

Genome-wide association study of varenicline-aided smoking cessation

Kayasha Coley PhD^{1*}, Qingning Wang PhD¹, Richard Packer PhD^{1,2}, Catherine John PhD^{1,2}, Erik Abner PhD³, Kadri Reis PhD³, Estonian Biobank Research Team³, Khaled F. Bedair PhD⁴, Sundararajan Srinivasan PhD⁴, Sara Paciga PhD⁵, Craig Hyde PhD⁵, Robert C. Free PhD^{2,6}, Nicola F. Reeve PhD¹, David J. Shepherd BM BCh¹, Tõnu Esko PhD^{3,7}, Colin Palmer PhD⁴, Ewan Pearson PhD⁴, Anders Malarstig PhD^{8,9}, Martin D. Tobin PhD^{1,2}, Chiara Batini PhD^{1,2*}

¹Department of Population Health Sciences, University of Leicester, Leicester, UK

²University Hospitals of Leicester NHS Trust, Groby Road, Leicester, UK

³Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

⁴Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK

⁵Pfizer Worldwide Research, Development and Medical, USA

⁶School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK

⁷Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

⁹Pfizer Worldwide Research, Development and Medical, Stockholm, Sweden

*Correspondence to KC (kc330@leicester.ac.uk; +44 (0)116 229 7251) and CB (cb334@leicester.ac.uk; +44 (0)116 229 7251)

Abstract

Introduction

Varenicline is an $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist with the highest therapeutic efficacy of any pharmacological smoking cessation aid and a 12-month cessation rate of 26%. Genetic variation may be associated with varenicline response, but to date no genome-wide association studies of varenicline response have been published.

Methods

In this study, we investigated the genetic contribution to varenicline effectiveness using two electronic health record-derived phenotypes. We defined short-term varenicline effectiveness (SVE) and long-term varenicline effectiveness (LVE) by assessing smoking status at 3 and 12 months, respectively, after initiating varenicline treatment. In Stage 1, comprising five European cohort studies, we tested genome-wide associations with SVE (1,405 cases, 2,074 controls) and LVE (1,576 cases, 2,555 controls), defining sentinel variants (the most strongly associated variant within 1 megabase) with p -value $<5 \times 10^{-6}$ to follow up in Stage 2. In Stage 2, we tested association between sentinel variants and comparable smoking cessation endpoints in varenicline randomised controlled trials. We subsequently meta-analysed Stages 1 and 2.

Results

No variants reached genome-wide significance in the meta-analysis. In Stage 1, 10 sentinel variants were associated with SVE and five with LVE at a suggestive significance threshold (p -value $<5 \times 10^{-6}$). None of these sentinels were previously implicated in varenicline-aided smoking cessation or in genetic studies of smoking behaviour.

Conclusions

We provide initial insights into the biological underpinnings of varenicline-aided smoking cessation, through implicating genes involved in various processes, including gene expression, cilium assembly and early-stage development.

Implications

Leveraging electronic health records, we undertook the largest genetic study of varenicline-aided smoking cessation to date, and the only such study to test genome-wide associations. We showed distinct genetic variants associated (p -value $<5 \times 10^{-6}$) with varenicline-aided smoking cessation which implicate diverse cellular functions, including transcriptional regulation, RNA modification and cilium assembly. These provide insights which, if independently corroborated, will improve understanding of varenicline response. The growing availability of biobank resources with genetic and varenicline response data will provide future opportunities for larger studies using the approach we developed.

Keywords

Varenicline; nicotine dependence; smoking cessation; electronic health records; pharmacogenomics; genetic epidemiology.

Introduction

Tobacco smoking is a major risk factor for many diseases including ischemic heart disease, stroke, chronic obstructive pulmonary disease, and lung cancer^{1,2}. Consequently, smoking is a leading cause of preventable morbidity and mortality in the UK³, and results in the premature death of over 74,000 smokers each year in England alone⁴. Despite the consistent decline in smoking prevalence in the UK in the last decade, the latest statistics on smoking report that 12.9% of the adult population were smokers in 2022, equating to approximately 6.4 million people⁵.

The substantial impact of smoking on health outcomes has highlighted the importance of quit attempts^{2,6}, less than 5% of which are successful without any behavioural support or pharmacological intervention⁶. Nicotine replacement therapy (NRT) and varenicline (Champix) have been prescribed as smoking cessation pharmacotherapies in the UK, with NRT also available over-the-counter⁷. Randomised controlled trials (RCTs) have shown the superiority of varenicline compared to NRT monotherapy for smoking cessation up to 12 months (OR [95% CI] = 1.57 [1.29, 1.91])⁸, and the largest observational study based on electronic health records (EHRs) reported 12-month smoking cessation rates of 26.3% for varenicline and 21.2% for NRT (OR [95% CI] = 1.34 [1.31, 1.38])⁹. The number of items of varenicline dispensed in England has decreased year-on-year since 2010-11, with 232,000 items being dispensed in 2020-21¹⁰.

Pharmacogenomic studies aim to identify genetic variants that influence how individuals respond to medications, helping to elucidate the mechanisms of drug activity and the reasons for varying effects and adverse reactions among individuals¹¹. Varenicline is an $\alpha_4\beta_2$ nicotinic acetylcholine receptor (nAChR) partial agonist which stimulates the release of dopamine in the brain, thus reducing nicotine cravings and withdrawal symptoms, while promoting smoking abstinence¹². Several candidate gene association studies, albeit with modest sample sizes, have been performed to investigate the contribution of genetic variation to the success of varenicline-aided quit attempts, and have reported associations with variants in genes encoding nAChR subunits, including *CHRNA4*, *CHRNA5*, *CHRNA7* and *CHRNA2*, as well as drug metabolisers (*CYP2B6*)¹³⁻¹⁶. However, candidate gene studies have been generally underpowered, subject to publication biases and are restricted to known biological pathways. In contrast, genome-wide association studies (GWASs), which survey associations with common genetic variants across the genome, have elucidated new biological knowledge¹⁷ and are increasingly used to study drug response¹⁸. Here we present the largest GWASs conducted to identify genetic variants associated with the success of varenicline-aided quit attempts, to provide insights into the biological processes underlying the most efficacious pharmacological smoking cessation therapy. We utilise varenicline prescription events and smoking status records in EHRs to define short-term and long-term varenicline effectiveness across five European cohort studies, and combine

these results with analyses undertaken in participants from varenicline RCTs conducted by Pfizer in a 2-stage design.

Methods

Stage 1 cohorts

Varenicline effectiveness phenotypes included in Stage 1 analyses were defined in Estonian Biobank¹⁹, the Extended Cohort for E-health, Environment and DNA (EXCEED) Study²⁰, Genetics of Diabetes Audit and Research, Tayside and Scotland (GoDARTS)²¹, Genetics of the NHS Scotland Health Research Register (GoSHARE)²² and UK Biobank²³ (further details in **Supplementary Material**).

Data extraction and phenotype definitions

Varenicline (Champix or Chantix) prescription events, of any pack type, size or dosage, were extracted from prescription EHRs linked to each Stage 1 cohort. The date of the earliest varenicline prescription recorded was defined as the *index date*. With the exception of Estonian Biobank, smoking status was extracted from clinical EHRs using unambiguous Read codes (**Supplementary Table 1**). In Estonian Biobank, smoking status was captured in questionnaire data collected at baseline, due to the limited availability of smoking status records in linked EHRs. Two binary phenotypes were derived from smoking status following an episode of varenicline treatment; cases responded positively to varenicline and quit smoking (i.e. non-smokers), and controls were continuing smokers. For short-term varenicline effectiveness (SVE), smoking status was inferred using the first record between 12 and 104 weeks after *index date*, while for long-term varenicline effectiveness (LVE) we used the first record after 52 weeks from *index date* (**Supplementary Figure 1**). These phenotypes were designed to be consistent with endpoints from previous RCTs and observational studies i.e. at the end of the 12-week standard treatment course, and at 1 year after starting treatment, while allowing sufficient time to capture smoking status due to the irregularity of general practitioner (GP) visitation and recording.

Descriptive characteristics

In the Stage 1 cohort studies, for SVE and LVE, statistical tests were conducted to examine the relationship between case-control status and several variables, including age (at *index date*), sex, mean number of varenicline prescriptions during a 12-week treatment episode after *index date* (UK Biobank and EXCEED only) and Heaviness of Smoking Index (HSI)²⁴ [UK Biobank only]. For categorical variables (sex and HSI), a Chi-square test was used. For continuous variables (age and number of varenicline prescriptions), t-tests were performed.

In UK Biobank, the HSI²⁴ was calculated for SVE and LVE cases and controls. HSI ranges from 0 to 6 and is the sum of two categorical variables:

- (1) Number of cigarettes smoked daily (CPD), where 1-10 is coded as 0, 11-20 as 1, 21-30 as 2, and 31+ as 3.
- (2) Time from waking to first cigarette (TTFC), where 61+ minutes is coded as 0, 31-60 minutes as 1, 6-30 minutes as 2, and ≤ 5 minutes as 3.

For both SVE and LVE, we used a Chi-squared test to compare the number of individuals with low-moderate dependence (HSI score ≤ 4) to those with high dependence (HSI score ≥ 5) within cases and controls to assess differences in nicotine dependence between these groups.

We also calculated the smoking cessation rate, defined as the proportion of individuals whom, after initiating their first varenicline treatment, were defined as a non-smoker, based on the first smoking status record within the outcome time-period for each phenotype definition. The Clopper-Pearson exact test based on the binomial distribution was used to calculate the 95% confidence interval.

Power calculations

We calculated power for conducting GWASs of SVE and LVE utilising cases and controls from the Stage 1 cohorts using the Genetic Association Study (GAS) Power Calculator²⁵. These calculations were based on an additive genetic model, and trait prevalence was defined as the proportion of cases among the total sample size (**Supplementary Figure 2**).

Genome-wide association studies

The 2-stage study design is outlined in **Figure 1**. Genotyping quality control and imputation was performed by each individual cohort following the parameters described in **Supplementary Table 2A**. Within each Stage 1 cohort, a GWAS of each varenicline effectiveness phenotype was performed (cohort-specific methodologies are described in **Supplementary Table 2A**). All analyses included only individuals of European ancestry and utilised imputed genomic data for variants with imputation quality ≥ 0.3 , minor allele frequency (MAF) $\geq 1\%$ and minor allele count (MAC) ≥ 20 . All association testing models utilised imputed dosages and assumed an additive genetic model. Covariates included age (at *index date*), age², sex, up to 10 principal components of genetic ancestry and genotyping array (UK Biobank only). Individuals in GoDARTS genotyped on different arrays were analysed separately and subsequently meta-analysed. All summary statistics were harmonised to genome build hg19 using *liftover*²⁶. For each phenotype, GWAS summary statistics from each Stage 1 cohort were meta-analysed using a fixed effect inverse variance weighted model implemented in METAL²⁷. Genomic control correction was applied. Only variants represented by either UK Biobank only, or at least two studies were included. Using the same methodology, we also performed a sensitivity analysis excluding GoDARTS from Stage 1.

Sentinel selection

We defined sentinel variants using an iterative procedure by selecting the variant with the lowest p -value in Stage 1, excluding all variants within a ± 1 megabase (Mb) region around it, and repeating this step until no variants with a p -value below the suggestive significance threshold (p -value $< 5 \times 10^{-6}$) remained. Open Targets Genetics²⁸ was used to determine the variant consequence and the gene with the highest 'overall Variant-to-Gene (V2G)' score. For variants which were not present in the Open Targets Genetics database, Variant Effect Predictor²⁹ was used to define the consequence and ANNOVAR³⁰ was used to determine the nearest gene. The biological functions of these genes were investigated using online databases (specifically, Open Targets Platform³¹ and GeneCards³²) and literature searches.

Fine-mapping

We fine-mapped each associated locus using the approximate Bayes Factor³³ method to calculate the posterior probability of inclusion (PIP) for each variant. We set the prior W to 0.04 and generated 95% credible sets for each locus by ordering variants by descending PIP, and selecting variants until the cumulative PIP was $\geq 95\%$.

Stage 2 dataset

Stage 2 included studies conducted by Pfizer utilising genetic and phenotypic data collected from nicotine-dependent individuals of European genetic ancestry who participated in RCTs assessing varenicline-related endpoints. Genotyping quality control and imputation was performed by Pfizer following the parameters described in **Supplementary Table 2B**. Following the methods described in **Supplementary Table 2B**, sentinel variants from Stage 1 were tested for association with the most comparable RCT smoking cessation endpoint: (i) for SVE, continuous abstinence during weeks 9-12, where cases were defined non-smokers between weeks 9-12 ($N_{cases} = 430$) and controls were defined as smokers with a positive carbon monoxide (CO) test at any point between weeks 9-12 ($N_{controls} = 808$); and (ii) for LVE, continuous abstinence at week 52 ('long-term quit'), where cases were defined non-smokers between weeks 9-52 ($N_{cases} = 367$), and controls were defined individuals who smoked at any point between weeks 9-52 ($N_{controls} = 1,222$). Where sentinels were not present in the Stage 2 genomic data, proxies within ± 100 kilobase (kb) in high linkage disequilibrium (LD) [$r^2 \geq 0.9$] were sought using PLINK³⁴. Stage 1 and Stage 2 results were meta-analysed using a fixed effect inverse variance weighted model implemented in METAL²⁷.

Comparison to published genetic studies

We performed two look-ups in the Stage 1 GWASs of SVE and LVE, including: (i) variants associated with varenicline efficacy and side effects reported in candidate gene association studies¹³⁻¹⁶; and (ii)

independent variants associated with smoking cessation (p -value $<5 \times 10^{-8}$) identified in the largest genomic study of smoking cessation to date in European individuals, conducted by the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN)³⁵. Additionally, a look-up of the Stage 1 sentinels (p -value $<5 \times 10^{-6}$) associated with SVE and LVE was performed in the European-only results of all the smoking behaviour phenotypes in the most recent GSCAN study³⁵; the results used for these look-up analyses were GWASs of smoking initiation, smoking cessation, CPD, and age at smoking initiation. Open Targets Genetics²⁸ was also used to determine whether any of the sentinel variants were also associated with other traits analysed in UK Biobank, FinnGen and the GWAS Catalog using the phenome-wide association approach in its 'PheWAS' module. The significance level for the PheWAS in Open Targets Genetics is defined after Bonferroni correction for the number of phenotypes tested, considering each as independent²⁸.

Results

Cohort characteristics

By leveraging varenicline prescription events and post-treatment smoking status records from clinical EHRs or baseline questionnaire data (Estonian Biobank only), we defined two phenotypes: short-term varenicline effectiveness (SVE) and long-term varenicline effectiveness (LVE) [**Supplementary Figure 1**]. **Table 1** outlines the characteristics of cases (successful quitters) and controls (continuing smokers) identified in each Stage 1 cohort (Estonian Biobank¹⁹, EXCEED²⁰, GoDARTS²¹, GoSHARE²² and UK Biobank²³) for inclusion in GWASs of each phenotype. Across both phenotypes defined in the five studies, there was no difference in the age of cases and controls, and only two studies (Estonian Biobank and GoDARTS) showed significant differences in the proportion of female cases compared to controls for SVE. For SVE, the smoking cessation rate ranged from 29.3% in GoDARTS to 44.4% in EXCEED, and for LVE, the smoking cessation rate was highest in Estonian Biobank at 42.7% and lowest in GoDARTS at 30.2%. GoDARTS had the lowest cessation rates in the short-term and long-term, suggesting barriers to sustained smoking abstinence, or an effect of the cohort recruitment strategy, whereby approximately half of the cohort have type 2 diabetes, for which smoking is an important risk factor³⁶. Interrogation of individual-level data in UK Biobank and EXCEED showed that, for both phenotypes, controls had fewer varenicline prescriptions within the 12-week treatment period than cases. Additionally, for both phenotypes defined in UK Biobank, there were more individuals with high nicotine dependence (corresponding to a HSI²⁴ score of 5 or 6) in the control group, compared to cases (p -value <0.001). For SVE, 13.3% of controls compared to 6.0% of cases had high nicotine dependence, while for LVE, 11.8% of controls compared to 6.8% of cases had high nicotine dependence.

Genome-wide association studies of varenicline-aided smoking cessation

For both varenicline effectiveness phenotypes, we conducted GWASs in each Stage 1 cohort and meta-analysed the results. **Figure 1** outlines the number of individuals and variants included in each analysis stage. The Stage 1 GWAS included up to 3,479 individuals (1,405 cases and 2,074 controls) for SVE, and up to 4,131 individuals (1,576 cases and 2,555 controls) for LVE. Both Stage 1 GWASs included over 10 million variants and did not show any evidence of test statistic inflation with λ_{GC} estimates of 0.996 and 0.994 for SVE and LVE, respectively (**Figure 2**; **Supplementary Figure 3**). We identified 10 sentinel variants associated with SVE and five sentinel variants associated with LVE that reached our pre-defined statistical threshold (p -value < 5×10^{-6}) [**Table 2**; further details in **Supplementary Table 3**]. There was no overlap (considering an LD threshold of $r^2 > 0.8$) between the sentinel variants for SVE and the sentinel variants for LVE. We compared effect size estimates for the sentinel variants after the exclusion of GoDARTS (as 50% of GoDARTS participants had type 2

diabetes) and found that this had no impact (**Supplementary Figure 4**). The 95% credible set for each locus (within which we expect the causal variant to lie with 95% probability) contained a median of 9 variants (range 1-105 variants) [**Supplementary Table 4**]. One locus (represented by sentinel rs2297038) had a single putative causal variant, and eight loci (represented by sentinel variants rs4364036, rs4686373, rs78597169, rs2297038, rs149318872, rs10421326, rs35443123 and rs766169) had a single variant with PIP >50%, all of which were also the sentinel variant.

The Stage 2 dataset comprised 430 cases and 808 controls for SVE and 367 cases and 1,222 controls for LVE from RCTs. Ten of the 15 sentinel variants (six SVE sentinel variants and four LVE sentinel variants) from Stage 1 were available for follow up in Stage 2 (**Figure 1; Table 2**). As five of the 15 Stage 1 sentinel variants were not represented in Stage 2 and as no variants reached genome-wide significance (p -value $<5 \times 10^{-8}$) in the meta-analysis of Stages 1 and 2 (**Table 2**), we describe all 15 variants taken into Stage 2 below (that is, reaching p -value $<5 \times 10^{-6}$ in Stage 1). These 15 sentinels collectively implicated 15 genes: *ALDH1A3*, *C2CD4C*, *DIO3*, *DPP4*, *LMO7*, *MTPAP*, *SULF2*, *THUMP2*, *TLL3* and *UTP23* for SVE, and *GLI2*, *PRDM13*, *SEMA3D*, *TBC1D22A* and *TNFAIP8* for LVE (**Table 2**). These genes are involved in the regulation of transcription (*GLI2* and *LMO7*), chromatin remodelling and methylation (*PRDM13*), processing of messenger RNA (*MTPAP*) and ribosomal RNA (*UTP23*), and transfer RNA methylation (*THUMP2*)^{31,32}. Additionally, some genes are essential for normal fetal development, including *ALDH1A3* for eye development, *DIO3* for thyroid hormone level regulation and *SEMA3D* for axon guidance and cell migration^{31,32}. Further gene functions include the regulation of apoptosis (*TNFAIP8*), glycoprotein metabolism (*SULF2*), glucose and insulin metabolism (*DPP4*), and cilium assembly (*TLL3* and *TBC1D22A*)^{31,32}. *TBC1D22A* is also likely to be involved in intracellular protein transport and GTPase activation³².

Phenome-wide association studies and comparison to published genetic studies

Using the repository of summary statistics from UK Biobank, FinnGen and the GWAS Catalog present in Open Targets Genetics²⁸, phenome-wide association studies (PheWASs) showed that two of the sentinels (*C2CD4C*-rs10421326 and *TNFAIP8*-rs7712227) were associated with various blood cell measurements. rs10421326-C (associated with increased varenicline effectiveness), was associated with increased lymphocyte count, reticulocyte count and related cellular phenotypes, while rs7712227-G (associated with increased varenicline effectiveness) was significantly associated with reduced mean reticulocyte volume.

Of the variants previously associated with varenicline efficacy in candidate gene studies¹³⁻¹⁶, only one reached a Bonferroni corrected threshold of p -value $<8.33 \times 10^{-3}$ (based on six independent tests considering variants as independent using an LD threshold of $r^2 < 0.2$) in our Stage 1 GWAS of LVE (**Supplementary Table 5**). However, this variant, *CHRNA4*-rs2236196, which was previously

associated with continuous abstinence at weeks 9-12 after varenicline treatment (OR [95% CI] = 1.54 [1.13, 2.09])¹³, had an opposite direction of effect to the LVE GWAS (OR [95% CI] = 0.86 [0.77, 0.96]) for allele G. Further, none of the variants associated with either nausea severity post-varenicline treatment in candidate gene association studies, or smoking cessation in GWAS, were associated with either SVE or LVE at a Bonferroni corrected threshold based on the number of independent tests (LD threshold of $r^2 < 0.2$) across each phenotype (**Supplementary Table 5**).

We determined whether any of our varenicline-mediated smoking cessation sentinel variants were previously associated with smoking behaviours (**Supplementary Table 6**). The LVE sentinel variant *PRDM13*-rs78695099 (not available to study in Stage 2) was associated with smoking behaviours, specifically smoking initiation (p -value = 0.00168), at a Bonferroni corrected p -value threshold based on 15 independent tests. Allele A of rs78695099, associated with a lower OR of quitting after varenicline treatment, conferred a reduced risk of smoking initiation.

Discussion

In this study, we leveraged routinely collected health data to undertake GWASs of varenicline-aided smoking cessation. Fifteen sentinel variants met our pre-defined statistical threshold (p -value $<5 \times 10^{-6}$) for SVE or LVE in Stage 1 (**Table 2; Supplementary Table 3**), and no overlap was observed between the variants associated with the two traits. None of these sentinels have been previously implicated in candidate gene association studies of varenicline-aided smoking cessation or in GWASs of smoking behaviour, and variants previously reported to be associated with varenicline efficacy, nausea, or smoking cessation were not associated with varenicline-aided smoking cessation in our study (**Supplementary Table 5**).

The 15 sentinels implicate genes that, individually, exhibit diverse cellular functions, with many involved in shared biological processes, including gene expression, specifically transcriptional regulation (*GLI2*, *LMO7* and *PRDM13*) and RNA modification (*MTPAP*, *UTP23*, and *THUMP2*), cilium assembly (*TLL3* and *TBC1D22A*), as well as processes essential for fetal growth and development (*ALDH1A3*, *DIO3*, *GLI2* and *SEMA3D*)^{31,32}. Further, *TBC1D22A*, which was implicated by LVE sentinel variant rs766169, has also been previously reported in smoking cessation by a statistically independent variant³⁵. With the caveat that future replication would be required, our investigations into the functions of the genes implicated by these variants provides initial insights into potential biological mechanisms.

Individuals who continue smoking after varenicline treatment may do so for various reasons beyond those relating to pharmacological efficacy, such as nicotine dependence or non-adherence potentially due to adverse reactions. We observed that controls (i.e. continuing smokers) had significantly higher nicotine dependence and received significantly fewer varenicline prescriptions, compared to cases. This suggests that non-adherence to treatment in the control group could be explained by lack of motivation or adverse reactions, thereby impacting on varenicline continuation and successful quit attempts.

By harnessing EHRs linked to five cohort studies, we have defined varenicline effectiveness with sample sizes almost 8-times larger than the largest candidate gene study of varenicline-aided smoking cessation¹³. To assess effectiveness, we considered only the first treatment episode to ensure a consistent approach to the clinical question being addressed by the study. Another strength of this study is the use of an unbiased GWAS approach, which allowed us to study over 10 million variants across the genome and identify novel associations. Previously, pharmacogenetic studies of this phenotype have taken a candidate gene approach where associations are limited to several hundred variants across a predefined set of likely implicated genes¹³⁻¹⁶.

Our study was well powered to detect variant associations of relatively large effect. For example, the study was well-powered at a genome-wide significance threshold (p -value $<5 \times 10^{-8}$) to detect common variants (MAF $>40\%$) with an odds ratio as low as 1.2 for both varenicline response phenotypes examined (**Supplementary Figure 2**). Whilst larger effect size estimates have been reported for some genetic determinants of drug response³⁷, common genetic variants associated with common complex traits have been typically more modest¹⁷. Replicable genetic associations of common complex traits have typically emerged in small numbers once several thousand cases were studied, and replicable associations increased dramatically once tens of thousands of cases became available to study¹⁷. Genetic studies of drug response have typically been much smaller, and whilst larger than previous candidate gene association studies of varenicline response¹³⁻¹⁶, fewer than 1,600 cases were available in each of our stages. In this context, the lack of replicable findings to date indicates the need for larger sample sizes. For studies utilising EHR-based phenotypes, this could be achieved through a combination of further data linkage in participants who have already consented to this (for example, UK Biobank so far has linked primary care EHRs on less than one-half of consenting participants) and through additional Biobanks (such as *All of Us*³⁸ and Our Future Health³⁹) as data becomes available. Denser genotyping arrays and new imputations would improve the availability of data for follow up; this is important to address given the unavailability of five of the 15 sentinels in Stage 2. Our Stage 1 studies were largely population-based studies or biobanks with healthy participants. This would be expected to impact on power where there is a lower proportion of smoking and could also impact on generalisability of findings to those with disease (although effect estimates did not alter when GoDARTS was excluded, in which 50% of participants had type 2 diabetes). Further, these analyses were restricted to individuals of European genetic ancestry and utilising similar data from developing initiatives with focused recruitment in minority ethnic populations^{38,39} will help to ensure the generalisability of findings. In summary, despite being the largest global study varenicline response, the associations we discovered did not reach genome-wide significance (p -value $<5 \times 10^{-8}$) and need to be viewed with caution pending independent replication or larger meta-analyses that corroborate these findings. Factors impacting on the power to detect genetic associations at this threshold include: (i) underlying effect sizes of common variants associated with varenicline response likely to be more modest than originally assumed; (ii) density of genotyping platforms and of imputation panels, such that not all studies contributed to meta-analysis of a given variant; (iii) possible measurement error in defining the phenotype from EHRs.

Varenicline was still being widely prescribed when we initiated this study, though its production was halted in July 2021 due to a nitrosamine impurity^{40,41}. In considering the future availability and use of varenicline, we note the following: (i) varenicline is the most efficacious smoking cessation therapy, was approved for use in 116 countries and has been prescribed to 24 million smokers⁴⁰; (ii) the

withdrawal of varenicline was not due to concerns about efficacy or side effects, but related to the theoretical risks of cancer attributable to the nitrosamine impurity⁴⁰; (iii) the health benefits of stopping smoking on cancer risk substantially outweigh cancer risk from the nitrosamine impurity⁴⁰; (iv) the Federal Drug Administration (FDA) stated its confidence in the ability to manufacture varenicline containing ≤ 37 ng/day of the N-nitroso-varenicline impurity (the agency's acceptable intake limit)⁴²; (v) varenicline is now off-patent and FDA-approved generic varenicline became available in September 2021, after which a rise in prescriptions was observed in the US⁴¹ and; (vi) generic varenicline was introduced in the UK in August 2024⁴³.

In conclusion, this is largest genome-wide association study of varenicline-aided smoking cessation to date, which incorporates definitions of short- and long-term cessation, and implicates a diverse range of processes, including gene expression, cilium assembly and fetal development. Larger and more diverse populations will be required for powerful, generalisable genetic association studies of varenicline-aided smoking cessation. The EHR-based phenotyping approach we present here can be applied once additional data becomes available in UK Biobank and other large studies under development, facilitating the corroboration of findings from this study.

Funding

This study was supported by specific personal funding from the following: Wellcome Trust Investigator Award (WT202849/Z/16/Z), Wellcome Trust Discovery Award (WT225221/Z/22/Z), MRC grant number MR/N011317/1 and NIHR Senior Investigator Award to MDT; and UKRI Innovation Fellowship at Health Data Research UK grant number MR/S003762/1 to CB. This research is funded by the National Institute for Health and Care Research (NIHR) Leicester Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of Interests

RP and CJ report funding from Orion Pharma outside of the submitted work. MDT has research collaborations with Orion Pharma and GlaxoSmithKline unrelated to the current work. SP, CH and AM are employees of Pfizer.

Data availability

Summary statistics from SVE and LVE Stage 1 meta-analyses are being deposited at the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and accession numbers will be added as soon as they are available.

Acknowledgements

We thank all participants of the cohort studies and randomised controlled trials used in this project. This study used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester. The Estonian Genome Center analyses were partially carried out in the High Performance Computing Center, University of Tartu.

Estonian Biobank: The activities of the Estonian Biobank are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of Estonian Biobank. Individual level data analysis in Estonian Biobank was carried out under ethical approval 1.1-12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 6-7/GI/33501 from the Estonian Biobank. The work of the Estonian Genome Center, University of Tartu was funded by the European Union through Horizon 2020 research and innovation program under grants no. 810645 and 894987, through the European Regional Development Fund projects GENTRANSMED (2014-2020.4.01.15-0012), MOBEC008, MOBERA21 and the Estonian Research Council grant PUT (PRG1291, PRG687 and PRG184).

EXCEED: EXCEED is supported by the University of Leicester, the National Institute for Health and Care Research Leicester Respiratory Biomedical Research Centre, the Wellcome Trust (WT 202849), and Cohort Access fees from studies funded by the Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council, National Institute for Health and Care Research, the UK Space Agency, and GlaxoSmithKline. It was previously supported by Medical Research Council grant G0902313. EXCEED is supported by BREATHE - The Health Data Research Hub for Respiratory Health (UKRI_PC_19004) in partnership with SAIL Databank. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK. EXCEED received ethical approval from the Leicester Central Research Ethics Committee (13/EM/0226). Substantial amendments have been approved by the same Research Ethics Committee for the collection of new data relating to the COVID-19 pandemic, including the COVID-19 questionnaires and antibody testing.

GoDARTS & GoSHARE: We are grateful to all the participants of GoDARTS and SHARE study, the general practitioners, the Scottish School of Primary Care for their help in recruiting the participants and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The study complies with the Declaration of Helsinki. We acknowledge the support of the Health Informatics Centre, University of Dundee, for managing and supplying the anonymized data and NHS Tayside, the original data owner. GoDARTS and SHARE studies were reviewed and approved by Tayside Medical Ethics Committee 053/04 and East of Scotland Ethics Committee NHS REC 13/ES/0020. The Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (supporting GoDARTS) was funded by the Wellcome Trust (072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT programme. GoSHARE is funded by NHS Research Scotland (NRS) (<http://www.nhsresearchscotland.org.uk/>), a partnership of Scottish Health Boards and the Scottish Government Chief Scientist Office (CSO) that oversees governance and management of clinical research.

UK Biobank: UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. The UK Biobank genetic and phenotypic data were obtained under UK Biobank Applications 4982 and 59822. UK Biobank has ethical approval from the UK National Health Service (NHS) National Research Ethics Service (11/NW/0382).

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. GBD Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. Nov 10 2018;392(10159):1923-1994. doi:10.1016/S0140-6736(18)32225-6
2. West R. Tobacco smoking: Health impact, prevalence, correlates and interventions. *Psychol Health*. Aug 2017;32(8):1018-1036. doi:10.1080/08870446.2017.1325890
3. Action on Smoking and Health. *Smoking Statistics*. 2023.
<https://ash.org.uk/resources/view/smoking-statistics>
4. NHS Digital. *Statistics on Smoking, England 2020*. 2020. *Statistics on Smoking*. Accessed 26th January 2022. <https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-smoking/statistics-on-smoking-england-2020>
5. Office for National Statistics. *Adult smoking habits in the UK: 2022*. 2023. *Statistical Bulletin*. Accessed 3rd November 2023.
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2022>
6. World Health Organization. *WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke*. 2023. Accessed 3rd November 2023.
<https://iris.who.int/bitstream/handle/10665/372043/9789240077164-eng.pdf?sequence=1>
7. National Institute for Health and Care Excellence. *Tobacco: preventing uptake, promoting quitting and treating dependence [NG209]*. 2021. Accessed 8th March 2022.
<https://www.nice.org.uk/guidance/ng209>
8. Cahill K, Stevens S, Perera R, Lancaster T. Pharmacological interventions for smoking cessation: an overview and network meta-analysis. *Cochrane Database Syst Rev*. May 31 2013;2013(5):Cd009329. doi:10.1002/14651858.CD009329.pub2
9. Taylor GMJ, Taylor AE, Thomas KH, et al. The effectiveness of varenicline versus nicotine replacement therapy on long-term smoking cessation in primary care: a prospective cohort study of electronic medical records. *Int J Epidemiol*. Dec 1 2017;46(6):1948-1957. doi:10.1093/ije/dyx109
10. NHS Digital. *Statistics on NHS Stop Smoking Services in England - April 2021 to March 2022*. 2022. *Statistics on NHS Stop Smoking Services in England*. Accessed 3rd November 2023.
<https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-nhs-stop-smoking-services-in-england/april-2020-to-march-2021/part-3---stop-smoking-services>
11. Pirmohamed M. Pharmacogenomics: current status and future perspectives. *Nat Rev Genet*. Jun 2023;24(6):350-362. doi:10.1038/s41576-022-00572-8

12. Ebbert JO, Wyatt KD, Hays JT, Klee EW, Hurt RD. Varenicline for smoking cessation: efficacy, safety, and treatment recommendations. *Patient Prefer Adherence*. Oct 5 2010;4:355-62.
doi:10.2147/ppa.s10620
13. King DP, Paciga S, Pickering E, et al. Smoking cessation pharmacogenetics: analysis of varenicline and bupropion in placebo-controlled clinical trials. *Neuropsychopharmacology*. Feb 2012;37(3):641-50. doi:10.1038/npp.2011.232
14. Rocha Santos J, Tomaz PR, Issa JS, et al. CHRNA4 rs1044396 is associated with smoking cessation in varenicline therapy. *Front Genet*. 2015;6:46. doi:10.3389/fgene.2015.00046
15. Tomaz PRX, Kajita MS, Santos JR, et al. Cytochrome P450 2A6 and 2B6 polymorphisms and smoking cessation success in patients treated with varenicline. *Eur J Clin Pharmacol*. Nov 2019;75(11):1541-1545. doi:10.1007/s00228-019-02731-z
16. Chen LS, Baker TB, Miller JP, et al. Genetic Variant in CHRNA5 and Response to Varenicline and Combination Nicotine Replacement in a Randomized Placebo-Controlled Trial. *Clin Pharmacol Ther*. Dec 2020;108(6):1315-1325. doi:10.1002/cpt.1971
17. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. Jul 6 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005
18. McInnes G, Yee SW, Pershad Y, Altman RB. Genomewide Association Studies in Pharmacogenomics. *Clin Pharmacol Ther*. Sep 2021;110(3):637-648. doi:10.1002/cpt.2349
19. Leitsalu L, Haller T, Esko T, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol*. Aug 2015;44(4):1137-47. doi:10.1093/ije/dyt268
20. John C, Reeve NF, Free RC, et al. Cohort profile: Extended Cohort for E-health, Environment and DNA (EXCEED). *Int J Epidemiol*. Oct 1 2019;48(5):1734. doi:10.1093/ije/dyz175
21. Hebert HL, Shepherd B, Milburn K, et al. Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *Int J Epidemiol*. Apr 1 2018;47(2):380-381j.
doi:10.1093/ije/dyx140
22. McKinstry B, Sullivan FM, Vasishta S, et al. Cohort profile: the Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. *BMJ Open*. Feb 1 2017;7(2):e013351. doi:10.1136/bmjopen-2016-013351
23. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. Oct 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
24. Borland R, Yong HH, O'Connor RJ, Hyland A, Thompson ME. The reliability and predictive validity of the Heaviness of Smoking Index and its two components: findings from the International Tobacco Control Four Country study. *Nicotine Tob Res*. Oct 2010;12 Suppl(Suppl 1):S45-50.
doi:10.1093/ntr/ntq038
25. Johnson JL, Abecasis GR. GAS Power Calculator: web-based power calculator for genetic association studies. *bioRxiv*. 2017;doi:10.1101/164343

26. Karolchik D, Hinrichs AS, Kent WJ. The UCSC Genome Browser. *Curr Protoc Bioinformatics*. Dec 2012;Chapter 1:1 4 1-1 4 33. doi:10.1002/0471250953.bi0104s40
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. Sep 1 2010;26(17):2190-1. doi:10.1093/bioinformatics/btq340
28. Ghossaini M, Mountjoy E, Carmona M, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*. Jan 8 2021;49(D1):D1311-D1320. doi:10.1093/nar/gkaa840
29. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. Jun 6 2016;17(1):122. doi:10.1186/s13059-016-0974-4
30. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. Sep 2010;38(16):e164. doi:10.1093/nar/gkq603
31. Ochoa D, Hercules A, Carmona M, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res*. Jan 6 2023;51(D1):D1353-D1359. doi:10.1093/nar/gkac1046
32. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. Jun 20 2016;54:1 30 1-1 30 33. doi:10.1002/cpbi.5
33. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol*. Jan 2009;33(1):79-86. doi:10.1002/gepi.20359
34. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. Sep 2007;81(3):559-75. doi:10.1086/519795
35. Saunders GRB, Wang X, Chen F, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. Dec 2022;612(7941):720-724. doi:10.1038/s41586-022-05477-4
36. Maddatu J, Anderson-Baucum E, Evans-Molina C. Smoking and the risk of type 2 diabetes. *Transl Res*. Jun 2017;184:101-107. doi:10.1016/j.trsl.2017.02.004
37. Maranville JC, Cox NJ. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. *Pharmacogenomics J*. Aug 2016;16(4):388-92. doi:10.1038/tpj.2015.47
38. All of Us Research Program Investigators. The "All of Us" Research Program. *N Engl J Med*. Aug 15 2019;381(7):668-676. doi:10.1056/NEJMs1809937
39. Our Future Health. Our Future Health. Accessed 17th November 2023, <https://ourfuturehealth.org.uk/>
40. Lang AE, Berlin I. Unavailability of varenicline: a global tragedy for the fight against the tobacco epidemic. *Lancet Respir Med*. Jun 2023;11(6):518-519. doi:10.1016/S2213-2600(23)00184-4
41. Lang AE, Patel U, Fitzpatrick J, Lee T, McFarland M, Good CB. Association of the Chantix Recall With US Prescribing of Varenicline and Other Medications for Nicotine Dependence. *JAMA Netw Open*. Feb 1 2023;6(2):e2254655. doi:10.1001/jamanetworkopen.2022.54655

42. Food and Drug Administration. FDA Updates and Press Announcements on Nitrosamine in Varenicline (Chantix). Accessed 4th June 2024, <https://www.fda.gov/drugs/drug-safety-and-availability/fda-updates-and-press-announcements-nitrosamine-varenicline-chantix>
43. National Centre for Smoking Cessation and Training. Varenicline. Accessed 16th September 2024, <https://www.ncsct.co.uk/library/view/pdf/NCST-Generic-varenicline.pdf>

Tables

Table 1 Case-control characteristics for **(A)** SVE and **(B)** LVE defined in Stage 1 cohorts.

Table 2 Sentinel variants associated with **(A)** SVE and **(B)** LVE in Stage 1 GWASs.

Table 1 Case-control characteristics for (A) SVE and (B) LVE defined in Stage 1 cohorts.

Cohort	N		Smoking cessation rate*, % [95% CI]	Mean age at <i>index date</i> , years (SD)			Female, n (%)			Mean number of varenicline prescriptions† (SD)		
	Cases	Controls		Cases	Controls	p-value	Cases	Controls	p-value	Cases	Controls	p-value
(A) Short-term varenicline effectiveness (SVE): Total of 1,405 cases and 2,074 controls												
Estonian Biobank	22	44	33.3 [22.2, 46.0]	43.1 (10.3)	44.1 (10.3)	0.71	17 (77.3)	10 (21.7)	<0.0001	–	–	–
EXCEED	64	80	44.4 [36.2, 52.9]	51.5 (8.9)	51.3 (9.3)	0.90	33 (51.6)	39 (48.8)	0.87	3.83 (1.72)	2.80 (1.53)	0.00027
GoDARTS	72	174	29.3 [23.7, 35.4]	61.9 (10.0)	59.4 (9.9)	0.82	25 (34.7)	92 (52.9)	0.01	–	–	–
GoSHARE	99	225	30.6 [25.6, 35.9]	56 (10.1)	54.8 (10.2)	0.34	51 (51.5)	108 (48.0)	0.64	–	–	–
UK Biobank	1,148	1,551	42.5 [40.7, 44.4]	55.7 (8.1)	55.6 (7.9)	0.75	654 (57.0)	863 (55.6)	0.52	3.72 (1.66)	2.70 (1.50)	<0.0001
(B) Long-term varenicline effectiveness (LVE): Total of 1,576 cases and 2,555 controls												
Estonian Biobank	85	114	42.7 [35.7, 49.9]	43.7 (12.0)	44.5 (9.5)	0.62	46 (54.1)	57 (49.1)	0.58	–	–	–
EXCEED	73	101	42.0 [34.5, 49.7]	50.6 (8.8)	50.5 (8.8)	0.94	39 (53.4)	48 (47.5)	0.54	3.56 (1.79)	2.94 (1.57)	0.019
GoDARTS	74	171	30.2 [24.5, 36.4]	60.5 (9.9)	59.3 (9.8)	0.42	35 (47.3)	82 (48.0)	1.00	–	–	–
GoSHARE	117	234	33.3 [28.4, 38.5]	55.4 (11.4)	53.8 (11.4)	0.22	59 (50.4)	118 (50.4)	1.00	–	–	–
UK Biobank	1,227	1,935	38.8 [37.1, 40.5]	55.2 (8.1)	55.2 (8.0)	1.00	697 (56.8)	1,069 (55.2)	0.41	3.53 (1.67)	2.88 (1.53)	<0.0001

Index date is defined as the date of the varenicline therapy initiation in first treatment episode. *The proportion of individuals prescribed varenicline and were non-smokers (defined from the first smoking status record) within the follow-up period for each phenotype. †During 12-week treatment episode after *index date*. *Abbreviations*: CI, confidence interval; SD, standard deviation.

Table 2 Sentinel variants associated with **(A)** SVE and **(B)** LVE in Stage 1, Stage 2 and Stage 1 + 2.

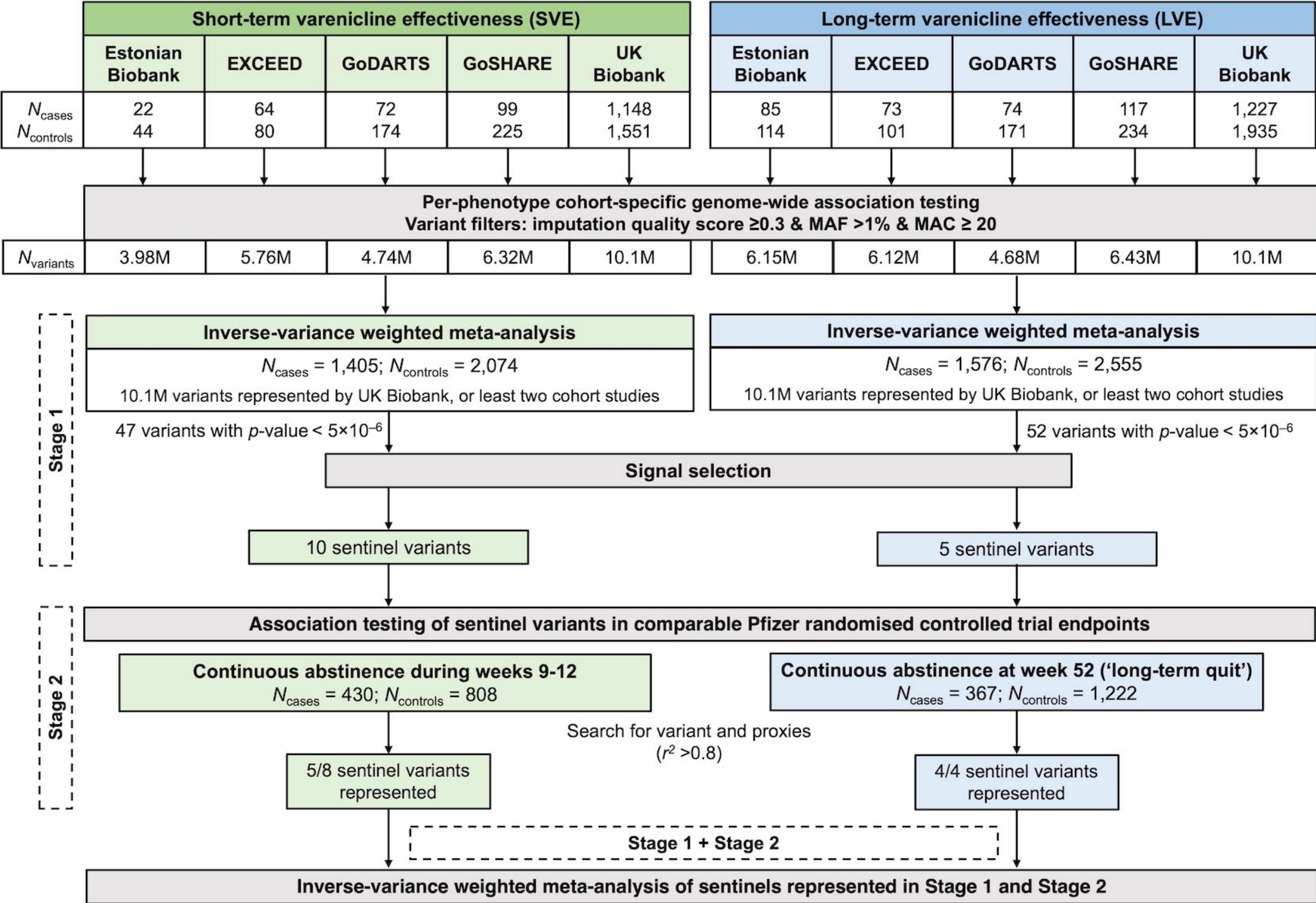
Sentinel variant	Tested variant*	Chr:Position (hg19)	EA	Stage 1			Stage 2			Stage 1 + 2			Consequence	Mapped Gene [†]
				EAF	OR [95% CI]	p-value	EAF	OR [95% CI]	p-value	EAF	OR [95% CI]	p-value		
(A) Short-term varenicline effectiveness (SVE)														
rs186444103		2:40054022	T	0.026	2.27 [1.61, 3.21]	3.20×10 ⁻⁶	–	–	–	0.026	2.27 [1.61, 3.21]	3.20×10 ⁻⁶	Intronic	<i>THUMPD2</i>
rs4364036		2:162904624	T	0.034	2.13 [1.55, 2.92]	2.55×10 ⁻⁶	0.027	0.73 [0.43, 1.24]	2.34×10 ⁻¹	0.033	1.61 [1.23, 2.12]	5.44×10 ⁻⁴	Intronic	<i>DPP4</i>
rs4686373		3:9853437	A	0.09	1.55 [1.29, 1.85]	2.50×10 ⁻⁶	0.078	1.21 [0.90, 1.64]	2.13×10 ⁻¹	0.087	1.45 [1.24, 1.69]	2.91×10 ⁻⁶	Intronic	<i>TTL3</i>
rs78597169		8:117830288	A	0.975	0.43 [0.30, 0.61]	3.12×10 ⁻⁶	–	–	–	0.975	0.43 [0.30, 0.61]	3.12×10 ⁻⁶	Intronic	<i>UTP23</i>
rs2297038 [‡]		10:30663464	T	0.434	1.33 [1.18, 1.50]	2.95×10 ⁻⁶	–	–	–	0.434	1.33 [1.18, 1.50]	2.95×10 ⁻⁶	Upstream	<i>MTPAP</i>
rs9600669 [§]		13:76776621	A	0.914	1.57 [1.30, 1.88]	1.94×10 ⁻⁶	0.900	0.80 [0.60, 1.06]	1.14×10 ⁻¹	0.91	1.28 [1.10, 1.50]	1.66×10 ⁻³	Intergenic	<i>LMO7</i>
rs8003262		14:102050501	A	0.941	1.73 [1.37, 2.19]	4.42×10 ⁻⁶	0.929	0.96 [0.70, 1.30]	7.71×10 ⁻¹	0.937	1.40 [1.16, 1.69]	4.46×10 ⁻⁴	Intergenic	<i>DIO3</i>
rs149318872		15:101162067	G	0.951	0.50 [0.37, 0.67]	3.42×10 ⁻⁶	–	–	–	0.951	0.50 [0.37, 0.67]	3.42×10 ⁻⁶	Intronic	<i>ALDH1A3</i>
rs10421326		19:473572	C	0.174	1.53 [1.31, 1.78]	5.74×10 ⁻⁸	0.211	0.93 [0.71, 1.21]	5.45×10 ⁻¹	0.185	1.32 [1.16, 1.50]	2.48×10 ⁻⁵	Intronic	<i>C2CD4C</i>
rs827937	rs228194	20:46793101	A	0.755	1.32 [1.18, 1.50]	2.33×10 ⁻⁶	0.739	1.04 [0.86, 1.27]	6.74×10 ⁻¹	0.751	1.25 [1.13, 1.38]	1.98×10 ⁻⁵	Intergenic	<i>SULF2</i>
(B) Long-term varenicline effectiveness (LVE)														
rs895545		2:121721467	T	0.943	0.59 [0.48, 0.73]	1.21×10 ⁻⁶	0.942	1.31 [0.88, 1.94]	1.74×10 ⁻¹	0.943	0.71 [0.58, 0.85]	2.72×10 ⁻⁴	Intronic	<i>GLI2</i>
rs7712227		5:118656769	A	0.06	0.59 [0.47, 0.73]	2.56×10 ⁻⁶	0.067	1.02 [0.72, 1.44]	9.28×10 ⁻¹	0.062	0.69 [0.57, 0.83]	8.08×10 ⁻⁵	Intronic	<i>TNFAIP8</i>
rs78695099		6:100195347	A	0.974	0.42 [0.30, 0.58]	2.05×10 ⁻⁷	–	–	–	0.974	0.42 [0.30, 0.58]	2.05×10 ⁻⁷	Intergenic	<i>PRDM13</i>
rs35443123 [‡]		7:85173800	A	0.434	0.79 [0.72, 0.88]	4.86×10 ⁻⁶	0.453	0.97 [0.82, 1.15]	7.27×10 ⁻¹	0.439	0.84 [0.77, 0.91]	4.15×10 ⁻⁵	Intergenic	<i>SEMA3D</i>
rs766169 [§]		22:47680100	T	0.051	1.84 [1.47, 2.29]	6.68×10 ⁻⁸	0.059	0.52 [0.34, 0.80]	1.30×10 ⁻³	0.052	1.41 [1.16, 1.71]	6.44×10 ⁻⁴	Intergenic	<i>TBC1D22A</i>

*Where sentinel variants were not present in Stage 2, tested variants were defined as a proxy within 100kb and $r^2 \geq 0.9$ (r^2 between rs827937 and rs228194 is 0.99); test statistics refer to the tested variant. [†]Gene with highest ‘overall V2G’ score in Open Targets Genetics. [‡]Variant not present in Open Targets Genetics database (nearest gene is reported in ‘Mapped Gene’ column). [§]No V2G evidence presented in Open Targets Genetics database (nearest gene is reported in ‘Mapped Gene’ column).
Abbreviations: Chr, chromosome; CI, confidence interval; EA, effect allele; EAF, effect allele frequency; OR, odds ratio.

Figures

Figure 1 Overview of study design. *Abbreviations:* MAC, minor allele count; MAF, minor allele frequency.

Figure 2 Miami plot for Stage 1 GWASs of short-term varenicline effectiveness (SVE) and long-term varenicline effectiveness (LVE). The grey solid line represents the genome-wide significance threshold (p -value $<5 \times 10^{-8}$), and the grey dotted line represents the suggestive significance threshold (p -value $<5 \times 10^{-6}$).



SVE

