Genomic Exploration of Essential Hypertension in African-Brazilian Quilombo Populations: A Comprehensive Approach with Pedigree Analysis and Family-Based Association Studies

Vinícius Magalhães Borges[1,2], Andrea R.V.R. Horimoto[3], Ellen Marie Wijsman[3]; Lilian Kimura[1]; Kelly Nunes[1], Alejandro Q. Nato, Jr.[2], Regina Célia Mingroni-Netto[1]

[1]Centro de Estudos sobre o Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-090, Brazil
[2]Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV 25755, USA
[3]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, 98105 USA

Corresponding authors:

Vinícius Magalhães Borges
Address: 1700 Third Ave (BBSC 239), Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV 25755, USA.

Regina Célia Mingroni-Netto
Address: Centro de Estudos sobre o Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-090, Brazil.

Genomic Exploration of Essential Hypertension in Brazilian Quilombo Populations: A Comprehensive Approach with Pedigree Analysis and Family-Based Association Studies

ABSTRACT

Essential Hypertension (EH) is a medical condition characterized by persistent high blood pressure, and it is one of the most significant public health problems globally, causing approximately 9.4 million deaths annually. The prevalence of EH varies according to genetic ancestry and affects differently specific populations, with 17% in the Americas, 19.2% in the Western Pacific, 23.2% in Europe, 25.1% in Southeast Asia, 26.3% in the Eastern Mediterranean, and 27.2% in Africa. EH is a multifactorial disease, that is, it is determined by genetic factors and influenced by environmental factors.

Although genetic factors are estimated to contribute to around 30-60% of the variation in blood pressure, the genetic complexity of hypertension is not yet fully understood due to the limited knowledge of candidate genes, susceptibility loci, and population-specific differences. Various approaches such as candidate gene-based, genome-wide linkage analysis (GWLA), and family- and population-based genome-wide association studies (GWAS) have been used to identify genetic factors contributing to EH, but part of the heritability of blood pressure-related phenotypes is not explained by the genetic factors known so far, mainly due to methodological limitations.

The main objective of this study was to investigate the genetic basis of EH by mapping regions of interest (ROIs) and investigate candidate genes and variants influencing EH in African-derived individuals from partially isolated populations of quilombo remnants in Vale do Ribeira (São Paulo - Brazil), which were previously well characterized from the clinical, genealogical and population genetics point of view.

Samples from 431 individuals (167 affected, 261 unaffected and 3 with an unknown phenotype) from eight quilombo remnant populations were genotyped using SNP array with approximately 650.000 SNPs. The global ancestry proportions of these populations were estimated to be 47%, 36%, and 16% for African, European, and Native American ancestries, respectively. In addition, genealogical information from 673 individuals was used to construct six pedigrees comprising a total of 1104 individuals. The mapping strategy consisted of a multi-level computational approach. Pedigrees were constructed (GenoPro v.3) based on interviews and kinship coefficient (King v.2.2, MORGAN v.3.4 and PBAP v.1). The dataset was pruned (King v.2.2 and PBAP v.1) to obtain three non-overlapping markers subpanels (PBAP v.2). Haplotype phasing and local ancestry (SNPFlip v.0.0.6, SHAPEIT2 and RFMIX v.2) were performed to obtain SNP allele frequency (ADMIXFRQ v.1) to account for admixture. Genome-wide and dense linkage analyses were performed using the three subpanels of markers (MORGAN v.3.4). Finally, we performed fine-mapping using family-based association studies (GENESIS v.2.24) based on population (MINIMAC v.4) and pedigree (GIGI2) imputed data, and EH-related genes and variant investigation (relying on publicly available databases).

The linkage analysis strategy resulted in the mapping of 22 ROIs with LOD score ≥ 1.45, which contain markers co-segregating with the phenotype. The LOD score range was 1.45-3.03 considering all the linked segments and these 22 ROIs encompassed 2363 genes. As our first fine-mapping strategy, we identified 60 EH-related genes as potential candidates that may contribute to high blood pressure in quilombo remnant pedigrees. In addition, as our second fine-mapping strategy, we identified 118 suggestive or significant variants through family-based association studies. Considering only the common results between the two strategies, we found that 14 genes — PHGDH and S100A10 (ROI1), MFN2 (ROI2), RYR2, EDARADD and MTR (ROI3), SERTAD2 (ROI4), LPP (ROI5), KCNT1 (ROI11), TENM4 (ROI13), P2RX1, ZZEF1 and RPA1 (ROI18), and ALPK2 (ROI20) — were identified within the mapped regions with strong and sufficient evidence in the literature attesting relatedness with the phenotype. These genes harbor 29 SNPs that were either within them or very close, and they were pointed out by family-based association studies as showing suggestive or significant association with hypertension. We also identified 46 other genes within the mapped regions, but with less evidence of relatedness to the phenotype since they were not replicated through family-based association studies.

Overall, the results obtained in this study revealed, through a complementary approach — combining admixture-adjusted genome-wide linkage analysis based on Markov chain Monte Carlo (MCMC) methods, association studies on imputed data, and in silico investigations — genetic regions, variants and candidate genes that shed light on the genetic basis of essential hypertension. These genes are responsible for encoding proteins that play crucial roles in regulating blood pressure, including the regulation of sodium and potassium levels, and the aldosterone pathway, among others. Our findings reveal genes and variants with distinct potential to explain the genetic etiology of essential hypertension observed in quilombo remnant populations.

INTRODUCTION

Essential Hypertension (EH) is a pervasive and sustained raise in arterial blood pressure (BP) and a major cause of premature death worldwide (K. T. Mills et al., 2020). Classified as the primary preventable risk factor for cardiovascular diseases (CVDs) (Fuchs & Whelton, 2020), EH is defined as systolic BP ≥140 mmHg and/or diastolic BP ≥90 mmHg (Barroso et al., 2021; Unger et al., 2021; Whelton et al., 2018).

EH, which affects 1.3 billion people worldwide annually (World Health Organization, 2023), demands comprehensive investigation. Global prevalence is 22.1% (WHO, 2015), 23.9% in Brazil (PNS, 2019). EH is a multifactorial chronic condition, intricately weaving together environmental factors, social determinants, and greatly (30-60%) genetic/epigenetic influences (Arnett & Claas, 2018; Niiranen et al., 2017; Patel et al., 2017). The risk for BP traits varies among ethnic groups (Keaton et al., 2021) and the genetic ancestry significantly influences hypertension risk (Aggarwal et al., 2021), particularly in Afro-descendant populations (Jones et al., 2017; Marden et al., 2016; K. T. Mills et al., 2020; Zhou et al., 2021; Zilbermint et al., 2019).

Yet, the genetic etiology of hypertension — encompassing genes, variants, susceptibility loci, and population disparities — remains elusive (Ehret & Caulfield, 2013; Seidel & Scholl, 2017). Despite advancements from the common disease-common variant and common disease-rare variant hypotheses, methodologies such as Genome-Wide Linkage Analysis (GWLA) and Genome-Wide Association Studies (GWAS) face limitations (Olczak et al., 2021). This gap leaves a segment of EH heritability unexplained by known genetic factors. Moreover, the existing (data as of May 2024) underrepresentation of African (0.21%), African American or Afro-Caribbean, and Native American (0.53%), Hispanic or Latin American (0.39%), Other/Mixed (0.67%) populations (M. C. Mills & Rahal, 2020) in worldwide genomic investigations imposes constraints on the generalizability of results to admixed populations (Buniello et al., 2019; Popejoy & Fullerton, 2016), such as the Brazilian one (68.1% European, 19.6% African, and 11.6% Native American) (de Souza et al., 2019).

This study focuses on the tri-hybrid admixed populations known as “quilombo remnants” in Vale do Ribeira region, São Paulo, Brazil. Quilombo remnants are communities established by runaway or abandoned African slaves, often exhibiting intricate mixtures with European and Native American ancestry. These populations represent a unique model for the study of diseases. They are marked by a high prevalence of EH (Borges & Kimura, 2023), well-defined clinical characterization, semi-isolation, background relatedness, high gene flow between populations, and founder effects (Lemes et al., 2014). They also exhibit relatively homogeneous environmental influences, including lifestyle, dietary habits, and natural habitat, thereby minimizing confounding factors found in larger urban populations. Studying EH in quilombo remnants helps to reduce biases associated with population heterogeneity. This approach improves the signal-to-noise ratio, thereby enhancing statistical power, while providing better representation of admixed populations in genomic studies. It also allows us to uncover chromosomal regions, genes, and variants that may contribute to EH.

Here, we employed an efficient multi-level computational approach that combines: (a) pedigree- and population-based methodologies to account for admixture (African, European, and Native American ancestries) through family analysis (Genome-Wide Linkage Analysis, GWLA) and (b) a two-step fine-mapping strategy based on family-based association studies (FBAS) and investigation of EH-related genes. By reducing the need for multiple tests, minimizing population stratification, and leveraging chromosomal location information provided by meiotic events, this approach aims to contribute to a deeper understanding of the genetic underpinnings of EH.

METHODS

SAMPLES AND SNP GENOTYPING

We conducted 51 trips to Vale do Ribeira (São Paulo, Brazil) from 2000-2020 to obtain samples (peripheral blood for DNA extraction), clinical data (average blood pressure, height, weight, waist circumference and hip circumference), and collect information (sex, age, family relationships, medical history, demographic information, and daily physical activity levels) from 431 consenting individuals aged 17 or older (Fig. 1A). Blood samples were drawn after signing informed consents approved by the ethics

committees (USP/Institute of Biomedical Sciences 111/2001 and USP/Institute of Biosciences 012/2004 and 034/2005).

Genomic DNA was extracted and quantified from each of the 431 blood samples and prepared for SNP genotyping through Axiom® Genome-Wide Human Origins 1 Array SNPs (Fig. 1B) according to Affymetrix requirements (refer to Supp. Material SS1 for details). Raw data was processed, annotated, and subjected to quality control according to Affymetrix Human v.5a threshold (Applied Biosystems, 2020).

From the combined collected data, we constructed six extended pedigrees encompassing 1104 individuals (Table 1) using GenoPro software: 431 genotyped (167 affected, 261 non-affected and 3 unk.) and 673 non-genotyped, from 8 different populations (Abobral [AB], André Lopes [AN], Galvão [GA], Ivaporunduva [IV], Nhunguara [NH], Pedro Cubas [PC], São Pedro [SP], and Sapatu [TU]) (Fig. 2). Pedigree structures underwent validation through calculation of multi-step pairwise kinship coefficients (Φ) conducted by KING-Robust (Manichaikul et al., 2010), MORGAN (Tong & Thompson, 2008; Wijsman et al., 2006), and PBAP (Nato et al., 2015) algorithms.

We filtered and trimmed the dataset (refer to Supp. Material SS1 for details) using KING-Robust (Manichaikul et al., 2010) and PLINK (Chang et al., 2015). Samples with a genotyping rate of less than 95%, as well as SNPs with a genotyping rate below 95%, were excluded. Monomorphic SNPs and SNPs resulting in heterozygous haploid calls across all remaining individuals were removed. SNPs not adhering to Hardy-Weinberg equilibrium ($P < 1x10^{-3}$) were also removed. SNP identification followed the dbSNP standard format (rsID), and their genetic locations (cM) were obtained through the Rutgers Combined Linkage-Physical Map v.3 (Matise et al., 2007). EH was considered a binary outcome, categorizing individuals as hypertensive (SBP ≥ 140 and/or DBP ≥ 90 mmHg) or normotensive (SBP < 140 and/or DBP < 90 mmHg). Individuals diagnosed and/or under medication for EH were classified as hypertensive.

LOCAL ANCESTRY ESTIMATION

We estimated local ancestry fractions (Fig. 1E), which allowed us to determine individual- and pedigree-specific ancestries. The reference dataset was composed of 189 samples from the 1000 Genomes Project (Fairley et al., 2020) and Stanford HGDP SNP Genotyping (Huang et al., 2011) data: 63 European (CEU - Northern Europeans from Utah), 63 African (YRI - Yoruba in Ibadan, Nigeria), and 63 Native American (Colombia, Maya, and Pima populations) samples. Overlapping markers (145,467 SNPs) present in both reference (189 samples) and inference (431 samples) datasets were extracted and both datasets were merged and pruned for missingness (<95% genotyping rate). Haplotypes were inferred using SHAPEIT2 (O'Connell et al., 2014). RFMix (Maples et al., 2013) was used to estimate local ancestry calls. We used in-house scripts to estimate global ancestry fractions for each sample and pedigree ancestries from RFMix output files (refer to Supp. Material SS2 for details).

PEDIGREE ANALYSIS

In our multipoint pedigree linkage analyses, we employed MORGAN (Monte Carlo Genetic Analysis) suite, leveraging its versatility and robust capabilities rooted in the Markov Chain Monte Carlo (MCMC) approach, allowing for simultaneously handling numerous markers and individuals within pedigrees through a sampling methodology (Wijsman et al., 2006).

We implemented a redundant and semi-independent design to yield results that are both independent and comparable. From the complete inference marker set, we selected three distinct non-overlapping subsets (or subpanels) of markers (Fig. 1F) employing specific selection criteria and parameters through PBAP (refer to Supp. Material SS3 for details). The first subpanel, labeled the gold standard, included top-quality markers meeting specific criteria. The other two subpanels, while informative, might have slightly lower quality and were applied for validation of the finding. All three subpanels were crucial for discovery analyses. Parametric linkage analysis was performed using an autosomal-dominant model with a risk allele frequency of 0.01, an incomplete penetrance of 0.70 (for genotypes with 1 or 2 copies of the risk allele), and a phenocopy rate of 0.05. The penetrance was estimated by assessing affected and unaffected individuals within pedigrees. All subsequent steps were performed concurrently for each pedigree (refer to Supp. Material SS3.1 for details).

We estimated allelic frequencies for each SNP marker (Fig. 1G) using ADMIXFRQ (Nafikov et al., 2018). This involved the generation of unique pedigree-specific files organized by ancestry based on local ancestry calls and subpanel information.

To compute LOD scores, we employed the approach in MORGAN outlined as $P_\gamma(Y_T|Y_M)= \sum_{S_M} P_\gamma(Y_T|_M)(S_M|Y)$, where SM denotes the meiosis indicators for all markers (Tong & Thompson, 2008). The LOD score calculation was a two-step process. In the initial step (Fig. 1H), we sampled inheritance vectors (IVs) for the gold standard subpanel using alternate SNP markers (2,500-3,000 SNPs) in a genome-wide “scan" analysis. Preliminary candidate regions were identified as those with a peak LOD score > 1.

In the second step, we conducted a "dense" analysis (Fig. 1J) restricted to the entire chromosomes containing each preliminary candidate region using all three subpanels and sampling IVs using all available SNP markers. LOD scores were calculated again for this second step, defining Regions of Interest (ROIs) as those with a maximum LOD score > 1.50 in at least one subpanel, with mandatory positive results for subpanel 1. ROI boundaries were determined based on LOD score > 1 marker position (refer to Supp. Material SS3.3 for details). Both steps (“scan” and “dense” analysis) were conducted separately for each pedigree.

To ensure the convergence of the sampling process, we performed diagnostic analysis of the MCMC runs (Fig. 1I), evaluating run length, autocorrelation of LOD scores, and run stability using three different graphical tools (refer to Supp. Material SS3.4 for details). The default setup for MCMC runs was 100,000 MC iterations, 40,000 burn-in iterations, 25 saved realizations, 1,000 identity-by-descent (ibd) graphs and output scores saved at every 25 scored MC iterations.

Results of this analysis were used to determine the appropriate running conditions. Once the correct setup for each pedigree was established, we repeated the genome-wide scan and dense analysis, which included sampling of IVs and calculation of LOD scores.

FINE-MAPPING STRATEGIES

Following pedigree analysis adjusted for admixture, we identified and fine-mapped 22 ROIs. We explored the 22 ROIS by analyzing suggestive/significant variants from the SNP array database through family-based association studies and by identifying EH-related genes through *in-silico* investigation.

Family-Based Association Studies

Initially, we addressed population structure through principal components analysis (PCA), first employing MORGAN Checkped (Thompson, 2011; Tong & Thompson, 2008; Wijsman et al., 2006) and PBAP Relationship Check (Nato et al., 2015) algorithms for sample relatedness verification. Pairwise kinship coefficients (Φ) were calculated using KING-robust (Manichaikul et al., 2010). The subsequent steps involved the iterative use of PC-AiR and PC-Relate functions conducted using the GENESIS R package (Gogarten et al., 2019). In the initial iteration, KING-robust estimates informed both kinship and ancestry divergence calculations, with resulting principal components (PCs) used to derive ancestry-adjusted kinship estimates (1st GRM) via PC-Relate. To further refine PCs for ancestry, a secondary PC-AiR run utilized the 1st GRM for kinship and KING-robust estimates for ancestry divergence, yielding new PCs. These new PCs informed a secondary PC-Relate run, culminating in a 2nd GRM. Subsequently, we evaluated variation using the top 10 PCs, selected in accordance with the Kaiser criterion on the Kaiser-Guttman rule (Kaiser, 1960).

We employed a comprehensive two-way independent imputation strategy for variants within each ROI identified through the family analysis (refer to Supp. Material SS4.2 for details). We utilized linkage disequilibrium (LD) information to perform population-based approach. This involved chromosome-wise dataset separation, phasing through SHAPEIT2+duoHMM (Delaneau et al., 2014), imputation using the MINIMAC4 (Fuchsberger et al., 2015) software and filtering through BCFtools (Danecek et al., 2021). Concurrently a pedigree-based strategy incorporating IVs information was applied. This involved chromosome-wise dataset separation and imputation conducted using GIGI2 (Cheung et al., 2013).

To account for the complex correlation structure of the data, we included the 2nd GRM as a random effect to fit the mixed models through fitNullModel function, conducted using the GENESIS R package (Gogarten et al., 2019). Comprehensive testing included 10 ancestry principal components and all available covariates as fixed effects (refer to Supp. Material SS4.1 for details). The final statistical model incorporated the top 8 PCs, along with sex, BMI, and age as significant variables. The next step was to fit the generalized linear mixed model (GLMM). Specifically for single-variant tests we employed a logistic mixed model expressed as:

$$\text{logit}(\pi) = X\alpha + G_j\beta_j + g$$

where $\pi = P(y = 1 \mid X, G_j, g)$ represents the $Nx1$ column vector of probabilities of being affected for the N individuals conditional to covariates, allelic dosages; and random effects; X is the vector of covariates; and α is the vector of fixed covariate effects. We assume that $g \sim N(0, \sigma_\alpha^2 \Phi)$ is a vector $g = (g_{1,\ldots,} g_N)$ of random effects for the N subjects, where σ_α^2 is the additive genetic variance and Φ is the GRM; G_j is a vector with the allelic dosages (0, 1, or 2 copies of the reference allele) or expected dose (in the case of imputed genotypes) at the locus j; and β_j is its corresponding effect size. The null hypothesis of $\beta_j = 0$ was assessed using a multivariate score test (Horimoto et al., 2023).

Finally, for each ROI, two independent FBAS tests were conducted using imputed variant datasets that are either (1) pedigree-based or (2) population-based. The dataset comprised all 431 samples from the combined 6 pedigrees. The single-variant association tests were conducted using the GENESIS R package (Gogarten et al., 2019), implementing the adjusted GLMM to perform Score tests.

We performed multiple testing correction using the effective number of independent markers (M_e) estimated using the Genetic Type I Error Calculator software (Li et al., 2012). We also evaluated adequacy of the analysis modeling through evaluation of the genomic inflation factor (λ) by dividing the median of the chi-square statistics by the median of the chi-square distribution with 1 degree of freedom (Devlin & Roeder, 1999). Analysis was carried out for each imputed variant dataset.

Suggestive or significant association variants underwent curation, assessment, and compilation into a comprehensive database using a custom R script. Data were sourced from NIH dbSNP (Sherry et al., 1999), CADD (Rentzsch et al., 2021) , NCBI PubMed (National Center for Biotechnology Information, 1988), Ensembl Variant Effect Predictor (McLaren et al., 2016), ClinVar (Landrum et al., 2018), Mutation Taster (Schwarz et al., 2014), SIFT (Vaser et al., 2016), PolyPhen (Adzhubei et al., 2010), and VarSome (Kopanos et al., 2019). Annotation included chromosome and physical positions (GRCh37/hg19), dbSNP rsID, associated genes, genomic alterations, variant consequences, exonic functions, and pathogenicity classifications. This procedure was executed independently for each ROI and for both imputed variant datasets.

Furthermore, we assessed LD patterns for variants located within a window size of 500,000 base pairs surrounding each identified suggestive or significantly associated variant. Leveraging the Ensembl Rest API for Linkage Disequilibrium and utilizing the ensemblQueryR tool (Fairbrother-Browne et al., 2023) we set thresholds for $r^2>0.7$ and D'>0.9. This analysis incorporated data from the 1000 Genomes Project (Fairley et al., 2020) for multiple populations, including European (CEU: Utah Residents with Northern and Western European Ancestry), African (YRI: Yoruba in Ibadan, Nigeria), and American populations: CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), PEL (Peruvian in Lima, Peru), and PUR (Puerto Rican in Puerto Rico). Additionally, we utilized the Ensembl REST API VEP to annotate and select LD patterns based on variant consequences.

Investigation of EH-related Genes

To elucidate the hypertension-related implications of genes identified within each ROI, we first identified all ROIs genes using biomaRt R(Durinck et al., 2009). Then we obtained and annotated genes associated with "essential hypertension" or "high blood pressure" according to NCBI PubMed (National Center for Biotechnology Information, 1988), MedGen (National Library of Medicine, n.d.), MalaCards (Rappaport et al., 2013), ClinVar (Landrum et al., 2018), Ensembl BioMart (Kinsella et al., 2011), and GWAS Catalog (Sollis et al., 2023). Finally, we matched both lists, systematically annotating based on

physical position (base pair), cytogenetic band, summary, molecular function, related phenotype, gene ontology, genetic location (cM; GRCh37/hg19), expression patterns, and publication data. Moreover, to prioritize these genes (refer to Supp. Material SS5 for details). we implemented VarElect (Stelzer et al., 2016).

RESULTS AND DISCUSSION

We examined age differences across pedigrees in affected and unaffected individuals, employing Welch Two Sample t-tests, one-way ANOVA, two-way ANOVA, and multiple regression analysis using in-house R scripts. The t-tests revealed significant differences in mean ages between affected and unaffected individuals across all pedigrees (Fig 3), consistently showing younger ages for unaffected individuals (ABDR: 36.52 vs. 48.02 years, P = 0.01478; ANNH: 38.27 vs. 55.95 years, P = 1.403e-05; GASP: 37.52 vs. 52.36 years, P < 0.001; PC: 36.03 vs. 55.40 years, P < 0.001; IV: 33.63 vs. 63.00 years, P = 1.072e-05; TU: 36.26 vs. 52.18 years, P = 0.00036), as seen in Supp. Table ST1. Overall, the combined data showed a mean age of 36.59 years for unaffected individuals versus 53.35 years for affected individuals (P < 2.2e-16). The one-way ANOVA was performed with 5 degrees of freedom for pedigrees and 184 for residuals, the sum of squares for pedigrees was 2837 and for residuals was 55278, resulting in mean squares of 567.5 and 300.4, respectively. Comparing the average age of affected individuals across pedigrees for both sexes, this analysis yielded an F value of 1.889 with a p-value of 0.0982, indicating marginally non-significant differences in average age across pedigrees. The two-way ANOVA highlighted a significant effect of affected status on average age (P < 0.001). However, neither sex nor the interaction between sex and affected status showed significant effects on average age, with p-values of 0.589 and 0.670 respectively. This suggests that while affected status significantly influences average age, sex and the interaction between sex and affected status do not have a significant impact. The multiple regression model showed that affected status significantly influenced average age (P < 0.001). The coefficients for sex and other pedigrees are not statistically significant. The overall model's performance is moderate, as indicated by the R-squared value of 0.2181, suggesting that about 21.81% of the variability in average age can be explained by the predictors in the model. Additionally, the F-statistic is significant (P < 2.2e-16), indicating that the model as a whole is significant in predicting average age.

We calculated ancestry proportions for individual admixture segments (local ancestry) using data from 145,467 SNPs considering all the 431 samples (Fig. 4A) and individually per pedigree (Fig. 4B). The three main components (PCs), PC1, PC2 and PC3, explain 12.42%, 6.17%, and 1.68% of the variance, respectively (Fig. 5A), allowing to visualize the genetic distance between the inference and the reference datasets (Fig. 5B). All six quilombo remnants pedigrees studied have a high degree of admixture. Among all the 431 individuals the estimates are 47.4%, 36.3%, and 16.1%, respectively, for African, European, and Native American ancestries (Table 2). The pedigrees from ABDR and PC had the highest (51.5%) estimates of African ancestral contribution. In comparison, TU had the highest (50.8%) estimate of European ancestral contribution, and ABDR had the highest estimate (16.9%) of Native American ancestral contribution. For comparison, the ancestry proportions for the general Brazilian population were estimated as 19.6%, 68.1%, and 11.6% for African, European, and Native American ancestral contributions (de Souza et al., 2019), respectively, with vivid contrasts among each Brazilian region (Supp. Table ST2-3).

The MCMC runs diagnosis were conducted, within each pedigree, on the smallest chromosome exhibiting a positive LOD score (Supp. Fig. SF2-7).

Through pedigree analysis, we identified 22 ROIs containing 2363 genes (Example in Fig. 6A). Employing family-based association studies (Fig. 6B-C), we uncovered 117 variants with suggestive/significantly association with hypertension (Supp. Table ST4). Furthermore, our investigation of EH-related genes yielded 60 promising candidate genes (Table 3). Highlighting the common results between both fine mapping strategies, 14 genes (highlighted in Table 3) were identified within the mapped regions with strong evidence of association with the phenotype. These regions harbor 29 SNPs (highlighted in Supp. Table ST4) implicated by our family-based association studies. These genes include *PHGDH* and *S100A10* (ROI1); *MFN2* (ROI2); *RYR2*, *EDARADD*, and *MTR* (ROI3); *SERTAD2* (ROI4); *LPP* (ROI5); *KCNT1* (ROI11); *TENM4* (ROI13); *P2RX1*, *ZZEF1*, and *RPA1* (ROI18); and *ALPK2* (ROI20). Our approach of combining admixture-adjusted linkage analysis using MCMC methods, association studies on imputed data, and in silico investigations yielded additional results to pursue, as different strategies occasionally

supported the role of the same gene in contributing to essential hypertension in quilombo remnants populations.

Twenty ROIs were supported by the investigation of EH-related genes as a fine-mapping strategy, and 17 of them were additionally supported by family-based association studies. To prioritize our results, we developed a comprehensive score based on the weights assigned to each strategy (Supp. Table ST5), including linkage analysis, EH genes investigation, and association studies. The scores allowed to classify the ROIs into three tiers based on their priority level (Table 4): high (top 20% of the ROIs), intermediate (30% of the ROIs), and low (50% of the ROIs).

Note that the SNPs linked to the EH phenotype identified in this study were genotyped using genomic arrays and are therefore common variants. Many of these variants are situated in non-coding or intergenic regions and may or may not have recognized functional or regulatory effects. Although these SNPs are not expected to affect the phenotype directly, several of these variants are in LD with variants of more significant impact. Notably, among these, some may be rare and remain ungenotyped. Specifically, we identified unique tag SNPs: 77 non-coding SNPs (3' UTR, 5' UTR, or TF binding), 196 regulatory, and 15 missense SNPs (Supp. Table ST6).

The study, as presented, has some limitations. One such limitation is that the investigation of EH-related genes, as part of the fine-mapping strategy, was focused on genes previously supported by literature as being related to blood pressure regulation. Consequently, it is possible that genes which have never been implicated in EH may contain rare or novel variants relevant to the origin of hypertension in quilombo remnant populations.

In conclusion, this study presented a unique multi-level computational approach that combined mapping strategies to deal with large family data, which provided reliable results. By using genome-wide linkage analyses based on MCMC methods and adjusted for admixture, association studies as the primary fine-mapping strategy, and limiting analyses to candidate genomic regions, the study took advantage of meiotic information provided by pedigrees while simultaneously reducing the need for multiple tests and avoiding population stratification. Therefore, the ROIs identified in this study are credible and provide valuable insights into the genetic basis of essential hypertension in the quilombo remnants populations.

Conducting analyses by merging all six pedigrees into only one would be a formidable challenge, and probably not feasible. However, the prospect of replicating these analyses using alternative computational packages is exciting and not an impossible task. Our study has demonstrated that blood pressure and hypertension in the quilombo remnant populations are likely influenced by multiple genes, possibly in a polygenic or oligogenic mechanism of inheritance. We have identified several loci across different chromosomes that contain genes and variants involved in the development of hypertension. Additionally, we have identified genomic regions of interest not previously associated with EH and will therefore be important targets for future research.

Furthermore, our study has the potential to shed light on important genomic regions, genes, and variants that are specific to African-derived populations. By focusing on this population, we have provided insights into the genetic factors that contribute to hypertension in a group that has been often underrepresented in genetic studies and databases.

In order to overcome the limitations, future steps of this investigation will involve the use of Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) data from this dataset. WGS and WES will enable the investigation of coding and non-coding variants within all ROIs, with a focus on rare variants that may have a higher impact on gene functioning. To optimize this process, we will prioritize high and intermediate ROIs when filtering variants detected after WES and WGS, addressing the low-priority ROIs afterward.

ACKNOWLEDGMENTS

The authors would like to thank the support of Dr. Diogo Meyer (IB/USP - Brazil), Dr. Julia M. Pavan Soler (IME/USP - Brazil), Dr. Suely Ruiz Giolo (UFPR - Brazil), Dr. Paulo Otto (IB/USP - Brazil) and Dr. Alexandre da Costa Pereira (Harvard University - USA) for valuable suggestions. The authors also thank

Dr. Christian Kubisch (University Medical Center Hamburg / Eppendorf - Germany) for ideas which stimulated the development of this project.

We acknowledge funding from CEPID-FAPESP (Research Center on the Human Genome and Stem Cells, grants 1998/14254-2 and 2013/08028-1, and FAPESP/INCT-CNPq 2014/50931-3 led by Dr. Mayana Zatz). This research also received support from grant FAPESP 2012/18010-0, titled “Balancing selection in the human genome: detection, causes, and consequences,” led by Dr. Diogo Meyer (IB/USP-Brazil). Additionally, we are grateful to CAPES for the sandwich Ph.D. fellowship #88887.371219/2019-00 and CNPq for the Ph.D. fellowship #142193/2017-8.

We also acknowledge funding from the Marshall University Joan C. Edwards School of Medicine, WV-INBRE grant (NIH P20GM103434), Bench-to-Bedside Pilot grant (Nato) under the West Virginia Clinical and Translational Science Institute (WV-CTSI) grant (NIH 5U54GM104942), and Nato startup fund.

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Aggarwal, R., Chiu, N., Wadhera, R. K., Moran, A. E., Raber, I., Shen, C., Yeh, R. W., & Kazi, D. S. (2021). Racial/Ethnic Disparities in Hypertension Prevalence, Awareness, Treatment, and Control in the United States, 2013 to 2018. *Hypertension*, *78*(6), 1719–1726. https://doi.org/10.1161/HYPERTENSIONAHA.121.17570

Applied Biosystems. (2020). *Axiom Analysis Suite 5.1 User Guide*. *703307*, 33. https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/axiom-analysis-suite.html

Arnett, D. K., & Claas, S. A. (2018). Omics of blood pressure and hypertension. *Circulation Research*, *122*(10), 1409–1419. https://doi.org/10.1161/CIRCRESAHA.118.311342

Barroso, W. K. S., Rodrigues, C. I. S., Bortolotto, L. A., Mota-Gomes, M. A., Brandão, A. A., Feitosa, A. D. de M., Machado, C. A., Poli-de-Figueiredo, C. E., Amodeo, C., Mion, D., Barbosa, E. C. D., Nobre, F., Guimarães, I. C. B., Vilela-Martin, J. F., Yugar-Toledo, J. C., Magalhães, M. E. C., Neves, M. F. T., Jardim, P. C. B. V., Miranda, R. D., … Nadruz, W. (2021). Diretrizes Brasileiras de Hipertensão Arterial – 2020. *Arquivos Brasileiros de Cardiologia*, *116*(3), 516–658. https://doi.org/10.36660/abc.20201238%0Ahttps://abccardiol.org/article/diretrizes-brasileiras-de-hipertensao-arterial-2020/

Borges, V. M., & Kimura, L. (2023). Panorama of arterial hypertension in quilombos in Brazil: a narrative review. *Physis: Revista de Saúde Coletiva*, *33*(e33050). https://doi.org/https://doi.org/10.1590/S0103-7331202333050.en

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Cheung, C. Y. K., Thompson, E. A., & Wijsman, E. M. (2013). GIGI: An approach to effective imputation of dense genotypes on large pedigrees. *American Journal of Human Genetics*, *92*(4), 504–516. https://doi.org/10.1016/j.ajhg.2013.02.011

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

de Souza, A. M., Resende, S. S., de Sousa, T. N., & de Brito, C. F. A. (2019). A systematic scoping review of the genetic ancestry of the brazilian population. *Genetics and Molecular Biology*, *42*(3), 495–508. https://doi.org/10.1590/1678-4685-gmb-2018-0076

Delaneau, O., Marchini, J., McVeanh, G. A., Donnelly, P., Lunter, G., Myers, S., Gupta-Hinch, A., Iqbal, Z., Mathieson, I., Rimmer, A., Xifara, D. K., Kerasidou, A., Churchhouse, C., Altshuler, D. M., Gabriel, S. B., Lander, E. S., Gupta, N., Daly, M. J., DePristo, M. A., … Peltonen, L. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nature Communications 2014 5:1*, *5*(1), 1–9. https://doi.org/10.1038/ncomms4934

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Ehret, G. B., & Caulfield, M. J. (2013). Genes for blood pressure: An opportunity to understand hypertension. *European Heart Journal*, *34*(13), 951–961. https://doi.org/10.1093/eurheartj/ehs455

Fairbrother-Browne, A., García-Ruiz, S., Reynolds, R. H., Ryten, M., & Hodgkinson, A. (2023). ensemblQueryR: fast, flexible and high-throughput querying of Ensembl LD API endpoints in R. *GigaByte*, *2023*, 1. https://doi.org/10.46471/GIGABYTE.91

Fairley, S., Lowy-Gallego, E., Perry, E., & Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research*, *48*(D1), D941–D947. https://doi.org/10.1093/NAR/GKZ836

Fuchs, F. D., & Whelton, P. K. (2020). High Blood Pressure and Cardiovascular Disease. *Hypertension*, *75*(2), 285–292. https://doi.org/10.1161/HYPERTENSIONAHA.119.14240

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2015). minimac2: faster genotype imputation. *Bioinformatics (Oxford, England)*, *31*(5), 782–784. https://doi.org/10.1093/BIOINFORMATICS/BTU704

Gogarten, S. M., Sofer, T., Chen, H., Yu, C., Brody, J. A., Thornton, T. A., Rice, K. M., & Conomos, M. P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics (Oxford, England)*, *35*(24), 5346–5348. https://doi.org/10.1093/BIOINFORMATICS/BTZ567

Horimoto, A. R. V. R., Boyken, L. A., Blue, E. E., Grinde, K. E., Nafikov, R. A., Sohi, H. K., Nato, A. Q., Bis, J. C., Brusco, L. I., Morelli, L., Ramirez, A., Dalmasso, M. C., Temple, S., Satizabal, C., Browning, S. R., Seshadri, S., Wijsman, E. M., & Thornton, T. A. (2023). Admixture mapping implicates 13q33.3 as ancestry-of-origin locus for Alzheimer disease in Hispanic and Latino populations. *Human Genetics and Genomics Advances*, *4*(3), 19. https://doi.org/10.1016/J.XHGG.2023.100207

Huang, L., Jakobsson, M., Pemberton, T. J., Ibrahim, M., Nyambo, T., Omar, S., Pritchard, J. K., Tishkoff, S. A., & Rosenberg, N. A. (2011). Haplotype variation and genotype imputation in African populations. *Genetic Epidemiology*, *35*(8), 766. https://doi.org/10.1002/GEPI.20626

Jones, E. S., Spence, J. D., McIntyre, A. D., Nondi, J., Gogo, K., Akintunde, A., Hackam, D. G., & Rayner, B. L. (2017). High frequency of variants of candidate genes in black Africans with low renin-resistant hypertension. *American Journal of Hypertension*, *30*(5), 478–483. https://doi.org/10.1093/ajh/hpw167

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Keaton, J. M., Hellwege, J. N., Giri, A., Torstenson, E. S., Kovesdy, C. P., Sun, Y. V., Wilson, P. W. F., O'donnell, C. J., Edwards, T. L., Hung, A. M., & Velez Edwards, D. R. (2021). Associations of biogeographic ancestry with hypertension traits. *Journal of Hypertension*, *39*(4), 633–642. https://doi.org/10.1097/HJH.0000000000002701

Kimura, L., Ribeiro-Rodrigues, E. M., De Mello Auricchio, M. T. B., Vicente, J. P., Batista Santos, S. E., & Mingroni-Netto, R. C. (2013). Genomic ancestry of rural African-derived populations from Southeastern Brazil. *American Journal of Human Biology*, *25*(1), 35–41. https://doi.org/10.1002/ajhb.22335

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database*, *2011*, 1–9. https://doi.org/10.1093/database/bar030

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics*, *35*(11), 1978–1980. https://doi.org/10.1093/bioinformatics/bty897

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., … Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lemes, R. B., Nunes, K., Meyer, D., Mingroni-Netto, R. C., & Otto, P. A. (2014). Estimation of inbreeding and substructure levels in African-derived Brazilian quilombo populations. *Human Biology*, *86*(4), 276–288. https://doi.org/10.13110/humanbiology.86.4.0276

Li, M. X., Yeung, J. M. Y., Cherny, S. S., & Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, *131*(5), 747–756. https://doi.org/10.1007/s00439-011-1118-2

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *American Journal of Human Genetics*, *93*(2), 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Marden, J. R., Walter, S., Kaufman, J. S., & Glymour, M. M. (2016). African Ancestry, Social Factors, and Hypertension among Non-Hispanic Blacks in the Health and Retirement Study. *Biodemography and Social Biology*, *62*(1), 19–35. https://doi.org/10.1080/19485565.2015.1108836

Matise, T. C., Chen, F., Chen, W., De La Vega, F. M., Hansen, M., He, C., Hyland, F. C. L., Kennedy, G. C., Kong, X., Murray, S. S., Ziegle, J. S., Stewart, W. C. L., & Buyske, S. (2007). A second-generation combined linkage-physical map of the human genome. *Genome Research*, *17*(12), 1783–1786. https://doi.org/10.1101/gr.7156307

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 1–14. https://doi.org/10.1186/s13059-016-0974-4

Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. *Nature Reviews Nephrology*, *16*(4), 223–237. https://doi.org/10.1038/s41581-019-0244-2

Mills, M. C., & Rahal, C. (2020). The GWAS Diversity Monitor tracks diversity by disease in real time. *Nature Genetics*, *52*(3), 242–243. https://doi.org/10.1038/S41588-020-0580-Y

Nafikov, R. A., Nato, A. Q., J., Sohi, H., Wang, B., Brown, L., Horimoto, A. R., Vardarajan, B. N., Barral, S. M., Tosto, G., Mayeux, R. P., Thornton, T. A., Blue, E., & Wijsman, E. M. (2018). Analysis of Pedigree Data in Populations with Multiple Ancestries: Strategies for Dealing with Admixture in Caribbean Hispanic Families from the ADSP. *Genet Epidemiol.*, *42*(6), 500–515. https://doi.org/10.1002/gepi.22133

National Center for Biotechnology Information. (1988). *National Center for Biotechnology Information*. Bethesda (MD): National Library of Medicine (US). https://www.ncbi.nlm.nih.gov/

National Library of Medicine. (n.d.). *MedGen*.

Nato, A. Q., Chapman, N. H., Sohi, H. K., Nguyen, H. D., Brkanac, Z., & Wijsman, E. M. (2015). PBAP: A pipeline for file processing and quality control of pedigree data with dense genetic markers. *Bioinformatics*, *31*(23), 3790–3798. https://doi.org/10.1093/bioinformatics/btv444

Niiranen, T. J., McCabe, E. L., Larson, M. G., Henglin, M., Lakdawala, N. K., Vasan, R. S., & Cheng, S. (2017). Risk for hypertension crosses generations in the community: Amulti-generational cohort study. *European Heart Journal*, *38*(29), 2300–2308. https://doi.org/10.1093/eurheartj/ehx134

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J. E., Rudan, I., McQuillan, R., Fraser, R. M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A. F., Vitart, V., … Marchini, J. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genetics*, *10*(4). https://doi.org/10.1371/JOURNAL.PGEN.1004234

Olczak, K. J., Taylor-Bateman, V., Nicholls, H. L., Traylor, M., Cabrera, C. P., & Munroe, P. B. (2021). Hypertension genetics past, present and future applications. *Journal of Internal Medicine*, *290*(6), 1130–1152. https://doi.org/10.1111/JOIM.13352

Patel, R. S., Masi, S., & Taddei, S. (2017). Understanding the role of genetics in hypertension. *European Heart Journal*, *38*(29), 2309–2312. https://doi.org/10.1093/EURHEARTJ/EHX273

Pesquisa Nacional de Saúde. (2019). Percepção do estado de saúde, estilos de vida, doenças crônicas e saúde bucal. In *Ibge*. http://www.pns.icict.fiocruz.br/arquivos/Portaria.pdf

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, *538*(7624), 161–164. https://doi.org/10.1038/538161a

Rappaport, N., Nativ, N., Stelzer, G., Twik, M., Guan-Golan, Y., Stein, T. I., Bahir, I., Belinky, F., Morrey, C. P., Safran, M., & Lancet, D. (2013). MalaCards: An integrated compendium for diseases and their annotation. *Database*, *2013*, 1–14. https://doi.org/10.1093/database/bat018

Rentzsch, P., Schubach, M., Shendure, J., & Kircher, M. (2021). CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Medicine*, *13*(1), 1–12. https://doi.org/10.1186/s13073-021-00835-9

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. https://doi.org/10.1038/nmeth.2890

Seidel, E., & Scholl, U. I. (2017). Genetic mechanisms of human hypertension and their implications for blood pressure physiology. *Physiological Genomics*, *49*(11), 630–652. https://doi.org/10.1152/physiolgenomics.00032.2017

Sherry, S. T., Ward, M., & Sirotkin, K. (1999). dbSNP - database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Research*, *9*(8), 677–679. https://doi.org/10.1101/gr.9.8.677

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., … Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(D1), D977–D985. https://doi.org/10.1093/nar/gkac1010

Stelzer, G., Plaschkes, I., Oz-Levi, D., Alkelai, A., Olender, T., Zimmerman, S., Twik, M., Belinky, F., Fishilevich, S., Nudel, R., Guan-Golan, Y., Warshawsky, D., Dahary, D., Kohn, A., Mazor, Y., Kaplan, S., Iny Stein, T., Baris, H. N., Rappaport, N., … Lancet, D. (2016). VarElect: The phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genomics*, *17*(Suppl 2). https://doi.org/10.1186/s12864-016-2722-2

Thompson, E. (2011). The structure of genetic linkage data: From LIPED to 1M SNPs. *Human Heredity*, *71*(2), 86–96. https://doi.org/10.1159/000313555

Tong, L., & Thompson, E. (2008). Multilocus lod scores in large pedigrees: combination of exact and approximate calculations. *Human Heredity*, *65*(3), 142–153. https://doi.org/10.1159/000109731

Unger, T., Borghi, C., Charchar, F., Khan, N. A., Poulter, N. R., Prabhakaran, D., Ramirez, A., Schlaich, M., Stergiou, G. S., Tomaszewski, M., Wainford, R. D., Williams, B., & Schutte, A. E. (2021). Hypertension: New Guidelines from the International Society of Hypertension. *American Family Physician*, *103*(12), 763–765. www.aafp.org/afpAmericanFamilyPhysician763PracticeGuidelines

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9. https://doi.org/10.1038/nprot.2015.123

Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., … Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Pr. *Journal of the American College of Cardiology*, *71*(19), e127–e248. https://doi.org/10.1016/j.jacc.2017.11.006

Wijsman, E. M., Rothstein, J. H., & Thompson, E. A. (2006). Multipoint linkage analysis with many multiallelic or dense diallelic markers: Markov chain-Monte Carlo provides practical approaches for genome scans on general pedigrees. *American Journal of Human Genetics*, *79*(5), 846–858. https://doi.org/10.1086/508472

World Health Organization. (2023). *Global report on hypertension The race against a silent killer* (World Health Organization, Ed.). https://www.who.int/publications/i/item/9789240081062

Zhou, B., Carrillo-Larco, R. M., Danaei, G., Riley, L. M., Paciorek, C. J., Stevens, G. A., Gregg, E. W., Bennett, J. E., Solomon, B., Singleton, R. K., Sophiea, M. K., Iurilli, M. L., Lhoste, V. P., Cowan, M. J., Savin, S., Woodward, M., Balanova, Y., Cifkova, R., Damasceno, A., … Ezzati, M. (2021). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, *398*(10304), 957–980. https://doi.org/10.1016/S0140-6736(21)01330-1

Zilbermint, M., Gaye, A., Berthon, A., Hannah-Shmouni, F., Faucz, F. R., Lodish, M. B., Davis, A. R., Gibbons, G. H., & Stratakis, C. A. (2019). ARMC 5 Variants and Risk of Hypertension in Blacks: MH- GRID Study. *Journal of the American Heart Association*, *8*(14). https://doi.org/10.1161/JAHA.119.012508

TABLES AND FIGURES

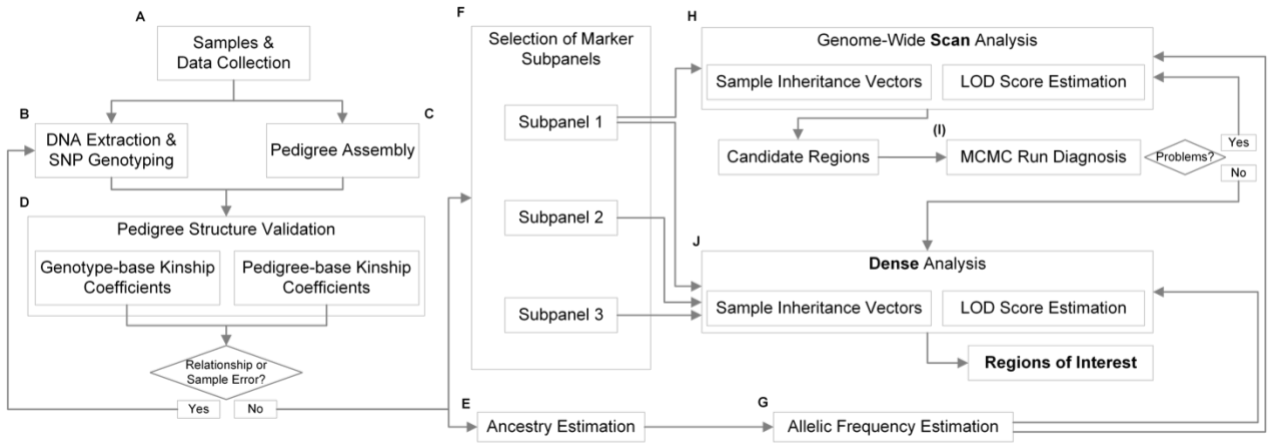


Figure 1. Schematic diagram of the pedigree analysis (GWLA) mapping workflow. A) Samples and collected data datasets; B) DNA extraction, quantification, SNP genotyping and quality control; C) Pedigree assembly; D) Pedigree structure validation analysis; E) Local ancestry estimation; F) Selection of markers for 3 subpanels; G) Allelic frequencies estimation; H) Genome-Wide scan analysis; I) MCMC convergence diagnostic analysis; J) Dense analysis resulting in the ROIs.

Table 1. Distribution of samples across pedigrees

Pedigree	**Genotyped**				**Non-genotyped**	**Total**
	Affected	Unaffected	Unknown	Subtotal		
ABDR	38	29	1	68	105	173
ANNH	37	54	-	91	117	208
GASP	45	49	1	95	88	183
IV	12	34	1	47	110	157
PC	16	51	-	67	130	197
TU	19	44	-	63	123	186
Total	**167**	**261**	**3**	**431**	**673**	**1104**

Total samples included by pedigree. Samples are separated into genotyped (and affected, unaffected and unknown phenotype) and non-genotyped. Abobral (ABDR), André Lopes and Nhunguara (ANNH), Galvão and São Pedro (GASP), Ivaporunduva (IV), Pedro Cubas (PC) and Sapatu (TU).

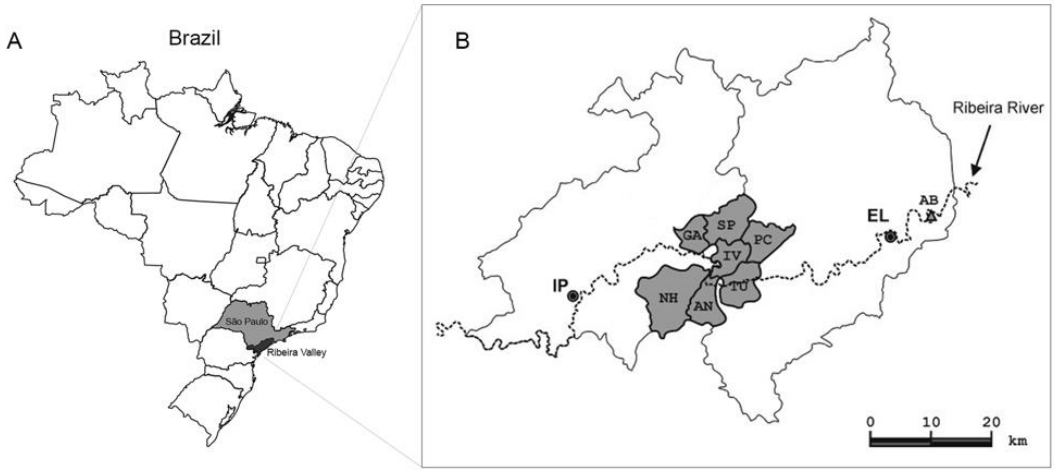


Figure 2. Geographical location of Quilombo populations. A) Brazilian territory in South America, with the State of São Paulo highlighted in gray and the Ribeira Valle in a darker shade of gray. B) Location of quilombo populations: AB (Abobral), AN (Andre Lopes), GA (Galvão), IV (Ivaporanduva), NH (Nhunguara), PC (Pedro Cubas), SP (São Pedro), and TU (Sapatu) (adapted from Kimura et al., (2013). The black dots denote the urban centers of Eldorado (EL) and Iporanga (IP).

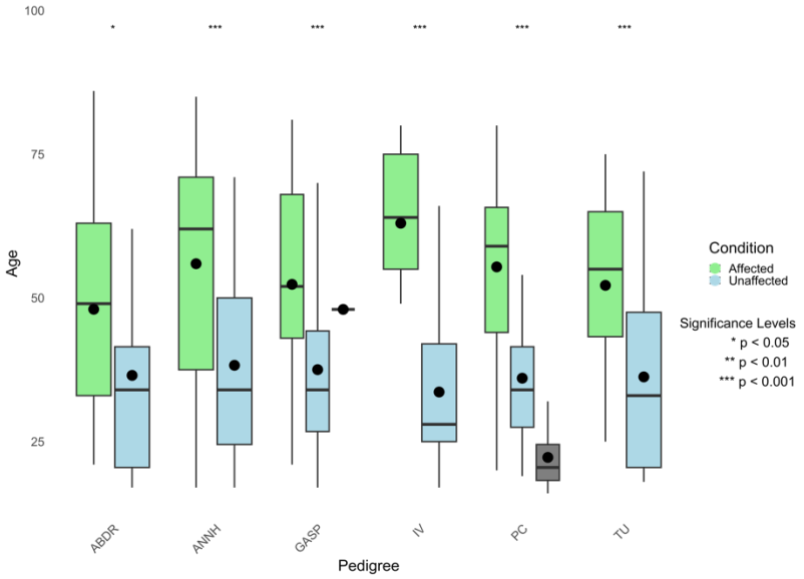


Figure 3. Distribution of ages for both affected and unaffected individuals across different pedigrees. The box plots represent the interquartile range (IQR) with the median age indicated by the line within each box. The whiskers extend to 1.5 times the IQR, with individual points outside the whiskers representing potential outliers. The small black dots indicate the mean ages for each group, and the error bars represent the standard deviation around these means. Affected individuals are shown in green, unaffected individuals are shown in blue, and unknown individuals are shown in gray.

Table 2. Ancestry proportions

Pedigrees	Sample size	Ancestry Proportions		
		African	European	Native American
ABDR	68	51.5%	31.4%	16.9%
GASP	95	46.8%	36.9%	16.2%
ANNH	91	50.8%	33.2%	15.8%
IV	47	47.3%	35.9%	16.7%
TU	63	33.4%	50.8%	15.7%
PC	67	51.5%	32.4%	16.0%
Total	**431**	**47.4%**	**36.3%**	**16.1%**

Data are presented by pedigree, with the corresponding sample size and ancestry proportions (%) that reflect the contribution of African, European, and Native American ancestries, respectively. The total dataset is described as “Total.”

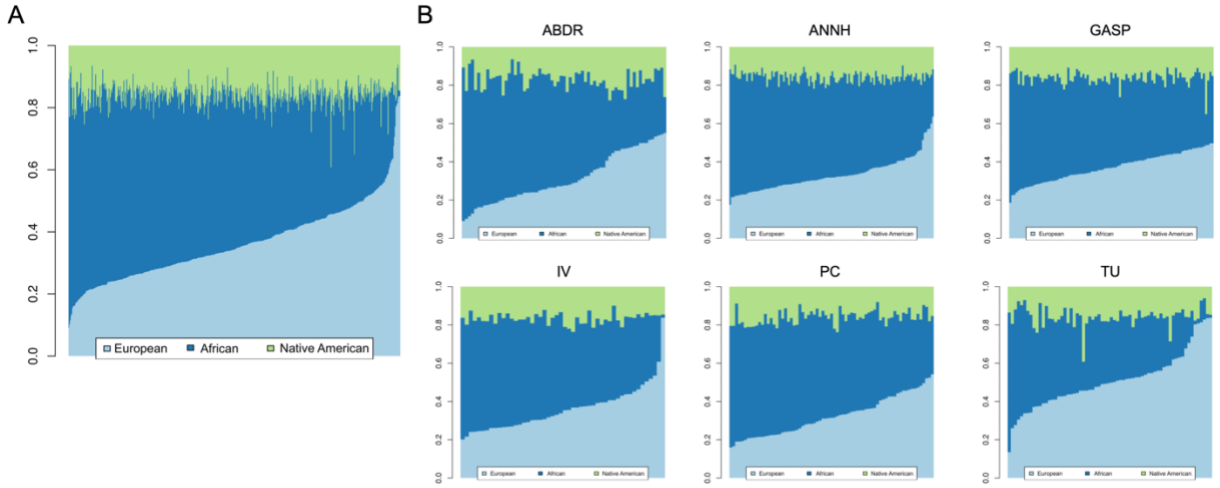


Figure 4. Graphic representation of global ancestry estimates: A) considering all 431 samples; B) considering each pedigree individually. The colors refer to the ancestral contribution: Green (Native American), dark blue (African) and light blue (European);

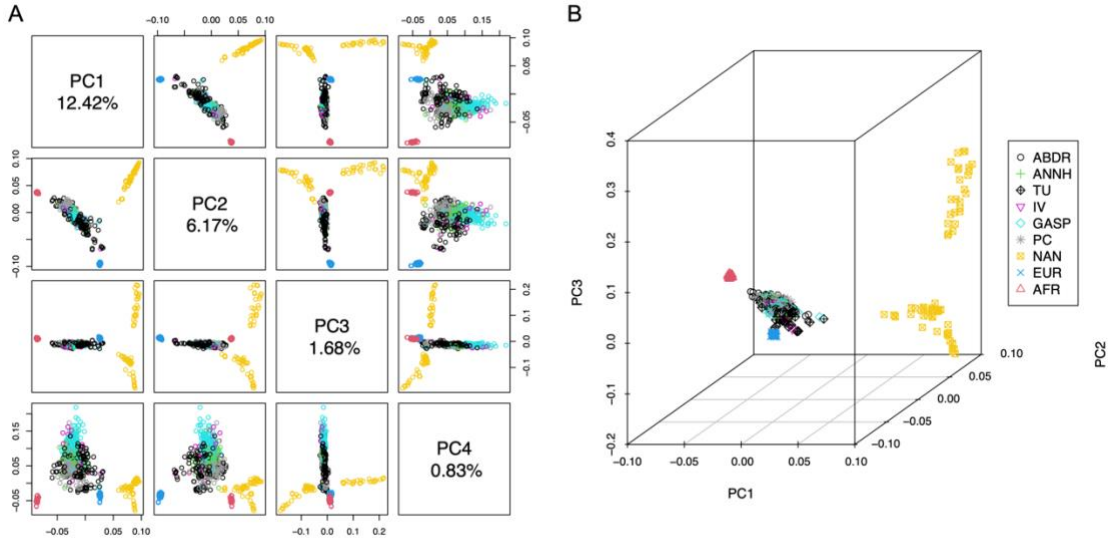


Figure 5. Graphic representation of the genetic distance between all individuals over Principal Component Analysis (PCA). The reference samples from European (EUR), African (AFR) and Native American (NAM) populations were represented by blue, red and yellow circles, respectively. A) PCA results for the first 4 PCs; B) 3D PCA results.

Table 3. Main genes resulting from the three strategies per region of interest (ROI): linkage analysis, EH-related gene investigation (EH) and association studies (AS).

Gene	Cytoband	ROI	EH	AS
ALPK2	18q21.31-q21.32	20	●	●
EDARADD	1q42.3	3	●	●
KCNT1	9q34.3	11	●	●
LPP	3q27.3-q28	5	●	●
MFN2	1p36.22	2	●	●
MTR	1q43	3	●	●
P2RX1	17p13.2	18	●	●
PHGDH	1p12	1	●	●
RPA1	17p13.3	18	●	●
RYR2	1q43	3	●	●
S100A10	1q21.3	1	●	●
SERTAD2	2p14	4	●	●
TENM4	11q14.1	13	●	●
ZZEF1	17p13.2	18	●	●
ABCC9	12p12.1	14	●	
ACE	17q23.3	19	●	
ADIPOQ	3q27.3	5	●	
APOE	19q13.32	21	●	
ARHGEF17	11q13.4	13	●	
BORCS7	10q24.32	12	●	
CASP3	4q35.1	7	●	
CDH13	16q23.3	17	●	
CNNM2	10q24.32	12	●	
CORIN	4p12	6	●	
CYGB	17q25.1	19	●	
CYP17A1	10q24.32	12	●	
CYP2C19	10q23.33	12	●	
CYP2C9	10q23.33	12	●	
EDN1	6p24.1	9	●	
HTR2A	13q14.2	16	●	
KDR	4p12	6	●	
KIT	4p12	6	●	
KLKB1	4q35.2	7	●	
KNG1	3q27.3	5	●	
LPL	8p21.3	10	●	
MC4R	18q21.32	20	●	
MEIS1	2p14	4	●	
NT5C2	10q24.32-q24.33	12	●	
PDE3A	12p12.2	14	●	
PDE4D	5q11.2-q12.1	8	●	
PHACTR1	6p24.1	9	●	
PLCE1	10q23.33	12	●	
PRKCA	17q24.2	19	●	
RBM47	4p14	6	●	
RBP4	10q23.33	12	●	
SGCZ	8p22	10	●	
SLC1A4	2p14	4	●	
SLC39A14	8p21.3	10	●	
SOCS3	17q25.3	19	●	
SST	3q27.3	5	●	
TBX2	17q23.2	19	●	
TGFB1	19q13.2	21	●	
TIMP2	17q25.3	19	●	
TLR3	4q35.1	7	●	
TMOD4	1q21.3	1	●	
TNFRSF1B	1p36.22	2	●	
UCP2	11q13.4	13	●	
VEGFC	4q34.3	7	●	
WBP1L	10q24.32	12	●	
ZDHHC2	8p22	10	●	

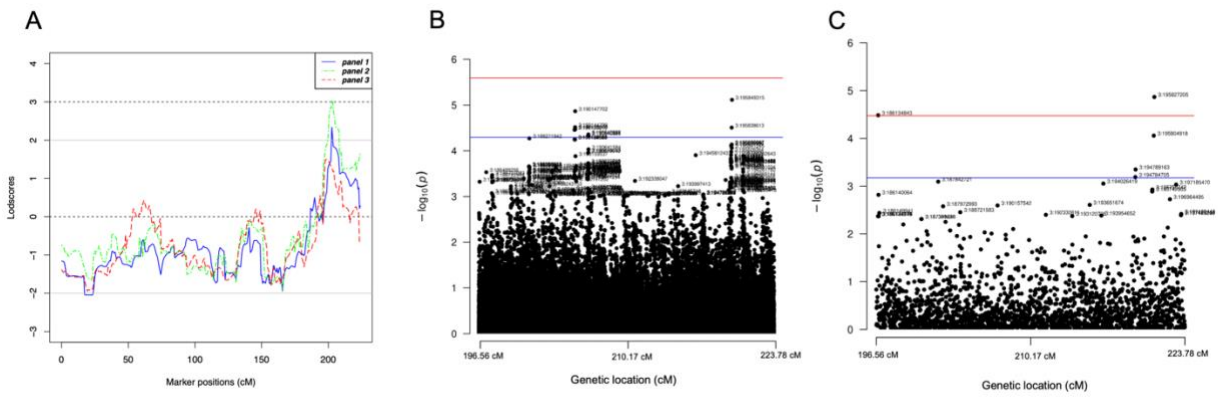
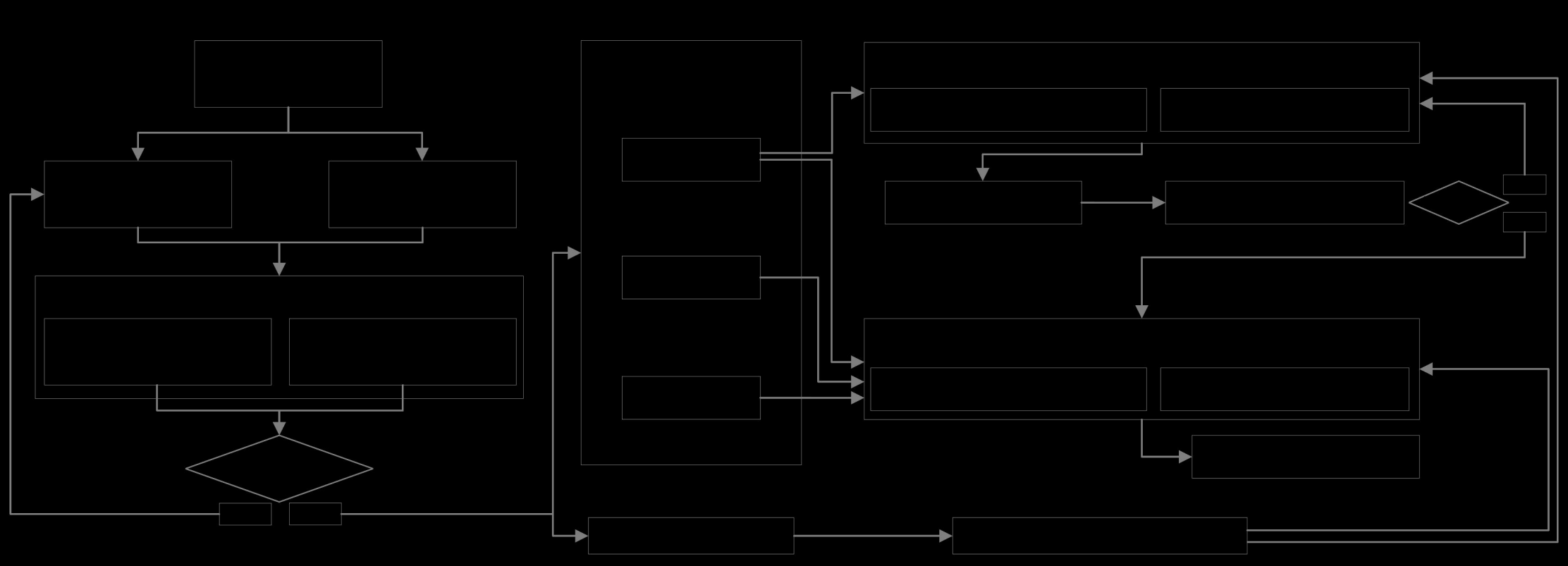


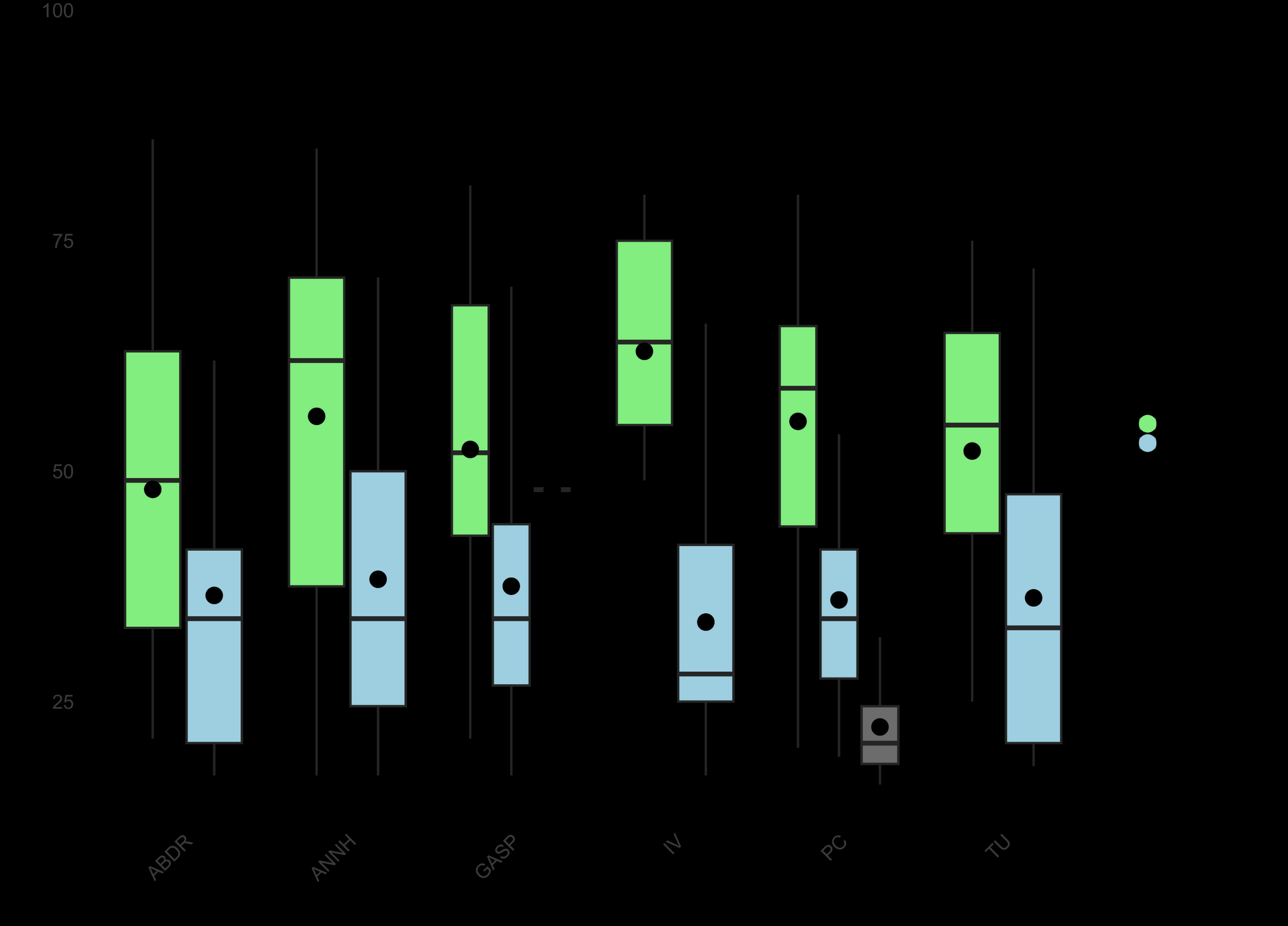
Figure 6. Illustrative representation of the results for ROI5. A) Dense linkage analysis for all 3 subpanels; B) Strategy performed with population-based imputed data; C) Strategy performed with pedigree imputed data. The black dots represent SNPs (identified by chr:physical position), the blue lines refer to the suggestive p-value (-log10(p)) and the red lines to the significant p-value (-log10(p)).

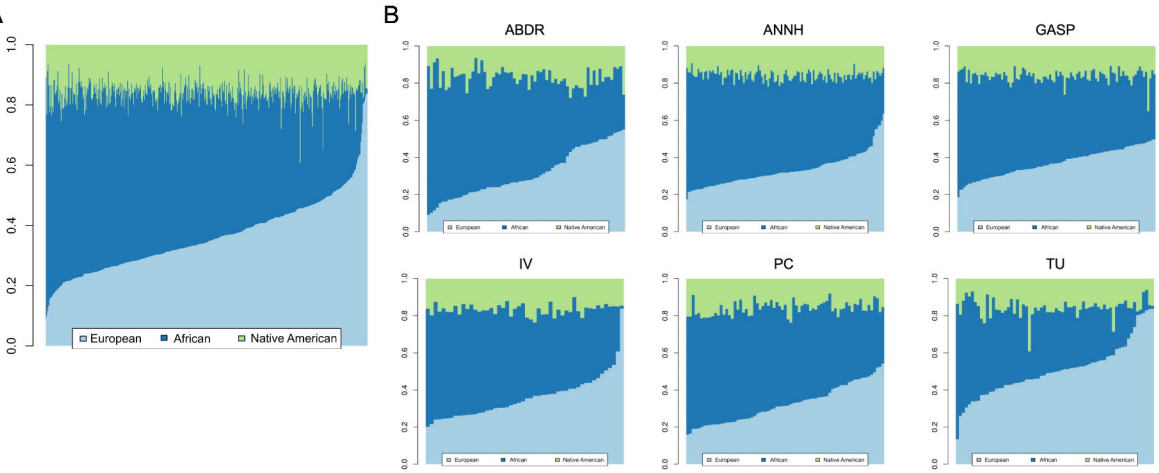
Table 4. Ranked Score-Weighted ROIs

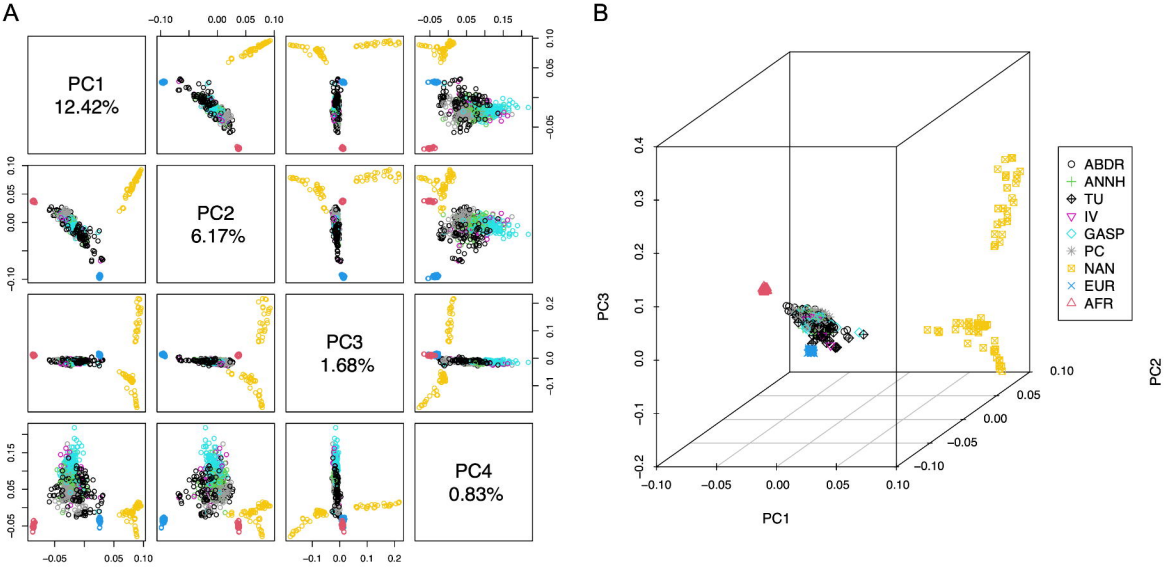
ROI	Cytoband	ROI size (Mb)	EH Genes	Linkage Analysis		Assoc. Studies Sugg./Sig.	Gene Relevance	TOTAL
				Peak LOD Score	Subpanels consensus			
5	3q27.3-q29	11.4 Mb	Yes | 4 genes (+1p)	3.036 (+6p)	3 (+3p)	Sig. (+3p)	Yes (+5p)	18
12	10q23.33-q25.1	11.2 Mb	Yes | 9 genes (+1p)	2.795 (+5p)	3 (+3p)	Sig. (+3p)	Yes (+5p)	17
13	11q13.4-q14.1	7.6 Mb	Yes | 3 genes (+1p)	2.414 (+4p)	3 (+3p)	Sugg. (+1p)	Yes (+5p)	14
19	17q23.2-25.3	19 Mb	Yes | 9 genes (+1p)	2.138 (+3p)	3 (+3p)	Sugg. (+2p)	Yes (+5p)	14
3	1q43	1.3 Mb	Yes | 3 genes (+1p)	2.161 (+3p)	2 (+2p)	Sugg. (+2p)	Yes (+5p)	13
9	6p24.1-p22.3	8 Mb	Yes | 2 genes (+1p)	2.621 (+5p)	2 (+2p)	Sugg. (+2p)	Yes (+3p)	13
10	8p23.1-p21.3	9.8 Mb	Yes | 4 genes (+1p)	2.658 (+5p)	3 (+3p)	Sugg. (+1p)	Yes (+3p)	13
7	4q32.3-q35.2	20.6 Mb	Yes | 4 genes (+1p)	2.322 (+4p)	2 (+2p)	Sugg. (+2p)	Yes (+3p)	12
21	19q13.12-13.32	10 Mb	Yes | 2 genes (+1p)	2.503 (+4p)	3 (+3p)	Sugg. (+1p)	Yes (+3p)	12
14	12p12.3-p11.23	11.1 Mb	Yes | 2 genes (+1p)	2.662 (+5p)	2 (+2p)	Sugg. (+1p)	Yes (+2p)	11
8	5q12.1-q13.2	11.1 Mb	Yes | 1 gene (+1p)	2.117 (+3p)	3 (+3p)	Sugg. (+2p)	Yes (+2p)	11
18	17p13.3-p13.2	4.3 Mb	Yes | 3 genes (+1p)	1.771 (+2p)	2 (+2p)	Sugg. (+1p)	Yes (+4p)	10
2	1p36.21-p36.22	3.2 Mb	Yes | 2 genes (+1p)	1.824 (+2p)	3 (+3p)	Sugg. (+1p)	Yes (+3p)	10
16	13q14.13-q21.33	25.6 Mb	Yes | 1 gene (+1p)	2.554 (+4p)	3 (+3p)	No (0p)	Yes (+2p)	10
1	1p12-q21.3	32.5 Mb	Yes | 3 genes (+1p)	1.503 (+1p)	2 (+2p)	Sugg. (+1p)	Yes (+4p)	9
4	2p14	3.4 Mb	Yes | 3 genes (+1p)	1.906 (+2p)	2 (+2p)	Sugg. (+1p)	Yes (+3p)	9
20	18q21.32	1.7 Mb	Yes | 2 genes (+1p)	1.608 (+1p)	3 (+3p)	Sugg. (+1p)	Yes (+3p)	9
6	4p15.1-q12	26.5 Mb	Yes | 4 genes (+1p)	1.938 (+2p)	3 (+3p)	No (0p)	Yes (+3p)	9
11	9q34.3	0.8 Mb	Yes | 1 gene (+1p)	1.418 (+1p)	3 (+3p)	Sugg. (+1p)	Yes (+2p)	8
17	16q23.3-q24.1	4.5 Mb	Yes | 1 gene (+1p)	2.175 (+3p)	2 (+2p)	No (0p)	Yes (+2p)	8
15	12q24.32-q24.33	4.6 Mb	No (0p)	2.087 (+3p)	2 (+2p)	Sugg. (+1p)	Yes (+1p)	7
22	19q13.41-13.42	3.2 Mb	No (0p)	1.860 (+2p)	3 (+3p)	Sugg. (+1p)	Yes (+1p)	7

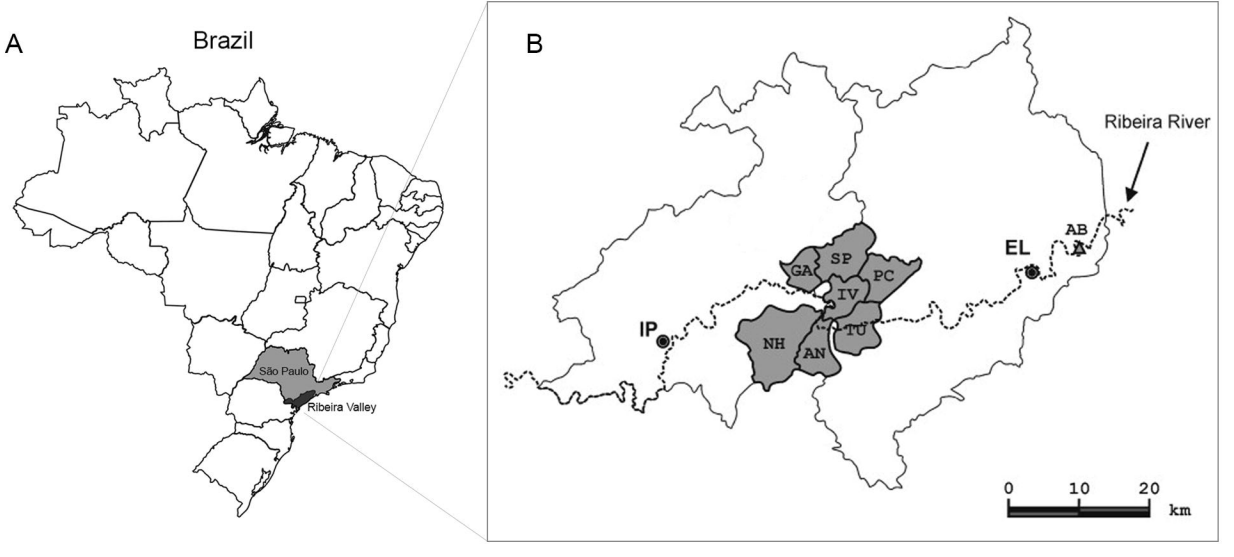
The ROIs are classified into three tiers according to their priority level: the top 20% are labeled as high priority (dark gray), the medium 30% as intermediate priority (medium gray), and the bottom 50% as low priority (light gray).

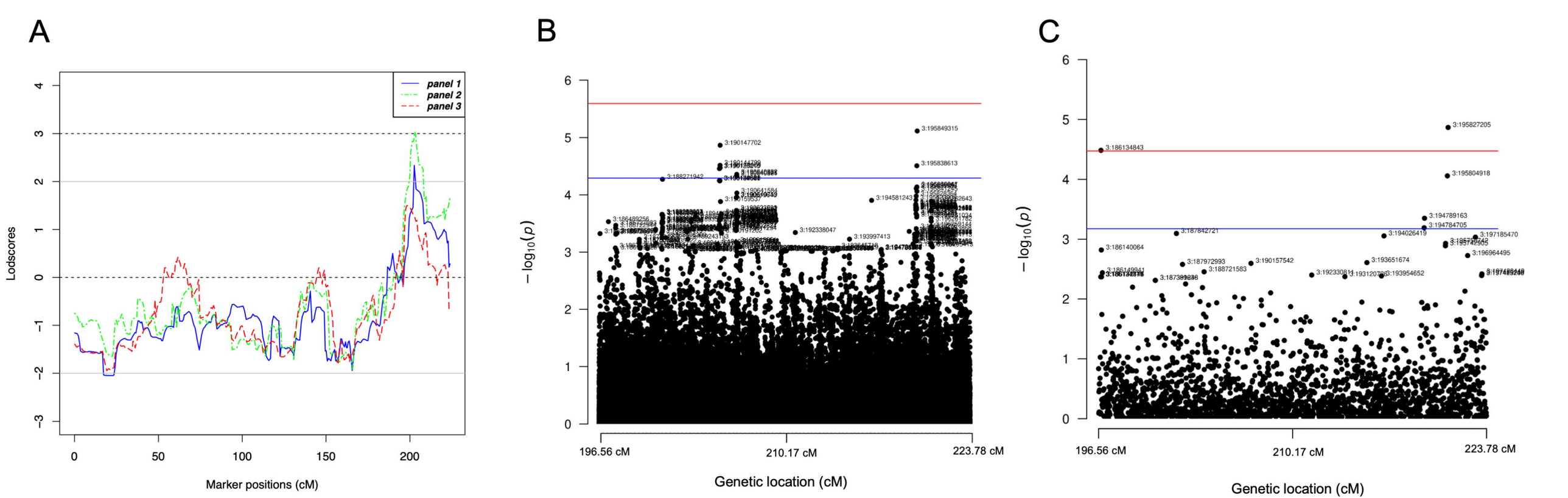



100
75
50
25
ABDR
ANNH
GASP
IV
PC
TU


A
B
ABDR
ANNH
GASP
IV
PC
TU
European
African
Native American
0.0
0.2
0.4
0.6
0.8
1.0


A
PC1
12.42%
PC2
6.17%
PC3
1.68%
PC4
0.83%
B
PC1
PC2
PC3
ABDR
ANNH
TU
IV
GASP
PC
NAN
EUR
AFR


A
Brazil
São Paulo
Ribeira Valley
B
Ribeira River
AB
EL
IP
SP
GA
PC
IV
TU
NH
AN
0
10
20
km


A
panel 1
panel 2
panel 3
Lodscores
Marker positions (cM)
B
$-\log_{10}(p)$
196.56 cM
210.17 cM
223.78 cM
Genetic location (cM)
C
$-\log_{10}(p)$
196.56 cM
210.17 cM
223.78 cM
Genetic location (cM)