Regeneron Genetics Center

Challenges and opportunities in the analysis of more than a million participants from diverse populations

2023 International Statistical Genetics Workshop

Timothy Thornton

Senior Director
Analytical Genetics
Regeneron Genetics Center

10.March.2023

REGENERON

NEVER STOP ASKING WHY

George Yancapoulos | Len Schleifer

RGC™

Regeneron Genetics Center

RGC is proud to sponsor the International Scholar and Cultural Exchange Program (ISCEP) for the International Statistical Genetics Workshop 2023!

Regeneron Genetics Center (RGC)

Established In 2014 And Is Now One Of The Largest Operational Human Sequencing Efforts

SAMPLE BIOBANKING

LIBRARY PREPARATION AND EXOME CAPTURE

ILLUMINA-BASED SEQUENCING

CLOUD BASED INFORMATICS & ANALYSIS

Mission:

Taking large scale human genetics to the next level for target discovery, support existing targets and identify novel indications

Regeneron Genetics Center:
Unprecedented Speed, Scale & Integration

~500,000 additional exomes to be sequenced annually

~2M exomes sequenced to date

120+ research collaborations

RGC
Regeneron Genetics Center

NUMEROUS existing targets & development programs validated

MULTIPLE potential new drug targets identified

>95% of genes with identified LOF carrier(s)

All accomplished in just the first 8 years!

RGC has the most diverse collection and catalogue of human coding variation to date

The New York Times

Hospital and Drugmaker Move to Build Vast Database of New Yorkers' DNA

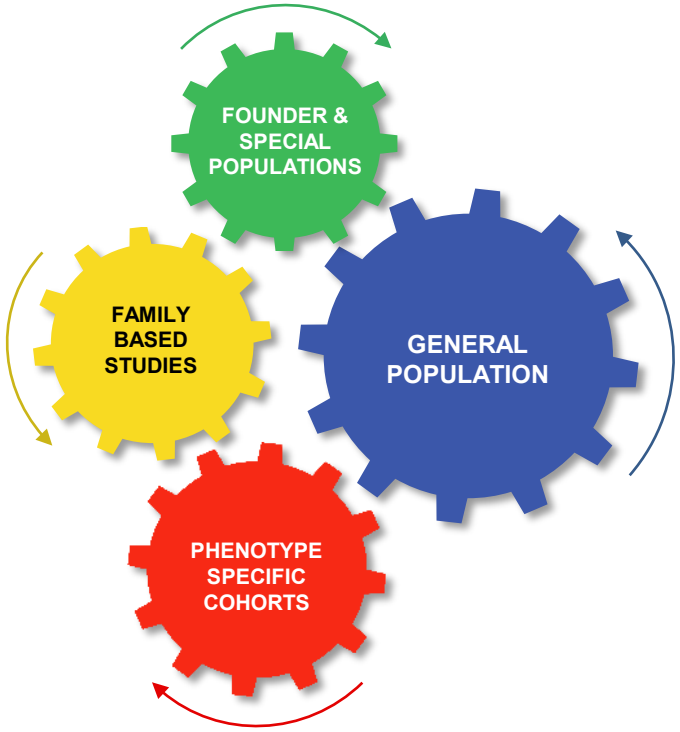
Patients will be asked if their genetic sequence can be added to a database — shared with a pharmaceutical company — in a quest to cure a multitude of diseases.

Wilbert Gibson is a Mount Sinai patient who agreed to let the hospital system use his genetic information in research for treatment of a variety of diseases. Hiroko Masuike/The New York Times

Leveraging Resources Across Genetic Architectures & Phenotypes

120+ Research Collaborations – Over 2,000,000 exomes sequenced to date

Integrated approaches across genetic trait architectures . . .



. . . will power genomic discovery

General Population

Geisinger

biobank uk — Improving the health of future generations

MCPS

LABioMed

Lund University

Phenotype Specific Cohorts

Feinstein Institute for Medical Research — Northwell Health

University of Utah Health Care

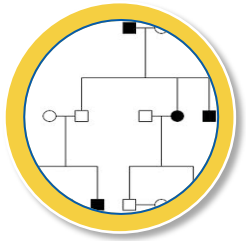
Mayo Clinic

UConn Health Center

UT Southwestern Medical Center

Duke University

Penn



Family Studies

Columbia University Medical Center

BCM — Baylor College of Medicine

The University of Iowa

Children's National Medical Center

Mario Negri Istituto di Ricerche Farmacologiche

SickKids

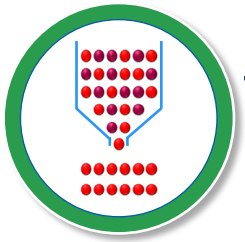
NIH — National Institutes of Health

Clinic for Special Children

DDC clinic — Center for Special Needs Children

Schneider Children's Medical Center of Israel

Rambam Medical Center



Founder & Special Populations

University of Maryland School of Medicine

Penn

NIH — National Institute of Mental Health

NIH — National Institute of Diabetes and Digestive and Kidney Diseases

Global Gene Corp

CNCD — Center for Non Communicable Diseases

Einstein — Albert Einstein College of Medicine of Yeshiva University

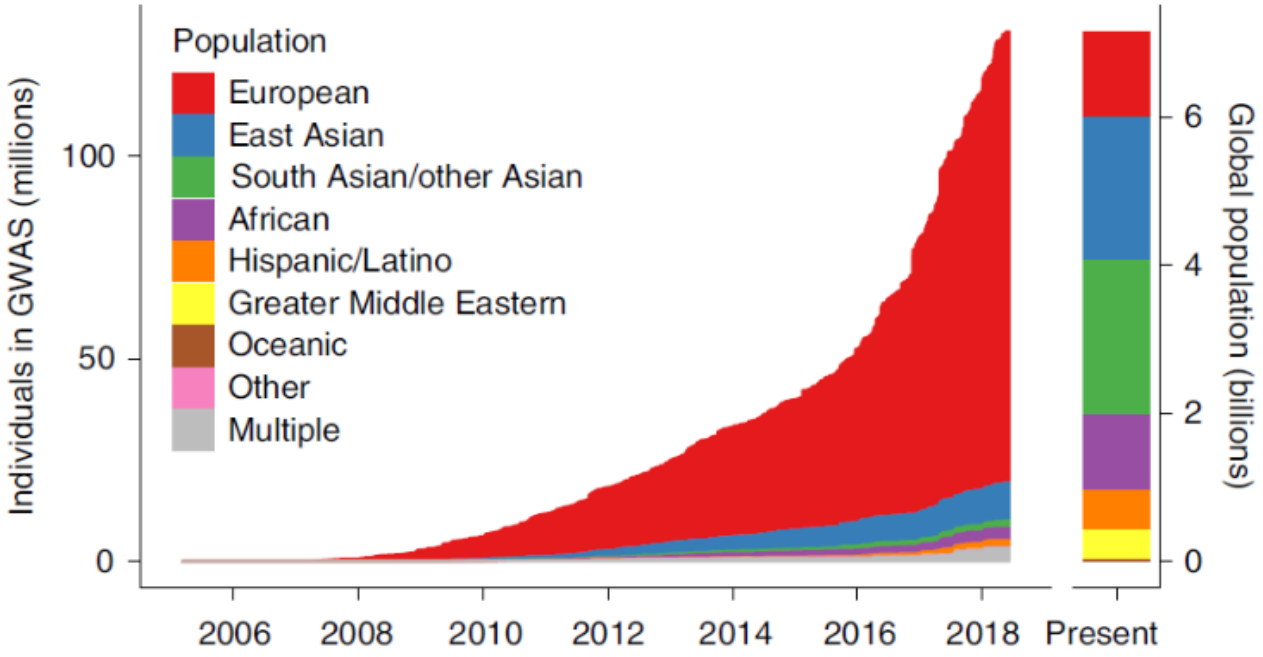
Case Western Reserve University — Est. 1826

ECOGENE-21 — The University of Edinburgh

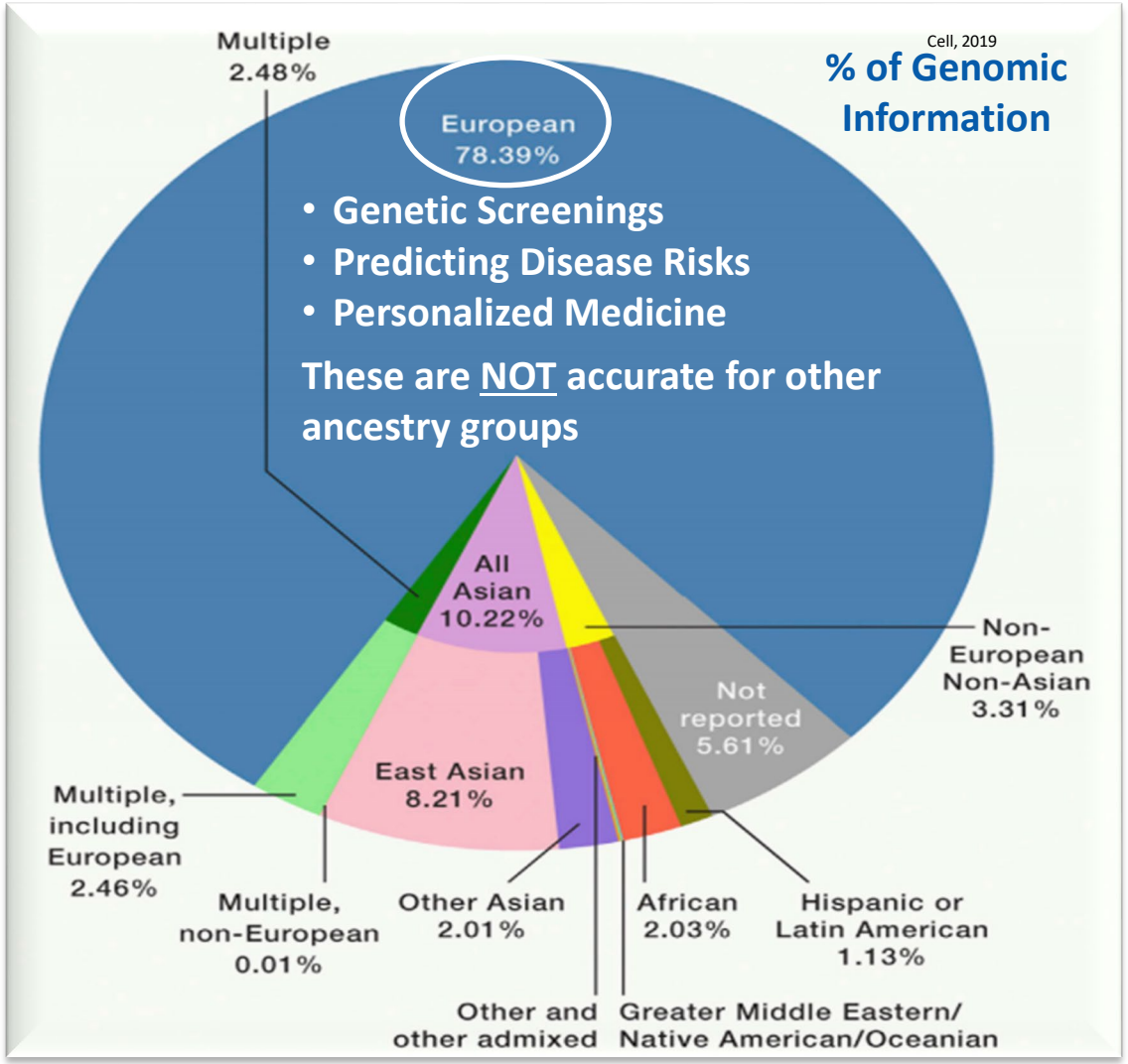
REGENERON

Genomic Diversity is Lacking…BUT the RGC is Building Diversity

Ancestry of GWAS Participants Over Time (compared with the global population)



Martin et al. (Nature Genetics, 2019)

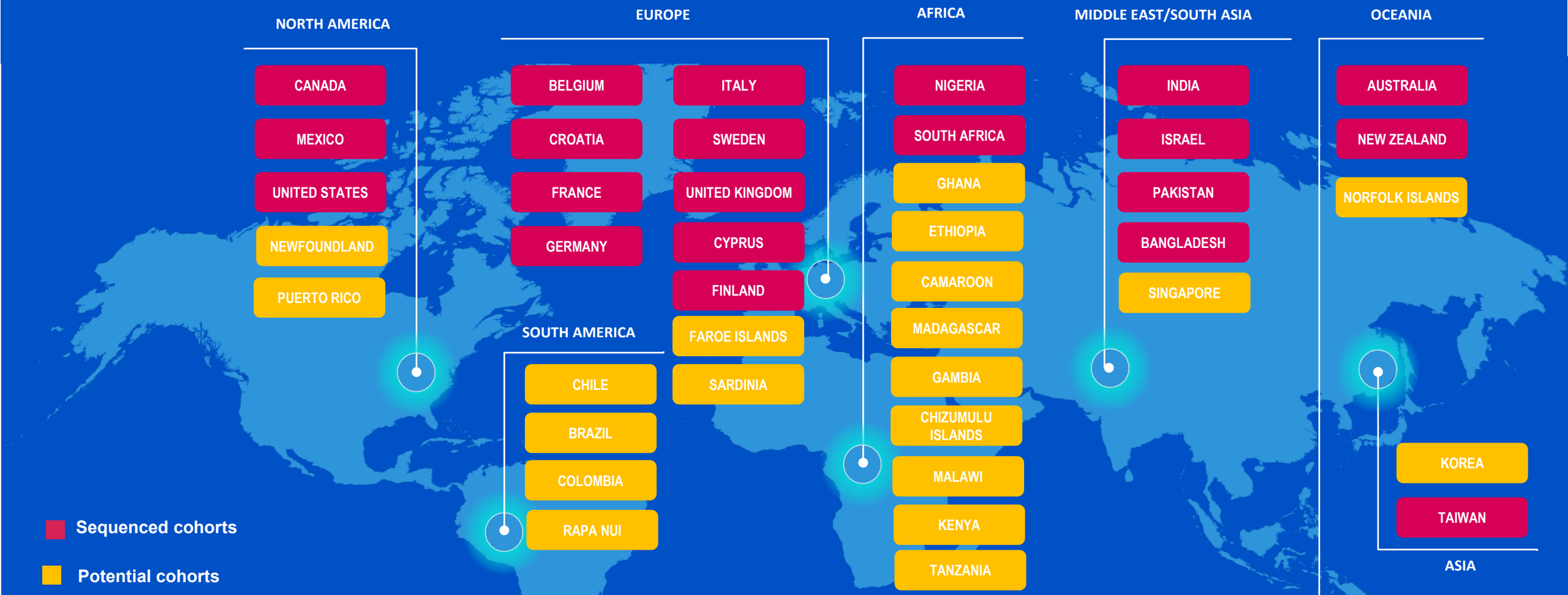


2021 Season

The Players Tribune Jun 4, 2021

RGC™

RGC COLLABORATIONS AROUND THE WORLD



- Over 300,000 individuals of African, South Asian, East Asian and Admixed American ancestry

Technologies at RGC Include:

EXOMES, ARRAYS & IMPUTATION

- Target protein coding exons at depth >20x
- Results in ~20,000 coding variants per individual

- Genotype or capture 0.5 – 1.5M common variants
- Impute remaining variants using reference panel

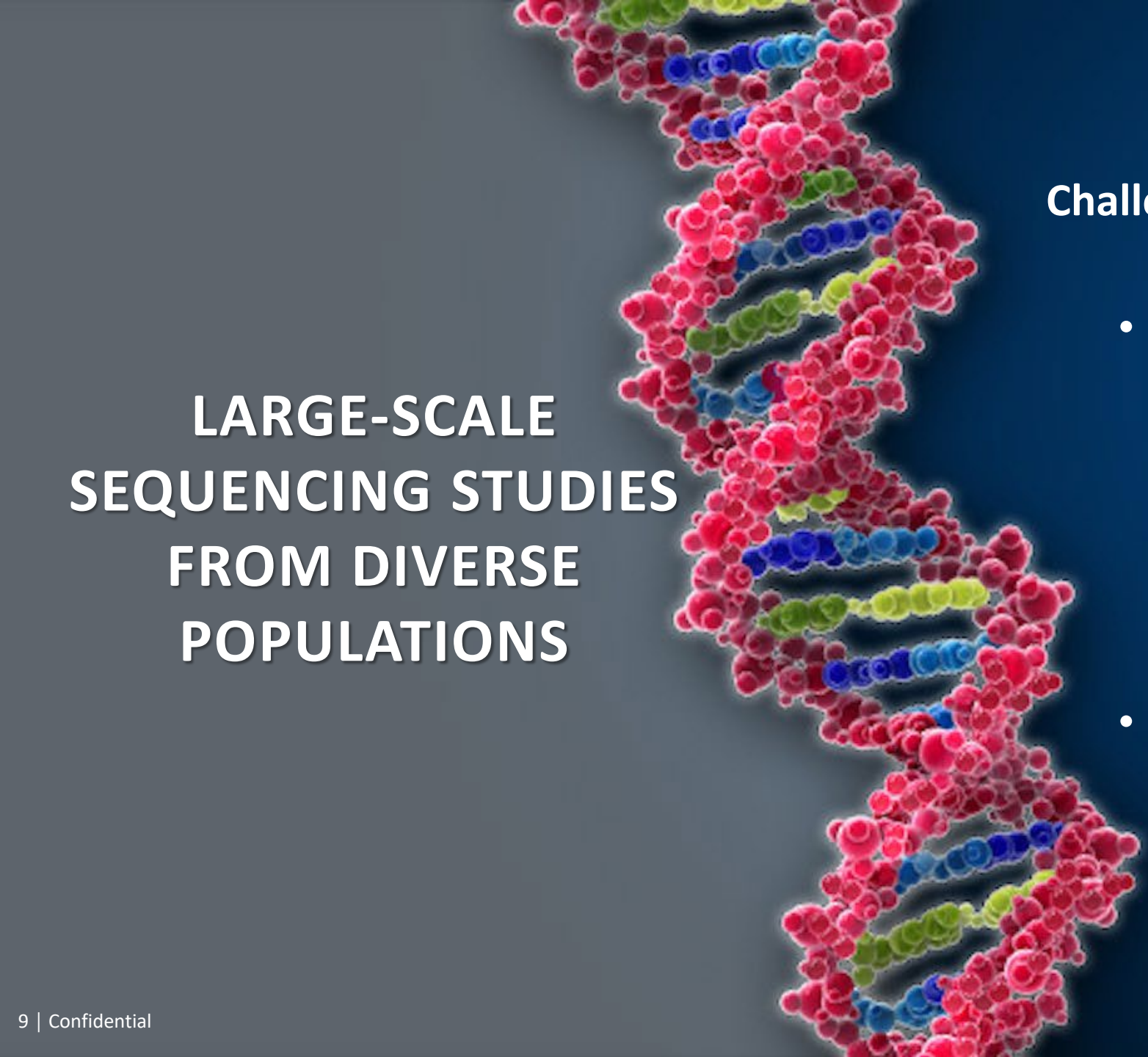
- Platforms are mature
- Analyses strategy evolving (imputation references)

WHOLE GENOME SEQUENCING

- Sequence entire genome at depth of ~30x

- Platforms evolving (e.g. read-length, amplification)
- Analyses strategies evolving (e.g. mapping, assembly)

LARGE-SCALE SEQUENCING STUDIES FROM DIVERSE POPULATIONS



Challenges:

- **Computational pipelines that can accommodate large-scale sequencing and genotyping data on more than a million study subjects from diverse populations**
- **Appropriately accounting for (and leveraging) diverse and admixed genomes that are essential for a variety of downstream genetic analyses**

LARGE-SCALE SEQUENCING STUDIES FROM DIVERSE POPULATIONS

Opportunities:

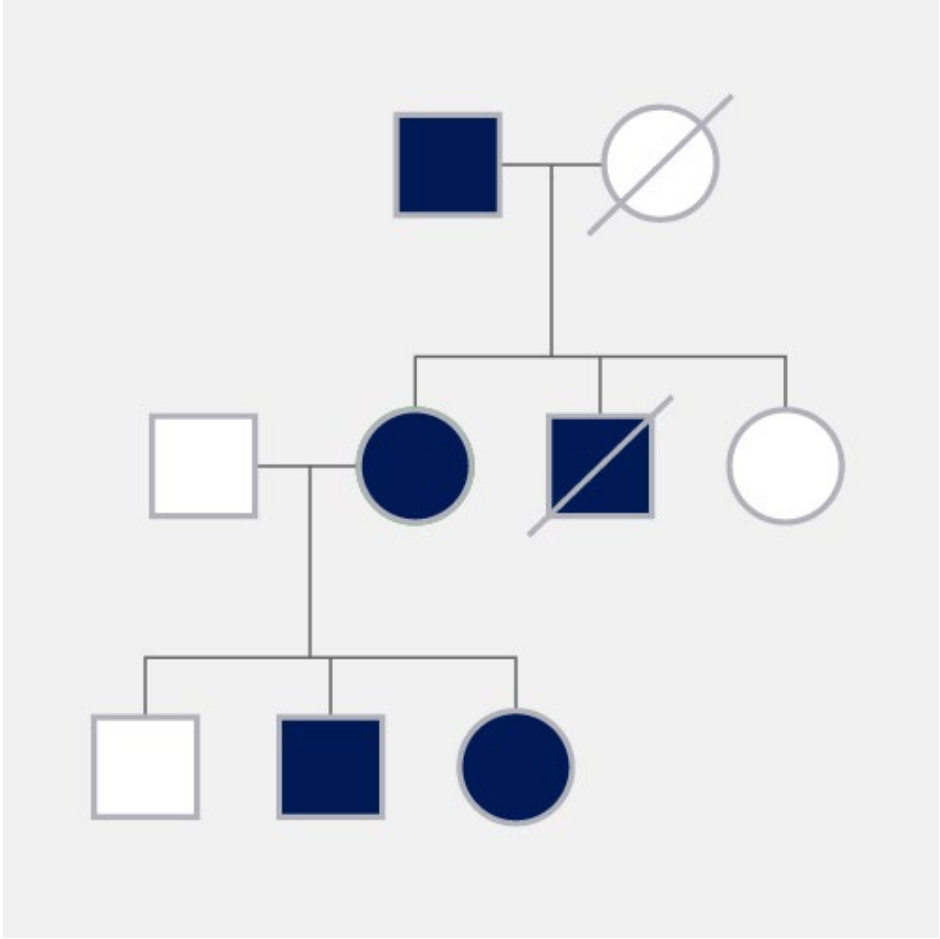
- **Identification of novel variants underlying phenotypic diversity and new therapeutic targets**
- **New insights into human health and health disparities, particularly for underserved populations**
- **Characterization of the genetic architecture of worldwide populations**

Computational pipelines for analysis of 1+ million samples from diverse populations

Relatedness estimation/inference and pedigree reconstruction in large-scale samples

Relatedness inference in complex studies

- Relatedness inference, estimation, and correction is essential for validity of many down-stream genetic analyses
- Large-scale studies often include related individuals
- Requires specific attention:
  - Accounting for relationships in PCA
  - Can induce spurious associations
- Challenging in many settings:
  - Can fail in settings with admixture
  - Computationally costly for biobank-scale



Relatedness Estimation Approaches

- There are two main approaches for estimating/inferring relatedness for pairs of individuals from genome-wide data
  - Methods based on average allele sharing statistics across the genome for a pair
  - Methods based on the length and number of segments inferred to be shared identical-by-descent (IBD) across the genome for a pair

Relatedness Estimation: Average Allele Sharing

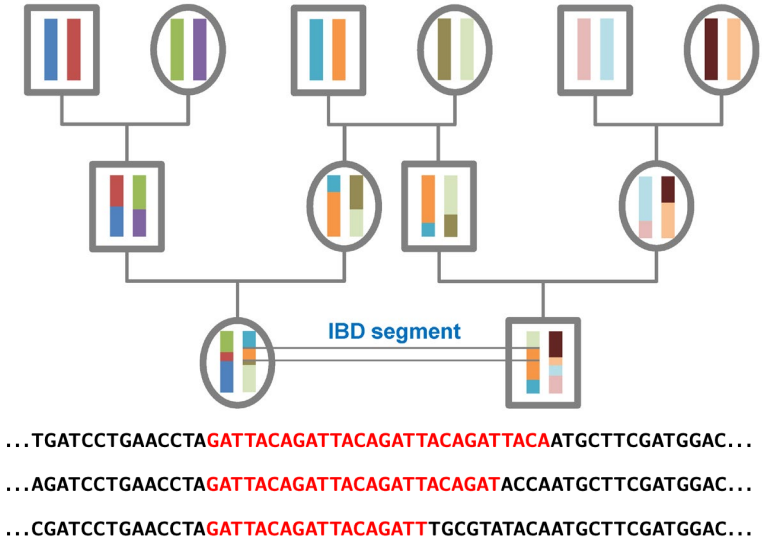
- PLINK (Purcell et al., *AJHG* 2007)
  - Uses allele frequencies estimated from the sample
  - Limitations: Biased relatedness estimates in samples with population structure

- KING-robust estimator (Manichaikul et al., 2010)
  - Estimator assumes discrete population structure (no admixture)
  - Limitations: Biased estimates in admixed populations

- REAP (Thornton et al. , *AJHG* 2012) and PC-Relate (Conomos et al., *AJHG* 2016)
  - Allows for admixture
  - Uses admixture proportions or PCS to calculate ancestry/population-specific allele frequencies
  - Limitations:
    - Requires reliable estimates of admixture proportions and ancestry/population-specific allele frequencies
    - Biased results if (1) admixture portions are mispecified or (2) PCs not fully capturing ancestry
    - **SCALABILITY ISSUES WITH LARGE-BIOBANK STUDIES!**

Relatedness Estimation: IBD SEGMENT DETECTION

- IBD segments inferred based on long segments of identical-by-state allele IBS sharing
  - methods call segments IBD based on pairs of individuals sharing many mega-bases of alleles IBS
  - Methods available for phased or unphased genetic data:
    - KING: unphased data with ibdseg
    - TRUFFLE: unphased data (Dimitromanolakis et al., 2019)
    - hap-IBD: phased data (Browning et al., 2021)
    - iLASH: phased data (Shemirani et al., 2021)
    - RaPID: phased genotype data (Naseri et al., 2019)



IBD detection performance in large-scale ancestrally diverse samples?

Relatedness pipeline architecture: Automated pipeline

DNAnexus applets

-Staples et al (AJHG, 2014): PRIMUS for reconstructing pedigrees, get nuclear families
-

Step1

QC: Filter inputs

- Missingness: 0.05 variant missingness threshold
- Heterozygosity: HWE 1E-20 (PLINK: keep few het)

- PLINK commands

Step 2

KING: Run IBD segmentation

Priority:

- Retain as much of sample as possible
- Need high quality variants for reliable IBD segment breakpoints
- KING allows parallelization by splitting data into K chunks; K included as input in the applet

- KING commands
- Data processing
- PLINK commands

Step 3

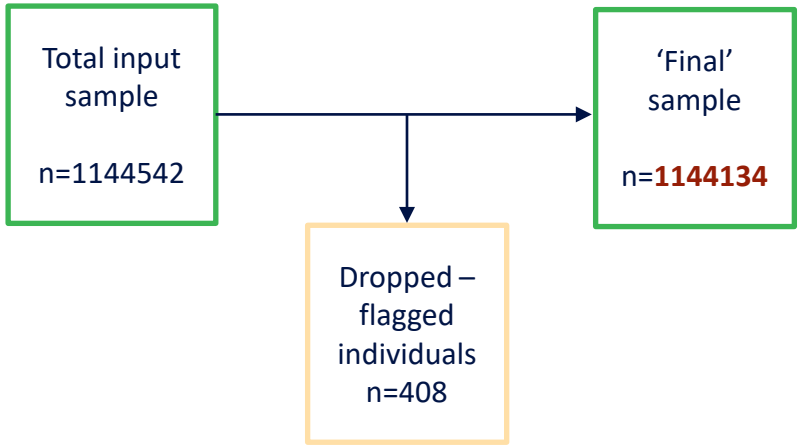
PRIMUS: Run pedigree reconstruction

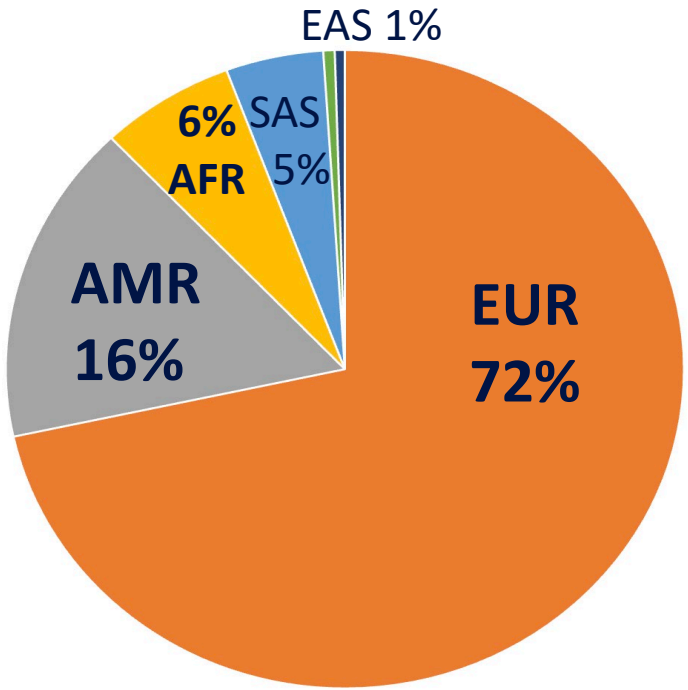
- Data processing
- PRIMUS commands

RGC

Scalable Relatedness Pipeline: Application to 1+ Million Diverse Samples

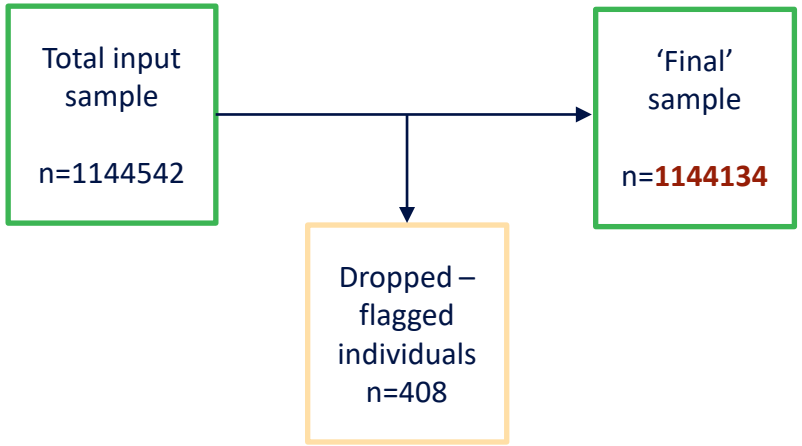
- Study of 46 cohorts with genome-wide data at RGC
- 1,144,542 individuals with shared variant set
- Relationships detected using pipeline





Ancestry distribution for 1+ Million samples at RGC

Scalable Relatedness Pipeline: Application to 1+ Million Samples



IBD segments and relatedness inferred for more than 654 billion pairs of individuals.

Pipeline completed in less than a day!

Sample size	Dup/MZ	Parent offspring	Full sibling	2nd	3rd
1144134	3178	137900	120217	325032	7398108

Important feature: No re-estimation of existing samples needed with future data freezes!

Computational pipelines for analysis of large bio-bank scale samples from diverse populations

Whole Genome Regression For Complex Trait Mapping

Computationally efficient WGR

nature genetics

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature genetics › technical reports › article

Technical Report | Published: 20 May 2021

Computationally efficient whole-genome regression for quantitative and binary traits

Joelle Mbatchou, Leland Barnard, Joshua Backman, Anthony Marcketta, Jack A. Kosmicki, Andrey Ziyatdinov, Christian Benner, Colm O'Dushlaine, Mathew Barber, Boris Boutkov, Lukas Habegger, Manuel Ferreira, Aris Baras, Jeffrey Reid, Goncalo Abecasis, Evan Maxwell & Jonathan Marchini

Nature Genetics **53**, 1097–1103 (2021) | Cite this article

14k Accesses | **27** Citations | **38** Altmetric | Metrics

REGENIE

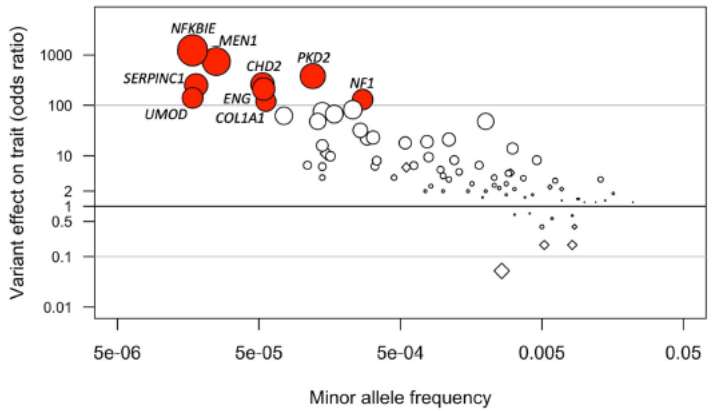
- Works on both quantitative and binary
  - Correction using penalized Firth/SPA for highly imbalanced binary traits
- Controls for population structure & relatedness through WGR framework
- Can process multiple phenotypes
- Decoupled WGR & association testing step
- Extended to gene-based testing
- Publicly available in C++ software on Github
- Apache Spark based implementation http://projectglow.io/

Exome sequencing and analysis of 454,787 UK Biobank participants

https://doi.org/10.1038/s41586-021-04103-z
Received: 9 July 2021
Accepted: 6 October 2021
Published online: 18 October 2021
Open access
Check for updates

Joshua D. Backman[1], Alexander H. Li[1], Anthony Marcketta[1], Dylan Sun[1], Joelle Mbatchou[1], Michael D. Kessler[1], Christian Benner[1], Daren Liu[1], Adam E. Locke[1], Suganthi Balasubramanian[1], Ashish Yadav[1], Nilanjana Banerjee[1], Christopher E. Gillies[1], Amy Damask[1], Simon Liu[1], Xiaodong Bai[1], Alicia Hawes[1], Evan Maxwell[1], Lauren Gurski[1], Kyoko Watanabe[1], Jack A. Kosmicki[1], Veera Rajagopal[1], Jason Mighty[1], Regeneron Genetics Center*, DiscovEHR*, Marcus Jones[1], Lyndon Mitnaul[1], Eli Stahl[1], Giovanni Coppola[1], Eric Jorgenson[1], Lukas Habegger[1], William J. Salerno[1], Alan R. Shuldiner[1], Luca A. Lotta[1], John D. Overton[1], Michael N. Cantor[1], Jeffrey G. Reid[1], George Yancopoulos[1], Hyun M. Kang[1], Jonathan Marchini[1,2], Aris Baras[1,2], Gonçalo R. Abecasis[1,2]✉ & Manuel A. R. Ferreira[1,2]✉

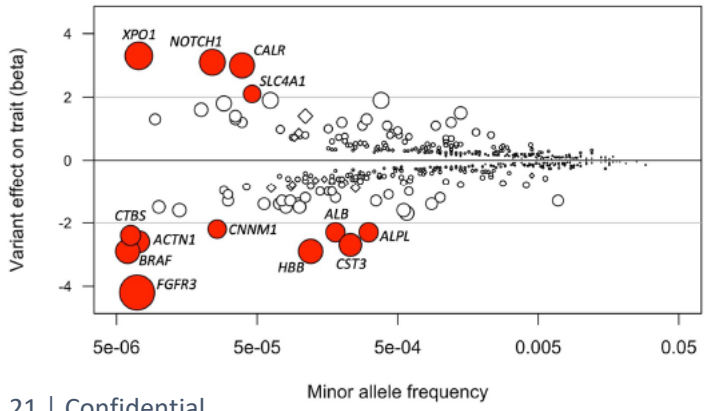
a **80 genes for which the lead association was with a binary trait**



Genes with an odds ratio >100:

Gene	Most associated binary trait
CHD2	Chronic lymphocytic leukemia of B-cell type
COL1A1	Bone disorder
ENG	Hereditary hemorrhagic telangiectasia
MEN1	Hyperparathyroidism
NF1	Benign neoplasm of peripheral nerves
NFKBIE	Chronic lymphocytic leukemia of B-cell type
PKD2	Cystic kidney disease
SERPINC1	Coagulation defects
UMOD	Chronic kidney disease

b **484 genes for which the lead association was with a quantitative trait**



Genes with leffectl >2:

Gene	Most associated quantitative trait
ACTN1	Platelet count
ALB	Albumin
ALPL	Alkaline phosphatase
BRAF	Neutrophil count
CALR	Platelet count
CNNM1	Aspartate aminotransferase
CST3	Cystatin C
CTBS	Peak expiratory flow
FGFR3	Height
HBB	Mean corpuscular volume
NOTCH1	Lymphocyte count
SLC4A1	Reticulocyte percentage
XPO1	Lymphocyte count

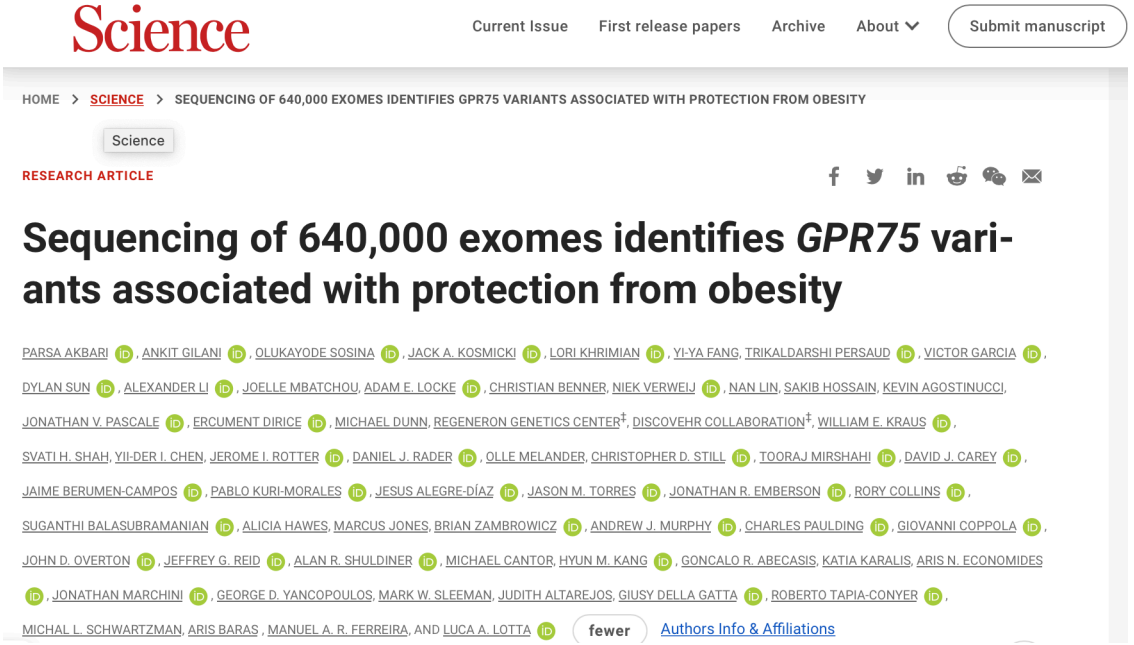
Table 1 | Number of coding variants discovered in exome sequencing data from 454,787 participants in the UK Biobank

Variant category	No. of variants (% with MAC=1)	Median number of variants per participant (IQR)
Coding regions[a]	12,326,144 (46.86)	19,895 (247)
Predicted function		
In-frame indels	75,096 (40.33)	115 (11)
Synonymous	3,457,173 (43.12)	10,273 (141)
Missense	7,878,586 (47.28)	9,292 (143)
Likely benign	1,532,129 (44.11)	6,561 (104)
Possibly deleterious	4,556,629 (47.23)	2,610 (70)
Likely deleterious	1,789,828 (50.1)	121 (16)
pLOF (any transcript)	915,289 (57.88)	214 (16)
Start lost	26,453 (47.94)	13 (4)
Stop gained	279,913 (54.02)	52 (8)
Stop lost	12,843 (56.51)	6 (3)
Splice donor	104,328 (58.67)	17 (5)
Frameshift	405,669 (60.41)	90 (10)
Splice acceptor	86,083 (60.79)	20 (5)

[a]Includes all coding variants: synonymous, in-frame indels, missense and pLOF variants.
MAC, minor allele count; IQR, interquartile range.

- Performed exome sequencing of 454,787 participants
- Identified ~12M coding variants, ~1M putative loss-of-function variants and ~ 1.8M deleterious missense variants
- Tested association with 3,994 health-related traits & found 564 genes with trait associations at $P \leq 2.18 \times 10\text{-}11$
- Rare variant associations were enriched in loci from genome-wide association studies, but 91% (most) were independent of common variant signals

Diverse RGC cohorts: Novel genetic discoveries and therapeutic targets


Science

Current Issue First release papers Archive About Submit manuscript

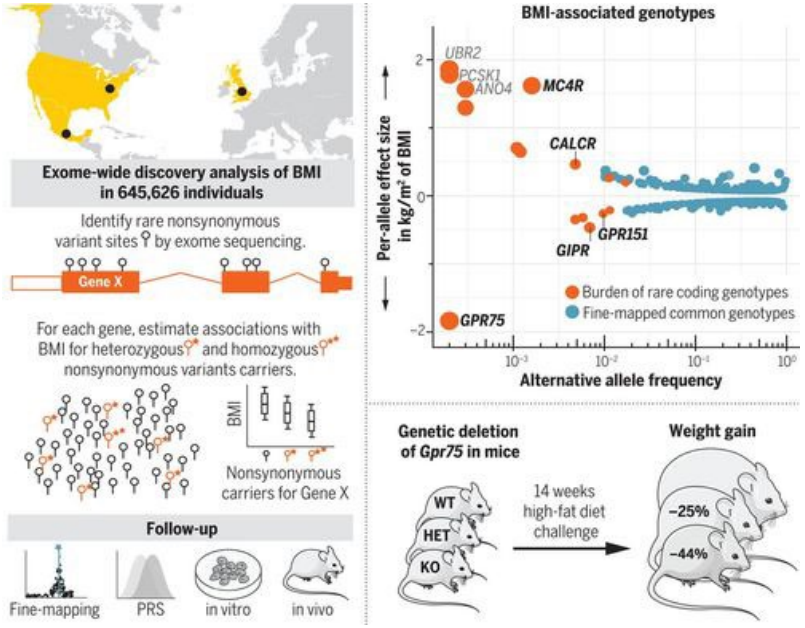
HOME > SCIENCE > SEQUENCING OF 640,000 EXOMES IDENTIFIES GPR75 VARIANTS ASSOCIATED WITH PROTECTION FROM OBESITY

Science

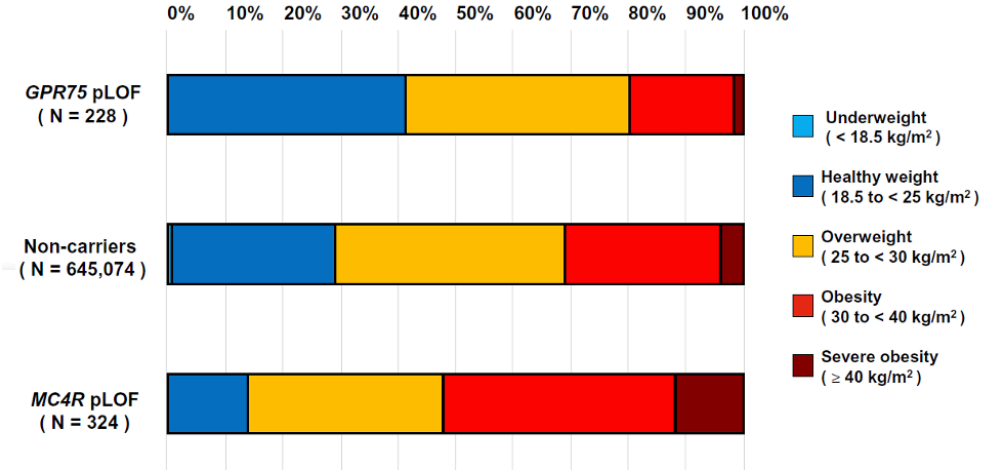
RESEARCH ARTICLE

Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection from obesity

PARSA AKBARI, ANKIT GILANI, OLUKAYODE SOSINA, JACK A. KOSMICKI, LORI KHRIMIAN, YI-YA FANG, TRIKALDARSHI PERSAUD, VICTOR GARCIA, DYLAN SUN, ALEXANDER LI, JOELLE MBATCHOU, ADAM E. LOCKE, CHRISTIAN BENNER, NIEK VERWEIJ, NAN LIN, SAKIB HOSSAIN, KEVIN AGOSTINUCCI, JONATHAN V. PASCALE, ERCUMENT DIRICE, MICHAEL DUNN, REGENERON GENETICS CENTER‡, DISCOVEHR COLLABORATION‡, WILLIAM E. KRAUS, SVATI H. SHAH, YII-DER I. CHEN, JEROME I. ROTTER, DANIEL J. RADER, OLLE MELANDER, CHRISTOPHER D. STILL, TOORAJ MIRSHAHI, DAVID J. CAREY, JAIME BERUMEN-CAMPOS, PABLO KURI-MORALES, JESUS ALEGRE-DÍAZ, JASON M. TORRES, JONATHAN R. EMBERSON, RORY COLLINS, SUGANTHI BALASUBRAMANIAN, ALICIA HAWES, MARCUS JONES, BRIAN ZAMBROWICZ, ANDREW J. MURPHY, CHARLES PAULDING, GIOVANNI COPPOLA, JOHN D. OVERTON, JEFFREY G. REID, ALAN R. SHULDINER, MICHAEL CANTOR, HYUN M. KANG, GONCALO R. ABECASIS, KATIA KARALIS, ARIS N. ECONOMIDES, JONATHAN MARCHINI, GEORGE D. YANCOPOULOS, MARK W. SLEEMAN, JUDITH ALTAREJOS, GIUSY DELLA GATTA, ROBERTO TAPIA-CONYER, MICHAL L. SCHWARTZMAN, ARIS BARAS, MANUEL A. R. FERREIRA, AND LUCA A. LOTTA fewer Authors Info & Affiliations



- Rare predicted loss of function coding variants in *GPR75* for heterozygous carriers found to be associated with
  - Lower BMI (-1.8 kg/m^2)
  - Lower body weight (~5.3 kg or 11.7 lbs lower)
  - Protection against obesity (54% lower odds)
- GPR75 knock-out mice show resistance to weight gain in high-fat diet challenge, as well as healthier insulin and fasting glucose profiles



Diverse RGC cohorts: Novel genetic discoveries and therapeutic targets

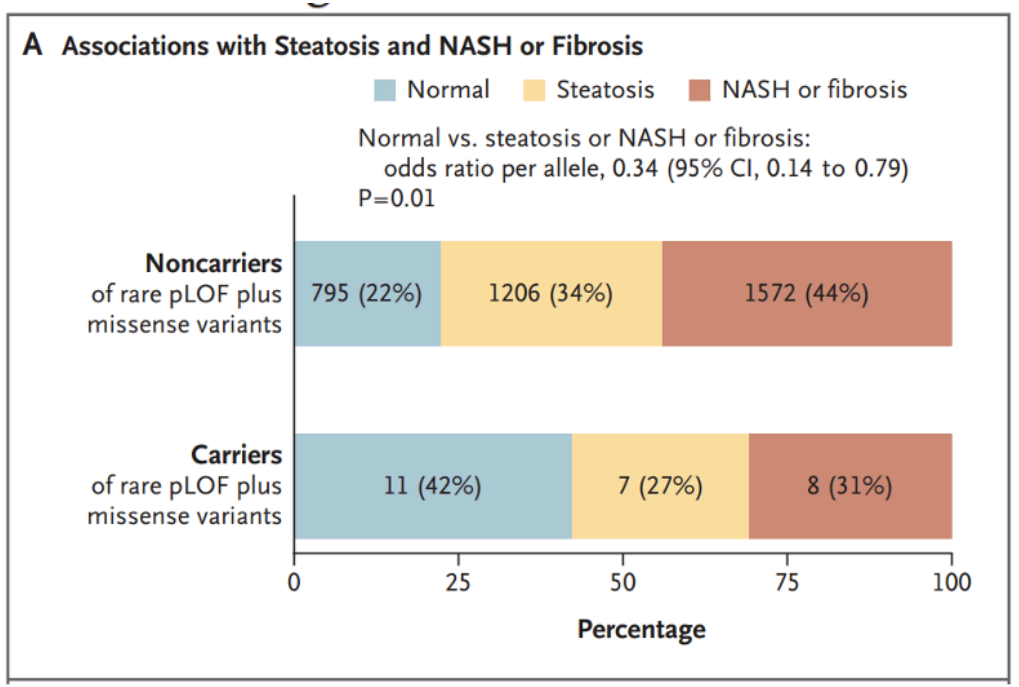
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Germline Mutations in *CIDEB* and Protection against Liver Disease

N. Verweij, M.E. Haas, J.B. Nielsen, O.A. Sosina, M. Kim, P. Akbari, T. De, G. Hindy, J. Bovijn, T. Persaud, L. Miloscio, M. Germino, L. Panagis, K. Watanabe, J. Mbatchou, M. Jones, M. LeBlanc, S. Balasubramanian, C. Lammert, S. Enhörning, O. Melander, D.J. Carey, C.D. Still, T. Mirshahi, D.J. Rader, P. Parasoglou, J.R. Walls, J.D. Overton, J.G. Reid, A. Economides, M.N. Cantor, B. Zambrowicz, A.J. Murphy, G.R. Abecasis, M.A.R. Ferreira, E. Smagris, V. Gusarova, M. Sleeman, G.D. Yancopoulos, J. Marchini, H.M. Kang, K. Karalis, A.R. Shuldiner, G. Della Gatta, A.E. Locke, A. Baras, and L.A. Lotta

Verweij et al. (2022) N Engl J Med 387:332-344



Rare predicted loss-of-function variants plus missense variants in CIDEB associated with 33% lower odds of liver disease of any cause

Leveraging Diverse and Admixed Genomes

The Mexico City Prospective Study

The Mexico City Prospective Study (MCPS)

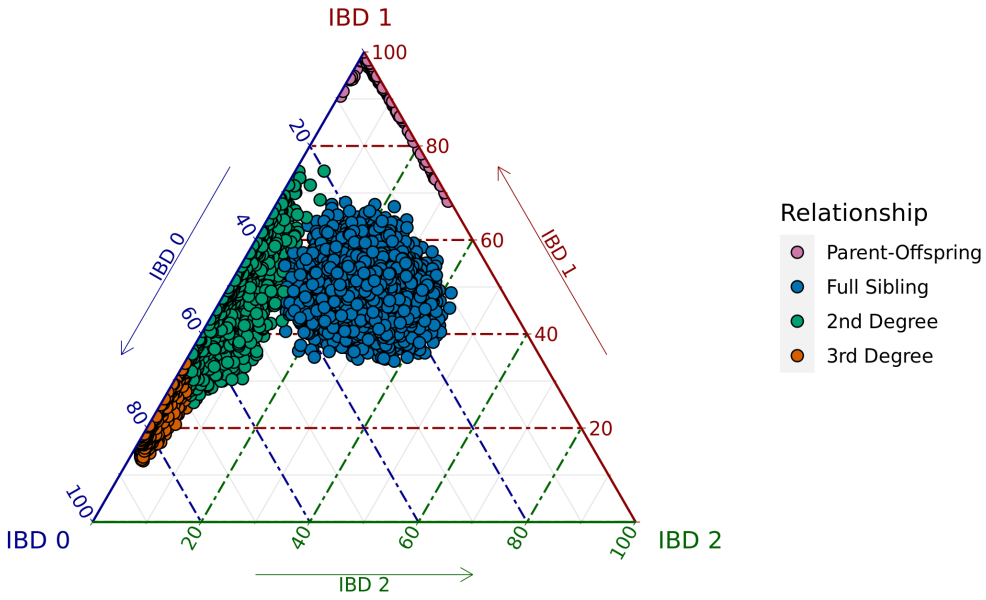
- Founded by epidemiologists from Mexico City and Oxford
- 159,755 adults enrolled by visiting 112,333 family households within two urban districts in 1998-2004

- Health questionnaires, physical measurements, blood, etc.
- Resurvey of ~10,000 participants in 2015-2019
- Linkage to mortality data ongoing

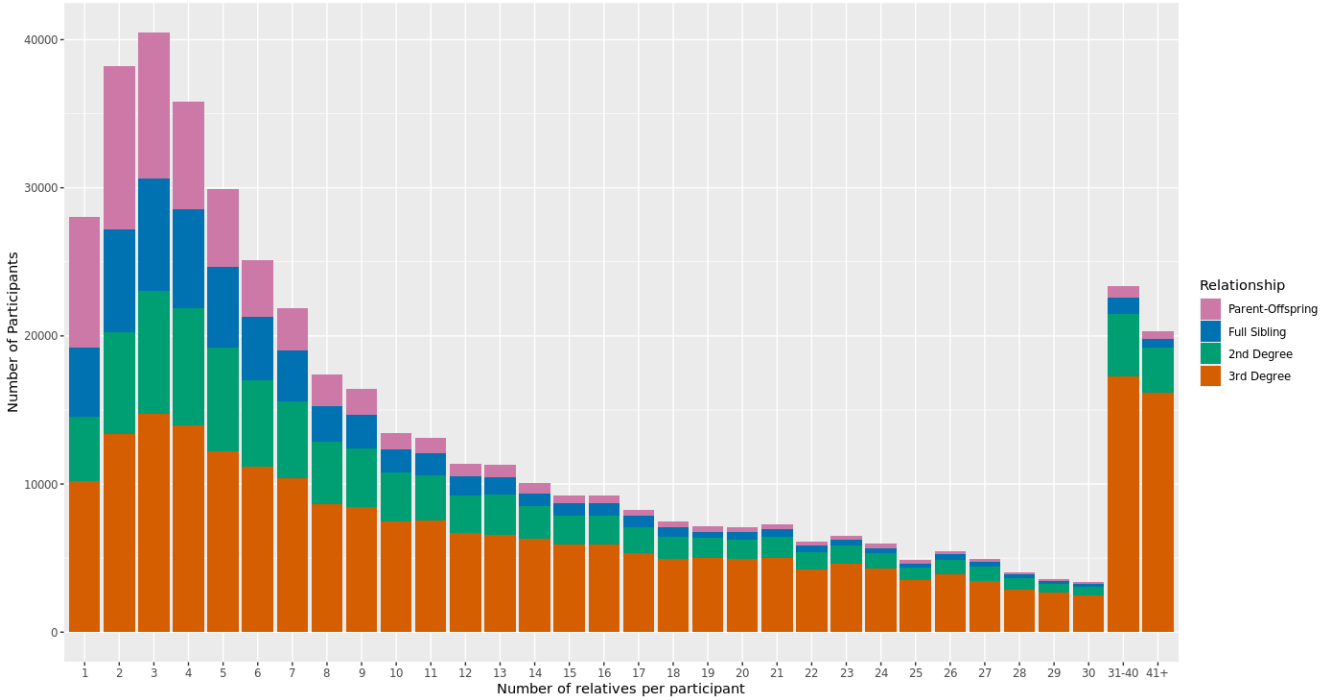
IBD-Based Relatedness Pipeline: MCPS

IRobust to admixture and scalable to large-scale bio-bank samples.

Proportion of the genome estimated to have 0, 1 or 2 alleles identical-by-descent (IBD)

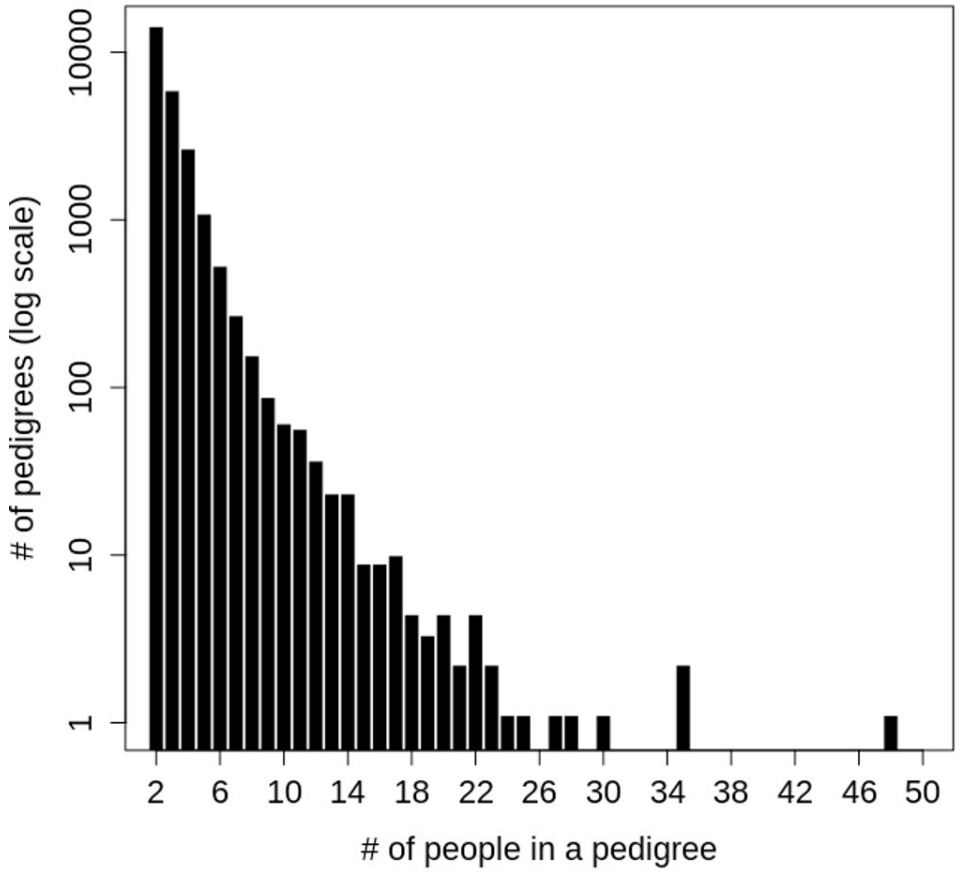


Distribution of the number of relatives per participant



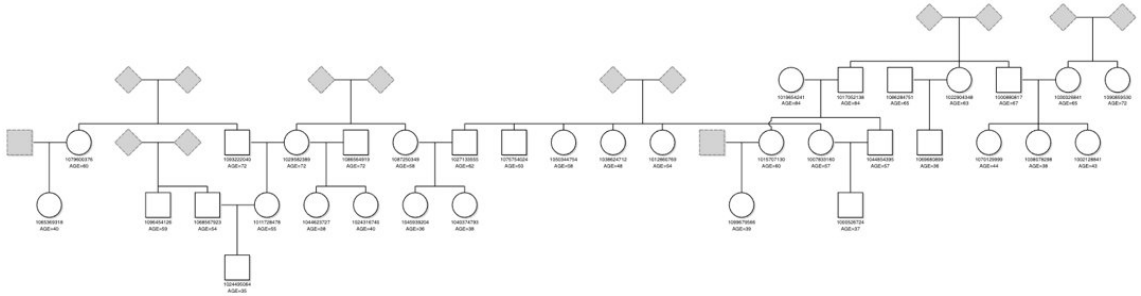
Parent-Offspring	Full-Siblings	Second-Degree	Third-Degree
31,597	29,482	47,080	120,180

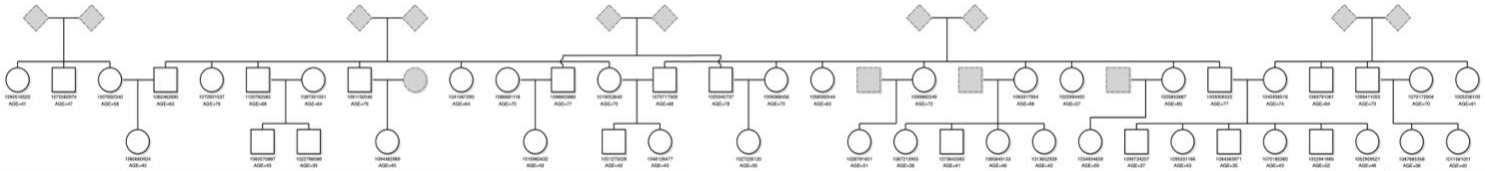
Pedigree Reconstruction in MCPS with PRIMUS



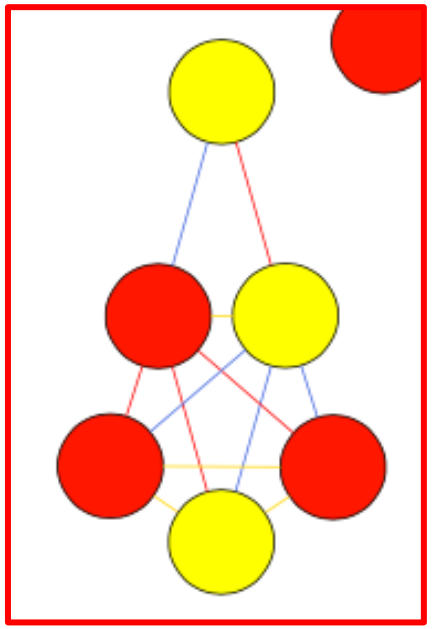
99.3% of 1st degree pedigrees reconstructed unambiguously with PRIMUS

3,595 nuclear families
- 2,268 trios
- 869 quartets
- 308 quintets
- 100 sextets
- 34 septet
- 11 octets
- 3 nonets
- 2 decets





MCPS 2nd degree network of families > 4 in size



Nodes are individuals

Red = Iztapalapa

Yellow = Coyoacan

Edges are relationships

Red = Parent/child

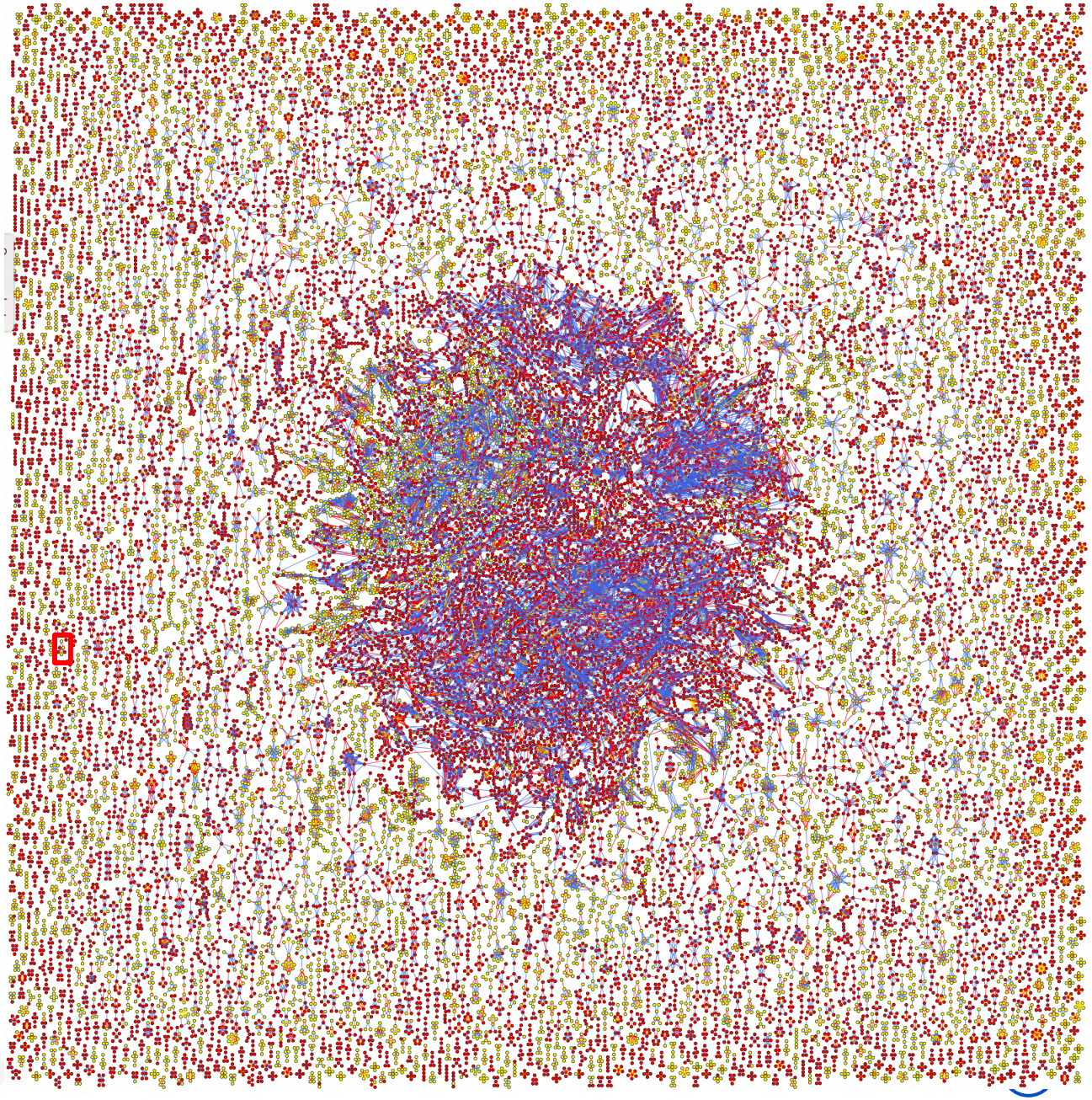
Gold = Full-sibling

Blue = 2nd degree

Render by Graphviz's sfdp layout engine

•Uses a "spring" model relying on a force-directed approach to minimize edge length

•Relatives are plotted closer to each other



MCPS is the largest non-European ancestry sequencing study

Collaboration between the National Autonomous University of Mexico, the National Institute of Genomic Medicine in Mexico, Oxford Population Health, Regeneron, Astra Zeneca and Abbvie.

	Genotyping Array	Whole Exome Sequencing (WES)	Whole Genome Sequencing (WGS)
Samples	140,829	141,046	9,950
All variants	0.56M	4.0M†	131.9M
Unique variants		**1.4M†,***	**31.5M****

†coding only
*not found in UK Biobank, TOPMed & gnomAD
**not found in TOPMed & gnomAD

Two resources derived from MCPS sequencing data

MCPS allele frequency browser https://rgc-mcps.regeneron.com/

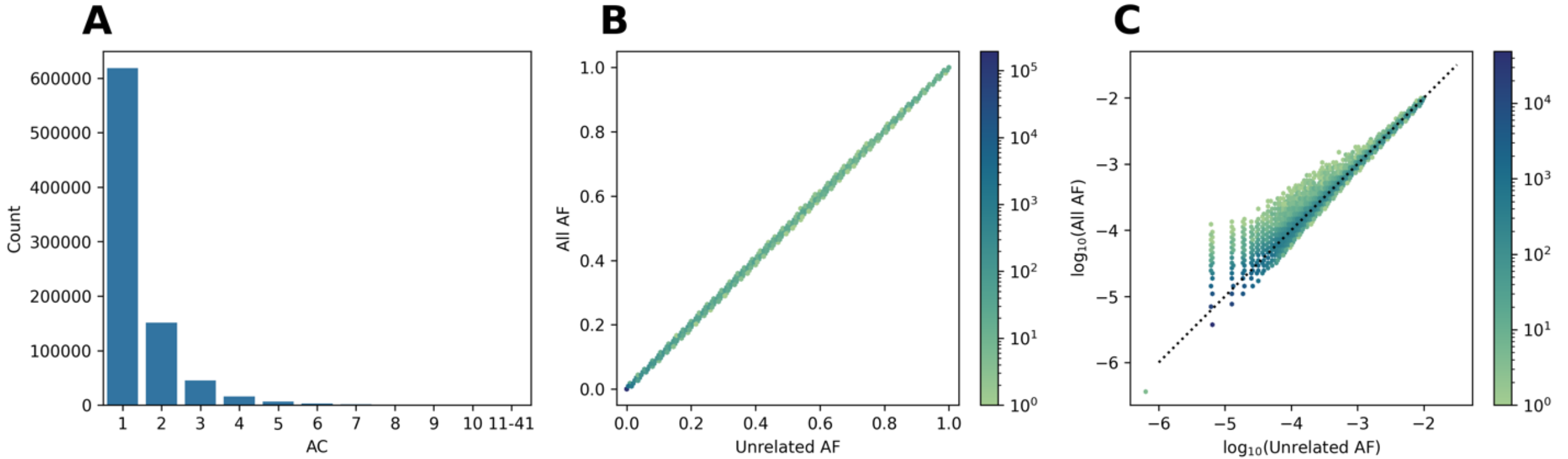
- 142 million variants in combined dataset of Array, WES and WGS
- Ancestry-specific allele frequencies from sequencing data

MCPS10K imputation reference panel

- Phasing of 9,950 WGS samples
- To be included in TOPMed reference panel

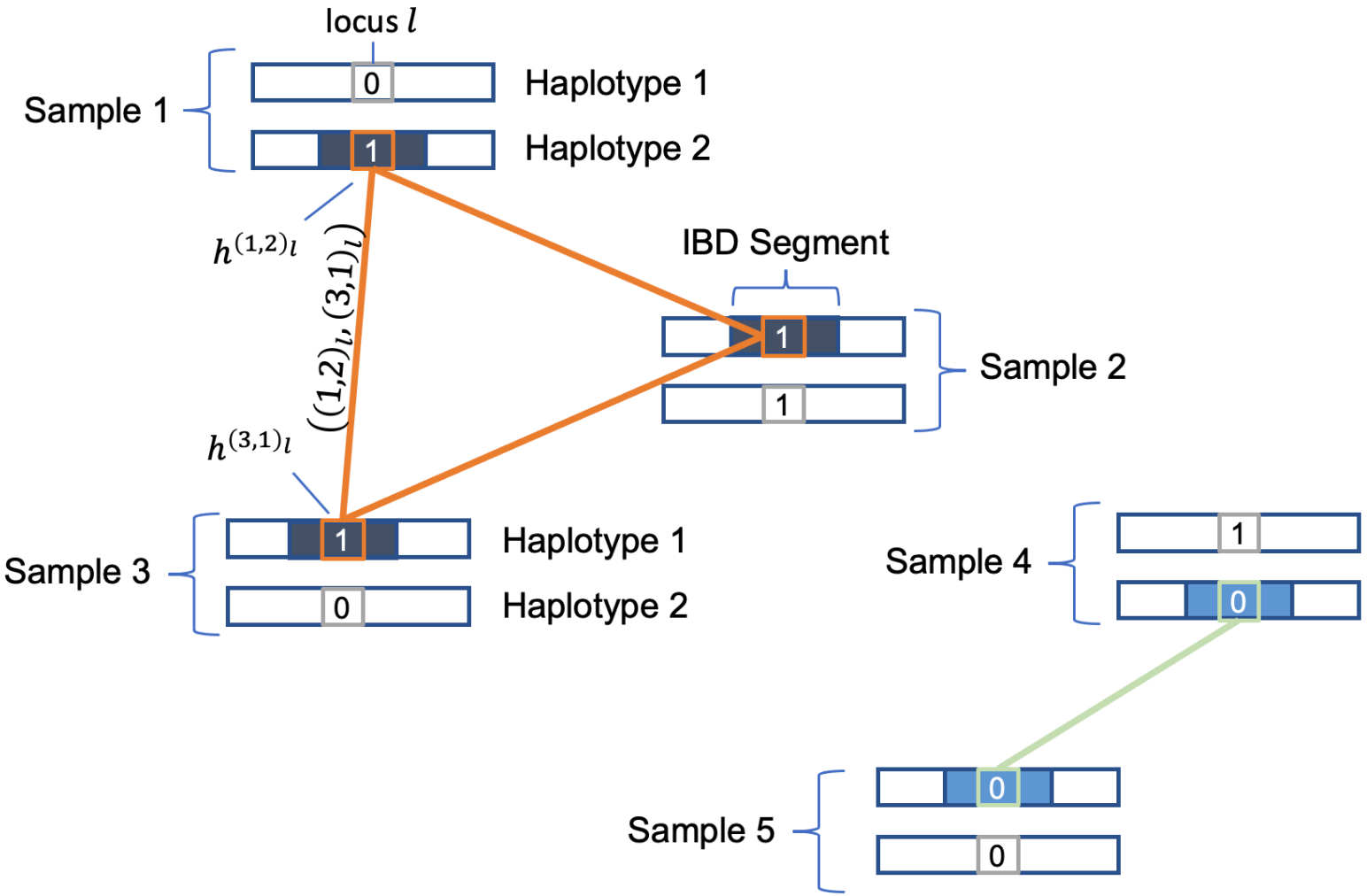
Schematic of ancestry specific allele frequencies

MCPS Allele Frequencies Browser: All Participants vs. Unrelated Subset?

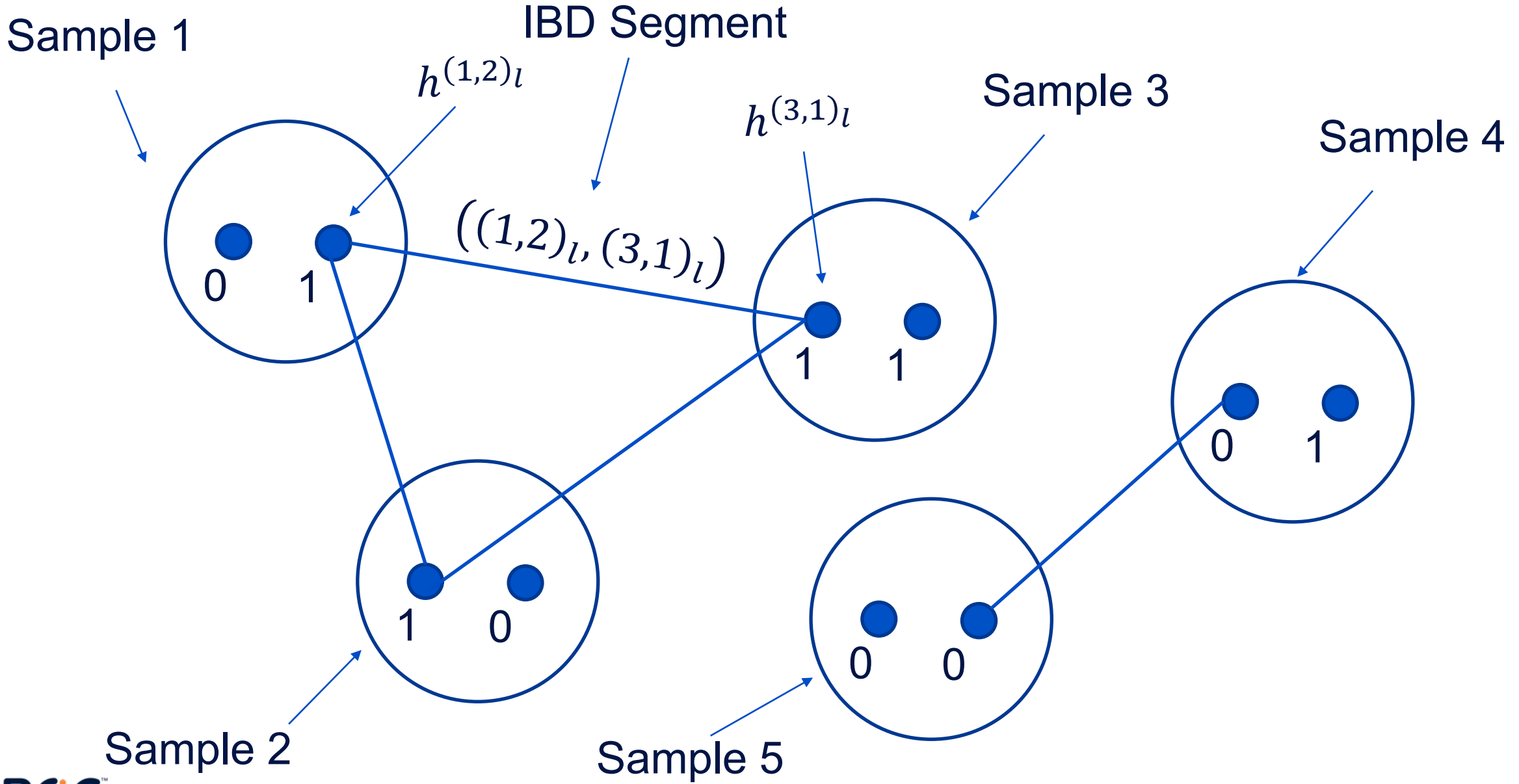


(A) Histogram of the alternate allele count of variants missing from the unrelated subset computed from the combined WES and array dataset for all chromosomes. **(B)** Hexbin plot of allele frequencies computed using the unrelated subset (*x*-axis*)* and all samples (*y*-axis) and for chromosome 22. **(C)** Hexbin plot of $\log_{10}$ allele frequencies of rare variants (AAF<0.01) on chromosome 22.

NEW APPROACH: Relatedness-Corrected Allele Frequencies Using IBD Segments

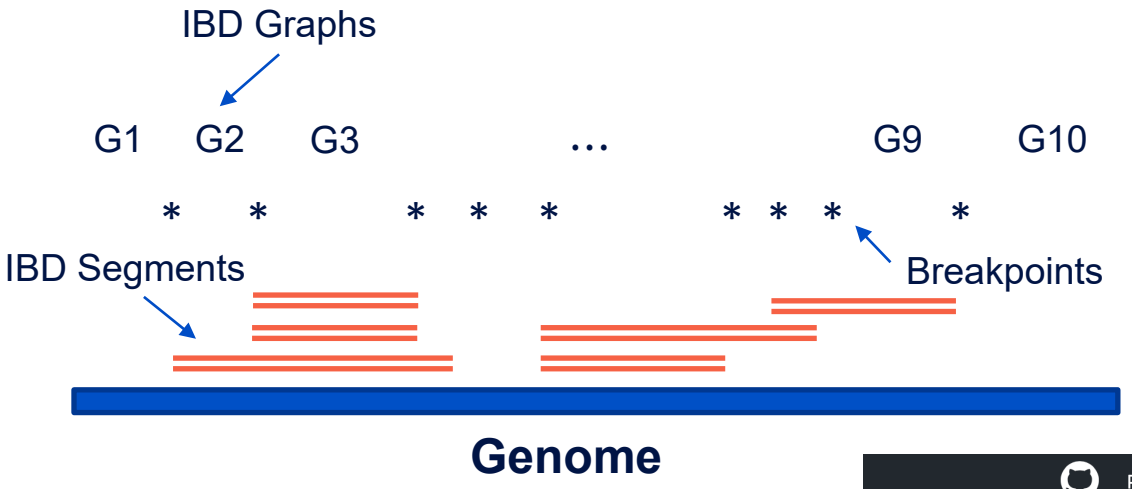


IBD Graph at Specific Genomic Location



NEW APPROACH: Relatedness-Corrected Allele Frequencies Using IBD Segments

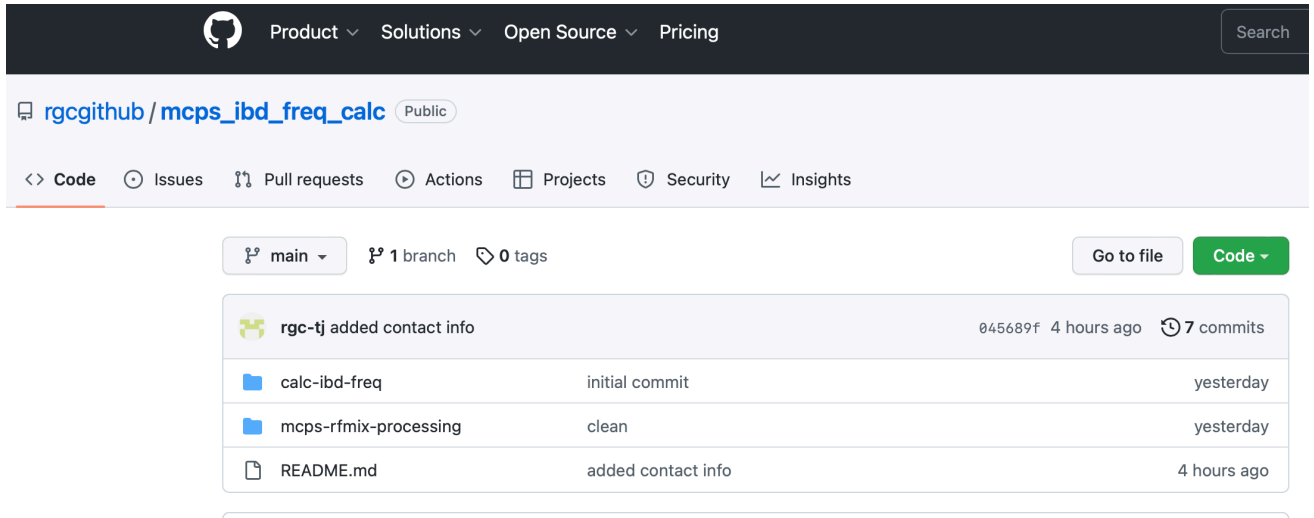
IBD segments define breakpoints where the relatedness structure changes along the genome

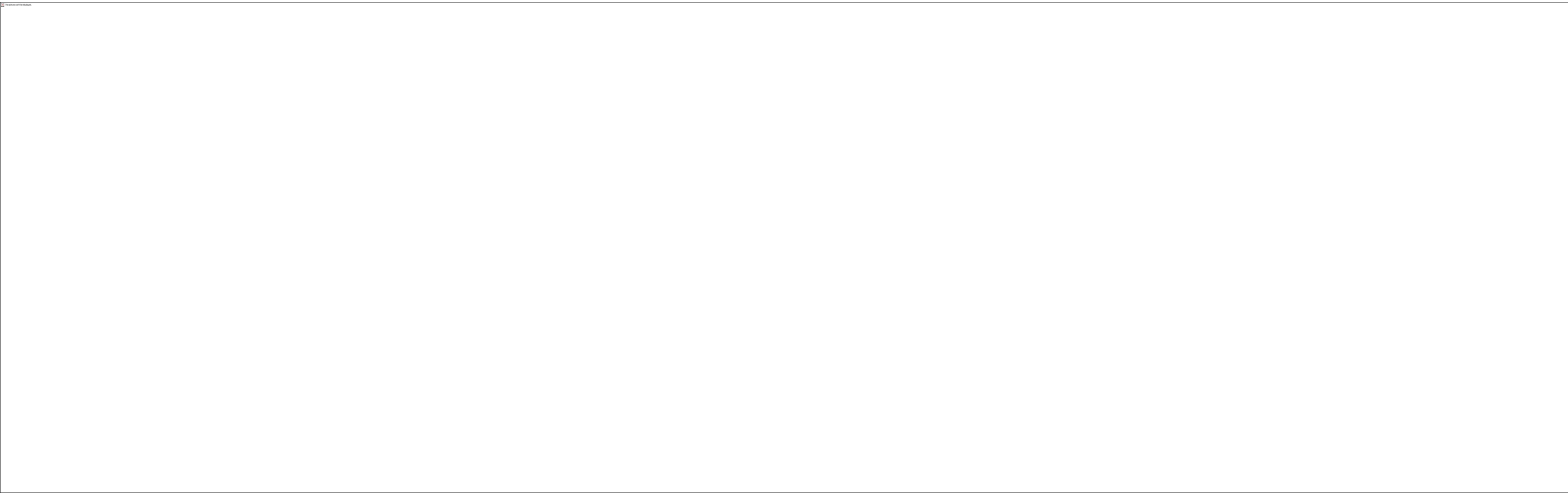


Algorithm

For each IBD graph along the genome:

1. Grab variants that fall in the span of the graph
2. Compute connected components of the graph
3. Set AC = # of alternate allele connected components
4. Set AN = total # of connected components





(A) Alternate allele frequencies computed for chromosome 22 correcting for IBD (*x*-axis) and computed from all samples (*y*-axis). **(B)** Log_{10} alternate allele frequencies for rare variants (AAF<0.01) computed for chromosome 22 correcting for IBD (*x*-axis) and computed from all samples (*y*-axis). **(C)** Log_{10} alternate allele frequencies for rare variants (AAF<0.01) computed for chromosome 22 correcting for IBD (*x*-axis) and computed from unrelated samples (*y*-axis).

>10-fold increase in size compared to gnomAD

Study	Source	# samples*	Effective # Indigenous samples	# variants
gnomAD[1]	WGS	7,612	4,610	14.8M**
MCPS	WGS	9,950	6,549	131.9M
MCPS	WES	141,046	91,856	9.3M

142M variants
→10-fold increase

→ 20-fold increase
in WES sample size

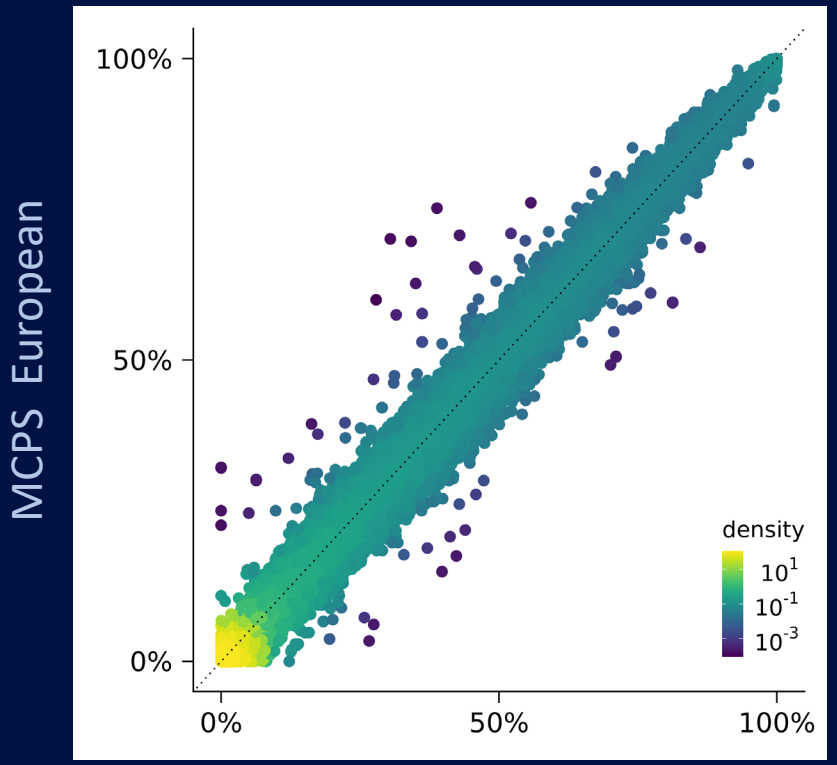
* Latino/Admixed American samples
**bi-allelic variants with low genotype missingness (≤10%) and an AF > 0.1%

[1]Wilson et al., AHGS 2021

MCPS WES AF estimates agree with gnomAD

European

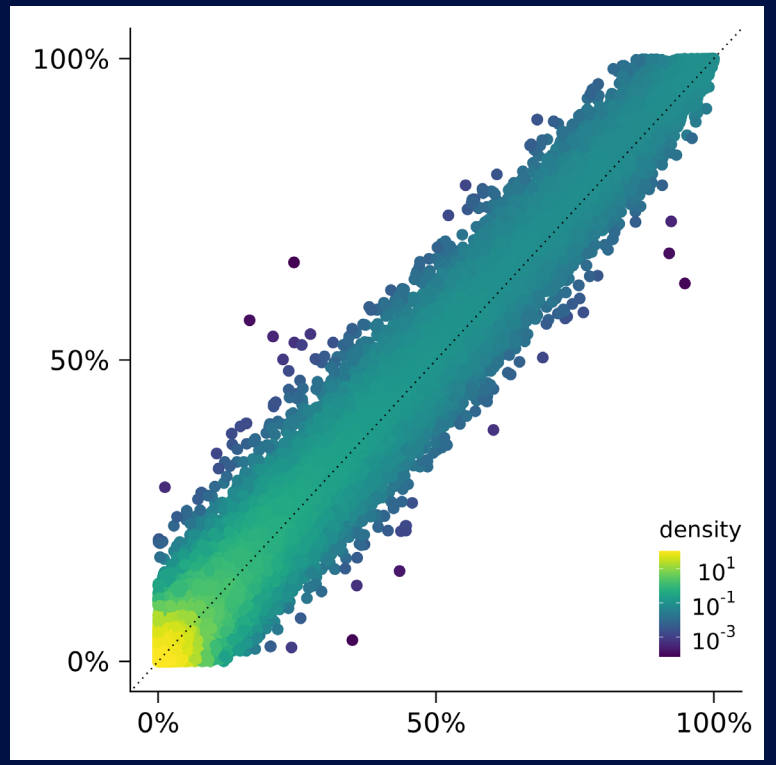
$r^2 = 0.99$



gnomAD Non-Finish European

African

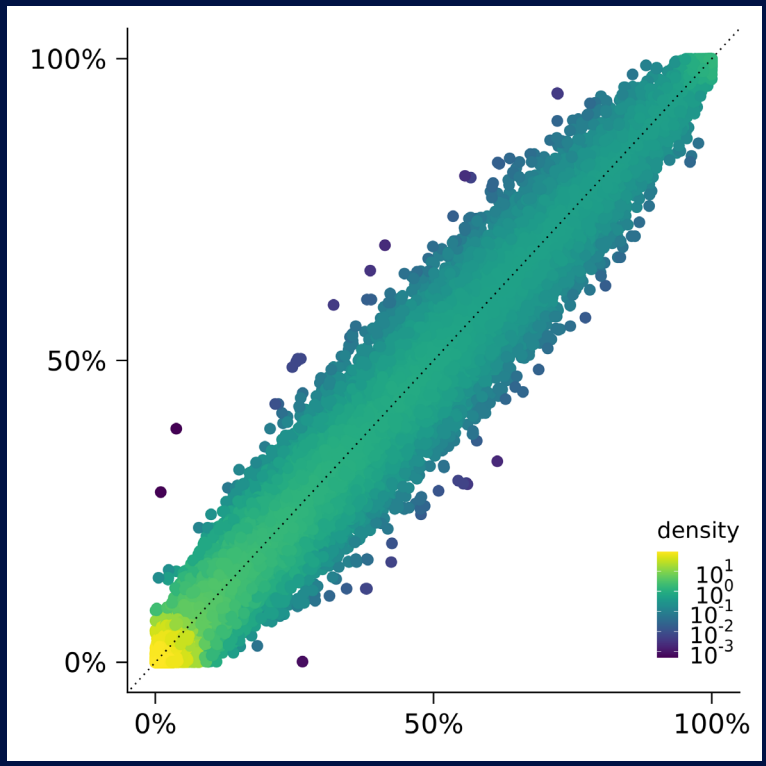
$r^2 = 0.98$



gnomAD African/African American

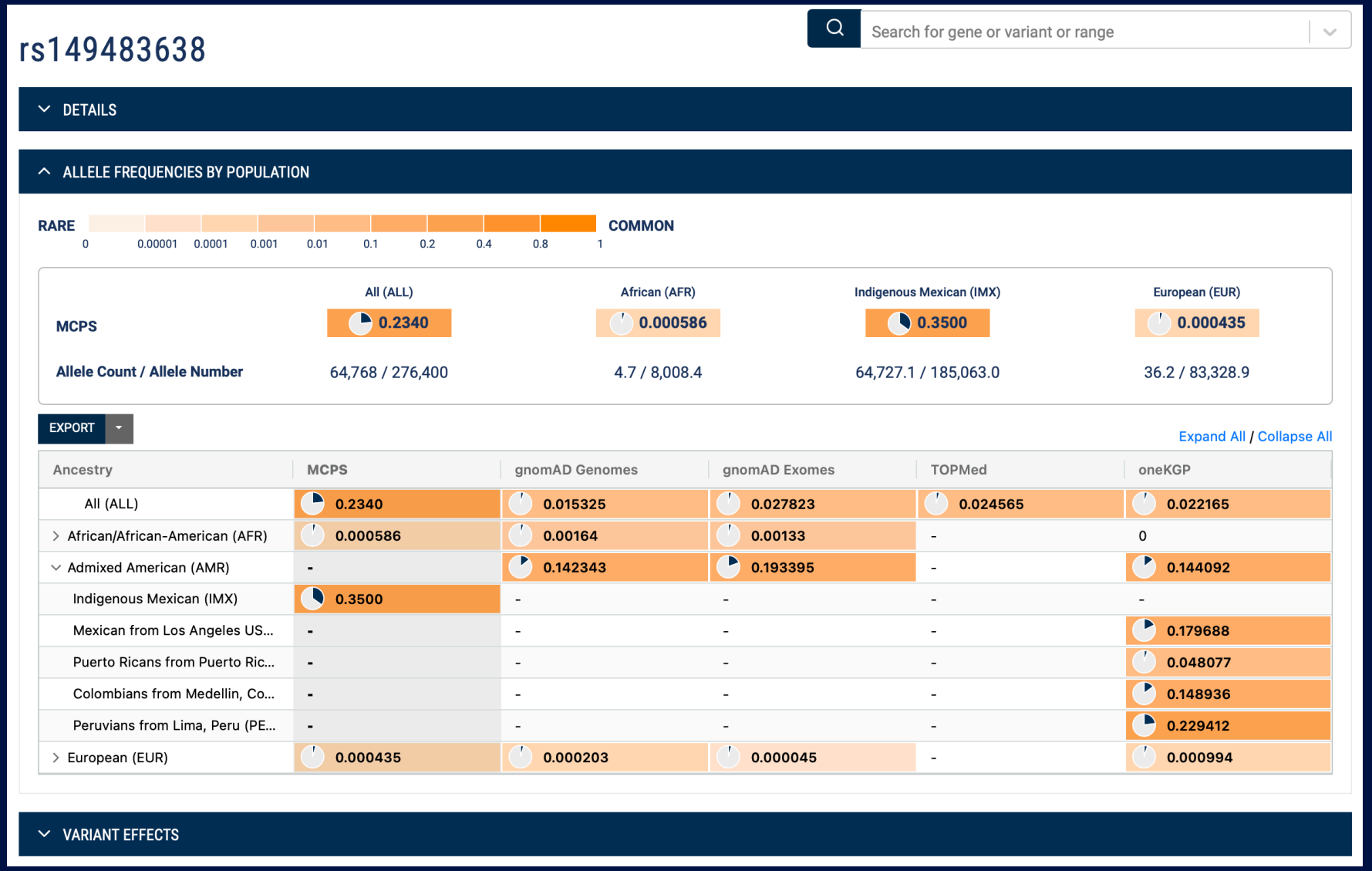
Indigenous

$r^2 = 0.99$



gnomAD Amerindigenous

MCPS Variant Browser https://rgc-mcps.regeneron.com/



Database of 142M variants

Raw VCF data files with allele frequencies are available for download

Acknowledgments

National Autonomous University of Mexico
Jesús Alegre-Díaz
Raul Ramirez-Reyes
Rogelio Santacruz-Benítez
Jaime Berumen
Pablo Kuri-Morales
Roberto Tapia-Conyer

Instituto Nacional de Medicina Genómica, Mexico City
Humberto García-Ortiz
Lorena Orozco-Orozco

Oxford
Jason Torres
Michael Turner
Rachel Wade
Rory Collins
Michael R Hill
Jonathan R Emberson

AstraZeneca
Abhishek Nag
Katherine Smith
Slavé Petrovski

Regeneron Genetics Center
Jonathan Marchini
Sheila M. Gaynor
Andrey Ziyatdinov
Tyler Joseph
Joshua Backman
Joelle Mbatchou
Yuxin Zou
Daren Liu
Jeffrey Staples
Razvan Panea
Alex Popov
Xiaodong Bai
Suganthi Balasubramanian
Lukas Habegger
Rouel Lanche
Alex Lopez
Evan Maxwell
Marcus Jones
Eric Jorgenson
Will Salerno
John Overton
Jeffrey Reid
Goncalo Abecasis
Lyndon Mitnaul
Aris Baras

AbbVie
Mark Reppell

University of Michigan
Sebastian Zollner

Regeneron Genetics Center

We are hiring! Visit our website for all posted positions: careers.regeneron.com

Open Position in Statistical Genetics: https://careers.regeneron.com/job/R20103/Associate-Manager-Statistical-Genetics

- Please contact Joelle Mbatchou for specific interest: joelle.mbatchou@regeneron.com

Associate Manager, Statistical Genetics

Tarrytown, New York, United States of America • Research and Development • R20103

☆ Save Apply Now ›

We are seeking a dedicated researcher to develop methods at the interface between machine learning and statistical genetics to provide a deeper understanding of human biology and aid in discovering novel therapeutic targets, enabling Regeneron to deliver novel medicines to patients in need. The role will include hands-on research and methods development, working across various applications involving large-scale genetic variation and electronic health record datasets at the Regeneron Genetics Center (RGC). You will design, implement, and refine methods and analyses to connect genetic variation to human health and disease.

In this Associate Manager role, a typical day might include the following:

- Develop and apply innovative methods at the interface between statistical genetics and MLand deploy them at scale to answer biological and disease genetics questions that cover all the areas of interest of RGC.

Get notified for similar jobs

Sign up to receive job alerts

your@email.com Submit

Get tailored job recommendations based on your interests.

THANK YOU!

Contact Information:

timothy.thornton@regeneron.com