BIOMEDICAL JOURNAL 44 (2021) 548–559

Available online at www.sciencedirect.com

ScienceDirect

Biomedical Journal

journal homepage: www.elsevier.com/locate/bj

Review Article

DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects

Ondrej Pös [a,b], Jan Radvanszky [b,c,d,**], Gergely Buglyó [e], Zuzana Pös [a,b,c], Diana Rusnakova [a,b], Bálint Nagy [e,*], Tomas Szemes [a,b,d]

[a] Department of Molecular Biology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia
[b] Geneton s.r.o., Bratislava, Slovakia
[c] Institute of Clinical and Translational Research, Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
[d] Comenius University Science Park, Bratislava, Slovakia
[e] Department of Human Genetics, Faculty of Medicine, University of Debrecen, Debrecen, Hungary


Dr.Jan Radvanszky


Dr.BALINT NAGY

ARTICLE INFO

Article history:
Received 27 November 2020
Accepted 5 February 2021
Available online 13 February 2021

Keywords:
Copy number variants
Structural variation
Human genome
CNV formation
Evolution
Genetic diseases

ABSTRACT

Copy number variants (CNVs) were the subject of extensive research in the past years. They are common features of the human genome that play an important role in evolution, contribute to population diversity, development of certain diseases, and influence host–microbiome interactions. CNVs have found application in the molecular diagnosis of many diseases and in non-invasive prenatal care, but their full potential is only emerging. CNVs are expected to have a tremendous impact on screening, diagnosis, prognosis, and monitoring of several disorders, including cancer and cardiovascular disease. Here, we comprehensively review basic definitions of the term CNV, outline mechanisms and factors involved in CNV formation, and discuss their evolutionary and pathological aspects. We suggest a need for better defined distinguishing criteria and boundaries between known types of CNVs.

Copy number variation (CNV) is a general term used to describe a molecular phenomenon in which sequences of the genome are repeated, and the number of repeats varies between individuals of the same species. Biological roles of resulting copy number variants (CNVs) range from seemingly no effect on common variability of physiological traits [1], through morphological variation [2,3], altered metabolic states [4], susceptibility to infectious diseases [5,6], and host–microbiome interactions [7–9], to a substantial contribution to common and rare genetic disorders/syndromes [10].

* *Corresponding author.* Department of Human Genetics, Faculty of Medicine, University of Debrecen, Debrecen, Egyetem tér 1, Hungary.
** *Corresponding author.* Institute for Clinical and Translational Research, Biomedical Research Centre Slovak Academy of Sciences, Dubravska cesta 9, 845 05, Bratislava, Slovakia.
E-mail addresses: jradvanszky@gmail.com (J. Radvanszky), nagy.balint@med.unideb.hu (B. Nagy).
Peer review under responsibility of Chang Gung University.
https://doi.org/10.1016/j.bj.2021.02.003

As such, they have a high potential to contribute to human population diversity [11] and also to micro- and macro-evolutionary processes [12]. In addition to their biological roles, their presence in our genomes may have several technical implications in biomedicine, either as biomarkers for certain pathological processes such as cancer, as biomarkers of environmental exposures such as radiation [13], or even as potential confounding factors when evaluating results of certain genetic diagnostic tests [14]. While it is not yet well described how many genes are absent from human reference genomes, approximately 100 genes were found to be homozygously deleted from the genomes of human individuals without causing apparent phenotypic consequences, likely due to the presence of redundant paralogs, the genes being limited to causing age-related phenotypes or being relevant only under certain environmental or physiological conditions [15]. These findings suggest that the field of pangenomics may open its doors to human pan-genomes [16] in addition to more commonly mentioned bacterial, archaeal, and plant pan-genomes [17]. Much effort has also been made to study the genomes of livestock and domestic animals in the context of CNV-associated, economically important traits [18,19]. These variants overlap genomic regions associated with traits such as feed conversion ratio [20], meat quality [21], milk production [22,23], and animal health [24], thus, may result in significant economic losses due to reduced production and quality of commodities or decreased commercial value of affected animals.

CNVs in the form of large insertions and deletions were reported among the first genetic "mutations" ever [25], well before the description of DNA structure and the birth of molecular biology. When searching PubMed for the term "copy number variation", the returned list contains 4759 results in a date-range from 1983 to 2020, while limiting the search to "humans" produces 3047 results dated between 1991 and 2020. As visible in publication timelines [Fig. 1], a sharp increase in publication numbers began around the year 2005, partly due to the term "copy number variation" getting widely adopted, but also to the research interest that CNVs started to gain at the time, resulting in rapidly accumulating knowledge in the field. Whether plateauing publication numbers following the year 2014 anticipate a permanent trend is not predictable at the moment, however, this does seem to suggest that CNVs are being readily prepared for routine biomedical applications. One may argue that clinical applications of CNV detection, particularly in the field of oncogenetics and severe congenital anomalies, were among the first in the history of genetic testing and are still in general use. However, considering the technical innovations now allowing population-scale genome-wide CNV screening, together with the still widening spectrum of known biological roles of CNVs, we have yet to take full advantage of CNV detection in routine clinical care [26]. In favor of this point-of-view, official systematic guidelines on the interpretation of CNVs and their classification by clinical impact were issued by the American College of Medical Genetics and Genomics for the first time only in 2019 [27]. Because of their important biological roles, relevant technical impact, and several associated uncertainties, we believe that the field of CNV research and application urgently calls not only for further investigation, but also for regular review of available knowledge and for constant revision of relevant definitions and classifications.

In this review, we outline basic definitions of the term CNV and discuss general perception of copy number variants, along with mechanisms and factors involved in their formation, their evolutionary aspects, and their influence on phenotype and disease.

Definitions, perception, and classification of CNVs

The first association of a CNV with a phenotype reported in a non-human species was the case of a reduced-eye mutant *Drosophila melanogaster* having the bar eyes phenotype [25] due to a single duplication of the *Bar* gene [28]. After the "general" human karyotype was established [29], reports of microscopically visible chromosomal aberrations in the human genome emerged as early as the 1960s, when cytogeneticists recognized the genetic background for many disorders, including

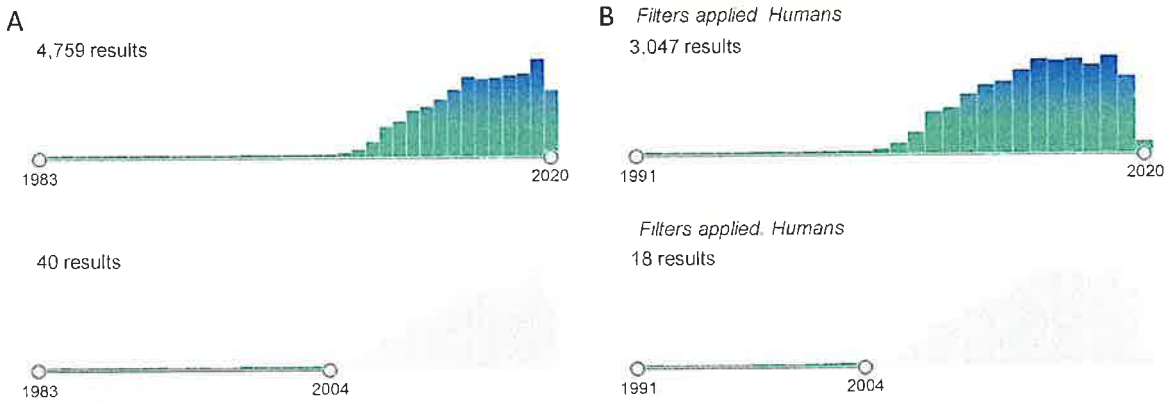

Fig. 1 Publication records. PubMed search for the term "copy number variation" as of July 24, 2020. Limiting results to the year 2004 reduced the number of entries considerably. Excluded entries are highlighted in grey; (A) no filter applied; (B) applying a built-in species-specific filter for "humans".

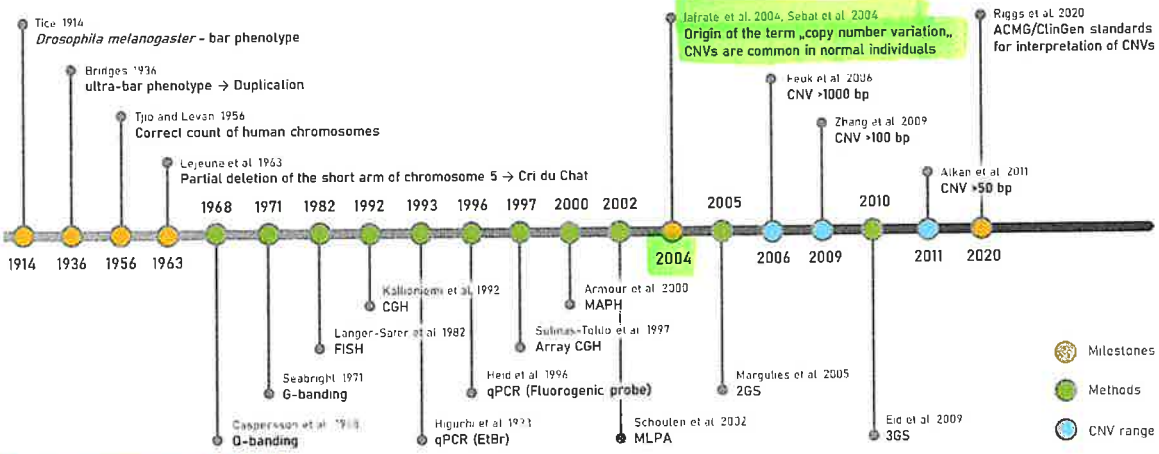

Fig. 2 Timeline of CNV research. It includes research milestones (orange) in the context of evolving methods for the assessment of CNVs (green) and the minimal length of variants to be considered as CNVs at a given time (blue). Main trends of CNV research are visible, e.g., the relatively wide time-frame following the first descriptions of large CNVs around the beginning of the 20th century. This nearly century-long phase was mainly about the development of methods, which finally allowed genome-wide, high-resolution CNV detection around the beginning of the 21st century, leading to the recognition of common features of CNVs and a subsequent shift in their definition, allowing shorter and shorter variants to be considered CNVs.

Cri du chat syndrome associated with a partial deletion of the short arm of chromosome 5 [30]. For decades, due to a limited resolution of microscopy-based cytogenetic methods, reports were mostly limited to large pathogenic abnormalities involving several genes and having prominent adverse effects on physiological processes. Although large structural changes including insertions and deletions were among the first genetic anomalies recognized, the concept of CNV as we perceive it today originates from less than two decades ago [Fig. 2], when large-scale differences between human genomes, previously considered rare mutational events, were shown to be common among normal human individuals, forming a considerable part of our intraspecies physiological variability [31,32]. It is estimated that around 4.8–9.5% of the genome is affected by CNVs [15], a larger fraction than by single nucleotide variants [11].

Although it is generally accepted that CNVs are a subtype of structural genome variants, their definition is still somewhat vague. In addition to basic criteria about the repeating nature and numerical variability among individuals, a "considerable" length is usually also required. Earlier studies defined a CNV as a DNA segment larger than 1000 bp [33], followed by ~100 bp [1], but currently, the size of CNVs is defined from 50 bp [34] to several Mb [35]. Despite CNVs being insertions or deletions, it is also debated in the literature whether a distinction should be made between indels and CNVs based on their size [36]. We did not find suggested criteria for the extent of homology between sequences to be considered CNVs, most likely because commonly used methods other than sequencing are not suitable to determine such a sequence homology. Mechanisms of formation are sometimes considered for distinction criteria, but there is no unambiguous relationship between the length of an insertion or deletion and its mechanism of formation, as models may be shared across different size-ranges [37]. Classification of identified variants is generally a routine procedure with no time or possibility to determine the exact mechanisms behind their formation. A discussion seems justified about whether there is any real reason to restrict the term CNV to variants having a "considerable length" instead of applying it to all variants meeting core criteria with no regard to molecular mechanisms or to the number and length of repeated elements.

For the reasons above, we favor the definition of "copy number variants" used for unbalanced structural rearrangements of the genome, which lead to variability, i.e., relative difference in copy numbers of particular DNA sequences among individuals of the same or distinct populations of a species. In line with this, when considering the definition in the light of the fact that there is no such thing as a "standard genome" [38], the term "copy number variant" is best applied in a general sense to describe variants contributing to "copy number variation" or "copy number variability", while duplications, insertions, and deletions may be considered as their particular molecular phenotypes. Terms such as "insertion", "deletion", "gain" or "loss" are most suitable to describe relative differences: i) against an artificial reference genome when the variant is *de novo*, or even if it is present in the population and was inherited from an ancestor with no way to determine the original copy number; ii) against a parental genome, when a *de novo* variant occurs in an offspring; or iii) against a germline genome of the individual, when a *de novo* somatic variant occurs; in each case with an aim to characterize the type of CNV more precisely.

Complying with these arguments, the list of variants belonging to CNVs can be extended (and is already extended by some authors) with several other variant types. On one end of the spectrum, we find changes involving entire sets of chromosomes and aneuploidies (i.e., numerical abnormalities defined as a loss or gain of an entire chromosome), as well as structural abnormalities represented as non-balanced chromosome rearrangements [39]. The opposite end of the

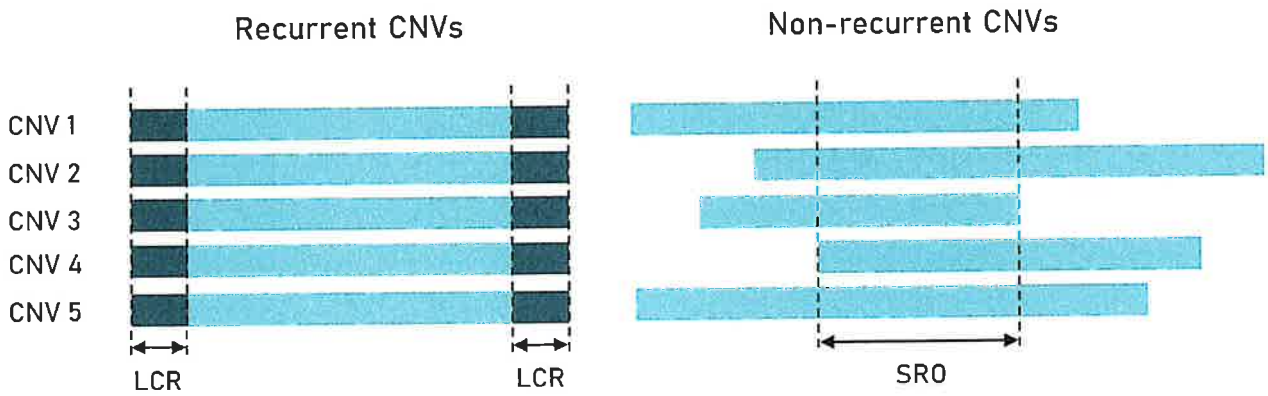

Fig. 3 Recurrent vs. non-recurrent rearrangements. Recurrent CNVs have the same size and common breakpoints enriched in low-copy repeats (LCRs). Non-recurrent CNVs with different sizes may share the smallest region of overlap (SRO). CNVs may occur as inherited or *de novo* events. Inherited CNVs are not recurrent events but always share the same breakpoints, resulting in a similar phenotypic effect. *De novo* CNVs are independently arising events that may have recurrent or distinct (non-recurrent) breakpoints. Even CNVs with non-recurrent breakpoints may show overlapping effects, as they disrupt the same SRO.

spectrum may range from interspersed and tandem repeats, including microsatellites, minisatellites, and macrosatellites, to few nucleotide insertions and deletions. Perhaps the most common and ancient benign CNV is one that is usually not listed as a CNV at all: the XY sex-determination system in humans and many other species, maintained through a strong balancing selection and determining extensive phenotypic differences between individuals of the same population.

Interspersed repeated sequences are generally composed of low-copy repeats, also called segmental duplications, which have greater than 95% sequence identity, and high-copy repeats, which include endogenous retroviruses, retrotransposons, and other transposable or mobile elements. Retrotransposons encompass long terminal repeat (LTR) and non-LTR retrotransposons, comprising of the well-known short interspersed nuclear elements (SINEs, such as the *Alu* element) and long interspersed nuclear elements (LINEs, such as L1). Low and high copy number repeats may act as mediators of the formation of other, larger CNVs, apart from being CNVs themselves.

Although the current classification of CNVs is satisfactory, distinguishing criteria and boundaries between different types of variants are not clearly defined, leading to some variants meeting criteria of more than one category at the same time. For core parts of this review, we focus on the conventional concept of CNVs, but where relevant, we discuss variants more difficult to classify.

Mechanisms of CNV formation

One of the main questions in understanding the biology of CNVs is connected to their origin and mechanisms of formation. To date, several different replication and repair mechanisms have been shown to be involved in the development of CNV. Apart from genomic factors, chemical and physical mutagens also drive their formation.

Genomic factors and mechanisms of CNV formation

CNVs emerge from different mutational mechanisms, including DNA recombination-, replication- and repair-associated processes. Mechanisms of change in gene copy number have been extensively studied through the analysis of CNV breakpoint junction sequences. Repeated sequences, including low-copy repeats (e.g., segmental duplications) and high-copy repeats (e.g., SINEs, LINEs, and endogenous retroviruses), are enriched in the vicinity of breakpoints, thus represent an important factor for CNV instability [40]. Such sequence motifs play a key role in triggering non-allelic homologous recombination (NAHR), one of the general mechanisms involved in the formation of recurrent CNVs [41]. Recurrent rearrangements share a common size, show clustering of breakpoints, and recur in multiple individuals. On the other hand, non-recurrent rearrangements with scattered breakpoints differ in size, but may share the smallest region of overlap among different patients [Fig. 3] [42] and might be formed by several different mechanisms: i) non-replicative mechanisms, including non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ); or ii) replicative mechanisms such as replication slippage, fork stalling and template switching or microhomology-mediated break-induced replication [41].

Short repetitive sequence motifs (e.g., inverted repeats) may adopt non-B DNA structures (e.g., cruciforms) [43] that may result in replication errors inducing CNV formation [40]. Such non-B-DNA forming sequences are also enriched in promoter regions, thus Conrad et al. suggested that the same properties that enable regulation of transcription may also be mutagenic for the formation of CNVs [44]. Consequently, CNVs may influence the evolution of gene regulation.

In addition to accurate replication of DNA during cell division, its equal distribution into daughter cells is essential in genome maintenance, specifically in ensuring the balance of genetic content in eukaryotic cells [45]. However, errors such

as nondisjunction, merotelic kinetochore-microtubule attachment, monopolar and multipolar spindle, and telomerase dysfunction may occur during this high-precision process. Nondisjunction is caused by weakened or completely inactive mitotic checkpoints. If a chromosome fails to separate correctly, it results in one daughter cell with a missing copy of the chromosome and the other one having an extra copy. If mitotic checkpoints are completely inactive, it causes the two daughter cells to have a disjoint set of chromosomes. Unbalanced chromosome segregation occurs when one kinetochore is attached to the microtubules emanating from both poles of the spindle [46]. In the case of dysfunction or absence of telomerase, the chromosome ends become uncapped. This condition triggers activation of the NHEJ DNA repair pathway in the cell, leading to a telomere fusion between the two chromosomes and the formation of a dicentric chromosome. Structural aneuploidy may occur by several mechanisms, including incorrect NAHR (the same mechanism that causes CNVs), misalignment of homologous chromosomes and/or unequal crossing over between non-sister chromatids during meiosis [39]. Such structural abnormalities may subsequently affect the development of chromosomal aneuploidies, suggesting a link between them.

When considering tandem repeats, both micro-, mini-, and macrosatellites display notable somatic and intergenerational dynamics leading to copy number changes. Microsatellites are typically altered during DNA repair and replication in mitosis mainly due to slippage by strand misalignment, primarily accounting for small-scale instability [47], however, unequal crossing over was also suggested to cause variable copy numbers of microsatellite motifs [48]. In the case of disease-associated repeats, which may gain up to a few thousands of repeat units in affected individuals, the mechanisms of expansion are different and may depend on the repeat sequence, length and location within a genome as well as the organism or cell type. They likely involve transient DNA secondary structures formed by the repetitive tracts as DNA unwinds for replication, transcription, or repair, during which these secondary structures may get excised or incorporated into the DNA, resulting in a reduction or expansion of the repetitive region [47]. These mechanisms are generally involved in expanding copy numbers over certain thresholds at single loci and do not lead to general microsatellite instability. Mechanisms may further be complemented by an impairment of the mismatch repair system leading to general hypermutability of microsatellite loci throughout the genome, called microsatellite instability, most commonly due to a failure to correct common errors during replication of repetitive DNA sequences [49]. While a link was reported between microsatellite instability and telomere shortening suggesting some association between the DNA mismatch repair system and telomere maintenance mechanisms, at least in colorectal cancer [50], telomeres *per se* may represent atypical mechanisms of CNV generation. Telomeres contain TTAGGG hexanucleotide repeats protecting the genetic content of chromosomes against chromosome shortening during cell division. In an active compensatory mechanism known as telomere length maintenance, telomerase increases copy numbers by adding TTAGGG repeat units to the ends of telomeres [51]. Minisatellites, on the other hand, generally mutate in the germline by complex conversion-like transfer of repeats between alleles. Inter-allelic unequal crossovers also occur between loci, although at low frequency [52].

Environmental factors contributing to the formation of CNVs

The contribution of environmental factors to the origin of CNVs is poorly understood. It was shown that perturbing DNA replication by chemical mutagens (e.g., aphidicolin or hydroxyurea) results in replication stress that induces the formation of CNVs across the human genome. Replication fork delay and subsequent error-prone repair seem to be important mechanisms in the process [53]. Hastings et al. proposed that under stress conditions, the repair of broken replication forks switches from high-fidelity homologous recombinational repair to non-homologous repair promoting CNVs [41]. Physical mutagens such as low-dose ionizing radiation also effectively induce *de novo* CNV mutations via a replication-dependent mechanism, since radiation-induced DNA strand breaks may collapse the replication fork [13]. It was shown that radiation induces duplications and deletions equally, while chemical mutagens induce copy number losses at a higher frequency compared to gains. On another note, radiation impacts random loci of the genome, while chemical mutagenesis targets specific genomic regions [54].

Costa et al. showed a 1.5-fold increase in the germline CNV mutation burden in offspring of parents accidently exposed to a low dose of ionizing radiation. The mutation rate of *de novo* CNVs was proven to be a biomarker of parental exposure useful in retrospective studies of human populations exposed to low doses of ionizing radiation [55]. It was also demonstrated that irradiation with laser-driven electrons may induce CNVs in human leukocytes *in vitro*. These CNVs are usually duplications or amplifications and tend to inversely correlate with chromosome size and gene density. CNVs may persist in the cell population as stable chromosomal changes for several days after exposure [56].

Evolutionary aspects of CNVs

As excessively static genomes are not able to conform to an ever-changing environment, genome plasticity driven by insertion sequence elements, transposons and integrons, together with DNA rearrangements, may determine whether an organism can survive changing conditions and compete for the resources it needs to reproduce, at least in prokaryotes. However, the main players of prokaryote genome plasticity, including CNV-associated mechanisms mentioned in the previous section, are also active in humans. Although changes to the genome, and thus to the phenotype, may threaten the ability to survive, gaining new phenotypes may also enhance chances of surviving even in previously disadvantageous environments [57]. Both are possible scenarios in the context of

the extent of change and environmental pressure, which define whether individuals fit the actual environment.

CNVs, like other changes in DNA, create phenotypes in which descendants slightly or largely differ from their predecessors. Small differences in phenotypes are generally considered normal variation, especially when relatively common in the population with small changes in physiology, i.e., benign traits or even common diseases and intolerances. Larger deviations, on the other hand, may be perceived as reproductional losses, malformations, or genetic/genomic disorders, which may result from large CNVs that are both rare and deleterious, as they tend to contain dosage sensitive genes [41]. It was shown that the number of events is steadily decreasing with increased CNV length, so there are many more small (common) CNVs than large (deleterious) ones [58].

CNVs in the light of evolutionary mechanisms

It was reported in several organisms that genes having tissue-specific expression tend to be more variable in copy numbers than widely expressed genes, which might have housekeeping functions [59–61]. The most commonly cited examples of evolutionarily important CNVs in humans fit well into this concept, including the CNV-related evolution of trichromatic vision [62], hemoglobins and myoglobins [63], and olfactory genes [64]. The gene family of amylase, a starch-digesting enzyme, may serve as a multipurpose example in understanding CNV-mediated evolutionary processes, from a mechanistic view of the generation of complex genomic features and novel functions to the explanation of certain adaptation models. The study of Axelsson et al. showed a 7.4-fold average increase of *AMY2B* copy numbers in dogs relative to wolves. The change correlates to an increase in both expression level and enzyme activity, indicating that duplications of the alpha-amylase locus conferred a selective advantage to early ancestors of modern dogs. An increase in amylase activity allowed them to thrive on a diet rich in starch compared to the carnivorous diet of wolves and constituted a crucial step in early dog domestication [65]. Mammals, in general, produce amylase in their pancreas, while primates, rodents, and lagomorphs also show salivary expression of the enzyme. The GRCh38/hg38 assembly of the human reference genome contains two pancreatic (*AMY2A*, *AMY2B*) and three salivary amylase genes (*AMY1A*, *AMY1B*, *AMY1C*), arrayed close to each other on chromosome 1, but *AMY1B* having an opposite orientation [66]. These genes were reported to be present in highly variable copy numbers in humans, ranging from 2 to 18 copies with an average of 6 copies per person [67]. Average copy numbers of *AMY1* were shown to be higher in populations with high-starch diets compared to those with traditionally low-starch diets, suggesting that when starch became a prominent component of the human diet, it led to the positive selection of the *AMY1* gene [66]. Evolution of this human multigene family likely involved several steps in primate, and later in human evolutionary lineages, with mechanisms involving unequal, homologous, and inter- and intrachromosomal crossovers [68].

Among typical examples of balancing selection maintaining CNVs in the population, the hemoglobin genes and their association to two severe human diseases, thalassemia and malaria are usually mentioned [69]. While homozygous deletions of one of the α-globin genes cause α-thalassemia on one arm of the balance, heterozygous deletions protect their carriers against malaria on the opposite arm, leading to a correlating frequency of α-thalassemia and malaria prevalence in human populations of malaria-endemic countries [70]. Such balancing selection may shape the distribution of certain CNVs for generations, and they may be further maintained by other selective agents in later generations. For such cases, CNVs involving the immunoregulatory and inflammatory cytokine *CCL3L1* may be used as models in which higher copy numbers were found to correlate with a lower risk of HIV infection and AIDS progression [71], but *CCL3L1* copy numbers may also be reduced due to their association with other severe diseases such as systemic lupus erythematosus [72], strongly suggesting that certain CNVs may have been subject to highly dynamic and heterogeneous forms of evolutionary pressure [69].

Duality of evolution and disease

The mechanisms by which CNVs contribute to evolution but also cause disease are highly similar and share a common basis. The most commonly mentioned mechanisms are gene duplications or multiplications, which are considered to be essential sources of evolutionary innovation, as redundant gene copies may acquire new functions. If the multiplicated gene is not dosage sensitive, so fitness is not reduced, one of the copies may keep its original function while the other one may escape selective pressure and silently undergo continuous changes by mutation. And although the original copy (or even a functional single-copy gene) may also undergo slight changes over generations, duplicated genes have a tendency to accumulate mutations faster. Some other CNV-related processes acting on disease and evolution [41] include: i) direct influence on the expression of gene products leading to changing levels of a protein, e.g., by changing the copy number of the coding sequence itself, as the gene and its regulatory regions may both be encompassed by the CNV; ii) direct influence on expression levels, or on localization or timing of expression through alteration – including creation or disruption – of regulatory regions when the coding sequence itself is not encompassed by the CNV; or iii) acquisition of new functions by forming new or modified products – or altered localization or timing of expression of pre-existing products – through recombination of functional domains of different genes.

Comparative analyses of human and chimpanzee genomes helped us understand the evolutionary significance of CNVs. Approximately one-third of the CNVs observed in the human genome, including some human disease-causing duplications, are not duplicated in chimpanzees [12]. It seems that the evolution of the mammalian genome during primate speciation has led to a genome architecture predisposing some regions to rearrangements, which also resulted in genomic disorders [73]. Genes that have likely evolved under purifying or positive selection for copy number changes have been identified, particularly those with inflammatory response functions such as *APOL1*, *APOL4*, *CARD18*, *IL1F7*, and *IL1F8*, which are deleted in chimpanzees. Moreover, a copy number loss of the oncogene *TBC1D3* involved in cell proliferation was observed in the chimpanzee compared with the human

reference genome [74]. In conclusion, although there is evidence that altered copies of specific genes offer a selective advantage, many variations in copy numbers are disadvantageous as they are involved in the formation and progression of several pathological conditions [41].

Pathological aspects of CNVs

As we mentioned above, many variations of the human genome, including variations in gene copy number, play important roles in human health and disease. Since CNVs often span across a large number of genes and regulatory regions, many of which are biomedically actionable and/or included in the OMIM database, pathogenic CNVs have been found to cause genomic disorders with Mendelian inheritance [75] or to be associated with complex, multifactorial diseases [76] including cancer [77], cardiovascular [78,79], autoimmune [80], neurodevelopmental and neurodegenerative disorders [81,82]. When considering aneuploidies and short tandem repeats (STRs) as CNVs, this list may be extended by aneuploidy-associated syndromes such as Down, Edwards, Patau, Klinefelter, or Turner [83], by repeat expansion disorders, which are typically severe neurodegenerative or neuromuscular conditions [84], and by many types of cancer that involve microsatellite instability [85] or a change in telomere length [86].

Pathomechanisms

CNVs are prevalent in both coding and non-coding regions of the genome. They can directly affect a coding sequence and cause disruption of gene function or alter gene dosage leading to the development of a disorder [87]. However, CNVs that encompass only non-coding regions may also have a functional impact on the human genome by a number of possible mechanisms. It was shown that CNVs may influence tissue-level gene expression through harboring small non-coding RNAs (sncRNAs) [88]. Studies have found that CNV-sncRNAs include miRNA [89], snoRNA [90], piRNA [91], and tRNA [92], and CNVs harboring long non-coding RNAs have also been reported [93]. CNVs may alter chromatin interaction domains, also known as topologically associated domains, which may disrupt spatial organization of the genome and result in pathogenic phenotypes [94].

Expansions of microsatellites – sometimes considered as CNVs – represent additional pathogenic mechanisms such as the formation of extremely stable aggregates of mRNA or protein [95]. Proteins may exert a cumulative direct toxic effect both after canonical translation and non-canonical non-ATG initiated translation [96]. At the RNA level, well-known indirect pathogenic effects act through an expanded RNA-repeat-mediated depletion or activation of regulators of various cellular functions, such as alternative splicing [97]. Telomeres and the highly dynamic nature of their repeat numbers also represent important problems, including their continuous shortening over a lifetime contributing to aging and their escape from shortening in cancerous cell lines [86].

Another conventional, although indirect cause of genomic rearrangement is the unmasking of recessive mutations or functional polymorphisms when a deletion occurs [98]. In contrast, duplication of a normal gene copy on one chromosome may mask a disease-causing mutation on the other chromosome, resulting in a healthy phenotype [99]. Copy number of the *SMN2* gene may indirectly modify severity in spinal muscular atrophy (SMA), a disease caused by the homozygous deletion of the *SMN1* gene. The gene *SMN2* differs from *SMN1* in a substitution causing exclusion of exon 7 during splicing, resulting in a truncated and unstable protein for most (~85%) *SMN2* transcripts. Although the absence of *SMN2* alone does not cause SMA, its copy number may alter the SMA phenotype: increased *SMN2* copy numbers were present in patients with milder (SMA type III) phenotype compared to patients with more severe SMA type I [100]. However, *SMN2* copy number is not a definitive modifier of SMA severity, and other modifying factors need to be taken into account for disease characterization [101].

Genomic disorders with Mendelian inheritance

In 1998, the term "genomic disorder" was defined: it refers to disorders caused by structural changes of the human genome. Such DNA rearrangements may lead to the loss or gain of dosage-sensitive gene(s) or disrupt a gene [10]. While in many monogenic diseases, an abnormal phenotype is caused by a point mutation, many of these genomic disorders are known as recombination-based conditions [102]. In addition, similarly to large chromosomal aberrations, smaller CNVs affecting only one exon or even smaller regions are also associated with human diseases. Therefore, depending on the size of the genomic segment involved, its position and genomic context, as well as the number of genes within the rearranged segment together with other risk factors, CNV-associated disorders may be classified as Mendelian diseases, contiguous gene syndromes, chromosomal disorders, or other sporadic or complex traits [73]. Among these, Mendelian genomic disorders may segregate as autosomal recessive, such as nephronophthisis 1, juvenile [103], autosomal dominant, such as Charcot–Marie–Tooth disease type 1A or hereditary neuropathy with liability to pressure palsies (both caused by one of the first identified disease-associated submicroscopic CNVs, i.e., a duplication or a deletion, respectively, which are the reciprocal products of the same non-homologous crossover event) [75], X-linked, such as hemophilia A [104], or even as Y-linked traits, such as azoospermia [105]. Repeat expansion disorders show a mode of inheritance depending on the chromosomal localization of the involved gene. Certain non-Mendelian manifestations are also typical for many of these disorders. These may include anticipation in myotonic dystrophy type 1 [106] or phenotype-modifying potential of certain repeat alleles when in combination with other pathogenic variants, like in the case of the suggested reciprocal interaction between myotonic dystrophy type 2 premutations and congenital myotonia caused by mutations of the *CLCN1* gene [107].

Multifactorial diseases

Another example where CNVs may play an important role are complex diseases with multifactorial etiology, caused by the combination of several genetic factors (each one having a low impact on the phenotype alone) and environmental factors.

Continuous progress in recent decades has increased our understanding of the pathophysiology of many complex diseases. However, there are still unanswered questions of risk factors or largely unknown genetic background, which prevent us from clearly uncovering and understanding the pathomechanism of such diseases. Therefore, it is assumed that the implication of CNVs in the pathogenesis of complex diseases could explain at least a fraction of the well-known "missing heritability" problem of these complex disorders [108]. Inflammatory bowel disease (IBD) represents a typical example of multifactorial disease. More than 200 IBD associated loci are known, yet the pathogenesis is still unclear [109]. Some authors assume that studying CNVs may shed more light on the mystery of IBD [76,110,111]. Recently, results of Frenkel et al. pointed not only to the important role of CNVs, but also to significant pathways in the pathogenesis of IBD [112]. Even though CNVs are heavily implicated, such large genetic variants are still understudied in IBD and other complex diseases [113]. This is probably due to the limitations of methods suitable for CNV detection at the time when major genome-wide association studies were carried out.

Infectious diseases may also be considered as complex multifactorial diseases as both genetic and environmental variability affect the susceptibility of individuals to infections. It was shown that host CNVs play an important role in clinical phenotypes related to infectious diseases. Examples include variants of α-globin [6] or the *CCL3L1* gene [5] as explained in chapter 4 (Evolutionary aspects of CNVs), among others [114].

CNV detection may also find application in the evaluation of microbiome balance through the analysis of CNVs in metagenomes in different body parts. The human microbiome interacts with the host and plays an important role in many host biological processes [7]. Host genomic variations influence the composition of the microbiome, which in turn affects the health of the individual. While numerous studies have been focused on associations between the gut microbiome and specific alleles of the host genome, gene copy number also varies. It was shown, for instance, that duplication of the human *AMY1* gene is associated with an increased number of oral *Porphyromonas* in saliva, which is linked to periodontitis. The gut microbiota of these individuals had an increased abundance of resistant starch-degrading microbes, produced higher levels of short-chain fatty acids, and drove increased adiposity when transferred to germ-free mice [8]. This case demonstrated that even seemingly harmless variants in the host genome may affect the health of individuals.

Current knowledge suggests the importance of analyzing CNVs not only in human cells but also in the microbiome. Taxonomic characterization of the human microbiota is often limited to the species level, but each microbial species represents a large collection of strains that may contain considerably different sets or copy numbers of genes resulting in potentially distinct functional features. This intra-species variation is caused by deletion and duplication events, which were shown to be prevalent in the human gut environment, with some species exhibiting CNVs in >20% of their genes. Variability is especially relevant in disease-associated genes involved in important functions such as transport and signaling. Greenblum et al. showed obesity to be associated with higher copy numbers of thioredoxin 1 in *Clostridium sp.*, an increased copy number of an MFS transporter gene in the *Roseburia inulinivorans* genome cluster, and increased HlyD in *Bacteroides uniformis* associated with IBD-afflicted individuals. According to the authors, the analysis of species composition alone is not sufficient to capture the true functional potential of the microbiome, as it may fail to capture important functional differences [9]. Hence, the analysis of intra-species variation in microbial communities is crucial.

Cancer

Although cancer may be monogenic or, more typically, multifactorial, oncogenetics tends to be considered as a special field and discussed separately. In cancer genetics, CNVs are divided into two classes based on their size: i) large-scale, also known as chromosome-arm level variants encompassing >25% [115] or $^1/_3$ of the chromosome arm [116]; and ii) focal variants defined as small regions of the genome, usually not more than 3 Mb in size, containing up to a few genes [117–119]. Since 25% of an average human chromosome arm contains more than 15 Mb of DNA, variants ranging from 3 to about 15 Mb do not meet the criteria of either one of the above categories. On the other hand, in the case of 21p or Yp, 25% of total length comprises less than 3 Mb, so a variant at such a location might fall into both classes at the same time. To our knowledge, there is no strict consensus in the literature, so we suggest to classify large-scale variants as >3 Mb, while focal variants should retain their current definition. Both types of CNVs are important in the context of disease, but the relatively small size and low gene content make focal CNVs more suitable for the identification of candidate driver genes. Analysis of CNVs is an important aspect of the molecular diagnosis of cancer. It was shown that recurring deletions are typically overrepresented in tumor suppressor genes and underrepresented in oncogenes [120]. Aberrations in gene copy numbers may reveal therapeutic targets or markers of drug resistance in several types of cancer [121]. High-resolution copy-number profiles of 3131 cancer specimens revealed the CNV landscape of the vast majority of cancer types. An average tumor sample consists of 17% genome amplification and 16% deletion, compared to averages of 0.35% and less than 0.1% in normal samples, respectively. Specific gene families and pathways have been shown to be overrepresented among focal somatic copy-number alterations. The most enriched ones are gene families important in cancer pathogenesis, such as kinases, cell cycle regulators and MYC family members [77], which may represent potential therapeutic targets.

Conclusions

The days when the Mendelian dogma of all our genes being present in two copies could be applied universally have long since passed. Normal and pathogenic CNVs have been described in eukaryotic cells as well as in the microbiome, and their phenotypic associations and causes involving replication, recombination and repair, along with environmental mutagens, are becoming more well-known as research methods evolve. And yet, general awareness of CNVs as

common polymorphisms and studies aiming to shed light on their contribution to the missing heritability problem seen in genome-wide association studies are relatively scarce, with SNPs still stealing most of the spotlight [58,122]. If we want to be careful not to "miss the forest for the trees" [123], an integrated approach is advised, taking different forms of polymorphism and gene expression data into account, rather than maintaining a sole focus on SNP analysis [124]. Having assessed recent developments in the field, we are of the view that copy number research deserves more attention as a vital and very interesting aspect of the paradigm shift currently underway in molecular and clinical genetics.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the OP Integrated Infrastructure for the project [ITMS: 313011V578] co-financed by the European Regional Development Fund; the Slovak Research and Development Agency [Grant ID: APVV-18-0319]; and the PANGAIA project H2020-MSCA-RISE-2019 [Grant agreement ID: 872539] funded under H2020-EU.1.3.3. Programme.

REFERENCES

[1] Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genom Hum Genet 2009;10:451–81.
[2] Wright D, Boije H, Meadows JR, Bed'hom B, Gourichon D, Vieaud A, et al. Copy number variation in Intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genet 2009;5:e1000512.
[3] Henkel J, Saif R, Jagannathan V, Schmocker C, Zeindler F, Bangerter E, et al. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLoS Genet 2019;15:e1008536.
[4] Elder PJD, Ramsden DB, Burnett D, Weickert MO, Barber TM. Human amylase gene copy number variation as a determinant of metabolic state. Expet Rev Endocrinol Metabol 2018;13:193–205.
[5] Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 2005;307:1434–40.
[6] Harteveld CL, Higgs DR. Alpha-thalassaemia. Orphanet J Rare Dis 2010;5:13.
[7] Mohajeri MH, Brummer RJM, Rastall RA, Weersma RK, Harmsen HJM, Faas M, et al. The role of the microbiome for human health: from basic science to clinical applications. Eur J Nutr 2018;57:1–14.
[8] Poole AC, Goodrich JK, Youngblut ND, Luque GG, Ruaud A, Sutter JL, et al. Human salivary amylase gene copy number impacts oral and gut microbiomes. Cell Host Microbe 2019;25:553–64. e7.
[9] Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. Cell 2015;160:583–94.
[10] Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 1998;14:417–22.
[11] Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006;444:444–54.
[12] Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 2005;437:88–93.
[13] Arlt MF, Rajendran S, Birkeland SR, Wilson TE, Glover TW. Copy number variants are produced in response to low-dose ionizing radiation in cultured cells. Environ Mol Mutagen 2014;55:103–13.
[14] Kubiritova Z, Gyuraszova M, Nagyova E, Hyblova M, Harsanyova M, Budis J, et al. On the critical evaluation and confirmation of germline sequence variants identified using massively parallel sequencing. J Biotechnol 2019;298:64–75.
[15] Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet 2015;16:172–83.
[16] Yang X, Lee WP, Ye K, Lee C. One reference genome is not enough. Genome Biol 2019;20:849.
[17] Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol 2015;23:148–54.
[18] Strillacci MG, Gorla E, Cozzi MC, Vevey M, Genova F, Scienski K, et al. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. PloS One 2018;13:e0204669.
[19] Lee YL, Bosse M, Mullaart E, Groenen MAM, Veerkamp RF, Bouwman AC. Functional and population genetic features of copy number variations in two dairy cattle populations. BMC Genom 2020;21:89.
[20] Strillacci MG, Gorla E, Ríos-Utrera A, Vega-Murillo VE, Montaño-Bermudez M, Garcia-Ruiz A, et al. Copy number variation mapping and genomic variation of autochthonous and commercial Turkey populations. Front Genet 2019;10:982.
[21] Bovo S, Ribani A, Muñoz M, Alves E, Araujo JP, Bozzi R, et al. Genome-wide detection of copy number variants in European autochthonous and commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. Anim Genet 2020;51:541–56.
[22] Durán Aguilar M, Román Ponce SI, Ruiz López FJ, González Padilla E, Vásquez Peláez CG, Bagnato A, et al. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. J Anim Breed Genet 2017;134:49–59.
[23] Di Gerlando R, Sutera AM, Mastrangelo S, Tolone M, Portolano B, Sottile G, et al. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. PloS One 2019;14:e0215204.
[24] Schurink A, da Silva VH, Velie BD, Dibbits BW, Crooijmans RPMA, François L, et al. Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity. BMC Genet 2018;19:49.
[25] Tice SC. A new sex-linked character in drosophila. Biol Bull 1914;26:221–30.
[26] Pös O, Radvanszky J, Styk J, Pös Z, Buglyó G, Kajsik M, et al. Copy number variation: methods and clinical applications. NATO Adv Sci Inst Ser E Appl Sci 2021;11:819.

[27] Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med 2020;22:245–57.
[28] Bridges CB. The bar "gene" a duplication. Science 1936;83:210–1.
[29] Tjio JH, Levan A. The chromosome number of man. Hereditas 2010;42:1–6.
[30] Lejeune J, Lafourcade J, Berger R, Vialatte J, Boeswillwald M, Seringe P, et al. 3 cases of partial deletion of the short arm of a 5 chromosome. C R Hebd Seances Acad Sci 1963;257:3098–102.
[31] Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet 2004;36:949–51.
[32] Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science 2004;305:525–8.
[33] Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006;7:85–97.
[34] Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet 2011;12:363–76.
[35] MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 2014;42:D986–92.
[36] Werdyani S, Yu Y, Skardasi G, Xu J, Shestopaloff K, Xu W, et al. Germline INDELs and CNVs in a cohort of colorectal cancer patients: their characteristics, associations with relapse-free survival time, and potential time-varying effects on the risk of relapse. Cancer Med 2017;6:1220–32.
[37] Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ, et al. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. Genome Res 2013;23:749–61.
[38] Ballouz S, Dobin A, Gillis JA. Is it time to change the reference genome? Genome Biol 2019;20:159.
[39] Orr B, Godek KM, Compton D. Aneuploidy. Curr Biol 2015;25:R538–42.
[40] Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its roles in human copy number variation. Genomics Inform 2014;12:136–44.
[41] Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet 2009;10:551–64.
[42] Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics 2008;1:4.
[43] Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem 2004;279:47411–4.
[44] Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature 2010;464:704–12.
[45] Andriani GA, Vijg J, Montagna C. Mechanisms and consequences of aneuploidy and chromosome instability in the aging brain. Mech Ageing Dev 2017;161:19–36.
[46] Gregan J, Polakova S, Zhang L, Tolić-Nørrelykke IM, Cimini D. Merotelic kinetochore attachment: causes and effects. Trends Cell Biol 2011;21:374–81.
[47] Khristich AN, Mirkin SM. On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. J Biol Chem 2020;295:4134–70.
[48] Bachinski LL, Czernuszewicz T, Ramagli LS, Suominen T, Shriver MD, Udd B, et al. Premutation allele pool in myotonic dystrophy type 2. Neurology 2009;72:490–7.
[49] Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. Nat Med 2016;22:1342–50.
[50] Takagi S, Kinouchi Y, Hiwatashi N, Nagashima F, Chida M, Takahashi S, et al. Relationship between microsatellite instability and telomere shortening in colorectal cancer. Dis Colon Rectum 2000;43:S12–7.
[51] Witzany G. The viral origins of telomeres and telomerases and their important role in eukaryogenesis and genome maintenance. Biosemiotics 2008;1:191–206.
[52] Jeffreys AJ, Neil DL, Neumann R. Repeat instability at human minisatellites arising from meiotic recombination. EMBO J 1998;17:4147–57.
[53] Arlt MF, Wilson TE, Glover TW. Replication stress and mechanisms of CNV formation. Curr Opin Genet Dev 2012;22:204–10.
[54] Hovhannisyan G, Harutyunyan T, Aroutiounian R, Liehr T. DNA copy number variations as markers of mutagenic impact. Int J Mol Sci 2019;20:4723.
[55] Costa EOA, Pinto IP, Gonçalves MW, da Silva JF, Oliveira LG, da Cruz AS, et al. Small de novo CNVs as biomarkers of parental exposure to low doses of ionizing radiation of caesium-137. Sci Rep 2018;8:5914.
[56] Harutyunyan T, Hovhannisyan G, Sargsyan A, Grigoryan B, Al-Rikabi AH, Weise A, et al. Analysis of copy number variations induced by ultrashort electron beam radiation in human leukocytes in vitro. Mol Cytogenet 2019;12:18.
[57] Bennett PM. Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. Methods Mol Biol 2004;266:71–113.
[58] Shen H, Li J, Zhang J, Xu C, Jiang Y, Wu Z, et al. Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. PloS One 2013;8:e59494.
[59] Dopman EB, Hartl DL. A portrait of copy-number polymorphism in Drosophila melanogaster. Proc Natl Acad Sci U S A 2007;104:19920–5.
[60] Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, et al. Segmental copy number variation shapes tissue transcriptomes. Nat Genet 2009;41:424–9.
[61] Keel BN, Lindholm-Perry AK, Snelling WM. Evolutionary and functional features of copy number variation in the cattle genome. Front Genet 2016;7:207.
[62] Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. PLoS Biol 2004;5:e148.
[63] Storz JF. Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. Physiology 2016;31:223–32.
[64] Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ. Extensive copy-number variation of the human olfactory receptor gene family. Am J Hum Genet 2008;83:228–42.
[65] Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 2013;495:360–4.
[66] Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet 2007;39:1256–60.
[67] Carpenter D, Mitchell LM, Armour JAL. Copy number variation of human AMY1 is a minor contributor to variation in salivary amylase expression and activity. Hum Genom 2017;11:2.
[68] Groot PC, Mager WH, Henriquez NV, Pronk JC, Arwert F, Planta RJ, et al. Evolution of the human α-amylase multigene family through unequal, homologous, and inter-

and intrachromosomal crossovers. Genomics 1990;8:97–105.
[69] Perry GH. The evolutionary significance of copy number variation in the human genome. Cytogenet Genome Res 2008;123:283–7.
[70] Farashi S, Harteveld CL. Molecular basis of α-thalassemia. Blood Cells Mol Dis 2018;70:43–53.
[71] Liu S, Yao L, Ding D, Zhu H. CCL3L1 copy number variation and susceptibility to HIV-1 infection: a meta-analysis. PloS One 2010;5:e15778.
[72] Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, et al. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. Ann Rheum Dis 2008;67:1076–83.
[73] Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet 2002;18:74–82.
[74] Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, et al. Copy number variation and evolution in humans and chimpanzees. Genome Res 2008;18:1698–710.
[75] Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 1991;66:219–32.
[76] Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number Predisposes to crohn disease of the colon. Am J Hum Genet 2006;79:439–48.
[77] Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463:899–905.
[78] Fahed AC, Gelb BD, Seidman JG, Seidman CE. Genetics of congenital heart disease: the glass half empty. Circ Res 2013;112:707–20.
[79] Zekavat SM, Ruotsalainen S, Handsaker RE, Alver M, Bloom J, Poterba T, et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nat Commun 2018;9:2606.
[80] Olsson LM, Holmdahl R. Copy number variation in autoimmunity-importance hidden in complexity? Eur J Immunol 2012;42:1969–76.
[81] Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. Neuron 2006;52:103–21.
[82] Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature 2016;530:177–83.
[83] Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. Clin Perinatol 2015;42:227–42. vii.
[84] Vittori A, Breda C, Repici M, Orth M, Roos RAC, Outeiro TF, et al. Copy-number variation of the neuronal glucose transporter gene SLC2A3 and age of onset in Huntington's disease. Hum Mol Genet 2014;23:3129–37.
[85] Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017;8:15180.
[86] Shammas MA. Telomeres, lifestyle, cancer, and aging. Curr Opin Clin Nutr Metab Care 2011;14:28–34.
[87] Nowakowska B. Clinical interpretation of copy number variants in the human genome. J Appl Genet 2017;58:449–57.
[88] Kumaran M, Krishnan P, Cass CE, Hubaux R, Lam W, Yasui Y, et al. Breast cancer associated germline structural variants harboring small noncoding RNAs impact post-transcriptional gene regulation. Sci Rep 2018;8:7529.
[89] Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. Copy number variation of microRNA genes in the human genome. BMC Genom 2011;12:183.
[90] Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat Genet 2008;40:719–21.
[91] Gould DW, Lukic S, Chen KC. Selective constraint on copy number variation in human piwi-interacting RNA Loci. PloS One 2012;7:e46611.
[92] Iben JR, Maraia RJ. tRNA gene copy number variation in humans. Gene 2014;536:376–84.
[93] Liu H, Gu X, Wang G, Huang Y, Ju S, Huang J, et al. Copy number variations primed lncRNAs deregulation contribute to poor prognosis in colorectal cancer. Aging 2019;11:6089–108.
[94] Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 2015;161:1012–25.
[95] Conlon EG, Manley JL. RNA-binding proteins in neurodegeneration: mechanisms in aggregate. Genes Dev 2017;31:1509–28.
[96] Pearson CE. Repeat Associated Non-ATG Translation Initiation: One DNA, Two Transcripts, Seven Reading Frames, Potentially Nine Toxic Entities! PLoS Genet 2011;7:e1002018.
[97] Radvanszky J, Kadasi L. Molecular genetic basis of myotonic dystrophy. eLS, vol. 72. Chichester, UK: John Wiley & Sons, Ltd; 2001. p. 490.
[98] Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med 2010;61:437–55.
[99] Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 2007;8:639–46.
[100] Kolb SJ, Kissel JT. Spinal muscular atrophy. Neurol Clin 2015;33:831–46.
[101] Wirth B, Karakaya M, Kye MJ, Mendoza-Ferreira N. Twenty-five years of spinal muscular atrophy research: from phenotype to genotype to therapy, and what comes next. Annu Rev Genom Hum Genet 2020;21:231–61.
[102] Shaffer LG, Lupski JR. Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu Rev Genet 2000;34:297–329.
[103] Konrad M, Saunier S, Heidet L, Silbermann F, Benessy F, Calado J, et al. Large homozygous deletions of the 2q13 region are a major cause of juvenile nephronophthisis. Hum Mol Genet 1996;5:367–71.
[104] Gitschier J, Wood WI, Tuddenham EG, Shuman MA, Goralka TM, Chen EY, et al. Detection and sequence of mutations in the factor VIII gene of haemophiliacs. Nature 1985;315:427–30.
[105] Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, et al. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 2001;29:279–86.
[106] Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell 1992;68:799–808.
[107] Radvanszky J, Surovy M, Polak E, Kadasi L. Uninterrupted CCTG tracts in the myotonic dystrophy type 2 associated locus. Neuromuscul Disord 2013;23:591–8.

[108] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.
[109] de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256–61.
[110] McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008;40:1107–12.
[111] Saadati HR, Wittig M, Helbig I, Häsler R, Anderson CA, Mathew CG, et al. Genome-wide rare copy number variation screening in ulcerative colitis identifies potential susceptibility loci. BMC Med Genet 2016;17:26.
[112] Frenkel S, Bernstein CN, Sargent M, Kuang Q, Jiang W, Wei J, et al. Genome-wide analysis identifies rare copy number variations associated with inflammatory bowel disease. PloS One 2019;14:e0217846.
[113] Ellinghaus D, Bethune J, Petersen BS, Franke A. The genetics of Crohn's disease and ulcerative colitis-status quo and beyond. Scand J Gastroenterol 2015;50:13–23.
[114] Hollox EJ, Hoh BP. Human gene copy number variation and infectious disease. Hum Genet 2014;133:1217–33.
[115] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.
[116] Serin Harmanci A, Harmanci AO, Zhou X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. Nat Commun 2020;11:89.
[117] Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A 2008;105:16224–9.
[118] Brosens RPM, Haan JC, Carvalho B, Rustenburg F, Grabsch H, Quirke P, et al. Candidate driver genes in focal chromosomal aberrations of stage II colon cancer. J Pathol 2010;221:411–24.
[119] Krijgsman O, Carvalho B, Meijer GA, Steenbergen RDM, Ylstra B. Focal chromosomal copy number aberrations in cancer-Needles in a genome haystack. Biochim Biophys Acta 2014;1843:2698–704.
[120] Zhang L, Yuan Y, Lu KH, Zhang L. Identification of recurrent focal copy number variations and their putative targeted driver genes in ovarian cancer. BMC Bioinf 2016;17:222.
[121] Peng H, Lu L, Zhou Z, Liu J, Zhang D, Nan K, et al. CNV detection from circulating tumor DNA in late stage non-small cell lung cancer patients. Genes 2019;10:926.
[122] Nagao Y. Copy number variations play important roles in heredity of common diseases: a novel method to calculate heritability of a polymorphism. Sci Rep 2015;5:17156.
[123] Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. Circulation 2007;115:3130–8.
[124] Momtaz R, Ghanem NM, El-Makky NM, Ismail MA. Integrated analysis of SNP, CNV and gene expression data in genetic association studies. Clin Genet 2018;93:557–66.