

REVIEW

Open Access



# Major influence of repetitive elements on disease-associated copy number variants (CNVs)

Ana R. Cardoso<sup>1,2,3</sup>, Manuela Oliveira<sup>1,2,3</sup>, Antonio Amorim<sup>1,2,3</sup> and Luisa Azevedo<sup>1,2,3\*</sup> 

## Abstract

Copy number variants (CNVs) are important contributors to the human pathogenic genetic diversity as demonstrated by a number of cases reported in the literature. The high homology between repetitive elements may guide genomic stability which will give rise to CNVs either by non-allelic homologous recombination (NAHR) or non-homologous end joining (NHEJ). Here, we present a short guide based on previously documented cases of disease-associated CNVs in order to provide a general view on the impact of repeated elements on the stability of the genomic sequence and consequently in the origin of the human pathogenic variome.

**Keywords:** Copy number variants (CNVs), Genetic diseases, Genomic structural variation, Low copy repeats, Retrotransposons, LINE, SINE, Non-allelic homologous recombination (NAHR)

## Background

Copy number variants (CNVs) are structural genomic markers (insertions or deletions) ranging in size from 1 kb to several megabytes for each copy. They are categorized as copy number polymorphisms (CNP) when multiple allelic states exist in the population or as rare copy number variants when they are found to be associated with genetic diseases (pathogenic copy number variants) [1, 2]. The origin of each repeated element of the CNV is influenced by the local genomic architecture which includes the presence of repetitive sequences within or flanking the repeated segment [3–7]. These repeated sequences drive non-allelic homologous recombination (NAHR) events which result in recurrent insertions and deletions with similar sequence sizes and clustered breakpoints [3, 6, 8] or non-homologous end joining (NHEJ) events that result in non-recurrent rearrangements that vary in terms of their size and breakpoint location [3, 6, 9]. Although several studies have been demonstrating the contribution of structural

variants to the genome architecture, few have specifically focused the influence of repeated sequences at breakpoint locations. With the aim to draw attention to these unstable regions and to establish their role in CNVs, we collated a number of cases of CNV-associated disorders proven to have been generated by low and high copy number repeats which may have influenced the degree of stability of the genomic sequence.

## Low copy repeats and their influence on pathogenic CNV formation

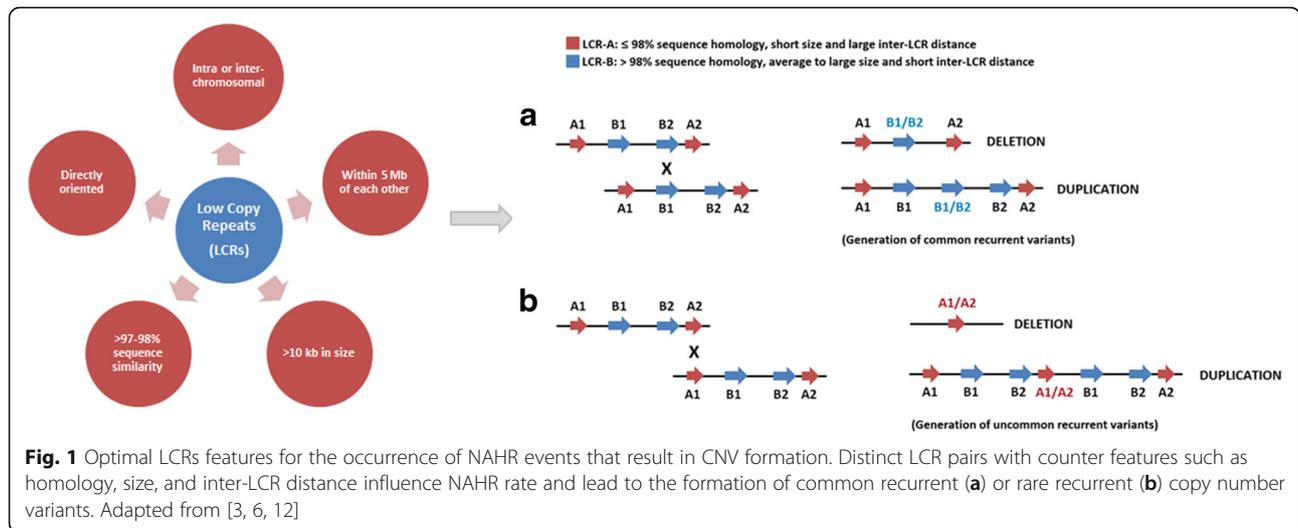
Low copy repeats (LCRs) are homologous sequences of  $\geq 1$  kb in length which are found in many copies throughout the genome since they are generated by duplication events [3, 10]. Large LCRs (>10 kb) with high sequence homology promote non-allelic homologous recombination (NAHR) [3–6, 10–12] and the misalignment of directly oriented sister chromatids carrying the LCR may promote NAHR thereby generating both duplications and deletions [4, 5] which in turn give rise to copy number variation. A schematic representation of this process is shown in Fig. 1.

Certain properties of the LCRs such as homology length, sequence similarity, and distance, serve to influence the frequency of NAHR events [3, 6, 12] (Fig. 1). As recently reviewed by Carvalho and Lupski [3], the NAHR

\* Correspondence: lazevedo@ipatimup.pt

<sup>1</sup>Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal

<sup>2</sup>IPATIMUP-Institute of Molecular Pathology and Immunology, University of Porto, Rua Júlio Amaral de Carvalho 45, 4200-135 Porto, Portugal  
Full list of author information is available at the end of the article



rate varies according to the length of the LCR sequence, the distance between distinct LCR sequences and the DNA sequence. The NAHR rate is, therefore, positively correlated with the LCR length but is inversely proportional to the distance between distinct LCRs [3, 9]. Since there is a high homology between distinct LCR sequences proximal to copy number variation regions; there is also an increased predisposition to NAHR events in these genomic regions [3, 4, 6, 9, 12].

A considerable number of disease-associated CNVs generated by LCRs have been documented and reviewed in previous works (e.g. [3, 6]), but for the purposes of this paper, we have only collated cases for which the specific repetitive element was found at the breakpoints of the structural variant and not those for which the causality of the repeats elements was only suggested. The

resulting set is presented in Table 1. For example, a complex array of LCRs spanning a 4-Mb region around the X-linked *MECP2* gene was associated with unique duplications ranging in size from 200 kb to 2.2 Mb in developmentally delayed males [13]. Duplications and deletions affecting the *PLP1* gene causing Pelizaeus-Merzbacher disease (OMIM #312080) are also associated with a specific LCR (LCR-PMD A/B pair) within a 3-Mb region flanking the gene in which a multitude of LCRs are located [14]. LCRs are also frequent at the 2q11-q21.1 locus [11], where recurrent deletions of the *NPHP1* gene (2q13) have been associated with nephronophthisis 1 (OMIM #256100). A 0.3-Mb copy number gain was detected in three X-linked intellectual disability (XLID) families and one sporadic patient [15]. The region overlapped the *GDI1* gene, an important XLID-associated

**Table 1** Repetitive elements detected at the breakpoints of CNVs associated with clinical phenotypes

Phenotype	Critical genes	Type of variant	Locus	Repetitive element involved	Ref.
MECP2 duplication syndrome	<i>MECP2, L1CAM</i>	Dup	Xq28	Several LCR-MECP2s pairs	[3, 6, 13, 44–46]
Rett syndrome	<i>MECP2</i>	Del	Xq28	Several LCR-MECP2 pairs	[6]
Neurofibromatosis type I	<i>NF1</i>	Del	17q11.2	NF1-REPs A/B/C	[3, 6]
Nephronophthisis	<i>NPHP1</i>	Del	2q13	Several LCR pairs	[11, 47–49]
Mental retardation, X-linked 41 (MRX41)	<i>GDI1</i>	Dup/Trip	Xq28	LCR-K1/L2 pair	[15]
Angelman and Prader-Willi syndromes	<i>UBE3A</i>	Del	15q11-q13	END-repeats (LCRs)	[6, 50]
Smith-Magenis syndrome	<i>RAI1</i> and <i>PMP22</i>	Del	17p11.2	SMS-REPs (LCRs)	[3, 6, 18]
Williams-Beuren syndrome	28 dosage-sensitive genes	Dup/Tripe/Del	7q11.23	A/B/C LCR blocks	[3, 6, 51]
15q13.1 microdeletion syndrome	<i>CHRNA7</i>	Dup/Trip	15q13.3	BP3/4/5	[3, 6, 52, 53]
3q29 microduplication or microdeletion syndrome	<i>DLG1, PAK2</i>	Dup/Del	3q29	A/B/C LCR blocks	[3, 54, 55]
Pelizaeus-Merzbacher disease	<i>PLP1</i>	Dup/Del	Xq22	LCR-PMD A/B pair	[3, 6, 14]
DiGeorge syndrome/velo-cardio facial syndrome	<i>COMT, TBX1</i>	Del	22q11.2	8 specific LCR22 repeats	[6, 17, 40]
Charcot-Marie-Tooth type 1A	<i>PMP22</i>	Dup	17p12	CMTA1-Reps (LCRs)	[3, 6, 37, 38]

gene highly expressed in the brain. The aberration was located in Xq28, a locus that includes other intellectual disability genes and that is frequently associated with recombination events caused by proximal LCRs (e.g., LCR K1/L2). The Angelman syndrome (AS) (OMIM #105830) and Prader-Willi syndrome (PWS) (OMIM #176270) are caused by recurrent 4-Mb deletions at the 15q11-q13 locus. The deleted region is flanked by LCRs [6] and accounts for 70 % of cases of AS and 70–75 % of cases of PWS [16]. The Smith-Magenis syndrome (SMS) (OMIM #182290) results from recurrent deletions of 3.7 Mb at 17p11.2 which account for more than 70 % of cases; about 25 % of affected individuals harbor deletions ranging from 1.5–9 Mb [6, 17]. The deletions are flanked by 200-kb highly homologous LCRs that play a role in generating meiotic NAHR events [16]. These deletions encompass the *RAI1* gene, which is critical in organ and neuronal development—patients with larger deletions manifest a more severe phenotype when the dosage-sensitive gene *PMP22* is deleted [18].

### Retrotransposons (high copy repeats) and their influence on pathogenic CNVs

Interspersed repeats are the most common type of high copy repeats, covering about 44 % of the human genome [4]. Retrotransposons account for the majority of transposable elements [5, 7, 19]. These are mobile elements that through reverse transcription have the ability to integrate into different regions [7, 19]. Long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and retrovirus-like elements (LTR transposons) are the three major categories of mammalian retrotransposons (Table 2).

Among LINEs, L1 is the most abundant element, typically of 6–8 kb in length, with the ability to increase genomic instability through NAHR events [4]. It is known that about 83 % of the human genome is prone to LINE-LINE recombination events that contribute to genomic instability and can give rise to unbalanced structural variants [20].

Alu elements are the most common SINEs and have been associated with NAHR events that lead to pathogenic duplications and deletions [3, 4, 21, 22]. Table 3

presents examples of high copy repeats that have been detected at the breakpoints of disease-associated CNVs.

Borun and colleagues [23] reported the presence of CNV breakpoints within Alu elements in the *STK11* gene which lead to the Peutz-Jeghers syndrome (OMIM #175200), where CNVs account for 30 % of cases. The 17p13.3 locus is enriched in copy number variations associated with genomic disorders, such as the Miller-Dieker syndrome (17p13.3 deletion syndrome) (OMIM #247200) and its reciprocal 17p13.3 duplication syndrome (OMIM #613215) [24]. The breakpoints of the reported CNVs at this locus are highly enriched in Alu elements, which mediate these junctions through an Alu-Alu mechanism. About 70 % of CNVs found in the *SPAST* gene have been associated with Alu recombination events [25]. Local Alu-rich architecture predisposes to the formation of pathogenic structural rearrangements associated with spastic paraplegia (OMIM #182601). An extra copy of the *LMNB1* gene at the 5q23 locus has been previously associated with autosomal dominant adult-onset demyelinating leukodystrophy (ADLD) (OMIM #169500). The analysis of twenty ADLD-affected families revealed sixteen duplications ranging from 128 to 475 kb in size, all of them spanning the *LMNB1* gene [26]. The centromeric region of the critical gene is enriched with SINE elements, particularly Alus. Alu-mediated recombination events were also found to be linked to pathogenic deletions at the *OTC* gene [27], a urea cycle gene for which a significant number of structural variants are known [28]. NAHR events between Alu repeats are also strongly correlated with the birth of structural rearrangements at the Alu-rich *BRCA1* locus [29] which is associated with breast cancer. Duplications (220 to 394 kb) and a triplication (1.61 to 2.04 Mb) of the *SNCA* gene located at 4q21 locus have been implicated in autosomal dominant Parkinson's disease (PD1 and PD4) (OMIM #168601, #605543). The phenotypic severity is consistent with a gene dosage effect [6]. Regarding recessive PD (OMIM #600116), about one third of pathogenic variants associated with the *PRKN* gene are CNVs occurring between exon 2 and exon 5, which may therefore be considered to be a

**Table 2** Main characteristics of the most abundant retrotransposons [4, 5, 7, 19]

	Retrotransposons (interspersed repeats)—44 % human genome		
	Non-long terminal repeat (LTR)		Long terminal repeat (LTR)
Repetitive element	Long interspersed nuclear repeats (LINEs)	Short interspersed nuclear repeats (SINEs)	Endogenous retroviruses (ERV)
Genomic coverage	20 %	11 %	8 %
Features	<ul style="list-style-type: none"> <li>• L1 is the most abundant class</li> <li>• Autonomous transposons</li> <li>• Reverse transcriptase (RT) encoded by LINE-1</li> </ul>	<ul style="list-style-type: none"> <li>• Alu is the most abundant class</li> <li>• Dependent on LINEs transposable machinery</li> <li>• Mobile polymerase III promoter</li> <li>• 100–400 bp in length</li> </ul>	<ul style="list-style-type: none"> <li>• Reduced transposable activity</li> <li>• Presence of <i>gag</i> and <i>pol</i> viral genes</li> </ul>

**Table 3** High copy repeats detected at the breakpoints of CNVs associated with clinical phenotypes

Phenotype	Critical genes	Type of variant	Locus	Repetitive elements involved	Ref.
Peutz-Jeghers syndrome	<i>STK11</i>	Del	19p13.3	Several AluY/AluY pairs	[23]
Spastic paraplegia 4	<i>SPAST, SLC30A6</i>	Dup/Del	2p22.3	Several Alu pairs	[25]
OTC deficiency	<i>OTC</i>	Del	Xp11.4	AluSx/AluSq pair	[27, 56]
Miller-Dieker syndrome and 17p13.3 duplication syndrome	<i>LIS1</i>	Del	17p13.3	Several Alu pairs	[6, 24]
Breast cancer	<i>BRCA1</i>	Del	17q21.31	AluSx/AluSc pair	[29, 57]
Autosomal dominant adult-onset demyelinating leukodystrophy (ADLD)	<i>LMNB1</i>	Dup/Trip	5q23.2	LIPA3 LINE repeats AluYA/AluYB pair	[26]
Azoospermia	<i>AZFa</i>	Del	Yq11	HERV15 A/B proviruses	[34, 58]
Mental retardation, X-linked 60 (MRX60)	<i>OPHN1</i>	Del	Xq12	AluY/AluY pair	[35]
Pelizaeus-Merzbacher disease	<i>PLP1</i>	Del	Xq22	AluSq/AluSx pair	[3, 6, 59]
DiGeorge syndrome/velo-cardio facial syndrome	<i>COMT, TBX1</i>	Del	22q11.2	Unclassified Alu/Alu pair	[6, 17, 40]
Charcot-Marie-Tooth type 1A	<i>PMP22</i>	Dup	17p12	AluY/AluY pair AluSg/AluSg pair	[39]
Williams-Beuren syndrome	28 dosage-sensitive genes	Dup/Del	7q11.23	AluS subfamily elements	[36]
Parkinson's disease	<i>SNCA</i>	Dup/Trip	4q21	Several Alu pairs	[32]

recombination hotspot [30, 31]. Ross and colleagues [32] reported the presence of Alu and LINE1 elements at the *SNCA* locus that may contribute to the genomic instability at this locus.

Human endogenous retroviruses (HERVs) represent about 4.9 % of the human genome [4]. Sequences with about 95 % sequence similarity were previously associated with NAHR events and recurrent CNVs, some of which with pathogenic implications [3, 33]. For example, the occurrence of NAHR between a particular set of HERV elements flanking the male fertility *AZFa* locus in the Y chromosome is strongly associated with pathogenic deletions associated with male infertility (OMIM #400042, #415000) [4, 34].

### Pathogenic copy number variants associated with both LCRs and retrotransposons

The breakpoints of some disease-associated CNVs have been reported to be caused by more than one type of repetitive elements which indicates that the same phenotype involves both low copy and high copy repeats that affect the stability of a target gene. Bergmann and colleagues [35] conducted a family study in which five brothers shared the same phenotypic pattern that included intellectual disability. The analysis of the *OPHN1* locus (Xq12) revealed the presence of a 17.6-kb intronic deletion and the breakpoints spanning the deletion revealed two highly homologous Alu repeats and additional repetitive sequences (interspersed and simple repeats).

A recurrent deletion of 1.6 to 1.8 Mb (>95 % of the patients) at the 7q11.23 locus causes the Williams-Beuren syndrome (OMIM #194050) [6]. Genes within this region are dosage-sensitive and the recurrently deleted region encompasses a total of 28 genes. This locus is characterized by highly homologous flanking LCRs that contribute to NAHR events [6]. Antonell and colleagues [36] reported the presence of Alu elements at the junctions of large duplicated blocks in 7q11.23 suggesting the influence of these retrotransposons in the generation of large LCRs.

Heterozygous duplication and reciprocal deletions of a 1.4–1.5-Mb segment at the 17p12 locus have been previously linked with the Charcot-Marie-Tooth type 1A syndrome (CMT1A) (OMIM #118220). About 70 % of CMT1A patients have a recurrent duplication of the dosage-sensitive *PMP22* locus and the NAHR event that gave rise to this copy number variation was mediated by LCRs [3, 6, 37, 38]. A study by Zhang and colleagues [39] revealed the presence of SINEs (Alu elements) and LINEs (L1 and L2) as well as LCRs within the breakpoints of rare non-recurrent deletions and duplications at the CMT1A locus.

About 96 % of the DiGeorge syndrome (DGS) (OMIM #188400)- and velo-cardio-facial syndrome (VCFS) (OMIM #192430)-affected patients harbor a 1.5–3 Mb deletion at the 22q11.2 locus that includes 24 to 30 genes [16]. The breakpoints of the common recurrent deletions at this locus are associated with LCRs [17] and one Alu sequence [40]. Both the deletions and duplications at this locus are generated by NAHR events

between the repeated regions flanking the CNV, specifically the low copy repeat known as LCR22 [41]. Furthermore, 20–25 % of individuals who harbor this deletion also show signs of schizophrenia, mood disorders, and other behavioral alterations [41].

## Conclusions

Although the majority of genetic diseases are caused by non-structural variants (e.g. [42, 43], an increasing number of causative mutations have been associated with CNVs and these cases were the focus of this short review. Low copy repeats and retrotransposons are the major contributors to CNV formation. Recurrent CNVs are mainly directed by NAHR events that occur between highly homologous LCR sequences. In terms of non-recurrent CNVs, NHEJ (among other molecular mechanisms [3]) generally occurs between sequences with a degree of homology lower than that observed between distinct LCRs. The diversity of breakpoint junctions of non-recurrent variants renders the establishment of phenotype-genotype relationships less reliable because the sequence that is deleted or duplicated in each patient is different and the affected region may also involve other genes. This review focused on disease-associated CNVs in order to show that although numerous cases of instability driven by repeated sequences around the affected locus (or loci) have been documented, we are still far from understanding all the phenotypic complexities associated with these unbalanced variants, mainly because the number of reported cases is still too small to draw general conclusions. Finally, it is important to mention that collated data, such as those presented in this paper, pertaining to the pathogenic structural variances are expected to drive future studies with the aim of establishing a map of unstable genomic hotspots which promises to be useful in the context of clinical genetic testing where the determination of the molecular basis of Mendelian and complex diseases (e.g., cancer) is of paramount importance.

## Abbreviations

CNV: Copy number variant; LCR: Low copy repeat; CNP: Copy number polymorphism; LINE: Long interspersed nuclear repeat; SINE: Short interspersed nuclear repeat; LTR: Long terminal repeat; NAHR: Non-allelic homologous recombination; NHEJ: Non-homologous end joining; XLID: X-linked intellectual disability

## Acknowledgements

Not applicable.

## Funding

IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology. This work is funded by FEDER funds through the Operational Programme for Competitiveness Factors - COMPETE and National Funds through the FCT-Foundation for Science and Technology, under the projects "PEst-C/SAU/LA0003/2013".

M. Oliveira (SFRH/BPD/66071/2009) was supported by FCT fellowships funded by POPH-QREN - Promotion of Scientific Employment, the European Social Fund, and National Funds of the Ministry of Education and Science.

## Availability of data and materials

Not applicable.

## Authors' contributions

ARC performed literature search. ARC and LA drafted the manuscript. MO and AA discussed the data and performed critical revisions to the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Author details

<sup>1</sup>Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal. <sup>2</sup>IPATIMUP-Institute of Molecular Pathology and Immunology, University of Porto, Rua Júlio Amaral de Carvalho 45, 4200-135 Porto, Portugal. <sup>3</sup>Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre S/N, 4169-007 Porto, Portugal.

Received: 19 July 2016 Accepted: 16 September 2016

Published online: 23 September 2016

## References

1. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet.* 2009;84(2):148–61.
2. Henrichsen CN, Chagnat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet.* 2009;18(R1):R1–8.
3. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016;17(4):224–38.
4. Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its roles in human copy number variation. *Genomics Inform.* 2014;12(4):136–44.
5. Jobling MA, Hollox E, Hurler M, Kivisild T, Tyler-Smith C. *Human evolutionary genetics: origins, peoples and disease.* 2nd ed. New York: Garland Science; 2014.
6. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron.* 2006;52(1):103–21.
7. Strachan T, Read A. *Human Molecular Genetics.* 4th ed. New York:Garland Science; 2011.
8. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics.* 2008;1(1):4.
9. Liu P, Lacia M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am J Hum Genet.* 2011;89(4):580–8.
10. Potier MC, Golfier G, Eichler EE. Chromosome-specific repeats (Low-copy Repeats). In: *eLS.* Chichester: Wiley; 2007.
11. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY, et al. NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. *Genome Res.* 2013;23(9):1395–409.
12. Peng Z, Zhou W, Fu W, Du R, Jin L, Zhang F. Correlation between frequency of non-allelic homologous recombination and homology properties: evidence from homology-mediated CNV mutations in the human genome. *Hum Mol Genet.* 2015;24(5):1225–33.
13. del Gaudio D, Fang P, Scaglia F, Ward PA, Craigen WJ, Glaze DG, et al. Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet Med.* 2006;8(12):784–92.
14. Lee JA, Inoue K, Cheung SW, Shaw CA, Stankiewicz P, Lupski JR. Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease. *Hum Mol Genet.* 2006;15(14):2250–65.

15. Vandewalle J, Van Esch H, Govaerts K, Verbeeck J, Zweier C, Madrigal I, et al. Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. *Am J Hum Genet.* 2009;85(6):809–22.
16. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet.* 2009;84(4):524–33.
17. Shaikh TH, Kurahashi H, Saïtta SC, O'Hare AM, Hu P, Roe BA, et al. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet.* 2000;9(4):489–501.
18. Madduri N, Peters SU, Voigt RG, Llorente AM, Lupski JR, Potocki L. Cognitive and adaptive behavior profiles in Smith-Magenis syndrome. *J Dev Behav Pediatr.* 2006;27(3):188–92.
19. Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev.* 1996;6(6):743–8.
20. Startek M, Szafranski P, Gambin T, Campbell IM, Hixson P, Shaw CA, et al. Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res.* 2015;43(4):2188–98.
21. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet.* 2002;3(5):370–9.
22. Teixeira-Silva A, Silva RM, Carneiro J, Amorim A, Azevedo L. The role of recombination in the origin and evolution of Alu subfamilies. *PLoS one.* 2013;8(6):e64884.
23. Borun P, De Rosa M, Nedoszytko B, Walkowiak J, Plawski A. Specific Alu elements involved in a significant percentage of copy number variations of the STK11 gene in patients with Peutz-Jeghers syndrome. *Fam Cancer.* 2015;14(3):455–61.
24. Gu S, Yuan B, Campbell IM, Beck CR, Carvalho CM, Nagamani SC, et al. Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Hum Mol Genet.* 2015;24(14):4061–77.
25. Boone PM, Yuan B, Campbell IM, Scull JC, Withers MA, Baggett BC, et al. The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am J Hum Genet.* 2014;95(2):143–61.
26. Giorgio E, Rolyan H, Kropp L, Chakka AB, Yatsenko S, Di Gregorio E, et al. Analysis of LMNB1 duplications in autosomal dominant leukodystrophy provides insights into duplication mechanisms and allele-specific expression. *Hum Mutat.* 2013;34(8):1160–71.
27. Quental R, Azevedo L, Rubio V, Diogo L, Amorim A. Molecular mechanisms underlying large genomic deletions in ornithine transcarbamylase (OTC) gene. *Clin Genet.* 2009;75(5):457–64.
28. Azevedo L, Stolnaja L, Tietzeova E, Hrebicek M, Hrubá E, Vilarinho L, et al. New polymorphic sites within ornithine transcarbamylase gene: population genetics studies and implications for diagnosis. *Mol Genet Metab.* 2003; 78(2):152–7.
29. Ewald IP, Ribeiro PL, Palmero EI, Cossio SL, Giugliani R, Ashton-Prolla P. Genomic rearrangements in BRCA1 and BRCA2: a literature review. *Genet Mol Biol.* 2009;32(3):437–46.
30. Toft M, Ross OA. Copy number variation in Parkinson's disease. *Genome Med.* 2010;2(9):62.
31. Hedrich K, Eskelson C, Wilmot B, Marder K, Harris J, Garrels J, et al. Distribution, type, and origin of Parkin mutations: review and case studies. *Mov Disord.* 2004;19(10):1146–57.
32. Ross OA, Braithwaite AT, Skipper LM, Kachergus J, Hulihan MM, Middleton FA, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol.* 2008;63(6):743–50.
33. Campbell IM, Gambin T, Dittwald P, Beck CR, Shuvarikov A, Hixson P, et al. Human endogenous retroviral elements promote genome instability via non-allelic homologous recombination. *BMC Biol.* 2014;12:74.
34. Bosch E, Jobling MA. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum Mol Genet.* 2003;12(3):341–7.
35. Bergmann C, Zerres K, Senderek J, Rudnik-Schoneborn S, Eggermann T, Hauser M, et al. Oligophrenin 1 (OPHN1) gene mutation causes syndromic X-linked mental retardation with epilepsy, rostral ventricular enlargement and cerebellar hypoplasia. *Brain.* 2003;126(Pt 7):1537–44.
36. Antonell A, de Luis O, Domingo-Roura X, Perez-Jurado LA. Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. *Genome Res.* 2005;15(9):1179–88.
37. Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat Genet.* 1992;2(4):292–300.
38. Weterman MA, van Ruissen F, de Wissel M, Bordewijk L, Samijn JP, van der Pol WL, et al. Copy number variation upstream of PMP22 in Charcot-Marie-Tooth disease. *Eur J Hum Genet.* 2010;18(4):421–8.
39. Zhang F, Seeman P, Liu P, Weterman MA, Gonzaga-Jauregui C, Towne CF, et al. Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am J Hum Genet.* 2010;86(6):892–903.
40. Uddin RK, Zhang Y, Siu VM, Fan YS, O'Reilly RL, Rao J, et al. Breakpoint Associated with a novel 2.3 Mb deletion in the VCFS region of 22q11 and the role of Alu (SINE) in recurring microdeletions. *BMC Med Genet.* 2006;7:18.
41. Shishido E, Aleksic B, Ozaki N. Copy-number variation in the pathogenesis of autism spectrum disorder. *Psychiatry Clin Neurosci.* 2014;68(2):85–95.
42. Ferreira F, Esteves S, Almeida LS, Gaspar A, da Costa CD, Janeiro P, et al. Trimethylaminuria (fish odor syndrome): genotype characterization among Portuguese patients. *Gene.* 2013;527(1):366–70.
43. Quelhas D, Quental R, Vilarinho L, Amorim A, Azevedo L. Congenital disorder of glycosylation type Ia: searching for the origin of common mutations in PMM2. *Ann Hum Genet.* 2007;71(Pt 3):348–53.
44. Van Esch H. MECP2 duplication syndrome. *Mol Syndromol.* 2012;2(3-5):128–36.
45. Bauters M, Van Esch H, Friez MJ, Boespflug-Tanguy O, Zenker M, Vianna-Morgante AM, et al. Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res.* 2008;18(6):847–58.
46. Carvalho CM, Zhang F, Liu P, Patel A, Sahoo T, Bacino CA, et al. Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum Mol Genet.* 2009;18(12):2188–203.
47. Saunier S, Calado J, Benessy F, Silbermann F, Heilig R, Weissenbach J, et al. Characterization of the NPHP1 locus: mutational mechanism involved in deletions in familial juvenile nephronophthisis. *Am J Hum Genet.* 2000; 66(3):778–89.
48. Yuan B, Liu P, Gupta A, Beck CR, Tejomurtula A, Campbell IM, et al. Comparative genomic analyses of the human NPHP1 locus reveal complex genomic architecture and its regional evolution in primates. *PLoS Genet.* 2015;11(12):e1005686.
49. Konrad M, Saunier S, Heidet L, Silbermann F, Benessy F, Calado J, et al. Large homozygous deletions of the 2q13 region are a major cause of juvenile nephronophthisis. *Hum Mol Genet.* 1996;5(3):367–71.
50. Amos-Landgraf JM, Ji Y, Gottlieb W, Depinet T, Wandstrat AE, Cassidy SB, et al. Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet.* 1999;65(2):370–86.
51. Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. *Am J Hum Genet.* 2003;73(1):131–51.
52. Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, et al. 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet.* 2005;77(1):154–60.
53. Ballif BC, Theisen A, Coppinger J, Gowans GC, Hersh JH, Madan-Khetarpal S, et al. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogenet.* 2008;1:8.
54. Soler-Alfonso C, Carvalho CM, Ge J, Roney EK, Bader PI, Kolodziejska KE, et al. CHRNA7 triplication associated with cognitive impairment and neuropsychiatric phenotypes in a three-generation pedigree. *Eur J Hum Genet.* 2014;22(9):1071–6.
55. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet.* 2008;40(3):322–8.
56. Shchelochkov OA, Li FY, Geraghty MT, Gallagher RC, Van Hove JL, Lichter-Konecki U, et al. High-frequency detection of deletions and variable rearrangements at the ornithine transcarbamylase (OTC) locus by oligonucleotide array CGH. *Mol Genet Metab.* 2009;96(3):97–105.
57. Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes. *Hum Mutat.* 2005;25(5):415–22.
58. Sun C, Skaletsky H, Rozen S, Gromoll J, Nieschlag E, Oates R, et al. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet.* 2000;9(15):2291–6.
59. Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, Rosenbarker L, et al. Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. *Am J Hum Genet.* 2002;71(4):838–53.