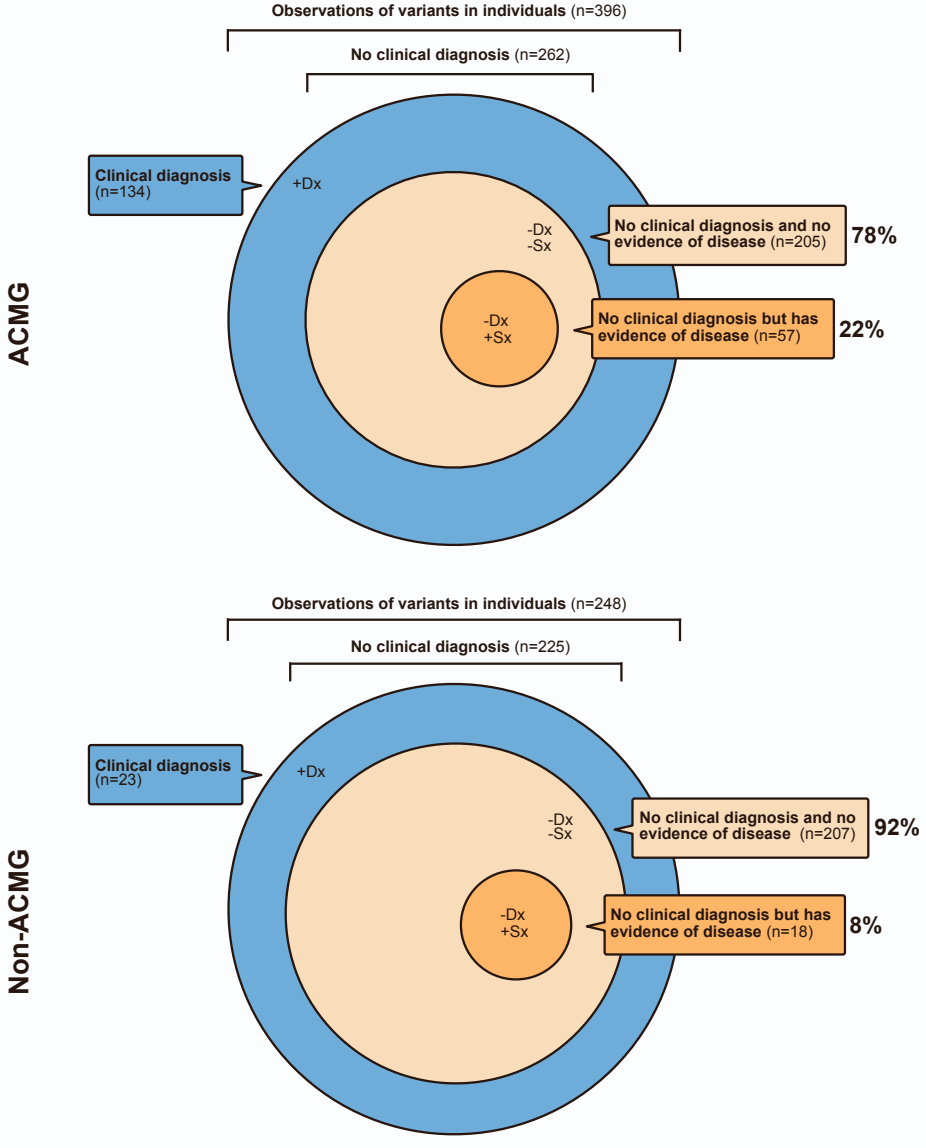
Cell Reports Medicine, Volume *5*

Supplemental information

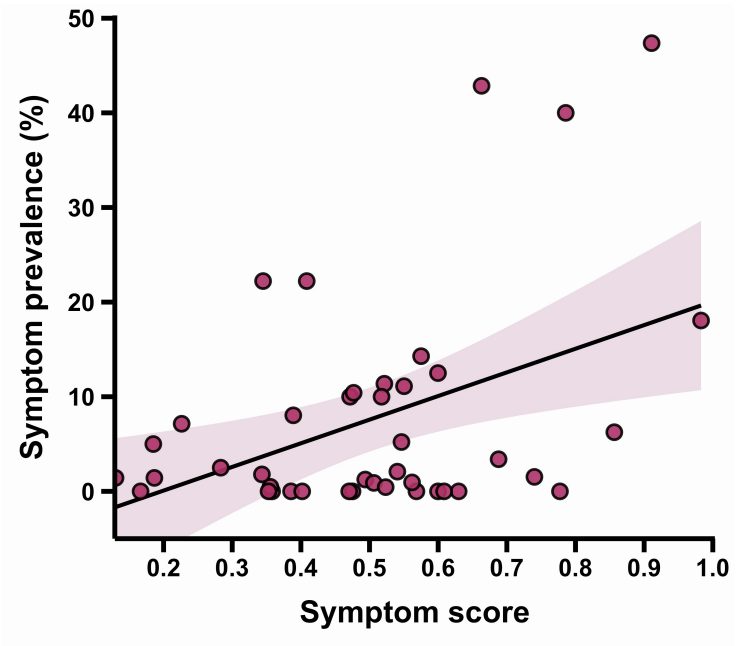
Genome-first evaluation with exome sequence and clinical data uncovers underdiagnosed genetic disorders in a large healthcare system

Iain S. Forrest, Áine Duffy, Joshua K. Park, Ha My T. Vy, Louis R. Pasquale, Girish N. Nadkarni, Judy H. Cho, and Ron Do

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Clinical diagnosis and phenotypes of observations in individuals of pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants in genes from the American College of Medical Genetics and Genomics' Recommendations for Reporting of Secondary Findings (ACMG SF v3.1). Observations of variants in individuals were categorized into three distinct groups: 1) clinically diagnosed (+Dx), the individual with a P/LP/LoF variant had a prior clinical diagnosis for the genetic disorder corresponding to the variant; 2) no clinical diagnosis and no evidence of disease (-Dx -Sx), the individual with a P/LP/LoF variant lacked a prior clinical diagnosis for the corresponding genetic disorder and had no electronic health record (EHR) evidence of disease symptoms or findings detected; and 3) no clinical diagnosis but has evidence of disease (-Dx +Sx), the individual with a P/LP/LoF variant lacked a clinical diagnosis for the corresponding genetic disorder, but has EHR evidence of disease symptoms or findings. Related to Figure 2.



Supplementary Figure 2. **Prevalence of symptoms detected in individuals without a clinical diagnosis carrying pathogenic/likely pathogenic or loss-of-function (P/LP/LoF) variants.** P/LP variants, reported in ClinVar with a minimum review status of two; LoF variants, variants with predicted LoF consequence annotated by Variant Effect Predictor; clinical diagnosis, clinical diagnosis of disease corresponding to a variant; symptom prevalence, prevalence of symptom detected in observations of variants in individuals without a clinical diagnosis; symptom score, on an ordinal scale ranging from 0 (lowest risk of hospitalization and death) to 1 (highest risk of hospitalization and death) from a previously published study[1]. Related to Figure 4.

1. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. *J Med Internet Res* 2015; **17**. DOI:10.2196/jmir.3962.

SUPPLEMENTARY TABLES

Supplementary Table 1. Summary of 54 monogenic disease predisposition genes containing pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants including a subset in the American College of Medical Genetics and Genomics' Recommendations for Reporting of Secondary Findings list of 78 genes (ACMG SF v3.1). Related to Table 2.

Disease	Gene	ACMG SF v3.1
Amyotrophic lateral sclerosis	*FIG4*	FALSE
Amyotrophic lateral sclerosis	*FUS*	FALSE
Amyotrophic lateral sclerosis	*OPTN*	FALSE
Amyotrophic lateral sclerosis	*SETX*	FALSE
Cardiomyopathy	*ABCC9*	FALSE
Cardiomyopathy	*ACTN2*	FALSE
Cardiomyopathy	*CSRP3*	FALSE
Cardiomyopathy	*DES*	TRUE
Cardiomyopathy	*DSG2*	TRUE
Cardiomyopathy	*DSP*	TRUE
Cardiomyopathy	*EYA4*	FALSE
Cardiomyopathy	*MYBPC3*	TRUE
Cardiomyopathy	*MYH6*	FALSE
Cardiomyopathy	*MYH7*	TRUE
Cardiomyopathy	*NEXN*	FALSE
Cardiomyopathy	*PKP2*	TRUE
Cardiomyopathy	*PRKAG2*	TRUE
Cardiomyopathy	*RBM20*	TRUE
Cardiomyopathy	*RYR2*	TRUE
Cardiomyopathy	*SCN5A*	TRUE
Cardiomyopathy	*SDHA*	FALSE
Cardiomyopathy	*TNNI3*	TRUE
Cardiomyopathy	*TNNT2*	TRUE
Cardiomyopathy	*TTN*	TRUE
Cardiomyopathy	*VCL*	FALSE
Colorectal cancer	*APC*	TRUE
Colorectal cancer	*BUB1B*	FALSE
Colorectal cancer	*FLCN*	FALSE
Colorectal cancer	*MLH1*	TRUE
Colorectal cancer	*MLH3*	FALSE
Colorectal cancer	*MSH2*	TRUE
Colorectal cancer	*TGFBR2*	FALSE
Endometrial cancer	*MSH6*	TRUE
Familial breast cancer	*BRCA1*	TRUE
Familial breast cancer	*BRCA2*	TRUE
Familial breast cancer	*PALB2*	TRUE
Familial breast cancer	*PPM1D*	FALSE
Familial hypercholesterolemia	*LDLR*	TRUE

Monogenic diabetes	*KCNJ11*	FALSE
Prostate cancer	*BRCA2*	FALSE
Retinitis pigmentosa	*AIPL1*	FALSE
Retinitis pigmentosa	*CRB1*	FALSE
Retinitis pigmentosa	*DHDDS*	FALSE
Retinitis pigmentosa	*KLHL7*	FALSE
Retinitis pigmentosa	*NR2E3*	FALSE
Retinitis pigmentosa	*NRL*	FALSE
Retinitis pigmentosa	*PDE6B*	FALSE
Retinitis pigmentosa	*PROM1*	FALSE
Retinitis pigmentosa	*PRPF8*	FALSE
Retinitis pigmentosa	*PRPH2*	FALSE
Retinitis pigmentosa	*RHO*	FALSE
Retinitis pigmentosa	*RLBP1*	FALSE
Retinitis pigmentosa	*RP1*	FALSE
Retinitis pigmentosa	*RPE65*	TRUE
Retinitis pigmentosa	*SEMA4A*	FALSE

Supplementary Table 2. **Summary of baseline demographic traits by genetic disorder.** Related to Figure 4 and Table 1.

Disease	Age, mean (SD) years	Female, n (%)	European, n (%)	African, n (%)	Hispanic, n (%)	Asian, n (%)	Other, n (%)
Alzheimer disease	75 (13)	464 (63)	161 (22)	185 (25)	317 (43)	11 (1.5)	65 (8.8)
Amyotrophic lateral sclerosis	72 (12)	5 (71)	4 (57)	0 (0)	3 (43)	0 (0)	0 (0)
Breast cancer	64 (14)	1488 (100)	580 (39)	335 (23)	409 (27)	43 (2.9)	121 (8.1)
Cardiomyopathy	67 (14)	344 (42)	226 (27)	248 (30)	250 (30)	21 (2.6)	78 (9.5)
Colorectal cancer	72 (11)	122 (56)	59 (27)	58 (27)	78 (36)	3 (1.4)	20 (9.2)
Endometrial cancer	69 (11)	115 (100)	25 (22)	30 (26)	51 (44)	2 (1.7)	7 (6.1)
Hypercholesterolemia	67 (13)	2777 (62)	1437 (32)	1114 (25)	1391 (31)	158 (3.5)	360 (8.1)
Prostate cancer	75 (9.7)	0 (100)	177 (37)	154 (32)	105 (22)	6 (1.3)	32 (6.8)
Retinitis pigmentosa	63 (15)	7 (47)	4 (27)	6 (40)	5 (33)	0 (0)	0 (0)
Type 2 diabetes	67 (13)	4113 (58)	1087 (15)	2208 (31)	2850 (40)	250 (3.5)	649 (9.2)

Clinical diagnoses were extracted as International Classification of Diseases 10 (ICD-10) codes from the EHR and ancestries were self-reported. Other ancestry, miscellaneous ancestries other than those listed; n, number; SD, standard deviation.

Supplementary Table 3. Study set of 303 pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants in 54 genes corresponding to nine genetic conditions. Related to Figures 1-4 and Tables 1-2.

Dataset is available at: http://dx.doi.org/10.17632/c2g66gycvx.1.

Supplementary Table 4. ICD-10 diagnosis codes for nine genetic disorders. Related to Figures 1, 2, and 4, and Tables 1 and 2.

Disease	ICD-10 diagnosis code
Amyotrophic lateral sclerosis	Amyotrophic lateral sclerosis (‘G12.21’)
Cardiomyopathy	Cardiomyopathies, including dilated and hypertrophic types (‘I42*’)
Colorectal cancer	Malignant neoplasm of the colon and/or rectum (‘C18*’, ‘C19’)
Endometrial cancer	Malignant neoplasm of the endometrium (‘C54.1’)
Breast cancer	Malignant neoplasm of the breast (‘C50*’)
Familial hypercholesterolemia	Hypercholesterolemia, unspecified or familial (‘E78.0*’)
Prostate cancer	Malignant neoplasm of the prostate, including personal history of prostate cancer and in situ carcinoma of the prostate (‘C61’, ‘Z85.46’, ‘D07.5’)
Retinitis pigmentosa	Retinitis pigmentosa (‘H35.52’)
Monogenic diabetes	Type 2 diabetes (‘E11*’)

, any ICD-10 diagnosis code below the stated parent level (e.g., ‘F90’ includes ‘F90.1’).

Supplementary Table 5. Disease-specific medications for nine genetic disorders with prescriptions identified in the electronic health record. Related to Figures 1 and 4, and Table 2.

Disease	Medications
Amyotrophic lateral sclerosis	riluzole
Cardiomyopathy	ivabradine, dexrazoxane
Colorectal cancer	oxaliplatin, panitumumab, tipiracil
Endometrial cancer	--
Familial breast cancer	abemaciclib, adriamycin, anastrozole, arimidex, aromasin, capecitabine, carboplatin, cisplatin, cyclophosphamide, docetaxel, doxorubicin, exemestane, femara, femring, fluorouracil, gemcitabine, goserelin, herceptin, ibandronate, ibrance, ixabepilone, lapatinib, letrozole, lynparza, megace, methotrexate, methyltestosterone, mitomycin, navelbine, neratinib, olaparib, otrexup, paclitaxel, palbociclib, pertuzumab, raloxifene, rasuvo, tamoxifen, taxotere, toremifene, trastuzumab, tucatinib, verzenio, vinblastine, vinorelbine, xeloda, zoladex
Familial hypercholesterolemia	atorvastatin, simvastatin, pravastatin, lovastatin, fluvastatin, pitavastatin, rosuvastatin, cholestyramine, colestipol, colesevelam, niacin, ezetimibe, alirocumab, evolocumab
Monogenic diabetes	acarbose, actos, admelog, albiglutide, alogliptin, amaryl, avandia, canagliflozin, chlorpropamide, cycloset, dapagliflozin, dulaglutide, empagliflozin, ertugliflozin, exenatide, glimepiride, glinide, glipizide, glucophage, glucotrol, glumetza, glyburide, glyset, humalog, insulin, insulin degludec, insulin detemir, insulin glargine, insulin lispro, invokana, januvia, jardiance, junior kwikpen, levemir, linagliptin, liraglutide, lixisenatide, metformin, miglitol, nateglinide, onglyza, ozempic, pioglitazone, pramlintide, precose, repaglinide, riomet, rosiglitazone, rybelsus, saxagliptin, saxenda, semaglutide, sitagliptin, starlix, steglatro, symlinpen, tanzeum, tradjenta, tresiba, trulicity, victoza
Prostate cancer	degarelix, triptorelin, bicalutamide, flutamide, abiraterone, enalautamide, nilutamide, apalutamide, darolutamide, sipuleucel
Retinitis pigmentosa	--

Disease-specific medication, medication with an indication specific to the disease (e.g., anastrozole is specifically indicated for breast cancer, whereas cyclophosphamide's indication is not specific to breast cancer); --, no disease-specific medication.

Supplementary Table 7. Clinical diagnosis and phenotype of individuals with pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants stratified by ancestry. Related to Figure 2, and Tables 1 and 2.

Ancestry	Observations, n	+Dx, n (%)	-Dx, n (%)	-Dx -Sx, n (%)	-Dx +Sx, n (%)
African	121	30 (25)	89 (74)	72 (81)	17 (19)
Asian	29	9 (31)	20 (69)	17 (85)	3 (15)
European	259	67 (26)	191 (74)	164 (86)	27 (14)
Hispanic	182	39 (21)	145 (80)	123 (85)	22 (15)
Other	53	12 (23)	42 (79)	36 (86)	6 (14)

Assessment of the clinical diagnosis and phenotypes of 644 observations of 303 pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants in 614 individuals stratified by self-reported ancestry. Observations were categorized into three distinct groups: 1) clinically diagnosed (+Dx), the individual with a P/LP/LoF variant had a prior clinical diagnosis for the genetic disorder corresponding to the variant; 2) no clinical diagnosis and no evidence of disease (-Dx -Sx), the individual with a P/LP/LoF variant lacked a prior clinical diagnosis for the corresponding genetic disorder and had no electronic health record (EHR) evidence of disease symptoms or findings; and 3) no clinical diagnosis but has evidence of disease (-Dx +Sx), the individual with a P/LP/LoF variant lacked a clinical diagnosis for the corresponding genetic disorder, but had EHR evidence of disease symptoms or findings.

Supplementary Table 8. Symptom scores for 39 symptoms and their prevalence among clinically undiagnosed observations of variants and disease in individuals with pathogenic/likely pathogenic or predicted loss-of-function (P/LP/LoF) variants. Related to Figure 4 and Table 2

Disease	Sign/symptom	Matched phenotype from Gotlieb *et al.*[1]	Symptom score	No clinical diagnosis, n	Evidence of disease, n (% of no clinical diagnosis)
Prostate cancer	Metastatic disease	Bone cancer metastatic	0.9829212	0	--
Cardiomyopathy	Heart failure	Cardiac failure acute	0.9108916	154	90 (58)
Amyotrophic lateral sclerosis	Respiratory failure	Respiratory failure	0.856541	20	4 (20)
Familial hypercholesterolemia	Coronary heart disease	Coronary artery disease	0.7857806	6	3 (50)
Cardiomyopathy	Valvular dysfunction	Cardiac valve disease	0.7401669	154	23 (15)
Cardiomyopathy	Cardiac remodeling	Cardiac hypertrophy	0.6879898	154	16 (10)
Endometrial cancer	Abnormal uterine bleeding	Uterine haemorrhage	0.6631796	12	2 (17)
Familial hypercholesterolemia	Peripheral artery disease	Peripheral vascular disorder	0.6298026	6	0 (0)
Prostate cancer	Anomalous DRE findings	Prostatic disorder	0.6086371	0	--
Endometrial cancer	Endometrial hyperplasia and disorganization on biopsy	Endometrial hyperplasia	0.5750928	12	1 (8.3)
Endometrial cancer	Thickened endometrium on imaging	Endometrial hypertrophy	0.5683344	12	0 (0)
Familial breast cancer	Axillary lymphadenopathy	Lymphadenopathy	0.5619201	106	20 (19)
Prostate cancer	Urethral compression	Urethral obstruction	0.5502927	0	--
Familial breast cancer	Mass	Breast mass	0.5465219	106	6 (5.7)
Prostate cancer	Elevated PSA	Prostatic specific antigen increased	0.5405282	0	--
Colorectal cancer	Exophytic colonic mass	Colonic polyp	0.5235142	17	1 (5.9)
Cardiomyopathy	Abnormal cardiac exam findings	Heart sounds abnormal	0.5215908	154	17 (11)
Familial hypercholesterolemia	Hypercholesterolemia	Hypercholesterolaemia	0.5175085	6	1 (17)
Colorectal cancer	Obstructed lumen	Colonic obstruction	0.5063074	17	0 (0)
Amyotrophic lateral sclerosis	Bulbar symptoms (dysarthria, dysphagia, facial weakness, tongue atrophy)	Facial paresis	0.493729	20	0 (0)
Prostate cancer	Infiltration to genitourinary system	Haematuria	0.4768623	0	--
Monogenic diabetes	Insulin perturbations	Insulin resistance	0.4752398	0	--
Familial hypercholesterolemia	Corneal arcus	Corneal deposits	0.4716786	6	1 (17)

Endometrial cancer	Enlarged uterus/abdominal mass	Uterine enlargement	0.470359	12	0 (0)
Monogenic diabetes	Dysregulated glycemic control	Blood glucose increased	0.4085338	0	--
Familial hypercholesterolemia	Xanthoma	Xanthoma	0.4019195	6	0 (0)
Colorectal cancer	Occult bleeding or melena	Faecal occult blood positive	0.3889438	17	3 (18)
Amyotrophic lateral sclerosis	Labile affect	Emotional disorder	0.3856494	20	0 (0)
Monogenic diabetes	Acanthosis nigricans	Acanthosis nigricans	0.358083	0	--
Familial breast cancer	Imaging anomalies	Breast microcalcification	0.3556099	106	5 (4.7)
Retinitis pigmentosa	Nyctalopia and glare sensitivity	Night blindness	0.3528767	168	0 (0)
Monogenic diabetes	Polyuria, polydipsia, polyphagia	Polyuria	0.3449517	0	--
Retinitis pigmentosa	Progressive, peripheral vision loss	Tunnel vision	0.3432615	168	3 (1.8)
Amyotrophic lateral sclerosis	Asymmetric limb weakness	Muscular weakness	0.2832084	20	1 (5)
Endometrial cancer	Pelvic pain	Pelvic pain	0.2263445	12	1 (8.3)
Familial breast cancer	Dermatological changes	Breast tenderness	0.1866282	106	0 (0)
Amyotrophic lateral sclerosis	Muscle fasciculations	Muscle twitching	0.185048	20	0 (0)
Familial breast cancer	Nipple abnormalities	Breast discharge	0.1665234	106	0 (0)
Familial breast cancer	Histopathological findings	Breast enlargement	0.1300308	106	2 (1.9)

No clinical diagnosis, observations of variants in individuals lacking a corresponding clinical diagnosis; evidence of disease, clinically undiagnosed observations that had electronic health record evidence of symptomatic disease; symptom score, on an ordinal scale ranging from 0 (lowest risk of hospitalization or death) to 1 (highest risk of hospitalization or death) from a previously published study[1]; --, not applicable as all observations for monogenic diabetes and prostate cancer were diagnosed with the corresponding disease.

1. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. *J Med Internet Res* 2015; **17**. DOI:10.2196/jmir.3962.