

Review

The 22q11.2 Low Copy Repeats

Lisanne Vervoort and Joris Robert Vermeesch * 

Center for Human Genetics, KU Leuven, 3000 Leuven, Belgium

* Correspondence: joris.vermeesch@kuleuven.be

Abstract: LCR22s are among the most complex loci in the human genome and are susceptible to nonallelic homologous recombination. This can lead to a variety of genomic disorders, including deletions, duplications, and translocations, of which the 22q11.2 deletion syndrome is the most common in humans. Interrogating these phenomena is difficult due to the high complexity of the LCR22s and the inaccurate representation of the LCRs across different reference genomes. Optical mapping techniques, which provide long-range chromosomal maps, could be used to unravel the complex duplication structure. These techniques have already uncovered the hypervariability of the LCR22-A haplotype in the human population. Although optical LCR22 mapping is a major step forward, long-read sequencing approaches will be essential to reach nucleotide resolution of the LCR22s and map the crossover sites. Accurate maps and sequences are needed to pinpoint potential predisposing alleles and, most importantly, allow for genotype–phenotype studies exploring the role of the LCR22s in health and disease. In addition, this research might provide a paradigm for the study of other rare genomic disorders.

Keywords: low copy repeats; genomic disorders; 22q11.2 deletion syndrome



Citation: Vervoort, L.; Vermeesch, J.R. The 22q11.2 Low Copy Repeats. *Genes* **2022**, *13*, 2101. <https://doi.org/10.3390/genes13112101>

Academic Editors: Beata Nowakowska and Donna M. McDonald-McGinn

Received: 23 September 2022

Accepted: 25 October 2022

Published: 11 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The acrocentric chromosome 22 is the second shortest human autosome with a length of 50 Mb of which 15 Mb belong to the p-arm. Together with the Y-chromosome, this chromosome has the highest density of low copy repeats (LCRs). LCRs are larger than 1 kb blocks of DNA that are duplicated across several inter- and intrachromosomal loci in the genome [1,2]. Due to their high sequence identity, they are subject to meiotic misalignment and recombination, a mechanism known as nonallelic homologous recombination (NAHR). Rearrangements between the LCRs on chromosome 22 (LCR22s) cause the 22q11.2 deletion syndrome (22q11.2DS), the most common genomic disorder in humans [3].

LCRs constitute 6.6% of the human genome [4], and several genes within the LCRs have played an important role in human adaptation [5]. LCRs are frequently removed in standard analysis pipelines. Due to the multiple mapping options, reads originating from LCRs are frequently wrongly assigned, creating errors and gaps in reference assemblies.

The study of the 22q11.2 deletion syndrome is limited by the lack of a proper genomic reference sequence in the 22q11.2 LCRs. The complexity of the LCR22s is characterized by a patchwork of subunits, present in multiple and variable copy numbers that share a high percentage of sequence identity. Consequently, genome sequences cannot be assembled properly, and the reference genomes have an inaccurate representation of the LCR22s.

2. Instability of the LCR22s in the Human Genome

The 22q11.2 locus is considered one of the most unstable loci in the human genome due to the presence of a cluster of eight LCR22s. They are commonly denoted as LCR22-A to -H and act as ideal substrates for nonallelic homologous recombination (NAHR), resulting in several chromosomal conditions (Figure 1).

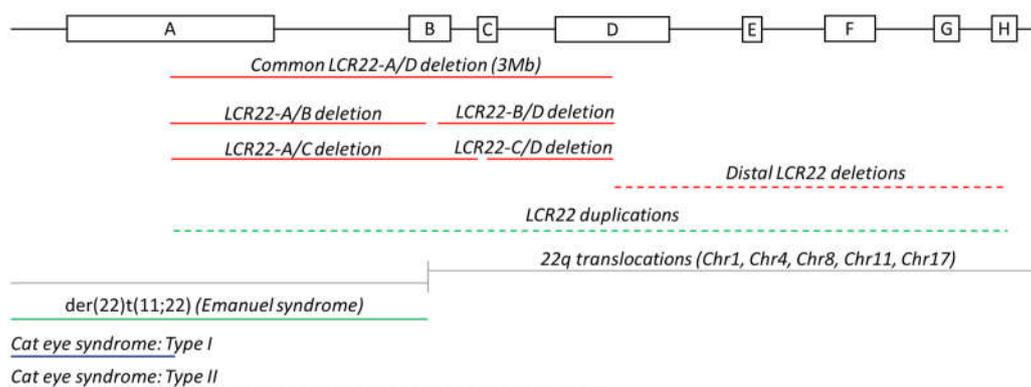


Figure 1. Rearrangements between low copy repeats on chromosome 22. Organization of LCR22s-A until -H. Schematic overview of the rearrangements observed between LCR22s with dotted lines indicating the options of crossovers between different blocks. The red lines indicate deletions. The green lines represent duplications. The grey lines show the rearrangement locus in the recurrent translocations involving chromosome 22. The dark blue lines show the triplication of a fragment in cat eye syndrome. The dashed lines cover different-sized CNVs.

2.1. Deletions and Duplications

Rearrangements among the four proximal LCR22s, -A, -B, -C, and -D, cause the 22q11.2DS and the reciprocal duplication syndrome [6] (Figure 1). Medical manifestations of the 22q11.2DS include congenital heart defects, immunodeficiency, and palatal anomalies. However, the syndrome is characterized by a high interpatient phenotypic variability. In 85% of individuals with the 22q11.2DS, NAHR occurs between LCR22-A and -D, the two largest LCR22s, resulting in a 3 Mb deletion (Figure 1). This region contains 46 protein-coding genes and less well characterized transcripts within the LCR22s. In 90–95% of patients, the deletion is de novo [6,7].

Nested deletions are defined as deletions where NAHR involves either LCR22-B or -C [6]. Of 22q11.2DS patients, 5% carry the 1.5 Mb LCR22-A/B deletion. The 2 Mb LCR22-A/C deletion, the 1.5 Mb LCR22-B/D deletion, and the 1 Mb LCR22-C/D deletion account for 2%, 4%, and 1% of patients, respectively (Figure 1) [7]. Given that these deletions cause a milder phenotype, they are more frequently inherited. Deletions encompassing the distal LCR22s (LCR22-E until -H) are less frequent, and the phenotypes differ from the traditional 22q11.2DS [8]. These deletions are referred to as the distal 22q11.2 deletion syndrome (MIM: 611867) [9]. Rearrangements where one of the breakpoints is located in the unique sequence between the LCR22s have been observed as well and are termed atypical deletions [8]. The prevalence of these nonstandard (nested, distal, and atypical) deletions is likely underestimated since they are associated with milder phenotypes and are, therefore, harder to diagnose. In addition, not all genetic tests are able to detect these uncommon deletions (for example, standard 22q11.2del FISH using probes TUPLE or N25 are located in the unique sequence between LCR22-A and -B).

Reciprocal to the standard 22q11.2DS, the 22q11.2 duplication syndrome (MIM: 608363) has been described [10]. Symptoms include velopharyngeal insufficiency, behavioral deficits, and dysmorphic features, and overlap with the 22q11.2DS.

2.2. Translocations

LCR22s are essential for the creation of several nonrandom constitutional translocations. Recurrent translocations include t(8;22)(q24.1;q11.2), t(11;22)(q23;q11.2), and t(17;22)(q11.2;q11.2), whereas t(1;22)(p21.1;q11.2) and t(4;22)(q35.1;q11.2) belong to the nonrecurrent class [11]. Translocations can directly affect the function of a gene; for example, t(17;22) disrupts *NF1* on chromosome 17, leading to neurofibromatosis type 1 [12].

The best known non-Robertsonian translocation is t(11;22)(q23;q11.2). Since there is no gain or loss of DNA, carriers are phenotypically normal, except for reproductive prob-

lems [13]. They can have phenotypically normal offspring when the child inherits the two normal chromosomes or the two translocated chromosomes, as a result of balanced 2:2 segregation during meiosis. Gametes and embryos following unbalanced 2:2 segregation can be conceived as well. Such conceptions lead to spontaneous miscarriage, explaining the high rates among carrier parents [14]. However, the offspring can inherit the derivative chromosome from the translocation-carrying parent in addition to the normal nontranslocation partner chromosomes. This mechanism is known as 3:1 meiotic malsegregation [13] and can lead to clinical symptoms since the copy number is changed. For example, carriers of the t(11;22)(q23;q11.2) translocation have a 10% risk of having a child with the supernumerary der(22)t(11;22) syndrome. This syndrome, also known as Emanuel syndrome (MIM: 609029), is caused by the presence of a partial trisomy of the 11q23 and the 22q11 locus. Clinical features of the syndrome include ear pits, micrognathia, heart malformations, and cleft palate [15]. The breakpoints of these constitutional 22q11.2 translocations all cluster in LCR22-B (Figure 1), described by Emanuel in 2008 as the most rearrangement-prone site of the human genome [16].

Several nonconstitutional translocations involving 22q were said to play a role in carcinogenesis as well. The chromosomal abnormality resulting from a translocation between chromosome 9q34 (ABL locus) and chromosome 22q11 (BCR locus) is called the Philadelphia chromosome. The *BCR-ABL1* fusion gene encodes a hybrid protein important in the disease pathway of several forms of leukemia [17]. In addition, the t(8;22) translocation is a variant of the t(8;14) translocation associated with Burkitt's lymphoma [18], and Ewing sarcoma (MIM: 612219) is caused by the t(11;22)(q24;q12) translocation fusing the *FLI1* and *EWS* genes on chromosomes 11 and 22, respectively [19]. These oncogenic translocations arise via mitotic rearrangement and are only present in the tumor or affected tissue.

2.3. Cat Eye Syndrome

A more complex constitutional LCR22-mediated rare disorder is cat eye syndrome (MIM: 115470), characterized by iris coloboma, anal atresia, and ear malformations [20]. The karyotype marks the presence of a small supernumerary dicentric, bisatellited chromosome representing an inv dup(22) (q11.2) [20,21]. This structure leads to a tetrasomy of the chromosome 22 p-arm and part of the 22q11.2 locus (from the centromere to the distal breakpoint). The breakpoints are mapped to LCR22-A and/or -D, and different cytogenetic cat eye syndrome types exist based on the specific LCR22 involvement (Figure 1) [20]: type I and IIb marker chromosomes are symmetrical with both breakpoints clustering in LCR22-A or -D, respectively. Asymmetric cat eye syndrome marker chromosomes characterized by one LCR22-A and one LCR22-D breakpoint are classified as type IIa. The supernumerary chromosome is predicted to be created via interchromosomal misalignment or intrachromosomal recombination, leading to an inversion with subsequent homologue crossover [21].

Hence, the presence of LCR sequence on chromosome 22 makes the locus prone to several types of genomic rearrangements, leading to disease and disease predisposition.

3. LCR22 Assembly and Representation in the Reference Genome

In 1999, chromosome 22 was the first finished human chromosome sequence [22]. This sequence covered the euchromatic part and consisted of 12 contigs spanning 33.4 Mb. Four of the 11 gaps were located in the LCR22-associated 22q11.2 locus [22]. In 2008, 8 of the 11 gaps were closed by combining different sequencing techniques [23]. However, the 22q11.2 locus still harbored two of the remaining gaps. The inability to differentiate between two alleles on homologous chromosomes or two (or more) paralogues on the same chromosome (or a combination) is the main cause for these failures [1]. The length of the LCR22 modules surpasses the length of second generation sequencing reads, and misalignments can lead to false structural variant calls.

As a consequence, comparisons between the different reference genomes showed several inconsistencies at the level of the LCR22s. A first organizational overview of

the LCR22 structure was provided even before the release of the first human reference genome assembly [24] (Figure 2). Based on sequencing data from a bacterial artificial chromosome (BAC), a P1-derived artificial chromosome, and cosmid insert clones, four LCR22s were differentiated, each represented as a mosaic patchwork of different modules. These modules are shared between one or more LCR22s via duplication and, therefore, have high sequence similarity (97–98%). For the two largest blocks, LCR22-A and -D, the lengths were estimated at 350 and 250 kb, respectively. They share long stretches of identical duplicon composition, except for the LCR22-A-specific proximal USP18 (yellow) and distal DGCR6/PRODH module (red) and the LCR22-D-specific PI4KA module (mustard) (Figure 2). The smaller LCR22-B and -C blocks are composed of modules that are present in LCR22-A and/or LCR22-D, but they do not share similar blocks between them, suggesting that rearrangements between the smaller LCR22s are not possible. Hence, the duplicated module structure of the LCR22s provides ideal elements for NAHR.

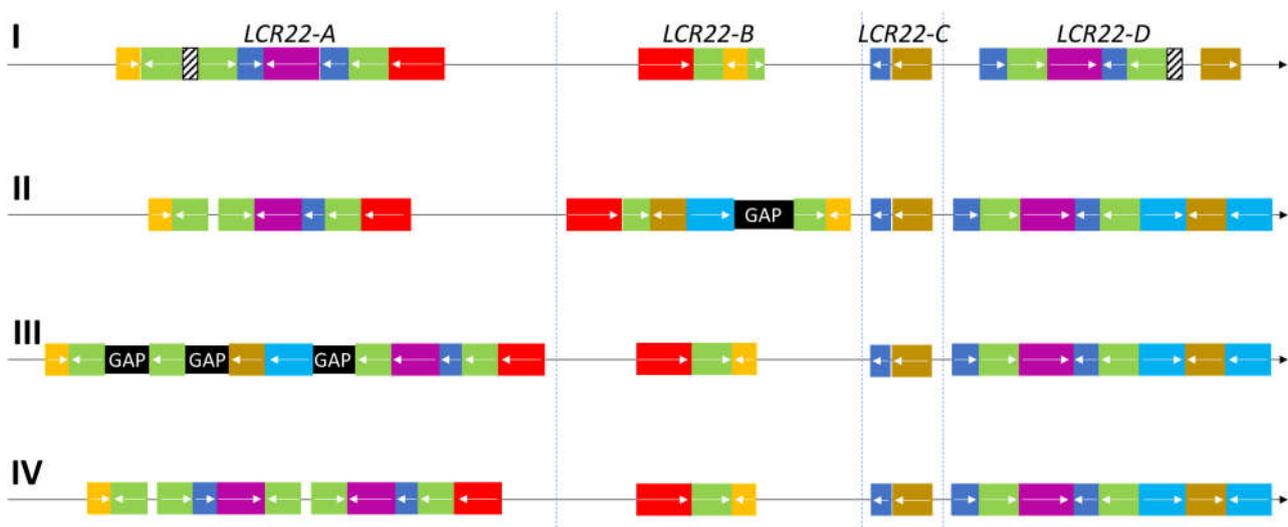


Figure 2. Organization of LCR22-duplicated modules in several references. (I) First description of the LCR22 structure in Shaikh et al. (2000). These modules serve as ‘essential duplicons’ in the comparison and composition of the LCR22s in the other reference genomes (II–IV). The yellow duplicon contains USP18 (or a paralogue). The red duplicon encompasses (paralogues of) the genes PRODH and DGCR6. The PI4KA gene or a paralogue is present in the mustard-colored blocks. The cyan duplicon always covers a RIMBP3 paralogue, the blue duplicon a BCR paralogue, and the magenta duplicon a GGT paralogue. A FAM230 paralogue can be found in the green duplicon, as well as a palindromic AT-rich repeat. (II) LCR22 organization in GRCh37/hg19. (III) LCR22 organization in GRCh38/hg38. (IV) LCR22 organization, for the first time without gaps, in the most recent reference genome (T2T-CHM13).

Despite the successful structural representation of LCR22 by Shaikh et al. (2000), difficulties were encountered when assembling the LCR22s during the generation of the human reference genome [25]. In hg19 (GRCh37, released February 2009), the LCR22s were delineated (Table 1) with a gap of 100 kb present in LCR22-B (Figure 2). Surprisingly, LCR22-A was identified as the second smallest LCR22. In the reference genome GRCh38, released in December 2013, relative LCR22 sizes were comparable to the first representation [24] with major changes compared with hg19 (Figure 2, Table 1). First, a 100 kb module, including the genes RIMBP3 and TMEM191B and pseudogene PI4KAP1, was shifted from LCR22-B to LCR22-A. Second, no gaps were present anymore in LCR22-B, but three new gaps appeared in LCR22-A with lengths of 100 (most proximal), 50, and 50 kb (most distal). Due to the variability and inconsistencies between the different genome assemblies, constructing the exact composition and sequence of the LCR22s is still an ongoing challenge.

Table 1. Exact chromosomal locations of the different LCR22s in hg19, hg38, and telomere-to-telomere reference genomes.

	GRCh37/hg19	GRCh38/hg38	T2T-CHM13
LCR22-A	chr22:18,639,043-19,022,986	chr22:18,156,276-19,035,473	CP068256.2:18,828,186-19,410,796
LCR22-B	chr22:20,128,537-20,731,921	chr22:20,141,014-20,377,631	CP068256.2:20,520,047-20,781,953
LCR22-C	chr22:21,021,564-21,092,560	chr22:20,667,276-20,738,272	CP068256.2:21,075,991-21,146,982
LCR22-D	chr22:21,363,668-21,916,380	chr22:21,009,379-21,562,091	CP068256.2:21,418,153-21,975,566

Despite their importance in disease predisposition, the extreme duplication nature of the LCR22s causes difficulties in standard mapping pipelines, assembly errors, and incomplete representation in the reference genomes. Global structural information is missing in the short reads and standard long reads, hindering a correct assembly of this structurally complex locus. There is a need for *de novo*, nonreference based approaches to correctly compile the LCR22 haplotype.

4. Entering the Era of Optical Mapping

Long-range information over the repeat is necessary to decipher the exact structural composition of the LCR22s. Applications to achieve this long-range structural conformation include optical mapping techniques, such as fiber-FISH (fluorescent in situ hybridization) and Bionano optical mapping. These tools enable the *de novo* assembly of structurally complex loci with a resolution of 5–10 kb [26]. In Bionano optical mapping, fluorescently labeled DNA fibers are scanned, and the *de novo* assembly is based on the labeling pattern of the reads (Figure 3). Due to the lengths of these fibers ($N_{50} > 150$ kb in general), large deletions, duplications, inversions, and translocations can be visualized. Using this technique, LCR haplotypes and additional structural variants were discovered in the 7q11.23, 15q13.3, and 16p12.2 locus [27]. The method is not targeted, and therefore, structural variants can be inferred at a whole-genome scale [28]. In fiber-FISH, one specific locus of the genome is scrutinized using fluorescently labeled probes targeting the region of interest (Figure 3). In brief, long DNA molecules (>300 kb) are stretched onto a glass surface and hybridized with probes [29]. This approach has proven to be successful for structural variation detection [30,31].

Optical mapping techniques were able to elucidate the LCR22 architecture for the first time [32]. To the same end, fiber-FISH was performed on the haploid cell line CHM1 and the diploid cell line GM12878. LCR22 haplotypes were *de novo* assembled by tiling fibers based on matching colors and distances between the probes [32]. The smaller LCR22s-B and -C were identical to the composition observed in the reference genome hg38. In both cell lines, the three haplotypes of LCR22-D were identical to each other, but differed from the hg38 reference by an inversion distal in LCR22-D. Three different haplotypes were compiled for LCR22-A, with lengths between 0.65 and 1.20 Mb and substantially different from the reference genome hg38. Bionano optical mapping was performed on these two cell lines as an orthogonal method to validate the results [32].

To evaluate whether the variation between these three different LCR22-A haplotypes in the three investigated alleles was a coincidence, a population-scale study was performed in 187 individuals and uncovered the presence of 6 LCR22-D variants and 26 complete LCR22-A configurations [32]. The LCR22 haplotyping of 57 diploid genomes (114 alleles of the parents of patients with 22q11.2DS) validated the presence of 5 LCR22-D configurations and extended the LCR22-A catalogue to 32 unique haplotypes [33]. Through optical mapping, visualization of the complex LCR22 repeat structure and LCR22-A variation was possible for the first time.

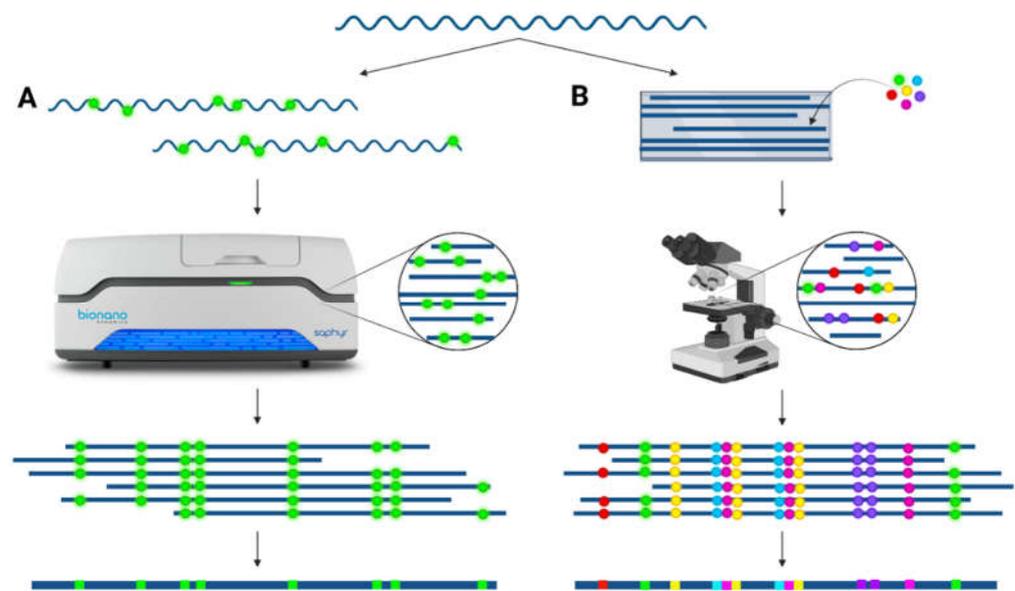


Figure 3. De novo assembly via optical mapping techniques. **(A)** Bionano optical mapping: ultralong (>100 kb) DNA fibers are fluorescently labeled on specific locations across the whole genome and subsequently linearized and scanned. These images are converted into molecules to visualize the labeling pattern across the DNA strand. By overlapping identical patterns, a consensus pattern can be created. **(B)** Fiber-FISH: ultralong (>300 kb) DNA fibers are stretched onto slides and hybridized with a combination of different fluorescently colored probes targeting the locus of interest. The slide is scanned using a microscope, allowing the detection of the different colors via scanning at different excitation levels. A consensus of the region of interest can be compiled based on overlapping colors and distances between the probes.

The expansion and variation of LCR22-A was examined in an evolutionary context, by mapping the LCR22 alleles of great apes (chimpanzee, bonobo, gorilla, orangutan) and one macaque (New World monkey) [34]. The optical mapping data were able to correct misassemblies and close gaps in the existing corresponding reference genomes. In addition, the haplotypes of the investigated samples from the *Pan* genus (five chimpanzees and one bonobo) showed no variation and represented a small, ‘minimal’ haplotype. Therefore, the expansion and variation observed in the human population can be considered human specific [34].

5. The Hunt for the Rearrangement Sites

Pinpointing the crossover sites at nucleotide resolution will give us insights into the mechanisms behind these deletions, the genes involved in the rearrangement, and eventually enable us to conduct phenotype–genotype correlation studies. Unfortunately, nucleotide level identification is complicated by the duplication nature.

Once again, optical mapping techniques have allowed for initial attempts at identifying rearrangements at the LCR22 subunit/duplicon level. In Demaerel et al. (2019), the seven LCR22-A/D recombinations all mapped to a 160 kb module (Figure 4). Pastor et al. (2020) mapped 30 LCR22-A/D deletion trios, and the same 160 kb locus or smaller loci were involved. All family maps demonstrate the association of the rearrangement with the FAM230 locus (Figure 4), raising the potential for this locus to be the preferred 22q11.2DS recombination hotspot.

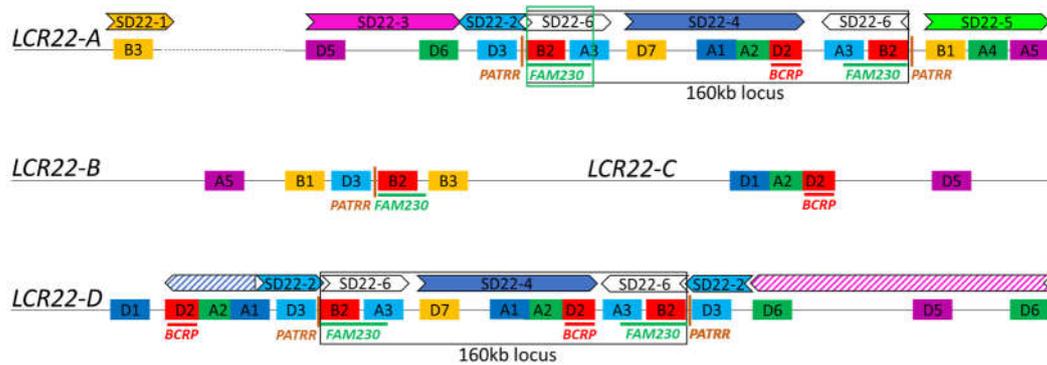


Figure 4. Recombination loci on the LCR22 map. The four proximal LCR22s are depicted schematically based on the fiber-FISH probe composition from Demaerel et al. (2019). LCR22-A is represented as the smallest haplotype including all possible SD22 duplicons. The dotted line reflects the variability in size, copy number, and composition. The SD22 duplicons are illustrated above the fiber-FISH probe compositions of LCR22-A and -D. The BCR module [35,36], FAM230 [33], PATRR [16], and 160 kb locus [32] are indicated in the LCR22s. Drawings are not to scale.

A subclass of 22q11.2 deletions is atypical rearrangements, where one of the breakpoints resides in an LCR22 and the second breakpoint is located in a unique sequence surrounding the LCR22s [24,37–45], offering an opportunity to map the rearrangement sites via the unique sequence. In the International 22q11.2 Brain Behavior Consortium [46], six individuals were identified with atypical deletions. A combination of whole-genome sequencing, fiber-FISH, patient-specific PCR, and targeted long-read sequencing was used to investigate these deletions. They were all located at different loci within the LCR22. Based on the presence of microhomology or polyA insertion at the breakpoints, they likely originated via replication-based mechanisms [47].

The distal LCR22 blocks (LCR22-E until -H) and LCR22-B and -C are substantially smaller compared with LCR22-A and -D, facilitating targeted approaches. Shaikh et al. (2007) used long-range PCR with paralogous-specific primers to amplify and subsequently sequence the breakpoints of two distal 22q11.2 deletions. The breakpoint cluster region (BCR) module was identified as the recombination locus in both an LCR22-D/E and an LCR22-E/F deletion [35]. The breakpoints of the constitutional 22q11.2 translocations all cluster in LCR22-B, described by Emanuel in 2008 as the most rearrangement-prone site of the human genome [16]. By scrutinizing these rearrangements, palindromic AT-rich repeats (PATRRs) were identified at the breakpoint locus in LCR22-B as well as in the involved partner chromosome locus, generating instability and an opportunity for rearrangement via secondary structure formation [11]. Although an important role in the generation of translocations, the contribution of palindrome-mediated pathways to deletions and duplications in the 22q11.2 locus cannot yet be confirmed [16].

To solve the complex LCR22-A/D rearrangement, Guo et al. (2016) charted shared and paralogous sequence polymorphisms between LCR22-A and -D by sequencing BACs containing (parts of) the LCR22s [35]. Based on this variation map, whole-genome sequencing data of two LCR22-A/D patients and their parents were screened for uniquely mapping read pairs within these LCR22s. The results suggested that the BCR module was involved in the recombination mechanism of two recurrent LCR22-A/D deletions (Guo et al. 2016). These studies propose that the BCR module is the hotspot for 22q11.2 chromosomal rearrangements (Figure 4).

The breakpoints of the 22q11.2 rearrangements coincide in the LCR22s, providing evidence for the presence of specific elements involved in nonhomologous rearrangement mechanisms. Both the FAM230 [33] and BCR module [35,36] were suggested as 22q11.2 recombination hotspot. Interestingly, neither of them are present in any of the four proximal LCR22s (Figure 4), and therefore, at least two recombination spots will be involved in the creation of the nested deletions involving LCR22-B and -C. It remains to be seen whether

other rearrangement sites are involved and whether there is variability observed at the nucleotide level. In order to paint the complete ‘recombination landscape’ at the 22q11.2 locus, more deletions should be mapped.

6. The Future: Ultralong Read Sequencing Approaches?

Long-range structural information at the base-pair level will be needed to map the LCR22s and pinpoint the nucleotide rearrangement positions. To reach this goal, long-read technology from Oxford Nanopore holds promise. Average read lengths are dependent on the input material and library preparation method, but reads have been reported to be as long as 2.27 Mb [48]. Error rates are lower than 5% and are continuously improving [49]. ONT sequencing of 3622 Icelanders has shown that the technique is able to accurately characterize structural variants at a population scale [50]. A revolution in CNV research is underway using high-throughput Nanopore sequencing.

Combinations of second- and third-generation sequencing tools and optical mapping techniques are the gold standard for whole-genome de novo assembly [51,52]. The telomere-to-telomere consortium generated the first complete assembly of a human genome, including unresolved loci, such as ribosomal rRNA, centromeric satellites, and segmental duplications. Using ultralong read Nanopore (39X coverage and N50–70 kb) and SMRT PacBio (70X coverage) data, de novo assembly of the homozygous complete hydatidiform mole CHM13 genome was achieved. This initial draft was polished by (shorter) Nanopore and PacBio, linked-read Illumina (10X Genomics), and Bionano optical mapping data. In April 2022, the first complete human genome [4] was assembled. Using this CHM13 haploid cell line with only a single LCR22-A allele, all LCR22s were completely sequenced. Projecting this sequence onto the duplicon structures of the optical mapping patterns described by Demaerel et al. (2019), this assembly represents a common LCR22-A haplotype. Unfortunately, T2T-CHM13 LCR22-A lacks the SD22-3 duplicon. The absence of this duplicon may stymie future analyses using T2T-CHM13 as a reference genome. Nevertheless, the release of this new reference genome makes it possible to compare novel sequences against a more accurate, reliable reference genome and is a major step forward.

7. Remaining Questions and Future Perspectives

Major progress in 22q11.2DS research has been made with the elucidation of the LCR22 structure via optical mapping techniques. New questions may now be probed in the future. What is the extent of this LCR22-A variability? What is the effect of this variability at the transcriptome level? Is there variability in the rearrangement locus, and is this genotype variability associated with the observed phenotypic variability? Are specific haplotype lengths or structures a risk factor for the rearrangements?

First, it will be essential to further catalogue the LCR22 variation at large scale. In addition, new resources and databases must be constructed to collect and visualize the human (LCR22) structural variation, as begun by the Human Pangenome Reference Consortium [53]. They will establish a database of 350 phased genomes, generating a total of 700 haplotypes [53]. The development of bioinformatic tools to solve and construct complex structural variants in a high-throughput way will accelerate future 22q11.2 research, including crossover site identification on a larger scale. This initiative will expand the catalogue of gross 22q11.2 structural variation at the duplicon and sequence level, thus generating a valuable control population dataset.

Second, the consequences of this discovered LCR22 variability must be examined. LCR22-specific structural variation could cause (I) gene-dosage effects in the LCR22-specific transcripts based on copy number variation or new fusion transcripts associated with specific rearrangements and (II) effects on gene expression in LCR22 flanking genes, and (III) exert a genome-wide impact. Of particular interest are the genes located within the LCR22-A sequence (FAM230, RIMBP3, PI4KAP, etc.). Given that an accurate gapless reference genome has only recently become available, genes and transcripts within the LCR22s are poorly characterized. Both regular and haplotype-resolved Hi-C analysis

mapping long-range chromosomal interactions on 22q11.2DS patients demonstrated that LCR22-A and -D act as topologically associated domains [54]. Thus, a nucleotide difference of over 1.75 Mb between the smallest and largest LCR22-A haplotype observed will likely affect topologically associated domain boundaries and may exert an effect at the transcriptional level.

With these major advances, genotype–phenotype association studies are now on the horizon and can be performed to associate haplotype configurations or transcriptomic alterations with a phenotypic feature. These studies may finally begin to interrogate a long-held hypothesis in the field. Since LCR genes in the genome are known to play a role in human neuronal development [5], the neuropsychiatric phenotypic consequences will be of particular interest.

Author Contributions: Conceptualization writing—original draft preparation, writing—review and editing, were performed by L.V. and J.R.V.; visualization by L.V.; supervision and funding acquisition by J.R.V.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fonds Wetenschappelijk Onderzoek (grant numbers G0E117N and G0A2622N to J.R.V.), KU Leuven (C14/18/092 to J.R.V.), and Stichting Marguerite-Marie Delacroix (GV/B-453 to L.V.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bailey, J.A.; Yavor, A.M.; Massa, H.F.; Trask, B.J.; Eichler, E.E. Segmental Duplications: Organization and Impact within the Current Human Genome Project Assembly. *Genome Res.* **2001**, *11*, 1005–1017. [[CrossRef](#)] [[PubMed](#)]
- Bailey, J.A.; Gu, Z.; Clark, R.A.; Reinert, K.; Samonte, R.V.; Schwartz, S.; Adams, M.D.; Myers, E.W.; Li, P.W.; Eichler, E.E. Recent Segmental Duplications in the Human Genome. *Science* **2002**, *297*, 1003–1007. [[CrossRef](#)] [[PubMed](#)]
- Blagojevic, C.; Heung, T.; Theriault, M.; Tomita-Mitchell, A.; Chakraborty, P.; Kernohan, K.; Bulman, D.E.; Bassett, A.S. Estimate of the Contemporary Live-Birth Prevalence of Recurrent 22q11.2 Deletions: A Cross-Sectional Analysis from Population-Based Newborn Screening. *CMAJ Open* **2021**, *9*, E802–E809. [[CrossRef](#)] [[PubMed](#)]
- Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The Complete Sequence of a Human Genome. *Science* **2022**, *376*, 44–53. [[CrossRef](#)] [[PubMed](#)]
- Dennis, M.Y.; Eichler, E.E. Human Adaptation and Evolution by Segmental Duplication. *Curr. Opin. Genet. Dev.* **2016**, *41*, 44–52. [[CrossRef](#)] [[PubMed](#)]
- McDonald-McGinn, D.; Sullivan, K.; Marino, B.; Philip, N.; Swillen, A.; Vorstman, J.; Zackai, E.; Emanuel, B.; Vermeesch, J.; Morrow, B.; et al. 22q11.2 Deletion Syndrome. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15071. [[CrossRef](#)]
- Campbell, I.M.; Sheppard, S.E.; Crowley, T.B.; McGinn, D.E.; Bailey, A.; McGinn, M.J.; Unolt, M.; Homans, J.F.; Chen, E.Y.; Salmons, H.I.; et al. What Is New with 22q? An Update from the 22q and You Center at the Children’s Hospital of Philadelphia. *Am. J. Med. Genet. Part A* **2018**, *176*, 2058–2069. [[CrossRef](#)]
- Burnside, R.D. 22q11.21 Deletion Syndromes: A Review of Proximal, Central, and Distal Deletions and Their Associated Features. *Cytogenet. Genome Res.* **2015**, *146*, 89–99. [[CrossRef](#)]
- Ben-Shachar, S.; Ou, Z.; Shaw, C.A.; Belmont, J.W.; Patel, M.S.; Hummel, M.; Amato, S.; Tartaglia, N.; Berg, J.; Sutton, V.R.; et al. 22q11.2 Distal Deletion: A Recurrent Genomic Disorder Distinct from DiGeorge Syndrome and Velocardiofacial Syndrome. *Am. J. Hum. Genet.* **2008**, *82*, 214–221. [[CrossRef](#)]
- Portnoi, M.F. Microduplication 22q11.2: A New Chromosomal Syndrome. *Eur. J. Med. Genet.* **2009**, *52*, 88–93. [[CrossRef](#)]
- Kato, T.; Kurahashi, H.; Emanuel, B.S. Chromosomal Translocations and Palindromic AT-Rich Repeats. *Curr. Opin. Genet. Dev.* **2012**, *22*, 221–228. [[CrossRef](#)] [[PubMed](#)]
- Hsiao, M.C.; Piotrowski, A.; Alexander, J.; Callens, T.; Fu, C.; Mikhail, F.M.; Claes, K.B.M.; Messiaen, L. Palindrome-Mediated and Replication-Dependent Pathogenic Structural Rearrangements within the NF1 Gene. *Hum. Mutat.* **2014**, *35*, 891–898. [[CrossRef](#)] [[PubMed](#)]
- Shaikh, T.H.; Budarf, M.L.; Celle, L.; Zackai, E.H.; Emanuel, B.S. Clustered 11q23 and 22q11 Breakpoints and 3:1 Meiotic Malsegregation in Multiple Unrelated t(11;22) Families. *Am. J. Hum. Genet.* **1999**, *65*, 1595–1607. [[CrossRef](#)] [[PubMed](#)]

14. Zenagui, R.; Bernicot, I.; Ranisavljevic, N.; Haquet, E.; Ferrieres-Hoa, A.; Pellestor, F.; Anahory, T. Inheritance of Imbalances in Recurrent Chromosomal Translocation t(11;22): Clarification by PGT-SR and Sperm-FISH Analysis. *Reprod. Biomed. Online* **2019**, *39*, 40–48. [[CrossRef](#)]
15. Carter, M.T.; St. Pierre, S.A.; Zackai, E.H.; Emanuel, B.S.; Boycott, K.M. Phenotypic Delineation of Emanuel Syndrome (Supernumerary Derivative 22 Syndrome): Clinical Features of 63 Individuals. *Am. J. Med. Genet. Part A* **2009**, *149A*, 1712–1721. [[CrossRef](#)]
16. Emanuel, B.S. Molecular Mechanisms and Diagnosis of Chromosome 22q11.2 Rearrangements. *Dev. Disabil. Res. Rev.* **2008**, *14*, 11–18. [[CrossRef](#)]
17. Groffen, J.; Stephenson, J.R.; Heisterkamp, N.; de Klein, A.; Bartram, C.R.; Grosveld, G. Philadelphia Chromosomal Breakpoints Are Clustered within a Limited Region, Bcr, on Chromosome 22. *Cell* **1984**, *36*, 93–99. [[CrossRef](#)]
18. Boerma, E.G.; Siebert, R.; Kluijn, P.M.; Baudis, M. Translocations Involving 8q24 in Burkitt Lymphoma and Other Malignant Lymphomas: A Historical Review of Cytogenetics in the Light of Today's Knowledge. *Leukemia* **2009**, *23*, 225–234. [[CrossRef](#)]
19. Riggi, N.; Suva, M.L.; Stamenkovic, I. Ewing's Sarcoma. *N. Engl. J. Med.* **2021**, *384*, 154–164. [[CrossRef](#)]
20. McDermid, H.E.; Morrow, B.E. Genomic Disorders on 22q11. *Am. J. Hum. Genet.* **2002**, *70*, 1077–1088. [[CrossRef](#)]
21. Emanuel, B.S.; Shaikh, T.H. Segmental Duplications: An “expanding” Role in Genomic Instability and Disease. *Nat. Rev. Genet.* **2001**, *2*, 791–800. [[CrossRef](#)] [[PubMed](#)]
22. Dunham, I.; Shimizu, N.; Roe, B.A.; Chissoe, S.; Dunham, I.; Hunt, A.R.; Collins, J.E.; Bruskiewich, R.; Beare, D.M.; Clamp, M.; et al. The DNA Sequence of Human Chromosome 22. *Nature* **1999**, *402*, 489–495. [[CrossRef](#)] [[PubMed](#)]
23. Cole, C.G.; McCann, O.T.; Collins, J.E.; Oliver, K.; Willey, D.; Gribble, S.M.; Yang, F.; McLaren, K.; Rogers, J.; Ning, Z.; et al. Finishing the Finished Human Chromosome 22 Sequence. *Genome Biol.* **2008**, *9*, R78. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Chaisson, M.J.P.; Huddleston, J.; Dennis, M.Y.; Sudmant, P.H.; Malig, M.; Hormozdiari, F.; Antonacci, F.; Surti, U.; Sandstrom, R.; Boitano, M.; et al. Resolving the Complexity of the Human Genome Using Single-Molecule Sequencing. *Nature* **2015**, *517*, 608–611. [[CrossRef](#)] [[PubMed](#)]
26. Chan, S.; Lam, E.; Saghbini, M.; Bocklandt, S.; Hastie, A.; Cao, H.; Holmlin, E.; Borodkin, M. Structural Variation Detection and Analysis Using Bionano Optical Mapping. *Methods Mol. Biol.* **2018**, *1833*, 193–203. [[CrossRef](#)]
27. Mostovoy, Y.; Yilmaz, F.; Chow, S.K.; Chu, C.; Lin, C.; Geiger, E.A.; Meeks, N.J.L.; Chatfield, K.C.; Coughlin, C.R.; Surti, U.; et al. Genomic Regions Associated with Microdeletion/Microduplication Syndromes Exhibit Extreme Diversity of Structural Variation. *Genetics* **2021**, *217*, iyaa038. [[CrossRef](#)]
28. Levy-Sakin, M.; Pastor, S.; Mostovoy, Y.; Li, L.; Leung, A.K.Y.; McCaffrey, J.; Young, E.; Lam, E.T.; Hastie, A.R.; Wong, K.H.Y.; et al. Genome Maps across 26 Human Populations Reveal Population-Specific Patterns of Structural Variation. *Nat. Commun.* **2019**, *10*, 1025. [[CrossRef](#)]
29. Louzada, S.; Komatsu, J.; Yang, F. Fluorescence In Situ Hybridization onto DNA Fibres Generated Using Molecular Combing. In *Fluorescence In Situ Hybridization (FISH)*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 275–293. [[CrossRef](#)]
30. Algady, W.; Louzada, S.; Carpenter, D.; Brajer, P.; Färnert, A.; Rooth, I.; Ngasala, B.; Yang, F.; Shaw, M.A.; Hollox, E.J. The Malaria-Protective Human Glycophorin Structural Variant DUP4 Shows Somatic Mosaicism and Association with Hemoglobin Levels. *Am. J. Hum. Genet.* **2018**, *103*, 769–776. [[CrossRef](#)]
31. Shi, W.; Massaia, A.; Louzada, S.; Handsaker, J.; Chow, W.; McCarthy, S.; Collins, J.; Hallast, P.; Howe, K.; Church, D.M.; et al. Birth, Expansion, and Death of VCY-Containing Palindromes on the Human Y Chromosome. *Genome Biol.* **2019**, *20*, 207. [[CrossRef](#)]
32. Demaerel, W.; Mostovoy, Y.; Yilmaz, F.; Vervoort, L.; Pastor, S.; Hestand, M.S.; Swillen, A.; Vergaelen, E.; Geiger, A.; Coughlin, C.R.; et al. The 22q11 Low Copy Repeats Are Characterized by Unprecedented Size and Structural Variability. *Genome Res.* **2019**, *29*, 1389–1401. [[CrossRef](#)]
33. Pastor, S.; Tran, O.; Jin, A.; Carrado, D.; Silva, B.A.; Uppuluri, L.; Abid, H.Z.; Young, E.; Crowley, T.B.; Bailey, A.G.; et al. Optical Mapping of the 22q11.2DS Region Reveals Complex Repeat Structures and Preferred Locations for Non-Allelic Homologous Recombination (NAHR). *Sci. Rep.* **2020**, *10*, 12235. [[CrossRef](#)]
34. Vervoort, L.; Dierckxsens, N.; Pereboom, Z.; Capozzi, O.; Rocchi, M.; Shaikh, T.H.; Vermeesch, J.R. 22q11.2 Low Copy Repeats Expanded in the Human Lineage. *Front. Genet.* **2021**, *12*, 706641. [[CrossRef](#)] [[PubMed](#)]
35. Shaikh, T.H.; O'Connor, R.J.; Pierpont, M.E.; McGrath, J.; Hacker, A.M.; Nimmakayalu, M.; Geiger, E.; Emanuel, B.S.; Saitta, S.C. Low Copy Repeats Mediate Distal Chromosome 22q11.2 Deletions: Sequence Analysis Predicts Breakpoint Mechanisms. *Genome Res.* **2007**, *17*, 482–491. [[CrossRef](#)] [[PubMed](#)]
36. Guo, X.; Delio, M.; Haque, N.; Castellanos, R.; Hestand, M.S.; Vermeesch, J.R.; Morrow, B.E.; Zheng, D. Variant Discovery and Breakpoint Region Prediction for Studying the Human 22q11.2 Deletion Using BAC Clone and Whole Genome Sequencing Analysis. *Hum. Mol. Genet.* **2016**, *25*, 3754–3767. [[CrossRef](#)] [[PubMed](#)]
37. Beaujard, M.P.; Chantot, S.; Dubois, M.; Keren, B.; Carpentier, W.; Mabboux, P.; Whalen, S.; Vodovar, M.; Siffroi, J.P.; Portnoi, M.F. Atypical Deletion of 22q11.2: Detection Using the FISH TBX1 Probe and Molecular Characterization with High-Density SNP Arrays. *Eur. J. Med. Genet.* **2009**, *52*, 321–327. [[CrossRef](#)] [[PubMed](#)]

38. Amati, F.; Conti, E.; Novelli, A.; Bengala, M.; Digilio, M.C.; Marino, B.; Giannotti, A.; Gabrielli, O.; Novelli, G.; Dallapiccola, B. Atypical Deletions Suggest Five 22q11.2 Critical Regions Related to the DiGeorge/Velo-Cardio-Facial Syndrome. *Eur. J. Hum. Genet.* **1999**, *7*, 903–909. [[CrossRef](#)] [[PubMed](#)]
39. Weksberg, R.; Stachon, A.C.; Squire, J.A.; Moldovan, L.; Bayani, J.; Meyn, S.; Chow, E.; Bassett, A.S. Molecular Characterization of Deletion Breakpoints in Adults with 22q11 Deletion Syndrome. *Hum. Genet.* **2007**, *120*, 837–845. [[CrossRef](#)] [[PubMed](#)]
40. Carlson, C.; Sirotkin, H.; Pandita, R.; Goldberg, R.; McKie, J.; Wadey, R.; Patanjali, S.R.; Weissman, S.M.; Anyane-Yeboah, K.; Warburton, D.; et al. Molecular Definition of 22q11 Deletions in 151 Velo-Cardio-Facial Syndrome Patients. *Am. J. Hum. Genet.* **1997**, *61*, 620–629. [[CrossRef](#)]
41. Levy, A.; Demczuk, S.; Aurias, A.; Depétris, D.; Mattei, M.; Philip, N. Interstitial 22q11 Microdeletion Excluding the ADU Breakpoint in a Patient with DiGeorge Syndrome. *Hum. Mol. Genet.* **1995**, *4*, 2417–2419. [[CrossRef](#)]
42. McQuade, L.; Christodoulou, J.; Budarf, M.L.; Sachdev, R.; Wilson, M.; Emanuel, B.; Colley, A. Patient with a 22q11.2 Deletion with No Overlap of the Minimal DiGeorge Syndrome Critical Region (MDGCR). *Am. J. Hum. Genet.* **1999**, *3*, 27–33. [[CrossRef](#)]
43. Nogueira, S.I.; Hacker, A.M.; Bellucco, F.T.S.; Christofolini, D.M.; Kulikowski, L.D.; Cernach, M.C.S.P.; Emanuel, B.S.; Melaragno, M.I. Atypical 22q11.2 Deletion in a Patient with DGS/VCFS Spectrum. *Eur. J. Med. Genet.* **2008**, *51*, 226–230. [[CrossRef](#)] [[PubMed](#)]
44. O'Donnell, H.; McKeown, C.; Gould, C.; Morrow, B.; Scambler, P. Detection of an Atypical 22q11 Deletion That Has No Overlap with the DiGeorge Syndrome Critical Region. *Am. J. Hum. Genet.* **1997**, *60*, 1544–1548. [[CrossRef](#)] [[PubMed](#)]
45. Uddin, R.K.; Zhang, Y.; Siu, V.M.; Fan, Y.S.; O'Reilly, R.L.; Rao, J.; Singh, S.M. 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. [[CrossRef](#)] [[PubMed](#)]
46. Gur, R.E.; Bassett, A.S.; McDonald-McGinn, D.M.; Bearden, C.E.; Chow, E.; Emanuel, B.S.; Owen, M.; Swillen, A.; den Bree, M.V.; Vermeesch, J.; et al. A Neurogenetic Model for the Study of Schizophrenia Spectrum Disorders: The International 22q11.2 Deletion Syndrome Brain Behavior Consortium. *Mol. Psychiatry* **2017**, *22*, 1664–1672. [[CrossRef](#)]
47. Vervoort, L.; Demaerel, W.; Rengifo, L.Y.; Odrzywolski, A.; Vergaelen, E.; Hestand, M.S.; Breckpot, J.; Devriendt, K.; Swillen, A.; McDonald-McGinn, D.M.; et al. Atypical Chromosome 22q11.2 Deletions Are Complex Rearrangements and Have Different Mechanistic Origins. *Hum. Mol. Genet.* **2019**, *28*, 3724–3733. [[CrossRef](#)]
48. Payne, A.; Holmes, N.; Rakyar, V.; Loose, M. BulkVis: A Graphical Viewer for Oxford Nanopore Bulk FAST5 Files. *Bioinformatics* **2019**, *35*, 2193–2198. [[CrossRef](#)]
49. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [[CrossRef](#)]
50. Beyter, D.; Ingimundardottir, H.; Oddsson, A.; Eggertsson, H.P.; Bjornsson, E.; Jonsson, H.; Atlason, B.A.; Kristmundsdottir, S.; Mehringer, S.; Hardarson, M.T.; et al. Long-Read Sequencing of 3,622 Icelanders Provides Insight into the Role of Structural Variants in Human Diseases and Other Traits. *Nat. Genet.* **2021**, *53*, 779–786. [[CrossRef](#)]
51. Gordon, D.; Huddleston, J.; Chaisson, M.J.P.; Hill, C.M.; Kronenberg, Z.N.; Munson, K.M.; Malig, M.; Raja, A.; Fiddes, I.; Hillier, L.D.W.; et al. Long-Read Sequence Assembly of the Gorilla Genome. *Science* **2016**, *352*, aae0344. [[CrossRef](#)]
52. Mao, Y.; Catacchio, C.R.; Hillier, L.D.W.; Porubsky, D.; Li, R.; Sulovari, A.; Fernandes, J.D.; Montinaro, F.; Gordon, D.S.; Storer, J.M.; et al. A High-Quality Bonobo Genome Refines the Analysis of Hominid Evolution. *Nature* **2021**, *594*, 77–81. [[CrossRef](#)] [[PubMed](#)]
53. Wang, T.; Antonacci-Fulton, L.; Howe, K.; Lawson, H.A.; Lucas, J.K.; Phillippy, A.M.; Popejoy, A.B.; Asri, M.; Carson, C.; Chaisson, M.J.P.; et al. The Human Pangenome Project: A Global Resource to Map Genomic Diversity. *Nature* **2022**, *604*, 437–446. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, X.; Zhang, Y.; Zhu, X.; Purmann, C.; Haney, M.S.; Ward, T.; Khechaduri, A.; Yao, J.; Weissman, S.M.; Urban, A.E. Local and Global Chromatin Interactions Are Altered by Large Genomic Deletions Associated with Human Brain Development. *Nat. Commun.* **2018**, *9*, 5356. [[CrossRef](#)] [[PubMed](#)]