INDIVIDUAL GENETIC HETEROGENEITY

Mauno Vihinen[1]

[1]Lunds Universitet

June 24, 2022

Abstract

Genetic variation is described from the perspective of an individual. All the different levels of genetic information and variation are covered, ranging from whether an organism is unmixed or hybrid, has variations in genome, chromosomes and more locally in DNA regions, epigenetic variants or alterations in selfish genetic elements. Genetic constitution and heterogeneity of microbiota are highly relevant for health and wellbeing of an individual. Mutation rates vary widely for variation types e.g. due to sequence context. Genetic information guides numerous aspects in organisms. Types of inheritance, whether Mendelian or non-Mendelian, zygosity, sexual reproduction and sex determination are covered. Functions of DNA and functional effects of variations are introduced, along with mechanism that reduce and modulate functional effects, including TARAR countermeasures and intraindividual genetic conflict. TARAR countermeasures for tolerance, avoidance, repair, attenuation and resistance are essential for life, integrity of genetic information and gene expression. The genetic composition, effects of variations and their expression are considered also in diseases and personalized medicine. The text synthesizes knowledge and insight on individual genetic heterogeneity and organizes and systematizes the central concepts.

INTRODUCTION

Genetic variation is pervasive, all individuals within a species differ somewhat in their genomes, and the genomes of individual cells in a multicellular organism contain differences. As an example, the genomes of humans are 99.9% identical (Shastry, 2002). There are some 3 million single nucleotide variations (SNVs) per haploid genome (about 1 variant per 1 000 bp (Brookes, 1999)), and numerous other types of alterations in comparison to reference genome. Genome in an organism is dynamic, somatic alterations are numerous and accumulate during life time. The genomic constitution of each case of cancer is unique (Burrell et al., 2013) and the number of variants in a cancer patient can exceed 1 million (Alexandrov et al., 2013). Variations are generated by several mechanisms. Evolution of sequences and species is based on genetic variation. Many variants are tolerated and do not have a phenotype, effect on fitness or other consequences. Harmful variants cause genetic diseases. Beneficial variants can be enriched by selection during evolution.

Genetic variation is one form of pervasive and normal biological heterogeneity called poikilosis (Vihinen, 2020a). There is a balance between variation generation and repair. In a normal situation, efficient DNA repair systems correct almost all lesions; however, it would be extremely costly to detect and correct all changes, thus somatic variations accumulate during the lifespan. Several mechanisms reduce and attenuate effects of variations. The last option for cells with excessive or very severe variation(s) is programmed cell death.

The goal of this survey is to provide a comprehensive view on individual genetic heterogeneity, its origin, implementation of genetic information, functional and phenotypic effects and constraints of functional effects. Population effects and phenomena are excluded.

Genetic variations can be described and grouped in many ways depending on the perspective and type of variation: type of nucleic acid; size or effect of variation; disease relevance; affected molecular level; variation generating mechanism; frequency; harmfulness (pathogenic/neutral/unknown); neutral, risk, and protective

alleles; type of inheritance; chromosome type (sex chromosomes vs autosomes); nuclear and organellar genetic variation; origin of variation (inherited/*de novo*/somatic/epigenetic); location in coding or noncoding DNA; conflict between different types of genetic elements and variations in them, etc.

Genetic information and material are in cells in the form of large oligonucleotides, DNA or RNA chains, as genes, pseudogenes and other expressed segments. Variation types include those in hybrids, genome, chromosomes and more locally in DNA regions, epigenetic variants and selfish genetic elements. All higher organisms form an ecosystem with microorganisms called holobiont. The genetic constitution and heterogeneity of microbiota is of great importance in health and in many diseases. Mutation rates vary widely for variation types dependent e.g. on the sequence context. Somatic variation is ubiquitous in multicellular organisms. Related to the implementation of genetic information in organisms, types of inheritance, both monogenic and non-Mendelian inheritance, zygosity, sexual reproduction and sex determination are covered. Functions of nucleic acid polymers and functional effects of variations are introduced, along with constraints of functional effects including TARAR countermeasures, individual genetic conflict and effects of microbiota. TARAR countermeasures affect the phenotype by processes in tolerance, avoidance, repair, attenuation and resistance. Intraindividual genetic conflict adds further layers to phenotypic expression. Thus, the genetic composition, effects of variations and their expression form a very complex system that is unique for each individual. Here, the topics mentioned above are also considered from the perspective of diseases.

Variations at population level are excluded. Population level concepts, such as mode of reproduction, types of fertilization, pregnant sex, type of breeder, mate choice, mating systems and interindividual genetic conflicts are not covered in here. Mechanisms that produce variations are also excluded apart from some notes, for recent reviews see (Chatterjee and Walker, 2017; Strauss, 2016). The goal of the presentation is to synthesize knowledge and insight in individual genetic heterogeneity and to organize and systematize the central concepts.

GENES AND GENOME

Genetic material is in most organisms in the form of long DNA strands. Exceptions are some viruses which have RNA genome (Payne, 2017). RNA viruses can be classified as those that possess double-stranded genome, and those with either positive or negative-sense single-stranded RNA genomes. There has been a long debate about RNA world and whether RNA or DNA was the original genetic material in the first living organisms (Higgs and Lehman, 2015). Genomes contain genes that code for gene products and other regions, which may function e.g. in regulation, for more detailed discussion see Functional effects of DNA variations.

Genomes of individuals in a species differ in many places. The length of the haploid human XY reference genome is 3,117,275,501 bp (T2T-CHM13v2.0 https://www.ncbi.nlm.nih.gov/assembly/GCF_-009914755.1/#/st Accessed June 14, 2022). The full complement of genes in a clade is called the pangenome. Individuals share the core genome but have variable numbers of dispensable genes. The size of the dispensable genome varies, for example, in soybean it accounts 20% and in rice 43% of the size (Ma et al., 2019). The products of many dispensable genes respond to stresses and various stimuli and are depleted in several cultivated barley strains (Ma et al., 2019). The human African pangenome is about 10% larger than the reference genome (Sherman et al., 2019). The other human pangenomes for several populations include smaller amounts of novel sequences, reviewed in (Sherman and Salzberg, 2020). 315 of the African insertions are within protein-coding genes and can lead to alternative forms of gene products.

The concept of a gene is instrumental for genetics and genetic variations. Many definitions have been provided, see e.g. (Gerstein et al., 2007; Portin and Wilkins, 2017; Waters, 2013) and references therein; however, none of them describes the current concept of gene. Here, *gene is defined as i) a continuous stretch of polynucleotide sequence, ii) consecutive segments, or iii) connected segments joined together to code for a functional gene product or products.* Essential for the gene function is that the nucleic acid sequence is transcribed. Gene products are either RNA or protein molecules.

The functions of RNA and protein molecules have been defined in Variation Ontology, VariO (Vihinen, 2014b). The RNA functions are amino acid transfer, catalytic activity, regulation, information transfer,

splicing function and translation. Protein functions comprise catalysis, information transfer, movement, recognition, storage, structural functions and transport. Together these functions form the basis for life and all cellular and organismal events, reactions and responses. The functions of DNA are described below.

Genetic material is either DNA or RNA. A single gene can code several gene products, which further can be modified in various ways. Proteoforms are protein forms coded by a single gene (Smith and Kelleher, 2013). Numerous mechanisms can generate different forms of the encoded protein, comprising genetic variants (substitutions, deletions, insertions, indels), transcription and translation errors, mRNA processing (alternative splicing, alternative start and stop codons, (programmed) ribosomal frameshifting), post translational modifications such as proteolytic processing and amino acid modifications, as well as protein splicing (Vihinen, 2021a). A genetic region may contain several overlapping genes. Genes in prokaryotes and certain genes in eukaryotes are intronless and thus continuous. Introns divide genes into segments in eukaryotes. Genes for immunological recognition molecules, immunoglobulins, and B and T cell receptors, are joined in VDJ recombination from independent segments. The location of a gene does not have to be fixed. Transposable elements can move within the genome and code for gene products.

The gene coding portion of a genome varies largely. Many viruses have so compact genomes that some genes overlap, usually in different reading frames. Overlap appears also in organisms with intergenic material, however, is quite rare. In human, the exons of protein coding genes account 2.94% and protein coding exons cover 1.22% of the genome (Consortium, 2012). Genes can be continuous or be split into exons and introns. When introns, promoters and poly(A) sites are also included, the protein coding genes cover 39.54% of human genome (Consortium, 2012). Despite the small portion of coding region of genes, the entire or almost entire genome is transcribed (Djebali et al., 2012). ENCODE project has charted the human noncoding regions (Moore et al., 2020), functions of which are still largely unknown.

Pseudogenes are copies of genes that are not functional or not at least in the same way as the original gene. They have been identified in all types of organisms ranging from bacteria to mammals. The definition in VariO is *nucleic acid segment that resembles a functional gene, but which is defective* . Pseudogenes originate with three mechanisms (Cheetham et al., 2020). Duplicated pseudogenes are produced via tandem duplication or uneven crossing over and processed pseudogenes arise by retrotransposition, when an mRNA sequence or part of it is reverse transcribed to DNA and inserted into the genome (Vihinen, 2014a). The human genome contains >14 700 duplicated or retroposed pseudogenes according to the GENCODE database (https://www.gencodegenes.org/human/stats.html Accessed June 14, 2022). The third form of pseudogenes, unitary pseudogene, is extremely rare. It it based on loosing gene function without a functional copy. Polymorphic pseudogenes are classified as pseudogenes in the reference genome for a species, but can be functional in some individuals (Cheetham et al., 2020).

A duplicate of a gene quickly accumulates variants that destroys its coding region, introduce stop codons etc. since there is no evolutionary pressure to keep pseudogenes active. However, some pseudogenes are expressed, also in a tissue-specific manner (Ma et al., 2021). Despite nonfunctional sequences, the accumulation of variations does not occur at random rate in some pseudogenes and display evolutionary selection. Pseudogenes have functions e.g. as regulatory elements at DNA, RNA and protein level, they provide a reservoir of genetic material for increased genetic diversity, some pseudogenes are translated (Cheetham et al., 2020; Vihinen, 2014a).

TYPES OF VARIATIONS

A very large number of variation types has been described at different levels. VariO facilitates systematic description of variations, their types, effects, consequences and mechanisms (Vihinen, 2014b). The ontology has been updated and it organizes systematically the concepts described in this text. VariO-based detailed descriptions of variation types have been presented previously for DNA (Vihinen, 2018), RNA (Vihinen, 2021b) and protein variants (Vihinen, 2015b) along with examples, often in relation to human diseases.

In addition to changes to DNA sequence, epigenetic variants that do not change the actual sequence are also common. A further layer of variation is introduced by selfish genetic elements that can multiply themselves.

Further, changes in microbiota, microorganisms found in and on all multicellular organisms, can be substantial and related also to diseases. Here, a summary is presented to highlight the multiple types and effects of different kinds of variants.

Hybrid

Organisms reproduce sexually or asexually. Sexual reproduction, which is discussed below, is important for the genetic composition of individuals and for the generation of new genetic combinations. Hybrids originate in sexual reproduction from combining two organisms of different breeds, varieties, species, genera or families. Hybrids are rather common, although many species have reproductive barriers that prevent hybrid offspring production by genetic, physiological and behavioral processes.

Many hybrids are sterile (Li et al., 2020), e.g. because of different number of chromosomes in comparison to parent species. In a hybrid, egg and sperm cells of different species are combined, therefore all the cells contain a mixture of genomes from both parental species. Intra-specific hybrids are produced by different subspecies of a species. Interspecific hybrids are hybrids between two different species within the same genus. Human (*Homo sapiens)* and Neanderthal hybrids were either intraspecific or interspecific. There has been a long debate whether Neanderthals were a separate species or a subspecies (*Homo neanderthalensis* or *Homo sapiens neanderthalensis*).

Intergeneric hybrid combines two genera within a family. Genera that produce intergeneric hybrids are always genetically related members of the same taxonomic tribe. An example is hybrid between gelada (*Theropithecus gelada*) and common baboon (*Papio hamadryas*), which interbreed occasionally in the wild (Jolly et al., 1997). Interfamilial hybrids are very rare; however, some instances are known, e.g. in kelps, brown algae seaweeds (Murúa et al., 2020). Hybridization has been used for development of novel species, especially in plant breeding to provide genetic variation and to improve properties (Alvarez and Guzmán, 2018).

Genome wide variations

Euploidy means change in the number of complete genome copies. Normal human and animal genomes are diploid i.e. contain two copies of all chromosomes (Figure 1A). Genome duplications are rather common; however, much higher forms of ploidy have been detected. Many organisms, including human, have underwent genome duplication during evolution (Singh et al., 2015). The resulting duplicated genes are called ohnologs.

The majority of human cells are anuclear and thus nulliploid (Sender et al., 2016). Higher ploidies in human appear in specialized cells, such as in liver, heart, bone marrow and placenta. In human liver about 50% of cells have more than two sets of chromosomes (Duncan, 2013). In plants, endopolyploidy levels up to 24 576C have been identified (Barow, 2006). 1C means haploid genome. In insects, polyploidy ranges from 4C to 2 048C, the latter detected in salivary glands in *Drosophila* (Ren et al., 2020). Polyploidy has been linked to high metabolic or synthetic activity, compensation for lack of nuclear DNA, response to environmental conditions and to relationship between cell volume and DNA content (Barow, 2006). Polyploidy reversal can reduce the degree of polyploidy by reductive mitosis, nuclear budding or nucleophagy (Wu et al., 2022).

There are two forms of polyploidy. In autopolyploidy, the chromosomes are from one taxon (Parisod et al., 2010), whereas in allopolyploidy they originate at least from two taxa. Mixoploid means an organism with different numbers of genome sets in different cells. This is very common in multicellular organisms.

Autoploidy and allopolyploidy can be further divided into categories (Tate et al., 2005). Strict autopolyploidy and interracial autopolyploidy are the two forms of autopolyploidy. Allopolyploidy can be subdivided as segmental, true, also called genomic allopolyploidy, and autoallopolyploidy, which is also a form of autopolyploidy.

Anuclear cells lacking DNA, like erythrocytes and platelets, are the most common in human body comprising about 90% of the total number of cells (Sender et al., 2016). These cells do not contain nucleus and are lacking

DNA. These cell types are highly differentiated and do not need DNA for their function, for example, erythrocytes are specialized oxygen transporters.

Chromosomal variations

Chromosomal variations can be classified to two major categories (Vihinen, 2018). Differences in the numbers of chromosomes are called aneuploidy. The other variant category includes those that change the chromosome structure. For detailed description and examples, see (Vihinen, 2018).

Aneuploidy is the most common cause of miscarriages and congenital birth defects. 20% of human oocytes are estimated to be aneuploid (Hassold and Hunt, 2001). Trisomy is an example of aneuploidy (Figure 1B). Down syndrome patients have a triplication of chromosome 21 (Antonarakis et al., 2020). The copy numbers of chromosomes vary in aneuploidy. In nullisomy, a certain chromosome is missing. The aneuploidy alterations range in size up to polysomy. Disomy, that is the normal state in many organisms, may still contain structural chromosomal variations. In heterodisomy, a pair of non-identical chromosomes is inherited from one parent, whereas in isodisomy a single chromosome has been duplicated from one parent. Uniparental disomy is a condition where both copies of a chromosome or a part of it are from one parent.

Numerous types of structural chromosomal alterations are known. Copy number variation, DNA mobile genetic element insertion and nucleotide expansion are different forms of chromosomal amplification (Figure 1C). These variants can be interspersed, i.e. scattered in genome or be in tandem, when following each other. DNA transposons, insertion sequences and retrotransposon insertions (LINE, LTR and SINE segments, etc.) are mobile genetic elements that can move and duplicate within chromosomes. Nucleotide expansions have been classified as microsatellites (mononucleotide to hexanucleotide expansion) and minisatellites of up to few tens of nucleotides long repeats. Chromosomal deletions (Figure 1D) are either structural copy number variations, interstitial deletions in the middle of the sequence or terminal deletions in the end of the chromosome arm.

Chromosomal translocations are either intra- or interchromosomal (Figure 1E). A dicentric translocation is formed when two chromosome segments fuse together and when they both contain a centromere. In reciprocal chromosomal translocation, genetic material is switched between homologous chromosomes. The third type is Robertsonian translocation, where the long arms of two chromosomes are fused. The p arms are in this case very short in both the original chromosomes. In intrachromosomal translocation, a chromosomal region is translocated within a chromosome. Inversion is a special case where the translocated segment is located back to original position but in reverse direction. If inversion appears within a chromosome arm, it is called paracentric. A pericentric inversion contains also the centromere.

Complex chromosomal variation is the outcome of at least four breakpoints (Figure 1F). Kataegis is a form of extremely complex chromosomal variation due to shattering of chromosome(s) into several pieces after which they are joined together. Immunological recognition molecule diversification of antibodies and B and T cell receptors (Figure 1G), is a complicated and special process that contains immunological receptor gene rearrangement, somatic hypermutation, gene conversion and class switch recombination.

Isochromosome is a variation that lacks one arm while the other has been duplicated (Figure 1H). Ring chromosome has a circular structure (Figure 1I). Telomere length change is also a chromosomal structural variation (Figure 1J). It can mean either lengthening or shortening of chromosome ends, telomeres. Telomere healing means addition of a telomere to chromosome that has double-stranded break and deletion of a terminal segment.

DNA and RNA chain variations

DNA chain variations (Figure 1K) are local alterations and include DNA deletion, insertion, indel, substitution and translocation (Vihinen, 2018). These are among the most common variants within genes. Indels are variants that contain both an insertion and deletion (Vihinen, 2015a). Substitutions are either transitions between similar types of nucleotides (purine to purine or pyrimidine to pyrimidine), or transversion where a purine (A, G) is changed to a pyrimidine (C, T) or *vice versa* . DNA translocations contain also inversions.

5

Variations in RNA (Figures 1L and 1M) comprise RNA deletions, insertions, indels, substitutions, translocations including inversions, as well as missing RNA, and changes in initiation and termination codon (Vihinen, 2021b). In the coding regions, insertions, indels and deletions are either in-frame or out-of-frame, indicating whether the amino acid code is retained or not. Missing protein is quite a common variant type. RNA quality control mechanisms, such as nonsense-mediated decay (NMD), destroy transcripts that contain premature stop codons due to substitutions, insertions, deletions or indels.

Substitutions can be divided into several categories. Missense variants change the coded amino acid while synonymous variants do not affect the coded protein sequence. Nonsense variants introduce premature stop codons and such transcripts are often destroyed unless the variant is towards the end of the transcript, no more than 50 bp from the end of penultimate exon (Kurosaki et al., 2019), or do not escape NMD due to various mechanisms (Lindeboom et al., 2019). Unsense variants (Vihinen, submitted) have largely been misclassified as synonymous although they change the mRNA sequence due to regulatory alterations to splicing or expression regulation, or due to aberrant splicing. Thus, these variants are not synonymous, they often lead to mRNA degradation and missing RNA, thus causing missing protein, as well.

Figure 1M shows effects of variations on mRNA splicing. Transcripts can be alternatively initiated, terminated or polyadenylated. These variants are common in many transcripts. Bases in mRNA can be modified. Alternative splicing is a very common phenomenon. Almost all human genes have alternatively spliced isoforms (Hu et al., 2015; Pan et al., 2008), but their biological significance is still largely elusive (Tress et al., 2017). *Trans* spicing joins two different transcripts to form an RNA chimera.

Cis splicing within the transcript is the normal mode of splicing. Constitutively spliced exons are spliced in the order they appear in mRNA. In exon skipping, splicing jumps over an exon. Mutually exclusive exons are a special form of exon skipping. In such a case, only one of a pair of exons appears in the final transcript. Loss of exon fragment means a case where a part of exon is spliced out. Inclusions can contain a fragment of intron or entire intron, called intron retention. Variation in the intron region can cause inclusion of a cryptic exon.

DNA variants can affect either exon, intron or intergenic DNA. Cells contain organelles, some of which contain DNA. Organellar DNA variants include those of mitochondrial or plastid DNA. Other extrachromosomal variation forms are those in plasmids and extrachromosomal circular or linear DNA or RNA.

Protein variations

Many variations in proteins are due to DNA and/or RNA alterations, therefore they are briefly introduced in here, see Figure 1N. Amino acid substitutions, insertions, deletions, indels and missing protein are the different forms of genetically caused amino acid changes (Vihinen, 2015b). Insertions and indels are either sequence retaining or amphigoric, when the end of the sequence is coded by a frameshifted RNA. Deletions are either truncations or sequence retaining. Missing protein is a frequent consequence e.g. because of a premature stop codon, out-of-frame alterations, aberrant splicing etc. Single gene encoded proteoforms can be altered at DNA, RNA and protein levels, see chapter Genes and genome.

Epigenetic variation

Epigenetics means heritable changes that do not change the DNA sequence, and which affect gene expression and gene silencing. These changes can even be transferred by inter- or transgenerational epigenetic inheritance across several generations (Bošković and Rando, 2018). Epigenetic inheritance can mean also transmission of the epigenetic changes to daughter cells. Epigenetic markers may originate from internal regulatory processes or as responses to the environment of the organism.

Epigenetic processes and mechanisms vary at different levels from DNA and protein up to organism (Figure 2). Epigenetics is based on modification of DNA or histone proteins in nucleosomes (Kumar et al., 2018). Although all types of DNA bases are modified in cells, epigenetic changes are known only in cytosine and adenine. They are incorporated by special enzymes: 5-methyl cytosine is the best-known example of epigenetic modifications (Figure 1O). It can be further enzymatically processed to 5-hydroxymethyl cytosine.

6

Methylation of cytosines is especially common in CpG dinucleotides, but also in triplets of these bases. As many of these segments have cytosine in both the DNA strands, methylation generates a strong signal. The methylation of adenine to N^6-methyladenine is the other epigenetic alteration. In addition to these base changes, epigenetic alteration can occur also in the sugar moiety of DNA. Phosphorothioate modifications have been identified in bacterial restriction-modification system and may be more widely spread (Tong et al., 2018).

Protein post translational modifications are common and hundreds of different modifications are known. Histones are proteins that form the core in nucleosomes, protein-DNA complexes that facilitate further compression of DNA. 145 to 147 nucleotides long stretch of double stranded DNA winds around a histone tetramer (Figure 1P). The formed nucleosomes appear like beads in a string. Charged and polar residues in the tails of the histone proteins are heavily modified (https://www.cellsignal.com/learn-and-support/reference-tables/histone-modification-table Accessed June 14, 2022). The epigenetic histone modifications include acetylation, ADP-ribosylation, biotinylation, citrullination, methylation, phosphorylation, serotonylation, sumoylation and ubiquitinylation. Changes in histones can affect chromatin remodeling and access of DNA with effects on transcription and further to gene expression (Zhang et al., 2015).

Protein level epigenetic processes include protein structural inheritance and proteinaceous infection of prions (Vihinen, 2015b). Structural inheritance occurs is centrioles, the main microtubule organizing centers and regulators of cell cycle progression (Wilson, 2008). Asymmetric cell division or centrosome inheritance may be related to cancer (Izumi and Kaneko, 2012). Prion formation occurs in transmissible spongiform encephalopathies in human and other species (Asher and Gregori, 2018). Prion proteins, which have undergone an irreversible structural conformational change, accumulate to amyloid fibrils in brains in the prion diseases. The transformed prion is insoluble and resistant to proteolytic degradation, it also stimulates the conversion of normal proteins. The normal prion has functions and is abundant in brain (Roucou et al., 2004).

Chromatin remodeling means dynamic modifications to chromatin architecture. These changes affect access of condensed genomic DNA to the transcription machinery, and thus controls gene expression. Epigenome editing refers to targeted modification of epigenetic sites. Systems comprising clustered regulatory interspaced short palindromic repeat (CRISPR)-Cas9, transcription activator-like effector nuclease (TALEN) and zinc finger nuclease (ZFN) processes are known (Harrison and Hart, 2018). They appear naturally in some organisms and have been the target of intense research and development for medical and other applications, including gene therapy.

Gene dosage compensation occurs both at cell and organism levels. X-chromosome inactivation (XCI) and genomic imprinting regulate which alleles are active in cells. This is important for the compensation of X-chromosome numbers in females and generally for silencing one allele. XCI turns entire chromosome off while genomic imprinting affects individual genes. Imprinting is maintained throughout the development of the individual or tissue. Human genome contains over 40 imprinted genomic regions (Court et al., 2013) of both protein-coding and non-coding genes. Imprinting can lead to diseases if there is a genetic variation(s) in the expressed allele (Monk et al., 2019). Products of imprinted genes are involved in development, metabolism, growth and behaviour. Imprinting relates to nine human diseases (Mackay and Temple, 2017), including Prader-Willi syndrome due to deletions in the 15q11-13 region and consequent defective *SNRPN* and *NDN* gene products.

XCI, also called lyonization, silences randomly one of the X chromosomes in females and reduces expression level similar to males who have just one X-chromosome (Fang et al., 2019). The dosage compensation of the additional X-chromosome in human females is virtually complete (Johnston et al., 2008).

There are several additional epigenetic processes. Bookmarking means transmission of gene expression patterns during mitosis when gene transcription is silenced (Michieletto et al., 2018). In paramutation, two alleles of a single locus interact leading to heritable change of one allele, induced by the other allele (Chandler, 2007). In transvection, an allele in a chromosome interacts with corresponding allele in the homologous

locus in *trans* (Fukaya and Levine, 2017). Structural (cortical) inheritance refers to transmission by spatial structures, such as the orientation of cilia in ciliates (Beisson, 2008).

Parental effect is a process where, in addition to the genotype and environmental effects, also the environment and phenotype of the parent determines the phenotype of an individual. Maternal effects are due to mRNA or proteins supplied to the egg. Nucleolar dominance appears when ribosomal assembly sites, nucleoli, are inherited only from one parent or progenitor in the case of a hybrid (Pikaard, 2018).

Lightly packed DNA in euchromatin can be expressed, whereas condensed DNA in heterochromatin is silenced. 56.1% of human chromatin is enriched in histone modifications (Consortium, 2012). Translocation or rearrangement of a gene within or near a heterochromatic region in a chromosome causes mosaic gene and protein expression patterns called position effect variegation. This is a common mechanism in plants but appears also in many other organisms. Genes within the heterochromatin are epigenetically silenced. An incontinentia pigmenti-like phenotype appears in patients whose Xq24-qter region is translocated to the 2q34 heterochromatin region (Genesio et al., 2011). The region contains a gene for the inhibitor of nuclear factor kappa B kinase subunit gamma (*IKBKG*), expression of which is silenced by epigenetic DNA and histone alterations (Genesio et al., 2011).

There is lots of variation in epigenetic marks. The changes differ between cells, the marks are written and removed by active systems. Further, genetic variations at the epigenetic modification sites and the genes for the activities for epigenetic changes can affect the epigenetic processes.

Selfish genetic elements

Selfish genetic elements enhance their number or transmission at the expense of other genes. This happen even when the action of a selfish genetic element has negative effect on the organismal fitness. There are three main types of selfish elements with subclasses (Figure 3).

First, over-replicating elements include mobile genetic elements and biased gene converters. Mobile genetic elements comprise transposable elements, plasmids and endogenous viruses. Transposons are capable of multiplication of themselves by copying to different part of the genome. 45.0% of human genome originates from transposons (Nurk et al., 2022). Second, homing endonucleases that insert preferentially into homologous uninserted sites are biased gene converters (Stoddard, 2014). They belong to group I introns and code for nucleases that are self-splicing.

Third, transmission distorters affect the distribution of homologous chromosomes in gametes (Lyttle, 1991; Wedell, 2020). Meiotic drivers manipulate meiotic process and transmission of traits. Driving chromosomes and driving sex chromosomes affect segregation of selfish gene-containing autosomes and sex chromosomes, respectively. Similar mechanism affects also B chromosomes, which are common in many organisms. They are extra chromosomes and not essential for the survival of the individual.

Killer meiotic drivers are ultra-selfish elements that kill meiotic products that do not contain the DNA element. Gamete or spore killers are forms of killer meiotic drivers (van der Gaag et al., 2000). The killers have two main mechanisms, either killer-target system or poison (toxin)-antidote system (Vogan et al., 2022).

The selfish gene in maternal effect dominant embryonic arrest (Medea) is composed of a toxin and an antidote in beetles (Beeman et al., 1992). A mother who carries *Medea* gene expresses the toxin in her germline and thereby kills her progeny. When also the progeny carries*Medea* , they produce the antidote and survive. When a mother has one *Medea* allele and one non-Medea allele, half of the offspring inherit the gene and survive.

Prokaryotic restriction-modification (RM) system can behave also as a selfish genetic element. RM systems are involved in e.g. genetic transformation and infection prevention. Restriction endonuclease degrades DNA unless it is modified by the methylase. RM systems can act in certain situations as mobile genetic elements (Kobayashi, 2001).

Genetic variation in microbiota

All higher organisms can be considered as symbionts as they constantly bear a very large number of microbes, archaea, bacteria, fungi, protists, and/or viruses. It has been estimated that the number of dmicroorganisms in human microbiota outnumbers the human cells by about 1.3 times (Sender et al., 2016). The soma and the microbiota interact in complex ways, they affect each other, and even genetic conflict of resources may appear. The microbiota varies greatly between individuals and even within an individual in different situations and times. Microbiota and consequently microbiome vary during life time and even during seasons depending e.g. on the available food, water and environmental conditions. Microbiome is the collective genome of the microbiota in a system, such as microbiota of a multicellular organism.

Although there is heritable component in the constitution of gut microbiota, the environmental effect is more important (Rothschild et al., 2018). We are only starting to understand the interactions between soma and microbiota and the significance e.g. for the immunity, metabolism, health and wellbeing of the organism (Sanna et al., 2022; Valdes et al., 2018). The constitution of gut microbiota depends on diet, water and other factors. It is possible to control microbiota to certain extent e.g. with probiotics and food fibre (Valdes et al., 2018). Antibiotics and other drugs, pesticides and other chemicals can have profound and long-lasting effects in the constitution of microbiota. The constitution of the microbiota may even contribute to parasite tolerance, as detected in *Caenorhabditis elegans*(Rafaluk-Mohr et al., 2022).

MUTATION RATE

DNA alterations are frequent and comprise lesions, adducts and structural variations. Several DNA repair mechanisms correct these errors and modifications, see TARAR countermeasures. Still some variants are inherited to daughter cells and to progeny. Lesion means an alteration to nucleic acid chain, variation is an alteration that is transmitted further to daughter cells or offspring. The numbers of lesions are elusive and difficult to evaluate due to the very large number of variation types and mechanisms. The total number may well be over 100 000 lesions per day per cell (Ségurel et al., 2014; Yousefzadeh et al., 2021), see Table 1. The number of abasic sites, one type of lesions, has been estimated to be about 30 000 every day in a mammalian cell (Lindahl, 1979; Lindahl and Barnes, 2000). The daily cellular frequencies of several other lesion types range from about 10 for double stranded breaks to 10 000 for single strand breaks (Yousefzadeh et al., 2021), see Table 1.

Mutation rate means the frequency of new variants within certain time, such as during a generation or between cell divisions. It is highly variable at genomic regions, and depends on sequence context, dinucleotides, affected nucleotides and of types of variations. Table 2 shows the mean numbers of autosomal variation sites per genome for variation types based on data from about 2 500 genomes representing 26 populations in the 1000 Genomes Project (Auton et al., 2015). More than 99.9% of human variants are SNVs and short deletions or insertions. However, structural variants affect more bases. For 99.4% of the variants the genotype is heterozygous, the corresponding number for deletions and insertions is 99.0% (Auton et al., 2015).

The sperm mutation rate is substantially higher than that for eggs due to larger number of cell divisions. Human sperm cells have undergone approximately 659 cell divisions in a 40 year-old male while oocytes have undergone just 31 cell divisions (Ségurel et al., 2014).

The average human germline mutation rate is about 1.2×10^{-8} per nucleotide per generation (Kessler et al., 2020; Milholland et al., 2017; Rahbari et al., 2016). The human germline mutation rate is higher than for other investigated multicellular organisms (Lynch, 2010). However, when the number of cell divisions is taken into account, the human mutation rate is the lowest. The spectra of female and male variants are practically identical (Rahbari et al., 2016). The human ratio of paternal to maternal variants is estimated as 3.5 (Rahbari et al., 2016).

Most studies on mutation rate have concentrated on SNVs. In human, the combined rate of insertions and deletions has been estimated to be about 6% of that for substitutions (Lynch, 2010). The number of instances inversely correlates with the insertion or deletion length. There are frequencies also for other types of variants in Table 2.

9

Based on the SNV mutation rate, the number of *de novo* variants is estimated as 76.9 (Rahbari et al., 2016) or 64.4 (Kessler et al., 2020). The number of *de novo* variants increases with the age of father by 2.87 variants per year (Kong et al., 2012; Rahbari et al., 2016). The rate of *de novo* short ([?]20 bp) insertions and deletions is 2.94 and for longer than 20 bp alterations 0.16 variants per generation (Kloosterman et al., 2015). These variants affect on average 4.1 kbp of sequence and 29 bases in coding region per generation.

Mutation rates are significantly higher in unicellular organisms than in multicellular ones. Yeast *Saccharomyces cerevisiae* can grow either in a haploid or diploid form. The mutation rate is substantially higher in the haploid cells, 4.04×10^{-10} vs. 2.89×10^{-10} for SNVs, i.e. about 40% difference (Sharp et al., 2018). Ploidy did not affect the rates of deletions and insertions, 2.03×10^{-11} and 1.63×10^{-11}, respectively.

Human somatic mutation rates vary for different types of cells. Germline cells have the lowest rates, while kidney tubules and appendiceal crypts have tens of times higher rate (Manders et al., 2021). It has been estimated that at the age of over 60, for example human intestinal epithelial cells that divide in every week or two, have generated a variation in every position in the genome (Lynch, 2010). Comparison of somatic mutation rates in 16 mammalian species indicated large differences in rates per year, however, the lifespan somatic mutation burden did not vary more than 3-fold (Cagan et al., 2022).

In addition to the genome, many organisms contain other genetic material. Endosymbiotic organelles contain DNA. The mutation rate of human mitochondria is about 3.6 times higher than germline rate, 4.33×10^{-8} per site per year (Cabrera, 2021). Heteroplasmy, co-existence of different mitochondrial DNA (mtDNA) genotypes, appears likely in all individuals (Stewart and Chinnery, 2021).

Several factors contribute to the mutation rate and include sequence context and the vulnerability of the site for variations, efficiency of DNA repair systems, fidelity of DNA replication, and chromatin structure (Arnheim and Calabrese, 2009; Baer et al., 2007; Hodgkinson and Eyre-Walker, 2011; Makova and Hardison, 2015).

The mutation rate of genomic regions varies widely, 100-fold differences have been reported (Michaelson et al., 2012), in the case of variants in 7-mers, over 400-fold differences are seen in human (Carlson et al., 2018). Y chromosome short tandem repeats (STRs) are widely used for paternity confirmation. The mutation rate of repetitive microsatellites ranges up to 7×10^{-3} per locus per gamete per generation in man (Brinkmann et al., 1998). Mutation rates in these segments vary over 700-fold, calculated from Fergus(s)on Y DNA Project data (http://dna.cfsna.net/HAP/Mutation-Rates.htm). Sequence context is an important factor (Hodgkinson and Eyre-Walker, 2011), there are mutation rate differences even in megabase sequence range (Terekhanova et al., 2017).

Related to the sequence context, certain segments are highly vulnerable for variants. These include R loops, G quadruplets and AT-rich segments (Arnheim and Calabrese, 2009; Nesta et al., 2021; Segurel et al., 2014). Of dinucleotides, CpGs are 10-times more common among variations than any other dinucleotides (Hodgkinson and Eyre-Walker, 2011; Ollila et al., 1996). Transitions are 1.7 times more common than transversions (Lynch, 2010).

The CpG sites have been depleted from genomes during evolution. Instead of the estimated 4% in human genome, there are less than 2% present. Depending on the site, CpG transition rate can be several hundred-folds greater than for non-CpG transversions (Smith et al., 2016).

VARIATION ORIGIN

Variants of genetic origin are either inherited, called germline variation, or somatic (Vihinen, 2014b). Somatic variants emerge in soma, cells of the body. These variants are due to numerous effects and mechanisms. DNA lesions appear in all cells that contain DNA and are largely repaired. However, some variants remain unrepaired and during the lifespan somatic variants accumulate (Alexandrov et al., 2015; Behjati et al., 2014; Vijg and Dong, 2020). Thus, the genomes in cells of an elderly person differ from the genomes the same individual had originally. There are also differences between cells and tissues.

de novo variants either appear in one of the germ cells or early during embryogenesis. The variants are then inherited by all subsequent daughter cells.

Non-genetic origin means either artificial, edited, epigenetic or modified variants (Vihinen, 2014b). Artificial variants are intentionally made. Genome editing with CRISPR-Cas9 system is an example of how edited variants are produced naturally in certain bacteria and archaea (Strich and Chertow, 2019). This mechanism is also widely used for artificial modifications. DNA methylation is the most common form of epigenetic DNA alteration (Weichenhan and Plass, 2013). Modified DNA is an outcome of e.g. several types of lesions, such as chemical adducts.

Inherited variants

Every individual inherits genetic material from parents. This material has undergone processes that have generated unique variations. The chromosomes that the parents have inherited from their parents are stochastically segregated during meiosis when gametes are generated, thereby producing new combinations. During the first meiotic division, homologous chromosomes are independently assorted, and during the second meiotic division, non-identical sister chromatids are independently assorted. Gametes are haploid, i.e. contain just one copy of each chromosome.

Further variations are produced during meiosis by genetic recombination by exchange of genetic material (Figure 4A). Homologous chromosomes form pairs in which DNA segments may be copied from one chromosome and attached to another without changing the donor chromosome. Another mechanism involves breaking and rejoining the DNA chains. Cross over between homologous chromatids leads to exchange of DNA segments. Genetic recombination can occur either during mitosis or meiosis. Mitotic recombination is rarer than meiotic recombination.

Homologous gene conversion is another mechanism that produces further variation by exchange of a homologous sequence between corresponding chromosomes or paralogous segments (Chen et al., 2007). Gene conversions can be of three types: interallelic between alleles, or non-allelic (interallelic, ectopic) gene conversion, either in *cis* or *trans* (Chen et al., 2007). Gene conversion appears predominantly during phases before mitosis whereas mitotic crossing overs mainly take place during interphase. VDJ recombination is a special mechanism for joining genetic fragments to generate genes for immunological recognition molecules (Chi et al., 2020).

Recombination rates vary widely between genetic loci, populations and organisms (Camara et al., 2016; Stapley et al., 2017). The autosome-wide average of recombination rate of Finns and non-Finnish Europeans are estimated to be 2.268+-0.4209 and 2.641+-0.5032 cM/Mb, respectively (Hassan et al., 2021). A 13 bp degenerate sequence motif is essential for recombination in large portion of the hotspot recombination sites and forms a binding site for PR/SET domain 9 (PRDM9), a histone-lysine N-methyltransferase (Baudat et al., 2010; Myers et al., 2010).

Single cell sequencing to investigate meiotic variation in over 30 000 sperm cells from 20 young donors indicated rates from 0.010 to 0.046 aneuploidies per gamete (Bell et al., 2020). There were from 22.2 to 28.17 crossing over events per cell.

Additional variation is produced by lesions at gametes or in the fertilized egg leading to *de novo* variants, variants that occur for the first time in a lineage. Cytoplasmic organelles are randomly distributed in mitotic segregation to daughter cells. Traditionally, organelles have been thought to be inherited in the egg.

Zygosity

Zygosity refers to different concepts. It is commonly used to describe similarity of alleles in a diploid system. Alleles are variants of a gene and can range in size from SNVs to large chromosomal rearrangements. In a homozygous organism, the alleles are similar, while in heterozygotes they differ. Compound heterozygous individual has two different alleles of a gene. When a diploid organism has only one copy of a gene, it is called hemizygous. Nullizygous organism has two loss-of-function null alleles or misses the gene completely.

Zygosity can refer also to allele origin. Autozygous alleles are due to non-random mating, whether due to inbreeding or consanguinity. They are called identical by descent. Identical by state is the name for allozygous alleles, which are due to random breeding.

Zygosity can mean also the genetic similarity of multiplets. For example, twins are either monozygotic or dizygotic whether due to single or double zygotes. In the case of monozygotic multiplets, the zygote has split to produce the siblings. Polyembryony is thus a form of clonal reproduction.

Somatic genetic heterogeneity

Human body consists on average of about 3.72 x 10^{13} cells (Bianconi et al., 2013), which are the outcome of about 10^{16} mitoses (Iourov et al., 2010). The accumulation of somatic variants has been assumed to be constant during life (Manders et al., 2021). Most of these variants do not have effect or major phenotype, they are just normal genetic variation and thus tolerated. Most forms of cancer are due to somatic variations, the number of which varies widely (Alexandrov et al., 2013). As the variation types have been discussed above, here the focus is on three somatic cell-specific phenomena: somatic mosaicism, microchimerism and aging-related somatic variants. Somatic differences appear in all multicellular organisms, although most studies are for human, there are data e.g. for plants (Diwan et al., 2014). Human somatic variation profiles are tissue-specific and may have strand asymmetries (Garcia-Nieto et al., 2019).

Mosaicism is due to postzygotic variations that differ in somatic cells, while chimerism means that the individual is derived from two or more zygotes. In microchimerism, an individual has a smaller number of cells stemming from another individual. Hybrids are formed by breeding of two (sub)species, see Variation types. Somatic hybrids are obtained when a nucleus is derived from one parent and cytoplasm is derived from both the parents, thus resulting in cytoplasmic hybrids, also called as cybrids. In heteroplasmy, there is more than one type of organellar genomes within a cell or individual. Mitochondrial cybrids have been used to study various diseases (Sazonova et al., 2018).

Somatic mosaicism

Somatic variations accumulate during the lifespan of all cells. Additional form of somatic variation is introduced in some organisms by programmed genome rearrangement (Rzeszutek et al., 2020). The size of chromosome(s) is reduced (chromatin diminution) or some chromosomes are entirely deleted in somatic cells. Consequently, the genome is different in different types of cells within an individual. In the case of some ciliates, up to 95% of genome is deleted in somatic cells during embryogenesis and individual cells show heterogeneity (Rzeszutek et al., 2020).

As discussed above, somatic mutation rate is substantially higher than germline rate, varying according to cell type. The rate in kidney tubules and appendiceal crypts is about 50 times higher than in sperms and oocytes (Manders et al., 2021). Despite the high-fidelity of DNA replication and DNA repair mechanisms, some new variants are introduced in every cell generation. Every single cell has thus somewhat different genotype, which is inherited by daughter cells, which again accumulate variations before dividing. Early variants of any size and type have large distribution and potential effect.

Variations in DNA repair system lead to accelerated aging and predisposes to cancers or both (Sharma et al., 2020). Variants that impair base excisions repair system are embryonic lethal (Hasty et al., 2003).

Somatic mosaicism is related to certain diseases when genes are expressed outside their normal context. Cancer is an example of extreme somatic micromosaicism, variants are different for each patient even for those having the same type of cancer. Certain cancers, such as lung cancer and melanoma, can contain even more than 1 million variations in comparison to other cells. Revertant mosaicism in a disease can restore partly or completely the wild type phenotype or activity by having reversion back to the original sequence (Hirschhorn, 2003).

Based on the distribution of mosaic cells, there are three categories: somatic, gonosomal and germline mosaicism (Biesecker and Spinner, 2013) which have been implicated in various diseases (Biesecker and

Spinner, 2013; Forsberg et al., 2017). Most of the mosaicism remains unnoticed as there are not severe signs or symptoms. The clinical manifestations include e.g. overgrowth or undergrowth, asymmetric growth, anomalies, streaky or patchy pigment. Cutaneous mosaicism can cause striking patterns (Biesecker and Spinner, 2013).

Microchimerism

Various forms of chimerism have been classified (Forsberg et al., 2017; Wu et al., 2016). Macrochimerism has been called also constitutional and classical chimerism and means that the cell lines from different zygotes are blended in an individual very early during embryogenesis. It is an extremely rare event. Microchimerism refers chimerism in smaller scale when there is a relatively small number of cells from another organism.

Microchimerism can be natural or artificial. Microchimerism due to medical procedures (e. g. blood transfusion or organ transplantation) is called iatrogenic. In orthotopic chimeras the transplantation has been made to cognate, related position of the cells. When donor cells are transplanted in different location, heterotopic chimera is formed. Chimerism is essential for transplant tolerance (Scandling et al., 2018).

Chimeras can be classified also based on temporal properties. In isochronic chimera the time of the cells matches with the recipient, otherwise it is a heterochronic event.

Offspring obtain cells from their mother and have thus maternal microchimerism (Loubiere et al., 2006). The cell transfer occurs also in the other direction so that mother has fetal microchimerism. Fetomaternal chimerism may persist for decades (Bianchi et al., 1996). There are even XY cells when the female has had a male pregnancy. In a multipregnancy with non-monozygotic multiplets, the siblings share the placenta and obtain blood cell transfer from each other, this is visible even when a sibling has aborted.

Gynandromorph is an organism with both male and female characteristics. They are chimeras that consist of genetically male and genetically female tissues. Gynandromorphs can be generated by several mechanisms: loss or damage of a sex chromosome, double fertilization of a binucleate egg, or symbionts (Narita et al., 2010). Depending on the stage of development when gynandromorphism occurred, the individual may have bilateral asymmetry or mosaic distribution of the different cell types. This shows in sexually dimorphic organisms, like butterflies and birds, as striking coloring patterns.

Chimera refers also to chimeric RNA transcripts that are formed by joining (parts) of two transcripts. This is a natural tissue-specific process (Frenkel-Morgenstern et al., 2012).

Aging-related somatic variation

Variations accumulate to cells during lifetime and are inherited to daughter cells. Progressive loss of physiological integrity along with impaired function are typical for aging. Out of the nine hallmarks for aging, three are directly related to DNA: genomic instability, telomere attrition and epigenetic alterations (Lopez-Otin et al., 2013).

It has been argued that DNA damage possibly affects all ageing phenotypes (Schumacher et al., 2021). DNA repair defects cause several progeroid, i.e. premature aging, syndromes such as Werner and Nestor-Guillermo progeria syndromes, along with other symptoms (Foo et al., 2019; Schumacher et al., 2021). DNA damages affect lesion correction, genome instability, mitochondrial function, metabolic alterations, epigenetics, telomere dysfunction, cell fate etc. (Schumacher et al., 2021; Vijg and Dong, 2020; Yousefzadeh et al., 2021).

INHERITANCE

Inheritance of genetic traits can be divided into two main categories: Mendelian and non-Mendelian inheritance. The former is called also monogenic inheritance due to being related to a single gene.

Monogenic inheritance

Mendelian inheritance is named after Gregor Mendel who revealed the bases of inheritance: segregation, independent assortment and dominance. Autosomal inheritance refers to inheritance of genes that are not located in sex chromosomes, i.e. in autosomes. Inheritance related to sex chromosomes is in humans and many other species either X chromosomal or Y-chromosomal (holandric). Pseudoautosomal inheritance refers to inheritance of genes in pseudoautosomal regions of the X and Y chromosome that can exchange between the two sex chromosomes. Human has 30 genes in this region (Ellison, 2009). Crossing over between X and Y chromosomes is usually restricted to these areas. Although located in sex chromosomes, these genes are inherited in autosomal fashion since both males and females have two copies of them.

Mendelian traits are either dominant or recessive in autosomes or sex chromosomes. In dominant diseases, one defective autosomal allele can cause an autosomal disorder. Huntington's disease is typically due to one defective copy of *HTT* gene for huntingtin that causes progressive degeneration of neurons in the brain. The number of CAG trinucleotide repeats in *HTT* gene in healthy individuals ranges from 10 to 26 while patients with the disease have extended repeat expansions (McColgan and Tabrizi, 2018). Autosomal recessive trait is due to variants in both autosome alleles. Individuals with only one copy of the defective allele are carriers and mostly healthy. Cystic fibrosis (CF) is due to deleterious alterations in both alleles coding for the cystic fibrosis transmembrane conductance regulator (CFTR) (De Boeck, 2020). CF affects lungs, pancreas, liver, kidneys, and intestine.

Sex-limited traits appear only in one sex, although both sexes may have the allele that determines the trait. These genes are expressed in only one sex and therefore there is no penetrance in the another. This term refers to autosomal qualities only. Variants in *CELSR1* gene for cadherin EGF LAG seven-pass G-type receptor 1 causes hereditary lymphedema only in females (Erickson et al., 2019).

Sex-linked inheritance is determined by a gene in a sex chromosome.*BTK* gene for Bruton tyrosine kinase is sex-linked as it is located in the X chromosome. Boys with deleterious variations in this gene have X-linked agammaglobulinemia, a primary immunodeficiency, that blocks B cell maturation and leads to susceptibility for infections (Vetrie et al., 1993).

non-Mendelian inheritance

Inheritance that does not follow Mendelian rules is called non-Mendelian and consists of several different effects and mechanisms. They can be grouped into nine categories: non-Mendelian dominance, multigenic inheritance, tropopeissis, polypleyri interactions, gene dosage compensation, genomic imprinting, uniparental disomy, extranuclear inheritance, and repetitive sequence inheritance. Non-Mendelian inheritance is equally important as Mendelian inheritance, the mechanisms are just more complex and there are numerous states.

non-Mendelian dominance

According to the rules of Mendelian inheritance, alleles are either recessive or dominant. Incomplete dominance and co-dominance are the two types of non-Mendelian dominance. The phenotype of a trait with incomplete dominance is intermediate. In codominance both the different alleles of a locus are expressed.

Tay-Sachs disease is an example of autosomal recessive inheritance. It is a rare neurodegenerative disease due to variations in *HEXA*gene for β-hexosaminidase (Dersh et al., 2016). Patients have two recessive alleles and therefore deficiency of the enzyme. Individuals with one defective allele are carriers and have incomplete dominance. They produce just half the amount of protein, enough for a normal life.

Human blood types describe red blood cells based on the presence or absence of inherited antigenic substances on their surface. ABO system describes blood groups A, B, and O. Alleles for the A and B blood groups are dominant. AB blood group bears the features of both the alleles due to codominance.

Multigenic inheritance

Multigenic inheritance, also called for quantitative or multifactorial inheritance, is common and in charge of numerous traits. Several genes contribute to these traits. Oligogenic inheritance is due to a few genes, whereas in polygenic inheritance a larger number of genes is involved. Many common diseases are due to

multigenic inheritance together with e.g. environmental factors (Crouch and Bodmer, 2020). These diseases include diabetes, cardiac diseases, cancer and many others.

Tropopeissis

Tropopeissis (τροποποίηση, Greek for modification) is a new term for gene interaction effects where genes jointly contribute to a phenotype. It is defined as an *interaction of genes or gene products that prevents partially or completely expression of the other gene.* It appears when one or more modifier genes mask, inhibit or suppress the expression of a gene. This phenomenon has traditionally been called epistasis, however, as there are different and contradictory definitions for epistasis (Cordell, 2002), the phenomenon is here renamed tropopeissis. Epistasis has often been considered as a dichotomous phenomenon; however, the suppression may not be complete. The capabilities of mathematical and statistical models for epistasis to explain biological phenomena are limited (Cordell, 2002). Many multigenic traits are more complicated than interactions between two loci can explain.

Theoretically, there are 512 two-locus, two-allele, two-phenotype, fully penetrant disease models (Li and Reich, 2000). The number of nonredundant models can be reduced based on different assumptions between 50 and 102 models. When continuous penetrance values are considered, there are 387 distinct types of two-locus models (Hallgrímsdóttir and Yuster, 2008). By accounting symmetry between loci and alleles, the number can be reduced to 69. Another classification concluded there to be 33 shape symmetry classes (Urbanowicz et al., 2014). There are many more interactions in systems that consist of more than two loci and where there is a range of effects. Here, some of the most common types of tropopeissis are discussed.

In dominant tropopeissis, a dominant allele at one locus can mask the expression of both alleles (dominant and recessive) at another locus. Recessive tropopeissis appears when recessive alleles at one locus mask the expression of both (dominant and recessive) alleles at another locus.

Dominant inhibitory tropopeissis means that a dominant allele at one locus masks the expression of both (dominant and recessive) alleles at a second locus. This is also known as inhibitory gene interaction. Duplicate recessive tropopeissis occurs when recessive alleles at either of the two interacting loci can mask the expression of dominant alleles at the two loci. In duplicate dominant tropopeissis, a dominant allele at either of two loci can mask the expression of recessive alleles at both the two loci. Polymeric gene interaction describes a situation where two dominant alleles have similar effect when they are separate but produce enhanced effect together.

Awns are active photosynthetic structures in barley (*Hordeum vulgare*) and contribute to grain yield. Four pairs of genes are involved in awn development. Their interactions show different forms of tropopeissis interactions, reviewed in (Huang et al., 2021).

Polypleyri interactions

Multifaceted (in Greek polypleyri, Πολύπλευρη) one-to-many interactions occur when one gene influences several traits. Polypleyri interaction is defined as *a gene or its product(s) affects at least to some extent two or more phenotypes or traits.* The phenomenon has been called pleiotropy in the past, but similar to epistasis, the definitions of pleiotropy are fuzzy, confusing and there is no consensus (Paaby and Rockman, 2013). Therefore, an unambiguous polypleyri interaction is introduced instead. Polypleyri interactions do not relate to fitness.

Polypleyri has several forms. See TARAR countermeasures for additional examples of polypleyri interaction mechanisms. For example, enzymes are promiscuous in two ways. Substrate and reaction promiscuity facilitate catalysis of various targets and processes, respectively (Vihinen, 2021a). Moonlighting activities are common additional functions that can be activated by variations (Lu and Hunter, 2018). A variant can have several different consequences even when affecting just one component. Albinism is an example of a polypleyri interaction mechanism. It is due to variants in a pigment producing enzyme. Tyrosinase (TYR) catalyzes the production of melanin and other pigments from tyrosine. The most severe form of albinism is called oculocutaneous albinism. People with this type of albinism have a wide spectrum of phenotypic

characteristics including white or pink hair, skin, and iris color and vision problems (Reissmann and Ludwig, 2013).

Gene dosage compensation

Diploid organisms have two copies of chromosomes, except for sex chromosomes in males. In organisms that have XY system, the Y chromosome is gene poor. Aneuploidies are largely harmful and detrimental; therefore, three different gene dosage mechanisms appear to correct for sex chromosome imbalance. Gene dosage compensation is important for balanced expression of proteins e.g. in protein complexes. Differences in abundances of complex-forming proteins can be harmful (Vihinen, 2021a).

In X-chromosome inactivation, one of the alleles in females is randomly silenced and it appears as tightly packed heterochromatin (Galupa and Heard, 2018). In human, the dosage compensation is virtually complete in females (Johnston et al., 2008). In marsupials, only the paternal X-chromosome is inactivated (Leeb and Wutz, 2010). XCI is based on epigenetic DNA modifications.

Another strategy to mitigate monosomy is upregulation of the X-chromosome in males. This process was initially detected in *Drosophila melanogaster* (Birchler et al., 2003), and was subsequently detected in several mammals (Nguyen and Disteche, 2006). The third process is down regulation of X-chromosomal genes in females and has been noticed in *Caenorhabditis elegans* (Jans et al., 2009). Condensin-like dosage compensation complex (DCC) binds to specific sequence motifs and facilitates the regulation.

Genomic imprinting

Constitutive monoallelic expression of the same allele in the entire organism or tissue is called imprinting and originates from gamete DNA methylation, which is maintained throughout the development of the individual or tissue. There are over 40 imprinted genomic regions in man (Court et al., 2013). Genes in these regions and their products are involved in development, metabolism, growth and behaviour. Variations in the expressed alleles can lead to diseases (Monk et al., 2019), nine diseases are known (Mackay and Temple, 2017). Deletions in the 15q11-13 region cause Prader-Willi syndrome by paternal inheritance as *SNRPN* and *NDN* gene products are defective (Kalsner and Chamberlain, 2015). Maternal alleles are imprinted in Prader-Willi syndrome. Random imprinting causes somatic genetic mosaicism.

Other forms of non-Mendelian inheritance

Uniparental disomy can also be considered as a form of non-Mendelian inheritance since there is genetic material for a chromosome region just from one of the parents. The structural characteristics (the presence of two chromosomes) are maintained although the chromosome is just from one parent. Uniparental disomy may remain unnoticed if there are no harmful variants in the duplicated chromosome. Paternal genome-wide uniparental disomy can cause Beckwith-Wiedemann syndrome when there is a harmful variant or variants (Darcy et al., 2015).

Extranuclear inheritance refers to inheritance due to genetic material outside the standard chromosomes of the organism. Organelles, mitochondria and plastids (chloroplasts, chromoplasts, leucoplasts and others), contain their own DNA and thus contribute to organellar inheritance. For example, mitochondria contain 13 protein, 22 tRNA and 2 rRNA-coding genes that are essential for the organelle function. Other proteins and e.g. tRNA molecules are imported from cytoplasm. The distribution of variant containing organelles to sister cells leads either to homo- or heteroplasmy (Nissanka and Moraes, 2020). In homoplasmy, both the cells contain variant-containing organelles, while in heteroplasmy only one of the cells and its daughter cells contain both the alleles.

Mitochondria and mitochondrial DNA are predominantly maternally inherited and has been called strict maternal inheritance. There are also other forms of mitochondrial inheritance including paternal leakage of mtRNA, maternal inheritance of stable heteroplasmy, and doubly uniparental inheritance (Breton and Stewart, 2015).

Doubly uniparental inheritance has been detected in various bivalve mollusks (Breton et al., 2007). Mitochondrial DNA is typically inherited from the mother; however, in doubly uniparental inheritance the situation is more complicated. Females inherit their mtDNA from the mother and transmit it to progeny of both sexes. Males inherit mtDNA from both parents, but they only transmit the DNA from their father. In female offspring the maternal mtDNA is degraded whereas it is maintained in the males. In angiosperm plants, potential biparental plastid inheritance in chloroplasts is widely distributed (Zhang and Sodmergen, 2010).

Plasmids are common in many organisms and introduce plasmid inheritance (Bouma and Lenski, 1988). Symbiotic infective particle inheritance (infectious heredity) refers to situation where the infectious particles, such as viruses, are transmitted to progeny (Arbuckle et al., 2010).

Transgenerational RNA inheritance is typically contributed by small RNAs: microRNAs, small intervening RNAs and piwi interacting RNAs (Bošković and Rando, 2018). However, the RNA types involved have not been extensively charted. The RNA molecules are involved in DNA methylation and histone modifications, there is not a clear picture on other effects and mechanisms.

Repetitive sequence inheritance is a special case of non-Mendelian inheritance. Certain short sequence stretches appear in variable numbers in genes and are linked to various diseases, more than 40 human diseases are known (Paulson, 2018). Repeats of three, four and five nucleotides and even longer repeated segments are linked to diseases. These sequences can be expanded in successive generations in a non-Mendelian manner. This increase in severity or earlier occurrence of symptoms is called genetic anticipation (Paulson, 2018).

SEXUAL REPRODUCTION AND SEX DETERMINATION

Reproduction of organisms can be either asexual or sexual. Parthenogenesis, asexual reproduction, is not reviewed in here although some genetic exchange is possible in prokaryotes including transformation (acquisition of genetic material from environment), transduction (phage or bacteriophage-mediated gene transfer) and conjugation (direct unidirectional gene transfer between prokaryotic cells). The focus here is in sexual reproduction since it contributes significantly to the genetics and individual heterogeneity and has a number of benefits (Crow, 1994). Sexual reproduction, the fusion of haploid gametes, is widely distributed, multicellular organisms that exclusively depend on asexual reproduction are very rare.

Gametes are reproductive cells produced in meiosis, reductional division, that generates cells with haploid genomes, i.e. containing just one copy of each chromosome pair (Figure 4A). In fertilization, haploid gametes fuse and form a zygote (Figure 4B). Thus, the new organism receives genetic material from both parents, which generates new genetic combinations and genetic diversity. Sexual reproduction produces genetically unique organisms. In endangered species and in inbred (in human called consanguineous) populations, the genetic pool is limited and leads to increase in genetic diseases and degeneration. The number of sexes varies, two is common, but there is wide range, *Schizophyllum commune*, a fungi, has > 20 000 sexes (mating types) (Brown and Casselton, 2001). Sexual reproduction generates variation that facilitates the survival of a species (Crow, 1994). The variant combinations allow adaptation to changing conditions. Due to sexual reproduction, risk of diseases is reduced. Thereby, the evolutionary fitness may be increased. Sexual reproduction can help to select beneficial variations and decrease harmful variations.

There are also some disadvantages (Crow, 1994). Sexual reproduction is energetically expensive and involves a long gestational period. It takes time and energy to find a mate and reproduce. Sexual reproduction is not a 100% successful method of creating offspring, since some chosen mates may be infertile or incompatible. Favorable variant may not always be passed to the offspring, because of inheritance patterns. There are typically fewer offspring than in organisms that have asexual reproduction.

In organisms with binary system of sexes, there are males and females; however, variations to the binary sexes are common. Intersex individuals have either both types of sex organs or have sex characteristics of both sexes (MacGillivray and Mazur, 2005). These individuals are often sterile. In human, about one out of 100 has some kind of differences in sex development (Arboleda et al., 2014). Numerous organisms are

hermaphrodites that have both types of reproductive organs and produce both types of gametes. In sequential hermaphroditism, the organism has one sex at a time. The other form is simultaneous hermaphroditism. Most flowering plants are hermaphroditic, there are also many hermaphroditic animals, mostly invertebrates. If hermaphrodites exist alongside males, the condition is called androdioecy and it appears e.g. in corals and roundworms (Weeks, 2012), while hermaphroditism alongside females is gynodioecy and is seen e.g. in angiosperms (Renner, 2014). In trioecy there are in the same species males, females and hermaphrodites. Some plants and animals have this sexual system (Chaudhuri et al., 2015). There are even cases where male flowers and hermaphrodite flowers or female flowers and hermaphrodite flowers appear on the same plant. The situations are called andromonoecy and gynomonoecy, respectively.

Some organisms can undergo sex change from one sex to another. Sequential hermaphroditism is one form. Several fish species can change their sex (Gemmell et al., 2019).

The sex of an individual is determined during sex-determination process, which is a special case of inheritance. Sex determination is genetic in many species but can be also determined environmentally and called metagamic sex determination (Figure 4B). This can happen due to various reasons, including temperature or salinity change, photoperiod or crowding in the population (Charnov and Bull, 1977), water pH, nutrition, body size etc. Temperature-dependent sex determination has been investigated in reptiles (Merchant-Larios and Díaz-Hernández, 2013). Cytoplasmic and organellar sex determination occur in some species.

Sex determination can be related in different ways to fertilization. In progamic sex determination, the sex is determined before fertilization, such as in male honeybees. The sex is determined during fertilization in syngamic sex determination, see e.g. (Mancebo Quintana and Mancebo Quintana, 2012). Most plants and animals have this type of sex determination. If the sex is determined after fertilization, it is called epigamic. Female honeybees are an example, the system is widely spread (Charnov and Bull, 1977).

There are several genetic (also called genotypic) sex-determination systems, see e.g. (Bachtrog et al., 2014). The following discussion relates to diploid organisms that have two sets of chromosomes and which is very common throughout taxon. XX/XY system (Figure 4C) is common in animals, including human, but appears also e.g. in some insects. X and Y are sex chromosomes, combinations of which determine the sex of the individual. Females have XX and males XY chromosomes. Some sex chromosome aneuploidies are known in human, such as X0 in Turner syndrome, XXY in Klinefelter syndrome, trisomy X (XXX), XYY, and XXYY (Skuse et al., 2018). In XX/XY system the sex is determined either in X- or Y-centered way. Humans are an example of Y-centered sex determination, SRY gene for sex determining region Y in Y-chromosome defines whether the individual is male or not. In many species sex determination is polygenic.

The homogametic sex can be also male. ZW/ZZ system appears in birds, reptiles and some insects. In this case, females have the two different sex chromosomes, ZW, while males have ZZ chromosomes.

In XX/X0 system females have XX genotype and males have X0 genotype, which means that they only have one sex chromosome (Figure 4D). This system is quite common in insects. Algae have U and V sex chromosomes. They have two alternating phases in their life cycle. Gametophyte is the sexual phase where U and V are assorted to spores. V is for male and U for female.

In haplodiploidy, males develop from unfertilized eggs and are haploid, and females develop from fertilized eggs and are diploid (Figure 4E). This type of sex determination is common in insects. About 15% of arthropod species reproduce through haplodiploidy (de la Filia et al., 2015). The sex of the haplodiploid organism is determined by the number of sets of chromosomes. Those formed from the union of a sperm and an egg develop as females. Interestingly, a male has no father and cannot have sons, but has a grandfather and can have grandsons. Recessive lethal and deleterious alleles are rapidly removed from such populations as the phenotype occurs always in males.

Paternal genome elimination appears in certain insects (Hodson et al., 2017) and some other organism. It is quite similar to haplodiploidy but has a different mechanism. Males develop from fertilized eggs, but they pass to the next generation only the maternally inherited haplotype. Paternal genome elimination is a form

18

of uniparental genome elimination.

In cytoplasmic sex determination, *Wolbachia,* an intracellular bacterial parasite, defines the sex in many arthropod species either due to killing or sterilizing males or by transforming them into phenotypic females (Ma et al., 2014). The genotype of mitochondria defines the sex e.g. in many gynodioecious plants (Chase, 2007).

FUNCTIONS OF DNA

DNA molecules have seven functions which have been systematized in VariO (Vihinen, 2014b). Information transfer is usually considered the most important, by some even the only DNA function. However, DNA has additional functions comprising regulation, transcription, DNA repair and replication, as well as reservoir of genetic material. Some DNA molecules have catalytic activity. The functions are discussed more in detail in (Vihinen, 2018).

Disease-causing variations are examples of alterations that change genetic information. Changes in 5' region of a gene such as promoter and transcription factor binding site (Horn et al., 2013) or enhancer (Soukup et al., 2019) affect transcription or DNA regulatory functions. DNA repair mechanisms are affected by several variation types and their context (Cai et al., 2010). DNA replication fidelity is affected by many factors (Bebenek and Ziuzia-Graczyk, 2018). Some DNA molecules have catalytic deoxyribozyme activities and can catalyze sequence-specific DNA depurination (Fresco and Amosova, 2017). Various genetic components can act as reservoir of genetic material (Johnson Pokorná and Reifová, 2021; Milligan and Lipovich, 2014).

The function of the gene is related to effects of variations in that gene. Essential genes are indispensable and harmful variants in them have effects (Rancati et al., 2018). Housekeeping genes (Eisenberg and Levanon, 2013) are common among essential genes, they are almost uniformly expressed in different cell types and tissues and independent of the condition.

FUNCTIONAL EFFECTS OF DNA VARIATIONS

Variants have a plethora of different consequences, effects and phenotypes. Many genetic alterations are tolerated and thus have no effect on phenotype or fitness.

Functional effects of DNA variants can be organized into five categories (Figure 5). They are defined according to the same principles as the protein functional effects, including the biological systems, processes and generalized consequences (Vihinen, 2021a). As the processes are described more in detail in other chapters, the mechanisms, effects and processes are only briefly mentioned when organized into categories in here.

Genetic variations either have no effect on the gene product, lead to gain-of-function, loss-of-function, antimorphic or neomorphic effect. Neomorphic means novel activity while antimorphic activity is antagonistic.

Gene dosage

Gene dosage has a deep impact on gene expression and gene product abundance. The abundance of genetic elements is affected by numerous mechanisms and processes and it has fundamental effects on the usage of genetic information. Euploidy and aneuploidy describe the numeral changes in entire genome copies and chromosomes, respectively. Structural variations within and between chromosomes contribute to chromosomal amplifications and deletions with consequent alterations to gene dosage.

Selfish genetic elements aim to increase their number in a genome. 45.0% of human genome consists of repeated selfish elements (Nurk et al., 2022), most of which have lost their capacity to translocate. In haplodiploidy, the ploidy of an individual depends on the sex of the individual, haploid for males and diploid for females. Programmed genomic rearrangements and (paternal) chromatin diminution reduce the number of genetic material and affect dosage of several genes.

Uniparental disomy reduces the number of different alleles to 1. Even location of a gene in a sex chromosome has a dosage effect. Harmful variants affect gene dosage also in somatic mosaicism.

19

Abundance of gene products

There is some correlation between the gene dosage and the abundance of gene products; however, it depends on the gene. Gene expression is usually tightly regulated, and the abundance of gene products can thus vary widely. Housekeeping genes have rather ubiquitous, constitutive and uniform expression in all normal cell types (Eisenberg and Levanon, 2013). These genes code for central functions e.g. in metabolism and signalling essential for all types of cells.

Variations that affect transcription, binding sites of transcription factors, promoter, start or stop sites, enhancers, silencers and other regulatory segments all affect transcription and thereby gene product generation. Inheritance affects how many alleles are needed for a trait to show in the phenotype. Somatic mosaicism affects the abundance when there are different alleles and one of them e.g. leads to mRNA degradation.

Missing RNA and missing protein are rather common outcomes of nucleic acid variants. Variants that impair the genetic code, such as stop codon-introducing variants at RNA level, out-of-frame insertions, deletions or indels, and variants that cause many types of aberrant splicing are recognized by RNA quality control mechanisms and degraded, therefore the transcripts are degraded and no protein is produced. In certain situations, variants can escape from NMD.

Epigenetic changes affect gene expression either by modifying DNA or by affecting histone structure. The non-Mendelian dominance differs from the Mendelian dominance. Polygenic inheritance is due to complex interactions of genes. Dosage compensation mechanisms, XCI and genomic imprinting, silence genes and chromosomes. Complicated genetic interactions, tropopeissis and polypleyri, along with extranuclear inheritance and repetitive sequence inheritance are other forms of non-Mendelian gene expression regulation.

The normal gene expression mode is biallelic. Many mammalian genes display random monoallelic expression (RME) (Chess, 2016), for details see (Vihinen, 2021a). They have mitotically stable allele-specific expression with different allelic states in clonal lineages. RME is widely spread and contributes to stochasticity in gene expression and heterogeneity in cells (Reinius and Sandberg, 2015). 0.5% to 15% of autosomal human genes exhibit monoallelic expression, at least in some cell types (Eckersley-Maslin and Spector, 2014).

Alteration of information

Variants that change the genetic information, such as protein coding regions, can change also the carried information and affect RNA splicing, protein structure, function or other properties. DNA lesions are very common - estimates reach up to hundreds of thousands per cell per day (Nakamura and Swenberg, 1999). Despite efficient and high-fidelity mechanisms, some lesions escape repair and can become variations. Large body of the variations are well tolerated, they do not have a phenotype and do not affect the fitness of the individual carrying them, because there is no major alteration in the carried information. This is normal genetic variation, which is constantly generated, and which provides the foundation for biodiversity and evolution.

Hybrids are offspring of different species and thus contain different genome sets with somewhat different information. In microchimerism, the genes of another organism are in small minority, however, still affect genetic information in the cells of foreign origin and in those interacting with them. Depending on the time when variants emerged, they are either inherited, *de novo* or somatic, and thus the spread of the potential alteration of information is either throughout the body or confined to a limited set of cells, tissues or organs.

The numbers of somatic variations increase during time as they are inherited to daughter cells and lead to somatic mosaicism. Genetic recombination produces variations by cross-over and gene conversion. Genome editing modifies genetic information in specific locations. Ribosomal frameshifting leads to different types of transcripts and protein products. Differences are generated also by alternative start and stop codons. All the changes and effects listed above are caused by changes to nucleic acid sequences. Epigenetic changes can be inherited, but they do not change the sequence.

Cell fate

DNA alterations affect also the fate of cells. Excessive number of variants within a cell leads to cell division arrest to allow the cell time to implement corrective processes. If not successful, cell becomes quiescent and cannot divide. Growth arrest is reversible in quiescence, while cellular senescence is non-reversible. In the most extreme case, programmed cell death clears cells that contain irreparable damage in their DNA. Uncontrolled growth leads to cancer.

Sex determination

For sex determination of progeny there are complex and organism-specific processes, which can be prone to DNA alterations. The consequences of such variants vary depending on the type of sex determination, whether genetic, environmental, cytoplasmic, or mitochondrial. Some factors can even cause sex change in certain organisms. Hermaphroditism is normal state in many organisms and enables sexual reproduction even despite of the original sex of individuals. Genetic conflict can affect both the viability (killing) and fertility of the offspring, especially males.

TARAR COUNTERMEASURES

Poikilosis in the form of genetic heterogeneity is pervasive and normal phenomenon and needed for evolution of species (Vihinen, 2020a). The extent of variation effects is restricted by TARAR countermeasures (Figure 6) that reduce the consequences of variations (Vihinen, 2020a, 2022). Various countermeasures reduce, restrict, tolerate and repair effects of genetic variations. Relevant countermeasures depend on the type of variation, its location and effects, the cells where it is expressed, environmental and other factors.

Tolerance

Many variants are tolerated and do not have a phenotypic effect. The ratio of neutral variants has been estimated and shown to vary largely depending on the gene and coded protein (Schaafsma and Vihinen, 2017) and references therein. Analysis of all possible single nucleotide substitutions in all the 22 human mitochondrial tRNAs indicated that 49.0% of the variants are tolerated (Niroula and Vihinen, 2016a). In short intervening RNAs (siRNAs) variants are more tolerated in the 5' end than in other regions (Amarzguioui et al., 2003).

The protein coding part of eukaryotic genomes is relatively small, in the case of human the exons of protein coding genes cover 2.94% of the genome (Consortium, 2012). We are only starting to gain insight on the significance of the other parts of the genomes. Genomes are mainly transcribed but are apparently quite tolerant to many variations.

Variation tolerance in the coded proteins is largely context dependent and relates to the functional relevance of the affected positions in gene products. Catalytic sites are among the most conserved positions in enzymes. Even in these sites, some variations are allowed depending on the role of the protein (Ribeiro et al., 2020). Enzymes with the same activity are in this respect most conserved followed by enzymes catalyzing the same reaction for different substrates and those with different activities. Note that DNA damage tolerance mechanisms refer to repair, discussed below.

Avoidance

Many genetic alterations, such as lesions, cannot be avoided, they appear stochastically at DNA sites, however, there are wide differences in mutation rates due to sequence vulnerabilities. Avoidance in genetics relates to the prevention of accumulation of potentially harmful variations due to inbreeding (Pusey and Wolf, 1996). Inbreeding avoidance reduces risk of inbreeding depression: reduced fitness of inbred individuals. Mechanisms to avoid inbreeding include kin recognition, dispersal of individuals, preference for extra-pair/extra-group copulations, delayed sexual maturation of offspring or reproductive suppression (Pusey and Wolf, 1996). In certain species, females mate with several males to increase genetic diversity of offspring.

Analysis of 41 species from six classes indicated that inbreeding avoidance is very general and depends on the risk of inbreeding depression in the species (Pike et al., 2021). Some species have a preference for inbreeding,

cichlid fish *Pelvicachromimis taeniatus* (Thünken et al., 2007) and ground tit *Parus humilis*(Wang and Lu, 2011) are examples.

Repair

Repair mechanisms and processes correct and rescue DNA variants. DNA is vulnerable for a very large number of substances, physical effects and biological processes that produce e.g. lesions and other changes. These changes can occur at any time and be due to internal or external factors. Alterations during DNA replication could be transferred to daughter cells. There are several DNA polymerases, depending on the organism, some of them have proofreading activity that substantially increases the fidelity of replication. For discussion on fidelity of DNA polymerases, see (Bebenek and Ziuzia-Graczyk, 2018). Polymerases incorporate a wrong base approximately once every 10^4 to 10^5 nucleotides polymerized; however, the fidelity is increased by proofreading activities by 10^2 to 10^3 -fold. Many factors affect the fidelity: polymerase selectivity and proofreading, mismatch repair, a supply of nucleotides, and the type of error, and the sequence context (Ganai and Johansson, 2016). The overall fidelity is one error per 10^8 to 10^{10} nucleotides polymerized.

Direct reversal repair mechanisms are specific to the damage. Although there are many types of DNA alterations, there are only three dedicated reversal processes (Yi and He, 2013). Photolyases reverse UV light-induced fusions of pyrimidine bases. O6-alkylguanine-DNA alkyltransferases reverse O-alkylated DNA damages. N-alkylated base adducts are reversed by AlkB family dioxygenases.

The double stranded nature of DNA facilitates many correcting measures. If the information is retained in one of the strands it is used as the template for repairing the altered strand.

Single strand damages are corrected either by base excision repair (BER), nucleotide excision repair (NER) or mismatch repair (Chatterjee and Walker, 2017). BER systems correct modified bases and nucleotides first by cleaving with glycosylase, then AP endonuclease nicks the strand, after which DNA polymerase removes the apurinic or apyrimidinic site and adds a new base by using the other strand as the template.

A larger stretch of DNA is repaired in NER. UV light induces formation of pyrimidine dimers between adjacent thymidines. These and other larger alterations are corrected by NER. Endonuclease removes the damaged region after which it is resynthesized with polymerase.

Mismatch repair system detects errors that have slipped through the polymerase proofreading activity. Depending on the organism, two or more proteins are involved in the detection and repair of these sites.

Double-strand breaks can be hazardous and lead to the deletion of substantial portion of a chromosome. Two different systems correct these variants (Chatterjee and Walker, 2017). In non-homologous end joining (NHEJ), a DNA ligase joins DNA strands. The process is often guided by short microhomologous segments in the ends of the strands. The process is not entirely accurate, it can introduce variations, as well as join segments that originally were not next to each other, not even in the same chromosome. Homologous recombination (HR) uses identical or nearly identical segment in a sister chromatid, or in a homologous chromosome.

DNA damage tolerance (DDT) pathways promote the bypass of single-stranded lesions during DNA replication (Pilzecker et al., 2019). The stalling of DNA replication fork can cause the fork to collapse and lead to DNA instability. In translesion synthesis (TLS), the DNA polymerase is temporarily replaced by a specialized TLS polymerase that can proceed through lesions. TLS is intrinsically error-prone. In error-free template switching (TS), the stalled nascent strand switches from the damaged template to the undamaged newly synthesized sister strand for extension past the lesion.

Telomeres are special repetitive regions in the ends of chromosomes essential for replication (Turner et al., 2019). They are shortened in every replication. Telomere healing means addition of telomere to chromosomes parts which have been deleted and lost the telomere or did not previously have a telomere. Without the healing process, these chromosomes would not replicate.

Attenuation

Attenuation mechanisms reduce and decrease the consequences of genetic variations. Traditionally genetic buffering and genetic suppression are presented as attenuation mechanisms. Genetic suppression refers to intragenic and intergenic interactions that reduce the effect of a variant (Hodgkin, 2005), whereas genetic buffering means various mechanisms and effects, such as modifier genes, gene interactions, environmental effects, phenotypic plasticity and genetic redundancy (Hartman et al., 2001; Rossi and Kontarakis, 2021). These mechanism act mainly at protein level, for discussion see (Vihinen, 2021a). Genetic redundancy functions at gene level and is thus discussed in here.

Many genomes, including that of human, have underwent one or more complete genome duplications during evolution. Thus, there have been lots of duplicated genes that have evolved further during time. Ancestral vertebrates had two rounds of whole genome duplication (Dehal and Boore, 2005). OHNOLOGS V2 database contains 7358 human two round ohnolog pairs (Singh and Isambert, 2020). Thus, many genomes contain redundant genes, products of which can attenuate effects of variants. However, even in yeast, a well studied model system, various estimates have been presented for the extent of buffering by duplicates, see (Hannay et al., 2008).

One yeast study indicated that out of 201 duplicate gene pairs, 34% were at least partially redundant in an growth rate analysis (Dean et al., 2008). 49 cases of the redundant pairs (24%) were synthetically lethal, i.e. lethal when the variants appeared together. Analysis of totally eleven species, both prokaryotes and eukaryotes, indicated that the chance of survival due to gene deletion significantly increased when there was a duplicate gene (Hannay et al., 2008). The effects of duplicates are not fully understood, in yeast, genetic interactions of the duplicates were found to be only partly overlapping (Ihmels et al., 2007). In addition, their expression patterns are not identical (Kafri et al., 2005). Many human disease-related genes have functionally redundant paralogs (Chen et al., 2013). Moonlighting activities and promiscuity of proteins further increase the complexity (Vihinen, 2021a).

Attenuation is related also to the essentiality of genes, as all genes are not indispensable and crucial for viability. Organisms can survive without many genes depending on the growth conditions. Studies of human essential genes have each indicated ˜2000 indispensable genes and proteins (Blomen et al., 2015; Hart et al., 2015; Wang et al., 2015), although the gene sets overlap only partly. The essential proteins are mainly involved in developmental processes and sequence-specific DNA binding (Lv et al., 2015). These genes are conserved, highly expressed, and they code for activities in DNA, RNA and protein synthesis, many of which are abundant and involved in macromolecular complexes.

Risk or susceptibility alleles have been identified in relation to risk to certain diseases. Protective alleles attenuate and prevent disease phenotype. They are usually found from healthy individuals where their presence prevents disease, despite other disease-promoting (susceptibility) alleles at genes elsewhere in the genome. Protective alleles have been identified in various diseases (Harper et al., 2015). Some of these alleles are evolutionarily old, for example, the genomic region that protects against severe form of COVID-19 was inherited from Neanderthals (Zeberg and Pääbo, 2021), similarly, another region that is a risk factor for severe COVID-19 also originates from Neanderthals (Zeberg and Pääbo, 2020).

Resistance

Resistance countermeasures actively support and protect the genetic integrity. DNA damage response (DDR) triggers various cellular processes (Ciccia and Elledge, 2010). Signaling pathways related to ATM, ATR and PRKDC are central for DDR (Blackford and Jackson, 2017). The three protein kinases regulate both DNA repair and related processes that control cell cycle checkpoints and apoptosis.

Cell cycle arrest in multicellular organisms prevents the division of a cell when it contains substantial amount of variations and allows time for the repair measures. There are three types of growth arrest (Pack et al., 2019). Quiescence is reversible, while cellular senescence is non-reversible and related to persistent stress due to e.g. telomere attrition or oncogene activation. As the most extreme measure, the cell undergoes programmed cell death (Roos and Kaina, 2013). Bacteria have a related system called SOS response, which varies between species (Kreuzer, 2013). It involves cell cycle arrest and DNA repair. The third growth arrest

type, terminal differentiation, is a permanent exit from the cell cycle.

Genetic resistance allows organisms to grow in stressful conditions. For example, in plants genetic resistance suppresses or retards the development of a pathogen or an insect. Pan genome means the combined genes in an organism. A pan genome can include resistance providing genes; however, they may lead to reduced fertility, in plants, seed production is reduced when a resistance gene is present (Tian et al., 2003).

Sex change or turn to hermaphroditism (Gemmell et al., 2019) in some organisms are examples of resistance when there either are not partners for sexual reproduction or when the balance of individuals with the different sexes is biased. These processes can be considered as types of genetic resistance counter mechanisms.

INTRAINDIVIDUAL GENETIC CONFLICT

The literature on genetic conflict has largely concentrated on transmission of genes (Gardner and Úbeda, 2017; McLaughlin and Malik, 2017; Werren, 2011). Intraindividual or intragenomic genetic conflict has been defined as a conflict between genetic elements when the phenotypic expression of genes promotes their own transmission (McLaughlin and Malik, 2017). However, intraindividual genetic conflict manifests also as differential gene expression varying depending on genes. Fitness is relevant for intergenomic genetic conflict but is not covered in here because it is population-related.

Cells contain various types of genes which may have conflicts. Nuclear genes in eukaryotes are the major genes, but there are also cytoplasmic genes in mitochondria, chloroplasts and other plastids, depending on the cell type and species. Genes can be classified as autosomal or sex chromosomal. In addition, there can be genes of mobile genetic elements, plasmids and intracellular microorganisms. *de novo* variants can contribute to genetic conflict. Selfish action of a certain gene and its products changes the fairness of meiosis and the probability of the traits to be transmitted to progeny. Genetic conflicts have various mechanisms.

Intragenomic conflicts have been classified in three categories based on origin, destination and situation (Gardner and Úbeda, 2017). The origin conflict means conflict between autosomal and X-chromosomal genes in males, between sex chromosomal genes, and between cytoplasmic and nuclear genes. Autosomal genes are inherited from both parents, whereas X-chromosome in males is just from the mother. Genomic imprinting silences genes and causes differential expression for genes of maternal and paternal origin. Cytoplasmic genes are either in organelles, or in cytoplasmic pathogens or plasmids.

Destination conflict is due to the transmission distortion caused by selfish genetic elements (Gardner and Úbeda, 2017). There is even a conflict between autosomal and X-linked genes in males. The autosomal genes are transmitted equally likely both to male and female offspring, while X-chromosomal genes are transmitted only to female progeny.

Selfish genetic elements can have numerous effects on their host. These include increased mutation rate, shaping of genome and its function, evolution, fitness, fertility, viability, gene flow, formation of sex chromosomes, behavior, mate preference, and sex determination (Wedell, 2020). Both over replicating elements and transmission distorters (see Selfish genetic elements) cause destination conflict. Homing endonucleases turn the affected alleles as homozygous. Medea is an example of lethal maternal effect by toxin-antidote system (Beeman et al., 1992). A mother who carries *Medea* gene expresses the toxin in her germline and thus kills her progeny. When offspring carry *Medea* they produce antidote and survive. If the mother has one copy of *Medea* allele, one half of the offspring inherit the gene and therefore survive.

Cytoplasmic components, such as mitochondria, are typically inherited from mother, which can lead to cytoplasmic genetic conflict. Organelles have various mechanisms to increase the production of female descendants. Human egg contains more than 100 000 mitochondria, while sperms have about 100 copies. There can be conflict between the mitochondria from the parents. Paternal mitochondrial DNA is rapidly eliminated after fertilization. Some rare cases of biparental mitochondrial inheritance have been reported (Luo et al., 2018) and can be called as paternal leakage in the case of maternal inheritance.

Male-killing can be caused by inherited protists or bacteria, such as *Wolbachia* , in cytoplasm. Eggs from

24

uninfected females can be inviable due to *Wolbachia* -containing sperms (Hoffmann et al., 1990). Another male affecting mechanism is feminization which converts males to females (Kageyama et al., 2002). Male sterility is still another mechanism of cytoplasmic conflict (Touzet and Budar, 2004).

The third type, situation conflict (Gardner and Úbeda, 2017) of greenbeard genes (Madgwick et al., 2019) is related to intergenomic genetic conflict and therefore not discussed in here.

GENETIC HETEROGENEITY AND DISEASES

The medical interpretation of effects of genetic variation is most straightforward in single gene Mendelian diseases. The phenotypic expression of variants is affected by several processes, mechanisms and factors (Figure 7). Thus, despite having a harmful genetic alteration, the phenotype can be attenuated by TARAR countermeasures. On the other hand, e.g. genetic conflict could even increase the effect of a variation. In addition to the genetic component and phenotypic expression, life style and environmental effects are pivotal in many conditions.

Any type and size of variation can cause disease. SNVs are the smallest genetic changes and large numbers of substitutions have been identified in thousands of diseases. Large part of variations, in many genes the majority, are benign and without phenotype. Experimental studies (James et al., 2022) and predictions (Niroula and Vihinen, 2016a; Schaafsma and Vihinen, 2017; Väliaho et al., 2015) have indicated that genes display a wide range of vulnerabilities for genetic alterations. Even in Cancer Gene Census -classified cancer-related genes (Sondka et al., 2018) with substitutions, just 40% of amino acid substitutions were predicted to be harmful (Niroula and Vihinen, 2015). Note that the results from massively parallel reporter assays, also called saturation mutagenesis studies, cannot be directly correlated to biological effect and disease relevance.

Genetic heterogeneity is pervasive in genetic diseases in many ways and levels. The inheritance pattern defines how inherited alleles interact. Figure 7 shows how genetic variation and introduced information alterations affect phenotypic expression and, once the effects are large enough, can lead to disease. The correlation between genetic variation and disease is most obvious in Mendelian diseases. Even in these cases, penetrance and phenotypes can vary even for individuals carrying exactly the same genetic alteration leading even to different diseases (Zhu et al., 2014) and even in individuals having the same genetic background (monozygotic multiplets) (Dulovic-Mahlow et al., 2021; Ketelaar et al., 2012; Zhu et al., 2021). Because of both genetic and phenotypic heterogeneity, it is necessary to increase sizes of cohorts in experimental investigations, e.g. in genome wide association studies (Manchia et al., 2013).

Incomplete penetrance and heterogeneous expressivity appear almost in all diseases, at least to some degree. Variable expressivity and consequent heterogeneous severity of a disease are due to several factors and mechanisms, discussed in (Gruber and Bogunovic, 2020; Zlotogora, 2003). From the point of disease, genetic variation has many dimensions in addition to the nucleic acid changes, it depends also on which kind of cell the variation appears, what kind of distribution the variant has, and how the gene is expressed in those cells. Species origin as well as the origin and type of gene can modify the phenotype. Many types of variations alter the dosage of genes, chromosomes and genomes and thus affect the abundance of gene products (Figure 5).

Genetic changes manifest phenotypic expression in multiple ways and levels. Sex determination is important for the individual as well as for the population. Inheritance mode defines how the genetic change is expressed. Genetic conflict and TARAR countermeasures modify and reduce the phenotypic expression, respectively. Severe genetic alterations are decisive for the cell fate. Functional effects and changes in abundance of gene products modify the type and extent of activity change.

Diseases are continua and combinations of effects, symptoms and mechanisms. Pathogenicity model (PM) describes diseases as the combination of three constituent factors: extent, modulation and severity, which are unique for each disease (Vihinen, 2017). In the PM, the pathogenicity of a disease is described jointly by the distribution of the three factors in a cohort of healthy and diseased individuals. Extent measures the breadth of disease appearance. Severity of the disease indicates the state or degree to which the disease

is expressed. Modulation means combined effects of many factors that modify disease phenotype, including TARAR countermeasures, genetic conflict and others.

Evidence-based cut-offs in the PM can be used for various purposes (Vihinen, 2017, 2020c). One such plane can be applied in diagnosis, another to define which cases are actionable, i.e. can be treated with existing regimes. The variability in the PM can explain also why some individuals benefit from a certain medication, while others do not respond or even have adverse drug reactions.

Somatic variations and those that emerge during development have diverse effects. Many of these variants are without phenotype, especially, if the gene with the variant is not expressed in cells that contain the variant. Variations accumulate during lifetime and contribute to somatic heterogeneity. The accumulation of genetic variations is one of the hallmarks of cancer (Hanahan and Weinberg, 2000) and important in many other diseases. In most cancers, majority of the variants are so called passengers that have little or no effect on the disease. Carcinogen-induced tumors, such as lung cancer and melanoma, can contain hundreds of thousands or more variations (Alexandrov et al., 2013), most of which are passengers. Cancer is the most heterogenic and personal disease, since the genetic basis is different in every patient (Burrell et al., 2013). The accumulation of variants to key genes contributes to the progression of cancer.

The functional effects of genetic variants can vary widely even within a single gene disease. There are examples where the change depends on the position and/or type of a variant and has either gain-of-function (gof) or loss-of-function (lof) effect or no effect at all. Lof and gof classifications should always be based on experimental evidence (Vihinen, 2021a). In some genes, certain variants have gof effect while others are of lof type. In *PPP3CA* gene for protein phosphatase 3 catalytic subunit alpha, lof and gof variants lead to early onset epileptic encephalopathy and multiple congenital abnormalities, respectively (Mizuguchi et al., 2018).

Most current therapies of genetic diseases are targeted to the functional consequences of the genetic alterations. In gene therapy, the goal is to modify cells to generate therapeutic effect, often by correcting the defective gene. Hematopoietic stem cell transplantation is widely used to correct gene defects by introducing cells with fully functional versions of genes in primary immunodeficiencies, certain cancers and some other diseases (Juric et al., 2016). Recombinant viruses (Lukashev and Zamyatnin, 2016) and non-viral vectors (polymers, lipids, peptides, inorganic materials, hybrid systems) (Zu and Gao, 2021) have been used to transform cells by inserting genetic material to chromosomes. More recently, techniques for genome editing have emerged and could allow direct correction of nonfunctional genes or gene products (Doudna, 2020). This is a form of curative treatment, but still in experimental phase.

Genetic counseling is an important part of genetic disease diagnosis and family planning (Patch and Middleton, 2018). The heterogeneity of the phenotypic expression and disease emergence make the task challenging. Consanguinity/inbreeding, which is common in some populations and species, influences the risk of inherited diseases.

FUTURE PERSPECTIVE

The gold standard methods for determining effects of variants are experimental, however, not possible to use to investigate all variants. 4.1-5.0 million sites differ in an individual in comparison to the reference genome, and about 10 000 – 12 000 of them cause amino acid substitutions (Auton et al., 2015). Therefore, computational solutions are instrumental for the prediction of disease-relevance, called pathogenicity or tolerance, of various types of genetic changes (Niroula and Vihinen, 2016b).

There is place for improvement in variation interpretation. Although the performance has increased during years, still many use even decades old methods with poor performance (Vihinen, 2020b). NGS technologies generate genetic data sets and now even longer structural variations can be identified with high reliability, which has facilitated filling the difficult to sequence gaps in human genome and obtaining the complete sequence (Nurk et al., 2022). Problems emerge in the processing of identified variants. Widely used annotation tools provide some erroneous descriptions, especially for synonymous, splicing and truncation variants

(Vihinen, submitted). Many disease-causing variants that are annotated as synonymous are nonsynonymous (Vihinen, submitted). The consequences of RNA quality control (Kurosaki et al., 2019) are largely ignored by the existing tools.

Most of the predictors for variation interpretation do not take the genetic and phenotypic heterogeneity into account, they have just two classes for benign and pathogenic variants. Only some methods consider the phenotype as a continuum. These tools include PON-P2 for amino acid substitutions in human (Niroula et al., 2015), PON-All for amino acid alterations in all organisms (Yang et al., 2022) and PON-PS, a variant severity predictor for human amino acid substitutions (Niroula and Vihinen, 2017). Depending on the gene and disease, some portion of variants cause heterogeneous phenotype in individuals and cannot be classified in binary categories of benign and pathogenic. The International Society of Gastrointestinal Hereditary Tumours (InSiGHT) has classified variants in mismatch repair system genes for many years. They collect world-wide information for patients with a certain genotype and then apply the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (Richards et al., 2015) and other criteria to classify the variants (Thompson et al., 2014). Currently there are 6454 unique variants (data from InSiGHT databases in LOVD) for four genes (*MLH1* , *MSH2* ,*MSH6* , *PSM2*). 2453 (38%) of them are unclassified variants (UV) after several years of systematic data collection and classification. There will always be heterogeneity and some cases will thus remain in the UV class in all disease-related genes.

Variations have been collected and made available in numerous publicly available databases. These resources are instrumental for variation interpretation, clinical diagnosis and research purposes. The funding of locus-specific variation databases (LSDBs) should be secured since they are typically the most reliable sources of variation information and curated by disease experts. LSDBs are too small to be considered and funded as infrastructures, and after initial publications, not eligible for research grants. Therefore, many of these valuable resources are maintained without dedicated funds, from the interest of the curators, situation that is not sustainable for long. Large variation databases such as ClinVar (Landrum et al., 2016), dbSNP (Sherry et al., 2001) and dbVar (Lappalainen et al., 2013) will continue as essential resources and have permanent funding. They should, however, pay more attention to record genetic heterogeneity and its phenotypic consequences.

The genetic heterogeneity and diversity have been largely ignored or considered as irrelevant in many studies. Wider ancestor distribution of cases is required, as the existing studies are often biased, dominated by individuals with European ancestry (Sirugo et al., 2019). Variation interpretation in many conditions requires better consideration of life style and environmental factors (Alfredsson and Olsson, 2019; Marras et al., 2019; Zhang et al., 2021). The inclusion of the effect of somatic variations in germline-based diseases and effects of microbiota are other important future research lines.

Our understanding is still very limited about the genetics of common, multigenic diseases. Rare monogenic diseases have been valuable in many ways in revealing genetic mechanisms. In common diseases, the interactions of genes and variants are relevant. In addition to coding genes, other types of functions have to be addressed more closely. For example, variants affecting regulatory functions need attention.

The investigations of variations and diseases need to step beyond the identification of genetic alterations. Functional studies to reveal mechanisms of the gene products (Raraigh et al., 2018) will, in addition to increasing understanding of variation consequences, phenotypic expression and disease mechanisms, pinpoint novel drug and other therapeutic targets.

The contemplation of genetic and phenotypic heterogeneity is essential in population screening, prenatal testing and newborn screening for improved detection rate. The availability of genomic sequencing is essential for implementing personalized/precision medicine (Carrasco-Ramiro et al., 2017; Goetz and Schork, 2018). Genetic heterogeneity is an essential factor and the complex interplay between numerous factors requires additional research and a paradigm shift from variation identification to variation effect understanding.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ACKNOWLEDGEMENTS

REFERENCES

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nature genetics *47* , 1402-1407.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L.*, et al.* (2013). Signatures of mutational processes in human cancer. Nature *500* , 415-421.

Alfredsson, L., and Olsson, T. (2019). Lifestyle and environmental factors in multiple sclerosis. Cold Spring Harb Perspect Med *9* .

Alvarez, J.B., and Guzmán, C. (2018). Interspecific and intergeneric hybridization as a source of variation for wheat grain quality improvement. Theor Appl Genet *131* , 225-251.

Amarzguioui, M., Holen, T., Babaie, E., and Prydz, H. (2003). Tolerance for mutations and chemical modifications in a siRNA. Nucleic acids research *31* , 589-595.

Antonarakis, S.E., Skotko, B.G., Rafii, M.S., Strydom, A., Pape, S.E., Bianchi, D.W., Sherman, S.L., and Reeves, R.H. (2020). Down syndrome. Nat Rev Dis Primers *6* , 9.

Arboleda, V.A., Sandberg, D.E., and Vilain, E. (2014). DSDs: genetics, underlying pathologies and psychosexual differentiation. Nat Rev Endocrinol *10* , 603-615.

Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., De Meirleir, K., Montoya, J.G.*, et al.* (2010). The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proceedings of the National Academy of Sciences of the United States of America *107* , 5563-5568.

Arnheim, N., and Calabrese, P. (2009). Understanding what determines the frequency and pattern of human germline mutations. Nature reviews. Genetics *10* , 478-488.

Asher, D.M., and Gregori, L. (2018). Human transmissible spongiform encephalopathies: historic view. Handb Clin Neurol *153* , 1-17.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526* , 68-74.

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R.*, et al.* (2014). Sex determination: why so many ways of doing it? PLoS biology *12* , e1001899.

Baer, C.F., Miyamoto, M.M., and Denver, D.R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. Nature reviews. Genetics *8* , 619-631.

Barow, M. (2006). Endopolyploidy in seed plants. BioEssays : news and reviews in molecular, cellular and developmental biology *28* , 271-281.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science (New York, N.Y.) *327* , 836-840.

Bebenek, A., and Ziuzia-Graczyk, I. (2018). Fidelity of DNA replication-a matter of proofreading. Curr Genet *64* , 985-996.

Beeman, R.W., Friesen, K.S., and Denell, R.E. (1992). Maternal-effect selfish genes in flour beetles. Science (New York, N.Y.) *256* , 89-92.

Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D.C., Tamuri, A.U., Martincorena, I., Petljak, M., Alexandrov, L.B., Gundem, G., *et al.* (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature *513* , 422-425.

Beisson, J. (2008). Preformed cell structure and cell heredity. Prion *2* , 1-8.

Bell, A.D., Mello, C.J., Nemesh, J., Brumbaugh, S.A., Wysoker, A., and McCarroll, S.A. (2020). Insights into variation in meiosis from 31,228 human sperm genomes. Nature *583* , 259-264.

Bianchi, D.W., Zickwolf, G.K., Weil, G.J., Sylvester, S., and DeMaria, M.A. (1996). Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. Proceedings of the National Academy of Sciences of the United States of America *93* , 705-708.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., *et al.* (2013). An estimation of the number of cells in the human body. Ann Hum Biol *40* , 463-471.

Biesecker, L.G., and Spinner, N.B. (2013). A genomic view of mosaicism and human disease. Nature reviews. Genetics *14* , 307-320.

Birchler, J.A., Pal-Bhadra, M., and Bhadra, U. (2003). Dosage dependent gene regulation and the compensation of the X chromosome in Drosophila males. Genetica *117* , 179-190.

Blackford, A.N., and Jackson, S.P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. Molecular cell *66* , 801-817.

Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., *et al.* (2015). Gene essentiality and synthetic lethality in haploid human cells. Science (New York, N.Y.) *350* , 1092-1096.

Bošković, A., and Rando, O.J. (2018). Transgenerational epigenetic inheritance. Annu Rev Genet *52* , 21-41.

Bouma, J.E., and Lenski, R.E. (1988). Evolution of a bacteria/plasmid association. Nature *335* , 351-352.

Breton, S., Beaupré, H.D., Stewart, D.T., Hoeh, W.R., and Blier, P.U. (2007). The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends in genetics : TIG *23* , 465-474.

Breton, S., and Stewart, D.T. (2015). Atypical mitochondrial inheritance patterns in eukaryotes. Genome *58* , 423-431.

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet *62* , 1408-1415.

Brookes, A.J. (1999). The essence of SNPs. Gene *234* , 177-186.

Brown, A.J., and Casselton, L.A. (2001). Mating in mushrooms: increasing the chances but prolonging the affair. Trends in genetics : TIG *17* , 393-400.

Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501* , 338-345.

29

Cabrera, V.M. (2021). Human molecular evolutionary rate, time dependency and transient polymorphism effects viewed through ancient and modern mitochondrial DNA genomes. Scientific reports *11* , 5036.

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T.H.H., Sanders, M.A., Lawson, A.R.J., Harvey, L.M.R., Bhosle, S., Jones, D., *et al.* (2022). Somatic mutation rates scale with lifespan across mammals. Nature *604* , 517-524.

Cai, Y., Patel, D.J., Broyde, S., and Geacintov, N.E. (2010). Base sequence context effects on nucleotide excision repair. J Nucleic Acids*2010* .

Camara, P.G., Rosenbloom, D.I., Emmett, K.J., Levine, A.J., and Rabadan, R. (2016). Topological data analysis generates high-resolution, genome-wide maps of human recombination. Cell Syst *3* , 83-94.

Carlson, J., Locke, A.E., Flickinger, M., Zawistowski, M., Levy, S., Myers, R.M., Boehnke, M., Kang, H.M., Scott, L.J., Li, J.Z., and Zöllner, S. (2018). Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. Nat Commun *9* , 3753.

Carrasco-Ramiro, F., Peiró-Pastor, R., and Aguado, B. (2017). Human genomics projects and precision medicine. Gene Ther *24* , 551-561.

Chandler, V.L. (2007). Paramutation: from maize to mice. Cell*128* , 641-645.

Charnov, E.L., and Bull, J. (1977). When is sex environmentally determined? Nature *266* , 828-830.

Chase, C.D. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends in genetics : TIG*23* , 81-90.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen *58* , 235-263.

Chaudhuri, J., Bose, N., Tandonnet, S., Adams, S., Zuco, G., Kache, V., Parihar, M., von Reuss, S.H., Schroeder, F.C., and Pires-daSilva, A. (2015). Mating dynamics in a nematode with three sexes and its evolutionary implications. Scientific reports *5* , 17676.

Cheetham, S.W., Faulkner, G.J., and Dinger, M.E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. Nature reviews. Genetics *21* , 191-201.

Chen, J.M., Cooper, D.N., Chuzhanova, N., Férec, C., and Patrinos, G.P. (2007). Gene conversion: mechanisms, evolution and human disease. Nature reviews. Genetics *8* , 762-775.

Chen, W.H., Zhao, X.M., van Noort, V., and Bork, P. (2013). Human monogenic disease genes have frequently functionally redundant paralogs. PLoS computational biology *9* , e1003073.

Chess, A. (2016). Monoallelic gene expression in mammals. Annu Rev Genet*50* , 317-327.

Chi, X., Li, Y., and Qiu, X. (2020). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. Immunology *160* , 233-247.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Molecular cell *40* , 179-204.

Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489* , 57-74.

Cordell, H.J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human molecular genetics*11* , 2463-2468.

Court, F., Martin-Trujillo, A., Romanelli, V., Garin, I., Iglesias-Platas, I., Salafsky, I., Guitart, M., Perez de Nanclares, G., Lapunzina, P., and Monk, D. (2013). Genome-wide allelic methylation analysis reveals disease-specific susceptibility to multiple methylation defects in imprinting syndromes. Human mutation *34* , 595-602.

30

Crouch, D.J.M., and Bodmer, W.F. (2020). Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proceedings of the National Academy of Sciences of the United States of America *117* , 18924-18933.

Crow, J.F. (1994). Advantages of sexual reproduction. Dev Genet *15* , 205-213.

Darcy, D., Atwal, P.S., Angell, C., Gadi, I., and Wallerstein, R. (2015). Mosaic paternal genome-wide uniparental isodisomy with down syndrome. Am J Med Genet A *167a* , 2463-2469.

De Boeck, K. (2020). Cystic fibrosis in the year 2020: A disease with a new face. Acta Paediatr *109* , 893-899.

de la Filia, A.G., Bain, S.A., and Ross, L. (2015). Haplodiploidy and the reproductive ecology of Arthropods. Curr Opin Insect Sci *9* , 36-43.

Dean, E.J., Davis, J.C., Davis, R.W., and Petrov, D.A. (2008). Pervasive and persistent redundancy among duplicated genes in yeast. PLoS genetics *4* , e1000113.

Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS biology *3* , e314.

Dersh, D., Iwamoto, Y., and Argon, Y. (2016). Tay-Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum-associated degradation. Mol Biol Cell *27* , 3813-3827.

Diwan, D., Komazaki, S., Suzuki, M., Nemoto, N., Aita, T., Satake, A., and Nishigaki, K. (2014). Systematic genome sequence differences among leaf cells within individual trees. BMC Genomics *15* , 142.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F.*, et al.* (2012). Landscape of transcription in human cells. Nature *489* , 101-108.

Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. Nature *578* , 229-236.

Dulovic-Mahlow, M., König, I.R., Trinh, J., Diaw, S.H., Urban, P.P., Knappe, E., Kuhnke, N., Ingwersen, L.C., Hinrichs, F., Weber, J.*, et al.* (2021). Discordant monozygotic Parkinson disease twins: role of mitochondrial integrity. Ann Neurol *89* , 158-164.

Duncan, A.W. (2013). Aneuploidy, polyploidy and ploidy reversal in the liver. Semin Cell Dev Biol *24* , 347-356.

Eckersley-Maslin, M.A., and Spector, D.L. (2014). Random monoallelic expression: regulating gene expression one allele at a time. Trends in genetics : TIG *30* , 237-244.

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends in genetics : TIG *29* , 569-574.

Ellison, J.W. (2009). Pseudoautosomal inheritance. In Encyclopedia of Life Sicences (ELS) (Chichester, John Wiley Sons).

Erickson, R.P., Lai, L.W., Mustacich, D.J., Bernas, M.J., Kuo, P.H., and Witte, M.H. (2019). Sex-limited penetrance of lymphedema to females with CELSR1 haploinsufficiency: A second family. Clin Genet *96* , 478-482.

Fang, H., Disteche, C.M., and Berletch, J.B. (2019). X Inactivation and escape: epigenetic and structural features. Front Cell Dev Biol *7* , 219.

Foo, M.X.R., Ong, P.F., and Dreesen, O. (2019). Premature aging syndromes: From patients to mechanism. J Dermatol Sci *96* , 58-65.

Forsberg, L.A., Gisselsson, D., and Dumanski, J.P. (2017). Mosaicism in health and disease - clones picking up speed. Nature reviews. Genetics *18* , 128-142.

Frenkel-Morgenstern, M., Lacroix, V., Ezkurdia, I., Levin, Y., Gabashvili, A., Prilusky, J., Del Pozo, A., Tress, M., Johnson, R., Guigo, R., and Valencia, A. (2012). Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. Genome research *22* , 1231-1242.

Fresco, J.R., and Amosova, O. (2017). Site-specific self-catalyzed DNA depurination: a biological mechanism that leads to mutations and creates sequence diversity. Annu Rev Biochem *86* , 461-484.

Fukaya, T., and Levine, M. (2017). Transvection. Curr Biol *27* , R1047-r1049.

Galupa, R., and Heard, E. (2018). X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. Annu Rev Genet *52* , 535-566.

Ganai, R.A., and Johansson, E. (2016). DNA replication - a matter of fidelity. Molecular cell *62* , 745-755.

García-Nieto, P.E., Morrison, A.J., and Fraser, H.B. (2019). The somatic mutation landscape of the human body. Genome Biol *20* , 298.

Gardner, A., and Úbeda, F. (2017). The meaning of intragenomic conflict. Nat Ecol Evol *1* , 1807-1815.

Gemmell, N.J., Todd, E.V., Goikoetxea, A., Ortega-Recalde, O., and Hore, T.A. (2019). Natural sex change in fish. Curr Top Dev Biol *134* , 71-117.

Genesio, R., Melis, D., Gatto, S., Izzo, A., Ronga, V., Cappuccio, G., Lanzo, A., Andria, G., D'Esposito, M., Matarazzo, M.R.*, et al.*(2011). Variegated silencing through epigenetic modifications of a large Xq region in a case of balanced X;2 translocation with Incontinentia Pigmenti-like phenotype. Epigenetics *6* , 1242-1247.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. Genome research *17* , 669-681.

Goetz, L.H., and Schork, N.J. (2018). Personalized medicine: motivation, challenges, and progress. Fertil Steril *109* , 952-963.

Gruber, C., and Bogunovic, D. (2020). Incomplete penetrance in primary immunodeficiency: a skeleton in the closet. Hum Genet *139* , 745-757.

Hallgrímsdóttir, I.B., and Yuster, D.S. (2008). A complete classification of epistatic two-locus models. BMC Genet *9* , 17.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell*100* , 57-70.

Hannay, K., Marcotte, E.M., and Vogel, C. (2008). Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation. BMC Genomics *9* , 609.

Harper, A.R., Nayee, S., and Topol, E.J. (2015). Protective alleles and modifier variants in human health and disease. Nature reviews. Genetics*16* , 689-701.

Harrison, P.T., and Hart, S. (2018). A beginner's guide to gene editing. Exp Physiol *103* , 439-448.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S.*, et al.* (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell *163* , 1515-1526.

Hartman, J.L.t., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. Science (New York, N.Y.) *291* , 1001-1004.

Hassan, S., Surakka, I., Taskinen, M.R., Salomaa, V., Palotie, A., Wessman, M., Tukiainen, T., Pirinen, M., Palta, P., and Ripatti, S. (2021). High-resolution population-specific recombination rates and their effect on phasing and genotype imputation. Eur J Hum Genet*29* , 615-624.

Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. Nature reviews. Genetics *2* , 280-291.

Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H., and Vijg, J. (2003). Aging and genome maintenance: lessons from the mouse? Science (New York, N.Y.) *299* , 1355-1359.

Higgs, P.G., and Lehman, N. (2015). The RNA World: molecular cooperation at the origins of life. Nature reviews. Genetics *16* , 7-17.

Hirschhorn, R. (2003). In vivo reversion to normal of inherited mutations in humans. J Med Genet *40* , 721-728.

Hodgkin, J. (2005). Genetic suppression. WormBook, 1-13.

Hodgkinson, A., and Eyre-Walker, A. (2011). Variation in the mutation rate across mammalian genomes. Nature reviews. Genetics *12* , 756-766.

Hodson, C.N., Hamilton, P.T., Dilworth, D., Nelson, C.J., Curtis, C.I., and Perlman, S.J. (2017). Paternal genome elimination in Liposcelis booklice (insecta: Psocodea). Genetics *206* , 1091-1100.

Hoffmann, A.A., Turelli, M., and Harshman, L.G. (1990). Factors affecting the distribution of cytoplasmic incompatibility in Drosophila simulans. Genetics *126* , 933-948.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K.*, et al.* (2013). TERT promoter mutations in familial and sporadic melanoma. Science (New York, N.Y.) *339* , 959-961.

Hu, Z., Scott, H.S., Qin, G., Zheng, G., Chu, X., Xie, L., Adelson, D.L., Oftedal, B.E., Venugopal, P., Babic, M.*, et al.* (2015). Revealing missing human protein isoforms based on ab initio prediction, RNA-seq and proteomics. Scientific reports *5* , 10940.

Huang, B., Wu, W., and Hong, Z. (2021). Genetic interactions of awnness genes in barley. Genes (Basel) *12* .

Ihmels, J., Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. (2007). Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Molecular systems biology *3* , 86.

Iourov, I.Y., Vorsanova, S.G., and Yurov, Y.B. (2010). Somatic genome variations in health and disease. Curr Genomics *11* , 387-396.

Izumi, H., and Kaneko, Y. (2012). Evidence of asymmetric cell division and centrosome inheritance in human neuroblastoma cells. Proceedings of the National Academy of Sciences of the United States of America*109* , 18048-18053.

James, P.A., Fortuno, C., Li, N., Lim, B.W.X., Campbell, I.G., and Spurdle, A.B. (2022). Estimating the proportion of pathogenic variants from breast cancer case-control data: Application to calibration of ACMG/AMP variant classification criteria. Human mutation.

Jans, J., Gladden, J.M., Ralston, E.J., Pickle, C.S., Michel, A.H., Pferdehirt, R.R., Eisen, M.B., and Meyer, B.J. (2009). A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. Genes Dev *23* , 602-618.

Johnson Pokorná, M., and Reifová, R. (2021). Evolution of B Chromosomes: From Dispensable Parasitic Chromosomes to Essential Genomic Players. Frontiers in genetics *12* , 727570.

Johnston, C.M., Lovell, F.L., Leongamornlert, D.A., Stranger, B.E., Dermitzakis, E.T., and Ross, M.T. (2008). Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. PLoS genetics *4* , e9.

Jolly, C.J., Woolley-Barker, T., Beyene, S., Disotell, T.R., and Phillips-Conroy, J.E. (1997). Intergeneric hybrid baboons. Int. J. Primatol. *18* , 597-627.

Juric, M.K., Ghimire, S., Ogonek, J., Weissinger, E.M., Holler, E., van Rood, J.J., Oudshoorn, M., Dickinson, A., and Greinix, H.T. (2016). Milestones of hematopoietic stem cell transplantation - from first human studies to current developments. Front Immunol *7* , 470.

Kafri, R., Bar-Even, A., and Pilpel, Y. (2005). Transcription control reprogramming in genetic backup circuits. Nature genetics *37* , 295-299.

Kageyama, D., Nishimura, G., Hoshizaki, S., and Ishikawa, Y. (2002). Feminizing Wolbachia in an insect, Ostrinia furnacalis (Lepidoptera: Crambidae). Heredity (Edinb) *88* , 444-449.

Kalsner, L., and Chamberlain, S.J. (2015). Prader-Willi, Angelman, and 15q11-q13 duplication syndromes. Pediatr Clin North Am *62* , 587-606.

Kessler, M.D., Loesch, D.P., Perry, J.A., Heard-Costa, N.L., Taliun, D., Cade, B.E., Wang, H., Daya, M., Ziniti, J., Datta, S., *et al.*(2020). De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. Proceedings of the National Academy of Sciences of the United States of America*117* , 2560-2569.

Ketelaar, M.E., Hofstra, E.M., and Hayden, M.R. (2012). What monozygotic twins discordant for phenotype illustrate about mechanisms influencing genetic forms of neurodegeneration. Clin Genet *81* , 325-333.

Kloosterman, W.P., Francioli, L.C., Hormozdiari, F., Marschall, T., Hehir-Kwa, J.Y., Abdellaoui, A., Lameijer, E.W., Moed, M.H., Koval, V., Renkens, I., *et al.* (2015). Characteristics of de novo structural changes in the human genome. Genome research *25* , 792-801.

Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic acids research *29* , 3742-3756.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., *et al.* (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature *488* , 471-475.

Kreuzer, K.N. (2013). DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. Cold Spring Harb Perspect Biol *5* , a012674.

Kumar, S., Chinnusamy, V., and Mohapatra, T. (2018). Epigenetics of modified DNA bases: 5-methylcytosine and beyond. Frontiers in genetics*9* , 640.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol *20* , 406-420.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., *et al.* (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic acids research *44* , D862-868.

Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J.D., Garner, J., Chen, C., Maguire, M., Corbett, M., Zhou, G., *et al.*(2013). DbVar and DGVa: public archives for genomic structural variation. Nucleic acids research *41* , D936-941.

Leeb, M., and Wutz, A. (2010). Mechanistic concepts in X inactivation underlying dosage compensation in mammals. Heredity (Edinb) *105* , 64-70.

Li, J., Zhou, J., Zhang, Y., Yang, Y., Pu, Q., and Tao, D. (2020). New insights into the nature of interspecific hybrid sterility in rice. Front Plant Sci *11* , 555572.

Li, W., and Reich, J. (2000). A complete enumeration and classification of two-locus disease models. Hum Hered *50* , 334-349.

34

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. Prog Nucleic Acid Res Mol Biol *22* , 135-192.

Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. Cold Spring Harb Symp Quant Biol *65* , 127-133.

Lindeboom, R.G.H., Vermeulen, M., Lehner, B., and Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nature genetics *51* , 1645-1651.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153* , 1194-1217.

Loubière, L.S., Lambert, N.C., Flinn, L.J., Erickson, T.D., Yan, Z., Guthrie, K.A., Vickers, K.T., and Nelson, J.L. (2006). Maternal microchimerism in healthy adults in lymphocytes, monocyte/macrophages and NK cells. Lab Invest *86* , 1185-1192.

Lu, Z., and Hunter, T. (2018). Metabolic kinases moonlighting as protein kinases. Trends in biochemical sciences *43* , 301-310.

Lukashev, A.N., and Zamyatnin, A.A., Jr. (2016). Viral vectors for gene therapy: current state and clinical perspectives. Biochemistry (Mosc)*81* , 700-708.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J.*, et al.* (2018). Biparental inheritance of mitochondrial DNA in humans. Proceedings of the National Academy of Sciences of the United States of America *115* , 13039-13044.

Lv, W., Zheng, J., Luan, M., Shi, M., Zhu, H., Zhang, M., Lv, H., Shang, Z., Duan, L., Zhang, R., and Jiang, Y. (2015). Comparing the evolutionary conservation between human essential genes, human orthologs of mouse essential genes and human housekeeping genes. Briefings in bioinformatics *16* , 922-931.

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. Proceedings of the National Academy of Sciences of the United States of America *107* , 961-968.

Lyttle, T.W. (1991). Segregation distorters. Annu Rev Genet *25* , 511-557.

Ma, W.J., Vavre, F., and Beukeboom, L.W. (2014). Manipulation of arthropod sex determination by endosymbionts: diversity and molecular mechanisms. Sex Dev *8* , 59-73.

Ma, Y., Liu, M., Stiller, J., and Liu, C. (2019). A pan-transcriptome analysis shows that disease resistance genes have undergone more selection pressure during barley domestication. BMC Genomics *20* , 12.

Ma, Y., Liu, S., Gao, J., Chen, C., Zhang, X., Yuan, H., Chen, Z., Yin, X., Sun, C., Mao, Y.*, et al.* (2021). Genome-wide analysis of pseudogenes reveals HBBP1's human-specific essentiality in erythropoiesis and implication in β-thalassemia. Dev Cell *56* , 478-493.e411.

MacGillivray, M.H., and Mazur, T. (2005). Intersex. Adv Pediatr*52* , 295-319.

Mackay, D.J.G., and Temple, I.K. (2017). Human imprinting disorders: Principles, practice, problems and progress. Eur J Med Genet *60* , 618-626.

Madgwick, P.G., Belcher, L.J., and Wolf, J.B. (2019). Greenbeard genes: theory and reality. Trends Ecol Evol *34* , 1092-1103.

Makova, K.D., and Hardison, R.C. (2015). The effects of chromatin organization on variation in mutation rates in the genome. Nature reviews. Genetics *16* , 213-223.

Mancebo Quintana, J.M., and Mancebo Quintana, S. (2012). A short-term advantage for syngamy in the origin of eukaryotic sex: effects of cell fusion on cell cycle duration and other effects related to the duration of the cell cycle-relationship between cell growth curve and the optimal size of the species, and circadian cell cycle in photosynthetic unicellular orrganisms. Int J Evol Biol *2012* , 746825.

Manchia, M., Cullis, J., Turecki, G., Rouleau, G.A., Uher, R., and Alda, M. (2013). The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PloS one *8* , e76295.

Manders, F., van Boxtel, R., and Middlekamp, S. (2021). The dynamics of somatic mutagenesis during life in humans. Frontiers in Aging *2:802407* .

Marras, C., Canning, C.G., and Goldman, S.M. (2019). Environment, lifestyle, and Parkinson's disease: Implications for prevention in the next decade. Mov Disord *34* , 801-811.

McColgan, P., and Tabrizi, S.J. (2018). Huntington's disease: a clinical review. Eur J Neurol *25* , 24-34.

McLaughlin, R.N., Jr., and Malik, H.S. (2017). Genetic conflicts: the usual suspects and beyond. J Exp Biol *220* , 6-17.

Merchant-Larios, H., and Díaz-Hernández, V. (2013). Environmental sex determination mechanisms in reptiles. Sex Dev *7* , 95-103.

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., *et al.* (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell *151* , 1431-1442.

Michieletto, D., Chiang, M., Colì, D., Papantonis, A., Orlandini, E., Cook, P.R., and Marenduzzo, D. (2018). Shaping epigenetic memory via genomic bookmarking. Nucleic acids research *46* , 83-93.

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. Nat Commun *8* , 15183.

Milligan, M.J., and Lipovich, L. (2014). Pseudogene-derived lncRNAs: emerging regulators of gene expression. Frontiers in genetics *5* , 476.

Mizuguchi, T., Nakashima, M., Kato, M., Okamoto, N., Kurahashi, H., Ekhilevitch, N., Shiina, M., Nishimura, G., Shibata, T., Matsuo, M., *et al.* (2018). Loss-of-function and gain-of-function mutations in PPP3CA cause two distinct disorders. Human molecular genetics *27* , 1421-1433.

Monk, D., Mackay, D.J.G., Eggermann, T., Maher, E.R., and Riccio, A. (2019). Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. Nature reviews. Genetics *20* , 235-248.

Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., *et al.*(2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583* , 699-710.

Murúa, P., Edrada-Ebel, R., Muñoz, L., Soldatou, S., Legrave, N., Müller, D.G., Patiño, D.J., van West, P., Küpper, F.C., Westermeier, R., *et al.* (2020). Morphological, genotypic and metabolomic signatures confirm interfamilial hybridization between the ubiquitous kelps Macrocystis (Arthrothamnaceae) and Lessonia (Lessoniaceae). Scientific reports *10* , 8279.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G., and Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science (New York, N.Y.) *327* , 876-879.

Nakamura, J., and Swenberg, J.A. (1999). Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. Cancer Res *59* , 2522-2526.

Narita, S., Pereira, R.A.S., Kjellberg, F., and Kageyama, D. (2010). Gynandromorphs and intersexes: potential to understand the mechanism of sex determination in arthropods. Terrestrial Arthropod Reviews *3* , 63-96.

Nesta, A.V., Tafur, D., and Beck, C.R. (2021). Hotspots of human mutation. Trends in genetics : TIG *37* , 717-729.

36

Nguyen, D.K., and Disteche, C.M. (2006). Dosage compensation of the active X chromosome in mammals. Nature genetics *38* , 47-53.

Niroula, A., Urolagin, S., and Vihinen, M. (2015). PON-P2: Prediction method for fast and reliable identification of harmful variants. PloS one *10(2):e0117380* .

Niroula, A., and Vihinen, M. (2015). Harmful somatic amino acid substitutions affect key pathways in cancers. BMC medical genomics*8* , 53.

Niroula, A., and Vihinen, M. (2016a). PON-mt-tRNA: a multifactorial probability-based method for classification of mitochondrial tRNA variations. Nucleic acids research *44* , 2020-2027.

Niroula, A., and Vihinen, M. (2016b). Variation interpretation predictors: Principles, types, performance, and choice. Human mutation*37* , 579-597.

Niroula, A., and Vihinen, M. (2017). Predicting severity of disease-causing variants. Human mutation *38* , 357-364.

Nissanka, N., and Moraes, C.T. (2020). Mitochondrial DNA heteroplasmy in disease and targeted nuclease-based therapeutic approaches. EMBO Rep*21* , e49612.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A.*, et al.* (2022). The complete sequence of a human genome. Science (New York, N.Y.) *376* , 44-53.

Ollila, J., Lappalainen, I., and Vihinen, M. (1996). Sequence specificity in CpG mutation hotspots. FEBS Lett *396* , 119-122.

Paaby, A.B., and Rockman, M.V. (2013). The many faces of pleiotropy. Trends in genetics : TIG *29* , 66-73.

Pack, L.R., Daigh, L.H., and Meyer, T. (2019). Putting the brakes on the cell cycle: mechanisms of cellular growth arrest. Curr Opin Cell Biol*60* , 106-113.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics *40* , 1413-1415.

Parisod, C., Holderegger, R., and Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. New Phytol *186* , 5-17.

Park, S.H., Ayoub, A., Lee, Y.T., Xu, J., Kim, H., Zheng, W., Zhang, B., Sha, L., An, S., Zhang, Y.*, et al.* (2019). Cryo-EM structure of the human MLL1 core complex bound to the nucleosome. Nat Commun*10* , 5540.

Patch, C., and Middleton, A. (2018). Genetic counselling in the era of genomic medicine. Br Med Bull *126* , 27-36.

Paulson, H. (2018). Repeat expansion diseases. Handb Clin Neurol*147* , 105-123.

Payne, S. (2017). Introduction to RNA viruses. Viruses, 97-105.

Pikaard, C.S. (2018). Nucleolar dominance. In Encyclopedia of Life Sciences (Chischester: John Wiley & Sons).

Pike, V.L., Cornwallis, C.K., and Griffin, A.S. (2021). Why don't all animals avoid inbreeding? Proc Biol Sci *288* , 20211045.

Pilzecker, B., Buoninfante, O.A., and Jacobs, H. (2019). DNA damage tolerance in stem cells, ageing, mutagenesis, disease and cancer therapy. Nucleic acids research *47* , 7163-7181.

Portin, P., and Wilkins, A. (2017). The evolving definition of the term "gene". Genetics *205* , 1353-1364.

Pusey, A., and Wolf, M. (1996). Inbreeding avoidance in animals. Trends Ecol Evol *11* , 201-206.

Rafaluk-Mohr, C., Gerth, M., Sealey, J.E., Ekroth, A.K.E., Aboobaker, A.A., Kloock, A., and King, K.C. (2022). Microbial protection favors parasite tolerance and alters host-parasite coevolutionary dynamics. Curr Biol.

Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Turki, S.A., Dominiczak, A., Morris, A., Porteous, D., Smith, B., *et al.* (2016). Timing, rates and spectra of human germline mutation. Nature genetics *48*, 126-133.

Rancati, G., Moffat, J., Typas, A., and Pavelka, N. (2018). Emerging and evolving concepts in gene essentiality. Nature reviews. Genetics *19*, 34-49.

Raraigh, K.S., Han, S.T., Davis, E., Evans, T.A., Pellicore, M.J., McCague, A.F., Joynt, A.T., Lu, Z., Atalar, M., Sharma, N., *et al.* (2018). Functional assays are essential for interpretation of missense variants associated with variable expressivity. Am J Hum Genet *102*, 1062-1077.

Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nature reviews. Genetics *16*, 653-664.

Reissmann, M., and Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. Semin Cell Dev Biol *24*, 576-586.

Ren, D., Song, J., Ni, M., Kang, L., and Guo, W. (2020). Regulatory mechanisms of cell polyploidy in insects. Front Cell Dev Biol *8*, 361.

Renner, S.S. (2014). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot *101*, 1588-1596.

Ribeiro, A.J.M., Tyzack, J.D., Borkakoti, N., Holliday, G.L., and Thornton, J.M. (2020). A global analysis of function and conservation of catalytic residues in enzymes. The Journal of biological chemistry *295*, 314-324.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405-423.

Roos, W.P., and Kaina, B. (2013). DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. Cancer Lett *332*, 237-248.

Rossi, A., and Kontarakis, Z. (2021). Beyond Mendelian inheritance: genetic buffering and phenotype variability. Phenomics.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., *et al.* (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210-215.

Roucou, X., Gains, M., and LeBlanc, A.C. (2004). Neuroprotective functions of prion protein. J Neurosci Res *75*, 153-161.

Rzeszutek, I., Maurer-Alcalá, X.X., and Nowacki, M. (2020). Programmed genome rearrangements in ciliates. Cellular and molecular life sciences : CMLS *77*, 4615-4629.

Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J., and Zhernakova, A. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. Nature genetics *54*, 100-106.

Sazonova, M.A., Sinyov, V.V., Ryzhkova, A.I., Galitsyna, E.V., Melnichenko, A.A., Postnov, A.Y., Orekhov, A.N., and Sobenin, I.A. (2018). Cybrid models of pathological cell processes in different diseases. Oxid Med Cell Longev *2018*, 4647214.

Scandling, J.D., Busque, S., Lowsky, R., Shizuru, J., Shori, A., Engleman, E., Jensen, K., and Strober, S. (2018). Macrochimerism and clinical transplant tolerance. Hum Immunol *79*, 266-271.

Schaafsma, G.C.P., and Vihinen, M. (2017). Large differences in proportions of harmful and benign amino acid substitutions between proteins and diseases. Human mutation *38* , 839-848.

Schumacher, B., Pothof, J., Vijg, J., and Hoeijmakers, J.H.J. (2021). The central role of DNA damage in the ageing process. Nature *592* , 695-703.

Ségurel, L., Wyman, M.J., and Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet *15* , 47-70.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. PLoS biology *14* , e1002533.

Sharma, R., Lewis, S., and Wlodarski, M.W. (2020). DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. Front Pediatr *8* , 570084.

Sharp, N.P., Sandell, L., James, C.G., and Otto, S.P. (2018). The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. Proceedings of the National Academy of Sciences of the United States of America *115* , E5046-e5055.

Shastry, B.S. (2002). SNP alleles in human disease and evolution. J Hum Genet *47* , 561-566.

Sherman, R.M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M.P., Chavan, S., Vergara, C., Ortega, V.E.*, et al.*(2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nature genetics *51* , 30-35.

Sherman, R.M., and Salzberg, S.L. (2020). Pan-genomics in the human genome era. Nature reviews. Genetics *21* , 243-254.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic acids research *29* , 308-311.

Singh, P.P., Arora, J., and Isambert, H. (2015). Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. PLoS computational biology *11* , e1004394.

Singh, P.P., and Isambert, H. (2020). OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. Nucleic acids research *48* , D724-d730.

Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic studies. Cell *177* , 26-31.

Skuse, D., Printzlau, F., and Wolstencroft, J. (2018). Sex chromosome aneuploidies. Handb Clin Neurol *147* , 355-376.

Smith, L.M., and Kelleher, N.L. (2013). Proteoform: a single term describing protein complexity. Nat Methods *10* , 186-187.

Smith, T., Ho, G., Christodoulou, J., Price, E.A., Onadim, Z., Gauthier-Villars, M., Dehainault, C., Houdayer, C., Parfait, B., van Minkelen, R.*, et al.* (2016). Extensive variation in the mutation rate between and within human genes associated with Mendelian disease. Human mutation *37* , 488-494.

Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nature reviews. Cancer *18* , 696-705.

Soukup, A.A., Zheng, Y., Mehta, C., Wu, J., Liu, P., Cao, M., Hofmann, I., Zhou, Y., Zhang, J., Johnson, K.D.*, et al.* (2019). Single-nucleotide human disease mutation inactivates a blood-regenerative GATA2 enhancer. J Clin Invest *129* , 1180-1192.

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc Lond B

Biol Sci *372* .

Stewart, J.B., and Chinnery, P.F. (2021). Extreme heterogeneity of human mitochondrial DNA from organelles to populations. Nature reviews. Genetics *22* , 106-118.

Stoddard, B.L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA *5* , 7.

Strauss, B.S. (2016). Mechanisms of mutation. In Genetic Diagnosis of Endocrine Disorders, R.E. Weiss, and S. Refetoff, eds. (Elsevier), pp. 3-18.

Strich, J.R., and Chertow, D.S. (2019). CRISPR-Cas biology and its application to infectious diseases. J Clin Microbiol *57* .

Tate, J.A., Soltis, D.E., and Soltis, P.A. (2005). Polyploidy in plants. In The Evolution of the Genome, T.R. Gregory, ed. (Elsevier), pp. 371-426.

Terekhanova, N.V., Seplyarskiy, V.B., Soldatov, R.A., and Bazykin, G.A. (2017). Evolution of local mutation rate and its determinants. Mol Biol Evol *34* , 1100-1109.

Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capella, G., den Dunnen, J.T.*, et al.* (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nature genetics*46* , 107-115.

Thünken, T., Bakker, T.C., Baldauf, S.A., and Kullmann, H. (2007). Active inbreeding in a cichlid fish and its adaptive significance. Curr Biol *17* , 225-229.

Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. Nature *423* , 74-77.

Tong, T., Chen, S., Wang, L., Tang, Y., Ryu, J.Y., Jiang, S., Wu, X., Chen, C., Luo, J., Deng, Z.*, et al.* (2018). Occurrence, evolution, and functions of DNA phosphorothioate epigenetics in bacteria. Proceedings of the National Academy of Sciences of the United States of America *115* , E2988-e2996.

Touzet, P., and Budar, F. (2004). Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? Trends Plant Sci *9* , 568-570.

Tress, M.L., Abascal, F., and Valencia, A. (2017). Alternative splicing may not be the key to proteome complexity. Trends in biochemical sciences *42* , 98-110.

Turner, K.J., Vasu, V., and Griffin, D.K. (2019). Telomere biology and human phenotype. Cells *8* .

Urbanowicz, R.J., Granizo-Mackenzie, A.L., Kiralis, J., and Moore, J.H. (2014). A classification and characterization of two-locus, pure, strict, epistatic models for simulation and detection. BioData Min *7* , 8.

Valdes, A.M., Walter, J., Segal, E., and Spector, T.D. (2018). Role of the gut microbiota in nutrition and health. Bmj *361* , k2179.

Väliaho, J., Faisal, I., Ortutay, C., Smith, C.I.E., and Vihinen, M. (2015). Characterization of all possible single nucleotide change –caused amino acid substitutions in the kinase domain of Bruton tyrosine kinase. Human mutation *36* , 638-647.

van der Gaag, M., Debets, A.J., Oosterhof, J., Slakhorst, M., Thijssen, J.A., and Hoekstra, R.F. (2000). Spore-killing meiotic drive factors in a natural population of the fungus Podospora anserina. Genetics*156* , 593-605.

Vetrie, D., Vořechovský, Sideras, P., Holland, J., Davies, A., Flinter, F., Hammarström, L., Kinnon, C., Levinsky, R., Bobrow, M., and et al. (1993). The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature *361* , 226-233.

Vihinen, M. (2014a). Contribution of pseudogenes to sequence diversity. Methods in molecular biology (Clifton, N.J.) *1167* , 15-24.

Vihinen, M. (2014b). Variation Ontology for annotation of variation effects and mechanisms. Genome research *24* , 356-364.

Vihinen, M. (2015a). Muddled genetic terms miss and mess the message. Trends in genetics : TIG *31* , 423-425.

Vihinen, M. (2015b). Types and effects of protein variations. Hum Genet*134* , 405-421.

Vihinen, M. (2017). How to define pathogenicity, health, and disease? Human mutation *38* , 129-136.

Vihinen, M. (2018). Systematics for types and effects of DNA variations. BMC Genomics *19* , 974.

Vihinen, M. (2020a). Poikilosis - pervasive biological variation. F1000Res *9* , 602.

Vihinen, M. (2020b). Problems in variation interpretation guidelines and in their implementation in computational tools. Mol Genet Genomic Med*8* , e1206.

Vihinen, M. (2020c). Strategy for disease diagnosis, progression prediction, risk group stratification and teatment - Case of COVID-19. Front Med (Lausanne) *7* , 294.

Vihinen, M. (2021a). Functional effects of protein variants. Biochimie*180* , 104-120.

Vihinen, M. (2021b). Systematics for types and effects of RNA variations. RNA biology *18* , 481-498.

Vihinen, M. (2022). Generic model for biological regulation. F1000Research, 11:419.

Vijg, J., and Dong, X. (2020). Pathogenic mechanisms of somatic mutation and genome mosaicism in aging. Cell *182* , 12-23.

Vogan, A.A., Martinossi-Allibert, I., Ament-Velásquez, S.L., Svedberg, J., and Johannesson, H. (2022). The spore killers, fungal meiotic driver elements. Mycologia *114* , 1-23.

Wang, C., and Lu, X. (2011). Female ground tits prefer relatives as extra-pair partners: driven by kin-selection? Mol Ecol *20* , 2851-2863.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science (New York, N.Y.) *350* , 1096-1101.

Waters, K. (2013). Molecular genetics. In The Stanford Encyclopedia of Philosophy (Fall 2013 Edition), E.N. Zalta, ed. (https://plato.stanford.edu/archives/fall2013/entries/molecular-genetics/, Metaphysics Research Lab, Stanford University).

Wedell, N. (2020). Selfish genes and sexual selection: the impact of genomic parasites on host reproduction. J. Zoology *311* , 1-12.

Weeks, S.C. (2012). The role of androdioecy and gynodioecy in mediating evolutionary transitions between dioecy and hermaphroditism in the animalia. Evolution *66* , 3670-3686.

Weichenhan, D., and Plass, C. (2013). The evolving epigenome. Human molecular genetics *22* , R1-6.

Werren, J.H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. Proceedings of the National Academy of Sciences of the United States of America *108 Suppl 2* , 10863-10870.

Wilson, P.G. (2008). Centriole inheritance. Prion *2* , 9-16.

Wu, C.S., Lu, W.H., Hung, M.C., Huang, Y.S., and Chao, H.W. (2022). From polyploidy to polyploidy reversal: its role in normal and disease states. Trends in genetics : TIG.

Wu, J., Greely, H.T., Jaenisch, R., Nakauchi, H., Rossant, J., and Belmonte, J.C. (2016). Stem cells and interspecies chimaeras. Nature *540* , 51-59.

Yang, Y., Shao, A., and Vihinen, M. (2022). PON-All, amino acid substitution tolerance predictor for all organisms. Front Mol Biosci *9:867572* .

Yi, C., and He, C. (2013). DNA repair by reversal of DNA damage. Cold Spring Harb Perspect Biol *5* , a012575.

Yousefzadeh, M., Henpita, C., Vyas, R., Soto-Palma, C., Robbins, P., and Niedernhofer, L. (2021). DNA damage-how and why we age? Elife *10* .

Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature *587* , 610-612.

Zeberg, H., and Pääbo, S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. Proceedings of the National Academy of Sciences of the United States of America *118* .

Zhang, Q., and Sodmergen (2010). Why does biparental plastid inheritance revive in angiosperms? J Plant Res *123* , 201-206.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. EMBO Rep *16* , 1467-1481.

Zhang, Y.B., Pan, X.F., Chen, J., Cao, A., Xia, L., Zhang, Y., Wang, J., Li, H., Liu, G., and Pan, A. (2021). Combined lifestyle factors, all-cause mortality and cardiovascular disease: a systematic review and meta-analysis of prospective cohort studies. J Epidemiol Community Health *75* , 92-99.

Zhu, A.Y., Costain, G., Cytrynbaum, C., Weksberg, R., Cohn, R.D., and Ali, A. (2021). Novel heterozygous variants in PXDN cause different anterior segment dysgenesis phenotypes in monozygotic twins. Ophthalmic Genet *42* , 624-630.

Zhu, X., Need, A.C., Petrovski, S., and Goldstein, D.B. (2014). One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nature neuroscience *17* , 773-781.

Zlotogora, J. (2003). Penetrance and expressivity in the molecular age. Genet Med *5* , 347-352.

Zu, H., and Gao, D. (2021). Non-viral vectors in gene therapy: recent development, challenges, and prospects. Aaps j *23* , 78.

TABLES

Table 1. Numbers of types of DNA lesions per cell per day, data from (Yousefzadeh et al., 2021).

DNA lesion	Frequency per cell per day
Endogeneous	Endogeneous
Cytosine deamination	100
Cyclopurine adduct	100
Depyrimidation	100
Depurination	10 000
8-oxoG	1 000
Monodialdehyde adducts	1 000
Alkylation adducts	1 000
Single strand breaks	10 000
Double strand breaks	10
Environmental	Environmental
Damaged bases	10

DNA lesion	Frequency per cell per day
Photodimers	100
Single strand breaks	2-5
Double strand breaks	0.25

Table 2. Autosomal variant numbers per genome, median values. Data from (Auton et al., 2015). The values indicate ranges in different populations.

Variation	Mean value, range
Variation type	Variation type
SNV	$3.53 - 4.31 \times 10^6$
Insertion or deletion	546 000 – 625 000
Large deletion	939 – 1 100
CNV	153 - 170
Inversion	9 -12
Mobile genetic element	1 012 - 1 222
Variation effect or site	Variation effect or site
Non-synonymous	10 200 - 12 200
Synonymous	11 200 – 13 800
Untranslated region	30 000 – 37 200
Intron	$1.68 - 2.06 \times 10^6$
Transcription factor binding site	748 - 927
Promoter	81 600 – 102 000
Insulator	57 700 - 70 900
Enhancer	288 000 – 354 000

FIGURE LEGENDS

Figure 1. Types of genetic variations. Genomic variation: A) Euploidy, altered number of genome sets. Chromosomal variation: B) Aneuploidy, chromosome number variation, trisomy as an example. C) Chromosomal amplification, D) interstitial chromosomal deletion, E) chromosomal translocations of type reciprocal translocation, F) complex chromosomal variation, G) Immunological recognition molecule diversification by immunological receptor gene rearrangement, H) isochromosome, I) ring chromosome, and J) telomere length variation. Types of DNA segment variation: K) DNA variation types. RNA variation: L) types of RNA variations, M) mRNA forms. Protein variation: N) types of protein variations. Epigenetic variation: O) DNA methylation, cytosine methylation at C5 as example, P) histone modification. Three-dimensional structure of nucleosome, DNA wound around histones (PDB code 6pwv (Park et al., 2019)). DNA is in cyan, histones H2A orange, H2B gray, H3 green, H4 yellow. Post translational modification sites in blue, the majority of the sites are not shown as the flexible termini are not included to the structure. Adapted from (Vihinen, 2015b, 2018, 2021b).

Figure 2. Epigenetic phenomena and processes. Data are presented at different levels from DNA chain to organism level.

Figure 3. Types and classification of selfish genetic elements. The three major types are over-replicating elements, biased gene converters and transmission distorters, each with more specific sublevels.

Figure 4. Sexual reproduction. A) Generation of sex haploid sex chromosomes during meiosis. Genetic recombination between chromosomes can occur during the process. B) The gametes are joined in fertilization. Sex determination of offsprng C) in XX/XY system. The homogametic sex can be also male, in ZW/ZZ

system. D) In XX/X0 system, only females have two sex chromosomes. The homogametic sex is male in ZZ/Z0 system. E) Haplodiploidy, diploid females develop from fertilized eggs and haploid males from unfertilized eggs. For simplicity, only one pair of chromosomes is shown.

Figure 5. Functional effects of DNA. Gene dosage describes differences in the number of gene copies. Abundance of genetic element is due to factors in gene expression, regulation, dosage compensation etc. Alteration of information class is for different variations and changes to the contents of genetic message. Effects to cell fate ranges from normal cell division to programmed cell death. In addition to the sex of the progeny and its possible change, sex determination category contains also effects on the progeny viability, fertility and hermaphroditism.

Figure 6. TARAR countermeasures for DNA variations.

Figure 7. Genetic variation, phenotypic expressions and their contribution to diseases. Genetic variations set the bases for phenotypic expression, which are then modified and affected by several factors and processes depending on the phenotypic type. Together with life style and environmental factors they contribute to diseases. The disease is described by its extent, modulation and severity.

Genomic
A

Chromosomal
B

C

D

E

F

G

H

I

J

DNA segment
K

ATGCAGTC
Deletion

ATGCCCAGTC
Insertion

AATCCAGTC
Indel

ATGCCAGTC

ATGCGACTC
Translocation
- inversion

ATGCCTGTC
Substitution
- transition
- transversion

RNA
L

AUGCAUAA
Deletion
- in-frame
- out-of-frame

AUGCCCAUAA
Insertion
- in-frame
- out-of-frame

AUGGCAUAA
Substitution
- missense
- nonsense
- synonymous
- unsense

AUACGUUAA
Indel
- in-frame
-out-of-frame

Missing RNA

AUGCCAUAA → ACGCCAUAA
Initiation codon
change

AACCGUUAUAA
Translocation
- inversion

AUGCCAAAA
Termination codon
change

M

Alternatively initiated

Alternatively terminated

Alternatively polyadenylated

Modified

Alternatively spliced

cis splicing

trans splicing

Constitutively spliced

Loss of exon fragment

Inclusion of intron fragment

Exon skipping

RNA chimera

Protein
N

AASEKQWENCE
Insertion
- amphigoric
- sequence
retaining

AASEQNCE
Deletion
- truncation
- sequence
retaining

AASENSELESS
Indel
-amphigoric
- sequence
retaining

AASEQWENCE

Missing protein

AASEKWENCE
Substitution

Epigenetic
O

P

Epigenetic phenomenon

Organism
    X-inactivation (lyonization)
    imprinting
    bookmarking
    paramutation
    parental effect
    transvection
    position effect variegation

Cell
    epigenetic regulation of gene expression

Genome
    epigenetic editing

Chromosome
    chromatin remodeling

DNA chain
    epigenetic DNA modification

Protein
    epigenetic protein modification
    proteinaceous infection
    protein structural inheritance

Epigenetic mechanism

DNA modification
gene translocation within chromosome
allele interaction
uniparental inheritance of organelles

gene dosage compensation

CRISPR-Cas9, TALEN, ZFN

DNA or histone modification

DNA base modification
DNA sugar modification

histone modification
infectious structural alteration

Epigenetic mechanisms and processes

DNA chain epigenetic mechanism
    epigenetic DNA modification
        epigenetic DNA base modification
        epigenetic DNA sugar modification

Protein epigenetic mechanism and phenomenon
    histone modification
    proteinaceous infection
    protein structural inheritance

Chromosome epigenetic mechanism
    chromatin remodeling

Genome epigenetic phenomenon
    epigenetic editing

Cell epigenetic phenomenon
    epigenetic gene expression regulation
    gene dosage compensation

Organism epigenetic phenomenon
    X-inactivation (lyonization)
    imprinting
    bookmarking
    paramutation
    parental effect
        maternal effect
        paternal effect
    transvection
    position effect variegation

A

Recombination   Meiosis   Fertilisation

B

C

Parents ♀ XX + XY ♂

Gametes   X X   X Y

Zygote   XX   XY

D

♀ XX + X0 ♂

XX   X

XX   X

E

♀   +   ♂

Gene dosage
euploidy
aneuploidy
structural variation
selfish genetic element
haplodiploidy
programmed genomic rearrangement
paternal genomic elimination
chromatin diminution
uniparental disomy
location in X chromosome

Sex determination
sex of progeny
intersex
sex change
hermaphroditism
viability
fertility

**Abundance of
genetic element**
transcription
translation
transcription factor binding
enhancer, silencer, regulator
inheritance
somatic mosaicism
missing RNA/protein
aberrant splicing
non-Mendelian dominance
polygenic inheritance
dosage compensation
tropopeissis
polypleyri
extranuclear inheritance
repeat sequence inheritance

Functional
effects

Cell fate
cell division
quiescence
senescence
programmed cell
death
uncontrolled growth

**Alteration of
information**
variation
hybrid
microchimerism
somatic mosaicism
genetic recombination
genome editing
ribosomal frameshifting
alternative start site
alternative stop site
epigenetic changes

48

Tolerance
neutral variation
variation tolerance

Avoidance
inbreeding
avoidance

TARAR
counter-
measures

Resistance
cell cycle arrest
pan genome
DNA damage
  response
programmed cell
  death

Repair
proofreading
direct reversal
excision repair
break correction
DNA damage
  tolerance
telomere healing

Attenuation
redundancy
genetic suppression
genetic interaction
number of sexes
hermaphroditism
sex change
protective allele
gene balance

Genetic variation

Species	Cell	Gene/genome		Variation	

- unmixed
- mixed
	type
	- germline
	- somatic
	variation distribution
	- uniform
	- somatic mosaicism
		species origin
		- (micro)chimerism
		- microbiota
		genetic origin
		- nuclear
		- organellar
		- cytoplasmic
		type
		- ausosomal
		- sex chromosomal
			dosage
			- euploidy
			- aneuploidy
			- structural variation
			- selfish element
			- uniparental disomy
			- haplodiploidy
			- sex chromosome
			- chromatin
			  diminution
			- programmed
			  genome rearrange-
			  ment
			- somatic mosaicism
				type
				size
				location
					genetic origin
					- inherited
					- somatic
					- *de novo*
					- microbiota
					non-genetic origin
					- artificial
					- edited
					- epigenetic
					- modified

Phenotypic expression

Sex determination	Inheritance mode	Genetic conflict	Cell fate	Functional effect	TARAR	Abundance
type - genetic - environmental - mitochondrial - cytoplasmic sex change hermaphroditism	- Mendelian - non-Mendelian	- origin conflict - destination conflict	- division - senescence - quiescence - cell death	- no effect - loss-of-function - gain-of-function - neomorphic - antagonistic	- tolerance - avoidance - repair - attenuation - resistance	- monoallelic expression - biallelic expression - regulation - constitutive expression - transcription - transcript quality control - epigenetics - inheritance

Life style   Environmental effects

Disease

Extent	Modulation	Severity