

A

TCGA individuals with reference populations
(SNPs used 54,436)

Smallest convex
sets identifying
reference ethnic
groups

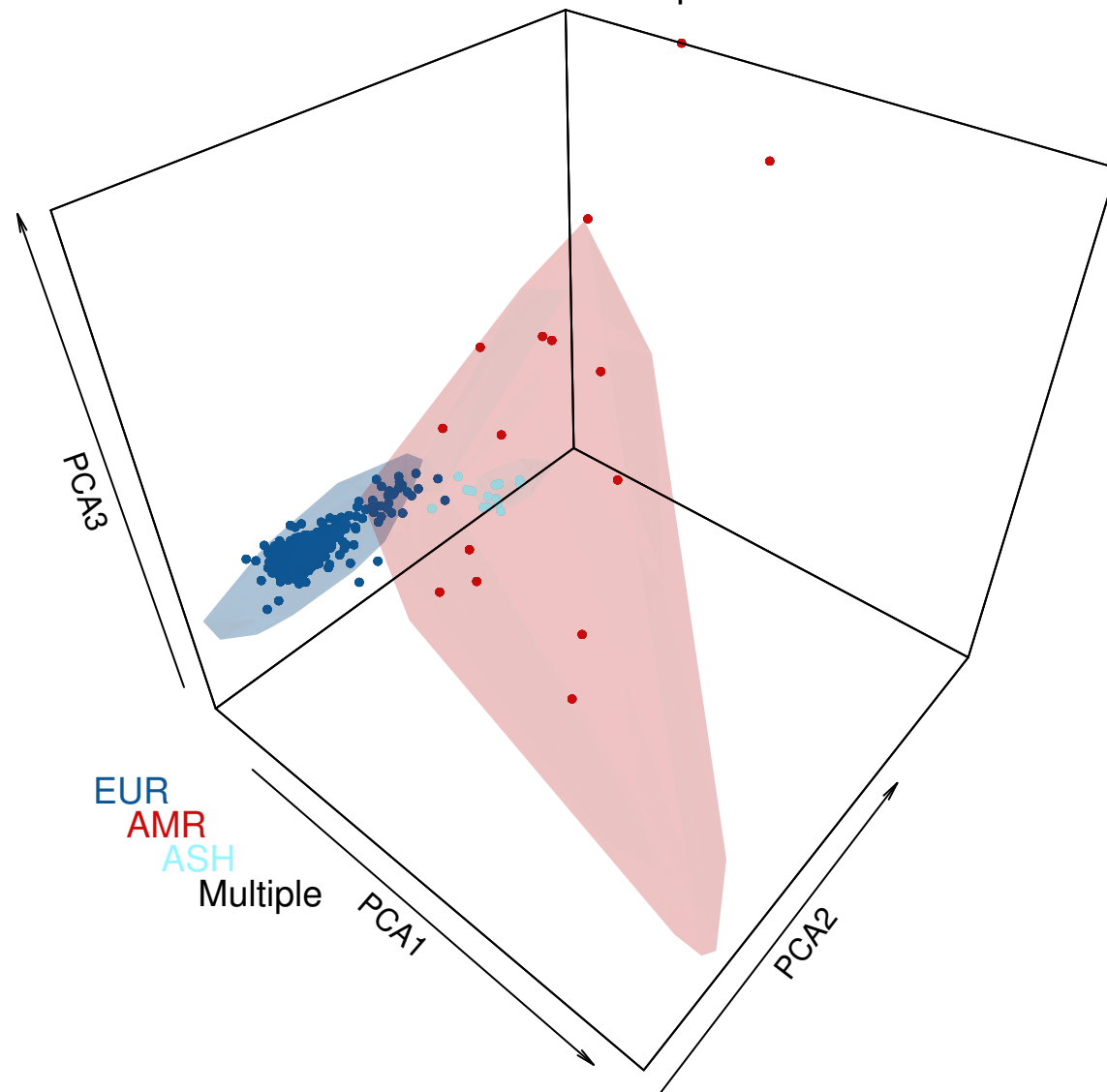
TCGA individuals with
colors representing
assigned ancestry

EUR
EAS
AMR
SAS
AFR
ASH
Multiple

B

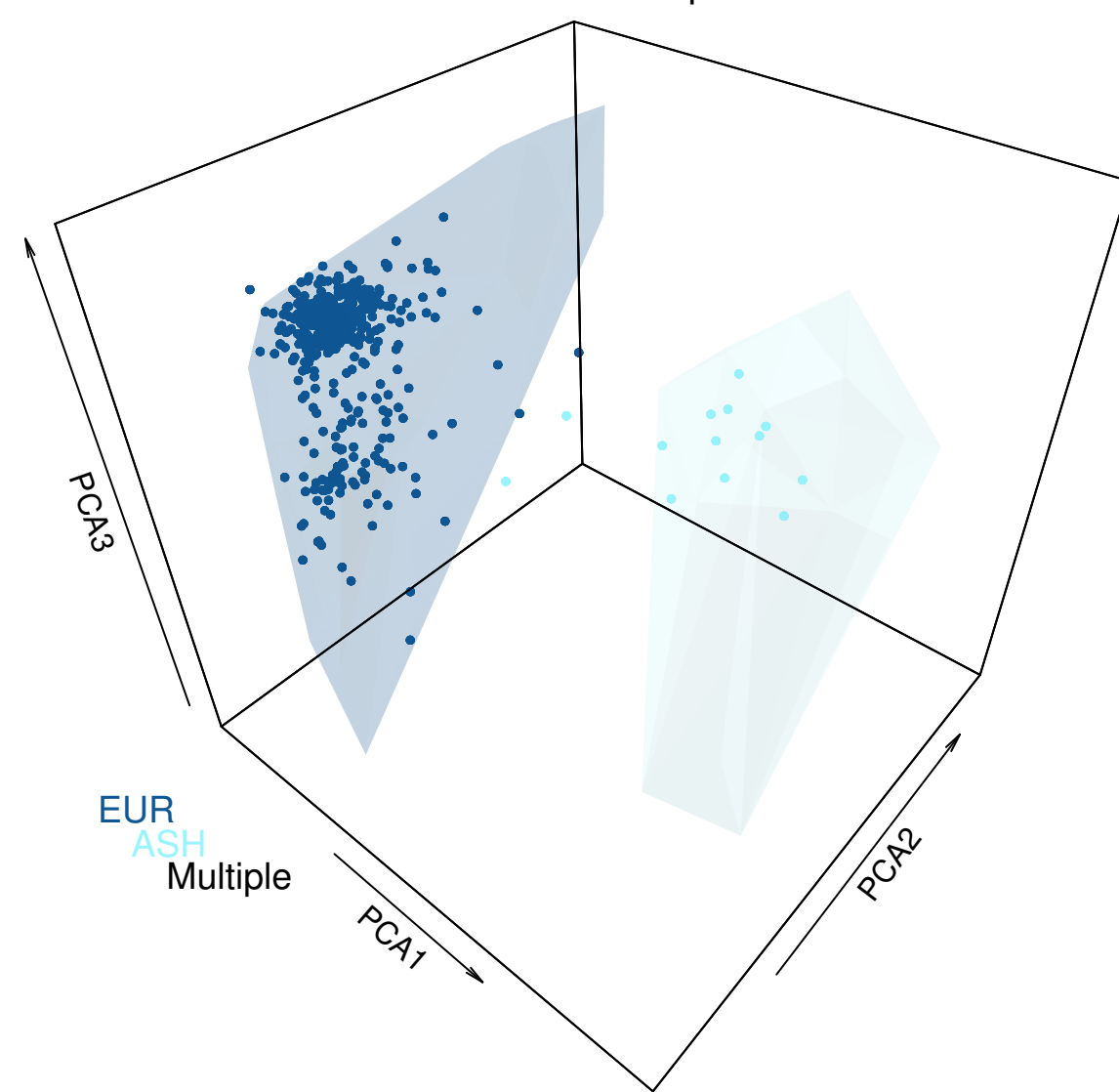
TCGA individuals with reference populations
(SNPs used 52,498)

Refinement step 1



TCGA individuals with reference populations
(SNPs used 51,533)

Refinement step 2



Example of EthSEQ WES-based ancestry analysis performed on the TCGA PRAD dataset. a) 3D space defined by the first three PCA components is shown. Colored polygons represent convex sets identifying reference ethnic groups, while point represent TCGA individuals. b) Two-step refinement analysis made by EthSEQ; the analysis is automatically refined by recursively run the method on reference and target models reduced to EUR, ASH and AMR groups (step refinement 1) and then on reference and target models reduced to EUR and ASH (step refinement 2). Final EthSEQ ancestry annotations are assigned according to the refinement analysis.