Resolving KIR Genotypes and Haplotypes Simultaneously Using Single Molecule, Real-Time Sequencing

Ekholm, J.M.[1], Pyo, C.W.[2], Eng, K.[1], Hall, R.[1], Vierra-Green, C.[4], Spellman, S.[4], Roe, D.[3] , Geraghty, D.E.[2], Maiers, M.[3], Ranade, S[1]

[1]PacBio, 1380 Willow Road, Menlo Park, CA 94025, [2]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [3]National Marrow Donor Program, Minneapolis, MN USA, [4]Center for International Blood and Marrow Transplant Research, Minneapolis, MN USA

Introduction

The killer immunoglobulin-like receptors (KIR) genes belong to the immunoglobulin superfamily and are widely studied due to the critical role they play in coordinating the innate immune response to infection and disease. KIR gene clusters are found on chromosome 19q13.4, in a span of ~150 kb region. The content of these clusters varies in number and type of gene alleles based on linkage disequilibrium and determines each haplotype as A (inhibitory) or B (activating) [1)]. Thus resolving individual genotypes of members of the KIR gene clusters, within and between the haplotypes of a given individual is important for biological interpretations of its function, phenotype, and disease.

Current genome-wide analysis methods or PCR based approaches for genotyping KIR genes in population studies, have been limited in their ability to acquire phased, extended, and complete genomic sequences that are long enough to assemble haplotypes with high confidence. Highly accurate, contiguous, long reads, like those generated by SMRT Sequencing, when combined with target-enrichment protocols, provide a straightforward strategy for generating complete *de novo* assembled KIR haplotypes. We have explored two different methods to capture the KIR region; one applying the use of fosmid clones and one using Nimblegen capture.

Complex Organization of KIR Haplotypes

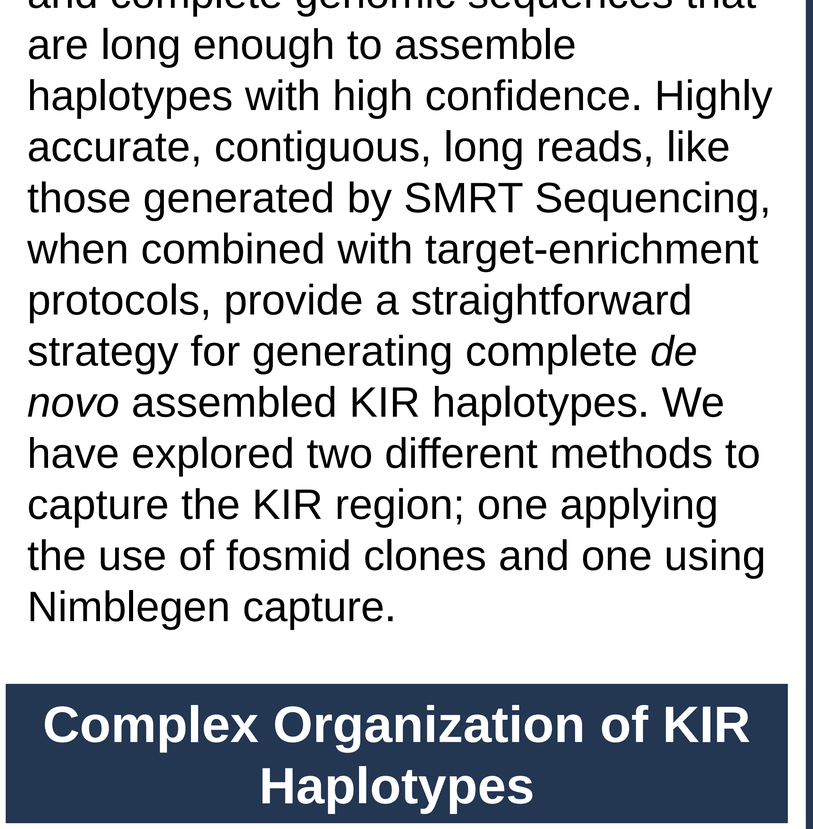

Figure 1. Polymorphic Gene Content of KIR Haplotypes: Gene Deletion, Insertions, and Hybridizations (CNVs)

References

1) Uhrberg M, Valiante N M, Shum B P, Shilling H G, Lienert-Weidenbach K, Corliss B, Tyan D, Lanier L L, Parham P. Human diversity in killer cell inhibitory receptor genes. Immunity. 1997;7:753. [PubMed: 9430221]

Method I: Fosmid Sequencing

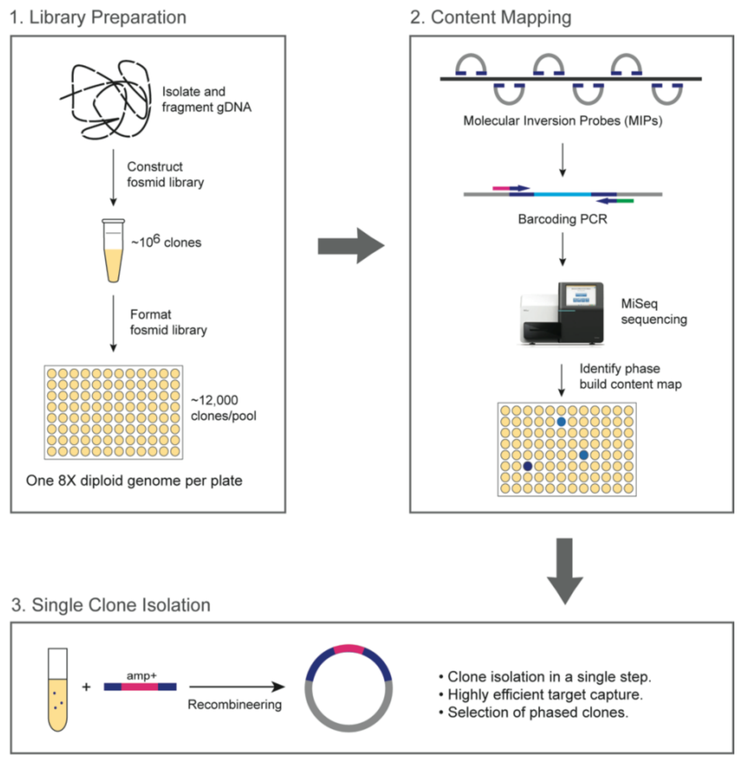

Figure 2. Recombineering approach for target enrichment of fosmids to capture haplotypes of interesting loci from a human genome in a tiling approach.

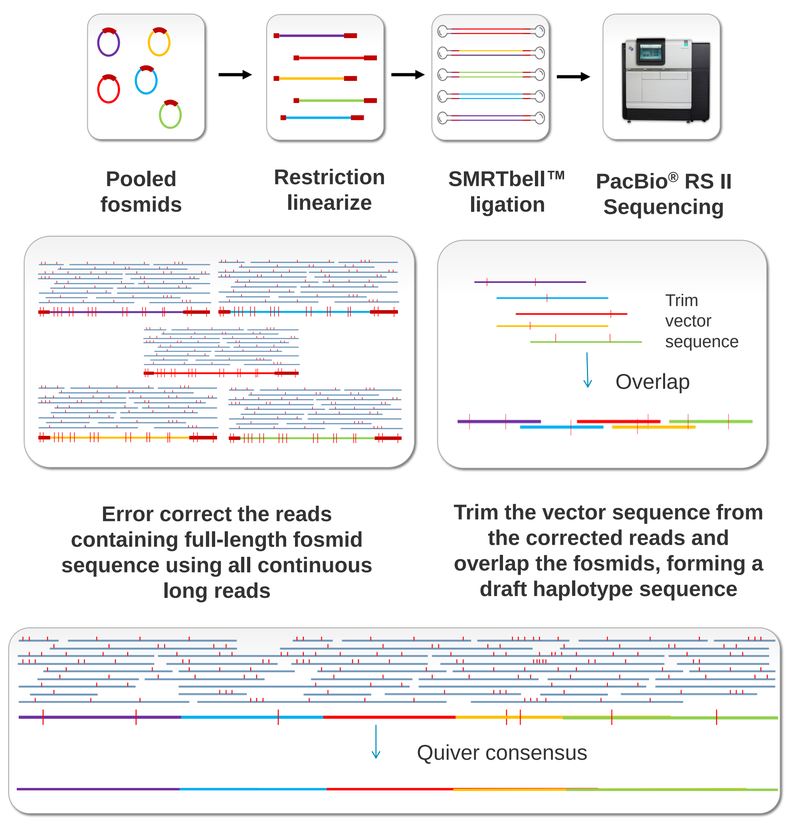

Figure 3. Workflow for SMRT® Sequencing of a full-length fosmid library preparation from a pool of fosmids belonging to a single haplotype or both haplotypes enriched from targeted regions of interest from genomic DNA and an automated DNA analysis pipeline.

Figure 4. Recombineering-based method for isolating a tiling of targeted fosmids (~35 kb – 50 kb)[3] was combined with SMRT Sequencing. 16 haplotypes assembled from 8 individuals (available in NCBI). Partial reference, available for homozygous sample 0071-9105 was 100% concordant with PacBio assembly.

Method II: NimbleGen SeqCap Target Enrichment

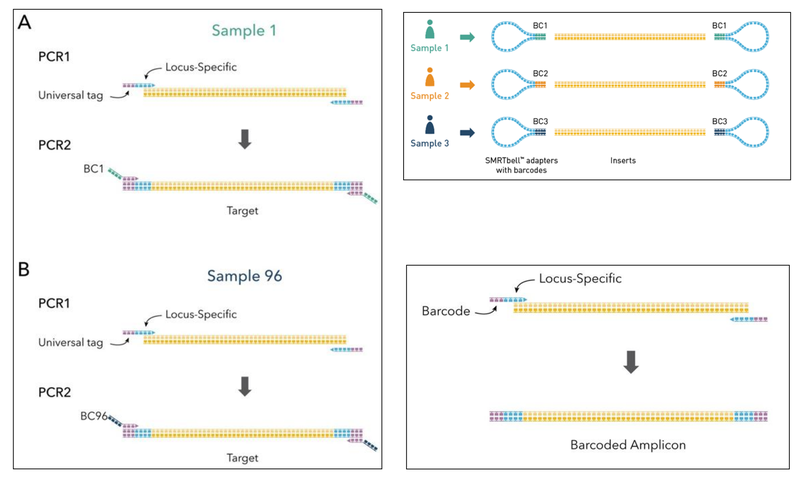

Figure 5. Barcoding options for targeted sequencing. *www.pacb.com/wp-content/uploads/2015/09/ProductNote-Barcoded-Adapters-Barcoded-Universal-Primers.pdf*

A. Barcoded Universal Primers: Barcode can be incorporated into the amplicon via a two-step tailed primer approach. Barcodes are commercially available from PacBio.

B. Barcoded Adapters: Barcodes are incorporated during ligation with barcoded adapters. Barcodes are commercially available from PacBio.

C. Locus-specific primers tailed with barcodes. Primers may be ordered from any oligo synthesis providers. The first 96 barcodes out of 384 sequences are available: *www.pacb.com/wp-content/uploads/PacBio-PCR-Primer-Barcodes- 0001-to-0096-IDT-Template.xlsx*.

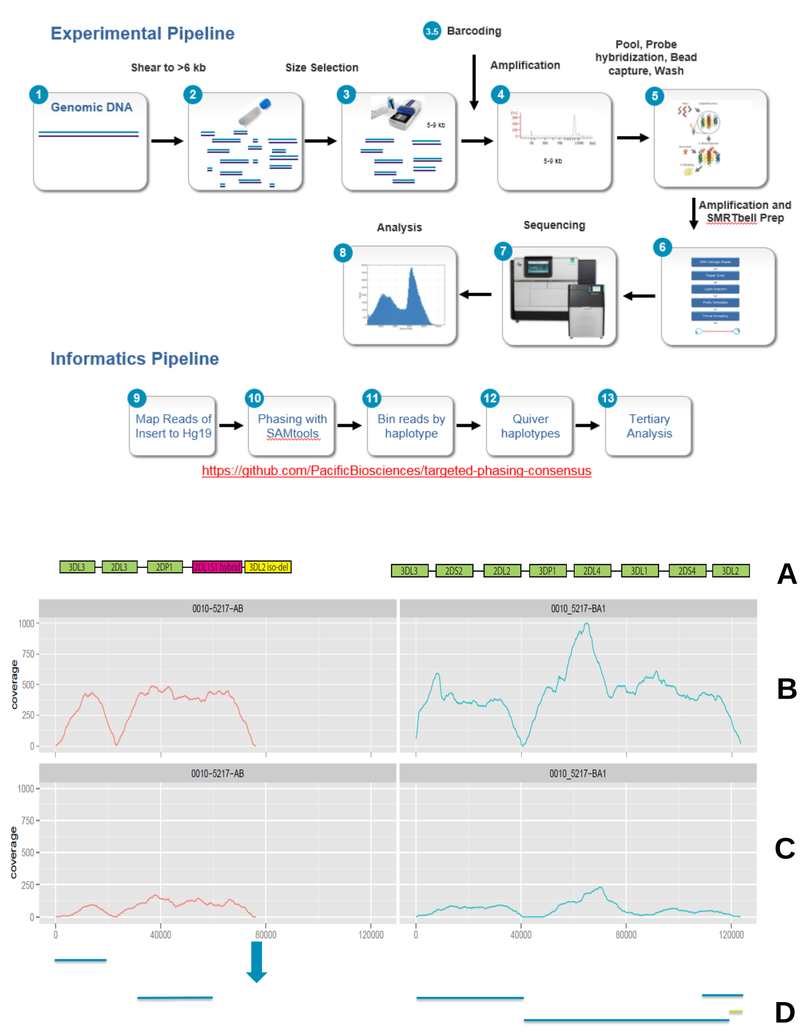

Figure 6. **A.** Known Haplotypes

B. Coverage showing alignment of captured reads >8.5 kb against the known reference haplotypes

C. Coverage of >8.5 kb reads against known haplotypes when data is aligned against all known full-length KIR sequences

D. *De novo* assembled phased sequence, red indicates assembly error that results in a contig of mixed phase

Conclusions

- SMRT Sequencing generated highly accurate long reads necessary for simultaneous genotyping and haplotyping of complex KIR regions
- SMRT Sequencing of target-enriched fosmids provided >35-50 kb contiguous sequences for improving/ establishing reference database with imputation-free information
- NimbleGen SeqCap, a scalable alternative, validated known KIR haplotypes of a known sample in a reference guided assembly of 5-8 kb enriched DNA