Icahn School of Medicine at Mount Sinai

Contributions of diverse populations and expanded catalogues of human variation to our understanding of low frequency and rare variants

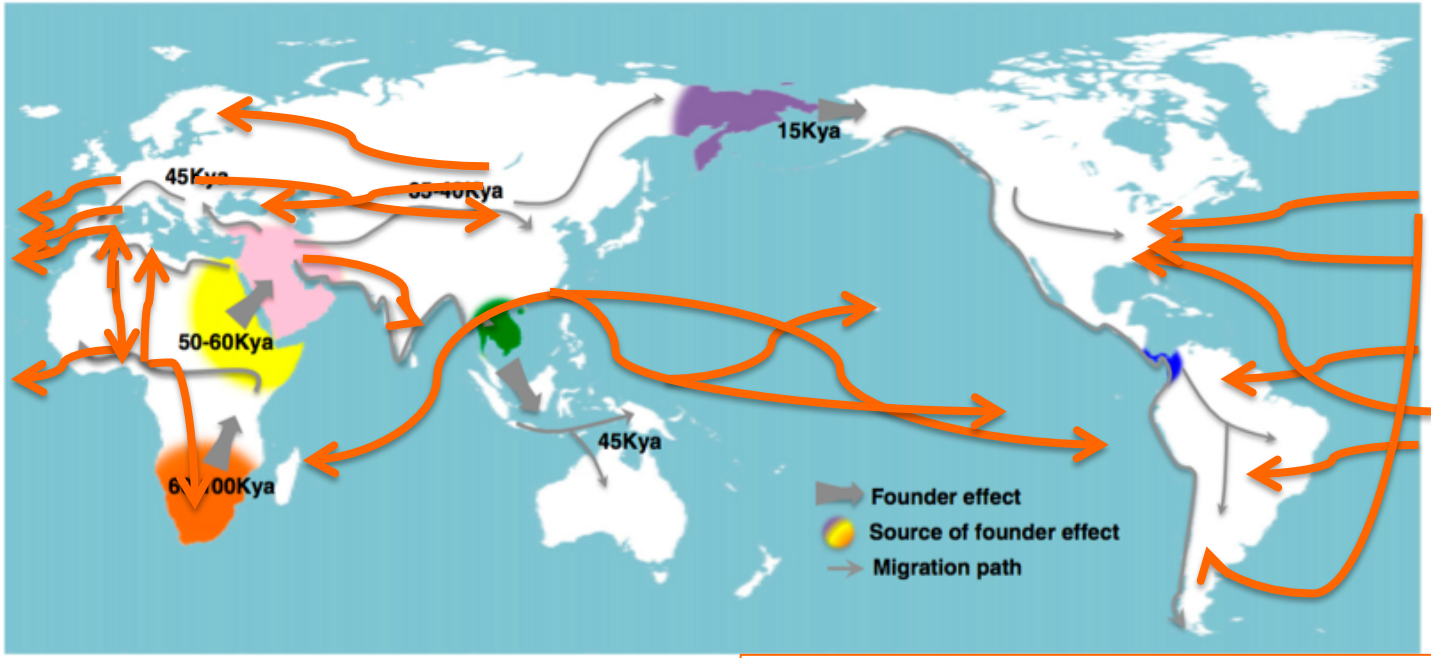
Missing Heritability Ten Years On

Eimear Kenny, PhD
The Charles Bronfman Institute of Personalized Medicine
Icahn School of Medicine at Mount Sinai
2018.05.02

eimear.kenny@mssm.edu
@EimearEKenny

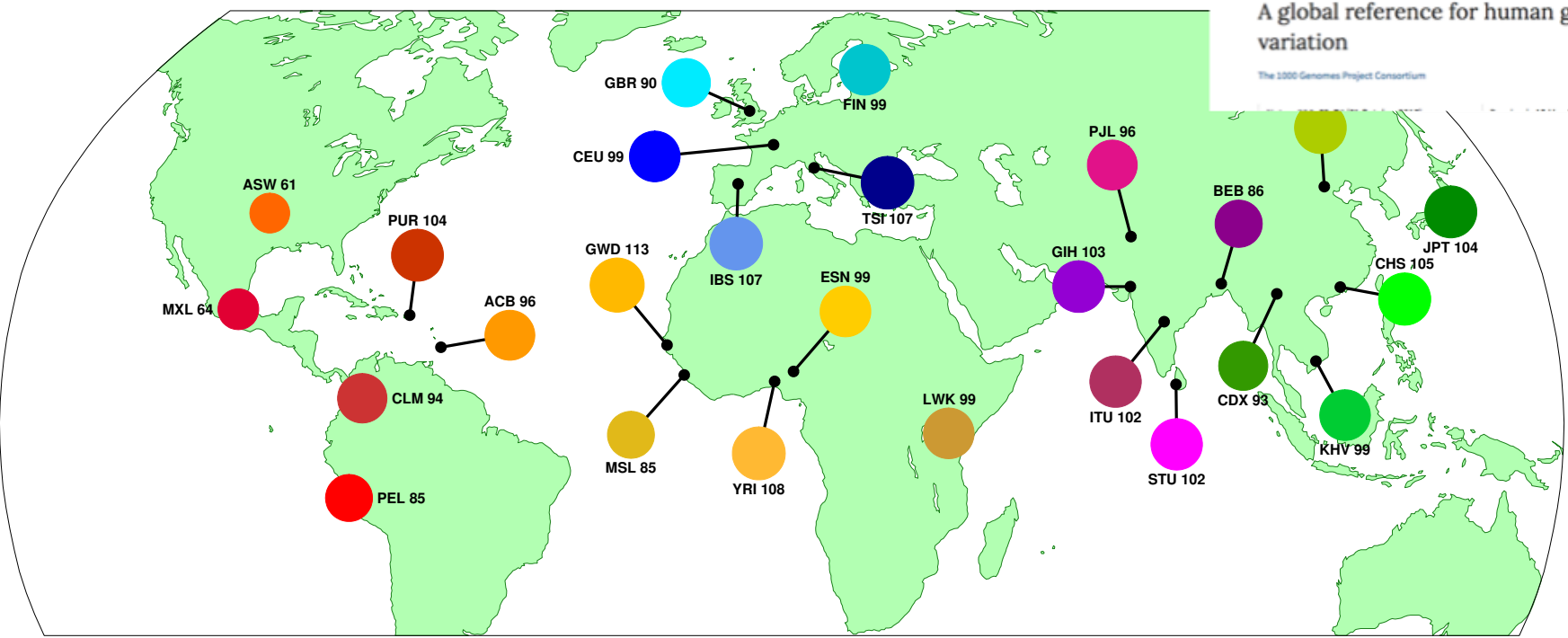
Human Genetic History is Complex!

And has changed dramatically recently:

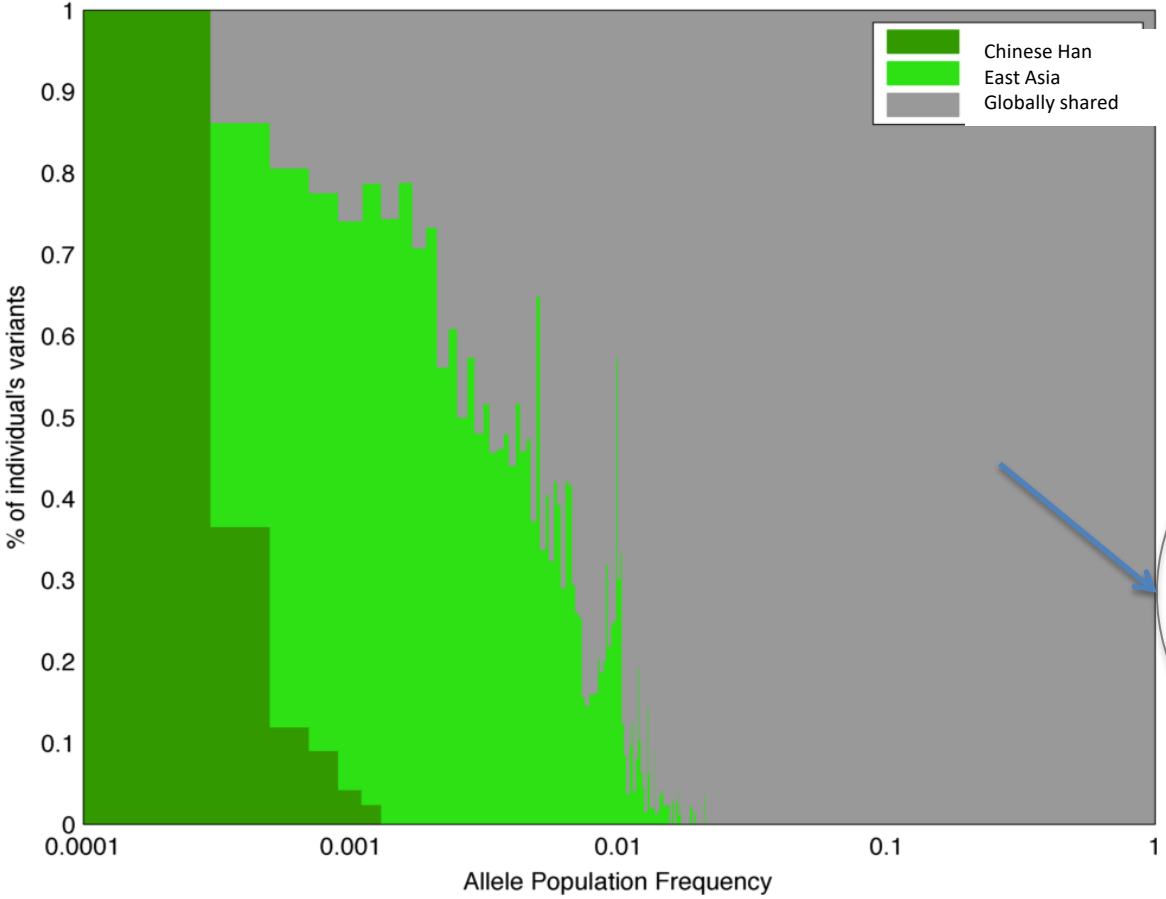


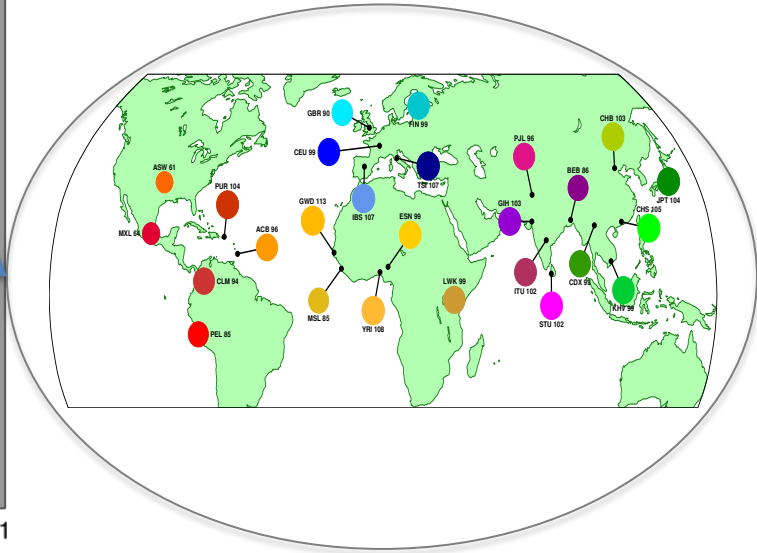
Henn, Cavalli-Sforza, Feldman, PNAS 2012

Genes mirror geography: lessons from the Thousand Genomes Project

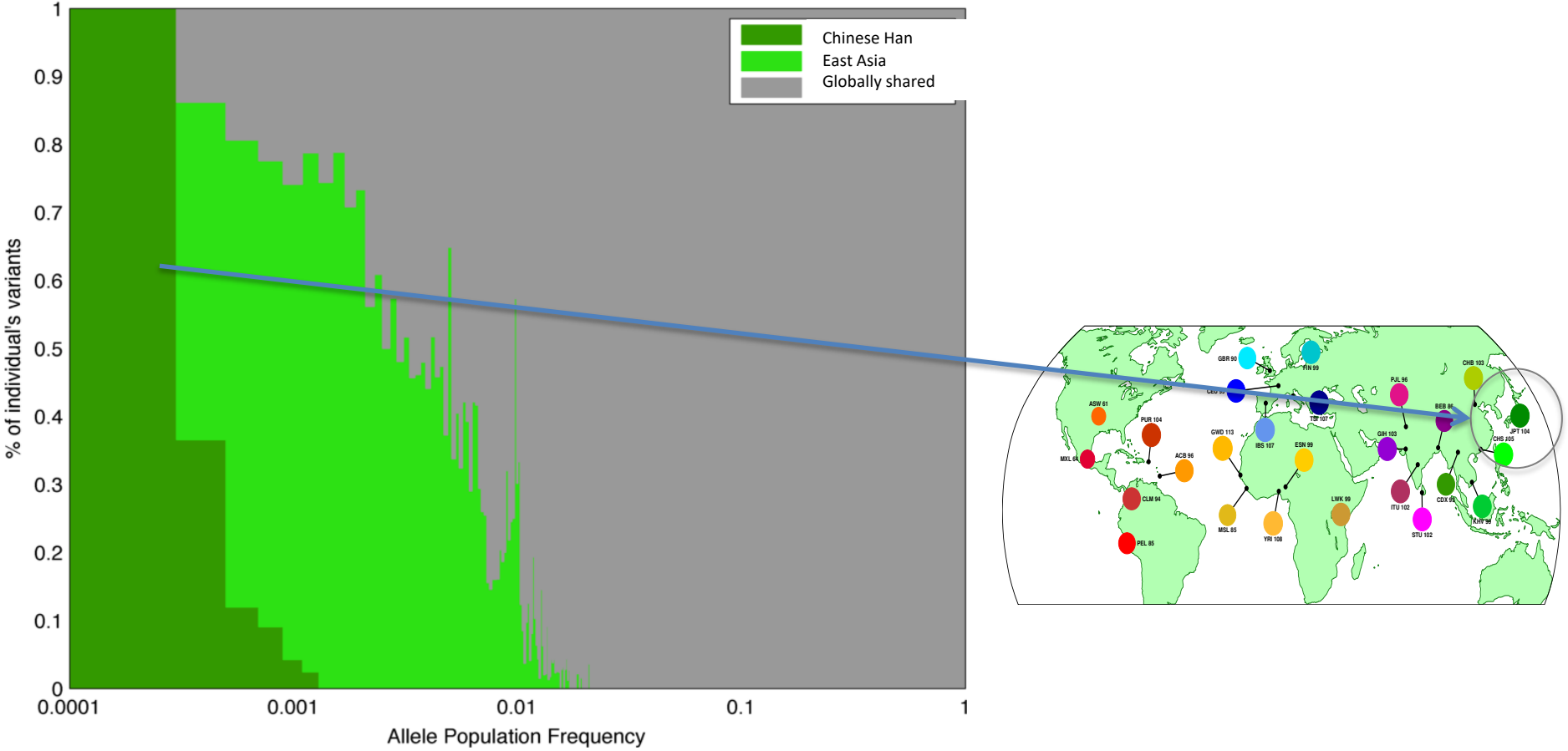


Common variants are shared globally

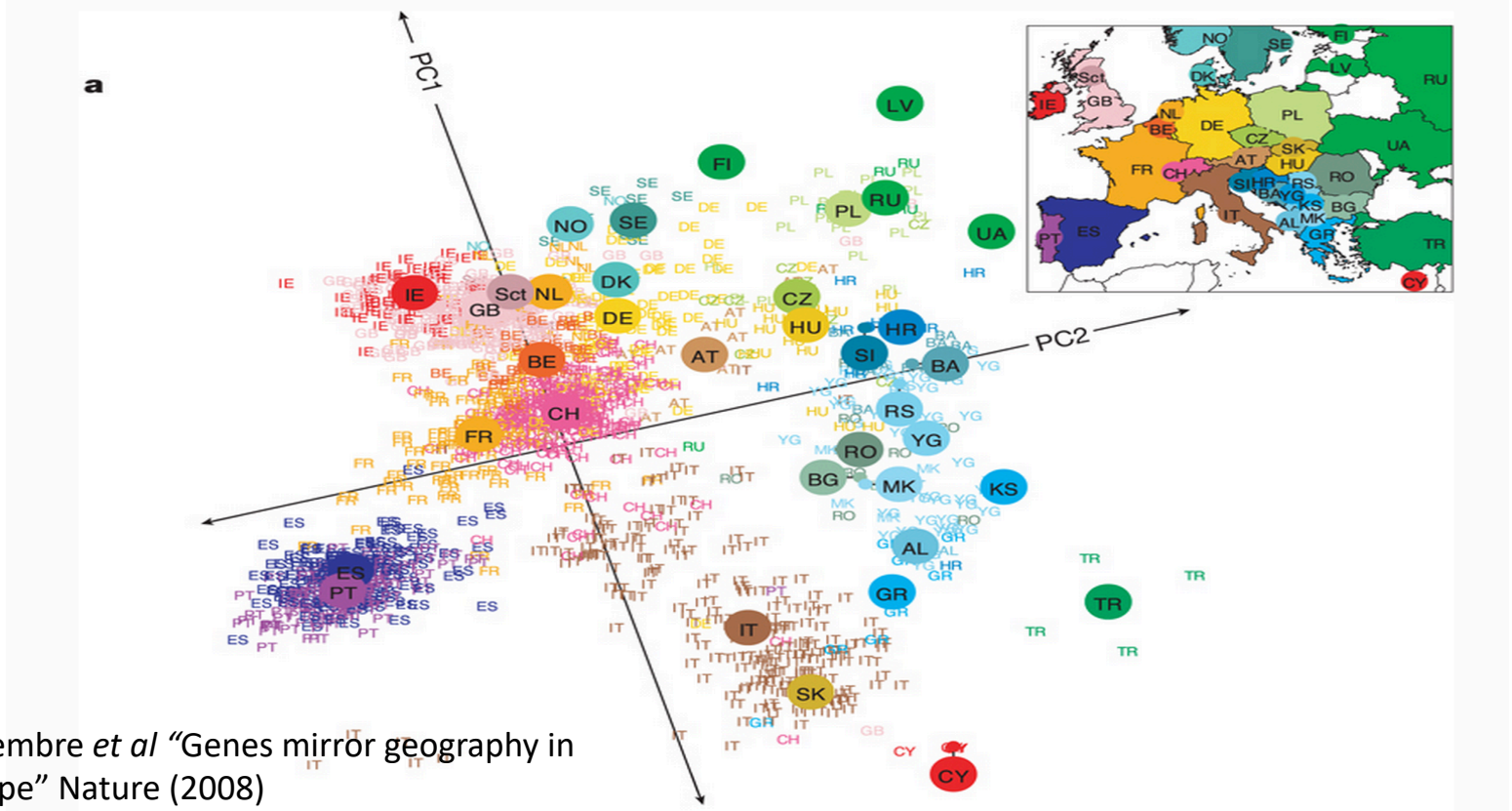




Rare variants are geospatially restricted

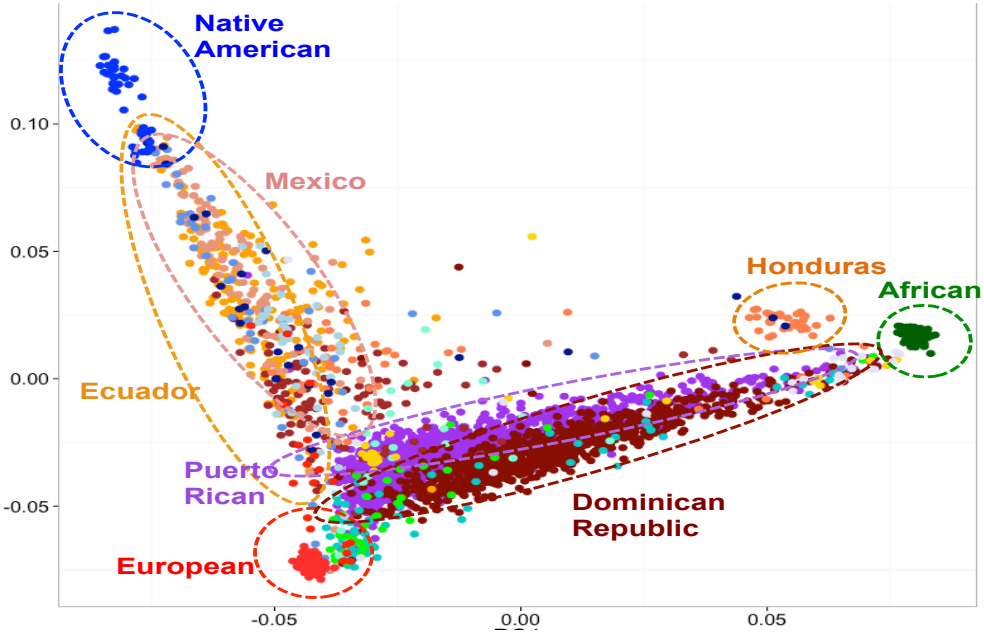


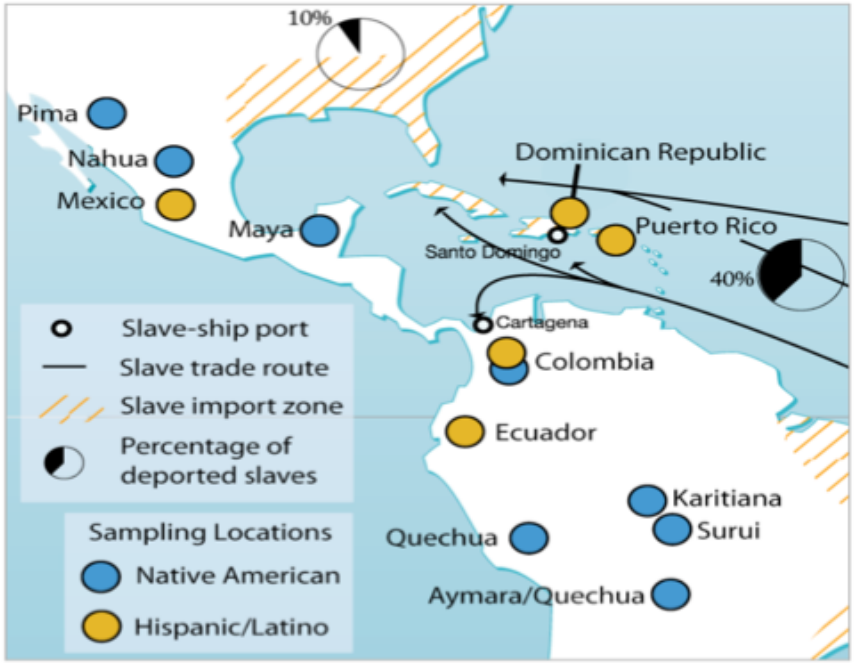
Common variants reveal population structure in Europe



Novembre *et al* "Genes mirror geography in Europe" Nature (2008)

Latin America genetic diversity show signatures of European, African and Native American ancestry





PLOS | GENETICS

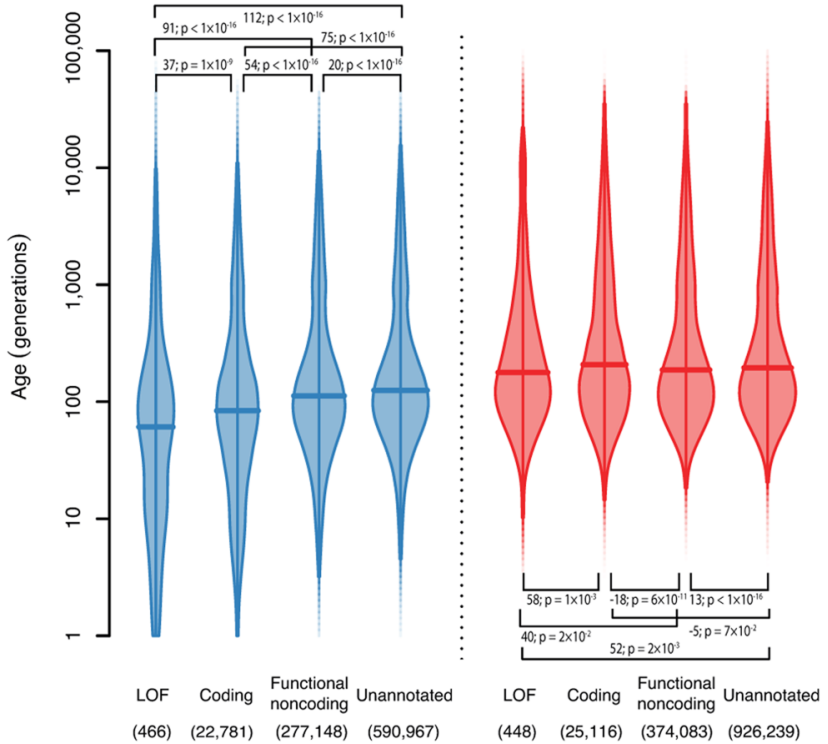
Browse Pul

OPEN ACCESS PEER-REVIEWED

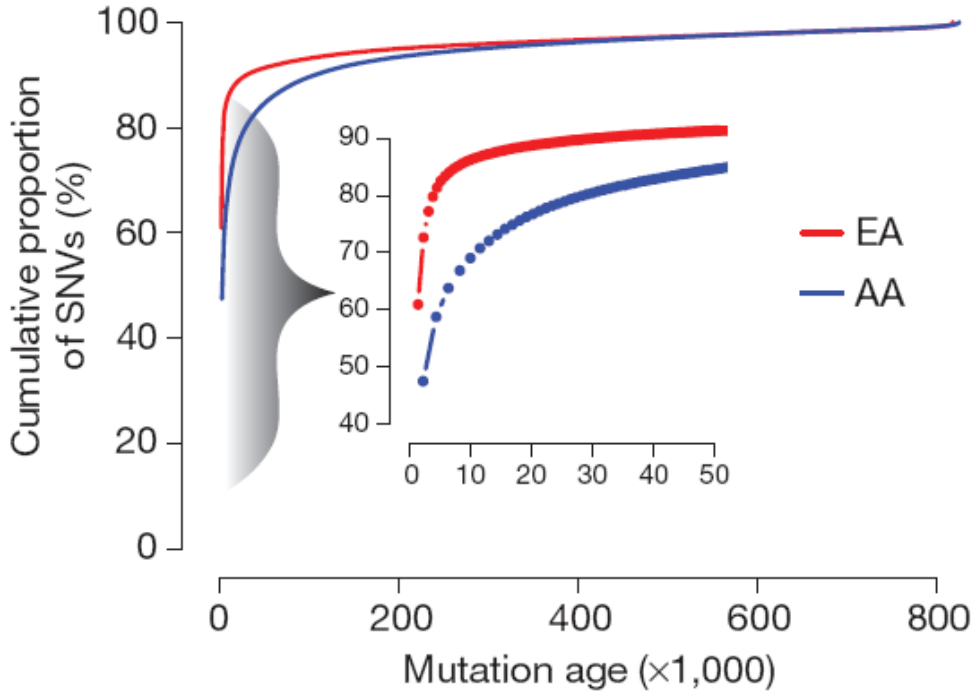
RESEARCH ARTICLE

Demography and the Age of Rare Variants

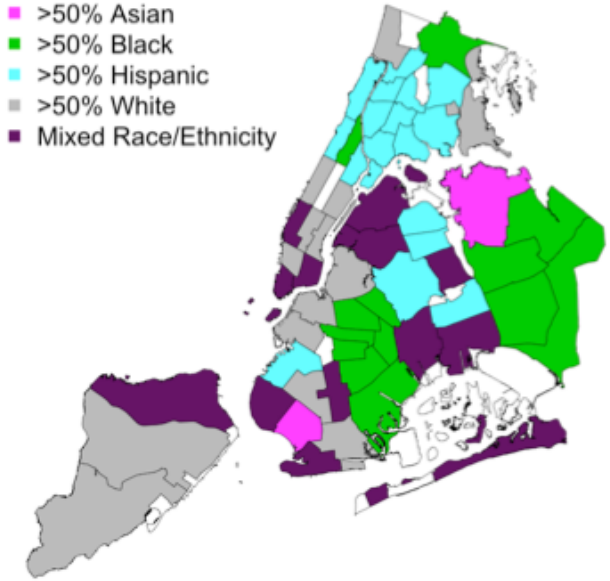
Iain Mathieson, Gil McVean



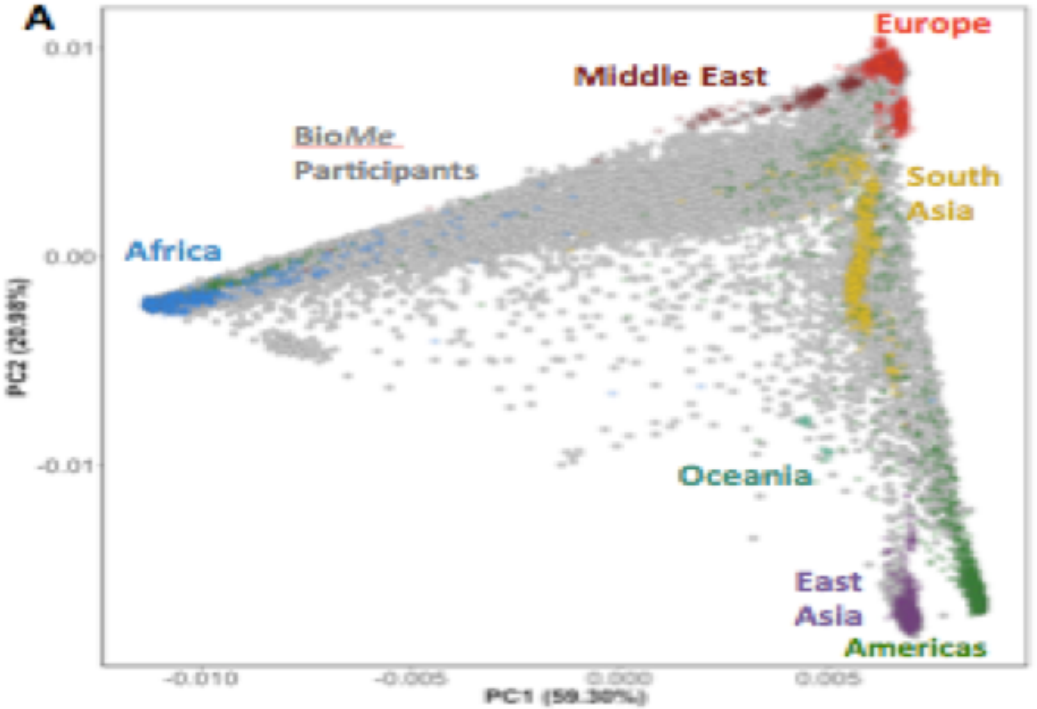
Fu *et al* "Analysis of 6,515 exomes reveals the recent origins of most human protein-coding variation" Nature (2013)



Recent demography expected to play a considerable role in rare variant mapping



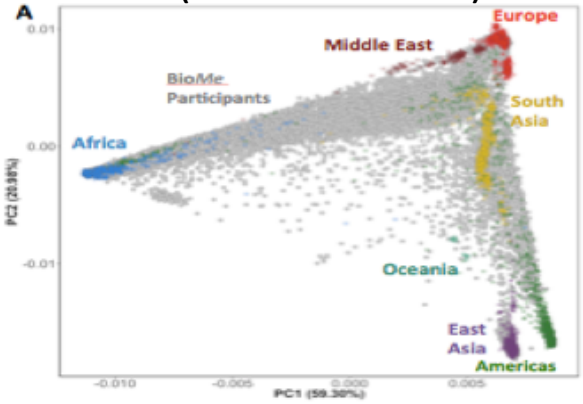
PCA (common SNPs)

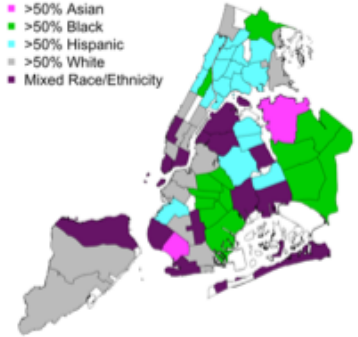


Recent demography expected to play a considerable role in rare variant mapping

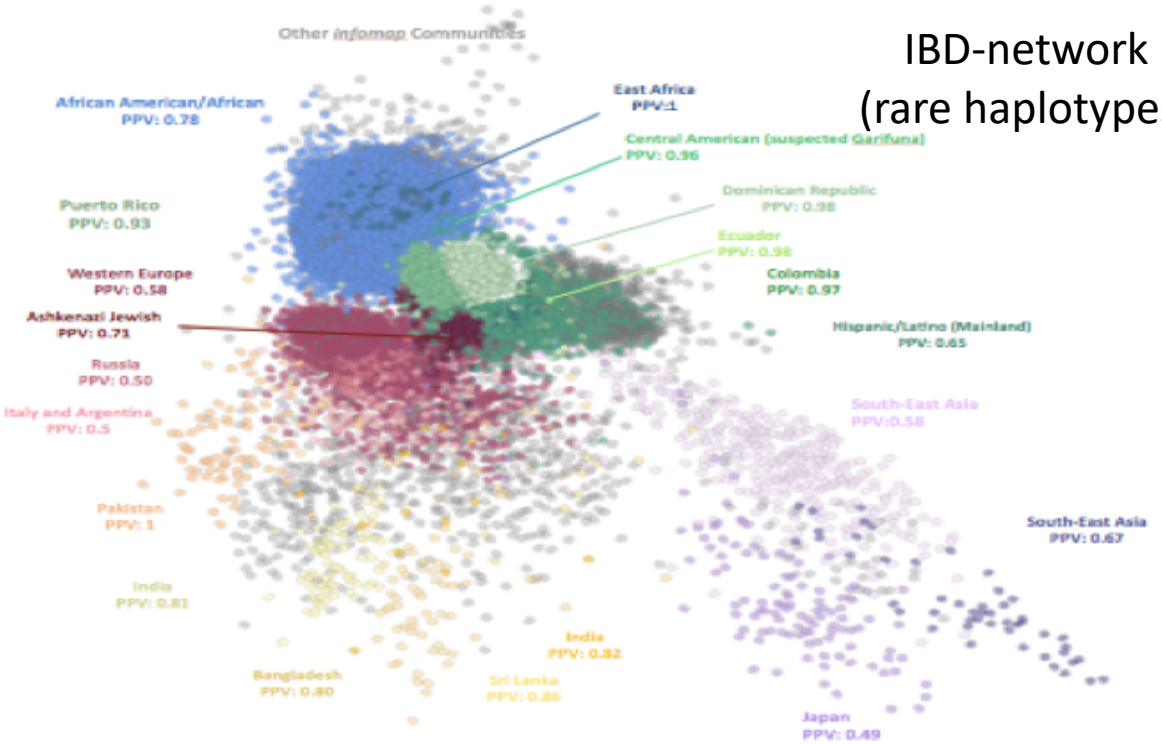
24K New Yorkers (BioMe Biobank)

PCA (common SNPs)

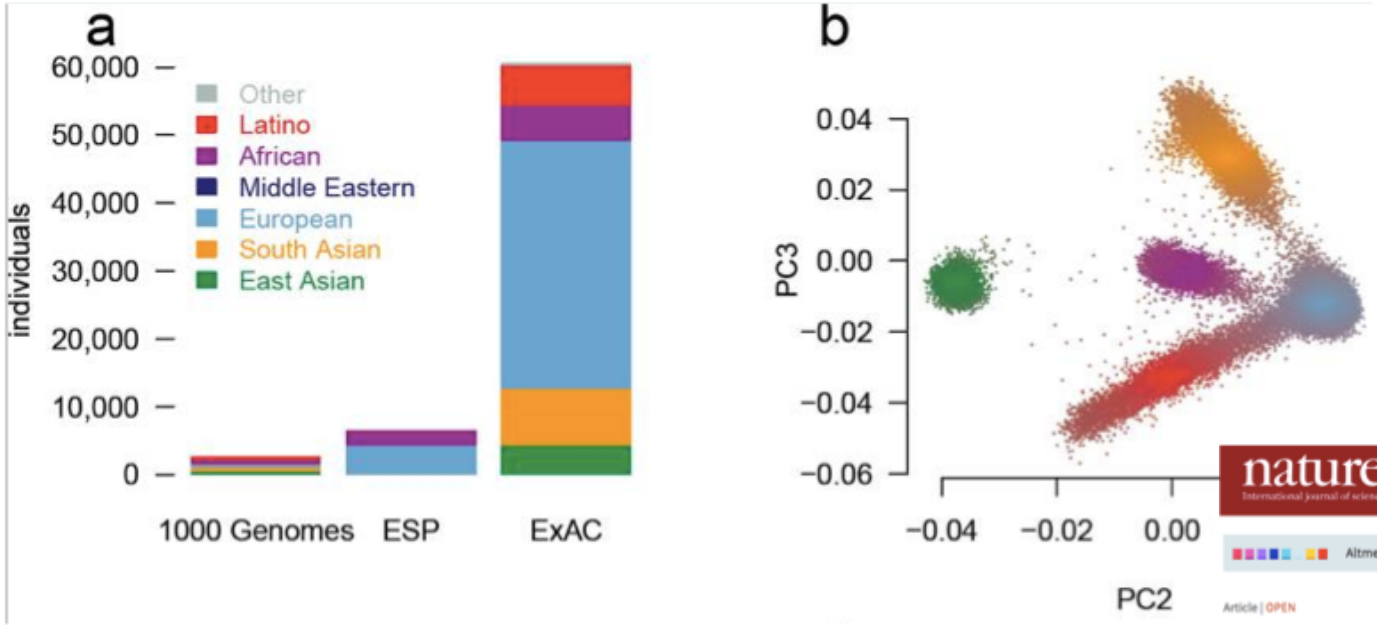


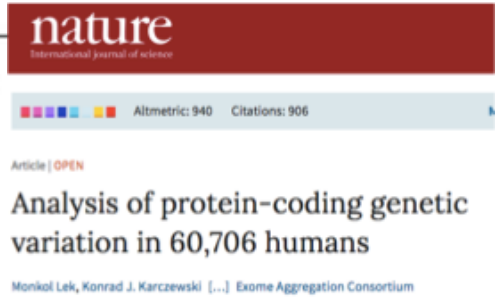


IBD-network
(rare haplotypes)



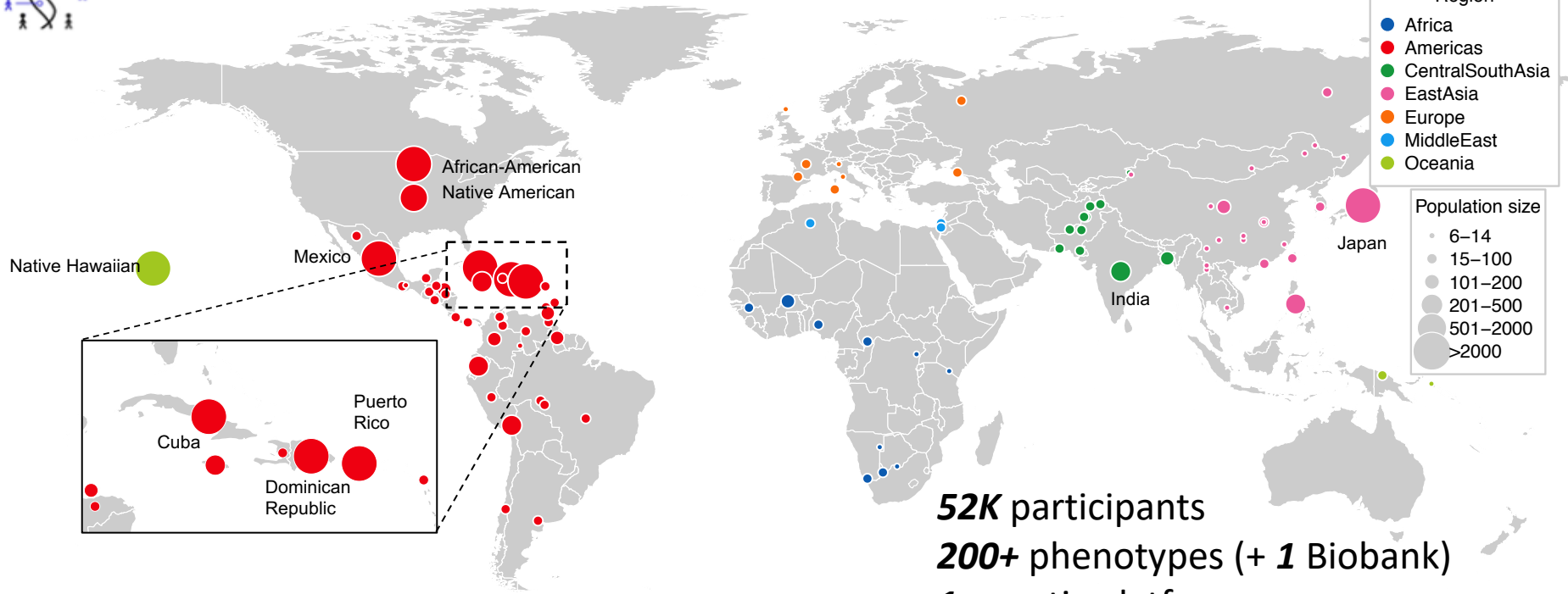
Large and increasingly diverse reference sequence databases





NHGRI Population Architecture using Genomics and Epidemiology (PAGE) Study

www.pagestudy.org



52K participants

200+ phenotypes (+ ***1*** Biobank)

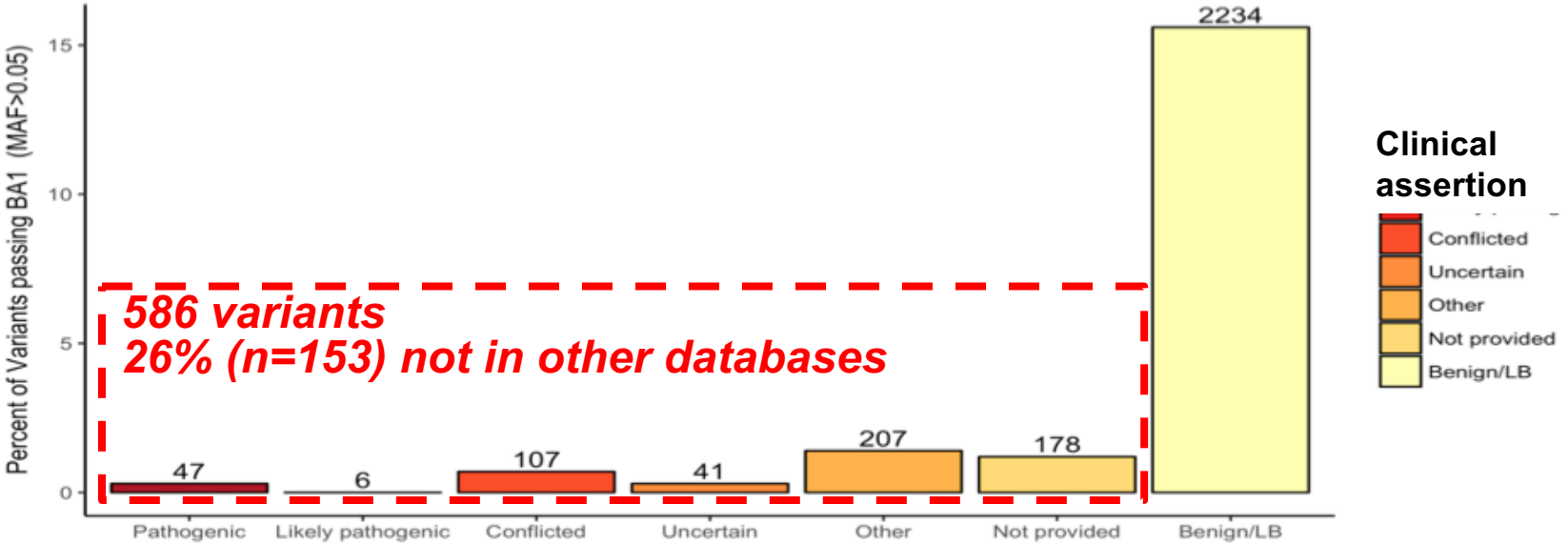
1 genetic platform

Wojcik et al. “Genetic diversity turns a new PAGE in our understanding of complex traits” (2017) *biorXiv*

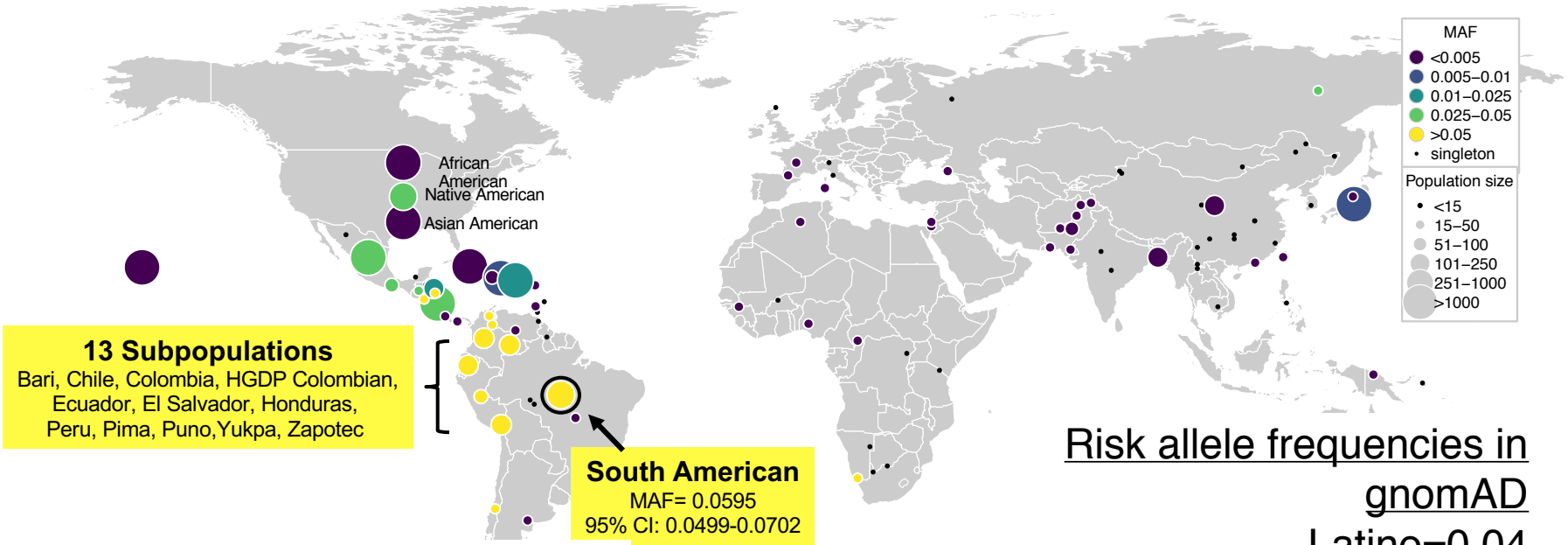
Over 30,000 segregating clinical variants from medical databases

		PAGE	ExAC	1000 Genomes
Study Populations				
Sample Size (N)		51,698	60,706	2,504
Population Labels		99	5	26
Phenotypes (Yes/No)		Yes	No	No
Clinical content (Total, polymorphic)				
Overall		63,902 (36,247)	234,585 (101,203)	237,620 (41,386)
ClinVar	Total	21,720 (15,793)	131,325 (62,350)	132,238 (30,519)
	Pathogenic	7,186 (2,276)	45,180 (9,779)	44,230 (1,643)
	Likely pathogenic	623 (320)		
	Uncertain significance	4,133 (3,805)	51,852 (24,705)	52,822 (8,473)
	Benign/Likely benign	6,483 (6,411)	33,306 (27,201)	34,077 (19,844)
	Conflicted	1,702 (1,567)	310 (294)	299 (216)
	Other	423 (351)	667 (371)	810 (343)
	Not provided	1,172 (1,063)		
	Within ACMG genes	2,496 (1,477)	42,023 (10,358)	43,580 (4,302)
HGMD		16,381 (10,445)	146,304 (32,208)	145,240 (12,279)
Predicted loss-of-function		38,950 (16,233)		
ACMG-59 genes		2,740 (1,596)		

Applying ACMG criteria to PAGE populations reveals 2,820 variants with MAF>0.05



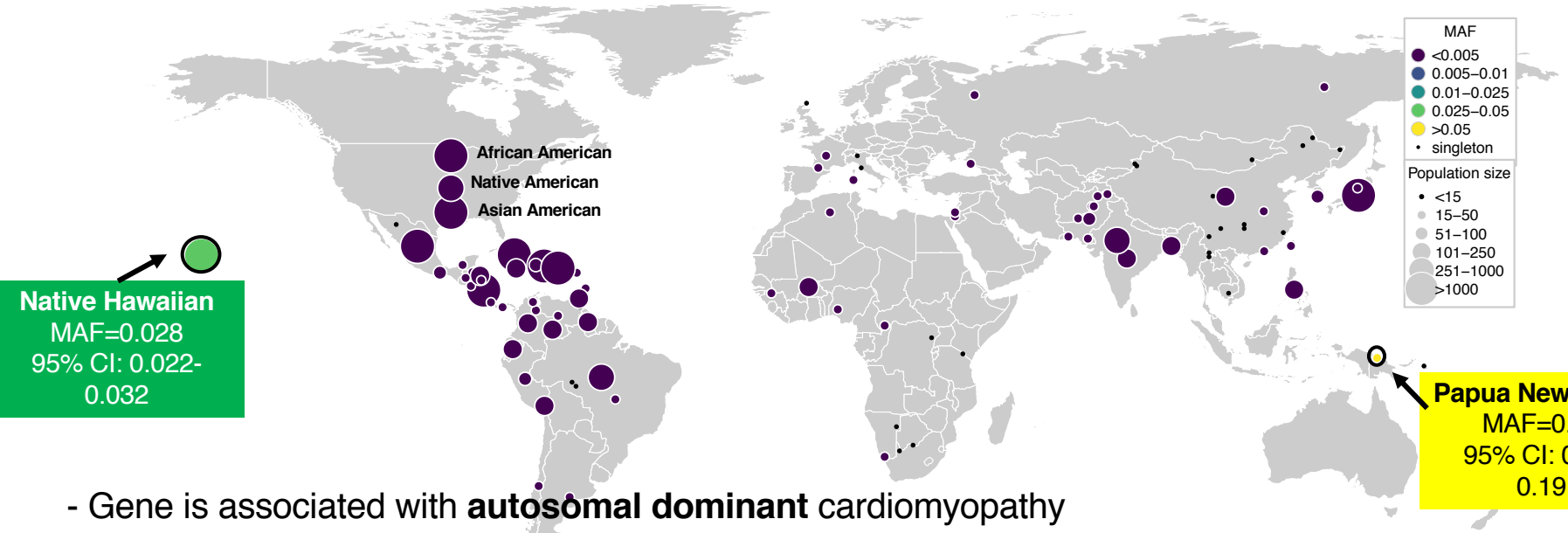
Worldwide frequencies of ClinVar conflicted variant *SCN5A*.pV195L



LongQT syndrome, autosomal dominant
Current status on ClinVar: **Conflicted**
Risk allele is allele shown here: **A**

Risk allele frequencies in gnomAD
Latino=0.04
African=0.003
East Asian=0.0029

Worldwide frequencies of ClinVar conflicted *TNNT2*.c5564G>A splice variant



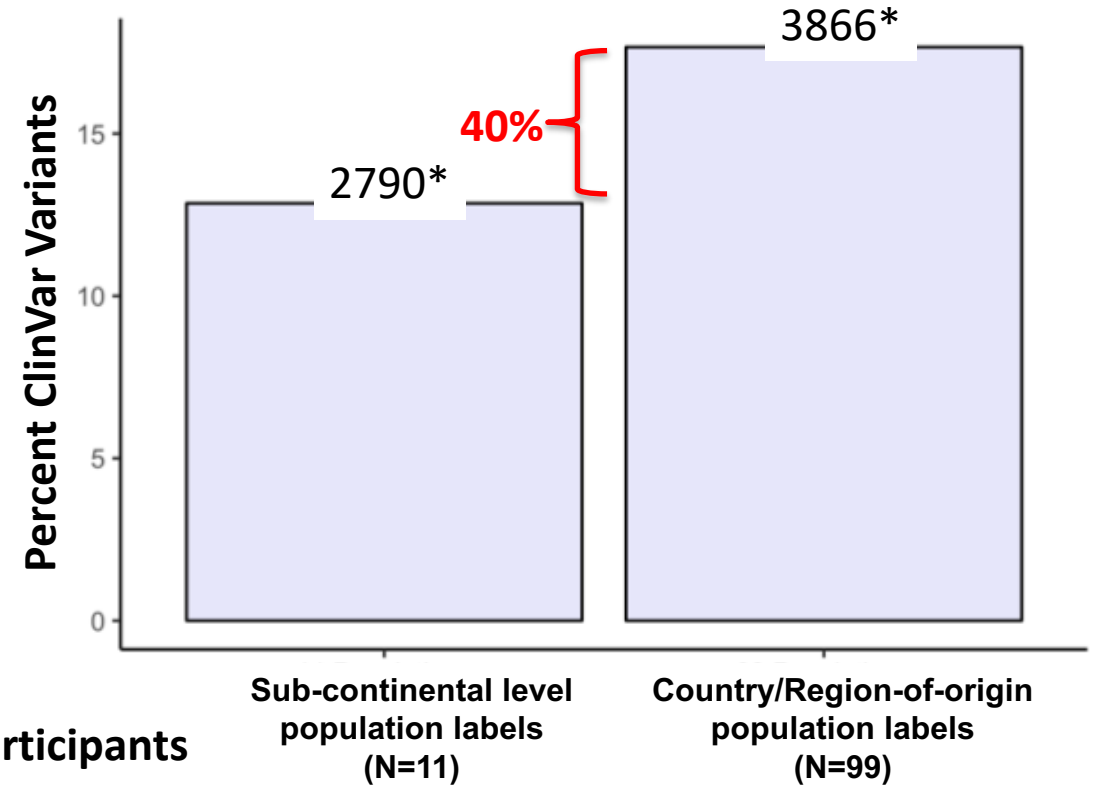
Native Hawaiian
MAF=0.028
95% CI: 0.022-0.032

Papua New Guinea
MAF=0.058
95% CI: 0.016-0.191

- Gene is associated with **autosomal dominant** cardiomyopathy
- Current status on ClinVar: **Conflicted**
- Risk allele is allele shown here: T

Only seen 5 copies in gnomAD
South Asian MAF=0.0001

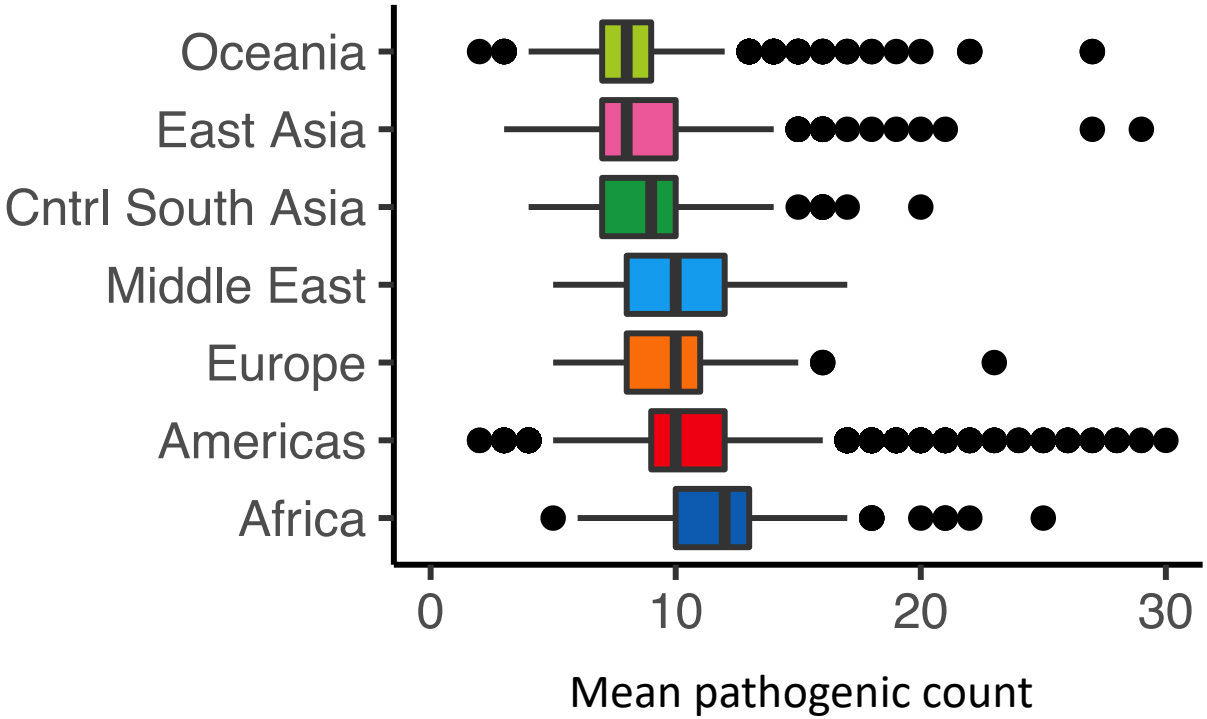
Additional 40% variants above 5% MAF using fine-grained population labels in PAGE



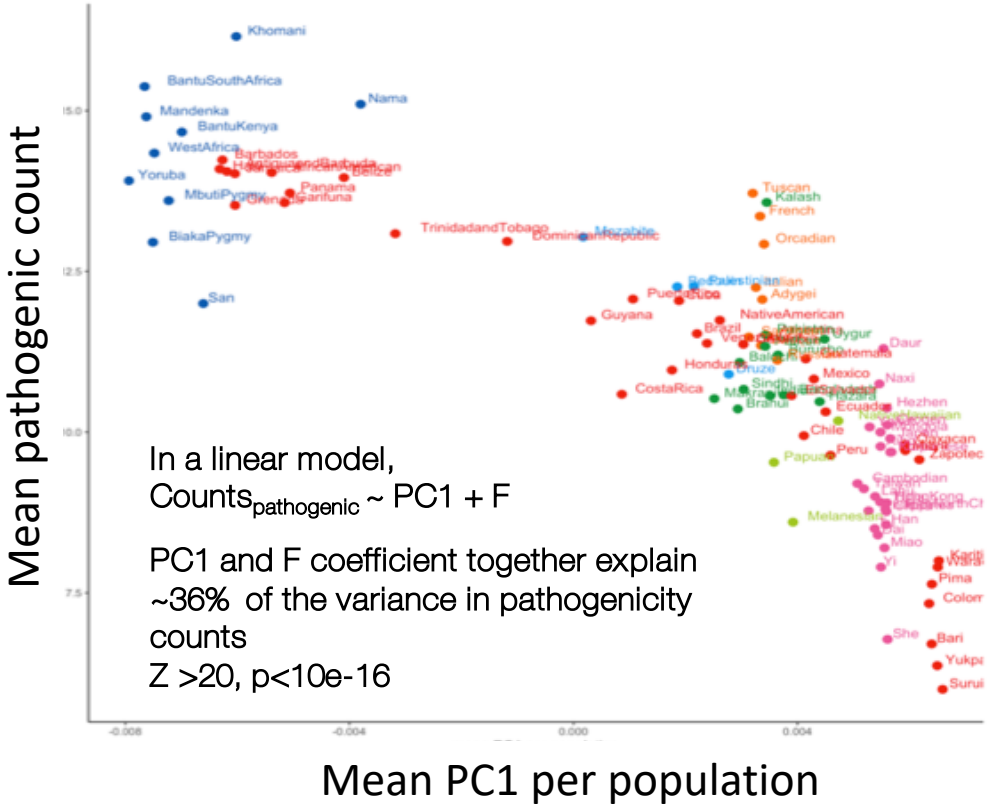
PAGE 52K participants

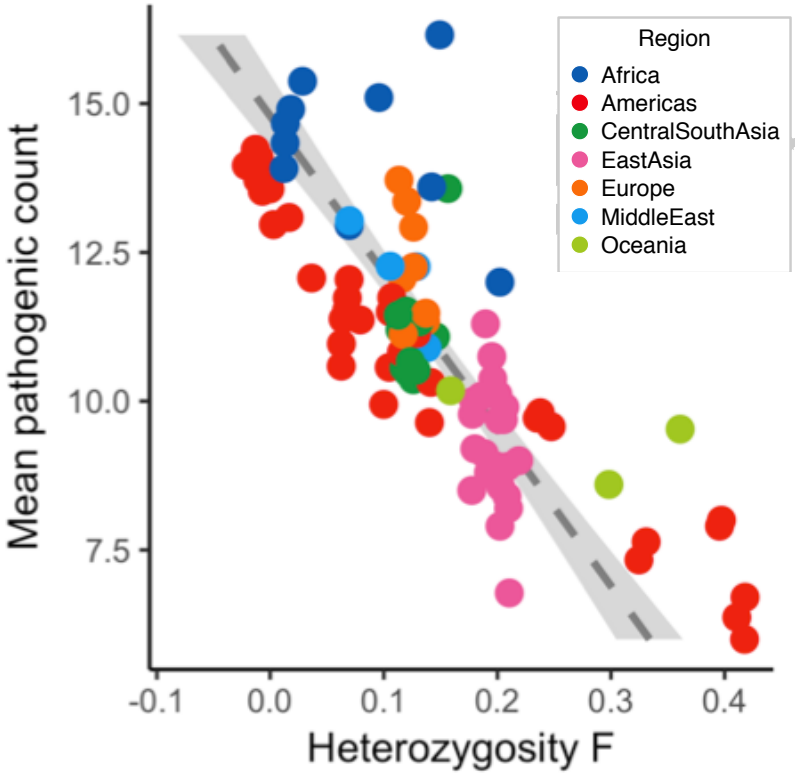
*To account for differences in sample size, we include only those variants where the ***lower bound of 95% confidence interval of risk allele frequency > 0.05***

Understanding genetic ancestry can help reduce false positives

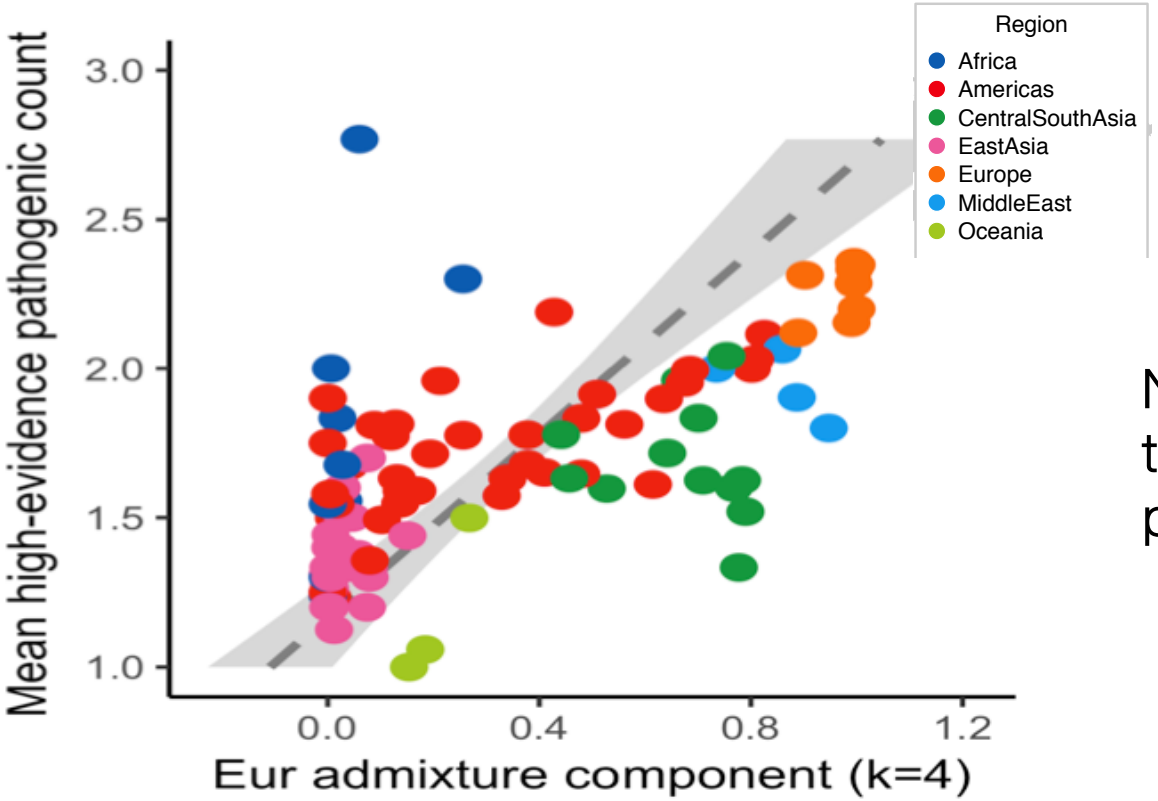


Understanding genetic ancestry can help reduce false positives





Understanding genetic ancestry can help reduce false positives

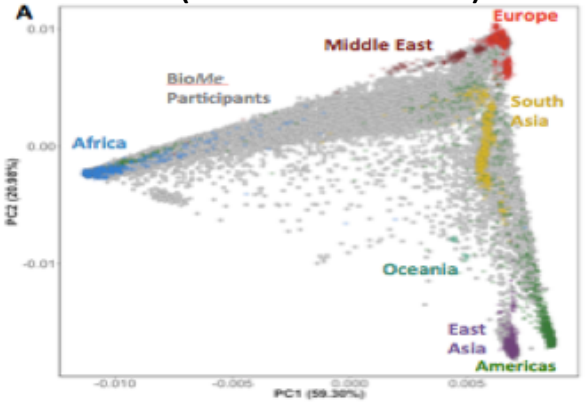


Non-Europeans more likely to have VUS *and* false positive pathogenics

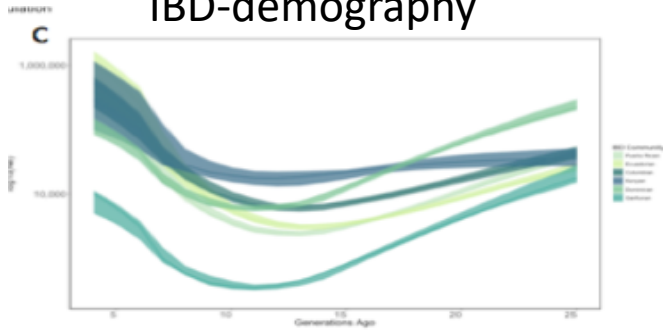
Founder populations can be hidden in plain sight

24K New Yorkers (BioMe Biobank)

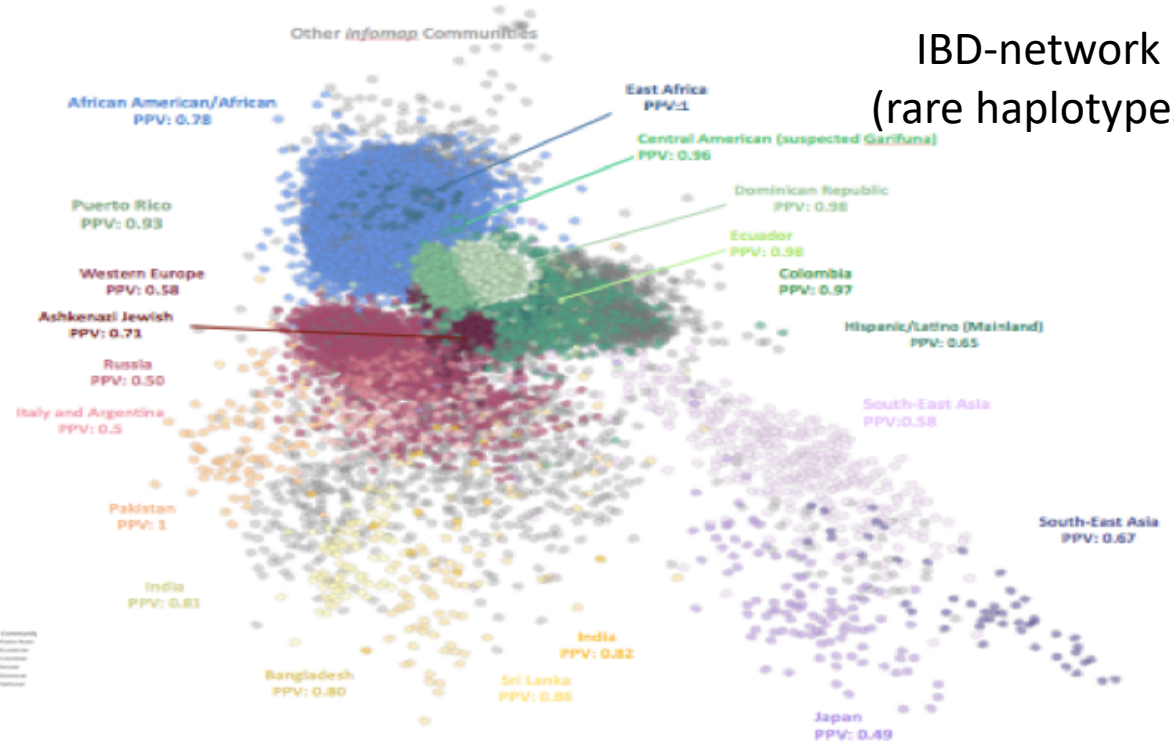
PCA (common SNPs)

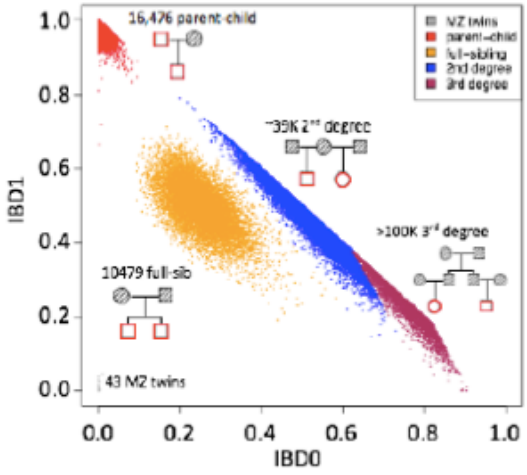


IBD-demography



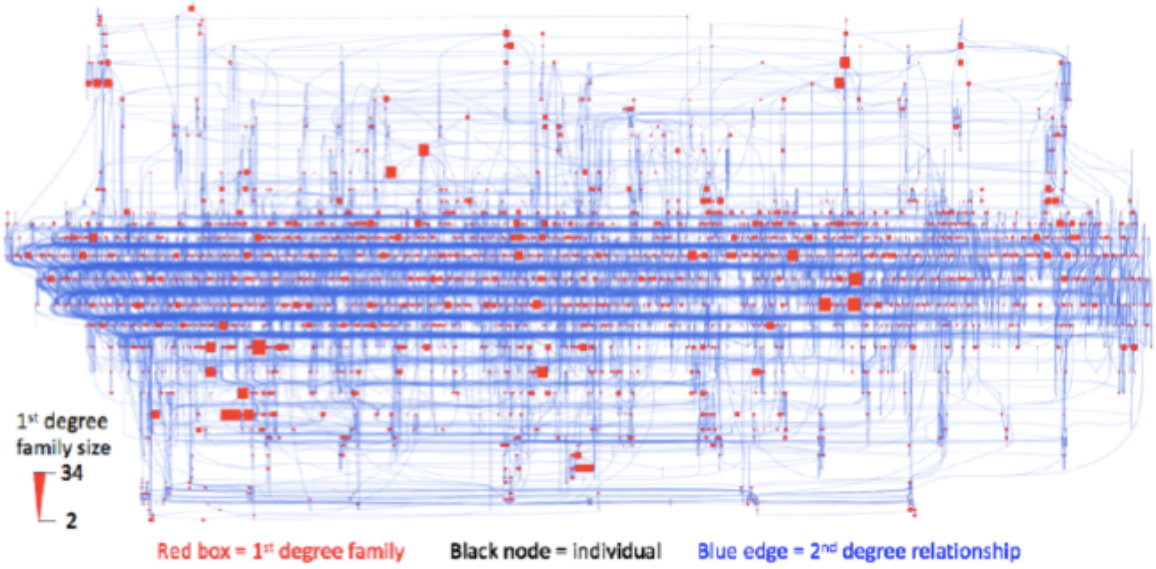
IBD-network
(rare haplotypes)





22% Geisinger cohort in large connected pedigree

D



Cold Spring Harbor Laboratory

bioRχiv

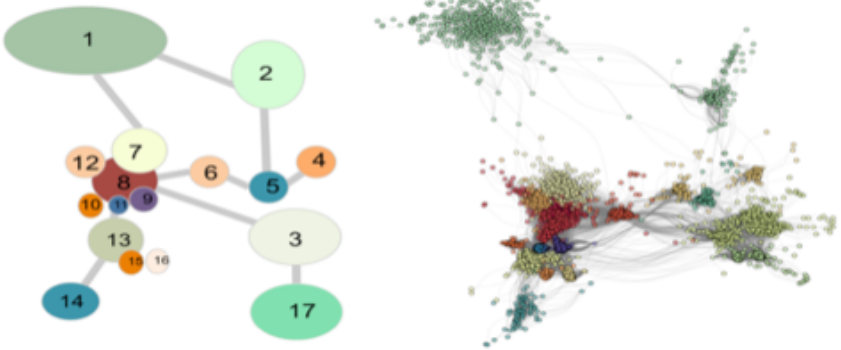
THE PREPRINT SERVER FOR BIOLOGY

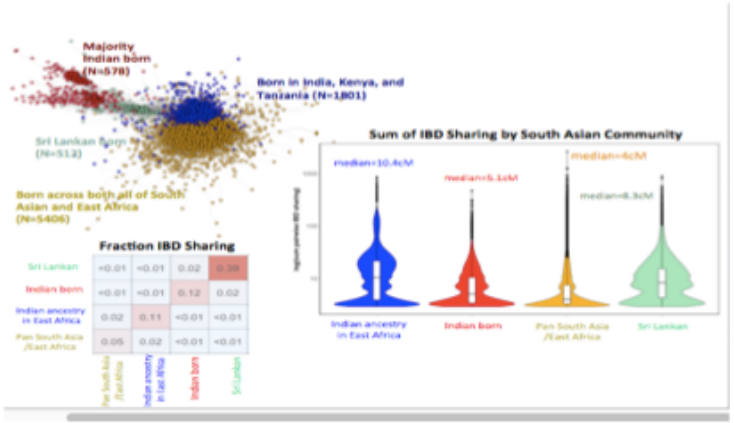
Profiling and leveraging relatedness in a precision medicine cohort of 92,455 exomes

Jeffrey Staples,[1] Evan K. Maxwell,[1] Nehal Gosalia,[1] Claudia Gonzaga-Jauregui,[1] Christopher Snyder,[2] Alicia Hawes,[1] John Penn,[1] Ricardo Ulloa,[1] Xiaodong Bai,[1] Alexander E. Lopez,[1]

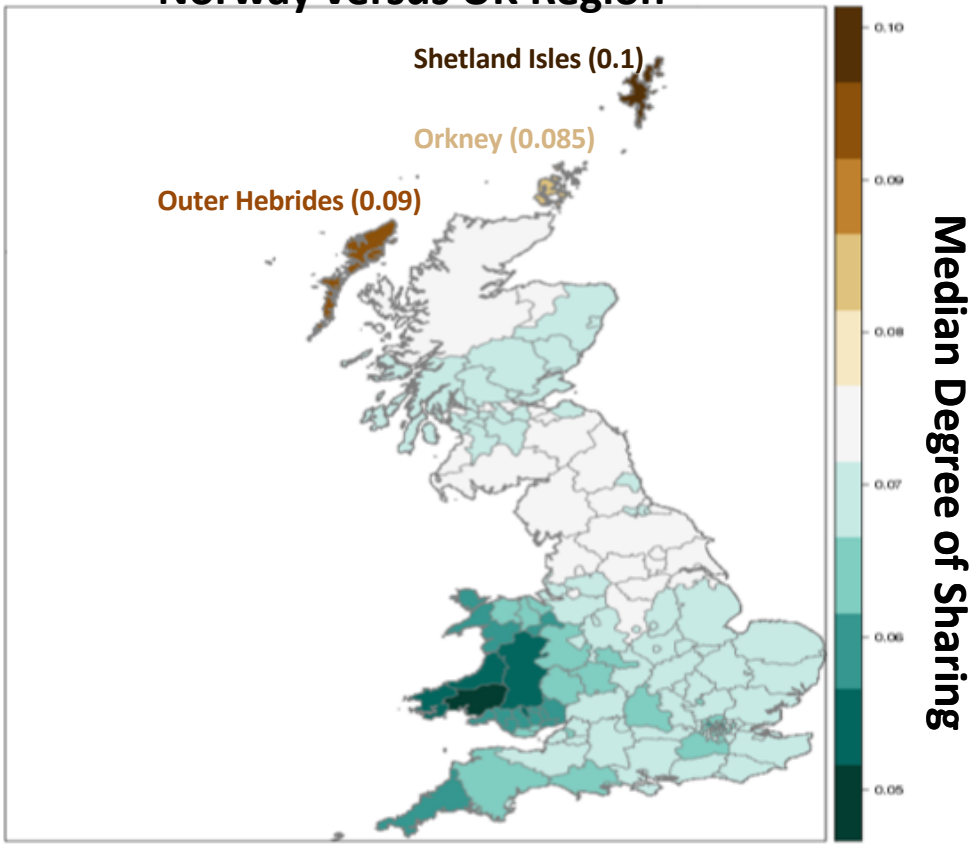
Recent demography expected to play a considerable role in rare variant mapping

14,181,681,196 IBD tracts present in 487,409 UK Biobank participants





Norway versus UK Region



Acknowledgements

Icahn School of Medicine at Mount Sinai

Gillian Belbin
Sinead Cullina
Stephane Wenric
Hannah Poisner
Muh-Ching Yee
Jackie Odgis

Institute of Personalized Medicine

Noura Abul-Husn
Girish Nadkarni
Tielman Van Vleck
Amanda Merkelson
Bernadette Liggayu
Patrick Shanley
Ruth Loos
Erwin Bottinger
Judy Cho

University of Colorado
Chris Gignoux

Stanford University
Genevieve Wojcik
Elena Sorokin
the Bustamante lab
Carlos Bustamante

PAGE-II Collaborators
Lucia Hindorff (NHGRI)
Kari North (UNC)
Chris Carlson (Fred Hutch)
Charles Kooperberg (Fred Hutch)
Chris Haiman (MEC)
Ruth Loos (MSSM)
Tara Matise (Rutgers)
Steve Buyske (Rutgers)
Jose-Luis Ambite (USC)

Funding NIH/NHGRI/NHLBI
U01HG109391
U01HG009080
X01HL134588