Detection of large inversions involving the Lynch syndrome gene PMS2 using Cas9-assisted targeted sequencing.

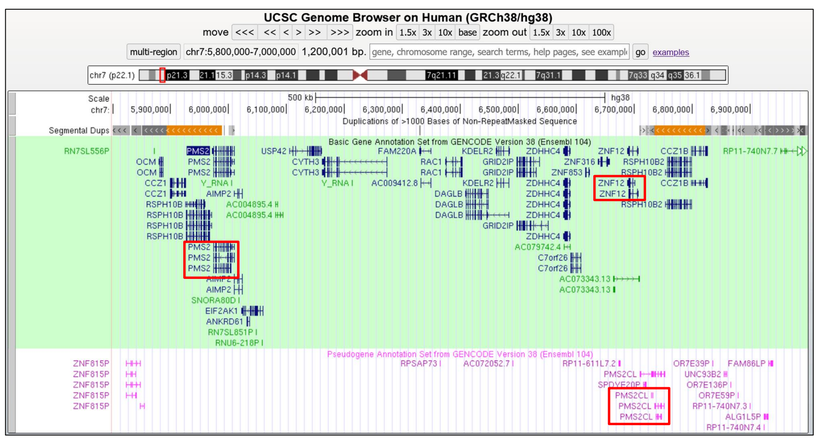
Chris Boles[1], Tina Huang[2], Feng Chen[2]

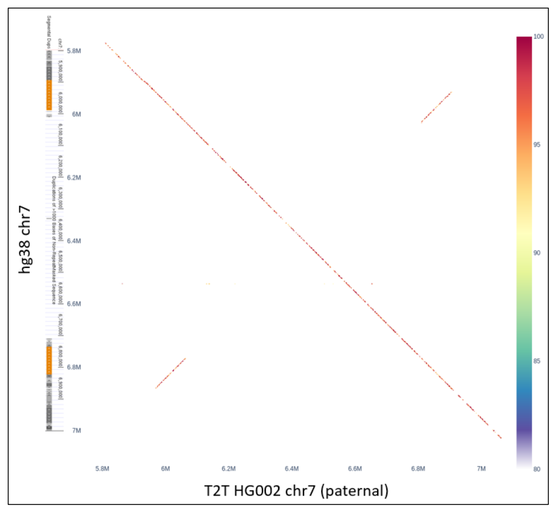
[1] Sage Science, Inc., Beverly, MA. [2] Illumina, Inc., Foster City, CA.

ABSTRACT: Deleterious alterations in the PMS2 mismatch repair gene on chromosome 7 (band p22.1) are associated with two hereditary cancer syndromes, constitutional mismatch repair deficiency (CMMRD), and Lynch syndrome (LS). Genetic tests for PMS2 alterations are complicated by the presence of 15 PMS2 pseudogenes scattered across chromosome 7, and a 100kb low copy number repeat (LCR, a.k.a. segmental duplication) sequence that covers the 3' portion of the PMS2 gene (exons 9-15). A second copy of the LCR sequence in inverted orientation is located 700kb toward the centromere and shares greater than 95% homology with the telomeric copy, including a PMS2CL pseudogene. Inversion events between the LCRs have been shown to cause PMS2 gene activation but are extremely difficult to detect by next generation sequencing methods due to the great length and sequence homology between the LCRs. We have used Cas9 endonuclease in vitro to selectively excise and electrophoretically purify a 200kb genomic DNA fragment spanning the entire telomeric LCR plus about 50kb of unique flanking sequence on both sides of the LCR – a fragment that encompasses entire unrearranged PMS2 gene. Purification of the 200kb fragment enables targeted short-read sequencing of the PMS2 locus without complications from the centromeric LCR or the PMS2 pseudogenes. To test for inversions involving the LCR regions, a different Cas9 digestion is performed using the same telomere-proximal cleavage site outside of the telomeric LCR but using a new second cleavage site in unique sequence about 50kb inside of the centromeric LCR. This second Cas9 digestion will also liberate a 200kb product from the inversion allele, but the sequencing reads will map in a very different pattern from the unrearranged allele, with high coverage of the region between the second cleavage site and the border of the centromeric LCR – a region approximately 650kb away from the location of the unrearranged PMS2 gene. We demonstrate the inversion detection method using Coriell cell lines from the Ashkenazi trio (HG002, HG003, HG004), in which the maternal haplotype inherited by the son carries the PMS2 inversion.

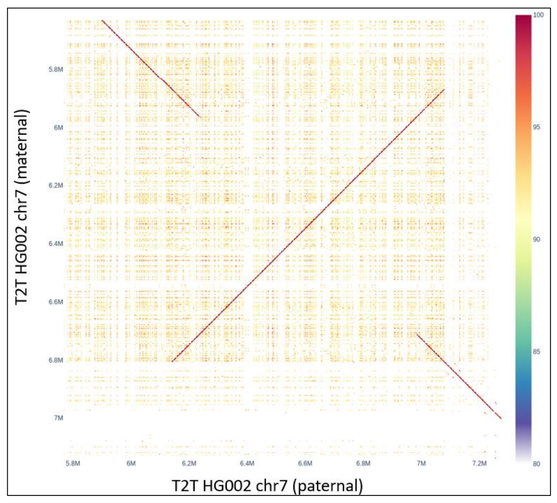
Overview of PMS2/PMS2CL region in the Ashkenzi trio (HG002, HG003, HG004)

The PMS2 gene is located approximately 6Mb from the telomeric end of the short arm of chr7. The two ~100kb LCRs are separated by about 730kb of single copy sequence. In the father (NA24149, HG003), the intervening single copy sequence is in the hg38 reference orientation in both alleles, whereas in the mother(NA24143, HG004) and son (NA24385,HG002), the single copy sequence is inverted on one chromosome, presumably due to recombination between the LCRs.





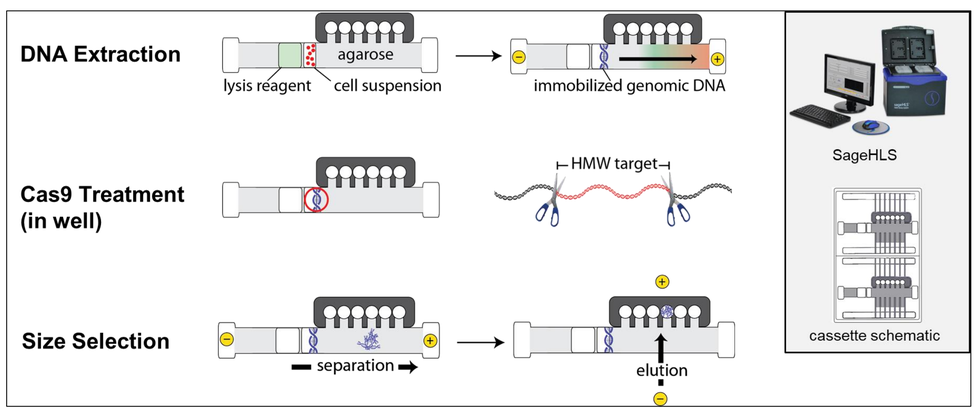
Dot plot of T2T HG002 paternal chr7 sequence against the hg38 chr7 reference. The positions of the LCRs in the hg38 reference are shaded orange on the Y axis.

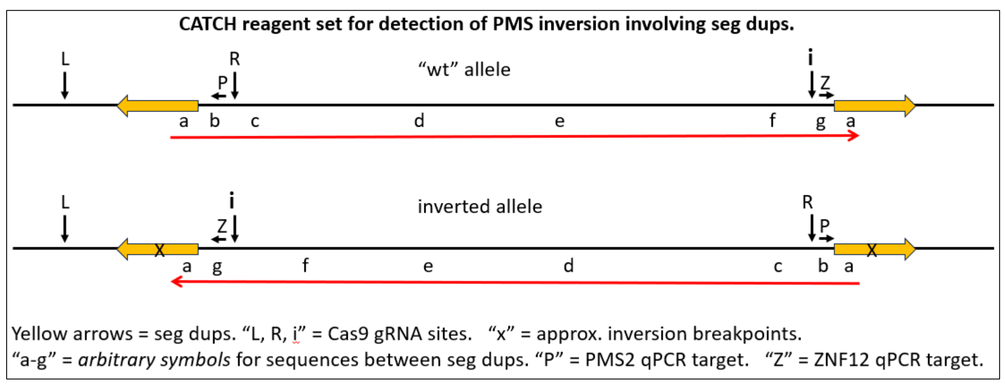


Dot plot of the two T2T HG002 chromosomes in the PMS2/PMS2CL region. The maternal allele carries a ~830kb inversion bounded by the LCRs.

CATCH long-fragment target preparation and analysis of the PMS2 region of the Ashkenazi trio

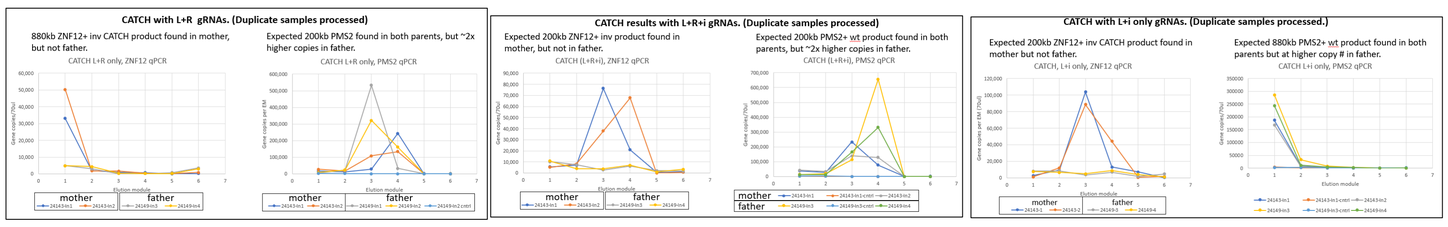
The SageHLS instrument was used for CATCH (Cas9-assisted targeting of chromosome segments; Jiang et al., 2015). Nuclei are loaded into the sample well of the Sage HLS cassette. DNA is extracted by electrophoresis of SDS through the sample well. Intact chromosomal DNA becomes entrapped in the wall of the sample well. Voltage is turned off and the DNA is digested with Cas9 complexes designed to excise large portions of the PMS2 region as shown below. The digestion products are purified by pulsed field electrophoresis and eluted from the gel in six consecutive size fractions. Targets are located and quantified by qPCR, and subsequently analyzed on Illumina sequencers.





qPCR analyses of CATCH products (parents)

(CATCH products from son (HG002) were identical to mother's: both are heterozygous for the chr7 inversion.)



Expected CATCH products from HG002

In mother and son, both wt and inv products are expected. In father, only two copies of wt fragments are expected.

CATCH with L+R gRNAs

wt 200kb PMS2+ (EM3,4)

inv 880kb ZNF12+ (EM1)

CATCH with L+i gRNAs

wt 880kb PMS2+ (EM1)

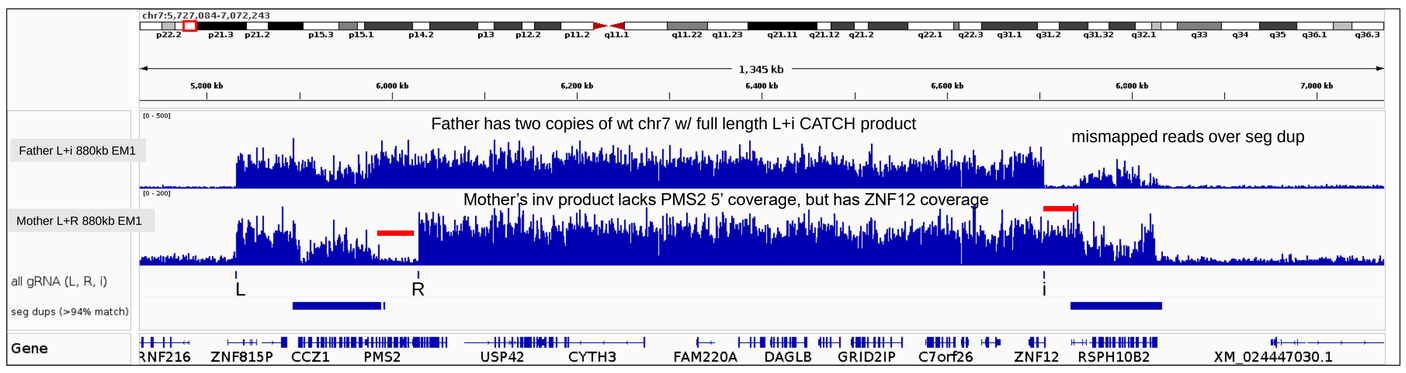
inv 200kb ZNF12+ (EM3,4)

CATCH with L+R+i gRNAs

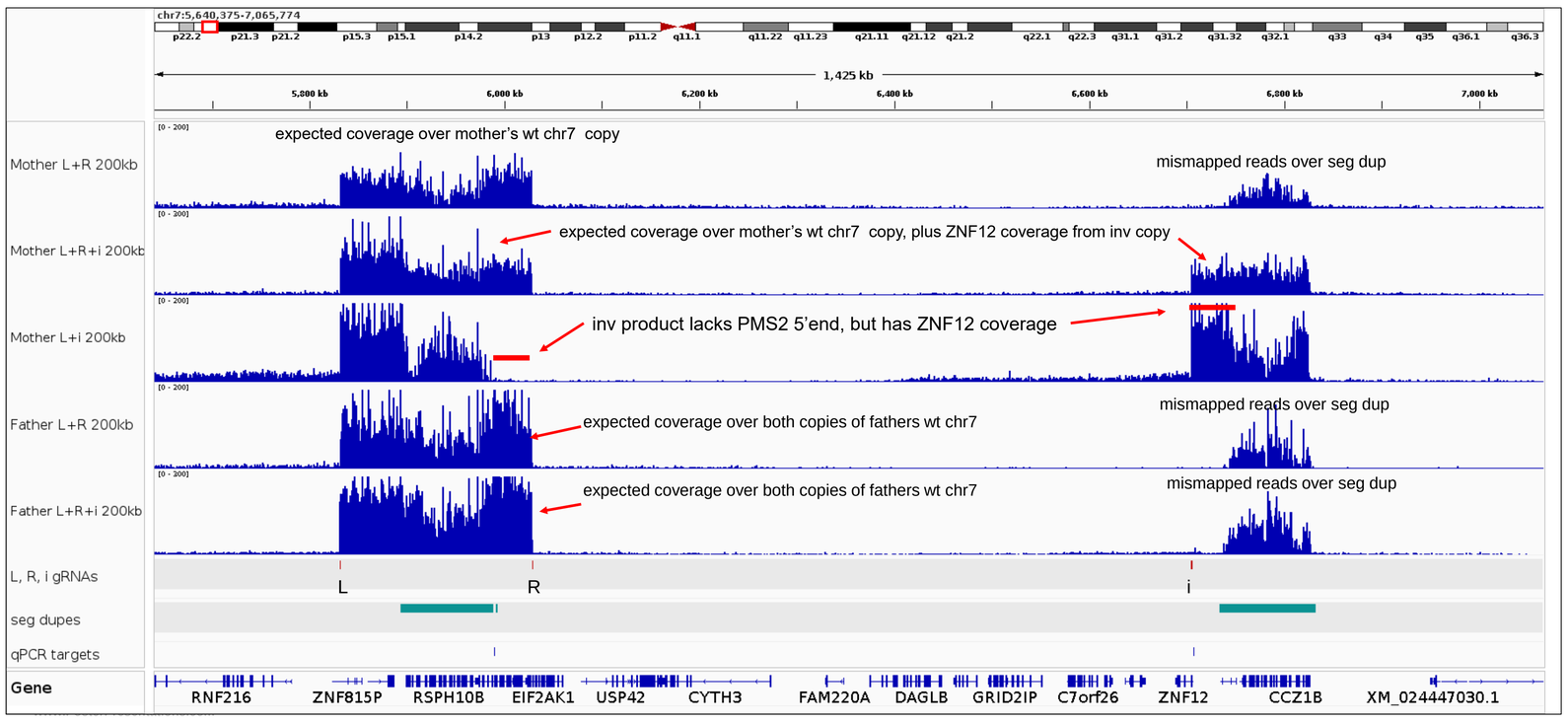
wt 200kb PMS2+ (EM3,4)

inv 200kb ZNF12+ (EM3,4)

Sequencing coverage of ~880kb elution module 1 products



Sequencing coverage of ~200kb elution module 3+4 products



Conclusions

Detection of large inversions bounded by long segmental duplications remains a difficult problem in genomic analysis. Recent methods for detecting such events include Strand-seq (Falconer et al. 2012), ultra-long Nanopore sequencing (Zhou et al. 2020), and optical mapping (Lam et al., 2012) all of which require specialized DNA extraction and handling techniques. Our CATCH-based method automates HMW DNA extraction from nuclei (or whole cells) and Cas9 digestion steps. Approximate size of digestion products is directly obtained from simple qPCR analyses of the HLS-CATCH elution products, and short-read Illumina sequencing gives a detailed view of their sequence composition. Our approach is readily applicable to other recurrent translocations using customized gRNAs and electrophoresis programs. The combination of custom genome cleavage, product sizing, and analysis of genome coverage from short-read Illumina sequencing make our method a relatively simple and cost-effective method to analyze such large inversions.

Acknowledgments: We thank Tiffany Truong and YuJue Chen for library prep during the initial stages of this work. We thank Robin Bombardi for sequencing and bioinformatic analyses.

References:

Jiang, et al. 2015. Nat Comms 6:8101. DOI:10.1038/ncomms9101.
Falconer, et al. 2012. Nat Meth 9:1107. DOI:10.1038/nmeth.2206.
Zhou, et al. 2020. bioRxiv DOI:10.1101/2020.10.23.349621.
Lam, et al. 2012. Nat Biotech 30:771. DOI:10.1038.nbt.2303.