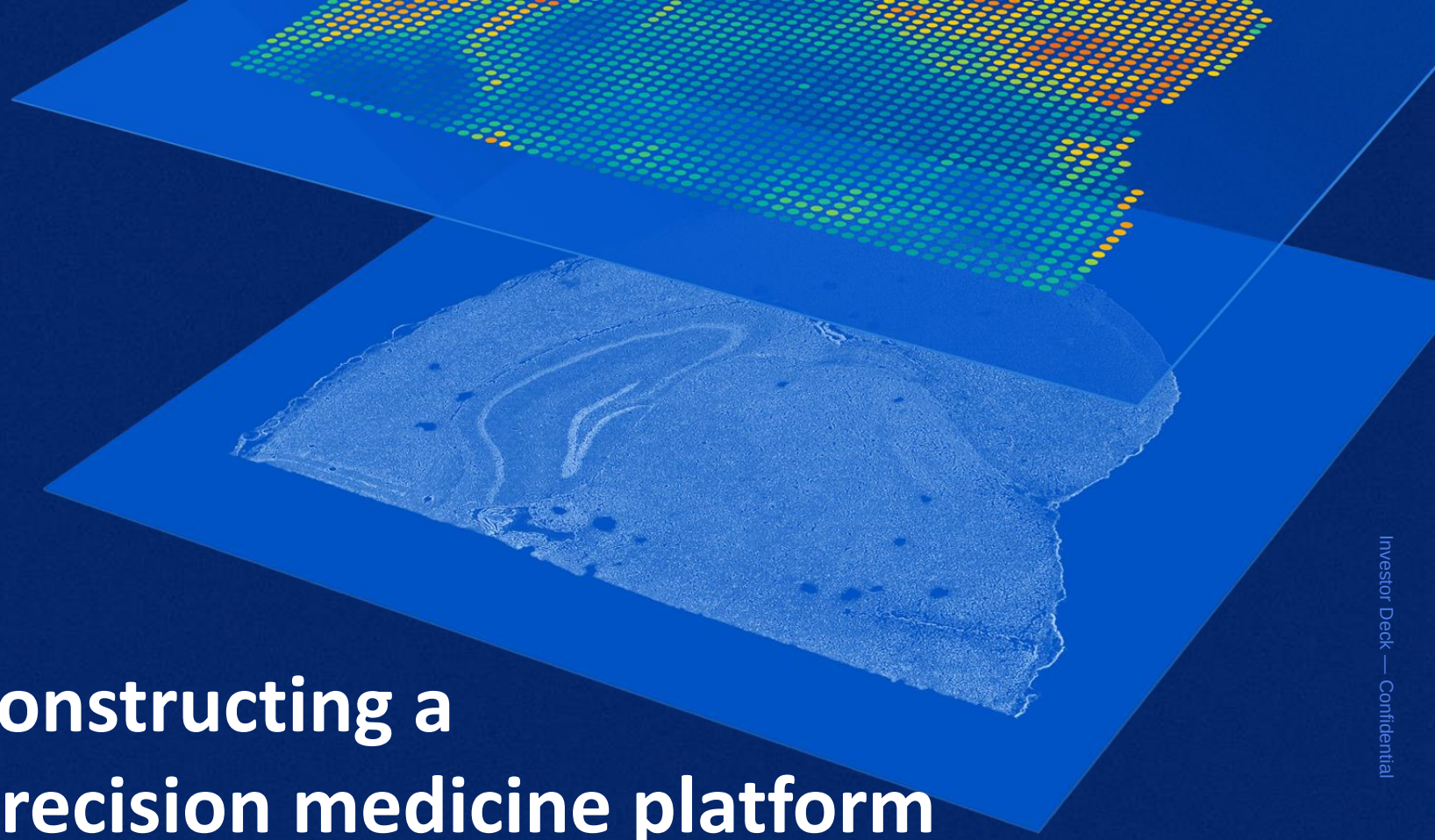
galatea.bio
powering the bio2bio revolution



Constructing a transnational precision medicine platform

Andy Larrea, PhD

Human genetics remains amongst the strongest predictors of success

How can we leverage “smart” genetics to improve biology and drug discovery?

Agenda

1. Building a diversity infrastructure – samples and software solutions.
2. Genotype/Phenotype based strategies
3. Drug discovery process and cohort development.

We are building a 10M sample biobank + data lake to accelerate precision health in LatAm

Underrepresented Groups

+

Electronic Health Records access w/ recontact for follow up

+

Genetics of Health & Disease

=

Superior Biological Insights for All

Case Study: UK Biobank Exome Sequencing Consortium

galatea.bio
powering the bio2bio revolution

Overview:

biobank uk

UK Biobank partnered with **8 pharmaceutical** companies to collect up to **500K samples and sequence genomes.**

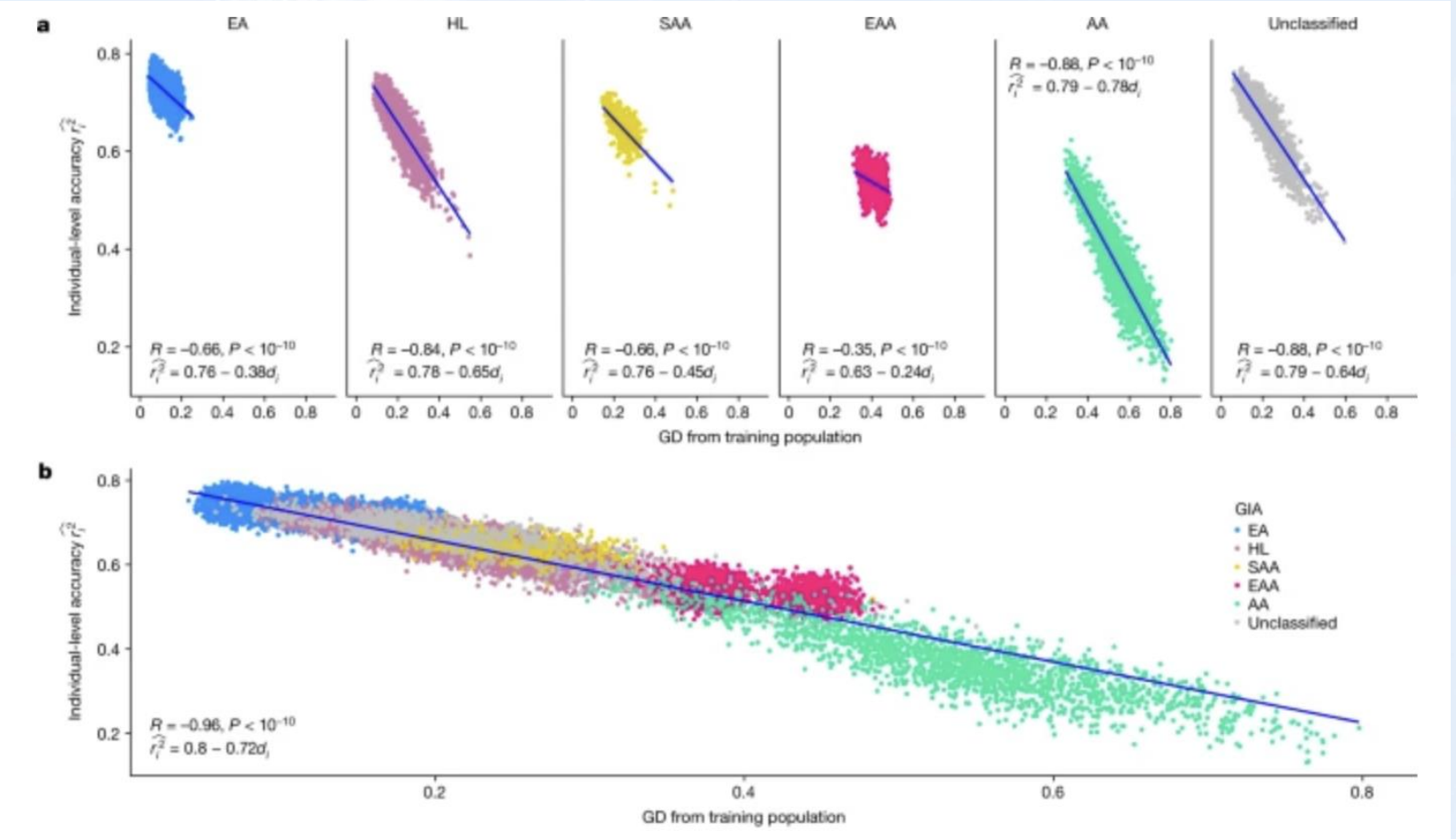
"The objective of the UKB-ESC is a comprehensive assessment of the protein-coding genetic variation in the **half-million UKB participants**"

Promising leads are helping identify the association between disease phenotypes, biomarker endo-phenotypes and functionally consequential genetic variants

Leads Discovered So Far	
Chronic Liver Disease	*HSD17B13*
Multiple Autoimmune Diseases	*TYK2*
Neuropsychiatric Disease	*NRXN1*
Cardiovascular Disease	*ASGR1*

Bottom Line: Large-scale sequencing partnerships are **driving transformative scientific discoveries**

Polygenic scoring accuracy varies across the genetic ancestry continuum

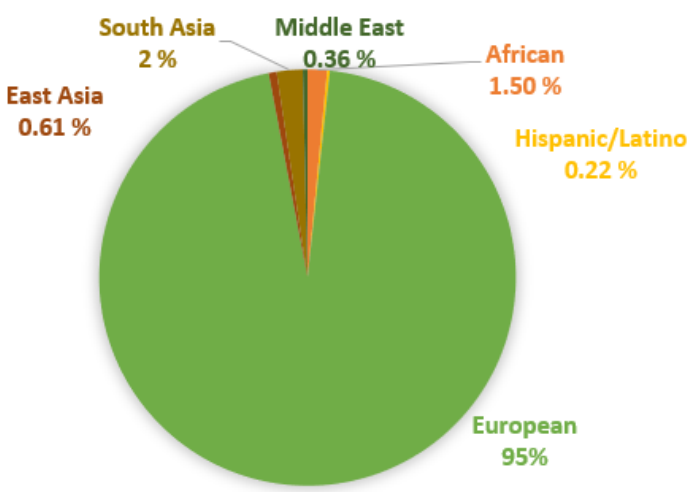


https://doi.org/10.1038/s41586-023-06079-4

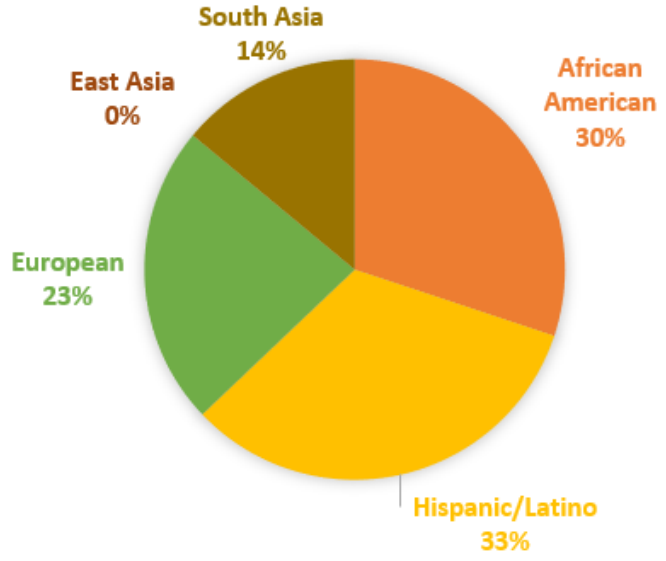
Early snapshot showed that we were moving in the right direction…

GalateaBio Biobank participants' ancestries are representative of broader Central/South American

UKBB

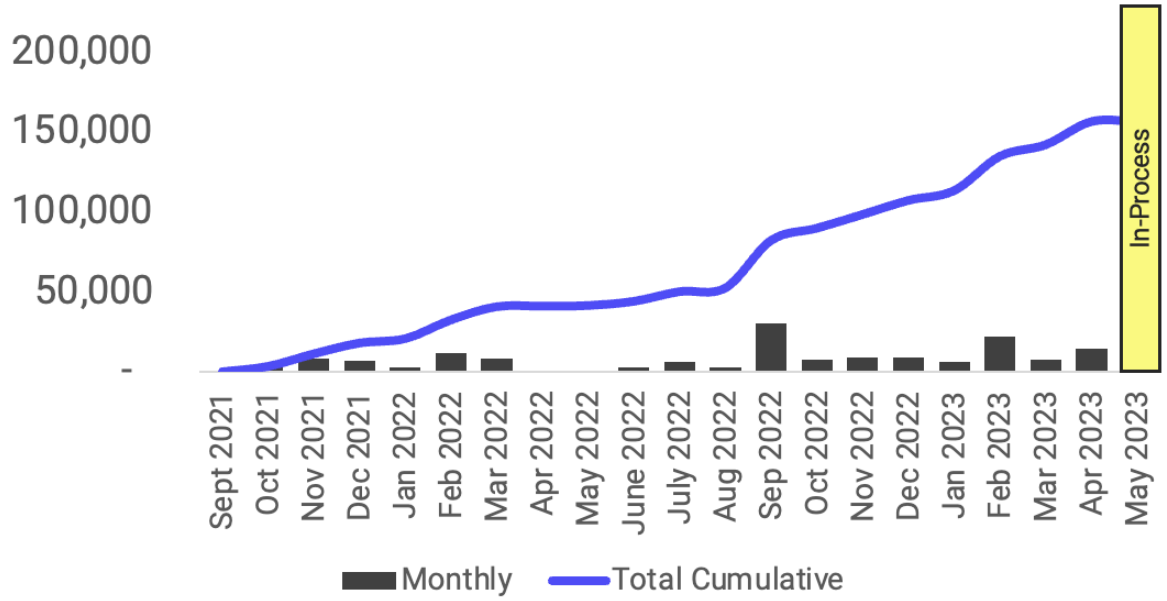


GalateaBio

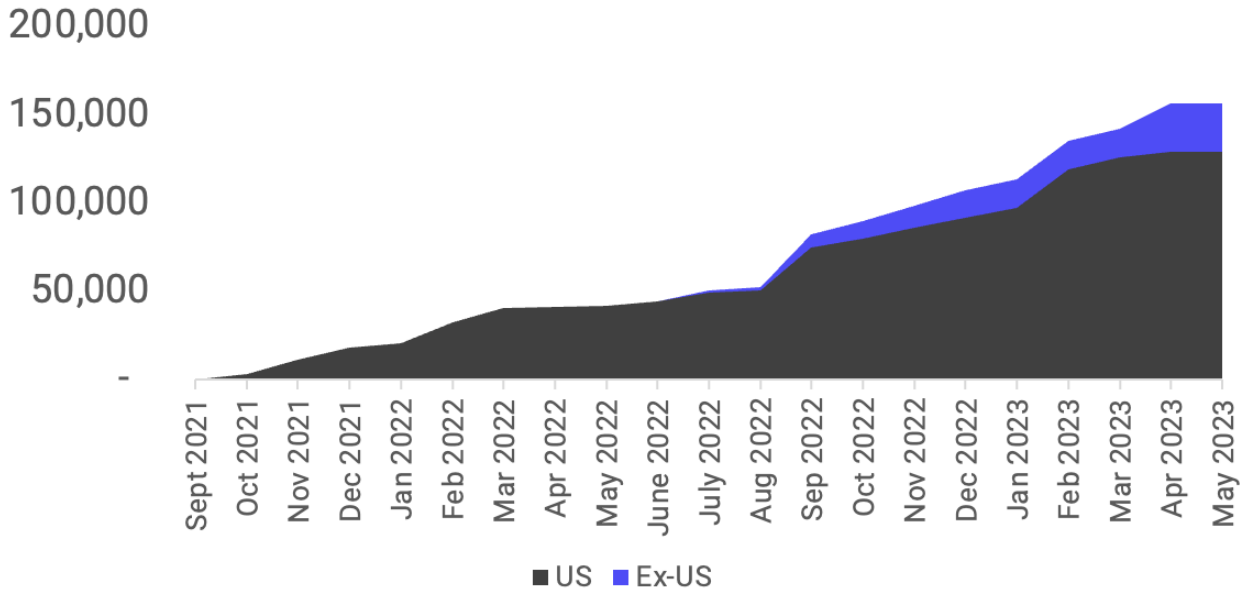


Rapidly growing biobank with samples from across network

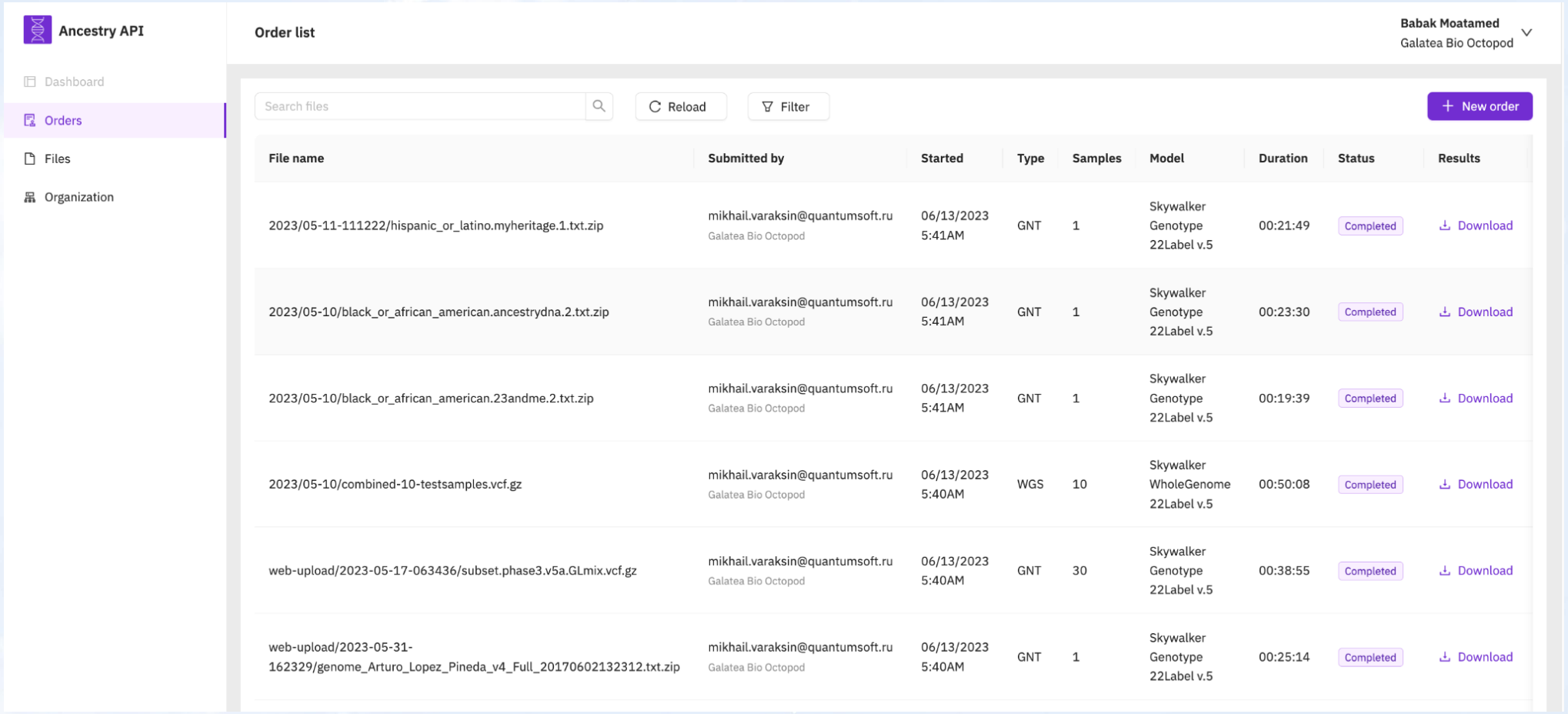
Total Sample (Cumulative, Monthly Volume)



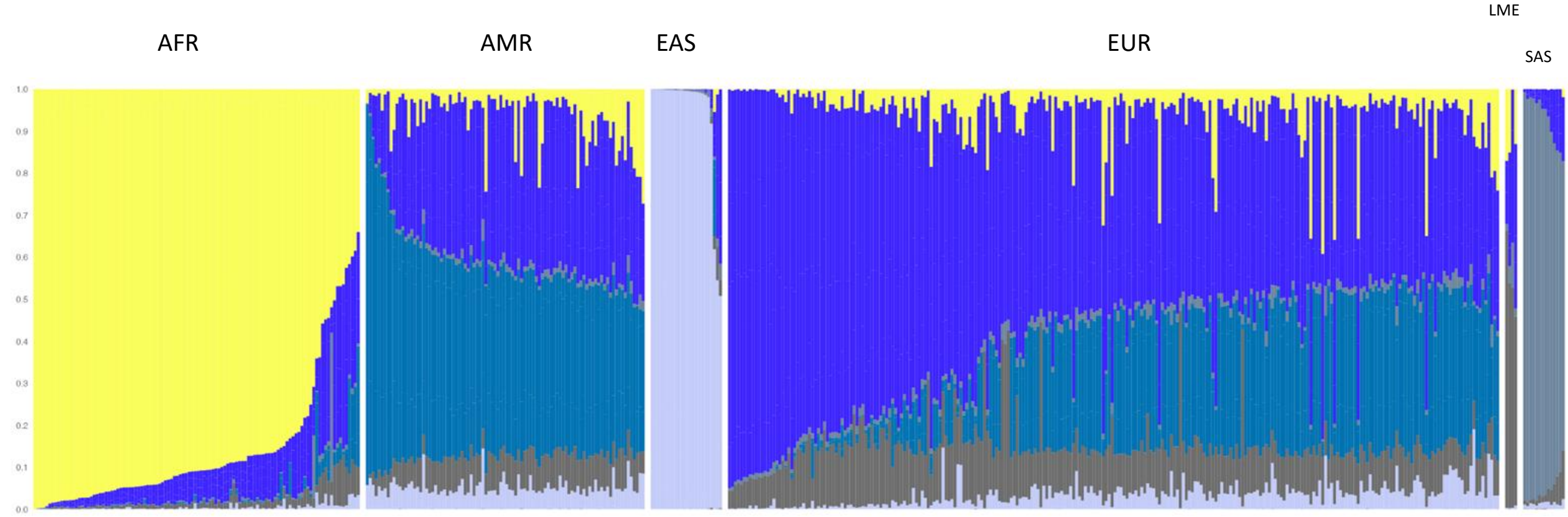
Total Sample (by Geography)



Ancestry API – diverse data inputs and rapidly improving models

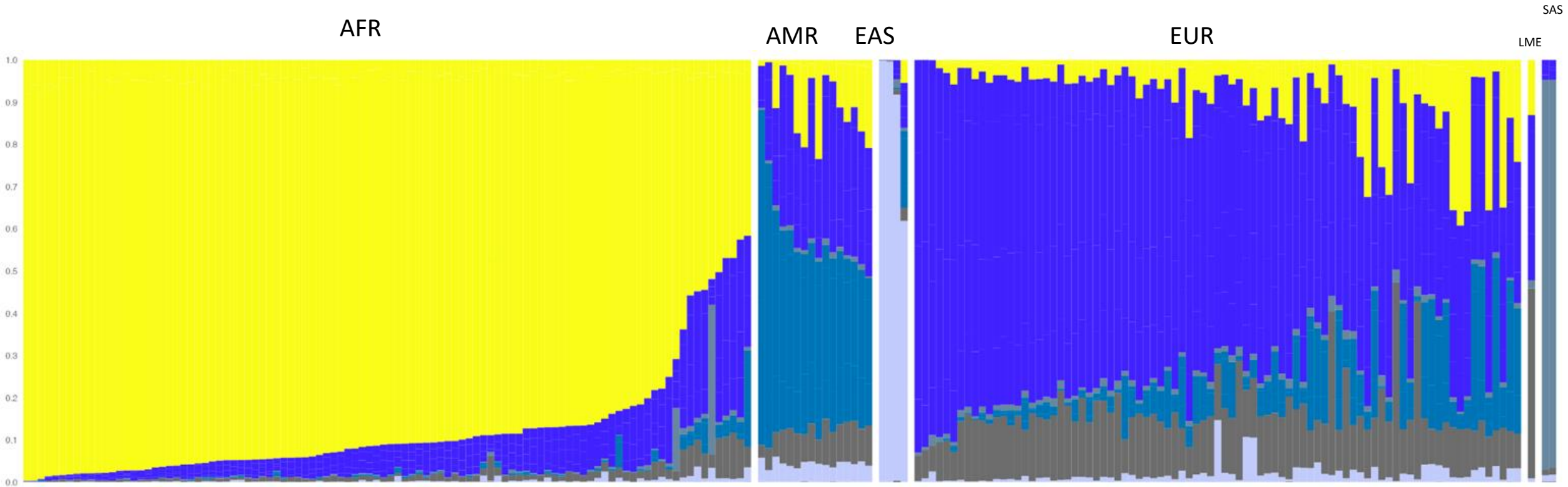


GALATEA BIO DATABASE



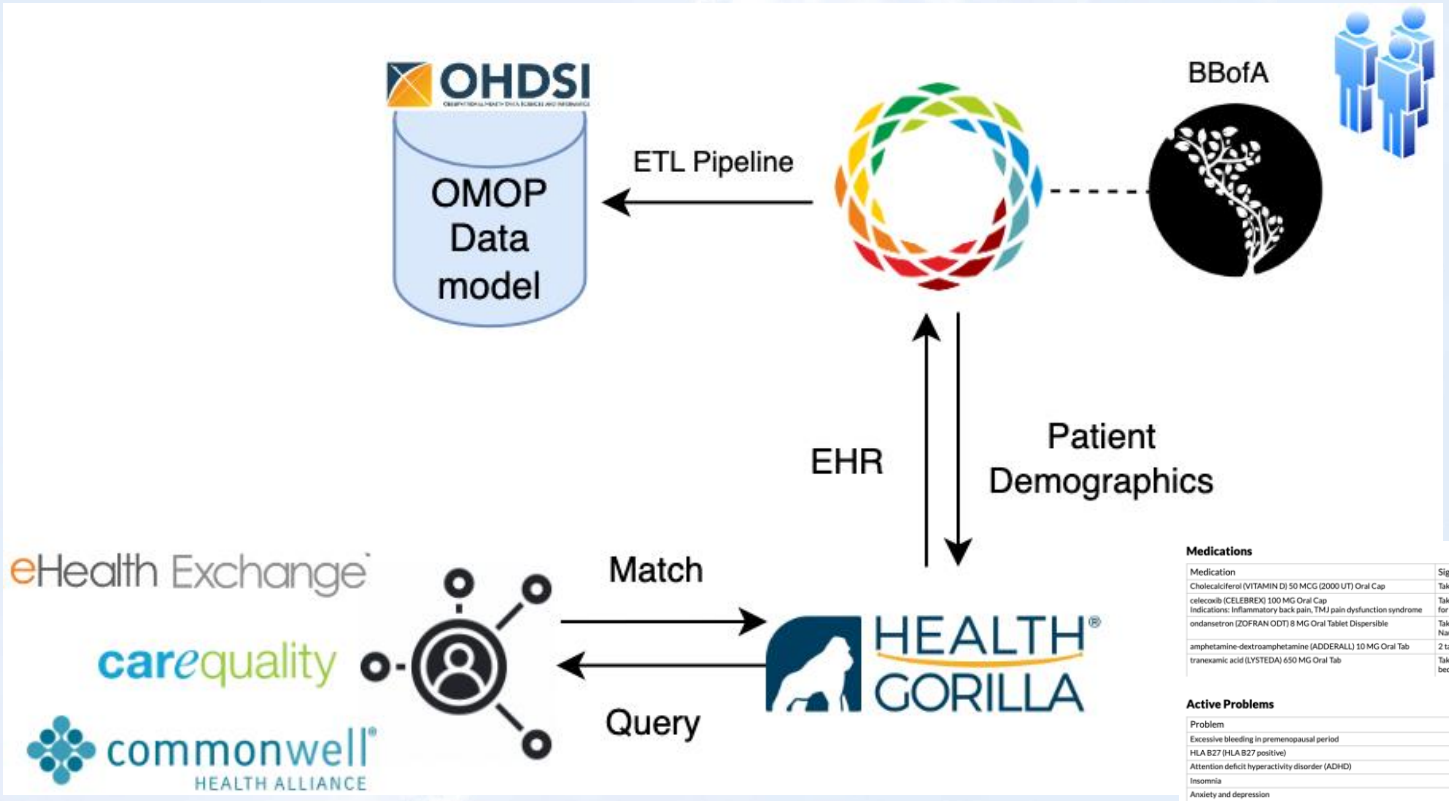
AFR African
AMR American Native
EAS East Asian
EUR European
LME Levantine & Middle East
SAS South Asian
OCE Oceanian

MDC



AFR African
AMR American Native
EAS East Asian
EUR European
LME Levantine & Middle East
SAS South Asian
OCE Oceanian

Automated process for connecting genetics to clinical data



Medications

Medication	Sig	Dispensed	Refills	Start Date	End Date	Status
Cholecalciferol (VITAMIN D) 50 MCG (2000 UT) Oral Cap	Take 1 capsule by mouth daily.		0			Active
celecoxib (CELEBREX) 100 MG Oral Cap Indications: Inflammatory back pain, TMJ pain dysfunction syndrome	Take 1 capsule (100 mg total) by mouth 2 (two) times daily as needed for Pain.	180 capsule	0	01/12/2022		Active
ondansetron (ZOFRAN ODT) 8 MG Oral Tablet Dispersible	Take 1 tablet (8 mg total) by mouth every 8 (eight) hours as needed for Nausea.	20 tablet	0	01/27/2022		Active
amphetamine-dextroamphetamine (ADDERALL) 10 MG Oral Tab	2 tab PO QAM then 1 tab PO between noon and 3pm as needed	90 tablet	0	02/15/2022		Active
tranexamic acid (LYSTEDA) 650 MG Oral Tab	Take 2 tablets (1,300 mg total) by mouth in the morning, at noon, and at bedtime.	30 tablet	3	03/22/2022		Active

Active Problems

Problem	Noted Date
Excessive bleeding in premenopausal period	03/23/2022
HLA B27 (HLA B27 positive)	02/19/2020
Attention deficit hyperactivity disorder (ADHD)	11/01/2017
Insomnia	11/01/2017
Anxiety and depression	12/07/2015
ADD (attention deficit disorder)	

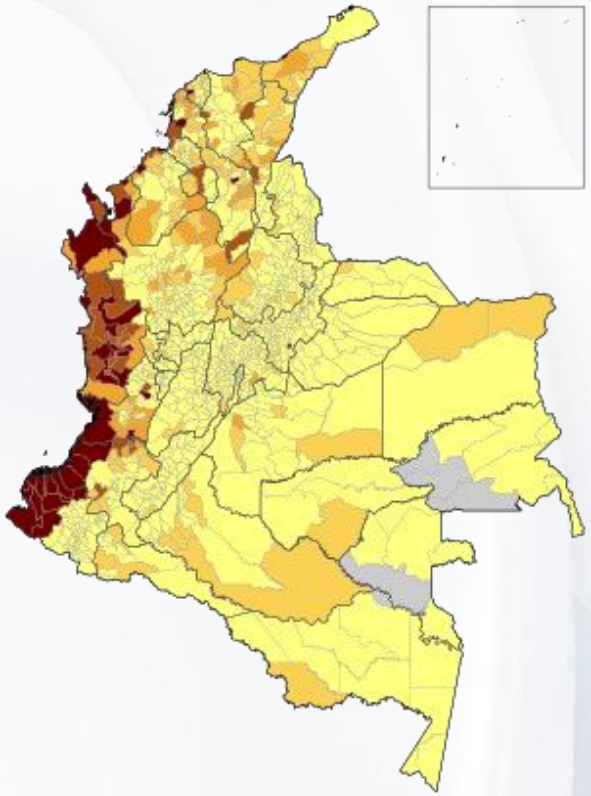
Immunizations

Name	Administration Dates	Next Due
Covid-19 Vaccine Pfizer 30 mcg/0.3 ml	10/07/2021, 01/08/2021, 12/18/2020	
HEP B, Ped/Adol	04/19/2004, 11/21/2003, 10/21/2003	
Hep B, Unspecified Formulation	04/19/2004, 11/21/2003, 10/21/2003	
Influenza	10/07/2019, 10/09/2018, 10/17/2017, 10/05/2016, 10/15/2014, 10/01/2013, 11/01/2007, 11/05/2003	
TDAP	12/20/2013, 10/18/2013	

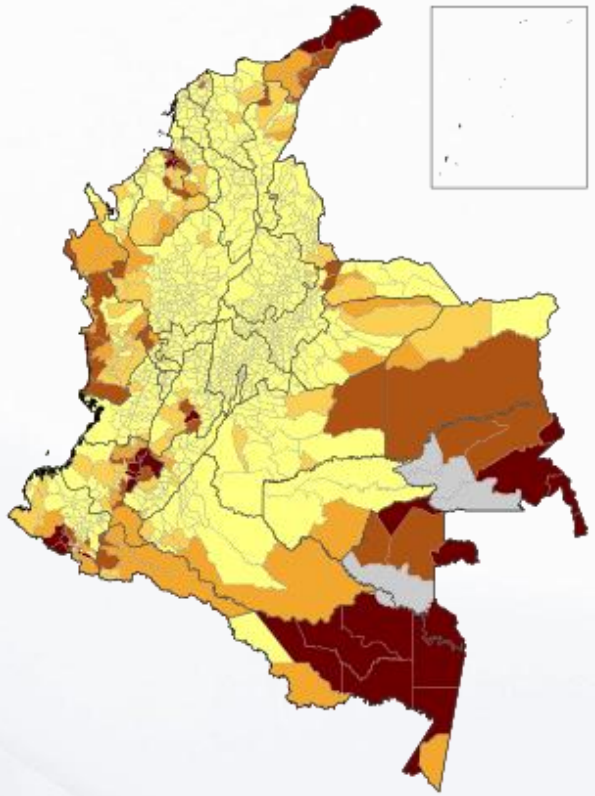
Family History

Medical History	Relation	Comments
Other[other]	Brother	ankylosing spondylitis
Hypertension	Father	
Diabetes	Maternal Grandfather	
Hypertension	Maternal Grandfather	
Stroke	Maternal Grandfather	

We have built a collaborative network in Colombia

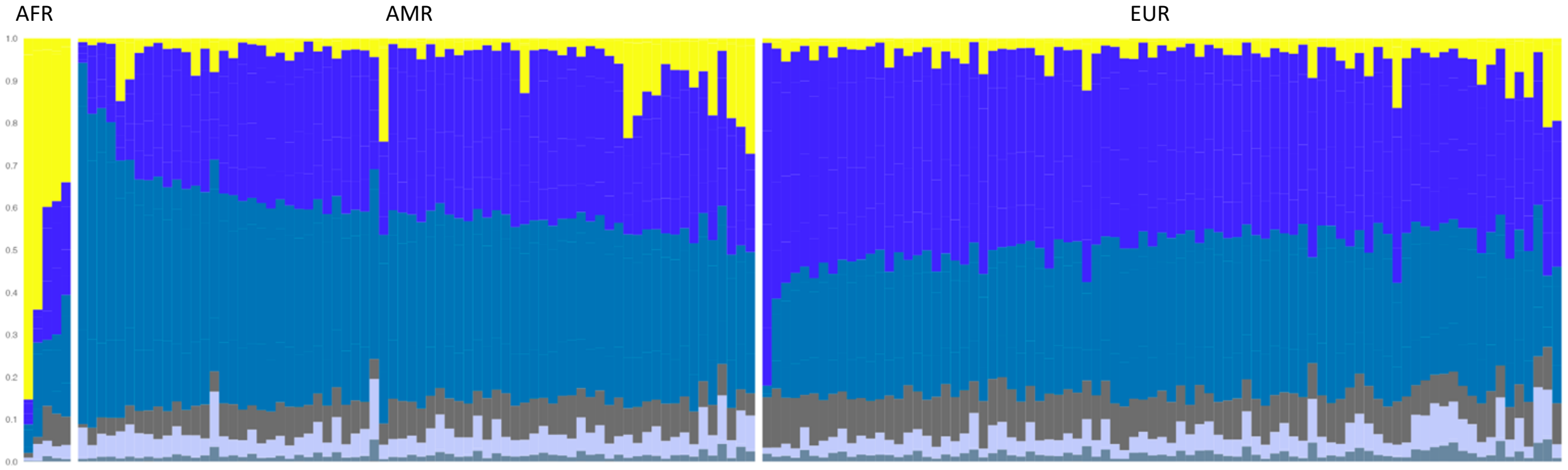


People with African ancestry in Colombia are concentrated mostly in coastal areas.



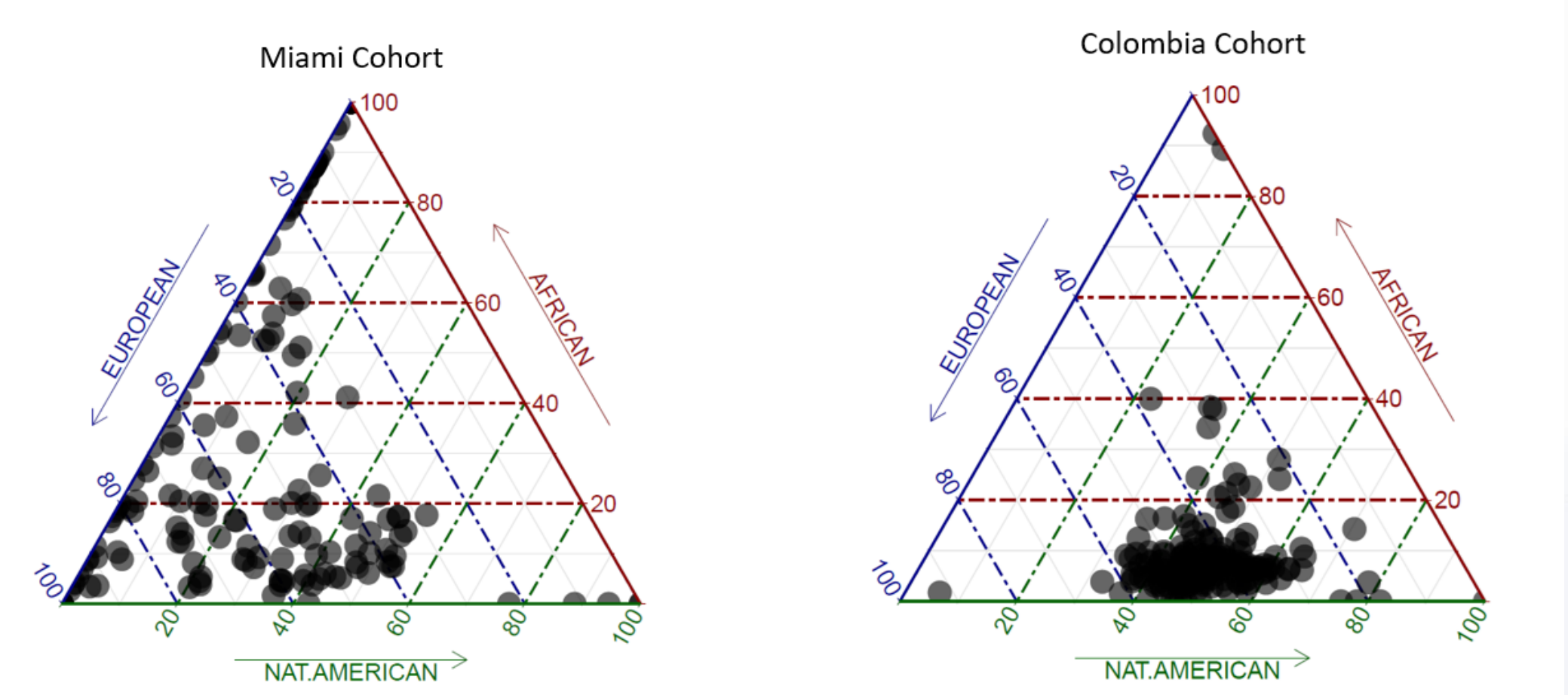
Amerindian population of Colombia by municipality in 2005.

Colombia



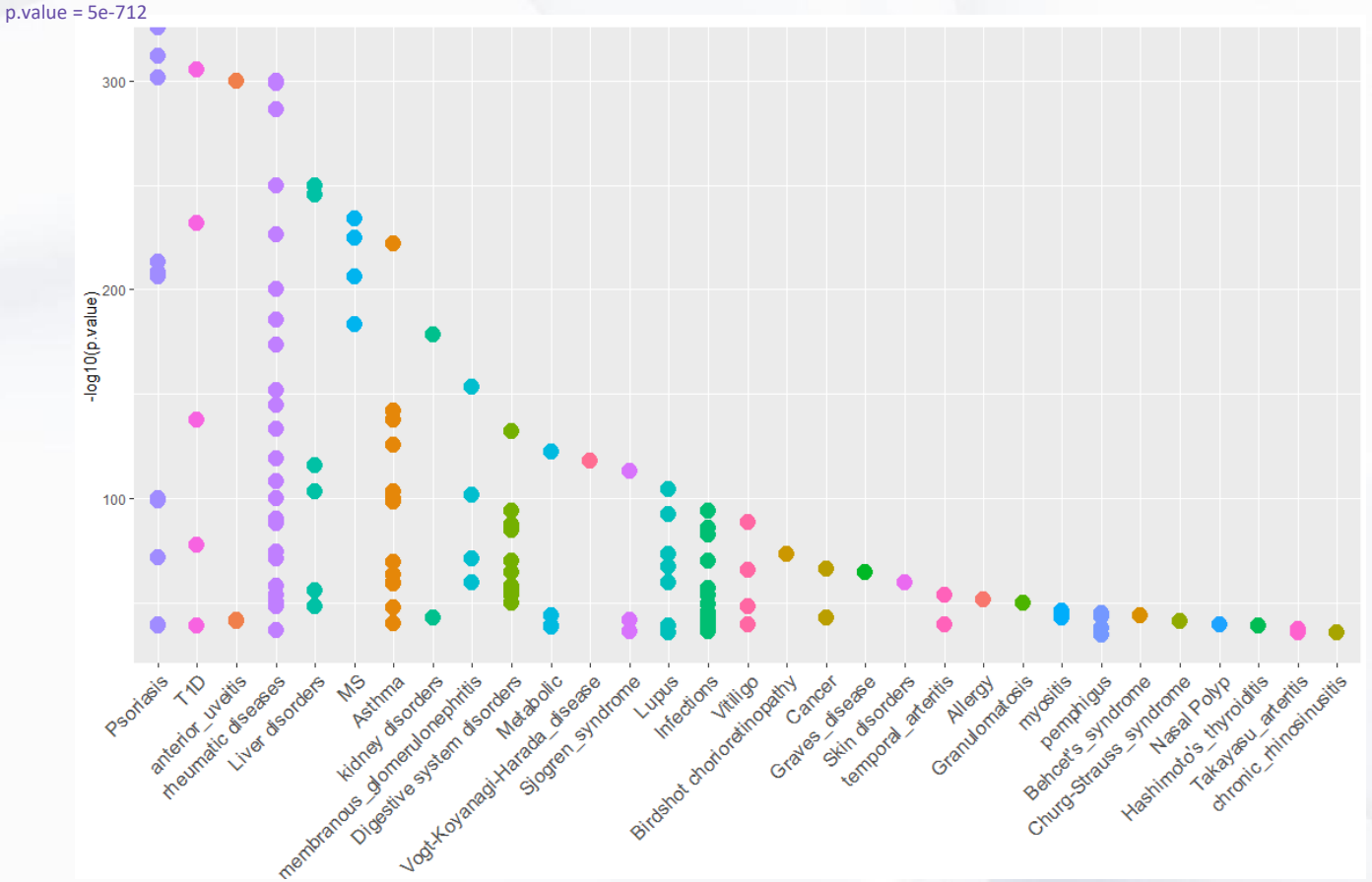
AFR African
AMR American Native
EAS East Asian
EUR European
LME Levantine & Middle East
SAS South Asian
OCE Oceanian

Colombian cohort is enriched in Native-American ancestry



Disease-related HLA associations

Autoimmune related disorders, including: psoriasis, T1D, anterior uveitis and rheumatic disorders; are leading to HLA-driven associations



HLA-associations mapped for: HLA-A, B, C, DRB1, DQB1, DPB1 genes

Information extracted and curated from GWAS Catalog on Oct 2022 (https://www.ebi.ac.uk/gwas/)

Seeding collaborative networks - Paraguay

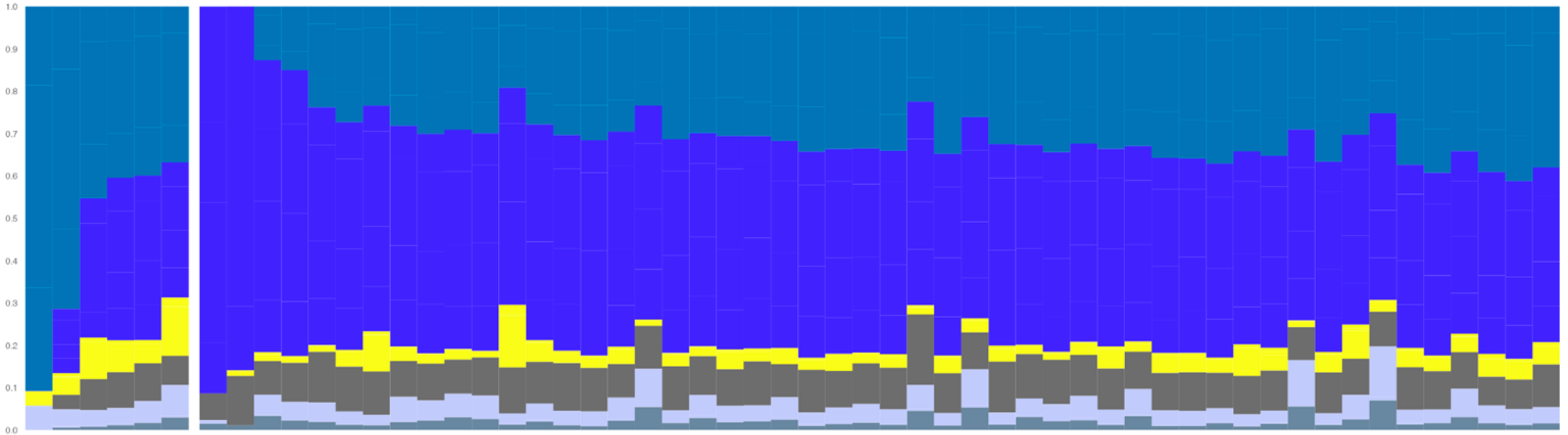
A triple alliance composed of three sites: CEDIC Research Institute, Hospital de Caaguazú, and CIMFACISA, will lead us to access to Paraguayan population

Paraguay population (~7M) is composed of an admixture of Native Americans, Africans, and Europeans (mostly migrants from the Iberian peninsula)

Paraguay

AMR

EUR



AFR African

AMR American Native

EAS East Asian

EUR European

LME Levantine & Middle East

SAS South Asian

OCE Oceanian

A growing network in Mexico

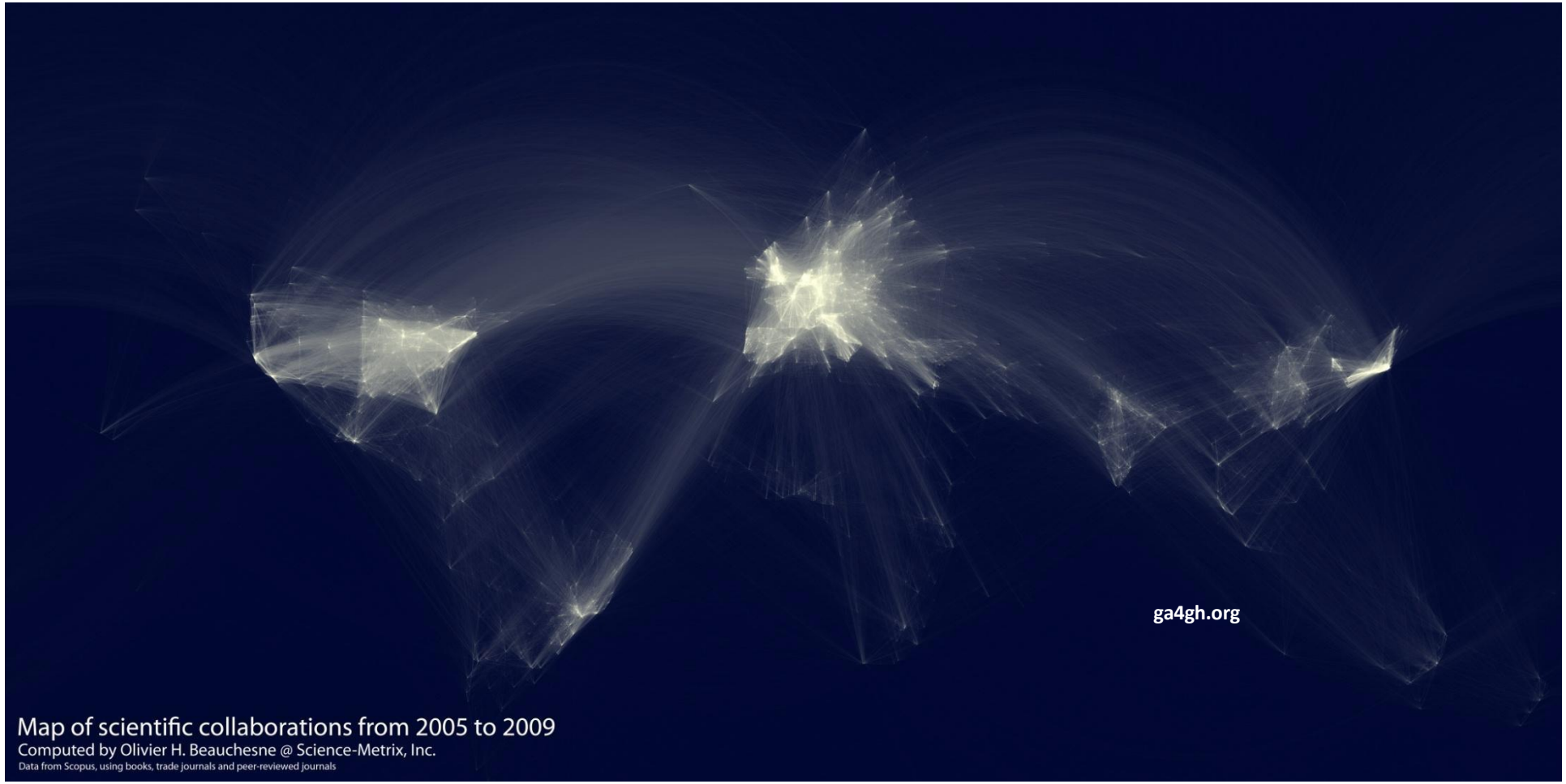
Genotype first – Convert sample into genetic data which may be used to identify interesting markers for additional studies

Pros: collecting genetic information on all samples can help drive discovery
Cons: mostly sequencing controls so the cost of discovery may be very high

Phenotype first – Collect clinical/phenotype information along with the biological sample. First mine the phenotype data to look for potentially interesting samples which can be followed up with richer genetic data.

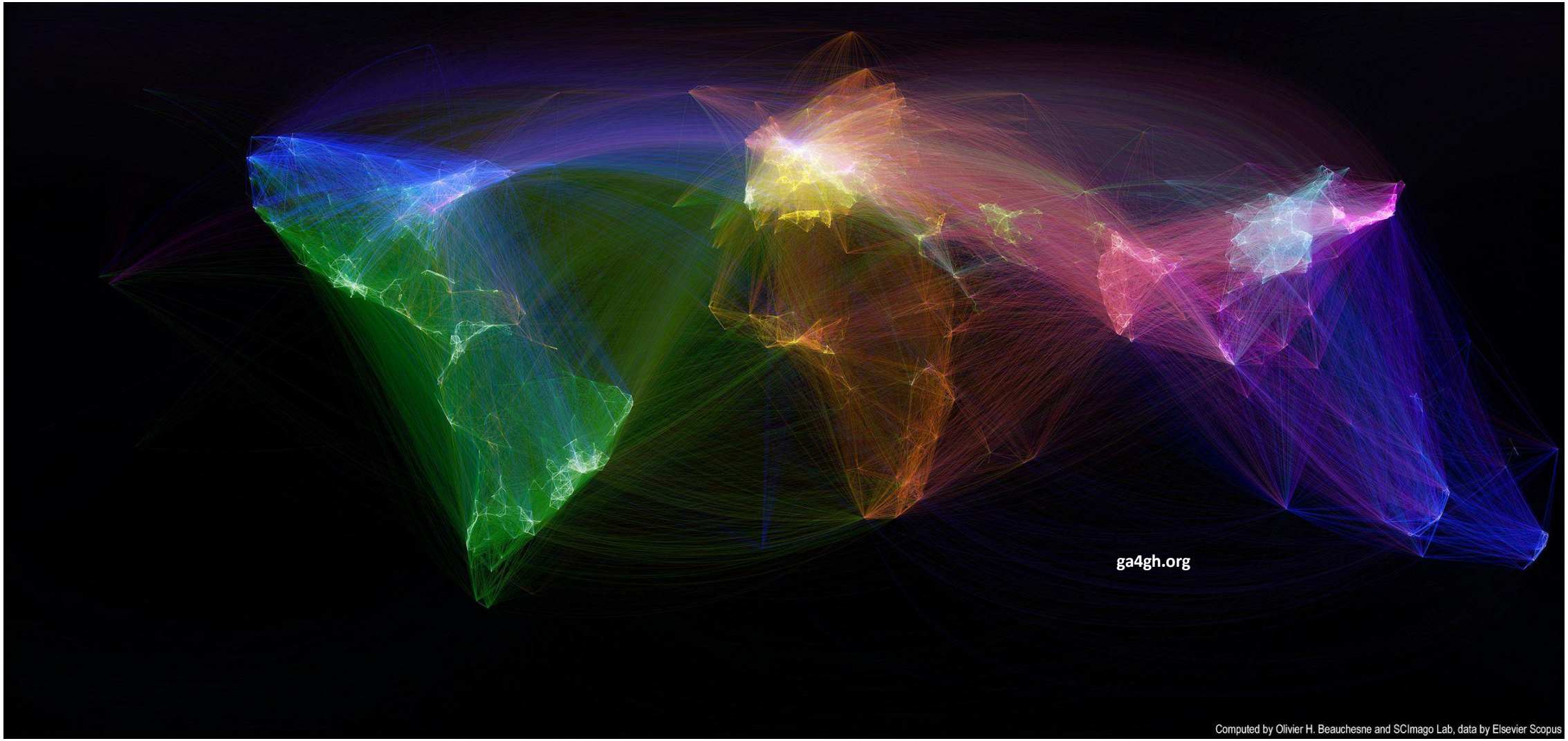
Pros: genetic information is focused on the most interesting samples from a clinical or phenotypic perspective.
Cons: requires access to a rich database of clinical information

Map of Scientific Collaborations from 2005-2009



http://olihb.com/2014/08/11/map-of-scientific-collaboration-redux/

Map of Scientific Collaborations from 2008-2014



http://olihb.com/2014/08/11/map-of-scientific-collaboration-redux/

The world is changing

Percentage of whole genomes and exomes that are funded by **healthcare** systems

Areas of clinical uptake: infectious disease, cancer, rare disease, common/chronic

Genomics enters healthcare

In **2017** active genomic medicine programmes are already underway in many countries. Finland, the UK, the US, and Australia are a few examples.

1 MILLION PEOPLE

The National Institutes of Health plans to recruit over 1 million people into its All of Us research programme.

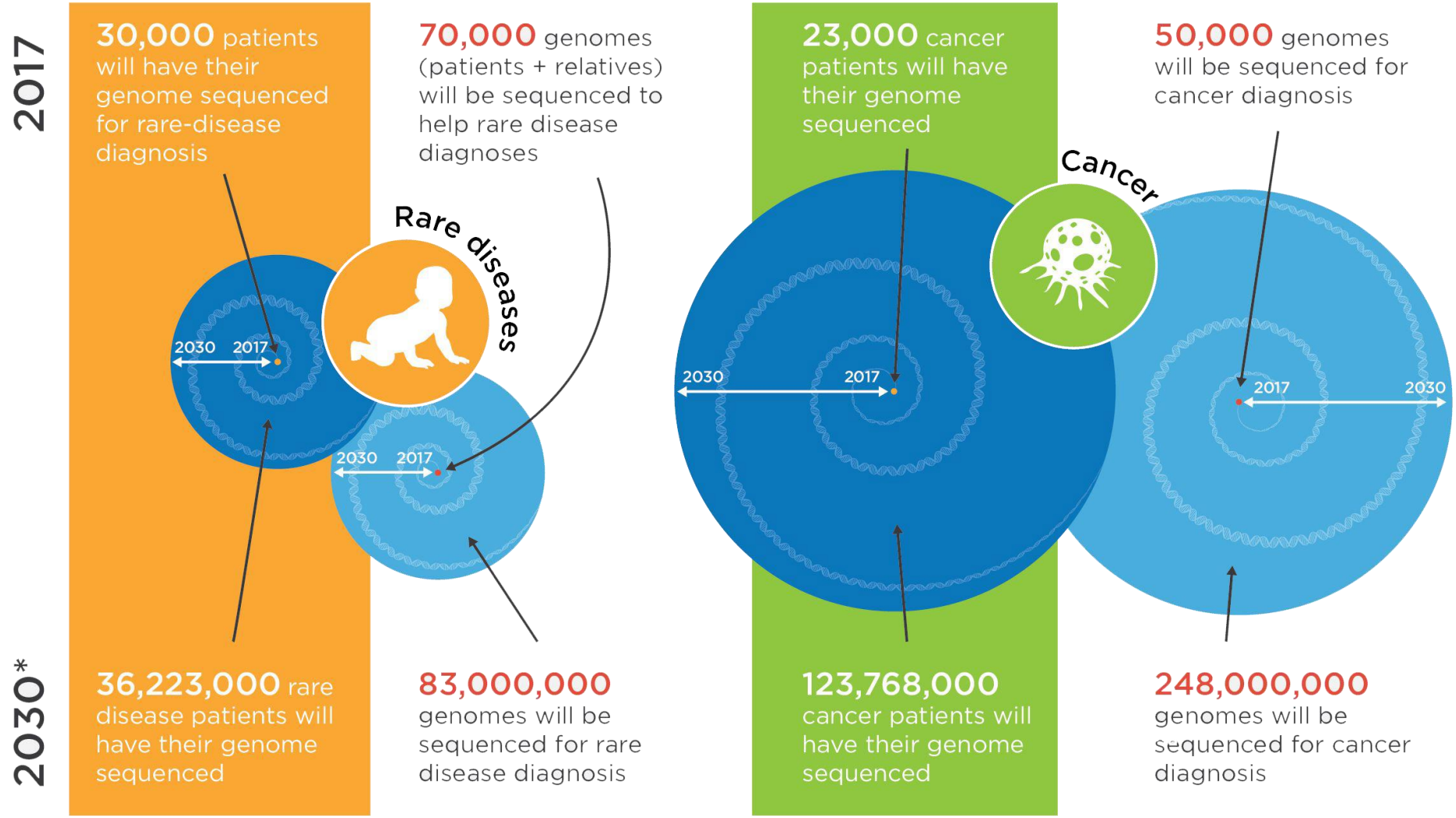
To improve patient outcomes and support research, the Australian healthcare system is building a Federation of clinical and genomic data.

10% of Finland's population is expected to have some genomic data in healthcare by 2020.

10%

The UK National Health Service plans to sequence 100,000 individuals by 2020.

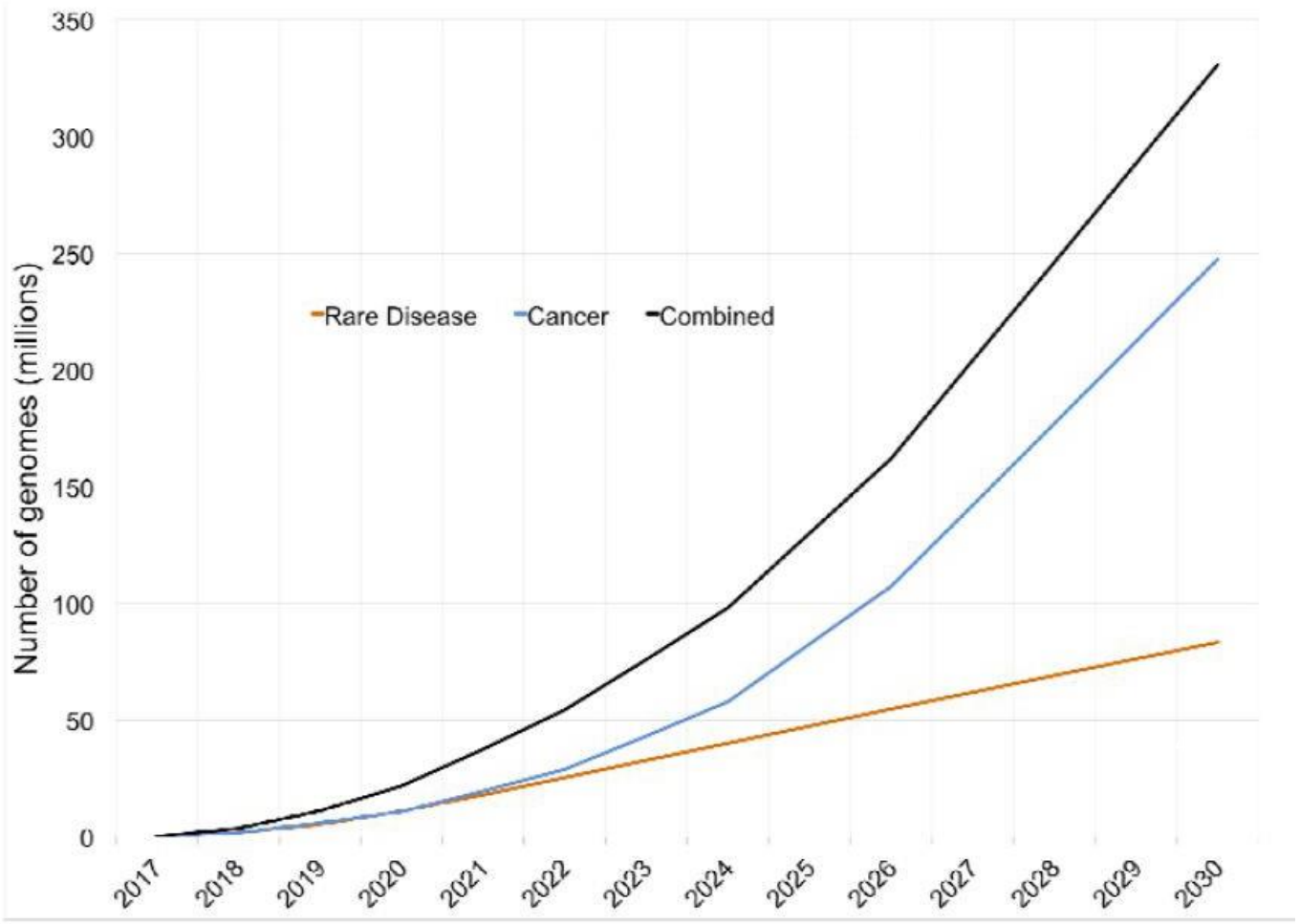
100,000 PATIENTS



* Projected figures, based on current data and known status of genomics initiatives worldwide.

From Op-ed on BioRxiv: doi.org/10.1101/203554

How many genomes?



From Op-ed on BioRxiv: doi.org/10.1101/203554

Opportunity

- If we can enable secondary use of clinical genomic data for research we will have a >60 million virtual cohort by 2025
- Data from millions of samples may be needed to show patterns in health/disease
- Humans will be the best studied organisms on the planet due to healthcare

Genetic analysis at the country-wide level

Andorra has an opportunity to lead and teach how to advance country-level health genomics

Demonstration project: Rare Genetic disorders

>10,000 rare diseases

>population incidence: 3-8% worldwide (2-5,000 Andorrans)

Diagnostic Odysseys

Average: 5-7 years to receive diagnosis

Horrendous cost to families and healthcare system

Tools like PheRS will lead the way in identifying patients at risk and create health care economic case

Phenotype risk scores identify patients with unrecognized Mendelian disease patterns

Lisa Bastarache[1], Jacob J. Hughey[1], Scott Hebbring[2], Joy Marlo[1], Wanke Zhao[3], Wanting T. Ho[3], Sara L.…

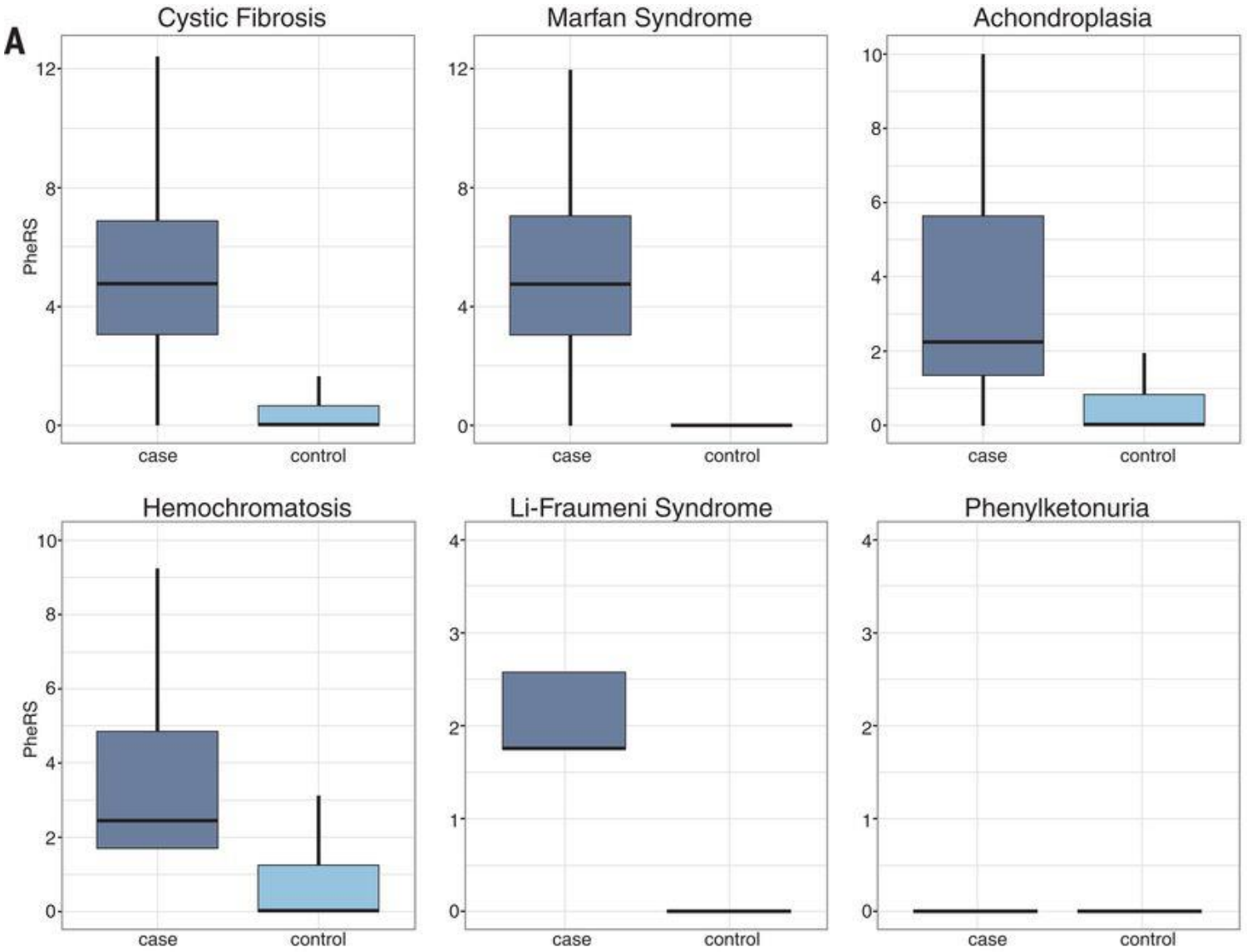
+ See all authors and affiliations

Science 16 Mar 2018:
Vol. 359, Issue 6381, pp. 1233-1239
DOI: 10.1126/science.aal4043

VANDERBILT UNIVERSITY
MEDICAL CENTER

Tools like PheRS will lead the way in identifying patients at risk and create health care economic case

A



B

Disease	Abbrev.	Cases	P
Cystic fibrosis	CF	733	5×10^{-220}
Marfan syndrome	MS	244	2×10^{-175}
Achondroplasia	ACH	101	4×10^{-51}
Hereditary hemochromatosis	HH	83	8×10^{-43}
Li-Fraumeni syndrome	LF	21	4×10^{-64}
Phenylketonuria	PKU	133	0.280

C

PheRS definition

Case/control definition	CF	MS	ACH	HH	LF	PKU
CF	4.50	0.00	0.00	1.43	0.00	0.00
MS	0.00	4.39	0.00	0.00	0.00	0.00
ACH	0.00	0.00	2.11	0.00	0.00	0.00
HH	0.00	0.00	0.00	1.70	0.00	0.00
LF	0.00	0.00	0.00	0.00	1.74	0.00
PKU	0.00	0.00	0.00	0.00	0.00	0.00

Establishing a discovery infrastructure

Table stakes:

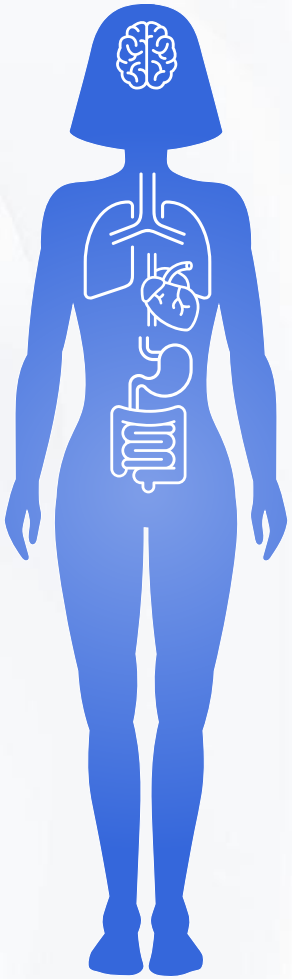
- Large network of clinics and partners able to recruit diverse populations
- Scalable bioinformatic tools able to measure and visualize sample diversity
- Scalable wet lab solutions to generate genetic information at low per-sample costs.
- Database to centralize clinical and genetic information (Geno:Pheno DB)

Rapidly build case:control cohorts that are disease specific and can drive new biomarker and druggable target discovery

Clinical cohort to drive discovery and therapeutics

Neurodevelopmental – ASD, Parkinson's, PANS/PANDAS, BIC

Infectious disease – Build on work done during COVID, build tools to understand host and pathogen genetics (zika, chikungunya, HIV, COVID, influenza, dengue …)



Cancer – (colorectal, esophageal, breast, ovarian, lung, stomach) Partner with early intervention clinics and physicians to access samples at average risk and look for genetic risk factors. Build the next generation of genetic tests to understand cancer risk.

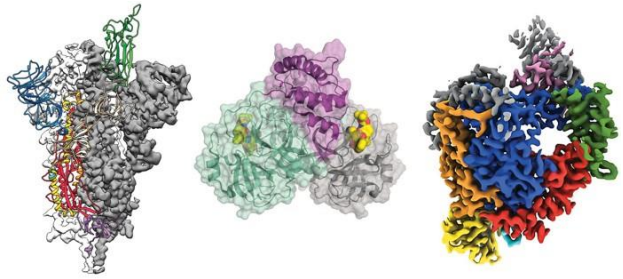
Cardiovascular – Familial hypercholesteremia, diabetes

Rheumatological – autoimmune disorders

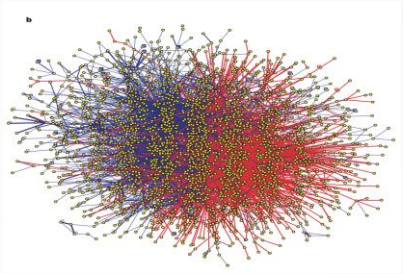
Larger drug discovery opportunity

Large cohorts with EMR and expanding -omics

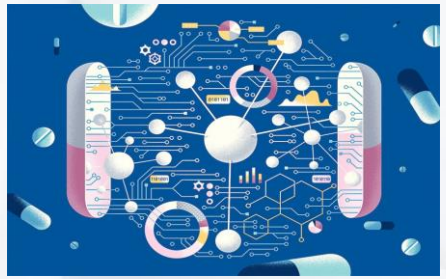
45,000 structures / speed



Massive (and more accurate) PPI networks

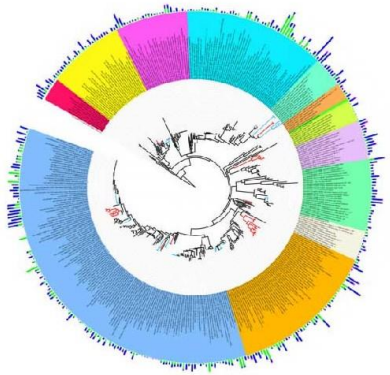


Machine learning to optimize lead discovery

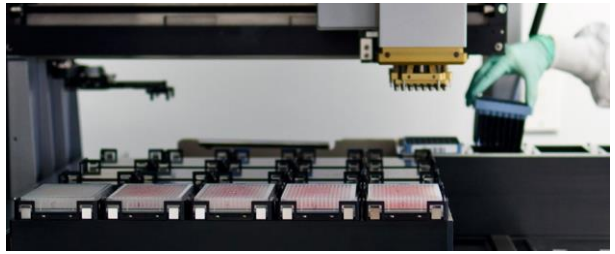


Larger drug discovery opportunity

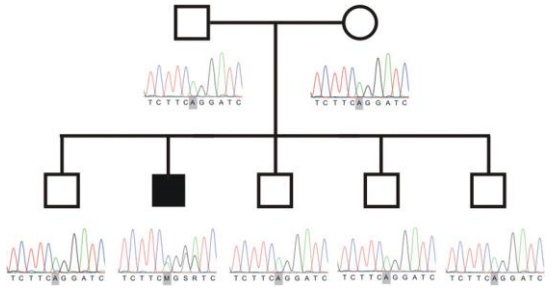
1) From evolutionary genomics to key protein elements



2) From functional screens to new leads



3) From gene discovery to new treatments



Supporting the work

Partner with us – If you have an interesting cohort, access to under-represented patient populations especially in disease areas that we are working on please reach out, or an interest/need in using our software tools and infrastructure.

Enroll with us – www.bbofa.org

Join us – andy.larrea@galatea.bio