Rev. Esp. Antrop. Fís. (2024)
Vol. 49: 89-97

Polygenic Risk Score for coronary artery disease and stroke in European populations

Mikel de la Peña Sanz[1], Neskuts Izagirre Arribalzaga[1,2], Santos Alonso Alegre[1,2]

[1]Department of Genetics, Physical Anthropology and Animal Physiology. Fac. Science and Technology. University of the Basque Country (UPV/EHU). Barrio Sarriena s/n 48940 Leioa-Bizkaia. Spain.
[2]Maria Goyri Building. Asnimal Biotechnology Center. Leioa Campus, Human Molecular Evolution Lab (2.08). Barrio Sarriena s/n 48940 Leioa-Bizkaia. Spain.

Corresponding Author: mdelapena007@ikasle.ehu.eus

Palabras claves:

Enfermedades complejas
Enfermedades cardiovasculares
ADN Neandertal
Genética de poblaciones

Recibido: 05-12-2023
Aceptado: 15-04-2024

Keywords:

Complex Diseases
Cardiovascular Diseases
Neanderthal DNA
Population Genetics

RESUMEN

Las enfermedades cardiovasculares componen la principal causa de muerte en el mundo, particularmente en Europa, donde existen ciertas disparidades entre países en cuanto a su impacto. Además de un componente ambiental, en esta enfermedad interviene un componente genético asociado a varios genes, cada uno con un pequeño efecto. Gracias a los análisis de riesgo poligénico es posible cuantificar el riesgo genético a poseer cierta enfermedad o rasgo a partir de ciertos marcadores genéticos. En este estudio se ha comparado el riesgo genético de diferentes poblaciones europeas a contraer dos tipos de enfermedades cardiovasculares: la enfermedad de las arterias coronarias y los accidentes cerebrovasculares. Los resultados mostraron que solamente el riesgo genético para la enfermedad de las arterias coronarias mostraba una diferencia significativa entre los valores medios de las distintas poblaciones. Además, se investigó la posible correlación entre el riesgo genético de ambas enfermedades y la cantidad de variantes de origen Neandertal presentes en cada individuo. A pesar de que la correlación resultaba estadísticamente no significativa, sí hay un indicio de que haya una relación con el riesgo genético a la enfermedad de las arterias coronarias. Finalmente, tras comparar el riesgo genético con la situación de hospitalización en cada país para las enfermedades cardiovasculares se concluye que, a pesar del efecto ambiental, es posible identificar la relación entre el componente genético y la enfermedad, resultando de interés para estudios evolutivos, así como clínicos.

ABSTRACT

Cardiovascular diseases are the main cause of death worldwide and particularly in Europe, where there are some disparities across the countries regarding its impact. In addition to an environmental component, this disease has also a genetic component by which many variants in different genes, contribute with a small effect each. Thanks to Polygenic Risk Scores it is possible to quantify the genetic risk for a certain illness or trait. In this study, we compared the genetic risk of different European populations for two types of cardiovascular diseases: coronary artery disease and stroke. The results showed that populations presented differences in genetic risk only for coronary artery disease. Moreover, we also explored for both traits the possible correlation between risk scores and the number of Neanderthal variants in the individuals. Although overall we did not find found any statistically significant correlation, some clue indicates there could be a correlation with the genetic risk score for coronary artery disease. Finally, after comparing the genetic risk scores with the hospitalization situation for cardiovascular diseases in each country, we conclude that, despite the possible effect of environmental factors, it is possible to identify a genetic component for the disease, which is relevant not only for evolutionary studies but also for clinical purposes.

ISSN: 2253-9921

Introduction

Recently, the World Health Organization (2020) estimated that the top 2 leading causes of death were a) ischaemic heart disease (IHD), also known as coronary artery disease (CAD), and b) stroke. CAD consists in the formation of an atherosclerotic plaque in the vessel lumen which hinders blood and oxygen flow to the myocardium (Shahjehan and Bhutta, 2023). Meanwhile, stroke happens when brain cells die due to a problem with blood and oxygen supply (Centers for Disease Control and Prevention, 2023). These diseases have caused 8.9 million and over 6 million deaths respectively in 2019, representing together 27% of world's total deaths, and the number of deaths for both diseases has kept increasing since the year 2000 (World Health Organization, 2020).

According to Townsend et al. (2021), CVD has remained the most common cause of death in Europe since the early 1980s. Previous studies have reported that CVD kills nearly 4 million people in Europe every year (44% of all deaths), with IHD accounting for 44% of these CVD deaths, and stroke accounting for an additional 25%. Disparities, however, are found across the continent, with large differences in current age-standardized and crude death rates for CVD among countries (Townsend et al., 2021).

Despite advances in risk prediction, most people's first sign of CVD risk is already a late event such as a stroke or myocardial infarction. Thus, when identifying biomarkers and risk factors that might add information to clinical risk factors, genetics seems to be a good path (Abraham, Rutten-Jacobs and Inouye, 2021). As Hüls and Czamara (2019) explained, since most complex diseases are influenced by several genes, each having only a small effect on its own, polygenic approaches, due to their genome-wide nature, often explain a greater proportion of the variance in complex traits than is possible with single variant approaches. The most popular polygenic approach is weighted polygenic risk scores (PRS) or genetic risk scores (GRS) (Hüls and Czamara, 2019). They integrate information from many common single nucleotide polymorphisms (SNPs) and provide a quantitative metric of inherited risk (Patel and Khera, 2023). This is done by weighted sums of risk alleles of many SNPs. Particularly for complex diseases, PRS are statistically powerful to test for marginal genetic effects and gene–environment interaction effects, as well as to predict individual trait values or risks of disease (Hüls and Czamara, 2019). Both CAD and stroke possess a multifactorial genetic basis, with independent genetic variants conferring small differences to the risk of both diseases (Chauhan and Debette, 2016; Patel and Khera, 2023).

As Abraham, Rutten-Jacobs and Inouye (2021) say, CVD PRSs seem to perform as well as or better than individual traditional risk factors and even being independent of clinical risk scores. This happens since genome-wide summary statistics (GWASs) have uncovered many genetic associations which implicate the role of novel biological pathways unexplained by traditional risk factors. Therefore, PRS that are based on GWAS tend to capture a broad range of genetic effects in a way that they are not so biased regarding to the biological pathways (Abraham, Rutten-Jacobs and Inouye, 2021).

What's more, in populations of European ancestry, a genome-wide association study of Neanderthal-introgressed alleles showed a wide range of associations with pathological outcomes, such as myocardial infarction and coronary atherosclerosis, among others (Simonti et al., 2016; Koller et al., 2022). Thus, a possible link between CVD and the Neanderthal introgression component is worth investigating.

Consequently, the aim of this study is to analyse whether different European populations show different PRS to CAD and stroke, as well as to study if the differential amount of introgressed Neanderthal DNA could represent an additional genetic risk. We hypothesize that the if CAD and stroke have a genetic basis, European populations in 1000 Genome Project (Auton et al., 2015) may show different genetic risk to these diseases. Based on the observations described above, we also hypothesize that the possible differential Neanderthal component could also be playing a role in this risk. We hope that elucidating this could have both an evolutionary and a clinical value

Methodology

To obtain PRS values for CAD and stroke, we first obtained from the NHGRI-EBI GWAS Catalog

(Sollis et al., 2022) a series of SNPs significantly associated to these diseases (p-value $<10^{-5}$). The search was done in December 2022, finding 3247 and 608 associations for CAD and stroke respectively. Then, those associations were filtered by the ancestry of the studied individuals, selecting those which had individuals with European ancestry, whether entirely or partially. Cases in which the articles did not provide ancestry data were also included, unless any clue could indicate no European ancestry individuals were present. Next, associated polymorphisms which were not SNPs (i.e., indels or repetitions) or SNPs for which the risk allele was unknown, were excluded. Moreover, repeated SNPs, indels or variants on the sexual chromosomes or mtDNA were also excluded, as well as those SNPs which had more than two alleles. Summary statistics of these SNPs were downloaded from the same NHGRI-EBI GWAS Catalog (Sollis et al., 2022) in December 2022 for trait EFO_0000712 and trait EFO_0001645.

Later, to perform a quality control (QC) of the data, we proceeded as described by Choi, Mak and O'Reilly (2020). In the absence of MAF value in GWAS, the frequency for European population was obtained from dbSNP (https://www.ncbi.nlm.nih.gov/snp/). Any SNP not found in the original article or found out to be an error was also discarded.

Genotypic population data was obtained from the 1000 Genomes Project (Auton et al., 2015), for the following European populations: Northern Europeans from Utah (CEU), Finnish in Finland (FIN), British from England and Scotland (GBR), Iberian population in Spain (IBS) and Tuscans from Italy (TSI), and a standard QC was applied (Choi, Mak and O'Reilly, 2020). Genomes were Downloaded from 1000 Genomes FTP on 08-06-2016. Since these populations were assembled as GRCh37/hg19 while SNPs were in GRCh38/hg38, a transformation was performed for SNPs using the webpage https://genome.ucsc.edu/cgi-bin/hgLiftOver. PRS calculation was accomplished by means of PLINK v1.90 b7 64-bit (Purcell et al., 2007). A clumping process was applied, requiring that the associations had a linkage disequilibrium (r2) lower than 0.1; 250 kb of distance and without restricting any SNP according to its p-value (threshold=1) (Choi, Mak and O'Reilly, 2020).

To calculate the Neanderthal introgression into these European populations, a total of 137853 SNPs identified previously as of Neanderthal origin were extracted (Vernot and Ackey, 2014; Gunz et al., 2019). Of these, those SNPs that were also present in the Yoruba in Ibadan (YRI) population were identified and removed. Finally, an estimate of Neanderthal ancestry was assessed by scoring 0, 1 or 2 if each locus was homozygous for the ancestral allele, heterozygote or homozygous for the derived, Neanderthal allele, respectively.

Statistical analyses were done by using R-Studio 4.1.1 (RStudio Team, 2020) and R commander 2.9-2 (Fox, 2005, 2017; R Core Team, 2023; Fox and Bouchet-Valat, 2024). First, a Shapiro-Wilk test was performed to test the normality of the data. Then, an ANOVA and Kruskal-Wallis test were performed to study the differences among the groups. Finally, the Pearson correlation coefficient between the estimated Neanderthal ancestry component and PRS was calculated. Further, QC summary statistics SNPs were filtered by using a Linkage Disequilibrium and MAF threshold of 0.2 and 0.1, respectively, to perform a PCA.

Finally, after analysing the differences of PRS results for both diseases in every population, we proceeded to compare the actual cardiovascular disease incidences as well as some health indicators. They were obtained from Eurostat database (https://ec.europa.eu/eurostat/data/database), selecting for the filtering all ages as well as both sexes from 2014, as all populations were present in the same study. The searched items were: 1) "Hospital discharges by diagnosis, in-patients, per 100 000 inhabitants", using "Diseases of the circulatory system (I00-I99)"; 2) "Causes of death - standardised death rate by NUTS 2 region of residence", for "Diseases of the circulatory system (I00-I99)"; 3) "Population structure indicators at national level", using "Median age of population"; 4) "Body mass index (BMI) by sex, age and educational attainment level", marking "obese" and "All ISCED 2011 levels"; 5) "Daily smokers of cigarettes by sex, age and educational attainment level", using "total" and "All ISCED 2011 levels"; 6) "Frequency of heavy episodic drinking by sex, age and educational attainment level", using "never or not in the last 12 months"; 7) "Frequency of fruit and vegetables

consumption by sex, age and educational attainment level", using "All ISCED 2011 levels and At least once a day" for both fruits and vegetables.

The databases World of Science (WOS) and PubMed were used for bibliographical research. The following keywords were used: coronary artery disease, cardiovascular disease, stroke, polygenic risk score or genetic risk score. These key words were integrated in the search bar by using "AND". Preferentially, we considered reviews from the last ten years. Moreover, we focused on journals with an impact index higher than 3 or belonging to the first 2 quartiles. In addition, articles cited in the bibliography of those references were also considered. Other sources such as Eurostat (https://ec.europa.eu/eurostat) and the Centers for Disease Control and Prevention (https://www.cdc.gov/index.htm) were also used.

Results and discussion

The list of SNPs after performing the QC along with their characteristics are available at Tables S1 and S2 (see Supplementary data). After filtering, the number of SNPs used for PRS calculation was 429-437 for CAD and 101-103 for Stroke (Table S3, see Supplementary data), as not all populations had information for all SNPs.

The two highest CAD PRS average scores correspond to CEU and TSI, while GBR and FIN had the two lowest values (Figure 1A).

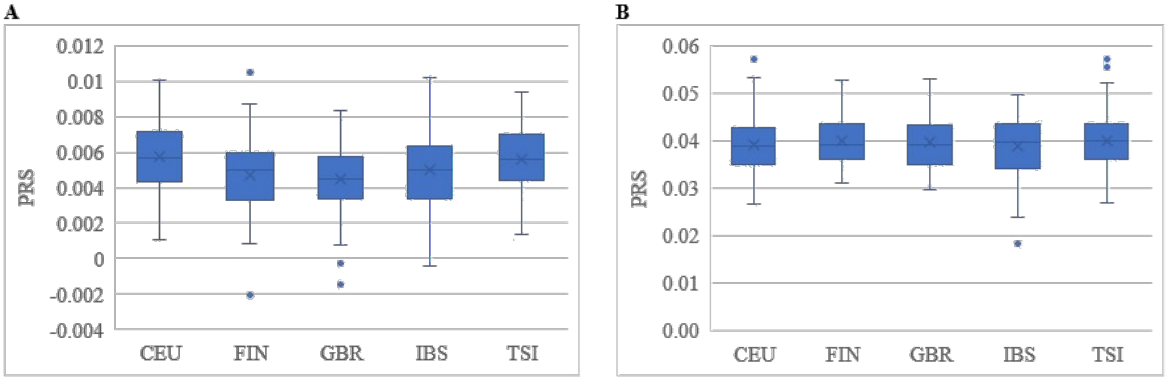

Figure 1. Box plot of PRS values (Y axis) for each population (across the X axis) for CAD (A) and stroke (B). Note that the scale for Y axis of both graphics is not the same. Mean values, marked as an x in the graphic, are 0.00578, 0.00472, 0.00451, 0.00498 and 0.00561 for CEU, FIN, GBR, IBS and TSI, respectively, for CAD. For stroke PRS, they are 0.0393, 0.0399, 0.0397, 0.0289 and 0.04, respectively. Standard deviations for CAD PRS are 0.00203, 0.00198, 0.00179, 0.00214 and 0.00178 for CEU, FIN, GBR, IBS and TSI, respectively. For stroke PRS standard deviations are 0.00548, 0.00486. 0.00577, 0.00603 and 0.00599, respectively.

According to the ANOVA (Table 1), these two groups (CEU-TSI vs GBR-FIN) were significantly different pairwise. Meanwhile, IBS with an intermediate risk score only showed significant difference with CEU. For stroke, the differences among populations were not significant for both ANOVA (Figure 1B, Table 1) and Kruskal-Wallis test (p score=0.8508). PRS scores for each individual are available at Table S4 (see Supplementary data). Differences in CAD and stroke PRS values may be due to difference in overall Odds Ratio of SNPs as well as different frequencies of risk alleles in populations.

We compared these results with actual data found in Eurostat (n.d.) (Table 2). The selected countries are thought to be a good proxy for 1000 Genome populations analysed: FIN for Finland, GBR for the UK, IBS for Spain, TSI, although being Tuscans, for Italy, and CEU for Germany, the Netherlands, Sweden, Norway and Denmark, as all these countries conform the Northern European region.

Table 1. P-values of ANOVA pairwise for difference among the European populations for CAD and stroke PRS. Values lower than 0.005 (Bonferroni correction for multiple tests for a nominal alfa 0,05 and 10 tests) are marked with 3 asterisks (* p-value <0.05; ** p-value <0.01; *** p-value < 0.005).

		FIN	GBR	IBS	TSI
CAD	**CEU**	0.00133***	< 0.001***	0.02828*	0.97213
	FIN		0.94603	0.86981	0.00957**
	GBR			0.43283	< 0.001***
	IBS				0.12612
Stroke	**CEU**	0.913	0.98	0.993	0.877
	FIN		0.999	0.697	1
	GBR			0.863	0.997
	IBS				0.630

Table 2. PRS values and actual data about the situation in 2014 for European countries that better represent the analysed populations Eurostat (n.d.). *Value for Toscana. **Data from 2019, since there was no information in 2014.

Country	Denmark	Finland	Germany	Italy	Netherlands	Norway	Spain	Sweden	The UK
Average CAD PRS	0.005783	0.004719	0.005783	0.005614	0.005783	0.005783	0.004982	0.005783	0.004507
Average Stroke PRS	0.039253	0.039942	0.039253	0.0400013	0.039253	0.039253	0.038922	0.039253	0.039716
Hospital discharges per 100,000	1629.3	2647.8	3772.5	1904.8	1385.2	2013.3	1307.1	2052.8	1206.6
Deaths per 100,000	256.73	378.86	403.67	310.15 281.25*	271.86	272.74	244.99	338.37	265.26
Mean age (years)	41.3	45.6	42.4	44.7	42	39	41.8	40.9	39.9
Obesity rate (%)	14.4	17.8	16.4	10.5	12.9	12.6	16.2	13.4	19.8
Smokers rate (%)	12.3	11.6	15	17.4	17.2	12.5	22.2	8.7	13.7
No heavy episodic drinking (%)	27.5	37.3	39.9	83.6	75.9**	56	77.5	53.5	54.4
Eating at least 1 fruit per day (%)	53	44	47.3	70.9	41	54.1	66.7	46.8	62.8
Eating at least 1 vegetable per day (%)	44.1	45.4	34.1	61.9	31.3	55.7	44.6	52.1	65.5

Although BMI is suggested not to be a perfect predictor of body composition, as occasionally two people with similar BMI could have different fat concentrations, it is often the best available measure (Rychter et al., 2020). By looking at Table 2, we can appreciate an overall tendency that as the obesity rate increases, so do the death and hospital discharges rates. However, some discrepancies can be seen, such as the fact that Germany despite having lower obesity rates compared to Finland, still has higher values of death and hospitalizations. This phenomenon could be explained by the lower CAD PRS value that FIN population has, indicating that genetics would be a factor compensating these values. Another anomaly would be the case for TSI, which despite having the lowest obesity rate as well as having a better eating habit and lower alcoholism, still has average heath results. This could be also explained by their high CAD PRS, along with their higher average population age. The contrary happens with GBR population, which in

spite of having the highest obesity rate, has a pretty low death and hospital discharges rates. Something similar can be appreciated in the case of Spanish people, that have low health issues regarding cardiovascular diseases although having higher obesity and smoking rates. In this case, genetics would not be explaining much of that improvement since PRS values were average. Instead, it could be attributed to healthier eating behaviour as well as lower alcoholism.

However, there are the cases of Sweden and Norway which by having similar obesity rates, lower smoking, lower average age, and better eating habits than the Netherlands, still have worse health results. The higher alcoholism could be contributing to this result, but there might be more risk factors which have not been gathered in Table 2. Moreover, there could be differences in the genetic risk score for all Norther European countries that have not been considered since CEU population is a mixture of all. Another limitation is that the comparison is being made mixing CVD incidences and CAD PRS values, therefore information about other CVD PRS is being overlooked.

Regarding the Neanderthal Component Value (NCV), when it was correlated with each disease for all European individuals and individuals for each population, in none of them there was a significant correlation (Table 3). However, in the case of CAD when using all individuals the effect of NCV on PRS value the p score reduced considerably up to 0.0562.

Table 3. Pearson correlation coefficient (r) and p value for relationship between NCV and PRS values for CAD and stroke.

	CEU	FIN	GBR	IBS	TSI	Total
			CAD			
r	-0.0274	-0.0266	0.0004	-0.0273	-0.1255	-0.0852
p value	0.7878	0.7942	0.9971	0.78	0.1976	0.0562
			Stroke			
r	-0.1184	-0.0313	-0.0444	0.0207	0.0301	-0.0096
p value	0.2432	0.7598	0.6776	0.8325	0.7583	0.8307

This low p value (scarcely higher than 0.05) could mean that the NCV might have a significant effect on CAD PRS, but more individuals would be required in order to prove this hypothesis. When observing each population, TSI showed lower p score for CAD PRS than others. We speculate that this might be due to the following idea described by Anagnostou et al. (2022): Italy is a highly diverse population, with a great genetic distance between northern and southern groups, and in this line, the risk of developing CAD also varies among Northern and Southern Italy. On the one hand, in Northern Italy the cold temperatures could have enhanced natural selection on genes involved in adipogenesis (*WDPCP*), HDL concentration (*RNMTL1P2*), cholesterol and triglyceride levels (*RAB3GAP1* and *R3HDM1*), as well as the sensibility of cells to insulin (*TMEM163*), for which Neanderthal derived adaptative alleles have been associated with lower risk of having CAD. On the other hand, in Central and Southern Italy, to where TSI population would belong to, the high presence of ancestral alleles for some of these loci might have increased susceptibility to CAD (Anagnostou et al., 2022).

In this sense, according to Khrameeva et al. (2014) contemporary Europeans have an excess of Neanderthal-like sites in genes involved in lipid catabolic processes. Neanderthals might have acquired changes in the lipid catabolism processes which would have resulted beneficial, possibly changing involved metabolites concentrations along with the expression of enzymes, and therefore have spread into European ancestors. In this regard, some diseases, among them CAD, appear to show an enrichment in genes involved in lipid catabolism which contain an excess of Neanderthal-like sites. However, alternative hypotheses suggest that this situation might have not been produced because of Neanderthal introgression, but rather it could be due to the presence of those genes in the ancestors of Neanderthals and out-of-Africa humans and then the frequencies had increased in both

groups because of the prehistoric European conditions (Khrameeva et al., 2014). Although p score was not statistically significant, as previously said, this could be due to a being a small sample. Nonetheless, limitations intrinsic to the method to obtain the NCV must be considered, as only ascertained SNPs were used for its estimation.

Regarding the risk for stroke, it was suggested that some Neanderthal genes were likely to increase risk for stroke due to a blood coagulation disorder (Gibbons, 2016; Groß, 2019). However, in this study that association has not been shown, as the p value suggests that there is no correlation between stroke PRS and NCV. By contrast, CEU population showed a much lower p value than other populations, but still has a high p value.

When performing a PCA with the all the risk SNPs, it seems that the main variance is explained by a North-South geographical tendency (Figure 2). Finns are shown at the left end of axis PCA1, followed by the north-western European populations (GBR and CEU), and finally southern European populations (IBS and TSI) at the right end of axis PCA1. Meanwhile, on the second axis intra-population variance is mostly appreciated. The ellipses manifest a great overlap between GBR and CEU, and another one between IBS and TSI for the stroke and CAD informative SNPs (Figure 2).

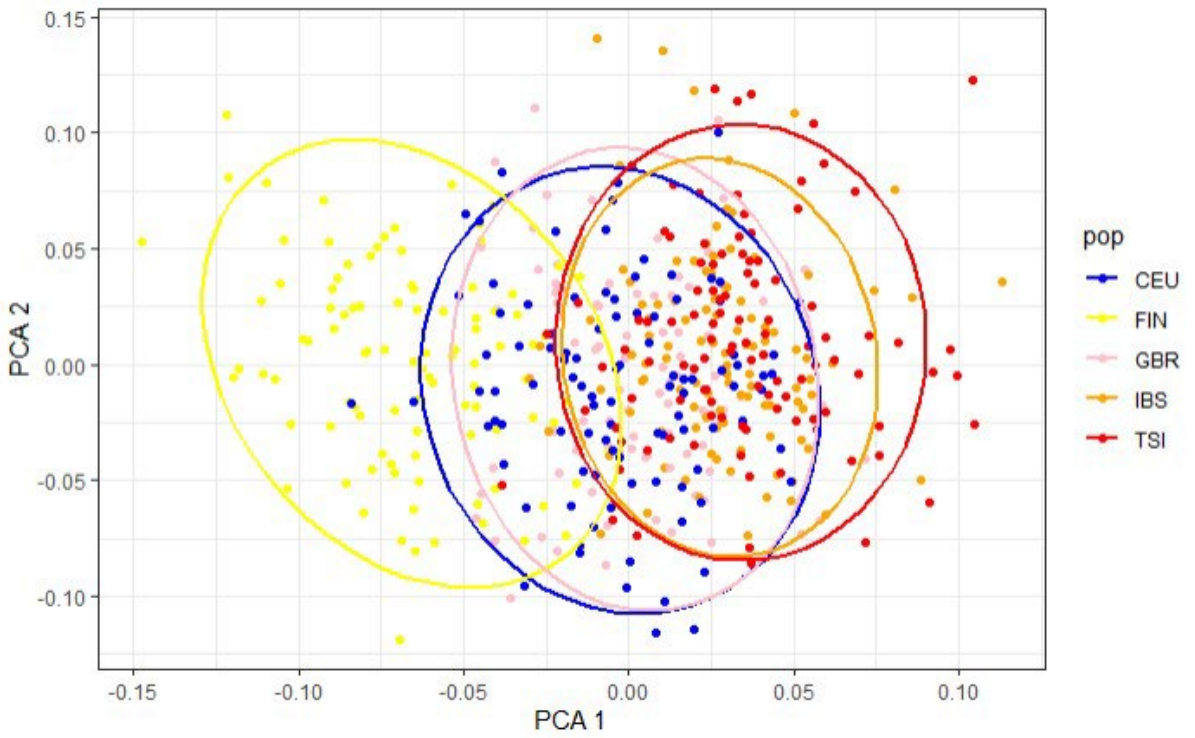

Figure 2. PCA constructed with CEU, FIN, GBR, IBS, TSI populations along with Russian population.

Finally, this study has several limitations. First, it has not been checked whether the chosen association p-value threshold ($p<10^{-5}$), standard in GWAS Catalog (Sollis et al., 2022), is the best for disease prediction by PRS. This could have been tested if a full GWAS study with the genotypes of control and cases as well as their phenotype would have been available. Second, only samples of 90-107 individuals were used, which could be limiting our power to detect relationships. Third, the fraction of the genetic variance not explained yet may still be substantial, so new informative SNPs might be discovered in the future that will allow a more precise inference of the underlying genetic risk. Finally, and most importantly, the poor transferability of PRS among people of different ancestries due to differences in linkage disequilibrium patterns, allele frequencies, heritability, and genetic architecture (Patel and Khera, 2023) makes it difficult to generalize the genetic risk of

certain disease obtained from GWAS data in one population to another population. However, there is consensus that PRS can be applicable within a continent, and consequently the results obtained in this work could be assumed to represent a reasonable inference.

Overall, we conclude that our first hypothesis can only be accepted for CAD, as stroke PRS values showed no difference among populations. Regarding the effect of Neanderthal component (second hypothesis), it cannot be completely discarded that there is a Neanderthal component involved in CAD genetic risk. Finally, research on this topic might be useful not only for evolutionary purposes, but also clinical ones, as this could indicate genetic risk of different populations and therefore, speed up preventive actions.

Supplementary data

https://docs.google.com/spreadsheets/d/1_AI-9POqbtLzwu7QiUlahY6KjAH63fx7FMHpLK1zDNY/edit?usp=sharing

Aknowledgements

Thanks to the Department of Education of the Government of the Basque Country (Gobierno Vasco/Eusko Jaurlaritza) for the student grant provided by the program Ikasiker collaboration in research. Research in SA lab is funded by Funds from the Basque Government to Research Groups of the Basque Country (IT-1693-22).

Bibliography

Abraham G., Rutten-Jacobs L.C., Inouye M. (2021). Risk prediction using polygenic risk scores for prevention of stroke and other cardiovascular diseases. Stroke 52: 2983-2991. https://doi.org/10.1161/strokeaha.120.032619

Anagnostou P., Montinaro F., Sazzini M., Di Vincenzo F., Bisol G.D. (2022). From the Alps to the Mediterranean and beyond: genetics, environment, culture and the «impossible beauty» of Italy. J Anthropol Sci 100: 267-294. https://doi.org/10.4436/jass.10010

Auton A., Abecasis G.R., Altshuler D., Durbin R., Bentley D., Chakravarti A., et al. (2015). A global reference for human Genetic Variation. Nature 526: 68-74. https://doi.org/10.1038/nature15393

Centers for Disease Control and Prevention (4th of May 2023). About Stroke. Consulted on 17/7/2023 in https://www.cdc.gov/stroke/about.htm

Chauhan G., Debette S. (2016). Genetic risk factors for ischemic and hemorrhagic stroke. Curr Cardiol Rep 18: 124. https://doi.org/10.1007/s11886-016-0804-z

Choi S.W., Mak T.S.H., O'Reilly P.F. (2020). Tutorial: A guide to performing polygenic risk score analyses. Nat Protoc 15: 2759-2772. https://doi.org/10.1038/s41596-020-0353-1

Eurostat. (s. f.). Eurostat. https://ec.europa.eu/eurostat/en/

Fox J. (2005). The R Commander: A Basic Statistics Graphical User Interface to R. J Stat Softw 14(9): 1-42. https://doi.org/10.18637/jss.v014.i09

Fox J. (2017). Using the R Commander: A Point-and-Click Interface or R. Boca Raton FL: Chapman and Hall/CRC Press.

Fox J., Bouchet-Valat M. (2024). Rcmdr: R Commander. R package version 2.9-2.

Gibbons A. (2016). Neandertal genes linked to modern diseases. Science 351(6274): 648-649. https://doi.org/10.1126/science.351.6274.648

Groß M. (2019). Mingling with Neanderthals. Curr Biol 29(4): R105-R107. https://doi.org/10.1016/j.cub.2019.01.070

Gunz P., Tilot A.K., Wittfeld K., Teumer A., Shapland C.Y., Van Erp T.G., et al. (2019). Neandertal introgression sheds light on modern human endocranial globularity. Curr Biol 29: 120-127.e5. https://doi.org/10.1016/j.cub.2018.10.065

Hüls A., Czamara D. (2019). Methodological challenges in constructing DNA methylation risk scores. Epigenetics 15: 1-11. https://doi.org/10.1080/15592294.2019.1644879

Koller D., Wendt F.R., Pathak G.A., De Lillo A., De Angelis F., Cabrera-Mendoza B., Tucci S., Polimanti R. (2022). Denisovan and Neanderthal archaic introgression differentially impacted the genetics of complex traits in modern populations. BMC Biology 20: 1-15. https://doi.org/10.1186/s12915-022-01449-2

Khrameeva E.E., Bozek K., He L., Zheng Y., Jiang X., Wei Y., et al. (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun 5: 3584. https://doi.org/10.1038/ncomms4584

Patel A.P., Khera A.V. (2023). Advances and applications of polygenic scores for coronary artery Disease. Annu Rev Med 74: 141-154. https://doi.org/10.1146/annurev-med-042921-112629

Purcell S., Neale B.M., Todd-Brown K., Thomas L.L., Ferreira M., Bender D., et al. (2007). PLINK: a tool set

for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575. https://doi.org/10.1086/519795

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Rychter A.M., Ratajczak A.E., Zawada A., Dobrowolska A., Krela-Kaźmierczak I. (2020). Non-systematic review of diet and nutritional risk factors of cardiovascular disease in obesity. Nutrients 12: 814. https://doi.org/10.3390/nu12030814

Shahjehan R.D., Bhutta B.S. (2023). Coronary artery disease. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK564304/

Simonti C.N., Vernot B., Bastarache L., Bottinger E.P., Carrell D., Chisholm R.L., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. Science 351: 737-741. https://doi.org/10.1126/science.aad2149

Sollis E., Mosaku A., Abid A., Buniello A., Cerezo M., Gil L.F., et al. (2022). The Nhgri-Ebi Gwas Catalog: Knowledgebase and Deposition Resource. Nucleic Acids Res 51: D977-D985. s

Townsend N., Kazakiewicz D., Wright L., Timmis A., Huculeci R., Torbica A., Gale C. P., Achenbach S., Weidinger F., Vardas P. E. (2021). Epidemiology of cardiovascular disease in Europe. Nat Rev Cardiol 19: 133-143. https://doi.org/10.1038/s41569-021-00607-3

Vernot B., Akey J. M. (2014). Resurrecting surviving neandertal lineages from modern human genomes. Science 343: 1017-1021. https://doi.org/10.1126/science.1245938

World Health Organization: WHO. (2020). The top 10 causes of death. Consulted on 10/7/2023 in https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death