IntechOpen

Methods in Molecular Medicine

Edited by Yusuf Tutar

Methods in Molecular Medicine

Edited by Yusuf Tutar

Published in London, United Kingdom



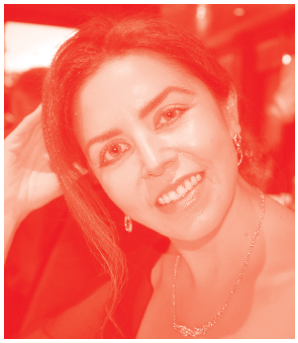




IntechOpen





Supporting open minds since 2005

Methods in Molecular Medicine
http://dx.doi.org/10.5772/intechopen.87364
Edited by Yusuf Tutar

Contributors
Enéas Ricardo Konzen, Maria Imaculada Zucchi, Luiz Santana Da Silva, Marcella Montenegro, Carlos Amaral, Malgorzata Rogalinska, Bhawana Singh, Donald Roy Love, Debra O. Prosser, Osman Ugur Sezerman, Tugce Bozkurt, Fatma Sadife Isleyen, Je-Yoel Cho, Kang-Hoon Lee, Yusuf Tutar, Berçem Yeman Kıyak, Adnan Batman, İrem Yalım Camcı, Elif Kadıoglu, Kezban Uçar Çiftçi, Servet Tunoglu, Ezgi Nurdan Yenilmez Tunoglu, Indu Raja, Kelly Kolkiewicz, Antonio Milano

First published in London, United Kingdom, 2021 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

Methods in Molecular Medicine
Edited by Yusuf Tutar
p. cm.
Print ISBN 978-1-83962-726-2
Online ISBN 978-1-83962-727-9
eBook (PDF) ISBN 978-1-83962-728-6

Meet the editor

Prof. Dr. Yusuf Tutar is the Head of the Biochemistry Division in the Faculty of Pharmacy, University of Health Sciences, Turkey and he is also adjunct faculty at the Molecular Oncology Division of Health Sciences Institutes. He obtained his MSc and Ph.D. degrees at Oregon State University and Texas Tech University, respectively. He pursued his postdoctoral studies at Rutgers University Medical School and National Institutes of Health (NIH/NIDDK), USA. Dr. Tutar's research is mainly focused on biochemistry, biophysics, genetics, and molecular biology with a specialization in the fields of prion, drug design, cancer, protein structure-function, protein folding, microRNA, pseudogenes, molecular cancer, proteomics, genomics, and protein expression and characterization by spectroscopic and calorimetric methods.

Contents

Preface

Understanding disease pathology and physiological pathways depends on research that examines how certain genes become abnormal. Disease processes with personal patterns can be elucidated through molecular medicine. The crux of the cure lies in precise detection and predisposition of the disease. Revealing molecular mechanisms helps create an effective personalized treatment and minimize side effects. Treatment risk and response to therapy prediction can be forecasted through early diagnosis, which improves prognosis reliability, and the effectiveness of therapies. Using this concept, monitoring health is a way to both make predictions and devise prevention. This book covers contemporary advances in molecular markers, disease-causing variants, retroelements, and the basis of distinct diseases.

Dr. Yusuf TUTAR
Professor,
University of Health Sciences,
Turkey

Section 1

Diseases

Chapter 1

Molecular Mechanisms of Distinct Diseases

Adnan Batman, İrem Yalım Camci, Elif Kadioglu, Kezban Uçar Çifçi, Berçem Yeman Kıyak, Servet Tunoglu, Ezgi Nurdan Yenilmez Tunoglu and Yusuf Tutar

Abstract

Molecular medicine describes molecular structures and mechanisms and this chapter focuses on molecular and genetics errors of diseases. Diseases can be classified into deficiency diseases, hereditary diseases, infectious diseases and physiological diseases and to get a glimpse of the mechanisms the chapter covers the most common disease of each class.

Keywords: chromosomal diseases, cancer, neurodegenerative diseases, pulmonary diseases, obesity-induced insulin resistance, lymphoblastic leukemia, viral immunology and infectious diseases

1. Introduction

Distinct diseases have different etiology pattern and this chapter covers the chromosomal diseases, cancer, neurodegenerative diseases, pulmonary diseases, obesity-induced insulin resistance, lymphoblastic leukemia, viral immunology and infectious diseases. These communicable and non-communicable diseases negatively affect structure-function of the organism and specific symptoms are associated with these conditions. Pathogens or internal dysfunctions may lead these diseases. The chapter provides pathology of selected diseases from each class along with the molecular mechanisms.

1.1 Chromosomal diseases

1.1.1 Down syndrome

Down syndrome (DS) is the most common chromosomal genetic disorder. The disease is caused by the trisomy of human chromosome 21 (HSA21) and is also the most genetic mental disability [1]. The HSA21 mosaic can also lead to DS. Maternity age is an important aspect in the formation of an individual with DS [2]. The main cause of this disease is the absence of normal chromosome separation during meiosis and the production of gametes with two copies of chromosome copies instead of a single copy. As a result, DS individuals have trisomy 21 in some body cells, and a normal number of chromosomes in others. This is called mosaicism and is seen in approximately 4% of DS individuals. The term mosaicism was first reported in 1961

[3] and can occur in two ways: either a normal zygote is exposed to an early mitotic error following fertilization, which results in trisomy 21 in some cells, or an early mitotic error in some cells allows it to return to normal karyotype [4].

HSA21 is the most studied human chromosome, and since the long arm of chromosome 21 has been fully sequenced, a significant progress has been made in understanding its functional genomic units. HSA21 is the smallest chromosome and the overall gene density per megabase is about 15 genes per Mb (for the human genome) [5]. HSA21 is also very rich in long encoding RNA (lncRNA) genes, and, one of the poorest for genes encoding microRNA (miRNA). Also, the gene density is average for pseudogenes encoding the protein per Mb [6]. HSA21 is a weak chromosome in non-encoding RNAs (ncRNAs) and long nuclear elements (LINE). Interestingly, HSA21 shows significant enrichment for proteins found in cytoskeleton structures. These cytoskeletal proteins are known to play a role in neurological disorders, especially Alzheimer's neuropathology [7].

Individuals with DS occasionally develop the myeloproliferative disorder (TMD), a disease that is mostly unique to DS. Almost all TMD cases were found to contain somatic mutations on the X chromosome, in the GATA1 transcription factor [8]. Certain features of DS contain genes on other chromosomes causing gene and trisomy mutations and working together to reveal the disorder in HSA21. Studies have shown that the formation of Trisomy 21 precedes the formation of GATA1 mutations [1]. This may indicate that Trisomy 21 either increases genomic discrepancy leading to GATA1 mutations, or it supplies a selected medium for hematopoietic cells containing GATA1 mutations.

1.1.2 Molecular mechanism

Many hypotheses have been proposed to explain the genotype–phenotype relationship in DS. One of these is the 'gene dosage effect' hypothesis putting forward that the phenotypes arise directly from the dosage imbalance of the genes. Overlapping this hypothesis, the 'DS Critical Region' (DSCR) was announced in the 1990s. [9, 10]. Many of the DS features can be called into a subset of the critical genes in the DSCR region, suggesting that DS phenotypes are mainly caused by the dosage imbalance of only a few genes on HSA21. Genomic regions affecting the presence of certain DS phenotypes have been identified and high-resolution genetic maps of DS features have been created [11]. Olson et al. studied the DSCR regions in mice to test its hypothesis. They concluded that dosage imbalance of some individual genes on HSA21 directly affects certain phenotypes, but they stated that more studies are needed [12].

The "amplified developmental instability" hypothesis suggests that dosage imbalance of the HSA21 gene leads to a non-specific impairment of cellular homeostasis [10]. Extra chromosome materials may also contribute to phenotypes by disrupting chromosomal regions. Some data on monozygotic twins for TS21 suggest that differential expression between normal and trisomic twins can be regulated across chromosome domains. This study shows that some DS phenotypes can be enlightened by the modification of the chromatin structure in the nucleus [13]. Monozygotic twins affected by DS but showing incompatible phenotypes have been reported in some cases, suggesting the role of epigenetics in the phenotypic variability of DS. For example, DNA methylation (controlling gen expression) has been shown to change in Trizomy of chromosome 21 (TS21) samples [14].

1.1.3 Turner syndrome

Turner syndrome (TS) is a disorder in mosaic karyotypes associated with complete or partial loss of the X chromosome. Seen especially in women, TS is

associated with short stature, delayed puberty, ovarian dysgenesis, infertility, congenital malformations of the heart, type 1 and type 2 diabetes mellitus, osteoporosis, and autoimmune disorders. It occurs in almost every 2500 live female births. Fetuses affected by TS are 99% estimated to result in fetal death. Approximately half have monosomy X (45, X) and 10% have a repeat (isochromosome) of the long arm of the X chromosome. Most of the rest has a mosaic in more cell lines for 45X. TS, which is associated with a missing X chromosome, was first identified about 100 years ago [15].

Related genes: Shox gene (short length homeobox protein-coding) located on X and Y chromosomes, it is a gene responsible for TS phenotype. This gene does not undergo X inactivation, and a decrease in the expression of SHOX explains some of the TS-related growth deficits. The gene product controls the expression of natriuretic peptide B (NPBB) and FGFR3 (fibroblast growth factor receptor 3) and regulates the proliferation and of chondrocytes, and also cooperates with SOX5, SOX6 and SOX9 and some other genes [16].

The TS genome is hypo-methylated with less hypermethylation sites and there are RNA expression changes that affect the X chromosome genes and autosomal genes compared to women who are 46 XX. Known escape genes are expressed differently in individuals with TS and other X chromosome genes such as RPS4X and JPX (CD40LG and KDM5C) in particularly, KDM5C (encoding lysine-specific demethylase 5C) can participate in the transcriptional profile of neuronal genes and play role in different neurocognitive profiles [17]. 40S ribosomal protein S_4 (RPS_4X) also plays an important role in TS, bringing together multiple protein complexes. In addition, the Y paralog of RPS4X (RPS4Y) may also have a role since it is normally expressed as duplicates [18].

Many different studies show that women with TS have increased mortality compared to the pool of a wide variety of related diseases [19]. The most obvious increase in morbidity is caused by autoimmunities like diabetes mellitus or thyroiditis, osteoporosis, cardiovascular diseases, hypertension, congenital malformations, especially endocrine diseases including heart diseases, digestive system and anemia [20].

1.1.4 Genotype-phenotype

It is still unclear which chromosomal regions or genes make up the phenotypical properties of TS. The physical symptoms of TS were thought to be due to the absence of normal sex chromosomes before inactivation of the X chromosome, or the haplo-insensitivity of the genes in the pseudo-autosomal regions of the aneuploidy [21]. It is thought that a complete phenotype results in the loss of short arm (Xp) in the X chromosome. Aneuploidy itself can cause growth failure. Loss of a region in Xp22.3 was found to be related to neurocognitive problems in TS [22]. Loss of the SRY gene locus in the short arm of the Y chromosome leads to the phenotype of TS, even if it does not cause a population of 45 X cells. It has also been suggested that an area in Xp11.4 is important for the development of lymphedema [23].

1.2 Cellular proliferation: cancer

Cancer can be defined as the uncontrolled cell growth with the most basic explanation. Cell stacks that grow uncontrollably are called tumors. Benign tumors grow much slower and usually do not metastasize, while malignant tumors can spread to other organs through metastasis, and lead to multiple organ damage and eventually death. Tumor cells acquire characteristic features such as sustaining growth signals in the process of cancer, avoiding growth suppressors, resisting cell death,

ensuring replicative immortality, initiating angiogenesis, and activating invasion and metastasis [24].

Cancer cells acquire these abilities in the process due to genetic instability and inflammation caused by environmental and hereditary effects. Many studies show that viruses, in addition to many environmental factors such as radiation and chemicals, induce cancer. Chronic inflammation has been shown to trigger oncogenic mutations, genetic instability, tumor growth, and angiogenesis through angiogenesis and cause local immunosuppression [25].

Two types of gene groups involved in cancer are oncogenes, which trigger cellular growth and uncontrolled proliferation, causing increased genetic instability with increased expression and tumor suppressor genes that cause cancer as a result of decreased control of their expression, cell division, and growth. Proto-oncogenes include RAS, WNT, MYC, ERK, and TRK genes. A mutation that may occur on a proto-oncogene or a regulatory region of the gene (e.g., promoter region) can cause an increase in the amount of protein with the change in protein structure [26]. Expressions of oncogenes can also be regulated with miRNAs [27]. Mutations occurring in these regulatory miRNAs can cause activation of oncogenes [28]. Cancer cells increase cell growth-division by activation of oncogenes, as well as suppress preventive control mechanisms of tumor suppressor genes that control this process.

Mutations in tumor suppressor genes cause loss of function. Therefore, they occur in both alleles. To inactivate the gene and its protein, wide-ranging effects, such as deletions, frame-shift mutations, insertions, should be seen rather than point mutations [29]. Tumor suppressor genes include retinoblastoma (RB) [30], TP53, BRCA1, BRCA2, APC, and PTEN. Many side factors such as transcription complexes, changes in cellular metabolism, microenvironment can guide the course of cancer [31].

The development of cancer is a multi-stage process consisting of initiation, promotion, and progression. Cancer-inducing events are usually caused by genetic mutations. Mutant cell proliferates rapidly in the promotion stage and acquires features that allow malignant behavior in the progression stage. Production of telomerase and expression of p53 are examples of malignant behavior [32]. Then, the process proceeds in the form of dysplasia formation, where new blood vessels are formed (angiogenesis) with cellular transformation. Angiogenesis facilitates the intravasation of cancer cells after undergoing an epithelial-mesenchymal transition (EMT) [33]. EMT gives an invasive phenotype to cancer cells and is managed by various transcription factors (such as SNAI, SLUG, ZEB2, ETS1, TWIST) [34]. These transcription factors also regulate each other for the protection of EMT [35].

1.2.1 Cancer cell metabolism

Normal cells only use anaerobic glycolysis when oxygen is absent or limited, while cancer cells can convert glucose to lactate in the presence of oxygen. Otto Warburg discovered that cancer cells exhibit a differentiated metabolism ability [36]. Warburg effect is biochemical properties that help identify cancer cells. On the other side, cancer cells are generally highly glucose-dependent. Glucose intake of cells is enabled by overexpression of different isoforms of membrane glucose transporters in cancer cells [37]. It has been shown that the benefit of the Warburg effect for cancer cells is not just the formation of glycolytic ATP, but also the production of many glycolytic intermediates before anabolic processes such as NADPH and amino acids [38]. Cancer cells are also able to metabolize glutamine to synthesize some amino acids they need, use it as a nitrogen source and for fatty acid synthesis in hypoxic conditions [39]. Therefore, blood glutamine levels increase

in some cancer cases [40]. Lactic acid is used to produce citric acid and maintain cancer progression in neighboring cancer cells. This is called the "Reverse Warburg effect" [41].

Tumor micro-environment, consisting of fibroblasts, adipocytes, endothelial cells, and macrophages, is a good source for tumor growth. Tumors "steal" energy-rich metabolites from their micro-environment [42]. Monocarboxylate carriers (MCTs) are used for L-lactate transfer between cancer cells and their microenvironment [43]. Tumors have heterogeneous structures with hypoxic and aerobic regions. A "metabolic symbiosis" behavior has recently been found between the two regions [44]. Lactate is produced by glycolysis in hypoxic tumor cells. This product is obtained by aerobic cancer cells by MCT1. Aerobic cells convert lactate to pyruvate with lactate dehydrogenase isoform B (LDH-B) enzyme.

When glucose consumption is not enough to meet the energy need of cancer cells, they begin the fatty acid oxidation (FAO) [45]. For example, prostate cancer, leukemia, and large B-cell lymphoma, increasing palmitate and FAO uptake in cells are among the most commonly used bioenergetic pathways [46–48]. Normal cells usually receive fatty acids by diet, while tumors show an increase in de novo fatty acid synthesis [45].

Pyruvate plays a pivotal role in the regulation of metabolic reprogramming, especially in tumors [49]. Pyruvate dehydrogenase (PDH) converts cytosolic pyruvate into mitochondrial acetyl-CoA, which is the first substrate of the Krebs cycle. Pyruvate dehydrogenase kinase (PDK) negatively regulates PDH. This reaction slides glucose from oxidative to glycolytic metabolism [50]. Lactate dehydrogenase (LDH) is the primary metabolic enzyme converting pyruvate into lactate. LDH plays an important role in arranging food interchange between stroma and tumor. Studies have shown that inhibition of LDH is important for treating advanced carcinomas [51]. Mitochondrial hyperpolarization is a mutual property of several tumor cells [52]. Tumor cells, which have more negative mitochondrial structures, are more selective targets in drug therapies [53].

1.2.2 Brain cancer

Brain tumors are cancer tissues that grow abnormally and prevent the brain or central spinal system from performing its normal functions. Primary brain tumors originating from brain tissue can usually spread only to other parts of the brain, and occasionally to other organs. Tumors that form in another tissue in the body migrate to the brain are called metastatic or secondary brain tumors. These types of tumors occur more frequently than primary brain tumors. They are termed after their tissue of origin [54].

The most prevalent primary tumor types in adults are glioma, astrocytomas, oligodendroglioma, meningioma, schwannoma, pituitary tumors, and central nervous system (CNS) lymphoma.

1.2.3 Genetic background of brain cancer

Retinoblastoma mutations are found in almost 75% of brain tumors and are mostly associated with glioblastoma, and Tp53 mutations are found in more than 80% of advanced gliomas [55]. Primary glioblastomas have EGFR tyrosine kinase mutations, tumor suppressor PTEN gene mutations, DNA repair protein O6-methylguanine-DNA methyltransferase (MGMT) protein abnormalities [56, 57]. While IDH1 mutations in the control mechanism of the citric acid cycle are seen in advanced glioblastomas, IDH2 mutations are usually shown in oligodendroglioma [58]. Mutations in the BRAF oncogene are common in pilocytic

astrocytomas, pleomorphic xanthoastrocytomas, and gangliogliomas [55]. In some glioblastoma tumors, telomere length is maintained by mutations in the TERT promoter and ATRX gene [59].

WHO groups glioma patients based on the presence of two genetic changes; first, mutations [60] in the family of genes encoding isocitrate dehydrogenase (IDH), and second, loss of two specific parts of the genome (1p and 19q co-deletion) [61]. The presence or absence of these changes gives a clue about the patient's prognosis and appropriateness of various kinds of treatments.

Approximately 40% of people with astrocytoma, oligodendroglioma, or IDH mutation bear a hereditary variation. This variation is a single nucleotide polymorphism (SNP) in the 8q24 region of the genome [62]. There is another SNP in the 11q23 region, which enhances the risk of IDH-mutant brain cancer. Approximately 5–8% of gliomas are familial, POT1 gene mutations have been found in 6 of 300 families with glioma [63].

Non-coding RNAs (ncRNAs) play important roles in regulating tumor malignancy in glioma [64–66]. According to healthy brain tissue, mir-21 expression increases in glioma and mir-21 acts as an oncogene [67, 68]. It has been reported that mir-124 and mir-137 act as tumor suppressors in glioblastoma multiform cells [69]. Hotair, SOX2ot, CRNDE, Malat1, H19, GAS are lncRNAs that have been recently shown to regulate glioma [70, 71]. Glioma cells also express the circRNAs, for example, circBRAF, bircFBXW7, circSMARCA5. These regulate proliferation, migration, and invasion of glioma cells [72–74]. The exosomal ncRNAs, mir-21, mir-148a, lncRNA PU03F3, lncRNACCAT2 can be used as circulating biomarkers of glioma patients [75–78]. circRNAs and the exosomal ncRNAs were also reported as potential biomarkers for the diagnosis and prognosis of glioma patients.

1.3 Molecular pathology of neurodegenerative diseases with Alzheimer's disease in focus

As a characteristic of almost all neurodegenerative diseases, abnormal protein assembly gathers these diseases under the prion concept [79]. Prion protein, known as PrP, was introduced to define protein pathogens and distinguish them from viruses and was identified as a proteinaceous infectious particle known to resist inactivation. Even back at that time, its importance was foreseen in terms of shedding light on the etiologies of chronic degenerative diseases [80]. Self-propagation is an important characteristic of prions that is also observed in abnormal protein assembly in Alzheimer's Disease (AD) [81, 82]. Aggregation of proteins in neurodegenerative diseases was believed to occur spontaneously in autonomous cells, however, it was later understood that this aggregation begins in a particular region and propagates across other regions developing the disease further. Transmission of these prion proteins across neuronal cells takes place trans-synaptically [82].

As described more than a 100 years ago, abnormal protein assembly forms the basis of neurodegeneration with AD being one of the most common neurodegenerative diseases. The pathology of abnormal protein assembly starts with misfolding of native proteins that gather to form seeds which eventually lead to aggregation and development of protein fibrils. The pathophysiology of AD involves amyloid plaque inclusions of β-amyloid (Aβ) peptides and neurofibrillary lesions of tau protein. Tau inclusions may also be characteristics of other neurodegenerative diseases, which do not necessarily show the same implications. Altering the native forms of this protein may contribute to its pathology and cause damage to its host cell.

Most cases of this disease are sporadic, while dominantly inherited mutations are also seen to a lesser extent. Back in the 1990s, missense mutations of APP, encoding amyloid precursor, were shown to cause AD [83–87]. Mutations in this

gene also increase the aggregation tendency of encoded proteins. Many studies have demonstrated phenotypes associated with neurodegeneration when this protein is overexpressed.

There are six isoforms of microtubule-associated protein tau ranging from 352 to 441 amino acids, encoded by the MAPT gene as a result of alternative mRNA splicing. One half has three repeats and the other has four repeats, altogether establishing the microtubule-binding domain and also the core of tau filaments in case of pathology [88]. All isoforms have been observed in the brains of AD patients. Diseases that have isoforms with only three or four repeats, but not both, lack the Aβ peptides seen in AD and therefore do not carry the symptoms specific to the disease [70]. Tau inclusions may be of a variety of conformations, which can also be caused by different mutations on the MAPT gene, explaining the existence of numerous tauopathies [89–93].

Aβ peptides are encoded by the amyloid precursor protein gene, APP, and are widely expressed as type 1 transmembrane glycoproteins. As a result of alternative mRNA splicing, there are three major transcripts named APP695, APP751, and APP770 [94, 95]. β- and γ-secretase enzymes take part in the production of Aβ peptides in sequential endoproteolytic cleavage. β-secretase is responsible for cleaving the N-terminus of the peptide thus removing the portion that remains on the extracellular side. This cleaved peptide is endocytosed and intracellular aggregation builds up which is later released into the extracellular space [79]. γ-Secretase is a membrane-embedded enzyme that is able to cleave many transmembrane proteins including C-terminus of the Aβ peptide. It a complex enzyme of four proteins; presenilin (PS) forming the catalytic core, presenilin enhancer-2 (Pen-2) enabling maturation of PS, anterior pharynx-defective (Aph-1) stabilizing the complex, and nicastrin possibly being the receptor for the enzyme's substrate [96, 97]. PS and Aph-1 each have two variants resulting in at least four different enzyme complexes, which give rise to various cleaved Aβ peptides. Additionally, γ-secretases cleave the peptide in three different sites. Different protein variants and cleavage sites produce Aβ peptides of different profiles, some of which may be more susceptible to aggregation [98].

Overall, it is important to target the pathways leading to abnormal protein assembly and only then treatments may be proposed based on these mechanisms. Once the first protein inclusion is formed, it is essential to keep an eye on the time frame until the disease symptoms come forth. When techniques sensitive enough to catch the first protein inclusion are developed, then tracking its transformation into filaments can be helpful in designing novel preventive approaches. Understanding this cascade will also contribute to planning more efficient therapeutic methods.

1.4 Cellular and molecular mechanisms of asthma and COPD

Asthma and chronic obstructive pulmonary disease (COPD) are common disorders characterized by progressive chronic inflammation in the lungs. They have unique characteristics with dissimilarly involved cells, mediators, and inflammation. They also have distinct responses to corticosteroid treatment. Roughly 15% of COPD patients have characteristics of asthma [99]. Also, a comparable ratio of asthma patients has traits of COPD that is currently the fifth leading cause of death worldwide [100]. Many risk factors are linked to COPD including smoking tobacco, air pollution, indoor cooking while tobacco smoking (including passive smoking) making up around 80% of the cases [101]. There are many types of cells and mediators that have a significant effect during the pathogenesis of asthma and COPD.

Macrophages have a crucial role in coordinating the inflammatory response activated by cigarette smoke extract in COPD cases [102]. They discharge inflammatory

mediators including tumor necrosis factor (TNF)-α, IL-8, other CXC chemokines, monocyte chemotactic peptide (MCP)-1, LTB4 and reactive oxygen species (ROS) [103]. However, the role of macrophages in asthma is not certain. Allergens via low-affinity IgE receptors may activate macrophages causing an inflammatory response through the discharge of a definite arrangement of cytokines. On the other hand, macrophages also excrete anti-inflammatory mediators, such as IL-10 that is thought to decrease in subjects with intense asthma [104].

Activated neutrophils were shown to be enhanced in some subjects with severe asthma and COPD in their sputum and airways [105]. Among the serine proteases secreted by neutrophils are neutrophil elastase (NE), cathepsin G, proteinase-3, matrix metalloproteinase (MMP)-8 and MMP-9, leading to alveolar destruction [103]. The mechanisms of neutrophilic inflammation in asthma and COPD are not clear. Demonstration of priming in COPD occurs at neutrophils in the peripheral circulation. Many chemotactic signals exhibit the capacity for neutrophil recruitment in COPD. These include LTB4, IL-8 and related CXC chemokines, comprising GRO-α (growth-related oncoprotein) and ENA-78 (epithelial neutrophil activating protein of 78 kDa) which are enhanced in COPD airways [106]. Although the mentioned mediators might be sourced from alveolar macrophages and epithelial cells, neutrophils have the capacity of being a vital source of IL-8 [107].

Airway and alveolar epithelial cells in COPD can be a vital point of source of inflammatory mediators and proteases 5. Cigarette smoke activates epithelial cells which produce inflammatory mediators, including TNF-α, IL-1β, GM-CSF and IL-8 [108]. Epithelial cells play an important role in airways defense and tissue repair processes. Goblet cells, a type of epithelial cell, in mucus catch bacteria and inhaled particulates [109]. Epithelial cells release antioxidants and antiproteases. Immunoglobulin A is carried by epithelial cells, hence involved in adaptive immunity [110]. On a side note, native and adaptive immune reactions of the airway epithelium are triggered by cigarette smoke and damage by other harmful agents, increasing sensitivity to infection.

The main role of dendritic cells is to introduce innate and adaptive immune reaction by activating macrophages, neutrophils, T and B lymphocytes among others [103].

Lymphocytes are directly involved in the pathogenesis of both asthma and COPD. Both airway and parenchymal inflammation exist in asthma and COPD patients [111]. Most lung lymphocytes are T cells which are in the respiratory tract of ordinary humans. Activated T lymphocytes are characteristic in both asthma and COPD, but CD4+ type-2 T lymphocytes are the major player in asthma whereas CD8+ type-1 lymphocytes are specific to COPD [111]. CD4+ T lymphocytes can generate many cytokines involved in mediating cell functions and cell–cell communications. This is done through impressing physiologic cell properties such as proliferation, differentiation and activation of other immunocompetent cells, chemotaxis, and connective tissue metabolism [112]. On the other hand, CD8+ T lymphocytes exist in the respiratory mucosa and are activated in response to foreign antigens [111]. Specifically, the cells in the respiratory mucosa have an important role in anti-viral immunity. Another lymphocyte type is B cells which are the minority (<5%) lymphocytes. The main function of B cells located in the lungs is the production of immunoglobulins for local defense mechanisms [113].

Apart from these mentioned cells, there are crucial molecular mediators in the pathogenesis of asthma and COPD. The first family of mediators is transforming growth factor (TGF) family. The TGF-β subfamily is composed of five parts that exhibits plenty of effects pertaining to asthma and COPD. A recent study shows that overexpression of TGF-β1 in mice causes Smad3-dependent pulmonary expression of procollagen, antiproteases and fibrosis [114]. TGF-β exhibits

chemotactic signatures for monocytes, macrophages and mast cells. Research shows an abnormal pulmonary expression of TGF-β1 in subjects suffering from COPD. Protein and mRNA expression of TGF- β1 are abundant in the lung tissue, including airway epithelial cells, of mild to moderate COPD patients. TGF-β1 has the role in pathogenesis of COPD because of its increased expression in parallel to the number of macrophages [115].

Another mediator family is the fibroblast growth factor (FGF) family with 23 members in humans. Their functional receptors are named from FGFR1 to FGFR5 [116]. FGFs have many functions such as development, tissue homeostasis, and repair. In addition to further growth factors, FGF-1, FGF-2, and FGF-7 and their receptors FGFR1 and FGFR2, are located abundantly in the lungs [101]. Research shows that increased expression degrees of FGF-1, FGF-2, and FGFR1 were detected in vascular and epithelial areas in the lungs of COPD patients. FGF-1 causes higher collagenase expression and lower collagen I expression in lung fibroblasts which prompt tissue remodeling.

Another family of mediators is the vascular endothelial growth factor (VEGF) family. There are seven units in this family capable of attaching to related cellular receptors. VEGFs have many functions including paracrine acting, angiogenic factors, prompting mitogenesis, emigration, and permeabilization of the vascular endothelium [101]. VEGF and its receptors assist in tissue remodeling as well as disease intensity in incessant lung diseases such as asthma [117]. COPD patients have increased pulmonary VEGF expression in bronchial and alveolar epithelial located around the vascular smooth muscle and alveolar macrophages. Additionally, unlike healthy subjects, COPD patients exhibit elevated levels of VEGFR-1 and VEGFR-2 expression inside the endothelium [118]. Furthermore, VEGFR-2 and VEGF expressions are decreased in COPD patients. Compared to VEGFR-2, VEGFR-1 has a higher affinity for VEGF which leads to VEGFR-1 scavenging VEGF from VEGFR-2. This phenomenon culminates VEGFR-1 activation and in the case of endothelial apoptosis, increased MMP activity as well as vascular and alveolar decimation [101]. This suggests the importance of harmony among VEGF, VEGFR-1, and VEGFR-2 during the pathogenesis of COPD subordinary types.

Finally, cytokines and chemokines are mediators supplying a chemotactic gradient which has the potential to activate macrophages, CD8+ T cells and neutrophils for COPD patients. It is known that inflammatory cells of both native and gained immune systems are significant in the COPD pathophysiology. This is where cytokines and chemokines are the key drivers [103, 119]. Different types of cytokines arrange chronic inflammation in asthma and COPD. T2 cytokines which are IL-4, IL-5, IL-9 and IL-13 interfere with allergic inflammation. Other types of cytokines including TNF-α and IL-1β accelerate the inflammatory response [120]. In asthma and COPD patients, chemokines are instrumental in drawing inflammatory cells from the circulation into the lungs [121].

1.5 The endocrine system

1.5.1 Molecular mechanism of obesity and obesity-induced insulin resistance

Obesity is a serious health problem that has become epidemic all over the world, especially in developed countries. It is characterized by hypertrophied adipocytes that secrete various adipokines and hormones, chronic inflammation in all tissues, and systemic insulin resistance resulting in type 2 diabetes, hypertension, and hyperlipidemia. In addition to these metabolic diseases, it can cause diseases such as cancer, atherosclerosis, obstructive sleep apnea syndrome, steatohepatitis, and musculoskeletal problems [122]. The obesity rate is 20% in women and 18% in men

in developed countries [123]. It affects complex metabolic pathways in all tissues as a result of chronic and progressive inflammation, leads to insulin resistance, endothelial dysfunction and lipotoxicity.

The pathophysiology of obesity includes complex interactions of numerous adipokines, hormones and pro-inflammatory cytokines with the central nervous system and metabolic organs (such as liver, pancreas, and muscle) as a result of genetic-environmental interactions.

Genetic etiology: Obesity is generally present in a polygenic etiology. Many studies have investigated the genetic background of body mass index (BMI) and waist/hip ratio (WHR), which are the best measurements of obesity. The results of these studies have been presented collectively in genome-wide association studies (GWAS) [124]. Although, single gene defects (monogenic) are rare in obesity, including especially melanocortin-4 receptor, leptin and leptin receptor genes [125].

Dysregulation in hypothalamic control: The center of food intake and energy regulation in the central nervous system is the arcuate nucleus (ARC) in the hypothalamus besides the autonomic nervous system and brain stem. The balance between the opposing effects of orexigenic and anorexigenic neurons is important. Agouti-related protein (AgRP) and neuropeptide Y (NPY) (AgRP/NPY) neurons are orexigenic that promotes appetite and eating. Pro-opiomelanocortin–producing (POMC) peptide and cocaine-and-amphetamine–regulated transcript (CART), collectively known as POMC/CART neurons are anorexigenic that suppress appetite and eating. Oxygenic pathways that increase energy balance become more effective in obesity [126].

Adipose tissue dysfunction and systemic inflammation; The most important pathophysiological mechanisms of obesity and obesity-related insulin resistance are adipocyte dysfunction (visceral adipose tissue; VAT) and low-grade chronic systemic inflammation. In particular, white adipocyte tissue in obese subjects contributes to the regulation of food intake, energy metabolism and other functions by secreting adipokines from adipose tissue, which provide the necessary signals to the central nervous system, hypothalamus, liver, pancreas, muscle tissue, and other systems to regulate appetite, food intake, and energy balance [125]. Leptin is the most important adipokine that stimulates anorexigenic POMC/CART neurons and induces production of pro-inflammatory cytokines (TNF-alpha and IL-6) by macrophages and monocytes. In the case of hyperleptinemia, leptin resistance develops by the inhibition of the JAK2/STAT3 signaling pathway, which later increases oxidative stress and inflammation, causing insulin resistance, hyperlipidemia and hypertension [127]. Resistin is a pro-inflammatory adipokine produced by the resistin gene (RETN), which activates SOCS3, causing the insulin signaling pathway to be inhibited and consequently induces insulin resistance [128]. Other adipokines like Retinol binding protein 4 (RBP4), Angiopoietin-like protein 2 (ANGPTL2), Visfantin, Adiponectin, Lipocalin 2, Serum Amyloid A, Angiotensinogen, Renin, Angiotensin-Converting Enzyme, Acylation-Stimulating Protein, and Vaspin, are increased, and adiponectin, and Apelin are decreased in obesity, altogether stimulating inflammation, lipolysis, releasing free fatty acid (FFA) and causing insulin resistance as a result [129].

Gastrointestinal hormones and microbiota: Gastrointestinal hormones and gut microbiota play a significant role in the complex pathophysiology of obesity. Ghrelin produced in the stomach induces starvation and food intake by stimulating orexigenic AgPR/NPY neurons in the hypothalamus. Although the effect of ghrelin cannot be fully explained, it is thought to increase in obesity, stimulate growth hormone release (GH), increase gastrointestinal motility and insulin secretion [130]. Decreased GLP-1, Peptide YY, pancreatic polypeptide, and increased amylin and cholecystokinin cause appetite inhibition and gastric emptying delay, resulting in excess energy [129]. Besides hormones in the gastrointestinal tract, changes in

microbiota-gut-brain axis and their effects on metabolic organs are also important. Occurring as a result of nutrition and gene–environment interactions; chronic systemic inflammation resulting from intestinal microbiota dysbiosis (increase in Firmicutes-Bacteroides ratio), microbial fermentation products, increase in short-chain fatty acid formation and intestinal permeability, decrease in butyrate-producing bacteria rate, leads to an increase in proinflammatory response in metabolic organs, impaired fat metabolism and glucose metabolism [131, 132].

İmpaired insulin sensitivity and oxidative stress; The beginning of insulin resistance is the first step in the pathophysiology of T2D. Anabolic effects such as glycogen and protein synthesis, glucose transport, adipogenesis are formed by phosphatidylinositol-3-kinase (PI3K)/Akt pathway activation as a result of insulin binding to its receptor (INSR) synthesized in the pancreas [133]. On the other hand, insulin shows mitogenic effects with mitogen-activated protein kinases/Ras pathway (MAPK/Ras).

Adipokines, FFA's, pro-inflammatory cytokines (TNF-a, IL-18, IL-1β, IL-6), synthesized as a result of inflammation in adipose tissue in obesity, also cause systemic inflammation in metabolic tissues such as liver and muscle. As a result, decreased GLUT-4 expression, activation of Ser/Thr kinases with insulin receptor substrate (IRS) phosphorylation, production of ceramides and proinflammatory cytokines, suppressing of cytokine signaling-3 (SOCS-3) expression, insulin pathways and effects. On the other hand, increased production of reactive oxygen radicals and production of toxic doses NO with inducible nitric oxide synthase (iNOS) activation, affect mitochondrial and endoplasmic reticulum functions. Activation of pro-inflammatory pathways increased oxidative stress, mitochondrial dysfunction, ER stress affects lipid metabolism, insulin mechanisms of action and other metabolic pathways, causing insulin resistance, Type 2 diabetes, hypertension, and hyperlipidemia [134].

Beta-cell dysfunction: In addition to peripheral insulin resistance in obesity, serious reductions in beta cell function are also observed. An increase in fat accumulation in islet cells due to chronic lipotoxicity disrupts the function of beta cells by blocking calcium channels. Chronic hyperglycemia due to disruption in glucose metabolism and systemic inflammation due to an increase in oxidative damage and lipotoxicity, disrupt insulin secretion pathways and cause changes in apoptosis gene expression. Hyperinsulinemia in obesity, impaired insulin signaling pathway, oxidative stress, lipotoxicity in islet cells, loss of beta-cell function and apoptosis may lead to the formation of type 2 diabetes [122, 135].

Obesity has become a pandemic all over the world as a result of rapidly changing lifestyles and genetic heritage in the last century. Despite the findings in recent studies on the development and complications of obesity, it is difficult to say that the subject of etiology and pathophysiology is still not fully understood. Especially omics technologies, big data on environmental gene interactions, neuroendocrinology, and neuropsychological studies will reveal findings that open up different horizons. However, due to its complications from deadly metabolic diseases to cancer, rapid preventive measures should be taken, and effective treatment models should be developed.

1.6 Alterations of blood cells: lymphoblastic leukemia

Acute lymphoblastic leukemia (ALL) is a heterogeneous malignancy emerging from lymphoid precursors. It is characterized by the proliferation of immature lymphoid cells with somatic mutations including chromosomal rearrangements, and aneuploidy [136]. ALL has two peak points; first point occurs at ~5 years of age (80%), and the second point occurs at the age of ~50 (20%) [137]. The basic mechanism underlying the development of ALL is similar in children and adults,

while they have the frequency of different genetic subtypes. Molecular analysis of genetic changes in leukemia disease provides a great advantage in order to understand prognosis and pathogenesis of ALL [138].

The diagnosis of ALL depends on the presence of at least 20% lymphoblast in bone marrow. Immunophenotyping by flow cytometry (FCM) identifies the subtype of ALL that may be B-cell precursor (BCP), mature B-cell types, or T-cell ALL. Chromosomal abnormalities are a characteristic of lymphoblastic leukemia, which are found in B or T cell lineage. The most common abnormality found in adult B precursor ALL is the t(9;22) BCR-ABL translocation, while the t(12;21)(p13;q22) TEL-AML1 translocation is most commonly found in childhood B precursor ALL [139]. On the other hand, the discovery of mutations in the receptor tyrosine kinase FLT3 contributes to the understanding of leukemogenesis mechanism in hyperdiploid ALL (20% of cases). Based on this finding, targeting specific tyrosine kinase inhibition may be useful in the management of leukemia [140].

1.6.1 Treatment in patients with lymphoblastic leukemia

Small-molecule kinase inhibitors have a clear benefit in the treatment of many cancer types including leukemia. Imatinib mesylate, a small-molecule inhibitor of BCR-ABL kinase, is highly effective in the treatment of chronic myelogenous leukemia (CML) [141]. Although the single kinase inhibitor is a remarkable treatment option in a different type of leukemia, it will need to be combined with either other targeted therapy or chemotherapy because of the resistance to small-molecule inhibitor [142]. Unlike ALL, Chronic Lymphocytic Leukemia (CLL) is defined the accumulation of monoclonal B cell with a special immunophenotype in the bone marrow, blood, and other lymphoid organs where B lymphocytes express CD19, CD23, CD5, low-level CD20 and surface immunoglobulins [143]. The standard treatment procedure of ALL and CLL includes consolidation therapy following chemotherapy in pediatric patients. For adult patients, unlike pediatric patients, the allogeneic hematopoietic stem cell transplantation is frequently preferred as consolidation therapy [144]. Because the patients resistant to chemotherapy are not respond to treatment well enough, novel therapy approaches such as Chimeric antigen receptor-modified T cell (CAR-T) therapy have developed in order to overcome chemotherapy resistance and improving the outcome of patients [145]. CAR-T cells, as immunotherapeutic tools, are genetically engineered to express a chimeric antigen receptor recognizing an antigen that is located in the special cells such as a tumor [146]. CD19 antigen on B lymphocytes was considered the initial target for CAR-T cell therapy. However, specific antigen loss might cause the failure of CAR-T cell therapy in CLL. CD19–20 co-targeting CAR-T cells were designed to kill both CD19-positive and CD19-negative CLL and it was shown that these cells were very effective in killing CLL cells. In one of the first reported in pediatric ALL the clinical trials, CAR-T cells targeted the CD19 antigen of B cells are designed with CD3ζ and CD28 costimulatory domain [147].

1.7 Viral immunology and infectious diseases

1.7.1 The differences between HIV-1 and SARS-CoV-2 genome

The origin of a pathogen has a crucial role in developing vaccines and blocking transmission. This may last many years due to its elusiveness as seen in HIV-1, SARS, and MERS [148–150]. According to a recent report, it was emphasized that

SARS-CoV-2 is able to infect T cells, which are targeted by HIV [151]. Another report alleged that the motif insertions of spike glycoprotein, similar to HIV-1, may help increase the range of host cells of SARS-CoV-2. HIV-1 envelope glycoprotein contains mutable insertions and deletions not necessary for biological function. Only 1 and 2 insertions are matched in only a few HIV-1 strains and this reveals that four insertions are scarce. Thus, HIV-1 cannot be assumed as the source for those insertion sequences in the SARS-CoV-2 genome due to their inefficient identities and scarceness in the HIV-1 sequences [152].

1.7.2 The origin of SARS-CoV2

The reported cases showed that there have been 3,162,284 COVID-19 cases in at least 212 countries and approximately 7.1% of which was resulted in death as of April 30, 2020 [153]. It is known that SARS-CoV, MERS-CoV, and SARS-CoV-2 are the members of coronoviridae family of the Nidovirales order, which comprises a relatively positive-sense, single-stranded RNA genome of around 26–32 kb [154]. 5o-methylguanosine cap at the beginning, a 3o-poly-A tail at the end, and a total of 6–10 genes in between exist in their genome [155, 156].

This family has extremely expressive instability and recombination rate, which is similar to RNA viruses, so it is practically unfeasible to prevent their distribution among humans and animals worldwide; nevertheless, the fact that the virus is exceedingly pathogenic to humans is closely related to random genetic recombination in the host. Although there is a strict genetical relation between SARS-CoV-2 and SARS-CoV, it is explicit that SARS-CoV-2 has a unique feature providing rapidly spread worldwide [157].

SARS-CoV-2 genome sequence is much more resembles a SARS-like bat rather than SARS-CoV [158, 159]. Two open reading frames translating the replication- and transcription-related gene into two large non-structural polyproteins [156]. Ribosomal frameshifting contributes to translate two different but overlapping open reading frames. Besides these nonstructural proteins, the subgenomic RNA also encodes the viral genome packaging protein N (nucleocapsid), and the viral coating proteins M (membrane), E (envelope), and S (spike) as the structural proteins. Viral coating proteins, which interact with host surface receptors, is generally preferred as the therapeutic target blocking protein–protein interaction [160, 161]. TMPRSS2, the human serine protease, enables S Protein of both SARS-CoV and SARS-CoV-2 to prime, and these two viruses use the angiotensin-converting enzyme 2 (ACE2) receptor in order to bind the host cell as the first step of the viral entry mechanism. Unlike SARS-CoV and SARS-CoV-2, the cell entry of MERS-CoV depends on the binding of its own spike protein to DPP4 (dipeptidyl peptidase 4). The RT-PCR analysis of the throat swabs is essential to the diagnosis of COVID19 pneumonia, and it takes 3.5 h to provide the results [162].

1.7.3 Current treatment approaches

Clinical management puts emphasis on the importance of supportive care and prevention of complications due to a lack of specific treatment for COVID-19 pneumonia. On the other hand, potential antiviral therapies for the purpose of rapidly dealing with this pandemic are taking place on several clinical trials. These trials focused on three main targets that include enhancing the host immune system, blocking the virus spike protein-host cell surface receptor interaction, and vaccine development [163].

1.7.4 Human Papillomavirus genome and treatment

HPV genome, which is a double-stranded circular DNA, has the early (E) genes that are responsible for replication and transcription, and the late (L) genes that are responsible for viral capsid proteins. In the early stage of HPV infection, the highly expressed E1 and E2 proteins provide the maintaining of viral replication and transcription within the cervical cell [164].

HPVs, unlike SARS coronaviruses, are non-enveloped viruses and don't have a specific host cell receptor that initiates the viral infection. Additionally, HPVs have many different genotypes such as HPV type 16 and type 18 which are known as the reason for cervical cancer. HPV infection may cause low-grade cytological changes on Papanicolaou smears, or low-grade squamous intraepithelial lesions [165]. When malignant conversion considered, viral oncoproteins E6 and E7 attach, respectively, tumor suppressor protein p53 and Rb have a crucial role [166]. Until today, many vaccine developments studies have been carried out to protect HPV malignant type 16 and 18. For example, the clinical vaccine Gardasil 9 provides effective protection against vaginal, cervical, and vulvar diseases caused by HPV type 16,18 and also its 5 other different types [167].

2. Conclusion

The chapter outlined the unique mechanism of each disease. Depending of the origin of the disease; deficiency, hereditary, infectious and physiological diseases may be treated diversely but the perturbation effect can only be eliminated with proper intervention. Current amelioration may be improved by biochemical methods only if the molecular mechanism is clearly understood. Therefore, molecular medicine provides unique solutions to diagnose and treat disease by elucidating macromolecular interaction and abnormalities in cells and tissues. The chapter summarizes current findings and methods to alleviate and cure the diseases.

Acknowledgements

BYK, EK, KUC and ENYT acknowledge YOK100/2000 bursary and thanks to Turkish Council of Higher Education (YOK).

Conflict of interest

None.

Notes/thanks/other declarations

Thanks to Assistant Prof. Dr. Lütfi Tutar for carefully reading the manuscript.

Author details

Adnan Batman[1], İrem Yalim Camci[2], Elif Kadioglu[1], Kezban Uçar Çifçi[1], Berçem Yeman Kıyak[1], Servet Tunoglu[3], Ezgi Nurdan Yenilmez Tunoglu[1] and Yusuf Tutar[1,4]*

1 Division of Molecular Medicine, University of Health Sciences, Hamidiye Health Sciences Institute, 34668, Istanbul, Turkey

2 Department of Molecular Biology and Genetics, Gebze Technical University, Kocaeli, 41400, Turkey

3 Department of Molecular Medicine, Institute of Health Sciences, Istanbul University, 34093, Istanbul, Turkey

4 Department of Basic Pharmaceutical Sciences, Division of Biochemistry, University of Health Sciences, Hamidiye Faculty of Pharmacy, 34668, Istanbul, Turkey

*Address all correspondence to: yusuf.tutar@sbu.edu.tr; ytutar@outlook.com

References

[1] D. Patterson, "Molecular genetic analysis of Down syndrome," Hum. Genet., vol. 126, no. 1, pp. 195-214, 2009.

[2] T. F. Williams and A. J. Dalton, "Dementia and aging adults with intellectual disabilities: A handbook," Dement. Aging Adults with Intellect. Disabil. A Handb., pp. 1-488, 2014.

[3] C. M. Clarke and J. H. Edwards, "21-Trisomy / Normal," pp. 1028-1030, 1961.

[4] A. Kuliev, Z. Zlatopolsky, I. Kirillova, J. Spivakova, and J. Cieslak Janzen, "Meiosis errors in over 20,000 oocytes studied in the practice of preimplantation aneuploidy testing," Reprod. Biomed. Online, vol. 22, no. 1, pp. 2-8, 2011.

[5] M. Hattori, A. Fujiyama, and Y. Sakaki, "The DNA sequence of human chromosome 21," Tanpakushitsu Kakusan Koso., vol. 46, no. 16 Suppl, pp. 2254-2261, 2001.

[6] A. Letourneau and S. E. Antonarakis, "Genomic determinants in the phenotypic variability of Down syndrome," Prog. Brain Res., vol. 197, pp. 15-28, 2012.

[7] F. I. M. Craik and R. S. Lockhart, "Levels of Processing and Zinchenko's Approach to Memory Research," J. Russ. East Eur. Psychol., vol. 46, no. 6, pp. 52-60, 2008.

[8] O. Tunstall-Pedoe et al., "Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations," Blood, vol. 112, no. 12, pp. 4507-4511, 2008.

[9] J. R. Korenberg et al., "Down syndrome phenotypes: The consequences of chromosomal imbalance," Proc. Natl. Acad. Sci. U. S. A., vol. 91, no. 11, pp. 4997-5001, 1994.

[10] M. Rachidi and C. Lopes, "Mental retardation in Down syndrome: From gene dosage imbalance to molecular and cellular mechanisms," Neurosci. Res., vol. 59, no. 4, pp. 349-369, 2007.

[11] K. J.O. et al., "The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies," Proc. Natl. Acad. Sci. U. S. A., vol. 106, no. 29, pp. 12031-12036, 2009.

[12] L. E. Olson, J. T. Richtsmeier, J. Leszl, and R. H. Reeves, "A chromosome 21 critical region does not cause specific down syndrome phenotypes," Science (80-.)., vol. 306, no. 5696, pp. 687-690, 2004.

[13] J. Salvador, M. Arigita, E. Carreras, A. Lladonosa, and A. Borrell, "Evolution of prenatal detection of neural tube defects in the pregnant population of the city of Barcelona from 1992 to 2006," Prenat. Diagn., vol. 31, no. 12, pp. 1184-1188, 2011.

[14] K. Kerkel et al., "Altered DNA methylation in leukocytes with trisomy 21," PLoS Genet., vol. 6, no. 11, 2010.

[15] N. C. Lonberg and J. Nielsen, "Letters to the Editors Sere evskij-Turner's Syndrome or Turner's Syndrome," vol. 364, pp. 363-364, 1977.

[16] M. Fukami, A. Seki, and T. Ogata, "SHOX Haploinsufficiency as a Cause of Syndromic and Nonsyndromic Short Stature," Mol. Syndromol., vol. 7, no. 1, pp. 3-11, 2016.

[17] C. Trolle et al., "Widespread DNA hypomethylation and differential gene expression in Turner syndrome," Sci. Rep., vol. 6, 2016.

[18] S. N. Rajpathak, S. K. Vellarikkal, A. Patowary, V. Scaria, S. Sivasubbu, and D. D. Deobagkar, "Human 45,X fibroblast transcriptome reveals distinct differentially expressed genes including long noncoding RNAs potentially associated with the pathophysiology of turner syndrome," PLoS One, vol. 9, no. 6, 2014.

[19] J. Rovet, "Turner syndrome: A review of genetic and hormonal influences on neuropsychological functioning," Child Neuropsychol., vol. 10, no. 4, pp. 262-279, 2004.

[20] C. H. Gravholt, S. Juul, R. W. Naeraa, and J. Hansen, "Morbidity in Turner syndrome," J. Clin. Epidemiol., vol. 51, no. 2, pp. 147-158, 1998.

[21] F. Haverkamp et al., "Growth retardation in Turner syndrome: Aneuploidy, rather than specific gene loss, may explain growth failure," J. Clin. Endocrinol. Metab., vol. 84, no. 12, pp. 4578-4582, 1999.

[22] A. R. Zinn and J. L. Ross, "Molecular analysis of genes on Xp controlling Turner syndrome and premature ovarian failure (POF)," Semin. Reprod. Med., vol. 19, no. 2, pp. 141-146, 2001.

[23] C. A. Boucher, C. A. Sargent, T. Ogata, and N. A. Affara, "Breakpoint analysis of Turner patients with partial Xp deletions: Implications for the lymphoedema gene location," J. Med. Genet., vol. 38, no. 9, pp. 591-598, 2001.

[24] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," Cell, vol. 100, no. 1, pp. 57-70, 2000.

[25] G. SI, G. FR, and K. M, "Immunity, Inflammation and Cancer," Cell, vol. 140, pp. 883-899, 2010.

[26] R. Todd and D. T. Wong, "Oncogenes.," Anticancer Res., vol. 19, no. 6A, pp. 4729-46, 1999.

[27] M. Negrini, M. Ferracin, S. Sabbioni, and C. M. Croce, "MicroRNAs in human cancer: From research to therapy," J. Cell Sci., vol. 120, no. 11, pp. 1833-1840, 2007.

[28] A. Esquela-Kerscher and F. J. Slack, "Oncomirs - MicroRNAs with a role in cancer," Nat. Rev. Cancer, vol. 6, no. 4, pp. 259-269, 2006.

[29] J. Mendelsohn, P. M. Howley, M. A. Israel, J. W. Gray, and C. B. Thompson, "The Molecular Basis of Cancer," Mol. Basis Cancer, 2008.

[30] D. L. Burkhart and J. Sage, "Cellular mechanisms of tumour suppression by the retinoblastoma gene," Nat. Rev. Cancer, vol. 8, no. 9, pp. 671-682, 2008.

[31] S. Masri and P. Sassone-Corsi, "The emerging link between cancer, metabolism, and circadian rhythms," Nat. Med., vol. 24, no. 12, pp. 1795-1803, 2018.

[32] R. G. McKinnell, R. E. Parchment, A. O. Perantoni, G. B. Pierce, and I. Damjanov, The Biological Basis of Cancer. Cambridge: Cambridge University Press, 2006.

[33] J. T. Buijs and G. van der Pluijm, "Osteotropic cancers: From primary tumor to bone," Cancer Lett., vol. 273, no. 2, pp. 177-193, Jan. 2009.

[34] M. A. Nieto, R. Y. Y. J. Huang, R. A. A. Jackson, and J. P. P. Thiery, "Emt: 2016," Cell, vol. 166, no. 1, pp. 21-45, 2016.

[35] I. Yalim-Camci et al., "ETS1 is coexpressed with ZEB2 and mediates ZEB2-induced epithelial-mesenchymal transition in human tumors," Mol. Carcinog., vol. 58, no. 6, pp. 1068-1081, 2019.

[36] W. H. Koppenol, P. L. Bounds, and C. V. Dang, "Otto Warburg's contributions to current concepts of

cancer metabolism," Nat. Rev. Cancer, vol. 11, no. 5, pp. 325-337, 2011.

[37] R. B. Hamanaka and N. S. Chandel, "Targeting glucose metabolism for cancer therapy," J. Exp. Med., vol. 209, no. 2, pp. 211-215, 2012.

[38] A. M. Hosios et al., "Amino Acids Rather than Glucose Account for the Majority of Cell Mass in Proliferating Mammalian Cells," Dev. Cell, vol. 36, no. 5, pp. 540-549, 2016.

[39] C. M. Metallo et al., "Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia," Nature, vol. 481, no. 7381, pp. 380-384, 2012.

[40] P. J. Masiar and E. Medekova, "The role of serine and glutamine in the metabolism of malignant bone tumors and their significance in the diagnosis and prognosis of bone tumors," Neoplasma, vol. 35, no. 2, pp. 197-206, 1988.

[41] S. Pavlides et al., "The reverse Warburg effect: Aerobic glycolysis in cancer associated fibroblasts and the tumor stroma," Cell Cycle, vol. 8, no. 23, pp. 3984-4001, 2009.

[42] L. Hui and Y. Chen, "Tumor microenvironment: Sanctuary of the devil," Cancer Lett., vol. 368, no. 1, pp. 7-13, 2015.

[43] P. Sanità et al., "Tumor-stroma metabolic relationship based on lactate shuttle can sustain prostate cancer progression," BMC Cancer, vol. 14, no. 1, 2014.

[44] G. L. Semenza, "Tumor metabolism: Cancer cells give and take lactate," J. Clin. Invest., vol. 118, no. 12, pp. 3835-3837, 2008.

[45] J. L. Yecies and B. D. Manning, "Chewing the Fat on Tumor Cell Metabolism," Cell, vol. 140, no. 1, pp. 28-30, 2010.

[46] S. Zha et al., "Peroxisomal branched chain fatty acid β-oxidation pathway is upregulated in prostate cancer," Prostate, vol. 63, no. 4, pp. 316-323, 2005.

[47] I. Samudio et al., "Pharmacologic inhibition of fatty acid oxidation sensitizes human leukemia cells to apoptosis induction," J. Clin. Invest., vol. 120, no. 1, pp. 142-156, 2010.

[48] P. Caro et al., "Metabolic Signatures Uncover Distinct Targets in Molecular Subsets of Diffuse Large B Cell Lymphoma," Cancer Cell, vol. 22, no. 4, pp. 547-560, 2012.

[49] P. W. Szlosarek, S. J. Lee, and P. J. Pollard, "Rewiring mitochondrial pyruvate metabolism: Switching off the light in cancer cells?," Mol. Cell, vol. 56, no. 3, pp. 343-344, 2014.

[50] M. J. Holness and M. C. Sugden, "Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation," Biochem. Soc. Trans., vol. 31, no. 6, pp. 1143-1151, 2003.

[51] D. Mishra and D. Banerjee, "Lactate dehydrogenases as metabolic links between tumor and stroma in the tumor microenvironment," Cancers (Basel)., vol. 11, no. 6, 2019.

[52] D. M. Hockenbery, "Targeting mitochondria for cancer therapy," Environ. Mol. Mutagen., vol. 51, no. 5, pp. 476-489, May 2010.

[53] M. P. Murphy and R. A. J. Smith, "Targeting Antioxidants to Mitochondria by Conjugation to Lipophilic Cations," Annu. Rev. Pharmacol. Toxicol., vol. 47, no. 1, pp. 629-656, 2007.

[54] National Brain Tumor Society (NBTS), "The Essential Guide to Brain Tumors," 2012.

[55] M. C. Mabray, R. F. Barajas, and S. Cha, "Modern Brain Tumor Imaging," Brain Tumor Res. Treat., vol. 3, no. 1, p. 8, 2015.

[56] H. M.E., M. A., L. W.L., and S. R., "Brain tumors: Molecular biology and targeted therapies," Ann. Oncol., vol. 17, no. SUPPL. 10, pp. x191–x197, 2006.

[57] E. Lee, R. L. Yong, P. Paddison, and J. Zhu, "Comparison of glioblastoma (GBM) molecular classification methods," Semin. Cancer Biol., vol. 53, pp. 201-211, 2018.

[58] H. Yan et al., "IDH1 and IDH2 mutations in gliomas," N. Engl. J. Med., vol. 360, no. 8, pp. 765-773, Feb. 2009.

[59] B. H. Diplas et al., "The genomic landscape of TERT promoter wildtype-IDH wildtype glioblastoma," Nat. Commun., vol. 9, no. 1, p. 2087, Dec. 2018.

[60] D. W. Parsons et al., "An integrated genomic analysis of human glioblastoma multiforme," Science (80-.)., vol. 321, no. 5897, pp. 1807-1812, 2008.

[61] R. B. Jenkins et al., "A t(1;19) (q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma," Cancer Res., vol. 66, no. 20, pp. 9852-9861, 2006.

[62] T. Rice et al., "Inherited variant on chromosome 11q23 increases susceptibility to IDH-mutated but not IDH-normal gliomas regardless of grade or histology," Neuro. Oncol., vol. 15, no. 5, pp. 535-541, 2013.

[63] M. N. Bainbridge et al., "Germline Mutations in Shelterin Complex Genes Are Associated With Familial Glioma," JNCI J. Natl. Cancer Inst., vol. 107, no. 1, p. c.30, Jan. 2015.

[64] W. Chen and C. Qin, "General hallmarks of microRNAs in brain evolution and development," RNA Biol., vol. 12, no. 7, pp. 701-708, 2015.

[65] R. E. Andersen and D. A. Lim, "Forging our understanding of lncRNAs in the brain," Cell Tissue Res., vol. 371, no. 1, pp. 55-71, 2018.

[66] M. Hanan, H. Soreq, and S. Kadener, "CircRNAs in the brain," RNA Biol., vol. 14, no. 8, pp. 1028-1034, 2017.

[67] C. H. Yang et al., "MicroRNA-21 promotes glioblastoma tumorigenesis by down-regulating insulin-like growth factor-binding protein-3 (IGFBP3)," J. Biol. Chem., vol. 289, no. 36, pp. 25079-25087, 2014.

[68] G. Gabriely et al., "MicroRNA 21 Promotes Glioma Invasion by Targeting Matrix Metalloproteinase Regulators," Mol. Cell. Biol., vol. 28, no. 17, pp. 5369-5380, 2008.

[69] Silber J. et al., "miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells," Bmc Med., vol. 6, no. 14, 2008.

[70] Y. Yan, Z. Xu, Z. Li, L. Sun, and Z. Gong, "An insight into the increasing role of LncRNAs in the pathogenesis of gliomas," Front. Mol. Neurosci., vol. 10, 2017.

[71] X. Q. Zhang and G. K. K. Leung, "Long non-coding RNAs in glioma: Functional roles and clinical perspectives," Neurochem. Int., vol. 77, pp. 78-85, 2014.

[72] Z. J. et al., "Differential Expression of Circular RNAs in Glioblastoma Multiforme and Its Correlation with Prognosis," Transl. Oncol., vol. 10, no. 2, pp. 271-279, 2017.

[73] L. Bolha and D. Glavač, "Circular RNA FBXW7: Implication in glioma tumorigenesis," Transl. Cancer Res., vol. 7, pp. S521–S524, 2018.

[74] D. Barbagallo et al., "CircSMARCA5 inhibits migration of glioblastoma multiforme cells by regulating a molecular axis involving splicing factors SRSF1/SRSF3/PTB," Int. J. Mol. Sci., vol. 19, no. 2, 2018.

[75] R. Shi et al., "Exosomal levels of miRNA-21 from cerebrospinal fluids associated with poor prognosis and tumor recurrence of glioma patients," Oncotarget, vol. 6, no. 29, pp. 26971-26981, 2015.

[76] Q. Cai, A. Zhu, and L. Gong, "Exosomes of glioma cells deliver miR-148a to promote proliferation and metastasis of glioblastoma via targeting CADM1," Bull. Cancer, vol. 105, no. 7-8, pp. 643-651, 2018.

[77] H. L. Lang et al., "Glioma cells promote angiogenesis through the release of exosomes containing long non-coding RNA POU3F3," Eur. Rev. Med. Pharmacol. Sci., vol. 21, no. 5, pp. 959-972, 2017.

[78] H. L. Lang et al., "Glioma cells enhance angiogenesis and inhibit endothelial cell apoptosis through the release of exosomes that contain long non-coding RNA CCAT2," Oncol. Rep., vol. 38, no. 2, pp. 785-798, 2017.

[79] M. Goedert, "Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Aβ, tau, and α-synuclein," Science (80-.)., vol. 349, no. 6248, 2015.

[80] S. B. Prusiner, "Novel proteinaceous infectious particles cause scrapie," Science (80-.)., vol. 216, no. 4542, pp. 136-144, 1982.

[81] M. Jucker and L. C. Walker, "Self-propagation of pathogenic protein aggregates in neurodegenerative diseases," Nature, vol. 501, no. 7465, pp. 45-51, 2013.

[82] M. Goedert, B. Falcon, F. Clavaguera, and M. Tolnay, "Prion-like mechanisms in the pathogenesis of tauopathies and synucleinopathies," Curr. Neurol. Neurosci. Rep., vol. 14, no. 11, pp. 1-11, 2014.

[83] A. Goate et al., "Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease," Nature, vol. 349, no. 6311, pp. 704-706, 1991.

[84] P. Poorkaj et al., "Tau is a candidate gene for chromosome 17 frontotemporal dementia," Ann. Neurol., vol. 43, no. 6, pp. 815-825, 1998.

[85] M. Hutton et al., "Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17," Nature, vol. 393, no. 6686, pp. 702-704, 1998.

[86] M. G. Spillantini, J. R. Murrell, M. Goedert, M. R. Farlow, A. Klug, and B. Ghetti, "Mutation in the tau gene in familial multiple system tauopathy with presenile dementia," Proc. Natl. Acad. Sci. U. S. A., vol. 95, no. 13, pp. 7737-7741, 1998.

[87] P. M.H. et al., "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.," Science (80-.)., vol. 276, no. 5321, pp. 2045-2047, 1997.

[88] M. Goedert, M. G. Spillantini, R. Jakes, D. Rutherford, and R. A. Crowther, "Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease," Neuron, vol. 3, no. 4, pp. 519-526, 1989.

[89] C. F. et al., "Brain homogenates from human tauopathies induce tau inclusions in mouse brain," Proc. Natl. Acad. Sci. U. S. A., vol. 110, no. 23, pp. 9535-9540, 2013.

[90] D. W. Sanders et al., "Distinct tau prion strains propagate in cells and mice and define different tauopathies," Neuron, vol. 82, no. 6, pp. 1271-1288, 2014.

[91] B. Falcon et al., "Conformation determines the seeding potencies of native and recombinant Tau aggregates," J. Biol. Chem., vol. 290, no. 2, pp. 1049-1065, 2015.

[92] S. Boluda, M. Iba, B. Zhang, K. M. Raible, V. M. Y. Lee, and J. Q. Trojanowski, "Differential induction and spread of tau pathology in young PS19 tau transgenic mice following intracerebral injections of pathological tau from Alzheimer's disease or corticobasal degeneration brains," Acta Neuropathol., vol. 129, no. 2, pp. 221-237, 2015.

[93] R. A. Crowther and M. Goedert, "Abnormal tau-containing filaments in neurodegenerative diseases," J. Struct. Biol., vol. 130, no. 2-3, pp. 271-279, 2000.

[94] J. Kang et al., "The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor," Nature, vol. 325, no. 6106, pp. 733-736, 1987.

[95] C. L. Masters and D. J. Selkoe, "Biochemistry of amyloid β-protein and amyloid deposits in Alzheimer disease," Cold Spring Harb. Perspect. Med., vol. 2, no. 6, 2012.

[96] P. Lu et al., "Three-dimensional structure of human γ-secretase," Nature, vol. 512, no. 7513, pp. 166-170, 2014.

[97] T. Xie et al., "Crystal structure of the γ-secretase component nicastrin," Proc. Natl. Acad. Sci. U. S. A., vol. 111, no. 37, pp. 13349-13354, 2014.

[98] H. Acx et al., "Signature amyloid β profiles are produced by different γ-secretase complexes," J. Biol. Chem., vol. 289, no. 7, pp. 4346-4355, 2014.

[99] P. J. Barnes, "Similarities and differences in inflammatory mechanisms of asthma and COPD," Breathe, vol. 7, no. 3, pp. 229-238, 2011.

[100] S. Y. A. Rafael Lozano, Mohsen Naghavi, Kyle Foreman, Stephen Lim, Kenji Shibuya, Victor Aboyans*, Jerry Abraham*, Timothy Adair*, Rakesh Aggarwal* et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," Lancet, 2012.

[101] W. I. De Boer, V. K. T. Alagappan, and H. S. Sharma, "Molecular mechanisms in chronic obstructive pulmonary disease: Potential targets for therapy," Cell Biochem. Biophys., vol. 47, no. 1, pp. 131-147, 2007.

[102] B. P.J., "Alveolar macrophages as orchestrators of COPD," COPD, vol. 1, no. 1, pp. 59-70, 2004.

[103] P. J. Barnes, S. D. Shapiro, and R. A. Pauwels, "Chronic obstructive pulmonary disease: Molecular and cellular mechanisms," Eur. Respir. J., vol. 22, no. 4, pp. 672-688, 2003.

[104] K. Tomita et al., "Attenuated production of intracellular IL-10 and IL-12 in monocytes from patients with severe asthma," Clin. Immunol., vol. 102, no. 3, pp. 258-266, 2002.

[105] V. M. Keatings, P. D. Collins, D. M. Scott, and P. J. Barnes, "Differences in Interleukin-8 and Tumor Necrosis Facfor-α in Induced Sputum from Patients with Chronic Obstructive Pulmonary Disease or Asthma," Am. J. Respir. Crit. Care Med., vol. 153, no. 2, pp. 530-534, 1996.

[106] S. L. Traves, S. V. Culpitt, R. E. K. Russell, P. J. Barnes, and L. E. Donnelly, "Increased levels of the chemokines GROα and MCP-1 in sputum samples from patients with COPD," Thorax, vol. 57, no. 7, pp. 590-595, 2002.

[107] F. Bazzoni, M. A. Cassatella, F. Rossi, M. Ceska, B. Dewald, and M. Baggiolini, "Phagocytosing neutrophils produce and release high amounts of the neutrophil-activating peptide 1/ interleukin 8," J. Exp. Med., vol. 173, no. 3, pp. 771-774, 1991.

[108] G. R. Hellermann, S. B. Nagy, X. Kong, R. F. Lockey, and S. S. Mohapatra, "Mechanism of cigarette smoke condensate-induced acute inflammatory response in human bronchial epithelial cells," Respir. Res., vol. 3, 2002.

[109] K. B. Adler and Y. Li, "Airway epithelium and mucus: Intracellular signaling pathways for gene expression and secretion," Am. J. Respir. Cell Mol. Biol., vol. 25, no. 4, pp. 397-400, 2001.

[110] C. Pilette, Y. Ouadrhiri, V. Godding, J. P. Vaerman, and Y. Sibille, "Lung mucosal immunity: Immunoglobulin-A revisited," Eur. Respir. J., vol. 18, no. 3, pp. 571-588, 2001.

[111] S. Baraldo, K. L. Oliani, G. Turato, R. Zuin, and M. Saetta, "The Role of Lymphocytes in the Pathogenesis of Asthma and COPD," Curr. Med. Chem., vol. 14, no. 21, pp. 2250-2256, 2007.

[112] B. Mehrad and T. J. Standiford, "Role of cytokines in pulmonary antimicrobial host defense," Immunol. Res., vol. 20, no. 1, pp. 15-27, 1999.

[113] C. A. Janeway, P. Travers, M. Walport, and E. Al, "Principles of innate and adaptive immunity," Immunobiol. Immune Syst. Heal. Dis. 5th Ed., pp. 1-9, 2001.

[114] P. Bonniaud et al., "Smad3 Null Mice Develop Airspace Enlargement and Are Resistant to TGF-β-Mediated Pulmonary Fibrosis," J. Immunol., vol. 173, no. 3, pp. 2099-2108, 2004.

[115] W. I. De Boer et al., "Transforming growth factor β1 and recruitment of macrophages and mast cells in airways in chronic obstructive pulmonary disease," Am. J. Respir. Crit. Care Med., vol. 158, no. 6, pp. 1951-1957, 1998.

[116] M. Mohammadi, S. K. Olsen, and O. A. Ibrahimi, "Structural basis for fibroblast growth factor receptor activation," Cytokine Growth Factor Rev., vol. 16, no. 2 SPEC. ISS., pp. 107-137, 2005.

[117] M. Hoshino, Y. Nakamura, and Q. A. Hamid, "Gene expression of vascular endothelial growth factor and its receptors and angiogenesis in bronchial asthma," J. Allergy Clin. Immunol., vol. 107, no. 6, pp. 1034-1038, 2001.

[118] A. R. Kranenburg, W. I. De Boer, V. K. T. Alagappan, P. J. Sterk, and H. S. Sharma, "Enhanced bronchial expression of vascular endothelial growth factor and receptors (Flk-1 and Flt-1) in patients with chronic obstructive pulmonary disease," Thorax, vol. 60, no. 2, pp. 106-113, 2005.

[119] W. I. De Boer, "Cytokines and therapy in COPD: A promising combination?," Chest, vol. 121, no. 5 SUPPL., pp. 209S-218S, 2002.

[120] P. J. Barnes, "The cytokine network in asthma and chronic obstructive pulmonary disease," J. Clin. Invest., vol. 118, no. 11, pp. 3546-3556, 2008.

[121] P. J. Barnes, "Cellular and molecular mechanisms of asthma and COPD," Clin. Sci., vol. 131, no. 13, pp. 1541-1558, 2017.

[122] B. Lauby-Secretan, C. Scoccianti, D. Loomis, Y. Grosse, F. Bianchini, and K. Straif, "Body fatness and cancer

- Viewpoint of the IARC working group," N. Engl. J. Med., vol. 375, no. 8, pp. 794-798, Aug. 2016.

[123] F. Johnson, L. Cooke, H. Croker, and J. Wardle, "Changing perceptions of weight in Great Britain: Comparison of two population surveys," Bmj, vol. 337, no. 7664, pp. 270-272, 2008.

[124] J. Sulc, T. W. Winkler, I. M. Heid, and Z. Kutalik, "Heterogeneity in Obesity: Genetic Basis and Metabolic Consequences," Curr. Diab. Rep., vol. 20, no. 1, 2020.

[125] D. Albuquerque, E. Stice, R. Rodríguez-López, L. Manco, and C. Nóbrega, "Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective," Mol. Genet. Genomics, vol. 290, no. 4, pp. 1191-1221, 2015.

[126] J. H. Yu and M. S. Kim, "Molecular mechanisms of appetite regulation," Diabetes Metab. J., vol. 36, no. 6, pp. 391-398, 2012.

[127] N. Sáinz, J. Barrenetxe, M. J. Moreno-Aliaga, and J. A. Martínez, "Leptin resistance and diet-induced obesity: Central and peripheral actions of leptin," Metabolism., vol. 64, no. 1, pp. 35-46, 2015.

[128] C. M. Steppan, J. Wang, E. L. Whiteman, M. J. Birnbaum, and M. A. Lazar, "Activation of SOCS-3 by Resistin," Mol. Cell. Biol., vol. 25, no. 4, pp. 1569-1575, 2005.

[129] S. Huether and K. McCance, Pathophysiology: The Biologic Basis for Disease in Adults and Children, vol. 13, no. 6. 2020.

[130] T. Sato, T. Ida, Y. Nakamura, Y. Shiimura, K. Kangawa, and M. Kojima, "Physiological roles of ghrelin on obesity," Obes. Res. Clin. Pract., vol. 8, no. 5, pp. e405–e413, 2014.

[131] A. Whang, R. Nagpal, and H. Yadav, "Bi-directional drug-microbiome interactions of anti-diabetics," EBioMedicine, vol. 39, pp. 591-602, 2019.

[132] "Pathophysiology of Obesity-Induced Health Complications," Pathophysiol. Obesity-Induced Heal. Complicat., 2020.

[133] N. Patel, C. Huang, and A. Klip, "Cellular location of insulin-triggered signals and implications for glucose uptake," Pflugers Arch. Eur. J. Physiol., vol. 451, no. 4, pp. 499-510, 2006.

[134] R. T. Atawia, K. L. Bunch, H. A. Toque, R. B. Caldwell, and R. W. Caldwell, "Mechanisms of obesity-induced metabolic and vascular dysfunctions," Front. Biosci. - Landmark, vol. 24, no. 5, pp. 890-934, 2019.

[135] G. Drews, P. Krippeit-Drews, and M. Duïfer, "Oxidative stress and beta-cell dysfunction," Pflugers Arch. Eur. J. Physiol., vol. 460, no. 4, pp. 703-718, Sep. 2010.

[136] S. P. Hunger and C. G. Mullighan, "Acute Lymphoblastic Leukemia in Children," N. Engl. J. Med., vol. 373, no. 16, pp. 1541-1552, Oct. 2015.

[137] S. Paul, H. Kantarjian, and E. J. Jabbour, "Adult Acute Lymphoblastic Leukemia," Mayo Clin. Proc., vol. 91, no. 11, pp. 1645-1666, 2016.

[138] D. G. Gilliland and M. S. Tallman, "Focus on acute leukemias," Cancer Cell, vol. 1, no. 5, pp. 417-420, 2002.

[139] E. J. Yeoh et al., "Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling," Cancer Cell, vol. 1, no. 2, pp. 133-143, 2002.

[140] S. A. Armstrong et al., "FLT3 mutations in childhood acute

lymphoblastic leukemia," Blood, vol. 103, no. 9, pp. 3544-3546, 2004.

[141] B. J. Druker et al., "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia," N. Engl. J. Med., vol. 344, no. 14, pp. 1031-1037, Apr. 2001.

[142] B. J. Druker, "Imatinib as a paradigm of targeted therapies," Adv. Cancer Res., vol. 91, pp. 1-30, 2004.

[143] N. Chiorazzi, K. R. Rai, and M. Ferrarini, "Chronic Lymphocytic Leukemia," N. Engl. J. Med., vol. 352, no. 8, pp. 804-815, Feb. 2005.

[144] D. Bhojwani and C.-H. Pui, "Relapsed childhood acute lymphoblastic leukaemia," Lancet Oncol., vol. 14, no. 6, pp. e205–e217, May 2013.

[145] J. H. Park et al., "Long-term follow-up of CD19 CAR therapy in acute lymphoblastic leukemia," N. Engl. J. Med., vol. 378, no. 5, pp. 449-459, 2018.

[146] C. H. June and M. Sadelain, "Chimeric antigen receptor therapy," N. Engl. J. Med., vol. 379, no. 1, pp. 64-73, 2018.

[147] A. Martyniszyn, A. C. Krahl, M. C. André, A. A. Hombach, and H. Abken, "CD20-CD19 Bispecific CAR T Cells for the Treatment of B-Cell Malignancies," Hum. Gene Ther., vol. 28, no. 12, pp. 1147-1157, 2017.

[148] B. Plenum et al., "Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*," Nature, vol. 397, no. February, pp. 436-441, 1999.

[149] W. Li et al., "Bats are natural reservoirs of SARS-like coronaviruses," Science (80-.)., vol. 310, no. 5748, pp. 676-679, 2005.

[150] E. I. Azhar et al., "Evidence for camel-to-human transmission of MERS coronavirus," N. Engl. J. Med., vol. 370, no. 26, pp. 2499-2505, 2014.

[151] P. Pradhan et al., "Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag," bioRxiv, p. 2020.01.30.927871, 2020.

[152] X. C., L. X., L. S., S. Y., G. S.-J., and G. F., "HIV-1 did not contribute to the 2019-nCoV genome," Emerg. Microbes Infect., vol. 9, no. 1, pp. 378-381, 2020.

[153] WHO., "WHO. Coronavirus Disease 2019 (COVID-19): Situation Report – 105.," WHO. Coronavirus Dis. 2019 Situat. Rep. – 105., p. 18, 2020.

[154] K. Pyrc et al., "Mosaic Structure of Human Coronavirus NL63, One Thousand Years of Evolution," J. Mol. Biol., vol. 364, no. 5, pp. 964-973, 2006.

[155] D. W. E., V. D. N., F. D., and M. V.J., "SARS and MERS: Recent insights into emerging coronaviruses," Nat. Rev. Microbiol., vol. 14, no. 8, pp. 523-534, 2016.

[156] "Coronaviridae - Figures - Positive Sense RNA Viruses - Positive Sense RNA Viruses (2011)," Int. Comm. Taxon. Viruses.

[157] J. S. Mani et al., "Natural product-derived phytochemicals as potential agents against coronaviruses: A review," Virus Res., vol. 284, 2020.

[158] A. Wu et al., "Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China," Cell Host Microbe, vol. 27, no. 3, pp. 325-328, 2020.

[159] P. Zhou et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," Nature, vol. 579, no. 7798, pp. 270-273, 2020.

[160] M. Hoffmann et al., "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor," Cell, vol. 181, no. 2, pp. 271-280.e8, 2020.

[161] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of SARS-CoV-2," Nat. Med., vol. 26, no. 4, pp. 450-452, 2020.

[162] T. Adhanom Ghebreyesus, "WHO Director-General's opening remarks at the media briefing on COVID-19," World Heal. Organ., no. March, p. 4, 2020.

[163] Z. Wu and J. M. McGoogan, "Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention," JAMA - J. Am. Med. Assoc., vol. 323, no. 13, pp. 1239-1242, 2020.

[164] S. A. Southern and C. S. Herrington, "Molecular events in uterine cervical cancer," Sex. Transm. Infect., vol. 74, no. 2, pp. 101-109, 1998.

[165] E.-K. Yim and J.-S. Park, "The Role of HPV E6 and E7 Oncoproteins in HPV-associated Cervical Carcinogenesis," Cancer Res. Treat., vol. 37, no. 6, p. 319, 2005.

[166] M. Julia Gargano, PhD; Elissa Meites, MD, MPH; Meg Watson, MPH; Elizabeth Unger, MD, PhD; Lauri Markowitz, Human Papillomavirus. .

[167] NCI, "National Cancer Institute. Human Papillomavirus (HPV) Vaccine Fact Sheet," 2015. [Online]. Available: http://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-vaccine-fact-sheet.

Chapter 2

Molecular Medicines for Parasitic Diseases

Bhawana Singh

Abstract

Being the cause for significant amount of morbidities and mortalities, parasitic diseases remain the major challenge for the healthcare community due to the limitations associated with the current chemotherapeutics. Drug discovery/invention can be achieved by collaborative efforts of biotechnologists and pharmacists for identifying potential candidates and successfully turn them into medicine for improving the healthcare system. Although molecular medicine for disease intervention is still in its infancy, however, significant research works and successful trials in short span of time have made it broadly accepted among the scientific community. This chapter identifies different molecular medicine approaches for dealing with parasites that have been coming up on the horizon with the new technological advances in bioinformatics and in the field of *omics*. With the better understanding of the genomics, molecular medicine field has not only raised hopes to deal with parasitic infections but also accelerated the development of personalized medicine. This will provide a targeted approach for identifying the druggable targets and their pathophysiological importance for disease intervention.

Keywords: CRISPR/Cas9 system, monoclonal antibody, immune checkpoint inhibitors, nanomedicine

1. Introduction

Parasitic diseases remain the threat to global healthcare sector, with considerable mortalities and morbidities associated with these diseases. Combating parasitic infection relies mostly on conventional chemotherapeutic approaches; however, an exponential rise in the number of recrudescent cases, lack of vaccines and toxicities issues associated with chemotherapies emphasized the need for the research to develop alternative strategies. It is noteworthy that emergence of drug resistance is not new; microbes have been evolving since ages, by knocking out one or the genes. In context of emerging drug resistance, World Health Organization (WHO) has warned for the upcoming "post-antibiotic era", therefore, molecular medicine has surged. Molecular medicine is the application of gene and/or DNA based information for therapeutic purpose. It involves the study of molecular mechanisms, identification of erroneous genetic and/or molecular pathways and development of molecular intervention with the aim to improve disease management.

This chapter provides an insight into how the anomalies in molecular pathways can be targeted leading to discovery of potential candidates for development of clinical medicine and innovative therapies to improve disease management strategies.

2. Evolution of molecular medicine

The field of molecular medicine evolved over a period of time since the discovery of DNA (in 1953) and recombinant DNA technology. Another major breakthrough in 1975, when it was discovered that DNA can be read base by base through the sequencing technique. Later, in 1985, it was known that DNA can be amplified with PCR, and this was major achievement in the field of molecular diagnostics. This journey gained pace with the advent of automated DNA sequencing in 1987 that served as the background for the human genome project (HGP) in 1990. The journey of HGP started the era of modern molecular medicine with the first successful gene therapy. In 1995, the success of unveiling the DNA sequence from the first model organism (*Hemophilus influenzae*) triggered the endeavor towards the completion of HGP. The completion of HGP (in 2000) and the free access to the human genome sequences provided the ground for the advent of omics era or post-genomic era (or functional genomics) [1].

This led to the use of increasing number of analytical platforms for DNA sequencing termed as the next generation sequencing platforms. Further, metagenomics approaches—omics and/or shot-gun approaches paved the way for the third-generation sequencing, that aimed to reduce the sequencing costs. These advances gained momentum with the computational approaches where synthetic biology has remarkably facilitated the DNA based analysis as well as the development of models for drug testing.

In order to deal with the limitation of the existing chemotherapeutic approaches there remains an urgent need for the conversion of biomedical knowledge into clinical application. Molecular medicine provides the opportunity to fill the gap between the basic research and the clinical application for the diagnosis, prevention and treatment of diseases. It involves the combinatorial application of pharmacology, biomedical and omics technologies for understanding and improving the molecular basis of the disease pathogenesis that will serve in designing disease intervention strategies. Development of molecular drug is a complex process that involves multidisciplinary effort including high throughput screening, chemical synthesis, modification, omics technologies, data mining, structure-based drug designing, phenotypic screening, target and lead identification and validation, etc. The development of molecular medicines involves following steps—first, the identification of target, potential tractability of target (i.e. identifying targets that are more druggable than others, depending upon their chemistry), establish genetic association of target with disease pathophysiology (some targets required for drug action may not necessarily associate with disease genetics) and validation of target (by establishing association of target with the disease development/persistence). Validation of target usually involves different molecular approaches to understand the role of target gene or protein in diseases pathophysiology. Overall, it is an interdisciplinary branch where recent technical advances have served as the milestone in gaining insight into the phenomenon of disease pathogenesis and development of innovative therapeutic measures.

Parasitic diseases are amongst the common infections in humans caused by protozoan and helminthic parasites. The causative agents, parasites, are diverse ranging from single celled protozoan to worms that be seen with naked eyes. Till the end of nineteenth century, parasitologists were mainly focused on understanding their life cycle; however, the concept took turn when some parasites were found to be associated with several human diseases that led to significant morbidity and mortality. Parasitic diseases are cosmopolitan, that may affect any part of world however, mostly the diseases are common in tropical countries, but tourism and migration can transmit them outside their geographical boundaries. The signs and symptoms of disease may not be obvious, and it may vary from mild abdominal pain to chronic

Molecular Medicines for Parasitic Diseases
DOI: http://dx.doi.org/10.5772/intechopen.91956

Diseases	Causative agent (pathogen)	Transmitting agent (vector)	Manifestation	Treatment options
Protozoan parasites				
Leishmaniasis (visceral, cutaneous and mucocutaneous)	*Leishmania* species	Sandfly (*Phlebotomus* & *Lutzomyia* species)	Fever, anemia, splenomegaly, lymphadenopathy; cutaneous forms manifests as skin lesions and ulcers	Liposomal amphotericin B, miltefosine, antimonials; fluconazole, itraconazole
Malaria	*Plasmodium* species	Female mosquito (*Anopheles* species)	Headache, fever, paroxysm, joint pain, anemia, jaundice; neurological symptoms in severe cases	Chloroquine, mefloquine, doxycycline
Chagas disease (American trypanosomiasis)	*Trypanosoma cruzi*	Kissing bugs (Triatominae)	Fever, malaise, enlargement (liver, spleen, lymph nodes), sometimes skin nodules (chagoma); chronic stages affects the brain, heart and digestive system	Benznidazole, nifurtimox
Human African Trypanosomiasis	*Trypanosoma brucei*	Tsetse fly (*Glossina* species)	First stage-intermittent fever, headache, swelling of lymph nodes, joint pain; second stage involves neurological symptoms	Pentamidine, suramin, fexinidazole, nifurtimox, eflornithine
Toxoplasmosis	*Toxoplasma gondii*	Oral route; transmitted by ingestion of parasite oocyst	Headache, fever, fatigue, muscle ache; skin manifestation includes erythema and roseola	Pyrimethamine, sulfadiazine, clindamycin, spiramycin
Trichomoniasis	*Trichomonas vaginalis*	Genital contacts	Pain, itchiness/burning in genitourinary organs, urethritis, prostatitis (in males) while frothy, foul-smelling discharge, vaginitis (in females)	Metronidazole
Giardiasis (beaver fever)	*Giardia* species	Feco-oral transmission by ingestion of cysts	Chronic diarrhea, abdominal cramps, nausea and vomiting	Nitroimdazole, quinacrine, furazolidone, paromomycin
Cryptosporidiosis	*Cryptosporidium* species	Oral transmission by consumption of contaminated water, undercooked food	Diarrhea, abdominal cramps, low-grade fever (in intestinal cryptosporidiosis); inflammation of nasal mucosa, cough, shortness of breath, hypoxemia(respiratory cryptosporidiosis)	Electrolyte replacement by rehydration therapy, nitazoxanide, azithromycin, paromomycin
Amoebiasis	*Entamoeba histolytica*	Feco-oral route	Diarrhea, severe abdominal pain	Amebicidals (metronidazole, tinidazole) and cysticidal agents (iodoquinol)

Diseases	Causative agent (pathogen)	Transmitting agent (vector)	Manifestation	Treatment options
Helminthic diseases				
Roundworm infection (in murine)	*Nippostrongylus braziliensis*	Skin penetration	Emphysema, loss of alveolar septa, lung hemorrhage	Tetramisole
Ascariasis (Roundworm infection, in human)	*Ascaris lumbricoides*	Feco-oral route	Fever, cough, weight loss, abdominal discomfort, intestinal ulcer accompanied with eosinophilia	Albendazole, mebendazole
Fasciolosis	*Fasciola hepatica*	Oral route, consumption of contaminated food	Acute phase marked by fever, nausea, skin rashes, abdominal pain; chronic phase manifests as jaundice, anemia and intermittent pain	Bromofenofos, triclabendazole, bithionol
Taeniasis	*Taenia* species	Consumption of undercooked pork or beef	Mild (abdominal pain and nausea) to no symptoms	Praziquantel, albendazole, niclosamide, mepacrine
Onchocerciasis (sub-cutaneous filariasis)	*Onchocerca* species	Blackfly (*Simulium* species)	Itchiness and bumps and depigmentation in skin to blindness	Ivermectin, moxidectin
Filariasis (lymphatic and serous cavity)	*Wuchereria bancrofti* and *Brugia* species	Blackflies and mosquitoes	Edema with skin thickening and underlying tissues	Diethylcarbamazine citrate (DEC)
Neural angiostrongyliasis (eosinophilic meningitis)	*Angiostrongylus cantonensis*	Oral route; upon ingestion larvae in undercooked prawn, snails, slugs, frogs	Headache, fever, malaise, nausea, neck stiffness, varying degree of neurological dysfunctions	No specific treatment, supportive care helps reduce the severity of symptoms
Schistosomiasis	*Schistosoma mansonii*	Contact with fresh water contaminated with parasites (released from fresh water snails)	Abdominal pain, diarrhea, fever, cough, bloody stool and/or blood in the urine	Praziquantel, oxamniquine, metrifonate, artesunate, mefloquine
Trichinosis	*Trichinella spiralis*	Consumption of undercooked pork	Nausea, vomiting, fever, diarrhea, facial swelling	Mebendazole, albendazole

Table 1.
Some parasitic diseases, their symptoms and treatment options.

hepatomegaly and eventually death. Some parasitic infections are easily treated while others are not. In the light of the lack of vaccine for parasitic infection, proper prophylactic measures (proper hygiene, prevention of contaminated food, water, preventing consumption of undercooked food, use of bednets, insecticide spraying to prevent vector borne diseases, etc.) and active disease surveillance remains the key for disease elimination. Unfortunately, poor disease management strategies have made parasitic infections a global healthcare challenge. In this article it's only possible to cover some important parasites (**Table 1**), for which research on molecular medicines are underway.

3. Molecular medicinal strategies and parasitic diseases

Parasitic infections (protozoan and helminthic infections) affect more than a quarter world population and cause chronic illness primarily in developing countries of world. These diseases affect the quality of life and treatment costs possess economic burden on families leading to viscous circle of poverty.

Molecular medicine is a broad field that includes insight into the molecular aspect of diseases. Recombinant DNA and cloning technologies are the conventional tools for studying the disease associated molecular profiles. Recent technical advances have paved the way for utilization of several molecular strategies for treating infectious diseases. Molecular medicine aims to understand the molecular basis of disease pathogenesis and allows the utilization of the information in designing specific diagnostic, therapeutic and prophylactic options. Mainly molecular medicine relies on two strategies—targeting genome and targeting signaling pathways, as targeted approach of disease management. Thus, it aims to improve the human health through the understanding of mechanism in human diseases.

3.1 Targeting genome

Apart from conventional approach of gene therapy (replacement of defective gene by exogenous DNA and editing mutated gene), recent technical advances have opened the arena for other strategies of manipulating the gene expression. Gene editing methods have gained limelight that involves the intrinsic molecular repair processes within the cell. The process of break repair in the DNA involves the homology-directed repair (HDR) and/or non-homologous end joining (NHEJ). The key step in gene-editing tool involves the precise introduction of double strand breaks. This process involves the use of engineered meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the recent CRISPR/Cas system [2]. Further, short antisense oligonucleotides potentially serve as tools for abrogating the transcription of target gene. As compared to other genome editing methods CRISPR/Cas system using guide RNA has shown immense potential for future of molecular medicine.

3.1.1 Engineered meganucleases

Although there remains plethora of meganucleases to choose from, however, most commonly used meganucleases include the ZFNs that have DNA binding zinc finger protein domain and nuclease domain. Cys_2-His_2 zinc finger domain is amongst the most abundant types of DNA binding motifs in eukaryotes [3, 4]. The ZFNs work by binding to the DNA and cleaving it, which then undergoes repair by either homologous recombination or error-prone NHEJ [5]. Site-specific cleavage is induced by manipulating the ZFN complex to recognize two sequences that are

on either side of target site. Upon identification, cleavage of genome is induced by restriction enzyme (FokI), thus generating double stranded breaks in genome that can be used for editing the region.

TALEN has been introduced as an alternative to ZFNs. These are similar to ZFNs in using restriction enzyme that incorporates with the DNA binding domain, but are of different origin. Similar to ZFNs, these protein structures lead to the double stranded breaks in DNA for genome editing [6]. Unfortunately, there are no evidences for the use of the aforementioned meganucleases for the development of molecular medicine for parasitic diseases.

3.1.2 RNA-guided engineered nucleases (CRISPR-Cas9 system)

After the discovery of clustered regularly interspaced palindromic repeats (CRISPR) (in 1987) as a part of bacterial immune system against invading viruses. This strategy has potential application in editing human chromosome with great accuracy. It is RNA guided gene editing tool that uses Cas9 endonuclease for generating double-stranded breaks at loci of interest, which are then repaired via. HDR (using a template) or NHEJ or error-prone microhomology-mediated end joining (MMEJ) [7]. Thus, leading to mutation (insertion, deletion or substitution) with no or minimal damage to host genome. Since the sequencing of the parasite genome, CRISPR/Cas9 genome editing tool accelerated the molecular research in parasitology.

Genome editing in malarial parasite (*P. falciparum* and *P. yoelii*) has provided the ground for development of molecular medicine [8, 9]. CRISPR system of genome editing has played crucial role in understanding the drug resistance and pathogen survival thus, genome editing of the parasite genome holds promises to trigger host immune responses while preventing disease pathology. There are several studies for CRISPR/Cas9 analyses in *T. gondii* however, the most remarkable one include the genome-wide screening that identified the genes involved in infection process [10, 11]. This has greatly accelerated the research to understand the parasite metabolic needs for survival and virulence, conversely, shedding light on dealing with drug resistance mechanisms. Similarly, CRISPR/Cas9 system for *T. cruzi* and *T. brucei* has facilitated functional studies of drug targets and/or vaccine candidates [12, 13]. Genetic manipulation has always remained arduous for *Leishmania* however, CRISPR/Cas9 system has proven efficacy for rapid genome editing and understanding gene functions with therapeutic implication [14, 15]. Similarly, genome editing for *T. vaginalis* has been recently introduced with potential *in vivo* toxicity issues which has been dealt by using nucleofusion based transfection [16, 17]. Large-scale functional genomic screening cannot rely on conventional CRISPR/Cas9 approach thus, CRISPRi and CRISPRa approaches have gained significant attention for generating knock-in/knock-down libraries.

Genome editing using CRISPR/Cas9 system in *Schistosoma mansoni* eggs reduced the infection induced granulomas. Similarly, CRISPR/Cas9 mediated deletion of granulin gene from liver fluke (*O. viverrini*) significantly improved the disease symptoms [18]. This is just the start of the use of CRISPR/Cas9 system in parasitic diseases, it has broader potential in designing molecular medicine for disease intervention.

3.1.3 RNA interference (RNAi)

RNA interference is the approach of inhibiting the gene expression or translation by insertion of double-stranded RNA into the cells and/or organism that mediates targeted degradation of its homologous RNA. It can be done *via*. Short interfering RNA (siRNAs), micro RNA (miRNA) and piwi-interacting (piRNA). It mediates

gene silencing by forming RNA-induced silencing complex (RISC) that eventually degrades target mRNA. This unique ability of silencing target gene holds hopes for controlling parasitic infections.

T. brucei was the first protozoan parasite where RNAi based targeting of β-tubulin gene, changed parasite morphology [19]. Conversely, *T. cruzi* [20], *L. donovani* [21] and *L. major* [22] lack the essential protein (Ago-1) for suppressing the gene expression. RNAi mediated targeting of cathepsin B reduced the disease progression [23]. RNAi mediated targeting of topoisomerases and farnesyl pyrophosphate synthase have proven efficacy of RNAi based molecular medicine for disease intervention [24, 25]. Later, in 2011, RNAi target sequencing (RIT) identified several potential targets for genome-scale functional analyses which could potentially therapeutic targets [26]. RNA aptamers (synthetic RNA and DNA molecules) binds the target ligands (RNA/DNA) with high-specificity, and have been developed as pharmaceutically active compounds against *T. brucei* [27].

Malarial parasites lack the conventional RNAi pathway however, antisense oligodeoxynucleotides treatment (against parasite topoisomerase) has shown to significantly reduce the parasite multiplication [28]. Further, topoisomerase targeting antisense nanoparticles and chitosan-based nanoformulation have also been used to inhibit *P. falciparum* growth [29, 30]. Recently, synthetic siRNA targeting the β-actinin and cysteine protease served as potential molecular target for *T. vaginalis* infection [31]. Although still scrappy, better understanding of RNAi pathway in protozoan parasites is likely to revolutionize the molecular medicine due to its genome homeostatic potential.

RNAi based silencing of key genes involved in regulating the parasite survival and development have been potential candidates for therapeutics. Several miRNA have been known to regulate the nematode development and survival in the host microenvironment due to their immunoregulatory potential. These have also been crucial players of host-parasite interaction and have been used as diagnostic marker of infection. *Nippostrongylus brasiliensis* (rat parasite) was the first nematode where RNAi was reported [32]. Much evidences for RNAi based knock-down studies have been seen in *Schistosoma* [33, 34], *Brugia* [35], *Trichinella spiralis* [36], *Ascaris suum* [37], *Angiostrongylus cantonensis* [38], *Taenia saginata* [39], *Echinococcus* [40], *H. contortus* [41] and *Onchocerca volvulus* [42]. Unfortunately, RNAi has not achieved the expected success in parasitic nematodes (schistosomes being an exception) [43]. This could be attributed to the lack of several key components of RNAi pathway [44]. Thus, targeted therapy with RNAi based approach (especially miRNA) is still in its infancy, in context of helminthic infections, that needs orchestrated support from the investors as well as the scientific community in order to stand alone as potential candidate for development of molecular medicine.

3.2 Targeting cells and signaling pathways

This area of molecular medicine remains the hottest area of research in the field of parasitic diseases after the World Health Organization (WHO) warning about the risk of post-antibiotic era. The search for novel therapeutic strategies intends to enhance pathogen killing by targeting regulatory molecules/pathways. Better understanding of disease immunobiology and cellular signaling will provide momentum to the identification of the pathways of therapeutic importance. This area of research towards the development of molecular medicine involves the use of genetically engineered antibodies, recombinant proteins, small molecules to alter signaling pathways, targeting the immunometabolic pathways, inflammasomes, etc.

3.2.1 Immunotherapeutic approach

Immunotherapy is use of biological substances (antigen/antibody, immunomodulators administration) to regulate host immune system in order to fulfill prophylactic and/or therapeutic purpose. Immunotherapy aims to trigger the immune power by directly (antigen based or active immunotherapy) or indirectly (antibody based or passive immunotherapy) [45, 46]. This section describes various immunotherapeutic strategies of molecular medicine that have been reported for the parasitic diseases.

3.2.1.1 Recombinant proteins/cytokine therapy

Cytokines are the small molecular weight, chemical messengers that regulates the immune responses in autocrine and paracrine manner. There is plethora of evidences for the involvement of cytokines in determining the pathophysiological consequences. Recombinant protein (cytokine) based therapy aims to trigger T-cell immune responses and induces parasite clearance.

In *T. cruzi* infection, TGF-β (transforming growth factor-β) has been implicated to yield pathological consequences however, treatment with TGF-β receptor kinase, SB-431542, has shown to restrict the entrance of parasite in the cardiomyocytes and disease associated cardiomyopathy [47, 48]. Additionally, cytokine combination therapy with recombinant IFN-γ and TNF-α has anti-parasitic potential [49]. In context of leishmaniasis, Murray et al. first proved the significance of targeting the cytokine, they showed that monoclonal antibody-based treatment targeting the IL-10 receptor (anti-IL-10 receptor) instigated parasite clearance by inducing NO (nitric oxide) production [50]. Likewise, combination of recombinant IFN-γ therapy with conventional chemotherapy yielded promising results in controlling the disease pathology [51]. Not much has been reported about the cytokine-based therapy in other protozoan diseases.

Conversely, in helminthic infection, MAb (monoclonal antibody) based blocking of IL-4 and IL-10 has shown disease improvement by reducing parasitic burden and inducing TH1 immune responses [52]. Likewise, IL-4 based MAb therapy, in schistosomiasis, has shown to inhibit granuloma formation and hepatic fibrosis [53]. Similar findings of marked reduction in granuloma and hepatic fibrosis were reported upon treatment with *Schistosoma* eggs along with recombinant IL-12 treatment [54]. Exogenous IL-25 based therapy has shown to potentially modulate intestinal functions by regulating IL-13 mediated STAT6 signaling in order to favor protective immune responses in intestinal nematode infection by *N. brasiliensis* [55]. Conversely, exogenous treatment with IL-13 and IL-25 triggered ILCs (innate lymphoid cells) responses and conferred protection against helminthic infections [56]. In schistosomiasis, IL-13 inhibitor, sIL-13Ralpha2-Fc has proved therapeutic benefit by preventing tissue fibrosis due to T_H2 dominated inflammatory responses [49] .

The significance of exogenous cytokine therapy has also been underlined in trichiasis, where IL-33 is known to induce thymic stromal lymphopoietin that generates polarized TH2 responses to confer protection against intestinal nematodes [57]. While, IL-25 treatment instigated TH2 responses and restricted infection induced gastrointestinal inflammation [58], MAb based blockade of IL-10 ameliorated disease pathology. There are evidences for the IL-27 mediated suppression of T-cell proliferation thus IL-27 receptor (WSX-1) knock down improved the mucosal immunity [59]. The use of immune triggering cytokines (IFN-γ, IL-12, GM-CSF) and/or blocking immunoregulatory cytokines that possesses pathological consequences holds hopes for the development of molecular medicine. Thus, therapeutic potential of cytokine therapy can be exploited alone

and/or in combination with conventional chemotherapy opening up the avenues for improving treatment outcomes.

3.2.1.2 Immune checkpoint therapy

Immune checkpoint molecules are involved in regulating the T-cell activation and functions. The expression of these molecules is enhanced during chronic infections as a result of immune subversion, thus, therapeutically targeting these molecules has shown promising results in cancer and infectious diseases vaccines [60, 61]. Indeed, T-cell dysfunctionality or exhaustion is the key for impaired T-cell responses during chronic infections; exhaustion is marked by loss of IL-2 production, reduced cytotoxicity, impaired production of pro-inflammatory mediators and reduced proliferative ability. The expression of multiple immune checkpoint molecules (PD-1, CTLA-4, LAG-3, Tim-3, TIGIT) remains the hallmark feature of exhausted cells; elevated expressions of these molecules are accompanied with progressive loss of T-cell functionality [62]. Immune checkpoint inhibitors have been novel strategy of reinvigorating the immune cell functions by abrogating the signaling by the immune checkpoint (or co-inhibitory molecules).

A number of immune checkpoint molecules have been reported in leishmaniasis including—LAG-3, Tim-3, CTLA-4, PD-1, etc. that negatively regulates T-cell functionality [63–65]. MAb based blockade of PD-1 and LAG-3 in malaria triggered pro-inflammatory cytokine responses and relieved T-cell inhibition [66]. Likewise, therapeutically targeting LAG-3 and PD-L1 restored CD4+ T-cells functions, restored follicular helper T-cells, plasma cells eventually cleared the blood stage of *Plasmodium* [67].

Unfortunately, this strategy of immune checkpoint therapy has been in its nascent stage for parasitic infections, and has yet not been used for HAT, Chagas disease, gastrointestinal protozoans as well as helminthic infections.

3.2.1.3 Immune cells and stem cell-based therapy

Immune cell manipulation offers another fascinating approach of molecular medicine to fight with parasitic diseases, when other treatment options fail to provide protective immunity [68, 69]. Direct transfer of immune cells has been holding great promises for conferring protection against protozoan, bacterial and viral infections [70]. Adoptive T-cell transfer therapy using tumor-infiltrating lymphocytes is the best example to clinical success of cellular therapy [71].

DC (dendritic cell) based vaccination approach using parasite peptide (KMP-11) elicited TH1 responses, reduced parasite load and induced lymphocyte proliferation in leishmaniasis infection [72]. Similarly, vaccination with DC along with histone H1 elicited pro-inflammatory responses (IFN-γ and IL-12), reduced the IL-10 and IL-4 producing cells and induced polarized TH1 responses [73]. Atypical progenitor cells (IL-7R+ c-kit+ cells) from malaria infected mice are potent fighters against infection, while transplantation of these cells had similar effects in disease recovery [74].

After the success of direct administration of MSCs (mesenchymal stromal cells) and antigen specific T-cells stem cell therapy has recently budded in the field of infectious diseases. MSCs have been shown to be equally important in conferring resistance against *P. berghei* infection, by suppressing IL-10, reducing the regulatory T-cells population and inducing the production of IL-12 [75, 76]. Likewise, autologous transplantation of MSCs and myoblasts has shown to significantly reduce the ventricular dysfunctions [77].

Transplantation of bone marrow mononuclear cells has marked effect on improving the inflammation and fibrosis in Chagas disease [78, 79]. Also, bone marrow transplantation holds promises for improving the quality of life in congestive heart failure due to Chagas disease [80, 81]. Adoptive immunotherapy in toxoplasmosis, by transferring CD8+ T-cells restricted parasite de-encystation; however, it failed to revert the T-cells exhaustion attributing to the short-lives of exhausted cells [82]. Further, MSCs therapy in toxoplasmosis has not been successful however, when used in combination with the spiramycin, pyrimethamine and folinic acid provided therapeutic benefits. Similarly, for coccidiosis, using adoptive transfer strategy, intraepithelial lymphocytes (IELs) and CD4+ T-cells from interferon gamma knock out (*Cryptosporidium parvum*-infected) mice has shown to provide protection against infection in naïve mice [83]. Likewise, adoptive transfer of sporozoites pulsed-DCs upon co-culture with CD4+ and CD8+ T-cells reduced parasite burden [84].

In helminthic diseases, MSC based therapy have been proven to be efficacious for reducing *Schistosoma japonicum* induced liver injury by using MSCs culture supernatant which inhibited macrophage activation by egg antigen. Macrophages primed with *N. brasiliensis* have been shown to clear the parasitic burden by neutrophil mediated mechanism of macrophage polarization [85] in strongyloidiasis. Filarial infections are associated with increased expression of Foxp3 expressing regulatory T-cells that impairs the CD4+ T-cell immunity. Regulatory T-cells targeted intervention using antibodies against CD25, glucocorticoid-induced TNF receptor family-related gene (GITR), provided cure for filarial infection [86]. In schistosomiasis, basophil depletion strategy has been shown to successfully ameliorate disease pathology and granulomatous lesions [87]. Similarly, *in vivo* DCs depletion has been an efficacious strategy to boost antigen specific T-cells expansion [88]. Antigen pulsed immune cell therapeutics has been extended to *F. hepatica* infection. DCs pulsed with parasite induces T_H1 responses and has been a viable vaccination option that protects against disease associated hepatic damage [89]. Similarly, transfer of *Hymenolepis diminuta* pulsed bone marrow derived DCs cells ameliorated colitis pathology by IL-4 signaling [90]. Therefore, cell based therapeutic strategy serves as potential molecular medicinal approach for parasitic infections.

3.2.1.4 Immunomodulators

Immunomodulators are small molecular inhibitors of signaling pathways that serve as molecular medicine for disease intervention. Imatinib, an Abl/Arg tyrosine kinase inhibitor, induces cytoskeleton remodeling to facilitate leishmanial parasite phagocytosis in the macrophages and reduces disease associated lesions [91]. Another signaling pathway inhibitor, AS-605240 (PI3K gamma inhibitor) has shown to be as efficacious as sodium stibogluconate (SSG) in the treating of *L. mexicana* infection [92]. Similarly, another PI3K inhibitor CAL-101 and IC87114 are known to effectively reduce parasitic burden by reducing the B-cells and regulatory T-cells populations [93, 94]. Another tyrosine kinase inhibitor, ibrutinib has been shown to treat leishmaniasis by triggering T_H1-polarized IFN-γ production [95]. Tellurium based immunomodulator (AS101) has shown to effectively revert t-cell anergy and promote NO production while inhibiting IL-10 signaling in *L. donovani* infection. In Chagas disease, inhibitors of GPCRs provide protection against the disease by preventing the parasite entry and infection [96]. Parasite derived thromboxane A2 signaling induces apoptosis, vasoconstriction and disease associated cardiomyopathy thus use of SQ29548, thromboxane A2 receptor antagonist abrogates the *T. cruzi* infection [97]. Conversely, platelet activating factor and leukotriene B4 induces NO production and effectively controls the parasite [98, 99].

Further, β-adrenergic receptor blockade along with carvedilol has been an effective strategy to improve clinical symptom of Chagas cardiomyopathy [100].

Tyrosine kinase inhibitors (lapatinib) have proven their efficacies in controlling Human African Trypanosomiasis (HAT) pathogenesis by blocking parasite endocytosis [101]. Furthermore, PI3Kγ/mTOR signaling inhibitors as NVP-BEZ235 restricts the *T. brucei* infection [102]. Lectin based therapy using parasite galactose-N-acetyl-D-galactosamine inhibitable lectin (Gallectin) instigates IL-12 production from DCs, T-cell proliferation and IFN-γ production [103].

Rosiglitazone, peroxisome proliferator-activator receptor gamma (PPARγ) agonist, is known to enhance phagocytic clearance of parasitized erythrocytes and reduce parasitic burden in malaria by inhibiting the mitogen-activated protein kinase (MAPK) and NF-κB signaling [104].

In helminthic infection (Strongyloidiasis), anakinra (IL-1β receptor antagonist) potentially improved innate cytokine responses (IL-33 and IL-25) eventually causing parasite expulsion [105]. Therefore, small molecular have shown potential therapeutic benefits in parasitic infection, here is just the tip of huge iceberg, research is underway to explore other molecules.

3.3 Nano-medicinal approach

The application of nanomaterials in the field of medicine for diagnosis and treatment received considerable attention in recent decades for parasitic diseases. The diagnostic potential of nanomaterials has been seen in malaria [106, 107], toxoplasmosis [108], cryptosporiodiosis [109], amebiasis [110] and leishmaniasis [111, 112].

Considering the nanoparticles as treatment option for parasitic diseases, these particles have proven efficacy in targeting the infected macrophages for treatment of VL [113]. Silver alone or in combination with chitosan nanoparticles exhibited anti-toxoplasma effects by exacerbating serum IFN-γ levels and lowering the parasitic burden [114]. Spiramycin loaded chitosan nanoparticles have shown to effectively treat toxoplasmosis [115]. In giardiasis, combination nanotherapy with silver, chitosan and curcumin nanoparticles have been shown to effectively clear the parasites from intestine and stool without any adverse effects [116]. Chitosan as nanosuspension has also shown lethal effects on *Cryptosporidium* oocysts [117]. Similarly, silver nanoparticles have also been shown to effectively reduced oocyst burden by triggering IFN-γ, without any adverse events as seen with standard therapeutic options [118].

The biodegradability and non-immunogenic properties of nanoparticles have made them suitable as delivery agents for drugs and vaccines. Nanoformulation of recombinant *P. falciparum* protein (Pfs25H) served as transmission blocking vaccine for malaria, by abrogating the parasite infectivity to mosquitoes. Similarly, polymeric vaccine using polymer poly(lactide-co-glycoside) acid (PLGA) nanoparticles with malaria antigen, VMP001, and immunostimulatory monophosphoryl A (MPL-A) triggered antigen-specific immune responses against *P. vivax* [119]. Furthermore, iron oxide nanoparticle conjugated with recombinant merozoite surface protein 1 (rMSP1) were efficiently engulfed by macrophages and DCs, that eventually triggered the pro-inflammatory responses [120]. In VL, conjugation of quercetin with gold nanoparticle [121], doxorubicin along with chitosan [122], amphotericin B as chitosan nanocapsule [123] and mannose-chitosan based nanoformulation of rifampicin served as effective delivery system for VL management [124]. Chitosan/poly (vinyl alcohol) based microspheres has also shown to abrogate the *Cryptosporidium* sporozoites attachment to the enterocytes thus served as potential oral chemotherapy for *Cryptosporidium* infection [125].

For helminthic infections, chitosan based albendazole formulation skewed the T-cell responses to T_H1 type and reduced the parasitic burden, which led to parasite clearance in echinococcosis [126, 127] as well as in toxocariasis [128]. Similarly, silver assembled on fungal (*Trichoderma harzianum*) cell wall in the form of nanoformulation improved the anti-fascioliasis potential of triclabendazole [129]. Liposomal nanoformulations (nanoparticles and nanocapsules) have been widely used for enhancing the efficacies and bioavailability of oral drugs for disease intervention. These formulations have gained significance as nanocarriers in helminthic infection due to their ability to diffuse through the intestinal mucosal layers. In schistosomiasis, liposome encapsulated praziquantel has shown significant reductions in parasitic burdens and hepatic granulomas due to increased affinity for parasite phospholipids [130]. Liposome based nanocapsules of praziquantel (PZQ-LNCs) improved the drug efficacy and ameliorated disease pathology [131]. Further, liposomal nanocapsules of miltefosine exerted potential schistosomal effects by ameliorating hepatic histology (reducing the granuloma size, number and inflammation) in single dose [132].

Nanoformulation has also been used for vaccine development and as adjuvants, self-assembling protein nanoparticles (SAPN) have shown to trigger protective antibodies and long-lived memory responses to confer sterile protection against malarial parasites (*P. berghei* and *P. falciparum*) [133]. SAPNs have also been used for delivering the epitopes to induce CD4+ and CD8+ T-cell responses against *Toxoplasma gondii* [134]. Archaea based nanoformulations (archaeosomes) have been used as adjuvant as a part of prophylactic vaccine against *T. cruzi*, instigated humoral as well as cell-mediated immune responses (T_H1 responses) leading to marked reduction in parasitic burdens [135]. Cationic solid lipid nanoparticles have been successfully used as adjuvant as part of prime-boost strategy to reduce the parasitic burdens during VL. The vaccination triggered IFN-γ production, NO production and high levels of immunoglobulins (IgG1 and IgG2a) [136]. Therefore, nanoparticles served as viable, safe and effective vaccine platform as well as development of molecular medicine for cost effective vaccine delivery.

4. Conclusion

In this world where cost of developing medicine for parasitic infections remain the greatest challenge, drug developers are embracing molecular medicine approach that promises to deal with the parasitic infections and improves the chances of successful treatment. Molecular medicine has revolutionized the field of drug discovery/development however, there are significant hurdles in turning the promise into reality. Perhaps, gradually but it is shaping the future of medicine with the help of molecular platforms, better bioinformatics services and better pharmacogenomic analyses has greatly facilitated the scientific community and the stakeholders to come on common platform to fight against the parasitic diseases.

Author details

Bhawana Singh
Department of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

*Address all correspondence to: bhavanasonali9@gmail.com

References

[1] Gannon F. Molecular medicine: Trendy title or new reality? EMBO Reports. 2003;**4**(8):733

[2] Porteus MH. Towards a new era in medicine: Therapeutic genome editing. Genome Biology. 2015;**16**:286

[3] Menoret S et al. Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. The FASEB Journal. 2013;**27**(2):703-711

[4] Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology. 2013;**31**(7):397-405

[5] Carlson DF, Fahrenkrug SC, Hackett PB. Targeting DNA with fingers and TALENs. Molecular Therapy—Nucleic Acids. 2012;**1**:e3

[6] Bogdanove AJ, Voytas DF. TAL effectors: Customizable proteins for DNA targeting. Science. 2011;**333**(6051):1843-1846

[7] Karimian A et al. CRISPR/Cas9 technology as a potent molecular tool for gene therapy. Journal of Cellular Physiology. 2019;**234**(8):12267-12277

[8] Wagner JC et al. Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. Nature Methods. 2014;**11**(9):915-918

[9] Ghorbal M et al. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. Nature Biotechnology. 2014;**32**(8):819-821

[10] Sidik SM et al. A genome-wide CRISPR screen in toxoplasma identifies essential Apicomplexan genes. Cell. 2016;**166**(6):1423-1435 e12

[11] Sidik SM, Huet D, Lourido S. CRISPR-Cas9-based genome-wide screening of *Toxoplasma gondii*. Nature Protocols. 2018;**13**(1):307-323

[12] Chiurillo MA et al. Different roles of mitochondrial calcium uniporter complex subunits in growth and infectivity of *Trypanosoma cruzi*. mBio. 2017;**8**(3). pii: e00574-17

[13] Lander N et al. CRISPR/Cas9-mediated endogenous C-terminal tagging of *Trypanosoma cruzi* genes reveals the Acidocalcisome localization of the inositol 1,4,5-trisphosphate receptor. The Journal of Biological Chemistry. 2016;**291**(49):25505-25515

[14] Soares Medeiros LC et al. Rapid, selection-free, high-efficiency genome editing in protozoan parasites using CRISPR-Cas9 ribonucleoproteins. mBio. 2017;**8**(6). pii: e01788-17

[15] Beneke T et al. A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. Royal Society Open Science. 2017;**4**(5):170095

[16] Peng D et al. CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. mBio. 2014;**6**(1):e02097-e02014

[17] Janssen BD et al. CRISPR/Cas9-mediated gene modification and gene knock out in the human-infective parasite *Trichomonas vaginalis*. Scientific Reports. 2018;**8**(1):270

[18] Arunsan P et al. Programmed knockout mutation of liver fluke granulin attenuates virulence of infection-induced hepatobiliary morbidity. eLife. 2019;**8**:e41463

[19] Ngo H et al. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proceedings of the National Academy of Sciences of the United States of America. 1998;**95**(25):14687-14692

[20] DaRocha WD et al. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. Molecular and Biochemical Parasitology. 2004;**133**(2): 175-186

[21] Zhang WW, Matlashewski G. Analysis of antisense and double stranded RNA downregulation of A2 protein expression in *Leishmania donovani*. Molecular and Biochemical Parasitology. 2000;**107**(2):315-319

[22] Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite Leishmania. Molecular and Biochemical Parasitology. 2003;**128**(2):217-228

[23] Abdulla MH et al. RNA interference of *Trypanosoma brucei* cathepsin B and L affects disease progression in a mouse model. PLoS Neglected Tropical Diseases. 2008;**2**(9):e298

[24] Prestrud P, Krogsrud J, Gjertz I. The occurrence of rabies in the Svalbard Islands of Norway. Journal of Wildlife Diseases. 1992;**28**(1):57-63

[25] Montalvetti A et al. Farnesyl pyrophosphate synthase is an essential enzyme in *Trypanosoma brucei*. In vitro RNA interference and in vivo inhibition studies. The Journal of Biological Chemistry. 2003;**278**(19):17075-17083

[26] Alsford S et al. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Research. 2011;**21**(6):915-924

[27] Goringer HU et al. RNA aptamers as potential pharmaceuticals against infections with African trypanosomes. Handbook of Experimental Pharmacology. 2006;**173**:375-393

[28] Noonpakdee W et al. Inhibition of plasmodium falciparum proliferation in vitro by antisense oligodeoxynucleotides against malarial topoisomerase II. Biochemical and Biophysical Research Communications. 2003;**302**(4):659-664

[29] Foger F et al. Inhibition of malarial topoisomerase II in *Plasmodium falciparum* by antisense nanoparticles. International Journal of Pharmaceutics. 2006;**319**(1-2):139-146

[30] Attasart P et al. Inhibition of *Plasmodium falciparum* proliferation in vitro by double-stranded RNA nanoparticle against malaria topoisomerase II. Experimental Parasitology. 2016;**164**:84-90

[31] Ravaee R et al. Synthetic siRNAs effectively target cystein protease 12 and alpha-actinin transcripts in *Trichomonas vaginalis*. Experimental Parasitology. 2015;**157**:30-34

[32] Hussein AS, Kichenin K, Selkirk ME. Suppression of secreted acetylcholinesterase expression in *Nippostrongylus brasiliensis* by RNA interference. Molecular and Biochemical Parasitology. 2002;**122**(1):91-94

[33] Tchoubrieva EB et al. Vector-based RNA interference of cathepsin B1 in *Schistosoma mansoni*. Cellular and Molecular Life Sciences. 2010;**67**(21): 3739-3748

[34] Guidi A et al. Application of RNAi to genomic drug target validation in Schistosomes. PLoS Neglected Tropical Diseases. 2015;**9**(5):e0003801

[35] Aboobaker AA, Blaxter ML. Use of RNA interference to investigate gene function in the human filarial nematode parasite *Brugia malayi*. Molecular and Biochemical Parasitology. 2003;**129**(1): 41-51

[36] Chen MX et al. Identification and characterization of microRNAs in *Trichinella spiralis* by comparison

with *Brugia malayi* and *Caenorhabditis elegans*. Parasitology Research. 2011;**109**(3):553-558

[37] Wang J et al. Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Research. 2011;**21**(9):1462-1477

[38] Chen MX et al. *Angiostrongylus cantonensis*: Identification and characterization of microRNAs in male and female adults. Experimental Parasitology. 2011;**128**(2):116-120

[39] Ai L et al. Characterization of microRNAs in *Taenia saginata* of zoonotic significance by Solexa deep sequencing and bioinformatics analysis. Parasitology Research. 2012;**110**(6):2373-2378

[40] Cucher M et al. High-throughput characterization of Echinococcus spp. metacestode miRNomes. International Journal for Parasitology. 2015;**45**(4):253-267

[41] Winter AD et al. Diversity in parasitic nematode genomes: The microRNAs of *Brugia pahangi* and *Haemonchus contortus* are largely novel. BMC Genomics. 2012;**13**:4

[42] Lustigman S et al. RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. Molecular and Biochemical Parasitology. 2004;**138**(2):165-170

[43] Knox DP et al. RNA interference in parasitic nematodes of animals: A reality check? Trends in Parasitology. 2007;**23**(3):105-107

[44] Viney ME, Thompson FJ. Two hypotheses to explain why RNA interference does not work in animal parasitic nematodes. International Journal for Parasitology. 2008;**38**(1):43-47

[45] Hsueh EC, Morton DL. Antigen-based immunotherapy of melanoma: Canvaxin therapeutic polyvalent cancer vaccine. Seminars in Cancer Biology. 2003;**13**(6):401-407

[46] Morsink LM, Walter RB. Novel monoclonal antibody-based therapies for acute myeloid leukemia. Best Practice & Research. Clinical Haematology. 2019;**32**(2):116-126

[47] Waghabi MC et al. SB-431542, a transforming growth factor beta inhibitor, impairs *Trypanosoma cruzi* infection in cardiomyocytes and parasite cycle completion. Antimicrobial Agents and Chemotherapy. 2007;**51**(8):2905-2910

[48] Waghabi MC et al. Pharmacological inhibition of transforming growth factor beta signaling decreases infection and prevents heart damage in acute Chagas' disease. Antimicrobial Agents and Chemotherapy. 2009;**53**(11): 4694-4701

[49] Munoz-Fernandez MA, Fernandez MA, Fresno M. Synergism between tumor necrosis factor-alpha and interferon-gamma on macrophage activation for the killing of intracellular *Trypanosoma cruzi* through a nitric oxide-dependent mechanism. European Journal of Immunology. 1992;**22**(2):301-307

[50] Murray HW et al. Determinants of response to interleukin-10 receptor blockade immunotherapy in experimental visceral leishmaniasis. The Journal of Infectious Diseases. 2003;**188**(3):458-464

[51] Badaro R et al. Treatment of visceral leishmaniasis with pentavalent antimony and interferon gamma. The New England Journal of Medicine. 1990;**322**(1):16-21

[52] Macedo MS et al. Immunomodulation induced by *Ascaris suum* extract in mice: Effect of anti-interleukin-4 and anti-interleukin-10 antibodies. Scandinavian Journal of Immunology. 1998;**47**(1):10-18

[53] Cheever AW et al. Anti-IL-4 treatment of *Schistosoma mansoni*-infected mice inhibits development of T cells and non-B, non-T cells expressing Th2 cytokines while decreasing egg-induced hepatic fibrosis. Journal of Immunology. 1994;**153**(2):753-759

[54] Wynn TA et al. An IL-12-based vaccination method for preventing fibrosis induced by schistosome infection. Nature. 1995;**376**(6541):594-596

[55] Zhao A et al. Critical role of IL-25 in nematode infection-induced alterations in intestinal function. Journal of Immunology. 2010;**185**(11):6921-6929

[56] Huang Y et al. IL-25-responsive, lineage-negative KLRG1(hi) cells are multipotential 'inflammatory' type 2 innate lymphoid cells. Nature Immunology. 2015;**16**(2):161-169

[57] Humphreys NE et al. IL-33, a potent inducer of adaptive immunity to intestinal nematodes. Journal of Immunology. 2008;**180**(4):2443-2449

[58] Owyang AM et al. Interleukin 25 regulates type 2 cytokine-dependent immunity and limits chronic inflammation in the gastrointestinal tract. The Journal of Experimental Medicine. 2006;**203**(4):843-849

[59] Artis D et al. The IL-27 receptor (WSX-1) is an inhibitor of innate and adaptive elements of type 2 immunity. Journal of Immunology. 2004;**173**(9):5626-5634

[60] Pauken KE, Wherry EJ. Overcoming T cell exhaustion in infection and cancer. Trends in Immunology. 2015; **36**(4):265-276

[61] Butt AQ, Mills KH. Immuno-suppressive networks and checkpoints controlling antitumor immunity and their blockade in the development of cancer immunotherapeutics and vaccines. Oncogene. 2014;**33**(38):4623-4631

[62] Blackburn SD et al. Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. Nature Immunology. 2009;**10**(1):29-37

[63] Singh B et al. A molecular signature for CD8(+) T cells from visceral leishmaniasis patients. Parasite Immunology. 2019;**41**(11):e12669

[64] Esch KJ et al. Programmed death 1-mediated T cell exhaustion during visceral leishmaniasis impairs phagocyte function. Journal of Immunology. 2013;**191**(11):5542-5550

[65] Murphy ML et al. B7-2 blockade enhances T cell responses to *Leishmania donovani*. Journal of Immunology. 1997;**159**(9):4460-4466

[66] Doe HT et al. Expression of PD-1/ LAG-3 and cytokine production by CD4(+) T cells during infection with Plasmodium parasites. Microbiology and Immunology. 2016;**60**(2):121-131

[67] Butler NS et al. Therapeutic blockade of PD-L1 and LAG-3 rapidly clears established blood-stage Plasmodium infection. Nature Immunology. 2011;**13**(2):188-195

[68] Fajardo-Moser M, Berzel S, Moll H. Mechanisms of dendritic cell-based vaccination against infection. International Journal of Medical Microbiology. 2008;**298**(1-2):11-20

[69] Delamarre L, Mellman I. Harnessing dendritic cells for immunotherapy. Seminars in Immunology. 2011;**23**(1):2-11

[70] Steinman RM. Dendritic cells in vivo: A key target for a new vaccine science. Immunity. 2008;**29**(3):319-324

[71] Rosenberg SA et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clinical Cancer Research. 2011;**17**(13):4550-4557

[72] Agallou M, Margaroni M, Karagouni E. Cellular vaccination with bone marrow-derived dendritic cells pulsed with a peptide of *Leishmania infantum* KMP-11 and CpG oligonucleotides induces protection in a murine model of visceral leishmaniasis. Vaccine. 2011;**29**(31):5053-5064

[73] Agallou M et al. Vaccination with Leishmania histone H1-pulsed dendritic cells confers protection in murine visceral leishmaniasis. Vaccine. 2012;**30**(34):5086-5093

[74] Belyaev NN et al. Induction of an IL7-R(+)c-Kit(hi) myelolymphoid progenitor critically dependent on IFN-gamma signaling during acute malaria. Nature Immunology. 2010;**11**(6):477-485

[75] Asami M et al. Susceptibility of multipotent haemopoietic stem cell deficient W/Wv mice to *Plasmodium berghei*-infection. Immunology and Cell Biology. 1991;**69**(**Pt 5**):355-360

[76] Thakur RS et al. Mesenchymal stem cells play an important role in host protective immune responses against malaria by modulating regulatory T cells. European Journal of Immunology. 2013;**43**(8):2070-2077

[77] Guarita-Souza LC et al. Simultaneous autologous transplantation of cocultured mesenchymal stem cells and skeletal myoblasts improves ventricular function in a murine model of Chagas disease. Circulation. 2006;**114**(1 Suppl):I120-I124

[78] Soares MB et al. Transplanted bone marrow cells repair heart tissue and reduce myocarditis in chronic chagasic mice. The American Journal of Pathology. 2004;**164**(2):441-447

[79] Topalis P et al. Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. Insect Molecular Biology. 2008;**17**(1):87-89

[80] Vilas-Boas F et al. Bone marrow cell transplantation in Chagas' disease heart failure: Report of the first human experience. Arquivos Brasileiros de Cardiologia. 2011;**96**(4):325-331

[81] Vilas-Boas F et al. Early results of bone marrow cell transplantation to the myocardium of patients with heart failure due to Chagas disease. Arquivos Brasileiros de Cardiologia. 2006;**87**(2):159-166

[82] Bhadra R, Cobb DA, Khan IA. Donor CD8+ T cells prevent *Toxoplasma gondii* de-encystation but fail to rescue the exhausted endogenous CD8+ T cell population. Infection and Immunity. 2013;**81**(9):3414-3425

[83] Tessema TS, Dauber E, Petry F. Adoptive transfer of protective immunity from *Cryptosporidium parvum*-infected interferon-gamma and interleukin-12-deficient mice to naive recipients. Vaccine. 2009;**27**(47):6575-6581

[84] Bedi B, McNair NN, Mead JR. Dendritic cells play a role in host susceptibility to *Cryptosporidium parvum* infection. Immunology Letters. 2014;**158**(1-2):42-51

[85] Chen F et al. Neutrophils prime a long-lived effector macrophage phenotype that mediates accelerated helminth expulsion. Nature Immunology. 2014;**15**(10):938-946

[86] Taylor MD et al. Removal of regulatory T cell activity reverses hyporesponsiveness and leads to filarial parasite clearance in vivo. Journal of Immunology. 2005;**174**(8):4924-4933

[87] Anyan WK et al. Basophil depletion downregulates *Schistosoma mansoni* egg-induced granuloma formation. Parasitology International. 2013;**62**(6):508-513

[88] Lundie RJ et al. A central role for hepatic conventional dendritic cells

in supporting Th2 responses during helminth infection. Immunology and Cell Biology. 2016;**94**(4):400-410

[89] Falcon CR et al. Adoptive transfer of dendritic cells pulsed with *Fasciola hepatica* antigens and lipopolysaccharides confers protection against fasciolosis in mice. The Journal of Infectious Diseases. 2012;**205**(3):506-514

[90] Matisz CE et al. Suppression of colitis by adoptive transfer of helminth antigen-treated dendritic cells requires interleukin-4 receptor-alpha signaling. Scientific Reports. 2017;**7**:40631

[91] Wetzel DM, McMahon-Pratt D, Koleske AJ. The Abl and Arg kinases mediate distinct modes of phagocytosis and are required for maximal Leishmania infection. Molecular and Cellular Biology. 2012;**32**(15):3176-3186

[92] Cummings HE et al. Critical role for phosphoinositide 3-kinase gamma in parasite invasion and disease progression of cutaneous leishmaniasis. Proceedings of the National Academy of Sciences of the United States of America. 2012;**109**(4):1251-1256

[93] Khadem F et al. Pharmacological inhibition of p110delta subunit of PI3K confers protection against experimental leishmaniasis. The Journal of Antimicrobial Chemotherapy. 2017;**72**(2):467-477

[94] Vishwakarma P et al. Ammonium trichloro [1,2-ethanediolato-O,O′]-tellurate cures experimental visceral leishmaniasis by redox modulation of Leishmania donovani trypanothione reductase and inhibiting host integrin linked PI3K/Akt pathway. Cellular and Molecular Life Sciences. 2018;**75**(3):563-588

[95] Dubovsky JA et al. Ibrutinib is an irreversible molecular inhibitor of ITK driving a Th1-selective pressure in T lymphocytes. Blood. 2013;**122**(15):2539-2549

[96] Croxford JL et al. Effects of cannabinoid treatment on Chagas disease pathogenesis: Balancing inhibition of parasite invasion and immunosuppression. Cellular Microbiology. 2005;**7**(11):1592-1602

[97] Silva JF et al. Mechanisms of vascular dysfunction in acute phase of *Trypanosoma cruzi* infection in mice. Vascular Pharmacology. 2016;**82**:73-81

[98] Talvani A et al. Leukotriene B(4) induces nitric oxide synthesis in *Trypanosoma cruzi*-infected murine macrophages and mediates resistance to infection. Infection and Immunity. 2002;**70**(8):4247-4253

[99] Aliberti JC et al. Platelet-activating factor induces nitric oxide synthesis in *Trypanosoma cruzi*-infected macrophages and mediates resistance to parasite infection in mice. Infection and Immunity. 1999;**67**(6):2810-2814

[100] Botoni FA et al. A randomized trial of carvedilol after renin-angiotensin system inhibition in chronic Chagas cardiomyopathy. American Heart Journal. 2007;**153**(4):544 e1-544 e8

[101] Woodring JL et al. Evaluation of aromatic 6-substituted Thienopyrimidines as scaffolds against parasites that cause Trypanosomiasis, Leishmaniasis, and Malaria. MedChemComm. 2015;**6**(2):339-346

[102] Diaz-Gonzalez R et al. The susceptibility of trypanosomatid pathogens to PI3/mTOR kinase inhibitors affords a new opportunity for drug repurposing. PLoS Neglected Tropical Diseases. 2011;**5**(8):e1297

[103] Ivory CP, Chadee K. Activation of dendritic cells by the Gal-lectin of *Entamoeba histolytica* drives Th1 responses in vitro and in vivo. European Journal of Immunology. 2007;**37**(2):385-394

[104] Serghides L et al. Rosiglitazone modulates the innate immune

response to *Plasmodium falciparum* infection and improves outcome in experimental cerebral malaria. The Journal of Infectious Diseases. 2009;**199**(10):1536-1545

[105] Zaiss MM et al. IL-1beta suppresses innate IL-25 and IL-33 production and maintains helminth chronicity. PLoS Pathogens. 2013;**9**(8):e1003531

[106] Guirgis BS et al. Gold nanoparticle-based fluorescence immunoassay for malaria antigen detection. Analytical and Bioanalytical Chemistry. 2012;**402**(3):1019-1027

[107] Thiramanas R et al. Sensitivity and specificity of PS/AA-modified nanoparticles used in malaria detection. Microbial Biotechnology. 2013;**6**(4):406-413

[108] Wang H et al. A piezoelectric immunoagglutination assay for *Toxoplasma gondii* antibodies using gold nanoparticles. Biosensors & Bioelectronics. 2004;**19**(7):701-709

[109] Weigum SE et al. Amplification-free detection of *Cryptosporidium parvum* nucleic acids with the use of DNA/RNA-directed gold nanoparticle assemblies. The Journal of Parasitology. 2013;**99**(5):923-926

[110] Hemadi A et al. Bioconjugated fluorescent silica nanoparticles for the rapid detection of *Entamoeba histolytica*. Acta Tropica. 2015;**145**:26-30

[111] Andreadou M et al. A novel non-amplification assay for the detection of Leishmania spp. in clinical samples using gold nanoparticles. Journal of Microbiological Methods. 2014;**96**:56-61

[112] de la Escosura-Muniz A et al. Magnetic bead/gold nanoparticle double-labeled primers for electrochemical detection of isothermal amplified Leishmania DNA. Small. 2016;**12**(2):205-213

[113] Kunjachan S et al. Physicochemical and biological aspects of macrophage-mediated drug targeting in anti-microbial therapy. Fundamental & Clinical Pharmacology. 2012;**26**(1):63-71

[114] Gaafar MR et al. Chitosan and silver nanoparticles: Promising anti-toxoplasma agents. Experimental Parasitology. 2014;**143**:30-38

[115] Hagras NA et al. Successful treatment of acute experimental toxoplasmosis by spiramycin-loaded chitosan nanoparticles. Experimental Parasitology. 2019;**204**:107717

[116] Said DE, Elsamad LM, Gohar YM. Validity of silver, chitosan, and curcumin nanoparticles as anti-Giardia agents. Parasitology Research. 2012;**111**(2):545-554

[117] Ahmed SA, El-Mahallawy HS, Karanis P. Inhibitory activity of chitosan nanoparticles against *Cryptosporidium parvum* oocysts. Parasitology Research. 2019;**118**(7):2053-2063

[118] Gaafar MR et al. Silver nanoparticles as a therapeutic agent in experimental cyclosporiasis. Experimental Parasitology. 2019;**207**:107772

[119] Moon JJ et al. Antigen-displaying lipid-enveloped PLGA nanoparticles as delivery agents for a *Plasmodium vivax* malaria vaccine. PLoS One. 2012;**7**(2): e31472

[120] Pusic K et al. Iron oxide nanoparticles as a clinically acceptable delivery platform for a recombinant blood-stage human malaria vaccine. The FASEB Journal. 2013;**27**(3):1153-1166

[121] Das S et al. One pot synthesis of gold nanoparticles and application in chemotherapy of wild and resistant type visceral leishmaniasis. Colloids and Surfaces. B, Biointerfaces. 2013;**107**:27-34

[122] Kunjachan S et al. Chitosan-based macrophage-mediated drug targeting

for the treatment of experimental visceral leishmaniasis. Journal of Microencapsulation. 2011;**28**(4):301-310

[123] Asthana S et al. Immunoadjuvant chemotherapy of visceral leishmaniasis in hamsters using amphotericin B-encapsulated nanoemulsion template-based chitosan nanocapsules. Antimicrobial Agents and Chemotherapy. 2013;**57**(4):1714-1722

[124] Chaubey P, Mishra B. Mannose-conjugated chitosan nanoparticles loaded with rifampicin for the treatment of visceral leishmaniasis. Carbohydrate Polymers. 2014;**101**:1101-1108

[125] Luzardo Alvarez A et al. In vitro evaluation of the suppressive effect of chitosan/poly(vinyl alcohol) microspheres on attachment of *C. parvum* to enterocytic cells. European Journal of Pharmaceutical Sciences. 2012;**47**(1):215-227

[126] Abulaihaiti M et al. Efficacy of albendazole-chitosan microsphere-based treatment for alveolar Echinococcosis in mice. PLoS Neglected Tropical Diseases. 2015;**9**(9):e0003950

[127] Liang W et al. Efficacy of albendazole chitosan microspheres against *Echinococcus granulosus* infection in mice. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 2014;**32**(3):188-192

[128] Barrera MG et al. In vivo evaluation of albendazole microspheres for the treatment of *Toxocara canis* larva migrans. European Journal of Pharmaceutics and Biopharmaceutics. 2010;**75**(3):451-454

[129] Gherbawy YA et al. The anti-fasciolasis properties of silver nanoparticles produced by *Trichoderma harzianum* and their improvement of the anti-fasciolasis drug triclabendazole. International Journal of Molecular Sciences. 2013;**14**(11):21887-21898

[130] Frezza TF et al. Liposomal-praziquantel: Efficacy against *Schistosoma mansoni* in a preclinical assay. Acta Tropica. 2013;**128**(1):70-75

[131] Amara RO et al. Praziquantel-lipid nanocapsules: An oral nanotherapeutic with potential *Schistosoma mansoni* tegumental targeting. International Journal of Nanomedicine. 2018;**13**: 4493-4505

[132] El-Moslemany RM et al. Miltefosine lipid nanocapsules: Intersection of drug repurposing and nanotechnology for single dose oral treatment of pre-patent *Schistosomiasis mansoni*. Acta Tropica. 2016;**159**:142-148

[133] Kaba SA et al. Protective antibody and CD8+ T-cell responses to the *Plasmodium falciparum* circumsporozoite protein induced by a nanoparticle vaccine. PLoS One. 2012;**7**(10):e48304

[134] El Bissati K et al. Effectiveness of a novel immunogenic nanoparticle platform for Toxoplasma peptide vaccine in HLA transgenic mice. Vaccine. 2014;**32**(26):3243-3248

[135] Higa LH et al. Archaeosomes display immunoadjuvant potential for a vaccine against Chagas disease. Human Vaccines & Immunotherapeutics. 2013;**9**(2):409-412

[136] Saljoughian N et al. Development of novel prime-boost strategies based on a tri-gene fusion recombinant L. tarentolae vaccine against experimental murine visceral leishmaniasis. PLOS Neglected Tropical Diseases. 2013;**7**(4):e2174

Section 2

Genes to Personalized Medicine and Molecular Therapie

Chapter 3

New Perspectives in Personalization of Therapy for Hematological Cancers

Małgorzata Rogalińska

Abstract

A progress in treatment of hematological cancers was achieved. Unfortunately, some youngsters, because of rare genetic alterations that are not easy to detect, as well as heavily pretreated old patients, because of coexisting diseases that lead to changes in patient metabolism, do not respond to therapy. Moreover, sometimes familiar diversities and alterations on genetic or epigenetic level that could be transferred on diversities in metabolism or cell signaling might be a reason why patients do not respond to therapy. Interestingly, for older patients a resistance to therapy could also occur as a reason of drug cross-reactivity. For designing of effective anticancer therapy for patient with chronic lymphocytic leukemia before drug administration, patient's leukemic cell response to anticancer drug(s) should be checked. Moreover, for patient response to treatment, also drugs prescribed previously by other medical doctors or even patients' diet could be important for achieving therapeutic success of therapy. Therefore it is important to choose the effective drugs before their administration to patient that will improve treatment efficacy and exclude resistance to therapy. It must be stated that the special attention for personalized therapy tests should be focused on patients previously resistant to therapy, more sensitive to drugs or heavily pretreated.

Keywords: personalization of therapy, resistance to treatment, anticancer drugs

1. Introduction

Several studies around the world are focus on molecular aspects related with developing of cancer. It seems to be not easy, because of high complexity of carcinogenesis and a difference between patients in disease progressing based on genetic, epigenetic or even environmental alteration. It is possible that in our bodies expression of proteins or external factors regulate hormone expression. Moreover, hormones could affect cell signal transduction and metabolism. It looks like we have some logical plan that coordinates gene expression, but because of potential involvement of many factors and familiar, sometimes even personal, predispositions for particular type of disease, diagnostic is not easy. The personal differences in disease progression activity and response to treatment reveal that even now cancers are the leading cause of death. The best way to avoid carcinogenesis is a health prophylactic since childhood. It is more likely that cancer or other diseases related with some disturbances in metabolism, except genetically related

documented cases, will develop in the group of overweight people with metabolic or hormonal disturbances and without physical exercises. For good health condition since childhood, we need to supplement probiotic bacteria to have bacterial flora in intestines that could cooperate somehow and save people against most infections and inflammation [1]. Moreover, it is very important what we consume. Food processing should be ecological, without unhealthy chemical additives [2, 3]. It must be underlined that health prophylactic might decrease the risk of getting ill or even save us from many diseases, including cancers. Proper screening tests that are able to find health problems in not advanced phase of disease are very important for prophylactic. It is extremely important to discover disease in early stage of development so we can effectively help patient to cure the disease, achieving success. It is always better and also cheaper to prevent than to cure diseases.

Before we had enough knowledge to cure successive cancers, the most effective in the war against cancers seems to be a prophylactic. The prophylactic tests should be designed for each person individually based on personal or genetic predispositions, work-related conditions, weight, or diet. Prophylactic test program should be organized by the Ministry of Health and paid from money directed to health service from our taxes.

The suitable prophylactic for professions, and increasing knowledge about disease development, could enlarge human life and decrease number of cases of cardiovascular diseases (e.g. stroke or heart attack). Therefore, we must remember that the special diet and physical exercises are able to regulate our hormone level, metabolism, and cell signaling leading to homeostasis [4, 5]. The disturbances in homeostasis could accompany the development of many diseases, including cancers.

There are several different types of lymphoproliferative disorders, characterized by disturbances in cell cycle and/or signal transduction that lead to fast proliferation of cancer cells (acute form) [6]. The other chronic type of neoplasm is caused, for example, by abrogation in apoptosis induction leading to accumulation of leukemic cells in peripheral blood. It was previously reported that in patient blood a quiescent and cycling cells were found that could be a problem in diagnostics and treatment [7–9].

Chronic lymphocytic leukemia (CLL), except bone marrow transplantation, for some patients is incurable. CLL usually occurs in older individuals but in youngsters could transform into a more aggressive form with fast developing disease [8].

Interestingly, a high heterogeneity in disease development, the characteristic accumulation in patient blood of mainly resting B lymphocytes, was observed. Leukemic cells accumulate in patient peripheral blood, because of inhibition in induction of apoptosis.

Several factors, even patient diet, could affect disease development and transformation into active form. It could lead to disturbances in cell cycle leading to increased cancer cell proliferation index. It could also induce changes in cell sensitivity to drug(s) [10]. The results of experiments directed toward personalization of therapy for CLL patients revealed that it is important to analyze before drug(s) administration if drug(s) will be active for this patient curing [10–14]. Moreover, during leukemic cell incubation with drugs, we are able to check also patients' cell sensitivity to in vitro incubation conditions. If in control untreated cells, cell viability decreases similarly like in treated ones, it is an indication that patient might need a reduction in drug dose in vivo [15]. Moreover, we need more knowledge on this topic to be able precisely to transform drug doses obtained in patient's blood into in vitro conditions. It must be underlined that for some patients more sensitive to drug(s) than others using above tests in vitro we are able to choose the

proper drug(s) dose to optimize drug(s) doses administration in vivo. For patients responding to treatment in nonstandard way and for patients more sensitive than others to drugs, a personalized therapy test could be a chance to choose effective treatment for curing to avoid drug resistance, patient fatigue or sometimes also secondary cancers. Moreover, the ineffective treatment might reduce patients' life longevity. Therefore it is very important to check before administration if drug will be active to patient or in the case of resistance to search for other potentially effective drug.

The special value of personalized therapy tests is presented in **Figure 1**. The representative combined results obtained for CLL patient by tests such as cell viability (flow cytometry; Vybrant Apoptosis Assay), analyses of thermal profiles of nuclei (differential scanning calorimetry) and PARP expression/proteolytic cleavage (Western blot) for patient confirm the value for such test in choosing effective therapy for CLL patient (**Figure 1**). The representative results for CLL patient confirm that leukemic cells were more sensitive to the combination of cladribine and cyclophosphamide/mafosfamide than the combination of fludarabine and cyclophosphamide/mafosfamide. For this patient adding rituximab does not change anything in cells' reactivity to drugs (compare CM and Rit20CM results). The combination of cladribine and cyclophosphamide/mafosfamide will be more effective for this patient curing than the combination including fludarabine. Moreover, the addition of monoclonal antibody to CM combination did not change anything when we analyze all results obtained for patient.

In **Figure 2** there is an explanation why we should analyze each patient results separately. It must be stated that in the median results of cell viability presented in **Figure 2** (Section A), there are also results of both patients presented in Sections B (better responder to CM than FM) and C (resistant to therapy). The median results of cell viability for patients B and C could be misleading and suggest that both patients should be administered with fludarabine and cyclophosphamide. For both patients (see **Figure 2B** and **C**) fludarabine will not be an optimal drug for curing.

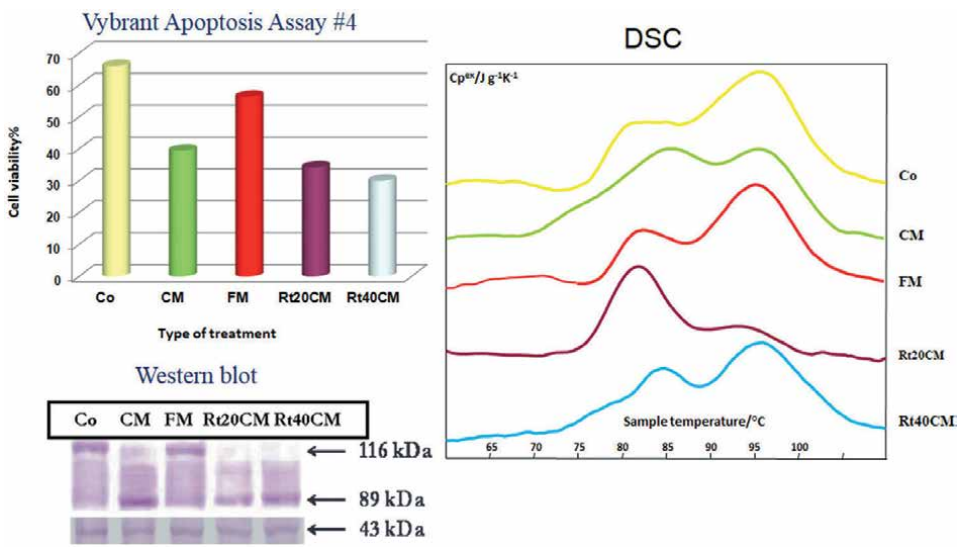


Figure 1.
Personalized therapy for CLL patient. Results of cell sensitivity with anticancer agents. Cell viability of chronic lymphocytic leukemia cells incubated for 48 h without Co and with anticancer drugs; Co, controlled untreated CLL cells; CM, cladribine + mafosfamide; FM, fludarabine + mafosfamide; RtCM (20 or 40 μM), rituximab + CM. Differences in thermal profiles were analyzed by differential scanning calorimetry (DSC). Analysis of protein expression related to apoptosis (PARP cleavage; 89 kDa) or actin (43 kDa) was studied by Western blot.

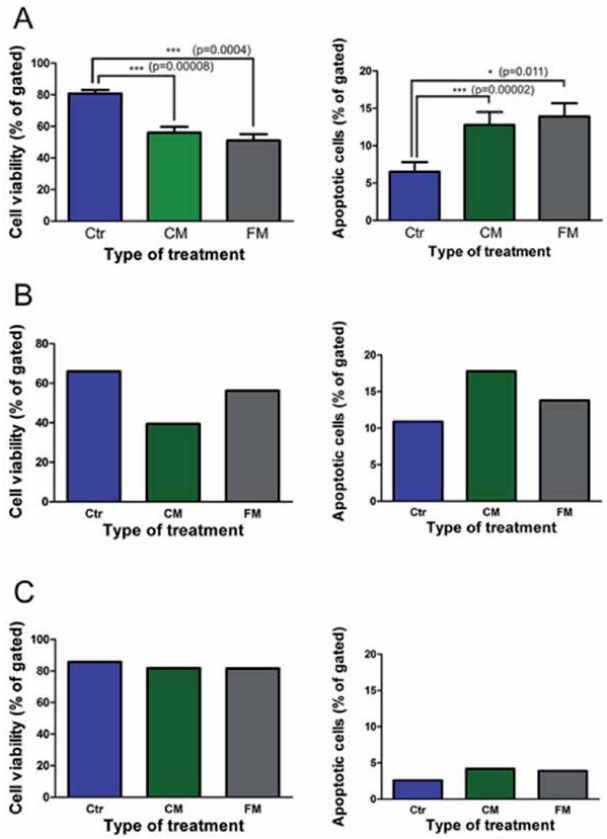


Figure 2.
Cell viability of chronic lymphocytic leukemia cells incubated for 48 h without Ctr and with anticancer drugs; CM, cladribine + mafosfamide; FM, fludarabine + mafosfamide; ***A****, median value of cell viability;* ***B****, results obtained for cells of CLL patient (better responded to CM than to FM);* ***C****, results obtained for cells of CLL patient resistant to treatment.*

Using above tests we can also monitor the differences in drug sensitivity during disease development (**Figure 3**) [10]. As presented in **Figure 3** results of cell sensitivity to anticancer agents could change during disease development. In active form of disease (year 2013), a higher leukocytosis than 18 months earlier results (year 2011) was noticed. For the same patient, we can observe differences in leukemic cell sensitivity to the same anticancer agents, i.e., combinations of cladribine + mafosfamide (cyclophosphamide), CM; fludarabine + mafosfamide (cyclophosphamide), FM; rituximab + CM (RCM); kinetin riboside (RK), that confirm the usefulness of personalized therapy tests also in monitoring of disease development and the changes in drug sensitivity [10].

Results of cell viability of chronic lymphocytic leukemia cells incubated for 48 h (24, 48 h) without Ctr and with anticancer drugs; CM, cladribine + mafosfamide; FM, fludarabine + mafosfamide; RCM, rituximab + CM; Rit, rituximab; RK, kinetin riboside [10].

The existing data confirm that in the development of CLL, some disturbances in cell signaling [16], apoptosis inhibition, or changes on epigenetic level [17–20] are observed. Moreover, the results of studies confirm that there are several molecules generated from encoded sequences of genes, characterized as miRNA that could affect gene expression, mainly by silencing. It could also be involved in regulation of protein synthesis important for cell cycle.

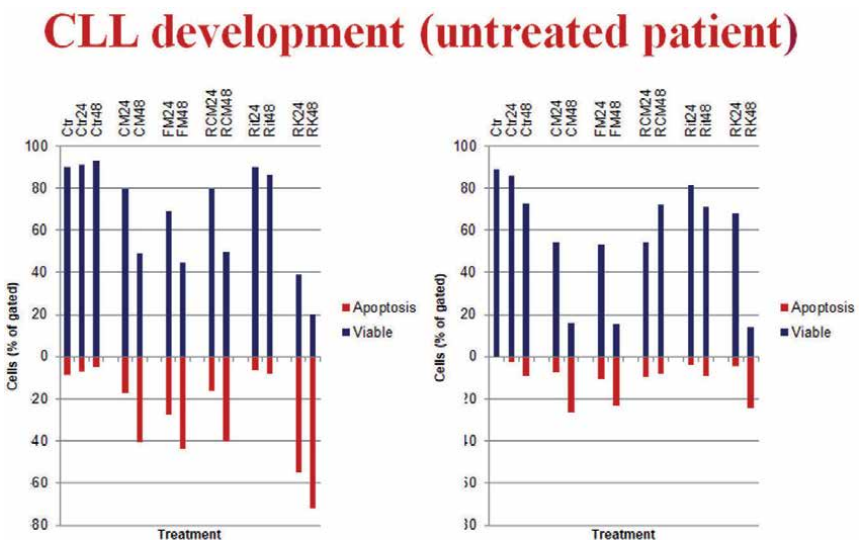


Figure 3.
Differences in drug sensitivity during disease development for the same CLL patient (results from year 2011 and 2013).

The background of chemoresistance includes one or more of the following mechanisms: induction of DNA repair, silencing of gene expression, some alterations in metabolism or in drug target structure, modifications in cell membrane or microenvironment composition, elevated expression of drug efflux pumps, and inhibition of apoptosis [21–24].

Everything is very complex, and on the one hand familiar diversities could affect cell signaling and metabolism; on the other hand also environmental factors could be important for final reaction of our cells and bodies to drugs, even human diet.

At the moment it is better to analyze the total effect of drug activity on leukemic cells before drug administration to CLL patient, preventing resistance to treatment.

Diagnostics for CLL are still based on the analysis of expression of characteristic clusters of differentiation (CD) as well as the presence of cytogenetic alterations (chromosomal aberrations) [25–27].

The currently used drugs in hematological clinics for CLL treatment are usually directed toward inhibition of Bcl-2 gene expression (venotoclax), inhibition of pathways related with signal transduction, for example, involved in inhibition of competitive binding of ATP to Bruton kinase (ibrutinib), or inhibition of PI3K signaling (idelalisib) [27–31]. Both kinases are involved in B-cell receptor signaling.

Venotoclax (VEN, ABT-199) is a selective inhibitor of antiapoptotic protein Bcl-2 expression. It is a BH3-mimetic molecule targeting BCL-2. VEN binds to BCL-2 and could activate BIM and induce apoptosis signaling. VEN demonstrates an activity in patients with poor prognostic, 17p-deleted chronic lymphocytic leukemia (CLL) [28]. VEN shows clinical activity on many hematological malignancies, lymphomas, acute myeloid leukemia, and early T-cell precursor ALL.

Ibrutinib directly works as an inhibitor of Bruton tyrosine kinase (BTK). Interestingly, there is crosstalk between Bruton tyrosine kinase signaling and bioenergetic stress responses. In primary chronic lymphocytic leukemia cells, a pharmacological interference between mitochondrial ATP synthesis and glucose metabolism could affect BTK activity. Moreover, ibrutinib could induce bioenergetic stress responses [32] that might affect for its resistance. Therefore ibrutinib activity could be regulated by glucose level, and patients with hyperglycemia might

be resistant to ibrutinib treatment in TP53 deficient chronic lymphocytic leukemia (CLL) lymphocytes [28].

Idelalisib is an inhibitor of the delta isoform of the phosphatidylinositol 3-kinase (PI3K). Drug could be active on cell proliferation, survival, or even induction of apoptosis. Moreover the strong heterogeneity of CLL feature and involvement in PI3K signaling could be a reason of differences between patients in clinical development of CLL and diversities in response to therapy. Idelalisib could induce several possible side effects, including hepatotoxicity, diarrhea, colitis, pneumonitis, or intestinal perforation. A special importance for personalize therapy is a chance to avoid resistance to treatment and reduce the development of secondary cancer (melanoma, head and neck, prostate, breast, or lung) [16].

Drugs usually for some patients lead to disease remission that could last for many years. It is also possible that for the group of patients, drugs could be ineffective and cause resistance to therapy [31]. For CLL, drugs based on higher generations of monoclonal antibodies could usually cause cytotoxicity in B-cells and fast reductions in the number of B lymphocytes from peripheral blood, decreasing patient's immunological strength. To increase drug activity, the combined therapy based on few drugs could in theory increase patients' response to therapy. Based on our experience with leukemic cell incubations with anticancer drug(s), sometimes it does work this way; for other patients, for example, the addition of second drug will not change cell response to anticancer agents (**Figure 1**). It could prove that sometimes one drug will be enough instead of combined therapy that will not be more effective, will cause patient's weakness, and will not improve the final effect of treatment. For the group of patients, who do not react at standard way to anticancer drugs or are resistant to treatment, choosing the optimal way of treatment for patient seems to be very important for his curing.

There are several molecules, for example, generated from encoded sequences of genes (pre-mRNA), characterized as miRNA, that could affect gene expression, mainly by their silencing, and could decrease or block protein expression. Because of a high complexity of metabolic reactions and several metabolic pathways included inside the cells that could be activated by drugs, hormonal regulation of human body metabolism and several different factors, previously taken medicaments or even diet, could affect patient's response to anticancer treatment [19, 21, 23, 32, 33]. Several potential targets for anticancer agents met in the human body are a reason of drug's side effect. Moreover, some familiar or personal diversities could also change patient reactivity to drugs. Drug reactivity could also be affected by environmental factors or even patient's diet.

At the moment it is better to analyze the total effect of drug activity on leukemic cells before drug administration to patient to avoid potential resistance to treatment.

2. Conclusion

The best way to prevent diseases is a prophylactic. The well-balanced healthy diet, physical exercises, supplementation of probiotic bacteria, or screening tests showing our health condition are the best way to prevent diseases and increase life expectancy. We have to start thinking about our health since childhood, and several wrong decisions could shorten our life. Because of a high complexity in cells and human body function and not enough knowledge related to disease etiology, currently it is better to analyze in vitro the apoptosis induction potential of leukemic cells incubated for 48 h with anticancer drugs before drug administration to CLL patient to exclude resistance to treatment. The resistance to treatment is usually confirmed in CLL cells incubated for 48 h with anticancer drug(s).

While, after 24 h of cell incubation in vitro with anticancer drug(s), the necessity of dose modification usually for lower values in the case of CLL patients' more sensitive cells to drug(s) was noticed. It must be underlined that sometimes the addition of drug (rituximab) as presented in **Figure 1** does not change anything in CLL cell reactivity to drugs (CM and Rit20CM).

Acknowledgements

The author would like to thank Emeritus Professor Zofia M. Kilianska and Professor Henryk Piekarski from University of Lodz and Professor Tadeusz Robak (Department of Hematology, Medical University of Lodz), as well as Professor J. Wesierska-Gadek from Medical University of Vienna (Austria) for help and valuable advices in studies.

Author details

Małgorzata Rogalińska
Faculty of Biology and Environmental Protection, Department of Cytobiochemistry, University of Lodz, Lodz, Poland

*Address all correspondence to: malgorzata.rogalinska@biol.uni.lodz.pl; gosiar@wp.pl

References

[1] Reid G. The development of probiotics for women's health. Canadian Journal of Microbiology. 2017;**63**(4):269-277

[2] Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. Journal of Translational Medicine. 2017;**15**(1):73

[3] Shafei A, Ramzy MM, Hegazy AI, Husseny AK, El-Hadary UG, Taha MM, et al. The molecular mechanisms of action of the endocrine disrupting chemical bisphenol A in the development of cancer. Gene. 2018;**647**:235-243

[4] Liu J. Irisin as an exercise-stimulated hormone binding crosstalk between organs. European Review for Medical and Pharmacological Sciences. 2015;**19**(2):316-321

[5] Ha MS, Son WM. Combined exercise is a modality for improving insulin resistance and aging-related hormone biomarkers in elderly Korean women. Experimental Gerontology. 2018;**114**:13-18

[6] Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. The New England Journal of Medicine;**373**(12):1136-1152

[7] Klein U, Dalla-Favera R. Germinal centres: Role in B-cell physiology and malignancy. Nature Reviews. Immunology. 2008;**8**(1):22-33

[8] Hallek M. Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. American Journal of Hematology. 2020;**94**:1266-1287

[9] Robak T, Stilgenbauer S, Tedeschi A. Front-line treatment of CLL in the era of novel agents. Cancer Treatment Reviews. 2017;**53**:70-78

[10] Rogalińska M, Góralski P, Błoński JZ, Robak P, Barciszewski J, Koceva-Chyła A, et al. Personalized therapy tests for the monitoring of chronic lymphocytic leukemia development. Oncology Letters. 2017;**13**(4):2079-2084

[11] Rogalińska M, Błoński JZ, Góralski P, Wawrzyniak E, Hartman M, Rogalska A, et al. Relationship between in vitro drug sensitivity and clinical response of patients to treatment in chronic lymphocytic leukemia. International Journal of Oncology. 2015;**46**:1259-1267

[12] Rogalińska M, Kiliańska ZM. Personalized therapy versus targeted therapy, differences in the meaning. Global Journal for Research Analysis. 2015;**4**(1):5-8

[13] Rogalińska M, Franiak-Pietryga I, et al. Toward personalized therapy for chronic lymphocytic leukemia: DSC and cDNA microarray assessment of two cases. Cancer Biology & Therapy. 2014;**14**(1):1-7

[14] Góralski P, Rogalińska M, Błoński JZ, Pytel E, Robak T, Kiliańska ZM, et al. The differences in thermal profiles between normal and leukemic cells exposed to anticancer drug evaluated by differential scanning calorimetry. Journal of Thermal Analysis and Calorimetry. 2014;**118**:1339-1344

[15] Kubczak M, Szustka A, Błoński JZ, Gucký T, Misiewicz M, Krystof V, et al. Dose and drug changes in chronic lymphocytic leukemia cell response in vitro: A comparison of standard therapy regimens with two novel cyclin-dependent kinase inhibitors. Molecular Medicine Reports. 2019;**19**(5):3593-3603

[16] Wu M, Akinleye A, Zhu X. Novel agents for chronic lymphocytic leukemia. Journal of Hematology & Oncology. 2013;**6**:36

[17] Tari K, Shamsi Z, Reza Ghafari H, Atashi A, Shahjahani M, Abroun S. The role of the genetic abnormalities, epigenetic and microRNA in the prognosis of chronic lymphocytic leukemia. Experimental Oncology. 2018;**40**:261-267

[18] Martín-Subero JI, López-Otín C, Campo E. Genetic and epigenetic basis of chronic lymphocytic leukemia. Current Opinion in Hematology. 2013;**20**(4):362-368

[19] Bayraktar R, Van Roosbroeck K, Calin GA. Cell-to-cell communication: MicroRNAs as hormones. Molecular Oncology. 2017;**11**(12):1673-1686

[20] Shah MY, Ferrajoli A, Sood AK, Lopez-Berestein G, Calin GA. MicroRNA therapeutics in cancer—An emerging concept. eBioMedicine. 2016;**12**:34-42

[21] Qu Y, Dou B, Tan H, Feng Y, Wang N, Wang D. Tumor microenvironment-driven non-cell autonomous resistance to antineoplastic treatment. Molecular Cancer. 2019;**18**:69

[22] Si W, Shen J, Zheng H, Fan W. The role and mechanisms of action of microRNAs un cancer drug resistance. Clinical Epigenetics. 2019;**11**:25

[23] Galicia-Vázquez G, Aloyz R. Metabolic rewiring beyond Warburg in chronic lymphocytic leukemia: How much do we actually know? Critical Reviews in Oncology/Hematology. 2019;**134**:65-70

[24] Filip AA, Grenda A, Popek S, Koczkodaj D, Michalak-Wojnowska M, Budzyński M, et al. Expression of circulating miRNAs associated with lymphocyte differentiation and activation in CLL-another piece in the puzzle. Annals of Hematology. 2017;**96**(1):33-50

[25] Rodrigues CA, Goncalves MV, Ikoma MRV, et al. Diagnosis and treatment of chronic lymphocytic leukemia: Recommendations from the Brazilian Group of chronic lymphocytic leukemia. Revista Brasileira de Hematologia e Hemoterapia. 2016;**38**(4):346-357

[26] Wawrzyniak E, Haus O. In: Robak T, Warzocha K, editors. Znaczenie badań cytogenetycznych w hematologii. Wyd. Via Medica: Hematologia; 2016

[27] Montserrat E, Bauman T, Delgado J. Present and future of personalized medicine in CLL. Best Practice & Research. Clinical Haematology. 2016;**29**(1):100-110

[28] Seyfried F, Demir S, Hörl RL, et al. Prediction of venetoclax activity in precursor B-ALL by functional assessment of apoptosis signaling. Cell Death & Disease. 2019;**e 10**:571

[29] Choudhary G, Al-Harbi S, Mazumder S, et al. MCL-1 and BCL-xL-dependent resistance to the BCL-2 inhibitor ABT-199 can be overcome by preventing PI3K/AKT/mTOR activation in lymphoid malignancies. Cell Death & Disease. 2015;**6**:e1593

[30] Burger JA, Tedeschi A, Barr PM. Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. The New England Journal of Medicine. 2015;**373**:2425-2437

[31] Furman RR, Sharman JP, Coutre SE, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. The New England Journal of Medicine. 2014;**370**:997-1007

[32] Sharif-Askari B, Doyon D, Paliouras M, et al. Bruton's tyrosine kinase is at the crossroads of metabolic adaptation in primary malignant

human lymphocytes. Scientific Reports. 2019;**9**:11069

[33] Rogalińska M. The role of mitochondria in cancer induction, progression and changes in metabolism. Mini Reviews in Medicinal Chemistry. 2016;**16**(7):524-530

Chapter 4

Genetic Risk Factors and Lysosomal Function in Parkinson Disease

Marcella Vieira Barroso Montenegro, Carlos Eduardo de Melo Amaral and Luiz Carlos Santana da Silva

Abstract

Parkinson disease is a complex disease that has multiple genetic and environmental factors. To achieve the early diagnosis and to be able to modify the disease progression, efforts are being made to identify individuals at risk. About 20 year ago, an evidence of major prevalence of Parkinsonism in patients with Gaucher Disease reported by studies worldwide led to the putative involvement of the *GBA* gene. Nowadays, the link from a rare disease with a common disease is well known and it is confirmed that mutations in the *GBA* gene are the most important genetic risk factor. Apart from rare mutations, genetic association studied appointed common variants in genes well associated with familial cases as *LRRK2* and *SNCA* may also contribute to the increased risk for sporadic cases. Other common variants in the *MAPT* gene were also reported. At least, genetic studies have been observed an excessive burden of relevant variants in genes with lysosomal function. Thus, a synergistic action of variants in genes that codifies proteins involved with the lysosome may be a mean of modulating the risk. In this chapter, we review the most robust genetic risk factor and the relevance of lysosomal function for Parkinson disease.

Keywords: Parkinson, GBA, risk factor, lysosome, GWAS

1. Introduction

Parkinson's disease (PD) is the second most common progressive neurodegenerative disease in humans and it is characterized by motor symptoms as muscular rigidity, resting tremor, bradykinesia, and postural instability and also by non-motor symptoms (hyposmia, constipation, depression, dementia, and postural hypotension, among others). These symptoms result primarily from the progressive loss of the dopaminergic neurons from the pars compacta of the mesencephalic substantia nigra and subsequent depletion of the dopamine neurotransmitter in the striatum, a central component of the basal ganglia that is responsible for the instigation and coordination of movements (**Figure 1**). The definitive diagnosis of PD is difficult being only confirmed with the presence of Lewy bodies, proteinaceous intracytoplasmic inclusion, in the reminiscent neurons of substantia nigra pars compacta and other regions in the brain postmortem analysis [1].

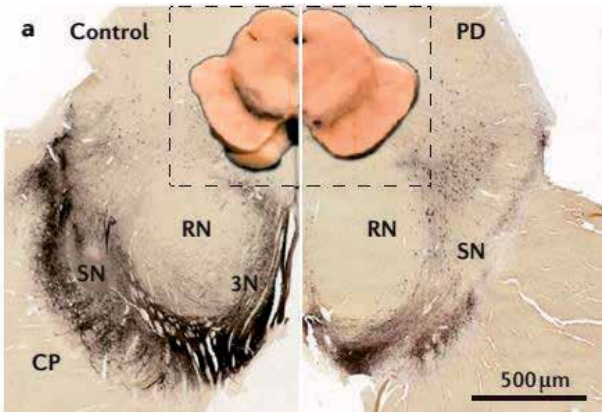


Figure 1.
Depigmentation of the substantia nigra (SN) (right panel) compared with control (left panel). Adapted from Ref. [1].

The etiology is not well understand, but PD is considered a complex disease, which counts with multiple genetic and environment factors. The most common is the sporadic PD for which the onset generally is late, after 60 years old. There is a rare form, the familial PD or monogenic PD (~10% of the cases), for which the disease is caused by mutations in a single gene and may present not only a late onset but also an earlier onset (below 45 years old) in some cases. Although less frequent, the study of monogenic forms of PD and their associated genes helps to understand the molecular basis of disease pathogenesis [1, 2].

Segregation studies of mutations in *SNCA* gene in large families with PD cases led to the discovery of the main protein involved in the disease pathogenesis, the α-synuclein. Shortly afterwards, postmortem studies in patients' brains revealed that this protein is the major component of Lewy bodies, in both sporadic and familial PD patients, reinforcing its important role in the development of PD [2–4]. Be it for genetic, environmental or both factors, the fact is that patients' brains do not have the soluble monomeric form of α-synuclein, which is easily degraded by lysosomal function, but the insoluble oligomeric forms [5, 6].

Nowadays, efforts have been directed to identify the individuals at risk of manifesting PD through clinical, genetic and biochemical markers in order to diagnose early and perhaps be possible to modify the progression of disease. For this purpose, genetic variations with the potential to alter the risk for PD have been widely researched. Both disease-causing variants and risk variants in genes associated with PD vary in frequency depending on ethnic background. Certain genetic variants may be a risk factor in an Asian population, but may not be statically significant in a European population, for example. The most robust and consistently replicated results are appointed to variants in the genes *LRRK2, MAPT, SNCA* and *GBA,* the last being the major genetic risk factor highlighting the importance of lysosomal pathway in the pathogenesis of PD [1–8].

2. Genetic risk factors

More than 20 years ago, Parkinson disease was understood as a disease caused by environmental factors only. It was from genetic analyzes in cases of familial PD that it was discovered that the genetic factor is also important and may even cause certain disease forms. Thus, we have monogenic PD, which can present a pattern

of inheritance defied as autosomal dominant or autosomal recessive. In 1997, the discovery of mutations in the *SNCA* gene as the cause of PD in certain families also helped scientists to better understand the etiopathology of this complex disease through the association between the duplication and triplication of the *SNCA* gene with altered α-synuclein protein expression and the disease progression. Since then, diverse other genes were associated with mutations that cause PD, among them the genes *LRRK2* and *VPS35,* that along with *SNCA* are altered in autosomal dominant cases, and *PINK1*, *PARKIN*, *DJ-1* that are altered in autosomal recessive forms [3, 8].

However, the genetics of PD are not simply composed of variants that cause the disease. More recently, the focus has been on genetic variants that do not lead to PD alone but increase the risk of developing the disease. Among the genetic risk factors associated with sporadic PD, rare high-impact variants and common low-impact variants have been identified by candidate gene studies and genome-wide association studies (GWAS). The complexity of the PD genetic increases even more due to several risk variants for PD that are heterogeneous and dependent on the genetic background of each population. Genetic variants can be associated with PD in some populations, but not in others [8, 9].

Two largest GWAS studies in 2014 and 2017 identified in total 28 independent PD-associated risk loci, mainly in *SNCA*, *LRRK2*, *MAPT* and *GBA*. Genetic risk factors are present in several genes involved in metabolic pathways that may be directly related to α-synuclein metabolism or involved in processes that affect cellular homeostasis [10, 11].

2.1 SNCA

It is worthy to emphasize the fact that genes that carry rare casual mutations of monogenic PD observed in previous family studies are not excluded of the possibility to also carry common variants that confer risk for developing PD. This is the case of *LRRK2* and *SNCA*.

So far, five point mutations (A53T, A30P and E46K) and two copy number variation (duplication and triplication) in the gene *SNCA* are well known to cause autosomal dominant PD indistinguishable of sporadic PD, or an early-onset PD if the triplication of the gene is present [12]. Recently, single nucleotide polymorphisms (SNP) in *SCNA* were reported in non-coding regions, suggesting that those variants play a role in the regulation of the genetic expression through modifications post-transcriptional as interacting with microRNA or altering alternative splicing mechanism [7].

In 1999, the association between REP1, a complex polymorphic microsatellite repeat in the promoter region, and PD was pointed out by [13]. Seven years later, Maraganore [14] confirmed this association with a larger meta-analysis study using more than 5000 samples from 11 sites. Further, functional analysis studies provided evidence that the length of alleles affects the protein expression: the 261 bp-long risk allele is associated with an upregulation of α-synuclein expression mimicking *SNCA* locus multiplication, whereas the 259 bp-long protective variant shows reduced gene expression [15].

In 2009, Simón-Sánchez et al. [16] used GWAS in a great sample and identified additional signals of association with PD from intro 4 to after the 3′ UTR. One year later, Mata et al. [17] showed possible association between rs356219 in the 3′ UTR region and α-synuclein plasma levels. To definitely ascertain which variants in this region alter the risk for PD, more studies are necessary in large and genetically diverse samples.

2.2 LRRK2

In the region of chromosome 12 is localized the *LRRK2* gene where several genetic variants have been found; however, segregation in families with monogenic PD and case-control studies demonstrated that only seven point variants (R1441G, R1441C, R1441H, Y1699C, G2019S, I2020T and N1437H) have enough evidence to be defined as cause of PD. This gene encodes a protein of the same name composed of domains with kinase activity, GTPase and several domains of interaction with other proteins, suggesting that its function changes depending on which proteins form complexes, the type of cells and the stage of development. Mutations in *LRRK2* gene are the most common cause of PD familial cases. The mutation G2019S is the most prevalent worldwide, and it is present in 4% of familial PD and is associated with an indistinguishable phenotype from the clinical manifestations of sporadic PD [12–18].

Besides the prevalence in rare monogenic PD, this mutation can also confer risk in the sporadic PD, being found in 1% of the cases. G2019S has a penetrance variable, and its carrier's risk to develop PD depends on age and ethnic background. The age-related risk has been estimated to be 28% at age 59, 51% at 69, and 74% at 79 years. The frequency is higher in North African, Middle Eastern and Ashkenazi Jewish PD patients [18, 19].

G2019S is frequent in most populations worldwide, but it is very rare in the Asian population where it accounts for less than 1% of *LRRK2* mutations. In contrast, most common genetic variants SNPs G2385R and R1628P are more frequent in Asian populations than in Caucasian populations. Those SNPs are associated with an increased risk of 2.2 fold and 1.84 fold to develop PD, respectively [20, 21]. Lately, regions close to *LRRK2* have been appointed by GWAS as increasing by 1.2 fold the risk for PD. This fact alerts that the regulation of the gene expression is important to develop this disease [8].

Those works reinforce the idea that genes can carry both rare disease-causing variants and common variants that increase the risk for PD, as seen in *SNCA* and *LRRK2*. Additionally, common variants present in those genes enhance the importance of their proteins' role in the disease and implicate that there is a common neurodegenerative process between sporadic and familial PD.

2.3 MAPT

The gene *MAPT* is frequently associated with other neurodegenerative diseases as Alzheimer disease and frontotemporal dementia (FDT). *MAPT* encodes for the microtubule-associated protein tau, whose role is to regulate microtubule dynamics and assemble microtubules into parallel arrays within axons, essential for normal axonal transport in neurons. Polymorphisms in this gene have been found to be an indisputable risk factor of the synucleinopathy. The H1 and H2 haplotypes represent two distinct clades of subhaplotypes ensued from an inversion of ~900 kb on chromosome 17q21, spanning the entire *MAPT* coding region, and are tagged, among others, by genotypes at two SNPs: rs9468 and rs1800547 [4–8].

The H1 haplotype and its subhaplotype H1c have been significantly associated with an increased risk for a number of neurodegenerative diseases. Several studies proposed the most common H1 haplotype as susceptibility factor for PD with an odds ratio of 1.5. Recent studies that investigated the association between H1 and specific PD clinical manifestation also observed the higher prevalence of H1 in patients with cognitive defects, as dementia and H1 homozygous PD patients showed an increased risk to manifest non-tremor dominant subtype, which is a worse clinical prognosis [22, 23].

The underlying biological mechanisms that link the *MAPT* locus (and tau protein) to neurodegeneration are not yet adequately characterized. Through the functional characterization of variants in the *MAPT* gene, some theories of the tau protein effect in PD involve increased tau expression; altered gene splicing promoting aggregation; and altered 4/3 repeated transcript ratio. The emerging concept of H1 pathogenicity points to the role of each tau isoforms expressed rather than the overall number of transcripts.

According to this model, the H1 haplotype is associated with an underexpression of a protective isoform and an overexpression of the detrimental variant, which lead to a subtle neuronal dysfunction that accumulates over the years and induces or accelerates cellular degeneration [22–24]. However this locus harbors many genes and the extended linkage disequilibrium means that the tau protein may be not the cause of neurodegeneration and its DNA sequence is just close to the casual locus. Thus, while *MAPT* is a candidate, we cannot be certain that this is the true biological mediator of risk [7–16].

Diverse studies identified *MAPT* locus variants as a risk factor for PD, but it may not be true to any population. In the Caucasian population, there was this association, while it was absent in the Japanese population. This observation has potential implications for the analysis of complex traits across populations such as genetic heterogeneity, particularly at minor risk loci, highlighting the power of comparing GWAS across different populations [9–16].

Despite having a modest effect (less than 30% of the change in risk), these common variants can have a considerable impact when combined. Results from a 2015 study revealed that patients at an early age of onset of symptoms had a higher polygenic combination of risk variants than patients with a late onset. This demonstrates the possible effect of the synergistic value of the changes caused by these variants to modulate the PD clinic, such as the age of onset of symptoms [12].

3. GBA: the principal genetic risk factor for Parkinson disease

The *GBA* gene was initially described in association with a rare lysosomal storage disease (LSD) called Gaucher disease (GD). When mutated in homozygosis, depending on the mutation present, the resulting enzyme is malformed or even no enzyme is synthesized leading to enzyme glucocerebrosidase (GCase) partial or total deficiency and glucosylceramide (GlcCer) accumulation. The symptoms are multisystemic, with the brain, spleen, liver and bone marrow being the main organs affected. The presence and intensity of those symptoms differ between the three types of GD (GD1, GD2 and GD3). The heterozygous individuals do not present any clinical manifestation; however, in the last years, this perspective has changed [25, 26].

Further a number of studies have recorded the occurrence of parkinsonian manifestations in patients with GD and their relatives [27, 28]. In ref. [29] was showed that *GBA* mutations homozygous individuals have 21.4 fold increased risk to develop PD with probability of 9–12% to manifest motor symptoms before 80 years old. Despite being low, this risk is considerably higher than in the same age group in general population, 3%. The *GBA* and PD association was confirmed in the Jewish Ashkenazi, which showed a prevalence of *GBA* mutations in heterozygosis and homozygosis individuals in the PD population that by far outweighs the reported prevalence of mutations in other susceptibility genes for PD, as *Parkin* and *SNCA* [30].

Researchers worldwide have attempted to validate the same association in populations from many different genetic backgrounds [31–41]. In 2009, an international

and multicenter study with a great sample of approximately 5000 PD patients and equal number of controls provided the definitive proof found for this association with an odds ratio greater than five (OR 5.3) and showed that mutations N370S and L444P are the most frequent in this gene. In other words, *GBA* mutation genes were recognized as the major genetic risk factor for PD until now [42].

In addition to alter the risk to manifest the disease, the presence of *GBA* mutations has also the potential to modify PD phenotype such as age of onset. The modulatory effects best described in the literature investigated the association with age of onset and declined cognitive. The association to other symptoms and patients survival rate has been not yet approached in more detail.

Whether in heterozygosis or in homozygosis individuals (GD patients), the age of onset of symptoms apparently occurs earlier than in PD patients without mutations, usually between the fourth and sixth decade of life [28–43]. In relation to symptoms is noticeable a cognitive decline earlier in PD patients with *GBA* mutations (PD-GBA) [26–45]. Dementia is one of the clinical manifestations that most affects the patient's quality of life and it is more frequent in GBA mutation carriers than in non-carriers. Longitudinal studies have shown that PD-GBA patients have a risk three times higher than patients without GBA mutations to present dementia [26, 44, 46]. Neuroimaging exams support this association by showing more expansive synucleinopathy in the neocortical and subcortical areas of PD-GBA patients, increasing the risk to dementia, psychosis and postural hypotension [46].

A few studies have evaluated the survival rate, if there is a greater risk of death in PD-GBA patients than in those without any mutations. A 2014 study found lower survival for the carrier group, but had a weak statistical value [47]. However, in 2016, a study with the largest sample number ever described replicated the same result with powerful statistical significance reinforcing this conclusion. It was defined in this study that there is a twice greater risk of mortality among PD-GBA patients. One explanation may be the increased presence of non-responsive levodopa motor impairments such as dysphagia and non-motor impairments such as orthostatic hypotension in the group of PD-GBA. There was no difference in disease duration compared to non-carriers, but patients were significantly younger at the time of death [46].

3.1 Genotype-phenotype correlations

Researchers have also observed different *GBA* mutations having a different impact in the clinical manifestations. Mutations generally associated with defined neuropathic forms of GD (GD type 2 and 3), such as L444P, are classified as severe mutations, while others associated with GD type 1, such as N370S, are classified as mild mutations [26–48].

A meta-analysis study included populations from North, Central and South America, Western and Eastern Europe, North Africa, Asia and Ashkenazi Jews, and its results showed a clear and significant differentiated effect comparing mild and severe mutations on the risk of developing PD and the age of onset of symptoms not only in Ashkenazi populations, but also worldwide. Severe mutations such as L444P confer a three to four times increased risk for its carriers to develop PD and are associated with the onset of symptoms 5 years earlier than mild mutations. The average age found for severe mutations was 53.1 (±11.2), whereas the average age for mild mutations was 58.1 (±10.6) [26].

The type of mutation was also relevant in modulating the cognitive impairments of PD patients. Ref. [46] observed that severe mutations conferred a higher risk of dementia for its carriers. The risk was three times greater compared to patients with mild mutations and five times greater when compared to the risk for PD without

GBA mutations. Another longitudinal study with a similar sample number corroborated with the increased risk of dementia for patients with severe mutations versus non-carriers [44]. With regard to other symptoms, motor impairments appear to be similar between patients with mild mutations and non-carriers, while those with severe mutations appear to be more aggravated and seem to have a higher frequency of non-motor symptoms such as psychosis, apathy and postural hypotension [46].

Patients' survival does not seem to differ when comparing the types of mutations with each other. However, when compared separately with non-carriers, mild mutations do not differ statistically from non-carriers, while mortality was shown to be greater for carriers of severe mutations than in non-carriers [46].

Due to the discovery and increasing number of proofs supporting the great influence of *GBA* mutations in PD, some authors consider the possibility of reclassifying them from risk factors for agents causing autosomal dominant PD [43, 49].

4. GBA-associated PD in different populations

Given the multifactorial etiology of PD, the different environment and ethnicity of a population may impact in the different results seen among the papers that investigated the frequency of *GBA* mutations in PD patients. Other possible causes of this variation can be the use of different techniques and methodologies.

The highest frequencies of mutations of the *GBA* gene have been found in PD patients of Ashkenazi Jewish ancestry, with rates of 13.7–31.3% in comparison with 4.5–6.2% in control groups [26–28]. The frequencies recorded in PD patients in non-Jewish populations representing other populations, such as Italians, Caucasian Americans, Greeks, Brazilians, British and Taiwanese, are invariably much lower—3.5% to 12.0%—while controls from the same populations range from 0% to 5.3% [31–41]. Previously, in North Africa, a study found no association between PD and mutations of the *GBA* gene; however, a more recent African study data suggested a risk association between mutations in the *GBA* gene and PD [37]. The lowest rate recorded to date was 2.3% in Norwegian PD patients, compared with 1.7% in the control [49].

The genetic background can also impact the frequency even within the same country. In Brazil, four studies evaluated the association between GBA and PD, with variances in frequencies depending on the region (**Table 1**). The North region had twice as many cases of PD patients with GBA mutations (7.4%) compared to the frequencies of the South and Southeast regions (3.5%) with a similar sample number. The fact may be explained by the different genetic composition of the North region, which, despite also has a European origin, has a higher Amerindian ancestry than the Southern Brazil, which is almost exclusively from European ancestry [31–50].

Greek and Italian studies have found significant differences comparing PD patients and controls from urban and rural areas, and from the North and South regions, respectively. In the Greek study, the frequency of *GBA* mutations between PD patients and controls was statically significant. However, when the cohorts were analyzed separately, there was a difference of frequencies. The difference between PD patients and controls was statistically significant only in the case of the patients of cohort A that is originated from Thessaly, a mainly rural area ($p = 0.021$, OR 4.2, 95% CI = 1.14–15.54) and not in the case of cohort B patients, the majority of which were residents and/or originated from the greater area of Athens, an urban environment ($p = 0.113$, OR 2.5, 95% CI = 0.77–8.42) [34].

In the Italian study, there was a lower frequency of mutations in PD patients (11/395, 2.8%) and in controls (1/483, 0.2%) from the Southern region, and the most common mutation was p.L444P. Conversely, in the Northern region, the most

Studies	Population studied	PD inclusion criteria	Method	Mutation analyzed	Patients mutation frequency	Control mutation frequency	Age of onset	*GBA* mutated PD and familiar history (FH)
Spitz et al. (2007)	65 PD patients and 267 control subjects from Southeastern Brazil	Early onset (<55 years).	PCR-RFLP, restriction endonucleases and electrophoresis	N370S and L444P	2/65 **(3%)**; L444P 2/2 (100%); N370S 0/2 (0%)	0/267	Patient 1 at 46 yr old and patient 2 at 42 yr old	The two patients had FH, no statistical test was used.
Socal et al. (2008)	62 PD patients from Southern Brazil	All patients diagnosed were included.	PCR-RFLP, restriction endonucleases	N370S, L444P and IVS2þ1	2/62 **(3.5%)**; L444P 1/2 (50%); N370S 1/2 (50%)	Not informed	Patients with mutation 37 ± 4 yr Patients without mutation 41.4 ± 10.8 yr	Not informed.
De Carvalho et al. (2012)	347 PD patients and 341 control subjects from Southeastern, Midwestern and Northern Brazil	All patients diagnosed were included.	Direct sequencing	N370S and L444P	13/347 **(3.7%)**; L444P 8/13 (62%); N370S 5/13 (38%)	0/341	Patients with mutation 49.9 ± 11.3 yr Patients without mutation 52.5 ± 13.3 yr	Those with FH and those without FH did not present statistical significance.
Amaral et al. (2018)	81 PD patients and 81 control subjects from Northern Brazil	All patients diagnosed were included.	Amplification of the exon 8–exon 11, PCR-RFLP for N370S and L444P, restriction endonucleases and direct sequencing of N370S and L444P	N370S and L444P	6/81 **(7.4%)**; L444P 3/6 (50%); N370S 3/6 (50%)	0/81	Patients with mutation 49.6 ± 17.4 yr Patients without 55.1 ± 11.6 yr	From the 6 patients, 2 had FH. No statistical test was used.

Bold valor are percentage.

Table 1.
GBA mutation among PD patients in different Brazilian regions.

frequent genetic defect found was p.N370S and the frequency of mutations in PD was 4.5% and 0.63% in controls. Therefore, the difference may be due to a particular frequency of *GBA* mutations in regions of Italy or to the sample size [36].

Although the frequency of *GBA* mutations in populations of PD patients has been well characterized worldwide, few data are available on the inverse relationship, that is, the risk of healthy heterozygous for *GBA* mutations in developing PD, since it is not known for sure the degree of influence that this genetic alteration has on the onset of the disease.

Anheim M et al. [43] published in 2012 an estimate of the penetration of PD in healthy heterozygous people for *GBA* mutations and reached a value of 7.6%, 13.7%, 21.4% and 29.7% for 50, 60, 70 and 80 years, respectively, based on a dominance model. However, in the same year, [51] found a lower value: 5% for 60 years and 15% for 80 years of age. Such difference is perhaps due to additional genetic factors or environmental factors, a fact that emphasizes the possibility of variance of the risk of developing PD according to the genetic background of the population. Families that have *GBA* mutations segregation through generations are a group at risk for developing PD and should be monitored for a possible early diagnosis to have better chances in modifying the disease.

5. Pathogenic mechanisms in PD

α-Synuclein is a key protein in the neuropathogenesis of PD, involved in several pathogenic processes. The physiological function of α-synuclein is not well understood, but studies show that it is normally located in presynaptic terminals where it binds to lipids and plays the role of regulating in more than one step the traffic of synaptic vesicles to be released. As cited above, the insoluble forms (oligomers and fibrils) of this protein accumulate and compound the Lewy bodies found in most PD patients and also contribute to neuronal cell death [5].

The reason behind this accumulation can be due to increased synthesis or decreased degradation (**Figure 2**). Mutations, as the triplications of the *SNCA* gene, can enhance the production of α-synuclein, while the interaction between the oligomers and fibrils formed with other mutant proteins can result in the deficiency of certain metabolic pathways and contribute to slowing down α-synuclein proteolysis. On the other hand, the initial deficiency of certain metabolic pathway caused by mutations in genes, advanced age or environmental factors can also be the trigger to accumulate α-synuclein resulting in the insoluble forms. This last theory has been

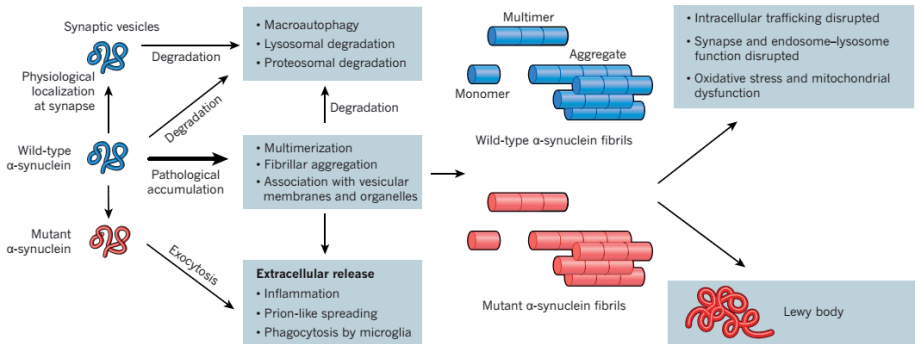


Figure 2.
The proposed physiological and PD-associated pathological functions of α-synuclein in neurons. Adapted from Ref. [6].

reinforced by the confirmed association of diverse genes involved in autophagy, endocytosis and lysosomal pathways as the *GBA* and *LRRK2* gene [1–6].

5.1 Lysosomal function-related genes and PD

The endosome-lysosome traffic processes, autophagy and lysosomal degradation, are essential functions for cell homeostasis, especially for neurons. The differentiated neurons have to maintain their homeostasis during the aging through degradation pathways since they do not divide in the same way as other eukaryotic cells. Moreover, cellular and animal models have also shown that the process of lysosome-autophagy and ubiquitin-proteasome has its activity reduced with natural aging. It may cause the accumulation of proteins whose homeostasis depends on those processes, such as α-synuclein. Indeed, the stimulation of degradation by macroautophagy through drugs proved to decrease intracellular levels of α-synuclein in experimental models [1–6].

Reciprocally, the accumulation of α-synuclein in the substantia nigra in experimental models leads to a reduction in lysosomal enzymes such as GCase, cathepsin B, β-galactosidase and hexosaminidase causing the inhibition of macroautophagy and ubiquitin-proteasome processes as a consequence enzyme transport to the lysosome interruption through dysfunction of vesicles and endosomes. The result is a vicious cycle where α-synuclein degradation mechanisms are inefficient resulting in the protein accumulation and it reinforces the inhibition of degradation activity [5–52].

Both GCase deficiency and the accumulation of its substrate (GlcCer) have been described to be associated with neurodegeneration (**Figure 3**). Feany M et al. [53] suggested that the connection of the α-synuclein to lipidic membranes would protect this protein from inadequate and clumped folding. Mutations of the *GBA* gene would alter the lipid composition of the membrane, which would favor a build-up of α-synuclein in the cytosol and subsequently in the Lewy bodies. Knockdown *GBA* in neuronal cells or in mouse models impairs α-synuclein clearance, whereas increasing glucocerebrosidase activity has the opposite effect, perhaps giving support to the loss-of-function theory in which the reduced or absent lysosomal enzyme is the trigger to α-synuclein accumulation [54].

Even excluding *GBA*, there was evidence for a burden damaging alleles in association with PD. In 2017, Ref. [55] performed a large study to examine the overlap between genes responsible for LSD and PD. More than half of PD cases in their cohort harbors one or more putative damaging variants among the 54 LSD genes. Specially, risk alleles in the genes *SMPD1* (Niemann-Pick type A/B), *GALC* (Krabbe disease), *SLC17A5* (Salla disease), *ASAH1* (Farber lipogranulomatosis) and *CTSD* (neuronal ceroid lipofuscinosis) have been candidate genes well replicated in different studies. *SMPD1* and *ASAH1,* along with *GBA*, participate in ceramide metabolism, and this fact can be evidence of the ceramide-associated process being relevant in a scenery of lysosomal dysfunction in PD [10–55].

The genes appointed as risk factor for PD to date explain only a fraction of PD heritability, suggesting the involvement of additional loci. Besides, the fact that *GBA* is the major genetic risk factor for PD makes other LSD genes attractive candidate risk factors. The results of [55] suggest that many genes that encode lysosomal enzymes besides *GBA* likely contribute to susceptibility for PD in Caucasian population.

Not only the lysosomal function is important, but also the previous steps necessary for the vesicle content to reach this organelle. In [10], a GWAS meta-analysis study found that PD-associated signals were enriched for autophagy and lysosomal function. *SCARB2* encodes a membrane protein (LIMP-2) required for correct targeting of GCase enzyme to the lysosome. Independent large GWAS have replicated

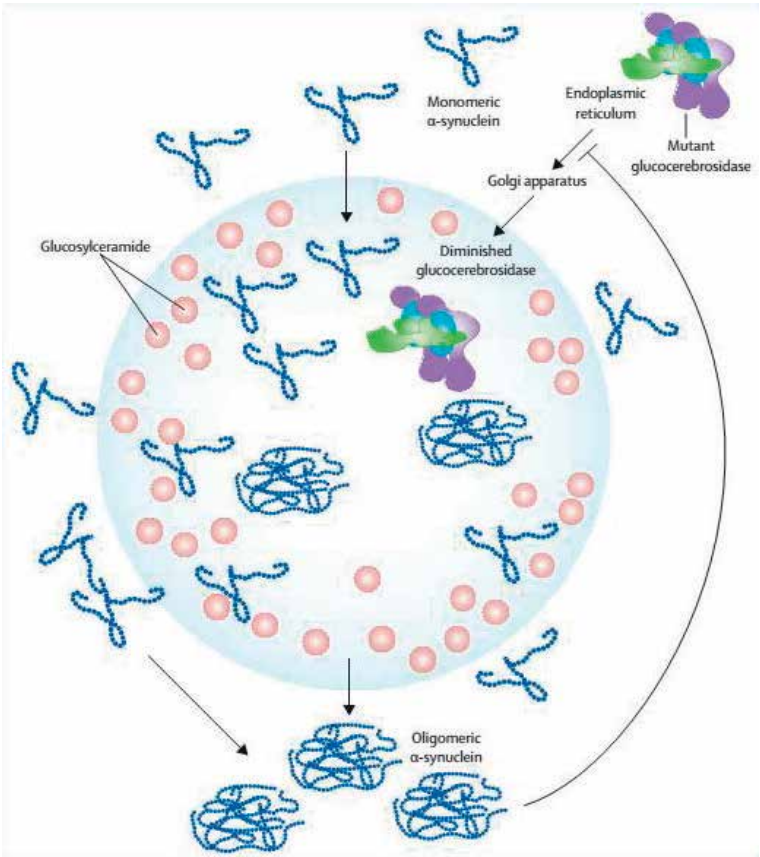


Figure 3.
The vicious cycle between the GCase and α-synuclein. Decreased glucocerebrosidase increases the lysosomal concentrations of glucosylceramide, which increases the formation of soluble α-synuclein oligomers. These oligomers also disrupt transport of newly synthesized glucocerebrosidase between the endoplasmic reticulum and Golgi apparatus, further compounding the problem. Adapted from reference [52].

common risk alleles in this gene. Functional analysis in cellular and animal model has shown that the reduction of LIMP-2 impairs the clearance of α-synuclein [55]. Those data reinforce that both the malfunction and the absence of the GCase, through mutations or impairment in the pathway, can result in α-synuclein accumulation.

The protein LRRK2 is complex and can work together with diverse proteins in different pathways, but for PD, the most relevant seems to be its endosome-to-lysosome trafficking function. Mutations in the kinase domain of the LRRK2 protein, such as the most common G2019S, compromise the traffic of the endosomal content to the lysosome through accentuated phosphorylation resulting in the dysregulation of proteins of the Rab family, responsible to target vesicles to the correct organelle membranes, including the lysosome [6–18].

In support of vesicular trafficking to lysosome impairment in PD, in 2009, two GWAS collaborative studies examining Caucasian and Asian subjects revealed significant risk alleles in *PARK16* locus for PD. This locus is a large linkage disequilibrium block that includes a Rab protein member of a subfamily that is implicated in vesicular transport to lysosomes and to lysosome-like organelles, the Rab-7 L1 (also known as RAB29) [9–16].

Mutations in the *VPS35* gene are one of the causes for autosomal dominant PD. Its protein is also involved in trafficking to lysosomes as a member of the

retromer complex, which has the role to regulate the delivery of the protein content within endosomes to organelles. Some of the proteins carried by this complex are cation-independent mannose-6-phosphate receptors, necessary for the transport of lysosomal enzymes to the lysosome. In the dysfunction of the retromer complex, the receptors are not returned to the Golgi complex, thus impairing the lysosomal function. In addition, mutations in *ATP13A2* are a rare cause of recessive juvenile-onset Parkinsonism and dementia and are associated to lysosomal dysfunction. This gene codifies a lysosomal P-type ATPase [1, 6].

Interestingly, potentiated retromer function might suppress the altered trafficking and toxicity that are associated with mutations in *LRRK2* or the overexpression of α-synuclein85, which suggests a potential therapeutic avenue. This fact emphasizes the possibility that different genes can interact with each other influencing the lysosomal function and as a consequence modifying the PD progression.

These common and rare risk alleles in *ATP13A1, RAB7L1*, *LRRK2* and *VPS35*, which support a model of partial loss-of-function variants in genes regulating lysosomal activity by cellular trafficking, result in an increased vulnerability to α-synuclein mechanisms in PD [55]. Ref. [10], the largest GWAS meta-analysis study, concluded that PD-associated signals were enriched for autophagy and lysosomal function. It replicated the results for *GBA* and *TMEM175* genes, which encode a potassium channel involved in the regulation of lysosome and identified three novel candidate genes, *CTSB* (a lysosomal cysteine protease), *ATP60A1* (an ATPase) and *GALC* (a lysosomal enzyme).

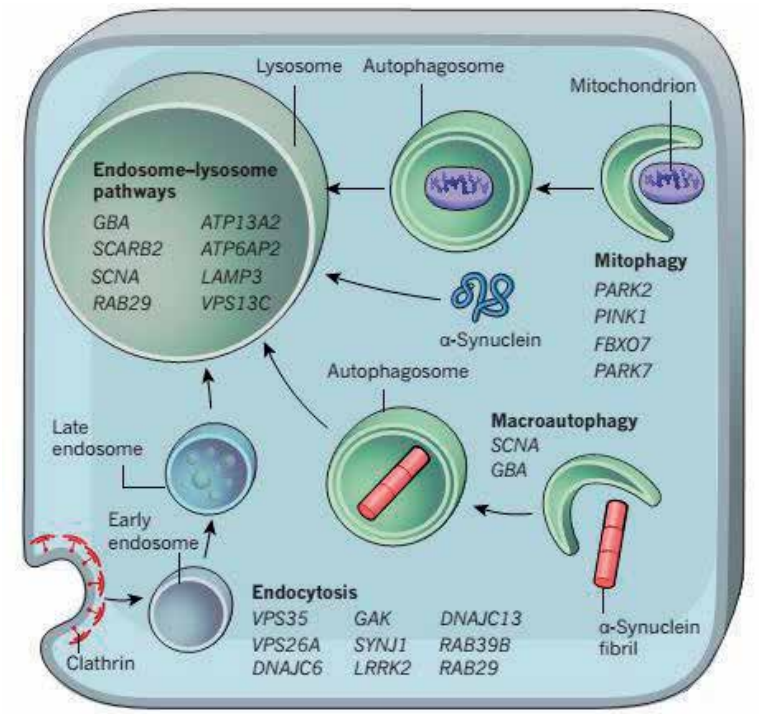


Figure 4.
Some of the PD-related genes associated with trafficking to the lysosome. Genes that encode intracellular trafficking components are associated with common sporadic and familial forms of PD, as well as related syndromes that share some of the clinical features of PD. Most of these genes are known to affect trafficking to the lysosome in the context of late endosome-to-lysosome pathways, clathrin-dependent endocytosis, macroautophagy or mitophagy. Wild-type α-synuclein (blue) can also enter lysosomes through chaperone-mediated autophagy. Adapted from Ref. [6].

Besides *GBA*, loss-of-function alleles are known as frequent PD risk factors, and some of those genes had the functional characterization made by analysis studies that showed knockout mice manifesting tremor phenotype with cerebral and cerebellar atrophy, thus corroborating with lysosome loss-of-function hypothesis to be involved with α-synuclein dysfunction and PD pathogenesis [10].

Therefore, advances in genetic and experimental model for PD have illuminated an important role for defects in intracellular transport pathways to lysosomes (**Figure 4**). The probability of discovering rare PD disease risk alleles at a single locus is low; however, if a set of lysosomal-related genes is investigated in conjunction, the chance of finding significant genetic variations is increased. Also, the candidate genes here appointed need further studies including even larger case–control studies and experiments in PD cellular or animal models.

6. Conclusion

Currently, genetic testing for PD is not a routine procedure, being restricted only to cases with a positive family history, with early onset or with the presence of specific atypical symptoms. In the future with the advance of genetic research, however, there is a possibility to use genetic variants to provide a perspective of the patient's clinical evolution. For this purpose, it is important to replicate risk variants for PD in large and genetically diverse samples due to the different results among populations. Genetic studies need to be a collaboration of the whole world to understand the genetics of a complex disease. In addition, candidate genes here appointed need further experiments in PD cellular or animal models understanding of the underlying pathology and molecular pathogenesis to provide perhaps the basis for the development of new therapies able to target mutated proteins that cause impairment in relevant pathways for PD as endosome trafficking, lysosome function and autophagy.

Acknowledgements

This study was performed with research grants from Instituto Nacional de Genética Médica e Populacional—INAGEMP (CNPq: 573993/2008-4), Fundação de Amparo à Pesquisa do Estado do Pará (FAPESPA) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) from Brazil.

Author details

Marcella Vieira Barroso Montenegro[1*], Carlos Eduardo de Melo Amaral[2] and Luiz Carlos Santana da Silva[1]

1 Laboratory of Inborn Errors of Metabolism, Federal University of Pará (LEIM/UFPA), Belém, Brazil

2 Department of Molecular Biology, Center for Hemotherapy and Hematology of the Pará Foundation (HEMOPA), Belém, Brazil

*Address all correspondence to: marcella_montenegro3@hotmail.com

IntechOpen

References

[1] Poewe W, Seppi K, Tanner C, Halliday G, Brundin P, Volkmann J, et al. Parkinson disease. Nature Reviews Disease Primers. 2017;**3**(1):1-21. DOI: 10.1038/nrdp.2017.13

[2] deSouza, Ruth-Mary, Schapira AH. Etiology and pathogenesis of Neurodegeneration. In: Schapira A, Wszolek Z, Dawson TM, Wood N, editors. 1st ed. 2017. pp. 46-52

[3] Polymeropoulos M. Mutation in the α-synuclein gene identified in families with Parkinson's disease. Science. 1997;**276**(5321):2045-2047. DOI: 10.1126/science.276.5321.2045

[4] Trotta L, Guella I, Soldà G, Sironi F, Tesei S, Canesi M, et al. SNCA and MAPT genes: Independent and joint effects in Parkinson disease in the Italian population. Parkinsonism & Related Disorders. 2012;**18**(3):257-262. DOI: 10.1016/j.parkreldis.2011.10.014

[5] Wong Y, Krainc D. α-Synuclein toxicity in neurodegeneration: Mechanism and therapeutic strategies. Nature Medicine. 2017;**23**(2):1-13. DOI: 10.1038/nm.4269

[6] Abeliovich A, Gitler A. Defects in trafficking bridge Parkinson's disease pathology and genetics. Nature. 2016;**539**(7628):207-216. DOI: 10.1038/nature20414

[7] Billingsley K, Bandres-Ciga S, Saez-Atienzar S, Singleton A. Genetic risk factors in Parkinson's disease. Cell and Tissue Research. 2018;**373**(1):9-20. DOI: https://doi.org/10.1007/s00441-018-2817-y

[8] Kalinderi K, Bostantjopoulou S, Fidani L. The genetic background of Parkinson's disease: Current progress and future prospects. Acta Neurologica Scandinavica. 2016;**134**(5):314-326. DOI: 10.1111/ane.12563

[9] Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nature Genetics. 2009;**41**(12):1303-1307. DOI: 10.1038/ng.485

[10] Chang D, Nalls M, Hallgrímsdóttir I, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nature Genetics. 2017;**49**(10):1511-1516. DOI: 10.1038/ng.3955

[11] Nalls M, Pankratz N, Lill C, Do C, Hernandez D, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature Genetics. 2014;**46**(9):989-993. DOI: 10.1038/ng.3043

[12] Ferreira M, Massano J. An updated review of Parkinson's disease genetics and clinicopathological correlations. Acta Neurologica Scandinavica. 2016;**135**(3):273-284. DOI: 10.1111/ane.12616

[13] Kruger R, Menezes Vieira-Saecker A, Kuhn W, Berg D, Muller T, Kohnl N, et al. Increased susceptibility to sporadic Parkinson's disease by a certain combined α-synuclein/apolipoprotein E genotype. Annals of Neurology. 1999;**45**(5):611-617

[14] Maraganore D. Collaborative analysis of α-synuclein gene promoter variability and Parkinson disease. Journal of the American Medical Association. 2006;**296**(6):661. DOI: 10.1001/jama.296.6.661

[15] Cronin K, Ge D, Manninger P, Linnertz C, Rossoshek A, Orrison B, et al. Expansion of the Parkinson disease-associated SNCA-Rep1 allele upregulates human α-synuclein in transgenic mouse brain. Human Molecular Genetics.

2009;**18**(17):3274-3285. DOI: 10.1093/hmg/ddp265

[16] Simón-Sánchez J, Schulte C, Bras J, Sharma M, Gibbs J, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature Genetics. 2009;**41**(12):1308-1312. DOI: 10.1038/ng.487

[17] Mata I, Shi M, Agarwal P, Chung K, Edwards K, Factor S, et al. SNCA variant associated with Parkinson disease and plasma α-synuclein level. Archives of Neurology. 2010;**67**(11):1350-1356. DOI: 10.1001/archneurol.2010.279

[18] Hernandez D, Reed X, Singleton A. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. Journal of Neurochemistry. 2016;**139**:59-74. DOI: 10.1111/jnc.13593

[19] Healy D, Falchi M, O'Sullivan S, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. The Lancet Neurology. 2008;7(7):583-590. DOI: 10.1016/s1474-4422(08)70117-0

[20] Farrer M, Stone J, Lin C, Dächsel J, Hulihan M, Haugarvoll K, et al. Lrrk2 G2385R is an ancestral risk factor for Parkinson's disease in Asia. Parkinsonism & Related Disorders. 2007;**13**(2):89-92. DOI: 10.1016/j.parkreldis.2006.12.001

[21] Ross O, Wu Y, Lee M, Funayama M, Chen M, Soto A, et al. Analysis of Lrrk2 R1628P as a risk factor for Parkinson's disease. Annals of Neurology. 2008;**64**(1):88-92. DOI: 10.1002/ana.21405

[22] Di Battista M, Pascale E, Purcaro C, Passarelli F, Passarelli E, Guglielmi R, et al. Clinical subtypes in Parkinson's disease: The impact of MAPT haplotypes. Journal of Neural Transmission. 2013;**121**(4):353-356. DOI: 10.1007/s00702-013-1117-7

[23] Williams-Gray C, Evans J, Goris A, Foltynie T, Ban M, Robbins T, et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. Brain. 2009;**132**(11):2958-2969. DOI: 10.1093/brain/awp245

[24] Wade-Martins R. The MAPT locus—A genetic paradigm in disease susceptibility. Nature Reviews Neurology. 2012;**8**(9):477-478. DOI: 10.1038/nrneurol.2012.169

[25] Mistry P, Lopez G, Schiffmann R, Barton N, Weinreb N, Sidransky E. Gaucher disease: Progress and ongoing challenges. Molecular Genetics and Metabolism. 2017;**120**(1-2):8-21. DOI: 10.1016/j.ymgme.2016.11.006

[26] Gan-Or Z, Amshalom I, Kilarski L, Bar-Shira A, Gana-Weisz M, Mirelman A, et al. Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology. 2015;**84**(9):880-887. DOI: 10.1212/WNL.0000000000001315

[27] Bembi B, Zambito Marsala S, Sidransky E, Ciana G, Carrozzi M, Zorzon M, et al. Gaucher's disease with Parkinson's disease: Clinical and pathological aspects. Neurology. 2003;**61**(1):99-101. DOI: 10.1212/01.WNL.0000072482.70963.D7

[28] Alcalay R, Dinur T, Quinn T, Sakanaka K, Levy O, Waters C, et al. Comparison of Parkinson risk in Ashkenazi Jewish patients with Gaucher disease and GBA heterozygotes. JAMA Neurology. 2014;**71**(6):752. DOI: 10.1001/jamaneurol.2014.313

[29] Bultron G, Kacena K, Pearson D, Boxer M, Yang R, Sathe S, et al. The risk of Parkinson's disease in type 1 Gaucher disease. Journal of Inherited Metabolic

Disease. 2010;**33**(2):167-173. DOI: 10.1007/s10545-010-9055-0

[30] Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the Glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. New England Journal of Medicine. 2004;**351**(19):1972-1977. DOI: 10.1056/nejmoa033277

[31] Amaral C, Lopes P, Ferreira J, Alves E, Montenegro M, Costa E, et al. GBA mutations p.N370S and p.L444P are associated with Parkinson's disease in patients from northern Brazil. Arquivos de Neuro-Psiquiatria. 2019;**77**(2):73-79. DOI: 10.1590/0004-282x20190006

[32] Bras J, Paisan-Ruiz C, Guerreiro R, Ribeiro M, Morgadinho A, Januario C, et al. Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal. Neurobiology of Aging. 2009;**30**(9):1515-1517. DOI: 10.1016/j.neurobiolaging.2007.11.016

[33] Nichols W, Pankratz N, Marek D, Pauciulo M, Elsaesser V, Halter C, et al. Mutations in GBA are associated with familial Parkinson disease susceptibility and age at onset. Neurology. 2008;**72**(4):310-316. DOI: 10.1212/01.wnl.0000327823.81237.d1

[34] Moraitou M, Hadjigeorgiou G, Monopolis I, Dardiotis E, Bozi M, Vassilatis D, et al. β-Glucocerebrosidase gene mutations in two cohorts of Greek patients with sporadic Parkinson's disease. Molecular Genetics and Metabolism. 2011;**104**(1-2):149-152. DOI: 10.1016/j.ymgme.2011.06.015

[35] Winder-Rhodes S, Evans J, Ban M, Mason S, Williams-Gray C, Foltynie T, et al. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Brain. 2013;**136**(2):392-399. DOI: 10.1093/brain/aws318

[36] Asselta R, Rimoldi V, Siri C, Cilia R, Guella I, Tesei S, et al. Glucocerebrosidase Mutations in Primary Parkinsonism. Parkinsonism & Related Disorders. 2014;**20**(11):1215-1220. DOI: 10.1016/j.parkreldis.2014.09.003

[37] Lesage S, Condroyer C, Hecham N, Anheim M, Belarbi S, Lohman E, et al. Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa. Neurology. 2011;**76**(3):301-303. DOI: 10.1212/wnl.0b013e318207b01e

[38] Eblan M, Nguyen J, Ziegler S, Lwin A, Hanson M, Gallardo M, et al. Glucocerebrosidase mutations are also found in subjects with early-onset parkinsonism from Venezuela. Movement Disorders. 2006;**21**(2):282-283. DOI: 10.1002/mds.20766

[39] Noreau A, Rivière J, Diab S, Dion P, Panisset M, Soland V, et al. Glucocerebrosidase mutations in a French-Canadian Parkinson's disease cohort. Canadian Journal of Neurological Sciences/Journal Canadien des Sciences Neurologiques. 2011;**38**(5):772-773. DOI: https://doi.org/10.1017/S0317167100012300

[40] Mitsui J, Mizuta I, Toyoda A, Ashida R, Takahashi Y, Goto J, et al. Mutations for Gaucher disease confer high susceptibility to Parkinson disease. Archives of Neurology. 2009;**66**(5):571-576. DOI: 10.1001/archneurol.2009.72

[41] Mao X, Burgunder J, Zhang Z, An X, Zhang J, Yang Y, et al. Association between GBA L444P mutation and sporadic Parkinson's disease from mainland China. Neuroscience Letters. 2010;**469**(2):256-259. DOI: 10.1016/j.neulet.2009.12.007

[42] Sidransky E, Nalls M, Aasly J, Aharon-Peretz J, Annesi G, Barbosa E, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. New England Journal of Medicine.

2009;**361**(17):1651-1661. DOI: 10.1056/NEJMoa0901281

[43] Anheim M, Elbaz A, Lesage S, Durr A, Condroyer C, Viallet F, et al. Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. Neurology. 2012;**78**(6):417-420. DOI: 10.1212/WNL.0b013e318245f476

[44] Liu G, Boot B, Locascio J, Jansen I, Winder-Rhodes S, Eberly S, et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. Annals of Neurology. 2016;**80**(5):674-685. DOI: 10.1002/ana.24781

[45] Beavan M, McNeill A, Proukakis C, Hughes D, Mehta A, Schapira A. Evolution of prodromal clinical markers of Parkinson disease in aGBAMutation–positive cohort. JAMA Neurology. 2015;**72**(2):201. DOI: 10.1001/jamaneurol.2014.2950

[46] Cilia R, Tunesi S, Marotta G, Cereda E, Siri C, Tesei S, et al. Survival and dementia inGBA-associated Parkinson's disease: The mutation matters. Annals of Neurology. 2016;**80**(5):662-673. DOI: 10.1002/ana.24777

[47] Brockmann K, Srulijes K, Pflederer S, Hauser A, Schulte C, Maetzler W, et al. GBA -associated Parkinson's disease: Reduced survival and more rapid progression in a prospective longitudinal study. Movement Disorders. 2014;**30**(3):407-411. DOI: 10.1002/mds.26071

[48] Thaler A, Bregman N, Gurevich T, Shiner T, Dror Y, Zmira O, et al. Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene. Parkinsonism & Related Disorders. 2018;**55**:45-49. DOI: https://doi.org/10.1016/j.parkreldis.2018.05.009

[49] Toft M, Pielsticker L, Ross O, Aasly J, Farrer M. Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population. Neurology. 2006;**66**(3):415-417. DOI: 10.1212/01.wnl.0000196492.80676.7c

[50] Santos N, Ribeiro-Rodrigues E, Ribeiro-dos-Santos Â, Pereira R, Gusmão L, Amorim A, et al. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Human Mutation. 2010;**31**(2):184-190. DOI: 10.1002/humu.21159

[51] McNeill A, Duran R, Hughes D, Mehta A, Schapira A. A clinical and family history study of Parkinson's disease in heterozygous-glucocerebrosidasemutation carriers. Journal of Neurology, Neurosurgery & Psychiatry. 2012;**83**(8):853-854. DOI: 10.1136/jnnp-2012-302402

[52] Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. The Lancet Neurology. 2012;**11**(11):986-998. DOI: https://doi.org/10.1016/S1474-4422(12)70190-4

[53] Feany M. New genetic insights into Parkinson's disease. New England Journal of Medicine. 2004;**351**(19):1937-1940. DOI: 10.1056/NEJMp048263

[54] Sardi S, Clarke J, Kinnecom C, Tamsett T, Li L, Stanek L, et al. CNS expression of glucocerebrosidase corrects α-synuclein pathology and memory in a mouse model of Gaucher-related synucleinopathy. Proceedings of the National Academy of Sciences. 2011;**108**(29):12101-12106. DOI: www.pnas.org/cgi/doi/10.1073/pnas.1108197108

[55] Robak L, Jansen I, van Rooij J, Uitterlinden A, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. Brain. 2017;**140**(12):3191-3203. DOI: 10.1093/brain/awx285

Chapter 5

Endogenous Retroelements in Cancer: Molecular Roles and Clinical Approach

Kang-Hoon Lee and Je-Yoel Cho

Abstract

Retroelements have been considered as "Junk" DNA although the encyclopedia of DNA elements (ENCODE) project has demonstrated that most of the genome is functional. Since the contribution of LINE1 (L1) and human endogenous retrovirus (HERV) has been suspected to cause human cancers, their regulations and putative molecular functions have been investigated in diverse types of cancer. Their diagnostic, prognostic, and therapeutic potentials have been incessantly proposed using cancer associated or specific properties, such as hypomethylation, increased transcripts, and reverse transcriptase, as well as cancer-associated antigens. This chapter presents the current knowledge on retroelements in various aspects during tumorigenesis and their clinical usage in many cancer studies.

Keywords: retrotransposons, repetitive elements, tumorigenesis, cancer, LINE, HERV, retroelement

1. Introduction

In recent decades, the development of genomic analysis technology has played an important role in the study and treatment of various diseases [1, 2]. However, these studies have been focused on genes that form proteins that account for about 1–2% of the entire genome, and the understanding of other parts remains relatively insufficient. A retroelement (RE), also called a retrotransposon, is a type I transposable element that replicates itself via RNA and reverse transcription and can be largely classified into two types based on the genome structure, including long terminal repeat sequences (LTRs). The intact endogenous retrovirus (ERVs) retains two LTRs at both ends of the genome, instead of long and short interspersed nuclear elements (LINE and SINE), which are non-LTR groups. LTRs compose ~8% of the human genome and most are known to be inactive due to accumulated mutations. Yet, interestingly, many are transcriptionally active [3]. The non-LTR groups can be divided again into autonomous LINEs and nonautonomous SINEs that need LINE's proteins [4]. The LINE1s (L1s), known as the only active REs, makes up ~17% of the human genome. Intact L1s retain ~6 kb of the genome, which encodes two proteins, ORF1 and ORF2, which are essential for replication and reverse transcription [5]. There are about 145 full-length, functional L1 elements in the human genome. On the other hand, SINEs, which are nonautonomous retroelements, have ~300 bp genomes without coding potential. Most SINEs are of the Alu type of which there are over one million copies in the human genome [6].

The association between REs and cancer has been suggested since 1950. As the presence of a viral-oncogene was unveiled and mouse mammary tumor virus (MMTV) became the accepted etiological agent of mammary tumors in mice, the possible carcinogenesis mechanism of ERV was also revealed, raising hope for overcoming cancer [7, 8]. Many studies have reported the association of RE expression with various cancer types, including breast cancer, melanoma, and kidney cancer [9]. However, the function of RE expression in cancer as a driver or passenger remains controversial [10, 11]. It is a chicken and egg situation, since the cancer-associated RE expression can cause malignant cell transformation and malignant cell transformation leads to global DNA hypomethylation, which in turn contributes to oncogenic RE expression [12–15]. In addition, the fact that most REs have lost their transposition activity due to accumulated mutations makes it difficult to evaluate the role of REs [16]. The RE sequences that occupies about half of the mammalian genome is known as "junk DNA," and, as the name suggests, little research has been done it [17]. However, in certain areas such as in the early embryogenesis process, degenerative disease, and cancer, the expression of REs have been studied relatively well [18, 19]. In particular, several studies have been conducted to reveal the relationship among the environmental stress, RE responses, and associated diseases [20, 21]. Although no direct relationship has been revealed yet, genome instability by activated RE is known to be the main mechanism linking RE with disease [22]. However, the transposition ratio of all the REs is about 0.02 germline events per generation [23], so it is too rare to explain their various roles.

In this chapter, we focus on the functional mechanisms of REs in various cancers from development to metastasis and from diagnosis to cancer therapy.

2. RE regulation in normal cells and abnormal reactivation and expansion in cancer

Fortunately, except for during the reprogramming process in early stage germ cells, most REs are strongly silenced by diverse epigenomic controls and their reactivation is molecularly inhibited [24, 25].

DNA methylation is a major epigenetic mechanism that contributes to retrotransposon silencing in both normal and cancer cells [26]. In early embryogenesis, a genome-wide DNA methylation is established by the DNA methyltransferase 3 (Dnmt3) and maintained by the methyltransferase1 (Dnmt1) [27]. Parental methylation pattern is genome-wide demethylated and methylated again at imprinted loci and REs by the Dnmt3, and these patterns are maintained by Dnmt1 in somatic cells [28–30]. Association between demethylation and RE expression was demonstrated in that the inactivation of DNMT3L, which is a non-catalytic homolog of DNMT3A/3B, causes the reactivation of L1 and IAP and leads to meiotic arrest as well as male sterility in male germ cells [31–33].

In cancer cells, a genome-wide DNA hypomethylation and the reactivation of REs that may result in the loss of chromosomal stability and imprinting patterns are well known [34]. Alteration of L1 methylation has been investigated in many types of cancers, including breast, colon, lung, ovarian, and prostate cancers [35–37]. Mostly, hypomethylation of the L1 promoter is associated with genome instability, aggressive histology, poor prognosis, and some metastasis [38]. Interestingly, some abnormal features, such as chromosome 8 abnormalities, are also associated with L1 hypomethylation [39]. In addition, due to their prevalent unmethylation in cancer samples, a moderate increase of Alu was also observed in cancer samples with a hypomethylated L1 promoter [40]. Similarly, hypomethylation of HERV has also been reported in various cancer cells [9, 12, 41–44]. Hypomethylation of its long terminal repeat (LTR)

where the promoter is located is associated with its overexpression in cancer [45]. Numerous HERV family members were expressed in cancer cell lines and primary tumor tissues. In a head and neck cancer study, tumor-specific methylation changes were found in HERV-H, HERV-W, and HERV-K families [24, 46]. Similarly, the hypomethylated CpGs resulting in high expression of HERV-K, -W, and L1 was reported in ovarian cancer [47]. Moreover, the hypomethylation of REs has been observed in specific stages or subtypes of cancer, such as during ovarian cancer progression and in the basal subtype of invasive ductal carcinoma breast cancer [48, 49]. Remarkably, individual RE expressions associated with cancer such as HERV-K at 22q11.23 (H22q), HERV-H5, HERV-H48–1, and HERV-E4 are highlighted in various cancers [46, 50, 51]. Their transcripts or viral proteins have been detected in sera from bladder, breast, liver, lung, ovarian, and prostate cancer patients [11].

The last cellular epigenomic regulation mechanism for silencing RE expression is histone modification [52]. In normal spermatogonia, one of the repressive histone modification marks, histone 3 lysine 9 dimethylation (H3K9me2), causes transcriptional repression and is sufficient to maintain L1 silencing in the absence of DNA methylation. Thus, the loss of H3K9me2 combined with the absence of DNA methylation may be the cause of LINE1 activation [53]. On the other hand, in the study of the association of histone modification with RE expression in cancer, two repressive histone modifications, H3K9me3 and H3K27me3, were more enriched at H22q, HERVK17, and L1 sequences in PC3 than in LNCaP prostate cell lines, of which RE expression levels are high and low, respectively. By contrast, the active modification H3K4me3 was the most enriched in LNCaP at the H22q LTR [54].

The expressed RE transcripts can eventually be knocked down by the PIWI system [55]. Piwi-interacting RNA (piRNA) is a well-studied mechanism that contributes to the silencing of REs in many animal germline cells [56, 57]. The piRNA system is a ribonucleoprotein complex consisting of a piRNA, and a P-element-induced wimpy testis (PIWI) subfamily of Argonaut nucleases protein [58]. The piRNA recognizes RE sequences and the PIWI protein destroys the RE transcripts [58, 59]. The piRNA system silences RE expression both at the transcriptional and posttranscriptional levels by modifying repressive chromatin modifications and by cleaving RE transcripts, respectively [57, 60]. However, the role of piRNA in post-transcriptional regulation is not similar to that of miRNA via providing sequence specificity because most piRNA sequences are found not to be complementary to target gene transcripts, suggesting that piRNAs may be involved in epigenetic regulation rather than posttranscriptional regulation of mRNA [61]. The deficient of the piRNA pathway causes overexpression of REs, significantly compromised genome structure and, invariably, germ cell death and sterility [58]. The aberrant expression of piRNAs has been reported in the development of cancer including the proliferation, apoptosis, metastasis, and invasion of cancer cells [62]. Moreover, the high expression of PIWI proteins has been documented in many cancer types, including gastric cancer, liver cancer, intestinal cancer, breast cancer, nonsmall cell lung cancer, bladder cancer, ovarian cancer, and melanoma and is furthermore associated with the aggressiveness of sarcomas, gliomas, and leukemia [61, 63]. The roles of PIWI proteins have been investigated separately in cancer invasion, migration, proliferation, division, and survival [64]. PIWIL1 has been known to induce epithelial-mesenchymal transition and confer migration and invasion of endometrial cancer cells [65]. The association of PIWIL2 via increasing the expression of CDK2 and cyclin A in cancer cells is reported in glioma and nonsmall lung cancer (NSCLC) cells [66]. PIWIL3 promotes the cancer proliferation, migration, and invasion through the JAK2/STAT3 signal pathway [67]. PIWIL4 can promote cancer cell division, migration, and survival of breast cancer by activating TGF-β, MAPK/ERK, and FGF signaling pathways [68].

The apolipoprotein B mRNA editing catalytic polypeptide 3 (APOBEC3) proteins are cytidine deaminases of which family consists of seven family members (APOBEC3-A through -H) with diverse activities against a variety of retroviruses and endogenous REs, even though the activity of L1 suppression does not correlate either with antiviral activity against Vif-deficient HIV-1 and murine leukemia virus, or with patterns of subcellular localization [69, 70]. Thus, the inhibitory effect of APOBEC3 family members, specifically APOBEC3G on L1 transposition might not be due to deaminase activity, but due to novel mechanism(s) [70].

Besides APOBEC3G, MOV10, SAMHD1, and ZAP have all been identified to be able to inhibit L1 activity through diverse mechanisms [71]. MOV10 inhibits L1 mobility through interacting with L1 RNP resulting in L1 transcript degradation [72]. SAMHD1 inhibits the L1 RT activity [73]. ZAP also restricts L1 activity through the loss of L1 transcripts and ribonucleoprotein integrity [74].

Together, it will be a universal explanation for the various epigenomic modifications that are directly associated with both genome-wide RE silencing and reactivation that is much more commonly found in diverse human cancers as frequent as 4–100 de novo insertions per tumor.

3. Roles of RE expressed in cancers

The genomic instability caused by de novo insertions of REs that frequently occur in cancer is the major pathophysiological role accepted by the public [75, 76]. However, this is a very limited explanation of the universal functions of REs, because most REs lose their ability to mobilize [16]. Although some retain their coding potentials, these are silenced tightly by various mechanisms and at various levels, such as epigenomic mechanisms, transcription, and posttranscription [77]. Thus, a more in-depth understanding of RE function is mandatory.

3.1 The source of genome instability

De novo insertions of REs, despite their defective form, can both directly and indirectly affect surrounding human genome sequences [78]. Some of these events occur at high enough frequency to result in vast amounts of rearrangement of the host genome sequence [16]. This does not happen only via the mechanism of transposition activity followed by reintegration but also via the homologous recombination between dispersed REs, resulting in large structural variations (SVs) including duplications, inversions, and deletions [79]. REs are also the source of small SVs such as single-nucleotide variants (SNVs) and short indels, which are caused by template switching during repair of replication errors [16]. The SVs derived from reactivation and expansion of REs via either mobilization activity or homologous recombination have been frequently found in many cancers (~50%) [80, 81]. A high enrichment was reported especially in certain types of cancers, such as esophageal cancers, colon cancers, and squamous cell lung cancers (> 90%) [82]. Although this result indicated that somatic L1 insertions are very frequently found in certain cancers, it is known that a majority of RE somatic integrations are passenger mutations with little or no effect on cancer development [83].

Nevertheless, specific SV loci derived from somatic L1 insertions have also been identified as drivers in most cancer types, including colorectal, breast, lung, and liver cancers [84–88]. For example, disruption of the APC gene by the insertion of L1 in colon cancer has been well studied [89]. Additionally, a recent study identified driver SV by L1 insertion in liver cancer [90]. L1 integration in the intron of the ST18 gene disrupted a cis-regulatory repressor element, resulting in increased expression of the ST18 gene [84].

Several algorithms have also been developed for the sensitive and precise detection of SVs from the whole genome sequence (WGS) and whole exome sequence (WES) data published in large international consortia such as The International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA), and driver SV events with remarkable functional consequences have been identified [82, 91]. The most SVs were generated by L1 (99%), followed by SINE VNTR Alu (SVA) and ERV [92]. Yet, few retrotranspositions of HERVs have been reported in human cancers [84, 93].

3.2 Epigenomic regulation and reactivation of REs in cancer

Since 1993 when the methylation status of L1 in cancer cells was first measured by Thayer et al., L1 hypomethylation has been reported in many types of human cancers, including prostate, ovarian, head and neck, lung, thyroid, and breast cancer [94, 95]. However, some controversial results showed no changes in L1 methylation levels of cancers including thyroid cancer, renal cancer, lymphoma, and leukemia [96]. This discrepancy may be due to differences in the tumor histological type, because association between L1 hypomethylation and clinical outcome has been demonstrated in melanoma patients. However, the mechanism of L1 hypomethylation effects on aggressive tumor behavior has not been fully investigated [49]. The most likely mechanism is the causing of DNA instability, which has been suspected as the main role of REs [92]. A DNA methyltransferase 1 (Dnmt1) mutation showed substantial genome-wide hypomethylation in all types of tissue and also known to be associated with aggressive T cell lymphomas [97, 98]. Notably, the mutation also showed a high frequency of chromosome 15 trisomy, which suggested that the DNA hypomethylation has a causal role in cancers by promoting genome instability [98]. Another possible mechanism is a dysregulation in transcription level, which activates proto-oncogenes and REs that affect tumor aggressiveness [99]. MicroRNAs, which are closely related to the development of human cancer, can be increased by global DNA hypomethylation, contributing to the acquisition of tumor aggressiveness [100]. In addition, it is possible that the L1 methylation state itself exerts a biological effect. It is known that L1 regulates the function of multiple genes by providing an alternative promoter and contributing to noncoding RNA expression [101, 102]. Therefore, further studies are needed to explain the mechanisms in which L1 hypomethylation affects tumor behavior.

3.3 REs, the origin of cancer associated non-coding transcripts

RNA sequencing using next-generation sequencing technology has provided a large amount of gene expression data in both normal and disease conditions, such as cancer [103]. Growing evidence suggests that REs in the intergenic regions of the human genome are sources of noncoding RNAs, including micro RNAs (miRNAs) and long noncoding RNAs (lncRNAs) [104]. Notably, about 30% of human lncRNAs originate from REs, specifically HERVs. In addition, about 80% of lncRNAs contain RE-originated sequences within or nearby their transcription start sites [105]. Importantly, a recent study has reported that many lncRNAs have a crucial role in a variety of fundamental cellular processes and diseases [106]. A recent study reported that a single-nucleotide polymorphism (SNP) in an L1-containing lncRNA sequence located in an intron of SLC7A2 leads to a decrease in its expression and results in a lethal encephalopathy phenotype [107]. Alu elements, which encode no functional proteins, are also frequently found at multiple locations in lncRNA sequences [108]. Recently, many studies have suggested that Alu sequence in lncRNAs can contribute to the function of lncRNAs. For example, Alu-mediated CDKN1A/p21 transcriptional regulator (APTR) negatively regulates p21 expression by recruiting polycomb

repressive proteins to the p21 promoter. The Alu sequence is crucial to the localization of APTR on the p21 promoter that regulates cell growth and proliferation [109].

Despite the limited contribution of L1 and Alu to lncRNAs, a close association between HERVs and ncRNAs was reported by Kelley and Rinn [110]. Hundreds of ncRNAs originated from HERV-H. For example, the lncRNA ROR known to promote the progression of human cancers is one of the ncRNAs promoted by a HERV-H element [111]. Moreover, the lncRNA produced by HERV-K11 directly binds to polypyrimidine tract-binding protein-associated splicing factor (PSF), of which the function is to repress proto-oncogene transcription, reversing the PSF-mediated repression of proto-oncogene transcription and subsequently driving tumorigenesis [46, 112]. Other HERV-related lncRNAs with tumor-suppressive potential have also been identified in the intronic RNAs arising from ERV-9 [45]. It has been reported that its antisense RNA at 3′-untranslated regions was found to physically bind to key transcription factors for cell proliferation such as NF-Y, p53, and sp1. This means that the HERV-related lncRNAs may have a function as decoy targets or traps for the transcription factors resulting in the growth retardation of cancer cells [113].

Another role of RE transcripts related to human disease is to form a complex with the cytoplasmic cDNA of the reactivated RE transcripts to trigger the signal of the inflammatory pathway [23]; for example, RE-derived cytosolic DNA accumulated in Aicardi-Goutières syndrome (AGS) [114]. IFNB1 expression also has an anticorrelation with L1 retrotransposition in cancer cells [115]. Moreover, the study by Ishak et al. showed that mutation of the RB1 gene causes both genome-wide upregulation of L1 expression in somatic cells as well as increased susceptibility to leukemia [116]. Gasche et al. reported that the IL-6 treatment of a cancer cell line induced genome-wide L1 promoter hypomethylation [117]. Altogether, the evidence indicates that REs modify an important aspect of human tumorigenesis.

3.4 RE proteins associated with tumorigenesis

ORF1 and ORF2 in L1 and GAG, POL, and ENV in HERV are proteins encoded by REs that are essential to complete the replication cycle, whereas Alu's are RNA polymerase III-transcribed sequences without coding potential [118]. Most REs lose their coding potential due to accumulated mutations; however, it is well known that hundreds of L1 are still active to produce two essential proteins, ORF1 (p40, RNA binding protein) and ORF2 (p109, endonuclease and reverse transcriptase activities) [119, 120]. Additionally, although no infectious virus formed by HERVs is reported, multiple protein expressions and their functions have been studied in various HERV families [46]. Most comprehensive studies have reported on envelop proteins (ENV) and their pathogenic properties. The transcripts encoding capsid and protease (GAG) and reverse transcriptase with RNase H domain and integrase (POL) ORFs have been detected in many cells and tissues from both diseased and healthy individuals [121]. Remarkably, HERV-W encodes an ENV protein known as ERVWE1 (Syncytin1), which has been adopted by the human to functionally contribute in placenta biogenesis [122]. Similarly, Syncytin2 encoded by ERVFRD1 is known to have a key role in the implantation of human embryos [123]. Aberrant expression of HERV-W has been known to be associated with various human diseases including cancer [122, 124, 125].

In cancer, an increase in retroviral protein expression was generally detected. Overexpression of L1 ORF1 protein was detected from more than 90% of breast, ovarian, and pancreatic cancers followed by tubular gastrointestinal tract, lung, and prostate cancers (about 50%) [126, 127]. However, the high expression of L1 ORF1p expression is dependent on tumor origin, and it differs case by case even within a similar histological type of cancer. For example, L1 ORF1p is detected in lung adenocarcinoma at greatly varying levels (about 20% are very high, about 30% are

moderate, and the rest are undetectable) [128]. Several antibodies targeting ORF2p have recently been produced, and thus, the overexpression of ORF2p was detected in many cancers. Although the functional effects of L1 proteins in human cancers remain unclear in most cancer contexts, this data suggests that L1 proteins are potential cancer biomarkers for the diagnosis of cancer development or the prognosis of clinical outcomes [126, 129]. On the other hand, the HERV-K ENV protein has been identified in various cancer tissues and several different mechanisms by which it associates with tumorigenesis have been proposed [130]. The melanocyte antigen HERV-K-MEL is expressed in about 85% of malignant melanocytes, whereas breast cancer, ovarian cancer, teratocarcinoma, sarcoma, and bladder cancer also express HERV-K ENV [131]. Other HERV families, HERV-E, and ERV3 have also been detected in more than 30% of ovarian cancer patients and are higher in patients with lymph-node-positive breast cancer [11, 132]. Moreover, some antibodies against HERV-K have been detected in serum samples with melanoma [133].

Despite HERVs being known to be incompetent in transposition, studies have shown that the protein-coding potentials can still promote neoplastic properties during tumorigenesis through diverse mechanisms [134]. The oncogenic role of HERV proteins is well investigated with NP9 and REC, which are accessory splice proteins of HERV-K [135]. The transcripts encoding these proteins are overexpressed in many tumors including breast cancers and both are known to interact with the promyelocytic leukemia zinc finger (PLZF) tumor suppressor, which is a transcriptional repressor and epigenetic modulator implicated in cancer. C-myc proto-oncogene is one of the major targets of PLZF. Interaction of NP9 and REC with PLZF abrogates the transcriptional repression of the c-Myc gene promoter, which results in c-Myc overproduction [136]. In addition, the abnormal cell-to-cell fusion activity of HERV-W ENV proteins has been shown to possibly contribute to tumor development and metastasis [130]. Further studies to characterize the expression and molecular functions of these HERV proteins in cancers are demanded.

4. Implementation of REs for cancer diagnosis and prognosis

4.1 Structural variations (SVs) associated with REs in cancer

Identification of somatic mutation hotspots associated with cancer is very important for functional analysis and diagnosis [137]. Several methods have been developed for the identification of somatic RE insertions in cancers (L1-seq, TIPseq, and ERVcaller), and many bioinformatics tools to discover somatic L1 insertions in silico using WGS or WES data have been developed [138, 139]. SVs via L1 insertion associated with cancer have been well investigated in a couple of genes, such as the APC gene that is considered to be a tumor suppressor of colorectal polyposis in colorectal cancer [89]. A potential suppressor of L1, TP53 mutation by L1 insertions, has been observed frequently in tumors. In addition, L1 insertional mutation of MOV10, which is a key L1 suppressor, decreased the expression of the MOV10 in tumors with high L1 insertions [140].

On the other hand, instead of cancer-associated SVs caused by RE insertion, genome variations that might be associated with HERVs or around gene expression in cancer have been identified. Chang et al. identified that four HERVs with mutation hotspots overlapped with exons of four human protein coding genes, which are TNN (HERV-9/LTR12), OR4K15 (HERV-IP10F/LTR10F), ZNF99 (HERV-W/HERV17/LTR17), and KIR2DL1 (MST/MaLR). They also evaluated the effect of each non-synonymous SNV on the survival of kidney cancer patients. Furthermore, they identified 788 HERVs harboring significantly increased the numbers of somatic single-nucleotide variations (SNVs) [141].

4.2 Global hypomethylation in cancer and identification of cancer associated RE methylation

Several studies have shown that global hypomethylation is very common in cancer [142]. The DNA methylation levels of L1 5′-untranslated region (UTR) in cancer have been extensively evaluated for potential use as an epigenomic marker for cancer diagnosis. The level of L1 hypomethylation increases in more advanced cancers; however, other types of REs, such as Alu and HERVs, have been lesser evaluated [143]. Since DNA methylation analysis has some benefits in handling tumor specimens, such as similar efficiency in fresh frozen and formalin-fixed paraffin-embedded tissue, many studies indeed have proposed DNA methylation as a diagnostic marker using fresh tumor biopsies or fixed tissue blocks [144]. Association between L1 hypomethylation and diagnostic and prognostic needs, such as tumor stage group, metastasis, the recurrence rate, and the survival rate, has been studied [145]. Also, L1 hypomethylation has been demonstrated to be a surrogate marker for predicting the response to cancer treatment [146]. Moreover, L1 hypomethylation is observed in very different types of specimen, including blood leukocyte DNA, serum, and oral rinse [147]. Hypomethylation of Alu was reported in several cancers, whereas hypomethylation of HERV-K and HERV-W genomes were found in urothelial cancer and ovarian cancer, respectively [47, 131, 148].

Classically, CpG methylation analyses have been performed in targeted sequence by discriminating between methylated and unmethylated DNA using bisulfite treatment followed by PCR amplification [149]. Although recent nanopore technology can separate between methylated and unmethylated DNA without any treatment, most analyses are usually based on methylation-specific PCR after bisulfite treatment (MSP) [150, 151]. Pyrosequencing-based analysis, specifically methylation-sensitive single-nucleotide-primer extension (MsSNuPE) and Methylight, is a promising method that can be used to reliably measure L1 methylation in paraffin-embedded cancer tissues with higher reproducibility [152]. Using this method, L1 hypomethylation has been tested in various human cancer patients, including gastric cancer, colon cancer, colorectal cancer, melanoma, and breast cancer, and its clinical implications have been suggested [153]. Recent studies have addressed that methylated L1 in circulating cell-free DNA (cfDNA) can be used as a potential prognostic and diagnostic target in cancers, and have promoted its potential as a minimally invasive screening technique. Lee et al. showed L1 hypomethylation in cfDNA of both human breast cancer and dog mammary tumor [154, 155].

Unfortunately, there are not many products in the marketplace that capitalize on the association between RE hypomethylation and diverse cancer types and features, even though many studies have provided evidence for it. Representatively, the only clinical test targeting methylation of L1 is used in the detection of bladder cancer in voided urine [156].

4.3 RE transcripts in cancer diagnosis

First of all, the quantitation of various HERV gene expressions was performed using a real-time PCR. The transcript expression of HERV-H, -K, -P, and -R ENV was significantly increased in the blood of lung cancer patients, and the level was generally much higher in the squamous cell carcinoma (SCC) subtype than the small-cell lung cancer (SCLC) subtype [157]. The level of HERV-K (HML-2) was found to be an independent prognostic factor for the overall survival rate of hepatocellular carcinoma patients [158]. The expression of HERV-H LTR-associating protein 2 (HHLA2) was significantly upregulated in bladder cancer, and it was suggested as a prognostic factor of tumor metastasis and poor survival of bladder

cancer patients [159]. The elevated HERV-K (HML-2) was detected in both protein and transcripts level in the blood of breast cancer patients at an early stage and was further increased with developing metastasis. Thus, HERV-K (HML-2) expression will be one of a best candidate for the early detection of an increased risk for breast cancer in women [160]. The expression of HERV-E transcripts is observed in von Hippel-Lindau (VHL)-deficient renal carcinomas. Interestingly, the introduced VHL gene suppressed HERV-E expression in VHL-deficient carcinoma [11]. In addition, high blood levels of the ENV transcripts of various HERV types have been detected in breast cancer patients and that are decreased by treatment of adjuvant chemotherapy which means that alteration of blood HERV transcripts is a very good candidate for diagnosis and is a prognosis marker of breast cancer [132].

4.4 Detection of RE proteins in cancer specimens

A correlation between HERV protein expression and human cancer has been described [11]. HERV proteins, GAG, POL, and ENV, have been identified in cancer tissues, and several factors from environment and hormone response, such as UV radiation, inflammation, estrogen and smoking, have been proposed as a cause of HERV protein expression in various cancer tissues [161]. Remarkably, the envelop protein, ENV, of HERV-K has been identified in melanoma by immunohistochemistry [162]. In melanomas, the expression of HERV-K ENV is higher than that in benign lesions, especially in metastatic tumors. Moreover, it has also been found in other types of cancers, such as breast, ovarian, and bladder cancer. Antibodies targeting HERV-E, HERV-K (HML-2), and ERV3 have also been detected in more than 30% of ovarian cancer patients and are higher in patients with lymph-node-positive breast tumors. In addition, the presence of serum antibodies against HERV-K proteins has been suggested as a prognostic factor for poor survival of melanoma patients [11].

In L1 proteins, high levels of ORF1 protein was prevalent in certain cancers, including breast and ovarian cancer, whereas no or little expression was detected from other cancers such as renal, liver, and cervical cancer [36]. Rodic et al. and Ardeljan et al. separately detected ORF1 protein via IHC in ~90% of ovarian cancer and in ~90% of the breast cancer samples examined [127, 163]. Chen et al. reported that the ORF1 protein level is very high in ductal carcinoma in situ (DCIS) [164]. Moreover, the ORF1 level was the highest in high-grade ovarian carcinoma, but the expression of ORF1 in prostate cancer has not been fully confirmed [36]. Ardeljan et al. reported ORF1 positivity in ~41% of all prostate cancer tissue samples examined [163]. ORF1 levels could be clinically measured using CT scans on the blood of lung cancer patients. On the other hand, ORF2 has only been limitedly tested as a diagnostic marker for cancer when compared to ORF1 expression. However, since ORF2 encodes a reverse transcriptase that is heavily associated with L1 activity, similar to L1 hypomethylation, it may yet be a better diagnostic marker for L1-associated disease development. High expression of ORF2 in transitional colon mucosa but no expression in normal colon mucosa was detected via IHC. ORF2 was also detected in prostate intraepithelial neoplasia [36]. However, since the ORF2 expression has been reported to be much less than that of ORF1, there are challenges to measure it in clinic.

5. RE in cancer therapy

Aberration of RE activities in various aspects has been suggested as a potential target for cancer therapy [165]. Several studies have shown that inhibiting RT

activity is a great therapeutic target for cancer. Sciamanna et al., 2005, uncovered that pharmacologic L1 inhibition by two reverse transcriptase inhibitors slows down the growth of malignant melanoma and prostatic cancer [166]. Carlini et al. similarly demonstrated the efficacy of reverse transcription inhibition of prostate cancer growth [167]. Furthermore, a clinical trial showed that Efavirenz, a non-nucleoside reverse transcriptase inhibitor (NNRTI), provides a therapeutic benefit by increasing the progression free survival in a high-stage castration-resistant prostate cancer cohort [168]. Recently, Efavirenz has been shown to suppress L1 activity and promote morphological differentiation in melanoma cells [169]. On the other hand, another class of RT inhibitor, the nucleoside reverse transcriptase inhibitor (NRTI), has also been shown to suppress L1 activity and induce anticancer activity in prostate cancer cell lines. Importantly, no significant effects were observed in normal cells [167]. Despite these successful findings, it is still unclear regarding the molecular function of RT inhibition in the gene expression regulation.

RNA interference (RNAi)-mediated downregulation of L1 generated identical effects to those observed with RT inhibitory drugs in human melanoma, which indicates that RT activity has a crucial role in L1 activity in human cancer [170]. Recently, a phase II human trial using Efavirenz on a cohort of metastatic patients with prostate cancer showed nonprogression when Efavirenz reached an optimal concentration in the blood [171]. Altogether, preclinical and clinical data provide evidence that RT inhibition is a potentially effective tool in a novel anticancer therapy against diverse human cancers with noncytotoxic effects on non-cancer cells [172].

Another approach regarding REs is an immunotherapy approach to target the pro-oncogenic effects of HERV ENV, which is possibly involved in tumor progression and in downstream metastatic spread, in a number of tissues. HERV ENVs exclusively upregulated in tumor tissues will be suitable targets to direct both passive and active immunotherapy against in cancer cells [130]. The antibodies recognizing the HERV ENVs has been developed, and currently, a monoclonal antibody against HERV-K (HML2) ENV successfully inhibits human breast cancer proliferation, with the activation of apoptosis [173]. On the other hand, various HERV-derived ENVs have been investigated as candidates of anticancer immunotherapy, either as tumor-associated or tumor-specific antigens in cancer cells [130]. ERVs were first used for antitumor immunization in the murine cancer models expressing ERV [9]. Similarly, in humans, protective immunity against the HERV-K MEL antigen in melanoma development has been investigated. This active immunotherapy is considered more advantageous with respect to passive immunization [130]. However, despite the antigenic similarity between HERV-K-MEL and yellow fever virus (YFV), no significant protective effects were shown in the 10 years post-anti-YFV vaccinations in the melanoma cohorts [174, 175]. HERV-H ENV (Xp22.3) is an another antigen significantly upregulated in a subset of gastrointestinal cancers. T cells that was sensitized to HERV-H ENV (Xp22.3) had lytic effects against colorectal cancer expressing the ENV. HERV-E ENV showed similar effects in renal carcinoma [130, 176].

In addition, demethylating drugs are commonly used as anticancer agents and are known to trigger RE reexpression [177]. Interestingly, DNA methyl transferase inhibitors are caused by immune attacks that increase the expression of HERV and thereby increase the viral dsRNA [178]. Accordingly, individual knocking down of MDA5, MAVS, or IRF7 inhibits the ability of DNA methyl transferase inhibitor to target colorectal cancers resulting in significantly reduced the anticancer activity [179]. Altogether, immunotherapy approaches targeting HERV ENV in a broad spectrum of cancers might be valuable for the expansion of target cancers and for use with other cancer therapies.

6. Conclusions

In this chapter, we reviewed and summarized the functions and regulatory mechanisms of retroelements in the development and progression of cancers, and further presented applications in the development of diagnosis and treatment targets using these characteristics (**Table 1**). We looked at the retrovirus as a functional genomic element that forms the genome, not as an ancient infected virus and its useless remnants. Reactivation of retroelements means that it affects various regulation processes of cells beyond not only controlling the functions of surrounding genes but also increasing the protein formed therefrom or its function, or prompting its reinsertion into a new position. Because of this, it is very important to analyze and understand retroelements' functions with regard to various target substances, for example, miRNA, transcription factors, epigenetic modifiers, and so on (**Figure 1**).

RE type	Cancer type	Experimental technique	References
L1	Colon, breast, lung, ovarian, prostate cancer	Bisulfite-pyrosequencing	[45]
		MCA/CpG island microarray	
L1	Intrahepatic cholangiocarcinoma	Bisulfite-pyrosequencing	[48]
L1	Prostate cancer	Southern blot analysis	[49]
Alu	Colon cancer	Next-generation sequencing of unmethylated Alu	[50]
HERV-K	Breast cancer	RT-PCR, northern blot, in situ hybridization	[51]
HERV-K	Breast cancer	TCGA RNA-seq, RPPA data anaylsis	[52]
HERV-K	Melanoma	IHC, immunoblotting	[53]
		Cell fusion-dependent colony formation assay	
HERV-K	Kidney cancer	RT-PCR, northern blot	[54]
HERV-K	Head and neck cancer	Microarray	[34]
LI	Ovarian cancer	Southern hybridization, RT-PCR	[57]
HERV-W	Ovarian cancer	Southern hybridization, RT-PCR	[57]
L1	Breast cancer	Absolute quantitative assessment of methylated alleles (AQAMA) PCR	[58]
HERV-K	Urothelial carcinoma	RT-PCT, bisulfite-pyrosequencing	[60]
L1	Urothelial carcinoma	RT-PCT, bisulfite-pyrosequencing	[60]
HERV-K	Pancreatic cancer	RT-PCR, IHC, IF, ELISA, female immunodeficient nude	[61]
HERV-K	Prostate cancer	RT-PCR, bisulfite-pyrosequencing, ChIP	[64]
LI	Colorectal, blood, brain, prostate, ovarian cancer	Tea (TE analyzer) from paired-end, whole-genome sequencing	[96]
L1	11 types of cancer	Whole genome, exome sequencing	[94]
L1	Lung, brain cancer	L1-seq	[97]
L1	Liver cancer	Retrotransposon capture sequencing (RC-seq)	[95]
L1	Colon cancer	Southern blot, isolation of the fragment containing the insertion	[98]

RE type	Cancer type	Experimental technique	References
L1	Liver cancer	RC-seq, whole genome sequencing	[99]
L1	Encephalopathy	DNA-seq, RT-PCR	[116]
Alu	Multiple cancer cell lines	RNA immunoprecipitation, RT-PCR	[118]
ERV-9	Multiple cancer cell lines	RT-PCR, western blot, RNA immunoprecipitation, xenograft models	[122]
L1	Multiple cancer cell lines	IF, LINE-1 activation assay, RT-PCR	[124]
L1	Leukemia	ChIP-seq, RNA-seq	[125]
L1	Oral cancer	Bisulfite-pyrosequencing	[126]
HERV-W	Testicular cancer	HERV GeneChip microarray, bisulfite sequencing PCR	[133]
HERV-W	Endometrial cancer	RT-PCR, DNA-seq, immunoblot	[134]
L1	Breast, ovarian, pancreatic, lung, prostate cancer	Immunohistochemistry	[136]
L1	Colon, prostate cancer	Immunoblot, IF, IHC	[138]
HERV	Breast cancer	RT-PCR	[141]
HERV-K	Teratocarcinoma	CRISPR/Cas9, immunoblot	[144]
HERV-K	Breast cancer	GST pull-down assay, Co-IP	[145]
L1	Gastrointestinal cancer	Tea (TE analyzer) from paired-end whole-genome sequencing, somatic SNV, indel call, RNA-seq for TCGA	[149]
HERV	Multiple cancer types	SNV, DNA functional elements analysis	[150]
L1	Liver cancer	Bisulfite pyrosequencing	[154]
L1	NSCLC	Methylation-specific real-time PCR assay	[155]
L1	Colon cancer	Bisulfite pyrosequencing	[162]
L1	Breast cancer	Bisulfite sequencing, MSRED, and RT-PCR	[163]
HERV	Lung cancer	RT-PCR	[166]
HERV-K	Liver cancer	RT-PCR	[167]
HERV-H	Multiple cancer types	Immunohistochemistry	[168]
HERV-K	Breast cancer	ELISA, RT-PCR	[169]
L1	Multiple cancer types	Immunohistochemistry	[172]
L1	Breast cancer	Western blot, IHC	[173]
L1	Melanoma, prostate cancer	IF, Western blot, xenograft model	[175]
L1	Prostate cancer	RT activity assay, RT-PCR	[176]
L1	Melanoma	IF, RT-PCR, western blot, xenograft model	[178]

Table 1.
RE expression in human cancers.

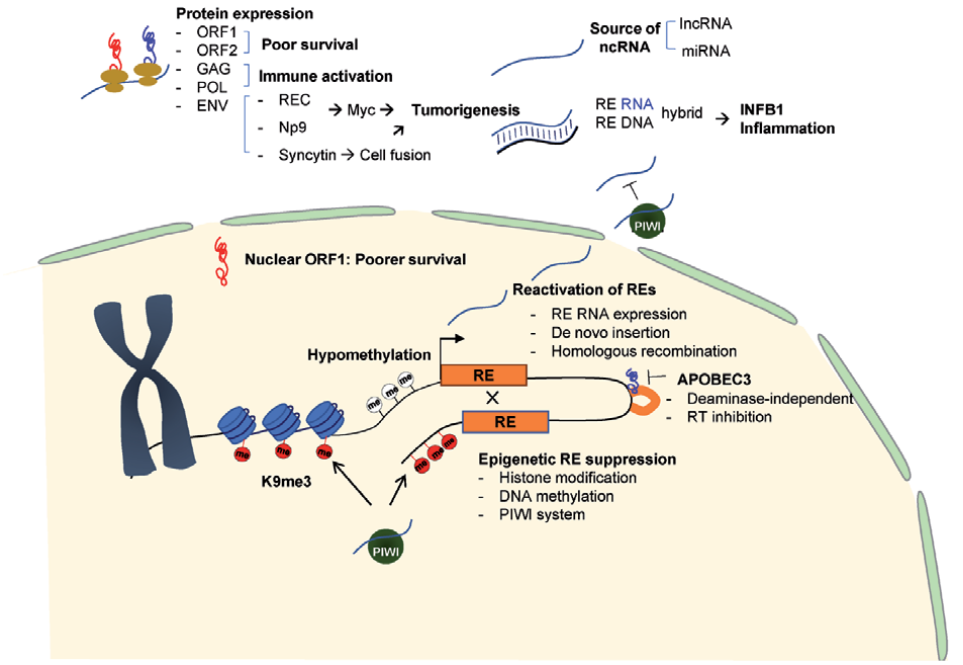


Figure 1.
Overall involvement of REs in cancer studies. RE expression was regulated by epigenomic controls such as histone modification and methylation. Reactivated RE by hypomethylation causes genome instability and the enrichment of cytoplasmic RE transcripts which may increase inflammatory signal. These may be involved in diverse biological process as a source of ncRNA including miRNAs. RE proteins are also involved in tumorigenesis process, and PIWI and APOBEC3 systems regulate RE activity in various ways.

Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF), funded by the Ministry of Science and ICT (#2016M3A9B6026771) and by (NRF-2019R1I1A1A01060265) at least partially. J.Y.C. conceived and developed the entire study and revised the chapter, and K.H.L. mainly wrote the first draft. We thank Hyeon-Ji Hwang for data acquisition, and Johannes Schabort for English editing.

Conflict of interest

The authors declare no conflict of interest.

Author details

Kang-Hoon Lee and Je-Yoel Cho*
Department of Biochemistry, BK21 PLUS Program for Creative Veterinary Science Research and Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul, South Korea

*Address all correspondence to: jeycho@snu.ac.kr

IntechOpen

References

[1] Collins FS et al. A vision for the future of genomics research. Nature. 2003;**422**(6934):835-847

[2] Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. Biochimica et Biophysica Acta. 2014;**1842**(10):1889-1895

[3] O'Donnell KA, Burns KH. Mobilizing diversity: Transposable element insertions in genetic variation and disease. Mobile DNA. 2010;**1**(1):21

[4] Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nature Reviews. Genetics. 2009;**10**(10):691-703

[5] Beck CR et al. LINE-1 retrotransposition activity in human genomes. Cell. 2010;**141**(7):1159-1170

[6] Richardson SR et al. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. Microbiology Spectrum. 2015;**3**(2) MDNA3-0061-2014

[7] Mason AL, Gilady SY, Mackey JR. Mouse mammary tumor virus in human breast cancer red herring or smoking gun? American Journal of Pathology. 2011;**179**(4):1588-1590

[8] Kassiotis G. Endogenous retroviruses and the development of cancer. Journal of Immunology. 2014;**192**(4):1343-1349

[9] Attermann AS et al. Human endogenous retroviruses and their implication for immunotherapeutics of cancer. Annals of Oncology. 2018;**29**(11):2183-2191

[10] Bannert N, Kurth R. Retroelements and the human genome: New perspectives on an old relation. Proceedings of the National Academy of Sciences of the United States of America. 2004;**101**(Suppl 2): 14572-14579

[11] Gonzalez-Cao M et al. Human endogenous retroviruses and cancer. Cancer Biology & Medicine. 2016;**13**(4):483-488

[12] Bannert N et al. HERVs new role in cancer: From accused perpetrators to cheerful protectors. Frontiers in Microbiology. 2018;**9**:178

[13] Yu HL, Zhao ZK, Zhu F. The role of human endogenous retroviral long terminal repeat sequences in human cancer (review). International Journal of Molecular Medicine. 2013;**32**(4):755-762

[14] Pfeifer GP. Defining driver DNA methylation changes in human cancer. International Journal of Molecular Sciences. 2018;**19**(4):1166

[15] Jansz N. DNA methylation dynamics at transposable elements in mammals. DNA Methylation. 2019;**63**(6):677-689

[16] Bourque G et al. Ten things you should know about transposable elements. Genome Biology. 2018;**19**:199

[17] Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genetics. 2014;**10**(5):e1004351

[18] Tokuyama M et al. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. Proceedings of the National Academy of Sciences of the United States of America. 2018;**115**(50):12565-12572

[19] Mita P, Boeke JD. How retrotransposons shape genome regulation. Current Opinion in Genetics & Development. 2016;**37**:90-100

[20] Casacuberta E, Gonzalez J. The impact of transposable elements in

environmental adaptation. Molecular Ecology. 2013;**22**(6):1503-1517

[21] Miousse IR et al. Response of transposable elements to environmental stressors. Mutation Research, Reviews in Mutation Research. 2015;**765**:19-39

[22] Jung YD et al. Retroelements: Molecular features and implications for disease. Genes & Genetic Systems. 2013;**88**(1):31-43

[23] Tam OH, Ostrow LW, Gale Hammell M. Diseases of the nervous system: Retrotransposon activity in neurodegenerative disease. Mobile DNA. 2019;**10**:32

[24] Szpakowski S et al. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. Gene. 2009;**448**(2):151-167

[25] Walter M et al. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife. 2016;**5**:e11418

[26] Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harbor Perspectives in Biology. 2014;**6**(5):a019133

[27] Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews. Genetics. 2010;**11**(3):204-220

[28] Adalsteinsson BT, Ferguson-Smith AC. Epigenetic control of the genome-lessons from genomic imprinting. Genes (Basel). 2014;**5**(3):635-655

[29] Leung D et al. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. Proceedings of the National Academy of Sciences of the United States of America. 2014;**111**(18):6690-6695

[30] Jones PA, Liang GN. OPINION rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics. 2009;**10**(11):805-811

[31] Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology. 2011;**411**(2):273-287

[32] Zeng Y, Chen TP. DNA methylation reprogramming during mammalian development. Genes. 2019;**10**(4):257

[33] Zamudio N et al. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes & Development. 2015;**29**(12):1256-1270

[34] Costello JF, Plass C. Methylation matters. Journal of Medical Genetics. 2001;**38**(5):285-303

[35] Estecio MRH et al. LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability. PLoS One. 2007;**2**(5):e399

[36] Lavasanifar A et al. Long interspersed nuclear element-1 mobilization as a target in cancer diagnostics, prognostics and therapeutics. Clinica Chimica Acta. 2019;**493**:52-62

[37] Cheung HH et al. DNA methylation of cancer genome. Birth Defects Research Part C-Embryo Today-Reviews. 2009;**87**(4):335-350

[38] Jeong S et al. Tumoral LINE-1 hypomethylation is associated with poor survival of patients with intrahepatic cholangiocarcinoma. BMC Cancer. 2017;**17**:588

[39] Schulz WA et al. Factor interaction analysis for chromosome 8 and DNA

methylation alterations highlights innate immune response suppression and cytoskeletal changes in prostate cancer. Molecular Cancer. 2007;**6**:14

[40] Jorda M et al. The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. Genome Research. 2017;**27**(1):118-132

[41] Wang-Johanning F et al. Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. Clinical Cancer Research. 2001;**7**(6):1553-1560

[42] Huang G et al. Human endogenous retroviral K element encodes fusogenic activity in melanoma cells. Journal of Carcinogenesis. 2013;**12**:5

[43] Cherkasova E et al. Detection of an immunogenic HERV-E envelope with selective expression in clear cell kidney cancer. Cancer Research. 2016;**76**(8):2177-2185

[44] Johanning GL et al. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. Scientific Reports. 2017;**7**:41960

[45] Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. Mobile DNA. 2016;**7**:24

[46] Zhang M, Liang JQ, Zheng S. Expressional activation and functional roles of human endogenous retroviruses in cancers. Reviews in Medical Virology. 2019;**29**(2):e2025

[47] Menendez L, Benigno BB, McDonald JF. L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. Molecular Cancer. 2004;**3**:12

[48] van Hoesel AQ et al. Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: A retrospective cohort study. Breast Cancer Research and Treatment. 2012;**134**(3):1103-1114

[49] Miousse IR, Koturbash I. The fine LINE: Methylation drawing the cancer landscape. BioMed Research International. 2015;**2015**:131547

[50] Kreimer U et al. HERV-K and LINE-1 DNA methylation and reexpression in urothelial carcinoma. Frontiers in Oncology. 2013;**3**:255

[51] Li M et al. Downregulation of human endogenous retrovirus type K (HERV-K) viral env RNA in pancreatic cancer cells decreases cell proliferation and tumor growth. Clinical Cancer Research. 2017;**23**(19):5892-5911

[52] Molaro A, Malik HS. Hide and seek: How chromatin-based pathways silence retroelements in the mammalian germline. Current Opinion in Genetics & Development. 2016;**37**:51-58

[53] Yang F, Wang PJ. Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. Seminars in Cell & Developmental Biology. 2016;**59**:118-125

[54] Goering W, Ribarska T, Schulz WA. Selective changes of retroelement expression in human prostate cancer. Carcinogenesis. 2011;**32**(10):1484-1492

[55] Sytnikova YA et al. Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in Drosophila ovarian cell cultures. Genome Research. 2014;**24**(12):1977-1990

[56] Ozata DM et al. PIWI-interacting RNAs: Small RNAs with big functions. Nature Reviews. Genetics. 2019;**20**(2):89-108

[57] Lim AK, Tao L, Kai T. piRNAs mediate posttranscriptional retroelement silencing and localization

to pi-bodies in the Drosophila germline. The Journal of Cell Biology. 2009;**186**(3):333-342

[58] Toth KF et al. The piRNA pathway guards the germline genome against transposable elements. Advances in Experimental Medicine and Biology. 2016;**886**:51-77

[59] Siomi MC et al. PIWI-interacting small RNAs: The vanguard of genome defence. Nature Reviews. Molecular Cell Biology. 2011;**12**(4):246-258

[60] Inoue K et al. Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. PLoS Genetics. 2017;**13**(7):e1006926

[61] Cheng Y et al. Emerging roles of piRNAs in cancer: Challenges and prospects. Aging (Albany NY). 2019;**11**(21):9932-9946

[62] Yu Y, Xiao J, Hann SS. The emerging roles of PIWI-interacting RNA in human cancers. Cancer Management and Research. 2019;**11**:5895-5909

[63] Maleki Dana P, Mansournia MA, Mirhashemi SM. PIWI-interacting RNAs: New biomarkers for diagnosis and treatment of breast cancer. Cell & Bioscience. 2020;**10**:44

[64] Liu Y et al. The emerging role of the piRNA/piwi complex in cancer. Molecular Cancer. 2019;**18**(1):123

[65] Chen Z et al. Stem cell protein Piwil1 endowed endometrial cancer cells with stem-like properties via inducing epithelial-mesenchymal transition. BMC Cancer. 2015;**15**:811

[66] Qu X et al. PIWIL2 promotes progression of non-small cell lung cancer by inducing CDK2 and Cyclin A expression. Journal of Translational Medicine. 2015;**13**:301

[67] Jiang L et al. Downregulation of Piwil3 suppresses cell proliferation, migration and invasion in gastric cancer. Cancer Biomarkers. 2017;**20**(4):499-509

[68] Wang Z et al. The role of PIWIL4, an Argonaute family protein, in breast cancer. The Journal of Biological Chemistry. 2016;**291**(20):10646-10658

[69] Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: An innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annual Review of Immunology. 2008;**26**:317-353

[70] Kinomoto M et al. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Research. 2007;**35**(9):2955-2964

[71] Liang W et al. APOBEC3DE inhibits LINE-1 Retrotransposition by interacting with ORF1p and influencing LINE reverse transcriptase activity. PLoS One. 2016;**11**(7):e0157220

[72] Choi J, Hwang SY, Ahn K. Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. Nucleic Acids Research. 2018;**46**(4):1912-1926

[73] Hu S et al. SAMHD1 inhibits LINE-1 retrotransposition by promoting stress granule formation. PLoS Genetics. 2015;**11**(7):e1005367

[74] Goodier JL et al. The broad-spectrum antiviral protein ZAP restricts human retrotransposition. PLoS Genetics. 2015;**11**(5):e1005252

[75] Kohnken R, Kodigepalli KM, Wu L. Regulation of deoxynucleotide metabolism in cancer: Novel mechanisms and therapeutic implications. Molecular Cancer. 2015;**14**:176

[76] Schneider AM et al. Roles of retrotransposons in benign and

malignant hematologic disease. Cell. 2009;**6**(2):121-145

[77] Crichton J et al. Defending the genome from the enemy within: Mechanisms of retrotransposon suppression in the mouse germline. Cellular and Molecular Life Sciences. 2014;**71**(9):1581-1605

[78] Huang CR, Burns KH, Boeke JD. Active transposition in genomes. Annual Review of Genetics. 2012;**46**:651-675

[79] Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annual Review of Genetics. 2007;**41**:331-368

[80] Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. Seminars in Cancer Biology. 2010;**20**(4):200-210

[81] Solyom S, Kazazian HH Jr. Mobile elements in the human genome: Implications for disease. Genome Medicine. 2012;**4**(2):12

[82] Yi K, Ju YS. Patterns and mechanisms of structural variations in human cancer. Experimental & Molecular Medicine. 2018;**50**(8):98

[83] Erwin JA et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. Nature Neuroscience. 2016;**19**(12):1583-1591

[84] Scott EC, Devine SE. The role of somatic L1 retrotransposition in human cancers. Viruses. 2017;**9**(6):131

[85] Helman E et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Research. 2014;**24**(7):1053-1063

[86] Shukla R et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 2013;**153**(1):101-111

[87] Lee E et al. Landscape of somatic retrotransposition in human cancers. Science. 2012;**337**(6097):967-971

[88] Iskow RC et al. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell. 2010;**141**(7):1253-1261

[89] Miki Y et al. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Research. 1992;**52**(3):643-645

[90] Schauer SN et al. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. Genome Research. 2018;**28**(5):639-653

[91] Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. Cancer Science. 2018;**109**(3):513-522

[92] Hancks DC, Kazazian HH Jr. Roles for retrotransposon insertions in human disease. Mobile DNA. 2016;**7**:9

[93] Cakmak Guner B et al. Detection of HERV-K6 and HERV-K11 transpositions in the human genome. Biomedical Reports. 2018;**9**(1):53-59

[94] Thayer RE, Singer MF, Fanning TG. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. Gene. 1993;**133**(2):273-277

[95] Baba Y et al. Long interspersed element-1 methylation level as a prognostic biomarker in gastrointestinal cancers. Digestion. 2018;**97**(1):26-30

[96] Piskareva O et al. The human L1 element: A potential biomarker in cancer prognosis, current status and future directions. Current Molecular Medicine. 2011;**11**(4):286-303

[97] Zhang W, Xu J. DNA methyltransferases and their roles in tumorigenesis. Biomarker Research. 2017;**5**:1

[98] Gaudet F et al. Induction of tumors in mice by genomic hypomethylation. Science. 2003;**300**(5618):489-492

[99] Lamprecht B, Bonifer C, Mathas S. Repeat-element driven activation of proto-oncogenes in human malignancies. Cell Cycle. 2010;**9**(21):4276-4281

[100] Veeck J, Esteller M. Breast cancer epigenetics: From DNA methylation to microRNAs. Journal of Mammary Gland Biology and Neoplasia. 2010;**15**(1):5-17

[101] Chishima T, Iwakiri J, Hamada M. Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs. Genes (Basel). 2018;**9**(1):23

[102] Tufarelli C, Cruickshanks HA, Meehan RR. LINE-1 activation and epigenetic silencing of suppressor genes in cancer: Causally related events? Mobile Genetic Elements. 2013;**3**(5):e26832

[103] Kukurba KR, Montgomery SB. RNA sequencing and analysis. Cold Spring Harbor Protocols. 2015;**2015**(11):951-969

[104] Zhang X et al. Non-coding RNAs and retroviruses. Retrovirology. 2018;**15**(1):20

[105] Kapusta A et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genetics. 2013;**9**(4):e1003470

[106] Sparber P et al. The role of long non-coding RNAs in the pathogenesis of hereditary diseases. BMC Medical Genomics. 2019;**12**(Suppl 2):42

[107] Cartault F et al. Mutation in a primate-conserved retrotransposon reveals a noncoding RNA as a mediator of infantile encephalopathy. Proceedings of the National Academy of Sciences of the United States of America. 2012;**109**(13):4980-4985

[108] Hadjiargyrou M, Delihas N. The intertwining of transposable elements and non-coding RNAs. International Journal of Molecular Sciences. 2013;**14**(7):13307-13328

[109] Negishi M et al. A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins. PLoS One. 2014;**9**(4):e95216

[110] Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biology. 2012;**13**(11):R107

[111] Pan Y et al. The emerging roles of long noncoding RNA ROR (lincRNA-ROR) and its possible mechanisms in human cancers. Cellular Physiology and Biochemistry. 2016;**40**(1-2):219-229

[112] Wang G et al. Regulation of proto-oncogene transcription, cell proliferation, and tumorigenesis in mice by PSF protein and a VL30 noncoding RNA. Proceedings of the National Academy of Sciences of the United States of America. 2009;**106**(39):16794-16798

[113] Xu L et al. A novel function of RNAs arising from the long terminal repeat of human endogenous retrovirus 9 in cell cycle arrest. Journal of Virology. 2013;**87**(1):25-36

[114] Stetson DB et al. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell. 2008;**134**(4):587-598

[115] Yu Q et al. Type I interferon controls propagation of long interspersed element-1. The Journal of Biological Chemistry. 2015;**290**(16):10191-10199

[116] Ishak CA et al. An RB-EZH2 complex mediates silencing of repetitive

DNA sequences. Molecular Cell. 2016;**64**(6):1074-1087

[117] Gasche JA et al. Interleukin-6 promotes tumorigenesis by altering DNA methylation in oral cancer cells. International Journal of Cancer. 2011;**129**(5):1053-1063

[118] Finnegan DJ. Retrotransposons. Current Biology. 2012;**22**(11):R432-R437

[119] Jiang F et al. Large-scale transcriptome analysis of retroelements in the migratory locust, *Locusta migratoria*. PLoS One. 2012;**7**(7):e40532

[120] Martin SL. The ORF1 protein encoded by LINE-1: Structure and function during L1 retrotransposition. Journal of Biomedicine & Biotechnology. 2006;**2006**(1):45621

[121] Garcia-Montojo M et al. Human endogenous retrovirus-K (HML-2): A comprehensive review. Critical Reviews in Microbiology. 2018;**44**(6):715-738

[122] Li F, Karlsson H. Expression and regulation of human endogenous retrovirus W elements. APMIS. 2016;**124**(1-2):52-66

[123] Soygur B, Sati L. The role of syncytins in human reproduction and reproductive organ cancers. Reproduction. 2016;**152**(5):R167-R178

[124] Gimenez J et al. Custom human endogenous retroviruses dedicated microarray identifies self-induced HERV-W family elements reactivated in testicular cancer upon methylation control. Nucleic Acids Research. 2010;**38**(7):2229-2246

[125] Strick R et al. Proliferation and cell-cell fusion of endometrial carcinoma are induced by the human endogenous retroviral Syncytin-1 and regulated by TGF-beta. Journal of Molecular Medicine (Berlin, Germany). 2007;**85**(1):23-38

[126] Burns KH. Transposable elements in cancer. Nature Reviews. Cancer. 2017;**17**(7):415-424

[127] Rodic N et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. The American Journal of Pathology. 2014;**184**(5):1280-1286

[128] Sur D et al. Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain. Mobile DNA. 2017;**8**:17

[129] De Luca C et al. Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. Oncotarget. 2016;**7**(4):4048-4061

[130] Grandi N, Tramontano E. HERV envelope proteins: Physiological role and pathogenic potential in cancer and autoimmunity. Frontiers in Microbiology. 2018;**9**:462

[131] Cegolon L et al. Human endogenous retroviruses and cancer prevention: Evidence and prospects. BMC Cancer. 2013;**13**:4

[132] Rhyu DW et al. Expression of human endogenous retrovirus env genes in the blood of breast cancer patients. International Journal of Molecular Sciences. 2014;**15**(6):9173-9183

[133] Cherkasova E, Weisman Q, Childs RW. Endogenous retroviruses as targets for antitumor immunity in renal cell cancer and other tumors. Frontiers in Oncology. 2013;**3**:243

[134] Xue B, He L. An expanding universe of the non-coding genome in cancer biology. Carcinogenesis. 2014;**35**(6):1209-1216

[135] Chan SM et al. The HERV-K accessory protein Np9 controls viability

and migration of teratocarcinoma cells. PLoS One. 2019;**14**(2):e0212970

[136] Denne M et al. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. Journal of Virology. 2007;**81**(11):5607-5616

[137] Baeissa H et al. Identification and analysis of mutational hotspots in oncogenes and tumour suppressors. Oncotarget. 2017;**8**(13):21290-21304

[138] Steranka JP et al. Transposon insertion profiling by sequencing (TIPseq) for mapping LINE-1 insertions in the human genome. Mobile DNA. 2019;**10**:8

[139] Ewing AD. Transposable element detection from whole genome sequence data. Mobile DNA. 2015;**6**:24

[140] Jung H, Choi JK, Lee EA. Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. Genome Research. 2018;**28**(8):1136-1146

[141] Chang TC et al. Investigation of somatic single nucleotide variations in human endogenous retrovirus elements and their potential association with cancer. PLoS One. 2019;**14**(4):e0213770

[142] Ehrlich M. DNA methylation in cancer: Too much, but also too little. Oncogene. 2002;**21**(35):5400-5413

[143] Kitkumthorn N, Mutirangura A. Long interspersed nuclear element-1 hypomethylation in cancer: Biology and clinical applications. Clinical Epigenetics. 2011;**2**:315-330

[144] Locke WJ et al. DNA methylation cancer biomarkers: Translation to the clinic. Frontiers in Genetics. 2019;**10**:1150

[145] Miyata T et al. Prognostic value of LINE-1 methylation level in 321 patients with primary liver cancer including hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Oncotarget. 2018;**9**(29):20795-20806

[146] Saito K et al. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. Clinical Cancer Research. 2010;**16**(8):2418-2426

[147] Barchitta M et al. LINE-1 hypomethylation in blood and tissue samples as an epigenetic marker for cancer risk: A systematic review and meta-analysis. PLoS One. 2014;**9**(10):e109478

[148] Bakshi A et al. DNA methylation variation of human-specific Alu repeats. Epigenetics. 2016;**11**(2):163-173

[149] Leontiou CA et al. Bisulfite conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. PLoS One. 2015;**10**(8):e0135058

[150] Kurdyukov S, Bullock M. DNA methylation analysis: Choosing the right method. Biology (Basel). 2016;**5**(1):3

[151] Shim J et al. Detection and quantification of methylation in DNA using solid-state Nanopores. Scientific Reports. 2013;**3**:1389

[152] Gonzalgo ML, Liang G. Methylation-sensitive single-nucleotide primer extension (Ms-SNuPE) for quantitative measurement of DNA methylation. Nature Protocols. 2007;**2**(8):1931-1936

[153] Irahara N et al. Precision of pyrosequencing assay to measure LINE-1 methylation in colon cancer, normal colonic mucosa, and peripheral blood cells. The Journal of Molecular Diagnostics. 2010;**12**(2):177-183

[154] Lee KH et al. Methylation of LINE-1 in cell-free DNA serves as a liquid biopsy biomarker for human breast cancers and dog mammary tumors. Scientific Reports. 2019;**9**(1):175

[155] Salvi S et al. Cell-free DNA as a diagnostic marker for cancer: Current insights. Oncotargets and Therapy. 2016;**9**:6549-6559

[156] Martinez VG et al. Epigenetics of bladder cancer: Where biomarkers and therapeutic targets meet. Frontiers in Genetics. 2019;**10**

[157] Zare M et al. Human endogenous retrovirus env genes: Potential blood biomarkers in lung cancer. Microbial Pathogenesis. 2018;**115**:189-193

[158] Ma WJ et al. Human endogenous retroviruses-K (HML-2) expression is correlated with prognosis and progress of hepatocellular carcinoma. BioMed Research International. 2016;**2016**:8201642

[159] Janakiram M et al. Expression, clinical significance, and receptor identification of the newest B7 family member HHLA2 protein. Clinical Cancer Research. 2015;**21**(10):2359-2366

[160] Wang-Johanning F et al. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. International Journal of Cancer. 2014;**134**(3):587-595

[161] Downey RF et al. Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? International Journal of Cancer. 2015;**137**(6):1249-1257

[162] Buscher K et al. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. Cancer Research. 2005;**65**(10):4172-4180

[163] Ardeljan D et al. The human long interspersed element-1 retrotransposon: An emerging biomarker of neoplasia. Clinical Chemistry. 2017;**63**(4):816-822

[164] Chen L et al. Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. Breast Cancer Research and Treatment. 2012;**136**(1):129-142

[165] Anwar SL, Wulaningsih W, Lehmann U. Transposable elements in human cancer: Causes and consequences of deregulation. International Journal of Molecular Sciences. 2017;**18**(5):974

[166] Sciamanna I et al. Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. Oncogene. 2005;**24**(24):3923-3931

[167] Carlini F et al. The reverse transcription inhibitor Abacavir shows anticancer activity in prostate cancer cell lines. PLoS One. 2010;**5**(12):e14221

[168] Hecht M et al. Efavirenz has the highest anti-proliferative effect of non-nucleoside reverse transcriptase inhibitors against pancreatic cancer cells. PLoS One. 2015;**10**(6):e0130277

[169] Oricchio E et al. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. Oncogene. 2007;**26**(29):4226-4233

[170] Sciamanna I, De Luca C, Spadafora C. The reverse transcriptase encoded by LINE-1 retrotransposons in the genesis, progression, and therapy of cancer. Frontiers in Chemistry. 2016;**4**:6

[171] Houede N et al. A phase II trial evaluating the efficacy and safety of efavirenz in metastatic castration-resistant prostate cancer. The Oncologist. 2014;**19**(12):1227-1228

[172] Sciamanna I et al. LINE-1-encoded reverse transcriptase as a target in

cancer therapy. Frontiers in Bioscience-Landmark. 2018;**23**:1360-1369

[173] Wang-Johanning F et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. Journal of the National Cancer Institute. 2012;**104**(3):189-210

[174] Hodges-Vazquez M et al. The yellow fever 17D vaccine and risk of malignant melanoma in the United States military. Vaccine. 2012;**30**(30):4476-4479

[175] Mastrangelo G et al. Does yellow fever 17D vaccine protect against melanoma? Vaccine. 2009;**27**(4):588-591

[176] Mullins CS, Linnebacher M. Endogenous retrovirus sequences as a novel class of tumor-specific antigens: An example of HERV-H env encoding strong CTL epitopes. Cancer Immunology, Immunotherapy. 2012;**61**(7):1093-1100

[177] Kong Y et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nature Communications. 2019;**10**:5228

[178] Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. Frontiers in Immunology. 2018;**9**:2039

[179] Roulois D et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell. 2015;**162**(5):961-973

Section 3

Genetics and Genomics in Clinical Practice

Chapter 6

Landscape Genetics: From Classic Molecular Markers to Genomics

Enéas Ricardo Konzen and Maria Imaculada Zucchi

Abstract

Landscape genetics combines population genetics and landscape ecology to understand processes that shape the distribution and organization of human, animal, or plant populations. This field of genetics emerged from the availability of several studies with classical molecular markers, such as isozymes, RAPD, AFLP, and microsatellites. Population genetic studies enabled the detection of population structure with those markers, but a more comprehensive analysis of natural populations was only possible with the development of statistical methods that combined both molecular data and environmental variables. Ultimately, the rapid development of sequencing technologies allowed studies at the genomic level, augmenting the resolution of association with environment factors. This chapter outlines basic concepts in landscape genetics, the main statistical methods used so far, and the perspectives of this field of knowledge into strategies for conservation of natural populations of plant and animal species. Moreover, we briefly describe the application of the field to understand historical human migration processes as well as how some diseases are spread throughout the world.

Keywords: molecular studies, environmental variables, population structure, genetic diversity, single nucleotide polymorphisms

1. Introduction

Population genetic studies deal with allele frequencies and processes that shape their variation within and among populations. Multiple studies have addressed genetic variation and their structure based on the screening of molecular markers such as allozymes (began with Lewontin and Hubby [1]), random amplified polymorphic DNA (RAPD) [2], amplified fragment length polymorphism (AFLP) [3], microsatellites or simple sequence repeats (SSR) [4], intersimple sequence repeats (ISSR) [5] and single nucleotide polymorphisms (SNP). The use of allozyme markers started up a series of population genetic studies, allowing relatively precise estimation of heterozygosity levels due to their codominance nature. Those markers were largely employed until the end of the 1990s. The development of techniques for screening directly at the level of DNA has accelerated the discovery of numberless markers in humans, animals, plants, fungi, and other organisms. RAPD, ISSR, and AFLP, in general, are more limited in describing genetic variation due to their dominance. In contrast, several SSR markers have been developed for studying a diverse set of species, enabling precise estimates of genetic diversity, gene flow, spatial genetic structure, paternity, linkage, and association mapping.

Ultimately, SNP markers have arisen as powerful markers for fine-scale genetic diversity, structure, and association mapping studies. The direct comparison among sequences of specific fragments generated by Sanger sequencing allowed the discovery of the first set of SNP. However, the revolution in sequencing technology of the last decade has provided numberless sequences for comparing individuals and deciphering population genetic mechanisms with high accuracy. The next-generation sequencing platforms generate millions of sequences that often result in thousands of SNP markers.

Nonetheless, the sole use of molecular data provides no definitive responses on evolutionary mechanisms operating in populations. An examination of the ecological factors, that drive the fate of individuals over generations or how current mechanisms impact in their adaptation or acclimation, is a much-needed task to better understand all species. Adequate statistical methods combining genetic and environmental variables are then necessary. Landscape genetics emerged as a field for the improvement of our understanding of the influence of geographical and environmental variables on the genetic structure of populations [6]. It diverges from the traditional basis of population genetics in the sense of more profound tests of the influence of landscape and environmental factors such as altitude, topography, and ground cover on population processes such as gene flow and population structure [7]. The rapid boost in genome-scale analyses also generated the terminology landscape genomics, as proposed by Joost et al. [8]. Landscape genomics differs from landscape genetics in the sense that it has become a powerful approach for scanning genes involved in complex adaptation mechanisms of species at populations and individual levels [9, 10].

This chapter is intended to provide brief concepts that cover the subject of landscape genetics and genomics. Furthermore, we outline potential applications of landscape genetic studies in the comprehension of adaptive traits of plants and animals and how such results may assist in the design of conservation strategies for endangered species. It is not our intent to provide an exhaustive panorama of landscape genetics studies so far, but rather contextualize concepts and applications with chosen case studies. Moreover, we briefly contextualize how landscape genetics is contributing in the comprehension of historical human migrations and the dispersion of human diseases.

2. Molecular markers and population structure studies

The most popular molecular markers employed in population genetic studies are SSR [4] and SNP. Simple sequence repeats are tandem repeated motifs with 1–6 bp [11] or up to 10 bp [12] with high frequency in genomes of all organisms. Plants commonly have AT-type repeats, whereas animals have the AC motif as the most common repeat unit [13]. High mutation rates are characteristics of microsatellite markers [12] providing markers with several alleles. SSR are codominant, hypervariable, and Mendelian inherited [14], which is implicated in high heterozygosity levels, increasing the discriminatory power among individuals and populations. Originally, SSR were developed from DNA libraries that required extensive laboratory work. Currently, however, the easiest way of discovering novel microsatellites if though direct sequencing of genomes and transcriptomes generated from NGS platforms [12]. With that available, SNP markers have actually been the most studied markers in recent years. SNP markers are the most abundant polymorphisms along plant and animal genomes. SNP consist on single base-pair changes present in the genome sequence that can occur as transitions or transversions, as nucleotide substitutions [15]. They can reach much higher density than all other types of markers in

genomes. Next-generation sequencing can generate large amounts of sequence data, enabling the detection of thousands of SNP [16].

Microsatellites and SNP markers are powerful tools for population genetic analyses. They have been extensively employed in studies with humans as well as animal and plant models and non-model species. The codominance and multiallelic nature of microsatellites make them suitable for estimating variables such as heterozygosity, inbreeding, gene flow, outcrossing rates, differentiation among populations and population structure [17]. SNP markers are generally employed for determining population structure as well, but with much higher density of markers and therefore genomic coverage to explain such subdivision. A series of studies have used SNP to dissect complex traits with QTL mapping and genome-wide association studies (GWAS) [15].

3. The concepts of landscape genetics and landscape genomics

Landscape genetics is concerned with testing the effects of landscape features on gene flow and genetic population structure. In general, the first studies of landscape genetics involved an exploratory phase, by geographically widespread sampling of populations and analysis of the effects of various landscape variables [18]. Landscape features or variables consist of any biotic, climatic, soil, or other conditions that comprise the habitat of organisms [6]. The population structure means the organization of genetic variation as influenced by a combination of evolutionary forces such as recombination, mutation, drift, natural selection, and historic demographic processes [19]. This leads to the idea that a group of subpopulations that exchange migrants in an occasional fashion are part of metapopulations [6].

The current status of genomic technologies allows the discovery of thousands of SNP markers, which has increased the resolution power for studying the association of environmental variables with specific genomics regions, also with a much deeper understanding of evolutionary processes. Genotyping-by-sequencing has enabled the discovery of SNP markers even in non-model species, which may lack a reference genome so far [20, 21]. This is where the concept of landscape genomics comes forward. Landscape genomics focuses on detecting candidate genes under selection as putative signals of local adaptation. The design of a landscape genomics experiment involves replicated sampling of environmental factors that might be driving selection, augmenting the resolution for detection of candidate loci under selection [10].

4. A briefing on statistical approaches in landscape genetics

In a landscape genetics study, two steps of analyses are normally required. The first involves the analysis of patterns of genetic variation. Next, such patterns are correlated with landscape variables based on statistical methods [22]. To test for association of environmental variables with genetic data, one of the simplest and commonly used methods is the Mantel's test, originally developed for identifying time-space clustering of diseases [23]. The test uses permutations to address the significance of the linear correlation coefficient between two pair-wise similarity or dissimilarity matrices [22]. One of the simplest examples of its application in landscape genetics is to correlate the genetic distances between individuals with their geographic location [24].

The methods for determining association of genetic data with environmental variables can be broadly categorized into approaches that deal with (i) pair-wise

landscape data and (ii) location-specific landscape data, as reviewed by Balkenhol et al. [22]. The development of methods in landscape genomics, however, expanded the range of tests for detecting loci under selection using genome scans, approaches for candidate gene discovery, QTL mapping and GWAS. Genome scans use two methods for detecting loci under selection, the differentiation outlier methods and the genetic-environmental association test, as reviewed by Storfer et al. [10]. Novel methods are continuously being developed, as more genomes are becoming sequenced or resequenced in populations.

5. Applications of landscape genetics

Several applications of landscape genetics or genomics can be described. We briefly account for case studies in plant and animal systems within this section. Moreover, a few examples of studies applied to humans are also given. In general, landscape genetics or genomics studies have provided association among geographic, abiotic, and biotic factors and genetic data provided by the screening of molecular markers in populations of diverse organisms. It has increased our power to detail inferences of movement and gene flow and potential adaptation to the landscape populations occur. However, studies for several organisms are still scarce or inexistent.

Cultivated crops such as maize, soybean, rice, and common bean were domesticated from wild progenitors which reflect their current adaptation to distinct environments. Landscape genomics studies have enabled a deeper understanding of processes shaping their distribution across multiple environments. Common bean (*Phaseolus vulgaris* L.) is an exceptional example of a widespread species original from America. Molecular data of wild germplasm identified two major gene pools, the Andean from Argentina to Colombia, and the Mesoamerican from Colombia to Mexico [25, 26]. A third smaller pool of wilds is also distinctive in a narrow area between Peru-Equador [27]. Microsatellites markers were broadly used to screen the genetic structure of wild and domesticated accessions of common bean (*Phaseolus vulgaris* L.), distinguishing from the broadest Andean and Mesoamerican gene pools to further subdivision within each one of them [25]. SNP markers from single fragments sequenced by Sanger also allowed an accurate distinction between Andean and Mesoamerican accessions, as well as their subdivisions [28]. The recognition of a parallel domestication event in each of the two major pools was also possible based on the detection of SNP markers in specific genomic regions of Andean and Mesoamerican genotypes [29]. Recent landscape genomics approaches enabled a more detailed description of the major events that determined the range expansion of *P. vulgaris* in America and how they were accompanied by environmental changes [26]. The climatic variability was also associated with differential drought adaptation and specific SNP markers were statistically related to root and shoot traits varying in a Mesoamerican panel of genotypes originated from regions with distinct precipitation regimes throughout the year [30].

Another application of landscape genomics concerns with the understanding of range expansion and ecological dominance of insect pests. The first step toward that is to know the population structure, gene flow and how natural selection is affecting adaptation. Zucchi et al. [31] described and addressed such problem by examining the population structure of *Piezodous guildiniis*, a soybean pest, in the United States and Brazil. A GBS-based set of SNP markers revealed genetic structure according to their geographic environment of origin. About 10% of loci were under positive selection, and their annotation revealed genes involved in genome

reorganization, neuropeptides, and energy mobilization [31]. Addressing such problem is to assist future endeavors at managing pest spreading in cultivated crops.

Another equally important questions addressed by landscape genomics are the consequences of climate change and human intervention to natural populations of wilds plants and animals. *Euterpe edulis* Martius is a palm species native to the Atlantic Rain Forest in Brazil, known as heart-of-palm [32]. The species is the list of endangered species to extinction [33]. Several studies have addressed the genetic diversity and structure of natural populations of this palm (for a compilation see [34]). Soares et al. [35] studied the genetic diversity and structure of remnant fragments of *E. edulis* in Bahia state and related the data to landscape metrics such as composition and configuration and local variables including the logging activity as human disturbance variable. No evidence of spatial genetic structure was detected, but distinct genetic clusters could be identified, suggesting a reduction in gene flow between the fragments of this study [35]. Natural populations located in other regions of Brazil, such as in Sao Paulo state, revealed to have high genetic diversity, as shown from microsatellite markers. Adjacent populations that have been generated though germplasm collection for management and cultivation showed similar genetic diversity. Those genetic materials could be used for recovering overexploited populations [36].

Landscape genetics studies with wild animals have been focused in recognizing their patterns of moving across their habitats. On terrestrial lands, landscape genetics of animals has particular features in comparison to aquatic environments or even to terrestrial plants. Landscape patterns interfere with organism behavior, thereby affecting mating and dispersal and reflecting on population processes [37].

6. Landscape genetics and human populations and diseases

Genomic technologies have also enabled studies to uncover historical human migrations and the genetic structure and diversity of human populations. For example, a genome-wide study of Malaysian ethnic groups using a SNP array revealed that humans from the peninsular area of Malaysia had higher genetic diversity, which the authors associated with a contact zone for recent human migrations in the Asian continent [38]. Such an example suggests the association between the genetic structure of human populations with geographic variables. In fact, Peter et al. [39] show that genetic differentiation generally tends to increase over higher geographic distances; however, distortions in those patterns also frequently occur. The human population structure, then, seems to be quite dynamic.

Landscape genetics also has been employed in epidemiological studies of human diseases. Statistical methods can be used in the identification of hotspot areas of disease movement [40]. This will have important implications in designing strategies for spread containment. One challenge, however, has been the application of landscape genetics methods in vector-borne diseases, which was reviewed by Hemming-Schroeder [40]. A few studies have been dedicated to such goal with human diseases. One interesting example is the correlation found between the genetic structure of *Aedes mcintochi*, a major vector for Rift Valley fever in Kenya, and mean precipitation values [41].

In 2020, one of the major global health issues concerns the new COVID-19. Sequencing technologies coupled with landscape genomics approaches have the potential to identify dispersal patterns of the virus in order to contain its spreading. Landscape genetic approaches have the power of assisting the decision-making process.

7. Concluding remarks

Climate change and human interference are no longer to be neglected on natural ecosystems. Among several fields of study devoted to deciphering the impact of these processes, landscape genetics will provide a better comprehension of the interaction between organisms and their environment of origin. The boost in sequencing technologies is enabling the study of the most diverse range of organisms. In fact, the Earth BioGenome Project is intended to sequence, catalog, and characterize all eukaryotic diversity in the forthcoming decade [42]. With that information available, resequencing to the level of population and their association with landscape variables will provide information for designing appropriate strategies for the conservation of endangered forms of life as well as any other species. The resequencing of several human genomes will also enable a better comprehension of the human population structure throughout the world and how the landscape shapes its organization. This has been and will be continuing valuable information to comprehending the dispersion of human diseases as well.

Conflict of interest

The authors declare no conflict of interest.

Author details

Enéas Ricardo Konzen[1*] and Maria Imaculada Zucchi[2,3]

1 Interdisciplinary Department, Center of Limnological, Coastal and Marine Studies, Federal University of Rio Grande do Sul, Imbé, Rio Grande do Sul, Brazil

2 Secretary of Agriculture and Supply of São Paulo State, Piracicaba, São Paulo, Brazil

3 Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

*Address all correspondence to: erkonzen@gmail.com

References

[1] Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics. 1966;**54**(2):595

[2] Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 1990;**18**(22):6531-6535

[3] Vos P, Hogers R, Bleeker M, Reijans M, Lee TV, Hornes M, et al. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research. 1995;**23**:4407-4414

[4] Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics. 1989;**44**(3):397

[5] Lagercrantz U, Ellegren H, Kakanuga T. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research. 1993;**21**:1111-1115

[6] Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: Combining landscape ecology and population genetics. Trends in Ecology & Evolution. 2003;**18**(4):189-197

[7] Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, et al. Putting the 'landscape' in landscape genetics. Heredity. 2007;**98**(3):128

[8] Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. Molecular Ecology. 2007;**16**(18):3955-3969

[9] Li Y, Zhang XX, Mao RL, Yang J, Miao CY, Li Z, et al. Ten years of landscape genomics: Challenges and opportunities. Frontiers in Plant Science. 2017;**8**:2136

[10] Storfer A, Patton A, Fraik AK. Navigating the interface between landscape genetics and landscape genomics. Frontiers in Genetics. 2018;**9**:68

[11] Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: An overview of the recent progress in plants. Euphytica. 2011;**177**(3):309-334

[12] Vieira MLC, Santini L, Diniz AL, Munhoz CDF. Microsatellite markers: What they mean and why they are so useful. Genetics and Molecular Biology. 2016;**39**(3):312-328

[13] Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, et al. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding. 1996;**2**(3):225-238

[14] Abdul-Muneer PM. Application of microsatellite markers in conservation genetics and fisheries management: Recent advances in population structure analysis and conservation strategies. Genetics Research International. 2014;**2014**:1-11

[15] Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. Biotechnology & Biotechnological Equipment. 2018;**32**(2):261-285

[16] Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, et al. Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in Elaeis guineensis. Frontiers in Plant Science. 2019;**10**:130

[17] Konzen ER. Towards conservation strategies for forest tree endangered species: The meaning of population genetic statistics. Advances in Forestry Science. 2014;**1**(1):45-51

[18] Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: Where are we now? Molecular Ecology. 2010;**19**(17):3496-3514

[19] Andam CP, Challagundla L, Azarian T, Hanage WP, Robinson DA. Population structure of pathogenic bacteria. In: Genetics and Evolution of Infectious Diseases. Amsterdam: Elsevier; 2017

[20] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;**6**(5):e19379

[21] Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 2012;**7**(2):e32253

[22] Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, et al. Identifying future research needs in landscape genetics: Where to from here? Landscape Ecology. 2009;**24**(4):455

[23] Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Research. 1967;**27** (2 part 1):209-220

[24] Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPDC, et al. Mantel test in population genetics. Genetics and Molecular Biology. 2013;**36**(4):475-485

[25] Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theoretical and Applied Genetics. 2009;**118**(5):979-992

[26] Ariani A, Mier B, Teran JC, Gepts P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. Molecular Biology and Evolution. 2017;**35**(1):119-131

[27] Gepts P, Osborn TC, Rashka K, Bliss FA. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for multiple centers of domestication. Economic Botany. 1986;**40**(4):451-468

[28] Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. Proceedings of the National Academy of Sciences. 2012;**109**(14):E788-E796

[29] Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. New Phytologist. 2013;**197**(1):300-313

[30] Teran JCBM, Konzen ER, Medina V, Palkovic A, Ariani A, Tsai SM, et al. Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.). Annals of Botany. 2019;**124**(6):917

[31] Zucchi MI, Cordeiro EM, Allen C, Novello M, Viana JPG, Brown PJ, et al. Patterns of genome-wide variation, population differentiation and SNP discovery of the red banded stink

bug (Piezodorus guildinii). Scientific Reports. 2019;**9**(1):1-11

[32] Reis MS, Guerra MP, Nodari RO, Ribeiro RJ, Reis A. Distribuição geográfica e situação atual das populações na área de ocorrência de Euterpe edulis Martius. Sellowia. 2000a;**49-52**:324-335

[33] Martinelli M, Moraes A. Livro Vermelho da Flora Do Brasil. Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro; 2013. p. 1100

[34] Konzen ER, Martins MP. Contrasting levels of genetic diversity among populations of the endangered tropical palm Euterpe edulis Martius. Cerne. 2017;**23**(1):31-42

[35] Soares LASS, Cazetta E, Santos LR, França DDS, Gaiotto FA. Anthropogenic disturbances eroding the genetic diversity of a threatened palm tree: A multi-scale approach. Frontiers in Genetics. 2019;**10**:1090

[36] Novello M, Viana JPG, Alves-Pereira A, de Aguiar Silvestre E, Nunes HF, Pinheiro JB, et al. Genetic conservation of a threatened Neotropical palm through community-management of fruits in agroforests and second-growth forests. Forest Ecology and Management. 2018;**407**:200-209

[37] Waits LP, Cushman SA, Spear SF. Applications of landscape genetics to connectivity research in terrestrial animals [chapter 12]. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape Genetics: Concepts, Methods, Applications. 1st ed. West Sussex: John Wiley and Sons Ltd; 2016. pp. 199-219

[38] Deng L, Hoh BP, Lu D, Fu R, Phipps ME, Li S, et al. The population genomic landscape of human genetic structure, admixture history and local adaptation in peninsular Malaysia. Human Genetics. 2014;**133**(9):1169-1185

[39] Peter BM, Petkova D, Novembre J. Genetic landscapes reveal how human genetic diversity aligns with geography. Molecular Biology and Evolution. 2019;**37**(4):943-951

[40] Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G. Landscape genetics: A toolbox for studying vector-borne diseases. Frontiers in Ecology and Evolution. 2018;**6**:21

[41] Campbell LP, Alexander AM. Landscape genetics of *Aedes mcintoshi* (Diptera: Culicidae), an important vector of rift valley fever virus in northeastern Kenya. Journal of Medical Entomology. 2017;**54**(5):1258-1265

[42] Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences. 2018;**115**(17):4325-4333

Chapter 7

Clinical Validation of a Whole Exome Sequencing Pipeline

Debra O. Prosser, Indu Raja, Kelly Kolkiewicz, Antonio Milano and Donald Roy Love

Abstract

Establishing whole exome sequencing (WES) in an accredited clinical diagnostic space is challenging. The validation (as opposed to verification) of an approach that will lead to clinical reports requires adhering to international guidelines and recommendations and developing a robust analytical pipeline that can scale due to the increasing clinical demand for comprehensive gene screening. This chapter will present a step-wise approach to WES validation that any laboratory can follow. The focus will be on highlighting the pivotal technical issues that must be addressed in validating WES and the analytical tools and QC metrics that must be considered before implementing WES in a clinical environment.

Keywords: whole exome sequencing, next-generation sequencing, validation, bioinformatics, diagnostics

1. Introduction

The decision as to which type of genetic test should be implemented by a clinical laboratory is largely driven by the type of referrals received by the laboratory and the complexity of patients' clinical phenotypes. In the main, testing has advanced from single-gene to multi-gene panels in which next-generation sequencing (NGS) has offered the technical means of undertaking this approach at low cost and high throughput. However, with the increasing awareness of genetic heterogeneity combined with gene discovery, whole exome sequencing (WES) offers laboratories a more streamlined approach. By implementing a single wet-work pipeline of exome capture coupled with the ability to analyze a virtual gene panel or report on the whole exome, laboratories can perform NGS in a more efficient manner.

Since the inception of NGS over a decade ago, multiple recommendations and guidelines have been published for NGS [1–3]. Using these guidelines, the College of American Pathologists (CAP) and Association for Molecular Pathology (AMP) published their Practical Framework for Designing and Implementing NGS Tests for Inherited Disorders in 2019 [4], and this is available through the CAP website (https://www.cap.org/member-resources/precision-medicine/next-generation-sequencing-ngs-worksheets).

We adopted this framework to establish a diagnostic NGS service using whole exome sequencing as our capture procedure and analyzing virtual gene panels or WES for reporting purposes.

The framework provides guidance and editable worksheets for the five steps involved in test establishment and validation.

1. Test design: setup

2. Assay design and optimization

3. Test validation

4. Quality management

5. Bioinformatics and IT

Throughout the validation process, it is essential that the NGS workflow is informed by the real-world local environment in which clinical testing will be performed.

2. Test design: setup

In view of the diverse range of referrals made to the authors' genetics laboratory (serving the needs of a 400-bed women and children's hospital in the Middle East), a whole exome capture solution was chosen for library preparation. The principal motivation behind this determination was to achieve an efficient workflow that would allow appropriate batching coupled with a time-limited turnaround time (TAT) for all referrals.

The limited number of staff in the authors' laboratory demanded a WES workflow that could be easily automated, twinned with a data analysis package that would allow secure remote access with a strong databasing function. The whole exome solution capture by SOPHiA™ Genetics was chosen for library preparation. This platform allows for the analysis of WES, clinical exome sequencing (CES) and clinical gene panels, together with the identification of single-nucleotide variants (SNVs) and copy number variants (CNVs) using SOPHiA™ DDM software.

3. Assay design and optimization

The validation pipeline needs to be grounded from the beginning in terms of the requirements of the test, which must take into account the sample types the laboratory will receive and the parameters that need to be satisfied (see **Table 1**).

Routinely, whole blood samples collected in EDTA are received by the authors' laboratory for testing. Therefore, our validation focused only on genomic DNA extracted from whole blood using our standard methods. The baseline validation of the WES data required the inclusion of two HapMap gDNA samples: the NIST control (NA12878) and the commercial control (SG063) supplied by SOPHiA™ Genetics.

The WES capture by SOPHiA™ Genetics was used for library preparation following all the steps as set out by the automated WES 32 reaction protocol. For instrumentation, our validation was restricted to automated library preparation using the PE Sciclone® G3 NGS workstation and sequencing using the Illumina® HiSeq4000 platform.

A critical additional consideration was the need for copy number variant calls to be made. This required a minimum batch number of eight patients and high coverage requirements, which involved restricting the number of samples per Illumina® HiSeq4000 lane to one pool of eight patients.

Test requirements	Must have	Nice to have
WES		Y
CES	Y	
Clinical panels	Y	
CNV detection	Y	
Necessary sample throughput per month	16	32
How deeply does each position need to be covered for accurate variant calling (if known—otherwise address during test optimization)	>20x	>50x
DNA from whole blood collected in EDTA	Y	
DNA from external/commercial sources (limitations)	Y	
Required/expected TAT	3 months	2 months
Combine different tests (existing or planned) within a sequencing run	Y	

WES, whole exome sequencing; CES, clinical exome sequencing; CNV, copy number variant; TAT, turnaround time.

Table 1.
Test requirements and limitations.

Importantly, the naming of the sequence files (.bam,. FASTQ, etc.) should be considered during the early phase of test design and validation. File conventions that are used for the bioinformatic process may be limited in terms of the type of special characters and/or character length. Following recommendations in the CAP/AMP-Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines [5], the identity of the sample must be preserved throughout all steps of the bioinformatic pipeline. These authors recommend the following four unique identifiers that should be applied to the sample file name:

i. Unique sample identifier

ii. Unique patient identifier

iii. Unique run identifier

iv. Laboratory location identifier

It is essential that the file naming convention that is decided upon for validation adheres to the above recommendations and can be universally implemented for all subsequent testing.

4. Test validation

Test validation mandates a need for accuracy, precision and stability. These assessments must be made in the context of expected clinical workloads and performance. For the authors' laboratory, the sample batch size was set at 16 samples per validation batch and a total of three validation runs performed over differing days with differing technologists.

Analytical performance was characterized by the assessment of precision, sensitivity and concordance of variant calls against previously validated data.

Inter-run and intra-run data were achieved by replicate analysis of two HapMap gDNAs, the NIST sample, NA12878, and the commercial control supplied by SOPHiA™ Genetics, SG063, as well as four well-characterized clinical samples previously reported by accredited laboratories. The remaining samples included a representative group of the clinical samples received by the authors' laboratory (see **Table 2**).

The complete NGS workflow should be included in the validation, from library preparation to bioinformatic analysis to report generation, which is highlighted below.

- Sample collection and DNA extraction. Genomic DNA is extracted and purified from blood samples using either the Gentra® PureGene® DNA Blood Mini Kit or the QIAsymphony® DSP DNA Midi kit (QIAGEN, Hilden, Germany). DNA quality is initially assessed by NanoDrop™ spectrophotometry.

  Genomic DNA preparation. The initial preparation of gDNA used in NGS library preparation is the most critical step in the NGS workflow, and the care and time taken here are key to successful library amplification and sequencing.

  High-quality gDNA can be by quantified using a Qubit™ fluorometer followed by sequential dilution with further quantification to the desired input concentration. It is essential to minimize pipetting gDNA volumes of less than 5 µl for dilution. In our study, gDNA is prepared to a working concentration of 40 ng/µl. After Qubit™ quantification, the integrity of the gDNA can be analyzed using an Agilent TapeStation 4200. Samples with a DNA integrity number (DIN) of greater than 7.5 can proceed to WES capture.

- Library preparation, targeted capture and sequencing. Whole exome sequencing was performed according to the SOPHiA™ Whole Exome Solution 32 Samples User Guide, in combination with the SOPHiA™ Library Preparation and Capture User Guide—automation with PerkinElmer Sciclone® G3 NGS workstation. Each validation run consists of 16 samples that are divided into 2 pools of 8 samples each, as shown in the validation grid in **Table 3**.

  The SOPHiA™ WES protocol for library construction subjects genomic DNA (200 ng) to enzymatic fragmentation, end repair and A-tailing. All these steps occur using a Sciclone® G3 NGS workstation. The adapter-ligated DNA is then amplified in a limited way via an eight-cycle PCR protocol.

  Post-amplification cleanup of the libraries is carried out using the Sciclone® G3 NGS workstation, and libraries are prepared for quantitation with a dilution factor of 4.

  Amplified libraries are analyzed using Qubit™ fluorometer and Agilent TapeStation 4200 to assess the quantity and quality of each individual library. Library DNA fragments should have a size distribution between 300 and 700 bp. Genomic DNA that has been fragmented, end repaired, A-tailed and adapter-ligated can then be considered library DNA, which is ready for pooling and then hybridization and capture. In the case of the SOPHiA™ WES protocol, eight samples are pooled (200 ng of each library) per capture.

  Prepared pools are hybridized for 4 h followed by post-capture amplification and cleanup on the Sciclone® G3 NGS workstation.

  Final library quantification is performed for each captured library pool using a Qubit™ fluorometer and Agilent TapeStation 4200. Subsequent pools are

Sample ID	Description	Purpose	Purpose (detail)	Specific variant/s of interest	Variant type	Measured metric
VAL-1	NA12878	Baseline validation	N/A	N/A	N/A	Intra-run variability Inter-run variability
VAL-2	SG063	Baseline validation	N/A	N/A	N/A	Intra-run variability Inter-run variability
VAL-3	Anonymized patient specimen	Baseline validation	Variant type	Ciliopathy gene panel CCDC39:c.2017G > T p.(Glu673*) CCDC39: Deletion of exons 14 to 20	SNV CNV	Inter-run variability Sensitivity
VAL-4	Anonymized patient specimen	Baseline validation	Variant type prevalent in gene	Single-gene analysis CFTR:c.1521_1523delCTT p.(Phe508del)	DEL	Inter-run variability Sensitivity
VAL-5	Anonymized patient specimen	Baseline validation	Variant type	Craniosynostosis gene panel CACNA1H:c.4318_4319delinsGC p.(Phe1440Ala)	DELINS	Inter-run variability Sensitivity
VAL-6	Anonymized patient specimen	Baseline validation	Variant type prevalent in gene	Tuberous sclerosis gene panel TSC2: Deletion of exons 2 to 16	CNV	Inter-run variability Sensitivity
VAL-7	Anonymized patient specimen	Gene-specific validation	Variant type	Arrhythmia cardiomyopathy gene panel SCN5A:c.4867C > T p.(Arg1623*)	SNV (stop)	Sensitivity
VAL-8	Anonymized patient specimen	Gene-specific validation	Variant type	Custom panel of 196 genes 200 genomic co-ordinates	SNV DEL/ DUP	Sensitivity
VAL-9	Anonymized patient specimen	Gene-specific validation	Variant type	Paroxysmal Dystonia gene panel Del 16p11.2 chr16:29,656,684-30,190,568	CNV	Sensitivity
VAL-10	Anonymized patient specimen	Gene-specific validation	Variant type	Leukodystrophy gene panel MLC1:c.908_918delinsGCA p.(Val303Glyfs*96)	DELINS	Sensitivity
VAL-11	Anonymized patient specimen	Gene-specific validation	Variant type	Epilepsy gene panel WWOX: Deletion of exons 1–5	CNV	Sensitivity
VAL-12	Anonymized patient specimen	Gene-specific validation	Variant range	Epilepsy gene panel	SNV DEL/ DUP	Sensitivity
VAL-13	Anonymized patient specimen	Gene-specific validation	Variant type	Single-gene analysis CFTR: deletion of exons 4–8	CNV	Sensitivity

Sample ID	Description	Purpose	Purpose (detail)	Specific variant/s of interest	Variant type	Measured metric
VAL-14	Anonymized patient specimen	Gene-specific validation	Variant range	Neuropathy gene panel	SNV DEL/ DUP	Sensitivity
VAL-15	Anonymized patient specimen	Gene-specific validation	Variant range	Cholestasis gene panel	SNV DEL/ DUP	Sensitivity
VAL-16	Anonymized patient specimen	Gene-specific validation	Variant type	Tuberous sclerosis gene panel (2 genes) TSC2:c.5238_5255del p.(His1746_Arg1751del)	DEL	Sensitivity
VAL-17	Anonymized patient specimen	Chromosomal CNV validation	Variant type	Molecular karyotype referral Dup 22q11.21 chr22:18,661,724-21,809,099	CNV	Sensitivity
VAL-18	Anonymized patient specimen	Gene-specific validation	Variant range	Primary ciliary dyskinesia gene panel DNAH5: Gain of exons 1 to 50 DNAH5:c.5503C > T p.(Gln1835*)	SNV CNV	Sensitivity
VAL-19	Anonymized patient specimen	Gene-specific validation (pseudogene)	Variant range	Inherited cancer gene panel CDKN2A:c.9_32dup p.(Ala4_Pro11dup)	SNV DEL	Sensitivity
VAL-20	Anonymized patient specimen	Gene-specific validation	Variant range	Custom panel of 196 genes 200 genomic coordinates	SNV DEL/ DUP	Blind analysis
VAL-21	Anonymized patient specimen	Chromosomal CNV validation	Variant type	Molecular karyotype referral Duplication at 16p13.11, deletion at 12p31 and duplication at Xp21.1	CNV	Sensitivity
VAL-22	Anonymized patient specimen	Gene-specific validation	Variant type prevalent in gene	Single-gene analysis DMD: duplication exons 45–62	CNV	Sensitivity
VAL-23	Anonymized patient specimen	Gene-specific validation	Variant type prevalent in gene	Dystrophinopathy gene panel DMD: deletion of exons 8–34	CNV	Sensitivity
VAL-24	Anonymized patient specimen	Gene-specific validation	Variant range	Custom panel of 196 genes 200 genomic co-ordinates	SNV DEL/ DUP	Sensitivity
VAL-25	Anonymized patient specimen	Gene-specific validation (pseudogene)	Pseudogene	Custom panel of nine genes	SNV DEL/ DUP	Sensitivity

Sample ID	Description	Purpose	Purpose (detail)	Specific variant/s of interest	Variant type	Measured metric
VAL-26	Anonymized patient specimen	Gene-specific validation	Variant type	Primary Immunodeficiency gene panel TBX1:c.1383_1421del p.(Ala464_Ala476del)	DEL	Sensitivity
VAL-27	Anonymized patient specimen	Gene-specific validation	Variant type	Dilated cardiomyopathy gene panel TTN:c.75984_75985insTACCA p.(Ala25329Tyrfs*32)	INS	Sensitivity
VAL-28	Anonymized patient specimen	Gene-specific validation	Variant type	Pediatric cancer gene panel SMARCB1:c.159_160delinsTATCTGGAGGCG (p.Leu54Ilefs*20)	DELINS	Sensitivity

DEL, deletion; INS, insertion; DUP, duplication; SNV, single-nucleotide variant; CNV, copy number variant.

Table 2.
Sample list.

		Run 001			Run 002			Run 003	
Pool A	**A**	VAL-1 NA12878	VAL-4	**A**	VAL-5	**VAL-13**	**A**	**VAL-21**	VAL-5
	B	**VAL-3**	VAL-10	**B**	VAL-2 SG-063	VAL-15	**B**	VAL-28	VAL-1 NA12878
	C	**VAL-11**	VAL-1 NA12878	**C**	**VAL-17**	VAL-2 SG-063	**C**	VAL-1 NA12878	VAL-24
	D	VAL-2 SG-063	VAL-12	**D**	VAL-16	VAL-19	**D**	**VAL-3**	VAL-25
Pool B	**E**	VAL-1 NA12878	**VAL-6**	**E**	VAL-2 SG-063	VAL-20	**E**	**VAL-22**	**VAL-6**
	F	VAL-7	VAL-8	**F**	**VAL-6**	**VAL-3**	**F**	**VAL-23**	VAL-2 SG-063
	G	VAL-2 SG-063	**VAL-9**	**G**	VAL-14	VAL-1 NA12878	**G**	VAL-4	VAL-27
	H	VAL-5	VAL-2 SG-063	**H**	**VAL-18**	VAL-4	**H**	VAL-1 NA12878	VAL-26

Copy number variant (CNV) samples are indicated in bold.

Table 3.
Validation grid.

diluted to 20 nM (in a total volume of 20 µl) and subjected to sequencing using an Illumina® HiSeq4000 Sequencing platform.

- Sequence analysis: performance metrics. Baseline performance metrics for the WES validation study must involve the analysis of well-characterized reference samples: the NIST sample (NA12878) and the SOPHiA™ Genetics control SG063. The sequence metrics for each sample in the run must be recorded and averages established using the reference samples. Samples must meet the sequencing metrics shown in **Table 4** in order to reach the threshold for clinical reporting.

  Analytical sensitivity and specificity must be calculated separately for each variant type (SNV, indel, CNV, etc.). Additional runs may be required to meet acceptable confidence intervals for less frequent variant types of insertions and deletions. For 95% confidence and 95% reliability, 59 variants of each type (and insertion/deletion range) should be analyzed [5]. The variant types that do not have strong confidence intervals must be listed in the test limitations of the clinical report until such time that the desired confidence levels have been achieved.

Selected sequencing metrics	Must have	Nice to have
Q30 score	>80	>85
Total number of reads per sample	>70 M	80–100 M
Percentage of mapped reads	>80%	>85%
Total percentage on-target reads	>90%	>95%
Coverage 10% quantile (at this depth, 90% target covered)	20x	50x

Table 4.
Sequencing metrics.

5. Quality management

The worksheets described by Santani et al. [4] set out very clear guidance for all quality aspects that need to be taken into consideration for the test to meet CAP requirements [4]. Through a validation study, the majority of a test's limitations will be discovered and can be recorded against the QC parameters. **Table 5** summarizes quality metrics that need to be addressed.

Section	Category	Criteria	Specific requirement Note that these may vary between tests and laboratories
Pre-analytical QC (per sample)	Specimen quality	Wrong specimen type	Whole blood
		Wrong type of tube	Purple top EDTA tube
		Insufficient quantity	≥0.5 ml
		Clotting (blood only)	No visible clots
		Insufficient labelling	Labelling contains name, DOB, barcode, date of collection
		Expired specimen	≤7 days since collection
		Expired collection tube	Collection tube not expired
	DNA quality and quantity	OD 260/280 ratio	>1.7
		Electrophoretic analysis	Shows intact high molecular weight DNA band
		Quantification	≥500 ng
		DNA integrity number (DIN)	>7.5
Analytical QC (per instrument run)	Instrument run QC	Cluster density	Not taken into account
		Base quality	Q30 ≥ 80
	Pipeline QC	Total reads passing filter	>280 M per lane
		% reads not assigned to any sample	<5%
	Control samples	Positive control	Expected variants found
Analytical QC (per sample)	Library preparation	Fragment size and distribution	>80% of fragments between 300 and 700 bp
		Pooled library concentration	>20 nM
	Sample de-multiplexing	% reads assigned to sample	8–12%
	Read alignment	% Reads aligned to target	>90%
		Distribution of coverage	>95% within 25–200×
		Coverage 10% quantile (at this depth 90% target covered at x)	>40×
		PCR duplicates	<20%
	Specimen identity	Accurate specimen identity, file names with 4 points of identification	All worksheets and transfers during bench work are witness checked for accurate specimen identification
	Data transfer Integrity	Data transfer to secure analysis platform	

Table 5.
Quality management.

	Considerations	Resources	Links
Gene selection	Clinical association	ClinGen	https://clinicalgenome.org/curation-activities/gene-disease-validity/ https://search.clinicalgenome.org/kb/gene-validity
		GeneReviews	https://www.ncbi.nlm.nih.gov/books/NBK1116/
Gene analysis	Appropriate transcripts	LRG RefSeq	https://www.lrg-sequence.org/ https://www.ncbi.nlm.nih.gov/refseq/rsg/
	Pseudogenes	Pseudogene PanelApp – Genes and Entities	http://pseudogene.org/ https://panelapp.genomicsengland.co.uk/panels/entities/?tag=locus-type-pseudogene
	Evaluated homopolymeric regions	Ivády et al. [6]	DOI: 10.1186/s12864-018-4544-x
	Mutation spectrum—reported deep intronic and/or promoter region variants	PanelApp—Genes and Entities	https://panelapp.genomicsengland.co.uk/panels/entities/?
	CNV analysis	ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
		Decipher	https://decipher.sanger.ac.uk/
	Establish if critical variants are not covered by assay		
Virtual panel creation	Expert reviewed panels	PanelApp	https://panelapp.genomicsengland.co.uk/
		ClinGen	https://www.clinicalgenome.org/data-sharing/clinvar/
	Phenotype-driven	HPO	https://hpo.jax.org/app/

Table 6.
Considerations for gene selection, analysis and virtual panel creation.

6. Bioinformatics and IT

To assess accuracy, genetic variants must be compared against publicly available reference data obtained from 1000 Genomes Project.

Clinical association, gene validity and mutation spectrum are applied to the creation of virtual gene panels in order to aid variant interpretation and reporting. The considerations associated with constructing virtual gene panels and the analysis of variants are shown in **Table 6**.

7. Conclusions

The decision to implement WES in a clinical diagnostic environment is one that must take into account local context, which encompasses clinical complexity, staff resources, equipment resources and bioinformatic expertise. The decisions described here were made based on the above considerations with a view to establishing opportunity, the most important of which was to have a WES pipeline that could scale over time in terms of patients tested and with the potential to be a regional resource.

It should be stressed, however, that a WES pipeline is sandwiched by two critical elements: first, the need to focus on the quality and accurate quantitation of genomic DNA; which dictates the quality of everything that happens downstream, and second, to understand that the identification of DNA variants is technically demanding but the classification of those variants is not currently a fully automated process. The former can sometimes be overlooked, while the latter can be a daunting exercise. It is perhaps the subject of another book chapter to discuss the approaches to variant classification.

Conflicts of interest

The authors declare no conflicts of interest.

Thanks

The authors wish to thank Mr. Duncan Kay of Custom Science (NZ) for his generous suggestions regarding commercial providers for WES data analysis and Javier Botet of Sophia Genetics for his advice regarding quality management considerations.

Author details

Debra O. Prosser, Indu Raja, Kelly Kolkiewicz, Antonio Milano
and Donald Roy Love*
Division of Pathology Genetics, Department of Pathology, Sidra Medicine, Doha, Qatar

*Address all correspondence to: dlove@sidra.org

References

[1] Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genetics in Medicine. 2013;**15**(9):733-747. DOI: 10.1038/gim.2013.92

[2] Hegde M, Santani A, Mao R, Ferreira-Gonzalez A, Weck KE, Voelkerding KV. Development and validation of clinical whole-exome and whole-genome sequencing for detection of germline variants in inherited disease. Archives of Pathology & Laboratory Medicine. 2017;**141**:798-805. DOI: 10.5858/arpa.2016-0622-RA

[3] Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. European Journal of Human Genetics. 2016;**24**:2-5. DOI: 10.1038/ejhg.2015.226

[4] Santani A, Simen BB, Briggs M, Lebo M, Merker JD, Nikiforova M, et al. Designing and implementing NGS tests for inherited disorders a practical framework with step-by-step guidance for clinical laboratories. The Journal of Molecular Diagnostics. 2019;**21**:369-374. DOI: 10.1016/j.jmoldx.2018.11.004

[5] Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. The Journal of Molecular Diagnostics. 2018;**20**(1):4-27. DOI: 10.1016/j.jmoldx.2017.11.003

[6] Ivády G, Madar L, Dzsudzsák E, Koczok K, Kappelmayer J, Krulisova V, et al. Analytical parameters and validation of homopolymer detection in a pyrosequencing-based next generation sequencing system. BioMed Central Genomics. 2018;**19**:158. DOI: 10.1186/s12864-018-4544-x

Chapter 8

Integrating Evolutionary Genetics to Medical Genomics: Evolutionary Approaches to Investigate Disease-Causing Variants

Ugur Sezerman, Tugce Bozkurt and Fatma Sadife Isleyen

Abstract

In recent years, next-generation sequencing (NGS) platforms that facilitate generation of a vast amount of genomic variation data have become widely used for diagnostic purposes in medicine. However, identifying the potential effects of the variations and their association with a particular disease phenotype is the main challenge in this field. Several strategies are used to discover the causative mutations among hundreds of variants of uncertain significance. Incorporating information from healthy population databases, other organisms' databases, and computational prediction tools are evolution-based strategies that give valuable insight to interpret the variant pathogenicity. In this chapter, we first provide an overview of NGS analysis workflow. Then, we review how evolutionary principles can be integrated into the prioritization schemes of analyzed variants. Finally, we present an example of a real-life case where the use of evolutionary genetics information facilitated the discovery of disease-causing variants in medical genomics.

Keywords: genomics, evolution, variant discovery

1. Introduction

NGS technologies can be integrated into medical diagnostics in several ways which vary in the number and type of sequenced regions. While targeted tests include sequencing particular disease-specific genes, sequencing all ~20,000 protein-coding genes by Whole-exome sequencing (WES) and entire genomes by Whole-genome Sequencing (WGS) are non-targeted approaches. These sequencing approaches are precise ways to detect genetic variation of a patient and in relation to a healthy population or healthy reference genome. However, sequencing-based diagnostic methods generate large amounts of genomic data. Approximately, 60,000–100,000 single nucleotide variations (SNV) and small insertions and deletions (indel) in each patient's personal genome can be detected on WES [1]. Translating these high numbers of genomic variants into useful clinical information is a crucial task. Although several methods have been introduced to help reduce the vast number of possible genes to clinically causative ones, this process still remains challenging.

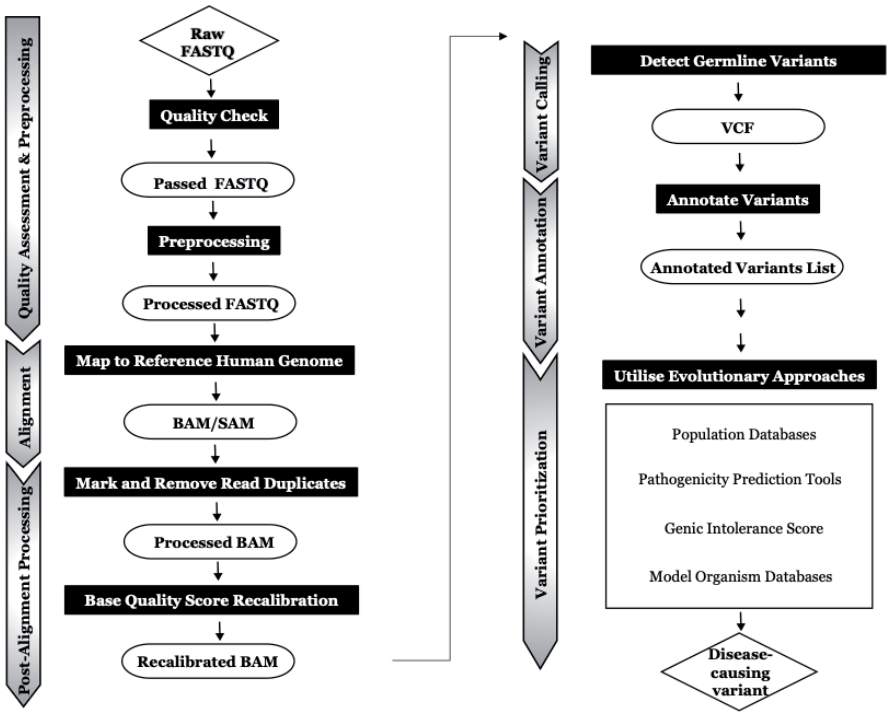


Figure 1.
A general workflow for WES data analysis. Six main steps, quality assessment & preprocessing, alignment, post-alignment processing, variant calling, variant annotation, and variant prioritization integrated with evolutionary approaches, are shown.

Disease-related genes show non-random distribution characteristics in the genome with the majority of them being already present in the eukaryotic ancestor [2]. Mendelian disease genes that underlie single-gene disorders tend to have a more ancient evolutionary origin [3]. Considering disease-related genes have evolved under the effect of natural selection like other genes, evolutionary approaches can provide powerful insight not only to understand human genetic diseases but also to detect genomic variants that cause them.

Here, we briefly describe the analysis workflow from raw data to genomic variants as the first step of the translation to the clinical outcome. We primarily focus on WES analysis because most variations that are responsible for Mendelian disorders disrupt protein-coding regions [4]. Then we give an insight into how evolutionary principles are integrated into the prioritization of detected variants. The framework of the chapter can be found in **Figure 1.**

2. From raw data to genomic variations

The common file format for the storage of data produced by sequencers is FASTQ [5]. FASTQ format stores both nucleotide sequence and its corresponding Phred quality scores [6, 7]. The Phred score related to the base-calling error probabilities indicates the quality of each nucleotide within a read. In a FASTQ file, each read is shown by four lines: The first line begins with a "@" and continues with a sequence identifier and an optional description. The second line consists of the raw sequence letters: A, T, G, C, and N (unknown). The third line starts with a "+"

character and can be followed by the same sequence identifier again. The "+" sign specifies the end of the sequence. The fourth line includes the quality scores for the sequence in the second line.

Here, we give an overview of the data analysis workflow from a FASTQ file to obtain annotated genomic variants.

2.1 Quality assessment and preprocessing

Although NGS platforms are capable of generating massively parallel sequences even in a single run, the quality of sequencing reads may not be perfect due to some reasons such as the failure in experimental processing and technical machine errors. The quality of raw FASTQ data should be assessed in the first step of the workflow since these errors affect downstream analysis.

A number of tools have been developed to evaluate raw FASTQ data. These tools generally take FASTQ files as input and generate summary statistics and graphs for a quick overview of the raw read quality. In addition to the most commonly used one FASTQC [8], developed by Simon Andrews at Babraham Institute, other tools are also available such as FQStat [9], Quack [10], SeqAssist [11], QC-Chain [12]. Based on the result of the quality check step, if there is a need, preprocessing is necessary before alignment.

The standard preprocessing step consists of trimming of low-quality bases and adapter sequence removal at the end of the reads. Adapter sequences can be ligated to 3′ and 5′ ends of reads depending on the used library preparation protocol during the sequencing. These adapter fragments should be removed correctly because of leading to either missed alignments or wrong genotyping in further downstream analyses. Many tools with different principles of implementation have been developed to perform preprocessing. Ktrim [13], PE-Trimmer [14], SeqPurge [15], AdapterRemoval [16], PEAT [17], Skewer [18], Trimmomatic [19], QcReads [20], AlienTrimmer [21], and Btrim [22] are tools can be used for adapter and quality trimming depending on the study design. In addition to these, some tools such as FastqCleaner [23], FastProNGS [24], EasyQC [25], fastp [26], TrimGalore, FASTX-Toolkit, afterQC, ClinQC, NGS QC Toolkit, PRINSEQ, fastQ_brew carry out both quality check and preprocessing functions.

2.2 Alignment of reads

After quality check and preprocessing of raw data, processed reads must be aligned to the reference genome. Both GRCh37 (hg19) and GRCh38 (hg38) are widely used as a reference for the human genome. Optimal alignment to reference sequences is not easy computational task and requires a fast and tolerant algorithm to obtain an imperfect alignment due to genomic variations. Several tools have been developed to align short reads. They mainly use the Burrows-Wheeler Transformation (BWT) algorithm, the Smith-Waterman (SW) dynamic programming algorithm or a combination of both of them. Bowtie2 [27] and BWA [28], which implement the BWT algorithm, are widely used for short reads alignment. Novoalign [29], MOSAIK [30], and SHRiMP2 [31] implement SW algorithm. For a comprehensive review of these methods and their differences, benchmark studies can be found in the literature [32, 33].

The output of the alignment step is the Sequence Alignment Map (SAM) file which contains mapped reads. BAM stands for Binary Alignment Map and is the binary version of a SAM file. Both BAM files and SAM files have the same information which include a header and an alignment section. The header section

provides some information such as reference sequence, read group, sequencing platform details and applied process information to the reads. The alignment section includes the genomic position with relevant descriptive information of each sequence.

SAMtools [34] and Integrative Genomics Viewer (IGV) [35] are also commonly used programs to view BAM/SAM files for further confirmation analysis of detected variants.

2.3 Post-alignment processing

Processing of aligned reads is recommended to improve the quality of downstream variant calling analysis. The processing step generally consists of marking read duplicates and base quality score recalibration (BQSR) to minimize technical biases.

During the sequencing, a library of DNA fragments from a particular genomic region is prepared using PCR amplification to provide adequate DNA fragments for the sequencing process. Therefore, some amplified fragments could share the same sequence and the same corresponding alignment position leading to bias in variant detection. These duplicates should be removed to eliminate PCR-introduced bias. MarkDuplicates available in the Picard [36] and SAMtools [34] are widely-used tools to detect read duplicates based on their identical 5′ region and position on the genome.

In addition to marking duplicates, base quality is also an important factor for variant detection. As mentioned in the section "Quality check and preprocessing", each sequence read has a Phred quality score generated by the sequencing machine. However, the machine could generate systematically biased scores. On the contrary, BQSR patterns errors empirically to recalibrate the base quality scores using a machine learning approach. Thus, technical bias is significantly minimized. The key point in this process is to exclude known variants before BQSR since they are true genomic variations. So, they should not be considered as sequencing errors. The most widely used tool for recalibration of base qualities is BaseRecalibrator available in Genome Analysis Toolkit (GATK) [37].

2.4 Variant calling

After the post-alignment processing step, variant analysis can be started on an analysis-ready BAM file. In the variant calling step, the differences between the reference genome and genome of interest are calculated. Variants can be categorized as germline and somatic variants while dealing with variant calling. Germline variants are inherited variations present in the germ cells. Somatic variants are present only in somatic cells and can be specific to a tissue. In this chapter, we focus on the identification of germline SNV and indels. Several tools based on different algorithms have been developed to call germline short variants. Tools such as HaplotypeCaller available in GATK [38], SAMtools [34], FreeBayes [39], and Platypus [40] are based on Bayesian approaches. VarScan [41] relies on a heuristic approach to identify variants, while SNVer [42] uses a frequentist approach. The performance of different tools has been evaluated by recent studies [43–45], yet, these tools mostly generate an analysis-ready VCF (Variant Call Format) file. A VCF file is a text file that contains header lines and data lines. The header lines begin with "##" symbol. The first header line is always the VCF format version and continues with lines defining the name, length, value type, and description of each item in relevant fields of each data line.

2.5 Variant annotation and prioritization

After variants are detected, biologically important features such as gene symbols, genomic position, amino acid change, and consequences of variants add to a VCF file in the annotation step. In addition to the basic annotation, several tools can be used to integrate the annotations from countless sources including information of known variants with minor allele frequency (MAF) found in public databases and pathogenicity prediction of variants. There are numerous variant annotation tools that implement different methods and most widely used ones are AnnoVar [46], VEP [47], SnpEff [48], GEMINI [49], VarAFT [50], AnnTools [51], SVA [52], NGS-SNP [53]. These annotation tools enable to filtering and prioritizing potential disease-causing mutations. The prioritization of clinically causative mutation among a vast amount of annotated variations is the most challenging part of the analysis and is not a fully automatized. In the next section, we are going to discuss how evolutionary approaches can be used to prioritize genomic variants.

3. Utilizing evolutionary information in variant prioritization

We have described the process of obtaining annotated variations from raw FASTQ data. Experimentally evaluation of each variant at a genomic scale would be an impractical process, but evolutionary principles can provide us a valuable set of an experiment from nature. Integrating evolutionary approaches into the prioritization step have the potential to distinguish the variant responsible for a particular disease among all annotated variants. Indeed, the association between disease and evolution has been attributed to natural selection [54, 55]. During evolution, variations at highly conserved genomic regions are exposed to natural selection because of their negative impact on fitness that make these conserved genes intolerant to variations [56]. On the contrary, at the faster-evolving regions of the genome, many variations have been tolerated over evolutionary time and accumulate in the population with high MAF. However, there is a predisposition for Mendelian disease genes to be more intolerant than the other genes [57]. These genes are also more conserved across species allowing us to compare the phenotypes of different mutant genes on a multispecies level [58].

In this part, we discuss the role of evolutionary approaches in variant prioritization. The first prioritization method aims to filter variants using information from allele frequencies in population databases. Then we introduce several pathogenicity prediction tools to interpret the rest of the variants, especially the ones with uncertain significance. Following that, we describe the usage of gene intolerance information while making inference the variant pathogenicity. Finally, we list commonly used model organism databases that can be used for the comparison of mutant gene phenotypes in several species.

3.1 Population databases

During human evolution, present and novel variations have been evaluated in terms of their biological impact. Population databases record the outcomes of genetic variations providing an extensive catalog that include thousands of individuals' genomic variations to researchers. At the end of the 1990s, the establishment of dbSNP has led to record genotype-phenotype associations via variant databases [59]. Latterly, large-scale projects such as gnomAD and 1000 Genome Project Databases that actively collect genomic data from various populations have become available MAF at population level found in these databases is one of the

primary guides to interpret that variant pathogenicity. Because causative variants related to most Mendelian disorders have deleterious effects on reproductive fitness. Generally, causative alleles are less likely to reside in these databases or are present with low frequencies. In any global population database, except for the well-known founder alleles, >5% MAF can be considered as benign [60]. Therefore, a subset of the total number of variants inside these databases can be used for variant filtration. This is often achieved according to three different approaches. The first approach, called discrete filtering, assumes that a disease-causing variant should not found in these databases [61, 62]. This approach can be useful for very rare Mendelian disorders, but it can be problematic in some cases. Excluding observed alleles, independent from their MAF, can lead to the elimination of truly pathogenic alleles found in the general population at low frequencies because of the increasing number of genomes in databases. Especially, elucidating autosomal recessive disorders are affected by this risk. The second approach, called 1%-approach, is based on allele frequency thresholds that change according to the inheritance model of variants. While the analysis of autosomal recessive variants MAF threshold can be set at 1%, MAF cutoff of 0.1% can be useful for autosomal dominant variants [62]. Alternatively, the third approach, called the quantile-based approach, employs frequency thresholds as in the previous method. However, the thresholds in the quantile-based method are variable and depend on disease prevalence, mode of inheritance, database size, and database characteristics [63].

Depending on the case, different approaches can be employed using population databases with different scopes and data collection. Here, we summarize the widely used population databases. 1000 Genome Project (1KGP) Database.

3.1.1 1000 Genome Project (1KGP) database

1KGP database provides a comprehensive set of human genetic variations from a diverse set of individuals of multiple populations. The database includes the reconstructed genomes of 2504 individuals from 26 populations obtained by combining low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. The database contains over 88 million variants, which consist of around 84.7 million SNPs, 3.6 million indels, and 60,000 structural variants [64, 65].

3.1.2 The Genome Aggregation Database (gnomAD)

gnomAD is an extensive collection of exome and genome sequencing data from several large-scale sequencing projects. The first release of gnomAD is also known as the Exome Aggregation Consortium (ExAC) dataset. gnomAD short variant v2 release contains 125,748 exomes, and 15,708 whole genomes mapped to the GRCh37/hg19 reference sequence. In contrast, the short variant v3 release contains 71,702 whole genomes, including most of the whole genomes from v2 release mapped to the GRCh38 reference sequence. Therefore, gnomAD v2 provides higher power for the analysis of the coding regions, while v3 offers a valuable resource for the analysis of non-coding regions. For the analysis of structural variants, gnomAD SV v2.1 data set grants access to a total of 10,847 genomes aligned against the GRCh37 reference sequence [66].

3.1.3 Database of short genetic variations (dbSNP) and the database of genomic structural variations (dbVar)

The National Center for Biotechnology Information (NCBI) maintains dbSNP and dbVar databases which together contain almost 2 billion submitted human

variants. Although dbVar does not have a reference structural variant database since the current technology cannot detect the precise breakpoints in the genome, dbSNP presents the reference variants as rs identifiers. Other contents of the dataset include population frequency, geographic origin of the population, population-specific genotype and allele frequencies as well as population-specific heterozygosity estimates. Besides serving as a human population database, dbSNP and dbVar also contain a variety of organisms´ genomic variations that can be a valuable resource for evolutionary studies [67, 68].

3.1.4 ClinVar

ClinVar is a public database that archives genetic variances of any type and the interpretations of their clinical significance for reported conditions. Unlike dbSNP and dbVar that are also maintained by NCBI, ClinVar only focuses on the medically relevant variations. Although ClinVar reviews the submissions of variants for validation, the clinical significance of the variants is reported directly from submitters. ClinVar displays any conflict between the interpretations for the same variant from different submitters or the consensus. In the strict comparison approaches, the algorithm evaluates submissions for a variant to be pathogenic and likely pathogenic as conflicting. In the more relaxed approach, the variants can be categorized as pathogenic/likely pathogenic, benign/likely benign, or uncertain significance [69].

3.1.5 Database of chromosomal imbalance and phenotype in humans using Ensembl resources (DECIPHER)

DECIPHER provides a catalog of common copy-number changes in healthy populations as well as chromosome rearrangements of patients and their phenotype record submitted by clinical researchers upon informed consent [70]. Therefore, DECIPHER can serve as a valuable platform during variant prioritization. Users can check both the healthy population database and the previously submitted clinical records within DECIPHER to understand the effect of the variant of interests better and to identify novel and potentially pathogenic variants.

3.2 Pathogenicity prediction tools

Even population MAF-based filtering, individuals generally have many variants that are not present in databases. Most of these variants do not classify definitively as benign or pathogenic according to criteria proposed by some clinical guidelines such as the American College of Medical Genetics and Genomics (ACMG) [60]. These types of alterations termed variants of uncertain significance (VUS). Further filtering approaches must use to reduce the number of VUS. For this purpose, numerous pathogenicity prediction tools based on different principles have been developed to evaluate the variant effect. ACMG and the European Society of Human Genetics (ESHG) [71] guidelines also recommend these in-silico methods to interpret variant pathogenicity.

The first methods were proposed to predict computationally whether an amino acid substitution will disturb the protein function. These methods, now part of the PolyPhen algorithm [72], use physical properties of the mutational change along with a multispecies alignment as a basis to evaluate mutations. Many methods have been derived from this idea and are based on different principles. Evolutionary conservation is among the most useful features for such predictions. Some methods such as SIFT [73], PROVEAN [74] and PANTHER [75] rely on sequence conservation. For example, SIFT, as the most widely used algorithm, compares the

alignments of related sequences by performing a PSI-BLAST search to check if the variant is tolerated in an evolutionary aspect. In addition to sequence conservation, another group of methods which take into account several features such as amino acid physicochemical properties, the context of variation position, protein structural features through machine learning algorithms are also available. CADD [76], MutationTaster2 [77], PolyPhen-2 [72], DANN [78] and VEST3 [79] are well-known examples of such tools.

The predicted impact of a variation obtained from different tools may not be the same. This problem led to researchers making efforts to develop meta predictors that combine the results from existing tools by using several approaches such as logistic regression, decision trees, random forests, and support vector machines to make their own decisions. MetaSVM and MetaLR [80], M-CAP [81] and REVEL [82] are well-known examples of meta-predictors.

Below, several useful tools are explained without a performance comparison. However, various benchmark studies that have extensively examined the accuracy of these tools can be found in the literature [83–85].

3.2.1 MutationTaster2

MutationTaster2, using a naive Bayes classifier, predicts the functional consequences of variants that are both in exonic and intronic regions by incorporating a scoring system for the evolutionary conservation around DNA variants. MutationTaster uses information from several variant databases, including 1KGP and ClinVar. The tool automatically predicts a variant as neutral if it is found more than four times in the homozygous state in these databases and as disease-causing if it is reported as pathogenic in ClinVar by listing the associated disease phenotypes [77].

3.2.2 Combined annotation-dependent depletion (CADD)

CADD combines 63 genomic features derived from evolutionary constraint, surrounding sequence context, and functional predictions to evaluate SNVs and short indels. The tool integrates all of these features into a single CADD score using a machine learning approach trained on a binary distinction between simulated variants and variants that have become fixed in human populations since the split between humans and chimpanzees. C scores correlate with pathogenicity of a variant and disease severity [76].

3.2.3 The Mendelian clinically applicable pathogenicity (M-CAP)

M-CAP uses a supervised learning classifier to interpret genomic variants and focus especially on coding mutations for Mendelian diseases. As a meta-predictor, it uses nine existing tools SIFT, PolyPhen-2, CADD, MutationTaster, MutationAssessor [86], FATHMM [87], LRT [88], MetaLR and MetaSVM. It also combines information of base-pair, amino acid, genomic region, and gene conservation from RVIS [89], PhyloP [90], PhastCons [91], SIPHY [92], GERP [93], PAM250 and BLOSUM62 [94]. Additionally, M-CAP establishes multiple-sequence alignments of 99 primate, mammalian, and vertebrate genomes to the human genome as a new feature [81].

3.2.4 PrimateAI

PrimateAI [95] is a deep neural network trained by a comprehensive dataset that includes around 380,000 common missense variants from humans and six

non-human primate species. PrimateAI categorizes the common missense mutations from other primate species as non-pathogenic for humans. Thus, it enables the identification of the pathogenic variants. PrimateAI has previously shown 88% accuracy in disease-causing variant identification and allowed the discovery of 14 novel candidate genes related to intellectual disability. PrimateAI also incorporates protein structure information as it learns to predict the secondary structure and solvent accessibility from amino acid sequences. PrimateAI provides a score to the user in which a threshold of >0.8 is for likely pathogenic classification, <0.6 is for likely benign, and 0.6–0.8 is as intermediate in genes with dominant modes of inheritance, and a threshold of >0.7 is for likely pathogenic and <0.5 for likely benign in genes with recessive modes of inheritance.

3.3 Genic intolerance

Genic intolerance is a gene-level assessment that has a potential to prioritize genomic variants. It has been developed as a scoring system to calculate tolerance of genes to a functional genetic variation on a genome-wide scale and rank them using 6503 WES data available in the National Heart, Lung, and Blood Institute-NHLBI Exome Sequencing Project [89]. This system predicts the expected common functional variation in the gene and compares them to apparently neutral variation found in the gene. The deviation from this prediction is attributed to the intolerance score, namely the Residual Variation Intolerance Score (RVIS). While genes with a positive RVIS score have more common functional variation than expected, genes with negative RVIS scores have less. A negative RVIS score indicates that the gene is intolerant. The scoring system also shows that the genes that cause Mendelian diseases are significantly more intolerant to functional variation than genes that do not cause any known disease.

3.4 Model organism databases

The evolutionary conservation of many biological processes among species allows the usage of several different model organisms to study human diseases. Although not all the human genes are conserved in invertebrate models such as worms and fruit flies, vertebrate models such as zebrafish and mouse provide valuable resources to study such genes. When evaluating the function of a conserved gene in model organisms, it is critical to keep in mind that orthologous genes usually cause different phenotypes in different species, although the gene products have a similar molecular function. The model organism databases listed below provide the related information on the molecular function of query genes so that they serve as a valuable resource during the variant prioritization process.

3.4.1 Mouse genome informatics (MGI)

MGI is the primary database that integrates genetic, genomic, and biological data for the laboratory Mouse. Mouse Genome Database (MGD) and Mouse Gene Expression Database (GXD) are the two largest contributors to MGI, both serving as valuable resources for the studies of human disease. MGD provides curated phenotypes and functional annotations for mouse genes and alleles, while GXD contains mouse gene expression data with an emphasis on endogenous gene expression during mouse development [96, 97]. The Human-Mouse Disease Connection tool within MGI is another important feature that facilitates exploring gene-phenotype-disease relationships between human and mouse. By simply searching the list of human genes on MGI, the algorithm finds matching mouse genes and their

homologs and displays the both human and mouse phenotypes associated with the genes of interest. MGI is updated once every week by adding new annotations from the literature.

3.4.2 International Mouse Phenotyping Consortium (IMPC)

IMPC aims to establish a comprehensive dataset of mouse genome and phenome by knocking out each gene individually and characterizing the physical and chemical changes, thus providing the foundations for the functional analysis of human genetic variation [98]. The project also aims to generate putative human pathogenic variants in both coding and non-coding regions of the mouse genome.

IMPC uses an algorithm that has been developed to detect phenotypic similarities between the mouse strains of IMPC and more than 7000 rare diseases. The algorithm evaluates a very diverse set of phenotyping parameters that comprise neurological, behavioral, metabolic, cardiovascular, pulmonary, reproductive, respiratory, sensory, musculoskeletal, and immunological parameters and provides a quantitative measure on how well a mouse model recapitulates disease features.

So far, over 3000 genes have already been cataloged and revealed models for 360 diseases, with 90% of the annotated phenotypes being novel [99]. By 2021, IMPC plans to analyze more than 9000 mouse genes to facilitate the prioritization and validation of variations obtained from clinical sequencing efforts.

3.4.3 Rat Genome Database (RGD)

RGD provides genetic, genomic, phenotypic, and disease-related data for the laboratory rat, *Rattus norvegicus*. Rats have been one of the most commonly used model organisms for human disease research. RGD catalogs the rat data and also serves as a comparative data analysis platform between species such as rat, mouse, and human by validating the orthologous relationships. The database currently contains more than 1300 rat strains with disease/phenotype annotations [100]. RGD contains several tools that facilitate the analysis of data in disease-related content. PhenoMiner is such a tool that standardizes the phenotype data obtained from different rat studies by using a variety of ontologies developed at RGD [101]. Users can select one of the Phenominer search categories that include rat strains, experimental conditions, clinical measurements, and measurement methods to begin their search. Then, the algorithm filters the data according to the selected conditions and displays the results.

3.4.4 FlyBase

FlyBase is the central resource for integrated Drosophila genetic and genomic data, including but not limited to sequence-level gene models, mutant phenotypes, mutant lesions and chromosome aberrations, as well as gene expression patterns [102]. The fruit fly—*Drosophila melanogaster*—is a member of the Drosophila family widely used as a model for human disease research.

FlyBase allows different approaches for data presentation to facilitate Drosophila translational research as the two main methods being the gene-centric and disease-centric ones. The Gene Report displays information on individual genes. The report also lists the mutant alleles of the gene and the expression pattern of the gene products. The Human Disease Model Report provides background information on a specific disease and presents summaries of the experimental data and results from previous fruit fly studies.

FlyBase also incorporates orthology prediction tools such as OrthoDB and DIOPT that have been developed to identify orthologs of fly genes in multiple organisms [103, 104]. Integrating the results of these tools to the Gene Reports provides users the identification of orthologs in up to 5000 species. The predicted orthologs serve as a valuable resource for the human disease gene variants prediction as FlyBase also indicates whether the human ortholog functionally complements the fly mutant upon transfer into the Drosophila genome.

3.4.5 WormBase

WormBase serves as the main database for genetic, genomic, and biological information on C.elegans and related nematodes. *C. elegans* is a widely used model for human disease variant research as over 40% of human genes have a C.elegans ortholog. WormBase catalogs the available mutant strains for each gene as well as related nematode studies. WS273 release of WormBase contains over 160,000 gene summaries for 10 nematode species. The gene summaries also include human ortholog diseases and phenotypes to aid the detection of human disease-causing variants [105].

3.4.6 Zebrafish Information Network (ZFin)

ZFIN is the main database that provides genetic, genomic, and phenotypic data from zebrafish studies [106]. Zebrafish—*Danio reriro*—is a model organism extensively used in biomedical research, especially for developmental and genomic studies. Powerful approaches are available to model human diseases using zebrafish. Genetic manipulation of zebrafish orthologs of human disease genes is a common strategy to model genetic disorders such as Duchenne muscular dystrophy [107] and Rett Syndrome [108]. Another strategy of disease modeling is generating transgenic zebrafish lines that express human genes. This approach allows testing the function of the potential disease-causative variant in disease pathology. For example, a transgenic zebrafish model confirmed the pathogenicity of two novel XPNPEP3 gene mutations predicted to be ciliopathy-causing in the clinic [109]. Users can easily search ZFIN to reach information on disease models, including the transgenic lines and mutant phenotypes related to their query.

4. Real-life case

4.1 Variation in the frizzled class receptor 6 (FZD6) protein found in individuals with the nail disorder

Nonsyndromic congenital nail disorder 1 (OMIM #1161050) is a condition affecting the fingernails and toenails characterized by extremely thick nails, onycholysis, hyponychia and claw-like appearance. Autosomal recessive mutations in the FZD6 gene (OMIM *603409) were found to be associated with this disorder [110]. FZD6 is a member of the highly conserved WNT receptors family crucial for developmental processes and differentiation. The study conducted on mice demonstrated that FZD6-mediated Wnt signaling has a regulatory role in the differentiation process of claw/nail formation [111].

In a previous study from our group, a Turkish family with three affected individuals reported. After performing WES on the index case, 96 de novo heterozygous, 421 homozygous, and 185 compound heterozygous variants were obtained

from data analysis. Employing population MAF frequency filtering according to the mode of inheritance has decreased the number of variants to 19, 46, 3 for de novo heterozygous, homozygous, compound heterozygous variants respectively. Further prioritization approaches were applied by integrating pathogenicity prediction scores provided by PrimateAI and other tools, model organism phenotypes, and gene intolerance scores. Ultimately, the FZD6 gene was found to be the most prominent gene even though the gene does not have a high intolerance score. However, the potential functional impact of the mutation was supported by the examination of the evolutionary conservation of the disturbed amino acid region. The region was found to be evolutionarily conserved in other FZD6 orthologues including *Pan troglodytes*, *Macaca mulatta*, *Pongo abelii*, *Bos taurus*, *Canis lupus familiaris*, *Rattus norvegicus*, *Mus musculus*, *Xenopus laevis*. The index case had a homozygous 8 bp deletion on the FZD6 gene caused p.Gly559Aspfs*16. Additionally, this mutation has previously been reported in two other Turkish families. It is also reported that all three families have a common ancestor. In this study, the pathogenicity mechanism for this mutation in nail dysplasia is provided for the first time. The mutation causes a frameshift and creates a premature stop codon at position 16 of the new reading frame [112].

This case study demonstrated that the promising applications of evolutionary approaches assist the clinical diagnosis.

5. Conclusion

Associating genomic variants with diseases is a multistep process. The early steps of this process are highly automated through the usage of several bioinformatics tools. However, the final prioritization step, which is the most critical step, is not completely automated. It requires a comprehensive interpretation together with integrative approaches. In this chapter, we aimed to explain the potential of integrating evolutionary principles into variant prioritization toward clinical utility. This chapter provides sufficient basic information to understand the required bioinformatics tools, various databases with increasing sequence data from individuals as well as model organism research. Finally, we conclude that the pre-evaluation of individual variations with evolutionary approaches can help shorten the diagnostic odyssey, hence saving time and resources. This chapter aims to contribute to the integration of evolutionary genetics to medical genomics. Further studies that combine theoretical and analytical approaches are needed to improve the field of precision medicine via the use of evolutionary insight.

Conflict of interest

The authors declare no conflict of interest.

DOI: http://dx.doi.org/10.5772/intechopen.92738

Author details

Ugur Sezerman*, Tugce Bozkurt and Fatma Sadife Isleyen
Graduate School of Health Sciences, Biostatistics and Bioinformatics Program,
Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

*Address all correspondence to: sezermanu@gmail.com

References

[1] Jamuar SS, Tan E-C. Clinical application of next-generation sequencing for Mendelian diseases. Human Genomics. 2015;**9**:10. DOI: 10.1186/s40246-015-0031-5

[2] Domazet-Loso T, Tautz D. An ancient evolutionary origin of genes associated with human genetic diseases. Molecular Biology and Evolution. 2008;**25**:2699-2707. DOI: 10.1093/molbev/msn214

[3] Cai JJ, Borenstein E, Chen R, Petrov DA. Similarly strong purifying selection acts on human disease genes of all evolutionary ages. Genome Biology and Evolution. 2009;**1**:131-144. DOI: 10.1093/gbe/evp013

[4] Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, Cooper DN. The human gene mutation database: Providing a comprehensive central mutation database for molecular diagnostics and personalised genomics. Human Genomics. 2009;**4**:69. DOI: 10.1186/1479-7364-4-2-69

[5] Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research. 2010;**38**:1767-1771. DOI: 10.1093/nar/gkp1137

[6] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Research. 1998;**8**:175-185. DOI: 10.1101/gr.8.3.175

[7] Ewing B, Green P. Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. Genome Research. 1998;**8**:186-194. DOI: 10.1101/gr.8.3.186

[8] Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[9] Chanumolu SK, Albahrani M, Otu HH. FQStat: A parallel architecture for very high-speed assessment of sequencing quality metrics. BMC Bioinformatics. 2019;**20**:424. DOI: 10.1186/s12859-019-3015-y

[10] Thrash A, Arick M, Peterson DG. Quack: A quality assurance tool for high throughput sequence data. Analytical Biochemistry. 2018;**548**:38-43. DOI: 10.1016/j.ab.2018.01.028

[11] Peng Y, Maxwell AS, Barker ND, Laird JG, Kennedy AJ, Wang N, et al. SeqAssist: A novel toolkit for preliminary analysis of next-generation sequencing data. BMC Bioinformatics. 2014;**15**:S10. DOI: 10.1186/1471-2105-15-S11-S10

[12] Zhou Q, Su X, Wang A, Xu J, Ning K. QC-chain: Fast and holistic quality control method for next-generation sequencing data. PLoS One. 2013;**8**:e60234. DOI: 10.1371/journal.pone.0060234

[13] Sun K. Ktrim: An extra-fast and accurate adapter- and quality-trimmer for sequencing data. Bioinformatics. 2020:btaa171. DOI: 10.1093/bioinformatics/btaa171

[14] Liao X, Li M, Zou Y, Wu F, Pan Y, Wang J. An efficient trimming algorithm based on multi-feature fusion scoring model for NGS data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2019:1-1. DOI: 10.1109/TCBB.2019.2897558

[15] Sturm M, Schroeder C, Bauer P. SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. BMC Bioinformatics. 2016;**17**:208. DOI: 10.1186/s12859-016-1069-7

[16] Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: Rapid adapter trimming, identification,

and read merging. BMC Research Notes. 2016;**9**:88. DOI: 10.1186/s13104-016-1900-2

[17] Li Y-L, Weng J-C, Hsiao C-C, Chou M-T, Tseng C-W, Hung J-H. PEAT: An intelligent and efficient paired-end sequencing adapter trimming algorithm. BMC Bioinformatics. 2015;**16**:S2. DOI: 10.1186/1471-2105-16-S1-S2

[18] Jiang H, Lei R, Ding S-W, Zhu S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014;**15**:182. DOI: 10.1186/1471-2105-15-182

[19] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;**30**:2114-2120. DOI: 10.1093/bioinformatics/btu170

[20] Ma Y, Xie H, Han X, Irwin DM, Zhang Y-P. QcReads: An adapter and quality trimming tool for next-generation sequencing reads. Journal of Genetics and Genomics. 2013;**40**:639-642. DOI: 10.1016/j.jgg.2013.11.001

[21] Criscuolo A, Brisse S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. Genomics. 2013;**102**:500-506. DOI: 10.1016/j.ygeno.2013.07.011

[22] Kong Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics. 2011;**98**:152-153. DOI: 10.1016/j.ygeno.2011.05.009

[23] Roser LG, Agüero F, Sánchez DO. FastqCleaner: An interactive bioconductor application for quality-control, filtering and trimming of FASTQ files. BMC Bioinformatics. 2019;**20**:361. DOI: 10.1186/s12859-019-2961-8

[24] Liu X, Yan Z, Wu C, Yang Y, Li X, Zhang G. FastProNGS: Fast preprocessing of next-generation sequencing reads. BMC Bioinformatics. 2019;**20**:345. DOI: 10.1186/s12859-019-2936-9

[25] Rangamaran VR, Uppili B, Gopal D, Ramalingam K. EasyQC: Tool with interactive user Interface for efficient next-generation sequencing data quality control. Journal of Computational Biology. 2018;**25**:1301-1311. DOI: 10.1089/cmb.2017.0186

[26] Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;**34**:i884-i890. DOI: 10.1093/bioinformatics/bty560

[27] Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nature Methods. 2012;**9**:357-359. DOI: 10.1038/nmeth.1923

[28] Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;**25**:1754-1760. DOI: 10.1093/bioinformatics/btp324

[29] Available from: http://www.novocraft.com/products/novoalign/

[30] Lee W-P, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One. 2014;**9**:e90581. DOI: 10.1371/journal.pone.0090581

[31] David M, Dzamba M, Lister D, Ilie L, Brudno M. SHRiMP2: Sensitive yet practical short read mapping. Bioinformatics. 2011;**27**:1011-1012. DOI: 10.1093/bioinformatics/btr046

[32] Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. Bioinformatics. 2012;**28**:3169-3177. DOI: 10.1093/bioinformatics/bts605

[33] Hatem A, Bozdag D, Toland AE, Çatalyürek ÜV. Benchmarking short sequence mapping tools. BMC Bioinformatics. 2013;**14**:184. DOI: 10.1186/1471-2105-14-184

[34] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;**25**:2078-2079. DOI: 10.1093/bioinformatics/btp352

[35] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;**29**:24-26. DOI: 10.1038/nbt.1754

[36] picard n.d. Available from: http://broadinstitute.github.io/picard/

[37] DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011;**43**:491-498. DOI: 10.1038/ng.806

[38] Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics. 2013;**43**. DOI: 10.1002/0471250953.bi1110s43

[39] Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. ArXiv:12073907 [q-Bio]; 2012

[40] Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, et al. Nature Genetics. 2014;**46**:912-918. DOI: 10.1038/ng.3036

[41] Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;**25**:2283-2285. DOI: 10.1093/bioinformatics/btp373

[42] Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H. SNVer: A statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. Nucleic Acids Research. 2011;**39**:e132-e132. DOI: 10.1093/nar/gkr599

[43] Sandmann S, de Graaf AO, Karimi M, van der Reijden BA, Hellström-Lindberg E, Jansen JH, et al. Evaluating variant calling tools for non-matched next-generation sequencing data. Scientific Reports. 2017;**7**:43169. DOI: 10.1038/srep43169

[44] Chen J, Li X, Zhong H, Meng Y, Du H. Systematic comparison of germline variant calling pipelines cross multiple next-generation sequencers. Scientific Reports. 2019;**9**:9345. DOI: 10.1038/s41598-019-45835-3

[45] Kumaran M, Subramanian U, Devarajan B. Performance assessment of variant calling pipelines using human whole exome sequencing and simulated data. BMC Bioinformatics. 2019;**20**:342. DOI: 10.1186/s12859-019-2928-9

[46] Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010;**38**:e164-e164. DOI: 10.1093/nar/gkq603

[47] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. Genome Biology. 2016;**17**:122. DOI: 10.1186/s13059-016-0974-4

[48] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide

polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118} ; iso-2; iso-3. Fly. 2012;**6**: 80-92. DOI: 10.4161/fly.19695

[49] Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative exploration of genetic variation and genome annotations. PLoS Computational Biology. 2013;**9**:e1003153. DOI: 10.1371/journal.pcbi.1003153

[50] Desvignes J-P, Bartoli M, Delague V, Krahn M, Miltgen M, Béroud C, et al. VarAFT: A variant annotation and filtration system for human next generation sequencing data. Nucleic Acids Research. 2018;**46**:W545-W553. DOI: 10.1093/nar/gky471

[51] Makarov V, O'Grady T, Cai G, Lihm J, Buxbaum JD, Yoon S. AnnTools: A comprehensive and versatile annotation toolkit for genomic variants. Bioinformatics. 2012;**28**:724-725. DOI: 10.1093/bioinformatics/bts032

[52] Ge D, Ruzzo EK, Shianna KV, He M, Pelak K, Heinzen EL, et al. SVA: Software for annotating and visualizing sequenced human genomes. Bioinformatics. 2011;**27**:1998-2000. DOI: 10.1093/bioinformatics/btr317

[53] Grant JR, Arantes AS, Liao X, Stothard P. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics. 2011;**27**:2300-2301. DOI: 10.1093/bioinformatics/btr372

[54] Miller MP. Understanding human disease mutations through the use of interspecific genetic variation. Human Molecular Genetics. 2001;**10**:2319-2328. DOI: 10.1093/hmg/10.21.2319

[55] Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. Proceedings of the National Academy of Sciences. 2004;**101**:15398-15403. DOI: 10.1073/pnas.0404380101

[56] Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, Dickson SP, et al. A genome-wide comparison of the functional properties of rare and common genetic variants in humans. The American Journal of Human Genetics. 2011;**88**:458-468. DOI: 10.1016/j.ajhg.2011.03.008

[57] Cooper DN, Kehrer-Sawatzki H. Exploring the potential relevance of human-specific genes to complex disease. Human Genomics. 2011;**5**:99. DOI: 10.1186/1479-7364-5-2-99

[58] Botstein D, Risch N. Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. Nature Genetics. 2003;**33**:228-237. DOI: 10.1038/ng1090

[59] Sherry ST, Ward M, Sirotkin K. dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. n.d.:4

[60] Richards S, Aziz N, Bale S, Bick D, Das S, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;**17**:405-423. DOI: 10.1038/gim.2015.30

[61] Stitziel NO, Kiezun A, Sunyaev S. Computational and statistical approaches to analyzing variants identified by exome sequencing. Genome Biology. 2011;**12**:227. DOI: 10.1186/gb-2011-12-9-227

[62] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for

Mendelian disease gene discovery. Nature Reviews. Genetics. 2011;**12**: 745-755. DOI: 10.1038/nrg3031

[63] Broeckx BJG, Peelman L, Saunders JH, Deforce D, Clement L. Using variant databases for variant prioritization and to detect erroneous genotype-phenotype associations. BMC Bioinformatics. 2017;**18**:535. DOI: 10.1186/s12859-017-1951-y

[64] The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;**526**:68-74. DOI: 10.1038/nature15393

[65] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;**526**:75-81. DOI: 10.1038/nature15394

[66] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Genomics. 2019. DOI: 10.1101/531210

[67] Sherry ST. dbSNP: The NCBI database of genetic variation. Nucleic Acids Research. 2001;**29**:308-311. DOI: 10.1093/nar/29.1.308

[68] MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The database of genomic variants: A curated collection of structural variation in the human genome. Nucl Acids Res. 2014;**42**:D986-D992. DOI: 10.1093/nar/gkt958

[69] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Research. 2016;**44**:D862-D868. DOI: 10.1093/nar/gkv1222

[70] Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using Ensembl resources. The American Journal of Human Genetics. 2009;**84**:524-533. DOI: 10.1016/j.ajhg.2009.03.010

[71] Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. European Journal of Human Genetics. 2016;**24**:2-5. DOI: 10.1038/ejhg.2015.226

[72] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature Methods. 2010;**7**:248-249. DOI: 10.1038/nmeth0410-248

[73] Ng PC. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research. 2003;**31**:3812-3814. DOI: 10.1093/nar/gkg509

[74] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and Indels. PLoS One. 2012;**7**:e46688. DOI: 10.1371/journal.pone.0046688

[75] Thomas PD. PANTHER: A library of protein families and subfamilies indexed by function. Genome Research. 2003;**13**:2129-2141. DOI: 10.1101/gr.772403

[76] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014;**46**:310-315. DOI: 10.1038/ng.2892

[77] Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: Mutation prediction for the deep-sequencing age. Nature Methods.

2014;**11**:361-362. DOI: 10.1038/nmeth.2890

[78] Quang D, Chen Y, Xie X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 2015;**31**:761-763. DOI: 10.1093/bioinformatics/btu703

[79] Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics. 2013;**14**:S3. DOI: 10.1186/1471-2164-14-S3-S3

[80] Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Human Molecular Genetics. 2015;**24**:2125-2137. DOI: 10.1093/hmg/ddu733

[81] Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nature Genetics. 2016;**48**:1581-1586. DOI: 10.1038/ng.3703

[82] Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. The American Journal of Human Genetics. 2016;**99**:877-885. DOI: 10.1016/j.ajhg.2016.08.016

[83] Li J, Zhao T, Zhang Y, Zhang K, Shi L, Chen Y, et al. Performance evaluation of pathogenicity-computation methods for missense variants. Nucleic Acids Research. 2018;**46**:7793-7804. DOI: 10.1093/nar/gky678

[84] Schaafsma GCP, Vihinen M. Representativeness of variation benchmark datasets. BMC Bioinformatics. 2018;**19**:461. DOI: 10.1186/s12859-018-2478-6

[85] Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? PLoS Computational Biology. 2019;**15**:e1006481. DOI: 10.1371/journal.pcbi.1006481

[86] Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Research. 2011;**39**:e118-e118. DOI: 10.1093/nar/gkr407

[87] Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human Mutation. 2013;**34**:57-65. DOI: 10.1002/humu.22225

[88] Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Research. 2009;**19**:1553-1561. DOI: 10.1101/gr.092619.109

[89] Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genetics. 2013;**9**:e1003709. DOI: 10.1371/journal.pgen.1003709

[90] Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Research. 2010;**20**:110-121. DOI: 10.1101/gr.097857.109

[91] Siepel A. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research. 2005;**15**:1034-1050. DOI: 10.1101/gr.3715005

[92] Garber M, Guttman M, Clamp M, Zody MC, Friedman N,

Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009;**25**:i54-i62. DOI: 10.1093/bioinformatics/btp190

[93] Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Computational Biology. 2010;**6**:e1001025. DOI: 10.1371/journal.pcbi.1001025

[94] Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences. 1992;**89**:10915-10919. DOI: 10.1073/pnas.89.22.10915

[95] Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nature Genetics. 2018;**50**:1161-1170. DOI: 10.1038/s41588-018-0167-z

[96] Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, The Mouse Genome Database Group, et al. Mouse Genome Database (MGD) 2019. Nucleic Acids Research. 2019;**47**:D801-D806. DOI: 10.1093/nar/gky1056

[97] Smith CM, Hayamizu TF, Finger JH, Bello SM, McCright IJ, Xu J, et al. The mouse gene expression database (GXD): 2019 update. Nucleic Acids Research. 2019;**47**:D774-D779. DOI: 10.1093/nar/gky922

[98] The International Mouse Phenotyping Consortium, Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, et al. High-throughput discovery of novel developmental phenotypes. Nature. 2016;**537**:508-514. DOI: 10.1038/nature19356.

[99] The International Mouse Phenotyping Consortium, Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, et al. Disease model discovery from 3,328 gene knockouts by the International Mouse Phenotyping Consortium. Nature Genetics. 2017;**49**:1231-1238. DOI: 10.1038/ng.3901.

[100] Smith JR, Hayman GT, Wang S-J, Laulederkind SJF, Hoffman MJ, Kaldunski ML, et al. The year of the rat: The rat genome database at 20: A multi-species knowledgebase and analysis platform. Nucleic Acids Research. 2019:gkz1041. DOI: 10.1093/nar/gkz1041

[101] Laulederkind SJF, Liu W, Smith JR, Hayman GT, Wang S-J, Nigam R, et al. PhenoMiner: Quantitative phenotype curation at the rat genome database. Database. 2013;**2013**:bat015. DOI: 10.1093/database/bat015

[102] Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al. FlyBase 2.0: The next generation. Nucleic Acids Research. 2019;**47**:D759-D765. DOI: 10.1093/nar/gky1003

[103] Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Research. 2017;**45**:D744-D749. DOI: 10.1093/nar/gkw1119

[104] Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics. 2011;**12**:357. DOI: 10.1186/1471-2105-12-357

[105] Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, et al. WormBase: A modern model organism information resource. Nucleic Acids Research. 2019:gkz920. DOI: 10.1093/nar/gkz920

[106] Ruzicka L, Howe DG, Ramachandran S, Toro S, Van Slyke CE, Bradford YM, et al. The Zebrafish information network: New support for non-coding genes, richer gene ontology annotations and the Alliance of genome resources. Nucleic Acids Research. 2019;**47**:D867-D873. DOI: 10.1093/nar/gky1090

[107] Berger J, Currie PD. Zebrafish models flex their muscles to shed light on muscular dystrophies. Disease Models & Mechanisms. 2012;**5**:726-732. DOI: 10.1242/dmm.010082

[108] Pietri T, Roman A-C, Guyon N, Romano SA, Washbourne P, Moens CB, et al. The first mecp2-null zebrafish model shows altered motor behaviors. Frontiers in Neural Circuits. 2013;**7**:1-10. DOI: 10.3389/fncir.2013.00118

[109] O'Toole JF, Liu Y, Davis EE, Westlake CJ, Attanasio M, Otto EA, et al. Individuals with mutations in XPNPEP3, which encodes a mitochondrial protein, develop a nephronophthisis-like nephropathy. The Journal of Clinical Investigation. 2010;**120**:791-802. DOI: 10.1172/JCI40076

[110] Fröjmark A-S, Schuster J, Sobol M, Entesarian M, Kilander MBC, Gabrikova D, et al. Mutations in frizzled 6 cause isolated autosomal-recessive nail dysplasia. The American Journal of Human Genetics. 2011;**88**:852-860. DOI: 10.1016/j.ajhg.2011.05.013

[111] Cui C-Y, Klar J, Georgii-Heming P, Fröjmark A-S, Baig SM, Schlessinger D, et al. Frizzled6 deficiency disrupts the differentiation process of nail development. Journal of Investigative Dermatology. 2013;**133**:1990-1997. DOI: 10.1038/jid.2013.84

[112] Saygi C, Alanay Y, Sezerman U, Yenenler A, Özören N. A possible founder mutation in FZD6 gene in a Turkish family with autosomal recessive nail dysplasia. BMC Medical Genetics. 2019;**20**:15. DOI: 10.1186/s12881-019-0746-6

Edited by Yusuf Tutar

Treatment risk and response to therapy prediction can be forecasted through early diagnosis, which improves prognosis reliability and effectiveness of therapies. This book covers contemporary advances in molecular markers, disease-causing variants, retroelements, and the basis of distinct diseases.

Published in London, UK

IntechOpen

ISBN 978-1-83962-728-6