

REVIEW

Open Access



# Dosage-sensitive genes in evolution and disease

Alan M. Rice and Aoife McLysaght\*

## Abstract

For a subset of genes in our genome a change in gene dosage, by duplication or deletion, causes a phenotypic effect. These dosage-sensitive genes may confer an advantage upon copy number change, but more typically they are associated with disease, including heart disease, cancers and neuropsychiatric disorders. This gene copy number sensitivity creates characteristic evolutionary constraints that can serve as a diagnostic to identify dosage-sensitive genes. Though the link between copy number change and disease is well-established, the mechanism of pathogenicity is usually opaque. We propose that gene expression level may provide a common basis for the pathogenic effects of many copy number variants.

A dramatic but transient form of dosage alteration occurs during every cell cycle where there is a drastic disruption to the relative ratios of gene copy number, with early-replicating DNA regions being twice as abundant as late-replicating regions during S-phase. Dosage sensitivity would predict that this should be compensated, and indeed elaborate mechanisms exist to mitigate this imbalance—it was recently discovered that in eukaryotes histone-mediated dosage compensation mechanisms dampen expression from early-replicating loci during cell replication, thus rebalancing the gene products with those from late-replicating loci [11, 12]. By contrast, in bacteria the differential ratio of early- and late-replicating genes is used as a cell cycle signal [12]. These systems have deep differences, but in both cases we see that even a transient change in relative gene dosage has noticeable consequences.

## Gene dosage matters

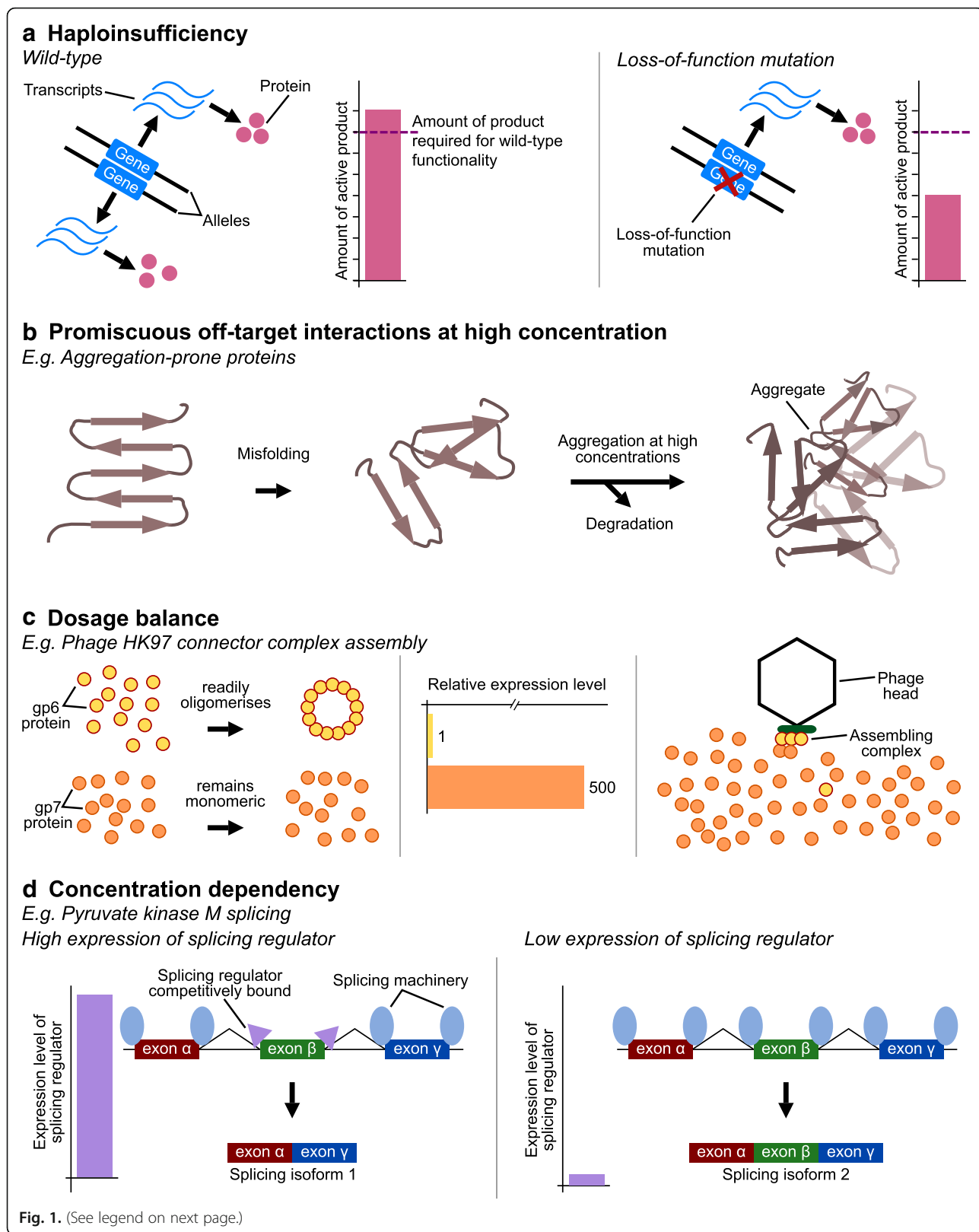
At the evolutionary level, gene duplication is an important and common process [1]; at the population level, copy number variation is the most abundant kind of genetic variation per base-pair [2]; and at the individual level, gene expression is often noisy [3]. All of these observations add up to the conclusion that for many of our genes there is good tolerance for changes in dosage. Indeed, Sewell Wright argued that even the very phenomenon of genetic dominance is suggestive of a tolerance of gene dosage changes [4]. However, for a significant fraction of the genome alteration of gene dosage has deleterious effects. This is most plainly seen in the association of copy number variants (CNVs) with human disease, including heart disease, cancers, diabetes and neuropsychiatric disorders, among others [5–8]. This dosage sensitivity reflects the generally linear relationship between gene copy number and protein product in most cases [9, 10].

## Dosage sensitivity

There are several different ways in which gene dosage can matter (Fig. 1). Haploinsufficiency, where a hemizygous state does not produce sufficient gene product for correct function, proposed by Wright as a source of dominant negative effects [4], is perhaps the most intuitive of the dosage constraints and has long been recognised as a cause of human disease [13], 22q11 deletion syndrome being just one well-studied example [14]. A recent large survey of human genetic variation identified over 3000 genes in the human genome with a near-total absence of loss-of-function alleles, suggesting that many of these are in fact haploinsufficient; over 70% of these were not previously associated with disease [15].

By contrast, why the presence of a surplus copy of a perfectly good gene should be deleterious is less obvious. Charcot-Marie-Tooth disease, a hereditary neuropathy, was one of the first human diseases shown to be due to duplication of a dosage-sensitive gene, *PMP22* [16, 17]. The fact that some phenotypically normal individuals carry both a duplication and a compensatory deletion of this gene supports the dosage sensitivity model rather than any other regulatory or structural effects as the underlying mechanism

\* Correspondence: aoife.mclysaght@tcd.ie  
Smurfit Institute of Genetics, Trinity College Dublin, University of Dublin,  
Dublin 2, Ireland



(See figure on previous page.)

**Fig. 1.** Various types of dosage sensitivity. Dosage sensitivity can be due to any of several different mechanisms. **a** For some proteins there is a minimum amount of active product required for normal function (haploinsufficiency). A hemizygous deletion or other loss of function allele will reduce the amount of active product below the threshold for functionality. **b** Some proteins form inappropriate interactions at high concentration, such as protein aggregation. These aggregates may themselves be toxic, or may phenocopy a deletion by removing the proteins from availability. **c** Dosage-balanced genes have constrained relative stoichiometry, for example the ratio of gp6 protein to gp7 protein in phage HK97 must be correct in order to achieve correct protein complex assembly. If gp6 is present in excess it preferentially forms large homomers, thus becoming unavailable to form the complex with gp7. **d** A simplified, hypothetical example of concentration-dependent activity based on the splicing of pyruvate kinase M, where the concentration of the splicing regulator determines its location of binding, which in turn determines which isoform is produced. Panel **d** is a modified from [25]

behind the disease condition [18]. In some cases, especially for intrinsically disordered proteins, the basis of pathogenicity may lie in an increased propensity for low-affinity off-target interactions at high concentrations [19, 20]. For example, extra copies of the  $\alpha$ -synuclein gene (*SNCA*) are associated with early-onset Parkinson's disease, possibly due to greater protein concentration increasing the likelihood of protein aggregation [21, 22]. Though, mechanistically speaking, why protein aggregates should have such devastating effects remains unclear [23].

Yet other genes are sensitive to both increases and decreases in copy number: many developmental morphogens act in a concentration-dependent manner [24]; whether pyruvate kinase M (*PKM*) is spliced into the adult or embryonic isoform depends on the concentration of hnRNP proteins, with high concentrations in cancer cells resulting in the ectopic production of the embryonic form [25]; and some sets of genes have constrained stoichiometry such that deviations from the normal ratio is deleterious, that is, they are dosage balanced [26]. Members of protein complexes may be particularly dosage balanced as deviations from the correct ratios of subunits can disrupt the biochemistry of protein complex assembly in non-linear ways, such that a 50% decrease in the amount of one component can result in a greater than 50% decrease in the amount of active product. Similarly, and somewhat counter-intuitively, even an increase in one of the components can result in a decrease in the amount of complete protein complex produced [20, 27–32].

The dosage balance model contends that, for genes that are in stoichiometric balance, any perturbation of their relative ratios is deleterious [26, 32, 33]. Under this model, sets of dosage-balanced genes can only change copy number in concert or not at all. There are multiple lines of evidence for this model, including: the copy number of different components of the ribosome co-vary to retain stoichiometric balance [34]; in some cases an artificially induced overexpression phenotype can be rescued by the overexpression of an interacting partner [35]; and genes whose products form protein complexes are not normally duplicated [36].

### Dosage-sensitive genes are duplicated by polyploidy or not at all

The knowledge that some genes in the genome are sensitive to copy number changes, should they occur, suggests a model where these genes have persistent sensitivity to evolutionary dosage changes, whereas other genes have no such sensitivity. The tendency of a gene to be duplicated or not is referred to as 'duplicability'. The observation that duplicability is a relatively stable property of a gene, with some genes consistently found as singletons and others repeatedly independently duplicated across distant lineages [37], supports the existence of ancient and persistent dosage constraints on genes. However, it is useful to remember that these dosage constraints are not a limit on the absolute number of copies of a gene, but on the ratio of a gene product to other components of the cell, either in terms of overall concentration or in terms of specific interacting partners [26, 27, 38–40]. So, the non-duplicability of dosage-sensitive genes relates to one-by-one duplication or loss and not to concerted events. Indeed, if a given set of dosage-balanced genes were to be linked on a chromosome, then a single segmental duplication including all of them may not be deleterious [41].

By definition, even genes that are not normally 'duplicable' are duplicated by whole genome duplication (WGD; or polyploidy). This event creates no deleterious imbalance, because, even though all of the genes are duplicated, the ratios remain unchanged. During the subsequent period of extensive gene loss that has followed every known polyploidy event [42] deletion of some, but not all, of a set of dosage-balanced genes would be deleterious, being imbalanced. This leads to the prediction that such genes should be preferentially retained through this period of purging of paralogs [43]. Consistent with expectations based on dosage sensitivity, we and others have observed that genes that are not generally duplicable by small-scale duplication (SSD) are in fact disproportionately retained after polyploidy [44–48]. The patterns of general duplicability by SSD and of retention after WGD are so contrasting as to result in almost completely non-overlapping groups of genes.

The dosage balance model prediction that WGD paralogs (termed 'ohnologs' [42, 49]) should be enriched for dosage-sensitive genes is supported by the observation

that most ohnologs are unduplicated even in lineages that diverged prior to the WGD event and do not experience subsequent duplications except if by another WGD event [44, 47]. This duplication constraint extends into recent population polymorphism: we found that ohnologs are rarely observed in CNVs in healthy individuals, whereas genes that are frequently copied by SSD also commonly have benign CNVs [47]. In other words, the general trend is that genes that can be individually duplicated, are; by contrast dosage-sensitive genes are duplicated by polyploidy, or not at all.

### CNVs and dosage sensitivity

The existence of CNVs has been long known, but since they were recognised as a significant category of genetic variation 13 years ago [50] our understanding of how they are generated and their phenotypic consequences has grown. Like any kind of genetic variation, the global distribution of CNVs is determined by demography and selection [51–53]. Many CNVs are selectively neutral [2], sometimes even when present as homozygous deletions [54]. Instances of positive selection include adaptations to diet [55] and pathogen resistance [56]. However, a significant number of CNVs are deleterious and associated with disorders [57, 58], including developmental delay [6], hearing loss [59], heart disease [7] and neuropsychiatric conditions [60, 61].

Obviously, any CNV changes the copy number of genes contained within its breakpoints, but it is not necessarily the case that the phenotype of any given CNV is due to gene dosage changes as the CNV will also potentially disrupt genes, uncouple genes from their regulatory sequences, or alter chromosome three-dimensional organisation [62–68] (Fig. 2). Nonetheless, dosage sensitivity of the encompassed genes is the most popular hypothesis to explain pathogenic CNVs, with multiple examples known [69]. This view was supported by an evolutionary analysis of gene duplication and loss across mammals where we found that genes in pathogenic CNVs have much more conserved copy number than genes observed in CNVs in healthy individuals [48]. This observation is uniquely explained by copy number constraints on the enclosed genes rather than any other model of CNV pathogenicity.

### Dosage sensitivity and genome evolution

Not surprisingly, dosage-sensitive genes have also played an enormous role in the evolution of gene content and gene expression of sex chromosomes. Not only did they precipitate the evolution of elaborate dosage compensation mechanisms [70–72], but they also have shaped gene content through both purifying selection and relocation of dosage-sensitive genes to autosomes [73–77]. For other dosage-sensitive genes that did not relocate to autosomes, especially members of large protein complexes, we and

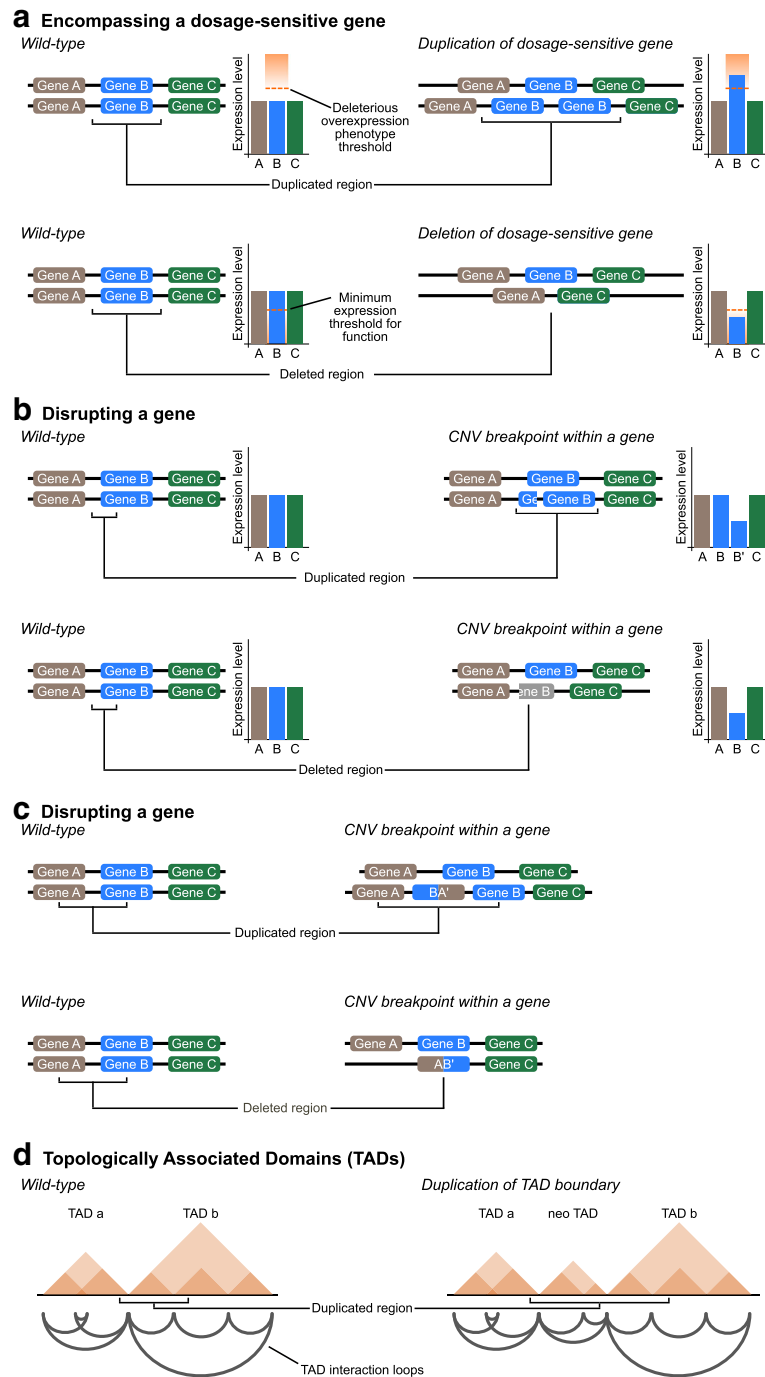
others have shown that the expression of sex-chromosome-linked genes and their autosomal interacting partners has evolved so as to maintain stoichiometric balance [78, 79].

Because dosage-sensitive genes are refractory to duplication events, and because duplications are often long, encompassing multiple genes [80], the simple presence of dosage-sensitive genes has incidental effects on neighbouring genes. The likelihood that a duplication of a given gene also includes a dosage-sensitive gene should decrease with the physical distance between them. Using ohnologs as a proxy for dosage-sensitive genes, we found evidence in support of this model, as the closer a gene is to an ohnolog the less likely it is to be duplicated [81]. This effect is sufficiently strong to create SSD and CNV deserts in the human genome [81].

One of the principal mechanisms of generation of CNVs is by non-allelic homologous recombination (NAHR; recombination events between different loci with high sequence similarity) [82]. Approximately 10% of the human genome is subject to recurrent CNVs due to the existence of NAHR hotspots [83]. These hotspots are created by the presence of segmental duplications (low-copy repeat sequences) that are at least 95% identical at the DNA level, at least 10 kb long, and located between 0.05 and 10 Mb apart [84]. At least 2129 known recurrent pathogenic CNVs occur at NAHR hotspots [85] and in at least two cases new human disease-associated NAHR hotspots were created by recent lineage-specific segmental duplication events [17, 86]. In both cases there is an as yet unproven claim that the duplication event was itself adaptive, thus compensating for the risk of disease in offspring [86, 87]. It remains unknown how the propensity of segmental duplications flanking dosage-sensitive genes to generate pathogenic CNVs has impacted upon genome evolution. One might expect purifying selection to destroy the NAHR hotspots around dosage-sensitive genes, perhaps by genome rearrangement events that eliminate the proximity of the segmental duplications.

### Evolutionary patterns are an informative trait to identify human disease genes

One of the recurrent pathogenic CNVs in humans occurs at 22q11, resulting in 22q11 deletion syndrome, which has a variable phenotype including heart defects, developmental disorders and schizophrenia [14]. Similarly, recurrent NAHR at 16p11.2 generates duplication and deletion CNVs, both of which are pathogenic, but which have different, 'mirrored' phenotypes impacting metabolism, developmental delay and neuropsychiatric traits [88, 89]. At both 22q11 and 16p11.2 there is a large CNV but also a smaller, 'critical' CNV with the same phenotype. As such, in each case the phenotype is considered to be due to dosage sensitivity of some of the genes within the smaller regions. However, even within the critical regions the number of genes is



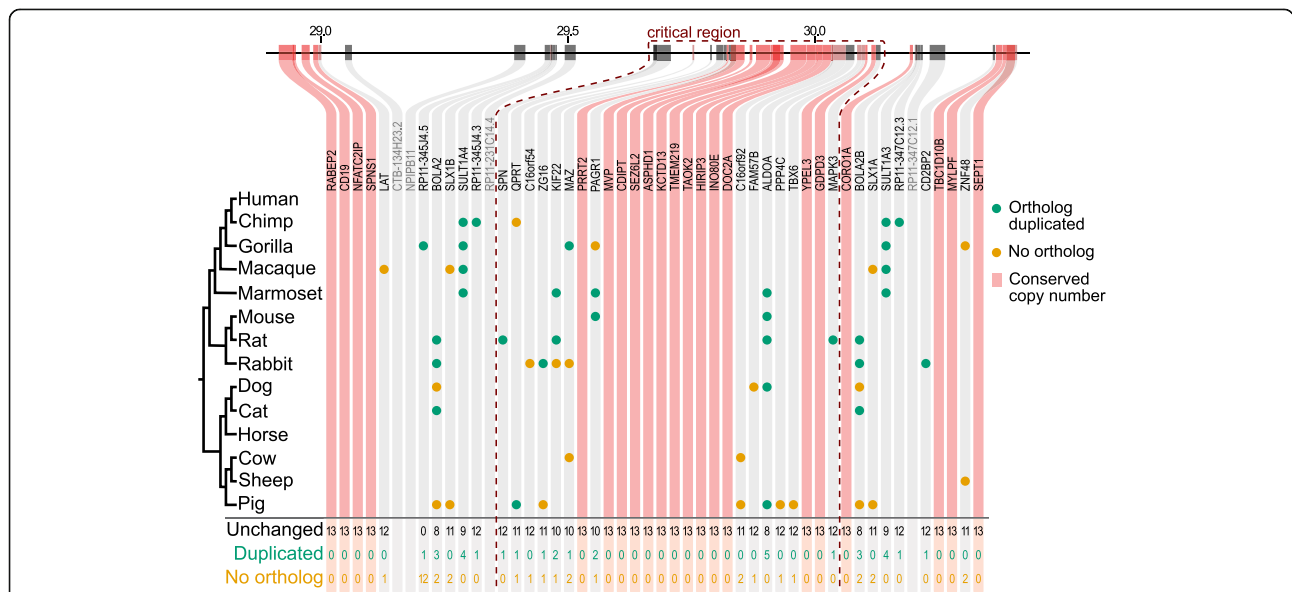
**Fig. 2.** Multiple different ways in which a CNV can have a pathogenic effect. **a** CNVs cause duplication and/or deletion of the enclosed genes. If one or more of those genes is dosage-sensitive then there will be a consequent phenotype, usually deleterious. **b, c** Alternatively, CNVs with breakpoints within a gene disrupt the gene by truncation (**b**) or formation of chimeras (**c**). Gene truncation will usually result in loss of function, but may alternatively result in a gain of function, dominant negative effect. Chimeric genes have unpredictable effects, and may be pathogenic. **d** Topologically associating domains (TADs) are structural units in the three-dimensional organisation of the genome and play a large role in mediating gene–enhancer interactions and other aspects of gene expression regulation. TADs are isolated from each other by TAD boundaries, which are determined by protein binding sites. CNVs encompassing TAD boundaries create new TADs. These can result in rewiring of gene enhancer interactions including the isolation of a gene from its regulator or the placement of a gene under the regulation of an inappropriate enhancer. Disruption of TADs has been associated with human disease [64–66, 68]

still quite large, numbering 28 and 26 for 22q11 and 16p11.2, respectively. An important challenge is to identify the most likely candidate genes for the phenotype from among these.

Evolutionary analysis provides a powerful method for pinpointing the dosage-sensitive genes within these and other CNV regions. Our detailed inspection of the evolutionary copy number conservation of the genes in the 22q11 deletion syndrome region revealed that orthologs of 16 out of the 28 genes are present (never lost) across 13 mammalian genomes analysed [48]. Similarly, in the 16q11.2 region, 13 out of the 26 genes in the critical region have completely conserved copy number (1:1 orthologs) in all mammalian genomes analysed (Fig. 3). These completely conserved genes fit the profile expected of dosage-sensitive genes, and as such are attractive candidate genes for the syndrome.

Trisomies are chromosomal abnormalities that are at least conceptually comparable to large CNVs. Trisomy 21, which results in Down's syndrome, is the most common human trisomy, occurring in approximately 1 in 700 live births [90]. The next most common human trisomies, 18 and 13, can survive briefly post-birth and result in Edward's syndrome and Patau syndrome, respectively. Other trisomies occur but are inviable [91]. The high frequency of trisomy 21 reflects the fact that it results in a relatively mild syndrome.

The pathogenic effects of the trisomy are likely to be due to a combination of the effects of the specific genes on the chromosome [47, 90] and general expression dysregulation [92]. In yeast the phenotype of an aneuploidy is largely independent of the identity of the chromosome [93], suggestive of a general disruption, not specific to the biochemical function of particular genes. Why human trisomy 21, 18 and 13 in particular are unusually viable is probably related to the low number of genes encoded on each chromosome. However, as with pathogenic CNVs, it is likely that only a subset of these are dosage sensitive. There is a very interesting correlation with the number of ohnologs (here again, a proxy for dosage-sensitive genes) and trisomy severity. Human chromosome 21 has the most common (least severe) trisomy and the smallest number of ohnologs. Similarly, chromosomes 18 and 13 have the next smallest numbers of ohnologs, and the next highest trisomy incidences, respectively (Table 1). By contrast, none of the mouse trisomies is viable (only mouse trisomy 19 survives 1–4 weeks post-birth [94]). However, there is a similar correlation between the number of ohnologs and the gestational survival of trisomies (Table 2). At least for human chromosome 21 we found that this number of ohnologs is surprisingly small even given its small number of genes [47]. Three-quarters of previously reported Down's syndrome candidate genes were independently discovered by this evolutionary analysis, but other genes were also



**Fig. 3.** Evolutionary conservation of copy number of genes in the 16p11.2 recurrent CNV region. The genes in the 16p11.2 region are illustrated across the top, with Mb co-ordinates indicated above and gene names below. The critical region (dashed outline) indicates a smaller CNV that exhibits the same phenotype as the larger CNV and so is considered sufficient for the syndrome. For each of the mammals, if the ortholog is duplicated it is represented by a green dot, and if not found (presumed deleted) it is represented by an orange dot. Otherwise the copy number is unchanged with respect to human. Where a given gene has 1:1 orthologs across all 13 mammals tested this is indicated by a red vertical stripe. Genes in the region that were not amenable to this analysis are indicated by greyed-out names. Copy number conservation data are from [48]. *BOLA2*, *SLX1* and *SULT1A* are part of a human-specific duplication with paralogs present on both flanks of the critical region and which increased the susceptibility to NAHR [86]

**Table 1** Human autosomes with common trisomies, in ascending order of number of ohnologs

Human chromosome number	Number of dosage-balanced ohnologs <sup>a</sup>	Trisomy syndrome	Frequency (number of live births) <sup>b</sup>
21	61	Down's	1/700
18	98	Edward's	1/5000
13	138	Patau	1/16,000

<sup>a</sup>Identified from gene trees in Ensembl v86 [109] as genes that duplicated at the base of the vertebrate tree with no subsequent SSD

<sup>b</sup>Frequency data from [90] and

ghr.nlm.nih.gov/trisomy-18; ghr.nlm.nih.gov/trisomy-13

identified as under evolutionary constraint that were not previously recognised to have an association with the syndrome. Thus, we argued that these genes, distinguished by their characteristic pattern of evolutionary copy number conservation, are interesting candidate genes for Down's syndrome [47]. Though aspects of the syndromes are clearly related to the specific genes on the chromosome, the correlation with number of genes encoded and perhaps specifically with the number of ohnologs suggests a general relationship with the extent of the burden of dosage-sensitive genes.

### Gene expression burden as a possible explanation for duplication phenotypes

As mentioned earlier, we and others found that genes that are successfully duplicated by SSD are usually not retained after WGD, and vice versa [46, 47, 81, 95, 96]. Curiously, these contrasts also carry over into the differences in gene expression level and coding sequence length. Whereas in human the median expression for SSD paralogs is 13.4 RPKM (reads per kilobase of transcript per million mapped reads) and the median CDS length is 1206 nucleotides, these values are much higher for ohnologs (23.7 RPKM and 1557 nucleotides, respectively). A similar pattern is seen in paramecium [97].

**Table 2** Mouse autosomes in ascending order of number of ohnologs

<i>Mus musculus</i> chromosome number	Number of dosage-balanced ohnologs <sup>a</sup>	Trisomy survival <sup>b</sup>
18	194	To term
16	201	14 days post-fertilisation—term
12	218	12–17 days post-fertilisation
13	232	13 days post-fertilisation—term
19	235	1–4 weeks post birth

<sup>a</sup>Identified from gene trees in Ensembl v86 [109] as genes that duplicated at the base of the vertebrate tree with no subsequent SSD

<sup>b</sup>Mouse trisomy survival data obtained from [94]. No unlisted trisomies survive past 19 days post-fertilisation

These contrasting patterns of expression for SSD paralogs and ohnologs could be explained if there are very different consequences of duplicating one (or a few) highly expressed gene(s) by SSD, compared to balanced duplication of all genes simultaneously by WGD. These different consequences might arise not only because some genes have dosage constraints (such as the requirement to maintain balanced ratios between specific gene products) and thus cannot be duplicated individually [38, 47, 98], but also because an extra copy of a highly expressed gene may be costly in terms of cellular resources [99–101].

This latter idea is consistent with observations from yeast where it was shown that overexpression of highly expressed genes has a greater negative effect than of less highly expressed genes [35]. The reported experiments linked the deleterious phenotype to the protein burden rather than the protein biochemical function (the phenotype was recapitulated when the protein sequence was replaced by green fluorescent protein (GFP)) but did not dissect the nature of that burden.

WGD defies this cost, as can be seen plainly in the readiness with which plant genome ploidy increases [102], an extreme example being oilseed rape which has a 72-fold increase since the origin of angiosperms owing to multiple WGD events [103]. This makes intuitive sense if the WGD increases and draws down cellular resources evenly, with no net difference compared to the pre-WGD genome. It has previously been shown that the greater the proportionate increase in copy number, the greater the phenotypic consequence; that is, adding one extra copy to a haploid is more dramatic than adding one extra copy to a diploid [30]. Thus, the greater the number of copies of a given gene, the lesser the impact of one more copy. One could therefore predict that the effects of SSD in a genome with a history of multiple WGD events, like that of oilseed rape, would be much reduced compared to an outgroup.

### Is gene expression a zero-sum game?

The idea of a cost associated with gene expression has rich theoretical support and experimental evidence [99–101, 104–106]. Furthermore, highly expressed genes are expected to be particularly dosage sensitive [35, 97]. Protein expression is a significant fraction of a cell's energy budget [99, 100, 107]. The cost of expression can be expanded to include a model where overexpression of one locus is not merely a waste, but actually sequesters cellular resources away from other genes. Clearly the number of RNA polymerases available for transcription and the number of ribosomes available for translation are both finite, but are they sometimes limiting? Experiments in *Escherichia coli* found the rather surprising result that overexpression of one locus could lead to a depletion of ribosomes [104]. Other

experiments showed that the cost of overexpression is not due to protein biochemical function or amino acid usage, but can be fully explained by the cost of the process of gene expression [101]. Rather than simply requiring resources, overexpression of some genes should naturally titrate out polymerases, ribosomes and other cellular resources so that they become unavailable for other genes. In other words, a duplication of a 'greedy', highly expressed gene might exert a deleterious effect by indirectly lowering the expression of other genes. This model views gene expression as a 'zero sum game', where an increase in one gene may cause decreased output from another. This is consistent with a recent model of human disease dubbed the 'omnigenic' model, where regulatory changes to any gene expressed in a disease-relevant tissue may contribute to disease whether or not there is a direct mechanistic link to the disease phenotype [108]. The authors of the omnigenic model suggest that gene regulatory networks are so interconnected as to allow changes in any expressed gene to affect any other.

Under the 'zero sum' model the barriers to duplication, that is, the costs of duplication, will sometimes lie not in the biochemical function of the gene that has been duplicated (though this is undoubtedly the case in many instances) but in the cost of the process of gene expression in terms of both energy (the cell spends seven ATPs for every amino acid of a protein) and the sequestration of the cell machinery such as polymerases and ribosomes. If RNA polymerases and ribosomes are limiting factors in the amount of gene expression possible from a given cell, then titration of these macromolecules by the doubling of a long, highly expressed gene would have knock-on consequences for the expression of other highly expressed genes, which may become reduced to pathogenic levels. Notably, this cost would not be incurred in WGD because all components of the system, including expression machinery, are duplicated equally, and all ratios remain constant.

### Outlook – Gene expression as a keystone to understanding copy number constraints?

This view of gene dosage sensitivity may sit alongside other more well-established modes of dosage sensitivity [30]; however, it is currently speculative. In particular one must query whether duplication of one highly expressed locus would have sufficient impact on cellular resources when expressed at physiological levels to have any effect. The costs ought to matter more in rapidly proliferating cells, which may limit this to microbial organisms [100], but they may also be relevant for quickly growing tissues. This is an interesting area to explore as it has the potential to explain some common trends in evolutionary gene duplication by different mechanisms and link this propensity to disease phenotypes of human CNVs and trisomies.

### Authors' contributions

AMcL and AR conceived the ideas, conducted analyses and wrote the manuscript. Both authors read and approved the final manuscript.

### Acknowledgements

We thank Laurence Hurst, Yoichiro Nakatani and all members of the McLysaght research group for valuable discussions. This work is supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/European Research Council grant agreement 309834.

### Competing interests

The authors declare that they have no competing interests.

Published online: 01 September 2017

### References

- Conrad B, Antonarakis SE. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet.* 2007;8:17–35.
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010;464:704–12.
- Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science.* 2012;336:183–7.
- Wright S. Physiological and evolutionary theories of dominance. *Am Nat.* 1934;68(714):24–53.
- Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature.* 2008;455:232–6.
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet.* 2011;43(9):838–46.
- Glessner JT, Bick AG, Ito K, Homsy JG, Rodriguez-Murillo L, Fromer M, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res.* 2014;115:884–96.
- Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature.* 2010;464:713–20.
- Khan Z, Ford MJ, Cusanovich DA, Mitran A, Pritchard JK, Gilad Y. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science.* 2013;342:1100–4.
- Ishikawa K, Makanae K, Iwasaki S, Ingolia NT, Moriya H. Post-translational dosage compensation buffers genetic perturbations to stoichiometry of protein complexes. *PLoS Genet.* 2017;13:e1006554.
- Voichek Y, Bar-Ziv R, Barkai N. Expression homeostasis during DNA replication. *Science.* 2016;351:1087–90.
- Bar-Ziv R, Voichek Y, Barkai N. Dealing with gene-dosage imbalance during S phase. *Trends Genet.* 2016;32:717–23.
- Fisher E, Scambler P. Human haploinsufficiency—one for sorrow, two for joy. *Nat Genet.* 1994;7:5–7.
- Karayiorgou M, Simon TJ, Gogos JA. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat Rev Neurosci.* 2010;11:402–16.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
- Lupski JR, Garcia CA. Molecular genetics and neuropathology of Charcot-Marie-Tooth disease type 1A. *Brain Pathol.* 1992;2:337–49.
- Keller MP, Seifried BA, Chance PF. Molecular evolution of the CMT1A-REP region: a human- and chimpanzee-specific repeat. *Mol Biol Evol.* 1999;16:1019–26.
- Hirt N, Eggemann K, Hyrenbach S, Lambeck J, Busche A, Fischer J, et al. Genetic dosage compensation via co-occurrence of PMP22 duplication and PMP22 deletion. *Neurology.* 2015;84:1605–6.
- Vavouri T, Semple JJ, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell.* 2009;138:198–208.



20. Cardarelli L, Maxwell KL, Davidson AR. Assembly mechanism is the key determinant of the dosage sensitivity of a phage structural protein. *Proc Natl Acad Sci U S A*. 2011;108:10168–73.
21. Irvine GB, El-Agnaf OM, Shankar GM, Walsh DM. Protein aggregation in the brain: the molecular basis for Alzheimer's and Parkinson's diseases. *Mol Med*. 2008;14:451–64.
22. Schulte C, Gasser T. Genetic basis of Parkinson's disease: inheritance, penetrance, and expression. *TACG*. 2011;4:67–80.
23. Ross CA, Poirier MA. Protein aggregation and neurodegenerative disease. *Nat Med*. 2004;10(Suppl):S10–7.
24. Rogers KW, Schier AF. Morphogen gradients: from generation to interpretation. *Annu Rev Cell Dev Biol*. 2011;27:377–407.
25. Chen M, David CJ, Manley JL. Concentration-dependent control of pyruvate kinase M mutually exclusive splicing by hnRNP proteins. *Nat Struct Mol Biol*. 2012;19:346–54.
26. Birchler JA, Bhadra U, Bhadra MP, Auger DL. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol*. 2001;234:275–88.
27. Veitia RA. Nonlinear effects in macromolecular assembly and dosage sensitivity. *J Theor Biol*. 2003;220:19–25.
28. Veitia RA. Exploring the molecular etiology of dominant-negative mutations. *Plant Cell*. 2007;19:3843–51.
29. Veitia RA, Bottani S, Birchler JA. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet*. 2008;24:390–7.
30. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A*. 2012;109:14746–53.
31. Veitia RA, Birchler JA. Models of buffering of dosage imbalances in protein complexes. *Biol Direct*. 2015;10:42.
32. Veitia RA, Potier MC. Gene dosage imbalances: action, reaction, and models. *Trends Biochem Sci*. 2015;40:309–17.
33. Veitia RA, Birchler JA. Dominance and gene dosage balance in health and disease: why levels matter! *J Pathol*. 2010;220:174–85.
34. Gibbons JG, Branco AT, Godinho SA, Yu S, Lemos B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc Natl Acad Sci U S A*. 2015;112:2485–90.
35. 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*. 2013;23:300–11.
36. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature*. 2003;424:194–7.
37. Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell*. 2016;28:326–44.
38. Birchler JA, Riddle NC, Auger DL, Veitia RA. Dosage balance in gene regulation: biological implications. *Trends Genet*. 2005;21:219–26.
39. Veitia RA. Exploring the etiology of haploinsufficiency. *Bioessays*. 2002;24:175–84.
40. Veitia RA. Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics*. 2004;168:569–74.
41. Teichmann SA, Veitia RA. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*. 2004;167:2121–5.
42. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*. 2001;2:333–41.
43. Veitia RA. Paralogs in polyploids: one for all and all for one? *Plant Cell*. 2005;17:4–11.
44. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, et al. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*. 2005;102:5454–9.
45. Freeling M, Thomas BC. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res*. 2006;16:805–14.
46. Makino T, Hokamp K, McLysaght A. The complex relationship of gene duplication and essentiality. *Trends Genet*. 2009;25:152–5.
47. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A*. 2010;107:9270–4.
48. Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat Commun*. 2017;8:14366.
49. Ohno S. Evolution by gene duplication. Berlin Heidelberg: Springer; 1970.
50. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36:949–51.
51. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–54.
52. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526:75–81.
53. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. *Science*. 2010;330:641–6.
54. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16:172–83.
55. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nat Genet*. 2007;39:1256–60.
56. Hardwick PH, Ménard A, Sironi M, Milet J, Garcia A, Sese C, et al. Haptoglobin (HP) and Haptoglobin-related protein (HPR) copy number variation, natural selection, and trypanosomiasis. *Hum Genet*. 2013;133:69–83.
57. Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet*. 2016;48(10):1107–11.
58. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. *Annu Rev Genet*. 2011;45:203–26.
59. Shearer AE, Kolbe DL, Azaiez H, Sloan CM, Frees KL, Weaver AE, et al. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med*. 2014;6:37.
60. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2016;49(1):27–35.
61. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnúsdóttir B, Morgen K, Arnarsdóttir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014;505:361–6.
62. Reymond A, Henrichsen CN, Harewood L, Merla G. Side effects of genome structural changes. *Curr Opin Genet Dev*. 2007;17:381–6.
63. Zhang F, Gu W, Hurler ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451–81.
64. Ibn-Salem J, Köhler S, Love MI, Chung H-R, Huang N, Hurler ME, et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol*. 2014;15:423.
65. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015;161:1012–25.
66. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet*. 2016;32:225–37.
67. Xie T, Yang Q-Y, Wang X-T, McLysaght A, Zhang H-Y. Spatial colocalization of human ohnolog pairs acts to maintain dosage-balance. *Mol Biol Evol*. 2016;33:2368–75.
68. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 2016;538:265–9.
69. Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. *Science*. 2015;349:aab3761–1.
70. Ohno S. Sex chromosomes and sex-linked genes. Berlin, Heidelberg: Springer; 1967.
71. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, et al. Sex determination: why so many ways of doing it? *PLoS Biol*. 2014;12:e1001899.
72. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. *Nat Commun*. 2016;7:12087.
73. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*. 2014;508:494–9.
74. Hughes JF, Skaletsky H, Koutseva N, Pyntikova T, Page DC. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. *Genome Biol*. 2015;16:104.
75. White MA, Kitano J, Peichel CL. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol Biol Evol*. 2015;32:1981–95.
76. Zimmer F, Harrison PW, Dessimoz C, Mank JE. Compensation of dosage-sensitive genes on the chicken Z chromosome. *Genome Biol Evol*. 2016;8:eww075–1242.

77. Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet.* 2017;49:387–94.
78. Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GAB. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A.* 2012; 109:5346–51.
79. Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, et al. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* 2012;10, e1001328.
80. Samonte RV, Samonte RV, Eichler EE, Eichler EE. Segmental duplications and the evolution of the primate genome. *Nat Rev Genet.* 2002;3:65–72.
81. Makino T, McLysaght A, Kawata M. Genome-wide deserts for copy number variation in vertebrates. *Nat Commun.* 2013;4:2283.
82. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016;17:224–38.
83. Mefford HC, Eichler EE. Duplication hotspots, rare genomic disorders, and common disease. *Curr Opin Genet Dev.* 2009;19:196–204.
84. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet.* 2005;77:78–88.
85. Dittwald P, Gambin T, Gonzaga-Jauregui C, Carvalho CMB, Lupski JR, Stankiewicz P, et al. Inverted low-copy repeats and genome instability—a genome-wide analysis. *Hum Mutat.* 2013;34:210–20.
86. Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, Sudmant PH, et al. Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature.* 2016;536(7615):205–9.
87. Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, et al. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.* 2001;11:1018–33.
88. Jacquemont S, Raymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature.* 2011;478:97–102.
89. Arbogast T, Ouagazzal A-M, Chevalier C, Kopanitsa M, Afinowi N, Migliavacca E, et al. Reciprocal effects on neurocognitive and metabolic phenotypes in mouse models of 16p11.2 deletion and duplication syndromes. *PLoS Genet.* 2016;12, e1005709.
90. Antonarakis SE. Down syndrome and the complexity of genome dosage imbalance. *Nat Rev Genet.* 2016;18(3):147–63.
91. van den Berg MMJ, van Maarle MC, van Wely M, Goddijn M. Genetics of early miscarriage. *Biochim Biophys Acta.* 1822:2012:1951–9.
92. Letourneau A, Santoni FA, Bonilla X, Sailani MR, Gonzalez D, Kind J, et al. Domains of genome-wide gene expression dysregulation in Down/s syndrome. *Nature.* 2014;508:345–50.
93. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science.* 2007;317:916–24.
94. Epstein CJ. Mouse monosomies and trisomies as experimental systems for studying mammalian aneuploidy. *Trends Genet.* 1985;1:129–34.
95. Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol.* 2007;8:R209.
96. Guan Y, Dunham MJ, Troyanskaya OG. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics.* 2007;175:933–43.
97. Gout J-F, Kahn D, Duret L, Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* 2010;6, e1000944.
98. Veitia RA. Gene dosage balance: deletions, duplications and dominance. *Trends Genet.* 2005;21:33–5.
99. Wagner A. Energy constraints on the evolution of gene expression. *Mol Biol Evol.* 2005;22:1365–74.
100. Wagner A. Energy costs constrain the evolution of gene expression. *J Exp Zool.* 2007;308:322–4.
101. Stoebel DM, Dean AM, Dykhuizen DE. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics.* 2008;178:1653–60.
102. Conant GC, Birchler JA, Pires JC. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol.* 2014;19:91–8.
103. Chalhouf B, Denoed F, Liu S, Parkin IAP, Tang H, Wang X, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science.* 2014;345:950–3.
104. Dong H, Nilsson L, Kurland CG. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol.* 1995;177:1497–504.
105. Hurst LD, Randerson JP. Dosage, deletions and dominance: simple models of the evolution of gene expression. *J Theor Biol.* 2000;205:641–7.
106. MacLean RC, Fuentes-Hernandez A, Greig D, Hurst LD, Gudelj I. A mixture of "cheats" and "co-operators" can enable maximal group benefit. *PLoS Biol.* 2010;8, e1000486.
107. Lane N, Martin W. The energetics of genome complexity. *Nature.* 2010;467: 929–34.
108. Boyle EA, Li Yi, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* 2017;169:1177–86.
109. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. *Nucleic Acids Res.* 2016;44:D710–6.