




OPEN

## Recent natural selection conferred protection against schizophrenia by non-antagonistic pleiotropy

Javier González-Peñas<sup>1,2,3</sup>, Lucía de Hoyos<sup>1,2,4</sup>, Covadonga M. Díaz-Caneja<sup>1,2,3,5</sup>, Álvaro Andreu-Bernabeu<sup>1,2,5</sup>, Carol Stella<sup>1,2</sup>, Xaquín Gurriarán<sup>1,2</sup>, Lourdes Fañanás<sup>3,6</sup>, Julio Bobes<sup>3,7</sup>, Ana González-Pinto<sup>3,8</sup>, Benedicto Crespo-Facorro<sup>3,9</sup>, Lourdes Martorell<sup>3,10</sup>, Elisabet Vilella<sup>3,10</sup>, Gerard Muntané<sup>3,10</sup>, María Dolores Molto<sup>3,11,12</sup>, Jose Carlos Gonzalez-Piqueras<sup>3,12,13</sup>, Mara Parellada<sup>1,2,3,5</sup>, Celso Arango<sup>1,2,3,5</sup> & Javier Costas<sup>14</sup>

Schizophrenia is a debilitating psychiatric disorder associated with a reduced fertility and decreased life expectancy, yet common predisposing variation substantially contributes to the onset of the disorder, which poses an evolutionary paradox. Previous research has suggested balanced selection, a mechanism by which schizophrenia risk alleles could also provide advantages under certain environments, as a reliable explanation. However, recent studies have shown strong evidence against a positive selection of predisposing loci. Furthermore, evolutionary pressures on schizophrenia risk alleles could have changed throughout human history as new environments emerged. Here in this study, we used 1000 Genomes Project data to explore the relationship between schizophrenia predisposing loci and recent natural selection (RNS) signatures after the human diaspora out of Africa around 100,000 years ago on a genome-wide scale. We found evidence for significant enrichment of RNS markers in derived alleles arisen during human evolution conferring protection to schizophrenia. Moreover, both partitioned heritability and gene set enrichment analyses of mapped genes from schizophrenia predisposing loci subject to RNS revealed a lower involvement in brain and neuronal related functions compared to those not subject to RNS. Taken together, our results suggest non-antagonistic pleiotropy as a likely mechanism behind RNS that could explain the persistence of schizophrenia common predisposing variation in human populations due to its association to other non-psychiatric phenotypes.

Schizophrenia is a complex mental disorder characterized by psychosis (i.e. delusions or hallucinations), social and emotional withdrawal, and cognitive deficits. Although the typical onset does not occur until adolescence or early adulthood, epidemiological and molecular evidence has consistently described the early neurodevelopmental nature of the disorder<sup>1–3</sup>. Affecting around 0.5% of the human population<sup>4</sup>, schizophrenia is associated with increased mortality<sup>5</sup> and a low rate of recovery of only 13.5%<sup>6</sup>. Moreover, schizophrenia has been associated with high rates of comorbid illnesses such as coronary heart disease, stroke, diabetes, respiratory diseases<sup>5,7</sup> and a 15% rate of unnatural deaths, including suicide<sup>5</sup>.

<sup>1</sup>Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, Calle Ibiza, 43, 28009 Madrid, Spain. <sup>2</sup>Instituto de Investigación Sanitaria Gregorio Marañón (IISGM), Madrid, Spain. <sup>3</sup>CIBERSAM, Centro Investigación Biomédica en Red Salud Mental, Madrid, Spain. <sup>4</sup>Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. <sup>5</sup>School of Medicine, Universidad Complutense, Madrid, Spain. <sup>6</sup>Department of Evolutionary Biology, Ecology and Environmental Sciences, Faculty of Biology, University of Barcelona, Barcelona, Spain. <sup>7</sup>Faculty of Medicine and Health Sciences - Psychiatry, Universidad de Oviedo, ISPA, INEUROPA, Oviedo, Spain. <sup>8</sup>BIOARABA Health Research Institute, OSI Araba, University Hospital, University of the Basque Country, Vitoria, Spain. <sup>9</sup>Department of Psychiatry, Hospital Universitario Virgen del Rocío, Universidad de Sevilla, Sevilla, Spain. <sup>10</sup>Hospital Universitari Institut Pere Mata, IISPV, Universitat Rovira i Virgili, Reus, Spain. <sup>11</sup>Department of Genetics, University of Valencia, Campus of Burjassot, Valencia, Spain. <sup>12</sup>Department of Medicine, Universitat de València, Valencia, Spain. <sup>13</sup>Fundación Investigación Hospital Clínico de Valencia, INCLIVA, 46010 Valencia, Spain. <sup>14</sup>Instituto de Investigación Sanitaria (IDIS) de Santiago de Compostela, Complejo Hospitalario Universitario de Santiago de Compostela (CHUS), Servizo Galego de Saúde (SERGAS), Santiago de Compostela, Galicia, Spain. ✉email: javier.gonzalez@iisgm.com

Family, twin, and adoption studies have estimated schizophrenia heritability to be 65–80%<sup>8</sup>, thus corroborating the major contribution of genetic variability to the development of the disorder. Although both rare copy number variants (CNVs)<sup>9</sup> and disruptive mutations<sup>10–12</sup> importantly contribute to the polygenic architecture of schizophrenia, genome-wide association studies (GWAS) have demonstrated that contribution from common genetic variation (mainly single nucleotide polymorphisms, SNPs) account for up to a half of the genetic variance in liability<sup>12,13</sup>. Methods capturing the cumulative risk such as polygenic risk scores (PGS)<sup>14</sup> or linkage disequilibrium (LD) score regression (LDSC)<sup>15</sup> have demonstrated an uneven distribution of schizophrenia genetic heritability across the genome, with enrichment of synaptic or brain-related pathways<sup>16–18</sup>, even though predisposing *loci* are genome-wide distributed.

Given the reduced fertility<sup>19</sup> and the early mortality associated with schizophrenia<sup>5</sup>, negative selection should result in the purge of deleterious alleles that contribute to the disorder. However, the persistence of common predisposing alleles and comparable prevalence rates of schizophrenia in human populations suggests an evolutionary paradox<sup>20,21</sup>. Some authors have attempted to explain this paradox by balanced selection, by which schizophrenia risk alleles could also provide advantages under certain environments<sup>22</sup>. In this sense, a recent study described that schizophrenia alleles are linked with creativity in the general population<sup>23</sup>, thus supporting classical theories that proposed schizophrenia vulnerability as a price to pay for the development of language and abstract or creative thinking<sup>24,25</sup>. Recent works, however, have shown strong evidence against a positive selection of predisposing *loci*<sup>26–28</sup>.

Most of the previous studies addressing the evolutionary nature of schizophrenia lack a clear dynamic perspective of the environment during human evolution. Classical theories stating schizophrenia as a by-product of human ancestral evolution may not be relevant to explain the recent evolution of schizophrenia predisposing alleles<sup>24,25</sup>. In this context, the study of human populations offers a unique opportunity to address recent natural selection (RNS)<sup>29,30</sup> that has taken place since African populations initiated dispersion into Eurasia around 60,000–100,000 years ago<sup>31–33</sup>. A recent study has demonstrated that genetic predisposing variants to schizophrenia exert similar risks across East Asian and European populations, thus pointing to a shared genetic basis for schizophrenia regardless of the ethnic or cultural backgrounds<sup>34</sup>. However, at the same time, SNPs associated with schizophrenia display greater population differentiation than matched control SNPs<sup>29</sup>, and some are reportedly even population-specific<sup>35,36</sup>. This scenario suggests that recent population differentiation events could have shaped the allelic distribution of schizophrenia polygenic variation across populations.

Moreover, given the high comorbidity rates of schizophrenia with other mental and somatic conditions and the pleiotropic effects of its predisposing *loci*<sup>37,38</sup>, selective forces for other traits could have acted to differentiate human populations across schizophrenia predisposing alleles, thus generating uneven schizophrenia genetic risks across populations. For instance, the risk allele of *SLC39A8*, a genome-wide significant gene in schizophrenia<sup>39</sup>, could have arisen after the human expansion to Europe to ease adaptation to colder climates by reducing blood pressure and risk of hypertension<sup>22,40</sup>. Older environments, however, could have promoted different genetic adaptations. In this sense, the schizophrenia alleles that have emerged and rose their frequency due to polygenic adaptation after human divergence from Neanderthal (500,000–700,000 years ago)<sup>41,42</sup> may not be adaptive throughout modern environments, and therefore subject to selective pressures in modern humans.

The classical approach to discover RNS signals has consisted of seeking variants with large allele frequency differences between populations<sup>29</sup>. Many of the available methods are based on  $F_{ST}$  statistical tests, which need an a priori classification of studied individuals in populations<sup>43</sup>. However, many methods do not account for the hierarchical population structure due to the unequal differentiation within populations, which may lead to detection of many false positives<sup>44</sup>. Moreover, classification in fixed population groups may be challenging when population ascertainment does not reveal differentiated clusters or individuals are misclassified<sup>45,46</sup>. Some individual-based methods performing genome-wide selection sweeps based on principal component analysis (PCA) have been recently developed to address these challenges<sup>47,48</sup>.

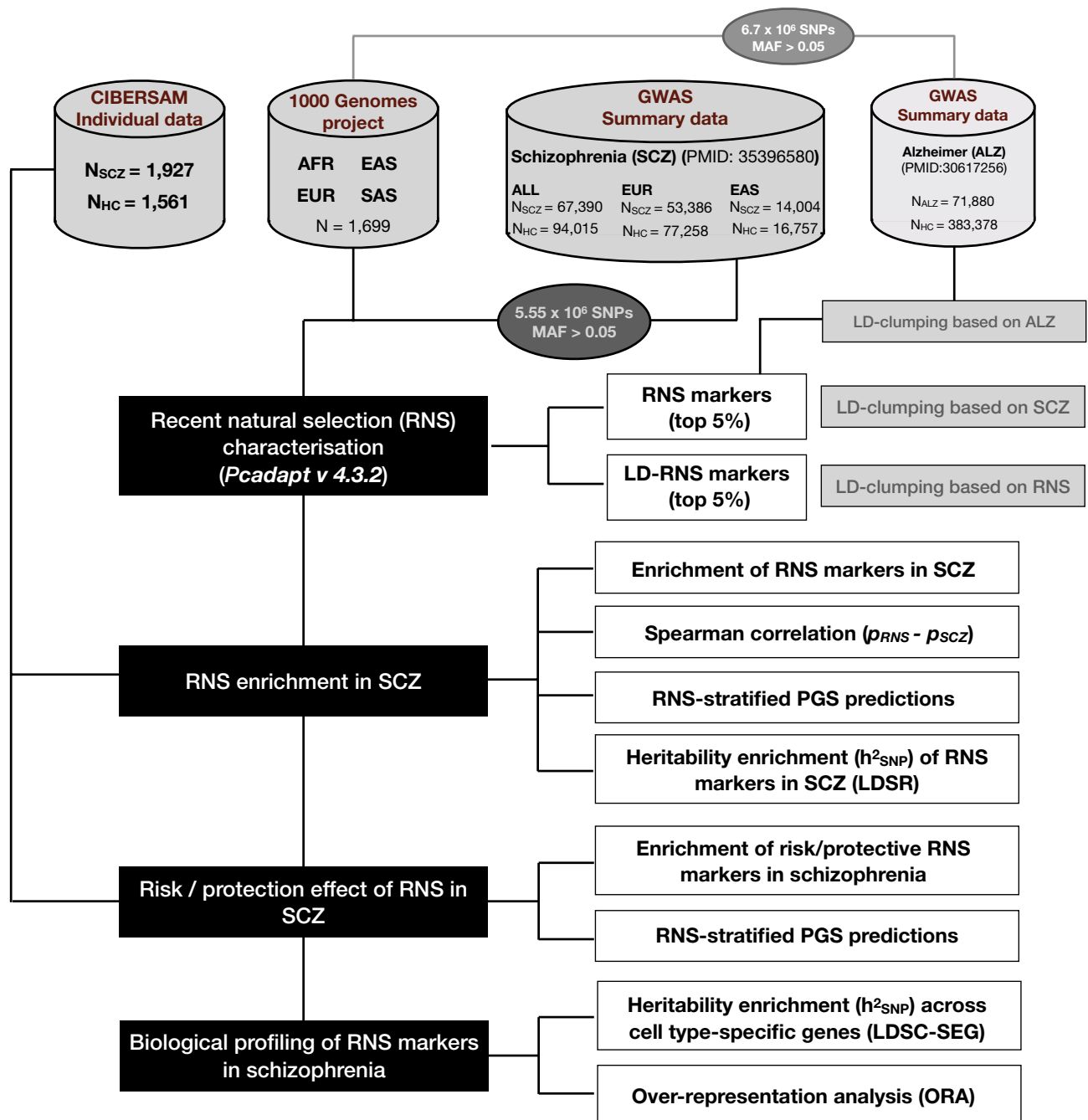
Here we aimed to perform a comprehensive approach to evaluate the effect of RNS on common genetic variation predisposing to schizophrenia (Fig. 1). First, by using *Pcadapt*<sup>48</sup>, a recently developed PCA-based method, we described a significant association between RNS signatures and schizophrenia. Second, by focusing on the derived alleles arisen during human evolution, we demonstrated a significant trend toward fixation of derived protective alleles. Third, we explored the biological and cell-type enrichments of the RNS markers associated with schizophrenia and observed less brain and neuronal specificity than across markers not subject to RNS. Finally, among the derived alleles arisen during recent evolution, we observed greater pleiotropy for protective than for risk alleles, thus suggesting that recent protective selection to schizophrenia is related to other phenotypes resulting from recent human adaptation to different environments.

## Results

**RNS signatures across the genome.** We assessed the signatures of recent polygenic adaptation across the genome after the human diaspora out of Africa (about 100 kya<sup>31</sup>). *Pcadapt*<sup>48</sup> was used to study RNS signals across human populations from 1000 Genomes Project data. Up to 3 principal components (PCs) were kept based on the proportion of variance explained (Supplementary Fig. 1). The first PC separates African populations from the rest, while the second and third PCs differentiate between Asian and European populations (Supplementary Fig. 2), similar to previous results described using 1000 Genomes phase 1 data<sup>47</sup>.

RNS was characterized across the genome (Supplementary Fig. 3) and RNS  $p$ -values for each variant were computed ( $p_{RNS}$ ). A marked accumulation of SNPs subject to RNS ( $p_{RNS} < 0.05$ ) was detected (Supplementary Fig. 4–5, one-sample Kolmogorov–Smirnov test  $p < 2 \times 10^{-16}$ ).

In addition, to rule out the possibility that many of the described RNS markers were false positives due to discontinuous PC space<sup>50</sup>, we estimated  $F_{st}$  statistical parameters to describe adaptation selection markers between

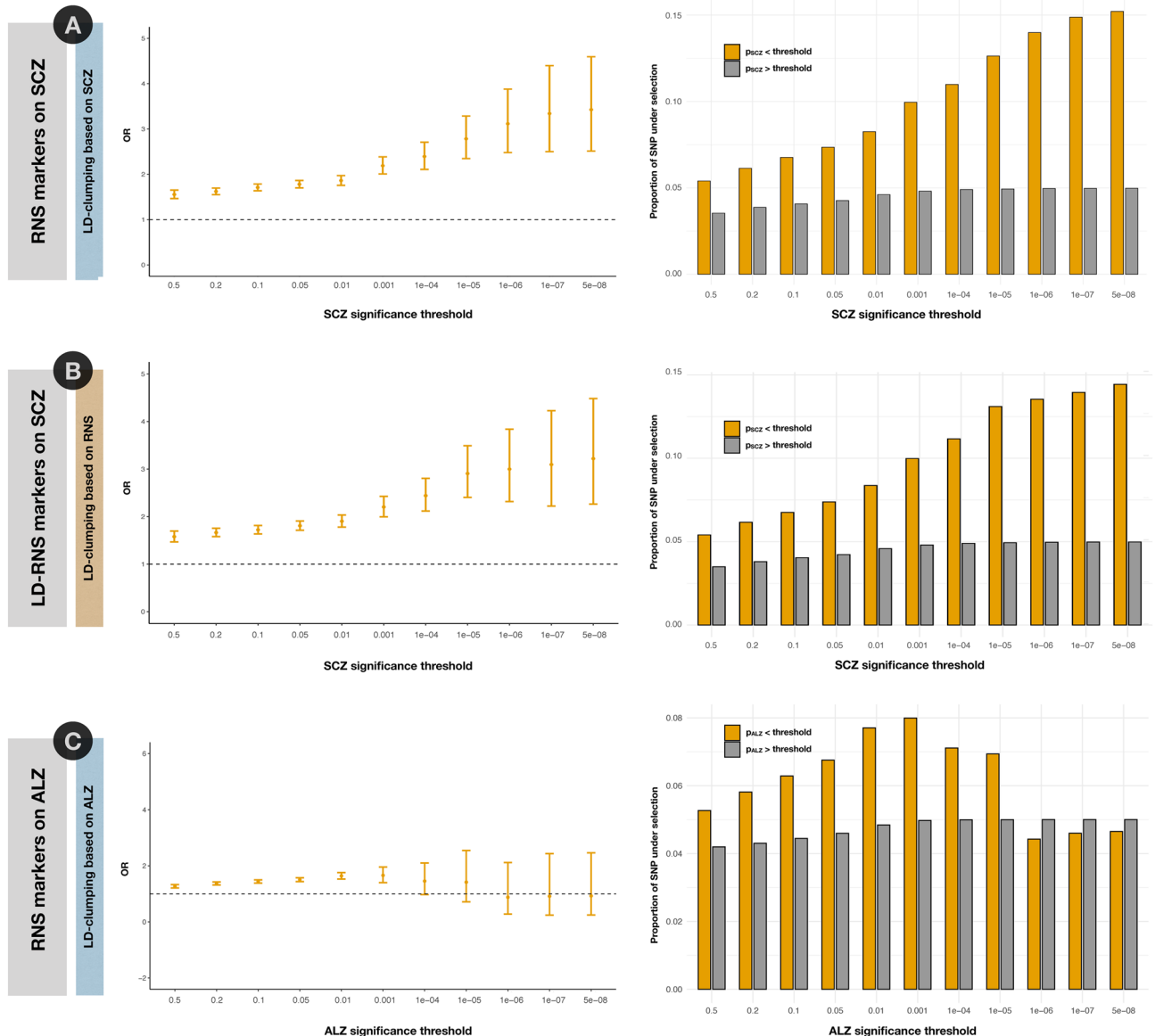


**Figure 1.** Workflow of the analytic pipeline. We used sequencing data from 1000 Genomes Project phase 3 outbred and unrelated subjects ( $N = 1699$  European (EUR), East Asian (EAS), Southern Asian (SAS) and African (AFR) individuals; American individuals were not used due to admixed ancestry) and GWAS summary statistics from schizophrenia<sup>49</sup>. ancestry-specific summary statistics for European and East Asian subjects were used for additional sensitivity analyses. Overlapping variants from both sources ( $5,554,437$  SNPs) were included in the subsequent analyses. Throughout the whole process, markers of RNS were considered as the top 5% LD-independent variants with the highest probability of being markers of natural selection (RNS markers;  $N_{\text{SNPs}} = 8679$ ; LD-clumping based on SCZ). As a sensitivity analysis, we used top 5% LD-independent variants as RNS markers after LD-clumping based on RNS (LD-RNS markers;  $N_{\text{SNPs}} = 6262$ ). Enrichment of RNS was also tested in Alzheimer's disease as a methodological control. Enrichment, effect-stratified analysis (bias towards risk or protection), and biological profiling of RNS in schizophrenia were performed as described in Methods. RNS-stratified polygenic score analyses ( $\text{PGS}_{\text{SCZ}}$ ) were performed in an independent Spanish case-control sample (CIBERSAM;  $N_{\text{SCZ}} = 1927$ ;  $N_{\text{HC}} = 1561$ ) and previous SCZ GWAS<sup>39</sup> as discovery sample. LDSC-SEG = LDSC applied to specifically expressed genes.

African and European populations (Supplementary Data 6). A great consistency was observed between *Fst* and *Pcadapt* selection signals: 98% and 73% of the top 100 selection markers estimated by *Fst* overlapped with the top 5% and 1% of *Pcadapt* RNS markers described (Supplementary Data 6).

**Enrichment of selection signatures across schizophrenia associated loci.** We evaluated the relationship between RNS and schizophrenia predisposing variation from the latest schizophrenia GWAS<sup>49</sup> using a comprehensive set of analyses (Fig. 1).

First, we analyzed the enrichment of selection signals across schizophrenia GWAS significance thresholds. To perform well-powered analyses, we considered the 5% of LD-independent SNPs with the lowest  $p_{RNS}$  as RNS markers ( $N_{SNPs} = 173,701$ ; LD-clumping based on  $p_{SCZ}$ , Supplementary Data 1) and compared them with the remaining LD-independent SNPs. We observed an increasingly significant enrichment for RNS markers ( $N_{SNPs} = 8679$ ) across the most stringent schizophrenia GWAS thresholds ( $p_{SCZ}$  threshold  $< 5 \times 10^{-8}$ , Fisher's exact test  $p = 4.40 \times 10^{-13}$ ; OR (CI 95%) = 3.45 (2.51, 4.59); Fig. 2A; Supplementary Data 3). To ensure that these



**Figure 2.** Enrichments for RNS signals in schizophrenia GWAS. Fisher's exact tests were performed to study the enrichment of the RNS markers (LD-clumping based on  $p_{SCZ}$ ; top 5% of most selected SNPs;  $N_{SNPs} = 8679$ ) and LD-markers (LD-clumping based on  $p_{RNS}$ ; top 5% of most selected SNPs;  $N_{SNPs} = 6262$ ) across schizophrenia GWAS thresholds<sup>49</sup>. (A) OR (95%CI) for the enrichment and proportion of 8679 RNS markers across the different schizophrenia (SCZ) GWAS thresholds. (B) OR (95%CI) for the enrichment and proportion of 6262 LD-markers markers across the different schizophrenia GWAS thresholds. (C) OR (95%CI) for the enrichment and proportion of RNS markers across the different Alzheimer's disease (ALZ)<sup>51</sup> GWAS thresholds. LD-independent SNPs were considered after clumping based on ALZ  $p$ -values. See Supplementary Data 3 for detailed results.

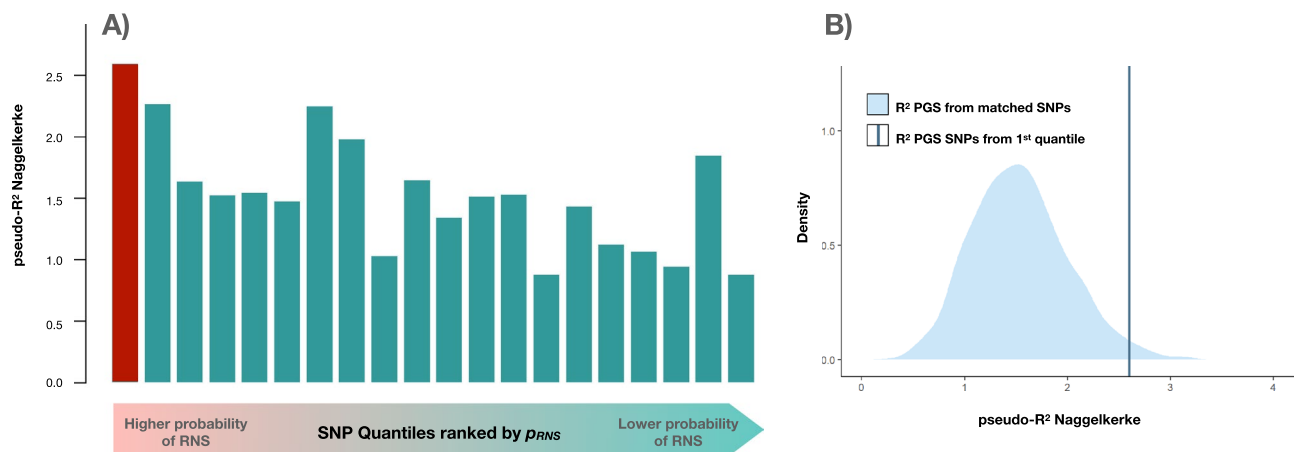
associations were not artefactual due to the LD-clumping based on SCZ  $p$ -values, we selected the 5% of LD-independent SNPs with the lowest  $p_{RNS}$  after LD-clumping based on RNS  $p$ -values (LD-RNS markers;  $N_{SNPs} = 6262$ ; Supplementary Data 1). Again, schizophrenia genome-wide significant (GWS) *loci* have a notably increased probability to overlap with one of the 6262 LD-RNS markers described (Fisher's exact test  $p = 8.38 \times 10^{-10}$ ; OR (CI 95%) = 3.22 (2.26, 4.48); Fig. 2B; Supplementary Data 3). Moreover, to ensure no methodology bias is present, enrichment analyses were repeated using GWAS data from Alzheimer's disease<sup>51</sup>, a related brain disorder<sup>52</sup> that may escape natural selection due to its late onset. LD-clumping was performed based on Alzheimer's GWAS  $p$ -values. No significant enrichment between RNS and Alzheimer's disease was observed (Supplementary Data 3; Fig. 2C) using the same procedure described above for schizophrenia.

We then assessed the partial correlation between schizophrenia and RNS across the set of LD-independent SNPs nominally associated with SCZ, while controlling for their minor allele frequency on the 1000 Genomes Project data ( $p_{SCZ} < 0.05$ ;  $N_{SNPs} = 41,079$ ; LD-clumping based on  $p_{SCZ}$ ) and found a significant correlation (Spearman-rho = 0.034;  $p = 7.03 \times 10^{-12}$ ). This result was confirmed by conducting 10,000 random permutations of the GWAS results with respect to the  $p_{RNS}$  values and testing whether the observed correlation was significantly higher than the null distribution ( $p_{perm} = 1.3 \times 10^{-3}$ ; Supplementary Data 4). A similar correlation pattern between schizophrenia and RNS was observed after LD-clumping based on RNS (Spearman-rho = 0.075;  $p = 5.12 \times 10^{-13}$ ). We repeated this analysis using GWAS data from Alzheimer's disease and observed no significant correlation with RNS (Supplementary Data 4).

Heritability enrichment analyses using LDSC showed a significant SNP-based heritability enrichment ( $h^2_{SNP}$ ) of schizophrenia across the RNS markers ( $h^2_{SNP}$  enrichment (CI 95%) = 1.31 (1.13 – 1.50);  $p = 1.19 \times 10^{-3}$ ; Supplementary Data 5).

We then assessed schizophrenia polygenic score prediction (PGS<sub>SCZ</sub>) in an independent case–control cohort (CIBERSAM;  $N_{SCZ} = 1927$ ;  $N_{HC} = 1561$ ; Supplementary Data 6). For this analysis, we used the PGS<sub>SCZ</sub> threshold with the most significant association with schizophrenia in the analysis using whole genome data ( $p_{SCZ} < 0.2$ ;  $N_{SNPs} = 61,040$ ). After dividing the resultant schizophrenia predisposing variation into 20-quantiles ( $N_{SNPs} = 3052$ ) ranked by  $p_{RNS}$ , we observed a reduction of the explained variance in the case–control cohort when moving to higher  $p_{RNS}$  (linear regression  $t = -3.78$ ;  $p = 0.0013$ ; Fig. 3A). To account for the likely bias of the genomic properties of selection markers, we also compared PGS<sub>SCZ</sub> using variants from the first quantile against PGS<sub>SCZ</sub> using 1000 sets including the same number of matched SNPs selected from the rest of variants. Matched SNPs were selected accounting for MAF, number of SNPs in LD (LD buddies), distance to nearest gene, and gene density (see Methods). PGS<sub>SCZ</sub> prediction from the first quantile was found to be significantly higher than the distribution of 1000 predictions from matched SNP selections ( $p = 0.019$ ; Fig. 3B).

Moreover, to ensure our PGS<sub>SCZ</sub> predictions using variants subject to RNS were not biased by the presence of Long-Range LD regions<sup>53</sup>, we repeated the PGS<sub>SCZ</sub> analysis after excluding Long-Range LD regions previously described<sup>53</sup> ( $p_{SCZ} < 0.2$ ;  $N_{SNPs} = 60,407$ ). Again, a similar reduction of the explained variance towards higher  $p_{RNS}$  was found ( $t = -4.64$ ;  $p = 2 \times 10^{-3}$ ). Also, PGS<sub>SCZ</sub> prediction from the first quantile was found to be significantly higher than the distribution of 1000 predictions from the remaining SNPs ( $p = 0.033$ ; Supplementary Data 6).



**Figure 3.** Schizophrenia Polygenic score predictions (PGS<sub>SCZ</sub>) in an independent cohort stratified by the probability of the SNPs to be subject to RNS. **(A)** PGS<sub>SCZ</sub> predictions (% pseudo-R<sup>2</sup>) on an independent schizophrenia case–control sample (CIBERSAM;  $N_{SCZ} = 1927$ ;  $N_{HC} = 1561$ ) using 20 subsets of SNPs ranked by their probability to be subject to RNS ( $p_{RNS}$ ) are displayed. For this analysis, we used the most significant schizophrenia threshold in the PGS<sub>SCZ</sub> analysis using whole genome data ( $p_{SCZ} < 0.2$ ;  $N_{SNPs} = 61,040$ ). LD-clumping based on  $p_{SCZ}$  ( $r^2 < 0.1$ ; 500 kb) was performed for PGS estimations. **(B)** PGS<sub>SCZ</sub> performance (% pseudo-R<sup>2</sup>) using variants from the first RNS quantile (SNPs with the highest probability to be subject to RNS;  $N_{SNPs} = 3052$ ) was compared against the distribution of PGS<sub>SCZ</sub> Nagelkerke's pseudo-R<sup>2</sup> obtained by selecting the same number of matched SNPs from the remaining 95%. A total of 1000 sets of matched SNPs were randomly drawn from SNPs matched to the variants from the first RNS quantile based on MAF, number of SNPs in LD, distance to nearest gene, and gene density. See Supplementary Data 6 for detailed results.

**Figure 4.** Directionality of RNS enrichment in schizophrenia considering the derived alleles emerged during human evolution. **(A)** Average  $\log_{10}(\text{OR}_{\text{SCZ}})$  of RNS markers (orange) and the remaining SNPs (grey), referred to the derived alleles that emerged during human evolution, across different SCZ GWAS thresholds. 95% confidence intervals of the means are displayed. **(B)** Proportion of SNPs within RNS markers and the remaining SNPs with a protective or risk effect of the derived allele, across different schizophrenia GWAS thresholds. **(C)**  $\text{PGS}_{\text{SCZ}}$  predictions in the case-control target sample ( $N_{\text{SCZ}} = 1927$ ;  $N_{\text{HC}} = 1561$ ) stratified by RNS. Schizophrenia GWAS summary statistics were subsetted based on the relationship of the SNP to RNS and the risk or protective effect of the derived allele as follows: (i) RNS markers (top 5% of the SNPs with the highest probability of being subject to RNS) whose derived alleles arisen through human evolution confer protection to schizophrenia (derived- $\text{OR}_{\text{SCZ}} < 1$ ; light brown), (ii) RNS markers (top 5% of the most selected SNPs) whose derived alleles arisen through human evolution confer risk to schizophrenia (derived- $\text{OR}_{\text{SCZ}} > 1$ ; dark brown), (iii) SNPs not subject to RNS (from the remaining 95% of SNPs) whose derived alleles arisen through human evolution confer protection to schizophrenia (derived- $\text{OR}_{\text{SCZ}} < 1$ ; light grey), (iv) SNPs not subject to RNS (from the remaining 95% of SNPs) whose derived alleles arisen through human evolution confer risk to schizophrenia (derived- $\text{OR}_{\text{SCZ}} > 1$ ; dark grey). Explained variance attributable to PGS was calculated as the increase in Nagelkerke's pseudo- $R^2$  between a linear model with and without the PGS variable.  $P$  values were obtained from the binomial logistic regression of SCZ phenotype on PGS, accounting for LD and including sex, age, and 10 MDS ancestry components as covariates. Significant PGS predictions after FDR correction ( $p_{\text{FDR}} < 0.05$ ) are marked with an asterisk. **(D)** Distribution of variance explained (% pseudo- $R^2$ ) from stratified  $\text{PGS}_{\text{SCZ}}$  predictions based on the abovementioned subsets, using a similar number of variants ( $N_{\text{SNPs}} = 1000$  from all available SNPs ( $P_{\text{SCZ}} < 1$ )) for each analysis. Wilcoxon test was used to compare 1000 pseudo- $R^2$  distributions from PGS predictions. See Supplementary Data 7 for detailed results.

### Enrichment of schizophrenia predisposing variation is also overrepresented in RNS markers within European populations.

Given the above-described relationship between schizophrenia and RNS markers across human populations, we aimed to study whether RNS markers related with local adaptation within European or East Asian populations were also enriched in schizophrenia predisposing *loci* across these populations.

Following a similar methodology, we considered the 5% of LD-independent SNPs with the lowest  $p_{\text{RNS}}$  as RNS markers within European ( $N_{\text{SNPs}} = 139,668$ ;  $N_{\text{samples}} = 130,644$ ; LD-clumping based on  $p_{\text{SCZ}}$ ; Supplementary Data 1C) or East Asian ( $N_{\text{SNPs}} = 179,669$ ;  $N_{\text{samples}} = 30,761$ ; LD-clumping based on  $p_{\text{SCZ}}$ ; Supplementary Data 1D) populations and compared them with the remaining LD-independent SNPs. We observed an increasingly significant enrichment for European-based RNS markers ( $N_{\text{SNPs}} = 6983$ ) across the most stringent schizophrenia GWAS thresholds ( $p_{\text{SCZ}}$  threshold  $< 5 \times 10^{-8}$ , Fisher's exact test  $p = 2.30 \times 10^{-6}$ ; OR (CI 95%) = 2.76 (1.83, 4.04); Supplementary Data 3). In case of the less powered test based on East Asian populations, there is also a trend for enrichment of schizophrenia *loci* across East Asian-based RNS markers ( $N_{\text{SNPs}} = 8983$ ;  $p_{\text{SCZ}}$  threshold  $< 5 \times 10^{-8}$ , Fisher's exact test  $p = 0.08$ ; OR (CI 95%) = 1.65 (0.88, 2.86); Supplementary Data 3).

Moreover, we described a great overlap between European ancestry based and global RNS markers. Across the 46,872 common LD-independent SNPs in both European and all population RNS data, 1176 out of 2249 RNS markers were among the top 5% with highest probability to be RNS marker (Chi-square test  $p < 1 \times 10^{-16}$ ).

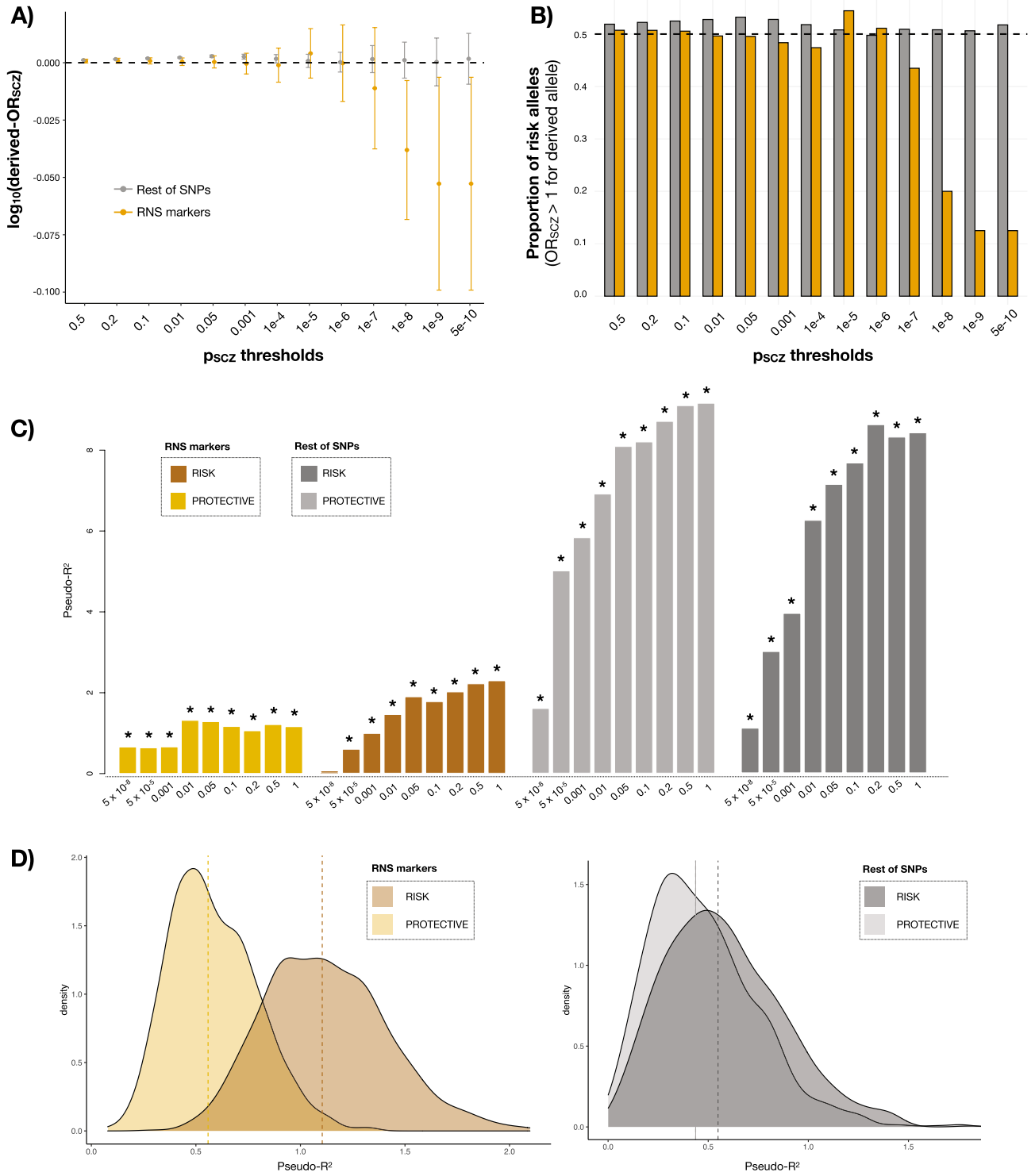
### Alleles arisen during human evolution and subject to RNS are biased towards protection against schizophrenia.

In order to study whether RNS markers confer risk or protection to schizophrenia, we focused on the derived alleles arisen during human evolution (i.e. derived selected alleles). We used LD-independent SNPs with available information of the ancestral allele (the allelic state of the most recent common ancestor of human and the closest primate) in the 1000 Genomes database ( $N_{\text{SNPs}} = 170,609$  (98.2% of total LD-independent SNPs)). We calculated GWAS derived- $\text{OR}_{\text{SCZ}}$  referred to the derived allele and evaluated whether RNS markers associated with schizophrenia were enriched for derived alleles conferring risk or protection to the illness.

We described a significant enrichment of RNS markers for derived alleles with a protective effect on schizophrenia risk at the most stringent GWAS thresholds (average  $\log_{10}(\text{derived-OR}_{\text{SCZ}})$  (CI95%) =  $-0.052$  ( $-0.099$ ,  $-0.006$ );  $p = 0.031$  at  $P_{\text{SCZ}}$  threshold =  $5 \times 10^{-10}$ ;  $N_{\text{protective-SNPs}} = 7$ ;  $N_{\text{risk-SNPs}} = 1$ ; Fig. 4A,B; Supplementary Data 7A). This enrichment for protective alleles was not observed for derived alleles of SNPs not subject to RNS (average  $\log_{10}(\text{derived-OR}_{\text{SCZ}})$  (CI95%) =  $0.001$  ( $-0.009$ ,  $0.013$ );  $p = 0.759$  at  $P_{\text{SCZ}}$  threshold =  $5 \times 10^{-10}$ ;  $N_{\text{protective-SNPs}} = 74$ ;  $N_{\text{risk-SNPs}} = 69$ ; Fig. 4A,B; Supplementary Data 7A).

Nine out of 13 RNS markers from schizophrenia GWS *loci* with derived protective alleles have many other pleiotropic effects such as body size, body mass index or systolic blood pressure (Supplementary Data 8). However, this ratio was not significantly greater than for schizophrenia GWS *loci* with derived risk alleles (4 out of 7 RNS markers; Fisher's exact test  $P = 0.47$ ).

Additionally, as pleiotropy might limit effect size<sup>54,55</sup>, we evaluated  $\text{PGS}_{\text{SCZ}}$  performance in the independent CIBERSAM case-control sample ( $N_{\text{SCZ}} = 1927$ ;  $N_{\text{HC}} = 1561$ ) for RNS markers and the rest of non-selected SNPs, by considering variants with protective or derived risk alleles separately (Fig. 4C, Supplementary Data 7B). Since there were different number of risk ( $N_{\text{SNPs}} = 2967$ ) and protective ( $N_{\text{SNPs}} = 3168$ ) SNPs subject to RNS, we then conducted  $\text{PGS}_{\text{SCZ}}$  comparisons after using random permutations of 1000 SNPs for each subset of SNPs (Fig. 4D, Supplementary Data 7B). Across RNS markers, risk variants explained more variance in the SCZ-HC status than protective variants do (Wilcoxon test;  $p < 10^{-16}$ ; median  $R^2_{\text{RISK}} = 1.10$ ; median  $R^2_{\text{PROTECTIVE}} = 0.54$ ). The difference in explained variance between risk and protective variants was smaller across non-RNS SNPs (Wilcoxon test;



$p = 4.5 \times 10^{-13}$ ; median  $R^2_{\text{RISK}} = 0.54$ ; median  $R^2_{\text{PROTECTIVE}} = 0.46$ ). We observed similar MAF across protective (MAF(CI95%) = 0.366 (0.358, 0.374)) and risk (MAF(CI95%) = 0.336 (0.329, 0.344)) RNS markers, thus ruling out that the results could be inflated by higher MAF of risk SNPs.

**Functional enrichment of schizophrenia GWS *loci* subject to RNS.** We finally aimed to describe the functional characteristics of the *loci* that were both subject to RNS and associated with schizophrenia. We used LDSC applied to specifically expressed genes (LDSC-SEG) to evaluate the schizophrenia heritability enrichments across RNS markers within 10 different tissues, 13 brain related tissues, and 3 brain cell-types and compared them to the LDSC-SEG enrichments across the rest of SNPs (Supplementary Data 4). While schizophrenia predisposing variation non-subject to RNS was found to be enriched at cortical brain tissue (FDR- $p = 0.029$ ) and neuronal cell-type (FDR- $p = 0.017$ ), schizophrenia predisposing variation within RNS markers was found to be enriched beyond brain-related tissues and not enriched in any brain-related cell types (Fig. 5, Supplementary Data 4).

We also characterized RNS signals by comparing genes mapped from schizophrenia GWS *loci* subject and not subject to RNS. Using fine mapping results from the latest SCZ GWAS<sup>49</sup>, we considered the broad set of 628 mapped from GWS *loci* with low numbers of expected causal SNPs ( $K < 3.5$ ; see Methods, Supplementary Data 9A). Genes mapped from GWS *loci* harboring (22 genes) and not harboring (468 genes) RNS markers were tested for functional over-representation analysis (ORA) using the whole gene-set as background. Although no brain-relevant overrepresentation was found in any case, we observed an enrichment of *loci* subject to RNS in brain unrelated functional signatures (miRNA metabolic process  $p_{\text{FDR}} = 0.006$ ; gland development  $p_{\text{FDR}} = 0.026$ ; epidermis development  $p_{\text{FDR}} = 0.026$ ; Supplementary Data 9B).

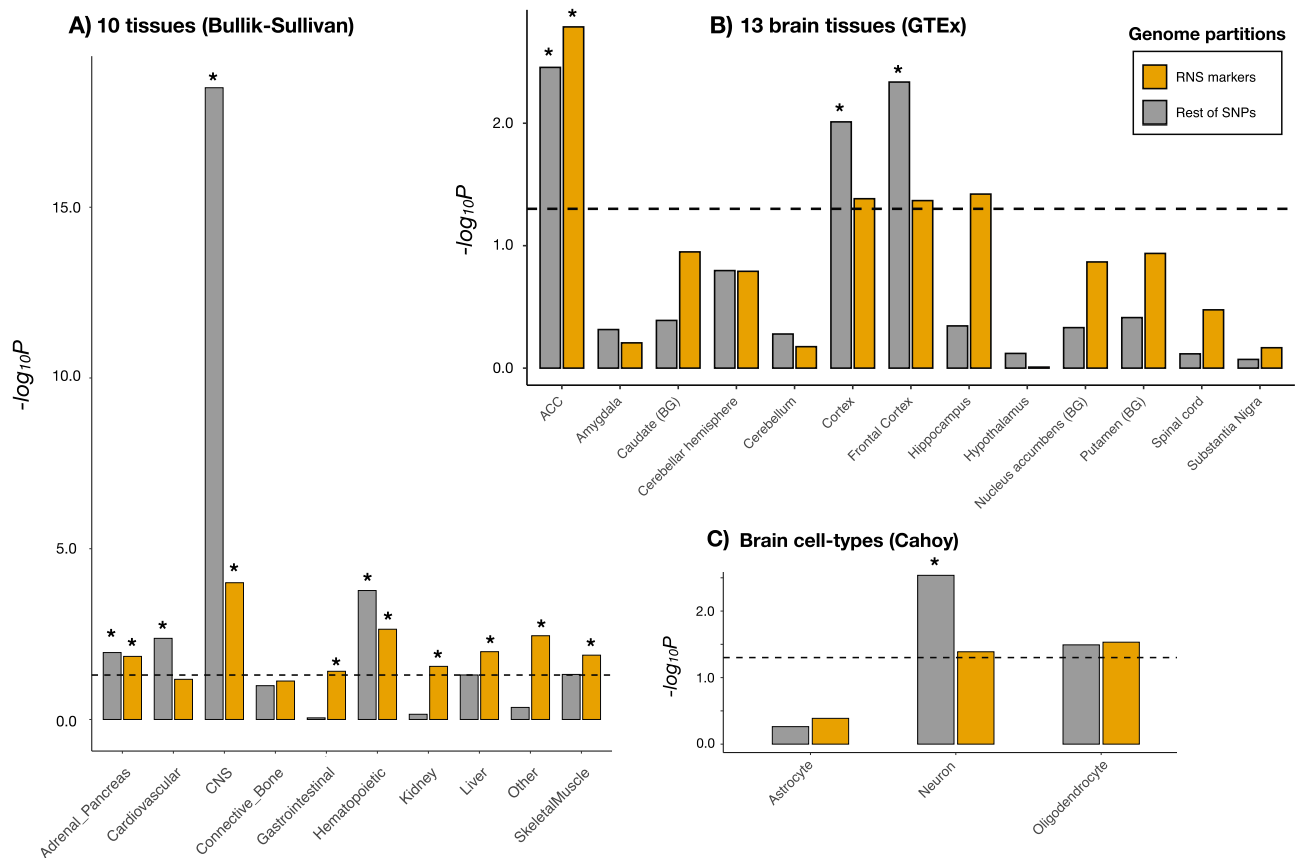
## Discussion

Here in this work, we characterized the RNS signatures that occurred after the human diaspora out of Africa across the whole genome and demonstrated a significant enrichment of RNS markers across schizophrenia predisposing *loci* as well as a significant enrichment in their contribution to schizophrenia heritability. Then, by analyzing the derived and ancestral alleles we observed that RNS markers were enriched for derived alleles conferring protection to schizophrenia. This tendency towards protective versus derived risk alleles was not found across variants not subject to RNS, thus suggesting a positive selection of protective schizophrenia alleles during recent evolution of human populations. The biological characterization of schizophrenia predisposing *loci* subject to RNS and their mapped genes showed an underrepresentation of brain and neuronal related functions as well as more pleiotropic associations relative to schizophrenia predisposing *loci* not subject to RNS. These results suggest that recent selection of schizophrenia alleles could be promoted by conferring selective advantage through other non-psychiatric phenotypes. Our results suggest pleiotropy as a likely mechanism behind RNS that could explain, at least in part, the recent evolutionary dynamics of schizophrenia predisposing genetic variation.

Previous studies have tried to explain why genetic predisposing variation to schizophrenia persists in the population, despite the reproductive fitness reduction in affected individuals<sup>20,21,24</sup>. In this sense, many authors have described the persistence of schizophrenia predisposing alleles as a result of balanced selection or because they could provide fitness advantages in certain environments<sup>22,41,59,60</sup>. However, in this study, we observed that schizophrenia GWS *loci* are enriched in variants that have rapidly expanded across modern human populations (and are therefore subject to RNS) and that the derived alleles confer protection rather than risk to schizophrenia. Thus, our results are in line with recent studies reporting negative selection of schizophrenia predisposing alleles<sup>26,61</sup>. Pardiñas et al.<sup>26</sup> have suggested a plausible explanation of the purported schizophrenia evolutionary paradox of persistent common variation and progressive removal of risk alleles through background selection<sup>62</sup>. By this mechanism, removal of rare haplotypes harboring deleterious mutations could reduce genetic diversity and allow common alleles with small deleterious effects to maintain a high frequency by drift<sup>26,62</sup>.

Apparent contradictory results regarding the positive or negative selection of alleles that predispose to schizophrenia need to be integrated using a dynamic perspective. Since our study has focused only on genetic selection initiated after African populations started their dispersion into Eurasia around 60,000–100,000 years ago<sup>31–33</sup>, a reliable explanation for the persistence of common alleles conferring risk to schizophrenia may reside on fitness advantages provided during earlier stages of the human evolution. In this sense, a previous study focusing on the Sapiens-Neanderthal divergence described a pattern of positive selection of schizophrenia predisposing variation that could have contributed to the *Homo sapiens* success against other human species<sup>41</sup>. Moreover, schizophrenia *loci* were also described to be enriched in genomic regions that had experienced specific evolutionary acceleration during early human evolution (HARs) in comparison with other non-human primates<sup>63</sup>. Although these evidences reinforce the idea of schizophrenia predisposing variation acquisition as a price to pay for human abilities such as language development<sup>24</sup> or even more adaptive immunological profiles<sup>64</sup>, the evolutionary pressures to retain schizophrenia risk alleles as an evolutionary advantage might have stopped acting before the dispersion of the *Homo sapiens* from Africa. In line with our results, studies exploring RNS patterns have not found a pattern of positive selection of schizophrenia predisposing alleles<sup>26,27,29</sup>. A recent study has proposed an evolutionary framework by which schizophrenia risk alleles increased their frequency with the development of the social brain and high-order cognitive functions, but after a turning point the trend was reversed and a negative pressure against schizophrenia alleles became the standard<sup>61</sup>. Interestingly, this turning point from positive to negative selection of predisposing alleles in schizophrenia could be shared by other psychiatric traits. In fact, a recent work observed a similar scenario for attention-deficit/hyperactivity disorder (ADHD): while introgressed Neanderthal alleles were enriched in ADHD risk variants, modern environments triggered the progressive elimination of ADHD-predisposing alleles<sup>65</sup>. This temporal perspective should also prevent readers from reaching conclusions from current evolutionary dynamics. While our approach explored





**Figure 5.** Comparative enrichments of tissue and cell-type signatures across RNS markers and the remaining non-selected SNPs. Tissue and cell-type sub-annotations were created from the intersection between the described annotations (RNS markers and the rest of SNPs) and gene expression data from 10 whole tissues<sup>56</sup>, 13 brain-related tissues (Brain GTEx<sup>57</sup>), and 3 brain cell-type annotation files (neurons, astrocytes, and oligodendrocytes<sup>57,58</sup>). See Methods for further details. **(A)** Heritability enrichment of schizophrenia across RNS markers and the rest of non-selected SNPs within genes expressed in 10 whole tissues<sup>56</sup>. *P* values are derived from one-sided *t*-tests evaluating whether the cell-type enrichment within a particular tissue sub-annotation is higher than the control background annotations. We considered all RNS markers and the rest of the SNPs from schizophrenia GWAS control annotations since they represent the genomic background in which tissue sub-annotations are embedded. **(B)** Heritability enrichment of schizophrenia across RNS markers and the rest of the non-selected SNPs within genes expressed in 13 brain-related tissues (Brain GTEx<sup>57</sup>). To control for brain expression bias, enrichments across brain-tissue sub-annotations were compared to “anti-target tissue” sub-annotations. The latter were created as the intersection between the RNS or the rest of SNPs’ annotations and the remaining 12 brain-related tissue annotations. *P* values are from one-sided *t*-tests evaluating whether the enrichment within a particular brain tissue sub-annotation is higher than the anti-target brain tissue sub-annotation. **(C)** Heritability enrichment of schizophrenia across RNS markers and the rest of the non-selected SNPs within genes expressed in 3 brain cell-type annotation files (neurons, astrocytes, and oligodendrocytes<sup>57,58</sup>). To control for brain expression bias, enrichments across cell-type sub-annotations were compared to “anti-target cell-type” sub-annotations. The latter were created as the intersection between the RNS or the rest of SNPs’ annotations and the remaining 2 cell-type annotations. One-sided *t*-tests were performed to evaluate whether the enrichment within a particular brain cell-type sub-annotation was higher than the anti-target brain cell-type sub-annotation. Dashed lines represent the nominal significance threshold ( $p < 0.05$ ). \* Significant results after Benjamini–Hochberg FDR correction. ACC: Anterior cingulate cortex; BG: Basal ganglia; CNS: central nervous system.

RNS from around 100,000 years ago, recent methods have been used to assess modern selection from up to 100 generations (around 2000 years ago)<sup>66,67</sup>. In these recent studies, schizophrenia was also affected by even more recent negative selection<sup>67</sup>.

Given the evidence here described about the relationship between predisposing variation to schizophrenia and RNS throughout human populations, we assessed whether these evolutionary dynamics were also present at the Asian or European intrapopulation level. While the association between RNS and the genetic predisposition to schizophrenia is replicated in European populations, this association appears to be absent among East Asian populations. This finding suggests that the observed relationship between schizophrenia and recent natural selection may be driven by selective pressures resulting from other traits, acting through specific environments

that are unique to certain populations. However, the considerably smaller sample size of the Asian schizophrenia cohorts compared to the European samples could be the main cause of the lack of association between recent natural selection and schizophrenia predisposing *loci*.

Upon assessing the whole SNP-based heritability enrichment across tissue and cell-type genome annotations, stratified by RNS markers, our results demonstrate a depletion of brain and neuronal-related functions in *loci* subject to RNS, as compared to those not subject to RNS. While common schizophrenia predisposing variation not subject to RNS exhibit clear patterns of heritability enrichment in brain-related tissues and neuronal cell types, aligning with expectations from prior studies, schizophrenia predisposing variation under RNS display lower enrichments in neuronal and brain-related functions, along with significant heritability enrichments in other tissues not directly linked to the central nervous system, such as the kidney or liver (Fig. 5). Similarly, functional enrichment analyses of fine-mapped genes derived from RNS-affected SNPs reveal enrichment in biological functions unrelated to the brain.

In fact, our results pointing to pleiotropic effects of schizophrenia *loci* under selection are in line with previous findings on specific schizophrenia related genes with marked evolutionary patterns. For instance, the GWS schizophrenia risk variant rs13107325 (C/T) within the *SLC39A8* gene has been described to be under positive selection of its derived risk T allele in Europeans<sup>36,39</sup>. This evolutionary event, however, has been reported to be driven by the migration of modern humans out of Africa and the consequent adaptation to colder climates by reducing blood pressure and the risk for hypertension<sup>22,40</sup>. Similarly, survival of the derived risk T allele of variant rs1150711 (C/T) of the *ZNF323* gene could be the result from the ameliorated lung function provided by this allele to European populations<sup>59</sup>. These and similar examples of antagonistic pleiotropy, in which the allele that predisposes to schizophrenia confers adaptive advantage or protects against another condition, have been used to explain the evolutionary paradox of schizophrenia. However, our study suggests that, on a genome-wide scale, non-antagonistic pleiotropy could be contributing to the RNS enrichment across schizophrenia common variation. In this sense, recent positive selection of schizophrenia derived protective alleles could also be a likely by-product of the pleiotropic effect of some genetic variants favoring adaptation to the environment during modern evolution of human populations<sup>68,69</sup>.

Differing genetic risks for schizophrenia described across distinct populations<sup>29</sup> could therefore be the result of distinct evolutionary pressures related to other phenotypes, with predisposing variants having also an impact on schizophrenia risk. Although most of schizophrenia alleles have been described to exert similar effects across populations<sup>70</sup>, allele frequencies could vary across populations, thus leading to differences in individual genetic risk profiles. In this sense, greater polygenic scores for schizophrenia have been described in African populations<sup>29,71</sup>. However, reportedly greater polygenic scores and our finding of positive selection of schizophrenia derived protective alleles do not imply that African populations are at higher genetic risk for schizophrenia. Since GWAS have been performed mainly in cohorts of European ancestry, there is a limited portability for the estimation of polygenic risk scores across African cohorts using the available GWAS data<sup>65,71</sup>, and these inferences should be avoided.

By evaluating the PGS predictions on schizophrenia within RNS with derived protective and risk alleles separately, we observed greater variance explained across similar numbers of risk than protective RNS markers (Fig. 4D). These findings suggest that risk variants under RNS are more likely to increase schizophrenia risk by acting in a cumulative way. It has been suggested that strong negative selection acting against highly pleiotropic deleterious variants of large effect may give rise to common pleiotropic variants of lower effect sizes<sup>54,55</sup>. Consequently, the predictive performance of a set of variants will decrease as the level of pleiotropy increases. If there is an excess of pleiotropic variants among the RNS derived protective alleles compared with derived risk alleles, this might explain the higher predictive capacity of a PGS based on derived risk alleles.

This study was subject to several limitations. First, we described RNS markers using whole 1000 Genomes data, while we used association data from mainly European-ancestry schizophrenia cohorts (around 80%<sup>49</sup>). In this sense, however, our sensitivity analyses considering population specific data lead to similar results in the case of European population, and therefore validating our findings, although this scenario was not replicated in East Asian populations. Therefore, the implications of our results suggesting protective selection by non-antagonistic pleiotropy may be limited, since environments shaping this evolutionary pattern could be absent in other populations. Nevertheless, this inconsistency between European and East Asian population specific analyses may be caused by a great difference in statistical power, and larger Asian cohorts should be used to rule out this possibility. Second, to have sufficient statistical power, our analysis is restricted to variants having a MAF higher than 5%. This filter could have removed many GWAS signals with remarkable RNS effects. For instance, a variant within *SLC39A8*, a schizophrenia-associated gene with well-described evolutionary properties, is absent from this study due to its low allele frequency. Third, along the lines of a previous related study<sup>72</sup>, we have used the top 5% markers with the highest probability of being subject to RNS for our main analyses in order to have enough statistical power. Finally, it is worth noting that while *pcadapt* has been employed in previous studies involving similar 1000 Genomes data<sup>47</sup>, PCA-based methods have been reported to have a susceptibility to false positives when applied to populations with pronounced stratification and a non-continuous principal component space. In this regard, the concurrence of selection outliers detected by alternative *Fst* methods served as validation for our identification of RNS markers using *pcadapt*.

These limitations notwithstanding, our results shed additional light on the relationship between RNS and schizophrenia. By taking advantage of the genome wide scans and statistical outputs provided by novel methods such as *Pcadapt*, we have described a clear enrichment of protective RNS markers across schizophrenia GWS *loci* and suggest non-antagonistic pleiotropy as a likely explanation. This novel perspective could help to integrate previous contradictory findings related to the evolutionary paradigm of schizophrenia and pave the way for further studies of the evolutionary patterns of other neuropsychiatric disorders and human behavioral traits.

## Methods

**RNS scans in human populations.** To evaluate the selective pressures that took place after the human diaspora out of Africa (from ca. 100,000 years ago to present), we performed scans detection of RNS signatures using a principle components-based approach implemented in *Pcadapt* v 4.3.2<sup>47,48</sup>. *Pcadapt* performs PCA and computes  $p$ -values to test for the presence of selection outliers, based on the correlations between genetic variation and the selected  $K$  principal components. Briefly, for a given SNP,  $z$ -scores are obtained by regressing the SNP position on the  $K$  principal components. The test implemented is based on the multi-dimensional Mahalanobis distance from the SNP to the  $K$  components, thus describing how distant the SNP is from the mean.

To study RNS signatures we used genetic data from the 1000 Genomes phase 3 sequencing database<sup>73</sup>, across African, European, and Asiatic populations. American populations, who show greater genetic heterogeneity due to recent admixture events, were eliminated from the analysis, as suggested by the *Pcadapt* developers<sup>47</sup>. Moreover, given the elevated inbreeding coefficients for some of the 10,000 Genomes subpopulations<sup>74</sup>, only 1699 subjects that have been previously described as outbred and unrelated<sup>74</sup> were finally considered. We retained common genetic variation (MAF > 0.05), biallelic, and overlapping with summary data from the schizophrenia GWAS used in this study, thus yielding a total of 5,554,437 SNPs.

Since LD can affect ascertainment of population structure<sup>75</sup>, *Pcadapt* accounts for LD-genome structure (window size = 200 SNP,  $r^2 = 0.1$ ) for the estimation of RNS probabilities for each of the 5,554,437 SNPs considered. The distribution of PC loadings was evaluated to ensure that selection signals correspond to regions subject to adaptation rather than to regions of lower recombination (high LD). We used Cattell's rule to select the appropriate number of PCs ( $K = 3$ ). *Pcadapt* test was then performed and derived  $p$ -values ( $p_{RNS}$ ) for every SNP were calculated to inform about their likelihood of being subject to RNS.

Throughout the subsequent analyses of the study, we consider the top 5% selected markers (5% of SNPs with the lower  $p_{RNS}$ ; RNS markers) after LD-clumping based on schizophrenia GWAS  $p$ -values (PLINK v1.9 *parameters-clump-r2 0.1-clump-kb 500*) and refer to them as RNS markers. We prioritized the association with schizophrenia and the top 5% RNS markers were selected to have enough SNPs with which to perform the subsequent analyses. However, to ensure that the reported results were not biased by LD-clumping based on SCZ  $p$ -values, we also considered the top 5% RNS markers after LD-clumping based on  $p_{RNS}$  and refer to them as LD-RNS markers.

We also used OutFLANK, that provides a robust estimation of the null distribution of a *Fst* test statistic<sup>76</sup>, between African and European 1000 thousand genomes populations, as an additional selection parameter to compare with *Pcadapt* estimates.

**Samples and GWAS summary data.** We used the latest summary statistics of the schizophrenia GWAS (PGC-SCZ3) conducted by the Psychiatric Genomics Consortium (PGC)<sup>49</sup> (data available at <https://www.med.unc.edu/pgc/results-and-downloads>). Main analyses were performed using the whole summary statistics (67,390 schizophrenia/schizoaffective disorder cases and 94,015 controls). Moreover, both ancestry-specific summary statistics for European (53,386 cases and 77,258 controls) and East Asian subjects (14,004 cases and 16,757 controls) were used for additional sensitivity analyses to explore association between RNS signals within European or East Asian populations and ancestry-specific predisposing variation to schizophrenia. As a specificity analysis for the detection of selection signals, we also used Alzheimer's disease<sup>51</sup> GWAS summary statistics in independent analyses. This disorder was selected based on its adequate statistical power comparable to that of schizophrenia, its described clinical similarities, and its later age at onset, after natural selection pressure has exerted its effect. Details regarding the summary statistics used are described in Supplementary data 10.

The Major Histocompatibility Complex (MHC) region was excluded in all analyses and only biallelic SNP and those with an imputation quality score > 0.9 were considered. Overall, 5,554,437 variants from 1000 Genomes overlapping with schizophrenia summary data were converted to PLINK format and used as the input for *Pcadapt* in the analysis. 6,693,073 variants were used in the case of Alzheimer's disease.

We used a case-control sample including 1927 schizophrenia cases (65% males) and 1561 healthy controls (HC) (55% males) from CIBERSAM (Centro de Investigación Biomédica en Red en Salud Mental, Spain) as an independent target sample for PGS predictions (SCZ\_CIBERSAM). All research was performed in accordance with relevant guidelines/regulations. Informed consent was obtained from all individual participants included in the study, and ethic boards/committees from the following involved Spanish hospitals approved the protocol: Clinical Research Ethics Committee of the Hospital Sant Joan de Reus, Research Ethics Committee of Asturias, Research Ethics Committee of Cantabria, Bioethics Commission of the University of Barcelona (CBUB), Galician Regional Research Ethics Committee, Scientific Research Ethics Committee of Hospital Gregorio Marañón and Research Ethics Committee of the Valencia's clinical Hospital. All participants were genotyped as part of the Psychiatric Genomics Consortium (PGC), and passed quality control (QC filters) per PGC-SCZ2 criteria<sup>39</sup>. There is no overlap between PGC-SCZ2 and SCZ\_CIBERSAM samples.

**Polygenic score (PGS) analyses.** PGS for schizophrenia (PGS<sub>SCZ</sub>) were calculated from PGC-SZ2<sup>39</sup> GWAS summary statistics as discovery sample and SCZ\_CIBERSAM case-control cohort as target sample. Summary data from the latest schizophrenia GWAS<sup>49</sup> was not used to estimate PGS because target sample was included as part the cohorts in the study. PGS<sub>SCZ</sub> were calculated using SNPs present in the 1000 genomes database for the studied populations. We included only biallelic variation, with imputation quality scores > 0.9, and excluded indels. 65PLINK 1.9 was used to calculate PGS across schizophrenia patients and healthy controls weighted by the logOR in the discovery sample. Standardized PGS were calculated, and significance was evaluated by logistic regression, using case-control status as dependent variable and sex, age, and the first 10

MDS ancestry components as covariates. We calculated explained variance attributable to PGS as the increase in Nagelkerke's pseudo- $R^2$  between a model with and without PGS variable.

Initially, we used several  $P$  thresholds ( $P < 5 \times 10^{-8}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-3}$ , 0.01, 0.05, 0.1, 0.2, 0.5 and 1) and the whole genome variation to estimate predictions in the SCZ\_CIBERSAM case-control cohort. The most significant  $P$  threshold ( $P < 0.2$ ) was selected for the subsequent PGS analyses stratified by RNS. We then performed stratified PGS predictions to compare SNPs that were subject to RNS to those that were not. Summary SNP data with  $P_{SCZ} < 0.2$  was divided into 20 quantiles of increasing  $p_{RNS}$ . PGS predictions were performed across the 20 SNP quantiles in the CIBERSAM case-control sample. Linear regression was performed to evaluate the change in the variance explained by PGS on case-control status in the CIBERSAM sample across the 20 quantiles. Nagelkerke's pseudo- $R^2$  from PGS predictions using the first quantile, corresponding to the 5% of SNPs with the highest probability of being subject to RNS, was compared against the distribution of pseudo- $R^2$  obtained by selecting the same number of SNPs from the remaining 95% of SNPs. On account of the likely bias in prediction due to the non-random genome features of SNPs subject to RNS, selection of SNPs from the remaining 95% was carried out by identification of sets of randomly drawn SNPs that were matched to the top 5% selected SNP based on allele frequency, number of SNPs in LD, distance to nearest gene, and gene density using default parameters from SNPsnip<sup>77</sup> (<https://data.broadinstitute.org/mpg/snpnsp/>). Since an earlier study by Price et al. suggested that recent selection signals could be partly explained as artifacts caused by long-range LD regions<sup>53</sup> that can lead to inflated variance explanations from variants subject to RNS, we repeat the previously explained stratified PGS predictions after removing 24 described Long-Range LD Regions as a sensitivity analysis to assess whether the results hold.

Additionally, we also calculated stratified PGS after subsetting SNP summary data into those for which the derived allele that emerged during human evolution conferred protection (derived- $OR_{SCZ} < 1$ ) or risk (derived- $OR_{SCZ} > 1$ ) to schizophrenia. Across variants subject (RNS markers) and non-subject to RNS, independently, we compared 1000 predictions from protective and risk variants, using a similar number of SNPs ( $N_{SNPs} = 1000$ ), on the SCZ\_CIBERSAM case-control sample. Wilcoxon test was used to compare the pseudo- $R^2$  distributions of predictions for SNPs with a protective (derived- $OR_{SCZ} < 1$ ) or risk (derived- $OR_{SCZ} > 1$ ) derived allele across SNPs subject (RNS markers) and non-subject to RNS.

**Partitioning heritability (LDSC).** We calculated SNP-based heritability ( $h^2_{SNP}$ ) estimates for the resulting genome partitions after dissections of schizophrenia summary genetic data based on RNS markers following the recommended procedure<sup>15,57</sup> (<https://github.com/bulik/ldsc/wiki/Partitioned-Heritability>).

First, we created per SNP annotation files (one per chromosome and desired annotation). Each file consisted of a row per SNP and a column for each sub-annotation (1 = a particular SNP is part of that sub-annotation, 0 = The SNP does not belong to the annotation). Annotation files were created for:

- i) RNS markers (SNPs from schizophrenia GWAS (PGC-SCZ3) summary statistics<sup>49</sup> that belong to the top 5% with the highest likelihood (lowest  $p_{RNS}$ ) of being a RNS marker) and the rest of the SNPs (the remaining 95% of the SNPs from schizophrenia GWAS summary data).
- ii) Sub-annotations from the intersection between the described annotations (RNS markers and the rest of SNPs) and gene expression data from 10 whole tissues<sup>56</sup>, 13 brain-related tissues (Brain GTEx<sup>57</sup>), and 3 brain cell-type annotation files (neurons, astrocytes, and oligodendrocytes<sup>57,58</sup>).

We used *ldsc v1.0.1*<sup>15,57</sup>, a command line tool for estimating heritability. We performed both heritability enrichment analyses across the described annotations ( $-h2$ ) and one-sided t-tests to evaluate whether the cell-type enrichment in schizophrenia within a particular annotation was higher than the same cell-type enrichment in schizophrenia outside the target annotation (anti-target) but within a background annotation (using  $-h2-cts$ ). Control background annotations from the original study were used<sup>56,57</sup>.

We ran LDSC using associated data files from phase 3 of the 1000 Genomes Project<sup>73</sup>. LD scores were computed for each annotation file using the recommended parameters: 1-cM window ( $-ld-wind-cm$  1), restriction to Hapmap3 SNPs, and exclusion of the MHC region due to its high gene density and exceptional LD, as recommended by the developers<sup>15</sup>. The  $'-overlap-annot'$  argument and 1000 Genomes phase 3—based frequency files ( $'1000G_Phase3_frq'$  files via  $-frqfile-chr$  argument) and LD weights ( $'weights_hm3_no_hla'$  files via  $-w-ld-chr$  argument) were used for LD score calculations.

Partitioned LDSC computes the proportion of SNP heritability associated with each annotation column while considering all the remaining annotations. This is performed by regression models using the estimated LD-scores jointly with other independent LD scores for baseline annotations to improve the model performance. We used the full baseline model v2.2, consistent of a full annotation column (1 per all SNPs) and 158 independent functional annotations, available at LDSC repository (<https://data.broadinstitute.org/alkesgroup/LDSCOPE/>), as independent LD scores.

### Characterization of schizophrenia genome-wide significant (GWS) loci with and without RNS markers.

We used FINEMAP results of the schizophrenia GWS loci from the latest schizophrenia GWAS<sup>49</sup>, and compared the functional overrepresentation of mapped genes from GWS loci harboring RNS markers against those genes not harboring RNS markers. Again, the MHC region was not considered due to its complex LD structure. In order to have sufficient statistical power, the broad fine-map set 628 genes (435 protein-coding) that contained at least one credible SNP from 249 regions with low numbers of expected causal SNPs ( $K < 3.5$ ) were used<sup>49</sup>. Genes mapped only from RNS markers (22 genes) were compared against genes mapped by SNPs not subject to RNS (468 genes).

Functional overrepresentation of genes from each category was assessed by over-representation analysis (ORA) with WEB-based GENE SeT AnaLysis Toolkit (WebGestalt; <http://www.webgestalt.org/>)<sup>78</sup>. All mapped genes from schizophrenia GWS *loci* studied here in this study were used as a background list. Enrichment across gene sets from Gene Ontology (GO) cellular components (CC) and biological functions (BF) categories was evaluated. FDR by Benjamini–Hocheberg adjustment was used to evaluate enrichment significance across GO gene sets, and a minimum threshold of 5 genes overlapping with each gene set was considered.

**Statistical analyses.** Statistical analyses were performed using R (<https://www.r-project.org/>). One-sample Kolmogorov–Smirnov test was performed to evaluate the overrepresentation of SNPs subject to RNS after principal components-based approach implemented in *Pcadapt*. Fisher’s exact tests were performed to study the enrichment for the RNS markers (top 5% RNS *loci*; LD-based on  $p_{SCZ}$ ;  $N_{SNPs} = 8679$ ) or the LD-RNS markers (top 5% RNS *loci*; LD-based on  $p_{RNS}$ ;  $N_{SNPs} = 6262$ ) across the different schizophrenia GWAS thresholds ( $p_{SCZ} < 5 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $1 \times 10^{-6}$ ,  $1 \times 10^{-5}$ ,  $1 \times 10^{-4}$ , 0.001, 0.01, 0.05, 0.1, 0.2, and 0.5). We also analyzed enrichment for the RNS markers for Alzheimer’s disease ( $N_{SNPs} = 6263$ ). Spearman correlation between  $P$  values from GWAS and RNS was performed ( $-\log_{10} p_{GWAS}$ , vs.  $-\log_{10} p_{RNS}$ ). Correlations were confirmed by conducting 10,000 random permutations of the GWAS results with respect to the  $p_{RNS}$  values and testing whether the observed correlation coefficients were significantly higher than the ones in the null distribution of the permuted datasets.

We calculated GWAS derived-OR<sub>SCZ</sub> (referred to the derived allele) for each SNP and evaluated the overrepresentation of protective or derived risk alleles across schizophrenia GWAS thresholds: chi-square tests were performed to compare the proportion of derived risk alleles (derived-OR<sub>SCZ</sub> > 1) within and outside RNS markers (top 5% RNS *loci*) across schizophrenia GWAS thresholds ( $p_{SCZ} < 5 \times 10^{-10}$ ,  $1 \times 10^{-9}$ ,  $1 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $1 \times 10^{-6}$ ,  $1 \times 10^{-5}$ ,  $1 \times 10^{-4}$ , 0.001, 0.01, 0.05, 0.1, 0.2, and 0.5). Additionally, one sample t-tests were performed to compare the average derived-OR<sub>SCZ</sub> within and outside RNS markers (top 5% RNS *loci*) to the neutrality across schizophrenia GWAS thresholds (we consider neutrality, i.e. no risk or protection bias is found when the confidence interval includes the value derived-OR<sub>SCZ</sub> = 1). Two sample t-tests were also performed to compare derived-OR<sub>SCZ</sub> within vs outside RNS markers.

We used Fisher’s exact tests to evaluate the enrichment of schizophrenia GWS *loci* subject to RNS with the derived protective or risk allele having the lead association with schizophrenia among all phenotypes included in GWAS Atlas (<https://atlas.ctglab.nl/>)<sup>79</sup>.

## Data availability

The results generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 November 2022; Accepted: 12 September 2023

Published online: 19 September 2023

## References

- Weinberger, D. R. Future of days past: Neurodevelopment and schizophrenia. *Schizophr. Bull.* **43**, 1164–1168 (2017).
- Hannon, E. *et al.* Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* **19**, 48–54 (2016).
- Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
- Saha, S., Chant, D., Welham, J. & McGrath, J. A systematic review of the prevalence of schizophrenia. *PLoS Med.* **2**, e141 (2005).
- Charlson, F. J. *et al.* Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophr. Bull.* **44**, 1195–1203 (2018).
- Jääskeläinen, E. *et al.* A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr. Bull.* **39**, 1296–1306 (2013).
- Laursen, T. M., Nordentoft, M. & Mortensen, P. B. Excess early mortality in schizophrenia. *Annu. Rev. Clin. Psychol.* **10**, 425–448 (2014).
- Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. *The Lancet* **373**, 234–239 (2009).
- Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* **204**, 108–114 (2014).
- Rees, E. *et al.* De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat. Neurosci.* **23**, 179–184 (2020).
- Singh, T., Neale, B. M., Daly, M. J. & Consortium, on behalf of the S. E. M.-A. (SCHEMA). *Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia*. 2020.09.18.20192815 (2020). <https://doi.org/10.1101/2020.09.18.20192815>.
- Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
- Gandal, M. J., Leppa, V., Won, H., Parikshak, N. N. & Geschwind, D. H. The road to precision psychiatry: Translating genetics into disease mechanisms. *Nat. Neurosci.* **19**, 1397–1407 (2016).
- The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Ramos, A., Gonzalez, L. A. N., Weinberger, D. R., Mitchell, K. J. & Nicodemus, K. K. The role of polygenic risk score gene-set analysis in the context of the omnigenic model of schizophrenia. *Neuropsychopharmacology* **44**, 1562–1569 (2019).
- Lee, P. H. *et al.* Partitioning heritability analysis reveals a shared genetic basis of brain anatomy and schizophrenia. *Mol. Psychiatry* **21**, 1680–1689 (2016).
- Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
- Power, R. A. *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse versus their unaffected siblings. *JAMA Psychiat.* **70**, 22–30 (2013).
- Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* **1**, 40–47 (2000).

21. Crespi, B., Summers, K. & Dorus, S. Adaptive evolution of genes underlying schizophrenia. *Proc. R. Soc. B Biol. Sci.* **274**, 2801–2810 (2007).
22. Li, M. *et al.* Recent positive selection drives the expansion of a schizophrenia risk nonsynonymous variant at SLC39A8 in Europeans. *Schizophr. Bull.* **42**, 178–190 (2016).
23. Power, R. A. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat. Neurosci.* **18**, 953–955 (2015).
24. Crow, T. J. Is schizophrenia the price that Homo sapiens pays for language?. *Schizophr. Res.* **28**, 127–141 (1997).
25. Huxley, J., Mayr, E., Osmond, H. & Hoffer, A. Schizophrenia as a genetic morphism. *Nature* **204**, 220–221 (1964).
26. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
27. Yao, Y. *et al.* No evidence for widespread positive selection signatures in common risk alleles associated with schizophrenia. *Schizophr. Bull.* **46**, 603–611 (2020).
28. Muntané, G. *et al.* The shared genetic architecture of schizophrenia, bipolar disorder and lifespan. *Hum. Genet.* **140**, 441–455 (2021).
29. Guo, J. *et al.* Global genetic differentiation of complex traits shaped by natural selection in humans. *Nat. Commun.* **9**, 1865 (2018).
30. Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nat. Genet.* **40**, 340–345 (2008).
31. Slatkin, M. & Racimo, F. Ancient DNA and human history. *Proc. Natl. Acad. Sci.* **113**, 6380–6387 (2016).
32. Haber, M. *et al.* a rare deep-rooting D0 African Y-chromosomal haplogroup and its implications for the expansion of modern humans out of Africa. *Genetics* **212**, 1421–1428 (2019).
33. Montinaro, F., Pankratov, V., Yelmen, B., Pagani, L. & Mondal, M. Revisiting the out of Africa event with a deep-learning approach. *Am. J. Hum. Genet.* **108**, 2037–2051 (2021).
34. Lam, M. *et al.* Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic pathways. *Am. J. Hum. Genet.* **105**, 334–350 (2019).
35. Ohi, K., Shimada, T., Yasuyama, T., Uehara, T. & Kawasaki, Y. Variability of 128 schizophrenia-associated gene variants across distinct ethnic populations. *Transl. Psychiatry* **7**, e988–e988 (2017).
36. Carrera, N. *et al.* Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. *Biol. Psychiatry* **71**, 169–177 (2012).
37. Zuber, V. *et al.* Identification of shared genetic variants between schizophrenia and lung cancer. *Sci. Rep.* **8**, 674 (2018).
38. Liu, H. *et al.* Integrated analysis of summary statistics to identify pleiotropic genes and pathways for the comorbidity of schizophrenia and cardiometabolic disease. *Front. Psychiatry* **11**, 256 (2020).
39. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
40. Costas, J. The highly pleiotropic gene SLC39A8 as an opportunity to gain insight into the molecular pathogenesis of schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **177**, 274–283 (2018).
41. Srinivasan, S. *et al.* Genetic markers of human evolution are enriched in schizophrenia. *Biol. Psychiatry* **80**, 284–292 (2016).
42. Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).
43. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
44. François, O., Martins, H., Caye, K. & Schoville, S. D. Controlling false discoveries in genome scans for selection. *Mol. Ecol.* **25**, 454–469 (2016).
45. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
46. Yang, W.-Y., Novembre, J., Eskin, E. & Halperin, E. A model-based approach for analysis of spatial structure in genetic data. *Nat. Genet.* **44**, 725–731 (2012).
47. Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E. & Blum, M. G. B. Detecting genomic signatures of natural selection with principal component analysis: Application to the 1000 genomes data. *Mol. Biol. Evol.* **33**, 1082–1093 (2016).
48. Privé, F., Luu, K., Vilhjálmsdóttir, B. J. & Blum, M. G. B. Performing highly efficient genome scans for local adaptation with R Package pcadapt Version 4. *Mol. Biol. Evol.* **37**, 2153–2154 (2020).
49. Trubetskov, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
50. Meisner, J., Albrechtsen, A. & Hanghøj, K. Detecting selection in low-coverage high-throughput sequencing data using principal component analysis. *BMC Bioinform.* **22**, 470 (2021).
51. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **51**, 404–413 (2019).
52. DeMichele-Sweet, M. A. A. *et al.* Genetic risk for schizophrenia and psychosis in Alzheimer disease. *Mol. Psychiatry* **23**, 963–972 (2018).
53. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
54. Shikov, A. E., Skitchenko, R. K., Predeus, A. V. & Barbitoff, Y. A. Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits. *Sci. Rep.* **10**, 1037 (2020).
55. Novo, I., López-Cortegano, E. & Caballero, A. Highly pleiotropic variants of human traits are enriched in genomic regions with strong background selection. *Hum. Genet.* **140**, 1343–1351 (2021).
56. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
57. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
58. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
59. Luo, X.-J. *et al.* Systematic integration of brain eQTL and GWAS identifies ZNF323 as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function. *Schizophr. Bull.* **41**, 1294–1308 (2015).
60. Nettle, D. & Clegg, H. Schizotypy, creativity and mating success in humans. *Proc. R. Soc. B Biol. Sci.* **273**, 611–615 (2006).
61. Liu, C., Everall, I., Pantelis, C. & Bousman, C. Interrogating the evolutionary paradox of schizophrenia: A novel framework and evidence supporting recent negative selection of schizophrenia risk alleles. *Front. Genet.* **10**, 389 (2019).
62. Charlesworth, B. The effects of deleterious mutations on evolution at linked sites. *Genetics* **190**, 5–22 (2012).
63. Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P. & Dudley, J. T. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol. Biol. Evol.* **32**, 1148–1160 (2015).
64. Carter, M. & Watts, C. A. H. Possible biological advantages among schizophrenics’ relatives. *Br. J. Psychiatry* **118**, 453–460 (1971).
65. Esteller-Cucala, P. *et al.* Genomic analysis of the natural history of attention-deficit/hyperactivity disorder using Neanderthal and ancient Homo sapiens samples. *Sci. Rep.* **10**, 8622 (2020).
66. Song, W. *et al.* A selection pressure landscape for 870 human polygenic traits. *Nat. Hum. Behav.* **5**, 1731–1743 (2021).
67. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760–764 (2016).
68. Amato, R., Pinelli, M., Monticelli, A., Miele, G. & Coccozza, S. Schizophrenia and vitamin D related genes could have been subject to latitude-driven adaptation. *BMC Evol. Biol.* **10**, 351 (2010).

69. Li, L. *et al.* Recent positive selection drives the expansion of a schizophrenia-associated variant within 10q24.33 in human populations through its pleiotropic effects on diverse human complex traits. *J. Psychiatry Brain Sci.* **2**, (2017).
70. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
71. Curtis, D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatry Genet.* **28**, 85–89 (2018).
72. Polimanti, R. & Gelernter, J. Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genet.* **13**, e1006618 (2017).
73. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
74. Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E. & Leutenegger, A.-L. High level of inbreeding in final phase of 1000 Genomes Project. *Sci. Rep.* **5**, 17453 (2015).
75. Abdellaoui, A. *et al.* Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21**, 1277–1285 (2013).
76. Whitlock, M. C. & Lotterhos, K. E. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *Am. Nat.* **186**, S24–S36 (2015).
77. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
78. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
79. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).

## Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation. Instituto de Salud Carlos III (SAM16PE07CP1, PI16/02012, PI17/00997, PI19/01024, PI20/00721), co-financed by ERDF Funds from the European Commission, “A way of making Europe”, CIBERSAM. Madrid Regional Government (B2017/BMD-3740 AGES-CM-2), European Union Structural Funds. European Union Seventh Framework Program under grant agreements FP7-4-HEALTH-2009-2.2.1-2-241909 (Project EU-GEI), FP7- HEALTH-2013-2.2.1-2-603196 (Project PSYSCAN) and FP7- HEALTH-2013-2.2.1-2-602478 (Project METSY); and European Union H2020 Program under the Innovative Medicines Initiative 2 Joint Undertaking (grant agreement No 115916, Project PRISM, and grant agreement No 777394, Project AIMS-2-TRIALS), Fundación Familia Alonso, Fundación Alicia Koplowitz and Fundación Mutua Madrileña. J González-Peñas holds a Sara Borrel from Instituto de Salud Carlos III (CD20/00118). C. M. Díaz-Caneja holds a Juan Rodés Grant from Instituto de Salud Carlos III (JR19/00024).

## Author contributions

J.G.-P. and J.C. designed research. J.C. supervised the work. J.G.P. performed all the analyses with contribution from L.H. J.G.P. wrote the manuscript with contribution from C.M.D.C. The rest of authors collaborated with the manuscript preparation and genetic data acquisition.

## Competing interests

Dr. Arango has been a consultant to or has received honoraria or grants from Acadia, Angelini, Gedeon Richter, Janssen Cilag, Lundbeck, Minerva, Otsuka, Roche, Sage, Servier, Shire, Schering Plough, Sumitomo Dainippon Pharma, Sunovion and Takeda. Dr. Crespo-Facorro has received honoraria (advisory board and educational lectures) and travel expenses from Takeda, Menarini, Angelini, Teva, Otsuka, Lundbeck and Johnson & Johnson. He has also received unrestricted research grants from Lundbeck. Dr. Díaz-Caneja has received honoraria from AbbVie, Sanofi, and Exeltis. The rest of the authors do not report any conflicts of interest related to this work.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-42578-0>.

**Correspondence** and requests for materials should be addressed to J.G.-P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023