Very large-scale biobank projects, experiences from FinnGen

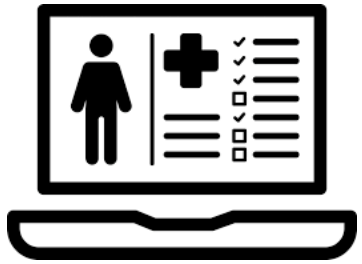
Aarno Palotie, M.D., Ph.D.

University of Helsinki, The Broad Institute of MIT and Harvard, Massachusetts General Hospital

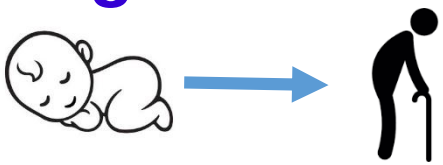
Nordic Biobank Conference 2022


FINNGEN

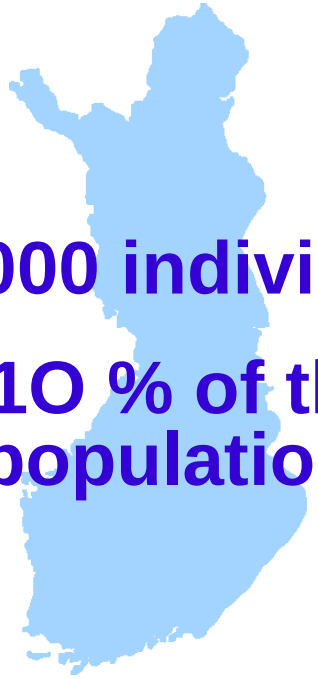
FinnGen



National Health Register data



FINNGEN



500 000 individuals

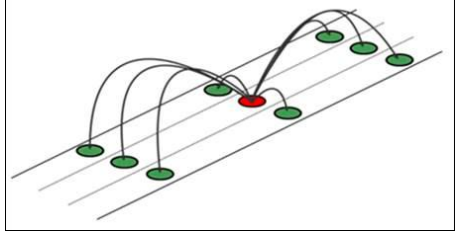
~1O % of the population

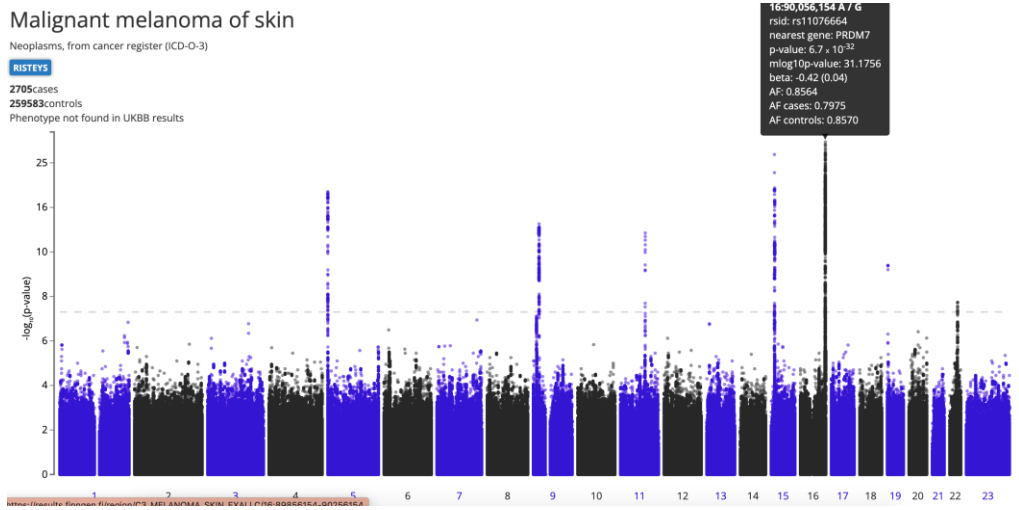
500 000 individuals Combined genotype and register data

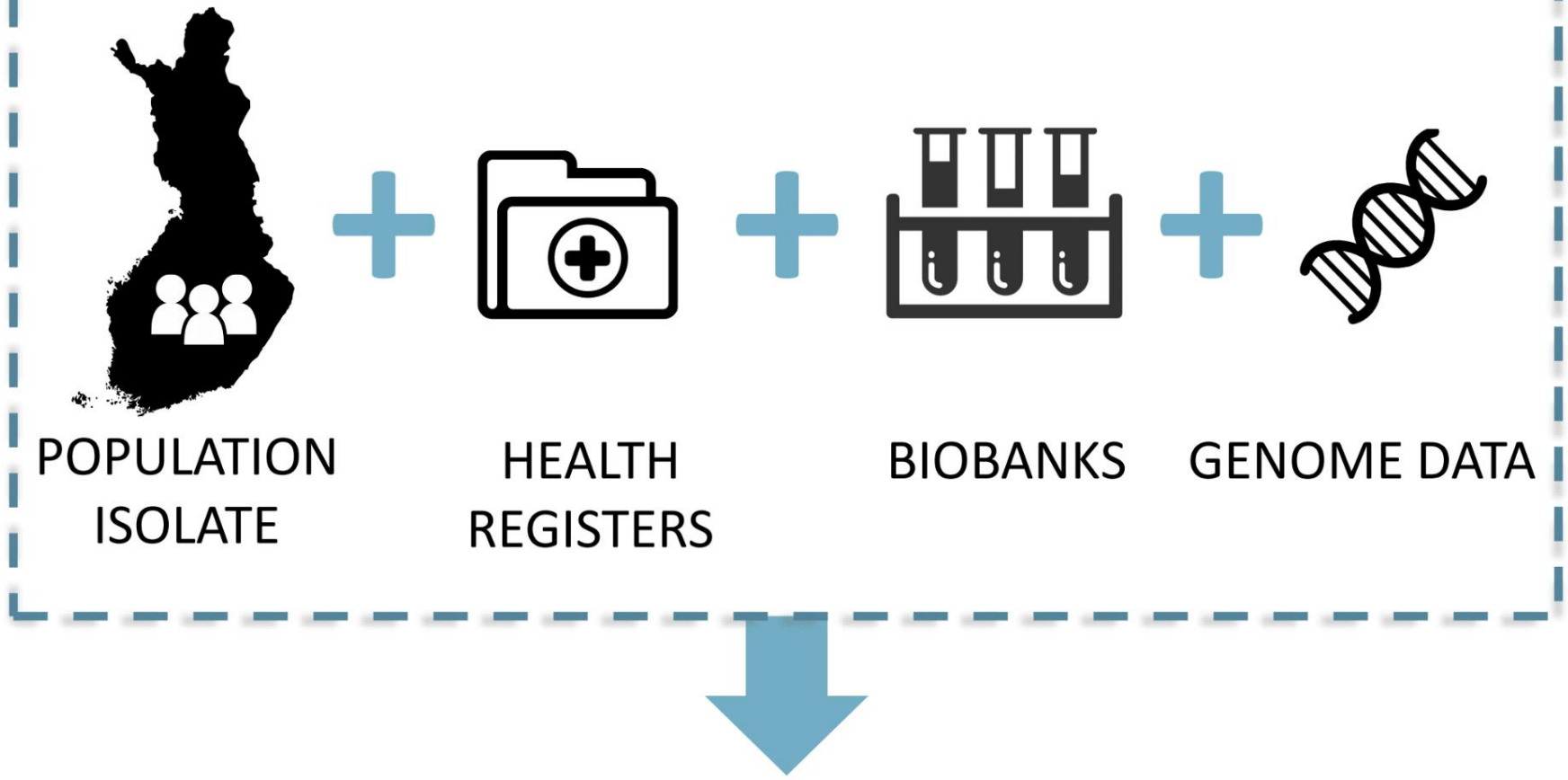
Association analyses

Axiom GWA array

Imputation

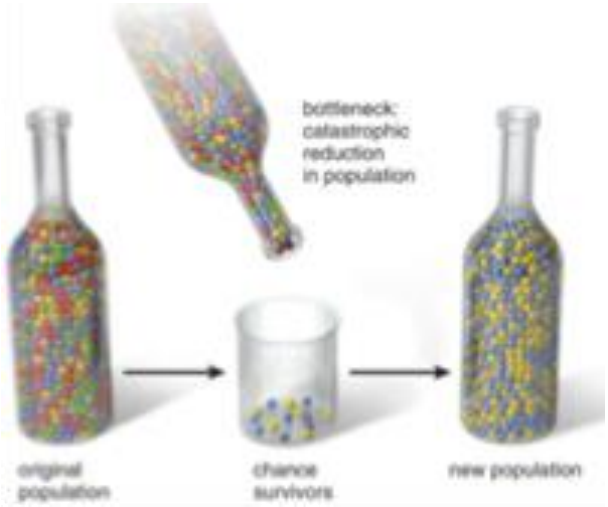


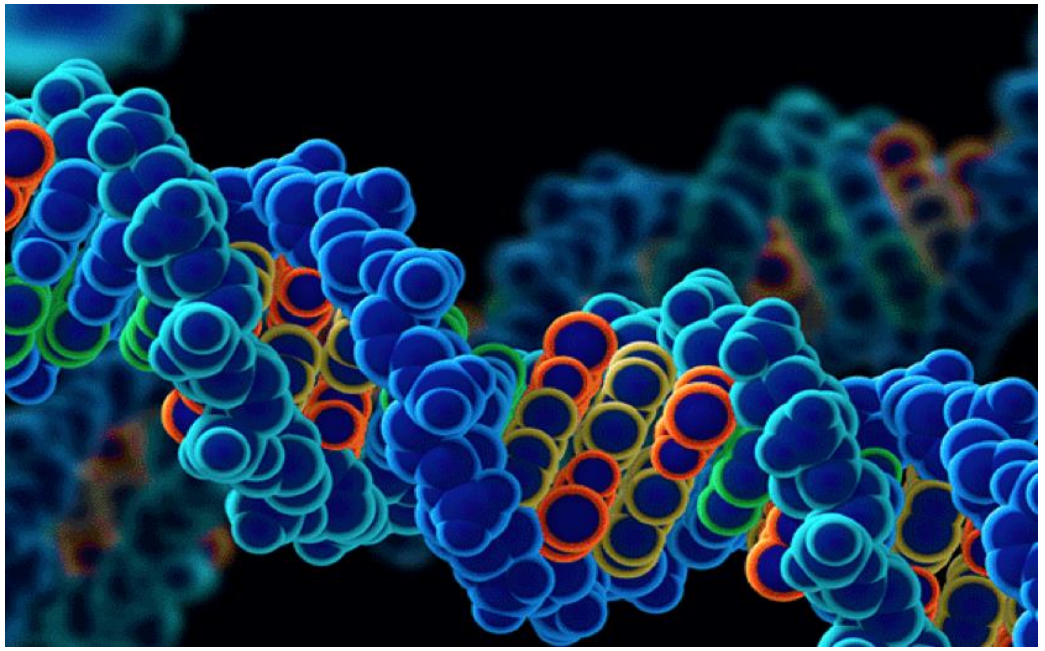




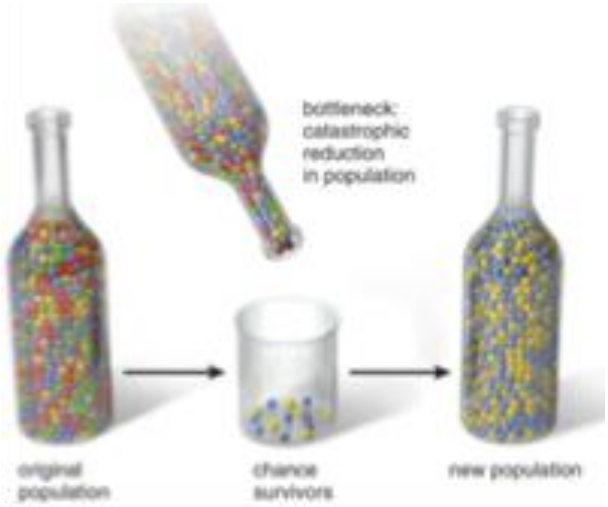
Innovative Study Designs

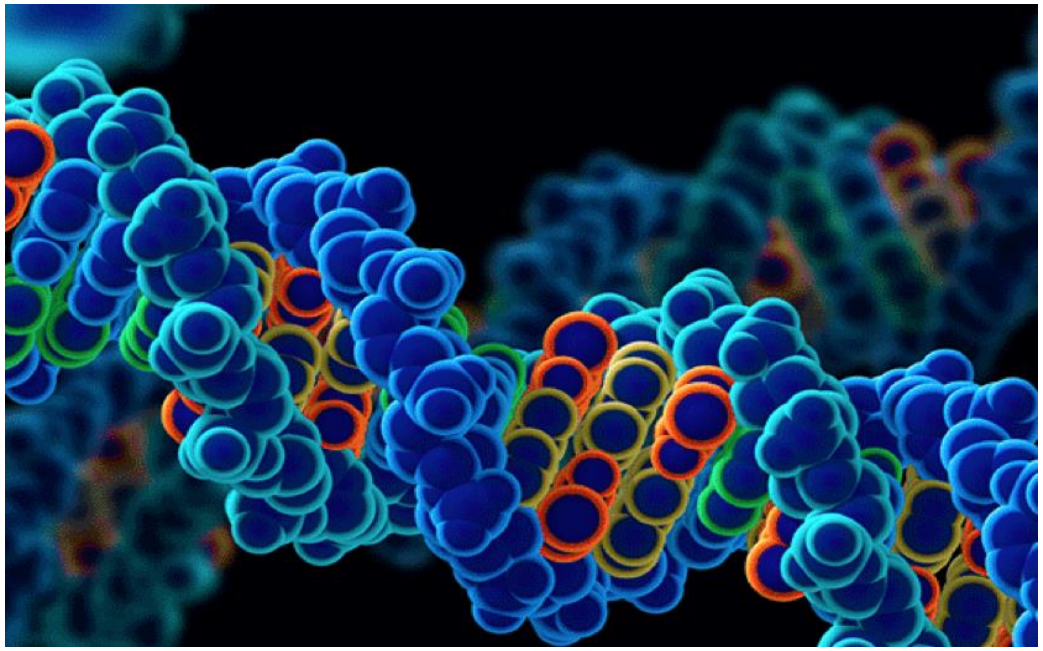
Power of a Genetic Isolate





Power of a Genetic Isolate

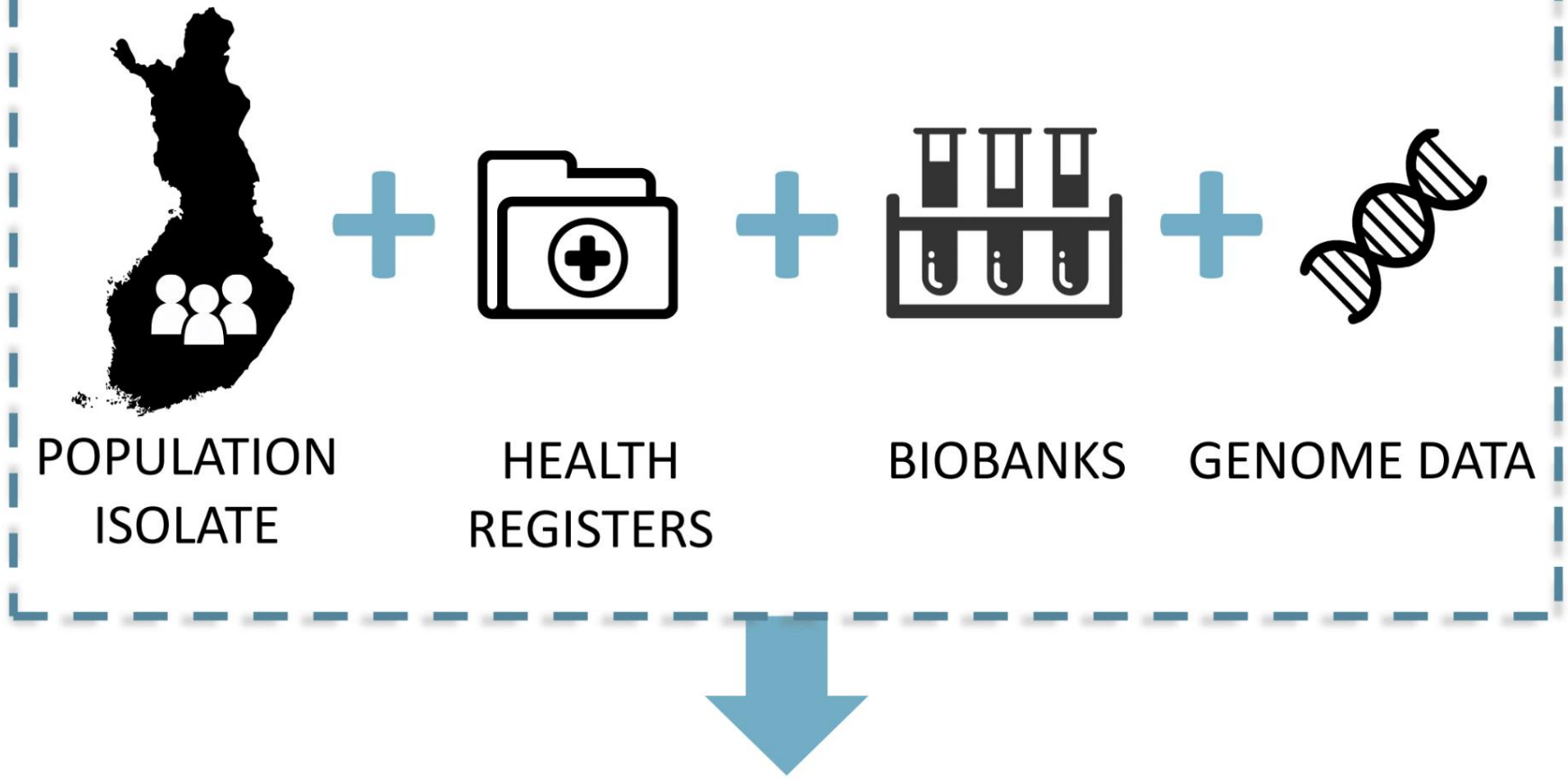




FINNGEN

Specific damaging genetic variants become enriched and easy to discover

Reconstructing genomes of Finns (’imputation’) from inexpensive genotype data is much more accurate than in the rest of the world

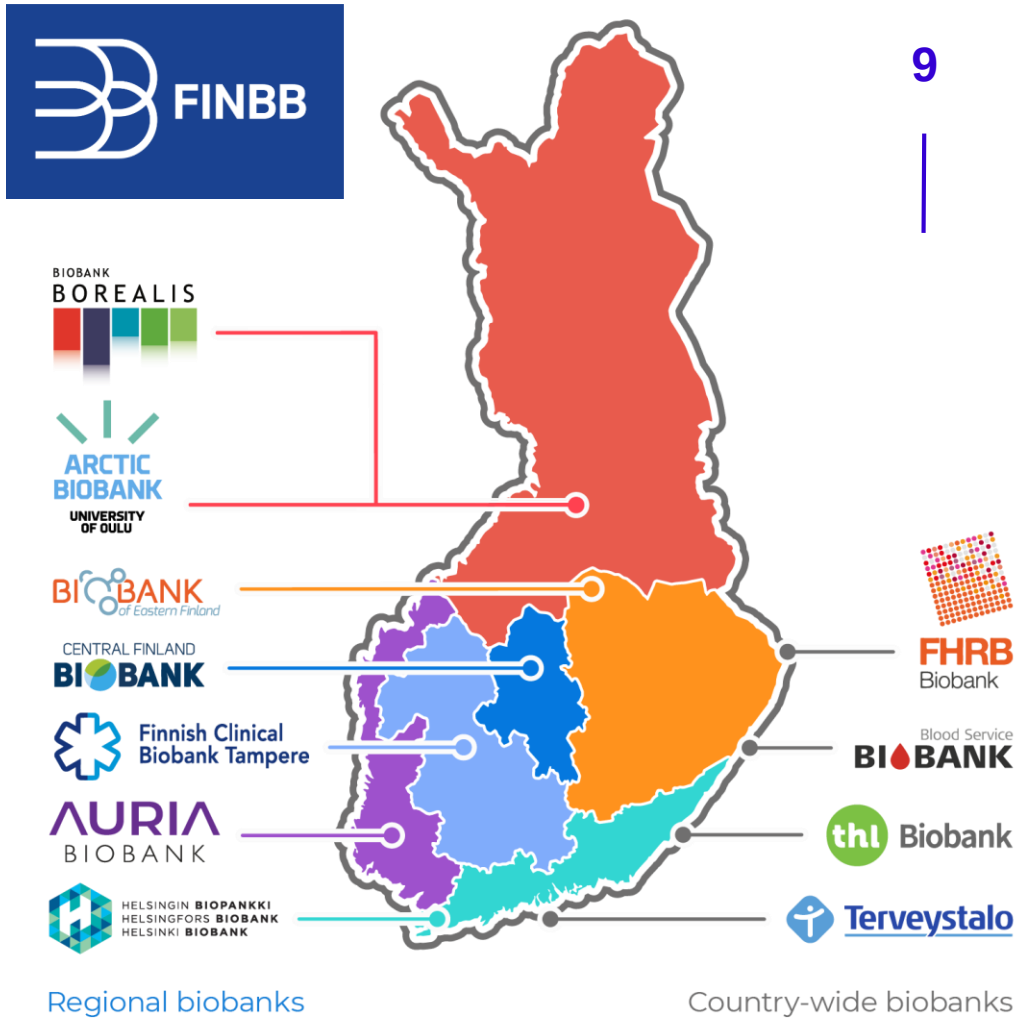


Innovative Study Designs

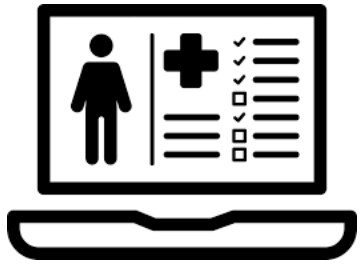
The Finnish Biobank Act is Unique

FINNGEN

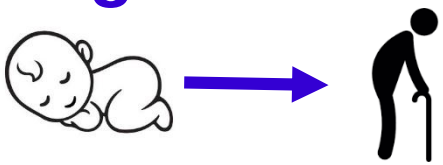
- Registration of biobanks, wide consent and protection of participants
- Transfer of existing sample and data collections to biobanks
- Possibility to recontact
- Possibility to collect samples and data from the health care
- **Collaboration with industry**



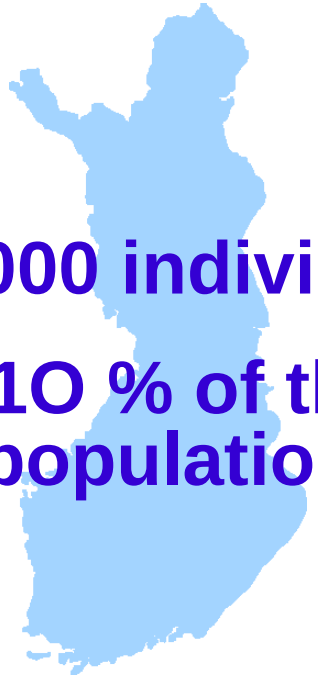
FinnGen



National Health Register data



FINNGENEN



500 000 individuals

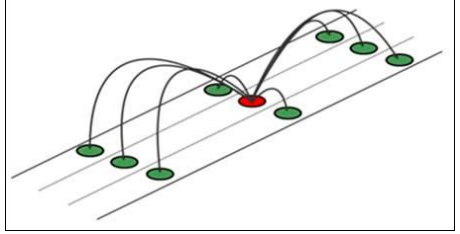
~1O % of the population

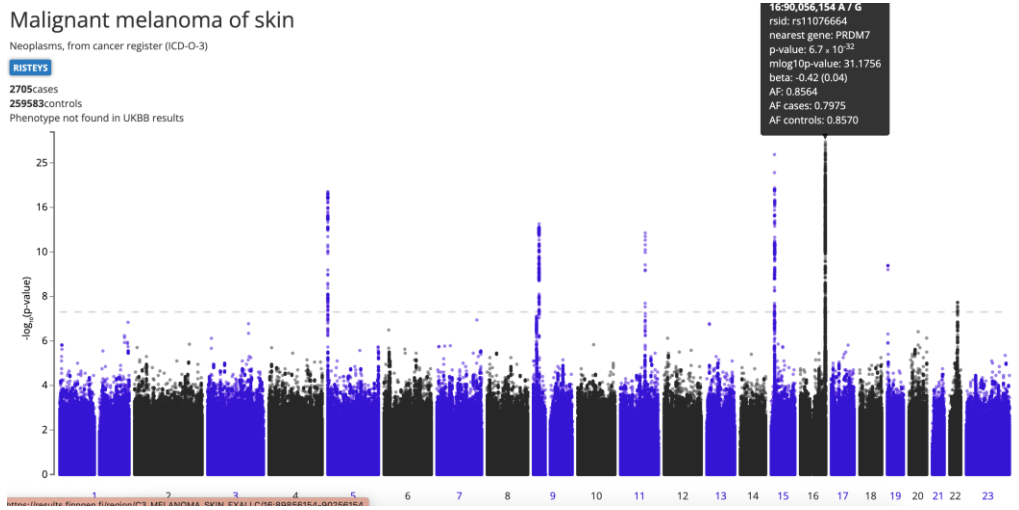
500 000 individuals Combined genotype and register data

Association analyses

Axiom GWA array

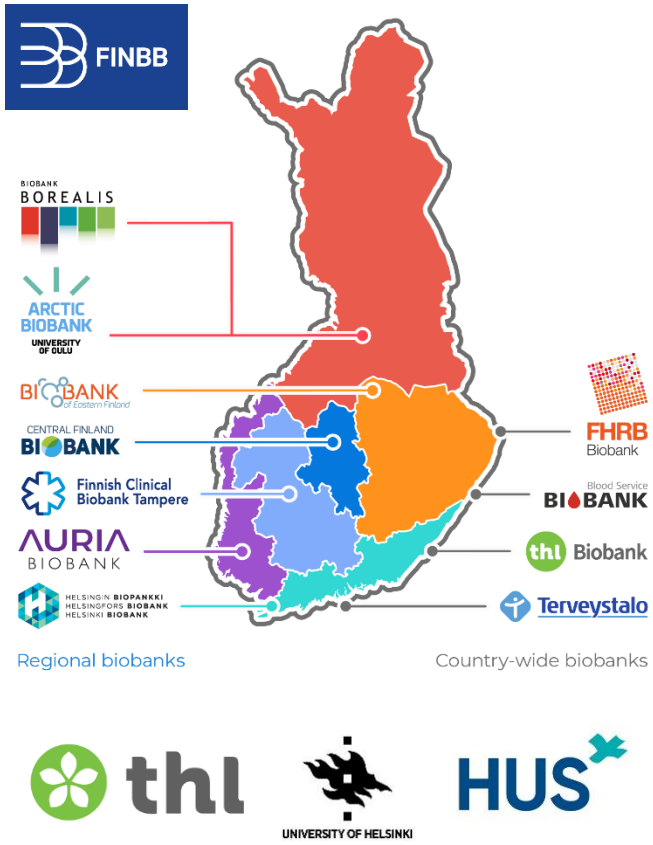
Imputation





Public-private research project

FINNGEN



thl

UNIVERSITY OF HELSINKI

HUS

Academic institutions and public health care

Pharmaceutical industry

Boehringer Ingelheim

abbvie

AstraZeneca

Biogen

Bristol-Myers Squibb

Genentech
A Member of the Roche Group

gsk

janssen
PHARMACEUTICAL COMPANIES OF Johnson & Johnson

maze

MSD

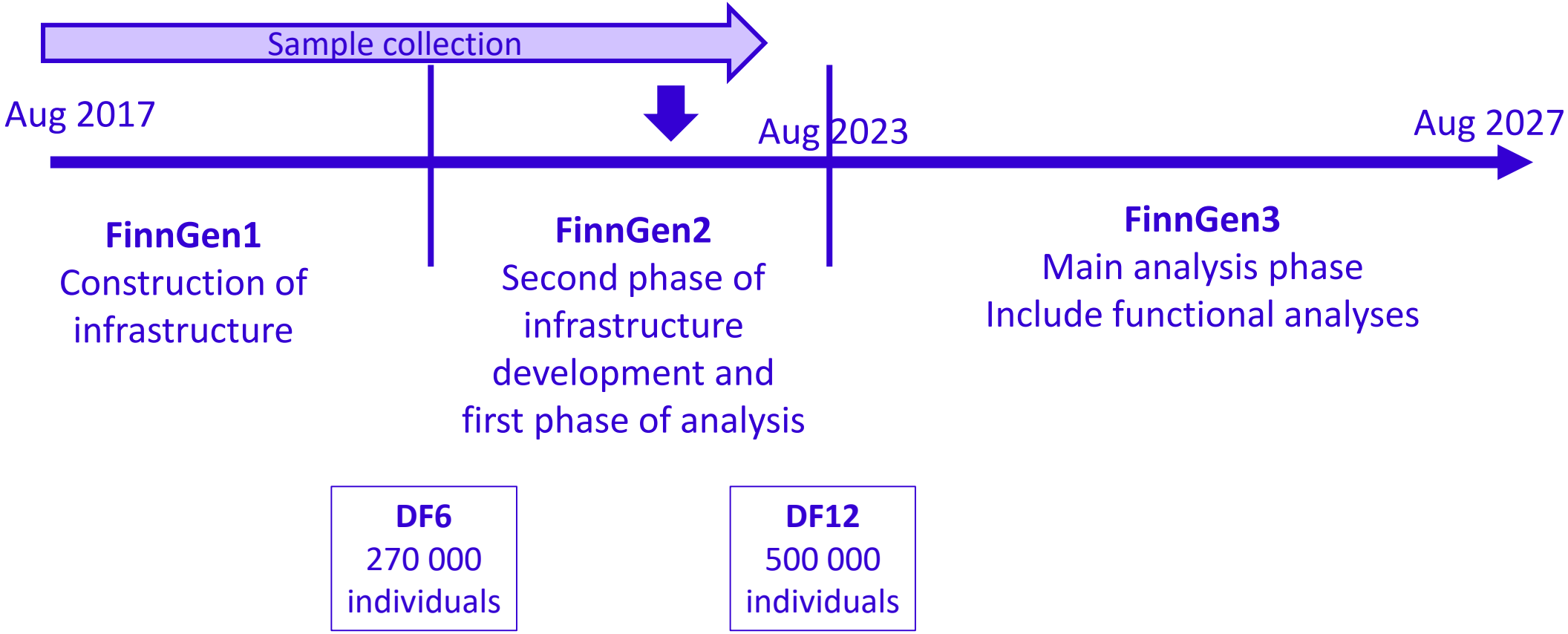
NOVARTIS

Pfizer

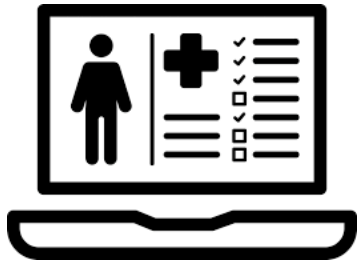
SANOFI

BUSINESS FINLAND

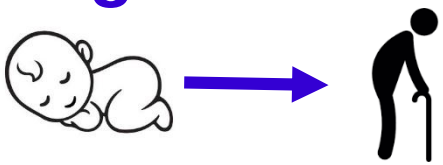
FinnGen, a 10 year project



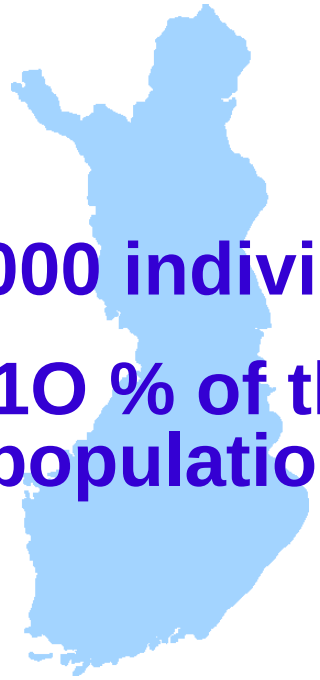
FinnGen



National Health Register data



FINNGENEN



500 000 individuals

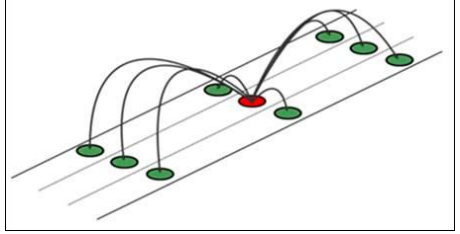
~1O % of the population

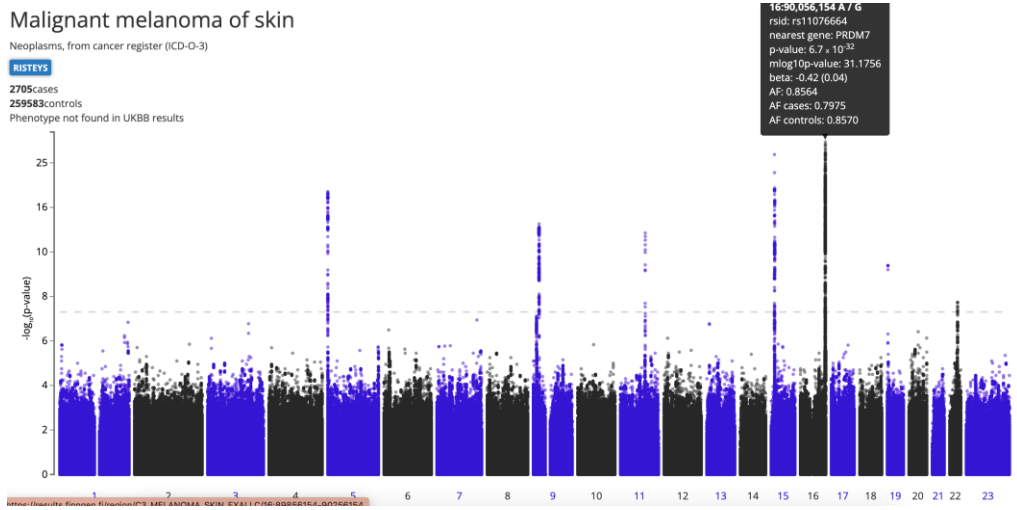
500 000 individuals Combined genotype and register data

Association analyses

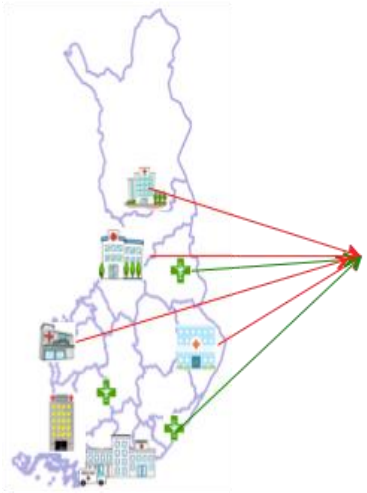
Axiom GWA array

Imputation





Nationwide health registries

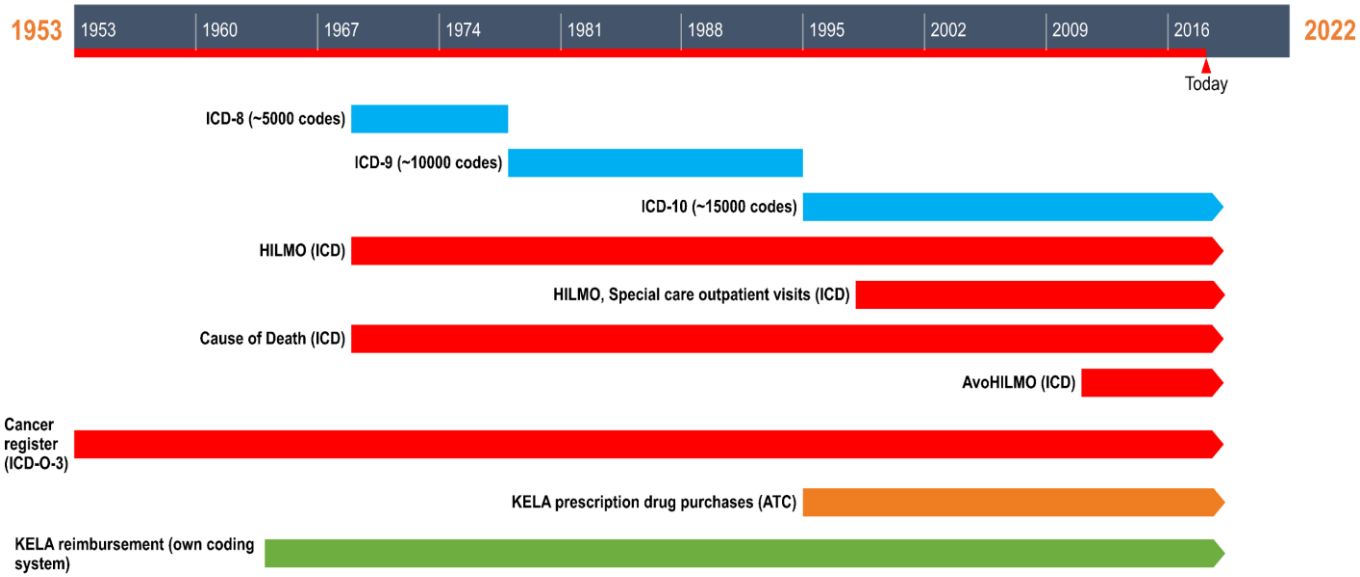


- Hospital discharge
- Hospital procedure
- Outpatient visit
- Outpatient procedure
- Primary care
- Primary care procedure
- Cancer register
- Cause of death
- Drug purchase
- Drug reimbursement

Register data for administration

Register data for administration

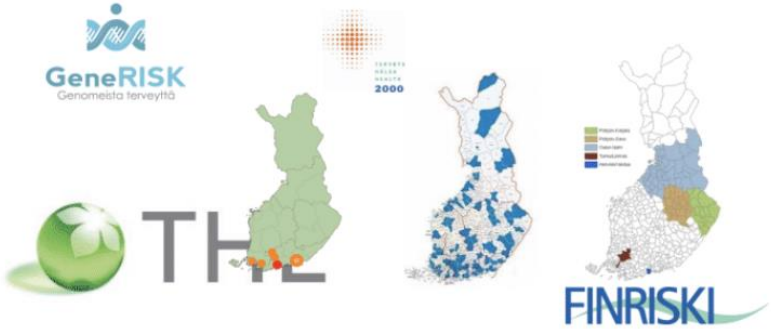
Register data for administration

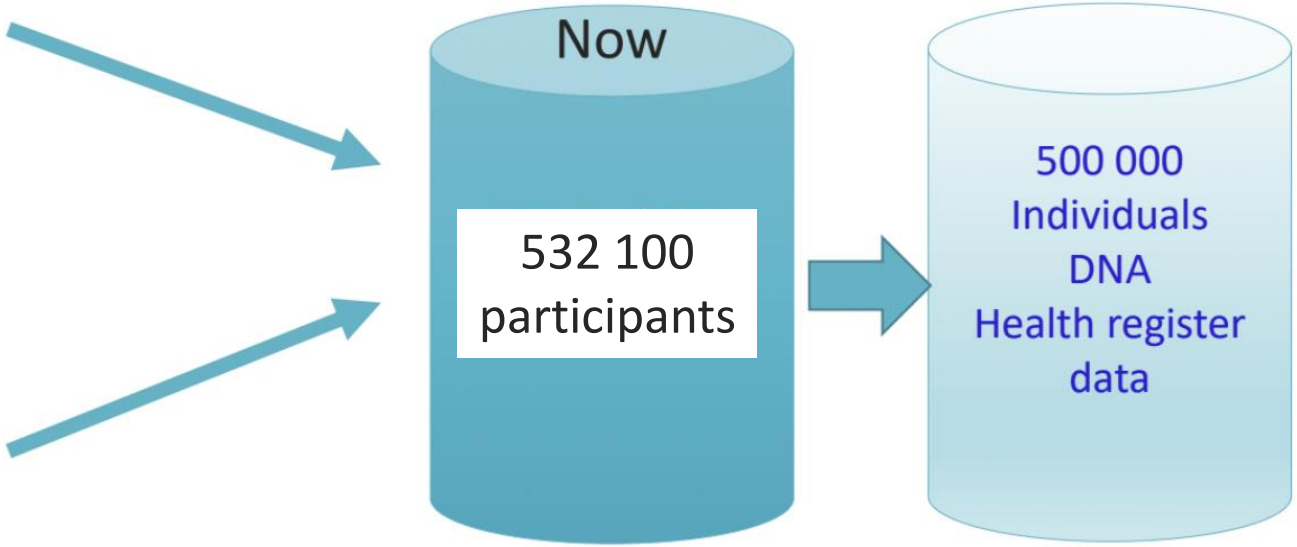


Prospective collections 300 000

Hospital biobanks

Legacy collections 200 000





532 100 participants, so far (target 525 000)

359 200 new s ted

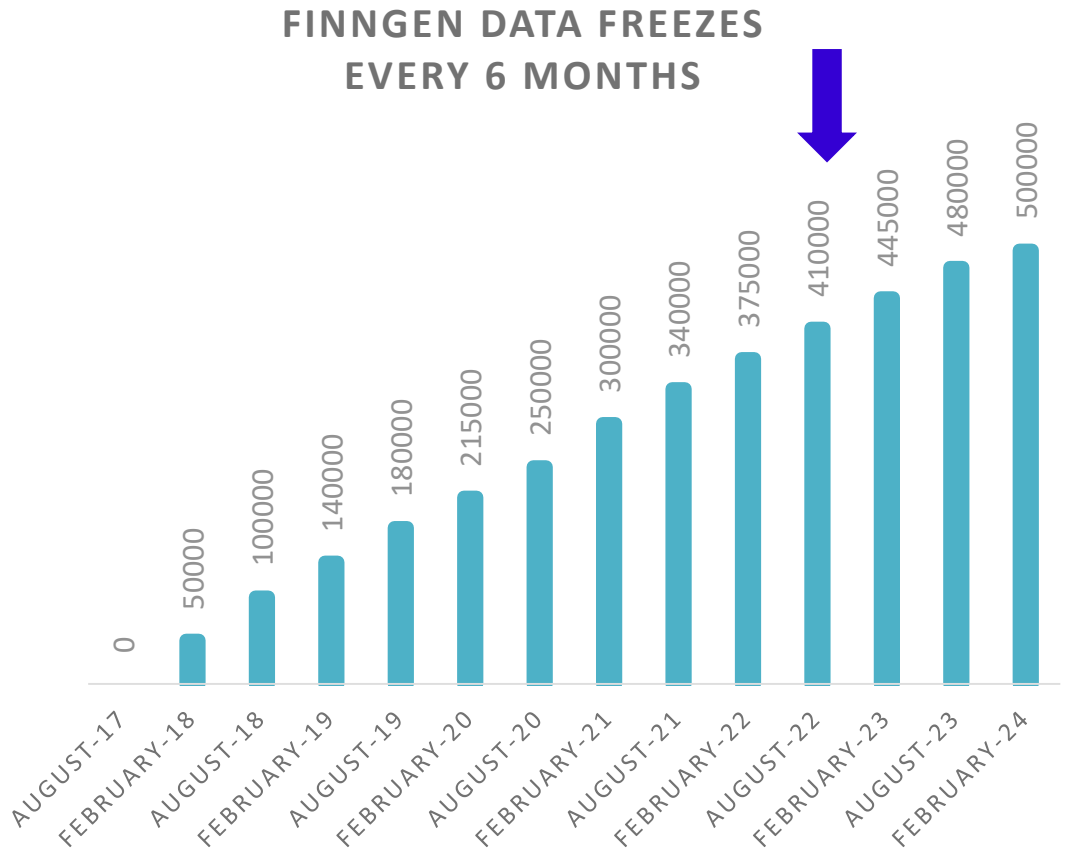
THL
TERVEYDEN JA HYVINVOINNIN LAITOS

New datasets produced every 6 mo

FINNGEN

Core analysis team deliverables

- GWAS and PheWas results of thousands of endpoints
- Finemapping of the GWAS hits
- Colocalization results of the GWAS hits
- Autoreporting (automatic annotation) of the GWAS results
- Meta-analyses (UKBB + Estonian BB)
- Variant annotations



Current dataset (Data Freeze 10):

- **429 209** pheno-geno data available
- 4519 disease endpoints
- 21.3 M variants

Providing an infrastructure for research

Tools to analyze data

FINNGEN Sandbox for Data Analysis

Approved users anywhere in world

FINNGEN

Industry:

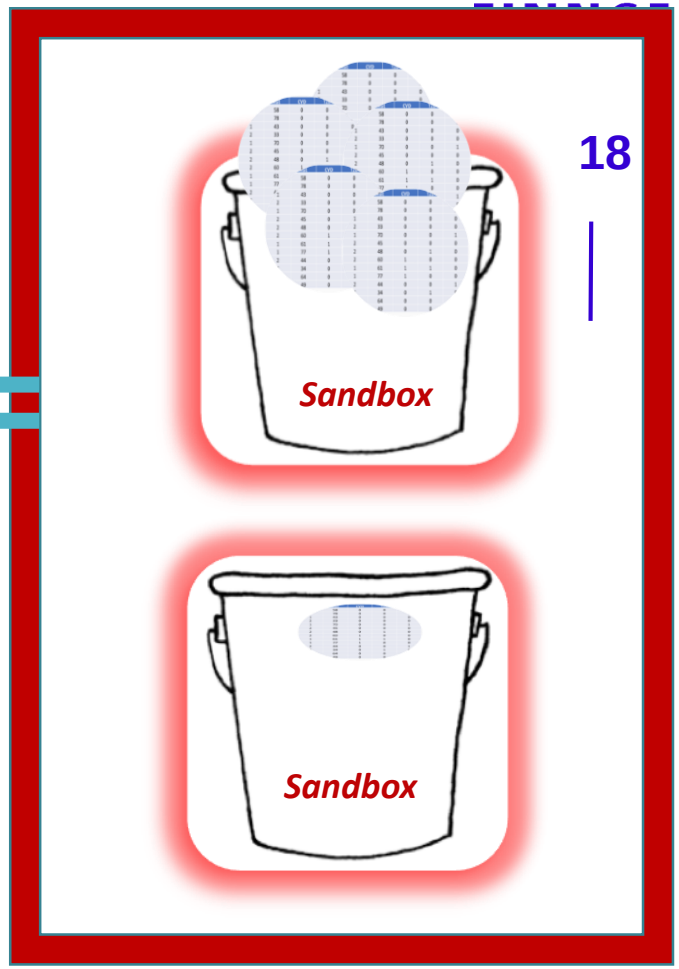
Abbvie Astra Zeneca
Biogen Celgene Pfizer
Genentech Merck/MSD
GSK Sanofi Janssen
Maze Novartis BI

Academic: UH, HUS, THL, Finnish Biobanks, Hospital Districts, Universities in Finland or abroad

finngen.fi
Google Cloud

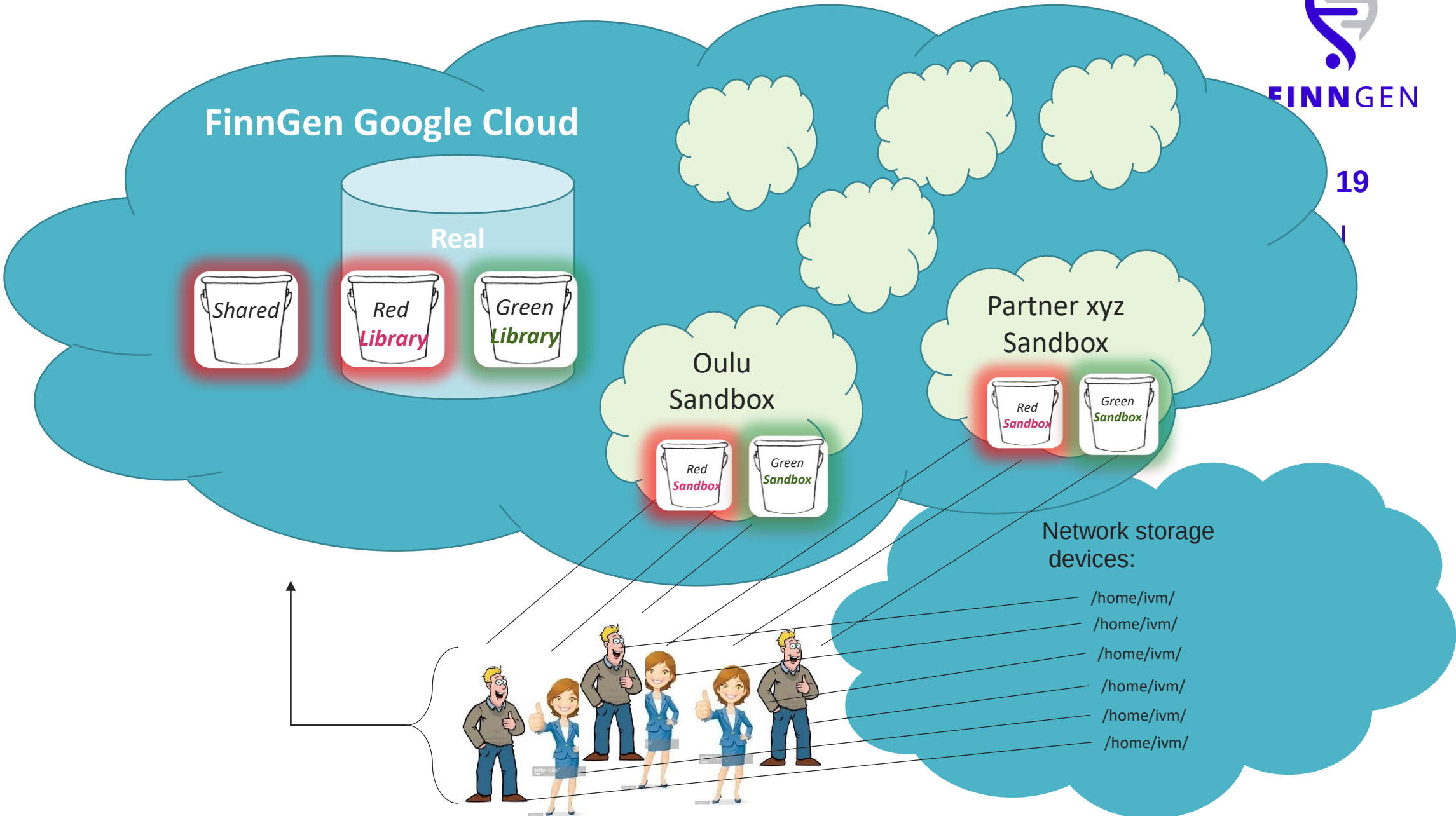
Individual and group workspaces
Summary and count level results for export, download, portals

standard or custom Tools



- **Secure access to individual level data**
- **Data cannot be copied**

Sandbox storage infrastructure



FinnGen Sandox has solved a problem that some have claimed unsolvable

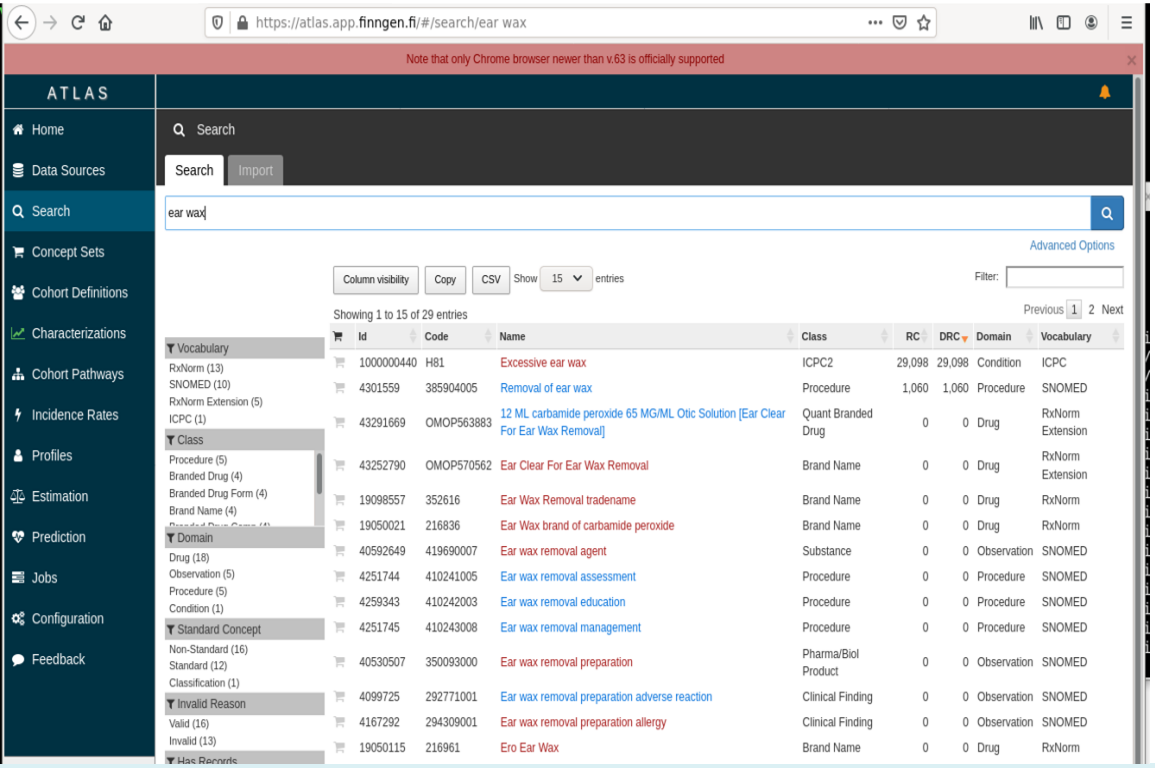
- We have developed an analysis environment for individual level data that can be accessed by researchers from outside EU
  - >150 useres outside EU
- It fullfils all national laws, regulations and the EU wide GDPR
- It provides an example and a potential solution for others
  - Already adopted for other projects (e.g., ELGH)
  - Has stimulated interest from several European countries
- Summary statistics downloaded by >2000 non-FinnGen scientists from all over the world

Providing an infrastructure for research

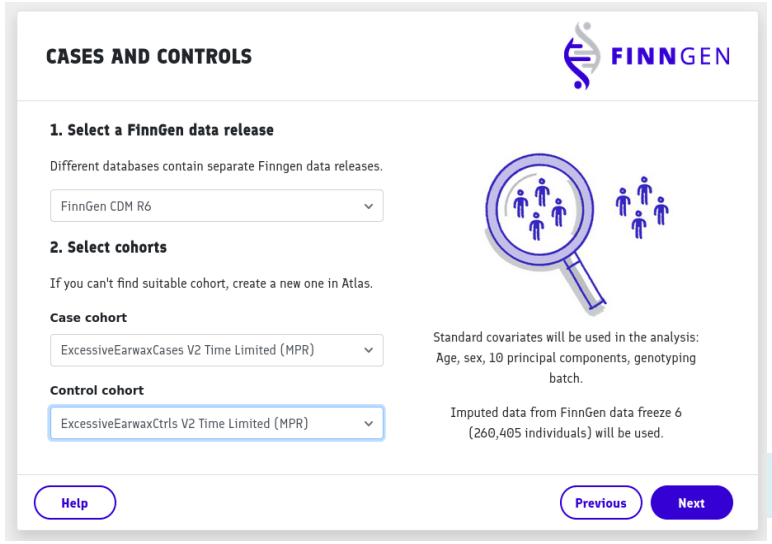
Tools to brows results

Streamlined custom phenotype definition and automated GWAS

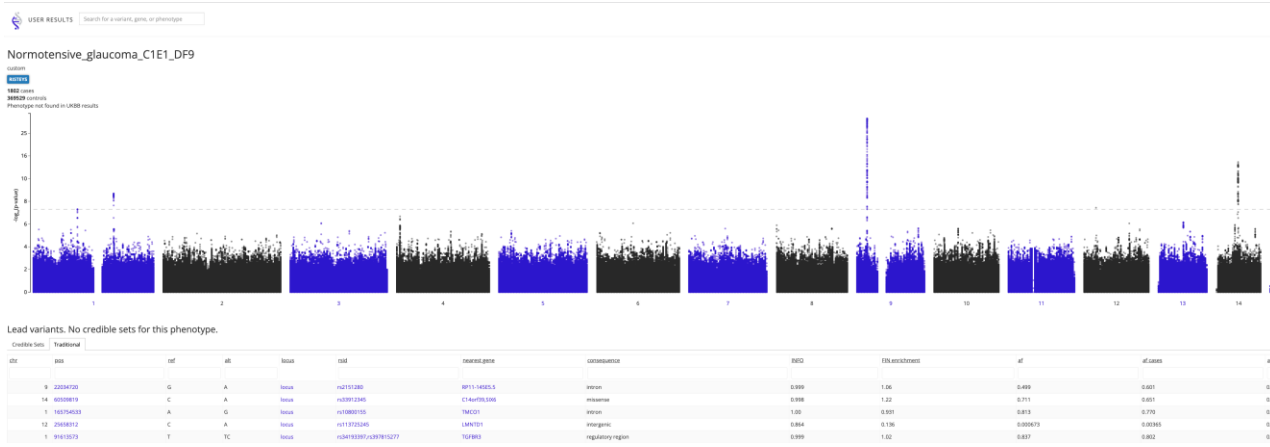
FINNGEN



1. ATLAS (OHDSI tk) for registry data exploration and custom cohort definition



2. Launch custom GWAS



3. Custom runs automatically appear in ‘userresults’ PheWeb


FINNGEN
FinnGen harvesting time

FinnGen - harvesting time

FINNGEN

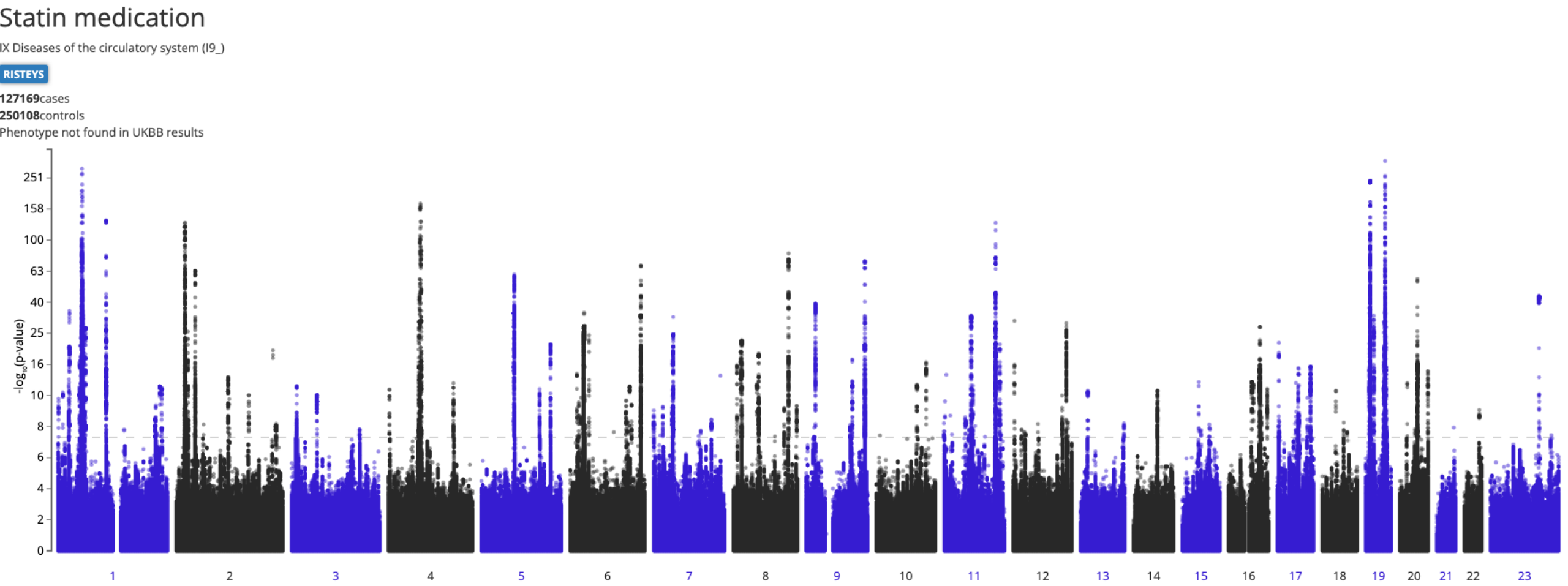
- Current data freeze (10) now has >429,200 Finns included with genome and >50 years of registry health data
- Hundreds of users throughout Finland and 13 subscribing pharma partners have individual level data access
- Public data releases (PheWeb through r7, downloads, OpenTargets)
  - 2256 full downloads of FinnGen results from around the world
  - FinnGen results pervasively used throughout the human genetics research community given their completely public and unencumbered access

A wealth of new discoveries

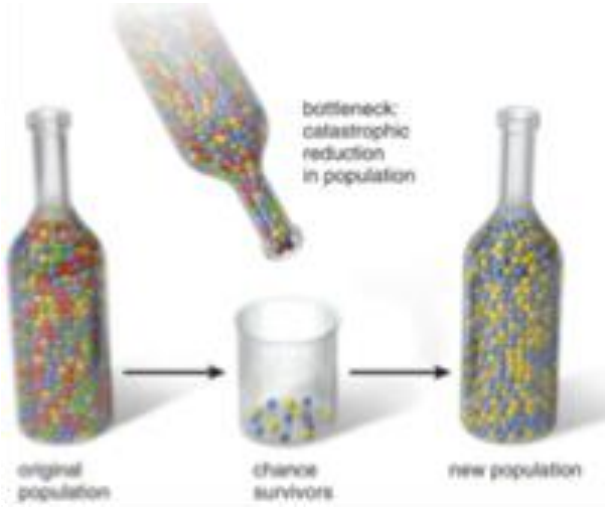
Current data freeze of 392 000 participants

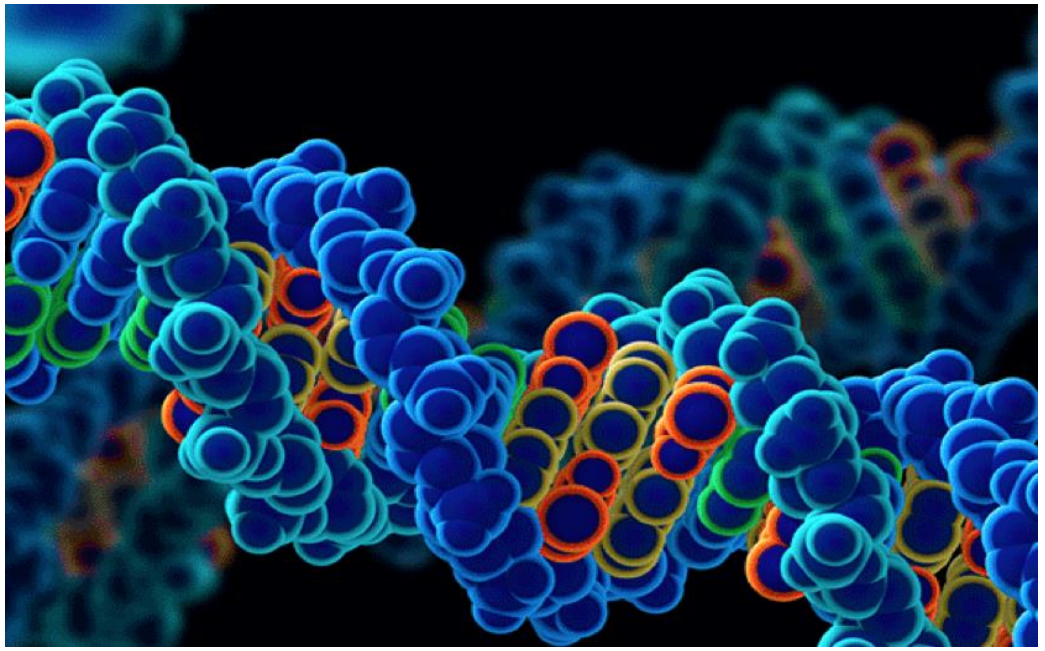
More than 400 Finnish specific gene associations

FINNGEN



Power of a Genetic Isolate





Variants enriched in Finland

New genes and new variants

New biological insight

gnomAD

genome aggregation database

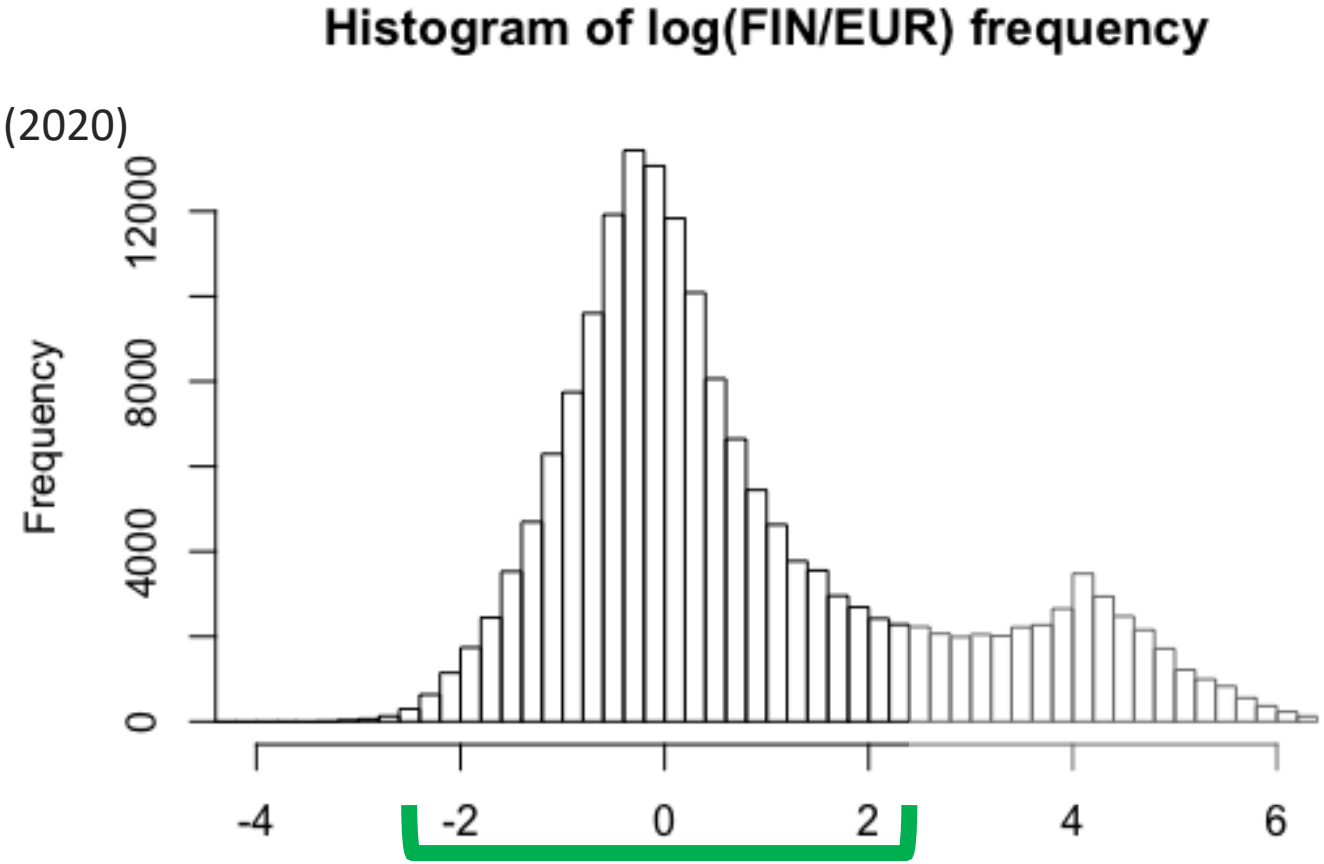
Plotted below: Alleles between 0.1% and 10%

Karczewski et al - *Nature* **volume 581**, pages434–443 (2020)

category	General European	Additional in FIN
pLOF	2070	829
missense	55250	15914
other disrupt.	3354	842

Konrad Karczewski

Founder and diverse populations have a huge and largely untapped potential for genetic discovery

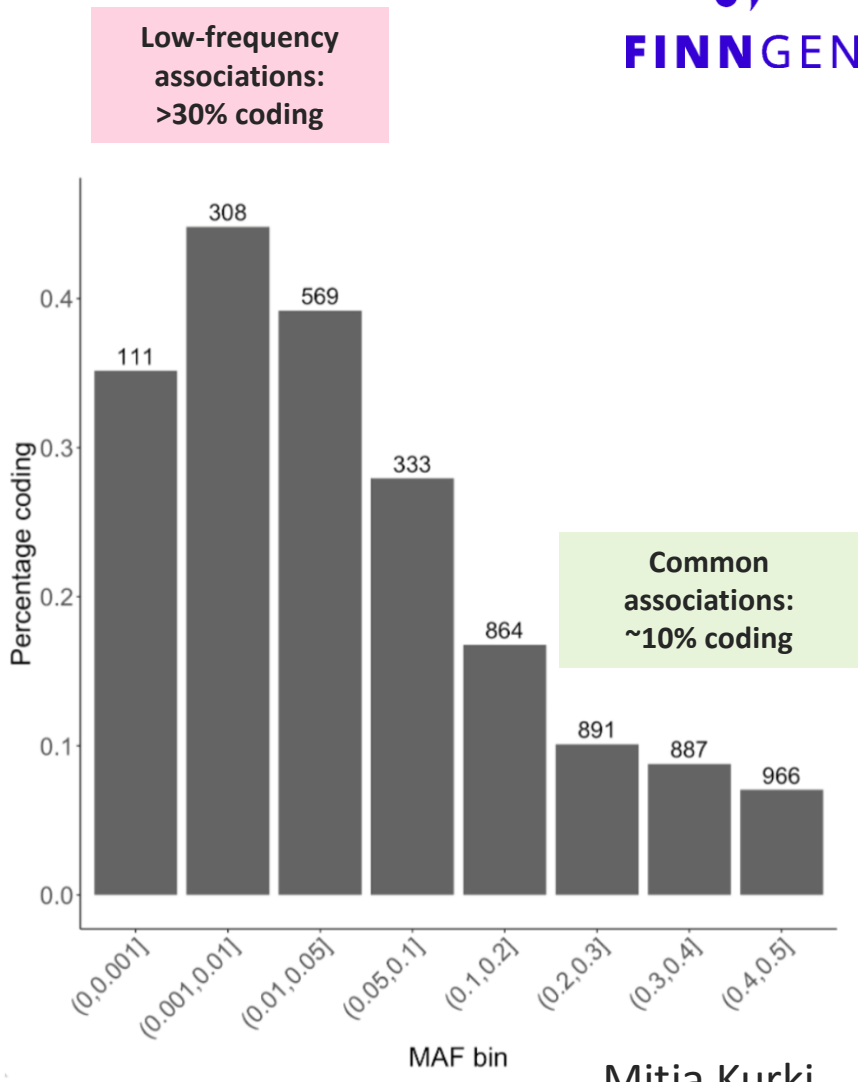


Normal variation – NFE and FIN frequencies within a factor of 10

Coding variant associations

- 11.8% significant associations contain a coding variant in credible set (PIP > 0.05)
- 538 unique coding variants across 497 genes
- More than half (299/538) have MAF < .05
- 2/3 of those are enriched in Finland by more than 2x frequency

- Signature of natural selection evident



Mitja Kurki

Example of a coding variant

- Finnish gene variant protects from glaucoma

= Finnish glaucoma gene association

Glaucoma

VII Diseases of the eye and adnexa (H7_)

RISTEYS

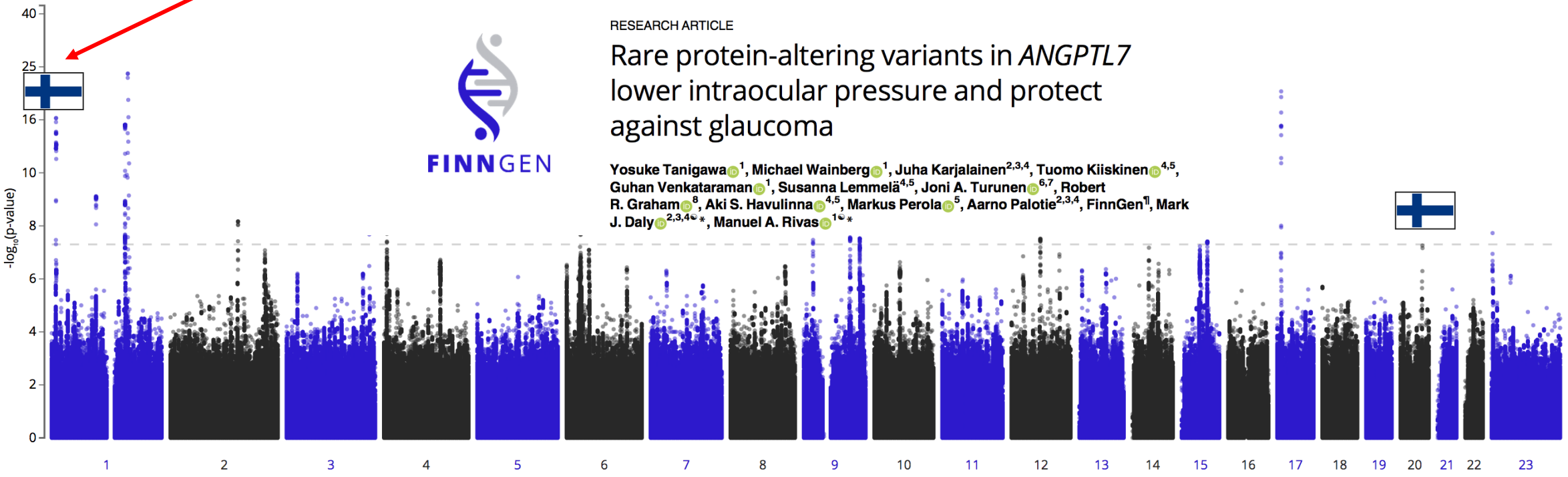
10485 cases
249920 controls
Phenotype not found in UKBB results

PLOS GENETICS

RESEARCH ARTICLE

Rare protein-altering variants in *ANGPTL7* lower intraocular pressure and protect against glaucoma

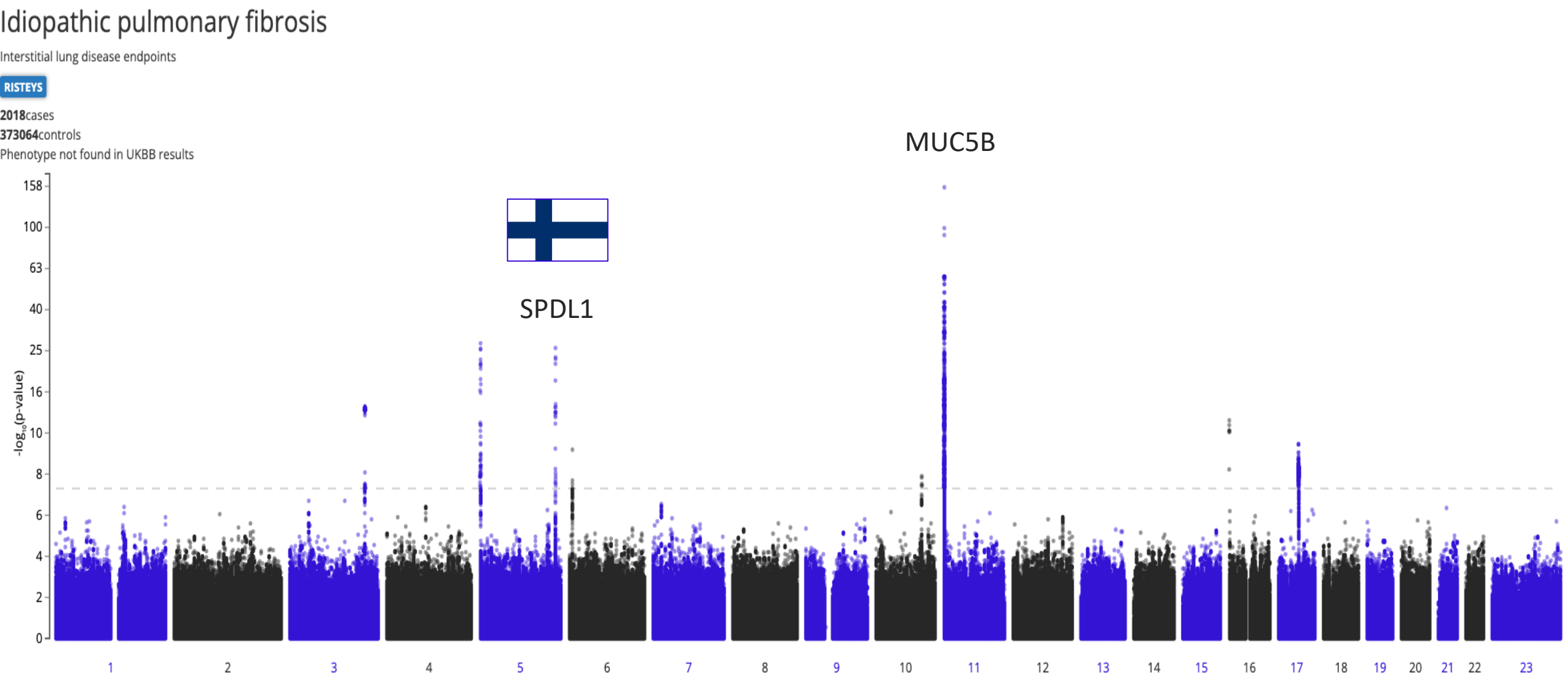
Yosuke Tanigawa[1], Michael Wainberg[1], Juha Karjalainen[2,3,4], Tuomo Kiiskinen[4,5], Guhan Venkataraman[1], Susanna Lemmelä[4,5], Joni A. Turunen[6,7], Robert R. Graham[8], Aki S. Havulinna[4,5], Markus Perola[5], Aarno Palotie[2,3,4], FinnGen¶, Mark J. Daly[2,3,4]*, Manuel A. Rivas[1]*



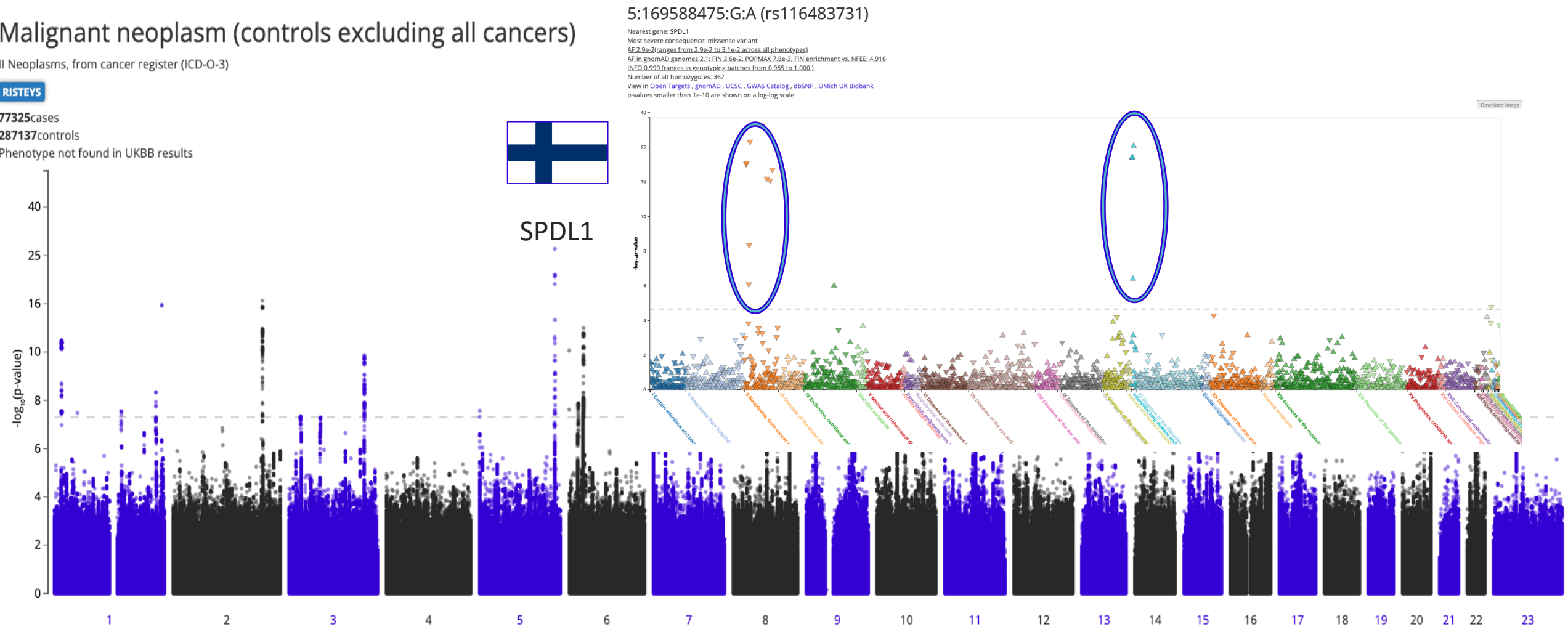
Pleiotropy: power of biobank studies

Same variant, different effect

Risk locus for IPF



Protective for “all cancers”



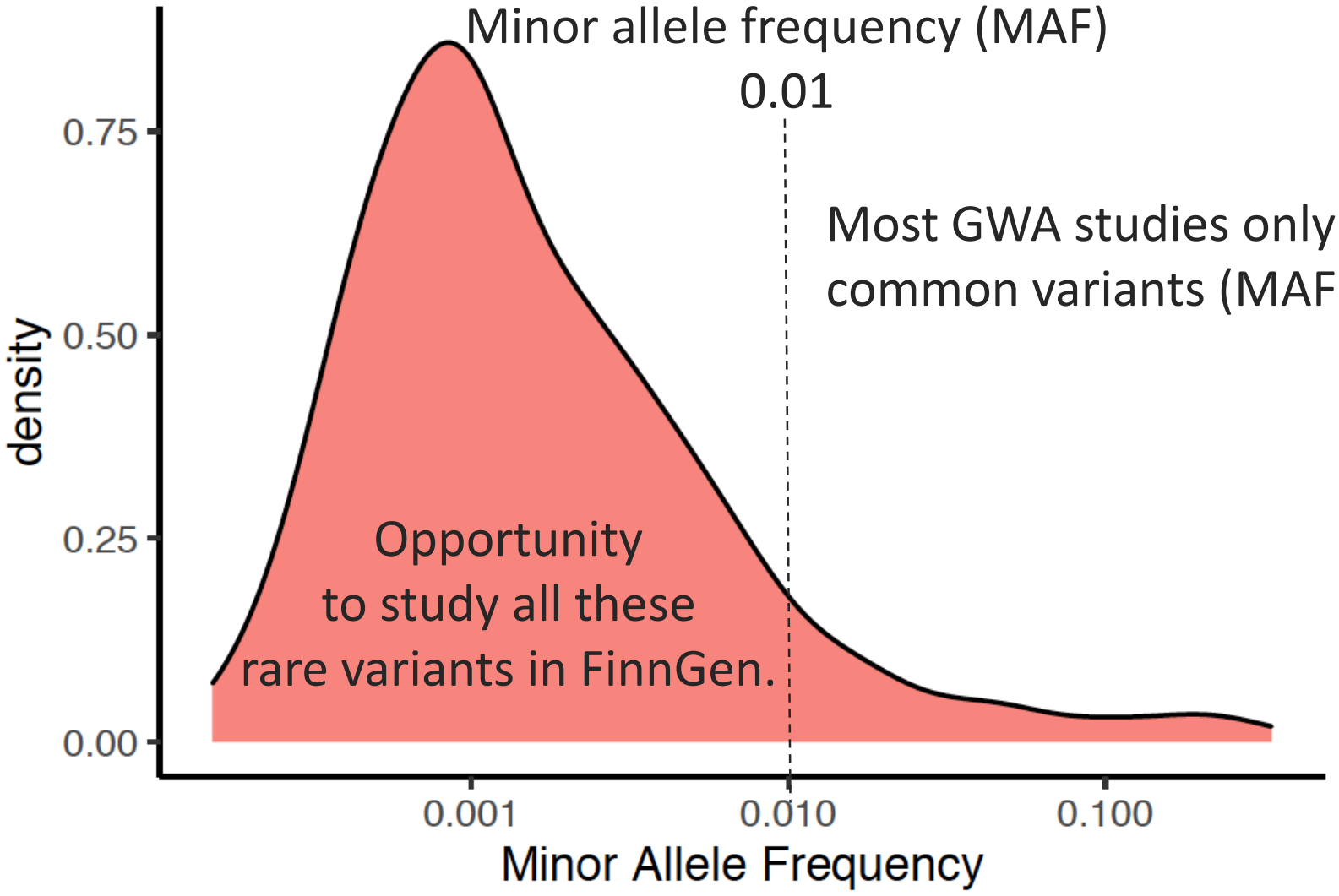
Finnish enriched variants Recessive??

Diseases considered recessive

Is that quite so?

FinnGen includes 608 known disease-causing variants (ClinVar database) - most are rare.

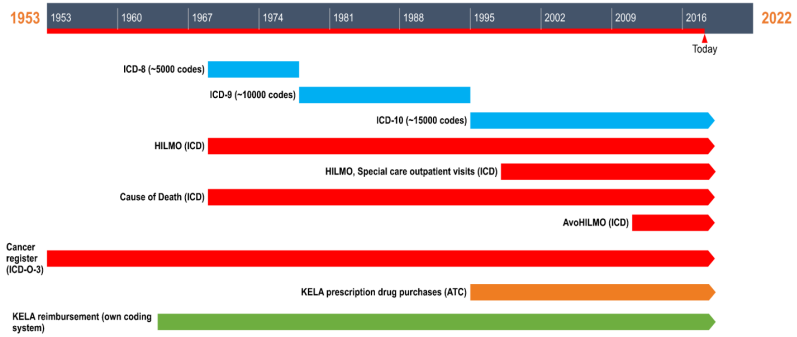
FINNGEN



Most GWA studies only investigate common variants (MAF > 0.01).

Henrike Heyne

Longitudinal analyses



3/5 recessive variants have effects in heterozygous state

- 5 known recessive disease variants have genomewide GWAS hits (p-value < $5x10^{-8}$)

  – 3/5 have effects in heterozygous state

    SERPINA1 → cholelithiasis, emphysema

    XPA → skin cancer

    NPHS1 → nephrotic syndrome

  – 2/5 no effects in heterozygous state → purely recessive effect?

    EYS → retinitis pigmentosa

    CLRN1 → retinitis pigmentosa

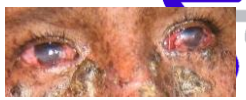
37

Known

Little evidence

Heyne, Nature, in press

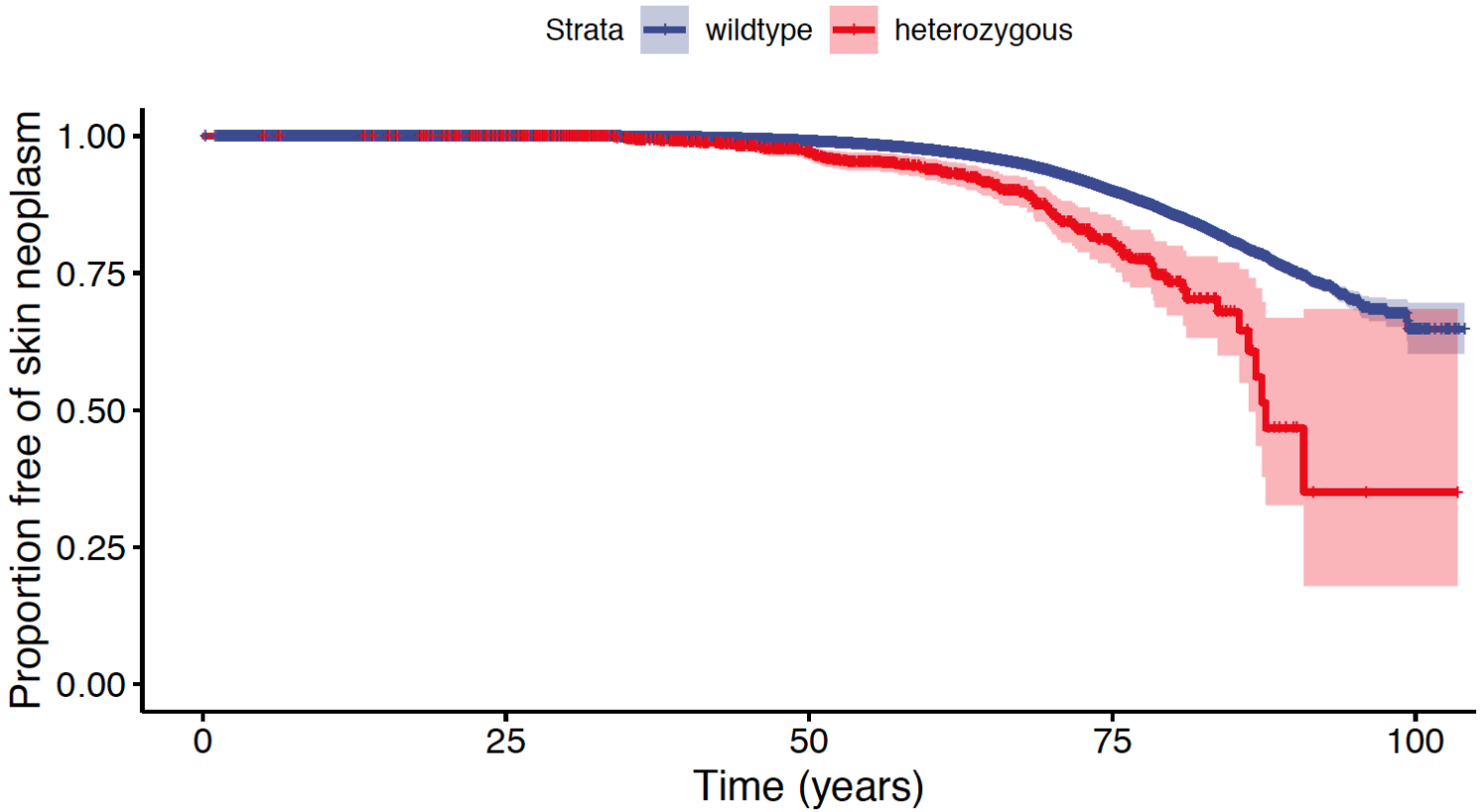
Known pathogenic PTV in *XPA*



FINNGEN

- two LOF mutations → **recessive** disease Xeroderma pigmentosum
- one LOF mutation → high risk for **skin cancer**

FinnGen: $p = 8x10^{-11}$, OR 3.5, in Japan: OR = 3.08, p = 0.01



Hirai Y et al, **J. Hum. Genet.** 2018

Heyne, Nature, in press

Opportunites provided by longitudinal data

Diseases susceptibility

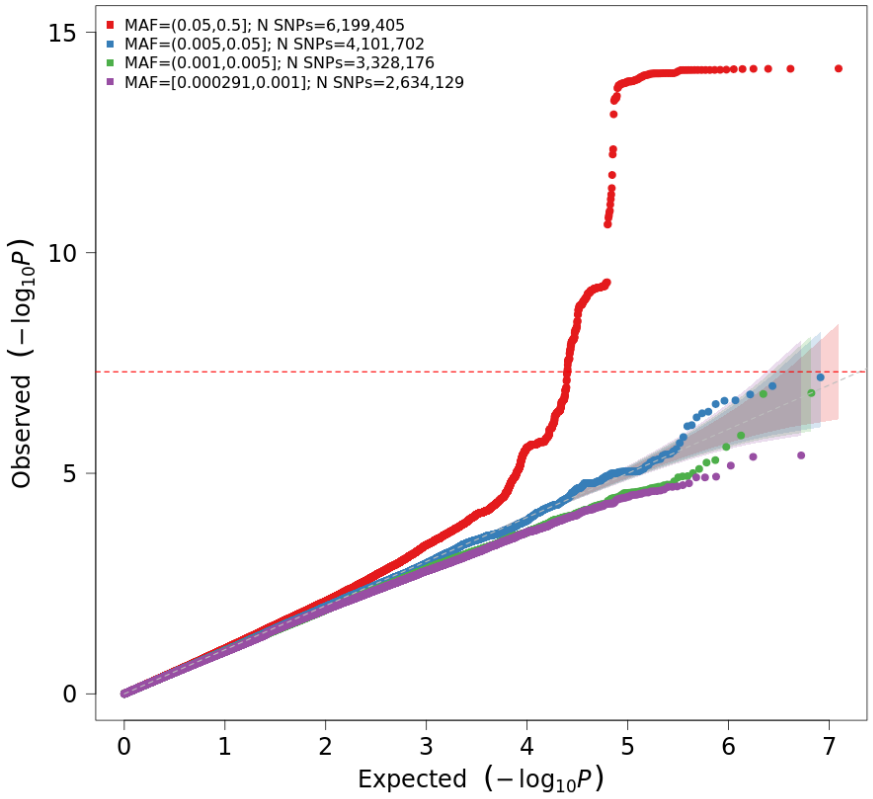
Disease progression

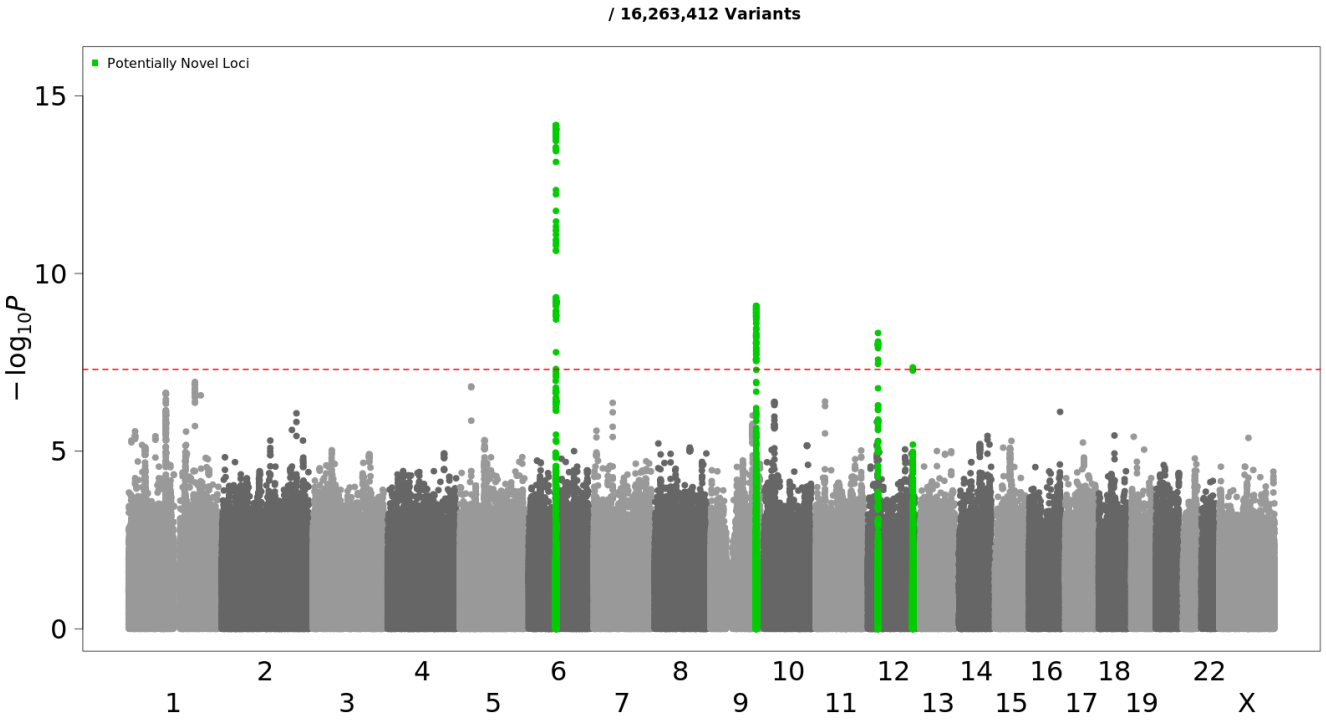
Disease Progression (Survival) Scans

Arthrosis -> Arthrosis + hip replacement surgery (6,825 failures/27,533 censored) – R5

Wei Zhou

40





An efficient and accurate frailty model approach for genome-wide survival association analysis controlling for population structure and relatedness in large-scale biobanks

https://www.biorxiv.org/content/10.1101/2020.10.31.358234v1.full

Rounak Dey, Wei Zhou, Tuomo Kiiskinen, Aki Havulinna, Amanda Elliott, Juha Karjalainen, Mitja Kurki, Ashley Qin, FinnGen, Seunggeun Lee, Aarno Palotie, Benjamin Neale, Mark Daly, Xihong Lin

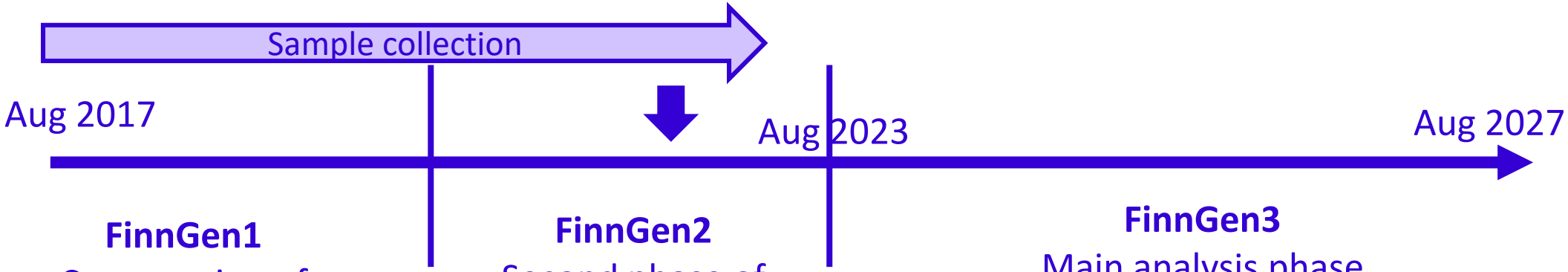
A centralized research resource can now be used by all Finnish researchers

Towards clinical application of risk assessment

So what is different in FinnGen

- Nationwide
- Special population
- Enrichment of disease cases
- Public-private partnership with an ”academic atmosphere”
  - 13 pharma working togeather
  - All national partners working together
- Strong core analysis group, providing browsable results

FinnGen, a 10 year project



DF6
270 000 individuals

DF12
500 000 individuals

FinnGen: moving beyond resource construction

FINNGEN

Characterize all the rare coding variants

study function & pleiotropy and learn therapeutic leverage points

Learn the genetics of disease progression and therapeutic response

lay the foundation for personalized medicine applications of genetics

From gene association to variant and function

- Having a list of associated genes is just the first step

- Next steps are much more challenging to scale:
  - From association to gene/variation
  - Which variation is causative
  - From variation to to understand the function
  - From function to potential clinical applications

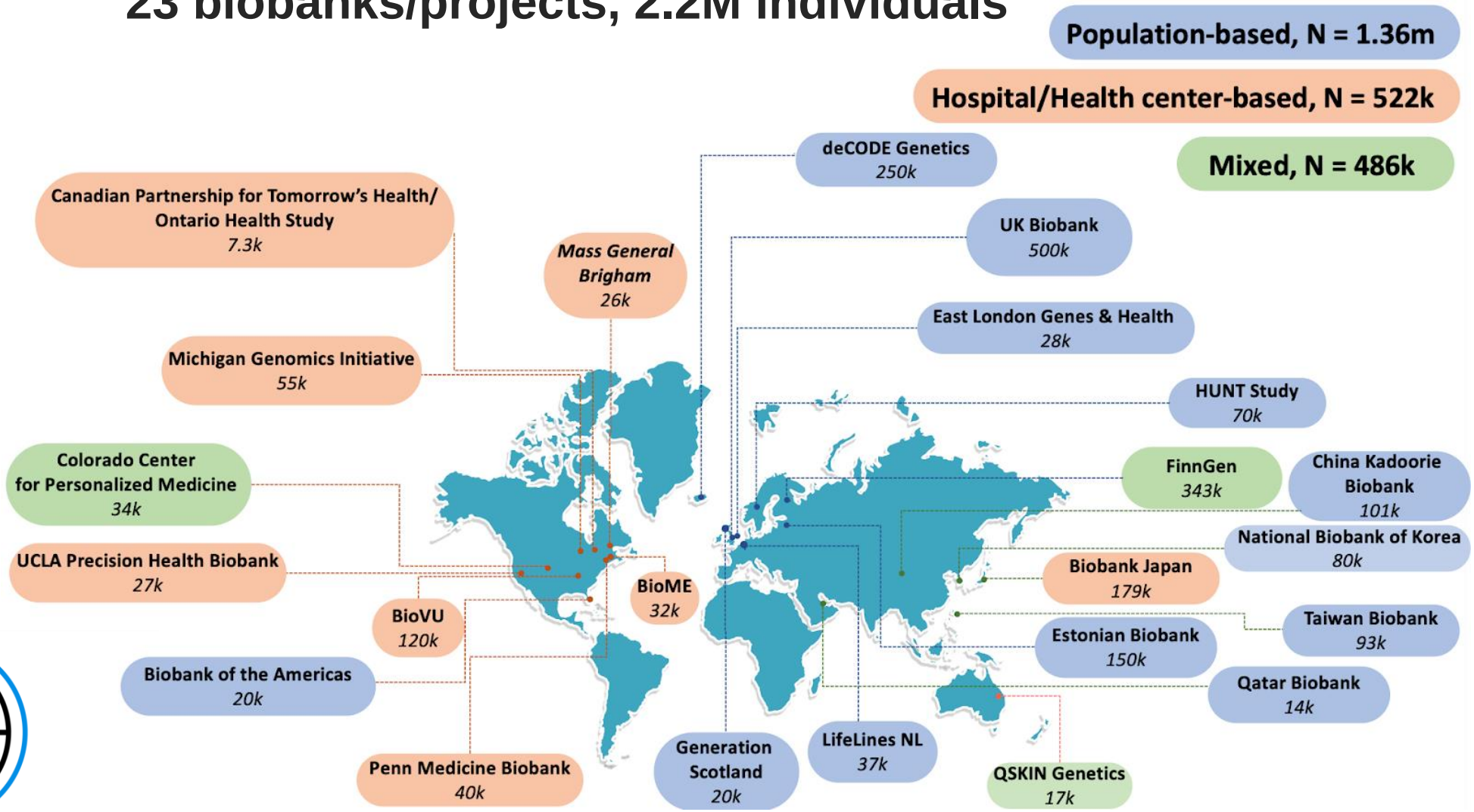
What do we need next in biobank studies

FINNGEN

- **Ethnic diversity**:
  - how to build evidence for potentially disease causative variants? Different variants in the same gene in different populations build confidence for causality
  - => collaboration across the globe
- **Recalling by genotype**:
  - Need to dive in to more specific and individual based phenotype data or samples for further characterization.
  - Needs to be rapid, and efficient, also cost efficient
  - Needs to meet the needs of the research projects (regulations or bureaucracy cannot guide what is the “right experiment”)

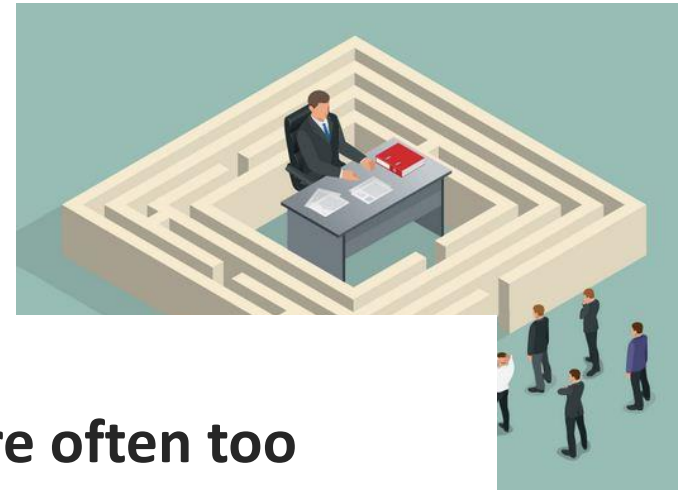
Global Biobank Meta-analysis Initiative (GBMI)

23 biobanks/projects, 2.2M individuals



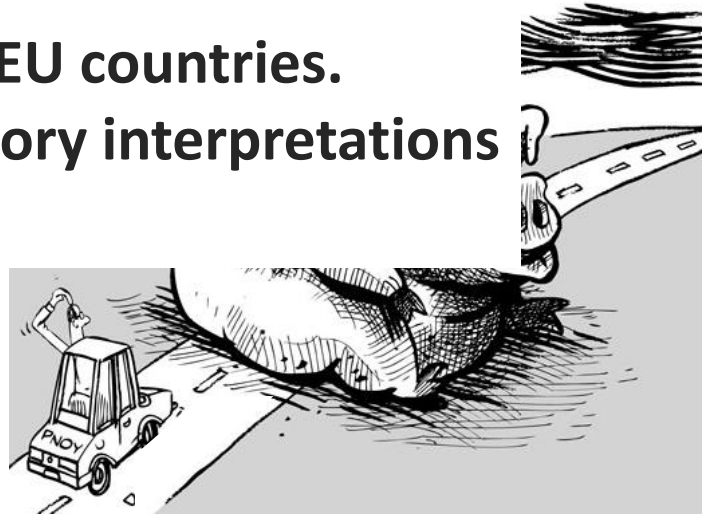
Global Biobank Meta-analysis Initiative

Zhou, W., Kanai, M., *et al*. *medRxiv* (2021)



FINNGEN

- I would like to argue that:
  - Regulations and discussions are often too narrowly focused on data protection and even that in a very narrow way.
  - Too much variation between EU countries.
  - Too much variation of regulatory interpretations between authorities



For biobanks to be viable also in future, important to:

FINNGEN

- Understand that biobanks are not, and should not, be money making machines
- Understand and acknowledge that biobanks are an **integral part of modern medical research infrastructure** that should have sufficient **core funding** to be viable
- Understand that biobanks are **to promote and help medical research**, not authorities to narrow research ideas
  - Attitude “yes, we should find a way how to do it”
- Understand that the most important people are those who provide biobanks with samples and data:
  - **Sample donators**
  - **Clinicians**

Nordic Society of Human Genetics and Precision Medicine

2022 CONFERENCE

3-4 NOVEMBER 2022

AT DOCKEN IN COPENHAGEN, DENMARK

WWW.NSHG-PM2022.ORG


Thank you!

NORDIC Biobank Conference

Gothenburg, Sweden

8–10 march 2022

Theme: Current trends and challenges in the Nordic countries

A unique opportunity to meet representatives in the area of human biobanking as well as other scientific experts from healthcare, academia, and industry.

Webbpage and registration: nbc.biobanksverige.se

The Nordic Biobank Conference 2022 is jointly organized by the Nordic countries, comprised of Denmark, Finland, Iceland, Norway and Sweden.