

Gene Hunting Approaches through the Combination of Linkage Analysis with Whole-Exome Sequencing in Mendelian Diseases: From Darwin to the Present Day

Seda Susgun^{a, b} Koray Kasan^c Emrah Yucesan^a

^aDepartment of Medical Biology, Faculty of Medicine, Bezmialem Vakif University, Istanbul, Turkey; ^bGraduate School of Health Sciences, Istanbul University, Istanbul, Turkey; ^cFaculty of Medicine, Bezmialem Vakif University, Istanbul, Turkey

Keywords

Gene hunting · Linkage analysis · Consanguineous marriages · Mendelian diseases · Whole-exome sequencing

Abstract

Background: In the context of medical genetics, gene hunting is the process of identifying and functionally characterizing genes or genetic variations that contribute to disease phenotypes. In this review, we would like to summarize gene hunting process in terms of historical aspects from Darwin to now. For this purpose, different approaches and recent developments will be detailed. **Summary:** Linkage analysis and association studies are the most common methods in use for explaining the genetic background of hereditary diseases and disorders. Although linkage analysis is a relatively old approach, it is still a powerful method to detect disease-causing rare variants using family-based data, particularly for consanguineous marriages. As is known that, consanguineous marriages or endogamy poses a social problem in developing countries, however, this same condition also provides a unique opportunity for scientists to identify and characterize pathogenic variants. The rapid advancements in sequencing technologies and their parallel implementa-

tion together with linkage analyses now allow us to identify the candidate variants related to diseases in a relatively short time. Furthermore, we can now go one step further and functionally characterize the causative variant through in vitro and in vivo studies and unveil the variant-phenotype relationships on a molecular level more robustly. **Key Messages:** Herein, we suggest that the combined analysis of linkage and exome analysis is a powerful and precise tool to diagnose clinically rare and recessively inherited conditions.

© 2021 The Author(s).
Published by S. Karger AG, Basel

Introduction

“As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it varies however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of sur-

Seda Susgun and Koray Kasan contributed equally to this review.

viving, and thus be naturally selected.” wrote Darwin and Kebler [1] in his book *On the Origin of Species*, highlighting the biological significance of inherited variations for populations and for the emergence of new species. The significance from the perspective of individuals, however, can manifest as suffering since most of the phenotypic translations of inherited variation produce detrimental outcomes [2]. Darwin also demonstrated through his inbreeding experiments that the incidence rate of the detrimental phenotypes or diseases and disorders increases tremendously in individuals with parentages that have low genetic diversity, which has led him to become concerned a great deal for his and his children’s health from the consequences of the long tradition of consanguineous marriages in his family [3]. Indeed, we now know that his worries were not misplaced, for the advancements in modern medical genetics have clearly shown the increased risks associated with consanguineous marriages in terms of inherited familial diseases and disorders, today we name them Mendelian diseases.

In medical genetics, gene hunting is the process of the identification and description of candidate genes and/or genetic variations that contribute to disease phenotypes. Great number of studies in the past have laid the conceptual foundations for the methods in use today, starting from Mendel’s [4] demonstration of heredity being transmitted in units in 1865; to Morgan’s [5] discovery that chromosomes carry the genes that allow offspring inherited traits from parents in 1910; Morgan’s [5] student Stuartvert’s [6] description of genetic linkage in 1913; and the discovery of the double-helix structure of the DNA by Crick and Watson [7]. However, the transformation of the medical genetics landscape came after the breakthroughs in the 1970–80s which include the development of Sanger sequencing in 1975 [8]; the discovery of naturally occurring polymorphic DNA markers in 1980 [9]; the genetic mapping of the very first disease gene, Huntington disease, in humans in 1983 [10]; and the discovery of positional cloning of disease genes in 1986 [11]. These discoveries facilitated the genetic mapping and identification of many previously established Mendelian diseases such as Huntington disease, cystic fibrosis, and Duchenne muscular dystrophy. However, the easing and acceleration of identifying specific genomic alterations associated with each Mendelian diseases case came only after the completion of the human genome project. The sequences of almost all the genetic content of the human genome were successfully determined, stored, and rendered publicly available at the end of the project in 2003 [12]. Today, we see rapid advancements on the forefront

of hunting Mendelian disease-causing genes, owing to the availability of open-access whole-genome/exome sequence data, through the innovations and cost reductions in sequencing protocols and next-generation sequencing (NGS) technologies. Moreover, we can now rapidly analyze the whole-genome sequences of individuals in one run.

Genetic linkage and association analyses are the most common methods in use for the identification of the genetic background of diseases. While genetic linkage analysis investigates the relationship between the transmission of a locus and the disease within families, genetic association analysis or genome-wide association studies (GWASs) investigate the relationship between a specific allele and the disease within the population, focusing on a larger individual sample size [13]. GWASs are the method of choice for analyzing common variants in complex traits using single nucleotide polymorphisms (SNPs) as marker for specific loci [14]. However, rare variants which could be responsible for a considerable proportion of complex diseases may be overlooked by GWAS. Although linkage analysis is a relatively older approach than association, it is still a powerful method to detect disease-causing rare variants using family-based data [15]. Furthermore, the increased availability of whole-exome sequencing (WES) and whole-genome sequencing analyses provides the opportunity of parallel implementation of sequencing together with linkage. In addition, combined analysis of linkage and WES is a powerful approach to diagnose clinically rare and recessively inherited conditions in consanguineous pedigrees.

The long tradition of consanguineous marriages and endogamy has been practiced since the evolution of human, with approximately 20% of the world population currently living in communities with a preference for consanguinity. There is a transverse belt of populations with 20–50% incidence rate of consanguinity that runs from Pakistan, Afghanistan, South India, and Morocco [16]. Additionally, according to a recent study by the Statistical Institute of Turkey, the frequency of consanguineous marriages in Turkey was revealed to be 18.5%, with 57.8% majority being first-cousin marriages [17]. It has been established that consanguineous marriages dramatically increase the probability of homozygosity at any genetic loci in the offspring by identical by descent, and hence very rare autosomal recessive disorders are predominantly observed in these families [18, 19]. Thus, this review focused on consanguineous family-based identification strategies for rare variants involved in disease etiology. Our previous studies demonstrated that linkage

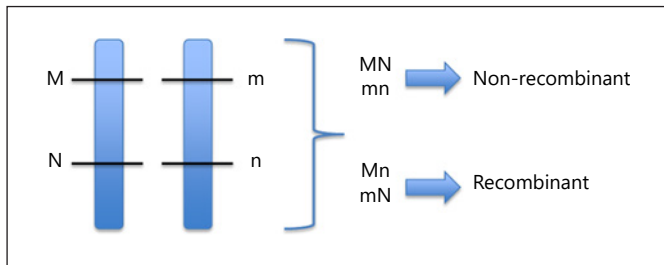


Fig. 1. The exchange of different chromosomal segments during meiosis results in recombinant regions.

and NGS analysis can unbiasedly detect ultra-rare variants in numerous genes specific almost to single consanguineous families [20, 21].

Linkage Analysis

Mendel’s law of “independent assortment” states that traits are inherited independent of each other. However, we now know that there are loci in the genome that display the tendency to be passed on together over generations. This tendency of loci positioned close to each other on the same chromosome being inherited together as an intact unit is termed genetic linkage. The independent or intact segregation of loci is established in the first prophase phase of meiotic cell division, depending on whether there is any exchange of genetic material between homologous chromosomes (*crossing over*) [22]. This process gives rise to gametes that carry recombinant and nonrecombinant chromosomes (Fig. 1). Thus, a near-unlimited number of new allelic combinations can be seen through this exchange. The frequency of the recombination that occurs is called the “recombination fraction” and is denoted by θ [23].

Linkage analysis is statistically a model-dependent (parametric) method. It involves the parametric statistical analysis of genomic regions that have the tendency to be inherited together with genetic markers (microsatellite, restriction site polymorphisms, SNPs, etc.) of known positions. Although microsatellites and restriction site polymorphisms were used as genetic markers for linkage analysis in the past, SNPs are now the marker of choice [24]. The main advantages SNPs present are their exceeding abundance in the human genome and their convenience for computational analyses [25]. The statistical value used in these linkage analyses is the logarithm of odds ratio (LOD) score.

Whole-Genome SNP Genotyping

Genetic variations are divided mainly into 3 groups based on the size of the affected sequence: short genomic variations, structural variations, and numerical chromosome anomalies [26]. Short genomic variations known as SNPs or single nucleotide variations are the most common type of genetic variations. These variations consist of an addition (insertion), a loss (deletion), or a changing (substitution) of single nucleotides at specific locations [27]. Copy number variations (CNVs) are a group of structural variations and refer to an intermediate-scale genetic change. CNVs include additional copies as named gains (duplications) or losses (deletions) of a DNA segment in the range from 1,000 base pairs (1 Kilobase) to several million nucleotides [28]. Variations that produce a change in chromosome number because of chromosome missegregation in cell division are termed numerical chromosome anomalies (e.g., aneuploidi or euploidi) [29].

SNP array technology has many advantages over conventional karyotyping techniques. These include the removal of DNA material limitation from isolated cells, making the tissue and/or cell culture step obsolete/redundant. Furthermore, samples for the SNP array do not need to be marked separately and hybridized with a control sample as is done with the array comparative genomic hybridization method, and each SNP array sample can be studied separately and independent of each other [30].

SNP genotyping method is applied with different commercial platforms using a microchip which includes attached oligo pieces ranging from a few thousand to millions, giving information from the millions of SNPs dispersed within the genome. Thus, providing the raw data required for population genetic studies, linkage analyses, and GWAS [31]. Consequently, the frequency distribution which indicates the genotype of each SNP and the total signal intensity data for each SNP is generated. This two-way analysis of the SNP array technique provides both the whole-genome genotype data and the CNV value of each SNP. After the DNA hybridization-based laboratory method, the obtained raw data is visualized with a software. Bioinformatics analyses are then conducted with consideration of B allele frequency (BAF) and LogR fraction (LRR) as mentioned below. Genotyping of a single polymorphism can result in 10 possible pair combinations: AA, TT, CC, GG, AT, AC, AG, TC, TG, and CG. However, due to the logic of the binary code of computers, the genotype outputs are only shown as AA, AB, or BB when calculating BAF. In normal condi-

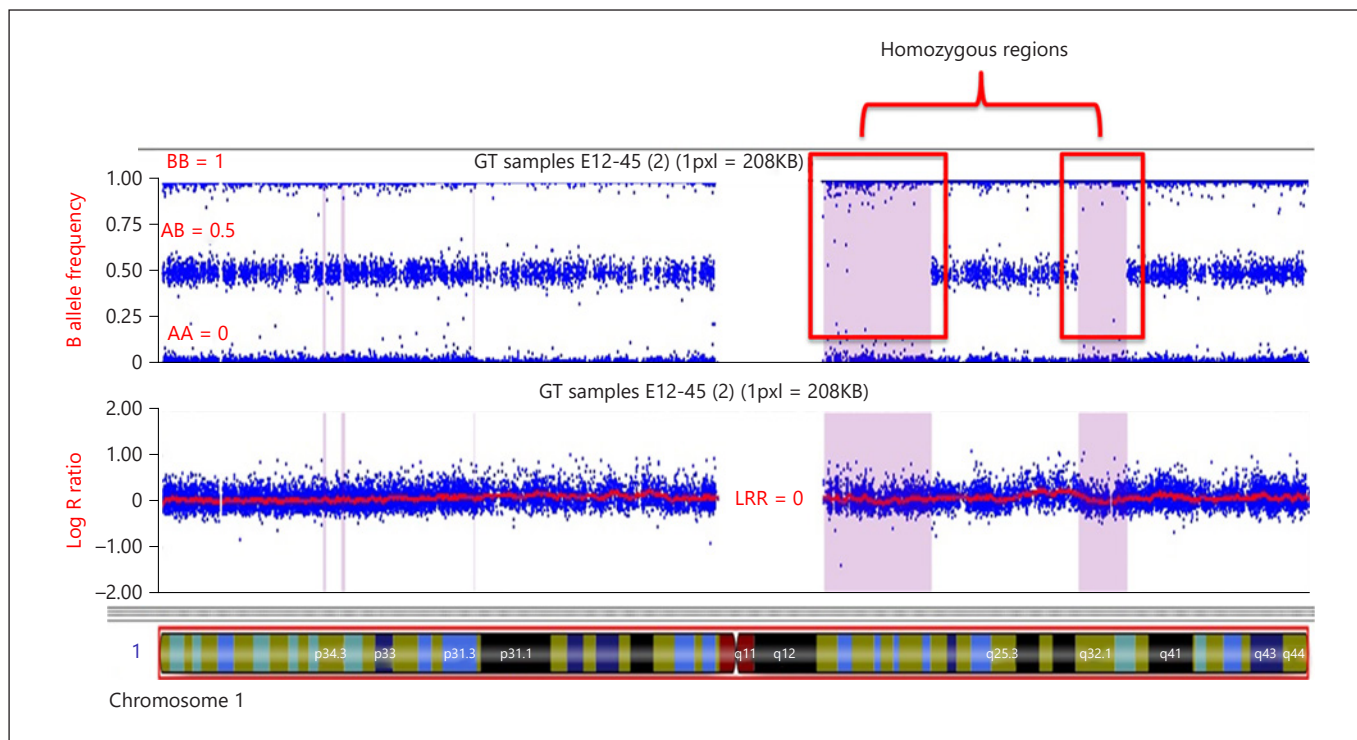


Fig. 2. LOH regions visualized by homozygosity detector. Purple boxes indicate true autozygous regions. A LRR of 0 indicates 2 identical copies of the same homolog in copy-neutral LOH. LRR, Log R ratio; LOH, loss of heterozygosity.

tions, the BAF profile is a value distributed at a certain rate within the range of 0–0.5 and 1. If the BAF value is outside the values which specified above, there is either true homozygosity/autozygosity or a loss/gain is suspected in that region of the genome. This situation can be clarified by examining the LRR. If a difference in BAF value is found, the signal intensity from the LRR data is normalized for each marker, and the region of focus is examined as the location of heterozygosity for each of the markers. Any loss or gain in the genome can also be detected by deviations from normal LRR values (Fig. 2) [30].

One exception to these results which are derived from the distribution of autosomes in a normal genome comes from increases or decreases in the number of copies in the genome. Increases (copy number 3 or 4) or decreases (copy number 0 or 1) of CNVs in the genome cause BAF distributions to change. Another exception is related to the X chromosome. X chromosome analysis in females is the same as autosomes for they normally contain 2 X chromosomes, while males under normal conditions are hemizygous for the X chromosome, making the CNV val-

ue of their X chromosome 1. However, pseudoautosomal regions (PAR1 and PAR2) of X and Y chromosomes show homology, and they can be evaluated just like autosomes. Thus, males are only hemizygous in terms of X chromosome regions excluding PARs (Table 1). Additionally, BAF and LRR values can also be utilized to analyze CNV regions.

LOD Score

Mendel’s law of “independent assortment or segregation” applies to genomic regions in different chromosomes or regions very distant from each other on the same chromosome; however, it is not always valid for regions relatively close to each other on the same chromosome. The value of θ is a measure of the frequency of recombination between 2 regions on the same chromosome and thus, an indicator for the proximity of these chromosomal regions. The θ value varies between 0 and 0.5 (Fig. 3). If the value of θ is 0, the 2 regions are very close to each other, indicating that they are linked and that there is no recombination

Table 1. CNV genotypes and BAF values in terms of autosomes and chromosome X

Copy, n	Definition	CNV genotypes	BAF values
2	Normal	AA, AB, BB	0-0.5-1
0	2 copies loss	Genotype data cannot be obtained	-
1	1 copy loss	A, B	0-1
2	LOH	AA, BB	0-1
3	1 copy gained	AAA, AAB, ABB, BBB	0-0.33-0.66-1
4	2 copies gained	AAAA, AAAB, AABB, ABBB, BBBB	0-0.25-0.5-0.75-1
<i>Chromosome X (excluding PAR1 and PAR2 regions)</i>			
1	Male (reference)	A, B	0-1
2	Female (reference)	AA, AB, BB	0-0.5-1

CNV, copy number variations; BAF, B allele frequency; LOH, loss of heterozygosity.

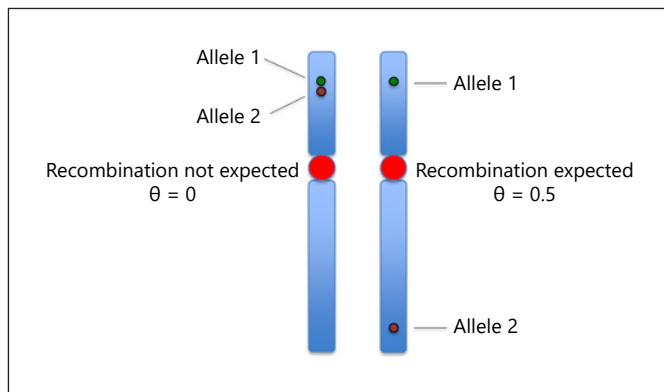


Fig. 3. The value of θ denoting to 0 means the 2 regions are very close to each other which indicates they are linked and that there is no recombination. The value of θ denoting 0.5 means the probability of assortment of distant loci on the same chromosome is 50%.

between them. If there is recombination between the regions, the value of θ differs between 0 and 0.5 with the distance of the 2 regions to each other being inversely proportional to the value of θ . Genetic maps are created based on this principle through systematic measurements of θ values for genomic markers such as SNPs and microsatellites. The unit of measurement used for genetic mapping in centimorgan (cM) with 1 cM corresponding to 1% recombination frequency ($\theta = 0.01$) and an area of 1 cM corresponding to an average of 1 Megabase in physical maps. However, a region of 1 cM can be less or >1 Megabase in physical maps depending on whether the recombination rate of that region is high or low, respectively. Based on the

principle of genetic linkage, one can examine genetic marker(s) that are close to the genome region associated with the trait of interest in families suffering from genetic diseases/disorders. For this purpose, healthy and/or affected parents, children, and other relatives are genotyped either for selected markers or for markers spread throughout the genome. The results obtained, with consideration of the disease phenotype and model, are assessed for whether there is meiotic recombination between the trait of interest and genetic markers. Any marker or group of markers found to be associated with the disease trait is an indication that the disease-related genomic disruption is either very close or falling somewhere in between these groups of markers. This observational approach that specifically marks the location of the disease-related genetic defects in the genome is called “Linkage Analysis” and is calculated by a statistical significance-likelihood ratio test called “LOD score analysis.”

The LOD score analysis examines the linkage between possible disease-phenotype-associated candidate regions located in unknown locations throughout the genome and a marker genotyped by the investigator. Since the exact location of the disease-associated region is not known, all values that θ can take should be examined and analyzed. In other words, LOD score is a function of θ , and for all θ values ($0 \leq \theta \leq 0.5$), the probability of recombination occurring is compared to the probability of it not occurring. The logarithm of the base 10 of the likelihood ratio is taken when calculating the LOD score which leads to the results being expressed as integers or fractions, unlike conventional p values. Due to LOD score being a function analysis, the LOD score result is reported together with the θ value used in the function. In a LOD

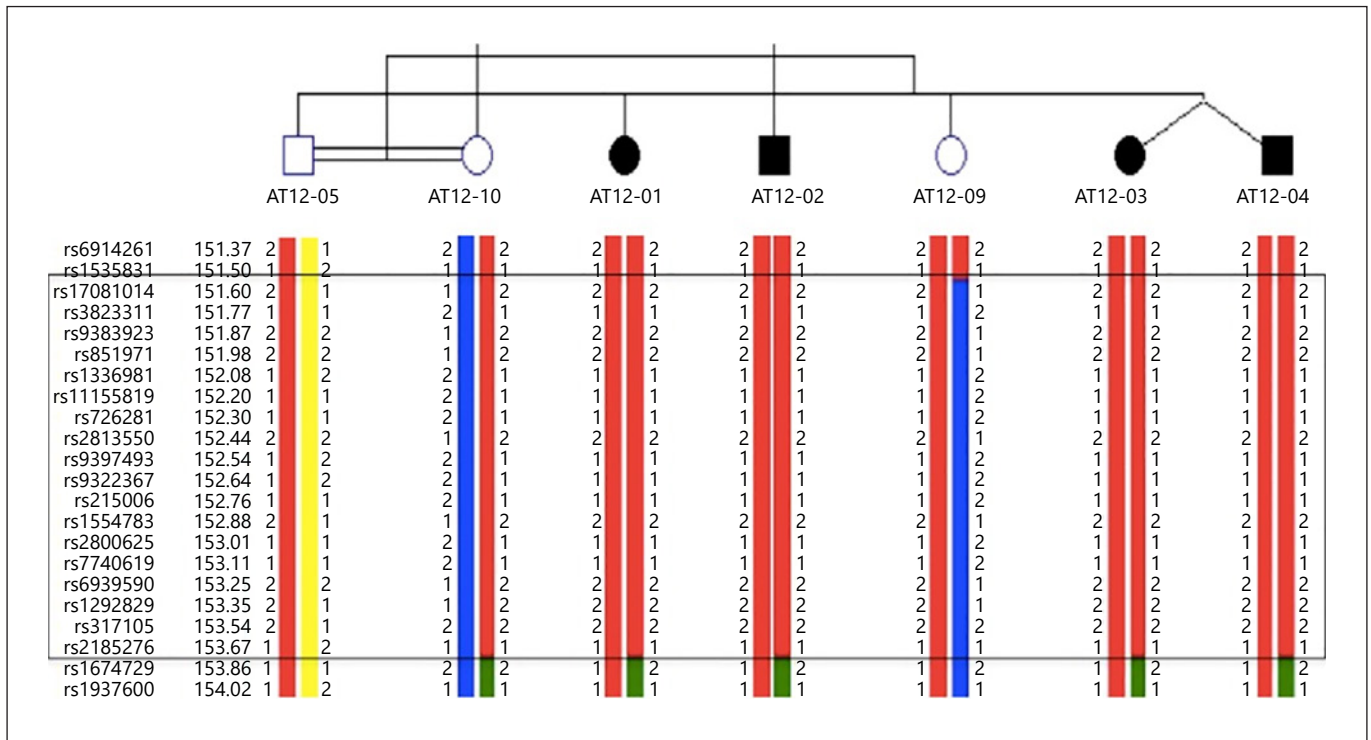


Fig. 4. The shared autozygous regions of 4 affected individuals visualized by the HaploPainter program. This shared region is not autozygous in parents and unaffected siblings [21].

score calculation, a graph encompassing all θ values is drawn, and the graph's peak is evaluated by considering θ values.

$$LOD\ score = Z = \log_{10} \left(\frac{\theta^r (1-\theta)^{nr}}{0.5^{r+nr}} \right)$$

r , recombinant, nr , nonrecombinant.

If the LOD score analysis result is -2 or less, linkage is rejected while a result of 3 , or more indicates linkage. LOD score analyses are divided into two-point LOD score analysis and multipoint LOD score analysis. In the two-point LOD score analysis, recombination between 2 regions associated with a genetic marker and a disease are observed. In the multipoint LOD score analysis, the relationship between relative genetic maps obtained through multiple marker usage and the disease-associated region is assessed. There are many open-access software, for example, Allegro [32], GeneHunter [33], Merlin [34], and SimWalk [35] which can perform the analyses. There are also interfaces such as easyLinkage [36], to collect different analysis platforms into one program.

Homozygosity Mapping

Homozygosity mapping is a model-dependent method that is used specifically in consanguineous pedigrees with recessive diseases. In this approach, haplotypes can be explained as the genotype map of a particular region on chromosomes. Homozygosity mapping is used to search for homozygous/autozygous haplotypes within particularly consanguineous families affected by a recessive disease [37]. Alkuraya [38] reported in 2010 consanguineous families permit the "reunion" of ancestral chromosomal segments in a pattern referred to as "autozygosity," which is essentially a special form of homozygosity. According to the recessive inheritance model, homozygous or autozygous haplotypes should be in a heterozygous state in the parents, heterozygous or wild type autozygous state in the healthy siblings, and the affected individuals should be in an autozygous state. There is a wide variety of computer programs that allow the detection of autozygous regions. These can be listed as follows: a commercial software called Homozygosity Detector plugin of Illumina Genome Viewer (Fig. 2), HaploPainter software (Fig. 4) [39] which is an open-source like Genehunter, Al-

Table 2. Databases used in analyses

Software	URL
Ensembl	http://www.ensembl.org/
NCBI	http://www.ncbi.nlm.nih.gov/genome/
UCSC	http://genome.ucsc.edu/
gnomAD	https://gnomad.broadinstitute.org/
DiscovEHR	http://www.discovehrshare.com/
Bravo/TOPMed	https://bravo.sph.umich.edu/freeze8/hg38/
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/
dbVar	http://www.ncbi.nlm.nih.gov/dbvar/
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/
OMIM	http://www.ncbi.nlm.nih.gov/Omim/
MutationTaster	http://www.mutationtaster.org/
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
SIFT	http://sift.jcvi.org/
CADD	https://cadd.gs.washington.edu/
Primer3plus	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

SNP, single nucleotide polymorphism.

legro, and SimWalk, whole-genome association, and population-based linkage analyses toolset named the PLINK [40] can be used to detect autozygous regions via identical by descent estimation. In addition “AutoZplotter” is another alternative software which takes a variant calling format (VCF) file as input and allows for manual inspection of long runs of homozygosity/autozygous regions, and is highly used/accessed in the clinical genetics field [41]. After the boundaries of autozygous regions are determined through genetic markers, a detailed analysis of the region for the associated phenotype is started. The determined autozygous region from linkage analysis can be used as a filter in WES analysis; therefore, this collective effort provides a significantly powerful step to identify the disease-causing gene/variant.

Whole-Exome Sequencing

WES is the targeted sequencing of the protein-coding regions of the genomic DNA and is applied by using NGS technologies [42]. The development of coupling targeted capture and massively parallel DNA sequencing has provided the means to analyze almost all of the coding sequences present in an individual human genome. This has allowed WES to become an efficient approach in situations where conventional methods fail to identify genes underlying Mendelian diseases [43]. However, even under circumstances where conventional methods offer

success, the parallel application of WES can still be more advantageous for accelerated discovery [44, 45]. In addition, this method can successfully work even with relatively small sample sizes acquired from a few family members. Nevertheless, fragment sequencing has higher sequencing error rates than conventional sequencing; thus, results require validation which Sanger sequencing can aptly provide [46].

WES analysis is carried out in 5 basic steps: raw data quality assessment, preprocessing, alignment, post-processing, and variant analysis (detection, annotation, and prioritization) [47]. These steps are completed through special software on the basis of compromise pipelines, and the results obtained are examined in VCF. VCF is a general format for storing genomic sequence data together with Rich Annotations [48]. Interpretation of genomic variants has been defined in more detail in the standards and guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology [49].

The first step of variant filtration and prioritization is to exclude less reliable variants, such as variants in low coverage or low quality according to quantitative scores. The variants are then restricted by the minor allele frequency threshold, considering common variants are less likely to cause disease than rare variants. The final step is to prioritize variants relative to the phenotypes of individuals. This step also includes *in silico* variant effect predictions. Possible effects of the determined variants are

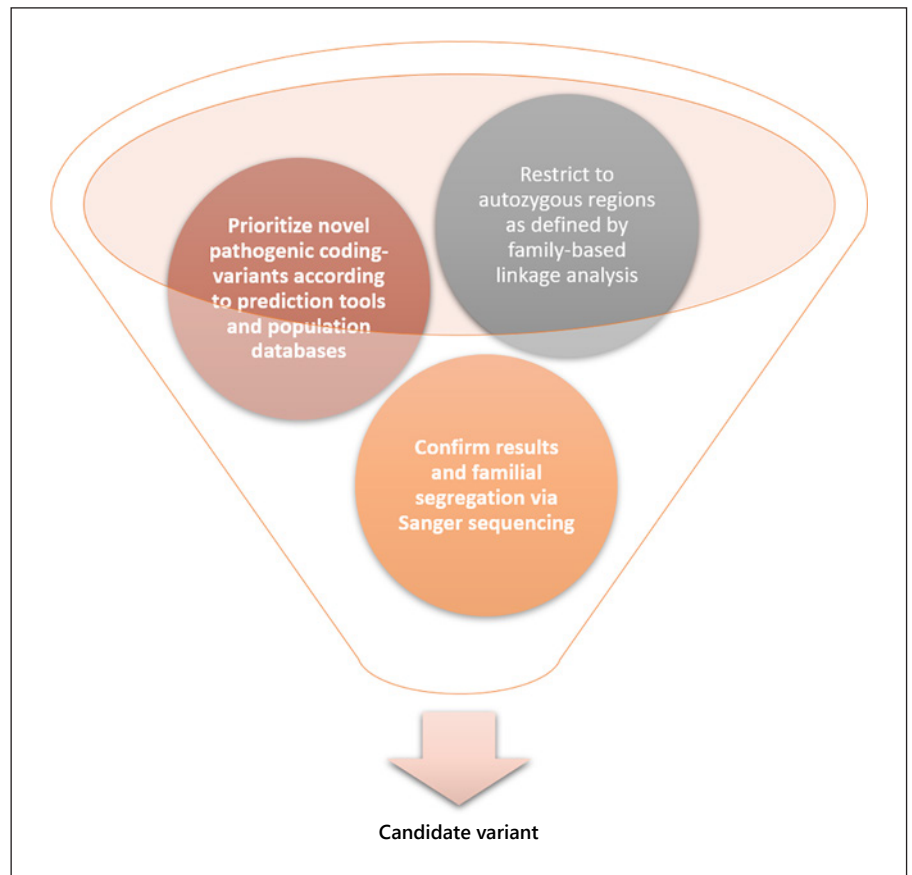


Fig. 5. Workflow of filtering approach in whole-exome sequencing to detect candidate genes by family-based data.

evaluated on the computer with online prediction tools such as MutationTaster, PolyPhen-2, SIFT, and CADD (Table 2). Additionally, establishing the inheritance pattern of disease from family history or previous studies is another useful step for the analyzes [50]. As previously mentioned, using the restricted region from linkage analysis as a filter provides significant advantages for determining candidate variants. In conclusion, these strategies are successful in detecting inherited or novel deleterious variants in familial diseases (Fig. 5).

Segregation Analysis and Population Screening

The candidate variants that are identified are further analyzed through Sanger sequencing for confirmation of their relation to the phenotype. The primers are designed with the appropriate computer programs such as Primer3Plus [51] for the following PCR step and are controlled through in silico PCR (<http://genome.ucsc.edu/>). After the variants are determined with Sanger se-

quencing of the amplified region, visualization of variants in chromatogram is done through CLC Workbench (<https://digitalinsights.qiagen.com>) or similar programs.

In addition, population screening should be performed to determine the minor allele frequency threshold and to demonstrate that the candidate pathogenic variant is not observed in any healthy individuals. For this purpose, using publicly available reference cohorts is key for the success of rare-variant analyses. gnomAD, DiscovEHR, and Bravo/TOPMed reference cohorts are publicly available for population screening which allow access to over 250,000 exomes (Table 2) [52].

Due to the nature of genomics studies, we cannot easily assume that the variant detected from the genomic analysis is the exact reason of the disease. The functional characterization of identified candidate variants must be done using in vitro and in vivo methods for confirmation. Afterward, the relationship between the variant and the disease will be unveiled clearly.

Functional Characterization of Candidate Variants

Genetic variants within the exome may alter protein function in many ways, such as enzymatic activity, thermodynamic stability, ligand binding, cellular localization, cell signaling, etc. [53]. Herein, we will discuss some of the functional studies. Mutagenesis studies are mostly preferred to determine various functions of variants, especially in the coding region. These studies include in vitro functional assays and genetic manipulation on model organisms. Patients own cells, specialized cell lines, peripheral blood cells, iPS cells (termed in vitro studies), or model organisms such as *Drosophila melanogaster* and *Danio rerio* (Zebrafish) (termed in vivo studies) can be used in these experimental approaches [54]. As an example, for genes that encode enzymes, an enzyme reporter assay can be used, and for proteins with unknown functions or if the measurement of interest is protein stability, then quantifying the protein levels may be preferred [55]. CRISPR/Cas9 system is the most popular genome-editing tool owing to its user-friendliness and rapidity [56]. This system can be used for several genomic regions editing, especially for loss of function mutations, and is used for targeting functional protein domains [57]. The RNA interference method is also preferred for sequence-specific gene silencing through double-stranded RNAs which initiates mRNA degradation [58]. Loss of function or dysfunction of ion channels can underlie human neurodegenerative diseases such as Alzheimer's and Parkinson disease, and in these cases, electrophysiological methods such as Patch-Clamp are the best choice to characterize the effects of variants [59]. Consequently, any specific characterization method can be designed for each candidate variant. When selecting the most appropriate method, it is important to consider the function of the gene and the possible effects of the variant. Additionally, the researcher's experience, laboratory facilities, and literature review are considered.

Conclusions

One of the earliest known scientific inquiries to understand the potential association between diseases and disorders with consanguinity was proposed by Darwin. He suggested the addition of questions about cousin matrimony in the 1871 British census, however, his proposal to scientifically test whether consanguinity has any detrimental effects in humans was rejected [3]. We have come a long way from trying to ascertain whether dis-

eases have underlying inherited genetic value or whether consanguinity is indeed risky. To date, many disease-causing variants have been successfully detected through the parallel implementation of linkage analysis and WES in consanguineous pedigrees, and this dual approach has provided the chance for genetic counseling to many families. Furthermore, the identification and functional characterization of these causal variants have shed light on a number of molecular genetic processes and pathways and have transformed our understanding of basic biology, developmental biology, evolutionary biology, and human physiology. In addition, understanding the spectrum of allelic variation in human genes and revealing the demographic and evolutionary forces that shape variations within and among populations are important for medical and population genetics research. Such information is critical for defining the architecture of diseases and disorders, ultimately facilitating the interpretation of personalized disease risk profiles that can open the door to the development of predictive and targeted therapies.

Statement of Ethics

The authors have no ethical conflicts to disclose.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding this work.

Funding Sources

Funding sources are not applicable for this study.

Author Contributions

All the authors made substantial contributions to the work and approved the final version to be published. S.S. and K.K. were responsible for the concept of the work and participated in drafting and revising the manuscript; E.Y. participated in supervising and revising the manuscript critically for important intellectual content.

References

- 1 Darwin C, Kebleur L. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London: John Murray; 1859.
- 2 Darwin C. *The variation of animals and plants under domestication*. London: John Murray; 1868.
- 3 Jones S. *Darwin's Island: the galapagos in the Garden of England*. Abacus; 2009.
- 4 Mendel G. Experiments in plant hybridization. In: "Versuche über Pflanzhybriden." *Abhandlungen, Verhandlungen des naturforschenden Vereins Brünn*. 1865. Available online 1996.
- 5 Morgan TH. Sex limited inheritance in drosophila. *Science*. 1910;32:120–2.
- 6 Sturtevant AH. The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. *J Exp Zool*. 1913; 14(1):43–59.
- 7 Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737–8.
- 8 Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94:441–8.
- 9 Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980;32:314–31.
- 10 Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 1983;306:234–8.
- 11 Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, et al. Cloning the gene for an inherited human disorder: chronic granulomatous disease: on the basis of its chromosomal location. *Nature*. 1986;322:32–8.
- 12 Collins FS, Morgan M, Patrinos A. The human genome project: lessons from large-scale biology. *Science*. 2003;300:286–90.
- 13 Ferreira MA. Linkage analysis: principles and methods for the analysis of human quantitative traits. *Twin Res*. 2004;7:513–30.
- 14 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24.
- 15 Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet*. 2015;16:275–84.
- 16 Bener A, Mohammad RR. Global distribution of consanguinity and their impact on complex diseases: genetic disorders from an endogamous population. *Egypt J Med Hum Genet*. 2017;18(4):315–20.
- 17 Kaplan S, Pinar G, Kaplan B, Aslantekin F, Karabulut E, Ayar B, et al. The prevalence of consanguineous marriages and affecting factors in Turkey: a national survey. *J Biosoc Sci*. 2016;48:616–30.
- 18 Erzurumluoglu AM, Shihab HA, Rodriguez S, Gaunt TR, Day IN. Importance of genetic studies in consanguineous populations for the characterization of novel human gene functions. *Ann Hum Genet*. 2016;80:187–96.
- 19 Thompson EA. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*. 2013;194:301–26.
- 20 Ugur Iseri SA, Yucesan E, Tuncer FN, Calik M, Kesim Y, Altiokka Uzun G, et al. Biallelic loss of EEF1D function links heat shock response pathway to autosomal recessive intellectual disability. *J Hum Genet*. 2019;64:421–6.
- 21 Yucesan E, Ugur Iseri SA, Bilgic B, Gormez Z, Bakir Gungor B, Sarac A, et al. SYNE1 related cerebellar ataxia presents with variable phenotypes in a consanguineous family from Turkey. *Neurol Sci*. 2017;38:2203–7.
- 22 Pulst SM. Genetic linkage analysis. *Arch Neurol*. 1999;56:667–72.
- 23 Taylor EW, Xu J, Jabs EW, Meyers DA. Linkage analysis of genetic disorders. *Methods Mol Biol*. 1997;68:11–25.
- 24 Weissenbach J. Microsatellite polymorphisms and the genetic linkage map of the human genome. *Curr Opin Genet Dev*. 1993;3: 414–7.
- 25 Strachan T, Goodship J, Chinnery P. *Genetics and genomics in medicine*. Oxfordshire, UK: Taylor & Francis; 2014.
- 26 Alzu'bi AA, Zhou L, Watzlaf VJM. Genetic variations and precision medicine. *Perspect Health Inf Manag*. 2019;16:1a.
- 27 Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*. 2000;407:513–6.
- 28 Eichler EE. Copy number variation and human disease. *Nat Educ*. 2008;1(3):1.
- 29 Nussbaum RL, McInnes RR, Willard HF. *Thompson & Thompson genetics in medicine e-book*. Elsevier Health Sciences; 2015.
- 30 Yucesan E. *Whole genome analysis in familial epilepsies*. Istanbul: Institute of Health Science, Department of Genetics, Istanbul University; 2016.
- 31 McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemes J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*. 2008;40:1166–74.
- 32 Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingólfssdóttir A. Allegro version 2. *Nat Genet*. 2005;37:1015–6.
- 33 Nyholt DR. GENEHUNTER: your "one-stop shop" for statistical genetic analysis? *Hum Hered*. 2002;53:2–7.
- 34 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002;30:97–101.
- 35 Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet*. 1996;58:1323–37.
- 36 Lindner TH, Hoffmann K. easyLINKAGE: a PERL script for easy and automated two-/ multi-point linkage analyses. *Bioinformatics*. 2005;21:405–7.
- 37 Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 1987;236:1567–70.
- 38 Alkuraya FS. Autozygome decoded. *Genet Med*. 2010;12:765–71.
- 39 Thiele H, Nürnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics*. 2005;21:1730–2.
- 40 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
- 41 Erzurumluoglu AM, Rodriguez S, Shihab HA, Baird D, Richardson TG, Day IN, et al. Identifying highly penetrant disease causal mutations using next generation sequencing: guide to whole process. *Biomed Res Int*. 2015;2015: 923491.
- 42 Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12:745–55.
- 43 Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet*. 2010;42: 13–4.
- 44 Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010;42:30–5.
- 45 Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*. 2010; 467:207–10.
- 46 Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Front Med*. 2014;8:42–57.
- 47 Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, et al. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform*. 2014;13:67–82.
- 48 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. Genomes project analysis G: the variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
- 49 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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;17:405–24.

- 50 Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, Nicolae DL, et al. Finding disease variants in Mendelian disorders by using sequence data: methods and applications. *Am J Hum Genet.* 2011;89:701–12.
- 51 Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 2007;35:W71–4.
- 52 Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet.* 2019;20:747–59.
- 53 Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 1973;181:223–30.
- 54 Rodenburg RJ. The functional genomics laboratory: functional validation of genetic variants. *J Inherit Metab Dis.* 2018;41:297–307.
- 55 Ipe J, Swart M, Burgess KS, Skaar TC. High-throughput assays to assess the functional impact of genetic variants: a road towards genomic-driven medicine. *Clin Transl Sci.* 2017;10:67–77.
- 56 Gaj T, Sirk SJ, Shui SL, Liu J. Genome-editing technologies: principles and applications. *Cold Spring Harb Perspect Biol.* 2016;8:a023754.
- 57 Campenhout CV, Cabochette P, Veillard AC, Laczik M, Zelisko-Schmidt A, Sabatel C, et al. Guidelines for optimized gene knockout using CRISPR/Cas9. *Biotechniques.* 2019;66:295–302.
- 58 Han H. RNA interference to knock down gene expression. *Methods Mol Biol.* 2018;1706:293–302.
- 59 Chen CC, Cang C, Fenske S, Butz E, Chao YK, Biel M, et al. Patch-clamp technique to characterize ion channels in enlarged individual endolysosomes. *Nat Protoc.* 2017;12:1639–58.