The American Journal of Human Genetics, Volume *99*

Supplemental Data

Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing

Paul J. Norman, Jill A. Hollenbach, Neda Nemat-Gorgani, Wesley M. Marin, Steven J. Norberg, Elham Ashouri, Jyothi Jayaraman, Emily E. Wroblewski, John Trowsdale, Raja Rajalingam, Jorge R. Oksenberg, Jacques Chiaroni, Lisbeth A. Guethlein, James A. Traherne, Mostafa Ronaghi, and Peter Parham

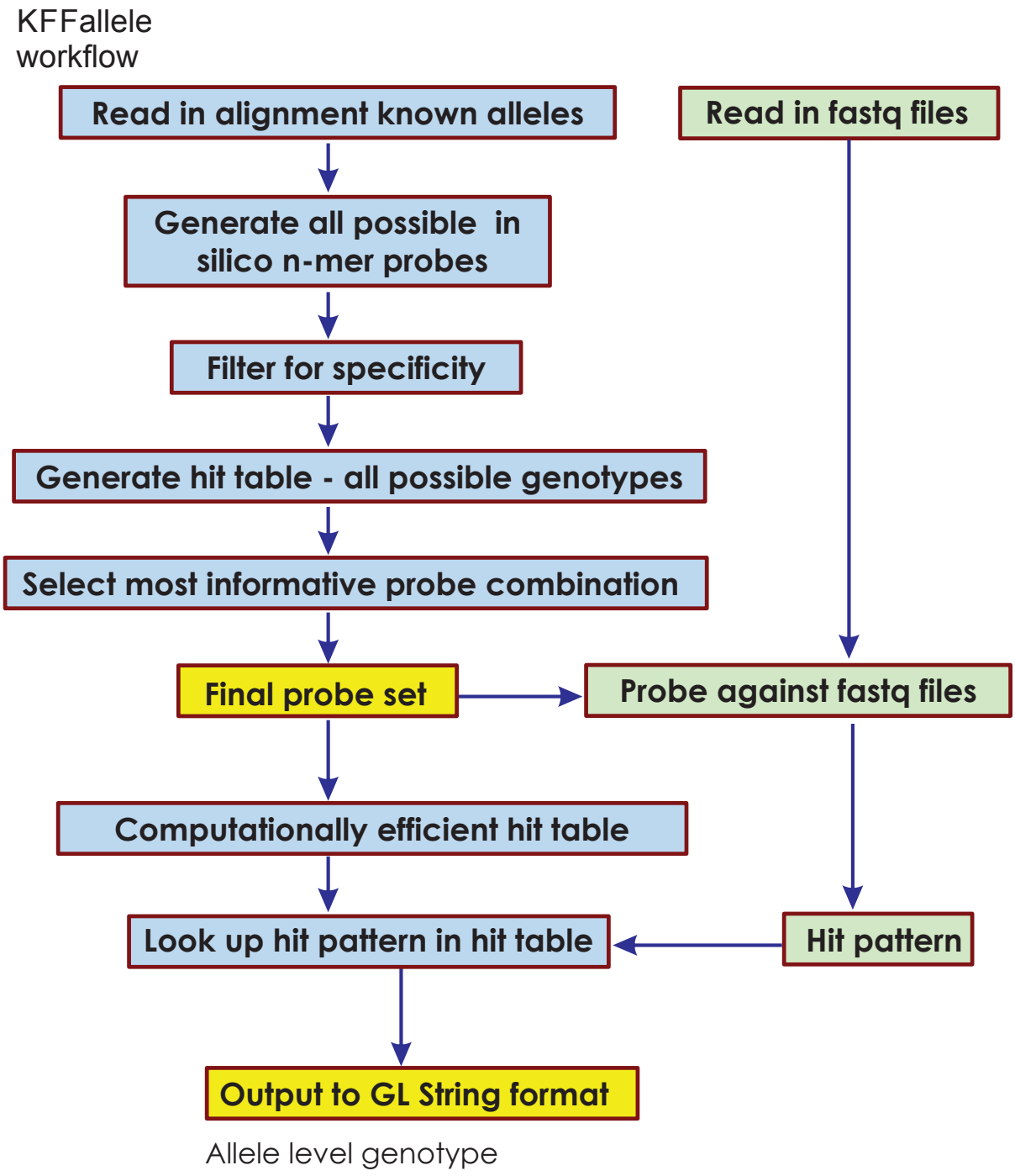
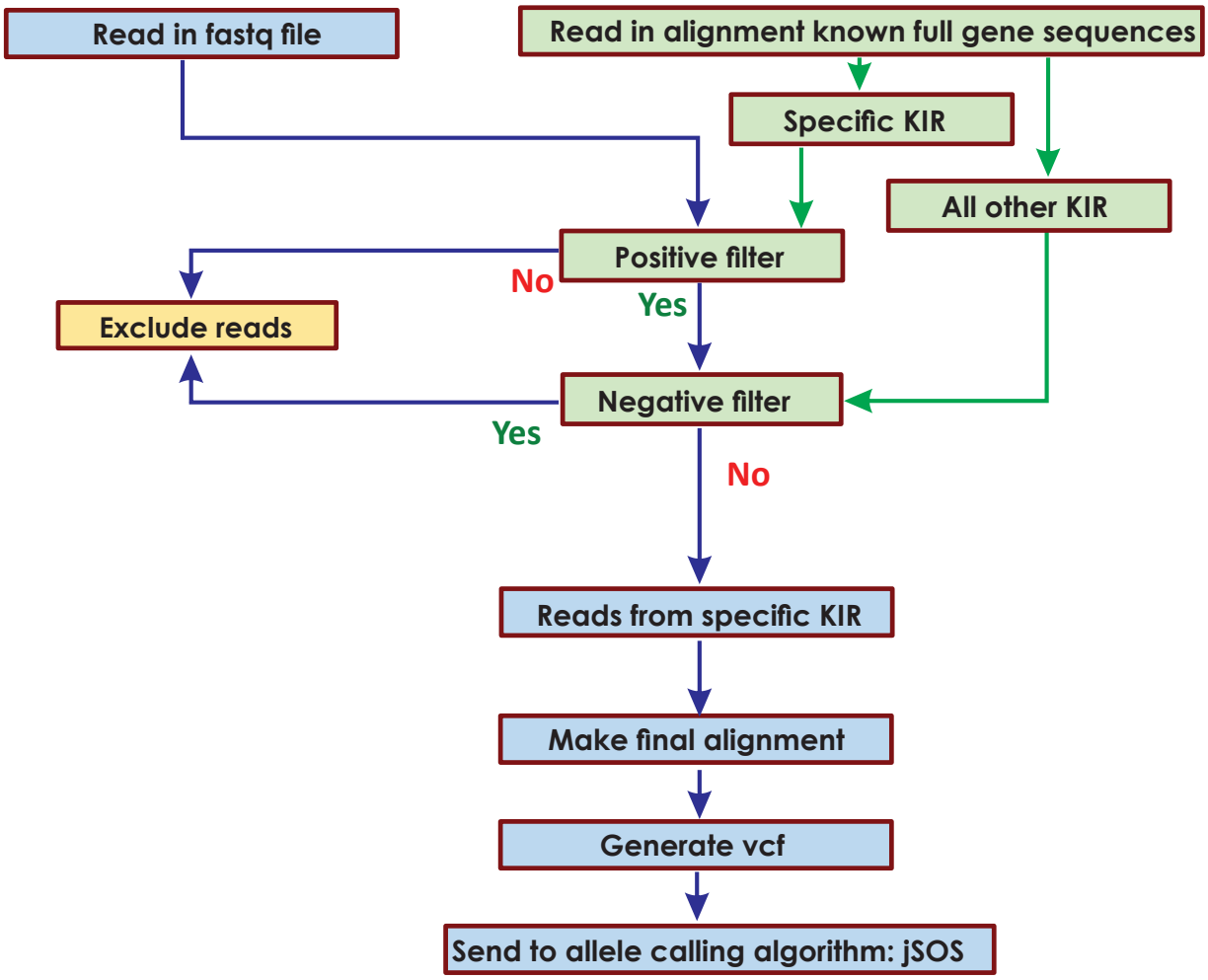
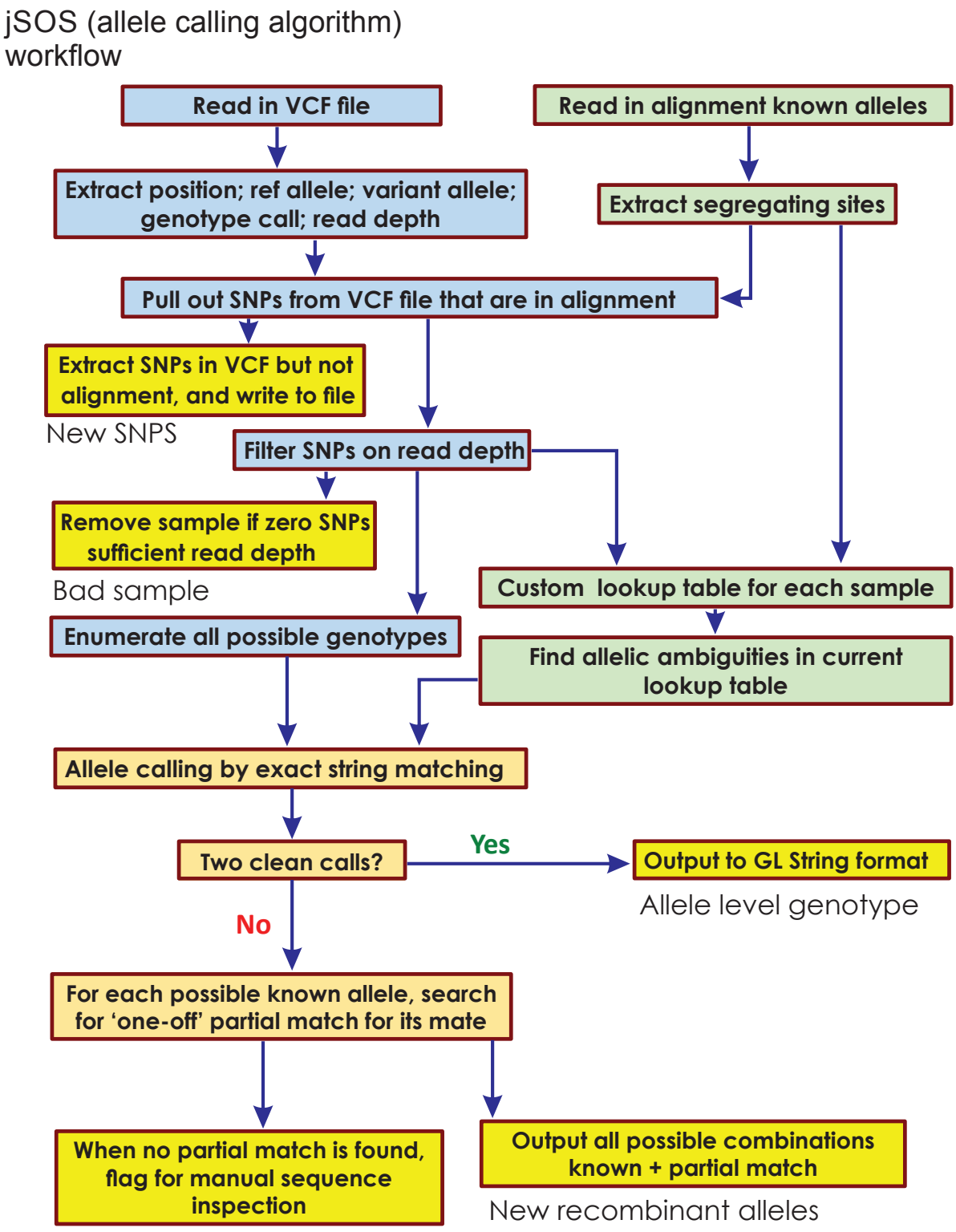


Figure S1. Description of the PING_allele workflow.
A. The KFFallele component, which generates *KIR* allele genotypes using sequence string searches of the unprocessed sequence read data.

SOS
workflow



B. The SOS component, which selects sequence reads specific for individual *KIR* genes, generates an alignment to a selected reference sequence and then determines the alleles present using a bespoke algorithm, termed jSOS (Figure S1C).



C. The jSOS allele-calling algorithm, which interprets the vcf files generated by SOS.

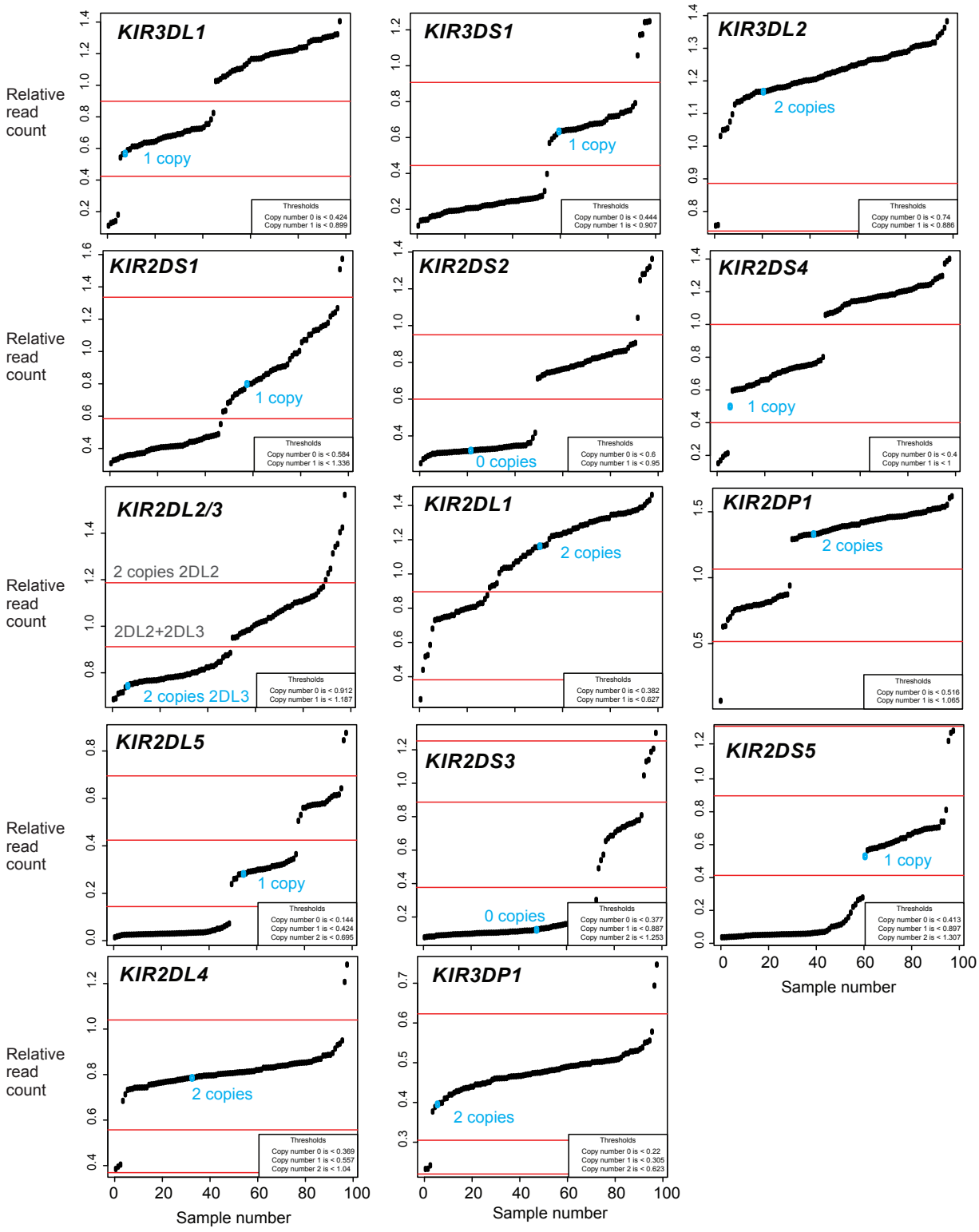


Figure S2. Read ratio groupings

Shown are MIRAgc plots for all *KIR* genes using 96 samples from DNA Set 3 (European controls) and sequenced using 2 x 300 bp MiSeq reads. Blue dots are the values obtained for the COX cell line, and the copy numbers (shown in blue text) match those obtained previously.[1; 2] Red lines are the threshold values set according to the groupings shown. The graphs were plotted using PING_gc V1.0 (See Web Resources). For *KIR2DL2/3* the groups also allow discrimination of the three possible genotypes as indicated (*KIR2DL2/2, KIR2DL2/3* and *KIR2DL3/3*).

KIR gene	Length (bp)	GenBank ID	Read ratio to determine copy number			
			1	2	3	4
*2DL1*00302*	14741	AC011501	0.4	0.78		
*2DS1*002*	14720	AL133414	0.6	1.2		
*2DL2*001*+	14782	AY320039		0.6		
*2DL4*00103*	11172	GU182338	0.2	0.8	1.2	
*2DL5A*001*	9695	AY320039	0.2	0.6	0.8	1.2
*2DP1*002*	13125	AC011501	0.2	0.7	1.3	
*2DS2*001*	14545	AL133414	0.6	0.95		
*2DS3*00103*	15071	AY320039	0.55	0.9	1.3	
*2DS4*00101*	15213	GU182338	0.4	1		
*2DS5*00201*	14996	AY320039	0.53	0.95	1.3	
*3DL1*00101*	14545	AC011501	0.5	0.78	1	
*3DS1*01301*	14933	AL133414	0.6	0.9		
*3DL2*00101*	17013	AC011501		1.9		
*3DL3*00201*	12360	AC011501				
*3DP1*002*	5713	AL133414	0.2	0.25	0.5	

Figure S3. *KIR* gene content calculations

Shows the reference gene sequences used to map sequence reads for gene content calculation for the 97 IHWG cell lines, which were sequenced using a 2 x 101 bp sequence run. The reads were filtered to be specific for the *KIR* region then mapped to all of the references simultaneously. The ratio of reads mapping to specific *KIR* / reads specific to *KIR3DL3* is used to calculate the copy number. The threshold values used to determine copy number are shown at the right. The accession numbers for reference alleles used for this purpose, as well as for PING_allele, are shown at the center. (+) for *KIR2DL2/3* discrimination of the three possible genotypes of the broad allele groups (*KIR2DL2* and *KIR2DL3*) was possible. The following threshold values were determined for the cell line data; 0.6 - *KIR2DL3* + *KIR2DL3*, 0.77 - *KIR2DL2* + *KIR2DL3*, 1.1 -*KIR2DL2* + *KIR2DL2*.

IHWG cell	Non-specific reads		Map to:	
	300 X 300 N	100 X 100 N	3DL2i6 %	Other KIR %
BOLETH	4	7,898	100	0
COX	1	7,497	100	0
DUCAF	2	6,801	100	0
ISH3	0	8,363	100	0
OMW	7	4,683	100	0
PGF	0	8,290	100	0
VAVY	0	5,624	100	0
WIN	2	3,966	100	0

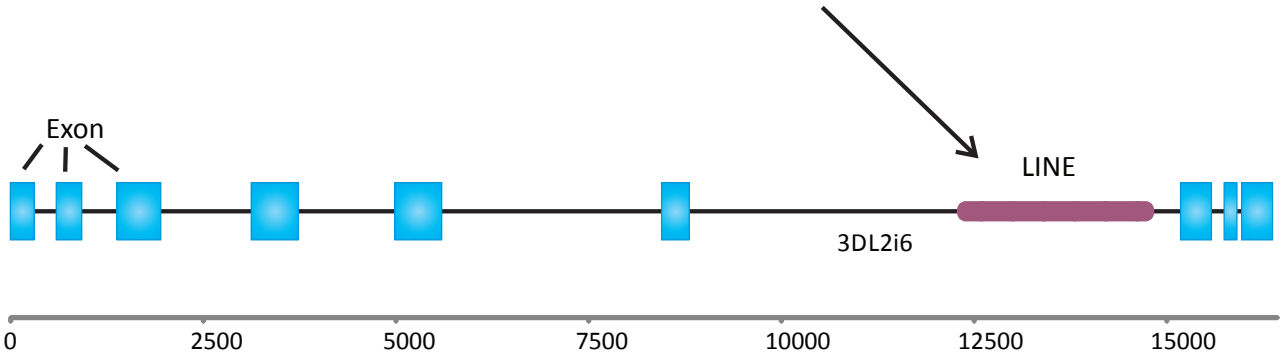


Figure S4. Potentially non-specific *KIR* reads
As part of the PING pipeline, sequence reads are filtered using a panel of reference haplotypes, in order to select those that originate from the *KIR* region. To test if any reads that may also map to elements outside the *KIR* region become selected in this process, they were re-mapped to the human genome (build 19). Shown are the results from eight cell lines that were each sequenced using two strategies (2 x 100 bp and 2 x 300 bp reads). All of those reads that could map outside *KIR* were selected and then mapped again to *KIR*, showing all of them originate from a single LINE element in intron 6 of *KIR3DL2*.

Supplemental References

1. Jiang, W., Johnson, C., Jayaraman, J., Simecek, N., Noble, J., Moffatt, M.F., Cookson, W.O., Trowsdale, J., and Traherne, J.A. (2012). Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. Genome research 22, 1845-1854.
2. Pyo, C.W., Guethlein, L.A., Vu, Q., Wang, R., Abi-Rached, L., Norman, P.J., Marsh, S.G., Miller, J.S., Parham, P., and Geraghty, D.E. (2010). Different patterns of evolution in the centromeric and telomeric regions of group A and B haplotypes of the human killer cell Ig-like receptor locus. PloS one 5, e15115.
3. Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., and Marsh, S.G. (2015). The IPD and IMGT/HLA database: allele variant databases. Nucleic acids research 43, D423-431.
4. Gomez-Lozano, N., Estefania, E., Williams, F., Halfpenny, I., Middleton, D., Solis, R., and Vilches, C. (2005). The silent KIR3DP1 gene (CD158c) is transcribed and might encode a secreted receptor in a minority of humans, in whom the KIR3DP1, KIR2DL4 and KIR3DL1/KIR3DS1 genes are duplicated. European journal of immunology 35, 16-24.
5. Norman, P.J., Abi-Rached, L., Gendzekhadze, K., Hammond, J.A., Moesta, A.K., Sharma, D., Graef, T., McQueen, K.L., Guethlein, L.A., Carrington, C.V., et al. (2009). Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes. Genome research 19, 757-769.
6. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.
7. Norman, P.J., Norberg, S.J., Nemat-Gorgani, N., Royce, T., Hollenbach, J.A., Shults Won, M., Guethlein, L.A., Gunderson, K.L., Ronaghi, M., and Parham, P. (2015). Very long haplotype tracts characterized at high resolution from HLA homozygous cell lines. Immunogenetics 67, 479-485.