

# Genomic insights for personalised care in lung cancer and smoking cessation: motivating at-risk individuals toward evidence-based health practices



Tony Chen,<sup>a,\*</sup> Giang Pham,<sup>b</sup> Louis Fox,<sup>b</sup> Nina Adler,<sup>c,d</sup> Xiaoyu Wang,<sup>e,f</sup> Jingning Zhang,<sup>g</sup> Jinyoung Byun,<sup>h</sup> Younghun Han,<sup>h</sup> Gretchen R. B. Saunders,<sup>i</sup> Dajiang Liu,<sup>j</sup> Michael J. Bray,<sup>k,l</sup> Alex T. Ramsey,<sup>b</sup> James McKay,<sup>m</sup> Laura J. Bierut,<sup>b</sup> Christopher I. Amos,<sup>h,n</sup> Rayjean J. Hung,<sup>d</sup> Xihong Lin,<sup>a,o</sup> Haoyu Zhang,<sup>e,p,\*\*</sup> and Li-Shiun Chen<sup>b,p,\*\*\*</sup>



<sup>a</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA

<sup>b</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, USA

<sup>c</sup>Department of Anthropology, University of Toronto, Toronto, ON, Canada

<sup>d</sup>Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, and University of Toronto, Toronto, Canada

<sup>e</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

<sup>f</sup>Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Rockville, MD, USA

<sup>g</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

<sup>h</sup>Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA

<sup>i</sup>Department of Psychology, University of Minnesota, Minneapolis, MN, USA

<sup>j</sup>Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA

<sup>k</sup>Department of Genetic Counseling, Bay Path University, Longmeadow, MA, USA

<sup>l</sup>ThinkGenetics, Inc, USA

<sup>m</sup>International Agency for Research on Cancer, World Health Organization, Lyon, France

<sup>n</sup>Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA

<sup>o</sup>Department of Statistics, Harvard University, Cambridge, USA

## Summary

**Background** Lung cancer and tobacco use pose significant global health challenges, necessitating a comprehensive translational roadmap for improved prevention strategies such as cancer screening and tobacco treatment, which are currently under-utilised. Polygenic risk scores (PRSs) may further motivate health behaviour change in primary care for lung cancer in diverse populations. In this work, we introduce the GREAT care paradigm, which integrates PRSs within comprehensive patient risk profiles to motivate positive health behaviour changes.

**Methods** We developed PRSs using large-scale multi-ancestry genome-wide association studies and standardised PRS distributions across all ancestries. We validated our PRSs in 561,776 individuals of diverse ancestry from the GISC Trial, UK Biobank (UKBB), and All of Us Research Program (AoU).

**Findings** Significant odds ratios (ORs) for lung cancer and difficulty quitting smoking were observed in both UKBB and AoU. For lung cancer, the ORs for individuals in the highest risk group (top 20% versus bottom 20%) were 1.85 (95% CI: 1.58–2.18) in UKBB and 2.39 (95% CI: 1.93–2.97) in AoU. For difficulty quitting smoking, the ORs (top 33% versus bottom 33%) were 1.36 (95% CI: 1.32–1.41) in UKBB and 1.32 (95% CI: 1.28–1.36) in AoU.

**Interpretation** Our PRS-based intervention model leverages large-scale genetic data for robust risk assessment across populations, which will be evaluated in two cluster-randomised clinical trials. This approach integrates genomic insights into primary care, promising improved outcomes in cancer prevention and tobacco treatment.

**Funding** National Institutes of Health, NIH Intramural Research Program, National Science Foundation.

\*Corresponding author.

\*\*Corresponding author.

\*\*\*Corresponding author.

E-mail addresses: [tonychen@g.harvard.edu](mailto:tonychen@g.harvard.edu) (T. Chen), [haoyu.zhang2@nih.gov](mailto:haoyu.zhang2@nih.gov) (H. Zhang), [li-shiun@wustl.edu](mailto:li-shiun@wustl.edu) (L.-S. Chen).

<sup>P</sup>These authors jointly supervised this work: Haoyu Zhang and Li-Shiun Chen.

eBioMedicine  
2024;110: 105441  
Published Online xxx  
<https://doi.org/10.1016/j.ebiom.2024.105441>

**Copyright** Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

**Keywords:** Lung cancer; Polygenic risk scores; Cancer prevention; Personalised interventions; Translational roadmap; Health behaviour change

### Research in context

#### Evidence before this study

We systematically searched PubMed for research articles published in English before June 21, 2024, using the search terms “polygenic risk score”, “lung cancer” and “clinical trials”. Several clinical trials, including the WISDOM, GenoVA, and eMERGE studies, have evaluated the utility of polygenic risk scores (PRSs) for diseases such as breast cancer and coronary artery disease. However, these studies do not include lung cancer or tobacco use, both of which present significant global health burdens. While genome-wide association studies (GWASs) have identified several genetic risk factors for both conditions, and PRSs for lung cancer have been developed, they have not yet been implemented in clinical settings.

#### Added value of this study

Our study introduces a new primary care paradigm that integrates the latest genetic data and research into precision

treatment of lung cancer and smoking cessation. We compute ancestry-adjusted PRS using large-scale GWASs and develop standardised risk models applicable across diverse ancestry. We validate our model’s accuracy in three ancestrally diverse cohorts, including two large-scale biobanks, finding significant odds ratios for both lung cancer and difficulty quitting smoking between the top and bottom risk groups.

#### Implications of all the available evidence

Our findings demonstrate that ancestry-adjusted PRS can effectively stratify patient risk for lung cancer and difficulty quitting smoking across diverse backgrounds. This framework will be implemented in two cluster-randomised clinical trials to evaluate the efficacy of PRS-based intervention tools in promoting health behaviour changes related to cancer prevention and tobacco treatment in patients of diverse ancestry.

## Introduction

The worldwide burden of lung cancer and tobacco smoking presents major challenges to global health.<sup>1</sup> Evidence-based practices to reduce their risk such as cancer screening and tobacco treatment (e.g. smoking cessation medication) have long existed but are infrequently used in most primary care practices. While it is well-established that quitting smoking can dramatically reduce the risk of developing lung cancer,<sup>2</sup> patients may face various barriers to participating in smoking cessation programs.<sup>3</sup> This presents a critical gap in preventive healthcare, particularly for individuals at high risk of lung cancer.

Polygenic risk scores (PRSs) have emerged as a valuable approach to assess disease susceptibility among populations and pinpoint individuals at higher risk.<sup>4–6</sup> By incorporating PRSs into clinical practice, we can provide patients with personalised information that not only highlights their heightened genetic risk for lung cancer but also expected difficulty quitting smoking. In our previous studies, large majorities of patients who smoke showed high interest in receiving genetically tailored tobacco treatment,<sup>7</sup> endorsed the importance of genetic results to guide treatment,<sup>8</sup> and found genetics-based tools useful in understanding health risk and taking control of their health.<sup>9</sup> With this personalised approach that makes risk more tangible, patients may be more motivated to engage in smoking cessation programs or undergo lung cancer screening.<sup>7</sup> However, substantial evidence suggests the complex interplay of

fear, fatalism, and risk perception when patients face lung cancer screening.<sup>10–13</sup> Therefore, it is important to evaluate the translation potential of personalised genetics in motivating patients for positive behaviour change.

Ongoing studies like eMERGE (electronic MEDical Records and GENomics),<sup>14</sup> GenoVA (Genomic Medicine at Veterans Affairs),<sup>15</sup> and WISDOM (Women Informed to Screen Depending on Measures of Risk)<sup>16</sup> are leading the implementation of PRS into genetic risk reports (Table 1). These studies aim to personalise medical reports and assess the impact of PRS on screening, diagnostic procedures, and patient behaviour. While the global burden of lung cancer is largely driven by tobacco smoking,<sup>17,18</sup> research studies have shown promise for PRS in lung cancer risk independent of traditional clinical risk models.<sup>19,20</sup> The unique value proposition of a lung cancer-specific PRS lies in leveraging clear, guideline-based recommendations for smoking cessation and lung cancer screening.<sup>20</sup>

Despite the potential benefits of PRS-based interventions, several practical challenges must be addressed for their effective implementation in the clinic. These include identifying the appropriate PRS to use, determining PRS cutoffs for defining patients at high-risk, understanding the corresponding odds ratios (ORs) for these patients, and standardizing PRS distributions across ancestries to ensure equitable application to diverse populations. Additionally, there remain

	GenoVA	eMERGE	Wisdom	PRECISE / MOTIVATE
<b>Goal</b>	Motivate early diagnosis	Increasing clinical actions to mitigate risk of future disease	Personalized screening frequency	Motivate lung cancer screening (LCS) / tobacco treatment (tx)
<b>Targets</b>				
<b>Target population</b>	All general Veterans' Affairs PCP* and patients	Patients in the large healthcare system	All women can sign up online	<b>High risk PC patients (LCS eligible / smokers) and their PCPs</b>
<b>Condition(s)**</b>	BrCa, PrCa, CRCa, Afib, CAD, T2D	BrCa, PrCa, CRCa, Afib, CAD, T1D, T2D, BMI, Asthma, CKD, HCL	BrCa	<b>LC, smoking cessation</b>
<b>Target outcome</b>	New diagnoses	Change screening practice or lifestyle	Screening compliance and cancer detection	<b>LCS, tobacco tx, and smoking cessation</b>
<b>Polygenic Risk Scores</b>				
<b>Genotyping</b>	Study team		Color genomics	23andMe
<b>Imputation</b>	1000 Genomes	All of Us		TOPMed
<b>Create PRS</b>	Top or many variants European weights Ancestry calibration	Terra cloud platform Ancestry calibration	Color genomics	Top independent variants Multi-ancestry weights Ancestry calibration
<b>Risk Stratification</b>	2 levels (OR of 2)	OR of 5-10		3 levels (top 20%, middle, bottom 20%) or tertiles
<b>Intervention Design</b>				
<b>Risk representation</b>	Only genetic Format of relative risk	Combined clinical + genetic Format of absolute risk	Combined clinical + genetic Format of absolute risk	Separate genetic + clinical Format of relative risk
<b>Messaging</b>	Average vs. elevated risk	Average vs. elevated risk	No discussion of risk Only recommend screening schedule	At risk, at high risk, or at very high risk
<b>Behavior mechanisms targeted</b>				<b>Perceived risk/benefit, Self efficacy, Personal relevance</b>

We compare the PRECISE and MOTIVATE trials, part of our GREAT framework, with existing PRS-informed trials: GenoVA, eMERGE, and WISDOM. Bolded text in the PRECISE/MOTIVATE column highlights the points where our trials differ from the current trials. Namely, the PRECISE and MOTIVATE trials investigate lung cancer and smoking and will focus on patients at high risk who are smokers or eligible for lung cancer screening. We also look at lung cancer screening, tobacco treatment, and smoking cessation as unique target outcomes. Finally, in addition to genetic and clinical risk messaging, the two trials have a unique emphasis on behaviour mechanisms around lung cancer and smoking. \* PCP, primary care provider. \*\* BrCa, breast cancer; PrCa, prostate cancer; CRCa, colorectal cancer; Afib, atrial fibrillation; CAD, coronary artery/heart disease; T2D, type 2 diabetes; T1D, type 1 diabetes; BMI, body mass index/obesity; CKD, chronic kidney disease; HCL, hypercholesterolemia; LC, lung cancer.

Table 1: Research on PRS use in clinical trials.

significant barriers of effectively communicating genetic risk information to patients and ensuring that is understood and acted upon.

To address these challenges and provide practical guidance for the implementation for PRSs in primary care, we introduce the Genomic Informed Care for Motivating High Risk Individuals Eligible for Evidence-based Prevention (GREAT) framework. The GREAT framework integrates PRSs to stratify disease risk and personalise interventions. By delivering personalised risk information, our goal is to empower patients to make informed decisions about their health, thereby increasing the uptake of evidence-based treatments. This personalised information may increase treatment engagement by activating mechanisms based on the

Capability-Opportunity-Motivation-Behaviour (COM-B) model and Theoretical Domains Framework (TDF).<sup>21-23</sup>

Our proposed approach will be implemented in two cluster randomised clinical trials, which will recruit individuals eligible for lung cancer screening and tobacco treatment. The first, PRECISE (“Precision Approaches to Lung Cancer Screening and Smoking Cessation Treatment in Primary Care”, NCT05627674), will evaluate the effectiveness of a multilevel intervention, *RiskProfile*, on increasing lung cancer screening and tobacco treatment utilization in primary care. The second trial, MOTIVATE (“Multilevel Intervention to Personalise and Improve Tobacco Treatment in Primary Care”, NCT05846841), will examine the effect of another multilevel intervention, *PrecisionTx*, on

promoting precision tobacco treatment in primary care. These trials aim to demonstrate how personalised risk assessments can significantly improve lung cancer prevention strategies and patient outcomes through PRS-enabled interventions.

**Methods**

**Study design**

*GREAT framework*

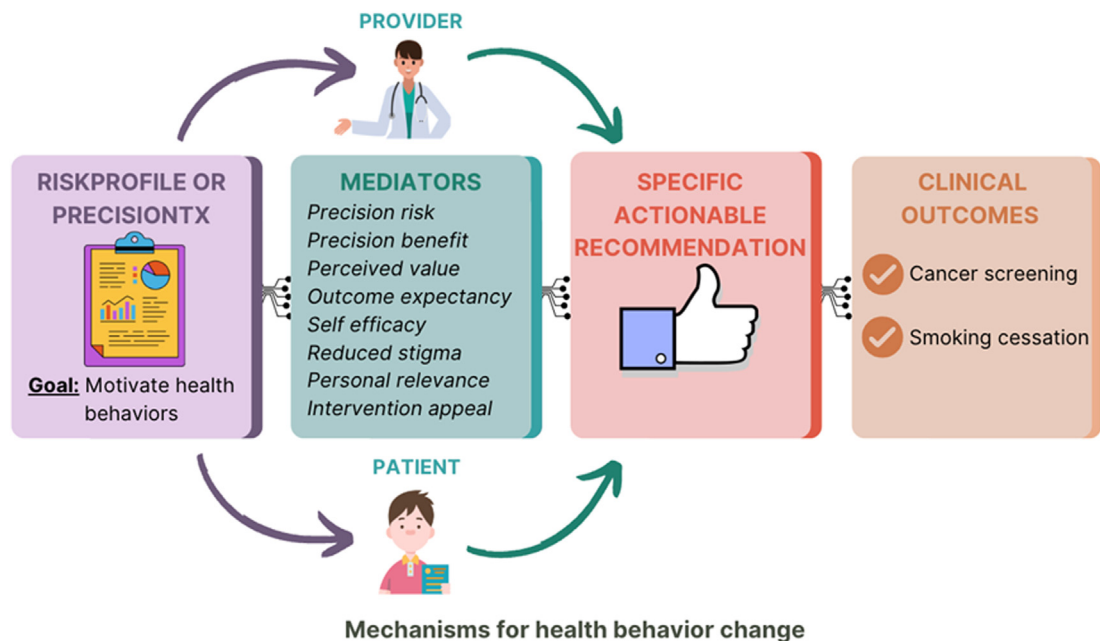
We have developed a GREAT framework to test whether personalised risk assessments can motivate high-risk patients who are eligible for cancer screening and tobacco treatment. The GREAT care paradigm is centred around multi-level intervention tools (*RiskProfile* and *PrecisionTx*) that utilise PRS to deliver precision patient risk (Fig. 1). Within the PRECISE and MOTIVATE trials, these PRS-based interventions will be communicated along with traditional clinical risk factors, such as smoking status and living with individuals who smoke.<sup>24-26</sup> The primary goal is to promote positive clinical outcomes, such as increased lung cancer screening rates, enhanced tobacco treatment adherence, and successful smoking cessation within primary care settings. Throughout the trials, we will evaluate shared

decision-making processes between patients and clinicians, focusing on actionable recommendations tailored to each patient’s risk profile. Additionally, we will assess intermediate health behaviour change mechanisms, including perceived benefit, self-efficacy, and outcome expectancy.

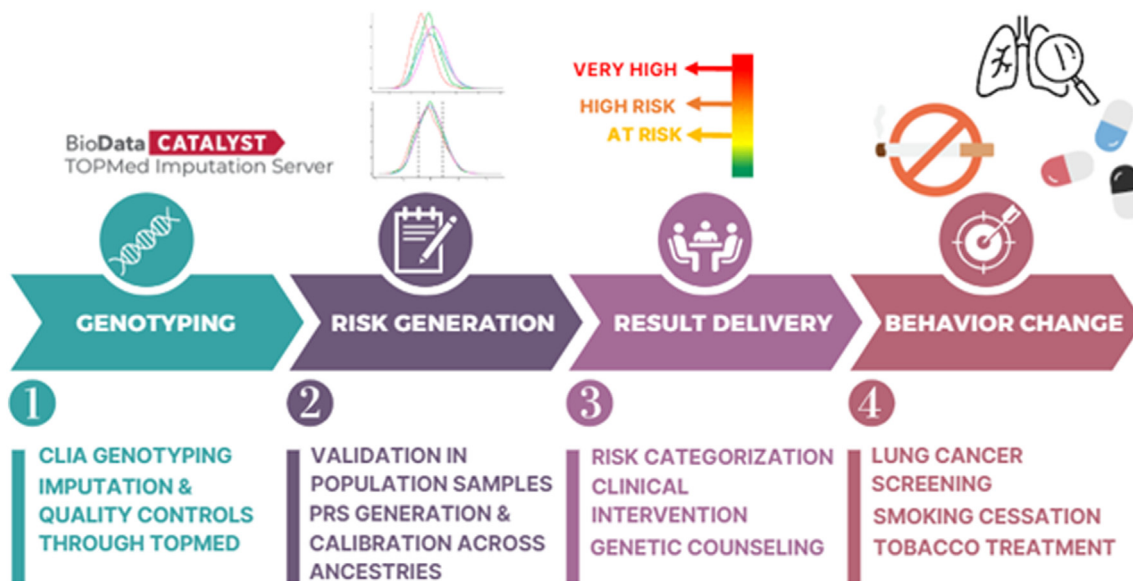
*Translation of genetic risk to clinical practice*

We present the GREAT paradigm within a broader roadmap for translating genetic risk into clinical practice (Fig. 2). This process begins with the Clinical Laboratory Improvement Amendments (CLIA) certified genotyping of enrolled participants’ genetic data, conducted by 23andMe. To ensure the integrity and reliability of the genetic data, imputation and quality controls were performed using the Trans-Omics for Precision Medicine (TOPMed) server, which also facilitates the imputation of GWAS variants.

Next, available GWAS variants and their corresponding weights were identified to create raw PRSs. These PRSs were then adjusted for genetic ancestry using reference data, such as the 1000 Genomes Project Phase 3 (1000G), and applied to validation datasets like the UK Biobank (UKBB) to establish risk categories and



**Fig. 1: Care Paradigm: Genomic Informed Care for Motivating High Risk Individuals Eligible for Evidence-based Prevention (GREAT).** The GREAT framework is a primary care paradigm that integrates genetic and clinical risk in precision health. Individuals and their providers in two upcoming trials (PRECISE and MOTIVATE) are enrolled and provided with multilevel interventions (e.g. RiskProfile and PrecisionTx) to promote clinical outcomes of lung cancer screening, tobacco treatment, and successful smoking cessation in primary care settings. Mechanisms of health behaviour changes (e.g. perceived benefit, self-efficacy, and outcome expectancy) will be evaluated. During the specific actionable recommendations phase, personalised shared decision-making will be facilitated by multilevel actions between patients and clinicians for better clinical outcomes.



**Fig. 2: Roadmap for translating genetic data to a genetic risk profile as a multilevel intervention in primary care.** In step 1, enrolled participants' genetic data are analysed by 23andMe's Clinical Laboratory Improvement Amendments (CLIA) certified genotyping process. Imputation and quality controls are conducted through the Trans-Omics for Precision Medicine (TOPMed) server to ensure the integrity and reliability of the genetic data, as well as to impute the GWAS variants. Step 2 involves identifying available GWAS variants and weights to create the raw Polygenic Risk Scores (PRS). The PRS is adjusted for genetic ancestry using reference data such as the 1000 Genomes Project Phase 3 and applied to validation data such as the UK Biobank to establish risk categories and compute ORs. In step 3, these scores are converted into 3 risk levels based on the established thresholds. In step 4, a report with precision treatment is created and communicated to both the participant and the provider to make informed and educated decisions. Behavioural interventionists (research staff who are trained, certified, and supervised by a team of genetic counsellor, psychologist, and psychiatrist) offer personalised guidance on behaviour change, leveraging the updated genetic insights. The outcome aims to increase lung cancer screening orders, improve participant adherence, promote smoking cessation, and highlight the benefits of tobacco treatment.

compute ORs. The PRSs were converted into three distinct risk levels based on these established thresholds.

Finally, these risk profiles were integrated into a multi-level intervention strategy, where precision treatment reports were generated and communicated to both participants and their healthcare providers. These reports serve as a foundation for informed, shared decision-making, aimed at motivating health behaviour changes such as increased lung cancer screening, improved participant adherence, and successful smoking cessation. Behavioural interventionists—trained, certified, and supervised by a multidisciplinary team including genetic counsellors, psychologists, and psychiatrists—offer personalised guidance to support behaviour change, leveraging the updated genetic insights.

### Ethics

This study was conducted using data from the Genetically Informed Smoking Cessation Trial (GISC),<sup>27</sup> UKBB,<sup>28</sup> AoU<sup>29</sup> and 1000G.<sup>30,31</sup> The 1000G was used for genetic ancestry inference. The GISC, UKBB, and AoU cohorts were used for PRS validation. The ethical approval for each dataset was obtained from the respective authorities as follows:

### GISC

The study was approved by the Institutional Review Board at Washington University in St Louis (Reference Number: 201305128).

### UKBB

Ethical approval was granted by the National Research Ethics Service Committee North West—Haydock (Reference Number: 11/NW/0382).

### AoU

The study was approved by the Institutional Review Board of the All of Us Research Program. Detailed information can be found [here](#).

### 1000G

Ethical approval was granted by the respective ethical review boards of the participating institutions. Detailed ethical considerations for each contributing institution can be found in the project's [documentation](#).

For UKBB, AoU and GISC, all participants provided written informed consent at the time of enrolment, which includes consent for genetic and health data usage in various research purposes. Participants in the

1000G provided written informed consent, allowing their genetic data to be used for research.

**Key datasets**

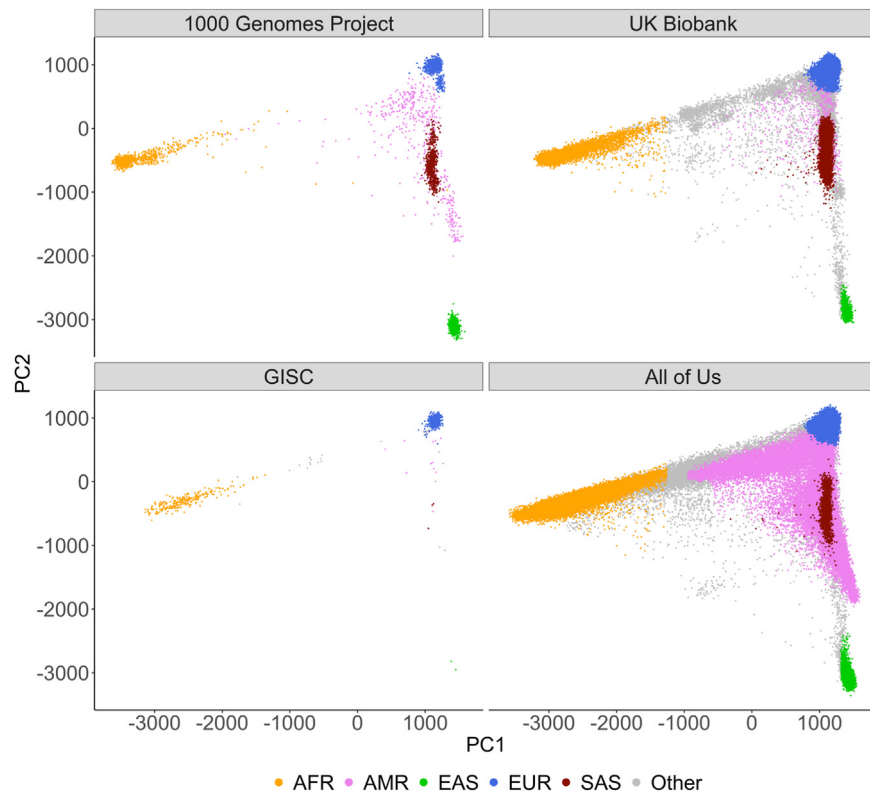
*1000 Genomes Project Phase 3 reference data and principal components analysis*

We use the 1000G<sup>30,31</sup> as a reference for genetic ancestry inference, as it is publicly accessible and includes genotype data from diverse populations. The 1000G dataset includes 3202 individuals, with 633 Europeans (EUR), 893 Africans (AFR), 585 East Asians (EAS), 601 South Asians (SAS), and 490 Admixed Americans (AMR). We conducted principal components analysis (PCA) in PLINK 2.0<sup>32</sup> on all 3202 samples, using 55,248 single-nucleotide polymorphisms (SNPs) from the recommended SNPs set by gnomAD<sup>33</sup> that were also found in the 1000G reference data, UKBB and GISC validation data, and 23andMe genotyping array used for the trial (Fig. 3, Supplementary Figure S1). From this PCA, we derived SNP weights that are used to compute ancestry PCs, so genotypes from different validation datasets could all be projected onto the same “PC space” (Supplementary Table S1).

We ran a random forest classifier to map individuals in 1000G to their respective populations using the first 5 PCs. We then projected our validation data to the 1000G-based “PC space” and applied the previously trained random forest classifier to produce “genetically inferred” ancestry labels. Individuals with a predicted probability less than 90% for any of the five ancestries were labelled as “Other.” While our proposed framework removes the need for labels in the clinic, we use them here for illustrative purposes in validation.

*Study dataset and PRS validation*

**GISC trial.** The GISC trial is a prospective, randomized, placebo-controlled trial conducted at Washington University in St. Louis, involving 822 current or previous smokers. Genetic data were available for 796 individuals, comprising 503 of EUR, 257 of AFR, and 36 individuals of “Other” self-reported ethnicity. The average age was 46.5 years (s.d. 11.3 years), with 54.4% female. While this dataset was not used for PRS prediction due to the small sample size and absence of lung cancer outcomes, it was utilized to validate the ancestry-adjustment procedure due to its resemblance to the



**Fig. 3: Cross-dataset distribution of genetic ancestry via PCA Projections in 1000G, GISC, UKBB, and AoU.** This figure illustrates the utility of principal components analysis (PCA) loadings obtained from the 1000 Genomes Project Phase 3 (1000G) in discriminating ancestries within external datasets, the Genetically Informed Smoking Cessation (GISC) trial. PCA was initially conducted on the globally diverse genotype data of 1000G. The resultant PCA-space was then used to project genotype data from GISC, UKBB, and AoU. The scatter plot displays the first and second PCs for each individual in these datasets, with points distinctly marked by genetically inferred ancestry.

expected populations in our PRECISE and MOTIVATE trials.

**UKBB.** The UKBB provides extensive genetic and clinical data from approximately 500,000 British individuals. Our analysis included 340,154 unrelated participants by genetically inferred ancestry: 6844 AFR, 730 AMR, 770 EAS, 313,279 EUR, 7197 SAS, and 11,334 Other. After excluding individuals with missing age or sex information, the average age of the cohort was 56.6 years (s.d. 8.2 years), with 54.1% female. The lung cancer analysis comprised 1830 cases (ICD10 codes C34.0–C34.9) and 338,334 controls, while the smoking cessation analysis involved 152,406 ever-smokers (117,483 former and 34,923 current). Sample characteristics regarding sex, age, and outcomes are provided in [Supplementary Table S2a and b](#).

**AoU.** The AoU dataset is focused on participants from diverse and historically under-represented backgrounds, with comprehensive genetic and health record data. Our study included 210,826 unrelated individuals with whole-genome sequencing data, comprising 45,108 AFR, 32,563 AMR, 3873 EAS, 110,712 EUR, 1689 SAS, and 16,881 Other by genetically inferred ancestry. After excluding those with missing age or sex information, the average age was 55.3 years (s.d. 14.5 years), with 59.7% female. We identified 1020 lung cancer cases using SNOMED codes and excluded secondary cancer cases. For the smoking analysis, we included 36,507 current and 116,409 former smokers. Additional details regarding sex, age, and outcomes can be found in [Supplementary Table 2c and d](#).

**Validation of Ancestry-Adjustment Procedure.** To evaluate the robustness of our PRS framework across different populations, we applied our procedure to the GISC, UKBB, and AoU datasets. The GISC trial, although not used for PRS prediction, was essential for validating the ancestry-adjustment procedure. In the UKBB and AoU datasets, we computed adjusted odds ratios (ORs) to assess the effectiveness of our risk stratification categories. Both datasets, with their rich genetic and clinical data, provided strong validation of our PRS model across diverse ancestry groups.

## Statistics

### Construction of polygenic risk scores

Our PRS models incorporate the latest findings by utilizing recently genome-wide association study (GWAS) summary statistics for lung cancer<sup>34</sup> and difficulty quitting smoking.<sup>35</sup> The lung cancer GWAS summary statistics were sourced from the International Lung Cancer Consortium (ILCCO), which includes 35,732 cases and 34,424 controls after excluding UKBB samples. The difficulty quitting smoking GWAS summary statistics

were derived from the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN), comprising about 1,193,150 individuals overall with 373,510 current smokers. Both sets of summary statistics specifically exclude UKBB samples to avoid overlapping with our validation data. While these GWAS summary statistics are derived from predominantly European ancestry, they also include a substantial proportion of non-European ancestry — about 26% for lung cancer and 21% for difficulty quitting smoking — which enhances the generalizability of the findings.<sup>36,37</sup>

For lung cancer risk, we started with 128 published SNPs found to be predictive of 5-year and lifetime cumulative risk for lung cancer.<sup>20</sup> Out of these, 101 SNPs overlapped with the published summary statistics, reference (1000G), and validation data (UKBB, GISC and AoU), and the 23andMe genotyping array used for the trial ([Supplementary Figure S1](#)). These SNPs were assigned effect sizes from the fixed-effect meta-analyses estimates in the latest lung cancer GWAS that includes EUR, AFR, and EAS ancestry.<sup>34</sup> For difficulty quitting smoking, we identified 175 SNPs predictive for smoking cessation following the same filtering procedure for lung cancer.<sup>35</sup>

The PRS construction began with the alignment of genotype data to the summary statistics, ensuring consistent PRS regardless of allele coding. Specifically, for any SNP  $G$  with reversed alleles, we recoded it as 2– $G$  to avoid discrepancies. The raw PRS for an individual  $i$  with  $M$  SNPs was computed as

$$PR S_i = \beta_1 G_{i1} + \beta_2 G_{i2} + \dots + \beta_M G_{iM}.$$

PRS calculations were performed using R, with genotype data input via the `genio` package.<sup>38</sup> PRS SNPs and weights for lung cancer and difficulty quitting smoking are provided in [Supplementary Tables S3–S4](#), respectively. We also used FAVOR (Functional Annotation of Variants Online Resources) to map each variant to a specific gene and functional annotation.<sup>39</sup>

### Standardizing PRS distributions across the continuum of genetic ancestry

We standardised the PRS distributions for lung cancer and difficulty quitting smoking using the 1000G dataset, employing a regression-based method to adjust for distributional differences across ancestries ([Supplementary Tables S5–S6](#)).<sup>40</sup> This adjustment process involves two key steps:

1. Mean adjustment: we conducted a linear regression of the raw PRS against the top five PCs derived from the PCA:

$$PR S_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \dots + \alpha_5 PC_{i5} + \epsilon_i^{mean}$$

We then computed residuals  $r_i$  of the raw PRS that account for mean differences in PRS distributions

across ancestry:

$$r_i = PR S_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 PC_{i1} - \dots - \widehat{\alpha}_5 PC_{i5}.$$

2. Variance adjustment: using the square residuals  $r_i^2$  as a proxy for PRS variance, we ran a secondary linear regression:

$$r_i^2 = \gamma_0 + \gamma_1 PC_{i1} + \gamma_2 PC_{i2} + \dots + \gamma_5 PC_{i5} + \epsilon_i^{var}$$

The final ancestry-adjusted PRS for each individual  $i$  was then computed as:

$$PRS_i^{adj} = \frac{PR S_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 PC_{i1} - \dots - \widehat{\alpha}_5 PC_{i5}}{\sqrt{\widehat{\gamma}_0 + \widehat{\gamma}_1 PC_{i1} + \widehat{\gamma}_2 PC_{i2} + \dots + \widehat{\gamma}_5 PC_{i5}}}$$

This standardization resulted in a PRS distribution with mean 0 and variance 1 across ancestries, ensuring that genetic risk is accurately reflected independent of ancestry. This method is crucial for individuals with admixed or unknown ancestry, where discrete ancestry-specific models are inappropriate.<sup>41,42</sup>

#### Definition of patient risk categories

We converted continuous genetic risks to categorical risk levels to support clinical communication and actionability.<sup>43,44</sup> The risk levels were designed for use in 2 ongoing trials in high-risk patients that evaluate the effect of personalised risks on motivating them to engage in cancer screening and tobacco treatment. Therefore, for lung cancer, we defined risk categories as the bottom 20%, middle 60%, and top 20% of the PRS distribution. These thresholds aligned with recent PRS risk stratification analyses for lung cancer.<sup>19</sup> It is important to note that all patients who receive this intervention are already at risk, due to factors such as heavy smoking or family history. Therefore, the bottom 20% is categorised as “at risk”, the middle 60% as “at high risk”, and the top 20% as “at very high risk”, which is expected to have approximately two-fold increased risk compared to the “at risk” group.

For difficulty quitting smoking, which is a behavioural phenotype related to substance dependence, we opted for a more agnostic approach by dividing the PRS distribution into equal thirds (top, middle, and bottom 33.3%). Since all patients receiving this intervention are active smokers, the bottom 33.3% are categorised as “at risk”, with the middle and top 33.3% categorised as “at high risk” and “at very high risk”, respectively.

To establish standardised thresholds, we computed PRS and ancestry PCs in the 1000G reference dataset. After running our ancestry-adjustment procedure, we identified percentiles for each ancestry-adjusted PRS distribution, which were applied to our UKBB and AoU validation data. These thresholds can also be used in future clinical applications with new validation datasets.

#### Adjusted odds ratios of genetic risk

We quantified patient risk as ORs of each outcome (lung cancer and difficulty quitting smoking) for individuals at “high risk” and “very high risk”, relative to individuals in the “at risk” category. Patients enrolled in either of our trials are considered at risk—for instance, meeting eligibility criteria for lung cancer screening or active tobacco smokers.<sup>26</sup> Therefore, we used the “at risk” category (i.e. “lowest genetic risk”) as the reference group to further identify individuals at particularly elevated genetic risk, who may benefit most from more comprehensive intervention and treatment strategies.

For each PRS model, we created a categorical variable  $PRS_{cat}$  taking values according to the three risk categories. Then, we ran logistic regressions with respect to  $PRS_{cat}$ , using “at risk” as the reference category. For smoking cessation analysis, we adjusted for age, sex, and 20 ancestry PCs. For lung cancer analysis, we additionally adjusted for smoking status (ever-smoker, never-smoker, or no-response). Each regression model yielded coefficients  $\widehat{\beta}$  corresponding to “high risk” and “very high risk”, and adjusted ORs for each PRS category (relative to “at risk”) are reported as  $e^{\widehat{\beta}}$ , along with corresponding 95% confidence intervals (CIs). These adjusted ORs will communicate the genetic risk of lung cancer or difficulty quitting smoking that is independent of the additional covariates. Within UKBB and AoU, we conducted these analyses across all ancestries combined, as well as within specific ancestry groups.

We also compared the risk stratification of our ancestry-adjusted PRS with “ancestry-matched” PRS. That is, we compared an individual’s raw PRS with the corresponding ancestry-specific raw PRS distribution in 1000G, i.e. European-only 1000G PRS distribution for individuals of genetically predicted European ancestry. For individuals with “Other” predicted ancestry, we used the overall raw PRS distribution among all 1000G samples.

#### Role of funders

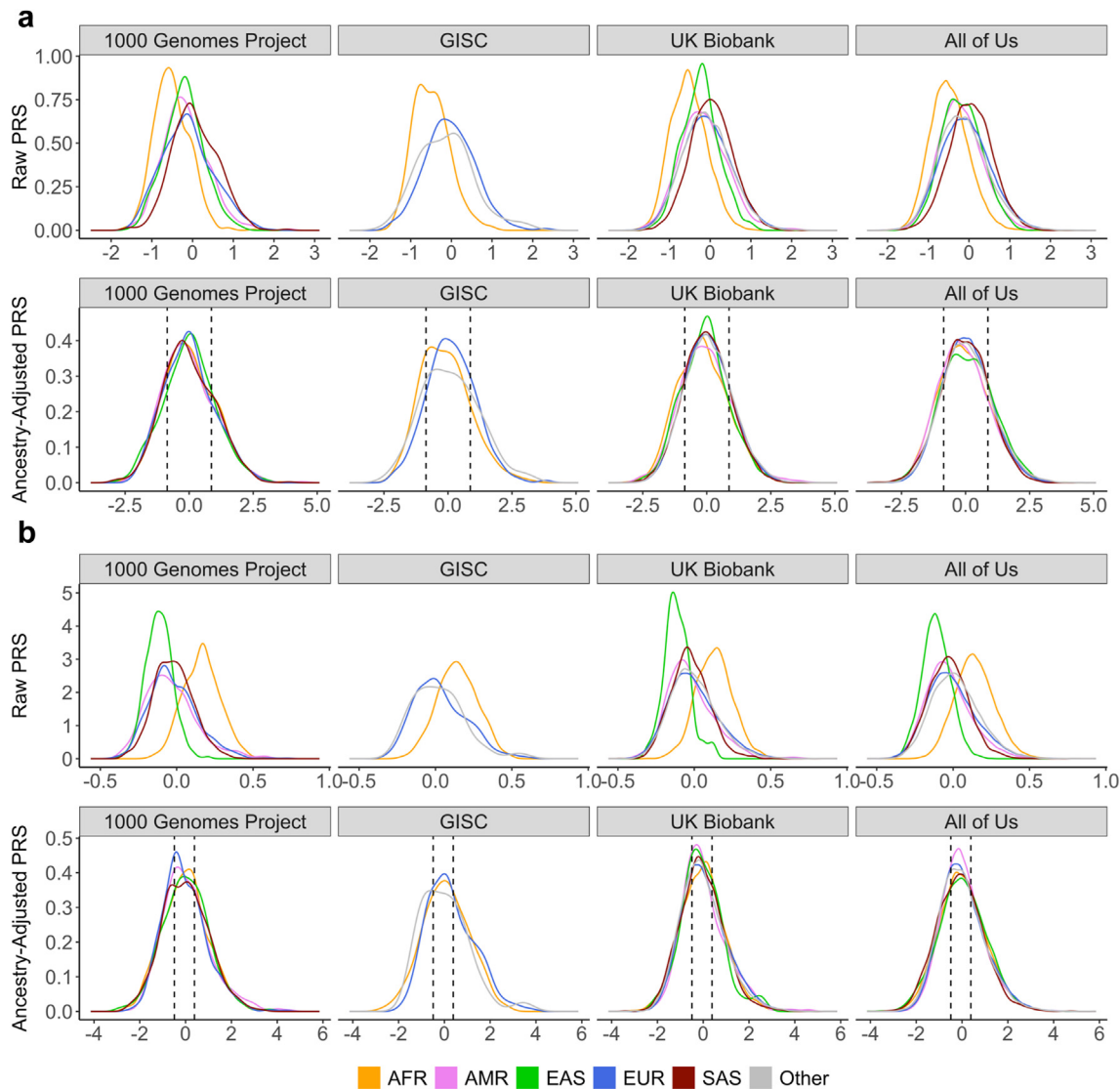
The funders played no role in the study design, analysis and collection of data, interpretation of results, or writing and submission of the paper.

## Results

### Harmonization PRS distributions across ancestry

We found notable variation in raw PRS distributions across ancestries, highlighting that applying a universal cutoff for raw PRS without accounting for ancestry can lead to biased risk profiling and inaccurate clinical recommendations (Fig. 4). However, our regression-based ancestry adjustment across all three datasets yields much more standardised distributions across ancestries. Specifically, the adjusted proportions of individuals within each risk category closely align with 20%-60%-20% for lung cancer, and 33.3%-33.3%-33.3% for



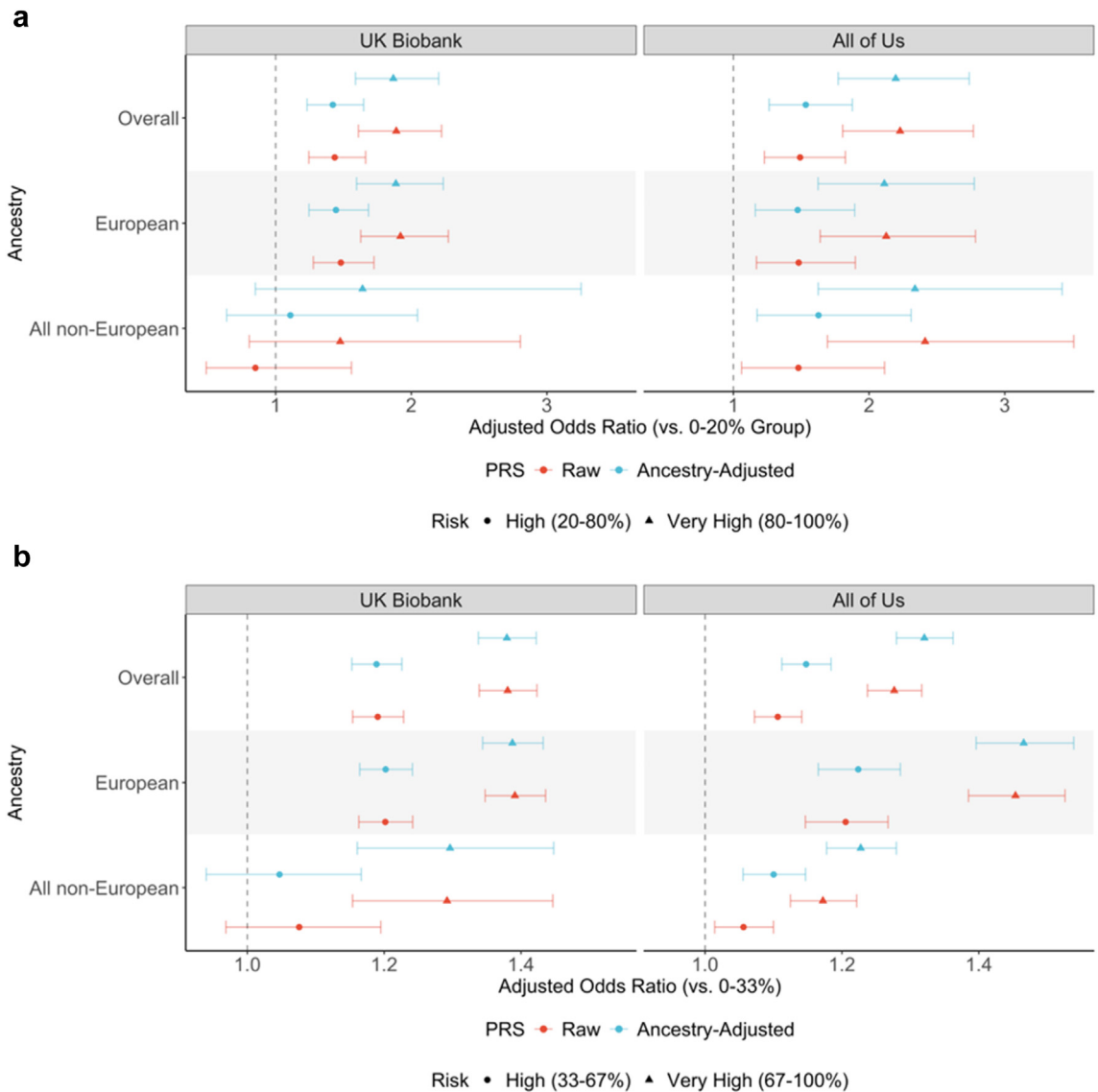


**Fig. 4: Ancestry adjustment of PRS for lung cancer and quit difficulty PRS across ancestral populations.** We showcase the adjustment process for PRS for (a) lung cancer and (b) difficulty quitting smoking within the 1000 Genomes Project, GISC Trial, UK Biobank, and All of Us datasets. It displays both raw and ancestry-adjusted PRS, with data points color-coded according to genetically inferred ancestries. EAS, AMR, and SAS ancestries were removed for GISC due to their small sample sizes. Ancestry adjustment effectively centres the PRS for different ancestries, mitigating the risk of incorrect stratification due to ancestry-related biases. Dotted vertical lines correspond to the 20th and 80th percentiles for lung cancer PRS distribution and 33rd and 67th percentiles for difficulty quitting smoking PRS among all 3202 samples in the 1000 Genomes Project.

difficulty quitting smoking. This adjustment ensures that patients of any background can be compared against a unified reference distribution for each outcome (Supplementary Table S6a–f). This standardisation places individuals across all ancestries on the same scale, allowing for a single risk stratification cutoff regardless of ancestral background. Such measures importantly enable fair risk assessment for individuals who may be labelled as “Other”, who may not fit well into a binned ancestry-specific risk model.

#### Risk stratification for lung cancer and smoking cessation in UKBB and AoU

After assigning UKB and AoU participants as “at risk”, “high risk”, or “very high risk” for lung cancer and difficulty quitting smoking, we identified significant ORs for both traits across different ancestry groups (Fig. 5, Supplementary Tables S7–S10). Our ancestry-adjusted PRS yielded similar ORs as using ancestry-matched distributions, demonstrating that a single PRS distribution can be appropriately applied to all



**Fig. 5: Risk stratification for lung cancer and difficulty quitting smoking using raw and ancestry-adjusted PRS.** This figure illustrates adjusted odds ratios with associated 95% confidence intervals of PRSs for (a) lung cancer and (b) difficulty quitting smoking among UK Biobank (N = 340,154 for lung cancer and N = 152,406 for difficulty quitting smoking), and All of Us (N = 210,826 for lung cancer and N = 152,916 for difficulty quitting smoking) participants. For difficulty quitting smoking, we adjusted for age, sex, and 20 ancestry PCs. For lung cancer, we additionally adjusted for smoking status (ever-smoker, never-smoker, or no-response). We compared risk stratification using a raw PRS with ancestry-matched percentiles, and our ancestry-adjusted PRS with the same percentiles for all individuals.

individuals. In UKBB, the overall adjusted ORs for lung cancer 1.42 (95% CI: 1.23–1.65) for the “high risk” group and 1.87 (95% CI: 1.59–2.20) for “very high risk” group compared to the “at risk” group (Supplementary Table S7a). In AoU, these adjusted ORs were slightly higher—1.49 (95% CI: 1.23–1.83) for “high risk” group and 2.23 (95% CI: 1.81–2.77) for “very high risk” group. While the adjusted ORs in non-European ancestries within UKBB were not significant, they were significant and even higher than those for European ancestries

within AoU (Supplementary Table S8a). Moreover, the ORs for “Other” ancestry group in AoU were much higher and more significant using the ancestry-adjusted PRS compared to raw PRS, where an ancestry-matched risk distribution is least suitable (Supplementary Table S8b).

We observed similar patterns for difficulty quitting smoking, where the ancestry-adjusted PRS performed similar or better than using ancestry-matched distributions for raw PRS. In UKBB, the overall adjusted ORs

for difficulty quitting smoking were 1.19 (95% CI: 1.15–1.23) for “high risk” group and 1.38 (95% CI: 1.34–1.42) for “very high risk” group (Supplementary Table S9a). In AoU, the adjusted ORs were 1.19 (95% CI: 1.15–1.23) for “high risk” group and 1.38 (95% CI: 1.34–1.42) for “very high risk” (Supplementary Table S10a) group. Notably, we observed slightly higher and more significant ORs in non-European AoU participants when using the ancestry-adjusted PRS compared to the raw PRS. Although the improvement within each specific ancestry group was not as substantial, the aggregate ORs in all non-European groups increased from 1.06 (95% CI: 1.01–1.10) to 1.10 (95% CI: 1.06–1.15) in the “high risk” group, and from 1.17 (95% CI: 1.12–1.22) to 1.23 (95% CI: 1.18–1.28) in the “very high risk” group.

Using ancestry-adjusted PRS ensures accurate risk stratification across all ethnic backgrounds, a critical consideration given the substantial variability in raw PRS distributions across diverse populations. The outcome-based validation in UKBB and AoU further verifies that the ancestry-adjusted PRS provides valid risk stratification, and often yields better risk stratification for non-European ancestries compared to ancestry-matched modelling. These findings collectively facilitate a more robust and standardised application of PRS in clinical reporting.

### Translating genetic risk into clinical reports

We have implemented our analytic framework in two recently launched trials—PRECISE and MOTIVATE—which are currently in the preliminary phases of recruitment and aim to engage over 100 physicians and 1600 patients. These trials are designed to promote health behaviour change using genetically informed multi-level interventions, RiskProfile and PrecisionTx, respectively. These interventions incorporate PRS to communicate precision risk of lung cancer and precision benefits of smoking cessation, promoting evidence-based practices such as cancer screening and tobacco treatment in individuals who smoke and/or are eligible for lung cancer screening. Based on functional information from FAVOR, our lung cancer PRS contains putative loss-of-function variants in genes such as *CHRNA5*<sup>45</sup> and *CHEK2* (Checkpoint kinase 2),<sup>46</sup> which are known to be involved in nicotine addiction and DNA damage repair, respectively (Supplementary Table S3). Our PRS for difficulty quitting does not contain similar loss-of-function variants, but still includes variants in smoking-related genes such as *DRD2* (Dopamine receptor D2)<sup>47</sup> and *UBXN2A* (UBX domain 2A),<sup>48</sup> which are associated with dopaminergic signaling and stress response mechanisms (Supplementary Table S4). These biological insights reinforce the medical relevance of our PRS models by linking genetic variants to key pathways involved in smoking-related diseases.

Most preventive practices such as lung cancer screening and tobacco treatment are still largely underutilised in primary care. Thus, we aim to evaluate whether personalised genetic risk increases provider practices on and patient engagement of these preventive practices. Access to 23andMe genotypes and expanded health information has been a motivating component for the research participants. Once patients provide samples for genotyping, we generate personalised risk profiles for both primary care providers and patients to support behaviour change. We inform patients about their genetic risk category according to the ancestry-adjusted PRS distributions, along with the ORs relative to the “at risk” group. We also provide personalised clinical risks using independent risk models and actionable recommendations to motivate cancer screening and tobacco treatment (Figs. 6 and 7).<sup>24–26</sup> This approach is designed to integrate seamlessly into routine diagnostic workflows, enhancing the utilization of preventive services in primary care.

Through our approach, we propose clear and patient-friendly communication strategies, including visual aids and educational materials, to facilitate understanding and meaningful interactions between patients and healthcare providers. Effectively communicating both the risk and precision of the PRS results is challenging but essential to empower patients to make informed decisions about their health. Moreover, it is crucial to consider patient perceived risk, perceived benefit, and personal relevance when discussing PRS results with patients. Patients’ understanding and interpretation of PRS may vary, leading to differing levels of engagement in preventive actions. Hence, comprehensive patient education programs can enhance awareness and knowledge about PRS, its implications, and available preventive measures.

### Discussion

In this study, we introduce a translational roadmap and analytical framework for implementing PRS within multilevel interventions to communicate precision risk and benefit and ultimately promote health behaviour change. Specifically, we frame our translational message specifically for patients at high risk who have not received guideline-recommended cancer screening or tobacco treatment,<sup>49–52</sup> taking special care to ensure inclusion of and fair risk assessment across diverse ancestries via PC-regression-based PRS adjustment. This framework will be evaluated in two cluster-randomised trials to evaluate the effectiveness of comprehensive risk profiles in motivating positive health decision-making from both patients and primary care providers.

A key feature of our framework is standardisation of PRS distributions across diverse ancestries using widely accessible data from the 1000G dataset, as an alternative to methods in the GenoVA (15) and eMERGE (14)



# RiskProfile: Genetic

MPI: Alex Ramsey, PhD  
 MPI: Li-Shiun Chen, MD  
 Phone: 314-747-7849

SMOKING STATUS: Current	REPORT TYPE: Genetic Risks and Benefits
TEST PERFORMED: Polygenic Risk Score	NOTE: This is a research report

\*Polygenic risk calculated using data from mostly European ancestry individuals. Results are known to be less accurate for non-European ancestry individuals.

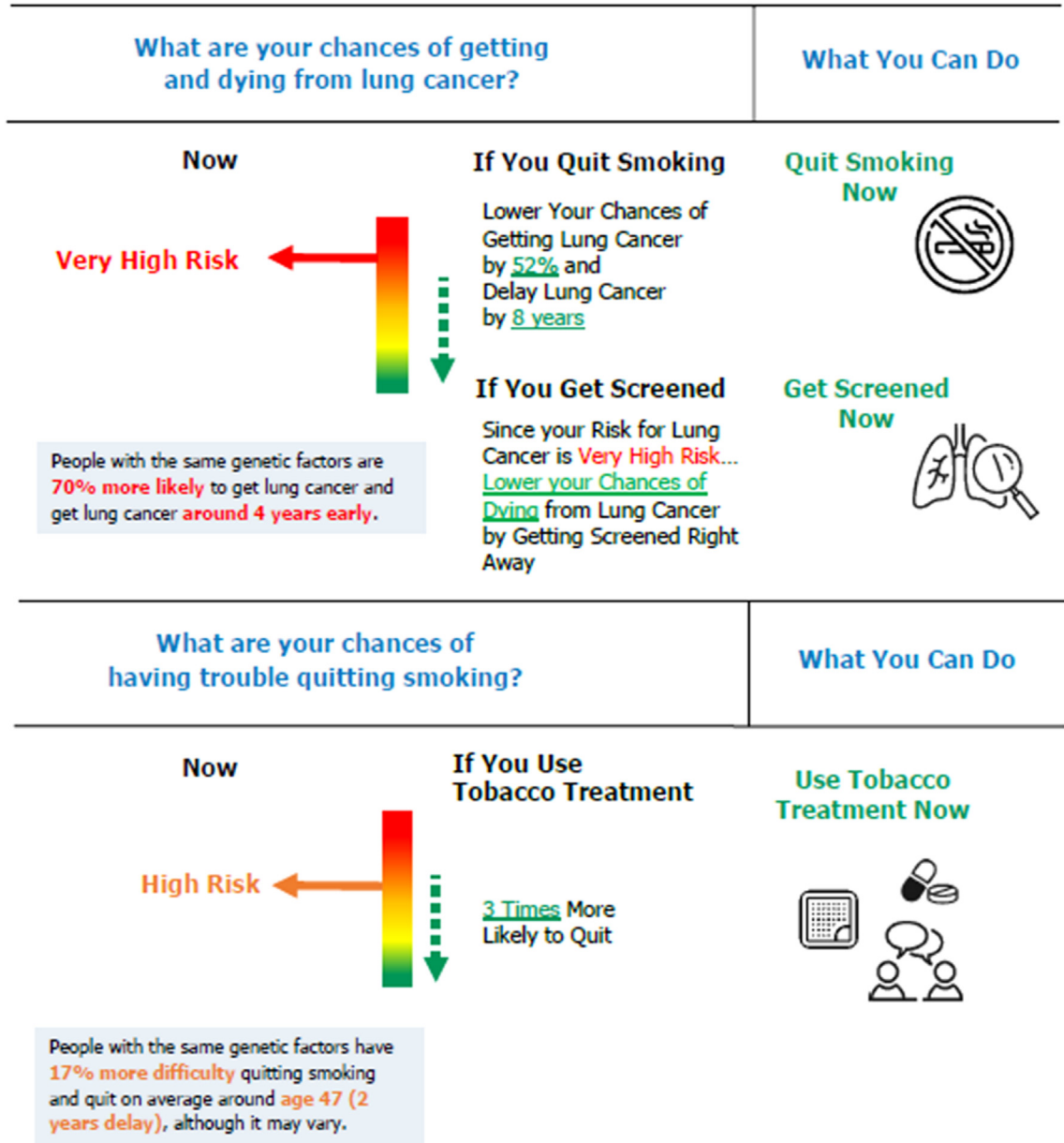


Fig. 6: Example 1 clinical report: RiskProfile. We present example 1 for genomically informed interventions using the GREAT framework. RiskProfile is designed to motivate lung cancer screening and tobacco treatment among screening-eligible patients. This intervention utilises ancestry-adjusted PRS to stratify patients into “at risk” (yellow), “high risk” (orange), and “very high risk” (red) genetic risk categories. RiskProfile focuses on prevention and expands beyond personalised risk to also provide personalised benefit of cancer screening and use a multilevel intervention design directed to both physicians and patients in clinical settings. In our PRECISE trial (NCT05627674), the effect of RiskProfile on clinician ordering and patient completion of lung cancer screening will be evaluated.



**Precision Treatment**  
*PrecisionTx*

MP: Li-Shan Chen, MD & Alex Ramsey, PhD MOTIVATE Trial,  
314.273.3826  
ID 000, Research Report

**Personalized Benefit of Quitting Smoking**

**LUNG CANCER RISK IN NEXT 6 YEARS**

**WHAT CAN YOU DO NOW?**



**Quit Smoking Now**



Lower your risk of lung cancer by 50% and delay by 8 years by quitting smoking now

**Personalized Benefit of Taking Tobacco Treatment**

**DIFFICULTY QUITTING RISK**

**WHAT CAN YOU DO NOW?**



**Take Tobacco Treatment Now**

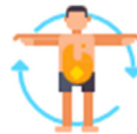


Take Tobacco Treatment now to improve your chances by 3-5 times

**Personalized Medicine**



Your Genetic Marker:  
rs16969968 GG/AG/AA



Your Nicotine Metabolism:  
Slow

**Good News! Based on your biological makeup, you will benefit most from:**

**Varenicline**

**Most Effective & Fewest Side Effects for You**

If you cannot take Varenicline, talk to your provider about other smoking cessation methods to help you quit

Fig. 7: Example 2 clinical report: PrecisionTx. We present example 2 for genomically informed interventions using the GREAT framework. PrecisionTx is designed to motivate tobacco treatment among patients who smoke. This intervention utilises ancestry-adjusted PRS to stratify

studies that use data from the Mass-General Brigham Biobank and AoU, respectively. We validated the transferability of our 1000G-based standardization in external datasets from the UKBB, GISC and AoU, allowing future trials to adopt a similar methodology irrespective of their specific genetic data. Furthermore, from large-scale biobank analysis, we verified that our ancestry-adjusted PRSs achieved similar or better risk stratification than naïve ancestry-matching, which importantly enables the inclusion of individuals of mixed or uncertain ancestry. By utilizing our provided PC loadings and PRS standardization formula for lung cancer and difficulty quitting smoking, new patients in these trials can receive accurate risk categorization reports, bypassing the inaccuracies of self-reported ethnicity and the need for re-training PCA models.

Unlike most current research that evaluates PRS-enabled interventions in general patient populations, our work uniquely focuses on designing and evaluating these interventions specifically among patients already at high risk due to factors like smoking or family history. While these individuals will benefit tremendously from lung cancer screening and smoking cessation, they may not be fully motivated to quit smoking or make other positive health changes from general medical advice alone. Therefore, we believe that within this context of a high-risk patient population, introducing an additional dimension of genetic risk may further encourage individuals to follow through with cancer screening or tobacco treatment. Importantly, we aim to follow best practices of communicating uncertainties and potential imprecision in risk estimates and thresholds to maintain transparency with patients.

Our work has several limitations, but we hope to contribute to the knowledge pool for the best practices in creating PRS-enabled interventions that may be disease-, population-, or context-specific. Our approach is tailored for unique outcomes, populations, and contexts to optimise health impact. While our framework is broadly applicable within primary care, the specific interplay between genetic and clinical risk, as well as modes of communicating and addressing patient risk, will vary across specific diseases and contexts.<sup>53,54</sup> For example, the individuals enrolled in our two trials share similar clinical profiles in line with eligibility for lung cancer screening and tobacco treatment. For other outcomes or populations of interest, domain expertise to identify the key baseline characteristics for high clinical risk is crucial to best communicate comprehensive risk and expected benefits from behaviour change.

Our PRS model also has two key limitations. First is the underrepresentation of non-European populations in the multi-ancestry GWAS used to derive the PRS weights, a challenge that persists across existing GWASs,<sup>55–58</sup> not limited to lung cancer and smoking cessation. This underrepresentation may reduce the predictive power of the PRS in non-European populations. However, with ongoing efforts to recruit individuals of diverse ancestry in genetic studies, we expect this challenge to be resolved in the coming years.

In addition, our current PRS models were constructed from a small set of significant predictive variants. While the PRSs include variants within genes associated with lung cancer and smoking,<sup>45–48</sup> only a small number were coding. Incorporating more biological information<sup>59,60</sup> into the model-building process, beyond evaluating genome-wide significance, may further enhance the interpretability and accuracy of the genetic risk models. Furthermore, with the development of new PRS approaches and toolkits to enhance predictive power in diverse populations from multi-ancestry data,<sup>42,61–64</sup> we can iteratively refine PRS implementation in our trial to synchronise with the latest advancements. Maintaining a dynamic PRS framework that aligns with the latest advancements will reduce the implementation gap and maximise impact in preventive healthcare outcomes. Following current genetic counselling recommendations, we have established a process to incorporate new evidence into our intervention for smoking cessation and lung cancer risk. This process will adjudicate new population-specific evidence on genetics and biomarkers, evaluating its impact on personal and population-level risk changes, and effectively communicating the dynamic nature of genetic evidence to patients and providers.

Looking to the future, we must also consider the scalability of precision interventions to real-world clinics and studies. A notable gap in current practice is the absence of genetic information in electronic health records (EHRs) for decision support and the lack of PRS generation in clinical labs. Implementing precision interventions in primary care necessitates a workflow that incorporates EHRs for recruitment, biomarker testing protocols, and standardised processes to generate personalised intervention reports.<sup>65</sup> This requires collaborations with primary care stakeholders, community advisory boards, genetic counselling, and health communication to improve intervention clarity, accuracy, and impact.<sup>8,9,66–68</sup> To reduce burden, we need to leverage existing EHR tools (e.g. Best Practice Advisories) and training to efficiently facilitate physician prescribing.<sup>69,70</sup>

---

patients into “at risk” (yellow), “high risk” (orange), and “very high risk” (red) genetic risk categories. PrecisionTx focuses on treatment and expands beyond personalised risk to also provide personalised benefit of tobacco treatment and use a multilevel intervention design directed to both physicians and patients in clinical settings. In our MOTIVATE trial (NCT05846841), the effect of PrecisionTx on clinician ordering, patient adherence, and smoking abstinence will be evaluated.

Understanding of mechanistic and implementation outcomes will guide scalable, efficient delivery components for integration into clinic workflows,<sup>65</sup> using trained embedded staff, and digital therapeutic tools to enable these PRS-informed behavioural interventions.<sup>71</sup>

In conclusion, we provide a roadmap that incorporates PRSs using multi-ancestry GWAS-based weights, translates risk into actionable categories, communicates comprehensive risk effectively, considers patient perspectives, and accommodates evolving science is essential for the equitable and pragmatic translation of PRS into clinical care. By addressing the barriers and implementing potential solutions at each stage, we can leverage PRS to improve preventive healthcare and significantly reduce the burden of lung cancer.

#### Contributors

T.C., H.Z., and L.C. conceived the project, T.C.; G.P., and L.F. carried out all data analyses under the supervision of H.Z. and L.C.; N.A. and X.W. organised data in AoU; G.S. and D.J. provided GWAS summary statistics for difficulty quitting smoking. T.C.; G.P., H.Z., and L.C. drafted the manuscript. All authors reviewed and approved the final version of the manuscript.

#### Data sharing statement

Genotype data from 1000G are publicly available and can be directly downloaded from the following links:

1. Genotype data: <https://www.cog-genomics.org/plink/2.0/resources>.
2. Population information: <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>.

[https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000G\\_2504\\_high\\_coverage/1000G\\_698\\_related\\_high\\_coverage.sequence.index](https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/1000G_698_related_high_coverage.sequence.index).

UKBB genotype and phenotype data are available via application through the UK Biobank website: <https://www.ukbiobank.ac.uk/>.

AoU genotype and phenotype data are available via application through the All of Us Research Program: <https://www.researchallofus.org/>.

GISC data is available via application through NIDA Center for Genetic Studies: <https://nidagenetics.org/>

No additional data were collected for this study. All results are presented in the paper's tables and figures.

#### Code sharing

R code and plink commands, as well as accompanying data, used for analysis are provided in a walkthrough available on GitHub at <https://github.com/chen-tony/GREAT>.

#### Declaration of interests

Laura J. Bierut (LJB) is listed as an inventor on Issued U.S. Patent 8,080,371, "Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction, LJB receives consulting fees from Research Triangle Institute for grant R01DA048824 "Identifying blood-based DNA methylation biomarkers of cannabis use" is a member of US Food and Drug Administration Tobacco Products Scientific Advisory Committee, and co-chair of National Comprehensive Cancer Network Smoking Cessation Panel. Michael J. Bray (MJB) was an employee at ThinkGenetic, Inc, where he had the option to receive stock options at the time the work was conducted. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

All other authors have no conflict of interests to report.

#### Acknowledgements

We would like to thank Reeya Joseph for her editorial support with the introduction, Peter Kraft for his advice on our manuscript, and Scott

Vrieze for his assistance with summary statistics for difficulty quitting smoking. We would also like to thank and acknowledge the participants enrolled in the UK Biobank (obtained under UK Biobank resource application 52008) and GISC trial for contributing vital data to this work.

This research was supported by NIH Training Grant T32GM135117 and NSF Graduate Research Fellowship DGE-2140743 (T.C.), R01HG011035, R01ES036042, R01HL173869 (D.L.), National Cancer Institute (NCI) R01CA268030, NIDA R01DA056050, National Institute on Drug Abuse (NIDA) R34DA052928, NIDA K12DA041449, Taylor Family Institute for Innovative Psychiatric Research (A.T.R.), NCI R01-CA268030, NIDA R01-DA056050 (L.J.B.), R35-3CA197449, R01-HL163560, U01-HG009088, and U01-HG012064 (X.L.), NIH Intramural Research Program (H.Z.), NIH 5T32-HL007776-25, R01-DA056050, R01-CA268030, P30-CA091842-19S5, P30-CA091842-16S2 and P50-CA244431 (L.C.) and NCI grant U19-CA203654 (Integrative Analysis of Lung Cancer Etiology and Risk Application and Translation).

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105441>.

#### References

- 1 Kocarnik JM, Compton K, Dean FE, et al. Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019. *JAMA Oncol.* 2022;8(3):420–444.
- 2 Wang X, Romero-Gutierrez CW, Kothari J, Shafer A, Li Y, Christiani DC. Prediagnosis smoking cessation and overall survival among patients with non-small cell lung cancer. *JAMA Netw Open.* 2023;6(5):e2311966.
- 3 Twyman L, Bonevski B, Paul C, Bryant J. Perceived barriers to smoking cessation in selected vulnerable groups: a systematic review of the qualitative and quantitative literature. *BMJ Open.* 2014;4(12):e006414.
- 4 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19:581–590.
- 5 Lewis CM, Vassos E. *Polygenic risk scores: from research tools to clinical instruments. Vol. 12, Genome medicine.* BioMed Central Ltd.; 2020.
- 6 Adeyemo A, Balaconis MK, Darnes DR, et al. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med.* 2021;27:1876–1884.
- 7 Chiu A, Hartz S, Smock N, et al. Most current smokers desire genetic susceptibility testing and genetically-efficacious medication. *J Neuroimmune Pharmacol.* 2018;13(4):430–437.
- 8 Ramsey AT, Bray M, Acayo Laker P, et al. Participatory design of a personalized genetic risk tool to promote behavioral health. *Cancer Prev Res.* 2020;13(7):583–592.
- 9 Ramsey AT, Bourdon JL, Bray M, et al. Proof of concept of a personalized genetic risk tool to promote smoking cessation: high acceptability and reduced cigarette smoking. *Cancer Prev Res.* 2021;14(2):253–262.
- 10 Quaife SL, Janes SM, Brain KE. The person behind the nodule: a narrative review of the psychological impact of lung cancer screening. *Transl Lung Cancer Res.* 2021;10(5):2427–2440.
- 11 Quaife SL, Waller J, Dickson JL, et al. Psychological targets for lung cancer screening uptake: a prospective longitudinal cohort study. *J Thorac Oncol.* 2021;16(12):2016–2028.
- 12 Quaife SL, Marlow LAV, McEwen A, Janes SM, Wardle J. Attitudes towards lung cancer screening in socioeconomically deprived and heavy smoking communities: informing screening communication. *Health Expect.* 2017;20(4):563–573.
- 13 Quaife SL, McEwen A, Janes SM, Wardle J. Smoking is associated with pessimistic and avoidant beliefs about cancer: results from the International Cancer Benchmarking Partnership. *Br J Cancer.* 2015;112(11):1799–1804.
- 14 Linder JE, Allworth A, Bland HT, et al. Returning integrated genomic risk and clinical recommendations: the eMERGE study. *Genet Med.* 2023;25(4):100006.
- 15 Hao L, Kraft P, Berriz GF, et al. Development of a clinical polygenic risk score assay and reporting workflow. *Nat Med.* 2022;28(5):1006–1013.

- 16 Shieh Y, Eklund M, Madlensky L, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst.* 2017;109(5):djw290.
- 17 Zhang P, Chen PL, Li ZH, et al. Association of smoking and polygenic risk with the incidence of lung cancer: a prospective cohort study. *Br J Cancer.* 2022;126(11):1637–1646.
- 18 Kanwal M, Ding XJ, Cao Y. Familial risk for lung cancer. *Oncol Lett.* 2017;13(2):535–542.
- 19 Zhu M, Lv J, Huang Y, et al. Ethnic differences of genetic risk and smoking in lung cancer: two prospective cohort studies. *Int J Epidemiol.* 2023;52(6):1815–1825.
- 20 Hung RJ, Warkentin MT, Brhane Y, et al. Assessing lung cancer absolute risk trajectory based on a polygenic risk model. *Cancer Res.* 2021;81(6):1607–1615.
- 21 Michie S. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care.* 2005;14(1):26–33.
- 22 Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implement Sci.* 2012;7(1):37.
- 23 Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci.* 2017;12(1):77.
- 24 Bray M, Chang Y, Baker TB, et al. The promise of polygenic risk prediction in smoking cessation: evidence from two treatment trials. *Nicotine Tob Res.* 2022;24(10):1573–1580.
- 25 Chen LS, Baker TB, Piper ME, et al. Interplay of genetic risk (CHRNA5) and environmental risk (partner smoking) on cigarette smoking reduction. *Drug Alcohol Depend.* 2014;143:36–43.
- 26 Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med.* 2013;368(8):728–736.
- 27 Chen L, 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.* 2020;108(6):1315–1325.
- 28 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
- 29 Bick AG, Metcalf GA, Mayo KR, et al. Genomic data in the all of us research program. *Nature.* 2024;627:340–346.
- 30 Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
- 31 Byrka-Bishop M, Evani US, Zhao X, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell.* 2022;185(18):3426–3440.e19.
- 32 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4(1):7.
- 33 Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature.* 2024;625(7993):92–100.
- 34 Byun J, Han Y, Li Y, et al. Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet.* 2022;54(8):1167–1177.
- 35 Saunders GRB, Wang X, Chen F, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature.* 2022;612(7941):720–724.
- 36 Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017;100(4):635–649.
- 37 Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med.* 2016;375(7):655–665.
- 38 Ochoa A. Genio: genetics input/output functions. Available from: <https://CRAN.R-project.org/package=genio>; 2023.
- 39 Zhou H, Arapoglou T, Li X, et al. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. *Nucleic Acids Res.* 2023;51(D1):D1300–D1311.
- 40 Ge T, Irvin MR, Patki A, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med.* 2022;14(1):70.
- 41 Lewis ACF, Molina SJ, Appelbaum PS, et al. Getting genetic ancestry right for science and society. *Science.* 2022;376(6590):250–252.
- 42 Kachuri L, Chatterjee N, Hirbo J, et al. Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet.* 2023;24.
- 43 Belkora J, Moore DH, Hutton DW. Assessing risk communication in breast cancer: are continuous measures of patient knowledge better than categorical? *Patient Educ Couns.* 2009;76(1):106–112.
- 44 Lautenbach DM, Christensen KD, Sparks JA, Green RC. Communicating genetic risk information for common disorders in the era of genomic medicine. *Annu Rev Genomics Hum Genet.* 2013;14(1):491–513.
- 45 Kraiss AM, Hautefeuille AH, Cros MP, et al. CHRNA5 as negative regulator of nicotine signaling in normal and cancer bronchial cells: effects on motility, migration and p63 expression. *Carcinogenesis.* 2011;32(9):1388–1395.
- 46 Cybulski C, Masojc B, Oszutowska D, et al. Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers. *Carcinogenesis.* 2008;29(4):762–765.
- 47 Comings DE, Ferry L, Bradshaw-Robinson S, Burchette R, Chiu C, Muhleman D. The dopamine D2 receptor (DRD2) gene: a genetic risk factor in smoking. *Pharmacogenetics.* 1996;6(1):73–79.
- 48 Teng Y, Rezvani K, De Biasi M. UBXLN2A regulates nicotinic receptor degradation by modulating the E3 ligase activity of CHIP. *Biochem Pharmacol.* 2015;97(4):518–530.
- 49 US preventive services task force. <https://www.uspreventiveserVICEStaskforce.org/uspstf/recommendation/lung-cancer-screening>; 2021. Lung Cancer.
- 50 Agency for healthcare research and quality; 2021. <https://www.ahrq.gov/prevention/guidelines/index.html> [Clinical Guidelines and Recommendations].
- 51 Tobacco Use and Dependence Guideline Panel, US Department of Health and Human Services. *Tobacco use and dependence guideline Panel. Treating tobacco use and dependence: 2008 update.* Rockville, MD: US Department of Health and Human Services; 2008.
- 52 Krist AH, Davidson KW, Mangione CM, et al. Screening for lung cancer. *JAMA.* 2021;325(10):962–970.
- 53 Arem H, Loftfield E. Cancer epidemiology: a survey of modifiable risk factors for prevention and survivorship. *Am J Lifestyle Med.* 2018;12(3):200–210.
- 54 Tran KB, Lang JJ, Compton K, et al. The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* 2022;400(10352):563–591.
- 55 Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 2019;179(3):589–603.
- 56 Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538(7624):161–164.
- 57 Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 2009;25(11):489–494.
- 58 Fitipaldi H, Franks PW. Ethnic, gender and other sociodemographic biases in genome-wide association studies for the most burdensome non-communicable diseases: 2005–2022. *Hum Mol Genet.* 2023;32(3):520–532.
- 59 Long E, Patel H, Golden A, et al. High-throughput characterization of functional variants highlights heterogeneity and polygenicity underlying lung cancer susceptibility. *Am J Hum Genet.* 2024;111(7):1405–1419.
- 60 Wang G, Zhou H, Strulovici-Barel Y, et al. Role of OSGIN1 in mediating smoking-induced autophagy in the human airway epithelium. *Autophagy.* 2017;13(7):1205–1220.
- 61 Zhang H, Zhan J, Jin J, et al. A new method for multi-ancestry polygenic prediction improves performance across diverse populations. *Nat Genet.* 2023;25.
- 62 Jin J, Zhan J, Zhang J, et al. MUSSEL: enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups. *Cell Genomics.* 2024;4(4):100539.
- 63 Zhang J, Zhan J, Jin J, et al. An ensemble penalized regression method for multi-ancestry polygenic risk prediction. *Nat Commun.* 2024;15(1):3238.
- 64 Hou K, Gogarten S, Kim J, et al. Admix-kit: an integrated toolkit and pipeline for genetic analyses of admixed populations. *Bioinformatics.* 2024;40(4).
- 65 Ayatollahi H, Hosseini SF, Hemmat M. Integrating genetic data into electronic health records: medical geneticists' perspectives. *Health Inform Res.* 2019;25(4):289.
- 66 Bourdon JL, Dorsey A, Zalik M, et al. In-vivo design feedback and perceived utility of a genetically-informed smoking risk tool among current smokers in the community. *BMC Med Genomics.* 2021;14(1):139.
- 67 Ramsey AT, Chen LS, Hartz SM, et al. Toward the implementation of genomic applications for smoking cessation and smoking-related diseases. *Transl Behav Med.* 2018;8(1):7–17.



- 
- 68 Chen LS, Baker TB, Ramsey A, Amos CI, Bierut LJ. Genomic medicine to reduce tobacco and related disorders: translation to precision prevention and treatment. *Addiction Neuroscience*. 2023;7:100083.
- 69 Chen LS, Baker TB, Korpecki JM, et al. Low-burden strategies to promote smoking cessation treatment among patients with serious mental illness. *Psychiatr Serv*. 2018;69(8):849–851.
- 70 Ramsey AT, Chiu A, Baker T, et al. Care-paradigm shift promoting smoking cessation treatment among cancer center patients via a low-burden strategy, Electronic Health Record-Enabled Evidence-Based Smoking Cessation Treatment. *Transl Behav Med*. 2019;10(6):1504–1514.
- 71 Kaphingst KA, Kohlmann W, Chambers RL, et al. Comparing models of delivery for cancer genetics services among patients receiving primary care who meet criteria for genetic evaluation in two healthcare systems: BRIDGE randomized controlled trial. *BMC Health Serv Res*. 2021;21(1):542.