# Supplementary Materials for

## Multiple paralogues and recombination mechanisms contribute to the high incidence of 22q11.2 Deletion Syndrome

**Authors**

Lisanne Vervoort[1], Nicolas Dierckxsens[1,2], Marta Sousa Santos[1], Senne Meynants[1], Erika Souche[1], Ruben Cools[1], Tracy Heung[3], Koen Devriendt[1], Hilde Peeters[1], Donna M. McDonald-McGinn[4,5,6], Ann Swillen[1], Jeroen Breckpot[1], Beverly S. Emanuel[4,5], Hilde Van Esch[1], Anne S. Bassett[3], Joris R. Vermeesch[1,*]

**Affiliations**

[1] Department of Human Genetics, KU Leuven, Leuven, Belgium
[2] Genomics and Regulatory Systems Unit & Marine Climate Change Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
[3] Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, ON, Canada
[4] Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[5] Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, PA, USA
[6] Department of Human Biology and Medical Genetics, Sapienza University, Rome, Italy

*Corresponding author: joris.vermeesch@kuleuven.be

**The supplementary materials include:**

**Table S1**: **Sample overview and information.**

Overview of the family samples used in the study with extra information on site of origin, deletion type, sequencing approach, and final breakpoint identification. POO = parent-of-origin, ONT = Oxford Nanopore Technologies, RL = recombination locus (based on Table 1).

| *Family* | *Site of origin* | *Individual* | *Deletion type* | *Sequencing* | *Breakpoint identification* |
|---|---|---|---|---|---|
| AD001 | Toronto | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD2 |
| | | Father (POO) | | / | |
| AD002 | Toronto | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Father (POO) | | / | |
| AD003 | Toronto | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Father (POO) | | / | |
| AD004 | Toronto | Patient | LCR22-A/D | Standard-long read sequencing (ONT) | Fiber-FISH and optical mapping: RL-AD1 |
| | | Father (POO) | | Standard-, Ultra-long and high duplex read sequencing (ONT) | |
| AD005 | Toronto | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD2 |
| | | Father (POO) | | / | |
| AD006 | Toronto | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Mother (POO) | | / | |
| AD007 | Toronto | Patient | LCR22-A/D | Standard-long read sequencing (ONT) | Fiber-FISH: RL-AD1<br>*De novo* assembly and analysis:<br>intronic/exonic in *GGT3P* (LCR22-A) and *GGT2P* (LCR22-D), L2 LINE element in locus. |
| | | Mother (POO) | | Standard-long read sequencing (ONT) | |

| | | | | | |
|---|---|---|---|---|---|
| AD008 | Philadelphia | Patient | LCR22-A/D | Ultra-long read sequencing (ONT) | Fiber-FISH: RL-AD1<br>*De novo* assembly and analysis:<br>intronic/exonic in *GGT3P* (LCR22-A) and *GGT2P* (LCR22-D), L2 LINE element in locus. |
| | | Mother (POO) | | Ultra-long read sequencing (ONT) | |
| AD009 | Leuven | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Father (POO) | | / | |
| AD010 | Leuven | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Father (POO) | | / | |
| AD011 | Leuven | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Mother (POO) | | / | |
| AD012 | Leuven | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Mother (POO) | | / | |
| AD013 | Leuven | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Mother (POO) | | / | |
| AD014 | New York | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Father (POO) | | / | |
| AD015 | New York | Patient | LCR22-A/D | / | Fiber-FISH: RL-AD1 |
| | | Mother (POO) | | / | |
| AB001 | Toronto | Patient | LCR22-A/B | Ultra-long read sequencing (ONT) | Fiber-FISH: RL-B2<br>manual investigation sequencing reads: PATRR |
| | | Father (POO) | | Ultra-long read sequencing (ONT) | |

| | | | | | |
|---|---|---|---|---|---|
| AB002 | Leuven | Patient (mosaic) | LCR22-A/B deletion and inversion | Ultra-long read sequencing (ONT) | Fiber-FISH: RL-B2<br>manual investigation sequencing reads: PATRR |
| | | Father (POO) | | Ultra-long read sequencing (ONT) | |
| AB003 | Leuven | Patient | LCR22-A/B | / | Fiber-FISH: RL-B1 |
| | | Father (POO) | | / | |
| AB004 | Leuven | Patient | LCR22-A/B | Long-range PCR and PacBio SMRT sequencing | Fiber-FISH: RL-B2<br>manual investigation amplicon: PATRR (same result observed in two children with LCR22-A/B deletion) |
| AB005 | Leuven | Patient | LCR22-A/B | / | Fiber-FISH: RL-B1 |
| | | Father (POO) | | / | |
| AC001 | Toronto | Patient | LCR22-A/C | Ultra-long read sequencing (ONT) | Fiber-FISH: RL-C<br>*De novo* assembly and analysis:<br>intronic in *BCRP* locus (AluJo element) |
| | | Father (POO) | | Ultra-long read sequencing (ONT) | |
| AC002 | Toronto | Patient | LCR22-A/C | Ultra-long read sequencing (ONT) | Fiber-FISH: RL-C<br>*De novo* assembly and analysis:<br>Proximal *POM121L15P/4P*, distal from AluS SINE element |
| | | Father (POO) | | Ultra-long read sequencing (ONT) | |
| AC003 | Leuven | Patient | LCR22-A/C | / | Fiber-FISH: RL-C |
| | | Mother (POO) | | / | |
| BD001 | Toronto | Patient | LCR22-B/D (LCR22-A/B rearrangement, Fig. 4A) | Standard- and Ultra-long read sequencing (ONT) | Fiber-FISH: RL-B1<br>*De novo* assembly and analysis:<br>intronic in *FAM230D* (LCR22-A) and *FAM230G* (LCR22-B) |
| | | Father (POO) | | Standard-long read sequencing (ONT) | |
| CD001 | Toronto | Patient | LCR22-C/D | Standard- and Ultra-long read sequencing (ONT) | Fiber-FISH: RL-C<br>*De novo* assembly and analysis:<br>intronic in *POM121L8P* (LCR22-D) |
| | | Mother (POO) | | Standard- and Ultra-long read sequencing (ONT) | |

**Suplementary Table S2: Sequencing run statistics**

| *Sample* | *Sequencing device* | *Approach* | *Sequencing output* | *N50* |
|---|---|---|---|---|
| **AD004** - Patient | Promethion (R9) - Leuven | Whole-genome standard-long | 106.7Gb | 21.5kb |
| **AD004** - Parent | Promethion (R10) - Leuven | Whole-genome standard-long<br>Whole-genome ultra-long<br>Whole-genome high duplex | 102.8Gb<br>20.7Gb<br>74.57Gb | 34.7kb<br>58.2kb<br>30.06kb |
| **AD007** - Patient | Promethion (R10) - Leuven | Whole-genome standard-long | 83.5Gb | 30.9kb |
| **AD007** - Parent | Promethion (R10) - Leuven | Whole-genome standard-long | 115.4Gb | 21.8kb |
| **AD008** - Patient | Promethion (R9) - Leuven | Whole-genome ultra-long | 52.9Gb / 12.6Gb / 82.7Gb | 51.1kb / 55.3kb / 61.9kb |
| **AD008** - Parent | Promethion (R9) - Leuven | Whole-genome ultra-long | 23.4Gb / 19.9Gb / 55.14Gb | 54.0kb / 52.8kb / 93.0kb |
| **AB001** - Patient | Promethion (R9) - Leuven | Whole-genome ultra-long | 42.9Gb / 75Gb | 61.2kb / 81.5kb |
| **AB001** - Parent | Promethion (R9) - Leuven | Whole-genome ultra-long | 29.9Gb / 74.6Gb | 104.0kb / 99.3kb |
| **AB002** - Patient | Promethion (R9) - Leuven | Whole-genome ultra-long | 52.6Gb / 111.4Gb | 75kb / 66kb |
| **AB002** - Parent | Promethion (R9) - Leuven | Whole-genome ultra-long | 26.6Gb / 14.8Gb | 126kb / 132kb |
| **AB004** - Patient | PacBio RSII - Leuven | Long-range PCR | 482 barcode reads | 1.8kb |
| **AC001** - Patient | Promethion (R9) - Leuven | Whole-genome ultra-long | 81.6Gb | 89.0kb |
| **AC001** - Parent | Promethion (R9) - Leuven | Whole-genome ultra-long | 88.2Gb | 76.0kb |
| **AC002** - Patient | Promethion (R9) - Leuven | Whole-genome ultra-long | 6Gb / 12.5Gb / 20.7Gb / 65.6Gb | 114.7kb / 132.3kb / 102.1kb / 105.8kb |
| **AC002** - Parent | Promethion (R9) - Leuven | Whole-genome ultra-long | 4.7Gb / 4.4 Gb / 55.7Gb | 32.5kb / 29.9 kb / 107.7kb |
| **BD001** - Patient | Promethion (R9) - Leuven<br>Promethion (R10) - Leuven | Whole-genome ultra-long<br>Whole-genome standard-long | 54.5Gb<br>52.6Gb / 39.5Gb | 114.1kb<br>20.6kb / 24.7kb |

| | | | | |
|---|---|---|---|---|
| **BD001** - Parent | Promethion (R10) - Leuven | Whole-genome standard-long | 47.1Gb / 55.1Gb | 29.6kb / 25.9kb |
| **CD001** - Patient | Promethion (R9) - Leuven<br>Promethion (R10) - Leuven | Whole-genome ultra-long<br>Whole-genome standard-long | 48.0Gb<br>106.8Gb | 90.5kb<br>26.1kb |
| **CD001** - Parent | Promethion (R9) - Leuven<br>Promethion (R10) - Leuven | Whole-genome ultra-long<br>Whole-genome standard-long | 15.5Gb / 2.2Gb / 56.5Gb<br>38.6 Gb / 36.3 Gb | 76.8kb / 80.9kb / 109.0kb<br>24.8kb / 24.4kb |

**Table S3: Inversion screening via interphase-FISH in families AB002 and BD001**

| Sample | Screener 1 | Screener 2 | Screener 3 |
|---|---|---|---|
| AB002 patient<br>(LCR22-A/B deletion and LCR22-A/B inversion) | N = 50<br>normal - deletion = 25 (50%)<br>normal - inversion = 19 (38%)<br>other = 6 (9%) | N = 21<br>normal - deletion = 7 (33%)<br>normal - inversion = 12 (57%)<br>other = 2 (10%) | - |
| Control individual<br>(negative control) | N = 50<br>normal - normal = 42 (84%)<br>other = 8 (16%) | N = 24<br>normal - normal = 22 (92%)<br>other = 2 (8%) | - |
| Individual with LCR22-A/B deletion<br>(positive control for deletion) | N = 50<br>normal - deletion = 43 (86%)<br>other = 7 (14%) | N = 39<br>normal - deletion = 34 (87%)<br>other = 5 (13%) | - |
| BD001 patient<br>(LCR22-A | - | - | N = 26<br>inversion = 100% |
| BD001 parent-of-origin<br>(LCR22-A/D inversion) | - | - | N = 17<br>inversion = 53% |
| BD001 non parent-of-origin<br>(no inversion) | - | - | N = 28<br>inversion = 15% |

**Table S4: CNV and size polymorphisms in PATRR-HSATI-AluY fragments**

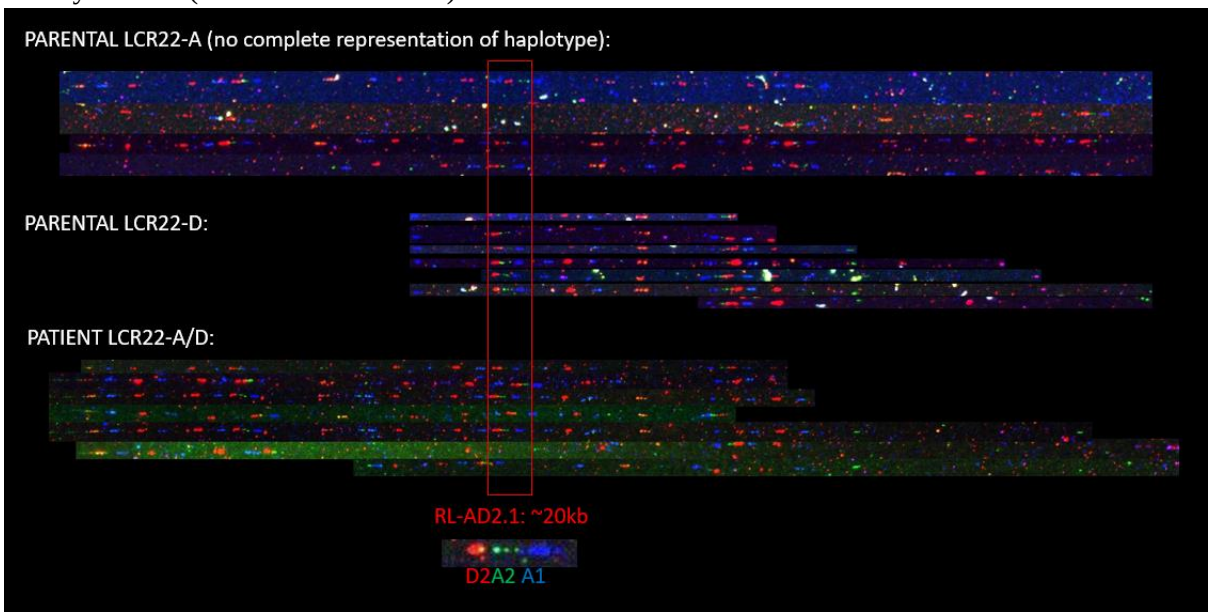
Copy numbers of PATRR-HSATI-AluY fragments determined by RepeatMasker on the first PATRR site in LCR22-A and all the PATRR sites in LCR22-B and –D of publicly available assemblies. The allele column is based on the assembly names extracted from public databases and is not applicable (NA) for GCh38/hg38 and T2T-CHM13v2.0 due to the nature of these assemblies.

| Assembly | Allele | LCR22-A #1 | LCR22-B | LCR22-D #1 | LCR22-D #2 |
|---|---|---|---|---|---|
| GCh38/hg38 | NA | 3<br>(6168 bp) | 2<br>(3268 bp) | 3<br>(4148 bp) | 3<br>(4049 bp) |
| T2T-CHM13v2.0 | NA | 3<br>(6312 bp) | 26<br>(25311 bp) | 3<br>(4288 bp) | 4<br>(6871 bp) |
| HG002 | Paternal | 3<br>(6312 bp | 4<br>(7458 bp) | 3<br>(4288 bp) | 3<br>(4288 bp) |
| | Maternal | 3<br>(6337 bp) | 5<br>(8550 bp) | 2<br>(3159 bp) | 4<br>(8269 bp) |
| HG005 | Paternal | 3<br>(6352 bp) | 5<br>(8308 bp) | 3<br>(4288 bp) | 4<br>(8068 bp) |
| | Maternal | 2<br>(3630 bp) | 4<br>(5515 bp) | 3<br>(4146 bp) | 4<br>(8039 bp) |
| HG00733 | Paternal | 2<br>(6336 bp) | 6<br>(9168 bp) | 3<br>(4288 bp) | 4<br>(8053 bp) |
| | Maternal | 3<br>(6312 bp) | 10<br>(12382 bp) | 5<br>(5690 bp) | 4<br>(8055 bp) |
| HG01109 | Paternal | 3<br>(6311 bp) | 4<br>(5560 bp) | 3<br>(4288 bp) | 4<br>(8073 bp) |
| | Maternal | 2<br>(3839 bp) | 5<br>(8250 bp) | 3<br>(3943 bp) | 4<br>(7903 bp) |

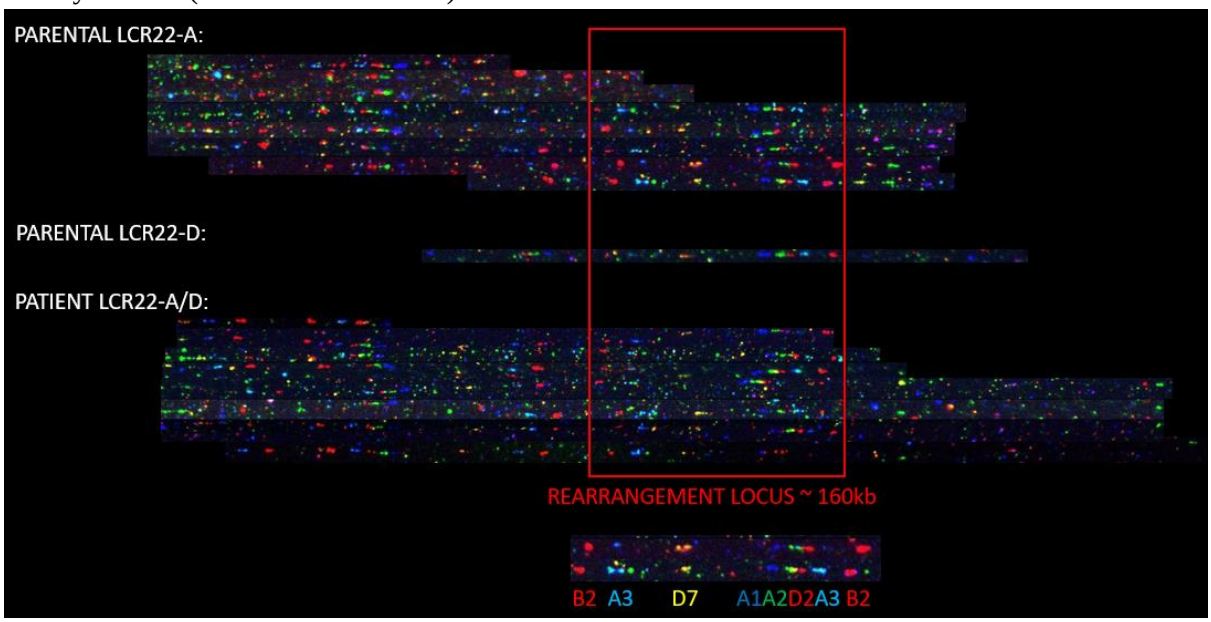
**Figure S1: Fiber-FISH recombination patterns**

Visualization of fiber-FISH recombination patterns. Alleles are *de novo* assembled compiling regions of interest based on matching colors and distances between the probes. The LCR22 haplotypes involved in the recombination are depicted: proximal and distal LCR22 for the parent-of-origin, the recombination for the patient. The red box indicates the region of overlap between the three alleles, indicating the region where the recombination should have taken place. In addition, the RL locus **(Table 1)** is reported, as well as a zoom of the locus with specific probe IDs, based on Demaerel et al, 2019. Only the rearranged parental alleles are presented. The other alleles are not shown.

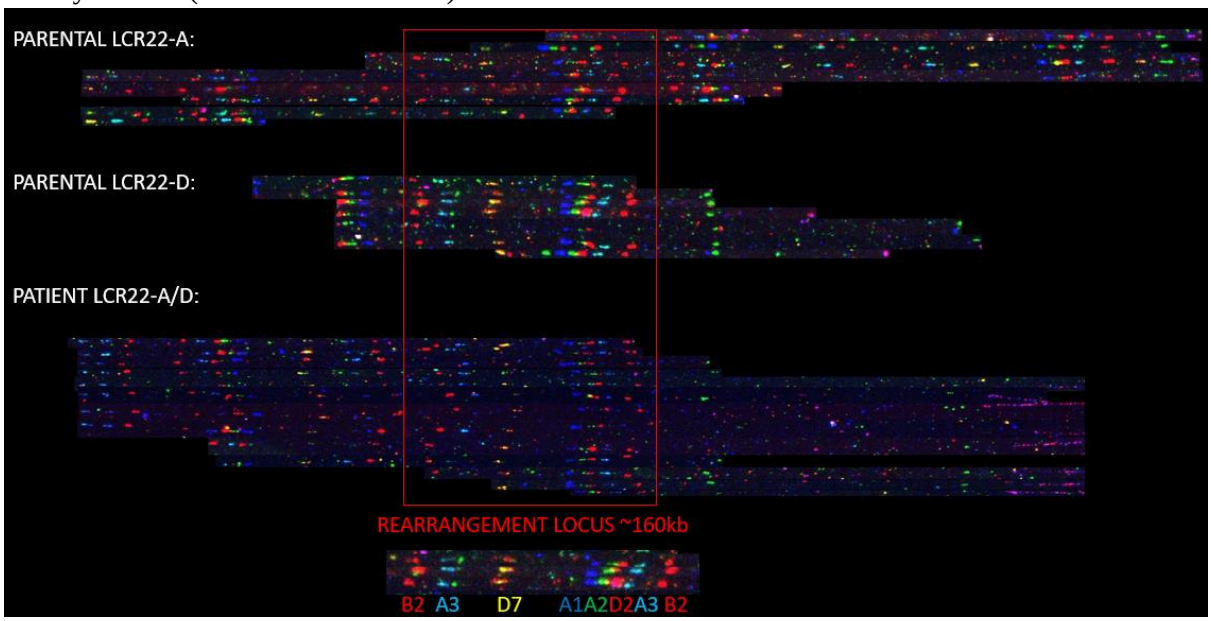
Family AD001 (LCR22-A/D deletion):



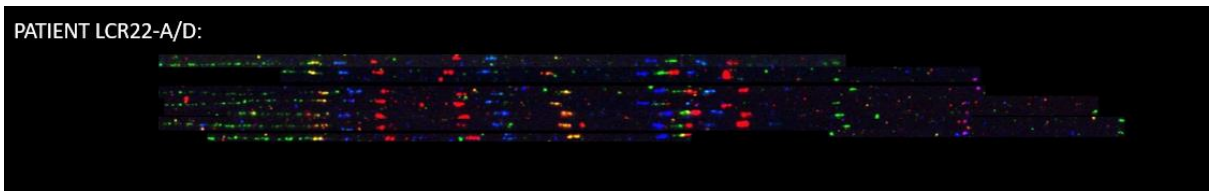

Family AD002 (LCR22-A/D deletion):



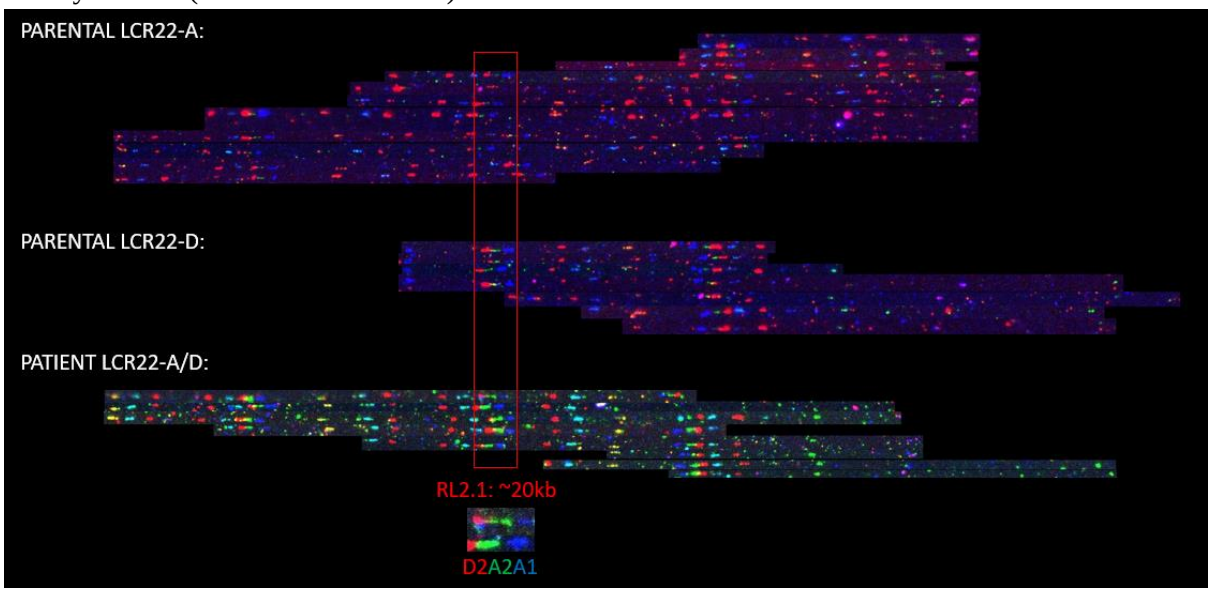
Family AD003 (LCR22-A/D deletion):



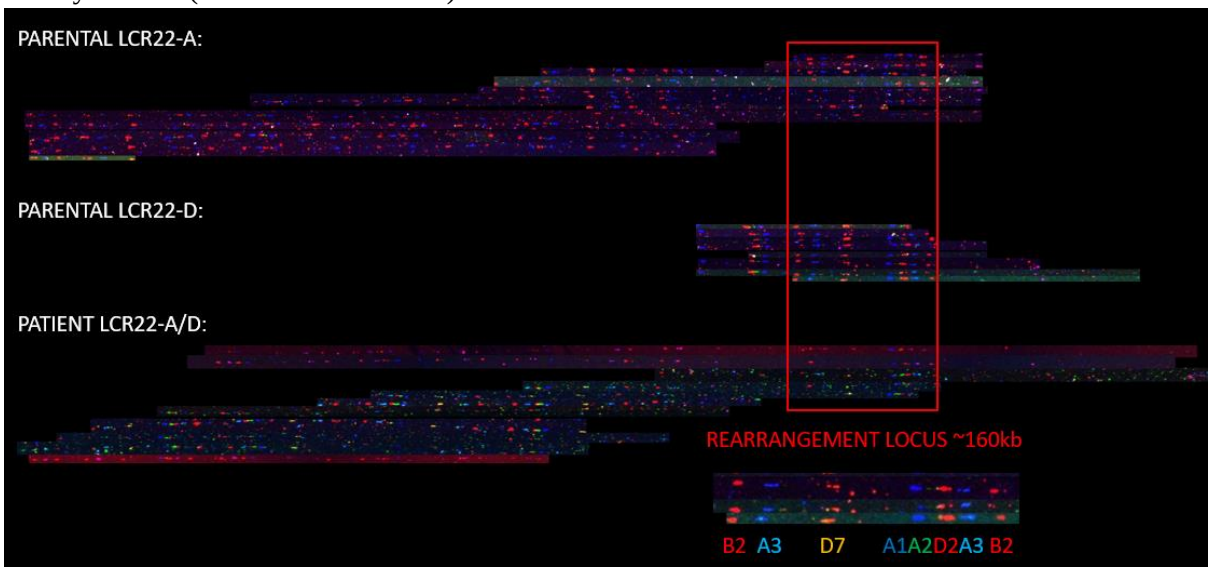

Family AD004 (LCR22-A/D deletion):



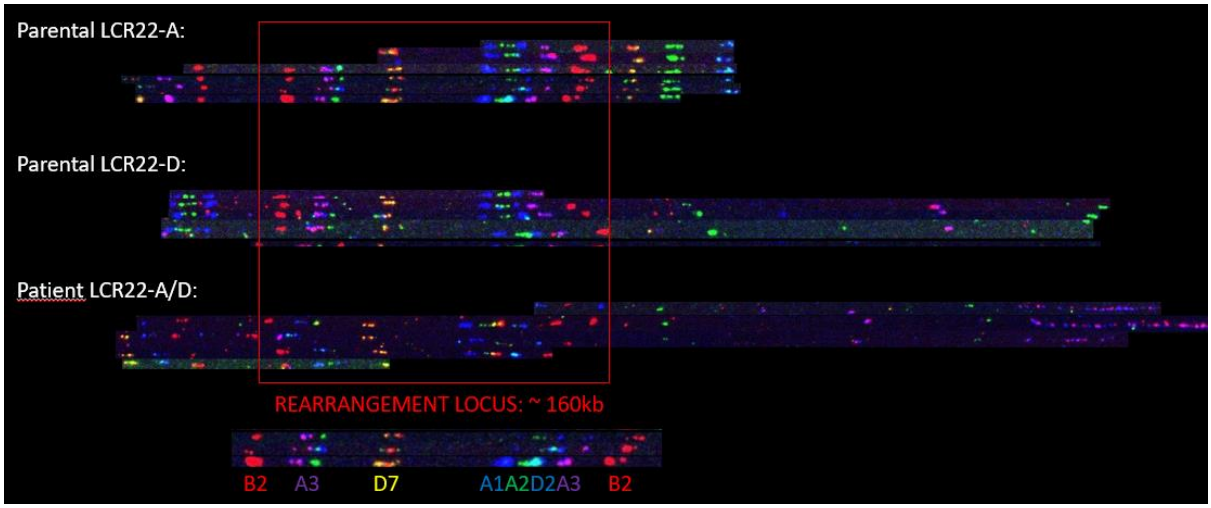

Family AD005 (LCR22-A/D deletion):



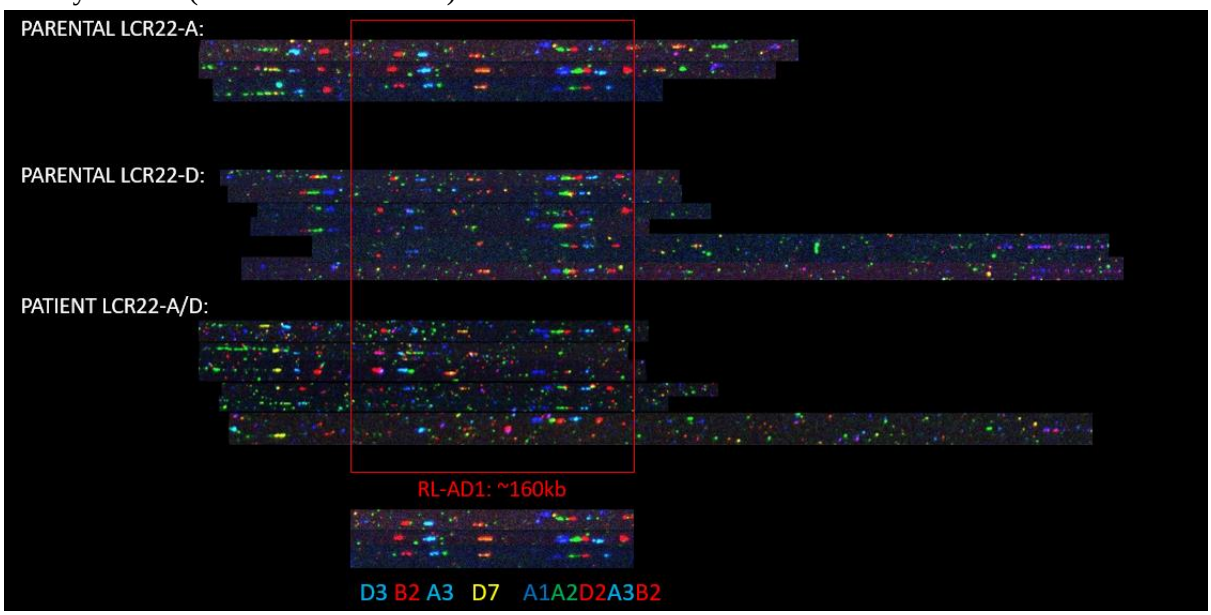
Family AD006 (LCR22-A/D deletion):



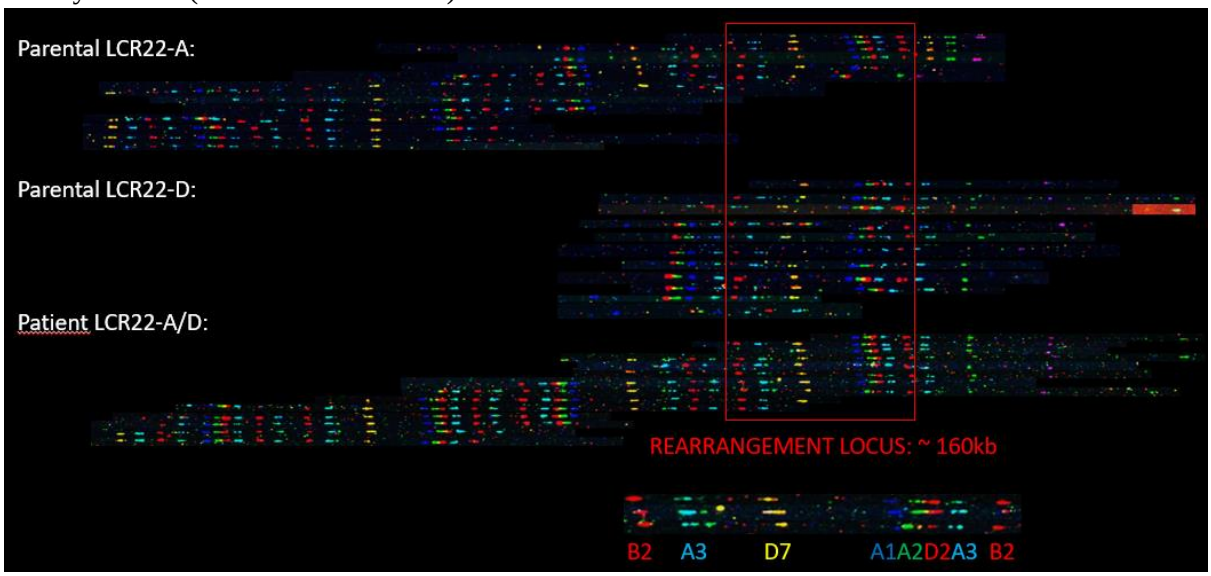

Family AD007 (LCR22-A/D deletion):



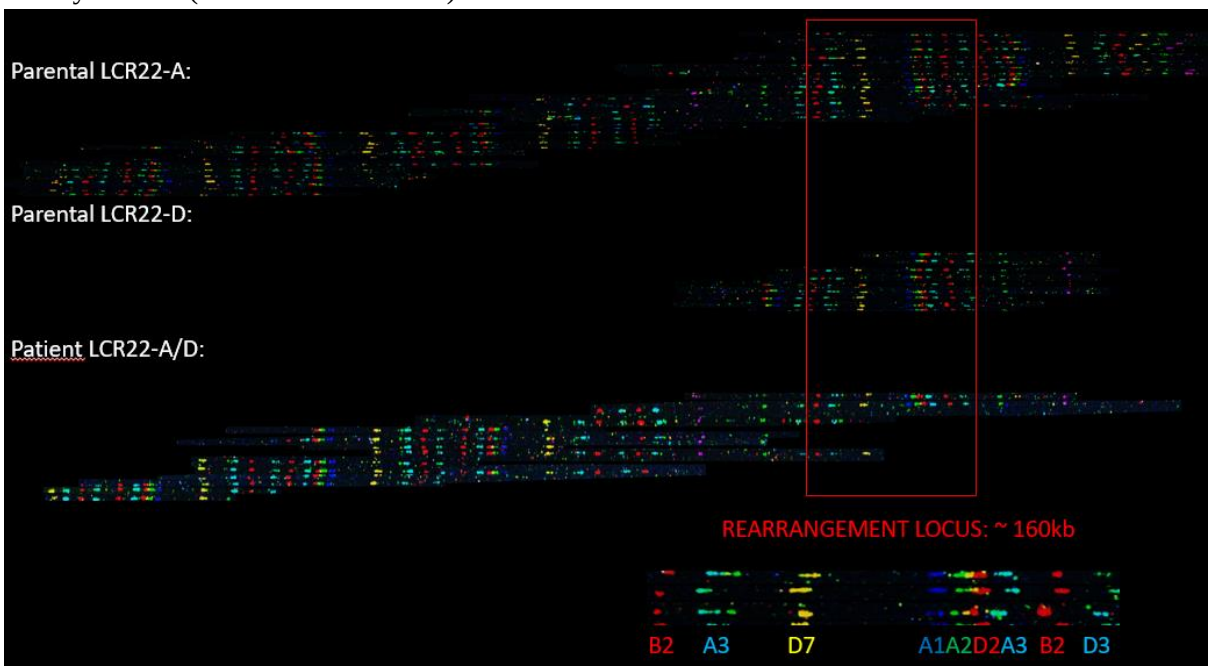

Family AD008 (LCR22-A/D deletion):



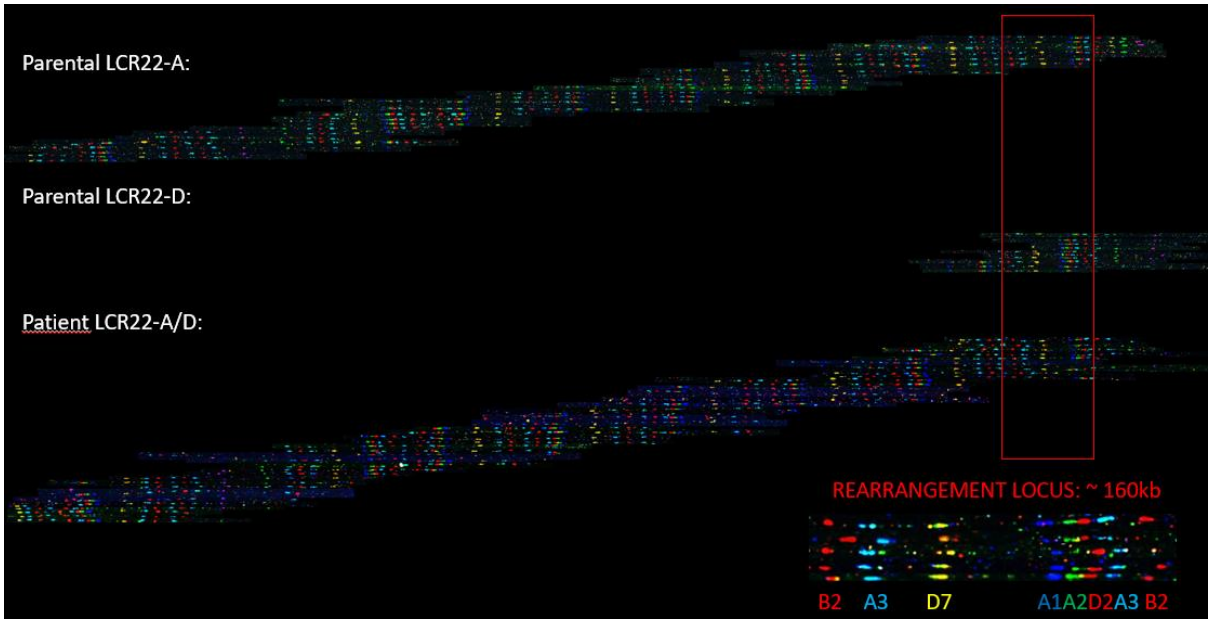
Family AD009 (LCR22-A/D deletion):



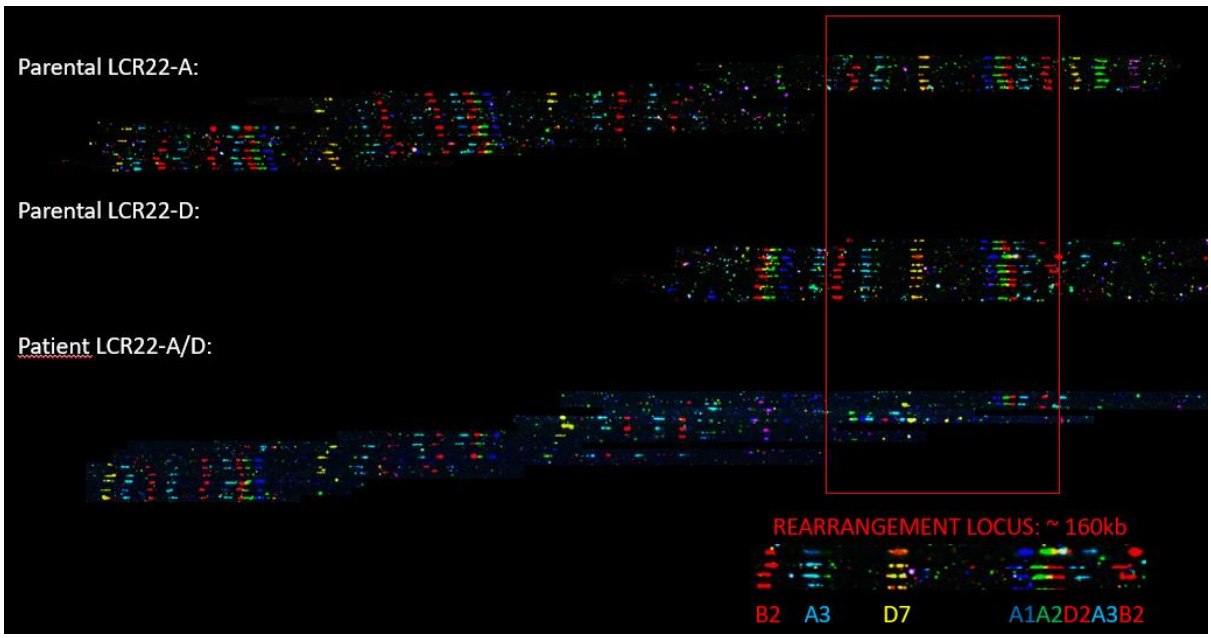

Family AD010 (LCR22-A/D deletion):



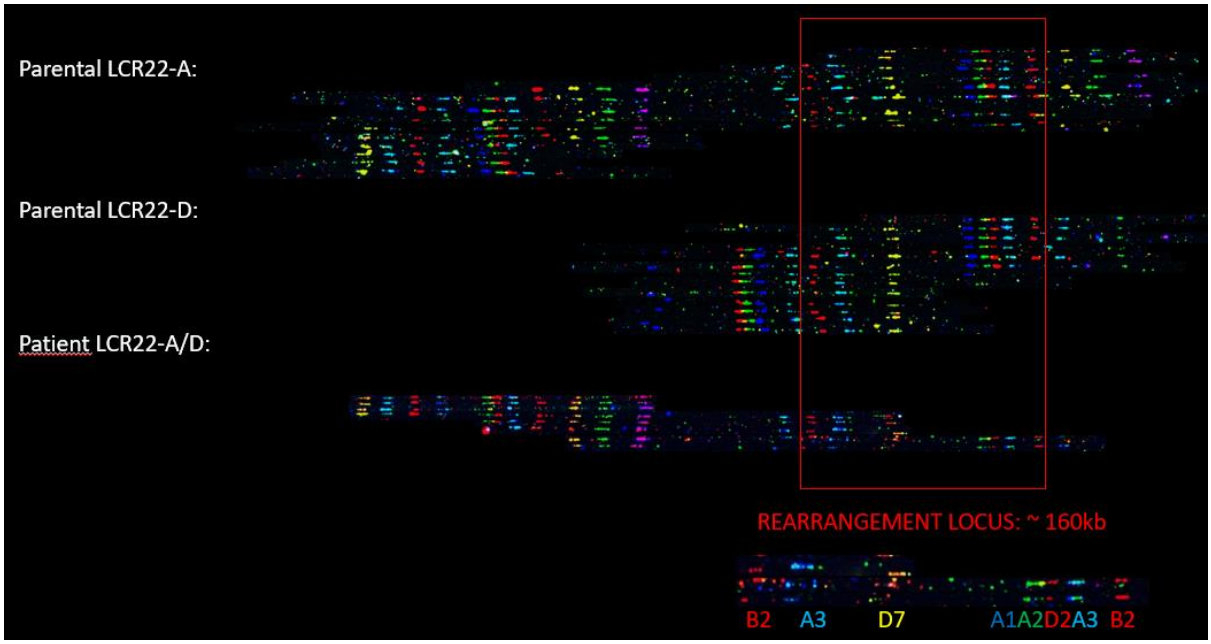
Family AD011 (LCR22-A/D deletion):



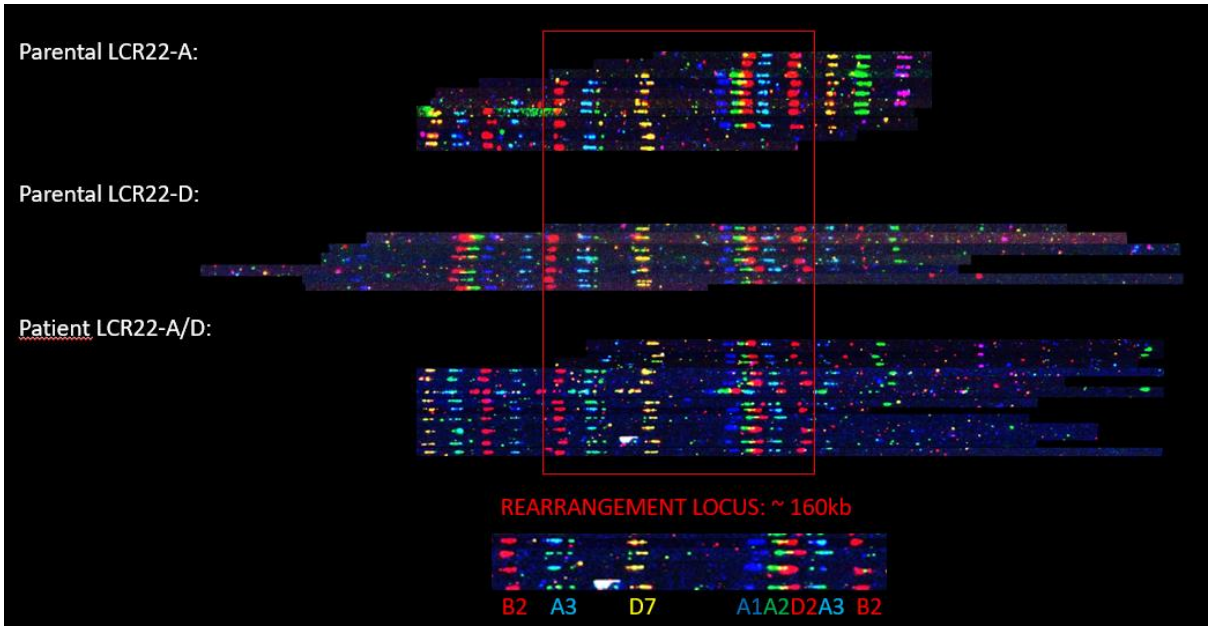

Family AD012 (LCR22-A/D deletion):



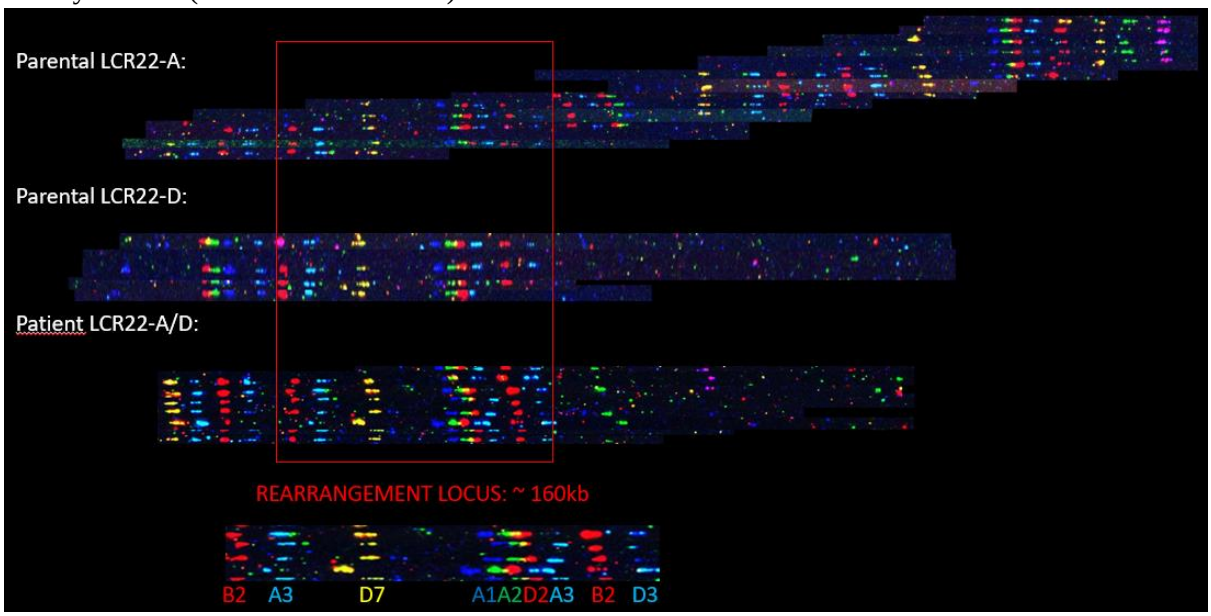
Family AD013 (LCR22-A/D deletion):



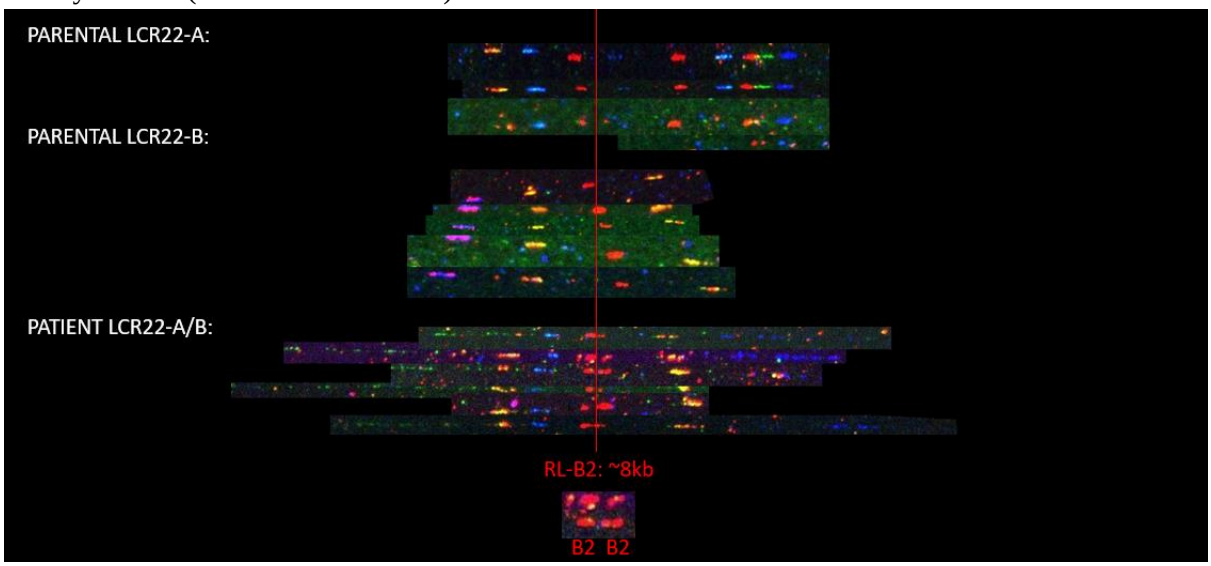

Family AD014 (LCR22-A/D deletion):



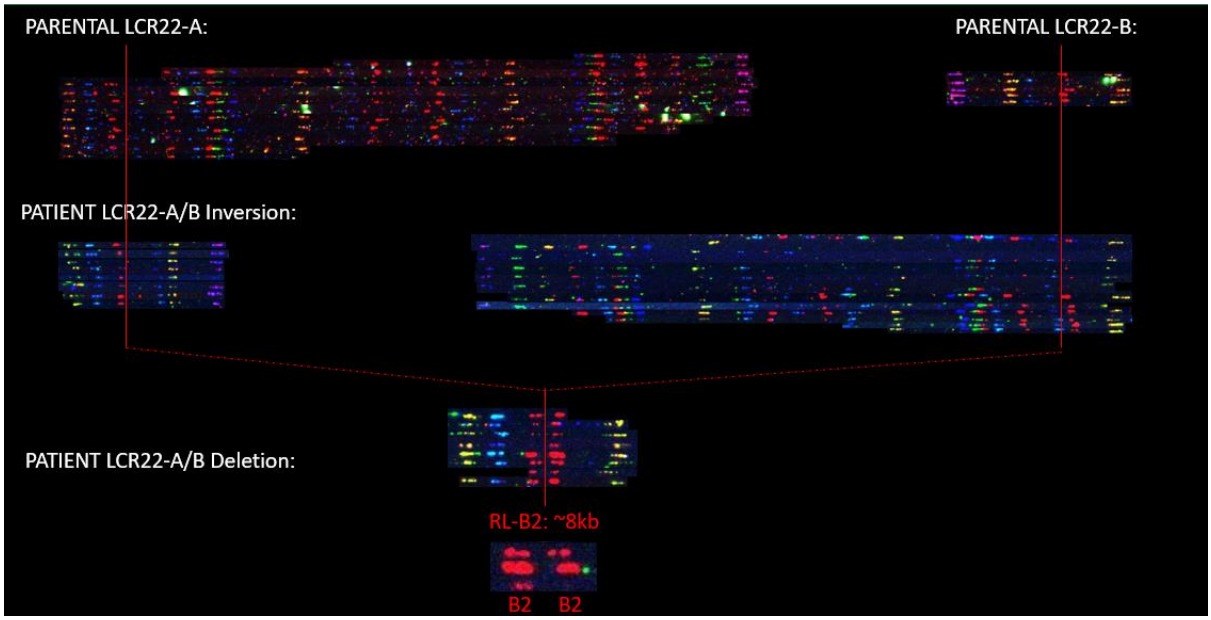
Family AD015 (LCR22-A/D deletion):



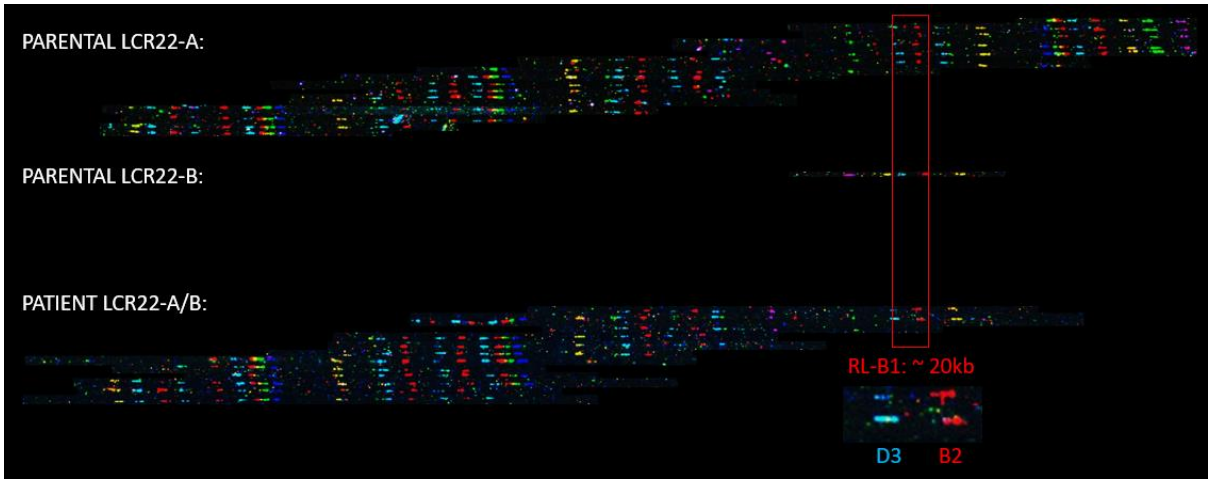

Family AB001 (LCR22-A/B deletion):



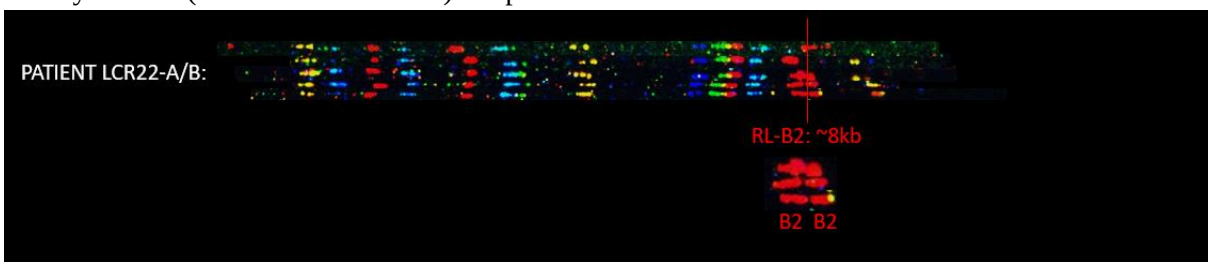
Family AB002 (LCR22-A/B inversion and LCR22-A/B deletion):



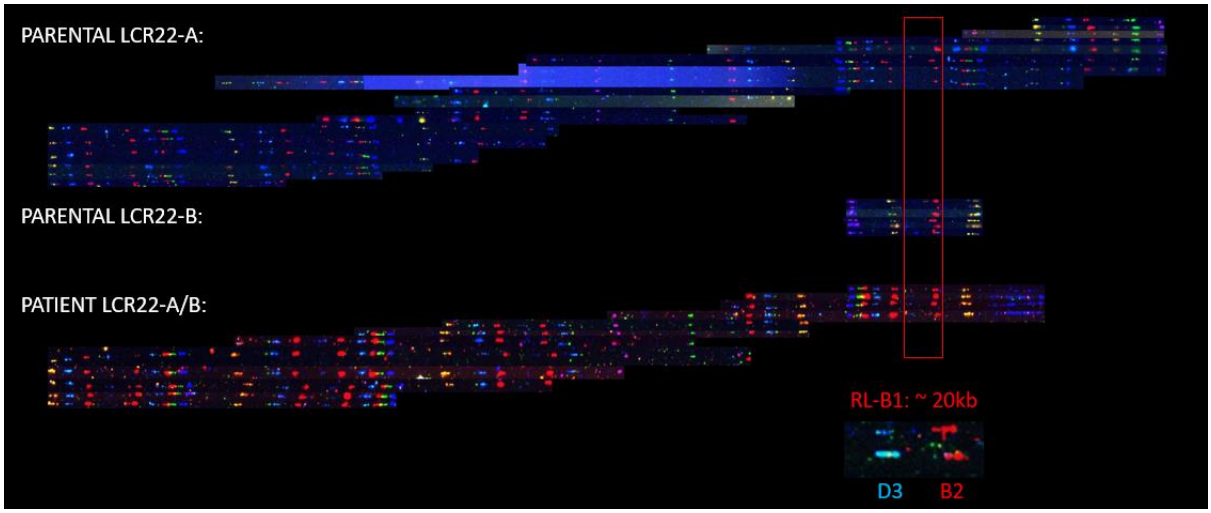

Family AB003 (LCR22-A/B deletion):



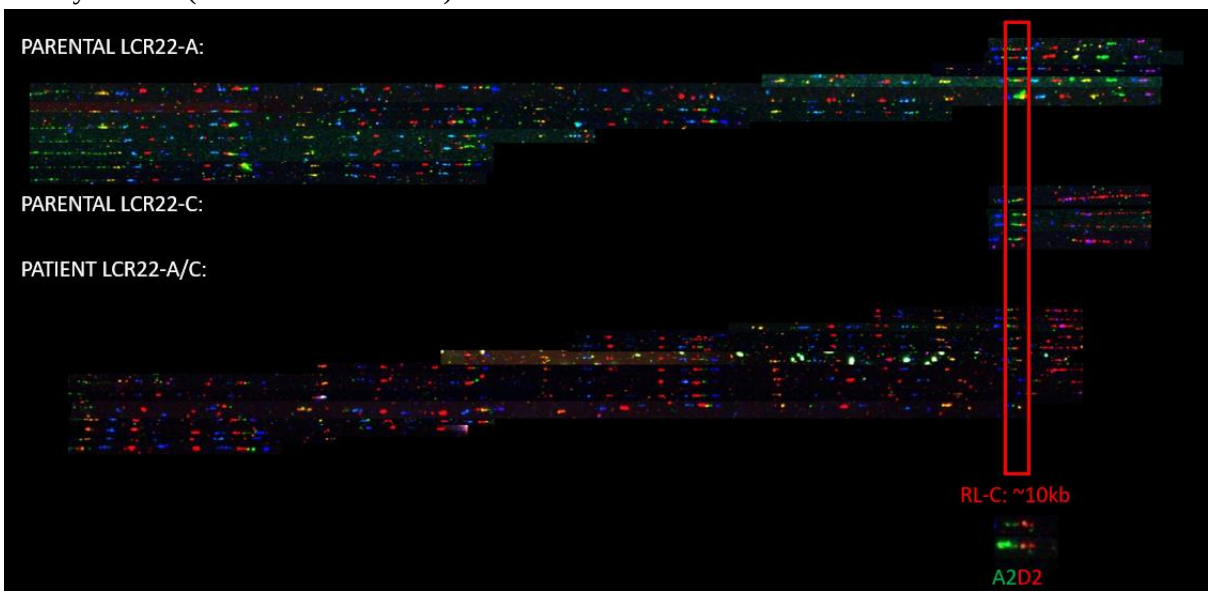

Family AB004 (LCR22-A/B deletion): no parent available



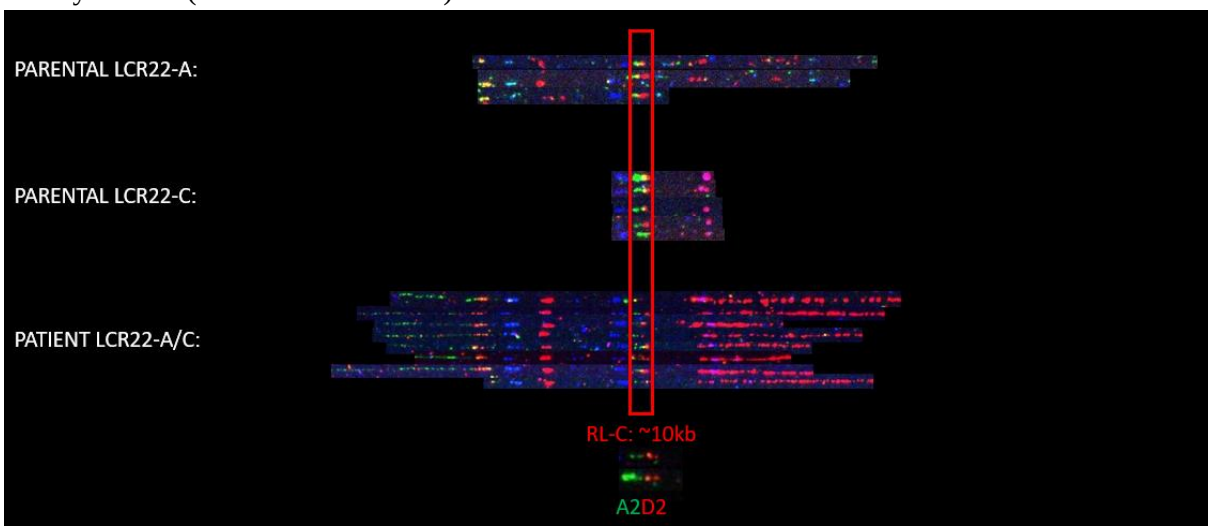
Family AB005 (LCR22-A/B deletion):



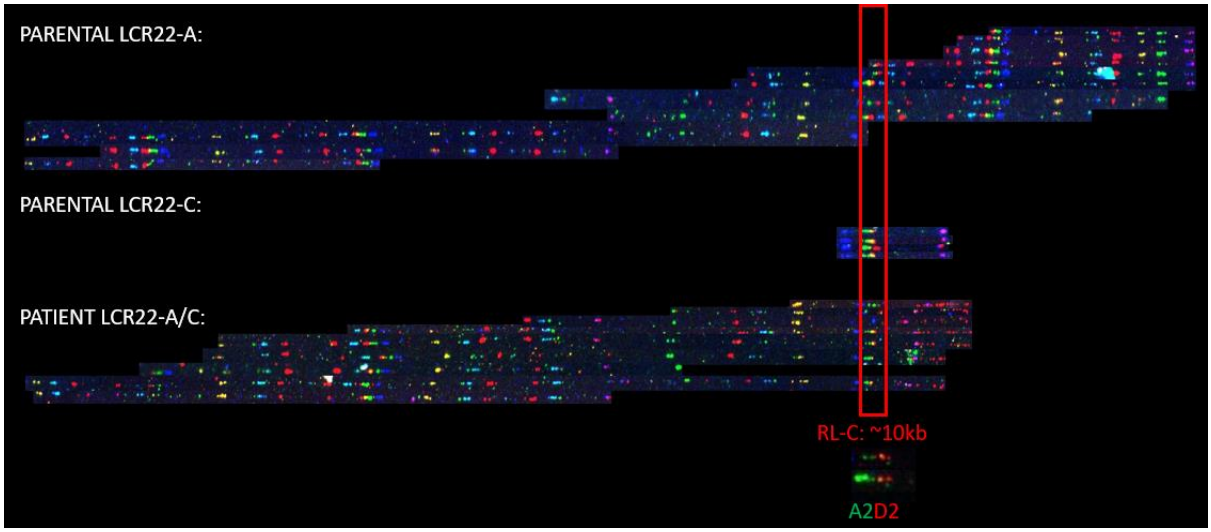

Family AC001 (LCR22-A/C deletion):



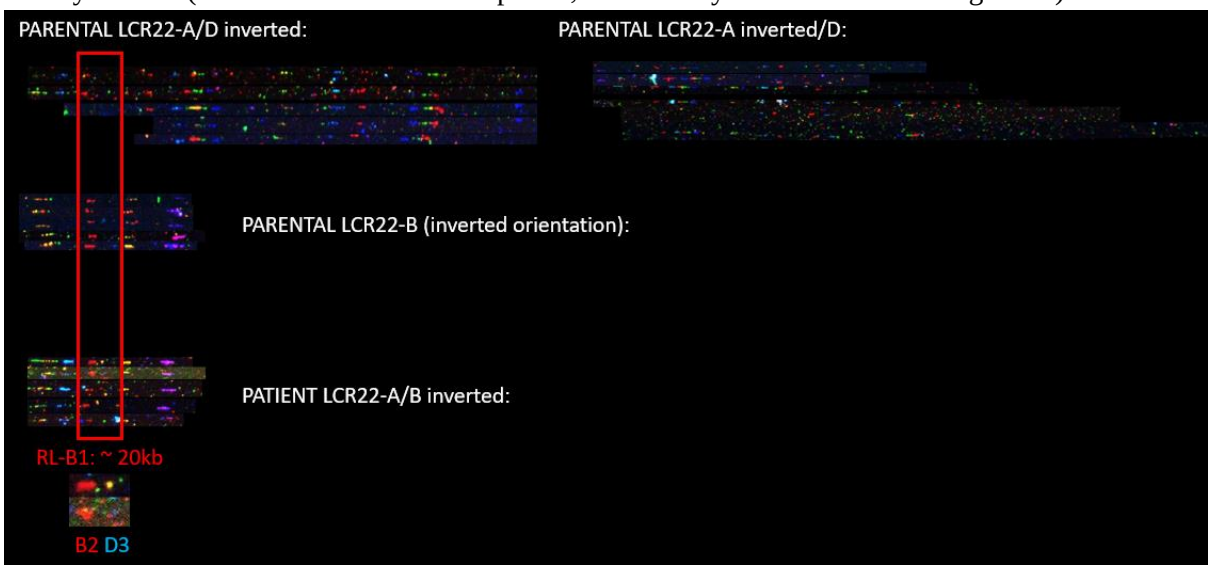

Family AC002 (LCR22-A/C deletion):



Family AC003 (LCR22-A/C deletion):




Family BD001 (LCR22-A/D inversion in parent, followed by LCR22-A/B rearrangement):




Family CD001 (LCR22-C/D deletion):

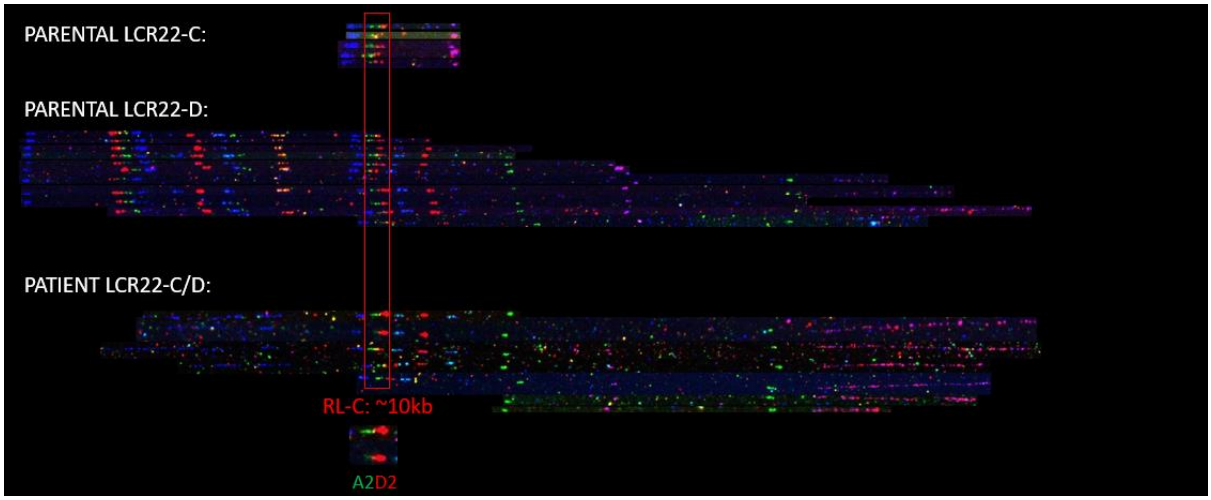


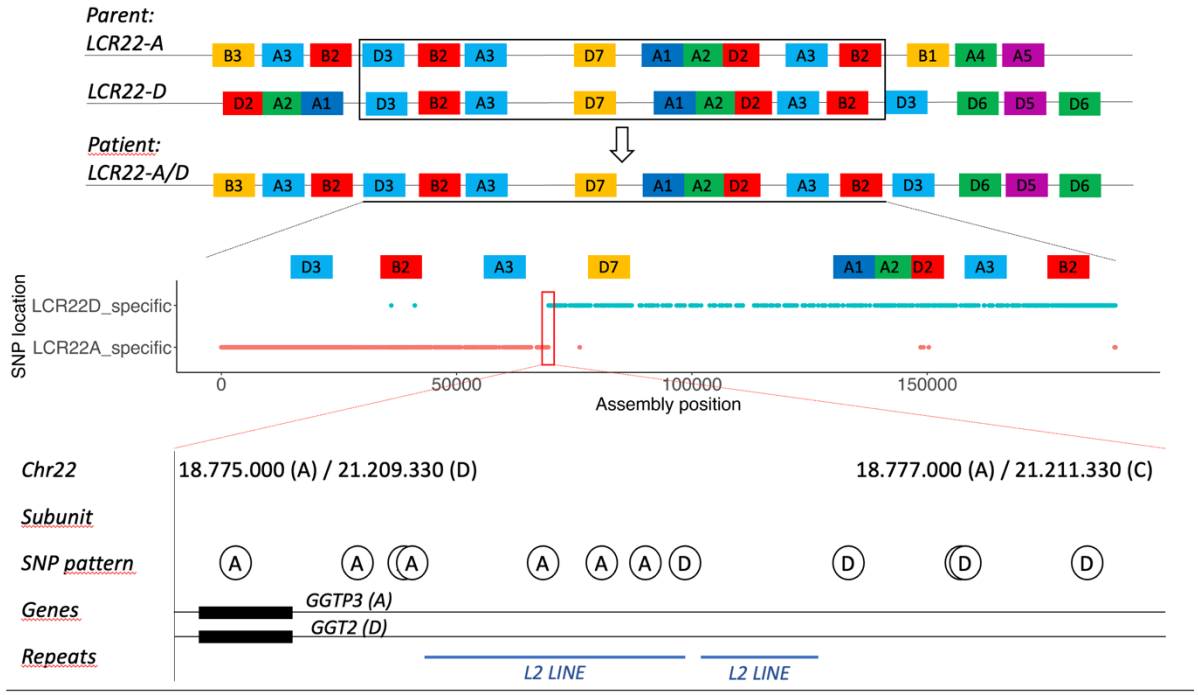
**Figure S2: Bionano optical papping of parental AD004**

Bionano optical mapping of the parent of AD004 family. The Bionano *de novo* assemblies are shown for the 22q11 locus (top picture) and specifically for LCR-A (middle picture) while the bottom picture represents the hg38 and parental LCR-A Fiber-FISH probes. All assemblies are mapped against the hg38 reference genome. The Bionano data shows a deletion starting between 18.18M and 18.34M and ending around 18.75M in hg38. Such a deletion corresponds to the one obtained using the ONT *de novo* assembly (blue rectangle, bottom picture).

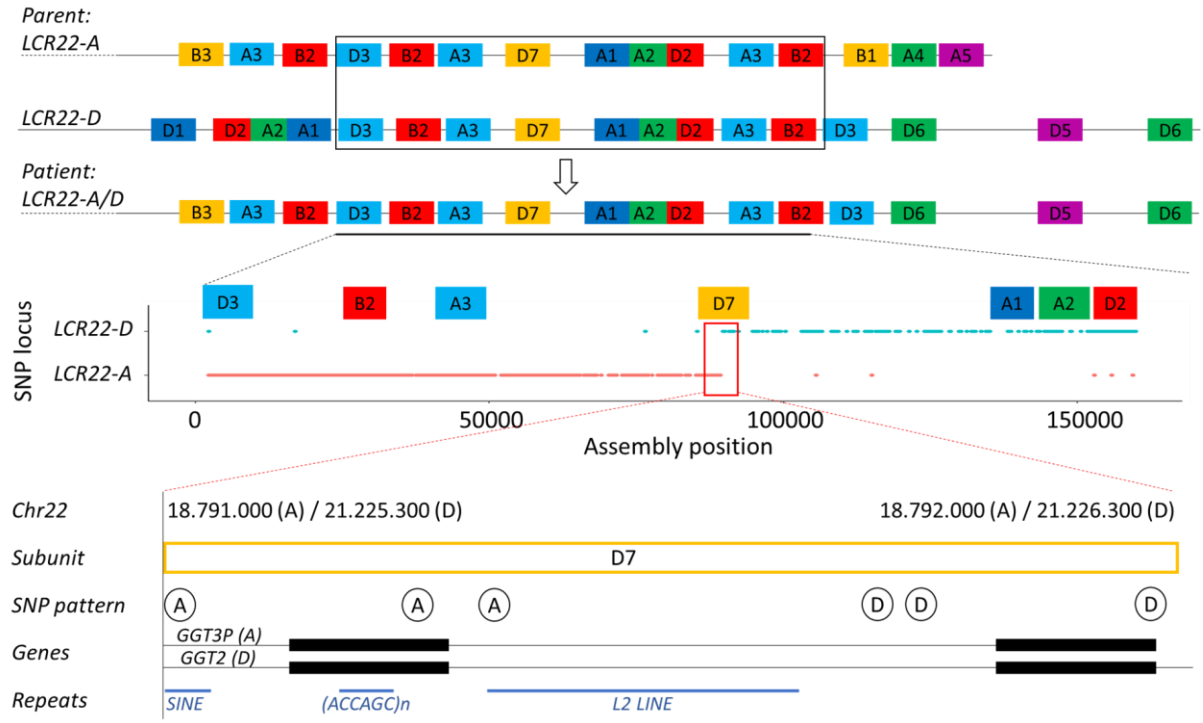
**Figure S3: Nucleotide-level recombination patterns**

Each family overview consists of three parts. (Upper part) Schematic representation of the Fiber-FISH haplotypes of the family. The black box indicates the shared subunits between the two parental alleles where crossover had taken place. Drawings are not to scale. (Middle part) Zoom of the shared LCR22 loci and representation of the LCR22-specific SNPs based on alignment of the LCR22-proximal (parent), LCR22-distal (parent), and LCR22-rearranged (patient) assemblies. SNPs present in the parental LCR22-proximal and patient LCR22-rearranged, but not in the parental LCR22-distal, are considered as LCR22-proximal specific SNPs (red, lower band). SNPs present in the parental LCR22-distal and patient LCR22-rearranged, but not in the parental LCR22-proximal, are considered as LCR22-distal specific SNPs (blue, upper band). The LCR22-specific SNPs are plotted along the patient assembly position. A SNP-specific density switch is observed from LCR22-proximal to LCR22-distal in a locus (red box), considered as the crossover site. (Lower part) This crossover site coordinates of the proximal and distal LCR22 in hg38 in combination with the observed fiber-FISH probes (subunit), SNP pattern, genes, and repeats.

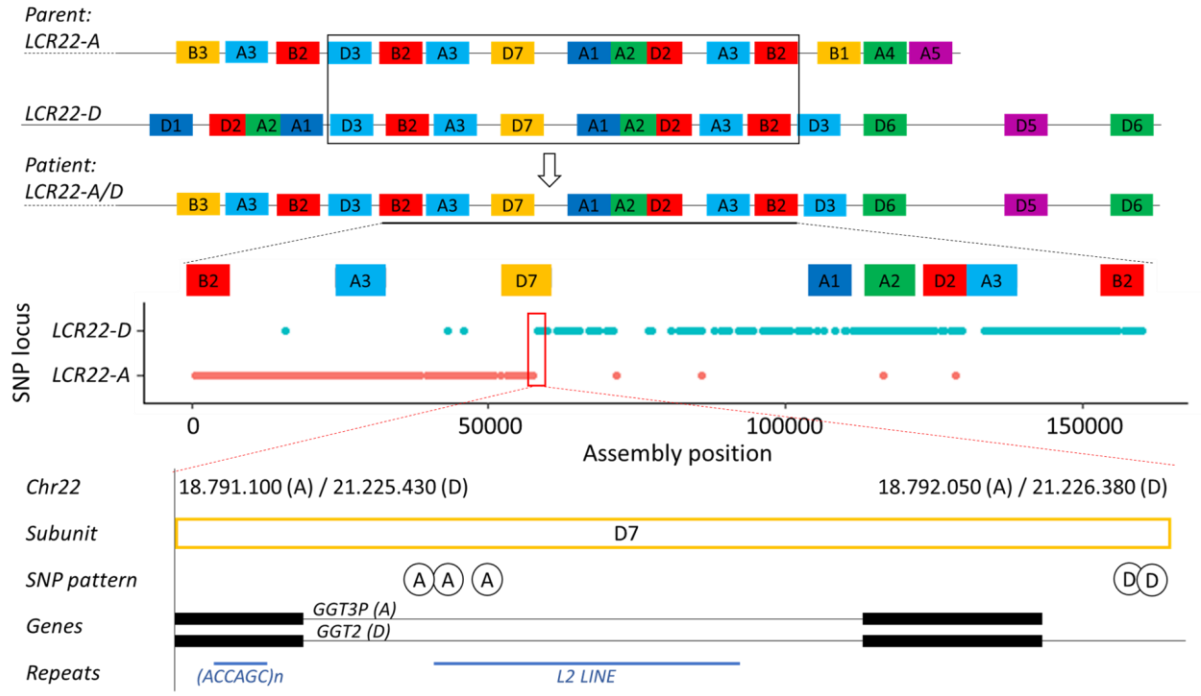
Family AD004 (LCR22-A/D):



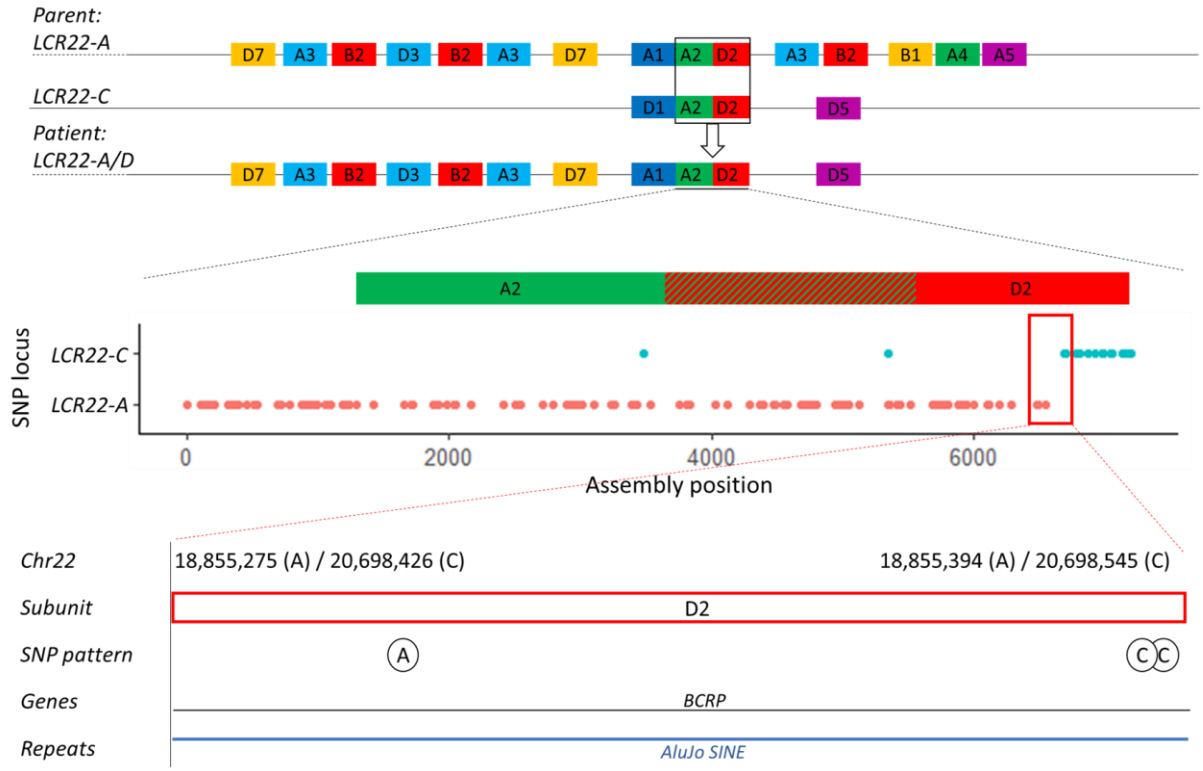
Family AD007 (LCR22-A/D):



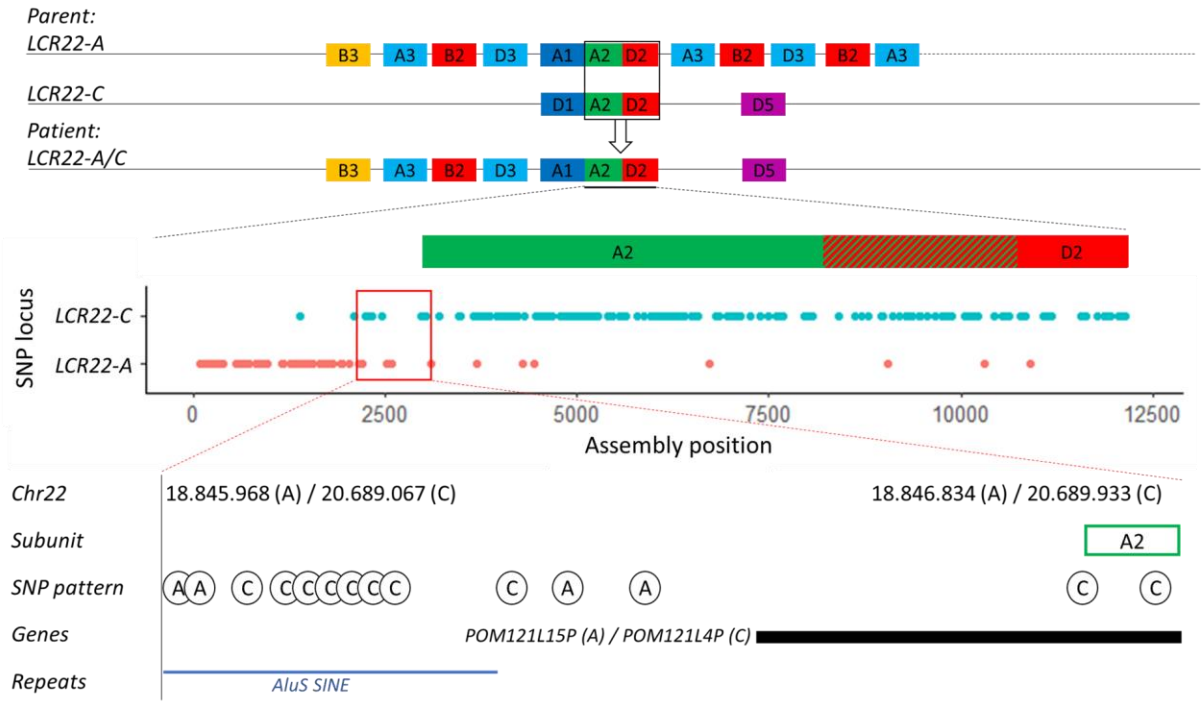

Family AD008 (LCR22-A/D):



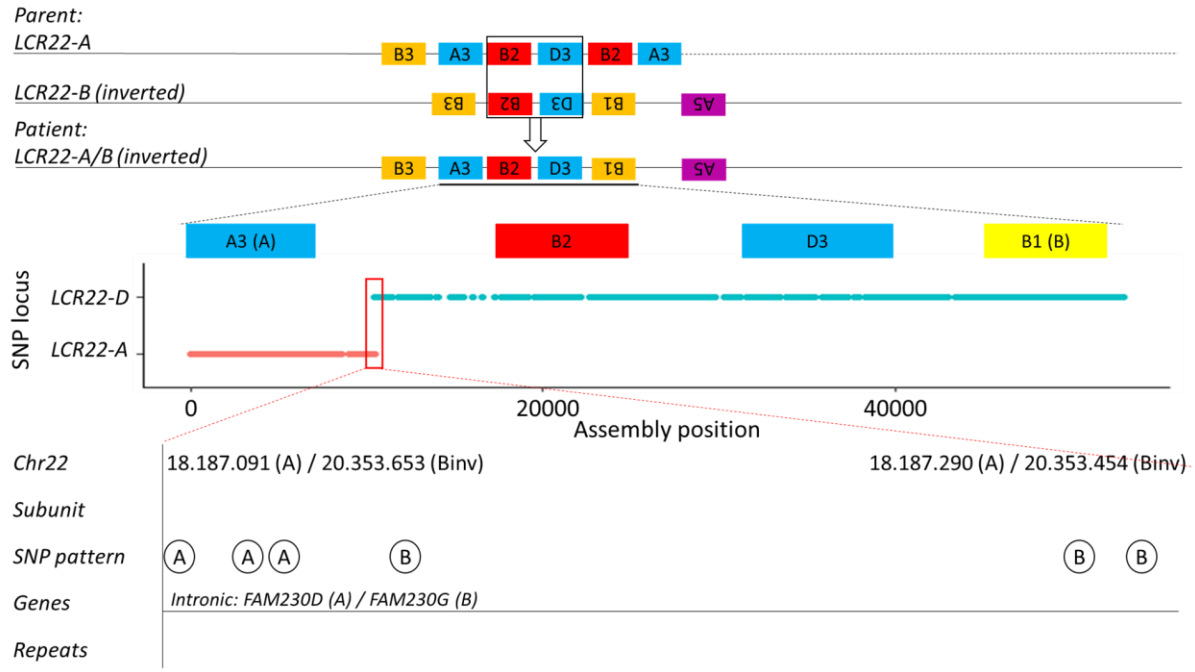
Family AC001 (LCR22-A/C):




Family AC002 (LCR22-A/C):



Family BD001 (LCR22-B/D):




Family CD001 (LCR22-C/D):

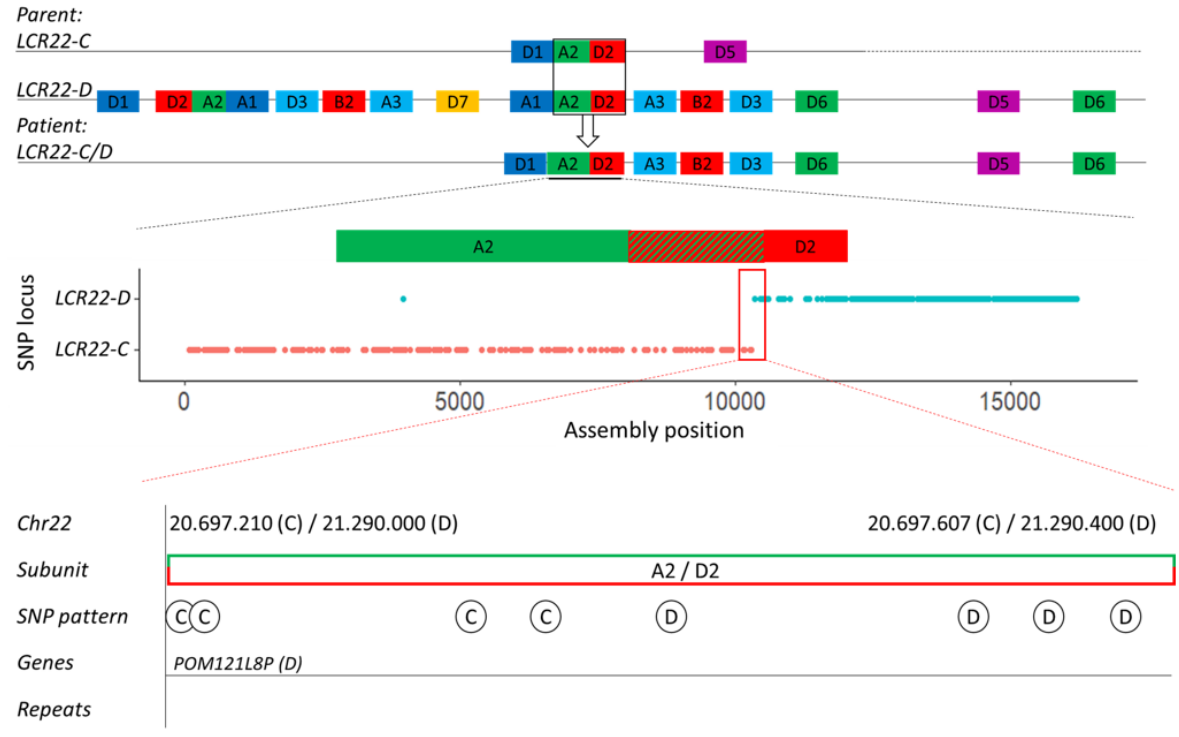


**Figure S4: Principal Component Analysis**

Principal Component Analysis clustering of our cohort (patients and parents of origin) and 1000genome individuals of African (AFR), Ad Mixed American (AMR), East Asian (EAS), European (EUR) and South Asian (SAS) populations, built using the first 2 components.

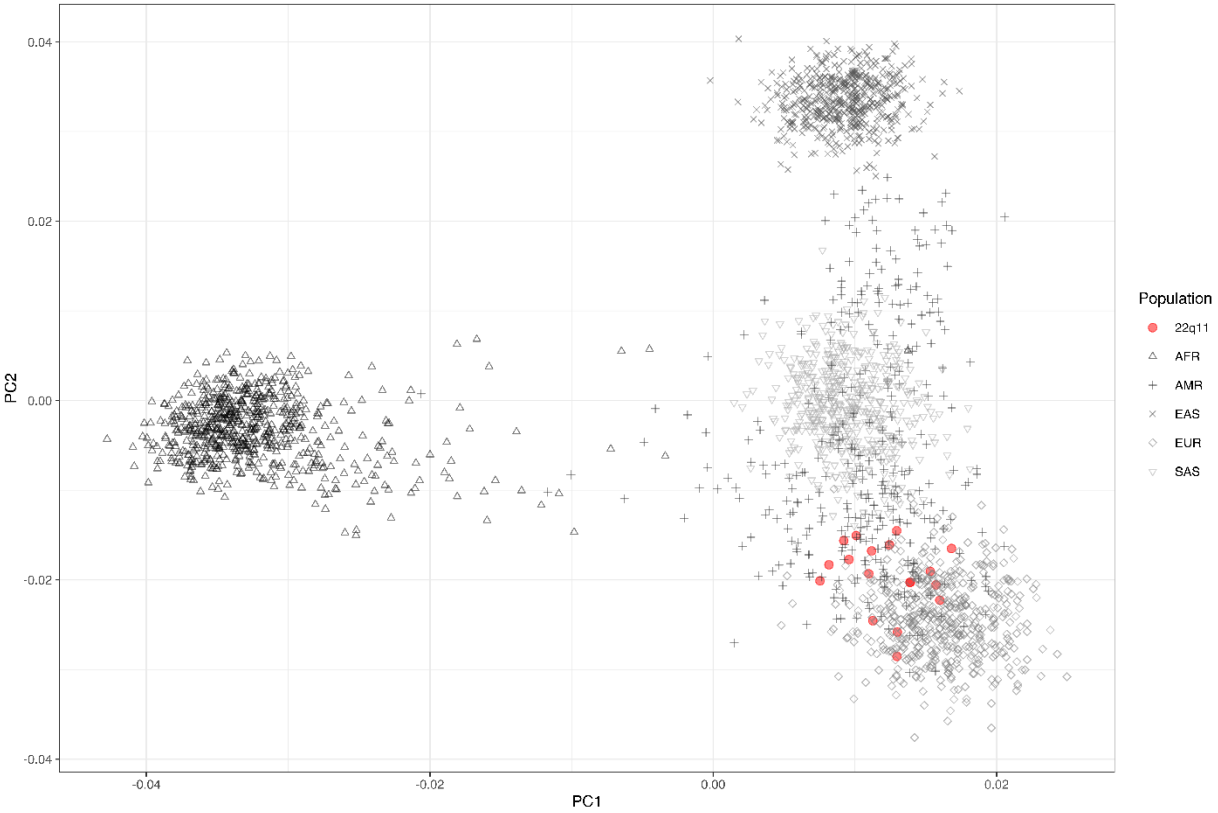


**Figure S5: Relatedness score**

The relatedness score of all pairs of samples for which genome wide long read sequencing data was available are represented as Parent-Child or Unrelated depending on the *a priori* relatedness knowledge. All duos consisting of patients only or parents only were considered as unrelated. Duos consisting of parents and patients from different families were also considered as unrelated.

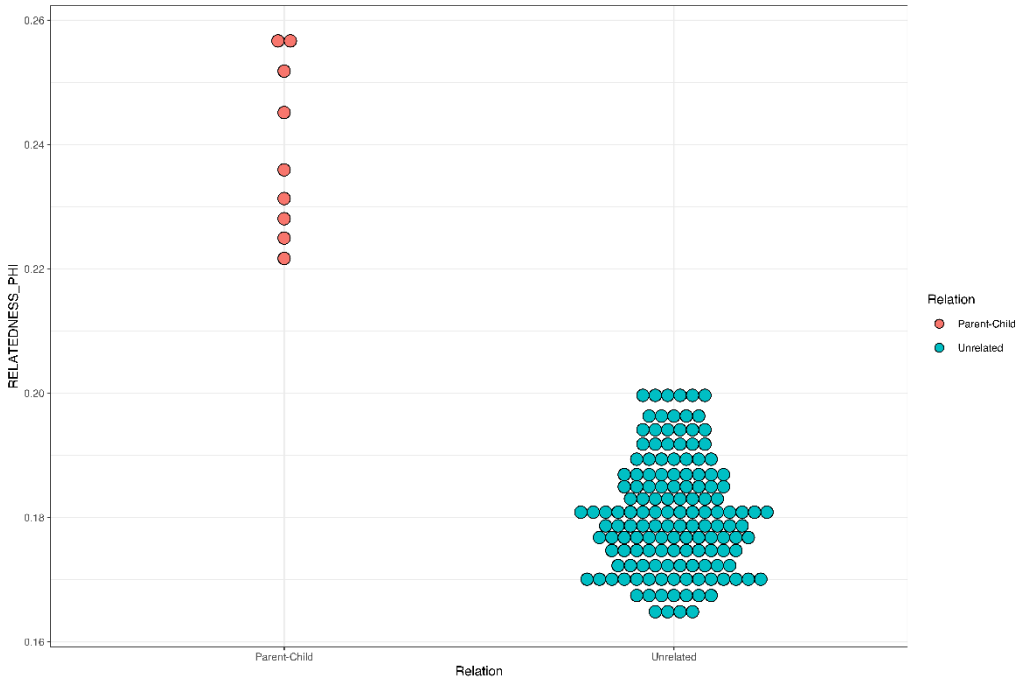


**Figure S6: Circos plot of the NOVOLoci and Verkko assemblies of the 22q11 regions**

Visualisation of the maternal and paternal assemblies of the HG002 22q11 regions produced by NOVOLoci with ONT data only, Flye using ONT data only and Verkko using ONT data only and a combination of ONT and PacBio data.


ONT
ONT
HiFi
HiFi + ONT
22q11
HG002 Paternal
NOVOLoci
Flye
Verkko
168
171
174
177
180
183
186
189
192
195
198
201
204
207
210
213