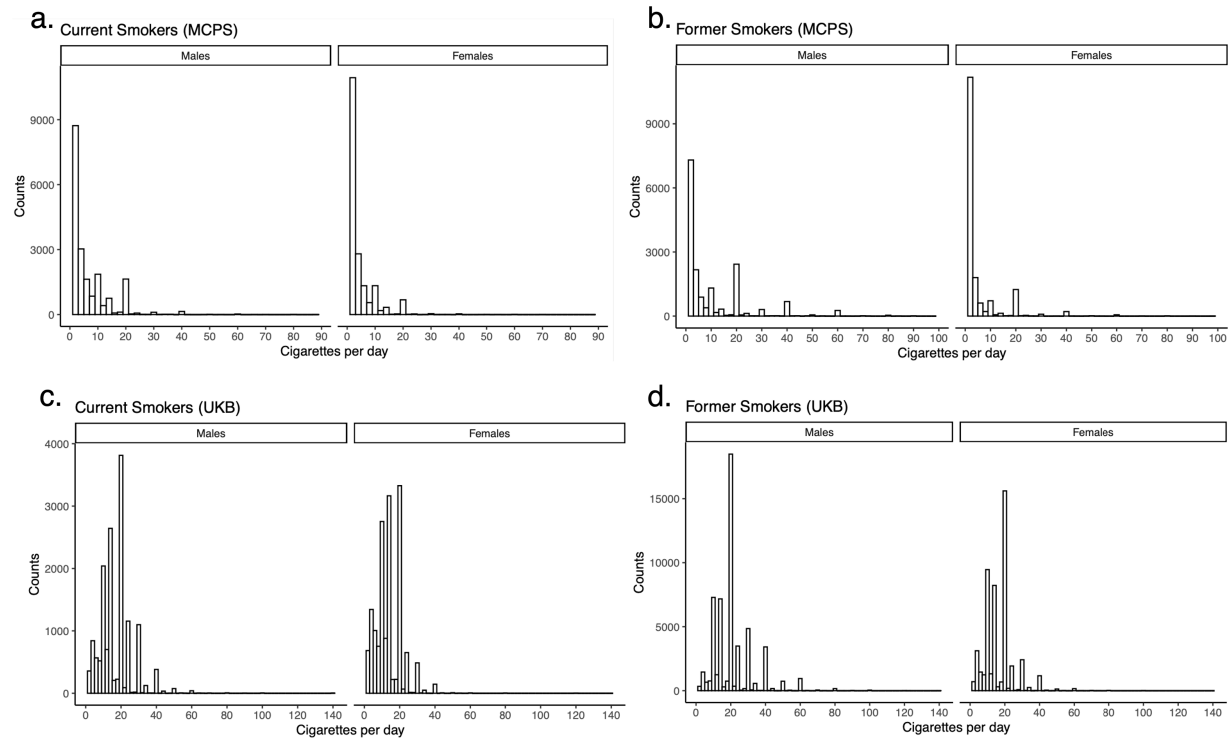


Supplementary Figures

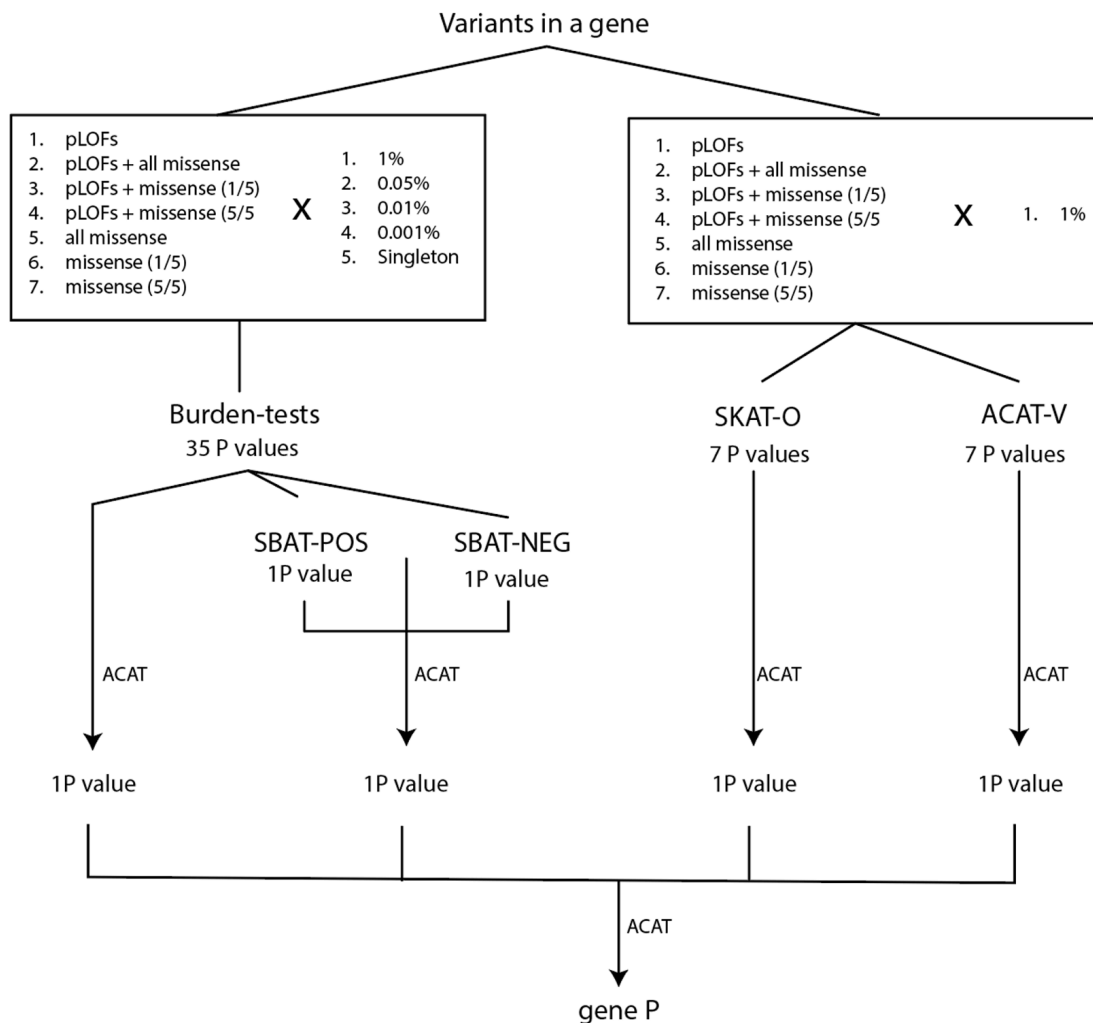
Table of Contents

Supplementary Figure 1. Distribution of cigarettes per day in the MCPS and UK Biobank cohorts.	2
Supplementary Figure 2. Gene P calculation workflow.....	3
Supplementary Figure 3. Quantile-quantile plots of exome-wide gene-level associations.	4
Supplementary Figure 4. Quantile-quantile plots of exome-wide single variant associations.....	5
Supplementary Figure 5. Gene-based association summary for CHRNA3.	6
Supplementary Figure 6. CHRNA3 rare variant association forest plot in MCPS.....	7
Supplementary Figure 7. CHRNA3 association with and without p.Glu284Gly.	8
Supplementary Figure 8. CHRNA3 associations in related versus unrelated participants.	8
Supplementary Figure 9. CHRNA3 associations in current versus former smokers.....	9
Supplementary Figure 10. CHRNA3 rare variant associations in UKB.	10
Supplementary Figure 11. CHRNA3 pLOF frequencies across ancestries.	11
Supplementary Figure 12. Phenome-wide association of rs147306385.....	12
Supplementary Figure 13. CHRNA3 locus common variant associations.	13
Supplementary Figure 14. CHRNA3 associations with binary smoking phenotypes.	14
Supplementary Figure 15. VSIR single variant association QQ plot.	15
Supplementary Figure 16. VSIR Ser280Leu association comparison.	16
Supplementary Figure 17. VSIR gene-based associations.	17
Supplementary Figure 18. DBH and FAM163B tissue expression.....	18
Supplementary Figure 19. GPR101 tissue expression.....	19
Supplementary Figure 20. CHRNA5 locus ancestry analysis.....	20
Supplementary Figure 21. CHRNA3 locus ancestry analysis.....	22



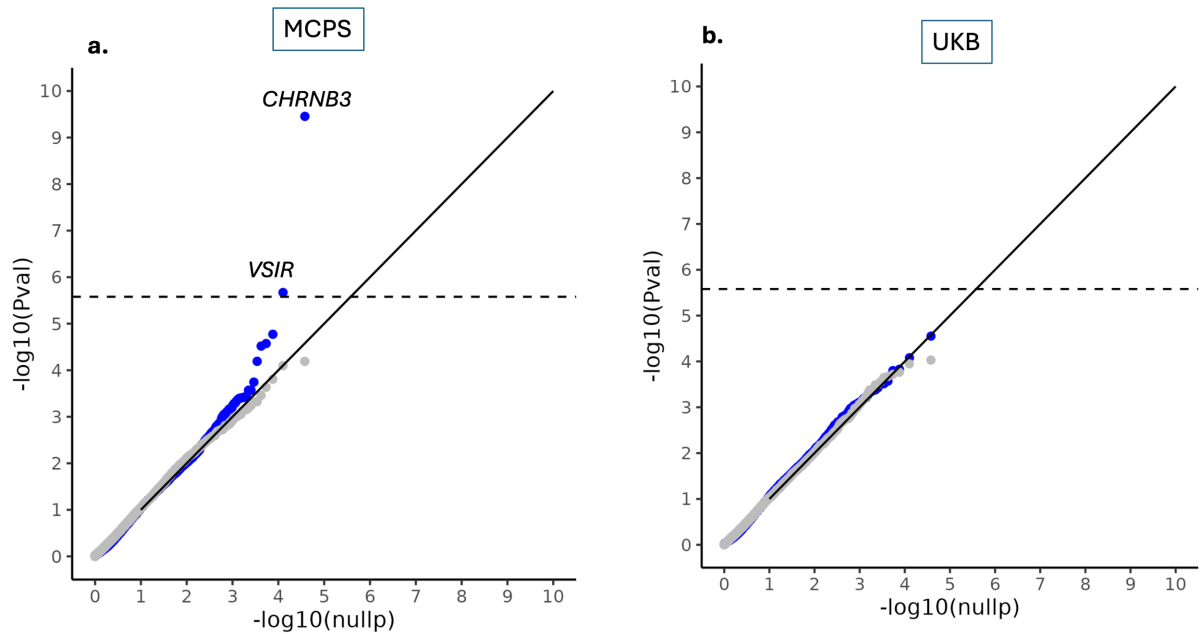
Supplementary Figure 1. Distribution of cigarettes per day in the MCPS and UK Biobank cohorts.

Histograms showing cigarettes per day distribution stratified by sex in current smokers (MCPS: a; UKB: c) and former smokers (MCPS: b; UKB: d). Y-axis represents count of participants. Sample sizes: MCPS current smokers n=37,897 (19,550 males, 18,347 females), MCPS former smokers n=33,164 (16,723 males, 16,441 females), UKB current smokers n=30,812 (14,994 males, 15,818 females), UKB former smokers n=102,321 (53,693 males, 48,628 females). Data collection: self-reported via interviewer-administered questionnaire (MCPS) or touchscreen/online questionnaire (UKB).



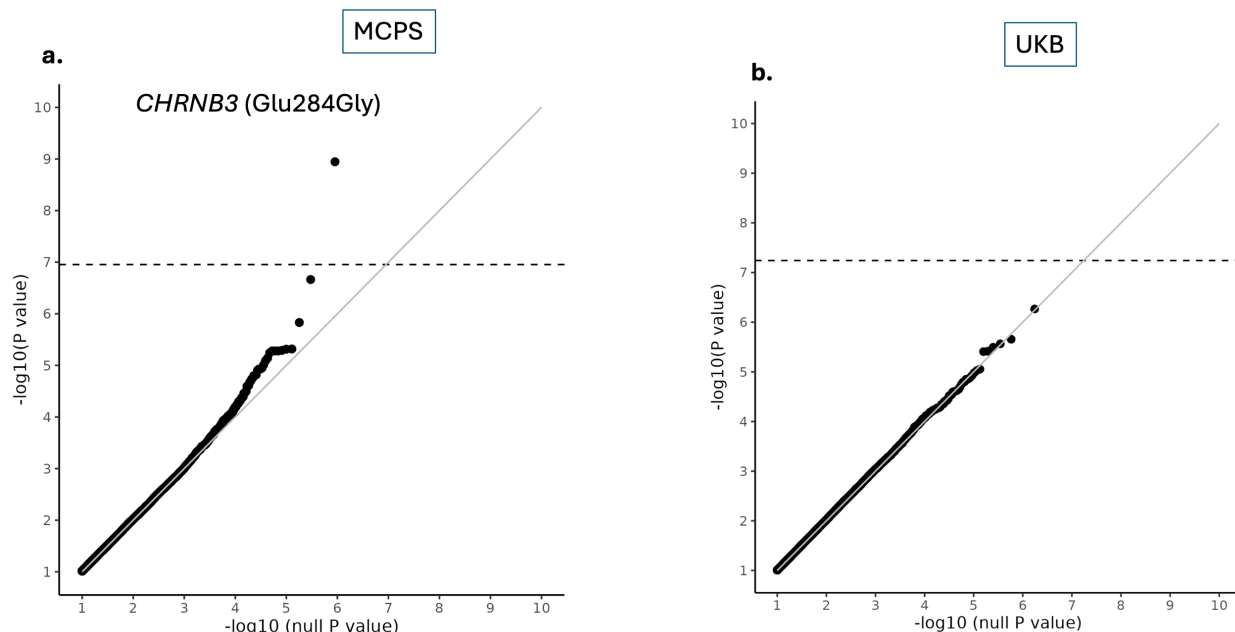
Supplementary Figure 2. Gene P calculation workflow.

The flow chart displays the schematic of gene P calculation. Aggregate rare variant tests were performed using various groupings and statistical approaches. Classic burden tests were performed by collapsing variants using 35 definitions (combinations of seven variant classes and five minor allele frequency thresholds) from which summary P values are calculated using Aggregated Cauchy Association Test (ACAT) and Sparse Burden Association Test (SBAT). Aggregate rare variant tests using Sequence Kernel Association Test (SKAT-O) and variant-level ACAT (ACAT-V) each yield 7 tests corresponding to 7 variant classes at 1% MAF threshold. The four summary P values are combined to produce a single gene P value. Statistical approach described in detail in Ziyatdinov et al. AJHG 2024 (ref.¹).



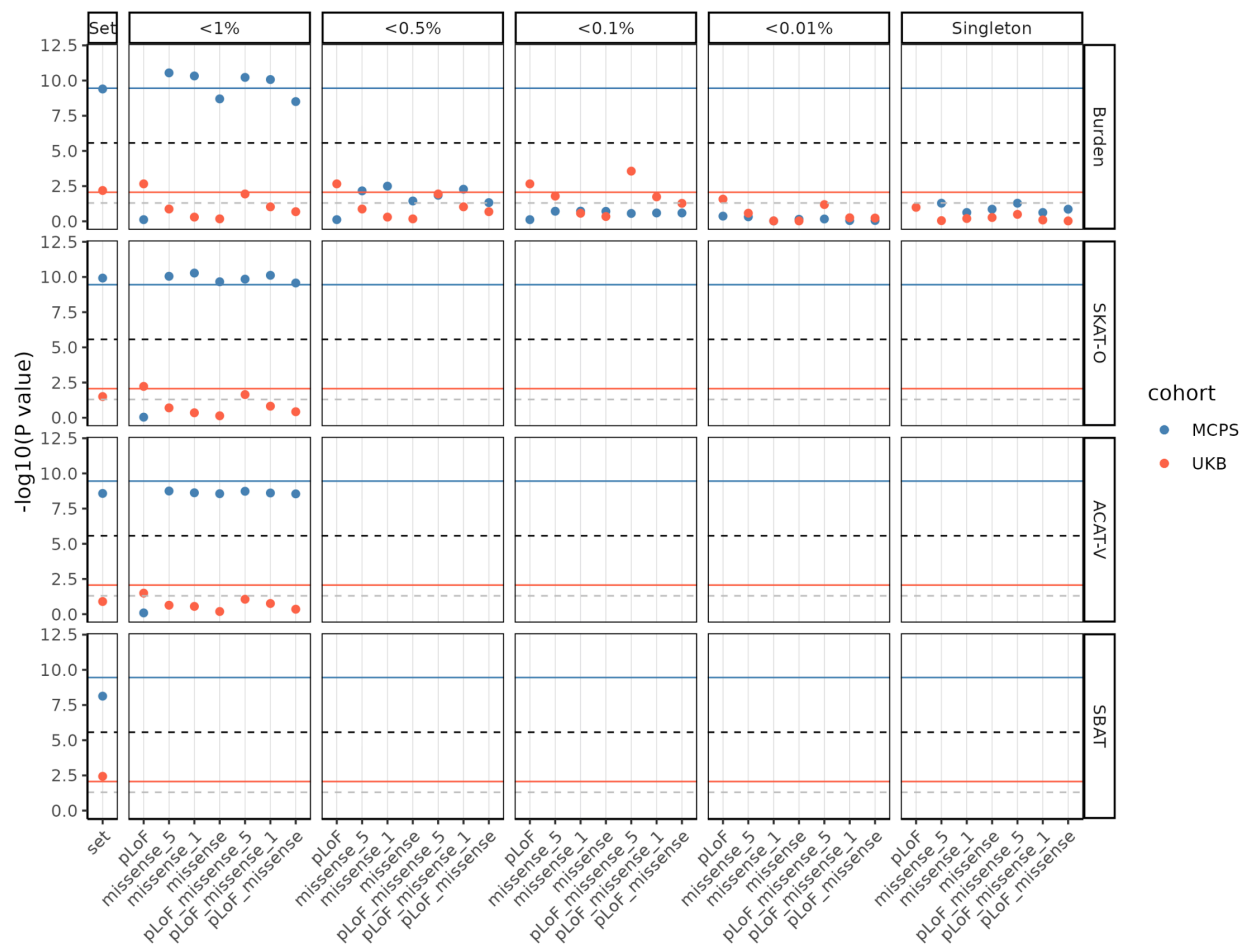
Supplementary Figure 3. Quantile-quantile plots of exome-wide gene-level associations.

Quantile-quantile plots of exome-wide gene-level rare variant associations with cigarettes per day in MCPS (a) and UKB (b). X-axes display expected P values under null distribution; Y-axes display observed P values (both in negative log10 scale). Dotted line corresponds to exome-wide significance threshold $P=2.6 \times 10^{-6}$, accounting for $n=18,955$ unique genes tested. Blue points represent gene tests based on rare missense and pLOFs; grey points represent synonymous variant controls to assess potential test statistic inflation. Statistical test: gene-level burden tests implemented in REGENIE².



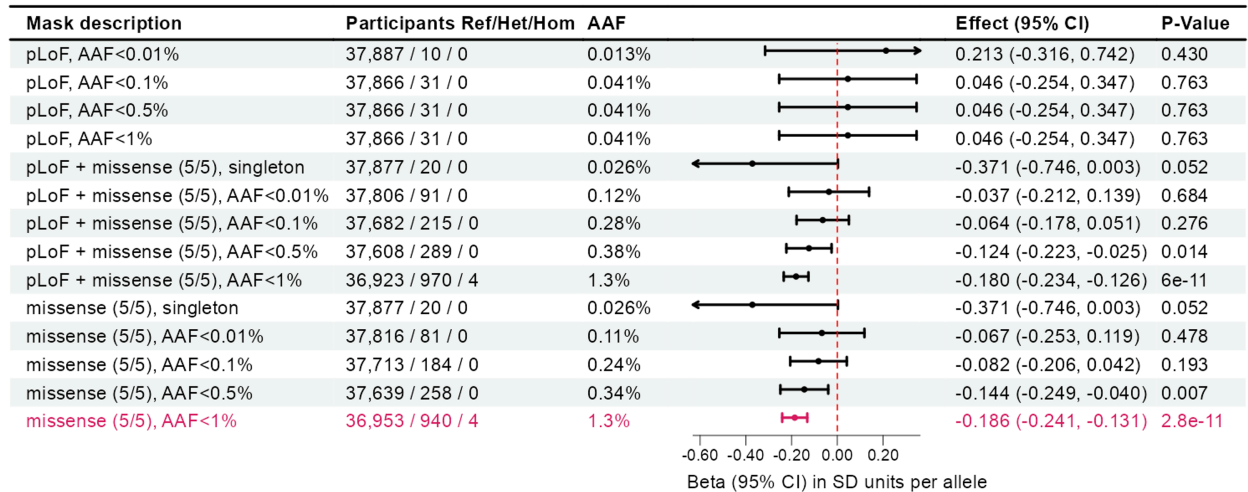
Supplementary Figure 4. Quantile-quantile plots of exome-wide single variant associations.

Quantile-quantile plots of exome-wide single rare variant associations with cigarettes per day in MCPS (a) and UKB (b). X-axes display expected P values under null distribution, Y-axes display observed P values (both in negative log₁₀ scale). Dotted lines correspond to exome-wide significance thresholds: $P=1.1\text{e-}7$ (MCPS, $n=449,606$ variants tested) and $P=5.7\text{e-}8$ (UKB, $n=871,359$ variants tested). *CHRNA3* p.Glu284Gly is highlighted in panel a. Statistical test: single variant association tests implemented in REGENIE².



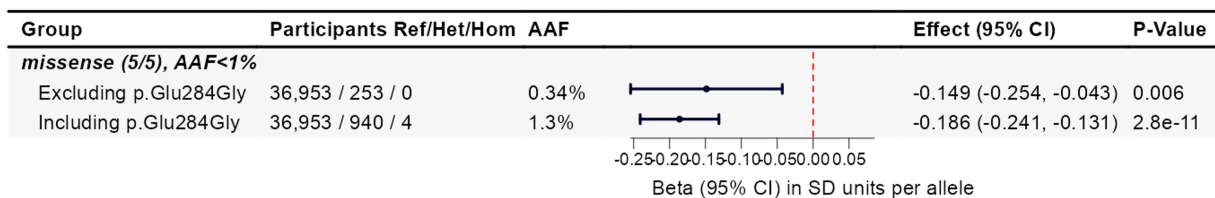
Supplementary Figure 5. Gene-based association summary for CHRN3.

Gene-based association results for CHRN3 with cigarettes per day in MCPS (blue) and UKB (red). X-axis corresponds to different variant aggregation masks (refer to Supplementary Table 2). Y-axis displays P values in negative log10 scale, grouped by statistical test type (burden tests across different MAF thresholds). Horizontal dotted line indicates $P=0.05$. Statistical tests: burden, SKAT-O, and ACAT-V as described in Supplementary Figure 2 and Ziyatdinov et al. AJHG 2024 (ref.¹).



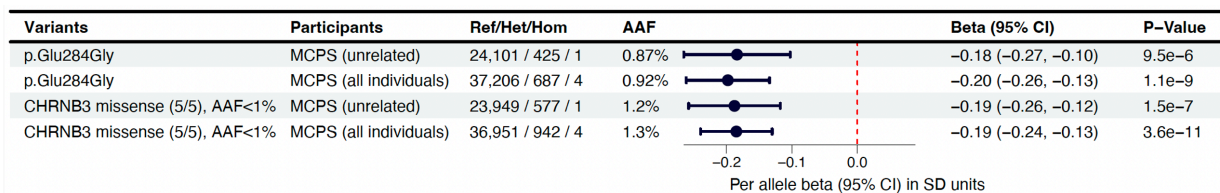
Supplementary Figure 6. CHRNA3 rare variant association forest plot in MCPS.

Forest plot summarizing aggregate rare variant associations of CHRNA3 with cigarettes per day in MCPS. Ref/Het/Hom counts refer to homozygous reference, heterozygous, and homozygous variant carriers respectively. Effect sizes in standard deviation units with 95% confidence intervals. Most significant burden test (missense 5/5, AAF<1%) highlighted in pink. n=37,897 participants analyzed. Statistical test: gene-level burden tests implemented in REGENIE².



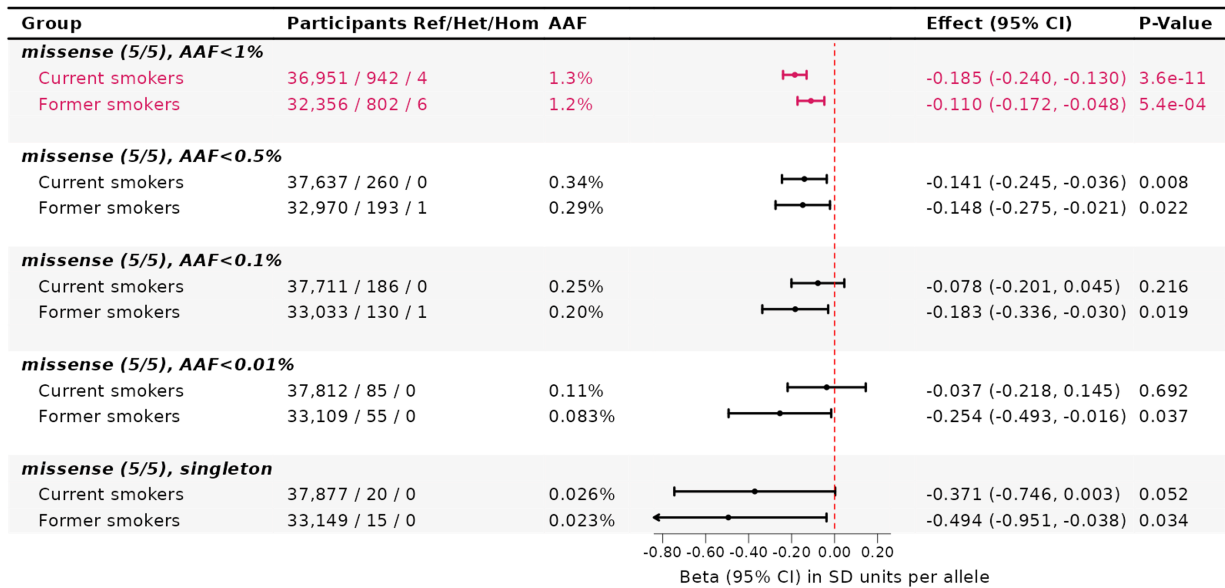
Supplementary Figure 7. CHRNA3 association with and without p.Glu284Gly.

Forest plot comparing aggregate rare deleterious missense variant associations of CHRNA3 with cigarettes per day in MCPS, excluding and including p.Glu284Gly. Effect sizes in standard deviation units with 95% confidence intervals. Excluding p.Glu284Gly: n=36,953 reference, n=253 carriers. Including p.Glu284Gly: n=36,953 reference, n=940 carriers. Statistical test: gene-based burden tests implemented in REGENIE².



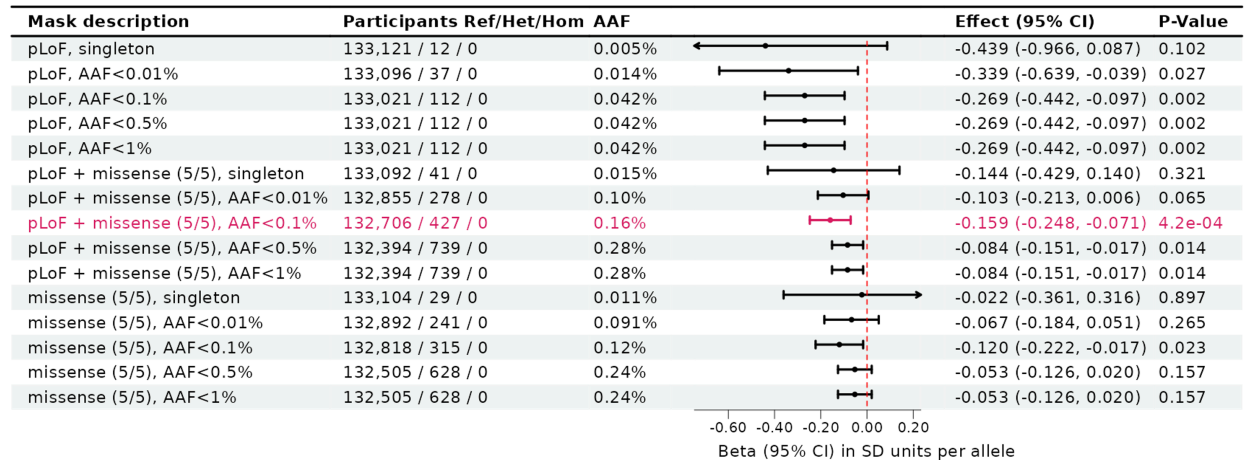
Supplementary Figure 8. CHRNA3 associations in related versus unrelated participants.

Forest plot comparing CHRNA3 rare variant associations with cigarettes per day between unrelated participants and all participants (including up to 3rd degree relatives) in MCPS. Effect sizes in standard deviation units with 95% confidence intervals. Sample sizes: unrelated participants n=24,527, all participants n=37,897. Statistical test: single variant and gene-level burden tests implemented in REGENIE².



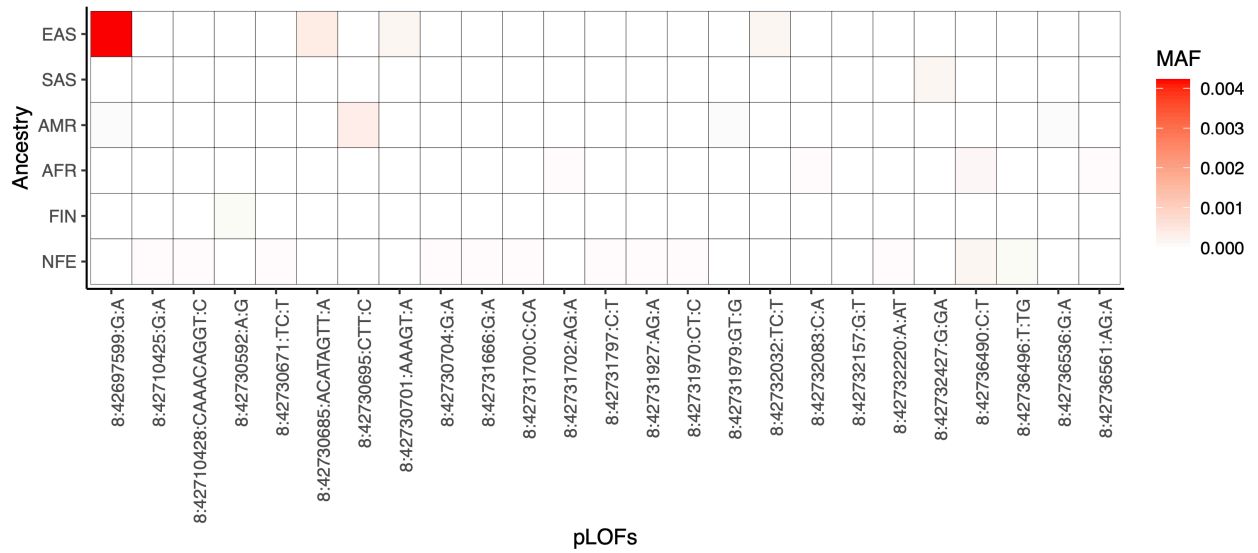
Supplementary Figure 9. CHRNA3 associations in current versus former smokers.

Forest plot summarizing aggregate rare deleterious missense variant associations of CHRNA3 with cigarettes per day in MCPS current smokers and former smokers. Effect sizes in standard deviation units with 95% confidence intervals. Current smokers: n=37,897 individuals. Former smokers: n=33,164 individuals. Statistical test: gene-level burden tests implemented in REGENIE².



Supplementary Figure 10. CHRNA3 rare variant associations in UKB.

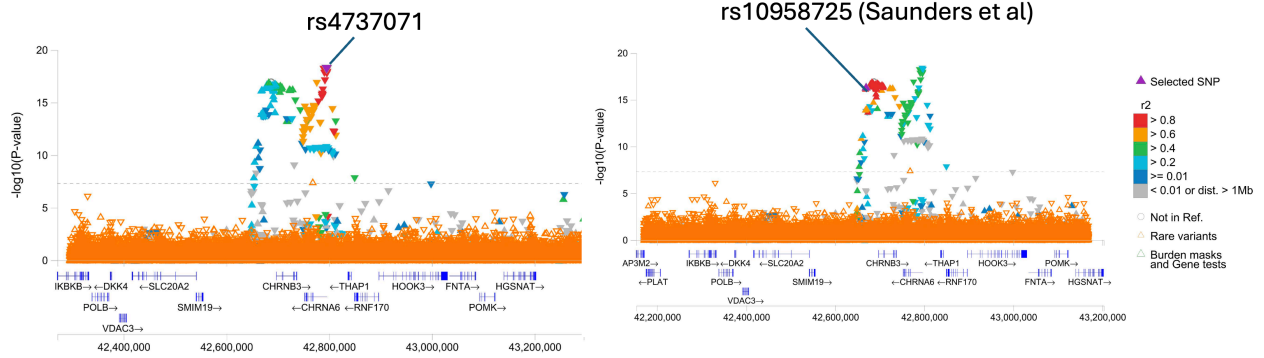
Forest plot summarizing aggregate rare variant associations of CHRNA3 with cigarettes per day in UKB. Most significant association highlighted in pink: pLoF + missense (5/5), AAF<0.1%. Effect sizes in standard deviation units with 95% confidence intervals. n=133,133 participants analyzed. Statistical test: gene-level burden tests implemented in REGENIE².



Supplementary Figure 11. CHRN3 pLOF frequencies across ancestries.

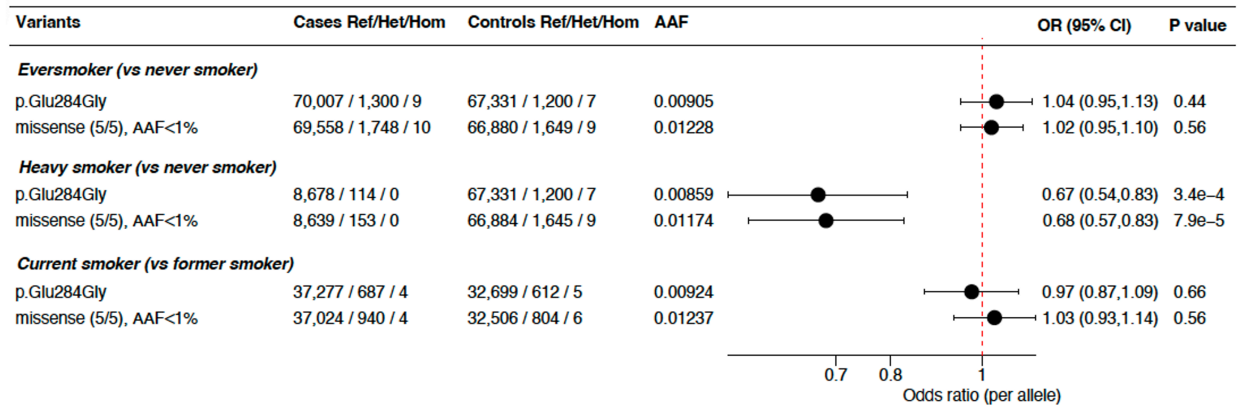
Heatmap displaying minor allele frequencies of predicted loss-of-function variants in CHRN3 with ≥ 10 carriers in gnomAD database. Variants labeled in Chromosome:Position:Allele1:Allele2 format (reference genome build 38). Color scale represents MAF from 0.000 to 0.004.

EAS=East Asians; SAS=South Asians; AMR=admixed Americans; AFR=Africans; FIN=Finnish Europeans; NFE=Non-Finnish Europeans. rs147306385 (8:42697599:G:A) shows East Asian-specific enrichment.



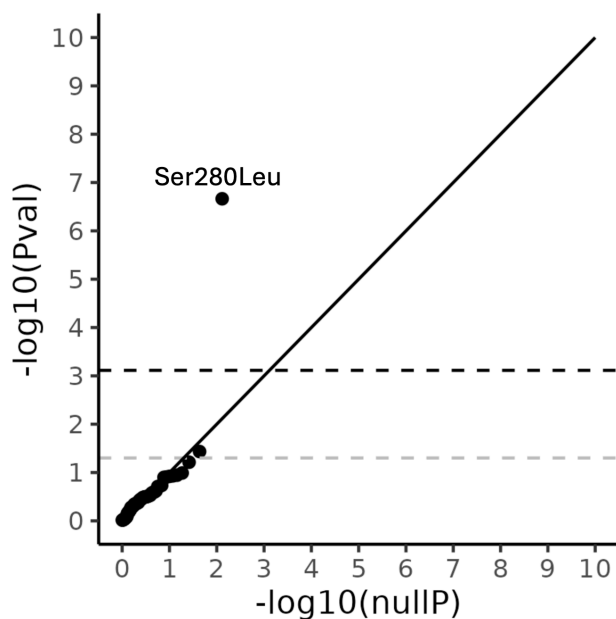
Supplementary Figure 13. CHRN3 locus common variant associations.

Locus zoom plots showing CHRN3 common variant associations from MCPS+UKB meta-analysis. Left panel: rs4737071 (top variant) and linkage disequilibrium patterns. Right panel: rs10958725 (Saunders et al.⁴ top variant) and LD patterns. Color coding represents r^2 values. The two top variants show weak LD ($r^2=0.40$), suggesting independent signals. Y-axis: $-\log_{10}(\text{P-value})$; X-axis: genomic position. Statistical test: linear regression meta-analysis. Two-sided tests.



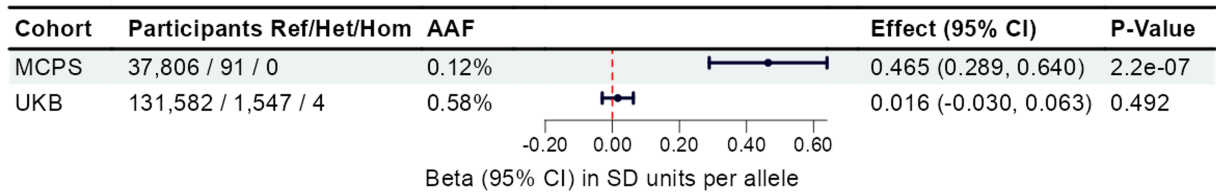
Supplementary Figure 14. CHRNA3 associations with binary smoking phenotypes.

Forest plot comparing CHRNA3 rare variant associations with binary smoking phenotypes in MCPS: ever smoking vs never smoking, heavy smoking vs never smoking, and current smoking vs former smoking. Effect sizes shown as odds ratios with 95% confidence intervals. Sample sizes vary by comparison (provided in Ref/Het/Hom format). Statistical test: single variant and gene-based burden association tests implemented in REGENIE²



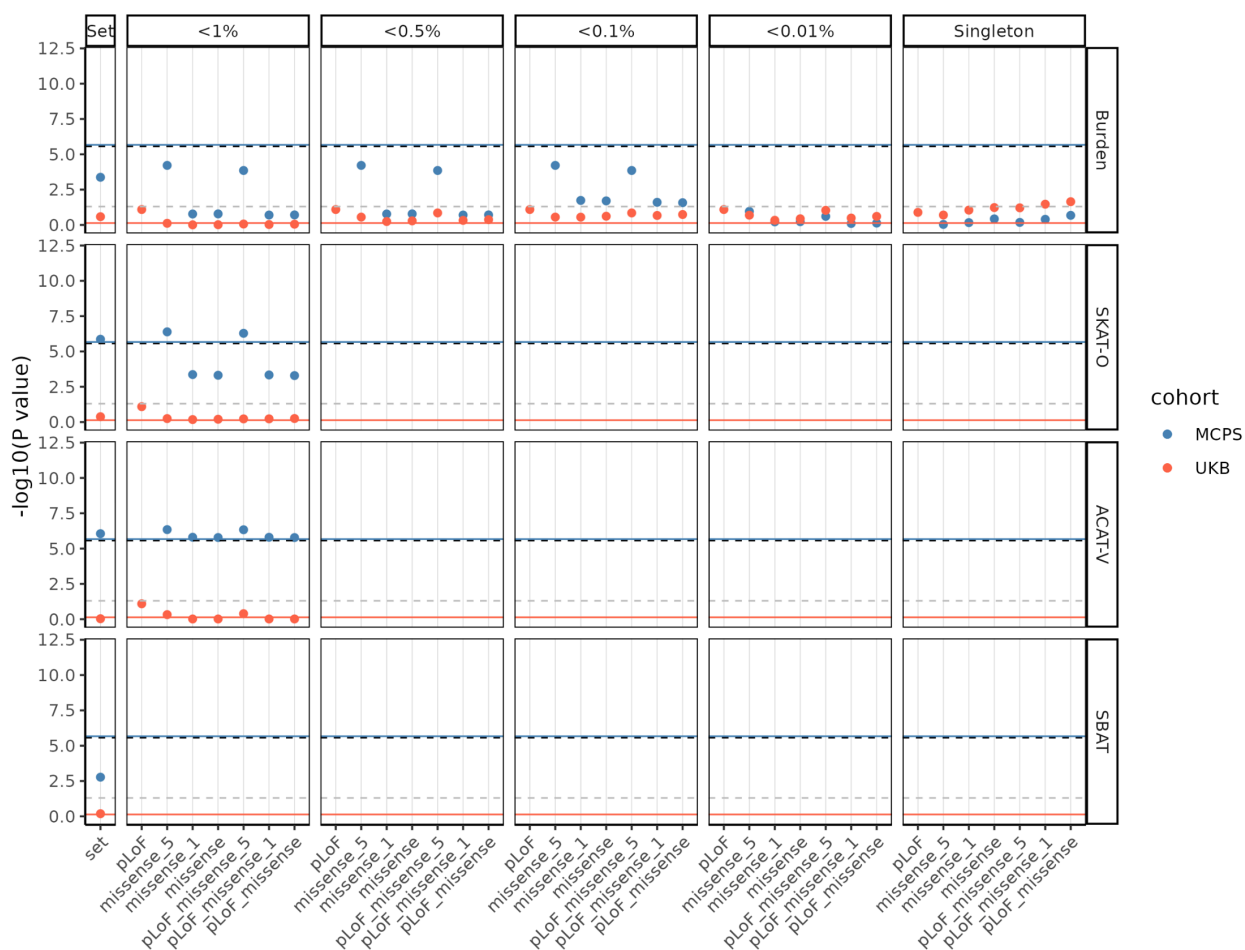
Supplementary Figure 15. VSIR single variant association QQ plot.

Quantile-quantile plot of VSIR rare variant associations with cigarettes per day in MCPS. X-axis: expected P values under null distribution; Y-axis: observed P values (negative log10 scale). Ser280Leu variant highlighted. Black dotted line: Bonferroni-adjusted significance threshold for VSIR variants tested; grey dotted line: $P=0.05$. Statistical test: single variant association tests implemented in REGENIE².



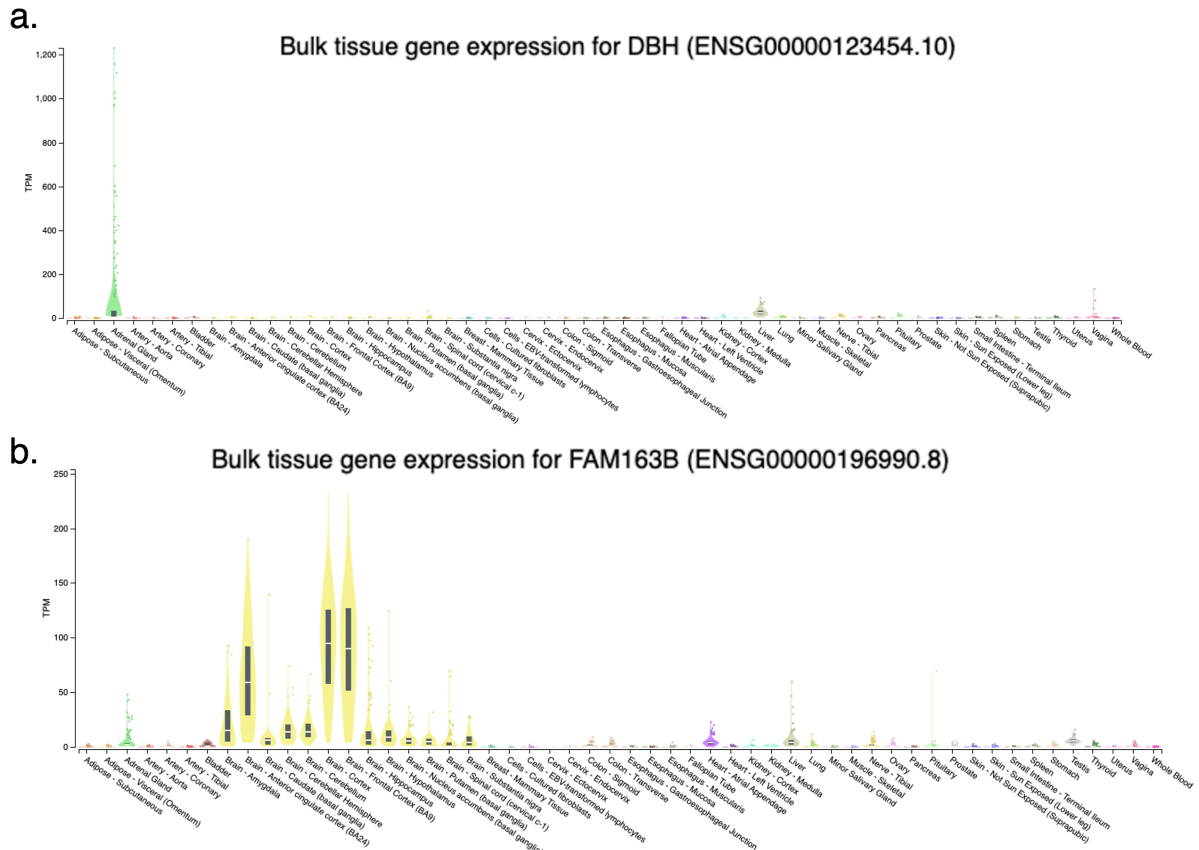
Supplementary Figure 16. VSIR Ser280Leu association comparison.

Forest plot comparing Ser280Leu association with cigarettes per day between MCPS and UKB. Despite higher MAF in UKB (0.58% vs 0.12%), no significant association observed in UKB. Effect sizes in standard deviation units with 95% confidence intervals. MCPS: n=37,897. UKB: n=133,133. Statistical test: single variant association tests implemented in REGENIE².



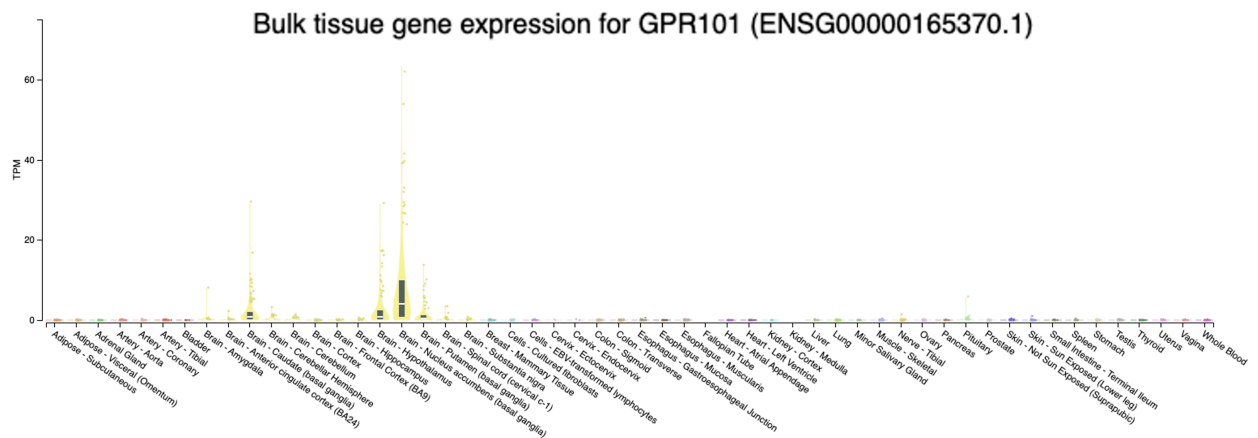
Supplementary Figure 17. VSIR gene-based associations.

Gene-based association results for VSIR with cigarettes per day in MCPS (blue) and UKB (red). X-axis: variant aggregation masks; Y-axis: P values (negative log10 scale) grouped by statistical test type. Horizontal dotted line: $P=0.05$. No significant associations observed in UKB despite larger sample size. Statistical tests: burden, SKAT-O, and ACAT-V tests implemented in REGENIE², described in Supplementary Figure 2 and Ziyatdinov et al. AJHG 2024 (ref.¹)



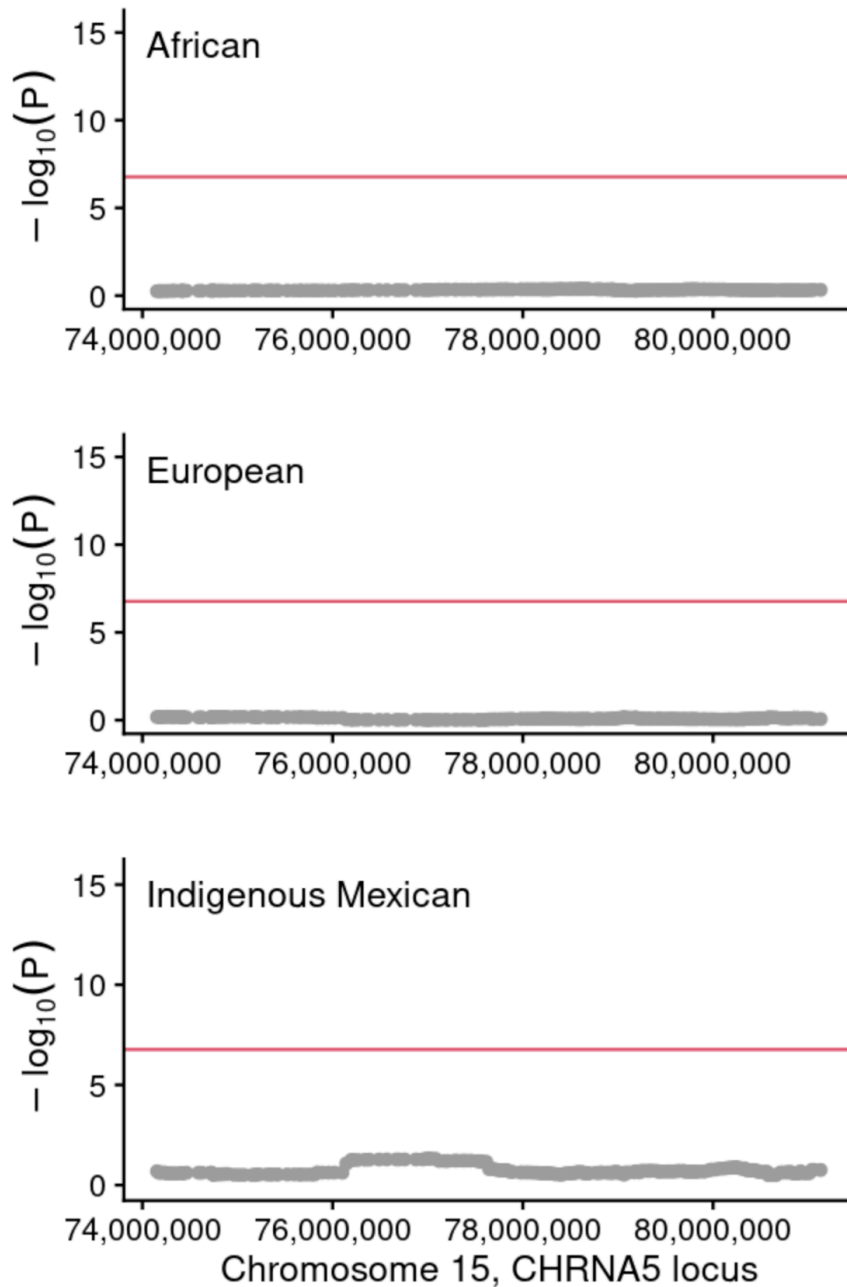
Supplementary Figure 18. DBH and FAM163B tissue expression.

Gene expression profiles for DBH (upper panel) and FAM163B (lower panel) across human tissues based on GTEx RNA-sequencing data. Y-axis: TPM (transcripts per million). Data source: Genotype-Tissue Expression (GTEx) Project database⁵.



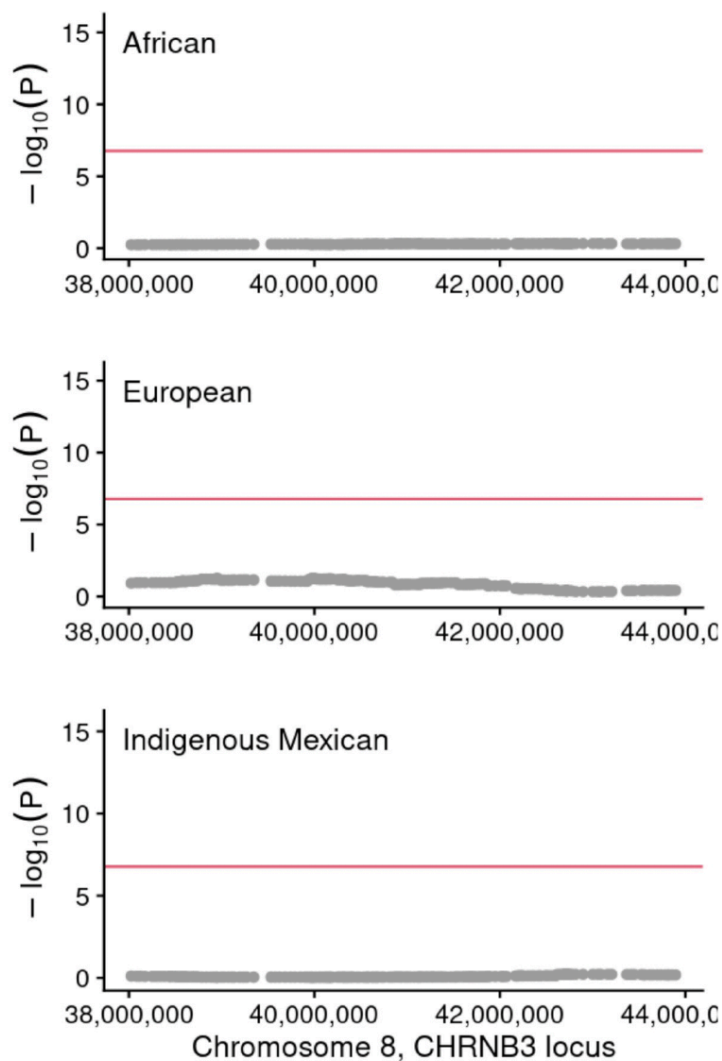
Supplementary Figure 19. GPR101 tissue expression.

Gene expression profile for GPR101 across human tissues based on GTEx RNA-sequencing data. Y-axis: TPM (transcripts per million). Highest expression observed in brain tissues. Data source: Genotype-Tissue Expression (GTEx) Project database.



Supplementary Figure 20. CHRNA5 locus ancestry analysis.

Local ancestry deviation analysis for CHRNA5 genomic region on chromosome 15. P values for deviation of local ancestry proportions (African, European, Indigenous Mexican) from global ancestry proportions. X-axis: genomic position; Y-axis: $-\log_{10}(P\text{-value})$. Red line: significance threshold. No significant deviations observed, indicating no evidence for local selection. Analysis methods described in Ziyatdinov et al. Nature 2023 (ref.⁶).



Supplementary Figure 21. CHRNA3 locus ancestry analysis.

Local ancestry deviation analysis for CHRNA3 genomic region on chromosome 8. P values for deviation of local ancestry proportions (African, European, Indigenous Mexican) from global ancestry proportions. X-axis: genomic position; Y-axis: $-\log_{10}(P\text{-value})$. Red line: significance threshold. Analysis methods described in Ziyatdinov et al. Nature 2023 (ref.⁶).

Supplementary References:

- 1 Ziyatdinov, A. *et al.* Joint testing of rare variant burden scores using non-negative least squares. *Am J Hum Genet* **111**, 2139-2149 (2024).
<https://doi.org/10.1016/j.ajhg.2024.08.021>
- 2 Mbatchou, J. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097-1103 (2021).
- 3 Sakaue, S. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415-1424 (2021). <https://doi.org/10.1038/s41588-021-00931-x>
- 4 Saunders, G. R. B. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720-724 (2022). <https://doi.org/10.1038/s41586-022-05477-4>
- 5 Lonsdale, J. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585 (2013). <https://doi.org/10.1038/ng.2653>
- 6 Ziyatdinov, A. Genotyping, sequencing and analysis of 140,000 adults from Mexico City. *Nature* **622**, 784-793 (2023). <https://doi.org/10.1038/s41586-023-06520-8>