

1 **Title:** A phenome-wide association study of four syndromic genes reveals pleiotropic  
2 effects of common and rare variants in the general population.

3

4 **Authors:** Catherine Tcheandjieu<sup>1,2,4</sup>, Matthew Aguirre<sup>2,3,4</sup>, Stefan Gustafsson<sup>1,4</sup>, Priyanka  
5 Saha<sup>1,2,4</sup>, Praneetha Potiny<sup>1,2,4</sup>, Melissa Haendel<sup>6</sup>, Erik Ingelsson<sup>1,4,5</sup>, Manuel A. Rivas<sup>3,4</sup>,  
6 James R. Priest<sup>\*1,2,4,7</sup>

7 **Affiliations:**

8 <sup>1</sup>Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305; <sup>2</sup>Department  
9 of Pediatric Cardiology Stanford University School of Medicine, Stanford, CA 94305;  
10 <sup>3</sup>Department of Biomedical Data Science, Stanford University School of Medicine,  
11 Stanford, CA 94305; <sup>4</sup>Department of Medicine, Division of Cardiovascular Medicine,  
12 Stanford University School of Medicine, Stanford, CA 94305. <sup>5</sup>Department of Medicine,  
13 Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford,  
14 CA 94305, Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305,  
15 <sup>6</sup>Department of Medical Informatics and Clinical Epidemiology, School of Medicine,  
16 Oregon Health & Science University (OHSU), <sup>7</sup>Chan-Zuckerberg Biohub, San Francisco,  
17 California

18

19 \*Corresponding author

20 Email: [jpriest@stanford.edu](mailto:jpriest@stanford.edu)

21

## 22 **Abstract**

23 The clinical evaluation of a genetic syndrome relies upon recognition of a characteristic  
24 pattern of signs or symptoms to guide targeted genetic testing for confirmation of the  
25 diagnosis. However, individuals displaying a few phenotypes of a complex syndrome may  
26 not meet criteria for clinical diagnosis or genetic testing. Here, we present a phenome-  
27 wide association study (PheWAS) approach to systematically explore pleiotropy of  
28 common and rare alleles in genes associated with four well-described syndromic  
29 diseases (Alagille (AS), Marfan (MS), DiGeorge (DS), and Noonan (NS) syndromes) in  
30 the general population.

31 Using human phenotype ontology (HPO) terms, we systematically mapped 60  
32 phenotypes related to AS, MS, DS and NS in 337,198 unrelated white British from the UK  
33 Biobank (UKBB) based on their hospital admission records, self-administrated  
34 questionnaires, and physiological measurements. We performed logistic regression  
35 adjusting for age, sex, and the first 5 genetic principal components, for each phenotype  
36 and each variant in the target genes (*JAG1*, *TBX1*, *FBN1*, *PTPN11*, *NOTCH2*, and  
37 *MAP2K1*) and performed a gene burden testing.

38 Overall, we observed multiple phenotype-genotype correlations, such as the  
39 association between variation in *JAG1*, *FBN1*, *PTPN11* and *SOS2* with diastolic and  
40 systolic blood pressure; and pleiotropy among multiple variants in syndromic genes. For  
41 example, rs11066309 in *PTPN11* was significantly associated with a lower body mass  
42 index, an increased risk of hypothyroidism and a smaller size for gestational age, all in  
43 concordance with NS-related phenotypes. Similarly, rs589668 in *FBN1* was associated  
44 with an increase in body height and blood pressure, and a reduced body fat percentage  
45 as observed in Marfan syndrome.

46 Our findings suggest that the spectrum of associations of common and rare  
47 variants in genes involved in syndromic diseases can be extended to individual  
48 phenotypes within the general population.

49

## 50 **Author Summary**

51 Standard medical evaluation of genetic syndromes relies upon recognizing a  
52 characteristic pattern of signs or symptoms to guide targeted genetic testing for  
53 confirmation of the diagnosis. This may lead to missing diagnoses in patients with silent  
54 or a low expressed form of the syndrome. Here we take advantage of a rich electronic  
55 health record, various phenotypic measurements, and genetic information in 337,198  
56 unrelated white British from the UKBB, to study the relation between single syndromic  
57 disease phenotypes and genes related to syndromic disease. We show multiple  
58 phenotype-genotypes associations in concordance with phenotypes variations found in  
59 syndromic diseases. For example, we show that mutation in *FBN1* was associated with  
60 high standing/sitting height ratio and reduced body fat percentage as observed in  
61 individuals with Marfan syndrome. Our findings suggest that common and rare alleles in  
62 SD genes are causative of individual component phenotypes present in a general  
63 population; further research is needed to characterize the pleiotropic effect of alleles in  
64 syndromic genes in persons without the syndromic disease.

65

## 66 **Introduction**

67 Genetic syndromes are rare diseases defined by a specific and clinically recognizable set  
68 of phenotypes across multiple organ systems. The era of next-generation sequencing has  
69 enabled substantial progress in linking syndromic disease to specific genetic loci, coupled  
70 with public databases of genotype-phenotype relationships to facilitate the classification  
71 of genetic variants from “benign” to “pathogenic” for use in clinical decision making. Large  
72 population-scale databases of genetic variation without phenotypes, such as ExAC, have  
73 provided additional context for characterizing genotype-phenotype relationships in  
74 genetic disease[1]. For mutations previously thought to cause disease, population  
75 databases have often suggested lower estimates of penetrance than initially  
76 recognized[2,3].

77

78 The diagnosis or classification of an individual with genetic syndrome relies upon expert  
79 recognition of a characteristic pattern of signs or symptoms or a set of defined diagnostic  
80 criteria. However, individuals displaying single phenotypes of a complex syndrome may  
81 not meet criteria for clinical diagnosis or genetic testing; expanding a binary definition of  
82 syndromic phenotypes to phenotype scores can identify more individuals with Mendelian  
83 disease patterns[4]. Similarly, individuals with clearly pathogenic mutations may be  
84 affected with only a single component phenotype of a genetic syndrome[5,6]. Recent  
85 descriptions of allelic heterogeneity, penetrance, and expressivity in syndromic disease  
86 genes have focused almost exclusively upon rare or familial alleles[7,8].

87

88 Here, we present a phenome-wide association study (PheWAS) approach to  
89 systematically explore pleiotropy of common and rare alleles in genes associated with

90 four well-described syndromic diseases in the general population. Using the UK Biobank,  
91 we linked individual-level medical and morphometric data to the characteristic phenotypes  
92 of Alagille (AS), Marfan (MS), DiGeorge (DS), and Noonan (NS) syndromes. These data  
93 allow a survey of the association of common and rare alleles to single component  
94 phenotypes of each syndrome within the general (non-syndromic) population.  
95

## 96 **Results**

97 Based on the Human Phenotype Ontology (HPO) – an ontology-based system developed  
98 using medical literature and other ontology-based systems[9] – we identified 196 HPO  
99 terms related to AS, MF, DS, and NS. Of these 196 HPO terms, 53 were shared between  
100 at least two syndromes, and seven terms were included in all four syndromes (S1 Table).  
101 After grouping the HPO terms into categories based on affected organ systems, there  
102 were 115 HPO terms of which 73 could be matched to 100 phenotypes available in the  
103 UKBB. Most of the unmatched phenotypes were related to specific abnormalities of body  
104 structure or the musculoskeletal system, which were poorly represented in clinical and  
105 billing codes, or measurements such as impaired T-cell function, not available in the  
106 UKBB.

107

### 108 **Characteristics of the study population**

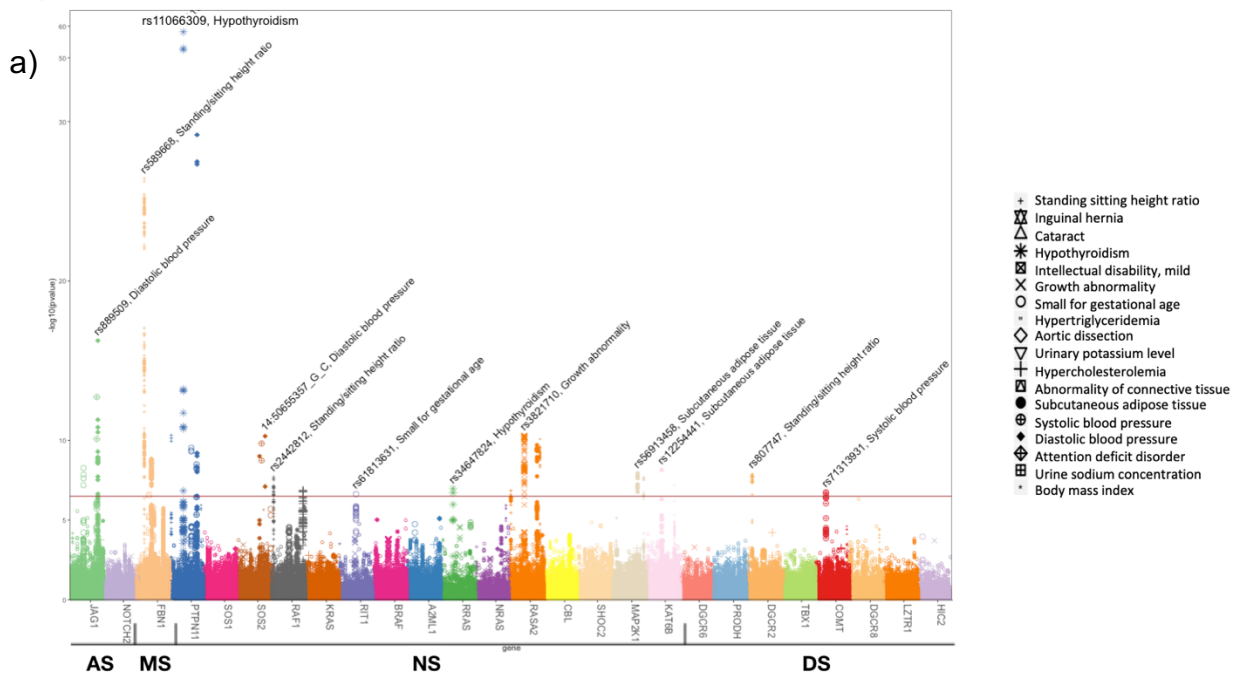
109 A total of 337,198 unrelated individuals were included in our analysis; the mean age was  
110 65.8 years (sd=8.0) and 53.7% of subjects were male. The number of subjects by  
111 phenotype is present in the S2 Table. Hypercholesterolemia (HP0003124),  
112 gastroesophageal reflux (HP0002020), premature osteoarthritis (HP0003088), and  
113 hypertriglyceridemia (HP0002155) were the most prevalent phenotypes with 12.8%  
114 (43,054 cases), 9% (30,229 cases), 8.9% (2,994 cases), and 8.6% (29,137 cases),  
115 respectively.

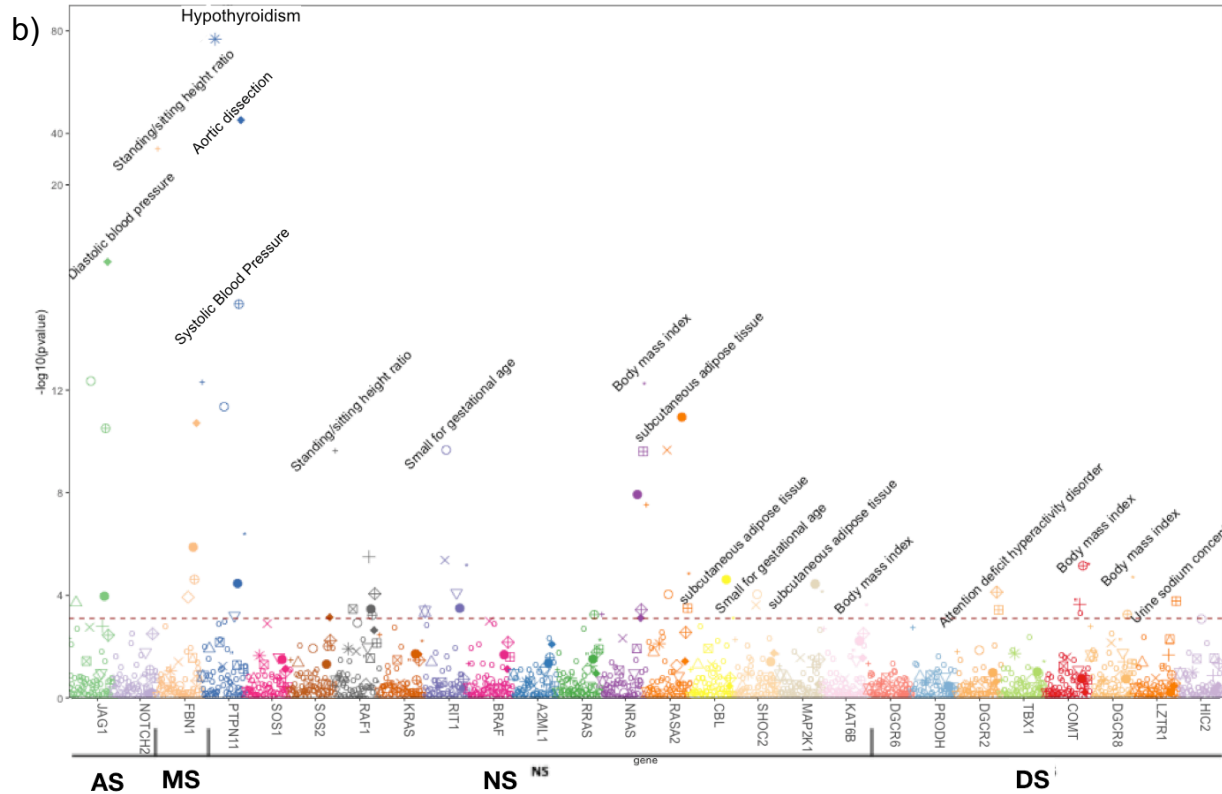
116

### 117 **Genotype-phenotype associations are common across syndromic genes**

118 Overall, 9 phenotypes: hypothyroidism (HP0000821), diastolic BP (HP0005117), systolic  
119 BP (HP0004421), standing/sitting height ratio (abnormality of body height; HP0000002),

120 birth weight (small for gestational age; HP0001518), amount of subcutaneous adipose  
 121 tissues or body fat percent (reduced subcutaneous adipose tissue; HP0003758), growth  
 122 abnormality (HP0001507), body mass index (abnormality of body mass; HP0045081),  
 123 and hyperlipidemia (HP0003124) were significantly associated with multiple SNPs across  
 124 *PTPN11*, *FBN1*, *JAG1*, *SOS2*, *RIT1*, *RAF1*, *KAT6B*, *RASA2*, *MAP2K1*, *DGCR2* and  
 125 *COMT* (Figure 1a, Table 1). The top SNP- phenotype association in each gene are  
 126 reported in Table 1. Diastolic BP and systolic BP along with body mass index displayed  
 127 a genetic association in each of the four syndromes. Birth weight, subcutaneous adipose  
 128 tissue (body fat percent), and tall stature for MS or short stature for NS are common  
 129 phenotypes for MS and NS; while hypothyroidism, and growth retardation are reported in  
 130 NS and DS. After correction for multiple testing, several other phenotypes reach  
 131 significance at the gene level (Figure 1b). When assessing the association by group of  
 132 phenotypes in each syndrome, we observed different patterns of association.





134

135 **Figure 1. Primary PheWAS results:** the variant level (a) and gene level (b). The red line

136 represents the level of significance after Bonferroni correction ( $p < 3.2 \times 10^{-7}$  at SNP level

137 and  $p < 9 \times 10^{-4}$  at gene level). Genes are represented by color and HPO terms are

138 indicated by shape.

SD genes	Phenotypes (HPO code)	N Cases	N controls	Total	OR or BETA	Se or IC [95%]	P
MS	<b>15:48835935_C_T (rs589668); FBN1 ALT=T Freq=0.25</b>						
	Standing and sitting height ratio (HP0000002)	-	-	334227	0.030	0.0028	3.54E-27
	Diastolic blood pressure (HP0005117)	-	-	334777	0.016	0.0028	7.00E-09
	Subcutaneous adipose tissue (HP0003758)	-	-	331110	-0.074	0.0179	3.52E-05
NS	<b>12:112883476_G_A (rs11066309); PTPN11 ALT=A Freq=0.41</b>						
	Hypothyroidism (HP0000821)	17746	256695	337198	1.190	[1.16-1.21]	6.62E-59
	Diastolic blood pressure (HP0005117)	-	-	334777	0.028	0.002	6.71E-30
	Standing and sitting height ratio (HP0000002)	-	-	334227	0.016	0.002	4.59E-11
	Small for gestational age (HP0001518)	-	-	192869	-0.020	0.003	2.95E-10
	Systolic blood pressure (HP0004421)	-	-	334544	0.014	0.002	3.24E-09
	Body mass index (HP0045081)	-	-	333117	-0.012	0.002	1.13E-06
	Stroke (HP0001297)	4930	269511	337198	1.080	[1.04-1.12]	2.96E-05
	<b>14:50655357_G_C (rs72681869); SOS2 ALT=C Freq=0.01</b>						
	Diastolic blood pressure (HP0005117)	-	-	334777	-0.076	0.0115	5.42E-11
Systolic blood pressure (HP0004421)	-	-	334544	-0.070	0.0110	1.58E-10	
Subcutaneous adipose tissue (HP0003758)	-	-	331110	-0.452	0.0739	9.83E-10	



	Body mass index (HP0045081)	-	-	333117	-0.055	0.0116	2.00E-06
	Small for gestational age (HP0001518)	-	-	192869	0.072	0.0152	2.03E-06
	Inguinal hernia (HP0000023)	367	7068	337198	1.230	[1.11-1.38]	1.61E-04
	Growth abnormality (HP0001507)	-	-	328546	-0.092	0.0257	3.61E-04
	<b>15:66702345_C_A (rs56913458); MAP2K1 ALT=A Freq=0.23</b>						
	Subcutaneous adipose tissue (HP0003758)	-	-	331110	-0.106	0.0185	1.14E-08
	Body mass index (HP0045081)	-	-	333117	-0.016	0.003	2.05E-08
	<b>1:155865042_T_C (rs61813631); RIT1 ALT=C Freq=0.02</b>						
	Small for gestational age (HP0001518)	-	-	192869	0.067	0.013	2.32E-07
	<b>3:12649576_A_G (rs2442812); RAF1 ALT=G Freq=0.45</b>						
	Standing and sitting height ratio (HP0000002)	-	-	334227	0.013	0.002	1.83E-08
	Hypercholesterolemia (HP0003124)	37775	262539	337198	0.970	[0.95-0.98]	2.46E-05
	Intellectual disability, mild (HP0001256)	-	-	108818	-0.016	0.004	1.50E-04
	<b>3:141301451_G_A (rs3821710); RASA2 ALT=A Freq=0.40</b>						
	Growth abnormality (HP0001507)	-	-	328546	0.036	0.006	4.80E-11
	Subcutaneous adipose tissue (HP0003758)	-	-	331110	0.099	0.016	3.29E-10
	Body mass index (HP0045081)	-	-	333117	0.010	0.002	8.33E-05
	Body height (HP0000002)	-	-	334227	-0.009	0.002	1.99E-04
	<b>10:76617857_C_T (rs12254441); KAT6B ALT=T Freq=0.37</b>						
	Subcutaneous adipose tissue (HP0003758)	-	-	331110	-0.096	0.017	7.12E-09
	Body mass index (HP0045081)	-	-	333117	-0.013	0.003	6.42E-07
	Hypothyroidism (HP0000821)	10387	172953	337198	0.940	[0.92-0.96]	1.07E-07
	<b>20:10624190_A_G (rs889509); JAG1 ALT=G Freq=0.42</b>						
AS	Diastolic blood pressure (HP0005117)	-	-	334777	-0.021	0.002	5.55E-17
	Small for gestational age (HP0001518)	-	-	192869	-0.012	0.003	1.86E-04
	<b>22:19104926_C_G (rs807747); DGCR2 ALT=G Freq=0.06</b>						
DS	Body height (HP0000002)	-	-	334227	-0.028	0.005	1.16E-08
	<b>22:19960184_C_G (rs71313931); COMT ALT=G Freq=0.29</b>						
	Systolic blood pressure (HP0004421)	-	-	334544	0.013	0.003	1.88E-07
	Body mass index (HP0045081)	-	-	333117	0.010	0.003	3.38E-04

139 Table 1: Table of association between the top SNPs in each gene and significant

140 (pvalue < 3.16x10<sup>-07</sup>) or suggestive (p-value < 5x10<sup>-04</sup>) associations with HPO

141 phenotypes.

142

### 143 Variation in syndromic genes are associated with component phenotypes

144 Marfan syndrome (MS) is a primary disorder of connective tissue with diagnostic criteria

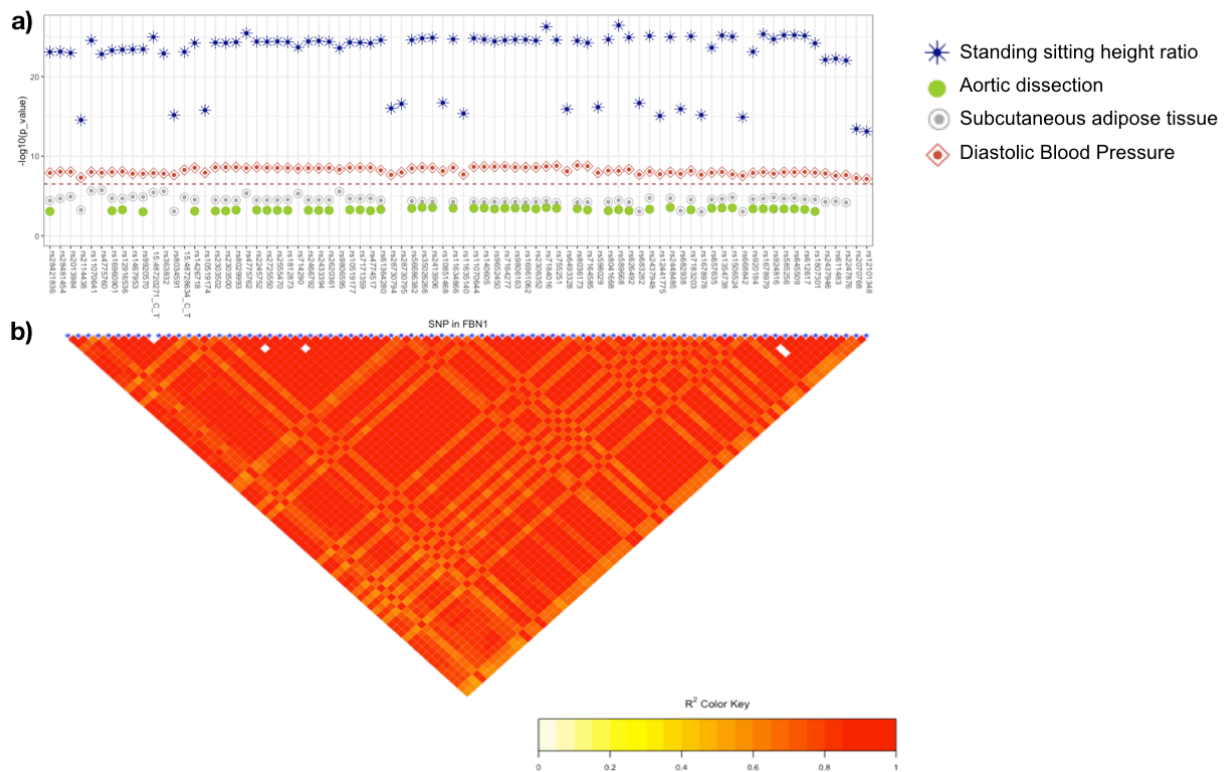
145 centered around cardiovascular, musculoskeletal, and ocular phenotypes linked to a

146 single gene *FBN1* which encodes an extracellular matrix protein. Several SNPs in *FBN1*

147 were significantly associated with increased standing/sitting height ratio and an elevated

148 diastolic BP. An increased risk of aortic dissection and a lower percent of body fat (two

149 major phenotypes in MS) were observed for several of these SNPs although the  
150 association was suggestive (Figure 2a and S3 Table). All SNPs in *FBN1* displaying  
151 associations were located within the same LD block and were highly correlated  
152 suggesting a single signal within the gene (Figure 2a and 2b). A common intronic variant  
153 rs589668 displays the top signal with high standing/sitting height ratio (beta=0.03,  
154 se=0.002, p=10<sup>-27</sup>, Table 1), an elevated diastolic BP (beta=0.02, se=0.002, p=7x10<sup>-09</sup>),  
155 and a lower percent of body fat (beta=-0.07, se=0.02, p=5x10<sup>-05</sup>, Table 1). Four additional  
156 SNPs (rs11070641, rs4775760, rs363832 and rs140605) that reach genome-wide  
157 significance with high standing/sitting height ratio and diastolic BP were correlated with  
158 several syndromic disease entities in CLINVAR including stiff skin syndrome, ectopia  
159 lentis, MASS syndrome, thoracic aortic aneurysm and aortic dissection (Supplementary  
160 Table 3). At the gene level association using SKAT test with SNPs allelic frequency  
161 weighted by their CADD (Combined Annotation Depletion Dependent) score,  
162 standing/sitting height ratio, systolic and diastolic BP, subcutaneous adipose tissue, and  
163 aortic dissection were significantly associated with *FBN1* (Figure 1b, S4 table).



164

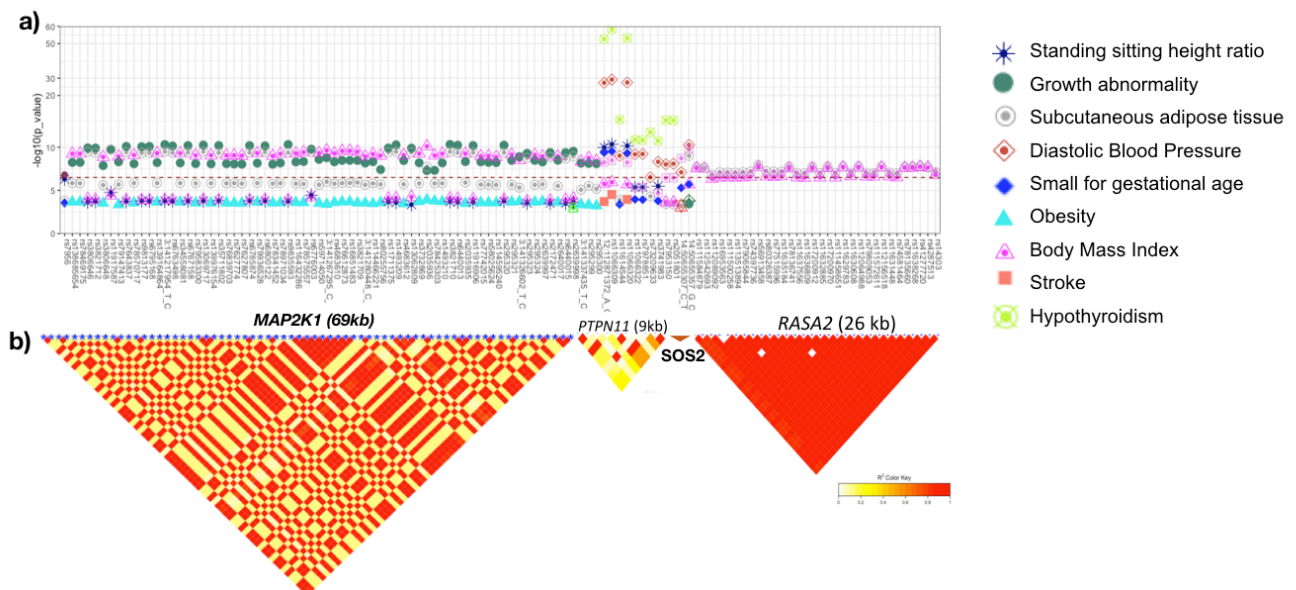
165 **Figure 2: PheWAS result and linkage plot for SNPs with pleiotropy in *FBN1*.**

166 Associations between SNPs with Pleiotropy and phenotypes (a) and linkage between the  
167 SNPs in *FBN1* (b). The red line represents the level of significance after Bonferroni  
168 correction ( $p=3.2 \times 10^{-07}$ )

169

170 For NS and RAS-opathy related phenotypes, SNPs in *PTPN11* were associated with  
171 increased risk of hypothyroidism, high diastolic and systolic BP, and high standing/sitting  
172 height ratio (Figure 1a, S3 Table). SNPs in *SOS2* were associated with lower systolic and  
173 diastolic BP, and lower percent of body fat (Figure 1a, Supplementary Table 3). SNPs in  
174 *MAP2K1* were associated with lower body mass index, as well as lower level of cutaneous  
175 adipocytes tissues (Figure 1a, Supplementary Table 3). SNPs in *PTPN11* display a  
176 moderate to low correlation with each other suggesting several independent signals within

177 the locus, while high correlation was observed between SNPs in *RASA2*, *SOS2* and  
 178 *MAP2K1* indicating that the association observed within each gene represents a single  
 179 signal (Figure 3b). Among SNPs with pleiotropic effects, rs11066309 in *PTPN11* displays  
 180 a strong association with increased risk for hypothyroidism (ALT freq=0.40; OR [95% CI]:  
 181 1.19; [1.16 – 1.21];  $p=6 \times 10^{-59}$ ) along with five other phenotypes, including decreased body  
 182 mass index (beta=-0.012,  $p=1.13 \times 10^{-06}$ ) and birth weight (beta=-0.020,  $p=2.95 \times 10^{-10}$ )  
 183 (Figure 3a, Table 1). In addition, among SNPs that reached significance, rs3741983  
 184 (*PTPN11*), rs72681869 (*SOS2*), rs61755579 (*SOS2*), and rs112542693 (*MAP2K2*) were  
 185 linked to Noonan syndrome and a number of abnormalities including global  
 186 developmental delay, short stature and abnormality of cardiovascular system in ClinVar.  
 187 At a gene level, *PTPN11*, *NRAS*, *RASA2*, *SOS2*, *MAP2K1*, and *RAF1* were significantly  
 188 associated with hypothyroidism, diastolic and systolic BP, birth weight, growth  
 189 abnormality, subcutaneous adipose tissue, standing/sitting height ratio and body mass  
 190 index (Figure 1b and S4 Table).



192 **Figure 3: PheWAS result and linkage plot for SNPs with pleiotropy in RAS-opathie**  
193 **genes.** Associations between SNPs with Pleiotropy effect in *MAP2K1*, *PTPN11*, *SOS2*  
194 and *RASA2*, and HPO terms (a) and linkage between the SNPs (b). The red line  
195 represents the level of significance after Bonferroni correction ( $p=3.2 \times 10^{-07}$ )

196  
197 Alagille syndrome is caused by mutations in *JAG1* and *NOTCH2* with major clinical  
198 manifestations in the heart and liver, and characteristic facial features. At a SNP level,  
199 none of the AS specific phenotypes reached significance after multiple testing correction.  
200 However, suggestive associations were observed between rs1051412 (reported in  
201 ClinVar as a benign variant) and cataracts ( $p=2.9 \times 10^{-04}$ , S1(a) Figure). Several SNPs in  
202 *JAG1* were significantly associated with diastolic BP, systolic BP and birth weight ( $p < 10^{-08}$ ,  
203 Figure 1a, S3 Table). The SNP rs889509 (20:10658882\_G\_C) displayed the most  
204 significant association with a lower diastolic BP (beta=-0.028;  $p=8.2 \times 10^{-11}$ , Table 1). At a  
205 gene level, *JAG1* was associated with diastolic BP ( $p=3.48 \times 10^{-15}$ ), birth weight  
206 ( $p=3.84 \times 10^{-10}$ ), systolic BP ( $p=3.32 \times 10^{-09}$ ) and cataracts ( $p=2.21 \times 10^{-04}$ , Figure 1b,  
207 Supplementary Table 4).

208  
209 DiGeorge syndrome encompasses a recurrent microdeletion of multiple genes at the  
210 22q11.2 locus due to the presence of segmental duplications, with affected individuals  
211 displaying neuropsychiatric, immunological, and cardiovascular phenotypes originating  
212 from defects in neural crest cell formation and migration. At the SNP level, rs807747 in  
213 *DGCR2* and rs71313931 in *COMT* were significantly associated with abnormal body  
214 height ( $p=2 \times 10^{-07}$ ) and systolic BP ( $p=5.6 \times 10^{-17}$ ), respectively (Figure 1a and Table 1). At

215 the gene level, *COMT* was associated with systolic BP, while *DGCR2* was associated  
216 with diastolic BP and standing/sitting height ratio ( $p < 5 \times 10^{-4}$ ) (S4 Table).

217

## 218 **Discussion**

219 Here, we systematically describe the association of variation at four syndromic loci with  
220 the component phenotypes of syndromic disease. We hypothesize that in the general  
221 population, common and rare alleles for syndromic diseases display pleiotropic effects  
222 with the phenotypes related to genetic syndromes. Using the UKBB, we linked individual-  
223 level data to the characteristic phenotypes of Alagille, Marfan, Noonan, and DiGeorge  
224 syndromes, showing clearly the association of common and rare alleles to single  
225 component phenotypes of each syndrome. While classical Mendelian forms of each  
226 syndrome may be uncommon, individual phenotypes attributable to these syndromic loci  
227 appear to be common within the general population. This finding support the findings of  
228 Bastarache et al[4], suggesting a scaling of Mendelian disease phenotypes into a  
229 continuous phenotyping score to improve the identification of individuals with rare  
230 diseases.

231

232 Within families of individuals affected by syndromic disease carrying the same pathogenic  
233 mutation, the expressivity of component phenotypes may vary in different  
234 individuals[10,11]. Here, we show that many common and rare variants at syndromic loci  
235 existing in the general population may result in expression of traits and phenotypes  
236 closely related to the syndrome of interest. For example, we observe associations of a  
237 common intronic variant in *FBN1* rs589668 (MAF=0.25 in Europeans populations) with  
238 increases in blood pressure and height and decreased subcutaneous fat distribution. In  
239 GTex[12], this variant is an eQTL strongly associated with decreased expression in whole  
240 blood ( $p=1.7 \times 10^{-37}$ ), which would be concordant with the known molecular mechanism of  
241 *FBN1* pathogenesis in MS: pathogenic alleles impairing gene function result in increased



242 height and abnormal fat distribution and increased arterial stiffness[13,14]. Modifiers of  
243 penetrance and phenotypic expressivity in Marfan syndrome have been proposed,[15,16]  
244 but our results suggest that common variants and local haplotype structure around  
245 syndromic genes may deserve more attention[17].

246

247 Noonan syndrome is caused by mutations in *PTPN11* and part of a group of related  
248 disorders arising from activating mutations in RAS-MAPK signaling pathway known as  
249 RASopathies which display many phenotypes across a variety of organ systems. A wide  
250 phenotypic variability and genetic heterogeneity have also been described in individuals  
251 with Noonan syndrome in relation to rare variants in *PTPN11*[18]. Here, we show that  
252 even in the general population, common and rare variants in *PTPN11* are independently  
253 associated with phenotypes such as hypothyroidism, small birth weight and low percent  
254 of body fat observed in some Noonan syndromes cases[19–21]. In GTex[12], numerous  
255 variants in *PTPN11*, such as rs11066309, rs3741983 and rs11066322 were significantly  
256 associated with a decreased expression in atrial appendage, adipose tissue, thyroid and  
257 skin and esophagus. Although in consistent with the role of *PTPN11* in thyroid function,  
258 cancer and autoimmunity[22–24], these variants are instead described as eQTLs with  
259 *TMEM116*, *ALDH2* and *MAPKAPK5-AS1* located up to 500kb upstream of *PTPN11*,  
260 suggesting that the association observed with rs11066309, rs3741983 and rs11066322  
261 may also be potentially linked to other genes.

262

263 Growth retardation, lower BMI and short stature are well-known characteristics of Noonan  
264 syndrome, and a recent study reported a phenotype-genotype variability of growth pattern



265 in patients with Noonan syndrome[25]. In concordance with this study, we showed that,  
266 in the general population, common and rare variants in *RASA2*, *SOS2* and *MAP2K1* are  
267 independently associated with growth characteristics (body mass index, height and  
268 growth abnormality) and the association driven by single or multiple haplotype in each  
269 gene.

270

271 When performing genetic testing, allele frequency is often used as a marker to determine  
272 pathogenicity of a genetic variant. Common variation in and around *JAG1* has previously  
273 been associated with such disparate phenotypes as pulse pressure, circulating blood  
274 indices, and birthweight, and none of the variants included in our analysis appeared to be  
275 directly associated with the component phenotypes of AS. However, the unifying  
276 molecular abnormality in AS are defects in vascular formation which lead to each of the  
277 component cardiovascular and liver phenotypes of Alagille syndrome[26,27].

278

279 Our study has some limitations. Our analysis was limited to phenotypes with more than  
280 100 cases; therefore, diseases with relatively rare prevalence were not analyzed. In  
281 addition, because our study cohort consist of adults from the general population, specific  
282 phenotype targeting facial and skeletal dysmorphism, such as butterfly vertebrae or broad  
283 forehead; specific abnormalities of organs, such as biliary disease; and phenotypes  
284 observed during childhood, such as developmental delay or attention deficit and  
285 hyperactivity disorder (ADHD) were not present. However, to correct on the lack of some  
286 of these phenotypes, we used certain phenotypes or measurement present in the UKBB,  
287 such as head circumference as a proxy for broad forehead, education level for ADHD,

288 weight and height at age 10 as proxy for growth abnormality. That said, the complexity of  
289 matching UKBB phenotypes to HPO terms may simply not capture some phenotypes,  
290 despite manual curation.

291

292 Key strengths of our study include the ability to systematically test multiple phenotype-  
293 genotype association and to highlight pleiotropy linked to syndromic genes. Our study  
294 maps UKBB phenotypes to HPO terms and shows that common and rare variants in  
295 genes responsible for Alagille, Marfan, Noonan, and DiGeorge syndromes, are also  
296 independently associated with component phenotypes of these syndromes in the general  
297 population.

298

299 Our findings suggest that the spectrum of association of common and rare variants in  
300 genes involved in syndromic diseases can be extended to individual component  
301 phenotypes in the general population. Further research is needed to characterize the  
302 pleiotropic effect of alleles in genes in persons without the corresponding genetic  
303 syndrome.

304

## 305 **Materials and Methods**

### 306 **Study population and data collection**

307 The study cohort was derived from the UK Biobank (UKBB), a large prospective cohort  
308 study with comprehensive health data from over 500,000 volunteer participants in the  
309 United Kingdom aged 37-73 years at recruitment in 2006-2010. The cohort has previously  
310 been described in detail[28,29]. Information on the UK biobank participants was collected  
311 at enrollment, and from electronic health record (EHR) information which includes  
312 diagnostic codes (ICD10, ICD9) and procedural code (OPCS) from hospital admission  
313 records dating to 1992, and cancer registries. Data collected at the assessment visit  
314 included information on a participant's health and lifestyle, hearing and cognitive function,  
315 collected through a touchscreen questionnaire and verbal interview. A range of physical  
316 measurements was also performed, including blood pressure; arterial stiffness; body  
317 composition measures (including impedance); hand-grip strength; ultrasound bone  
318 densitometry; spirometry; and an exercise/fitness test with ECG. Samples of blood, urine,  
319 and saliva were also collected. Medical phenotypes were aggregated as previously  
320 described, incorporating available information including a broad set of medical  
321 phenotypes defined using computational matching and manual curation of on hospital in-  
322 patient record data (ICD10 and ICD9 codes), self-reported verbal questionnaire data, and  
323 cancer and death registry data[30,31].

324

### 325 **Phenotypes of target syndromes**

326 We identified phenotypes related to syndromic diseases through the Human Phenotype  
327 Ontology (HPO). HPO is an ontology-based system developed using medical literature,  
328 and other ontology-based systems such as Orphanet, and OMIM[9]. HPO provides a

329 standardized vocabulary of phenotypic and abnormalities encountered in human  
330 diseases. The HPO has link symptoms/phenotypes to diseases or genetic disorders, and  
331 the causing genes. As an example, Alagille syndrome (AS) is linked to *JAG1*, and  
332 *NOTCH2* genes as well as all the phenotypes or symptoms observed in AS, such as atrial  
333 septal defect, hypertelorism, and butterfly vertebra.

334  
335 HPO terms were directly matched to UKBB phenotypes when phenotypes in both  
336 systems had similar terminology. The direct phenotype matching was conducted using a  
337 semi-automatic mapping system which combines semantic and lexical similarity between  
338 word[32] followed by manual curation. When the HPO terms were not present, we  
339 performed an indirect matching by hand to find in the UKBB, the phenotype that best  
340 reflects the target HPO terms. For example, abnormality of body structure or body  
341 morphology such as abnormal body height, reduced sub cutaneous adipose tissues, bone  
342 density or broad forehead, were respectively matched to sitting/standing height ratio;  
343 body fat percentage; bone mineral density, and head bone area. For psychiatric diseases  
344 such as depression and neurodevelopmental disorders such as attention deficit and  
345 hyperactivity disorder (ADHD), we used a score of depressive symptoms and self-  
346 reported educational level respectively as proxies for these terms.

347  
348 To increase the number of subjects in some subgroup of phenotypes, we combined  
349 subcategories of HPO terms into a group or category. For example, 39 HPO terms  
350 representing an abnormality of head, ears, and eyes such as low-set ears, strabismus,  
351 macrotia, webbed neck, short neck, abnormality of the eye, microcornea, down-slanted

352 palpebral fissure and other congenital abnormality of ears, were grouped into  
353 “Abnormality of head or neck (HP0000152)” and mapped to icd10 targeting congenital  
354 malformations of eye, ear, face, and neck and other organs especially facial appearance  
355 (ICD10: Q10 to Q18 and Q87). Ten HPO terms for congenital abnormality of  
356 cardiovascular system including Ventricular septal defect, Atrial septal defect, Tetralogy  
357 of Fallot, Patent ductus arteriosus, Bicuspid aortic valve, Truncus arteriosus, Coarctation  
358 of aorta, Tricuspid valve prolapse were combined into abnormality of the cardiovascular  
359 system (HP0001626) and mapped to Congenital malformations of the circulatory system  
360 (ICD10: Q20 to Q28).

361

## 362 **Genotyping data**

363 Genotyping was performed using the Affymetrix UK BiLEVE Axiom array on an initial  
364 50,000 participants; the remaining 450,000 participants were genotyped using the  
365 Affymetrix UK Biobank Axiom® array. The two arrays are extremely similar (with over  
366 95% common content). Quality control and imputation to over 90 million SNPs, indels and  
367 large structural variants was performed[33]. 5647 SNPs in 26 genes related to Alagille  
368 syndrome, Marfan syndrome, Noonan syndrome, and DiGeorge syndrome were selected  
369 for our study (table2). The selected SNPs had a MAF>0.005 and an imputation  
370 measurement ( $R^2$ )  $\geq 0.8$

371

372

373

374

	<b>Genes</b>	<b>MAF&gt;0.005, R2&lt;0.8</b>	<b>HPO terms</b>	<b>matched HPO term</b>
Alagille Syndrome	<i>JAG1, NOTCH2</i>	302	61	36
Marfan syndrome	<i>FBN1</i>	518	64	42
Noonan syndrome	<i>PTPN11, SOS1, RAF1, KRAS, RIT1, BRAF, A2ML1, RRAS, SOS2, NRAS, RASA2, CBL, SHOC2, MAP2K1, KAT6B</i>	3062	58	24
DiGeorge syndrome	<i>22q11.2 deletion (TBX1, DGCR2, DGCR8, DGCR6, COMT, PRODH, HIC2, LZTR1)</i>	1228	66	35

375

376 Table 2: table summarizing number of genes, SNPs and phenotypes (HPO terms) for  
377 each syndrome, include in our analysis

378

### 379 **Statistical analysis**

380 SNP level

381 For binary traits, logistic regression with adjustment on age, sex, batch, and the top 5  
382 principal components were used. First, regression was used in a situation of unbalanced  
383 numbers of cases and controls, especially when the number of cases was very small (less  
384 than 200 cases). For continuous traits, we performed linear regression with adjustment  
385 on age, sex, batch, and the top 5 principal components. Our analysis was restricted to  
386 individuals of Europeans descent, due to the relatively small number of individuals from  
387 other ethnic groups in the UKBB. Bonferroni correction based on the number of  
388 independent tests was used to correct on multiple testing. Given the high correlation  
389 between SNPs within gene or regions, Bonferroni correction is often stringent when to  
390 number of tests considered is number of SNPS time the number of phenotypes. To take  
391 in account the correlation between SNPs, we estimate the number of independent SNPs

392 in a block of 50 kb with a correlation > 0.8 using the pairwise pruning method implemented  
393 in [PLINK](#) which estimated 2166 independent SNPs within our target regions. We apply  
394 a threshold of  $3.16 \times 10^{-07} = 0.05 / (2166 \times 73)$  independent tests.

395 Gene Level

396 The Sequence Kernel association test (SKAT) was performed for gene-level association  
397 using allelic frequency methods and weighted by the CADD score (Combined Annotation  
398 Depletion Dependent)[34,35]. A Bonferroni correction of  $1 \times 10^{-03}$  based on 26 genes  
399 analyzed using allelic frequency methods and weighted by the CADD score was applied  
400 for multiple testing correction.

401

## 402 **Acknowledgments**

403 This study is supported by data from the UKBB, applications 24983 (Dr Rivas), 15860 (Dr  
404 Priest), and 13721 (Dr Ingelsson).

405

## 406 **Source of funding**

407 This project was funded by the National Institutes of Health (R00HL130523) and  
408 Stanford CVI-MCHRI Seed Grant to Dr Priest

## 409 **Disclosures**

410 All co-authors have nothing to disclose

## 411 **Supporting information**

412 **S1 Table:** Overall HPO term present in Alagille, Noonan, Marfan and Digeorge  
413 syndromes

414 **S2 Table:** description binary and continuous phenotypes

415 **S3 table:** Association between all SNPs and all phenotypes

416 **S4 table:** Gene level association with SKAT with SNPs weighted using the CADD score  
417 or using the MAF

418 **S1 figure:** diagram phenotype matching system between the UK- biobank (UKBB)  
419 phenotypes and HPO terms

420 **S2 figure:** Association between SNPs in *FBN1* and all HPO terms at the SNPS and  
421 gene levels

422 **S3 figure:** Association between SNPs in *PTPN11*, and *gene in RAS/MAKP2* and all  
423 HPO terms at the SNPS and gene levels

424 **S4 figure:** Association between SNPs in *NOTCH2*, and *JAG1* and all HPO terms at the  
425 SNPS and gene levels

426 **S4 figure:** Association between SNPs in *22q11 locus* and all HPO terms at the SNPS  
427 and gene levels

428 **S6 figure:** Forest plots showing association between Clinvar SNPs in *FBN1*, *JAG1*,  
429 *PTPN11*, *MAP2K1* and *SOS2* and, MF, AS, NS-related phenotypes, respectively.

430



431 **References**

- 432 1. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al.  
433 Genetic Misdiagnoses and the Potential for Health Disparities. *N Engl J Med*.  
434 2016;375: 655–665. doi:10.1056/NEJMsa1507092
- 435 2. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, et al.  
436 Using high-resolution variant frequencies to empower clinical genome  
437 interpretation. *Genet Med*. 2017;19: 1151–1158. doi:10.1038/gim.2017.26
- 438 3. Wright CF, West B, Tuke M, Jones SE, Patel K, Laver TW, et al. Assessing the  
439 Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing  
440 Variants in a Population Setting. *Am J Hum Genet*. 2019;104: 275.  
441 doi:10.1016/J.AJHG.2018.12.015
- 442 4. Bastarache L, Hughey JJ, Hebring S, Marlo J, Zhao W, Ho WT, et al. Phenotype  
443 risk scores identify patients with unrecognized mendelian disease patterns.  
444 *Science (80- )*. 2018;359: 1233–1239. doi:10.1126/science.aal4043
- 445 5. Priest JR, Ceresnak SR, Dewey FE, Malloy-Walton LE, Dunn K, Grove ME, et al.  
446 Molecular diagnosis of long QT syndrome at 10 days of life by rapid whole  
447 genome sequencing. *Hear Rhythm*. 2014;11: 1707–13.  
448 doi:10.1016/j.hrthm.2014.06.030
- 449 6. Priest JR, Osoegawa K, Mohammed N, Nanda V, Kundu R, Schultz K, et al. De  
450 Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of  
451 Atrioventricular Septal Heart Defects. *PLoS Genet*. 2016;12: e1005963.  
452 doi:10.1371/journal.pgen.1005963
- 453 7. Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF,

- 454 et al. Quantifying prion disease penetrance using large population control cohorts.  
455 Sci Transl Med. 2016;8: 322ra9-322ra9. doi:10.1126/scitranslmed.aad5169
- 456 8. Mallawaarachchi AC, Furlong TJ, Shine J, Harris PC, Cowley MJ. Population data  
457 improves variant interpretation in autosomal dominant polycystic kidney disease.  
458 Genet Med. 2018; 1. doi:10.1038/s41436-018-0324-x
- 459 9. Lobo M, Lamurias A, Couto FM. Identifying Human Phenotype Terms by  
460 Combining Machine Learning and Validation Rules. Biomed Res Int. 2017;2017:  
461 1–8. doi:10.1155/2017/8565739
- 462 10. Moncini S, Bonati MT, Morella I, Ferrari L, Brambilla R, Riva P. Differential allelic  
463 expression of SOS1 and hyperexpression of the activating SOS1 c.755C variant  
464 in a Noonan syndrome family. Eur J Hum Genet. 2015;23: 1531–7.  
465 doi:10.1038/ejhg.2015.20
- 466 11. Kamath BM, Bason L, Piccoli DA, Krantz ID, Spinner NB. Consequences of JAG1  
467 mutations. J Med Genet. 2003;40: 891–5. doi:10.1136/JMG.40.12.891
- 468 12. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel  
469 Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project.  
470 Biopreserv Biobank. 2015;13: 311–319. doi:10.1089/bio.2015.0032
- 471 13. Davis MR, Arner E, Duffy CRE, De Sousa PA, Dahlman I, Arner P, et al.  
472 Expression of FBN1 during adipogenesis: Relevance to the lipodystrophy  
473 phenotype in Marfan syndrome and related conditions. Mol Genet Metab.  
474 2016;119: 174–85. doi:10.1016/j.ymgme.2016.06.009
- 475 14. Salvi P, Grillo A, Marelli S, Gao L, Salvi L, Viecca M, et al. Aortic dilatation in  
476 Marfan syndrome. J Hypertens. 2018;36: 77–84.

- 477 doi:10.1097/HJH.0000000000001512
- 478 15. Aubart M, Gazal S, Arnaud P, Benarroch L, Gross M-S, Buratti J, et al.  
479 Association of modifiers and other genetic factors explain Marfan syndrome  
480 clinical variability. *Eur J Hum Genet.* 2018;26: 1759–1772. doi:10.1038/s41431-  
481 018-0164-9
- 482 16. Hutchinson S, Furger A, Halliday D, Judge DP, Jefferson A, Dietz HC, et al. Allelic  
483 variation in normal human FBN1 expression in a family with Marfan syndrome: a  
484 potential modifier of phenotype? *Hum Mol Genet.* 2003;12: 2269–2276.  
485 doi:10.1093/hmg/ddg241
- 486 17. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al.  
487 Modified penetrance of coding variants by cis-regulatory variation contributes to  
488 disease risk. *Nat Genet.* 2018;50: 1327–1334. doi:10.1038/s41588-018-0192-y
- 489 18. Jongmans M, Otten B, Noordam K, van der Burgt I. Genetics and Variation in  
490 Phenotype in Noonan Syndrome. *Horm Res Paediatr.* 2004;62: 56–59.  
491 doi:10.1159/000080500
- 492 19. Freire BL, Homma TK, Funari MFA, Lerario AM, Vasques GA, Malaquias AC, et  
493 al. Multigene sequencing analysis of children born small for gestational age with  
494 isolated short stature. *J Clin Endocrinol Metab.* 2019 [cited 29 Mar 2019].  
495 doi:10.1210/jc.2018-01971
- 496 20. Lee MJ, Kim BY, Ma JS, Choi YE, Kim YO, Cho HJ, et al. Hashimoto thyroiditis  
497 with an unusual presentation of cardiac tamponade in Noonan syndrome. *Korean*  
498 *J Pediatr.* 2016;59: S112–S115. doi:10.3345/kjp.2016.59.11.S112
- 499 21. Şıklar Z, Berberoğlu M. Syndromic disorders with short stature. *J Clin Res Pediatr*

- 500 Endocrinol. 2014;6: 1–8. doi:10.4274/Jcrpe.1149
- 501 22. Cote GJ, Grubbs EG, Hofmann M-C. Thyroid C-Cell Biology and Oncogenic  
502 Transformation. *Recent Results Cancer Res.* 2015;204: 1–39. doi:10.1007/978-3-  
503 319-22542-5\_1
- 504 23. Quaió CRDC, Carvalho JF, da Silva CA, Bueno C, Brasil AS, Pereira AC, et al.  
505 Autoimmune disease and multiple autoantibodies in 42 patients with  
506 RASopathies. *Am J Med Genet A.* 2012;158A: 1077–82.  
507 doi:10.1002/ajmg.a.35290
- 508 24. Couser NL, Keelean-Fuller D, Davenport ML, Haverfield E, Masood MM, Henin M,  
509 et al. Cleft palate and hypopituitarism in a patient with Noonan-like syndrome with  
510 loose anagen hair-1. *Am J Med Genet A.* 2018;176: 2024–2027.  
511 doi:10.1002/ajmg.a.40432
- 512 25. Cessans C, Ehlinger V, Arnaud C, Yart A, Capri Y, Barat P, et al. Growth patterns  
513 of patients with Noonan syndrome: Correlation with age and genotype. *Eur J*  
514 *Endocrinol.* 2016;174: 641–650. doi:10.1530/EJE-15-0922
- 515 26. Loerakker S, Stassen OMJA, Ter Huurne FM, Boareto M, Bouten CVC, Sahlgren  
516 CM. Mechanosensitivity of Jagged-Notch signaling can induce a switch-type  
517 behavior in vascular homeostasis. *Proc Natl Acad Sci U S A.* 2018;115: E3682–  
518 E3691. doi:10.1073/pnas.1715277115
- 519 27. Hofmann JJ, Zovein AC, Koh H, Radtke F, Weinmaster G, Iruela-Arispe ML.  
520 Jagged1 in the portal vein mesenchyme regulates intrahepatic bile duct  
521 development: insights into Alagille syndrome. *Development.* 2010;137: 4061–72.  
522 doi:10.1242/dev.052118

- 523 28. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank:  
524 An Open Access Resource for Identifying the Causes of a Wide Range of  
525 Complex Diseases of Middle and Old Age. *PLOS Med*. 2015;12: e1001779.  
526 doi:10.1371/journal.pmed.1001779
- 527 29. Collins R. What makes UK Biobank special? *Lancet*. 2012;379: 1173–1174.  
528 doi:10.1016/S0140-6736(12)60404-8
- 529 30. DeBoever C, Tanigawa Y, Lindholm ME, McInnes G, Lavertu A, Ingelsson E, et  
530 al. Medical relevance of protein-truncating variants across 337,205 individuals in  
531 the UK Biobank study. *Nat Commun*. 2018;9: 1612. doi:10.1038/s41467-018-  
532 03910-9
- 533 31. McInnes G, Tanigawa Y, DeBoever C, Lavertu A, Olivieri JE, Aguirre M, et al.  
534 Global Biobank Engine: enabling genotype-phenotype browsing for biobank  
535 summary statistics. *bioRxiv*. 2018; 304188. doi:10.1101/304188
- 536 32. Pang C, Sollie A, Sijtsma A, Hendriksen D, Charbon B, de Haan M, et al. SORTA:  
537 a system for ontology-based re-coding and technical annotation of biomedical  
538 phenotype data. *Database*. 2015;2015: bav089. doi:10.1093/database/bav089
- 539 33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-  
540 wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017 [cited 15  
541 Sep 2017]. Available: <http://www.biorxiv.org/content/early/2017/07/20/166298>
- 542 34. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for  
543 sequencing data with the sequence kernel association test. *Am J Hum Genet*.  
544 2011;89: 82–93. doi:10.1016/j.ajhg.2011.05.029
- 545 35. Richardson TG, Campbell C, Timpson NJ, Gaunt TR. Incorporating Non-Coding

546 Annotations into Rare Variant Analysis. Wang J, editor. PLoS One. 2016;11:

547 e0154181. doi:10.1371/journal.pone.0154181

548

549