RESOLVING THE STRUCTURE OF LARGE NEUROPSYCHIATRIC COPY NUMBER VARIANTS (CNVs) WITH CRISPR-CATCH/LONG-READ SEQUENCING

Bo Zhou[1], GiWon Shin[2], Lisanne Vervoort[3], Stephanie U. Greer[2], Tanmoy Roychowdhury[4], Reenal Pattni[1], Alexej Abyzov[4], Joris R. Vermeesch[3], Hanlee P. Ji[2], Alexander E. Urban[1]

[1]Stanford University, Departments of Psychiatry and Behavioral Sciences, Genetics, and [2]Medicine
[3]Katholieke Universiteit Leuven, Department of Human Genetics (Belgium); [4]Mayo Clinic, Department of Health Sciences Research

CRISPR-Catch/Long-Read sequencing (CLR-Seq) is a novel method to analyze at base-pair resolution, for the first time, the segmental duplication regions (SegDups) that are frequently forming the boundaries of large CNVs.

About 5% of the human genome consists of SegDup region also known as low copy repeats (LCRs), which are stretches of complex sequences that are ≥95% identical to at least one other locus. However, they are very often highly homologous to multiple other SegDups. Many SegDups range into the tens of thousands of base-pairs (and even into the millions of base-pairs, as is the case for the SegDups bounding the neuropsychiatric 22q11.2 deletion region).

SegDups mediate the formation of large genomic deletions or duplications non-allelic homologous recombination that can encompass hundreds of thousands or millions of basepairs, often containing multiple genes, and that are associated with various developmental defects and in particular with neurodevelopmental, neuropsychiatric phenotypes. The complex nature of SegDups and their high degree of sequence homology makes it impossible to determine the exact extent of a given large CNV as they are impenetrable to the current genome analysis technologies. In practice this means that the SegDup regions are simply excluded from analysis, both in basic research and in clinical diagnosis.

We have been developing a CLR-Seq approach to resolve the SegDups that contain the breakpoints of large neurodevelopmental CNVs, at base-pair resolution. CRISPR-Catch/LRS isolates the SegDup sequence surrounding the breakpoint-junction of a given large CNV from the rest of the genome, by *in vitro* cutting in the unique sequence on both sides outside of the SegDups, followed by pulse-field gel-electrophoresis to obtain intact gDNA fragments of up to several hundred kb in size, which are then sequenced using long-read sequencing technology (here Oxford Nanopore Technologies, ONT).

We are using as a test case the 22q11 deletion syndrome (22q11DS), which gives a ~30x elevated risk for schizophrenia. The vast majority of 22q11DS have both endpoints located in SegDup regions LCR22A and D.

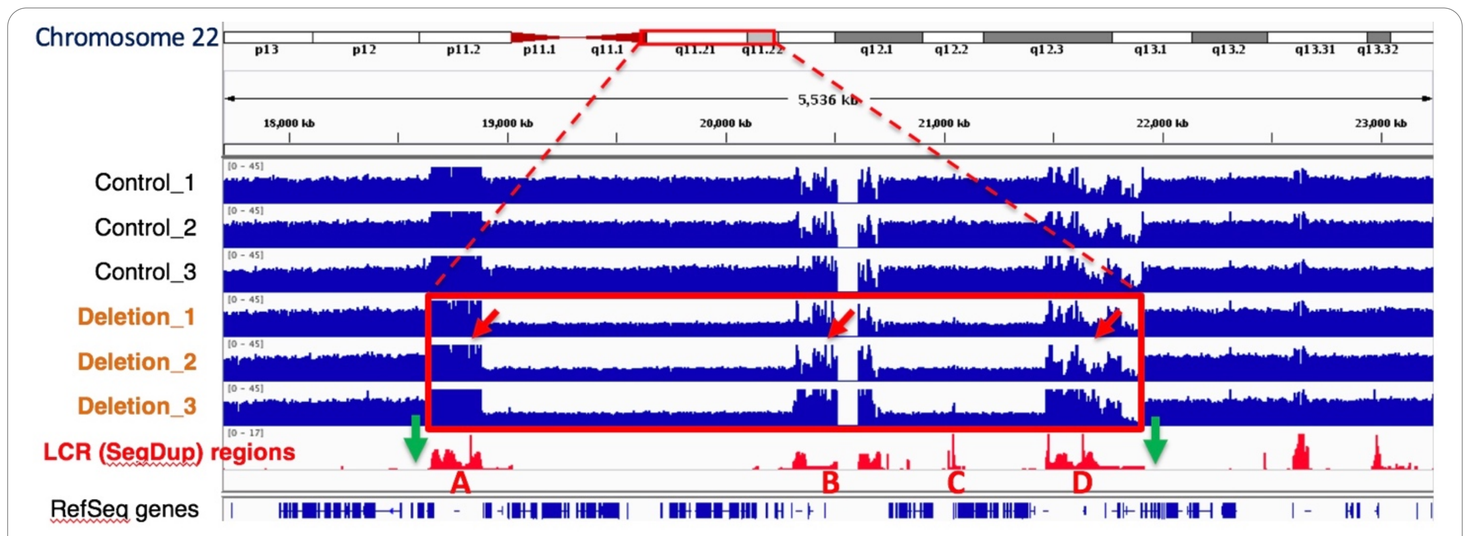

Figure 1. The LCR/SegDup regions flanking the 22q11 deletion are impenetrable even to state-of-the-art whole-genome sequencing (WGS). Dark blue signal tracks: WGS sequencing data from 3 control individuals and 3 patients with 22q11DS, plotted for depth-of-coverage. For the heterozygous 22q11DS deletion the read-depth signal drops to about half, as expected (red box). However, for the LCR/SegDup regions (marked in red in 2nd-to-bottom track), the WGS signal becomes an uninterpretable 'firestorm' (indicated by red arrows for patient Deletion_2 as an example) because of ambiguous mapping to the highly homologous LCRs. The actual breakpoints of the deletions are concealed within the regions of uninterpretable signal located at LCRs A and D. CRISPR-Catch cutting at the sites indicated by green arrows allows to physically isolate the breakpoint-junction region as a contiguous DNA fragment from the rest of the genome, making it amenable to long-read DNA sequencing and computational assembly, to be followed by computational resolution of the exact breakpoints for each patient.

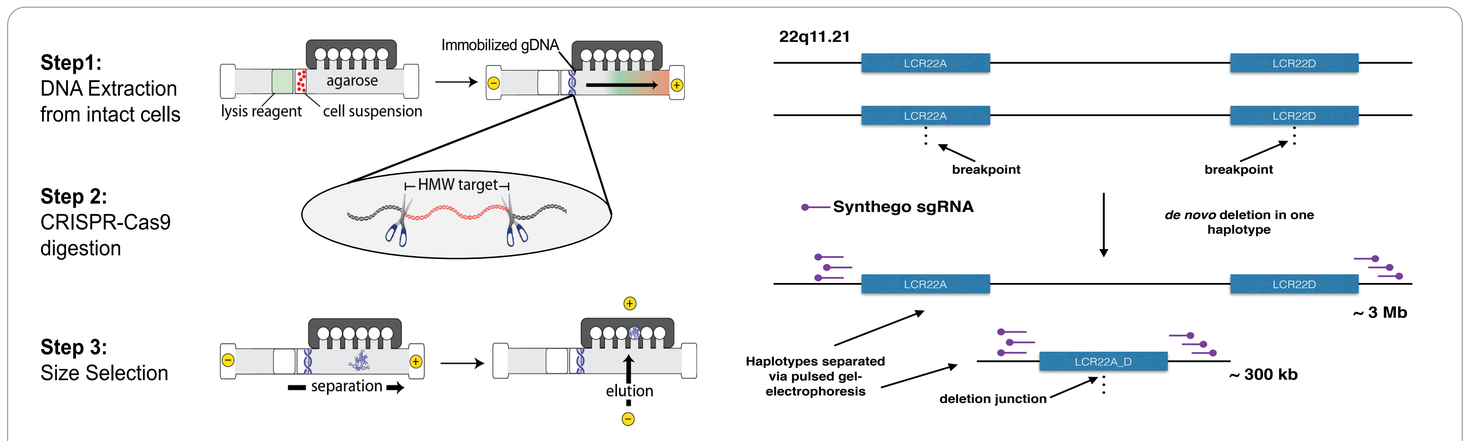

Figure 2. The principle of CRISPR-Catch and its application to the 22q11 deletion region. Left panel: the principle of CRISPR-Catch. *Step 1.:* Cells are lysed while located within an agarose cell, exposing their genomic DNA to *in vitro* CRISPR-cutting while maintaining ultra-long DNA fragment lengths (up to several hundred kbp). *Step 2.*: CRISPR-components and sgRNA guide-RNAs are added to the gel-pocket and cuts take place (i.e. at green arrows in Fig. 1, see also sgRNAs in bottom panel of this Figure, HMW is high-molecular weight, i.e. very large genomic DNA fragment). *Step 3.:* Electrophoresis using the pulse-field principle is carried out, only the (very large) DNA fragment containing the breakpoint-junction region is small enough to run into the gel and is thus separated from the rest of the genome, including from the chromosome 22 without deletion; switching the direction of the current then leads to elution of this fragment into a pocket, followed by long-read DNA sequencing. Right panel: CRISPR-Catch will cut outside the LCRs (to where Synthego sgRNAs guide the cuts, also as indicated by green arrows in Fig. 1) on both chromosomes 22. On the chromosome with deletion a fragment of several hundred kbp in size (i.e. containing region LCR22A_D which resulted from the formation of the deletion) will be generated, on the chromosome without deletion the CRISPR-generated fragment will be more than 3 Mbp in size, too large to enter the pulse-field gel.

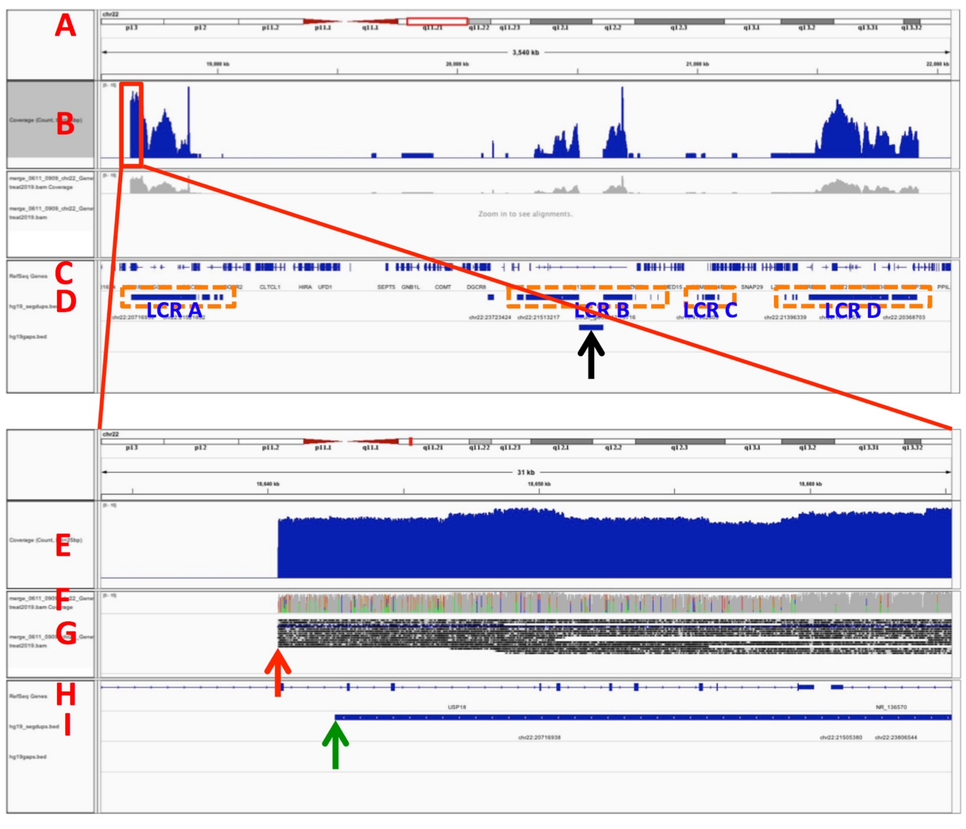

Figure 3. Preliminary data showing alignment of long-read-sequencing data after CRISPR-Catch capture of the breakpoint-junction region in a 22q11DS patient with typical 3 Mbp deletion from LCR-A to LCR-D. A.: chromosome 22; B.: zoomed in plot of the 22q11.2 deletion region showing read-depth of Oxford Nanopore reads from CRISPR-Catch fragment for 22q11.2 deletion region; C.: genes in the region (also H), including genes within SegDups; D.: SegDup clusters in the 22q11.2 deletion region, the typical deletion (as in this patient) extends from within LCR-A to within LCR-D; E.: zoomed-in view of read-depth plot from track B; F.: collapsing the Oxford Nanopore reads from track G; G.: raw Oxford Nanopore long-read sequencing reads after sequencing the CRISPR-Catch fragment that resulted from cutting in the unique sequence to the left of LCR-A (red arrow) and to the right of LCR-D (not indicated); H.: zoomed-in view of genes in the region; I.: zoomed-in view of track D showing the left end of SegDup cluster LCR-A (green arrow).

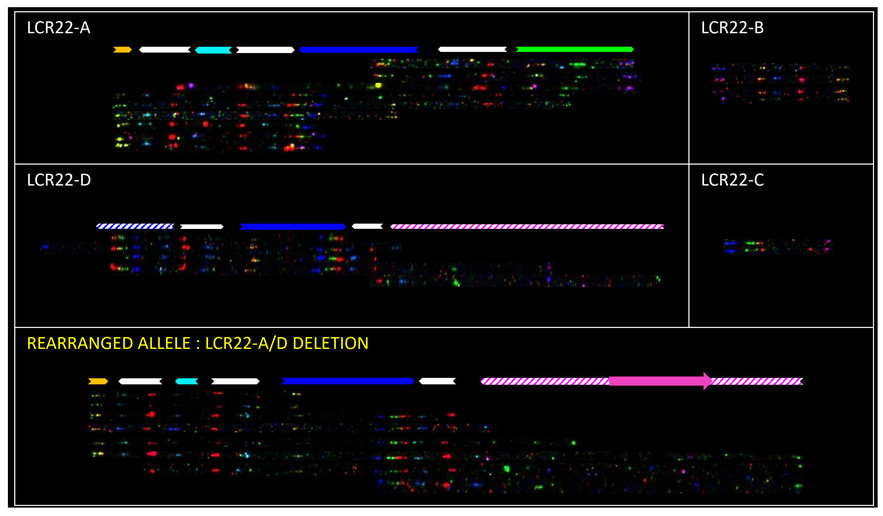

Figure 4. CRISPR-Catch/LRS-derived assembly of LCR22A/D Deletion rearrangement haplotype validated Using fiber-FISH optical mapping. Colors: FISH probes targeting specific duplicons within 22q11.2 SegDup regions. Fiber-FISH optical mapping for one individual of all LCR22 alleles A,B,C,D on the normal haplotype (without deletion) as well as of the haplotype with the LCR22 A/D deletion rearrangement (bottom panel). Sizes of each duplicon in bottom left panel.