Chapter 2

DNA Polymorphisms: DNA-Based Molecular Markers and Their Application in Medicine

Salwa Teama

Additional information is available at the end of the chapter

http://dx.doi.org/ 10.5772/intechopen.79517

Abstract

DNA polymorphisms are the different DNA sequences among individuals, groups, or populations. Polymorphism at the DNA level includes a wide range of variations from single base pair change, many base pairs, and repeated sequences. Genomic variability can be present in many forms, including single nucleotide polymorphisms (SNPs), variable number of tandem repeats (VNTRs, e.g., mini- and microsatellites), transposable elements (e.g., Alu repeats), structural alterations, and copy number variations. Different forms of DNA polymorphisms can be tracked using a variety of techniques; some of these techniques include restriction fragment length polymorphisms (RFLPs) with Southern blots, polymerase chain reactions (PCRs), hybridization techniques using DNA microarray chips, and genome sequencing. During the last years, the recent advance of molecular technologies revealed new discoveries of DNA polymorphisms. DNA polymorphisms are endless, and more discoveries continue at a rapid rate. Mapping the human genome requires a set of genetic markers. DNA polymorphism serves as a genetic marker for its own location in the chromosome; thus, they are convenient for analysis and are often used as in molecular genetic studies.

Keywords: copy number variations, genetic polymorphism, microsatellite, minisatellite, molecular markers, single nucleotide polymorphisms (SNPs), variable number tandem repeats (VNTRs)

1. Introduction

Genetic polymorphism is the existence of at least two variants with respect to gene sequences, chromosome structure, or a phenotype (gene sequences and chromosomal variants are seen at the frequency of 1% or higher), typical of a polymorphism, rather than the focus being on rare variants [1].

The human genome comprises 6 billion nucleotides of DNA packaged into two sets of 23 chromosomes, one set inherited from each parent. The probability of polymorphic DNA in humans is great due to the relatively large size of human genome. Genomic variability includes a wide range of variations from single base pair change, many base pairs, and repeated sequences [2].

Single nucleotide polymorphisms are the most common type of genetic variations in humans [3], due to their abundance across the human genome; single nucleotide polymorphisms (SNPs) have become important genetic markers for mapping human diseases, population genetics, and evolutionary studies. SNPs have become very important since technologies for DNA sequencing have become feasible and widely available. Advance continues at a rapid rate [4].

A major step forward in genome identification is the discovery of about 30–90% of the genome which is constituted by regions of repetitive DNA which are highly polymorphic in nature [5]. Polymorphic tandem repeated sequences have emerged as important genetic markers and initially, variable number tandem repeats (VNTRs) were used in DNA fingerprinting. In recent years, evidence has been accumulated for the involvement of VNTR repeats in a wide spectrum of pathological states [6].

Throughout the past years, scientists have believed that genes strictly came in two copies in a genome. However, with the recent advancement in molecular technology, discoveries have revealed substantial segments of DNA, ranging in size from thousands to millions of DNA bases that could vary in copy number. Such copy number variations (or CNVs) encompass gene copies, newly discovered CNVs are important sources of genomic diversity [7, 8].

The development and use of DNA-based molecular markers is one of the most significant developments in the field of molecular genetics that facilitate the study of genetic variations in health and diseases [5].

This chapter reviews the DNA-based genetic markers and their application in medicine, with a particular emphasis on common DNA-based genetic markers, including single nucleotide polymorphisms and short tandem repeats (STRs).

2. Polymorphisms at DNA level

Genomic variability at DNA level can be present in many forms including: single nucleotide polymorphisms, variable number of tandem repeats (e.g., mini- and microsatellites), transposable elements (e.g., Alu repeats), structural alterations, and copy number variations. It can occur in the nucleus or mitochondria. Two major sources: (1) mutations that may result as chance processes or have been induced by external agents such as radiation and (2) recombination. Once formed, it can be inherited, allowing its inheritance to be tracked from parent to child [3].

The genomes of humans may be divided into different parts based on known functional properties; the coding and noncoding regions mostly do not code for protein [2, 9]. The coding

regions contain DNA sequences which determine primarily the amino acid sequences of the proteins for which they code. Noncoding DNA generally containing DNA sequences with no function has not yet been discovered or possibly no function exists [10]; such sequences may be either single copy or exist as multiple copies called repetitive DNA [10]. Indeed, regions of DNA that do not code for proteins tend to have more polymorphisms. Recently, there has been substantial progress in understanding genome content which centered on discovered protein-coding genes which considered a functional DNA sequence moving away for discoveries of many repeat families, and various copy number variations encompass gene copies leading to dosage imbalance that plays an important role in genome structure, evolution, and diversity [11, 12]. "The Human Genome Project has revealed that humans have only 20,000–30,000 structural genes (protein-coding genes) (International Human Genome Sequencing Consortium, 2004)" [13].

3. Type of polymorphisms

3.1. Single nucleotide polymorphisms

Single base change is "high-density natural sequence variations in human genome" [14]. SNPs are mostly formed when errors occur (substitution, insertion and deletion). SNPs are prominent sources of variation in human genome and serve as excellent genetic markers. Some regions of the genome are richer in SNPs than others. SNPs may occur within gene sequences or in intergenic sequences. SNPs mostly are located in noncoding regions of the genome and have mostly no direct known impact on the phenotype of an individual but their role till now remains elusive, and depending on where SNPs occurs, it might have different consequences at the phenotypic level [3].

3.2. Insertion/deletion polymorphisms

It is a type of DNA variation in which a specific nucleotide sequence of various lengths ranging from one to several 100 base pairs is inserted or deleted. Indels are widely spread across the genome. Some authors consider one base pair as SNPs or repeat insertion/deletion as indels.

3.3. Polymorphic repetitive sequences

DNA repeats can be classified as interspersed repeats or tandem repeats. This can comprise over two-thirds of the human genome [15]. Interspersed repeats are dispersed across the genome within gene sequences or intergenic and include retro (pseudo) genes and transposons. Tandem repeats or variable number tandem repeats (≥2 bp in length) that are adjacent to each [16] can involve as few as two copies or many thousands of copies. Centromeres and telomeres largely comprise tandem repeats. Despite increasing evidence on the functionality of DNA repeats, their biologic role is still elusive and under frequent debate [11]. Tandem repeats are organized in a head-to-tail orientation; based on the size of each repeat unit, satellite repeats can be further divided into macrosatellites, minisatellites, and microsatellites [17]. Some of these repeats are

described as follows: macrosatellites, with sequence repeats longer than 100 bp, are the largest of the tandem DNA repeats, located on one or multiple chromosomes [11], minisatellites, stretches of DNA, are characterized by moderate length patterns, 10–100 bp usually less than 50 bp [9, 18], and microsatellites also known as short tandem repeats (STRs) repeat units of less than 10 bp, [3].

3.4. Structural and copy number variations

Structural and copy number variations (CNVs) are another frequent source of genome variability [6, 19, 20]. The term CNVs therefore encompasses previously introduced terms such as large-scale copy number variants (LCVs) [19], copy number polymorphisms (CNPs) [20], and intermediate-sized variants (ISVs) [21]. Some currently used terms are structural variations; a genomic alteration (e.g., an inversion) that involves segments of DNA > 1 kb, copy number polymorphisms; a duplication or deletion event involving >1 kb of DNA [22], intermediate-sized structural variant; and a structural variant that is ~8–40 kb in size, this can refer to a CNVs or a balanced structural rearrangement (e.g., an inversion) [21].

4. Common DNA-based molecular markers

The development and use of molecular methods for the detection of DNA molecular markers is one of the most significant progresses in the field of molecular genetics. Mapping the human genome requires a set of genetic markers to which we can relate the position of genes. Some of these markers are genes, others SNPs and VNTRs. Molecular markers can be used to mark in genomes for various purposes such as mapping human diseases, pharmacogenetics, and human identification.

4.1. Single nucleotide polymorphisms

Single base pair change leads to single nucleotide variant, probably accounting for many genetic conditions caused by single gene or multiple genes. SNPs represent the major source of human genomic variability. Due to the lack of knowledge on exact SNP number, it is difficult to give a direct estimate of the number of the SNPs in the human genome but in different public and private data bases, more than 5 million have been recorded and about 4 million validated [23]. “The data from the Human Genome project revealed that that human nucleotide sequence differs every 1000-1500 bases from one individual to another” [24]. “The SNP Map working group observed that two haploid genomes differ at 1 nucleotide per 1331 bp”. Over 60,000 however are within genes and some of them associated with diseases [2].

Single nucleotide polymorphisms within protein-coding regions either synonymous polymorphisms; those that do not have any effect on the organism and are said to be selectively silent as the substitution causes no amino acid change in the protein produced (silent mutation) or nonsynonymous substitution results in change in encoded amino acids either missense mutation; change the protein through codon alteration or nonsense mutation results in a chain termination codon [3].

Single nucleotide polymorphisms within a coding sequence cause genetic diseases including sickle cell anemia. SNPs responsible for a disease can also occur in any genetic region that can *eventually* affect the expression activity of genes, for example, in promoter regions. SNPs in the noncoding region of the gene, though their effect is still debatable, most of the genome mostly consists of regulatory elements that control gene expression, but these regions have remained largely unexplored in clinical diagnostics due to the high cost of whole genome sequencing and interpretive challenges. Clinical diagnostic sequencing currently focuses on identifying causal mutations in the exome, where most disease-causing mutations are known to occur.

Another important group of SNPs is the one that alters the primary structure of a protein involved in drug metabolism; these SNPs are targets for pharmacogenetics studies.

However, some SNPs are not causative, some SNPs are in close association with, and therefore segregate with, a disease-causing sequence so, the presence of SNP correlates with the presence or an increased risk of developing the disease; these SNPs are useful in diagnostics, disease prediction, and other applications [3].

Single nucleotide polymorphisms can be used as genetic markers for constructing high genetic maps and to carry out association studies related to diseases because of their abundance and the availability of high throughput analysis technologies. SNPs have become an important application in the development and research of genetic markers [14].

There are numerous strategies that can be implemented to new single nucleotide variant (SNVs) discoveries; the most common and well-known method is by direct sequencing and in comparison to a puplic or other sequence date base [25, 26] or locus specific amplification of target genomic region followed by sequence comparison [27, 28]; prescreening prior to sequence determination is needed. SNV detection encompasses two broad areas: (1) scanning DNA sequences for previously unknown polymorphisms and (2) screening (genotyping) individuals for known polymorphisms. Scanning for new SNVs can be further classified to two different types of approaches, the first one being the global (or random approach) and the other one being the regional (targeted approach) [14]. There are certain methods which have been developed for using SNVs randomly in the genome; "such as representation shotgun sequencing [14, 29], primer-ligation-mediated PCR [14, 30] and degenerate oligonucleotide–primed PCR" [14, 31].

Haplotypes are groups of SNPs that are generally inherited together. Haplotypes can have stronger correlations with diseases or other phenotypic effects compared with individual SNPs and may therefore provide increased diagnostic accuracy in some cases [32].

4.2. Microsatellites (short tandem repeats)

Microsatellites are short tandem repeats (STRs), repeat units, or motifs of less than 10 bp; because of high variability, microsatellite loci are often used in forensics, population genetics, and genetic genealogy. Significant associations were demonstrated between microsatellite variants and many diseases [15].

Depending on the search algorithm, there are approximately 700,000–1,000,000 microsatellite loci which are 2–6 bp long in the human reference genome [33, 34]. Di- and tetra-nucleotides constitute about 75% of microsatellites, with the remaining loci containing tri-, penta, and hexanucleotide. Within genes, STRs are nonrandomly distributed across protein-coding sequences, untranslated regions (UTRs), and introns. STRs containing dinucleotide repeat units that are much more abundant in the regulatory or UTR regions than in other genomic regions. In the coding regions of the genes, repeats mostly have either trimeric or hexameric repeat unit, likely as a result of selection against frameshift mutations [34, 35]. "The mutation rates of STRs often lie between 10^3 and 10^6 per cell generation which is 10- to 10^5-fold higher than the average mutation rates observed in nonrepeated regions of the genome"[36, 37].

"Polymorphism of tandem repeats within protein-coding regions reveals that tandem repeat variation is an important source of variation in many proteins, many of this variation is of significant impact on protein function. Tandem repeats has been associated with a number of diseases and phenotypic conditions, changes in the protein products of genes, leading to diseases, other tandem repeat polymorphisms in noncoding regions are known to modify function through their impact on gene regulation". "These polymorphisms can arise from events such as unequal crossover, replication slippage or double-strand break repair" [38].

Variations in the STR length play a significant role in modulating gene expression and STRs are likely to be general regulatory elements; regulatory STRs manifest significant polymorphism because of their high intrinsic mutation rate [15].

There are examples for distinctive phenotypic changes and diseases that are directly associated with the increases or decreases of microsatellite repeat arrays; for example, considering Huntington disease gene, triplet nucleotide mutations, the mutation that causes the disease, is an expansion of CAG repeats from the normal range of 11–14 copies to abnormal range of at least 38 copies. The extra CAG repeats that causes extra glutamine is produced [9] and there are more than 40 neurological diseases in humans, such as spinocerebellar ataxia with polyglutamine tracts, which are caused by microsatellite motif length changes in trinucleotide arrays [39].

Testing candidate genes for polymorphisms in exons, promoters, splice sites, or other regulatory regions will have to be done using SNP testing, because it is the most common polymorphisms and more likely responsible for phenotypic variations. For complex phenotypic traits and candidate loci, single-loci SNP analyses present less information due to the bi-allelic nature of the markers, as compared to the multi-allelic microsatellites. However, performing haplotype frequency may improve the accuracy [40]. Recently, polymorphic tandem repeated sequences and coy number variations have emerged as important sources of genomic diversity that facilitate the study of genetic variations in health and diseases.

5. The major technique for DNA-based molecular marker detection

Different forms of DNA-based molecular markers can be tracked using a variety of techniques. Some of these techniques include RFLPs with Southern blots and polymerase chain reactions (PCRs). Recently great advances in methodology for DNA polymorphisms detection using

real time PCR, hybridization techniques using DNA microarray chips, genome sequencing each technique has its own advantage and disadvantage.

5.1. Restriction fragment length polymorphism with southern blot

DNA digestion with restriction enzyme endonuclease cuts DNA at a specific sequence pattern known as a restriction endonuclease recognition site. Thus, the alleles differ in length and can be distinguished by gel electrophoresis, which can arise from a number of genetic events including point mutation in restriction sites, mutation that creates a new restriction site, insertion, deletion, and repeated sequences. The first polymorphic RFLP was described in 1980. RFLPs were the original DNA targets used for human identification, parentage testing, and gene mapping.

The method of hybridization of DNA with probes is called Southern blotting, after the name of the inventor, Southern [41]. RFLP requires relatively large amounts of DNA. Hence, it cannot be performed with the samples degraded by environmental factors and also takes longer time to get the results [42, 43]. PCR-RFLP is now replaced to avoid using Southern blot.

5.2. Polymerase chain reaction

In-vitro amplification of particular DNA sequences with the help of specifically chosen primers and DNA polymerase enzyme is done. The amplified fragments are separated electrophonically and detected by different staining methods. Real-time PCR useful modification of PCR can detect polymorphisms by various methodologies using real-time PCR chemistries, for example, TaqMan assay or molecular beacons.

5.3. Genomic array technology

Genomic array technology is a type of hybridization analysis allowing simultaneous study of large numbers of targets or samples. In 1987, macroarray evolved into the microarray. Tens of thousands of targets can be screened simultaneously in a very small area. Automated depositing systems (arrayers) can place thousands of spots on glass substrate of the size of a microscope slide (chip) with spotting representative sequences of each gene in triplicate, simultaneous screening of the entire human genome on a single chip. This technique facilitates the process of identifying specific homozygous and heterozygous alleles, by comparing the disparity of hybridization of the target DNA with each redundant probe. Microarray is also used to characterize genetic diversity and drug responses, to identify new drug targets, and to assess the toxicological properties of chemicals and pharmaceuticals [44].

5.4. Sequencing

Since technologies for rapid DNA sequencing have become available they are now widely used. There is a great progression for the detection of single nucleotide variants (SNVs) by direct sequencing, but intermediate-sized (from 50 bp to 50 kb) structural variants (SVs) remain a challenge. Such variants are too small to detect with cytogenetic methods but too large to reliably discover with short-read DNA sequencing. Recent high-quality genome

assemblies using long-read sequencing have revealed that each human genome has approximately 20,000 structural variants, spanning 10 million base pairs, more than twice the number of bases affected by SNVs. New long-read sequencing approaches are needed to meet this challenge, as short-read sequencing technologies only detect 20% of the SVs present in the human genome [45–48].

6. The major application for DNA-based genetic markers

DNA-based molecular markers are such powerful tools for mapping human diseases and discover many multifactorial diseases and disorders.

6.1. Mapping human diseases and risk prediction

Genetic mapping and linkage: The mapping of the human genome has made possible to develop a haplotype map in order to better define human SNV variability. The haplotype map or HapMap will be a tool for the detection of human genetic variation that can affect health and diseases [23]. The HapMap project is far more useful because it will reduce the number of SNVs required to examine the entire genome for association with a phenotype or diseases from the 10 million SNPs that are expected to exist to approximately tag 500,000 SNPs [38]. The first large-scale effort to produce a human genetic map was performed mainly using RFLP; other several projects are underway to identify more markers in humans and to make this data publicly available to scientists worldwide. Many groups that are involved in these massive efforts through DNA polymorphisms discovery resource include the SNP consortium (TSC) http://snp.cshl.org [49, 50]. The reason for the current enormous interest in SNPs is the hope that they could be used as markers to identify genes that predispose individuals to common, multifactorial disorders by using linkage disequilibrium (LD) mapping.

"The HapMap Project (http://hapmap.ncbi.nlm.nih.gov/), and other approaches, such as genome wide association studies, have been widely reported for complex polygenic diseases, with some interesting novel genes affecting disease susceptibility now identified. Genome Wide Association; the GWAS has now been used for a large range of traits and diseases e.g. baldness and eye color" [51, 52].

6.2. Quantitative trait loci mapping, candidate genes, and complex traits

The identification of genes affecting complex trait is a very difficult task. For many complex traits, the observable variation is quantitative, and loci affecting such traits are generally termed quantitative trait loci (QTL). (SNVs) can be used as genetic markers for constructing high-density genetic maps and to carry out association studies related to complex traits and diseases [14].

6.3. Pharmacogenetics

Individual response to a drug is governed by many factors such as genetics, age, sex, environment, and disease. The influence of genetic factors on the response of a drug is a known fact.

Polymorphic STRs, together with SNPs and CNVs, can explain variability in response to pharmacotherapy because of their prevalence in the human genome and their functional role as regulators of gene expression and its applications. Pharmacogenetics is the study of the influence of genetics factors on drug response and metabolism. The science of pharmacogenetics when applied can be used to evade adverse drug reactions, predict toxicity and therapeutic failure, and refine therapeutic efficiency and improve clinical outcomes [53].

7. DNA fingerprinting and human identification

Establishing an individual's identity is one of the uses of DNA sequence information that highlights uniqueness of a particular sample [5], also known as genetic fingerprinting; DNA typing and DNA profiling are molecular genetic methods that enable the identification of individuals using hair, blood, semen, or other biological samples, based on unique patterns in their DNA. This uniqueness in each individual is the basis of human identification at the DNA level, forensic identification, determination of genetic variation, determination of family relationship, and one important instance is identifying good genetic matches for organ or marrow donation. When first described in 1984 by British scientist Alec Jeffreys, the technique used was minisatellites; these sequences are unique to each individual, with the exception of identical twins. Different DNA fingerprinting methods exist, using either restriction fragment length polymorphism (RFLP) or PCR or both. More than 200 RFLP loci have been described in human DNA. Initially, forensic medicine used minisatellite testing; however, this method requires a large amount of material and yield low-quality results especially when only little amount of materials are available. Nowadays, in most forensic samples, the study of DNA is usually performed by microsatellite analysis. The most useful microsatellite for human identification is those with a greater number of alleles, smaller size, higher frequency of heterozygotes (higher than 90%), and low frequency of mutations [43]. Among others, the microsatellite DNA marker has been the most widely used, due to its easy use by simple PCR, followed by a denaturing gel electrophoresis [40]. Each person has some STRs that were inherited from the father and some from mother, useful in paternity testing but however no person has STRs that are identical to those of either parent. The uniqueness of an individual's STR provides the scientific marker of identity and hence is helpful in forensic identification [54]. Genomic and mitochondrial are two types of DNA which are used in forensic sciences. The genomic DNA is found in the nucleus of each cell in the human body and represents a DNA source for most forensic applications. Mitochondrial DNA (mt DNA) is another source of material that can be used; various biological samples such as hair, bones, and teeth that lack nucleate cellular materials can be analyzed with mt DNA [43, 55].

7.1. Sex-chromosome STR testing

"Majority of the length of the human Y chromosome is inherited as a single block in linkage from father to male offspring as a haploid entity. DNA genetic markers on the human Y chromosome are valuable tools for understanding human evolution, migration and for tracing relationships among males" [43, 56]. "Chromosome X specific STRs is used in the

identification and the genomic studies of different ethnic groups worldwide, because the small size of X-chromosome STR alleles; about 100–350 nucleotides, it is relatively easy to be amplified and detected with high sensitivity" [43].

7.2. DNA typing and engraftment monitoring

DNA typing becomes the method of choice for engraftment monitoring, donor cells are examined by following donor polymorphisms in the recipient blood and bone marrow. Although RFLP can efficiently differentiate donor and recipient cells, the detection of RFLP requires the use of southern blot methods, which is too labor intensive and has limited sensitivity for this application, in comparison with small minisatellites or microsatellites that are easily detected by PCR amplification, because of increased rapidity and the 0.5–1% sensitivity achievable with PCR. Sensitivity can be raised to 0.01% using Y-STR, but this approach is limited to that transplant from sex mismatched donor recipient pairs preferably from a female donor to a male recipient [2].

Nowadays, DNA fingerprinting is used as a tool for designing "personalized" medical treatments for cancer patients.

8. Conclusion and future perspectives

Single nucleotide polymorphisms (SNPs) have become an important application in the development and research of genetic diseases or other phenotypic traits. Haplotypes are groups of SNPs that are generally inherited together. Haplotypes can have stronger correlations with diseases or other phenotypic effects compared with individual SNPs and may therefore provide increased diagnostic accuracy in some cases.

Polymorphic tandem repeated sequences have emerged as important genetic markers and initially, variable number tandem repeats (VNTRs) were used in DNA fingerprinting; in recent years, evidence has been accumulated for the involvement of VNTR repeats in a wide spectrum of pathological states.

The new global CNV map will transform medical research in four main areas: detection for genes underlying common diseases, study of familial genetic conditions, exclude variation found in unaffected individuals, helping researchers to target the region that might be involved and the data generated will also contribute to a more accurate and complete human genome reference sequence used by all biomedical scientists. Currently, approximately 2000 CNVs have been described; there could be thousands more CNVs in the human population. About 100 CNVs were detected in each genome tested with the average size being 250,000 bases (an average gene is 60,000 bases). With advanced molecular technologies more CNVs will be discovered and more DNA samples from worldwide populations are examined.

Recently, there has been substantial progress in understanding genome content which centered on protein-coding genes which considered a functional DNA sequence moving away for many discoveries, many repeat families, and various copy number variations that play

an important role in genome structure, evolution, and diversity. Additional efforts are being placed to develop strategies that would overcome the obstacles in alignment next-generation sequencing data. "Future precision medicine efforts will direct to connect genotypes to phenotypes and distinguish common, from rare or potentially disease linked variants. New long-read sequencing approaches are needed to meet this challenge."

Other important applications of genetic polymorphism knowledge are improving health care through gene therapy, discovery of new drugs and drug targets, and upgradation of the discovery processes with advanced technologies.

Advances in molecular technologies, DNA sequencing technology, and microarray, coupled with novel, efficient computational analysis tools, have made it possible to analyze sequence-based experimental data, more discoveries, and development at a rapid rate.

Conflict of interest

The author declares that there is no conflict of interest.

Author details

Salwa Teama

Address all correspondence to: salwateama2004@yahoo.com

Molecular Genetics Unit, Medical Ain Shams Research Institute, Faculty of Medicine, Ain Shams University, Cairo, Egypt

References

[1] Daly AK. Pharmacogenetics and human genetic polymorphisms. The Biochemical Journal. 2010;**429**(3):435-449. DOI: 10.1042/BJ20100522

[2] Buckingham L. Chromosomal structure and chromosomal mutation. In: Buckingham L, editor. Molecular Fundamentals Methods and Clinical Applications. 2nd ed. Philadelphia: F.A. Davis Company; 2012. Chapter 8. ISBN.0-8036-2677-0

[3] Ismail S, Essawi M. Genetic polymorphism studies in humans. Middle East Journal of Medical Genetics. 2012;**1**:57-63

[4] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M,

Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, KJ MK, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. 2011;**475**(7356):348-352. DOI: 10.1038/nature10242

[5] Rao SR, Trivedi S, Emmanuel D, Merita K, Hynniewta M. DNA repetitive sequences types, distribution and function: A review. Journal of Cell and Molecular Biology. 2010; **7**(2) & **8**(1):1-11

[6] Bruce HA, Sachs N, Rudnicki DD, Lin SG, WillourVL, Cowell JK, Conroy J, McQuaid DE, Rossi M, Gaile DP, Nowak NJ, Holmes SE, Sklar P, Ross CA, Delisi LE, Margolis RL. Long tandem repeats as a form of genomic copy number variation: Structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. Psychiatric Genetics. 2009;**19**(2):64-71. DOI: 10.1097/YPG.0b013e3283207ff6

[7] Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery LN, Ishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature. 2006;**444**(7118):444-454. DOI: 10.1038/nature05329

[8] Clancy S. Copy number variations (CNVs) have been linked to dozens of human diseases, but can they also represent the genetic variation that was so essential to our evolution. Nature Education. 2008;**1**(1):95. https://www.nature.com/scitable/topicpage/copy-number-variation-445

[9] Weaver R. Transmission genetics. In: Weaver R, editor. Molecular Biology. Fourth ed. McGraw Hill International Edition; 2008. Chapter 1. ISBN. 978-0-07-110216-2

[10] Fowler JCS, Burgoyne LA, Scott AC, Harding HWJ. Repetitive deoxyribonucleic acid DNA and human genome variation—A concise review relevant to forensic biology. Journal of Forensic Science. 1988;**33**:1111-1126

[11] Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. Biology Direct. 2011;**6**:19. DOI: 10.1186/1745-6150-6-19

[12] Dumbovic G, Forcales S, Perucho M. Emerging roles of macrosatellite repeats in genome organization and disease development. Epigenetics. 2017;**12**(7):515-526. DOI: 10.1080/15592294.2017.1318235

[13] Miller RD, Kwok P-Y. Single nucleotide polymorphisms in the public domain: How useful are they? Nature Genetics. 2001;**27**:371-372

[14] Xu J, Xu G, Chen S. A new method for SNP discovery. BioTechniques. 2009;**46**(3):201-208. DOI: 10.2144/000113075

[15] Krynetskiy E. Beyond SNPs and CNV: Pharmacogenomics of polymorphic tandem repeats. 2017;**8**:170. DOI: 10.4172/2153-0645.1000170

[16] O'Dushlaine CT, Edwards RJ, Park SD, Shields DC. Tandem repeat copy number variation in protein-coding regions of human genes. Genome Biology. 2005;**6**:R69. DOI: 10.1186/gb-2005-6-8-r69

[17] Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiology and Molecular Biology Reviews. 2008;**72**(4):686-727. DOI: 10.1128/MMBR.00011-08

[18] Jeffreys AJ, Wilson V, Thein SL. Individual specific "fingerprints" of human DNA. Nature. 1985;**316**:76-79. DOI: 10.1038/316076a0

[19] Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nature Genetics. 2004; **36**(9):949-951. DOI: 10.1038/ng1416

[20] Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M. Strong association of de novo copy number mutations with autism. Science. 2007;**316**(5823):445-449. DOI: 10.1126/science.1138659

[21] Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. Nature Genetics. 2005;**37**(7):727-732. DOI: 10.1038/ng1562

[22] Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nature Reviews. Genetics. 2006;**7**(2):85-97. DOI: 10.1038/nrg1767

[23] Sobrinoa B, Mb B'n, Carracedo A. SNPs in forensic genetics: A review on SNP typing methodologies. Forensic Science International. 2005;**154**(2-3):181-194. DOI: 10.1016/j. forsciint.2004.10.020

[24] Kruglyak L, Nickerson DA. Variation is the spice of life. Nature Genetics. 2001;**27**(3):234-236. DOI: 10.1038/85776

[25] Guryev VE, Berezikov E, Cuppen E. CASCAD: A database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. BMC Genomics. 2005;**6**:10. DOI: 10.1186/1471-2164-6-10

[26] Matukumalli LK, Grefenstette JJ, Hyten DL, Choi IY, Cregan PB, Van Tassell CP. SNP PHAGE—High throughput SNP discovery pipeline. BMC Bioinformatics. 2006;**7**:468. DOI: 10.1186/1471-2105-7-468

[27] Twyman RM. SNP discovery and typing technologies for pharmacogenomics. Current Topics in Medical Chemistry. 2004;**4**:1423-1431

[28] Suh Y, Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. Mutation Research. 2005;**573**:41-53. DOI: 10.1016/j.mrfmmm.2005.01.005

[29] Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;**407**(6803):513-516. DOI: 10.1038/35035083

[30] Makrigiorgos MG. PCR based detection of minority point mutations. Human Mutation. 2004;**23**:406-412. DOI: 10.1002/humu.20024

[31] Gupta PK, Roy JK, Prasad M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science. 2001;**80**(4):524-535

[32] Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF. Haplotype variation and linkage disequilibrium in 313 human genes. Science. 2001;**293**(5529):489-493. DOI: 10.1126/science.1059431

[33] Ellegren H. Microsatellites: Simple sequences with complex evolution. Nature Reviews. Genetics. 2004;**5**(6):435-445. DOI: 10.1038/nrg1348

[34] Willems T, Gymrek M, Highnam G, Mittelman D, Erlich Y. 1000 genomes project consortium, the landscape of human STR variation. Genome Research. 2014;**24**:1894-1904. DOI: 0.1101/gr.177774.114

[35] Bolton KA, Ross JP, Grice DM, Bowden NA, Holliday EG, Avery-Kiejda KA, Scott RJ. STaRRRT: A table of short tandem repeats in regulatory regions of the human genome. BMC Genomics. 2013;**14**:795. DOI: 10.1186/1471-2164-14-795

[36] Legendre M, Pochet N, Pak T, Verstrepen KJ. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Research. 2007;**17**(12):1787-1796. DOI: 10.1101/gr.6554007

[37] Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nature Genetics. 2005;**37**(9):986-990. DOI: 10.1038/ng1618

[38] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002;**296**(5576):2225-2229. DOI: 10.1126/science.1069424

[39] Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: Mechanisms of dynamic mutations. Nature Reviews Genetics. 2005;**6**(10):729-742. DOI: 10.1038/nrg1689

[40] Vignal A, Milana D, Sancristobala M, Eggenb A. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution. 2002;**34**(3):275-305. DOI: 10.1051/gse: 2002009

[41] Southern E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology. 1975;**98**(3):503-517

[42] Zehner R, Zimmermann S, Mebs D. RFLP and sequence analysis of the cytochrome B gene of selected animals and man: Methodology and forensic application. International Journal of Legal Medicine. 1998;**111**(6):323-327

[43] Datta P, Sood S, Rastogi P, Bhargava K, Bhargava D, Yadav M. DNA profiling in forensic dentistry. Journal of the Indian Academy of Forensic Sciences. 2012;**34**(2):156-159. ISSN: 09710973

[44] Rapley R, Harbron S. Molecular Analysis and Genome Discovery. Chichester: John Wiley & Sons Ltd; 2005. DOI: 10.1002/0470020202.index. Online ISBN: 9780470020203

[45] Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, Fu A, Li Q, Li N, Gong S, Lintner KE, Ding Q, Wang Z, Hu J, Wang D, Wang F, Wang L, Lyon GJ, Guan Y, Shen Y, Evgrafov OV, Knowles JA, Thibaud-Nissen F, Schneider V, Yu CY, Zhou L, Eichler EE, So KF, Wang K. Long-read sequencing and de novo assembly of a Chinese genome. Nature Communications. 2016;**7**:12065. DOI: 10.1038/ncomms12065

[46] Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, Kuk J, Park GH, Kim J, Ryu H, Kim J, Roh M, Baek J, Hunkapiller MW, Korlach J, Shin JY, Kim C. De novo assembly and phasing of a Korean human genome. Nature. 2016;**538**(7624):243-247. DOI: 10.1038/nature20098

[47] Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stütz AM, Stedman W, Anantharaman T, Alex Hastie A, Dai H, Fritz MH, Cao H, Cohain A, Deikus G, Durrett RE, Blanchard SC, Altman R, Chin C, Guo Y, Paxinos E, Korbel JO, Darnell RB, McCombie WR, Kwok PY, Mason CE, Schadt EE, Bashir A. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nature Methods. 2015;**12**(8):780-786. DOI: 10.1038/nmeth.3454

[48] Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, Graves-Lindsay TA, Munson KM, Kronenberg ZN, Vives L, Peluso P, Boitano M, Chin CS, Korlach J, Wilson RK, Eichler EE. Discovery and genotyping of structural variation from long-read haploid genome sequencing data. Genome Research. 2017;**27**(5):677-685. DOI: 10.1101/gr.214007.116

[49] Thorisson GA, Stein LD. The SNP consortium website: Past, present and future. Nucleic Acids Research. 2003;**31**(1):124-127. PMID: 12519964

[50] Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Research. 1998;**8**(12):1229-1223. PMID: 9872978

[51] Hillmer AM, Brockschmidt FF, Hanneken S, Eigelshoven S, Steffens M, Flaquer A, Herms S, Becker T, Kortüm AK, Nyholt DR, Zhao ZZ, Montgomery GW, Martin NG, Mühleisen TW, Alblas MA, Moebus S, Jöckel KH, Bröcker-Preuss M, Erbel R, Reinartz R, Betz RC, Cichon S, Propping P, Baur MP, Wienker TF, Kruse R, Nothen MM. Susceptibility variants for male-pattern baldness on chromosome 20p11. Nature Genetics. 2008;**40**(11):1279-1281. DOI: 10.1038/ng.228

[52] Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D, Zhu G, Larsson M, Duffy DL, Montgomery GW, Mackey DA, Walsh S, Lao O, Hofman A, Rivadeneira F, Vingerling JR, Uitterlinden AG, Martin NG, Hammond GJ, Kayser M. Digital quantification of human eye color highlights genetic association of three new loci. PLoS Genetics. 2011;**6**(e1000934):1-15. DOI: 10.1371/journal.pgen.1000934

[53] Abraham BK, Adithan C. Genetic polymorphism of CYP2D6. Indian Journal of Pharmacology. 2001;**33**:147-169

[54] Butler, JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. Journal of Forensic Sciences. 2006;**51**(2):253-265. DOI: 10.1111/j.1556-4029.2006.00046.x

[55] Sweet D, Sweet CHW. DNA analysis of dental pulp to link incinerated remains of homicide victim to crime scene. Journal of Forensic Sciences. 1995;**40**(2):310-314

[56] Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, Su B. Global distribution of Y chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. Journal of Human Genetics. 2010;**55**(7):428-435. DOI: 10.1038/jhg.2010.40