

Commentary

Formation of a Family of Long Intergenic Noncoding RNA Genes with an Embedded Translocation Breakpoint Motif in Human Chromosomal Low Copy Repeats of 22q11.2—Some Surprises and Questions

Nicholas Delihias

Department of Molecular Genetics and Microbiology, School of Medicine, Stony Brook University, Stony Brook, New York, NY 11794-5222, USA; Nicholas.delihias@stonybrook.edu; Tel.: +1-631-286-9427

Received: 5 July 2018; Accepted: 17 July 2018; Published: 20 July 2018



Abstract: A family of long intergenic noncoding RNA (lincRNA) genes, *FAM230* is formed via gene sequence duplication, specifically in human chromosomal low copy repeats (LCR) or segmental duplications. This is the first group of lincRNA genes known to be formed by segmental duplications and is consistent with current views of evolution and the creation of new genes via DNA low copy repeats. It appears to be an efficient way to form multiple lincRNA genes. But as these genes are in a critical chromosomal region with respect to the incidence of abnormal translocations and resulting genetic abnormalities, the 22q11.2 region, and also carry a translocation breakpoint motif, several intriguing questions arise concerning the presence and function of the translocation breakpoint sequence in RNA genes situated in LCR22s.

Keywords: formation of lincRNA genes; chromosomal low copy repeats; segmental duplications; 22q11.2; translocation breakpoint sequence; directed mutations; palindromic AT-repeats; human satellite 1; HSAT I; formation of RNA exons

As thousands of long noncoding RNA (lncRNA) genes have recently been detected [1–4], one of the interesting problems is their formation. These RNA genes are highly diverse and several different pathways concerning their origins have been outlined [5–12]. Here we analyze and discuss the formation of a family of long intergenic noncoding RNA (lincRNA) genes via gene sequence duplication [13]. We concentrate on duplications that occurred specifically in chromosomal low copy repeats (LCR22) in chromosome 22 (chr22) [14–16] that are in or close to the 22q11.2 chromosomal region. These duplications evolved into eight lincRNA genes that form part of the *FAM230* lincRNA gene family [13] (see also Table S1 for the characteristics of these genes). Although LCR22s provide the means for formation of multiple lincRNA genes, 22q11.2 is a critical chromosomal region, prone to deletions that are mediated by LCR22s and result in genetic disorders such as DiGeorge Syndrome and velo-cardio-facial syndrome [14,15,17]. In addition, the DNA translocation type A breakpoint motif (TBTA), which is directly involved in 22q11.2 deletions [18,19], is incorporated in newly formed lincRNA genes in LCR22s. Significantly, the TBTA sequence present in these genes is modified by highly selective deletion mutations. This offers intriguing questions but presents a possible paradox.

Fifteen lincRNA genes originated by duplication of the sequence of lincRNA gene *FAM230C*, which is situated in chromosome 13 (chr13) [13]. The *FAM230C* gene sequence is the source of formation of two primary groups of genes, where the group category depends on whether the gene originated from the 5' half or 3' half of *FAM230C* sequence. Eight lincRNA genes originated from copies of the 3' end sequence of *FAM230C* and were formed specifically in LCR22s in chr22. These genes include the TBTA motif originally derived from copies of the *FAM230C* gene (Figure 1). Translocation breakpoint

type A motif and its related translocation breakpoint sequences have been shown to undergo DNA strand breakage via palindromic hot spot stem loop structures and cruciforms leading to chromosomal translocation [19–21].

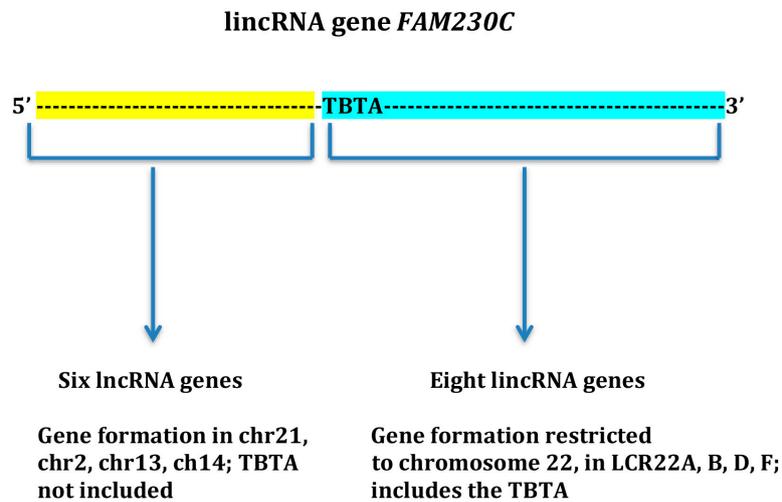


Figure 1. A schematic of 5' and 3' sections of the long intergenic noncoding RNA (lincRNA) gene *FAM230C* present in chr13 that form two distinct groups of long noncoding RNA (lincRNA) genes. Based on reference [13]. Abbreviations: chr, chromosome; TBTA, translocation breakpoint type A; LCR22, low copy repeats in chr22.

An exception is a ninth gene, *AP000552.3* ENSG00000237407, which is a separate type—a small gene that is antisense to *AP000552.1*. ENSG00000206142 (one of the eight genes), and it does not harbor the TBTA. The eight lincRNA genes all display a tissue specificity of RNA transcript expression with major expression only in the testes [13,22]; however, RNA transcript functions are not known.

On the other hand, a heterogeneous group of six genes was formed from the 5' half of the *FAM230C* gene sequence and none contain the TBTA motif. These genes reside in chromosomes other than chr22 (Figure 1). In addition, RNA expression from these genes is varied, with some such as *DUXAP9* that shows RNA expression in multiple tissues (http://useast.ensembl.org/Homo_sapiens/Gene/ExpressionAtlas?db=core;g=ENSG00000225210;r=14:19062316-19131167) [23]. Thus there appear to be two very different categories of genes formed from the lincRNA *FAM230C* gene sequence, whereby cellular regulatory processes determine where the genes are formed and what sequences they contain.

Multiple copies of the *FAM230C* sequence in LCR22s are the result of a large expansion of the sequence involving DNA segmental duplications, with the subsequent formation of the eight lincRNA genes. Although only the 3' half sequence of *FAM230C* is used for gene formation, remnants of the 5' half sequence of *FAM230C* are present in LCR22s and these are not part of the new RNA genes [13].

Thus, LCR22s are a vehicle for creation of multiple lincRNA genes. This is in keeping with the concept that LCRs, or segmental duplications are a major force in human evolution and formation of new genes [24–28]. Genes formed from copies of the 5' half of *FAM230C* (Figure 1) do not appear to involve segmental duplications and for the most part, these are single genes formed in different chromosomes.

Another aspect of this process is more difficult to understand. The *FAM230C* sequence carries the TBTA motif and *FAM230C* sequence duplications spread multiple copies of the TBTA motif in LCR22s. The TBTA motif is sequestered within RNA genes that are formed in LCR22A, B, D and F [13] (Figure 1). These LCR22s are close to each other and less than 10 megabase-pairs apart, in or near the 22q11.2 region [29]. Low copy repeats closer than 10 megabase-pairs are prone to misalignment with resultant chromosomal deletions or duplications [30]. LCR22s are known to participate in meiotic nonallelic homologous recombinations that lead to 22q11.2 deletions and subsequent genetic

diseases [31]. The 22q11.2 region displays the most common chromosomal microdeletion genetic disorder “estimated to result mainly from de novo nonhomologous meiotic recombination events occurring in approximately 1 in every 1000 fetuses” [32]. It is also estimated that ~1 in 3000 to 4000 infants are born with the 22q11.2 deletion [33]. Thus, the 22q11.2 region is associated with a significant incidence of genetic abnormalities that involve participation by LCR22s.

At the molecular level, the TBTA and its related repeat sequences contain palindromic AT-rich repeat sequences (PATRR) that form a very long stem loop (Figure 2). These have loop breakpoint sites directly associated with translocations that can result in genetic disorders involving 22q11.2 [18,19,34,35]. Specifically, PATRR breakpoint sites have been found in LCR22B [18,19,36]. This raises the question of the presence of the TBTA motif in lincRNA genes situated in LCR22s.

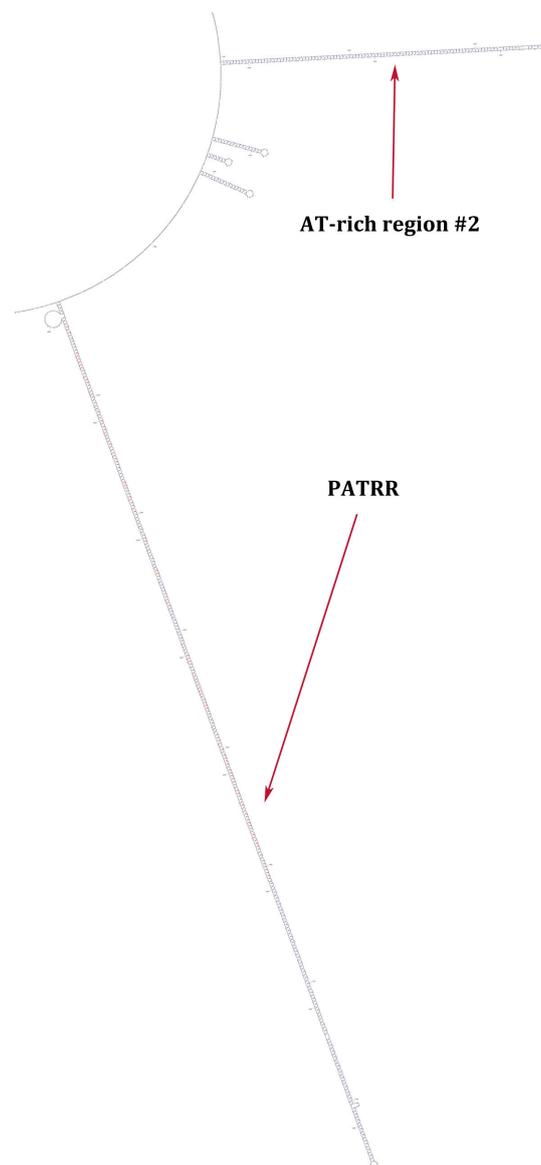


Figure 2. A segment of the secondary structural model of the TBTA (GenBank sequence ID: AB261997.1) showing two long stem loops: the 294 bp palindromic AT-rich repeat sequences (PATRR) and the 104 bp stem loop formed by AT sequences from AT-rich region #2. The PATRR has a significant number of G:C bonds in the lower portion of the stem but is A:T base pair rich in the upper portion close to the loop side. The AT-rich region #2 stem loop consists entirely of A:T bonds with the exception of two G:T bonds. The Zuker DNA folding program [37] was used to generate the secondary structure.

In addition to the PATRR, the TBTA has another section that can form a long stem loop, the AT-rich region #2 (Figure 2). Tong et al. [39] showed that AT-rich region #2 present in a translocation breakpoint element that is related to the TBTA, displays translocation activity, albeit representing a rare translocation event and shows a low frequency of translocation (1.52×10^{-7}) as opposed to the TBTA PATRR (ID: AB261997.1), which has a 10^{-4} – 10^{-5} frequency of translocation [19]. Analysis of TBTA sequences from the eight lincRNA genes shows deletions of the AT-rich region #2, but surprisingly, only in two of the eight lincRNA genes, *LINC01658* and *LINC01662*. As an example, Figure 4 shows the AT-rich region #2 and the PATRR-associated AT-rich sequences totally deleted in *LINC01658*. The sequences between the arrows in Figure 4 denote deleted areas. In addition, a complete elimination of AT-rich sequences occurred in this gene with an additional deletion, that of the smaller AT-rich region #1, with the exception of positions 360–366 (Figure 4). Essentially, *LINC01658* is devoid of AT-rich sequences. In contrast, an example of the presence of the entire AT-rich #2 motif is in *LINC01663*, one of the six RNA genes that have the AT-rich region #2 (Figure S1). As expected, the AT sequences are highly variable. There is also a very large number of AT bases present in *LINC01663* relative to the TBTA AT-rich region (Figure S1), indicating a robust expansion of AT sequences in this gene. This is in sharp contrast to *LINC01658* that is devoid of AT-rich sequences.

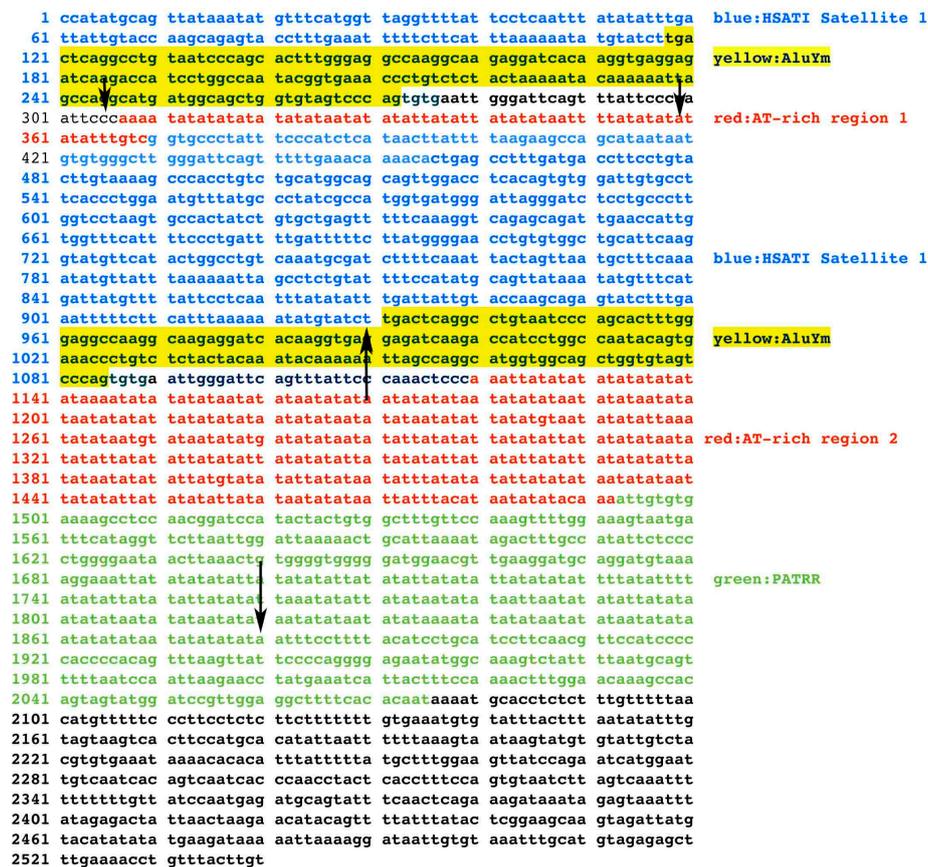


Figure 4. The TBTA sequence, highlighting the various elements it harbors, including the PATRR and AT-rich region #2 and AT-rich region #1. The sequence is from National Center of Biotechnology information (NCBI) GenBank: AB261997.1. The figure is modified from reference [13]. Two arrows (bottom) delineate the regions deleted in the RNA gene *LINC01658* sequence that is equivalent to TBTA positions 930–1880, and encompasses the AluYm transposable element (in yellow), the entire AT-rich region #2 (in red), the 5' side of the PATRR and includes all of the PATRR-associated AT-rich sequences up to position 1880 (in green). The AT-rich region #1 is also deleted (delineated by arrows at positions 305 and 359) with the exception of the very small AT sequence, positions 360–366. These deletions essentially result in RNA gene *LINC01658* having no AT-rich sequences.

This seemingly is paradoxical, as one would expect AT-rich #2 sequences to be deleted in all lincRNA genes to eliminate the possibility of a translocation breakpoint sequence evolving in RNA genes present in LCR22s. AT-rich sequences are highly unstable, undergo extensive point mutations, insertions, deletions, and readily form long stem-loop secondary structures. For example, random DNA sequences with 500 bases containing 95% A + T can generate 50–100 base pair stems of a stem loop structure. None of the lincRNA genes display very long, perfect AT stem loops, but one cannot rule out potential breakpoint sequences evolving from AT-rich variable sequences in one or more lincRNA genes. Does the cell tolerate a probability of a rare translocation event occurring within RNA genes in LCR22s that carry AT-rich #2 sequences? This is in a background of the 22q11.2 region that already is highly problematic in terms of incidence of genetic disease stemming from abnormal translocations.

Why does the *FAM230C* gene, which is in chr13, harbor the TBTA? And why is the translocation breakpoint motif specifically added to newly formed RNA genes in LCR22s? We do not have enough information to answer or comment on the first question, but in terms of the second, the TBTA carries the human satellite 1, HSAT I (Figure 4) [13,21,40]. Segments of the HSAT I sequence form the entire exon1 of several annotated RNA transcripts from lincRNA genes [41] (see Figure S2). Thus, the TBTA sequence helps form lincRNA gene structure by carrying the HSAT I satellite, and this demonstrates a role for a satellite sequence in development of lincRNA gene and transcript exon sequence. This may be a secondary or separate role of the TBTA and does not address the addition of the motif to the RNA genes that are specifically in LCR22s. Protein genes are known to carry breakpoint sequences; about 2000 genes have been detected that harbor purine/pyrimidine tracts that form long stem loops [42]. Perhaps genes are storage and protection sites for these elements.

The TBTA and its related motifs are present in nonhuman primates [21,43]. In addition, the 3' half of the *FAM230C* sequence is found in chr22 of the chimpanzee genome with a sequence identity of 97% compared with the human *FAM230C* [41] (Figure S3). It appears to be an ancient and highly conserved sequence present in a common ancestor of humans and chimpanzees. However, neither the entire *FAM230C* sequence or the eight related gene sequences have been detected as complete sequences, or as yet have been annotated in the chimpanzee or other primates. There must be more complete genomic sequences from the chimpanzee and other primates, as well as better lincRNA gene annotations to determine if these genes are present in the chimpanzee or other primate genomes, or if they are specific to humans.

We also do not know the evolutionary relationship between *FAM230C* and the putative protein gene *FAM230A* (Ensembl: ENSG00000277870) in human chr22. The *FAM230A* gene is only partially defined as there is a remaining 50,000 bp unsequenced gap in the central portion of the gene.

It is surprising that deletion mutations eliminated the PATRR structure from the eight lincRNA genes in LCR22s, yet six genes are left with AT-rich sequences, some having an excess of AT residues that we hypothesize may potentially evolve into breakpoint structures. 22q11.2 is a complex region. We have some, but not a full understanding of the relationship of these lincRNA genes to the region, to the translocation breakpoint motif, and of RNA transcript function. What does appear obvious is the high specificity in types of mutations that occurred in the RNA genes, the regulation in chromosomal placement of genes, the mechanism of gene formation, and the importance of a DNA satellite in RNA exon sequence formation. However, to our knowledge, the eight genes of the *FAM230* lincRNA family are the first lincRNA genes known to be formed by DNA segmental duplications. As segmental duplications are considered a major factor in human evolution and creation of new genes [24–28], there may be other lincRNA gene families formed by this pathway.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2311-553X/4/3/16/s1>. Table S1: Properties of lincRNA genes in chr22 (from reference [13]), Figure S1: Nucleotide sequence alignment of the TBTA AT-rich region #2 with the AT-rich region of *LINC01663*, Figure S2: Nucleotide sequence alignment of the exon 1 sequence from RNA transcript *LINC01660-203* with the TBTA, Figure S3: Nucleotide sequence alignment of the RNA gene *FAM230C* with a segment of chr22 from the chimpanzee genome.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)] [[PubMed](#)]
2. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [[CrossRef](#)] [[PubMed](#)]
3. Goff, L.A.; Rinn, J.L. Linking RNA biology to lncRNAs. *Genome Res.* **2015**, *25*, 1456–1465. [[CrossRef](#)] [[PubMed](#)]
4. Morris, K.V.; Mattick, J.S. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15*, 423–437. [[CrossRef](#)] [[PubMed](#)]
5. Kapusta, A.; Feschotte, C. Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends Genet.* **2014**, *30*, 439–452. [[CrossRef](#)] [[PubMed](#)]
6. Guo, X.; Lin, M.; Rockowitz, S.; Lachman, H.M.; Zheng, D. Characterization of human pseudogene-derived non-coding RNAs for functional potential. *PLoS ONE* **2014**, *9*, e93972. [[CrossRef](#)] [[PubMed](#)]
7. Ulitsky, I. Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **2016**, *17*, 601–614. [[CrossRef](#)] [[PubMed](#)]
8. Liu, W.H.; Tsai, Z.T.; Tsai, H.K. Comparative genomic analyses highlight the contribution of pseudogenized protein-coding genes to human lincRNAs. *BMC Genom.* **2017**, *18*, 786. [[CrossRef](#)] [[PubMed](#)]
9. Terracciano, D.; Terreri, S.; de Nigris, F.; Costa, V.; Calin, G.A.; Cimmino, A. The role of a new class of long noncoding RNAs transcribed from ultraconserved regions in cancer. *Biochim. Biophys. Acta* **2017**, *1868*, 449–455. [[CrossRef](#)] [[PubMed](#)]
10. Espinosa, J.M. On the Origin of lncRNAs: Missing Link Found. *Trends Genet.* **2017**, *33*, 660–662. [[CrossRef](#)] [[PubMed](#)]
11. Wu, H.; Yang, L.; Chen, L.L. The Diversity of Long Noncoding RNAs and Their Generation. *Trends Genet.* **2017**, *33*, 540–552. [[CrossRef](#)] [[PubMed](#)]
12. Awan, H.M.; Shah, A.; Rashid, F.; Shan, G. Primate-specific Long Non-coding RNAs and MicroRNAs. *Genom. Proteom. Bioinform.* **2017**, *15*, 187–195. [[CrossRef](#)] [[PubMed](#)]
13. Delihias, N. A family of long intergenic non-coding RNA genes in human chromosomal region 22q11.2 carry a DNA translocation breakpoint/AT-rich sequence. *PLoS ONE* **2018**, *13*, e0195702. [[CrossRef](#)] [[PubMed](#)]
14. Edelmann, L.; Pandita, R.K.; Morrow, B.E. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am. J. Hum. Genet.* **1999**, *64*, 1076–1086. [[CrossRef](#)] [[PubMed](#)]
15. Shaikh, T.H.; Kurahashi, H.; Saitta, S.C.; O'Hare, A.M.; Hu, P.; Roe, B.A.; Driscoll, D.A.; McDonald-McGinn, D.M.; Zackai, E.H.; Budarf, M.L.; 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*, 489–501. [[CrossRef](#)] [[PubMed](#)]
16. Eichler, E.E. Masquerading repeats: Paralogous pitfalls of the Human Genome. *Genome Res.* **1998**, *8*, 758–762. [[CrossRef](#)] [[PubMed](#)]
17. Stankiewicz, P.; Lupski, J.R. Molecular-evolutionary mechanisms for genomic disorders. *Curr. Opin. Genet. Dev.* **2002**, *12*, 312–319. [[CrossRef](#)]
18. Edelmann, L.; Spiteri, E.; Koren, K.; Pulijaal, V.; Bialer, M.G.; Shanske, A.; Goldberg, R.; Morrow, B.E. AT-rich palindromes mediate the constitutional t(11;22) translocation. *Am. J. Hum. Genet.* **2001**, *68*, 1–13. [[CrossRef](#)] [[PubMed](#)]
19. Kato, T.; Kurahashi, H.; Emanuel, B.S. Chromosomal translocations and palindromic AT-rich repeats. *Curr. Opin. Genet. Dev.* **2012**, *22*, 221–228. [[CrossRef](#)] [[PubMed](#)]
20. Kurahashi, H.; Inagaki, H.; Hosoba, E.; Kato, T.; Ohye, T.; Kogo, H.; Emanuel, B.S. Molecular cloning of a translocation breakpoint hotspot in 22q11. *Genome Res.* **2007**, *17*, 461–469. [[CrossRef](#)] [[PubMed](#)]
21. Babcock, M.; Yatsenko, S.; Stankiewicz, P.; Lupski, J.R.; Morrow, B.E. AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. *Genome Res.* **2007**, *17*, 451–460. [[CrossRef](#)] [[PubMed](#)]

22. Fagerberg, L.; Hallström, B.M.; Oksvold, P.; Kampf, C.; Djureinovic, D.; Odeberg, J.; Habuka, M.; Tahmasebpour, S.; Danielsson, A.; Edlund, K.; et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteom.* **2014**, *13*, 397–406. [[CrossRef](#)] [[PubMed](#)]
23. Zerbino, D.R.; Achuthan, P.; Akanni, W.; Amode, M.R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Girón, C.G.; et al. Ensembl 2018. *Nucleic Acids Res.* **2017**, *46*, D754–D761. [[CrossRef](#)] [[PubMed](#)]
24. Samonte, R.V.; Eichler, E.E. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **2002**, *3*, 65–72. [[CrossRef](#)] [[PubMed](#)]
25. Stankiewicz, P.; Shaw, C.J.; Withers, M.; Inoue, K.; Lupski, J.R. Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res.* **2004**, *14*, 2209–2220. [[CrossRef](#)] [[PubMed](#)]
26. Dennehey, B.K.; Gutches, D.G.; McConkey, E.H.; Krauter, K.S. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **2004**, *83*, 493–501. [[CrossRef](#)] [[PubMed](#)]
27. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **2013**, *92*, 155–161. [[CrossRef](#)] [[PubMed](#)]
28. Dennis, M.Y.; Eichler, E.E. Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* **2016**, *41*, 44–52. [[CrossRef](#)] [[PubMed](#)]
29. Guo, X.; Freyer, L.; Morrow, B.; Zheng, D. Characterization of the past and current duplication activities in the human 22q11.2 region. *BMC Genom.* **2011**, *12*, 71. [[CrossRef](#)] [[PubMed](#)]
30. Harel T Pehlivan, D.; Caskey, C.T.; Lupski, J.R. Mendelian, Non-Mendelian, Multigenic Inheritance, and Epigenetics. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease*, 5th ed.; Academic Press: Waltham, MA, USA, 2015; Chapter 1, pp. 3–27.
31. Shaikh, T.H.; Kurahashi, H.; Emanuel, B.S. Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: An update and literature review. *Genet. Med.* **2001**, *3*, 6–13. [[CrossRef](#)] [[PubMed](#)]
32. McDonald-McGinn, D.M.; Sullivan, K.E.; Marino, B.; Philip, N.; Swillen, A.; Vorstman, J.A.; Zackai, E.H.; Emanuel, B.S.; Vermeesch, J.R.; Morrow, B.E.; et al. 22q11.2 deletion syndrome. *Nat. Rev. Dis. Primers* **2015**, *1*, 15071. [[CrossRef](#)] [[PubMed](#)]
33. Kobrynski, L.J.; Sullivan, K.E. Velocardiofacial syndrome, DiGeorge syndrome: The chromosome 22q11.2 deletion syndromes. *Lancet* **2007**, *370*, 1443–1452. [[CrossRef](#)]
34. Kurahashi, H.; Inagaki, H.; Ohye, T.; Kogo, H.; Kato, T.; Emanuel, B.S. Chromosomal translocations mediated by palindromic DNA. *Cell Cycle* **2006**, *5*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
35. Inagaki, H.; Kato, T.; Tsutsumi, M.; Ouchi, Y.; Ohye, T.; Kurahashi, H. Palindrome-Mediated Translocations in Humans: A New Mechanistic Model for Gross Chromosomal Rearrangements. *Front. Genet.* **2016**, *7*, 125. [[CrossRef](#)] [[PubMed](#)]
36. Gotter, A.L.; Shaikh, T.H.; Budarf, M.L.; Rhodes, C.H.; Emanuel, B.S. A palindrome-mediated mechanism distinguishes translocations involving LCR-B of chromosome 22q11.2. *Hum. Mol. Genet.* **2004**, *13*, 103–115. [[CrossRef](#)] [[PubMed](#)]
37. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
38. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [[CrossRef](#)]
39. Tong, M.; Kato, T.; Yamada, K.; Inagaki, H.; Kogo, H.; Ohye, T.; Tsutsumi, M.; Wang, J.; Emanuel, B.S.; Kurahashi, H. Polymorphisms of the 22q11.2 breakpoint region influence the frequency of de novo constitutional t(11;22)s in sperm. *Hum. Mol. Genet.* **2010**, *19*, 2630–2637. [[CrossRef](#)] [[PubMed](#)]
40. Frommer, M.; Prosser, J.; Vincent, P.C. Human satellite I sequences include a male specific 2.47 kb tandemly repeated unit containing one Alu family member per repeat. *Nucleic Acids Res.* **1984**, *12*, 2887–2900. [[CrossRef](#)] [[PubMed](#)]
41. Delihias, N. Complexity of a small non-protein coding sequence in chromosomal region 22q11.2: Presence of specialized DNA secondary structures and RNA exon/intron motifs. *BMC Genom* **2015**, *16*, 785. [[CrossRef](#)] [[PubMed](#)]

42. Bacolla, A.; Collins, J.R.; Gold, B.; Chuzhanova, N.; Yi, M.; Stephens, R.M.; Stefanov, S.; Olsh, A.; Jakupciak, J.P.; Dean, M.; et al. Long homopurine·homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res.* **2006**, *34*, 2663–2675. [[CrossRef](#)] [[PubMed](#)]
43. Inagaki, H.; Ohye, T.; Kogo, H.; Yamada, K.; Kowa, H.; Shaikh, T.H.; Emanuel, B.S. Kurahashi Palindromic AT-rich repeat in the *NF1* gene is hypervariable in humans and evolutionarily conserved in primates. *Hum. Mutat.* **2005**, *26*, 332–342. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).