University of Tartu
Bioinformatics (MTAT.03.239)

Project Report

Principal Component Analysis of Genotype Data from Various Populations

Kateryna Peikova, Marija Samoviča
13.06.2019.

Introduction

Human genetic variation is the set of genetic differences which distinguish our genomes from one another[1]. Although the differences between individuals within one population makes up the most part of human genetic variation, the geographically and ancestrally distant populations differ from each other[2]. In the last decades, advanced technology has become available that has allowed to study the extent and pattern of human genetic variation on a large scale. Such studies are of fundamental interest to evolutionary research and medical applications[3].

Anthropologists had previously hypothesized that humans originated in sub-Saharan Africa and left to colonize the rest of the world in several waves. This theory, known as "Out of Africa", had been previously supported, but not universally accepted, until recently when several large-scale population genetics studies showed that populations have less genetic diversity the further that population is from Africa[4] (Figure 1). Following Africa, the highest heterozygosities appear in populations from the Middle East, Europe, and Central and South Asia. Populations of East Asia have still lower genetic variation, and Pacific Islander and Native American populations, at the greatest geographic distance from Africa over migration paths traversed in human evolution, are the least heterozygous[5].

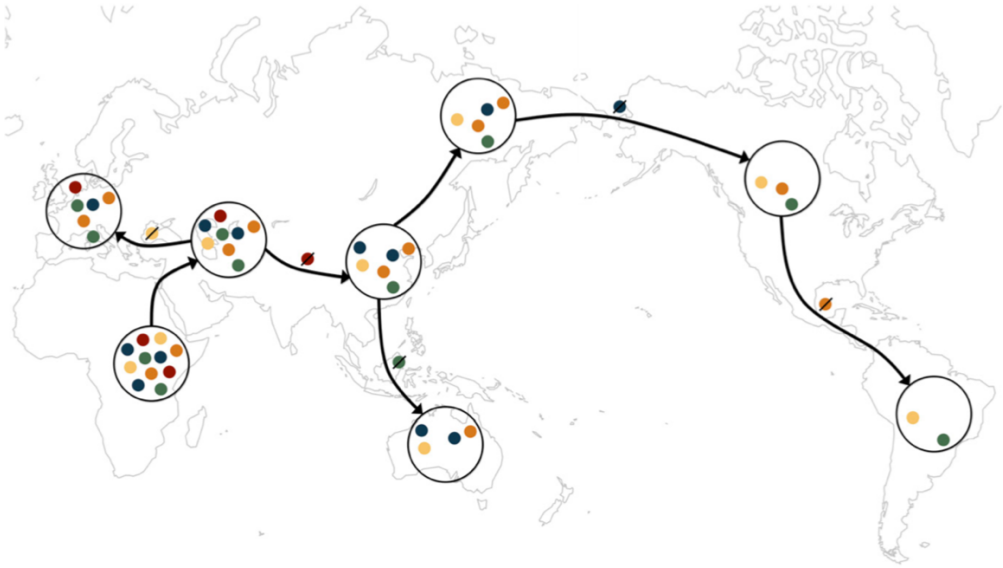

Figure 1. Schematic of the serial founder model in human evolution, where each color represents a distinct allele. Migration events outward from Africa tend to carry with them only a subset of the genetic diversity from the original population, and some alleles are lost during migration events (adapted from Rosenberg & Kang (2015)).

A powerful motivation to study human genetic variation is the discovery and description of the genetic contribution to many human diseases. As the vast majority of human diseases are

multifactorial, such as heart disease, cancer, diabetes and psychiatric disorders, researchers nowadays are trying to uncover genetic variations associated with these common diseases[6]. It is known that in some populations particular alleles can be more common than in others, and, as a result of variation in frequencies of both genetic and nongenetic risk factors, rates of diseases and their characteristics vary across populations[2]. Therefore, it is of high necessity to examine such associations in diverse human populations[5].

Single-nucleotide polymorphisms (SNPs) are the most frequent type of genetic variation among humans, occurring on average about every 100 to 300 base pairs[7], and SNPs are responsible for most of the variation in human phenotypes[8]. The largest initiative up to date to investigate human genome sequence variation is 1000 Genomes Project, and the final published dataset contains genomes of 2504 individuals from 26 populations. Altogether, the researchers characterized a broad spectrum of genetic variation, in total over 88 million variants, of which 84,7 million were SNPs. Although most common variants are shared across the world, rarer variants are typically restricted to closely related populations, and researchers found that 86% of variants were restricted to a single continental group[9].

In large-scale genetic variation studies involving hundreds and thousands of individuals, understanding the genetic structure of sampled individuals is of high necessity as many applications assume either a single panmictic population or knowledge of population structure[10]. Therefore, bioinformatics is essential for handling such data. Principal component analysis (PCA) is a crucial step in quality control of genomic data and a common approach for understanding population genetic structure[11]. PCA does not require a priori knowledge of population structure because it acts to project an individual's multilocus genotype onto a small number of dimensions (usually two) that maximally separate the data. When applying PCA to genome sequencing data, it is important to prune genotype data based on patterns of linkage disequilibrium (LD), as PCA assumes that markers are independent[10].

Aims of the Project:

- Compare existing tools for performing principal component analysis (PCA) on genotype data;
- Map new datasets to reference superpopulations from 1000 Genomes Project dataset;
- Create a convenient Nextflow pipeline for mapping new datasets on a reference principal component (PC) plot.

Significance:

- Detection of mislabeled samples;
- Prediction of superpopulation affiliation of unknown samples and new datasets, for example, when incorporating previously published datasets in other studies.

Methods

During work on the project, overall four different tools for principal component analysis (PCA) of genotype data were tested with the aim to project new datasets on a reference principal component (PC) plot to determine the superpopulation structure in the unknown datasets.

QTLTools

First, command line executable tool set QTLTools[12] was used to perform PCA analysis on 1000 Genomes Project dataset.

The QTLtools mode PCA allows performing PCA either on molecular phenotype quantifications or genotype data. It is typically used to detect outliers in the data, to stratify the data or to build a covariate matrix before QTL mapping. QTLtools provides parameters that help to avoid correlation of SNPs. The --maf 0.05 option ensures that only common variants with minor allele frequency higher than 5% are included in the analysis. The --distance 50000 option ensures that only genetic variances separated by at least 50000 nucleotides are included in the analysis.

QTLTools approach was successful in performing the PCA analysis on the dataset (Figure 2), the tool was convenient as it was fast and as input allowed the data to be in the original format (VCF), but it was not possible to map new datasets on an existing one. Also, QTLTools uses only an indirect approach for linkage disequilibrium (LD) pruning by selecting single-nucleotide polymorphisms (SNPs) within a user defined distance.

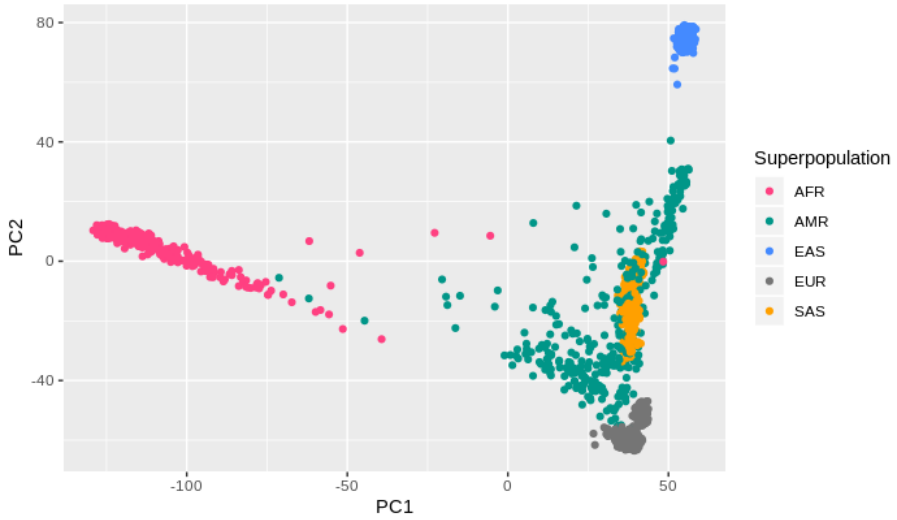

Figure 2. PCA of 1000 Genomes Project dataset, using QTLTools.

SNPRelate

The second tool used was SNPRelate – R package for computations of SNP data adapted for genome-wide association studies (GWAS)[13]. The functions in SNPRelate for PCA include calculating the genetic covariance matrix from genotypes, computing the correlation coefficients between sample loadings and genotypes for each SNP, calculating SNP eigenvectors (loadings) and estimating the sample loadings of a new dataset from specified SNP eigenvectors. SNPRelate allows to perform LD based SNP pruning.

Unlike QTLTools, SNPRelate is easiliy usable in R environment and is available in Bioconda. Unfortunately, the tool performed PCA of genotype data extremely slowly – during work on this project, we did not manage to perform LD pruning and PCA using SNPRelate due to enormous time

requirement. Additional drawbacks are that the input data need to be converted to GDS format, and this tool does not allow to map new datasets onto a reference dataset.

FlashPCA2

Third tool used was FlashPCA2 – both command line and R package tool for performing PCA of large SNP datasets[14]. In literature, FlashPCA2 is described as one of the fastest tools for PCA of genotype data as it outperforms other competing tools in terms of computation time[15]. This tool only computes the number of PCs the user has specified. Although all input data need to be converted to BED format, it is only a slight disadvantage as it allows to project new datasets on a reference PC plot. Alas, the tool failed to precisely map new datasets as slight shifts of clusters of known ancestry were observed when examining the output (Figure 3).

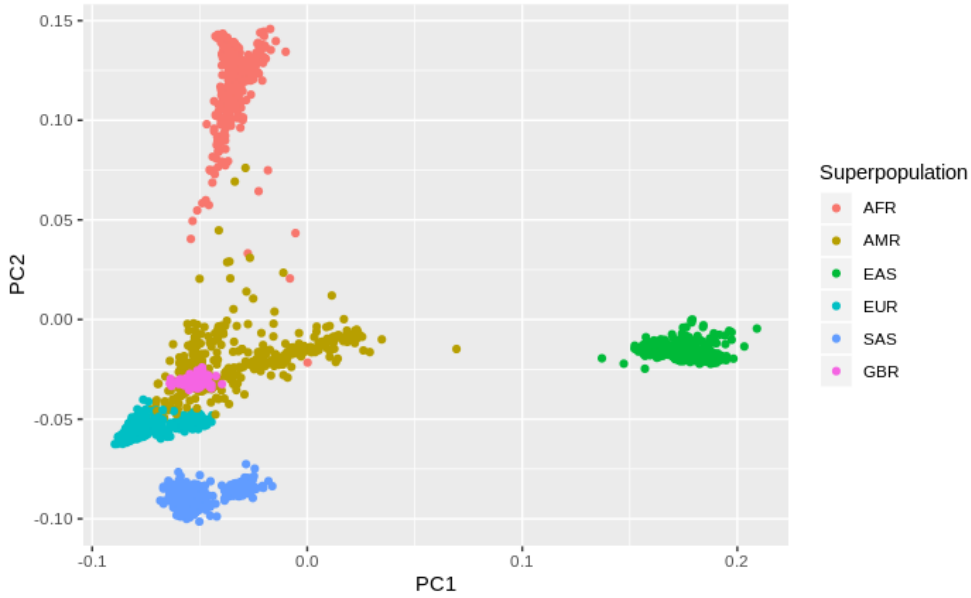

Figure 3. PCA performed with FlashPCA2. Onto the PC plot of reference dataset (1000 Genomes Project) was mapped an extracted subset of the same dataset - British population in England and Scotland (GBR, shown in pink) – belonging to European (EUR) superpopulation (shown in turqoise).

LDAK5

Lastly, LDAK5 was used – a command line tool set for obtaining LD Adjusted Kinships, which has further applications in diverse genetic variation studies[16]. This tool calculates a kinship matrix which can then be used to perform PCA of genotype data. LDAK5 performance is fast, as the number of PCs is chosen by user. LDAK5 worked successfully for mapping new datasets onto a reference PC plot and it was therefore applied in this project.

Nextflow Pipeline

As LDAK5 is a stand-alone command line tool which also requires the data to be converted to BED format and PCA plots to be created separately, the workflow needed to be compiled into an effective pipeline. For this purpose, Nextflow was used – a bioinformatics workflow manager that enables the development of portable and reproducible workflows[17]. Nextflow supports different scripting languages, tools and environments, and a Nextflow script is defined by composing many different processes[18]. Therefore, in this project a Nextflow pipeline was created to perform mapping and assigning of new datasets by using a reference PC plot.

Before performing the PCA of genotype data established in the pipeline, all datasets were preprocessed using BCFtools – command line utility for manipulating Variant Call Format (VCF) and its binary version (BCF) files[19]:

1. Only autosomes were included in the analysis, therefore all chromosomes except X and Y were concatenated into one sorted file for each dataset
2. SNP identificators were converted in one consequent format, containing chromosome, position, reference allele and alternative allele (for example, chr1_629731_T_C):

```
bcftools annotate source_data/GRCh38_22_chrs_sorted_2.vcf.gz --
set-id 'chr%CHROM\_%POS\_%REF\_%FIRST_ALT' -Oz -o
source_data/GRCh38_renamed_ids.vcf.gz
```

3. Only biallelic SNPs were extracted and used for analysis:

```
bcftools view -m2 -M2 -v snps
source_data/GRCh38_renamed_ids.vcf.gz -Oz -o
source_data/GRCh38_renamed_ids_no_multiallelic2.vcf.gz --
threads=20
```

Established Nextflow pipeline:

- In the beginning of the pipeline, Plink – a toolset for whole genome association analysis[20] – is used for several processes:
    1. 1000 Genomes Project reference dataset and the new dataset are converted from VCF to binary format BED;
    2. The Ad Mixed American (AMR) population is excluded from the 1000 Genomes Project reference dataset;
    3. Duplicated SNPs are removed from the datasets;
    4. LD pruning is performed on the reference dataset;
    5. List of SNPs that are shared between the reference dataset after LD pruning and the new dataset is extracted;
    6. Overlapped SNPs are extracted from the new dataset that will be mapped;
    7. Overlapped SNPs are extracted from the reference dataset;
- LDAK5:
    1. Kinship matrix for the reference dataset is calculated;
    2. PCA of reference dataset is performed;
    3. New dataset is mapped on the main dataset;
- R script, using ggplot2, dplyr and DMwR packages:
    1. PC plots are created;
    2. Distances for each of the individuals of the new dataset to cluster centers of the reference dataset are calculated and normalized to sum up to one;
    3. 5-nearest neighbors approach is used to assign individuals of the new dataset to existing superpopulations in the reference dataset;
    4. To verify the assigning, a threshold is set on the calculated values of the distances to cluster centers – if the distance to the closest cluster center exceeds 0,1, then the individual is assigned to the admixed category.

Results

In this project 1000 Genomes Project dataset and six other datasets (Cedar, Fairfax, Gencord, Imwar, Nedelec, Quatch) were used to project the new datasets onto the reference principal component (PC) plot. For all datasets only autosomes were used, they were filtered to contain only biallelic single-nucleotide polymorphisms (SNPs), and also SNP identificators (IDs) were converted to be in one format. All principal component analyses (PCAs) were performed, using the created Nextflow pipeline.

1) Nedelec

Nedelec dataset consists of 171 individuals. The dataset was preprocessed as described previously, and, using the final subset of SNPs, PCA analysis of the dataset was performed. The results indicate that Nedelec contains individuals from at least two distinct populations and admixed individuals with ancestry from both of these populations (Figure 4).

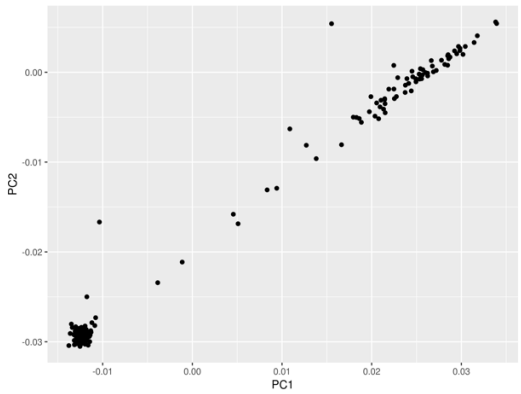

Figure 4. PCA plot of Nedelec dataset, using a filtered set of SNPs that are shared between 1000 Genomes and Nedelec datasets.

1000 Genomes dataset was preprocessed as described in the methods section and SNPs that were shared with Nedelec dataset were used for PCA. Also, the American superpopulation (AMR) was excluded from the reference dataset because of mixed ancestry. The PCA plot of the final set of SNPs and superpopulations shows that the previously observed clustering of superpopulations is not compromised by using a subset of all SNPs after LD pruning (Figure 5).

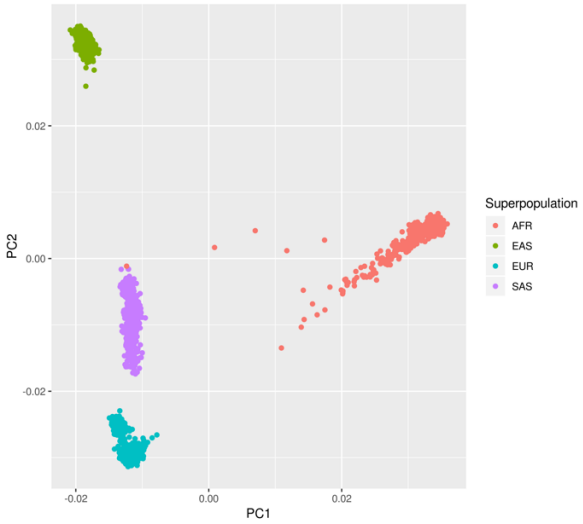

Figure 5. PCA analysis of 1000 Genomes dataset, where only shared SNPs from preprocessed 1000 Genomes and Nedelec datasets were used.

Mapping of Nedelec dataset onto the PCA plot of 1000 Genomes Project dataset shows that not only individuals from European and African ancestry are present, but also an individual with South Asian ancestry. In addition, there are several individuals for which a clear admixed ancestry can be observed (Figure 6).

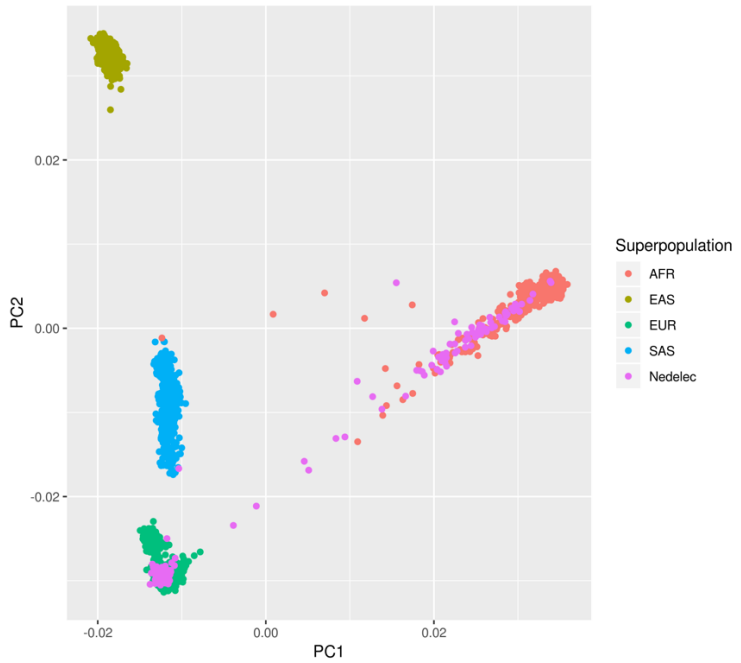

Figure 6. Nedelec dataset mapped onto 1000 Genomes reference dataset.

Using the k-nearest neighbors approach, all individuals from the Nedelec dataset were assigned to one of the four superpopulations present in the reference dataset. The result was imprecise, as several individuals that have a similar distance to two superpopulation clusters were assigned to one of them (Figure 7, Appendix 2).

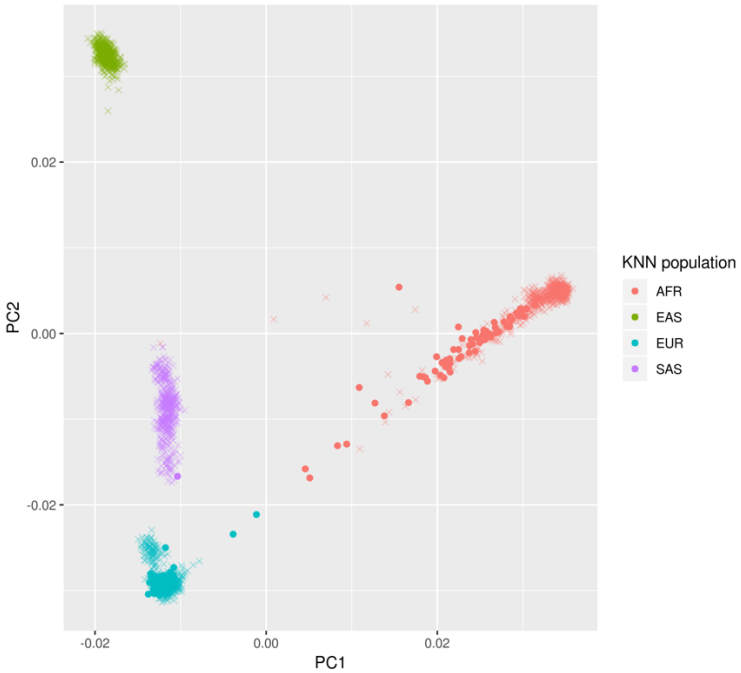

Figure 7. Assigning individuals of Nedelec dataset to superpopulations from the reference 1000 Genomes dataset, using k-nearest neighbors approach. Reference data are shown with crosses, individuals from Nedelec dataset – with dots.

Because there is a clear presence of admixed individuals in the Nedelec dataset, another layer of assigning was employed. By applying a threshold on the distances to the centers of each superpopulation cluster, 17 individuals were identified as admixed instead of assigning them to one of the superpopulations (Figure 8, Appendix 3).

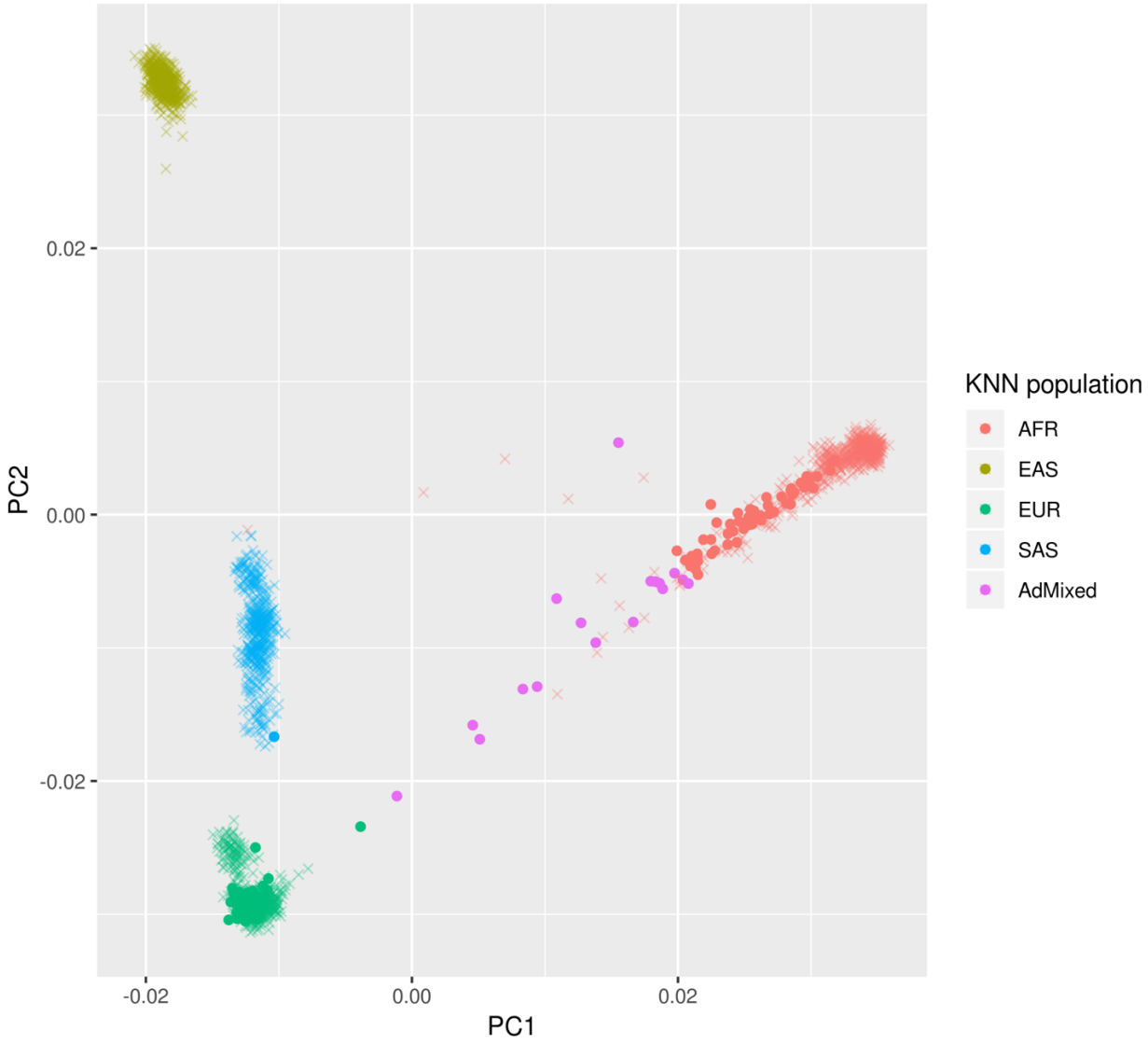

Figure 8. Final assignment of individuals from the Nedelec dataset (shown as dots) to either existing superpopulations or admixed category, using the 1000 Genomes dataset (shown as crosses) as reference.

2) Cedar

Cedar dataset contains SNP information of 341 individuals. The PCA of Cedar dataset alone shows no clear separate populations (Figure 9 A). Mapping the dataset onto the reference dataset indicates that all individuals have European ancestry (Figure 9 B). When assigning the individuals to reference superpopulations, both k-nearest neighbors approach and also distance threshold setting assigns all individuals to European superpopulation (Figure 9 C, D, Appendices 2, 3).

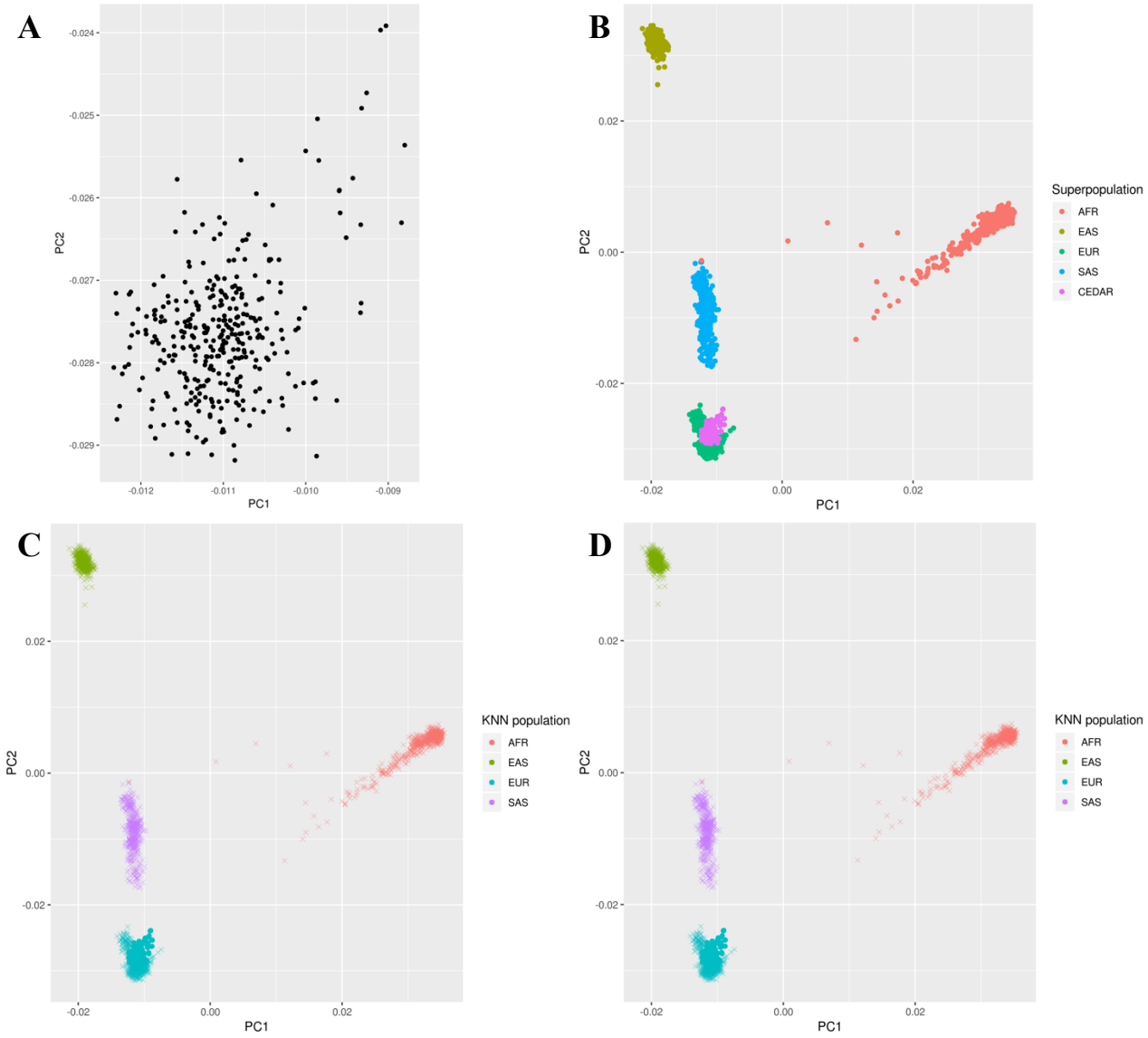

Figure 9. Assigning individuals from Cedar dataset to superpopulations, using 1000 Genomes Project dataset as reference. **A** PCA of Cedar dataset. **B** Projecting Cedar onto PC plot of reference dataset. **C** Assigning individuals to superpopulations with k-nearest neighbors method. **D** Verification of assignment to superpopulations by setting a threshold on distances to cluster centers.

3) Fairfax

Fairfax dataset holds genotype data of 432 individuals. PCA plot shows one dense cluster with possibly few admixed individuals or people from different superpopulations present in the dataset (Figure 10 A). When mapped onto the reference dataset, the "outliers" are identified as four individuals from South Asian superpopulation (Figure 10 B-D, Appendices 2, 3).

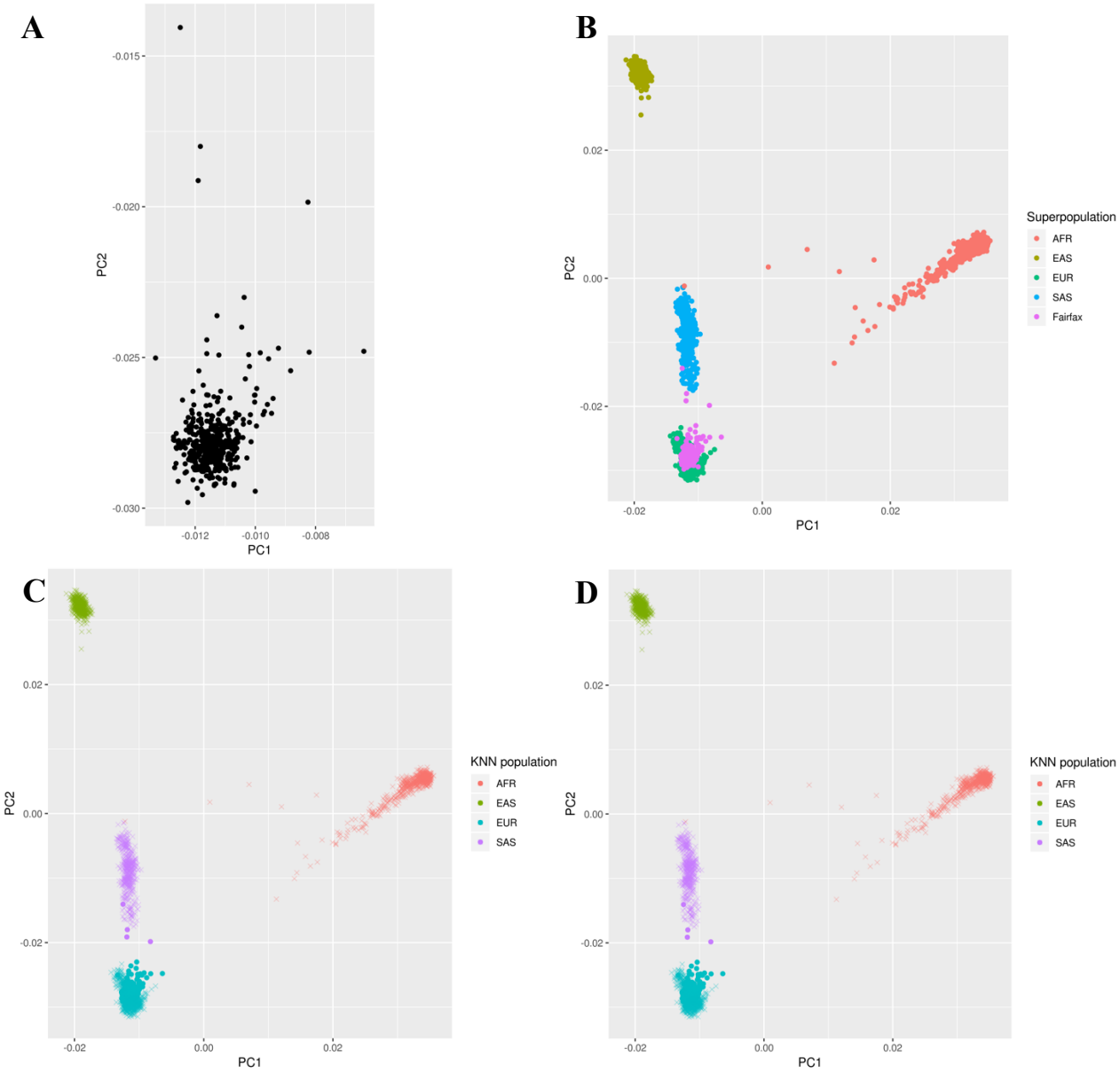

Figure 10. Assigning individuals from Fairfax dataset to superpopulations, using 1000 Genomes Project dataset as reference. **A** PCA of Fairfax dataset. **B** Mapping Fairfax onto the reference dataset. **C** Assigning individuals to superpopulations with k-nearest neighbors method. **D** Verification of assignment by setting a threshold on distances to cluster centers.

4) Gencord

PCA of Gencord dataset (204 individuals) shows one dense cluster with several data points being scattered further from the cluster (Figure 11 A). Mapping Gencord onto the 1000 Genomes Project dataset assigns most of the individuals to European superpopulation (196), but eight – to South Asian (Figure 11 B). Although several data points are located relatively far from the centers of both European and South Asian clusters, no individual is identified as having admixed ancestry (Figure 11 C, D, Appendices 2, 3).

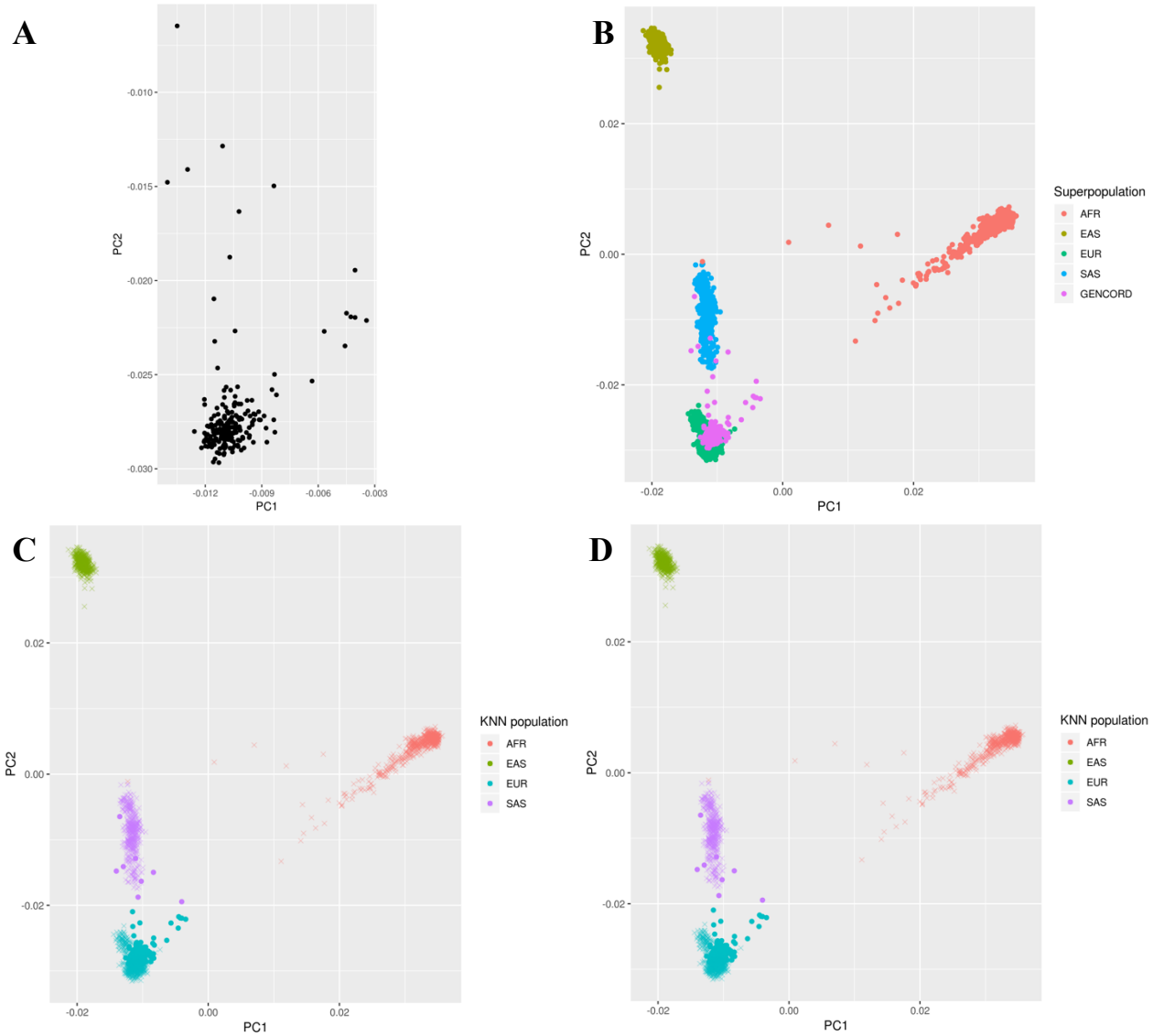

Figure 11. Assigning individuals from the Gencord dataset to superpopulations with 1000 Genomes Project dataset as reference. **A** PCA of Gencord dataset. **B** Mapping Gencord onto the reference dataset. **C** Assigning individuals to superpopulations with k-nearest neighbors method. **D** Verification of assignment by setting a threshold on distances to cluster centers.

5) Quatch

Quatch dataset (200 individuals) contains two distinct populations (Figure 12 A). Mapping the dataset onto the reference indicates that both European and African superpopulations are represented in the new dataset (Figure 12 B). When assigning the individuals of the Quatch dataset to the superpopulations, all are distinctly assigned to either European or African superpopulations (Figure 12 C, D, Appendices 2, 3).

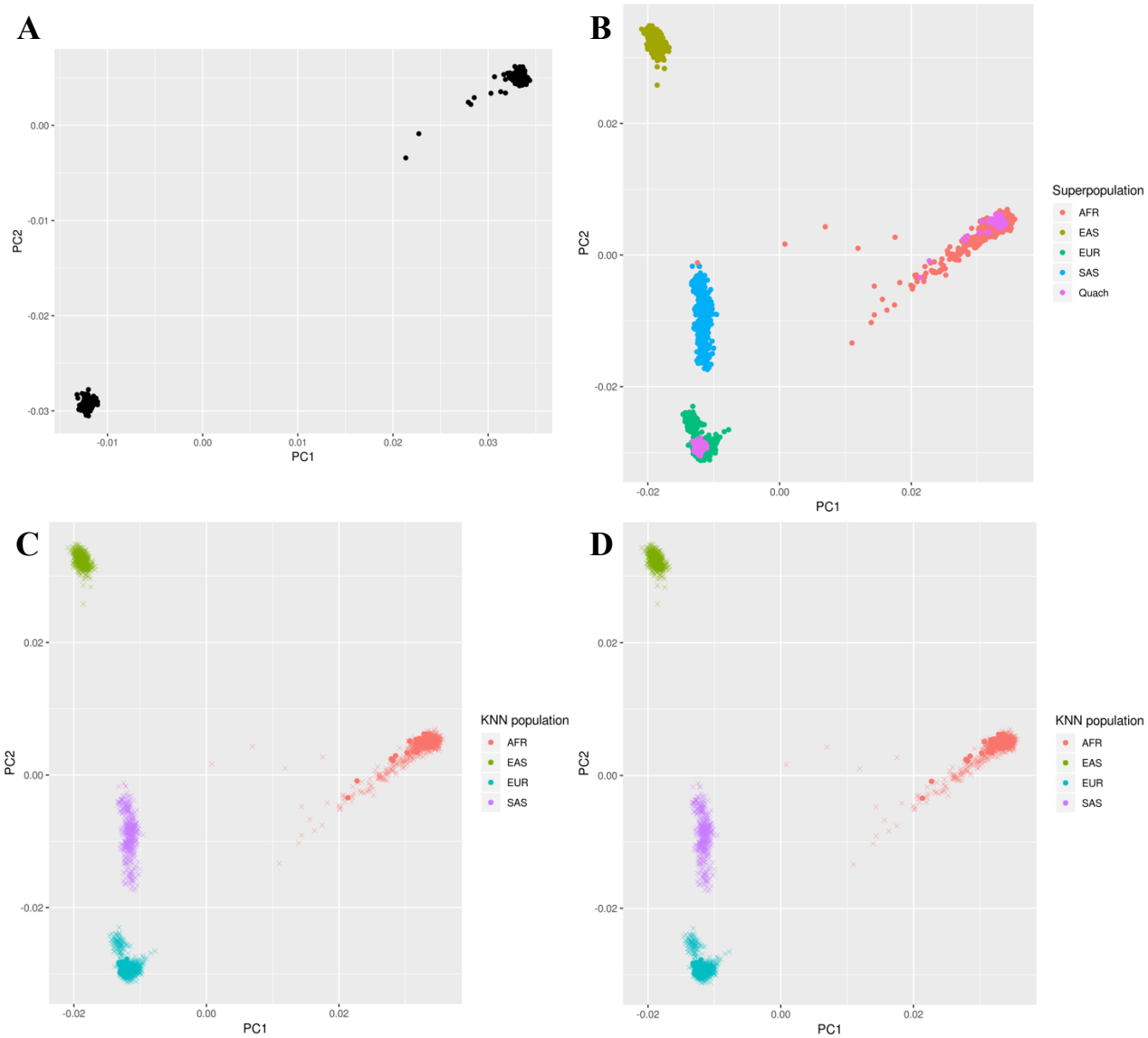

Figure 12. Mapping and assigning individuals from the Quatch dataset onto the reference dataset (1000 Genomes Project). **A** PCA of Quatch dataset. **B** Quatch dataset mapped onto the reference. **C** Assigning of individuals, using k-closest neighbors method. **D** Verification of assignment by setting a threshold on distances to cluster centers.

6) Imwar

Imwar dataset contains genotype information of total 687 individuals. PCA shows that Imwar is composed of individuals of several different ancestries (Figure 13 A). When mapped onto the reference dataset, it can be clearly seen that there are also individuals present with diverse admixed ancestries (Figure 13 B). Therefore, assigning individuals to superpopulations using only k-nearest neighbors method was insufficient, and final assignment resulted with 49 individuals marked as admixed (Figure 13 C, D, Appendices 2, 3).

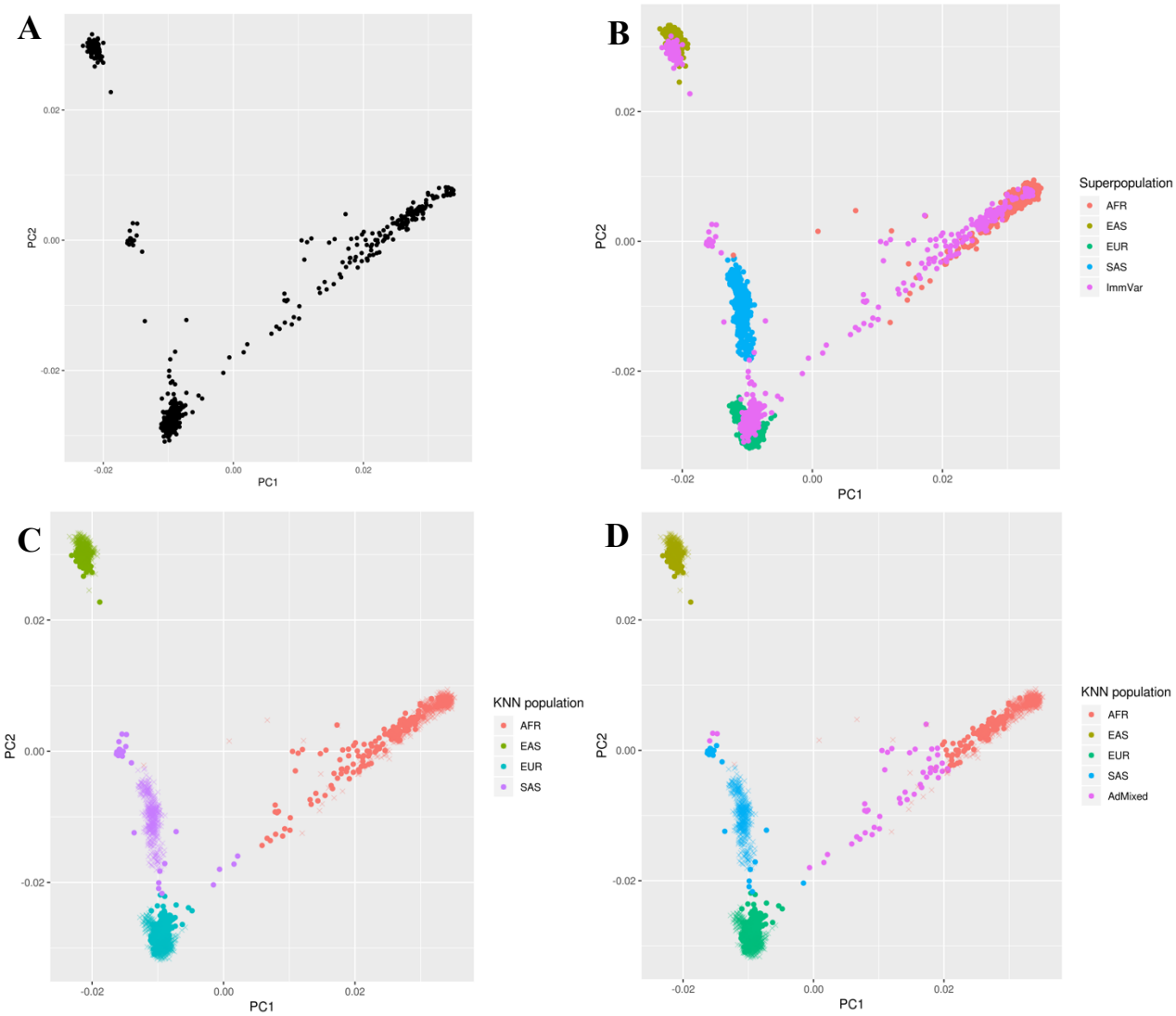

Figure 13. Mapping and assigning individuals from the Quatch dataset onto the 1000 Genomes Project dataset. **A** PCA plot of Imwar dataset. **B** Imwar dataset mapped onto the reference dataset. **C** Assigning individuals to superpopulations with k-nearest neighbors method. **D** Verification of assignment by setting a threshold on distances to cluster centers.

Discussion

In this project four different tools for principal component analysis (PCA) of single nucleotide polymorphism (SNP) data were evaluated and compared. As PCA has long been an important tool for inferring and visualizing genetic structures of populations, a variety of diverse tools have been made open access. Widely used and competing tools for PCA on genotype data include QTLTools and SNPRelate, of which QTLTools is preferable in terms of computation time when using a large genome-scale dataset. As the final goal of this project was to develop a workflow to map new genotype datasets onto a reference principal component (PC) plot, none of the two tools were applicable.

In order to examine the population structures of new datasets and assign individuals to superpopulations, FlashPCA2 and LDAK5 were tested. As FlashPCA2 is both command line and R package toolset adapted for fast performance when using large datasets, it would be preferred over the stand-alone LDAK5 toolset. However, FlashPCA2 produced imprecise mapping of new datasets onto the reference, therefore LDAK5 was applied in further work together with additional tools as it resulted in fast and precise mapping.

To create a reproducible workflow, a Nextflow pipeline for mapping and assigning new datasets was designed and created. The input for the created pipeline is two preprocessed datasets – reference dataset and a new dataset to map. Using the established pipeline, the superpopulation structure of six new datasets was studied. All six datasets were successfully mapped onto the reference 1000 Genomes Project dataset.

The Ad Mixed American (AMR) superpopulation in the 1000 Genomes Project dataset is of mixed ancestry with recent admixture from two populations that were previously genetically isolated for tens of thousands of years. Therefore, AMR was excluded from the reference dataset before the mapping of new datasets as admixture induces complications in analyzing population structures from PCA results. It has been shown that in PCA admixed individuals are dispersed along the line segment connecting the clusters of the two parental populations in the two-dimensional space of the first two PCs. This kind of dispersion along a line has been used as a diagnostic of admixture. The relative distances of an individual from the centroids of the clusters of the parental populations have been used to estimate the admixture proportions of the individual[11].

In the project, six datasets were projected on the reference 1000 Genomes Project PC plot and all individuals were assigned to one of the four superpopulations present in the reference dataset by implementing 5-nearest neighbors approach. In some cases, examples of dispersion indicating admixture were observed, for example, in Imwar and Nedelec datasets. In order to assign such individuals, another layer of assignment verification was used – distances to each of the cluster centers were calculated and normalized, and then a threshold of 0,1 to the closest cluster was set to assign individuals to admixed group if exceeded. In order to improve the assignment, different thresholds for each of the clusters could be introduced based on previous literature.

Although specific methods and tools for estimation of admixture proportions have been developed, such analyses are limited and produce relatively wide confidence intervals[11]. As the significance of this project was to examine population structures of unknown datasets and assign individuals to superpopulations, not estimate proportional affiliation to several superpopulations, determination of admixture proportions is therefore out of scope for this project. Overall, the established approach is well suited for identification of swapped data, as well as examination of superpopulation structure of new datasets with unknown genetic structure of sampled individuals.

References

1. Casillas, S., Mulet, R., Villegas-Mirón, P., Hervas, S., Sanz, E., Velasco, D., ... & Barbadilla, A. (2017). PopHuman: the human population genomics browser. *Nucleic acids research*, *46*(D1), D1003-D1010.
2. Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *science*, *298*(5602), 2381-2385.
3. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... & Stephens, M. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98.
4. https://news.stanford.edu/news/2008/february27/med-genemap-022708.html
5. Rosenberg, N. A., & Kang, J. T. (2015). Genetic diversity and societally important disparities. *Genetics*, *201*(1), 1-12.
6. National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series [Internet]. Bethesda (MD): National Institutes of Health (US); 2007. Understanding Human Genetic Variation.Available from: https://www.ncbi.nlm.nih.gov/books/NBK20363/
7. Ke, X., Taylor, M. S., & Cardon, L. R. (2008). Singleton SNPs in the human genome and implications for genome-wide association studies. *European Journal of Human Genetics*, *16*(4), 506.
8. Sunyaev, S., Kondrashov, F. A., Bork, P., & Ramensky, V. (2003). Impact of selection, mutation rate and genetic drift on human genetic variation. *Human molecular genetics*, *12*(24), 3325-3330.
9. 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.
10. Schraiber, J. G., & Akey, J. M. (2015). Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*, *16*(12), 727.
11. Ma, J., & Amos, C. I. (2012). Principal components analysis of population admixture. *PloS one*, *7*(7), e40115.
12. https://qtltools.github.io/qtltools/
13. https://github.com/zhengxwen/SNPRelate
14. https://github.com/gabraham/flashpca
15. Abraham, G., Qiu, Y., & Inouye, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics*.
16. http://dougspeed.com/ldak/
17. https://github.com/nextflow-io/nextflow
18. Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature biotechnology*, *35*(4), 316.
19. https://samtools.github.io/bcftools/bcftools.html
20. http://zzz.bwh.harvard.edu/plink/

Appendix

Table 1. Overview of populations and superpopulations included in the 1000 Genomes Project dataset (http://www.internationalgenome.org/faq/which-populations-are-part-your-study).

Population Code	Population Description	Superpopulation Code	Superpopulation Name
CHB	Han Chinese in Beijing, China	EAS	East Asian
JPT	Japanese in Tokyo, Japan	EAS	East Asian
CHS	Southern Han Chinese	EAS	East Asian
CDX	Chinese Dai in Xishuangbanna, China	EAS	East Asian
KHV	Kinh in Ho Chi Minh City, Vietnam	EAS	East Asian
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	EUR	European
TSI	Toscani in Italia	EUR	European
FIN	Finnish in Finland	EUR	European
GBR	British in England and Scotland	EUR	European
IBS	Iberian Population in Spain	EUR	European
YRI	Yoruba in Ibadan, Nigeria	AFR	African
LWK	Luhya in Webuye, Kenya	AFR	African
GWD	Gambian in Western Divisions in the Gambia	AFR	African
MSL	Mende in Sierra Leone	AFR	African
ESN	Esan in Nigeria	AFR	African
ASW	Americans of African Ancestry in SW USA	AFR	African
ACB	African Caribbeans in Barbados	AFR	African
MXL	Mexican Ancestry from Los Angeles USA	AMR	Ad Mixed American
PUR	Puerto Ricans from Puerto Rico	AMR	Ad Mixed American
CLM	Colombians from Medellin, Colombia	AMR	Ad Mixed American
PEL	Peruvians from Lima, Peru	AMR	Ad Mixed American
GIH	Gujarati Indian from Houston, Texas	SAS	South Asian
PJL	Punjabi from Lahore, Pakistan	SAS	South Asian
BEB	Bengali from Bangladesh	SAS	South Asian
STU	Sri Lankan Tamil from the UK	SAS	South Asian
ITU	Indian Telugu from the UK	SAS	South Asian

Table 2. Assignments to superpopulations from reference dataset, using k-nearest neighbors method.

Dataset	Total Number of Individuals	East Asian (EAS)	European (EUR)	African (AFR)	South Asian (SAS)
Cedar	341	0	341	0	0
Fairfax	432	0	428	0	4
Gencord	204	0	196	0	8
Imwar	687	141	358	164	24
Nedelec	171	0	98	72	1
Quatch	200	0	100	100	0

Table 3. Assignments to superpopulations from reference dataset and admixed category, using both k-nearest neighbors approach and distance threshold to centers of superpopulation clusters.

Dataset	Total Number of Individuals	East Asian (EAS)	European (EUR)	African (AFR)	South Asian (SAS)	Admixed
Cedar	341	0	341	0	0	0
Fairfax	432	0	428	0	4	0
Gencord	204	0	196	0	8	0
Imwar	687	141	358	121	18	49
Nedelec	171	0	97	56	1	17
Quatch	200	0	100	100	0	0