

High-dimensional analysis of SNP array data reveals major subdivisions in the Canary Island (Spain) population



Beatriz Guillen-Guio¹, Itahisa Marcelino-Rodríguez¹, Adrián Muñoz-Barrera², Luis A. Rubio-Rodríguez², Almudena Corrales^{1,3}, Ana Díaz-de Usera², Rafaela González-Montelongo², Jose M. Lorenzo-Salazar², Carlos Flores^{1,2,3}

¹Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; ²Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain; ²Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; ³CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

Introduction

The Canary Islands is a Spanish archipelago located in the Atlantic Ocean, 100 kilometers off Morocco. Due to its volcanic origin, it has not been connected to the mainland which implies an outstanding genetic makeup characterized by isolation and historical admixture.

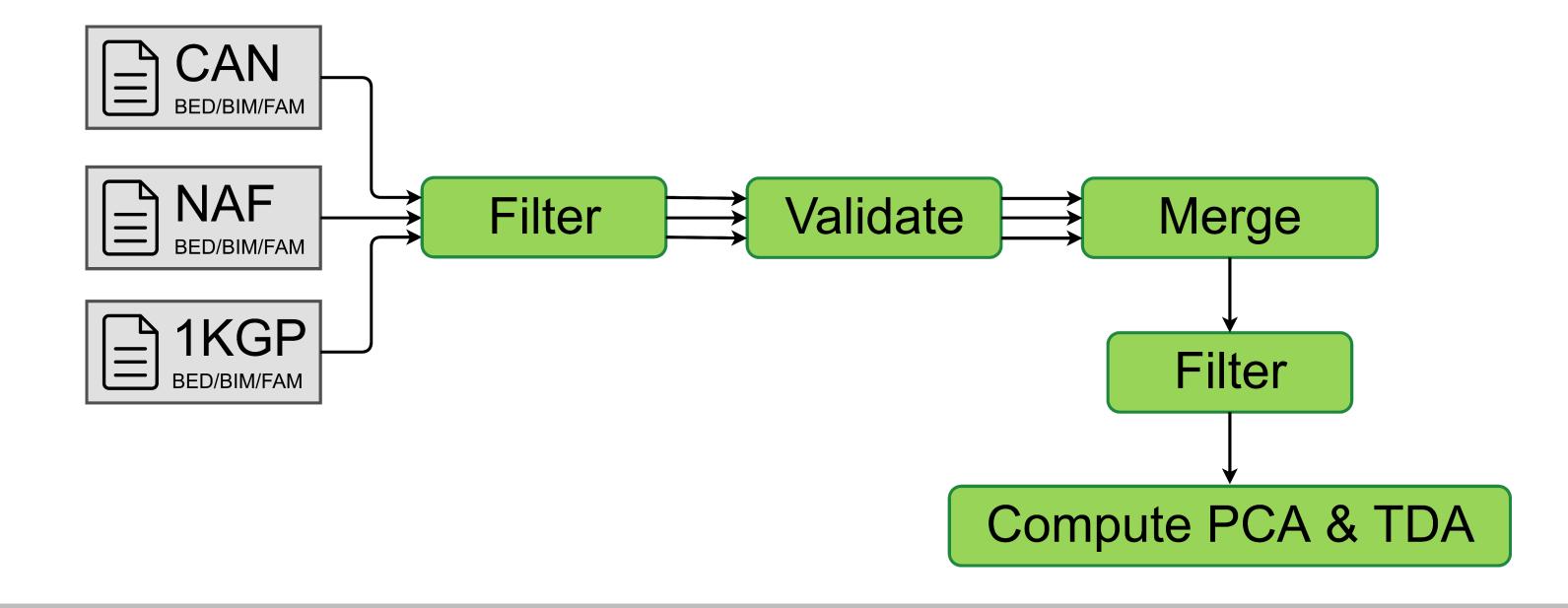
Recent studies1 based on SNP array data, model-based ancestry estimations, and principal component analysis (PCA) have revealed the existence of a significant African influence and population isolation in current inhabitants of the Canary Islands. Furthermore, the smaller islands have showed an outlier behavior when runs of homozygosity are considered. This issue could be interpreted as the result of genetic isolation, reduced population size and inbreeding events.

Here, we reassess the Canary Islanders genomic diversity and structure based on alternative high-dimensional analysis techniques in order to increase the accuracy in genetic subdivision within the islands.

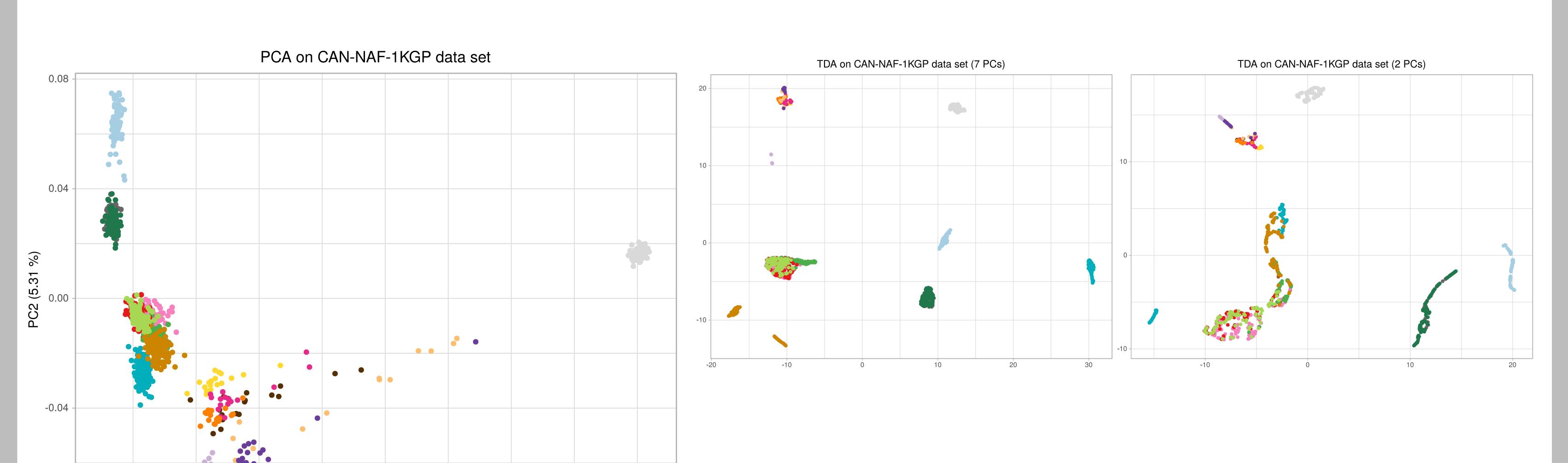
Materials and Methods

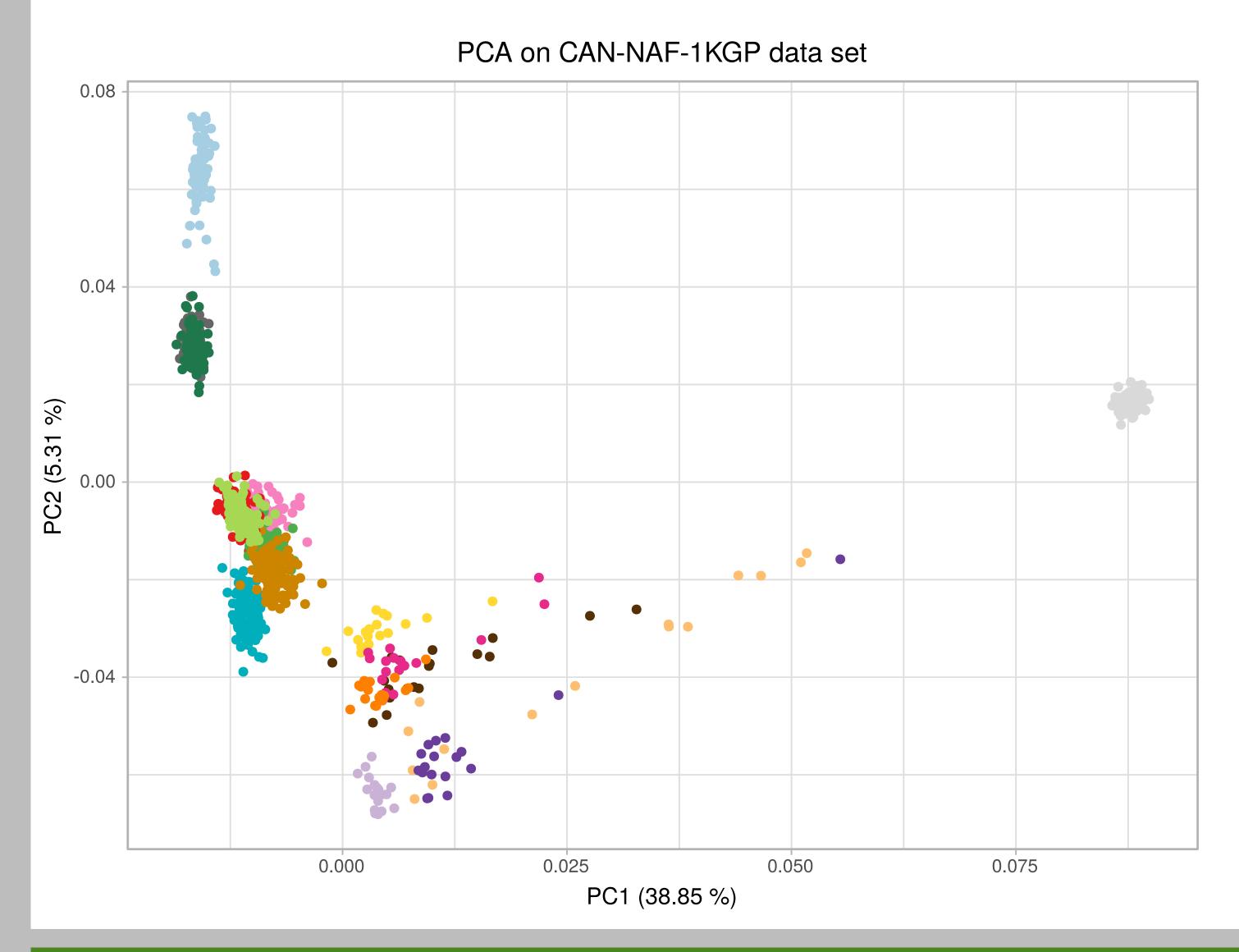
SNP array data from 618 Canary Islanders (CAN), and from 125 North-Africans (NAF) and SNP WGS data from 108 Yoruba and 289 European individuals from 1000 Genomes Project (1KGP) have been considered.

Data manipulation, principal components analysis (PCA) and topological data analysis (TDA) have been performed using PLINK v1.9, R software and umap3 v0.2.0 library for R as shown on fig. 1.



Results





0.025

PC1 (38.85 %)

0.050

0.075

Contact Funding

0.000

Ministerio de Ciencia, Innovación y Universidades (RTC-2017-6471-1; MINECO/AEI/FEDER, UE), agreement OA17/008 with ITER. Fellowship by ACIISI co-funded by European Social Fund to B.G.G (TESIS2015010057) and Spanish Ministry of Education, Culture, and Sports to ADU (FPU16/01435).

The authors declare no conflict of interest













References Guillen-Guio et al. *Mol Biol Evol* 2018; 35(12): 3010-3026

The 1000 Genomes Project Consortium. Nature 2015; 526: 68-74

Fregel et al. Eur J Hum Genet 2015; 23(9): 1236-1243.

Conclusions

TDA has allowed us to disentangle in a better way the inherent subdivision of Canary Islands, where La Gomera shows a signature pattern in three different clusters. This has been previously trend observed in other studies4 so our next steps will be aimed to unravel uniqueness of La Gomera subpopulations.

