

Abstract

Determination of unique individual haplotypes is an essential first step toward understanding how identical genotypes having different phases lead to different biological interpretations of function, phenotype, and disease. Genome-wide methods for identifying individual genetic variation have been limited in their ability to acquire phased, extended, and complete genomic sequences that are long enough to assemble haplotypes with high confidence. We explore a recombineering approach for isolation and sequencing of a tiling of targeted fosmids to capture interesting regions from human genome. Each individual fosmid contains large genomic fragments (~35 kb) that are sequenced with long-read SMRT® technology to generate contiguous long reads. These long reads can be easily *de novo* assembled for targeted haplotype resolution within an individual's genomes. The P5-C3 chemistry for SMRT Sequencing generated contiguous, full-length fosmid sequences of 30 to 40 kb in a single read, allowing assembly of resolved haplotypes with minimal data processing. The phase preserved in fosmid clones spanned at least two heterozygous variant loci, providing the essential detail of precise haplotype structures. We show complete assembly of haplotypes for various targeted loci, including the complex haplotypes of the KIR locus (~150 to 200 kb) and conserved extended haplotypes (CEHs) of the MHC region. This method is easily applicable to other regions of the human genome, as well as other genomes.

Complex Organization of KIR Haplotypes

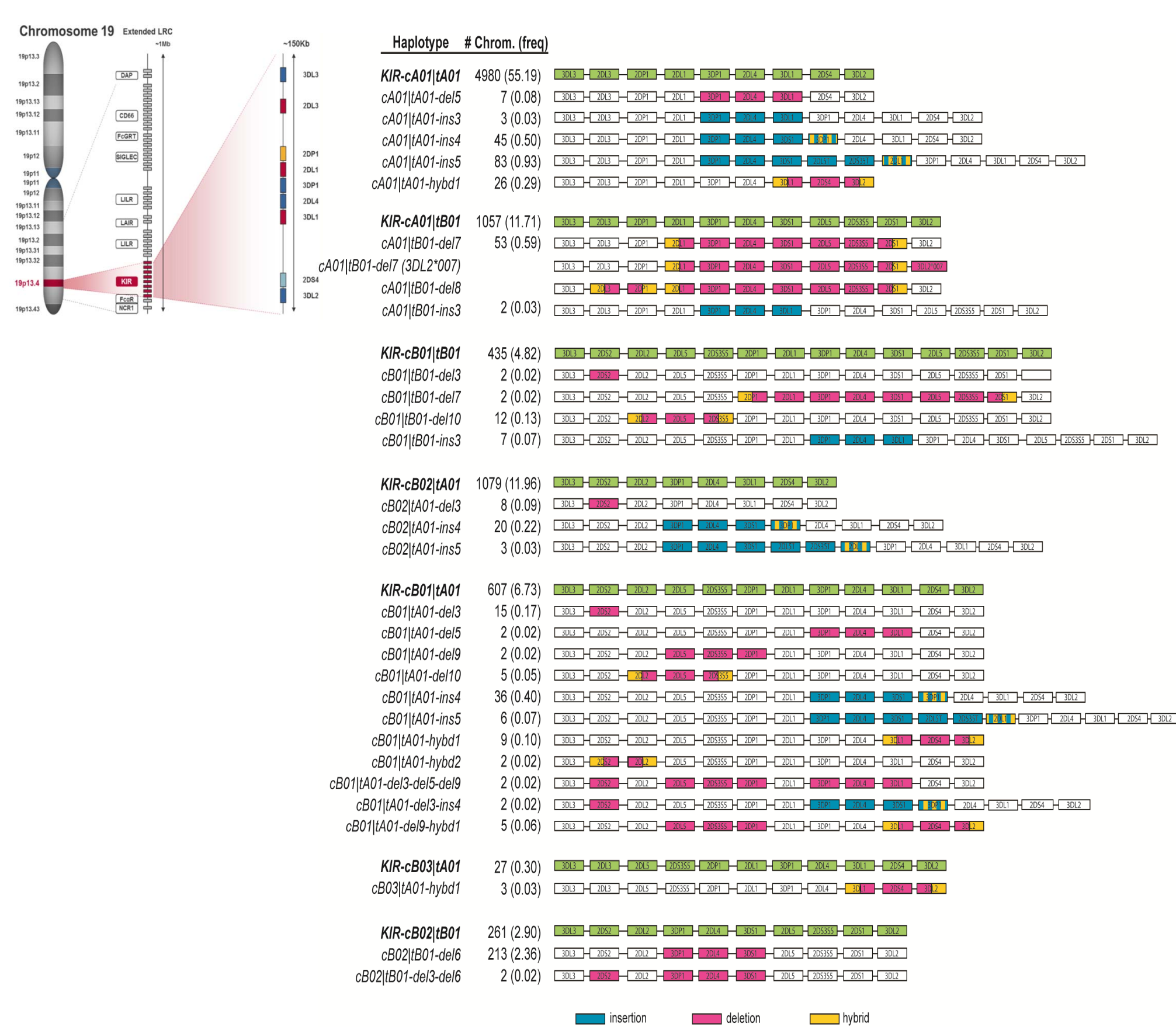


Figure 1: Polymorphic Gene Content of KIR Haplotypes: Gene Deletion, Insertions, and Hybridizations (CNVs)

Method

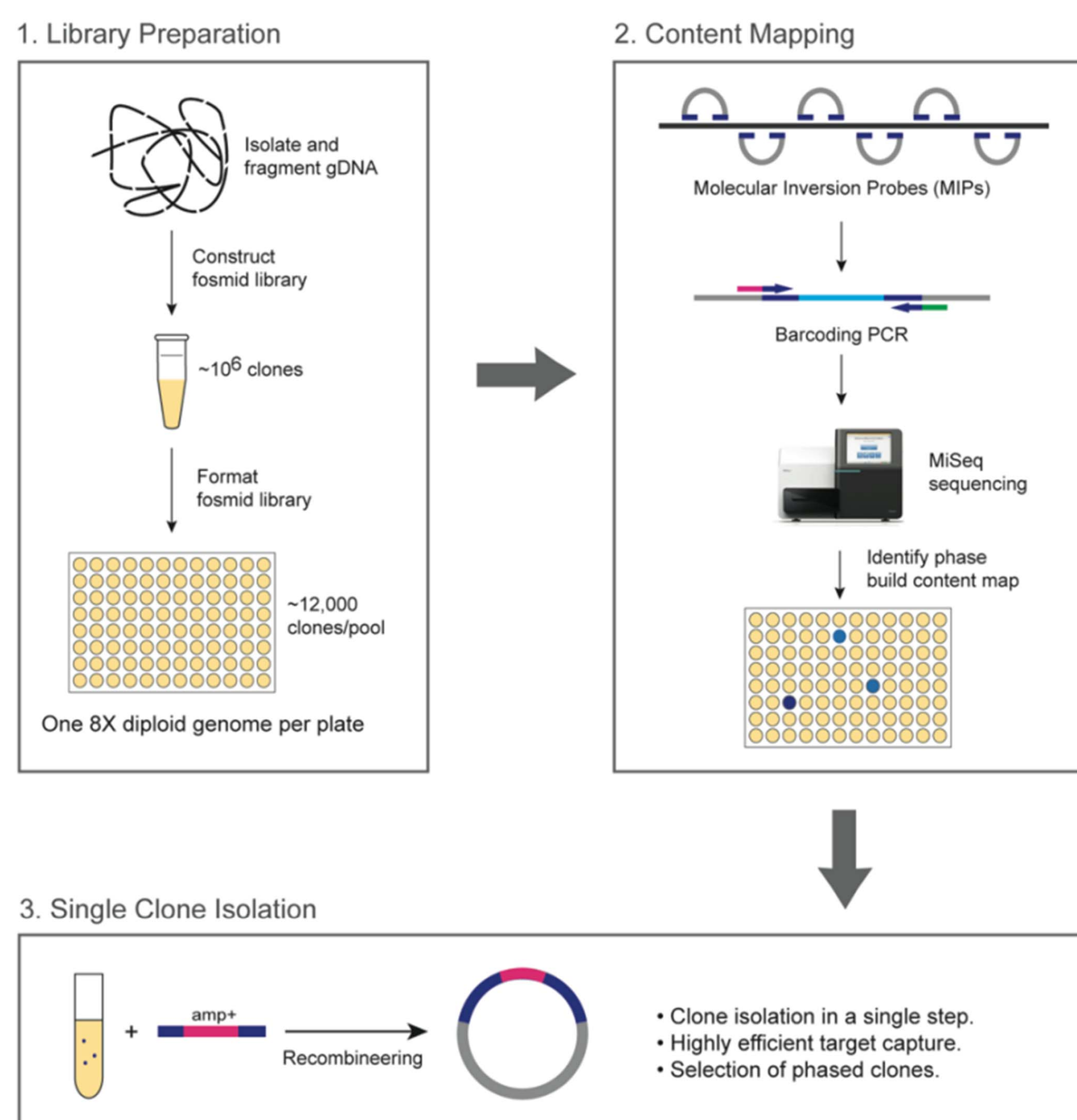


Figure 2: Recombineering approach for target enrichment of fosmids to capture haplotypes of interesting loci from a human genome in a tiling approach.



Figure 3: Workflow for SMRT® Sequencing of a full-length fosmid library preparation from a pool of fosmids belonging to a single haplotype or both haplotypes enriched from targeted regions of interest from genomic DNA and an automated DNA analysis pipeline.

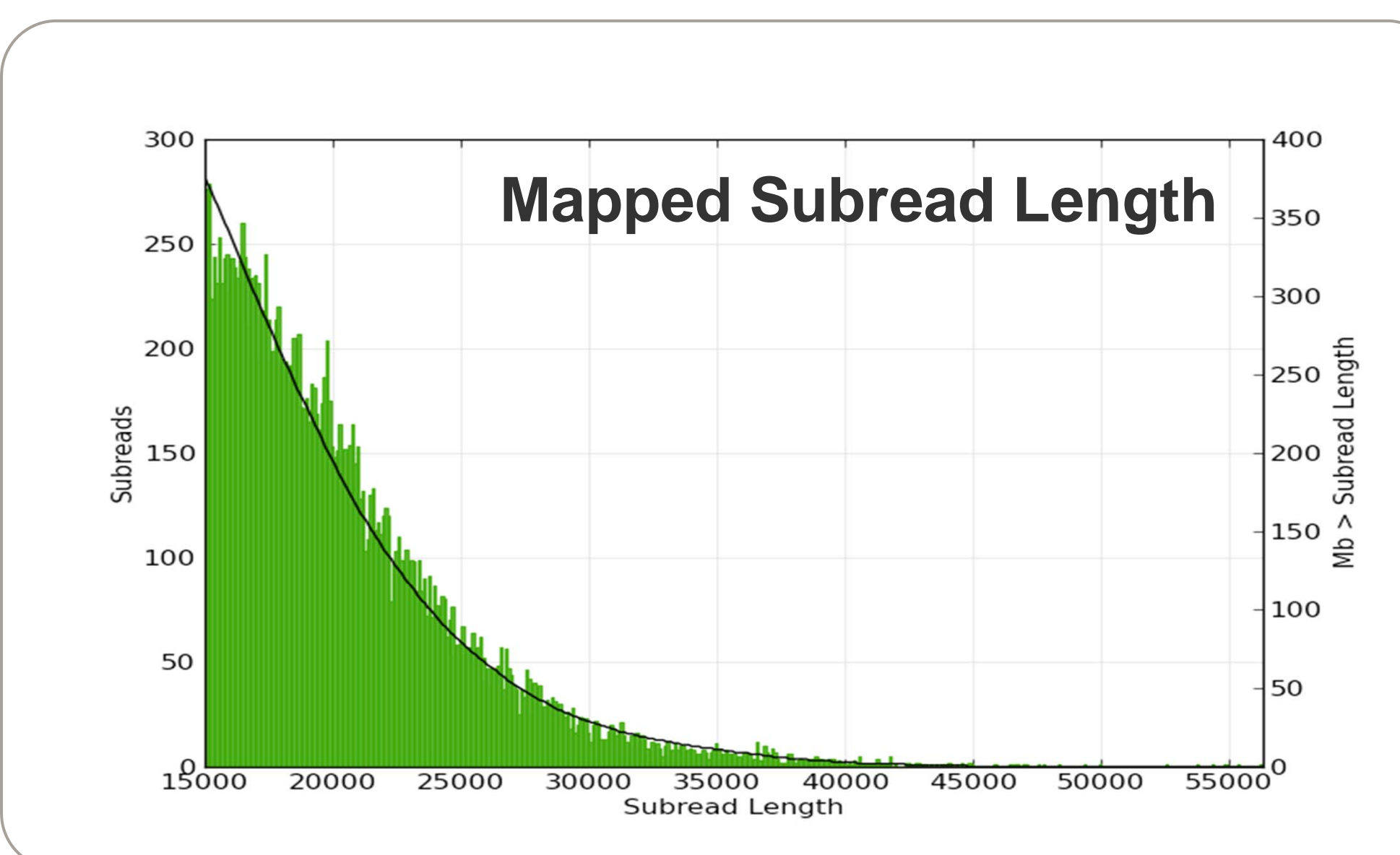


Figure 4: Mapped subread length of reads >15 kb in a single SMRT Cell® from a size-selected full-length fosmid library.

Results

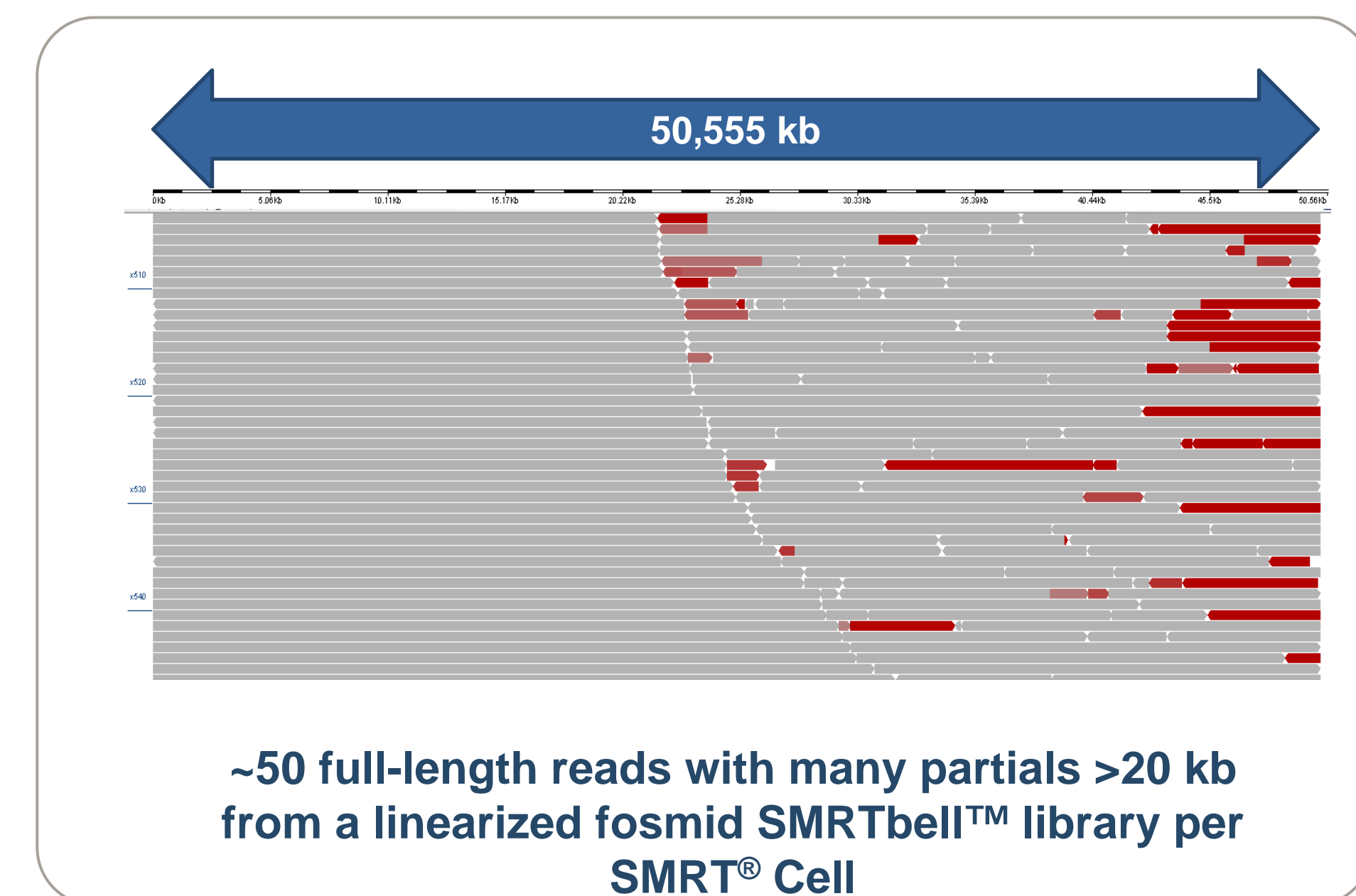


Figure 5: Ultra-long ~50-kb read lengths essential for assembly of complex haplotypes achieved by target enrichment and sequencing of full-length fosmids, using P5-C3 chemistry and 6-hour movies.

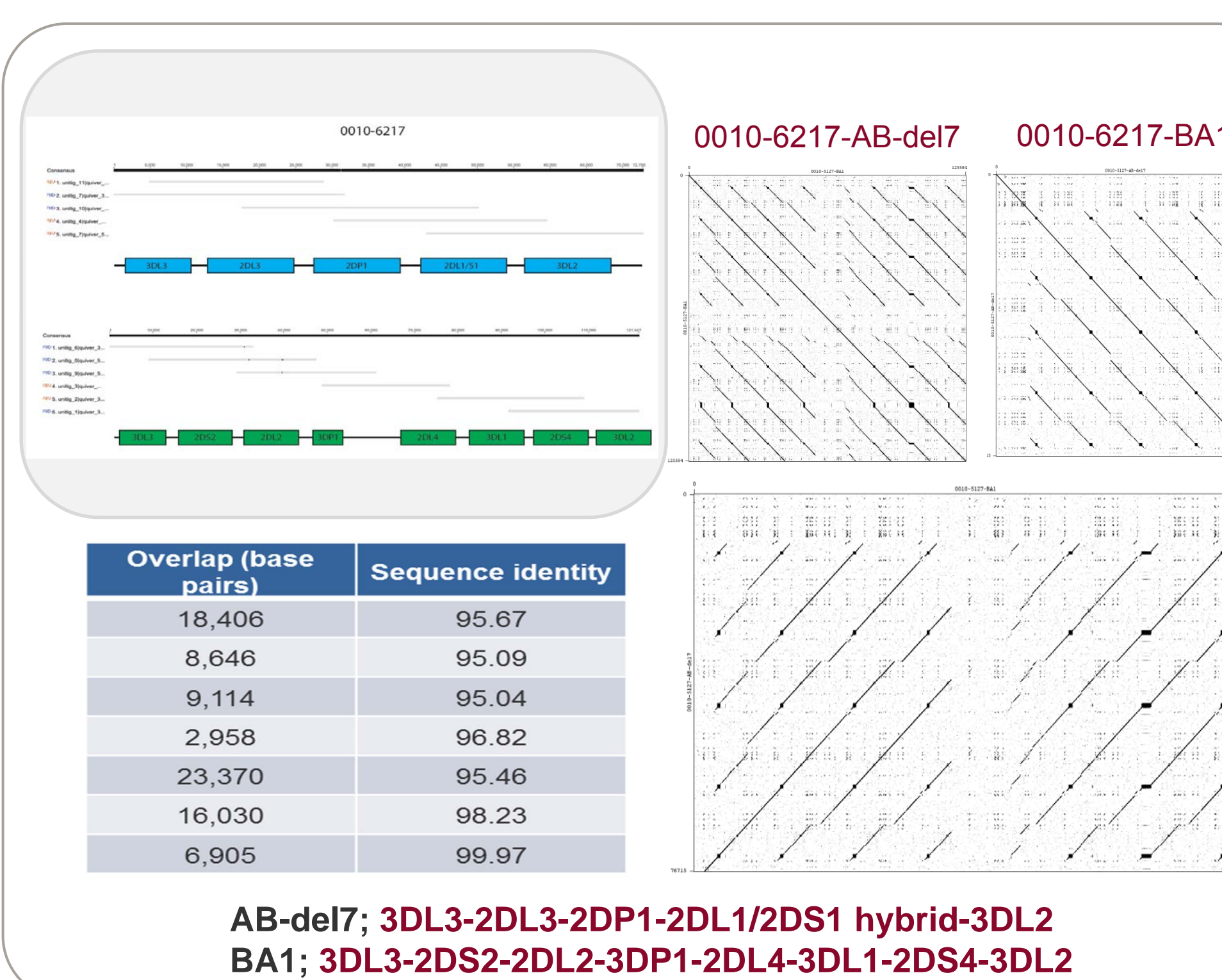


Figure 6: Haploid assembly of KIR-region-enriched fosmids for both haplotypes from human sample.

KIR Haplotype Sample	Length bp	Contigs	Coverage of >35 kb Subreads	Sample Type
KIR_G085-BA1	149,678	1	1,011	Haploid
KIR_G085-A	142,599	1	256	Haploid
KIR_LUCE-B_del	136,740	1	71	Haploid
KIR_FH05-A	155,305	1	28	Haploid
0071_9105_9-Hap2	142,664	1	880	Haploid
5029_6000_8 Haploid A	142,037	1	219	Haploid
5029_6000_8 Haploid BA2 mu	152,007	1	134	Haploid
5057_0197_9	141,674	1	209	Diploid (Known Homozygous)
0010_6217_3	76,716 125,585	2	278	Diploid (Heterozygous)
0116_9149_4	106,087 91,032	2	186	Diploid (Heterozygous)

Figure 7: KIR haplotypes assembled using full-length fosmid sequencing.

Conclusion

- PacBio's ultra-long reads provide a way for simultaneous genotyping and haplotyping of complex KIR regions.
- The recombineering approach for target enrichment followed by full-length fosmid sequencing is well-suited for haplotype and extended haplotype sequencing.

Acknowledgements

The authors thank Cheryl Heiner, Brett Bowman and Lawrence Hon for helpful suggestions during this work.

