



Published in final edited form as:

Hum Genet. 2016 February ; 135(2): 253–256. doi:10.1007/s00439-015-1614-x.

Chimeric transcripts resulting from complex duplications in chromosome Xq28

Luciana W. Zuccherato¹, Benjamin Alleva², Marjorie A. Whitters¹, Claudia M. B. Carvalho^{1,3}, and James R. Lupski^{1,4,5}

¹Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604 BG, Houston, TX 77030, USA

²Biology Department, University of Iowa, Iowa City, IA, USA

³Centro de Pesquisas René Rachou-FIOCRUZ, Belo Horizonte, MG, Brazil

⁴Pediatrics, Baylor College of Medicine, Houston, TX, USA

⁵Texas Children's Hospital, Houston, TX, USA

Abstract

Gene fusions have been observed in somatic alterations in cancer and in schizophrenia. However, the underlying mechanism(s) for their formation are poorly understood. We experimentally demonstrated the expression of splicing variants of in silico predicted chimeric genes *F8/CSAG1* and *BCAP31/TEX28* in two individuals with de novo complex genomic rearrangements of Xq28; *F8/CSAG1* includes exonization of an ERVL-MaLR intronic repetitive element. We provide evidence that replicative repair may contribute to exon shuffling processes and diversify the repertoire of expressed transcripts.

The spawning of new genes constitutes an important substrate for evolutionary diversity between organisms. Gene fusions and exon shuffling are processes in which independent genes or exons can fuse in a single transcriptional unit, resulting in a chimeric protein available to potentially explore new functions and increase phenotypic variation (Long et al. 2003).

The origin of new combinations of genes can be a consequence of the disturbance of the genomic integrity. The DNA replication-based mechanism fork stalling and template switching or microhomology-mediated break-induced replication (FoSTeS/MMBIR) has been proposed as potentially contributing to exon shuffling events (Zhang et al. 2009).

Correspondence to: James R. Lupski.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-015-1614-x) contains supplementary material, which is available to authorized users.

Compliance with ethical standards

Conflict of interest J. R. L. has stock ownership in 23 and Me, is a paid consultant for Regeneron Pharmaceuticals, has stock options in Lasergen Inc., is a member of the Scientific Advisory Board of Baylor Miraca Genetics Laboratories, and a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the chromosomal microarray analysis (CMA) and clinical exome sequencing offered in the Baylor Miraca Genetics Laboratories.

Nevertheless, the assessment of transcriptional consequences of complex genomic rearrangements (CGRs) has been rarely reported, and its role in evolutionary studies of gene diversity remains rudimentary (Inoue et al. 2001). We hypothesize that CGRs play a role in the rapid increase in the diversity of transcripts and proteins, as novel breakpoint junctions and inverted DNA segments can occur in a single rearrangement event.

We now show the presence of novel chimeric genes *BCAP31/TEX28* and *F8/CSAG1* expressed in lymphoblastoid cell lines (LCLs) in two individuals with CGRs in Xq28, likely generated via the replicative repair mechanisms FoSTeS/MMBIR (Fig. 1). From a cohort of 38 unrelated males with *MECP2* duplication syndrome (Carvalho et al. 2009, 2011, 2013), 10 individuals were investigated further based on characteristics of the CGR and breakpoint regions predicted by in silico analyses to result in potential chimeric genes (Table 1). The expression of the chimeric cDNAs was experimentally confirmed in two of seven subjects for whom LCLs were available (Supplementary Methods, Supplementary Fig. 1). The lack of fusion gene transcripts from the other five individuals may be a result of disturbance of the 5' and 3' regulatory elements in the chimeric genes, absence of specific transcripts in LCLs, or RNA expression levels not detected by our assay design.

Analysis of the *BCAP31/TEX28* fusion transcript predicted the formation of a 277-amino acid protein, in which 113 amino acids correspond to the four exons of *BCAP31*, followed by one new amino acid in-frame with 163 amino acids from exons 4 and 5 of *TEX28* (Fig. 1a–d). The newly expressed *F8/CSAG1* transcripts include 239 bp of a partial sequence from an LTR (Long terminal repeat) ERVL-MaLR (Mammalian apparent LTR-retrotransposons) intronic repetitive element, located upstream exon 3 of *CSAG1* in the chimeric gene. Also, 20 bp from an intronic sequence is added between exons 4 and 5 of *CSAG1* (Fig. 1e–h). These data indicate an “exonization” event not previously observed for these genes.

The *F8/CSAG1* chimeric transcripts are predicted by conceptual translation to generate novel proteins composed of eight and 2143 amino acids of the short and long *F8* transcripts, respectively, followed by the insertion of 62 amino acids encoded from the ‘exonized’ ERVL-MaLR intronic repetitive element. Noteworthy, although the exonization of the LTR created a premature stop codon, the chimeric transcripts likely escape the nonsense-mediated decay mechanism surveillance, a quality control of eukaryotic mRNA responsible for inhibiting the production of truncated proteins with deleterious effects (Khajavi et al. 2006).

The fusion genes are predicted to retain important protein domains, including the copper ion binding and oxidoreductase activities in *F8/CSAG1* long transcript, and the B cell receptor-associated protein 31 domain in *BCAP31/TEX28*. No putative known conserved domains or robust protein similarities were detected in the inserted LTR sequence. Additional experiments will be needed to evaluate the impact of these transcripts at the translational level.

Our findings suggest that CGR formed by FoSTeS/MMBIR may contribute significantly to the formation of new genes and proteins during gene and genome evolution (Carvalho et al. 2011; Zhang et al. 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the individuals for their participation in the study. Supported in part by National Human Genome Research Institute/National Heart Lung and Blood Institute (NHGRI/NHLBI) grant to the Baylor Hopkins Center for Mendelian Genomics (U54 HG006542), National Institute of Neurological Disorders and Stroke (NINDS) (R01 NS058529) to J. R. L., and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Young Investigator fellowship from the Science without Borders Program to C. M. B. C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NINDS, NHGRI/NHLBI and NIH.

References

- Carvalho CM, Zhang F, Liu P, Patel A, Sahoo T, Bacino CA, Shaw C, Peacock S, Pursley A, Tavyev YJ, Ramocki MB, Nawara M, Obersztyn E, Vianna-Morgante AM, Stankiewicz P, Zoghbi HY, Cheung SW, Lupski JR. Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum Mol Genet.* 2009; 18:2188–2203. [PubMed: 19324899]
- Carvalho CM, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, McCall A, Pivnick EK, Hines-Dowell S, Seaver LH, Friehling L, Lee S, Smith R, Del Gaudio D, Withers M, Liu P, Cheung SW, Belmont JW, Zoghbi HY, Hastings PJ, Lupski JR. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat Genet.* 2011; 43:1074–1081. [PubMed: 21964572]
- Carvalho CM, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, Belmont JW, Hastings PJ, Lupski JR. Replicative mechanisms for CNV formation are error prone. *Nat Genet.* 2013; 45:1319–1326. [PubMed: 24056715]
- Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, Wyman DW, Lupski JR, Birren B. 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; 11:1018–1033. [PubMed: 11381029]
- Khajavi M, Inoue K, Lupski JR. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet.* 2006; 14:1074–1081. [PubMed: 16757948]
- Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet.* 2003; 4:865–875. [PubMed: 14634634]
- Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet.* 2009; 41:849–853. [PubMed: 19543269]

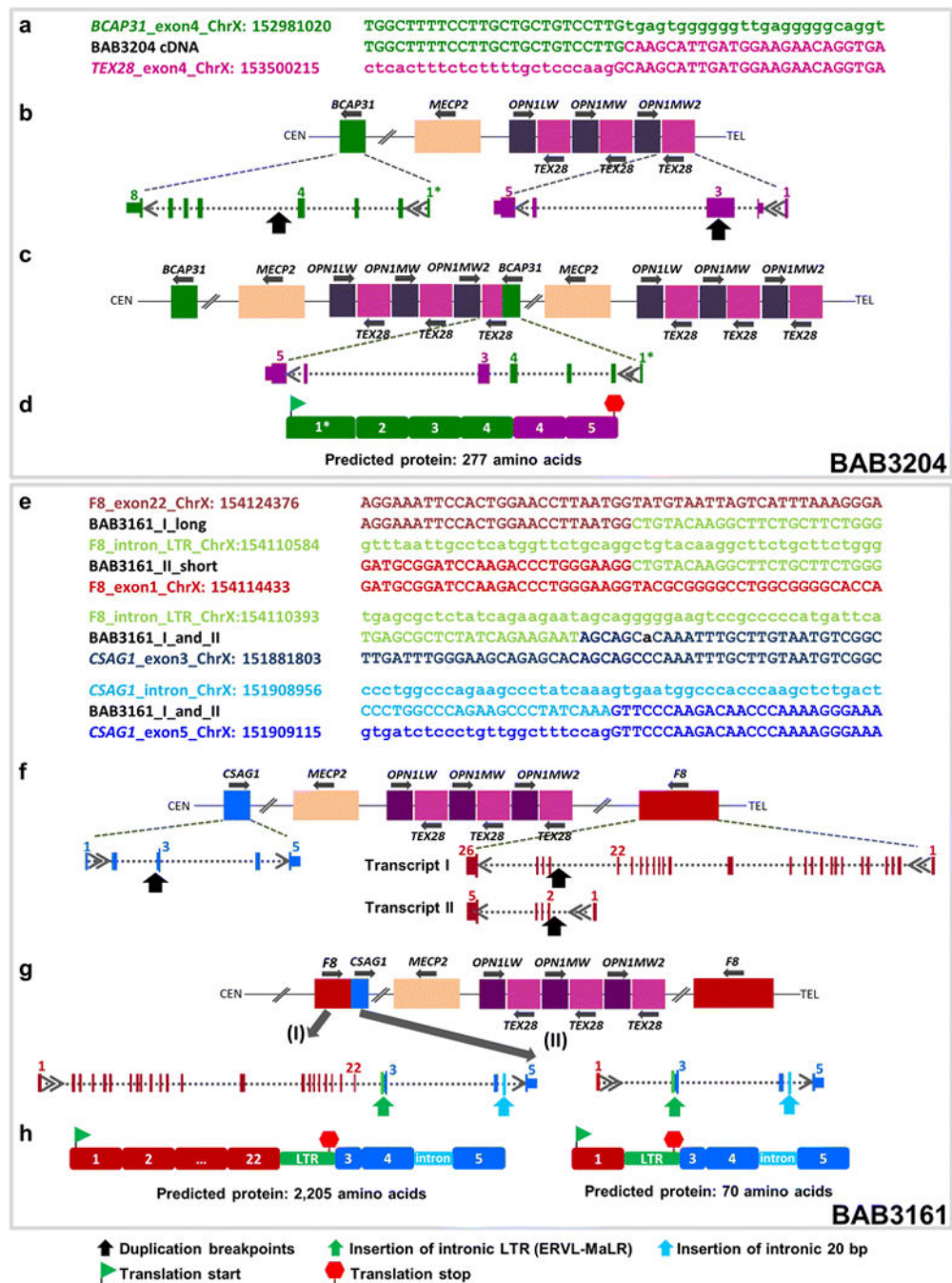


Fig. 1. Formation of fusion genes in patients BAB3204 and BAB3161. *Top* Sanger sequencing of RT-PCR product using individual-specific primers targeting exons that flank genomic breakpoint junction. The cDNA sequences are aligned to the genomic sequences of *BCAP31* and *TEX28* in patient BAB3204 (*a*), and with the genomic sequences of *F8* and *CSAG1* in BAB3161 (*e*). For the genomic sequences, the uppercase letters represent the exons and the lowercase letters, introns. Genomic wild-type structure (*b*, *f*), post-duplication and breakpoint junction structure resulting in novel transcripts (*c*, *g*) and cDNA structures (*d*, *h*).

The latter cDNA structures were obtained by RT-PCR and sequencing of the entire transcripts. In BAB3204, a duplication of ~ 462 Kb encompassing Xq28 region led to the formation of a transcribed fusion gene containing exons 1–4 from the *BCAP31*, plus exons 4 and 5 from *TEX28* (NM_001205201.1) (*c*). The *I** symbol refers to the four splicing variants deposited in the RefSeq database for *BCAP31* (NM_001139457.2, NM_005745.7, NM_001139441.1 and NM_001256447.1). The duplication breakpoint contains part of exon 3 of *TEX28*, although it is not transcribed in the cDNA. In BAB3161, a complex rearrangement constituted by two interspersed duplications of 1.45 Mb and 1.1 Mb (centromeric and telomeric, respectively) or a DUP-NML-DUP (duplication-normal-duplication) led to the formation of chimeric genes in both of the two known alternative splicing transcripts of *F8*, transcripts I (NM_000132.3, 26 exons) and II (NM_019863.2, five exons), fused to exons 3–5 of *CSAG1* (NM_001102576.2) (*g*). Importantly, the genomic rearrangement generated an inversion of the *F8*, leading to a transcript with the same transcriptional orientation as *CSAG1*. Sequencing of the RT-PCR products revealed that a segment of an ERVL-MaLR repetitive element is incorporated in both transcripts. In addition, both chimeric transcripts have 20 bp from an intronic sequence between exons 4 and 5 of *CSAG1* (*g*). The *green* and *red flags* represent computationally predicted start and termination codons in the cDNA sequences (*d, h*)

Table 1

Individuals with potential candidate chimeric events

Sample	Chimeric gene	Detection of fusion gene transcript	References
BAB2623	(<i>TEX28/U52111.14</i>)	No	Carvalho et al. (2009, 2013)
BAB2624	(<i>L1CAM/ZNF185</i>)	No cell line	Carvalho et al. (2009, 2013)
BAB2769	(<i>BGN/PDZD4</i>)	No	Carvalho et al. (2009, 2011)
BAB2772	(<i>TEX28/AVPR2</i>)	No	Carvalho et al. (2009, 2011)
BAB2799	(Opsin/ <i>ABCD1</i>)	No cell line	Carvalho et al. (2009)
BAB2801	(<i>RENBP/SLC6A8</i>)	No	Carvalho et al. (2009, 2011)
BAB3027	(Opsin/ <i>BGN</i>)	No	Carvalho et al. (2013)
BAB3161	(<i>F8/CSAG1</i>)	Yes (2 transcripts)	Carvalho et al. (2013)
BAB3172	(<i>TEX28/PDZD4</i>)	No cell line	Carvalho et al. (2013)
BAB3204	(<i>BCAP31/TEX28</i>)	Yes (4 transcripts)	Carvalho et al. (2013)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript