nature reviews genetics

https://doi.org/10.1038/s41576-023-00574-0

Review article

Check for updates

# Mendelian inheritance revisited: dominance and recessiveness in medical genetics

Johannes Zschocke [1] ✉, Peter H. Byers [2,3] & Andrew O. M. Wilkie [4]

## Abstract

Understanding the consequences of genotype for phenotype (which ranges from molecule-level effects to whole-organism traits) is at the core of genetic diagnostics in medicine. Many measures of the deleteriousness of individual alleles exist, but these have limitations for predicting the clinical consequences. Various mechanisms can protect the organism from the adverse effects of functional variants, especially when the variant is paired with a wild type allele. Understanding why some alleles are harmful in the heterozygous state — representing dominant inheritance — but others only with the biallelic presence of pathogenic variants — representing recessive inheritance — is particularly important when faced with the deluge of rare genetic alterations identified by high throughput DNA sequencing. Both awareness of the specific quantitative and/or qualitative effects of individual variants and the elucidation of allelic and non-allelic interactions are essential to optimize genetic diagnosis and counselling.

## Sections



[1]Institute of Human Genetics, Medical University Innsbruck, Innsbruck, Austria. [2]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [3]Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA. [4]MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ✉e-mail: johannes.zschocke@i-med.ac.at

## Introduction

Gregor Mendel's observation that some physical traits are inherited as discrete units that can completely disappear and reappear over successive generations[1] was a crucial step in the identification of the gene as the core unit of inheritance[2]. With the rediscovery of Mendel's experiments in 1900, it was immediately recognized that some human diseases – later referred to as monogenic, single gene or Mendelian diseases – can be understood in the Mendelian terms of dominant and recessive[3,4]. Medical genetics defines these terms based on the clinical consequences in the heterozygote, when there is a pathogenic variant on only one copy of a biallelic gene: a condition is denoted dominant if it manifests in the heterozygous state, irrespective of the clinical features in the homozygote[5]. This definition differs from the original approach used by Mendel, who deliberately chose common traits in the garden pea that were identical in the hybrid (heterozygote) and one type of homozygote[2], but is more suitable for rare inherited diseases that interfere with normal life. The term recessive is used when a heterozygous variant is asymptomatic, with a disease arising only with pathogenic variants on both copies of a gene.

The introduction of massively parallel (also known as next-generation) sequencing into medical care over the past dozen years, in particular for the simultaneous analysis of the coding sequences of all protein-coding human genes (the exome) in a routine diagnostic setting, has expanded our understanding of the genetic basis and variability of diseases. The Online Mendelian Inheritance in Man (OMIM) database currently lists 6,209 single gene disorders and traits (updated 8 November 2022), and these represent more than 70% of the 'rare diseases' (conditions with a prevalence of <1:2,000) that, in total, are estimated to affect 4–5% of the global general population[6]. However, the exact functional and clinical consequences remain to be elucidated for the majority of sequence[7,8] and structural[9] variants in the human genome. Different genetic alterations in the same gene may have diverse consequences, which are often difficult to predict. A substantial number of genes traditionally associated with either dominant or recessive diseases are now linked to both inheritance patterns, based on functionally different pathogenic variants. Indeed, of the 4,658 autosomal disease genes currently listed in OMIM, about 53% ($n$ = 2,464) are associated with dominant conditions, 35% ($n$ = 1,643) with recessive conditions and 12% ($n$ = 551) with both patterns of inheritance.

To understand inheritance patterns and the mechanisms of dominance and recessiveness in medical genetics, it is important to keep in mind that there is a long path from a genetic variant to a noticeable manifestation in the whole organism. Diploid organisms have a built-in redundancy (two active copies) of most autosomal genes and, over the space of evolutionary time, have developed a portfolio of mechanisms that help compensate for genetic alterations and adapt to changes in environmental or endogenous demands. Simply put, no gene or protein works in isolation. As a result, the functional consequences of a particular genotype often differ when examined at different phenotype levels, that is, the measurable effect on protein, cell, organ or clinical (or other) function (Fig. 1). Understanding these connections with regard to heritable diseases and variable clinical manifestations is the essence of medical genetics, and assessing variant effects in heterozygotes is one of the most fundamental challenges in genetic diagnostics.

This Review provides an overview of the mechanisms that determine whether a monogenic disease is caused by monoallelic or biallelic alterations, corresponding to pathogenic variants on one or both copies of a gene, respectively. Highlighting the important distinction between qualitative and quantitative variant effects, the Review explains why genes show differences in the ability to compensate dosage changes, and describes the principles underlying pathogenic functional gene and protein alterations. Various clinical examples are provided to illustrate complex pathogenic constellations, the overlap between monogenic and multifactorial conditions, and why the Mendelian terms are not suitable for gonosomal and mitochondrial genes. Finally, we suggest a framework for using the concepts of dominance and recessiveness in the prediction of genetic variant effects. Recognizing the different functional effects of genetic changes is essential for classifying and codifying novel variants in diagnostic molecular genetic analyses, and for estimating the probability that a disease will occur or recur in family members.

## Functional effects of genetic variants

One of the central aspects of explaining monogenic inheritance patterns is the distinction between quantitative and qualitative variant effects. Many variants have primarily quantitative 'dosage' consequences for the transcript and protein, that is, they cause a loss, reduction or increase of the gene product without introducing novel functional characteristics. The term loss of function (LoF) variant or null variant refers to the complete loss of a protein encoded from the allele, due to loss of the allele, unstable mRNA or unstable and inactive protein. Hypomorphic (Hyp) variants[10] do not completely abolish the gene product but cause quantitatively diminished, qualitatively unaltered protein function. By contrast, qualitative variants result in structural or regulatory alterations that change the normal function of the encoded protein. Different types of qualitative variant effects can be distinguished and are described in detail in the respective section below. Qualitative effects must be considered for all variants that produce a stable abnormal protein with potentially disrupted regulation, modification, processing, secretion, cellular localization or other function. The pathogenetic mechanisms may vary and be specific for different alterations in the same gene.

Whether the loss, reduction or alteration of protein function has an effect on cell or organ function, and whether it causes disease manifestations in the monoallelic or the biallelic state, depends on the normal and abnormal function of the affected protein in the context of compensatory mechanisms, individual differences, environmental influences and chance. As a general rule for proteins that are not tightly regulated, heterozygous variants that cause altered structure and function (qualitative) are more likely to result in phenotypic effects than heterozygous variants with reduced function (quantitative). The organism is better equipped to compensate for a deficiency than to counteract abnormal actions of a protein.

### Quantitative variant effects

The clinical consequences of quantitative variants depend on the combined functional effects in the biallelic genotype (Fig. 2a). There are many examples of quantitative variants resulting in an autosomal recessive condition, such as phenylketonuria (Fig. 1). Autosomal dominant diseases caused by quantitative variants (that is, there is a clinical manifestation in the heterozygote) are generally semi-dominant, with more severe consequences or lethality in the homozygous state.

**Loss of function variants in most genes are recessive.** Studies in several species, such as *Drosophila melanogaster*[11], *Chlamydomonas reinhardtii*[12] and yeast[13], indicate that organisms can compensate heterozygous LoF variants in the majority of genes, with no organismal

phenotypic difference between homozygous wild type (WT) and heterozygous LoF constellations. This phenomenon – denoted haplosufficiency and representing Mendelian recessiveness – is also characteristic of the majority of human genes. In a study that compared the observed and expected frequencies of predicted LoF variants in exome sequencing data sets from more than 60,000 individuals, and calculated the probability of being LoF intolerant (pLI) for >18,000 protein-coding genes, 10,374 genes were identified as likely haplosufficient (LoF tolerant, pLI ≤ 0.1) whereas only 3,230 genes were likely intolerant of heterozygous LoF variants (pLI ≥ 0.9)[14]. These findings were corroborated in a larger study of 125,748 exomes and 15,708 genomes with the calculation of a more elaborate metric denoted the LoF observed/expected upper bound fraction (LOEUF)[8,15]. Bioinformatic approaches to identify genes and variants associated with haplosufficiency, and thus autosomal recessive inheritance, from large-scale human population data are being developed[16].

Exome data from 454,787 mostly middle-aged adults in the UK Biobank showed that for >80% of genes, at least 50 individuals carried a predicted LoF variant[17]. In some instances, alternative mRNA splicing may skip particular exons of a gene at least in some organs and, therefore, counteract LoF variants in these exons[18], but such effects seem to be rare. Evidence that a high proportion of genes are haplosufficient was also provided by a comprehensive survey of the clinical effects of copy number variants (CNVs) and heterozygous gene deletions on chromosome 18. This study identified only 19/263 genes (7%) with an abnormal phenotype in more than 50% of affected individuals, whereas a heterozygous deletion of 146/263 genes (56%) was asymptomatic[19]. Thus, the majority of gene products have a relatively high dosage tolerance, and clinical effects are seen primarily with biallelic alterations. In summary, recessiveness of LoF variants (that is, haplosufficiency) seems to be the norm rather than the exception. However, it remains possible that heterozygous LoF variants in haplosufficient genes can sometimes produce subtle clinical manifestations, which may be associated with measurable quantitative abnormalities at other phenotypic levels[20]. Vice versa, predicted deleterious variants in genes associated with autosomal dominant developmental disorders may sometimes cause lifelong subclinical or attenuated related phenotypes in the adult general population[21].

**Haplosufficiency reflects compensatory mechanisms of the cell.** To explain the prevalence of haplosufficiency, Ronald Fisher proposed the evolutionary selection of modifiers that counteract the effects of recurring deleterious variants[22]. As an alternative explanation, Sewall Wright reasoned that dominance and recessiveness reflect differences in the redundancy of cellular functions[23]. He postulated a safety factor in enzymatic activity, which may also be the product of evolutionary selection. Subsequent studies have mostly supported Wright's concept. Kacser and Burns pointed out that combining multiple enzymes into a multistep biochemical pathway can buffer variation at individual steps[24]. A large change in the activity of a single enzyme therefore has only a limited impact on metabolic flux. This principle is also true for other regulatory networks[25]. For example, post-translational mechanisms assist in the regulation of protein concentrations[26], which are more conserved across species than the concentration of transcripts[27]. In summary, for genetic variants that have primarily quantitative dosage effects, recessiveness of the variant allele reflects the presence of compensatory mechanisms that secure clinically normal function. These mechanisms may also improve adaptation to environmental changes, and therefore may have been selected during evolution. They may also serve as possible targets for therapeutic interventions (Fig. 1).

**Haploinsufficiency is associated with highly regulated cellular functions.** In general, genes linked to highly regulated cellular functions – such as those encoding transcription factors or enzyme regulators – are more likely to be dosage-sensitive[28]. Haploinsufficient genes show much narrower ranges in cell to cell variability of expression for a given cell type than other genes in the genome, and it has been postulated that their expression may be limited by the toxicity of their overexpression[29]. Haploinsufficient genes also have a higher number of paralogues, as duplication of genes with highly regulated functions tends to be beneficial because of improved control options[28]. Transcription factor families with very few members are more likely to be dosage-sensitive than transcription factor families with more members[30]. The existence of a non-linear relationship between genetic and phenotypic variation, with threshold effects, can explain allelic dominance in the control of some developmental processes[31]. For cooperative binding of transcription factors to target regulatory sequences, a 'half dose' of a transcription factor may be insufficient to reach the normal transcriptional response threshold in specific cellular and/or developmental contexts[32,33]. A similar non-linear threshold effect based on dosage alterations of several interacting proteins has also been discussed for macromolecules[34]. Yeast studies give examples of haploinsufficiency mechanisms that depend on external factors[13].

**Triplosensitivity and haploinsufficiency are related.** Limited data are available with regard to the clinical consequences of an increased copy number of individual genes[35]. In yeast, >80% of genes do not show a reduced growth rate when overexpressed[36]. There is growing evidence that haploinsufficient genes are also more likely to be sensitive to increased dosage (gain of quantity), that is triplosensitive. For example, genes that have been shown to contribute to a Down syndrome phenotype in triosomy 21 are more likely to have a high pLI score, indicating haploinsufficiency[37]. Genome-based structural variant data showed a strong correlation between tolerance of LoF variants and occurrence of copy-gain structural variants[9]. Dosage sensitivity patterns for regulatory elements such as enhancers, repressors or transcription factor binding sites do not show obvious differences between deletions and duplications[38]. A recent analysis of CNV data in nearly one million individuals identified 3,635/18,641 autosomal protein-coding genes with high probability of dosage sensitivity. Of these, 2,076 were haploinsufficient, 911 both haploinsufficient and triplosensitive, and 648 triplosensitive only. Genes more sensitive to deletion than duplication were more likely to show features of tightly regulated developmentally critical genes, that is, they tended to be larger, farther from other genes and had a greater number of poised enhancers in *cis*[39]. An example of a gene associated with both haploinsufficiency and triplosensitivity phenotypes is *PMP22*, duplication of which leads to Charcot–Marie–Tooth disease type 1A (CMT1A), whereas its deletion causes hereditary neuropathy with liability to pressure palsies[40]. It is important to distinguish triplosensitivity from the effects of gain of function (GoF) variants that cause clinical manifestation through other mechanisms, as discussed below (see 'Qualitative variant effects'). Also, there may be non-specific effects of overexpression of some genes, as increased dosage may increase low-affinity abnormal, promiscuous interactions with off-target molecular partners[41]. In some instances, the effects of inappropriate gene overexpression may be unrelated to the specific protein function; for example, overexpression might cause depletion of

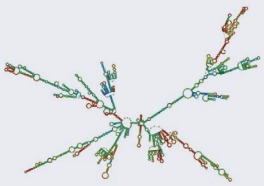
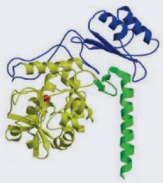
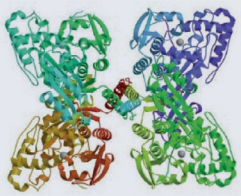
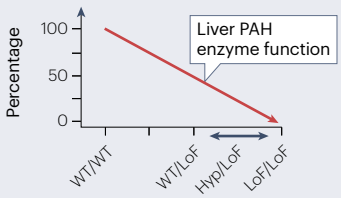
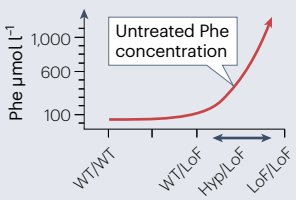
| Level | Variable functions and effects | | Abnormalities in PAH deficiency | Treatment strategies in PAH deficiency |
|---|---|---|---|---|
| **Genotype** | | | | |
| DNA ↓ | Genetic and epigenetic variants |  | Variants in the *PAH* gene | Potential gene therapy |
| **Transcript** | | | | |
| ↓ mRNA ↓ | Transcript quantity and sequence |  | Abnormal transcript amount and/or sequence | Potential mRNA replacement therapy |
| **Different phenotype levels** | | | | |
| ↓ Protein | Protein quantity, structure and function |  | Abnormal amount, stability, activity and/or regulation of PAH enzyme | Pharmaceutical supplementation of the cofactor tetrahydrobiopterin stabilizes mutant proteins |
| ↓ Multimer | Structural interactions |  | Some variants interfere with active tetramer formation | |
| ↓ Enzyme | Molecular function |  | Heterozygous variants cause reduced enzyme activity | |
| ↓ Cell metabolism ↓ | Regulatory and/or compensatory effects<br>Cellular function<br>Biochemical and network function |  | Increased levels of Phe and its metabolites and reduced levels of Tyr and its products | Dietary reduction of Phe and supplementation with Tyr to prevent abnormalities<br>Exogenous Phe ammonia lyase |
| ↓ Organ ↓ | Organ function |  | Increased levels of Phe interferes with transport processes at the blood–brain barrier | Supplementation with large neutral amino acids |
| ↓ Organism | Clinical manifestation |  | Progressive intellectual disability | Symptomatic therapy |

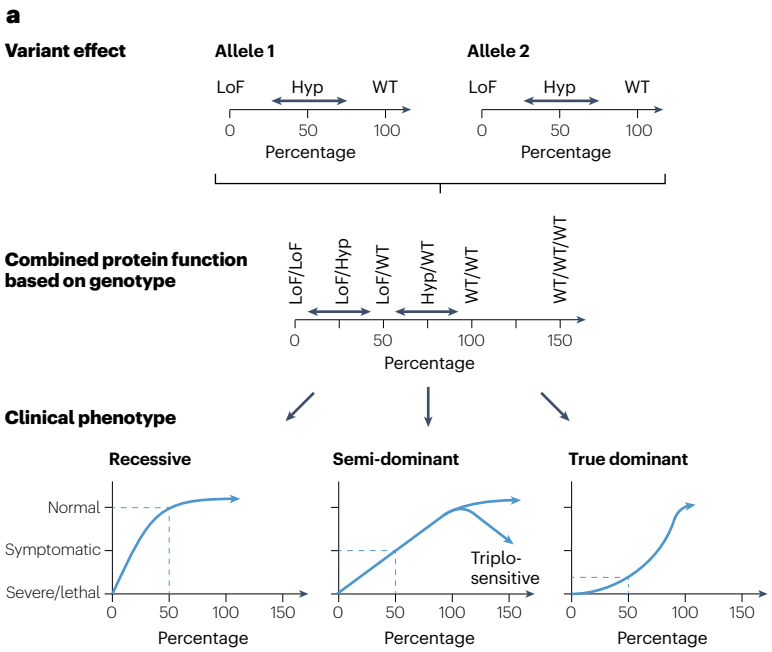
**Fig. 1 | Genotype and phenotypes – from DNA sequence to clinical manifestation in phenylalanine hydroxylase deficiency.** The pathway from DNA sequence to clinical presentation is depicted for phenylketonuria, an autosomal recessive disease caused by pathogenic variants in the phenylalanine hydroxylase (*PAH*) gene. The PAH enzyme encoded by this gene catalyses the conversion of the amino acid phenylalanine (Phe) to tyrosine (Tyr). There are a large number of different pathogenic *PAH* gene variants, including loss of function (LoF; null) variants that completely remove protein PAH function and hypomorphic (Hyp) variants with residual enzyme activity. WT denotes the normal (wild type) allele. The overall PAH activity for an individual results from the combined effect of two *PAH* alleles. At the enzymatic phenotype level, heterozygous *PAH* null variants (WT/LoF) are associated with 50% reduction of PAH activity in liver biopsies, representing an intermediate effect. Heterozygotes may show minor abnormalities at the metabolic phenotype level (amino acid concentrations in blood: Phe > Tyr, whereas normally Tyr > Phe) but are completely asymptomatic at the clinical phenotype level, in line with recessive inheritance. Individuals with biallelic null variants (LoF/LoF) in the *PAH* gene – when untreated – have highly elevated Phe concentrations in the blood and develop severe intellectual disability (clinical phenotype). Compound heterozygosity for a LoF variant and some Hyp variants (Hyp/LoF) is associated with sufficient enzyme function to maintain blood Phe concentrations below the therapeutic threshold of 360–600 µmol $l^{-1}$; this condition is denoted mild hyperphenylalaninaemia, which does not require treatment. Diagnosis and treatment of PAH deficiency are possible at various levels. Phenylketonuria is usually diagnosed prior to the onset of symptoms based on elevated Phe concentrations in newborn screening (that is, the metabolic phenotype level) but can also be identified through gene variant analysis. PAH enzyme studies require a liver biopsy. The standard therapy is a controlled low-Phe diet with supplementation of Tyr and other amino acids as well as vitamins and cofactors, which allows normal development and intellectual function. Other treatment strategies have been established or are being explored, as indicated in the last column. The availability of highly effective treatments changes the clinical phenotype of phenylketonuria from a 'disease' with severe neurological manifestation and intellectual disability to a 'disease risk' that necessitates a specialist diet to avoid the development of neurological symptoms. The structural image of the *PAH* mRNA was generated using RNAfold[174], and the PAH protein structures are reprinted with permission from ref. [175], ELSEVIER.

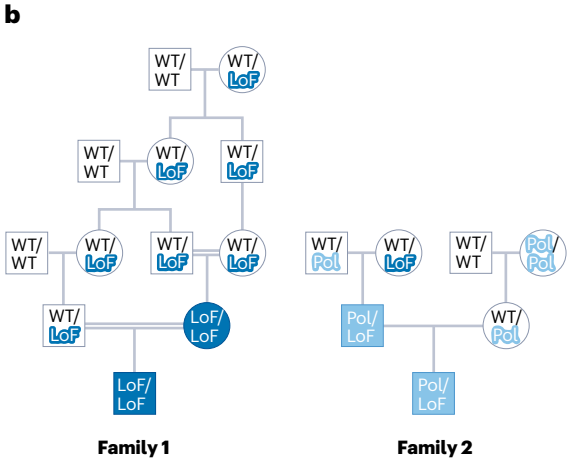
cellular resources for transcription and/or translation of other genes[42], or stoichiometric imbalances among multi-protein complexes that may affect the expression of other genes[43]. This indirect mechanism has been invoked for an 'omnigenic' model of complex diseases, in which regulatory changes to any gene expressed in a disease-relevant tissue might contribute to disease regardless of whether there is a direct mechanistic link to the specific phenotype[44,45].

**Variants with residual function can explain attenuated phenotypes.** Hyp variants are frequent causes of attenuated manifestations in dominant and recessive disorders (Fig. 2a; the term 'mild' should be avoided, particularly when patients still experience serious clinical consequences). An example is isolated congenital bilateral absence of the vas deferens, which is observed in 1–2% of infertile males. Affected individuals are usually compound heterozygous for a LoF variant and a Hyp variant in the *CFTR* gene, which encodes a chloride channel required for normal secretion in several epithelial tissues. Biallelic LoF variants cause cystic fibrosis, which affects the lungs, digestive system and other organs in the body. Residual channel function prevents clinical manifestations of classical cystic fibrosis but is insufficient for normal development of the vasa deferentia, causing azoospermia and infertility[46]. Using Mendel's terminology for recessive diseases, Hyp variants dominate over LoF variants as they determine disease severity in compound heterozygous individuals, just as the monoallelic WT confers a normal phenotype in the heterozygous variant carrier. The concept of functional hemizygosity in autosomal recessive diseases caused by quantitative variants helps determine the clinical impact of a novel or rare variant when a known LoF variant is on the other allele. In this circumstance, the uncharacterized variant is likely LoF if the phenotype is 'severe', whereas an attenuated phenotype must be mediated by residual function of that variant. The severity in the functionally hemizygous genotype may resemble homozygosity for the Hyp variant. This approach was first developed for the classification of phenylalanine hydroxylase (*PAH*) gene variants and the prediction of disease severity in phenylketonuria[47]. Some variants are associated with specific treatment options that enhance residual function[48]. Benign variants that in conjunction with a severe variant on the other allele cause laboratory abnormalities but no clinical disease have been recognized for many genes; examples include mild hyperphenylalaninaemia (Fig. 1) or Duarte galactosaemia detected by biochemical newborn screening[49].

**Pseudo-dominance is a special form of recessive inheritance.** The occurrence of a recessive (biallelic) disease in successive generations is denoted pseudo-dominant inheritance (Fig. 2b). It is sometimes observed in consanguineous families (Fig. 2b, family 1; 50% recurrence risk for future siblings of the affected boy) or populations with high prevalence of recessive disease variants. It can also be caused by prevalent Hyp variants in the general population (Fig. 2b, family 2). For example, erythropoietic protoporphyria is mostly caused by compound heterozygosity for a *FECH* LoF variant and the common Hyp splice variant c.315-48T>C, which has an allele frequency of up to 33% in East Asians[50]. Thus, inheritance is recessive and not dominant; disease risk in children of a couple with an affected person and a partner with the heterozygous splice variant is 25% as homozygosity for the splice variant is benign. *PRPF31*-associated retinitis pigmentosa is often described as an autosomal dominant condition with incomplete penetrance because only some heterozygous individuals develop symptoms. However, detailed analyses showed that individuals heterozygous for a pathogenic variant remain asymptomatic only if they carry four copies of a regulatory minisatellite repeat element (MSR1) adjacent to the *PRPF31* promoter in *trans*. This 4-copy variant – which has an allele frequency of 15% in Europeans – confers markedly increased *PRPF31* expression compared with the 'normal' 3-copy allele[51]. It is open for discussion whether retinitis pigmentosa associated with *PRPF31* LoF should be regarded as a dominant disease, with absence of symptoms mediated by an increased-function variant in *trans*, or as recessive, with a common Hyp variant on one allele. Assuming that the four-copy MSR1 allele, which is more frequent in Europeans than in other populations[51], is evolutionarily more recent, it could potentially represent a compensatory mechanism to avoid haploinsufficiency in line with the concept proposed by Fisher[22].

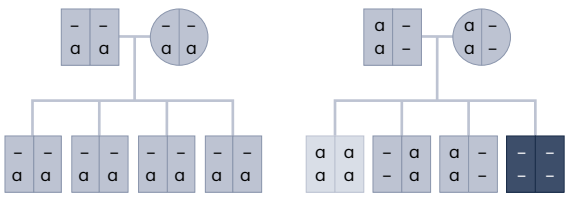
A related phenomenon has been reported for autosomal recessive *PEX6*-associated Zellweger syndrome. The *PEX6* gene has two common alleles: the reference allele that is actually the minor allele in African and European populations, and an allele with a 3′ untranslated region variant c.*442_445delTAAA that is associated with increased expression and over-representation among transcripts when in *trans* with




c

| Disease or trait name | Phenotype | Number of normal α-gene copies | Genotype constellation |
|---|---|---|---|
| | Normal | 4 | αα/αα |
| α-Thalassaemia minima (trait) | Asymptomatic | 3 | αα/α– |
| α-Thalassaemia minor (trait) | Mild microcytic anaemia | 2 | αα/-- (*cis* functional)<br>α–/α– (*trans* functional) |
| HbH disease | Microcytic anaemia, hepatosplenomegaly<br>Hb electrophoresis: HbH = β4 elevated | 1 | α–/-- |
| Hb Bart's hydrops fetalis syndrome | Fetal oedema, prenatal or postnatal lethal<br>Hb electrophoresis: Hb Bart's = γ4 | 0 | --/-- |




Different inheritance patterns in couples with α-thalassaemia minor

the reference allele. Because of the allelic expression imbalance, heterozygosity for the missense variant c.2578C>T (p.Arg860Trp, which also seems to have a dominant negative effect) on a c.*442_445delTAAA background causes Zellweger syndrome when in *trans* with the (effectively hypomorphic) reference allele, but is essentially asymptomatic when c.*442_445delTAAA is homozygous[52].

**Biallelic LoF variants in some recessive diseases are prenatal lethal.** In some conditions, homozygosity for LoF variants is rarely or never observed because it is usually incompatible with life. One example is thrombocytopenia-absent radius syndrome caused by compound heterozygosity for a *RBM8A* LoF variant (frequently a chromosomal microdeletion) and a Hyp regulatory variant (often c.–21G>A, allele frequency up to 3% in Europeans) that leads to a reduced amount of structurally normal gene product, the RNA-binding protein Y14 (ref.[53]). This constellation also explains why, despite recessive inheritance, thrombocytopenia-absent radius syndrome is not associated with consanguinity and (similar to other recessive diseases with predominantly compound heterozygous inheritance patterns) was not found through autozygosity mapping. Similarly, in autosomal recessive Smith–Lemli–Opitz syndrome, homozygosity for the prevalent DHCR7 LoF variant c.964-1G>C is less frequently observed in surviving affected individuals than expected based on its allele frequency, because of early miscarriage or fetal demise[54]. This phenomenon needs to be taken into consideration when carrier couples with LoF variants receive genetic counselling, as the probability of an affected live-born child is less than 25%.

**Quantitative variant effects underlie many digenic and polygenic traits.** Some traits, such as α-thalassaemia (Fig. 2c,d), are caused by quantitative variant effects in two genes; this is known as digenic inheritance and is discussed in more detail below (see 'Digenic inheritance and epistasis'). Similarly, many common disorders or quantitative traits are caused by a combination of quantitative variants in many different genes[55]; such conditions can be regarded as polygenic traits.

**Fig. 2 | Quantitative variant effects. a**, Quantitative variant genotypes in autosomal genes are often associated with a spectrum of functional severities determined by the combined quantitative effects of variants on each copy of the gene ('Percentage' denotes protein function, with normal function represented as 100%; simplified without consideration of modifying factors such as regulatory changes). Manifestation of the phenotype in a person heterozygous for a variant allele and the wild type (WT) allele defines inheritance as dominant, regardless of the clinical features or disease severity in the homozygote. Heterozygosity for a loss of function (LoF) variant (LoF/WT, typically 50% protein function) is often asymptomatic, reflecting recessive inheritance (lower left). In a semi-dominant constellation (lower centre), LoF heterozygotes usually show clinical symptoms (haploinsufficiency), but homozygotes (LoF/LoF) are more severely affected. Triplosensitivity indicates clinical manifestation if there are three copies of an autosomal gene (WT/WT/WT, shown as 150% protein function), for example in trisomy or microduplications; this is often — but not always — correlated with haploinsufficiency of the corresponding gene. True dominant diseases (in the definition used by Mendel) have similar clinical manifestation in heterozygotes and homozygotes (lower right); this phenomenon is not usually observed for quantitative variants although, per definition, embryonic lethal heterozygous traits may be regarded as true dominant. Hypomorphic (Hyp) variants on at least one gene copy are associated with variable residual function and attenuated disease manifestation. **b**, Pseudo-dominant inheritance: multiple consanguineous unions in the same family can sometimes lead to the occurrence of an autosomal recessive disease in successive generations (family 1). Family 2 illustrates pseudo-dominance associated with a prevalent Hyp variant — denoted as a functional polymorphism (Pol) because it has an allele frequency >1% — that is disease-causing in *trans* with a LoF variant, but usually asymptomatic in the homozygous state. Erythropoietic protoporphyria is an example of a disease that can be inherited in this pattern. Functional variants are indicated by white text with blue outline (dark blue for LoF variants, light blue for the Hyp Pol), affected individuals are indicated by blue circles (for females) or squares (for males; dark blue homozygous LoF, light blue for residual function Pol/LoF). **c**, α-Thalassaemia as a quantitative trait with digenic inheritance: the human genome contains two paralogous tandem α-globin genes on chromosome 16p13.3, *HBA1* and *HBA2*, which produce identical haemoglobin (Hb) α-chains[176]. Consequently, five different phenotypes are associated with the presence of zero to four functional alleles, not considering additional spectrum effects of Hyp alleles. The absence of one or two copies of these genes is asymptomatic or causes mild microcytic anaemia, usually described together as 'α-thalassaemia trait' but sometimes denoted as α-thalassaemia minima and α-thalassaemia minor, respectively. Absence of three copies causes HbH disease with hepatosplenomegaly and sometimes transfusion-dependent anaemia, whereas absence of all four *HBA1/2* copies results in severe prenatally lethal Hb Bart's hydrops fetalis syndrome. **d**, Couples in which each individual has only two functional *HBA1/2* copies (αα) may have very different disease risks for offspring, depending on whether the functional gene copies are in *trans* (left pedigree) or in *cis* (right pedigree). Two out of several possibilities are shown. Clinical severity is reflected by the colour (darker colour represents increased severity). The monogenic terms dominant and recessive are not suitable for α-thalassaemia, which in the absence of an adequate term is better regarded as a digenic trait.

## Qualitative variant effects

Many inherited diseases are caused by genetic variants that cause functional alterations rather than quantitative loss of the protein product. In one large study of de novo pathogenic variants in developmental disorders, it was estimated that 57–59% of de novo pathogenic missense or truncating variants had quantitative effects reflecting haploinsufficiency, and 41–43% acted by qualitatively altering function[56]. For the purpose of this Review, we wish to distinguish the following — sometimes overlapping — variant effects that are regarded as qualitative: gain of function effects (Table 1); dominant negative effects (Fig. 3); subfunction effects (Fig. 4a,b); moonlighting-function effects (Fig. 4c,d); and tissue-specific transcript effects.

**Gain of function encompasses a great variety of functional mechanisms.** Historically, variants that cause uncontrolled activation of the encoded protein (with the same amount of protein) were designated GoF variants, but the term was subsequently expanded to include other variant effects. In the medical genetics literature, 'gain of function' does not usually differentiate between simple quantitative gains in the 'same' activity and more complex qualitative alterations. However, conceptually it is useful to limit the term to variants that have abnormal or novel functional effects, as opposed to 'gain of quantity' variants with simple gene or protein dosage effects resembling those caused by increased gene copy number (discussed above in 'Triplosensitivity and haploinsufficiency are related'). Some variants cause complex combinations of qualitative and quantitative alterations that lead to specific cellular and clinical phenotypes, explaining different disease manifestations.

GoF mechanisms are often highly specific for the respective gene and the exact amino acid alteration of the encoded protein (Table 1). For example, uncontrolled cleavage of complement 1 subunit C1s triggered by missense and/or in-frame GoF variants in *C1R* or *C1S* causes autosomal dominant periodontal Ehlers–Danlos syndrome: intracellular autoactivation of the normally blocked serine protease domain leads to connective tissue alterations[57]. By contrast, biallelic LoF variants in *C1R* or *C1S* cause an autosomal recessive systemic lupus erythematosus-like syndrome that has no clinical overlap with the GoF phenotype. The GoF variant p.E325K in the *KLF1* gene alters the DNA-binding specificity of the encoded transcription factor and causes congenital dyserythropoietic anaemia type IV[58]. Autosomal dominant exercise-induced hyperinsulinism caused by a promoter variant in *SLC16A1* — which codes for the monocarboxylate (pyruvate/lactate) transporter 1 — is an example of pathogenic ectopic gene expression. The gene is normally silenced in pancreatic β-cells, but its abnormal expression caused by the promoter variants leads to increased insulin release and hypoglycaemia when blood lactate concentrations are increased, for example, after anaerobic exercise[59]. A conceptually similar process can affect enhancers; for example, variants in the zone of polarizing activity regulatory sequence almost 1 Mb away from the *SHH* gene cause ectopic *SHH* expression in the anterior margin of the developing limb bud leading to pre-axial polydactyly[60]. Disruption of normal splicing can lead to the expression of abnormal transcripts with GoF effects. For example, variants that interfere with normal splicing of the organ-specific exon 9 (IIIc) of the *FGFR2* gene cause Apert and Pfeiffer syndromes by driving ectopic expression of an alternative splice form (IIIb), leading to an autocrine signalling loop[61,62]. Another example is provided by intron 9 splice donor variants in the *WT1* gene in Frasier syndrome, which cause an imbalance of two alternative transcripts that generate protein isoforms varying by three amino acids; these isoforms differ in their relative transcriptional versus post-transcriptional activities[63]. The majority of nucleotide repeat expansions manifest through GoF mechanisms such as detrimental interaction with repeat RNA-binding proteins and aberrant 'repeat-associated non-ATG translation', for example in myotonic dystrophy[64], or generation of abnormal proteins that form insoluble aggregates, the main pathogenic mechanism in Huntington disease[65].

**Table 1 | Examples of different gain of function effects**

| Principle | Basic mechanism | Detailed mechanism | Disease/trait | Gene(s) | Refs. |
|---|---|---|---|---|---|
| Protein activation and/or loss of protein control | Ligand-independent signalling increase | Constitutive activation by intermolecular cross-linking or loss of negative regulation | Osteoglophonic dysplasia, encephalocutaneous lipomatosis | *FGFR1* | 171,172 |
| | Ligand-dependent signalling increase | Increased binding affinity for physiological or non-physiological ligands | Pfeiffer syndrome | *FGFR1* | 173 |
| | Uncontrolled enzyme function | Intracellular autoactivation of the normally blocked serine protease domain | Periodontal Ehlers–Danlos syndrome | *C1R, C1S* | 57 |
| | Uncontrolled ion channel function | Loss of gating | Paramyotonia congenita, hypokalaemic and hyperkalaemic periodic paralysis | *SCN4A* | 90 |
| | | | Long QT syndrome type 3 | *SCN5A* | 91 |
| | Transcription factor binding promiscuity | Mixed gain and loss of transcription factor binding specificity | Congenital dyserythropoietic anaemia type IV | *KLF1* | 58 |
| | Activation of other protein functions | Decrease in the activation threshold of the pyrin inflammasome | Familial Mediterranean fever | *MEFV* | 70 |
| Loss of expression control | Ectopic gene expression | Promoter activation | Exercise-induced hyperinsulinism | *SLC16A1* | 59 |
| | | Enhancer activation | Pre-axial polydactyly | *SHH* | 60 |
| | Alteration of splicing | Disruption of alternative splicing | Apert and Pfeiffer syndromes | *FGFR2* | 61 |
| | | | Frasier syndrome | *WT1* | 63 |
| | Alteration of topologically associating domains | Novel regulatory landscape, enhancer adoption | Acropectoral syndrome | *SHH* | 153 |
| Non-specific effects of abnormal gene product | Abnormal mRNA effects | Detrimental interaction with repeat RNA-binding proteins, aberrant repeat-associated non-ATG translation | Myotonic dystrophy | *DMPK, CNBP* | 64 |
| | Toxic protein effect | Coding triplet repeat expansion (polyglutamine disorders) | Huntington disease | *HTT* | 69 |
| | | Protein aggregation disorders (amyloidoses) | Hereditary transthyretin-related amyloidosis | *TTR* | 66 |
| Other functional effects | Novel protein function | Different substrate binding based on size of active centre | ABO blood groups | *ABO* | 71 |

Note that this table is not exhaustive, and additional gain of function (GoF) mechanisms are well recognized, for example in tumour development.

Structural variants that trigger direct or indirect protein aggregation and amyloid formation cause a range of autosomal dominant GoF diseases, such as transthyretin amyloidosis[66]. GoF is also a key driver of allelic series (see below 'Allelic series').

Autosomal GoF variants are usually associated with clinical manifestation in the heterozygote, as normal function or regulation of the WT protein is often unable to counteract the abnormal variant. The effect is not usually true dominant in the original Mendelian definition, as the homozygous state — when observed — is generally more severe than the heterozygous state[2,67]. An example is achondroplasia caused by the heterozygous activating *FGFR3* variant p.Gly380Arg; homozygosity for this variant is usually associated with prenatal or perinatal lethality. Huntington disease is one of the very few known diseases with largely similar clinical manifestation in heterozygotes and homozygotes, representing true dominance[68,69]. Occasionally, heterozygous GoF variants may be asymptomatic, with clinical manifestation predominantly in the biallelic state. This is the case with familial Mediterranean fever caused by *MEFV* variants that decrease the activation threshold of the pyrin inflammasome[70]. The term co-dominance refers to the phenomenon whereby different alleles of a gene encode functionally distinct proteins that cause alternative phenotypes, both of which can be recognized in the compound heterozygous state. The best-known example is the blood group AB caused by compound heterozygosity of the A and B alleles of the *ABO* gene, each of which codes for fully active but functionally different enzymes[71].

In rare circumstances, heterozygous individuals may have a clinically more adverse phenotype than WT or variant homozygotes. This phenomenon is denoted underdominance, negative overdominance, or metabolic interference, and may occur if two different alleles of a gene produce different proteins with adverse interaction effects[72]. It is related to — but conceptually different from — cellular interference in X-chromosomal traits, whereby interference is between rather than within cells (see below 'Gonosomal and mitochondrial inheritance'). Underdominance is not well recognized in humans. One example involves myocilin, which is encoded by the *MYOC* gene: variants in this gene may cause glaucoma in the heterozygous but not homozygous state[73]. Absence of myocilin is asymptomatic, and pathogenesis in *MYOC*-associated glaucoma involves GoF variant protein accumulation causing endoplasmic reticulum stress[74], which might be enhanced or exclusively present in heterozygous individuals.

**Dominant negative effects should be considered for multimeric proteins.** A dominant negative pathogenic mechanism involves direct interference with or blockage of WT function by the variant, for example by disrupting the formation or function of homotypic or heterotypic multimers or other interactions involving the WT protein and

other molecules (Fig. 3a). The clinical consequences exceed those of the heterozygous loss of one gene copy, and sometimes resemble biallelic LoF variants. Dominant negative effects have a major role in variants of structural proteins or multimeric channel proteins but are not usually expected for variants of proteins that act as monomers. A well-known structural protein example is collagen I, where the variant effects differ for the α1-chains and α2-chains that have a 2:1 ratio in the final triple helix (Fig. 3b). Dominant negative variants can be distinguished conceptually from toxic protein variants, in that the toxic protein causes the adverse effect on its own, irrespective of the WT protein, whereas in the dominant negative context the aberrant protein compromises the function of the normal protein. The same principle applies to toxic RNA variants.

**Variants can differentially affect protein subfunctions.** Some proteins have sequential functions in cellular processes — such as sequential steps in an enzymatic reaction or transport processes — that may be differentially affected by genetic variants (Fig. 4a). One example is apolipoprotein B (ApoB)-mediated lipid transport: variants in the *APOB* gene that prevent or reduce low density lipoprotein (LDL) production cause reduced blood cholesterol concentrations, whereas variants that interfere with LDL receptor (LDLR) binding cause familial hypercholesterolaemia (Fig. 4b). A similar mechanism underlies hawkinsinuria, an autosomal dominant disease caused by the missense variant c.722A>G (p.Asn241Ser) in the *HPD* gene, which codes for 4-hydroxyphenylpyruvate dioxygenase in tyrosine (Tyr) breakdown. The variant inhibits one step of the complex HPD reaction and leads to the production of hawkinsin, an unusual sulfur-containing amino acid[75]. Complete or attenuated general deficiency of HPD function causes autosomal recessive tyrosinaemia type 3 without accumulation of hawkinsin.

**Moonlighting functions may explain diverse variant effects.** Some proteins have taken on several roles in different pathways such as metabolism, gene regulation or signal transduction, with the non-canonical functions denoted as moonlighting functions[76,77]. As a result, genetic variants that selectively affect only one particular function of the protein may differ fundamentally in their phenotypic effect from LoF variants that completely remove all of its functions (Fig. 4c). Inheritance of a phenotype may be variant-specific and function-specific dominant or recessive, depending on whether the loss of the affected function can be compensated in the heterozygous state. Moonlighting differs from pleiotropy, which usually refers to the relevance of one particular protein function for different cellular processes or pathways. A well-characterized X-chromosomal example of a moonlighting protein is the mitochondrial 17β-hydroxysteroid dehydrogenase type 10 (HSD10) protein encoded by the *HSD17B10* gene (Fig. 4d). As a homotetramer, it functions as a 2-methyl-3-hydroxybutyryl-CoA dehydrogenase in isoleucine breakdown; it also serves as a non-enzymatic scaffold for the RNaseP complex required for mitochondrial DNA (mtDNA) transcript processing. Different *HSD17B10* variants differentially affect both functions, with complete protein loss incompatible with life due to RNaseP disruption. This type of gene and/or protein functional complexity illustrates the difficulty in dissecting the effects of 'pathogenic' variants on different protein functions in genetic analyses.

**Transcription differences modify the impact of some variants.** Alternative splicing and differential start codon usage allow the production of different proteins from the same gene and can regulate organ-specific gene functions[78]. The relevance of this mechanism for understanding inherited diseases and cancer is becoming increasingly recognized[79]. Pathogenic variants that affect only some transcripts may cause attenuated or atypical clinical phenotypes. For example, general deficiency of the cytoskeletal linker protein plectin causes an autosomal recessive multisystem disease with epidermolysis bullosa and variable other manifestations including pyloric atresia and muscular dystrophy, whereas clinical features in individuals with pathogenic variants in the skin-specific exon 1a are limited to the skin[80]. Homozygous LoF variants in the breast cancer predisposition gene *BRCA1* are not usually compatible with life but several individuals with a Fanconi syndrome subtype had homozygous nonsense variants in the distal part of the large exon 11. The attenuated phenotype is mediated by an alternative splice donor sequence within exon 11, which allows the production of a shorter, partially functional transcript without the pathogenic variant[81]. A similar mechanism has been reported for BRCA2 (ref. [82]). Heterozygous single-nucleotide variants in the promoter 1B of the adenomatous polyposis coli (*APC*) gene, associated with isoform 1B of the APC protein and predominantly expressed in the gastric mucosa, cause gastric adenocarcinoma and proximal polyposis of the stomach but rarely lead to familial adenomatosis polyposis[83]. Complete loss of UDP-glucose pyrophosphorylase (*UGP2*) gene function seems to be incompatible with life, whereas pathogenic variants selectively affecting the start codon of a brain-specific isoform cause a severe autosomal recessive developmental and epileptic encephalopathy[84]. Similar effects may be expected for genes with different transcripts that correspond to different targeted organelles, such as fumarate hydratase[85].

## Allelic series

Considering the often diverse and variable effects of genetic alterations on the encoded protein, it is not surprising that functional variants in the same gene may manifest in a range of very different — sometimes opposite — phenotypes. This phenomenon is called an allelic series. Sometimes different variants affect the same cellular function, reflecting for example a reduced versus increased activity spectrum. An example is glucokinase (*GK*) variants that cause a spectrum of reduced through to increased insulin secretion, respectively resulting in dominant or recessive monogenic diabetes mellitus versus dominant hyperinsulinism[86]. By contrast, recessive LoF and dominant GoF variants in the *FAR1* gene — encoding peroxisomal fatty acyl-CoA reductase 1 required for plasmalogen biosynthesis — cause opposite biochemical phenotypes but overlapping clinical manifestations including developmental delay, spastic paresis, seizures and cataracts[87]. Heterozygous activating variants of proteins involved in growth stimulation may represent oncogenic drivers or inherited risk factors for tumour development, whereas LoF variants predominantly affect morphogenesis. For example, inherited or somatic activating variants in the *RET* proto-oncogene trigger the development of thyroid carcinomas and other malignancies[88], whereas LoF variants are risk factors for Hirschsprung aganglionosis. Some variants associated with co-segregation of both conditions in the same family have been called 'Janus mutations'[89]. Variants in the fibroblast growth factor receptor 1 (*FGFR1*) gene have different effects on protein characteristics and signalling function, which are reflected in a range of consequent phenotypes (Fig. 5). Allelic series with often overlapping phenotypes have been described for many channelopathies. Notable examples are the sodium channels in skeletal and cardiac muscle that are encoded by *SCN4A* and *SCN5A*, respectively. Different *SCN4A* variants cause enhanced or impaired channel activation, and in consequence myofibre hyperexcitability or hypoexcitability, leading to a phenotypic spectrum from myotonia to

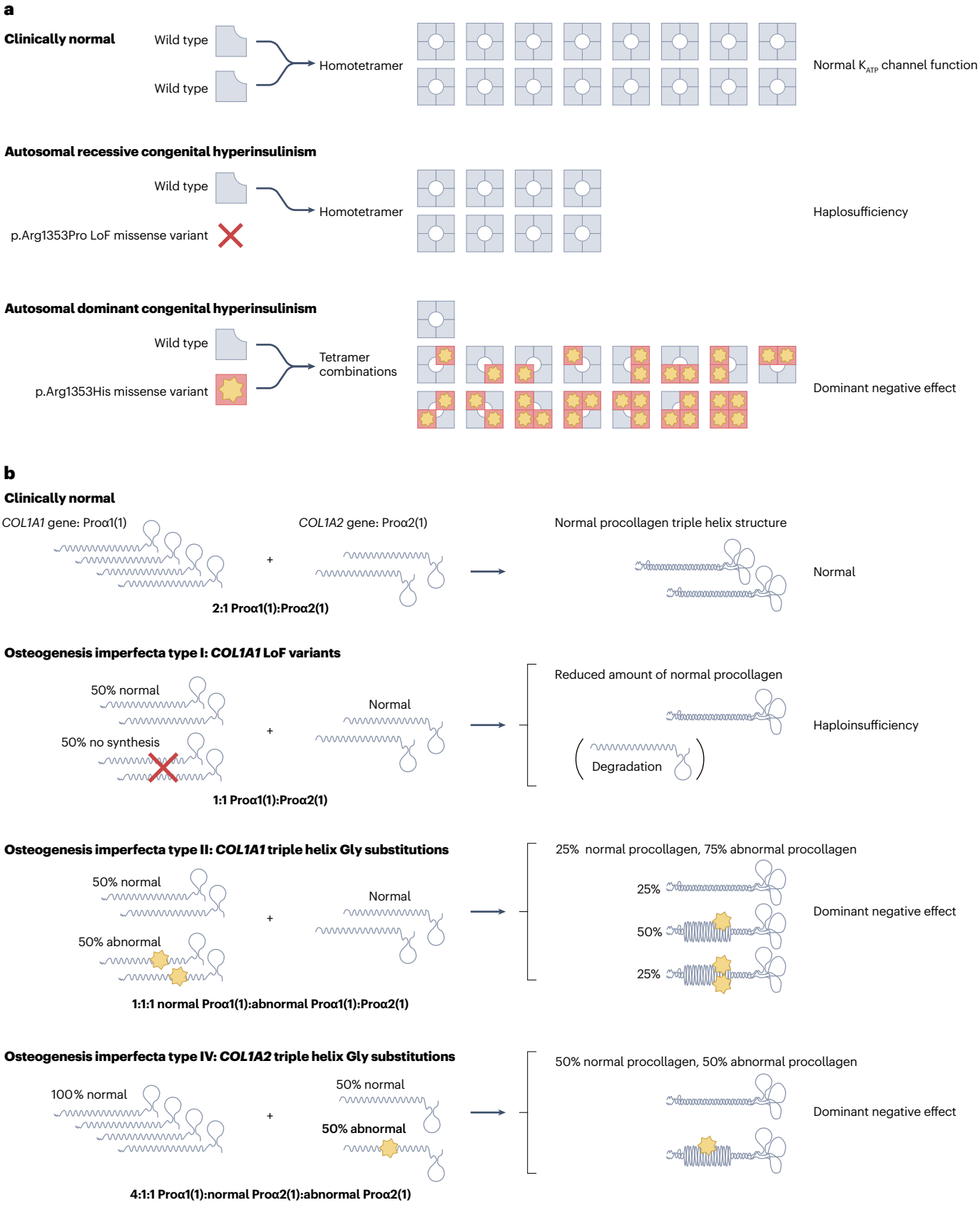

a
Clinically normal
Wild type
Wild type
Homotetramer
Normal $K_{ATP}$ channel function
Autosomal recessive congenital hyperinsulinism
Wild type
p.Arg1353Pro LoF missense variant
Homotetramer
Haplosufficiency
Autosomal dominant congenital hyperinsulinism
Wild type
p.Arg1353His missense variant
Tetramer combinations
Dominant negative effect
b
Clinically normal
COL1A1 gene: Proα1(1)
COL1A2 gene: Proα2(1)
2:1 Proα1(1):Proα2(1)
Normal procollagen triple helix structure
Normal
Osteogenesis imperfecta type I: COL1A1 LoF variants
50% normal
50% no synthesis
Normal
1:1 Proα1(1):Proα2(1)
Reduced amount of normal procollagen
Degradation
Haploinsufficiency
Osteogenesis imperfecta type II: COL1A1 triple helix Gly substitutions
50% normal
50% abnormal
Normal
1:1:1 normal Proα1(1):abnormal Proα1(1):Proα2(1)
25% normal procollagen, 75% abnormal procollagen
25%
50%
25%
Dominant negative effect
Osteogenesis imperfecta type IV: COL1A2 triple helix Gly substitutions
100% normal
50% normal
50% abnormal
4:1:1 Proα1(1):normal Proα2(1):abnormal Proα2(1)
50% normal procollagen, 50% abnormal procollagen
Dominant negative effect

**Fig. 3 | Dominant negative effects. a**, $K_{ATP}$ channel-related congenital hyperinsulinism: blood sugar regulation by pancreatic β-cells depends on the normal function of an octameric $K_{ATP}$ channel that consists of four Kir6.2 proteins and four SUR1 proteins, respectively encoded by the *KCNJ11* and *ABCC8* genes (top; Kir6.2 not depicted for simplicity). ATP-mediated closure of the $K_{ATP}$ channel causes insulin release. Pathogenic *ABCC8* variants include two missense variants at the same amino acid residue that are associated with different inheritance patterns of congenital hyperinsulinism. p.Arg1353Pro is a non-functional recessive loss of function (LoF) missense variant that in the heterozygous state leads to a reduced amount of normal protein, which is sufficient for normal function (haplosufficiency; middle). By contrast, p.Arg1353His results in a stable abnormal SUR1 protein that interferes with the wild type (WT) protein and has a dominant negative effect: statistically, only 1/16 channels consist of 4 WT SUR1 proteins required for normal channel function (bottom)[177]. **b**, Osteogenesis imperfecta: type 1 collagen, the main structural component of connective tissue, has a fibrillar structure based on a heterotrimer composed of two Proα1(1) chains encoded by *COL1A1* and one Proα2(1) chain encoded by *COL1A2* (top). These molecules are proteolytically converted to collagen after secretion by removing the propeptides. Variants in each type 1 collagen gene result in osteogenesis imperfecta or brittle bone disease. The clinical effects depend on the affected gene and, for missense variants, on the location of the substitution and the substituting amino acid. The variant effect is amplified in the process of triple helix cross-linking and the formation of fibrils in the extracellular matrix, with the abnormal molecules disrupting the overall function and structure of the collagen fibres. Heterozygous LoF variants in *COL1A1* result in the production of half the amount of normal, fully functional type 1 collagen, which leads to osteogenesis imperfecta type I, a relatively mild form of osteogenesis imperfecta (haploinsufficiency, second row). By contrast, missense variants that interfere with triple helix formation cause more severe phenotypes through dominant negative effects. *COL1A1* glycine (Gly) substitutions at the repetitive Gly-Xaa-Yaa motif required for triple helix formation typically cause lethal osteogenesis imperfecta type II (third row). *COL1A2* Gly changes are rarely lethal and usually produce the moderate osteogenesis imperfecta type IV phenotype (bottom)[178].

muscle weakness[90]. Similarly, based on channel kinetics, *SCN5A* variants can cause different cardiac manifestations such as long QT syndrome type 3 (GoF variants), Brugada syndrome or sick sinus syndrome (LoF variants), as well as overlap syndromes[91].

### Complex mechanisms

**Being heterozygous may have selective advantages.** Heterozygous variants in some genes can have advantageous clinical effects not observed in homozygous WT or variant states. This phenomenon — also denoted overdominance — is well recognized as a likely explanation for the high prevalence of certain autosomal recessive diseases such as the haemoglobinopathies, cystic fibrosis or phenylketonuria[92–94]. Heterozygous *CFTR* variants, which cause cystic fibrosis when homozygous, may protect against severe fluid loss in gastrointestinal infections — a beneficial overdominant effect — but also increase the risk for some disease manifestations such as bronchiectasis or pancreatitis[95]. Thus, a pathogenic variant in a recessive disease gene may formally relate to more than one trait with different inheritance patterns and manifestation probabilities influenced by other genetic and exogenous factors. The balance between beneficial and adverse effects of heterozygote and homozygote status guides evolutionary adaptation in changing environments.

**Anticipation is not restricted to repeat expansion disorders.** Earlier or more severe manifestation of some hereditary diseases in the offspring than in the parent — denoted anticipation — has long been recognized but was frequently argued to be due to ascertainment bias[96]. It is now well established that expansions of unstable repeat sequences (examples in Table 1) lead to anticipation in repeat diseases. Anticipation unrelated to unstable repeats is observed in heritable disorders of telomere maintenance (telomeropathies)[97]. This was first shown for autosomal dominant dyskeratosis congenita type 1 caused by pathogenic variants in *TERC*, the gene encoding the telomerase non-coding RNA component: failure to restore normal telomere length leads to progressive telomere shortening and increased disease severity over successive generations[98]. This mechanism has since been described in other telomeropathies[99].

**Epigenetic factors interfere with Mendelian inheritance patterns.** Non-Mendelian monogenic inheritance patterns beyond dominance and recessiveness are observed for diseases linked to imprinted genes with monoallelic parent of origin-dependent expression[100]. One example is Angelman syndrome, a severe neurodevelopmental disease linked to the *UBE3A* gene on chromosome 15q11.2, which in the brain is expressed only from the maternal allele[101]. Pathogenic LoF variants cause disease only when inherited from the mother, and can be transmitted asymptomatically along the male line over many generations. Another example is the paternal dominant inheritance observed in some tumour predispositions (see below 'Tumour predisposition syndromes').

## Disease variation beyond single genes

It is well recognized that individuals with the 'same' monogenic disease may show quite divergent clinical features[102]. The terms penetrance and expressivity, respectively, are used for the phenomena that not all individuals with a particular gene variant develop the associated disease and that individuals may differ in the specific manifestation. In many cases this is due to the different molecular effects of allelic variants and biallelic genotype combinations, as outlined above. Careful functional assessment of the identified variants may allow the delineation of different disease groups linked to the same gene[103]. In other instances, special monogenic circumstances such as mosaicism have a central role[104,105]. More frequently, however, the combined effects of all variants (including regulatory variants[106]) in a 'causative' gene do not explain the phenotypic variability observed in the individual. In such cases, other relevant factors need to be considered, including modifier genes or digenic inheritance, common familial and/or polygenic elements[107–109], epigenetic alterations, environmental influences and chance. Combining ultra-rare high-penetrance variants with more common functional variants may allow improved calculation of polygenic scores for individual risk assessment in monogenic and multifactorial traits[110,111]. From a conceptual point of view, the terms penetrance and expressivity are best used for factors that are not linked to the primary disease gene and cannot be determined by the comprehensive characterization of that gene, and thus represent modifying factors beyond monogenic inheritance. In particular, the terms should not be used to describe the variable clinical presentation caused by functionally different variants in the same gene.

### Digenic inheritance and epistasis

There is an increasing number of examples of digenic inheritance, that is diseases that manifest only with pathogenic variants in two genes[112,113].

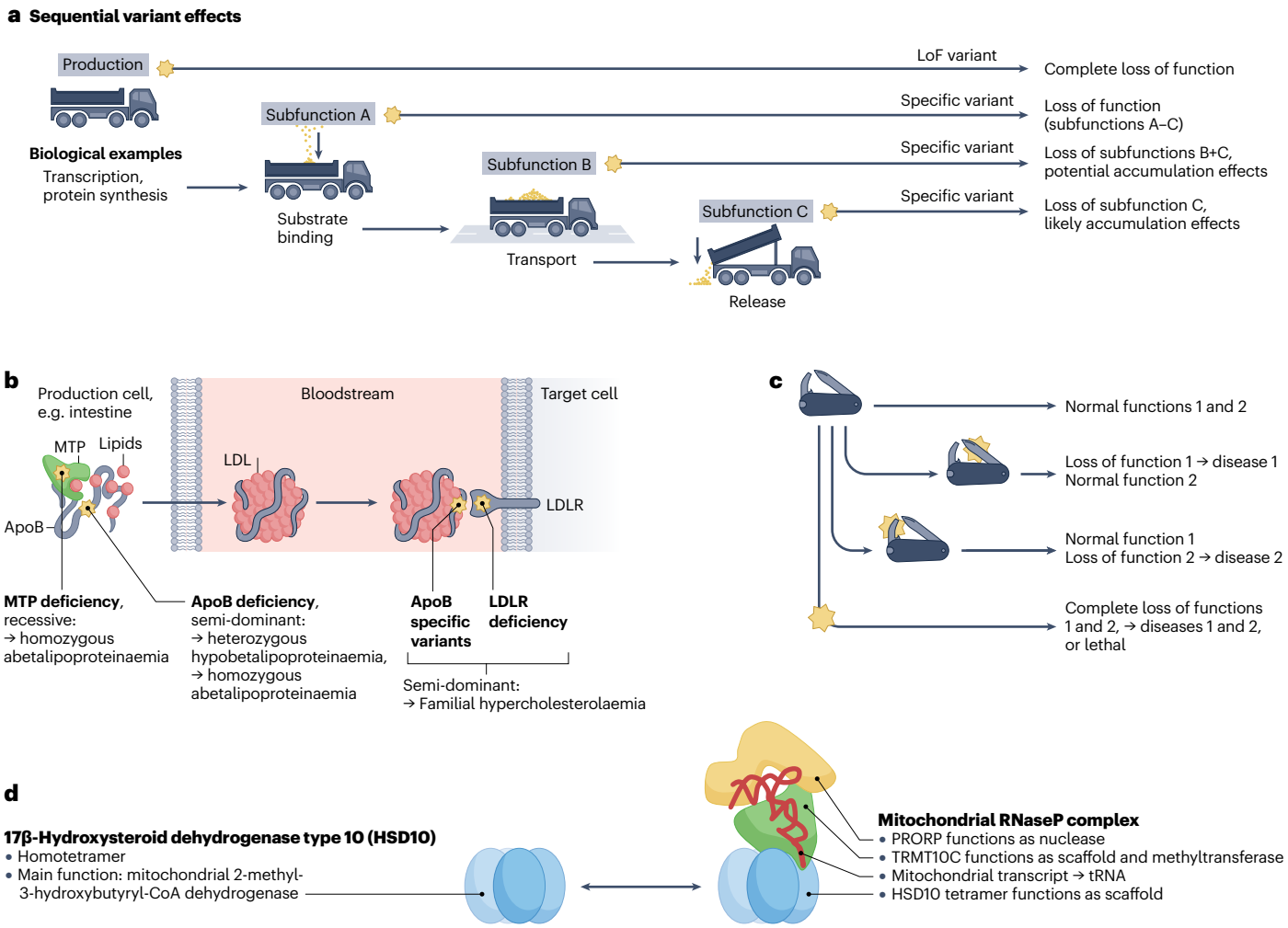



**Fig. 4 | Protein subfunction and moonlighting-function effects. a**, Variants may differentially affect production or sequential subfunctions of a protein, leading to specific functional alterations depicted here for a putative transport protein (subfunctions A–C). Different variants may be associated with different inheritance patterns. Clinical consequences of heterozygous loss of function (LoF) variants that prevent protein synthesis depend on possible haploinsufficiency. Transport function is also lost for variants that interfere with substrate binding or transport itself. By contrast, deficient release of the cargo is likely to cause variant-specific accumulation effects, which may be observed in the heterozygote. **b**, An example for sequential variant effects is apolipoprotein B (ApoB), a major component of low density lipoproteins (LDLs) that transport lipids and fat-soluble vitamins between organs and cells. It serves as a structural component, a ligand for cell surface receptors and a cofactor for lipid-metabolizing enzymes. Variants in the *APOB* gene can differentially affect these functions, with variable inheritance patterns. Complete ApoB deficiency due to homozygous *APOB* LoF variants results in abetalipoproteinaemia. This condition may also be caused by homozygous LoF variants in the microsomal transfer protein (*MTP*) gene. Heterozygous LoF variants of *APOB* cause autosomal semi-dominant familial hypobetalipoproteinaemia with reduced cholesterol, triglyceride and ApoB concentrations in blood, which is mostly asymptomatic[179]. By contrast, *MTP* LoF heterozygotes have no clinical or biochemical abnormalities[180]. ApoB deficiency (semi-dominant) and MTP deficiency (recessive) are phenotypically indistinguishable in homozygotes but can be differentiated in heterozygotes, representing two different inheritance patterns. Some missense variants in the LDL receptor (LDLR) binding domain of ApoB have the opposite effect of LoF variants or most other pathogenic *APOB* variants: they do not affect lipoprotein assembly or secretion but prevent its uptake by the liver. The consequence is autosomal dominant familial hypercholesterolaemia, the same phenotype as caused by LoF variants in the *LDLR* gene[180]. **c**, Some genes code for moonlighting proteins with unrelated functions in different cellular processes. This can be imagined as a Swiss pocketknife with different purposes that can be altered individually or lacking in combination. Complete loss of the protein, or alteration of specific functions by qualitative variants, may result in different diseases with potentially different inheritance patterns. **d**, The mitochondrial protein 17β-hydroxysteroid dehydrogenase type 10 (HSD10; SDR5C1), encoded by the *HSD17B10* gene, is an example of a moonlighting protein. HSD10 serves as mitochondrial 3-methyl-2-hydroxybutyryl-CoA dehydrogenase (MHBD) in isoleucine breakdown, and as a non-enzymatic component of mitochondrial RNaseP in mitochondrial DNA (mtDNA) transcript processing. The two different functions are thought to have had different evolutionary constraints, leading to optimization of the essential structural role while maintaining a relatively non-specific enzymatic function[181]. Deficiency of the MHBD activity causes accumulation of certain urinary organic acids[182], but the neurodegenerative clinical features of HSD10 disease are due to deficient RNaseP. *HSD17B10* variants can differentially affect both functions, and urinary organic acid analysis is unable to reliably confirm or exclude pathogenic effects on RNaseP function in suspected patients[183,184]. The representation of HSD10/RNaseP is based on the structure reported in ref. [185].

For example, α-thalassaemia has a unique non-Mendelian inheritance pattern that cannot be adequately described with the terms dominant and recessive, and is best regarded as digenic (Fig. 2c,d). Similarly, hyperbilirubinaemia in Rotor syndrome only occurs with deficiency of all four copies of two adjacent genes on chromosome 12p12.1, *SLCO1B1* and *SLCO1B3*, which show 87% cDNA sequence (80% amino acid) identity (plus short different carboxy-terminal tails) and code for functionally similar liver-specific organic anion transporters. Inheritance of Rotor syndrome is regarded as digenic recessive[114], although it usually resembles 'Mendelian' recessive inheritance because of tight linkage between the genes that are sometimes both disrupted by a single large deletion.

The term epistatic was coined by Bateson for non-allelic factors that suppress the manifestation of a particular so-called hypostatic trait[115]. It is now used for various types of gene–gene interactions in the production of a phenotype[116,117] and the variable manifestation of monogenic diseases depending on other genes[113]. Whereas the term digenic implies a more or less equal role of two genes in disease

## Glossary

### Anticipation
The earlier or more severe manifestation of a hereditary disease in the offspring than in the parent.

### Allelic series
Different variants in the same gene often acting through different pathogenetic pathways cause a range of phenotypes that may be associated with different inheritance patterns.

### Autozygosity mapping
A method to identify disease-causing variants in consanguineous families by focusing on autosomal regions with runs of homozygous genotypes inherited from a shared ancestor (that is, autozygous regions).

### Co-dominance
Different alleles of the same autosomal gene yield functionally distinct proteins with alternative phenotypes, both of which can be recognized in the (compound) heterozygous state.

### Compound heterozygous
Different pathogenic variants on the two alleles of the same autosomal gene, causing biallelic loss or modification of gene function.

### Dominant negative effect
A heterozygous variant codes for a structurally altered protein that interferes with the wild type (WT) protein.

### Exome
The transcribed sequences of all protein-coding human genes.

### Expressivity
The severity or extent of clinical manifestation in persons with a particular genotype. Variable expressivity indicates that individuals with the same genotype (for example, in the same family) may have quite different phenotypes.

### Genotype
The constellation of genetic variants in a particular gene or in the genome.

### Gain of function (GoF) variant
A genetic variant that causes inappropriate or novel protein functions such as uncontrolled activation or loss of regulation of the encoded protein, ectopic and/or illegitimate organ and/or cell-specific expression patterns, or novel (including toxic) protein or mRNA functions.

### Gonosomal
A gene or variant on one of two sex chromosomes, as opposed to autosomes.

### Haplosufficiency
Complete loss of one copy of a particular autosomal gene is usually asymptomatic. Haplosufficiency is a typical feature of recessive diseases and Mendelian wild type (WT) dominance.

### Haploinsufficient
Complete loss of one copy of a particular autosomal gene causes noticeable clinical effects; a single (haploid) normal allele does not suffice for normal development or homeostasis. Haploinsufficiency is the central pathogenetic mechanism in semi-dominant diseases caused by loss of function (LoF) variants.

### Hypomorphic (Hyp) variants
Genetic variants that reduce but do not completely abolish the function of the encoded protein.

### Loss of function (LoF) variant
A genetic variant that causes complete loss of the encoded protein. Also known as a null variant.

### Moonlighting-function effects
A multifunctional protein performs two or more autonomous, independent and mechanistically different functions; specific impairment of one function by a pathogenic variant does not necessarily affect the other function(s).

### Overdominance
The phenomenon whereby some heterozygous variants have advantageous clinical effects not observed in homozygous wild type (WT) or homozygous variant states.

### Penetrance
The probability of clinical manifestation in persons with a particular genotype, which is frequently age-dependent. Penetrance may be complete (100%) or incomplete (reduced).

### Phenotype
The measurable consequences of a genetic variant on the protein, cell, organ or clinical level.

### Pleiotropy
A protein function is required in two or more different cellular processes or pathways; pathogenic variants in the respective gene cause different seemingly unrelated phenotypic traits.

### Pseudo-dominant
The occurrence of an autosomal recessive (biallelic) disease in successive generations.

### Semi-dominant
Describes the phenomenon whereby an autsomal genetic variant in its heterozygous state is associated with a less severe, intermediate phenotype than its homozygous state. Most quantitative variants are semi-dominant, but semi-dominance is also commonly observed for variants that have qualitative effects.

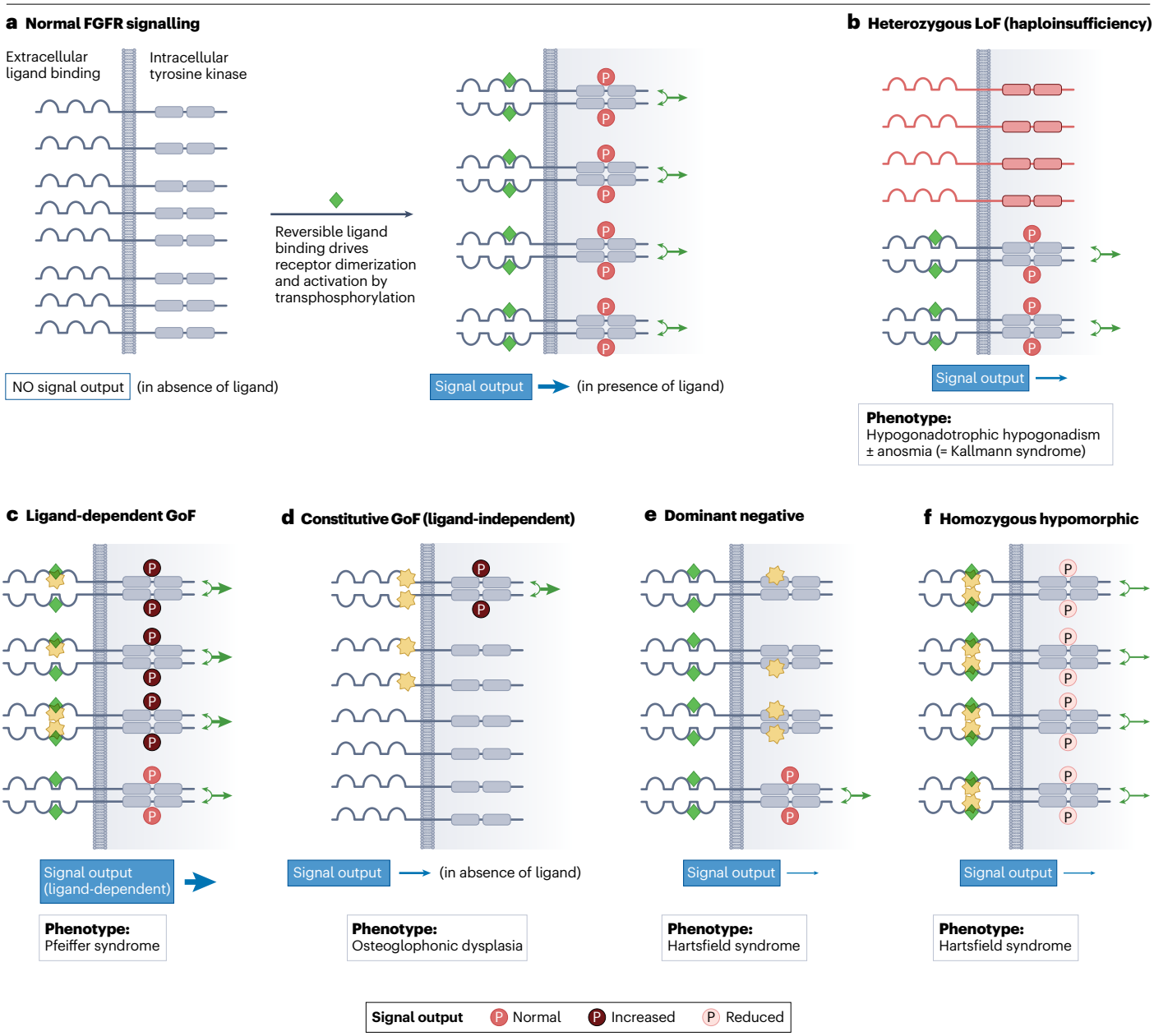
### Subfunction effects
A protein has successive functions in cellular processes, such as sequential steps in an enzymatic reaction or transport processes, which are differently affected by genetic variants.

### Triplosensitive
An additional copy of a particular autosomal gene (for example through a duplication that causes three instead of two gene copies) that has adverse clinical consequences.

### Tissue-specific transcript effects
A gene shows functional variability in expression or splicing, which is differently affected by genetic variants.

### Underdominance
The phenomenon whereby some heterozygous variants cause a clinically adverse phenotype that is absent or less severe in wild type (WT) or variant homozygotes.

### Wild type
(WT). The genetic sequence or function defined as normal.




development, the term epistasis describes how genetic modifiers influence the phenotype associated with a primary gene. The traditional human example of epistasis is the Bombay blood group, which is observed when expression of the *ABO* alleles is prevented by the autosomal recessive deficiency of fucosyltransferase 1, the protein that catalyses the immediately preceding enzymatic step[118]. Epistasis may also account for discrepant dominant and recessive inheritance patterns for the same disease. For example, spastic paraplegia type 7 is usually caused by biallelic pathogenic variants in the *SPG7* gene but may also be caused by a heterozygous *SPG7* variant in combination with variants in genes coding for SPG7 interaction partners[119]. A similar effect, 'synergistic heterozygosity', has been suggested for some inherited metabolic phenotypes that reflect the combined effect of several heterozygous, normally recessive variants in genes linked to the same metabolic pathway[120]. This concept mirrors the proposed combined gene dosage effects of interacting proteins in macromolecules[34] (see 'Haploinsufficiency is associated with highly regulated cellular functions'). With regard to complex traits, however, genetic variation seems to be mostly due to additive effects, with little evidence for epistasis[121,122].

## Tumour predisposition syndromes

**Genetic risk versus genetic disease.** Most cancer risk syndromes caused by heterozygous variants in tumour suppressor genes run in families as autosomal dominant traits with reduced penetrance and variable expressivity: that is, not all individuals with a heterozygous pathogenic variant develop a tumour in their lifetime, and the manifestation varies between affected persons. It is not the particular tumour

**Fig. 5 | Allelic variants in the fibroblast growth factor receptor 1 gene.** Fibroblast growth factor receptor 1 (FGFR1) is a tyrosine kinase (TK) cell surface receptor that dimerizes in response to binding extracellular fibroblast growth factors (FGFs) in a 2:2 multimeric complex to activate intracellular signalling pathways leading to cell differentiation, proliferation, migration and other functions. Somatic activating variants, including gene amplifications and fusion proteins, are involved in various malignant tumours. Sequence variants in *FGFR1* result in several different clinical phenotypes that depend on the nature of the alteration in the protein. **a**, Simplified representation of normal FGFR1 signalling[186]: FGFR1 monomers (eight are shown) span the cell membrane. The extracellular domains are structurally related to immunoglobulins and bind FGF ligands (green diamonds). The intracellular domains have TK activity. In the absence of FGF (left), receptors exist as monomers, the TK domains remain unphosphorylated and signal output is zero. In the presence of FGF (right), a productive 2FGF:2FGFR1 complex is generated that activates the kinase and results in *trans*-phosphorylation (small double-headed green arrows and annotated P). This leads to TK activation and downstream signalling (blue arrows). Total signal output yielded by the eight monomers is indicated by the thickness of the blue arrow. **b**, Haploinsufficiency: heterozygous loss of function (LoF) variants (pink monomers) caused by diverse molecular lesions (complete or partial deletions, nonsense or frameshift variants, or missense variants resulting in protein instability) reduce signal output by 50%. This is tolerated in most developmental contexts but causes hypogonadotropic hypogonadism (sometimes accompanied by anosmia, defining Kallmann syndrome) because of the particular dosage sensitivity to FGF signalling of the embryonic olfactory and gonadotrophin-releasing hormone neurons during development[187]. **c**, Ligand-dependent gain of function (GoF): a specific heterozygous missense variant, p.Pro252Arg (yellow stars), within the extracellular domain leads to enhanced binding affinity for a specific repertoire of FGF ligands, and hence enhanced ligand-dependent signalling[173]. The associated phenotype, Pfeiffer syndrome, is characterized by craniosynostosis and medially deviated first digits. **d**, Constitutive activation: heterozygous GoF variants (yellow stars) close to the transmembrane domain lead to constitutive activation in the absence of ligand, potentially by covalent dimerization of monomers. This leads to ectopic activation in normally quiescent cells and results in osteoglophonic dysplasia, a severe congenital dysplasia of the long bones and skull associated with secondary bony changes[171]. Other distinct heterozygous pathogenic variants that cause autoactivation of the TK domain in the absence of ligand occur in the mosaic state in encephalocraniocutaneous lipomatosis; such variants are presumed lethal in the constitutionally heterozygous state[172]. **e**, Dominant negative: localized heterozygous missense variants (yellow stars) within the TK domains can interfere with normal activation. In consequence, three quarters of these dimers are locked into unproductive complexes that remain unphosphorylated. Signal output is reduced by up to 75%, which causes dominantly inherited Hartsfield syndrome with brain malformation (holoprosencephaly), split-hand/foot malformation (ectrodactyly), cleft lip/palate, and variable other features such as intellectual disability and, sometimes, hypogonadotropic hypogonadism[188]. **f**, Homozygous hypomorphic: missense variants (yellow stars) that reduce, but do not eliminate, FGFR1 function (potentially by reduction in FGF binding affinity or protein stability) may diminish TK activation and total signalling output to a similar extent as the dominant negative scenario, leading to a recessive type of Hartsfield syndrome[189].

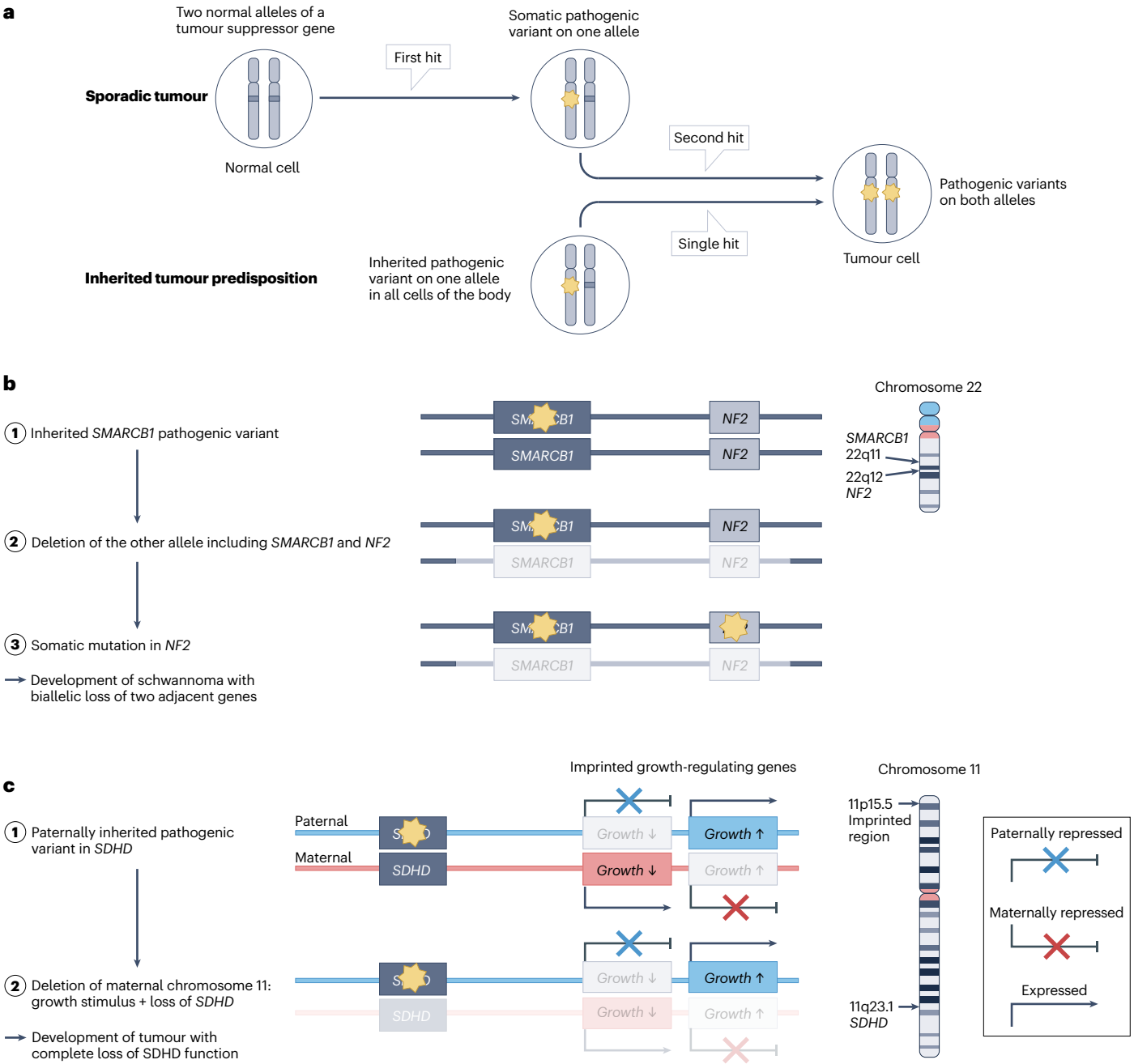
that is inherited as a genetic disease, but the increased probability of tumour development that is influenced by other genetic and environmental factors as well as by chance somatic events. This concept is reflected in the two-hit hypothesis developed by Alfred G. Knudson based on the statistical study of retinoblastoma[123] (Fig. 6a) and in the more complex digenic three-step pathogenesis involving biallelic loss of two adjacent genes in the development of schwannomas (Fig. 6b). The continuum between monogenic and polygenic risk factors is an increasingly recognized challenge for the determination of individual disease risks[111,124].

**Monoallelic versus biallelic tumour predispositions.** There is an overlap between autosomal dominant tumour predispositions and childhood-onset syndromes (sometimes associated with developmental anomalies) caused by biallelic inherited pathogenic variants in some tumour suppressor genes. Fanconi anaemia subtypes caused by biallelic *BRCA2* or *BRCA1* pathogenic variants were discussed above with regard to qualitative variants that differentially affect different transcripts (see 'Transcription differences modify the impact of some variants'). Similarly, Lynch syndrome and constitutional mismatch repair deficiency caused by heterozygous and biallelic variants in mismatch repair genes, respectively, may represent a phenotypic continuum determined by individual variant effects[125]. Some variants associated with constitutional mismatch repair deficiency in compound heterozygosity are regarded as benign concerning tumour risk in the heterozygous state, which is challenging for the classification of pathogenicity and the assignment of inheritance patterns.

**Genomic imprinting in monogenic tumour predispositions.** Some tumours show a monogenic risk only if a heterozygous pathogenic variant is inherited from a specific parent, typically the father. This is not usually due to imprinting of the respective gene itself but to the presence of growth-related imprinted genes in the chromosomal context. If the heterozygous pathogenic variant is on the same chromosomal strand as an imprinted growth-promoting (usually paternal) allele, loss of the other (usually maternal) allele at both loci causes proliferation of a cell with complete loss of the WT protein. This mechanism underlies *SDHD*-related paraganglioma–pheochromocytoma syndrome (Fig. 6c), or congenital focal hyperinsulinism caused by insulin-secreting pancreatic adenomas associated with heterozygous paternally inherited pathogenic variants in *KCNJ11* or *ABCC8*, which are adjacent genes on chromosome 11p15.1 (ref. [126]) (discussed in a different context in Fig. 3a).

## Gonosomal and mitochondrial inheritance

The inheritance mechanisms outlined above apply to autosomal genes, which normally are biallelic in all individuals. Cells generally express only single copies of most gonosomal genes, and the pathogenetic principles associated with gonosomal variants do not reflect the functional relationship of two differing alleles within a cell. Instead, clinical manifestation in females depends on the effect of an expressed variant on cell survival and function, compensation by cells with normal function and, sometimes, rare effects such as cellular interference (Fig. 7). Variant-specific protein effects – as outlined in this Review – are similar and influence disease presentation, but sex-specific effects have a major role in cellular and clinical phenotypes. It should be noted that not all X-chromosomal genes undergo stable X-inactivation: 25–30% of X-chromosomal protein-coding genes show complete or variable escape from X-inactivation and, therefore, are biallelically expressed in females, potentially explaining some of the differences between males and females[127,128]. Outside the pseudoautosomal regions PAR1 and PAR2 (which undergo X–Y recombination in males), 17 genes have functional homologues on the X and Y chromosomes. These genes code for regulatory proteins, are highly dosage-sensitive[129,130] and are candidates for explaining clinical manifestations in persons with




Turner or Klinefelter syndrome[131]. Homologue function or expression in males and females may not be exactly equal, as pathogenic variants in these genes show the usual X-chromosomal manifestation patterns with a normal or less severe clinical phenotype in females compared with males. A notable exception is the *TBL1X*/*TBL1Y* gene pair: variants in both homologous genes have been associated with sensorineural deafness, and a missense variant in *TBL1Y* was suggested as the cause of Y-chromosomal adult-onset sensorineural deafness in a large family[132]. However, this observation requires confirmation in additional families.

It has been convincingly argued that the terms dominant and recessive should be avoided for X-chromosomal traits outside the pseudoautosomal regions[133]. With the exception of the pseudoautosomal regions, X and Y-chromosomal genes are present as only a single copy in XY males. Variants in these genes are hemizygous and have full impact on the encoded transcript and protein. The terms dominant and recessive thus do not apply to inheritance patterns in males. Similarly, in females, cellular X-inactivation causes epigenetic silencing of one copy of most X-chromosomal genes, implying that single gene copies determine the cellular phenotype. For the majority of X-chromosomal genes, therefore, only one copy is transcriptionally active in both XY and XX cells, and XX individuals can be regarded as mosaics[134] who are functionally hemizygous at the cellular level.

The terms dominant and recessive are also inappropriate for pathogenic variants in mtDNA, which codes for 13 respiratory chain

**Fig. 6 | Multiple steps in tumour predisposition syndromes. a**, Two-step pathogenesis in tumour development: most monogenic tumour predispositions are caused by heterozygous pathogenic variants in tumour suppressor genes and, thus, represent dominant traits. However, the cellular pathogenetic mechanism is recessive: biallelic loss of gene function triggers tumour development[190]. Normally two independent mutation events (hits) in the same cell are required for this effect. With an inherited loss of function (LoF) variant on one allele, a single mutation event is sufficient for complete loss of tumour suppressor function. The high probability of this happening in one of the relevant cells represents the autosomal dominant risk trait. Reduced penetrance and variable expressivity in tumour predisposition syndromes are thus true stochastic phenomena linked to chance events. It should be noted, however, that specific pathogenic variants in some tumour suppressor genes such as *TP53* (ref. [191]) may have gain of function (GoF) effects with a potentially dominant cellular pathogenetic mechanism. **b**, Three-step pathogenesis in monogenic schwannomatosis: inherited heterozygous variants in any of three currently known genes on chromosome 22q11 — *SMARCB1, LZTR1* and *DGCR8* — cause an autosomal dominant predisposition to develop multiple Schwann cell tumours. Schwannomas also occur in neurofibromatosis type 2 (NF2) caused by inherited heterozygous variants in the *NF2* gene on chromosome 22q12. Somatic *NF2* mutations are frequently found in schwannomatosis-related schwannomas, which is explained with a three-step pathogenesis: (1) inheritance of a pathogenic variant (yellow star) in one of the relevant genes (*SMARCB1* in 50% of cases, illustrated here); (2) somatic loss of the other 22q allele, including wild type (WT) copies of *SMARCB1*, *NF2* and, potentially, other relevant genes; and (3) somatic mutation (yellow star) in the remaining WT *NF2* allele on the same chromosomal strand as the inherited *SMARCB1* variant[192,193]. Schwannomas in NF2 may develop through a complementary genetic mechanism. The location of several functionally related genes in the same chromosomal region explains the overlapping clinical features of NF2 and schwannomatosis, with different probabilities of associated tumours in both conditions. **c**, Sex-dependent (paternal) dominant inheritance: predisposition to phaeochromocytoma and paraganglioma (PPGL) is caused by heterozygous LoF variants in succinate dehydrogenase (*SDH*; respiratory chain complex 2) subunit genes, which comprise four paralogues (*SDHA*, *SDHB*, *SDHC* and *SDHD*). In contrast to PPGL associated with *SDHA*, *SDHB* and *SDHC* variants, PPGL linked to *SDHD* (or the assembly factor gene *SDHAF2*, not shown) on chromosome 11q is almost exclusively limited to variants inherited from the father (yellow star). Cell proliferation and tumour development involve dysregulation of the imprinted region on chromosome 11p15.5, which contains active growth-supporting genes on the paternal allele and active growth-restricting genes on the maternal allele. Tumours usually show loss of maternal chromosome 11 as a somatic event[194], causing loss of the remaining WT *SDHD* allele and increased growth stimulus most likely mediated by loss of the maternally imprinted chromosome 11p15.5 region. PPGL associated with maternally inherited *SDHD* variants is associated with more complex chromosome 11 alterations[195].

proteins as well as mitochondrial tRNAs and rRNAs, and has numerous copies located in the multiple mitochondria present in each cell. In most mtDNA-related diseases, variable clinical manifestation is due to either variant effects or heteroplasmy, that is, a variable proportion of variant mtDNA copies in different organs. Inheritance is strictly maternal; published data suggesting biparental inheritance[135] probably reflect mtDNA segments embedded within the nuclear genome[136].

## Diagnostic challenges in genetic medicine

The annotation of diseases and inheritance patterns associated with individual genes is often challenging[137]. Understanding the specific effects of genetic variants on the transcript and various phenotype levels, and the functional interaction with the WT allele and/or other allelic and non-allelic variants, is key to the correct interpretation of pathogenicity in medical genetics.

### Available resources

Functional variant information in many current databases is limited to the estimated likelihood of general pathogenicity. The variant interpretation guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG–AMP) use 28 criteria based on population, computational, functional, de novo and other data to define variants as (likely) pathogenic or benign, or of uncertain significance[138]. They are the general standard for diagnostic laboratories but do not prevent conflicting classifications for many variants by different laboratories[139]. Detailed refinements as well as specific guidelines for certain (groups of) genes have been developed[140]. Additional work focuses on validation of gene–disease associations and dosage sensitivity[35]. A complementary ABC grading system including functional criteria and clinical aspects aims to address some limitations of the ACMG–AMP classification particularly with regard to low-penetrant and recessive disease-associated variants[141]. Nevertheless, current classification systems do not fully address the complex functional and clinical effects that determine different inheritance patterns for different variants in some genes.

**The nomenclature of genetic variants can make functional assumptions.** In addition to specifying the altered nucleotide(s) in relation to the coding sequence of a gene (denoted with c.), or the nuclear or mitochondrial genomic sequence (denoted g. or m.), variant names may also describe the predicted functional effect on transcript level (denoted r.) or the encoded protein (denoted p.). It is important to recognize that in the absence of additional data such as transcript sequencing results, these functional interpretations are sometimes incorrect. Some variants may have more than one potential functional consequence on transcript splicing and the encoded amino acid sequence[142,143], which need to be carefully disentangled. Also, there may be unexpected effects of coding variants on mRNA structure and processing[144], and predicted silent (synonymous) variants that are not expected to alter the amino acid sequence may have functional consequences, for example, through disruption of splicing enhancers[145].

**Enhancing variant and disease databases.** Integrating the functional concepts summarized in this Review may provide further assistance for variant interpretation in genetic diagnostics. Variant databases could be complemented with information on specific quantitative or qualitative functional effects, and disease databases would benefit from providing the respective mechanisms of pathogenesis. In many instances, additional computational and experimental methods are helpful for confirming or excluding a presumed pathogenic effect[146]. Careful phenotyping of a patient is often essential for the correct interpretation of genetic–genomic data, for example when a single heterozygous variant is identified in a gene associated with an autosomal recessive disease.

### Prediction of monogenic variant effects

**Is a variant suggestive of a quantitative effect?** Quantitative effects may be expected for gene CNVs (such as whole gene deletions or chromosomal CNVs), major structural gene alterations, regulatory or promoter variants that prohibit transcription, most variants that introduce early translation frameshifts or termination codons, and

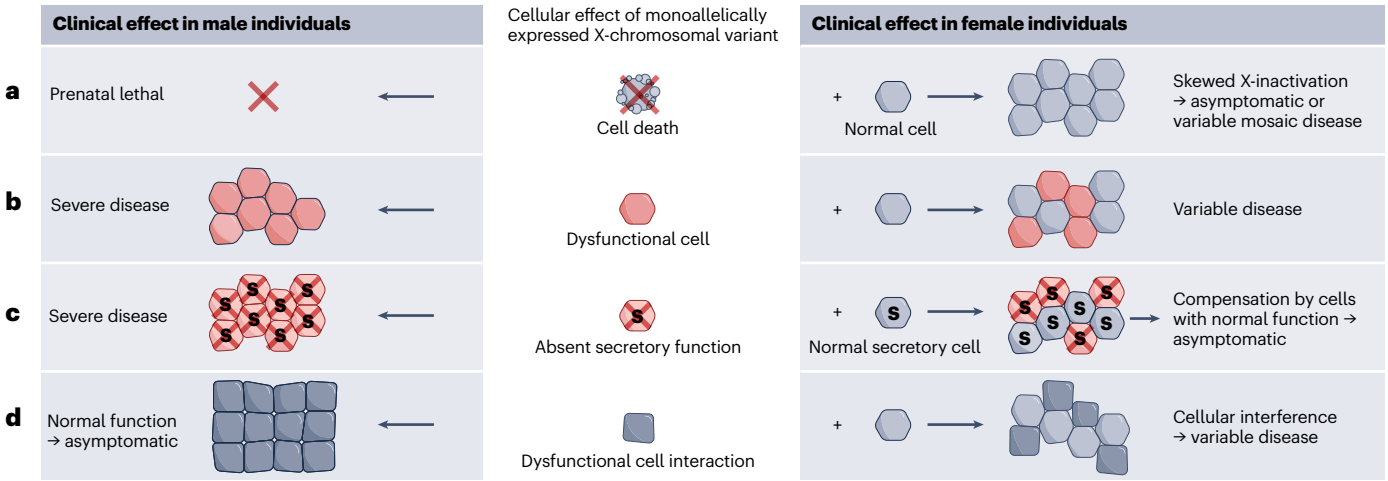



**Fig. 7 | Gonosomal inheritance.** The manifestation of pathogenic X-chromosomal variants depends on cell survival and intercellular compensatory mechanisms, with four main pathogenic principles. **a**, Variants that in the (functionally) hemizygous state are incompatible with cellular survival are lethal in males and cause skewed X-inactivation in females; this is sometimes recognizable by mosaic skin or organ manifestation patterns, for example in incontinentia pigmenti or oral–facial–digital syndrome type 1. Surviving affected males have either post-zygotic mosaic pathogenic variants or Klinefelter (XXY) syndrome. **b**, Variants that disturb essential cellular functions without causing cell death show a less severe manifestation in females than in males, with clinical variability in females often due to variable X-inactivation. It is important to note that heterozygous mothers of affected boys may show ascertainment bias with regard to severity; that is, by good fortune they have a milder disease phenotype, otherwise they would not have children, and their heterozygous daughters may be more likely to be more severely affected. An example is ornithine transcarbamylase (OTC) deficiency, an X-linked urea cycle disorder. **c**, Variants in genes for secreted proteins or other cellular functions that upon wild type (WT) inactivation in some cells are compensated by normal function of adjacent cells are usually asymptomatic in females. Examples include haemophilia and many X-linked metabolic diseases; manifesting females usually show unfavourable skewed X-inactivation, have Turner syndrome or happen to be homozygous. **d**, Very rarely, female heterozygosity or male mosaicism causes cellular interference leading to a more severe phenotype than in hemizygous males; known examples are *EFNB1*-associated craniofrontonasal syndrome[196] and *PCDH19*-associated epilepsy[197]. None of these principles relates to the functional relationship of two differing alleles within a cell, and the manifestation mechanisms in heterozygous females are very different from heterozygous autosomal variants. Also considering the variable manifestation in heterozygous females, the terms dominant and recessive should be avoided for X-chromosomal traits outside the pseudoautosomal regions[133].

many missense variants that result in protein degradation or remove protein function without other structural effects. Computational stability predictors may assist in identifying protein-degrading variants particularly in haploinsufficient genes[147], but their ability to determine pathogenicity is hampered by the possibility of non-quantitative protein effects[148]. Most variants that introduce a premature stop codon cause LoF through nonsense-mediated decay of the transcript; however, it is important to recognize possible nonsense-mediated decay escape with termination codons, for example, in the last exon or the 3′ region of the second last exon, or when a translated frameshift transcript extends towards the last exon–exon junction[149]. Thus, C-terminal nonsense variants may have more severe, dominant altered-function effects than amino-terminal variants that may be recessive in a haplosufficient gene. Alternative splicing may counteract variants in specific exons of some genes, as shown for example for *BRCA1* and *BRCA2* (refs. [81,82]). Splice site variants that cause incomplete disruption of splicing may represent Hyp variants that lead to reduced production of functionally normal protein. Variants located in non-coding regions of a gene (the promoter, 5′ and 3′ untranslated regions, and introns) and the start codon may have adverse effects on transcription and often — but not always — cause complete loss of gene function. An interesting phenomenon is the translation-suppressing effect caused by variants that introduce a functional out-of-frame translation initiation codon 5′ of the authentic start codon, or which extend short upstream open reading frames[150,151]. Reduced or absent gene function may also be linked to loss of *cis*-regulatory (for example, enhancer) elements[152]. Although it might be assumed that CNVs exert predominantly quantitative effects, this is not always the case particularly because some CNVs disturb the underlying architecture of topologically associating domains. Both deletions and duplications that span topologically associating domain boundaries can bring genes into novel regulatory landscapes, resulting in GoF through ectopic expression[153].

**Estimating gene dosage sensitivity.** Overall, the metrics discussed above (pLI and LOEUF scores, see ‘Loss of function variants in most genes are recessive’)[8,14], together with new empiric measures emerging for triplosensitivity[39], are reasonably predictive of the consequences of heterozygous CNVs or clearly disruptive intragenic variants. Nevertheless, it is important to be aware of potential pitfalls[154,155]. On the other hand, identifying Hyp effects of missense and other variants, and predicting their consequences for phenotype, remains heavily reliant on empiric observations from segregation of phenotype and/or bespoke experimentation. In well-understood situations, massively scalable assays are beginning to provide valuable supportive data[156,157].

**Is there a potential qualitative effect?** Predicting adverse functional effects of variants that may result in a stable, altered protein is difficult. Bioinformatic approaches have been developed to estimate the likely pathogenicity of missense variants in particular gene regions through the comparison of observed and expected variant frequencies[158–161].

Refinements include hierarchical analyses using Bayesian regression[162], aggregation of homologous protein domains[163] and inclusion of structural protein data[164]. Some algorithms have been integrated into software solutions for the prediction of variant pathogenicity[165]. Only a few studies differentiate missense LoF and GoF effects[166,167], and they do not usually consider functionally different or overlapping quantitative and qualitative effects, which would be essential for adequate diagnosis and assessing inheritance patterns. Recent advances in computational strategies have greatly improved the accuracy of protein structure predictions[168]. Integrating this knowledge can aid interpretation, as LOF and other variants differ in terms of their location within structures, their predicted effects on protein stability and their clustering in three-dimensional space[169].

Variants that cause substantial qualitative protein effects are mostly associated with dominant manifestation, and segregation analyses in families may provide decisive additional information. Unfortunately, current disease databases rarely provide the different pathogenic mechanisms such as LoF, Hyp, GoF, dominant negative effects, subfunction loss, or others, when a gene is associated with different – sometimes opposing – phenotypes. Similarly, variant databases do not usually provide systematic information on specific functional effects. A thorough knowledge of the disease-related literature, understanding of protein structure and function as well as associated multimers, and integration of clinical information are essential for the correct interpretation of rare and de novo pathogenic variants.

### Variant-unrelated factors

For the majority of protein-coding genes, the Matched Annotation from NCBI and EMBL-EBI (MANE) project – a collaboration between the National Center for Biotechnology Information (NCBI) and the European Molecular Biology Laboratories European Bioinformatics Institute (EMBL-EBI) – has identified single 'MANE Select' transcripts that represent the biology of the respective gene. These will be supplemented with additional 'MANE Plus Clinical' transcripts when necessary for clinical variant reporting. The importance of considering alternative transcripts for variant interpretation was illustrated for *SCN8A* gene analyses: pathogenic variants in the alternatively spliced exon 5A were initially missed in diagnostic genetic testing because the respective transcript had not been included in the Consensus Coding Sequence database[170].

Many apparently monogenic conditions have strong genetic or epigenetic elements or modifiers, or become manifest only after additional somatic genetic events, explaining variable or low penetrance. Designating inheritance patterns associated with specific phenotypes or novel and rare variants in these constellations remains challenging.

## Conclusions

A major goal for future decades will be to describe the complete range of functional effects and phenotypes associated with variation at every position in the human genome. Daunting as this challenge is, the useful application of this information to biallelic genes in a medical genetics context will obligatorily require additional consideration of the allele that the variant is partnered with – whether this is the normal WT allele, the identical variant or a different variant in the same gene. A robust framework for the codification and classification of variant properties – such as presented in this Review – should assist in understanding and describing the dominant or recessive behaviour of variant alleles in different contexts.

Published online: 20 February 2023

## References


1. Mendel, G. Versuche über Pflanzenhybriden [German]. *Verh. Naturforschenden Vereines Brünn.* **4**, 3–47 (1866). English translation in Bateson, W. Mendel's Principles of Heredity: A Defence. (Cambridge University Press, 1902).
2. Zschocke, J., Byers, P. H. & Wilkie, A. O. M. Gregor Mendel and the concepts of dominance and recessiveness. *Nat. Rev. Genet.* **23**, 387–388 (2022).
3. Bateson, W. & Saunders, E. R. The facts of heredity in the light of Mendel's discovery. *Rep. Evol. Comm. R. Soc.* **1**, 125–160 (1902).
4. Garrod, A. E. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* **2**, 1616–1620 (1902).
5. McKusick, V. A. *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, 1966).
6. Nguengang Wakap, S. et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur. J. Hum. Genet.* **28**, 165–173 (2020).
7. Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).
   **This article summarizes exome sequence data from almost 50,000 individuals in the UK biobank, including information on the prevalence of LoF variants and pathogenic variants of clinical importance per participant in the general population.**
8. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
   **This study aggregates human sequence data from the Genome Aggregation Database (gnomAD), and describes the LOF observed/expected upper bound fraction score for estimating the likelihood of haplosufficiency versus haploinsufficiency for most protein-coding genes.**
9. Collins, R. L. et al. A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020).
   **This article describes a sequence-resolved analysis of all types of structural variants in almost 15,000 genomes across diverse global populations in gnomAD, and provides estimates of dosage sensitivity in the non-coding genome.**
10. Muller, H. J. in *Proc. Sixth Int. Congress of Genetics* (ed. Jones, D. F.) 213–255 (Brooklyn Botanic Gardens, 1932).
11. Morgan, T. H., Bridges, C. B. & Sturtevant, A. H. The genetics of *Drosophila*. *Bibliographica Genet.* **II**, 1–262 (1925).
12. Orr, H. A. A test of Fisher's theory of dominance. *Proc. Natl Acad. Sci. USA* **88**, 11413–11415 (1991).
13. Deutschbauer, A. M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–1925 (2005).
14. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
15. Minikel, E. V. et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459–464 (2020).
16. Balick, D. J., Jordan, D. M., Sunyaev, S. & Do, R. Overcoming constraints on the detection of recessive selection in human genes from population frequency data. *Am. J. Hum. Genet.* **109**, 33–49 (2022).
17. Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
18. Cummings, B. B. et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020).
19. Cody, J. D. The consequences of abnormal gene dosage: lessons from chromosome 18. *Trends Genet.* **36**, 764–776 (2020).
20. Barton, A. R., Hujoel, M. L. A., Mukamel, R. E., Sherman, M. A. & Loh, P. R. A spectrum of recessiveness among Mendelian disease variants in UK Biobank. *Am. J. Hum. Genet.* **109**, 1298–1307 (2022).
21. Kingdom, R. et al. Rare genetic variants in genes and loci linked to dominant monogenic developmental disorders cause milder related phenotypes in the general population. *Am. J. Hum. Genet.* **109**, 1308–1316 (2022).
22. Fisher, R. A. The possible modification of the response of the wild type to recurrent mutations. *Am. Nat.* **62**, 679 (1928).
23. Wright, S. Physiological and evolutionary theories of dominance. *Am. Nat.* **68**, 24–53 (1934).
24. Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97**, 639–666 (1981).
25. Omholt, S. W., Plahte, E., Oyehaug, L. & Xiang, K. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* **155**, 969–980 (2000).
26. Ishikawa, K., Makanae, K., Iwasaki, S., Ingolia, N. T. & Moriya, H. Post-translational dosage compensation buffers genetic perturbations to stoichiometry of protein complexes. *PLoS Genet.* **13**, e1006554 (2017).
27. Khan, Z. et al. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).
28. Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* **20**, 287–290 (2004).
29. Morrill, S. A. & Amon, A. Why haploinsufficiency persists. *Proc. Natl Acad. Sci. USA* **116**, 11866–11871 (2019).
    **This paper analyses the functional characteristics of haploinsufficient genes, and proposes a dosage-stabilizing hypothesis of haploinsufficiency to explain its persistence over evolutionary time.**
30. Ni, Z., Zhou, X. Y., Aslam, S. & Niu, D. K. Characterization of human dosage-sensitive transcription factor genes. *Front. Genet.* **10**, 1208 (2019).
31. Gilchrist, M. A. & Nijhout, H. F. Nonlinear developmental processes as sources of dominance. *Genetics* **159**, 423–432 (2001).

32. Veitia, R. A., Caburet, S. & Birchler, J. A. Mechanisms of Mendelian dominance. *Clin. Genet.* **93**, 419–428 (2018).
33. Johnson, A. F., Nguyen, H. T. & Veitia, R. A. Causes and effects of haploinsufficiency. *Biol. Rev. Camb. Philos. Soc.* **94**, 1774–1785 (2019).
34. Veitia, R. A. A generalized model of gene dosage and dominant negative effects in macromolecular complexes. *FASEB J.* **24**, 994–1002 (2010).
35. Thaxton, C. et al. Utilizing ClinGen gene–disease validity and dosage sensitivity curations to inform variant classification. *Hum. Mutat.* **43,** 1031-1040 (2021).
36. Sopko, R. et al. Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
37. Antonarakis, S. E. Down syndrome and the complexity of genome dosage imbalance. *Nat. Rev. Genet.* **18**, 147–163 (2017).
38. Abel, H. J. et al. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**, 83–89 (2020).
39. Collins, R. L. et al. A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041–3055.e25 (2022).
40. van Paassen, B. W. et al. PMP22 related neuropathies: Charcot–Marie–Tooth disease type 1A and hereditary neuropathy with liability to pressure palsies. *Orphanet J. Rare Dis.* **9**, 38 (2014).
41. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208 (2009).
42. Stoebel, D. M., Dean, A. M. & Dykhuizen, D. E. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics* **178**, 1653–1660 (2008).
43. Makanae, K., Kintaka, R., Makino, T., Kitano, H. & Moriya, H. Identification of dosage-sensitive genes in *Saccharomyces cerevisiae* using the genetic tug-of-war method. *Genome Res.* **23**, 300–311 (2013).
44. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
45. Rice, A. M. & McLysaght, A. Dosage-sensitive genes in evolution and disease. *BMC Biol.* **15**, 78 (2017).
46. Rudnik-Schöneborn, S. et al. Andrological findings in infertile men with two (biallelic) CFTR mutations: results of a multicentre study in Germany and Austria comprising 71 patients. *Hum. Reprod.* **36**, 551–559 (2021).
47. Guldberg, P. et al. A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. *Am. J. Hum. Genet.* **63**, 71–79 (1998).
48. Ramsey, B. W. et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N. Engl. J. Med.* **365**, 1663–1672 (2011).
49. Carlock, G. et al. Developmental outcomes in duarte galactosemia. *Pediatrics* **143,** e20182516 (2019).
50. Gouya, L. et al. The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. *Nat. Genet.* **30**, 27–28 (2002).
51. Rose, A. M. et al. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Sci. Rep.* **6**, 19450 (2016).
52. Falkenberg, K. D. et al. Allelic expression imbalance promoting a mutant PEX6 allele causes Zellweger spectrum disorder. *Am. J. Hum. Genet.* **101**, 965–976 (2017).
53. Boussion, S. et al. TAR syndrome: clinical and molecular characterization of a cohort of 26 patients and description of novel noncoding variants of RBM8A. *Hum. Mutat.* **41**, 1220–1225 (2020).
54. Daum, H. et al. Smith–Lemli–Opitz syndrome: what is the actual risk for couples carriers of the DHCR7:c.964-1G>C variant? *Eur. J. Hum. Genet.* **28**, 938–942 (2020).
55. Plomin, R., Haworth, C. M. & Davis, O. S. Common disorders are quantitative traits. *Nat. Rev. Genet.* **10**, 872–878 (2009).
56. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
**This study describes an analysis and meta-analysis of de novo pathogenic variants in more than 7,500 individuals with developmental disorders, estimation of their prevalence in the general population and the relative frequency of variants with quantitative and qualitative pathogenic effects, and identification of factors that influence the diagnostic yield**.
57. Gröbner, R. et al. C1R mutations trigger constitutive complement 1 activation in periodontal Ehlers–Danlos syndrome. *Front. Immunol.* **10**, 2537 (2019).
58. Ilsley, M. D. et al. Corrupted DNA-binding specificity and ectopic transcription underpin dominant neomorphic mutations in KLF/SP transcription factors. *BMC Genomics* **20**, 417 (2019).
59. Otonkoski, T. et al. Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic β cells. *Am. J. Hum. Genet.* **81**, 467–474 (2007).
60. Koyano-Nakagawa, N. et al. Etv2 regulates enhancer chromatin status to initiate Shh expression in the limb bud. *Nat. Commun.* **13**, 4221 (2022).
61. Oldridge, M. et al. De novo alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *Am. J. Hum. Genet.* **64**, 446–461 (1999).
62. Yu, K. & Ornitz, D. M. Uncoupling fibroblast growth factor receptor 2 ligand binding specificity leads to Apert syndrome-like phenotypes. *Proc. Natl Acad. Sci. USA* **98**, 3641–3643 (2001).
63. Tsuji, Y. et al. Systematic review of genotype–phenotype correlations in Frasier syndrome. *Kidney Int. Rep.* **6**, 2585–2593 (2021).
64. Swinnen, B., Robberecht, W. & Van Den Bosch, L. RNA toxicity in non-coding repeat expansion disorders. *EMBO J.* **39**, e101112 (2020).
65. Malik, I., Kelley, C. P., Wang, E. T. & Todd, P. K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).
66. Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
67. Zlotogora, J. Dominance and homozygosity. *Am. J. Med. Genet.* **68**, 412–416 (1997).
68. Cubo, E. et al. Clinical manifestations of homozygote allele carriers in Huntington disease. *Neurology* **92**, e2101–e2108 (2019).
69. Tabrizi, S. J., Flower, M. D., Ross, C. A. & Wild, E. J. Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. *Nat. Rev. Neurol.* **16**, 529–546 (2020).
70. Jamilloux, Y. et al. Familial Mediterranean fever mutations are hypermorphic mutations that specifically decrease the activation threshold of the Pyrin inflammasome. *Rheumatology* **57**, 100–111 (2018).
71. Jeyakanthan, M. et al. Chemical basis for qualitative and quantitative differences between ABO blood groups and subgroups: implications for organ transplantation. *Am. J. Transplant.* **15**, 2602–2615 (2015).
72. Johnson, W. G. Metabolic interference and the +– heterozygote. a hypothetical form of simple inheritance which is neither dominant nor recessive. *Am. J. Hum. Genet.* **32**, 374–386 (1980).
73. Morissette, J. et al. Homozygotes carrying an autosomal dominant TIGR mutation do not manifest glaucoma. *Nat. Genet.* **19**, 319–321 (1998).
74. Huard, D. J. E. et al. Trifunctional high-throughput screen identifies promising scaffold to inhibit Grp94 and treat myocilin-associated glaucoma. *ACS Chem. Biol.* **13**, 933–941 (2018).
75. Brownlee, J. M., Heinz, B., Bates, J. & Moran, G. R. Product analysis and inhibition studies of a causative Asn to Ser variant of 4-hydroxyphenylpyruvate dioxygenase suggest a simple route to the treatment of Hawkinsinuria. *Biochemistry* **49**, 7218–7226 (2010).
76. Xu, D. et al. The evolving landscape of noncanonical functions of metabolic enzymes in cancer and other pathologies. *Cell Metab.* **33**, 33–50 (2021).
77. Pan, C., Li, B. & Simon, M. C. Moonlighting functions of metabolic enzymes and metabolites in cancer. *Mol. Cell* **81**, 3760–3774 (2021).
78. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
79. Kim, H. K., Pham, M. H. C., Ko, K. S., Rhee, B. D. & Han, J. Alternative splicing isoforms in health and disease. *Pflug. Arch. Eur. J. Physiol.* **470**, 995–1016 (2018).
80. Gostyńska, K. B. et al. Mutation in exon 1a of PLEC, leading to disruption of plectin isoform 1a, causes autosomal-recessive skin-only epidermolysis bullosa simplex. *Hum. Mol. Genet.* **24**, 3155–3162 (2015).
81. Seo, A. et al. Mechanism for survival of homozygous nonsense mutations in the tumor suppressor gene BRCA1. *Proc. Natl Acad. Sci. USA* **115**, 5241–5246 (2018).
82. Mesman, R. L. S. et al. Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. *Genet. Med.* **22**, 1355–1365 (2020).
83. Li, J. et al. Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. *Am. J. Hum. Genet.* **98**, 830–842 (2016).
84. Perenthaler, E. et al. Loss of UGP2 in brain leads to a severe epileptic encephalopathy, emphasizing that bi-allelic isoform-specific start-loss mutations of essential genes can cause genetic diseases. *Acta Neuropathol.* **139**, 415–442 (2020).
85. Dik, E., Naamati, A., Asraf, H., Lehming, N. & Pines, O. Human fumarate hydratase is dual localized by an alternative transcription initiation mechanism. *Traffic* **17**, 720–732 (2016).
86. Sternisha, S. M. & Miller, B. G. Molecular and cellular regulation of human glucokinase. *Arch. Biochem. Biophys.* **663**, 199–213 (2019).
87. Ferdinandusse, S. et al. An autosomal dominant neurological disorder caused by de novo variants in FAR1 resulting in uncontrolled synthesis of ether lipids. *Genet. Med.* **23**, 740–750 (2021).
88. Salvatore, D., Santoro, M. & Schlumberger, M. The importance of the RET gene in thyroid cancer and therapeutic implications. *Nat. Rev. Endocrinol.* **17**, 296–306 (2021).
89. Arighi, E. et al. Biological effects of the dual phenotypic Janus mutation of ret cosegregating with both multiple endocrine neoplasia type 2 and Hirschsprung's disease. *Mol. Endocrinol.* **18**, 1004–1017 (2004).
90. Nicole, S. & Lory, P. New challenges resulting from the loss of function of Na(v)1.4 in neuromuscular diseases. *Front. Pharmacol.* **12**, 751095 (2021).
91. Liu, M., Yang, K. C. & Dudley, S. C. Jr Cardiac sodium channel mutations: why so many phenotypes? *Nat. Rev. Cardiol.* **11**, 607–615 (2014).
**This paper provides an exemplary review of the diverse functional effects of different pathogenic variants in the major cardiac sodium channel gene, based on the direct effect of the mutation on channel biophysics as well as age, sex, body temperature, cardiac regions and additional modifiers of channel behaviour**.
92. Wiuf, C. Do ΔF508 heterozygotes have a selective advantage? *Genet. Res.* **78**, 41–47 (2001).
93. Allison, A. C. Polymorphism and natural selection in human populations. *Cold Spring Harb. Symp. Quant. Biol.* **29**, 137–149 (1964).
94. Krawczak, M. & Zschocke, J. A role for overdominant selection in phenylketonuria? Evidence from molecular data. *Hum. Mutat.* **21**, 394–397 (2003).
95. Miller, A. C. et al. Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. *Proc. Natl Acad. Sci. USA* **117**, 1621–1627 (2020).

96. Penrose, L. S. The problem of anticipation in pedigrees of dystrophia myotonica. *Ann. Eugen.* **14**, 125–132 (1948).
97. Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* **20**, 299–309 (2019).
98. Vulliamy, T. et al. Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nat. Genet.* **36**, 447–449 (2004).
99. Mangaonkar, A. A. & Patnaik, M. M. Short telomere syndromes in clinical practice: bridging bench and bedside. *Mayo Clin. Proc.* **93**, 904–916 (2018).
100. Monk, D., Mackay, D. J. G., Eggermann, T., Maher, E. R. & Riccio, A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **20**, 235–248 (2019).
101. Buiting, K., Williams, C. & Horsthemke, B. Angelman syndrome — insights into a rare neurogenetic disorder. *Nat. Rev. Neurol.* **12**, 584–593 (2016).
102. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
103. Martinelli, S. et al. Functional dysregulation of CDC42 causes diverse developmental phenotypes. *Am. J. Hum. Genet.* **102**, 309–320 (2018).
104. Cohen, A. S., Wilson, S. L., Trinh, J. & Ye, X. C. Detecting somatic mosaicism: considerations and clinical implications. *Clin. Genet.* **87**, 554–562 (2015).
105. Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* **14**, 307–320 (2013).
106. Smedley, D. et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* **99**, 595–606 (2016).
107. Turner, H. & Jackson, L. Evidence for penetrance in patients without a family history of disease: a systematic review. *Eur. J. Hum. Genet.* **28**, 539–550 (2020).
108. Niemi, M. E. K. et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268–271 (2018).
109. Goodrich, J. K. et al. Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. *Nat. Commun.* **12**, 3505 (2021).
110. Fallerini, C. et al. Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity. *Hum. Genet.* **141**, 147–173 (2022).
111. Mars, N. et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat. Commun.* **11**, 6383 (2020).
112. Schäffer, A. A. Digenic inheritance in medical genetics. *J. Med. Genet.* **50**, 641–652 (2013).
113. Deltas, C. Digenic inheritance and genetic modifiers. *Clin. Genet.* **93**, 429–438 (2018).
114. Jirsa, M. et al. Rotor syndrome in *GeneReviews* [Internet] (Univ. of Washington, 2019).
115. Bateson, W. Facts limiting the theory of heredity. *Science* **26**, 649–660 (1907).
116. Miko, I. Epistasis: gene interaction and phenotype effects. *Nat. Educ.* **1**, 197 (2008).
117. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
118. Kelly, R. J. et al. Molecular basis for H blood group deficiency in Bombay (Oh) and para-Bombay individuals. *Proc. Natl Acad. Sci. USA* **91**, 5843–5847 (1994).
119. Estiar, M. A. et al. Evidence for non-Mendelian inheritance in spastic paraplegia 7. *Mov. Disord.* **36**, 1664–1675 (2021).
120. Vockley, J. et al. Complex patterns of inheritance, including synergistic heterozygosity, in inborn errors of metabolism: implications for precision medicine driven diagnosis and treatment. *Mol. Genet. Metab.* **128**, 1–9 (2019).
121. Wei, W. H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* **15**, 722–733 (2014).
122. Hivert, V. et al. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *Am. J. Hum. Genet.* **108**, 786–798 (2021).
123. Knudson, A. G. Jr Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68**, 820–823 (1971).
124. Turnbull, C., Sud, A. & Houlston, R. S. Cancer genetics, precision prevention and a call to action. *Nat. Genet.* **50**, 1212–1218 (2018).
125. Fernandez-Rozadilla, C. et al. Early colorectal cancers provide new evidence for a Lynch syndrome-to-CMMRD phenotypic continuum. *Cancers* **11,** 1081 (2019).
126. Verkarre, V. et al. Paternal mutation of the sulfonylurea receptor (*SUR1*) gene and maternal loss of 11p15 imprinted genes lead to persistent hyperinsulinism in focal adenomatous hyperplasia. *J. Clin. Invest.* **102**, 1286–1291 (1998).
127. Tukiainen, T. et al. Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
**This paper describes a comprehensive study of the pattern of X-chromosome inactivation in human female cells, including data on genes with variable 'escape' from inactivation and a discussion of the potential clinical relevance**.
128. Balaton, B. P., Dixon-McDougall, T., Peeters, S. B. & Brown, C. J. The exceptional nature of the X chromosome. *Hum. Mol. Genet.* **27**, R242–R249 (2018).
129. Naqvi, S., Bellott, D. W., Lin, K. S. & Page, D. C. Conserved microRNA targeting reveals preexisting gene dosage sensitivities that shaped amniote sex chromosome evolution. *Genome Res.* **28**, 474–483 (2018).
130. Bellott, D. W. et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
131. Zhang, X. et al. Integrated functional genomic analyses of Klinefelter and Turner syndromes reveal global network effects of altered X chromosome dosage. *Proc. Natl Acad. Sci. USA* **117**, 4864–4873 (2020).
132. Di Stazio, M. et al. TBL1Y: a new gene involved in syndromic hearing loss. *Eur. J. Hum. Genet.* **27**, 466–474 (2019).
133. Dobyns, W. B. et al. Inheritance of most X-linked traits is not dominant or recessive, just X-linked. *Am. J. Med. Genet. A* **129A**, 136–143 (2004).
134. Migeon, B. R. X-linked diseases: susceptible females. *Genet. Med.* **22**, 1156–1174 (2020).
135. Luo, S. et al. Biparental inheritance of mitochondrial DNA in humans. *Proc. Natl Acad. Sci. USA* **115**, 13039–13044 (2018).
136. Pagnamenta, A. T., Wei, W., Rahman, S. & Chinnery, P. F. Biparental inheritance of mitochondrial DNA revisited. *Nat. Rev. Genet.* **22**, 477–478 (2021).
137. Strande, N. T. et al. Evaluating the clinical validity of gene–disease associations: an evidence-based framework developed by the clinical genome resource. *Am. J. Hum. Genet.* **100**, 895–906 (2017).
138. Richards, S. et al. 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**, 405–424 (2015).
**This paper reports the standard ACMG–AMP guidelines for classifying variants in molecular genetic diagnostics as (likely) pathogenic, (likely) benign or of uncertain significance, based on various criteria such as population data, computational data, functional data and segregation data**.
139. Amendola, L. M. et al. Performance of ACMG–AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research consortium. *Am. J. Hum. Genet.* **98**, 1067–1076 (2016).
140. Harrison, S. M., Biesecker, L. G. & Rehm, H. L. Overview of specifications to the ACMG/AMP variant interpretation guidelines. *Curr. Protoc. Hum. Genet.* **103**, e93 (2019).
141. Houge, G. et al. Stepwise ABC system for classification of any type of genetic variant. *Eur. J. Hum. Genet.* **30**, 150–159 (2022).
**This paper is an extension of the ACMG–AMG variant classification system, encompassing functional effect and clinical importance grades that are combined into a joint variant class to provide better estimates of variant significance in a clinical setting**.
142. Zhang, S. et al. Base-specific mutational intolerance near splice sites clarifies the role of nonessential splice nucleotides. *Genome Res.* **28**, 968–974 (2018).
143. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
144. Manning, K. S. & Cooper, T. A. The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.* **18**, 102–114 (2017).
145. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683–691 (2011).
**This paper comprehensively reviews the extent to which synonymous variants influence disease, the various molecular mechanisms that underlie these effects and the implications for future research and biomedical applications**.
146. Spielmann, M. & Kircher, M. Computational and experimental methods for classifying variants of unknown clinical significance. *Cold Spring Harb. Mol. Case Stud.* **8,** a006196 (2022).
147. Birolo, G. et al. Protein stability perturbation contributes to the loss of function in haploinsufficient genes. *Front. Mol. Biosci.* **8**, 620793 (2021).
148. Gerasimavicius, L., Liu, X. & Marsh, J. A. Identification of pathogenic missense mutations using protein stability predictors. *Sci. Rep.* **10**, 15387 (2020).
149. Coban-Akdemir, Z. et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am. J. Hum. Genet.* **103**, 171–187 (2018).
**This article describes an exemplary study of genes with premature termination codons that are predicted to escape nonsense-mediated decay and may be disease-causing through GoF effects**.
150. Ferdinandusse, S. et al. A mutation creating an upstream translation initiation codon in SLC22A5 5′UTR is a frequent cause of primary carnitine deficiency. *Hum. Mutat.* **40**, 1899–1904 (2019).
151. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106**, 7507–7512 (2009).
152. Bhatia, S. et al. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* **93**, 1126–1134 (2013).
153. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
**This article provides a comprehensive review of structural and quantitative chromosomal rearrangements that may affect the expression of distant genes through copy number alteration of regulatory elements or modification of the 3D genome by disrupting higher-order chromatin organization such as topologically associating domains**.
154. Fuller, Z. L., Berg, J. J., Mostafavi, H., Sella, G. & Przeworski, M. Measuring intolerance to mutation in human genetics. *Nat. Genet.* **51**, 772–776 (2019).
155. Ziegler, A., Colin, E., Goudenège, D. & Bonneau, D. A snapshot of some pLI score pitfalls. *Hum. Mutat.* **40**, 839–841 (2019).
156. Findlay, G. M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
157. Przybyla, L. & Gilbert, L. A. A new era in functional genomics screens. *Nat. Rev. Genet.* **23**, 89–103 (2022).
158. Gussow, A. B., Petrovski, S., Wang, Q., Allen, A. S. & Goldstein, D. B. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. *Genome Biol.* **17**, 9 (2016).

159. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
160. Havrilla, J. M., Pedersen, B. S., Layer, R. M. & Quinlan, A. R. A map of constrained coding regions in the human genome. *Nat. Genet.* **51**, 88–95 (2019).
161. Li, G. C., Forster-Benson, E. T. C. & Sanders, C. R. Genetic intolerance analysis as a tool for protein science. *Biochim. Biophys. Acta Biomembr.* **1862**, 183058 (2020).
162. Hayeck, T. J. et al. Improved pathogenic variant localization via a hierarchical model of sub-regional intolerance. *Am. J. Hum. Genet.* **104**, 299–309 (2019).
163. Wiel, L. et al. MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum. Mutat.* **40**, 1030–1038 (2019).
164. Silk, M. et al. MTR3D: identifying regions within protein tertiary structures under purifying selection. *Nucleic Acids Res.* **49**, W438–W445 (2021).
165. Qi, H. et al. MVP predicts the pathogenicity of missense variants by deep learning. *Nat. Commun.* **12**, 510 (2021).
166. Heyne, H. O. et al. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci. Transl. Med.* **12,** eaay6848 (2020).
167. Bayrak, C. S. et al. Identification of discriminative gene-level and protein-level features associated with pathogenic gain-of-function and loss-of-function variants. *Am. J. Hum. Genet.* **108,** 2301–2318 (2021).
168. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
169. Gerasimavicius, L., Livesey, B. J. & Marsh, J. A. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat. Commun.* **13**, 3895 (2022).
**This study investigates protein-level effects of pathogenic missense variants associated with different molecular mechanisms, showing that qualitative variant effects may be missed by current variant prioritization strategies, and highlighting ways to improve computational predictions through consideration of molecular disease mechanisms**.
170. Epilepsy Genetics Initiative. De novo variants in the alternative exon 5 of SCN8A cause epileptic encephalopathy. *Genet. Med.* **20**, 275–281 (2018).
171. White, K. E. et al. Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *Am. J. Hum. Genet.* **76**, 361–367 (2005).
172. Bennett, J. T. et al. Mosaic activating mutations in FGFR1 cause encephalocraniocutaneous lipomatosis. *Am. J. Hum. Genet.* **98**, 579–587 (2016).
173. Ibrahimi, O. A., Zhang, F., Eliseenkova, A. V., Linhardt, R. J. & Mohammadi, M. Proline to arginine mutations in FGF receptors 1 and 3 result in Pfeiffer and Muenke craniosynostosis syndromes through enhancement of FGF binding affinity. *Hum. Mol. Genet.* **13**, 69–78 (2004).
174. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
175. Erlandsen, H. & Stevens, R. C. The structural basis of phenylketonuria. *Mol. Genet. Metab.* **68**, 103–125 (1999).
176. Higgs, D. R. et al. A review of the molecular genetics of the human α-globin gene cluster. *Blood* **73**, 1081–1104 (1989).
177. Magge, S. N. et al. Familial leucine-sensitive hypoglycemia of infancy due to a dominant mutation of the β-cell sulfonylurea receptor. *J. Clin. Endocrinol. Metab.* **89**, 4450–4456 (2004).
178. Marini, J. C. et al. Osteogenesis imperfecta. *Nat. Rev. Dis. Prim.* **3**, 17052 (2017).
179. Lee, J. & Hegele, R. A. Abetalipoproteinemia and homozygous hypobetalipoproteinemia: a framework for diagnosis and management. *J. Inherit. Metab. Dis.* **37**, 333–339 (2014).
180. Ramasamy, I. Update on the molecular biology of dyslipidemias. *Clin. Chim. Acta* **454**, 143–185 (2016).
181. Wohlfarter, Y. et al. Lost in promiscuity? An evolutionary and biochemical evaluation of HSD10 function in cardiolipin metabolism. *Cell. Mol. Life Sci.* **79**, 562 (2022).
182. Zschocke, J. et al. Progressive infantile neurodegeneration caused by 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency: a novel inborn error of branched-chain fatty acid and isoleucine metabolism. *Pediatr. Res.* **48**, 852–855 (2000).
183. Zschocke, J. HSD10 disease: clinical consequences of mutations in the *HSD17B10* gene. *J. Inherit. Metab. Dis.* **35**, 81–89 (2012).
184. Rauschenberger, K. et al. A non-enzymatic function of 17β-hydroxysteroid dehydrogenase type 10 is required for mitochondrial integrity and cell survival. *EMBO Mol. Med.* **2**, 51–62 (2010).
185. Bhatta, A., Dienemann, C., Cramer, P. & Hillen, H. S. Structural basis of RNA processing by human mitochondrial RNase P. *Nat. Struct. Mol. Biol.* **28**, 713–723 (2021).
186. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).
187. Chung, W. C., Moyle, S. S. & Tsai, P. S. Fibroblast growth factor 8 signaling through fibroblast growth factor receptor 1 is required for the emergence of gonadotropin-releasing hormone neurons. *Endocrinology* **149**, 4997–5003 (2008).
188. Hong, S. et al. Dominant-negative kinase domain mutations in FGFR1 can explain the clinical severity of Hartsfield syndrome. *Hum. Mol. Genet.* **25**, 1912–1922 (2016).
189. Simonis, N. et al. FGFR1 mutations cause Hartsfield syndrome, the unique association of holoprosencephaly and ectrodactyly. *J. Med. Genet.* **50**, 585–592 (2013).
190. Cavenee, W. K. et al. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**, 779–784 (1983).
191. Stein, Y., Rotter, V. & Aloni-Grinstein, R. Gain-of-function mutant p53: all the roads lead to tumorigenesis. *Int. J. Mol. Sci.* **20**, 6197 (2019).
192. Kehrer-Sawatzki, H. et al. Phenotypic and genotypic overlap between mosaic NF2 and schwannomatosis in patients with multiple non-intradermal schwannomas. *Hum. Genet.* **137**, 543–552 (2018).
193. Nogué, C. et al. DGCR8 and the six hit, three-step model of schwannomatosis. *Acta Neuropathol.* **143**, 115–117 (2022).
194. Hoekstra, A. S. et al. Loss of maternal chromosome 11 is a signature event in SDHAF2, SDHD, and VHL-related paragangliomas, but less significant in SDHB-related paragangliomas. *Oncotarget* **8**, 14525–14536 (2017).
195. Burnichon, N. et al. Risk assessment of maternally inherited SDHD paraganglioma and phaeochromocytoma. *J. Med. Genet.* **54**, 125–133 (2017).
196. Twigg, S. R. et al. Cellular interference in craniofrontonasal syndrome: males mosaic for mutations in the X-linked EFNB1 gene are more severely affected than true hemizygotes. *Hum. Mol. Genet.* **22**, 1654–1662 (2013).
197. Mincheva-Tasheva, S., Nieto Guil, A. F., Homan, C. C., Gecz, J. & Thomas, P. Q. Disrupted excitatory synaptic contacts and altered neuronal network activity underpins the neurological phenotype in PCDH19-clustering epilepsy (PCDH19-CE). *Mol. Neurobiol.* **58**, 2005–2018 (2021).



## Acknowledgements
The authors wish to thank U. Moog and S. Rudnik for helpful comments on earlier drafts of this manuscript, and J. Amberger for help with Online Mendelian Inheritance in Man (OMIM) statistics. A.O.M.W. acknowledges part-funding from NIHR Oxford Biomedical Research Centre.


## Author contributions
All authors contributed to all aspects of the article; J.Z. developed the concept, wrote the initial draft of the manuscript and carried out major revisions.

## Competing interests
The authors declare no competing interests.

## Additional information

**Correspondence** should be addressed to Johannes Zschocke.



**Peer review information** *Nature Reviews Genetics* thanks Alessandra Renieri, Zornitza Stark and Hans Waterham for their contribution to the peer review of this work.




## Related links
**Matched Annotation from NCBI and EMBL-EBI (MANE):** https://www.ncbi.nlm.nih.gov/refseq/MANE/
**Online Mendelian Inheritance in Man:** https://omim.org/statistics/geneMap