

1 **The penetrance of rare variants in cardiomyopathy-associated genes:**
2 **a cross-sectional approach to estimate penetrance for secondary findings**

3 Kathryn A. McGurk^{1,2}, Xiaolei Zhang^{1,15}, Pantazis Theotokis^{1,2}, Kate Thomson⁴, Andrew
4 Harper^{4,16}, Rachel J. Buchan^{1,2,3}, Erica Mazaika¹, Elizabeth Ormondroyd⁴, William T.
5 Wright⁵, Daniela Macaya⁶, Chee Jian Pua⁷, Birgit Funke⁸, Daniel G. MacArthur^{9,10},
6 Sanjay Prasad^{1,3}, Stuart A. Cook^{2,7}, Mona Allouba^{1,11}, Yasmine Aguib^{1,11}, Magdi H.
7 Yacoub^{1,3,11}, Declan P. O'Regan², Paul J. R. Barton^{1,2,3}, Hugh Watkins⁴, Leonardo
8 Bottolo^{12,13,14}, James S. Ware^{1,2,3,*}.

9
10 ¹National Heart and Lung Institute, Imperial College London, London, UK.

11 ²MRC London Institute of Medical Sciences, Imperial College London, London, UK.

12 ³Royal Brompton & Harefield Hospitals, Guy's and St. Thomas' NHS Foundation Trust,
13 London, UK.

14 ⁴Division of Cardiovascular Medicine, Radcliffe Department of Medicine and the
15 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

16 ⁵Northern Ireland Regional Genetics Centre, Belfast Health and Social Care Trust,
17 Belfast City Hospital, Belfast, Northern Ireland, UK.

18 ⁶GeneDx LLC, Gaithersburg, Maryland, US.

19 ⁷National Heart Research Institute Singapore and Duke-National University of
20 Singapore, Singapore.

21 ⁸Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized
22 Genetic Medicine, Boston, Massachusetts, USA.

23 ⁹Centre for Population Genomics, Garvan Institute of Medical Research and UNSW,
24 Sydney, NSW, Australia.

25 ¹⁰Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne,
26 VIC, Australia.

27 ¹¹Aswan Heart Centre, Aswan, Egypt.

28 ¹²Department of Medical Genetics, University of Cambridge, Cambridge, UK.

29 ¹³The Alan Turing Institute, London, UK.

30 ¹⁴MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.

31 ¹⁵Current address: European Bioinformatics Institute, Wellcome Genome Campus,
32 Hinxton, UK.

33 ¹⁶Current address: Centre for Genomics Research, Discovery Sciences,
34 BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK.

35

36 ***Corresponding author:** Professor James S. Ware, j.ware@imperial.ac.uk

37 **Word count:** 5,499

38 **Keywords:** Penetrance, Cardiomyopathy, Prevalence, Secondary findings

39 **Abstract**

40 Understanding the penetrance of pathogenic variants identified as secondary findings
41 (SFs) is of paramount importance with the growing availability of genetic testing. We
42 estimated penetrance through large-scale analyses of patients referred for diagnostic
43 sequencing for hypertrophic cardiomyopathy (HCM; 10,400 cases, 1,340 variants) and
44 dilated cardiomyopathy (DCM; 2,564 cases, 665 variants), using a cross-sectional
45 approach comparing allele frequencies against reference populations (293,226
46 participants from UK Biobank and gnomAD). We generated updated prevalence
47 estimates for HCM (1:543) and DCM (1:220).

48 In aggregate, the penetrance by late adulthood of rare, pathogenic variants (23% for
49 HCM, 35% for DCM) and likely pathogenic variants (7% for HCM, 10% for DCM) was
50 substantial for dominant CM. Penetrance was significantly higher for variant subgroups
51 annotated as loss of function or ultra-rare, and for males compared to females for
52 variants in HCM-associated genes.

53 We estimated variant-specific penetrance for 316 recurrent variants most likely to be
54 identified as SFs (51% HCM and 17% DCM cases). 49 variants were observed at least
55 ten times (14% of cases) in HCM-associated genes. Median penetrance was 14.6%
56 ($\pm 14.4\%$ SD). We explore estimates of penetrance by age, sex, and ancestry, and
57 simulate the impact of including future cohorts.

58 This dataset is the first to report penetrance of individual variants at scale and will
59 inform the management of individuals undergoing genetic screening for SFs. While
60 most variants had low penetrance and the costs and harms of screening are unclear,
61 some carriers of highly penetrant variants may benefit from SFs.

62

63 **Graphical Abstract**

64 A flowchart of the estimated penetrance for dominant cardiomyopathy by late adulthood
65 for a variant of interest. The estimates of penetrance in this study are for carriers
66 identified from unselected populations (e.g., consumer-initiated elective genomic testing
67 or as secondary (2°) findings in clinical settings). If the variant is ultra-rare (i.e.,
68 identified once or less in population datasets), only estimates by variant subgroup in
69 aggregate are available. If the variant is identified multiple times in both case and
70 population datasets, variant-specific penetrance estimates may be available. If the
71 variant is curated as a variant of uncertain significance (VUS), the penetrance estimate
72 is low. If the variant is a likely pathogenic predicted loss of function (pLoF) variant and is
73 identifiable multiple times in cases and population cohorts, penetrance estimates vary
74 by gene (**Figure 2**). High aggregate penetrance represents an estimate of $>25\%$;
75 moderate aggregate penetrance represents $10\%-25\%$, and low penetrance represents
76 $<10\%$. This flowchart was created with draw.io.

77 **Introduction**

78 Cardiomyopathies (CM) are diseases of the heart muscle, characterised by abnormal
79 cardiac structure and function that is not due to coronary disease, hypertension, valve
80 disease, or congenital heart disease. Many cases have a monogenic etiology with
81 autosomal dominant inheritance. Penetrance is incomplete and age-related, and
82 expressivity is highly variable. These features present huge challenges for disease
83 management. In particular, the penetrance of variants in CM-associated genes is
84 incompletely characterised and poorly understood, especially when identified in an
85 asymptomatic individual without family history of CM. With the growing availability of
86 whole exome sequencing in wider clinical settings and consumer-initiated elective
87 genomic testing¹, the importance of estimating the penetrance of individual variants
88 identified as secondary findings (SFs) to guide intervention is ever-increasing.

89 SFs are genetic variants that are actively sought out (as opposed to incidental findings),
90 but which are unrelated to the clinical indication for genetic testing and can therefore be
91 considered as opportunistic genetic screening. Genes associated with inherited CMs
92 make up one-fifth of the 78 genes recommended by the American College of Medical
93 Genetics and Genomics (ACMG SF v3.1) for reporting SFs during clinical sequencing².
94 It is recommended to return variants that would be classified as pathogenic or likely
95 pathogenic in an affected individual and where there is >90% confidence that the variant
96 is causing the observed disease³. This is independent of the probability that an
97 individual carrying the variant will develop disease (penetrance). The ACMG SF
98 guidelines have not yet been adopted globally, with the European Society of Human
99 Genetics recommending a cautious approach with limited capacity, but responsive to
100 accumulating evidence^{4,5}.

101 We are concerned that the costs, harms, and benefits, have not been fully
102 characterised. We have previously discussed issues with the recommendations based
103 on the lack of estimates of the harms and cost of this approach for variants in specific
104 genes⁶. These estimates are required to conform to the ninth rule of Wilson and
105 Jugner's principles of screening⁷. The burden of the implementation of reporting SFs in
106 specific healthcare systems remains unassessed. There is little evidence for clinical
107 utility and limited justification for use of resources⁵. Research is beginning to become
108 available on patient perspectives and impact⁸⁻¹².

109 Subclinical phenotypic expressivity of rare variants in CM-associated genes has been
110 demonstrated in the UK Biobank population cohort¹³⁻¹⁵. Causes of variability in
111 penetrance may include: i) genetic and allelic heterogeneity, with different alleles having
112 different consequences on protein function, ii) environmental modifiers altering genetic
113 influence (e.g., age, sex, hypertension, lifestyle), and iii) additional genetic modifiers
114 with additive or epistatic interactions with the variant of interest (other variants or
115 combinations of genetic factors, e.g. polygenic risk, variants in *cis* that drive allelic
116 imbalance, imprinting, epigenetic regulation, compensation, threshold model, and
117 transcript isoform expression)¹⁶⁻²².

118 Variant-specific estimates of penetrance are required to appropriately inform clinical
119 practice and to fully utilise genetics as a tool to individualise the risk of developing
120 disease in asymptomatic carriers^{6,23}. It is challenging to estimate the penetrance of
121 individual rare variants through other study methods as longitudinal population studies
122 require very large sample sizes and long-term follow-up is required if penetrance is age-
123 related. Where data is available for rare variants in CM-associated genes, reported
124 penetrance is mostly estimated from family-based studies. These may be affected by
125 ascertainment biases and secondary genetic and environmental factors²⁴ and thus less
126 applicable to SFs. Penetrance has been estimated in aggregate by gene and by
127 disease^{13,25,26}. Variant-specific penetrance in the general adult population for rare
128 variants in CM-associated genes is unknown.

129 Here, we apply a cross-sectional approach, using a method that compares the allele
130 frequency of individual rare variants in large cohorts of cases and reference populations
131 to estimate penetrance. As well as providing aggregated penetrance estimates for
132 groups of rare variants (e.g., those curated as pathogenic), this approach can estimate
133 the penetrance of *individual* rare alleles. Importantly, these estimates represent variants
134 in the general population, rather than in families ascertained for disease.

135 **Methods**

136 *Case cohort*

137 Sequencing data for 10,400 individuals referred for HCM gene panel sequencing, and
138 2,564 individuals referred for DCM gene panel sequencing was collected from seven
139 international testing centres: three UK-based - the NIHR Royal Brompton Biobank,
140 Oxford Molecular Genetics Laboratory, and Belfast Regional Genetics Laboratory; two
141 US-based – the Partners Laboratory of Molecular Medicine and GeneDx; and the
142 National Heart Centre, Singapore, and Aswan Heart Centre, Egypt. Although the
143 diagnosis cannot directly be reconfirmed, given genetic testing guidelines (e.g.,^{3,27}), a
144 clinical diagnosis of CM is implicit. For information on DNA sequencing and data
145 obtained for analyses, see the supplementary materials.

146 For each variant observed in one or more individuals referred for CM sequencing, we
147 calculated the allele count (AC) and allele number (AN), and further stratified by
148 reported age, sex, and ancestry, where the data allowed. All patients gave written
149 informed consent, and the study was approved by the relevant regional research ethics
150 committees.

151 *Population cohort*

152 167,478 participants of the UK Biobank (UKBB) with whole exome sequencing data
153 available for analyses and 125,748 exome sequenced participants of the Genome
154 Aggregation Database (gnomAD; version v2.1.1) were included in this study.

155 Briefly, the UK Biobank recruited participants aged 40–69 years old from across the UK
156 between 2006 and 2010²⁸, of which the 200,571 exome tranche of individuals that had
157 not withdrawn, were included in this study²⁹. The maximal subset of unrelated
158 participants was used, identified by those included in the UKBB PCA analysis (S3.3.2²⁸,
159 n=167,478). Age at recruitment, genetic sex, and genetic (for European (EUR) and
160 British ancestry) or reported ancestry information (for other global ancestries: AFR,
161 African, Caribbean (n=2,903); SAS, Indian, Pakistani, Bangladeshi (n=3,136); EAS,
162 Chinese (n=605)) were incorporated.

163 The gnomAD database contains sequencing information for unrelated individuals
164 sequenced as part of various disease-specific and population genetic studies³⁰. The
165 version 2 short variant dataset spans 125,748 exomes from unrelated individuals
166 sequenced as part of various disease-specific and population genetic studies. Ensembl
167 Variant Effect Predictor³¹ (VEP, version 105) was used to incorporate the variant-
168 specific summary counts. Variants flagged by gnomAD as AC0 were excluded from
169 gnomAD counts. For more information on the incorporation of these datasets, please
170 see the supplementary materials.

171 *Variant annotation*

172 VEP (105) was used to annotate the case and population datasets, with additional
173 plugins: gnomAD³⁰ (version r2.1), LOFTEE³⁰, SpliceAI³² (1.3.1), REVEL³³ (1.3), and

174 ClinVar³⁴ (20220115). The data was organised using PLINK³⁵ (1.9) and the VEP output
175 was analysed using R (4.1.2).

176 Protein altering variants, defined with respect to MANE transcripts, that were annotated
177 as high or moderate impact by SO and ENSEMBL were included in the analysis.
178 Analysis was restricted to genes with strong or definitive evidence of causing CM
179 following ClinGen guidance^{36,37} and expert curation (gene2phenotype Cardiac Panel;
180 <https://www.ebi.ac.uk/gene2phenotype/>), to include 8 sarcomeric HCM-associated
181 genes (*MYH7*, *MYBPC3*, *MYL2*, *MYL3*, *ACTC1*, *TNNI3*, *TNNT2*, *TPM1*) and 11 DCM-
182 associated genes (*BAG3*, *DES*, *DSP*, *LMNA*, *MYH7*, *PLN*, *RBM20*, *SCN5A*, *TNNC1*,
183 *TNNT2*, *TTN_{PSI>90%}*), with the exception of *FLNC* that was not included on the panel
184 sequencing of the DCM case cohort (**Table S4**). Variants with consequences consistent
185 with the known disease-causing mechanism were retained.

186 Further manual annotation was undertaken following ACMG guidelines using ClinVar³⁴
187 and Cardioclassifier³⁸, as previously published¹³. For analyses of variants in aggregate,
188 the UKBB data was filtered following the same thresholds and used to estimate
189 aggregate penetrance.

190 *Statistical analysis*

191 Estimation of penetrance and 95% confidence interval

192 Penetrance is expressed as a probability function on a scale of 0%-100%. Penetrance
193 was estimated from case-population data in a Binomial framework following Bayes'
194 theorem for conditional probability²⁶

$$P(D|A) = P(D) \frac{P(A|D)}{P(A)},$$

$$\text{penetrance} = \text{population prevalence} \frac{\text{case allele frequency}}{\text{population allele frequency}},$$

195 where, *D*, disease; *A*, allele; and *P*, probability, $P(D|A)$ = penetrance (probability of
196 disease given a risk allele), $P(D)$ = fixed population baseline risk of disease (probability
197 of disease), $P(A|D)$ = allele frequency in the case cohort (probability of the allele given
198 disease), $P(A)$ = allele frequency in the population cohort (probability of the allele).

199 We define penetrance in this setting as the probability of dominant CM by late adulthood
200 (UKBB had a mean age of 56 years old at recruitment). We assume the independence
201 of the random variables in the penetrance equation above to derive the 95% confidence
202 interval for penetrance as the product and ratio of binomial proportions. We used the
203 specialised version of the Central Limit Theorem, the Delta method, on the log-
204 transformed random variable $\log(D|A) = \log(D) + \log(A|D) - \log(A)$ with an improved
205 mean approximation and adjustment for degeneracy (as allele frequency tends to 0 for
206 rare variants). Please see additional methods and alternative approaches considered
207 (**Supplementary methods, Table S3, Lists S1-S2, Figures S4-S5**).

208 For estimates of penetrance by sex, we adjusted all terms of the penetrance equation
209 by values for sex-specific parameters. For estimates of penetrance by ancestry, we kept
210 $P(D)$ as estimated for CM (there are few estimates of the prevalence of CM in specific
211 ancestries) and proportioned $P(A|D)$ and $P(A)$ by reported ancestry. For estimates of
212 penetrance by age, we normalised $P(D)$ by the number diagnosed in the case cohort by
213 a particular age in a cumulative fashion, with $P(A|D)$ by a particular age, and $P(A)$ fixed
214 as total population allele frequency (**Supplementary methods**).

215 Estimated cardiomyopathy prevalence

216 To incorporate $P(D)$ in our penetrance analysis, we estimated the uncertainty
217 surrounding the reported prevalence of CM (**Tables S1-S2, Figures S1-S3**). For HCM,
218 we meta-analysed four imaging-based prevalence estimates^{13,39-41} excluding studies
219 with potential selection biases. From the meta-analysis estimate ($p_D \equiv P(D)$) and its
220 confidence interval, we derived values of allele count, x_D , and allele number, n_D (where
221 $p = \frac{x}{n}$). A literature review was also completed for DCM, but there were not enough
222 imaging-based prevalence estimates in literature, so 39,003 participants of the UK
223 Biobank imaging cohort were used to estimate phenotypic DCM⁴²⁻⁴⁴ (**Supplementary**
224 **methods**). Using the same methods and included studies, we derived estimates for
225 male- and female-specific HCM and DCM prevalence.

226 Results

227 *Case cohort summary information*

228 Sequencing data for 10,400 individuals referred for HCM genetic panel sequencing and
229 2,564 individuals referred for DCM genetic panel sequencing were included in the
230 analysis. Aggregated frequency of rare protein altering variants in well-established
231 disease genes was 42% for HCM and 33% for DCM in the respective case cohorts
232 (**Tables S6-S7**). Of the cohorts with age, sex, and ancestry information available (20%
233 of HCM cases, 42% of DCM cases), 35% and 32% were female, 93% and 91% were of
234 EUR ancestry, and mean age was 48 and 49 years old, for HCM and DCM, respectively
235 (**Table S5**).

236 *New estimates of the prevalence of cardiomyopathies*

237 To estimate the prevalence of CMs, a literature review and meta-analysis were
238 undertaken (**Tables S1-S2, Figures S1-S3**). Prevalence is underestimated when
239 derived from national cohorts using coding systems such as ICD codes, due to
240 incomplete ascertainment through diagnostic and procedure coding⁴⁵. We would
241 therefore expect the most accurate estimates of the prevalence of CM to come from
242 imaging studies in populations, where echocardiogram or cardiac magnetic resonance
243 imaging was used to identify CM within a population sample that is representative. The
244 estimates are not generalisable if the prevalence is estimated for selected subgroups of
245 individuals, such as young, elderly, or athletic cohorts. We therefore meta-analysed four
246 imaging-based prevalence estimates which resulted in a HCM population prevalence
247 estimate of 1 in 543 individuals ($p_D = 0.18\%$ (95% $CI_D = 0.15\%-0.23\%$))^{13,39-41}. The well
248 reported estimate of 1 in 500 individuals for HCM prevalence (0.20%) is within this
249 confidence interval.

250 A literature review revealed insufficient imaging-based estimates to undertake a direct
251 meta-analysis of the prevalence of DCM. Instead, we used 39,003 participants of the
252 UK Biobank imaging cohort to estimate phenotypic DCM⁴²⁻⁴⁴. This derived a DCM
253 population prevalence of 1 in 220 individuals ($p_D = 0.45\%$ (95% $CI_D = 0.39\%-0.53\%$)),
254 which includes the well reported estimate of 1 in 250 (0.40%)⁴⁶ within the confidence
255 interval.

256 We also estimated sex-specific CM prevalence. This resulted in a HCM population
257 prevalence of ~1 in 1,300 females ($p_D = 0.08\%$ (95% $CI_D = 0.04\%-0.12\%$)) and ~1 in
258 360 males ($p_D = 0.28\%$ (95% $CI_D = 0.22\%-0.35\%$)), and a DCM population prevalence
259 of ~1 in 340 females ($p_D = 0.30\%$ (95% $CI_D = 0.23\%-0.38\%$)) and ~1 in 160 males
260 ($p_D = 0.63\%$ (95% $CI_D = 0.52\%-0.75\%$)).

261 *Estimated penetrance of rare variants in aggregate*

262 In individuals with cardiomyopathy referred for diagnostic sequencing, we identified
263 1,340 rare (inclusive population allele frequency of <0.1%) variants in HCM-associated
264 genes (4,314 observations, case frequency 42%) and 665 rare variants in DCM-

265 associated genes (833 observations, case frequency 33%) (**Tables S6-S9**). The UKBB
266 dataset was filtered following the same pipeline. 1,735 rare variants in HCM-associated
267 genes (9,196 observations, 5.5% population frequency) and 4,586 rare variants in
268 DCM-associated genes (22,232 observations; 13.3% population frequency) were used
269 to estimate aggregate penetrance of rare variant subgroups. Variants with a pathogenic
270 classification in ClinVar were the most penetrant subgroup by ACMG classification⁴⁵
271 (HCM 22.5% (17.5%-28.8%), DCM 35.0% (21.6%-56.8%); **Figure 1, Table S15**).

272
273 An estimate of the aggregate penetrance of both pathogenic and likely pathogenic
274 variants in HCM was 10.7% (8.6%-13.3%) using this approach, concordant to a recent
275 estimate derived using direct assessment of cardiac imaging in UKBB (10.8%; variant
276 carriers with LVH ≥ 13 mm without hypertension or valve disease; binomial 95%
277 confidence interval of 3.0%-25.4%; $n=4/37$)¹¹. This concordance was also observed for
278 other variants in the same paper (e.g., VUS) which we estimated penetrance as 0.55%
279 (0.45%-0.68%) compared to 0.57% (0.07%-2.03%, $n=2/353$)¹¹.

280
281 The aggregate penetrance of pathogenic and likely pathogenic variants in DCM was
282 11.3% (9.3%-13.6%). Population penetrance of rare variants in DCM-associated genes
283 in UKBB has previously been estimated at $\leq 30\%$ ⁴⁷ for a clinical or subclinical diagnosis
284 in an analysis of 44 DCM-associated genes, and in the range of 5%-6% for TTNtvs
285 (1.9%-12.8%; 877 carriers)⁶ depending on the definition used. We report a concordant
286 penetrance estimate from our analysis of strong- and definitive-evidence DCM genes
287 only, and 9.8% (8.0%-12.1%) for TTNtvs (**Figure 2, Figure S12**).

288
289 Variants predicted to result in premature termination codons (PTCs; nonsense mediated
290 decay competent or incompetent⁴⁸) in *MYBPC3*, *BAG3*, *DSP*, and *LMNA*, were the most
291 penetrant (**Figure 2, Figure S12, Tables S13-S14, Tables S18-S19**). Inframe deletions
292 in *TNNT2* were highly penetrant for both HCM and DCM. TTNtvs and missense variants
293 predicted to be damaging in *TPM1* and *TNNC1*, had moderate penetrance.

294
295 Stratification by variant rarity showed that variants absent from the gnomAD database
296 were the most penetrant subgroup (HCM pathogenic 91.9% (57.3%-100.0%), HCM
297 likely pathogenic 20.2% (14.2%-28.7%), DCM pathogenic 100.0% (56.3%-100.0%),
298 DCM likely pathogenic 13.7% (11.2%-16.8%); **Figure 1, Table S16**). Stratification of
299 penetrance by sex identified increased penetrance for males compared to females for
300 rare variants in HCM-associated genes (**Figure 1, Figure S13, Table S20**). Penetrance
301 was estimated as $<20\%$ up to 50 years of age by modelling the penetrance of CM as an
302 age-related cumulative frequency using the proportion of cases referred at each age
303 decile (**Figure 1, Table S17**).

304
305 While there are limitations to the cohort size when split by reported ancestry and we are
306 unable to rule out local ancestry mismatches between case and population datasets,
307 there was no significant difference in the penetrance of truncating variants in *TTN*
308 (TTNtvs_{PSI>90%}) between African (5.7% (2.9%-10.9%)), European (6.9% (5.5-8.5%)),
309 East Asian (6.1% (3.0%-12.4%)), and South Asian (5.7% (2.1%-15.8%)) ancestries, as
310 previously suggested⁴⁹.

311 *Estimated penetrance of individual rare variants*

312 Of the variants identified and used to estimate penetrance in aggregate, we report four
313 subgroups of variants in our case series (**Figure 3**):

314 Group 1 consisted of 339 variants that were found in more than one case (case allele
315 count (AC) ≥ 2) and were ultra-rare in population reference sets (population AC (popAC)
316 ≤ 1). Penetrance cannot be estimated with precision for individual variants in this group,
317 since the population allele frequency (AF) cannot be estimated with precision. When
318 considered in aggregate this group have high penetrance (**Figures S14**). For HCM, 294
319 variants in group 1 were identified 1,322 times (13% case frequency, 31%
320 observations). 29% were curated as pathogenic (P, n=84, 41% of HCM group 1
321 observations), 34% were likely pathogenic (LP, n=100, 36% observations), and 37%
322 were curated as uncertain significance (VUS, n=110, 23% observations). For DCM, 45
323 variants in group 1 were identified 132 times (5% case frequency, 16% observations).
324 18% of these were P (n=8, 20% DCM group 1 observations), 49% LP (n=22, 55%
325 observations), and 33% VUS (n=15, 25% observations).

326 Group 2 included 316 variants found multiple times in both cases and population
327 reference datasets (caseAC ≥ 2 , popAC ≥ 2). This group is expected to include variants
328 with intermediate penetrance, including founder effect variants. For this group, we can
329 estimate AF in both populations, and therefore can estimate penetrance (**Figure 4**,
330 **Interactive Figure 4, Tables S10-S11**). These account for more than half of all variants
331 identified in HCM-associated genes and include those most likely to be identified as
332 SFs. For HCM, 257 variants were identified a total of 2,203 times (21% case frequency,
333 51% observations). 11% were P (n=29, 37% HCM group 2 observations), 25% LP
334 (n=64, 31% observations), 59% VUS (n=151, 29% observations), and 5% likely benign
335 (LB, n=13, 3% observations). 49 of these variants were recurrent at least ten times and
336 described a large portion of observations (case AC ≥ 10 ; found 1,424 times, 33.0% of
337 case cohort observations, case frequency of 13.7%). The median penetrance of these
338 was 14.6% ($\pm 14.4\%$ SD). For DCM, 59 variants were identified 140 times (6% case
339 frequency, 17% observations). None were curated as P, 24% were LP (n=14, 22% DCM
340 group 2 observations), 56% VUS (n=33, 53% observations), 17% LB (n=10, 21%
341 observations), and 3% B (n=2, 4% observations). With the current DCM case cohort
342 size, no variant was identified ten or more times.

343 The final two groups consisted of 1,350 variants with only a single observation in our
344 case series. This does not provide a reliable of case frequency so penetrance estimates
345 would lack precision. Group 3 variants were those identified multiple times in the
346 population (popAC ≥ 2), and consisted mostly of VUSs: For HCM, 206 variants were
347 identified (2% case frequency, 5% of case observations). This included 0.5% P (n=1;
348 MYBPC3:c.3297dup:p.(Tyr1100Valfs*49)), 5% LP (n=10), 92% VUS (n=189), and 3%
349 LB (n=6). For DCM, 232 variants were identified (9% case frequency, 28%
350 observations). 1% were P (n=3), 7% LP (n=17), 79% VUS (n=183), 12% LB (n=27), and
351 1% B (n=2).

352 Group 4 variants are those observed once in cases, and rarely in the population
353 reference dataset ($\text{popAC} \leq 1$). A substantial portion of these were P/LP: for HCM, 583
354 variants were identified (6% case frequency, 13% observations). 10% were P ($n=59$),
355 24% LP ($n=142$), and 66% VUS ($n=382$). For DCM, 329 variants were identified (13%
356 case frequency, 40% observations). 3% were P ($n=10$), 58% LP ($n=192$), and 39% VUS
357 ($n=127$).

358 *The impact of age, sex, and ancestry, on variant-specific penetrance estimates*

359 For group 2, where age-related penetrance could be derived, we estimated the
360 penetrance of specific variants by decade of age (e.g., **Figure 5**). For some variants
361 (e.g., MYBPC3:c.1624G>C:p.(Glu542Gln)), the age-related penetrance curve shows
362 infrequent onset before middle age. These curves may inform surveillance strategies in
363 variant carriers unaffected at first assessment.

364 We identified rare variants in HCM-associated genes where estimated penetrance for
365 males was significantly increased compared to females (**Figure S13**). Identification of
366 such variants allows for future investigations of modifiers protecting female carriers from
367 disease.

368 For estimates of penetrance by ancestry, variants that were nominally more common in
369 AFR, EAS, or SAS ancestries compared to EUR ancestry were identified (**Table S12**).
370 We interpret these as more consistent with inaccurate penetrance estimation arising
371 from ancestries where the variant is sparsely observed, rather than true differences in
372 penetrance on different ancestral background. For example,
373 MYBPC3:c.1544A>G:p.(Asn515Ser) was identified 5/492 times in AFR cases
374 ($\text{AF}=0.005$) and 33/10,655 times in the AFR population participants ($\text{AF}=0.0016$;
375 penetrance of 0.6% (0.2%-1.5%)) compared to 1/9,692 times in EUR cases
376 ($\text{AF}=0.00005$) and not observed in 211,532 EUR population participants. Even when
377 ancestry is nominally matched, broad continental groupings hide great diversity and
378 results may be misleading due to stratification between cases datasets (mostly North
379 AFR from Egypt) and population reference datasets (e.g., UK Biobank participants from
380 the Caribbean) (**Box 1**).

381 *Clinical impact of specific variants now shown to have low penetrance*

382 We can define the upper bound of the penetrance estimate for some variants. 162 rare
383 variants in HCM-associated genes (63% of variants, observed 745 times (7% case
384 frequency; 17% of observations) have a penetrance of $\leq 10\%$, according to the upper
385 limit (UCI) of the 95% CI for our estimate. These included two variants previously
386 curated as definitively pathogenic, and 25 variants curated as likely pathogenic.
387

388 One of the pathogenic variants is splice acceptor MYBPC3:c.26-2A>G which has an
389 estimated penetrance of 1.0% (0.4%-2.8%) or 0.9% (0.3%-2.5%) in EUR ancestry, as it
390 was identified 4 times in EUR cases and 20 times in population participants (90% were
391 EUR). The potential for this variant to have incomplete penetrance has been noted
392 previously through identified asymptomatic carriers (see ClinVar ID 42644). There is in

393 silico evidence of an alternate splice site downstream that could result in an in-frame
394 deletion of two amino acids.

395
396 The second pathogenic variant identified with a UCI of $\leq 10\%$ is the missense variant
397 MYH7:c.3158G>A:p.R1053Q which is a Finnish founder mutation. This variant had an
398 estimate penetrance of 2.2% (0.9%-5.2%), as it was identified 7 times in EUR cases
399 and 17 times in the population cohort (16 Finnish from gnomAD, 1 NWE from UK
400 Biobank). Estimates of penetrance are sensitive to allele frequency differences across
401 ancestries. Analysis of founder mutations in the population they derive from would
402 provide additional confidence in their penetrance estimates.

403
404 For DCM, 17 rare variants (29% of variants) observed 45 times (2% case frequency; 5%
405 of observations) met this criterion. None of the 17 variants were curated as P/LP.

406 **Box 1**

407 *Case study: The MYBPC3:c.1504C>T:p.(Arg502Trp) Northwestern European variant*

408 The variant MYBPC3:c.1504C>T:p.R502W was found in our cohort 159 times in
409 patients referred for HCM genetic panel sequencing (3.7% of total observations; 1.5%
410 total case frequency). To date, the variant has been classified on ClinVar fifteen times
411 as pathogenic (ClinVar ID 42540). Penetrance has been previously estimated as ~50%
412 (increased relative risk of 340) by 45 years old in a clinical setting, with major adverse
413 clinical events in carriers significantly more likely when another sarcomeric variant is
414 present⁵⁰.

415
416 In our case cohort, carriers of this variant were reported as broadly European ancestry
417 (Oxford, n=59; London, n=11; Belfast, n=30; LMM, n=45; GDX, n=14). In gnomAD, the
418 variant was identified 10 times, of which 7 carriers were non-Finnish Northwestern
419 Europeans (NWE; plus 1 African; 1 South Asian, and 1 other), and in the UK Biobank,
420 the variant was found 77 times, of which 68 carriers were NWE (plus 8 other
421 Europeans, and 1 other). The population frequency of the variant in Ensembl population
422 genetics showed that the variant (rs375882485) is only found multiple times in NWE
423 ancestry sub cohorts. Thus, the variant is most common in NWE populations: the UK,
424 Ireland, Belgium, the Netherlands, Luxembourg, Northern France, Germany, Denmark,
425 Norway, Sweden, and Iceland.

426
427 We use this relatively common variant to highlight the effect of ancestry on estimated
428 variant penetrance (**Figure 6**): the penetrance was estimated as 6.4% (4.6%-9.0%)
429 using the UK Biobank cohort (93% European) and this is inflated to 35.1% (18.2%-
430 67.5%) when estimating the penetrance using the gnomAD dataset (45% European),
431 due to the difference in the proportion of individuals with NWE ancestry. In individuals of
432 NWE ancestry only, the penetrance of this variant is 6.4% (4.6%-9.0%). Penetrance
433 estimated from the NWE subset of gnomAD or UKB do not differ significantly.

434
435 As access to larger genomic datasets becomes available, including more diverse
436 ancestries, we can increase the precision of these variant-specific penetrance estimates
437 by gaining further confidence in maximum population allele frequencies⁵¹.

438 [end of box 1]

439

440 *Penetrance estimate simulations of increased cohort sizes*

441 We anticipate two benefits to estimating the penetrance of rare variants from increasing
442 cohort sizes: i) there will be more variants that are observed recurrently in cases and
443 populations, permitting AF estimates and hence penetrance estimates, and ii) the
444 precision of our penetrance estimates will increase as AF of rare variants is ascertained
445 with greater precision.

446 We sought to understand whether it would be more valuable to focus resources on
447 aggregating data from larger numbers of cases (plausible ~100,000 cases with global
448 collaboration efforts), and/or from larger numbers of population participants with near-
449 term publicly available population datasets (~5,000,000 participants).

450 Efforts to increase reference population sample size will provide additional confidence in
451 penetrance estimates once case aggregation to 10,000 cases is reached (**Figure S6**).
452 There is substantial confidence to be gained by increasing the population cohort size:
453 we found that increasing the population dataset from 300,000 participants to 4.5 million
454 participants could provide ~20% certainty, depending on the penetrance of the variant
455 (**Figures S7-S11**). The increase in confidence gained from increasing the case cohort
456 sample size from 10,000 cases to 100,000 cases was limited (with the caveat that more
457 variants will be identified).

458 Discussion

459

460 We show that some subgroups of rare variants in the population are penetrant and for
461 these it may be reasonable to return as SFs. These include ultra-rare variants, predicted
462 PTCs in certain genes where loss of function is a known disease mechanism, and
463 variants with enough evidence to have been classified as definitively pathogenic
464 previously (**Graphical abstract**).

465

466 There is still uncertainty regarding the penetrance of individual ultra-rare variants, and
467 the implications of returning SFs in healthcare systems have yet to be estimated. While
468 we have previously attempted to assess the burden of long-term surveillance for DCM⁶,
469 cost-effect analyses are vital to fully understand the risks and benefits of reporting SFs
470 in different healthcare systems. For variant types with low penetrance, it is very
471 uncertain that the benefit of returning SFs will outweigh harms and justify costs.

472

473 Here, we provide the first at-scale estimates of variant-specific penetrance for variants
474 in CM-associated genes that include those likely to be most frequently identified as SFs.
475 Most have low estimated penetrance, where an asymptomatic individual without family
476 history for disease may choose no or less frequent surveillance, depending on the
477 healthcare system and follow-up cost.

478

479 Population penetrance estimates derived from unselected individuals (with certain
480 caveats⁵²) that are agnostic to personal or family history of disease, should provide a
481 better estimate of the probability of manifesting disease when a variant is identified as a
482 SF. Importantly, the penetrance of variants found in CM patients and relatives in a
483 clinical setting is increased compared to the penetrance of variants estimated for those
484 identified through SFs (e.g., MYBPC3:c.1504C>T:p.R502W with estimated penetrance
485 of 50% in HCM patients and 6% here, in the population).

486

487 While published data is sparse and heterogeneous, overall estimates of penetrance by
488 adulthood in the general population are lower than family-based studies. An abstract
489 reported the penetrance of asymptomatic carriers referred to hospital for predictive
490 testing after identification of a genotype- and CM-positive relative⁵³. For HCM, 17 of 65
491 carriers (26.2%) were diagnosed with HCM (10 on first clinical evaluation, 7 during 2
492 years of follow up). For DCM, 2 of 22 carriers (9.1%) were diagnosed with DCM (2 on
493 first clinical evaluation, 0 during 2 years of follow up (excluding 5 with hypokinetic non-
494 dilated cardiomyopathy and 4 with isolated left ventricular dilatation))⁵³. Additionally, a
495 study of carriers identified during family screening who did not fulfill diagnostic criteria
496 for HCM at first evaluation identified HCM or an abnormal ECG in 127 of 285 carriers
497 (44.6%; 82 at baseline, 45 over a median of 8 years follow-up)²⁵. First degree relatives
498 in the same household may be at increased risk of disease due to shared environment
499 and other genetic factors.

500

501 The ACMG guidelines for reporting “medically actionable” variants in 78 genes come
502 with the caution that evaluating SFs requires an increased amount of supportive
503 evidence of pathogenicity, given the low prior likelihood that variants unrelated to the

504 indication are pathogenic⁵⁴. Here, we show that variants with a definitive pathogenic
505 assertion in ClinVar had the highest penetrance estimates. This may be because
506 penetrant variants are more likely to yield sufficient evidence for confident
507 interpretations, especially family segregation data.

508
509 Genetic laboratories communicate their confidence on whether a variant has a role in
510 disease (i.e., pathogenicity), but do not consistently indicate the penetrance.
511 Pathogenicity addresses whether a variant explains the aetiology of a patient.
512 Penetrance addresses the probability of future disease in carriers. The ClinGen
513 consortium Low-Penetrance/Risk Allele Working Group recommends providing
514 penetrance estimates on clinical reports (aggregate gene-level or individual variants)
515 and noting when penetrance is assumed or where current information is
516 limited/unavailable.

517
518 TTNtv are common in the general population (~1 in 250 for variants in exons
519 constitutively expressed in the adult heart), and we show that the penetrance in
520 aggregate of TTNtvs is reduced compared to predicted loss of function variants in other
521 CM-associated, haploinsufficient genes. While recent work has increased our
522 understanding of the functional mechanisms of TTNtvs in disease^{55,56}, future work is
523 required to identify modifiers of TTNtvs to understand this reduced penetrance in the
524 population.

525
526 The penetrance of a variant may depend on characteristics of the variant itself and
527 modulating effects of genetic background and environment. This study characterises
528 individual variants, while ongoing work is dissecting the role of secondary genetic
529 influences. Polygenic scores may identify individuals at particular risk of disease,
530 modifying the estimated penetrance of a single dominant variant.

531
532 We present two novel dimensions to estimates of penetrance: the penetrance in the
533 general population and variant-specific penetrance. As described, the results of this
534 method are concordant with previous population estimates of aggregated penetrance in
535 the UKBB population derived using independent approaches, providing confidence in
536 the methods. In addition, we provide updated estimates for the population prevalence of
537 HCM and DCM and stratify by sex. The addition of future, publicly available, large-scale,
538 global population datasets and biobanks will aid this area of research by allowing for
539 increased confidence in ancestry-specific population allele frequencies and CM
540 prevalence. We provide the summary counts for each variant via an online browser and
541 the function to estimate penetrance in R for transferability and use in other diseases and
542 datasets.

543 *Limitations*

544 This study has not been undertaken without careful consideration of the limitations. This
545 method cannot quantify the penetrance of pathogenic variants that are absent/singleton
546 in the population, while in aggregate the penetrance of this group of variants is
547 significant.

548 Comparisons of case and control allele frequency are vulnerable to confounding by
549 population stratification, and we have explored some examples in this manuscript. We
550 do not have genome-wide variation data to directly assess genetic ancestry for the case
551 cohort, so this is based on data reported by the referring clinician. As the EUR
552 participants dominate our case and population datasets, greater representation of
553 diverse ancestral backgrounds is essential for equitable access to genomic medicine.
554 Estimates of the penetrance of variants and the prevalence of cardiomyopathies in more
555 ancestral groups are required. The current data for both comes from UK Biobank which
556 has limitations⁵².

557 In the absence of genome-wide data we cannot exclude the possibility of unrecognised
558 or cryptic relatedness within the case cohort. As described by Minikel et al. (2016)²⁶,
559 when a variant is highly penetrant, cryptically related individuals are likely included in
560 case series and if a disease is fatal, population cohorts are likely depleted of causal
561 variants.

562 Case allele frequency in unrelated cases may not be a fair estimate of the case allele
563 frequency in all cases observed in the clinic. Our estimate of case allele frequency, and
564 therefore of penetrance, is influenced by genetic testing referral practice. If clinicians are
565 cautious and only refer selected high confidence cases for testing, case allele frequency
566 and estimated penetrance will be high, whereas if clinicians were to test widely and
567 indiscriminately then our apparent case allele frequency would be lower, resulting in
568 lower penetrance estimates¹⁸.

569 Current diagnostic data assumes that the testing centre obtained complete coverage of
570 the gene. Limited data was available on age and sex for large portions of the case
571 cohorts. Our DCM referred cohort was only moderate in size and thus, increases in
572 sample size here through global collaboration would aid our estimates of penetrance for
573 variants in DCM-associated genes.

574 Finally, the UKBB volunteer population cohort is healthier than the average individual⁵²,
575 and the gnomAD consortium includes some individuals with severe disease but likely at
576 a frequency equivalent to or lower than the general population³⁰. The proposed
577 penetrance model is an approximation since in reality the three parameters used on the
578 right-hand side of the penetrance equation share some degree of dependence.

579 **Conclusion**

580 This is the first evaluation of the penetrance of individual rare variants in CM-associated
581 genes at scale. These recurrent variants are those that are likely to generate SFs.
582 Variants previously annotated as pathogenic, loss-of-function variants in specific genes
583 susceptible to haploinsufficiency, and those that are the rarest in the population, have
584 high penetrance, similar to observations from family studies. This first attempt at
585 estimating the penetrance of rare variants has highlighted the requirement for large
586 case and population datasets with known genetic ancestry. We are now able to start
587 putting bounds on the estimate of penetrance for a specific variant identified as a

588 secondary finding: for some, including those expected to be most penetrant, we do not
589 currently have enough data, for others, we can provide asymptomatic variant carriers
590 with an estimated probability of manifesting disease.

591 **Data and code availability**

592 All case cohort data arising from this analysis is available through DECIPHER
593 (<https://www.deciphergenomics.org/>). Both gnomAD (<https://gnomad.broadinstitute.org/>)
594 and UK Biobank (<https://www.ukbiobank.ac.uk/>) population reference datasets are
595 publicly available. Analysis code is available on GitHub
596 (<https://github.com/ImperialCardioGenetics/variantfx/tree/main/PenetrancePaper>).

597 **Competing interests**

598 J.S.W. has consulted for MyoKardia, Inc., Foresite Labs, and Pfizer. A.H. now works for
599 AstraZeneca, UK. D.P.O. has consulted for Bayer. L.B. has consulted for Roche.
600 D.G.M. is a paid advisor to GlaxoSmithKline, Insitro, Variant Bio and Overtone
601 Therapeutics, and has received research support from AbbVie, Astellas, Biogen,
602 BioMarin, Eisai, Merck, Pfizer, and Sanofi-Genzyme; none of these activities are directly
603 related to the work presented here.

604 **Acknowledgements and funding sources**

605 This work was supported by the Wellcome Trust [107469/Z/15/Z; 200990/A/16/Z],
606 Medical Research Council (UK) [MC_UP_1605/13], British Heart Foundation
607 [RG/19/6/34387, RE/18/4/34215, FS/IPBSRF/22/27059], and the NIHR Imperial College
608 Biomedical Research Centre. The views expressed in this work are those of the authors
609 and not necessarily those of the funders. For open access, the authors have applied a
610 CC BY public copyright license to any Author Accepted Manuscript version arising from
611 this submission.

612 **Credit statement**

613 Conceptualization: J.S.W, K.A.M.; Methodology: J.S.W., K.A.M., L.B., X.Z., P.T.; Formal
614 Analysis: K.A.M., L.B., P.T., K.T., A.H.; Resources: K.T., R.B., W.T.W., D.M., P.C.J.,
615 B.F., D.M., S.P., S.C., M.A., Y.A., M.H.Y., D.O'R., H.W., J.S.W.; Data curation: K.A.M.,
616 P.T., E.M.; Writing – original draft: K.A.M.; Writing – review & editing: (all authors);
617 Visualization: K.A.M., J.S.W.; Supervision: J.S.W., H.W.; Project administration: J.S.W.,
618 P.B.

619 **Ethics declaration**

620 All research participants provided written informed consent, and the studies were
621 reviewed and approved by the relevant research ethics committee (*Aswan Heart*
622 *Centre*: FWA00019142, Research Ethics Committee code
623 20130405MYFAHC_CMV_20130330; *NIHR Royal Brompton Biobank*: South Central –
624 Hampshire B Research Ethics Committee, 09/H0504/104+5, 19/SC/0257; *National*
625 *Heart Centre Singapore*: Singhealth Centralised Institutional Review Board 2020/2353
626 and Singhealth Biobank Research Scientific Advisory Executive Committee SBRSA
627 2019/001v1; *UK Biobank*: National Research Ethics Service 11/NW/0382, 21/NW/0157,
628 under terms of access approval number 47602).

629 In addition, diagnostic laboratories (Oxford Molecular Genetics Laboratory, Belfast
630 Regional Genetics Laboratory, the Partners Laboratory of Molecular Medicine, and

631 GeneDx) provided aggregated (and therefore fully anonymous) cohort-level summaries
632 of variant data collected for clinical purposes during routine healthcare. Secondary use
633 of this data did not require research consent from individuals, and approval for public
634 release of the data followed local governance procedures. Data are publicly available,
635 through DECIPHER (<https://www.deciphergenomics.org/>). Analyses of these data do
636 not require Research Ethics Committee approval.

637 **References**

- 638 1. Blout Zawatsky, C.L., Bick, D., Bier, L., Funke, B., Lebo, M., Lewis, K.L., Orlova, E.,
639 Qian, E., Ryan, L., Schwartz, M.L.B., et al. (2023). Elective genomic testing: Practice
640 resource of the National Society of Genetic Counselors. *J. Genet. Couns.*
- 641 2. Miller, D.T., Lee, K., Abul-Husn, N.S., Amendola, L.M., Brothers, K., Chung, W.K.,
642 Gollob, M.H., Gordon, A.S., Harrison, S.M., Hershberger, R.E., et al. (2022). ACMG SF
643 v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A
644 policy statement of the American College of Medical Genetics and Genomics (ACMG).
645 *Genet. Med.* 24, 1407–1414.
- 646 3. Wilde, A.A.M., Semsarian, C., Márquez, M.F., Shamloo, A.S., Ackerman, M.J.,
647 Ashley, E.A., Sternick, E.B., Barajas-Martinez, H., Behr, E.R., Bezzina, C.R., et al.
648 (2022). European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia
649 Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS)
650 Expert Consensus Statement on the state of genetic testing for cardiac diseases . *EP*
651 *Eur.* 24, 1307–1367.
- 652 4. de Wert, G., Dondorp, W., Clarke, A., Dequeker, E.M.C., Cordier, C., Deans, Z., van
653 El, C.G., Fellmann, F., Hastings, R., Hentze, S., et al. (2021). Opportunistic genomic
654 screening. Recommendations of the European Society of Human Genetics. *Eur. J.*
655 *Hum. Genet.* 29, 365–377.
- 656 5. Ormondroyd, E., Mackley, M.P., Blair, E., Craft, J., Knight, J.C., Taylor, J.C., Taylor,
657 J., and Watkins, H. (2018). “Not pathogenic until proven otherwise”: Perspectives of UK
658 clinical genomics professionals toward secondary findings in context of a Genomic
659 Medicine Multidisciplinary Team and the 100,000 Genomes Project. *Genet. Med.* 20,
660 320–328.
- 661 6. McGurk, K.A., Zheng, S.L., Henry, A., Josephs, K., Edwards, M., de Marvao, A.,
662 Whiffin, N., Roberts, A., Lumbers, T.R., O’Regan, D.P., et al. (2022). Correspondence
663 on "ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome
664 sequencing: a policy statement of the American College of Medical Genetics and
665 Genomics. *Genet. Med.* 24, 744–746.
- 666 7. Wilson, J.M.G., and Jungner, G. (1968). Principles and practice of screening for
667 disease: public health papers no. 34 (Geneva: World Health Organization).
- 668 8. Wynn, J., Martinez, J., Bulafka, J., Duong, J., Zhang, Y., Chiuzan, C., Preti, J.,
669 Cremona, M.L., Jobanputra, V., Fyer, A.J., et al. (2018). Impact of Receiving Secondary
670 Results from Genomic Research: A 12-Month Longitudinal Study. *J. Genet. Couns.* 27,
671 709–722.
- 672 9. Hart, M.R., Biesecker, B.B., Blout, C.L., Christensen, K.D., Amendola, L.M.,
673 Bergstrom, K.L., Biswas, S., Bowling, K.M., Brothers, K.B., Conlin, L.K., et al. (2019).
674 Secondary findings from clinical genomic sequencing: prevalence, patient perspectives,
675 family history assessment, and health-care costs from a multisite study. *Genet. Med.*
676 *Off. J. Am. Coll. Med. Genet.* 21, 1100–1110.
- 677 10. Thauvin-Robinet, C., Thevenon, J., Nambot, S., Delanne, J., Kuentz, P., Bruel, A.-

- 678 L., Chassagne, A., Cretin, E., Pelissier, A., Peyron, C., et al. (2019). Secondary
679 actionable findings identified by exome sequencing: expected impact on the
680 organisation of care from the study of 700 consecutive tests. *Eur. J. Hum. Genet.* *27*,
681 1197–1214.
- 682 11. Mackley, M.P., Fletcher, B., Parker, M., Watkins, H., and Ormondroyd, E. (2017).
683 Stakeholder views on secondary findings in whole-genome and whole-exome
684 sequencing: A systematic review of quantitative and qualitative studies. *Genet. Med.* *19*,
685 283–293.
- 686 12. Ormondroyd, E., Harper, A.R., Thomson, K.L., Mackley, M.P., Martin, J., Penkett,
687 C.J., Salatino, S., Stark, H., Stephens, J., and Watkins, H. (2020). Secondary findings in
688 inherited heart conditions: a genotype-first feasibility study to assess phenotype,
689 behavioural and psychosocial outcomes. *Eur. J. Hum. Genet.* *28*, 1486–1496.
- 690 13. de Marvao, A., McGurk, K.A., Zheng, S.L., Thanaj, M., Bai, W., Duan, J., Biffi, C.,
691 Mazzarotto, F., Statton, B., Dawes, T.J.W., et al. (2021). Phenotypic Expression and
692 Outcomes in Individuals With Rare Genetic Variants of Hypertrophic Cardiomyopathy. *J.*
693 *Am. Coll. Cardiol.* *78*, 1097–1110.
- 694 14. Pirruccello, J.P., Bick, A., Wang, M., Chaffin, M., Friedman, S., Yao, J., Guo, X.,
695 Venkatesh, B.A., Taylor, K.D., Post, W.S., et al. (2020). Analysis of cardiac magnetic
696 resonance imaging in 36,000 individuals yields genetic insights into dilated
697 cardiomyopathy. *Nat. Commun.* *11*, 2254.
- 698 15. Pirruccello, J.P., Bick, A., Chaffin, M., Krishna, G., Choi, S.H., Lubitz, S.A., Carolyn,
699 Y., Ng, K., Philippakis, A., Ellinor, P.T., et al. (2021). Titin truncating variants in adults
700 without known congestive heart failure. *J. Am. Coll. Cardiol.* *75*, 1239–1241.
- 701 16. Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A.,
702 Guigo, R., Iossifov, I., Vasileva, A., and Lappalainen, T. (2018). Modified penetrance of
703 coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* *50*,
704 1327–1334.
- 705 17. Goodrich, J.K., Singer-Berk, M., Son, R., Sveden, A., Wood, J., England, E., Cole,
706 J.B., Weisburd, B., Watts, N., Caulkins, L., et al. (2021). Determinants of penetrance
707 and variable expressivity in monogenic metabolic conditions across 77,184 exomes.
708 *Nat. Commun.* *12*, 3505.
- 709 18. Kingdom, R., and Wright, C.F. (2022). Incomplete Penetrance and Variable
710 Expressivity: From Clinical Studies to Population Cohorts. *Front. Genet.* *13*, 920390.
- 711 19. Glazier, A.A., Thompson, A., and Day, S.M. (2019). Allelic imbalance and
712 haploinsufficiency in MYBPC3-linked hypertrophic cardiomyopathy. *Pflügers Arch. - Eur.*
713 *J. Physiol.* *471*, 781–793.
- 714 20. Roberts, A.M., Ware, J.S., Herman, D.S., Schafer, S., Baksi, J., Bick, A.G., Buchan,
715 R.J., Walsh, R., John, S., Wilkinson, S., et al. (2015). Integrated allelic, transcriptional,
716 and phenomic dissection of the cardiac effects of titin truncations in health and disease.
717 *Sci. Transl. Med.* *7*, 270ra6.

- 718 21. Tadros, R., Francis, C., Xu, X., Vermeer, A.M.C., Harper, A.R., Huurman, R., Kelu
719 Bisabu, K., Walsh, R., Hoorntje, E.T., te Rijdt, W.P., et al. (2021). Shared genetic
720 pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite
721 directions of effect. *Nat. Genet.* 53, 128–134.
- 722 22. Harper, A.R., Goel, A., Grace, C., Thomson, K.L., Petersen, S.E., Xu, X., Waring,
723 A., Ormondroyd, E., Kramer, C.M., Ho, C.Y., et al. (2021). Common genetic variants
724 and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and
725 expressivity. *Nat. Genet.* 53, 135–142.
- 726 23. McGurk, K.A., and Halliday, B.P. (2022). Dilated cardiomyopathy – details make the
727 difference. *Eur. J. Heart Fail.* 24, 1197–1199.
- 728 24. Gail, M.H., Pee, D., Benichou, J., and Carroll, R. (1999). Designing studies to
729 estimate the penetrance of an identified autosomal dominant mutation: Cohort, case-
730 control, and genotyped-proband designs. *Genet. Epidemiol.* 16, 15–39.
- 731 25. Lorenzini, M., Norrish, G., Field, E., Ochoa, J.P., Cicerchia, M., Akhtar, M.M., Syrris,
732 P., Lopes, L.R., Kaski, J.P., and Elliott, P.M. (2020). Penetrance of Hypertrophic
733 Cardiomyopathy in Sarcomere Protein Mutation Carriers. *J. Am. Coll. Cardiol.* 76, 550–
734 559.
- 735 26. Minikel, E.V., Vallabh, S.M., Lek, M., Estrada, K., Samocha, K.E.,
736 Sathirapongsasuti, J.F., McLean, C.Y., Tung, J.Y., Yu, L.P.C., Gambetti, P., et al.
737 (2016). Quantifying prion disease penetrance using large population control cohorts.
738 *Sci. Transl. Med.* 8, 322ra9.
- 739 27. Ackerman, M.J., Priori, S.G., Willems, S., Berul, C., Brugada, R., Calkins, H.,
740 Camm, A.J., Ellinor, P.T., Gollob, M., Hamilton, R., et al. (2011). HRS/EHRA Expert
741 Consensus Statement on the State of Genetic Testing for the Channelopathies and
742 Cardiomyopathies. *EP Eur.* 13, 1077–1109.
- 743 28. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A.,
744 Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with
745 deep phenotyping and genomic data. *Nature* 562, 203–209.
- 746 29. Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G.,
747 Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., et al. (2021). Advancing
748 human genetics research and drug discovery through exome sequencing of the UK
749 Biobank. *Nat. Genet.* 53, 942–948.
- 750 30. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q.,
751 Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational
752 constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443.
- 753 31. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek,
754 P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.*
755 17, 122.
- 756 32. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F.,
757 Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019).

- 758 Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.
- 759 33. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti,
760 S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble
761 Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum.*
762 *Genet.* 99, 877–885.
- 763 34. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M.,
764 and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence
765 variation and human phenotype. *Nucleic Acids Res.* 42, D980-5.
- 766 35. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D.,
767 Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for
768 whole-genome association and population-based linkage analyses. *Am. J. Hum.*
769 *Genet.* 81, 559–575.
- 770 36. Ingles, J., Goldstein, J., Thaxton, C., Caleshu, C., Corty, E.W., Crowley, S.B.,
771 Dougherty, K., Harrison, S.M., McGlaughon, J., Milko, L. V., et al. (2019). Evaluating the
772 Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ. Genomic Precis. Med.* 12,
773 57–64.
- 774 37. Jordan, E., Peterson, L., Ai, T., Asatryan, B., Bronicki, L., Brown, E., Celeghin, R.,
775 Edwards, M., Fan, J., Ingles, J., et al. (2021). Evidence-Based Assessment of Genes in
776 Dilated Cardiomyopathy. *Circulation* 144, 7–19.
- 777 38. Whiffin, N., Walsh, R., Govind, R., Edwards, M., Ahmad, M., Zhang, X., Tayal, U.,
778 Buchan, R., Midwinter, W., Wilk, A.E., et al. (2018). CardioClassifier: disease- and
779 gene-specific computational decision support for clinical genome interpretation. *Genet.*
780 *Med.* 20, 1246–1254.
- 781 39. Zou, Y., Song, L., Wang, Z., Ma, A., Liu, T., Gu, H., Lu, S., Wu, P., Zhang, Y., Shen,
782 L., et al. (2004). Prevalence of idiopathic hypertrophic cardiomyopathy in China: A
783 population-based echocardiographic analysis of 8080 adults. *Am. J. Med.* 116, 14–18.
- 784 40. Maron, B.J., Spirito, P., Roman, M.J., Paranicas, M., Okin, P.M., Best, L.G., Lee,
785 E.T., and Devereux, R.B. (2004). Prevalence of hypertrophic cardiomyopathy in a
786 population-based sample of American Indians aged 51 to 77 years (the Strong Heart
787 Study). *Am. J. Cardiol.* 93, 1510–1514.
- 788 41. Maron, B.J., Gardin, J.M., Flack, J.M., Gidding, S.S., Kurosaki, T.T., and Bild, D.E.
789 (1995). Prevalence of Hypertrophic Cardiomyopathy in a General Population of Young
790 Adults. *Circulation* 92, 785–789.
- 791 42. Petersen, S.E., Aung, N., Sanghvi, M.M., Zemrak, F., Fung, K., Paiva, J.M., Francis,
792 J.M., Khanji, M.Y., Lukaschuk, E., Lee, A.M., et al. (2017). Reference ranges for cardiac
793 structure and function using cardiovascular magnetic resonance (CMR) in Caucasians
794 from the UK Biobank population cohort. *J. Cardiovasc. Magn. Reson.* 19, 18.
- 795 43. Mestroni, L., Maisch, B., McKenna, W.J., Schwartz, K., Charron, P., Rocco, C.,
796 Tesson, F., Richter, A., Wilke, A., and Komajda, M. (1999). Guidelines for the study of
797 familial dilated cardiomyopathies. Collaborative Research Group of the European

- 798 Human and Capital Mobility Project on Familial Dilated Cardiomyopathy. *Eur. Heart J.*
799 *20*, 93–102.
- 800 44. McNally, E.M., and Mestroni, L. (2017). Dilated cardiomyopathy: Genetic
801 determinants and mechanisms. *Circ. Res.* *121*, 731–748.
- 802 45. O'Malley, K.J., Cook, K.F., Price, M.D., Wildes, K.R., Hurdle, J.F., and Ashton, C.M.
803 (2005). Measuring diagnoses: ICD code accuracy. *Health Serv. Res.* *40*, 1620–1639.
- 804 46. Hershberger, R.E., Hedges, D.J., and Morales, A. (2013). Dilated cardiomyopathy:
805 The complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* *10*, 531–547.
- 806 47. Shah, R.A., Asatryan, B., Dabbagh, G.S., Aung, N., Khanji, M.Y., Lopes, L.R., van
807 Duijvenboden, S., Holmes, A., Muser, D., Landstrom, A.P., et al. (2022). Frequency,
808 Penetrance, and Variable Expressivity of Dilated Cardiomyopathy-Associated Putative
809 Pathogenic Gene Variants in UK Biobank Participants. *Circulation* *146*, 101–124.
- 810 48. Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of
811 nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* *48*, 1112–1118.
- 812 49. Haggerty, C.M., Damrauer, S.M., Levin, M.G., Birtwell, D., Carey, D.J., Golden,
813 A.M., Hartzel, D.N., Hu, Y., Judy, R., Kelly, M.A., et al. (2019). Genomics-First
814 Evaluation of Heart Disease Associated With Titin-Truncating Variants. *Circulation* *140*,
815 42–54.
- 816 50. Saltzman, A.J., Mancini-DiNardo, D., Li, C., Chung, W.K., Ho, C.Y., Hurst, S., Wynn,
817 J., Care, M., Hamilton, R.M., Seidman, G.W., et al. (2010). Short communication: the
818 cardiac myosin binding protein C Arg502Trp mutation: a common cause of hypertrophic
819 cardiomyopathy. *Circ. Res.* *106*, 1549–1552.
- 820 51. Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A.H., Karczewski, K., Ing, A.Y.,
821 Barton, P.J.R., Funke, B., Cook, S.A., Macarthur, D., et al. (2017). Using high-resolution
822 variant frequencies to empower clinical genome interpretation. *Genet. Med.* *19*, 1151–
823 1158.
- 824 52. Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins,
825 R., and Allen, N.E. (2017). Comparison of Sociodemographic and Health-Related
826 Characteristics of UK Biobank Participants With Those of the General Population. *Am.*
827 *J. Epidemiol.* *186*, 1026–1034.
- 828 53. Mahmood, A., Morris-Rosendahl, D., Edwards, M., Fleming, A., Homfray, T., Mason,
829 S., Quinn, E., Ware, J., Baksi, J., Prasad, S., et al. (2022). 10 Disease penetrance in
830 asymptomatic carriers of familial cardiomyopathy variants. *Heart* *108*, A9 LP-A9.
- 831 54. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W.,
832 Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the
833 interpretation of sequence variants: A joint consensus recommendation of the American
834 College of Medical Genetics and Genomics and the Association for Molecular
835 Pathology. *Genet. Med.* *17*, 405–424.
- 836 55. McAfee, Q., Chen, C.Y., Yang, Y., Caporizzo, M.A., Morley, M., Babu, A., Jeong, S.,
837 Brandimarto, J., Bedi, K.C., Flam, E., et al. (2022). Truncated titin proteins in dilated

838 cardiomyopathy. *Sci. Transl. Med.* *13*, eabd7287.

839 56. Fomin, A., Gärtner, A., Cyganek, L., Tiburcy, M., Tuleta, I., Wellers, L., Folsche, L.,
840 Hobbach, A.J., von Frieling-Salewsky, M., Unger, A., et al. (2022). Truncated titin
841 proteins and titin haploinsufficiency are targets for functional recovery in human
842 cardiomyopathy due to TTN mutations. *Sci. Transl. Med.* *13*, eabd3079.

843

844 **Figure legends**

845

846

847 **Figure 1** Aggregated penetrance of rare variants by variant curation, rarity, age, and
848 sex.

849 In aggregate, variants curated as pathogenic and variants that are particularly rare
850 (gnomAD AC=0) were most penetrant. The plot depicts aggregated estimated
851 penetrance for rare variants in HCM- (A, B, C, D) and DCM-associated (E, F, G, H)
852 genes. Variant curation was assessed following ACMG guidelines through ClinVar and
853 CardioClassifier software with additional manual curation of variants with conflicting
854 evidence (A, E (for HCM cases: 173 P variants, 316 LP, 832 VUS, 19 LB; for the UK
855 Biobank: 30 P, 98 LP, 1552 VUS, 53 LB, 2 B; for DCM cases: 21 P, 245 LP, 358 VUS,
856 37 LB, 4 B; for the UK Biobank: 15 P, 505 LP, 3951 VUS, 108 LB, 7 B)). The variants
857 were assessed for rarity by gnomAD allele count (AC) bins, where 0 is not identified in
858 the gnomAD dataset (B, F). Age was assessed in decades based on the cumulative
859 proportion of cases analysed by each age timepoint (C, G). Sex was estimated using all
860 parameters stratified by reported sex (D, H).

861

862 **Figure 2** The aggregate estimates of penetrance of “loss of function” variants are high
863 for specific genes.

864 The plot depicts estimated penetrance of HCM-associated (left) and DCM-associated
865 (right) rare variants. pLoF and non-pLoF variant groups are plotted in green and blue,
866 respectively. pLoF, predicted loss of function variants; *, TTNtv that are PSI>90%.
867 Pathogenic *TNNT2* inframe deletions caused an increased penetrance signal for
868 inframe deletions for both HCM and DCM (see **Figure S12**).

869

870 **Figure 3** Penetrance of individual variants could be estimated for 316 recurrently
871 observed rare variants from group 2.

872 The figure shows variant counts and subgroups for rare variants in HCM-associated
873 (left) and DCM-associated (right) genes. The colour of the description matches the
874 subgroup of variants depicted. The pie charts plot the proportion of all variant
875 observations in each subgroup (also denoted as ‘G+’). The observations approximate to
876 the number of variant carriers, though a small number of individuals may carry more
877 than one variant. G+, denotes the proportion of all variants identified; all, frequency of
878 the variant in cases; obs, observations of allele count.

879

880 **Figure 4** Variant-specific estimates of penetrance for the 316 recurrently observed rare
881 variants in CM-associated genes from group 2.

882 An interactive widget is available for browsing the individual variants in this figure (**see**
883 **Figure4_interactive.html**). The variants depicted (HCM n=257 (top), DCM n=59
884 (bottom)) were identified multiple times in cases and population reference datasets and
885 penetrance could therefore be estimated. The x-axis denotes the number of times the
886 variant was observed in each case cohort. AC, allele count; B/LB, benign/likely benign;
887 VUS, variant of uncertain significance; LP, likely pathogenic; P, pathogenic.

888

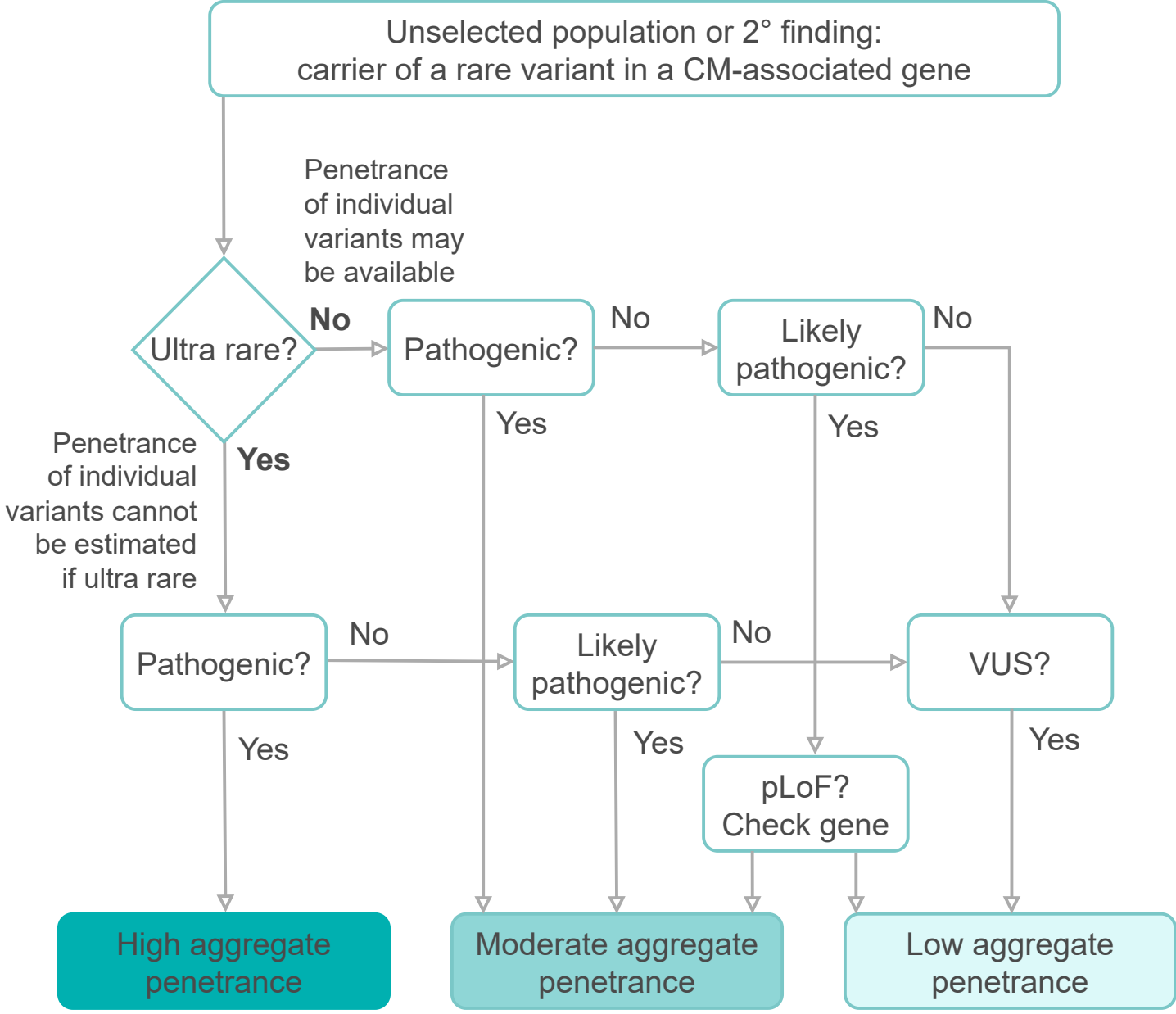
889 **Figure 5** Variant-specific estimates of penetrance by age can now be derived.

890 The plot depicts the age-related cumulative penetrance of five HCM-associated rare
891 variants across age deciles from variant group 2. The x-axis starts in the decade of the
892 20s as the analysis of these variants was underpowered for teens and younger. “20s”
893 here means “by 30 years old”.

894

895 **Figure 6** Penetrance estimates are inflated with underestimated population frequency.

896 The map of the world emphasises the large proportion of observations of
897 *MYBPC3:c.1504C>T:p.(Arg502Trp)* in HCM cases of Northwestern European (NWE)
898 ancestry. The numbers on the map are the counts of rare variant genotype positive
899 observations (n≈cohort participants) from each cohort with the specified ancestry, and
900 the percentages derive the proportion of observations that are due to the
901 *MYBPC3:c.1504C>T:p.(Arg502Trp)* variant. The graph shows the penetrance estimated
902 for the variant based on subgroups of reference dataset participants included. The
903 penetrance is inflated when estimated using gnomAD due to the variant being most
904 common in participants with NWE ancestry (which dominates the UKBB dataset).
905 Population frequency of gnomAD, UK Biobank, and Ensembl population genetics
906 showed that this variant (rs375882485) is only found multiple times in NWE ancestry
907 sub cohorts. The map excludes Antarctica for figure clarity. A limitation is the low
908 sample sizes for AFR, SAS, and EAS, ancestries.



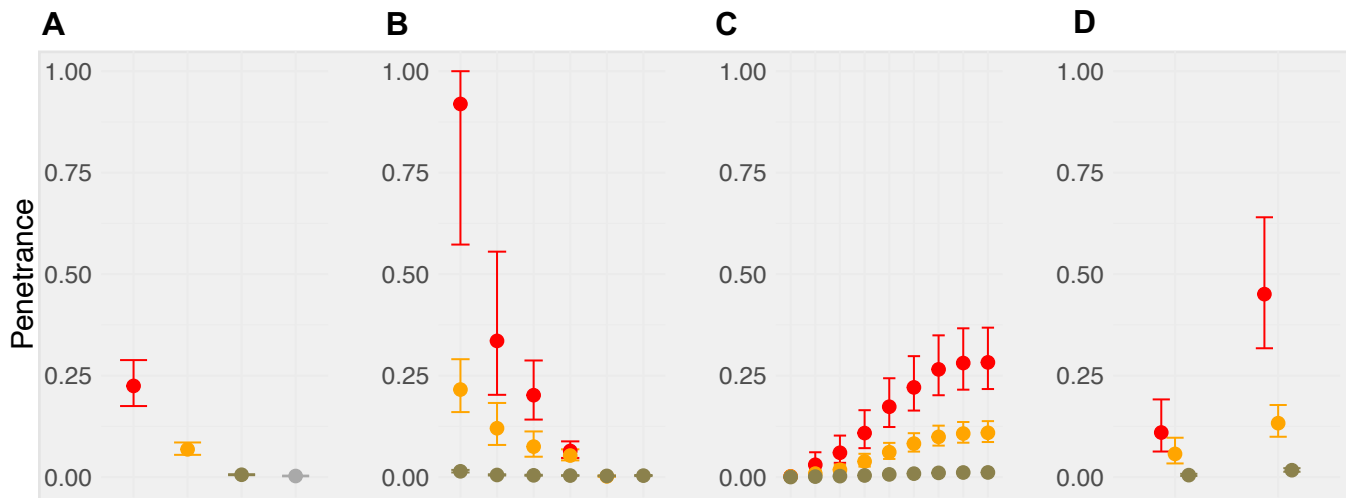
Variant curation

gnomAD AC
(allele count)

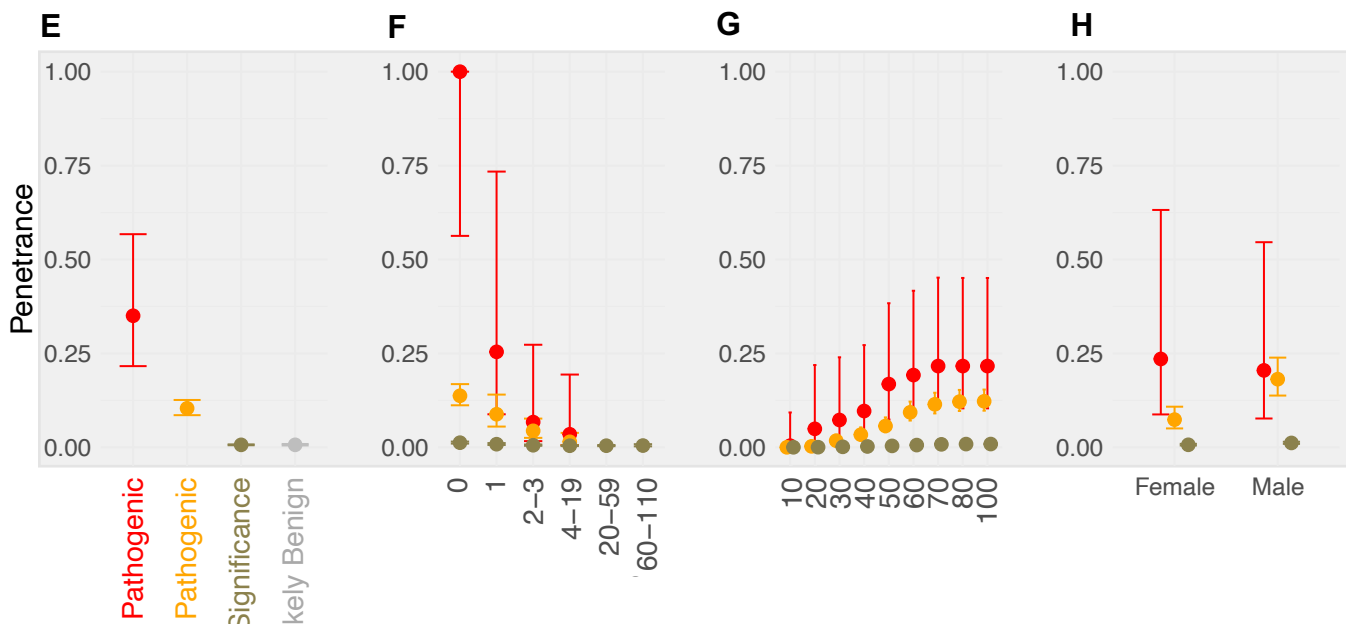
Age (years)

Sex

HCM

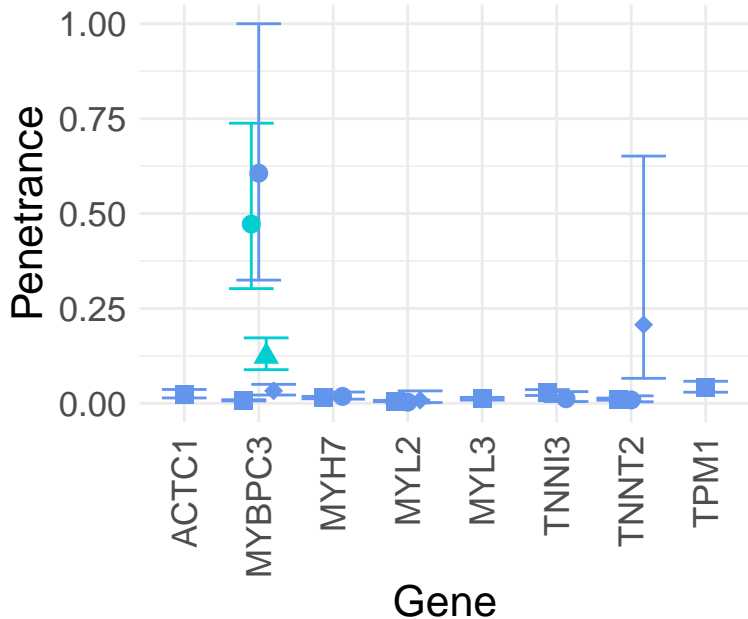


DCM

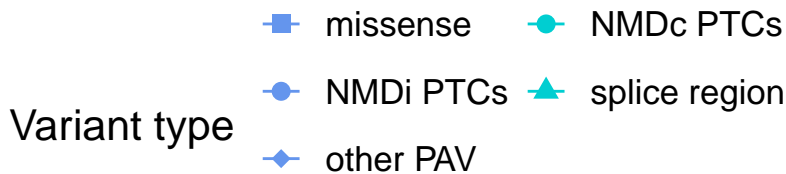
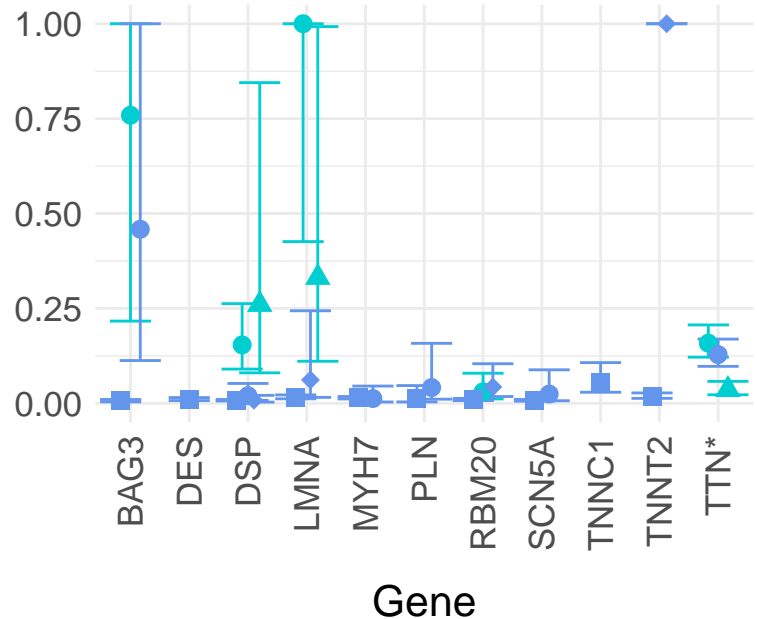


Curation P LP VUS

HCM



DCM



HCM

1,340 distinct variants:
4,314 observations (42% case frequency)
amongst 10,400 cases

Variants identified in the population

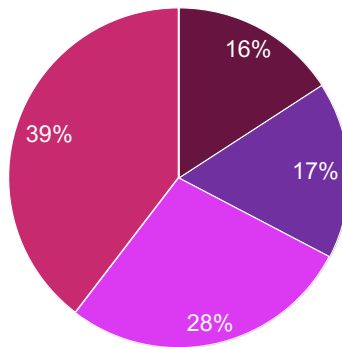
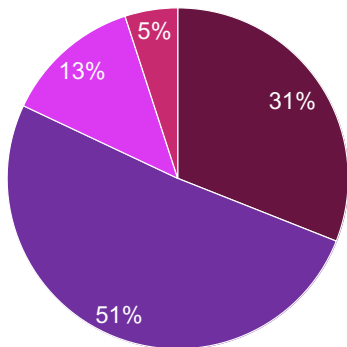
		≤1 obs	≥2 obs	Total
Variants identified in cases	≥2 obs	Group 1 294 1,322 obs (13% all, 31% G+)	Group 2* 257 2,203 obs (21% all, 51% G+)	551
	1 obs	Group 4 583 583 obs (6% all, 13% G+)	Group 3 206 206 obs (2% all, 5% G+)	789
Total		877	463	1,340

DCM

665 distinct variants:
833 observations (33% case frequency)
amongst 2,564 cases

Variants identified in the population

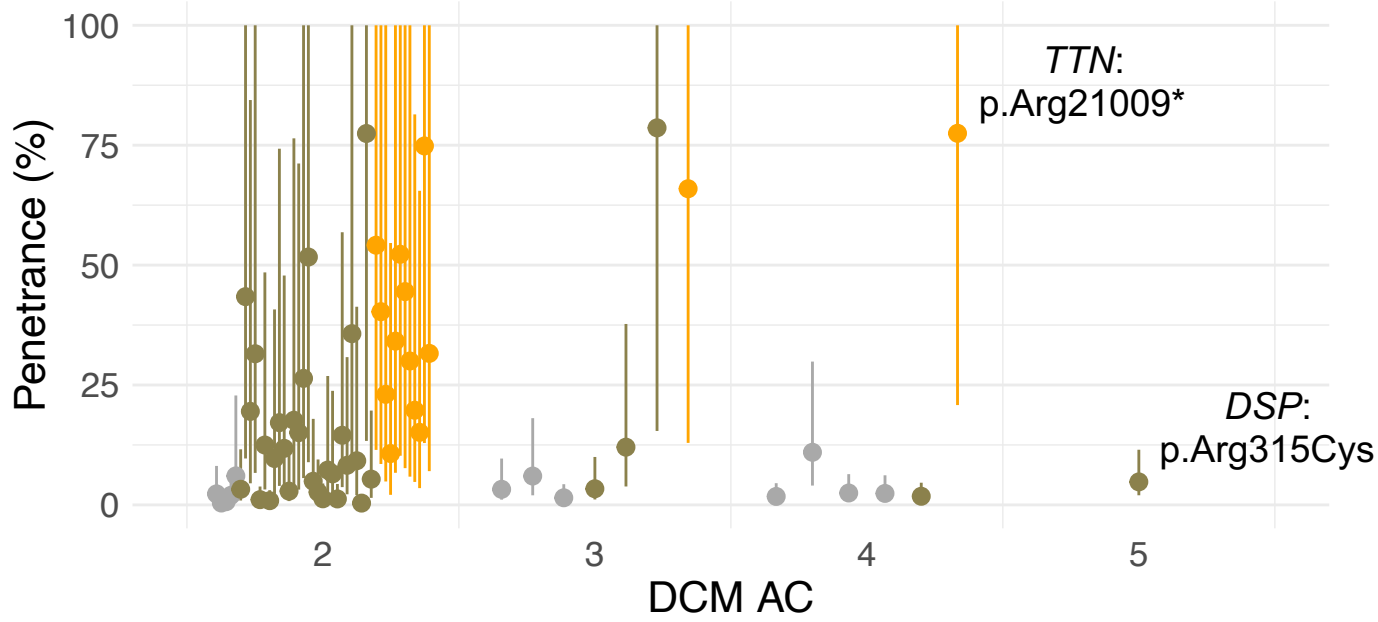
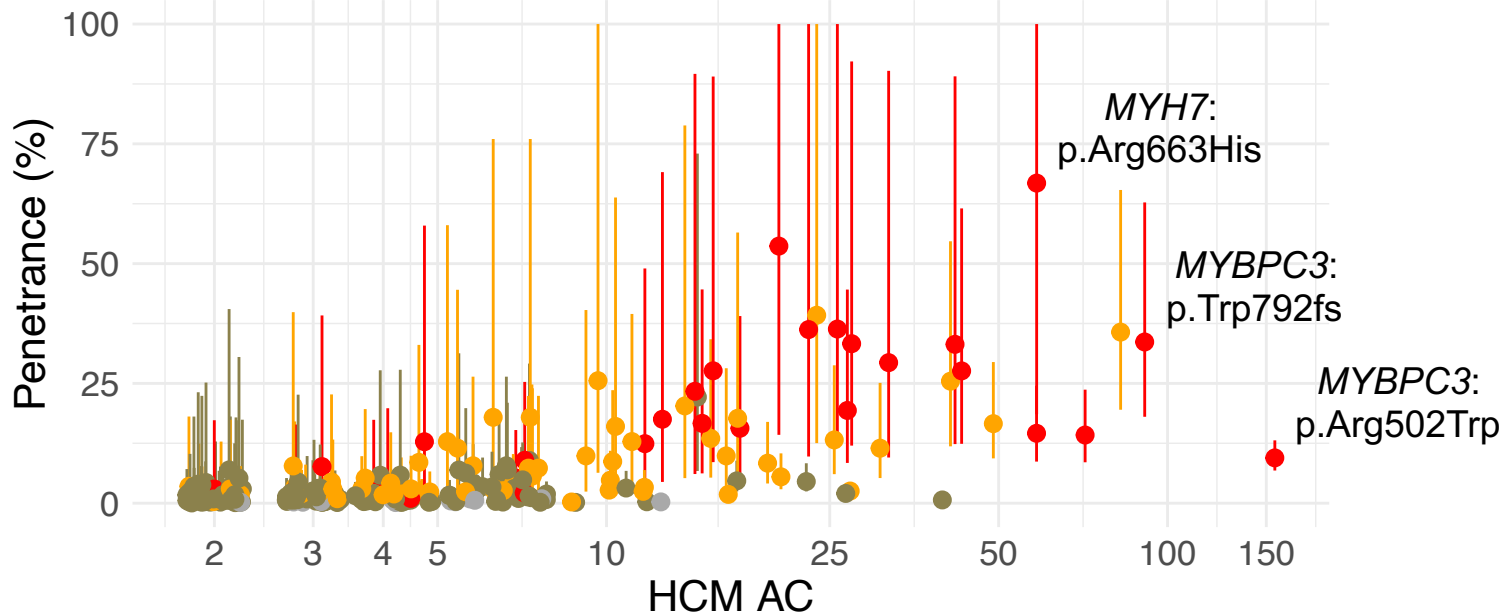
		≤1 obs	≥2 obs	Total
Variants identified in cases	≥2 obs	Group 1 45 132 obs (5% all, 16% G+)	Group 2* 59 140 obs (6% all, 17% G+)	104
	1 obs	Group 4 329 329 obs (13% all, 39% G+)	Group 3 232 232 obs (9% all, 28% G+)	561
Total		374	291	665

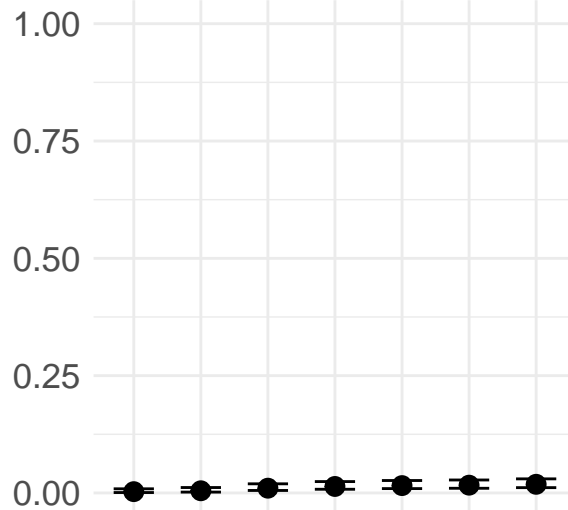
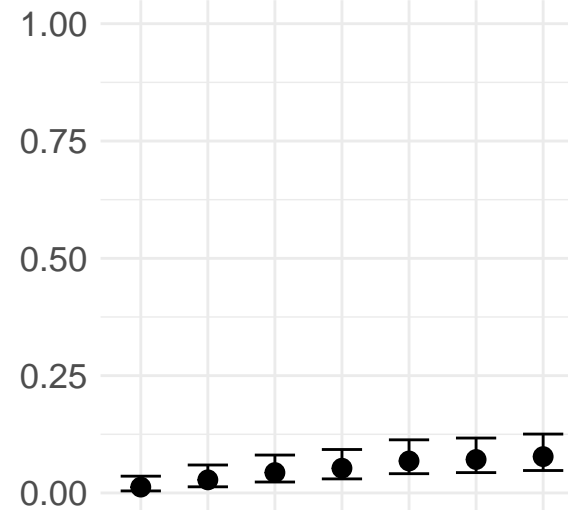
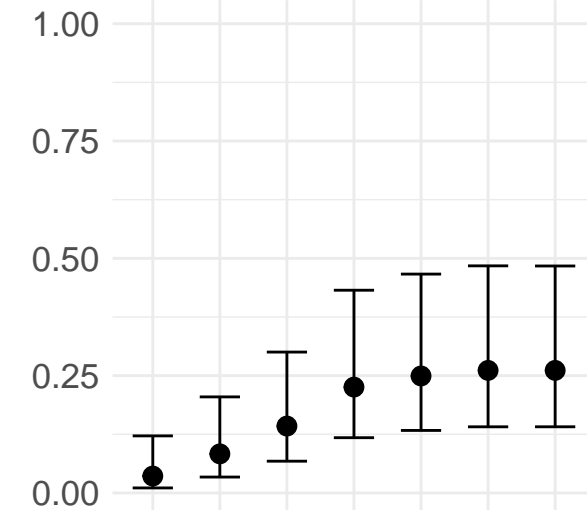
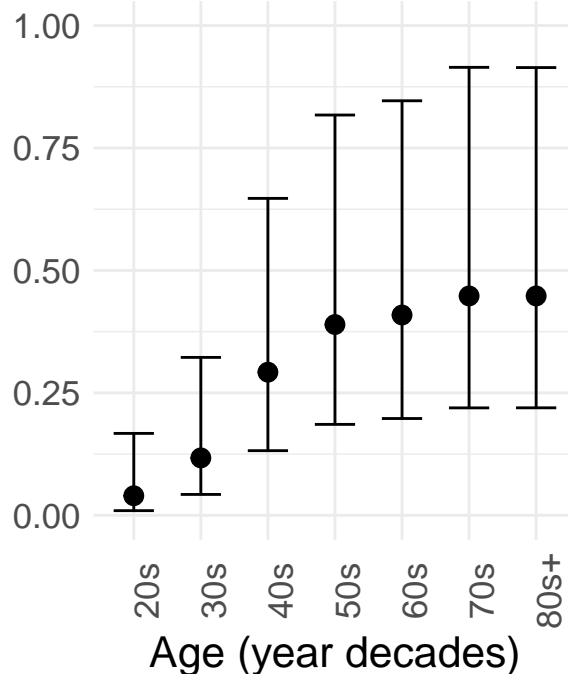


- Group 1:** Variants observed recurrently in cases and absent or singleton in the population: penetrance estimates are unreliable as the population frequency is uncertain. This group is expected to include most definitively pathogenic, high penetrance variants.
- Group 2*:** Variants observed recurrently in cases and the wider population: these are the variants most likely to be observed as secondary findings. *Penetrance can be estimated.
- Group 3:** Variants observed once in cases, and recurrently in the population: penetrance estimates are unreliable as the case frequency is uncertain. Variants in this group are likely either not pathogenic or have low penetrance.
- Group 4:** Variants are singleton in cases and absent or singleton in the population: current data is too sparse to estimate penetrance.

Group 2

Curation P LP VUS B/LB



TNNT2:c.862C>T:p.R288C**MYBPC3:c.1504C>T:p.R502W****MYBPC3:c.1624G>C:p.E542Q****MYBPC3:c.2373dup:p.-792*fs*41****MYH7:c.1988G>A:p.R663H**