Genetics inMedicine | ARTICLE

A genotype-first approach to exploring Mendelian cardiovascular traits with clear external manifestations

Brittany M. Wenger, BS[1], Nihir Patel, MS[2], Madeline Lui, BA[1], Arden Moscati, PhD[3], Ron Do, PhD[3], Douglas R. Stewart, MD[4], Marco Tartaglia, PhD[5], Laura Muiño-Mosquera, MD, PhD[6,7], Julie De Backer, MD, PhD[7,8], Amy R. Kontorovich, MD, PhD[2,9] and Bruce D. Gelb, MD[2,10]

Purpose: The purpose of this study is to use a genotype-first approach to explore highly penetrant, autosomal dominant cardiovascular diseases with external features, the RASopathies and Marfan syndrome (MFS), using biobank data.

Methods: This study uses exome sequencing and corresponding phenotypic data from Mount Sinai's Bio*Me* ($n = 32{,}344$) and the United Kingdom Biobank (UKBB; $n = 49{,}960$). Variant curation identified pathogenic/likely pathogenic (P/LP) variants in RASopathy genes and *FBN1*.

Results: Twenty-one subjects harbored P/LP RASopathy variants; three (14%) were diagnosed, and another 46% had ≥1 classic Noonan syndrome (NS) feature. Major NS features (short stature [9.5% $p = 7e\text{-}5$] and heart anomalies [19%, $p < 1e\text{-}5$]) were less frequent than expected. Prevalence of hypothyroidism/autoimmune disorders was enriched compared with biobank populations ($p = 0.007$). For subjects with *FBN1* P/LP variants, 14/41 (34%) had a MFS diagnosis or highly suggestive features. Five of 15 participants (33%) with echocardiographic data had aortic dilation, fewer than expected ($p = 8e\text{-}6$). Ectopia lentis affected only 15% ($p < 1e\text{-}5$).

Conclusions: Substantial fractions of individuals harboring P/LP variants with partial or full phenotypic matches to a RASopathy or MFS remain undiagnosed, some not meeting diagnostic criteria. Routine population genotyping would enable multidisciplinary care and avoid life-threatening events.

Genetics in Medicine (2021) 23:94–102; https://doi.org/10.1038/s41436-020-00973-2

Keywords: exome sequencing; Mendelian disorders; cardiovascular system; genotype–phenotype correlations; precision medicine

INTRODUCTION

Monogenic disorders have traditionally been studied by ascertaining individuals phenotypically and then discovering causal genetic variation. With the decreasing cost of exome sequencing of cohorts from large biobanks, a genotype-first approach whereby individuals are studied based on their genetic variation, irrespective of phenotype, has emerged as a powerful tool.[1] This approach can uncover a wide spectrum of phenotypic heterogeneity and disease severity. Clinically, it may aid in the diagnosis and treatment of patients with mild or nonclassic presentations of monogenic disorders and uncover at-risk family members.[2] Genotype-first studies anticipate an expected future when exome or genome sequencing is routinely performed, perhaps in all newborns.

In this study, we focused on highly penetrant, autosomal dominant Mendelian cardiovascular disorders (CVDs) with external manifestations: the RASopathies and Marfan syndrome (MFS).[3,4] These disorders are well situated for electronic health record (EHR) review as they often present with external manifestations such as height abnormalities that can be captured in patients who have not undergone cardiac workup. Additionally, their relatively high frequency and high penetrance were anticipated to result in more subjects for study than most other Mendelian cardiovascular genetic traits. The RASopathies include Noonan syndrome (NS), NS with multiple lentigines (NSML), Costello syndrome, cardiofaciocutaneous syndrome, and NS with loose anagen hair (NSLAH). They are fully penetrant autosomal dominant disorders with variable expressivity caused by pathogenic variation in genes encoding components of the RAS/mitogen-activated protein kinase (RAS/MAPK) pathway.[5,6] NS is the commonest RASopathy, with an estimated incidence of ~1:1000–2500 live births. Characteristic features include cardiovascular defects (especially pulmonary valve stenosis and hypertrophic cardiomyopathy [HCM]), craniofacial abnormalities, short stature, and intellectual disability.

MFS is a highly penetrant autosomal dominant disorder with an estimated population incidence of 1:5000–10,000 and

[1]Icahn School of Medicine at Mount Sinai, New York, NY, USA; [2]Mindich Child Health and Development Institute, Icahn School of Medicine, New York, NY, USA; [3]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [4]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; [5]Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy; [6]Division of Pediatric Cardiology. Department of Pediatrics, Ghent University Hospital, Ghent, Belgium; [7]Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; [8]Department of Cardiology, Ghent University Hospital, Ghent, Belgium; [9]Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [10]Departments of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine, New York, NY, USA. Correspondence: Bruce D. Gelb (bruce.gelb@mssm.edu)
These authors contributed equally: Amy R. Kontorovich, Bruce D. Gelb

Submitted 11 June 2020; revised 8 September 2020; accepted: 9 September 2020
Published online: 29 September 2020

is caused by pathogenic *FBN1* variation.[4] Key features include aortic aneurysm and dissection, ectopia lentis, and a distinctive body habitus with tall stature. Natural history studies show a mean life expectancy of 30–40 years for individuals diagnosed with MFS,[7] but longevity approaches normal with proper medical and surgical management.[4]

In this genotype-first study of RASopathies and MFS, we determined clinical features and diagnostic rates of individuals harboring pathogenic variants from two large biobanks.

MATERIALS AND METHODS

Study populations

Mount Sinai's Bio*Me* is an ancestrally diverse biobank linking 32,344 patients' exome sequences with EHRs. Patients have been recruited nonselectively from >30 diverse sites across the Mount Sinai Health System since 2007 (91% of recruitment from primary care and unrelated subspecialty sites as well as generic ones such as annual DNA Day fairs; 6.7% of recruitment has been at cardiac care settings. The United Kingdom Biobank (UKBB), which began enrolling subjects in 2006, is a prospective cohort study with sociodemographic, baseline characteristic, and phenotypic data from 502,543 participants aged 40–69 at time of recruitment. A subset of these participants (49,960) have exome sequencing. Individuals selected for exome sequencing were enriched modestly for hospital admission for asthma (unpublished data).

Ethics statement

Informed consent for sequencing, phenotype assessment, and publication of results was obtained at time of enrollment for both Bio*Me* and UKBB biobank participants. Further details are located in the Bio*Me* researcher FAQ (https://icahn.mssm.edu/research/ipm/programs/biome-biobank/researcher-faqs) and UKBB resource catalog (http://biobank.ndph.ox.ac.uk/showcase/catalogs.cgi), respectively. EHR review for this study was approved by the Institutional Review Board (IRB) of the Icahn School of Medicine at Mount Sinai in New York, New York.

Variant identification

We identified variants for genes paired to the RASopathies or MFS with definitive or strong levels of evidence as rated by the Clinical Genomic Resource (ClinGen). We included variants with genotype quality ≥30, read depth ≥7, and heterozygous B-allele frequency (BAF). Due to differing depth of coverage between the two cohorts, we defined the acceptable BAF range for Bio*Me* participants as 0.2–0.8 and as 0.3–0.7 for UKBB participants. For Bio*Me*, variants were confirmed visually using the Integrative Genomics Browser (IGV).[8] All variants of interest were validated with mutalyzer[9] or VariantValidator.[10]

To identify pathogenic or likely pathogenic (P/LP) for RASopathies, two authors with ClinGen RASopathy Expert Curation Panel expertise (B.D.G. and M.T.) curated all nonsynonymous variants in *BRAF*, *CBL*, *KRAS*, *NRAS*, *PTPN11*, *RAF1*, *RRAS*, *RRAS2*, *RIT1*, *SHOC2*, *SOS1*, *SOS2*. For this curation, they used the ClinGen RASopathy-specific refinement of the American College of Medical Genetics and Genomics–Association for Molecular Pathology (ACMG-AMP) guidelines.[11] Two RASopathy genes, *HRAS* and *LZTR1*, had no available data in the BioMe exome sequencing set so were excluded. P/LP variants were then classified as either definitively or likely causal. Definitively causal P/LP variants were ones previously associated with RASopathies; likely causal ones were P/LP variants that have not been previously observed but, based on their positions in areas of the protein known to harbor multiple pathogenic variants and available information about protein structure and function, were expected to engender gain of function in a manner similar to nearby definitively causal variants.

For MFS, variant selection was based on two recent studies[12,13] that refined accuracy relative to the revised Ghent MFS variant criteria.[14] Two authors with ClinGen expertise (B.D.G. and L.M.-M., the latter being on the *FBN1* Variant Curation Expert Panel) curated all *FBN1* variants. We included splicing variants only within 1–2 base pairs of a canonical splice donor or acceptor.[13,15–17] Variants had to be appropriately rare (minor allele frequency [MAF] $\leq 10^{-4}$) in population databases with computational lines of evidence in support of pathogenicity.[18]

Phenotype assessment

For Bio*Me* subjects, height *Z*-scores were calculated using subjects' age, sex, and self-reported ethnicity based on averages and standard deviations for individuals matching in age, sex, and ethnicity across the entire Bio*Me* population. For the UKBB, we calculated height *Z*-score based on averages and standard deviations for individuals with matching gender across the entire UKBB population.

For inclusion of Bio*Me* subjects, EHRs needed to contain sufficient encounters for analysis, defined as an established patient with a primary care physician, encounter with an ambulatory cardiologist or encounters with at least two other specialists. EHRs of Bio*Me* subjects harboring P/LP variants and meeting these criteria were reviewed, including progress and consult notes, operative reports, and reports from diagnostic procedures, to identify traits known to be associated with either RASopathies or MFS. For UKBB participants harboring P/LP variants, we analyzed International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) diagnosis codes, surgical procedure codes, hearing loss assessments, ophthalmologic assessments, and self-reported medical and surgical histories for traits related to RASopathies or MFS.

RASopathies

EHRs (Bio*Me*) or recorded data points (UKBB) of subjects with P/LP RASopathy variation were assessed for associated features including height, ophthalmologic abnormalities, congenital heart disease, QRS axis in the frontal plane on most recent electrocardiogram,

echocardiographic data including left ventricular (LV) posterior wall and interventricular septal thickness at end-diastole, learning disabilities, undescended testes for males, infertility, personal history of malignancy, hydronephrosis, bleeding abnormalities, dysmorphic facial features, abnormal pigmentation, relevant family history, skeletal deformities, documented hearing loss, and evidence of Chiari malformation when available.

MFS

EHRs (Bio*Me*) or available database entries (UKBB) of participants with P/LP *FBN1* were assessed for height, ectopia lentis (EL), myopia or other ophthalmologic abnormalities, abnormal aorta findings on any available imaging study, mitral valve prolapse, spinal curvature, dural ectasia, skin striae, protusio acetabuli, pes planus, and relevant family history. For Bio*Me* individuals who had echocardiographic data in their EHRs, we recorded age, body surface area, height, and aortic diameter measurements at end-diastole in the parasternal long axis view using the leading-edge-to-leading-edge technique at the sinus of Valsalva (SOV), tubular ascending aorta and aortic arch. SOV *Z*-scores were determined based on participants' body surface area, age, and sex.[19]

Statistical methods

Binomial tests were used to assess whether the rates of observed features differed from the prevalence of the feature in the biobank at large or expected rates for the syndrome based on published literature. For *Z*-scores, *t*-tests were used to test for significance. For all testing, $p < 0.05$ was used as the threshold for significance.

RESULTS

RASopathies

We identified 15 unique P/LP RASopathy variants in Bio*Me* participants and seven unique variants in UKBB participants (Supplementary Table 1). For the commonest RASopathy gene, *PTPN11*, there were ten unique P/LP variants in Bio*Me* participants with one, p. M504V, found in two individuals. Of these ten variants, all but one was NS-associated (T468M is NSML-associated). Furthermore, seven were rated as definitely causal, and three were rated likely causal. Additionally, 1–2 P/LP variants were identified in each of four other RASopathy genes (*KRAS*, *NRAS*, *RAF1*, and *RIT1*). While all of these variants were classified as definitely causal, individuals harboring *KRAS* p. G12D and *NRAS* p. G12D variants were excluded. Both alleles are associated with severe RASopathy phenotypes including early lethality when inherited germline and have been observed as somatic variants in cancer.[20,21] Based on the diagnoses of myelofibrosis documented in two of three individuals' EHRs, these variants were deemed most likely due to somatic changes in hematopoietic malignancy.

Among the UKBB participants, we identified four *PTPN11*, two *SOS1*, and one *SHOC2* P/LP variants. Three *PTPN11* and both *SOS1* variants are associated with NS, while the fourth *PTPN11* variant is associated with NSML and the *SHOC2* variant causes NSLAH.

In total, 21 participants (14 from Bio*Me*, 7 from UKBB) were found to harbor a P/LP RASopathy variant. Their ages and ethnicities were comparable with those of the overall biobank populations (Table 1). Only three, all with underlying *PTPN11* variation and from Bio*Me*, had a diagnosis of NS. Of these individuals, two were born with congenital heart defects and the third was diagnosed after her daughter was born with a congenital heart defect. For the remaining 18 participants without a RASopathy diagnosis, robust application of NS clinical criteria[22] was not possible due to limitations in capturing facial or sternal features from EHR review or ICD codes. Nonetheless, many, but not all, undiagnosed participants seemingly met the requirements for a diagnosis of NS. Six of the 11 undiagnosed participants from Bio*Me* fulfilled the major cardiac criterion, most due to findings on electrocardiograms (ECGs) but also one with HCM.

Including diagnosed individuals, 4 of 21 individuals (19%, all from Bio*Me*) were diagnosed with congenital heart disease or HCM, less than the expected 80–90% reported rate of cardiac involvement in RASopathies ($p < 0.00001$).[23]

Table 1 Demographic data.

	Bio*Me* all	Bio*Me* RASopathies	Bio*Me* *FBN1*	UKBB all	UKBB RASopathies	UKBB *FBN1*
Age (years), mean (range)	58	53 (25–80)	55 (31–84)	57 (40–69)	60 (46–69)	58 (46–68)
Female	58%	67%	75%	54%	43%	47%
Height *Z*-score (*p* value)	0 ± 1	−0.8 (0.0013)	0.8 ± 1.2 (0.0001)	0 ± 1	−0.3 ± 1.4 (0.43)	0.6 ± 1.2 (0.013)
Fluid intelligence *Z*-score (*p* value)	n/a	n/a	n/a	0 ± 1	−0.7 ± 1.0 (0.08)	0.3 ± 1.2 (0.39)
Ethnicity						
White/European	31%	20%	25%	94%	86%	82%
African American/Black	24%	33%	25%	2%	0%	12%
Asian	2%	0%	8%	2%	0%	6%
Hispanic/Latin American	35%	33%	41%	n/a	n/a	n/a
Other	8%	12%	0%	1%	14%	0%

n/a not available, *UKBB* UK Biobank.

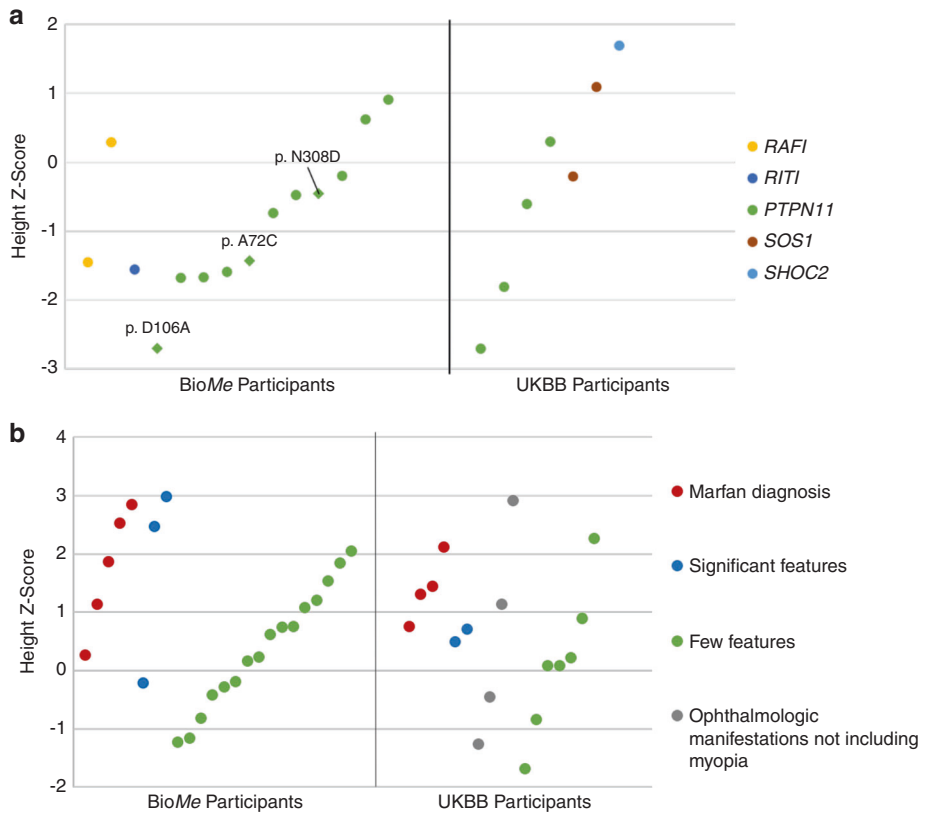

Fig. 1 Height of participants with pathogenic/likely pathogenic (P/LP) variants. (**a**) RASopathy genes. Diamonds, individuals diagnosed with Noonan syndrome; circles, undiagnosed individuals. As indicated in the key, colors correspond to the genes harboring the P/LP variants. (**b**) *FBN1*. As indicated in the key, colors correspond to the clinical status of the individuals. Significant features include ≥2 of the following: aortic dilation, height *Z*-score >2, ectopia lentis, retinal detachment, congenital skeletal abnormalities, family history of Marfan syndrome (MFS) phenotype, nonrheumatic mitral valve abnormalities. *UKBB* UK Biobank.

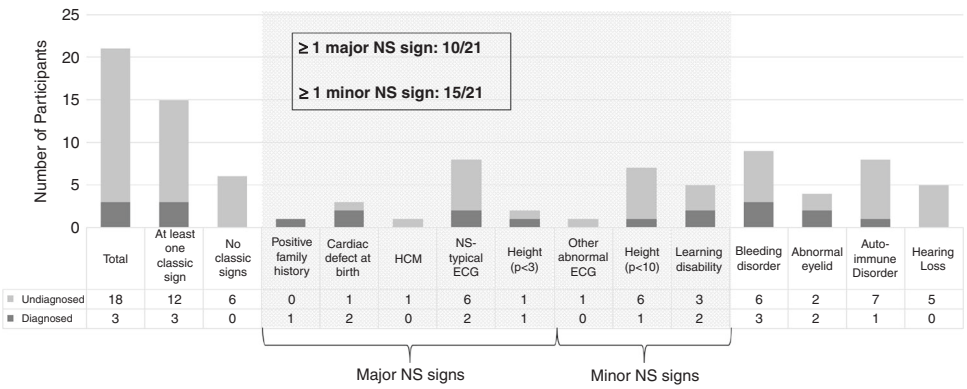

Fig. 2 Phenotypes of individuals with underlying pathogenic/likely pathogenic RASopathy variation by diagnosis status. The major and minor signs for diagnosing are per van der Burgt.[22] *ECG* electrocardiogram, *HCM* hypertrophic cardiomyopathy, *NS* Noonan syndrome.

Average height *Z*-scores were −0.86 and −1.2 in Bio*Me* and UKBB participants harboring P/LP RASopathy variants, respectively, significantly lower than biobank populations ($p = 0.0013$ and $p = 0.016$, respectively) (Fig. 1a). Short stature (height below the third percentile) was found in 2 of 21 (9.5%) participants harboring P/LP RASopathy variants, of which one was undiagnosed with a RASopathy (Fig. 2). An additional six participants fulfilled the minor height criterion for NS (<10th percentile), of whom five were undiagnosed. Neither individual with an underlying *SOS1* P/LP variant had short stature ($Z = -0.2$ and 1.1), consistent with the *SOS1* NS phenotype.[6]

Three individuals had a learning disability, a minor criterion for NS. The fluid intelligence *Z*-score for the UKBB RASopathy cohort trended low (mean of −0.66; $p = 0.08$). Although *SOS1*-related NS is associated with near-normal cognitive function, both UKBB individuals with P/LP variants in this gene had below average fluid intelligence on exam (*Z*-scores of −0.5 and −1.8).

Within Bio*Me*, at least one major or minor sign of NS was detected in 8 of the 11 (73%) undiagnosed individuals with relevant P/LP variants (Fig. 2). The three not meeting any criteria for NS did exhibit some associated features associated including bleeding abnormalities. Only one of seven participants from the UKBB had electrocardiographic (ECG) or echocardiographic data available, limiting our ability to assess cardiac phenotypes. Nevertheless, four of seven individuals exhibited at least one minor NS sign.

In the Bio*Me* RASopathy cohort, 9 of 14 participants had hematological abnormalities, including storage pool disease, thrombocytopenia, and prolonged bleeding time after surgeries. While UKBB participants with P/LP RASopathy variants did not have increased prevalence of bleeding abnormalities, they did demonstrate increased rates of hearing loss when compared with the biobank at large (71% vs. 29%, $p = 0.01$).

A notable finding replicated in both cohorts was a high prevalence of hypothyroidism and other autoimmune disorders. In Bio*Me*, 5 of 14 (36%) participants harboring P/LP RASopathy variants were affected by at least one autoimmune-related disorder (hypothyroidism/Hashimoto disease, 2; hypothyroidism plus discoid lupus, 1; systemic lupus erythematosus [SLE], 2), greater than the prevalence of 17% for autoimmune disorders in Bio*Me* overall. In the UKBB RASopathy cohort, we observed three individuals (43%) with autoimmune disorders (psoriasis, 1; hypothyroidism, 2). Combining the two biobank cohorts, 8 of 21 individuals had autoimmune disorders, significantly more than expected ($p = 0.007$).

Finally, myelofibrosis is a rare disorder, estimated to affect 0.5–1.5 in 100,000 individuals. We observed one individual, harboring *SOS1* G434R, who was diagnosed with myelofibrosis and a second, harboring *RIT1* A41G, with an unspecified blood dyscrasia. Of note, that *SOS1* variation is not present in the COSMIC cancer database so is unlikely to represent a somatic variant. The cardiac phenotype and stature of the individual with blood dyscrasia were consistent with NS. Myelofibrosis has been associated with *SHOC2* germline NSLAH,[24] so may be an emerging age-related manifestation of RASopathies more broadly.

MFS

We identified 21 unique *FBN1* variants meeting P/LP criteria in Bio*Me* participants and 14 unique variants among UKBB participants (Supplementary Table 2). Among them, 3 variants interrupted splice sites, 12 created cysteine residues in cb-EGF or TB domains, 7 removed cysteine residues in cb-EGF or TB domains, 12 altered critical amino acids, and 1 had significant clinical evidence for pathogenicity in ClinVar.

Overall, 44 participants (Bio*Me*, 27; UKBB, 17) harbored P/LP *FBN1* variants. Three Bio*Me* subjects had insufficient EHR encounters so were excluded from phenotypic analyses. Demographics for individuals with P/LP variation are shown in Table 1.

Of the 41 subjects harboring P/LP *FBN1* variants with phenotype data, 9 (22%) were diagnosed with MFS. While there were insufficient clinical data available to apply the revised Ghent criteria[14] for MFS for any of the 32 undiagnosed participants based on EHR review or biobank data, five participants had a clinical picture strongly suggestive of undiagnosed MFS (Fig. 3). One Bio*Me* participant had a family history of aortic aneurysm in both her mother and sister, personal history of thoracoabdominal aneurysm repair, and thickened mitral valve leaflets with a ruptured chorda. A pair of participants whom we concluded were closely related based on chart review, most likely mother and daughter, shared the same underlying genotype and expressed a MFS-like phenotype. The presumed mother had a height *Z*-score of 3.0 and pes planus. Her presumed daughter had a height *Z*-score of 2.5 and scoliosis. While SOV was not significantly dilated for either participant (*Z*-scores of 1.33 and 0.23 respectively), both had dilated ascending aortas above the 95th percentile for age. In the UKBB, one individual had nonrheumatic mitral valve disease, a musculoskeletal disorder, and a father who died at age 30. Another had EL.

Average height *Z*-scores for Bio*Me* and UKBB subjects with P/LP *FBN1* variants were 0.8 ± 1.2 and 0.6 ± 1.2, respectively, significantly taller than expected ($p = 0.0001$ and $p = 0.013$, respectively) (Fig. 1b). Across both cohorts, 8 of 41

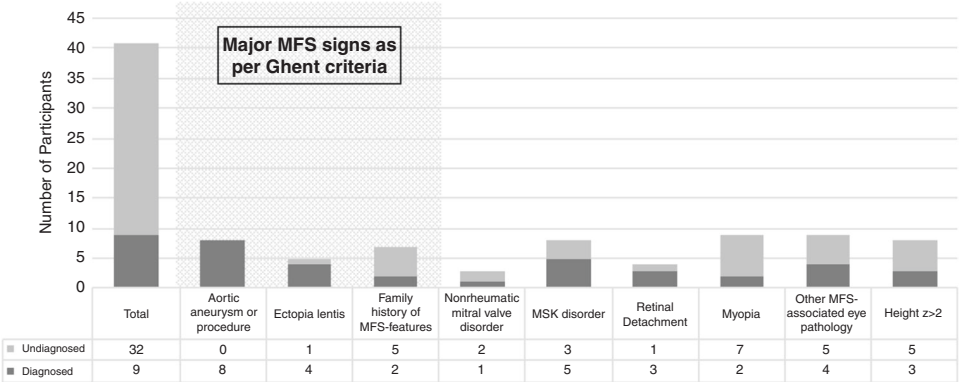

Fig. 3 Phenotypes of individuals with underlying pathogenic/likely pathogenic *FBN1* variation by diagnosis status. The diagnostic features are per the revised Ghent criteria. *MFS* Marfan syndrome, *MSK* musculoskeletal.[14]

Table 2 Prevalence of MFS-associated features in the UKBB *FBN1* cohort.

Phenotypic manifestation	*FBN1* cohort	UKBB	*p* value
MFS diagnosis	23%	0.02%	2e-15
Aortic root aneurysm	18%	0.2%	5e-6
Mitral valve abnormality	6%	0.2%	0.04
Congenital lens malformation	12%	0.02%	5.4e-9
Blindness and low vision	12%	0.4%	0.002
Retinal detachments and breaks	18%	0.5%	8e-5
Other retinal disorders	18%	1.6%	0.002
Dislocation of the lens	18%	0.02%	4e-12
Myopia	47%	0.5%	9e-15
Glaucoma	29%	2.5%	5e-06
Presenile cataract	18%	n/a	n/a

MFS Marfan syndrome, *n/a* not available, *UKBB* UK Biobank.

participants (3 with an MFS diagnosis) had height *Z*-scores >2.0.

Echocardiographic data were available for 15 Bio*Me* subjects (mean age 56 years, range 31–84). Five (33%) of these participants, all diagnosed with MFS, had aortic root aneurysm or past thoracic aortic surgery, statistically significantly fewer than the expected prevalence of 85% in MFS patients aged 51–60 ($p = 8e\text{-}6$).[25] Two of 15 subjects with echocardiographic data had mitral valve involvement, the previously noted one with mitral valve leaflet thickening and a rupture chorda and another with mitral valve prolapse. Among the 17 UKBB individuals with *FBN1* P/LP variants (mean age, 57 years), only 3 (18%) had recorded aortic aneurysm, dissection, or past aortic surgery, also significantly fewer than expected in MFS ($p = 0.0002$) (Table 2). One individual had an ICD code for a mitral valve abnormality.

EL, a seminal feature of MFS and major Ghent criterion, was noted in 2 of 24 Bio*Me* participants (8%) and 4 of 17 UKBB participants (23%) harboring P/LP *FBN1* variants, statistically fewer than expected (*p* values of <0.0001 and 0.03, respectively), based on EL's estimated prevalence in MFS of 50%.[26]

In the UKBB, there was a high incidence of non-EL, MFS-associated ophthalmologic abnormalities, including myopia, glaucoma, early-onset cataracts, retinal detachment, and other lens and retinal pathologies (Table 2).[27] Myopia (undiagnosed, 6; MFS diagnosis, 2) and glaucoma (undiagnosed, 2; MFS diagnosis, 3) were found at a statistically significant increased rates compared to the entire UKBB (*p* values of <0.0001 and < 0.00001, respectively). Overall, 8 of the 14 undiagnosed UKBB participants had at least one ophthalmologic feature associated with MFS. Five of those had at least one nonmyopia MFS-related phenotype, including two with retinal pathology and two with presenile cataracts (occurring at ages 17 and 33 years). Finally, two individuals in UKBB had a parent who died at least two standard deviations below the average age of death for their gender.

Genotype-first versus phenotype-first syndrome prevalence

We identified 15 individuals in Bio*Me* harboring RASopathy-associated P/LP variants out of 32,344 enrollees with exome sequencing data (1 in 2156), consistent with population estimates for NS of 1 in 1000–2500 individuals. Separately, the diagnosis code that includes NS (ICD-10 Q87.19; eight disorders with NS being commonest) was linked to 13 of the 51,692 Bio*Me* enrollees (1 in 3976). In the UKBB, only 7 of 49,960 participants were found to harbor RASopathy pathogenic variation (1 in 7137), significantly fewer than expected ($p = 0.021$). The higher rate of likely causal RASopathy variants in Bio*Me* relative to the UKBB ($p = 0.011$) is probably attributable to the former's cohort being drawn from a health-care system while the latter has a healthy volunteer bias. Indeed, the UKBB only has 21 individuals among its 502,000 participants with a Q87.1 ICD code, which includes Q87.19 plus Prader–Willi syndrome (1 in 23,905 subjects).

Twenty-seven Bio*Me* participants (1 in 1198) harbored P/LP *FBN1* variation. Even using the upper limit for MFS prevalence of 1 in 5000, the prevalence is significantly higher than expected ($p = 0.0006$). The diagnosis codes that include MFS (ICD-10 Q87.40, Q87.410, and Q87.418) were linked to 15 Bio*Me* participants (1 in 3446). In the UKBB, 17 individuals (1 in 2939) had P/LP *FBN1* variants, somewhat higher than the 10 in 49,960 that would be expected if the UKBB subject were representative of the population ($p = 0.25$). Interestingly, 96 UKBB participants (1 in 5229) bore an ICD code for MFS, consistent with the expected population frequency of 1 in 5000–10,000.

DISCUSSION

Using a genotype-first approach to ascertain individuals within biobanks at risk for a RASopathy or MFS, we found that a minority carrying associated pathogenic variation were diagnosed with these presumed highly penetrant CVDs. Only 14% of participants with causative RASopathy variation were assigned a syndrome diagnosis. An additional 46% of this cohort exhibited at least one classic NS feature, many of whom would likely receive a RASopathy diagnosis upon focused clinical examination. Among individuals harboring P/LP *FBN1* variants, only 22% were diagnosed with MFS but an additional 12% bore sufficient phenotype features for an MFS diagnosis. Unsurprisingly, higher rates of pathogenic variation were observed for both syndromes in Bio*Me* compared with the UKBB, likely reflecting differences between health-care-centric versus population-based data sets. Previous CVD genetics-first studies have focused on cardiomyopathies, where the diagnosis rates have varied widely,[28–30] due to lack of external features and incomplete penetrance. That a minority of individuals in our study were diagnosed despite external features is revealing.

Genotype-first approaches depend upon the a priori likelihood that variants assessed as P/LP are truly disease-causing, which varies with disease penetrance and variant

curation accuracy. The RASopathies are ideal because their pathogenic mechanism necessitates functional changes through specifically positioned missense substitutions, limiting the universe of possible pathogenic variants. As a result, nearly all causal variants are known or readily discerned, providing high confidence for the variants retained for this study. Despite this, few participants harboring relevant variants were diagnosed with a RASopathy. As many were born before or not long after NS was first described in 1963,[31] their physicians were likely unaware of the diagnosis during their early childhood when the condition is most readily recognized.[3] Pathogenic variants in most RASopathy genes were not observed for several proposed reasons: lack of genotype data for Costello syndrome (*HRAS*), rarity of most RASopathies, survival bias, and, for the UKBB, reliance on healthy volunteers.

Predicting pathogenic variation is more challenging for *FBN1*. Considering our findings, some retained *FBN1* variants may not be disease-causing. The high prevalence of P/LP *FBN1* variants relative to expectation in both biobanks suggests the inclusion of some false positives. Moreover, the rates of likely MFS varied with variant type. Variants that created cysteine residues in important *FBN1* regions had a particularly low MFS/likely MFS yield (1 of 17). Our experience with MFS highlights the value of variant curation accuracy for genotype-first approaches.

The RASopathies have a 80–90% cardiac involvement, but the prevalence of cardiovascular diagnoses among individuals with RASopathy P/LP variation was notably low (Bio*Me*, 4 of 15; UKBB, 0 of 7). Pulmonary valve stenosis and atrial septal defects, which have nearly normal survival, are the most prevalent congenital heart defects in NS; severe heart defects with poor survival are exceptional in NS. While HCM can be lethal in NS, its prevalence is only 20%. Thus, the low cardiac diagnosis rate among individuals harboring RASopathy causal variants is probably attributable to NS phenotypic heterogeneity, with low clinical suspicion for cases without CVD manifestations and a higher likelihood of individuals with minimal or no cardiac involvement surviving to adulthood. The findings with height were similar. While the average height of the RASopathy pathogenic allele-bearing group was less than the population, only 10% of these individuals exhibited short stature, less than the expected 40–50%.[32] While ~60% of subjects with a P/LP RASopathy variants had at least one clinical feature related to NS, most would be insufficient to raise suspicion for NS. The added insight of genotype would presumably prompt an NS-focused workup, potentially uncovering undiagnosed medical issues.

Incomplete phenotype information likely plagued our ability to fully characterize subjects with P/LP *FBN1* variants. Without echocardiographic data, absent for all UKBB participants and a substantial fraction of Bio*Me* subjects, assessment of MFS-related cardiovascular involvement proved unsatisfactory. Despite that, variable expressivity in MFS was evident in these data, particularly when comparing individuals with the same P/LP *FBN1* genotype. For example, three undiagnosed Bio*Me* individuals with *FBN1* p.A2620P were all tall (height Z-score range = 1.8–3.0). Two had at least one other skeletal manifestation and aortic dilation, whereas the third had no documented cardiac or skeletal abnormality despite multiple health encounters. Two UKBB individuals harbored *FBN1* p.R240C variants. One was diagnosed with MFS with a height Z-score of 0.8 and eye involvement (EL, retinal detachment, and severe myopia). The other was undiagnosed with a height Z-score of −0.4 but was postprocedural for eye surgery and had dilated cardiomyopathy, which can be observed in MFS.[33]

Another possible explanation for the low diagnosis rates is that the penetrance of RASopathies and MFS is incomplete, with the current tenet of complete and high penetrance, respectively, having been skewed by ascertainment bias. The issues with potential false positive *FBN1* P/LP variants and some phenotype data limitations notwithstanding, it is striking that 51% of individuals harboring these variants had no evidence for cardiac, ophthalmologic, or skeletal abnormalities associated with MFS and 76% had a height Z-score <1.5. Moreover, these participants were predominantly between 55 and 68 years old, which is beyond the average survival for individuals recognized to have MFS who go untreated.[4] Similarly, 29% of individuals with underlying P/LP RASopathy-associated variation had no discernible major or minor features of NS and 67% had a height Z-score > −1.5. Clinically, it is not unusual to diagnose NS or MFS in a first-degree relative, even with minor features, in part because the positive family history contributes to the weight of evidence. Those same individuals had already gone unnoticed based on their own phenotypes alone. The observation of less severe phenotypes raises the possibility that gene modifiers contribute to this broader range of variable expressivity. If one redefines disease penetrance as a causal variant plus any minor feature associated with the relevant disorder, then it is possible that penetrance could reach completeness. Without broad exome or genome sequencing, even the most astute clinicians would not be able to achieve a high diagnostic rate among sporadic cases.

Implementation of a genotype-first approach using biobanks with adult subjects for disorders principally characterized in children and young adults enables observation of features more prevalent later in life. Indeed, we observed hypothyroidism and autoimmune diseases in 38% of individuals harboring P/LP RASopathy-associated variants. Hypothyroidism is observed in 4% of children with NS.[34] This is likely attributable to autoimmune thyroiditis, based on case reports and increased rates of antithyroid antibodies.[35–37] Relatedly, other autoimmune disorders, particularly SLE, have been described in individual patients with NS.[36–39] In one pediatric and young adult NS series, 14% had an autoimmune disorder but autoantibodies were present in 30%.[40] Thus, our results show that thyroid disease and autoimmune disorders become more prevalent in the RASopathies with advancing age, suggesting that surveillance should be incorporated into the multidisciplinary care of these patients.

Overall, our findings suggest that the defined seminal features of both MFS and the RASopathies may be less prevalent in affected individuals than previously thought, adversely impacting the likelihood of receiving a diagnosis. As a result, these undiagnosed individuals are not privy to multidisciplinary care, which can extend lifespan and improve health. As variant curation for the relevant genes, particularly for *FBN1*, continues to become more accurate, genotype-first identification of individuals affected with these disorders is likely to predominate. This, in turn, will enable a time when universal exome or genome sequencing is offered as part of routine health care, perhaps during the newborn period.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-020-00973-2) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We acknowledge the participants enrolled in Bio*Me* and UKBB, without whom this research would not be possible. We also acknowledge the ongoing work of the Mount Sinai's Charles Bronfman Institute for Personalized Medicine for ongoing curation of Bio*Me*. This work was supported in part by United States Public Health Service (USPHS) grant to R.D. (GM124836 and HL139865), B.D.G. (HL135742), and A.R.K. (HL140083) and grants from Italian Association for Cancer Research (IG21614), European Joint Programme on Rare Diseases (NSEuroNet) to M.T. J.D.B is supported as senior clinical researcher by the Research Foundation Flanders and holds a Grant for Medical Research from the Baillet Latour Funds. Research reported in this paper was supported by the Office of Research Infrastructure of the National Institutes of Health under award numbers S10OD018522 and S10OD026880. This work was also supported in part through the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland. This research has been conducted using the UK Biobank Resource under application number 16218.

DISCLOSURE

B.D.G. and M.T. receive royalties for genetic testing of the RASopathies from Correlegan, GeneDx, LabCorp, and Prevention Genetics. D.R.S. performs contract telegenetics services for Genome Medical, Inc., in accordance with relevant National Cancer Institute ethics policies. The other authors declare no conflicts of interest.

REFERENCES

1. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. Cell. 2014;156:872–877.
2. Alver M, et al. Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. Genet Med. 2019;21:1173–1180.
3. Roberts AE, Allanson JE, Tartaglia M, Gelb BD. Noonan syndrome. Lancet. 2013;381:333–342.
4. Pyeritz RE. Recent progress in understanding the natural and clinical histories of the Marfan syndrome. Trends Cardiovasc Med. 2016;26:423–428.
5. Aoki Y, Niihori T, Inoue S-i, Matsubara Y. Recent advances in RASopathies. J Hum Genet. 2016;61:33–39.
6. Tartaglia M, Gelb BD, Zenker M. Noonan syndrome and clinically related disorders. Best Pract Res Clin Endocrinol Metab. 2011;25:161–179.
7. Von Kodolitsch Y, Raghunath M, Nienaber C. Marfan syndrome: prevalence and natural course of cardiovascular manifestations. Z Kardiol. 1998;87:150.
8. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–192.
9. Wildeman M, Van Ophuizen E, Den Dunnen JT, Taschner PE. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum Mutat. 2008;29:6–13.
10. Freeman PJ, Hart RK, Gretton LJ, Brookes AJ, Dalgleish R. VariantValidator: accurate validation, mapping, and formatting of sequence variation descriptions. Hum Mutat. 2018;39:61–68.
11. Gelb BD, et al. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. Genet Med. 2018;20:1334–1345.
12. Baudhuin LM, Kluge ML, Kotzer KE, Lagerstedt SA. Variability in gene-based knowledge impacts variant classification: an analysis of FBN1 missense variants in ClinVar. European J Hum Genet. 2019;27:1550–1560.
13. Muiño-Mosquera L, et al. Tailoring the American College of Medical Genetics and Genomics and the Association for Molecular Pathology Guidelines for the Interpretation of Sequenced Variants in the FBN1 Gene for Marfan Syndrome: proposal for a disease-and gene-specific guideline. Circ Genom Precis Med . 2018;11:e002039.
14. Loeys BL, et al. The revised Ghent nosology for the Marfan syndrome. J Med Genet. 2010;47:476–485.
15. Desmet F-O, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37:e67–e67.
16. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol. 1997;4:311–323.
17. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11:377–394.
18. Kopanos C, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35:1978.
19. Devereux RB, et al. Normal limits in relation to age, body size and gender of two-dimensional echocardiographic aortic root dimensions in persons≥ 15 years of age. Am J Cardiol. 2012;110:1189–1194.
20. Santos FP, et al. Prognostic impact of RAS-pathway mutations in patients with myelofibrosis. Leukemia. 2020;34:799–810.
21. Lo F-S, et al. Noonan syndrome caused by germline KRAS mutation in Taiwan: report of two patients and a review of the literature.Eur J Pediatr. 2009;168:919.
22. Van Der Burgt I, et al. Clinical and molecular studies in a large Dutch family with Noonan syndrome. Am J Med Genet. 1994;53:187–191.
23. Prendiville TW, et al. Cardiovascular disease in Noonan syndrome. Arch Dis Child. 2014;99:629–634.
24. Gripp KW, et al. Expanding the SHOC2 mutation associated phenotype of Noonan syndrome with loose anagen hair: structural brain anomalies and myelofibrosis. Am J Med Genet A. 2013;161:2420–2430.
25. Rybczynski M, et al. Frequency and age-related course of mitral valve dysfunction in the Marfan syndrome. Am J Cardiol. 2010;106:1048–1053.
26. Lacro RV, et al. Characteristics of children and young adults with Marfan syndrome and aortic root dilation in a randomized trial comparing atenolol and losartan therapy. Am Heart J. 2013;165:828–35. e3.
27. Nahum Y, Spierer A. Ocular features of Marfan syndrome: diagnosis and management. Isr Med Assoc J. 2008;10:179.
28. Haggerty CM, et al. Genomics-first evaluation of heart disease associated with titin-truncating variants. Circulation. 2019;140:42–54.
29. Haggerty CM, et al. Electronic health record phenotype in subjects with genetic variants associated with arrhythmogenic right ventricular cardiomyopathy: a study of 30,716 subjects with exome sequencing. Genet Med. 2017;19:1245–1252.
30. Park J. et al. A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. Genet Med. 2020;22:102–111.

31. Noonan JA. Associated noncardiac malformations in children with congenital heart disease. J Pediatr. 1963;63:468–470.
32. Noonan JA, Raaijmakers R, Hall B. Adult height in Noonan syndrome. Am J Med Genet A. 2003;123:68–71.
33. Alpendurada F, et al. Evidence for Marfan cardiomyopathy. Eur J Heart Fail. 2010;12:1085–1091.
34. Sharland M, Burch M, McKenna W, Paton M. A clinical study of Noonan syndrome. Arch Dis Child. 1992;67:178–183.
35. Lee MJ, et al. Hashimoto thyroiditis with an unusual presentation of cardiac tamponade in Noonan syndrome. Korean J Pediatr. 2016;59: S112.
36. Amoroso A, et al. The unusual association of three autoimmune diseases in a patient with Noonan syndrome. J Adolesc Health. 2003;32:94–97.
37. Svensson J, Carlsson A, Ericsson UB, Westphal O, Ivarsson SA. Noonan's syndrome and autoimmune diseases. J Pediatr Endocrinol Metab. 2003;16:217–218.
38. Uehara T, Hosogaya N, Matsuo N, Kosaki K. Systemic lupus erythematosus in a patient with Noonan syndrome-like disorder with loose anagen hair 1: more than a chance association. Am J Med Genet A. 2018;176:1662–1666.
39. Bader-Meunier B, et al. Are RASopathies new monogenic predisposing conditions to the development of systemic lupus erythematosus? Case report and systematic review of the literature. Semin Arthritis Rheum. 2013;43:217–219.
40. Quaio CR, et al. Autoimmune disease and multiple autoantibodies in 42 patients with RASopathies. Am J Med Genet A. 2012;158:1077–1082.