

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

3

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

8

9 <sup>1</sup> Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA

10 <sup>2</sup> Department of Psychiatry, Washington University School of Medicine, St. Louis, USA

11 <sup>3</sup> Regeneron Genetics Center, Tarrytown, NY, USA

12 <sup>4</sup> Department of Medicine, Section of Epidemiology and Population Science, Institute for  
13 Clinical and Translational Research, Houston, TX, USA

14 <sup>5</sup> Department of Psychology, University of Minnesota, Minneapolis, MN, USA

15 <sup>6</sup> Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA,  
16 USA

17 <sup>7</sup> Department of Genetic Counseling, Bay Path University, Longmeadow, MA, USA

18 <sup>8</sup> ThinkGenetics, Inc, USA

19 <sup>9</sup> International Agency for Research on Cancer, World Health Organization, Lyon, France

20 <sup>10</sup> Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston,  
21 TX, USA

22 <sup>11</sup> Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research  
23 Institute, Sinai Health, and University of Toronto, Toronto, Canada

24 <sup>12</sup> Department of Statistics, Harvard University, Cambridge, USA

25 <sup>13</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda,  
26 MD, USA

27

28 \*Correspondence to: Tony Chen ([tonychen@g.harvard.edu](mailto:tonychen@g.harvard.edu)), Haoyu Zhang  
29 ([haoyu.zhang2@nih.gov](mailto:haoyu.zhang2@nih.gov)) and Li-Shiun Chen ([li-shiun@wustl.edu](mailto:li-shiun@wustl.edu)).

30 †These authors jointly supervised this work: Haoyu Zhang and Li-Shiun Chen.

31 **Abstract**

32 Lung cancer and tobacco use pose significant global health challenges and require a  
33 comprehensive translational roadmap for improved prevention strategies. We propose  
34 the GREAT care paradigm (Genomic Informed Care for Motivating High Risk Individuals  
35 Eligible for Evidence-based Prevention), which employs polygenic risk scores (PRSs) to  
36 stratify disease risk and personalize interventions, such as lung cancer screening and  
37 tobacco treatment. We developed PRSs using large-scale multi-ancestry genome-wide  
38 association studies and adjusted for genetic ancestry for standardized risk stratification  
39 across diverse populations. We applied our PRSs to over 340,000 individuals of diverse  
40 ethnic background and found significant odds ratios for lung cancer and difficulty quitting  
41 smoking. These findings enable the evaluation of PRS-based interventions in ongoing  
42 trials aimed at motivating health behavior changes in high-risk patients. This pioneering  
43 approach enhances primary care with genomic insights, promising improved outcomes in  
44 cancer prevention and tobacco treatment, and is currently under assessment in clinical  
45 trials.

46

47 **Introduction**

48 The worldwide burden of lung cancer and tobacco smoking presents major  
49 challenges to global health<sup>1</sup>. Evidence-based practices to reduce their risk such as cancer  
50 screening and tobacco treatment (e.g. smoking cessation medication) have long existed  
51 but are infrequently used in most primary care practices. Communication of the precision  
52 risk of lung cancer and precision benefit of smoking cessation is a promising but untested  
53 strategy to promote health behavior changes to reduce cancer risk. To address this gap,  
54 polygenic risk scores (PRSs) emerge as a valuable approach to assess disease  
55 susceptibility among populations and pinpoint individuals at higher risk<sup>2-4</sup>. PRSs can be  
56 derived from large-scale genome-wide association studies (GWAS) to estimate individual  
57 disease risk and have shown promise in predicting health outcomes and promoting  
58 preventive healthcare<sup>5-9</sup>. Despite their potential, PRSs' implementation in primary care is  
59 limited, especially in diverse populations. Implementing a PRS-based precision  
60 intervention is crucial in order to address the multifaceted needs of different communities  
61 and individuals. Harnessing PRSs effectively can make significant progress in mitigating  
62 lung cancer's public health impact.

63 Ongoing studies like eMERGE<sup>10</sup>, GenoVA<sup>11</sup>, and WISDOM<sup>12</sup> are leading the  
64 implementation of PRS into genetic risk reports (**Table 1**). They aim to personalize  
65 medical reports and understand the impact of PRS on screening, diagnostic procedures,  
66 and patient behavior. Notably, a gap persists as these initiatives have not yet formulated  
67 a PRS specifically for lung cancer. A likely reason is that the global burden of lung cancer  
68 is primarily driven by tobacco smoking rather than genetics<sup>13,14</sup>. However, accounting for  
69 the genetic basis of lung cancer may provide patients and clinicians with additional  
70 actionable information. The unique value proposition of a lung cancer-specific PRS lies  
71 in leveraging established and clear guideline-based prevention strategies, including  
72 smoking cessation treatment and lung cancer screening<sup>15</sup>. By incorporating PRSs for lung  
73 cancer and difficulty quitting smoking without treatment, there is an opportunity to  
74 revitalize and enhance these often under-utilized prevention practices.

75 We introduce the Genomic Informed Care for Motivating High Risk Individuals  
76 Eligible for Evidence-based Prevention (GREAT) framework as a novel approach to  
77 incorporate PRS-enabled interventions in clinical settings (**Figure 1**). The core of GREAT  
78 is the use of PRSs that offer precise risk estimates for lung cancer and difficulty quitting.  
79 By providing patients with personalized risk information, we aim to activate behavior  
80 change mechanisms that promote preventive actions. The primary targets of this  
81 intervention are high-risk individuals eligible for evidence-based prevention practices,  
82 such as lung cancer screening and smoking cessation. By integrating precision risk  
83 information with the benefits of timely interventions, GREAT empowers patients to make  
84 informed decisions about their health and motivates them to take proactive steps towards  
85 prevention. Ultimately, our objective is not only to motivate, but to significantly reduce lung  
86 cancer morbidity and mortality through this innovative care paradigm.

87 Effectively translating PRSs into clinical practice requires a comprehensive and  
88 pragmatic translational roadmap for equitable and effective implementation (**Figure 1**).  
89 First, to address ancestry diversity, we take a two-step approach of (1) constructing PRS  
90 based on large-scale multi-ancestry GWAS, and (2) standardizing PRS distributions  
91 across the continuum of genetic ancestry<sup>16,17</sup> by leveraging reference data from the 1000  
92 Genomes Project Phase 3 (1000G)<sup>18</sup>. Given the variations in allele frequencies (AF) and  
93 linkage disequilibrium (LD) across ancestries, this step is critical to ensure fair risk  
94 stratification based on PRS distributions<sup>10,19</sup>. Second, to document accuracy and  
95 transportability of the PRSs to diverse populations, we perform large-scale validation  
96 using data from individuals of diverse self-reported ethnicities in the UK Biobank (UKBB)  
97 and Genetic Informed Smoking Cessation (GISC) trial. Third, we translate risk into  
98 actionable categories by setting appropriate thresholds for the PRS. The alignment of  
99 these thresholds with clinical significance involves many considerations for meaningful  
100 risk stratification. Fourth, we propose clear and patient-friendly communication strategies,  
101 including visual aids and educational materials, to facilitate understanding and meaningful  
102 interactions between patients and healthcare providers. Effectively communicating both  
103 the risk and precision of the PRS results is challenging but essential to empower patients  
104 to make informed decisions about their health. Moreover, it is crucial to consider patient  
105 perceived risk, perceived benefit, and personal relevance when discussing PRS results  
106 with patients. Patients' understanding and interpretation of PRS may vary, leading to  
107 differing levels of engagement in preventive actions. Hence, comprehensive patient  
108 education programs can enhance awareness and knowledge about PRS, its implications,  
109 and available preventive measures.

110 In this paper, we introduce the design for two cluster randomized clinical trials  
111 (RCT): (1) PRECISE, which evaluates the effectiveness of a multilevel intervention,  
112 *RiskProfile*, on increasing lung cancer screening and tobacco treatment utilization in  
113 primary care (NIDA Grant 5R01CA268030-02); and (2) MOTIVATE, which evaluates the  
114 effect of PrecisionTx, a multilevel intervention to promote precision tobacco treatment in  
115 primary care (NIDA Grant 5R01DA056050-02). Through the innovative use of PRS, our  
116 aim is to motivate lung cancer screening and tobacco treatment among high-risk patients.  
117 We present a new care paradigm (**Figure 1**) and outline a translational roadmap (**Figure**  
118 **2**) that discusses potential barriers and solutions for implementation. By incorporating  
119 personalized risk assessments, such PRS-enabled interventions have the potential to  
120 significantly improve lung cancer prevention strategies and patient outcomes.

121

## 122 **Results**

### 123 *Sample Characteristics*

124 For primary validation, we used data from 340,154 unrelated individuals in the UK  
125 Biobank (UKBB)<sup>20</sup>, given its large sample size, rich clinical data, and inclusion of  
126 individuals from diverse ancestry backgrounds (**Methods**). Lung cancer validation

127 involved 1,830 cases and 338,324 controls across five self-reported ethnic backgrounds:  
128 European (EUR, N = 318,043, N<sub>case</sub> = 1,762), African (AFR, N = 6,409, N<sub>case</sub> = 19), East  
129 Asian (EAS, N = 599, N<sub>case</sub> = 2), and South Asian (SAS, N = 7,520, N<sub>case</sub> = 10), with 7,583  
130 (N<sub>case</sub> = 37) with “Other” self-reported ethnic backgrounds such as mixed or unknown  
131 (**Supplementary Table 1a**). Lung cancer occurrence was slightly higher among men  
132 (53.3% of cases) compared to women (46.7% of cases).

133 For difficulty quitting, the cohort comprised 34,923 current smokers and 117,483  
134 individuals who had previously smoked. The breakdown by self-reported ethnicity for this  
135 analysis was as follows: EUR (N = 145,483, 95.4%), AFR (N = 1,874, 1.2%), EAS (N =  
136 131, 0.1%), and SAS (N = 1,733, 2.2%) (**Supplementary Table 1b**). Among those who  
137 had quit smoking, 50.3% were male among Europeans, while quitting in non-European  
138 males varied between 51.2% and 81.5%.

139 In addition, we validated the PRSs using data from the Genetically Informed  
140 Smoking Cessation (GISC) trial, which more accurately reflects the patient demographics  
141 anticipated in the PRECISE and MOTIVATE trials. The difficulty quitting analysis  
142 encompassed 647 current smokers and 149 former smokers, with the ancestry  
143 distribution as follows: 503 of European descent, 257 of African descent, and 36 of other  
144 ancestries (**Supplementary Table 1c**).

145

#### 146 *Harmonization PRS distributions across ancestry*

147 To harmonize the polygenic risk scores (PRS) across diverse ancestries, we first  
148 projected genotypes of individuals from the UKBB and GISC onto a principal components  
149 analysis (PCA) space using PC loadings derived from 55,248 variants within 1000G  
150 dataset (**Methods**). The resulting PC scores aligned closely with the continental  
151 ancestries represented in the 1000G, confirming that projecting genotypes to an  
152 externally-defined PC-space still maintains similar clustering by ethnicity (**Figure 3**). The  
153 SNPs used and their corresponding loadings for the top five PCs are detailed in  
154 **Supplementary Table 4**.

155 We constructed PRSs for lung cancer and smoking cessation using large multi-  
156 ancestry GWAS with publicly accessible summary statistics (**Methods, Supplementary**  
157 **Tables 2-3**). Variations in the raw PRS across ancestries were notable (**Figure 4**). For  
158 instance, only 6% of AFR individuals in the UKBB cohort had a raw lung cancer PRS  
159 above the 80th percentile when benchmarked against the 1000G distribution. Conversely,  
160 57% of EAS individuals ranked below the 33rd percentile for the difficulty quitting PRS  
161 (**Supplementary Table 6a-b**). Such differences indicate that applying a universal cutoff  
162 for PRS without ancestry adjustment could lead to skewed risk profiling and inaccurate  
163 clinical recommendations. Even with the smaller sample size of the GISC dataset, there  
164 were noticeable difference in PRS distribution similar to those in the UKBB data  
165 (**Supplementary Tables 6c-d**).

166 To address this, we use a two-step ancestry adjustment procedure that regresses  
167 out ancestry PCs from the raw PRS such that the mean and variance of the PRS  
168 distribution are consistent across all populations (**Methods, Figure 4, Supplementary**  
169 **Table 5**). This adjustment step places individuals from different ancestries on a  
170 standardized scale, enabling the use of a single risk stratification cutoff irrespective of an  
171 individual's ancestral background. After ancestry adjustment, the corresponding  
172 proportions of the UKBB individuals within each risk category closely match 20%-60%-  
173 20% for lung cancer, and 33.3%-33.3%-33.3% for difficulty quitting, so that patients of  
174 any background can be appropriately compared against a single reference distribution for  
175 each outcome (**Supplementary Table 6a-b**).

176

#### 177 *Risk stratification for lung cancer and smoking cessation*

178 All participants receiving interventions in our two ongoing trials are high-risk  
179 primary care patients who meet the criteria for lung cancer screening with elevated risks  
180 of lung cancer. Thus, we will assign patients to one of three PRS-based risk categories –  
181 “at risk”, “high risk”, and “very high risk” – using percentile cutoffs of ancestry-adjusted  
182 PRS distributions based on 1000G. To quantify patient risk, we calculated odds ratios  
183 (ORs) relative to the “at risk” group using 350,154 UKBB participants for lung cancer and  
184 152,406 for difficulty quitting (**Figure 5, Supplementary Tables 4-5**). For lung cancer,  
185 individuals within the 0-20th percentiles of the adjusted PRS distribution were categorized  
186 as “at risk”, those in the 20-80th percentiles as “high risk”, and the 80-100th percentiles  
187 as “very high risk”. These percentiles yielded overall ORs of 1.42 (95% CI: 1.24 – 1.65)  
188 for “high risk” and 1.85 (95% CI: 1.58 – 2.18) for “very high risk” group compared to the  
189 “at risk” group (**Supplementary Table 7a**). Notably, the ORs derived from our ancestry-  
190 adjusted PRS were nearly identical to those obtained by matching individuals' raw PRS  
191 values with ancestry-specific distributions in 1000G (**Supplementary Table 7b**). This  
192 indicates that ancestry adjustment not only preserves similar results as the ancestry-  
193 matched approach but also crucially supports the inclusion and robust risk stratification  
194 of individuals with mixed or unknown ethnic backgrounds.

195 However, differences in ORs were still observed between EUR and non-EUR  
196 participants, potentially due to the limited number of non-EUR cases (67 out of 1,830 total  
197 lung cancer cases) in the UKBB cohort, which is predominantly EUR. The OR for the EUR  
198 “very high risk” group was 1.85 (95% CI: 1.57 – 2.19), which is consistent with the  
199 combined odds ratios across all ancestries (**Supplementary Table 7a**). The OR for all  
200 non-EUR samples in the “very high risk” category was modestly reduced to 1.63 (95% CI:  
201 0.78 – 3.51) (**Supplementary Table 7b**). Given the small number of non-European cases,  
202 enhanced diversity in biobank-scale validation data should better illustrate the difference  
203 in risk stratification between raw and ancestry-adjusted PRS.

204 Since difficulty quitting is a behavior trait with no established absolute risk rates  
205 like cancer, we use terciles (0-33<sup>rd</sup>, 33<sup>rd</sup>-67<sup>th</sup>, and 67-100<sup>th</sup> percentiles) of the ancestry-

206 adjusted PRS distribution to provide provides with slightly more agnostic risk information.  
207 The resulting ORs among UKBB participants were 1.19 (95% CI: 1.15 – 1.22) for "high  
208 risk" and 1.36 (95% CI: 1.32 – 1.41) for "very high risk" relative to "at risk"  
209 (**Supplementary Table 8a-b**). Further validation using smoking status outcomes from the  
210 GISC trial assessed showed higher overall ORs in both risk categories, but risk  
211 stratification using the ancestry-adjusted PRS distribution still reflected similar odds ratios  
212 as ancestry-matched PRS distributions (**Supplementary Table 9a-b**). However, since the  
213 GISC trial have much smaller sample size compared to UKBB, the confidence intervals  
214 were notably wider. Similar to the lung cancer analysis, the ORs using our ancestry-  
215 adjusted PRS aligned closely those derived from ancestry-matched raw PRS.

216 Using ancestry-adjusted PRS ensures equitable risk stratification across all ethnic  
217 backgrounds, a critical consideration given the substantial variability in raw PRS  
218 distributions across diverse populations (**Figure 4**). The outcome-based validation in  
219 UKBB and GISC further verify that the ancestry-adjusted PRS yields valid risk  
220 stratification. These findings collectively facilitate more robust and standardized  
221 application of PRS in clinical reporting.

222

#### 223 *Translating polygenic risk scores into clinical reports*

224 We highlight two example trials: PRECISE (NIDA Grant 5R01CA268030-02) and  
225 MOTIVATE (NIDA Grant 5R01DA056050-02), designed to promote health behavior  
226 change using genetically-informed interventions, RiskProfile and PrecisionTx,  
227 respectively. These interventions incorporate PRS in communicating precision risk of lung  
228 cancer and precision benefits of smoking cessation to promote evidence-based practices  
229 such as cancer screening and tobacco treatment in high-risk individuals who smoke  
230 and/or are eligible for lung cancer screening. PRS risk stratification from either RiskProfile  
231 or PrecisionTx and clinical information are delivered within a comprehensive report, along  
232 with actionable recommendations to reduce lifetime risk (**Figure 6**). Access to 23andme  
233 genotypes and expanded health information has been a motivating component for the  
234 research participants. Both PRECISE and MOTIVATE are currently in the preliminary  
235 phases of recruiting primary care providers and patients. The recruitment strategy aims  
236 to engage over 100 physicians and 1600 patients in these trials.

237

#### 238 **Discussion**

239 In this study, we introduce the GREAT framework in primary care. The application  
240 of PRSs in the two example trials offers precise risk estimates for lung cancer and difficulty  
241 quitting to high-risk individuals to activate behavior change mechanisms that promote  
242 health. To enable the interventions, we present a feasible translational roadmap to  
243 transform genetic data, implemented in two example PRS-enabled interventions  
244 designed to promote health behaviors. These behavior-change tools will be evaluated for

245 implementation and effectiveness in motivating patients at high risk to reduce their risk by  
246 increased cancer screening and smoking cessation.

247 Importantly, we have accomplished our goals of generating behavior-change  
248 interventions by a) framing our translational message specifically for high-risk patients  
249 who have not received guideline-recommended cancer screening or tobacco treatment<sup>21-</sup>  
250 <sup>24</sup>, b) translating risk categories into precision risk and benefit that are designed to  
251 motivate health behavior changes, and c) ensuring inclusion of diverse ancestry with PC-  
252 regression-based PRS adjustment.

253 Our goal is to enable a robust implementation of PRS in currently funded clinical  
254 trials to evaluate the efficacy of these relatively novel interventions. The GREAT  
255 framework guides the implementation of PRS-enabled interventions in primary care  
256 settings. Critical questions such as timing, methodology, and location of these  
257 interventions' delivery to patients and providers are addressed to optimize its acceptability,  
258 understanding, and potential impact.

259 Our approach to ancestry adjustment of PRS employs widely accessible data from  
260 the 1000G dataset, as an alternative to methods in the GenoVA<sup>11</sup> and eMERGE<sup>10</sup> studies  
261 that use data from the Mass-General Brigham Biobank and All of Us, respectively. We  
262 have validated the transferability of our 1000G-based standardization in external datasets  
263 from the UKBB and GISC, allowing future trials to adopt a similar methodology  
264 irrespective of their specific genetic data. By utilizing our provided PC loadings and PRS  
265 standardization formula for lung cancer and difficulty quitting based on 1000G data, new  
266 patients in these trials can receive accurate risk categorization reports, bypassing the  
267 potential inaccuracies of self-reported ethnicity and the need for re-training PCA models.

268 A notable gap in current practice is the absence of genetic information in electronic  
269 health records (EHRs) for decision support and the lack of PRS generation in clinical labs.  
270 Implementing multilevel precision interventions in primary care necessitates a workflow  
271 that incorporates the use of EHRs for recruitment, protocols of biomarker testing, and a  
272 standardized process to generate the personalized intervention reports<sup>25</sup>. This requires  
273 collaborations with primary care stakeholders, community advisory boards, genetic  
274 counseling, and health communication to improve the messaging and visualization for  
275 intervention clarity, accuracy, and impact<sup>26-30</sup>. Patients expressed a notable interest in  
276 receiving personalized interventions. In our previous study, 85% of smoking patients  
277 reported a high interest in receiving genetically tailored tobacco treatment<sup>31</sup>. Further, a  
278 substantial majority (95%) of individuals who smoke endorsed the importance of receiving  
279 genetic results, in particular to guide their treatment<sup>27</sup>. Following receipt of a personalized  
280 genetic risk profile for smoking cessation, 91% of participants who smoke found the tool  
281 to be highly useful, most notably to better understand their health, cope with health risks,  
282 and feel more in control of their health<sup>28</sup>. Such pronounced interest and the perceived  
283 significance of genetic data highlight the growing demand for personalized interventions  
284 among patients who smoke. Personalized interventions may further increase patient



285 compliance. For example, our study found that patients reported higher interest in taking  
286 medication (97.5% vs. 61.0%,  $p < .0001$ ) when medication was personalized based on  
287 their genetics<sup>31</sup>. These data demonstrate the translation potential of personalized  
288 genetics in motivating patients for positive behavior change.

289 Unlike most current research that evaluates PRS-enabled interventions in general  
290 patient populations, our work provides a unique aspect by designing and evaluating these  
291 interventions specifically among high-risk patients who will benefit tremendously from the  
292 recommended health behaviors (lung cancer screening and smoking cessation) when  
293 general medical advice is not enough to motivate such behaviors.

294 Here we share three key design considerations for best practices. First, for  
295 equitable implementation of precision health interventions, tools must be designed with  
296 racial/ethnic minority communities engaged in the development process at the outset,  
297 rather than solely examining whether these interventions work for these communities post  
298 hoc. We engage in ongoing participatory sessions with racial and ethnic minority  
299 communities and advisory boards across all phases through iterative cycles of  
300 intervention development, feedback, and testing so that innovative genomics-informed  
301 tools are designed for use and benefit across diverse populations.

302 Second, we have chosen clinically meaningful thresholding for PRS risk categories  
303 in communicating personalized risks and benefits with patients in our research. The  
304 categories were selected because all participants receiving interventions in our two  
305 ongoing trials are high-risk primary care patients eligible for lung cancer screening, who  
306 are current or previous heavy smokers. We defined lung cancer risk by the bottom 20%,  
307 middle 60% and 20%, and difficulty quitting smoking by slightly more agnostic tertiles, to  
308 motivate positive behavior change. We aim to follow best practices of communicating  
309 uncertainties. Furthermore, we need to make decision on options of risk presentation  
310 such as a continuous or categorical assignment. Importantly, we strive to be transparent  
311 about the imprecision in both risk estimates and action thresholds for PRS.

312 Third, we expect to update our workflow to adapt to new GWAS and evolving  
313 methodologies. As scientific knowledge rapidly progresses, outdated or inaccurate PRS  
314 predictions can hinder effective implementation. To address this challenge, a dynamic  
315 PRS framework that allows for regular updates based on new scientific discoveries is  
316 necessary. This will ensure that the PRS-based intervention remains current and aligned  
317 with the latest advancements, ultimately shortening the implementation gap and  
318 maximizing its impact on preventive healthcare outcomes. Leveraging current  
319 recommendations on genetic counseling, we have established a process and threshold  
320 to incorporate new evidence into our intervention regarding smoking cessation and lung  
321 cancer risk. This process will a) adjudicate population specific new evidence regarding  
322 genetics and biomarkers, b) evaluate its impact on changes in risk levels at personal and  
323 population level, and c) develop effective communication regarding the dynamic nature  
324 of genetic evidence with patients and providers.

325           There are several limitations in our work as we hope to share our experiences to  
326 help inform the knowledge pool for the best practices in creating PRS-enabled  
327 interventions that may be disease-, population-, or context-specific. We have tailored our  
328 approach to our unique outcomes (lung cancer and difficulty quitting smoking), population  
329 (patients eligible for lung cancer screening and tobacco treatment), and context (primary  
330 care settings) to optimize the potential health impact of our intervention tools. Another  
331 notable limitation is the underrepresentation of non-European populations in the multi-  
332 ancestry GWAS employed to derive the PRS weights, as well as in the UKBB used to  
333 evaluate the PRSs. These factors may reduce the predictive power of the PRS in non-  
334 European populations. A dominance of European populations persists in most existing  
335 GWAS<sup>32–35</sup>, not limited to lung cancer and smoking cessation. With the burgeoning  
336 emphasis on incorporating minority populations in GWAS analyses and the ongoing  
337 development of new PRS approaches<sup>7,8,36–39</sup> that focus on enhancing predictive power in  
338 diverse populations, we can iteratively refine the PRS implementation in our trial to  
339 synchronize with the most contemporary advancements.

340           Many questions need to be answered in the near future. First, how can we reduce  
341 the time lag from evidence to implementation? One challenge is the constant evolution of  
342 evidence that identifies new biomarkers for treatment. For instance, our recent work  
343 highlights the potential of polygenic risk scores in guiding future treatment approaches<sup>40</sup>.  
344 However, despite the presence of actionable precision treatment findings, the ever-  
345 changing evidence base and the perception that even better data are on the horizon have  
346 hindered effective implementation<sup>29,41</sup>. In this proposal, we seek to overcome this  
347 challenge by utilizing cutting-edge, biology-based metabolic and genetic markers that  
348 offer robust evidence for precision treatment. The motivation behind this approach is to  
349 reduce the time lag from evidence generation to practical implementation, particularly in  
350 the context of precision medicine, where the evidence base is continuously evolving and  
351 dynamic. This affords the unique opportunity to measure and report on the time from  
352 landmark publications to implementation of key findings, an approach that is being  
353 increasingly called for in translational science<sup>42</sup>. By leveraging state-of-the-art markers,  
354 we aim to enhance the efficiency and effectiveness of precision treatment and ensure that  
355 patients can benefit from the latest and most accurate recommendations.

356           Second, can we truly evaluate the effect of precision interventions? We expect that  
357 precision interventions may activate multiple mechanistic pathways to the uptake and  
358 efficacy of lung cancer screening or tobacco treatment. Understanding potential plausible  
359 mechanisms is needed to improve or refine the intervention for intended outcomes and  
360 contexts. Third, can these precision interventions be scaled in the real-world clinics?  
361 Evidence has shown that physicians are highly receptive to guidance on medication  
362 recommendations based on biomarkers<sup>43</sup>. To reduce burden, we need to leverage  
363 existing EHR tools (e.g. Best Practice Advisories) to efficiently facilitate physician  
364 prescribing<sup>44,45</sup>. Understanding of mechanistic and implementation outcomes will guide

365 scalable, efficient delivery components for integration into clinic workflows<sup>25</sup>, trained  
366 embedded staff, and digital therapeutic tools to enable these PRS-informed behavioral  
367 interventions<sup>46</sup>.

368 In conclusion, a well-designed roadmap that validates the PRS, creates it using  
369 TE weights, translates risk into actionable categories, communicates effectively,  
370 considers patient perspectives, and accommodates evolving science is essential for the  
371 equitable and pragmatic translation of PRS into clinical care. By addressing the barriers  
372 and implementing potential solutions at each stage, we can harness the power of PRS to  
373 improve preventive healthcare and make a meaningful difference in reducing the burden  
374 of diseases like lung cancer.

375 **Acknowledgements**

376 We would like to thank Reeya Joseph for her editorial support with the introduction, Peter  
377 Kraft for his advice on our manuscript, and Scott Vrieze for his assistance with summary  
378 statistics for difficulty quitting. We would also like to thank and acknowledge the  
379 participants enrolled in the UK Biobank (obtained under UK Biobank resource application  
380 52008) and GISC trial for contributing vital data to this work.

381  
382 This research was supported by NIH Training Grant T32GM135117 and NSF Graduate  
383 Research Fellowship DGE-2140743 (T.C.), R35-3CA197449, R01-HL163560, U01-  
384 HG009088, and U01-HG012064 (X.L.), NIH Intramural Research Program (H.Z.), NIH  
385 5T32-HL007776-25, R01-DA056050, R01-CA268030, P30-CA091842-19S5, P30-  
386 CA091842-16S2 and P50-CA244431 (L.C.) and U19-CA203654.

387

388 **Author Contributions**

389 T.C., H.Z., and L.C. conceived the project, T.C., G.P., and L.F. carried out all data analyses  
390 under the supervision of H.Z. and L.C. G.S. and D.J. assisted with obtaining and  
391 processing GWAS summary statistics for difficulty quitting. T.C., G.P., H.Z., and L.C.  
392 drafted the manuscript. All co-authors reviewed and approved the final version of the  
393 manuscript.

394

395 **Conflicts of Interest**

396 Laura J. Bierut is listed as an inventor on Issued U.S. Patent 8,080,371, “Markers for  
397 Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and  
398 treatment of addiction. Michael J. Bray is an employee at ThinkGenetic, Inc. Where  
399 authors are identified as personnel of the International Agency for Research on  
400 Cancer/World Health Organization, the authors alone are responsible for the views  
401 expressed in this article and they do not necessarily represent the decisions, policy, or  
402 views of the International Agency for Research on Cancer/World Health Organization.  
403 All other authors have no conflict of interests to report.

404

405 **Data and Code Availability**

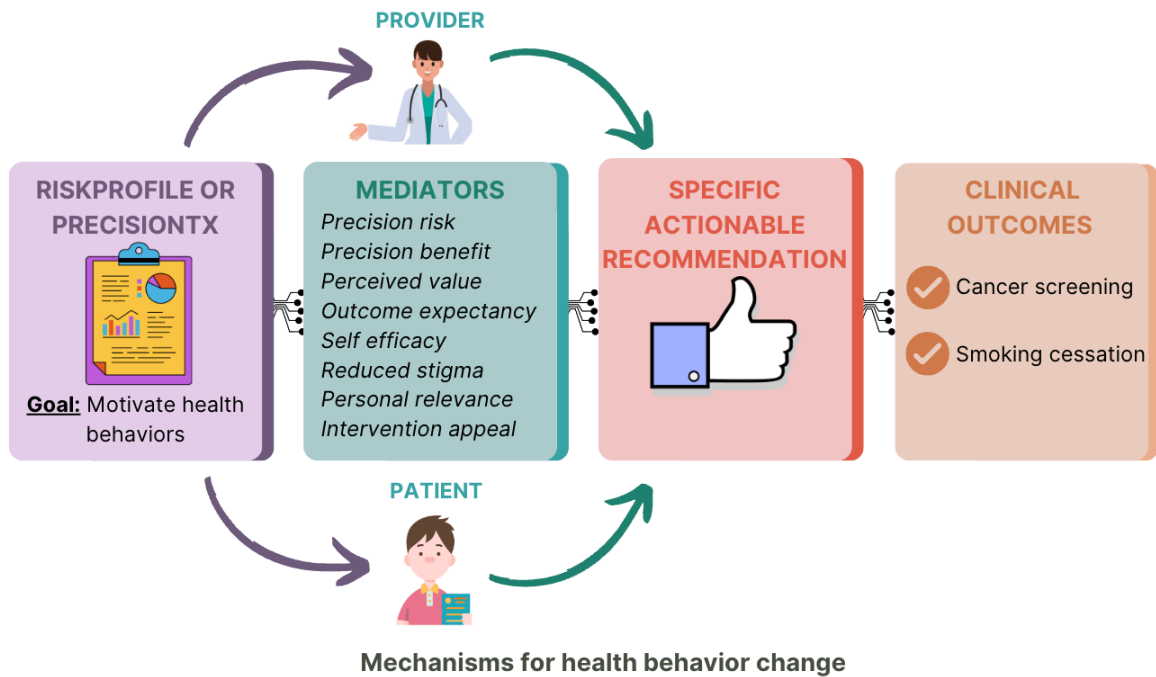
406 R code and plink commands, as well as accompanying data, used for analysis are  
407 provided in a walkthrough available on GitHub at <https://github.com/chen-tony/GREAT>.

408

409 Genotype data from the 1000 Genomes Project (Phase 3) were acquired via the plink  
410 website <https://www.cog-genomics.org/plink/2.0/resources>. Population information were  
411 obtained from public resources:

412 <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/> and  
413 [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).

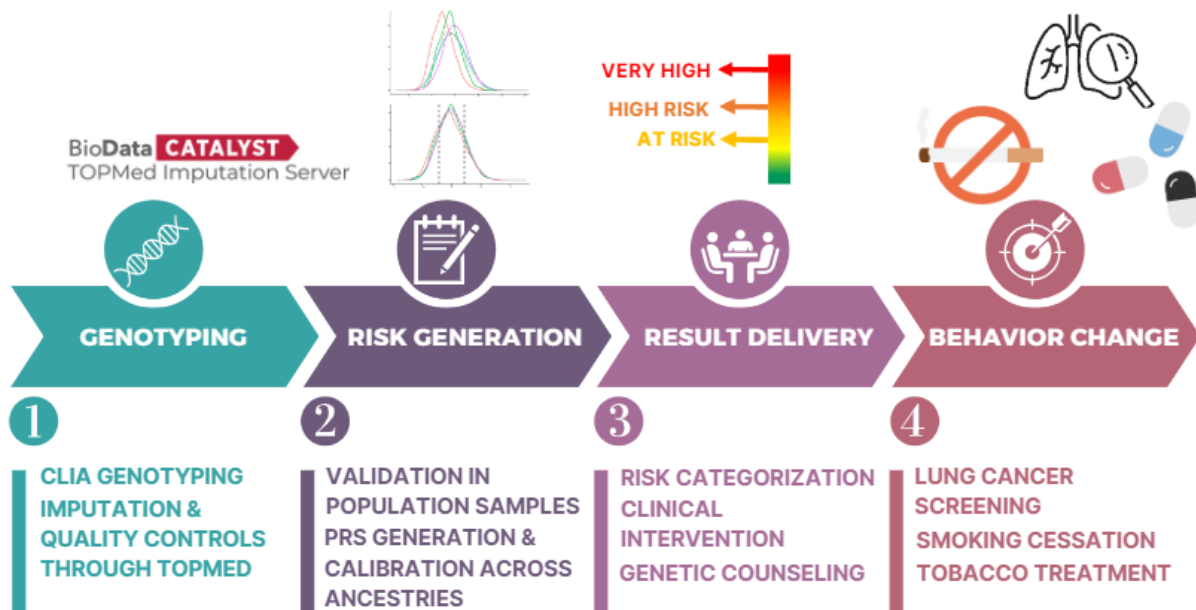
415 **Figures**



416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431

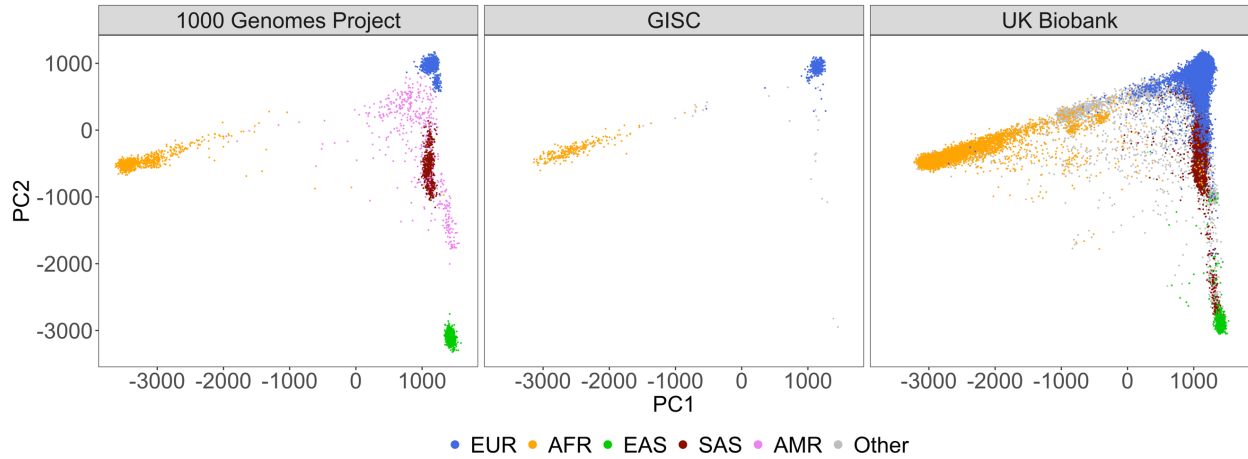
**Figure 1. Care Paradigm: Genomic Informed Care for Motivating High Risk Individuals Eligible for Evidence-based Prevention (GREAT).** The GREAT framework is a primary care paradigm that integrates genetic and clinical risk in precision health. Individuals and their providers in two upcoming trials (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 behavior changes (e.g., perceived benefit, self-efficacy, and outcome expectancy) will be evaluated. During the specific actionable recommendations phase, personalized shared decision-making will be facilitated by multilevel actions between patients and clinicians for better clinical outcomes.

432  
433



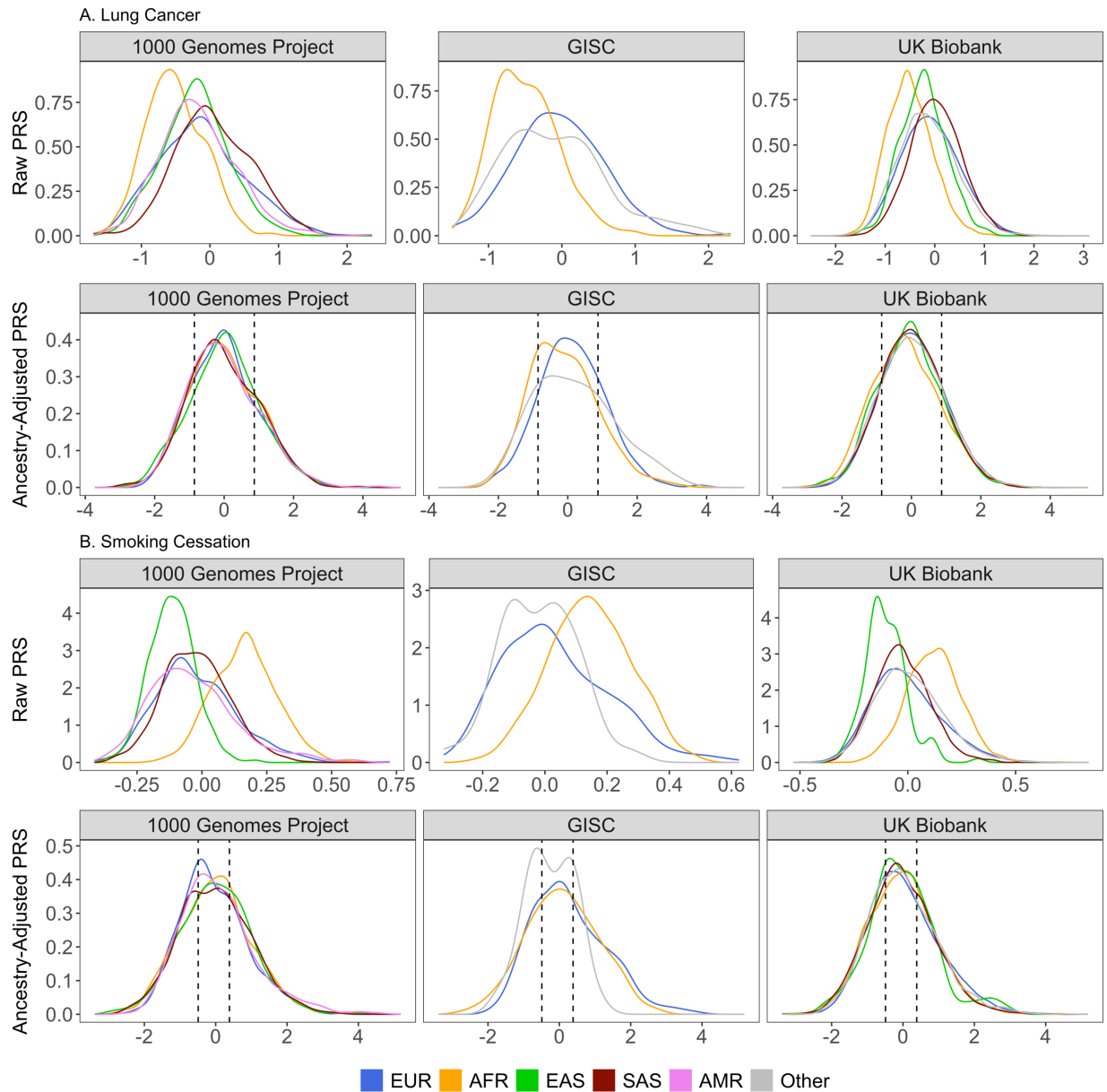
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450

**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 odds ratios. 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 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.



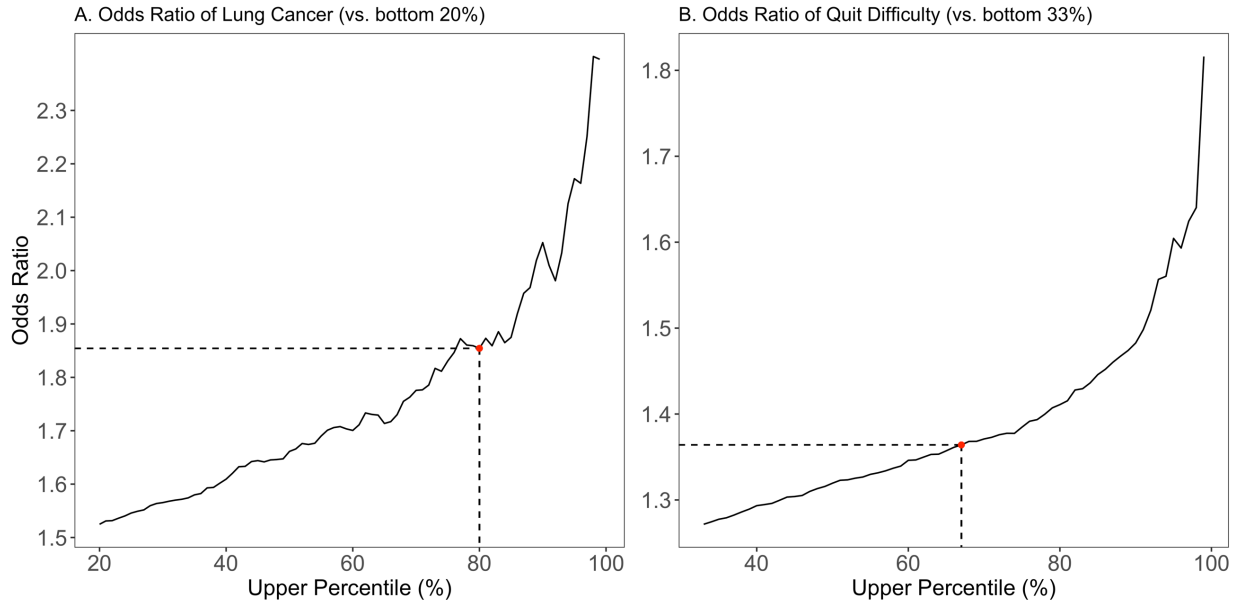
451  
 452 **Figure 3. Cross-dataset discrimination of self-reported ethnicity via PCA**  
 453 **Projections in 1000G, UKBB and GISC.** This figure illustrates the utility of principal  
 454 components analysis (PCA) loadings obtained from the 1000 Genomes Project Phase 3  
 455 (1000G) in discriminating ancestries within external datasets, specifically the UK Biobank  
 456 (UKBB) and the Genetically Informed Smoking Cessation (GISC) trial. PCA was initially  
 457 conducted on the globally diverse genotype data of 1000G. The resultant PCA-space was  
 458 then used to project genotype data from both the UKBB and GISC. The scatter plot  
 459 displays the first and second PCs for each individual in these datasets, with points  
 460 distinctly marked by self-reported ethnicity.

461  
 462



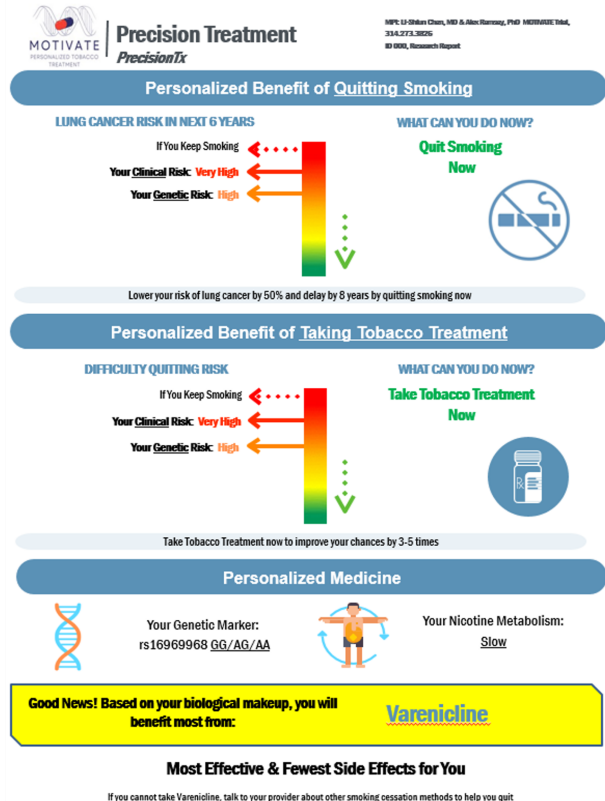
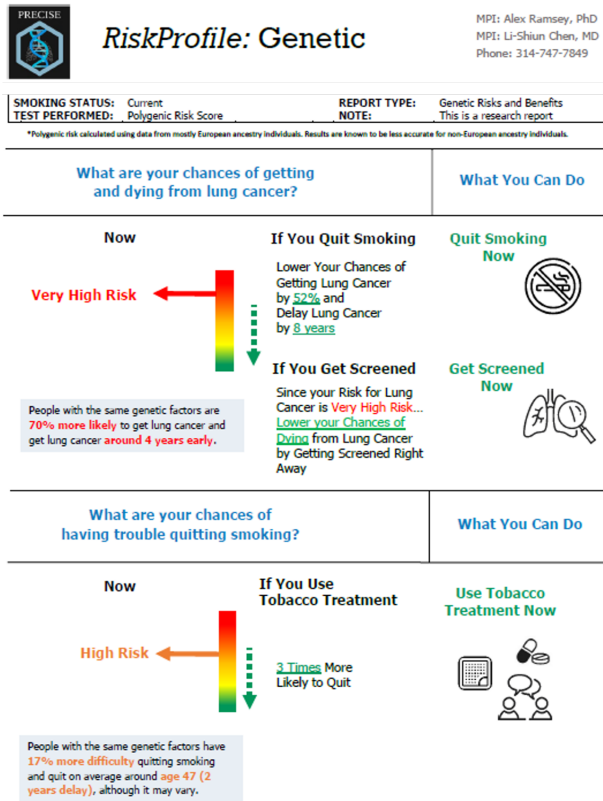
463  
 464 **Figure 4 Ancestry adjustment of PRS for lung cancer and quit difficulty PRS across**  
 465 **ancestral populations.** We showcase the adjustment process for polygenic risk scores  
 466 (PRS) for lung cancer (Panel A) and difficulty quitting smoking (Panel B) within the 1000  
 467 Genomes Project (1000G) and UK Biobank datasets. It displays both raw and ancestry-  
 468 adjusted PRS, with data points color-coded according to self-reported continental  
 469 ancestries. Ancestry adjustment effectively centers the PRS for different ancestries,  
 470 mitigating the risk of incorrect stratification due to ancestry-related biases. Dotted vertical  
 471 lines correspond to the 20<sup>th</sup> and 80<sup>th</sup> percentiles for lung cancer PRS distribution and 33<sup>rd</sup>  
 472 and 67<sup>th</sup> percentiles for difficulty quitting PRS among all 3,202 samples in the 1000  
 473 Genomes Project.





474

475 **Figure 5. Risk stratification through ancestry-adjusted PRS percentiles and**  
 476 **associated odds ratios.** This figure illustrates the odds ratios (ORs) calculated for lung  
 477 cancer (Panel A) and difficulty quitting smoking (Panel B) among UK Biobank participants  
 478 (N=340,154 for lung cancer and N=152,406 for smoking), based on selected cut points  
 479 within the ancestry-adjusted PRS distribution. The dashed lines mark the upper  
 480 percentiles used for defining risk categories in our study: the 80th percentile for lung  
 481 cancer, correlating with an OR of 1.85 (95% CI: 1.58 – 2.18) and the 67<sup>th</sup> percentile for  
 482 quit difficulty, corresponding to an OR of 1.36 (95% CI: 1.32-1.41).



483  
 484 **Figure 6. Example clinical reports for lung cancer (left) and smoking (right).** We  
 485 present two genomically-informed interventions using the GREAT framework. RiskProfile  
 486 (left) is designed to motivate lung cancer screening and tobacco treatment among eligible  
 487 patients. PrecisionTx (right) is designed to motivate precision tobacco treatment and  
 488 smoking cessation. Both tools utilize ancestry-adjusted PRS to stratify patients into “at  
 489 risk” (yellow), “high risk” (orange), and “very high risk” (red) genetic risk categories. While  
 490 RiskProfile focuses more on prevention and PrecisionTx focuses more on treatment, both  
 491 interventions expand beyond personalized risk to also provide personalized benefit of  
 492 cancer screening and personalized medication recommendation, and use a multilevel  
 493 intervention design directed to both physicians and patients in clinical settings.  
 494

495 **Tables**

|             | <b>GenoVA</b>            | <b>eMERGE</b>  | <b>Wisdom</b>                    | <b>PRECISE / MOTIVATE</b>                                |
|-------------|--------------------------|--|----------------------------------|--|
| <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 |

**Targets**

|                          |   |  |   |  |
|--------------------------|---|--|---|--|
| <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</b> PC patients (LCS eligible / smokers) and their PCPs |
| <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>                        |

**Polygenic Risk Scores**

|                            |  |  |                |  |
|----------------------------|--|--|----------------|--|
| <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<br>European weights<br>Ancestry calibration | Terra cloud platform<br>Ancestry calibration | Color genomics | Top independent variants<br>Multi-ancestry weights<br>Ancestry calibration |
| <b>Risk Stratification</b> | 2 levels (OR of 2)   | OR of 5-10                                   |                | 3 levels (top 20%, middle, bottom 20%) or tertiles                         |

**Intervention Design**

|                                     |   |  |  |  |
|-------------------------------------|---|--|--|--|
| <b>Risk representation</b>          | Only genetic<br>Format of relative risk | Combined clinical + genetic<br>Format of absolute risk | Combined clinical + genetic<br>Format of absolute risk     | Separate genetic + clinical<br>Format of relative risk           |
| <b>Messaging</b>                    | Average vs. elevated risk               | Average vs. elevated risk                              | No discussion of risk<br>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> |

496

497 \* PCP=primary care provider

498 \*\* BrCa=breast cancer, PrCa=prostate cancer, CRCa=colorectal cancer, Afib=atrial  
499 fibrillation, CAD=coronary artery/heart disease, T2D=type 2 diabetes, T1D=type 1  
500 diabetes, BMI=body mass index/obesity, CKD=chronic kidney disease,  
501 HCL=hypercholesterolemia, lung cancer=lung cancer

502

503 **Table 1. Research on PRS use in clinical settings.** We compare the PRECISE and  
504 MOTIVATE trials, part of our GREAT framework, with existing PRS-informed trials:  
505 GenoVA, eMERGE, and WISDOM. Bolded text in the PRECISE / MOTIVATE column  
506 highlight the points where our trials differ from the current trials. Namely, the PRECISE  
507 and MOTIVATE trials investigate lung cancer and smoking and will focus on high risk  
508 patients who are smokers or eligible for lung cancer screening. We also look at lung  
509 cancer screening, tobacco treatment, and smoking cessation as unique target outcomes.  
510 Finally, in addition to genetic and clinical risk messaging, the two trials have a unique  
511 emphasis on behavior mechanisms around lung cancer and smoking.

512

513 **References**

- 514 1. Kocarnik, J. M. *et al.* Cancer Incidence, Mortality, Years of Life Lost, Years Lived  
515 With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From  
516 2010 to 2019. *JAMA Oncol* **8**, 420 (2022).
- 517 2. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of  
518 polygenic risk scores. *Nature Reviews Genetics* vol. 19 581–590 Preprint at  
519 <https://doi.org/10.1038/s41576-018-0018-x> (2018).
- 520 3. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical  
521 instruments. *Genome Medicine* vol. 12 Preprint at [https://doi.org/10.1186/s13073-](https://doi.org/10.1186/s13073-020-00742-5)  
522 [020-00742-5](https://doi.org/10.1186/s13073-020-00742-5) (2020).
- 523 4. Adeyemo, A. *et al.* Responsible use of polygenic risk scores in the clinic: potential  
524 benefits, risks and gaps. *Nature Medicine* vol. 27 1876–1884 Preprint at  
525 <https://doi.org/10.1038/s41591-021-01549-6> (2021).
- 526 5. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of  
527 Polygenic Risk Scores. *Am J Hum Genet* **97**, 576–592 (2015).
- 528 6. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk  
529 to disease from genome-wide association studies. *Genome Res* **17**, 1520–1528  
530 (2007).
- 531 7. Zhang, H. *et al.* A new method for multiancestry polygenic prediction improves  
532 performance across diverse populations. *Nat Genet* (2023).
- 533 8. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations.  
534 *Nat Genet* **54**, 573–580 (2022).
- 535 9. Chen, T., Zhang, H., Mazumder, R. & Lin, X. Ensembled best subset selection  
536 using summary statistics for polygenic risk prediction. *bioRxiv* (2023).
- 537 10. Linder, J. E. *et al.* Returning integrated genomic risk and clinical  
538 recommendations: The eMERGE study. *Genetics in Medicine* **25**, 100006 (2023).
- 539 11. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting  
540 workflow. *Nat Med* **28**, 1006–1013 (2022).
- 541 12. Shieh, Y. *et al.* Breast Cancer Screening in the Precision Medicine Era: Risk-  
542 Based Screening in a Population-Based Trial. *J Natl Cancer Inst* **109**, djw290  
543 (2017).
- 544 13. Zhang, P. *et al.* Association of smoking and polygenic risk with the incidence of  
545 lung cancer: a prospective cohort study. *Br J Cancer* **126**, 1637–1646 (2022).
- 546 14. Kanwal, M., Ding, X.-J. & Cao, Y. Familial risk for lung cancer. *Oncol Lett* **13**, 535–  
547 542 (2017).
- 548 15. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory Based on a  
549 Polygenic Risk Model. *Cancer Res* **81**, 1607–1615 (2021).
- 550 16. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society.  
551 *Science (1979)* **376**, 250–252 (2022).

- 552 17. Ding, Y. *et al.* Polygenic scoring accuracy varies across the genetic ancestry  
553 continuum. *Nature* (2023).
- 554 18. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74  
555 (2015).
- 556 19. Ge, T. *et al.* Development and validation of a trans-ancestry polygenic risk score  
557 for type 2 diabetes in diverse populations. *Genome Med* **14**, 70 (2022).
- 558 20. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the  
559 Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*  
560 **12**, e1001779 (2015).
- 561 21. US Preventive Services Task Force. Lung Cancer: Screening.  
562 [https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-](https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening)  
563 [cancer-screening](https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening) (2021).
- 564 22. Agency for Healthcare Research and Quality. Clinical Guidelines and  
565 Recommendations. <https://www.ahrq.gov/prevention/guidelines/index.html> (2021).
- 566 23. Tobacco Use and Dependence Guideline Panel. *US Department of Health and*  
567 *Human Services. (2008). Tobacco Use and Dependence Guideline Panel.*  
568 *Treating Tobacco Use and Dependence: 2008 Update.* (US Department of Health  
569 and Human Services, Rockville, MD, 2008).
- 570 24. Krist, A. H. *et al.* Screening for Lung Cancer. *JAMA* **325**, 962 (2021).
- 571 25. Ayatollahi, H., Hosseini, S. F. & Hemmat, M. Integrating Genetic Data into  
572 Electronic Health Records: Medical Geneticists' Perspectives. *Healthc Inform Res*  
573 **25**, 289 (2019).
- 574 26. Bourdon, J. L. *et al.* In-vivo design feedback and perceived utility of a genetically-  
575 informed smoking risk tool among current smokers in the community. *BMC Med*  
576 *Genomics* **14**, 139 (2021).
- 577 27. Ramsey, A. T. *et al.* Participatory Design of a Personalized Genetic Risk Tool to  
578 Promote Behavioral Health. *Cancer Prevention Research* **13**, 583–592 (2020).
- 579 28. Ramsey, A. T. *et al.* Proof of Concept of a Personalized Genetic Risk Tool to  
580 Promote Smoking Cessation: High Acceptability and Reduced Cigarette Smoking.  
581 *Cancer Prevention Research* **14**, 253–262 (2021).
- 582 29. Ramsey, A. T. *et al.* Toward the implementation of genomic applications for  
583 smoking cessation and smoking-related diseases. *Transl Behav Med* **8**, 7–17  
584 (2018).
- 585 30. Chen, L.-S., Baker, T. B., Ramsey, A., Amos, C. I. & Bierut, L. J. Genomic  
586 medicine to reduce tobacco and related disorders: Translation to precision  
587 prevention and treatment. *Addiction Neuroscience* **7**, 100083 (2023).
- 588 31. Chiu, A. *et al.* Most Current Smokers Desire Genetic Susceptibility Testing and  
589 Genetically-Efficacious Medication. *Journal of Neuroimmune Pharmacology* **13**,  
590 430–437 (2018).

- 591 32. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse  
592 Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**,  
593 589–603 (2019).
- 594 33. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**,  
595 161–164 (2016).
- 596 34. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics:  
597 concerns and remedies. *Trends in Genetics* **25**, 489–494 (2009).
- 598 35. Fitipaldi, H. & Franks, P. W. Ethnic, gender and other sociodemographic biases in  
599 genome-wide association studies for the most burdensome non-communicable  
600 diseases: 2005–2022. *Hum Mol Genet* **32**, 520–532 (2023).
- 601 36. Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores  
602 across global populations. *Nat Rev Genet* (2023).
- 603 37. Jin, J. *et al.* MUSSEL: Enhanced Bayesian Polygenic Risk Prediction Leveraging  
604 Information across Multiple Ancestry Groups. *bioRxiv* (2023).
- 605 38. Zhou, G., Chen, T. & Zhao, H. SDPRX: A statistical method for cross-population  
606 prediction of complex traits. *The American Journal of Human Genetics* **110**, 13–22  
607 (2023).
- 608 39. Zhang, J. *et al.* An Ensemble Penalized Regression Method for Multi-ancestry  
609 Polygenic Risk Prediction. *bioRxiv* (2023).
- 610 40. Bray, M. *et al.* The Promise of Polygenic Risk Prediction in Smoking Cessation:  
611 Evidence From Two Treatment Trials. *Nicotine & Tobacco Research* **24**, 1573–  
612 1580 (2022).
- 613 41. Ramsey, A. T. *et al.* Designing for Accelerated Translation (DART) of emerging  
614 innovations in health. *J Clin Transl Sci* **3**, 53–58 (2019).
- 615 42. Proctor, E. *et al.* FAST: A Framework to Assess Speed of Translation of Health  
616 Innovations to Practice and Policy. *Global Implementation Research and*  
617 *Applications* **2**, 107–119 (2022).
- 618 43. Scoville, E. A. *et al.* Precision nicotine metabolism-informed care for smoking  
619 cessation in Crohn’s disease: A pilot study. *PLoS One* **15**, e0230656 (2020).
- 620 44. Chen, L.-S. *et al.* Low-Burden Strategies to Promote Smoking Cessation  
621 Treatment Among Patients With Serious Mental Illness. *Psychiatric Services* **69**,  
622 849–851 (2018).
- 623 45. Ramsey, A. T. *et al.* Care-paradigm shift promoting smoking cessation treatment  
624 among cancer center patients via a low-burden strategy, Electronic Health  
625 Record-Enabled Evidence-Based Smoking Cessation Treatment. *Transl Behav*  
626 *Med* (2019).
- 627 46. Kaphingst, K. A. *et al.* Comparing models of delivery for cancer genetics services  
628 among patients receiving primary care who meet criteria for genetic evaluation in  
629 two healthcare systems: BRIDGE randomized controlled trial. *BMC Health Serv*  
630 *Res* **21**, 542 (2021).

- 631 47. Byun, J. *et al.* Cross-ancestry genome-wide meta-analysis of 61,047 cases and  
632 947,237 controls identifies new susceptibility loci contributing to lung cancer. *Nat*  
633 *Genet* **54**, 1167–1177 (2022).
- 634 48. Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and  
635 alcohol use. *Nature* **612**, 720–724 (2022).
- 636 49. Chen, S. *et al.* A genome-wide mutational constraint map quantified from variation  
637 in 76,156 human genomes. *bioRxiv* 2022.03.20.485034 (2022).
- 638 50. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the  
639 expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-  
640 3440.e19 (2022).
- 641 51. Chen, L. *et al.* Genetic Variant in *CHRNA5* and Response to Varenicline and  
642 Combination Nicotine Replacement in a Randomized Placebo-Controlled Trial.  
643 *Clin Pharmacol Ther* **108**, 1315–1325 (2020).
- 644 52. Dey, R. *et al.* Efficient and accurate frailty model approach for genome-wide  
645 survival association analysis in large-scale biobanks. *Nat Commun* **13**, 5437  
646 (2022).
- 647 53. Hinrichs, A. S. The UCSC Genome Browser Database: update 2006. *Nucleic*  
648 *Acids Res* **34**, D590–D598 (2006).
- 649 54. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction  
650 across Diverse Populations. *The American Journal of Human Genetics* **100**, 635–  
651 649 (2017).
- 652 55. Manrai, A. K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities.  
653 *New England Journal of Medicine* **375**, 655–665 (2016).
- 654 56. Ochoa, A. *genio*: Genetics Input/Output Functions. Preprint at [https://CRAN.R-](https://CRAN.R-project.org/package=genio)  
655 [project.org/package=genio](https://CRAN.R-project.org/package=genio) (2023).
- 656 57. Price, A. L. *et al.* Principal components analysis corrects for stratification in  
657 genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
- 658 58. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and  
659 richer datasets. *Gigascience* **4**, 7 (2015).
- 660 59. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify  
661 individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219–  
662 1224 (2018).
- 663 60. Krainc, T. & Fuentes, A. Genetic ancestry in precision medicine is reshaping the  
664 race debate. *Proceedings of the National Academy of Sciences* **119**, (2022).
- 665 61. Malina, D. *et al.* *Race and Genetic Ancestry in Medicine-A Time for Reckoning*  
666 *with Racism. n engl j med* vol. 384 (2021).
- 667
- 668

669 **Methods**

670 *Reference and validation data*

671 We conducted our analyses using data from three prominent datasets: UK Biobank  
672 (UKBB)<sup>20</sup>, 1000 Genomes Project (1000G)<sup>50</sup>, and Genetically Informed Smoking  
673 Cessation Trial (GISC)<sup>51</sup>. The UKBB, a widely recognized dataset, encompasses rich  
674 genetic and clinical data from approximately 500,000 British individuals. Our study  
675 specifically used data from 340,154 unrelated multi-ancestry individuals, up to third-  
676 degree relatives<sup>52</sup>, who had consented as of September 5, 2023 (**Supplementary Table**  
677 **1**). The 1000G dataset provides a globally diverse genetic reference of 3,202 individuals,  
678 with 633 Europeans (EUR), 893 Africans (AFR), 585 East Asians (EAS), 601 South  
679 Asians (SAS), and 490 Admixed Americans (AMR). We synchronized our data by using  
680 the latest data release on hg38 reference genome and using liftOver<sup>53</sup> to convert to hg37  
681 and align with our UKBB genotype data.

682 UKBB was used for the primary validation of our PRS due to its considerable  
683 sample size and inclusion of non-European ethnicity. We used self-reported ethnicity  
684 using UKBB Field 21000, defining European (EUR) as White, British, Irish, or any other  
685 white background; African (AFR) as Black, Caribbean, African, Black or Black British, or  
686 any other black background; East Asian (EAS) as Chinese; and South Asian (SAS) as  
687 Indian, Pakistani, Bangladeshi, Asian or Asian British, or any other Asian background.  
688 From the 340,154 participants, the breakdown was 318,043 EUR, 6,409 AFR, 599 EAS,  
689 7,520 SAS, and 7,583 with other self-reported ethnicity. Participants' mean age was 56.6  
690 years (SD 8.2; range 38-81 years), and the cohort was 54.1% female (183,969  
691 individuals). Our lung cancer analyses included 1,830 lung cancer cases in the UKBB,  
692 defined by whether a patient had at least one ICD10 code in C34.0-CD34.9 under Field  
693 40006, and 338,334 controls with no ICD10 codes recorded (**Supplementary Table 1**).  
694 The smoking cessation analysis involved 152,406 "ever-smokers", including 117,483  
695 former and 34,923 current smokers, with the latter defined as having difficulty quitting  
696 based on Field 20116. We excluded 186,040 'never smokers' and 1,312 participants who  
697 opted for 'prefer not to answer'.

698 The Genetically Informed Smoking Cessation (GISC) trial is a prospective,  
699 randomized, placebo-controlled smoking cessation trial conducted at Washington  
700 University in St Louis<sup>51</sup>. This study includes 822 total individuals, all of whom are smokers.  
701 We focused on 796 individuals with genetic information, including 503 of European self-  
702 reported ethnicity, 257 African self-reported ethnicity, and 36 self-reported as "Other".  
703 GISC data was used for secondary validation, as the patient population more closely  
704 resembles the expected patients enrolled in our PRECISE and MOTIVATE trials.

705

706 *Construction of polygenic risk scores*

707 Our trial incorporates the latest findings by utilizing recently genome-wide  
708 association study (GWAS) summary statistics for lung cancer<sup>47</sup> and difficulty quitting<sup>48</sup>



709 which exclude samples from the UK Biobank to avoid overlapping with our validation data.  
710 While these meta-analyses predominantly consist of individuals of European ancestry,  
711 they also include a substantial proportion of non-European ethnic background—about 26%  
712 for lung cancer and 21% for difficulty quitting—which enhances the generalizability of the  
713 findings<sup>54,55</sup>.

714 For lung cancer risk, we started with a set of 128 published SNPs found to be  
715 predictive of 5-year and lifetime cumulative risk for lung cancer<sup>15</sup>. Out of these, 101 SNPs  
716 overlapped with the published summary statistics, reference, and validation data (1000G,  
717 UKBB, and GISC), and the 23andMe genotyping array used for the trial (**Supplementary**  
718 **Figure 1**). These SNPs were then assigned effect sizes from the fixed-effect meta-  
719 analyses estimates in the most recent lung cancer GWAS that includes EUR, AFR, and  
720 EAS ancestry<sup>47</sup>. Use of 23andme has been an incentive for patient participation.

721 For difficulty quitting, we started with 206 SNPs and SNP effects identified as  
722 predictive of smoking cessation<sup>48</sup>. Among these 206, we identified 177 SNPs following  
723 the same filtering procedure for lung cancer and used a final list of 175 SNPs after  
724 removing 2 multiallelic SNPs.

725 The PRS construction began with the alignment of genotype data to the summary  
726 statistics, ensuring consistent PRS regardless of initial reference and alternative allele  
727 coding. Specifically, for any SNP  $G$  with reversed alleles, we recoded it as  $2 - G$  to avoid  
728 discrepancies. If we were to instead change the sign of the corresponding effect size  $\beta$ ,  
729 there would be an added constant of  $2\beta$  within the PRS, which can alter the overall PRS  
730 distribution and subsequent risk stratification if patient genotypes are coded differently.  
731 Once flipped SNPs were recoded, we generated the PRS as a weighted sum of SNP  
732 dosages, utilizing the effect sizes  $\beta$  from the published summary statistics. The raw PRS  
733 for an individual patient  $i$  with  $M$  SNPs was computed as

$$734 \quad PRS_i = \beta_1 G_{i1} + \beta_2 G_{i2} + \dots + \beta_M G_{iM}.$$

735 PRS calculations were performed using R, with genotype data input via the genio  
736 package<sup>56</sup>. PRS SNPs and weights for lung cancer and difficulty quitting are provided in  
737 **Supplementary Tables 2-3**, respectively.

738

### 739 *Principal components analysis of the 1000 Genomes Project*

740 To ensure our PRS can be applied universally across ancestries, we conducted ancestry  
741 adjustment using the 1000G dataset, which provides a representative cross-section of  
742 the five major global superpopulations: AFR, AMR, EAS, EUR and SAS<sup>18</sup>. Principal  
743 components analysis (PCA) is popular tool in ancestry inference, as it can capture  
744 continental genetic diversity and provide a continuous, label-free quantification of genetic  
745 ancestry<sup>16,17,57</sup>. We performed PCA on all 3,202 1000G samples, using 55,248 SNPs that  
746 are shared among the recommended SNPs set by gnomAD<sup>49</sup>, 1000G reference data,  
747 UKBB and GISC validation data, and the 23andMe genotyping array used for the trial  
748 (Error! Reference source not found.). PCA was performed using plink 2.0<sup>58</sup> with the

749 following command to generate the top five PCs: “--pca allele-wts 5”. This process  
 750 resulted in a set of loadings or weights for each of the 55,248 SNPs corresponding to  
 751 each principal component. We then applied these loadings to genotype data from 1000G,  
 752 UKBB, and GISC within plink 2.0, employing the command: “--score [i] [j] header  
 753 cols=+scoresums,-scoreavgs,-dosagesum,-nallele --score-col-nums [k1]-[k2]” to  
 754 generate the PC scores.

755

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

757 We standardized the PRS distributions for lung cancer and difficulty quitting using  
 758 data from the 1000G dataset, employing a regression-based method to adjust for  
 759 distributional differences across ancestries<sup>11,19,59</sup>. This adjustment process involves two  
 760 key steps:

761 First, we conducted a linear regression of the raw PRS against the top five PCs  
 762 derived from the PCA, such that the PRS of individual  $i$  is a linear model of their PCs with  
 763 random noise:

$$764 \quad PRS_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \dots + \alpha_5 PC_{i5} + \epsilon_i^{mean}.$$

765 We obtained an estimated intercept  $\widehat{\alpha}_0$ , and weights  $(\widehat{\alpha}_1, \dots, \widehat{\alpha}_5)$  for each PC. The  
 766 residuals of the raw PRS, accounting for the linear effects of the PCs, were calculated as:

$$767 \quad R_i = PRS_i - \widehat{\alpha}_1 PC_{i1} - \dots - \widehat{\alpha}_5 PC_{i5}.$$

768 This first step is designed to remove mean differences in the PRS distribution across  
 769 ancestries. Subsequently, we used the square residuals  $R_i^2$  as a measure for PRS  
 770 variance, and ran a secondary linear regression model with:

$$771 \quad R_i^2 = \gamma_0 + \gamma_1 PC_{i1} + \gamma_2 PC_{i2} + \dots + \gamma_5 PC_{i5} + \epsilon_i^{var}.$$

772 From this, we derived a second estimated intercept ( $\widehat{\gamma}_0$ ) and a new set of weights  
 773  $(\widehat{\gamma}_1, \dots, \widehat{\gamma}_5)$  for each PC's effect on the variance of the PRS. The final ancestry-adjusted  
 774 PRS for each individual  $i$  was then computed as:

$$775 \quad PRS_i^{cal} = \frac{PRS_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 PC_{i1} - \dots - \widehat{\alpha}_5 PC_{i5}}{\sqrt{\widehat{\gamma}_0 + \widehat{\gamma}_1 PC_{i1} + \widehat{\gamma}_2 PC_{i2} + \dots + \widehat{\gamma}_5 PC_{i5}}}$$

776 By scaling the residuals with the fitted values from the second regression, we  
 777 standardized the variance of the PRS distribution across ancestries to mean 0 and  
 778 variance 1.

779 We validated this ancestry adjustment procedure in the UKBB and GISC datasets  
 780 by applying the PC coefficients  $(\widehat{\alpha}_1, \dots, \widehat{\alpha}_5, \widehat{\gamma}_1, \dots, \widehat{\gamma}_5)$  to the raw PRS. This ancestry-  
 781 adjusted PRS accurately reflects an individual's genetic risk independent of their ancestry,  
 782 facilitating a unified risk stratification methodology. This is especially crucial for individuals  
 783 with admixed or unknown ancestry, for whom discrete ancestry-specific prediction models  
 784 may be unsuitable or invalid<sup>4,16,36,60,61</sup>.

785

### 786 *Risk categories determination*

787 To stratify patients by genetic risk for lung cancer and difficulty quitting, we  
788 calculated odds ratios (OR). These ratios compare the probability of an outcome  
789 occurring in individuals within a percentile range  $p$  (i.e. 80-100%) of the ancestry-adjusted  
790 PRS distribution with the probability of the same outcome occurring in individuals within  
791 another percentile range  $q$  (i.e. 0-20%).

$$792 \quad OR_{pq} = \frac{P(Y = 1|PRS \in p)/P(Y = 0|PRS \in p)}{P(Y = 1|PRS \in q)/P(Y = 0|PRS \in q)}$$

793 Following the determination of the desired OR for each health outcome, we established  
794 cut points within the PRS distribution to categorize individuals into three distinct risk  
795 groups: “at risk”, “high risk”, and “very high risk”.

796 We chose clinically meaningful thresholds to define three risk categories – “at risk”,  
797 “high risk”, and “very high risk” – in communicating personalized risks and benefits with  
798 patients in our research. We use these category names because all participants in these  
799 two ongoing trials are high-risk patients eligible for lung cancer screening with active  
800 heavy smoking. For lung cancer, we categorize patients by the bottom 20%, middle 60%,  
801 and top 20%. For difficulty quitting, we divide the PRS distribution into thirds – using the  
802 bottom, middle, and top 33%. Since difficulty quitting is a behavior trait with no established  
803 absolute risk rates like cancer, we use these percentiles to provide slightly more agnostic  
804 risk information.

805 For our ancestry-adjusted PRS, we use the distribution among all 1000G samples  
806 to set percentile ranges and evaluate corresponding odds ratios among UKBB  
807 participants. For comparison, we also evaluate odds ratios using ancestry-matched raw  
808 PRS distributions, i.e. European-only 1000G PRS distribution for self-reported European  
809 UKBB participants. For UKBB participants with “Other” ethnic background, we use the  
810 raw PRS distribution among all 1000G samples, rather than matching to a specific group.