

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

3
4 Tony Chen^{1,*}, Giang Pham², Louis Fox², Nina Adler^{3,4}, Xiaoyu Wang^{5,6}, Jingning Zhang⁷,
5 Jinyoung Byun⁸, Younghun Han⁸, Gretchen R.B. Saunders⁹, Dajiang Liu¹⁰, Michael J.
6 Bray^{11,12}, Alex T. Ramsey², James McKay¹³, Laura Bierut², Christopher I. Amos^{8,14},
7 Rayjean J. Hung⁴, Xihong Lin^{1,15}, Haoyu Zhang^{5,†,*}, Li-Shiun Chen^{2,†,*}

8
9 ¹ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA

10 ² Department of Psychiatry, Washington University School of Medicine, St. Louis, USA

11 ³ Department of Anthropology, University of Toronto, Mississauga, ON, Canada

12 ⁴ Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research
13 Institute, Sinai Health, and University of Toronto, Toronto, Canada

14 ⁵ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda,
15 MD, USA

16 ⁶ Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer
17 Research, Leidos Biomedical Research Inc, Rockville, MD, USA

18 ⁷ Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health,
19 Baltimore, USA.

20 ⁸ Department of Medicine, Section of Epidemiology and Population Science, Institute for
21 Clinical and Translational Research, Houston, TX, USA

22 ⁹ Department of Psychology, University of Minnesota, Minneapolis, MN, USA

23 ¹⁰ Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA,
24 USA

25 ¹¹ Department of Genetic Counseling, Bay Path University, Longmeadow, MA, USA

26 ¹² ThinkGenetics, Inc, USA

27 ¹³ International Agency for Research on Cancer, World Health Organization, Lyon,
28 France

29 ¹⁴ Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston,
30 TX, USA

31 ¹⁵ Department of Statistics, Harvard University, Cambridge, USA

32
33 *Correspondence to: Tony Chen (tonychen@g.harvard.edu), Haoyu Zhang
34 (haoyu.zhang2@nih.gov) and Li-Shiun Chen (li-shiun@wustl.edu).

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

36 **Abstract**

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

42
43 **Methods:** We propose the GREAT care paradigm, which employs PRSs to stratify
44 disease risk and personalize interventions. We developed PRSs using large-scale multi-
45 ancestry genome-wide association studies and standardized PRS distributions across
46 all ancestries. We applied our PRSs to 796 individuals from the GISC Trial, 350,154
47 from UK Biobank (UKBB), and 210,826 from All of Us Research Program (AoU), totaling
48 561,776 individuals of diverse ancestry.

49
50 **Results:** Significant odds ratios (ORs) for lung cancer and difficulty quitting smoking
51 were observed in both UKBB and AoU. For lung cancer, the ORs for individuals in the
52 highest risk group (top 20% versus bottom 20%) were 1.85 (95% CI: 1.58 – 2.18) in
53 UKBB and 2.39 (95% CI: 1.93 – 2.97) in AoU. For difficulty quitting smoking, the ORs
54 (top 33% versus bottom 33%) were 1.36 (95% CI: 1.32 – 1.41) in UKBB and 1.32 (95%
55 CI: 1.28 – 1.36) in AoU.

56
57 **Conclusion:** Our PRS-based intervention model leverages large-scale genetic data for
58 robust risk assessment across populations. This model will be evaluated in two cluster-
59 randomized clinical trials aimed at motivating health behavior changes in high-risk
60 patients of diverse ancestry. This pioneering approach integrates genomic insights into
61 primary care, promising improved outcomes in cancer prevention and tobacco treatment.

62
63 **Keywords:** Lung Cancer, Polygenic risk scores, Cancer prevention, Personalized
64 interventions, Translational roadmap, Health behavior change

65 Introduction

66 The worldwide burden of lung cancer and tobacco smoking presents major challenges
67 to global health¹. Evidence-based practices to reduce their risk such as cancer
68 screening and tobacco treatment (e.g. smoking cessation medication) have long existed
69 but are infrequently used in most primary care practices. Communication of the
70 precision risk of lung cancer and precision benefit of smoking cessation is a promising
71 but untested strategy to promote health behavior changes to reduce cancer risk. To
72 address this gap, polygenic risk scores (PRSs) emerge as a valuable approach to
73 assess disease susceptibility among populations and pinpoint individuals at higher risk²⁻
74 ⁴. PRSs can be derived from large-scale genome-wide association studies (GWAS) to
75 estimate individual disease risk and have shown promise in predicting health outcomes
76 and promoting preventive healthcare⁵⁻⁹. Despite their potential, PRSs' implementation in
77 primary care is limited, especially in diverse populations. Implementing a PRS-based
78 precision intervention is crucial in order to address the multifaceted needs of different
79 communities and individuals¹⁰. Harnessing PRSs effectively can make significant
80 progress in mitigating lung cancer's public health impact.

81
82 Ongoing studies like eMERGE¹¹, GenoVA¹², and WISDOM¹³ are leading the
83 implementation of PRS into genetic risk reports (**Table 1**). They aim to personalize
84 medical reports and understand the impact of PRS on screening, diagnostic procedures,
85 and patient behavior. Notably, a gap persists as these initiatives have not yet formulated
86 a PRS specifically for lung cancer. A likely reason is that the global burden of lung
87 cancer is primarily driven by tobacco smoking rather than genetics^{14,15}. However,
88 accounting for the genetic basis of lung cancer may provide patients and clinicians with
89 additional actionable information. The unique value proposition of a lung cancer-specific
90 PRS lies in leveraging established and clear guideline-based prevention strategies,
91 including smoking cessation treatment and lung cancer screening¹⁶. By incorporating
92 PRSs for lung cancer and difficulty quitting smoking without treatment, there is an
93 opportunity to revitalize and enhance these often-under-utilized prevention practices.

94

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

111
112 Effectively translating PRSs into clinical practice requires a comprehensive and
113 pragmatic translational roadmap for equitable and effective implementation (**Figure 1**).
114 First, to address ancestry diversity, we take a two-step approach of (1) constructing
115 PRS based on large-scale multi-ancestry GWAS, and (2) standardizing PRS
116 distributions across the continuum of genetic ancestry to ensure accurate risk
117 stratification based on PRS distributions for all patients²⁰. Second, to document
118 accuracy and transportability of the PRSs to diverse populations, we perform large-
119 scale validation using diverse data from the Genetic Informed Smoking Cessation
120 (GISC) trial²¹, UK Biobank (UKBB)²², and All of Us Research program (AoU)²³. Third,
121 we translate risk into actionable categories by setting appropriate thresholds for the
122 PRS. The alignment of these thresholds with clinical significance involves many
123 considerations for meaningful risk stratification. Fourth, we propose clear and patient-
124 friendly communication strategies, including visual aids and educational materials, to
125 facilitate understanding and meaningful interactions between patients and healthcare

126 providers. Effectively communicating both the risk and precision of the PRS results is
127 challenging but essential to empower patients to make informed decisions about their
128 health. Moreover, it is crucial to consider patient perceived risk, perceived benefit, and
129 personal relevance when discussing PRS results with patients. Patients' understanding
130 and interpretation of PRS may vary, leading to differing levels of engagement in
131 preventive actions. Hence, comprehensive patient education programs can enhance
132 awareness and knowledge about PRS, its implications, and available preventive
133 measures.

134
135 In this paper, we introduce the design for two cluster randomized clinical trials (RCT): (1)
136 PRECISE, which evaluates the effectiveness of a multilevel intervention, *RiskProfile*, on
137 increasing lung cancer screening and tobacco treatment utilization in primary care
138 (NIDA Grant 5R01CA268030-02); and (2) MOTIVATE, which evaluates the effect of
139 PrecisionTx, a multilevel intervention to promote precision tobacco treatment in primary
140 care (NIDA Grant 5R01DA056050-02). Through the innovative use of PRS, our aim is to
141 motivate lung cancer screening and tobacco treatment among high-risk patients. We
142 present a new care paradigm (**Figure 1**) and outline a translational roadmap (**Figure 2**)
143 that discusses potential barriers and solutions for implementation. By incorporating
144 personalized risk assessments, such PRS-enabled interventions have the potential to
145 significantly improve lung cancer prevention strategies and patient outcomes.

146

147 **Materials and Methods**

148 1000 Genomes Project Phase 3 reference data and principal components analysis

149 We use the 1000 Genomes Project Phase 3 (1000G)^{24,25} as a reference for genetic
150 ancestry inference, as it is publicly accessible and includes genotype data from diverse
151 populations. The 1000G dataset includes 3,202 individuals, with 633 Europeans (EUR),
152 893 Africans (AFR), 585 East Asians (EAS), 601 South Asians (SAS), and 490 Admixed
153 Americans (AMR). We conducted principal components analysis (PCA) plink 2.0²⁶ on all
154 3,202 samples, using 55,248 SNPs that are shared among the recommended SNPs set
155 by gnomAD²⁷, 1000G reference data, UKBB and GISC validation data, and the

156 23andMe genotyping array used for the trial (**Figure 3, Supplementary Figure 1,**
157 **Supplementary Table 4**).

158
159 We ran a random forest classifier to map each individual in 1000G to their respective
160 populations using the first 5 PC scores. We then projected our validation data to this PC
161 space and applied the same random forest classifier to produce “genetically-inferred”
162 ancestry labels. Individuals with a predicted probability less than 90% for any of the five
163 ancestries were labeled as “Other.” While our proposed framework removes the need
164 for labels in the clinic, we use them here for illustrative purposes in validation.

165
166 Validation in GISC, UK Biobank, and All of Us

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

173
174 GISC Trial: The GISC trial, conducted at Washington University in St
175 Louis²¹, included 822 current or previous smokers. Genetic data were available for 796
176 individuals (503 European, 257 African, 36 Other self-reported ethnicities), This dataset
177 was primarily used for ancestry inference validation due to its relevance to our
178 PRECISE and MOTIVATE trials.

179
180 UKBB: The UKBB dataset included 340,154 unrelated individuals with rich genetic and
181 clinical data. This cohort consisted of 6,844 African, 730 Admixed American, 770 East
182 Asian, 313,279 European, 7,197 South Asian, and 11,334 Other individuals by
183 genetically-inference ancestry. The lung cancer analysis included 1,830 cases (ICD10
184 codes C34.0-C34.9) and 338,334 controls. The smoking cessation analysis involved
185 152,406 ever-smokers (117,483 former and 34,923 current), excluding 186,040 never
186 smokers and 1,312 prefer-not-to-answer responses.

187

188 AoU: The AoU dataset included 210,826 unrelated individuals with whole-genome
189 sequencing data. This cohort comprised 45,108 African, 32,563 Admixed American,
190 3,873 East Asian, 110,712 European, 1,689 South Asian, and 16,881 Other individuals
191 by genetically-inference ancestry. We focused on 152,916 current or previous smokers
192 for our smoking analysis.

193

194 Construction of polygenic risk scores

195 Our trial incorporates the latest findings by utilizing recently genome-wide association
196 study (GWAS) summary statistics for lung cancer²⁸ and difficulty quitting smoking²⁹,
197 excluding UK Biobank samples to avoid overlapping with our validation data. While
198 these meta-analyses predominantly consist of individuals of European ancestry, they
199 also include a substantial proportion of non-European ancestry — about 26% for lung
200 cancer and 21% for difficulty quitting smoking — which enhances the generalizability of
201 the findings^{30,31}.

202

203 For lung cancer risk, we started with 128 published SNPs found to be predictive of 5-
204 year and lifetime cumulative risk for lung cancer¹⁶. Out of these, 101 SNPs overlapped
205 with the published summary statistics, reference, and validation data (1000G, UKBB,
206 GISC and AoU), and the 23andMe genotyping array used for the trial (**Supplementary**
207 **Figure 1**). These SNPs were assigned effect sizes from the fixed-effect meta-analyses
208 estimates in the latest lung cancer GWAS that includes EUR, AFR, and EAS ancestry²⁸.
209 For difficulty quitting smoking, we identified 175 SNPs predictive for smoking cessation
210 following the same filtering procedure for lung cancer²⁹.

211

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

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

216 PRS calculations were performed using R, with genotype data input via the genio
217 package³². PRS SNPs and weights for lung cancer and difficulty quitting smoking are
218 provided in **Supplementary Tables 2-3**, respectively.

219

220 Standardizing PRS distributions across the continuum of genetic ancestry

221 We standardized the PRS distributions for lung cancer and difficulty quitting smoking
222 using the 1000G dataset, employing a regression-based method to adjust for
223 distributional differences across ancestries (**Supplementary Tables 4-5**)²⁰. This
224 adjustment process involves two key steps:

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

$$227 PRS_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \dots + \alpha_5 PC_{i5} + \epsilon_i^{mean}.$$

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

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

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

$$232 r_i^2 = \gamma_0 + \gamma_1 PC_{i1} + \gamma_2 PC_{i2} + \dots + \gamma_5 PC_{i5} + \epsilon_i^{var}.$$

233

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

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

235 This standardization resulted in a PRS distribution with mean 0 and variance 1 across
236 ancestries, ensuring that genetic risk is accurately reflected independent of ancestry.
237 This method is crucial for individuals with admixed or unknown ancestry, where discrete
238 ancestry-specific models are inappropriate^{33,34}.

239

240 Definition of patient risk categories

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

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

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

245 We established cut points in the PRS distribution to categorize individuals into three risk
246 groups: “at risk”, “high risk”, and “very high risk”. For lung cancer, we defined these
247 categories as the bottom 20%, middle 60%, and top 20% of the PRS distribution. For
248 difficulty quitting smoking, categories were the bottom, middle, and top 33%. These
249 thresholds were chosen to communicate personalized risks and benefits to high-risk
250 patients in our trials, all of whom are eligible for lung cancer screening due to heavy
251 smoking. Since difficulty quitting smoking lacks established absolute risk rates, we used
252 these percentiles to provide a more agnostic risk assessment.

253

254 We use the distribution of our ancestry-adjusted PRSs among all 1000G samples to set
255 percentile ranges and evaluate corresponding odds ratios among UKBB and AoU
256 participants. For comparison, we also evaluate odds ratios using ancestry-matched raw
257 PRS distributions, i.e. European-only 1000G PRS distribution for UKBB and AoU
258 participants with genetically predicted European ancestry. For individual with “Other”
259 predicted ancestry, we use the overall raw PRS distribution among all 1000G samples.

260

261 **Results**

262 Harmonization PRS distributions across ancestry

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

272 across all ancestries on the same scale, allowing for a single risk stratification cutoff
273 regardless of ancestral background. Such measures importantly enable fair risk
274 assessment for individuals who may be labelled as “Other”, who may not fit well into a
275 binned ancestry-specific risk model.

276
277 Risk stratification for lung cancer and smoking cessation in UK Biobank and All of Us

278 After assigning UKB and AoU participants as “at risk”, “high risk”, or “very high risk” for
279 lung cancer and difficulty quitting smoking, we identified significant odds ratios for both
280 traits across different ancestry groups (**Figure 5, Supplementary Tables 7-10**). Our
281 ancestry-adjusted PRS yielded similar ORs as using ancestry-matched distributions,
282 demonstrating that a single PRS distribution can be appropriately applied to all
283 individuals. In UKBB, the overall ORs for lung cancer 1.42 (95% CI: 1.24 – 1.65) for the
284 “high risk” group and 1.85 (95% CI: 1.58 – 2.18) for “very high risk” group compared to
285 the “at risk” group (**Supplementary Table 7a**). In AoU, these ORs were slightly higher –
286 1.65 (95% CI: 1.36 – 2.02) for “high risk” group and 2.39 (95% CI: 1.93 – 2.97) for “very
287 high risk” group. While the ORs in non-European ancestries within UKBB were not
288 significant, they were significant and even higher than those for European ancestries in
289 AoU (**Supplementary Table 8a**). Moreover, the ORs for “Other” ancestry group in AoU
290 were much higher and more significant using the ancestry-adjusted PRS compared to
291 raw PRS, where an ancestry-matched risk distribution is least suitable (**Supplementary**
292 **Table 8b**).

293
294 We observed similar patterns for difficulty quitting smoking, where the ancestry-adjusted
295 PRS performed similar or better than using ancestry-matched distributions for raw PRS.
296 In UKBB, the overall ORs for difficulty quitting smoking were 1.19 (95% CI: 1.15 – 1.22)
297 for “high risk” group and 1.36 (95% CI: 1.32 – 1.41) for “very high risk” group
298 (**Supplementary Table 9a**). In AoU, the ORs were 1.12 (95% CI: 1.08 – 1.15) for “high
299 risk” group and 1.32 (95% CI: 1.28 – 1.36) for “very high risk” (**Supplementary Table**
300 **10a**) group. Notably, we observed higher and more significant ORs in non-European
301 AoU participants when using the ancestry-adjusted PRS compared to the raw PRS.
302 Although the improvement within each specific ancestry group was not as substantial,

303 the aggregate ORs in all non-European groups increased from 0.89 (95% CI: 0.86 –
304 0.93) to 1.05 (95% CI: 1.01 – 1.09) in the “high risk” group, and from 1.07 (95% CI: 1.03
305 – 1.11) to 1.23 (95% CI: 1.18 – 1.28) in the “very high risk” group.

306
307 Using ancestry-adjusted PRS ensures accurate risk stratification across all ethnic
308 backgrounds, a critical consideration given the substantial variability in raw PRS
309 distributions across diverse populations. The outcome-based validation in UKBB and
310 AoU further verifies that the ancestry-adjusted PRS provides valid risk stratification, and
311 often yields better risk stratification for non-European ancestries compared to ancestry-
312 matched modeling. These findings collectively facilitate a more robust and standardized
313 application of PRS in clinical reporting.

314
315 Translating polygenic risk scores into clinical reports

316 We highlight two example trials: PRECISE (NIDA Grant 5R01CA268030-02) and
317 MOTIVATE (NIDA Grant 5R01DA056050-02), designed to promote health behavior
318 change using genetically-informed interventions, RiskProfile and PrecisionTx,
319 respectively. These interventions incorporate PRS to communicate precision risk of lung
320 cancer and precision benefits of smoking cessation, promoting evidence-based
321 practices such as cancer screening and tobacco treatment in high-risk individuals who
322 smoke and/or are eligible for lung cancer screening. PRS risk stratification from either
323 RiskProfile or PrecisionTx, along with clinical information, is delivered within a
324 comprehensive report that includes actionable recommendations to reduce lifetime risk
325 (**Figure 6** and **Figure 7**). Access to 23andMe genotypes and expanded health
326 information has been a motivating component for the research participants. Both
327 PRECISE and MOTIVATE are currently in the preliminary phases of recruiting primary
328 care providers and patients. The recruitment strategy aims to engage over 100
329 physicians and 1600 patients in these trials.

330
331 **Discussion**

332 In this study, we introduce the GREAT framework in primary care, applying PRSs in two
333 trials to provide precise risk estimates for lung cancer and difficulty quitting smoking to

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

339
340 Importantly, we have accomplished our goals of generating behavior-change
341 interventions by a) framing our translational message specifically for high-risk patients
342 who have not received guideline-recommended cancer screening or tobacco
343 treatment³⁵⁻³⁸, b) translating risk categories into precision risk and benefit that are
344 designed to motivate health behavior changes, and c) ensuring inclusion of diverse
345 ancestry with PC-regression-based PRS adjustment.

346
347 Our goal is to enable a robust implementation of PRS in currently funded clinical trials to
348 evaluate the efficacy of these relatively novel interventions. The GREAT framework
349 guides the implementation of PRS-enabled interventions in primary care. Future
350 research need to address critical questions such as timing, methodology, and location of
351 these interventions' delivery to patients and providers to optimize its acceptability,
352 understanding, and impact.

353
354 Our approach to ancestry adjustment of PRS employs widely accessible data from the
355 1000G dataset, as an alternative to methods in the GenoVA¹² and eMERGE¹¹ studies
356 that use data from the Mass-General Brigham Biobank and AoU, respectively. We
357 validated the transferability of our 1000G-based standardization in external datasets
358 from the UKBB, GISC and AoU, allowing future trials to adopt a similar methodology
359 irrespective of their specific genetic data. By utilizing our provided PC loadings and PRS
360 standardization formula for lung cancer and difficulty quitting smoking, new patients in
361 these trials can receive accurate risk categorization reports, bypassing the inaccuracies
362 of self-reported ethnicity and the need for re-training PCA models.

363

364 A notable gap in current practice is the absence of genetic information in electronic
365 health records (EHRs) for decision support and the lack of PRS generation in clinical
366 labs. Implementing precision interventions in primary care necessitates a workflow that
367 incorporates EHRs for recruitment, biomarker testing protocols, and standardized
368 processes to generate personalized intervention reports³⁹. This requires collaborations
369 with primary care stakeholders, community advisory boards, genetic counseling, and
370 health communication to improve intervention clarity, accuracy, and impact⁴⁰⁻⁴⁴. Patients
371 expressed a notable interest in receiving personalized interventions. In our previous
372 study, 85% of smoking patients reported a high interest in receiving genetically tailored
373 tobacco treatment⁴⁵. Further, a substantial majority (95%) of individuals who smoke
374 endorsed the importance of receiving genetic results to guide their treatment⁴¹.
375 Following receipt of a personalized genetic risk profile for smoking cessation, 91% of
376 participants who smoke found the tool highly useful for understanding their health, cope
377 with health risks, and feeling more in control of their health⁴². Such pronounced interest
378 and the perceived significance of genetic data highlight the growing demand for
379 personalized interventions among patients who smoke. Personalized interventions may
380 increase patient compliance. For example, study found that patients reported higher
381 interest in taking medication (97.5% vs. 61.0%, $p < .0001$) when medication was
382 personalized based on their genetics⁴⁵. However, substantial evidence suggests the
383 complex interplay of fear, fatalism, and risk perception when patients face lung cancer
384 screening⁴⁶⁻⁴⁹. Therefore, it is important to evaluate the translation potential of
385 personalized genetics in motivating patients for positive behavior change.

386
387 Unlike most current research that evaluates PRS-enabled interventions in general
388 patient populations, our work uniquely focuses on designing and evaluating these
389 interventions specifically among high-risk patients who will benefit tremendously from
390 recommended health behaviors such as lung cancer screening and smoking cessation,
391 when general medical advice is not enough to motivate such behaviors.

392
393 Here we share three key design considerations for best practices. First, for equitable
394 implementation of precision health interventions, tools must be designed with active

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

400
401 Second, we have chosen clinically meaningful thresholding for PRS risk categories in
402 communicating personalized risks and benefits to patients. These categories were
403 selected because all participants in our ongoing trials are high-risk primary care patients
404 eligible for lung cancer screening, who are current or former heavy smokers. We
405 defined lung cancer risk by the bottom 20%, middle 60% and top 20%, and difficulty
406 quitting smoking by slightly more agnostic tertiles, to motivate positive behavior change.
407 We aim to follow best practices of communicating uncertainties and deciding between a
408 continuous or categorical risk presentation. Transparency about the imprecision in both
409 risk estimates and action thresholds for PRS is essential.

410
411 Third, we plan to update our workflow to adapt to new GWAS findings and evolving
412 methodologies. As scientific knowledge advances, outdated PRS predictions can hinder
413 effective implementation. A dynamic PRS framework that allows for regular updates
414 based on new discoveries is essential. This ensures that PRS-based interventions
415 remain current and aligned with the latest advancements, reducing the implementation
416 gap and maximizing their impact on preventive healthcare outcomes. Following current
417 genetic counseling recommendations, we have established a process to incorporate
418 new evidence into our intervention for smoking cessation and lung cancer risk. This
419 process will adjudicating new population-specific evidence on genetics and biomarkers,
420 evaluating its impact on personal and population-level risk changes, and effectively
421 communicating the dynamic nature of genetic evidence to patients and providers.

422
423 Our work have several limitations, but we hope contribute to the knowledge pool for the
424 best practices in creating PRS-enabled interventions that may be disease-, population-,
425 or context-specific. Our approach is tailored for our unique outcomes (lung cancer and

426 difficulty quitting smoking), population (patients eligible for lung cancer screening and
427 tobacco treatment), and context (primary care) to optimize health impact. A notable
428 limitation is the underrepresentation of non-European populations in the multi-ancestry
429 GWAS employed to derive the PRS weights. This underrepresentation may reduce the
430 predictive power of the PRS in non-European populations. The dominance of European
431 populations persists in existing GWAS⁵⁰⁻⁵³, not limited to lung cancer and smoking
432 cessation. However, with the burgeoning emphasis on incorporating minority
433 populations in GWAS analyses and the development of new PRS approaches and
434 toolkits to enhance predictive power in diverse populations^{7-9,34,54-56}, we can iteratively
435 refine PRS implementation in our trial to synchronize with the latest advancements.

436
437 Many questions remain for the near future. First, how can we reduce the time lag from
438 evidence to implementation? The constant evolution of evidence identifying new
439 biomarkers for treatment is a challenge. For instance, our recent work highlights the
440 potential of PRSs in guiding future treatment approaches⁵⁷. Despite actionable precision
441 treatment findings, the ever-changing evidence base and the perception of better data
442 have hindered effective implementation^{43,58}. To address this, we aim to use cutting-edge
443 metabolic and genetic markers that offer robust evidence for precision treatment,
444 reducing the time from evidence generation to practical application. This approach
445 allows us to measure and report the time from landmark publications to the
446 implementation of key findings⁵⁹, enhancing the efficiency and effectiveness of
447 precision treatment and ensuring patients benefit from the latest recommendations.

448
449 Second, can we truly evaluate the effect of precision interventions? We expect that
450 precision interventions may activate multiple mechanistic pathways to the uptake and
451 efficacy of lung cancer screening or tobacco treatment. Understanding these
452 mechanisms is essential to refine for intended outcomes and contexts. Third, can these
453 precision interventions be scaled in the real-world clinics? Evidence shows that
454 physicians are highly receptive to medication recommendations based on biomarkers⁶⁰.
455 To reduce burden, we need to leverage existing EHR tools (e.g. Best Practice
456 Advisories) and training to efficiently facilitate physician prescribing^{61,62}. Understanding

457 of mechanistic and implementation outcomes will guide scalable, efficient delivery
458 components for integration into clinic workflows³⁹, using trained embedded staff, and
459 digital therapeutic tools to enable these PRS-informed behavioral interventions⁶³.

460
461 In conclusion, a well-designed roadmap that validates the PRS, incorporates multi-
462 ancestry GWAS-based weights, translates risk into actionable categories,
463 communicates effectively, considers patient perspectives, and accommodates evolving
464 science is essential for the equitable and pragmatic translation of PRS into clinical care.
465 By addressing the barriers and implementing potential solutions at each stage, we can
466 leverage PRS to improve preventive healthcare and significantly reduce the burden of
467 lung cancer.

468 **Acknowledgements**

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

474
475 This research was supported by NIH Training Grant T32GM135117 and NSF Graduate
476 Research Fellowship DGE-2140743 (T.C.), R35-3CA197449, R01-HL163560, U01-
477 HG009088, and U01-HG012064 (X.L.), NIH Intramural Research Program (H.Z.), NIH
478 5T32-HL007776-25, R01-DA056050, R01-CA268030, P30-CA091842-19S5, P30-
479 CA091842-16S2 and P50-CA244431 (L.C.) and U19-CA203654.

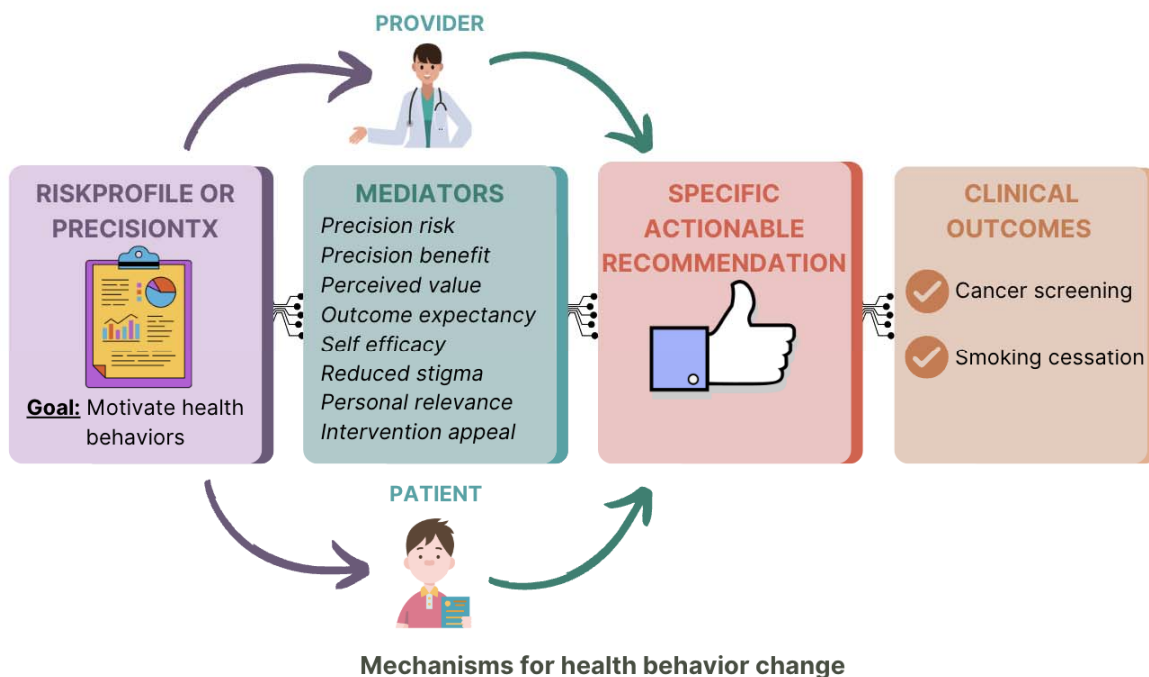
480
481 **Author Contributions**
482 T.C., H.Z., and L.C. conceived the project, T.C., G.P., and L.F. carried out all data
483 analyses under the supervision of H.Z. and L.C. G.S. and D.J. assisted with obtaining
484 and processing GWAS summary statistics for difficulty quitting smoking. T.C., G.P., H.Z.,
485 and L.C. drafted the manuscript. All co-authors reviewed and approved the final version
486 of the manuscript.

487
488 **Conflicts of Interest**
489 Laura J. Bierut is listed as an inventor on Issued U.S. Patent 8,080,371, “Markers for
490 Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis,
491 and treatment of addiction. Michael J. Bray is an employee at ThinkGenetic, Inc. Where
492 authors are identified as personnel of the International Agency for Research on
493 Cancer/World Health Organization, the authors alone are responsible for the views
494 expressed in this article and they do not necessarily represent the decisions, policy, or
495 views of the International Agency for Research on Cancer/World Health Organization.
496 All other authors have no conflict of interests to report.

497
498 **Data and Code Availability**
499 R code and plink commands, as well as accompanying data, used for analysis are
500 provided in a walkthrough available on GitHub at <https://github.com/chen-tony/GREAT>.

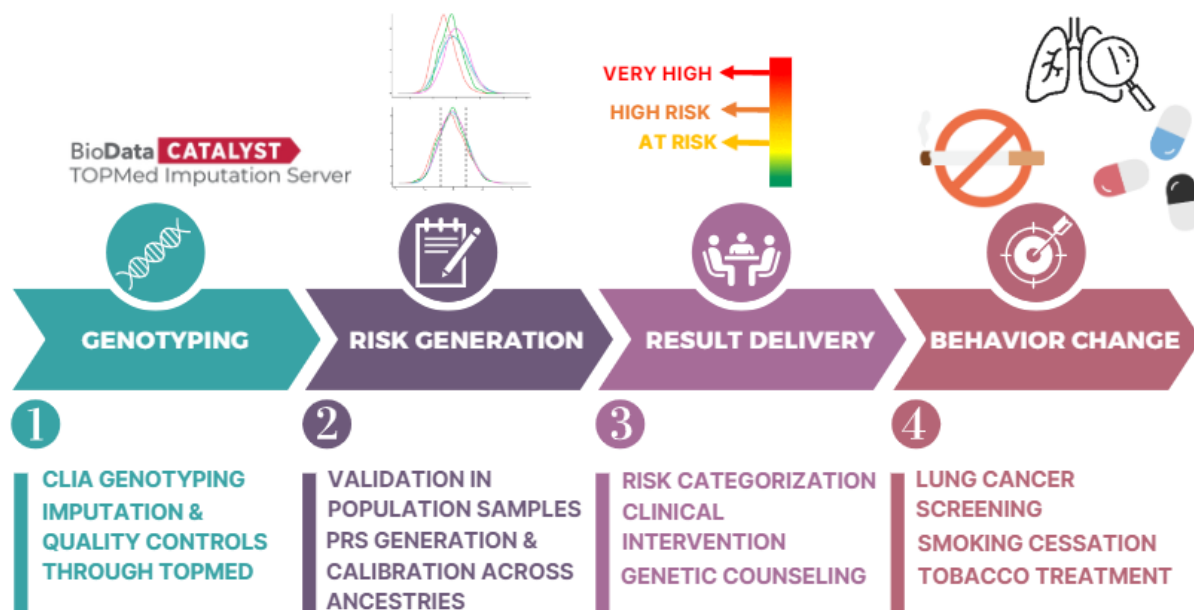
501
502 Genotype data from the 1000 Genomes Project (Phase 3) were acquired via the plink
503 website <https://www.cog-genomics.org/plink/2.0/resources>. Population information were
504 obtained from public resources:
505 <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/> and
506 https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/1000G_698_related_high_coverage.sequence.index.

508 **Figures**

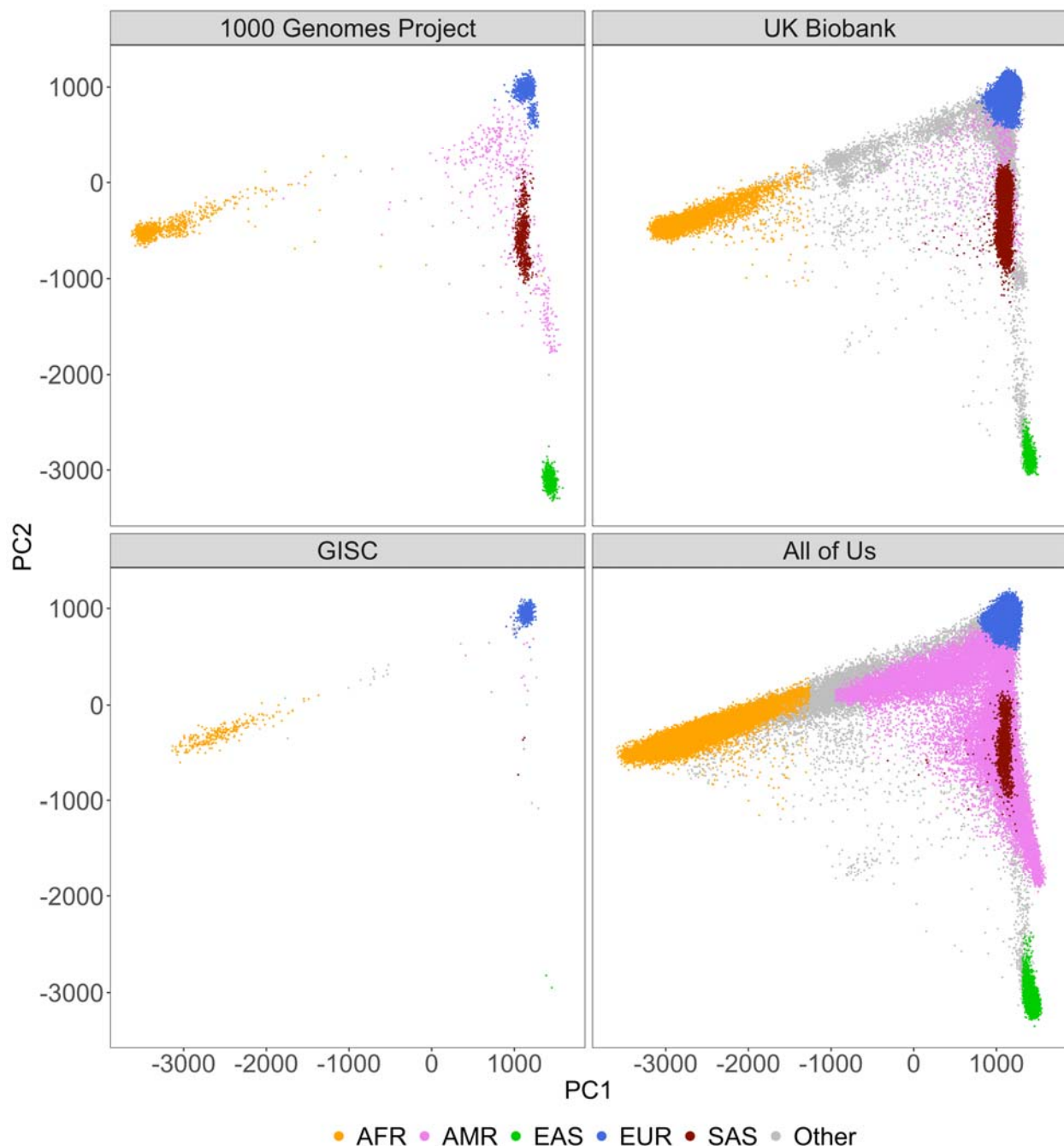


509
510 **Figure 1. Care Paradigm: Genomic Informed Care for Motivating High Risk**
511 **Individuals Eligible for Evidence-based Prevention (**GREAT**).** The GREAT
512 framework is a primary care paradigm that integrates genetic and clinical risk in
513 precision health. Individuals and their providers in two upcoming trials (PRECISE and
514 MOTIVATE) are enrolled and provided with multilevel interventions (e.g. RiskProfile and
515 PrecisionTx) to promote clinical outcomes of lung cancer screening, tobacco treatment,
516 and successful smoking cessation in primary care settings. Mechanisms of health
517 behavior changes (e.g., perceived benefit, self-efficacy, and outcome expectancy) will
518 be evaluated. During the specific actionable recommendations phase, personalized
519 shared decision-making will be facilitated by multilevel actions between patients and
520 clinicians for better clinical outcomes.

521
522

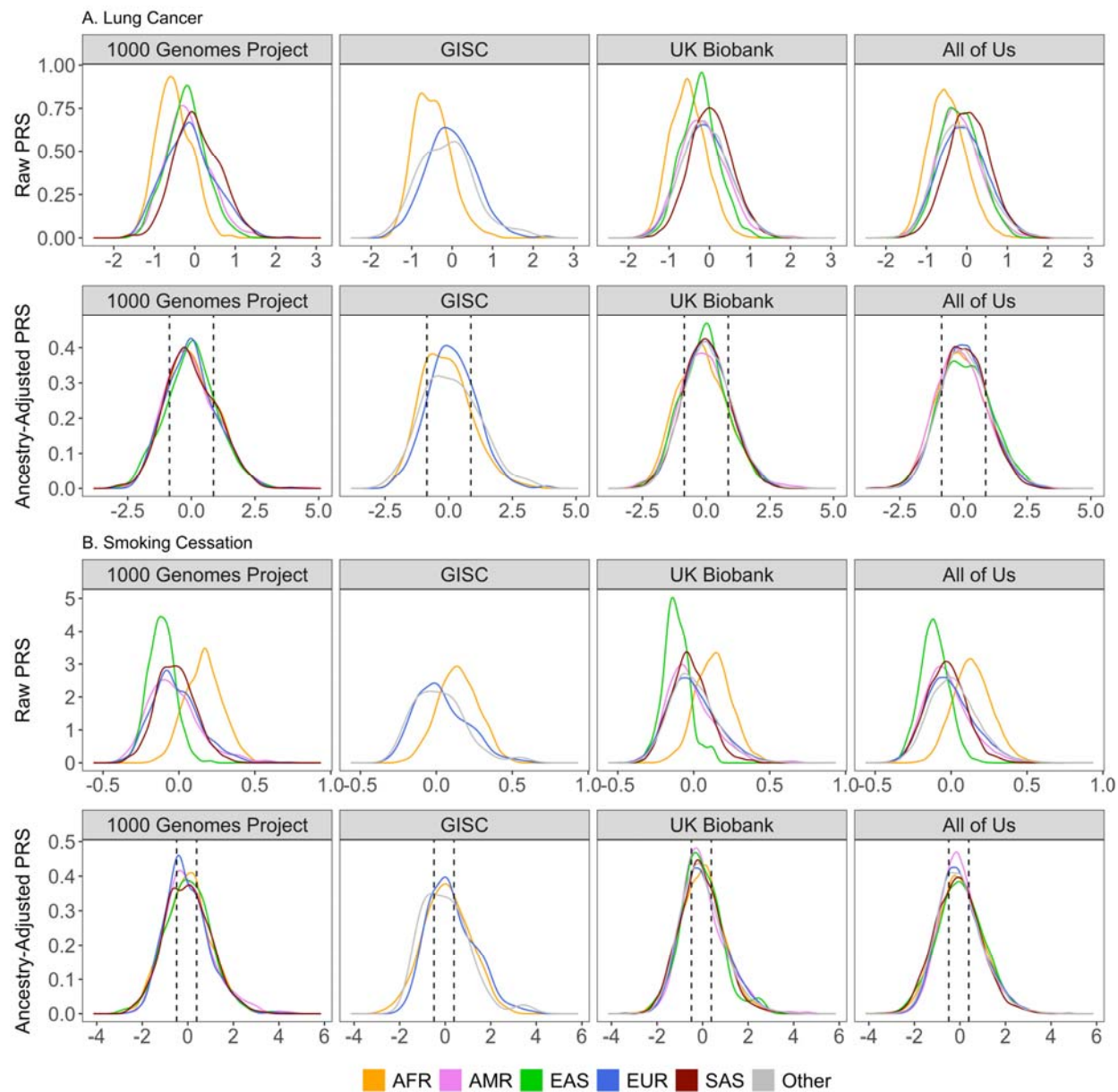


523
524 **Figure 2 Roadmap for translating genetic data to a genetic risk profile as a**
525 **multilevel intervention in primary care.** In step 1, enrolled participants' genetic data
526 are analyzed by 23andMe's Clinical Laboratory Improvement Amendments (CLIA)
527 certified genotyping process. Imputation and quality controls are conducted through the
528 Trans-Omics for Precision Medicine (TOPMed) server to ensure the integrity and
529 reliability of the genetic data, as well as to impute the GWAS variants. Step 2 involves
530 identifying available GWAS variants and weights to create the raw Polygenic Risk
531 Scores (PRS). The PRS is adjusted for genetic ancestry using reference data such as
532 the 1000 Genomes Project Phase 3 and applied to validation data such as the UK
533 Biobank to establish risk categories and compute ORs. In step 3, these scores are
534 converted into 3 risk levels based on the established thresholds. In step 4, a report with
535 precision treatment is created and communicated to both the participant and the
536 provider to make informed and educated decisions. Behavioral interventionists
537 (research staff who are trained, certified, and supervised by a team of genetic counselor,
538 psychologist, and psychiatrist) offer personalized guidance on behavior change,
539 leveraging the updated genetic insights. The outcome aims to increase lung cancer
540 screening orders, improve participant adherence, promote smoking cessation, and
541 highlight the benefits of tobacco treatment.

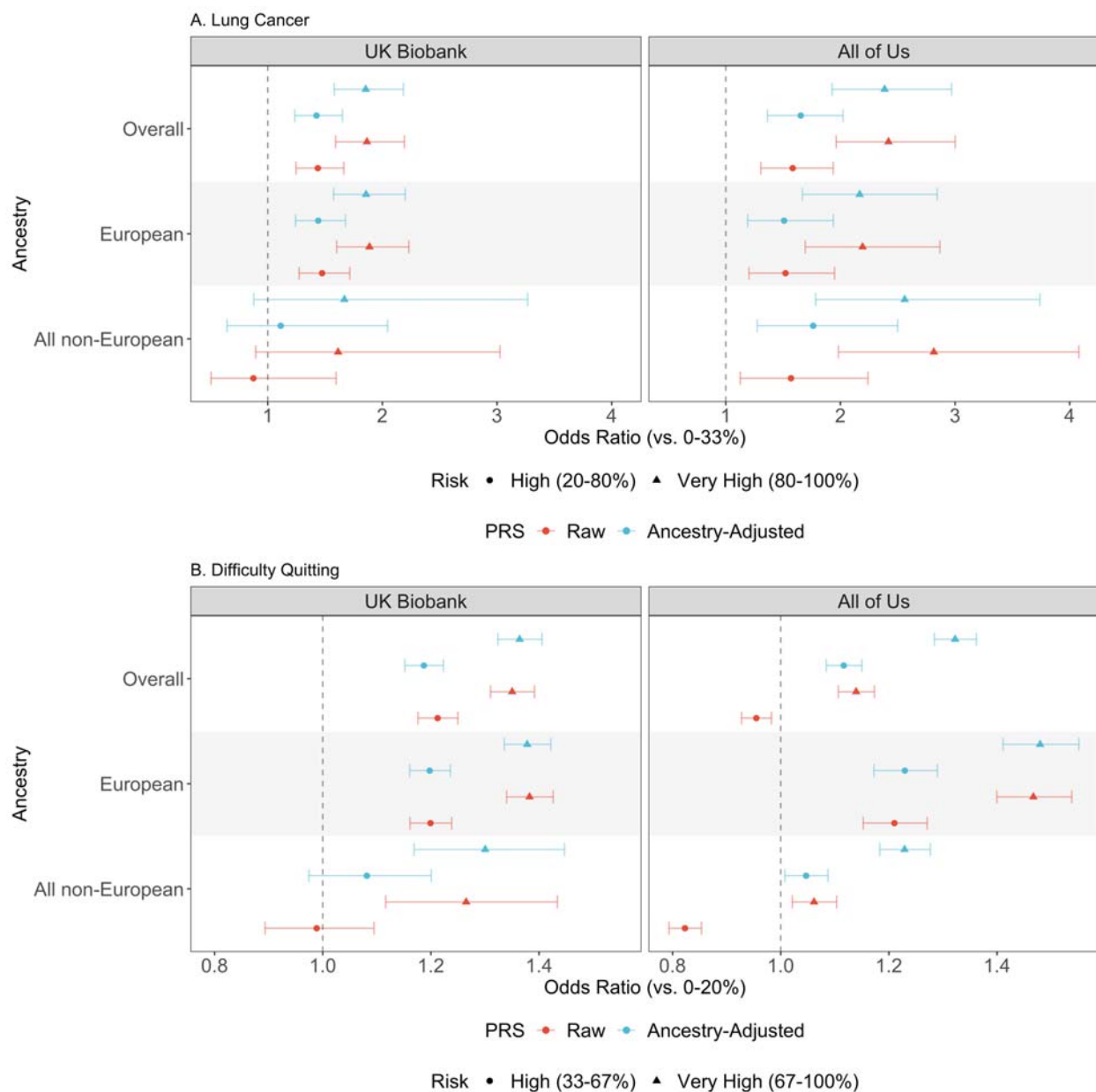


542
543 **Figure 3. Cross-dataset distribution of genetic ancestry via PCA Projections in**
544 **1000G, GISC, UKBB, and AoU.** This figure illustrates the utility of principal components
545 analysis (PCA) loadings obtained from the 1000 Genomes Project Phase 3 (1000G) in
546 discriminating ancestries within external datasets, the Genetically Informed Smoking
547 Cessation (GISC) trial. PCA was initially conducted on the globally diverse genotype
548 data of 1000G. The resultant PCA-space was then used to project genotype data from
549 GISC, UKBB, and AoU. The scatter plot displays the first and second PCs for each

550 individual in these datasets, with points distinctly marked by genetically-inferred
551 ancestry.



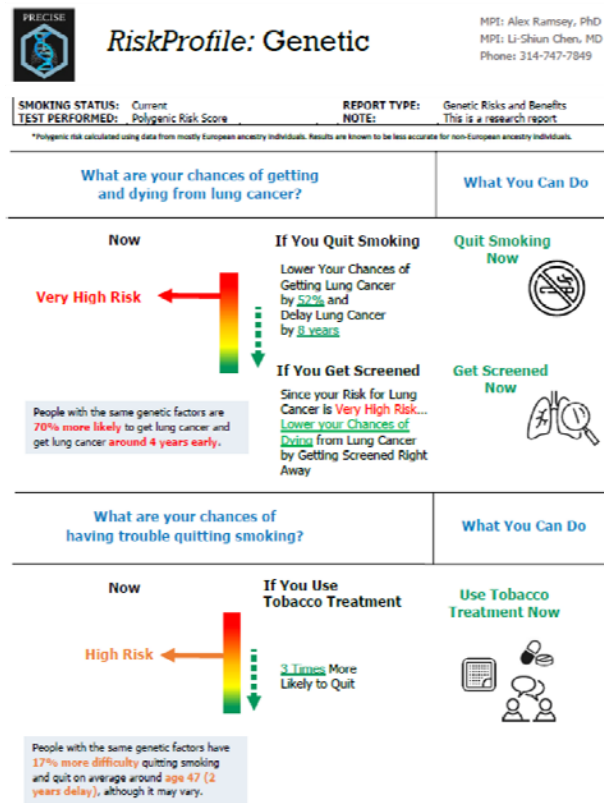
552
 553 **Figure 4 Ancestry adjustment of PRS for lung cancer and quit difficulty PRS**
 554 **across ancestral populations.** We showcase the adjustment process for polygenic
 555 risk scores (PRS) for lung cancer (Panel A) and difficulty quitting smoking (Panel B)
 556 within the 1000 Genomes Project (1000G), GISC Trial, UK Biobank, and All of Us
 557 datasets. It displays both raw and ancestry-adjusted PRS, with data points color-coded
 558 according to genetically-inferred ancestries. EAS, AMR, and SAS ancestries were
 559 removed for GISC due to their small sample sizes. Ancestry adjustment effectively
 560 centers the PRS for different ancestries, mitigating the risk of incorrect stratification due
 561 to ancestry-related biases. Dotted vertical lines correspond to the 20th and 80th
 562 percentiles for lung cancer PRS distribution and 33rd and 67th percentiles for difficulty
 563 quitting smoking PRS among all 3,202 samples in the 1000 Genomes Project.



564

565

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



573
574 **Figure 6. Example 1 clinical report: RiskProfile.** We present example 1 for
575 genomically-informed interventions using the GREAT framework. RiskProfile is
576 designed to motivate lung cancer screening and tobacco treatment among screening-
577 eligible patients. This intervention utilizes ancestry-adjusted PRS to stratify patients into
578 “at risk” (yellow), “high risk” (orange), and “very high risk” (red) genetic risk categories.
579 RiskProfile focuses on prevention and expands beyond personalized risk to also provide
580 personalized benefit of cancer screening and use a multilevel intervention design
581 directed to both physicians and patients in clinical settings. In an ongoing NCI-funded
582 trial (NCT05627674), the effect of RiskProfile on clinician ordering and patient completion of
583 lung cancer screening will be evaluated.



584
585 **Figure 7. Example 2 clinical report: PrecisionTx.** We present example 2 for
586 genomically-informed interventions using the GREAT framework. PrecisionTx is
587 designed to motivate tobacco treatment among patients who smoke. This intervention
588 utilizes ancestry-adjusted PRS to stratify patients into “at risk” (yellow), “high risk”
589 (orange), and “very high risk” (red) genetic risk categories. PrecisionTx focuses on
590 treatment and expands beyond personalized risk to also provide personalized benefit of
591 tobacco treatment, and use a multilevel intervention design directed to both physicians
592 and patients in clinical settings. In an ongoing NIDA-funded trial (NCT05846841), the
593 effect of PrecisionTx on clinician ordering, patient adherence, and smoking abstinence
594 will be evaluated.

595 **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

596
597 * PCP=primary care provider

598 ** BrCa=breast cancer, PrCa=prostate cancer, CRCa=colorectal cancer, Afib=atrial
599 fibrillation, CAD=coronary artery/heart disease, T2D=type 2 diabetes, T1D=type 1
600 diabetes, BMI=body mass index/obesity, CKD=chronic kidney disease,
601 HCL=hypercholesterolemia, lung cancer=lung cancer
602

603 **Table 1. Research on PRS use in clinical trials.** We compare the PRECISE and
604 MOTIVATE trials, part of our GREAT framework, with existing PRS-informed trials:
605 GenoVA, eMERGE, and WISDOM. Bolded text in the PRECISE / MOTIVATE column
606 highlight the points where our trials differ from the current trials. Namely, the PRECISE
607 and MOTIVATE trials investigate lung cancer and smoking and will focus on high risk
608 patients who are smokers or eligible for lung cancer screening. We also look at lung
609 cancer screening, tobacco treatment, and smoking cessation as unique target outcomes.
610 Finally, in addition to genetic and clinical risk messaging, the two trials have a unique
611 emphasis on behavior mechanisms around lung cancer and smoking.
612

613 References

- 614 1. Kocarnik, J. M. *et al.* Cancer Incidence, Mortality, Years of Life Lost, Years Lived With
615 Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019.
616 *JAMA Oncol* **8**, 420 (2022).
- 617 2. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic
618 risk scores. *Nature Reviews Genetics* vol. 19 581–590 Preprint at
619 <https://doi.org/10.1038/s41576-018-0018-x> (2018).
- 620 3. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments.
621 *Genome Medicine* vol. 12 Preprint at <https://doi.org/10.1186/s13073-020-00742-5>
622 (2020).
- 623 4. Adeyemo, A. *et al.* Responsible use of polygenic risk scores in the clinic: potential benefits,
624 risks and gaps. *Nature Medicine* vol. 27 1876–1884 Preprint at
625 <https://doi.org/10.1038/s41591-021-01549-6> (2021).
- 626 5. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic
627 Risk Scores. *Am J Hum Genet* **97**, 576–592 (2015).
- 628 6. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to
629 disease from genome-wide association studies. *Genome Res* **17**, 1520–1528 (2007).
- 630 7. Zhang, H. *et al.* A new method for multiancestry polygenic prediction improves
631 performance across diverse populations. *Nat Genet* (2023) doi:10.1038/s41588-023-
632 01501-z.
- 633 8. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat*
634 *Genet* **54**, 573–580 (2022).
- 635 9. Chen, T., Zhang, H., Mazumder, R. & Lin, X. Ensembled best subset selection using
636 summary statistics for polygenic risk prediction. *bioRxiv* (2023)
637 doi:10.1101/2023.09.25.559307.
- 638 10. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health
639 disparities. *Nat Genet* **51**, 584–591 (2019).
- 640 11. Linder, J. E. *et al.* Returning integrated genomic risk and clinical recommendations: The
641 eMERGE study. *Genetics in Medicine* **25**, 100006 (2023).
- 642 12. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow.
643 *Nat Med* **28**, 1006–1013 (2022).
- 644 13. Shieh, Y. *et al.* Breast Cancer Screening in the Precision Medicine Era: Risk-Based
645 Screening in a Population-Based Trial. *J Natl Cancer Inst* **109**, djw290 (2017).
- 646 14. Zhang, P. *et al.* Association of smoking and polygenic risk with the incidence of lung
647 cancer: a prospective cohort study. *Br J Cancer* **126**, 1637–1646 (2022).
- 648 15. Kanwal, M., Ding, X.-J. & Cao, Y. Familial risk for lung cancer. *Oncol Lett* **13**, 535–542
649 (2017).
- 650 16. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory Based on a Polygenic Risk
651 Model. *Cancer Res* **81**, 1607–1615 (2021).
- 652 17. Michie, S. Making psychological theory useful for implementing evidence based practice:
653 a consensus approach. *Qual Saf Health Care* **14**, 26–33 (2005).
- 654 18. Cane, J., O'Connor, D. & Michie, S. Validation of the theoretical domains framework for
655 use in behaviour change and implementation research. *Implementation Science* **7**, 37
656 (2012).

- 657 19. Atkins, L. *et al.* A guide to using the Theoretical Domains Framework of behaviour change
658 to investigate implementation problems. *Implementation Science* **12**, 77 (2017).
- 659 20. Ge, T. *et al.* Development and validation of a trans-ancestry polygenic risk score for type 2
660 diabetes in diverse populations. *Genome Med* **14**, 70 (2022).
- 661 21. Chen, L. *et al.* Genetic Variant in *CHRNA5* and Response to Varenicline and Combination
662 Nicotine Replacement in a Randomized Placebo-Controlled Trial. *Clin Pharmacol Ther* **108**,
663 1315–1325 (2020).
- 664 22. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a
665 Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* **12**, e1001779 (2015).
- 666 23. Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* (2024)
667 doi:10.1038/s41586-023-06957-x.
- 668 24. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- 669 25. Byrská-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000
670 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).
- 671 26. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
672 datasets. *Gigascience* **4**, 7 (2015).
- 673 27. Chen, S. *et al.* A genome-wide mutational constraint map quantified from variation in
674 76,156 human genomes. *bioRxiv* 2022.03.20.485034 (2022)
675 doi:10.1101/2022.03.20.485034.
- 676 28. Byun, J. *et al.* Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237
677 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet* **54**,
678 1167–1177 (2022).
- 679 29. Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use.
680 *Nature* **612**, 720–724 (2022).
- 681 30. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across
682 Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).
- 683 31. Manrai, A. K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities. *New*
684 *England Journal of Medicine* **375**, 655–665 (2016).
- 685 32. Ochoa, A. genio: Genetics Input/Output Functions. Preprint at [https://CRAN.R-](https://CRAN.R-project.org/package=genio)
686 [project.org/package=genio](https://CRAN.R-project.org/package=genio) (2023).
- 687 33. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society. *Science* (1979)
688 **376**, 250–252 (2022).
- 689 34. Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores across
690 global populations. *Nat Rev Genet* (2023) doi:10.1038/s41576-023-00637-2.
- 691 35. US Preventive Services Task Force. Lung Cancer: Screening.
692 [https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-](https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening)
693 [screening](https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening) (2021).
- 694 36. Agency for Healthcare Research and Quality. Clinical Guidelines and Recommendations.
695 <https://www.ahrq.gov/prevention/guidelines/index.html> (2021).
- 696 37. Tobacco Use and Dependence Guideline Panel. *US Department of Health and Human*
697 *Services. (2008). Tobacco Use and Dependence Guideline Panel. Treating Tobacco Use and*
698 *Dependence: 2008 Update.* (US Department of Health and Human Services, Rockville, MD,
699 2008).
- 700 38. Krist, A. H. *et al.* Screening for Lung Cancer. *JAMA* **325**, 962 (2021).

- 701 39. Ayatollahi, H., Hosseini, S. F. & Hemmat, M. Integrating Genetic Data into Electronic
702 Health Records: Medical Geneticists' Perspectives. *Healthc Inform Res* **25**, 289 (2019).
- 703 40. Bourdon, J. L. *et al.* In-vivo design feedback and perceived utility of a genetically-informed
704 smoking risk tool among current smokers in the community. *BMC Med Genomics* **14**, 139
705 (2021).
- 706 41. Ramsey, A. T. *et al.* Participatory Design of a Personalized Genetic Risk Tool to Promote
707 Behavioral Health. *Cancer Prevention Research* **13**, 583–592 (2020).
- 708 42. Ramsey, A. T. *et al.* Proof of Concept of a Personalized Genetic Risk Tool to Promote
709 Smoking Cessation: High Acceptability and Reduced Cigarette Smoking. *Cancer Prevention*
710 *Research* **14**, 253–262 (2021).
- 711 43. Ramsey, A. T. *et al.* Toward the implementation of genomic applications for smoking
712 cessation and smoking-related diseases. *Transl Behav Med* **8**, 7–17 (2018).
- 713 44. Chen, L.-S., Baker, T. B., Ramsey, A., Amos, C. I. & Bierut, L. J. Genomic medicine to reduce
714 tobacco and related disorders: Translation to precision prevention and treatment.
715 *Addiction Neuroscience* **7**, 100083 (2023).
- 716 45. Chiu, A. *et al.* Most Current Smokers Desire Genetic Susceptibility Testing and Genetically-
717 Efficacious Medication. *Journal of Neuroimmune Pharmacology* **13**, 430–437 (2018).
- 718 46. Quaife, S. L., Janes, S. M. & Brain, K. E. The person behind the nodule: a narrative review
719 of the psychological impact of lung cancer screening. *Transl Lung Cancer Res* **10**, 2427–
720 2440 (2021).
- 721 47. Quaife, S. L. *et al.* Psychological Targets for Lung Cancer Screening Uptake: A Prospective
722 Longitudinal Cohort Study. *Journal of Thoracic Oncology* **16**, 2016–2028 (2021).
- 723 48. Quaife, S. L., Marlow, L. A. V., McEwen, A., Janes, S. M. & Wardle, J. Attitudes towards
724 lung cancer screening in socioeconomically deprived and heavy smoking communities:
725 informing screening communication. *Health Expectations* **20**, 563–573 (2017).
- 726 49. Quaife, S. L., McEwen, A., Janes, S. M. & Wardle, J. Smoking is associated with pessimistic
727 and avoidant beliefs about cancer: results from the International Cancer Benchmarking
728 Partnership. *Br J Cancer* **112**, 1799–1804 (2015).
- 729 50. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations:
730 Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
- 731 51. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164
732 (2016).
- 733 52. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns
734 and remedies. *Trends in Genetics* **25**, 489–494 (2009).
- 735 53. Fitipaldi, H. & Franks, P. W. Ethnic, gender and other sociodemographic biases in genome-
736 wide association studies for the most burdensome non-communicable diseases: 2005–
737 2022. *Hum Mol Genet* **32**, 520–532 (2023).
- 738 54. Jin, J. *et al.* MUSSEL: Enhanced Bayesian polygenic risk prediction leveraging information
739 across multiple ancestry groups. *Cell Genomics* **4**, 100539 (2024).
- 740 55. Zhang, J. *et al.* An ensemble penalized regression method for multi-ancestry polygenic
741 risk prediction. *Nat Commun* **15**, 3238 (2024).
- 742 56. Hou, K. *et al.* Admix-kit: an integrated toolkit and pipeline for genetic analyses of admixed
743 populations. *Bioinformatics* **40**, (2024).

- 744 57. Bray, M. *et al.* The Promise of Polygenic Risk Prediction in Smoking Cessation: Evidence
745 From Two Treatment Trials. *Nicotine & Tobacco Research* **24**, 1573–1580 (2022).
- 746 58. Ramsey, A. T. *et al.* Designing for Accelerated Translation (DART) of emerging innovations
747 in health. *J Clin Transl Sci* **3**, 53–58 (2019).
- 748 59. Proctor, E. *et al.* FAST: A Framework to Assess Speed of Translation of Health Innovations
749 to Practice and Policy. *Global Implementation Research and Applications* **2**, 107–119
750 (2022).
- 751 60. Scoville, E. A. *et al.* Precision nicotine metabolism-informed care for smoking cessation in
752 Crohn’s disease: A pilot study. *PLoS One* **15**, e0230656 (2020).
- 753 61. Chen, L.-S. *et al.* Low-Burden Strategies to Promote Smoking Cessation Treatment Among
754 Patients With Serious Mental Illness. *Psychiatric Services* **69**, 849–851 (2018).
- 755 62. Ramsey, A. T. *et al.* Care-paradigm shift promoting smoking cessation treatment among
756 cancer center patients via a low-burden strategy, Electronic Health Record-Enabled
757 Evidence-Based Smoking Cessation Treatment. *Transl Behav Med* (2019)
758 doi:10.1093/tbm/ibz107.
- 759 63. Kaphingst, K. A. *et al.* Comparing models of delivery for cancer genetics services among
760 patients receiving primary care who meet criteria for genetic evaluation in two
761 healthcare systems: BRIDGE randomized controlled trial. *BMC Health Serv Res* **21**, 542
762 (2021).
- 763
- 764