Genomic Insights for Personalized Care: Motivating At-Risk Individuals Toward Evidence-Based Health Practices

Tony Chen[1,*], Giang Pham[2], Louis Fox[2], Nina Adler[3,4], Xiaoyu Wang[5,6], Jingning Zhang[7], Jinyoung Byun[8], Younghun Han[8], Gretchen R.B. Saunders[9], Dajiang Liu[10], Michael J. Bray[11,12], Alex T. Ramsey[2], James McKay[13], Laura Bierut[2], Christopher I. Amos[8,14], Rayjean J. Hung[4], Xihong Lin[1,15], Haoyu Zhang[5,†,*], Li-Shiun Chen[2,†,*]

[1] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA
[2] Department of Psychiatry, Washington University School of Medicine, St. Louis, USA
[3] Department of Anthropology, University of Toronto, Mississauga, ON, Canada
[4] Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, and University of Toronto, Toronto, Canada
[5] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
[6] Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Rockville, MD, USA
[7] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA.
[8] Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA
[9] Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[10] Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA
[11] Department of Genetic Counseling, Bay Path University, Longmeadow, MA, USA
[12] ThinkGenetics, Inc, USA
[13] International Agency for Research on Cancer, World Health Organization, Lyon, France
[14] Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA
[15] Department of Statistics, Harvard University, Cambridge, USA

*Correspondence to: Tony Chen (tonychen@g.harvard.edu), Haoyu Zhang (haoyu.zhang2@nih.gov) and Li-Shiun Chen (li-shiun@wustl.edu).
†These authors jointly supervised this work: Haoyu Zhang and Li-Shiun Chen.

Abstract

Background: Lung cancer and tobacco use pose significant global health challenges, necessitating a comprehensive translational roadmap for improved prevention strategies. Polygenic risk scores (PRSs) are powerful tools for patient risk stratification but have not yet been widely used in primary care for lung cancer, particularly in diverse patient populations.

Methods: We propose the GREAT care paradigm, which employs PRSs to stratify disease risk and personalize interventions. We developed PRSs using large-scale multi-ancestry genome-wide association studies and standardized PRS distributions across all ancestries. We applied our PRSs to 796 individuals from the GISC Trial, 350,154 from UK Biobank (UKBB), and 210,826 from All of Us Research Program (AoU), totaling 561,776 individuals of diverse ancestry.

Results: 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.

Conclusion: Our PRS-based intervention model leverages large-scale genetic data for robust risk assessment across populations. This model will be evaluated in two cluster-randomized clinical trials aimed at motivating health behavior changes in high-risk patients of diverse ancestry. This pioneering approach integrates genomic insights into primary care, promising improved outcomes in cancer prevention and tobacco treatment.

Introduction

The worldwide burden of lung cancer and tobacco smoking presents major challenges to global health[1]. 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. Communication of the precision risk of lung cancer and precision benefit of smoking cessation is a promising but untested strategy to promote health behavior changes to reduce cancer risk. To address this gap, polygenic risk scores (PRSs) emerge as a valuable approach to assess disease susceptibility among populations and pinpoint individuals at higher risk[2–4]. PRSs can be derived from large-scale genome-wide association studies (GWAS) to estimate individual disease risk and have shown promise in predicting health outcomes and promoting preventive healthcare[5–9]. Despite their potential, PRSs' implementation in primary care is limited, especially in diverse populations. Implementing a PRS-based precision intervention is crucial in order to address the multifaceted needs of different communities and individuals[10]. Harnessing PRSs effectively can make significant progress in mitigating lung cancer's public health impact.

Ongoing studies like eMERGE[11], GenoVA[12], and WISDOM[13] are leading the implementation of PRS into genetic risk reports (**Table 1**). They aim to personalize medical reports and understand the impact of PRS on screening, diagnostic procedures, and patient behavior. Notably, a gap persists as these initiatives have not yet formulated a PRS specifically for lung cancer. A likely reason is that the global burden of lung cancer is primarily driven by tobacco smoking rather than genetics[14,15]. However, accounting for the genetic basis of lung cancer may provide patients and clinicians with additional actionable information. The unique value proposition of a lung cancer-specific PRS lies in leveraging established and clear guideline-based prevention strategies, including smoking cessation treatment and lung cancer screening[16]. By incorporating PRSs for lung cancer and difficulty quitting smoking without treatment, there is an opportunity to revitalize and enhance these often-under-utilized prevention practices.

We introduce the Genomic Informed Care for Motivating High Risk Individuals Eligible for Evidence-based Prevention (GREAT) framework as a novel approach to incorporate PRS-enabled interventions in clinical settings (**Figure 1**). The core of GREAT is the use of PRSs that offer precise risk estimates for lung cancer and difficulty quitting smoking. By providing patients with personalized risk information, we aim to activate behavior change mechanisms that promote preventive actions. The primary targets of this intervention are high-risk individuals eligible for evidence-based prevention practices, such as lung cancer screening and smoking cessation. By integrating precision risk information with the benefits of timely interventions, GREAT empowers patients to make informed decisions about their health and motivates them to take proactive steps towards prevention. The personalized information may increase the patient uptake of evidence-based treatment by activating potential mechanisms such as treatment. Guided by enhanced outcome expectancy and self-efficacy based on the Capability-Opportunity-Motivation-Behavior (COM-B) model and Theoretical Domains Framework (TDF)[17–19]. Ultimately, our objective is not only to motivate, but to significantly reduce lung cancer morbidity and mortality through this innovative care paradigm.

Effectively translating PRSs into clinical practice requires a comprehensive and pragmatic translational roadmap for equitable and effective implementation (**Figure 1**). First, to address ancestry diversity, we take a two-step approach of (1) constructing PRS based on large-scale multi-ancestry GWAS, and (2) standardizing PRS distributions across the continuum of genetic ancestry to ensure accurate risk stratification based on PRS distributions for all patients[20]. Second, to document accuracy and transportability of the PRSs to diverse populations, we perform large-scale validation using diverse data from the Genetic Informed Smoking Cessation (GISC) trial[21], UK Biobank (UKBB)[22], and All of Us Research program (AoU)[23]. Third, we translate risk into actionable categories by setting appropriate thresholds for the PRS. The alignment of these thresholds with clinical significance involves many considerations for meaningful risk stratification. Fourth, we propose clear and patient-friendly communication strategies, including visual aids and educational materials, to facilitate understanding and meaningful interactions between patients and healthcare

125 providers. Effectively communicating both the risk and precision of the PRS results is
126 challenging but essential to empower patients to make informed decisions about their
127 health. Moreover, it is crucial to consider patient perceived risk, perceived benefit, and
128 personal relevance when discussing PRS results with patients. Patients' understanding
129 and interpretation of PRS may vary, leading to differing levels of engagement in
130 preventive actions. Hence, comprehensive patient education programs can enhance
131 awareness and knowledge about PRS, its implications, and available preventive
132 measures.
133
134 In this paper, we introduce the design for two cluster randomized clinical trials (RCT):
135 (1) PRECISE, which evaluates the effectiveness of a multilevel intervention, *RiskProfile*,
136 on increasing lung cancer screening and tobacco treatment utilization in primary care
137 (NIDA Grant 5R01CA268030-02); and (2) MOTIVATE, which evaluates the effect of
138 PrecisionTx, a multilevel intervention to promote precision tobacco treatment in primary
139 care (NIDA Grant 5R01DA056050-02). Through the innovative use of PRS, our aim is to
140 motivate lung cancer screening and tobacco treatment among high-risk patients. We
141 present a new care paradigm (**Figure 1**) and outline a translational roadmap (**Figure 2**)
142 that discusses potential barriers and solutions for implementation. By incorporating
143 personalized risk assessments, such PRS-enabled interventions have the potential to
144 significantly improve lung cancer prevention strategies and patient outcomes.
145
146 **Materials and Methods**
147 <u>1000 Genomes Project Phase 3 reference data and principal components analysis</u>
148 We use the 1000 Genomes Project Phase 3 (1000G)[24,25] as a reference for genetic
149 ancestry inference, as it is publicly accessible and includes genotype data from diverse
150 populations. The 1000G dataset includes 3,202 individuals, with 633 Europeans (EUR),
151 893 Africans (AFR), 585 East Asians (EAS), 601 South Asians (SAS), and 490 Admixed
152 Americans (AMR). We conducted principal components analysis (PCA) plink 2.0[26] on all
153 3,202 samples, using 55,248 SNPs that are shared among the recommended SNPs set
154 by gnomAD[27], 1000G reference data, UKBB and GISC validation data, and the

23andMe genotyping array used for the trial (**Figure 3, Supplementary Figure 1, Supplementary Table 4**).

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

Validation in GISC, UK Biobank, and All of Us

We validated our PRS approaches in three datasets: Genetically Informed Smoking Cessation Trial (GISC)[21], UK Biobank (UKBB)[22], and All of Us Research Program (AoU)[23] Here we present are brief background of the three studies. More details of the study background are provided in the **Supplementary Note**, and sample characteristics regarding sex, age, and outcomes information in UKBB and AoU are provided in **Supplementary Table 1.**

GISC Trial: The GISC trial, conducted at conducted at Washington University in St Louis[21], included 822 current or previous smokers. Genetic data were available for 796 individuals (503 European, 257 African, 36 Other self-reported ethnicities), This dataset was primarily used for ancestry inference validation due to its relevance to our PRECISE and MOTIVATE trials.

UKBB: The UKBB dataset included 340,154 unrelated individuals with rich genetic and clinical data. This cohort consisted of 6,844 African, 730 Admixed American, 770 East Asian, 313,279 European, 7,197 South Asian, and 11,334 Other individuals by genetically-inference ancestry. The lung cancer analysis included 1,830 cases (ICD10 codes C34.0-C34.9) and 338,334 controls. The smoking cessation analysis involved 152,406 ever-smokers (117,483 former and 34,923 current), excluding 186,040 never smokers and 1,312 prefer-not-to-answer responses.

AoU: The AoU dataset included 210,826 unrelated individuals with whole-genome sequencing data. This cohort comprised 45,108 African, 32,563 Admixed American, 3,873 East Asian, 110,712 European, 1,689 South Asian, and 16,881 Other individuals by genetically-inference ancestry. We focused on 152,916 current or previous smokers for our smoking analysis.

Construction of polygenic risk scores

Our trial incorporates the latest findings by utilizing recently genome-wide association study (GWAS) summary statistics for lung cancer[28] and difficulty quitting smoking[29], excluding UK Biobank samples to avoid overlapping with our validation data. While these meta-analyses predominantly consist of individuals of 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[30,31].

For lung cancer risk, we started with 128 published SNPs found to be predictive of 5-year and lifetime cumulative risk for lung cancer[16]. Out of these, 101 SNPs overlapped with the published summary statistics, reference, and validation data (1000G, UKBB, GISC and AoU), and the 23andMe genotyping array used for the trial (**Supplementary Figure 1**). 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[28]. For difficulty quitting smoking, we identified 175 SNPs predictive for smoking cessation following the same filtering procedure for lung cancer[29].

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

$$PRS_i = \beta_1 G_{i1} + \beta_2 G_{i2} + \cdots + \beta_M G_{iM}.$$

PRS calculations were performed using R, with genotype data input via the genio package[32]. PRS SNPs and weights for lung cancer and difficulty quitting smoking are provided in **Supplementary Tables 2-3**, respectively.

Standardizing PRS distributions across the continuum of genetic ancestry

We standardized 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 4-5**)[20]. 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:

$$PRS_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \cdots + \alpha_5 PC_{i5} + \epsilon_i^{mean}.$$

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

$$r_i = PRS_i - \widehat{\alpha_0} - \widehat{\alpha_1} PC_{i1} - \cdots - \widehat{\alpha_5} PC_{i5}.$$

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

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

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

$$PRS_i^{adj} = \frac{PRS_i - \widehat{\alpha_0} - \widehat{\alpha_1} PC_{i1} - \cdots - \widehat{\alpha_5} PC_{i5}}{\sqrt{\widehat{\gamma_0} + \widehat{\gamma_1} PC_{i1} + \widehat{\gamma_2} PC_{i2} + \cdots + \widehat{\gamma_5} PC_{i5}}}$$

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[33,34].

Definition of patient risk categories

We stratified patients by genetic risk for lung cancer and difficulty quitting smoking by calculating odds ratios (ORs). ORs compare the probability of an outcome occurring in

245 individuals within a percentile range p (i.e. 80-100%) of the ancestry-adjusted PRS
246 distribution to those within a lower percentile range q (i.e. 0-20%).

247 $$OR_{pq} = \frac{P(Y = 1|PRS \in p)/P(Y = 0|PRS \in p)}{P(Y = 1|PRS \in q)/P(Y = 0|PRS \in q)}$$

248 We established cut points in the PRS distribution to categorize individuals into three risk
249 groups: “at risk”, “high risk”, and “very high risk”. For lung cancer, we defined these
250 categories as the bottom 20%, middle 60%, and top 20% of the PRS distribution. For
251 difficulty quitting smoking, categories were the bottom, middle, and top 33%. These
252 thresholds were chosen to communicate personalized risks and benefits to high-risk
253 patients in our trials, all of whom are eligible for lung cancer screening due to heavy
254 smoking. Since difficulty quitting smoking lacks established absolute risk rates, we used
255 these percentiles to provide a more agnostic risk assessment.
256
257 We use the distribution of our ancestry-adjusted PRSs among all 1000G samples to set
258 percentile ranges and evaluate corresponding odds ratios among UKBB and AoU
259 participants. For comparison, we also evaluate odds ratios using ancestry-matched raw
260 PRS distributions, i.e. European-only 1000G PRS distribution for UKBB and AoU
261 participants with genetically predicted European ancestry. For individual with “Other”
262 predicted ancestry, we use the overall raw PRS distribution among all 1000G samples.
263

264 **Results**

265 <u>Harmonization PRS distributions across ancestry</u>

266 We found notable variation in raw PRS distributions across ancestries, highlighting that
267 applying a universal cutoff for raw PRS without accounting for ancestry can lead to
268 biased risk profiling and inaccurate clinical recommendations (**Figure 4**). However, our
269 regression-based ancestry adjustment across all three datasets yields much more
270 standardized distributions across ancestries. Specifically, the adjusted proportions of
271 individuals within each risk category closely align with 20%-60%-20% for lung cancer,
272 and 33.3%-33.3%-33.3% for difficulty quitting smoking. This adjustment ensures that
273 patients of any background can be compared against a unified reference distribution for
274 each outcome (**Supplementary Tables 6a-f**). This standardization places individuals

275 across all ancestries on the same scale, allowing for a single risk stratification cutoff
276 regardless of ancestral background. Such measures importantly enable fair risk
277 assessment for individuals who may be labelled as “Other”, who may not fit well into a
278 binned ancestry-specific risk model.
279
280 Risk stratification for lung cancer and smoking cessation in UK Biobank and All of Us
281 After assigning UKB and AoU participants as “at risk”, “high risk”, or “very high risk” for
282 lung cancer and difficulty quitting smoking, we identified significant odds ratios for both
283 traits across different ancestry groups (**Figure 5**, **Supplementary Tables 7-10**). Our
284 ancestry-adjusted PRS yielded similar ORs as using ancestry-matched distributions,
285 demonstrating that a single PRS distribution can be appropriately applied to all
286 individuals. In UKBB, the overall ORs for lung cancer 1.42 (95% CI: 1.24 – 1.65) for the
287 “high risk” group and 1.85 (95% CI: 1.58 – 2.18) for “very high risk” group compared to
288 the “at risk” group (**Supplementary Table 7a**). In AoU, these ORs were slightly higher –
289 1.65 (95% CI: 1.36 – 2.02) for “high risk” group and 2.39 (95% CI: 1.93 – 2.97) for “very
290 high risk” group. While the ORs in non-European ancestries within UKBB were not
291 significant, they were significant and even higher than those for European ancestries in
292 AoU (**Supplementary Table 8a**). Moreover, the ORs for “Other” ancestry group in AoU
293 were much higher and more significant using the ancestry-adjusted PRS compared to
294 raw PRS, where an ancestry-matched risk distribution is least suitable (**Supplementary**
295 **Table 8b)**.
296
297 We observed similar patterns for difficulty quitting smoking, where the ancestry-adjusted
298 PRS performed similar or better than using ancestry-matched distributions for raw PRS.
299 In UKBB, the overall ORs for difficulty quitting smoking were 1.19 (95% CI: 1.15 – 1.22)
300 for “high risk” group and 1.36 (95% CI: 1.32 – 1.41) for “very high risk” group
301 (**Supplementary Table 9a)**. In AoU, the ORs were 1.12 (95% CI: 1.08 – 1.15) for “high
302 risk” group and 1.32 (95% CI: 1.28 – 1.36) for “very high risk” (**Supplementary Table**
303 **10a**) group. Notably, we observed higher and more significant ORs in non-European
304 AoU participants when using the ancestry-adjusted PRS compared to the raw PRS.
305 Although the improvement within each specific ancestry group was not as substantial,

306 the aggregate ORs in all non-European groups increased from 0.89 (95% CI: 0.86 –
307 0.93) to 1.05 (95% CI: 1.01 – 1.09) in the “high risk” group, and from 1.07 (95% CI: 1.03
308 – 1.11) to 1.23 (95% CI: 1.18 – 1.28) in the “very high risk” group.
309
310 Using ancestry-adjusted PRS ensures accurate risk stratification across all ethnic
311 backgrounds, a critical consideration given the substantial variability in raw PRS
312 distributions across diverse populations. The outcome-based validation in UKBB and
313 AoU further verifies that the ancestry-adjusted PRS provides valid risk stratification, and
314 often yields better risk stratification for non-European ancestries compared to ancestry-
315 matched modeling. These findings collectively facilitate a more robust and standardized
316 application of PRS in clinical reporting.
317
318 Translating polygenic risk scores into clinical reports
319 We highlight two example trials: PRECISE (NIDA Grant 5R01CA268030-02) and
320 MOTIVATE (NIDA Grant 5R01DA056050-02), designed to promote health behavior
321 change using genetically-informed interventions, RiskProfile and PrecisionTx,
322 respectively. These interventions incorporate PRS to communicate precision risk of lung
323 cancer and precision benefits of smoking cessation, promoting evidence-based
324 practices such as cancer screening and tobacco treatment in high-risk individuals who
325 smoke and/or are eligible for lung cancer screening. PRS risk stratification from either
326 RiskProfile or PrecisionTx, along with clinical information, is delivered within a
327 comprehensive report that includes actionable recommendations to reduce lifetime risk
328 (**Figure 6** and **Figure 7**). Access to 23andMe genotypes and expanded health
329 information has been a motivating component for the research participants. Both
330 PRECISE and MOTIVATE are currently in the preliminary phases of recruiting primary
331 care providers and patients. The recruitment strategy aims to engage over 100
332 physicians and 1600 patients in these trials.
333
334 **Discussion**
335 In this study, we introduce the GREAT framework in primary care, applying PRSs in two
336 trials to provide precise risk estimates for lung cancer and difficulty quitting smoking to

high-risk individuals. These estimates aim to activate behavior change mechanisms that promote health. To enable the interventions, we present a translational roadmap to implement genetic data in PRS-enabled interventions designed to promote health behaviors, evaluating their effectiveness in motivating high-risk patients to increase cancer screening and smoking cessation.

Importantly, we have accomplished our goals of generating behavior-change interventions by a) framing our translational message specifically for high-risk patients who have not received guideline-recommended cancer screening or tobacco treatment[35–38], b) translating risk categories into precision risk and benefit that are designed to motivate health behavior changes, and c) ensuring inclusion of diverse ancestry with PC-regression-based PRS adjustment.

Our goal is to enable a robust implementation of PRS in currently funded clinical trials to evaluate the efficacy of these relatively novel interventions. The GREAT framework guides the implementation of PRS-enabled interventions in primary care. Future research need to address critical questions such as timing, methodology, and location of these interventions' delivery to patients and providers to optimize its acceptability, understanding, and impact.

Our approach to ancestry adjustment of PRS employs widely accessible data from the 1000G dataset, as an alternative to methods in the GenoVA[12] and eMERGE[11] 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. 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.

367 A notable gap in current practice is the absence of genetic information in electronic
368 health records (EHRs) for decision support and the lack of PRS generation in clinical
369 labs. Implementing precision interventions in primary care necessitates a workflow that
370 incorporates EHRs for recruitment, biomarker testing protocols, and standardized
371 processes to generate personalized intervention reports[39]. This requires collaborations
372 with primary care stakeholders, community advisory boards, genetic counseling, and
373 health communication to improve intervention clarity, accuracy, and impact[40–44]. Patients
374 expressed a notable interest in receiving personalized interventions. In our previous
375 study, 85% of smoking patients reported a high interest in receiving genetically tailored
376 tobacco treatment[45]. Further, a substantial majority (95%) of individuals who smoke
377 endorsed the importance of receiving genetic results to guide their treatment[41].
378 Following receipt of a personalized genetic risk profile for smoking cessation, 91% of
379 participants who smoke found the tool highly useful for understanding their health, cope
380 with health risks, and feeling more in control of their health[42]. Such pronounced interest
381 and the perceived significance of genetic data highlight the growing demand for
382 personalized interventions among patients who smoke. Personalized interventions may
383 increase patient compliance**.** For example, study found that patients reported higher
384 interest in taking medication (97.5% vs. 61.0%, $p<.0001$) when medication was
385 personalized based on their genetics[45]. However, substantial evidence suggests the
386 complex interplay of fear, fatalism, and risk perception when patients face lung cancer
387 screening[46–49] Therefore, it is important to evaluate the translation potential of
388 personalized genetics in motivating patients for positive behavior change.
389
390 Unlike most current research that evaluates PRS-enabled interventions in general
391 patient populations, our work uniquely focuses on designing and evaluating these
392 interventions specifically among high-risk patients who will benefit tremendously from
393 recommended health behaviors such as lung cancer screening and smoking cessation,
394 when general medical advice is not enough to motivate such behaviors.
395
396 Here we share three key design considerations for best practices. First, <u>for equitable</u>
397 <u>implementation</u> of precision health interventions, tools must be designed with active

engagement from racial/ethnic minority communities, rather than solely examining their effectiveness for these communities post hoc. We engage in ongoing participatory sessions with racial and ethnic minority communities and advisory boards across all phases through iterative cycles of intervention development, feedback, and testing so that innovative genomics-informed tools are beneficial across diverse populations.

Second, we have chosen clinically meaningful thresholding for PRS risk categories in communicating personalized risks and benefits to patients. These categories were selected because all participants in our ongoing trials are high-risk primary care patients eligible for lung cancer screening, who are current or former heavy smokers. We defined lung cancer risk by the bottom 20%, middle 60% and top 20%, and difficulty quitting smoking by slightly more agnostic tertiles, to motivate positive behavior change. We aim to follow best practices of communicating uncertainties and deciding between a continuous or categorical risk presentation. Transparency about the imprecision in both risk estimates and action thresholds for PRS is essential.

Third, we plan to update our workflow to adapt to new GWAS findings and evolving methodologies. As scientific knowledge advances, outdated PRS predictions can hinder effective implementation. A dynamic PRS framework that allows for regular updates based on new discoveries is essential. This ensures that PRS-based interventions remain current and aligned with the latest advancements, reducing the implementation gap and maximizing their impact on preventive healthcare outcomes. Following current genetic counseling recommendations, we have established a process to incorporate new evidence into our intervention for smoking cessation and lung cancer risk. This process will adjudicating 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.

Our work have several limitations, but we hope 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 our unique outcomes (lung cancer and

429 difficulty quitting smoking), population (patients eligible for lung cancer screening and
430 tobacco treatment), and context (primary care) to optimize health impact. A notable
431 limitation is the underrepresentation of non-European populations in the multi-ancestry
432 GWAS employed to derive the PRS weights. This underrepresentation may reduce the
433 predictive power of the PRS in non-European populations. The dominance of European
434 populations persists in existing GWAS[50–53], not limited to lung cancer and smoking
435 cessation. However, with the burgeoning emphasis on incorporating minority
436 populations in GWAS analyses and the development of new PRS approaches and
437 toolkits to enhance predictive power in diverse populations [7–9,34,54–56], we can iteratively
438 refine PRS implementation in our trial to synchronize with the latest advancements.
439
440 Many questions remain for the near future. First, <u>how can we reduce the time lag from</u>
441 <u>evidence to implementation?</u> The constant evolution of evidence identifying new
442 biomarkers for treatment is a challenge. For instance, our recent work highlights the
443 potential of PRSs in guiding future treatment approaches[57]. Despite actionable precision
444 treatment findings, the ever-changing evidence base and the perception of better data
445 have hindered effective implementation[43,58]. To address this, we aim to use cutting-edge
446 metabolic and genetic markers that offer robust evidence for precision treatment,
447 reducing the time from evidence generation to practical application. This approach
448 allows us to measure and report the time from landmark publications to the
449 implementation of key findings [59], enhancing the efficiency and effectiveness of
450 precision treatment and ensuring patients benefit from the latest recommendations.
451
452 <u>Second, can we truly evaluate the effect of precision interventions?</u> We expect that
453 precision interventions may activate multiple mechanistic pathways to the uptake and
454 efficacy of lung cancer screening or tobacco treatment. Understanding these
455 mechanisms is essential to refine for intended outcomes and contexts. <u>Third, can these</u>
456 <u>precision interventions be scaled in the real-world clinics?</u> Evidence shows that
457 physicians are highly receptive to medication recommendations based on biomarkers[60].
458 To reduce burden, we need to leverage existing EHR tools (e.g. Best Practice
459 Advisories) and training to efficiently facilitate physician prescribing[61,62]. Understanding

460 of mechanistic and implementation outcomes will guide scalable, efficient delivery
461 components for integration into clinic workflows[39], using trained embedded staff, and
462 digital therapeutic tools to enable these PRS-informed behavioral interventions[63].
463
464 In conclusion, a well-designed roadmap that validates the PRS, incorporates multi-
465 ancestry GWAS-based weights, translates risk into actionable categories,
466 communicates effectively, considers patient perspectives, and accommodates evolving
467 science is essential for the equitable and pragmatic translation of PRS into clinical care.
468 By addressing the barriers and implementing potential solutions at each stage, we can
469 leverage PRS to improve preventive healthcare and significantly reduce the burden of
470 lung cancer.

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.), 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 U19-CA203654.

Author Contributions

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. G.S. and D.J. assisted with obtaining and processing GWAS summary statistics for difficulty quitting smoking. T.C., G.P., H.Z., and L.C. drafted the manuscript. All co-authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

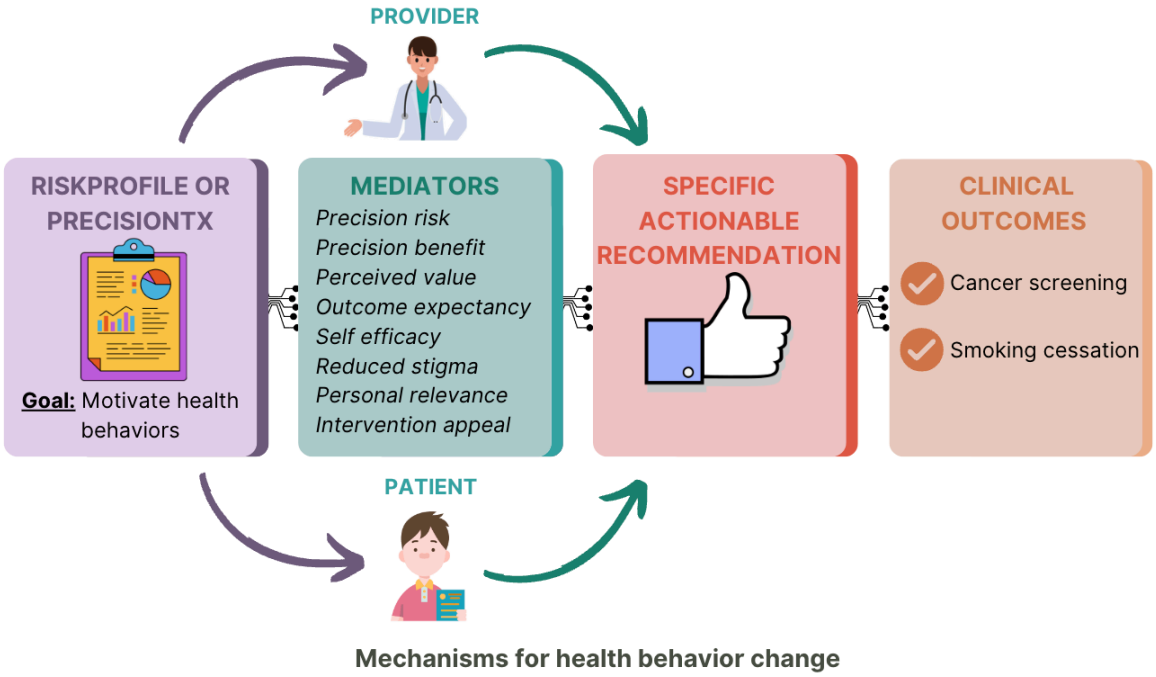
Laura J. Bierut is listed as an inventor on Issued U.S. Patent 8,080,371, “Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Michael J. Bray is an employee at ThinkGenetic, Inc. Where authors are identified as personnel of the 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.

Data and Code Availability

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.

Genotype data from the 1000 Genomes Project (Phase 3) were acquired via the plink website https://www.cog-genomics.org/plink/2.0/resources. Population information were obtained from public resources: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ and https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/1000G_698_related_high_coverage.sequence.index.

511 **Figures**



512

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

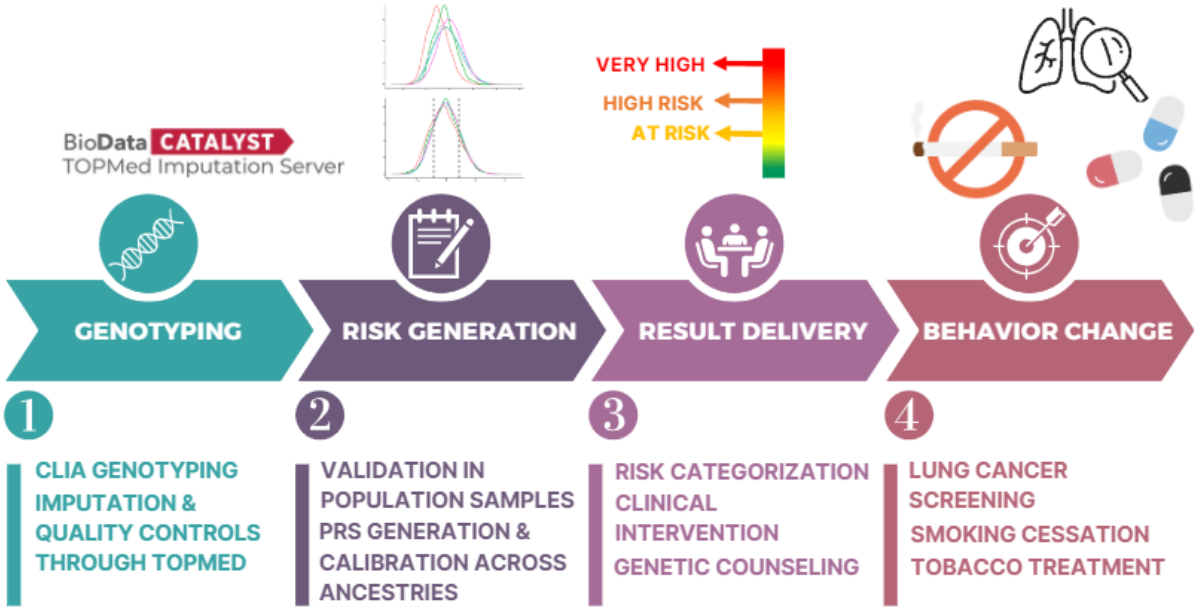


Figure 2 Roadmap for translating genetic data to a genetic risk profile as a multilevel intervention in primary care. In step 1, enrolled participants' genetic data are analyzed by 23andMe's Clinical Laboratory Improvement Amendments (CLIA) certified genotyping process. Imputation and quality controls 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. Behavioral interventionists (research staff who are trained, certified, and supervised by a team of genetic counselor, psychologist, and psychiatrist) offer personalized guidance on behavior 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.

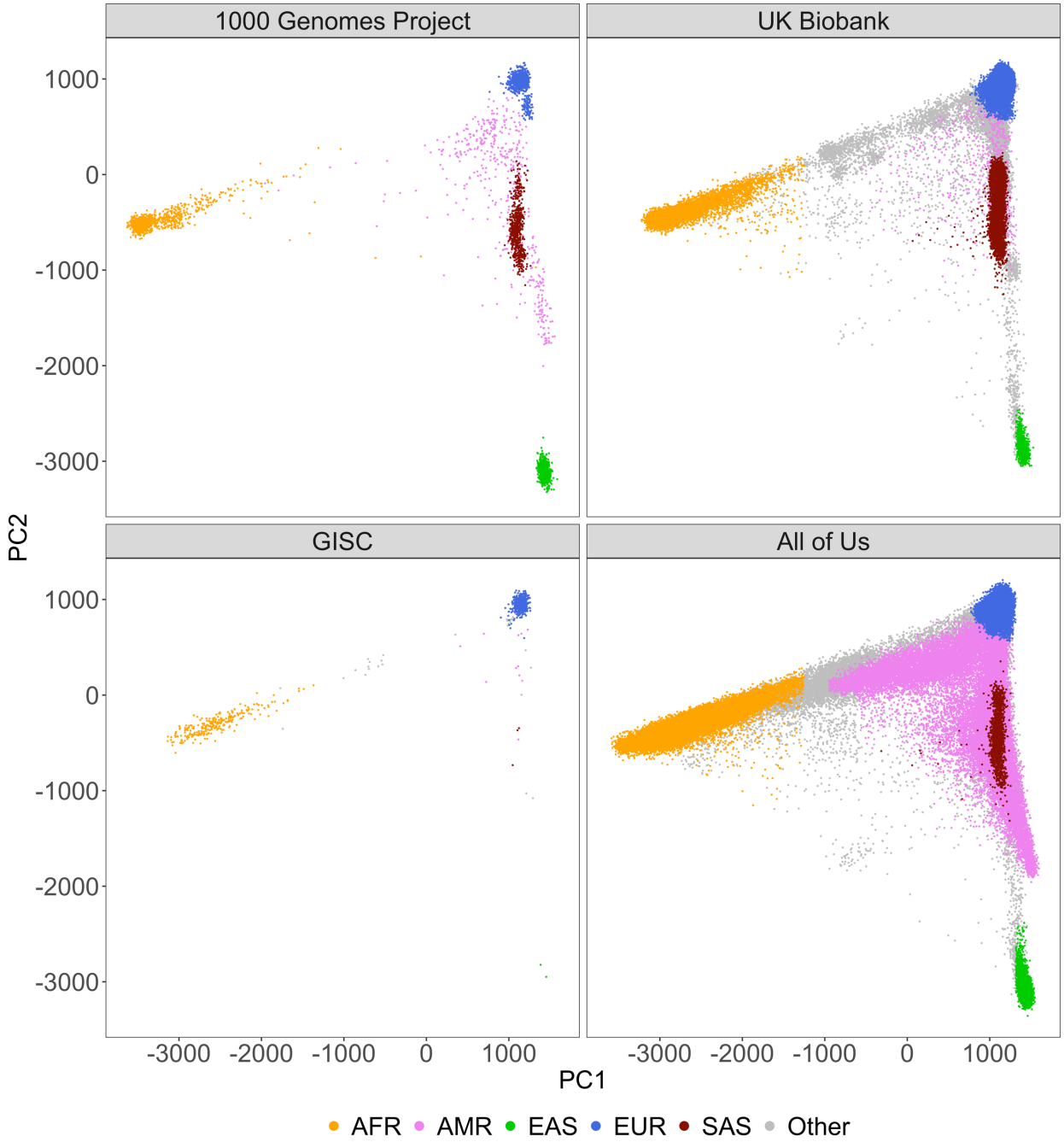


Figure 3. Cross-dataset distribution of genetic ancestry via PCA Projections in 1000G, GISC, UKBB, and AoU. This figure illustrates the utility of principal components analysis (PCA) loadings obtained from the 1000 Genomes Project Phase 3 (1000G) in discriminating ancestries within external datasets, 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

553 individual in these datasets, with points distinctly marked by genetically-inferred
554 ancestry.

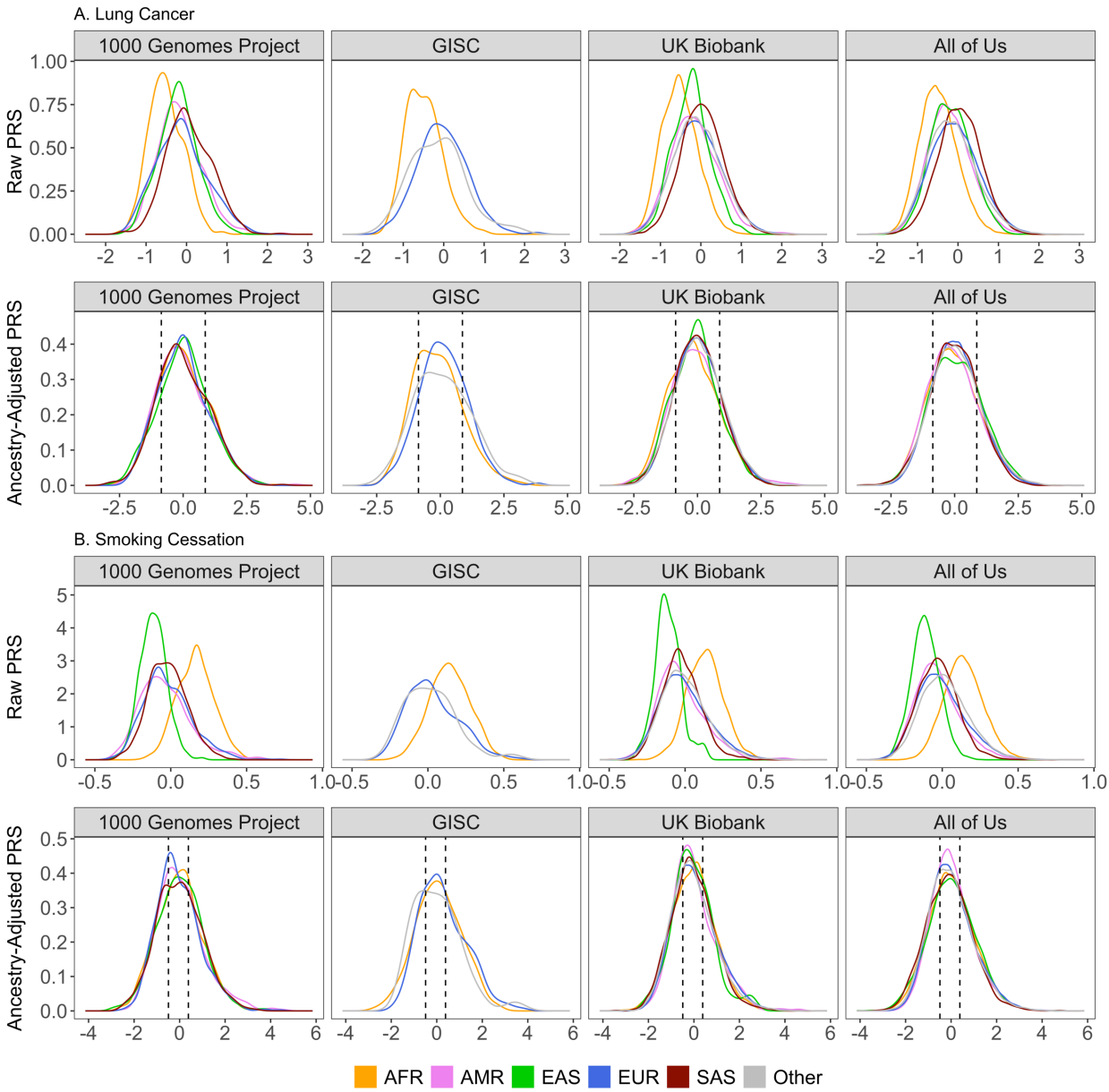


Figure 4 Ancestry adjustment of PRS for lung cancer and quit difficulty PRS across ancestral populations. We showcase the adjustment process for polygenic risk scores (PRS) for lung cancer (Panel A) and difficulty quitting smoking (Panel B) within the 1000 Genomes Project (1000G), GISC Trial, UK Biobank, 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 centers 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 3,202 samples in the 1000 Genomes Project.

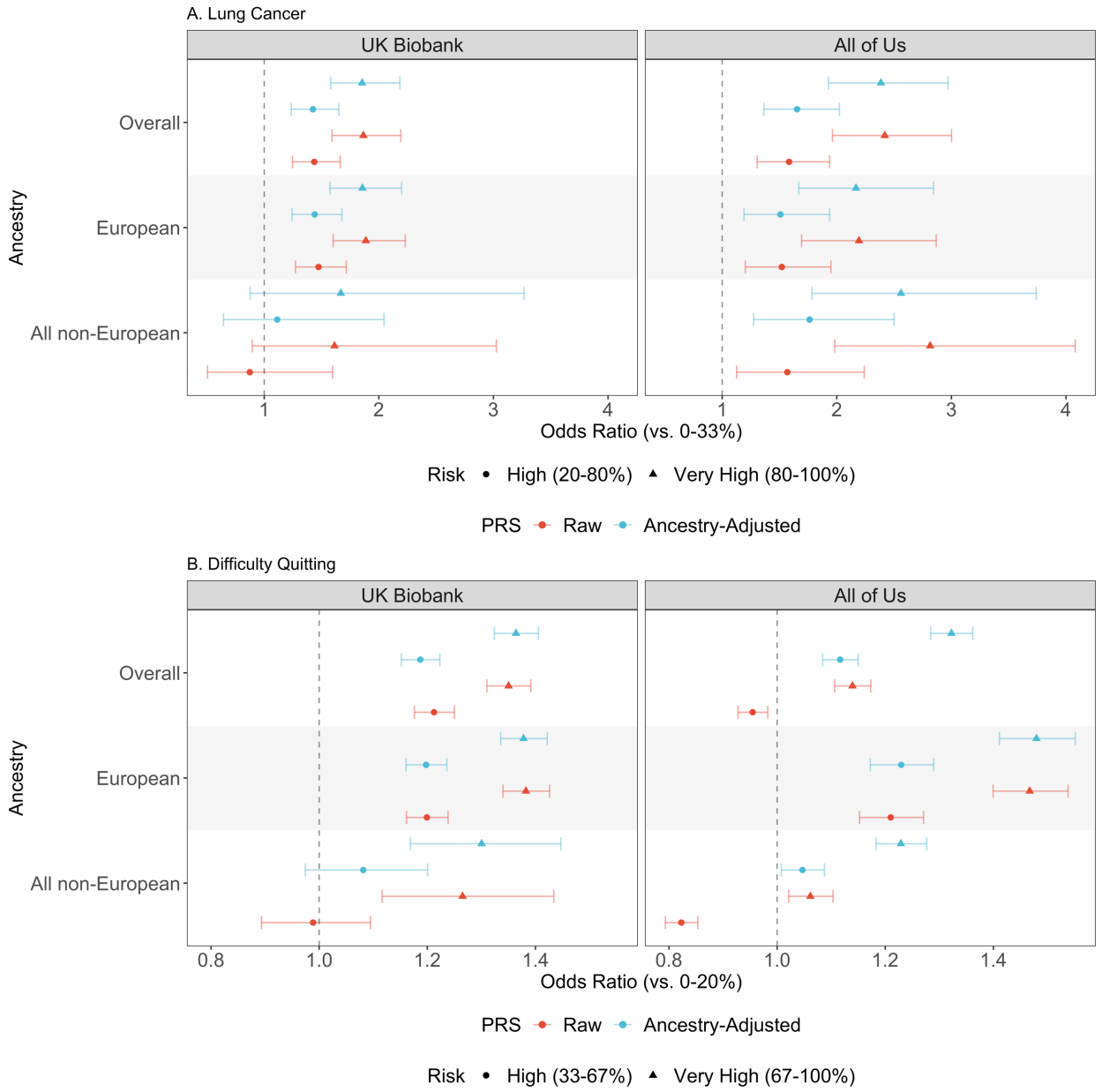


Figure 5. Risk stratification for lung cancer and difficulty quitting smoking using raw and ancestry-adjusted PRS. This figure illustrates the odds ratios (ORs) with associated 95% confidence intervals of PRSs for lung cancer (A) and difficulty quitting smoking (B) among UK Biobank (N=340,154 for lung cancer and N=152,406 for smoking), and All of Us (N=210,826 for lung cancer and N=152,916 for smoking) participants. We compare risk stratification using a raw PRS with ancestry-matched percentiles, and our ancestry-adjusted PRS with the same percentiles for all individuals.

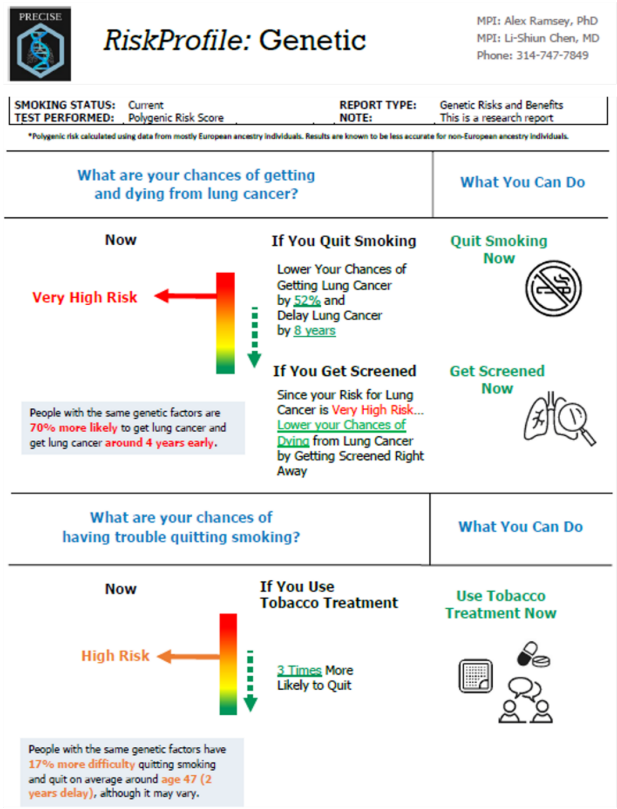


Figure 6. Example 1 clinical report: RiskProfile. We present example 1 for genomically-informed interventions using the GREAT framework. RiskProfile is designed to motivate lung cancer screening and tobacco treatment among screening-eligible patients. This intervention utilizes ancestry-adjusted PRS 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 personalized risk to also provide personalized benefit of cancer screening and use a multilevel intervention design directed to both physicians and patients in clinical settings. In an ongoing NCI-funded trial (NCT05627674), the effect of RiskProfile on clinician ordering and patient completion of lung cancer screening will be evaluated.

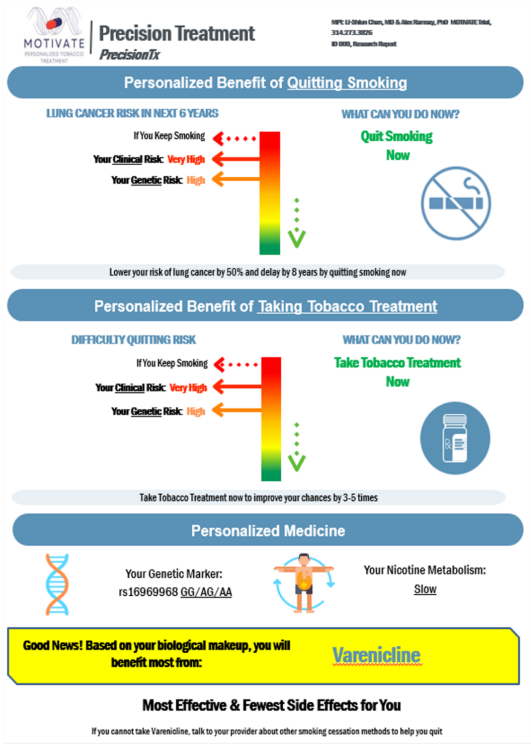


Figure 7. Example 2 clinical report: PrecisionTx. We present example 2 for genomically-informed interventions using the GREAT framework. PrecisionTx is designed to motivate tobacco treatment among patients who smoke. This intervention utilizes ancestry-adjusted PRS to stratify patients into “at risk” (yellow), “high risk” (orange), and “very high risk” (red) genetic risk categories. PrecisionTx focuses on treatment and expands beyond personalized risk to also provide personalized benefit of tobacco treatment, and use a multilevel intervention design directed to both physicians and patients in clinical settings. In an ongoing NIDA-funded trial (NCT05846841), the effect of PrecisionTx on clinician ordering, patient adherence, and smoking abstinence will be evaluated.

Tables

	GenoVA	eMERGE	Wisdom	PRECISE / MOTIVATE
Goal	Motivate early diagnosis	Increasing clinical actions to mitigate risk of future disease	Personalized screening frequency	Motivate lung cancer screening (LCS) / tobacco treatment
Targets				
Target population	All general Veterans' Affairs PCP* and patients	Patients in the large healthcare system	All women can sign up online	**High risk** PC patients (LCS eligible / smokers) and their PCPs
Condition(s)**	BrCa, PrCa, CRCa, Afib, CAD, T2D	BrCa, PrCa, CRCa, Afib, CAD, T1D, T2D, BMI, Asthma, CKD, HCL	BrCA	**LC, smoking cessation**
Target outcome	New diagnoses	Change screening practice or lifestyle	Screening compliance and cancer detection	**LCS, tobacco tx, and smoking cessation**
Polygenic Risk Scores				
Genotyping	Study team		Color genomics	23andMe
Imputation	1000 Genomes	All of Us		TOPMed
Create PRS	Top or many variants European weights Ancestry calibration	Terra cloud platform Ancestry calibration	Color genomics	Top independent variants Multi-ancestry weights Ancestry calibration
Risk Stratification	2 levels (OR of 2)	OR of 5-10		3 levels (top 20%, middle, bottom 20%) or tertiles
Intervention Design				
Risk representation	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
Messaging	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
Behavior mechanisms targeted				**Perceived risk/benefit, Self efficacy, Personal relevance**

* 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, lung cancer=lung cancer

Table 1. **Research on PRS use in clinical trials**. 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 highlight 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 high risk patients 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 behavior mechanisms around lung cancer and smoking.

References

1. Kocarnik, J. M. *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* **8**, 420 (2022).
2. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* vol. 19 581–590 Preprint at https://doi.org/10.1038/s41576-018-0018-x (2018).
3. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine* vol. 12 Preprint at https://doi.org/10.1186/s13073-020-00742-5 (2020).
4. Adeyemo, A. *et al.* Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nature Medicine* vol. 27 1876–1884 Preprint at https://doi.org/10.1038/s41591-021-01549-6 (2021).
5. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576–592 (2015).
6. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* **17**, 1520–1528 (2007).
7. Zhang, H. *et al.* A new method for multiancestry polygenic prediction improves performance across diverse populations. *Nat Genet* (2023) doi:10.1038/s41588-023-01501-z.
8. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573–580 (2022).
9. Chen, T., Zhang, H., Mazumder, R. & Lin, X. Ensembled best subset selection using summary statistics for polygenic risk prediction. *bioRxiv* (2023) doi:10.1101/2023.09.25.559307.
10. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584–591 (2019).
11. Linder, J. E. *et al.* Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genetics in Medicine* **25**, 100006 (2023).
12. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow. *Nat Med* **28**, 1006–1013 (2022).
13. Shieh, Y. *et al.* Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst* **109**, djw290 (2017).
14. Zhang, P. *et al.* Association of smoking and polygenic risk with the incidence of lung cancer: a prospective cohort study. *Br J Cancer* **126**, 1637–1646 (2022).
15. Kanwal, M., Ding, X.-J. & Cao, Y. Familial risk for lung cancer. *Oncol Lett* **13**, 535–542 (2017).
16. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory Based on a Polygenic Risk Model. *Cancer Res* **81**, 1607–1615 (2021).
17. Michie, S. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* **14**, 26–33 (2005).
18. Cane, J., O'Connor, D. & Michie, S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science* **7**, 37 (2012).

19. Atkins, L. *et al.* A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation Science* **12**, 77 (2017).
20. Ge, T. *et al.* Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med* **14**, 70 (2022).
21. Chen, L. *et al.* Genetic Variant in *CHRNA5* and Response to Varenicline and Combination Nicotine Replacement in a Randomized Placebo-Controlled Trial. *Clin Pharmacol Ther* **108**, 1315–1325 (2020).
22. Sudlow, C. *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* **12**, e1001779 (2015).
23. Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* (2024) doi:10.1038/s41586-023-06957-x.
24. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
25. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440.e19 (2022).
26. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
27. Chen, S. *et al.* A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv* 2022.03.20.485034 (2022) doi:10.1101/2022.03.20.485034.
28. Byun, J. *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* **54**, 1167–1177 (2022).
29. Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720–724 (2022).
30. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).
31. Manrai, A. K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities. *New England Journal of Medicine* **375**, 655–665 (2016).
32. Ochoa, A. genio: Genetics Input/Output Functions. Preprint at https://CRAN.R-project.org/package=genio (2023).
33. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society. *Science (1979)* **376**, 250–252 (2022).
34. Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet* (2023) doi:10.1038/s41576-023-00637-2.
35. US Preventive Services Task Force. Lung Cancer: Screening. *https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening* (2021).
36. Agency for Healthcare Research and Quality. Clinical Guidelines and Recommendations. *https://www.ahrq.gov/prevention/guidelines/index.html* (2021).
37. Tobacco Use and Dependence Guideline Panel. *US Department of Health and Human Services. (2008). Tobacco Use and Dependence Guideline Panel. Treating Tobacco Use and Dependence: 2008 Update*. (US Department of Health and Human Services, Rockville, MD, 2008).
38. Krist, A. H. *et al.* Screening for Lung Cancer. *JAMA* **325**, 962 (2021).

39. Ayatollahi, H., Hosseini, S. F. & Hemmat, M. Integrating Genetic Data into Electronic Health Records: Medical Geneticists' Perspectives. *Healthc Inform Res* **25**, 289 (2019).
40. Bourdon, J. L. *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* **14**, 139 (2021).
41. Ramsey, A. T. *et al.* Participatory Design of a Personalized Genetic Risk Tool to Promote Behavioral Health. *Cancer Prevention Research* **13**, 583–592 (2020).
42. Ramsey, A. T. *et al.* Proof of Concept of a Personalized Genetic Risk Tool to Promote Smoking Cessation: High Acceptability and Reduced Cigarette Smoking. *Cancer Prevention Research* **14**, 253–262 (2021).
43. Ramsey, A. T. *et al.* Toward the implementation of genomic applications for smoking cessation and smoking-related diseases. *Transl Behav Med* **8**, 7–17 (2018).
44. Chen, L.-S., Baker, T. B., Ramsey, A., Amos, C. I. & Bierut, L. J. Genomic medicine to reduce tobacco and related disorders: Translation to precision prevention and treatment. *Addiction Neuroscience* **7**, 100083 (2023).
45. Chiu, A. *et al.* Most Current Smokers Desire Genetic Susceptibility Testing and Genetically-Efficacious Medication. *Journal of Neuroimmune Pharmacology* **13**, 430–437 (2018).
46. Quaife, S. L., Janes, S. M. & Brain, K. E. The person behind the nodule: a narrative review of the psychological impact of lung cancer screening. *Transl Lung Cancer Res* **10**, 2427–2440 (2021).
47. Quaife, S. L. *et al.* Psychological Targets for Lung Cancer Screening Uptake: A Prospective Longitudinal Cohort Study. *Journal of Thoracic Oncology* **16**, 2016–2028 (2021).
48. Quaife, S. L., Marlow, L. A. V., McEwen, A., Janes, S. M. & Wardle, J. Attitudes towards lung cancer screening in socioeconomically deprived and heavy smoking communities: informing screening communication. *Health Expectations* **20**, 563–573 (2017).
49. Quaife, S. L., McEwen, A., Janes, S. M. & Wardle, J. Smoking is associated with pessimistic and avoidant beliefs about cancer: results from the International Cancer Benchmarking Partnership. *Br J Cancer* **112**, 1799–1804 (2015).
50. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
51. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
52. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics* **25**, 489–494 (2009).
53. Fitipaldi, H. & Franks, P. W. Ethnic, gender and other sociodemographic biases in genome-wide association studies for the most burdensome non-communicable diseases: 2005–2022. *Hum Mol Genet* **32**, 520–532 (2023).
54. Jin, J. *et al.* MUSSEL: Enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups. *Cell Genomics* **4**, 100539 (2024).
55. Zhang, J. *et al.* An ensemble penalized regression method for multi-ancestry polygenic risk prediction. *Nat Commun* **15**, 3238 (2024).
56. Hou, K. *et al.* Admix-kit: an integrated toolkit and pipeline for genetic analyses of admixed populations. *Bioinformatics* **40**, (2024).

57. Bray, M. *et al.* The Promise of Polygenic Risk Prediction in Smoking Cessation: Evidence From Two Treatment Trials. *Nicotine & Tobacco Research* **24**, 1573–1580 (2022).
58. Ramsey, A. T. *et al.* Designing for Accelerated Translation (DART) of emerging innovations in health. *J Clin Transl Sci* **3**, 53–58 (2019).
59. Proctor, E. *et al.* FAST: A Framework to Assess Speed of Translation of Health Innovations to Practice and Policy. *Global Implementation Research and Applications* **2**, 107–119 (2022).
60. Scoville, E. A. *et al.* Precision nicotine metabolism-informed care for smoking cessation in Crohn's disease: A pilot study. *PLoS One* **15**, e0230656 (2020).
61. Chen, L.-S. *et al.* Low-Burden Strategies to Promote Smoking Cessation Treatment Among Patients With Serious Mental Illness. *Psychiatric Services* **69**, 849–851 (2018).
62. Ramsey, A. 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) doi:10.1093/tbm/ibz107.
63. Kaphingst, K. A. *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* **21**, 542 (2021).