two polymorphisms of this gene (namely rs1800468G>A and rs1800471G>C) have been shown to be associated with TGF-β1 dysfunction and increased tumor risk [215, 216]. Several evidence indicates that ESCC risk may be increased by the genotypes with rs1800471 C alleles and rs1800471G>C polymorphism may play an important role in ESCCC tumorigenesis [217]. There were some drawbacks to this analysis, however. In contrast to the above findings, several studies have shown that RS#1800468 polymorphism is not associated with an increased risk of ESCC, but with a reduction in the survival of this tumor, consistent with other studies of various cancers [216, 218].

3.1.6 Pancreatic cancer

3.1.6.1 TGF-β1 gene polymorphism

TGF-β functions not only as a powerful epithelial, endothelial and hematopoietic cell proliferation inhibitor but also acts as an effective pro-inflammatory cytokinetic cell proliferator. In cellular proliferation, angiogenesis, differentiation, migration, and apoptosis, TGF-β pathway has important roles. TGF-β expression levels have been reported to correlate with the period of postoperative survival in various malignancies [219]. A study by Zhang et al. [220] showed the TT genotype to be more common in patients with leakage of pancreatic anastomosis in TGF-β. In addition, TT genotype cases had an increased level of bilio-digestive anastomosis leakage.

3.1.7 Bladder cancer

3.1.7.1 IL-6 gene polymorphism

Interleukin-6 (IL6), a pleiotropic inflammatory cytokine released by different types of lymphoid/non-lymphoid cells, is important for immune response, survival of cells, proliferation, and apoptosis [113]. In the IL-6 promoter region, different polymorphic variants have been identified that are associated with IL-6 transcription activity [115], IL-6 influences the release of acute-phase proteins in the acute inflammatory response and regulates the anti-inflammatory cytokines, thereby influencing the strength of inflammatory response. IL-6 is active in multiple cancer growth pathways and promotes neo-angiogenesis. IL-6 polymorphism has been shown to modulate changes in its expression to induce cancer risk, like bladder cancer. Association studies of IL-6 gene polymorphisms conducted in India and globally support the risk of CC genotype and C allele in many cancers, especially

BC. Fishman et al. have stated that IL-6-174 G/C polymorphic heterogeneity affects transcription and IL-6 protein expression. The variant genotype of IL-6-174 G/C has been shown to be substantially associated with an increased risk of BC but other authors have not identified any risk association. In addition, there was no proven association between IL-6-572 G/C, -596 A/G polymorphisms, and BC susceptibility [117].

3.1.7.2 IL-4 gene polymorphism

Interleukin-4 (IL-4), an anti-inflammatory cytokine produced mainly by activated CD4+T cells, plays a significant role in the production of Th2 to examine and destroy distorted cells and eradicate extracellular pathogens [221]. The risk conferred by IL-4-590 C/T polymorphism in different cancers has been substantiated by studies from different quarters [222]. In order to modulate the incidence of BC, a polymorphic variance of IL-4-590 C/T was found and another analysis similarly observed a substantial difference in variant allele distribution between cases and controls [223].

3.1.7.3 TGF- β gene polymorphism

TGF-β1 is the most abundant type of TGF-β, which includes numerous polymorphic variants that regulate the expression of TGF-β1 proteins. The risk association of TGF-β1 polymorphisms in a number of cancers has been verified. TGF-β1 was shown to be associated with the possibility of BCC in 41 SNPs. In comparison, 3 polymorphisms in TGF-β1 and 4 in the TGFβR1 gene showed no correlation yet another study reported comparable outcomes of negligible interaction with BC [210]. Gautam et al., on the other hand, observed an important association between TGF-β1 c.29 C/T and the risk of BCF [118].

3.1.8 Gliomas

3.1.8.1 TGF-β gene polymorphism

TGF-β functions as an oncogenic factor that contributes to the growth and invasion of cells and lowers host tumor immune responses [224]. The −509C/T TGF-β1 gene polymorphism can theoretically control the transcription of TGF-β1. The 869T/C polymorphism may have decreased cancer survival in patients carrying the −509C/T T allele and in patients with the 869T/C C allele, as both of these alleles are associated with elevated levels of TGF-β1. In comparison, TT genotype-carrying glioma patients and CC genotype patients have longer average survival. Therefore, in patients with glioma, the TT genotype of the −509C/T polymorphism and the CC genotype of the 869T/C polymorphism have the ability to be used as predictors of improved survival.

3.1.8.2 IL-10 gene polymorphism

IL-10 has pleiotropic effects on inflammation and immunoregulation and can facilitate carcinogenesis [54]. In gliomas, several IL-10 gene polymorphisms have been shown to influence disease susceptibility and severity [134]. The association of the IL-10 rs1800871 C/T genotype with increased survival in low-grade glioma patients was confirmed by Mingjun et al. [225]. Another research found an important protective interaction of variant IL-10 (−1082A/G) G allele inst glioma whereby It has been shown to induce an increase in the development of

IL-10 In 2012, Tanikawa et al. [226] noted that high serum IL-10 levels increased tumor-specific immune response and decreased tumor growth. The exact role of IL-10 in glioma is unclear; further studies are needed to elucidate the mechanisms underlying the relationship between IL-10 polymorphisms and the prognosis of glioma patients (**Table 3**).

3.1.9 Lymphomas

Studies regarding the role of interleukin gene polymorphisms in Hodgkin's Lymphoma have revealed anti-inflammatory functions of IL-1R and IL-10 [240, 241].

3.1.10 Leukemias

3.1.10.1 IL-10 gene polymorphism in ALL

In cancer growth and progression, interleukin-10, a pleiotropic cytokine serves as an immune stimulant factor. In ALL IL-10 SNP rs1800896 was correlated with the progression of the disease and also affect the cytokine expression. Many experiments, however, studied IL-10 rs100896 T/C polymorphism to determine the relevant relationship with susceptibility and prognosis.

3.1.11 Myelomas

3.1.11.1 IL-1Ra gene polymorphism

In response to similar stimuli that induce IL-1 release, the IL1-Receptor Antagonist (IL-lRa) is developed and released. According to Liang Zheng et al., carriers of IL1B rs16944 GA variant heterozygotes had a slightly reduced chance of multiple myeloma compared to IL-1B rs16944 AA homozygotes, had a slightly reduced [213].

3.1.11.2 IL-10gene polymorphism

IL-10 suppresses immune responses as an immunosuppressive cytokine by functioning simultaneously on the innate and the adaptive immune system. IL-10 can also inhibit pro-inflammatory cytokine secretion, antigen presentation, and cell growth. In the pathogenesis of hematological disorders, both IL-10 and IL-10R SNPs are involved. IL-10-592G/A and IL-10-1,082G>A SNPs, -592(C) or -1,082(G) are considered to be correlated with a strong IL-10 expression and a low -592(A) or -1,082(G) expression (A). However, studies have shown, that the C allele of IL10-592 has no clear interaction with MM [242].

3.1.11.3 TGF-β gene polymorphism

TGF-β is an active regulatory cytokine with divergent hemopoietic cell effects. Usually, TGF-β serves to reduce the release of immunoglobulin by B cells. Studies have shown that TGFβ1 genotypes with rs1800471 C alleles may raise the risk of MM and rs1800471G>C polymorphism may play a significant role in MMM tumorigenesis [217].

Activity	Cancer type		Cytokines	Polymorphisms	Association	Allele	References
Anti-inflammatory	Solid tumors	Breast	IL-1β	−511C>T	High risk	−511 CC	[23]
			TGFβ1	Codon 29T>C	Significant risk	29TT	[194]
					No	—	[196]
		HCC	IL-1β	511T>C	Risk Factor	511T>C	[49]
			IL1-RN	Intron 286bp	Risk Factor	Intron 286bp	[47]
				Allele 2 VNTR		Allele 2 VNTR	[227]
			TGF-β1	509C>T	Risk Factor	509C>T	[228]
			IL-4	2590C>T	Risk Factor	2590C>T	[229]
				233C>T	Risk Factor	233C>T	[230]
			IL-6	−597G>A	No	−597G>A	[231, 232]
				−572G>C	Risk factor	−572G>C	[54]
				−174G>C	Risk factor	−174G>C	[233]
			IL-10	−1082G>A	Risk factor	−1082G>A	[156]
				−819T>C	Risk factor	−819T>C	[156]
		Gastric	IL-4	IL-4 –590C/T	Asso	T allele	[234]
				IL-4–168T/C	Neg assoc	C allele	[234]
			IL-10	IL-10–1082 A/G	Lack of assn	AA	[206]
				IL-10–819 C/T	Lack of assn	TC	[235]
				IL-10–592 A/C	Lack of assn	AC	[207]

Activity	Cancer type	Cytokines	Polymorphisms	Association	Allele	References
	Prostate	IL-10	IL-10 (rs1800871)	Not favouring		[209]
			IL-10rs1800871)/(rs1800872)	Favouring		[210]
			IL-10 −1082 A/G	No association		[210]
			−819 C/T	No association		
			−592 C/A	No association		
	Esophageal	IL-1β	rs16944 G>A	Protective	rs16944 GA	[213]
		TGFβ1	rs1800468G>A +915G>C	Significant risk	rs1800468AA	[216]
				Significant risk	+915 CC	[227, 236]
	Pancreatic	TGFβ	TGF-β	Association	TT genotype	[237]
	Glioma	TGFβ1	−509C/T,869T/C	Better prognosis	−509TT,869CC	[238]
		IL-10	−819 C>T	Improved survival	−819 CC	[225]
			−1082A/G	Protective role	−1082GG	[239]

Table 3.
A meta-data of different anti-inflammatory cytokine polymorphisms and their association in various solid tumors.

4. Conclusion

Germline constitutional mutations or changes in the polymorphic sequence are key instruments used as molecular tools to predict the susceptibility of a person to risk numerous cancers. Cytokines are also thought to be essential molecules that have a dual face due to pro-inflammatory as well as anti-inflammatory mechanisms in carcinogenesis. These molecules regulate the development of cancer when involved in the activation of the immune response, while their recurrent inflammatory reaction will envisage the development of malignancy and tumor formation. Any differences in the polymorphic sequence of cytokine genes cause heterogeneity in their expression to deregulate genetic regulation, which eventually makes a person vulnerable to multiple cancers. As stated in the t segment, considerable evidence that links cytokine gene susceptibility polymorphism with various cancers has materialized their position as important bimolecular genetic cancer susceptibility and prognosis determinants. In studying candidate genes implicated in particular pathways for specific cancer; we conclude there is a desperate need for more case-control association studies involving additional variables haplotypic gene-gene associations and clinical-pathological features to unearth factual association for cancer. Because cytokine polymorphisms play a crucial role in immunological processes in which a rich supply of inflammatory cytokines are produced by various intensities of inflammation and cancer microenvironment, research into genetic polymorphisms in cytokine genes and their reaction to chemo-radiotherapy is believed to assist patients with cancer treatment and management.

Conflict of interest

The authors declare no conflict of interest.

Author details

Arshad A. Pandith[1†], Ina Bhat[1†], Sheikh Mansoor[1†], Aabid Koul[1†], Usma Manzoor[1†], Iqra Anwar[1†], Fozia Mohammad[1†], Qurat Ul Aein[1†], Shahid M. Baba[1†] and Carmen Vladulescu[2†]

1 Advanced Center for Human Genetics, Sher-I-Kashmir Institute of Medical Sciences (SKIMS), Srinagar, J and K, India

2 Department of Biology and Environmental Engineering, University of Craiova, Craiova, Romania

*Address all correspondence to: arshaajiz@gmail.com

† These authors contributed equally.

References

[1] Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;**144**(5):646-674

[2] Crusz SM, Balkwill FR. Inflammation and cancer: Advances and new agents. Nature Reviews. Clinical Oncology. 2015;**12**:584-596

[3] Chanmee T, Ontong P, Konno K, Itano N. Tumor-associated macrophages as major players in the tumor microenvironment. Cancers. 2014;**6**: 1670-1690

[4] Wu Y, Zhou B. TNF-α/NF-κB/Snail pathway in cancer cell migration and invasion. British Journal of Cancer. 2010;**102**:639-644

[5] Allavena P, Mantovani A. Immunology in the clinic review series; focus on cancer: Tumour-associated macrophages: Undisputed stars of the inflammatory tumour microenvironment. Clinical and Experimental Immunology. 2012; **167**(2):195-205

[6] Smith KA. Toward a molecular understanding of adaptive immunity: A chronology–part II. Frontiers in Immunology. 2012;**3**:364

[7] Dantzer R, Meagher MW, Cleeland CS. Translational approaches to treatment-induced symptoms in cancer patients. Nature Reviews. Clinical Oncology. 2012;**9**(7):414

[8] Swiatek BJ. Is interleukin-10 gene polymorphism a predictive marker in HCV infection? Cytokine and Growth Factor Reviews. 2012;**23**(1–2):47-59

[9] Candido J, Hagemann T. Cancer-related inflammation. Journal of Clinical Immunology. 2013;**33**(1):79-84

[10] Eil R, Vodnala SK, Clever D, Klebanoff CA, Sukumar M, Pan JH, et al. Ionic immune suppression within the tumour microenvironment limits T cell effector function. Nature. 2016; **537**(7621):539-543

[11] Berraondo P, Sanmamed MF, Ochoa MC, Etxeberria I, Aznar MA, Pérez-Gracia JL, et al. Cytokines in clinical cancer immunotherapy. British Journal of Cancer. 2019;**120**(1):6-15

[12] Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;**140**(6):883-899

[13] Chen Z, Bouamar R, Van Schaik RH, De Fijter JW, Hartmann A, Zeier M, et al. Genetic polymorphisms in IL-2, IL-10, TGF-β1, and IL-2 RB and acute rejection in renal transplant patients. Clinical Transplantation. 2014;**28**(6): 649-655

[14] Yang Y, Luo C, Feng R, Bi S. The TNF-α, IL-1B and IL-10 polymorphisms and risk for hepatocellular carcinoma: A metaanalysis. Journal of Cancer Research and Clinical Oncology. 2011; **137**(6):947-952

[15] Stowe RP, Peek MK, Cutchin MP, Goodwin JS. Plasma cytokine levels in a population-based study: Relation to age and ethnicity. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences. 2010; **65**(4):429-433

[16] Chou IC, Lin WD, Wang CH, Tsai CH, Li TC, Tsai FJ. Interleukin (IL)-1β, IL-1 receptor antagonist, IL-6, IL-8, IL-10, and tumor necrosis factor α gene polymorphisms in patients with febrile seizures. Journal of Clinical Laboratory Analysis. 2010;**24**(3): 154-159

[17] Bandil K, Singhal P, Sharma U, Hussain S, Basu S, Parashari A, et al. Impacts of TNF-LTA SNPs/haplotypes and lifestyle factors on oral carcinoma in

an Indian population. Molecular Diagnosis & Therapy. 2016;**20**(5): 469-480

[18] Skibola CF, Bracci PM, Nieters A, Brooks-Wilson A, De Sanjosé S, Hughes AM, et al. Tumor necrosis factor (TNF) and lymphotoxin-α (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph Consortium. American Journal of Epidemiology. 2010;**171**(3):267-276

[19] Chiurillo MA. Role of gene polymorphisms in gastric cancer and its precursor lesions: Current knowledge and perspectives in Latin American countries. World Journal of Gastroenterology. 2014;**20**(16):4503

[20] Hong JB, Zuo W, Wang AJ, Lu NH. Helicobacter pylori infection synergistic with IL-1β gene polymorphisms potentially contributes to the carcinogenesis of gastric cancer. International Journal of Medical Sciences. 2016;**13**(4):298

[21] Haura EB, Turkson J, Jove R. Mechanisms of disease: Insights into the emerging role of signal transducers and activators of transcription in cancer. Nature Clinical Practice Oncology. 2005;**2**:315-324

[22] Berger, Franklin G. "The interleukin-6 gene: a susceptibility factor that may contribute to racial and ethnic disparities in breast cancer mortality." Breast cancer research and treatment. 2004;**88**(3):281-285.

[23] Ito LS, Iwata H, Hamajima N, Saito T, et al. Significant reduction in breast cancer risk for Japanese women with interleukin 1B–31 CT/TT relative to CC genotype. Japanese Journal of Clinical Oncology. 2002;**32**: 398-402

[24] DeMichele A, Martin AM, Mick R, et al. Interleukin-6-174 G to C polymorphism is associated with improved outcome in high-risk breast cancer. Cancer Research. 2003;**63**: 8051-8056

[25] Iacopetta B, Grieu F, Joseph D. The _ 174 G/C gene polymorphism in interleukin-6 is associated with an aggressive breast cancer phenotype. British Journal of Cancer. 2004;**90**: 419-422

[26] Blankenstein T. The role of tumor 53 stroma in the interaction between tumor 54 and immune system. Current Opinion in 55 Immunology. 2005;**17**: 180-186

[27] Broser EEG. 2016;**85**. Available from: http://www.ensembl.org/index.html

[28] Giordani L, Bruzzi P, Lasalandra C, Quaranta M, et al. Polymorphisms of Interleukin-10 and tumour necrosis factor-a gene promoter and breast cancer risk. Clinical Chemistry. 2003; **49**:1664-1667

[29] Langsenlehner U, Krippl P, Renner W, Yazdani-Biuki B, Eder T, Koppel H, et al. Interleukin-10 promoter polymorphism is associated with decreased breast cancer risk. Breast Cancer Research and Treatment. 2005; **90**:113-115

[30] Karakus N, Kara N, Ulusoy AN, Ozaslan C, Bek Y. Tumor necrosis factor alpha and beta and interferon gamma gene polymorphisms in Turkish breast cancer patients. DNA and Cell Biology. 2011;**30**(6):371-377

[31] He JR, Chen LJ, Su Y, Cen YL, Tang LY, Yu DD, et al. Joint effects of Epstein-Barr virus and polymorphisms in interleukin-10 and interferon-gamma on breast cancer risk. The Journal of Infectious Diseases. 2012;**205**(1):64-71

[32] Erdei E, Kang H, Meisner A, White K, et al. Polymorphisms in cytokine genes and serum cytokine

levels among New Mexican women with and without breast cancer. Cytokine. 2010;**51**(1):18-24

[33] Luo JL, Maeda S, Hsu LC, Yagita H, Karin M. Inhibition of NF-kB in cancer cells converts inflammation-induced tumor growth mediated by TNF-α to TRAIL-mediated tumor regression. Cancer Cell. 2004;**6**:297-305

[34] Liu J, Zhai X, Jin G, Hu Z, et al. Functional variants in the promoter of interleukin-1beta are associated with an increased risk of breast cancer: A case-control analysis in a Chinese population. International Journal of Cancer. 2006; **118**:2554-2558

[35] Mi YY, Yu QQ, Xu B, Zhang LF, et al. Interferon gamma +874 T/A polymorphism contributes to cancer susceptibility: A meta-analysis based on 17 case-control studies. Molecular Biology Reports. 2011;**38**(7):4461-4467

[36] Correlations of IFN-γ genetic polymorphisms with susceptibility to breast cancer: a meta-analysis. Tumour Biol. 2015;**36**(9):7307. DOI: 10.1007/s13277-015-3699-1. PMID: 26281989

[37] Mocellin S, Rossi CR, Pilati P, Nitti D. Tumor necrosis factor, cancer and anticancer therapy. Cytokine & Growth Factor Reviews. 2005;**16**: 35-53

[38] Neben K et al. Polymorphisms of the tumor necrosis factor-alpha gene promoter predict for outcome after thalidomide therapy in relapsed and refractory multiple myeloma. Blood. 2002;**100**:2263-2265

[39] Wilson AG, Symons JA, McDowell TL, McDevitt HO, et al. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America. 1997;**94**: 3195-3199

[40] Mestiri S, Bouaouina N, Ahmed SB, et al. Genetic variation in the tumor necrosis factor-alpha promoter region and in the stress protein hsp 70-2: Susceptibility and prognostic implications in breast carcinoma. Cancer. 2001;**91**(4):672-678

[41] Azmy IA, Balasubramanion SP, Wilson AG, et al. Role of tumor necrosis factor gene polymorphisms (–308 and –238) in breast cancer susceptibility and severity. Breast Cancer Research. 2004; **6**(4):395-400

[42] Smith KC, Bateman AC, Fussell HM, Howell WM. Cytokine Gene polymorphisms and breast cancer susceptibility and prognosis. European Journal of Immunogenetics. 2004;**31**(4): 167-173

[43] Scola L, Vaglica M, Crivello A, et al. Cytokine gene polymorphisms and breast cancer susceptibility. Annals. New York Academy of Sciences. 2006;**1 89**:104-109

[44] Wang J et al. Tumour necrosis factor alpha -308G/A polymorphism and risk of the four most frequent cancers: A meta-analysis. International Journal of Immunogenetics. 2011;**38**: 311-320

[45] Jeng JE, Tsai HR, Chuang LY, Tsai JF, et al. Independent and additive interactive effects among tumor necrosis factor-alpha polymorphisms, substance use habits, and chronic hepatitis B and hepatitis C virus infection on risk for hepatocellular carcinoma. Medicine (Baltimore). 2009; **88**:349-357. DOI: 10.1097/MD.0b013e31 81c10477

[46] Skerrett DL, Moore EM, Bernstein DS, Vahdat L. Cytokine genotype polymorphisms in breast carcinoma: Associations of TGF-beta 1 with relapse. Cancer Investigation. 2005;**23**:208-214

[47] Chen CC, Yang SY, Liu CJ, Lin CL, et al. Association of cytokine and DNA repair gene polymorphisms with hepatitis B-related hepatocellular carcinoma. International Journal of Epidemiology. 2005;**34**:1310-1318

[48] El-Omar EM, Carrington M, Chow WH, McColl KE, et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature. 2000;**404**:398-402

[49] Hirankarn N, Kimkong I, Kummee P, Tangkijvanich P, et al. Interleukin-1beta gene polymorphism associated with hepatocellular carcinoma in hepatitis B virus infection. World Journal of Gastroenterology. 2006;**12**:776-779

[50] Wang Y, Kato N, Hoshida Y, Yoshida H, et al. Interleukin-1beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection. Hepatology. 2003;**37**:65-71

[51] Akkiz H, Bayram S, Bekar A, Ozdil B, Sümbül AT, et al. G-308A TNF-alpha polymorphism is associated with an increased risk of hepatocellular carcinoma in the Turkish population: Case-control study. Cancer Epidemiology. 2009;**33**:261-264. DOI: 10.1016/j.canep.2009.06.001

[52] Hösel M, Quasdorff M, Wiegmann K, Webb D, et al. Not interferon, but interleukin-6 controls early gene expression in hepatitis B virus infection. Hepatology. 2009;**50**: 1773-1782. DOI: 10.1002/hep.23226

[53] Ataseven H, Bahcecioglu IH, Kuzu N, Yalniz M, et al. The levels of ghrelin, leptin, TNF-alpha, and IL-6 in liver cirrhosis and hepatocellular carcinoma due to HBV and HDV infection. Mediators of Inflammation. 2006;**2006**:78380

[54] Park BL, Lee HS, Kim YJ, Kim JY, et al. Association between interleukin 6 promoter variants and chronic hepatitis B progression. Experimental & Molecular Medicine. 2003;**35**:76-82

[55] Xue H, Lin B, An J, Zhu Y, et al. Interleukin-10–819 promoter polymorphism in association with gastric cancer risk. BMC Cancer. 2012; **12**:102

[56] Williams TM, Eisenberg L, Burlein JE, Norris CA, Pancer S, Yao D, et al. Two regions within the human IL-2 gene promoter are important for inducible IL-2 expression. The Journal of Immunology. 1988;**141**(2):662-666

[57] Wu GY, Hasenberg T, Magdeburg R, Bonninghoff R, et al. Association between EGF, TGF-β1, VEGF gene polymorphism and colorectal cancer. World Journal of Surgery. 2009;**33**(1):124-129

[58] Togawa S, Joh T, Itoh M, Katsuda N, et al. Interleukin-2 gene polymorphisms associated with increased risk of gastric atrophy from *Helicobacter pylori* infection. Helicobacter. 2005;**10**:172-178

[59] Savage SA, Abnet CC, Haque K, Mark SD, et al. Polymorphisms in interleukin -2, -6, and -10 are not associated with gastric cardia or esophageal cancer in a high-risk Chinese population. Cancer Epidemiology, Biomarkers & Prevention. 2004;**13**: 1547-1549

[60] Matsuo K, Oka M, Murase K, Soda H, Isomoto H, Takeshima F, et al. Expression of interleukin 6 and its receptor in human gastric and colorectal cancers. Journal of International Medical Research. 2003;**31**(2):69-75

[61] Putoczki TL, Thiem S, Loving A, Busuttil RA, Wilson NJ, Ziegler PK, et al. Interleukin-11 is the dominant IL-6 family cytokine during gastrointestinal tumorigenesis and can be targeted therapeutically. Cancer Cell. 2013;**24**(2): 257-271

[62] Brat DJ, Bellail AC, Van Meir EG. The role of interleukin-8 and its receptors in gliomagenesis and tumoral angiogenesis. Neuro-Oncology. 2005; **7**(2):122-133

[63] Kido S, Kitadai Y, Hattori N, Haruma K, Kido T, Ohta M, et al. Interleukin 8 and vascular endothelial growth factor—Prognostic factors in human gastric carcinomas? European Journal of Cancer. 2001;**37**(12):1482-1487

[64] Yoshino N, Ishihara S, Rumi MAK, Ortega-Cava CF, Yuki T, Kazumori H, et al. Interleukin-8 regulates expression of Reg protein in Helicobacter pylori-infected gastric mucosa. American Journal of Gastroenterology. 2005; **100**(10):2157-2166

[65] Hull J, Rowlands K, Lockhart E, Sharland M, Moore C, Hanchard N, et al. Haplotype mapping of the bronchiolitis susceptibility locus near IL8. Human Genetics. 2004;**114**:272-279

[66] Lu Y, Wang ZD, Shen J, Xu YC. Meta-analysis on the relationship between IL8-251 gene polymorphism and gastric cancer. Zhonghua Yu Fang Yi Xue Za Zhi. 2007;**41**(suppl):39-42

[67] Kang JM, Kim N, Lee DH, Park JH, Lee MK, Kim JS, et al. The effects of genetic polymorphisms of IL-6, IL-8, and IL-10 on *Helicobacter pylori*-induced gastroduodenal diseases in Korea. Journal of Clinical Gastroenterology. 2009;**43**:420-428

[68] Zhang B, Rong G, Wei H, Zhang M, Bi J, Ma L, et al. The prevalence of Th17 cells in patients with gastric cancer. Biochemical and Biophysical Research Communications. 2008;**374**:533-537

[69] Wu GY, Lu Q, Hasenberg T, Niedergethmann M, et al. Association between EGF, TGF-β1, TNF-α gene polymorphisms and cancer of the pancreatic head. Anticancer Research. 2010;**30**(12):5257-5261

[70] Noach LA, Bosma NB, Jansen J, Hoek FJ, Van Deventer SJH, Tytgat GNJ. Mucosal tumor necrosis factor-or, interleukin-1/3, and interleukin-8 production in patients with *Helicobacter pylori* infection. Scandinavian Journal of Gastroenterology. 1994;**29**(5): 425-429

[71] Beales ILP, Calam J. Interleukin 1β and tumour necrosis factor α inhibit acid secretion in cultured rabbit parietal cells by multiple pathways. Gut. 1998; **42**(2):227-234

[72] Forones NM, Mandowsky SV, Lourenço LG. Serum levels of interleukin-2 and tumor necrosis factor-alpha correlate to tumor progression in gastric cancer. Hepato-Gastroenterology. 2001;**48**(40):1199-1201

[73] Beales IL, Calam J. Pathogenic mechanisms in *Helicobacter pylori* infection. Hospital Medicine (London, England: 1998). 1998;**59**(3):186-190

[74] El-Omar EM, Rabkin CS, Gammon MD, Vaughan TL, Risch HA, Schoenberg JB, et al. Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. Gastroenterology. 2003;**124**(5):1193-1201

[75] Rad R, Dossumbekova A, Neu B, Lang R, Bauer S, Saur D, et al. Cytokine gene polymorphisms influence mucosal cytokine expression, gastric inflammation, and host specific colonisation during *Helicobacter pylori* infection. Gut. 2004;**53**(8):1082-1089

[76] Gorouhi F, Islami F, Bahrami H, Kamangar F. Tumour-necrosis factor-A polymorphisms and gastric cancer risk: A meta-analysis. British Journal of Cancer. 2008;**98**(8):1443-1451

[77] Xu H, Ding Q, Jiang HW. Genetic polymorphism of interleukin-1A

(IL-1A), IL-1B, and IL-1 receptor antagonist (IL-1RN) and prostate cancer risk. Asian Pacific Journal of Cancer Prevention. 2014;**15**(20):8741-8747

[78] Mandić S, Sudarević B, Marczi S, et al. Interleukin-6 polymorphism and prostate cancer risk in population of Eastern Croatia. Collegium Antropologicum. 2013;**37**:907-911

[79] Pierce BL, Biggs ML, DeCambre M, et al. C-reactive protein, interleukin-6, and prostate cancer risk in men aged 65 years and older. Cancer Causes & Control. 2009;**20**:1193-1203

[80] Mandal S, Abebe F, Chaudhary J. 174G/C polymorphism in the interleukin-6 promoter is differently associated with prostate cancer incidence depending on race. Genetics and Molecular Research. 2014;**13**: 139-151

[81] Yang M, Li C, Li M. Association of interleukin-6 (-174 G/C) polymorphism with the prostate cancer risk: A meta-analysis. Biomedical Reports;**2**:637-643. DOI: 10.3892/Br.2014.300

[82] Grivennikov S, Karin M. Autocrine IL-6 signaling: A key event in tumorigenesis? Cancer Cell. 2008;**13**:7-9

[83] Nguyen TN, Baaklini S, Koukouikila-Koussounda F, et al. Association of a functional TNF variant with *Plasmodium falciparum* parasitaemia in a congolese population. Genes and Immunity. 2017;**18**:152-157

[84] Satoru M, Toshinobu N, Masatomo M, Yudai H, et al. Interleukin-6-634G >C genetic polymorphism is associated with prognosis following surgery for advanced thoracic esophageal squamous cell carcinoma. Digestive Surgery. 2012, 2012;**29**:194-201

[85] Oka M, Yamamoto K, Takahashi M, Hakozaki M, et al. Relationship between serum levels of interleukin 6, various disease parameters and malnutrition in patients with esophageal squamous cell carcinoma. Cancer Research. 1996;**56**: 2776-2780

[86] Buraczynska M, Jozwiak L, Ksiazek P, Borowicz E, Mierzicki P. Interleukin-6 gene polymorphism and faster progression to endstage renal failure in chronic glomerulonephritis. Translational Research. 2007;**150**:101-105

[87] Kitamura A, Hasegawa G, Obayashi H, Kamiuchi K, et al. Interleukin-6 polymorphism (-634C/G) in the promotor region and the progression of diabetic nephropathy in type 2 diabetes. Diabetic Medicine. 2002;**19**:1000-1005

[88] Sansone P, Storci G, Tavolari S, Guarnieri T, et al. IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. The Journal of Clinical Investigation. 2007; **117**:3988-4002

[89] Trinchieri G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. National Review. 2003;**3**:133-146

[90] Yi PT, Wan LW, Song YL, Jian Z, et al. Associations between polymorphisms in IL-12A, IL-12B, IL-12Rb1, IL-27 gene and serum levels of IL-12p40, IL-27p28 with esophageal cancer. Journal of Cancer Research and Clinical Oncology. 2012;**138**: 1891-1900

[91] Hsieh YY, Chang CC, Tsai CH, Lin CC, Tsai FJ. Interleukin (IL)-12 receptor beta 1 codon 378 G homozygote and allele, but not IL-1 (beta-511 promoter, 3953 exon 5, receptor antagonist), IL-2 114, IL-4-590 intron 3, IL-8 30-UTR 2767, and IL-18 105, are associated with higher susceptibility to leiomyoma. Fertility and Sterility. 2007; **87**(4):886-895

[92] Imboden M, Nicod L, Nieters A, Glaus E, et al. The common G-allele of interleukin-18 single-nucleotide polymorphism is a genetic risk factor for atopic asthma. The SAPALDIA Cohort Study. Clinical and Experimental Allergy. 2006;**36**:211-218

[93] Takagawa T, Tamura K, Takeda N, Tomita T, et al. Association between IL-18 gene promoter polymorphisms and inflammatory bowel disease in a Japanese population. Inflammatory Bowel Diseases. 2005;**11**:1038-1043

[94] Ye SW, Yan L, Yun GL, Hui T, et al. Interleukin-18 gene promoter polymorphisms and the risk of esophageal squamous cell carcinoma. Acta Oncologica. 2007;**46**: 1090-1096

[95] Watanabe M, Tanaka K, Takizawa T, et al. Characterization of a canine tetranucleotide microsatellite marker located in the first intron of the tumor necrosis factor alpha gene. The Journal of Veterinary Medical Science. 2014;**76**:119-122

[96] Zhu G, Du Q, Wang X, Tang N, She F, Chen Y. TNF-alpha promotes gallbladder cancer cell growth and invasion through autocrine mechanisms. International Journal of Molecular Medicine. 2014;**33**:1431-1440

[97] Piotrowski P, Wudarski M, Sowinska A, Olesinska M, Jagodzinski PP. TNF-308 G/A polymorphism and risk of systemic lupus erythematosus in the Polish population. Modern Rheumatology. 2015;**25**:719-723

[98] Hai-feng W, Hong-bo L, Hare A, Apizi PL, et al. p 21 and TNF-α gene polymorphisms and susceptibility of Xinjiang Kazakh esophageal family. Chinese Journal of Cancer Prevention and Treatment. 2014;**21**: 329-332

[99] Deans C, Rose-Zerilli M, Wigmore S, Ross J, et al. Host cytokine genotype is related to adverse prognosis and systemic inflammation in gastro-oesophageal cancer. Annals of Surgical Oncology. 2007;**14**:329-339

[100] Cui X-B, Wang D-D, Zhang H-Y, Li T-T, et al. Tumor necrosis factor-α gene 308G/A polymorphism is not associated with esophageal squamous cell carcinoma risk in Kazakh patients. International Journal of Clinical and Experimental Pathology. 2015;**8**:9293

[101] Guo W, Wang N, Li Y, Zhang JH. Polymorphisms in tumor necrosis factor genes and susceptibility to esophageal squamous cell carcinoma and gastric cardiac adenocarcinoma in a population of high incidence region of North China. Chinese Medical Journal. 2005;**118**: 1870-1878

[102] Luo M, Yang Y, Luo D, Liu L, Zhang Y, Xiao F, ... & Luo Z. (2016). Tumor necrosis factor-alpha promoter polymorphism 308 G/A is not significantly associated with esophageal cancer risk: a meta-analysis. Oncotarget. 2016;**7**(48):79901.

[103] Filipe SO, Bustamante J, Chapgier A, Vogt G, de BL, Feinberg J, et al. Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: Molecular, cellular, and clinical features. Seminars in Immunology. 2006;**18**:347-361

[104] Du W, Ye W, Chen M, Li D, et al. Association research between polymorphism of IFN-gamma and IL-10, environmental risk factors, and susceptibility to esophageal cancer. Journal of Hygiene Research. 2013; **42**(5):770-776

[105] Motoyama S, Miura M, Hinai Y, Maruyama K, Usami S, Nakatsu T, ... & Ogawa J. I. (2009). Interferon-gamma 874A>T genetic polymorphism is associated with infectious complications

following surgery in patients with thoracic esophageal cancer. Surgery. 2009;**146**(5):931-938

[106] Chen G, Hohmeier HE, Newgard CB. Expression of the transcription factor STAT-1a in insulinoma cells protects against cytotoxic effects of multiple cytokines. The Journal of Biological Chemistry. 2001;**276**:766-772

[107] Berkovic MC, Ivkovic TC, Marout J, Zjacić-Rotkvic V, et al. Interleukin 1β gene single-nucleotide polymorphisms and susceptibility to pancreatic neuroendocrine tumors. DNA and cell biology. 2012;**31**(4):531-536

[108] Hoffman SC, Stanley EM, Darrin Cox E, Craighead N, et al. Association of cytokine polymorphic inheritance and in vitro cytokine production in anti-CD3/CD28-stimulated peripheral blood lymphocytes. Transplantation. 2001;**72**: 1444-1450

[109] Reichert TE, Watkins S, Stanson J, Johnson JT, et al. Endogenous IL-2 in cancer cells: A marker of cellular proliferation. The Journal of Histochemistry and Cytochemistry. 1998;**46**:603-612

[110] Hofsli E, Thommesen L, Yadetie F, Langaas M, et al. Identification of novel growth-factor responsive genes in neuroendocrine gastrointestinal tumor cells. British Journal of Cancer. 2005;**92**: 1506-1516

[111] Lu PH, Tang Y, Li C, Shen W, et al. Metaanalysis of association of tumor necrosis factor alpha-308 gene promoter polymorphism with gastric cancer. Chinese Journal of Preventive Medicine. 2010;**44**(3):209-214

[112] Zhou P, Li JP, Zhang C. Polymorphisms of tumor necrosis factor-alpha and breast cancer risk: Appraisal of a recent metaanalysis. Breast Cancer Research and Treatment. 2011;**126**(1):253-254; author reply 255-256

[113] Kishimoto T. Interleukin-6: From basic science to medicine—40 years in immunology. Annual Review of Immunology. 2005;**23**:1-21

[114] Tahara E. Growth factors and oncogenes in human gastrointestinal carcinomas. Journal of Cancer Research and Clinical Oncology. 1990;**116**(2): 121-131

[115] Morgan MD, Harper L, Williams J, Savage C. Anti-neutrophil cytoplasm-associated glomerulonephritis. Journal of the American Society of Nephrology. 2006;**17**(5):1224-1234

[116] Paule B, Terry S, Kheuang L, Soyeux P, Vacherot F, de la Taille A. The NF-κB/IL-6 pathway in metastatic androgen-independent prostate cancer: New therapeutic approaches? World Journal of Urology. 2007;**25**(5):477-489

[117] Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, et al. Association of common polymorphisms in inflammatory genes interleukin (IL) 6, IL8, tumor necrosis factor α, NFKB1, and peroxisome proliferator-activated receptor γ with colorectal cancer. Cancer Research. 2003;**63**(13): 3560-3566

[118] Gautam KA, Tripathi M, Sankhwar SN, et al. Functional polymorphisms in the IL6 gene promoter and the risk of urinary bladder cancer in India. Cytokine. 2016;**77**: 152-156

[119] Fishman D, Faulds G, Jeffery R, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. The Journal of Clinical Investigation. 1998;**102**: 1369-1376

[120] Leibovici D, Grossman HB, Dinney CP, Millikan RE, Lerner S, Wang Y, et al. Polymorphisms in inflammation genes and bladder cancer: From initiation to recurrence, progression, and survival. Journal of Clinical Oncology. 2005;**23**(24): 5746-5756

[121] Aben KK, Witjes JA, Schoenberg MP, Hulsbergen-van de Kaa C, Verbeek AL, Kiemeney LA. Familial aggregation of urothelial cell carcinoma. International Journal of Cancer. 2002;**98**:274-278. DOI: 10.1002/ijc.10191

[122] Ebadi N, Jahed M, Mivehchi M, et al. Interleukin-12 and interleukin-6 gene polymorphisms and risk of bladder cancer in the Iranian population. Asian Pacific Journal of Cancer Prevention. 2014;**1518**:7869-7873

[123] Ahirwar D, Kesarwani P, Manchanda PK, et al. Anti- and proinflammatory cytokine gene polymorphism and genetic predisposition: Association with smoking, tumor stage and grade, and bacillus Calmette-Guérin immunotherapy in bladder cancer. Cancer Genetics and Cytogenetics. 2008;**184**:1-8

[124] Mcdermott MF. TNF and TNFR biology in health and disease. Cellular AND Molecular Biology-Paris-Wegmann. 2001;**47**(4):619-636

[125] Kohaar I, Tiwari P, Kumar R, Nasare V, Thakur N, Das BC, et al. Association of single nucleotide polymorphisms (SNPs) in TNF-LTA locus with breast cancer risk in Indian population. Breast Cancer Research and Treatment. 2009;**114**(2):347-355

[126] Sáenz-López P, Carretero R, Cózar JM, Romero JM, Canton J, Vilchez JR, et al. Genetic polymorphisms of RANTES, IL1-A, MCP-1 and TNF-A genes in patients with prostate cancer. BMC Cancer. 2008;**8**(1):1-8

[127] Marsh H, Haldar N, Bunce M, et al. Polymorphisms in tumour necrosis factor (TNF) are associated with risk of bladder cancer and grade of tumour at presentation. British Journal of Cancer. 2003;**89**:1096-1101. DOI: 10.1038/sj.bjc.6601165

[128] Ahirwar DK, Mandhani A, Dharaskar A, Kesarwani P, & Mittal RD. Association of tumour necrosis factor-α gene (T-1031C, C-863A, and C-857T) polymorphisms with bladder cancer susceptibility and outcome after bacille Calmette-Guérin immunotherapy. BJU international. 2009;**104**(6):867-873.

[129] Kim EJ, Jeong P, Quan C, Kim J, Bae SC, Yoon SJ, et al. Genotypes of TNF-α, VEGF, hOGG1, GSTM1, and GSTT1: Useful determinants for clinical outcome of bladder cancer. Urology. 2005;**65**(1):70-75

[130] Nonomura N, Tokizane T, Nakayama M, Inoue H, Nishimura K, Muramatsu M, et al. Possible correlation between polymorphism in the tumor necrosis factor-beta gene and the clinicopathological features of bladder cancer in Japanese patients. International Journal of Urology. 2006; **13**(7):971-976

[131] Lima L, Silva J, Amaro T, et al. IL-4 and TNF-α polymorphisms are associated with risk of multiple superficial tumors or carcinoma in situ development. Urologia Internationalis. 2011;**87**:457-463

[132] Brenner AV, Butler MA, Wang SS, Ruder AM, et al. Single-nucleotide polymorphisms in selected cytokine genes and risk of adult glioma. Carcinogenesis. 2007;**28**(12):2543-2547

[133] Rahaman SO, Vogelbaum MA, Haque SJ. Aberrant Stat 3 signaling by

interleukin-4 in malignant glioma cells: Involvement of IL-13Rα2. Cancer Research. 2005;**65**:2956

[134] Choi WA, Kang MJ, Kim YJ, Seo JH, et al. Gene-gene interactions between candidate gene polymorphisms are associated with total IgE levels in Korean children with asthma. The Journal of Asthma. 2012;**49**(3): 243-252

[135] Jin T, Li X, Zhang J, Wang H, et al. Genetic association between selected cytokine genes and glioblastoma in the Han Chinese population. BMC Cancer. 2013;**13**:236

[136] Schwartzbaum J, Ding B, Johannesen TB, Osnes LT, et al. Association between prediagnostic IgE levels and risk of glioma. Journal of the National Cancer Institute. 2012;**104**: 1251-1259

[137] Ruan Z, Zhao Y, Yan L, Chen H, et al. Single nucleotide polymorphisms in IL-4Ra, IL-13 and STAT6 genes occurs in brain glioma. Frontiers in Bioscience (Elite Edition). 2011;**3**:33

[138] Ben CA, Busson M, Douik H, Boukouaci W, et al. Association of IL-12p40 +1188 A/C polymorphism with nasopharyngeal cancer risk and tumor extension. Tissue Antigens. 2011; **78**(2):148-151

[139] Kaarvatn MH, Vrbanec J, Kulic A, Knezevic J, et al. Single nucleotide polymorphism in the interleukin 12B gene is associated with risk for breast cancer development. Scandinavian Journal of Immunology. 2012;**76**(3): 329-335

[140] Shamran HA, Ghazi HF, AL-Salman A, et al. Singh single nucleotide polymorphisms in IL-10, IL-12p40, and IL-13 genes and susceptibility to glioma. International Journal of Medical Sciences. 2015;**12**(10):790-796

[141] Chen H, Cheng S, Wang J, Cao C, et al. Interleukin-12 rs 3212227 polymorphism and cancer risk: A meta-analysis. Molecular Biology Reports. 2012;**39**(12):10235-10242

[142] Terabe M, Park JM, Berzofsky JA. Role of IL-13 in regulation of anti-tumor immunity and tumor growth. Cancer Immunology, Immunotherapy. 2004;**53**: 79-85

[143] Husain SR, Puri RK. Interleukin-13 receptor-directed cytotoxin for malignant glioma therapy: From bench to bedside. Journal of Neuro-Oncology. 2003;**65**:37-48

[144] Graves PE, Kabesch M, Halonen M, Holberg CJ, et al. A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children. The Journal of Allergy and Clinical Immunology. 2000;**105**: 506-513

[145] Su Y, Tang LY, Chen LJ, He JR, et al. Joint effects of febrile acute infection and an interferon-gamma polymorphism on breast cancer risk. PLoS ONE. 2012;**7**(5):e37275

[146] Wei Y, Liu F, Li B, Chen X, et al. Polymorphisms of tumor necrosis factor-alpha and hepatocellular carcinoma risk: A HuGE systematic review and meta-analysis. Digestive Diseases and Sciences. 2011;**56**: 2227-2236. DOI: 10.1007/s10620-011-1617-y

[147] Nieters A, Yuan JM, Sun CL, Zhang ZQ, et al. Effect of cytokine genotypes on the hepatitis B virus-hepatocellular carcinoma association. Cancer. 2005;**103**:740-748

[148] Karra VK, Gumma PK, Chowdhury SJ, Ruttala R, et al. IL-18 polymorphisms in hepatitis B virus related liver disease. Cytokine. 2015;**73**:

277-282. DOI: 10.1016/j.cyto.2015.02.015

[149] Zhang PA, Wu JM, Li Y, Yang XS. Association of polymorphisms of interleukin-18 gene promoter region with chronic hepatitis B in Chinese Han population. World Journal of Gastroenterology. 2005;**11**: 1594-1598

[150] Kim YS, Cheong JY, Cho SW, Lee KM, et al. A functional SNP of the Interleukin-18 gene is associated with the presence of hepatocellular carcinoma in hepatitis B virus-infected patients. Digestive Diseases and Sciences. 2009;**54**:2722-2728. DOI: 10.1007/s10620-009-0970-6

[151] Li S, Deng Y, Chen ZP, Huang S, et al. Genetic polymorphism of interleukin-16 influences susceptibility to HBV-related hepatocellular carcinoma in a Chinese population. Infection, Genetics and Evolution. 2011; **11**:2083-2088. DOI: 10.1016/j.meegid.2011.09.025

[152] Romani S, Hosseini SM, Mohebbi SR, Kazemian S, Derakhshani S, Khanyaghma M, ... & Zali MR. Interleukin-16 gene polymorphisms are considerable host genetic factors for patients' susceptibility to chronic hepatitis B infection. Hepatitis research and treatment; 2014.

[153] Saxena R, Chawla YK, Verma I, Kaur J. IL-6(-572/-597) polymorphism and expression in HBV disease chronicity in an Indian population. American Journal of Human Biology. 2014;**26**:549-555. DOI: 10.1002/ajhb.22562

[154] Lee SG, Kim B, Yook JH, Oh ST, Lee I, Song K. TNF/LTA polymorphisms and risk for gastric cancer/duodenal ulcer in the Korean population. Cytokine. 2004;**28**(2): 75-82

[155] Zabaleta J, Camargo MC, Piazuelo MB, Fontham E, Schneider BG, Sicinschi LA, et al. Association of interleukin-1β gene polymorphisms with precancerous gastric lesions in African Americans and Caucasians. American Journal of Gastroenterology. 2006;**101**(1):163-171

[156] Wang Y, Lerner S, Leibovici D, et al. Polymorphisms in the inflammatory genes IL-6, IL-8, TNF-α, NFKB1, and PPARG and bladder cancer risk. Proceedings of the American Association for Cancer Research Abstract. 2004;**3979**

[157] Purdue MP, Lan Q, Kricker A, Grulich AE, Vajdic CM, Turner J, et al. Polymorphisms in immune function genes and risk of non-Hodgkin lymphoma: Findings from the New South Wales non-Hodgkin Lymphoma Study. Carcinogenesis. 2007;**28**(3): 704-712

[158] Sugimoto M, Yamaoka Y, Furuta T. Influence of interleukin polymorphisms on development of gastric cancer and peptic ulcer. World Journal of Gastroenterology: WJG. 2010;**16**(10): 1188

[159] Shin WG, Jang JS, Kim HS, Kim SJ, Kim KH, Jang MK, et al. Polymorphisms of interleukin-1 and interleukin-2 genes in patients with gastric cancer in Korea. Journal of Gastroenterology and Hepatology. 2008;**23**(10):1567-1573

[160] Tao YP, Wang WL, Li SY, Zhang J, Shi QZ, Zhao F, et al. Associations between polymorphisms in IL-12A, IL-12B, IL-12Rβ1, IL-27 gene and serum levels of IL-12p40, IL-27p28 with esophageal cancer. Journal of Cancer Research and Clinical Oncology. 2012; **138**(11):1891-1900

[161] Suganuma M, Okabe S, Marino MW, Sakai A, Sueoka E, Fujiki H. Essential role of tumor necrosis factor α (TNF-α) in tumor promotion as

revealed by TNF-α-deficient mice. Cancer Research. 1999;**59**(18): 4516-4518

[162] Sethi G, Sung B, Aggarwal BB. TNF: A master switch for inflammation to cancer. Frontiers in Bioscience. 2008; **13**(2):5094-5107

[163] Konwar R, Gara R, Singh M, Singh V, Chattopadhyay N, Bid HK. Association of interleukin-4 and interleukin-1 receptor antagonist gene polymorphisms and risk of benign prostatic hyperplasia. Urology. 2008; **71**(5):868-872

[164] Jin T, Li X, Zhang J, Wang H, Geng T, Li G, et al. Genetic association between selected cytokine genes and glioblastoma in the Han Chinese population. BMC Cancer. 2013;**13**(1):1-6

[165] Hackstein H, Jakobs C, Peric A, Michel G, Baal N, Bein G, Bruene B, Hornung V, Yakut E. "Extracorporeal photopheresis promotes IL-1β production (VAC3P. 1054)." 2015:71-1.

[166] Gu X, Shen Y, Fu L, Zuo HY, Yasen H, He P, et al. Polymorphic variation of inflammation-related genes and risk of non-Hodgkin lymphoma for Uygur and Han Chinese in Xinjiang. Asian Pacific Journal of Cancer Prevention. 2014;**15**(21):9177-9183

[167] Rothman N, Skibola CF, Wang SS, Morgan G, Lan Q, Smith MT, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: A report from the InterLymph Consortium. The Lancet Oncology. 2006;**7**(1):27-38

[168] Hiroki CH, Amarante MK, Petenuci DL, Sakaguchi AY, Trigo FC, Watanabe MAE, et al. IL-10 gene polymorphism and influence of chemotherapy on cytokine plasma levels in childhood acute lymphoblastic leukemia patients: IL-10 polymorphism and plasma levels in leukemia patients. Blood Cells, Molecules, and Diseases. 2015;**55**(2):168-172

[169] Pehlivan M, Sahin HH, Pehlivan S, Ozdilli K, Kaynar L, Oguz FS, et al. Prognostic importance of single-nucleotide polymorphisms in IL-6, IL-10, TGF-β1, IFN-γ, and TNF-α genes in chronic phase chronic myeloid leukemia. Genetic Testing and Molecular Biomarkers. 2014;**18**(6):403-409

[170] Nursal AF, Pehlivan M, Sahin HH, Pehlivan S. The associations of IL-6, IFN-γ, TNF-α, IL-10, and TGF-β1 functional variants with acute myeloid leukemia in Turkish patients. Genetic Testing and Molecular Biomarkers. 2016;**20**(9):544-551. DOI: 10.1089/gtmb.2016.0036

[171] Akdis M, Burgler S, Crameri R, Eiwegger T, Fujita H, Gomez E, et al. Interleukins, from 1 to 37, and interferon-γ: Receptors, functions, and roles in diseases. Journal of Allergy and Clinical Immunology. 2011;**127**(3):701-721

[172] Davies FE, Rollinson SJ, Rawstron AC, Roman E, Richards S, Drayson M, et al. High-producer haplotypes of tumor necrosis factor alpha and lymphotoxin alpha are associated with an increased risk of myeloma and have an improved progression-free survival after treatment. Journal of Clinical Oncology. 2000;**18**(15):2843-2851

[173] Kadar K, Kovacs M, Karadi I, Melegh B, Pocsai Z, Mikala G, et al. Polymorphisms of TNF-alpha and LT-alpha genes in multiple myeloma. Leukemia Research. 2008;**32**:1499-1504

[174] Juan D, Yuan Z, Zhang C, Hua WF, Baoan J, Hou CJ. Role of the TNF-α promoter polymorphisms for development of multiple myeloma and clinical outcome in thalidomide plus dexamethasone. Leukemia Research. 2010;**34**:1453-1458

[175] Vignali DAA, Kuchroo VK. IL-12 family cytokines: Immunological playmakers. Nature Immunology. 2012; **13**:722-728

[176] Martino A, Buda G, Maggini V, Lapi F, Lupia A, Di Bello D, et al. Could age modify the effect of genetic variants in IL6 and TNF α genes in multiple myeloma? Leukemia Research. 2012;**36**: 594-597

[177] Tan, D.E., Foo, J.N., Bei, J.X., Chang, J., Peng, R., Zheng, X., Wei, L., Huang, Y., Lim, W.Y., Li, J. and Cui, Q., 2013. Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. Nature genetics. **45**(7):804-807.

[178] Zhu B, Xiao C, Zhu B, Zheng Z, Liang J. Little association between the interleukin 10-3575T/A polymorphism and cancer risk: Pooled analysis of 15608 cancer cases and 17539 controls. International Journal of Clinical and Experimental Medicine. 2015;**8**(8):14335

[179] Zhang C, Han F, Yu J, Hu X, Hua M, Zhong C, et al. Investigation of NF-κB-94ins/del ATTG and CARD8 (rs 2043211) Gene Polymorphism in Acute Lymphoblastic Leukemia. Frontiers in Endocrinology. 2019;**10**:501

[180] Wang LJ, Chou P, Gonzalez-Ryan L, Huang W, Haut PR, Kletzel M. Evaluation of mixed hematopoietic chimerism in pediatric patients with leukemia after allogeneic stem cell transplantation by quantitative PCR analysis of variable number of tandem repeat and testis determination gene. Bone Marrow Transplantation. 2002; **29**(1):51-56

[181] Zhang XJ, Yan KL, Wang ZM, Yang S, Zhang GL, Fan X, et al. Polymorphisms in interleukin-15 gene on chromosome 4q31. 2 are associated with psoriasis vulgaris in Chinese population. Journal of Investigative Dermatology. 2007;**127**(11):2544-2551

[182] Gonzalez B, Flores-A H, Sanchez D, Alaez C, Gorodezky C. IL6 (− 174G/C) SNP but not IL4 (− 590C/T) polymorphisms contributes to the development of acute lymphoblastic leukemia, and is involved in gender specific risk in Mexican children. Human Immunology. 2015;**76**:94

[183] Mondet, J., Hussein, K., & Mossuz, P. (2015). Circulating cytokine levels as markers of inflammation in Philadelphia negative myeloproliferative neoplasms: diagnostic and prognostic interest. Mediators of inflammation, 2015

[184] Jevtovic-Stoimenov T, Cvetkovic T, Despotovic M, Basic J, Cvetkovic J, Marjanovic G, et al. The influence of TNF alpha-308 G/A polymorphism on oxidative stress in patients with chronic lymphocytic leukemia. Leukemia Research. 2017;**54**:66-72

[185] Martino A, Campa D, Jurczyszyn A, Martínez-López J, Moreno MJ, Varkonyi J, et al. Genetic variants and multiple myeloma risk: IMMEnSE validation of the best reported associations—An extensive replication of the associations from the candidate gene era. Cancer Epidemiology and Prevention Biomarkers. 2014;**23**(4):670-674

[186] Wu Z, Sun Y, Zhu S, Tang S, Liu C, & Qin W. Association of Interferon gamma+ 874T/A polymorphism and leukemia risk: a meta-analysis. Medicine. 2016;**95**(12)

[187] Nass J, Efferth T. Drug targets and resistance mechanisms in multiple myeloma. Cancer Drug Resistance. 2018;**1**(2):87-117

[188] Peng X, Shi J, Sun W, Ruan X, Guo Y, Zhao L, et al. Genetic polymorphisms of IL-6 promoter in cancer susceptibility and prognosis: A meta-analysis. Oncotarget. 2018;**9**(15):12351

[189] Basmaci C, Pehlivan M, Tomatir AG, Sever T, Okan V,

Yilmaz M, et al. Effects of TNF, NOS3, MDR1 gene polymorphisms on clinical parameters, prognosis and survival of multiple myeloma cases. Asian Pacific Journal of Cancer Prevention. 2016; **17**(3):1009-1014

[190] Dinarello CA. Biologic basis for interleukin-1 in disease. Blood. 1996;**87**: 2095-2147

[191] Ito LS, Iwata H, Hamajima N, et al. K Significant reduction in breast cancer risk for Japanese women with interleukin 1B-31CT/TTrelative to CC genotype. Japanese Journal of Clinical Oncology. 2002;**32**:398-402

[192] Alliston T, Choy L, Ducy P, Karsenty G, et al. TGFbeta- induced repression of CBFA1 by Smad 3 decreases cbfa 1 and osteocalcin expression and inhibits osteoblast differentiation. The EMBO Journal. 2001;**20**(9):2254-2272

[193] Adorno M, Cordenonsi M, Montagner M, Dupont S, et al. A Mutant-p 53/Smad complex opposes p 63 to empower TGFbeta-induced metastasis. Cell. 2009;**137**(1):87-98

[194] Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, et al. A transforming growth factorbeta 1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. Cancer Research. 2003;**63**(10):2610-2615

[195] Hishida A, Iwata H, Hamajima N, Matsuo K, et al. Transforming growth factor B1 T29C polymorphism and breast cancer risk in Japanese women. Breast Cancer. 2003;**10**(1): 63-69

[196] Yongsheng H, Binghui L, Ji Q, Jingtian X, et al. TGF-b1 29T/C polymorphism and breast cancer risk: A meta-analysis involving 25, 996 subjects. Breast Cancer Research and Treatment. 2010;**123**:863-868

[197] Turner DM, Williams DM, Sankaran D, Lazarus M, et al. An investigation of polymorphism in the interleukin-10 gene promoter. European Journal of Immunogenetics. 1997;**24**:1-8

[198] Keijsers V, Verweij CL, Westendorp RGJ, Breedveld FC, et al. IL-10 polymorphisms in relation to production and rheumatoid arthritis. Arthritis and Rheumatism. 1997;**40** (Suppl 9):S179

[199] Watanabe Y, Kinoshita A, Yamada T, Ohta T, et al. A catalog of 106 single-nucleotide polymorphisms (SNPs) and 11 other types of variations in genes for transforming growth factor-beta 1 (TGF- beta 1) and its signaling pathway. Journal of Human Genetics. 2002;**47**:478-483

[200] Talaat RM, Dondeti MF, El-Shenawy SZ, Khamiss OA. Transforming growth factor-β 1 gene polymorphism (T29C) in Egyptian patients with hepatitis B virus infection: A preliminary study. Hepatitis Research and Treatment. 2013;**2013**:293274. DOI: 10.1155/2013/293274

[201] Perrey C, Turner SJ, Pravica V, Howell WM, et al. ARMS-PCR methodologies to determine IL-10, TNF-alpha, TNF- beta and TGF-beta 1 gene polymorphisms. Transplant Immunology. 1999;**7**:127-128

[202] Yamada Y, Miyauchi A, Goto J, Takagi Y, Okuizumi H, et al. Association of a polymorphism of the transforming growth factor-beta 1 gene with genetic susceptibility to osteoporosis in postmenopausal Japanese women. Journal of Bone and Mineral Research. 1998;**13**:1569-1576

[203] Opal SM, DePalo VA. Anti-inflammatory cytokines. Chest. 2000; **117**:1162-1172

[204] Rosenwasser LJ, Klemm DJ, Dresback JK, Inamura H, Mascali JJ,

Klinnert M, et al. Promoter polymorphisms in the chromosome 5 gene cluster in asthma and atopy. Clinical and Experimental Allergy. 1995; **25**(suppl 2):74-78; discussion: 95-6

[205] Wu J, Lu Y, Ding YB, Ke Q, Hu ZB, Yan ZG, et al. Promoter polymorphisms of IL2, IL4, and risk of gastric cancer in a high-risk Chinese population. Molecular Carcinogenesis. 2009; **48**(626):32

[206] Zhou Y, Li N, Zhuang W, Liu GJ, Wu TX, Yao X, et al. Interleukin-10-1082 promoter polymorphism associated with gastric cancer among Asians. European Journal of Cancer. 2008;**44**(17):2648-2654

[207] Zhuang W, Wu XT, Zhou Y, Liu L, Liu GJ, Wu TX, et al. Interleukin 10-592 promoter polymorphism associated with gastric cancer among Asians: A meta-analysis of epidemiologic studies. Digestive Diseases and Sciences. 2010; **55**(6):1525-1532

[208] Chen KF, Li B, Wei YG, Peng CJ. Interleukin-10-819 promoter polymorphism associated with gastric cancer among Asians. The Journal of International Medical Research. 2010; **38**:1-8

[209] Yu Z, Liu Q, Huang C, et al. The interleukin 10-819C/T polymorphism and cancer risk: A HuGE review and meta-analysis of 73 studies including 15, 942 cases and 22, 336 controls. OMICS. 2013;**17**:200-214

[210] Shao N, Xu B, Mi YY, Hua LX. IL-10 polymorphisms and prostate cancer risk: A meta-analysis. Prostate Cancer and Prostatic Diseases. 2011;**14**(2): 129-135

[211] He B, Zhang Y, Pan Y, Xu Y, Gu L, Chen L, et al. Interleukin 1 beta (IL1B) promoter polymorphism and cancer risk: Evidence from 47 published studies. Mutagenesis. 2011;**26**:637-642

[212] Pociot F, Molvig J, Wogensen L, Worsaae H, et al. Taq I polymorphism in the human interleukin-1 beta (IL-1 beta) gene correlates with IL-1 beta secretion in vitro. European Journal of Clinical Investigation. 1992;**22**:396-402

[213] Liang Z, Jun Y, Liming W, Xu W, et al. Interleukin 1B rs 16944 G>A polymorphism was associated with a decreased risk of esophageal cancer in a Chinese population. Clinical Biochemistry. 2013;**46**:1469-1473

[214] Yamada S, Matsuhisa T, Makonkawkeyoon L, Chaidatch S, Kato S, Matsukura N. Helicobacter pylori infection in combination with the serum pepsinogen I/II ratio and interleukin-1beta-511 polymorphisms are independent risk factors for gastric cancer in Thais. Journal of Gastroenterology. 2006;**41**:1169-1177

[215] Mazur G, Bogunia-Kubik K, Wrobel T, Kuliczkowski K, Lange A. TGF-beta 1 gene polymorphisms influence the course of the disease in non-Hodgkin's lymphoma patients. Cytokine. 2006;**33**:145-149

[216] Peters CA et al. TGFB1 single nucleotide polymorphisms are associated with adverse quality of life in prostate cancer patients treated with radiotherapy. International Journal of Radiation Oncology, Biology, Physics. 2008;**70**:752-759

[217] Wei YS, Xu QQ, Wang CF, Pan Y, et al. Genetic variation in transforming growth factor-beta 1 gene associated with increased risk of esophageal squamous cell carcinoma. Tissue Antigens. 2007;**70**:464-469

[218] Guan X, Zhao H, Niu J, Tan D, et al. Polymorphisms of TGFB1 and VEGF genes and survival of patients with gastric cancer. Journal of Experimental & Clinical Cancer Research. 2009;**28**:94

[219] Shu XO, Gao YT, Cai Q, Pierce L, Cai H, Ruan ZX, et al. Genetic polymorphisms in the TGF-β1 gene and breast cancer survival: A report from the Shanghai Breast Cancer Study. Cancer Research. 2004;**64**(3):836-839

[220] Zhang L, Wu G, Herrle F, Niedergethmann M, et al. Single nucleotide polymorphisms of genes for egf, tgf-β and tnf-α in patients with pancreatic carcinoma. Cancer genomics and proteomics. 2012;**9**:287-296

[221] Muller-Hermelink N, Braumuller H, Pichler B, et al. TNFR1 signalling and IFN-gamma signalling determine whether T cells induce tumor dormancy or promote multistage carcinogenesis. Cancer Cell. 2008;**13**:507-518

[222] Gomes M, Coelho A, Araujo A, Teixeira AL, Catarino R, Medeiros R. Influence of functional genetic polymorphism (-590C/T) in non-small cell lung cancer (NSCLC) development: The paradoxal role of IL4. Gene. 2012; **504**:111-115

[223] Bozdogan ST, Erol B, Dursun A, et al. The IL-1RN and IL-4 gene polymorphisms are potential genetic markers of susceptibility to bladder cancer: A case-control study. World Journal of Urology. 2015;**33**:389-395

[224] Kaminska B, Kocyk M, Kijewska M. Tgf beta signaling and its role in glioma pathogenesis. Advances in Experimental Medicine and Biology. 2013;**986**:171-187

[225] Mingjun H, Jieli D, Lihong C, Tingqin H, et al. IL-10 and PRKDC polymorphisms are associated with glioma patient survival. Oncotarget. 2016;**7**(49):80680-80687

[226] Tanikawa T, Wilke CM, Kryczek I, Chen GY, Kao J, Nunez G, et al. Interleukin-10 ablation promotes tumor development, growth, and metastasis. Cancer Research. 2012;**72**:420-429

[227] Saxena R, Chawla YK, Verma I, Kaur J. Interleukin-1 polymorphism and expression in hepatitis B virus-mediated disease outcome in India. Journal of Interferon & Cytokine Research. 2013; **33**:80-89. DOI: 10.1089/jir.2012.0093

[228] Ma J, Liu YC, Fang Y, Cao Y, Liu ZL. TGF-beta 1 polymorphism 509 C> T is associated with an increased risk for hepatocellular carcinoma in HCV-infected patients. Genetics and Molecular Research. 2015;**14**(2):4461-4468

[229] Lu Y, Wu Z, Peng Q, Ma L, Zhang X, Zhao J, et al. Role of IL-4 gene polymorphisms in HBV-related hepatocellular carcinoma in a Chinese population. PLoS One. 2014;**9**(10): e110061

[230] Wu, Z., Qin, W., Zeng, J., Huang, C., Lu, Y., & Li, S. (2015). Association between IL-4 polymorphisms and risk of liver disease: an updated meta-analysis. Medicine, **94**(35)

[231] Tang S, Yuan Y, He Y, Pan D, Zhang Y, Liu Y, et al. Genetic polymorphism of interleukin-6 influences susceptibility to HBV-related hepatocellular carcinoma in a male Chinese Han population. Human Immunology. 2014;**75**:297-301. DOI: 10.1016/j.humimm.2014.02.006

[232] Li Y, Du Z, Wang X, Wang G, Li W. Association of IL-6 promoter and receptor polymorphisms with multiple myeloma risk: A systematic review and meta-analysis. Genetic Testing and Molecular Biomarkers. 2016;**20**(10):587-596

[233] Ognjanovic S, Yuan JM, Chaptman AK, Fan Y, et al. Genetic polymorphisms in the cytokine genes and risk of hepatocellular carcinoma in low-risk non-Asians of USA. Carcinogenesis. 2009;**30**:758-762. DOI: 10.1093/carcin/bgn286

[234] Wu J, Lu Y, Ding YB, Ke Q, Hu ZB, Yan ZG, et al. Promoter polymorphisms

of IL2, IL4, and risk of gastric cancer in a high-risk Chinese population. Molecular Carcinogenesis. 2009;**48**(7): 626-632

[235] Chen GY, Nuñez G. Sterile inflammation: Sensing and reacting to damage. Nature Reviews Immunology. 2010;**10**(12):826-837

[236] Namkung JH, Lee JE, Kim E, Park GT, Yang HS, Jang HY, et al. An association between IL-9 and IL-9 receptor gene polymorphisms and atopic dermatitis in a Korean population. Journal of Dermatological Science. 2011;**62**(1):16-21

[237] Zhang, L., Wu, G., Herrle, F., Niedergethmann, M., & Keese, M. (2012). Single nucleotide polymorphisms of genes for EGF, TGF-β and TNF-α in patients with pancreatic carcinoma. Cancer genomics & proteomics. **9**(5):287-295

[238] Gonzalez-Zuloeta Ladd AM, Arias-Vasquez A, Siemes C, Coebergh JW, et al. Transforming growth factor beta 1 leu 10pro polymorphism and breast cancer morbidity. European Journal of Cancer. 2007;**43**:371-374

[239] Schaaf BM, Boehmke F, Esnaashari H, Seitzer U, et al. Pneumococcal septic shock is associated with the interleukin-10-1082 gene promoter polymorphism. American Journal of Respiratory and Critical Care Medicine. 2003;**168**(4):476-480

[240] Kornman KS. Interleukin 1 genetics, inflammatory mechanisms, and nutrigenetic opportunities to modulate diseases of aging. The American Journal of Clinical Nutrition. 2006;**83**(2):475S-483S

[241] Schoof N, Franklin J, Fürst R, Zander T, von Bonin F, Peyrade F, et al. Interleukin-10 gene polymorphisms are associated with freedom from treatment failure for patients with Hodgkin lymphoma. The Oncologist. 2013;**18**: 80-89

[242] Li G, Li D. Relationship between IL-10 gene polymorphisms and the risk of non-Hodgkin lymphoma: A meta-analysis. Human Immunology. 2016; **77**(5):418-425

Chapter 5

Genetic Polymorphism and Prostate Cancer: An Update

Surayya Siddiqui, Sridevi I. Puranik, Aimen Akbar and Shridhar C. Ghagane

Abstract

Genetic polymorphism and prostate cancer (PC) are the most pernicious and recurrently malignancy worldwide. It is the most dominating cause of cancer related casualty among men in the US. Asian countries are inflicted with PC at an alarming rate though still the prevalence of PC is lower than European and American men. Some of the genetic and environmental factors that might play a role in PC risk include: age genetic predilection, family history, race/ethnicity, lifestyle, and dietary habits and non-dietary environmental risk factors such as smoking. Socio-economic factors including economic, scholastic and intellectual factors do not, intrinsically seem to straight away influence the risk of acquiring PC. Other genetic changes that may support an increased risk of developing PC include *HPC1*, *HPC2*, *HPCX*, *CAPB*, *ATM,s HOXB13* and mismatch repair genes. PC occurrence rates are highly variable. Almost all PC mortalities are due to metastatic disease, generally through tumors the progress to be hormone refractory or castrate resistant. PC, developing research has acknowledged a number of candidate genes and biological pathways associated with PC. Indirect pathways such as P13K/AKT signaling pathway is one of most well known alternate pathway in PC Vascular endothelial growth factor (VEGF) is widely known to be potent stimulator of angiogenesis. The over expression of EGFR in a very large majority of cases is accompanied by the succession of PC, implying that this may play a mechanistic role. Numerous occupational factors have been proposed to cause PC. Some of the risk factors include; farmers/agricultural workers, pesticides, shift work and flight personnel. PC treatment can be done through surgery, radical prostatectomy is the main type of surgery. Risks of injury are many – reactions to anesthesia, loss of blood, blood clumps in the legs/lungs, injury to surrounding organs, infection at the site of surgery and many more. The other treatments are hormone therapy, chemotherapy and radio therapy chemotherapy. Chemotherapeutic drugs are typically used one at a time for PC such as transurethral resection of prostate (TURP). Some of the chemotherapeutic drugs are Docetaxel, Cabazitaxel, Mitoxantrone and Estramustine. Among the score of biomarkers being studied, numerous markers and techniques deserve awareness and acceptability for both patients and urologists in clinical practice.

Keywords: prostate, *BRCA2*, *P13K/AKT* signaling pathway, single nucleotide polymorphisms (SNPs)

1. Introduction

Cancer is both genetically and phenotypically sophisticated. Among all the cancers known so far, prostate cancer (PC) is the most pernicious and recurrently diagnosed malignancy worldwide (after lung cancer), counting 1,414,259 new cases (7.3% of all cancer associated diseases) and causing 375,304 deaths (3.8% of all deaths caused by cancer in men) in 2020 [1]. It is the second most dominating cause of cancer related casualty among men in the United States. Prostate cancer is expected to cause 248,530 new cases and 34,130 deaths in the United States in 2021, making it the most common cancer among American men [2]. Prostate cancer is thought to contain a genetic makeup that includes somatic copy number changes, point modifications, structural rearrangements, and chromosomal number differences [3]. The somatic transformation rate of prostate cancer lies between $1x10^{-6}$ and $2x10^{-6}$ which is identical to breast, renal and ovarian cancer [4]. However, in addition to somatic genetic alterations, it has been discovered that prostate cancer is associated with the loss of tumor suppressor gene activity due to epigenetic changes in their expression. Prostatic adenocarcinoma is the most frequent type of prostate cancer, accounting for nearly all occurrences. It grows in the cells of the glands that produce prostatic fluid. A digital rectal exam (DRE), a prostate-specific antigen (PSA) blood test, a prostate biopsy, or a computed tomography (CT) scan are all common ways for a clinician to detect prostate cancer [5].

Prostate cancer is categorized as the third highest among cancers of the male urinary and reproductive system in China. In Chinese populations, this condition impairs the quality of life and longevity of men over the age of 50 [6]. India presently has a population of over a billion people, with an estimated 1.5 million cancer cases reported each year. Asian Indians were diagnosed at an average age of 65.9 years (range: 46–86 years). India has a low incidence of prostate cancer. However, due to a dearth of prostate cancer screening, the majority of individuals are detected at an advanced stage [7]. The percentage of individuals with late-stage prostate cancer who have both high-grade illness and bone metastases is 65 percent. African-American men had a greater occurrence rate and a more severe kind of prostate cancer than White men. Prostate cancer is currently afflicting Asian countries at an alarming rate. Still, the prevalence of prostate cancer in Asian men is comparatively lower than European and American men. One reason that is endorsed by epidemiologic studies of immigrant populations and by recent somatic genomic alteration analysis is due to distinct genetic backgrounds. The etiological factors correlated with prostate cancer are still unresolved and ambiguous [8]. Variations in social, environmental, and genetic factors are thought to be the cause of this disparity. Some of the factors that may influence the risk of prostate cancer include:

- Age, genetic predisposition,
- family history
- race/ethnicity
- lifestyle and food habits are all factors to consider.

Although hereditary factors have been suspected to have a role in prostate cancer etiology, only a few genetic markers linked to prostate cancer risk have been identified. The strongest risk factor is acknowledged to be a hereditary liability [9]. Variations in the usage of diagnostic tests are reflected in the incidence rates around

the world. The importance of family history in obtaining evidence for major genetic determinants in cancer prevalence cannot be overstated. Prostate cancer progression and progression are influenced by lifestyle and nutritional habits. Future research is needed to better understand the interaction of environmental and diaspora factors [10]. Prostate cancer is generally asymptomatic at the initial stage and may require minimal or no treatment at all. However, the most recurring grievance is complications with urination and nocturia. Prostatic hypertrophy causes all of the symptoms listed above. Because the axial skeleton is the most prevalent site of bone metastatic disease, more advanced stages of the disease may exist with urine retention and back pain [11].

Genetics has an important role in the initiation and progression of prostate cancer. Tumor biomarkers are used to detect, treat, and diagnose many tumors in their early stages. Understanding the genesis and causative risk factors for prostate cancer can help to provide some solutions. This can aid in the expansion of significant screening and precluding methods for prostate cancer. Currently, no evidence supports how to prevent this cancer [12]. A promising understanding of the etiology and causative risk components is crucial for understanding cancer prevention. This cancer has a complex, multiple etiology, with 42 percent genetic factors and 58 percent environmental/lifestyle factors estimated to be involved. Prostate cancer is a growing threat to Asian men's health. Although the rate of incidence of prostate cancer is far greater in Western cultures (120 per 100,000 in Northern America) than in East Asian cultures (less than 10 per 100,000 in Asia), when Asians migrate to Western countries, their rate of prostate cancer incidence increases [13]. Prostate cancer has all of the criteria of an excellent chemoprevention target illness, including a lengthy latency, high incidence, a huge proportion of tumor markers, and identifiable preneoplastic lesions.

The increased morbidity and mortality from prostate cancer necessitate the implementation of current and effective preventive measures in everyday life. In conclusion, no vital known lifestyle or infectious agent is recognized as a risk factor for prostate cancer, categorizing it as one of the limited widespread cancers with unidentified risk factor [14].

2. Occurrence rates

Prostate cancer occurrence rates are highly variable. Occurrence, stringency, and mortality rates vary with ethnicities, geological location, and age. Its risk increases with an increase in age with the youngest being least susceptible. Also, this cancer is more widespread and invasive in African-Americans [15]. Native Asians, on the other hand, have a remarkably low incidence of prostate cancer, which is likely due to a combination of environmental and hereditary factors. Prostate cancer has the highest occurrence of 83.4% per 1000,000 incidences in Northern Europe, notably Ireland, and the lowest occurrence rate in South Central Asia, at 6.3 per 100,000. In Asia, however, prostate cancer accounts for only 1–10% of cases, but the incidence of occurrence is rapidly increasing [16].

Prostate cancer ranks second (37.5 per 100,000) in nations with a high Human Development Index (HDI), and vice versa in countries with a low HDI (11.3 per 100,000). The highest occurrence rates of prostate cancer are found among black men in the United States and the Caribbean, confirming the importance of Western African ancestry in controlling prostate cancer risk. The variation in prostate cancer incidence rates is mainly due to different prostate cancer diagnostic practices globally. In the early 1980s and late 1990s, there was a drastic increase in prostate cancer cases in the United States, Australia, and Canada due to which the detection

of pre-symptomatic cancer was feasible [17]. The adoption of prostate-specific antigen (PSA) testing was a major factor in this. One of the most striking features of prostate cancer is the wide range of occurrence rates caused by differences in race, geography, or age, as well as variations in environmental factors such as work environment, nutrition, and lifestyle, all of which can increase the risk of precocious prostate cancer. Environmental factors are one of the reasons for varied incidence rates of different cancers globally [18]. The cause of prostate cancer is unknown, even though it is more common in some populations. However, it is clear that prostate cancer is a multifaceted illness with various genetic and environmental variables contributing to its etiology.

3. Mortality rates

Almost all deaths from prostate cancer are caused by metastatic illness, which occurs when tumors become hormone-refractory or castrate-resistant. Prostate cancer mortality rates have dropped dramatically in Northern America, Northern, and Western Europe, indicating improved treatment and early detection through increased screening [19]. The highest mortality rate of prostate cancer is 27.9 per 100,000 in the Caribbean, with the lowest mortality rate being 3.1 per 100,000 in South Central Asia. The highest occurrence and mortality rates are found in African-American men. As a result, we can speculate that African-American males may maintain some genes that are more susceptible to mutations in prostate cancer and that these mutations are linked to a more aggressive type of cancer. African-American men, as well as men with a family history, should be screened at the age of 45. Undeniable death is the most important prostate cancer endpoint [20].

4. Genes associated with prostate cancer risk

Developing research has acknowledged a number of candidate genes and biological pathways associated with increased susceptibility to cancer. A significantly recent direction for scientists to identify genes involved in human disease is Genome-wide association studies (GWAS) [21]. This approach looks for tiny differences in the genome known as SNPs (single-nucleotide polymorphisms). Because genome-wide association studies examine SNPs across the genome, they promise a promising technique to research complex, prevalent diseases in which numerous genetic variations contribute to a person's risk. Genome-wide association studies provide an enormous approach for the identification of genetic markers correlated with prostate cancer risk. Hundreds of thousands of SNPs can be examined simultaneously in each analysis [22]. In genome-wide association studies (GWAS), case–control surveys, multiple linkage estimations, next-generation sequencing (NGS), and admixture mapping studies, several genes and chromosomal areas have been identified to be associated with prostate cancer. As more prostate cancer risk variations are identified, the cumulative effects of these variants may become increasingly important clinically. Unfortunately, attempts at specifying a reliable biomarker have thus far proved futile (**Table 1**) [15–17].

There is a significant role of genetics in prostate cancer that is indicated through epidemiological studies. In the advancement and succession of prostate cancer, there is an involvement of a heightening number of single nucleotide polymorphisms (SNPs). Individual SNPs show a moderate connection with prostate cancer risk, but when they are combined, they have a larger, dose-dependent association that presently accounts for 30% of prostate cancer family risk. An SNP is a

Genes	Population	N (Total Cases Studied)	Prostate cancer risk	Reference
BRCA1	UK and IRELAND	376	Weakly validated	Nyberg et al., 2020
BRCA2	UK and IRELAND	447	Validated	Nyberg et al., 2020
AR	NON-HISPANIC WHITE	226	Validated	Ingles et al., 1997
HOXB13	NORTHERN EUROPEAN	3508	Validated	Cooney et al., 2016
CYP17A1	TUNISIAN	250	Validated	Souiden et al., 2011
SRD5A2	CHINESE	495	Validated	Hsing et al., 2001
TLR4	NORTH INDIAN Asian	398	Validated	Singh et al., 2013
HNF1B	EUROPEAN AMERICANS, JAPANESE and AFRICAN AMERICANS	483	Validated	Grisanzio et al., 2012
MSMB	SOUTHERN CHINESE HAN	509	Validated	Xu et al., 2010
JAZF1	EUROPEAN	19421	Validated	Prokunina-Olsson et al., 2010
ESR1	CAUCASIAN	1859	Validated	Nicolaiew et al., 2009
(*RNase L*) *HPC1*	SWEDISH	2425	Not validated	Wiklund et al., 2004
ELAC2/HPC2	AUSTRALIAN	1557	Not validated	Severi et al., 2003
MSR1	CHINESE	410	Validated	Hsing et al., 2007
PSA	JAPANESE	782	Not validated	Wang et al., 2003
HPCX	ASHKENAZI JEWISH	2230	Validated	Agalliu et al., 2010

Table 1.
Genes associated with prostate cancer.

deviation from the predicted nucleotide in a DNA sequence that arises when a single nucleotide (A, T, C, or G) in the genome changes [20–24]. SNPs are thought to make a significant impact on disease vulnerability. SNPs are simple to find and only exist once, making them an ideal biomarker. Because of the growing interest in the role of SNPs in prostate cancer progression and succession, a large number of studies on SNPs in prostate cancer are being published [25].

Many case–control studies have recognized innumerable single nucleotide polymorphisms (SNPs) correlated with prostate cancer but the clinical role of these SNPs remains ambiguous. Some SNPs also affect levels of the prostate-secreted proteins, prostate-specific antigen (PSA), hexokinase (HK) 2, and β-microseminoprotein (β-MSP); prostate cancer risk was also related with some of these SNPs. Numerous SNPs associated with prostate cancer affect the role and/or generation of a prostate cancer marker. For example, rs198977 in *KLK2* may affect hK2 function and decrease hK2 levels in blood [26]. The SNP rs10993994 in *MSMB* reduces levels of β-MSP and is correlated with increased levels of PSA in the blood or semen of an active young man. The function of cell cycle dysregulation in prostate cancer vulnerability has been described via changing the cell's ability to respond effectively to DNA damage. Patients with advanced or metastatic prostate cancer have been reported to exhibit modifications in the *CDKN2* gene but not in primary tumors (**Table 2**) [27].

SNP	Chromosome	Population	N (Total Cases Studied)	Prostate Cancer Risk	Reference
rs6983267	8q24	Caucasians and Asians	50854	Identified	Li et al., 2015
rs1447295	8q24	Caucasians and Asians	50854	Identified	Li et al., 2015
rs16901979	8q24	Africans Americans	50854	Identified	Li et al., 2015
rs138213197	HOXB13	Europeans	9012	Identified	Beebe-Dimmer et al., 2015
rs4242382	8q24	Asian and Caucasian	3657	Identified	Zhao et al., 2014
rs4430796	17q12	Non-Hispanic White	421	Identified	Levin et al., 2008
rs7501939	17q12	Non-Hispanic white	421	Identified	Levin et al., 2008
rs10896449	11q13	European	19395	Identified	Chung et al., 2011
rs10486567	7p15.2	Finnish	947	Not identified	Chen et al., 2014
rs1938781	11q12	Japanese	5560	Identified	Akamatsu et al., 2012
rs2252004	10q26	Japanese	5560	Identified	Akamatsu et al., 2012
rs2055109	3p11.2	Japanese	5560	Identified	Akamatsu et al., 2012
rs1859962	17q24	Non-Hispanic white	421	Not Identified	Levin et al., 2008
rs12793759	11q13	European	19395	Identified	Chung et al., 2011
rs3737559	2p15	Icelander	23205	Identified	Gudmundsson et al., 2008
rs5945572	Xp11.22	Icelander	23205	Identified	Gudmundsson et al., 2008

Table 2.
SNPs associated with prostate cancer.

The t-allele of *MDM2* was found to be the most strongly linked to prostate cancer. The presence of at least one copy of the t allele of MDM2 tSNP309g increases the risk of advanced prostate cancer. The association of the rs1447295 A allele with prostate cancer was reported. However, no substantial relation of the rs6983267 G allele with prostate cancer patients was recorded. Seven single nucleotide polymorphisms (SNPs) on chromosome *17q*, 3 SNPs on *17q12*, and 4 SNPs on *17q24.3* are associated with the risk of prostate cancer was recognized in the European population by a genome-wide association study (GWAS) [25–28]. Research on susceptibility genes is one of the hottest subjects in prostate cancer risk factors. Even still, confirming the prostate cancer susceptibility genes has proven difficult. In 2007, in a GWAS conducted in the United States based on 3 known risk loci of

prostate cancer, SNPs rs4242382 and rs6983267 on 8q24 and SNP rs4430796 on 17q in the hepatocyte nuclear factor 1B (*HNF1B*) gene, 4 SNPs associated with prostate cancer, rs10993994 on chromosome 10 in the microseminoprotein-beta (*MSMB*) gene, rs4962416 on chromosome 10 in the C-terminal binding protein 2 (*CTBP2*) gene, rs10896449 on 11q13, and rs10486567 on chromosome 7 in the juxtaposed with another zinc finger protein 1 (*JAZF1*) gene was observed [29]. Two loci within the gene, SNPs rs1799950, and rs3737559, have been associated with early-onset prostate cancer and genealogical prostate cancer, respectively.

The steroid *SRD5A2* gene, located on chromosome 2p23 is a major gene that makes it prone to prostate cancer. Evidence supporting the linkage of this gene to prostate cancer is provided [30]. The transformation of testosterone (T) leads to dihydrotestosterone (DHT), the most powerful androgen receptor adversary in prostate cells by 5α-reductase (5-AR) enzymes (types 1 and 2). Under normal physiological conditions, a 5-AR enzyme that is encoded by *SRD5A2* is inevitably expressed over *SRD5A1* in the prostate. Attractive targets for preventing prostate cancer advancement are represented by 5-AR enzymes. The hypothetical basis for chemoprevention strategies is the inhibition of 5-AR enzymes and the reduction of prostate cancer succession in low-risk disease was also lately shown [30–33]. To potentially advance cancer growth and proliferation, *SRD5A* genetic variations alter sex-steroid exposure. Further studies are required to fully comprehend the influence of functional variations in *SRD5A* genes in both normal and prostate cancer cells.

Androgens are widely known to play a significant role in the progression of prostate cancer, even until it reaches advanced stages. In prostate cancer, especially castration-resistant prostate cancer, the androgen receptor (AR) plays a critical role (CRPC) [34]. Androgen deprivation therapy suppresses hormone-naive prostate cancer; however, AR is altered by prostate cancer and adapts for survival at castration levels of androgen. The AR gene can be found on chromosome X (Xq11–12). Shorter glutamine repeats have been linked to increased AR transcriptional activity [35]. The role of epithelial AR is to send secretory proteins to the prostate gland, such as prostate-specific antigen (PSA), whereas stromal AR is involved in prostate development. Most prostate cancers can be suppressed with androgen deprivation therapy, however, some high-risk prostate tumors progress to castration-resistant prostate cancer, which then thrives under castrated testosterone levels [36]. AR is the most frequent anomalous gene in metastatic CRPC. Several mechanisms play a fundamental role in the development of metastatic CRPC: Point mutations in the androgen receptor, Androgen receptor amplification, Variations of androgen biosynthesis, Variations in androgen receptor cofactor in the prostate cancer, and Androgen receptor variants [37].

The binding of anti-androgens such as bicalutamide and flutamide to the ligand-binding domain (LBD) of AR inhibits androgen binding to LBD. In the progression of prostate cancer to castration-resistant prostate cancer, prostate cancer survives and restarts its growth under castration levels of androgen. A second-generation non-steroidal anti-androgen with greater affinity for the LBD of AR is Enzalutamide [38]. Androgens play a critical role in the progression of both normal prostate epithelium and stromal prostate cancer, and activation of genes involved in androgen metabolism may be linked to an increased prostate cancer risk.

A fundamental regulatory enzyme in the steroidogenic pathway is Cytochrome P450 17α-hydroxylase/17,20-lyase (*CYP17A1*); Both 17α-hydroxylase and 17,20-lyase activities are catalyzed by *CYP17A1* and it is crucial for the production of both androgens and glucocorticoids [39]. *CYP17A* is a target for the hormonal treatment of prostate cancer. Many recent investigations have found that *CYP17A1* is

significantly expressed in more than 50% of human prostate carcinomas, indicating that cancer cells synthesize androgen intracellularly. The degree of expression was directly linked to nuclear expression of the phosphorylated active form of ARs. *CYP17A1* plays a crucial role in adrenal and intratumoral de novo biosynthesis of androgens. Abiraterone is a suppressor of *CYP17A1*. Current genes of interest that have been identified as genealogical tumor suppressor genes and act as biomarkers for prostate cancer include *RNase L* (*HPC1*, *1q22*), *MSR1* (8p), *ELAC2*/*HPC2* (17p11). Pathogenic variations in genes, such as *BRCA1*, *BRCA2*, the mismatch repair genes, and *HOXB13* exchange views on subtle to a mild lifetime risk of prostate cancer [40].

In the breast cancer predisposition gene 2 (*BRCA2*), germline mutations are the genetic events known to date that confer the highest risk of prostate cancer (8.6-fold in men ≤65 years). BRCA gene mutations, particularly in *BRCA2*, enhance the likelihood of developing PCa and have implications for disease prognosis and management [41]. *BRCA2* and *BRCA1*'s roles in prostate tumorigenesis are still unknown. The gene *BRCA1* has been linked to an increased risk of sporadic prostate cancer. The relevance of *BRCA1* and *BRCA2* to male cancer has been extensively researched since families with these mutations demonstrate clustering of cancer in men. Both *BRCA1* and *BRCA2* are tumor suppressor genes that are inherited in an autosomal dominant, incomplete penetrance manner [42]. Although some studies have begun to assess the involvement of these genes in prostate cancer, the precise role of *BRCA1* and *BRCA2* in prostate cancer progression and progression has yet to be determined. BRCA1 is one of the co-regulators of the androgen receptor (AR), which regulates a signaling pathway important for prostate cancer progression and progression. *BRCA2* was more common in castrate-resistant prostate cancer than primary prostate cancer, further inferring the invasive nature of prostate cancer in *BRCA2* mutation carriers [43]. Some, like *BRCA2*, are beginning to show clinical promise in the treatment and screening of prostate cancer. Numerous studies have shown that the common *BRCA1* and *BRCA2* mutations found in breast and ovarian cancer families also increase the risk of prostate cancer in Ashkenazi Jewish populations [41–44]. The odds ratios range from 2.1 to 4.8 and principally reach statistical significance for *BRCA2* but not *BRCA1*. Further, studies in various populations assist a role for *BRCA2*. *RNase L* (2′–5′ oligoadenylate-dependent ribonuclease L) (*HPC1*), located at chromosome 1q22; *MSR1*, with a linkage region on chromosome 8p; and *ELAC2* (*HPC2*), on chromosome 17p11 are the most pertinent candidate genes for prostate cancer [45]. In prostate cancer, these three genes have been identified as inherited tumor suppressor genes. The key changes in the *RNase L* gene linked to prostate cancer include D541E, R462Q, and I97L missense mutations. R462Q and D541E are antonymous variations of *RNase L* that exhibit a decrease in enzymatic activity [46]. In hereditary situations, the R462Q or Arg462G ln variation is linked to an increased prostate cancer risk in the Finnish population, American Caucasians, and Japanese men. I97L is a missense mutation in the ankyrin domain's third and seventh repeat sites. There is no clear link between this mutation and the risk of prostate cancer [47].

Hereditary prostate cancer 1 gene (*HPC1*) (*1q24–25*), *CAPB* (*1p36*), *PCAP* (*1q42–43*), *HPC2* (*17p12*), *HPC20* (*20q13*), and *HPCX* (*Xq27–28*) are the genes responsible for the advancement of sporadic and, in particular, familial prostate cancer. Certain regions in the *MSR1* gene appear to play a role in prostate cancer. *rs12718376*, in the terminal 3= regions of gene *MSR1*, was characterized as linked to a heightened prostate cancer risk in the Caucasian population [48]. With *HOXB13* bearers, a subjective history of invasive cancer is strongly retained. The findings

of Beebe-Dimmer et al. show that the G84E mutation in the *HOXB13* gene affects about 0.5 percent of people of European heritage, confirming previous claims that the mutation is linked to prostate cancer [49].

5. Alternate pathways of PC risk

5.1 *PI3K/Akt* signaling pathway

PI3K/Akt signaling pathway is one of the most well-known alternate pathways in prostate cancer. This pathway's activation appears to be common in many aggressive prostate tumors. Additionally, as prostate cancer develops toward a resistant, metastatic condition, the *PI3K/Akt* pathway is more frequently activated [43–45]. Cellular metabolism, tumor genesis, growth, proliferation, metastasis, and cytoskeletal remodeling are all controlled by this signaling system. The activation of growth and survival pathways is one way that the PI3K pathway might cause cancer.

Further, there is an involvement of the *PI3k/Akt* pathway in prostate cancer with modulation of DNA damage repair pathway [46]. Furthermore, the *PI3K/Akt* pathway is involved in modifying a more invasive phenotype in prostate cancer cells through modulating cholesterol ester production. Phosphatases such as phosphatase and tensing homolog gene (PTEN), PH and leucine-rich repeat protein phosphatase (PHLPP), cellular prostatic acid phosphatase, PP2A, and INPP4B activate the *PI3K/Akt* pathway. The *PI3K/Akt* pathway has also been demonstrated to be important in the survival and proliferation of prostate cancer stem cells. The link between the *PI3K/Akt* and AR pathways has piqued researchers' curiosity as a potential co-targeting method in prostate cancer. In various preclinical studies, reciprocal connections between these pathways have been demonstrated. These preclinical discoveries have aided the advancement of clinical trials including the combined inhibition of both the AR and the PI3k/Akt/mTOR pathways. With the invasive oncogenic properties of the *PI3K/Akt* pathway, there has been a lot of interest in using it as a biomarker to discriminate more significantly [48].

Although present research suggests that this pathway gives predictive information, it is unclear whether it offers many significant advantages over currently employed clinical and pathologic indicators. However, there is continued interest in using this pathway as a predictive biomarker for newly targeted medicines of this pathway. Activation of the *PI3K/Akt* pathway is undoubtedly important in the invasive character of many prostate tumors. As more individuals develop non-AR-driven tumors, this non-androgen receptor pathway may become increasingly relevant with the use of contemporary AR pathway inhibitors and combination therapy.

5.2 Vascular endothelial growth factor (VEGF)

In both healthy and pathological settings, vascular endothelial growth factor (*VEGF*) is widely recognized as a potent activator of angiogenesis. Most solid tumors, including prostate cancer, have strong evidence of it. Prostate cancer is aided by vascular endothelial growth factor (*VEGF*) and pigment epithelium-derived factor (PEDF) [49]. Angiogenesis appears to play a significant role in prostate cancer, according to sufficient data. Secretion of protein factors such as the vascular endothelial growth factor (*VEGF*) is exuded by prostate cancer cells which are widely studied and known as the major angiogenic marker. Many variables control and regulate the *VEGF* pathway, including local environmental hypoxia and different hormones, growth factors, and cytokines.

5.3 Epidermal growth factor receptor (EGFR)

The overexpression of *EGFR* (epidermal growth factor receptor) is associated with the progression of prostate cancer in the vast majority of cases, showing that this may play a mechanistic role. The first known member of the HER receptor family was the Epidermal growth factor receptor (EGFR or HER-1). After total prostatectomy, the risk of recurrence and succession to hormone resistance is associated with the expression of *EGFR*. One hundred percentages of metastases of hormone-refractory prostate cancers express *EGFR*, implying that this receptor is a primary transduction pathway for tumor growth [50]. Only a few clinical trials using combos with anti-EGFR drugs have lately been conducted. Erlotinib is a reversible and orally active *EGFR* tyrosine kinase inhibitor that stops the cell cycle in the G1 phase. The preclinical results imply that combining additional targeted treatments with *EGFR*, particularly antiangiogenics, should be studied.

5.4 Single nucleotide polymorphisms (SNPs)

In addition, GWAS has identified more than 150 SNVs related to the advancement of prostate cancer, but the clinical advantage of these findings remains skeptical.

5.4.1 8q24

The *8q24* polymorphisms have been implicated in various cancers. Overall, each of the *8q24* polymorphisms was individually correlated with prostate cancer risk. Substantial relations were also observed in an analysis by ethnicity, source of control, and quality score. Interestingly, the effect of rs1447295 on prostate cancer risk was observed among Caucasians and Asians, but not Africa-Americans. The effect of rs16901979 was more eminent among Africa-Americans than Asians [37]. Similarly, rs6983267 conferred a higher prostate cancer risk among Caucasians than Asians. Collaboratively, these *8q24* variants (s) may regulate prostate cancer risk in an ethnic-specific manner. *8q24* rs4242382-A polymorphism was associated with prostate cancer risk in Chinese men.

5.4.2 17q12 and 17q24

Gudmundsson and colleagues identified two prostate cancer susceptibility loci on chromosome 17q in a recent genome-wide association study. Precisely, there was an association of two intronic SNPs in the *TCF2* gene (rs4430796 and rs7501939) at *17q12* and a third SNP (rs1859962) at *17q24* with the sporadic prostate cancer risk. The strongest evidence of prostate cancer association was observed in SNPs rs4430796 and rs7501939. Moreover, the results obtained by Levin et al., suggest that the increased risk associated with these SNPs is approximately doubled in individuals making them susceptible to develop the timely-onset disease. While the results for correlation of SNP rs1859962 at *17q24* were not statistically substantial, there was symbolic evidence that the "G" allele was over-transmitted to affected men [38–41].

5.4.3 HOXB13 gene

In the *HOXB13* gene, an exotic modification named as *G84E* (*rs138213197*) has been identified according to the studies conducted in 2012, which has been categorized as the first major genetic variation linked with familial prostate cancer.

The function of *HOXB13* in prostate cancer is not well recognized [43]. The *HOXB13* protein is known to be critical in the embryonic development of the prostate gland and is expressed in normal prostate tissue toward adulthood. There is an indication that *HOX* genes may act as both, an oncogene and a tumor suppressor gene in the prostate as well as other cancers. The carriers of the G84E mutation represent 0.5% of the total population and are found almost exclusively among patients of known European ancestry (**Figures 1** and **2**).

5.4.4 11q13

The prostate cancer susceptibility alleles on chromosome *11q13* have been identified by genome-wide association studies. Chung et al. reported the results of fine-mapping the region surrounding the most notable SNP, rs10896449, originally associated with prostate cancer risk in *11q13*. Numerous locus models that included consequential SNPs consecutively identified that the second association of rs12793759 is independent of rs10896449 and remained significant [8].

5.4.5 JAZF1 gene

Chen et al. demonstrated that two loci (rs4242382 and rs10486567) are highly associated with familial multiple prostate cancer. A stronger effect of the risk allele

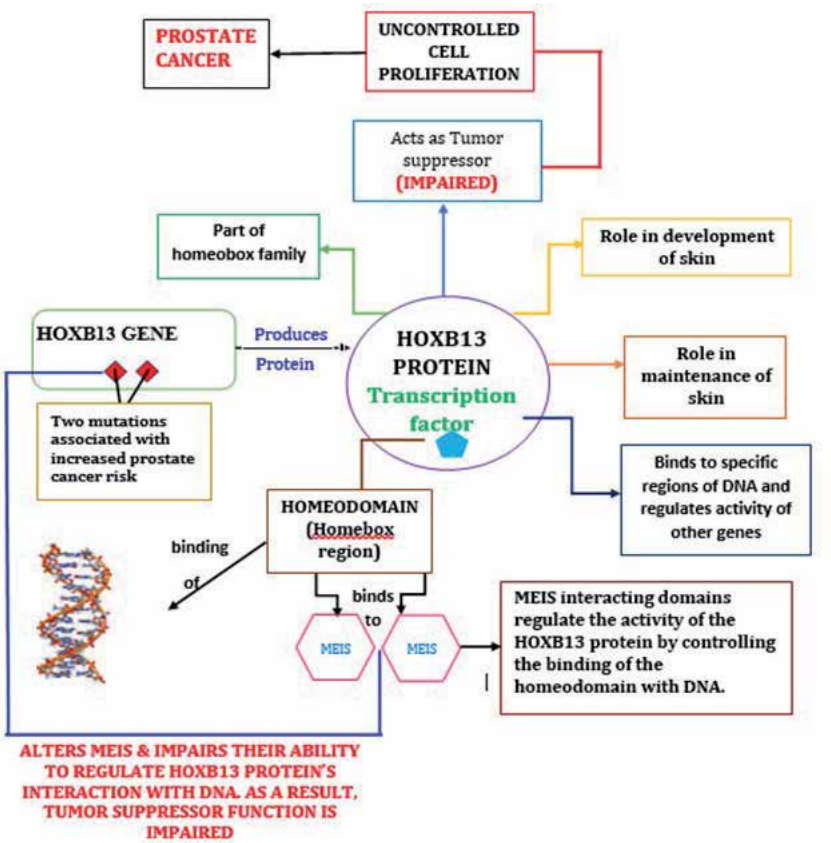


Figure 1.
Mechanism of action of HOXB13 gene.

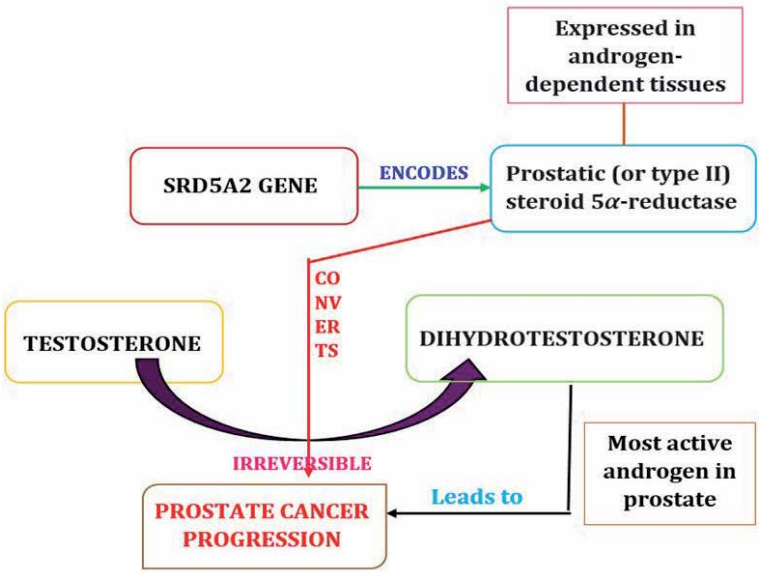


Figure 2.
Mechanism of action of SRD5A *gene.*

A at rs10486567 was observed in the presence of the risk allele A at rs4242382. The SNP rs10486567 at 7p15.2 was consistent with the genome-wide study. A three C2-H2-type zinc finger protein, which is a transcriptional repressor of *NR2C2*, a nuclear orphan receptor is encoded by the SNP rs10486567 located within intron 2 of the *JAZF*1 gene on chromosome 7p15.2. This protein is highly expressed in prostate tissue and interacts with the androgen receptor [9].

5.4.6 11q12*,* 10q26 *and* 3p11.2

We have identified three new loci, 11q12, 10q26, and 3p11.2, that are associated with prostate cancer susceptibility at a genome-wide significance level in the Japanese population. These findings provide additional support for conducting GWAS for prostate cancer in diverse populations to identify risk loci for this genetically heterogeneous cancer. Akamatsu et al. showed that there are three new loci, 11q12, 10q26, and 3p11.2 that are associated with prostate cancer vulnerability at a genome-wide significance level in the Japanese population. These findings provide additional assistance for conducting Genome-wide association studies for prostate cancer in diverse populations to recognize risk loci for this genetically heterogeneous cancer. Further functional studies are required to understand the biological consequences of the newly recognized vulnerable loci in prostate cancer carcinogenesis [10].

For SNP rs5945572 A on *Xp11* the findings were the same for cases with younger age at onset (≤65) or aggressive phenotype and the entire group. On the other hand, there was significantly higher frequency of rs721048 A, among individuals diagnosed with aggressive prostate cancer than among those with less aggressive disease. Both the *Xp11* and the 2p15 variants are familiar and confer a moderate risk, resulting in an estimated 7% of individuals of European descent to prostate cancer risk. However, the cardinal causative biological perturbation associated with these variants remains to be illuminated. Certainly, subsequent studies have not been able to confirm numerous considerable gene regions that are recognized by association

analysis and the strategies of identifying candidate genes have not worked effectively. Regardless of this, the advances in genomic technologies and techniques of detection have enhanced the proficiency to study the cardinal genetics of prostate cancer risk. Still, no other biomarkers are used at the moment for the identification of prostate cancer. Regardless, numerous studies for finding alternative biomarkers are in process, such as *PCA3* (prostate cancer antigen 3), *TMPRSS2: ETS* (transmembrane protease, serine 2: transcription factor ETS-related gene), *GSTP1* (glutathione-S-transferase enzyme), and *C-reactive protein*, but they still need to be examined in clinical trials [50–52].

A risk factor is anything that increases a person's chance of developing cancer. Although risk factors often influence the chance to develop cancer, most of them do not directly or by themselves cause cancer. Some people with several known risk factors never develop cancer, while others with no known risk factors do [53]. In prostate cancer, it is normal to distinguish between epidemiologic endogenic factors (e.g., race, hormonal factors, genealogical information, age, etc.) and exogenic factors (e.g., sustenance, ecological factors, and lifestyle) which also include smoking, excess body fat accumulation.

6. Genetic and environmental risk factors

6.1 Age

One of the most important risk factors for prostate cancer is age. Between the ages of 55 and 74, a person's risk of acquiring prostate cancer is estimated to be. The occurrence and mortality rates of prostate cancer are significantly linked to age, with the highest prevalence found in older men over 65 years of age [54].

6.2 Race/ethnicity

African-Americans have the highest prostate cancer risk, followed by whites, Hispanics, American Indians/Alaska Natives, and Asian/Pacific Islanders. According to the American Cancer Society, African-Americans have a death rate that is double that of white men. The risk of prostate cancer in Scandinavian males may also be higher. Historically, the occurrence rate in East Asia (Japan and China) has been low. However, when Chinese and Japanese men immigrate to the U.S., they have an elevated risk of developing prostate cancer when compared with their native populations [55].

6.3 Diet

The high intake of calories and fats in western diets is undoubtedly one of the primary causes of prostate cancer. Saturated fat, which is often ingested from animal sources, may be linked to the development of prostate cancer, however the link is small. Red meat is a key component of animal-fat consumption, and some research suggests red meat consumption has risk ratios. The biological causes for this link are still a mystery. Alpha-linolenic acid, an 18-carbon fatty acid found in meat and some vegetable oils, is significant because it is a necessary precursor for the creation of prostaglandins and leukotrienes. More research is needed to see if dietary substitution of fatty acids implicated in the prostaglandin production pathways affects prostate cancer development [56].

There are more than 20 epidemiological studies that have evaluated the role of dairy food intake in prostate cancer. These studies are compatible with a positive association, unrelated to the contribution of dairy foods to total and saturated fat intake. Although β-carotene intake has not been related with prostate cancer risk, another crucial dietary carotenoid compound—lycopene—has become the priority of considerable notoriety. Lycopene, which is mostly absorbed through the ingestion of tomato-based foods, is the most important carotenoid in most Americans' diets. Consuming either a lycopene-rich diet or a β-carotene additive is thought to reduce the incidence of prostate cancer by about 40% [57]. Legumes, which include but are not limited to soy, have also been researched concerning to prostate cancer risk. So far, nothing has been corroborated by the findings. Green tea, another plant product that includes several polyphenolic chemicals with possible anti-carcinogenic qualities, is also being investigated.

6.4 Obesity

Obesity has been found to not affect the overall risk of prostate cancer, according to numerous studies. Obese men, on the other hand, are more likely to develop more aggressive forms of prostate cancer.

6.5 Family history

Numerous investigations, including familial prostate cancer gatherings, twin studies, and illness incidence in young individuals, support the importance of genetic variables. Prostate cancer is classified into two types: familial and sporadic. Familial/genealogical prostatic disease is a malignancy that occurs in affected members of one family at an early age (55 years old). This type of cancer has at least one first-degree relative with prostate cancer [48–53]. However, in sporadic prostate cancer, the genetic material is damaged over time due to external environmental sensitivity. A family can be affected by this type of cancer if three generations are affected, three first-degree relatives are affected, or two relatives are affected before the age of 55. Both forms of cancer have different rates of occurrence. Familial/genetic cancer has a rate of 15%, while sporadic carcinoma has a rate of 80 to 90%. A meta-analysis of several studies found that the cancer risk is higher for men who have a brother who has been diagnosed with cancer rather than a father. The most and that the genealogical risk is higher for early-onset disease [52].

7. Non-dietary environmental risk factors

7.1 Smoking

Smoke from burning tobacco is considered a carcinogen for many human malignancies, both submissive and non-submissive. Regardless, determining its causal link with prostate cancer has been a lengthy procedure. Because tobacco smoke carcinogens work explicitly, producing DNA mutations, and indirectly, causing hormone metabolism alterations, the causal link with prostate cancer is biologically reasonable. Some researchers have discovered a link between smoking and specific genetic variants that are linked to an increased risk of prostate cancer [54].

7.2 Socio-economic factors

All social elements, including economic, lifestyle, scholastic, and intellectual aspects, do not appear to influence the risk of prostate cancer on their own. Despite

this, they are indirectly involved by influencing dietary factors, occupational exposure, and access to health systems, both for timely detection and appropriate treatment, and they are unquestionably involved by influencing dietary factors, occupational exposure, and acquisition to health systems [12–15].

7.3 Other genetic changes

Other genes that may support a heightened risk of developing prostate cancer include *HPC1*, *HPC2*, *HPCX*, *CAPB*, *ATM*, *FANCA*, *HOXB13*, and mismatch repair genes. None of these, however, has been proven to cause prostate cancer or to be specific to this illness. Researchers are working to find genes linked to an increased risk of prostate cancer, and they are always attempting to learn more about how specific genetic changes can influence the spread of prostate cancer [23].

7.4 Occupational risk factors

Prostate cancer is thought to be caused by a variety of occupational variables. Farmers/agricultural workers, chemicals (organochlorine insecticides), shift work, and flight personnel are also risk factors [45].

7.5 Insecticides/pesticides

The most often used pesticide/insecticide was lead arsenate. Its use was reduced, but not stopped, until 1988, due to later discoveries of its harmful health consequences. In the United States, however, the usage of lead arsenate has poisoned most of the farmed land. Many workers are expected to develop cancer as a result of its carcinogenic qualities. Most frequently used insecticides are organochlorine pesticides, organophosphorus pesticides, carbamates, concerning to and triazines [51–56]. Ancestry is also important in the interplay between pesticide susceptibility and ancestry. When exposed to pesticides, a person with a family history of prostate cancer has a significantly increased risk of prostate cancer, which has been confirmed for carbamates, fonofos, chlorpyrifos, and phorate. Evidence is also presented that pesticide exposure causes cancer to become more aggressive.

7.6 Organochlorine pesticides

In Swedish and American research, rational results demonstrated a link between organochlorine pesticides and a significantly elevated incidence of prostate cancer. In a Swedish study, chlordane was shown to be the most dangerous organochlorine pesticide, whereas heptachlor, 2,4-dichlorophenoxyacetic acid (2,4-D), and lindane were found to be the most dangerous. Certain review articles support a bit that pesticides may elevate the prostate cancer risk minimally. The specific pesticide that is responsible for the increased risk is not yet discovered [57].

7.7 Shift work

Shift work is usually classified as work that is not done during the day. As a result, continual night shift work is classified as shift work. There is a strong link between shift work exposure and prostate cancer. Because of the various exposure conditions, the degree of variability is fairly significant. A meta-analysis conducted by Krstev et al. and prior meta-analyses provided evidence that shift employment, including night work, can raise the risk of prostate cancer [48].

7.8 Flight personnel

A great number of researchers have looked into the risk of cancer in-flight crew members. Because flight crews frequently travel across multiple time zones, they are sometimes thought to be representatives for circadian rhythm disturbance. They are, nonetheless, subjected to carcinogenic ionizing cosmic radiation [57]. They are, nonetheless, subjected to carcinogenic ionizing cosmic radiation. Cancer risk might be increased as a result of several exposures. According to a meta-analysis undertaken by Krstev et al., pilots had a considerably increased risk of prostate cancer. It is difficult to explain the variation between the pilots and cabin crew by circadian rhythm disorder or cosmic radiation and could be due to unrestricted distraction [58]. The evidence implies that there may be a link between pilots and prostate cancer.

7.9 Farming

A farmer's job entails a variety of responsibilities, including animal care, dealing with feed, seed, and animal waste, salvaging hay and other grains, driving tractors and other vehicles, operating various machines, and performing maintenance and repairs. As a result, they are exposed to organic and inorganic dust, pesticides, fungus, germs, viruses, lubricants, diesel exhaust, and welding fumes, and ultraviolet light [49–53]. Due to a variety of exposures, numerous studies have suggested an increased prostate cancer risk in farmers compared to other occupations.

7.10 Screening

PSA (prostate-specific antigen) is a substance produced by the prostate gland. High PSA readings can indicate prostate cancer, prostatitis (a noncancerous illness), or an enlarged prostate gland. The prostate-specific antigen (PSA) test is the most widely used prostate cancer screening method. This is a straightforward blood test that determines the amount of PSA in your circulation. This test is usually the first step in any prostate cancer diagnosis. Nonetheless, PSA screening alone cannot determine whether cancer is present. The PSA test is also used to monitor the effects of treatment for prostate cancer, such as surgery, radiation, hormone therapy, and chemotherapy [55]. When a man undergoes treatment for prostate cancer, his PSA level will plunge significantly. Conventional screening with PSA is one of the tools the physician will use to measure if cancer has recurred. Biochemical recurrence occurs when PSA levels rise to a certain threshold following prostate cancer treatment. This means that some cancer cells have managed to survive and are now releasing PSA. If this occurs, the doctor will schedule additional testing and suggest additional treatment options.

8. Prostate cancer treatment

8.1 Surgery for prostate cancer

A radical prostatectomy is the main type of surgery for prostate cancer. In this operation, the surgeon removes the complete prostate gland and additionally some of the tissue around it, including the seminal vesicles [56]. This includes:

- Open/radical prostatectomy
- Laparoscopic prostatectomy

In the more conventional approach to prostatectomy, called an open prostatectomy, the surgeon operates through a single long skin incision (cut) to eliminate the prostate and surrounding tissues whereas, in a laparoscopic prostatectomy, the surgeon makes numerous minor incisions and uses extraordinary long surgical tools to eliminate the prostate. The surgeon either holds the equipment straight away or uses a control panel to specifically move robotic arms that clasp the equipment [57]. This technique of prostatectomy has become more common in recent times. If done by experienced surgeons, the laparoscopic radical prostatectomy can provide outcomes analogous to the open approach.

8.1.1 Open prostatectomy

Open prostatectomy comprises Radical retropubic prostatectomy and Radical perineal prostatectomy.

8.1.1.1 Radical retropubic prostatectomy

For this open operation, the surgeon creates an incision in the lower abdomen, from the umbilicus down to the pubic bone. Either general anesthesia (asleep) or spinal or epidural anesthesia (drugging the lower half of the body) is given together with sedation during the surgery. If there is a plausible chance that cancer might have spread to surrounding lymph nodes (based on your PSA level, prostate biopsy results, and other conditions), the surgeon may also eliminate some of these lymph nodes at this time which is known as a pelvic lymph node dissection. The nodes are then dispatched to the laboratory to identify if they have cancer cells in them. If the cancer cells are found in any of the nodes, the surgeon might not proceed with the surgery [59]. This is because it is doubtful if cancer can be cured with surgery and eliminating the prostate could assist with severe side effects. After the prostate is eliminated, still under anesthesia, a catheter which is a thin, flexible tube will be put in the penis to help exude the bladder. The catheter will usually stay in place for 1 to 2 weeks until healed. After the catheter is removed, patients can urinate on their own [60].

8.1.1.2 Radical perineal prostatectomy

In this open operation, the surgeon makes an incision in the skin between the anus and scrotum (the perineum). This strategy is used less often because it is more plausible to lead to erection problems and because the surrounding lymph nodes cannot be eliminated. But it is usually a shorter operation and might be an alternative if erections are not concerned and there is no need for the lymph nodes to be removed. It also might be used if any other medical conditions make retropubic surgery dangerous. It can be just as remedial as the retro public approach if done accurately [57]. The perineal operation may cause less pain and an easier recuperation than the retropubic prostatectomy. After the surgery, still under anesthesia, a catheter will be put in the penis to assist exude the bladder. The catheter usually stays in place for 1 to 2 weeks until healed. After the catheter is removed, patients can urinate on their own.

8.1.2 Laparoscopic prostatectomy

If treatment with laparoscopic surgery is taken under consideration, it is crucial to comprehend what is familiar and what is not yet familiar about this approach. The most significant factors are likely to be the aptitude and experience of the

surgeon. If it is decided that laparoscopic surgery is the right treatment, be sure to look for a surgeon with a lot of experience. Laparoscopic prostatectomy comprises Laparoscopic radical prostatectomy and Robotic-assisted laparoscopic radical prostatectomy [59].

8.1.2.1 Laparoscopic radical prostatectomy

For a laparoscopic radical prostatectomy (LRP), the surgeon inserts special long equipment through numerous small incisions in the abdominal wall to eliminate the prostate. One of the pieces of equipment has a small video camera on the end, which lets the surgeon see the interior of the body. Laparoscopic prostatectomy has some advantages over open radical prostatectomy, encompassing limited blood loss and pain, shorter hospital stays (usually no more than a day), rapid recovery times, and the catheter will need to remain in the bladder for less time [60]. The rates of major side effects from LRP, such as erection difficulties and discomfort holding urine (incontinence) seem to be about as identical as for open prostatectomies. Recovery of bladder control may be hindered slightly with this approach. However, more long-term studies are required to compare side effects and risks of recurrence. Between open prostatectomy and laparoscopic radical prostatectomy, the success of either technique seems to be determined primarily by the experience and aptitude of the surgeon [55].

8.1.2.2 Robotic-assisted laparoscopic radical prostatectomy

This technique is also known as robotic prostatectomy. In this technique, laparoscopic surgery is done using a robotic system. The surgeon settles down at a control panel in the operating room and robotic arms are moved to operate through numerous small incisions in the patient's abdomen. Robotic prostatectomy has superiority over the open approach in terms of less pain, blood loss, and healing time. But with regards to the side effects, men are most concerned about urinary and/or erection difficulties [53]. There does not seem to be a discrepancy between robotic prostatectomy and other techniques. For the surgeon, the robotic system may provide more maneuverability and more accuracy when moving the equipment than a conventional laparoscopic radical prostatectomy. Regardless, the most important factor in the accomplishment of either type of laparoscopic surgery is the surgeon's experience and aptitude.

8.2 Risks of prostate surgery

The risks with any type of radical prostatectomy are much like those of any major surgery. Difficulties throughout or momentarily after the operation can encompass:

- Reactions to anesthesia
- Loss of blood from the surgery
- Blood clumps in the legs or lungs
- Injury to surrounding organs
- Infections at the site of surgery.

8.3 Side effects of prostate surgery

The major possible side effects of radical prostatectomy include incontinence of urine i.e., being unable to control urine and erectile dysfunction i.e., impotence; problems getting or keeping erections. These side effects can also occur with other forms of prostate cancer treatment [57–59].

Incontinence of urine: not being able to control urine or having leakage or dribbling. Being incontinent can influence not only physically but emotionally and socially as well. Some of the major types of incontinence include:

1. Men who lack self-restraint of stress might leak urine when they cough, laugh, sneeze, or exercise. Lacking self-restraint of stress is the most popular after prostate surgery. It's usually induced by trouble with the bladder sphincter valve. Prostate cancer treatments can damage this valve that keeps urine in the bladder or the nerves that keep the valve functioning.

2. Men who lack self-restraint of overflow have trouble emptying their bladder. They take a long time to urinate and have a dribbling stream with little force. This is usually inflicted by blockage or narrowing of the bladder opening by scar tissue.

3. Men who lack self-restraint of urge have an instant need to urinate. This happens when the bladder becomes too susceptible to stretching as it fills with urine.

4. Barely after surgery, men lost all proficiency to control their urine. This is called continuous incontinence.

9. Chemotherapy

Chemotherapy (chemo) utilizes anti-cancer drugs injected into a vein or given orally. These drugs travel through the bloodstream to reach cancer cells throughout the body. Chemo is periodically used if prostate cancer has spread outside the prostate gland and hormone therapy has been proven futile. Contemporary research has also indicated that chemo might be beneficial if provided along with hormone therapy. Chemo is still not a conventional treatment for timely prostate cancer [54].

9.1 Chemo drugs in treatment of prostate cancer

Chemo drugs are typically used one at a time for prostate cancer. Some of the chemo drugs used to treat prostate cancer comprise of:

- Docetaxel (Taxotere)
- Cabazitaxel (Jevtana)
- Mitoxantrone (Novantrone)
- Estramustine (Emcyt)

Extensively, the first chemo drug given is docetaxel, in combination with the steroid drug prednisone. If this does not function or stops functioning, cabazitaxel is usually the following chemo drug attempted (Although there may be additional treatment options as well).

Docetaxel and cabazitaxel have been indicated to help men live longer, on average than former chemo drugs. They may slow cancer's development and also decrease its symptoms, resulting in a better quality of life. However, chemo is very improbable to remedy prostate cancer. Other chemo drugs being researched for utilization in prostate cancer include carboplatin, oxaliplatin, and cisplatin [56–58].

Chemo drugs for prostate cancer are typically given intravenously (IV), as an infusion over a certain period and some drugs, such as estramustine, are given as a pill orally. Usually, a slightly larger and sturdier IV is required in the vein system to dispense chemo. They are known as central venous catheters (CVCs), central venous access devices (CVADs), or central lines. They are used to provide medicines, blood products, nutrients, or fluids right into the blood and can also be used to take out blood for testing. Many different kinds of CVCs are available. The most common types are the port and the PICC line [55].

Doctors usually give chemo in cycles, with each period of treatment followed by a relaxation period to give some time to recuperate from the impact of the drugs. Cycles are more frequently 2 or 3 weeks long. The schedule differs depending on the drugs used. For example, with certain drugs, the chemo is provided only on the first day of the cycle. With other drugs, it is provided for a few days in succession or once a week. Then, at the verge of the cycle, there are repetitions of the chemo scheduled to begin the next cycle. The duration of medication for progressive prostate cancer is based on how well it is functioning and what side effects it causes [56–58].

9.2 Apparent side effects of chemotherapy

Chemo drugs invade cells that are dividing instantly, which is why they function against cancer cells. But various cells in the body, such as those in the bone marrow (where new blood cells are made), the lining of the mouth and intestines, and the hair follicles, also divide rapidly. These cells can also be influenced by chemo, which can lead to side effects. The side effects of chemo are determined by the kind and dose of drugs provided and for how long they are consumed [59]. Various common side effects can include:

- Loss of hair
- Mouth sores
- Anorexia
- Nausea and vomiting
- Diarrhea
- Heightened chance of infections (from experiencing inadequate white blood cells)
- Manageable bruising or bleeding (from experiencing inadequate blood platelets)
- Exhaustion (from experiencing inadequate red blood cells)

These side effects are temporary and normally disappear once the treatment is finished. There are usually ways to reduce these side effects. For example, drugs can be provided to help intercept or decrease nausea and vomiting. Together with these risks, various side effects are observed more usual with particular chemo drugs. For example:

- Docetaxel and cabazitaxel occasionally cause drastic allergic reactions. Medicines are provided before each treatment to help preclude this. These drugs can also destruct nerves leading to a phenomenon known as peripheral neuropathy, which can cause numbness, tingling, or burning sensations in the hands or feet.

- Mitoxantrone can, very unusually, cause leukemia after several years.

- Estramustine conveys an intensified risk of blood clots.

If any side effects are noticed while getting chemo, report them to the cancer care team so that they can be treated quickly. In certain cases, the doses of the chemo drugs may need to be decreased or medication may need to be impeded or ceased to obstruct the effects from getting worse.

10. Chemoprevention

Chemoprevention, a prophylactic strategy that utilizes non-poisonous natural or artificial compounds to alter, impede, or avert cancer by targeting certain steps in the carcinogenic pathway, is gaining adhesion among health care practitioners. Soy isoflavones and curcumin, staples of the Asian diet, have demonstrated convincing results as practical factors for the chemoprevention of prostate cancer [60]. This is because of their proficiency to regulate various intracellular signaling pathways which comprise cellular proliferation, apoptosis, inflammation, and androgen receptor signaling. Contemporary information has disclosed that the DNA damage response (DDR) is one of the timeliest incidents in the multistep advancement of human skin cancers to aggressive malignancy. Soy isoflavones and curcumin stimulate the DDR, providing an alternative and justification for their clinical application in prostate cancer chemoprevention.

The prostate cancer risk can be ameliorated by approximately 30% by soy food consumption. Elevated soy consumption is correlated with localized prostate cancer reduction even among Japanese men with a significantly high soy intake concerning Caucasians. Moreover, a lower mortality rate from prostate cancer was correlated with soy consumption. The level of PSA decreases in prostate cancer patients simply because of the customary consumption of soy. Lately, it has been reported that the widespread presence of equal production, a potential risk factor for prostate cancer, is decreasing in the young generations of Korea and Japan [53–58].

Some of the important points to take into consideration are that chemoprevention must be safe, sustain the quality of life, reduce the occurrence, consequence, and harshness of the disease and be economically feasible. Of all the drugs that have demonstrated greater scientific evidence in clinical trials, worth mentioning are inhibitors of the enzyme 5-α reductase, which converts testosterone into dihydrotestosterone and two isoforms exist, type 1 and type 2 exist for this.

Dutasteride is an influential inhibitor of 5-α reductase and is 45 times stronger than finasteride in hindering type 1 isoform and twice as strong on isoform 2. Undoubtedly, this is a guaranteeing drug in the prevention of prostate cancer in

the risk population [59]. Among the scores of biomarkers being studied, numerous markers and techniques deserve awareness because of the promising published data indicating that better prostate cancer screening methods will be available in the coming times that will maintain non-aggressiveness and acceptability for both patients and urologists in clinical practice.

Conflict of interest

The authors declare a conflict of interest as none.

Author details

Surayya Siddiqui[1], Sridevi I. Puranik[2], Aimen Akbar[3] and Shridhar C. Ghagane[4,5]*

1 Department of Life Sciences, St. Ann's College for Women, Hyderabad, India

2 Department of Zoology, KLES B.K. Arts, Science and Commerce College, Chikodi, Chikodi, Karnataka, India

3 Department of Parasitology, McGill University, Montreal, Quebec, Canada

4 Division of Urologic-Oncology, Department of Urology, KLES Dr. Prabhakar Kore Hospital and Medical Research Centre, Belagavi, Karnataka, India

5 Department of Urology, JN Medical College, KLE Academy of Higher Education and Research (Deemed-to-be-University), JNMC Campus, Belagavi, Karnataka, India

*Address all correspondence to: shridhar.kleskf@gmail.com

IntechOpen

References

[1] Gudmundsson, J., Sulem, P., Rafnar, T., Bergthorsson, J., Manolescu, A., and Gudbjartsson, D. et al. (2008). Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nature Genetics, 40(3), 281-283. DOI:10.1038/ng.89

[2] Beebe-Dimmer, J., Hathcock, M., Yee, C., Okoth, L., Ewing, C., and Isaacs, W. et al. (2015). The HOXB13 G84E mutation is associated with an increased risk for prostate Cancer and other malignancies. Cancer Epidemiology Biomarkers and Prevention, 24(9), 1366-1372. DOI:10.1158/1055-9965.epi-15-0247

[3] li, Q., Liu, X., Hua, R., Wang, F., An, H., Zhang, W., and Zhu, J. (2015). Association of three 8q24 polymorphisms with prostate cancer susceptibility: Evidence from a meta-analysis with 50,854 subjects. Scientific Reports, 5(1). DOI:10.1038/srep12069

[4] Zhao, C., Liu, M., Xu, Y., Yang, K., Wei, D., and Shi, X. et al. (2014). 8q24 rs4242382 Polymorphism is a Risk Factor for Prostate Cancer among Multi-Ethnic Populations: Evidence from Clinical Detection in China and a Meta-analysis. Asian Pacific Journal of Cancer Prevention, 15(19), 8311-8317. DOI:10.7314/apjcp.2014.15.19.8311

[5] Van den Broeck, T., Joniau, S., Clinckemalie, L., Helsen, C., Prekovic, S., and Spans, L. et al. (2014). The role of single nucleotide polymorphisms in predicting prostate Cancer risk and therapeutic decision making. Biomed Research International, 2014, 1-16. DOI:10.1155/2014/627510

[6] Ferrís-i-Tortajada, J., Berbel-Tornero, O., Garcia-i-Castell, J., López-Andreu, J., Sobrino-Najul, E., and Ortega-García, J. (2011). Non-dietary environmental risk factors in prostate cancer. Actas Urológicas Españolas (English Edition), 35(5), 289-295. DOI:10.1016/j.acuroe.2010.12.001

[7] Levin, A., Machiela, M., Zuhlke, K., Ray, A., Cooney, K., and Douglas, J. (2008). Chromosome 17q12 variants contribute to risk of early-onset prostate Cancer. Cancer Research, 68(16), 6492-6495. DOI:10.1158/0008-5472.can-08-0348

[8] Chung, C., Ciampa, J., Yeager, M., Jacobs, K., Berndt, S., and Hayes, R. et al. (2011). Fine mapping of a region of chromosome 11q13 reveals multiple independent loci associated with the risk of prostate cancer. Human Molecular Genetics, 20(14), 2869-2878. DOI:10.1093/HMG/ddr189

[9] Chen, L., Fann, J., Chiu, S., Yen, A., Wahlfors, T., and Tammela, T. et al. (2014). Assessing interactions of two loci (rs4242382 and rs10486567) in familial prostate Cancer: Statistical evaluation of epistasis. Plos ONE, 9(2), e89508. DOI:10.1371/journal.pone.0089508

[10] Akamatsu, S., Takata, R., Haiman, C., Takahashi, A., Inoue, T., and Kubo, M. et al. (2012). Common variants at 11q12, 10q26, and 3p11.2 are associated with prostate cancer susceptibility in Japanese. Nature Genetics, 44(4), 426-429. DOI:10.1038/ng.1104

[11] Murabito, J., Rosenberg, C., Finger, D., Kreger, B., Levy, D., and Splansky, G. et al. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham heart study. BMC Medical Genetics, 8(Suppl 1), S6. DOI:10.1186/1471-2350-8-s1-s6

[12] Wang, L., Sato, K., Tsuchiya, N., Yu, J., Ohyama, C., and Satoh, S. et al. (2003). Polymorphisms in prostate-specific antigen (PSA) genes, risk of prostate cancer, and serum PSA levels in

the Japanese population. Cancer Letters, 202(1), 53-59. DOI:10.1016/j.canlet.2003.08.001

[13] Penney, K., Salinas, C., Pomerantz, M., Schumacher, F., Beckwith, C., and Lee, G. et al. (2009). Evaluation of 8q24 and 17q risk loci and prostate Cancer mortality. Clinical Cancer Research, 15(9), 3223-3230. DOI:10.1158/1078-0432.ccr-08-2733

[14] Liu, F., Hsing, A., Wang, X., Shao, Q., Qi, J., and Ye, Y. et al. (2011). Systematic confirmation study of reported prostate cancer risk-associated single nucleotide polymorphisms in Chinese men. Cancer Science, 102(10), 1916-1920. DOI:10.1111/j.1349-7006.2011.02036.x

[15] Rawla, P. (2019). Epidemiology of prostate Cancer. World Journal of Oncology, 10(2), 63-89. DOI:10.14740/wjon1191

[16] Dianat, S., Margreiter, M., Eckersberger, E., Finkelstein, J., Kuehas, F., and Herwig, R. et al. (2009). Gene polymorphisms and prostate cancer: The evidence. BJU International, 104(11), 1560-1572. DOI:10.1111/j.1464-410x.2009.08973.x

[17] Types of Prostate Cancer: Common, Rare, and More. (2021), from https://www.cancercenter.com/cancer-types/prostate-cancer/types

[18] Roberts, E., Cossigny, D., and Quan, G. (2013). The role of vascular endothelial growth factor in metastatic prostate Cancer to the skeleton. Prostate Cancer, 2013, 1-8. DOI:10.1155/2013/418340

[19] Genetics, H., and Research, G. (2021). What Is Genome-Wide Association Studies? MedlinePlus Genetics, from https://medlineplus.gov/genetics/understanding/genomicresearch/gwastudies/

[20] Krstev, S., and Knutsson, A. (2019). Occupational risk Factors for prostate Cancer: A meta-analysis. Journal Of Cancer Prevention, 24(2), 91-111. DOI:10.15430/jcp.2019.24.2.91

[21] Alvarez-Cubero, M., Saiz, M., Martinez-Gonzalez, L., Alvarez, J., Lorente, J., and Cozar, J. (2013). Genetic analysis of the principal genes related to prostate cancer: A review. Urologic Oncology: Seminars and Original Investigations, 31(8), 1419-1429. DOI:10.1016/j.urolonc.2012.07.011

[22] Tan, Y., Zeigler-Johnson, C., Mittal, R., Mandhani, A., Mital, B., Rebbeck, T., and Rennert, H. (2008). Common 8q24 sequence variations are associated with Asian Indian advanced prostate Cancer risk. Cancer Epidemiology Biomarkers and Prevention, 17(9), 2431-2435. DOI:10.1158/1055-9965.epi-07-2823

[23] Dean, M., and Lou, H. (2013). Genetics and genomics of prostate cancer. Asian Journal of Andrology, 15(3), 309-313. DOI:10.1038/aja.2013.29

[24] Fujita, K., and Nonomura, N. (2019). Role of androgen receptor in prostate Cancer: A review. The World Journal of Men's Health, 37(3), 288. DOI:10.5534/wjmh.180040

[25] Giatromanolaki, A., Fasoulaki, V., Kalamida, D., Mitrakas, A., Kakouratos, C., Lialiaris, T., and Koukourakis, M. (2019). CYP17A1 and androgen-receptor expression in prostate carcinoma tissues and Cancer cell lines. Current Urology, 13(3), 157-165. DOI:10.1159/000499276

[26] Horie, S. (2012). Chemoprevention of prostate Cancer: Soy Isoflavones and Curcumin. Korean Journal of Urology, 53(10), 665. DOI:10.4111/kju.2012.53.10.665

[27] Hsing, A., Sakoda, L., Chen, J., Chokkalingam, A., Sesterhenn, I., and Gao, Y. et al. (2007). MSR1 variants and

the risks of prostate cancer and benign prostatic hyperplasia: A population-based study in China. Carcinogenesis, 28(12), 2530-2536. DOI:10.1093/carcin/bgm196

[28] Wiklund, F., Jonsson, B., Brookes, A., Strömqvist, L., Adolfsson, J., and Emanuelsson, M. et al. (2004). Genetic analysis of the RNASEL gene in hereditary, familial, and sporadic prostate Cancer. Clinical Cancer Research, 10(21), 7150-7156. DOI:10.1158/1078-0432.ccr-04-0982

[29] BRATT, O. (2002). Hereditary prostate Cancer: Clinical aspects. The Journal of Urology, 906-913. DOI:10.1097/00005392-200209000-00004

[30] Cheng, I., Chen, G. K., Nakagawa, H., He, J., Wan, P., Laurie, C. C., Shen, J., Sheng, X., Pooler, L. C., Crenshaw, A. T., Mirel, D. B., Takahashi, A., Kubo, M., Nakamura, Y., Al Olama, A. A., Benlloch, S., Donovan, J. L., Guy, M., Hamdy, F. C., Kote-Jarai, Z., Haiman, C. A. (2012). Evaluating genetic risk for prostate cancer among Japanese and Latinos. Cancer epidemiology, biomarkers & prevention: A publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 21(11), 2048-2058. DOI:10.1158/1055-9965.

[31] Nyberg, T., Frost, D., Barrowdale, D., Evans, D. G., Bancroft, E., Adlard, J., Ahmed, M., Barwell, J., Brady, A. F., Brewer, C., Cook, J., Davidson, R., Donaldson, A., Eason, J., Gregory, H., Henderson, A., Izatt, L., Kennedy, M. J., Miller, C., Morrison, P. J., Antoniou, A. C. (2020). Prostate Cancer risks for male BRCA1 and BRCA2 mutation carriers: A prospective cohort study. European urology, 77(1), 24-35. DOI:10.1016/j.eururo.2019.08.025.

[32] Guérin, O., Fischel, J., Ferrero, J., Bozec, A., and Milano, G. (2010). EGFR targeting in hormone-refractory prostate Cancer: Current appraisal and prospects for treatment. Pharmaceuticals, 3(7), 2238-2247. DOI:10.3390/ph3072238.

[33] Genetics of Prostate Cancer (PDQ®)—Health Professional Version. (2021), from https://www.cancer.gov/types/prostate/hp/prostate-genetics-pdq#:~:text=Pathogenic%20variants%20in%20genes%2C%20such,and%20 screening%20for%20prostate%20cancer.

[34] Klein, R., Hallden, C., Gupta, A., Savage, C., Dahlin, A., and Bjartell, A. et al. (2012). Evaluation of multiple risk-associated single nucleotide polymorphisms versus prostate-specific antigen at baseline to predict prostate Cancer in unscreened men. European Urology, 61(3), 471-477. DOI:10.1016/j.eururo.2011.10.047

[35] Hsing, A. W., Chen, C., Chokkalingam, A. P., Gao, Y. T., Dightman, D. A., Nguyen, H. T., Deng, J., Cheng, J., Sesterhenn, I. A., Mostofi, F. K., Stanczyk, F. Z., and Reichardt, J. K. (2001). Polymorphic markers in the SRD5A2 gene and prostate cancer risk: A population-based case-control study. Cancer epidemiology, biomarkers & prevention: A publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 10(10), 1077-1082.

[36] Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209-249. DOI:10.3322/caac.21660

[37] Zhou, C., Wang, J., Cao, S., Shi, X., Zhang, Y., and Liu, M. et al. (2011). Association between single nucleotide polymorphisms on chromosome 17q and

the risk of prostate cancer in a Chinese population. Chinese Journal of Cancer, 30(10), 721-730. DOI:10.5732/cjc.011.10070

[38] Severi, G., Giles, G., Southey, M., Tesoriero, A., Tilley, W., and Neufing, P. et al. (2003). ELAC2/HPC2 polymorphisms, prostate-specific antigen levels, and prostate Cancer. JNCI Journal of The National Cancer Institute, 95(11), 818-824. DOI:10.1093/jnci/95.11.818

[39] Nicolaie, N., Cancel-Tassin, G., Azzouzi, A., Grand, B., Mangin, P., and Cormier, L. et al. (2009). Association between estrogen and androgen receptor genes and prostate cancer risk. European Journal of Endocrinology, 160(1), 101-106. DOI:10.1530/eje-08-0321

[40] Singh, S., Priyadarshini, A., Chakraborti, A., and Mandal, A. (2013). Asp299Gly and Thr399Ile polymorphism of TLR-4 gene in patients with prostate cancer from North India. Indian Journal of Urology, 29(1), 37. DOI:10.4103/0970-1591.109982

[41] Surgery for Prostate Cancer. (2021). Retrieved from https://www.cancer.org/cancer/prostate-cancer/treating/surgery.html

[42] Ingles, S., Ross, R., Yu, M., Haile, R., Irvine, R., La Pera, G., and Coetzee, G. (1997). Association of Prostate Cancer Risk with genetic polymorphisms in vitamin D receptor and androgen receptor. JNCI Journal of The National Cancer Institute, 89(2), 166-170. DOI:10.1093/jnci/89.2.166

[43] Grisanzio, C., Werner, L., Takeda, D., Awoyemi, B., Pomerantz, M., and Yamada, H. et al. (2012). Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis. Proceedings of the National Academy of Sciences, 109(28), 11252-11257. DOI:10.1073/pnas.1200853109

[44] Prokunina-Olsson, L., Fu, Y., Tang, W., Jacobs, K., Hayes, R., and Kraft, P. et al. (2010). Refining the Prostate Cancer Genetic Association within the JAZF1 Gene on Chromosome 7p15.2. Cancer Epidemiology Biomarkers and Prevention, 19(5), 1349-1355. DOI:10.1158/1055-9965.epi-09-1181

[45] HOXB13gene: MedlinePlus Genetics. (2021). Retrieved from https://medlineplus.gov/genetics/gene/hoxb13/

[46] Souiden, Y., Mahdouani, M., Chaieb, K., Elkamel, R., and Mahdouani, K. (2011). CYP17 gene polymorphism and prostate cancer susceptibility in a Tunisian population. Cancer Epidemiology, 35(5), 480-484. DOI:10.1016/j.canep.2010.11.008

[47] Cooney, K., Pilie, P., and Giri, V. (2016). HOXB13 and other high penetrant genes for prostate cancer. Asian Journal of Andrology, 18(4), 530. DOI:10.4103/1008-682x.175785

[48] Agalliu, I., Leanza, S., Smith, L., Trent, J., Carpten, J., Bailey-Wilson, J., and Burk, R. (2010). Contribution of HPC1 (RNASEL) and HPCX variants to prostate cancer in a founder population. The Prostate, 70(15), 1716-1727. DOI:10.1002/pros.21207

[49] Xu, B., Wang, J., Tong, N., Mi, Y., Min, Z., and Tao, J. et al. (2010). A functional polymorphism in MSMB gene promoter is associated with prostate cancer risk and serum MSMB expression. The Prostate, 70(10), 1146-1152. DOI:10.1002/pros.21149

[50] Wallis, C. J., and Nam, R. K. (2015). Prostate Cancer Genetics: A review. EJIFCC, 26(2), 79-91.

[51] Agalliu, I., Wang, Z., Wang, T., Dunn, A., Parikh, H., and Myers, T.

et al. (2013). Characterization of SNPs associated with prostate Cancer in men of Ashkenazi descent from the set of GWAS identified SNPs: Impact of Cancer family history and cumulative SNP risk prediction. Plos ONE, 8(4), e60083. DOI:10.1371/journal.pone.0060083

[52] Prostate Cancer Risk Factors. (2021), from https://www.hopkinsmedicine.org/health/conditions-and-diseases/prostate-cancer/prostate-cancer-riskfactors#:~:text=Environmental%20Exposures,are%20exposed%20to%20agricultural%20chemicals.

[53] Gann P. H. (2002). Risk factors for prostate cancer. Reviews in urology, 4 Suppl 5(Suppl 5), S3–S10.

[54] Lévesque, É., Laverdière, I., Lacombe, L., Caron, P., Rouleau, M., and Turcotte, V. et al. (2013). Importance of 5α-Reductase gene polymorphisms on circulating and Intraprostatic androgens in prostate Cancer. Clinical Cancer Research, 20(3), 576-584. DOI:10.1158/1078-0432.ccr-13-1100

[55] Prostate Cancer – Risk Factors and Prevention. (2020), from https://www.cancer.net/cancer-types/prostate-cancer/risk-factors-and-prevention#:~:text=Other%20genes%20that%20may%20carry,be%20specific%20to%20this%20disease.

[56] Castro, E., and Eeles, R. (2012). The role of BRCA1 and BRCA2 in prostate cancer. Asian Journal of Andrology, 14(3), 409-414. DOI:10.1038/aja.2011.150

[57] TOREN, P., and ZOUBEIDI, A. (2014). Targeting the PI3K/Akt pathway in prostate cancer: Challenges and opportunities (review). International Journal of Oncology, 45(5), 1793-1801. DOI:10.3892/ijo.2014.2601

[58] Prostate Cancer Detection: PSA Screening. (2021). from https://zerocancer.org/learn/about-prostate-cancer/detection-diagnosis/psa-test/

[59] Rivera-Pï¿½rez, J., Monter-Vera, M., Barrientos-Alvarado, C., Toscano-Garibay, J., Cuesta-Mejï¿½as, T., and Flores-Estrada, J. (2017). Evaluation of VEGF and PEDF in prostate cancer: A preliminary study in serum and biopsies. Oncology letters. DOI:10.3892/ol.2017.7374

[60] Makridakis, N., Akalu, A., and Reichardt, J. (2004). Identification and characterization of somatic steroid 5α-reductase (SRD5A2) mutations in human prostate cancer tissue. Oncogene, 23(44), 7399-7405. DOI:10.1038/sj.onc.1207922

Chapter 6

HER2^{Ile655Val} Polymorphism and Risk of Breast Cancer

Tung Nguyen-Thanh, Thong Ba Nguyen and Thuan Dang-Cong

Abstract

HER2 plays a vital role in the development and progression of several types of human cancer, so the HER2 becomes one of major targets for HER2-positive breast cancer treatment. Several reports have shown that the HER2 oncogene expression relates to clinicopathological factors in cancer patients. HER2^{Ile655Val} single nucleotide polymorphism associates with malignant tumors, including prostate cancer, colorectal cancer, osteosarcoma, gastric cancer, uterine cervical carcinoma, fibroadenoma, and breast cancer. To understand the precise association, this chapter was described to estimate the association between HER2^{Ile655Val} single nucleotide polymorphism and susceptibility to breast cancer. Our findings suggest that the Val allele in HER2 codon 655 single nucleotide polymorphism is strongly associated with the risk of breast cancer. HER2^{Ile655Val} single nucleotide polymorphism might also be a susceptibility factor that favors early-onset breast cancer.

Keywords: single nucleotide polymorphism (SNP), HER2^{Ile655Val}, rs1136201, breast cancer, early-onset, meta-analysis

1. Introduction

Breast cancer is the most common cancer among women, increasing incidence in most countries, representing a public health threat [1, 2]. Breast cancer is considered the leading cause of women's deaths worldwide [3]. More than two million were newly diagnosed with breast cancer in women worldwide in 2018 [3, 4]. There will be an estimated 18.1 million new cancer cases and 9.6 million cancer deaths in 2018. In the United States, breast cancer caused 42,000 deaths in 2017 [5]. There is a link with aging, especially among women aged 45 to 65, and it is increasing among younger women [6–8].

Human epidermal growth factor receptor 2 (HER2), also known as c-erbB2 and neu, is located on human chromosome 17q21 and is responsible for encoding a 185-kDa cross-membrane glycoprotein receptor. HER2 belongs to the ErbB family of growth factor receptors with intrinsic tyrosine kinase activity. The members of this family take the form of homodimer and heterodimer when activated via cell growth, specifically chemical and invasion [9, 10]. HER2 overexpression is seen in breast cancer, gastric cancer, and ovarian cancer [11]. HER2-targeted therapies have significantly enhanced HER2-positive breast cancer patients [12, 13]. Targets in downstream or resistant pathways of particular interest in HER2-positive breast

cancer include mTOR, PI3K, IGF-1R, Akt, HSP90, and VEGF that allow cell development, survival, and differentiation [14–17].

Several studies have independently discovered the association between $HER2^{Ile655Val}$ single nucleotide polymorphism and different benign and malignant tumors. The association between $HER2^{Ile655Val}$ SNP and the risk of breast cancer, especially early-onset breast cancer, has also been investigated; however, these results are inconclusive and controversial. Several articles have shown the association of $HER2^{Ile655Val}$ SNP with an increased risk of early-onset breast cancer in Chinese, Australian, and Taiwanese women [18–21]. Nonetheless, the association has not been observed in other studies [22–24]. In the present chapter, we aimed to obtain a more reliable estimate of the association between $HER2^{Ile655Val}$ SNP and the risk of breast cancer and susceptibility to early-onset breast cancer.

2. The role of HER2 in breast cancer

The proto-oncogene HER2/neu (C-erbB-2) has been localized to chromosome 17q21.1 and encodes a transmembrane tyrosine kinase growth factor receptor [25]. HER2 (human epidermal growth factor receptor 2) is a member of the epidermal growth factor receptor family, encodes a 185 kDa transmembrane glycoprotein with tyrosine kinase activity [26]. Most studies on HER2 found this gene was involved in inducing mammary carcinogenesis. HER2/neu gene amplification has been associated with the development of breast cancer [27].

HER2 is one of the biomarkers that play an essential role in breast cancer classification. Based on ER, PR, HER2, Ki-67 marker, breast cancer can be divided into five groups: Luminal A (ER+ or PR+; HER2-; Ki67 low), Luminal B HER2-negative (ER+ or PR+; HER2-; Ki67 high), Luminal B HER2-positive (ER+ or PR+; HER2+; Ki67 any), HER2-overexpression (ER-; PR-; HER2+; Ki67 any), Triple-negative (ER-; PR-; HER2-; Ki67 any) [28]. In addition, classification based on HER2 expression provides enhanced and essential therapeutic guidance [29]. Patients with subtype absence HER2 expression will have a poor prognosis and not receive the most benefit from chemotherapy [30].

HER2 is a human epidermal growth factor receptor (HER/EGFR/ERBB) family member. The basic structure of the epidermal growth factor receptor was described in **Figure 1**. In the extracellular domain, LD1 and LD2 are two repeated ligand-binding domains. CR1 and CR2 are two repeated cysteine-rich regions. TM indicates the short transmembrane spanning sequences. In the intracellular domain, TK is a catalytic tyrosine kinase, and CT is the carboxyl-terminal tail. Circled Ps are the phosphorylation sites within the TK and CT regions [31].

Schematic diagram of HER2 signaling pathways is shown in **Figure 1**. Upon ligand binding, dimerization between receptors of the EGFR family and HER2 receptor is induced. The homodimers or heterodimers after that, stimulate a serial of signaling cascades. Among various signaling pathways, the phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways are the two major and most studied pathways which take a pivotal role in tumor proliferation and anti-apoptosis. The whole signal transduction process can be divided into three sections: signal input (ligand-binding and dimerization), signal processing (a series of signaling cascades), and signal output (corresponding cellular processes) [31].

HER2 molecular pathways approach for HER2-targeted therapeutic strategies have the potential to adopt the HER2 overexpression tumor cells. Drugs targeting HER2 may include monoclonal antibodies that downregulate HER2 expression by binding to the extracellular domain (such as Trastuzumab, Pertuzumab ...), small

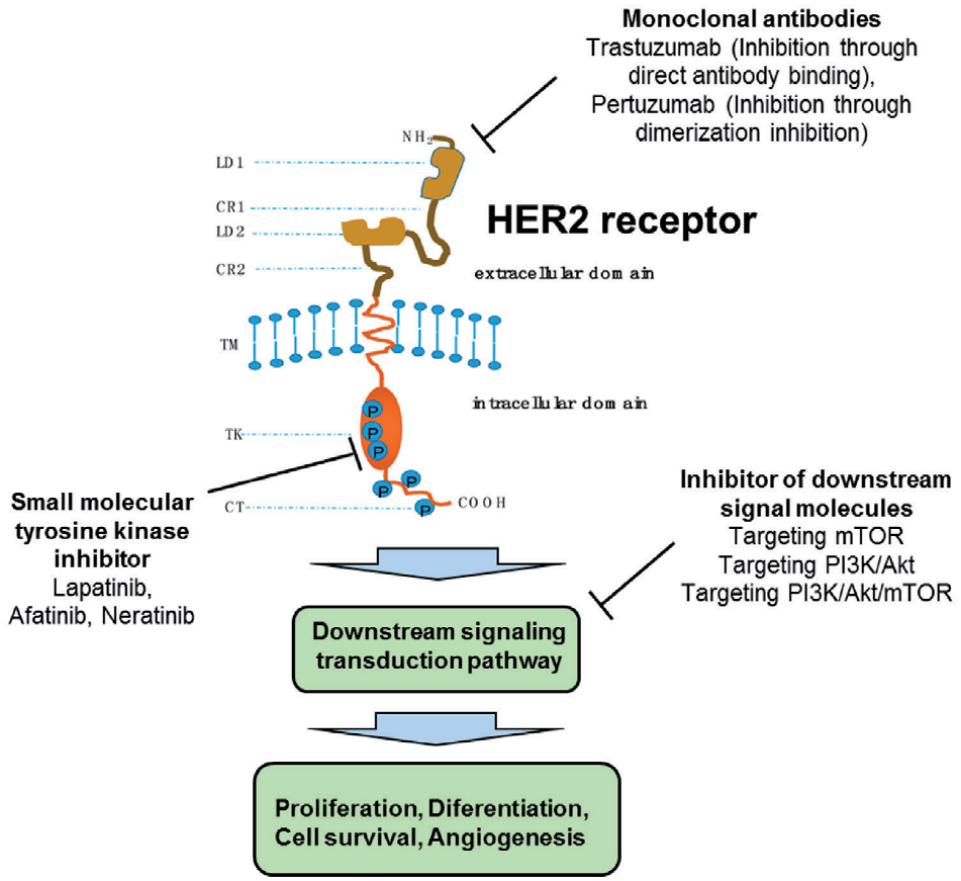


Figure 1.
HER2 molecular pathways approaches for HER2 positive breast cancer therapy (modified from Lv et al. 2016) [31].

molecular tyrosine kinase inhibitors (Lapatinib, Afatinib, Neratinib ...), which compete for ATP-binding to block HER2 signaling, antibody-drug conjugates (such as trastuzumab emtansine, called T-DM1), heat shock protein 90 inhibitors (hsp90) and inhibitors of downstream signal molecules (Targeting mTOR, PI3K/Akt ...) (**Figure 1**) [31, 32].

HER2 amplified or overexpressed in approximately 15–20% of breast cancer and associated with the aggressive clinical feature of absence therapy [33–35]. The IHC method of HER2 protein detection is advantageous, convenient, inexpensive, and only requires conventional microscopy. However, the results may be influenced by the time and the fixation protocol, the antibody clone, and it is challenging to apply the score sheet to have accurate conclusions. Therefore, for an ambiguous result obtained by immunohistochemical staining, the actual state of HER2 gene amplification should be assessed by performing fluorescent in situ hybridization (FISH) or dual chromogenic in-situ hybridization (DISH) because of its high accuracy and reliability, although expensive [36]. HER2 gene amplification and protein overexpression are important not only as a prognostic factor but also as a predictive clinical response to anti-HER2 therapeutic [37, 38]. Moreover, overall survival in patients treating with HER2-targeted therapy was higher than in patients without treatment [39].

The correlation between HER2 gene expression and breast cancer is not clear. According to most studies, it is thought that HER2 overexpression is a poor

prognostic factor [40]. The breast cancer parameters such as lymph node metastases, tumor grade, cell proliferation were reported that associated with the gene expression of HER2 [41, 42]. In the initial reports, HER2/neu amplification was a significant predictor of early relapse and death in breast cancer. HER2 gene amplification has a significant predictor of both disease-free survival and time to relapse in breast cancer patients [43, 44]. However, several authors demonstrated that amplification of the HER2 gene relate to ER, PR status but not correlate with age, tumor size, and lymph node [45].

3. HER2^{Ile655Val} single nucleotide polymorphism

Several studies have independently discovered the association between HER2^{Ile655Val} SNP and different types of benign and malignant tumors, including breast cancer, prostate cancer, colorectal cancer, osteosarcoma, gastric cancer, uterine cervical carcinoma, and fibroadenoma [46–53]. Riaz et al. conducted meta-analysis research and concluded that HER2^{Ile655Val} SNP is significantly correlated with six significant types of cancer (breast, ovarian, uterine, lung, thyroid, and gastric), suggesting that carriers of the Val allele and Val/Val genotype may be linked with an elevated risk of these cancers [54].

Single nucleotide polymorphisms residing in regulatory or functionally relevant gene regions may affect protein function [55]. HER2^{Ile655Val} SNP has been identified in the transmembrane domain-coding region of the HER2 gene at codon 655, encoding either isoleucine (Ile: ATC) or valine (Val: GTC) [53, 56]. Substitution of these two amino acids can alter the hydrophobicity of proteins, affecting the shape stability of the regions in the protein [57, 58]. Fleishman et al. found that substitution of Val for Ile in this position of the transmembrane region will destabilize the formation of active HER2 dimers, leading to reduced receptor activation and tyrosine kinase activity, even under conditions of HER2 overexpression [59]. The association between the HER2^{Ile655Val} SNP and the risk of breast cancer has been widely investigated in populations worldwide [18, 19, 23, 24, 53, 55, 60–92].

4. HER2^{Ile655Val} single nucleotide polymorphism contributes to breast cancer risk

Previous meta-analyses by Tao W et al. 2009 [93], Lu S et al. 2010 [94], Wang H et al. 2013 [95], Chen W et al. 2014 [47], and Krishna BM et al. 2018 [96] found that HER2^{Ile655Val} SNP is associated significantly with an increased risk of breast cancer, particularly in young women. However, a meta-analysis by Ma Y et al. 2011 revealed that HER2^{Ile655Val} SNP is not associated with breast cancer susceptibility [97].

Our recent meta-analyses further demonstrate the potential contribution of HER2^{Ile655Val} single nucleotide polymorphism to the oncogenesis of breast cancer [98]. The meta-analysis showed that the HER2 codon 655 Val allele was significantly associated with an increased risk of breast cancer in an allele genetic model (additive model, RR 1.21, 95% CI 1.07–1.36; I^2 = 61.0%; n = 16). There was a 21% significant increase in the risk of breast cancer in subjects who were Val carriers (Ile/Val and Val/Val) (dominant model, RR 1.21, 95% CI 1.06–1.38; I^2 = 58.0%; n = 16). The recessive model HER2 codon 655 was not associated with the risk of breast cancer (RR 1.26, 95% CI 0.99–1.60; I^2 = 23.6%; n = 15). There was publication bias in the studies (Begg's funnel plot was symmetric; additive model, Egger's test t = 5.44, P for bias = 0.000, n = 16; dominant model, Egger's test t = 4.92, P for bias = 0.000, n = 16; recessive model, Egger's test t = 4.35, P for bias = 0.001, n = 15) (**Figure 2**).

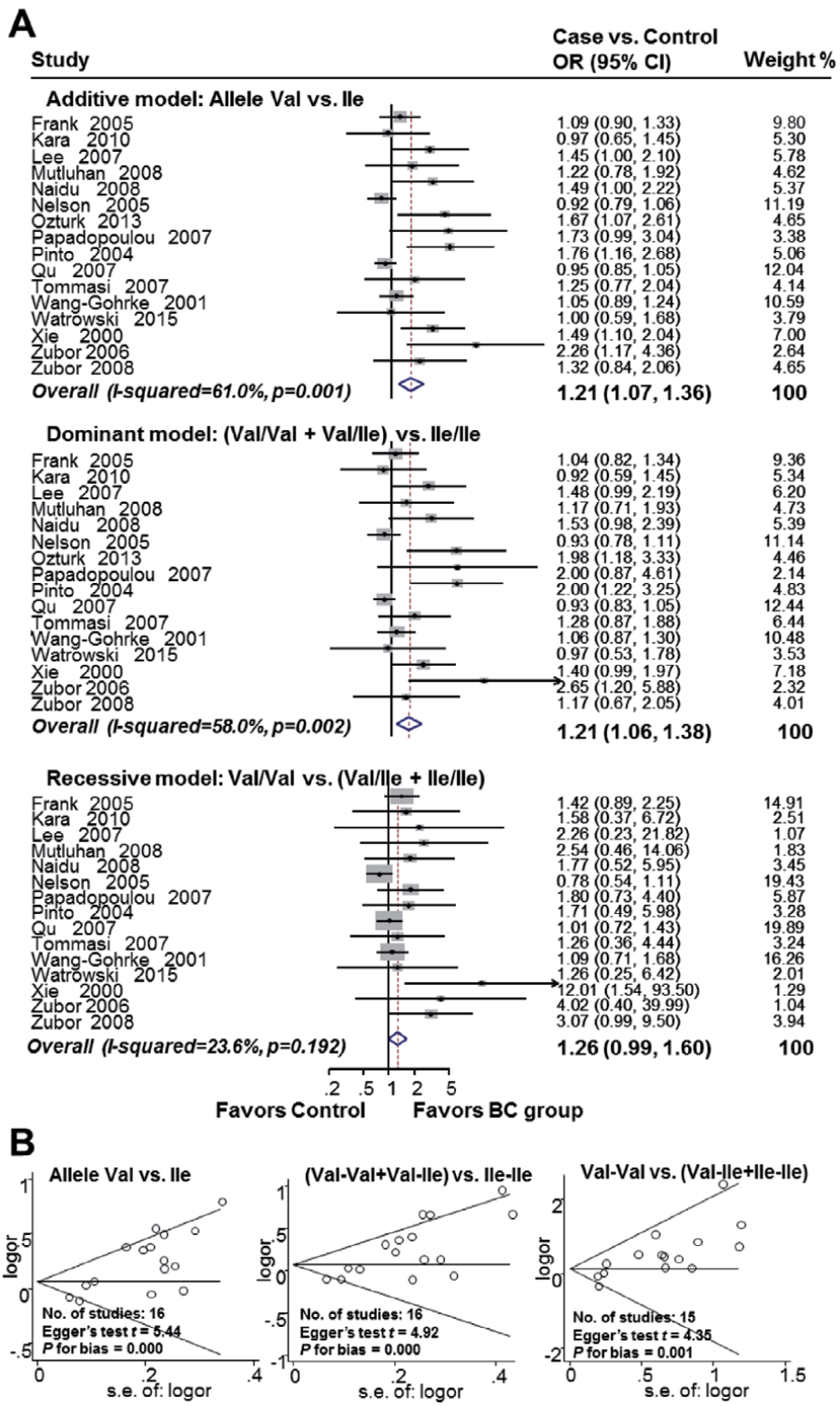


Figure 2.
The association between HER2^{Ile655Val} *SNP and the risk of breast cancer in worldwide populations (modified from Nguyen Thanh et al. 2021) [98]. A. Forest plot for the association between HER2*^{Ile655Val} *SNP and breast cancer risk; B. Funnel plot evaluating publication bias among studies included in the meta-analysis.*

The molecular mechanism of HER2^{Ile655Val} SNP, a non-polar to non-polar amino acid mutation, has been investigated in previous studies. Using computational exploration, Fleishman et al. 2002 proposed that the transmembrane region of the HER2 homodimer can exist in two stable conformations, either in an active or inactive form. The dimer mediated by the C-terminal dimerization motif is more durable than the dimer formed by the N-terminal motif. The authors found that substitution of Val for Ile in this position of the transmembrane region will destabilize the formation of active HER2 dimers mediated by the N-terminal dimerization motif and lead to reduced receptor activation and tyrosine kinase activity. However, the presence of the Val allele could reinforce the stabilization of the receptor's active state, which results in augmentation of autophosphorylation, hyper-active tyrosine kinase, and cellular proliferation [59]. Bocharov and colleagues researched the spatial structure of the dimeric transmembrane domain of the HER2 protein. They found that the Ile655Val variant can excessively stabilize the ErbB2 active dimeric state due to substituting the bulk side chain of Ile with the smaller Val, thus allowing tighter TM helix packing [99]. In another experiment, Tanaka et al. assessed the role of amino acid substitutions in the conformational stability of human lysozyme protein via thermodynamic analysis at high temperature and very low pH. They showed that in constructed isoleucine to valine mutants, the strength of mutant proteins was reduced compared to that of the wild-type protein [100].

5. Association between HER2^{Ile655Val} single nucleotide polymorphism and early-onset breast cancer susceptibility

Prior studies independently found a correlation between the high presence of the Val allele in the codon 655 of the HER2 gene and the onset of breast cancer [101–104]. Additionally, Millikan et al. 2003 and Tommasi et al. 2007 reported a strong association between the variant allele 655Val and breast cancer in younger women when combined with family history [105, 106]. However, other published data reveal the opposite results, with no significant correlation between the HER2^{Ile655Val} genotype and the risk of early-onset breast cancer in patients 40 to 50 years [63, 90, 107–109].

We conducted the meta-analysis, which collected 17 age-stratified articles with a cut-off age value from 40 to 55 years old (46.53 ± 3.84). We found that in young women, HER2 codon 655 polymorphism was strongly and significantly associated with breast cancer in all genetic models (additive, dominant, and recessive), which was in contrast to the results from a subgroup of older women (Data not shown).

Subgroup meta-analysis in the breast cancer population was utilized to investigate the association between HER2^{Ile655Val} SNP and the age at onset of breast cancer, calculating adjusted RRs comparing older and younger participants in the breast cancer population (**Figure 3**). There was a significant 17% increase in the risk of early onset of breast cancer in patients who were Val carriers (Ile/Val and Val/Val) (dominant model, RR 0.83, 95% CI 0.72 to 0.97; I^2 = 36.5%; n = 14). Meanwhile, no significant association of HER2^{Ile655Val} SNP polymorphism with the age of onset was found in the subgroup of breast cancer women under an additive model (additive, RR 0.87, 95% CI 0.75–1.01, I^2 = 39.8%; n = 10) and a recessive model (RR 0.96, 95% CI 0.73 to 1.26, I^2 = 0.0%; n = 9). There was no publication bias in the studies (Begg's funnel plot was symmetric; additive model, Egger's test t = −2.02, P for bias = 0.078, n = 10; dominant model, Egger's test t = −1.91, P for bias = 0.08, n = 14; recessive model, Egger's test t = −2.27, P for bias = 0.057, n = 9). Furthermore, subgroup meta-analysis of the control population indicated no significant association between HER2^{Ile655Val} SNP and the age of participants was

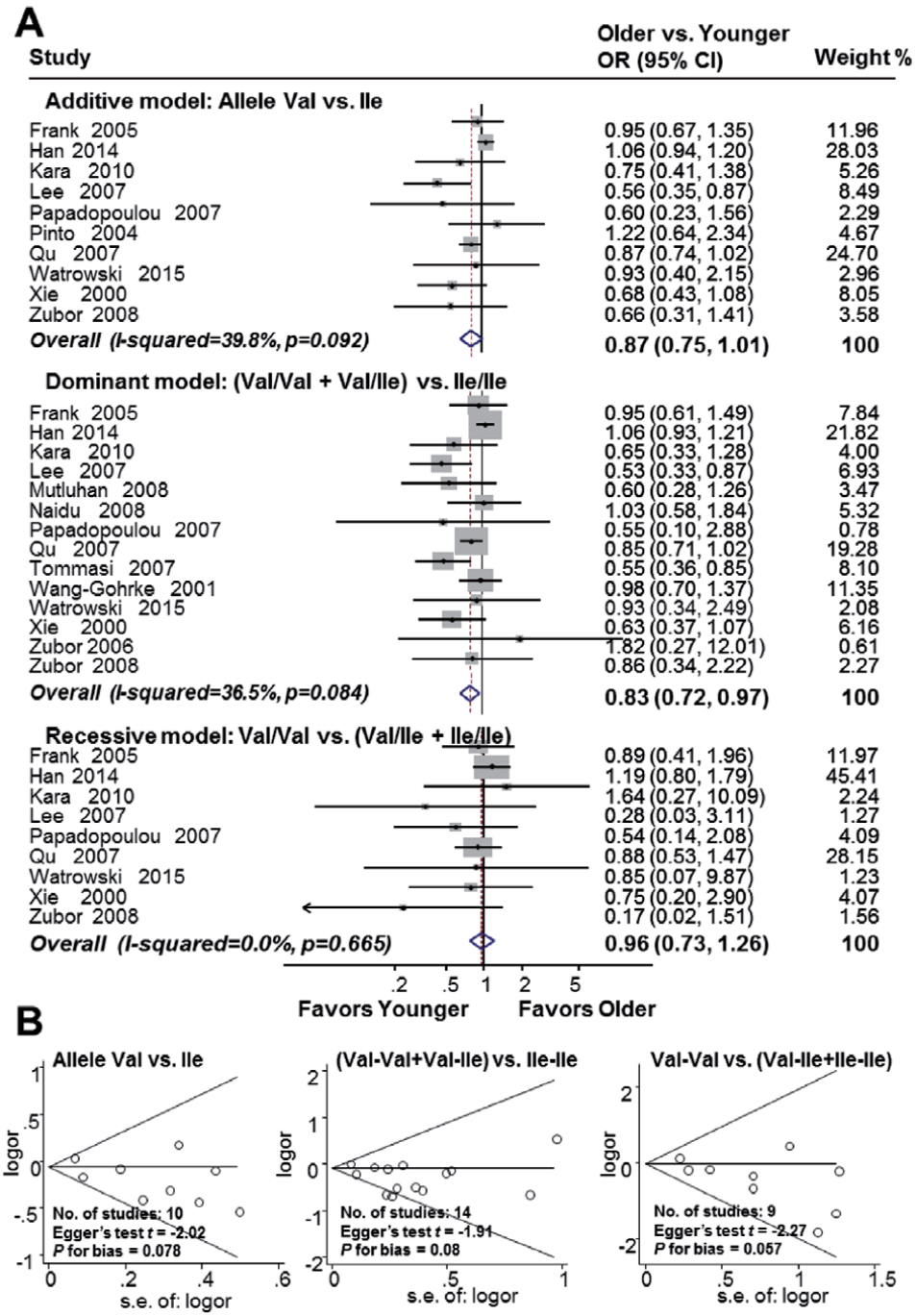


Figure 3.
HER2^{Ile655Val} *SNP is associated with an increased risk of early-onset breast cancer (modified from Nguyen Thanh et al. 2021) [98]. (A) Forest plot for the association between HER2*^{Ile655Val} *SNP and an increased risk of early-onset breast cancer; (B) Funnel plot evaluating for publication bias among studies included in the meta-analysis.*

found in control women under all genetic models (additive, RR 0.94, 95% CI 0.75 to 1.18; I^2 = 28.1%; n = 6; dominant, RR 0.94, 95% CI 0.83 to 1.07; I^2 = 0.0%; n = 9, and recessive, RR 1.19, 95% CI 0.70 to 2.02; I^2 = 2.0%; n = 3) (Data not shown). In the breast cancer population, the dominant model of HER2^{Ile655Val} SNP was shown

to be significantly associated with younger age. Therefore, the high presence of the Val allele in codon 655 of the HER2 gene might explain the increasing frequency of younger age onset of breast cancer.

Molecular mechanisms of the early onset state have been studied in certain types of cancer. Several candidate genes and signaling pathways have been found in the early onset of colorectal cancer. REG1A, CK20, and MAP3K8 gene expression were related to early-onset colorectal cancer formation [110]. Using PPI network analysis, Zhao and colleagues suggested that early-onset colorectal cancer is associated with vascular smooth muscle contraction signaling pathways. They also identified seven hub genes, namely, ACTA2, ACTG2, MYH11, CALD1, MYL9, TPM2, and LMOD1, along with this signaling pathway [111]. Recently, using weighted gene co-expression network analysis and other analysis methods, Mo et al. identified seven genes (SPARC, DCN, FBN1, WWTR1, TAGLN, DDX28, and CSDC2) associated with the development and prognosis of early-onset colorectal cancer. These genes may serve as novel biomarkers for diagnosing early-onset colorectal cancer [112]. In prostate cancer, a study conducted by Weischenfeldt et al. found the genomic alteration landscapes of early-onset prostate cancer compared to older-onset cancer. They discovered that early-onset prostate cancer possesses a higher frequency of balanced structural rearrangements, with a specific abundance of androgen-regulated ETS gene fusions, and concluded that ETS fusion genes are signs of early-onset prostate cancer [113]. Furthermore, Gerhauser et al. demonstrated the role of androgen receptor-driven rearrangements, an early APOBEC-driven mutational mechanism, and ESRP1 gene duplication that contributed to the pathogenesis seen in early-onset prostate cancer [114]. Nevertheless, the molecular mechanism involved in the pathogenesis of early-onset breast cancer is, to date, poorly understood and needs to be studied further.

6. Conclusion

HER2 was involved in the development and progression of mammary carcinogenesis, including breast cancer. HER2-targeted therapies have significantly enhanced the clinical outcome for HER2-positive breast cancer patients. $HER2^{Ile655Val}$ single nucleotide polymorphism associate with an increased risk of breast cancer. In addition, $HER2^{Ile655Val}$ SNP might be considered as a susceptibility factor for early-onset breast cancer. Further molecular studies are required to reveal the mechanism of this correlation.

Author details

Tung Nguyen-Thanh[1,2*], Thong Ba Nguyen[3] and Thuan Dang-Cong[4]

1 Faculty of Basic Science, Hue University of Medicine and Pharmacy, Hue University, Hue, Vietnam

2 Institute of Biomedicine, Hue University of Medicine and Pharmacy, Hue University, Hue, Vietnam

3 Department of Anatomy, Biochemistry, and Physiology, John A. Burns School of Medicine, The University of Hawaii at Manoa, Honolulu, Hawaii, USA

4 Department of Histology, Embryology, Pathology and Forensic, Hue University of Medicine and Pharmacy, Hue University, Hue, Vietnam

*Address all correspondence to: nguyenthanhtung@hueuni.edu.vn; nttung@huemed-univ.edu.vn

IntechOpen

References

[1] Zahmatkesh B, Keramat A, Alavi N, Khosravi A, Kousha A, Motlagh AG, Darman M, Partovipour E, Chaman R: Breast cancer trend in Iran from 2000 to 2009 and prediction till 2020 using a trend analysis method. Asian Pacific Journal of Cancer Prevention 2016, 17(3):1493-1498.

[2] Momenimovahed Z, Salehiniya H: Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer: Targets and Therapy 2019, 11:151.

[3] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 2018, 68(6):394-424.

[4] Jenkins C, Minh LN, Anh TT, Ngan TT, Tuan NT, Giang KB, Hoat LN, Lohfeld L, Donnelly M, Van Minh H *et al*: Breast cancer services in Vietnam: a scoping review. Global health action 2018, 11(1):1435344.

[5] Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. CA: A Cancer Journal for Clinicians 2020, 70(1):7-30.

[6] Dobi Á, Kelemen G, Kaizer L, Weiczner R, Thurzó L, Kahán Z: Breast cancer under 40 years of age: increasing number and worse prognosis. Pathology & Oncology Research 2011, 17(2):425-428.

[7] Zubor P, Vojvodova A, Danko J, Kajo K, Szunyogh N, Lasabova Z, Biringer K, Visnovsky J, Dokus K, Galajda P: HER-2 [Ile655Val] polymorphism in association with breast cancer risk: a population-based case-control study in Slovakia. Neoplasma 2006, 53(1):49.

[8] Bouchardy C, Fioretta G, Verkooijen H, Vlastos G, Schaefer P, Delaloye J, Neyroud-Caspar I, Majno SB, Wespi Y, Forni M: Recent increase of breast cancer incidence among women under the age of forty. British journal of cancer 2007, 96(11):1743-1746.

[9] Yarden Y, Sliwkowski MX: Untangling the ErbB signalling network. Nature reviews Molecular cell biology 2001, 2(2):127-137.

[10] Friedlander E, Barok M, Szollosi J, Vereb G: ErbB-directed immunotherapy: Antibodies in current practice and promising new agents (vol 116, pg 126, 2008). Immunology Letters 2009, 124(1):55-56.

[11] Tai W, Mahato R, Cheng K: The role of HER2 in cancer therapy and targeted drug delivery. Journal of controlled release 2010, 146(3):264-275.

[12] Incorvati JA, Shah S, Mu Y, Lu J: Targeted therapy for HER2 positive breast cancer. Journal of hematology & oncology 2013, 6(1):38.

[13] Sidaway P: HER2-targeted agents overcome resistance. Nature reviews Clinical oncology 2020, 17(3):133.

[14] Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L: Treatment of HER2-positive breast cancer: current status and future perspectives. Nature reviews Clinical oncology 2012, 9(1):16-32.

[15] Franklin MC, Carey KD, Vajdos FF, Leahy DJ, De Vos AM, Sliwkowski MX: Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. Cancer cell 2004, 5(4):317-328.

[16] Agus DB, Gordon MS, Taylor C, Natale RB, Karlan B, Mendelson DS, Press MF, Allison DE, Sliwkowski MX, Lieberman G: Phase I clinical study of pertuzumab, a novel HER dimerization inhibitor, in patients with advanced

cancer. Journal of clinical oncology 2005, 23(11):2534-2543.

[17] Nahta R, Yu D, Hung M-C, Hortobagyi GN, Esteva FJ: Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. Nature clinical practice Oncology 2006, 3(5):269-280.

[18] Xie D, Shu XO, Deng Z, Wen WQ, Creek KE, Dai Q, Gao YT, Jin F, Zheng W: Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. Journal of the National Cancer Institute 2000, 92(5):412-417.

[19] Montgomery KG, Gertig DM, Baxter SW, Milne RL, Dite GS, McCredie MR, Giles GG, Southey MC, Hopper JL, Campbell IG: The HER2 I655V polymorphism and risk of breast cancer in women < age 40 years. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2003, 12(10):1109-1111.

[20] Lee S-C, Hou M-F, Hsieh P-C, Wu S-H, Hou LA, Ma H, Tsai S-M, Tsai L-Y: A case–control study of the HER2 Ile655Val polymorphism and risk of breast cancer in Taiwan. Clinical biochemistry 2008, 41(3):121-125.

[21] Naidu R, Yip C, Taib NA: Polymorphisms of HER2 Ile655Val and cyclin D1 (CCND1) G870A are not associated with breast cancer risk but polymorphic allele of HER2 is associated with nodal metastases. Neoplasma 2008, 55(2):87-95.

[22] Kara N, Karakus N, Ulusoy AN, Ozaslan C, Gungor B, Bagci H: P53 codon 72 and HER2 codon 655 polymorphisms in Turkish breast cancer patients. DNA and cell biology 2010, 29(7):387-392.

[23] An HJ, Kim NK, Oh D, Kim SH, Park MJ, Jung MY, Kang H, Kim SG, Lee KP, Lee KS: Her2 genotype and breast cancer progression in Korean women. Pathology international 2005, 55(2):48-52.

[24] Zubor P, Vojvodova A, Danko J, Kajo K, Szunyogh N, Lasabova Z, Biringer K, Visnovsky J, Dokus K, Galajda P *et al*: HER-2 [Ile655Val] polymorphism in association with breast cancer risk: a population-based case-control study in Slovakia. Neoplasma 2006, 53(1):49-55.

[25] Muleris M, Almeida A, Malfoy B, Dutrillaux B. Assignment of v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (ERBB2) to human chromosome band 17q21. 1 by in situ hybridization. Cytogenetic and Genome Research 1997, 76(1/2):34.

[26] Saini KS, Azim Jr HA, Metzger-Filho O, Loi S, Sotiriou C, de Azambuja E, Piccart M: Beyond trastuzumab: new treatment options for HER2-positive breast cancer. The Breast 2011, 20:S20-S27.

[27] Piechocki MP, Yoo GH, Dibbley SK, Lonardo F: Breast cancer expressing the activated HER2/neu is sensitive to gefitinib in vitro and in vivo and acquires resistance through a novel point mutation in the HER2/neu. Cancer research 2007, 67(14):6825-6843.

[28] Onitilo AA, Engel JM, Greenlee RT, Mukesh BN: Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clinical medicine & research 2009, 7(1-2):4-13.

[29] Iqbal N, Iqbal N: Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications. Molecular biology international 2014, 2014.

[30] Knutson KL, Clynes R, Shreeder B, Yeramian P, Kemp KP, Ballman K, Tenner KS, Erskine CL, Norton N,

Northfelt D: Improved survival of HER2+ breast cancer patients treated with trastuzumab and chemotherapy is associated with host antibody immunity against the HER2 intracellular domain. Cancer research 2016, 76(13):3702-3710.

[31] Lv Q, Meng Z, Yu Y, Jiang F, Guan D, Liang C, Zhou J, Lu A, Zhang G: Molecular Mechanisms and Translational Therapies for Human Epidermal Receptor 2 Positive Breast Cancer. International journal of molecular sciences 2016, 17(12):49.

[32] Di Modica M, Tagliabue E, Triulzi T: Predicting the Efficacy of HER2-Targeted Therapies: A Look at the Host. Disease markers 2017, 2017:7849108.

[33] Krishnamurti U, Silverman JF: HER2 in breast cancer: a review and update. Advances in anatomic pathology 2014, 21(2):100-107.

[34] Figueroa-Magalhaes MC, Jelovac D, Connolly R, Wolff AC: Treatment of HER2-positive breast cancer. Breast (Edinburgh, Scotland) 2014, 23(2):128-136.

[35] Varga Z, Noske A, Ramach C, Padberg B, Moch H: Assessment of HER2 status in breast cancer: overall positivity rate and accuracy by fluorescence in situ hybridization and immunohistochemistry in a single institution over 12 years: a quality control study. BMC cancer 2013, 13(1):615.

[36] Ellis C, Dyson M, Stephenson T, Maltby E: HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques. Journal of clinical pathology 2005, 58(7):710-714.

[37] Parkes E, McKenna S, McAleer J, Clarke J, Clayton A, James C: HER2 as a prognostic factor in metastatic breast cancer treated with taxanes. Journal of Clinical Oncology 2010, 28(15_supply):1150-1150.

[38] Schneeweiss A, Chia S, Hegg R, Tausch C, Deb R, Ratnayake J, McNally V, Ross G, Kiermaier A, Cortés J: Evaluating the predictive value of biomarkers for efficacy outcomes in response to pertuzumab-and trastuzumab-based therapy: an exploratory analysis of the TRYPHAENA study. Breast Cancer Research 2014, 16(4):R73.

[39] Mendes D, Alves C, Afonso N, Cardoso F, Passos-Coelho JL, Costa L, Andrade S, Batel-Marques F: The benefit of HER2-targeted therapies on overall survival of patients with metastatic HER2-positive breast cancer–a systematic review. Breast Cancer Research 2015, 17(1):140.

[40] Allred DC, Clark GM, Molina R, Tandon AK, Schnitt SJ, Gilchrist KW, Osborne CK, Tormey DC, McGuire WL: Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. Human pathology 1992, 23(9):974-979.

[41] Pengfei S, Cheng C, Yufeng Y: Correlation Between HER-2 Gene Amplification or Protein Expression and Clinical Pathological Features of Breast Cancer. Cancer Biotherapy and Radiopharmaceuticals 2019, 34(1):42-46.

[42] Aman NA, Doukoure B, Koffi KD, Koui BS, Traore ZC, Kouyate M, Effi AB: HER2 overexpression and correlation with other significant clinicopathologic parameters in Ivorian breast cancer women. BMC clinical pathology 2019, 19:1.

[43] Iqbal N, Iqbal N: Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and

Therapeutic Implications. Molecular biology international 2014, 2014:852748.

[44] Parkes E, McKenna SM, McAleer JJ, Clarke J, Clayton AJ, James CR: HER2 as a prognostic factor in metastatic breast cancer treated with taxanes. Journal of Clinical Oncology 2010, 28(15_supply): 1150-1150.

[45] Shokouh TZ, Ezatollah A, Barand P: Interrelationships Between Ki67, HER2/neu, p53, ER, and PR Status and Their Associations With Tumor Grade and Lymph Node Involvement in Breast Carcinoma Subtypes: Retrospective-Observational Analytical Study. Medicine 2015, 94(32):e1359-e1359.

[46] Yokomizo A, Koga H, Kinukawa N, Tsukamoto T, Hirao Y, Akaza H, Mori M, Naito S: Association of HER-2 polymorphism with Japanese sporadic prostate cancer susceptibility. The Prostate 2005, 62(1):49-53.

[47] Chen W, Yang H, Tang WR, Feng SJ, Wei YL: Updated meta-analysis on HER2 polymorphisms and risk of breast cancer: evidence from 32 studies. Asian Pacific journal of cancer prevention : APJCP 2014, 15(22):9643-9647.

[48] Liang X, Zhang Y-j, Liu B, Ni Q, Jin M-j, Ma X-y, Yao K-y, Li Q-l: Association between HRE-2 gene polymorphism at codon 655 and genetic susceptibility of colorectal cancer. Chinese journal of medical genetics 2009, 26:302-305.

[49] Xin DJ, Shen GD, Song J: Single nucleotide polymorphisms of HER2 related to osteosarcoma susceptibility. Int J Clin Exp Pathol 2015, 8(8):9494-9499.

[50] Kuraoka K, Matsumura S, Hamai Y, Nakachi K, Imai K, Matsusaki K, Oue N, Ito R, Nakayama H, Yasui W: A single nucleotide polymorphism in the transmembrane domain coding region of HER-2 is associated with development and malignant phenotype of gastric cancer. Int J Cancer 2003, 107(4):593-596.

[51] Kruszyna L, Lianeri M, Roszak A, Jagodzinski PP: HER2 codon 655 polymorphism is associated with advanced uterine cervical carcinoma. Clin Biochem 2010, 43(6):545-548.

[52] Zubor, Karol Kajob AS, Norbert Szunyogha, Silvester Galoa, Carlos A. Dussanad, Gabriel Minarikc, Jozef Visnovskya and Jan Danko: Human-epithelial-growth-factor-receptor-2Ile655Val-polymorphism-and-risk-of-breast-fibroadenoma European Journal of Cancer Prevention 2008, 17(1):33-38.

[53] Ozturk O, Canbay E, Kahraman OT, Fatih Seyhan M, Aydogan F, Celik V, Uras C: HER2 Ile655Val and PTEN IVS4 polymorphisms in patients with breast cancer. Molecular biology reports 2013, 40(2):1813-1818.

[54] Riaz SK, Rashid MM, Kayani MA, Malik MFA: Role of HER-2 Ile655Val Polymorphism as Universal Cancer Susceptibility Marker among Different Cancers. Archives of Iranian medicine 2016, 19(6):430 -438.

[55] Frank B, Hemminki K, Wirtenberger M, Bermejo JL, Bugert P, Klaes R, Schmutzler RK, Wappenschmidt B, Bartram CR, Burwinkel B: The rare ERBB2 variant Ile654Val is associated with an increased familial breast cancer risk. Carcinogenesis 2005, 26(3):643-647.

[56] Krishna BM, Chaudhary S, Panda AK, Mishra DR, Mishra SK: Her2 Ile 655 Val polymorphism and its association with breast cancer risk: an updated meta-analysis of case-control studies. Scientific reports 2018, 8(1):1-19.

[57] Ameyaw M-M, Tayeb M, Thornton N, Folayan G, Tariq M, Mobarek A, Evans DP, Ofori-Adjei D, McLeod HL: Ethnic variation in the

HER-2 codon 655 genetic polymorphism previously associated with breast cancer. Journal of human genetics 2002, 47(4):172-175.

[58] Papewalis J, Nikitin A, Rajewsky MF: G to A polymorphism at amino acid codon 655 of the human erbB-2/HER2 gene. Nucleic acids research 1991, 19(19):5452.

[59] Fleishman SJ SJ, Ben-Tal N.: A putative molecular-activation switch in the transmembrane domain of erbB2. Proceedings of the National Academy of Sciences (PNAS) 2002, 99(25): 15937-15940.

[60] AbdRaboh NR, Shehata HH, Ahmed MB, Bayoumi FA: HER1 R497K and HER2 I655V polymorphisms are linked to development of breast cancer. Disease markers 2013, 34(6):407-417.

[61] Akisik E, Dalay N: Estrogen receptor codon 594 and HER2 codon 655 polymorphisms and breast cancer risk. Experimental and molecular pathology 2004, 76(3):260-263.

[62] Al-Janabi AM, Algenabi AHA, Kamoona TH, Alkhafaji SM: Association of HER2 [ILe655Val] gene polymorphism and Breast Cancer Risk in Iraqi females Population. International Journal of Advanced Research 2015, 3(12):1483 – 1489.

[63] Baxter SW, Campbell IG: Re: Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. Journal of the National Cancer Institute 2001, 93(7):557-559.

[64] Benusiglio PR, Lesueur F, Luccarini C, Conroy DM, Shah M, Easton DF, Day NE, Dunning AM, Pharoah PD, Ponder BA: Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case-control study. Breast cancer research : BCR 2005, 7(2): R204-R209.

[65] Cox DG, Hankinson SE, Hunter DJ: The erbB2/HER2/neu receptor polymorphism Ile655Val and breast cancer risk. Pharmacogenetics and genomics 2005, 15(7):447-450.

[66] de Almeida FC, Banin Hirata BK, Ariza CB, Losi Guembarovski R, de Oliveira KB, Suzuki KM, Guembarovski AL, Oda JMM, Vitiello GAF, Watanabe MAE: HER2 Ile655Val polymorphism is negatively associated with breast cancer susceptibility. Journal of clinical laboratory analysis 2018, 32(6):e22406.

[67] Han X, Diao L, Xu Y, Xue W, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B *et al*: Association between the HER2 Ile655Val polymorphism and response to trastuzumab in women with operable primary breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology 2014, 25(6):1158-1164.

[68] Hishida A, Hamajima N, Iwata H, Matsuo K, Hirose K, Emi N, Tajima K: Re: Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. Journal of the National Cancer Institute 2002, 94(23):1807-1808.

[69] Kalemi TG, Lambropoulos AF, Gueorguiev M, Chrisafi S, Papazisis KT, Kotsis A: The association of p53 mutations and p53 codon 72, Her 2 codon 655 and MTHFR C677T polymorphisms with breast cancer in Northern Greece. Cancer letters 2005, 222(1):57-65.

[70] Kallel I, Kharrat N, Al-fadhly S, Rebai M, Khabir A, Boudawara TS, Rebai A: HER2 polymorphisms and breast cancer in Tunisian women. Genetic testing and molecular biomarkers 2010, 14(1):29-35.

[71] Kamali-Sarvestani E, Talei AR, Merat A: Ile to Val polymorphism at codon 655 of HER-2 gene and breast

cancer risk in Iranian women. Cancer letters 2004, 215(1):83-87.

[72] Kara N, Karakus N, Ulusoy AN, Ozaslan C, Gungor B, Bagci H: P53 codon 72 and HER2 codon 655 polymorphisms in Turkish breast cancer patients. DNA and cell biology 2010, 29(7):387-392.

[73] Keshava C, McCanlies EC, Keshava N, Wolff MS, Weston A: Distribution of HER2(V655) genotypes in breast cancer cases and controls in the United States. Cancer letters 2001, 173(1):37-41.

[74] Lemieux J, Diorio C, Cote MA, Provencher L, Barabe F, Jacob S, St-Pierre C, Demers E, Tremblay-Lemay R, Nadeau-Larochelle C *et al*: Alcohol and HER2 polymorphisms as risk factor for cardiotoxicity in breast cancer treated with trastuzumab. Anticancer research 2013, 33(6):2569-2576.

[75] Lee SC, Hou MF, Hsieh PC, Wu SH, Hou LA, Ma H, Tsai SM, Tsai LY: A case-control study of the HER2 Ile655Val polymorphism and risk of breast cancer in Taiwan. Clinical biochemistry 2008, 41(3):121-125.

[76] Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, Geradts J, Iacocca M, Cowan D, Conway K *et al*: HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast cancer research and treatment 2003, 79(3):355-364.

[77] Mutluhan H, Akbas E, Erdogan NE, Soylemez F, Senli MS, Polat A, Helvaci I, Seyrek E: The influence of HER2 genotypes as molecular markers on breast cancer outcome. DNA and cell biology 2008, 27(10):575-579.

[78] Naidu R, Yip CH, Taib NA: Polymorphisms of HER2 Ile655Val and cyclin D1 (CCND1) G870A are not associated with breast cancer risk but polymorphic allele of HER2 is associated with nodal metastases. Neoplasma 2008, 55(2):87-95.

[79] Nelson SE, Gould MN, Hampton JM, Trentham-Dietz A: A case-control study of the HER2 Ile655Val polymorphism in relation to risk of invasive breast cancer. Breast cancer research : BCR 2005, 7(3): R357-R364.

[80] Papadopoulou E, Simopoulos K, Tripsianis G, Tentes I, Anagnostopoulos K, Sivridis E, Galazios G, Kortsaris A: Allelic imbalance of HER-2 codon 655 polymorphism among different religious/ethnic populations of northern Greece and its association with the development and the malignant phenotype of breast cancer. Neoplasma 2007, 54(5):365-373.

[81] Parvin S, Islam MS, Al-Mamun MM, Islam MS, Ahmed MU, Kabir ER, Hasnat A: Association of BRCA1, BRCA2, RAD51, and HER2 gene polymorphisms with the breast cancer risk in the Bangladeshi population. Breast cancer 2017, 24(2):229-237.

[82] Pinto D, Vasconcelos A, Costa S, Pereira D, Rodrigues H, Lopes C, Medeiros R: HER2 polymorphism and breast cancer risk in Portugal. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation 2004, 13(3):177-181.

[83] Qu S, Cai Q, Gao YT, Lu W, Cai H, Su Y, Wang SE, Shu XO, Zheng W: ERBB2 genetic polymorphism and breast cancer risk in Chinese women: a population-based case-control study. Breast cancer research and treatment 2008, 110(1):169-176.

[84] Rajkumar T, Samson M, Rama R, Sridevi V, Mahji U, Swaminathan R, Nancy NK: TGFbeta1 (Leu10Pro), p53

(Arg72Pro) can predict for increased risk for breast cancer in south Indian women and TGFbeta1 Pro (Leu10Pro) allele predicts response to neo-adjuvant chemo-radiotherapy. Breast cancer research and treatment 2008, 112(1):81-87.

[85] Roca L, Dieras V, Roche H, Lappartient E, Kerbrat P, Cany L, Chieze S, Canon JL, Spielmann M, Penault-Llorca F *et al*: Correlation of HER2, FCGR2A, and FCGR3A gene polymorphisms with trastuzumab related cardiac toxicity and efficacy in a subgroup of patients from UNICANCER-PACS 04 trial. Breast cancer research and treatment 2013, 139(3):789-800.

[86] Sezgin E, Sahin FI, Yagmurdur MC, Demirhan B: HER-2/neu gene codon 655 (Ile/Val) polymorphism in breast carcinoma patients. Genetic testing and molecular biomarkers 2011, 15(3):143-146.

[87] Siddig A, Mohamed AO, Kamal H, Awad S, Hassan AH, Zilahi E, Al-Haj M, Bernsen R, Adem A: HER-2/neu Ile655Val polymorphism and the risk of breast cancer. Annals of the New York Academy of Sciences 2008, 1138:84-94.

[88] Tommasi S, Fedele V, Lacalamita R, Bruno M, Schittulli F, Ginzinger D, Scott G, Eppenberger-Castori S, Calistri D, Casadei S *et al*: 655Val and 1170Pro ERBB2 SNPs in familial breast cancer risk and BRCA1 alterations. Cellular oncology : the official journal of the International Society for Cellular Oncology 2007, 29(3):241-248.

[89] Wang-Gohrke S, Chang-Claude J: Re: Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. Journal of the National Cancer Institute 2001, 93(21):1657-1659.

[90] Watrowski R, Castillo-Tong DC, Wolf A, Schuster E, Fischer MB, Speiser P, Zeillinger R: HER2 Codon 655 (Ile/Val) Polymorphism and Breast Cancer in Austrian Women. Anticancer research 2015, 35(12):6667-6670.

[91] Zhang M, Guo LL, Cheng Z, Liu RY, Lu Y, Qian Q, Lei Z, Zhang HT: A functional polymorphism of TGFBR2 is associated with risk of breast cancer with ER(+), PR(+), ER(+)PR(+) and HER2(−) expression in women. Oncology letters 2011, 2(4):653-658.

[92] Zubor P, Kajo K, Stanclova A, Szunyogh N, Galo S, Dussan CA, Minarik G, Visnovsky J, Danko J: Human epithelial growth factor receptor 2[Ile655Val] polymorphism and risk of breast fibroadenoma. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation 2008, 17(1):33-38.

[93] Tao W, Wang C, Han R, Jiang H: HER2 codon 655 polymorphism and breast cancer risk: a meta-analysis. Breast cancer research and treatment 2009, 114(2):371-376.

[94] Lu S, Wang Z, Liu H, Hao X: HER2 Ile655Val polymorphism contributes to breast cancer risk: evidence from 27 case-control studies. Breast Cancer Res Treat 2010, 124(3):771-778.

[95] Wang H, Liu L, Lang Z, Guo S, Gong H, Guan H, Zhang J, Liu B: Polymorphisms of ERBB2 and breast cancer risk: a meta-analysis of 26 studies involving 35,088 subjects. Journal of surgical oncology 2013, 108(6):337-341.

[96] Krishna BM, Chaudhary S, Panda AK, Mishra DR, Mishra SK: Her2 (Ile)655(Val) polymorphism and its association with breast cancer risk: an updated meta-analysis of case-control studies. Sci Rep 2018, 8(1):7427.

[97] Ma Y, Yang J, Zhang P, Liu Z, Yang Z, Qin H: Lack of association between HER2 codon 655

polymorphism and breast cancer susceptibility: meta-analysis of 22 studies involving 19,341 subjects. Breast cancer research and treatment 2011, 125(1):237-241.

[98] Nguyen Thanh T, Nguyen Tran BS, Hoang Thi AP, Tran Binh T, Ba Nguyen T, Le Minh T, Nguyen Vu QH, Dang Cong T: HER2Ile655Val Single Nucleotide Polymorphism Associated with Early-Onset Breast Cancer Susceptibility: A Systematic Review and Meta-Analysis. Asian Pacific journal of cancer prevention : APJCP 2021, 22(1):11-18.

[99] Bocharov EV, Mineev KS, Volynsky PE, Ermolyuk YS, Tkach EN, Sobol AG, Chupin VV, Kirpichnikov MP, Efremov RG, Arseniev AS: Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J Biol Chem 2008, 283(11):6950-6956.

[100] Takano K OK, Kaneda H, Yamagata Y, Fujii S, Kanaya E, Kikuchi M, Oobatake M, Yutani K: Hydrophobic Residues to the Stability of Human Lysozyme Calorimetric Studies and X-ray Structural Analysis of the Five Isoleucine to Valine Mutants. J Mol Biol 1995, 254:62 - 76.

[101] Papadopoulou E, Tripsianis G, Tentes I, Anagnostopoulos K, Sivridis E, Galazios G, Kortsaris A: Allelic imbalance of HER-2 codon 655 polymorphism among different religious/ethnic populations of northern Greece and its association with the development and the malignant phenotype of breast cancer. *Neoplasma* 2007, 54(6):365 - 373.

[102] Xie Dawen X-OS, Zonglin Deng, Wan-Qing Wen, Kim E. Creek, Qi Dai, Yu-Tang Gao, Fan Jin, Wei Zheng: Population-Based, Case–Control Study of HER2 Genetic Polymorphism and Breast Cancer Risk. Journal of the National Cancer Institute, 2000, 92(5):412 - 417.

[103] Karen G. Montgomery DMG, Simon W. Baxter, Roger L. Milne, Gillian S. Dite, Margaret R.E. McCredie, Graham G. Giles, Melissa C. Southey, John L. Hopper, Ian G. Campbell: The HER2 I655V Polymorphism and Risk of Breast Cancer in Women < Age 40 Years. Cancer Epidemiology, Biomarkers & Prevention 2003, 12:1109-1111.

[104] Joni L. Rutter, Nilanjan Chatterjee fSW, f and Jeffrey Struewing: The HER2 I655V polymorphism and Breast cancer risk in Ashkenazim. Epidemiology 2003, 14(6): 694-700.

[105] Robert Millikan AE, Kendra Worley, Lorna Biscocho, Elizabeth Hodgson, Wen-Yi Huang, Joseph Geradts, Mary Iacocca, David Cowan, Kathleen Conway, Dressler L: HER2 codon 655 polymorphism and risk of breast cancer in African American and whites. Breast Cancer Research and Treatment 2003, 79:355 - 364.

[106] Tommasi Vita Fedele, Michele Bruno, Francesco Schittulli, David Ginzinger, Gery Scott, Serenella Eppenberger-Castori, Daniele Calistri, Silvia Casadei, Ian Seymour, Salvatore Longo, Gianluigi Giannelli, Brunella Pilato, Giovanni Simone, Christopher C. Benz and Angelo Paradiso: 655Val-and-1170Pro-ERBB2-SNPs-in-familial-breast-cancer-risk-and-BRCA1-alterations. Cellular Oncology 2007, 29:241 - 248.

[107] Nurten Kara NK, Ali Naki Ulusoy, Cihangir Ozaslan, Bulent Gungor, and Hasan Bagci: P53-codon-72-and-HER2-codon-655-polymorphisms-in-Turkish-breast-cancer-patients. DNA and cell biology 2010, 29(7):387-392.

[108] An Hee Jung, Haeyoun Kang, Nam Keun Kim SGK: Her2 genotype and breast cancer progression in Korean

women. Pathology International 2005, 55:48 - 52.

[109] Chan KY, Cheung AN, Yip SP, Ko HH, Lai TW, Khoo US: Population-based case-control study of HER2 genetic polymorphism and breast cancer risk. Journal of the National Cancer Institute 2002, 94(20):1581-1582.

[110] Tezcan G, Tunca B, Ak S, Cecener G, Egeli U: Molecular approach to genetic and epigenetic pathogenesis of early-onset colorectal cancer. World J Gastrointest Oncol 2016, 8(1):83-98.

[111] Zhao B, Baloch Z, Ma Y, Wan Z, Huo Y, Li F, Zhao Y: Identification of Potential Key Genes and Pathways in Early-Onset Colorectal Cancer Through Bioinformatics Analysis. Cancer Control 2019, 26(1):1073274819831260.

[112] Mo X, Su Z, Yang B, Zeng Z, Lei S, Qiao H: Identification of key genes involved in the development and progression of early-onset colorectal cancer by co-expression network analysis. Oncol Lett 2020, 19(1):177-186.

[113] Weischenfeldt J, Simon R, Feuerbach L, Schlangen K, Weichenhan D, Minner S, Wuttig D, Warnatz HJ, Stehr H, Rausch T *et al*: Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. Cancer Cell 2013, 23(2):159-170.

[114] Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, Heckmann D, Sidiropoulos N, Waszak SM, Hubschmann D *et al*: Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell 2018, 34(6):996-1011 e1018.

Chapter 7

Human Genetic Polymorphisms Associated with Susceptibility to COVID-19 Infection and Response to Treatment

Necla Benlier, Nevhiz Gundogdu and Mehtap Ozkur

Abstract

Clinicians and researchers observing the natural history of endemic and epidemic infections have always been fascinated by the vagaries of these diseases, in terms of both the changing nature of the disease severity and phenotype over time and the variable susceptibility of hosts within exposed populations. SARS-CoV-2, the virus that causes COVID-19 and is believed to originate from bats, quickly transformed into a global pandemic. The pandemic of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been posing great threats to the global health in many aspects. Currently, there are no proven effective vaccines or therapeutic agents against the virus. Comprehensive understanding of the biology of SARS-CoV-2 and its interaction with hosts is fundamentally important in the fight against SARS-CoV-2. Advanced age, male sex, and comorbidities such as hypertension and cardiovascular disease as well as diabetes and obesity have been identified as risk factors for more severe COVID-19. However, which and to what extent specific genetic factors may account for the predisposition of individuals to develop severe disease or to contract the infection remains elusive. The increasing availability of data from COVID-19 patient populations is allowing for potential associations to be established between specific gene loci and disease severity, susceptibility to infection, and response to current/future drugs.

Keywords: Human Genetic Polymorphisms, SARS-COV-2, COVID-19, Genetic Susceptibility

1. Introduction

Variability in response to drugs, both in terms of efficacy and tolerability, and ways to customize treatments according to their characteristics have become important topics of medical research. It was determined that genetic variations in some ethnic groups may affect the response to drug and the outcomes of treatment [1]. Various enzymes are often involved in drug absorption/distribution/metabolism/excretion (ADME) processes and show multiple interactions [2].

Gene polymorphisms are sequence variations at specific locations within the genome and are observed in more than 1% of the population. Genetic polymorphisms can alter the coding of proteins or their expression, and affect natural

or acquired immunity [3]. Single nucleotide polymorphisms (SNPs) have been studied in relation to various diseases, which are associated with variations of DNA sequence with phenotypic changes [4]. Variations in the genes control competence in the cellular and humoral immune systems, which define the individual risk level for diseases [5, 6].

2. SARS-CoV-2

After the investigation that started with the World Health Organization (WHO) China Country Office reporting a cluster of pneumonia cases of unknown cause in Wuhan city of Hubei province of China on 31 December 2019, it was identified on January 7, 2020 that the agent was a new coronavirus that causes infection in humans. The causative virus was designated as 2019-nCoV (2019-novel coronavirus) by the WHO and SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2) by the International Virus Taxonomy Committee, and the disease caused by the virus was named as COVID-19 (coronavirus disease- 2019) [7, 8].

The disease is highly contagious and its main clinical symptoms are fever, dry cough, fatigue, muscle pain and shortness of breath. Since the first reported case of COVID-19 in the city of Wuhan, China, at the end of 2019, COVID-19 has rapidly spread all over China and then to all countries of the world [9]. It was confirmed that the virus spreads from person to person, through close contact and via respiratory particles that are generated by coughing or sneezing [10]. Most of the studies conducted in Wuhan at the beginning of the epidemic showed that the first patients worked in or visited a seafood market in Wuhan. Initially, it was thought to be caused by snakes, but later studies have shown it to be related to bats. As the pandemic progressed, it was shown that this viral infection is transmitted from one person to another through droplets and by touching the face with hands exposed to contaminated surfaces [11]. The virus can be found in respiratory secretions of patients 1–2 days before the onset of clinical symptoms and two weeks after disease symptoms [12].

Coronaviruses belong to the subfamily *Coronavirinae* of the family *Coronaviridae* and include four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus. The genome of CoVs (27–32 kb) is a single-stranded positive-sense RNA (+ ssRNA) that is larger than other RNA viruses [13]. The genome size of SARS-CoV-2 is approximately 29.9 kb [14].

Coronaviruses contain four structural proteins, namely S (Spike) protein, E (Envelope) protein, M (Membrane) protein and N (Nucleocapsid) protein, which are involved in the formation of mature virus particles (virions) and the emergence of infection [15]. The S protein is responsible for the formation of spikes on the surface of the coronavirus, and these protrusions play a key role in the attachment of the virus to host cell receptors and its entry into the cell. The spike protein is also the main antigenic component of the virus [16].

Depending on viral strains and cell types, coronavirus S proteins can be cleaved by one or more host proteases, including cathepsins, furin, neutrophil elastase (ELANE), transmembrane protease serine protease-2 (TMPRSS-2), and possibly TMPRSS11A [17–19]. The presence of these proteases on target cells largely determines whether coronavirus particles enter cells via the plasma membrane or endocytosis. Hoffmann et al. showed that SARS-CoV-2 uses transmembrane protease 2 (TMPRSS2) for S protein priming and the SARS-CoV receptor angiotensin-converting enzyme 2 (ACE2) for entry into target cells [20]. Likewise, Ou et al. found that cathepsin L (CTSL) is essential to the virus entry in the host cell [21]. Moreover, it was reported that the S protein of the A2a subtype possesses an additional elastase-specific proteolytic cleavage site that enhances the virus's ability

to penetrate host cells [19]. This virus subtype has been reported in China and has spread rapidly in North America and Europe [22, 23].

Coronaviruses contain sixteen nonstructural proteins (Nsp1–16). Nsp1 mediates RNA processing and replication and Nsp2 modulates the host cell's survival signaling pathway. Nsp3 separates the translated polyprotein into its distinct proteins. Nsp4 contains transmembrane domain 2 (TM2) and modifies ER membranes. Nsp5 participates in the polyprotein processing during replication. Nsp6 is a putative transmembrane domain. A heterodimer of Nsp7 and Nsp8 significantly increases nsp12 and template-primary RNA combination. Nsp9 participates in viral replication by acting as an ssRNA-binding protein. Nsp10 is crucial for cap methylation of viral mRNAs. Nsp11 is identical to the first segment of Nsp12 and its function is unknown. Nsp12 contains RNA-dependent RNA polymerase (RdRp), a critical component of coronavirus replication/transcription. Nsp13 binds to ATP and the zinc-binding domain of Nsp13 is involved in replication and transcription. Nsp14 is an exoribonuclease domain. Nsp15 has Mn (2+)-dependent endoribonuclease activity. Nsp16 is a 2'-O-ribose methyltransferase [24]. Entry of coronaviruses into host cells is mediated by an increase in glycoprotein (S protein) [16, 25, 26]. Transmembrane spike glycoproteins form homotrimers that protrude from the viral surface. The spike glycoprotein is critical for the entry of coronaviruses, thus making it an attractive antiviral target. A six-helical bundle (6-HB) is formed by HR1 and HR2 which are vital for spike protein-dominated membrane fusion of SARS-CoV or SARS-CoV-2, making HR1 and HR2 a promising drug target [27, 28]. The spike protein of SARS-CoV-2 contains a receptor binding domain that specifically recognizes the ACE2 receptor. The receptor-binding domain is a critical target for antiviral compounds and antibodies [29].

Most patients with COVID-19 develop symptoms such as dyspnea, fever, dry cough, muscle pain, fatigue and diarrhea. However, the rate of complications such as sepsis, septic shock and multiple organ dysfunction syndrome (MODS) has been found to range between 2 and 20% in line with the data collected to date [30–32]. Pathophysiological findings in the lungs, which are the entry route of the virus, show embolisms caused by thrombus formation at the micro and macro level associated with extensive endothelial involvement, as well as intra-alveolar hyaline membranes and alveolar edema [33]. Widespread pulmonary damage is attributed to a cytokine storm that shows local and then systemic effects, together with the damage that starts directly with the virus-ACE2 receptor relationship [34].

3. ACE2

ACE2 cleaves Ang II to angiotensin [1–7], which produces vasodilating, anti-inflammatory and anti-fibrotic effects via binding to the Mas receptor [35–37]. Tissue-bound or membrane-bound ACE2 is a type of transmembrane protein with a single metalloprotease active site and a transmembrane domain [38, 39]. ACE2 is expressed in virtually all human organs in varying degrees. ACE2 expression is present in type II alveolar cells (AT2), respiratory epithelial cells, bronchial transient epithelial secretory cells,, myocardial cells, endothelial cells and artery smooth muscle cells, esophagus epithelial cells, tongue epithelial cells, neurons and glia, stomach, cholangiocytes, adipose tissue, pancreatic exocrine glands and islets, bladder urothelial cells, renal proximal tubule cells, podocytes, testis (Leydig and Sertoli cells and spermatogonia), uterus epithelial cells, ovary and breast, maternal–fetal interface, enterocytes from ileum and colon and rectum cells [40, 41]. In the lung, ACE2 is abundantly expressed by Clara cells, type I and II alveolar epithelial cells, macrophages, bronchial epithelium, endothelium, and vascular smooth muscle cells [42].

ACE2 is encoded on chromosome Xp22 and spans 39.98 kb genomic DNA. This gene generates two transcripts originating from the same 805 amino acid residue protein: one transcript consisting of 18 exons and 17 introns (transcript length: 3339 bps), and the other consisting of 19 exons and 18 introns (transcript length: 3507 bps). The ACE2 gene displays high levels of polymorphism; in fact, some single nucleotide polymorphisms (SNPs) of this gene have been reported to be associated with susceptibility to diseases such as hypertension and type 2 diabetes [43, 44].

Single-cell RNA sequencing analysis has shown that ACE2 mRNA is expressed at a higher level in the Asian population than in the Caucasian and African-American populations, and also Asian men have a higher ACE2 mRNA expression compared to Asian women [45, 46].

There is a negative correlation between ACE2 expression and COVID-19 severity. ACE2 is secreted from membranes at different levels. Two cell membrane proteases are important for this secretion: transmembrane protease serine 2 (TMPRSS2) and protein 17 (ADAM17) containing disintegrin and metalloproteinase domain [47].

More specifically, ADAM17 acts directly on ACE2 and leads to the shedding of ACE2 into the extracellular cellular space, whereas TMPRSS2 affects not only ACE2, but also the S protein of SARS-CoV-2, resulting in membrane fusion and cellular uptake of the virus. As a result of an extensive database analysis, Cao et al. identified 1700 variants in the ACE2 gene region on the X chromosome. They identified 15 (14 SNPs and 1 insertion/deletion (INDEL)) uniquely expressed variants with higher minor allele frequencies (MAF) in the Asian population than in the European population [48].

It is still debated whether these differences should be taken into account in epidemiological studies on COVID-19, which includes ethnic associations with disease development [49]. Importantly, diseases correlated with high levels of SARS-Cov-2 infection, such as hypertension and diabetes, have been found to have a lower expression of ACE2 in relation to the SNPs in the ACE2 genes.

Individuals with rs383510/T and rs2070788/G genotypes of TMPSSRSS2 located on chromosome 21q22.3 were found to be more prone to develop a severe form of influenza A (H1N1) and acute respiratory distress syndrome (ARDS) [50]. Remarkably, males have been shown to be more likely to develop severe H1N1 influenza [51].

The ADAM17 gene region on chromosome 2p25.1 has been determined to differ in allele profiles between Asian and European populations and these SNPs are associated with hypertension [52] and/or sepsis [53].

When the S1 protein, located at the spikes of SARS-CoV and SARS-CoV2, attaches to the enzymatic domain of ACE2 on the cell membrane, both the virus and the enzyme are taken up into the host cell by endocytosis [54, 55]. This led to the idea that lowering the amount of ACE2 in cells could help fight against coronavirus infection. As an antithesis to this, ACE2 has also been shown to have a protective effect against viral lung injury by increasing the production of the vasodilator angiotensin 1–7. Also, some studies in mice have demonstrated that the interaction of the coronavirus terminal protein with ACE2 causes a reduction in ACE2 levels in the cell membrane, with the protein being pulled into and degraded, and therefore may increase lung damage [56].

Studies in rodents have shown that both ACE inhibitors and angiotensin receptor blockers (ARBs) used to treat high blood pressure increase the amount of ACE2 and therefore may increase the severity of coronavirus infections [57]. However, scientific societies have recommended continuing standard ACE inhibitor and ARB treatment [58]. A systematic review and meta-analysis published on July 11, 2012

found that the use of ACE inhibitors resulted in a 34% reduction in pneumonia risk compared to controls. Besides, it was observed that the risk of pneumonia is reduced by treatment with ACE inhibitors in patients with a high risk of pneumonia, especially those with stroke and heart failure [59].

ACE2 limits the adverse vasoconstrictor and profibrotic effects of AngII. Hydrolysis of AngII to Ang (1e7) reduces the oxidative stress of AngII on endothelial cerebral arteries. Disruption of ACE2 results in increased AngII levels and impaired cardiac function. Decreased cardiac ACE2 levels have been reported in hypertension (HT) and diabetic heart disease, and low ACE2 mRNA expression has been associated with HT, dyslipidemia, and/or heart failure [60, 61].

It was reported that ACE2 gene polymorphisms can influence both susceptibility to SARS-CoV-2 and the prognosis of COVID-19 disease. The S1 domain of the SARS-CoV2 spike protein mediates its binding to the ACE2 receptor site, while the S2 domain mediates membrane fusion at the membrane-associated portion undergoing postbinding transconformational modifications. In a study conducted by Li et al. in rats, they produced a conformational change in the ACE2 α-helix 1 structure by changing the His353 amino acid of the ACE2'receptor and modifying a glycosylation site (Asp 90), and as a result, they determined that this receptor became more suitable for the binding of SARS-CoV [62]. They also found that Leu584Ala, a point mutation in ACE2, significantly increased the binding of the enzyme, thus facilitating the entry of SARS-CoV into target cells [63]. In some studies, ACE2 expression was found to be low in cells infected with SARS-CoV, and recombinant SARS-CoV spike protein decreased ACE2 expression and thus increased lung damage [64, 65]. In a study by Cao et al. in different populations, seven of the 32 ACE2 variants (Lys26Arg, Ile486Val, Ala627Val, Asn638Ser, Ser692Pro, Asn720Asp and Leu731Ile/Phe) were found to be effective for SARS-CoV2 [48].

In one study, Stawiski et al. reported that while human ACE2 variants K26R, T27A, N64K, S19P, I21V, E23K, T92I, Q102P and H378R are predicted to increase host susceptibility, other ACE2 variants K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, M62V, K68E, F72V, Y83H, G326E, G352V, D355N, Q388L and D509Y are putative protective variants predicted to show decreased binding to SARS-CoV-2 S-protein. Among these, T92I variant, part of a consensus NxS/T N-glycosylation motif, exhibited increased affinity for S-protein [66].

4. Transmembrane serine protease

Transmembrane protease serine type 2 (TMPRSS2) belongs to the type II transmembrane serine protease family. It facilitates the entry and activation of the virus by making proteolytic cleavage in the spike protein. After SARS-CoV binds to ACE2, proteolytic cleavage of the S protein via the cysteine protease cathepsin B/L or TMPRSS2 is required for the virus to enter the cell. Although both activate the SARS CoV Spike protein, TMPRSS2 activity has been shown to be required for the spread of the virus in the host. While rodents given TMPRSS2 serine protease inhibitor were protected from SARS-CoV, the same effect was not observed in those given cathepsin B/L cysteine prosthetic inhibitor [67]. The transmembrane serine protease TMPRSS2 is an essential enzyme capable of degrading the hemagglutinin of many influenza virus subtypes and coronavirus S protein [68, 69]. Studies have shown that mice with TMPRSS2 deficiency are more resistant to infections with H1N1 and H7N9 influenza A virus [68, 70]. TMPRSS2 has been demonstrated to help SARS-CoV-2 enter host cells by cleaving the S protein [20]. Matsuyama et al. showed that cell lines expressing TMPRSS2 are highly susceptible to MERS-CoV, SARS-CoV and SARS-CoV-2 [71]. The gene encoding TMPRSS2 is polymorphic and

is regarded as a susceptibility gene for H1N1 and H7N9 influenza [50]. TMPRSS2 is expressed in numerous tissues that are targets of COVID-19, such as the lung, heart, kidney, and digestive tract. It is also expressed in microvascular endothelial cells, suggesting that it may play a role in endothelial dysfunction, thrombosis and related complications. TMPRSS2 variants are also thought to contribute to the clinical diversity of COVID-19. In the study of Asselta et al., it was reported that TMPRSS2 exonic variant p.Val160Met and two haplotypes were detected more frequently in the Italian population than in East Asians [72]. Besides, in another study, the presence of TMPRSS2-ERG fusion in prostate cancer and the strong regulation of TMPRSS2 by androgens led to the hypothesis that TMPRSS2 could partly explain the greater involvement of males in the COVID-19 pandemic [73].

5. Elastase

SARS-Cov-2 enters the cell by binding its S protein with cellular receptors [20]. Some proteases, such as TMPRSS2, cathepsin L, neutrophil elastase, and probably TMPRSS11A are involved in this process. As a matter of fact, polymorphisms in their encoding genes could not only have an impact in the expression and/or structure of these proteases but also be associated with susceptibility to SARS-CoV-2 infection. Elastase is secreted by neutrophils as part of an inflammatory response to a viral infection and is also produced by opportunistic bacteria that can colonize virally infected respiratory tissue [74]. Increased elastase activity as a result of the aberrant inflammatory process produces considerable pulmonary damage that contributes significantly to the pathogenesis of chronic obstructive pulmonary disease, cystic fibrosis, ARDS and pulmonary fibrosis [75, 76]. Moreever, the dramatic increase in neutrophil elastase (NE) in severe COVID-19 may be related to neutrophil activation by the IL-8/CXCR2 pathways [77]. The ELANE gene encoding neutrophil elastase is located on chromosome 19p13.3. Two transcripts have been reported for this gene, which produce consensus coding sequence. The first transcript contains 5 exons and 4 introns and 5 exons encode this 267 amino acid protein (transcript length, 909 bps). The second transcript consists of 6 exons and 5 introns, and 5 exons encode this 267 amino acid protein (transcript length, 1028) [78]. In the ELANE gene, 12 polymorphisms with potential functional effects were identified: ten in the promoter region, two in the 5′ region near the gene and two in the 3′ region near the gene. These 12 polymorphisms produce binding sites for various transcription factors and microRNAs [78].

The levels of NE expression are known to be affected by the polymorphisms in the promoter region of the neutrophil elastase (ELANE) gene. Several polymorphisms were identified to date in the six repetitive tandem motifs of the ELANE gene promoter region: -903 T/G, −741G/A, −832G/T, -789C/T, and extra 52 bp between the fourth and fifth repeats. Polymorphisms -903 T/G and -741G/A have been associated with risk of lung cancer [79]. Luciferase activity assays have shown higher activity for ELANE gene promoter constructs carrying -903 T/−741G compared to the constructs carrying -903G/−741A. Based on these findings, predicted activity of ELANE genotypes was classified as low (−903TG), intermediate (−903TT/−741AG and -903TT/−741AA), or high (−903TT/−741GG) [80].

6. Cathepsin L

Cathepsin L is a peptidase that cleaves peptide bonds, preferably with aromatic residues at the P2 position and hydrophobic residues at the P3 position [81]. It was

reported previously that cathepsin L participates in the viral glycoprotein processing of Ebola virus and SARS-CoV, and this viral process has been found to be critical for cell membrane fusion and host cell entry [82]. Using cathepsin B and L inhibitors in HEK 293/hACE2 cells, Ou et al. showed that treatment with cathepsin L inhibitor reduced the entry of SARS-CoV-2 into cells [21]. This finding suggests that cathepsin L may be crucial for S protein priming in the lysosome for viral entry.

Six polymorphisms with possible functional implications were identified in the cathepsin L gene. These polymorphisms have been found in various regions of the gene and form binding sites for transcription factors [78]. Among these polymorphisms, rs41307457 has a high frequency only in the African population and rs41312184 is present with a high frequency especially in the European population. The authors suggested that the relationship of these polymorphisms with SARS-CoV-2 infection should be analyzed in these populations [78].

7. Human alpha-1 antitrypsin (A1AT)

Human alpha-1 antitrypsin (A1AT) is a 52 kDa glycoprotein synthesized in the liver and circulates in the blood, and is a natural inhibitor of several proteases. Adequate A1AT activity is crucial for the prevention of proteolytic tissue damage [83]. In people with one of many inherited mutations in A1AT, low circulating A1AT levels increase the risk of devastating diseases, particularly emphysema [84]. Infusion of plasma-purified A1AT protein has proven therapeutic benefits in patients with A1AT deficiency [85]. The pharmacokinetics and safety of A1AT have been well studied. It was accepted as safe with its rare and generally well tolerated side effects [86]. Studies have shown that human A1AT has both anti-inflammatory and anti-SARS-CoV-2 viral effects [87]. This dual role makes it a unique and excellent candidate for the treatment of COVID-19. Alpha-1-antitrypsin (AAT) is a serine protease inhibitor (SERPIN) and the third most abundant circulating protein. AAT plasma level may increase 3 to 5-fold in states of systemic inflammation and/or infection, perhaps an indication of the homeostatic role of AAT, but has been found to be insufficient in severe cases of COVID-19 [88, 89]. Alpha-1-antitrypsin (AAT) has been shown to antagonize various pathophysiological mechanisms induced by SARS-CoV-2. It has been demonstrated that these pathophysiological mechanisms inhibit TMPRSS-2, the host serine protease that degrades the spike protein of SARS-CoV-2, SARS-CoV-2 [20]. AAT also has antiviral activity against other RNA viruses, influenza and HIV. It also induces autophagy, a known host effector mechanism against MERS-CoV, a related coronavirus that causes Middle East Respiratory Syndrome [90]. Additionally, AAT has potent anti-inflammatory properties, in part by inhibiting both nuclear factor-kappa B (NFκB) activation and ADAM17 (also known as tumor necrosis factor-alpha converting enzyme) and may therefore reduce the hyperinflammatory response to COVID-19 [91]. Moreover, AAT inhibits neutrophil elastase, a serine protease that helps recruit potentially harmful neutrophils and is implicated in acute lung injury [92]. AAT inhibition of ADAM17 prevents ACE2 from being scavenged, thus preserving ACE2 inhibition of bradykinin and reducing bradykinin's ability to cause capillary leakage in COVID-19 [93]. AAT also inhibits thrombin and venous thromboembolism, and microthrombi and macrothrombi in situ are increasingly recognized to play a role in COVID-19 [94]. Furthermore, AAT inhibition of elastase results in the formation of neutrophil extracellular traps (NETs), a complex extracellular structure composed of neutrophil-derived DNA, histones, and proteases, and involved in the immunothrombosis of COVID-19 can antagonize. In fact, AAT has been shown to alter the shape and adherence of NETs not associated with COVID-19 [95]. AAT inhibition of endothelial cell apoptosis may

also limit endothelial damage associated with severe COVID-19-related acute lung injury, multi-organ dysfunction, and preeclampsia-like syndrome in gravid women [96]. Although it is well known that alpha 1 antitrypsin deficiency is quite common in Europeans, rs17580 is the most common deficiency variant as reported by most of the studies conducted to date [97].

8. HLA

Human leukocyte antigens (HLA) are encoded by major histocompatibility complex (MHC) genes and are highly polymorphic. MHC molecules act as receptors for viral peptides. Studies have shown that polymorphisms in the HLA region are associated with susceptibility to many common infectious diseases [98]. In a study conducted in 28 patients with severe respiratory failure, HLA-DR expression was found to be very low, suggesting that HLA is an important immune regulator in COVID-19 [99]. In addition, several studies have attempted to determine HLA alleles that are associated with increased or reduced susceptibility. HLA genes are important in olfactory perception. Loss of sense of smell differs among individuals who had COVID-19 [100]. The olfactory receptor gene is located at the same locus as the MHC and is co-transmitted [101]. Therefore, variations in HLA genes are known to play a role in differences in immune response against pathogens.

9. Conclusion

Today, many genetic polymorphisms are known to be involved in pathways that play an important role in the attachment of the microbiological agent to the host cell, resulting in variations in the susceptibility to disease and disease severity. Currently, genetic polymorphisms are used in molecular medicine for many purposes. Better understanding of the mechanisms caused by genetic polymorphisms is expected to allow for the development of new treatments and discovery of preventive drugs. The COVID-19 pandemic shows marked geographical differences in its prevalence and mortality. This variability may be due both to the presence of several subtypes of the virus and to genetic differences in human populations. Given this fact and the important roles of ACE2, TMPRSS2, cathepsin L and elastase in the process of virus entry into the host cell, this article aims to suggest possible variants at these loci for genetic association studies in SARS patients. Although there appears to be a multifactorial genetic influence on the risk of SARS-Cov-2 infection and possible disease severity, SNP profiling of the ACE2, ADAM17 and TMPRSS2 genes is recommended to identify potentially vulnerable populations at risk with a relatively simple and easy-to-perform test such as PCR [102] or MASSarray [103]. Thus, a 'multiSNP risk score' applicable to large populations can be determined, and therefore it may be possible to identify subjects carrying a combination of fewer suitable alleles for ACE2, ADAM17 and TMPRSS2. Such an analytical strategy has recently been developed based on patient genetics for immunogenetic profiling designed to individualize immunotherapy [104]. It is known that macrophage activation syndrome (MAS) is an important cause of mortality and morbidity in patients with COVID-19. In a study investigating gene polymorphisms in the pentraxin 3 (PTX3), a molecule that is synthesized by a number of inflammatory cells and considered to be associated with mortality, MAS was found to be less common in COVID-19 patients with the AG genotype (rs1840680 (1449A/G) polymorphism) and PTX3 levels were higher in patients carrying the A allele [103]. In a study on interferon-induced membrane protein- 3 (IFITM-3) gene variants,

an established risk factor in severe viral infections, the IFITM3-SNP, rs12252-G allele was found to be significantly associated with hospitalization and mortality in COVID-19 patients and lower IFNγ levels were lower patients with the AG/GG genotype [105]. Homocysteine can be used as a potential biomarker to predict the severity of a number of infections in COVID-19. A study in the Latino population suggested that the MTHFR 677 T allele may contribute to the mortality from COVID-19 [106]. Dipeptidylpeptidase-4 (DPP4) is known to be a key protein for the entry of SARS-CoV-2 into the host cell as well as in obesity and hypertension, which are associated with worse prognosis in COVID-19. In light of these data, a study suggested that the DPP4 rs3788979 polymorphism might be a risk factor for COVID-19 disease [107]. Recently, in a south Asian population, the frequency of Human leukocyte antigen (HLA) variants HLA-B*51 and HLA class II, DRB1*13 was found to be high in patients with fatal COVID-19 [108].

In summary, until now, genetic influences on the interindividual susceptibility of COVID-19 have been largely underestimated; therefore, we hope that this review will fill this gap and pave the way for validation in studies at the experimental and clinical levels. Taken together, these data suggest that several gene variants may have an effect on susceptibility to COVID-19 disease, its prognosis and possibly the efficacy of vaccines. As SARS-CoV-2 continues to threaten global health, it is essential to elucidate the molecular mechanisms involved in this infection to develop specific treatment and prevention strategies.

Author details

Necla Benlier[1]*, Nevhiz Gundogdu[2] and Mehtap Ozkur[1]

1 Department of Medical Pharmacology, Faculty of Medicine, University of SANKO, Gaziantep, Turkey

2 Department of Pulmonary Medicine, Faculty of Medicine, University of SANKO, Gaziantep, Turkey

*Address all correspondence to: nbenlier@hotmail.com.tr

References

[1] Calcagno A, Cusato J, D'Avolio A,rt al. Genetic Polymorphisms Affecting the Pharmacokinetics of Antiretroviral Drugs. Clin Pharmacokinet. 2017 Apr;56(4):355-369.

[2] Ikediobi O, Aouizerat B, Xiao Y, et al. Analysis of pharmacogenetic traits in two distinct South African populations. Hum Genomics. 2011;5(4):265-282.

[3] Kinane DF, Shiba H, Hart TC. The genetic basis of periodontitis. Periodontol 2000 2005;39:91-117.

[4] Mout R, Willemze R, Landegent JE. Repeat polymorphisms in the interleukin-4 gene (IL4). Nucleic Acids Res 1991;19:3763.

[5] Kinane DF, Hodge P, Eskdale J, et al. Ellis R, Gallagher G. Analysis of genetic polymorphisms at the interleukin-10 and tumour necrosis factor loci in early onset periodontitis. J Periodont Res 1999;34:379-386.

[6] Nunn ME. Understanding the etiology of periodontitis: An overview of periodontal risk factors. Periodontol 2000 2003;32:11-23.

[7] https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[8] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5(4):536-544.

[9] Debnath M, Banerjee M, Berk M. Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes, The FASEB Journal.2020; 34:8787– 8795.

[10] Du RH, Liang LR, Yang CQ, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. Eur Respir J. 2020; 55(5).

[11] Imai Y, Kuba K, Rao S, et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. Nature 2005 ;436(7047), 112-116.

[12] Yang XH, Deng W, Tong Z, et al. Mice transgenic for human angiotensin-converting enzyme 2 provide a model for SARS coronavirus infection. Comparative Medicine 2007;57(5), 450-459.

[13] Brian DA, Baric RS. Coronavirus genome structure and replication. Curr Top Microbiol Immunol 2005; 287():1-30.

[14] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet (London England) 2020; 395, 565-574.

[15] Fehr AR, S. Perlman S. Coronaviruses: an overview of their replication and pathogenesis Methods Mol. Biol 2015; 1282, pp. 1-23

[16] Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus Nature 2003; 426, pp. 450-454

[17] Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein Proc. Natl. Acad. Sci. U. S. A. 2014; 111 , pp. 15214-15219,

[18] Bertram S, Dijkman R, Habjan M, et al. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the

respiratory epithelium. J Virol. 2013;87(11):6150-6160.

[19] Bhattacharyya C, Das C, Ghosh A, et al. SARS-CoV-2 mutation 614G creates an elastase cleavage site enhancing its spread in high AAT-deficient regions. Infect Genet Evol. 2021;90:104760.

[20] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271-280.e8.

[21] Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV [published correction appears in Nat Commun. 2021;12(1):2144].

[22] Biswas NK, Majumder PP. Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type Indian J. Med. Res 2020; 10.4103/ijmr.IJMR_1125_20

[23] Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med. 2020;382(24):2302-2315.

[24] Naqvi AAT, Fatima K, Mohammad T, et al. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochim Biophys Acta Mol Basis Dis. 2020;1866(10):165878.

[25] Li F, Li W, Farzan M, et al. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Sci. (N. Y. N.Y.)2005; 309, 1864-1868.

[26] Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu. Rev. Virol 2016; 3, 237-261.

[27] Liu S, Xiao G, Chen Y, et al. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. Lancet (London England) 2004; 363, 938-947.

[28] Xia S, Zhu Y, Liu M, et al. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell Mol Immunol. 2020;17(7):765-767.

[29] Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020;5(4):562-569.

[30] Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study [published correction appears in Lancet Respir Med. 2020 ;8(4):e26]. L

[31] Sun P, Qie S, Liu Z, et al. Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. J Med Virol 2020;92:612-617.

[32] Rodriguez-Moralesa AJ, CardonaOspinaa JA, Gutiérrez-Ocampoa E, et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020;34:101623.

[33] Al-Ania F, Chehade S, Lazo-Langner A. Thrombosis risk associated with COVID19 infection. A scoping review. Thromb Res 2020;192:152-160.

[34] Han H, Ma Q, Li C, et al. Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. Emerg Microbes Infect 2020;9:1123-1130.

[35] Sanchis-Gomar F, Lavie CJ, Perez-Quilis C, et al. Angiotensin-Converting Enzyme 2 and Antihypertensives (Angiotensin Receptor Blockers and Angiotensin-Converting Enzyme Inhibitors) in Coronavirus Disease 2019. Mayo Clin Proc. 2020;95(6):1222-1230.

[36] Hamming I, Cooper ME, Haagmans BL, et al. The emerging role of ACE2 in physiology and disease. J Pathol 2007; 212: 1-11.

[37] Bernardi S, Toffoli B, Zennaro C, et al. High-salt diet increases glomerular ACE/ACE2 ratio leading to oxidative stress and kidney damage. Nephrol Dial Transplant 2012; 27: 1793- 1800.

[38] Lambert DW, Yarski M, Warner FJ, et al. Tumor necrosis factor-alpha convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2). J Biol Chem. 2005;280(34):30113-30119.

[39] Xia H, Sriramula S, Chhabra KH, et al. Brain angiotensin-converting enzyme type 2 shedding contributes to the development of neurogenic hypertension. Circ Res. 2013;113(9):1087-1096.

[40] Zou X, Chen K, Zou J, et al. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Frontiers of Medicine. 2020;14:185-192.

[41] Beyerstedt S, Casaro EB, Rangel ÉB. COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. Eur J Clin Microbiol Infect Dis. 2021;40(5):905-919.

[42] Santos RA, Frezard F, Ferreira AJ. Angiotensin-(1-7): blood, heart, and blood vessels Curr. Med. Chem. Cardiovasc. Hematol. Agent 2005; 3, pp. 383-391.

[43] Zhang Q, Cong M, Wang N, et al. Association of angiotensin-converting enzyme 2 gene polymorphism and enzymatic activity with essential hypertension in different gender: A case-control study. Medicine (Baltimore). 2018;97(42):e12917.

[44] Liu C, Li Y, Guan T, et al. ACE2 polymorphisms associated with cardiovascular risk in Uygurs with type 2 diabetes mellitus. Cardiovasc Diabetol. 2018;17(1):127.

[45] Zhao Y, Zhao Z, Wang Y, et al. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. Am J Respir Crit Care Med. 2021; 203(6):782.

[46] Cai G. Tobacco-use disparity in gene expression of ACE2, the receptor of 2019-nCov. 2020.

[47] Heurich A, Hofmann-Winkler H, Gierer S, et al. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. J Virol. 2014;88(2):1293-1307.

[48] Cao Y, Li L, Feng Z, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov. 2020;6:11.

[49] Yi Y, Lagniton PNP, Ye S, et al. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci. 2020;16(10):1753-1766.

[50] Cheng Z, Zhou J, To KK, et al. Identification of TMPRSS2 as a Susceptibility Gene for Severe 2009 Pandemic A(H1N1) Influenza and

A(H7N9) Influenza. J Infect Dis. 2015;212(8):1214-1221.

[51] Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res. 1999;59(17):4180-4184.

[52] Li Y, Cui LL, Li QQ, et al. Association between ADAM17 promoter polymorphisms and ischemic stroke in a Chinese population. J Atheroscler Thromb. 2014;21(8):878-893.

[53] Shao Y, He J, Chen F, et al. Association Study Between Promoter Polymorphisms of ADAM17 and Progression of Sepsis. Cell Physiol Biochem. 2016;39(4):1247-1261.

[54] Wang H, Yang P, Liu K, et al. SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. Cell Res. 2008;18(2):290-301.

[55] Millet JK, Whittaker GR. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. Virology. 2018;517:3-8.

[56] Akhmerov A, Marbán E. COVID-19 and the Heart. Circ Res. 2020 May 8;126(10):1443-1455.

[57] Jia H. Pulmonary Angiotensin-Converting Enzyme 2 (ACE2) and Inflammatory Lung Disease. Shock. 2016;46(3):239-248.

[58] Diaz JH. Hypothesis: angiotensin-converting enzyme inhibitors and angiotensin receptor blockers may increase the risk of severe COVID-19. J Travel Med. 2020;27(3):taaa041.

[59] Bozkurt B, Kovacs R, Harrington B. Joint HFSA/ACC/AHA Statement Addresses Concerns Re: Using RAAS Antagonists in COVID-19. J Card Fail. 2020;26(5):370.

[60] Chen YY, Liu D, Zhang P, et al. Impact of ACE2 gene polymorphism on antihypertensive efficacy of ACE inhibitors. J Hum Hypertens 2016;30:766e71.

[61] Luo Y, Liu C, Guan T, et al. Association of ACE2 genetic polymorphisms with hypertension-related target organ damages in south Xinjiang. Hypertens Res 2019;42(5):681e9.

[62] Li W, Zhang C, Sui J, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J 2005;24:1634e43.

[63] Xiao F, Zimpelmann J, Agaybi S, et al. Characterization of angiotensin-converting enzyme 2 ectodomain shedding from mouse proximal tubular cells. PLoS One 2014;9(1):e85958.

[64] Kuba K, Imai Y, Rao S, et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. Nat Med 2005;11(8):875e9.

[65] Glowacka I, Bertram S, Herzog P, et al. Differential downregulation of ACE2 by the spike proteins of severe acute respiratory syndrome coronavirus and human coronavirus NL63. J Virol 2010;84(2):1198e205.

[66] Stawiski EW, Diwanji D, Suryamohan K, et al. Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility. BioRxiv 2020.

[67] Zhou Y, Vedantham P, Lu K, et al. Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. 2015;116:76-84.

[68] Sakai K, Ami Y, Tahara M, et al. The host protease TMPRSS2 plays a major role in in vivo replication of emerging H7N9 and seasonal influenza viruses. J Virol. 2014;88(10):5608-5616.

[69] Glowacka I, Bertram S, Müller MA, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J Virol. 2011;85(9):4122-4134.

[70] Tarnow C, Engels G, Arendt A, et al. TMPRSS2 is a host factor that is essential for pneumotropism and pathogenicity of H7N9 influenza A virus in mice. J Virol. 2014;88(9): 4744-4751.

[71] Matsuyama S, Nao N, Shirato K, et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc Natl Acad Sci U S A. 2020;117(13): 7001-7003.

[72] Asselta R, Paraboschi EM, Mantovani A,et al. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging (Albany NY). 2020;12(11):10087-10098.

[73] Stopsack KH, Mucci LA, Antonarakis ES, et al. TMPRSS2 and COVID-19: Serendipity or Opportunity for Intervention?. Cancer Discov. 2020;10(6):779-782.

[74] Barrett AJ, Rawlings ND, Woessner JF. Handbook of Proteolytic Enzymes Elsevier Academic Press, London 2004.

[75] Kawabata K, Hagio T, Matsuoka S. The role of neutrophil elastase in acute lung injury. Eur J Pharmacol. 2002;451(1):1-10.

[76] Hashimoto S, Okayama Y, Shime N, et al. Neutrophil elastase activity in acute lung injury and respiratory distress syndrome. Respirology. 2008;13(4):581-584.

[77] 77.Didangelos A. COVID-19 Hyperinflammation: What about Neutrophils?. *mSphere*. 2020;5(3): e00367-20.

[78] Vargas-Alarcón G, Posadas-Sánchez R, Ramírez-Bello J. Variability in genes related to SARS-CoV-2 entry into host cells (ACE2, TMPRSS2, TMPRSS11A, ELANE, and CTSL) and its potential use in association studies. Life Sci. 2020;260:118313.

[79] Taniguchi K, Yang P, Jett J, et al. Polymorphisms in the promoter region of the neutrophil elastase gene are associated with lung cancer development. Clin Cancer Res.2002; 8:1115-1120

[80] Park JY, Chen L, Lee J, Sellers T, et al. Polymorphisms in the promoter region of neutrophil elastase gene and lung cancer risk. Lung Cancer.2005; 48:315-321

[81] Kirschke H. L. Cathepsin Handb. Proteolytic Enzym. 2013; pp. 1808-1817,

[82] Elshabrawy HA, Fan J, Haddad CS, et al. Identification of a broad-spectrum antiviral small molecule against severe acute respiratory syndrome coronavirus and Ebola, Hendra, and Nipah viruses by using a novel high-throughput screening assay. J Virol. 2014;88(8):4353-4365.

[83] Bristow CL, Di Meo F, Arnold RR. Specific activity of alpha1proteinase inhibitor and alpha2macroglobulin in human serum: application to insulin-dependent diabetes mellitus. Clin Immunol Immunopathol. 1998;89(3): 247-259.

[84] Chapman KR, Burdon JG, Piitulainen E, et al. Intravenous augmentation treatment and lung density in severe α1 antitrypsin deficiency (RAPID): a randomised, double-blind, placebo-controlled trial. Lancet. 2015;386(9991):360-368.

[85] Stoller JK, Aboussouan LS. Alpha1-antitrypsin deficiency. Lancet. 2005; 365 (9478), 2225-2236

[86] Petrache I, Hajjar J, Campos M. Safety and efficacy of alpha-1-antitrypsin augmentation therapy in the treatment of patients with alpha-1-antitrypsin deficiency. Biologics. 2009;3:193-204.

[87] Gao W, Zhao J, Kim H, et al. α1-Antitrypsin inhibits ischemia reperfusion-induced lung injury by reducing inflammatory response and cell death. J Heart Lung Transplant. 2014;33(3):309-315.

[88] de Serres FJ, Blanco I, Fernández-Bustillo E. Genetic epidemiology of alpha-1 antitrypsin deficiency in North America and Australia/New Zealand: Australia, Canada, New Zealand and the United States of America. Clin Genet. 2003;64(5):382-397.

[89] Hazari YM, Bashir A, Habib M, et al. Alpha-1-antitrypsin deficiency: Genetic variations, clinical manifestations and therapeutic interventions. Mutat Res. 2017;773: 14-25.

[90] Harbig A, Mernberger M, Bittel L, et al. Transcriptome profiling and protease inhibition experiments identify proteases that activate H3N2 influenza A and influenza B viruses in murine airways. J Biol Chem. 2020;295(33): 11388-11407.

[91] Bergin DA, Reeves EP, Meleady P, et al. α-1 Antitrypsin regulates human neutrophil chemotaxis induced by soluble immune complexes and IL-8. J Clin Invest. 2010;120(12):4236-4250.

[92] Ishii T, Doi K, Okamoto K, et al. Neutrophil elastase contributes to acute lung injury induced by bilateral nephrectomy. Am J Pathol. 2010;177(4):1665-1673.

[93] Sodhi CP, Wohlford-Lenane C, Yamaguchi Y, et al. Attenuation of pulmonary ACE2 activity impairs inactivation of des-Arg9 bradykinin/BKB1R axis and facilitates LPS-induced neutrophil infiltration. Am J Physiol Lung Cell Mol Physiol. 2018;314(1):L17-L31.

[94] Gans H, Tan BH. Alpha-1-antitrypsin, an inhibitor for thrombin and plasmin. Clin Chim Acta. 1967;17(1):111-117.

[95] Frenzel E, Korenbaum E, Hegermann J, et al. Does augmentation with alpha1-antitrypsin affect neutrophil extracellular traps formation?. Int J Biol Sci. 2012;8(7): 1023-1025.

[96] Petrache I, Fijalkowska I, Zhen L, et al. A novel antiapoptotic role for alpha1-antitrypsin in the prevention of pulmonary emphysema. Am J Respir Crit Care Med. 2006;173(11):1222-1228.

[97] Dutta AK, Goswami K. Host genomics of COVID-19: Evidence point towards Alpha 1 antitrypsin deficiency as a putative risk factor for higher mortality rate. Med Hypotheses. 2021;147:110485.

[98] Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun. 2017;8(1):599.

[99] Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe. 2020;27(6): 992-1000.e3.

[100] Kanjanaumporn J, Aeumjaturapat S, Snidvongs K, et al. Smell and taste dysfunction in patients with SARS-CoV-2 infection: A review of epidemiology, pathogenesis, prognosis, and treatment options. Asian Pac J Allergy Immunol. 2020;38(2):69-77.

[101] Ehlers A, Beck S, Forbes SA, et al. MHC-linked olfactory receptor loci

exhibit polymorphism and contribute to extended HLA/OR-haplotypes. Genome Res. 2000;10(12):1968-1978.

[102] Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013;14(6):415-426.

[103] Kerget F, Kerget B, Kahraman ÇY, et al. Evaluation of the relationship between pentraxin 3 (PTX3) rs2305619 (281A/G) and rs1840680 (1449A/G) polymorphisms and the clinical course of COVID-19 [published online ahead of print, 2021 Jul 27]. J Med Virol.

[104] Refae S, Gal J, Ebran N, et al. Germinal Immunogenetics predict treatment outcome for PD-1/PD-L1 checkpoint inhibitors [published correction appears in Invest New Drugs. 2021 Feb;39(1):287-292].

[105] Alghamdi J, Alaamery M, Barhoumi T, et al. Interferon-induced transmembrane protein-3 genetic variant rs12252 is associated with COVID-19 mortality. Genomics. 2021;113(4):1733-1741.

[106] Ponti G, Pastorino L, Manfredini M, et al. COVID-19 spreading across world correlates with C677T allele of the methylene tetrahydrofolate reductase (MTHFR) gene prevalence. J Clin Lab Anal. 2021;35(7):e23798.

[107] Posadas-Sánchez R, Sánchez-Muñoz F, Guzmán-Martín CA, et al. Dipeptidylpeptidase-4 levels and DPP4 gene polymorphisms in patients with COVID-19. Association with disease and with severity. Life Sci. 2021;276:119410.

[108] Naemi FMA, Al-Adwani S, Al-Khatabi H, et al. Association between the HLA genotype and the severity of COVID-19 infection among South Asians. J Med Virol. 2021;93(7): 4430-4437.

Chapter 8

Characterization, Comparative, and Phylogenetic Analyses of Retrotransposons in Diverse Plant Genomes

Aloysius Brown, Orlex B. Yllano, Leilani D. Arce, Ephraim A. Evangelista, Ferdinand A. Esplana, Lester Harris R. Catolico and Merbeth Christine L. Pedro

Abstract

Retrotransposons are transposable elements that use reverse transcriptase as an intermediate to copy and paste themselves into a genome via transcription. The presence of retrotransposons is ubiquitous in the genomes of eukaryotic organisms. This study analyzed the structures and determined the comparative distributions and relatedness of retrotransposons across diverse orders (34) and families (58) of kingdom Plantae. *In silico* analyses were conducted on 134 plant retrotransposon sequences using ClustalW, EMBOSS Transeq, Motif Finder, and MEGA X. So far, the analysis of these plant retrotransposons showed a significant genomic relationship among bryophytes and angiosperms (216), bryophytes and gymnosperms (75), pteridophytes and angiosperms (35), pteridophytes and gymnosperms (28), and gymnosperms and angiosperms (70). There were 13 homologous plant retrotransposons, 30 conserved domains, motifs (reverse transcriptase, integrase, and gag domains), and nine significant phylogenetic lineages identified. This study provided comprehensive information on the structures, motifs, domains, and phylogenetic relationships of retrotransposons across diverse orders and families of kingdom Plantae. The ubiquitousness of retrotransposons across diverse taxa makes it an excellent molecular marker to better understand the complexity and dynamics of plant genomes.

Keywords: transposable elements, retrotransposon, genetic polymorphism, phylogenetic analysis, genome

1. Introduction

Retrotransposons can move within genomes due to their highly effective transposition mechanism. Because of this high level of transposition, their presence is a significant feature of plant genomes and other eukaryotic organisms. Since the discovery of transposable elements (TE) by Barbara McClintock more than seven decades ago, there have been several challenges in studying the structures of retrotransposons due to their repetitive structure, diversity in form, their large

number in a genome, and their ability to replicate so frequently [1]. Even studying closely related genomes does not overcome this problem since retrotransposons also tend to be highly species-specific, a trait that makes them difficult to classify. Research has shown that they are not merely transient components of a genome but are instrumental in genomic development and adaptation, influencing these genomes from how chromosomes are structured to helping activate certain genes under certain conditions [2]. The interaction of retrotransposons with a host genome is not a simple one. Pieces of evidence have shown that they have helped shaped genomes for an extended period. In some cases, this has imparted important genetic traits to their host organisms. In others, they have been linked to mutagenesis and disease, prompting their host to develop regulatory safeguards to suppress and limit their activities [3].

Recent advances in sequencing technologies have come a long way in helping unravel the structure of plant genomes. Plant genomes are some of the most complex and diverse among known eukaryotic kingdoms [4] and vary widely in size across kingdom Plantae, with the smallest genomes sequenced so far being from green algae species [5] and the largest being *Pinus taeda*, which is around 22 Gbp in length [6]. A significant portion of the plant genome comprises transposable elements, the so-called "jumping genes" [7]. The diversity and size variation across plant genomes is primarily attributed to the activity of these transposable elements [8]. The transposable elements are known to have viral origins; in particular, retrotransposons structures closely resemble retroviruses without the gene for the viral envelope or with a nonfunctional envelope gene. It is hypothesized that transposable elements enter the genomes of eukaryotes through infection by ancient viruses and remained as parasitic elements in their host genomes [9]. More studies are needed to understand better the complexity of plant retrotransposons and unravel its salient features.

1.1 Classes and types of transposable elements

The complexity and diversity of transposable elements coupled with the availability of recent genomic sequences in the genebanks have generated various groupings of TEs. However, concerted efforts have been made to come up with a generally accepted and unified nomenclature. The replication process employed by transposable elements are used to classify them into two large groups [10]. Retrotransposons or Class I transposable elements use the enzyme reverse transcriptase to copy and paste themselves in the genome and are the most abundant type in plant genomes. DNA transposons or Class II transposable elements use other enzymes, including DNA polymerase and transposase, to copy and insert themselves into genomes [11]. This copy and paste mechanism is responsible for the significant number of transposable elements in eukaryotic genomes.

Class I Transposable Elements or Retrotransposons consists of the long terminal repeats (LTRs) retrotransposons and the non-long terminal repeats (non-LTRs) retrotransposons. These LTR retrotransposons and non-LTR retrotransposons are further subdivided based on their dynamics in the genome. The autonomous retrotransposons can be independently mobile, while the nonautonomous retrotransposons necessitate the presence of TEs for their movement. Some of the LTR retrotransposons in eukaryotes include Gypsy, Copia, BEL, DIRS, ERVI, ERV2, and ERV3 superfamilies. In contrast, superfamilies of non-LTR retrotransposons includes SINE1,2,3, LINES, CR1, CRE, I, RTE, TX1, Jockey, Penelope, R2, R4, RandI, Rex1, L1, and NeSL [12, 13].

A less well-studied class of retrotransposons in plant genomes are non-LTR retrotransposons. These are the LINEs-Long Interspersed Nuclear Elements and

the SINEs-Short Interspersed Nuclear Elements. They do not exhibit much activity in plant genomes and constitute around 33.5% or about one third of the human genome [13]. More so, they contribute to new insertions in the human genome and have been linked to mutagenesis and human diseases [14].

LINEs are considered the oldest class of retrotransposons in plant genomes. Evidence suggests that they are highly regulated or inactive since their transcription is rarely observed in plant genomes [15]. In contrast, studies have shown that the ancient activity of SINEs helped shaped the genomic diversification of some monocot species [16] and the heterogeneity of many eukaryotic genomes, but apart from this, little is known so far of their activity in plant genomes [17]. With this, there is a need to study and characterize the diverse retrotransposons and understand how and to what extent they influence changes in a host genome.

1.2 Characterization of retrotransposons

The presence of transposable elements in an organism has many implications for its genomic activity. Depending on the region of the chromosome they are located on, they may affect what type of genes are expressed in the genome and the functions of these genes [18]. Gypsy retrotransposons have a widespread and more diverse position on the chromosomes in plant genomes, while Copia retrotransposons tend to cluster in proximal regions of the chromosomes they are located on [19]. However, it is worth pointing out that LTR retrotransposons tend to group in different chromosomal regions regardless of their lineages [20]. Research into plant genomic structures has yielded valuable insight into the characterization of retrotransposons due to their ubiquitous presence in plant genomes [21]. They are subclassified into LINES and SINES [22]. The LTR-retrotransposons are further classified into "superfamilies" based on their genetic sequences, namely, the Copia superfamily, the Gypsy superfamily, Bel-Pao, retrovirus, and endogenous retrovirus superfamilies [23]. Of these, the most widespread in plant genomes and the most well studied are the Gypsy and Copia superfamilies. Gypsy retrotransposons are differentiated from Copia retrotransposons by the position of the integrase protein in their genetic sequence. In gypsy retrotransposons, integrase is situated after the reverse transcriptase in the genetic sequence and before the reverse transcriptase in Copia retrotransposons [24]. Phylogenetic analyses and time of divergence are used to further divide these superfamilies into different lineages. The Copia superfamily comprised TORK, Bianca, Ale, Maximus lineages Gypsy superfamily of Attila, CRM, Del, and Galadriel lineages [25]. LTR-retrotransposons showcase such variety in number, position, and distribution in their host genome due to their unique ability to express the independent activity and replicate themselves numerous times on chromosomes [26].

A key feature of LTR retrotransposons and the structure that gives them their name is the presence of two homologous structures called long terminal repeats at both ends of their genetic sequence. These DNA sequences can vary in size from a hundred bps to thousands of bps [27]. These LTRs are non-coding regions that bracket the internal coding regions and are also a component of retroviral sequences [28]. LTR retrotransposons vary widely in size and functional characteristics. In plants, they have been documented as short as four kbp in *Helianthus* species [29] to over 23 kbp in *Populus trichocarpa* [30]. The structures of LTR retrotransposons are organized into one or several Open Reading Frames (ORF) [31]. The ORFs contains genetic information for the pol and gag genes and are integral to transcription in the host genome [32]. Like their retroviral counterparts, the gag genes encode functional polyproteins, and the pol gene usually contains the reverse transcriptase. These genes are typically separated by stop codons [33]. The pol gene encodes three

important proteins, each of which has a crucial role in retrotransposal replication in the genome [34]. These proteins are Integrase, Protease, and Reverse Transcriptase [35]. Because retrotransposons replicate similarly to viruses, and their replication can lead to mutations and disrupt DNA repair, there are genomic mechanisms in place to silence their activity [36]. To escape this silencing, LTR retrotransposons may possess another region called the chromodomain. One mechanism the cell uses to silence retrotransposons is the formation of heterochromatin near areas of retrotransposon activity [37]. The presence of heterochromatin makes it difficult for the retrotransposon proteins to access the cell DNA, suppressing replication [38]. The chromodomain region encodes a protein that helps the retrotransposon escape silencing by manipulating these heterochromatins. Chromodomains are found upstream of the 3′ end of the genetic sequence in retrotransposons [39].

1.3 Mechanism of action

Retrotransposons insert and reinsert themselves in a host genome by transcription. This process is accomplished by the reverse transcription of an RNA intermediate transcript. This transcript is the template that is used to generate new copies of the retrotransposon [40]. The reverse transcription of retrotransposons is a complex procedure. In LTR retrotransposon, the process is helped by the long terminal repeats at each end of their structure that acts as start sites for replicating the internal region. The replication of this internal region occurs in opposite directions to produce two DNA strands. At the 3′ end, tRNA binds to the initiation site of the left LTR and replicates one of the two DNA strands. At the right LTR, a Polypurine Tract, which acts as a primer, binds immediately upstream of this region and replicates the second of the two DNA strands [41].

The mRNA template is synthesized first in the replication of retrotransposons. This mRNA template is then translated into proteins utilized in the process. The mRNA template has a U region and a short repeat sequence at each end. tRNA acts as a primer and binds to a primer binding site on the mRNA. This initiates the production of minus (−) strand DNA through the catalyzation of Reverse Transcriptase. The synthesized DNA reaches the U5 region at the 5′ end of the template and pairs with the repeat sequence at the 3′ end of the genomic RNA. Once synthesis of this first DNA strand is complete, the enzyme RNase H deteriorates the genomic RNA template, leaving only fragments. These fragments then prime the synthesis of the second DNA strand. As with the first strand, Reverse Transcriptase synthesizes another DNA strand but uses the first DNA strand as a template. At the end of this process, a linear double-stranded DNA is made with an LTR region (comprised of the repeat sequence, U5, and U3 regions) at each end. The enzyme integrase then inserts this new retrotransposon DNA into the host chromosomal DNA by using the 3' OH of each strand to integrate at target sites a few base pairs apart in the genome [42].

1.4 Role of retrotransposons

Retrotransposons are known to be major drivers of genomic diversity and homogeneity during the development of eukaryotic genomes. Presently, their activity in plant genomes is regulated by different mechanisms, but they are still capable of bursts of activity when reactivated by mutations, adjacent gene expression, or environmental factors [43]. Grandbastien [44] has noted that all the retrotransposons that are known to be active in plant genomes are usually dormant during their host development but become active in response to environmental stressors. This could be linked to retrotransposons being proliferators of genomic diversity

since their activation by stresses induces survival genes to turn on. The study by Hilbricht et al. [45] on *Craterostigma plantagineum* dehydration led to the isolation and identification of a retroelement gene, the Craterostigma desiccation-tolerant (CDT-1) gene, that is turned on by dehydration and imparts drought-resistant properties to the plant. This is also in line with Zhao et al. [46], which found a potential link of the OAR1 gene to the tolerance of osmotic and alkaline stresses in *Arabidopsis thaliana*. Though often characterized by their propensity to initiate mutagenesis, retrotransposons have been shown to affect the expression of genes they are adjacent to in the genome and even help regulate the structure of centromeres [47], as noted in an investigation of maize species by Gao et al. [48]. Analysis of tomato plants demonstrated that differences in volatile esters between two different colored fruits of different species of these plants are linked to the placement of retrotransposons near the family of esterases that exhibits a high level of enzyme activity. This placement results in a higher expression of the esterase, resulting in the reduced levels of multiple esters [49]. Retrotransposons have also been linked to disease resistance in plants. A study showed that activation of athila LTR retrotransposons led to genome expansion in *Capsicum baccatum* by increasing the number of a disease-resistant gene family [50] and analysis of *Phaeodactylum tricornutum* cells showed the activity of LTR-retrotransposon initiate a plant response to a decrease in nitrate and when exposed to reactive aldehydes that stress diatoms and leads to cell death [51]. Analysis of retrotransposon families in sorghum species shows that their activity influences genomic adaptation and diversity [52]. This finding suggests that retrotransposons play vital roles in regulating genes that encode functional proteins [53]. A study of Thale Cress and Adzuki bean seedlings treated with the DNA methylation inhibitor zebularine increased activity and accumulation of the retrotransposon ONSEN in the seedlings treated than in the control seedlings [54]. These studies point to the pivotal role of retrotransposons in plants' adaptation to their environment and their contribution to genomic diversity.

This study compared, characterized, identified shared patterns, and determined the relationships of different retrotransposons across diverse plant taxa.

2. Materials and methods

To assemble the plant retrotransposon library, we collected genomic DNA sequences deposited at the National Center for Biotechnology Information (NCBI) nucleotide database. These were then further sorted to include only sequences with 300 to 800 base pairs in length. In total, 134 retrotransposon sequences were selected and analyzed in this study. Of these, 54 were angiosperms, 46 were gymnosperms, 11 were pteridophytes, three were liverworts, and 20 were bryophytes. The sequences were downloaded in the FASTA format and saved in a text document for further analyses. To study the characteristics of the plant retrotransposon sequences and identify homogeny, multiple sequence alignment (ClustalW) program was utilized. The parameters of the ClustalW analysis were defined as follows: Pairwise Alignment was set to slow and accurate for DNA sequences only. The Gap Open Penalty was set to 15 and the Gap Extension Penalty to 6.66. The Weight Matrix used was the International Union of Biochemistry (IUB) matrix for DNA sequences. These same parameters were used for the multiple sequence analysis with hydrophilic gaps included in the computation.

Motif analyses were performed on the plant retrotransposon sequences to identify motifs, protein domains, and conserved domains. The nucleotide sequences were translated into their corresponding amino acid (aa) sequences with the EMBOSS Transeq tool developed by the European Bioinformatics Institute. The algorithm was

set to translate the nucleotide sequences into the three possible reading frames using the standard codon table. The translated aa sequences were then analyzed for protein domains, families, and functional sites using the PROSITE tool developed by the Swiss Institute of Bioinformatics [55] and the MOTIF Finder program of the Kyoto University Bioinformatics Center [56]. All three reading frames were analyzed to ensure the proper frame would be used for motif identification. The aligned retrotransposon sequences were analyzed using the MEGA-X. The software was used to construct a maximum likelihood phylogenetic tree with the Tamura-Nei method used to account for the substitution rate differences between nucleotides and the inequality of nucleotide frequencies. The Nearest-Neighbor-Interchange was used as the heuristic method to improve the likelihood of the tree. The phylogenetic tree generated by the MEGA-X program was then modified in the MEGA X Tree Topology Editor to produce a circular phylogenetic diagram for better data visualization.

3. Results and discussion

3.1 Multiple sequence alignment

Figure 1 shows the alignment scores of sequences produced from the multiple sequence alignment analysis performed in the clustalW program. These scores represent the pairwise alignment between each pair of retrotransposon sequences. The cutoff alignment score was set at 50 percent identity between two aligned sequences.

In total, there were 870 pairwise alignments with a 50 to 100 percent alignment score. Fifty-five percent (476) of the alignments had a percent identity in the range of 50 to 59. Thirty-two percent (281) had a percent identity in the range of 60 to 69. Seven percent (65) had a percent identity in the range of 70 to 79, 4% (35) had a percent identity in the range of 80 to 89, and 2% (13) had a percent identity in the range of 90 to 100. The multiple sequence alignment scores of 40% and higher are considered significant. However, an alignment less than 40% is considered too divergent [57]. The alignment score for this multiple sequence analysis was set to 50% to include only highly significant alignments.

3.2 Identification of homologous sequences

Table 1 contains the aligned sequences with the highest alignment score. There is a diversity in the relationship of these sequences. *T. pellucida* 1 to *T. pellucida* 2

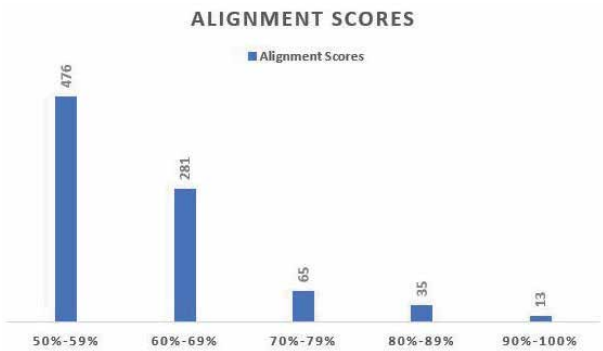


Figure 1.
Significant pairwise alignment scores of 134 plant retrotransposon sequences.

Sequences Aligned	Aligned Score
A. concolor: *A. veitchii*	90
L. saxicola: *P. schreberi*	91
S. cooperi: *D. truncatula*	91
D. polysetum: *L. glaucum*	93
P. cuspidatum: *R. canescens*	93
L. gmelinii: *L. czekanowskii*	94
A. araucana: *A. brownii*	94
A. sativa: *A. sterilis*	94
A. ipaensis: *A. hypogaea*	95
P. patens: *M. polymorpha 1*	99
*T. pellucida*1: *T. pellucida 2*	100
V. dubyana: *F. antipyretica*	100
N. tetragona: *M. grandiflora 2*	100

Table 1.
Aligned sequences with an alignment score of 90 to 100.

are of the same species but clones. Each plant in the sequence pairs alignments of *A. concolor* to *A. veitchii*, *L. gmelinii* to *L. czekanowskii*, *A. sativa* to *A. sterilis*, *A. ipaensis* to *A. hypogaea*, and *V. dubyana* to *F. antipyretica*, belong to the same genus. *A. araucana* and *A. brownii* belong to the same family. The sequences aligned in each alignment pair of *L. saxicola* to *P. schreberi* and *D. polysetum* to *L. glaucum* belong to the same order, while those in the pairs of *S. cooperi* to *D. truncatula* and *P. cuspidatum* to *R. canescens* belong to the same class. Sequences belonging to only the same division can be closely related, as in the case of *P. patens* to *M. polymorpha* with a 99% identity and *N. tetragona* to *M. grandflora* with a 100% identity. The pair of sequences aligned in the same genus had the highest number of aligned pairs.

The results above confirm the highly conserved nature of retrotransposons. This was supported by the study of retrotransposons in mammals [58]. Despite their enormous size and diversity, it has been noted that similar retrotransposons tend to cluster together in similar genomes of hosts belonging to the same order, family, or class [59]. Specific types of retrotransposons belonging to the same family or lineage can be conserved across a particular kingdom or division [60]. The presence of homologs can be inferred from these aligned sequences considering their high percent identity and their distribution to different species [61]. An alignment of 90 and higher was used as the cutoff value for homolog identification [62].

3.3 Conservation of retrotransposons

Table 2 is a summation of retrotransposons sequences with an alignment score of 80 to 89. This is the pairwise alignment score between pairs of sequences.

Aligned sequence pairs in the same genus were: *L. occidentalis* to *L. sibirica*, *A. concolor* to *A. balsamea*, *L. occidentalis* to *L. kaempferia*, *P. rubens* to *P. schrenkiana*, *A. veitchii* to *A. balsamea*, and *L. kaempferi* to *L. sibirica*. More so, the aligned sequences pairs that had sequences in the same family were: *L. sibirica* to *P. rubens*, *L. sibirica* to *P. schrenkiana*, *L. occidentalis* to *P. contorta*, *P. contorta* to *L. sibirica*, *L. occidentalis* to *P. schrenkiana*, *P. contorta* to *P. schrenkiana*, *L. kaempferi* to *P. rubens*, *P. contorta* to *L. kaempferi*, *P. contorta* to *P. reubens*, *L. occidentalis* to *P. rubens*, and *L. kaempferi* to *P. schrenkiana*. At the same order level, the following were the aligned

Sequences Aligned	Aligned Score	Sequences Aligned	Aligned Score
L. saxicola: *P. polyantha*	80	*P. contorta*: *L. sibirica*	86
L. saxicola: *D. polysetum*	80	*S. obtusum*: *A. rupestris*	87
D. polysetum: *R. canescens*	80	*L. occidentalis*: *L. kaempferi*	87
L. glaucum: *P. cuspidatum*	80	*L. occidentalis*: *P. schrenkiana*	87
P. polyantha: *L. glaucum*	81	*P. contorta*: *P. schrenkiana*	87
P. polyantha: *P. cuspidatum*	81	*L. kaempferi*: *P. rubens*	87
P. polyantha: *R. canescens*	81	*P. rubens*: *P. schrenkiana*	87
P. polyantha: *D. polysetum*	82	*J. communis*: *T. baccata*	87
P. polyantha: *H. ciliata*	82	*S. cooperi*: *N. exaltata*	88
D. polysetum: *P. cuspidatum*	82	*D. truncatula*: *N. exaltata*	88
G. biloba2: *P. rubens*	82	*P. contorta*: *L. kaempferi*	88
P. schreberi: *P. polyantha*	83	*P. contorta*: *P. rubens*	88
G. biloba2: *P. contorta*	83	*A. veitchii*: *A. balsamea*	88
L. occidentalis: *L. sibirica*	83	*L. occidentalis*: *P. rubens*	89
L. sibirica: *P. rubens*	84	*L. kaempferi*: *L. sibirica*	89
L. sibirica: *P. schrenkiana*	84	*L. kaempferi*: *P. schrenkiana*	89
G. biloba2: *P. schrenkiana*	85		
A. concolor: *A. balsamea*	85		
L. occidentalis: *P. contorta*	86		

Table 2.
Aligned sequences with an alignment score of 80 to 89.

sequence pairs: *L. saxicola* to *P. polyantha*, *J. communis* to *T. baccata*, and *D. tuncatula* to *N. exaltata*. Aligned sequence pairs that had sequences in the same class were: *L. saxicola* to *D. polysetum*, *D. polysetum* to *R. canescens*, *L. glaucum* to *P. cuspidatum*, *P. polyantha* to *L. glaucum*, *P. polyantha* to *P. cuspidatum*, *P. polyantha* to *R. canescens*, *P. polyantha* to *D. polysetum*, *P. polyantha* to *H. ciliata*, *D. polysetum* to *P. cuspidatum*, *P. schreberi* to *P. polyantha* and *S. cooperi* to *N. exaltata*. Likewise, the aligned sequence pairs with sequences in the same division were: *G. biloba* to *P. schrenkiana*, *G. biloba* to *P. contorta*, *G. biloba* to *P. rubens*, and *S. obtusum* to *A. rupestris*.

3.4 Motifs and domains

Molecular characterization is important in understanding the nature of any genetic element and its insertion origin in a genome. Molecular characterization provides a detailed description of the structure of a genetic sequence, changes that it induces in a genome, and how it affects genetic expression [63]. Characterization is an important feature in the study of retrotransposons. It is also used for classifying retrotransposons [64], uncovering their associations in a genome [65, 66], and discovering new types of retrotransposons (**Table 3**) [66].

The identification of the reverse transcriptase motif in these retrotransposon sequences is significant because it is not only integral to the replication process of retrotransposons but is one of the most significant parts of their structure [67]. The reverse transcriptase type identified in these sequences was only found in LTR retrotransposons and retroviruses. The presence of this reverse transcriptase type

Reverse transcriptase (RNA-dependent DNA polymerase)	**Simian taste bud-specific gene product family**
Reverse transcriptase (RNA-dependent DNA polymerase)	Simian taste bud-specific gene product family
Tsi6	BAFF-R, TALL-1 binding
RNase H-like domain found in reverse transcriptase	Zinc knuckle
Tc5 transposase DNA-binding domain	GAG-polyprotein viral zinc-finger
Peptidase propeptide and YPEB domain	Mis6
Integrase zinc-binding domain	Protein prenyltransferase alpha subunit repeat
Integrase core domain	Chromatin remodeling factor Mit1 C-terminal Zn finger 2
H_2C_2 zinc finger	5′-3′ exonuclease, N-terminal resolvase-like domain
gag-polypeptide of LTR copia-type	Retrotransposon gag protein
Aspartyl protease	C2H2 zinc-finger
gag-polyprotein putative aspartyl protease	GAG-pre-integrase domain
Retroviral aspartyl protease	Eukaryotic translation initiation factor 3 subunit G
Domain of unknown function	3′ exoribonuclease family, domain 2
Putative peptidase (DUF1758)	HicA toxin of bacterial toxin-antitoxin,
Fimbrial assembly protein (PilN)	BRK domain

Table 3.
Motifs and domains identified by the MOTIF finder.

usually indicates that the sequence is a retrotransposon mobile element or a retrovirus [68]. Reverse transcriptase gene identification could be used to identify retrotransposon sequences due to their high specificity. Reverse transcriptases are known to be multidomain enzymes, with notable domains being the catalytic domain and the RNase H domain [69]. The Tc5 transposase DNA-binding domain is a structural motif found in many proteins that regulate gene expression. The RNase H-like domain found in these retrotransposon sequences belongs to a reverse transcriptase subfamily that shares sequence similarity with reverse transcriptases from endogenous retroviruses of the zebrafish and the Moloney mouse leukemia retroviruses [69, 70]. This finding strengthens the viral origins of retrotransposons in eukaryotes.

The presence of the zinc-binding domain indicates the presence of integrase since it is one of the domains in the integrase enzyme. Integrase allows retroviruses and retroelements to insert their DNA into a host genome [71]. The integrase core domain that was also detected in this sequence is one of the three known domains of the integrase enzyme. It is the catalytic domain that catalyzes the transfer of retroviral or retrotransposal DNA made by reverse transcriptase to the site in the genome where it will be inserted [72]. GAG-Pre-Integrase domain lies upstream of the integrase region in retroviral polyproteins. They are usually connected to elements that assist in retroviral insertion [73].

The Copia family of retrotransposons is a large retrotransposon family active in the genomes of plants. It is classified under the long terminal repeats retrotransposons along with the Gypsy family [74]. The GAG Polypeptide of the LTR-Copia type domain is highly conserved and found only in Copia retrotransposons [75]. This domain was identified in seven species: *G. biloba*, *L. occidentalis*, *P. contorta*, *L. kaempferi*, *L. sibirica*, *P. rubens*, *P. schrenkiana*, definitively identifying them as Copia family retrotransposons.

Some domains were identified that are not generally associated with retrotransposons. The Hic A toxin functions as an mRNA interferase in bacteria and archaea species [76], Tsi6 is a bacterial immunity protein, and the Fimbrial Assembly Protein functions in the production of bacterial fimbria used for cellular attachment [77]. The Simian taste-bud specific gene is found in primates, and mutations of this gene have been linked to follicular lymphomas [78]. The Mis6 protein is integral for chromosome segregation during mitosis, and the protein prenyltransferase alpha subunit repeat functions in protein prenylation. In contrast, the eukaryotic translation initiation factor 3 subunit G initiates protein synthesis [79], and the BAFF-R is a polypeptide that binds to the ligands of TALL-1, a tumor necrosis factor that initiates inflammation in humans [80]. Zinger finger proteins are a large family of proteins noted for their role as transcription factors and their ability to bind Zn ions. Several of these protein types were identified from the plant retrotransposon sequences, including: H2C2 zinc finger, zinc knuckle, GAG-polyprotein viral zinc-finger, chromatin remodeling factor Mit1 C-terminal Zn finger 2, and C2H2 zinc-finger. Recent studies revealed that they are highly involved in regulating plant response to abiotic stressors in their environment [81]. Peptidase propeptide and YPEB domain, putative peptidase (DUF1758), 5′-3′ exonuclease, N-terminal resolvase-like domain, and the BRK domain are all hypothetical proteins of which little to nothing is known of their activity presently [82].

3.5 Patterns and profiles

The PROSITE database has an extensive collection of protein families, subfamilies, domains, and motifs managed by the Swiss Institute of Bioinformatics [83]. The database is organized into unique protein profiles and patterns to identify functional sites, domains, and protein families [84].

Table 4 contains the PROSITE patterns of four motifs found in the PROSITE database. IPNS_1 was found in *E. arvense*, ASP_PROTEASE in *G. biloba*, ZINC_PROTEASE in *P. contorta*, and TONB_DEPENDENT REC 1 in *T. aestivum*. Isopenicillin N synthase signature 1 is an enzyme found in bacterial and fungal species instrumental in the production of cephalosporin and penicillin [85]. TonB-dependent receptor proteins signature 1 is a type of protein found in *E. coli* involved in cellular transportation of substrates into the periplasmic space by active transport [86]. The presence of these bacterial domains in plant retrotransposons supports their role as genetic reservoirs. Because of their transposable nature, they can "jump" from bacterial plasmids onto chromosomes, carrying genes with them [87].

Aspartyl proteases are a family of enzymes that hydrolyzes peptide bonds [88]. They are very diverse and can be found in species including humans, retroviruses, plants, and fungi. In retroviruses, they are usually encoded in the pol gene as part of a polypeptide [89]. The zinc protease utilizes zinc in its catalytic function to break down polyproteins. Retrotransposon's polyproteins are very important elements

Found Motif	Description
IPNS_1	PS00185, Isopenicillin N synthase signature 1
ASP_PROTEASE	PS00141, Eukaryotic and viral aspartyl proteases active site
ZINC_PROTEASE	PS00142, Neutral zinc metallopeptidases, zinc-binding region signature
TONB_DEPENDENT_REC_1	PS00430, TonB-dependent receptor proteins signature 1

Table 4.
Patterns identified from plant retrotransposons.

of their replication mechanism, and these proteases enable the hydrolysis of these larger proteins into smaller functional polypeptides [90]. The Pol polyproteins and proteases are needed in retrotransposon replication to form mRNA and its packaging in the transposition of retrotransposons [91].

Table 5 contains the four PROSITE profiles identified in the retrotransposon sequences. The Reverse Transcriptase catalytic domain profile was detected in 25 different species, the Integrase catalytic domain profile in four species, and the zinc finger CCHC-type profile, and the zinc finger SWIM-type profile in one species each. Reverse Transcriptase is a multidomain enzyme consisting of two domains: The Catalytic Domain and the RNase H binding domain. These two domains are used to perform the three enzymatic actions of Reverse Transcriptase [92]. The Catalytic Domain carries out the polymerase activities using DNA-dependent polymerase and RNA-dependent polymerase. The RNase H domain is responsible for the ribonuclease enzymatic activity [93]. Together, these two reverse transcriptase domains enable the "copy" part of the retrotransposon replication mechanism.

The integrase is also a multidomain enzyme (**Table 5**). Its structure consists of three domains integral to its function: An N-terminal zinc finger domain, a C-terminal DNA binding domain, and the Integrase core domain between them [94]. These integrase domains are responsible for the "paste" part of retrotransposon replication, allowing them to transpose themselves into other sites of their host genome [95]. The CCHC zinc finger is associated with retroviruses. They are found in the capsid protein and aids the virus in host infection [96]. The presence of this protein confirms the relationship between retroviruses and retrotransposons. They have developed from retroviruses and still retain proteins for the viral capsids and envelopes [97]. These proteins have been repurposed from aiding in viral infection to assisting in DNA and RNA binding [98].

The SWIM-type zinc finger was isolated from a retrotransposon sequence of *Manihot esculenta* (**Table 5**). The SWIM zinc finger is found in all major eukaryotic groups. It has a strong association with the plant MuDR family of transposases. These enzymes belong to the MuDR transposon, a part of one of the largest families of transposons in plants, the Mu family [99]. They are known mutagens, which is in line with one of the characteristics of transposable elements as instigators of mutagenesis in their host genomes [100].

3.6 Phylogenetic analysis

The phylogenetic analysis uses characters like nucleotide or amino acid sequences to construct a tree to show the relationship among different taxa at the molecular level. This analysis can also investigate domain relationships within an individual taxon [101], and this has become an essential tool for comparing genetic data between different species and groups [102].

Found Motif	Description
RT_POL	PS50878, Reverse transcriptase (RT) catalytic domain profile
INTEGRASE	PS50994, Integrase catalytic domain profile
ZF_CCHC	PS50158, Zinc finger CCHC-type profile
ZF_SWIM	PS50966, Zinc finger SWIM-type profile

Table 5.
Profiles identified from plant retrotransposons.

The history of these retrotransposons was analyzed and created using the Maximum Likelihood method and Tamura-Nei model [103]. The initial tree and guide tree for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Tamura-Nei model. All the codon positions included were 1st+2nd+3rd+noncoding translated proteins. The final dataset consisted of 892 positions. The MEGA X program was used to investigate relationship analyses [104]. The neighbor-joining tree algorithm was tested with bootstrap replicates of 1000 [105] and the resulting bootstrap values displayed above the tree's nodes. The cutoff value for the tree branches was set at 70% [106] to identify lineage clusters. The largest of these clusters with values above the cutoff is the group "C," which contained well-supported branches of retrotransposon lineages. All the plant sequences in this group were from bryophytes. Well-supported groups were group "B" (*M. grandiflora 1* and *M. polymorpha 2*), group "E" (*A. sativa* and *A. sterilis*), group "F" (*S. cooperi* and *D. truncatula*), group "G" (*M. esculenta* and *F. virosa*), and group "I" (*N. tetragona* and *M. grandiflora 2*) (**Figure 2**). Likewise, moderately supported groups (**Figure 3**) were group "A" (*M. polymorpha 3* and *M. notabilis*), group "D" (*V. speciosa 2* and *B. papyrifera*), and group "H" (*P. patens 2* and *L. lagopus 2*) [107].

Figure 4 shows the circular ideogram of diverse retrotransposons across range-wide orders and families of the kingdom Plantae. This ideogram was constructed to ensure holistic visualization of large-scale data and efficiently visualize enormous amounts of genomic information.

The "red" group on the upper right was represented by a cluster of retrotransposons from gymnosperms, while the "blue" group had retrotransposons originating from angiosperm. The "green" group had two novel retrotransposons, namely, Silava and Romani, distinct for gymnosperms. The "yellow" group comprises Gypsy family retrotransposons from angiosperms except for M. polymorpha and *P. massoniana*, a liverwort and gymnosperm, respectively. The "orange" group is the largest cluster composed of Gypsy family retrotransposons from the bryophytes. The "purple" group is a clade of two gymnosperm retrotransposons from the Gypsy and Copia families. In contrast, the "brown" group is a clade of two gymnosperms Copia retrotransposons, and the "pacific blue" group is a clade of non-LTR retrotransposons from two eudicots. The "ruby" group is a cluster of Copia family retrotransposons, and the "Davidson orange" group comprises mostly Gypsy retrotransposons with some notable novel-type families (Cereba, N1, Osr30, and Silava). Osr30 is distinct to *O. sativa*, Cereba to cereal plants, and Silava to gymnosperms. The "pink" group is a cluster of angiosperm Gypsy retrotransposons, the "medium green" is a cluster of giant ferns Cassandra retrotransposons, and the "tyrian purple" is a cluster of Poaceae family retrotransposons. The "lochmara blue" is a cluster of Copia-like retrotransposons, and the "deep red" group is a cluster of

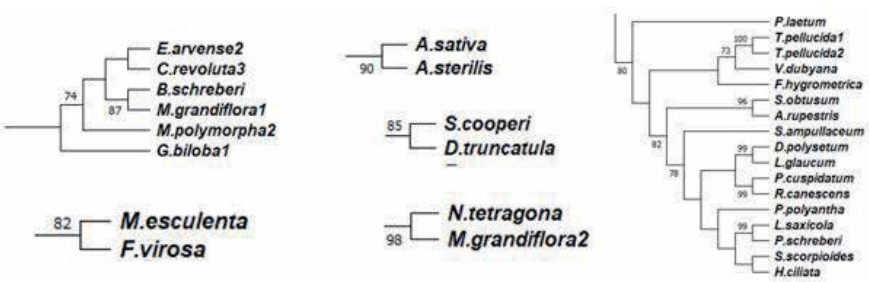


Figure 2.
Well supported bootstrap branches based on the phylogenetic analysis.

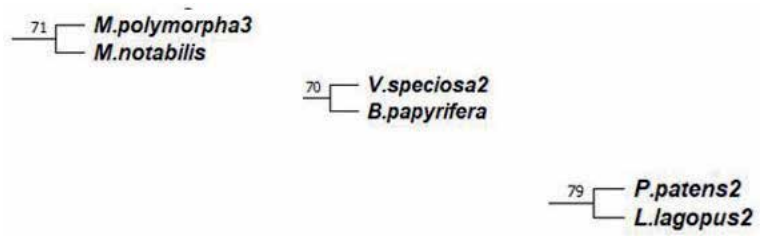


Figure 3.
Moderately supported bootstrap branches based on the phylogenetic analysis.

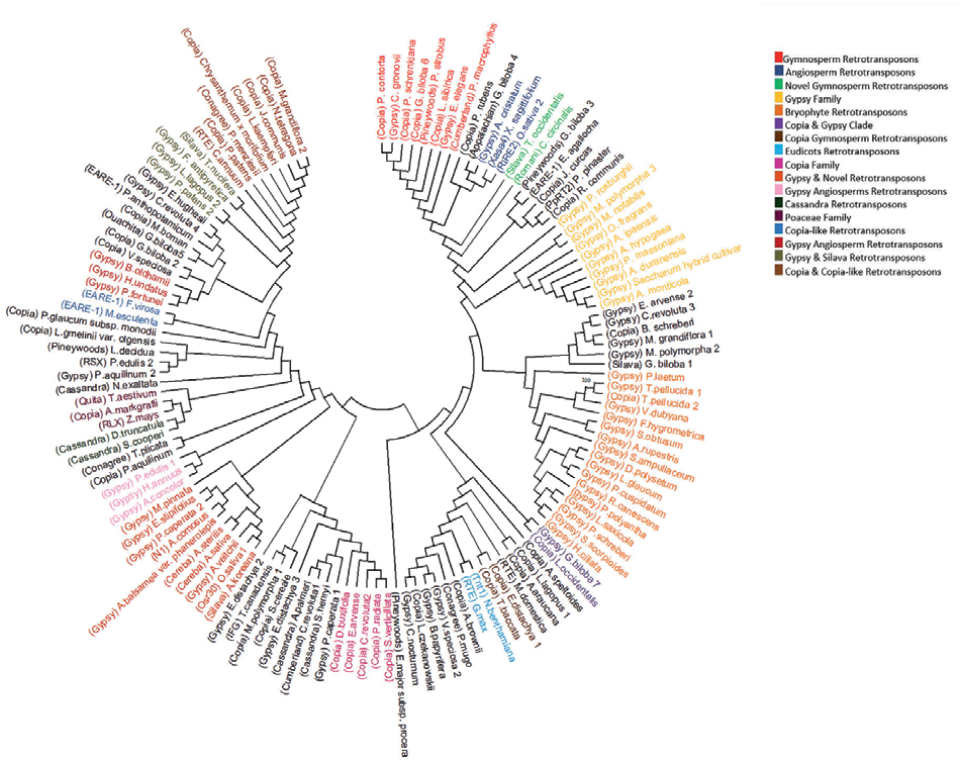


Figure 4.
Circular ideogram of retrotransposons across diverse plant genomes.

angiosperm retrotransposons [108]. The "verdigris" group is a cluster of Gypsy retrotransposons with the inclusion of Silava retrotransposons. It was noted that Silava retrotransposons tend to cluster with Gypsy retrotransposons. The "saddle brown" group is a cluster of Copia retrotransposons with two novel Copia-like retrotransposons (RTE & Conagree). All black clusters formed the mixed groups.

Retrotransposons of the gypsy family tend to cluster together (**Figure 4**). The Gypsy family is the largest group, forming a large cluster of bryophyte sequences and eudicot sequences with few liverworts and gymnosperms sequences forming outgroups of these clades. Gypsy retrotransposons are very diversified and more widespread in plant genomes than Copia retrotransposons [109]. Retrotransposons of the Copia family tend to be grouped based on the plant group they belong to. These retrotransposons are interspersed with novel families of retrotransposons that are Copia-like in structure. Copia-like retrotransposons are common in plant genomes and are identified by their reverse transcriptase, similar in structure to the

Copia family retrotransposons [110]. Gymnosperm retrotransposons are grouped together regardless of family, and they are associated with monocot retrotransposons. Possibly, this attribute could be the result of retrotransposal duplication events in these genomes [111]. Notably, retrotransposons are more active in the Poaceae family [112], leading to the genesis of more unique and novel retrotransposon families.

4. Conclusions

Retrotransposons are such a significant part of plant genomes that they warrant more studies to understand them better. Retrotransposons were conserved in nature, tended to cluster in different plant families and classes, and revealed significant genome relationships between different families within a plant division. Retrotransposons were characterized by certain motifs and domains useful in classifying them and helping understand their role in plant genomes. Plant retrotransposons exhibited much diversification while also retaining the conservation of certain parts of their structures. Retrotransposons in plant genomes retained genes from other life domains, just as they reserved harmful genes. They can also keep useful genes essential in helping their hosts survive adverse conditions. Findings in the PROSITE amino acid patterns and profiles found that some of these plant retrotransposons contain viral, bacterial, fungal, and mammalian genes. The high specificity of retrotransposal Reverse Transcriptase could be used as an important tool in identifying retrotransposons. More so, phylogenetic analysis revealed the relationships of the retrotransposons and unveiled their diversification into several lineages. This study provided valuable information on the characteristics, patterns, profiles, diversity, and phylogenetic relationship of retrotransposons across the range-wide plant orders and families and are necessary in understanding the functions, complexity, and dynamics of plant genomes.

Acknowledgements

We would like to thank the faculty members of the Department of Biology, College of Science and Technology, Adventist University of the Philippines, and reviewers for the valuable comments. The National Center for Biotechnology Information, Bethesda MD, USA for the DNA sequences. We are grateful to Sir Owen E. Pitakia, Dr. Edwin Balila, and Dr. Lorcelie Taclan for their indispensable counsels and support.

Conflict of interest

The authors declare no conflict of interest.

DOI: http://dx.doi.org/10.5772/intechopen.99074

Author details

Aloysius Brown[1], Orlex B. Yllano[1,2*], Leilani D. Arce[3], Ephraim A. Evangelista[4], Ferdinand A. Esplana[4], Lester Harris R. Catolico[5] and Merbeth Christine L. Pedro[5]

1 Department of Biology, College of Science and Technology, Adventist University of the Philippines, Silang, Cavite, Philippines, Brewerville City, Liberia

2 Cell and Molecular Biology Laboratory, Department of Biology, CST, Adventist University of the Philippines, Silang Cavite, Philippines

3 Botany and Systematics Laboratory, Department of Biology, CST, Adventist University of the Philippines, Silang Cavite, Philippines

4 Microbiology Laboratory, Department of Biology, CST, Adventist University of the Philippines, Silang Cavite, Philippines

5 Anatomy and Physiology Laboratory, Department of Biology, CST, Adventist University of the Philippines, Silang Cavite, Philippines

*Address all correspondence to: obyllano@aup.edu.ph

IntechOpen

References

[1] Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. 2018;46(21):e126.

[2] Mustafin RN, Khusnutdinova EK. The Role of Transposable Elements in Emergence of Metazoa. Biochemistry (Mosc). 2018;83(3):185-99.

[3] Mita P, Boeke JD. How retrotransposons shape genome regulation. Curr Opin Genet Dev. 2016;37:90-100.

[4] Bennett MD. Variation in Genomic Form in Plants and Its Ecological Implications. New Phytologist. 1987;106(s1):177-200.

[5] Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev. 2015;35:119-25.

[6] Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. 2014;15(3):R59.

[7] Gao L, McCarthy EM, Ganko EW, McDonald JF. Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. BMC Genomics. 2004;5(1):18.

[8] Orozco-Arias S, Isaza G, Guyot R. Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning. Int J Mol Sci. 2019 Aug 6;20(15).

[9] Malik HS, Henikoff S, Eickbush TH. Poised for Contagion: Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. Genome Res. 2000;10(9):1307-18.

[10] Kazazian HH. Mobile elements: drivers of genome evolution. Science. 2004;303(5664):1626-32.

[11] Neumann P, Novák P, Hoštáková N, Macas J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification [Internet]. Vol. 10, Mobile DNA. Mob DNA; 2019. Available from: https://pubmed.ncbi.nlm.nih.gov/30622655/

[12] Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet. 2008 May;9(5):411-2; author reply 414.

[13] Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009;10(10):691-703.

[14] Schmidt T. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol. 1999. 40(6):903-10.

[15] Mao H, Wang H. Distribution, Diversity, and Long-Term Retention of Grass Short Interspersed Nuclear Elements (SINEs). Genome Biol Evol. 2017;9(8):2048-56.

[16] Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell. 2011;23(9):3117-28.

[17] Sahebi M, Hanafi MM, van Wijnen AJ, Rice D, Rafii MY, Azizi P, et al. Contribution of transposable elements in the plant's genome. Gene. 2018;665:155-66.

[18] Nagaki K, Shibata F, Kanatani A, Kashihara K, Murata M. Isolation of centromeric-tandem repetitive DNA sequences by chromatin affinity purification using a HaloTag7-fused centromere-specific histone H3 in tobacco. Plant Cell Rep. 2012;31(4):771-9.

[19] Joly-Lopez Z, Bureau TE. Diversity and evolution of transposable elements in Arabidopsis. Chromosome Res. 2014;22(2):203-16.

[20] Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326(5956):1112-5.

[21] Xiong Y, Eickbush TH. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol. 1988;5(6):675-90.

[22] Zhang L, Yan L, Jiang J, Wang Y, Jiang Y, Yan T, et al. The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. Virulence. 2014;5(6):655-64.

[23] Janicki M, Rooke R, Yang G. Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosome Res. 2011;19(6):787-808.

[24] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973-82.

[25] Bennetzen JL, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 2005;95(1):127-32.

[26] Du D, Du X, Mattia MR, Wang Y, Yu Q, Huang M, et al. LTR retrotransposons from the *Citrus x clementina* genome: characterization and application. Tree Genetics and Genomes. 2018;14(4):43.

[27] Rho M, Choi J-H, Kim S, Lynch M, Tang H. De novo identification of LTR retrotransposons in eukaryotic genomes. BMC Genomics. 2007;8:90.

[28] Mascagni F, Giordani T, Ceccarelli M, Cavallini A, Natali L. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). BMC Genomics. 2017;18(1):634.

[29] Cossu RM, Buti M, Giordani T, Natali L, Cavallini A. A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genetics and Genomes. 2012;8(1):61-75.

[30] Chang W, Jääskeläinen M, Li S, Schulman AH. BARE retrotransposons are translated and replicated via distinct RNA pools. PLoS One. 2013;8(8):e72270.

[31] Mascagni F, Barghini E, Giordani T, Rieseberg LH, Cavallini A, Natali L. Repetitive DNA and Plant Domestication: Variation in Copy Number and Proximity to Genes of LTR-Retrotransposons among Wild and Cultivated Sunflower (*Helianthus annuus*) Genotypes. Genome Biol Evol. 2015;7(12):3368-82.

[32] Joly-Lopez Z, Bureau TE. Exaptation of transposable element coding sequences. Curr Opin Genet Dev. 2018;49:34-42.

[33] Piednoël M, Carrete-Vega G, Renner SS. Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. Plant J. 2013;75(4):699-709.

[34] Paz RC, Kozaczek ME, Rosli HG, Andino NP, Sanchez-Puerta MV. Diversity, distribution and dynamics of

full-length Copia and Gypsy LTR retroelements in *Solanum lycopersicum*. Genetica. 2017;145(4-5):417-30.

[35] Usai G, Mascagni F, Natali L, Giordani T, Cavallini A. Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. Tree Genetics & Genomes. 2017;13(5):96.

[36] Sanchez DH, Gaubert H, Drost H-G, Zabet NR, Paszkowski J. High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. Nat Commun. 2017;8(1):1283.

[37] Ma J, Devos KM, Bennetzen JL. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. 2004;14(5):860-9.

[38] Ragupathy R, Banks T, Cloutier S. Molecular characterization of the Sasanda LTR copia retrotransposon family uncovers their recent amplification in *Triticum aestivum* (L.) genome. Mol Genet Genomics. 2010;283(3):255-71.

[39] Curcio MJ, Garfinkel DJ. Regulation of retrotransposition in *Saccharomyces cerevisiae*. Mol Microbiol. 1991;5(8):1823-9.

[40] Boeke JD, Corces VG. Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol. 1989;43:403-34.

[41] Nishihara H. Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation. Genes Genet Syst. 2020 Jan 30;94(6):269-81.

[42] Finnegan DJ. Retrotransposons. Current Biology. 2012;22(11):R432-7.

[43] Hirochika H, Okamoto H, Kakutani T. Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation. Plant Cell. 2000;12(3):357-69.

[44] Grandbastien M-A. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochim Biophys Acta. 2015;1849(4): 403-16.

[45] Hilbricht T, Varotto S, Sgaramella V, Bartels D, Salamini F, Furini A. Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant *Craterostigma plantagineum*. New Phytol. 2008;179(3):877-87.

[46] Zhao Y, Xu T, Shen C-Y, Xu G-H, Chen S-X, Song L-Z, et al. Identification of a retroelement from the resurrection plant *Boea hygrometrica* that confers osmotic and alkaline tolerance in *Arabidopsis thaliana*. PLoS One. 2014;9(5):e98098.

[47] Defraia C, Slotkin RK. Analysis of retrotransposon activity in plants. Methods Mol Biol. 2014;1112:195-210.

[48] Gao D, Jiang N, Wing RA, Jiang J, Jackson SA. Transposons play an important role in the evolution and diversification of centromeres among closely related species. Front Plant Sci. 2015;6:216.

[49] Goulet C, Mageroy MH, Lam NB, Floystad A, Tieman DM, Klee HJ. Role of an esterase in flavor volatile variation within the tomato clade. Proc Natl Acad Sci U S A. 2012;109(46):19009-14.

[50] Kim S, Park J, Yeom S-I, Kim Y-M, Seo E, Kim K-T, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. 2017;18(1):210.

[51] Maumus F, Allen AE, Mhiri C, Hu H, Jabbari K, Vardi A, et al. Potential impact of stress activated

retrotransposons on genome evolution in a marine diatom. BMC Genomics. 2009;10:624.

[52] Guo H, Jiao Y, Tan X, Wang X, Huang X, Jin H, et al. Gene duplication and genetic innovation in cereal genomes. Genome Res. 2019;29(2):261-9.

[53] Lin J, Cai Y, Huang G, Yang Y, Li Y, Wang K, et al. Analysis of the chromatin binding affinity of retrotransposases reveals novel roles in diploid and tetraploid cotton. J Integr Plant Biol. 2019;61(1):32-44.

[54] Boonjing P, Masuta Y, Nozawa K, Kato A, Ito H. The effect of zebularine on the heat-activated retrotransposon ONSEN in *Arabidopsis thaliana* and *Vigna angularis*. Genes Genet Syst. 2020;95(4):165-72.

[55] ExPASy – PROSITE [Internet]. [cited 2021 Apr 8]. Available from: https://prosite.expasy.org/

[56] MOTIF: Searching Protein Sequence Motifs [Internet]. [cited 2021 Apr 8]. Available from: https://www.genome.jp/tools/motif/

[57] Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. Current Protocols in Bioinformatics. 2003;00(1):2.3.1-2.3.22.

[58] Buckley RM, Kortschak RD, Raison JM, Adelson DL. Similar Evolutionary Trajectories for Retrotransposon Accumulation in Mammals. Genome Biol Evol. 2017;9(9):2336-53.

[59] Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst. 2020;94(6):233-52.

[60] Suguiyama VF, Vasconcelos LAB, Rossi MM, Biondo C, de Setta N. The population genetic structure approach adds new insights into the evolution of plant LTR retrotransposon lineages. PLoS One. 2019;14(5):e0214542.

[61] Masuta Y, Kawabe A, Nozawa K, Naito K, Kato A, Ito H. Characterization of a heat-activated retrotransposon in *Vigna angularis*. Breed Sci. 2018;68(2):168-76.

[62] Reznikoff WS, Bordenstein SR, Apodaca J. Comparative sequence analysis of IS50/Tn5 transposase. J Bacteriol. 2004;186(24):8240-7.

[63] Oliva N, Florida Cueto-Reaño M, Trijatmiko KR, Samia M, Welsch R, Schaub P, et al. Molecular characterization and safety assessment of biofortified provitamin A rice. Scientific Reports. 2020;10(1):1376.

[64] Cao Y, Jiang Y, Ding M, He S, Zhang H, Lin L, et al. Molecular characterization of a transcriptionally active Ty1/copia-like retrotransposon in Gossypium. Plant Cell Rep. 2015;34(6):1037-47.

[65] Cavalcante MG, Souza LF, Vicari MR, de Bastos CEM, de Sousa JV, Nagamachi CY, et al. Molecular cytogenetics characterization of *Rhinoclemmys punctularia* (Testudines, Geoemydidae) and description of a Gypsy-H3 association in its genome. Gene. 2020;738:144477.

[66] Rezende-Teixeira P, Siviero F, Brandão AS, Santelli RV, Machado-Santelli GM. Molecular characterization of a retrotransposon in the *Rhynchosciara americana* genome and its association with telomere. Chromosome Res. 2008;16(5):729-42.

[67] Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 1990 Oct;9(10):3353-62.

[68] CDD Conserved Protein Domain Family: RT_LTR [Internet]. [cited 2021

Apr 18]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cd01647

[69] Das D, Georgiadis MM. The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. Structure. 2004;12(5):819-29.

[70] Nowak E, Potrzebowski W, Konarev PV, Rausch JW, Bona MK, Svergun DI, et al. Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. Nucleic Acids Res. 2013;41(6):3874-87.

[71] Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature. 2010;464(7286):232-6.

[72] Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science. 1994;266(5193):1981-6.

[73] CDD Conserved Protein Domain Family: gag_pre-integrs [Internet]. [cited 2021 Apr 19]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cl16514

[74] Yang S, Zeng K, Chen K, Zhao X, Wu J, Huang Y, et al. Sequence Evolution, Abundance, and Chromosomal Distribution of Ty1-copia Retrotransposons in the *Saccharum spontaneum* Genome. Cytogenet Genome Res. 2020;160(5):272-82.

[75] Miller K, Rosenbaum J, Zbrzezna V, Pogo AO. The nucleotide sequence of *Drosophila melanogaster* copia-specific 2.1-kb mRNA. Nucleic Acids Res. 1989;17(5):2134.

[76] Jørgensen MG, Pandey DP, Jaskolska M, Gerdes K. HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. J Bacteriol. 2009;191(4):1191-9.

[77] Martin PR, Watson AA, McCaul TF, Mattick JS. Characterization of a five-gene cluster required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. Mol Microbiol. 1995;16(3):497-508.

[78] Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41(8):873-5.

[79] Boguski M, Murray A, Powers S. Novel repetitive sequence motifs in the alpha and beta subunits of prenyl-protein transferases and homology of the alpha subunit to the MAD2 gene product of yeast. New Biol. 1992 Apr 1;4(4):408-11.

[80] Liu Y, Hong X, Kappler J, Jiang L, Zhang R, Xu L, et al. Ligand-receptor binding revealed by the TNF family member TALL-1. Nature. 2003;423(6935):49-56.

[81] Han G, Lu C, Guo J, Qiao Z, Sui N, Qiu N, et al. C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. Front Plant Sci [Internet]. 2020 [cited 2021 May 30];11. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2020.00115/full

[82] InterPro [Internet]. [cited 2021 May 30]. Available from: https://www.ebi.ac.uk/interpro/

[83] Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, et al. The PROSITE database. Nucleic Acids Res. 2006;34(Database issue):D227-230.

[84] Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V,

Bairoch A, et al. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res. 2010;38(Database issue):D161-166.

[85] Cohen G, Shiffman D, Mevarech M, Aharonowitz Y. Microbial isopenicillin N synthase genes: Structure, function, diversity and evolution. Trends in Biotechnology. 1990;8:105-11.

[86] TonB-dependent receptor, conserved site (IPR010917) – InterPro entry – InterPro [Internet]. [cited 2021 Apr 18]. Available from: https://www.ebi.ac.uk/interpro/entry/InterPro/IPR010917/

[87] Babakhani S, Oloomi M. Transposons: the agents of antibiotic resistance in bacteria. J Basic Microbiol. 2018;58(11):905-17.

[88] Rao JK, Erickson JW, Wlodawer A. Structural and evolutionary relationships between retroviral and eucaryotic aspartic proteinases. Biochemistry. 1991;30(19):4663-71.

[89] Davies DR. The structure and function of the aspartic proteinases. Annu Rev Biophys Biophys Chem. 1990;19:189-215.

[90] Gazda LD, Joóné Matúz K, Nagy T, Mótyán JA, Tőzsér J. Biochemical characterization of Ty1 retrotransposon protease. PLoS One. 2020;15(1):e0227062.

[91] Checkley MA, Mitchell JA, Eizenstat LD, Lockett SJ, Garfinkel DJ. Ty1 gag enhances the stability and nuclear export of Ty1 mRNA. Traffic. 2013;14(1):57-69.

[92] Katz RA, Skalka AM. The retroviral enzymes. Annu Rev Biochem. 1994;63:133-73.

[93] Herschhorn A, Hizi A. Retroviral reverse transcriptases. Cell Mol Life Sci. 2010;67(16):2717-47.

[94] Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem. 1998;67:1-25.

[95] Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, et al. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci U S A. 2000;97(15):8233-8.

[96] Katz RA, Jentoft JE. What is the role of the cys-his motif in retroviral nucleocapsid (NC) proteins? Bioessays. 1989;11(6):176-81.

[97] Dodonova SO, Prinz S, Bilanchone V, Sandmeyer S, Briggs JAG. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. Proc Natl Acad Sci U S A. 2019;116(20):10048-57.

[98] Sandmeyer SB, Clemens KA. Function of a retrotransposon nucleocapsid protein. RNA Biol. 2010;7(6):642-54.

[99] Makarova KS, Aravind L, Koonin EV. SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes. Trends Biochem Sci. 2002;27(8):384-6.

[100] Sabot F, Schulman AH. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity (Edinb). 2006;97(6):381-8.

[101] Zhang D, Kan X, Huss SE, Jiang L, Chen L-Q, Hu Y. Using Phylogenetic Analysis to Investigate Eukaryotic Gene Origin. J Vis Exp. 2018;(138).

[102] Hillis DM. Phylogenetic analysis. Current Biology. 1997;7(3):R129-31.

[103] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial

DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512-26.

[104] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547-9.

[105] Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010;17(3):337-54.

[106] Sacks-Davis R, Daraganova G, Aitken C, Higgs P, Tracy L, Bowden S, et al. Hepatitis C virus phylogenetic clustering is associated with the social-injecting network in a cohort of people who inject drugs. PLoS One. 2012;7(10):e47335.

[107] Efron B, Halloran E, Holmes S. Bootstrap confidence levels for phylogenetic trees. PNAS. 1996;93(23):13429-13429.

[108] Huang J, Wang Y, Liu W, Shen X, Fan Q, Jian S, et al. EARE-1, a Transcriptionally Active Ty1/Copia-Like Retrotransposon Has Colonized the Genome of *Excoecaria agallocha* through Horizontal Transfer. Front Plant Sci [Internet]. 2017 [cited 2021 Jun 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2017.00045/full

[109] Suoniemi A, Tanskanen J, Schulman AH. Gypsy-like retrotransposons are widespread in the plant kingdom. Plant J. 1998;13(5):699-705.

[110] Voytas DF, Cummings MP, Koniczny A, Ausubel FM, Rodermel SR. copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci U S A. 1992;89(15):7124-8.

[111] Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, Bousquet J. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. BMC Biology. 2012;10(1):84.

[112] Vicient CM, Jääskeläinen MJ, Kalendar R, Schulman AH. Active Retrotransposons Are a Common Feature of Grass Genomes. Plant Physiology. 2001;125(3):1283-92.

Chapter 9

Sex Determination

Rakesh Choudhary, Subhash Chand, Tejveer Singh, Rajesh K. Singhal, Vinay K. Chourasiya and Indu

Abstract

A wide array of sex determination mechanisms, encompassing genetic and non-genetic pathways (i.e., hormonal, environmental, and epigenetic factors), have been found among different organisms. The presence of two complementary sexes, male and female, is an ancient feature in biology. Triggering the differentiation of male and female reproductive organs is a conserved ontogenic process, and sex determination is an inherently fascinating process. Sex determination is dependent on molecular signaling whether the male and the female differentiating pathway is activated, and different triggering elements such as genetic, non-genetic, and epigenetic factors control the whole process. This chapter describes various aspects of sex determination, such as historical development, the evolution of sex chromosomes, and different sex determination systems in other organisms.

Keywords: genic sex determination, Haplo-diploidy, environmental sex determination

1. Introduction

Sexual reproduction is a historical process of life on earth, and the most popular heterogametic system (X and Y sex chromosomes) in humans and many other organisms leaves an imprint that sex determination mechanism is ancient and conserved [1]. Sex determination is inherently an integral part of reproduction that separates reproductive organs responsible for male and female gamete production [2, 3]. It is an intricate developmental process that describes whether the individual will be developed as male or female. At the same time, sexual differentiation is the subsequent development of phenotypic differences (primary and secondary sexual characters) between male and female individuals from an undifferentiated zygote [2]. Sex differentiation stages are decided by the sex determination, that is, the gender-specific response of different tissues to hormones produced by the gonads (male or female reproductive organs) distinctly in both genders [4, 5].

Various pathways decide males and females, and these pathways have been evolved rapidly in many species or genera/taxa. Sex determination is regulated by several different genetic (i.e., sex chromosomes) and non-genetic pathways (i.e., hormonal, environmental, and epigenetic factors). In different animal and plant species, genetic systems have been classified as homogametic or heterogametic sex types [2]. In most species, heteromorphic sex chromosomes are present, which are the results of evolutionary changes in size or shape of the sex chromosomes. Similarly, non-genetic pathways also play a key role in determining the sex in fern species (hormonal regulation) and crocodiles, alligators, and turtles (thermo-regulation).

Sex determination is an important evolutionary process as it encourages the genetic fitness of an individual. The ultimate aim of sex determination is to promote

the heterozygosity or accumulation of diverse alleles in a species, which is vital for creating genetic variation in living organisms. It is pivotal in plant breeding to design a specific breeding program as per the need and demands of diverse stakeholders to improve plant productivity and nutritional quality. The exact mechanism or genes determining the sex or reproductive organ is unclear. Therefore, it is an important area of study in developmental and evolutionary biology, as well as in ecology. Sex determination in various plant and animal species is not under the control of the universal model. Thus, this chapter provides a brief overview of the different mechanisms of sex determination in plant and animal species.

2. Historical development

Whether a plant or animal will become a male, a female, or bisexual is determined during the initial development of an organism. Hundreds of years ago, researchers have started studying the mechanism of sex determination. For example, in 335 B.C.E., **Aristotle** anticipated that sex is controlled or ruled by the heat of the male partner during intercourse. A male child will be born when the male parent's heat overwhelms the female parent's coldness and *vice versa*. Environmental theories of sex determination (i.e., Aristotle's theory, in reptiles, temperature during embryo development regulates the gender) were popular until about 1900 and meticulous scientific research began after the discovery of sex chromosomes during 1900, and gradual scientific improvements followed during the next century. In 1891, **Hermann Henking** [6], a German Biologist studying spermatogenesis in the insect firebug (*Pyrrhocoris apterus;* 2n = 24), detected that, as a result of meiosis, half of the spermatozoa have not received all 12 chromosomes and ended with 11 chromosomes only. It means one of the chromosomes was not involved in meiosis. This chromosome seemed and behaved differently from others; he was not able to speculate the significance of this element and named it **"X element or X body"**. In 1902, **Clarence Erwin McClung** [7] cleared the Henking assumption through cytological observations on several grasshopper species and demonstrated that the somatic cells in female grasshoppers are different in chromosome number than do corresponding cells in the male, which he referred to as "**accessory or supernumerary chromosome**" and demonstrated their association with sex determination [8]. Later on, American geneticist **Edmund Beecher Wilson** (1905) observed differences, either in the presence or in the absence of one chromosome or in the size of one chromosome pair in germ cells of both the sexes of *protenor* species [9]. Another American geneticist **Nettie Maria Stevens** (1905) studied germ cells of both sexes of mealworm beetle (*Tenebrio molitor*) and found that in males, one chromosome was smaller than the other chromosome, and she confirmed that this chromosome must be regulating the sex in males [8]. Later on, this smaller chromosome was named as "**Y chromosome**" and the larger one as "**X chromosome**" by Stevens (1905). After Stevens died in 1912, Wilson was the first to designate the name "**sex chromosome**" for the pair of XX and XY chromosomes. **Hermann Joseph Muller** (1914), an American geneticist, speculated that differentiation of sex chromosomes would arise from lack of recombination due to the appearance of sex-determining genes on the Y or W chromosomes [10]. By the end of the 1950s, the male-determining function was established on the small arm of the Y chromosome and was named "**testis determining factor"** (*TDF* in humans and *Tdf* in mice). Further, **Ohno** (1967) proposed the concept of ancestral sex chromosomes and their progress to evolve modern-day sex chromosomes by degeneration of the Y or W chromosomes [11]. As science progressed, the major breakthrough in sex determination was achieved through sequencing and transgenic approaches. Gene-sequencing approaches revealed an open-reading frame (ORF) coding a single

exon gene in a male mouse (XY) and named it as a **sex-determining region** of the Y chromosome (*SRY* in humans and *Sry* in mice). Conclusive evidence of functionality of *SRY* gene was developed through transgenic approach by generating sex-reversed mouse (transgenic XX mouse having *Sry* gene).

3. Evolutionary differentiation of sex chromosomes

Sex chromosome evolution is linked with dosage compensation of sex-linked genes [11]. For example, human sex chromosomes evolved around 300 million years ago. The Y chromosome underwent inversions that inhibited large regions from recombination between homologous regions of X and Y chromosomes. This leads to the gradual spread of regions with reduced recombination. Sequence-based analysis shows the six evolutionary strata on the X chromosome, and each gene on it diverged from their Y paralogs for the same length of time. Same evolutionary strata were also found in other mammals and even in birds. It is evident that Y chromosome in mammals and W chromosome in birds are poor in gene richness, and also have lost several functional genes. The human X chromosome maintains 98% of genes, while Y chromosome retains only 3% of the genes located on the proto-sex chromosomes. The human Y chromosome is rich in palindromic duplicated sequences that help in the retention of specific Y-linked genes, which are essential for male fertility. These sequences also endorse deletions in chromosomes and tend to male sterility due to functional gene loss. Thus, these sequences maintain the integrity of Y chromosomes. **H.J. Muller** [10] suggested the origin of sex chromosomes from a pair of autosomes (**Figure 1**).

4. Sexual differentiation in animals and plants

In animals, **primary sex characters** are associated with male and female gametes producing organs such as gonads, and their development depends on the genes of their zygotes. **Secondary sex characters** are associated with different attributes, which differentiate males and females, such as the development of mammary gland,

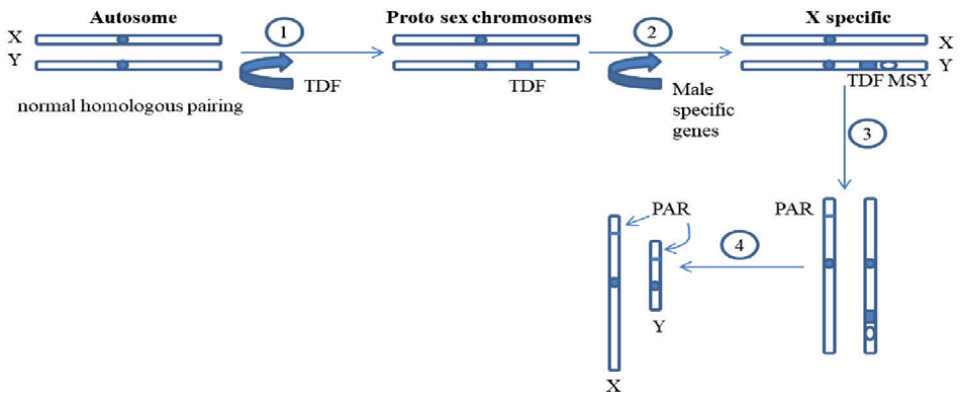


Figure 1.
Evolutionary differentiation of X and Y sex chromosomes from ancient autosomes [12]. There was end-to-end pairing between ancient X and Y chromosomes. During evolution sex-determining locus such as TDF (testis-determining factor) accumulated in one chromosome of pair (step-1). Further, there was accumulation of male-specific gene/s (step-2), which was responsible for chromosomal recombination repression (step-3) and led to development of male-specific region (MSY) on Y chromosome. Mutations and deletions (step-4) in the non-recombining region rapidly degraded the sex-specific chromosomes. The pseudoautosomal region (PAR) is present on both X and Y chromosomes in small portion, which helps in partial chromosomal pairing between X and Y chromosomes at anaphase I (modified from Graves, 2006 [5]).

genital duct, pitch of voice. The development of these traits is mainly due to hormones produced by diploid gonads.

In most plant species, both male and female reproductive organs are present in same flower (i.e., **hermaphrodite or bisexual** plants) or in different flowers of same plant (**monoecious** plants), and in some cases such as papaya, date palm, spinach, asparagus, male and female sex organs are present on flower of different plants (i.e., **dioecious** or **unisexual** plants). The monoecious plants produce either **staminate** (male)- or **pistillate** (female)-type flower in the same plant such as maize [13]. However, there are different kinds of flower combinations in monoecious plants such as **andromonoecious** (many Umbelliferae)—has staminate and hermaphrodite flowers, **gynomonoecious** (Atriplex and many Compositae)—has pistilate and hermaphrodite flowers, and **androgynomonoecious** or **trimonoecious** (*Acer campestre*)—has staminate, pistillate, and hermaphrodite flowers [13, 14].

5. Sex determination systems

Conventionally sex determination systems are classified based on the mechanism or causative factors involved in the specification of individual sex. Broadly, it has been classified into four categories:

- **Genetic sex determination (GSD)—**when sex is determined early in the development by genetic factors (sex chromosomes, genes or alleles).

- **Environmental sex determination (ESD)**—when the sex of an individual is influenced by environmental parameters such as temperature, photoperiod, nutrition.

- **Maternal sex determination (MSD)**—when the sex of an offspring is determined by genotype or physiological condition of the mother.

- **Mixed sex determination**—when both genetic and environmental factors determined the sex of an individual.

- Further, each sex determination system has been classified into different categories as depicted in **Figure 2** and detailed elaboration is given below.

5.1 Genetic sex determination (GSD)

Genetic sex determination system is also recognized as **genotypic sex determination**, and the development of an individual as male or female is triggered by the presence or absence of one or more genes or chromosomal segment or the entire chromosomal complement. These gene/s or chromosome/s is responsible for the primary and secondary sexual characters associated with each sex. Indeed, genes responsible for the development of male or female sex are located on a single pair of homologous chromosomes (sex chromosomes). In both sexes, they occur distinctly and are characterized by specific genes or by a different allelic constitution at homologous loci. The evolution of separate sexes (male and female) is the result of the evolution of **anisogamy** that is, sexual reproduction by the fusion of dissimilar gametes. The **hermaphroditism/gynandromorphism** (male or female sex organs within an individual) is common phenomenon in most of flowering plants (more than 90%), whereas it is a very rare phenomenon in animals (in some individuals of *Drosophila*). Separate sexes

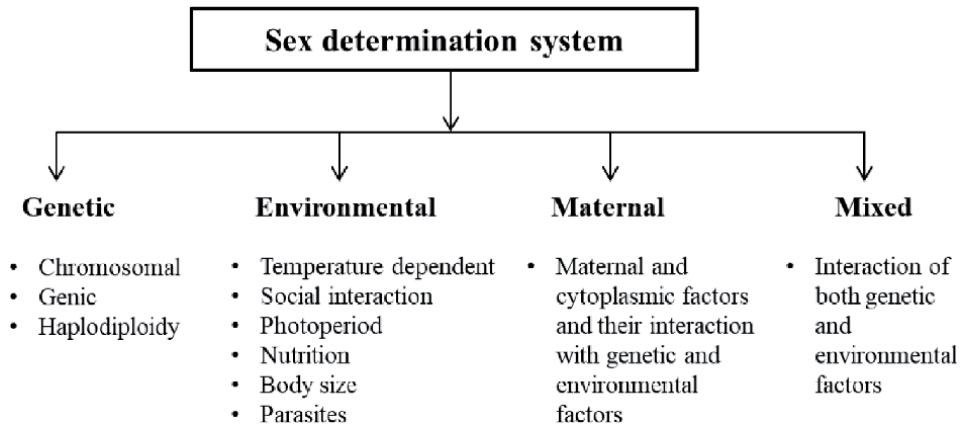


Figure 2.
General classification of sex determination system.

have evolved individually in both plants and animals, which suggests that there must be an evolutionary penalty for hermaphroditism. In animals, genetic sex determination is very well established in most of the species and this system is very well studied in *Drosophila melanogaster* flies, *Caenorhabditis elegans* nematodes, and humans. Commonalities among these bring us to a general impression of genetic regulations of sex determination and conservation of sex determination mechanism.

Sex determination mechanisms are very complex and evolved with a remarkably diverse array among plant species than among animals. In plants, a number of forms are present in functional hermaphroditism in flowering plants, varying from "perfect flower- male and female reproductive organ in each flower" to "monoecy-separate sex flowers on same individuals." Many other forms are like gynomonoecious (both female and hermaphrodite flowers), andromonoecious (both male and hermaphrodite flowers), dioecious (separate sex individuals), gynodioecious (either female or hermaphrodite), and androdioecious (either male or hermaphrodite). In plants, rapid progress is achieved in learning genetics and molecular mechanism of sex determination by comparing the monoecy and dioecy.

5.1.1 Chromosomal sex determination system

In chromosomal sex determination systems, male and female individuals differ from each other by either in morphology or in a number of one pair of chromosomes these are known as **sex chromosomes or allosomes** or **heterochromosomes,** which are dissimilar to the normal chromosomes (autosomes). On the basis of structure, there are two types of sex chromosomes such as: **i) homomorphic**—both X and Y chromosomes are structurally similar and **ii) heteromorphic**—both X and Y chromosomes are distinct morphologically. In diploid species, where male or female individual produces different types of gametes is known as **heterogametic sex** (**Table 1**), whereas individual producing similar kind of gametes is known as **homogametic sex.**

There are different chromosomal mechanisms for sex determination and illustrated below:

Chromosomal sex determination is widespread, but not ubiquitous, in the animal kingdom. Autosomes are present in two copies in diploid organisms, three copies in triploids, and so on. Generally, males are XY and females are XX in most mammalian species. In **XY system**, X chromosome is large and gene-rich, while

S.N.	Chromosomal mechanism	Male	Female	Example (Animals)	Example (Plants)
1.	XX (female) and XY (male)	Heterogametic (XY)	Homogametic (XX)	Humans, mice, Diptera, Hemiptera, Coleoptera, most common in animals	Asparagus, Spinach, Hemp, White Campion, Sorrel, Humulus
2.	XX (male) and XY (female) or ZZ and ZY system	Homogametic (XX)	Heterogametic (XY)	Birds, silkworm	Maidenhair tree, California poplar, Wild strawberry
3.	XX (female) and X0 (male)	Heterogametic (X0)	Homogametic (XX)	Grasshopper, protenor, Orthopteran insects	—
4.	X0 (female) and XX (male)	Homogametic (XX)	Heterogametic (X0)	Insects such as *Fumea*	—

Table 1.
Different mechanisms of the chromosomal sex determination in animals and plants.

Y chromosome is small and heterochromatic, that is, almost devoid of genes. Generally, animal cells comprise two types of sex chromosomes, that is, X chromosomes present in both male and female, while Y chromosomes present in male only. Homogametic parent produces one type of gametes, while heterogametic two different types of gametes (**Figure 3a** and **b**).

In land plant species, heteromorphic sex chromosomes are found in most of the species and homomorphic sex chromosomes are restricted only to the gymnosperm and angiosperm. In asparagus, papaya, and spinach, the X and Y chromosomes are homomorphic but functionally distinct. Asparagus also shows distinct YY male, which is unique in its type. In the case of spinach, females are homogametic (XX) and males are heterogametic (XY) as mammals. The Y chromosome consists of genes that are responsible for the suppression of carpel development and for activation of stamen development. In *Ginkgo biloba* (Maidenhair tree) and *Populus trichocarpa* (California poplar), *Fragraria elateria* (wild strawberry) female is heterogametic, while male is homogametic.

The heteromorphic sex chromosomes system is present in *Cannabis sativa (Hemp), Silene latifolia L.* (white Campion or liverwort), *Rumex acetosa* (Sorrel),

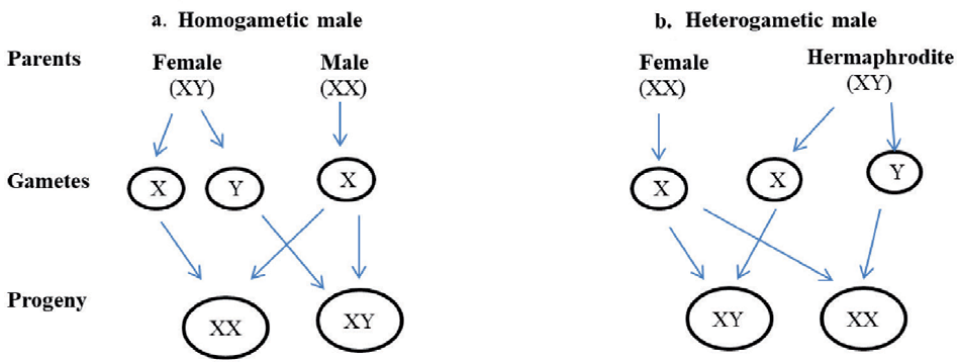


Figure 3.
a and b. Homogametic and heterogametic nature of males in birds and humans, respectively. Homogametic male (XX) produces one kind of haploid gametes only, that is, X type, while heterogametic male produces two different types of gametes, that is, X and Y type during gametogenesis. In both case, male and female progenies are produced in equal proportion.

Humulus spp., etc. The X and Y chromosomes are morphologically and functionally dissimilar and show a lack of complete pairing with each other during meiosis. In the above species, females are homogametic (XX) and males are heterogametic (XY). Some of the organisms have multiple heterochromosomes either in one or in both sex. By analogy (partially or non-homologous) in homogametic and heterogametic (XX and XY or ZZ and ZW) conditions, the gender is defined. Multiple heterochromosomes are the result of mutation (translocations) and these chromosomes are derived from the existing heterochromosomes systems (XX or ZW). Female (XX) and male (XY_1Y_2) are found in *Humulus japonicas* and *R. acetosa* etc., whereas female produces one type of gamete (X) and male produces two types of gametes (X and Y_1Y_2). Mating between egg cell (X) and male gamete (X) tend to produce diploid female (XX), while a fusion of egg cell (X) and male gamete (Y_1Y_2) tends to produce male (XY_1Y_2). Some strains of *Humulus lupulus* show homogametic female ($X_1X_1X_2X_2$)—produces one type of gamete (X_1X_2), and heterogametic male ($X_1X_2Y_1Y_2$)—producing two different type of gametes (X_1X_2 and Y_1Y_2). These chromosomes are generally present in orthopteran, crustacean, coleopterans, and mammals.

5.1.2 Sex determination due to active Y chromosome in plants

In contrast to the animal Y chromosome, plant Y chromosome is large and contains a high proportion of genomic DNA. In *Rumex acctosa*, the Y chromosome is rich in condensed heterochromatin, while it is rich in euchromatin in *S. latifolia*. In case of *S. latifolia,* X and Y chromosomes contain sex-determining genes; however, autosomal genes also play a significant role in sex determination. The Y chromosome can be divided into the four major functional fragments and they rule the sex differentiation (**Figure 4**) such as **i) female suppressor region**—contains genes of female suppression and positioned at one end of chromosome; **ii) male promoter region—**contains genes for promotion of maleness (*i.e.*, development of stamen); **iii) male fertility region**—contains genes for initiation of male fertility and anther maturation; **iv) pairing region**—helps in chromosomal pairing with one end of X chromosomes. Thus, normal disjunction of X chromosomes and Y chromosomes occurs during anaphase I. Mutation in first region leads to production of both male and female flowers on same plant. Mutation in second region only leads to the development of asexual flower. Mutation in third region only tends to develop male sterile XY plant. Mutation in fourth region only leads to chromosomal anomalies during meiotic cell division. The X chromosome consists of two functional regions. **First region** covers major portion of X chromosome and is nonhomologous to the Y chromosome. It consists of genes that are responsible for the development of pistillate flower. However, function of this segment is suppressed by first fragment

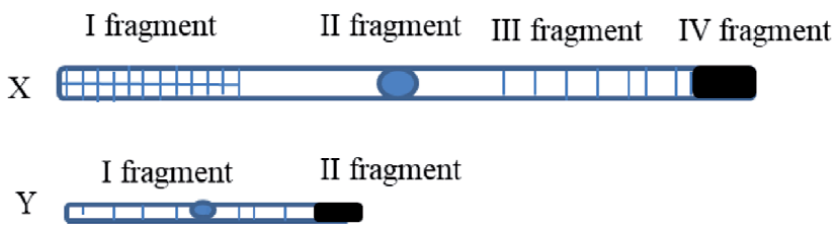


Figure 4.
Structure of X and Y chromosomes in plants. The length and genome content is higher in X chromosomes than Y chromosomes. The X and Y chromosomes partially pair with IV fragment of X chromosome and II fragment of Y chromosome, and helps to regular normal segregation of sex chromosomes at anaphase I during gametogenesis.

of Y chromosome. Sometime, one Y chromosome can inhibit the effect of four X chromosomes and produce male flower in XY plants. In **second region**, small end portion of X chromosome helps to pair with Y chromosome due to their homology. Thus, two genes are essential for sex determination in plants—one gene for suppression of carpel development and other gene for the development of stamen. However, mammalian cell carries single gene (*SRY)*, which controls sex determination.

There is an abundant diversity in chromosomal sex determination systems; however, there are some different chromosomal sex determination systems which fall under the category of "**Miscellaneous or Other category.**" The **UV system** is also a part of this category. UV system chromosomal sex determination is determined at haploid phase of the life cycle. Females and males are haploid and characterized by the possession of a sex chromosomes U and V, respectively. This system is generally present in organisms with haplontic and haplodiplontic system (some algae and bryophytes) having anisogamous and heterosporous condition. In UV system, sex is determined during meiosis not at the time of fertilization. In case of fungus gnat *Sciara,* all zygotes have similar genotypes (XXXAA) and the loss in one or more paternal chromosomes will determine that the zygote will develop into a female (XXAA) or male (XAA).

The Y chromosome in XY system and W chromosome in ZW system may have gone through the degeneration process and lost some of the original genes that are present in the another sexual chromosome (X or Z). Therefore, in homogametic sex (XX or ZZ), some genes are present in double copy (as like the autosomes), whereas in heterogametic sex (XY or ZW) they are in single copy. Genetic imbalance affects all the genes on sex chromosome in XO and ZO system. The genes not involved in sexual differentiation require identical level of expression in the two sexes. **Dosage compensation** is the phenomena, which balances the level of expression of genes in both the genders [15]. Dosage compensation phenomenon is very well understood in the *Drosophila*, *Caenorhabditis*, birds, and lepidopterans. In multiple heterochromosomes, the dosage compensation system becomes more complex, for example, platypus and birds.

5.1.3 Genic sex determination system

Sex determination is governed by separate genes or alleles present on specific locus of the chromosomes of both males and females [16]. In this system, sex determination is in control of distinct alleles rather than sex chromosomes; therefore, such a system may also refer to as a **multiple allele sex determination system**. In case of polygenic sex determination system, a set of the factors (genes) distributed on several chromosomes were involved and have masculinizing or feminism effects and collectively, they govern one sex or other.

Genic balance theory (GBT) was given by **Calvin Blackman Bridges** (1921) [17] for sex determination in *Drosophila melanogaster* (2n = 2x = 8). In *Drosophila*, instead of XY sex chromosome, sex is determined by the genic balance or sex index ratio between X-chromosomes and autosome genomes (sets).

$$Sex \text{ index ratio} = \frac{No.\text{of X chromosmes}}{No.\text{of autosomal sets}} = \frac{X}{A} \tag{1}$$

In *Drosophila*, Y chromosome is heterochromatic. Thus, it is not active in sex determination (**Table 2**). However, gene for male fertility is located on Y chromosome and Y chromosome also plays a major role in spermatogenesis and

Sex index ratio	Sex type	Fertility status	Examples
X/A = <0.5	Super male or meta-male	Sterile male	3A + X0; 3A + XY
X/A = 0.5	Male	Fertile (Y chromosome present) or sterile male (Y chromosome absent)	2A + XY (fertile); 2A + X0 (sterile); 4A + XX (sterile); 4A + XXY (fertile)
X/A= >0.5 and < 1.0	Intersex	Sterile	3A + XX; 3A + XXY; 4A + XXX
X/A = 1.0	Female	Fertile female	2A + XX; 3A + XXX; 2A + XXY; 3A + XXXY
X/A= >1.0	Super female or meta-female	Sterile female	3A + XXXX; 2A + XXX

Table 2.
Sex index ratio of genic balance mechanism in Drosophila [Bridges, 1921].

development of male reproductive organ. Hence, Y chromosome is essential for restoring male fertility. The gene of femaleness is located on X-chromosome and gene associated with maleness is located on autosomes. It is also applicable to some other animal species such as nematodes (*Caenorhabditis elegans*).

The sterile meta-females and meta-males have been entitled as **glamour girls and boys** of fly world by **Dodson**.

5.1.4 Male haploidy or haplodiploidy sex determination system

Haplodiploidy is most commonly used in insects of Order-Hymenoptera (honey bees, ants, and wasps) and Thysanoptera (thrips) for sex determination. Sex determination takes place by sets of chromosomes of an individual receives [18]. Two sets of chromosomes (diploid) tend to female and one set (haploid) tends to male sex formation [18, 19]. For example, in honey bee male individual (*i.e.*, drone) formed from unfertilized egg cells (*i.e.*, haploid). Thus, male develops from the process of parthenogenesis and called as **arrhenotoky** (where haploid egg cell develops males rather than females through parthenogenesis). However, female (queen and worker bees) develops from diploid egg cells (*i.e.*, fertilized egg cell). Thus, male has half number of chromosomes than female and is haploid. The male (drone) is solely derived from queen and in some cases from worker honey bees. The chromosomes number in diploid queen is 32, while 16 chromosomes in haploid drones. Drone produces sperm cells that consist of whole genome and sperm cells are genetically identical. Thus, the genetic makeup of female workers is derived half from mother and other half from father, while genetic makeup of drone is solely derived from mother. **Byes and coworkers** [18] cloned complementary sex-determining (*cds*) locus in the *Apis meliifera* and proved that this gene is responsible for sex determination cascade of honeybees. Interestingly, firstly, in haplodiploidy system male has no father and cannot have son but it has grandfather and can have grandson. If there is only one queen in a hive, then the relatedness between workers will be ¾ rather than ½, which is common between siblings in other sex determination systems. Thus, it shows more **eusocial behavior** of honey bees. Secondly, there will be rapid elimination of recessive lethal and deleterious alleles from the population due to haploid genomic nature of males, while dominant lethal and deleterious alleles will be removed every time of their occurrence because of their phenotypic expression in each stage (**Figure 5**).

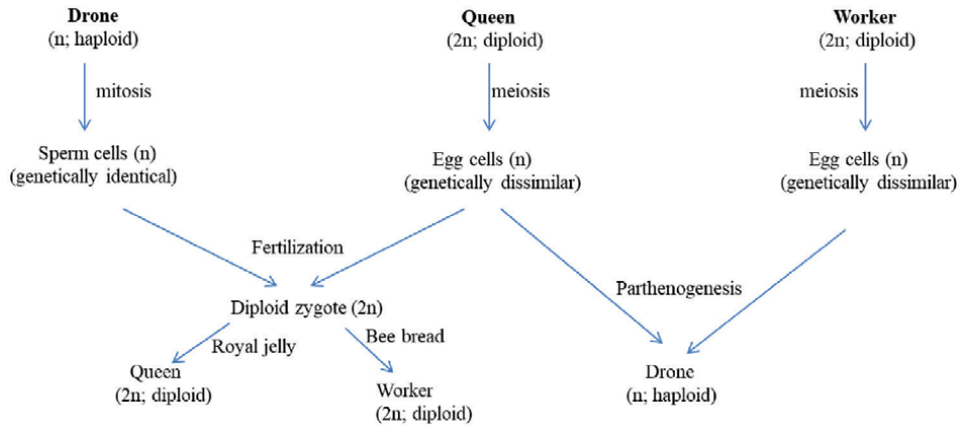


Figure 5.
Haplodiploidy system of sex determination in honey bees. Drones are haploid and produce genetically similar sperm cells by mitosis, while queen is diploid and produces egg cells that are genetically dissimilar and generated through meiotic cell division. Formation of queen or worker will be controlled by the feed stuff to the developing zygote after fertilization between male sperm cell and female egg cell. Drones are formed through parthenogenesis.

5.1.5 Single gene sex determination

There are evidences where single autosomal genes affect the sex type in animals. For example in *Drosophila*, one autosomal recessive gene-*transformer* (*tra*) affects the pattern of sex. If this is present in homozygous recessive state in XX zygotes, then it convert females into males but sterile. However, *tra* gene does not affect in male (XY) or when it is present in heterozygous state (*Tra/tra*) in female, when a female *Drosophila* having heterozygous *tra* gene (XX *Tra tra*) was mated with male having homozygous *tra* genes (XY *tra tra*). In F_1 generation, 1/4 progeny will be normal female (XX *Tra tra*), while 3/4 progeny will be male. Among male progenies, 1/3 progenies comprises XX chromosomes but found to be sterile male due to recessive homozygous *tra* genes. Another example is human, where recessive autosomal gene—*testicular feminization*—induces breast and vagina in males (XY). These male individuals also have rudimentary testis and are sterile. Single gene sex determination also occurs in dioecious plant species. For example, in papaya (**Figure 6**), sex determination occurs due to single gene with three alleles (*m, M1* and *M2*).

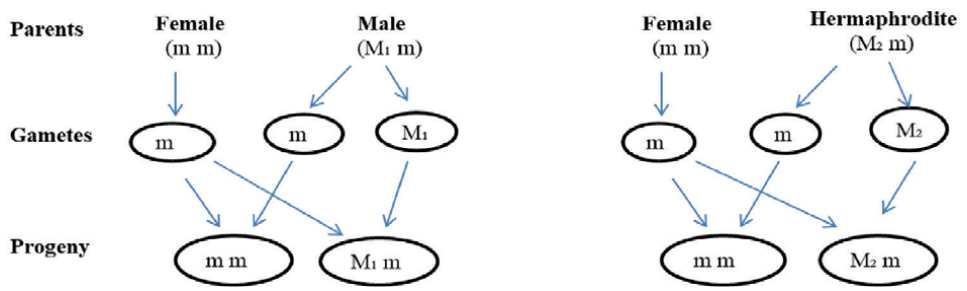


Figure 6.
Determination of sex in dioecious papaya through single gene. Genetic constitution of female, male, and hermaphrodite plants is mm, M_1m, and M_2m, respectively. Crossing between female and male plants tends to produce 50% female (mm) and 50% male (M_1m) progenies. Crossing between female and hermaphrodite plants tends to produce 50% female (mm) and 50% hermaphrodite (M_2m) progenies. Selfing in hermaphrodite tends to produce 2/3 hermaphrodite and 1/3 female progenies, while 1/4 progenies will be nonviable due to expression of lethal genes (M_2M_2).

5.2 Environmental sex determination (ESD)

In many species, sex of an individual is governed by the environmental circumstances on a zygote of unstated sex and sex is determined by the effect of environmental factors on embryonic and post-embryonic developmental stages. The ESD generally occurs in unicellular eukaryotes and among multicellular organisms, it is found mainly in non-avian reptiles, amphibians, and some fishes. Among the different environmental factors, temperature plays a key role on sex determination. However, other environmental factors such as social environment, nutrition, and pH also play decisive role in sex determination. These ESD systems are more often labile than the genetic sex determination system; evolutionary drivers can force to shift ESD to the GSD system and changes are due to variation in the threshold temperature and nutrition, etc. Mainly, sexual liability is encountered in lizards (*Bassiana duperreyi* by temperature) and ferns (gametophtyic age). Based on different environmental factors, ESD mechanisms are classified, as mentioned below:

5.2.1 Temperature-dependent sex determination

Sex is irreversibly determined by the incubation temperature during embryogenesis. Temperature affects the sex in most of the species of the turtle, crocodiles, lizards, and snakes. Based on incubation temperature for eggs, there are following three different reactions that may occur (**Table 3**).

5.2.2 Size of egg or body size

Egg size decides sex differentiation in many species like sea worm (*Dinophilus*). Big egg size tends to develop female, while small egg size tends to produce male progeny. In many plants of the genus *Arisaema* (Araceae), the sex depends on the body size of plants (small plants only bear male flowers, and large plants only female flowers, while intermediate ones will have both male and female flowers).

5.2.3 Interaction with conspecifics or social sex determination

The fate of an individual as a male or female will be decided based on spatial proximity of an individual relative to other members of its own species or the interaction with other conspecifics. Chemical (pheromones) or other communication channels (tactile or visual) stimulates the developmental response for one sex or the other. *Bonellia viridis* (marine annelid) will develop as a female if its larvae settle

High temperature	Low temperature	Intermediate	Examples	Temperature range
Male	Female	Both male and female in variable proportion	Crocodiles, alligator, and lizards	High—30–35°C Low—23–28°C
Female	Male	Both male and female in variable proportion	Most species of turtles	High—30–35°C Low—23–28°C
Female	Female	Males	*Chleydra serpentine* (turtle spp.) and few crocodile spp.	High—30–35°C Low—23–28°C
Female	Female	Both male and female in variable proportion	Australian crocodile	High—>30°C Low—< 25°C

Table 3.
Sex determination reactions based on incubation temperature for eggs.

on a sea floor area in isolation from other individuals. In contrast, if larva attached to proboscis of an adult female, it starts to progress into a male through the effect of male pheromones released by female. In case of many sequential hermaphrodite fishes, they start their life as one sex and later on converted to another sex based on social interactions such as anemone fish (the largest male in the group become the dominant female after the death of dominant female) and homosporous fern (*Ceratopteris richardii*).

5.2.4 Photoperiod, nutrition, parasitism, water pH, and social interaction

Under the ESD system, photoperiod is also a sex-decisive factors and in case of brackish water amphipod *(Gammarus duebeni*), the sex ratio varies according to photoperiod exposure during the post-hatching. During elongated dark period, the proportions of male individuals will be higher in comparison with the female. Nutrition, parasitism and water pH also determine the sex in few species. Nutritional control of sex determination occurs in calanoid copepods and mermithid nematodes, whereas parasitism plays as decisive role in sex determination of some of the isopods, coenopods, and copepods. In the South American cichlid fishes and some poecilids, **water acidity** has an effect on sex determination and individuals growing in acid waters will be predominantly males and those develop in neutral or slightly basic water will be females. Many fishes are sequential hermaphrodites, where they start their life as one sex, but change sex later in development. In the anemone fish (*Amphiprion akallopisos*), which lives in social groups with one dominant breeding pair as well as several subordinate males, sex change occurs when the dominant female dies and the largest male in the group becomes the dominant female.

Environmental factors (such as light, temperature, humidity, day length, GA_3, and ethylene) have effect on limited plant species in their sex determination. For example, plant-*equisetum* develops as female under normal environmental condition, while as male under stress condition. In cucurbits such as melons and cucumbers, sex is also affected by the application of growth hormones such as GA_3 and ethylene induce femaleness. Thus, environmental factor has more impact on males rather than female.

5.3 Maternal and cytoplasmic sex determination system

In this category, the sex of the progeny depends on the mother of the individual and their interaction effect with genetic and environmental factors. Maternal sex determination occurs in two different forms—one where sex is established by its mother rather than the individual's genotype, whereas in second form, physiological conditions of the mother and their specific signals determine that an offspring will be male or female. In the dipterans insects such as *Chrysomya albiceps* and *Calliphora rufifacies,* two forms of female are present, one is producing only male offspring known as **androgenic females** (ff) and the other is exclusively producing female offspring known as **gynogenic females** (Ff). In cecidomyid midge, the sex is governed by the nutritional condition of the mother and in response to the nutritional conditions, female brain secretes a factor and it reaches to ovaries to determine the gender of an individual.

In genetic sex determination (GSD), sex-determining factors are typical nuclear genes and show the Mandelian inheritance. On the contrary, some of the sex-determining causes are inherited by the cytoplasm and transmitted only from mother to daughter and not by the males. Sex ratio distortion toward the female sex has been observed in members of crustaceans group and is unique in this group [20].

5.4 Mixed sex determination system

When the sex of an individual is ruled by the combined effect of genetic and environmental factors at various degrees, in case of American salamander *(Pleurodeles)* and fish (*Menidia menidia*), sex is determined by the joint effect of both the sex karyotype (ZW and ZZ) and incubation temperature of eggs.

6. Conclusion

The several mechanisms of sex determination reveal the diverse pathways governing sex determination in both plants and animals and these pathways are also very well understood in various model organisms. The highly evolved system of sex determination is heterogametic sex determination in animals, that is, XX/XY. Even though there are several unsolved mysteries related to the sex determination system such as why heterogamety is more common in male then female? Why degeneration of sex chromosomes occurs only in few organisms not in all? With the progress in molecular techniques over the past decades, several puzzles were solved like discovery of *Sry* and a ray of hope arises to learn more about molecular basis of sex determination, evolution of sex chromosome, mapping of gene, sequencing, gender-dependent expression of sex-regulating gene, and relationship between the evolution of genetic degeneration and dosage compensation. Multiple "-omics" data and integrative approaches will allow scientists to address the unresolved questions and finding the new sex-determining genes as well as genetic networks involved in sex determination.

Author details

Rakesh Choudhary[1], Subhash Chand[2], Tejveer Singh[2], Rajesh K. Singhal[2], Vinay K. Chourasiya[3] and Indu[2*]

1 Rani Lakshmi Bai Central Agricultural University, Jhansi, India

2 ICAR-Indian Grassland and Fodder Research Institute, Jhansi, India

3 Depatment of Seed Science and Technology, CSA University of Agriculture and Technology, Kanpur, India

*Address all correspondence to: agrico.rakesh@gmail.com

References

[1] Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex determination: why so many ways of doing it? PLoS Biology. 2014;**12**(7):e1001899

[2] Singh BD. Sex determination in plants (chapter-20). In: Genetics. New Delhi: Kalyani publisher; 2009. pp. 255-263

[3] Fusco G, Minelli A. Determination of Sex and Mating Type. In: The Biology of Reproduction. Cambridge: Cambridge University Press; 2019. pp. 297-341. DOI: 10.1017/9781108758970.008

[4] Grimaldi D, Engel MS, Engel MS, Engel MS. Evolution of the Insects. Cambridge, UK: Cambridge University Press; 2005

[5] Graves JAM. Sex chromosome specialization and degeneration in mammals. Cell. 2006;**124**(5):901-914

[6] Henking H. Spermatogenese und deren Beziehung zur Entwicklung bei Pyrrhocoris apterus L. Zeitschrift für wissenschaftliche Zoologie. 1891;**51**: 685-736

[7] McClung CE. The accessory chromosome—sex determinant? The Biological Bulletin. 1902;**3**(1-2):43-84

[8] Stevens NM. Studies in Spermatogenesis with Especial Reference to the "Accessory Chromosome". Washington D.C.: Carnegie Institution of Washington; 1905 http://www.archive.org/details/studiesinspermat01stevrich

[9] Wilson EB. The chromosomes in relation to the determination of sex in insects. Science. 1905;**22**(564): 500-502

[10] Muller HJ. A Factor for the Fourth Chromosome of Drosophila. Science. 1914;**39**:906

[11] Ohno S. Sex chromosomes and sex-linked genes. Berlin, New York etc: Springer-Verlag; 1967

[12] Vyskot B. Y Chromosome Evolution. In: Brenner's Encyclopedia of Genetics. 2nd ed. Vol. 7. USA: Academic Press; 2013. DOI: 10.1016/B978-0-12-374984-0.01659-4

[13] Charlesworth D. Plant sex determination and sex chromosomes. Heredity. 2002;**88**(2):94-101

[14] Dellaporta SL, Calderon-Urrea A. Sex determination in flowering plants. The Plant Cell. 1993;**5**(10):1241-1251

[15] Disteche CM. Dosage compensation of the sex chromosomes and autosomes. In: Seminars in cell and developmental biology. 2016; (Vol. 56, pp. 9-18). Academic Press

[16] Lebedeff GA. Genetics of hermaphroditism in Drosophila virilis. Proceedings of the National Academy of Sciences of the United States of America. 1934;**20**(12):613

[17] Bridges CB. Sex in relation to chromosomes and genes. The American Naturalist. 1925;**59**:127-137

[18] Beye M, Hasselmann M, Fondrk MK, Page RE Jr, Omholt SW. The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell. 2003;**114**(4):419-429

[19] Evans JD, Shearman DC, Oldroyd BP. Molecular basis of sex determination in haplodiploids. Trends in Ecology & Evolution. 2004;**19**(1):1-3

[20] Subramoniam T. Sex Determination (chapter-1). In: Sexual Biology and Reproduction in Crustaceans. USA: Academic Press; 2017. pp. 1-28. DOI: 10.1016/B978-0-12-809337-5.00001-0

Edited by Mahmut Çalışkan

This book provides a glimpse into the dynamic process of genetic polymorphism by presenting studies carried out on different kinds of organisms at the DNA level or gene expression level. Chapters address such topics as genetic polymorphism in animals, gametocyte biomarkers, thrombotic disorders, prostate cancer, and more.

Published in London, UK

IntechOpen

ISBN 978-1-83968-812-6
9 781839 688126