

1 A novel method for predicting Lp(a) levels from 2 routine outpatient genomic testing identifies those at 3 risk of cardiovascular disease across a diverse 4 cohort

5 Natalie Telis^{*1}, Hang Dai^{*1}, Ashley Waring², David Kann³, Dana Wyman¹, Simon White¹, Basil
6 Khuder¹, Francisco Tanudjaja¹, Alexandre Bolze¹, Matthew E. Levy¹, Cassie Hajek¹, Lisa M.
7 McEwen¹, Douglas Stoller³, Christopher N. Chapman⁴, C. Anwar A. Chahal⁵, Daniel P. Judge²,
8 Douglas A. Olson⁶, Joseph J. Grzymalski^{7,8}, Nicole L. Washington¹, William Lee¹, Elizabeth T.
9 Cirulli¹, Shishi Luo¹, Kelly Schiabor Barrett¹

10

11 ¹Helix, 101 S Ellsworth Ave Suite 350, San Mateo, California 94401

12 ²Division of Cardiology, Medical University of South Carolina, 30 Courtenay Drive, MSC 592, Charleston,
13 SC 29425

14 ³Center for Inherited Cardiovascular Diseases, WellSpan Health, 157 North Reading Road, Ephrata,
15 Pennsylvania 17522

16 ⁴University of Nebraska Medical Center, Omaha, Nebraska, USA

17 ⁵Department of Pathology & Laboratory Medicine; St. Luke's University Health Network; Bethlehem PA

18 ⁶HealthPartners, Minneapolis, MN 55125

19 ⁷University of Nevada, School of Medicine; 1664 N. Virginia Reno, Nevada 89557

20 ⁸Renown Health, Reno, Nevada 89521

21 Abstract

22

23 **Background:** Lipoprotein(a) (Lp(a)) levels are a largely genetically determined and often an
24 unmeasured predictor of future Atherosclerotic Cardiovascular Disease (ASCVD). With the
25 increased use of exome sequencing in the clinical setting, there is opportunity to identify
26 patients who have a high chance of having elevated Lp(a) and are therefore at risk of ASCVD.
27 However, accurate genetic predictors of Lp(a) are challenging to design. In addition to single
28 nucleotide variants (SNVs), which are often summarized as a combined genetic risk score,
29 Lp(a) levels are significantly impacted by copy number variation in repeats of the kringle IV
30 subtype 2 domain (KIV-2), which are challenging to quantify. KIV-2 copy numbers are highly
31 variable across populations, and understanding their impact on Lp(a) levels is important to
32 creating an equitable and reliable genetic predictor of Lp(a)-driven cardiovascular risk for all
33 individuals.

34

35 **Methods:** We develop a novel method to quantify individuals' total number of KIV-2 repeats
36 from exome data, validate this quantification against measured Lp(a) levels, and then use this
37 method, combined with a SNV-based genetic risk score, to genotype an entire all-comers cohort

38 of individuals from health systems across the United States (Helix Research Network; N =
39 76,147) for an estimated Lp(a) level.

40
41 **Results:** Our combined genotyping strategy improved prediction of those with clinically-elevated
42 Lp(a) measurements across the genetically diverse cohort, especially for individuals not
43 genetically similar to European reference populations, where GRS-based estimates fall short
44 ($r^2= 0.04$ for GRS, $r^2 = 0.34$ KIV2+GRS in non-European). Importantly, high combined genetic
45 risk of high Lp(a) genotypes are significantly associated with earlier onset and increased
46 incidence in ASCVD, compared to average and low combined genetic risk genotypes in a
47 retrospective analysis of atherosclerotic diagnoses derived from electronic health records
48 (EHRs). This holds in the cohort at large (CAD HRs=1.29, 1.58), in the European subcohort
49 (HRs=1.30,1.61) as well as at trending levels of significance in individuals not genetically similar
50 to Europeans (HRs=1.22,1.31). In addition, high combined genetic risk for high Lp(a) genotypes
51 are at least 2-fold enriched amongst individuals with ASCVD diagnosis despite a lack of EHR-
52 based evidence of traditional risk factors for cardiovascular disease.

53
54 **Conclusions:** Our study demonstrates that genetically predicted Lp(a) levels, incorporating
55 both SNV and our novel KIV-2 repeat estimate, may be a practical method to predict clinically
56 elevated Lp(a). Supporting this, individuals with high combined genetic risk for high Lp(a) have
57 an increased risk for ASCVD, as evidenced across data from seven US-based health systems.

58 Introduction

59 Lipoprotein(a), or Lp(a), is an atherogenic lipoprotein particle comprised of two core
60 components: low-density lipoprotein (LDL) and apolipoprotein(a), the latter of which is
61 genetically encoded by the *LPA* gene (Marcovina and Koschinsky 1998). Circulating Lp(a)
62 levels are predominantly determined by genetics, with the majority of variability in these levels
63 linked to variations at the *LPA* locus (Utermann *et al.* 1987; Kraft *et al.* 1996; Clarke *et al.* 2009;
64 Schmidt *et al.* 2016; Tsimikas and Marcovina 2022). Lp(a) levels can vary up to a hundred-fold
65 among individuals, with different population subgroups exhibiting shifted distributions and
66 medians(Varvel *et al.* 2016; Tsimikas 2017; Tsimikas and Marcovina 2022; Dudum *et al.* 2024).

67
68 Elevated Lp(a) levels—typically cited as roughly 120 nmol/L (50 mg/dL) or above—are present
69 in 20% of the population and have been identified as a predictor of Atherosclerotic
70 Cardiovascular Disease (ASCVD), independently of established ASCVD risk factors like LDL
71 (Kamstrup *et al.* 2009; Emerging Risk Factors Collaboration *et al.* 2009; Nordestgaard *et al.*
72 2010; Koschinsky *et al.* 2024). While some studies suggest different clinical risk level thresholds
73 for primary and secondary prevention of ASCVD may be warranted, there is consensus that
74 higher levels of Lp(a) are associated with increased ASCVD risk at the population level and
75 across ethnicities, gender, baseline ASCVD status, and subpopulations such as those with Type
76 2 Diabetes (T2D) (Erqou *et al.* 2010; Varvel *et al.* 2016; Patel *et al.* 2021; Berman *et al.* 2024;
77 Wong *et al.* 2024). Aligned with this evidence, there are also emerging professional guidelines
78 that support incorporating Lp(a) screening strategies. Both universal, one-time Lp(a) screening
79 for adults or screening of those at intermediate or high risk via established CAD risk

80 assessments have been recommended (Grundy *et al.* 2019; Cegla *et al.* 2019; Tsimikas and
81 Stroes 2020; Koschinsky *et al.* 2024).

82
83 Lp(a) is not typically measured in the course of care, and accurate measurement of Lp(a) levels
84 across the spectrum of Lp(a) isoform sizes is difficult and a known limitation of current
85 laboratory tests (Wilson *et al.* 2019). Given the high heritability of Lp(a) levels, genetic
86 estimation of Lp(a) has the potential to be a practical alternative to more traditional lab
87 measurement, especially when screening populations (Reyes-Soffer *et al.* 2022; Trinder *et al.*
88 2022). Both an SNV-based genetic risk score (GRS) and KIV-2 copy number have
89 independently been used to genetically estimate clinically-relevant Lp(a) levels (Kamstrup *et al.*
90 2009; Burgess *et al.* 2018; Trinder *et al.* 2020). While both axes of variation contribute to
91 circulating Lp(a) levels, due to genotyping and sequencing prerequisites for each variation type,
92 the SNV-based GRS and KIV-2 copy number variation have seldom been studied together,
93 particularly in large cohorts (Mack *et al.* 2017; Zekavat *et al.* 2018).

94
95 It has been shown that the SNV-based GRS associates with ASCVD for individuals who self-
96 reported their ethnicity to be European; however, this same GRS was not evaluated in those
97 who self-reported other ethnicities, limiting the clinical utility of such estimations for Lp(a)-related
98 disease risk (Trinder *et al.* 2020). Recent research datasets, comprising whole genome or
99 targeted *LPA* locus sequencing data and measured Lp(a) levels, illustrate the potential benefits
100 of incorporating both SNV and KIV-2 copy number variations to genetically estimate
101 Lp(a) (Zekavat *et al.* 2018; Coassin and Kronenberg 2022). The improvements in *LPA* genetic
102 estimations are particularly noticeable in participants with self-reported non-European
103 ancestries, where smaller KIV-2 copy number alleles, which contribute to clinically significant
104 Lp(a) elevations, are more prevalent (Mehta *et al.* 2022; Tsimikas and Marcovina 2022).

105
106 We genotype a large, diverse clinicogenomic cohort derived from seven health systems across
107 the United States—the Helix Research Network (HRN)—for both *LPA* GRS and estimated total
108 KIV-2 CNV count. We show that, in the subset of individuals with measured Lp(a) values in their
109 medical records, using the combination of these genetic elements to predict those with Lp(a)
110 levels in the clinically-elevated range is much more accurate than using either element alone.
111 To test clinical utility of these genetic-based Lp(a) predictions across ancestries, we then use
112 GRS and the estimated number of KIV-2 repeats to establish low, average and high combined
113 genetic risk Lp(a) groups across the cohort and associate these risk groups to cardiovascular
114 phenotypes with established connections to Lp(a) levels (Clarke *et al.* 2009; Klarin *et al.* 2019;
115 Kronenberg *et al.* 2022).

116 Methods

117 Subjects and genetic data

118 Our study comprised 76,147 genetic samples from participants in the Helix Research Network
119 (HRN) that were sequenced and analyzed at Helix using the Exome+[®] assay as previously
120 described, recruited from the Healthy Nevada Project (n=40,430); myGenetics (n=23,113); and

121 In Our DNA SC (n=12,604) (see Supplementary Table 1) (Cirulli *et al.* 2020). The HRN studies
122 were reviewed and approved by their applicable Institutional Review Boards (IRB, projects
123 956068-12 and 21143). All participants gave their informed, written consent prior to
124 participation. All data used for research were deidentified. Demographic distributions of these
125 individuals may be found in Supplementary Tables 1-3.

126 Phenotypes

127 HRN phenotypes were processed from Epic/Clarity Electronic Health Records (EHR) data as
128 previously described and updated as of January 2024 (Cirulli *et al.* 2020). International
129 Classification of Diseases, Ninth and Tenth Revision, US Clinical Modification (ICD-9-CM and
130 ICD-10-CM) codes and associated dates were collected from available diagnosis tables (from
131 problem lists, medical histories, admissions data, surgical case data, account data, and
132 invoices). All datasets were transformed into the OMOP Common Data Model v5.4. Therefore,
133 diagnosis codes were mapped to standard SNOMED terms and represented by concept ids. To
134 be comprehensive in our cohort definition, we used both the source concept id and the standard
135 concept id to extract relevant diagnoses.

136
137 Coronary artery disease (CAD), peripheral artery disease (PAD), aortic stenosis (AS), Ischemic
138 Stroke (IS), and myocardial infarction (MI) were defined through review of each respective
139 disease diagnosis in the phecode X map(Shuey *et al.* 2023), physician input, as well as a
140 literature search around clinical diagnosis codes (ICD-10, ICD-10-CM, ICD-9-CM, SNOMED)
141 given the different natures of the source datasets. We identified the first record of each included
142 ICD-10, ICD-10-CM, ICD9-CM or SNOMED term for each disease phenotype to represent the
143 first occurrence of disease for the associated individual. A complete list of diagnosis codes
144 included in each phenotype definition may be found in Appendix Table 1.

145 Clinical measurements

146 HDL, LDL and Lp(a) measurements were identified in the HRN as any measurement matching a
147 set of codes to exhaustively capture blood assays encoded in different units (see Appendix
148 Table 2). Each measurement was transformed into a unified set of units. When measured in
149 mg/dL, Lp(a) was converted to nmol/L using a conversion factor of 2.15 as established in the
150 literature (see for example (Khera *et al.* 2014).

151 Statistical analyses

152 Statistical analyses were run using the statsmodel package in Python 3.7.3. For binary
153 variables, logistic regression or Firth logistic regression was used; for quantitative variables,
154 linear regression was used after rank-based inverse normal transformation. For time to event
155 analyses, the lifelines package was used, using age of first diagnosis received anywhere in their
156 medical record as an end point. Cox proportional hazards were adjusted for sex, age,
157 population group, and BMI, a common cardiovascular disease risk factor available for most
158 participants (n with BMI= 101,986 [94%]). Linear models were adjusted for the exome assay
159 version and KIV-2 copy number estimate (CNE) was normalized for each exome assay subset
160 (see Supplementary Figure 3).

161 Genetic similarity

162 Genetic similarity was assigned using ADMIXTURE with the 1000 Genomes Project (1KGP)
163 population set as a reference (Alexander *et al.* 2009; 1000 Genomes Project Consortium *et al.*
164 2015). To facilitate comparisons with previous studies, and to report results in populations
165 previously underrepresented cohorts in genetics research, we divided our cohort into European
166 (EUR) and non-European (non-EUR) groups as follows. For each participant, we used exome-
167 wide variant calls to calculate Admixture coefficients for five continental supergroups -- AFR,
168 AMR, EAS, EUR, SAS -- derived from data unifying 26 separate populations in the Phase 3
169 data of the 1000 Genomes project. Our EUR cohort includes only individuals with EUR > 0.85 (a
170 majority of their genetic similarity assigned to the European populations; and non-EUR < 0.1),
171 and non-EUR describes individuals with a majority of their ancestry assigned to any of the other
172 population groups ('AFR', 'AMR', 'EAS', 'SAS'). A full breakdown of population assignment and
173 demographics for the full cohort, as well as the subset with Lp(a) measurements, can be found
174 in Supplementary Table 1.

175
176 This approach follows guidelines on the use of population descriptors in genetics research,
177 where the EUR group represents individuals genetically similar to the EUR reference group from
178 the 1000 Genomes project and the non-EUR group represents individuals that are not
179 genetically similar to the EUR reference group.

180 Genetic measures associated with Lp(a)

181 We calculate a genetic risk score (GRS) to predict Lp(a) levels, using genetic variants and
182 corresponding effect estimates from (Trinder *et al.* 2020), PGS Catalog # PGP000127).

183 KIV-2 normalized coverage based copy number quantification

184
185 Exact KIV-2 copy numbers for both alleles are challenging to determine from short-read whole
186 exome sequencing data. However, a quantification of the total KIV-2 copy number difference
187 among individuals sequenced by the same assay can still be used as a numerical variable to
188 stratify variation in KIV-2 copy number for each individual. Here we developed a method to
189 quantify such a difference.

190
191 For any individual, an *LPA* allele is expected to have one copy for each of the KIV-1 and KIV-3
192 to KIV-10 domains, but an unknown number of copies of the KIV-2 domain. On reference
193 genome hg38, the *LPA* gene has six copies of KIV-2 domain with identical or slightly different
194 sequences, so we masked the second to sixth KIV-2 domains; we also masked the second
195 exon of the KIV-1 domain and the first exon of the KIV-3 domain due to their similarity to KIV-2.
196 Reads extracted from *LPA* are re-aligned to the masked chr6 reference sequence, and the
197 coverage is counted for unmasked exons of KIV domains.

198
199 The total copy number of KIV-2, estimated through KIV-2 exon 1, is quantified as the following
200 normalized coverage of KIV-2 exon 1:

201 Use $KIV-i-1$ to denote the coverage of KIV-i domain exon 1,
202 $CNE_{KIV-2-1} = KIV-2-1 / (1/8 \times \sum_i KIV-i-1)$, where $i \in \{1, 4, 5, \dots, 10\}$;
203 Similarly, that estimated through KIV-2 exon 2 is quantified as:
204 $CNE_{KIV-2-2} = KIV-2-2 / (1/8 \times \sum_{i=3}^{10} KIV-i-1)$;
205 and the final total copy number of KIV-2 is quantified as:
206 $CNE_{KIV-2} = 1/2 \times (CNE_{KIV-2-1} + CNE_{KIV-2-2})$.
207 Afterwards, this copy number of KIV-2 is normalized by exome assay version to account for
208 variability in region coverage, forming the final KIV-2 CNE.

209 Assigning genetic risk of clinically elevated Lp(a)

210 In the distribution of individuals with empirically measured Lp(a), we set out to determine the
211 GRS as well as KIV-2 CNE genetic profiles of those at the extreme ends of the spectrum, which
212 represent clinically relevant ranges. We define combined cutoffs for the GRS as well as KIV-2
213 CNE to identify individuals with a high, medium, and low genetic risk of clinically elevated Lp(a).
214 For the GRS, for high risk we choose a threshold of 120 nmol/L, as the GRS effect sizes are
215 measured in units of nmol/L, to reflect the clinically significant level. For KIV-2 CNE, as it is
216 normalized per version of our exome (see above), we choose a threshold of +/-0.8 (KIV-2 CNE
217 above or below 0.8 standard deviations from the mean) to match the proportion of individuals
218 with a measured Lp(a) above 120 nmol/L. High risk is defined as either a GRS > 120 nmol/L or
219 a KIV-2 CNE below 0.8 standard deviations from the mean. Low risk is defined as either a GRS
220 <10nmol/L or a KIV-2 CNE above 0.8 standard deviations from the mean. All other genotype
221 combinations are considered average risk.

222
223 To extrapolate to the general population, a slightly different approach is needed, as the
224 distribution of Lp(a) measures taken due to clinical care is ascertained and therefore lopsided
225 upward relative to the expected range. Nonetheless, we continue to focus on the extremes of
226 the genetic distributions on either end. We continue using the GRS thresholds of >120 nmol/L
227 and <10 nmol/L for high and low risk and match a KIV-2 CNE cutoff to reflect this proportion in
228 the general population, resulting in a cutoff of +/- 1.5 SD, somewhat more conservative than that
229 used in analyses of the subcohort with Lp(a) measures. This ensures individuals at high and low
230 ends of genetic risk of high Lp(a) are included at a roughly similar rate and that GRS and KIV-2
231 CNE are both contributing to the combined genotypes. These high and low combined genotypes
232 reflect our best understanding of clinical risk of ASCVD, but we also evaluated a threshold
233 corresponding to 150nmol/L, to represent a significant clinical endpoint (for example, see
234 (Yeang *et al.* 2022). The demographic distribution of the resulting risk groups calculated using
235 either a GRS cutoff of 120 nmol/L and 150nmol/L determining the KIV-2 CNE distribution
236 cutoffs can be found in Supplementary Table 3.

237 Results

238 Lp(a) is seldom measured in clinical practice

239

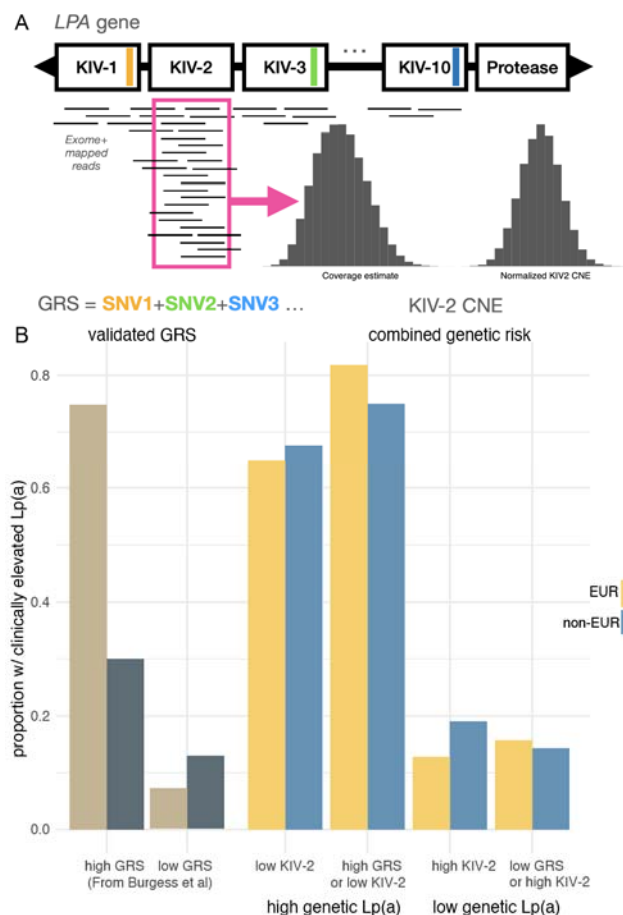
240 Within an all-comers cohort recruited from three health systems in the Helix Research Network
241 (HRN, n=76,147), we identify only 1584 (2.0%) individuals with 1777 clinically ascertained Lp(a)
242 measurements (for a full breakdown of cohort demographics, see Supplementary Table 1). In
243 contrast, 99% of individuals across this cohort have an LDL measurement. As these Lp(a)
244 measurements were taken during routine care in response to a clinical observation, we expect
245 them to be non-representative of the total population (Varvel *et al.* 2016; Dudum *et al.* 2024).
246 Indeed, the mean Lp(a) measured in the cohort is 50 nmol/L, slightly below the upper threshold
247 for normal (<75 nmol/L) and higher than the expected population mean (28-45 nmol/L)
248 (Langsted *et al.* 2014; Patel *et al.* 2021). This cohort also has a slightly elevated LDL when
249 measured (mean 101.4 mg/dL; for full comparison see Supplementary Figure 4).

250 SNV and KIV-2-based genetic predictor performance in those with Lp(a) levels

251
252 To facilitate comparisons with previous studies, and to report results in populations previously
253 underrepresented cohorts in genetics research, we divided our cohort into European (EUR) and
254 non-European (non-EUR) groups (see Methods). As expected, a previously-validated genetic
255 risk score (GRS), which outputs a genetically predicted Lp(a) level based on SNVs in *LPA*, is
256 associated with Lp(a) levels in EUR individuals in our cohort (n=1346) with an r^2 of 0.4 (see
257 Supplemental Figure 1). However, the r^2 in non-EUR individuals (n=238) is 0.04, indicating little
258 predictive value of the GRS for Lp(a) levels in non-EURs. In parallel, KIV-2 copy number
259 estimation (CNE) is anticorrelated with Lp(a) levels in both subsets, with an r^2 of 0.32 and 0.29
260 for EUR and non-EUR cohorts, respectively (see Supplemental Figure 2). This negative
261 correlation is expected as a lower KIV-2 CNE implies a smaller total molecule size, which eases
262 export of the molecule, increasing measured blood levels of Lp(a) (Reyes-Soffer *et al.* 2022).
263 Accordingly, a linear model combining both genetic predictors to genetically estimate Lp(a)
264 improves performance, with an r^2 of 0.49 for EUR individuals ($p < 2.2e-16$ for both variables) and
265 0.34 for non-EURs ($p < 2.2e-16$ for KIV-2 CNE; $p = 5.5e-6$ for GRS) in the dataset; this model
266 performs better than either alone.

267
268 As a performance metric, the correlation coefficient measures the ability to predict the specific
269 Lp(a) values in a patient population. However, the key clinical outcome for population-wide
270 genetic screening is whether an individual is expected to have an Lp(a) level high enough to be
271 associated with excess risk of future cardiovascular morbidity (Grundy *et al.* 2019; Kronenberg
272 *et al.* 2022). Accordingly, we calculate an additional performance metric to determine if KIV-2
273 CNE and GRS genotype groups can be established for predicting clinically relevant Lp(a)
274 elevations (>120 nmol/L). We identify individuals with a high GRS or a low KIV-2 CNE by
275 extrapolating from the distribution for individuals with a clinically elevated Lp(a) (see Methods).
276 In the subset of individuals with low KIV-2 CNE (approximately the lower 20% of the
277 distribution), 70% of these individuals have a clinically elevated Lp(a), compared to only 17% of
278 people without this genetic risk. Likewise for individuals with high GRS (above 120 nmol/L
279 predicted), 80% have a clinically elevated Lp(a), compared to only 16% without (Figure 1, Chi
280 squared p -value $< 2.2e-16$). The GRS performance differs significantly by population; 88% of
281 EUR individuals with a high GRS have an elevated Lp(a), compared to only 43% of non-EUR
282 individuals. When unifying across the presence of either genetic risk, we find 89% of individuals

283 with either a high GRS or a low KIV-2 CNE have a clinically elevated Lp(a), compared to only
284 9.8% of individuals with neither genetic risk predictor. This combined model performed similarly
285 well for individuals of both the EUR and non-EUR cohorts, in stark contrast to the ancestry-
286 biased performance of the GRS.
287



288
289 **Figure 1:** GRS and KIV-2 each contribute independently to predicting Lp(a) levels.

290 A) Schematic representation of bioinformatic process for KIV-2 copy number estimate (CNE) and calculation of the
291 *LPA* GRS score. B) Proportion of individuals in each category with an Lp(a) > 120nmol/L, divided by membership in
292 either EUR ('European') or non-EUR ('Non Eur') cohort. From left to right, 'high GRS' is a GRS > 120 nmol/L (as the
293 GRS is in units of predicted Lp(a)); 'low KIV-2' denotes a normalized KIV-2 less than 0.8 standard deviations from the
294 mean; and 'high GRS or low KIV-2' indicates that a patient was included for either reason. Although some individuals
295 with high GRS may have a non-risk KIV-2, and vice versa, 'low PRS or high KIV-2' are individuals with neither genetic
296 risk factor.

297

298 Combined genetic risk for predicting clinically-relevant Lp(a) levels

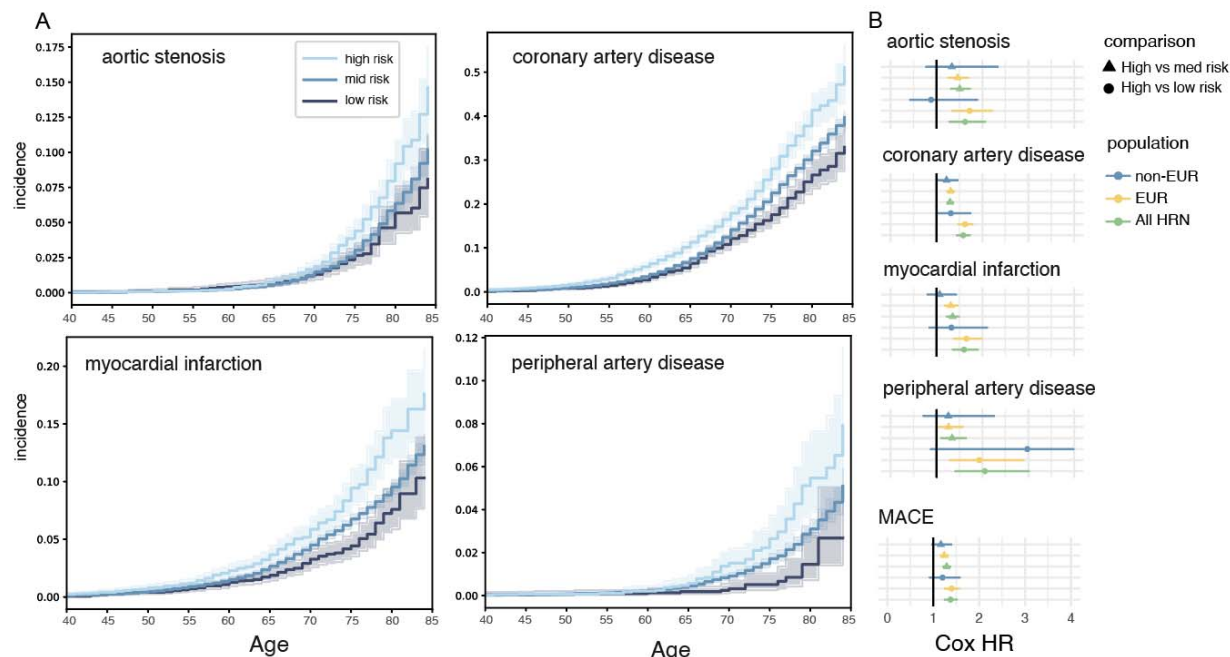
299 Guided by this data, we developed a population-level, ancestry-agnostic genotyping approach
300 for categorizing genetic Lp(a) risk leveraging both our KIV-2 CNE and the *LPA* GRS.
301 Specifically, we assign individuals to have a combined genetic risk for a high Lp(a) ("high
302 combined genetic risk") if they have either a GRS >120 nmol/L or a KIV-2 CNE below 1.5
303 standard deviations from the mean (n=14,355, or 18% of the participants). The remaining

304 individuals are considered to have a low combined genetic risk for a high Lp(a) (“low combined
305 genetic risk”) if they have a GRS < 10 nmol/L or a KIV-2 CNE above 1.5 standard deviations
306 from the mean (n=8,659). The remaining individuals harbor genetic profiles indicative of
307 average Lp(a) levels (n=53,133). Notably, the average LDL for individuals with recorded lab
308 values is consistent between these genetic groups, ranging between 104-107 (Supplementary
309 Figure 4).

310 Combined genetic risk, predicting clinically-elevated Lp(a), is associated with increased
311 ASCVD diagnoses in both EUR and non-EUR individuals

312 Next, we investigate association of our combined genetic risk groups with cardiovascular
313 disease diagnoses associated with clinically-elevated Lp(a) including coronary artery disease
314 (CAD), peripheral artery disease (PAD), myocardial infarction (MI), aortic stenosis (AS),
315 ischemic stroke (IS), as well as the five-point Major Adverse Cardiovascular Event (MACE)
316 index. For each phenotype, we compare the rate and age of onset of these diagnoses between
317 the high and average and high and low combined genetic risk groups using a Cox regression
318 controlling for sex, age, population group, and BMI to calculate a proportional hazard ratio (HR)
319 for each comparison. We find significant proportional hazards for CAD, PAD, MI, AS and MACE,
320 with an increase in the magnitude of these differences when comparing the high combined
321 genetic risk group to the average combined genetic risk group (e.g. CAD HR=1.29 [1.22-1.38],
322 $p=2.01e-16$; PAD HR=1.34 [1.08-1.66], $p=0.0015$) and low combined genetic risk group (CAD
323 HR= 1.58[1.43-1.75], $p=4.89e-18$; PAD HR= 2.05 [1.39-3.03], $p=6.8e-5$) (See Figure 2 and
324 Appendix 2 for all pairwise combinations). IS was only moderately enriched in the high
325 combined genetic risk group compared to the low combined genetic risk group (HR = 1.19[1.02-
326 1.41], $p=0.03$).

327
328 Next, we re-test for association in EUR (n=61,406) and non-EUR (n=14,741) cohorts
329 separately. As expected, given sample size enrichment (81% of cohort) and established
330 performance of GRS against similar outcomes, all comparisons remain significant in the EUR
331 subset(Figure 3B, Appendix 2). Despite the substantially smaller sample sizes of the high
332 (n=3,349) and low (n=1,798) combined genetic risk groups for non-EUR cohort, we observe
333 similar stepwise increases in hazard ratio confidence intervals for the suite of phenotypes with
334 significant HR differences in the cohort at large (Figure 3B, Appendix 2). In particular,
335 differences in diagnoses trends for CAD between the high and average combined genetic risk
336 groups and high and low combined genetic risk groups trend towards significance
337 (HR=1.22[1.0-1.48], $p=0.04$; HR=1.31[0.98-1.76], $p=0.06$). The exception was AS in the high vs.
338 the low combined genetic risk group for non-EUR, where low power was an issue (36% power
339 to detect the effect size seen in the EUR group in the non-EUR group). Nonetheless, without the
340 use of a clinical Lp(a) measurement, these genetically predicted Lp(a) risk groups identify
341 individuals with an increased incidence of future cardiovascular morbidity.

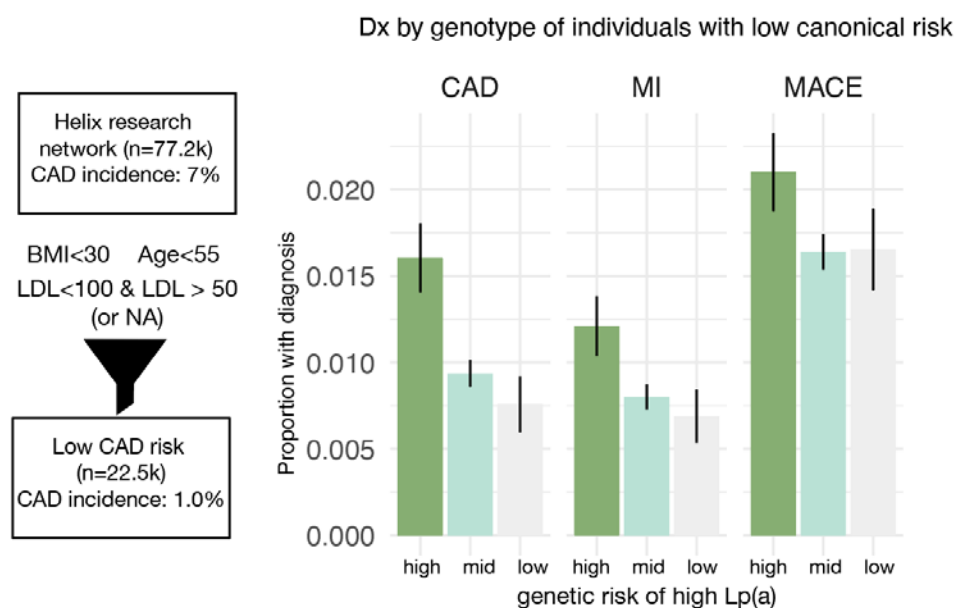


342
343 **Figure 2.** Elevated combined genetic risk of a high Lp(a) is associated with increased phenotype incidence after
344 adjusting for age, sex, BMI and population of origin. A) the cumulative fraction of individuals with coronary artery
345 disease, divided by their Lp(a) combined genetic risk group (high, average, or low). B) The Cox proportional hazard
346 ratio and 95% confidence intervals for high versus average or high versus low combined genetic risk of high Lp(a)
347 groups across statistical significant phenotypes in the unified cohort, as well as these estimations in subcohorts
348 separated by genetic similarity (EUR or non-EUR cohorts).
349

350 Predicted Lp(a) and incidence of ASCVD in individuals with otherwise low risk

351 In addition to increasing ASCVD risk independently of other risk factors like LDL, Lp(a) is also
352 suspected to contribute to the incidence of premature or otherwise unexpected ASCVD
353 (Verbeek *et al.* 2018). To measure the latent contribution of a high combined genetic risk of high
354 Lp(a) to unexpected ASCVD, we subsetted our cohort to individuals with a lower risk of
355 cardiovascular disease using key risk factors available from EHR records. We included
356 individuals with either no LDL measurement or a low clinically-ascertained measurement (<100
357 mg/dL, but >50mg/dL, to exclude people with other possible lipid disorders), low age (<55), and
358 most recent BMI < 30 (Zhou *et al.* 2021; Wong *et al.* 2022). As expected, this subset has a
359 substantially lowered incidence of CAD (Figure 3). Next, we compared the percentage of
360 individuals in each combined genetic risk category with an early ASCVD diagnosis (age <55).
361 The mean age in each risk category in this analysis was similar (41.4, 41.5, and 41.4 years
362 across high, medium, and low risk groups; see Supplemental Figure 5) so comparing incidence
363 across groups is not biased by age of onset or followup time. We find that 1.2-1.6% of
364 individuals with a high combined genetic risk of a high Lp(a) in this low traditional risk cohort
365 harbor a diagnosis of CAD or MI, compared to 0.6-0.7% of low traditional risk individuals with a
366 low combined genetic risk of high Lp(a) (Figure 3).
367

368 Another way to summarize this contribution to excess risk is to calculate the proportion of
369 individuals in high combined genetic risk of high Lp(a) receiving premature diagnoses compared
370 to the overall representation of the group in the population. Using CAD as an example, we find
371 that while individuals with a high combined genetic risk of high Lp(a) are 17% of the low risk
372 population, they comprise 27% of low canonical risk individuals who are nonetheless diagnosed
373 with CAD (binomial test, p -value = 0.014). Across ASCVD diagnoses in this low risk population,
374 there is 2-fold enrichment in those with high combined genetic risk of high Lp(a) compared to
375 those with low combined genetic risk of Lp(a) (Figure 3). While these enrichments may not
376 account for the majority of cardiovascular morbidity, elevated Lp(a) represents an unobserved
377 determinant of risk, and may explain nearly a third of individuals diagnosed with disease who
378 lack other canonical risk factors.
379



380
381 **Figure 3:** Proportion of individuals at low canonical risk with disease diagnoses. The proportion of individuals in each
382 Lp(a) genetic risk group receiving a diagnosis. These individuals are all restricted to age <55, BMI < 30, and either no
383 LDL measurement or a low measurement, to include only individuals with no canonical risk factors for cardiovascular
384 disease.

385 Discussion

386 Genetic prediction of Lp(a) levels presents significant challenges as well as significant potential
387 for precision medicine and prevention of future risk of Atherosclerotic Cardiovascular Disease
388 (ASCVD). In this work, we develop a practical method to estimate the number of copies of KIV-2
389 (KIV-2 CNE) from exome sequencing in the Helix Research Network, an all-comers cohort
390 collected from multiple health system sites across the United States. While genetic risk scores
391 (GRS) applied in the clinic are known to perform poorly in understudied and underrepresented
392 genetic groups, KIV-2 CNE is a population-agnostic genetic risk predictor (Privé *et al.* 2022;
393 Kullo *et al.* 2022). KIV-2 CNE performed well to predict Lp(a) in both the European (EUR) and
394 non-European (non-EUR) subsets of our cohort with Lp(a) measurements (Figure 1).
395 Independent of genetic similarity group, individuals with a low KIV-2 CNE are 5-fold more likely

396 to present with elevated Lp(a) levels (80% vs 16%, Chi squared p-value < 2.2e-16). This
397 demonstrates the opportunity to develop a scalable genetic risk predictor and with equitable
398 performance, which we test across a diverse health system cohort.

399
400 We use KIV-2 CNE in combination with a GRS to develop a combined genetic risk predictor of
401 high Lp(a) and genotype a cohort of 76,417 individuals for this genetic risk. This exome-based
402 categorization of combined genetic risk of high Lp(a) is associated with increased incidence and
403 earlier onset of coronary artery disease, myocardial infarction, heart failure, and other
404 phenotypes associated with ASCVD with strong statistical significance in both the cohort at
405 large, the EUR subset, and with trending significance and similar direction of effects in the non-
406 EUR subcohort (Figure 2). Encouragingly, the hazard ratio estimate for our combined high
407 genetic risk group compared to the low risk group for coronary artery disease (HR=1.6, whole
408 cohort; p-value = 2e-19) is similar to the established hazard ratios from association studies
409 based on actual measures of clinically-elevated Lp(a), confirming that this genetic estimate is a
410 useful clinical instrument for predicting elevated Lp(a) levels (Finneran *et al.* 2021).

411
412 Lp(a) has been shown to predict cardiovascular morbidity risk independently of leading risk
413 factors, such as LDL, age, and BMI, as well as in individuals who develop ASCVD without any
414 of these canonical risk factors. Amongst patients at low canonical risk in our cohort, there is a 2-
415 fold enrichment of individuals with a high combined genetic risk of high Lp(a) compared to low
416 combined genetic risk individuals, and this high risk group accounts for 27% of unexpected
417 cases of coronary artery disease (Figure 3). This suggests that our genetic measure may help
418 identify additional individuals with elevated ASCVD risk not identified based on standard risk
419 scoring measures.

420
421 This study does not exhaustively look at all possible genetic variation in Lp(a) with an impact on
422 Lp(a) levels; recent work has shown that SNVs in the KIV-2 repeat itself are impactful, although
423 likely less impactful on Lp(a) than population-specific variation in the abundance of KIV-2 (Di
424 Maio *et al.* 2020; Schachtl-Riess *et al.* 2021; Behera *et al.* 2023). Future work incorporating
425 SNVs, particularly in the KIV-2 region, as well as tailoring the distribution thresholds for high and
426 low KIV-2 copy number estimates may improve the precision of these genetic estimations. One
427 of the major limitations of studies that rely on retrospective analyses of EHR is the lack of
428 detailed phenotyping over a longitudinal timespan and inherent selection bias in building a
429 dataset from participants actively engaging with the healthcare system. Although EHRs are very
430 comprehensive, many data points are not recorded with the desired frequency or detail, and
431 coding patterns vary by both physician and health system and will be biased by the overall
432 engagement of each participant with the health system, making it challenging to
433 comprehensively apply known ASCVD risk estimates (D'Agostino *et al.* 2008). Despite these
434 limitations, using EHR data for phenotyping provides an opportunity to evaluate genetic
435 associations with health outcomes at a population level, creating evidence to support more
436 detailed clinical follow up projects. For example, based on this retrospective analysis, we
437 believe a prospective analysis validating this combined genetic estimation of high Lp(a) against
438 clinically-measured Lp(a) levels more broadly across populations is warranted based on these
439 preliminary findings.

440
441 Therapy options for patients with elevated Lp(a) are limited. Current available lipid lowering
442 drugs do not provide targeted Lp(a) lowering. PCSK9 monoclonal antibody inhibitors modestly
443 decrease Lp(a) levels(Burgess *et al.* 2018; O'Donoghue *et al.* 2019; Lamina *et al.* 2019; Bittner
444 *et al.* 2020). In patients with clinical ASCVD and recurrent events, if Lp(a) is elevated, apheresis
445 can be considered; however, this is a cumbersome and undesirable option for most patients
446 (Leebmann *et al.* 2013; Pokrovsky *et al.* 2016). Multiple novel nucleic acid therapies that lower
447 Lp(a) are currently being studied in phase III clinical outcomes trials (Tsimikas *et al.* 2020;
448 O'Donoghue *et al.* 2022). Pelacarsen, an antisense oligonucleotide, and olezarsen, a small
449 interfering RNA (siRNA), have been found to safely reduce Lp(a) by 40-80% depending on the
450 dose and frequency of injections. The HORIZON and OCEAN(a) trials will determine if these
451 medications reduce cardiovascular events in patients with a history of clinical ASCVD; these
452 results are expected as early as 2025. Additionally, the ACCLAIM-Lp(a) trial will investigate
453 lepodisiran, another siRNA that reduces Lp(a), in a phase III clinical outcomes trial for both
454 primary and secondary prevention patients. Currently, Lp(a) measurement is used mostly for
455 ASCVD risk stratification, but if these pharmaceutical options prove to reduce cardiovascular
456 events, there will be an even greater interest in identifying patients with elevated Lp(a) for
457 prevention.

458
459 Lp(a) is rarely measured as part of clinical practice: in our clinical cohort, only 1.4% of patients
460 had a documented Lp(a) measurement taken. In contrast, every one of them has been
461 sequenced. As sequencing becomes an established part of routine care, the impact of genetic
462 data on treatment and preventative medicine will be determined by the efficacy and fairness of
463 genomic predictors of future morbidity. Accordingly, we develop a predictor which performs
464 similarly to the value of the actual lab measurement in patients where it has been ascertained,
465 and which is highly sensitive regardless of ancestry and population of origin. This cross-
466 population performance is crucial to the fair and equitable development of genetic medicine, and
467 is rarely achieved in the development of polygenic predictors (Wang *et al.* 2020; Privé *et al.*
468 2022; Kullo *et al.* 2022; Aw *et al.* 2024). This predictor enables us to identify patients at
469 significantly higher risk of future cardiovascular disease, which can both compound traditional
470 risk as well as identify at risk individuals who present without canonical risk factors. In short, this
471 genomic-based predictor for high Lp(a), which may be calculated for any patient regardless of
472 current diagnoses, family history, or other measured risk factors, is both a fair future predictor of
473 cardiovascular morbidity and demonstrates that 27% of unexplained coronary artery disease
474 may be attributed to a combined genetic risk of elevated Lp(a). This work advances the
475 evidence base supporting Lp(a) estimation as a potential equitable baseline health measure
476 from genetic data, especially as exome sequencing for disease risk estimation and prevention
477 becomes more common in medicine.

478

479 **Declaration of Interests**

480 KMSB, ETC, AB, NT, WL, SL, BK, HD, and NLW are all employees of Helix.

481

482 **Acknowledgments**

483 We acknowledge the entire Helix bioinformatics and lab teams for their contributions to the
484 production of exome sequencing pipeline as well as the research administration team for
485 coordinating the project. We thank all of the participants of the Helix Research Network.

486

487 **Funding**

488 Funding was provided to the Desert Research Institute by the Renown Institute for Health
489 Innovation and the Renown Health Foundation. Funding was provided to DRI by the Nevada
490 Governor's Office of Economic Development. Funding was provided to the myGenetics program
491 by HealthPartners.

492

493 **Ethics Declaration**

494 The Helix cohorts were reviewed by Salus IRB (Reliance on Salus for all sites) and approved
495 (approval number 21143), the WCG IRB (Western Institutional Review Board, WIRB-
496 Copernicus Group) and approved (approval number 20224919), the MUSC Institutional Review
497 Board for Human Research and approved (approval number Pro00129083), and the University
498 of Nevada, Reno Institutional Review Board and approved (approval number 7701703417). All
499 participants gave their informed, written consent before participation. All data used for research
500 were de-identified.

501

502 **References**

503

504 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, *et al.*,
505 2015 A global reference for human genetic variation. *Nature* 526: 68–74.

506 Alexander D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in
507 unrelated individuals. *Genome Res.* 19: 1655–1664.

508 Aw A. J., J. McRae, E. Rahmani, and Y. S. Song, 2024 Highly parameterized polygenic scores
509 tend to overfit to population stratification via random effects. *bioRxiv*.

510 <https://doi.org/10.1101/2024.01.27.577589>

- 511 Behera S., J. R. Belyeu, X. Chen, L. F. Paulin, N. Q. H. Nguyen, *et al.*, 2023 Identification of
512 allele-specific KIV-2 repeats and impact on Lp(a) measurements for cardiovascular disease
513 risk. *bioRxiv.org*. <https://doi.org/10.1101/2023.04.24.538128>
- 514 Berman A. N., D. W. Biery, S. A. Besser, A. Singh, A. Shiyovich, *et al.*, 2024 Lipoprotein(a) and
515 major adverse cardiovascular events in patients with or without baseline atherosclerotic
516 cardiovascular disease. *J. Am. Coll. Cardiol.* 83: 873–886.
- 517 Bittner V. A., M. Szarek, P. E. Aylward, D. L. Bhatt, R. Diaz, *et al.*, 2020 Effect of alirocumab on
518 lipoprotein(a) and cardiovascular risk after acute coronary syndrome. *J. Am. Coll. Cardiol.*
519 75: 133–144.
- 520 Burgess S., B. A. Ference, J. R. Staley, D. F. Freitag, A. M. Mason, *et al.*, 2018 Association of
521 LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-
522 Lowering Therapies: A Mendelian Randomization Analysis. *JAMA Cardiol* 3: 619–627.
- 523 Cegla J., R. D. G. Neely, M. France, G. Ferns, C. D. Byrne, *et al.*, 2019 HEART UK consensus
524 statement on Lipoprotein(a): A call to action. *Atherosclerosis* 291: 62–70.
- 525 Cirulli E. T., S. White, R. W. Read, G. Elhanan, W. J. Metcalf, *et al.*, 2020 Genome-wide rare
526 variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat.*
527 *Commun.* 11: 542.
- 528 Clarke R., J. F. Peden, J. C. Hopewell, T. Kyriakou, A. Goel, *et al.*, 2009 Genetic variants
529 associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* 361: 2518–
530 2528.
- 531 Coassin S., and F. Kronenberg, 2022 Lipoprotein(a) beyond the kringle IV repeat polymorphism:
532 The complexity of genetic variation in the LPA gene. *Atherosclerosis* 349: 17–35.

- 533 D'Agostino R. B. Sr, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, *et al.*, 2008 General
534 cardiovascular risk profile for use in primary care: the Framingham Heart Study: The
535 Framingham heart study. *Circulation* 117: 743–753.
- 536 Di Maio S., R. Grüneis, G. Streiter, C. Lamina, M. Maglione, *et al.*, 2020 Investigation of a
537 nonsense mutation located in the complex KIV-2 copy number variation region of
538 apolipoprotein(a) in 10,910 individuals. *Genome Med.* 12: 74.
- 539 Dudum R., Q. Huang, X. S. Yan, M. A. Fonseca, P. Jose, *et al.*, 2024 Lipoprotein(a) levels in
540 disaggregated racial and ethnic subgroups across atherosclerotic cardiovascular disease
541 risk levels. *JACC Adv.* 3: 100940.
- 542 Emerging Risk Factors Collaboration, S. Erqou, S. Kaptoge, P. L. Perry, E. Di Angelantonio, *et*
543 *al.*, 2009 Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and
544 nonvascular mortality. *JAMA* 302: 412–423.
- 545 Erqou S., A. Thompson, E. Di Angelantonio, D. Saleheen, S. Kaptoge, *et al.*, 2010
546 Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies
547 involving 58,000 participants. *J. Am. Coll. Cardiol.* 55: 2160–2167.
- 548 Finneran P., A. Pampana, S. A. Khetarpal, M. Trinder, A. P. Patel, *et al.*, 2021 Lipoprotein(a)
549 and Coronary Artery Disease Risk Without a Family History of Heart Disease. *J Am Heart*
550 *Assoc* 10: e017470.
- 551 Grundy S. M., N. J. Stone, A. L. Bailey, C. Beam, K. K. Birtcher, *et al.*, 2019 2018
552 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the
553 management of blood cholesterol: Executive summary: A report of the American college of
554 cardiology/American heart association task force on clinical practice guidelines. *J. Am. Coll.*
555 *Cardiol.* 73: 3168–3209.

- 556 Kamstrup P. R., A. Tybjaerg-Hansen, R. Steffensen, and B. G. Nordestgaard, 2009 Genetically
557 elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* 301: 2331–2339.
- 558 Khera A. V., B. M. Everett, M. P. Caulfield, F. M. Hantash, J. Wohlgemuth, *et al.*, 2014
559 Lipoprotein(a) concentrations, rosuvastatin therapy, and residual vascular risk: an analysis
560 from the JUPITER Trial (Justification for the Use of Statins in Prevention: an Intervention
561 Trial Evaluating Rosuvastatin). *Circulation* 129: 635–642.
- 562 Klarin D., J. Lynch, K. Aragam, M. Chaffin, T. L. Assimes, *et al.*, 2019 Genome-wide association
563 study of peripheral artery disease in the Million Veteran Program. *Nat. Med.* 25: 1274–
564 1279.
- 565 Koschinsky M. L., A. Bajaj, M. B. Boffa, D. L. Dixon, K. C. Ferdinand, *et al.*, 2024 A focused
566 update to the 2019 NLA scientific statement on use of lipoprotein(a) in clinical practice. *J.*
567 *Clin. Lipidol.* 18: e308–e319.
- 568 Kraft H. G., A. Lingenhel, R. W. Pang, R. Delport, M. Trommsdorff, *et al.*, 1996 Frequency
569 distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a)
570 levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-
571 random. *Eur. J. Hum. Genet.* 4: 74–87.
- 572 Kronenberg F., S. Mora, E. S. G. Stroes, B. A. Ference, B. J. Arsenault, *et al.*, 2022
573 Lipoprotein(a) in atherosclerotic cardiovascular disease and aortic stenosis: a European
574 Atherosclerosis Society consensus statement. *Eur. Heart J.* 43: 3925–3946.
- 575 Kullo I. J., C. M. Lewis, M. Inouye, A. R. Martin, S. Ripatti, *et al.*, 2022 Polygenic scores in
576 biomedical research. *Nat. Rev. Genet.* 23: 524–532.
- 577 Lamina C., F. Kronenberg, and Lp(a)-GWAS-Consortium, 2019 Estimation of the required
578 lipoprotein(a)-lowering therapeutic effect size for reduction in coronary heart disease

- 579 outcomes: A Mendelian randomization analysis: A Mendelian randomization analysis.
580 JAMA Cardiol. 4: 575–579.
- 581 Langsted A., P. R. Kamstrup, and B. G. Nordestgaard, 2014 Lipoprotein(a): fasting and
582 nonfasting levels, inflammation, and cardiovascular risk. *Atherosclerosis* 234: 95–101.
- 583 Leebmann J., E. Roeseler, U. Julius, F. Heigl, R. Spitthoever, *et al.*, 2013 Lipoprotein apheresis
584 in patients with maximally tolerated lipid-lowering therapy, lipoprotein(a)-
585 hyperlipoproteinemia, and progressive cardiovascular disease: prospective observational
586 multicenter study: Prospective observational multicenter study. *Circulation* 128: 2567–2576.
- 587 Mack S., S. Coassin, R. Rueedi, N. A. Yousri, I. Seppälä, *et al.*, 2017 A genome-wide
588 association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a)
589 isoforms. *J. Lipid Res.* 58: 1834–1844.
- 590 Marcovina S. M., and M. L. Koschinsky, 1998 Lipoprotein(a) as a risk factor for coronary artery
591 disease. *Am. J. Cardiol.* 82: 57U–66U; discussion 86U.
- 592 Mehta A., V. Jain, A. Saeed, J. J. Saseen, M. Gulati, *et al.*, 2022 Lipoprotein(a) and ethnicities.
593 *Atherosclerosis* 349: 42–52.
- 594 Nordestgaard B. G., M. J. Chapman, K. Ray, J. Borén, F. Andreotti, *et al.*, 2010 Lipoprotein(a)
595 as a cardiovascular risk factor: current status. *Eur. Heart J.* 31: 2844–2853.
- 596 O’Donoghue M. L., S. Fazio, R. P. Giugliano, E. S. G. Stroes, E. Kanevsky, *et al.*, 2019
597 Lipoprotein(a), PCSK9 inhibition, and cardiovascular risk. *Circulation* 139: 1483–1492.
- 598 O’Donoghue M. L., J. A. G López, B. Knusel, B. Gencer, H. Wang, *et al.*, 2022 Study design
599 and rationale for the Olpasiran trials of Cardiovascular Events And lipoprotein(a) reduction-
600 DOSE finding study (OCEAN(a)-DOSE). *Am. Heart J.* 251: 61–69.

- 601 Patel A. P., M. Wang (汪敏先), J. P. Pirruccello, P. T. Ellinor, K. Ng, *et al.*, 2021 Lp(a)
602 (Lipoprotein[a]) Concentrations and Incident Atherosclerotic Cardiovascular Disease: New
603 Insights From a Large National Biobank. *Arterioscler. Thromb. Vasc. Biol.* 41: 465–474.
- 604 Pokrovsky S. N., O. I. Afanasieva, and M. V. Ezhov, 2016 Lipoprotein(a) apheresis. *Curr. Opin.*
605 *Lipidol.* 27: 351–358.
- 606 Privé F., H. Aschard, S. Carmi, L. Folkersen, C. Hoggart, *et al.*, 2022 Portability of 245
607 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from
608 the same cohort. *Am. J. Hum. Genet.* 109: 12–23.
- 609 Reyes-Soffer G., H. N. Ginsberg, L. Berglund, P. B. Duell, S. P. Heffron, *et al.*, 2022
610 Lipoprotein(a): A Genetically Determined, Causal, and Prevalent Risk Factor for
611 Atherosclerotic Cardiovascular Disease: A Scientific Statement From the American Heart
612 Association. *Arterioscler. Thromb. Vasc. Biol.* 42: e48–e60.
- 613 Schachtl-Riess J. F., A. Kheirkhah, R. Grüneis, S. Di Maio, S. Schoenherr, *et al.*, 2021 Frequent
614 LPA KIV-2 Variants Lower Lipoprotein(a) Concentrations and Protect Against Coronary
615 Artery Disease. *J. Am. Coll. Cardiol.* 78: 437–449.
- 616 Schmidt K., A. Noureen, F. Kronenberg, and G. Utermann, 2016 Structure, function, and
617 genetics of lipoprotein (a). *J. Lipid Res.* 57: 1339–1359.
- 618 Shuey M. M., W. W. Stead, I. Aka, A. L. Barnado, J. A. Bastarache, *et al.*, 2023 Next-generation
619 phenotyping: Introducing PhecodeX for enhanced discovery research in medical
620 phenomics. medRxiv.
- 621 Trinder M., M. M. Uddin, P. Finneran, K. G. Aragam, and P. Natarajan, 2020 Clinical Utility of
622 Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic

- 623 Cardiovascular Disease. *JAMA Cardiol* 6: 1–9.
- 624 Trinder M., K. Paruchuri, S. Haidermota, R. Bernardo, S. M. Zekavat, *et al.*, 2022 Repeat
625 measures of lipoprotein(a) molar concentration and cardiovascular risk. *J. Am. Coll.*
626 *Cardiol.* 79: 617–628.
- 627 Tsimikas S., 2017 A test in context: Lipoprotein(a): Diagnosis, prognosis, controversies, and
628 emerging therapies. *J. Am. Coll. Cardiol.* 69: 692–711.
- 629 Tsimikas S., E. Karwatowska-Prokopczuk, I. Gouni-Berthold, J.-C. Tardif, S. J. Baum, *et al.*,
630 2020 Lipoprotein(a) reduction in persons with cardiovascular disease. *N. Engl. J. Med.* 382:
631 244–255.
- 632 Tsimikas S., and E. S. G. Stroes, 2020 The dedicated “Lp(a) clinic”: A concept whose time has
633 arrived? *Atherosclerosis* 300: 1–9.
- 634 Tsimikas S., and S. M. Marcovina, 2022 Ancestry, lipoprotein(a), and cardiovascular risk
635 thresholds: JACC review topic of the week. *J. Am. Coll. Cardiol.* 80: 934–946.
- 636 Utermann G., H. J. Menzel, H. G. Kraft, H. C. Duba, H. G. Kemmler, *et al.*, 1987 Lp(a)
637 glycoprotein phenotypes. Inheritance and relation to Lp(a)-lipoprotein concentrations in
638 plasma. *J. Clin. Invest.* 80: 458–465.
- 639 Varvel S., J. P. McConnell, and S. Tsimikas, 2016 Prevalence of elevated LP(a) mass levels
640 and patient thresholds in 532 359 patients in the United States. *Arterioscler. Thromb. Vasc.*
641 *Biol.* 36: 2239–2245.
- 642 Verbeek R., R. M. Hoogeveen, A. Langsted, L. C. A. Stiekema, S. L. Verweij, *et al.*, 2018
643 Cardiovascular disease risk associated with elevated lipoprotein(a) attenuates at low low-
644 density lipoprotein cholesterol levels in a primary prevention setting. *Eur. Heart J.* 39:

645 2589–2596.

646 Wang Y., J. Guo, G. Ni, J. Yang, P. M. Visscher, *et al.*, 2020 Theoretical and empirical
647 quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat.*
648 *Commun.* 11: 3865.

649 Wilson D. P., T. A. Jacobson, P. H. Jones, M. L. Koschinsky, C. J. McNeal, *et al.*, 2019 Use of
650 Lipoprotein(a) in clinical practice: A biomarker whose time has come. A scientific statement
651 from the National Lipid Association. *J. Clin. Lipidol.* 13: 374–392.

652 Wong N. D., M. J. Budoff, K. Ferdinand, I. M. Graham, E. D. Michos, *et al.*, 2022 Atherosclerotic
653 cardiovascular disease risk assessment: An American Society for Preventive Cardiology
654 clinical practice statement. *Am. J. Prev. Cardiol.* 10: 100335.

655 Wong N. D., W. Fan, X. Hu, C. Ballantyne, R. C. Hodgeveeen, *et al.*, 2024 Lipoprotein(a) and
656 long-term cardiovascular risk in a multi-ethnic pooled prospective cohort. *J. Am. Coll.*
657 *Cardiol.* 83: 1511–1525.

658 Yeang C., E. Karwatowska-Prokopczuk, F. Su, B. Dinh, S. Xia, *et al.*, 2022 Effect of pelacarsen
659 on lipoprotein(a) cholesterol and corrected low-density lipoprotein cholesterol. *J. Am. Coll.*
660 *Cardiol.* 79: 1035–1046.

661 Zekavat S. M., S. Ruotsalainen, R. E. Handsaker, M. Alver, J. Bloom, *et al.*, 2018 Deep
662 coverage whole genome sequences and plasma lipoprotein(a) in individuals of European
663 and African ancestries. *Nat. Commun.* 9: 2606.

664 Zhou Z., J. Macpherson, S. R. Gray, J. M. R. Gill, P. Welsh, *et al.*, 2021 Are people with
665 metabolically healthy obesity really healthy? A prospective cohort study of 381,363 UK
666 Biobank participants. *Diabetologia* 64: 1963–1972.

