

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# DNA Sequencing Resolves Misdiagnosed and Rare Genetic Disorders

*Alice Abdel Aleem*

## Abstract

This chapter focuses on the mandatory requirement of DNA sequencing approaches for genetic diagnosis and recurrence prevention of inherited diseases. Sequencing the DNA and coded transcripts has intensely promoted our understanding of functional genomics and the fundamental importance of non-coding genomic sequences in causing heritable diseases, when mutated. Though Sanger sequencing, the first employed approach in identifying genetic mutations has been replaced nowadays in many laboratories with the highly robust massive parallel sequencing techniques, “Sanger” remains vital in countries with limited resources and also of essential importance in validating the results of large scale sequencing technologies. Next generation sequencing (NGS) enabled the parallel sequencing of the whole exome (WES) and whole genome (WGS) regions of human genome and has revolutionized the field of genetic and genomic research in human. WES and WGS have facilitated the identification of the role of previously unrecognized genes in causing neurologic phenotypes, brain structural malformation, and resolved the causal genes in puzzling and misdiagnosed genetic phenotypes. Role of fusion genes and non-coding RNA in causing neurogenetic recessive diseases has been uncovered by the application of NGS platforms, published examples are presented in this chapter. Extensive phenotypic variability that retained patients either as misdiagnosed or undiagnosed cases for years has been correctly diagnosed through NGS research applications.

**Keywords:** DNA sequencing in human genetic disorders, NGS platforms in rare diseases, leukodystrophies, neuromuscular disorders, muscle dystrophy, non-coding genetic mutation, puzzling phenotypes, NGS, WES, WGS

## 1. Introduction

Since the significant discovery made by Watson and Crick [1] delineating the DNA double helical structure of alternate units (nucleotides) composed of deoxyribose sugar phosphate backbone and nitrogen bases pyrimidines (Cytosine, C and Thiamine, T) and purines (Adenine, A and Guanine, G). And the following crucial findings, Chargaff’s rules [2] informing that the quantity of nitrogen bases differs in between species and the numbers of A equal to T, same for C and G [concluding the pairing status), the field of genetics, genomics, and hereditary is magnificently progressed.

The biology of the genetic code “central dogma” describes the flow of heritable genetic information from the nuclear DNA through the transcription process into the mRNA that is further translated into proteins or families of proteins. Central

dogma of noncoding regions of DNA has also its influences on the stability of mRNA, Exon-intron splicing machinery, and translational efficiency [3–6].

The order (sequence) of nucleotides within a known or yet undiscovered set of genes is the first check point that dictates the coded messenger message and translated proteins. DNA-regulatory sequences including promoters, un-translated regions, DNA-methylation related (epigenetic and posttranscriptional splicing modifications) interactively play in defining the transcriptome and proteome expression profiles in different tissues of the body. Newly developed sequencing technologies have enabled the discovery of these regulatory and expression-modifier sequences [7–9].

Changes in the sequence of DNA-nucleotides located at the coding, non-coding, or splicing regions of the genome are anticipated to amend, in different ways, the genetic message as well as properties of the coded proteins and hence its functions in the cell.

These sequence variations are either inherited (passing from a generation to the next through the germline's cells; ovum/sperm) or spontaneous (de novo) in a subject germ cells. Spontaneous mutations will be further potentially inherited, mostly in a dominant pattern, through the subject's descent when his/her reproduction ability is not affected by the mutation. Changes (polymorphic variations or disease underlying mutations) in the DNA sequence may arise through base substitution, small insertion or deletion of bases, structural variations (large deletions or complex rearrangements), dynamic mutations (expansion of repetitive elements of the genome). DNA, cDNA, or RNA sequencing tools are the evidence based investigations that help us as scientists or physicians to identify or “see in Sanger's chart” these nucleotide changes and accurately allocate its genomic position [10, 11].

Monogenetic (Mendelian) disorders caused by single gene defect(s) are regularly counted under rare (orphan) diseases. Population with high rate of consanguineous marriages described to have extended multiple generational families that harbor rare monogenetic diseases. The single gene defect can occur on the two copies (alleles) of a gene (homozygous mutant) or on one allele only (heterozygous). Inheritance of Mendelian disorders may be autosomal recessive (the two alleles of an autosomal gene should carry the causative mutation to produce the disease phenotype), dominant (one mutant allele will be enough to cause the genetic disease), or X-linked (the mutant gene is located on the X chromosome) with the disease transmission occurring mostly through the females who are obligate carrier of the X-linked mutation [12, 13].

Monogenetic diseases may affect various body systems; cardiovascular, central nervous, peripheral nerves, endocrine, renal, or pulmonary, etc. The clinical phenotypic spectrum of the different distinct categories of these diseases is likely heterogeneous or overlapping which harden the clinicians' decision in making a definitive diagnosis. Academic studies in the field of human genetic diseases as well as diagnostics has been complicated for a long period of time by the remarkable clinical and genetic heterogeneity that were evident for the subgroups of a bunch of familial recurrent diseases involving: congenital muscle dystrophies, limb girdle muscle dystrophies, cortical brain malformation, hereditary spastic paraplegias, hereditary sensory neuropathies, neurodevelopmental, or others. With the evolving NGS technologies progress and discovery has been promptly started.

DNA and cDNA high throughput and validation sequencing tools are fundamental approaches that should be implemented in laboratories to reach a correct genetic diagnosis and provide accurate genetic counseling for rare heritable diseases. Genetic diseases may remain for decades undiagnosed or incorrectly managed when sequencing technologies are either not available or not accessible to patients due to its high cost. Genes and mutations identification in patients services all family members; siblings, cousins, nephews, or other relatives allowing carrier detection, premarital planning when first cousin or relative marriage is considered, prenatal diagnosis or preimplantation genetics. These sequencing outcomes mark

the long term goal of reducing the occurrence or recurrence of genetic disorders in the community achievable.

Discovery of new genes and novel genetic “mutations/etiologies” for rare diseases, has exposed the basis of genetic heterogeneity, increased depth of genomic investigations and been intensely empowered, starting 2005, by the emerged technologies of next generation sequencing (NGS) that enabled the massive parallel sequencing (MPS) of millions of DNA or RNA nucleotides at a time [11, 14].

Whole Exome Sequencing (WES), one of the NGS platforms, grew into a widely used genetic diagnostic test in certified diagnostic labs over the world as well as a research tool in academic studies. WES targets the variants located in the coding regions and splicing boundaries of genes simultaneously at a time. The protein coding genes have been estimated to constitute ~2% of the human genome. Though WES is a powerful tool for the identification of underlying genetic defect in Mendelian disorders, obviously it lacks the capacity to detect non-coding or regulatory disease causing genetic variations [15–18].

Whole Genome Sequencing (WGS), most extensive NGS’ platform, has the capacity to interrogate the whole genome of a subject; the promoters, the un-translated upstream and downstream genomic ends, intragenic and intergenic regions in addition to the coding and splicing parts. Its applications in monogenetic diseases are still mostly at the level academic research. Its value in discovering new causal roles of previously unrecognized genes in rare inherited diseases came from its nature in detecting non-coding, regulatory and large structural variations arise in subjects’ genome [19].

Advances in NGS wet lab methodologies, improvements in informatics pipelines (read alignment, variants call), and the huge released data annotation and analysis platforms lead the new genes discovery, the identification of new etiologies for rare diseases, and new cellular mechanisms contributing genetic syndromes and disorders. The better understanding of molecular biology of gene’s mutation constitutes the essentials for new therapeutics [20].

In this chapter we shed the light on live recent examples demonstrating the role of DNA sequencing tools in gene discovery and in resolving the dilemma of certain genetic phenotypes that were undiagnosed for years.

## **2. Sanger sequencing**

### **2.1 Advances in diagnostics and research of Monogenetic diseases**

To late nineties Sanger sequencing (Chain Terminator Method) was the tool we used to use both in service and research to identify gene mutations or recognize polymorphic sequence variations in particular gene(s). Sanger Sequence is named after Frederick Sanger and his colleagues who had developed the method in late seventies [21, 22]. This sequencing method enabled the identification of nucleotides sequence in a single DNA or RNA amplified fragments and hence the changes (variations) from the reference genomes. Sanger Sequencing was highly applicable in diagnostics when a particular gene or few alternative genes are in question.

### **2.2 Demonstrative example from author’s experience: a well-defined genetic phenotype with two alternative claimed causative genes confirmed true by Sanger sequencing**

Here, we show the value of Sanger sequencing in resolving, in a fairly good turn-around time, the genetic defect in a group of patients with a phenotype of abnormal cerebral white matter associated with subcortical cysts. The leukodystrophies are a

group of diseases, collectively characterized by primarily white matter involvements at variable degrees of severity ranged from a change in signal intensity, on brain images, to cystic cavitation or vanishing of the brain white matter contents [23, 24]. This group of diseases is genetically heterogeneous, however with a good clinical history, examination and high resolution brain imaging, a differential diagnosis can be set and Sanger sequencing can be applied for the few differential genes. The association of distinctive clinical features of macrocephaly (large sized head) detected since birth or shortly thereafter, motor developmental delay, seizures and ataxia precipitated by trauma as well as brain images of diffusely swollen white matter with the very characteristic finding of subcortical cysts preferentially occurring in brain temporal or frontoparietal lobes (**Figure 1**) suggested a clinical diagnosis of megalencephalic leukoencephalopathy (MLC), an autosomal recessive disease [OMIM # 604004]. A long list of metabolic disorders can be listed for a differential diagnosis.

In 75% of these patients, MLC1 gene's mutations are causal for the disease phenotype, whereas in ~20% of cases it is another gene, the HEPACAM/Glia-CAM that contributes the MLC phenotype. Both MLC1 and Glia-CAM are of a reasonable coding regions' size. Application of direct Sanger sequencing had helped several of such patients to get a solid genetic diagnosis of their diseases and allowed their families to use the Sanger sequencing results in performing premarital counseling and preventive measures through the carrier detection and prenatal diagnosis. Thus in cases feature a rather defined phenotype, average sized coding region of genes are in claim, and few alternative candidate causative genes, application of Sanger sequencing empowers the genetic diagnosis in a fair short turnaround time and makes the disease primary prevention quite possible [25].

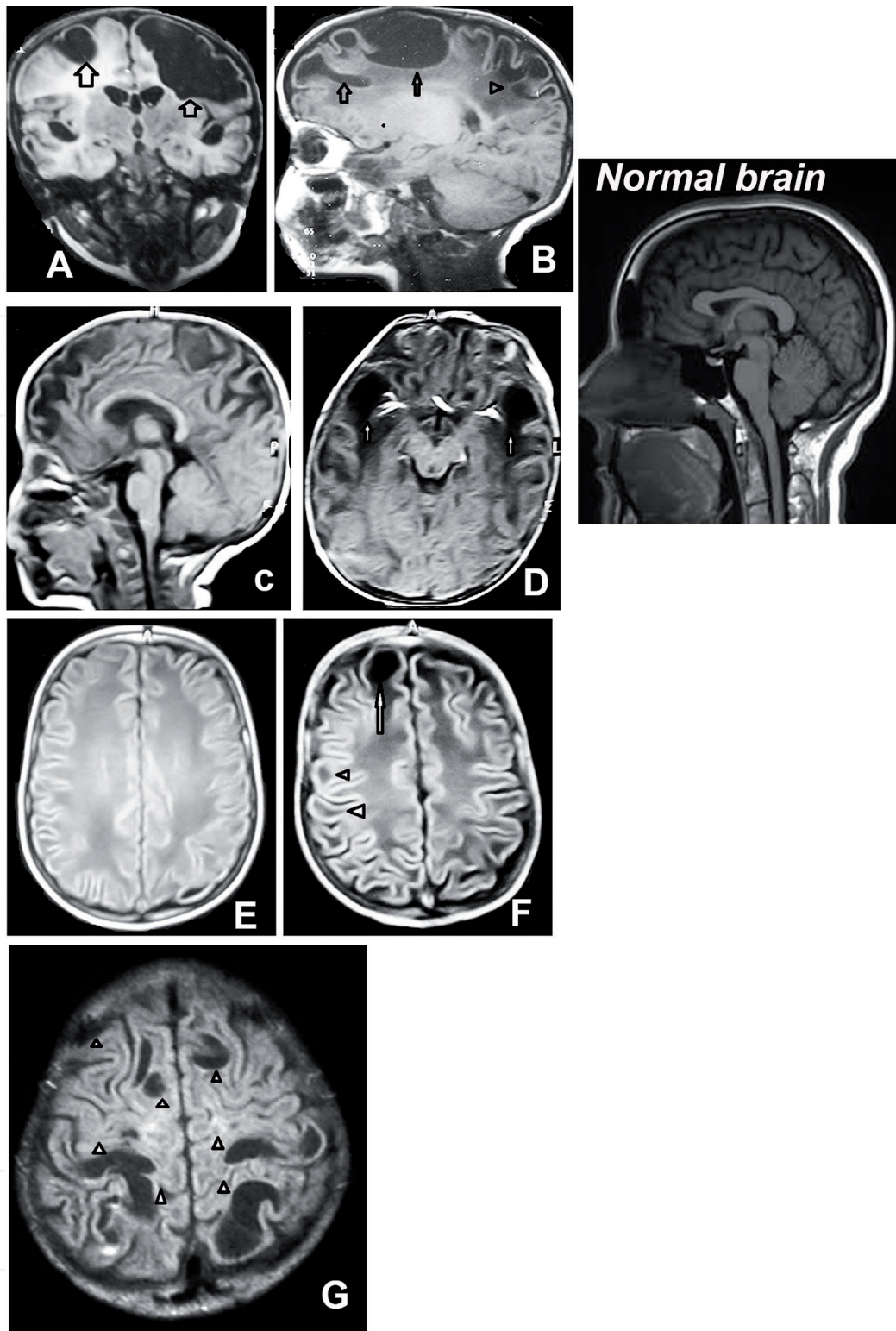
### 2.3 Immunohistochemistry-guided Sanger sequencing

In some other diseases due to a known contributing family of proteins coded by a subset of genes, the roundabout time may be quite consuming to resolve the specific causal gene and hence Sanger sequencing may not be the suitable diagnostic tool particularly when there is a large flow of samples. A good such example is the Limb Girdle muscle dystrophies (LGMDs) which constitute a large group of progressive muscle weakness and wasting. Each of the several main groups of LGMD possesses a list of several subtypes caused by genetic mutations in many of muscle proteins related genes.

Muscle biopsy (a specimen of muscle fibers) used in immunohistochemical staining is an invasive diagnostic approach applied in patients with LGMDs aiming to detect the specific missing (deficient) muscle protein, secondary to gene's alteration using mono- or poly-clonal antibodies.

Sarcoglycanopathies is a known genetic group of LGMDs. It is comprised of a family of four proteins forming four subgroups of sarcoglycanopathies; alpha, beta, gamma, and delta annotated according to the encoded protein and the corresponding gene [26].

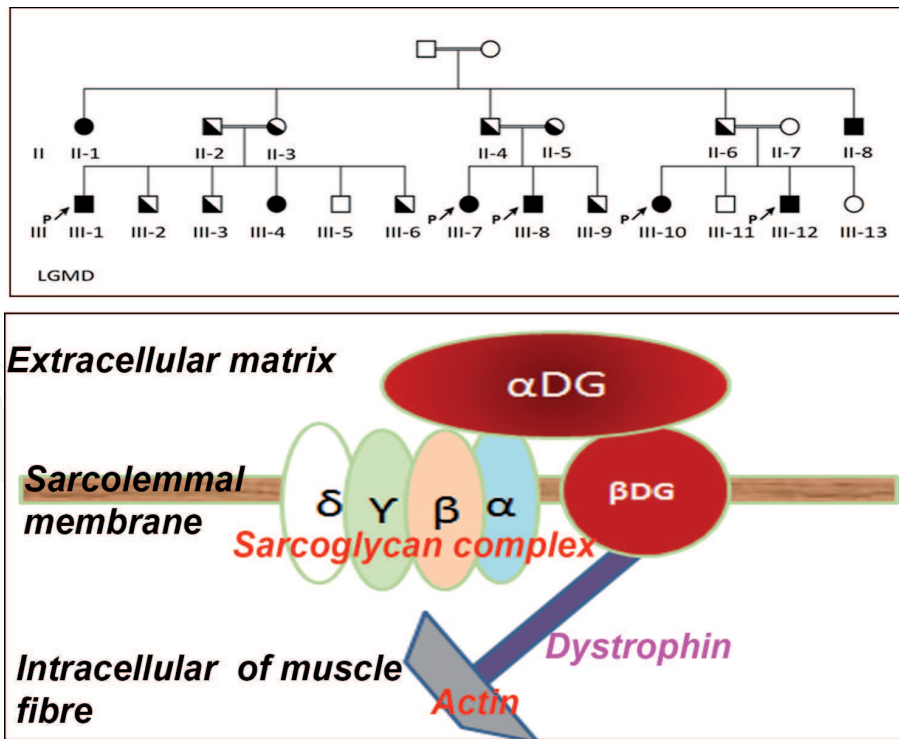
The antibodies implemented in the immunohistochemistry procedure are anticipated to have the capacity to confirm the diagnosis of sarcoglycanopathy-LGMD and the level of the specific protein expression in the muscles, or in the best case scenario may also suggest the specific type of deficient sarcoglycan, whether alpha or beta, etc. However, in order to confidently determine which of the four sarcoglycan genes,  $\alpha$ ,  $\beta$ ,  $\gamma$ , or delta harbors a heritable causative pathogenic mutation, gene sequencing should follow the immunohistochemistry. In such cases, Sanger sequencing guided by the immunohistochemistry results possibly will be a valuable diagnostic approach in areas of limited resources, particularly in extended families with multiple affected subjects across successive generations (**Figure 2**). However, many of the times this is not the case since the antibodies cross react to its different proteins subtypes. In such situation,



**Figure 1.**

Brain magnetic resonance imaging (MRI) in Egyptian patients with *MLC1* mutations [25]. Permission obtained from the copyright owner. Images are of different *MLC* patients captured at ages between 2 and 3 years old. Images T1-weighted sagittal (A and C), coronal (B), and axial (D): show extensive large vacuoles (A), widespread cystic changes involving frontal, temporal, most of the parietal, and occipital white matter (B), diffusely abnormal and swollen cerebral white matter (C and D). Cystic changes in frontal, parietal (C), and temporal subcortical regions (D). Images T1 weighted (E and F): display diffusely abnormal and swollen cerebral, subcortical, and periventricular, white matter. Cystic changes are visible in the frontal lobes (F). Axial T2 image at high level (G) demonstrates cystic changes (vacuolation) affecting different brain cortical regions. Sagittal T1 image of normal brain is included to support perception of the striking differences.

though the time required to interrogate multiple related genes, each separately and release the results may be relatively long, however the sequencing outcomes' significance in disease's prevention and recurrence worth the time and efforts.



**Figure 2.** Extended pedigree with autosomal recessive LGMD alpha-Sarcoglycanopathy. Diagrammatic representation of sarcoglycans complex. The pedigree displayed the effect of consanguineous marriages in producing multiple generations with AR-limb girdle muscle dystrophy (LGMD): subtype alpha sarcoglycan that was confirmed only on DNA sequencing. Immunohistochemistry to the core family reported delta sarcoglycan [cross reactivity pitfall]. The diagram demonstrates the sarcoglycans complex (alpha, beta, gamma, and delta), a family of skeletal muscle sarcolemma proteins that are connected to the extracellular matrix through the alpha dystroglycans ( $\alpha$ DG) which connects to the intracellular dystrophin and actin proteins (muscle cytoskeleton) via  $\beta$ DG forming the dystrophin-glycoprotein complex (DGC).

## 2.4 Challenges for the diagnostic application of Sanger sequencing

Genes of extensively large coding regions like the FBN1, Titin, dystrophin, and many others constitute a challenge to use Sanger direct sequencing as a robust tool to characterize the underlying mutations. As a kind of solution, numerous commercial labs are limiting their molecular diagnostic service to specific gene's mutations' hot spots reported in the populations, when applicable. However, this approach is of a limited value when the case harbors a new or rare gene mutation.

In rather complex or non-specific clinical genetic presentations that are either of un-determined causative genes or of negative gene panel's results for a particular group of diseases, the Sanger sequencing remains unaccommodating.

The evolving roles of non-coding RNA and regulatory sequences alterations in causing heritable genetic diseases toughen the value of Sanger sequencing in diagnostics and human genetic research academic studies.

For all of these essentials new accommodating approaches were in need to satisfy the health care providers' goals to better serve patients with genetic diseases and the researchers need toward discovery of new genes and new etiologies for undiagnosed or misdiagnosed genetic disorders.

Targeted genes panel is a designed approach aiming to collectively sequencing a group of genes of a known causative relation to a particular inherited genetic disease or a group of closely related diseases. Examples involve panels for Limb girdle muscle dystrophy, hereditary spastic paraplegias (HSPs), inherited deafness, etc. This approach essentially and basically requires a continuous update of the designed panel to involve newly discovered genes aiming at avoiding false negative results. HSPs are a large group of diseases characterized by progressive lower limb spasticity, raised heel

(tip toes) gait and associated in its complex phenotype with brain images abnormalities, developmental delay, ataxia, and other features. The list for HSPs associated gene defects is huge involving around 80 genes and continues to expand further [27]. Commercial HSP gene's panel are offered by various diagnostic laboratories, however pitfalls of negative results that falsely decline the diagnosis of HSPs is not uncommon.

Academic studies discover newly characterized HSP related genes yearly; this has to be regularly updating the diagnostic market. A proper alternative tool will be one of the cut edge NGS technologies.

### 3. Next Generation Sequencing (NGS)

#### 3.1 NGS role in mapping genes and mutations to monogenetic diseases' phenotypes

WES and WGS yield a high throughput set of data. Of the interpretation process, these raw sequencing data/reads should be aligned to human reference nuclear genome. Differences between the subjects' sequencing reads and the reference genome are annotated as "variations" which may be counted either as common "polymorphic" or rare variants. The file contains all annotated variants of subject's sample is designated as the variants calling files (VCF).

The NGS' chemistry and nucleotide capture efficiency, depth of sequencing coverage, as well as bioinformatics pipelines employed in calling the variants of subjects' genome including the quality of mapping/alignment to the reference genome govern the potentials of the NGS' output [VCFs] in genes identification [28–30].

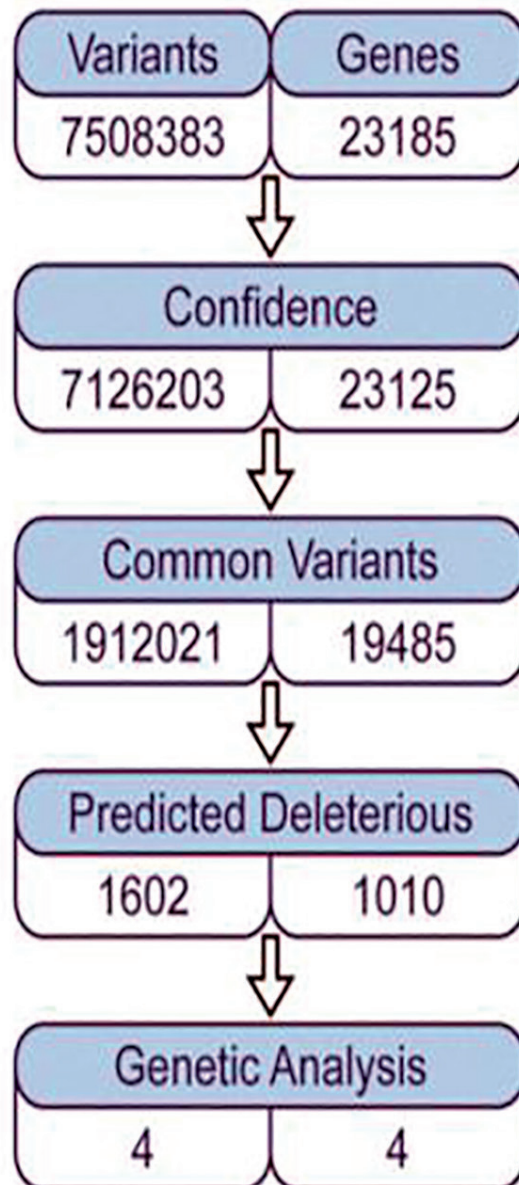
The key challenge in NGS data analysis is to identify the disease causal variants against the tremendous number of variants that are present at a low/rare frequency in genome or annotated, in-silico, as deleterious/pathogenic. Variants prioritization is the protocol employed to select the most potential disease causing variants. The diagram below (**Figure 3**) represents the number of variants originally called in WGS data of a subject and the filters sequentially applied aiming to highlight the most potential candidate disease related variants.

#### 3.2 Gene discovery: identification of genes underlying a worldwide known clinical diagnosis

Kabuki syndrome (KS), OMIM # 147920 is a developmental, musculoskeletal, and intellectual disability with distinctive facial features genetic syndrome. This syndrome was first described, clinically, in families from Japan in 1981 [31] then described worldwide in patients from different ethnic groups. Intensive research has been made using the emerged high throughput sequencing technology to identify the KS causative gene, however unsuccessfully. The sporadic nature of KS (affected patients had negative family history and unaffected parent) harden the path of gene identification. The first Kabuki-associated gene (Lysine methyltransferase 2D, KMT2D, originally named as MLL2, a gene that regulates the expression of several downstream targets) was discovered only by late 2010 [32] along with the further developments made to WES and the process of variants identification and interpretation. KMT2D spontaneous gene mutations were found in over 75% of patients. A second X linked functionally related gene lysine demethylase 6A (KDM6A) contributes 20% of KS cases [33].

This illustrates how it took about 30 years to identify the underlying gene(s) of a well-defined inherited genetic phenotype. Though the most modern high throughput technology was available for quite number of years, however refinement and





**Figure 3.**

*Diagrammatic representation of filters applied in WGS-VCF data analysis. The first line showed number of variants and corresponding number of genes called out of WGS in that particular sample. The confidence filter reveals the number of variants and corresponding genes passed the quality settings selected for this filter including the reads coverage/depth. The common variants filter presents the number of variants and genes after exclusion of “common” variants presented in public data base at frequency high  $\geq 1\%$ . The predicted deleterious filter has several options in its setting; the user can opt in the parameters wished to be considered in the analysis. The genetic filter has alternative options to select pattern of inheritance/transmission; homozygosity vs. heterozygosity in cases vs. controls among multiple other options.*

optimization of variants calling pipeline and variants analysis was recurrently visited to evolve into successful gene discovery for KS.

### 3.3 NGS approach resolves puzzling clinical phenotypes

With the author experience and the clinical examples discussed below we are aiming to outline the significance of NGS in driving research's discovery into clinical implementation and patients care.

Hereditary sensory and autonomic neuropathies, HSANs, are a genetically heterogeneous group of diseases, its phenotypic characteristics involve pain sensitivity (sensory loss) with its sequels, decreased sweating (hypohidrosis/autonomic function), plus mild motor weakness in a subset of patients [34]. Though the mechanism of development of disease pathology is not well understood, however; a

known, short list of underlying genes were characterized and sequenced when the unique HSAN phenotype is suspected.

A consanguineous pedigree had two children, a boy and a girl aged 14 and 10 years respectively displayed a phenotype resembled that of hereditary sensory and autonomic neuropathies (HSANs). The clinical presentations characterized by two distinct features, severe pain insensitivity associated with hypohydrosis since birth along with the sequels of impaired pain sensation and severe aseptic destruction of large and small joints as well as the vertebrae (**Figure 4**). The two affected siblings had been examined by multiple local and international experts, the clinical diagnosis given was a general one describing an immune inflammatory disease (due to the joints destruction), however the association of the severely remarkable pain insensitivity remained unexplained in the context of immune-inflammation.

WGS revealed, unexpectedly, a homozygous mutation in LIFR. LIFR mutations have been associated with Stüve-Wiedemann syndrome (SWS), a lethal autosomal recessive skeletal dysplasia that may be associated with mild reduced pain sensation in atypical long survivors.

The complexity (overlapping phenotypes) as well as the striking severity of pain insensitivity phenotype, which phenocopy HSANs and atypically associated with extensive bone destruction challenges the diagnosis. The WGS had resolved this case dilemma, provided the family opportunities for preimplantation genetics as well as premarital counseling for other family members. Not only had that, but also reveals a new mechanism of LIFR's functional alteration (defective glycosylation of the mutant protein) [35]. WGS finding in these cases warrant the attention to consider LIFR testing in genetically unresolved phenotypes mimics HSAN.

### **3.4 NGS maps neurodevelopmental axonal guidance phenotype to a previously unrecognized gene**

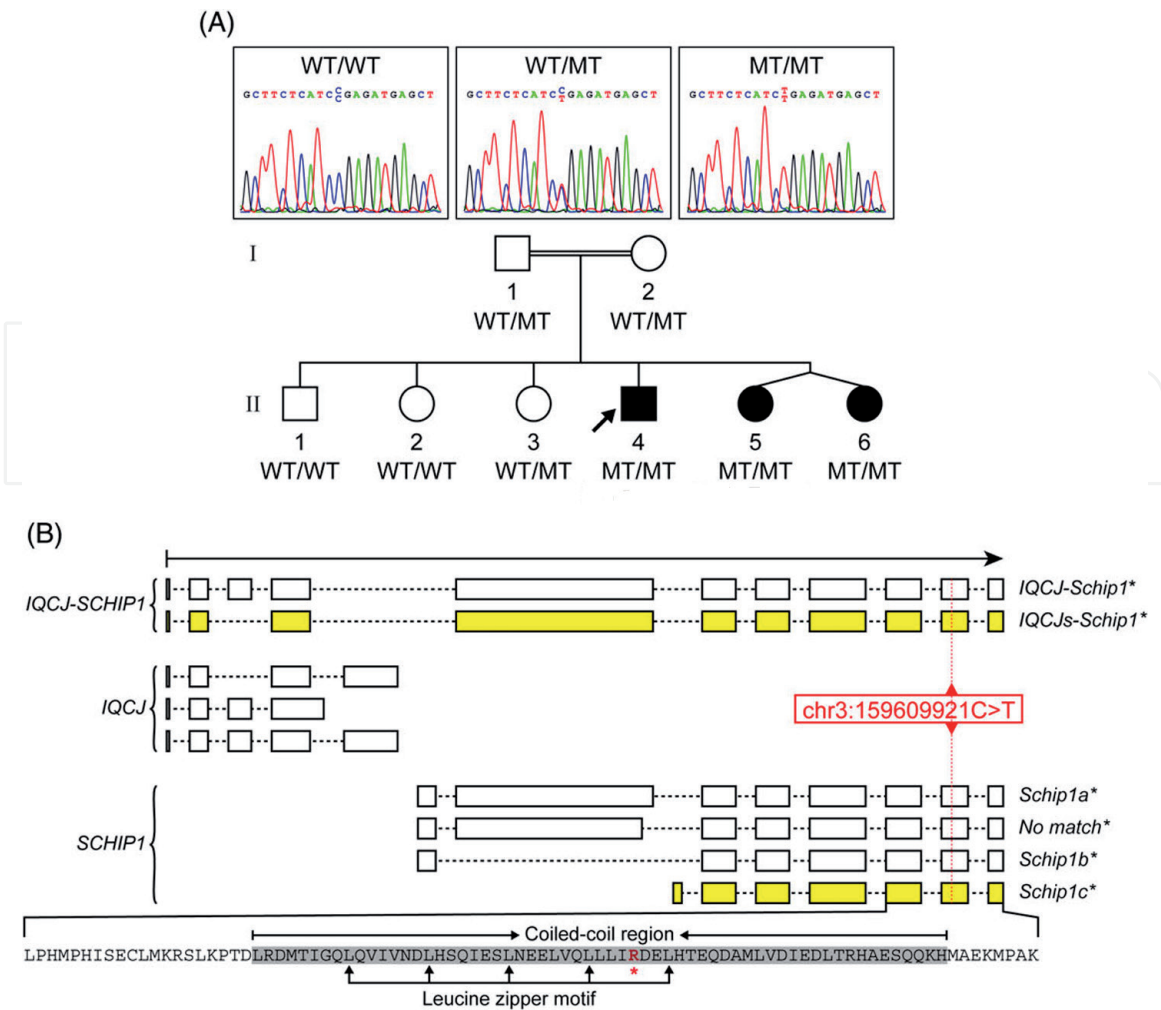
Neurodevelopmental disorders associated with brain malformation are the most extensively large group of neurological disorders. This group incorporates a broad spectrum of manifestations primarily involving the central nervous system and variably associated with motor and/or psychomotor delay, microcephaly, epilepsy, specific behavior, abnormal movements, eye symptoms, dysmorphic features, or hypotonia. Brain imaging is very helpful for the clinical diagnosis; however it remains challenging to reach a firm genetic diagnosis without NGS approaches. Each individual disease of this group is of the rare diseases. Some underlying genes have been identified and characterized; many others stay unknown or uncharacterized for its role in causing such diseases, waiting further research and discoveries.

We present here such example of a family with three affected siblings, a boy and a twin sister born to a consanguineous parent. The clinical phenotype of global developmental delay, learning difficulties associated with mild dysmorphism, hearing impairment was presented at variable severity between the older boy and the two affected female siblings. This clinical phenotype though can be categorized as neurodevelopmental disorder, however is very nonspecific. The older boy was given a provisional diagnosis of autistic spectrum hyperactivity due to some related features. The brain imaging of cortical malformation (polymicrogyria-cobblestone complex), central atrophy, and axonal guidance defects were variably shown in the three siblings. WGS applied for 8 members of this family (6 siblings: 3 affected and 3 unaffected plus parent) followed by bioinformatic variants analysis and genes functional reviews have successfully filtered the SNVs yield and identified a novel nonsense mutation in a previously unrecognized gene, Schwanomin-Interacting Protein1 (SCHIP1) (**Figure 5**) [36]. SCHIP1 was not previously associated to human neurodevelopmental disorders or brain malformation. However, mouse studies knocked out



**Figure 4.** Phenotypic sequels of severe pain insensitivity, aseptic painless fractures and inflammation of large and small joints in patients with LIFR mutation [35]. Permission obtained from the copyright owner. Multiple images display: prominent spine kyphoscoliosis. Swollen knee joints; the scar in the left knee is due to surgical procedure treating joint's inflammation. Neuropathic chronic planter ulcerations: extensive and penetrating at the right heel, healing at the left big toe; evident hyperkeratosis of surrounding skin. Inflamed and painless distal inter-phalangeal joint of the left finger. Kyphoscoliosis, obvious corneal opacity (abnormal white band), tongue ulceration, and fissured lips (due to hypohydrosis-dryness) featured the pain insensitivity in the younger affected subject. This figure aims to demonstrate how the clinical picture mimics the diseases of pain insensitivity however the gene sequencing reveals a different disease category.

schip1 isoforms produced a phenotype of brain axonal guidance defects, similarly to that detected in these patients. This gene has multiple isoforms including a fused gene (IQCJ-SCHIP1) isoform with variable tissue expression pattern and reported to have a role in axonogenesis during brain development. This example demonstrated the significant role of massive parallel sequencing approach as well as reviews of studies developed in mice with rather similar brain imaging phenotype in characterizing a new gene contributing neurodevelopmental-brain malformation phenotype.



**Figure 5.** NGS maps neurodevelopmental axonal guidance phenotype in a consanguineous family to *SCHIP1* [36]. Permission obtained from the copyright owner. Images showed (A) the family pedigree, the Sanger sequencing validating the nonsense mutation positioned in *SCHIP1* and fused *IQCJ-SCHIP1* isoform and its recessive segregation in family members. (B) Diagrammatic assembly of the *IQCJ/SCHIP1* locus (exons illustrated as boxes, introns as lines) and the alternatively spliced transcriptional isoforms. The mutation position is marked in red. The asterisks indicate a mouse isoform of the human protein.

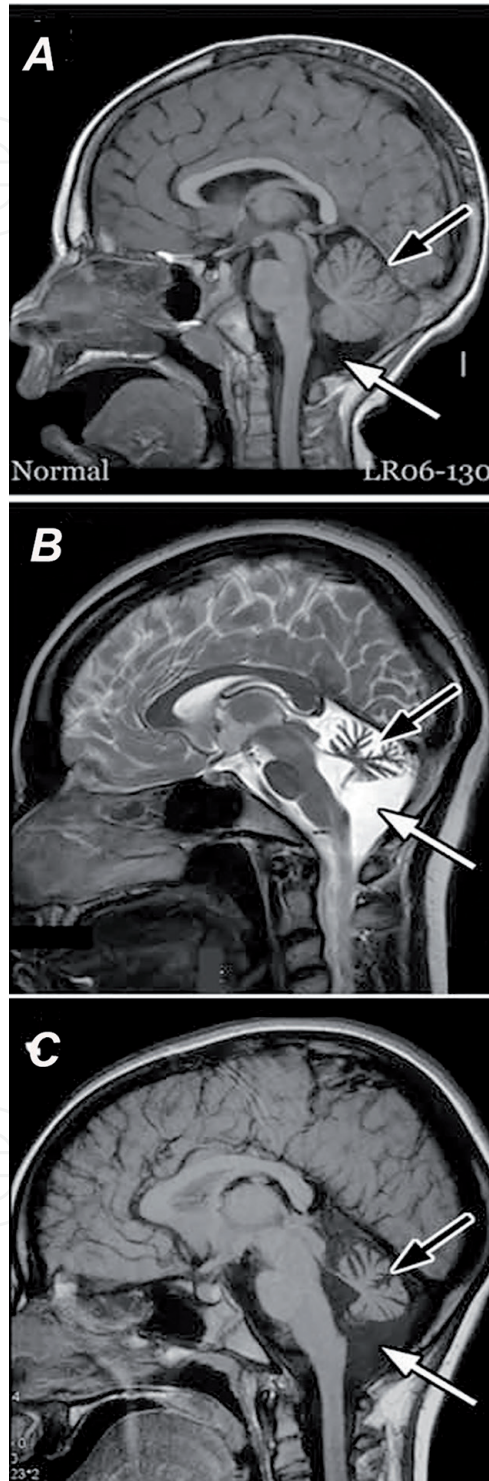
### 3.5 WGS reveals new non-coding RNA minor splicing component's machinery that maps to a pure congenital cerebellar ataxia phenotype

Hereditary Cerebellar Ataxias (HCAs), the uncoordinated gait and body movements, can be inherited as autosomal dominant or recessive traits or in association with other neurological diseases. Hereditary ataxias are due to degeneration of cerebellar neurons or spinocerebellar tracts dysfunction [37]. Many several genes, its coding regions have been identified as causatives for the HCAs.

The emerging regulatory role of small non-coding RNA is evolving as a new mechanism leading human genetic diseases. WGS is particularly relevant to the identification of mutations in non-coding regions of the genome. An example, the 2nd worldwide of such condition was recently published [38]. In this referenced article, a large interrelated kindred had 6 patients with hereditary ataxias of unknown genetic etiology. Delayed speech and developmental milestones, congenital hypotonia, dysarthric speech, intention tremor, head nodding, and ataxic gait with a falling tendency were the main complains, however at variable severity among the affected patients. Brain images support the cerebellar involvements (**Figure 6**). Clinical diagnosis of an autosomal recessive cerebellar ataxia

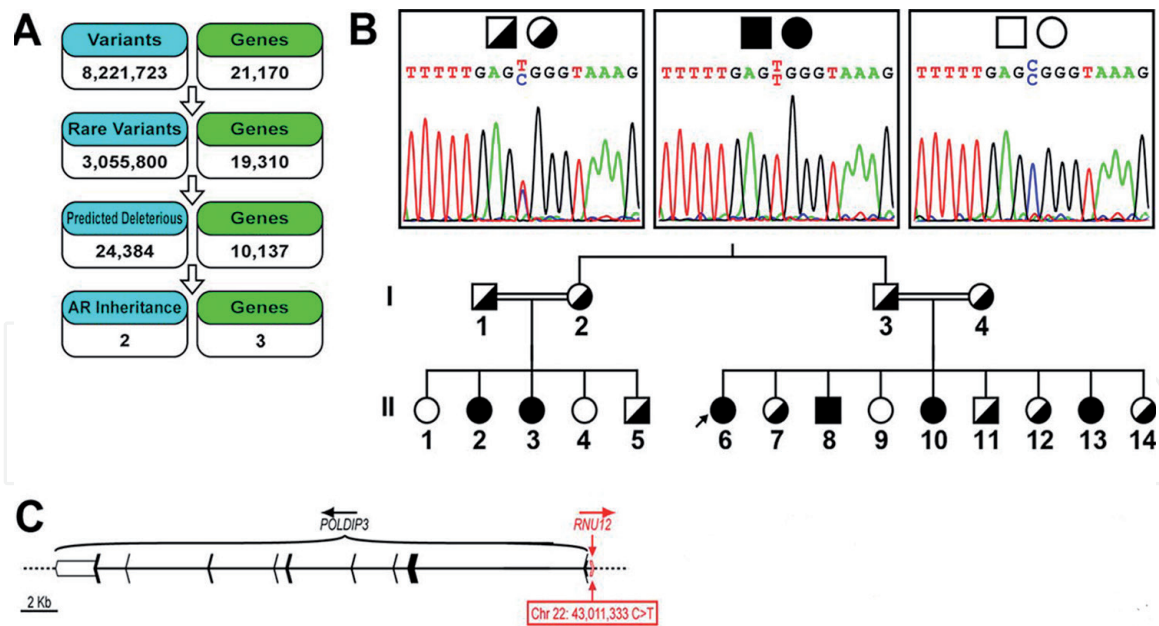
was suggested. Genetic investigations involving gene panel test and WES were performed; however results came back as negative.

WGS performed, on research basis, for 11 members of two branches of the extended family revealed interesting, nevertheless complex result that required functional testing to verify the causative gene and the biological impact of the



**Figure 6.**

*Brain MRI-sagittal views in normal and patients with ncRNU12 mutations [38]. Permission from the copyright owner was obtained. Normal sagittal T1-weighted midline image (A) showing normal cerebellar foliation as well as normal brainstem proportions. A normal variant of prominent cisterna magna, the space at the posterior fossa (white arrow). Sagittal T2 weighted images (B and C) are of two affected female cousins, with variable degrees of clinical severity: pictures displayed dilatation of the cerebellar interfolia spaces indicating cerebellar atrophy or hypoplasia, reduction in the superior and inferior vermis' sizes (arrows). Brainstem is moderately affected (smaller in size). Widened CSF spaces around the posterior fossa, 2ndry to cerebellar and brain stem atrophy, more obvious in image B.*



**Figure 7.** WGS variant identification, Sanger validation, and genomic organization of the region encompassing the RNU12 variant [38]. Permission from the copyright owner was obtained. (A) Filters of ingenuity variant analysis (IVA) tool used to analyze the WGS data of 11 members of the extended family concluded two variants (one pathogenic, the 2nd was a kind of haplotype-linked polymorphism) passed all filters, corresponding to 3 genes, 2 that were recognized by IVA and a third gene identified by direct inspection of the locus. (B) Sanger sequencing of subjects' DNA validating the single nucleotide substitution on chromosome 22 and its recessive segregation with the phenotypes (normal/wild, normal/carrier, patient/mutant). (C) Diagrammatic representation of the chromosome 22 region, incorporating the variant that maps to the proximal promoter of POLDIP3 and a functional sequence of RNU12. POLDIP3 and RNU13 are transcribed from opposite strands (arrows). Multiple functional experiments were done and confirmed RNU12 contribution to the phenotype and disease pathomechanism.

genomic mutation. WGS data analysis identified a variant (SNV) that was located in the promoter region of a protein coding gene POLDIP3 and fell as well in a small nuclear non-coding RNA gene (RNU12) that was transcribed from the opposite strand (**Figure 7**). Interestingly, RNU 12 was reported as a component of the U12-minor splicing machinery that functions in splicing of genes containing minor introns. Experimental investigations involved quantitative expression of the genes, RNA seq, semi-quantitative analysis of retention of minor introns containing genes (due to defective splicing machinery) established the causal relation of RNU12 to the disease phenotype in this large family. This story underscores the value of WGS in uncovering the unrecognized regulatory role of snRNU12 gene in human brain development and function. And the value in identifying the molecular gene defect in an example of monogenetic diseases that would have been remained uncovered when only WES was undertaken. This gene's result has been used by healthy family members in carrier detection, premarital counseling and prenatal diagnosis.

The ages at which patients of this kindred have getting the genetic diagnosis of their disease were as of 25 year old (for the female proband), 22 year old for her brother, 15 and 10 years old of her sisters (first branch), 19 and 13 years old of female siblings of second branch. This highlights how NGS empowered the diagnostic odyssey of monogenetic diseases translating research into clinic improving targeted patients care and prevention of diseases' recurrence in family and community.

#### 4. Conclusion

Advancement of new therapeutics for genetic diseases is definitely influenced by research and technologies that support swift, reliable, and interpretable OMICs

(genomic, transcriptomic and proteomic) research. DNA and RNA sequencing are of such technologies that greatly advanced the discoveries in human genetics. However, still further improvements of big-data pipeline analysis and functional investigations are mandatory to maximize and empower discoveries made by the “Sequencing.”

## **Acknowledgements**

The author thanks the team and colleagues participated in the work of original images that are re-used in this chapter following the permissions of the Copyright Owners. This chapter was made possible by funds received from the Qatar National Research Fund [grants: PPM1-1206-150013 and NPRP4-099-3-039; principal investigator: Alice Abdelaleem], a member of Qatar Foundation. The findings achieved herein are solely the responsibility of the author.

Author's appreciation extends to the institutes of Weill Cornell Medicine Qatar, Brain and Mind Research Institute-NY-USA, Hamad Medical Corporation Qatar, Pediatric Neurology-Cairo University Hospitals Egypt, and National Research Centre Egypt.

## **Conflict of interest**

The author has nothing to declare.

## **Notes**

Permissions for re-published figures have been obtained from the Copyright Owner(s).

IntechOpen

## **Author details**

Alice Abdel Aleem

Weill Cornell Medicine Qatar, Brain and Mind Research Institute, Education City, Doha, Qatar

\*Address all correspondence to: [aka2005@qatar-med.cornell.edu](mailto:aka2005@qatar-med.cornell.edu)

## **IntechOpen**

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Watson JD, Crick FH. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*. 1953;**171**:737-738
- [2] Chargaff E, Lipshitz R, Green C. Composition of the deoxyribose nucleic acids of four genera of sea-urchin. *The Journal of Biological Chemistry*. 1952;**195**:155-160. PMID 14938364
- [3] Nirenberg M. Historical review: Deciphering the genetic code—A personal account. *Trends in Biochemical Sciences*. 2004;**29**:46-54. DOI: 10.1016/j.tibs.2003.11.0090
- [4] Nirenberg M, Leder P. RNA code words and protein synthesis. *Science*. 1964;**145**:1399-1407. DOI: 10.1126/science.145.3639.1399
- [5] Jimenez-Sanchez G, Childs B, Valle D. Human disease genes. *Nature*. 2001;**409**:853-855. DOI: 10.1038/35057050
- [6] Stranger B, Dermitzakis E. From DNA to RNA to disease and back: The 'central dogma' of regulatory disease variation. *Human Genomics*. 2006;**2**:383-390. DOI: 10.1186/1479-7364-2-6-383
- [7] 1000 Genomes Proj. Consort. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;**491**:56-65
- [8] Goodwin S, McPherson J, McCombie W. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*. 2016;**17**:333-351
- [9] Flusberg B, Webster D, Lee J, Travers K, Olivares E, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*. 2010;**7**:461-465
- [10] Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*. 2013;**110**:3-24
- [11] Levy S, Myers R. Advancements in next-generation sequencing. *Annual Review of Genomics and Human Genetics*. 2016;**17**:95-115
- [12] McKusick V. Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders. 12th ed. USA: Baltimore Johns Hopkins University Press; 1998
- [13] Amberger J, Bocchini C, Schiettecatte F, Scott A, Hamosh A. OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*. 2015;**43**:D789-D798. DOI: 10.1093/nar/gku1205
- [14] Shendure J, Balasubramanian S, Church G, Gilbert W, Rogers J, Schloss J, et al. DNA sequencing at 40: Past, present and future. *Nature*. 2017;**550**:345-353
- [15] Voelkerding K, Dames S, Durtschi J. Next-generation sequencing: From basic research to diagnostics. *Clinical Chemistry*. 2009;**55**:4641-4658
- [16] Bamshad M, Ng S, Bigham A, Tabor H, Emond M, Nickerson D, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews-Genetics*. 2011;**12**:745-755
- [17] Kaiser J. Human genetics. Affordable 'exomes' fill gaps in a catalogue of rare diseases. *Science*. 2010;**330**:903-911
- [18] Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*. 2010;**42**:30-35



- [19] Mardis E, Whole-genome sequencing: New technologies, approaches, and applications. Genomic and Personalized Medicine. 2nd ed. Elsevier, UK: Ginsburg & Willard; 2013. pp. 87-93. Chapter 7. eBook ISBN: 9780123822284
- [20] Salk J, Schmitt M, Loeb L. Enhancing the accuracy of next generation sequencing for detecting rare and subclonal mutations. Nature Reviews Genetics. 2018;**19**:269-285
- [21] Sanger F, Nicklen S, Coulson A. DNA sequencing with chain-terminating inhibitors. Proceedings of the National academy of Sciences of the United States of America. 1977;**1077**(74):5463-5467
- [22] Maxam AM, Gilbert W. A new method for sequencing DNA. PNAS. 1977;**74**:560-564
- [23] van der Knaap M, Wolf N, Heine V. Leukodystrophies. Neurology clinical Practice. 2016;**6**:506-514. DOI: 10.1212/CPJ.0000000000000289
- [24] van der Knaap M, Bugiani M. Leukodystrophies—Much more than just diseases of myelin. Nature Reviews | Neurology. 2018;**14**:747-748
- [25] Mahmoud I, Mahmoud M, Refaat M, Girgis M, Waked N, El Badawy A, et al. Clinical, neuroimaging, and genetic characteristics of megalencephalic leukoencephalopathy with subcortical cysts in Egyptian patients. Pediatric Neurology. 2014;**50**:140-148
- [26] Nalini A, Gayathri N, Thaha F, Das S, Shylashree S. Sarcoglycanopathy: Clinical and histochemical characteristics in 66 patients. Neurology India. 2010;**58**:691-696
- [27] Bis-Brewer D, Züchner S. Perspectives on the genomics of HSP beyond mendelian inheritance. Frontiers in Neurology. 2018;**9**:1-11. DOI: 10.3389/fneur.2018.00958
- [28] Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. PNAS. 2009;**106**:96-101
- [29] Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, et al. Clinical interpretation and implications of whole-genome sequencing. JAMA. 2014;**311**:1035-1045
- [30] Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Human Mutation. 2015;**36**:815-822
- [31] Niikawa N, Matsuura N, Fukushima Y, Ohsawa T, Kajii T. Kabuki make-up syndrome: A syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. The Journal of Pediatrics. 1981;**99**:565-569
- [32] Ng S, Bigham A, Buckingham K, Hannibal M, McMillin M, Gildersleeve H, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics. 2010;**42**:790-793
- [33] Lederer D, Grisart B, Digilio MC, et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. American Journal of Human Genetics. 2012;**90**:119-124
- [34] Rotthier A, Baets J, Timmerman V, Janssens K. Mechanisms of disease in hereditary sensory and autonomic neuropathies. Nature Reviews Neurology. 2012;**8**:73-85
- [35] Elsaid M, Chalhoub N, Kamel H, Ehlayel M, Ibrahim N, Elsaid A,

et al. Non-truncating LIFR mutation:  
Causal for prominent congenital pain  
insensitivity phenotype with progressive  
vertebral destruction? *Clinical Genetics*.  
2016;**89**:210-216

[36] Elsaid M, Chalhoub N,  
Ben-Omran T, Kamel H, AlMureikhi M,  
Ibrahim K, et al. Homozygous nonsense  
mutation in SCHIP1/IQCJ-SCHIP1  
causes a neurodevelopmental brain  
malformation syndrome. *Clinical  
Genetics*. 2018;**93**:387-391

[37] Anheim M, Tranchant C, Koenig M.  
The autosomal recessive cerebellar  
ataxias. *The New England Journal of  
Medicine*. 2012;**366**:636-646

[38] Elsaid M, Chalhoub N,  
Ben-Omran T, Kumar P, Kamel H,  
Ibrahim I, et al. Mutation in noncoding  
RNA RNU12 causes early onset  
cerebellar Ataxia. *Annals of Neurology*.  
2017;**81**:68-78

IntechOpen