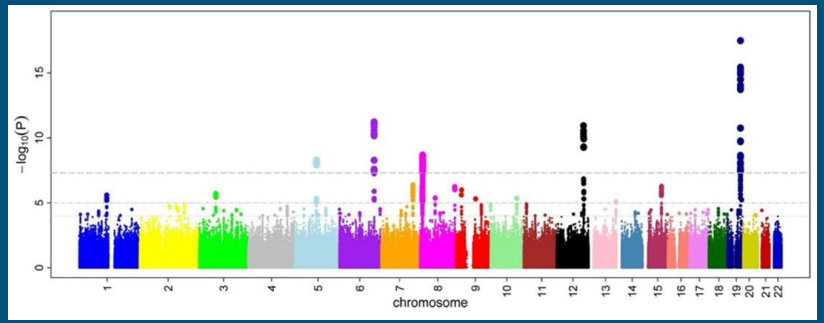
Methods for GWAS on admixed individuals

David Bass, 8 Dec. 2022

GWAS

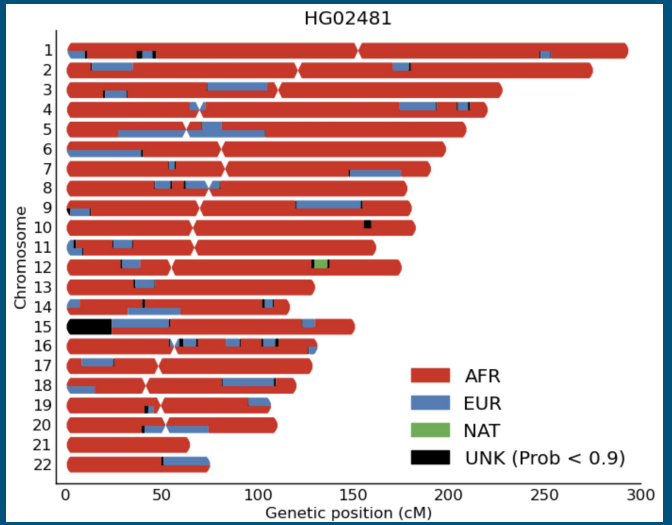
- Genome-wide association studies (GWAS) are a class of methods for describing correlations between single nucleotide polymorphisms (SNPs) and traits
- GWAS is invaluable for genomic medicine because it allows us to describe potential genetic causes of disease
- It is important that we are able to perform GWAS for all individuals, regardless of their ancestry



Ikram et al. (2010)

GWAS on admixed individuals

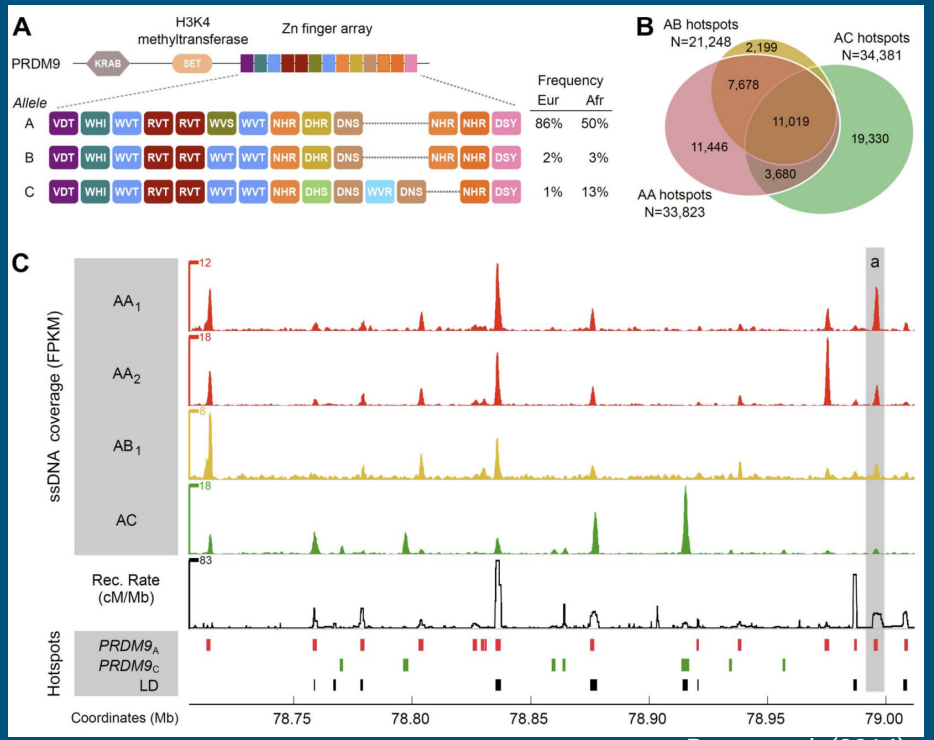
- Traditionally, GWAS requires the assumption that individuals belong to a distinct ancestry group
- This assumption fails for mixed-race individuals and individuals from admixed populations, e.g. African Americans and Latinos, leading to false positives



Martin et al. (2017)

Local ancestry inference

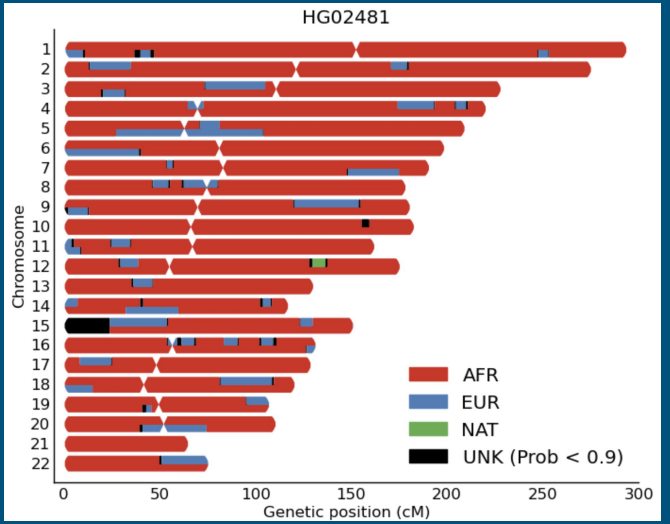
- Conserved patterns of recombination in humans allow for accurate haplotype inference of haplotypes via phasing
- We can resolve “local” ancestry at the haplotypic scale via local ancestry inference (LAI) methods like RFMIX



Pratto et al. (2014)

GWAS via local ancestry

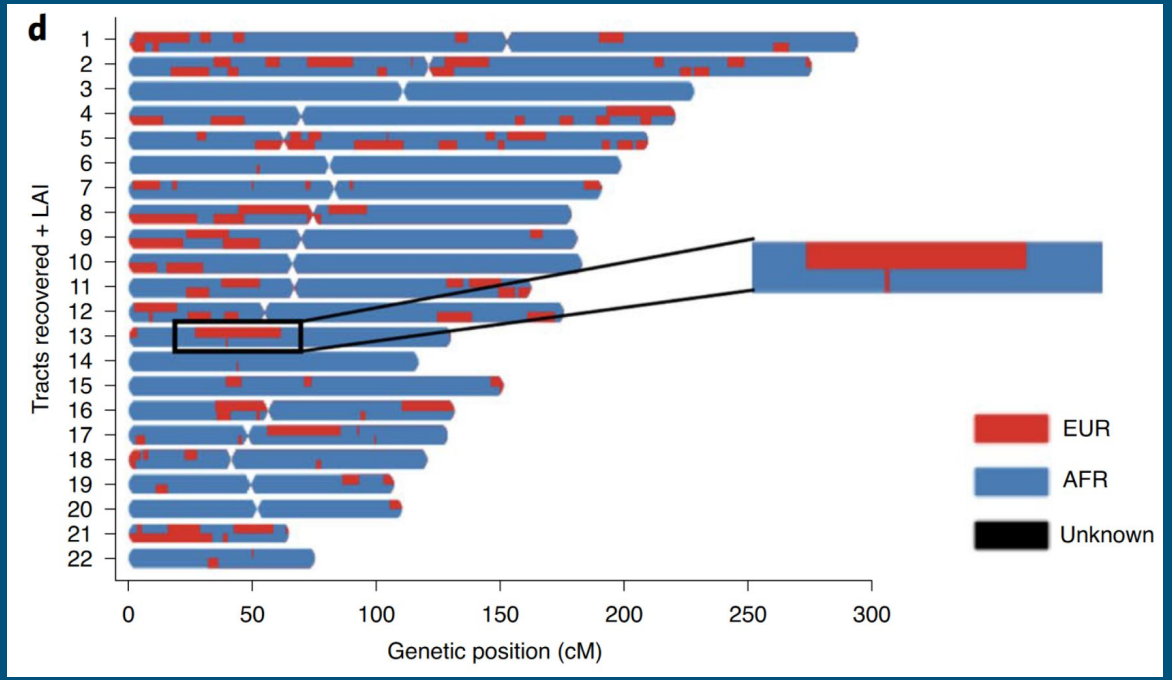
- With local ancestry information, we can subdivide the genome into subsets of haplotype blocks belonging to single constituent ancestries
- Methods like Tractor and asaMAP model local ancestry in order to make GWAS viable on all individuals



Martin et al. (2017)

Local ancestry inference methods

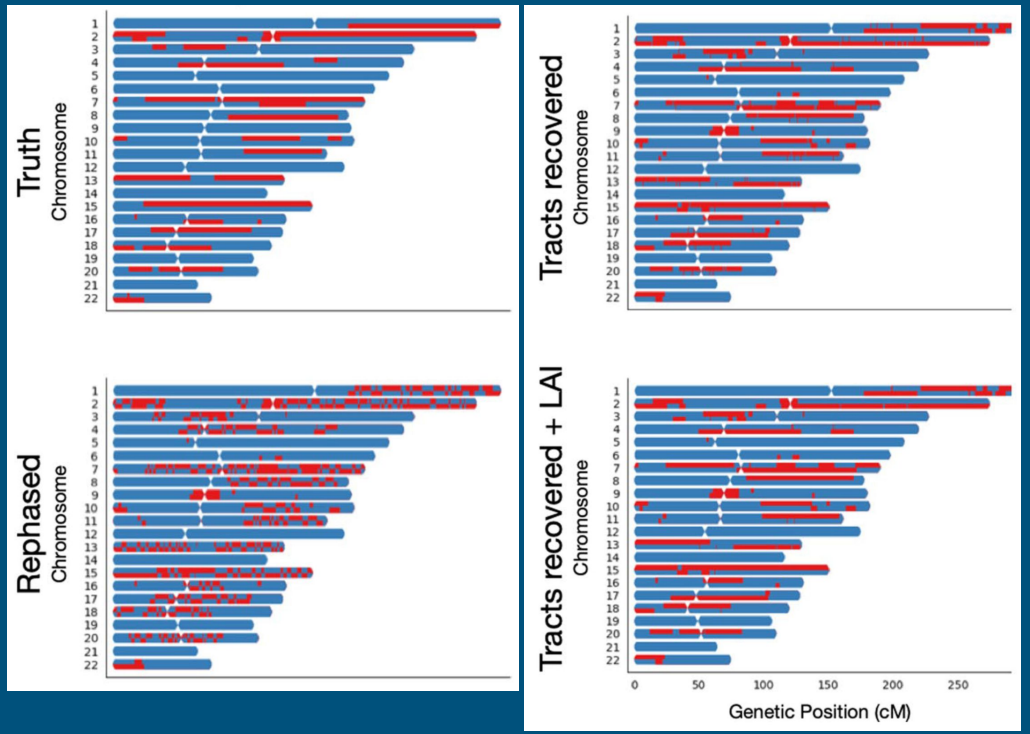
- RFMIX is an LAI method that uses random forests trained on reference panels
- LAI methods require phased genome data as input



Atkinson et al. (2021)

Improving RFMIX with iterated phasing

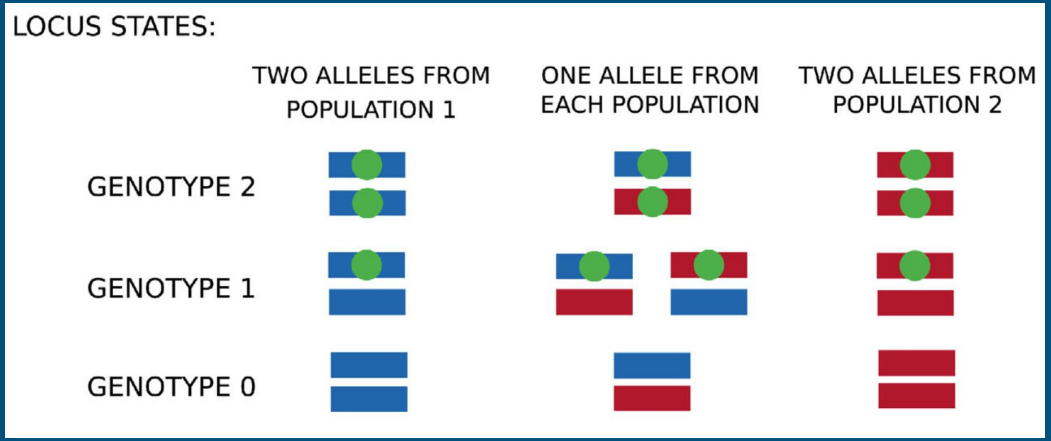
- Haplotype phasing often erroneously “chops up” haplotype blocks via “switch errors”
- Atkinson et al. (2021) perform phasing followed by LAI twice, with switch errors in the second round of phasing corrected with local ancestry information



Atkinson et al. (2021)

Ancestry-specific allele effects in asaMAP

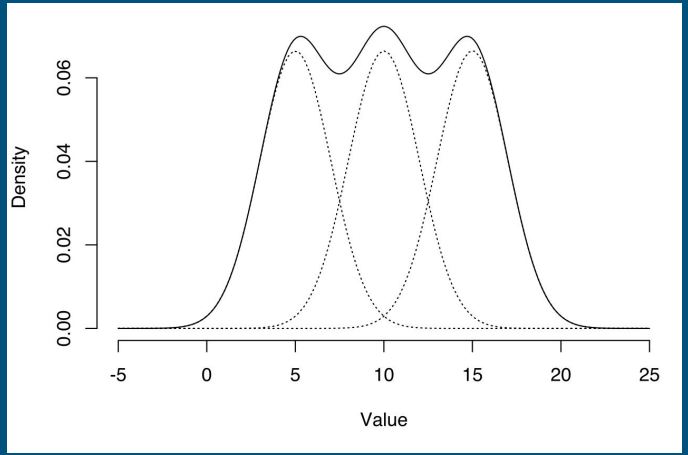
- We typically treat genotype and ancestry as distinct variables
- asaMap incorporates ancestry-specific differences in allele effect by considering 11 “locus states,” each of which is a distinguishable combination of genotype and ancestry



Skotte et al. (2019)

Refresher on mixture distributions

- A mixture distribution is defined by mixture weights (probabilities) and mixture components (probability distributions)
- Method for generating a random value from a mixture distribution:
  - Choose a mixture component with probability equal to its mixture weight
  - Generate a random value from that mixture distribution
- asaMAP models trait values with a mixture distribution in which each mixture component corresponds to a locus state



Smason79 (2012)

asaMAP's statistical model

- asaMAP describes the distribution of a trait as a mixture distribution with normally distributed locus state-specific mixture components
  - η_i is the trait value in individual *i*
  - $x_j(s)$ is the count of effect alleles at locus state *s* from population *j*
  - z_{ic} is a covariate, like population structure, in individual *i*
- This model assumes additive allele effects, but it can be easily adapted to describe dominant/recessive alleles

$$\eta_i = \alpha + \beta_1 x_1(s) + \beta_2 x_2(s) + \sum_c \gamma_c z_{ic}$$

Skotte et al. (2019)

asaMAP's statistical model

Locus state-specific mixture weights are calculated as the conditional probability of locus state *s* given the individual's genotype *g*, ancestry-specific allele frequencies *f*, and admixture proportions *q*

$$
\begin{aligned}
p(s|g, f, q) &= p(a, t|g, f, q) \\
&= p(a|t, g, f, q)p(t|g, f, q) \\
&= p(a|t, f, q)p(t|g).
\end{aligned}
$$

Skotte et al. (2019)

Separating ancestry from allele in Tractor

- While asaMAP combines genotype and ancestry into locus state, Tractor considers them with separate variables in its linear model
  - *Y* is the trait value in a given individual

$$Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots + \beta_k X_k$$

X_1 is the number of copies of ancestry 1 at the site

X_2 is the number of copies of the risk allele coming from ancestry 1

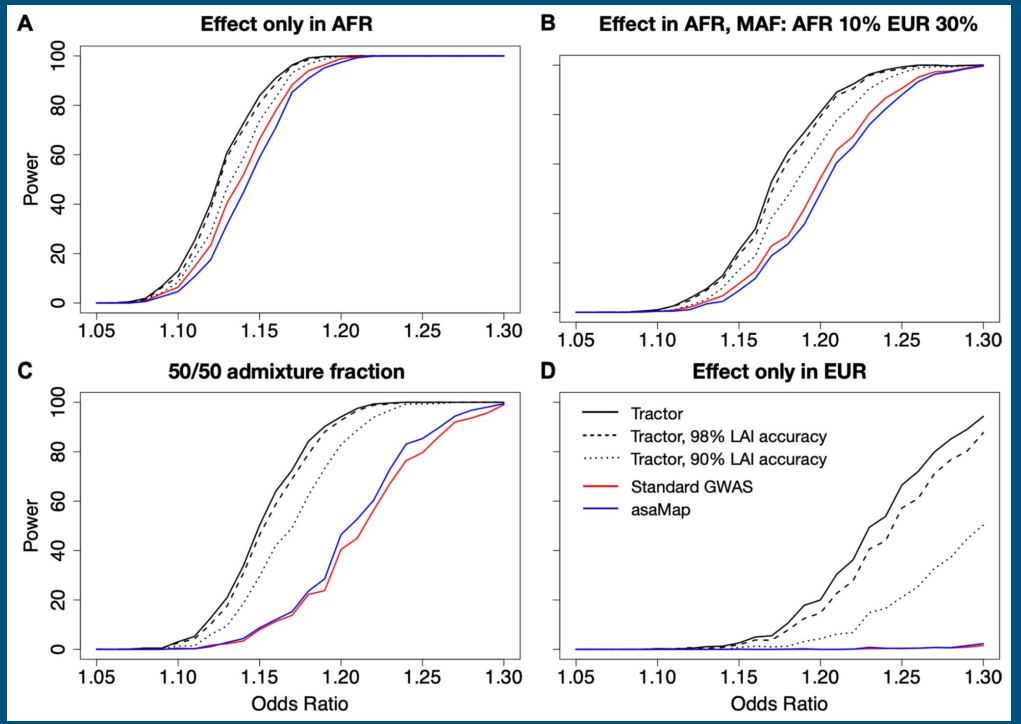
X_3 is copies coming from ancestry 2

X_4 to X_k are other covariates (such as PCs)

Broad Institute (2020)

Statistical power of asaMAP and Tractor

- Tractor significantly outperforms both asaMAP and standard GWAS across several simulated scenarios, even when LAI accuracy is suppressed
- Tractor's power gains are especially large when admixture is maximized (C) and when the effect allele only causes the trait in the rare ancestry (D)



Atkinson et al. (2021)

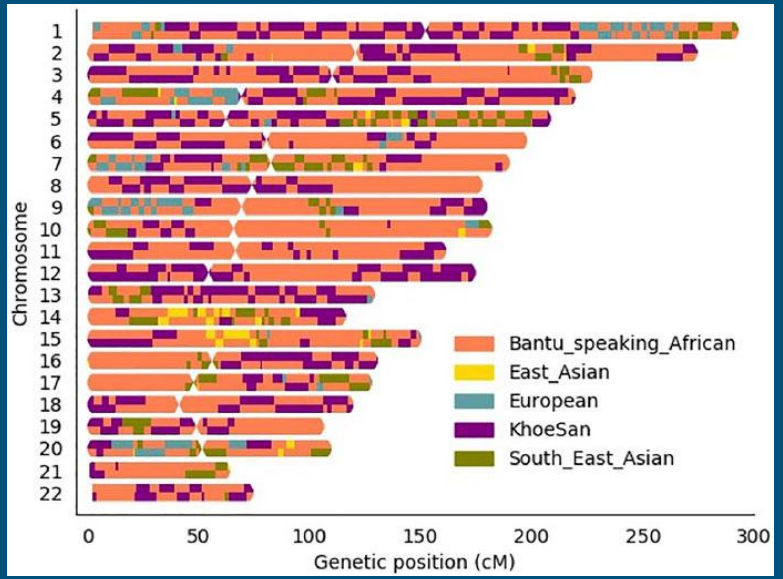
Comparing asaMAP and Tractor

- Overall, Tractor is significantly more powerful than asaMAP
- Despite not taking advantage of the locus states introduced by asaMAP, Tractor still achieves highly accurate estimates of ancestry-specific effects
- Tractor is better than asaMAP at identifying the locations of causal variants
- Both methods improve upon admixture mapping, a previous approach that performs poorly when trait value is not stratified by ancestry

Atkinson (2022)

Room for improvement

- Quality of LAI is poor for individuals with more than two ancestries
- asaMAP and Tractor employ simple models of allele effect
- The performance of causal inference methods, like fine-mapping, on the results of either method have yet to be studied
- Both methods enforce conformity to pre-established ancestry groups



Swart et al. (2021)

Works cited

Atkinson, E. G., Maihofer, A. X., Kanai, M., Martin, A. R., Karczewski, K. J., Santoro, M. L., Ulirsch, J. C., Kamatani, Y., Okada, Y., Finucane, H. K., Koenen, K. C., Nievergelt, C. M., Daly, M. J., & Neale, B. M. (2021). Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nature Genetics*, *53*(2), 195–204. https://doi.org/10.1038/s41588-020-00766-y

Broad Institute (Director). (2020, October 6). *Stanley Center Primer: An Introduction to GWAS + ancestry-specific gene discovery with Tractor*. https://www.youtube.com/watch?v=ngQYuk_1y9g

Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., & Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. *PLoS Genetics*, *14*(4), e1007308. https://doi.org/10.1371/journal.pgen.1007308

Duello, T. M., Rivedal, S., Wickland, C., & Weller, A. (2021). Race and genetics versus 'race' in genetics. *Evolution, Medicine, and Public Health*, *9*(1), 232–245. https://doi.org/10.1093/emph/eoab018

Gignoux, C. R., Torgerson, D. G., Pino-Yanes, M., Uricchio, L. H., Galanter, J., Roth, L. A., Eng, C., Hu, D., Nguyen, E. A., Huntsman, S., Mathias, R. A., Kumar, R., Rodriguez-Santana, J., Thakur, N., Oh, S. S., McGarry, M., Moreno-Estrada, A., Sandoval, K., Winkler, C. A., … Burchard, E. G. (2019). An admixture mapping meta-analysis implicates genetic variation at 18q21 with asthma susceptibility in Latinos. *The Journal of Allergy and Clinical Immunology*, *143*(3), 957–969. https://doi.org/10.1016/j.jaci.2016.08.057

Graham, S. E., Clarke, S. L., Wu, K.-H. H., Kanoni, S., Zajac, G. J. M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T. W., Locke, A. E., Marouli, E., Hwang, M. Y., Han, S., Narita, A., Choudhury, A., Bentley, A. R., Ekoru, K., Verma, A., … Willer, C. J. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature*, *600*(7890), Article 7890. https://doi.org/10.1038/s41586-021-04064-3

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Home · Atkinson-Lab/Tractor Wiki. (n.d.). GitHub. Retrieved December 8, 2022, from https://github.com/Atkinson-Lab/Tractor

Korunes, K. L., & Goldberg, A. (2021). Human genetic admixture. *PLoS Genetics*, *17*(3), e1009374. https://doi.org/10.1371/journal.pgen.1009374

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *American Journal of Human Genetics*, *93*(2), 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., & Kenny, E. E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics*, *100*(4), 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Mixture distribution. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Mixture_distribution&oldid=1091582237

Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G. V., & Camerini-Otero, R. D. (2014). Recombination initiation maps of individual human genomes. *Science (New York, N.Y.)*, *346*(6211), 1256442. https://doi.org/10.1126/science.1256442

Skotte, L., Jørsboe, E., Korneliussen, T. S., Moltke, I., & Albrechtsen, A. (2019). Ancestry-specific association mapping in admixed populations. *Genetic Epidemiology*, *43*(5), 506–521. https://doi.org/10.1002/gepi.22200

Swart, Y., Uren, C., van Helden, P. D., Hoal, E. G., & Möller, M. (2021). Local Ancestry Adjusted Allelic Association Analysis Robustly Captures Tuberculosis Susceptibility Loci. *Frontiers in Genetics*, *12*. https://www.frontiersin.org/articles/10.3389/fgene.2021.716558

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-021-00056-9

Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., & Jin, L. (2002). Distribution of Recombination Crossovers and the Origin of Haplotype Blocks: The Interplay of Population History, Recombination, and Mutation. *American Journal of Human Genetics*, *71*(5), 1227–1234.

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., Highland, H. M., Patel, Y. M., Sorokin, E. P., Avery, C. L., Belbin, G. M., Bien, S. A., Cheng, I., Cullina, S., Hodonsky, C. J., Hu, Y., Huckins, L. M., Jeff, J., Justice, A. E., … Carlson, C. S. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, *570*(7762), Article 7762. https://doi.org/10.1038/s41586-019-1310-4

Conclusions

- In order to achieve equity in genomic medicine, we need good methods for GWAS on admixed individuals
- Even if we fix the Eurocentric bias in genomic reference panels, tools like asaMAP and Tractor will still be useful for leveraging information about ancestries to study the results of their admixture
- Better methods for phasing, LAI, and GWAS will improve our predictions of the genetic sources of disease in admixed individuals