**Docket #:** S13-268

# PCAmask: a novel approach for elucidating the evolutionary history of human populations through local ancestry deconvolution in admixed individuals

The genomes of admixed individuals are unique in that they are comprised of a mosaic of DNA segments derived from different ancestral populations. The advent of local ancestry deconvolution algorithms--designed to assign specific regions of admixed chromosomes to their respective ancestral origin--has made it possible to explore the genetic structure of ancient populations through the lens of modern admixed genomes.

While a number of algorithms have been developed to tackle the local ancestry problem, few tools have been devised to make meaningful use of the resulting information in the context of human evolutionary biology. Furthermore, this approach has suffered from a lack of resolution, providing information on continental variations in ancestry (e.g. whether a genome derives more of its ancestry from European versus African ancestors), but with limited insights into finer scale variation (e.g. whether a European segment may have corresponded to an Italian vs. a Scandinavian ancestor).

Stanford researchers have discovered an approach that provides a rigorous statistical framework for a more refined analysis of the genetics of ancestral populations. Specifically, the method interfaces three novel algorithms, acting in the following capacities:

1. A local ancestry deconvolution algorithm (***RFMix***, S13-178) is run to classify different regions of the admixed genomes by continental ancestry.

2. A PCA-based classification method (***PCAmask***) is used to further resolve the ancestral origins of the ancestral blocks identified in the first step, by projecting specific ancestral blocks against diverse panels of human populations deriving from that ancestral continental group.
3. A novel demographic inference algorithm (***Tracts***) is used to identify demographic events that are likely to have shaped variation in the ancestral populations that gave rise to the admixed population under study.

## Applications

- Can be used to make inferences on the natural history of human populations
- Can be used to identify ancestry-specific health risk factors in specific populations
- Can readily be used to empower current admixture mapping and association studies

## Advantages

- Integration of local ancestry deconvolution with downstream genetic clustering and demographic inference approaches in a single, seamless framework
- Increased resolution into the genetic structure of components of admixture in admixed populations versus other methods, such as running PCA or ADMIXTURE on non-deconvolved genomes

## Publications

- "Reconstructing the population genetic history of the Caribbean." Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martínez RJ, Hedges DJ, Morris RW, Eng C, Sandoval K, Acevedo-Acevedo S, Norman PJ, Layrisse Z, Parham P, Martínez-Cruzado JC, Burchard EG, Cuccaro ML, Martin ER, Bustamante CD. PLoS Genet. 2013 Nov;9(11):e1003925. doi: 10.1371/journal.pgen.1003925. Epub 2013 Nov 14.

## Innovators

- Carlos Bustamante
- Andres Moreno
- Simon Gravel
- Fouad Zakharia

## Licensing Contact

### Cheryl Cathey

Senior Licensing and Strategic Alliance Manager

Email