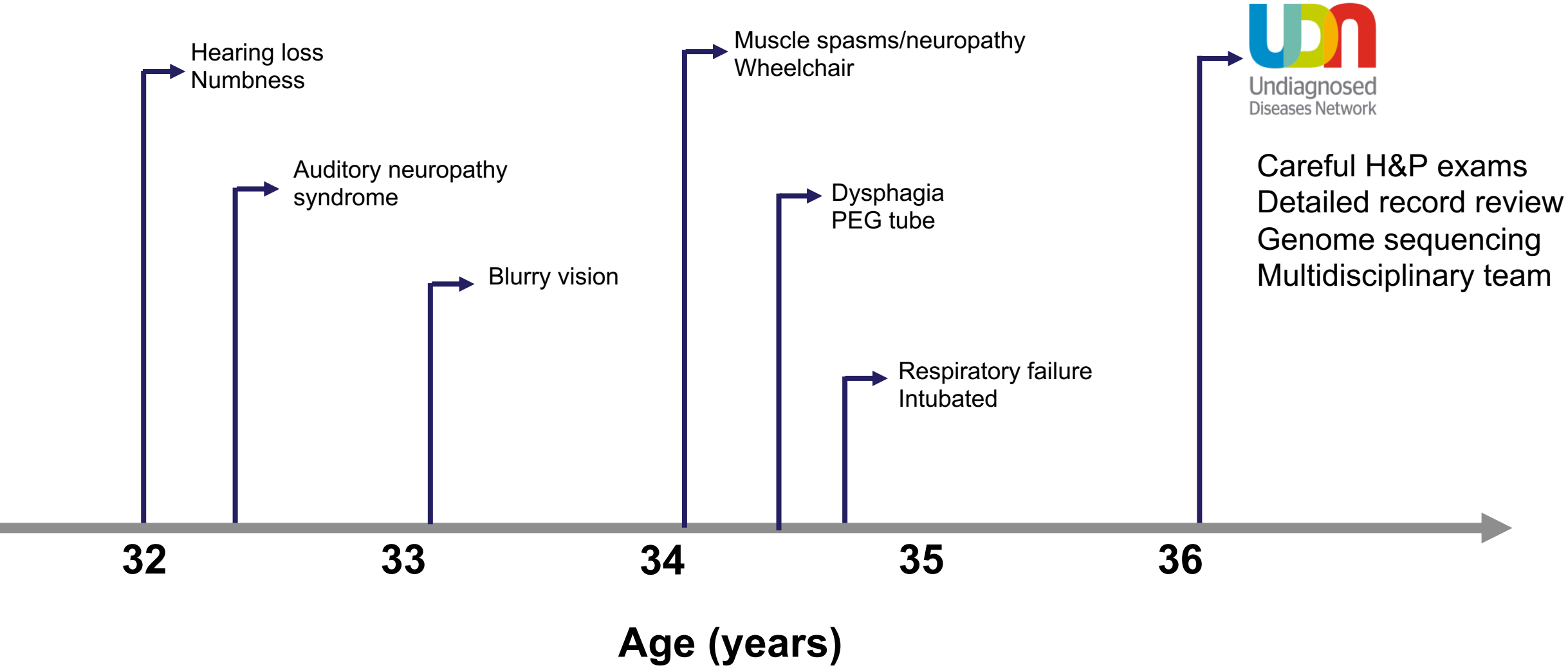
Phenotype risk scores (PheRS) for risk prediction

Josh Denny, MD MS

Vanderbilt University Medical Center

5/6/2019

Case: What is wrong with this patient?



Various methods of testing for gene/disease associations

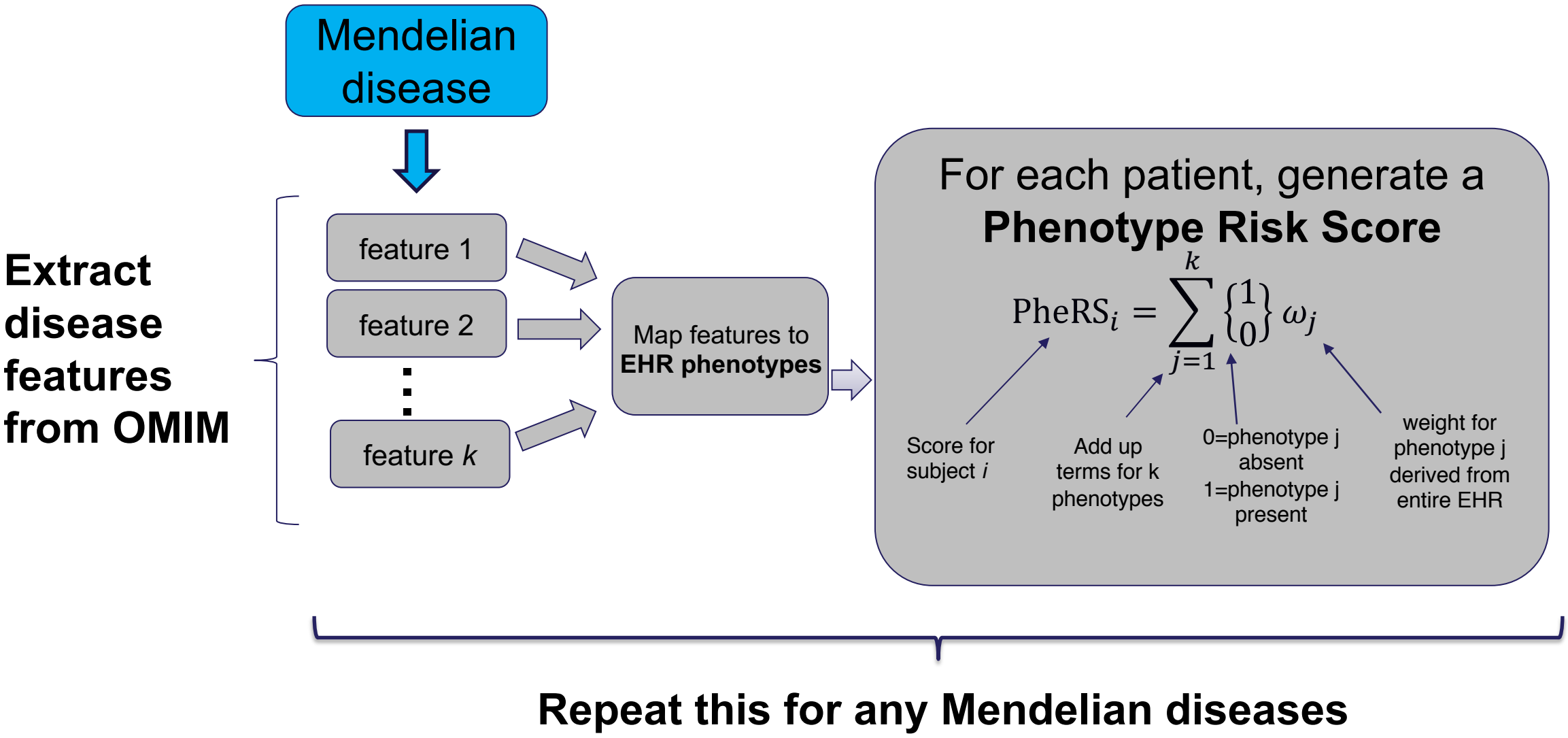
Looping through association tests

- GWAS/NGS – one phenotype, many genotypes
- PheWAS – one genotype, many phenotypes

Aggregation approaches

- GRS/PRS/GPS – one phenotypes, many genotypes aggregated
- **Phenotype risk scores (PheRS) – one genotype, many phenotypes aggregated**

Automating assessments of "phenotype patterns" in the EHR

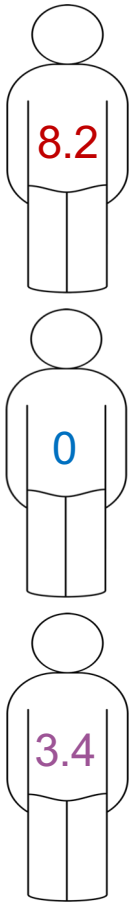


Bastarache et al, Science 2018

CYSTIC FIBROSIS; CF

	HPO	Phecodes	
INHERITANCE			
- Autosomal recessive			
GROWTH			
Other			
- Failure to thrive	1508	264.2	Failure to thrive……..1.62
CARDIOVASCULAR			
Heart			
- Cor pulmonale	1648	415.1	Acute pulmonary heart disease……1.49
RESPIRATORY			
Airways			
- Chronic bronchopulmonary infection	6538	483	Acute bronchitis & bronchiolitis.....1.00
- Bronchiectasis	2110	496.3	Bronchiectasis……..1.80
- Asthma	2099	495	Asthma…….0.98
- Pulmonary blebs	-	-	
- Pseudomonas colonization	-	-	
ABDOMEN			
Pancreas			
- Pancreatic insufficiency in 80%	1738	577	Diseases of pancreas…….1.42
Biliary Tract			
- Biliary cirrhosis	2613	571.6	Primary biliary cirrhosis……....2.06

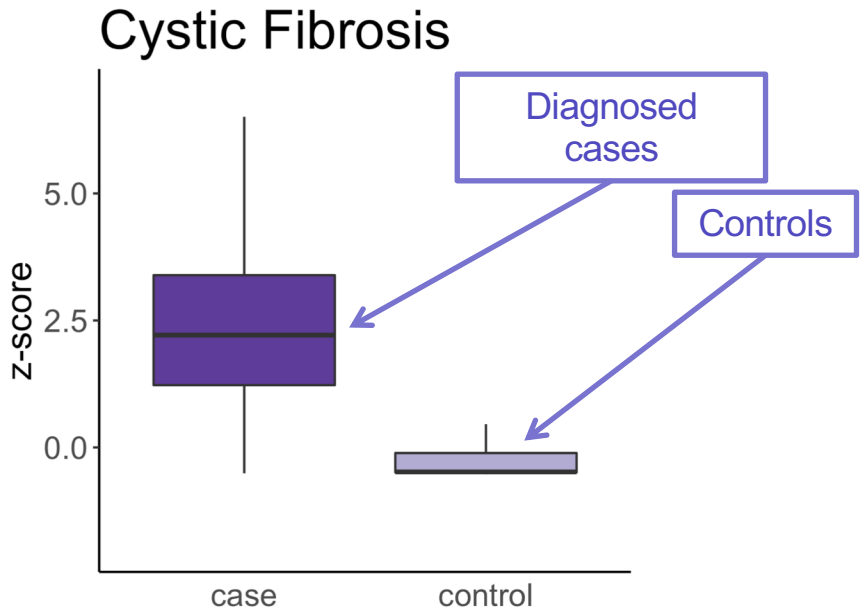
Many diseases have “phenotype patterns”: consider Cystic Fibrosis



Bronchiectasis
Pancreatitis
Asthma
Failure to thrive

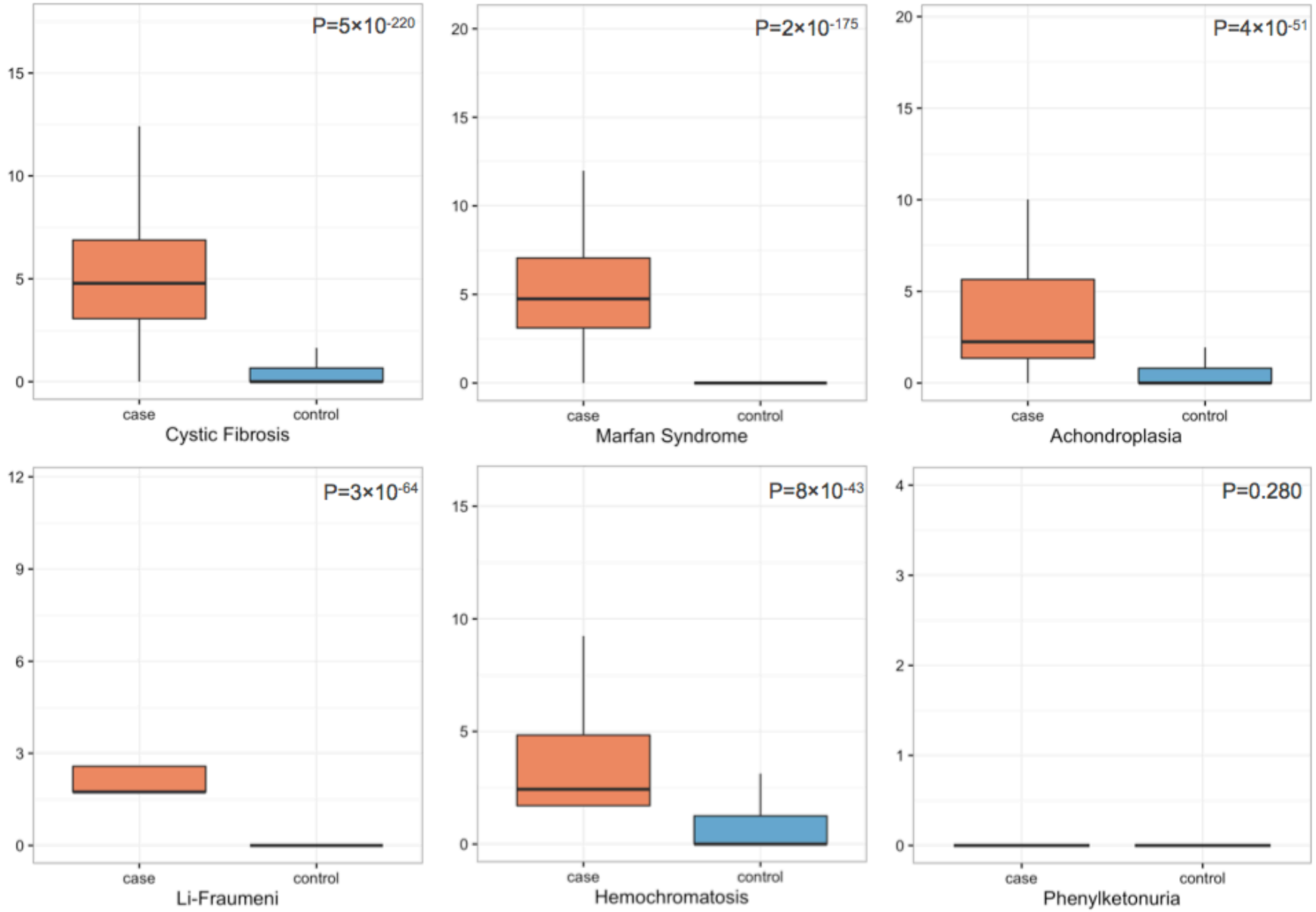
No matching symptoms

Pancreatitis
Bronchitis



*You can differentiate a group individuals diagnosed with a disease using **only the features** of the disease*

Validating Phenotype Risk Scores on diagnosed individuals



Bastarache et al, Science 2018

RESEARCH ARTICLE

HUMAN GENOMICS

Phenotype risk scores identify patients with unrecognized Mendelian disease patterns

Gene	Variant	rsID	HOM/ HET	Associated Mendelian Disease	OMIM Reported inheritance	Phenotype categories in PheRS	Beta	P	ClinVar	HGMD	ACMG
CFTR	c.1624G>T p.Gly542Ter	rs113993959	1/27	Cystic fibrosis	AR		1.39	2.9×10^{-8}	P	Y	P
CHRNA4	c.1448G>A p.Arg483Gln	rs55855125	1/21	Nocturnal frontal lobe epilepsy, 1	AD		0.58	9.0×10^{-8}	U		U
DGKE	c.966G>A p.Trp322Ter	rs138924661	1/14	Nephrotic syndrome, type 7	AR		1.31	2.8×10^{-7}	LP	Y	LP→P
SUOX	c.228G>T p.Arg76Ser	rs202085145	0/24	Sulfocysteinuria	AR		0.82	1.7×10^{-6}	U		U→P
CFTR	c.1657C>T p.Arg553Ter	rs74597325	0/12	Cystic fibrosis	AR		1.81	2.1×10^{-6}	P	Y	P
KIF1B	c.2021C>T p.Thr674Ile	rs41274468	0/21	Charcot-Marie-Tooth disease, 2A1	AD		0.79	5.3×10^{-6}			U
VWF	c.5851A>G p.Thr1951Ala	rs144072210	0/21	Von Willebrand disease	AR*		0.53	8.6×10^{-6}		Y	U
KIF1A	c.2676C>T p.Ala993=	rs116297894	1/25	Spastic paraplegia-30	AR		0.84	1.3×10^{-5}	LB		LB→U
F10	c.872G>A p.Arg291Gln	rs149212700	0/15	Factor X deficiency	AR*		0.62	1.9×10^{-5}			U
HFE	c.502G>C p.Glu168Gln	rs146519482	0/40	Hemochromatosis	AR		1.08	4.0×10^{-5}	U	Y	U
TG	c.229G>A p.Gly77Ser	rs142698837	0/69	Thyroid dyshormonogenesis	AR		0.26	6.0×10^{-5}		Y	U→P
SH2B3	c.1183G>A p.Glu395Lys	rs148636776	0/22	Familial erythrocytosis, 1	AD		1.48	6.1×10^{-5}			U→P
SPTBN2	c.7109G>A p.Arg2370His	rs145522851	0/11	Spinocerebellar ataxia	AR*		0.75	9.0×10^{-5}			U→LP
FAN1	c.1520G>A p.Arg507His	rs150393409	0/434	Interstitial nephritis, karyomegalic	AR		0.15	9.9×10^{-5}			LB→U
PANK2	c.1561G>A p.Gly521Arg	rs137852959	0/26	HARP syndrome	AR		0.58	1.1×10^{-4}	P	Y	P
SH2B3	c.1183G>A p.Glu395Lys	rs148636776	0/22	Essential thrombocythemia	AD		0.33	1.4×10^{-4}			U→P
AGXT	c.883G>A p.Ala295Thr	rs13408961	1/35	Primary hyperoxaluria, type I	AR		0.82	1.7×10^{-4}	U/LB		LB→U
PLCG2	c.751A>G p.Ile251Val	rs190840748	0/10	Familial cold autoinflammatory syn. 3	AD		0.70	1.9×10^{-4}			U

Neoplastic
Endocrine/Metabolic/Blood
Nervous/Psychiatric/Sensory
Circulatory/Respiratory
Digestive/Genitourinary
Musculoskeletal/Dermatologic
Other symptoms/Injuries

Tested

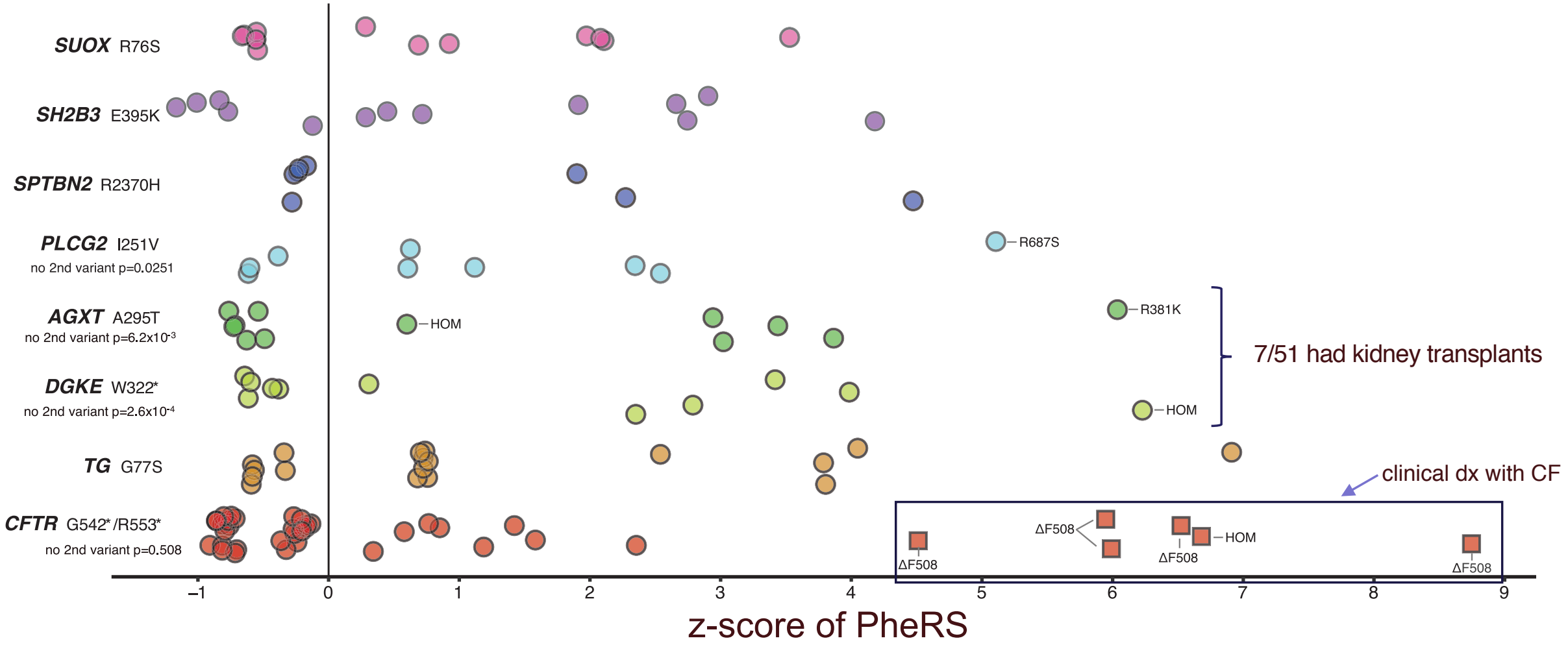
6644 variant/disease pairs

Found

18 significant associations (14 novel)

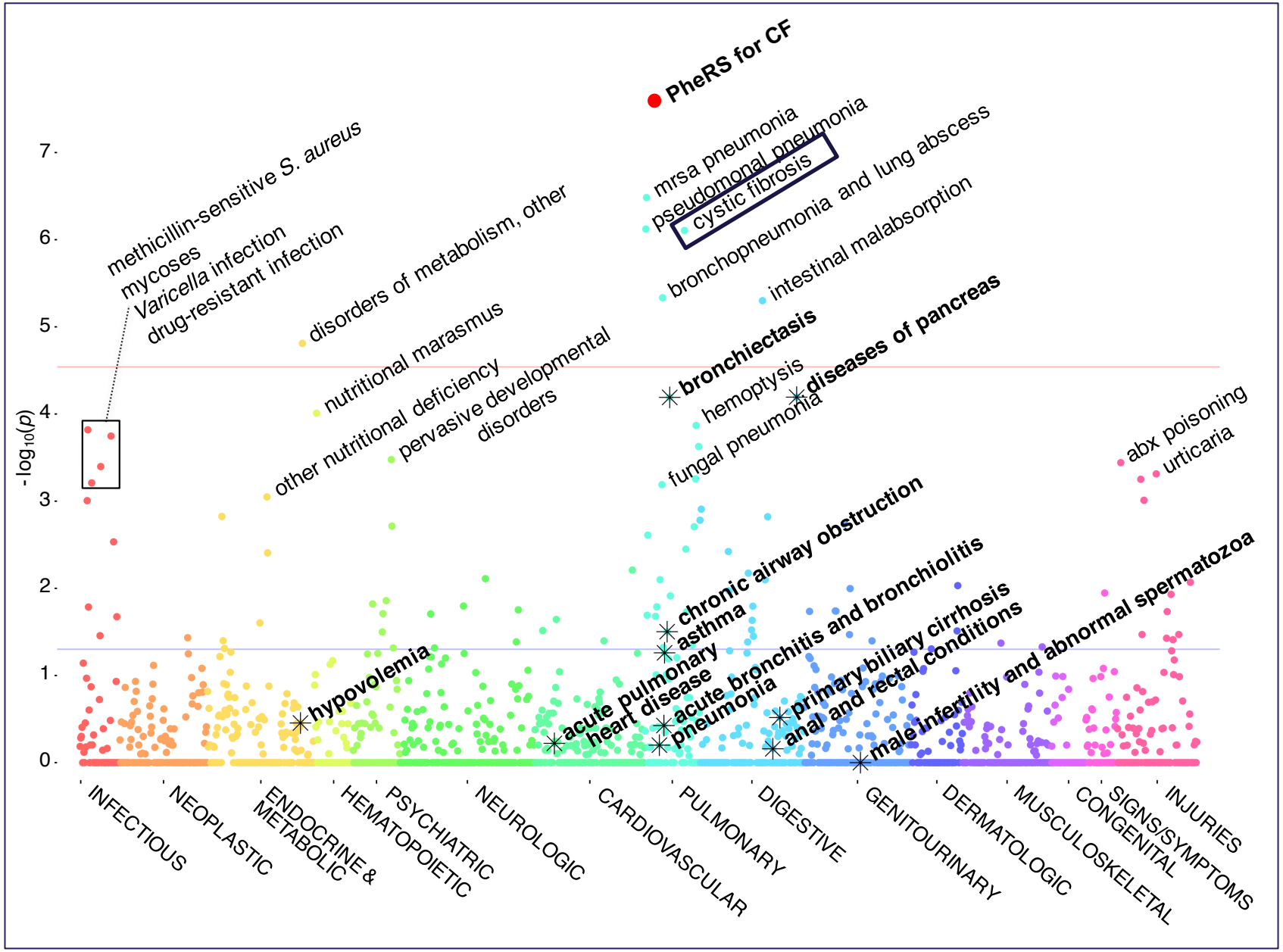
Bastarache et al, Science 2018

PheRS identifies novel pathogenic variants with clinical impact



Bastarache et al, Science 2018

PheWAS of a CF variant



Bastarache et al, Science 2018

A case of a rare disease

The NEW ENGLAND JOURNAL of MEDICINE

CASE RECORDS of the MASSACHUSETTS GENERAL HOSPITAL

Her PheRS for Cystic Fibrosis – top 2%-ile

Case 40-2018: A 47-Year-Old Woman with Recurrent Sinusitis, Cough, and Bronchiectasis

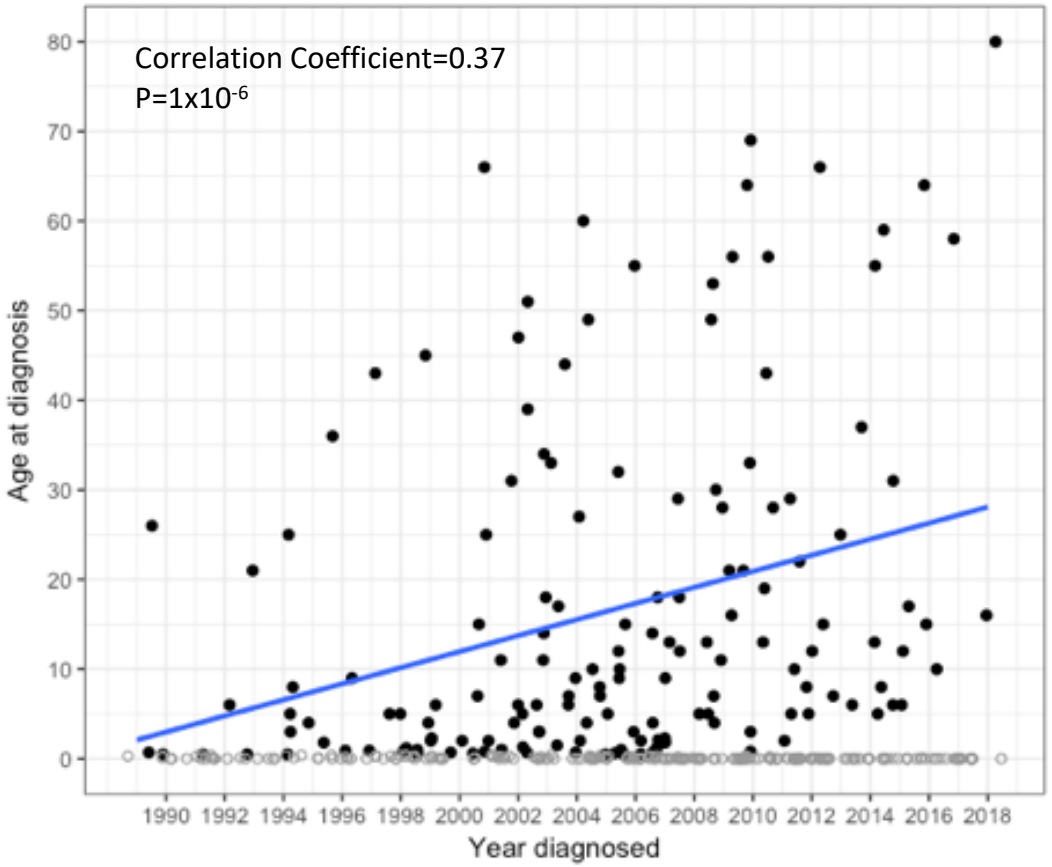
James E. Mojica, M.D., Christopher J. Richards, M.D., Jad S. Husseini, M.D., and Lida P. Hariri, M.D., Ph.D.

PRESENTATION OF CASE

Dr. Christopher J. Richards: A 47-year-old woman was evaluated at the outpatient pulmonary clinic of this hospital because of recurrent sinusitis with progressive cough and bronchiectasis.

Since her mid-20s, the patient had had recurrent episodes of sinus congestion, with two or three sinus infections annually, which had prompted treatment with

Average age of new CF diagnoses over time at VUMC (n=368)



Mojica NEJM 2018, Bastarache NEJM 2019

Interpreting variants

Premise

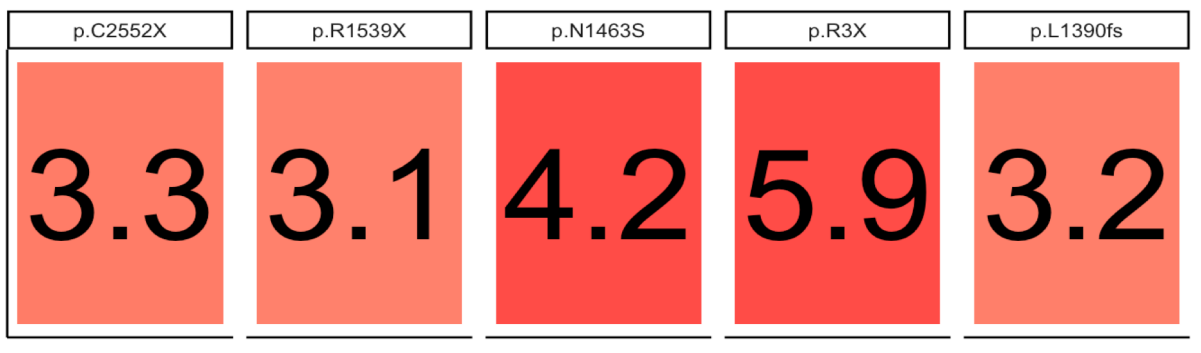
- ACMG59 variants are becoming increasingly returned (eMERGE, *All of Us*)
- Most rare variants are VUS

eMERGEseq

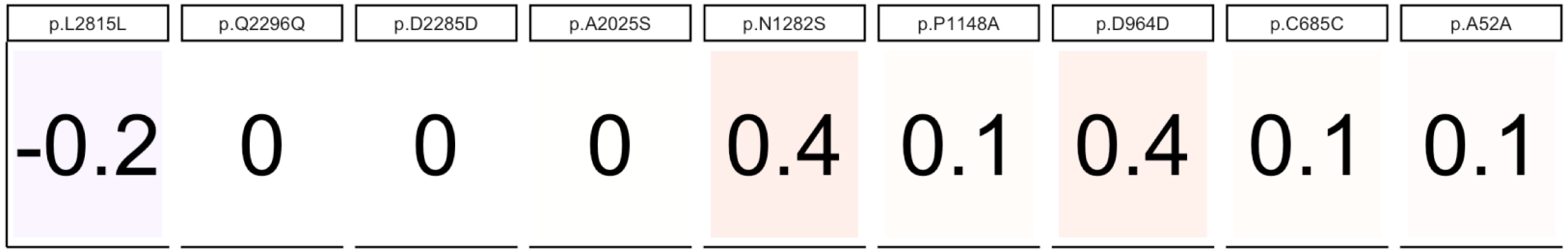
- 25,000 patients sequenced for 109 genes (including 58 ACMG "returnable" genes)
- 10 clinical sites
- All linked to EHR data

PheRS for high-throughput variant interpretation using EHRs: ***FBN1*** **(Marfan Syndrome)**

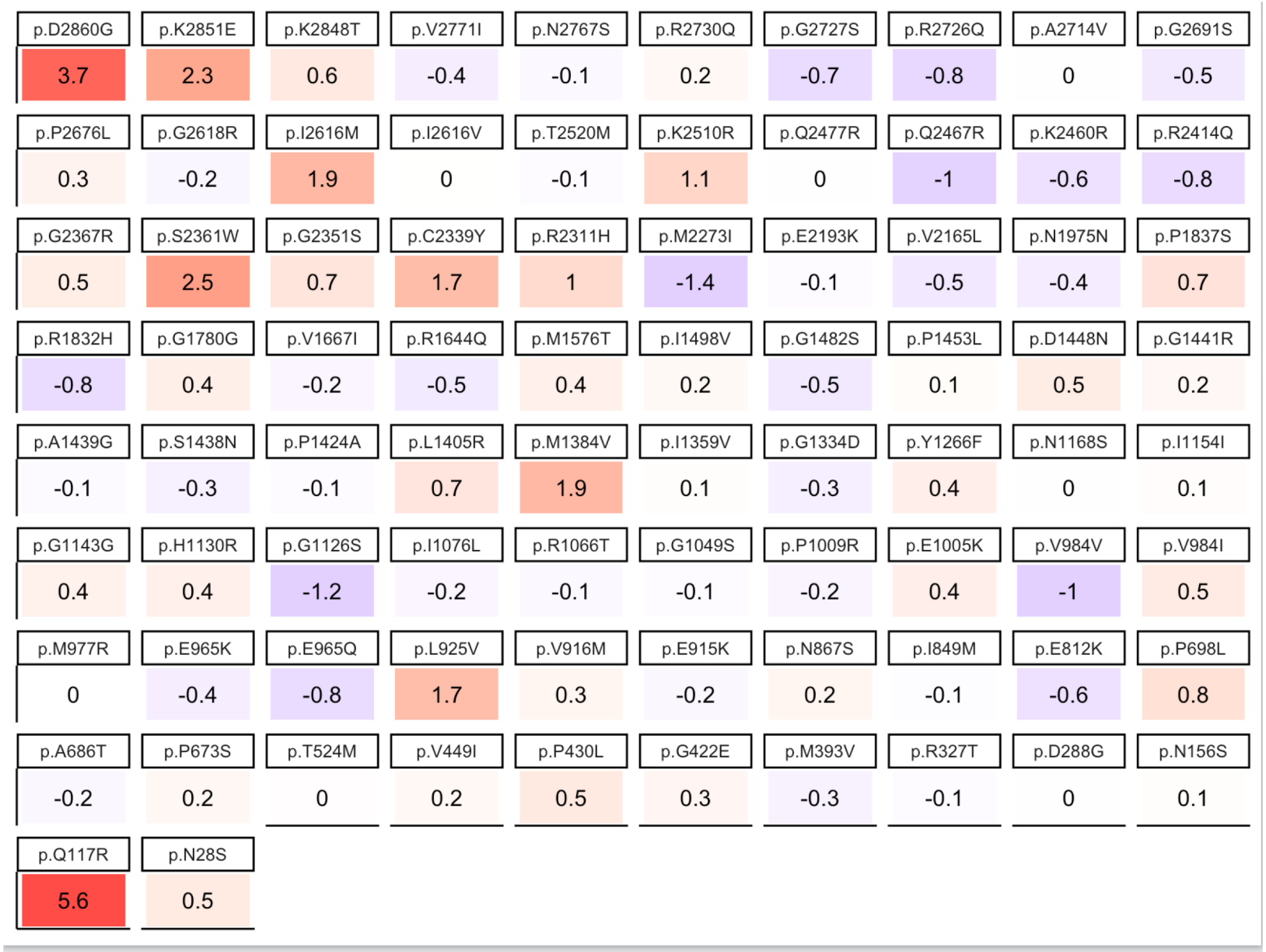
Known pathogenic variants



Known benign variants

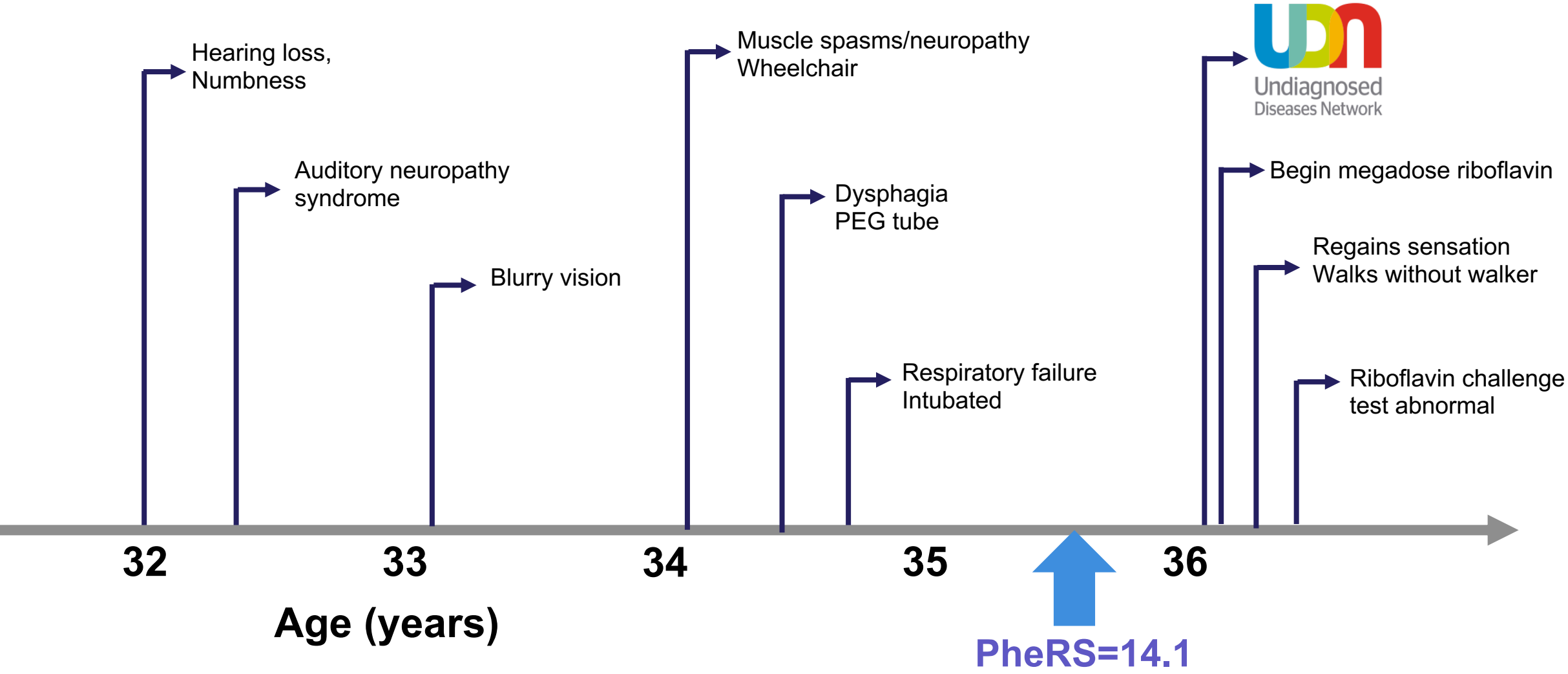


What about *FBN1* Variants of Uncertain Significance



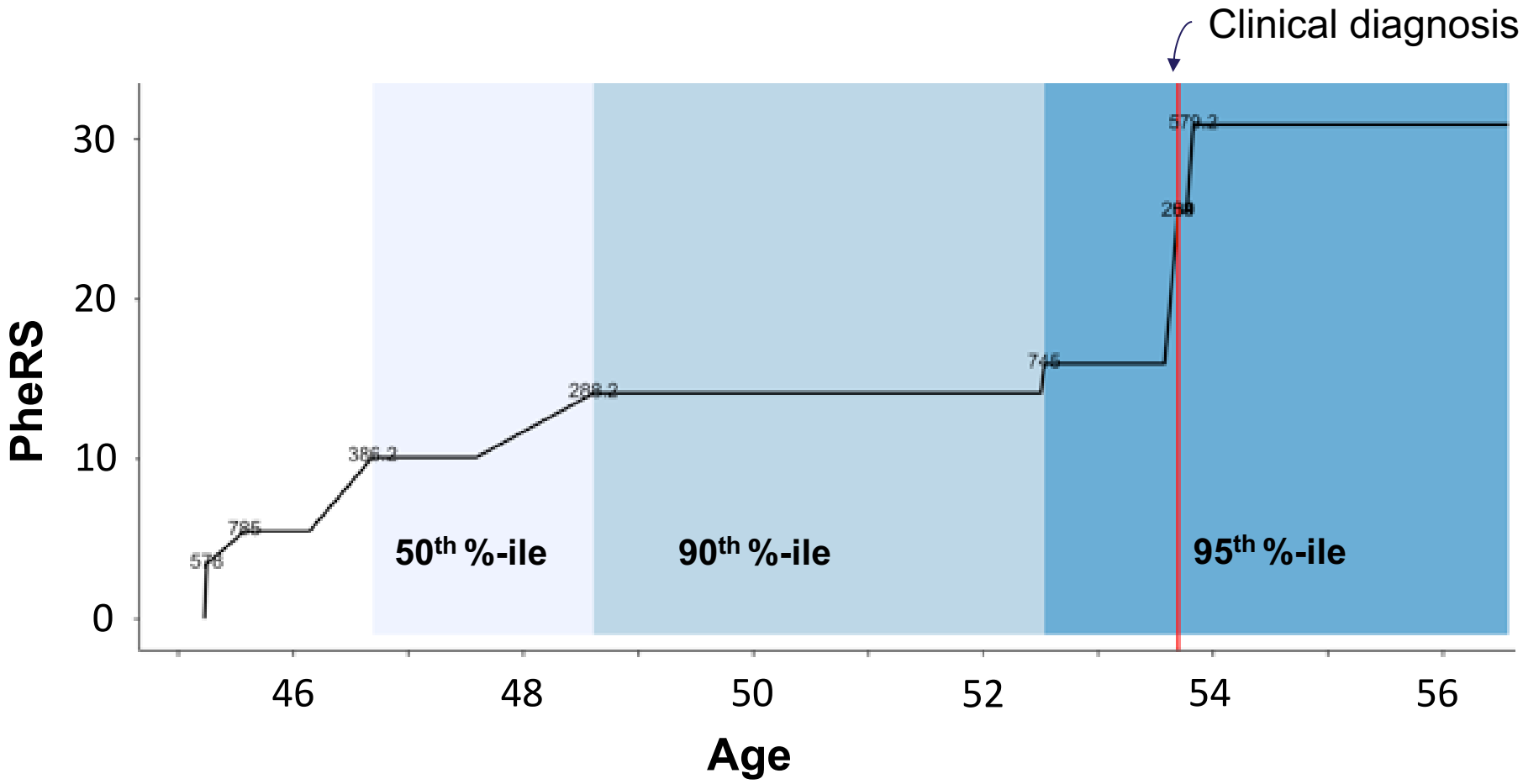
p.D2860G	p.K2851E	p.K2848T	p.V2771I	p.N2767S	p.R2730Q	p.G2727S	p.R2726Q	p.A2714V	p.G2691S
3.7	2.3	0.6	-0.4	-0.1	0.2	-0.7	-0.8	0	-0.5
p.P2676L	p.G2618R	p.I2616M	p.I2616V	p.T2520M	p.K2510R	p.Q2477R	p.Q2467R	p.K2460R	p.R2414Q
0.3	-0.2	1.9	0	-0.1	1.1	0	-1	-0.6	-0.8
p.G2367R	p.S2361W	p.G2351S	p.C2339Y	p.R2311H	p.M2273I	p.E2193K	p.V2165L	p.N1975N	p.P1837S
0.5	2.5	0.7	1.7	1	-1.4	-0.1	-0.5	-0.4	0.7
p.R1832H	p.G1780G	p.V1667I	p.R1644Q	p.M1576T	p.I1498V	p.G1482S	p.P1453L	p.D1448N	p.G1441R
-0.8	0.4	-0.2	-0.5	0.4	0.2	-0.5	0.1	0.5	0.2
p.A1439G	p.S1438N	p.P1424A	p.L1405R	p.M1384V	p.I1359V	p.G1334D	p.Y1266F	p.N1168S	p.I1154I
-0.1	-0.3	-0.1	0.7	1.9	0.1	-0.3	0.4	0	0.1
p.G1143G	p.H1130R	p.G1126S	p.I1076L	p.R1066T	p.G1049S	p.P1009R	p.E1005K	p.V984V	p.V984I
0.4	0.4	-1.2	-0.2	-0.1	-0.1	-0.2	0.4	-1	0.5
p.M977R	p.E965K	p.E965Q	p.L925V	p.V916M	p.E915K	p.N867S	p.I849M	p.E812K	p.P698L
0	-0.4	-0.8	1.7	0.3	-0.2	0.2	-0.1	-0.6	0.8
p.A686T	p.P673S	p.T524M	p.V449I	p.P430L	p.G422E	p.M393V	p.R327T	p.D288G	p.N156S
-0.2	0.2	0	0.2	0.5	0.3	-0.3	-0.1	0	0.1
p.Q117R	p.N28S								
5.6	0.5								

Case: A diagnosis – **riboflavin deficiency**



PheRS may help diagnose people earlier, but it will be hard

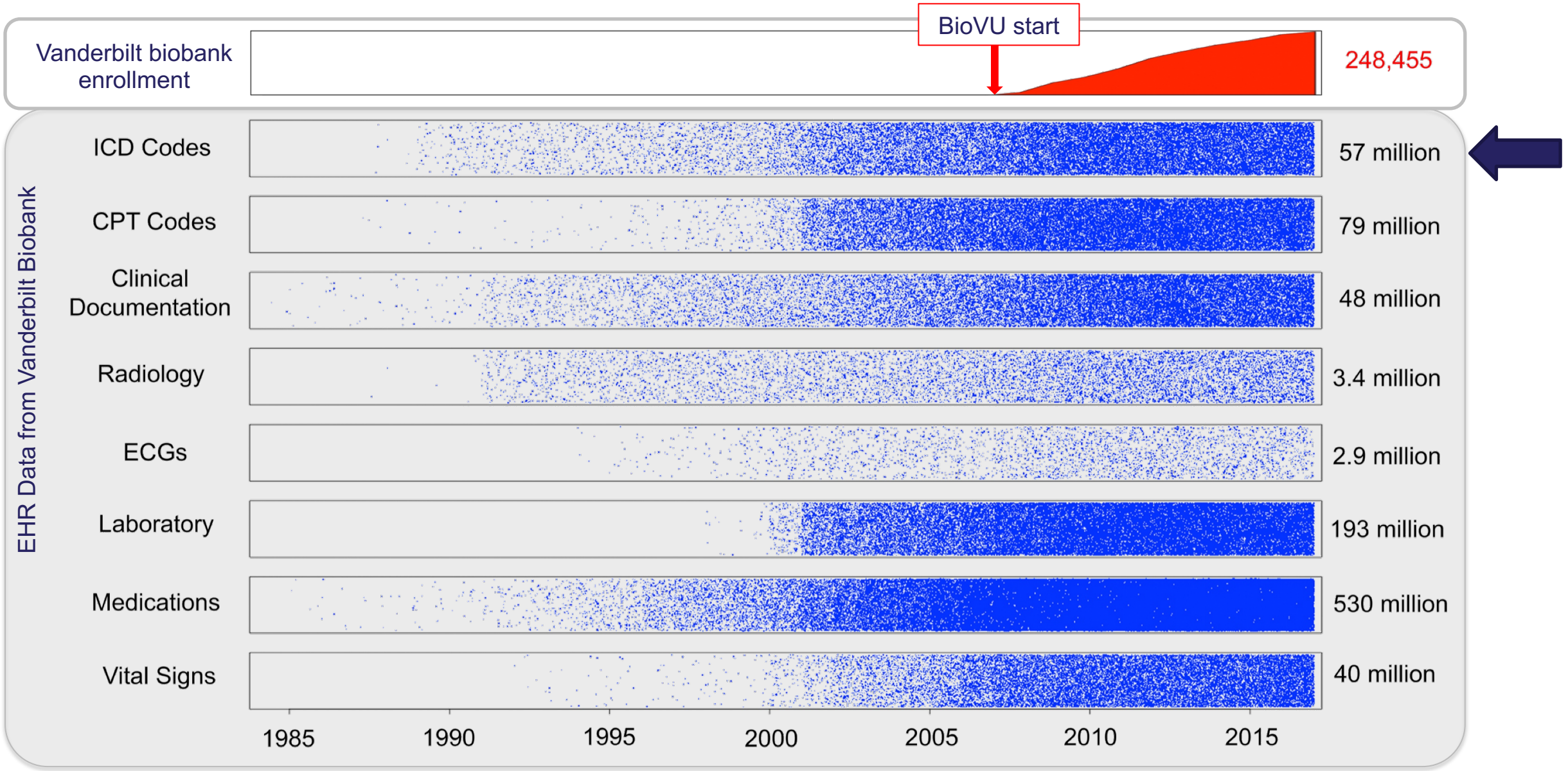
Polycythemia vera due to a *JAK2* mutation

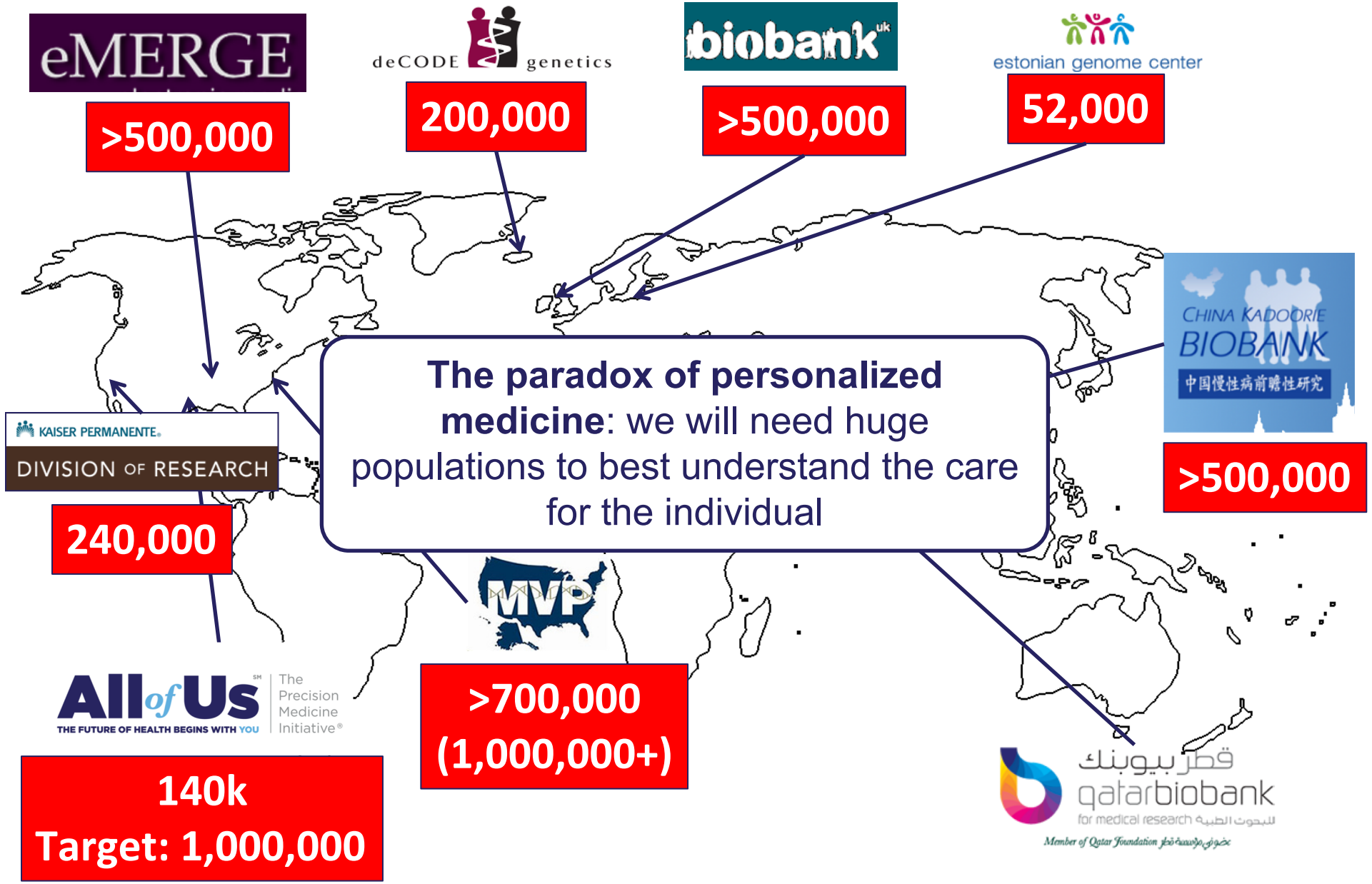


PT/ET prevalence ~ 40/100k

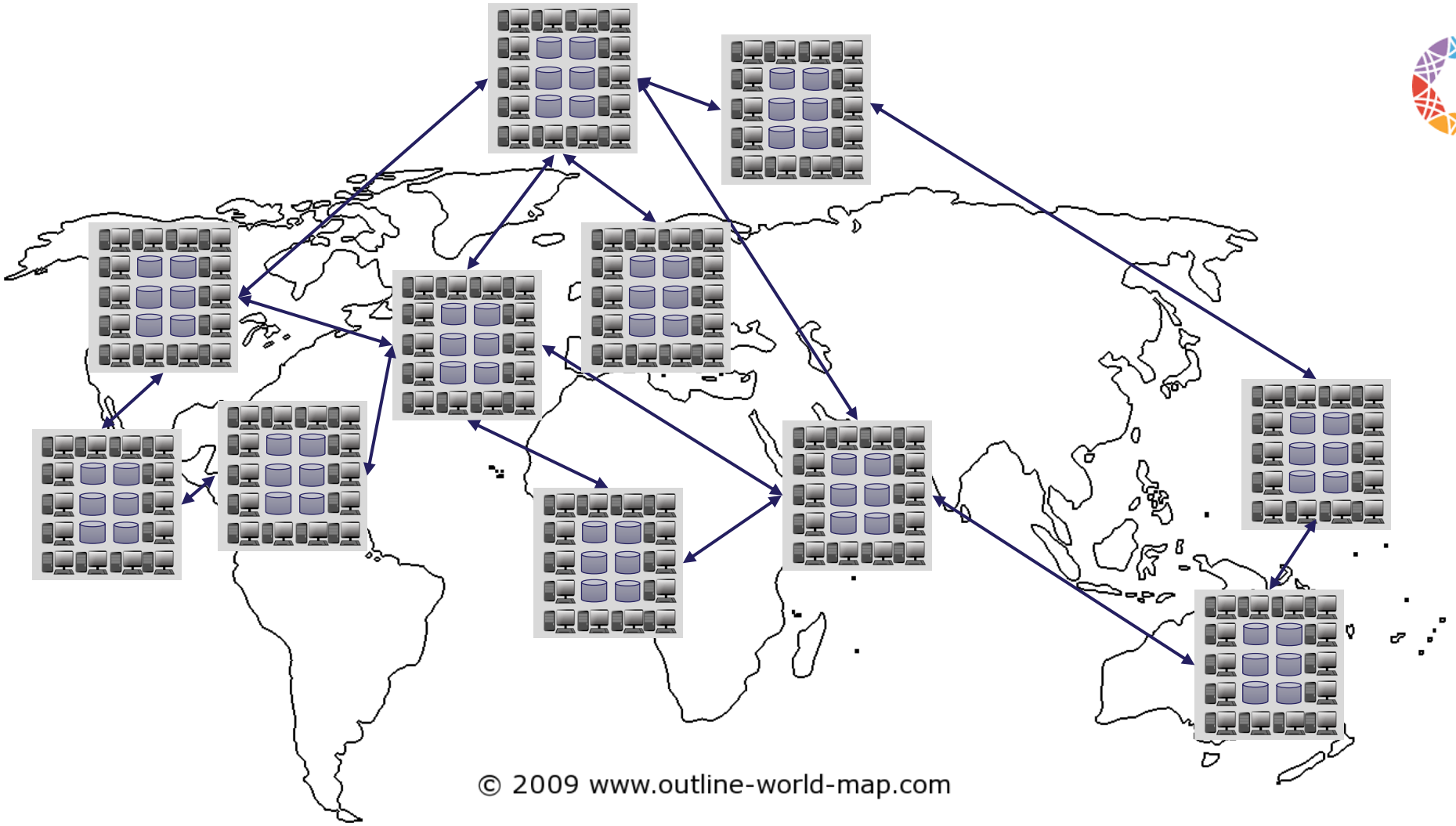
200-fold enrichment ~ 1/10

EHRs provides dense resource for efficient discovery: BioVU's example




eMERGE
>500,000
deCODE genetics
200,000
biobank uk
>500,000
estonian genome center
52,000
CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究
>500,000
KAISER PERMANENTE
DIVISION OF RESEARCH
240,000
The paradox of personalized medicine: we will need huge populations to best understand the care for the individual
MVP
>700,000
(1,000,000+)
All of Us
THE FUTURE OF HEALTH BEGINS WITH YOU
The Precision Medicine Initiative
140k
Target: 1,000,000
قطر بيوبنك
qatarbiobank
for medical research
Member of Qatar Foundation

The power of a data biosphere of common semantics and APIs



Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

IHCC

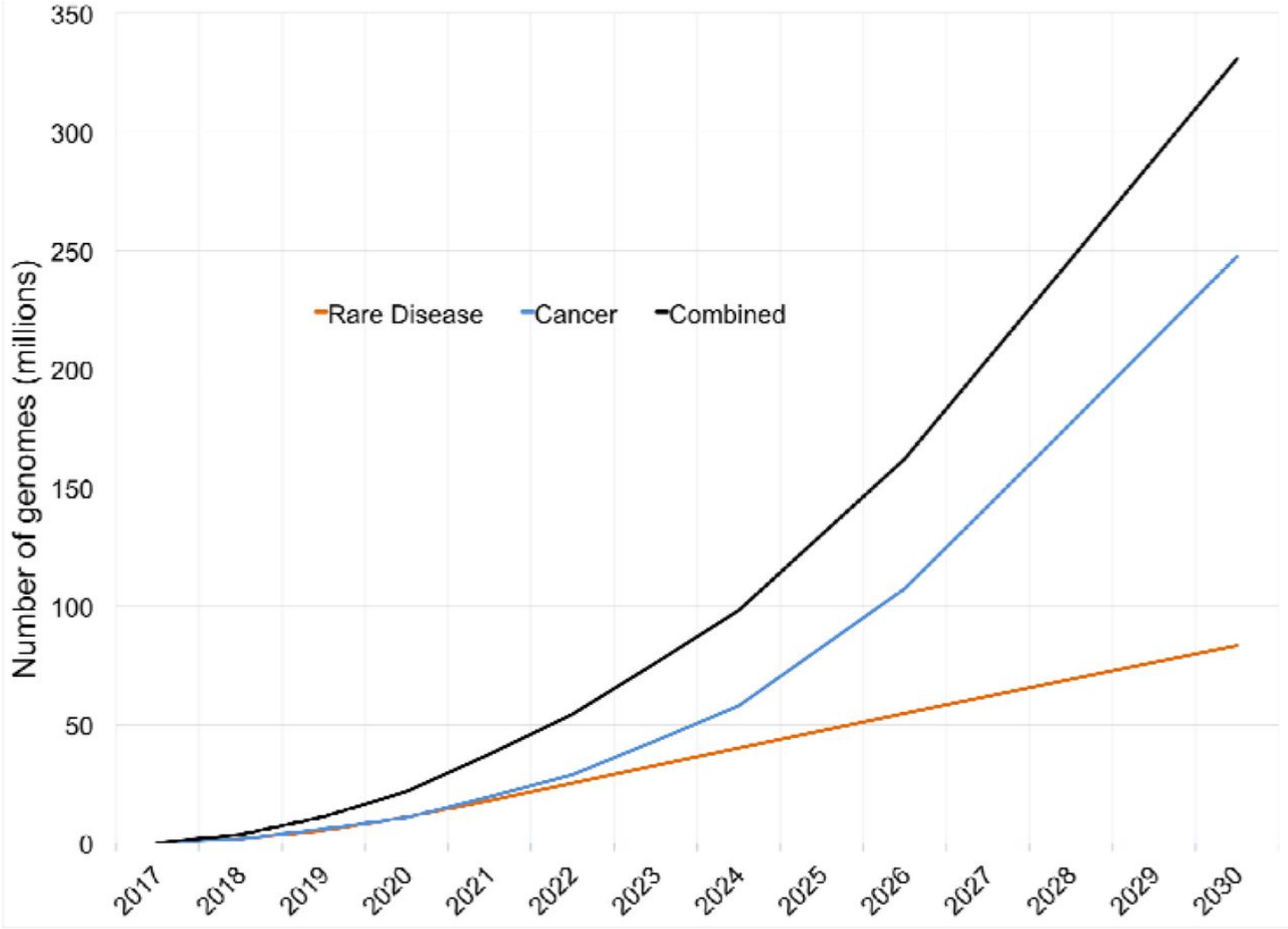
Global Genomic Medicine Collaborative

All of Us RESEARCH PROGRAM

DataSTAGE

Sources for genetic data

- Research cohorts
- Clinical testing will explode
  - Cancer testing
  - Increasing rare disease and expanding indications (UHC, BCBS Evidence Street)
  - Yet another reason to store genetic data not as PDFs…
  - Genomics as the tool to really move the Learning Healthcare System into standard practice?



Birney et al. biorXiv 2017. https://doi.org/10.1101/203554

A small sampling of the VUMC Team

All of Us℠ The Precision Medicine Initiative®
THE FUTURE OF HEALTH BEGINS WITH YOU

PGRN Pharmacogenomics Research Network

emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

+many others!