

# Novel high-resolution multi-ethnic HLA imputation reference panels constructed based on high-coverage whole-genome sequencing data

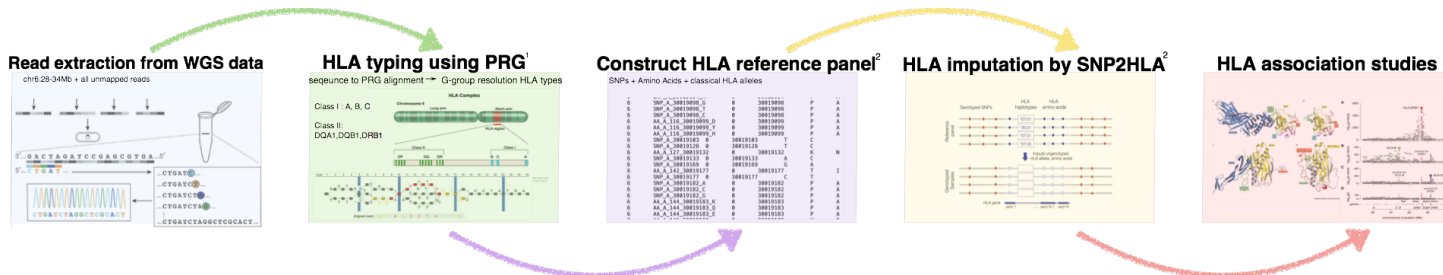
Y.Luo<sup>1,2</sup>, M. Gutierrez-Arcelus<sup>1,2</sup>, A. Metspalu<sup>3</sup>, J.G. Wilson<sup>4</sup>, S. Kathiresan<sup>5</sup>, T. Esko<sup>2,3</sup>, S. Raychaudhuri<sup>1,2</sup>, the TOPMed project



## Background & Study Design

The human leukocyte antigen (HLA) region harbors genes that are crucial to a wide range of human diseases. However it has been challenging to identify and fine-map HLA associations from GWAS data. HLA imputation approaches have been restricted by limitations in reference panels to European populations and to amino acid (4-digit) resolution. Here, we infer HLA types from large scale whole-genome sequencing (WGS) data, and build the **first multi-ethnic high resolution (at G-group resolution) HLA reference panels**. These references can be applied to a wide range of worldwide data sets, where HLA imputation was not previously possible. The computational strategies that we present can also be applied in any high coverage WGS data, thereby allowing users to build their own HLA reference panels.

## Workflow of Constructing Multi-ethnic High Resolution HLA Reference Panel

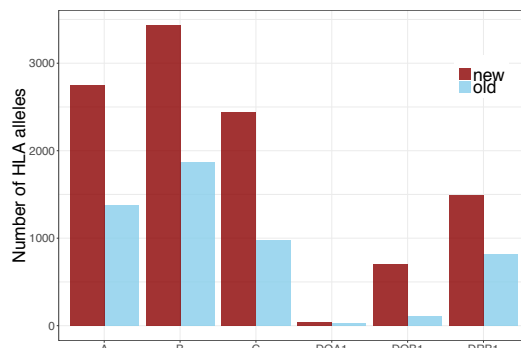


## Reference Panel Description and Validation

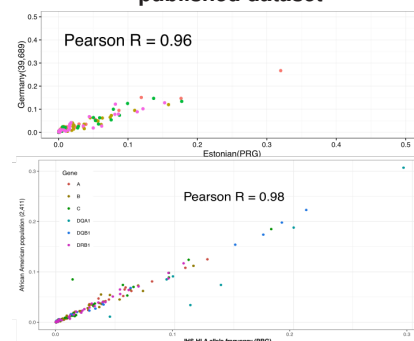
### Reference description

Ancestry	Size	Platform
Estonian	2,244	WGS@25x
African-American	3,076	WGS@25x
British	5,225	genotyping

### New enriched HLA dictionary



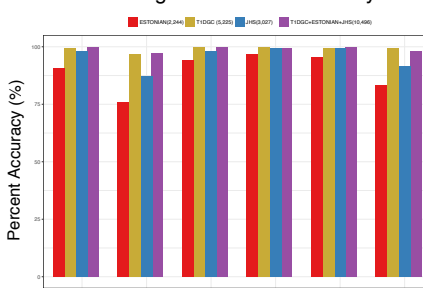
### High frequency concordance with published dataset



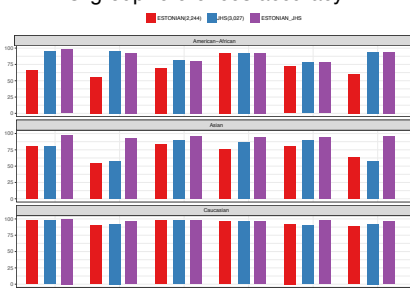
## Imputation Performance Validation

- Imputed into 2,291 Long Island samples from GaP registry
- 189,672 variants on ImmunoChip
- a multi-ethnic cohort
- 75 individuals (25 Asian, 25 European and 25 African) went on for gold-standard HLA sequencing
- validation of six classical HLA genes - A, B, C, DQA1, DQB1, DRB1

### 4-digit references accuracy

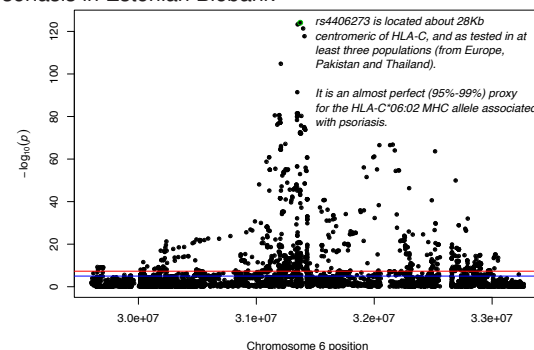


### G-group references accuracy



## Fine-Mapping HLA Association in Estonian Biobank

Confirmation of previously established HLA-C association of Psoriasis in Estonian Biobank



## Contact Information

Yang Luo, PhD  
yangluo@broadinstitute.org



## References

- Dilthey, A. T. et al. High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. PLoS Comput. Biol. 12, e1005151 (2016)
- Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 8, e64683 (2013)

<sup>1</sup>Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.; <sup>2</sup> Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; <sup>3</sup> Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; <sup>4</sup> Physiology and Biophysics, University of Mississippi Medical Center; <sup>5</sup> Center for Genomic Medicine, Massachusetts General Hospital