

Reconstructing ancient HLA haplotypes, using variation graphs and pangenome assemblies

Alba Refoyo Martínez, Fernando Racimo, Evan K. Irving-Pease
Lundbeck GeoGenetics Centre, Globe Institute, University of Copenhagen

UNIVERSITY OF COPENHAGEN

GEOGENETICS CENTRE
LUNDBECK FOUNDATION

Evolution of thousands of traits over the last 12k years

- Western Eurasia underwent a dramatic social transformation with agricultural development, and migration from Anatolian people to Europe, which is known as the Mesolithic–Neolithic transition.
- We study the genomic impact of this transition across time and space by reconstructing polygenic scores for phenotypes of ancient hunter-gatherers, farmers and steppe nomad populations, using effect size estimates from genome-wide association studies conducted on the UK Biobank

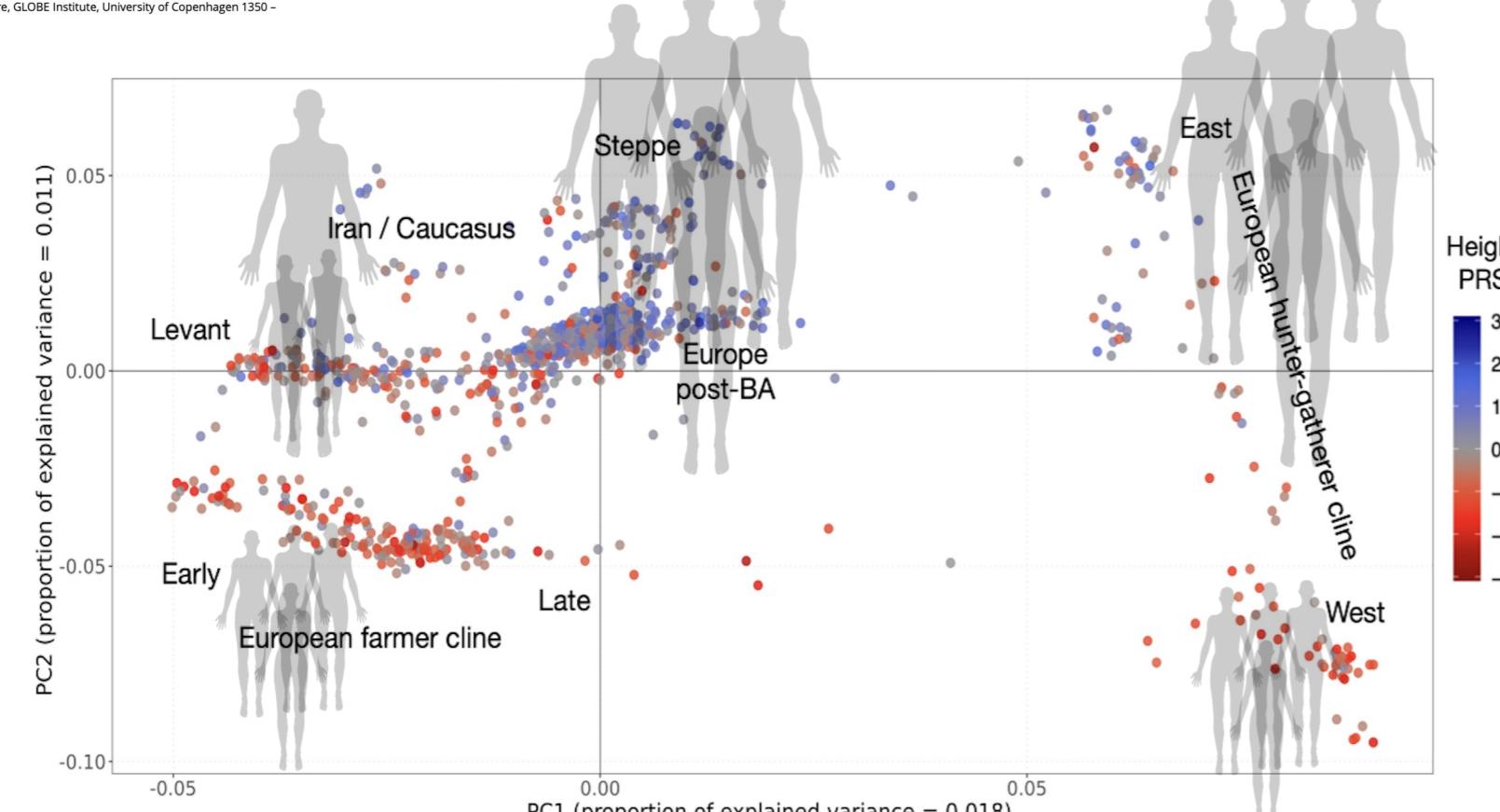


RESEARCH ARTICLE
Open Access
Open Peer-Review
Open Code

How robust are cross-population signatures of polygenic adaptation in humans?

Alba Refoyo-Martínez^{1,*}, Siyang Liu^{2,3}, Anja Moltke Jørgensen⁴, Xin Jin⁵, Anders Albrechtsen⁵, Alicia R. Martin^{5,6,7}, Fernando Racimo¹

¹ Lundbeck GeoGenetics Centre, GLOBE Institute, University of Copenhagen 1350 – Denmark; ² Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States; ³ Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States; ⁴ Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States; ⁵ Lundbeck GeoGenetics Centre, GLOBE Institute, University of Copenhagen 1350 – Denmark; ⁶ Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States; ⁷ Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States



- GWAS tells us how individual locations in the genome are associated with different disease traits.
- However, genomic variation gathered from people of European ancestry do not reflect all human genomic differences
- Important to use a GWAS panel from a closely related ancestry to the target population
- Polygenic risk scores (PRS) provide information about the risk of an individual or a population for having a trait or disease

HLA region on ancient samples

Number of HLA alleles	
HLA Class I Alleles	23002
HLA Class II Alleles	8673
HLA Alleles	31675
Other non-HLA Alleles	655

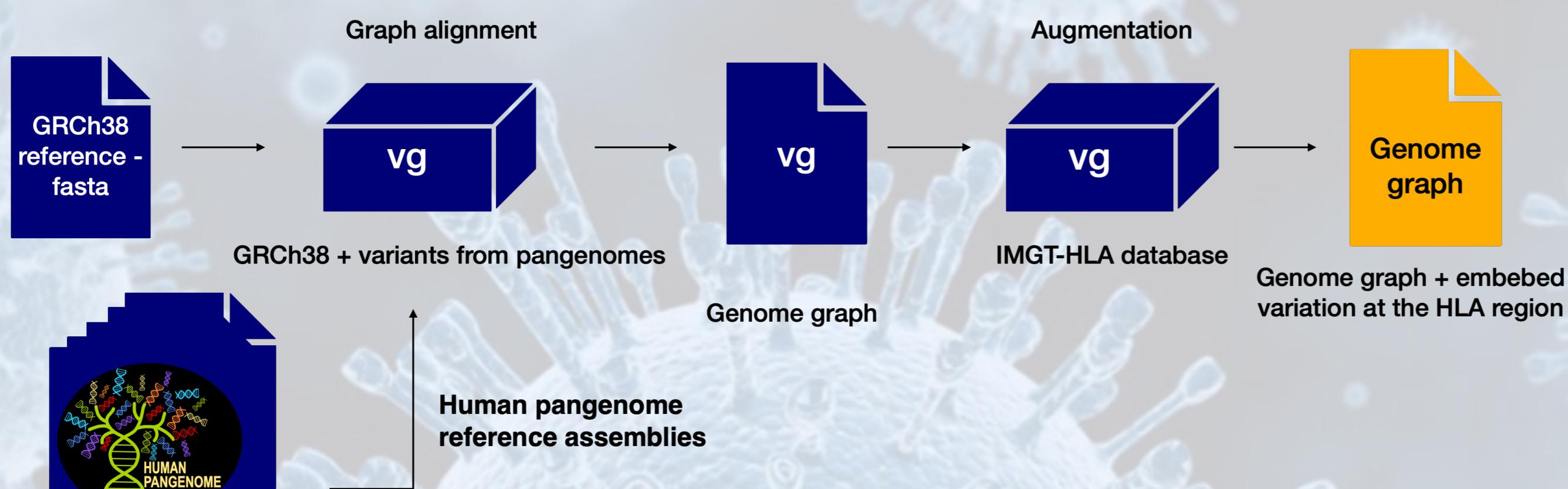
HLA Class I						
Gene	A	B	C	E	F	G
Allele	7114	8464	6855	278	47	92

HLA Class II													
Gene	DRA	DRB	DQA1	DQA2	DQB1	OPA1	OPA2	DPB1	DPB2	DMA	DMB	DOA	DOB
Allele	32	3841	363	40	2136	315	5	1890	6	7	13	12	13



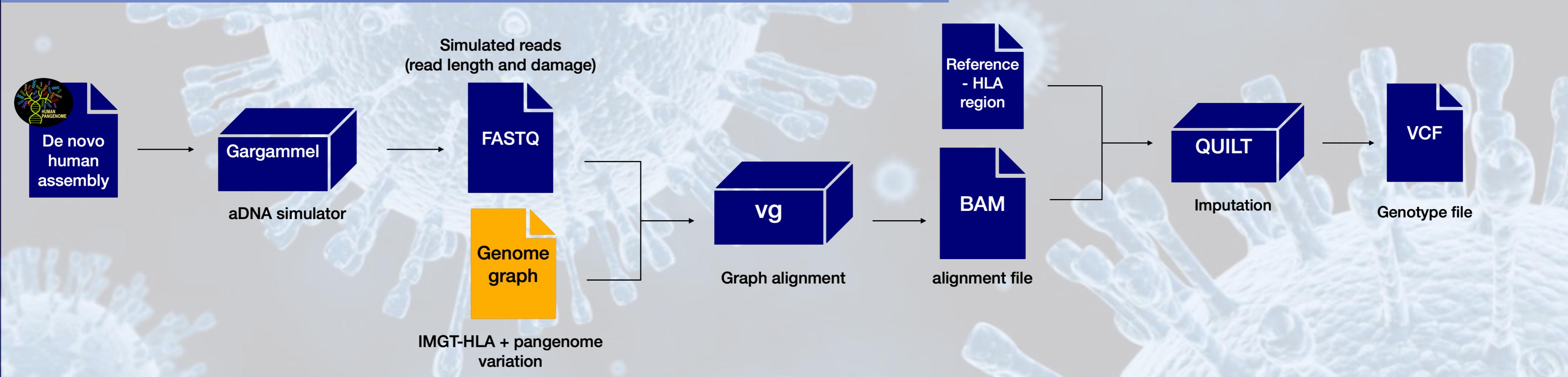
- HLA region encodes key surface proteins for the human body's defense against pathogens and immunity
- We want to get a better understanding of the immune system, its evolutionary struggle against foreign microorganism and their functional associations
- Looking at how trait-associated variants on HLA genes evolve and their haplotype distribution across different time periods and space
- Problem: Highly variable region containing complex structural variants with high sequence similarity between the polymorphic alleles that makes the alignment difficult, and therefore, the typing.
- General aim: Optimize ancient short read data alignments to improve HLA alignment and typing on ancient DNA samples using graph alignments

Augmentation with pangenesomes and IPD-IMGT/HLA database



- Pangenesomes provide additional genetic information as they represent alternative sequences not present in linear reference genomes
- Variation graphs improve the representation of structurally complex regions and reduces reference bias when aligning low-coverage, highly-damaged ancient DNA sequencing reads
- LOOV Genome graphs: leave-one-out-cross-validation

Alignment to the VG graph and imputation using QUILT



Conclusion

- Ancient genomes provide a highly detailed picture of changes in genetic variation over several millennia, and of ancestral relationships between ancient humans and humans living in Western Eurasia today
- We found that the most significantly over-dispersed scores correspond to variants associated with traits related to pigmentation, anthropometric traits and disorders associated to diet and sugar levels, suggesting strong population trait differences preceding the transition, followed by trait homogenization via subsequent admixture.
- Increase power combining graph aligners and pangenesomes

Acknowledgements

Meso-neo consortium, Geogenetics Centre and the Racimo group.

AR-M is supported by the Lundbeck Foundation GeoGenetics Centre grant (R302-2018-2155) and a Novo Nordisk grant to the GeoGenetics Centre (NNF18SA0035006).



DOWNLOAD ME



alba.martinez@sund.ku.dk



@albarma