David Bass

Traditional genome-wide association studies (GWAS) require that individuals be grouped by ancestry (Uffelmann et al. 2021), making them inapplicable to individuals that are admixed, i.e. individuals belonging to multiple ancestry groups (Korunes and Goldberg 2021). Despite the lack of genetic evidence for racial categories (Duello et al. 2021), the existence of distinct human subpopulations is a necessary assumption for many statistical models (e.g. Gutenkunst et al. 2009). Fortunately, the existence of haplotypes resulting from patterns of recombination (Wang et al. 2002) alleviates the need to assign individuals to a single ancestry group by enabling the resolution of ancestry at a local, as opposed to whole-genome, scale. Tools like RFMix perform local-ancestry inference (LAI), in which ancestry is inferred at each of many tracts of the genome associated with haplotypes (Maples et al. 2013). For example, while there may not be a genomic reference panel to which a mixed-race individual can be compared in GWAS, LAI can assign single-ancestry identities to windows of their genome so that subsets of their genome can be compared to corresponding extant reference panels.

The software package Tractor implements this procedure with extraordinary success, increasing statistical power for GWAS on admixed individuals with minimal losses in statistical power for individuals from a single ancestry (Atkinson et al. 2021). For a given quantitative trait and SNP, Tractor constructs a linear regression model in which the value of the trait is predicted by the frequencies of haplotypes at that SNP locus, the frequencies of risk alleles of that SNP in each admixed ancestry, and demographic covariates like age and sex. This model improves upon previous admixture mapping approaches by not directly associating the trait with ancestry so that it remains applicable when the trait is not stratified by ancestry (Gignoux et al. 2018).

Additionally, Atkinson et al. improve the resolution of haplotypes via alternating applications of LAI and statistical phasing. Statistical phasing, the task of determining haplotype blocks, often performs switch errors, in which a given haplotype block is disrupted by the erroneous inference of a small, distinct haplotype block within it, splitting it into two haplotype blocks (Choi et al. 2018). By rapidly “switching” between haplotype blocks, switch errors deflate the average length of inferred haplotype blocks. In the context of LAI on two ancestry groups, transitions between haplotype blocks represent “switches” from one ancestry to the other, so switch errors inflate the number of ancestry tracts, harming the power of GWAS by contributing additional gaps to portion of the genome from each constituent ancestry. By reducing the frequency of switch errors, Atkinson et al. set the stage for Tractor to achieve optimal performance on well-inferred local ancestry information.

Atkinson et al. claim improvement over asaMAP, a statistical method that accounts for the multiple ancestries of admixed individuals by allowing alleles’ effect sizes to vary among ancestries, but does not take full advantage of the accurate LAI achieved by Atkinson et al. (Skotte et al. 2019). In addition to benchmarking Tractor on simulated data, Atkinson et al. use empirical blood lipid data for which the value of studying multiple ancestry groups was established by Graham et al. (2019). Wojcik et al. (2019) emphasize the significance of the capacity for GWAS on diverse populations that Tractor enables.

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

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

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

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

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