

1 **A spectrum of recessiveness among Mendelian disease variants in UK Biobank**

2

3 Alison R. Barton,^{*1,2,3} Margaux L.A. Hujoel,^{1,2} Ronen E. Mukamel,^{1,2} Maxwell A. Sherman,^{1,2,4} Po-Ru
4 Loh^{*1,2}

5

6 1 Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical
7 School, Boston, MA, 02115, USA

8 2 Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

9 3 Bioinformatics and Integrative Genomics Program, Department of Biomedical Informatics, Harvard
10 Medical School, Boston, MA, 02115, USA

11 4 Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology,
12 Cambridge, MA, 02139, USA

13 * Correspondence should be addressed to A.R.B. (alisonbarton@g.harvard.edu) or P.-R.L.
14 (poruloh@broadinstitute.org)

15

16 **Abstract**

17 Recent work has found increasing evidence of mitigated, incompletely penetrant phenotypes in
18 heterozygous carriers of recessive Mendelian disease variants. We leveraged whole-exome imputation
19 within the full UK Biobank cohort ($N \sim 500K$) to extend such analyses to 3,481 rare variants curated from
20 ClinVar and OMIM. Testing these variants for association with 57 quantitative traits yielded 103
21 significant associations involving variants previously implicated in 35 different diseases. Notable
22 examples included a *POR* missense variant implicated in Antley-Bixler syndrome that associated with a
23 1.76 (s.e. 0.27) cm increase in height, and an *ABCA3* missense variant implicated in interstitial lung
24 disease that associated with reduced FEV1/FVC ratio. Association analyses with 1,257 disease traits
25 yielded five additional variant-disease associations. We also observed contrasting levels of
26 recessiveness between two more-common, classical Mendelian diseases. Carriers of cystic fibrosis
27 variants exhibited increased risk of several mitigated disease phenotypes, whereas carriers of spinal
28 muscular atrophy alleles showed no evidence of altered phenotypes. Incomplete penetrance of cystic
29 fibrosis carrier phenotypes did not appear to be mediated by common allelic variation on the functional

30 haplotype. Our results show that many disease-associated recessive variants can produce mitigated
31 phenotypes in heterozygous carriers and motivate further work exploring penetrance mechanisms.

32

33 **Introduction**

34

35 Since the advent of next generation sequencing, the number of variants identified as contributing to
36 Mendelian disease has grown rapidly¹. Roughly 20% of all protein-coding genes in humans have been
37 associated with at least one Mendelian disease². Increasingly, studies of recessive disease variants
38 have begun observing that these variants can sometimes cause mitigated phenotypes in heterozygous
39 carriers, thereby contributing to population variation in complex traits and disease susceptibility³⁻¹¹.
40 However, the rarity of most such variants together with their unavailability in most SNP-array-based
41 genotyping studies has limited attempts to explore this phenomenon at scale. Early work focused on
42 smaller cohorts recruited for specific diseases, such as a series of studies that demonstrated increased
43 risk of male infertility^{12,13}, bronchiectasis¹⁴⁻¹⁶, and asthma^{15,17} among other phenotypes in cystic fibrosis
44 carriers. More recently, larger data sets have enabled extending the breadth of such analyses to more
45 phenotypes^{4,5} and to more recessive disease variants³.

46

47 With increasing exome sequencing of population biobank cohorts¹⁸⁻²⁰, a new opportunity to search for
48 carrier phenotypes in a phenome-wide, exome-wide manner has emerged. Furthermore, biobank data
49 sets present an opportunity to ameliorate ascertainment biases by assessing phenotypes in population
50 cohorts, complementing analyses of patients and their families. Family-based studies have been
51 observed to be susceptible to ascertainment biases that inflate observed effects²¹, while the opposite
52 “healthy volunteer” phenomenon has been observed in biobank cohorts²². Genome-wide genotyping
53 and imputation in biobank data sets also provide opportunities to investigate potential genetic modifiers
54 of incompletely penetrant carrier phenotypes²³.

55

56 Here we leveraged exome-wide imputation within the UK Biobank cohort²⁴ to power a broad
57 investigation of quantitative and disease phenotypes amongst carriers of recessive disease variants.
58 Next, we performed a focused analysis of two relatively more common severe recessive Mendelian
59 diseases, using the power afforded by high carrier frequencies to characterize carrier phenotypes or
60 establish a truly recessive pattern of phenotypes. Finally, we considered the molecular mechanisms
61 underlying incomplete penetrance observed amongst carriers, evaluating a previously proposed model
62 of modified penetrance²³.

63

64 **Subjects and methods**

65

66 **Imputed carriers of recessive Mendelian disease variants in UK Biobank**

67 We previously used the first tranche of whole exome sequencing data released by the UK Biobank
68 ($N=49,960$)¹⁸ to impute coding variants into SNP-array data available for $N=487,409$ participants in the
69 full UK Biobank cohort²⁵, achieving accurate imputation of rare variant genotypes at minor allele
70 frequencies (MAF) down to ~ 0.00005 ²⁴. Here, we analyzed a subset of imputed variants that were
71 annotated in ClinVar²⁶ as “pathogenic” or “likely pathogenic” for diseases annotated in OMIM² as
72 “autosomal recessive”. We further restricted to rare variants ($MAF < 0.01$) with a minimum MAF of
73 0.00001 and estimated imputation accuracy of $R^2 > 0.5$, leaving 3,481 variants for analysis.

74

75 **Association tests with quantitative traits**

76 We tested imputed genotype dosages for association with 57 quantitative traits using linear mixed
77 models implemented in BOLT-LMM v2.3.4. These traits included the 54 quantitative traits we previously
78 analyzed²⁴ and three additional traits (skin pigmentation, tanning ability, and hair color). We performed
79 quantitative trait association analyses on $N=459,327$ UK Biobank participants who reported European

80 ancestry and had not withdrawn from the study at the time of analysis. We did not attempt to filter
81 homozygotes or compound heterozygotes from these analyses, reasoning that such individuals would
82 account for negligible numbers of carriers of the rare variants we analyzed (both based on allele
83 frequencies and on the “healthy volunteer” ascertainment bias of UK Biobank).

84

85 **Association tests with binary traits**

86 We tested the same imputed variants for association with 1,139 binary disease phenotypes curated by
87 UK Biobank. These consisted of the complete set of “first-occurrence” of disease traits in the UK
88 Biobank converted to simple case and control status as well as the set of 8 “algorithmically defined
89 health outcomes” disease categories provided by the UK Biobank. We tested variants for association
90 with binary traits using the BinomiRare test²⁷ to obtain p -values robust to case-control imbalance while
91 adjusting for age (stratified into five-year tranches) and sex. For computational efficiency, we re-
92 implemented the BinomiRare test and applied a binomial approximation when the number of observed
93 cases among carriers exceeded 100. We estimated odds ratios as xw / yz , where x , y , z , w denote
94 ratios of observed versus expected cases among carriers, cases among noncarriers, controls among
95 carriers, and controls among noncarriers, respectively. We estimated 95% confidence intervals using a
96 normal approximation, i.e., converting p -values to z -scores and then taking the 95% CI of the log odds
97 ratio (OR) to be $\log(\text{OR}) \pm 1.96 * \log\text{OR} / z$. We performed association analyses on an unrelated
98 subset of $N=415,291$ UK Biobank participants who reported European ancestry and had not withdrawn
99 from the study²⁸.

100

101 **Analyses of cystic fibrosis carriers**

102 We identified cystic fibrosis carriers in UK Biobank using SNP-array genotypes for the Phe508del
103 variant and (in auxiliary analyses) the missense SNP rs78655421, excluding participants with a cystic
104 fibrosis report (according to the “first occurrences” data field). We applied the same analysis pipeline as

105 above to test for associations with the 1,139 binary traits and applied a significance threshold of FDR <
106 5% (q -value < 0.05).

107

108 **Analyses of spinal muscular atrophy carriers**

109 We identified spinal muscular atrophy carriers in the UK Biobank $N=200K$ exome sequencing release
110 as individuals with evidence of only one functional copy of *SMN1*. We estimated the number of
111 functional copies of each of *SMN1* and *SMN2* based on depth of coverage of exome sequencing reads
112 that mapped uniquely to the exon 7-intron 7 region of each gene (chr5:70,951,800-70,952,600 for
113 *SMN1* and chr5:70,076,400-70,077,100 for *SMN2* in hg38 coordinates; these regions contain four
114 paralogous sequence variants that distinguish the highly homologous genes and were captured by
115 exome sequencing). This approach accounted for deletions of exons 7-8 that commonly inactivate
116 copies of *SMN2* and occasionally *SMN1*²⁹. We computed exome sequencing read-depth using
117 mosdepth v0.3.1³⁰ and normalized each sample's read-depth measurements against corresponding
118 measurements from other samples with similar exome-wide sequencing depth profiles using a pipeline
119 we recently described²⁸.

120

121 We analyzed SMA carriers for evidence of changes in three traits related to neuromuscular function:
122 walking speed, hand grip strength (maximum of left- and right-hand measurements), and FEV1 / FVC
123 ratio (a measure of lung function). Using age, age squared, and sex as covariates, we performed linear
124 regressions to test for an association between SMA carrier status and each trait.

125

126 **Testing a model of modified penetrance in carriers of loss-of-function variants**

127 To further investigate potential molecular mechanisms underlying why some recessive variant carriers
128 display mild phenotypes but others do not, we considered a model of modified penetrance proposed by
129 Castel *et al.* (2018)²³. This model proposes that the penetrance of a deleterious variant can be affected
130 by variants on the allele on the homologous chromosome, particularly in the case of common *cis*-

131 eQTLs that modulate expression of the functional copy of the gene. To evaluate this model, we
132 analyzed heterozygous carriers of relatively common disease variants in two genes, *FLG* and *CFTR*.
133 To perform association tests on variants carried on the haplotypes opposite the disease variants, we
134 imputed variants from the Haplotype Reference Consortium panel (r1.1) using Minimac3 v2.0.1 (run on
135 genomic windows including 3 Mb flanks of each gene) and analyzed these variants together with the
136 variants we previously imputed from whole-exome sequencing²⁴. Next, we extracted the genotypes for
137 these carriers at variants within 1 Mb up- and downstream of each gene. We then recoded the
138 genotypes for each carrier to be hemizygous for the alleles sitting on the haplotype opposite from the
139 deleterious variant. We performed association tests on these recoded hemizygous variants using the
140 Fisher's exact test implemented in plink (v1.9)³¹ (--fisher-midp) (which could perform this analysis after
141 we recoded the chromosome as "X" and coded all individuals as male). We assessed the power of
142 these analyses to detect associations between common variants on the opposite haplotype using the
143 wp.logistic function in the WebPower R package.

144

145 **Results**

146

147 **Quantitative phenotypes in carriers of recessive disease variants**

148 Testing 3,481 rare recessive disease variants for association with 57 quantitative traits measured in the
149 UK Biobank identified 103 significant ($p < 2.52 \times 10^{-7}$; Bonferroni-corrected) variant-trait associations
150 (**Fig. 1** and **Table S1**). These associations involved variants reported to be pathogenic for 35 distinct
151 recessive diseases. For many of these diseases (19/35 diseases), carriers exhibited significant
152 deviations in multiple quantitative traits. Some of these multiple associations partly reflected correlated
153 measurements of blood, lipid, or pigmentation traits, such as associations between a variant believed to
154 cause Bernard-Soulier syndrome type C (a recessive bleeding disorder) and mean platelet volume
155 (0.68 ± 0.03 SD), platelet distribution width (0.54 ± 0.03 SD), and platelet count (-0.65 ± 0.03 SD).

156 However, others pointed to distinct manifestations of pleiotropy such as associations of a variant for
157 McArdle disease (a recessive glycogen storage disorder that interferes with muscle function) with both
158 increased urate levels (0.15 ± 0.02 SD) and increased waist-hip ratio (0.10 ± 0.02 SD).

159

160 For some of the diseases, carriers exhibited traits that might be expected based on the known
161 biological mechanisms of the disease, supporting the validity of our analytical approach. For example,
162 when considering Mendelian disorders where the production of a particular protein or compound is
163 altered, one might expect a carrier to have reduced levels of that same molecule. We observed this
164 phenomenon with infantile hypophosphatasia, which is defined by errors in alkaline phosphatase³². In
165 UK Biobank, carriers of several variants in the *ALPL* gene reported as pathogenic for recessive
166 hypophosphatasia exhibited decreased alkaline phosphatase (ranging from -2.63 ± 0.20 SD to -
167 0.71 ± 0.12 SD) and increased phosphate, as might be expected. Another example involved two
168 variants in *ANGPTL3* that have been implicated in hypobetalipoproteinemia 2, a recessive disorder in
169 which individuals experience low levels of several lipid biomarkers³³. Carriers showed decreases in
170 apolipoprotein A levels (-0.56 ± 0.04 SD; -0.63 ± 0.07 SD), cholesterol levels (-0.52 ± 0.04 SD; -
171 0.52 ± 0.07 SD), and triglyceride levels (-0.67 ± 0.04 SD; -0.52 ± 0.07 SD).

172

173 Other diseases with more complex biological mechanisms yielded less straightforward carrier
174 phenotypes. Here we highlight three such examples. First, a missense variant in *POR* implicated in
175 Antley-Bixler syndrome³⁴, a recessive skeletal disorder in which bones fuse prematurely, associated
176 with a 0.27 ± 0.04 SD increase in height (i.e., 1.76 ± 0.27 cm). Second, a frameshift variant in
177 *ADAMTSL4* implicated in ectopia lentis 2, a recessive disorder of the fibers in the eyes that can lead to
178 vision problems, associated with a 0.13 ± 0.02 SD (0.84 ± 0.12 cm) decrease in height. Decreased height
179 has been observed in patients with ectopia lentis 2, but the mechanism by which *ADAMTSL4* causes
180 this change has not been extensively examined³⁵. Third, a missense variant in *ABCA3* implicated in

181 pulmonary surfactant metabolism dysfunction 3, a recessive interstitial lung disease caused by
182 disruptions in the surface tension of lung surfactant³⁶, associated with a 0.12 ± 0.01 SD decrease in
183 FEV1/FVC ratio, a measure of lung function. These examples add to the growing body of evidence that
184 rare variants that cause severe disease in homozygotes or compound heterozygotes can often produce
185 mild, subclinical phenotypes in heterozygous carriers^{3,4}.

186

187 **Disease phenotypes in heterozygous carriers of recessive variants**

188 We next tested the same set of 3,481 rare recessive disease variants for association with 1,139 binary
189 traits in UK Biobank, identifying five associations that reached significance ($p < 1.26 \times 10^{-8}$; Bonferroni-
190 corrected) (**Table 1**). As with the quantitative traits, some associations were expected from previous
191 literature. Carriers of a frameshift variant in *HBB* exhibited increased risk of thalassemia³⁷, and carriers
192 of a stop gain variant in *COL4A4* implicated in Alport syndrome 2, a recessive disorder that involves
193 kidney dysfunction, exhibited increased risk of hematuria (OR=10.5; 95% CI, 5.2-21.2)^{38,39}, as we and
194 others have recently reported^{24,40,41}. Carriers of a missense variant in *TYR* (tyrosinase) implicated in
195 recessive oculocutaneous albinism type IA exhibited increased risk of disorders of aromatic amino-acid
196 metabolism (OR=63.3; 95% CI, 16.3-245.1)⁴². A missense variant in *TG* (thyroglobulin) implicated in
197 recessive thyroid dyshormonogenesis 3 increased risk of hypothyroidism in carriers (OR=2.20; 95% CI,
198 1.68-2.88)^{43,44}.

199

200 A more intriguing association involved a splice donor variant in *IFT140* previously implicated in
201 recessive short-rib thoracic dysplasia 9 and retinitis pigmentosa 80, often with accompanying renal
202 disease^{45,46}. Carriers of this variant exhibited increased risk of cystic kidney disease (OR=18.4; 95% CI,
203 8.5-40.1), corroborating recent findings from analyses of directly-sequenced individuals and imputation
204 using the TOPMed reference panel^{47,48}. Loss-of-function of both copies of *IFT140* appears to be
205 inviable based on murine studies⁴⁹, such that this canonical splice variant has been observed in cases
206 of recessive disease only in compound heterozygotes with partial function of the other copy of the

207 gene⁵⁰. While retinitis pigmentosa 80 primarily manifests in visual symptoms and recessive short-rib
208 thoracic dysplasia 9 in skeletal symptoms, *IFT140* encodes a protein related to cilia function that also is
209 expressed in the kidney, and renal symptoms have been noted in both diseases. The observed
210 association between carriers of the splice variant and cystic kidney disease suggests partial
211 haploinsufficiency of *IFT140* in its role in the kidney.

212

213 **Contrasting recessiveness of cystic fibrosis and spinal muscular atrophy**

214 In light of the diversity of autosomal recessive Mendelian diseases for which we observed mitigated
215 phenotypes in carriers, we decided to more closely investigate two relatively common recessive
216 Mendelian diseases to ask whether mitigated phenotypes were a ubiquitous feature of recessive
217 disease carriers. To do so, we identified diseases with sufficiently high carrier frequencies in UK
218 Biobank that we would be well-powered to identify mitigated carrier phenotypes or lack thereof. The two
219 diseases we identified based on these criteria were cystic fibrosis (CF) and spinal muscular atrophy
220 (SMA).

221

222 Previous studies have identified mitigated phenotypes in carriers of cystic fibrosis variants related to
223 phenotypic manifestations of the disease^{4,5}. To further explore the extent of this phenomenon utilizing
224 the deep phenotyping of UK Biobank, we tested our full set of quantitative and binary traits for
225 associations with carriers of the most common CF mutation, *CFTR* Phe508del (MAF=1.6%), which was
226 directly genotyped by UK Biobank SNP-arrays. Carriers of this variant showed significant associations
227 (q -value < 0.05) with asthma (OR=1.12; 95% CI,1.06-1.17), aspergillosis (OR=2.60; 95% CI,1.63-4.13),
228 bronchiectasis (OR=1.40; 95% CI,1.20-1.61), and duodenal ulcer (OR=1.30; 95% CI,1.15-1.45) (**Fig.**
229 **2a** and **Table S2**). Four additional associations reached significance at a relaxed FDR threshold of
230 10%: COPD (OR=1.17; 95% CI 1.07-1.27), cholelithiasis (OR=1.13; 95% CI 1.06-1.22), male infertility
231 (OR=2.10; 95% CI,1.40-3.15), and other prostate disorders (OR=1.39; 95% CI 1.15-1.67) (**Table S2**).

232 We also tested carriers of the next most common cystic fibrosis mutation, *CFTR* Arg117His
233 (MAF=0.2% in UK Biobank) but concluded that power was insufficient (**Table S2**).

234
235 The odds ratios we calculated for carriers of Phe508del, while significant, were much smaller than
236 those recently reported in an analysis of CF carriers ascertained from a database of insurance claims
237 from individuals who had been tested for carrier status⁴ (**Fig. S1**). Furthermore, several reported
238 associations did not replicate in our analysis of UK Biobank. For example, whereas the claims analysis
239 showed a strong association between carrier status and short stature⁴, we did not observe an
240 association between Phe508del carrier status and height in UK Biobank despite ample power (effect
241 size = -0.000 ± 0.006 SD). The odds ratios we computed were more consistent with those reported by
242 Çolak et al. (2020) using Phe508del genotyping in the Copenhagen General Population Study⁵. These
243 results underscore the importance of understanding issues of ascertainment bias when studying
244 penetrance in population studies^{51–53}.

245
246 In contrast to CF, potential phenotypes of SMA carriers have not (to our knowledge) previously been
247 explored, in part because of the difficulty of genotyping SMA carrier mutations, most of which arise from
248 structural variation at the *SMN1–SMN2* locus^{29,54}. SMA is usually caused by loss-of-function mutations
249 in both copies of the *SMN1* gene, with disease severity then determined by the number of functional
250 copies of the paralogous *SMN2* gene. The availability of whole-exome sequencing (WES) data for
251 $N \sim 200K$ UK Biobank participants¹⁹ enabled us to estimate the number of functional copies of *SMN1*
252 and *SMN2* in each sequenced sample from WES depth-of-coverage (**Fig. 2b**). We ascertained 3,462
253 SMA carriers (i.e., individuals likely to carry only one functional copy of *SMN1*) in this way from the set
254 whole-exome sequenced individuals of European-ancestry ($N=187,720$). Interestingly, we found no
255 significant associations between SMA carrier status and potential manifestations of muscle weakness –
256 walking speed, grip strength, and FEV1/FVC ratio (**Fig. 2c**) – even when stratifying for *SMN2* copy
257 number (**Table S3**). These results suggest that SMA is a truly recessive disease in which muscle

258 weakness phenotypes only manifest in individuals who carry two *SMN1* alleles inactivated by loss-of-
259 function variants.

260

261 **Testing a model of modified penetrance**

262 In all the instances in which we observed mitigated phenotypes in carriers of recessive disease
263 variants, the associated phenotypes exhibited variable penetrance in heterozygotes. We therefore
264 sought to explore the possible molecular mechanisms underlying this incomplete penetrance.

265

266 Castel *et al.* (2018) previously proposed a model of modified penetrance in which the haplotype
267 arrangement of loss-of-function and expression-modifying variants in an individual might affect overall
268 phenotype (**Fig. 3a**)²³. In this model, the phenotypic impact of a deleterious variant inactivating one
269 copy of a gene is mediated by the amount of expression of the functional copy (on the opposite
270 haplotype), such that a common *cis*-eQTL influencing expression of the functional allele can influence
271 the severity of the phenotype. Explicitly, if the *cis*-eQTL increases expression of the functional, wildtype
272 protein, this could partially ameliorate the loss of the other copy; in contrast, if the *cis*-eQTL decreases
273 expression of the functional copy, one might expect the carrier to have a more severe phenotype.

274

275 To explore this hypothesis, we considered two genes, *FLG* and *CFTR*, in which variants known to
276 cause both recessive disease and produce mitigated phenotypes in carriers are sufficiently common to
277 power analysis. Loss-of-function variants in *FLG* are known to cause ichthyosis vulgaris in
278 homozygotes or compound heterozygotes, and carrier status has been associated with asthma and
279 atopic dermatitis^{55,56}. In UK Biobank, 10.3% of participants carried a loss-of-function variant in *FLG* that
280 was associated with asthma or atopic dermatitis in heterozygotes. As discussed in the previous section,
281 mutations in *CFTR* are responsible for cystic fibrosis, as well as several mitigated phenotypes in
282 carriers. Approximately 3.1% of individuals in the UK Biobank are carriers for the Phe508del mutation in
283 *CFTR* that we considered for this analysis.

284

285 To determine whether variants on the opposite (putatively functional) haplotype in carriers might affect
286 their susceptibility to mitigated phenotypes, we restricted our analysis just to carriers of these
287 deleterious variants (**Fig. 3b**). For each nearby variant at each locus, we then ran an association test
288 between opposite-haplotype genotypes and mitigated phenotypes. No tested variant at either locus
289 significantly associated with phenotype (**Fig. 3c**). Given that we were well-powered to detect common
290 variant associations with an odds ratio >1.2 in both scenarios (**Fig. 3d**), these results suggest that the
291 modified penetrance model is unlikely to underlie incomplete penetrance of these carrier phenotypes.

292

293 **Discussion**

294

295 Our results demonstrate that for a wide range of Mendelian diseases, variants traditionally considered
296 to be recessive can cause milder phenotypes in heterozygous carriers. We also observed that entirely
297 recessive effects do exist: heterozygous carriers for spinal muscular atrophy exhibited no evidence of
298 even a subtle effect on phenotypes related to muscle strength. These observations suggest a spectrum
299 of recessiveness that is now becoming visible in very large population cohorts.

300

301 Our study did have several limitations. First, even with the large sample size provided by exome
302 sequencing in UK Biobank, we still lacked power to assess potential effects of many very rare variants
303 that are known to cause Mendelian recessive diseases. Second, our examination of potential
304 interactions between variants on opposite haplotypes was even more power-constrained, such that we
305 could only assess this model for two diseases involving common variants. Third, the effects we
306 estimated are likely to be influenced by the “healthy volunteer” ascertainment bias observed in analyses
307 of population biobank cohorts²².

308

309 As even larger, well-phenotyped cohorts with whole-exome or whole-genome sequencing become
310 available, our ability to determine the extent of mild carrier phenotypes will increase. More
311 comprehensive phenome-wide and genome-wide studies will allow for an assessment of how common
312 the phenomenon of incomplete recessivity is amongst severe Mendelian diseases and the spectrum of
313 phenotypes that can manifest. Moreover, the higher power afforded by extremely large studies will also
314 enable more extensive exploration of potential interactions between variants that could help to explain
315 incomplete penetrance and shed light on the molecular mechanisms that underlie mitigated
316 phenotypes.
317

318 **Supplemental data** include one figure and three tables.

319

320 **Acknowledgments**

321 We thank A. Gusev, A. Price and S. Sunyaev for helpful discussions. This research was conducted
322 using the UK Biobank Resource under application no. 10438. A.R.B. was supported by US NIH grant
323 T32 HG229516 and fellowship F31 HL154537. M.L.A.H. was supported by US NIH fellowship F32
324 HL160061. M.A.S. was supported by the MIT John W. Jarve (1978) Seed Fund for Science Innovation
325 and US NIH fellowship F31 MH124393. R.E.M. was supported by US NIH grant K25 HL150334 and
326 NSF grant DMS-1939015. P.-R.L. was supported by US NIH grant DP2 ES030554, a Burroughs
327 Wellcome Fund Career Award at the Scientific Interfaces, the Next Generation Fund at the Broad
328 Institute of MIT and Harvard, and a Sloan Research Fellowship. The funders had no role in study
329 design, data collection and analysis, decision to publish or preparation of the manuscript.
330 Computational analyses were performed on the O2 High Performance Compute Cluster, supported by
331 the Research Computing Group, at Harvard Medical School (<http://rc.hms.harvard.edu>).

332

333 **Declaration of interests**

334 The authors declare no competing interests.

335

336 **Data and code availability**

337 Access to the UK Biobank Resource is available by application (<http://www.ukbiobank.ac.uk/>). BOLT-
338 LMM (v2.3.4) is available at <https://data.broadinstitute.org/alkesgroup/BOLT-LMM/>. mosdepth (v0.3.1)
339 is available at <https://github.com/brentp/mosdepth>. Minimac4 (v.1.0.1) is available at
340 <https://genome.sph.umich.edu/wiki/Minimac4>. plink (v1.9) is available from [https://www.cog-
342 genomics.org/plink1.9/](https://www.cog-
341 genomics.org/plink1.9/).

342

343

344 **Web resources**

345 ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>

346 OMIM, <https://omim.org/>

347

348 **References**

349 1. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure,
350 J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* *12*, 745–
351 755.

352 2. Hamosh, A., Amberger, J.S., Bocchini, C., Scott, A.F., and Rasmussen, S.A. (2021). Online
353 Mendelian Inheritance in Man (OMIM®): Victor McKusick’s magnum opus. *Am. J. Med. Genet. A*
354 *185*, 3259–3265.

355 3. Wright, C.F., West, B., Tuke, M., Jones, S.E., Patel, K., Laver, T.W., Beaumont, R.N., Tyrrell, J.,
356 Wood, A.R., Frayling, T.M., et al. (2019). Assessing the Pathogenicity, Penetrance, and Expressivity of
357 Putative Disease-Causing Variants in a Population Setting. *Am. J. Hum. Genet.* *104*, 275–286.

358 4. Miller, A.C., Comellas, A.P., Hornick, D.B., Stoltz, D.A., Cavanaugh, J.E., Gerke, A.K., Welsh, M.J.,
359 Zabner, J., and Polgreen, P.M. (2020). Cystic fibrosis carriers are at increased risk for a wide range of
360 cystic fibrosis-related conditions. *Proc. Natl. Acad. Sci. U.S.A.* *117*, 1621–1627.

361 5. Çolak, Y., Nordestgaard, B.G., and Afzal, S. (2020). Morbidity and mortality in carriers of the cystic
362 fibrosis mutation *CFTR* Phe508del in the general population. *Eur. Respir. J.* *56*, 2000558.

363 6. Mäkitie, O., Pereira, R.C., Kaitila, I., Turan, S., Bastepe, M., Laine, T., Kröger, H., Cole, W.G., and
364 Jüppner, H. (2010). Long-term clinical outcome and carrier phenotype in autosomal recessive
365 hypophosphatemia caused by a novel *DMP1* mutation. *J. Bone Miner. Res.* *25*, 2165–2174.

366 7. Dagher, H., Buzza, M., Colville, D., Jones, C., Powell, H., Fassett, R., Wilson, D., Agar, J., and
367 Savige, J. (2001). A comparison of the clinical, histopathologic, and ultrastructural phenotypes in
368 carriers of X-linked and autosomal recessive Alport’s syndrome. *Am. J. Kidney Dis.* *38*, 1217–1228.

369 8. Watts, J.A., Morley, M., Burdick, J.T., Fiori, J.L., Ewens, W.J., Spielman, R.S., and Cheung, V.G.
370 (2002). Gene expression phenotype in heterozygous carriers of ataxia telangiectasia. *Am. J. Hum.*
371 *Genet.* *71*, 791–800.

372 9. Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A.,
373 Berg, D., Bras, J., Brice, A., et al. (2009). Multicenter analysis of glucocerebrosidase mutations in
374 Parkinson’s disease. *N. Engl. J. Med.* *361*, 1651–1661.

375 10. Vieira, S.R.L., and Morris, H.R. (2021). Neurodegenerative Disease Risk in Carriers of Autosomal
376 Recessive Disease. *Front. Neurol.* *12*, 679927–679927.

- 377 11. Perez, Y., Wormser, O., Sadaka, Y., Birk, R., Narkis, G., and Birk, O.S. (2017). A Rare Variant in
378 *PGAP2* Causes Autosomal Recessive Hyperphosphatasia with Mental Retardation Syndrome, with a
379 Mild Phenotype in Heterozygous Carriers. *Biomed. Res. Int.* 2017, 3470234.
- 380 12. Chillón, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., Romey, M.-C., Ruiz-
381 Romero, J., Verlingue, C., Claustres, M., et al. (1995). Mutations in the Cystic Fibrosis Gene in Patients
382 with Congenital Absence of the Vas Deferens. *N. Engl. J. Med.* 332, 1475–1480.
- 383 13. Yu, J., Chen, Z., Ni, Y., and Li, Z. (2012). *CFTR* mutations in men with congenital bilateral absence
384 of the vas deferens (CBAVD): a systemic review and meta-analysis. *Hum. Reprod.* 27, 25–35.
- 385 14. Girodon, E., Cazeneuve, C., Lebargy, F., Chinet, T., Costes, B., Ghanem, N., Martin, J., Lemay, S.,
386 Scheid, P., Housset, B., et al. (1997). *CFTR* gene mutations in adults with disseminated bronchiectasis.
387 *Eur. J. Hum. Genet.* 5, 149–155.
- 388 15. Tzetis, M., Efthymiadou, A., Strofalis, S., Psychou, P., Dimakou, A., Pouliou, E., Doudounakis, S.,
389 and Kanavakis, E. (2001). *CFTR* gene mutations – including three novel nucleotide substitutions – and
390 haplotype background in patients with asthma, disseminated bronchiectasis and chronic obstructive
391 pulmonary disease. *Hum. Genet.* 108, 216–221.
- 392 16. Casals, T., De-Gracia, J., Gallego, M., Dorca, J., Rodríguez-Sanchón, B., Ramos, M., Giménez, J.,
393 Cisteró-Bahima, A., Oliveira, C., and Estivill, X. (2004). Bronchiectasis in adult patients: an expression
394 of heterozygosity for *CFTR* gene mutations? *Clin. Genet.* 65, 490–495.
- 395 17. Nielsen, A.O., Qayum, S., Bouchelouche, P.N., Laursen, L.C., Dahl, R., and Dahl, M. (2016). Risk
396 of asthma in heterozygous carriers for cystic fibrosis: A meta-analysis. *J. Cyst. Fibros.* 15, 563–567.
- 397 18. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey, A.K., Gonzaga-
398 Jauregui, C., Khalid, S., Ye, B., Banerjee, N., et al. (2020). Exome sequencing and characterization of
399 49,960 individuals in the UK Biobank. *Nature* 586, 749–756.
- 400 19. Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G., Sasson, A., Wong,
401 E., Liu, D., Wade Davis, J., Haefliger, C., et al. (2021). Advancing human genetics research and drug
402 discovery through exome sequencing of the UK Biobank. *Nat. Genet.* 53, 942–948.
- 403 20. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D.,
404 Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK
405 Biobank participants. *Nature* 599, 628–634.
- 406 21. Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013).
407 Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of
408 reduced penetrance in human inherited disease. *Hum. Genet.* 132, 1077–1130.
- 409 22. Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen,
410 N.E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank
411 Participants With Those of the General Population. *Am. J. Epidemiol.* 186, 1026–1034.

- 412 23. Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov,
413 I., Vasileva, A., and Lappalainen, T. (2018). Modified penetrance of coding variants by cis-regulatory
414 variation contributes to disease risk. *Nat. Genet.* *50*, 1327–1334.
- 415 24. Barton, A.R., Sherman, M.A., Mukamel, R.E., and Loh, P.-R. (2021). Whole-exome imputation
416 within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nat. Genet.* *53*,
417 1260–1269.
- 418 25. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D.,
419 Delaneau, O., O’Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and
420 genomic data. *Nature* *562*, 203–209.
- 421 26. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J.,
422 Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting
423 evidence. *Nucleic Acids Res.* *46*, D1062–D1067.
- 424 27. Sofer, T. (2017). BinomiRare: a robust test of the association of a rare-variant with a disease for
425 pooled and meta-analysis, with application to the HCHS/SOL. *Genet. Epidemiol.* *41*, 388–395.
- 426 28. Mukamel, R.E., Handsaker, R.E., Sherman, M.A., Barton, A.R., Zheng, Y., McCarroll, S.A., and
427 Loh, P.-R. (2021). Protein-coding repeat polymorphisms strongly shape diverse human phenotypes.
428 *Science* *373*, 1499–1505.
- 429 29. Chen, X., Sanchis-Juan, A., French, C.E., Connell, A.J., Delon, I., Kingsbury, Z., Chawla, A.,
430 Halpern, A.L., Taft, R.J., Bentley, D.R., et al. (2020). Spinal muscular atrophy diagnosis and carrier
431 screening from genome sequencing data. *Genet. Med.* *22*, 945–953.
- 432 30. Brent S. Pedersen, and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes
433 and exomes. *Bioinformatics* *34*, Pages 867-868.
- 434 31. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-
435 generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* *4*, s13742-015-
436 0047–0048.
- 437 32. Weiss, M.J., Cole, D.E., Ray, K., Whyte, M.P., Lafferty, M.A., Mulivor, R.A., and Harris, H.
438 (1988). A missense mutation in the human liver/bone/kidney alkaline phosphatase gene causing a lethal
439 form of hypophosphatasia. *Proc. Natl. Acad. Sci. U.S.A.* *85*, 7666–7669.
- 440 33. Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V.,
441 Fisher, S., Abreu, J., Barry, A.J., et al. (2010). Exome Sequencing, *ANGPTL3* Mutations, and Familial
442 Combined Hypolipidemia. *N. Engl. J. Med.* *363*, 2220–2227.
- 443 34. Flück, C.E., Tajima, T., Pandey, A.V., Arlt, W., Okuhara, K., Verge, C.F., Jabs, E.W., Mendonça,
444 B.B., Fujieda, K., and Miller, W.L. (2004). Mutant P450 oxidoreductase causes disordered
445 steroidogenesis with and without Antley-Bixler syndrome. *Nat. Genet.* *36*, 228–230.

- 446 35. Neuhann, T.M., Stegerer, A., Riess, A., Blair, E., Martin, T., Wieser, S., Kläs, R., Bouman, A.,
447 Kuechler, A., and Rittinger, O. (2015). *ADAMTSL4*-associated isolated ectopia lentis: Further patients,
448 novel mutations and a detailed phenotype description. *Am. J. Med. Genet. A* *167*, 2376–2381.
- 449 36. Shulenin, S., Noguee, L.M., Annilo, T., Wert, S.E., Whitsett, J.A., and Dean, M. (2004). *ABCA3* Gene
450 Mutations in Newborns with Fatal Surfactant Deficiency. *N. Engl. J. Med.* *350*, 1296–1303.
- 451 37. Taher, A.T., Musallam, K.M., and Cappellini, M.D. (2021). β -Thalassemias. *N. Engl. J. Med.* *384*,
452 727–743.
- 453 38. Longo, I., Porcedda, P., Mari, F., Giachino, D., Meloni, I., Deplano, C., Brusco, A., Bosio, M.,
454 Massella, L., Lavoratti, G., et al. (2002). *COL4A3/COL4A4* mutations: From familial hematuria to
455 autosomal-dominant or recessive Alport syndrome. *Kidney Int.* *61*, 1947–1956.
- 456 39. Buzza, M., Wang, Y.Y., Dagher, H., Babon, J.J., Cotton, R.G., Powell, H., Dowling, J., and Savige,
457 J. (2001). *COL4A4* mutation in thin basement membrane disease previously described in Alport
458 syndrome1. *Kidney Int.* *60*, 480–483.
- 459 40. Yang, C., Song, Y., Chen, Z., Yuan, X., Chen, X., Ding, G., Guan, Y., McGrath, M., Song, C., Tong,
460 Y., et al. (2019). A Nonsense Mutation in *COL4A4* Gene Causing Isolated Hematuria in Either
461 Heterozygous or Homozygous State. *Front. Genet.* *10*, 628.
- 462 41. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman,
463 G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine
464 biomarkers in the UK Biobank. *Nat. Genet.* *53*, 185–194.
- 465 42. Grønskov, K., Ek, J., and Brøndum-Nielsen, K. (2007). Oculocutaneous albinism. *Orphanet J. Rare*
466 *Dis.* *2*, 43.
- 467 43. Ieiri, T., Cochaux, P., Targovnik, H.M., Suzuki, M., Shimoda, S., Perret, J., and Vassart, G. (1991).
468 A 3' splice site mutation in the thyroglobulin gene responsible for congenital goiter with
469 hypothyroidism. *J. Clin. Invest.* *88*, 1901–1905.
- 470 44. Grasberger, H., and Refetoff, S. (2011). Genetic causes of congenital hypothyroidism due to
471 dysmorphogenesis. *Curr. Opin. Pediatr.* *23*, 421–428.
- 472 45. Perrault, I., Saunier, S., Hanein, S., Filhol, E., Bizet, A.A., Collins, F., Salih, M.A.M., Gerber, S.,
473 Delphin, N., Bigot, K., et al. (2012). Mainzer-Saldino syndrome is a ciliopathy caused by *IFT140*
474 mutations. *Am. J. Hum. Genet.* *90*, 864–870.
- 475 46. Schmidts, M., Frank, V., Eisenberger, T., al Turki, S., Bizet, A.A., Antony, D., Rix, S., Decker, C.,
476 Bachmann, N., Bald, M., et al. (2013). Combined NGS Approaches Identify Mutations in the
477 Intraflagellar Transport Gene *IFT140* in Skeletal Ciliopathies with Early Progressive Kidney Disease.
478 *Hum. Mutat.* *34*, 714–724.
- 479 47. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G.,
480 Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the
481 NHLBI TOPMed Program. *Nature* *590*, 290–299.

- 482 48. Senum, S.R., Li, Y. (Sabrina) M., Benson, K.A., Joli, G., Olinger, E., Lavu, S., Madsen, C.D.,
483 Gregory, A.V., Neatu, R., Kline, T.L., et al. (2021). Monoallelic *IFT140* pathogenic variants are an
484 important cause of the autosomal dominant polycystic kidney-spectrum phenotype. *Am. J. Hum. Genet.*
- 485 49. Jonassen, J.A., SanAgustin, J., Baker, S.P., and Pazour, G.J. (2012). Disruption of IFT Complex A
486 Causes Cystic Kidneys without Mitotic Spindle Misorientation. *J. Am. Soc. Nephrol.* 23, 641–651.
- 487 50. Online Mendelian Inheritance in Man, OMIM® (2019). MIM Number 614620. (Baltimore, MD:
488 Johns Hopkins University).
- 489 51. Ganguli, M., Lytle, M.E., Reynolds, M.D., and Dodge, H.H. (1998). Random versus volunteer
490 selection for a community-based study. *J. Gerontol. A Biol. Sci. Med. Sci.* 53, M39-46.
- 491 52. Mirshahi, U.L., Colclough, K., Wright, C.F., Wood, A.R., Beaumont, R.N., Tyrrell, J., Laver, T.W.,
492 Stahl, R., Golden, A., Goehringer, J.M., et al. (2021). The penetrance of age-related monogenic disease
493 depends on ascertainment context. *MedRxiv* 2021.06.28.21259641.
- 494 53. Forrest, I.S., Chaudhary, K., Vy, H.M.T., Bafna, S., Jordan, D.M., Rocheleau, G., Loos, R.J.F., Cho,
495 J.H., and Do, R. (2021). Ancestrally and Temporally Diverse Analysis of Penetrance of Clinical
496 Variants in 72,434 Individuals. *MedRxiv* 2021.03.11.21253430.
- 497 54. Alías, L., Bernal, S., Fuentes-Prior, P., Barceló, M.J., Also, E., Martínez-Hernández, R., Rodríguez-
498 Alvarez, F.J., Martín, Y., Aller, E., Grau, E., et al. (2009). Mutation update of spinal muscular atrophy in
499 Spain: molecular characterization of 745 unrelated patients and identification of four novel mutations in
500 the *SMN1* gene. *Hum. Genet.* 125, 29–39.
- 501 55. Weidinger, S., O’Sullivan, M., Illig, T., Baurecht, H., Depner, M., Rodriguez, E., Ruether, A.,
502 Klopp, N., Vogelberg, C., Weiland, S.K., et al. (2008). Filaggrin mutations, atopic eczema, hay fever,
503 and asthma in children. *J. Allergy Clin. Immunol.* 121, 1203-1209.e1.
- 504 56. Wang, Q., Dhindsa, R.S., Carss, K., Harper, A., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.,
505 Mackay, A., Muthas, D., et al. (2020). Surveying the contribution of rare variants to the genetic
506 architecture of human disease through exome sequencing of 177,882 UK Biobank participants. *BioRxiv*
507 2020.12.13.422582.

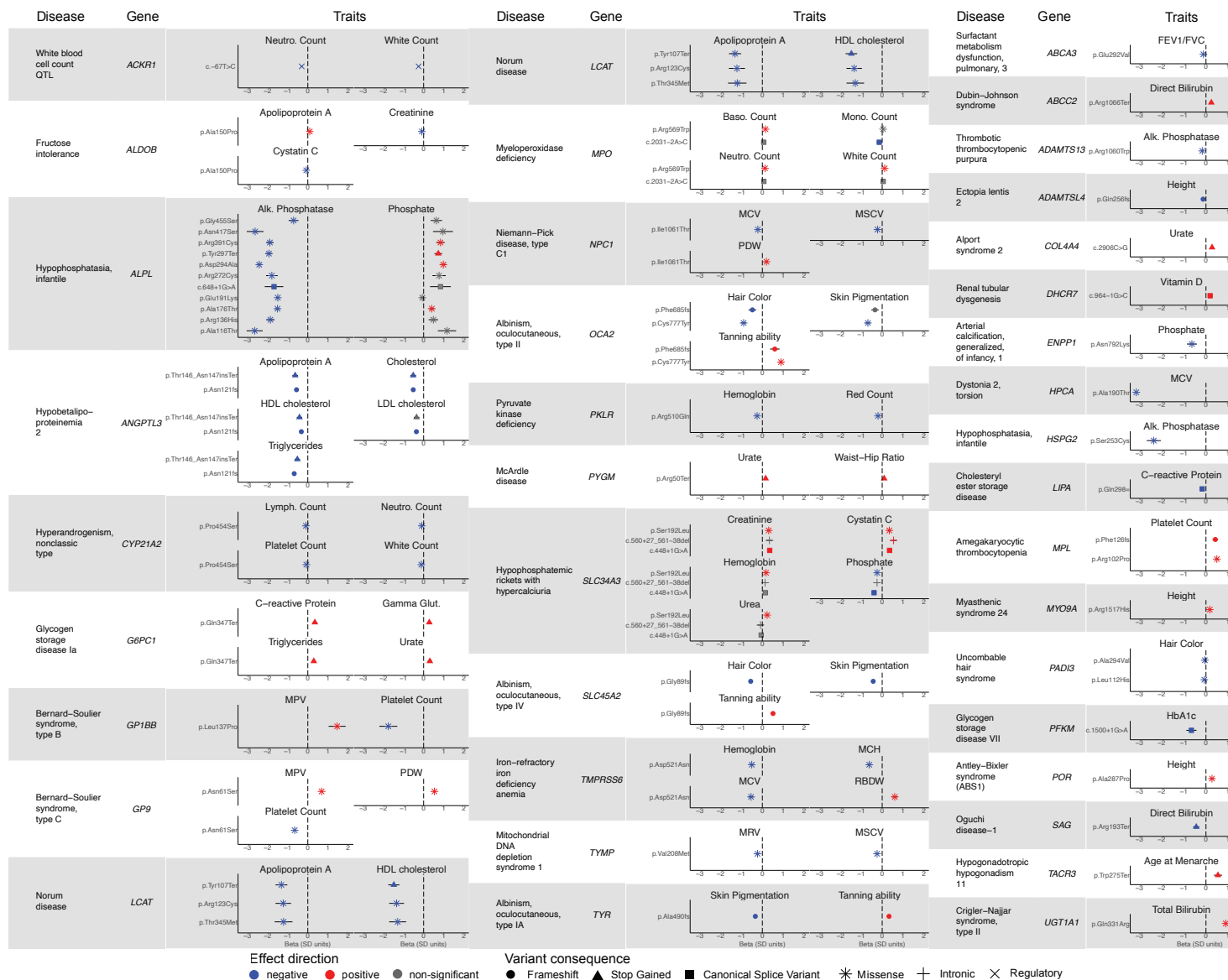
508

509 Figures

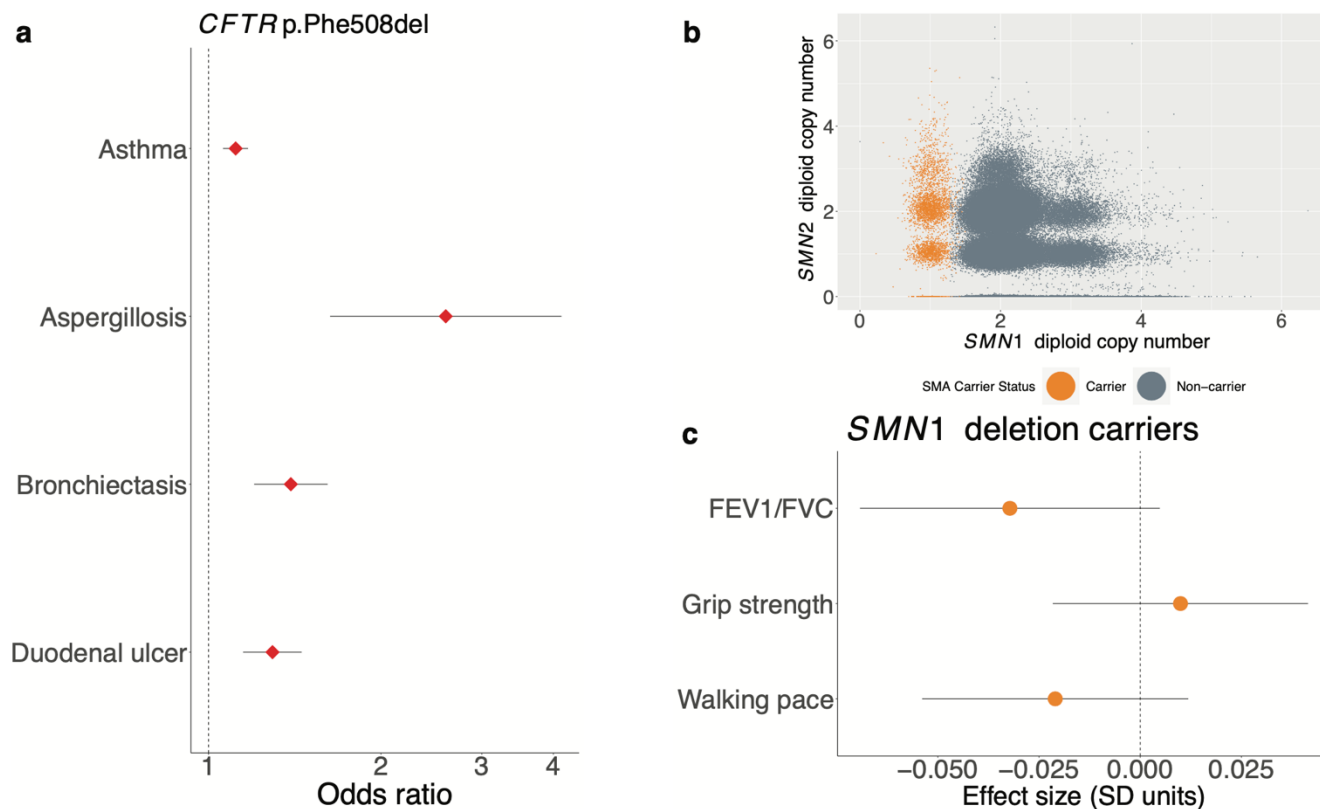
510

511 Figure 1. Carriers of recessive Mendelian disease variants display quantitative phenotypes.

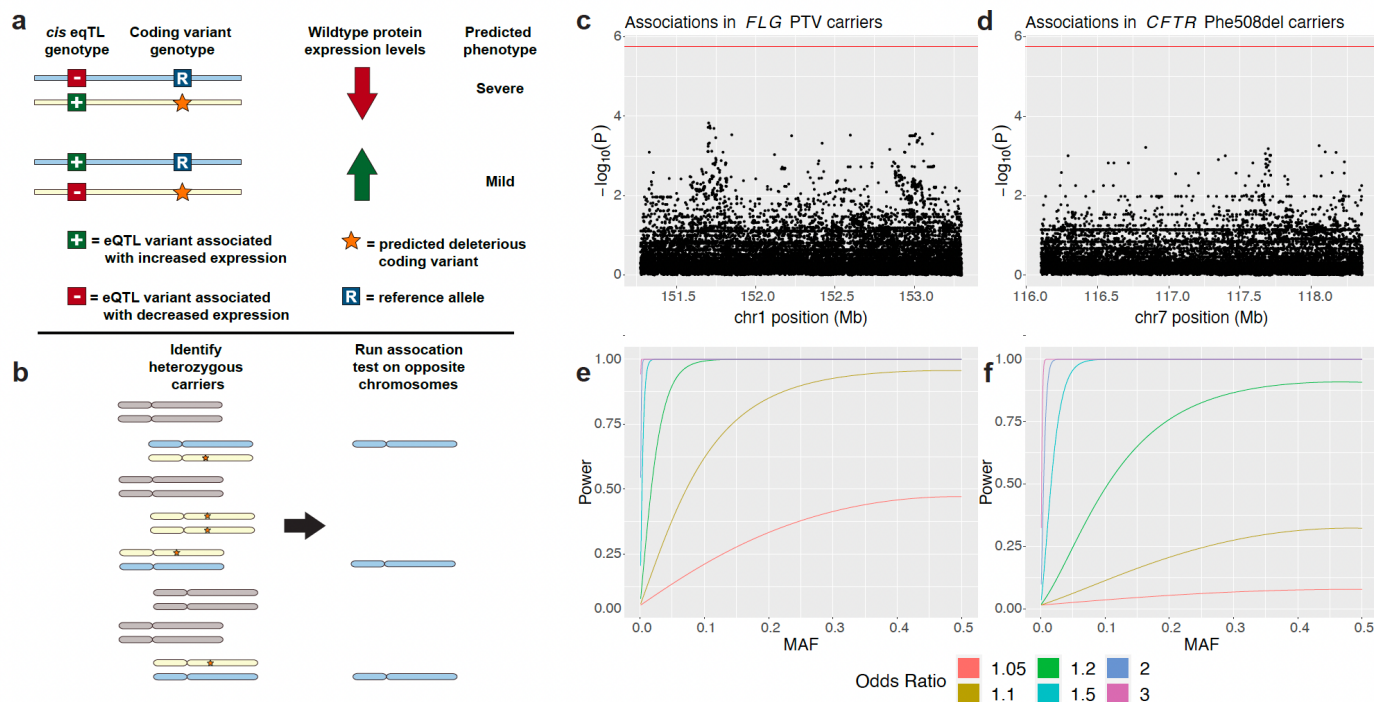
512 Mendelian diseases and their associated genes are listed to the left in each column, and the mean
 513 effect size is plotted on the right for each associated quantitative trait in units of standard deviation
 514 (error bars, 95% CIs). Positive-effect variants are shown in red, negative-effect variants in blue, and
 515 variants not Bonferroni-significant for one of the displayed traits in gray. Marker shapes correspond to
 516 effects on the gene and gene product as reported in ClinVar.



517 **Figure 2. Cystic fibrosis carriers show mitigated phenotypes, but spinal muscular atrophy**
518 **carriers do not.** (a) Carriers with the *CFTR* Phe508del inframe deletion exhibit increased risk of
519 several mitigated disease phenotypes (data points, odds ratios; error bars, 95% CIs). (b) Genotyping
520 for SMA carrier status using exome sequencing data from the UK Biobank. *SMN1* and *SMN2* copy
521 numbers were estimated based on sequencing read depth, and SMA carriers (orange) were identified
522 as those individuals estimated to have one functional copy of *SMN1* (with a deletion allele on the
523 homologous chromosome). (c) *SMN1* deletion carriers did not display evidence of changes in any of
524 three traits related to neuromuscular function (data points, mean values in units of standard deviations;
525 error bars, 95% CIs).
526



527 **Figure 3. Common variants on haplotypes opposite recessive alleles in carriers of two**
 528 **Mendelian diseases do not appear to modify penetrance of carrier phenotypes.** (a) The “modified
 529 penetrance” model of Castel *et al.* (2018) posits that *cis*-eQTLs can increase or decrease severity of a
 530 deleterious variant by modulating the quantity of functional protein produced by the opposite haplotype.
 531 (b) To test this hypothesis, we examined opposite haplotypes (blue chromosomes) in heterozygous
 532 carriers of recessive disease variants (orange star on yellow chromosome). We tested variants carried
 533 on these opposite haplotypes for association with mitigated phenotypes observed in carriers. (c,d)
 534 Manhattan plots showing association test results for variants on the opposite haplotype of deleterious
 535 variants in *FLG* and *CFTR* with asthma (the phenotype most strongly associated with carrier status in
 536 each case). No association reached Bonferroni significance (red line). (e,f) Power analyses for the tests
 537 conducted in c,d indicate that these tests were well-powered to detect common variant effects with
 538 odds ratios >1.2.
 539



540 **Tables**

541

542 **Table 1. Carriers of Mendelian recessive disease variants exhibit increased risk of less-severe**
 543 **disease phenotypes.** Odds ratios and p-values are reported for the five associations that reached
 544 Bonferroni significance.
 545

Recessive disease association	ClinVar reported variant				Disease association in carriers	Trait category	OR (95% CI)	P-Value
	Gene	Variant	Variant impact	MAF				
Alport syndrome 2	<i>COL4A4</i>	2:227917083:G:C	stop gained	6.16E-04	Recurrent and persistent haematuria (N02)	Genitourinary system disorders	10.47 (5.17-21.2)	6.75E-11
Thyroid dysmorphogenesis 3	<i>TG</i>	8:133894854:C:T	stop gained	6.38E-04	Other hypothyroidism (E03)	Endocrine, nutritional and metabolic diseases	2.20 (1.68-2.88)	9.72E-09
Sickle cell anemia	<i>HBB</i>	11:5248233:CAG:C	frameshift	2.15E-05	Thalassaemia (D56)	Blood, blood-forming organs and certain immune disorders	3,183 (440-23,030)	1.36E-15
Albinism, oculo-cutaneous, type IA	<i>TYR</i>	11:88961072:C:A	missense	1.10E-03	Disorders of aromatic amino-acid metabolism (E70)	Endocrine, nutritional and metabolic diseases	63.27 (16.33-245.10)	1.94E-09
Retinitis pigmentosa 80 & short rib thoracic dysplasia 9	<i>IFT140</i>	16:1607935:C:A	splice donor	5.11E-04	Cystic kidney disease (Q61)	Congenital disruptions and chromosomal abnormalities	18.42 (8.47-40.05)	1.98E-13