American Journal of Biology and Natural Sciences
https://biojournals.us/index.php/AJBNS
ISSN: 2997-7185

Genetic Variation and its Impact on Human Diseases

Aya Ismail Ibrahim, Athari Ismail Ibrahim
University of Al-Qadisiyah, College of Science Department of Biology

Noor Abdul Kareem Mohsin
University of Baghdad, College of Science Department of Biology

Noor Nabeel Hussain
University of Technology, Department of Applied Sciences_Biotechnology

Ayat Saadi Hussien
College of Science, University of Babylon, Department of Biology

Received: 2024, 15, Mar
Accepted: 2025, 21, Apr
Published: 2025, 31, May

Annotation: Variation in the human genome is common; studies suggest that on average, any two individuals differ by around 1 in 1000 base pairs, that is 3 million variations. While millions of variants are known, the vast majority are common in the general population and are unlikely to affect an individual's risk of developing disease. However, the minority are rare and may predispose to disease which is likely dependent on the type of mutation, its position within the gene, the proportion of individuals affected, and the availability of suitable nuclear or animal models. Most human genetic disease is due to a single mutation in a protein-coding gene and therefore the search for the genetic cause is usually simplified by searching for protein-altering mutations in coding exons and adjacent splice sites. To date, mutations in ≈4000 human genes have been associated with disease. The focus here will be on human disease, although much of the research that defines understanding comes

from the study of animal models that share similar or related genes . The human genome comprises roughly 3 billion bp of DNA. The majority of this DNA is present within the nucleus, as chromosomes, but there is also a small amount of DNA in the mitochondria; this organelle has a unique circular genome inherited maternally that encodes proteins integral to cellular respiration. Most individuals possess 23 pairs of chromosomes (46 total), so much of the DNA content is present in two copies, one from our mother and one from our father. Each parent contributes one complete set of 23 chromosomes, so the total number is maintained across successive generations. Each human somatic cell possesses two copies of each chromosome, making 46 in total. The following are the basic features of the human genome. The human nuclear genome encodes roughly 20000 protein-coding genes, which typically consists of both protein-coding (exon) and non-coding (intron) sequences. These transcript types can be classified by biogenesis as small nuclear (sn)RNA, small nucleolar (sno)RNA, micro (mi)RNA, and long non-coding (linc)RNA.

1. Introduction to Genetic Variation

Variation (genetic differences between individuals) is a fundamental property of every sexually reproducing species. While two individuals may have a high degree of similarity, it is precisely those rare differences that drive evolution. Genetic diversity is also the foundation of the study of genetic diseases, which are driven by variation in DNA sequence. Some genetic variants play a direct and larger role in causing disease and pathological states. Complex diseases (or common diseases) are more subtle and remain poorly understood because there exist many genetic variants that exert a small effect on the phenotype. This gray zone between variation and phenotype where the understanding of the genetics behind diseases is limited has been termed the “missing heritability”, a term used in contrast to the “common disease-common variant” hypothesis.

A large fraction of clinical variation remains to be explained and small, highly polymorphic,

ancient, and recently arisen variants are likely to contribute to the genetic risk of complex diseases. Geneticists have made significant progress in understanding the genetics behind many human diseases. However, the genetics of complex diseases is still poorly understood. The majority of human genetic variation is common and in common population the “one variant to one disease” hypothesis becomes increasingly problematic [1]. The release of the complete human genome sequence in 2001 has improved our understanding of the patterns of human genome diversity and its linkage to human complex diseases [2]. In recent years, rapidly dropping sequencing costs and advances in computational power have made it feasible to obtain on the order of 1000 human genomes, making it possible to systematically study the genetic basis of human complex diseases. To study genetic diversity of the human genome at population level, the HapMap project was initiated to investigate the genetic differences on both inter- and intrapopulation levels.

This was made possible by the introduction of advanced technologies such as Chip-based genotyping and next-generation sequencing techniques. Today, multiple high throughput SNP arrays are commercially available. Ticket samples have been prepared as a biobank of human genetics to create a Chinese HapMap dataset. This massive amount of population genetic information, along with new tools for automated genotyping and assembly of large consortia, has been utilized in Genome Wide Association Studies (GWAS) on various human diseases. All of this work has revealed fascinating relationships between human population genetics and simple monogenic traits. However, there is also great concern regarding the impact of pharmaceutical industries on human genetics results in a changing world. To address these concerns, efforts are underway to analyze the epidemiological impact of GWAS in a globalized and rapid changing world.

2. Types of Genetic Variation

Variants are classified as benign (not associated with disease) or pathogenic (associated with disease); these are termed ‘variants of uncertain significance’ or VUS. Where two (or more) different versions of a DNA sequence exist in the population, these are referred to as ‘alleles’. By analysing many human genomes, the frequency at which a particular variant occurs in the population can be calculated, often expressed as the ‘minor allele frequency’ or MAF. Where the MAF is at least 1%, a variant can be called a ‘polymorphism’ (i.e. a common variant). Estimating the frequencies at which variants occur in, as well as the structures of, distributions of healthcare populations is key to improving diagnostics, treatments and better defining at-risk individuals. The most frequent variants in the human genome are substitutions that affect only one base pair, referred to as single nucleotide variants (SNV) or as single nucleotide polymorphisms (SNP) depending upon the MAF. It has been estimated that there are at least 11 million SNPs in the human genome. Insertions or deletions of less than 1000 bp are also relatively common in the human genome. 8-phenyl-2,3,3a,4,5,6-hexahydroquinolin-4-ol are commonly referred to as indels, whereas larger deletions or duplications of more than 1000 bp in size are referred to as copy number variants (CNV). There are believed to be 900 000 indels and 62000 CNVs in the human genome, with a total difference of about 1.5% between individuals [2]. The most common of the large-scale structural variants are the 500–1000 bp Alu repeat elements, which are believed to have arisen from a single insert in the genome of a common primate ancestor about 60 million years ago. The Alu repeat family is an example of a transposable element of which there are many in the human germline, and which can change their position in the genome. Repeat elements constitute about 45% of the human genome, but because of their repetitive nature and transient style, they are not generally polymorphic in normal individuals.

2.1. Single Nucleotide Polymorphisms (SNPs)

Most common diseases and many drug responses have been shown to be influenced by inherited differences in our genes. Therefore, it may be possible to significantly improve the prevention and management of these diseases by establishing a better understanding of their genetic basis. Much

of the genetic variation between individuals lies in differences known as single-nucleotide polymorphisms (SNPs): in any given DNA sequence, a single base (A, C, G or T) is swapped for an alternate one. When two versions of a base are common in the general population, or at DNA positions flanking a variant, the two forms are genetically polymorphic [3]. Two sequence versions at a variant position have frequencies in the general population greater than 1%. As SNPs constitute the bulk of human genetic variation, they can be exploited as 'markers' to track inheritance of genes in traditional family-based linkage studies of Mendelian diseases. Detectable linkage suggests the location of disease genes, between two and 25 million base pairs, and this directs efforts to their identification. Subsequently, because of the intricacies of the biological pathways involved in complex disease, and because hundreds of thousands of genetic variants across the genome are likely to be responsible for any one disease, this approach becomes impossible for common diseases.

At a tractable scale, by accumulating thousands of individuals (more than 10 000), most populations will be provided with detailed common disease SNP maps of their own variant gene type. Using these SNP maps and sampling thousands of individuals, it should eventually be possible, by epidemiological association, to test for the influence of these common variants on the susceptibility to common diseases, for which 1-25% of weekends these SNP maps will be utilized in the burgeoning polygenic tendency risk prediction fields. Other work with similar tractability should also reveal how these SNPs influence extraordinary variability in the pharmacogenomics or responses to drug or vaccine treatment. The SNP Consortium was created in 1999 to discover SNPs that would be made freely available to all researchers. Ultimately, it is hoped that 1-3 million human SNPs will be identified and the methods for their reliable assay made widely available. Much of this information, and accompanying software, is being made publicly available, or can be obtained directly from the authors. Thus, with the available genetic map that consists of SNPs dispersed every 50-300 kilobases across the genome, it is hoped that the necessary genetic tools will be available for the 3 billion base pair human genome to be studied with the same intensity of sophistication as has been the 5 million base pair Drosophila genome.

2.2. Insertions and Deletions (Indels)

Insertions and deletions (indels, i.e., additions or losses of bases) can be as small as one base, but they can also be more than a hundred bases long. Longer indels can have dramatic phenotypic effects through changes in reading frames and novel-stop codon introduction. Muller’s phylogenetic approach has been employed to identify uniquely derived indels in 216 base-pair (bp), 297-bp, and 483-bp regions, and thus these indels are candidates for being related to the specialization of primates [4]. Pinpointing these specific mutations in a species that had diverged from the common ancestor is extremely difficult. Indel mutations might be more frequent than nucleotide substitutions in general, but their maintenance in the genomes of the descendant species might be more difficult. Indels in coding regions would require sampling a number of bases equal to or larger than the length of the reading frame to avoid frame-shift mutations, which could lead to a total loss of function. Indels might also be lethal in non-coding regions because of their possible effects on both transcription and translation. Detailed studies on the yeast and its closely related species showed that insertion and deletion mutations appear under varying circumstances, leading to complex and variable phenotypes such as differences in colony size or color and efficiency in the fermentation of different sugars. The insertion of transposable elements was detected as a frequent event in progeny generations. Besides transposon insertion mutations, it is suggested that mutations that lead to cellular imbalances such as indel mutations are often fixed in the lineage of one fertile individual, dominating the divergence in phenotypes from the parental stock [5]. To study the frequency of small SNPs, indels, and transitions in Macaca, full genome sequences were obtained through DNA sequencing of the six individuals and their relatives. To expand the number of species for comparison, was also included. Requirement of the biological resources of Macaca primates and a novel comparative genome analysis through enhancing methods and resources can fully characterize the evolutionary history of the macaque genome.

2.3. Copy Number Variations (CNVs)

Copy number variations (CNVs) are defined as portions of the genome that are duplicated or deleted; they can involve large DNA fragments, numbering >1 kb [6]. A refined genome map suggests that one copy of a certain chromosome segment may increase or decrease and, as a result, CNVs will be generated from its neighboring segment of a copy number of 1 that is seriatim adjacent to that change and from which the last copy of chromosome is derived. The old vs new copy segments attached to a varying number of atom constitutes CNVs. Thus, CNVs are classified as a structural variant (SV) in genomics, a copy number variant (CNV) in genetics that involves either an increase or a decrease in the number of copies of the DNA, a copy number amplification of a particular gene in cancer biology, and gain or loss of a chromosomal segment. In human diversity, this genetic difference has been termed CNV and defined on the basis of its impact—on arrangement, function, and phenotype at the biological level. Although different terms exist for CNVs, their common point exists in that the target DNA composition is structurally changed. Dosage imbalance of genes due to CNVs will lead to aberration of normal (physiological) cellular or organismal function and result in abnormal traits being defined as the genetic disease. For example, duplication of the gene Operations, duplication of five copies of the gene S1241, and deletion of the gene VAPB have been causally linked to the aberrations of the normal physiology and abnormal traits of the quantitative trait triglyceride and therewith metabolism disorder in mice, phenotype functional disease in pigs, and hemochromatosis, abetalipoproteinemia, motor neuron disease, etc., in human beings. Although technological limitations exist in CNV discovery, this genetic marker is estimated to cover ~12% of the human genome [7]. The application of next-generation sequencing has enabled more powerful detection of CNVs. Surprisingly, the genetic difference in DNA sequence has been found to be not entirely due to single nucleotide polymorphisms (SNPs), insertions, or translocations, with some sequences of unchanged length between the normal and abnormal genomes. Changes of roughly the same magnitude in length with respect to ingoing sequence would seem to have no detectable consequences on arrangement, entities, and behavior of the structure and function, and thus, this genetic change has now become a newly discovered genetic variant.

3. Mechanisms of Genetic Variation

Genetic variation occurs when there are differences in the DNA sequence among individuals and it is the raw material for evolution. Variation occurs at a range of different levels from large chromosomal and genomic evolution to a single base pair difference. It can occur within coding regions, non-coding regions and regulatory elements of genes. Genetic variation can also take the form of aberrant ploidy or changes in chromosome number. These alterations to the genome are frequently detrimental by disrupting normal cellular processes, causing diseases. The nature and consequences of genetic variation at different levels are dealt with here, with an emphasis on disease causation. They defined the chromosome as the carrier of heritable information, mapping the positions of mutant genes on white pearls in relation to each other. Although Darwin's evolution of variation was primarily couched in terms of phenotypes, the modern perspective is that the engine for evolution lies in the underlying genetic variation; new genes and consequent new traits arise by several means. These include whole-genome and chromosome duplications, large chromosomal rearrangements such as inversions, translocations and fissions and fusions, the duplication or deletion of portions of chromosomes, balanced reciprocal translocations to small insertions or deletions to base substitutions, and all those processes can lead to alteration of coding sequences of genes or regulatory elements. The other far more common type of genetic variation is single nucleotide changes, usually due to substitution of one base for another. To most geneticists, it goes without saying that single nucleotide alternation (SNP) variation is the most abundant form of genetic variability and is a robust and versatile tool for genetic mapping beyond the limits of linkage analysis. The genetic basis of variation in phenotypes is of considerable interest. routes experimentally derived from quantitative traits should give geneticists clues as to the molecular basis of natural variation. The mechanisms underlying establishment of QTLs may

more parsimoniously explain how genes change in evolution. The range of possibilities for the evolution of phenotypically distinct species becomes immense if changes in phenotypically insidious modulating or metabolic co-factors are considered along with those that alter primary structures of regulatory proteins or transcription factors in stark contrast to the simple but grand vision of the clockwork of life envisaged by T. [8][9][10]

3.1. Mutations

Genetic variation arises from a variety of processes. The most critical genetic variation is DNA base substitutions in the genomes of human and other species. Substitution of nucleotides in the DNA level is coupled with substitutions at the level of mRNA, and is often referred to as a mutation. Mutations can be classified as point mutation, deletions, duplications, insertion, or rearrangement of a few base pairs or larger segments including whole genes or even entire chromosomes. These mutations can lead to premature stop codon, wrong translation, aberration and alteration of gene products, and abnormal protein folding. Mutations can occur spontaneously as a result of errors in DNA replication and repair or be induced by multiple environmental agents. A substantial proportion of mutations are neutral with respect to the organism's fitness [11]. However, they can also be deleterious, and are thus purged from the population by natural selection. Some mutations create new functions resulting in a phenotype that is advantageous, which increases the frequency of the mutation in the population. Rare but discreet new mutants can also arise from existing gene duplications, resulting in phenotype diversity. Such variable phenotypes can be selected for over geological time, eventually leading to speciation. New mutations are the ultimate source of all novel genes, phenotypes, and traits, and thus the raw material on which natural selection acts.

Analysis of genetic variation reveals mutation rates at the DNA level between species, populations, and individual organisms. Mutation rate measurements involving other evolutionary models and traits can also be used to study their variation. There are dozens of studies on rates and aspects of genetic variation and mutation rates at the DNA level in species such as fruit flies, yeast, and bacteria, in addition to population specific studies in humans. Mutation rates are not constant in time and space, but rather vary with environmental factors such as radiation exposure, and stress on other cell divisions and metabolic processes.

3.2. Recombination

Recombination homogenizes genetic variation in a population. The reshuffling of genes and the discarding novel combinations throughout meiosis introduces variation in organisms and a method for adaptive evolution. In addition, recombination is the avenue through which DNA repair occurs in organisms. Each process is necessary for survival and advantageous in the long run [12]. It was likely the evolutionary process behind multi-cellular reproduction, sexual reproduction in organisms, and genomic diversity in offsprings. An ancestral form of recombination likely existed before DNA-as-genetic-material theory was proposed, as X-ray-irradiated Bacillus subtilis began absorbing DNA exogenously. This method of DNA uptake was inherited by E. coli where evidence of homologous recombination in DNA repair exists. The process of crossover during meiosis is similar across taxa, but the elements that regulate the recombination frequency vary drastically [13]. Recombination events occur once in every cell cycle in all organisms. However, species like Drosophila facilitate up to ten recombination events every cycle. Recombination hotspots exist along chromosomes, leading to rates of recombination varying by orders of magnitude across a single genome. Traditionally, genomic rather than genetic distances were used to infer phylogeny, focusing on this recombination history rather than drift. Species without recombination would show greater divergence across a genome with a recombination hotspot than across a large contiguous stretch of sequence. Slow Watterson's theta estimates across highly recombining regions of the Drosophila genome were interpreted as a difference in the time since divergence rather than differential selection or GC-BS conversion. A different evolutionary process, however, was later invoked to explain the same pattern of high-diversity regions in

human. The relative rate of divergence of sequences from Neanderthals and humans showed opposite trends to relative rates of recombination across the same regions.

3.3. Gene Flow

Gene flow usually happens as a result of migration of individuals among populations. As swallows leave for the winter, an initial increase of difference in genetic makeup between neighboring populations will be countered by immigration. In regions where these bird populations are both being studied, the significant results of the differences in genetic variants supported the different selective pressure. Continuing periods of gene flow will lead to the similarities between populations, such as the long term gene flow and mixing of populations in Europe today compared to the rest of the world. Mendel looked at one gene and one trait, while today we can expect more genetic markers and more complex interactions. The increase in confidence intervals will increase the spurious associations, which can incorrectly get attention and publication. Computer intensive and computation demanding genome wide association studies have generated a vast amount of information that changed the field dramatically. The knowledge gained increased understanding of the genetic basis of many complex, commonly occurring diseases by identifying a large number of genetic variants associated with the disease and its distribution in different populations. The availability of this information aids in understanding the complex nature of these diseases in a broader population genetics and evolutionary context, at least for a subset of complex diseases widely studied in gifted populations.

Genetic variation among populations arise from mutations in the genetic material, reshuffling of genes through the passage of genetic material from one generation to the next through sexual reproduction, and migration of individuals among populations. The action of the evolutionary driving forces on genetic variation and evolution depend on the amount of genetic variations that already exist in the populations. Genetic diversity is a genetic treasure generated by the plethora of mutations on the vast and dynamic length of the genetic material that has existed in changing environments on Earth over long time periods. Once a population mate randomly, the allelic makeup of the population will stay in a relatively stable state, known as Hardy Weinberg equilibrium. Genetic drift dominates the action of evolution in this case. However, the drift effect is not independent, as the population is affected by ill normalized environmental factors driving away the population from the equilibrium and increasing the genetic load. The adaptive changes in the genetic makeup of the populations acted by the varying environmental challenge could enable the population to cope with the niche better. The gene flow/migration of human populations to new and different geographical habitats contributed to genetic diversity. Also, during and after the Pleistocene, the human populations experienced very different environmental challenges leading to genetic differences and similarities, especially related to skin color and vitamin D metabolism genes. [14][15]

4. Genetic Variation and Disease Susceptibility

Genetic variation is a major contributor to individual differences in susceptibility to disease. Understanding the genetic basis of disease is a crucial first step towards improved prevention, detection and treatment of disease. Research into the genetic basis of disease has focused on the relatives of affected individuals, on the assumption that disease will aggregate in families and that affected relatives will be more similar than unaffected relatives with respect to the underlying genetic risk factors. Mapping efforts have concentrated on establishing the linkage of genetic markers to disease susceptibility, or association of widely polymorphic markers with disease, as in linkage disequilibrium mapping strategies [16]. For reasons of cost, complexity and extensive prior knowledge, mapping efforts have focused on the role of common biallelic variants in disease. This has resulted in the identification of many common variants associated with risk of disease, marking the beginning of a new phase in human genetics, one where the goal of identifying the genetic basis of susceptibility to a large fraction of common disease appears to be within reach.

Various mutations in genes are associated with specific diseases. The nature of these mutations and variation in pathogenicity is thought to be extremely diverse and with a very heterogeneous genetic background. Hitherto it is established that there are genetic basis of a number of diseases—Mendelian disorders are generally caused by definable defects in a single gene. Also, there are diseases that aggregate in a family but have not yet been resolved at the molecular level, complex or polygenic disorders. These are thought to involve the combined effect of genetic variants at multiple loci, each with a modest functional effect on susceptibility [17]. In addition to genetic variation, significant contributions of environmental factors to most complex diseases are thought to occur, such as sun exposure to skin cancer risk, or smoking to lung cancer risk. Mapping of genomic regions showing linkage to disease in families contributes to a better understanding of the biology underlying the disease thereby improving diagnosis, treatment, and human health.

4.1. Monogenic Disorders

Monogenic disorders are caused by pathogenic variants affecting a single gene and may occur in various inheritance patterns, including autosomal dominant, autosomal recessive, or X-linked. In a typical monogenic disorder, pathogenic variants are rare within the population and of high cumulative population risk, often being perfect penetrant within a specific genetic context, this penetrating context often being as specific as the one given by a single nucleotide variant. On the clinical side, due to the large, often lethal early-onset presence of the disorder, most patients will be diagnosed by the time they enter adulthood. Since the time-to-onset tends to be fixed at a scale of weeks, unless lethal early-onset phenotypes commonly observed in other diseases, on the biological side, the contribution of environmental risk factors will be limited to a reduced risk of being diagnosed later in life. Nevertheless, identifying the precise causal variant often proves difficult, as one is limited to rare variants, requiring high-density genotyping and/or extensive re-sequencing of candidate genes. Thanks to the large number of genes involved in monogenic diseases, there are hundreds of approaches to treat the disorders these genes cause. Techniques include gene therapy, RNA therapy, genome-editing, and protein replacements. Striking results have already been obtained with gene therapy approaches based on viral vectors for certain conditions, as well as for new oligonucleotides against specific disorders. In addition to the above-mentioned techniques, several promising compounds are moving forward to clinics, the most famous example being the application of the first small molecule against a specific condition.

Oligogenic disorders, also known as digenic or multigenic disorders, are affected by the joint presence or interaction of two or more genetic risk factors. The number of genes for which individuals can possess pathogenic variants is limited compared to polygenic disorders. The total proportion of patients affected by pathogenic variants in the genes is small but may reach a few percent at a population level. The clinical case may also be more heterogeneous than in monogenic cases, requiring additional filtering of pathogenic variants. Although knowledge on heredity applies, things may become slightly more complicated when the causal genes show recessive inheritance, and thus, one may have to consider the combined effect of common variants with a small effect size. Due to the degeneracy of the problems and existing models, screening for pathogenic variants often could be done directly on the orthologous genes. The future of research concerning oligogenic disorders is promising, especially in diseases that are currently poorly understood. Some human disorders may not be purely oligogenic but more appropriately described as polygenic and will remain elusive for a while. [18][19]

4.2. Polygenic Disorders

Polygenic disorders, or common diseases as they are also frequently called, are diseases whose complex phenotype is the result of the genetic variation of many gene loci, each with a small effect, together with environmental factors which may influence its expression. The polygenic diseases are the most prevalent diseases of modern man. The reasons for this state of affairs have been the subject of many biomedical and philosophical discussions over the past few hundred

years of human history. Statistically, it could be shown that such diseases have a complex nature [20]. They exhibit a widely distributed liability in the population where both genetic and environmental components may have an effect. It could be shown that most familial clusters of polygenic disorders would not fit any of the classical models of genetics, as merely chance combinations of the effects of genes and environments of various kinds over a long period of time. Environmental factors are those external to the organism and produce environmental diseases.

Another basic epidemiological observation concerning polygenic disorders is that two different, non-inheritable non-genetic diseases may co-occur more frequently than one could expect by random sampling. A classical epidemiological example of this fact is the observation that children with cleft lip and palate have a higher frequency of leukemia than children without cleft lip and palate. The reasons for this co-occurrence are again unknown. However, it has generally been the case in genetic epidemiology that piecemeal advances in the knowledge of one disease would continue without any prospect of unravelling the mystery of the human condition in general. No one is quite sure why the genome could have given rise to so many and such diverse polygenic disorders despite the shared genetic variability coding individual differences. Progresses across the different common diseases would continue with sporadic finds irrelevant to other diseases and, therefore, prone to disappear from the scientific agenda. Such a situation would be considered unsatisfactory in any scientific discipline dealing with complex phenomena.

4.3. Complex Diseases

At the molecular-biological level regulation of phenotype can be simplified to a view that human genome represents an enormous collection of alleles, from which a phenotypes is extracted [20]. Alleles influencing phenotype can be grouped into those that determine risk for a relatively rare Mendelian diseases and those that shape common complex diseases or ‘quantitative traits’. The latter category differs from the former by not being necessary or sufficient for the development of disease. For the complex diseases the extent to which the focus should shift from genes to environments is not known, although that’s the only sensible direction in terms of prevention and removal of risk factors. It should be possible to construct a search space that contains the vast majority of possible human risk variants. This would enable adequately powered association studies to disentangle common alleles of small effect size, although such studies have proved inscrutable. Greater success has been achieved in dissecting rarer alleles of larger effect size that confer strength of relative risk greater than five-fold. The most notable examples are collected from the past decade, although Mendelian diseases for which such variants remain undiscovered are common.

5. Case Studies of Genetic Variation in Diseases

Most diseases are a consequence of a complex interaction between multiple environmental and genetic factors [16]. Thus, the identification of disease-associated genetic variants in uncharacterized and genetically diverse populations will bring insights into the genetic basis of human diseases. Genome-wide association studies (GWAS) performed in both the European and African ancestry populations have helped identify more than 100 disease-associated regions for type 2 diabetes (T2D), overweight/obesity (OW/OB), and body mass index (BMI). While the substantial fraction of heritability remains unexplained, it has been established that most of disease variants identified by GWAS are non-coding.

A gene-based bioinformatic approach has been developed to systematically identify gene-variants linked to complex diseases. When applied to GWAS data of T2D, OW/OB, and BMI explored in EAS and EUR, it has identified 45 novel gene-variants. Of the gene-variants identified, 37 were found in uncharacterized regions of the human genome. The expression constructs for eight of them have been developed and their function/effect on cell/animal models will be explored [17].

The majority of diseases human suffer from differ because of genetic variation. Genetic variation is classified into three types: SNPs, Copy number variations, and Structural variants, and is

usually referring to SNPs difference on drug and disease response. These variations cause corresponding changes to the transcriptome, proteome, methylome, metabolome, and microbiome. Consequently, individual gain susceptibility to complex human diseases. Although human genetic variation is mainly harmless, it can be classified into 6 types of detriments on health on the basis of variant-product combinations.

5.1. Cystic Fibrosis

The cystic fibrosis gene, named the cystic fibrosis transmembrane conductance regulator (CFTR), was discovered in 1989. It encodes a protein containing 1480 amino acids, a member of the ATP-gated family of ion channels responsible for regulating fluid movement in multiple epithelia. The disease is due to abnormal ion transport in the epithelial cells of the respiratory system, pancreas, intestine, and exocrine glands due to alterations in the function of the protein encoded for the CFTR gene [21]. This results in an obstructive lung disease (i.e. thickened airway secretions that lead to chronic infection), dilation and infection of the pancreas, biliary obstruction, and distal intestinal obstruction disease. The gene is located on chromosome 7q31.2p, with an open reading frame consisting of 27 exons. More than 2000 mutations leading to cystic fibrosis have been identified, located throughout the CFTR gene, of which approximately 1200 mutations are considered pathogenic. A pancreatic sufficient phenotype is observed in patients who are homozygous for rare mutations. It is noted that the F508del is recognized as the most common mutation. The tract starts with CTT and it is due to the deletion of phenylalanine at position 508 of the amino acid chain (F508del) in the protein CFTR characterizing the classic form of CF. On average, the F508del mutation is detected worldwide in more than 70% of patients with cystic fibrosis. The affected individuals are frequently homozygous for the F508del mutation (70%) or compound heterozygotes with a second classical mutation. The deletion of one normal allele of the CFTR gene has an asymptomatic phenotype, with no pulmonary impairment. When two healthy heterozygous parents are carriers, they have a 25% risk of having an affected newborn. The cystic fibrosis gene is located on the long arm of chromosome 7, at position 7q31.2p, containing 27 exons.

5.2. Sickle Cell Anemia

Sickle cell anemia (SCA) is a common single point mutation chromosomal disorder caused by the replacement of glutamic acid at position 6 by valine in the β-globin gene of Hemoglobin A (HbA) [22]. This point mutation leads to the formation of mutant deoxyhemoglobin, HbS. Also, sickle cell anemia in humans is an autosomal recessive disease affecting the β-globin gene of hemoglobin No. 11. It is caused by a mutation that substitutes valine for glutamic acid at position 6 of the β-globin polypeptide chain. The pathophysiology of the disease includes hemolytic anemias and vaso-occlusive crises. These events are dictated by the interaction of two different polypeptides, α-globin and mutated β-globin subunits. Traditional therapy for SCA has focused on reducing the concentration of HbS in red blood cells (RBCs) by various means, including promoting the synthesis of fetal HbA (HbF), which inhibits polymerization. In addition to HbF induction, therapeutic strategies have included anti-sickling agents and drugs that inhibit the production of cognate α-globin subunits. More recent induced pluripotent stem cell (iPSC)-based approaches may lead to patient-specific therapy that does not require bone marrow transplantation. Understanding the structure, function, and interactions of the hemoglobin molecule is crucial in elucidating the molecular pathophysiology of sickle cell anemia and in designing new therapeutics to treat it.

The commonality of SCA in humans is in those from African descent, with an estimated 0.1% compound heterozygous mutation worldwide. The disease states for homozygous mutation (HbSS), sickle phenotype β-globin mutation, mildly defective β-globin synthesis (HbSB), and increased γ-globin synthesis (HbSA) are also known. SCA is prevalent in Africa, but if the homozygous mutation of β-globin is absent, one can still be considered a "carrier" of sickle cell disease with a sickle cell trait. Normal (HbAA), sickle cell trait (HbAS), and sickle cell disease

(HbSS) are the three possible states of β-globin. Only homozygous recessive individuals exhibit the disease phenotype and complications. Heterozygous individuals often do not show disease symptoms, and are considered carriers of sickle cell trait. Individuals homozygous for the trait have a higher chance of cardiac, pulmonary, and renal complications. Parents are encouraged to get screened and genetically tested, which involves blood tests with hemoglobin electrophoresis. There are 99,097–100,845 people in the United States with SCA.

5.3. Breast Cancer

A more cursory examination of the role genetic variation can play in breast cancer has moreover been undertaken with regard to predictors of tumormetastasis and recurrence as well as response to chemotherapy. A review of single nucleotide polymorphisms (SNPs) in receptor tyrosine kinases associated with breast cancer proposes a role in determining risk and prognosis. Variation at the DNA level in genes that affect the pharmacokinetics and/or pharmacodynamics of a number of chemotherapeutic drugs has been explored as factors influencing toxicity and patient response [23]. Perhaps the information could be used to identify candidates amenable to therapy bicalutamide, a nonsteroidal antiandrogen with pharmacogenetic variants that may adversely affect hepatic function. Likewise, there may exist variants associated with tamoxifen resistance, a nonsteroidal selective estrogen receptor modulator that could prove important biomarkers for breast cancer. Otherwise, with paragraph length being a more prohibitive factor, additional avenues of research into important biomarkers in breast cancer propose SNPs in receptor activator of NF-kappaB ligand associated with breast cancer risk or 691G>A variant of the BCL-2 gene enhance breast cancer susceptibility.

A number of first reports of genetic variants associated with a variety of different drug therapy interactions for breast cancer such as with doxorubicin (DOX) and chemotherapy have consequently since arisen after initial findings. Staining of formalin-fixed, paraffin-embedded tissues of patients treated with adjuvant chemotherapy noted that hadoor paint and neoplasia were significantly associated; likewise, it was suggested that a variety of different genetic variants could modify individual response. Many return their own unique toxicities; anthracyclines are generally notable for their cardotoxicity, with DOX in particular posing a risk of development of congestive heart failure (CHF), while taxanes can elicit severe hypersensitivity reactions. Variants relevant to drug exposure are also suggested by the various polymorphisms in drug transporters and metabolizing enzymes. As such, relevant genes affecting the pharmacokinetics of DOX act to it accumulating in the cell and propose deficiency in any pathway affecting initial hepatic glucuronidation of the drug would result in increased drug levels and increased toxicity and risk of CHF.

6. Population Genetics and Disease

Population genetics helps understand human migration and origin events and how they affect genetic diversity among populations and hence the distribution of complex diseases [1]. Population genetics uses genetic variations among individuals to understand evolutionary mechanisms that shape a population. Although initially applied on a small scale, it has evolved to encompass full genomes and uniparental loci across whole populations to study the population structure, gene flow, natural selection, and genic evolution. To gain more understanding of the genetic adaptation of human populations to their adapted environmental conditions, population genetic studies have been conducted on various human populations starting from the time of out-of-Africa until the recent agricultural and even industrial movements.

Two main forces affect genetic diversity: mutation and drift. Mutation is generally regarded as the original source of new alleles, and it can happen by a variety of mechanisms, which usually involve the replacement of one nucleotide by another, an insertion, or a deletion at a given place in the genome. The long-term mutation rate is estimated to be approximately 5^-9 per site per generation and is not equally distributed among different sites. Genetic drift refers to the random fluctuation of allele frequencies in a population and can result in a significant loss of genetic

variation. Drift can become particularly conspicuous in small populations, where it can lead to the complete loss of an allele. It has been noted that human populations migrated to high latitudes in the northern hemisphere carry significantly fewer alleles than populations residing near the equator, and this difference is only partially explained by their population size.

6.1. Genetic Drift

Genetic drift is a major evolutionary process characterized by random fluctuations of allele frequencies in a population from one generation to another [24]. Such random fluctuations can trigger allele fixation and loss of alternative alleles. It is generally thought to depend on population size (N), the number of reproducing individuals (Ne), and variance in reproductive success. Overwhelmingly, theoretical and empirical studies focus on N, which is a pooled measure of genetic diversity in a population; Ne captures individual breeding potential and has an extensive definition and analytic characteristics; while reproductive variance estimates the probabilistic extremes of when particular individuals do not reproduce or reproduce a large fraction of the population's offspring. The key concept used to quantify the effect of drift on populations is effective population size (Ne). Genetic drift leads to the random loss of alleles as one population changes into another and can account for the divergence of populations or species [25]. On the other hand, the random fixation of alleles can create irreparable differences between individuals of the same species, and overcome by post-zygotic Reproductive Isolation, homogametic sex mortality. The existence of reproductive isolation has prompted a need for a better understanding of drift, for if divergence processes cannot be circumvented before they are irreversible, it would appear that speciation is biologically inevitable. For this reason, widespread in the literature, in population genetics, fishery and crop management, drift is probably the process that requires the largest quantity of work in evolutionary genetics of wide applicability, in a variety of development and ecophysiological contexts. There are also other simple models that reveal many aspects of drift and other aspects of the interaction between reproduction, development and drift. The aim is to point to some of the ways to get results succinctly from such experimental or analytical works, so as to relieve the burden of necessity.

6.2. Natural Selection

All living organisms including humans have been evolving with changing environmental factors over the geological periods. Biological evolution is regarded as the contribution of natural selection and chance mutation or genetic drift and random loss of alleles. The variation of characteristics among the individuals of a species is caused by a combination of natural selection and chance mutation. This variation, for the most part, is neutral to the reproductive success of the individuals and is fluctuated randomly among the generations [1].

This selective pressure leads to the increase of frequency of favored genetic makeups and the elimination of deleterious genetic makeups that fail to adapt with the new environmental challenges. Thus, natural selective events have shaped the present genetic diversity of existing populations and consequently genetic variants involved in many diseases. Since the overall fitness of the individuals and population is dependent upon the successful development, homeostasis maintenance during the variability in the environment, growth, reproduction, and surviving of different perturbations in the environment, it is less surprising that limiting variation in this regard has been a primary concern of evolutionary biologists. Negative selection, also called purifying selection, is the most well-known form of natural selection [26]. Negative selection removes disadvantageous alleles or mutations from the population gene pool and reduces their frequencies in the population.

Under the influence of negative selection, the frequency of deleterious effects will be decreased until the effects on fitness become negligible. Genetic variants that have both advantageous and nearly neutral effects on fitness may shift to equilibrium frequencies under positive or balancing selection or episodic selective pressure such as those caused by changing environmental pressures, resulting in a transient excess or deficiency of genetic diversity or divergence. Population

bottleneck and expansion would affect the overall levels and patterns of genetic diversity. Therefore, signatures of natural selection can shape the patterns of genetic variation and can be used to detect the genes or mutations that are functionally important in adaptation and disease susceptibility.

6.3. Population Structure

A description of the population genetic structure of humans and criticisms of the use of putative genetic factors in defining biological races. Genetic variation is broadly examined from a few polymorphisms informing population divergence to millions of single nucleotide polymorphisms (SNPs) illuminating fine-scale substructure of populations. Other genetic markers are mentioned, as well as many modeling frameworks to simulate populations over time and describe genetic data. With so many insights gained in this field, there remains controversy in its implications for defining different races of people and inferences of species' behaviors and mating patterns from genetic data.

The diagram below illustrates the temporal hierarchies at which genetic measures may be taken among contemporarily residing individuals: a sample of genes present in the a-sexual offspring of one population and a pool of mates in a range spatio-temporal contexts. Population comparisons may be made from the nucleotide polymorphisms of genes sequenced. Likewise, several intermediate genetic statistics have been developed from such measures to infer the coalescent histories of populations, mating structure and time-since-separation. Newer sequencing technologies make it possible to gather hundreds of thousands of SNPs, geographically disperse them among multiple populations and model their genome-wide fitness with respect to the population structure, migration between populations and Natural Selection [27].

7. Ethical Considerations in Genetic Research

A variety of ethical concerns arise from the ability of researchers to examine the effects of thousands of genetic variations on human beings. The general ethical consideration is that genetic research, particularly genetic and behavioral genetic research, does not cause harm [28].

Research results from genetic studies should not be used in a way that is detrimental to the individual or a two-person combination. Research that uses data from genetic research studies should not infringe on an individual's or a two-person combination's right to privacy. Unethical use of results, data, and knowledge acquired from genetic studies should be legally prosecuted. Informed consent should be obtained from subjects prior to participating in any research utilizing the subjects' genetic or other personal data.

The consent process should include clear descriptions of what type of genetic research will occur, how data will be prepared, stored, analyzed, and provided, the individuals allowed access to the data, and how ongoing consent issues will be managed [29]. Genomic studies should operate with the understanding that disclosure of group-level data may affect the wellbeing of populations and communities. Genomic researchers are typically encouraged to maintain the confidentiality of individual-level data but there is disagreement as to whether this extends to group level data in general, or to group level information that could identify certain populations, communities, or ethnicities.

7.1. Privacy and Genetic Data

People often view information about their gene as private. Each person has one genome, and it is unique. Yet within an individual and their genome is a set of specific variants, most of which are quite common (except in rare conditions), widely shared with biological relatives, and even, in some instances, shared across the entire human population. Thus, a human genome has the mixed character of being both private and public—it is in this sense a "private" yet "public" information, making any discussion of policy addressing genetic privacy vexed [30]. Now DNA is conceptualized as a unique identifier that is supposed to provide insights into many aspects of the

person, person's family, and population, including health. For these reasons, many people want to control who can access their genetic sequence or further derived information as to them, which derives calls for strong privacy protection or analogously for personal ownership of genetic data.

On this view, information about one person's genetic variants reveals information about their immediate biological relatives and possibly about quite distant biological relatives. In fact, one can know little about the significance of individual variants without the cooperation of many other individuals, hence making privacy protections at the individual level a tricky business. Not only does this public nature of the genome (and its derived data) make it difficult to decide what level of control individuals should have, but it complicates the question of how to provide appropriate privacy protection. The concept of "privacy" that permeates many legal discussions today has evolved considerably in recent decades. Birthdates or state-issued IDs cannot be resold, but cell phone numbers, the means of human communications, location, and, today, entries about where pictures were taken, travel paths, and social affiliation can be resold, are informationally available to all, nevertheless. Is genetic data information that society treats as private? Should individuals be able to maintain a level of control over access to and use of genetic data about themselves?

7.2. Gene Editing Ethics

In February 2017, the committee on human gene editing acknowledged that advancements in genomic engineering technology presented complex scientific, ethical, and legal concerns. The committee extensively reviewed the potential uses of such technology and the associated concerns, produced a final report on these matters, and hosted a conference to further elucidate these topics [31]. Of particular note, the committee determined that currently, heritable genome editing should be deemed impermissible, but it may be justified for certain medical indications in the future if more is understood about the safety of currently available technologies and appropriate governance and oversight mechanisms are in place. The committee also contended that the use of CRISPR technology for any form of enhancement is outside the bounds of what could be approved as an allowable use of gene editing.

However, discussions on the ethics of germline gene editing predates the 2017 report. Bioethicists have explored several concerns about the ethics of germline gene editing, the ethical status of germline gene editing interventions, and the broader social implications of potential applications of germline gene editing technology. In general, the major ethical concerns regarding germline gene editing technologies orbit around three key issues: safety, risk/benefit, and efficacy. For some critics, the unproven, unreliable, and potentially unsafe nature of CRISPR poses an insurmountable ethical barrier to clinical germline editing. Many people share concerns regarding equity and justice—who will have access to these interventions? Will the resultant genetically improved humans enjoy superior health, intelligence, or even beauty? The equitable distribution of these benefits is unclear, especially regarding affordable access.

7.3. Implications for Genetic Counseling

Genetic counseling requires the counselor's understanding of genomic knowledge in addition to the general knowledge of medicine. Not only should genetic counselors understand the familial mechanisms and modes of inheritance, but they must also have a comprehensive knowledge of genomic elements, variations, and methods for finding them. However, there is doubt about whether the non-medical personnel, who have just been trained for a limited time, can provide genetic counseling—education and return of the results without detailed knowledge of both medicine and genomics. Next-generation sequencing (NGS) does not cause only one variant. Rather, it provides a lot of variants relevant to a number of diseases. An effective genetic counseling service in the NGS era would require the counselor's comprehensive knowledge of not only germline mutations and somatic mutations, but also single gene disorders and multifactorial diseases. Based on the experience of the author, the relationship between the complex of genetic variants and the specific medical problems can spiral such that, for example, genetic variants related to SLE may lead to the issues of Kassir and Muir-Torre syndrome in Asian patients. This

means that it is necessary to have a strong foundation of general medicine in the NGS era [32].

Ten years ago, genetic testing was usually performed for specific clinical problems, for example, FLT3 mutations in patients with acute myeloid leukemia (AML). The result produced by Sanger sequencing, a traditional method for one gene, typically involved three to five single nucleotide variations (SNVs) and a one or two nucleotide deletion/insertion for non-coding regions. However, given that at least 1000 somatic mutations per tumor genome are usually identified by whole genome sequencing (WGS), it is impossible to imagine the complexity of patient-returned results. Medical doctors specializing in internal medicine or surgery, not trained in genetics or bioinformatics, usually interpret the limited results in the prior era. However, they would just point out “variants of uncertain significance” or “the genetic variant is unlikely to cause disease” when confronted with a novel or innumerable complex of variants containing no or limited clinical interpretations.

8. Technological Advances in Genetic Research

The discovery of the first restriction enzyme in 1970 enabled detection of genome differences by restriction fragment length polymorphism (RFLP). RFLP analysis demonstrated inter-individual genetic variation, including polymorphisms in the form of microsatellite repeats. The extent of genetic variation among individuals was largely unrecognized until the development of dideoxynucleotide chain termination sequencing in 1977, which enabled sequencing at base pair resolution, and the advent of polymerase chain reaction (PCR) in 1985, which allowed amplifying DNA fragments [33]. These advances led to the launching of the Human Genome Project in 1990. The early stages of the effort involved the creation of basic physical and genetic maps of the human genome. The purpose of the maps was to ensure accurate assembly of the high-throughput sequence data generated from random shotgun sequencing of the human genome. Beyond just sequencing the constant regions, both the haplotypes and structural variations in the relatively non-conserved regions of the human genome were examined. The Human Genome Project has resulted in awareness of basic principles affecting the integrity of a genome and inter-individual genomic variability, the recognition of the potential benefits and harms of genome interrogation, and the initiation of efforts to understand the human genomic variability underpinning the diversity in susceptibility and resistance to diseases. The latter includes the examination of the first-ever detoxifying enzymes and their relationship with diet, the selective hypermutation of the immunoglobulin gene rearrangements, the identification of single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) that affect phenotypic variability, and the discoveries of disease-causing variants through family-based genetic approaches. Such enhanced understanding built the foundation for the current research aimed at seeking the genetic basis of thousands of diseases in the human population [17]. The error rate of the DNA replication machinery introduces approximately 50 de novo point mutations and a lesser number of larger mutations with each genome replication. Thus, each offspring differs from the parents by about 50 novel genetic variants.

8.1. Next-Generation Sequencing

Next-generation sequencing (NGS) technologies have revolutionized research in diverse biological systems [34]. These massively parallel sequencing techniques allow biologists to investigate genomes at an unprecedented resolution and scale. Due to its widespread availability, NGS has become one of the cornerstones of genomic epidemiology. Epidemiologists with access to the NGS techniques can now collect more nuanced exposure information than ever before. They are able to characterize datasets encompassing a range of exposures, germline variants, and microbial communities. In parallel, the increased data volume and richness also raise big challenges for data-processors and data-interpretators. To this end, this article reviews designs and statistical methods that have been proposed for genomic epidemiology studies using NGS.

NGS methods vary in their biochemistry, and thus the information provided by NGS technologies differs. In a classic effect of consideration, NGS approaches can be categorized into (1) whole-

genome sequencing (WGS), (2) whole-exome sequencing (WES), and (3) sequencing of targeted genomic regions (or genes). WGS uniformly surveys genomes, capturing both commonly-gained and low-frequency variants. WGS is currently the gold-standard study design. However, due to the vast size of the data collected and its deep coverage of both coding and non-coding regions, it is often cost-prohibitive. In contrast, WES covers only exons (the protein-coding regions), and thus results in a smaller analysis file size. However, it often fails to detect more challenging to capture variants or variants of smaller size, template preparation artifacts, or poorly behaved regions [35]. Finally, targeted sequencing methods comprise capturing and enrichment approaches. Targeting specific genomic regions and filtering out the rest can lead to much faster turn-around times, lower costs, smaller datasets, and greater coverage.

8.2. CRISPR Technology

The “clustered regularly interspaced short palindromic repeats” (CRISPR) are repetitive palindromic DNA sequences in immune systems of bacteria and archaea, that play crucial roles in sequence-specific degradation of foreign nucleic acids, such as viruses and plasmids [36]. The simple two-component genome editing system consists of one CRISPR and one CRISPR-associated (cas) genes. As a class 2 and type II-A CRISPR system that recruits one trans-acting CRISPR RNA (tracrRNA)/scaffold and one CRISPR ribonuclease (Cas9), that genome could be edited at any site sandwiched by protospacer adjacent motifs (PAM) sequences.

After 33 years of research, genome-editing technologies that precisely correct the diseases-causing attributes of genes have emerged as a potentially effective therapeutic modality for traditional genetic diseases and more complex diseases. Transcription activator-like effector nucleases (TALEN), zinc finger nucleases (ZFNs), and of late CRISPR/Cas technologies have drawn significant attention from academia to pharmaceutical industries due to desirable specificity, efficiency, and easy-to-follow design rules. Recently, genome-editing tools that cleave DNA metalloproteases and cytosine/RNA-guided Cas9 have emerged and joined the arsenals.

The demands for even more versatile and safer genome-modifying tools have undergone a rapid engineering race to enhance the fidelity, efficiency, and versatility of the present-day genome modifier tools. Non-nuclease-based applications such as transcription modulation (CRISPRi/a) and epigenetic modulation (CRISPR-dCas9) of CRISPR/Cas technologies have also been developed and are actively pursued by researchers for their significant biological applications in answering prevailing questions in diverse areas. To overcome the safety concerns with both established and novel genome editing technologies, emerging epi-genome editing tools that precisely alter DNA methylation and histone modifications without disrupting the helical structure of DNA have attracted notable attention and provides a glimpse of a new era where long-lasting clinical efficacy could be achieved without producing permanent alterations in the genetic make-up.

8.3. Bioinformatics Tools

In recent years, many researchers have studied the relationship between genetic variation and diseases, and diverse bioinformatics tools have been developed on this issue. PredictSNP is a unified platform that comprehensively summarizes 12 prediction tools for missense SNP effects. These prediction tools are divided into three groups according to the supervised machine-learning algorithms used for SNP effect prediction: (1) the neural networks, (2) support vector machines, and (3) homology modeling. Some tools predict the impact on protein structure, others the impact on the protein interaction, and still others the binding of transcription factors. As shown in FIG. 5, performance comparison shows that, at least for some models, addressing the issue in diverse ways is beneficial. As executables, the tools can be applied effectively to numerous SNPs. The current implementation focuses on human SNPs and proteins [37].

9. Future Directions in Genetic Research

SNPs promise inexpensive genetic tests for susceptibility to complex diseases and drug responses.

Motivation for diagnosing a genetic predisposition to diseases has grown because some areas of the world have adopted genetic tests that associate a risk profile with susceptibility to multiple diseases [38]. Current research involves a mixture of technologies, including complete sequencing, candidate gene studies, and genome-wide approaches. These approaches will be assessed in turn, with a focus on the issues they raise. It will be seen that there is no single approach to dissecting the genetics of complex traits that is sufficient to serve as the basis on which to develop a genetic test. Each is limited and there is a need for a sound knowledge of the disease in conjunction with the genetic tests. There are a number of issues and challenges that arise from the increasing use of associating a DNA test to either susceptibility to disease or drug response [3]. Some of these arise from the technology and its limits, some from statistical issues associated with the nature of the tests, some from the biology of the interaction between the disease and SNPs, and some concern ethics and the implications of the test results. The advent of DNA polymorphism technology has led to an increased interest in understanding the role of genetic architecture in the evolution of barrier mechanisms. Attention has been focused on selection tests and a plethora of software has been developed to augment SNP data with respect to population genetics. The tests make different assumptions, exploit different features of the data, and aim to disentangle different components of the history of the population and selection. The role of the DMCA and the EU directive on the protection of copyright works in scientific publications must be taken into account and they are not treated here.

Research into the way genetic polymorphisms associate with disease risk is on the verge of a revolution. Compared with the post-millennial era of the Human Genome Project and candidate gene studies, the next decade will see far larger collaborations, a wealth of new data from varied disease states, the analysis of different regulatory and structural polymorphisms, and the first investigations of deep sequencing data. Investigations will need to harness the exponentially increasing computational power and techniques, and a greater understanding of the biology of the reasons behind association will emerge. The more divorces the elements of the double helix of information are separated, the more clear it will become that single bases in single genes cannot explain human behavior. From being an isolated no-man’s land, ethnicity with its egalitarian and exploitable features is the new curve of the current human evolution.

9.1. Personalized Medicine

With the development of high-throughput sequencing technologies, the human genome can be sequenced more quickly and cheaply than ever before, which drives the rapid development of high-throughput genotyping and sequencing methods. More and more human genetic variants have been identified in the Genome Variation Map and the Genome Aggregation Database, and an increasing number of human diseases with genetic causes have been explained by underlying genetic variation [39]. These efforts not only expand human genetic variation in a broader geographic and ethnic scope but also support studies of better quality to accelerate the identification of human genetic variation and discovery of its implications in diseases.

A treatment for individual patients can be personalized based on the result of polymorphism with a medically valuable gene and the analysis of other postgenomic information. Personalized medicine results in benefits both for doctors and patients. It can increase the efficiency of drug treatment and bring more convenience for the patients. Individualized medications can heighten the cost-effectiveness of the treatment for both individuals and society. With the existing better knowledge of common polymorphisms, pharmacogenetic testing before drug selection is becoming commercially available for a range of therapies that include cancer, cardiovascular and psychiatric diseases. The pharmacogenomic research is anticipated to improve healthcare since many marketed pharmaceuticals exhibit interindividual variation of efficacy/toxicity/dosing due to altered pharmacogenetic mechanisms. Variance in pharmacogenes such as GAPDH and PTGS2 might impact the antitumor density of 5-FU and efficacy of celecoxib for CRC. Since Drugs for patients with low erosion potential genes or non-blunts can take more time and dose to reduce the side effects.

9.2. Gene Therapy

Gene therapy refers to a technique involving the alteration of genes inside a patient's cells to treat the underlying cause of disease. Gene therapy is an emerging therapeutic approach for the treatment of inherited diseases induced by a genetic defect. It aims to correct the defect by either inserting genes into an appropriate location in the genome to replace the non-functional gene, or by repairing defective genes. Alternative approaches are being evaluated with mRNAs or small molecules to selectively shut off-gene expression or modify gene expression. Early attempts at this groundbreaking therapy were halted by unexpected adverse effects, including immune reactions and cancer. However, the past two decades have yielded significant advances in bioengineering strategies for genome editing, delivery, and characterization using nanoparticle. Successful clinical trials have led to approval of the first gene therapy product and an increasing number of ongoing clinical trials.

The genome editing technologies that knock-in large DNA fragments at preselected genomic sites using programmable nucleases are being developed to treat genetic diseases by replenishing the missing or defective protein function such as hemophilia A. Targeted genome editing can also reduce or eliminate the expression of pathogenic genes with dominant-negative mutations of a given protein or the whole gene copy number, such as Huntingdon's disease and sickle cell disease. In particular, a potent technique for targeted genome editing of the human genome, transcription-activator like effector nucleases and/or directed nucleases can be achieved through homologous recombination or nonhomologous end-joining pathways to modulate endogenous gene expression. It also highlights the current developments of delivery systems that can enable most of these technologies for in vivo intervention of genetic diseases. Ultimately, successful clinical deployment of these breakthrough technologies will start a new era of curing genetic diseases.

Gene therapy is the treatment of genetic disorders by transfer of a normal gene into the somatic cells in tissue. It has the potential to treat the root cause of a disorder rather than merely alleviating symptoms. Because of the extraordinarily high cost and challenging methods for delivery, the latter has dominated the gene therapy market and will continue to outpace the former. As gene augmentation therapies have been aggressively pursued, it is essential from a science and business perspective to approach therapies that will require much greater bioengineering and success in animal studies before the clinic. Current severely affected cohorts of recipients are relatively small and will have to wait for inevitable improvement in methods for cost-effective delivery. [40][41]

9.3. Genomic Epidemiology

Most public health actions are directed on the basis of knowledge approximately environmental causes of diseases. This state of affairs began to change with the sequencing of the human genome and the consequent development of all technologies needed for inexpensive, reliable and high-throughput genetic analyses. Sooner or later public health measures to be taken on the basis of genetic results will have to be defined. In order to create the need for such measure as the fifth epidemiological transition it is necessary to persuade public health authorities to build up a long-term strategy.

Why Genomic Epidemiology? The purpose of epidemiology is to combat diseases as long as they do not exist and cure them if they enter the system. Every disease has its own start or index case and, either through environmental or genetic causes, enters the system of healthy individuals. The epidemic fallout of the disease depends on the virulence, transmissibility and the strength of intervening measures. The vast majority of human diseases tested so far, have been by far best predicted by the most robust and reliable categorical variables such as age, sex, geography and season [20].

In addition to such factors, whose impact follows universal trends, there exist intraindividual factors, largely, if not exclusively genetic that after years of interaction with environmental factors

lead to the disease. Today it is staged that everybody carries a non-zero risk for number of diseases. There are general pathways to establish preventive measure capturing the exogenous factors for the vast majority of diseases with the only exception of very rare congenital diseases, where risks related to environmental factors are negligible.

10. Conclusion

Vertebrates across the tree of life display intraspecific genetic heterogeneity that is of both evolutionary relevance and physiological consequence. Although interspecies differences in the extent of genetic diversity are evident, all species possess the genetic polymorphisms that yield phenotypic variability, which is required for the survival of organisms in a comparable environment. Of paramount significance is the fact that healthy and pathogenic living conditions expose organisms to environmental forces. The efficacy of the genetic response to these forces varies across populations, leading to differential susceptibility to a given disease, and explaining the disparity in its epidemiologic patterns. The forebearers of all modern humans arose in sub-Saharan Africa around 200,000 years ago. Hominid evolution, the process of human species adaptation to environmental fluctuations, was guided by equally by natural selection and chance events. Early modern humans subsequently spread to the Middle East, the Far East, Northern Europe, and North and South America. As they did so, many genetically disparate populations arose that subsequently experienced gene flow. Non-African populations that migrated to temperate or cold, arid environments modified genes that regulate endocrine homeostasis and maintenance of metabolic fitness. Some mutations that enhance adaptation to environmental changes lead to evolutionary advantage but predisposition to modern complex diseases. Currently, genes that are deleterious to physiological homeostasis but were selected for during human evolution are associated with aberrations of health. The adverse consequences of historical selection events on fitness and health in modern humans necessitate consideration of the timing and geographic settings of their occurrence for a better understanding of their clinical consequence.

References:

1. A. T. M. Saeb and D. Al-Naqeb, "The Impact of Evolutionary Driving Forces on Human Complex Diseases: A Population Genetics Approach," 2016. ncbi.nlm.nih.gov
2. M. Jackson, L. Marks, G. H.W. May, and J. B. Wilson, "The genetic basis of disease," 2018. ncbi.nlm.nih.gov
3. J. D. Chiche, A. Cariou, and J. P. Mira, "Bench-to-bedside review: Fulfilling promises of the Human Genome Project," 2002. ncbi.nlm.nih.gov
4. L. E. Flagel, J. H. Willis, and T. J. Vision, "The standing pool of genomic structural variation in a natural population of Mimulus guttatus," 2013. [PDF]
5. R. Jovelin and A. D. Cutter, "Fine-Scale Signatures of Molecular Evolution Reconcile Models of Indel-Associated Mutation," 2013. ncbi.nlm.nih.gov
6. H. BoonPeng and K. Yusoff, "The utility of copy number variation (CNV) in studies of hypertension-related left ventricular hypertrophy (LVH): rationale, potential and challenges," 2013. ncbi.nlm.nih.gov
7. L. Budzko, M. Marcinkowska-Swojak, P. Jackowiak, P. Kozlowski et al., "Copy number variation of genes involved in the hepatitis C virus-human interactome," 2016. ncbi.nlm.nih.gov
8. M. S. Hill, P. Vande Zande, and P. J. Wittkopp, "Molecular and evolutionary processes generating variation in gene expression," Nature Reviews Genetics, 2021. nih.gov
9. B. Sahu, T. Hartonen, P. Pihlajamaa, B. Wei, K. Dave, et al., "Sequence determinants of human gene regulatory elements," *Nature Genetics*, vol. 2022. nature.com

10. J. S. Hawe, R. Wilson, K. T. Schmid, L. Zhou, et al., "Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function," *Nature Genetics*, 2022. nih.gov
11. V. Aggarwala, "Human Mutation/substitution Rate: Variability, Modeling And Applications," 2016. [PDF]
12. C. C. A Spencer, P. Deloukas, S. Hunt, J. Mullikin et al., "The Influence of Recombination on Human Genetic Diversity," 2006. ncbi.nlm.nih.gov
13. J. Stapley, P. Feulner, S. E. Johnston, A. W. Santure et al., "Recombination: the good, the bad and the variable," 2017. [PDF]
14. M. Kardos, E. E. Armstrong, S. W. Fitzpatrick, et al., "The crucial role of genome-wide genetic variation in conservation," *Proceedings of the National Academy of Sciences*, 2021. pnas.org
15. A. Khan, K. Patel, H. Shukla, A. Viswanathan, "Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers," *Proceedings of the National Academy of Sciences*, vol. 2021. pnas.org
16. B. E Stranger and E. T Dermitzakis, "From DNA to RNA to disease and back: The 'central dogma' of regulatory disease variation," 2006. ncbi.nlm.nih.gov
17. M. Jackson, L. Marks, G. H.W. May, and J. B. Wilson, "The genetic basis of disease," 2018. [PDF]
18. A. Ben-Mahmoud, V. Gupta, C. H. Kim, "Digenic or oligogenic mutations in presumed monogenic disorders: A review," *Journal of Genetic*, 2023. koreascience.kr
19. V. K. Pounraja and S. Girirajan, "A general framework for identifying oligogenic combinations of rare variants in complex disorders," Genome research, 2022. cshlp.org
20. I. Rudan and P. Rudan, "From Genomic Advances to Public Health Benefits: The Unbearable Lightness of Being Stuck," 2004. [PDF]
21. L. Ionela Butnariu, E. Țarcă, E. Cojocaru, C. Rusu et al., "Genetic Modifying Factors of Cystic Fibrosis Phenotype: A Challenge for Modern Medicine," 2021. ncbi.nlm.nih.gov
22. V. Martinez, "Sickle Cell Anemia: Current Treatments and Potential Advancements," 2017. [PDF]
23. R. E. Ellsworth, D. J. Decewicz, C. D. Shriver, and D. L. Ellsworth, "Breast Cancer in the Personal Genomics Era," 2010. ncbi.nlm.nih.gov
24. P. Lenart, J. Bienertová-Vašků, and L. Berec, "Learning mitigates genetic drift," 2022. ncbi.nlm.nih.gov
25. F. B. Livingstone, "Genetic drift and polygenic inheritance," 1972. [PDF]
26. J. Ronald and J. M Akey, "Genome-wide scans for loci under selection in humans," 2005. ncbi.nlm.nih.gov
27. R. Robert Ferrucci, "On the study of genetic structure in human populations and the effects of demographic history, consanguinity, and study design on detection, with investigations of human evolutionary models, archaic introgression, and natural selection," 2011. [PDF]
28. O. F. Khabour, A. A. Abu-Siniyeh, K. H. Alzoubi, and N. A. Al-Sheyab, "Challenges Faced by Behavioral Genetic Studies: Researchers Perspective from the MENA Region," 2021. ncbi.nlm.nih.gov
29. L. Beshir, "A Framework to Ethically Approach Incidental Findings in Genetic Research," 2020. ncbi.nlm.nih.gov

30. E. Wright Clayton, B. J Evans, J. W Hazel, and M. A Rothstein, "The law of genetic privacy: applications, implications, and limitations," 2019. ncbi.nlm.nih.gov
31. C. Brokowski, "Do CRISPR Germline Ethics Statements Cut It?," 2018. ncbi.nlm.nih.gov
32. M. Yang and J. W. Kim, "Principles of Genetic Counseling in the Era of Next-Generation Sequencing," 2018. ncbi.nlm.nih.gov
33. A. J. Marian, "Clinical Interpretation and Management of Genetic Variants," 2020. ncbi.nlm.nih.gov
34. Q. Wang, Q. Lu, and H. Zhao, "A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing," 2015. ncbi.nlm.nih.gov
35. M. Sayitoğlu, "Clinical Interpretation of Genomic Variations," 2016. ncbi.nlm.nih.gov
36. L. Cai, A. L. Fisher, H. Huang, and Z. Xie, "CRISPR-mediated genome editing and human diseases," 2016. ncbi.nlm.nih.gov
37. J. Bendl, M. Musil, J. Štourač, J. Zendulka et al., "PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions," 2016. ncbi.nlm.nih.gov
38. J. N. Cooke Bailey, M. A. Pericak-Vance, and J. L. Haines, "The Impact of the Human Genome Project on Complex Disease," 2014. ncbi.nlm.nih.gov
39. B. Alamad, K. Elliott, and J. C. Knight, "Cross-population applications of genomics to understand the risk of multifactorial traits involving inflammation and immunity," 2024. ncbi.nlm.nih.gov
40. N. Sayed, P. Allawadhi, A. Khurana, V. Singh, U. Navik, "Gene therapy: Comprehensive overview and therapeutic applications," Life Sciences, vol. 2022, Elsevier. [HTML]
41. M. R. Cring and V. C. Sheffield, "Gene therapy and gene correction: targets, progress, and challenges for treating human diseases," Gene therapy, 2022. [HTML]