

# Rare coding variants in *CHRNA3* associate with reduced daily cigarette smoking across ancestries

Received: 6 December 2024

Accepted: 8 January 2026

Published online: 24 February 2026

 Check for updates

Veera M. Rajagopal<sup>1</sup>✉, Andrey Ziyatdinov<sup>1</sup>, Tyler Joseph<sup>1</sup>, Ariane Ayer<sup>1</sup>, Mohsin Ahmed<sup>2</sup>, Joelle Mbatchou<sup>1</sup>, Yuxin Zou<sup>1</sup>, Amelia J. Averitt<sup>1</sup>, Nilanjana Banerjee<sup>1</sup>, Michael Cantor<sup>1</sup>, Jason M. Torres<sup>3</sup>, Esteban Chen<sup>1</sup>, Jennifer Rico Varela<sup>1</sup>, Regeneron Genetics Center\*, Marcus Jones<sup>1</sup>, John Overton<sup>1</sup>, Olivier Harari<sup>2</sup>, Luca Lotta<sup>1</sup>, Gonçalo Abecasis<sup>1</sup>, Aris Baras<sup>1</sup>, Jaime Berumen<sup>4</sup>, Pablo Kuri-Morales<sup>5</sup>, Jesús Alegre-Díaz<sup>4</sup>, Roberto Tapia-Conyer<sup>4</sup>, Rory Collins<sup>3</sup>, Jonathan R. Emberson<sup>3</sup>, Jonathan Marchini<sup>1</sup> & Giovanni Coppola<sup>1</sup>✉

Rare coding variants that alter protein function and confer beneficial health effects can suggest potential drug targets. *CHRNA3* encodes the  $\beta 3$  subunit of nicotinic acetylcholine receptors that bind nicotine and mediate its action in the brain. Here we report an exome-wide association study of number of cigarettes smoked per day (cig per day) in 37,897 current smokers from the Mexico City Prospective Study. We identify a deleterious missense variant in *CHRNA3*, p.Glu284Gly, that associates with a significant reduction in daily cigarette consumption. The missense variant is enriched in people of Indigenous Mexican ancestry but rare in other ancestries. We further identify a predicted loss-of-function variant in *CHRNA3* that significantly associates with reduction in number of smoked cigarettes per day in participants of Japan Biobank. This variant is enriched in people of East Asian ancestry but is rare in other ancestries. Finally, we find that rare deleterious missense and predicted loss-of-function variants in aggregate associate with a reduction in the number of smoked cigarettes per day in individuals of European ancestry from the UK Biobank. Our results suggest that loss of function of *CHRNA3* significantly associates with daily cigarette smoking, proposing  $\beta 3$  inhibition as a potential therapeutic strategy for nicotine addiction.

Despite its decline in recent years, cigarette smoking continues to be one of the leading preventable causes of death worldwide<sup>1</sup>. Limited progress has been made in the past two decades in identifying new drugs for smoking addiction<sup>2</sup>. Human genetic association studies offer

a powerful approach to discovering new drug targets<sup>3</sup>. Naturally occurring loss of function genetic variants conferring protection from human diseases are among the best lines of evidence to identify new drug targets. Drug designs mimicking such natural protection

<sup>1</sup>Regeneron Genetics Center, Tarrytown, NY, USA. <sup>2</sup>Regeneron Pharmaceuticals Inc, Tarrytown, NY, USA. <sup>3</sup>Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>4</sup>Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico. <sup>5</sup>Tecnológico de Monterrey, Monterrey, Mexico. \*A list of authors and their affiliations appears at the end of the paper.

✉ e-mail: [veera.rajagopal@regeneron.com](mailto:veera.rajagopal@regeneron.com); [giovanni.coppola@regeneron.com](mailto:giovanni.coppola@regeneron.com)

mechanisms have delivered successful medicines, e.g., *CCRS* inhibitors for the treatment of human immunodeficiency virus infection and *PCSK9* inhibitors for the treatment of hypercholesterolemia<sup>3</sup>.

Cigarette smoking has been extensively characterized at the molecular level compared to other addiction-related behaviors, thanks to decades of animal and human genetic studies<sup>4</sup>. Some of the first loss-of-function protective genetic variants in a human behavioral trait were discovered for smoking in the 1990s. A null missense variant in *CYP2A6* with a minor allele frequency (MAF) of ~2.5% in Europeans<sup>5</sup> and a *CYP2A6* gene deletion variant with an MAF of ~20% in East Asians<sup>6</sup> associated with lower average number of cigarettes smoked daily by slowing the nicotine metabolism in the blood. With rapidly increasing population-level sequencing studies, similar protective associations are beginning to emerge. For example, through a large-scale exome-wide association study (ExWAS) of smoking behavior, we have discovered that rare deleterious coding variants in *CHRN2*, encoding the  $\beta 2$  subunit of nicotinic acetylcholine receptors (nAChRs), associate with reduced likelihood of heavy smoking<sup>7</sup>. The findings validated the well-characterized role of  $\beta 2$  subunit in mediating the rewarding effects of nicotine<sup>8</sup>.

The nAChRs are ion channels made of five subunits. There are at least nine nAChR subunit types ( $\alpha 2$ –7,  $\beta 2$ –4) known to be expressed in the human brain, most of which were linked to smoking behavior-related phenotypes such as ever smoking (vs never), smoking quantity (measured as number of cigarettes per day (cig per day)), nicotine addiction, and tobacco use disorder through common (MAF  $\geq 1\%$ ) variant-based genome-wide association studies (GWAS)<sup>9–13</sup>; some of the strongest associations include loci 15q25.1 containing *CHRNA5*, *CHRNA3* and *CHRN4* (ref. 14), 8p11.21 containing *CHRN3* and *CHRNA6* (ref. 15) However, with few exceptions such as  $\beta 2$  and  $\alpha 5$ , the phenotypic consequences of loss or gain of these subunits on smoking behavior in humans are poorly understood<sup>7,16,17</sup>. Sequencing large

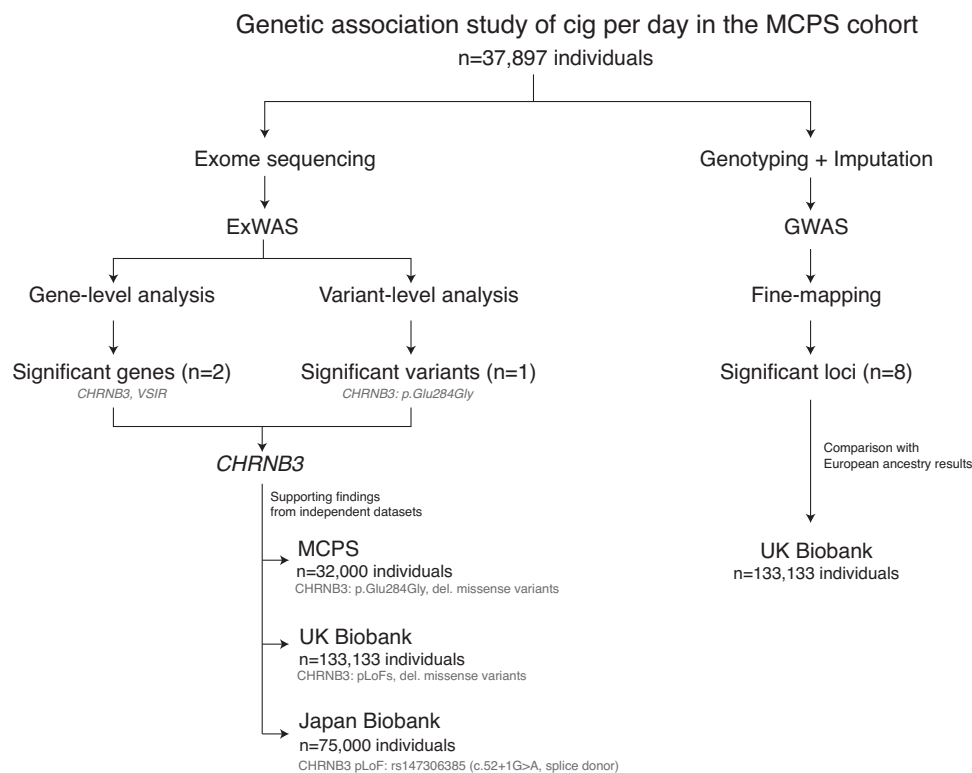
numbers of individuals from diverse ancestries can help discover rare coding variants in these subunits, and investigating associations with smoking behavior provides insight into their functional roles within addiction-related neural pathways.

We have recently exome sequenced around 150,000 Mexican residents who were participants of the Mexico City Prospective Study (MCPS) established in the late 1990s<sup>18,19</sup>. In continuation of our previous work<sup>7</sup> on rare-variant (MAF < 1%) associations of smoking behavior, we now report an ExWAS of cig per day in individuals of Indigenous Mexican ancestry from the MCPS cohort (throughout this article, the usage of the word ‘ancestry’ indicates genetically predicted ancestry of the participants). The ExWAS revealed a significant association with *CHRN3* driven by an Indigenous Mexican-enriched missense variant. We replicate the *CHRN3* protective signal using independent samples from three ancestries: Indigenous Mexicans<sup>18</sup>, Europeans<sup>20</sup>, and East-Asians<sup>21</sup>. We further report a GWAS of cig per day in Indigenous Mexican ancestry, compare the results with Europeans-based results, and discuss the similarities and differences in the common variant architecture between the two ancestry groups.

## Results

### Exome-wide association study

The overall study design is shown in Fig. 1, and the demographics of the study participants, including the distribution of cig per day, are reported in Supplementary Data 1 and Supplementary Fig. 1. We studied 37,897 individuals (19,550 males and 18,347 females) from the MCPS cohort who were current smokers and for whom information on cig per day, rare variants (based on exome sequencing), and common variants (based on array genotyping combined with imputation) were available (“Methods”). Our primary goal was to discover potential drug targets for smoking addiction based on genetic associations driven by rare coding variants, which point to the causal genes with high



**Fig. 1 | Study design.** The flow chart summarizes the study design. The primary study is a genetic association analysis of cigarettes per day in 37,897 individuals from the MCPS cohort. The secondary studies include genetic association analyses of cigarettes per day in MCPS (independent sample), UK Biobank, and Japan

Biobank. In addition to cigarettes per day, other smoking phenotypes (ever smoker and heavy smoker) were analyzed, which are described in the manuscript. ExWAS Exome-wide association study, GWAS Genome-wide association study, MCPS Mexico City Prospective Study.

confidence. To achieve that goal, we performed an ExWAS of cig per day in the MCPS cohort. We tested the associations of rare (MAF < 1%) missense and predicted loss-of-function (pLOF) variants grouped under seven combinations of variant classes and five categories of allele frequency bins (Supplementary Data 2). The association analyses were performed using the latest version of REGENIE<sup>22</sup>, which tests variant associations using different statistical approaches and combines the results using an ensemble approach to calculate a single, unified  $P$  value per gene called “gene  $P$ ” (Supplementary Fig. 2 and Supplementary Data 3). We used the gene  $P$  values to identify exome-wide significant genes using a significance threshold of  $P < 2.6 \times 10^{-6}$  based on 18,955 genes tested ( $0.05/18,955$ ). We also calculated gene  $P$  based on only synonymous variants as a reference for null distribution with the assumption that synonymous variants overall do not associate with the phenotype; this helped assess any potential inflation in the test statistics. After discovering significant genes, we studied the individual association tests in detail to understand the direction of the effect and the classes of variants and allele frequency bins contributing to the association signals. All the association tests were adjusted for important confounders, including age, sex, relatedness, genetic principal components, and any nearby common variant signals (“Methods”). Our primary analysis was focused on gene-level associations as aggregate testing offers greater statistical power than testing one variant at a time, and further, our gene  $P$  approach is expected to capture even scenarios where gene associations are driven by a single variant<sup>22</sup>. However, we also explored the single-variant associations as a secondary analysis. We tested the associations of 449,606 variants including pLOFs and missense variants mapped to canonical transcripts. Variants with  $P < 1.1 \times 10^{-7}$  were considered statistically significant based on a Bonferroni-adjusted significance threshold ( $0.05/449,606$ ). We also performed ExWAS of cig per day in 133,133 individuals of European ancestry from the UK Biobank (UKB) using the same analysis workflow for comparison of results between Indigenous

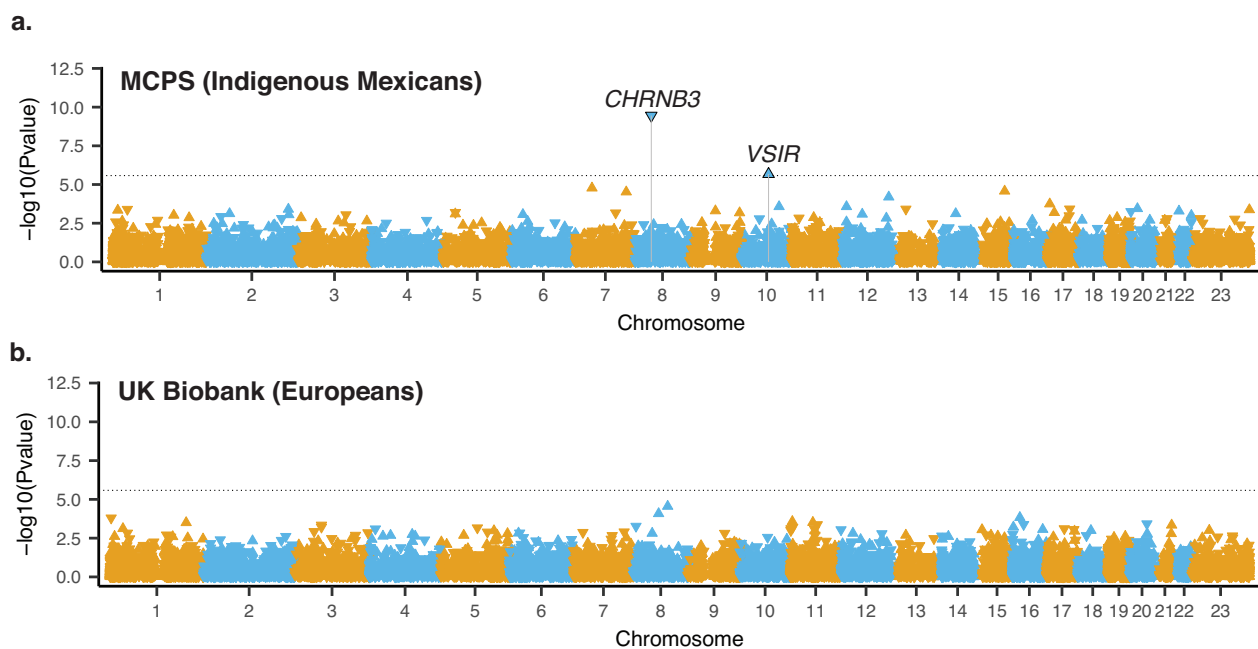
Mexican and European ancestries and cross-ancestry validation of significant findings.

### Exome-wide significant genes and variants

ExWAS of cig per day in the MCPS cohort using gene-based testing identified two genes significantly associated with cig per day: *CHRNA3* (gene  $P = 3.5 \times 10^{-10}$ ) and *VSIR* (gene  $P = 2.1 \times 10^{-6}$ ) (Fig. 2a and Supplementary Fig. 3a). Testing individual rare coding variants across the exome, we identified a missense variant, p.Glu284Gly (rs75384358), in *CHRNA3* significantly associated with cig per day at exome-wide significance (beta =  $-0.19$ ;  $P = 1.1 \times 10^{-9}$ ; Supplementary Fig. 4a). Neither the gene-level nor variant-level testing yielded any significant rare-variant associations with cig per day in the UKB, despite a 3.5-fold larger sample size compared to MCPS (Fig. 2b and Supplementary Figs. 3b and 4b).

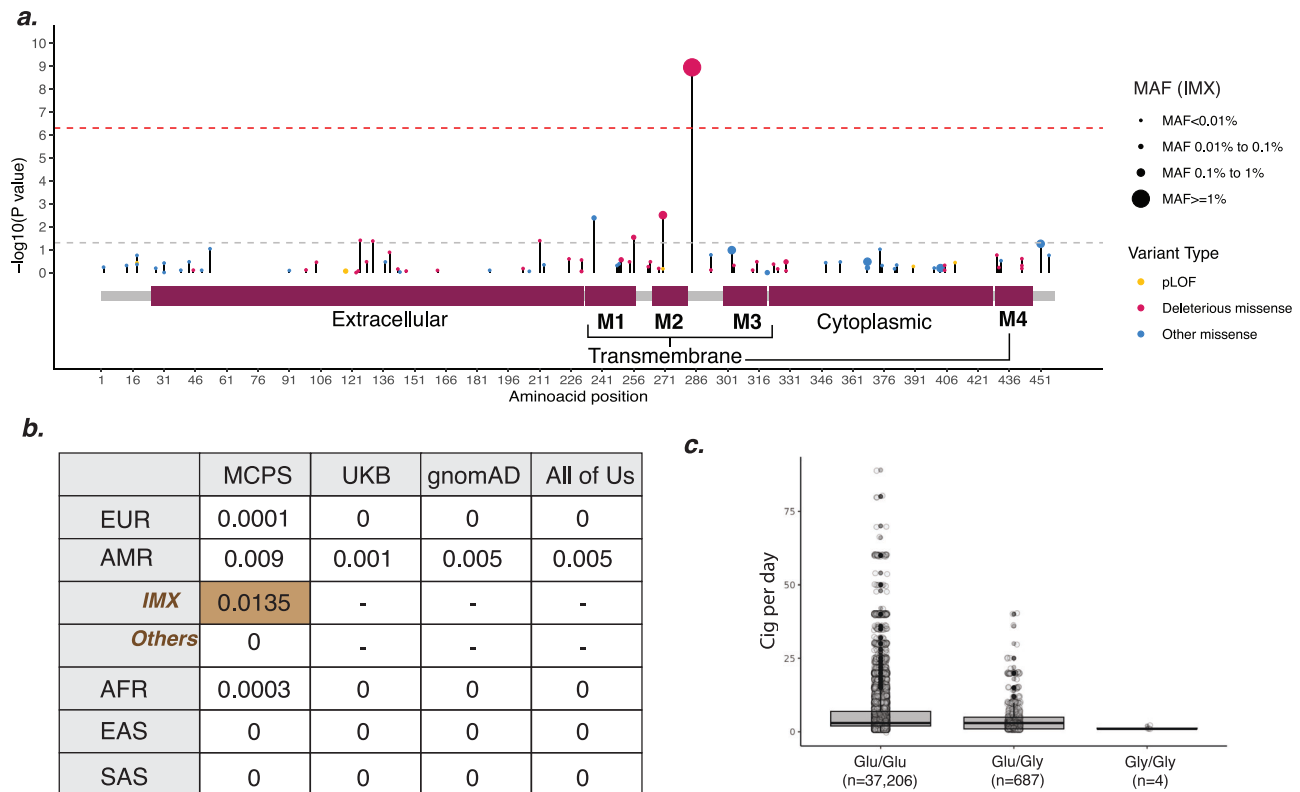
### Association of rare coding variants in *CHRNA3*

*CHRNA3* showed the strongest association with cig per day both at the gene-level and variant-level analysis in the MCPS cohort. Comparing the individual gene-level association tests, we found that burden tests that aggregate carriers across all deleterious missense variants in *CHRNA3* showed the strongest statistical association (beta =  $-0.19$ ;  $P = 2.8 \times 10^{-11}$ ) (Supplementary Figs. 5 and 6 and Supplementary Data 4). We found in total, 944 carriers (including 4 homozygotes) of deleterious missense variants in *CHRNA3* who smoked significantly fewer cig per day on average compared to 36,951 non-carriers (beta =  $-0.19$ ;  $P = 3.6 \times 10^{-11}$ ). Most of these individuals (73%;  $n = 691$  including 4 homozygotes) were carriers of a single missense variant, p.Glu284Gly, the only variant that showed exome-wide significance in our variant-level analysis (Supplementary Fig. 4a and Supplementary Data 5). Therefore, the gene-level burden association was primarily driven by this single missense variant. However, excluding p.Glu284Gly, we found that the remaining 253 carriers in aggregate also smoked fewer cig per day compared to non-carriers at nominal significance, suggesting other



**Fig. 2 | ExWAS Manhattan plots.** Manhattan plots showing results from exome-wide rare-variant association studies of cigarettes per day in the MCPS (a) and UKB (b). Number of individuals analyzed: MCPS – 37,897; UK Biobank – 133,133. Only aggregate level associations are plotted. The X axes correspond to chromosomal base pair start positions of individual genes and Y axes correspond to  $P$  values (expressed in negative log scale) from burden association tests performed using REGENIE. The dotted line corresponds to a statistical significance threshold of

$P = 2.6 \times 10^{-6}$ . The direction of the triangle represents the direction of association (downward—cig per day decreasing; upward—cig per day increasing); MCPS Mexico City Prospective Study. Statistical test used: burden association test implemented in REGENIE software. Two-sided tests were performed with adjustment for multiple tests using Bonferroni method adjusting for the number of unique genes tested ( $n$  genes = 18,955).



**Fig. 3 | CHRNB3 p.Glu284Gly characterization.** **a** Illustrates the distribution of rare pLOF, deleterious missense and other missense variants (differentiated using colors) in CHRNB3 identified in the MCPS cohort. The *X* axis corresponds to the amino acid position on canonical transcript. The *Y* axis corresponds to *P* values of variant associations with cigarettes per day in the MCPS. The thin grey bar represents the full protein, and the thick colored bars correspond to protein domains. M1 to M4 are parts of the transmembrane domain of the  $\beta$ 3 subunit. The dotted grey line corresponds to  $P = 0.05$  and the red dotted line corresponds to  $P = 5e-7$ , exome-wide association threshold, multiple-tests-adjusting for number of unique rare variants tested across the whole exome. The size of the circles corresponds to different minor allele frequency (MAF) thresholds. **b** displays the MAF of Glu284Gly

in MCPS and other databases (UKB, gnomAD, and All of Us) in five continental ancestries: Europeans (EUR), admixed Americans (AMR), Africans (AFR), East Asians (EAS), and South Asians (SAS). In the MCPS, using fine-scale ancestry estimation, we further report the Indigenous Mexican (IMX) specific MAF. **c** displays box plots overlaid on a scatter plot of cigarettes per day distribution across three genotype groups of p.Glu284Gly. Data are presented as individual data points with box plots showing median (center line), first and third quartiles (box bounds), and whiskers extending to 1.5 times the interquartile range.  $n = 37,206$  participants (homozygous reference),  $n = 687$  participants (heterozygous),  $n = 4$  participants (homozygous variant). Statistical test used: linear regression adjusted for ancestry, age, and sex. Two-sided test. No multiple tests adjustment.

deleterious missense variants in CHRNB3 also influenced smoking (beta =  $-0.15$ ;  $P = 0.006$ ; Supplementary Fig. 7). Focusing on the pLOFs, we did not observe a significant association, possibly due to low statistical power as pLOFs in CHRNB3 were rare in the MCPS cohort; in total we found only 31 carriers of pLOFs (aggregate MAF = 0.04%) in the analyzed sample, fewer than the 944 carriers of deleterious missense variants (aggregate MAF = 1.3%; Supplementary Fig. 6 and Supplementary Data 4). We tested the CHRNB3 association with cig per day after removing related individuals (up to 3rd degree relatives) from the MCPS cohort. The results still showed significant protective associations, confirming that the high relatedness in the MCPS cohort did not affect our findings (Supplementary Fig. 8).

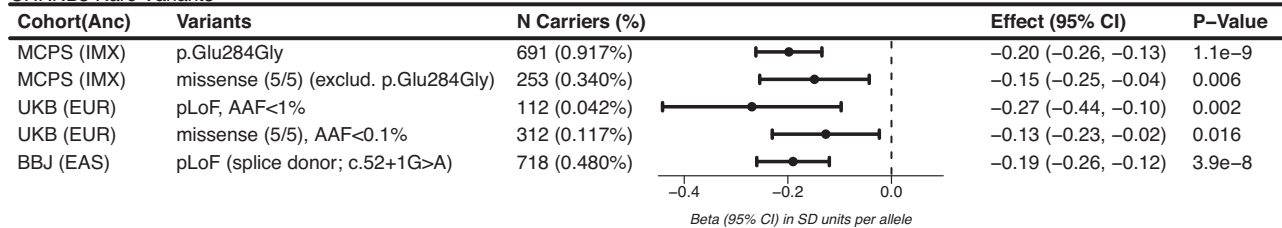
The variant p.Glu284Gly corresponds to a glutamate to glycine substitution at position 284 in the  $\beta$ 3 subunit protein based on the canonical transcript (Fig. 3a). It is located between the second (M2) and third (M3) transmembrane domain, more proximal to M2 that forms the ion channel part of the nAChR receptor. The MAF of p.Glu284Gly was 0.9% in the MCPS cohort. Fine-scale ancestry analysis showed that all the carriers were of Indigenous Mexican ancestry<sup>18</sup>. The haplotype-resolved Indigenous Mexican-specific MAF of p.Glu284Gly was 1.4% (Fig. 3b). The variant is either absent or extremely rare in other ancestry groups, hence it appears to be specific to Indigenous Mexican ancestry. Notably, it was the most frequent coding variant in CHRNB3 (considering pLOFs and all missense) in the MCPS cohort (Fig. 3a). The

heterozygous and homozygous carriers smoked on average 4.6 and 1.25 cig per day, respectively, which were 21% and 78% lower compared to non-carriers who smoked on average 5.6 cig per day (Fig. 3c). The homozygous effect size exceeded the heterozygous effect size beyond what would be expected in an additive model (expected beta =  $-0.40$ ; observed beta =  $-0.90$ ). The cig per day reduction by p.Glu284Gly was more pronounced in males (beta =  $-0.26$ ;  $P = 5.7e-8$ ) than females (beta =  $-0.14$ ;  $P = 0.002$ ); this is likely due to the fact that females smoked fewer cigarettes than males in this cohort owing to cultural factors, which in turn attenuated the genetic effects.

To validate the rare-variant association of CHRNB3 with cig per day, we used three datasets (Fig. 1). Firstly, we tested the association in an independent set of 33,162 individuals in the MCPS cohort who were former smokers and reported on the number of cigarettes they smoked on average per day in the past. We found a nominally significant gene-level association for CHRNB3 (gene  $P = 0.001$ ; Supplementary Data 4). Focusing on individual tests, we found a nominally significant protective association for both rare deleterious missense variants in aggregate (beta =  $-0.11$ ,  $P = 5.e-4$ ) and p.Glu284Gly (beta =  $-0.10$ ,  $P = 0.007$ ) in isolation with cig per day, thus supporting our discovery findings (Supplementary Fig. 9). Despite having comparable sample size, we noticed that the effect size and *P* value were modest in the former smokers compared to current smokers. We suspect that this is likely due to factors such as recall bias (as participants tend to

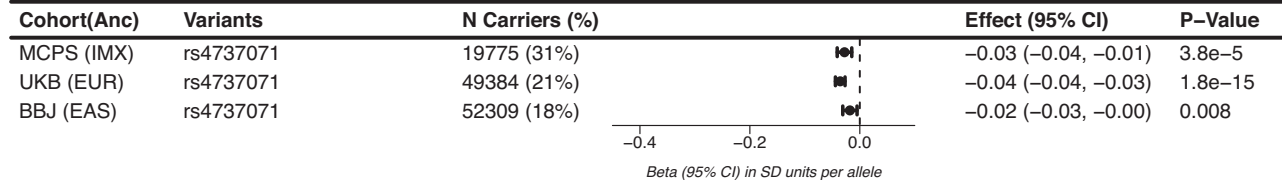
a.

## CHRNA3 Rare Variants



b.

## CHRNA3 Common Variants



**Fig. 4 | CHRNA3 summary forest plots.** The figure displays forest plots summarizing the rare (a) and common variant associations (b) of CHRNA3 with cigarettes per day in diverse ancestries. The rare-variants plot (a) displays an allelic series of associations of rare missense and pLOFs from MCPS, UKB, and BBJ. *N* carriers represent total number of heterozygous and homozygous carriers for effect allele. The effect size in standard deviation units along with 95% confidence intervals are

plotted. The dashed vertical line corresponds to null effect ( $\beta = 0$ ). IMX Indigenous Mexicans, EUR Europeans, EAS East Asians, UKB UK Biobank, MCPS Mexico City Prospective Study, BBJ Biobank Japan. Statistical tests used: linear or logistic regression models, testing association with individual variant or aggregate of variants, adjusted for ancestry, age, and sex for MCPS; Two-sided tests. No multiple tests adjustment.

report their current smoking behavior more accurately than their past behavior) and dilution of genetic effects by environmental factors (e.g., reduced smoking on doctor's advice).

Secondly, we studied the CHRNA3 rare-variant association with cig per day in the UKB. Overall, we found a nominally significant gene-level association for CHRNA3 (gene  $P = 0.008$ ; Supplementary Data 4). Comparing individual gene-based tests, the most significant association was observed for gene-burden test that aggregated ultra-rare (MAF < 0.1%) deleterious missense variants and rare pLOFs (beta = -0.16;  $P = 4.2e-4$ ; Supplementary Fig. 10). Notably, we observed nominal protective association when aggregating only pLOFs (beta = -0.27;  $P = 0.002$ ), which was not observed in the MCPS. The larger sample size in UKB compared to MCPS yielded more pLOF carriers (112 carriers in UKB vs 31 carriers in MCPS), hence boosting the statistical power to capture at least a nominal signal.

Thirdly, we attempted to validate our findings by leveraging any publicly available GWAS datasets. Allele frequencies of certain coding variants can differ dramatically between ancestries due to phenomena like genetic drift. As a result, a rare variant in one ancestry group that can be detected only via sequencing is sometimes common in another ancestry group and detectable via genotyping and imputation. We surveyed the allele frequencies of pLOFs in CHRNA3 across major ancestries. We focused only on pLOFs as their functional effects are relatively more homogeneous, hence interpretable with more confidence compared to missense variants. The variant survey identified a pLOF in CHRNA3 that was enriched in East Asian ancestry. The pLOF was a splice donor variant (c.52+1G>A, rs147306385) seen at a MAF of 0.5% in East Asian ancestry, but not observed in other ancestry groups (Supplementary Fig. 11). Querying the phenome-wide association results of this variant in a public repository of genetic association results hosted by the Japan biobank<sup>21</sup>, we found that the top association of the pLOF variant (which was the only significant association across the phenome after adjusting for multiple testing) was cig per day, with an effect size comparable to that of p.Glu284Gly in the MCPS cohort (beta = -0.19;  $P = 3.9e-8$ ; Supplementary Fig. 12).

In summary, we provide evidence supporting the protective association of rare coding variants in CHRNA3 with cig per day in three datasets independent of our discovery sample. Altogether, we found

rare pLOFs and deleterious missense variants in CHRNA3 associate with significant reduction in cig per day in diverse ancestries—Indigenous Mexican, European, and East Asian (Fig. 4a).

### Association of common variants near CHRNA3

One of the first GWAS loci for cig per day discovered by early studies of smoking behavior was 8p11 encompassing two nAChR genes: CHRNA3 and CHRNA6, first reported by Thorgeirsson et al. (ref. 15) and later replicated by multiple GWASs<sup>10,11</sup>. The most recent trans-ancestry GWAS of cig per day by Saunders et al. (ref. 11) reported a single fine-mapped locus near CHRNA3 where the top variant was rs10958725; this is in high LD with rs13280604 ( $r^2 = 0.99$ ), the top variant reported by Thorgeirsson et al. (ref. 15), hence both represent the same signal. To characterize the common variant signal at CHRNA3, we meta-analyzed the common variants (MAF > 0.01) results of UKB and MCPS and found a single genome-wide significant locus near the CHRNA3/CHRNA6 locus (independent of the rare variant p.Glu284Gly) where the top variant among the fine-mapped SNPs was rs4737071, an intronic variant in CHRNA6. This was only moderately correlated ( $r^2 = 0.40$ ) with the top variant, rs10958725, reported by Saunders et al. (ref. 11) in both EUR and admixed American (AMR) ancestry groups. We studied this signal using conditional analyses and found that our common variant signal retains statistical significance after conditioning on the previously reported signal by Saunders et al. (ref. 11) ( $P = 1.5e-6$ ). Similarly, when conditioning on our top variant, the Saunders et al. (ref. 11) signal remains statistically significant in our meta-analysis ( $P = 0.0004$ ). These results indicate that, despite a moderate correlation, the two associations at the CHRNA3 locus may represent independent signals (Supplementary Fig. 13). We report the associations of our top SNP (Fig. 4) as well as the association of the top SNP reported by Saunders et al. (ref. 11) after conditioning on our top SNP and vice versa (Supplementary Data 6). For both variants, the carriers of the minor allele smoked fewer cig per day compared to carriers of the major allele; the effect size was expectedly substantially lower compared to p.Glu284Gly. Although the common variant signal at the 8p11 locus has been known for a long time, the causal gene was not clear. Our rare-variant findings favor CHRNA3 as the likely causal gene at this locus, driving the common variant associations.

In summary, the findings so far demonstrate an allelic series of genetic associations of *CHRNA3* with cig per day comprising both common non-coding and rare coding variants from diverse ancestries (Indigenous Mexican, European, and East Asian), implicating *CHRNA3* strongly in the biology of smoking addiction (Fig. 4a, b).

### Association of other smoking phenotypes with *CHRNA3*

The primary association of rare and common variants in *CHRNA3* was observed with cig per day, which indexes the tolerance to nicotine dose. We next evaluated if rare variants in *CHRNA3* also influence the risk of becoming a smoker. We studied *CHRNA3* associations with two binary phenotypes, namely ever vs never smokers (cases are those who ever smoked regularly in their lifetime and controls are never smokers) and heavy vs never smokers (cases are those who smoked 10 or more cigarettes per day) in the MCPS cohort. Neither p.Glu284Gly nor rare deleterious missense variants in *CHRNA3* in aggregate were significantly associated with ever smoking (vs never) (Supplementary Fig. 14). However, both showed nominally significant associations with heavy smoking (vs never). Rare variants in *CHRNA3* associated with decreased risk of heavy smoking (vs never) by -32%. We also tested the association of rare variants in *CHRNA3* with likelihood of quitting smoking by comparing current vs former smokers, but found no significant association (Supplementary Fig. 14).

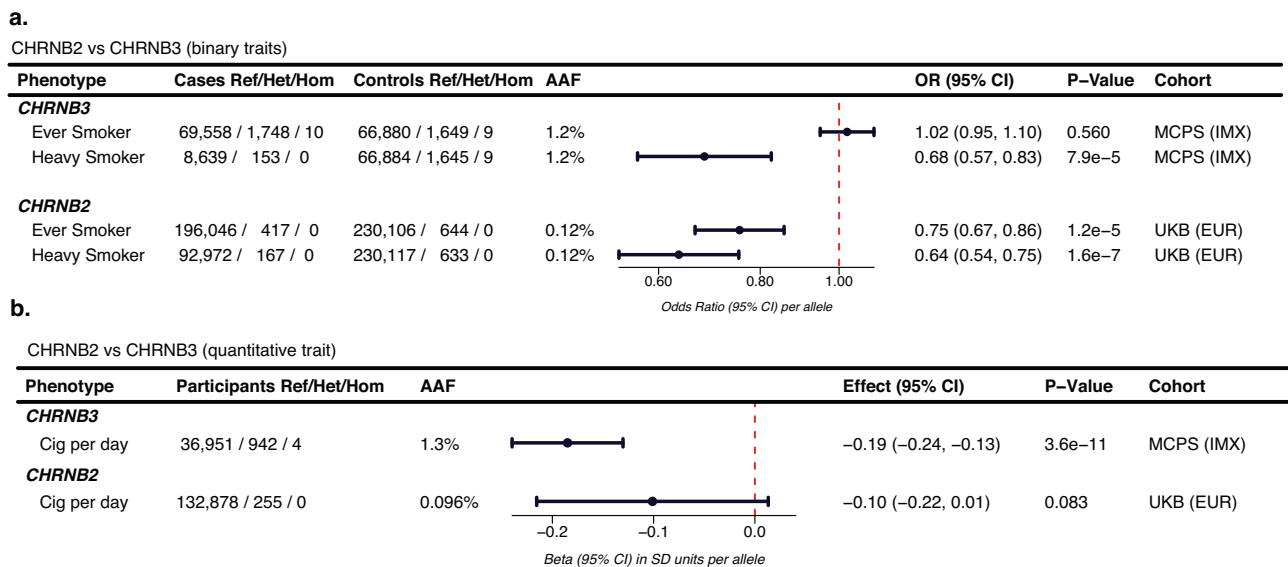
### *CHRNA2* vs *CHRNA3*

We previously reported a protective association between rare variants in *CHRNA2* and smoking in the UKB<sup>7</sup>. Specifically, we have shown that rare pLOFs and deleterious missense variants in *CHRNA2* in aggregate associate with reduction in the risk of ever smoking (vs never) by 25% and heavy smoking (vs never) by 35%, recapitulating the well-characterized role of *CHRNA2* in mediating the rewarding effects of nicotine<sup>8</sup>. Motivated by the current discovery, we compared the rare-variant associations of *CHRNA2* with smoking in the UKB against that of *CHRNA3* in the MCPS (Fig. 5a). We found similar levels of reduction (32%–34%) in the risk of heavy smoking (vs never) in both the *CHRNA3*

and *CHRNA2* rare-variant carriers. However, a reduction in the risk of ever smoking (vs never) (i.e., becoming a smoker) was seen only in the rare-variant carriers of *CHRNA2* but not *CHRNA3* (Fig. 5a). Regarding cig per day, while rare variants in *CHRNA3* were associated with lower cig per day, rare variants in *CHRNA2* were not significantly associated with cig per day (Fig. 5b). We note these comparisons should be read with two caveats in mind. Firstly, the *CHRNA2* associations were based on European ancestry from the UKB (the main rare variant driving the *CHRNA2* association occurs only in Europeans) and the *CHRNA3* associations were based on Indigenous Mexican ancestry from the MCPS (the main rare variant driving the *CHRNA3* association occurs only in Indigenous Mexicans). Secondly, there were -tenfold more rare-variant carriers for *CHRNA3* (aggregate MAF = 1.2%) in the MCPS than for *CHRNA2* (aggregate MAF = 0.1%) in the UKB, hence we were statistically better powered to detect associations for *CHRNA3* than *CHRNA2*.

### Associations of rare coding variants in *VSIR*

The second gene that was exome-wide significantly associated with cig per day in the MCPS cohort based on gene-based testing was *VSIR*, which encodes an immune-related protein, V-set immunoregulatory receptor (Fig. 2a and Supplementary Data 7). Focusing on individual coding variant associations, we found one missense variant, Ser280Leu, in *VSIR*, which showed a significant association with cig per day (Supplementary Fig. 15 and Supplementary Data 8). After excluding this variant, the gene-level tests were no longer significant (gene  $P > 0.05$ ), suggesting that the gene-level associations were driven by Ser280Leu. We studied MAF of Ser280Leu across ancestries, expecting that it may be enriched in the Indigenous Mexican ancestry. On the contrary, we found that it is less frequent among Indigenous Mexicans (MAF = 0.12%) than Europeans (MAF = 0.58%). Further analysis of Ser280Leu in the UKB revealed no significant association with cig per day (beta = 0.01,  $P = 0.49$ ; Supplementary Fig. 16), despite 17-fold more carriers than in the MCPS (1547 vs 91). Gene-based analysis of *VSIR* in the UKB did not show a significant association with cig per day either (Supplementary Fig. 17). The lack of replication in a better-



**Fig. 5 | *CHRNA2* vs *CHRNA3*.** The figure displays forest plots contrasting the rare-variant associations of *CHRNA3* vs *CHRNA2* for binary (a) and quantitative (b) smoking phenotypes. Aggregate level associations based on the rare deleterious missense variants for *CHRNA3* (which showed the strongest association with its primary phenotype, cigarettes per day, in the MCPS) and rare pLOFs plus deleterious missense variants for *CHRNA2* (which showed the strongest association with its primary phenotype, heavy smoker, in the UKB). The “Ref/Het/Hom” counts refer to number of homozygous for reference allele, heterozygous, and homozygous for

effect allele respectively. The effect size in standard deviation units along with 95% confidence intervals are plotted. The dashed vertical line corresponds to null effect ( $\beta = 0$ ). IMX Indigenous Mexicans, EUR Europeans, UKB UK Biobank, MCPS Mexico City Prospective Study. Statistical tests used: linear regression for cigarettes per day; logistic regression for smoking status phenotypes. All models adjusted for ancestry, age, and sex. Two-sided tests. No multiple tests adjustment.

powered analysis suggested that the association of *VSIR* with cig per day in the MCPS is a false positive.

### Genome-wide association study

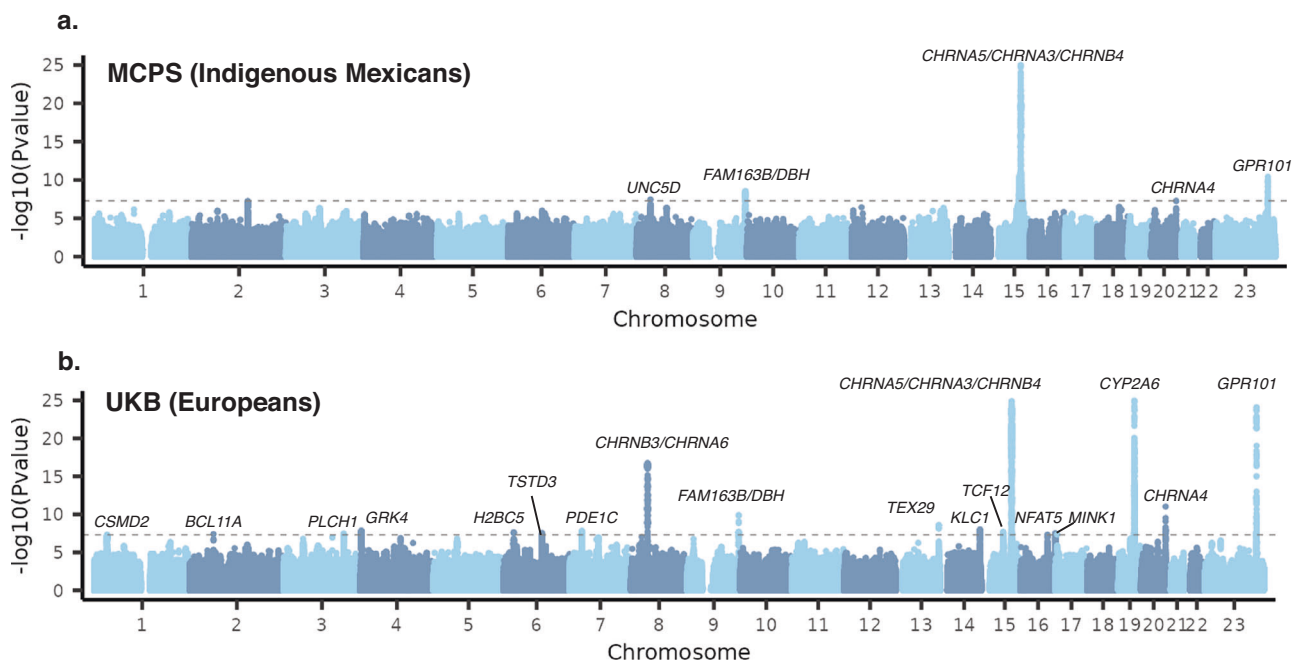
Common variant associations of cig per day in Indigenous Mexican ancestry have not been investigated in detail so far. The most recent GWAS of cig per day in a broader admixed American ancestry group, reported by Saunders et al. (ref. 11) in a sample of 35,129 individuals (85% from 23andMe), identified one genome-wide signal at the well-established 15q25.1 locus containing the *CHRNA3*, *CHRNA5*, and *CHRNB4* gene cluster<sup>12</sup>. Therefore, to further understand the common variant architecture of cig per day in Indigenous Mexican ancestry, we performed a GWAS of cig per day in 37,897 individuals (current smokers) from the MCPS cohort, focusing on common variants (MAF > 1%). We identified eight genome-wide significant loci in chromosomes 15 ( $n$  loci = 5), 9 ( $n$  = 1), 8 ( $n$  = 1), and X ( $n$  = 1); three of the loci contained two independent signals (Fig. 6 and Supplementary Data 9). To know how many of these loci had been previously reported in European ancestry-based GWAS of cig per day, we studied the association of these eight loci with cig per day in the UKB ( $N$  = 133,133).

Three of the eight loci identified in the MCPS were also genome-wide significant in the UKB, all of which had been reported previously<sup>10,11</sup>. These include loci 15q25.1, 9q34.2, and Xq26.3.

At the locus 15q25.1, we found two independent signals in the MCPS within the *CHRNA3*-*AS*-*B4* cluster. At the primary signal, the index variant (rs55781567; beta = 0.14 SD units;  $P$  = 1.2e-50) was in high LD with the top variant at the primary signal in UKB (rs55853698;  $r^2$  = 0.90;  $P$  = 1.1e-164), the top variant at this locus reported by Thorgerirsson et al. (ref. 15) in 2010 (rs1051730;  $r^2$  = 0.86;  $P$  = 2e-69) and the well-established *CHRNA5* missense variant<sup>16</sup> (rs16969968;  $r^2$  = 0.84;  $P$  = 7e-46). At the secondary signal, the index variant (rs138804289; beta = -0.09;  $P$  = 1e-38) was in high LD with the top variant at the secondary signal in UKB (rs13329271;  $r^2$  = 0.95) and the top variant (rs2456020;  $r^2$  = 0.80;  $P$  = 1e-26) reported in the AMR-based GWAS of cig per day by Saunders et al. (ref. 11).

At the locus 9q34.2, we found two independent signals in the MCPS close to genes *FAMI63B* and *DBH*. At the primary signal, the index variant (rs3174179; beta = 0.10;  $P$  = 3.7e-9) was in high LD with the top variant at the single signal at this locus in UKB (rs56116178;  $r^2$  = 0.72;  $P$  = 1.3e-10) and the top variant at one of the two signals reported in the previous European ancestry-based GWAS of cig per day by Saunders et al. (ref. 11) (rs113067637;  $r^2$  = 0.93;  $P$  = 7.7e-12). The second signal we identified at this locus appears to be novel as the index variant, rs3025336 (beta = -0.04;  $P$  = 2.7e-9), located within an intron of *FAMI63B*, is not in LD with either the top variant in the UKB or any of the top variants reported by Saunders et al. (ref. 11) ( $r^2$  < 0.02). Notably, rs3025336 had an MAF of 28% in Indigenous Mexican ancestry but only 4% in European ancestry, hence offering higher statistical power for discovery in the MCPS. Also, rs3025336 replicates in the UKB at nominal significance (beta = -0.03;  $P$  = 5.5e-4), suggesting a genuine association. Hence, by discovering an additional new signal, we further strengthen the association of this locus with smoking. Although the causal gene at this locus is still unclear, traditionally *DBH* (encoding dopamine beta-hydroxylase) has been viewed as causal due to its role in dopamine metabolism<sup>21</sup>. However, *FAMI63B* (the closest gene to the new signal), encoding a protein of unknown function expressed densely in the brain<sup>23</sup>, warrants further investigation for its potential involvement in smoking behavior (Supplementary Fig. 18).

At the locus Xq26.3, we found one independent signal in the MCPS close to genes *RBMX* and *GPR101*. The index variant at this locus (rs34269802; beta = -0.05;  $P$  = 3.9e-11) was in high LD with the top variant at this locus in the UKB (rs55958410;  $r^2$  = 0.79;  $P$  = 1.6e-23) and the top variant reported previously by Erzurumluoglu et al. (ref. 24) in 2019 (rs1190736;  $r^2$  = 0.79;  $P$  = 6.4e-10). Notably, Xq26.3 is the third strongest locus in the GWAS of cig per day in European ancestry groups after the *CYP2A6* and *CHRNA5* loci on chromosomes 19 and 15, respectively. Yet, it was not identified until 2019 as most early GWASs of smoking behavior did not include analyses of the X chromosome. The causal gene at this locus is not clear, but *GPR101*, encoding a brain-specific G protein-coupled transmembrane receptor expressed



**Fig. 6 | GWAS Manhattan plots.** Manhattan plots showing results from genome-wide common variant association studies of cigarettes per day in the MCPS (a) and UKB (b). The X axes correspond to chromosomal base pair start positions of individual variants and Y axes correspond to  $P$  values (expressed in negative log scale) from variant association tests performed using REGENIE. The dotted line

corresponds to the genome-wide statistical significance threshold of  $P = 5e-8$ . UKB UK Biobank, MCPS Mexico City Prospective Study. Statistical test used: linear regression. Two-sided tests with no multiple tests adjustment, beyond genome-wide significant threshold.

densely in the nucleus accumbens<sup>23</sup>, a key brain region involved in the reward pathway, is a prime candidate (Supplementary Fig. 19).

Five of the eight loci we identified in the MCPS were not genome-wide significant in the UKB or reported previously. Four of these five new loci were located on chromosome 15 adjacent to the *CHRNA5* cluster. Notably, the locus 15q24.1 contains multiple genes encoding the cytochrome P450 superfamily of enzymes, namely *CYP1A1*, *CYP1A2*, and *CYP11A1*. The index variant rs12437562 (intergenic variant) was significantly associated with cig per day in the MCPS but not in the UKB. The MAF of rs12437562 was more than 50 times enriched in Indigenous Mexican ancestry compared to European ancestry. The lack of association in European ancestry is likely due to the low statistical power, as rs12437562 was rare in the UKB. This chromosome 15 cluster of cytochrome genes is distinct from the chromosome 19 cluster of cytochrome genes, such as *CYP2A6*, *CYP2A7*, etc., classically associated with smoking in the European ancestry-based studies<sup>15</sup>. Importantly, we found no genetic signal for cig per day in the *CYP2A6* locus in the Indigenous Mexican ancestry group (Fig. 6). For example, the well-characterized *CYP2A6* null missense variant Leu160His (ref. 5) associated with lower cig per day in European ancestry in the UKB (beta = -0.17;  $P = 8.8 \times 10^{-57}$ , MAF = 2.8%) but had no association in Indigenous Mexican ancestry in the MCPS (beta = -0.03;  $P = 0.29$ ; MAF = 1.2%) despite having similar allele frequencies in both populations. The distinct genetic signals of cytochrome enzymes associated with cig per day in European and Indigenous Mexican ancestries warrant further investigation concerning nicotine metabolism.

Altogether, five of the eight loci-associated cig per day were clustered on chromosome 15 across a five Mbp region. Given the high density of smoking signals in the long arm of chromosome 15 in the Indigenous American ancestry group, we explored the admixture pattern in the associated regions<sup>18</sup>. We did not find significant enrichment of specific ancestry groups in the chromosome 15 signals (Supplementary Fig. 20).

## Discussion

Individuals of non-European ancestries have been historically under-represented in human genomic studies<sup>25</sup>. Fortunately, this diversity gap is slowly being filled by researchers from both academia and industry<sup>11,18</sup>. Here we have provided the first look at the rare-variant architecture of cig per day in Indigenous Mexicans, revealing a notable discovery—the p.Glu284Gly missense variant in *CHRNA5* associated with a strong reduction in the number of cigarettes smoked per day. Using a haplotype-resolved local ancestry approach, we find the MAF of this variant in Indigenous Mexicans is -1.4%, based on which we estimate 2.8% and 0.02% of individuals of this population group are heterozygous and homozygous carriers, respectively. Based on recent population estimates, this would correspond to ~400,000 and ~3000 individuals, respectively. Although local ancestry analysis suggested that p.Glu284Gly occurred exclusively in the Indigenous Mexican haplotypes, based on the admixture-based selection analysis of the MCPS cohort reported previously, we did not find an over-representation of Indigenous Mexican ancestry in the region containing *CHRNA5* to support purifying selection (Supplementary Fig. 21). Hence, the variant may have drifted in time to a higher frequency, benefitting its carriers by guarding them from smoking more cigarettes.

When investigating the *CHRNA5* coding variants, we found a drifted pLOF in the East Asian participants in the Japan Biobank<sup>21</sup>. The effect size of this pLOF was remarkably similar to the effect size of p.Glu284Gly, which is 0.2 SD fewer cigarettes per minor allele. This corresponds to approximately 1.3 cigarettes per day in the current smokers in the MCPS cohort who smoked, on average, 5.8 (SD = 6.4) cigarettes per day and 1.7 cigarettes per day in the current smokers in the UKB who smoked, on average, 15.5 cigarettes per day (SD = 8.4). Since the mean and SD of cig per day can vary across ancestries, we

preferred SD units over actual numbers of cig per day as the standard measure of effect size in our reporting. Benchmarking this effect size against well-established cig per day coding variants, the per-allele effect size of p.Glu284Gly is similar to the per-allele effect size of *CYP2A6* null missense variant (p.Leu160His)<sup>5</sup> and 1.7 times larger than the *CHRNA5* missense variant (p.Asp398Asn)<sup>16</sup>. Importantly, the effect size in p.Glu284Gly homozygotes appears to be much larger. Based on the similar heterozygote-effect size of p.Glu284Gly compared to the EAS-specific pLOF (c.52+1 G > A), we suspect that p.Glu284Gly has a loss-of-function consequence, hence p.Glu284Gly homozygotes are likely *CHRNA5* knockouts. Our findings will inspire future call-back studies in Mexico to deeply profile the phenotypic characteristics of *CHRNA5* human knockouts. Such studies will provide insights into the safety and viability of *CHRNA5* as a therapeutic target for smoking cessation.

Understanding the mechanism through which loss of the  $\beta 3$  subunit associates with reduced smoking is important for drug development. Unlike the  $\beta 2$  subunit, which has been extensively studied using gene manipulation in animal models for many decades, the  $\beta 3$  subunit is relatively poorly studied<sup>26</sup>. Our data show distinct *CHRNA5* genetic associations with smoking behavior. We observed that rare variants in *CHRNA5* were significantly associated with reduced cig per day but were not associated with risk of ever smoking (vs never). This phenotypic pattern differs from what we previously reported for *CHRNA2*, where rare variants were associated with decreased risk of ever smoking (vs never) but not with reduced cig per day<sup>7</sup>. These differential effects align with knowledge based on animal studies that nAChR subunits contribute to distinct aspects of smoking behavior<sup>8,17</sup>.

The observed pattern for *CHRNA5* inversely mirrors what has been documented for *CHRNA5*, where the hypomorphic p.Asp398Asn variant associated with increased cig per day (effect = 0.18 SD;  $P = 2.3 \times 10^{-200}$ ) but has minimal effect on ever smoking status (odds ratio = 0.99;  $P = 0.01$ ) (variant associations statistics from publicly available GWAS summary statistics from the largest GWAS of cig per day and smoking initiation to date, published by Saunders et al. (ref. 11)). This differential effect was experimentally validated by Fowler et al. (ref. 17), showing that deleting the  $\alpha 5$  subunit, expressed densely in medial habenula, in mouse predominantly affected nicotine consumption quantity rather than initial reinforcement<sup>17</sup>. Similarly, for *CHRNA3*, Jackson et al. (ref. 27) reported that  $\beta 3$  deletion in mice did not affect nicotine reward but influenced withdrawal<sup>27</sup>. Interestingly, like  $\alpha 5$ ,  $\beta 3$  is also expressed densely in medial habenula, a brain region critical for nicotine consumption regulation<sup>28-30</sup>.

These findings suggest that targeting *CHRNA5* may contribute to smoking cessation through mechanisms distinct from those targeted by current pharmacotherapy, which is focused on  $\beta 2$ -containing receptors<sup>31</sup>. The current first-line pharmacotherapy, varenicline, primarily acts as a partial  $\alpha 4\beta 2$  agonist/antagonist<sup>32</sup>. Given the selective effect of *CHRNA5* variants on cig per day rather than smoking initiation, and its expression in circuits implicated in aversion,  $\beta 3$ -containing receptors may represent a complementary therapeutic target to existing approaches for reducing smoking intensity in established smokers.

Our study has several limitations. Although we replicated the findings in former smokers from the MCPS cohort and found overall consistent associations of *CHRNA5* rare variants with cig per day across different ancestries, we lacked an independent Mexican ancestry cohort for direct replication of the p.Glu284Gly variant. Our phenotypic characterization focused primarily on cig per day rather than comprehensive clinical measures of nicotine dependence or tobacco use disorder, potentially limiting the direct assessment of clinical implications of the genetic findings. While the genetic associations were robust, functional characterization of *CHRNA5* rare variants was not performed to precisely understand the molecular

mechanisms of through which loss of  $\beta 3$  subunit influences smoking behavior. Finally, the population-specific nature of the key variants (p.Glu284Gly in Indigenous Mexicans, c.52+1G>A in East Asians) limits replication of specific variant associations across ancestries. However, the consistent phenotypic impact of these different variants from diverse ancestries strengthens the overall conclusion that disrupting *CHRNA3* function associates with reduced cigarette consumption. This demonstrates that, although genetic variants diverge, the underlying biology, and therefore the translational impact, often converges across ancestries.

In summary, by combining genetic data from diverse ancestries, we have shown that naturally occurring deleterious coding variants in *CHRNA3* strongly associate with reduced daily cigarette smoking in humans. The finding adds *CHRNA3* to the portfolio of human genetics-validated, potential drug targets for smoking.

## Methods

### Study cohorts

The MCPS was established in the late 1990s through collaboration between Mexican and British scientists to investigate health effects of tobacco and other factors in Mexico<sup>18,19</sup>. From 1998 to 2004, over 150,000 adults aged 35 and older (mean age 50) were recruited from two historic Mexico City districts: Coyoacán and Iztapalapa. Participants completed questionnaires, underwent physical measurements, provided blood samples, and agreed to mortality follow-up. The study areas have diverse populations, including long-term residents and more recent migrants from various regions of Mexico. The MCPS represents one of the largest prospective cohort studies in Latin America and provides valuable data on disease risk factors in the Mexican population.

UKB is a large-scale biomedical database and research resource containing in-depth genetic and health information from half a million UK participants<sup>33,34</sup>. The project was established in 2006–2010, recruiting 500,000 volunteers aged 40–69 from across the UK. Participants provided blood, urine, and saliva samples for future analysis, underwent various physical measurements, and answered questions about their lifestyle, environment, and medical history. UKB enables research into the genetic and environmental determinants of common diseases of middle and old age. The resource is made available to all bona fide researchers for health-related research that is in the public interest, without preferential access.

The Biobank Japan (BJ) Project was launched in 2003 in Japan with the aim of developing personalized medicine through the creation of a large patient-based biobank of East Asian populations<sup>21,35</sup>. The study enrolled approximately 200,000 patients with any of 47 target common diseases between 2003 and 2008, collected DNA, serum samples, and detailed clinical information from participants, conducted annual follow-up surveys, stored biological samples and clinical data to enable genomic and biomarker research. The project combines genomic data with detailed clinical information to study the genetic and environmental factors influencing common diseases in the Japanese population. It serves as a valuable resource for researchers investigating the genetic basis of diseases and drug responses.

### Smoking phenotypes

The smoking phenotypes used in the current study were self-reported, collected via interviewer-administered questionnaire (MCPS) or touchscreen/online follow-up questionnaire (UKB). In the MCPS, participants reported if they had ever smoked tobacco (ever smoker); if they currently regularly smoke tobacco (current smoker); among current smokers, number of cigarettes smoked per day, on the days when participant smokes (cig per day), among participants who ever quit smoking, number of cigarettes smoked per day, on a day when participant used to smoke (cig per day). In the UKB, participants reported about their current (data field: 1239) and past tobacco

smoking (data field: 1249); among current smokers, number of cigarettes smoked per day (data field: 3456); among former smokers, number of cigarettes smoked per day (data field: 2887). In both MCPS and UKB, heavy smokers were defined as those who smoked 10 or more cigarettes per day.

### Exome sequencing

Exome sequencing for all cohorts was conducted at the Regeneron Genetics Center in Tarrytown in the US using standardized protocols<sup>20</sup>. DNA from peripheral blood was fragmented into 200-bp libraries, barcoded, and enriched for exome regions using a modified xGen probe. Sequencing was performed on Illumina NovaSeq 6000 platforms, achieving >20 $\times$  coverage for over 90% of target sequences in 99% of samples. Reads were aligned to the hg38 reference genome using BWA-MEM. Variant calling was done using WeCall to generate individual gVCF files. Quality control removed samples with low coverage, excess heterozygosity, sex mismatches, and genetic duplicates. High-quality gVCFs were merged into a project-level VCF (pVCF) using GLnexus. Further filtering removed SNVs with read depth <7 and indels with depth <10. Variants were required to have at least one homozygous genotype or a heterozygous genotype with allele balance  $\geq 0.15$  ( $\geq 0.20$  for indels). The final pVCF files were converted to PGEN format using PLINK version 2 for analysis.

### Variant annotation

The SnpEff tool (v5.3) was used to annotate variants identified through exome sequencing. Annotations were based on the most severe consequence in the canonical transcript with defined start and end positions, according to Ensembl release 85. Putative loss-of-function (pLOF) variants were defined as those with non-ancestral alleles annotated as stop gain, start lost, splice donor, splice acceptor, stop lost, or frameshift. To assess missense variant deleteriousness, five prediction algorithms were employed: SIFT(v6.2.1)<sup>36</sup>, PolyPhen-2 HDIV(v2)<sup>37</sup>, PolyPhen-2 HVAR (v2)<sup>37</sup>, LRT<sup>38</sup>, and MutationTaster(GRCh38)<sup>39</sup>. Missense variants were grouped into three categories: all missense variants irrespective of their deleteriousness prediction, missense (1/5) variants, if predicted as deleterious by at least one of the five algorithms and missense (5/5) variants, if predicted as deleterious by all the five algorithms.

### Genotyping and imputation

Genotyping methods varied across cohorts, as detailed in cohort-specific publications<sup>18,34</sup>. The UKB participants were genotyped using the Applied Biosystems UK BiLEVE Axiom Array or UKB Axiom Array and the MCPS participants were genotyped using the Global Screening Array. After standard quality control, common variants were imputed using the TOPMed LD reference panel via the TOPMed Imputation Server (release-3)<sup>40</sup>. For analysis, only variants with MAF > 0.01 and imputation INFO score > 0.8 were retained. Totally, 9,835,118 and 10,343,511 variants were analyzed in the MCPS and UKB cohorts, respectively. To account for potential batch effects, variables representing genotyping arrays and imputation batches were included as covariates in all analyses of imputed variants.

### Genetic ancestry inference

The genetic ancestry of participants was determined using a set of common variants directly genotyped on the arrays. The process involved: (a) Computing principal components (PCs) for HapMap3 individuals using a public genotype reference panel<sup>41</sup>; (b) Selecting high-confidence variants (MAF > 0.10, genotype missingness < 5%, Hardy-Weinberg equilibrium  $P > 1 \times 10^{-5}$ ) common to both the study dataset and HapMap3; (c) Projecting study samples onto PCs computed from HapMap3 samples; (d) Assigning individuals to one of five ancestral groups (EUR, African, AMR, East Asian, South Asian) if their likelihood of belonging to that ancestry exceeded 0.3; (e) Calculating likelihood estimates using a kernel density estimator trained on

HapMap3 PCs. This approach allowed for consistent ancestry classification across all participating cohorts.

### Fine-scale and local genetic ancestry inference in MCPS

Fine-scale and local ancestry analyses are described in detail in the MCPS flagship publication<sup>18</sup>. In brief, ancestry analysis in the MCPS cohort utilized both fine-scale haplotype-based methods and local ancestry inference. Two complementary haplotype-based approaches were applied: one using identical-by-descent (IBD) segments identified by hap-IBD software (v1), and another leveraging a scalable haplotype-copying hidden Markov model implemented in Impute5. These methods produced low-dimensional representations with distinct “star-like” structures, effectively delineating Mesoamerican and European ancestral contributions. Local ancestry inference was performed using RFMix, which estimated that 66.0% of autosomal ancestry was attributable to Indigenous Mexican populations, with the majority (35.6%) coming from central Mexico, followed by southern (15.9%), southeastern (11.8%), northern (1.6%), and northwestern (1.1%) parts of Mexico. European and African ancestries accounted for 31.1% and 2.9%, respectively. The combined results from these approaches revealed that the Indigenous American ancestry in the MCPS cohort predominantly derived from Mesoamerican populations in central, southern, and southeastern Mexico. This multi-faceted analysis provided a high-resolution view of the complex admixture patterns in this cohort, offering greater insight than broad continental classifications alone.

### Genetic association analysis

Genetic association analyses were conducted separately for each cohort using the REGENIE software (v3.2.1)<sup>42</sup>. REGENIE employs a two-step whole-genome regression framework that efficiently controls for population stratification and sample relatedness. In the first step, REGENIE calculates trait-prediction values (local PGS) using a sparse set of genotypes, typically from array data. The second step involves computing variant associations with phenotypes using either logistic or linear regression. This step incorporates the trait-prediction values from step one as covariates, along with additional covariates including the first 20 genetic PCs from common variants, the first 20 genetic PCs from rare variants, age, age squared, sex, an age-sex interaction term, and genotyping batch indicators. In burden analysis, REGENIE creates a pseudo-genotype (burden mask) by collapsing a set of variants into a single categorical variable, which is then treated as a variant genotype for association statistics computation. Fine mapping of common variants was performed using the sum of single effects (SuSiE, v0.12.10) statistical approach<sup>43</sup>. All rare-variant association analyses were adjusted for nearby common variant signals by including all the fine-mapped index variant genotypes as covariates in the association runs.

### Gene *P* calculation

We employed the GENE-P approach<sup>22</sup> to assess gene-level associations, which combines multiple tests of genetic association within a gene into a single *p*-value as illustrated in the Supplementary Fig. 2. This method classifies variants into different groups based on variant annotations (pLOF, all missense, missense (1/5), missense (5/5)) and allele frequency thresholds (1%, 0.05%, 0.01%, 0.001%, and singletons) and applies four main testing strategies: BURDEN (classic burden test across allele frequency bins), SBAT (Set-Based Association Test assuming same effect direction), SKATO (Sequence Kernel Association Test Optimal for mixed effect directions), and ACATV (Aggregated Cauchy Association Test for Variants). The BURDEN test produced 35 *p*-values (7 masks × 5 AF bins), SBAT yielded 2 *p*-values for positive and negative direction, while SKATO and ACATV each generated 7 *p*-values (7 masks × 1% bin). The Aggregated Cauchy Association Test (ACAT) was then used to combine *p*-values within each approach. Finally, the four ACAT-combined *p*-values from BURDEN, SBAT, SKATO, and

ACATV were further combined using ACAT to produce a single GENE-P value, providing a comprehensive assessment of genetic associations within each gene.

### Japan Biobank analysis

We queried the association of rs147306385 in the Pheweb portal hosted by the Biobank Japan (BBJ) Project<sup>21</sup>. The query returned associations with 287 phenotypes including smoking-related phenotypes. The results were based on prior GWAS conducted on the ~180,000 BBJ participants using either SAIGE<sup>44</sup> for binary traits or BOLT-LMM<sup>45</sup> for quantitative traits. The top association observed was cig per day based on a GWAS in 74,893 individuals. The downloaded summary statistics contained information on effect size, *P* value, MAF, and total number of samples studied. The genotype-specific counts were calculated based on the MAF, assuming no deviation from Hardy Weinberg Equilibrium.

### Ethical approval and informed consent

All study participants have provided informed consent, and all participating cohorts have received ethical approval from their respective institutional review board. The MCPS has received ethical approval from the Mexican Ministry of Health, the Mexican National Council for Science and Technology, the UNAM, and the University of Oxford<sup>18,19</sup>. The UKB project has received ethical approval from the Northwest Centre for Research Ethics Committee (11/NW/0382)<sup>20,34</sup>. The work described here has been approved by the UKB (application no. 26041)<sup>20</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data supporting the findings of this manuscript are reported in the main text, figures, and Supplementary Data. Genome-wide association summary statistics for common and rare variants are also available for download at <https://doi.org/10.6084/m9.figshare.30840641>. UKB individual-level genotypic and phenotypic data are available to approved investigators via the UK Biobank study [[www.ukbiobank.ac.uk/](http://www.ukbiobank.ac.uk/)]. Additional information about registration for access to the data are available at [www.ukbiobank.ac.uk/register-apply/](http://www.ukbiobank.ac.uk/register-apply/). Data access for approved applications requires a data transfer agreement between the researcher's institution and UK Biobank, the terms of which are available on the UK Biobank website [[www.ukbiobank.ac.uk/media/ezrderzw/applicant-mta.pdf](http://www.ukbiobank.ac.uk/media/ezrderzw/applicant-mta.pdf)]. The MCPS represents a long-standing collaboration between researchers at the National Autonomous University of Mexico (UNAM) and the University of Oxford. The investigators welcome requests from researchers in Mexico and elsewhere who wish to access MCPS data. If you are interested in obtaining data from the study for research purposes, or in collaborating with MCPS investigators on a specific research proposal, please visit <https://www.ctsu.ox.ac.uk/research/mcps>, where you can download the study's Data and Sample Access Policy in English or Spanish. The policy lists the data available for sharing with researchers in Mexico and in other parts of the world. Full details of the data available may also be viewed at <https://datashare.ndph.ox.ac.uk/>. MCPS ancestry-specific allele frequencies are available at <https://rgc-mcps.regeneron.com>. Data supporting the findings related to Japan Biobank is freely available at <https://pheweb.jp/>. The results reported in this manuscript in Supplementary Fig. 12 can be accessed by querying the server using variant identifier: rs147306385.

### Code availability

All the genetic association analyses were performed using the REGENIE software v.3.2.1 developed in-house. The REGENIE software is freely

available at GitHub [<https://github.com/rgcgithub/regenie>] and Zenodo [<https://doi.org/10.5281/zenodo.6789126>].

## References

- Reitsma, M. B. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet* **397**, 2337–2360 (2021).
- Lengel, D. & Kenny, P. J. New medications development for smoking cessation. *Addict. Neurosci.* **7**, 100103 (2023).
- Trajanoska, K. From target discovery to clinical drug development with human genetics. *Nature* **620**, 737–745 (2023).
- Wills, L. Neurobiological mechanisms of nicotine reward and aversion. *Pharmacol. Rev.* **74**, 271–310 (2022).
- Pianezza, M. L., Sellers, E. M. & Tyndale, R. F. Nicotine metabolism defect reduces smoking. *Nature* **393**, 750–750 (1998).
- Oscarson, M. Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. *FEBS Lett.* **448**, 105–110 (1999).
- Rajagopal, V. M. Rare coding variants in CHRN2 reduce the likelihood of smoking. *Nat. Genet.* **55**, 1138–1148 (2023).
- Piccio, M. R. Acetylcholine receptors containing the  $\beta 2$  subunit are involved in the reinforcing properties of nicotine. *Nature* **391**, 173–177 (1998).
- Dani, J. A. & Bertrand, D. Nicotinic acetylcholine receptors and nicotinic cholinergic mechanisms of the central nervous system. *Annu. Rev. Pharmacol. Toxicol.* **47**, 699–729 (2007).
- Liu, M. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
- Saunders, G. R. B. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720–724 (2022).
- Toikumo, S. et al. Multi-ancestry meta-analysis of tobacco use disorder identifies 461 potential risk genes and reveals associations with multiple health outcomes. *Nat. Hum. Behav.* **8**, 1177–1193 (2024).
- Quach, B. C. et al. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nat. Commun.* **11**, 5562 (2020).
- Ware, J. J., van den Bree, M. & Munafò, M. R. From men to mice: CHRNA5/CHRNA3, smoking behavior and disease. *Nicotine Tob. Res.* **14**, 1291–1299 (2012).
- Thorgerirsson, T. E. Sequence variants at CHRN3–CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42**, 448–453 (2010).
- Berrettini, W. H. & Doyle, G. A. The CHRNA5–A3–B4 gene cluster in nicotine addiction. *Mol. Psychiatry* **17**, 856–866 (2012).
- Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J. & Kenny, P. J. Habenular  $\alpha 5$  nicotinic receptor subunit signalling controls nicotine intake. *Nature* **471**, 597–601 (2011).
- Ziyatdinov, A. Genotyping, sequencing and analysis of 140,000 adults from Mexico City. *Nature* **622**, 784–793 (2023).
- Tapia-Conyer, R. Cohort profile: the Mexico City prospective study. *Int. J. Epidemiol.* **35**, 243–249 (2006).
- Backman, J. D. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
- Sakaue, S. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
- Ziyatdinov, A. et al. Joint testing of rare variant burden scores using non-negative least squares. *Am. J. Hum. Genet.* **111**, 2139–2149 (2024).
- Lonsdale, J. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
- Understanding Society Scientific, G., Epic, C. V. D., Gscan, Consortium for Genetics of Smoking, B. & consortium, C. H. D. E. Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. *Mol. Psychiatry* **25**, 2392–2409 (2020).
- Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
- Piccio, M. R. & Kenny, P. J. Mechanisms of nicotine addiction. *Cold Spring Harb. Perspect. Med.* **11**, a039610 (2021).
- Jackson, A. B., Toma, W., Contreras, K. M., Alkhlaif, Y. & Damaj, M. I. The  $\beta 3$  subunit of the nicotinic acetylcholine receptor is required for nicotine withdrawal-induced affective but not physical signs or nicotine reward in mice. *Pharmacol. Biochem. Behav.* **183**, 1–5 (2019).
- Boulos, L. J., Darcq, E. & Kieffer, B. L. Transcriptomic characterization of the human habenula highlights drug metabolism and the neuroimmune system. *Front. Neurosci.* <https://doi.org/10.3389/fnins.2018.00742> (2018).
- Shih, P. Y. Differential expression and function of nicotinic acetylcholine receptors in subdivisions of medial habenula. *J. Neurosci.* **34**, 9789–9802 (2014).
- Quik, M. Localization of nicotinic receptor subunit mRNAs in monkey brain by in situ hybridization. *J. Comp. Neurol.* **425**, 58–69 (2000).
- Jordan, C. J. & Xi, Z. X. Discovery and development of varenicline for smoking cessation. *Expert Opin. Drug Discov.* **13**, 671–683 (2018).
- Coe, J. W. Varenicline: an  $\alpha 4\beta 2$  nicotinic receptor partial agonist for smoking cessation. *J. Med. Chem.* <https://doi.org/10.1021/jm050069n> (2005).
- Sudlow, C. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- Nagai, A. Overview of the BioBank Japan project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
- Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
- Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* <https://pubmed.ncbi.nlm.nih.gov/23315928/> (2013).
- Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
- Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
- Consortium, N. T.-O. fP. M. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* **590**, 290–299 (2021).
- The International HapMap, C. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
- Mbatchou, J. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
- Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet.* **18**, e1010299 (2022).
- Zhou, W. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
- Loh, P. R. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

## Acknowledgements

We thank the UK Biobank team, their funders, the dedicated professionals from the member institutions who contributed to and supported this work, and the UK Biobank participants. The exome sequencing was funded by the UK Biobank Exome Sequencing Consortium (i.e., Bristol Myers Squibb, Regeneron, Biogen, Takeda, AbbVie, Alnylam,

AstraZeneca, and Pfizer). This research has been conducted using the UK Biobank Resource under application number 2604. We thank the participants of the MCPS cohort. The MCPS has received funding from the Mexican Health Ministry, the National Council of Science and Technology for Mexico, the Wellcome Trust (058299/Z/99), Cancer Research UK, British Heart Foundation, and the UK Medical Research Council (MC\_UU\_00017/2, MR/Z504543/1). These funding sources had no role in the design, conduct, or analysis of the study or the decision to submit the manuscript for publication. Genotyping, exome sequencing and whole genome sequencing was funded through an academic partnership between the National Autonomous University of Mexico, the University of Oxford, Regeneron, AstraZeneca, and Abbvie. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the UK Department of Health. We thank the participants and investigators of the Biobank Japan study.

## Author contributions

All authors reviewed the manuscript for important intellectual content and approved the manuscript submitted for publication. Conceptualization: V.M.R., J.Marchini, G.C. Genetic analysis: V.M.R., A.Z., T.J., A.A., J.Mbatchou, Y.Z., J.O., G.A., J.Marchini. Phenotype preparation and harmonization: V.M.R., A.J.A., N.B., M.C., J.T., J.R.E., J.Marchini. Statistical methods development: A.Z., T.J., J.Mbatchou, Y.Z., G.A., J.Marchini. Analytical pipeline development: V.M.R., A.Z., T.J., A.A., J.M.<sup>1</sup>, Y.Z., A.J.A., N.B., J.T., J.O., G.A. Clinical development: M.A., O.H. Data curation: M.A., M.C., J.T., J.O., O.H., G.A., J.B., P.K.-M., J.A.-D., R.T.-C., R.C., J.R.E., J.Marchini. Funding acquisition: A.B., J.B., P.K.-M., J.A.-D., R.T.-C., R.C., J.R.E., J.Marchini. Project administration: M.J., E.C., J.R.V. Supervision: L.L., G.A., A.B., P.K.-M., J.A.-D., R.T.-C., R.C., J.R.E., J.Marchini, G.C. Writing—original draft: V.M.R. and G.C. All authors: Members of the Regeneron Genetics Center and collaborating institutions contributed to securing the funding, study design and oversight; reviewed the final version of the manuscript; performed and were responsible for sample genotyping and exome sequencing; conceived and were responsible for laboratory automation, sample tracking and the library information management system; were responsible for the development and validation of the clinical phenotypes used to identify study participants and (when applicable) controls; performed and were responsible for the analysis needed to produce exome and genotype data; provided the computing infrastructure development and operational support; provided variant and gene annotations and their functional interpretation of variants and conceived and were responsible for creating, developing and deploying the analysis platforms and computational methods used to analyze the genomic data; developed the statistical analysis plans; contributed to the quality control of the genotype and phenotype files and the generation of the analysis-ready datasets; developed the statistical genetics pipelines and tools and use thereof in the generation of the association results; contributed to the quality control of the review and the interpretation of the results and generated and formatted the results to create the manuscript figures; contributed to the development of the study design and analysis plans

and the quality control of the phenotype definitions; quality-controlled, reviewed and interpreted the association results; contributed to the management and coordination of all research activities, planning and execution and managed the review of the project.

## Competing interests

V.M.R., A.Z., T.J., A.A., M.A., J.Mbatchou, Y.Z., A.J.A., N.B., M.C., E.C., J.R.V., M.J., J.O., O.H., L.L., G.A., A.B., J.Marchini and G.C. are current or former employees and/or stockholders of Regeneron Pharmaceuticals. J.R.E. declares grants to the University of Oxford from Regeneron and AstraZeneca. The other authors declare no competing interests. V.M.R., G.C., A.B., J.Marchini, and A.Z. are inventors on provisional patent application (63/593,164) and US non-provisional application (18/809,138) relating to CHOLINERGIC RECEPTOR NICOTINIC BETA 3 SUB-UNIT (*CHRN3*) genetics.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68825-2>.

**Correspondence** and requests for materials should be addressed to Veera M. Rajagopal or Giovanni Coppola.

**Peer review information** *Nature Communications* thanks Paul Kenny, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2026

---

## Regeneron Genetics Center

Aris Baras<sup>1</sup>, Gonçalo Abecasis<sup>1</sup>, Adolfo Ferrando<sup>1</sup>, Giovanni Coppola<sup>1</sup>✉, Andrew Deubler<sup>1</sup>, Luca A. Lotta<sup>1</sup>, John D. Overton<sup>1</sup>, Jeffrey G. Reid<sup>1</sup>, Alan Shuldiner<sup>1</sup>, Katherine Siminovitch<sup>1</sup>, Jason Portnoy<sup>1</sup>, Marcus B. Jones<sup>1</sup>, Lyndon Mitnaul<sup>1</sup>, Alison Fenney<sup>1</sup>, Jonathan Marchini<sup>1</sup>, Manuel Allen Revez Ferreira<sup>1</sup>, Maya Ghousaini<sup>1</sup>, Mona Nafde<sup>1</sup>, William Salerno<sup>1</sup>, Cristen J. Willer<sup>1</sup>, Lourdes Crane<sup>1</sup>, Christina Beechert<sup>1</sup>, Erin Fuller<sup>1</sup>, Laura M. Cremona<sup>1</sup>, Eugene Kalyuskin<sup>1</sup>, Hang Du<sup>1</sup>, Caitlin Forsythe<sup>1</sup>, Zhenhua Gu<sup>1</sup>, Kristy Guevara<sup>1</sup>, Michael Lattari<sup>1</sup>, Alexander Lopez<sup>1</sup>, Kia Manoochehri<sup>1</sup>

Prathyusha Challa<sup>1</sup>, Manasi Pradhan<sup>1</sup>, Raymond Reynoso<sup>1</sup>, Ricardo Schiavo<sup>1</sup>, Maria Sotiropoulos Padilla<sup>1</sup>, Chenggu Wang<sup>1</sup>, Sarah E. Wolf<sup>1</sup>, Manan Goyal<sup>1</sup>, George Mitra<sup>1</sup>, Sanjay Sreeram<sup>1</sup>, Rouel Lanche<sup>1</sup>, Vrushali Mahajan<sup>1</sup>, Sai Lakshmi Vasireddy<sup>1</sup>, Gisu Eom<sup>1</sup>, Krishna Pawan Punuru<sup>1</sup>, Sujit Gokhale<sup>1</sup>, Benjamin Sultan<sup>1</sup>, Pooja Mule<sup>1</sup>, Mudasar Sarwar<sup>1</sup>, Muhammad Aqeel<sup>1</sup>, Xiaodong Bai<sup>1</sup>, Lance Zhang<sup>1</sup>, Sean O’Keeffe<sup>1</sup>, Razvan Panea<sup>1</sup>, Evan Edelstein<sup>1</sup>, Ayesha Rasool<sup>1</sup>, Evan K. Maxwell<sup>1</sup>, Boris Boutkov<sup>1</sup>, Alexander Gorovits<sup>1</sup>, Ju Guan<sup>1</sup>, Lukas Habegger<sup>1</sup>, Alicia Hawes<sup>1</sup>, Olga Krasheninina<sup>1</sup>, Samantha Zarate<sup>1</sup>, Adam J. Mansfield<sup>1</sup>, Joshua Backman<sup>1</sup>, Kathy Burch<sup>1</sup>, Adrian Campos<sup>1</sup>, Liron Ganel<sup>1</sup>, Sheila Gaynor<sup>1</sup>, Benjamin Geraghty<sup>1</sup>, Arkopravo Ghosh<sup>1</sup>, Salvador Romero Martinez<sup>1</sup>, Christopher Gillies<sup>1</sup>, Lauren Gurski<sup>1</sup>, Eric Jorgenson<sup>1</sup>, Tyler Joseph<sup>1</sup>, Michael Kessler<sup>1</sup>, Jack Kosmicki<sup>1</sup>, Adam Locke<sup>1</sup>, Priyanka Nakka<sup>1</sup>, Karl Landheer<sup>1</sup>, Olivier Delaneau<sup>1</sup>, Anthony Marcketta<sup>1</sup>, Joelle Mbatchou<sup>1</sup>, Arden Moscati<sup>1</sup>, Anita Pandit<sup>1</sup>, Jonathan Ross<sup>1</sup>, Carlo Sidore<sup>1</sup>, Eli Stahl<sup>1</sup>, Timothy Thornton<sup>1</sup>, Sailaja Vedantam<sup>1</sup>, Rujin Wang<sup>1</sup>, Kuan-Han Wu<sup>1</sup>, Bin Ye<sup>1</sup>, Blair Zhang<sup>1</sup>, Andrey Ziyatdinov<sup>1</sup>, Yuxin Zou<sup>1</sup>, Jingning Zhang<sup>1</sup>, Kyoko Watanabe<sup>1</sup>, Mira Tang<sup>1</sup>, Frank Wendt<sup>1</sup>, Suganthi Balasubramanian<sup>1</sup>, Suying Bao<sup>1</sup>, Kathie Sun<sup>1</sup>, Chuanyi Zhang<sup>1</sup>, Sean Yu<sup>1</sup>, Aaron Zhang<sup>1</sup>, David Corrigan<sup>1</sup>, Dhruv Shidhaye<sup>1</sup>, Chen Wang<sup>1</sup>, Keyrun Adhikari<sup>1</sup>, Alexander Lachmann<sup>1</sup>, Brian Hobbs<sup>1</sup>, Jon Silver<sup>1</sup>, William Palmer<sup>1</sup>, Rita Guerreiro<sup>1</sup>, Amit Joshi<sup>1</sup>, Antoine Baldassari<sup>1</sup>, Sarah Graham<sup>1</sup>, Ernst Mayerhofer<sup>1</sup>, Erola Pairo Castineira<sup>1</sup>, Mary Haas<sup>1</sup>, Niek Verweij<sup>1</sup>, George Hindy<sup>1</sup>, Jonas Bovijn<sup>1</sup>, Tanima De<sup>1</sup>, Luanluan Sun<sup>1</sup>, Olukayode Sosina<sup>1</sup>, Arthur Gilly<sup>1</sup>, Peter Dornbos<sup>1</sup>, Juan Rodriguez-Flores<sup>1</sup>, Moeen Riaz<sup>1</sup>, Manav Kapoor<sup>1</sup>, Gannie Tzoneva<sup>1</sup>, Momodou W. Jallow<sup>1</sup>, Anna Alkelai<sup>1</sup>, Ariane Ayer<sup>1</sup>, Veera Rajagopal<sup>1</sup>, Sahar Gelfman<sup>1</sup>, Vijay Kumar<sup>1</sup>, Jacqueline Otto<sup>1</sup>, Jose Bras<sup>1</sup>, Silvia Alvarez<sup>1</sup>, Jessie Brown<sup>1</sup>, Hossein Khiabani<sup>1</sup>, Joana Revez<sup>1</sup>, Kimberly Skead<sup>1</sup>, Valentina Zavala<sup>1</sup>, Jae Soon Sul<sup>1</sup>, Lei Chen<sup>1</sup>, Sam Choi<sup>1</sup>, Amy Damask<sup>1</sup>, Nan Lin<sup>1</sup>, Charles Paulding<sup>1</sup>, Sameer Malhotra<sup>1</sup>, Joseph Herman<sup>1</sup>, Michelle G. LeBlanc<sup>1</sup>, Nadia Rana<sup>1</sup>, Jennifer Rico-Varela<sup>1</sup>, Jaimee Hernandez<sup>1</sup>, Larizbeth Romero<sup>1</sup>, Ashley Paynter<sup>1</sup>, Randi Schwartz<sup>1</sup>, Jody Hankins<sup>1</sup>, Anna Han<sup>1</sup>, Samuel Hart<sup>1</sup>, Ryan Smith<sup>1</sup>, Ann Perez-Beals<sup>1</sup>, Gina Solari<sup>1</sup>, Johannie Rivera-Picart<sup>1</sup>, Michelle Pagan<sup>1</sup> & Sunilbe Siceron<sup>1</sup>