Copy number variants, diseases and gene expression

Charlotte N. Henrichsen, Evelyne Chaignat and Alexandre Reymond[*]

The Center for Integrative Genomics, Genopode Building, University of Lausanne, CH-1015 Lausanne, Switzerland

Copy number variation (CNV) has recently gained considerable interest as a source of genetic variation likely to play a role in phenotypic diversity and evolution. Much effort has been put into the identification and mapping of regions that vary in copy number among seemingly normal individuals in humans and a number of model organisms, using bioinformatics or hybridization-based methods. These have allowed uncovering associations between copy number changes and complex diseases in whole-genome association studies, as well as identify new genomic disorders. At the genome-wide scale, however, the functional impact of CNV remains poorly studied. Here we review the current catalogs of CNVs, their association with diseases and how they link genotype and phenotype. We describe initial evidence which revealed that genes in CNV regions are expressed at lower and more variable levels than genes mapping elsewhere, and also that CNV not only affects the expression of genes varying in copy number, but also have a global influence on the transcriptome. Further studies are warranted for complete cataloguing and fine mapping of CNVs, as well as to elucidate the different mechanisms by which they influence gene expression.

INTRODUCTION

A fundamental question in current biomedical research is to establish a link between genomic variation and phenotypic differences, which encompasses both the seemingly neutral polymorphic variation, as well as the pathological variation that causes or predisposes to disease. DNA copy number variants, defined as stretches of DNA larger than 1 kb that display copy number differences in the normal population (1), have recently gained considerable interest as a source of genetic diversity likely to play a role in functional variation (2–5).

Extensive studies were performed to establish CNV catalogs for multiple species, however little is known about the functional impact of CNVs at the cellular and organismal level. Here, we review recent works on the identification of CNVs, their association with disease and their global impact on gene expression, as well as discuss possible mechanisms by which CNVs exert this influence.

COPY NUMBER VARIATION IN HUMAN AND MODEL ORGANISMS

Whole-genome CNV catalogs have been established for the human (4,6–8), the mouse (9–16), the rat (17), the chimpanzee (7,18), the rhesus macaque (19) and *Drosophila melanogaster* (20,21) (Table 1). CNVs were identified by array comparative genome hybridization (aCGH) (22), either with bacterial artificial chromosome or oligonucleotide probes. Alternatively, bioinformatics mining of whole genome shotgun (WGS) sequences was used to predict CNVs in human, macaque and rat (17,23,24). As observed in humans, distinct array platforms yielded considerably different datasets with little overlap [43% and 37% overlap, respectively (4,16)], whereas studies performed using the same platform were shown to be highly reproducible (12,16). Similarly, Guryev *et al.* (17) found a very good correspondence between WGS-based predictions and hybridization methods. On the other hand, changes in calling methods were shown to modify the number of CNV predictions, with methods that congruently analyze aCGH data from multiple samples significantly increasing the number of detectable CNVs (14,16).

General characteristics of the CNVs identified in human and model organisms are summarized in Table 1. Interestingly, the CNVs reported in *D. melanogaster* are one order of magnitude shorter (maximum 33 kb or 0.017% of the fly genome) than those found in mammals (up to 3 Mb or 0.12% of the mouse genome). A certain synteny has been observed among primates with 22% and 25% of chimpanzee and macaque

---

[*]To whom correspondence should be addressed. Tel: +41 21 692 3960, Fax: +41 21 692 3965, Email: alexandre.reymond@unil.ch

Table 1. Genome-wide copy number variation (CNV) surveys in human and model organisms

Model organism	Platform	Probe density	Resolution	% of genome covered by CNVs	Size range of CNVs	References
Human	BAC arrays		1 Mb		Up to 2 Mb	(2,8)
Human	Representational Oligonucleotide Microarray Analysis (ROMA)	1/17 kb	35 kb	~1.49%	700 bp to 3.19 Mb	(5)
Human	BAC arrays and SNP genotyping arrays			12%	Up to 6 Mb	(4)
Human	Bioinformatics and BAC arrays				8–329 kb	(6)
Drosophila	Spotted PCR array	1/4.1 kb			Up to 33 kb	(20)
Drosophila	Euchromatic genome tiling array	1/36 bp	<1 kb	~2%	Up to 35 kb	(21)
Mouse	BAC arrays		1 Mb			(9,11,15)
Mouse	Oligonucleotide arrays	1/6kb	50 kb	Up to 10%	40–3000 kb	(10,12,16)
Mouse (single strain C57BL/6J)	Representational oligonucleotide microarray analysis (ROMA)				4–4000 kb	(13)
Mouse	Segmental duplication-specific oligonucleotide array			60% of segmental duplications		(14)
Rat	Bioinformatics and oligonucleotide arrays	1/5 kb		1.4%		(17)
Macaque	Oligonucleotide arrays	1/6 kb	40 kb		32–710 kb	(19)
Chimpanzee	Human BAC arrays (2)	1/1 Mb	1 Mb			(7)

BAC, bacterial artificial chromosome; PCR, polymerase chain reaction.

CNVs overlapping human CNVs (7,19). In addition, CNVs identified in multiple macaques were frequently observed in multiple human samples, suggesting the existence of hotspots for CNV (19). Indeed, structural genomic rearrangements are facilitated by repetitive sequences that provide the substrate for non-allelic homologous recombination (25). Consistently, CNVs are associated with segmental duplications (SDs) in mammals (2,4,7,8,11,12,19) and to centromeres and telomeres (17,26), structures also associated with SDs. For example, as much as 60% of the basepairs mapping inside C57BL/6J SDs were recently shown to be polymorphic among classical mouse inbred strains (14). In other words, an average of 20 Mb of DNA is variable in number of copies between two laboratory mice in these regions. Similarly, in *D. melanogaster*, CNVs cluster in regions of the fly genome enriched in tandem repeats, and over half the transposons assessed were demonstrated to vary in number of copies between strains (20).

On the other hand, the lower boundaries of the ranges of CNV size reflect the resolution of the platforms used, as well as the power and stringency of the prediction algorithms, rather than the actual minimal size of CNVs. A complete cataloguing of CNVs and other structural changes in mammals and other animal models still awaits further analyses with increased resolution and the advent of different approaches that take advantage of high-throughput sequencing (27,28). For example, paired-end sequence tag mapping allows the detection of smaller structural variants (down to a few kilobases), however, this approach is biased towards small events, meaning that the different detection techniques should be used in parallel (28,29).

Several personal genomes from different ethnicity and of both sexes have been sequenced, each dataset reinforcing the catalog of variants that exist in human genomes (29–32). For example, 1.4% of the total sequencing data from the genome of James Watson does not map with the NCBI36 reference assembly, and 23 new CNV regions were identified (32). Similarly, 2500 and 5000 structural variants were uncovered in an Asian and an African genome, respectively (29,30),

while Kidd and colleagues (27) found 525 CNV regions in a sample of four African and four non-African genomes. In one Asian male, 34% of the structural variants discovered are novel, and 33 genes were found to be completely or partially disrupted, a change that is homozygous for 30% of these loci. For example, *CYP4F12* on chromosome 19 was broken in two segments by an inversion (29). This epitomizes that the sequencing of large numbers of individuals is needed to fully capture the diversity and complexity of the genome.

The number of polymorphisms of copy numbers identified in the mouse, the only animal model for which we have accumulated a fair amount of data at the moment, appears to be similar, if not greater than that reported in human (14,16). However, this type of variation appears to be far more clustered in discrete regions in this model than in humans, probably reflecting the unnatural man-driven establishment of inbred lines. Of note, CNVs and single nucleotide polymorphisms (SNPs) cluster in different regions of the domesticated mouse genome (10,33). Consistently, Cutler *et al.* (10) found that CNV segments were significantly enriched among sequences with low and moderate SNP content. The artificially repeated brother–sister mating of mice over generations has eliminated recessive lethal alleles, but may have increased the fixation of potentially (slightly) deleterious CNV alleles. Consistently, significantly more CNVs were identified in classical inbred strains than in wild-caught *Mus musculus domesticus* mice (16).

CNV regions are not only polymorphic within a species, but spontaneous *de novo* rearrangements can occur even in an inbred population, as shown by recent studies of the classical C57BL/6J strain (13,16,34). Egan *et al.* and Henrichsen *et al.* identified and validated 38 and 39 CNVs in this isogenic strain, respectively, 18 of which occurred through recurrent mutations (13,16). In addition, a single-copy and a duplication allele of the CNV spanning the *Ide* locus were found to be in Hardy–Weinberg equilibrium 14 years after their appearance in the Jackson Laboratory colony (34). These findings illustrate the highly dynamic nature of CNV regions and challenge the notion of the isogenicity of inbred strains (35).

PHENOTYPIC CONSEQUENCES OF CNVS

With their prevalence, e.g. 10% of the mouse autosomal genome and 60% of its duplicated regions (14,16), CNVs should constitute significant contributors to intraspecific genetic variation. Consistently, an initial case of adaptive CNV alleles under positive selection was recently uncovered (36). Similarly, copy numbers of individual CNVs have been associated with diseases or susceptibility to diseases, either through dosage of a single gene (37–44), a contiguous set of genes [e.g. Williams-Beuren syndrome and infantile spasms (45,46), DiGeorge Syndrome (47), Smith-Magenis syndrome (48), Potocki-Lupski syndrome (49)] or allele combinations in the case of complex diseases, especially of the central nervous system. Recent studies in schizophrenia (50,51), autism (52,53) and mental retardation (54–56) revealed multiple disease-causing rearrangements, but also lead to the description of novel, clinically recognizable microdeletion syndromes and recurrent rearrangements causing variable phenotypes (57–65). Structural rearrangements in three regions, namely 1q21.1, 15q11.2 and 15q13.3 were shown to be associated with both mental retardation and schizophrenia. A more detailed investigation of cases might shed light on a possible link between these conditions. Interestingly, while deletions at all three loci were linked to schizophrenia and related psychoses (50,51), the maternal duplication of the entire 15q11-q13 region was shown to cause autism spectrum disorder (ASD) (66). Similarly, duplication of band 1q21.1 was identified in patients with ASD, whereas both deletions and duplications were associated with mental retardation. These examples suggest that some conditions might not be associated with a specific number of copies of a particular CNV, but rather that the simple presence of a structural change at a given position of the human genome may cause perturbation in particular pathways regardless of gene dosage.

CNV has also been shown to induce phenotypes in non-mammalian vertebrates and in invertebrates. For example, copy number amplification of regions containing the mannose-binding lectin and the *pfmdr1* and *gch1* genes influence the zebrafish susceptibility to bacterial infection and malaria parasite *Plasmodium falciparum* drug resistance, respectively (67,68).

FROM GENOTYPE TO PHENOTYPE—BRIDGING THE GAP

Phenotypic effects of genetic differences, both at the single-nucleotide and large-scale level, are supposedly brought about by changes in expression levels, either directly affecting the genes concerned by the genetic change, or indirectly through position effects or downstream pathways and regulatory networks (69,70). Thus, transcriptome analyses and single-gene expression studies seem appropriate proxies to study the consequences of CNV.

A first attempt to estimate the contribution of CNVs to gene expression variation found that in human lymphoblastoid cell lines, changes in copy numbers capture about one-fifth of the detected genetic variation (71). This type of approach, however, does not allow to follow the influence of CNVs in multiple tissues nor developmental stages, as an individual will have been the source of a limited number of cell lines, typically a skin fibroblast and a lymphoblastoid cell line or, more rarely, different cell lines from umbilical cord. This difficulty can be overcome using animal models, as recent analyses suggest that CNVs are also important contributors to their phenotypic variation (Table 1). In some cases given CNVs were shown to cause the same phenotype in a model organism and in human patients, further exemplifying the validity of models studies. For example, Aitman *et al*. (40) showed that lower copy number of the *Fcgr3* gene predisposed both rats and humans to immunologically mediated glomerulonephritis. In addition to natural CNV, structural modifications can be engineered into model organisms, e.g. mice, and thus allow the study of different alleles of a CNV, as well as position effects, in a genetically homogenous background (72,73).

Two recent reports attempted to assess the effects of CNVs on tissue transcriptomes of rats and mice (16,17). Both studies report an over-representation of differentially expressed genes among CNV-mapping transcripts. Globally, a weak yet significant positive correlation was found between relative expression level and gene dosage (see examples in Fig. 1). This correlation was driven by a restricted number of genes, as only a fraction of genes comprised between 5% and 18% in function of the tissue and the rodent species analyzed showed a strong correlation between number of copies and relative expression levels (16,17) (see example in Fig. 1A and possible mechanism in Fig. 2B). In about two-third of the genes, however, the number of copies had no effect on relative expression levels in any tissue (see example in Fig. 1C), suggesting either dosage compensation mechanisms or the incomplete inclusion of regulatory elements in the deletion/duplication event (see possible mechanisms in Fig. 2). In addition, genetic imprinting may also modulate the expression at a CNV locus, in a parental allele-dependent manner (74). The first hypothesis is supported by the observation that expression of some genes correlated with gene dosage in some tissues but not in others (see examples in Fig. 1D and mechanisms in Fig. 2B and C), whereas the second is supported by studies of the quantitative collinearity of *Hoxd* genes, which showed that their expression levels are correlated with their rank in the Hoxd cluster (75) (see mechanism in Fig. 2D). Alternatively, the multiple copies of a gene might be mapping to different chromatin environments and thus be regulated differentially (Fig. 2E); indeed, aCGH does not allow discriminating tandem from non-tandem duplications. Furthermore, for 2–15% of the genes depending on the tissue and the rodent assessed, the recorded relative expression levels were significantly inversely correlated with CNV (16,17) (see examples in Fig. 1B). The mechanism of this effect is still poorly understood, but may be explained by two models summarized in Fig. 2F and 2G. In the first model, a negative correlation between number of copies and relative expression is explained by immediate early genes (IEG, Fig. 2F). These genes are initially expressed at levels proportional to their number of copies (Fig. 2A and B). These IEG induce directly or indirectly the expression of a repressor which, by a negative feedback loop, reduces or even abolishes the expression of the CNV gene (Fig. 2F).
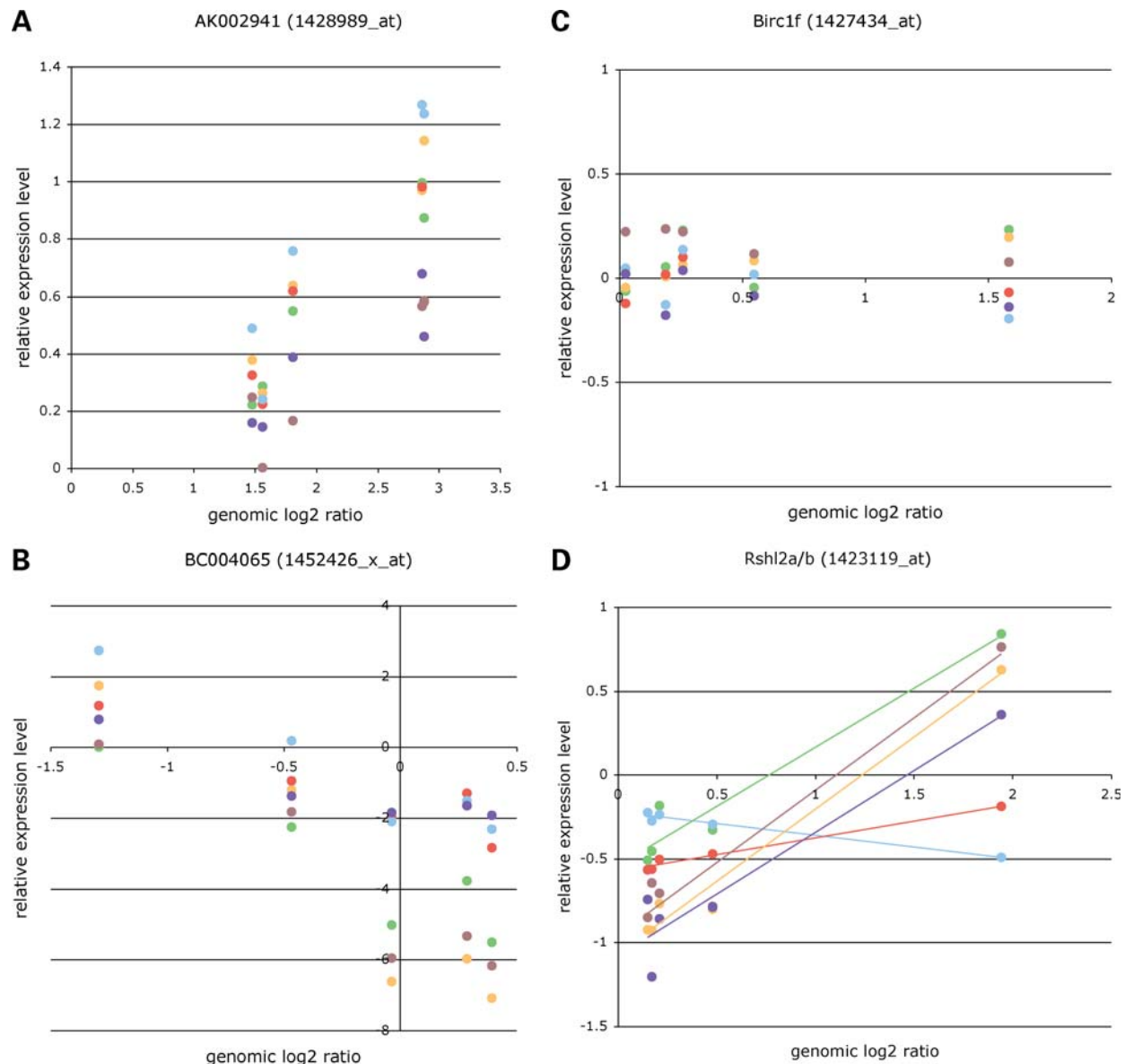
Figure 1. Relative expression levels as a function of relative copy number. Examples of positive correlation [*AK002941* gene, 1428989_at probeset; (**A**)], negative correlation [*BC004065* gene, 1452426_x_at probeset; (**B**)], absence of correlation [*Birc1f* gene, 1427434_at probeset; (**C**)] and multiple trend correlation [*Rshl2a/b* gene, 1423119_at probeset; (**D**)] between relative expression levels and array CGH log2 ratio measured in testis (blue), brain (red), liver (brown), lung (green), heart (purple) and kidney (orange).

In the second hypothesis, the extra copies of a gene impair through steric hindrance their access to a specific transcription factory, where this particular locus should be transcribed (76) (Fig. 2G).

Genes mapping to CNVs show lower expression levels than transcripts that do not vary in numbers of copies, both in fly and mouse (16,20). These observations suggest that CNV genes might be more specific in their expression patterns than other genes. In support of this notion, CNV genes were detectable in a smaller number of tissues than non-CNV genes both in the mouse and *Drosophila* (16,20). In addition, Dopman and Hartl (20) observed that the CNV genes expressed in the fly midgut and male accessory glands were enriched for functions relevant to these tissues, defense

response in both cases and post-mating behavior genes in the latter case.

Large-scale rearrangements do not only affect expression by altering gene dosage. The deletion associated with Williams-Beuren syndrome was shown to modify the expression of some of the normal-copy number neighboring genes in human lymphoblastoid and skin fibroblast cell lines (77). Consistently, Stranger *et al*. (71) observed that although CNVs capture 18% of the detected genetic variation in lymphoblastoid cell lines, more than half of the identified associations between copy numbers and expression levels involved genes mapping outside the CNV intervals. Likewise, an engineered duplication of a segment of mouse chromosome 11, that models the rearrangement present in Potocki-Lupski syndrome patients,
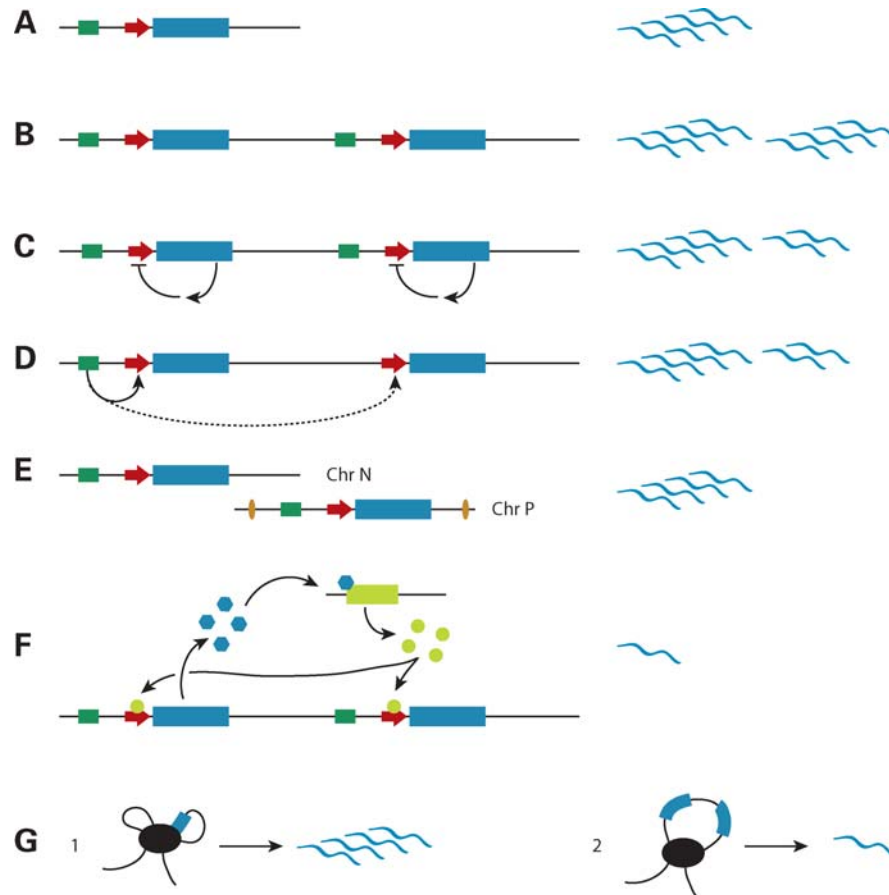
Figure 2. Duplication scenarios and their influence on expression. (**A**) Single copy gene locus. The gene intron–exon region (blue box), the gene promoter (red arrow) and its enhancer (green box) are shown. Transcript levels are indicated schematically on the right. (**B**) Complete tandem duplication including the regulatory region. (**C**) Complete tandem duplication including the regulatory region of a gene under a compensatory mechanism. A negative feedback loop reduces the gene transcription level. (**D**) Complete tandem duplication excluding the regulatory region. The single enhancer only weakly influences expression of the second copy of the gene, which is expressed at a lower level. (**E**) Complete non-tandem duplication including the regulatory region. The duplicated locus maps to another chromosome region where a different chromatin context, e.g. insulators (yellow ellipses), modifies its expression level. (**F**) Immediate early gene model. In the presence of a duplication, the concentration of CNV gene product (blue hexagons) is sufficient to induce a repressor (light green box), the product of which (light green disks) blocks the expression of the CNV gene. (**G**) A tandem duplication (2) physically impairs the access of the CNV genes copies to the transcription factory where it should be transcribed (1).

was shown to affect the expression of both the genes mapping within the duplicated interval and its flanks (72). Thus, CNVs also profoundly affect the expression of genes located in their vicinity. This effect extends over half a megabase into their genomic neighborhoods (16,17). But how may changes in copy number of CNV regions alter the expression of genes in their vicinity? Different CNV-induced mechanisms that include the physical dissociation of the transcription unit from its *cis*-acting regulators (78,79), modification of transcriptional control through alteration of chromatin structure (80–83) and modification of the positioning of chromatin within the nucleus and/or within a chromosome territory of a genomic region (84,85) might play a role, both individually or in combination. Copy number changes might also influence gene expression through perturbation of transcript structure (70,86,87). Similar mechanisms might explain how some pathways are perturbed without correlation with a particular number of copies of a CNV (see above). Detailed investigations of the different mechanisms by which CNVs influence gene expression are warranted to shed light on how CNVs alter the architecture of chromosomal segments and thus influence the expression of genes.

CONCLUSION

The identification of CNVs in both human and model organisms and their association with diseases and phenotypes have confirmed that these genome structural changes are important contributors to variation. They exert their influence by modifying the expression of genes mapping within and close to the rearranged region.

Conflict of Interest statement. None declared.

FUNDING

REFERENCES

1. Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E. and Feuk, L. (2007) Challenges and standards in integrating surveys of structural variation. *Nat. Genet.*, **39**, S7–S15.
2. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.
3. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R. *et al.* (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, **451**, 998–1003.
4. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.
5. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.
6. Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D. *et al.* (2005) Fine-scale structural variation of the human genome. *Nat. Genet.*, **37**, 727–732.
7. Perry, G.H., Tchinda, J., McGrath, S.D., Zhang, J., Picker, S.R., Caceres, A.M., Iafrate, A.J., Tyler-Smith, C., Scherer, S.W., Eichler, E.E. *et al.* (2006) Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl Acad. Sci. USA*, **103**, 8006–8011.
8. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R. *et al.* (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.*, **77**, 78–88.
9. Adams, D.J., Dermitzakis, E.T., Cox, T., Smith, J., Davies, R., Banerjee, R., Bonfield, J., Mullikin, J.C., Chung, Y.J., Rogers, J. *et al.* (2005) Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nat. Genet.*, **37**, 532–536.
10. Cutler, G., Marshall, L.A., Chin, N., Baribault, H. and Kassner, P.D. (2007) Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.*, **17**, 1743–1754.
11. Li, J., Jiang, T., Mao, J.H., Balmain, A., Peterson, L., Harris, C., Rao, P.H., Havlak, P., Gibbs, R. and Cai, W.W. (2004) Genomic segmental polymorphisms in inbred mouse strains. *Nat. Genet.*, **36**, 952–954.
12. Graubert, T.A., Cahan, P., Edwin, D., Selzer, R.R., Richmond, T.A., Eis, P.S., Shannon, W.D., Li, X., McLeod, H.L., Cheverud, J.M. *et al.* (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.*, **3**, e3.
13. Egan, C.M., Sridhar, S., Wigler, M. and Hall, I.M. (2007) Recurrent DNA copy number variation in the laboratory mouse. *Nat. Genet.*, **39**, 1384–1389.
14. She, X., Cheng, Z., Zollner, S., Church, D.M. and Eichler, E.E. (2008) Mouse segmental duplication and copy number variation. *Nat. Genet.*, **40**, 909–914.
15. Snijders, A.M., Nowak, N.J., Huey, B., Fridlyand, J., Law, S., Conroy, J., Tokuyasu, T., Demir, K., Chiu, R., Mao, J.H. *et al.* (2005) Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res.*, **15**, 302–311.
16. Henrichsen, C.N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H. and Reymond, A. (2009) Segmental copy number variation shapes tissue transcriptome. *Nat. Genet.*, doi:10.1038/ng.345.
17. Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S.A., Cook, S., Pravenec, M., Aitman, T., Jacob, H., Shull, J.D. *et al.* (2008) Distribution and functional impact of DNA copy number variation in the rat. *Nat. Genet.*, **40**, 538–545.
18. Kehrer-Sawatzki, H. and Cooper, D.N. (2007) Structural divergence between the human and chimpanzee genomes. *Hum. Genet.*, **120**, 759–778.
19. Lee, A.S., Gutierrez-Arcelus, M., Perry, G.H., Vallender, E.J., Johnson, W.E., Miller, G.M., Korbel, J.O. and Lee, C. (2008) Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum. Mol. Genet.*, **17**, 1127–1136.
20. Dopman, E.B. and Hartl, D.L. (2007) A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **104**, 19920–19925.
21. Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O. and Long, M. (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, **320**, 1629–1631.
22. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **20**, 207–211.
23. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003–1007.
24. Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L., Venter, J.C., Wilson, R.K. *et al.* (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**, 222–234.
25. Gu, W., Zhang, F. and Lupski, J.R. (2008) Mechanisms for human genomic rearrangements. *PathoGenetics*, **1**, 4.
26. Nguyen, D.Q., Webber, C. and Ponting, C.P. (2006) Bias of selection on human copy-number variants. *PLoS Genet.*, **2**, e20.
27. Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
28. Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L. *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*, **318**, 420–426.
29. Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J. *et al.* (2008) The diploid genome sequence of an Asian individual. *Nature*, **456**, 60–65.
30. Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.
31. Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G. *et al.* (2007) The diploid genome sequence of an individual human. *PLoS Biol.*, **5**, e254.
32. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–876.
33. Lakshmi, B., Hall, I.M., Egan, C., Alexander, J., Leotta, A., Healy, J., Zender, L., Spector, M.S., Xue, W., Lowe, S.W. *et al.* (2006) Mouse genomic representational oligonucleotide microarray analysis: detection of copy number variations in normal and tumor specimens. *Proc. Natl Acad. Sci. USA*, **103**, 11234–11239.
34. Watkins-Chow, D.E. and Pavan, W.J. (2008) Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res.*, **18**, 60–66.
35. Stevens, J.C., Banks, G.T., Festing, M.F. and Fisher, E.M. (2007) Quiet mutations in inbred strains of mice. *Trends Mol. Med.*, **13**, 512–519.
36. Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R. *et al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.*, **39**, 1256–1260.
37. Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J. *et al.* (2005) The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, **307**, 1434–1440.
38. Hollox, E.J., Armour, J.A. and Barber, J.C. (2003) Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am. J. Hum. Genet.*, **73**, 591–600.
39. Aldred, P.M., Hollox, E.J. and Armour, J.A. (2005) Copy number polymorphism and expression level variation of the human alpha-defensin genes *DEFA1* and *DEFA3*. *Hum. Mol. Genet.*, **14**, 2045–2052.
40. Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E. *et al.* (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature*, **439**, 851–855.
41. Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M. *et al.* (2008) Psoriasis is associated with increased beta-defensin genomic copy number. *Nat. Genet.*, **40**, 23–25.

42. Breunis, W.B., van Mirre, E., Bruin, M., Geissler, J., de Boer, M., Peters, M., Roos, D., de Haas, M., Koene, H.R. and Kuijpers, T.W. (2008) Copy number variation of the activating *FCGR2C* gene predisposes to idiopathic thrombocytopenic purpura. *Blood*, **111**, 1029–1038.

43. Yang, Y., Chung, E.K., Wu, Y.L., Savelli, S.L., Nagaraja, H.N., Zhou, B., Hebert, M., Jones, K.N., Shu, Y., Kitzmiller, K. *et al.* (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.*, **80**, 1037–1054.

44. Shao, W., Tang, J., Song, W., Wang, C., Li, Y., Wilson, C.M. and Kaslow, R.A. (2007) CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. *Genes Immun.*, **8**, 224–231.

45. Bayes, M., Magano, L.F., Rivera, N., Flores, R. and Perez Jurado, L.A. (2003) Mutational mechanisms of Williams-Beuren syndrome deletions. *Am. J. Hum. Genet.*, **73**, 131–151.

46. Marshall, C.R., Young, E.J., Pani, A.M., Freckmann, M.L., Lacassie, Y., Howald, C., Fitzgerald, K.K., Peippo, M., Morris, C.A., Shane, K. *et al.* (2008) Infantile spasms is associated with deletion of the *MAGI2* gene on chromosome 7q11.23-q21.11. *Am. J. Hum. Genet.*, **83**, 106–111.

47. Baldini, A. (2004) DiGeorge syndrome: an update. *Curr. Opin. Cardiol.*, **19**, 201–204.

48. Bi, W., Yan, J., Stankiewicz, P., Park, S.S., Walz, K., Boerkoel, C.F., Potocki, L., Shaffer, L.G., Devriendt, K., Nowaczyk, M.J. *et al.* (2002) Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse. *Genome Res.*, **12**, 713–728.

49. Potocki, L., Chen, K.S., Park, S.S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboa, K., Kashork, C.D. *et al.* (2000) Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.*, **24**, 84–87.

50. Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, **455**, 237–241.

51. Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E. *et al.* (2008) Large recurrent microdeletions associated with schizophrenia. *Nature*, **455**, 232–236.

52. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y. *et al.* (2008) Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.*, **82**, 477–488.

53. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. *et al.* (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**, 445–449.

54. Lu, X., Shaw, C.A., Patel, A., Li, J., Cooper, M.L., Wells, W.R., Sullivan, C.M., Sahoo, T., Yatsenko, S.A., Bacino, C.A. *et al.* (2007) Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. *PLoS ONE*, **2**, e327.

55. de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N. *et al.* (2005) Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.*, **77**, 606–616.

56. Friedman, J.M., Baross, A., Delaney, A.D., Ally, A., Arbour, L., Armstrong, L., Asano, J., Bailey, D.K., Barber, S., Birch, P. *et al.* (2006) Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am. J. Hum. Genet.*, **79**, 500–513.

57. Koolen, D.A., Sharp, A.J., Hurst, J.A., Firth, H.V., Knight, S.J., Goldenberg, A., Saugier-Veber, P., Pfundt, R., Vissers, L.E., Destree, A. *et al.* (2008) Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J. Med. Genet.*, **45**, 710–720.

58. Sharp, A.J., Hansen, S., Selzer, R.R., Cheng, Z., Regan, R., Hurst, J.A., Stewart, H., Price, S.M., Blair, E., Hennekam, R.C. *et al.* (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.*, **38**, 1038–1042.

59. Hannes, F.D., Sharp, A.J., Mefford, H.C., de Ravel, T., Ruivenkamp, C.A., Breuning, M.H., Fryns, J.P., Devriendt, K., Van Buggenhout, G., Vogels, A. *et al.* (2008) Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the

60. Sharp, A.J., Mefford, H.C., Li, K., Baker, C., Skinner, C., Stevenson, R.E., Schroer, R.J., Novara, F., De Gregori, M., Ciccone, R. *et al.* (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat. Genet.*, **40**, 322–328.

61. Sharp, A.J., Selzer, R.R., Veltman, J.A., Gimelli, S., Gimelli, G., Striano, P., Coppola, A., Regan, R., Price, S.M., Knoers, N.V. *et al.* (2007) Characterization of a recurrent 15q24 microdeletion syndrome. *Hum. Mol. Genet.*, **16**, 567–572.

62. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D. *et al.* (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.*, **359**, 1685–1699.

63. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Genevieve, D., Cormier-Daire, V., van Esch, H., Fryns, J.P. *et al.* (2006) Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am. J. Hum. Genet.*, **79**, 370–377.

64. Ballif, B.C., Hornor, S.A., Jenkins, E., Madan-Khetarpal, S., Surti, U., Jackson, K.E., Asamoah, A., Brock, P.L., Gowans, G.C., Conway, R.L. *et al.* (2007) Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat. Genet.*, **39**, 1071–1073.

65. Shaffer, L.G., Theisen, A., Bejjani, B.A., Ballif, B.C., Aylsworth, A.S., Lim, C., McDonald, M., Ellison, J.W., Kostiner, D., Saitta, S. *et al.* (2007) The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. *Genet. Med.*, **9**, 607–616.

66. Veenstra-Vanderweele, J., Christian, S.L. and Cook, E.H. Jr (2004) Autism as a paradigmatic complex genetic disorder. *Annu. Rev. Genom. Hum. Genet.*, **5**, 379–405.

67. Jackson, A.N., McLure, C.A., Dawkins, R.L. and Keating, P.J. (2007) Mannose binding lectin (MBL) copy number polymorphism in Zebrafish (D. rerio) and identification of haplotypes resistant to *L. anguillarum*. *Immunogenetics*, **59**, 861–872.

68. Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.C., Krishna, S., Nosten, F. and Anderson, T.J. (2007) Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.*, **24**, 562–573.

69. Dermitzakis, E.T. and Stranger, B.E. (2006) Genetic variation in human gene expression. *Mamm. Genome*, **17**, 503–508.

70. Reymond, A., Henrichsen, C.N., Harewood, L. and Merla, G. (2007) Side effects of genome structural changes. *Curr. Opin. Genet. Dev.*, **17**, 381–386.

71. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C. *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853.

72. Molina, J., Carmona-Mora, P., Chrast, J., Krall, P.M., Canales, C.P., Lupski, J.R., Reymond, A. and Walz, K. (2008) Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki-Lupski syndrome. *Hum. Mol. Genet.*, **17**, 2486–2495.

73. Yan, J., Bi, W. and Lupski, J.R. (2007) Penetrance of craniofacial anomalies in mouse models of Smith-Magenis syndrome is modified by genomic sequence surrounding rai1: not all null alleles are alike. *Am. J. Hum. Genet.*, **80**, 518–525.

74. Hogart, A., Leung, K.N., Wang, N.J., Wu, D.J., Driscoll, J., Vallero, R.O., Schanen, C. and Lasalle, J.M. (2008) Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *J. Med. Genet.*, doi:10.1136/jmg.2008.061580.

75. Montavon, T., Le Garrec, J.F., Kerszberg, M. and Duboule, D. (2008) Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev.*, **22**, 346–359.

76. Sexton, T., Umlauf, D., Kurukuti, S. and Fraser, P. (2007) The role of transcription factories in large-scale structure and dynamics of interphase chromatin. *Semin. Cell Dev. Biol.*, **18**, 691–697.

77. Merla, G., Howald, C., Henrichsen, C.N., Lyle, R., Wyss, C., Zabot, M.T., Antonarakis, S.E. and Reymond, A. (2006) Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am. J. Hum. Genet.*, **79**, 332–341.

duplication may be a rare benign variant. *J. Med. Genet.*, doi:10.1136/jmg.2007.055202.

78. Kleinjan, D.A. and van Heyningen, V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.*, **76**, 8–32.

79. Kleinjan, D.J. and van Heyningen, V. (1998) Position effect in human genetic disease. *Hum. Mol. Genet.*, **7**, 1611–1618.

80. Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R. *et al.* (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature*, **439**, 973–977.

81. Gabellini, D., Green, M.R. and Tupler, R. (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell*, **110**, 339–348.

82. Lee, J.A. and Lupski, J.R. (2006) Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, **52**, 103–121.

83. Muncke, N., Wogatzky, B.S., Breuning, M., Sistermans, E.A., Endris, V., Ross, M., Vetrie, D., Catsman-Berrevoets, C.E. and Rappold, G. (2004) Position effect on PLP1 may cause a subset of Pelizaeus-Merzbacher disease symptoms. *J. Med. Genet.*, **41**, e121.

84. Fraser, P. and Bickmore, W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413–417.

85. Heard, E. and Bickmore, W. (2007) The ins and outs of gene regulation and chromosome territory organisation. *Curr. Opin. Cell. Biol.*, **19**, 311–316.

86. Denoeud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., Chrast, J. *et al.* (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.*, **17**, 746–759.

87. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.