bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Allele imputation for the Killer cell Immunoglobulin-like Receptor KIR3DL1/S1

Genelle F Harrison[1,2+], Laura Ann Leaton[1,2+], Erica A Harrison, Marte K Viken[3,4], Jonathan Shortt[1], Christopher R Gignoux[1], Benedicte A Lie [3,4], Damjan Vukcevic[5,6], Stephen Leslie[5,6,7], and Paul J Norman[1,2*]

1. Division of Biomedical Informatics and Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA.
2. Department of Immunology and Microbiology, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA.
3. Department of Immunology, University of Oslo and Oslo University Hospital, Norway.
4. Department of Medical Genetics, University of Oslo and Oslo University Hospital, Norway.
5. School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia
6. Melbourne Integrative Genomics, University of Melbourne, Parkville, Victoria, Australia
7. School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

+ equal contribution

Correspondence: *Paul Norman: paul.norman@cuanschutz.edu

Keywords: KIR3DL1, KIR3DS1; HLA; imputation; immune-mediated disease

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Highly polymorphic interactions of KIR3DL1 and KIR3DS1 with HLA class I ligands modulates the effector functions of natural killer (NK) cells and some T cells. This genetically determined diversity affects severity of infections, immune-mediated diseases, and some cancers, and impacts the course of cancer treatment, including transplantation. KIR3DL1 is an inhibitory receptor, and KIR3DS1 is an activating receptor encoded by the *KIR3DL1/S1* gene that has more than 200 diverse and divergent alleles. Determination of *KIR3DL1/S1* genotypes for medical application is hampered by complex sequence and structural variation that distinguishes individuals and populations, requiring targeted approaches to generate and analyze high-resolution allele data. To overcome these obstacles, we developed and optimized a model for imputing *KIR3DL1/S1* alleles at high-resolution from whole-genome SNP data, and designed to represent a substantial component of human genetic diversity. We show that our Global model is effective at imputing *KIR3DL1/S1* alleles with an accuracy ranging from 89% in Africans to 97% in East Asians, with mean specificity of 99.8% and sensitivity of 99% for named alleles >1% frequency. We used the established algorithm of the HIBAG program, in a modification named Pulling Out Natural killer cell Genomics (PONG). Because HIBAG was designed to impute *HLA* alleles also from whole-genome SNP data, PONG allows combinatorial diversity of *KIR3DL1/S1* and *HLA-A* and *B* to be analyzed using complementary techniques on a single data source. The use of PONG thus negates the need for targeted sequencing data in very large-scale association studies where such methods might not be tractable. All code, imputation models, test data and documentation are available at https://github.com/NormanLabUCD/PONG.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author Summary

Natural killer (NK) cells are cytotoxic lymphocytes that identify and kill infected or malignant cells and guide immune responses. The effector functions of NK cells are modulated through polymorphic interactions of KIR3DL1/S1 on their surface with the human leukocyte antigens (HLA) that are found on most other cell types in the body. KIR3DL1/S1 is highly polymorphic and differentiated across human populations, affecting susceptibility and course of multiple immune-mediated diseases and their treatments. Genotyping *KIR3DL1/S1* for direct medical application or research has been encumbered by the complex sequence and structural variation, which requires targeted approaches and extensive domain expertise to generate and validate high-resolution allele calls. We therefore developed Pulling Out Natural Killer Cell Genomics (PONG) to impute *KIR3DL1/S1* alleles from whole genome SNP data, and which we implemented as an open-source R package. We assessed imputation performance using data from five broad population groups that represent a substantial portion of human genetic diversity. We can impute *KIR3DL1/S1* alleles with an accuracy ranging from 89% in Africans and South Asians to 97% in East Asians. Globally, imputation of *KIR3DL1/S1* alleles having frequency >1% has a mean sensitivity of 94% and specificity of 99.8%. Thus, the PONG method both enables highly sensitive individual-level calling and makes large scale medical genetic studies of *KIR3DL1/S1* possible.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The *KIR3DL1/S1* gene encodes highly polymorphic receptors that are expressed by natural killer (NK) cells and some T cells to modulate their effector functions in immunity (1, 2). The receptors interact with HLA class I ligands that are expressed by most nucleated cells to signify their health status to the immune system (3, 4). KIR3DL1 allotypes are inhibitory receptors, specific for subsets of highly polymorphic HLA-A and B (5, 6). The KIR3DS1 allotypes are activating receptors, specific for non-polymorphic HLA-F and a smaller subset of HLA-A and B (7-9). Sequence diversity of KIR3DL1/S1 and HLA class I allotypes diversifies human immune responses to specific infections, cancers, cancer treatment and transplantation (10-19). Accordingly, this genetically determined diversity also associates with differential susceptibility and severity for multiple immune-mediated diseases (20-26). Although these factors render it imperative to genotype *KIR3DL1/S1* allotypes accurately for medical research and applications that include therapy decisions (27, 28), the high complexity of the genomic region presents obstacles for standard ascertainment methods (29). The ability to impute alleles from whole-genome SNP genotype (WG-SNP) data will decrease expense and effort, and greatly increase the capacity of research or applications where knowledge of KIR3DL1/S1 and HLA class I combinatorial diversity is critical.

The *KIR* locus, on human chromosome 19, is highly divergent in sequence and structure (29). As defined by the extensively curated ImmunoPolymorphism Database (IPD), *KIR3DL1/S1* has 220 alleles characterized (release 2.10.0: December 2020), with large numbers continuing to be discovered (30). As observed for polymorphic *HLA*, the *KIR3DL1/S1* alleles both distinguish individuals and characterize broad ancestral human populations (31, 32). As a likely consequence

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of selective pressure providing resistance to infectious diseases (33), specific combinations of KIR3DL1/S1 and HLA associate, differentially across populations, with severity of specific viral infections or autoimmune diseases (34-41). Likewise, specific combinations of KIR3DL1/S1 with HLA class I influence cancer susceptibilities non-uniformly across populations (42). In this regard, two key areas of human health significantly impacted by the population differentiation of KIR3DL1/S1 and HLA combinatorial diversity are HIV research and treatment, and cancer therapy (43-46). In particular, specific combinations of KIR3DL1/S1 and HLA allotypes influence rejection and relapse rates following transplantation (47-50). For these reasons, it is critical to establish methods for elucidating genetic variation in *KIR3DL1/S1* that can accommodate the full range of human genetic diversity.

KIR3DL1 specifically binds to subsets of HLA-A or B that carry a five amino acid motif, termed Bw4, on their external facing α1-helix (51). Expression of KIR3DL1 gives NK cells the ability to detect diseased cells that may have lost or altered expression of these HLA class I molecules, and likely serves as an immune checkpoint inhibitor for functionally mature T cells (52-54). KIR3DL1 polymorphism, and polymorphism both within and outside the Bw4 motif of HLA affects the specificity and strength of the interaction (55-57). Polymorphism also determines the expression level or signal transduction abilities of the receptor (58, 59). KIR3DL1/S1 segregates into three ancient lineages (015, 005 and 3DS1) that have distinct expression and function phenotypes (31). The 015 lineage comprises inhibitory receptors having high expression and high affinity for $Bw4^{+}$HLA-B. The 005 lineage are inhibitory receptors having low expression and preferential affinity for $Bw4^{+}$HLA-A. The 3DS1-lineage are activating receptors specific for HLA-F and some $Bw4^{+}$HLA-B allotypes expressed by infected cells (60-62). As defined by these phenotypes, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

lineages differentially associate with distinct pathological phenotypes (15, 63-65). Exceptions to these broad rules (e.g. 3DL1*007 belongs to the 015 lineage but has low expression) contribute to a hierarchy of receptor allotype strengths and reinforce the need to genotype *KIR3DL1/S1* to high resolution (66-68).

Multiple methods are available to impute *HLA class I* genotypes with high accuracy from WG-SNP data (69-73). We chose to adapt one of these programs so that *KIR3DL1/S1* and *HLA-A* and *B* genotypes could be imputed from the same data source, using an identical algorithm. In the current study we have adapted the HIBAG framework (73) to impute *KIR3DL1/S1* alleles, in a modification we have named Pulling Out NK cell Genomics (PONG). There are two components to the process: 1) model building that employs machine-learning to determine which combinations of SNPs correlate with known alleles, and 2) imputation that uses this model to determine allele genotypes from study cohorts (73, 74). Construction of the imputation models required high-resolution *KIR3DL1/S1* alleles and WG-SNP data obtained from the same set of individuals. PONG thus serves as a complement to PING (Pushing Immunogenetics to the Next Generation), which can determine *KIR3DL1/S1* alleles from high throughput sequence data (75). With a goal to create a model representing a substantial component of human genetic diversity, we compiled and rigorously tested the imputation using data from the 1,000 Genomes populations (76). The R package PONG is freely available, as are the data sets and imputation models described in this study.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

Method Overview

KIR3DL1/S1 exhibits exceptional sequence polymorphism as well as variation in gene content (**Figure 1A**). Here, we adapted and optimized the framework of HIBAG (73) to impute *KIR3DL1/S1* alleles, in a modification we have named Pulling Out NK cell Genomics (PONG). The development of PONG was focused on building a robust training model that could be used to impute unknown *KIR3DL1/S1* alleles from WG-SNP data across diverse global populations. Training an imputation model requires an input of individuals having known *KIR3DL1/S1* genotypes, coupled with high-density SNP data from the *KIR* region, as typically obtained through whole-genome SNP analysis (**Figure 1B**). We optimized the process using 1,000 Genomes individuals, because we had previously determined their *KIR3DL1/S1* alleles (77) and high density SNP data is available from this cohort (76). We distributed the 1,000 Genomes individuals into the designated five major population groups (termed 'superpopulations' by 1,000 Genomes): Africa and African-descent (AFR), Americas (AMR), East Asia (EAS), Europe (EUR) and South Asia (SAS). We first optimized the model building parameters using the EUR group. We randomly divided each population group into two parts, building an imputation model using the first part and testing the imputation accuracy with the second part. We then built a global model by combining all the 1,000 genomes individuals and repeating the process. Finally, we tested the global model on an independent population having both high-resolution sequence and WG-SNP data.

Samples and Genomic Data

We obtained high density SNP data for the *KIR* genomic region (chromosome 19: 55247563 – 55361930, Hg19) from the 1,000 Genomes Project Phase 3 individuals (76). The data had been

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

obtained using the Illumina Omni 2.5 platform, which has 4,093 SNPs in the *KIR* genomic region (76). To determine the *KIR3DL1/S1* alleles present in each individual we used the Pushing Immunogenetics to the Next Generation (PING) pipeline, as previously described (77). Included were a subset of 143 individuals from whom Sanger sequences of *KIR3DL1/S1* were obtained (77). In total, there were 2,083 individuals from the 1,000 Genomes data set from whom we had independently derived *KIR* sequence and chromosome 19 SNP data available (**Table S1a**), and these were divided into the designated five major population groups as indicated (**Table S1b**). We also analyzed SNP data obtained using the Infinium Immunoarray 24v2 (78) from 397 Norwegians (79), from whom we also determined high resolution *KIR3DL1/S1* genotypes through targeted sequencing (**Table S2**).

Modifications to HIBAG to Impute KIR3DL1/S1

We modified the HIBAG package version 1.2.4. The package name and relevant C++ functions were changed from HIBAG to KIRpong to avoid any conflict when both programs are installed. We removed genome build Hg18 and included Hg19 and Hg38 instead. We maintained many of the HIBAG functions while adjusting the selected chromosome positions to target the *KIR* gene cluster on chromosome 19. We modified the 'hlaBED2Geno' function to sample chromosome 19 positions 50247563 – 59128983 for Hg19 and 46457117 – 58617616 for Hg38. The 'hlaLociInfo' function was updated to target the *KIR* gene cluster and was specified as 55247563 – 55361930 for Hg19 and 54734034 – 54853884 for Hg38. Also in this function, the name of the gene was changed to *KIR3DL1/S1*. Finally, the printout messages were changed from HIBAG and *HLA* to PONG and *KIR3DL1/S1* to avoid confusion if both programs are active. The HIBAG functions

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were maintained, as extensive documentation for these functions is available. The modified package is available on Github (https://github.com/NormanLabUCD/PONG).

Optimization and Testing of Model Building

The input data for model building is a text file containing the *KIR3DL1/S1* allele information, and SNP data in PLINK (80) binary format (.bed, .bim, .fam files) from the same individuals (**Figure 1B**). The first column of the text file contains the sample name (Sample.ID), the second column, *KIR3DL1/S1* allele 1 (Allele1) and the third column, allele 2 (Allele2). We optimized the model parameters using the 1,000 Genomes European populations group (EUR), comprising 353 individuals from five countries (76) and having 26 distinct *KIR3DL1/S1* alleles (77). We then expanded model building and testing to include populations from Africa (AFR, 558 individuals, 46 distinct *KIR3DL1/S1* alleles), the Americas (AMR, 298 individuals, 34 *KIR3DL1/S1* alleles), East Asia (EAS, 406 individuals, 28 *KIR3DL1/S1* alleles) and South Asia (SAS, 467 individuals, 30 *KIR3DL1/S1* alleles). The cohort of 397 individuals from Norway contained 18 distinct *KIR3DL1/S1* alleles (**Table S2**), 14 of which were also present in the 1,000 Genomes data set.

We randomly selected 50% of individuals from the EUR group to be used for model building. The remaining 50% of individuals were used to test the accuracy of the model. We first optimized the parameters to be used for filtering SNPs prior to model building. We compared the imputation accuracy of models built after removing SNPs with minor allele counts (MAC) < 2, or < 3, or a minor allele frequency (MAF) < 1% or < 5%. We also tested the impact of removing individuals carrying any *KIR3DL1/S1* allele having MAC < 3 in the full EUR group (model + test). Once a robust model was established for the EUR population, we expanded the model to include all

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

205 populations using the pre-filtering parameters and procedures established above. Each model was
206 evaluated based on the time needed for model building as well as the accuracy of imputation.
207
208 ***Imputation of KIR3DL1/S1 alleles from Immunochip data***
209 To increase SNP density for the Norwegian cohort, we first imputed 1,000 Genomes WG-SNP
210 data using the Michigan imputation server (81). Although this process produced 2,882 SNPs in the
211 *KIR* region, it was insufficient to adequately improve accuracy of imputation (53% to 75%; data
212 not shown). We therefore expanded the target region to chr19: 55,100,000 – 55,500,000 (hg19) to
213 match that used for the KIR*IMP program that can be used to impute *KIR* gene content genotypes
214 (82), and built and tested *KIR3DL1/S1* allele imputation models as described above.
215
216 ***Computational Capabilities***
217 All experiments were performed using a server with 512 GB 2400MHz RAM, running Ubuntu
218 18.04, R 3.5.1, R-server 1.1.456, and using a single core from a 2.3GHz Xeon E5-2697 CPU.
219
220 ***Evaluation of imputation models***
221 Overall accuracy of a given imputation model was determined as the number of correct allele calls
222 made per individual (0, 1 or 2) divided by 2N. Sensitivity and specificity of a given model were
223 determined per *KIR3DL1/S1* allele. Sensitivity was measured as the percentage of individuals
224 known to be positive for a given allele who were also called positive for that allele by imputation.
225 Specificity was determined as the percentage of individuals known to be negative for a given allele
226 that were also called as negative for that allele by imputation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Parameters for frequency filtering of SNP and allele data

We designed, tested and optimized a model to impute *KIR3DL1/S1* alleles from WG-SNP data using a modification to the HIBAG framework and algorithm (73). We used SNP data from the 1,000 Genomes project (76) and *KIR3DL1/S1* genotypes that we had previously determined from sequence data from the same individuals (77). We focused first on the EUR group, comprised of 353 individuals and having 26 distinct *KIR3DL1/S1* allele sequences, ranging from 0.14% to 20% allele frequency (**Figure 2A**). We randomly selected 50% of the EUR individuals to be used for model building and used the other 50% to test the accuracy of the model. With the goal of maximizing the imputation accuracy of the test dataset, while preserving computational efficiency in model building, we first determined the effect of removing low-frequency SNPs. We measured the imputation accuracy of models that were built following removal of SNPs having a minor allele count (MAC) of < 2 (1,286 SNPs remaining in the *KIR* region) or MAC < 3 (1,044 SNPs remaining in the *KIR* region) in the full set of 353 individuals. We also measured the accuracy following removal of SNPs with a minor allele frequency (MAF) < 1% (941 SNPs remaining in the *KIR* region) or < 5% (645 SNPs remaining in the *KIR* region) in the full set of 353 individuals. A model was also built with no filtering of the genotype data for comparison (4,089 SNPs in *KIR* region). The *KIR* region SNPs used for testing the model accuracy were not filtered, and the models took from 6-9 seconds to impute *KIR3DL1/S1* alleles from the test data set of 177 individuals.

For the purposes of this test, accuracy was determined from the number of correct allele calls per individual in the test set. The lowest imputation accuracy was obtained using a MAF < 5%, with 91% of *KIR3DL1/S1* alleles called correctly, whereas models built with all other filtering

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

parameters imputed the alleles with 92% accuracy (**Figure 2B**, **Table S3**). Thus, imputation accuracy was similar across all SNP frequency filtering parameters tested (**Figure 2B**). The model building run time ranged from 29 minutes, when SNPs were filtered at MAF < 5%, to 84 minutes when no filtering was used. We selected MAC < 3 as this was the fastest build time (66 minutes) for models of 92% accuracy (**Figure 2B**). Of the 26 *KIR3DL1/S1* alleles observed in the full EUR group (N=353), twelve were observed less than three times (**Figure 2A**). Following removal of individuals possessing at least one of these twelve infrequent alleles, 14 *KIR3DL1/S1* alleles and 339 individuals remained in the population. As above, we removed SNPs having MAC < 3, divided this population in half, built a model and tested it on the other half. Compared with using MAC < 3 alone, the time required to build this model decreased from 66 minutes to 44 minutes, and the time to run the model reduced to 5 seconds (165 individuals), whereas the imputation accuracy increased from 92% to 96% (**Figure 2B**). Thus, this combination of filtering parameters produced the fastest time for model building and running, with the highest accuracy for imputing *KIR3DL1/S1* alleles. We therefore implemented these parameters in all subsequent analyses.

We next evaluated the sensitivity and specificity of the final EUR imputation model, as described in Methods. Of 14 alleles in the model data set, 13 were also present in the test set (**Figure 2C**). We observed a modal specificity of 100%, and a mean of 99%. The two alleles having 98% specificity were *KIR3DL1*00101* and *KIR3DS1*01301*, thus for every 100 individuals imputed to have either allele, two were not shown as present through sequencing. We observed similarly high modal sensitivity of 100%, with a mean of 77%. All alleles with a frequency in the EUR group greater than 1.6% were imputed with >99% sensitivity. Below 1.6% allele frequency, two alleles (*3DL1*00402* and *3DS1*049N*) were imputed with 50% sensitivity and two (*KIR3DL1/S1* neg,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and *3DL1*009*) with 0% sensitivity. *KIR3DL1*00402* and *3DS1*049N* are each distinguished from their closest (parental) alleles by a single or a doublet nucleotide substitution, respectively (30). Accordingly, in each case these alleles were imputed as the parental allele (not shown). *KIR3DL1*009* represents a double recombination having exons 2-3 identical to *3DS1*01301* and exons 1 and 4-9 identical to *3DL1*001* (83, 84). The haplotype that lacks *KIR3DL1/S1* represents a large-scale deletion encompassing up to seven *KIR* genes (**Figure 1A**), and likely has very few identifying SNPs within the *KIR* locus. Thus, we observe a clear relationship between *KIR3DL1/S1* allele frequency and accuracy of imputation, with all high-frequency alleles being imputed with high accuracy, and those imputed with lower accuracy attributed both to their low frequency and lack of additional identifying characteristics.

Development of a trained Global model for KIR3DL1/S1 imputation

After establishing the most robust filtering parameters for model building in the EUR population group, we expanded the analysis to the four other major population groups from the 1,000 Genomes project (Africa - AFR, Americas - AMR, East Asia - EAS and South Asia - SAS). We also combined all five population groups to form an additional 'Global' group (**Table 1**). As above, *KIR* locus SNPs and *KIR3DL1/S1* alleles having MAC < 3 in each respective group were removed, imputation models were then built using 50% individuals, and tested on the remaining 50%. Following the filtering based on *KIR3DL1/S1* allele counts, the African population group had the highest diversity with 31 alleles and the East Asian group the lowest with 13 alleles (**Figure 3A**). A total of 90 distinct *KIR3DL1/S1* alleles were present in the Global group, 42 of these occurred more than twice in total and were therefore included in model building. This allele filtering process resulted in 58 of the 2,082 individuals being removed. The Global model included 1,017

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

296 individuals and took 10 weeks to build. This process also increased the number of target alleles
297 within all the individual population groups (**Figure 3A**).
298
299 **Table 1. Number of *KIR3DL1/S1* alleles and individuals in data sets.**

	Number of Individuals in Data Set			
1000 Genomes Population Group	All	Global *KIR3DL1/S1* MAC < 3	In Model Set	In Test Set
Africa (AFR)	558	541	272	269
Americas (AMR)	298	292	146	146
East Asia (EAS)	406	389	196	193
Europe (EUR)	353	345	174	171
South Asia (SAS)	467	457	229	228
Global	2,082	2,024	1,017	1,007

300
301
302 In testing models built within each respective population group, imputation accuracy ranged from
303 87.8% in the SAS group to 96.6% in EAS group (**Figure 3B** and **Table S3**). When using the Global
304 model, however, imputation accuracy increased for all groups, ranging from 89.0% in SAS and to
305 97.2% in EAS (**Figure 3B**). When the test group was comprised of individuals drawn from all five
306 of the population groups, an accuracy of 92.3% was achieved. This latter finding gives an estimate
307 of the accuracy of *KIR3DL1/S1* allele imputation for individuals of unknown genetic ancestry.
308 Using the Global model, the imputation time ranged from 2 min 9 s for AMR (N=146) to 4 min 5
309 s for SAS (N =228), and it took 16 min 1 s to impute the Global test set of 1,007 individuals (**Table**
310 **S3**).
311

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We next evaluated the specificity and sensitivity of the Global imputation model. The mean specificity across the 42 alleles was 99.8%, with 40 of them having a specificity above 99% (**Figure 3C**). The lowest specificities were observed for *3DS1*01301* at 96% and 3DL1*01502 at 98.5%. Of the individuals falsely imputed as having *3DS1*01301*, 84% were due to a *KIR3DL1/S1* deletion haplotype. This finding is consistent with the suggestion that the parental haplotype for the deletion carried *3DS1*01301* (84). The individuals falsely imputed as having *3DL1*01502*, possessed either *3DL1*01702*, **051* or **025* (33% each). All these alleles fall into the same ancestral lineage as *3DL1*015*, and likely exhibit similar phenotypes of high expression and ligand binding (1). In the final Global population group (2N = 4,068) there were 15 *KIR3DL1/S1* alleles with a frequency above 1% and 27 alleles with a frequency below 1% (**Figure 3C**). For those *KIR3DL1/S1* alleles having allele frequency below 1%, we observed a modal sensitivity of 0%, and a mean of 29%. Conversely, *KIR3DL1/S1* alleles with a frequency above 1% had a modal sensitivity of 100% and a mean of 94%. The sensitivity rose to 99% when the allele representing the absence of *KIR3DL1/S1* was excluded. Despite a frequency of less than 1% the alleles *006, *092, *035 and *089 were imputed with 100% sensitivity (**Figure 3C**).

In total 27 *KIR3DL1/S1* alleles had a frequency less than 1% in the Global population. When the global frequency was above 1%, PONG was able to impute the alleles 91 to 100% of the time (**Figure 3D**). Thus, similar to the model built using the European population group, low frequency alleles were more likely to be incorrectly imputed than high frequency alleles. An exception to this was the allele representing the absence of *KIR3DL1/S1* (*00000) in which the frequency was 4% but PONG was only able to impute the absent allele correctly 35% of the time. Together this shows that PONG is effective for imputing common *KIR3DL1/S1* alleles and some rare alleles across a

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

335 diverse set of human populations. Consistent with their overall lower accuracy, the African and
336 South Asian population groups had the highest number of *KIR3DL1/S1* alleles with a frequency
337 less than 1% (17 and 9, respectively). By contrast, only three low frequency alleles were present
338 in the East Asian population. In summary, the accuracy of PONG is affected by the frequency of
339 *KIR3DL1/S1* alleles and is therefore less effective in more diverse human populations given a
340 similar-sized training sample.
341
342 ***Testing the Global model using less dense genotyping datasets.***
343 We analyzed a cohort of 397 individuals from Norway, from whom we generated Infinium
344 Immunoarray SNP and high-resolution *KIR3DL1/S1* allele sequence data. For this test, we
345 extended the window in which classifiers are sampled to match that of the KIR*IMP program
346 (chr19: 55,100,000 – 55,500,000: hg19), which contains 294 SNPs on this chip. We observed a
347 strong correlation between allele frequencies in Norway and the 1,000 Genomes EUR group ($r = 0.96$). In total, 18 *KIR3DL1/S1* alleles were identified in the Norwegian cohort through nucleotide
349 sequencing, including one rare allele (*3DL1*044*: 0.02%) that is absent from the 1,000 Genomes
350 Global population. After filtering for MAC < 3, there were 13 *KIR3DL1/S1* alleles present. In
351 testing the model built using 50% of the Norwegian cohort against the other 50% of the cohort, we
352 observed 92% accuracy, sensitivity of mode 100% and mean 75%, and specificity of mode 100%,
353 mean 99% (**Figure 4**). As in previous analyses, *KIR3DL1/S1* alleles having allele frequency > 1%
354 have greater imputation accuracy than those < 1% (**Figure 4**). In this analysis the modal sensitivity
355 of alleles with a frequency < 1% was 0% with a mean of 33%. By contrast, *KIR3DL1/S1* alleles
356 with a frequency > 1% had a modal sensitivity of 100% and a mean of 85%. These experiments

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

357 show that high resolution KIR3DL1/S1 genotypes can be imputed from low-density SNP arrays,
358 and with similar accuracy to high-density arrays.
359
360 ***Obtaining and Running PONG***
361 The PONG program is installed using the command line and opened as a library in R (R version
362 2.14.0 – 4.0.0.) (85). PONG can be run using WG-SNP data mapped either to hg19 or hg38. The
363 imputation algorithm does not require data to be phased (73). The Global model (hg19) and the
364 model built with the EUR group (hg38) are available for download. Other models will be added
365 as they become available. Using our Global model, we estimate that 1,000 individuals could be
366 imputed every 15 minutes using a single core on a laboratory server, such as the one we have used.
367 Users can also create their own models when WG-SNP data and *KIR3DL1/S1* allele genotypes are
368 available, and modify the data input and filtering parameters, as described in the tutorial.
369 • The imputation models are available at https://github.com/NormanLabUCD/PONG
370 • The 1,000 Genomes test data can be found in ref (77) and **Table S1**.
371 • A tutorial describing the pipeline for model building and testing is available at:
372 https://github.com/NormanLabUCD/PONG/inst/doc/
373

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Knowledge of *KIR3DL1/S1* diversity can help predict the course of specific infections, immune-mediated diseases, and their therapies (86-92). However, by virtue of the polymorphic and structural complexity at the locus, it is often ignored in genome-wide association studies. The primary goal of this study was to develop a model trained to impute *KIR3DL1/S1* alleles rapidly from WG-SNP data encompassing a wide range of human genetic diversity. We built imputation models using high-density WG-SNP data (76) and high-resolution *KIR3DL1/S1* allele calls (77) from the five broadly defined 1,000 Genomes population groups, and then built a model for the Global group. To achieve these goals, we adapted the coding framework and algorithm from HIBAG (73) in a modification that we have named PONG. We determined that the imputation models are most effective when both the WG-SNP data and *KIR3DL1/S1* alleles have been filtered to remove alleles that occur infrequently. The former filter to reduce the model building run time and the latter to increase imputation accuracy. The resulting range of imputation accuracies of the final Global model was 89% for Africans and South Asians, to 97% for East Asians. The 1,000 Genomes WG-SNP data has a dense set of genotypes, including 1,832,506 SNPs from chromosome 19 (76). Other genotype chips used for disease association studies have less dense sets of SNPs, including the Infinium Immunoarray, which targets markers associated with autoimmune disease and inflammatory disorders (78). Because *KIR3DL1/S1* diversity is associated with development or severity of multiple autoimmune diseases (20, 22, 23, 25, 93), we tested the accuracy of imputation using results generated from this genotyping chip, and achieved similar imputation accuracy as achieved from the high-density array.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Although PONG is effective in imputing *KIR3DL1/S1* alleles, there are a few limitations to this program that we are optimistic will improve over time. As observed for *HLA* (94), we found a negative correlation between the accuracy of PONG and the diversity of *KIR3DL1/S1* alleles in a population. For example, African populations have the highest number of distinct *KIR3DL1/S1* alleles as well as the highest number with a frequency below 1%. The result is that imputation accuracy is lowest in Africans. Conversely, the East Asian group has the lowest number of *KIR3DL1/S1* alleles of allele frequency below 1%, and the highest imputation accuracy. Therefore, PONG is most effective at imputing the most frequent alleles. The imputation accuracy of the model will improve over time as more immunogenetic studies of *KIR3DL1/S1* are conducted, thus expanding our sample set for building more robust and diverse models. Given that the model is open source, and that PONG has a model building function available, this can be achieved both by the developers and users. PONG is also less accurate at imputing the absence of the *KIR3DL1/S1* gene (which we designated *00000), and we were only able to impute this null *KIR3DL1/S1* allele at an accuracy of 35% using a global model. However KIR*IMP, which is targeted to *KIR* gene content diversity, is able to impute the presence or absence of *KIR3DL1/S1* with an accuracy above 90% (82). Therefore, PONG can be coupled with KIR*IMP to improve the accuracy of imputing the '*KIR3DL1/S1* absent' allele.

Accurate sequencing and assembly can be challenging for highly polymorphic and structurally diverse regions of the genome (95). Both these phenomena are characteristics of the *KIR* locus (29). Therefore, PONG relies on high quality WG-SNP data with robust quality control measures implemented in SNP calling pipelines. New techniques to improve the identification of structural variation are being created, including long-range optical mapping, which uses the optical signal

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

419 strength from each SNP genotype to identify deletions and duplications (96). Together, an
420 increased sampling of individuals having rare *KIR3DL1/S1* alleles and better characterization of
421 structural variation from WG-SNP data will likely improve the imputation accuracy of PONG.
422
423 Highly polymorphic interactions of KIR3DL1/S1 with HLA-A and B modulate the critical
424 functions of NK cells in immunity, which include the destruction of infected or cancerous cells
425 (2). Combinatorial diversity of KIR3DL1/S1 with HLA-A and B allotypes thus affects the
426 susceptibility and course of multiple immune-mediated diseases. Several methods are available to
427 impute *HLA* alleles (70-73), but large-scale genetic studies often exclude analysis of *KIR3DL1/S1*
428 due to the exceptional polymorphism and structural diversity of the genomic region. A secondary
429 goal of this study was thus to produce imputation models that could be used in conjunction with
430 existing models to impute the combinatorial diversity of KIR3DL1/S1 and HLA allotypes. By
431 comparison with *KIR3DL1/S1*, the mean imputation accuracy for HIBAG across seven *HLA* genes
432 was 81.2% in African populations and 91.1% in East Asians (73). In African populations, *HLA-*
433 *DPB1* had the lowest imputation accuracy at 74.2% and *HLA-A* had the highest observed accuracy
434 at 92.4%. The corresponding imputation accuracies of these *HLA* genes in East Asians were 89.8%
435 and 92.1% respectively (73). Therefore, the mean accuracy of *KIR3DL1/S1* allele imputation
436 described herein is equivalent, and likely better than that obtained for *HLA class I* and *II* using the
437 same underlying algorithm. We therefore propose that using this algorithm to impute both
438 KIR3DL1/S1 and HLA-A and B genotypes from WG-SNP data presents a considerable advantage
439 over other approaches. This approach is particularly applicable for studies of Biobank data, where
440 targeted sequencing of *KIR3DL1/S1* and *HLA-A* and *B* from many thousands of individuals is not

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

currently tractable. Utilizing our pre-built models, PONG can be implemented to make genetic association studies of *KIR3DL1/S1* in combination with *HLA-A* and *B* possible at very large scale.

Acknowledgements

This study was performed with support from National Institutes of Health of the USA, R56 AI151549 and R01 AI128775 (PN), and R01HG010297 (CG).

Data availability statement

All code written in support of this publication, imputation models, test data and documentation on installing and running are publicly available at https://github.com/NormanLabUCD/PONG

Conflict of Interest

SL is a partner in Peptide Groove LLP. All other authors declare no competing interest.

Author Contributions

Conceptualization: Paul J. Norman, Damjan Vukcevic and Stephen Leslie

Data Curation: Genelle F Harrison, Laura Ann Leaton, Marte K Viken and Paul J. Norman

Formal analysis: Genelle F Harrison and Laura Ann Leaton

Funding acquisition: Paul J. Norman

Investigation: Laura Ann Leaton, Genelle F Harrison and Paul J. Norman

Methodology: Genelle F Harrison, Laura Ann Leaton, Jonathan Shortt, Christopher R Gignoux and Paul J. Norman

Project Administration: Paul J. Norman

Resources: Marte K Viken, Benedicte A Lie, Christopher R Gignoux and Paul J. Norman

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

466
467 **Software:** Genelle F Harrison, Laura Ann Leaton, Jonathan Shortt and Erica A Harrison

468 **Supervision:** Paul J. Norman

469 **Validation:** Genelle F Harrison, Laura Ann Leaton and Erica A Harrison

470 **Visualization:** Genelle F Harrison, Laura Ann Leaton and Paul J. Norman

471 **Writing – Original Draft Preparation:** Genelle F Harrison, Laura Ann Leaton and Paul J.
472 Norman

473 **Writing – Review & Editing:** Genelle F Harrison and Paul J. Norman and all authors

474

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure legends

Figure 1. Genomic location of *KIR3DL1/S1* and overview of allele imputation workflow.

A. Shows the location of the *KIR3DL1/S1* gene on five examples of common *KIR* haplotypes. *KIR3DL1/S1* is shaded in blue, and other *KIR* genes are shaded grey. The *KIR3DL1/S1* gene can be absent (haplotype 4) or fused in-frame with *KIR3DL2* (haplotype 5) (84). The human genome coordinates (build hg19) from which classifiers were drawn for imputation are given at the top.

B. Schematic of model building, testing and output for the imputation of *KIR3DL1/S1* alleles using PONG. Shown are the required input files and their format for model building (blue) and testing (green). Red boxes give an example of the output from the imputation.

Figure 2. Optimization of *KIR3DL1/S1* allele imputation using data from Europeans.

A. Bar graph shows the *KIR3DL1/S1* allele frequencies in the combined EUR population group comprised of 353 individuals from Italy, Finland, United Kingdom, Spain, or Utah. The alleles were determined from high-throughput sequence data (77).

B. Shown is a summary of the results obtained using models tested during optimization. From left to right are the filtered criteria (SNPs or *KIR3DL1/S1* alleles), the filtering threshold values, resulting model build time, and accuracy of the imputed genotypes. Grey dotted arrow indicates that the final model that was built using MAC < 3 for SNPs and for *KIR3DL1/S1* alleles.

C. Shows the imputation accuracy for each *KIR3DL1/S1* allele present in the final filtered EUR data set. Blue bars indicate the sensitivity (% of times a given allele was called as present when known to be present). Red line indicates specificity (% of times a given allele was called as absent when known to be absent).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3. Accurate imputation of *KIR3DL1/S1* alleles using a Global population model.

A. Bar graphs shows the number of *KIR3DL1/S1* alleles present in each of the five broad population groups of the 1,000 Genomes database. The bar colors indicate: (pink) the number of alleles present before filtering, (ruby) by MAC < 3 filtering, and (burgundy) by combining the five groups to form a Global population and then MAC < 3 filtering. The population groups are East Asian (EAS), European (EUR), South Asian (SAS), American (AMR) and African (AFR).

B. Shows the imputation accuracy obtained for each of the population group and the Global models. (Within group) the model was built using 50% of the indicated group and tested on the other 50%. (Global) the model was built using 50% of all individuals and tested on the remaining 50% of the specified group.

C. and **D.** Show the imputation efficacy for each allele present in the final Global data set. Blue bars indicate the sensitivity (% of times a given allele was called as present when known to be present). Red line indicates specificity (% of times a given allele was called as absent when known to be absent). Blue dots indicate the KIR3DL1/S1 allele frequencies in the Global population.

Figure 4. Accurate imputation of *KIR3DL1/S1* alleles from Immunochip SNP data.

Bar graph shows the efficiency of *KIR3DL1/S1* allele imputation using a model built and tested on a cohort from Norway who also had their *KIR3DL1/S1* alleles genotyped to high resolution. Blue bars indicate the sensitivity (% of times a given allele was called as present when known to be present). Red line indicates specificity (% of times a given allele was called as absent when known to be absent).

Supporting Information

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

523 **S1 Table** (xlsx) *KIR3DL1/S1* genotypes of 1,000 Genomes Individuals

524 **S2 Table** (xlsx) *KIR3DL1/S1* genotypes of Norwegian Individuals

525 **S3 Table** (xlsx) Parameters and Output statistics of Imputation Models

526

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Parham P, Norman PJ, Abi-Rached L, Guethlein LA. Variable NK cell receptors exemplified by human KIR3DL1/S1. The Journal of Immunology. 2011;187(1):11-9.
2. O'Connor GM, McVicar D. The yin-yang of KIR3DL1/S1: molecular mechanisms and cellular function. Critical reviews in immunology. 2013;33(3):203-18.
3. Quatrini L, Chiesa MD, Sivori S, Mingari MC, Pende D, Moretta L. Human NK cells, their receptors and function. European journal of immunology. 2021.
4. Hammer Q, Rückert T, Romagnani C. Natural killer cell specificity for viral infections. Nature immunology. 2018;19(8):800-8.
5. Litwin V, Gumperz J, Parham P, Phillips JH, Lanier LL. NKB1: a natural killer cell receptor involved in the recognition of polymorphic HLA-B molecules. The Journal of experimental medicine. 1994;180(2):537-43.
6. Colonna M, Samaridis J. Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural killer cells. Science (New York, NY). 1995;268(5209):405-8.
7. Garcia-Beltran WF, Hölzemer A, Martrus G, Chung AW, Pacheco Y, Simoneau CR, et al. Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. Nat Immunol. 2016;17(9):1067-74.
8. Kiani Z, Bruneau J, Geraghty DE, Bernard NF. HLA-F on Autologous HIV-Infected Cells Activates Primary NK Cells Expressing the Activating Killer Immunoglobulin-Like Receptor KIR3DS1. Journal of virology. 2019;93(18).
9. O'Connor GM, Vivian JP, Gostick E, Pymm P, Lafont BA, Price DA, et al. Peptide-Dependent Recognition of HLA-B*57:01 by KIR3DS1. Journal of virology. 2015;89(10):5213-21.
10. Ahlenstiel G, Martin MP, Gao X, Carrington M, Rehermann B. Distinct KIR/HLA compound genotypes affect the kinetics of human antiviral natural killer cell responses. The Journal of clinical investigation. 2008;118(3):1017-26.
11. Digitale JC, Callaway PC, Martin M, Nelson G, Viard M, Rek J, et al. Inhibitory KIR ligands are associated with higher P. falciparum parasite prevalence. The Journal of infectious diseases. 2020.
12. Forlenza CJ, Boudreau JE, Zheng J, Le Luduec JB, Chamberlain E, Heller G, et al. KIR3DL1 Allelic Polymorphism and HLA-B Epitopes Modulate Response to Anti-GD2 Monoclonal Antibody in Patients With Neuroblastoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2016;34(21):2443-51.
13. López-Vázquez A, Rodrigo L, Martínez-Borra J, Pérez R, Rodríguez M, Fdez-Morera JL, et al. Protective effect of the HLA-Bw4I80 epitope and the killer cell immunoglobulin-like receptor 3DS1 gene against the development of hepatocellular carcinoma in patients with hepatitis C virus infection. The Journal of infectious diseases. 2005;192(1):162-5.
14. MacFarlane AWt, Jillab M, Smith MR, Alpaugh RK, Cole ME, Litwin S, et al. NK cell dysfunction in chronic lymphocytic leukemia is associated with loss of the mature cells expressing inhibitory killer cell Ig-like receptors. Oncoimmunology. 2017;6(7):e1330235.
15. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F, et al. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nature Genetics. 2007;39(6):733-40.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

570 16. Ruggeri L, Mancusi A, Urbani E, Velardi A. Identifying NK Alloreactive Donors for
571 Haploidentical Hematopoietic Stem Cell Transplantation. Methods in molecular biology (Clifton,
572 NJ). 2016;1393:141-5.
573 17. Trefny MP, Rothschild SI, Uhlenbrock F, Rieder D, Kasenda B, Stanczak MA, et al. A
574 Variant of a Killer Cell Immunoglobulin-like Receptor Is Associated with Resistance to PD-1
575 Blockade in Lung Cancer. Clinical cancer research : an official journal of the American
576 Association for Cancer Research. 2019;25(10):3026-34.
577 18. Greenland JR, Sun H, Calabrese D, Chong T, Singer JP, Kukreja J, et al. HLA Mismatching
578 Favoring Host-Versus-Graft NK Cell Activity Via KIR3DL1 Is Associated With Improved
579 Outcomes Following Lung Transplantation. American journal of transplantation : official journal
580 of the American Society of Transplantation and the American Society of Transplant Surgeons.
581 2017;17(8):2192-9.
582 19. van der Ploeg K, Le Luduec JB, Stevenson PA, Park S, Gooley TA, Petersdorf EW, et al.
583 HLA-A alleles influencing NK cell function impact AML relapse following allogeneic
584 hematopoietic cell transplantation. Blood advances. 2020;4(19):4955-64.
585 20. Ahn RS, Moslehi H, Martin MP, Abad-Santos M, Bowcock AM, Carrington M, et al.
586 Inhibitory KIR3DL1 alleles are associated with psoriasis. Br J Dermatol. 2016;174(2):449-51.
587 21. Anderson KM, Augusto DG, Dandekar R, Shams H, Zhao C, Yusufali T, et al. Killer Cell
588 Immunoglobulin-like Receptor Variants Are Associated with Protection from Symptoms
589 Associated with More Severe Course in Parkinson Disease. J Immunol. 2020.
590 22. Augusto DG, Lobo-Alves SC, Melo MF, Pereira NF, Petzl-Erler ML. Activating KIR and
591 HLA Bw4 ligands are associated to decreased susceptibility to pemphigus foliaceus, an
592 autoimmune blistering skin disease. PloS one. 2012;7(7):e39991.
593 23. Lorentzen AR, Karlsen TH, Olsson M, Smestad C, Mero IL, Woldseth B, et al. Killer
594 immunoglobulin-like receptor ligand HLA-Bw4 protects against multiple sclerosis. Annals of
595 neurology. 2009;65(6):658-66.
596 24. Petrushkin H, Norman PJ, Lougee E, Parham P, Wallace GR, Stanford MR, et al.
597 KIR3DL1/S1 Allotypes Contribute Differentially to the Development of Behçet Disease. The
598 Journal of Immunology. 2019;203(6):1629-35.
599 25. Vendelbosch S, Heslinga SC, John M, van Leeuwen K, Geissler J, de Boer M, et al. Study
600 on the protective effect of the KIR3DL1 gene in ankylosing spondylitis. Arthritis & rheumatology
601 (Hoboken, NJ). 2015;67(11):2957-65.
602 26. Yawata N, Shirane M, Woon K, Lim X, Tanaka H, Kawano YI, et al. Molecular Signatures
603 of Natural Killer Cells in CMV-Associated Anterior Uveitis, A New Type of CMV-Induced
604 Disease in Immunocompetent Individuals. International journal of molecular sciences. 2021;22(7).
605 27. Erbe AK, Wang W, Carmichael L, Hoefges A, Grzywacz B, Reville PK, et al. Follicular
606 lymphoma patients with KIR2DL2 and KIR3DL1 and their ligands (HLA-C1 and HLA-Bw4)
607 show improved outcome when receiving rituximab. Journal for immunotherapy of cancer.
608 2019;7(1):70.
609 28. Shaffer BC, Hsu KC. Selection of allogeneic hematopoietic cell transplant donors to
610 optimize natural killer cell alloreactivity. Seminars in hematology. 2020;57(4):167-74.
611 29. Béziat V, Hilton HG, Norman PJ, Traherne JA. Deciphering the killer-cell
612 immunoglobulin-like receptor system at super-resolution for natural killer and T-cell biology.
613 Immunology. 2017;150(3):248-64.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

30. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43(Database issue):D423-31.
31. Norman PJ, Abi-Rached L, Gendzekhadze K, Korbel D, Gleimer M, Rowley D, et al. Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans. Nat Genet. 2007;39(9):1092-9.
32. Nunes K, Maia MHT, Dos Santos EJM, Dos Santos SEB, Guerreiro JF, Petzl-Erler ML, et al. How natural selection shapes genetic differentiation in the MHC region: A case study with Native Americans. Hum Immunol. 2021.
33. Deng Z, Zhen J, Harrison GF, Zhang G, Chen R, Sun G, et al. Adaptive Admixture of HLA class I Allotypes Enhanced Genetically Determined Strength of Natural Killer Cells in East Asians. Molecular biology and evolution. 2021.
34. Koyro TF, Kraus E, Lunemann S, Hölzemer A, Wulf S, Jung J, et al. Upregulation of HLA-F expression by BK polyomavirus infection induces immune recognition by KIR3DS1-positive natural killer cells. Kidney international. 2020.
35. Townsley E, O'Connor G, Cosgrove C, Woda M, Co M, Thomas SJ, et al. Interaction of a dengue virus NS1-derived peptide with the inhibitory receptor KIR3DL1 on natural killer cells. Clinical and experimental immunology. 2016;183(3):419-30.
36. Beltrame LM, Sell AM, Moliterno RA, Clementino SL, Cardozo DM, Dalalio MM, et al. Influence of KIR genes and their HLA ligands in susceptibility to dengue in a population from southern Brazil. Tissue antigens. 2013;82(6):397-404.
37. Carrington M, Wang S, Martin MP, Gao X, Schiffman M, Cheng J, et al. Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. The Journal of experimental medicine. 2005;201(7):1069-75.
38. Luo M, Czarnecki C, Nebroski M, Kimani J, Bernard N, Plummer FA. KIR3DL1 alleles and their epistatic interactions with human leukocyte antigen class I influence resistance and susceptibility to HIV-1 acquisition in the Pumwani sex worker cohort. AIDS (London, England). 2018;32(7):841-50.
39. Hollenbach JA, Pando MJ, Caillier SJ, Gourraud PA, Oksenberg JR. The killer immunoglobulin-like receptor KIR3DL1 in combination with HLA-Bw4 is protective against multiple sclerosis in African Americans. Genes Immun. 2016;17(3):199-202.
40. Saito H, Hirayama A, Umemura T, Joshita S, Mukawa K, Suga T, et al. Association between KIR-HLA combination and ulcerative colitis and Crohn's disease in a Japanese population. PLoS One. 2018;13(4):e0195778.
41. Umemura T, Joshita S, Saito H, Yoshizawa K, Norman GL, Tanaka E, et al. KIR/HLA genotypes confer susceptibility and progression in patients with autoimmune hepatitis. JHEP reports : innovation in hepatology. 2019;1(5):353-60.
42. Deng Z, Zhao J, Cai S, Qi Y, Yu Q, Martin MP, et al. Natural Killer Cells Offer Differential Protection From Leukemia in Chinese Southern Han. Frontiers in immunology. 2019;10:1646.
43. Fadda L, O'Connor GM, Kumar S, Piechocka-Trocha A, Gardiner CM, Carrington M, et al. Common HIV-1 peptide variants mediate differential binding of KIR3DL1 to HLA-Bw4 molecules. Journal of virology. 2011;85(12):5970-4.
44. Gooneratne SL, Richard J, Lee WS, Finzi A, Kent SJ, Parsons MS. Slaying the Trojan horse: natural killer cells exhibit robust anti-HIV-1 antibody-dependent activation and cytolysis against allogeneic T cells. Journal of virology. 2015;89(1):97-109.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

659 45. Song R, Lisovsky I, Lebouché B, Routy JP, Bruneau J, Bernard NF. HIV protective
660 KIR3DL1/S1-HLA-B genotypes influence NK cell-mediated inhibition of HIV replication in
661 autologous CD4 targets. PLoS pathogens. 2014;10(1):e1003867.
662 46. Koehler RN, Alter G, Tovanabutra S, Saathoff E, Arroyo MA, Walsh AM, et al. Natural
663 killer cell-mediated innate sieve effect on HIV-1: the impact of KIR/HLA polymorphism on HIV-
664 1 subtype-specific acquisition in east Africa. The Journal of infectious diseases. 2013;208(8):1250-
665 4.
666 47. Prakash S, Sarangi AN, Alam S, Sonawane A, Sharma RK, Agrawal S. Putative role of
667 KIR3DL1/3DS1 alleles and HLA-Bw4 ligands with end stage renal disease and long term renal
668 allograft survival. Gene. 2017;637:219-29.
669 48. Stern M, Hadaya K, Hönger G, Martin PY, Steiger J, Hess C, et al. Telomeric rather than
670 centromeric activating KIR genes protect from cytomegalovirus infection after kidney
671 transplantation. American journal of transplantation : official journal of the American Society of
672 Transplantation and the American Society of Transplant Surgeons. 2011;11(6):1302-7.
673 49. Venstrom JM, Gooley TA, Spellman S, Pring J, Malkki M, Dupont B, et al. Donor
674 activating KIR3DS1 is associated with decreased acute GVHD in unrelated allogeneic
675 hematopoietic stem cell transplantation. Blood. 2010;115(15):3162-5.
676 50. Foley BA, De Santis D, Van Beelen E, Lathbury LJ, Christiansen FT, Witt CS. The
677 reactivity of Bw4+ HLA-B and HLA-A alleles with KIR3DL1: implications for patient and donor
678 suitability for haploidentical stem cell transplantations. Blood. 2008;112(2):435-43.
679 51. Vivian JP, Duncan RC, Berry R, O'Connor GM, Reid HH, Beddoe T, et al. Killer cell
680 immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. Nature.
681 2011;479(7373):401-5.
682 52. Boudreau JE, Hsu KC. Natural killer cell education in human health and disease. Current
683 opinion in immunology. 2018;50:102-11.
684 53. Malmberg KJ, Sohlberg E, Goodridge JP, Ljunggren HG. Immune selection during tumor
685 checkpoint inhibition therapy paves way for NK-cell "missing self" recognition. Immunogenetics.
686 2017;69(8-9):547-56.
687 54. Parsons MS, Zipperlen K, Gallant M, Grant M. Killer cell immunoglobulin-like receptor
688 3DL1 licenses CD16-mediated effector functions of natural killer cells. J Leukoc Biol.
689 2010;88(5):905-12.
690 55. Saunders PM, MacLachlan BJ, Widjaja J, Wong SC, Oates CVL, Rossjohn J, et al. The
691 Role of the HLA Class I α2 Helix in Determining Ligand Hierarchy for the Killer Cell Ig-like
692 Receptor 3DL1. J Immunol. 2021;206(4):849-60.
693 56. Saunders PM, Pymm P, Pietra G, Hughes VA, Hitchen C, O'Connor GM, et al. Killer cell
694 immunoglobulin-like receptor 3DL1 polymorphism defines distinct hierarchies of HLA class I
695 recognition. The Journal of experimental medicine. 2016;213(5):791-807.
696 57. Kim S, Sunwoo JB, Yang L, Choi T, Song YJ, French AR, et al. HLA alleles determine
697 differences in human natural killer cell responsiveness and potency. Proceedings of the National
698 Academy of Sciences of the United States of America. 2008;105(8):3053-8.
699 58. Li H, Pascal V, Martin MP, Carrington M, Anderson SK. Genetic control of variegated
700 KIR gene expression: polymorphisms of the bi-directional KIR3DL1 promoter are associated with
701 distinct frequencies of gene expression. PLoS genetics. 2008;4(11):e1000254.
702 59. Yawata M, Yawata N, Draghi M, Little AM, Partheniou F, Parham P. Roles for HLA and
703 KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function.
704 The Journal of experimental medicine. 2006;203(3):633-45.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

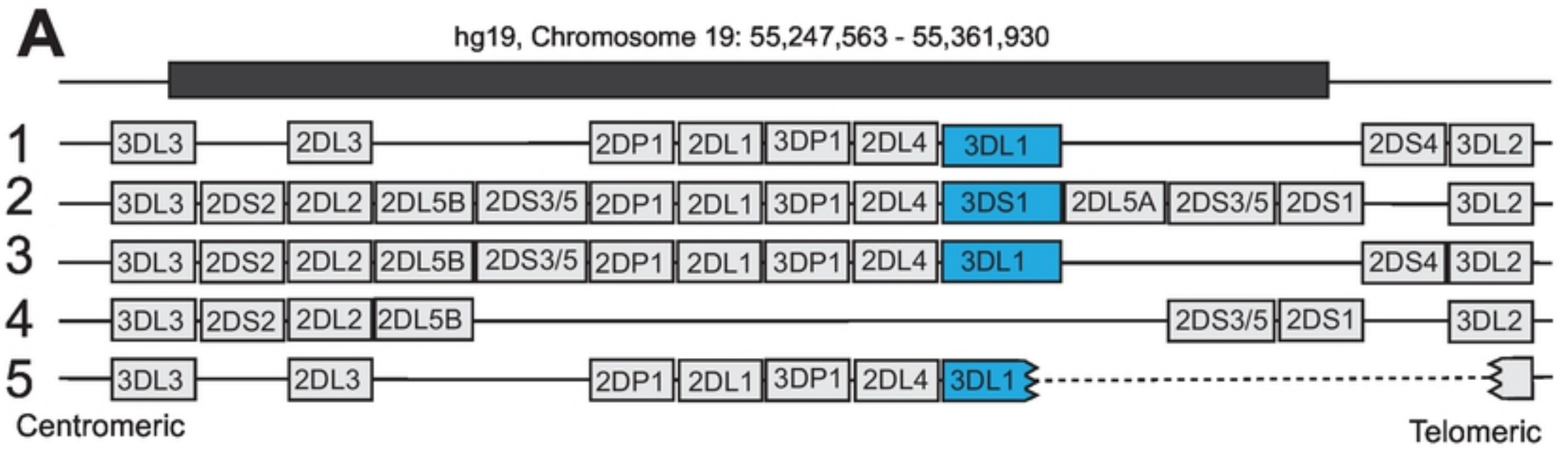
60. Carlomagno S, Falco M, Bono M, Alicata C, Garbarino L, Mazzocco M, et al. KIR3DS1-Mediated Recognition of HLA-*B51: Modulation of KIR3DS1 Responsiveness by Self HLA-B Allotypes and Effect on NK Cell Licensing. Frontiers in immunology. 2017;8:581.
61. O'Connor GM, Yamada E, Rampersaud A, Thomas R, Carrington M, McVicar DW. Analysis of binding of KIR3DS1*014 to HLA suggests distinct evolutionary history of KIR3DS1. J Immunol. 2011;187(5):2162-71.
62. Thananchai H, Gillespie G, Martin MP, Bashirova A, Yawata N, Yawata M, et al. Cutting Edge: Allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. J Immunol. 2007;178(1):33-7.
63. Körner C, Altfeld M. Role of KIR3DS1 in human diseases. Frontiers in immunology. 2012;3:326.
64. Kamya P, Boulet S, Tsoukas CM, Routy JP, Thomas R, Côté P, et al. Receptor-ligand requirements for increased NK cell polyfunctional potential in slow progressors infected with HIV-1 coexpressing KIR3DL1*h/*y and HLA-B*57. Journal of virology. 2011;85(12):5949-60.
65. Berinstein J, Pollock R, Pellett F, Thavaneswaran A, Chandran V, Gladman DD. Association of variably expressed KIR3dl1 alleles with psoriatic disease. Clinical rheumatology. 2017;36(10):2261-6.
66. Boudreau JE, Giglio F, Gooley TA, Stevenson PA, Le Luduec JB, Shaffer BC, et al. KIR3DL1/HLA-B Subtypes Govern Acute Myelogenous Leukemia Relapse After Hematopoietic Cell Transplantation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2017;35(20):2268-78.
67. Boudreau JE, Mulrooney TJ, Le Luduec JB, Barker E, Hsu KC. KIR3DL1 and HLA-B Density and Binding Calibrate NK Education and Response to HIV. J Immunol. 2016;196(8):3398-410.
68. Martin MP, Naranbhai V, Shea PR, Qi Y, Ramsuran V, Vince N, et al. Killer cell immunoglobulin-like receptor 3DL1 variation modifies HLA-B*57 protection against HIV-1. The Journal of clinical investigation. 2018;128(5):1903-12.
69. Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. American journal of human genetics. 2008;82(1):48-56.
70. Li SS, Wang H, Smith A, Zhang B, Zhang XC, Schoch G, et al. Predicting multiallelic genes using unphased and flanking single nucleotide polymorphisms. Genetic epidemiology. 2011;35(2):85-92.
71. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS computational biology. 2013;9(2):e1002877.
72. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One. 2013;8(6):e64683.
73. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG--HLA genotype imputation with attribute bagging. Pharmacogenomics J. 2014;14(2):192-200.
74. Breiman L. Heuristics of instability and stabilization in model selection. Annals of Statistics. 1996;24(6):2350-83.
75. Marin WM, Dandekar R, Augusto DG, Yusufali T, Heyn B, Hofmann J, et al. High-throughput Interpretation of Killer-cell Immunoglobulin-like Receptor Short-read Sequencing Data with PING. bioRxiv. 2021:2021.03.24.436770.
76. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

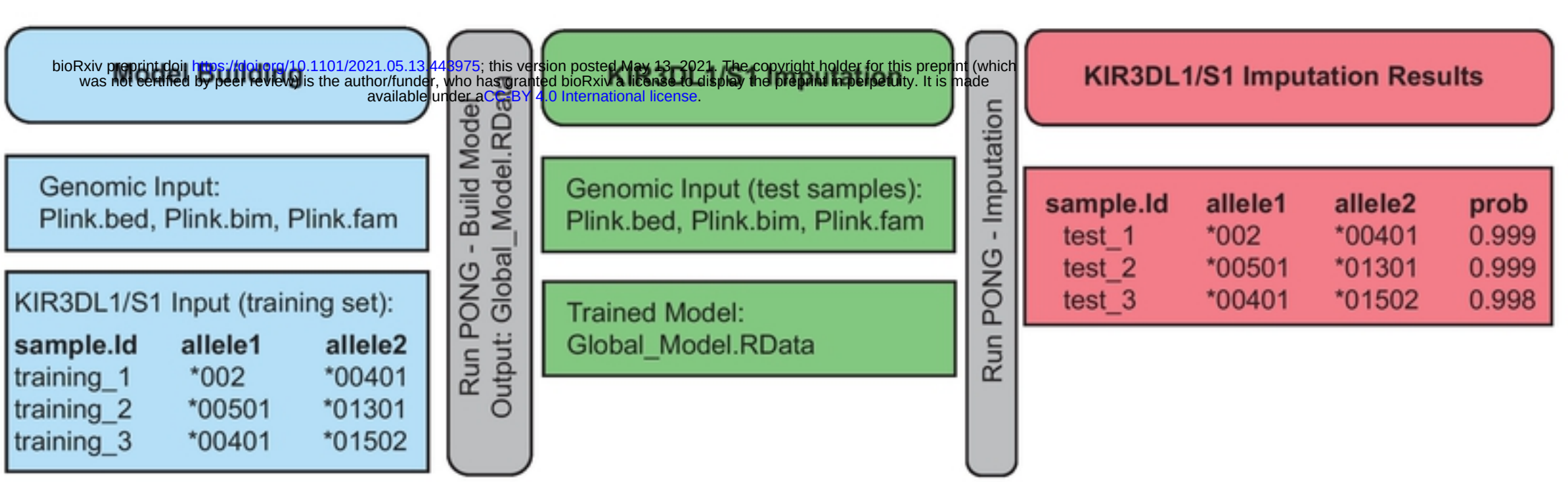
bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

77. Norman PJ, Hollenbach JA, Nemat-Gorgani N, Marin WM, Norberg SJ, Ashouri E, et al. Defining KIR and HLA class I genotypes at highest resolution via high-throughput sequencing. The American Journal of Human Genetics. 2016;99(2):375-91.
78. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nature genetics. 2011;43(12):1193.
79. Lande A, Fluge Ø, Strand EB, Flåm ST, Sosa DD, Mella O, et al. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). Scientific reports. 2020;10(1):5267.
80. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75.
81. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. BioRxiv. 2019:563866.
82. Vukcevic D, Traherne JA, Næss S, Ellinghaus E, Kamatani Y, Dilthey A, et al. Imputation of KIR Types from SNP Variation Data. American journal of human genetics. 2015;97(4):593-607.
83. Crum KA, Logue SE, Curran MD, Middleton D. Development of a PCR-SSOP approach capable of defining the natural killer cell inhibitory receptor (KIR) gene sequence repertoires. Tissue antigens. 2000;56(4):313-26.
84. Norman PJ, Abi-Rached L, Gendzekhadze K, Hammond JA, Moesta AK, Sharma D, et al. Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes. Genome Res. 2009;19(5):757-69.
85. Team RC. R: A language and environment for statistical computing. 2013.
86. Caocci G, Martino B, Greco M, Abruzzese E, Trawinska MM, Lai S, et al. Killer immunoglobulin-like receptors can predict TKI treatment-free remission in chronic myeloid leukemia patients. Experimental hematology. 2015;43(12):1015-8.e1.
87. Erbe AK, Wang W, Reville PK, Carmichael L, Kim K, Mendonca EA, et al. HLA-Bw4-I-80 Isoform Differentially Influences Clinical Outcome As Compared to HLA-Bw4-T-80 and HLA-A-Bw4 Isoforms in Rituximab or Dinutuximab-Based Cancer Immunotherapy. Frontiers in immunology. 2017;8:675.
88. Gabriel IH, Sergeant R, Szydlo R, Apperley JF, DeLavallade H, Alsuliman A, et al. Interaction between KIR3DS1 and HLA-Bw4 predicts for progression-free survival after autologous stem cell transplantation in patients with multiple myeloma. Blood. 2010;116(12):2033-9.
89. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet. 2002;31(4):429-34.
90. Alicata C, Pende D, Meazza R, Canevali P, Loiacono F, Bertaina A, et al. Hematopoietic stem cell transplantation: Improving alloreactive Bw4 donor selection by genotyping codon 86 of KIR3DL1/S1. European journal of immunology. 2016;46(6):1511-7.
91. Gagne K, Busson M, Bignon JD, Balère-Appert ML, Loiseau P, Dormoy A, et al. Donor KIR3DL1/3DS1 gene and recipient Bw4 KIR ligand as prognostic markers for outcome in unrelated hematopoietic stem cell transplantation. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2009;15(11):1366-75.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

794 92. Ureshino H, Shindo T, Kojima H, Kusunoki Y, Miyazaki Y, Tanaka H, et al. Allelic
795 Polymorphisms of KIRs and HLAs Predict Favorable Responses to Tyrosine Kinase Inhibitors in
796 CML. Cancer immunology research. 2018;6(6):745-54.
797 93. Díaz-Peña R, Vidal-Castiñeira JR, Alonso-Arias R, Suarez-Alvarez B, Vicario JL, Solana
798 R, et al. Association of the KIR3DS1* 013 and KIR3DL1* 004 alleles with susceptibility to
799 ankylosing spondylitis. Arthritis & Rheumatism. 2010;62(4):1000-6.
800 94. Pappas DJ, Lizee A, Paunic V, Beutner KR, Motyer A, Vukcevic D, et al. Significant
801 variation between SNP-based HLA imputations in diverse populations: the last mile is the hardest.
802 Pharmacogenomics J. 2018;18(3):367-76.
803 95. Vollger MR, Dishuck PC, Sorensen M, Welch AE, Dang V, Dougherty ML, et al. Long-
804 read sequence and assembly of segmental duplications. Nature methods. 2019;16(1):88-94.
805 96. Mostovoy Y, Yilmaz F, Chow SK, Chu C, Lin C, Geiger EA, et al. Genome mapping
806 resolves structural variation within segmental duplications associated with
807 microdeletion/microduplication syndromes. bioRxiv. 2020.
808

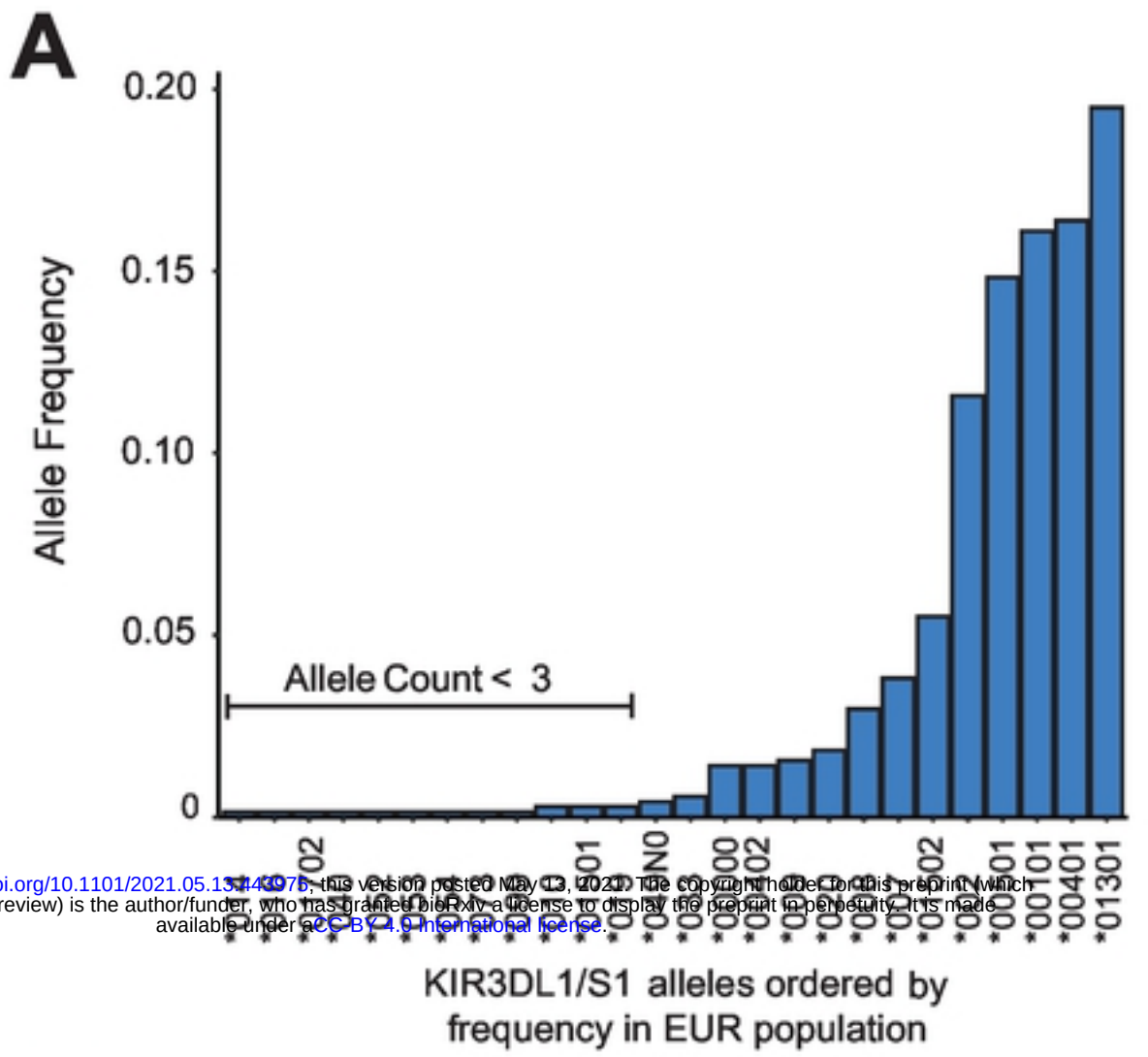




bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC BY 4.0 International license.

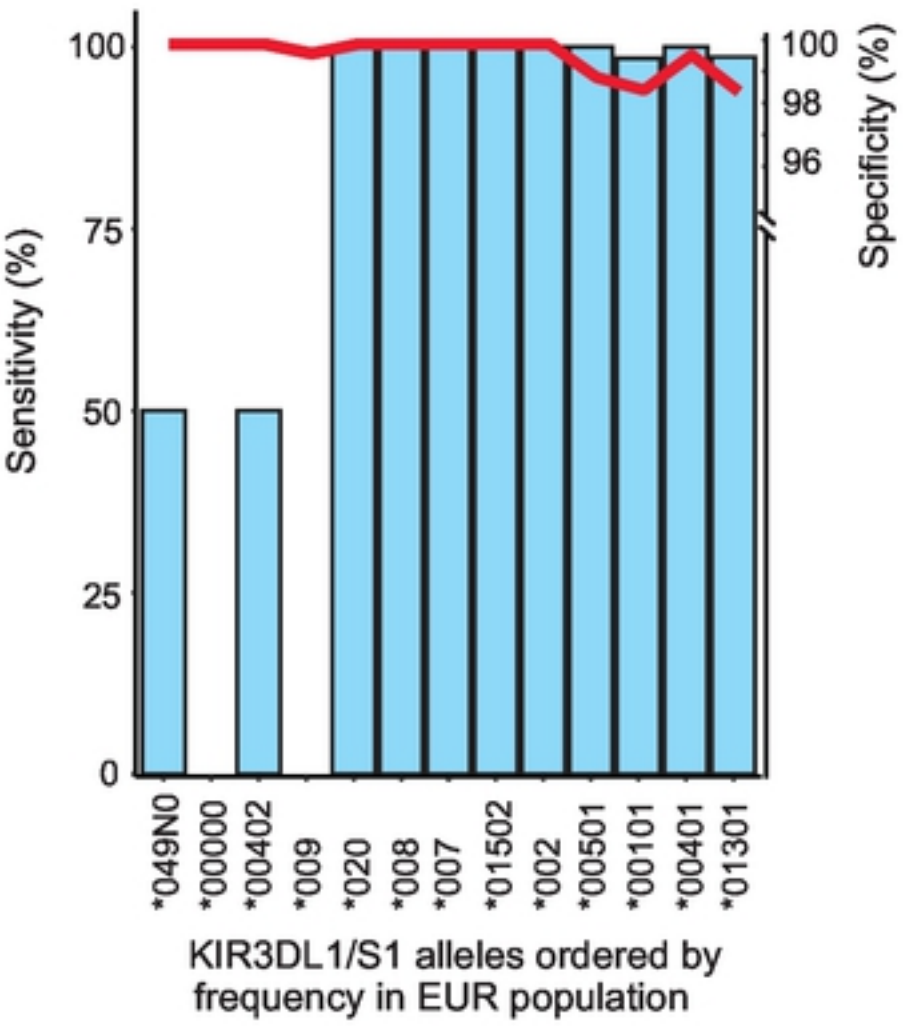
A



B

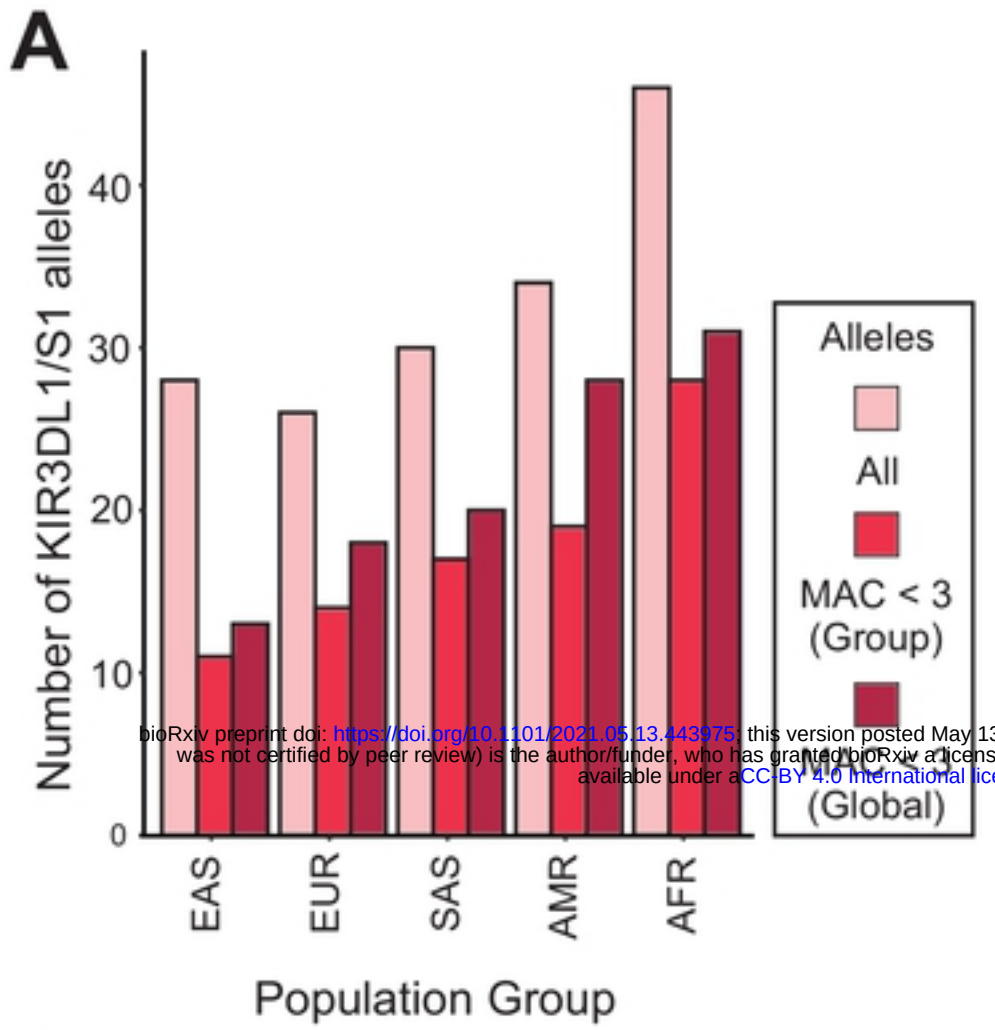
Filtered	Input Threshold	Building Time (hr : min)	Accuracy (%)
SNPs	none	1:24	92
	MAC < 2	1:13	92
	MAC < 3	1:06	92
	MAF < 1%	1:09	92
	MAF < 5%	0:29	91
KIR3DL1/S1 alleles	MAC < 3	0:49	96

C



bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

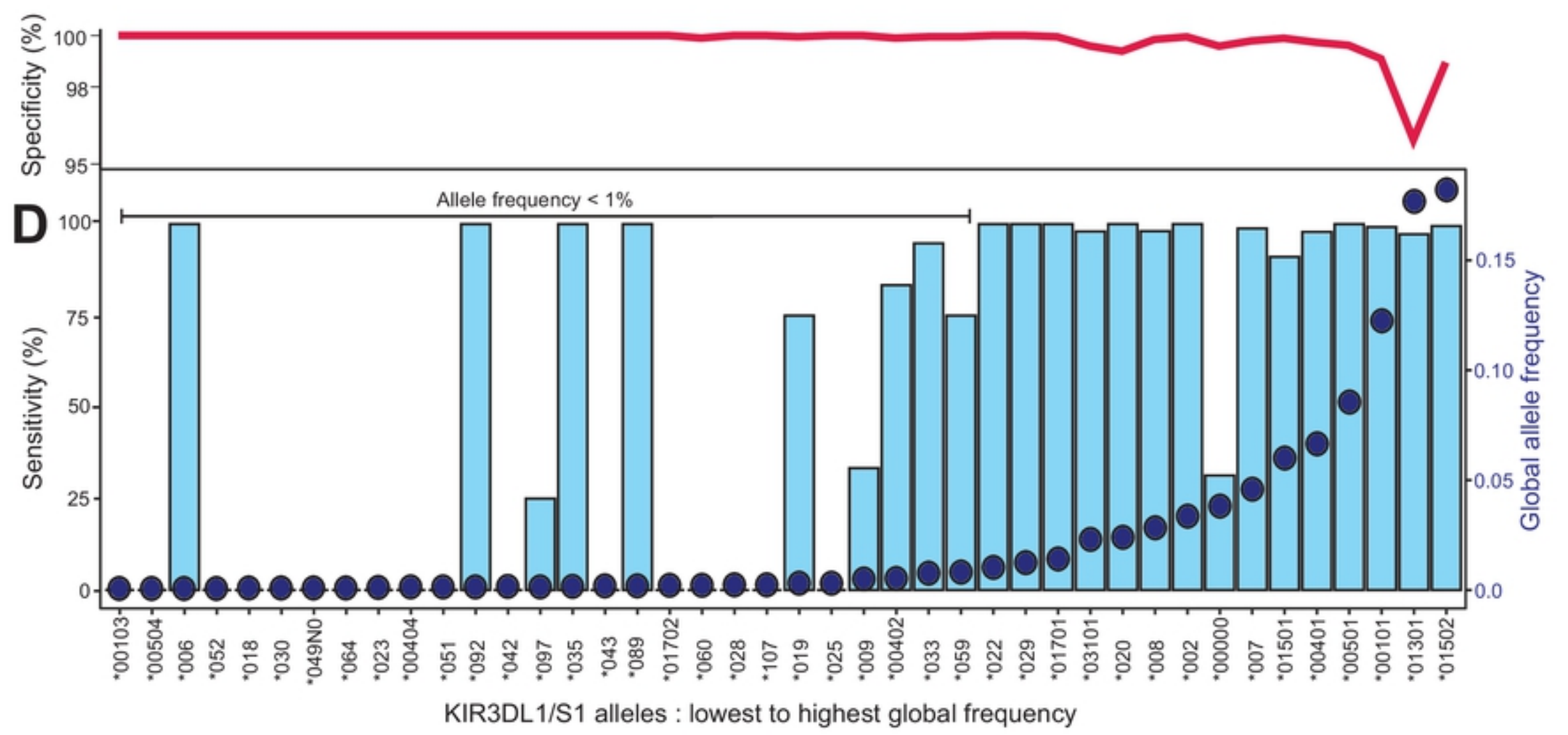
A



B

Population Group	Model Build (accuracy) Within Group	Global
African (AFR)	88.2%	89.2%
American (AMR)	88.6%	92.5%
East Asian (EAS)	96.6%	97.2%
European (EUR)	94.0%	95.0%
South Asian (SAS)	87.8%	89.0%
Global		92.3%

C



D

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.443975; this version posted May 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.