Glossary of some genetic terms (in alphabetical order) (3,4)

Allele: One of two or more versions of DNA sequence or a gene at a given locus.

Allele-specific oligonucleotide (ASO): A short DNA probe synthesized to match a particular DNA sequence, allowing the discrimination of alleles that differ only by a single base.

Allelic heterogeneity: Different variants at the same gene or same locus resulting in a similar phenotype.

Analytic validity: The ability a test to identify the present or absence of a specific variant or allele as the test was designed to detect.

Aneuploidy: A form of heteroploidy in which the chromosome number is not an exact multiple of haploid number (human haploid number: 23).

Aneusomy: Abnormal number of chromosomes in the cell.

Anticipation: The progressive earlier onset and increased severity of a disease in successive generations in a family. The phenomenon is a feature of trinucleotide repeat (TNR) disorders.

Autosome: Any numbered chromosome or non-sex chromosomes (chromosome 1 to chromosome 22).

Autosomal dominant inheritance: Inheritance pattern in which a child inherits one copy of an affected variant or affected allele from one parent which is enough to cause a condition (50% or ½ chance of inheriting abnormal allele).

Autosomal recessive inheritance: Inheritance pattern in which a child inherits one copy of an affected variant or an affected allele from each parent to cause a condition (25% or ¼ chance of inheriting one abnormal allele from each parent or both affected alleles).

Breakpoint: A locus on a chromosome or a DNA sequence where a DNA change (alteration) or a break occurs such as deletion, inversion, or translocation.

Chimera: An individual composed of cells derived from two genetically different zygotes. In humans, blood group chimeras result from an exchange of hematopoietic stem cells by dizygotic twins in utero.

Chromosome: A thread-like structure in which DNA is packaged in orderly fashion in the cell.

Chromosomal syndrome: A disorder caused by an abnormality in the number or structure of chromosomes

Cis configuration (Coupling): Both variants are on one gene or homologous chromosome.

CLIA (Clinical Laboratory Improvement Amendments): Established quality standards for laboratory testing performed on specimens from humans, such as blood, body fluids and tissues, for the purpose of diagnosis, prevention, treatment of disease, or assessment of health.

Clinical heterogeneity: Multiple and variable clinical characteristics associated with a specific disorder.

Clinical utility: The ability of the test to improve the medical care of an individual.

Clinical validity: The ability of a clinical test to determine the presence or absence of a particular disease that the test was designed to ascertain.

Codon (triplet nucleotides): A triplet, three bases or trinucleotides in a DNA or RNA molecule, specifying a single amino acid.

Compound heterozygote (compound heterozygous): The presence of two different alleles of a particular gene (two different affected alleles, each inherited from one parent).

Conservative substitution (mutation): A replacement of one amino acid with another one with similar properties.

Consanguinity: The state of being related by descent from a common ancestor.

Contiguous gene syndrome: A syndrome resulting from a deletion or duplication of chromosomal DNA extending over two or more continuous genes.

Copy number variant (CNV): An unbalanced variation in DNA sequence defined by the loss or addition of a segment of DNA, typically defined as larger than 1 kb and ranging up to 3 Mb. CNVs may be alleles that involve tandem replications that involve two or more copies of a DNA segment.

Co-segregation: The transmission of 2 or more genes on the same chromosome, as a result of the physical proximity to each other. The distance is measured in map units, or centi Morgans (cM). One map unit is equivalent to 1% crossing-over between two linked genes.

Clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated protein 9 (Cas9) (CRISPR-Cas 9): Genome editing (aka., gene editing) was developed using a naturally occurring genome editing system from *Escherichia* coli as an immune defense system. This technology allows for introducing or

correcting mutations by cutting DNA at specific nucleotides and replacing them with desired nucleotides.

The bacteria use this system for the defense against a viral attack. Bacteria capture small regions of the virus' DNA and insert them into their own DNA, creating a segment known as CRISPR arrays. If the same virus attacks the bacteria, the bacteria can produce RNA (nucleotide) sequences from the CRISPR arrays that recognize and attach to specific regions of the virus' DNA. Then, the bacteria are able to use a similar enzyme to cut the DNA and disable the virus. Currently, this technology is being investigated to be used in human diseases.

Crossover (Crossing over): The process of reciprocal chromosomal segment exchange between chromatids of homologous chromosome, during prophase of the first meiotic division.

Deletion: A loss of DNA segment.

***De novo* mutation:** A new mutation which is not inherited from a parent or the previous generation.

Differentially methylated region (DMR): Genomic regions or genes where different methylation patterns occur, during gametogenesis, fertilization, or embryogenesis. It can also refer to different tissues or different developmental stages.

Digenic inheritance: A situation in which a phenotype is explained by a combination of genotypes at two independent loci.

Diploid: The presence of two sets of chromosomes, as in most somatic cells or zygotes.

Disomy: Two copies of chromosomes.

DNA methylation: The addition of a methyl residue to the 5-position of the pyrimidine ring of a cytosine base in DNA to form 5-methylcytosine.

Dominant: A trait is phenotypically expressed in heterozygotes.

Dominant negative: A variant or allele whose protein product disrupts the function of a wild-type or "normal" allele in the same cell.

Duplication: The presence of one or more copies of a small segment of gene, gene, or locus of a chromosome.

Exon: A region of a gene that is represented in mature messenger RNA after removal of introns or a protein coding region.

Expressivity: The extent to which a genetic trait is expressed.

Founder effect: A high frequency of a variant allele in a population founded by a small group when one of more the founders was a carrier of the allele.

Frameshift mutation or variant: A deletion or insertion of DNA (nucleotides) that is not an exact multiple of three base pairs, shifting the reading-frame of the gene, starting at the alteration.

Fluorescence in situ hybridization (FISH): A laboratory technique for detecting and identifying a specific DNA sequence within a chromosome using a fluorescent probe.

Gain-of-function mutation or variant: A genetic change associated with an increase in one or more of the normal functions of a protein (hypermorphic) or a new function for the protein (neomorphic).

Gene: A hereditary unit or a sequence of chromosomal DNA that is required to produce a functional product.

Gene mutation: Alternation of the sequence of DNA (nucleotides) in a gene. The change may or may not result in any phenotypic effect.

Genetic marker: A locus that has readily classifiable alleles or with known location that can be used in genetic studies (polymorphism).

Genetic heterogeneity: The same condition or disorder caused by different genetic variants or alleles.

Genome (Human): The complete DNA sequence containing the entire genetic information of an individual or a population.

Genome sequencing: The process of determining the sequence of DNA bases or nucleotides in the genome, including coding and noncoding.

Genome-wide association study (GWAS): A genetic study using hundreds to millions of variants at polymorphic loci distributed throughout the genome to identify certain variants (SNPs) or gene-related SNPs that are associated with a particular trait or a condition.

Genomic imprinting: An epigenetic phenomenon by which monoallelic expression is determined by the parental origin (paternal vs maternal).

Genotype: The genetic constitution of an individual or particular alleles that are present at one or more loci.

Germline: Cell lines in gametes which are heritable to the next generation.

Germline mosaicism (Human): Presence of two or more genetically different germlines in an individual, resulting from somatic mutation during cellular proliferation and differentiation.

Haplotype: A specific variant of mtDNA that is associated with the phylogenetic origins of maternal lineages.

Hemizygous: A term for the genotype of an individual with only one representative a chromosome or chromosome segment, rather than the usual two (refers especially, but not exclusively to X-linked genes in the male).

Heteroplasmy: The presence of one or more variant(s) in mitochondrial DNA resulting in more than one type of mitochondrial DNA that may differ in percentage by a cell type in an individual.

Heterozygote (heterozygous): An individual or genotype with two different alleles, one of which is wild-type or "normal" allele, at a given locus among a pair of homologous chromosomes.

Homoplasmy: The presence of only one type of mitochondrial DNA in the mitochondria of an individual.

Homozygosity by decent (HBD): A phenomenon in which an individual inherits two identical alleles from two individuals (parents) who are from a common ancestor.

Homozygote (homozygous): An individual or genotype with identical alleles of a given locus on a pair of homologous chromosomes.

Human Genome Project: A major research project, international in scope, which took place from 1990 to 2003 and resulted in the mapping, sequencing, and assembly of a representative human genome, extended to the genomes of many model organisms.

Imprinting center (IC): A regulatory region in the germline that acts as a master cis-regulatory element to local regions of imprinted genes, also known as imprinting control region.

Incompletely dominant (semi dominant): A trait that is inherited in a dominant manner but is more severe in a homozygote than in a heterozygote.

Indel: Abbreviation for "insertion/deletion" a small structural variant defined by the presence or absence of a segment of DNA, ranging from one base to a few hundred nucleotide base pairs.

In-frame deletion, duplication or insertion: A deletion, duplication, or insertion that does not destroy the normal reading frame of the gene (triplet codon).

Insertion: A structural variation in which a DNA segment from another locus, chromosome, or an exogenous source (nucleotides) such as a retrovirus, is inserted or integrated into the genome.

Intergenic DNA (region): The mostly un-transcribed DNA or nucleotides between genes that make up a large portion of the total DNA in the genome.

Intron: A segment of a gene that is initially transcribed but then removed from the primary RNA transcript by splicing and joining the sequences (exons) on either side of it.

Insertion: A type of mutation in which an addition of one or more nucleotides base pairs or a segment of DNA occurs and the length of DNA sequence or of chromosome is altered (changing in RNA reading frame).

Inversion: A balanced chromosomal rearrangement in which a segment of a chromosome is reversed end to end. If the centromere is included in the inversion, pericentric inversion.

Karyotype: The chromosomal constitution or a display of all chromosomes in an individual.

Linkage analysis: a statistical method to determine whether two or more loci are assorted independently or are transmitted together during meiosis because of the proximity on the chromosome.

Linkage disequilibrium (LC): The occurrence of combination of allele in coupling phase at two or more linked loci (haplotypes), more frequently than expected from the frequency of the alleles in the population.

Locus: The physical location occupied by a gene on a chromosome.

Locus heterogeneity: Variants in different genes or loci that result in the same phenotype or condition.

Loss of function mutation or variant: A change in DNA associated with a reduction or a complete loss of one or more of the normal functions of a protein.

Loss of heterozygosity (LOH): A type of variant that results in the loss of one copy of a segment of DNA, containing a gene or genes.

Maternal inheritance: The transmission of genetic information through the mother (mitochondrial inheritance).

Meiosis: Cell division occurring in the diploid germ cells, yielding gametes (ova and sperms) containing the haploid chromosome number.

Meiosis I error: Nondisjunction is the most common error observed in meiosis II in which two homologous chromosomes fail to separate during anaphase I, resulting in an abnormal number of chromosomes in daughter cells.

Meiosis II error: Nondisjunction is the most common error observed in meiosis II in which sister chromatids fail to separate, resulting in one gamete with an extra chromosome, one gamete without chromosome, and two normal gametes.

Messenger RNA (mRNA): An RNA transcribed from the DNA of a gene that determines or directs the sequence of amino acids of the encoded polypeptide.

Methylation: An epigenetic mechanism that regulates gene expression and tissue differentiation. It is also important for other biological functions, such as heterochromatin formation and transcriptional regulation. Methylation typically results in turning off of a gene.

Microdeletion: A chromosomal deletion which is too small to be detectable using conventional cytogenetic method.

Minor allele frequency: The frequency of the second most common allele (less common allele) at a locus used in population genetics to differentiate between common and rare variants.

Missense mutation or variant: A mutation or variant in nucleotide that changes a codon to another codon for another amino acid or results in an amino acid change, resulting in polypeptide sequence alteration.

Mitochondrial DNA (mtDNA): The DNA in the circular chromosome of the mitochondria, and in humans, containing 37 genes and multiple copies per cell. MtDNA is maternally inherited and evolves 5 to 10 times as rapidly as genomic DNA (nuclear DNA).

Mitosis: The process of typical cell division, resulting in two cells genetically identical to the parent cell.

Monosomy: A chromosome constitution in which one of a chromosome pair is missing, such as seen in 45, X, Turner syndrome.

Mosaicism, human: A condition in which an individual has two of more genetically distinct cell lines or within tissues that were derived from a single zygote, due to a post-zygotic mutation.

Multiplex ligation-dependent probe amplification (MLPA): A polymerase chain reaction (PCR)-based laboratory technique used to detect copy number variants (CNVs) in multiple samples (~120-500 nucleotides length).

Mutation: Genetic change that gives rise to a variation DNA sequence or the process by which genetic change occurs.

Next-generation sequencing (NGS): A high-throughput massively parallel or simultaneous sequencing technique.

Nonconservative substitution (mutation): A replacement of amino acid with another amino acid with dissimilar properties.

Nondisjunction: The failure of two members of a chromosome pair to disjoin during meiosis I, or of paired chromatids to disjoin during meiosis II or mitosis, resulting in one daughter cell with both and the other with neither or none.

Nonsense-mediated mRNA decay (NMD): A cellular mechanism to prevent translation into truncated proteins by recognizing and degrading mRNAs that carry premature termination (nonsense) codons.

Nonsense mutation or variant: A single-base substitution in DNA resulting in a termination codon.

Nonsynonymous: A single nucleotide variant (SNV) that alters a codon resulting in a replacement of amino acid with another amino acid and amino acid sequence alteration (a change in the peptide sequence).

Novel property mutation or variant: A genetic change that confers a new property on the protein. Also called gain-of-function or neomorphic mutation or variant.

Nulle allele: An allele that results either in total absence of the gene product or total loss of function of the product.

Obligate heterozygotes: An individual who may be phenotypically unaffected, but on the basis of pedigree analysis, must carry a specific variant.

Out-of-frame deletion or duplication or insertion: A deletion, duplication, or insertion that shifts the reading frame of the gene or the change is not in a multiple of triplet codon, resulting in a change in polypeptide or protein sequence.

Partial aneusomy: Sub-chromosomal variant leading to loss of one copy of a segment of a chromosome (partial monosomy) or to gain of a third, extra copy of a segment of a chromosome (partial trisomy).

Pathogenic variant: A genetic variant that causes a disease, either alone (for dominant or X-linked disorders) or in combination with another pathogenic variant on the other allele (for recessive disorders).

Penetrance: The fraction of individuals who have any type of signs or symptoms of the trait, carrying with a variant that is known to cause the trait (all or none phenomenon).

Phenocopy: A mimic of a genetically determined phenotype, caused instead by an environmental or nongenetic factor.

Phenotype: The observed biochemical, physiological, and morphological characteristics of an individual as determined by a genotype and the environment in which it is expressed. Also, the constellation of features that can be recognized as a disease, disorder, or complex trait.

Phenotypic threshold effect: Primarily applies to mtDNA, a threshold proportion (%) of heteroplasmy for a given variant of mitochondrial DNA or gene when phenotypic expression or disease manifests.

Pleiotropic: A single gene variant or allele causing multiple seemingly unrelated phenotypic features (usually in different organ systems).

Point mutation: Single nucleotide variant or one nucleotide alteration.

Polygenic: A trait or condition thought to be caused by a culminative effects of many genes or alleles, typically, each effect is small or negligible. When epigenic or environmental features are recognized, it is called multifactorial or polygenic.

Polygenic (risk) score (PRS): A calculated number that represents an estimated effect of many genetic variants on a phenotype or trait. The score, which is typically calculated as a weighted sum of trait-associated alleles, is usually comprised of common variants but may also include rare variants.

Polymorphism: The occurrence of two or more alternative alleles in a population, each of which at a frequency greater than that could be maintained by recurrent mutation alone. A locus is arbitrarily considered to be polymorphic if a rarer allele (minor allele) has a frequency of at least 0.001 so that the heterozygote frequency is at least 0.02. By convention, any allele with frequency less than 0.001 is considered a rare variant. Despite common usage of polymorphism to mean an allele at a polymorphic locus, or (mistakenly) to benign changes, the term “common variant”, “single nucleotide variant”, and “nonpathogenic variant” are more accurate.

Population genetics: The study of genetic variants in populations and of how their frequencies change over time in response to forces such as mutation, selection, genetic drift, and gene flow.

Premutation: In disorders associated with unstable repeats (e.g., fragile X syndrome, Huntington disease), a moderate expansion of the number of repeats, with increased risk for further expansion during meiosis. Disorders may have distinctive names for

premutation disorders such as X-associated tremor/ataxia syndrome (FXTAS) or premature ovarian insufficiency (FXPOI).

Proband: The affected family member through whom the family is ascertained, also called propositus or index case.

Qualitative trait: A descriptive property-related trait in in nature.

Quantitative trait: A measurable or countable trait.

Rearrangement: Chromosome breakage followed by reconstitution into an abnormal combination. The outcome may be balanced or unbalanced, depending on whether any genetic material is gained or lost.

Recessive: A trait only expressed in homozygous, compound heterozygous, or hemizygous state of variants or both alleles are affected or biallelic.

Recombination: A process by which pieces of DNA within chromosome are broken and recombined to produce new combinations of alleles which creates genetic diversity.

RNA editing: Posttranscriptional modification of any type of RNA transcript, including base changes and other modifications that can affect function or stability of the molecule. Also called RNA epigenetics. Associated with functions such as neuronal regulation and immune defense.

Sanger sequencing: Developed in the laboratory of Fred Sanger in 1977, which remains as the gold standard for accuracy of DNA sequence determination. However, it is impractical for large-scale sequencing, but it is still used for some targeted sequencing or confirmation analyses. Also known as the chain termination method.

Segmental aneusomy: A portion or locus with abnormal number of chromosomes (non-diploid).

Segmental chromosomal syndrome: A disorder caused by a part or segment of chromosomal alteration (either by deletion, duplication, inversion, and insertion)

Sex chromosome: A chromosome that determine individual's genetic sex, X or Y (XY for male and XX for female).

Single nucleotide variant (SNV): A single base pair alteration in DNA or nucleotide sequence.

Single nucleotide polymorphism (SNP): The term single nucleotide polymorphism (SNP) is used widely to imply a relatively common variant (present in >1%), and although entrenched, more precise terminology such as the use of common SNV is encouraged instead.

Single-gene disorder: A disorder due to one or a pair of pathogenic alleles at a single locus. Also called monogenic disorder or Mendelian disorder.

SNP array: A type of microarray or DNA array that uses oligonucleotides corresponding to high frequency genomic single nucleotide variants from polymorphic loci to detect a chromosomal or sub-chromosomal deletion or duplication (i.e., CNV). Alternative to comparative genomic hybridization (CGH), SNP array is used for GWAS and LOH studies.

Somatic cell: Any cell of the body, excluding germline cells or non-germline cells.

Structural rearrangement or variant: A change in structure of one or more chromosomes. The alteration may be balanced if there is no change in genomic content (e.g., balanced translocation or inversion), or unbalanced, if genomic content is changed (e.g., duplication or deletion).

Synonymous mutation (Silent mutation): A single nucleotide variant (SNV) that does not alter the amino acid of the encoded peptide because of redundancy (degeneracy) of triplet codons for proteins so there may be no functional consequence of the variant.

Termination codon: One of the three codons (UAG, UAA, and UGA) that terminate synthesis of amino acids, or a polypeptide. Also called a stop codon or chain-termination codon.

Translocation: The transfer of a segment of one chromosome to another locus of the same chromosome or to a homologous chromosome or to a nonhomologous chromosome. If two nonhomologous chromosomes exchange pieces, the translocation is reciprocal.

Triplet codon: A trinucleotide sequence of DNA or RNA that corresponds to a specific amino acid.

Trisomy: The state of having three representatives of a given chromosome instead of the usual pair (2), such as trisomy 21 (Down syndrome).

Unbalanced (skewed) X inactivation: Substantial deviation from the expected equal distribution of inactivation between the two X chromosomes in female.

Uniparental disomy: The presence of two copies of a given chromosome (or part thereof), both inherited from one parent, with no copy from the other parent. When both homologues of the parental pair are present, the situation is heterodisomy. When one parental homologue is present in duplicate, the situation is isodisomy. Each parent has a pair or two of each type of chromosomes (1-22 autosomes and XY or XX).

Unstable repeat expansion: A diverse group of disorders that are caused by the expansion of trinucleotide, tetranucleotide, or pentanucleotide repeat sequences which are susceptible to expand in subsequent generations (dynamic variant).

Untranslated region (UTR): The segment of a gene and corresponding mRNA either precedes the initiator codon (5'-UTR) or follows the stop codon (3'-UTR) which will not be translated into protein products.

Variant: An allele that differs from wild type or "normal" type. Referring to an altered allele.

Variant of uncertain significance (VUS): A genomic sequence variant detected whose pathogenic significance is currently unknown. This consequence of relatively untargeted screening by sequencing or microarrays has created challenges for diagnosis – particularly prenatal, but it is not a static designation. Further knowledge and experience can move the designation in the direction of either "benign" or "pathogenic"

Whole exome sequencing (WES): Also known as exome sequencing (ES). Use of high-throughput methods to sequence all the exons of protein-coding genes (the exome), which comprise about 1.5% of the genome.

Whole genome sequencing (WGS): Use of high-throughput method to determine the sequence of an individual's entire genome (minus the few percent that current technologies are not capable of sequencing).

X-inactivation, (Lyonization): Inactivation of genes on one X chromosome in somatic cells of female, generally occurring randomly, early in embryonic life.

X-linkage: The presence of alleles of or a variant on the X chromosome and/or the characteristic inheritance patterns associated with X chromosome. Gene on the X chromosome, or traits determined by such genes are X-linked.

Y-linkage: Genes on the Y chromosome, or traits (e.g., the male sex) determined by such genes, are Y-linked.

Zygosity: The number of zygotes from which a multiple birth is derived. Monozygotic vs dizygotic. Also, the status of an allele in a gene or region, homozygous, heterozygous, hemizygous.

Zygote: The first diploid cell resulting from fertilization between two gametes or a fertilized ovum (ovum and sperm).