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



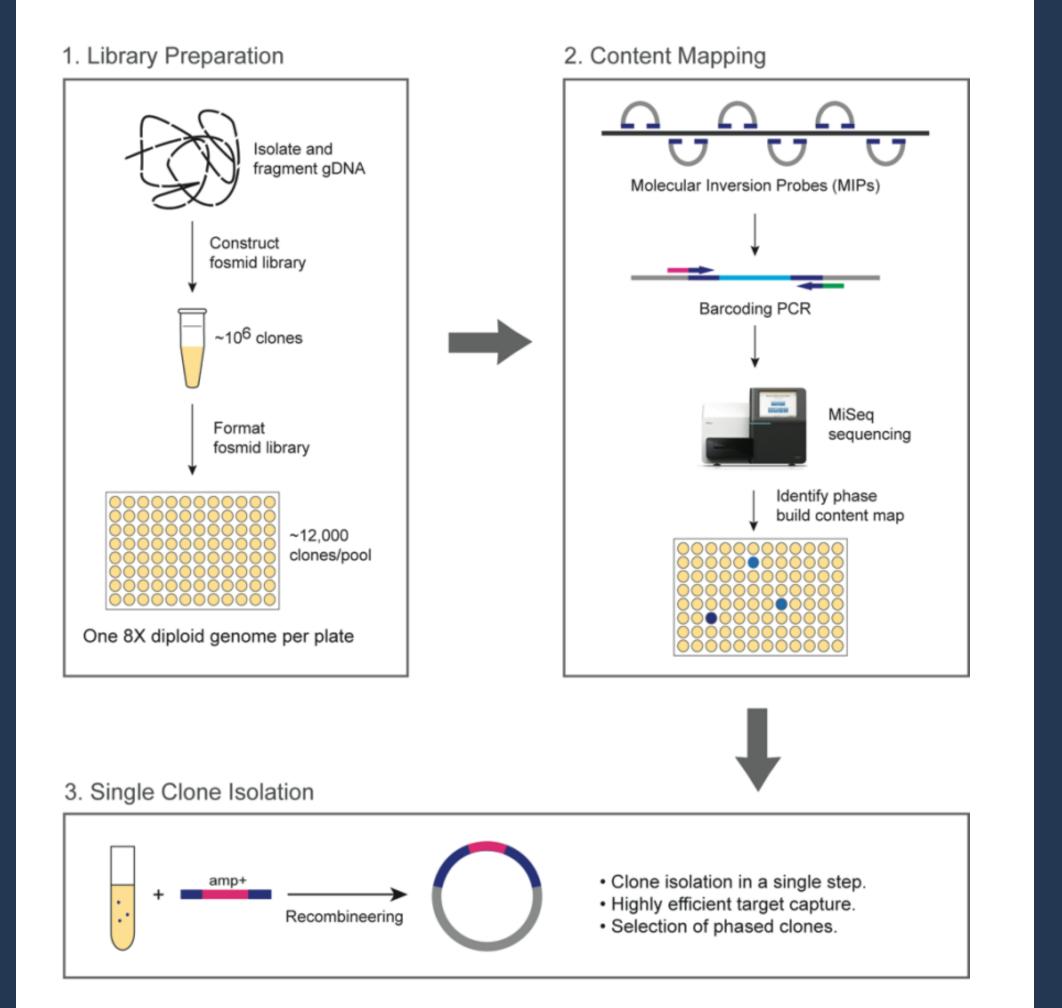
Ekholm, J.M.<sup>1</sup>, Pyo, C.W.<sup>2</sup>, Eng, K.<sup>1</sup>, Hall, R.<sup>1</sup>, Vierra-Green, C.<sup>4</sup>, Spellman, S.<sup>4</sup>, Roe, D.<sup>3</sup>, Geraghty, D.E.<sup>2</sup>, Maiers, M.<sup>3</sup>, Ranade, S<sup>1</sup> <sup>1</sup>PacBio, 1380 Willow Road, Menlo Park, CA 94025, <sup>2</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, <sup>3</sup>National Marrow Donor Program, Minneapolis, MN USA, <sup>4</sup>Center for International Blood and Marrow Transplant Research, Minneapolis, MN USA



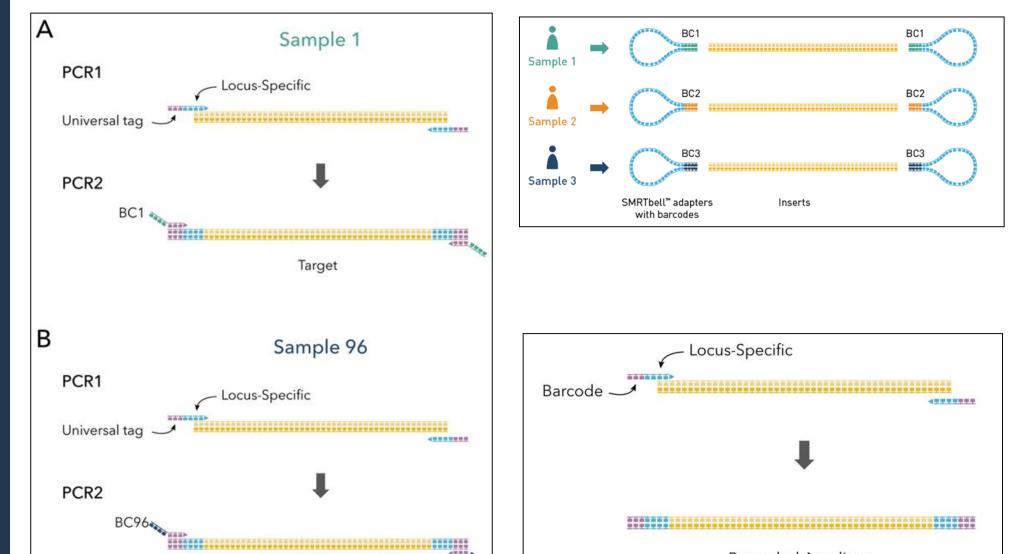
### Introduction

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)<sup>1)</sup>. 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.

# Method I: Fosmid Sequencing



## Method II: NimbleGen SeqCap Target Enrichment



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

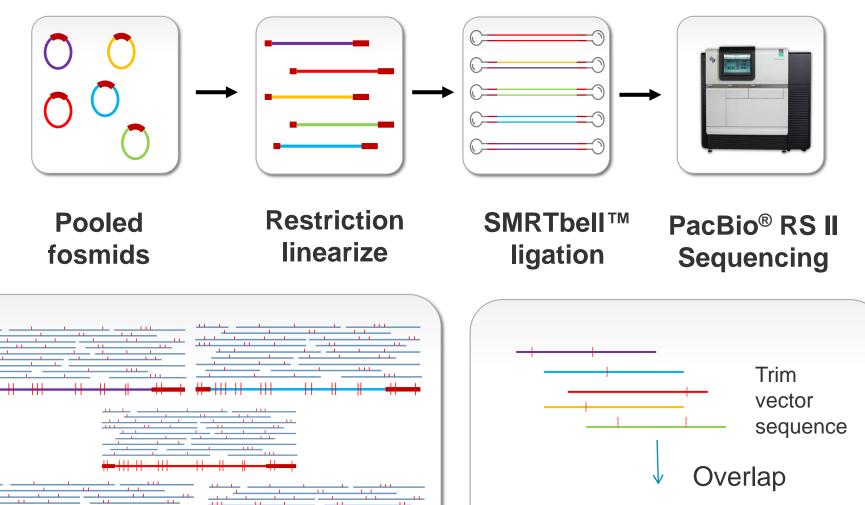
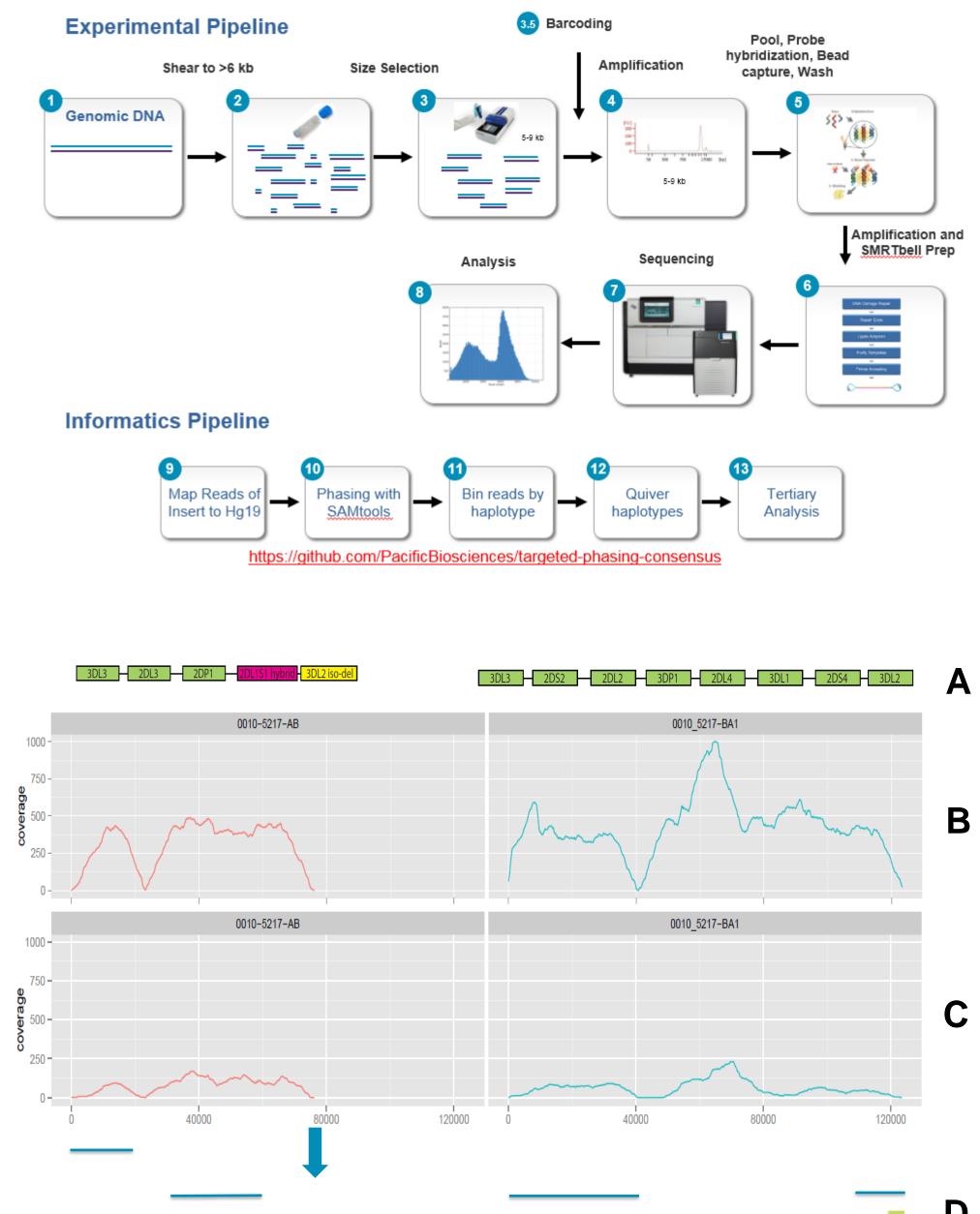


Figure 5. Barcoding options for targeted sequencing. <u>www.pacb.com/wp-content/uploads/2015/09/ProductNote-</u> <u>Barcoded-Adapters-Barcoded-Universal-Primers.pdf</u>

Target

- 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: <u>www.pacb.com/wp-content/uploads/PacBio-PCR-</u> <u>Primer-Barcodes- 0001-to-0096-IDT-Template.xlsx</u>.



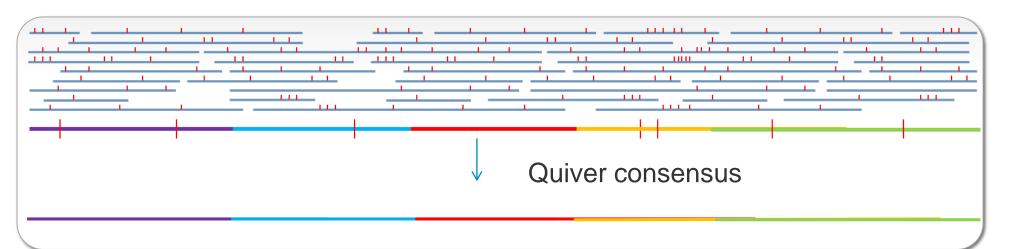
## Complex Organization of KIR Haplotypes

	ome 19 Extended LF ~1N	~150Kb	 Haplotype	# Chrom. (freq)	
13.3		3DL3	KIR-cA01 tA01	4980 (55.19)	
3.13			cA01 tA01-del5	7 (0.08)	
3.12	CD66	 2DL3	cA01 tA01-ins3	3 (0.03)	
3.11	DAP CD55 FCGRT SIGLEC		cA01 tA01-ins4	45 (0.50)	
12	SIGLEC E	2DP1 2DL1	cA01 tA01-ins5	83 (0.93)	
p11 p11 p12	/ T	3DP1	cA01 tA01-hybd1	26 (0.29)	
s.11 /		2DL4			
8.12 / 8.13		3DL1	KIR-cA01 tB01	1057 (11.71)	3DL3 - 2DL3 - 2DP1 - 2DL1 - 3DP1 - 2DL4 - 3DS1 - 2DL5 - 2DS355 - 2DS1 - 3DL2
3.2 3.31			cA01 tB01-del7	53 (0.59)	
.32		2DS4	cA01 tB01-del7 (3DL2*007)		
3.4	FcgR	3DL2	cA01 tB01-del8		
3.43		⊤↓	cA01 tB01-ins3	2 (0.03)	
			KIR-cB01 tB01	435 (4.82)	3DL3 - 2DS2 - 2DL2 - 2DL5 - 2DS355 2DP1 - 2DL1 - 3DP1 - 2DL4 - 3DS1 - 2DL5 2DS355 20S1 - 3DL2
			cB01 tB01-del3	2 (0.02)	
			cB01 tB01-del7	2 (0.02)	
			cB01 tB01-del10	12 (0.13)	30L3 - 2052 - 20 <mark>L2 - 20L5 - 203355 - 2021 - 20L1 - 30P1 - 20L4 - 3D51 - 20L5 - 205355 - 2051 - 3DL2 -</mark>
			cB01 tB01-ins3	7 (0.07)	
			KIR-cB02 tA01	1079 (11.96)	3DL3 2D52 2DL2 3DP1 2DL4 3DL1 2D54 3DL2
			cB02 tA01-del3	8 (0.09)	
			cB02 tA01-ins4	20 (0.22)	
			cB02 tA01-ins5	3 (0.03)	



Error correct the reads containing full-length fosmid sequence using all continuous long reads

Trim the vector sequence from the corrected reads and overlap the fosmids, forming a draft haplotype sequence



Polish using the Quiver algorithm generating a highly accurate haplotype sequence

**Figure 3.** Workflow for SMRT<sup>®</sup> Sequencing of a fulllength 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.

#### NMDP-KIR-project

01-1096 cA01|tA01 3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DL1 2D54 3DL2

cA01|tB01-ins5 3dl3 2dl3 2dl1 2dl1 2dl1 3dp1 2dl4 3d51 2dl5 2dl5 2dl3 2dp1 2dl1 3dp1 2dl4 3d51 2dl5

0010-6217 cB02|tA01 3DL3 2D52 2DL2 3DP1 2DL4 3DL1 2D54 3DL2

cA01|tB01-del7 3DL3 2DL3 2DP1 2DL151 hybrid 3DL2 Iso-del

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



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

References

<sup>1)</sup> 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]



**Figure 4.** Recombineering-based method for isolating a tiling of targeted fosmids (~35 kb – 50 kb)<sup>3</sup> 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.

#### 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

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks are the sole property of their respective owners.