WEBINAR:

CanPath's genomic data landscape: Insights from Ontario and Quebec

Dr. Philip Awadalla
National Scientific Co-Director, *CanPath*
Executive Scientific Director, *Ontario Health Study*

Dr. Guillaume Lettre
Scientific Co-Director, *CARTaGENE*

Tuesday, March 11

11:00 am EDT

Dedicated to reconciliation with Indigenous Peoples

We acknowledge the traditional territories of the Mississauga of the New Credit First Nation, Anishnawbe, Wendat, Huron, and Haudenosaunee Indigenous Peoples on which the CanPath National Coordinating Centre at the University of Toronto now stands. The territory was the subject of the Dish With One Spoon Wampum Belt Covenant, an agreement between the Iroquois Confederacy and Confederacy of the Ojibwe and allied nations to peaceably share and care for the resources around the Great Lakes.

With CanPath teams working across Canada from coast to coast, we also acknowledge the ancestral territories that are home to many Indigenous people from across Turtle Island, and we are grateful to have the opportunity to work on this land.

Speakers

Moderator

Dr. John McLaughlin
CanPath Executive Director (2018-2023)

Presenter

Dr. Philip Awadalla
CanPath National Co-Scientific Director & OHS Executive Scientific Director

Presenter

Dr. Guillaume Lettre
CARTaGENE Co-Scientific Director

CanPath's Genomic Data Landscape: Insights from Ontario and Quebec

Philip Awadalla
Nuffield Department of Population Health, University of Oxford
Department of Molecular Genetics, University of Toronto

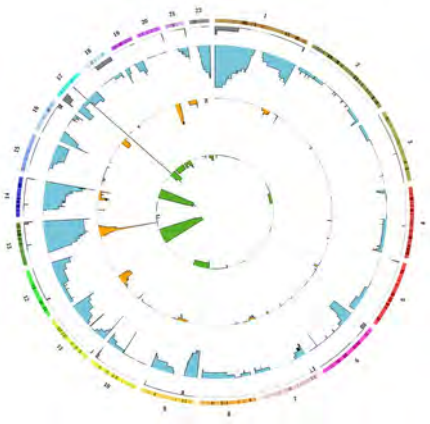
CanPath Webinar March 11, 2025

CanPath
Canadian Partnership for Tomorrow's Health

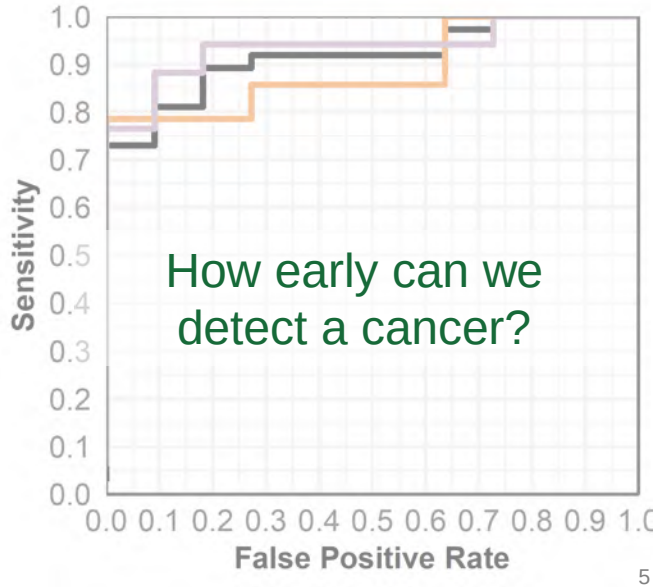
Overview

CanPath
Canadian Partnership for Tomorrow's Health

Mutation accumulation in blood with age



Mechanisms of healthy blood aging



AWADALLA LAB

Pioneering Genomics for Precision Health.

Phillip Awadalla

Vanessa Bruat

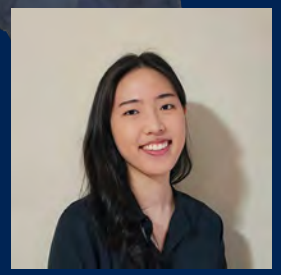
Mawussé Agbessi

Marie-Julie Favé

Elias Gheba

Kimberly Skead

Nicholas Cheng



Jasmine Kang

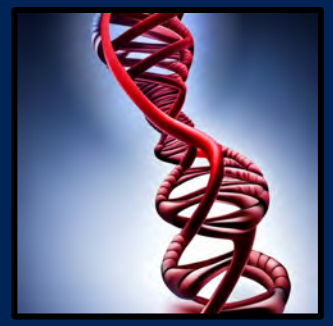
Ido Nofech-Mozes

Tom Ouellette

Yiran Shao

Zixuan Lan

Omics' & Seq

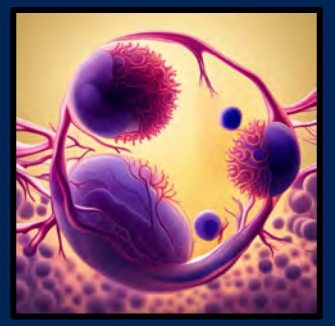


Blood & cfDNA

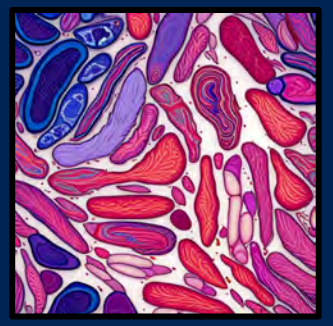
Environment

Evolution

Cells

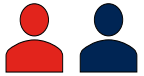


Imaging



Population cohorts or laboratories unlocks potential to improve health

1 in 2 Canadians will die from **cancer or a chronic disease**



1 in 2 Canadians will be diagnosed with **cancer**

1 in 10 Canadians live with **asthma or COPD**

1 in 12 Canadians are with diagnosed with **heart disease**

Questions that can be answered:

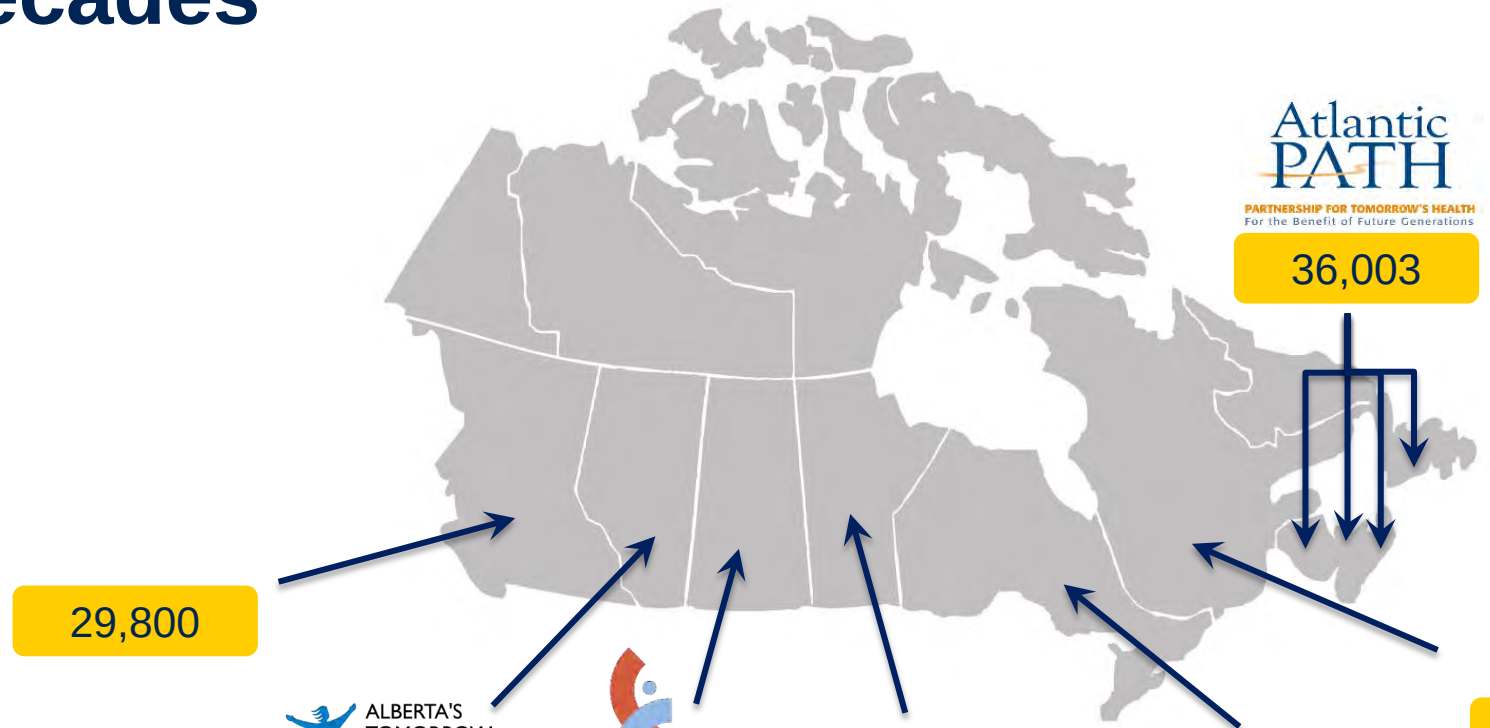
- How do we address the root causes of health and disease in the population?
- What can we do to improve our health?
- What can we do together to build healthier communities? Impact of environment?
- Can cancer and other serious illnesses be detected years earlier?
- How do we build learning health systems that improve outcomes?

*Manolio et al, Nature Reviews Genetics 2006 (re: value of prospective cohorts).

CanPath

CanPath is following the health of over 360,000 adult Canadians for decades

- Questionnaire Data
- Data Linkage
- Long-term Follow Up
- Biological Samples
- Physical Measures
- Genomics



CanPath

National Leadership Team

Dr. Philip Awadalla
National Scientific Director, CanPath; Executive Scientific Director, Ontario Health Study

Dr. Jennifer Brooks
Executive Director, CanPath

Dr. Trevor Dummer
National Scientific Co-Director, CanPath

Dr. Parveen Bhatti
Scientific Director, BC Generations Project

Dr. Jennifer Vena
Scientific Director, Alberta's Tomorrow Project

Ms. Shandra Harman
Strategic Director, Alberta's Tomorrow Project

Mr. David Tran
Interim Scientific Director, Healthy Future Sask

Dr. Donna Turner
Scientific Director, Manitoba Tomorrow Project

Dr. Simon Gravel
Scientific Co-Director, CARTaGENE

Dr. Vikki Ho
Scientific Co-Director, CARTaGENE

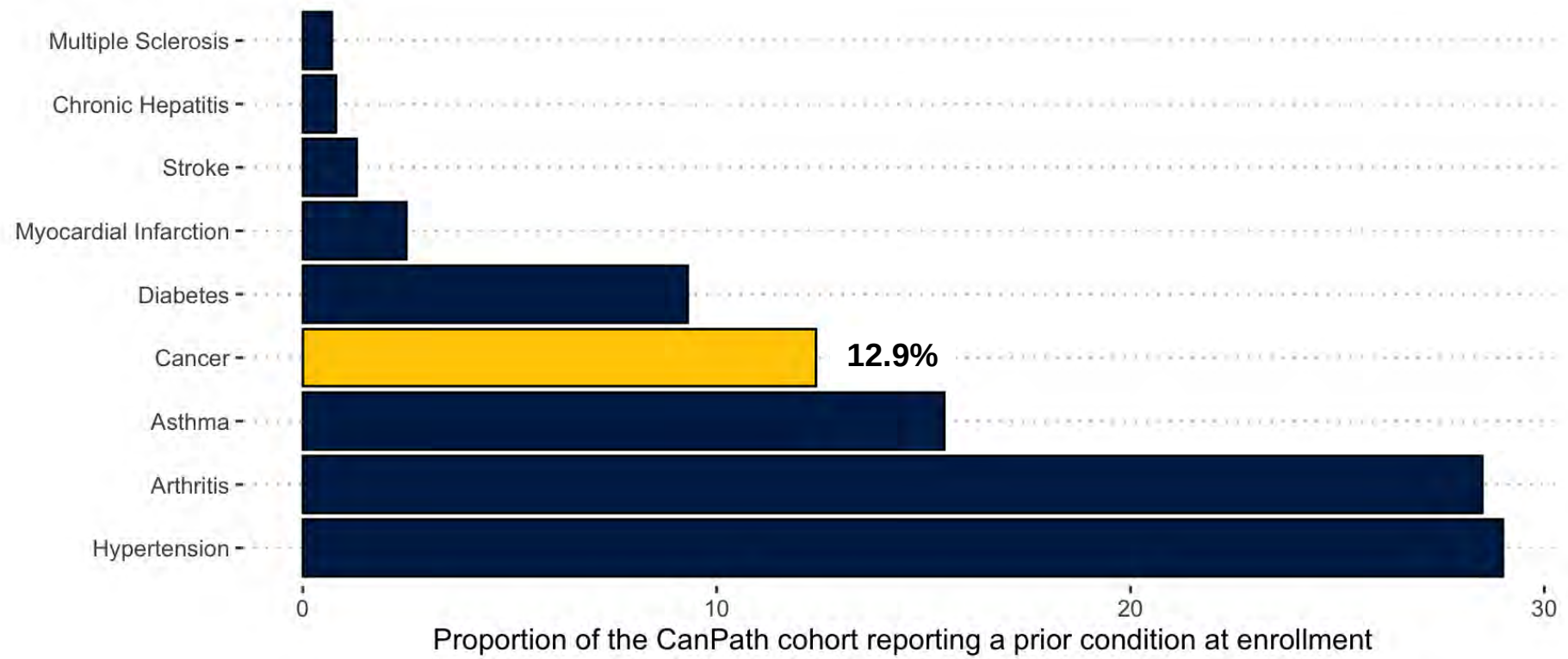
Dr. Guillaume Lettre
Scientific Co-Director, CARTaGENE

Dr. Robin Urquhart
Scientific Director, Atlantic PATH

Mr. Jason Hicks
Executive Director, Atlantic PATH

CanPath equips researchers to understand the causes of cancer development and progression

- Over **one in ten CanPath participants report a history of cancer** at enrollment



Enabling research breakthroughs to improve the health of Canadians

- CanPath enables research across health domains to improve disease **prevention, detection, treatment and health services**
- CanPath data and biological samples are available to researchers to study **a wide range of exposures (environment, lifestyle, etc.)** and **outcomes (common chronic disease, rare disease, infectious disease, etc.)**
- The longitudinal nature of CanPath enable scientists to perform health-related research **today and for years to come**
- **CanPath enables a healthier Canada** by building and hosting harmonized national self-reported health data alongside linked administrative health data

CanPath

Solving the Canadian Problem:

- **Canada has some of the most comprehensive healthcare datasets in the world, but…**
- **Linking and sharing data across jurisdictions is challenging** and represents a major research and public health limitation.
- **Accessing data is cumbersome, costly, and there are major barriers around where data can reside**; all of which limit the extent to which data can be utilised.

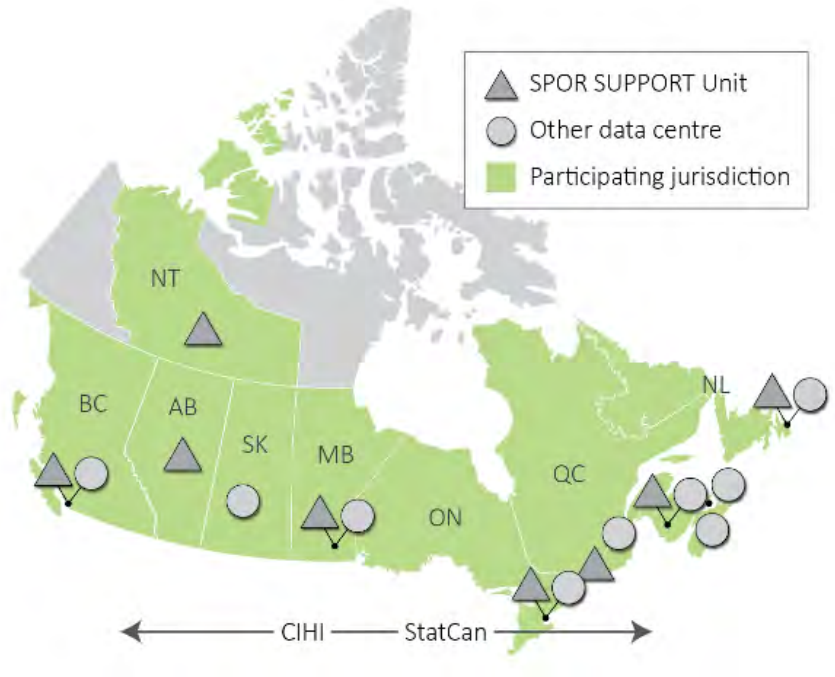
Key Partnership:

Health Data Research Network Canada

CanPath was **Health Data Research Network Canada (HDRN Canada)'s first external partner.**

CanPath and HDRN Canada partnered to facilitate multi-jurisdictional linkage between CanPath cohorts and regional data holders.

This will allow researchers and policy/decision makers to use linked and linkable administrative (real-world) data holdings for multi-province studies and initiatives.



CanPath will be the first Canadian cohort to host national cohort data and administrative data at a central location

- Linkages between the CanPath cohort and the Canadian Institute for Health Information (CIHI) administrative health data are underway.
- Individual-level linked CIHI data (N=290,000) will be hosted alongside the harmonized national CanPath dataset and made available to approved researchers through the trusted research environment
- CanPath will be the first Canadian program to be able to combine the wealth of cohort resources with national administrative level data in a central/national location

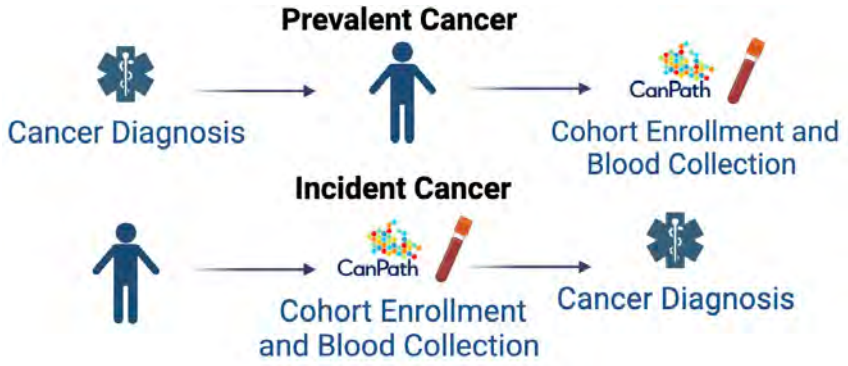
ED visits | Day Surgery | Outpatient Clinics | Inpatient visits | Physician Billing | Medications

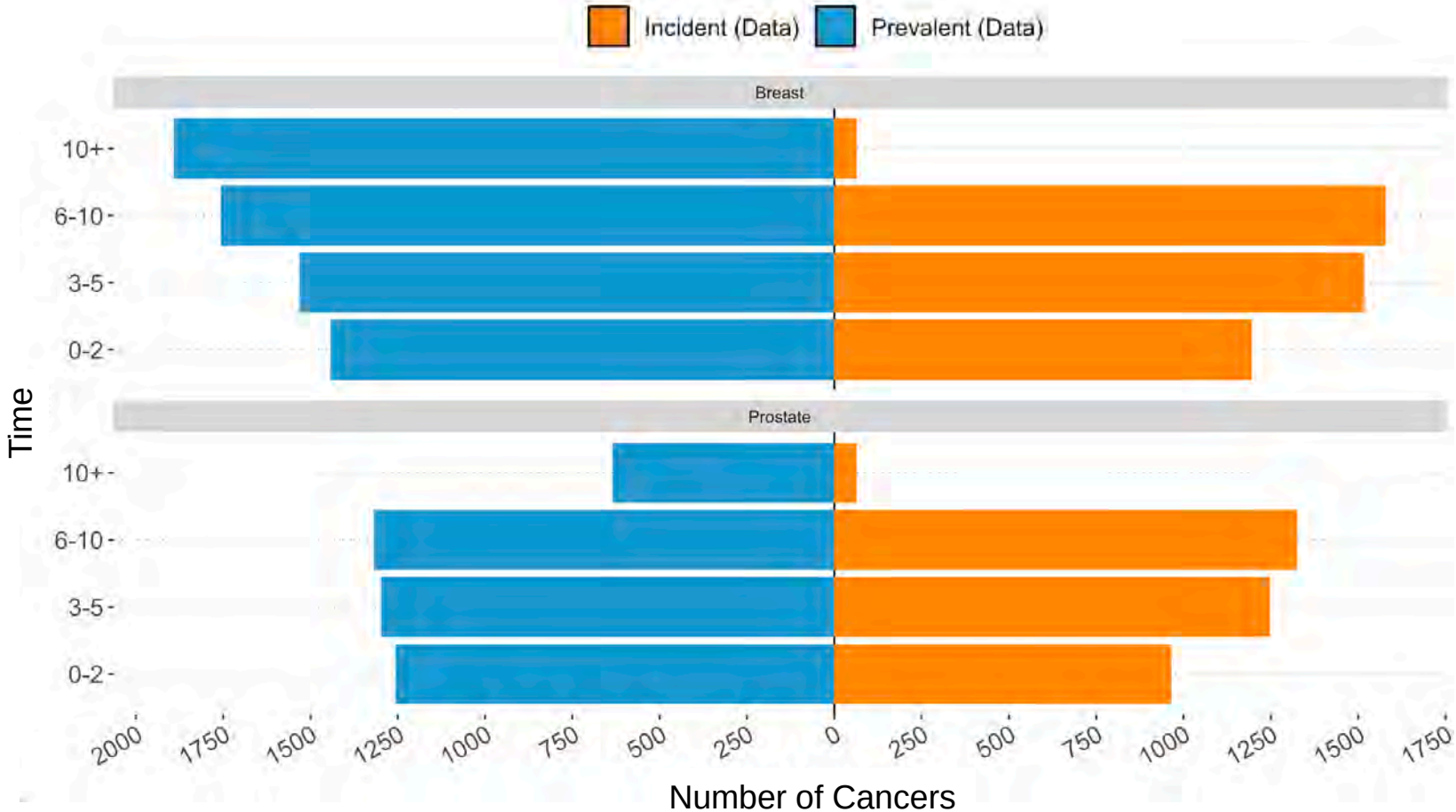
www.cihi.ca
At the heart of data

Canadian Institute for Health Information
Institut canadien d'information sur la santé

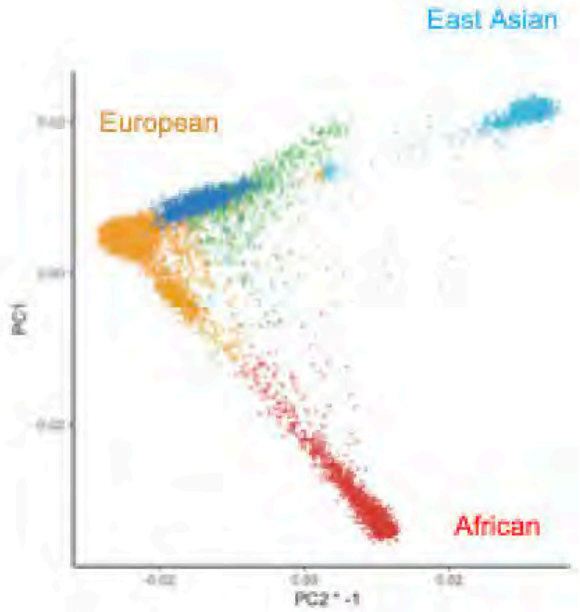
CanPath

Linkages enable us to map the time between participant enrollment in CanPath and cancer development





Genotyping of CanPath



Ontario Health Study
~20,300 participants

Québec Cartagene cohort
~ 30,000 participants

Majority of European descent (~92%)

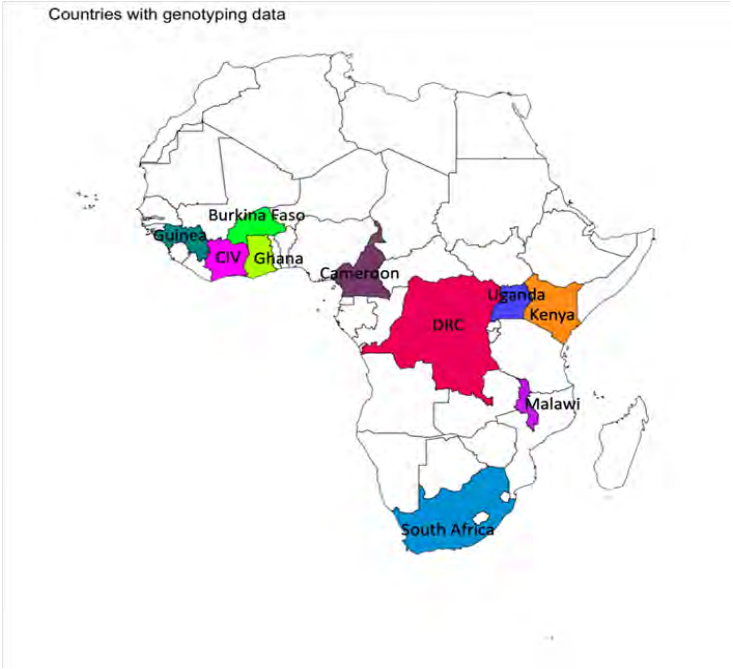
Awadalla, P, et al. (2013)
Kirsh, V, et al. (2017)

Projected on 1000Genomes

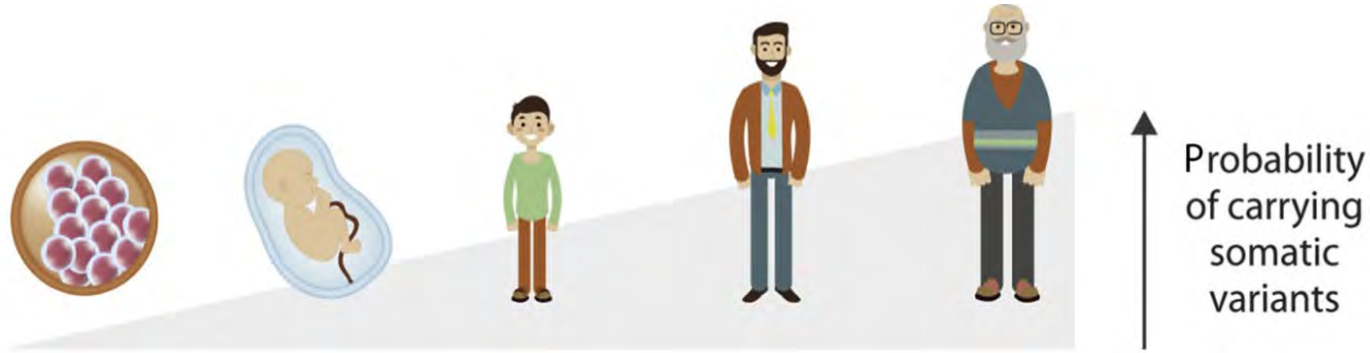
PRAIRIEGEN: A MULTI-OMICS APPROACH TO ADVANCING DATA INTEGRATION FROM MANITOBA AND SASKATCHEWAN POPULATIONS INTO THE PAN-CANADIAN GENOME LIBRARY

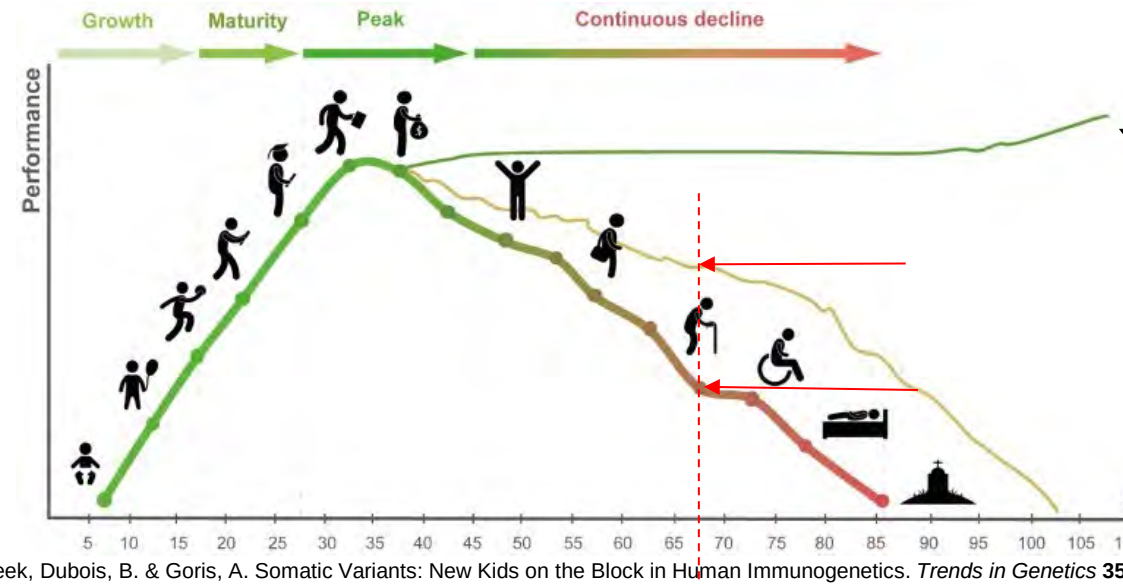
Mechanisms of aging in blood using population cohorts and single-cell -omics

Mechanisms of aging in blood using population cohorts and single-cell -omics



Somatic evolution varies in aging



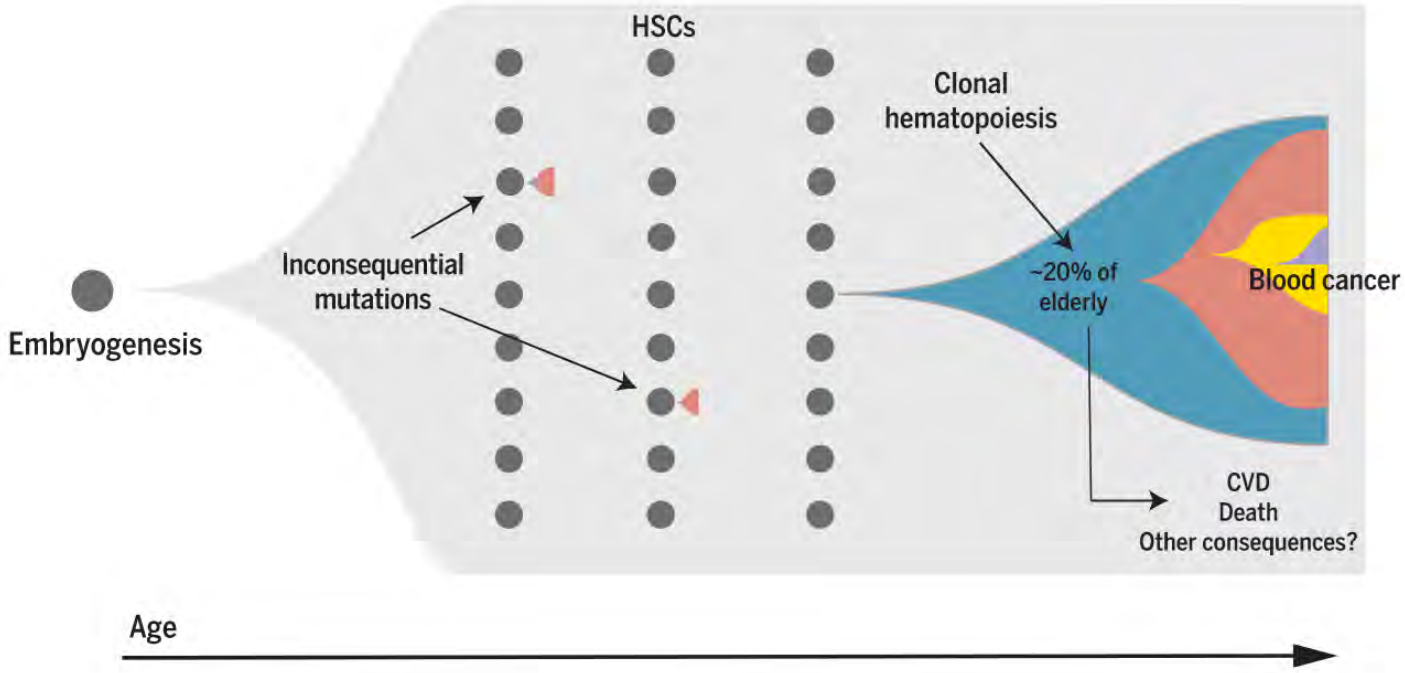


Healthy aging population

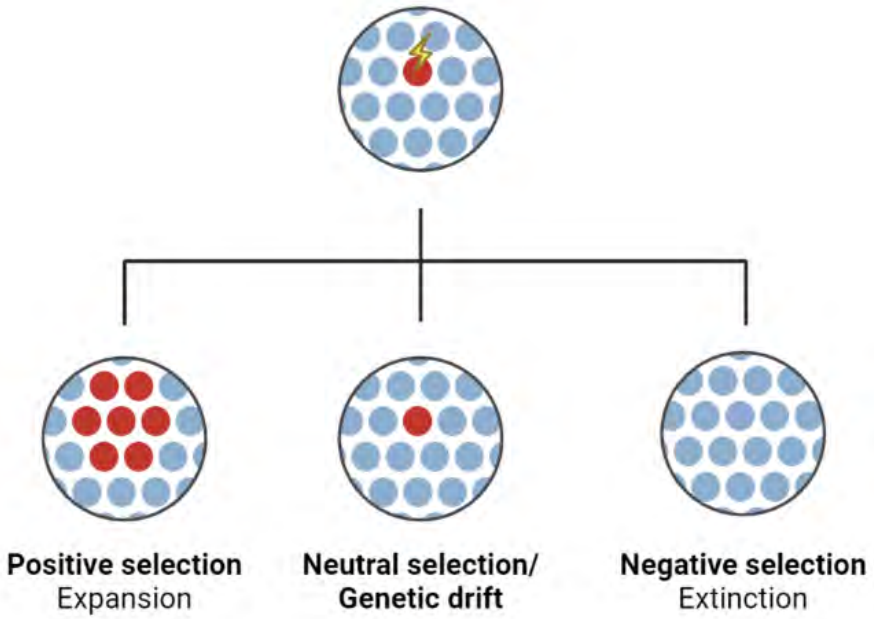
- Maintain optimal physiological function
- Delayed onset of age-related illness

Lies Van Horebeek, Dubois, B. & Goris, A. Somatic Variants: New Kids on the Block in Human Immunogenetics. *Trends in Genetics* **35**, 935–947 (2019).
Zhavoronkov, A., Li, R., Ma, C. & Mamoshina, P. Deep biomarkers of aging and longevity: from research to applications. *Aging* 11, 10771–10780 (2019).

Somatic evolution in blood and aging



Jaiswal, S. & Ebert, B. L. Clonal hematopoiesis in human aging and disease. *Science* 366, eaan4673 (2019).



Modified from: Kapadia, C. D. & Goodell, M. A. Tissue mosaicism following stem cell aging: blood as an exemplar. *Nature Aging* 4, 295–308 (2024).

Somatic mutations accumulate in our blood over time

- Blood cell hierarchy derived from **population of stem cells** (HSCs)
- HSC populations are very tightly regulated
- **Age-Related Clonal Hematopoiesis:** the preferential expansion of blood cells that carry recurrent somatic mutations
- ARCH almost inevitable in elderly
- Increased risk of **cancers and cardiovascular disease**

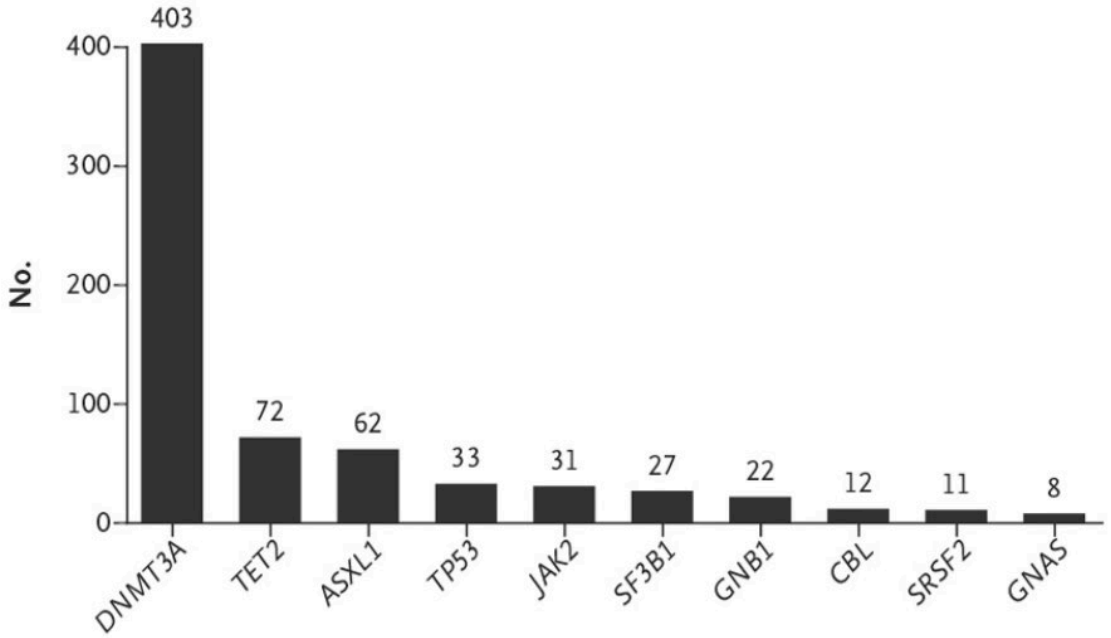


Figure from Jaiswal et al. NEJM (2014), Jaiswal et al. 2014, Genovese et al. 2014, Xie et al. 2014, Loh et al. 2018

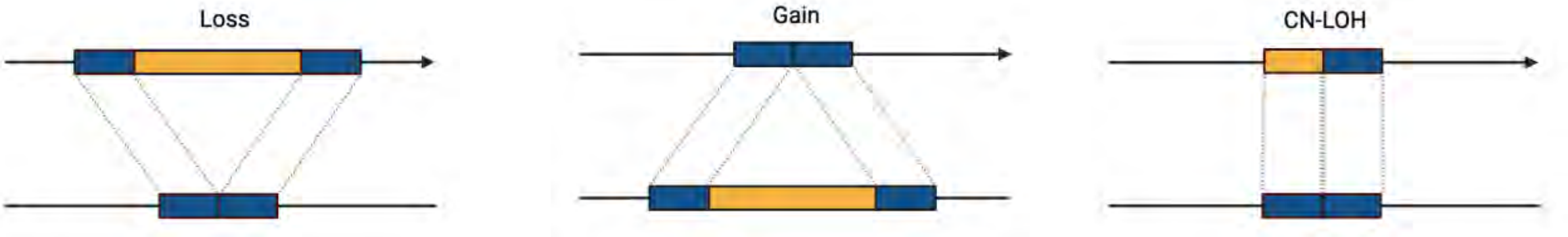
Why are large mutations tolerated in our blood?

Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations

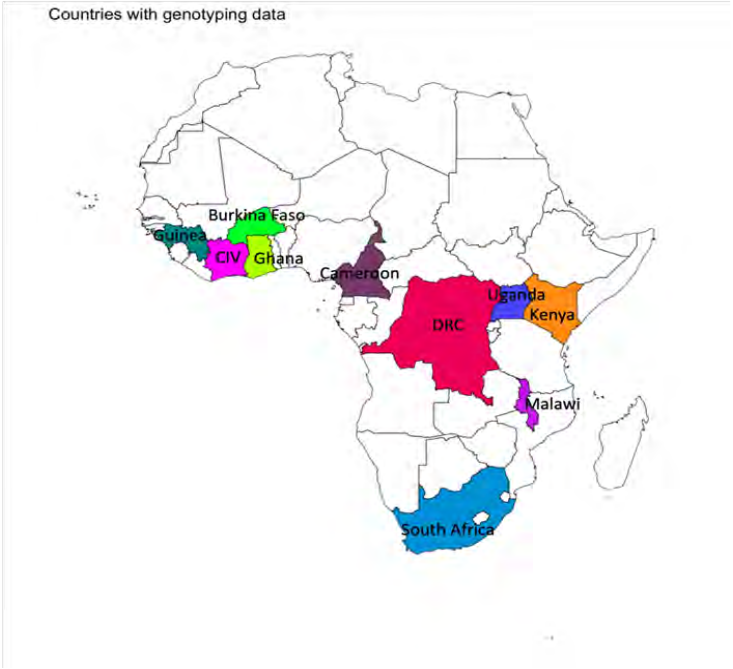
Po-Ru Loh, Giulio Genovese, Robert E. Handsaker, Hilary K. Finucane, Yakir A. Reshef, Pier Francesco Palamara, Brenda M. Birmann, Michael E. Talkowski, Samuel F. Bakhoum, Steven A. McCarroll & Alkes L. Price

Nature **559**, 350–355 (2018) | Cite this article

- Mosaic chromosomal alterations (mCAs) were found in approximately **5% of the population**
- **Is selection is playing a role in maintaining somatic mutations in blood, why are large mCAs tolerated?**



Mechanisms of aging in blood using population cohorts and single-cell -omics



Population cohorts for mCAs

CanPath
Canadian Partnership for Tomorrow's Health

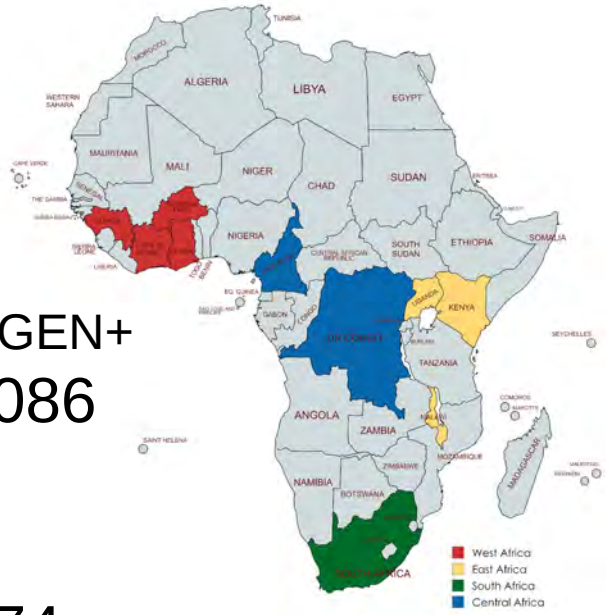
>360,000 samples

(European-ancestry) Ontario Health Study
n = 15,227

(European-ancestry) CARTaGENE
n = 25,843

AWI-Gen
n = 9,925

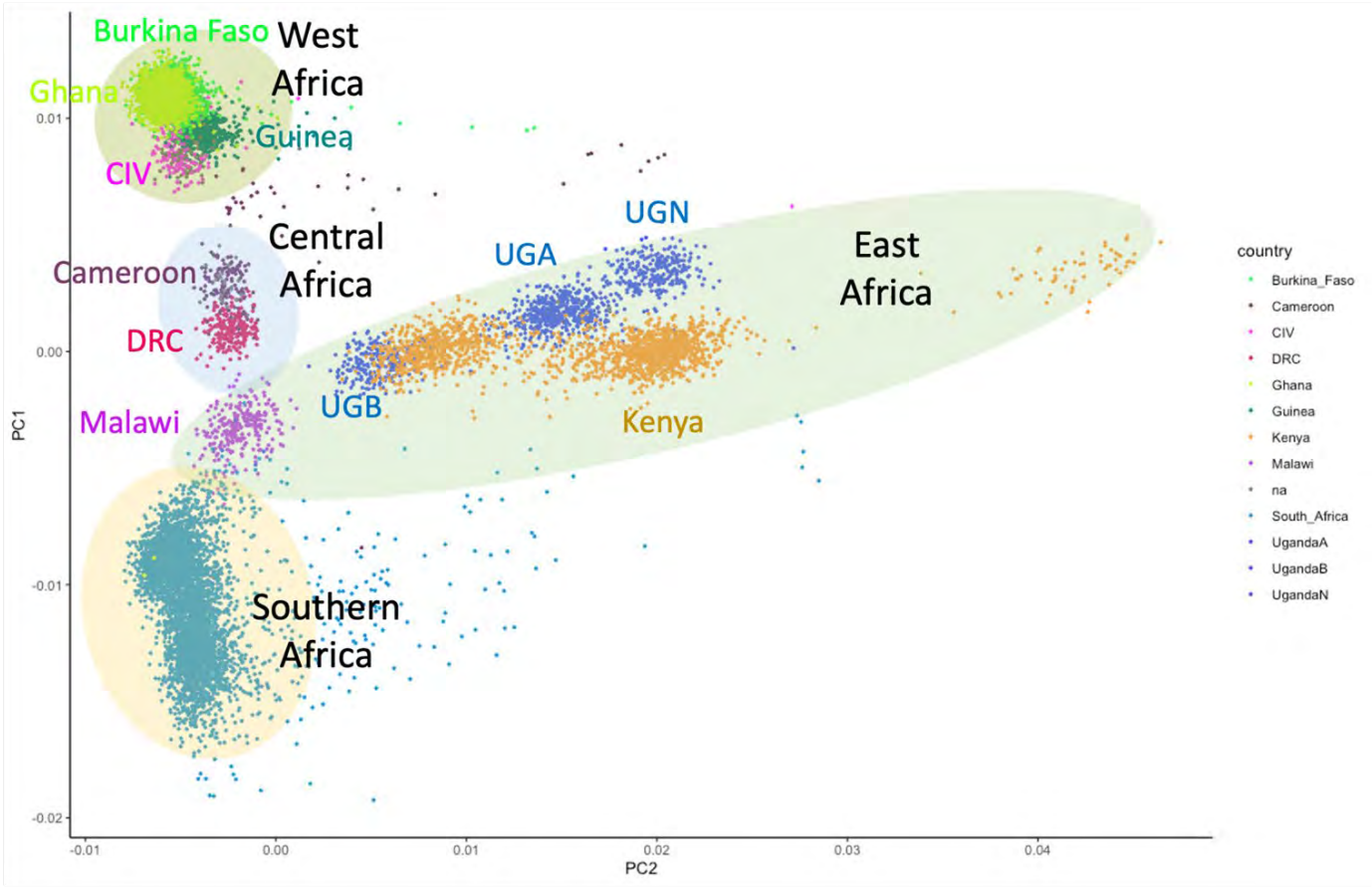
TrypanoGEN+
n = 3,086

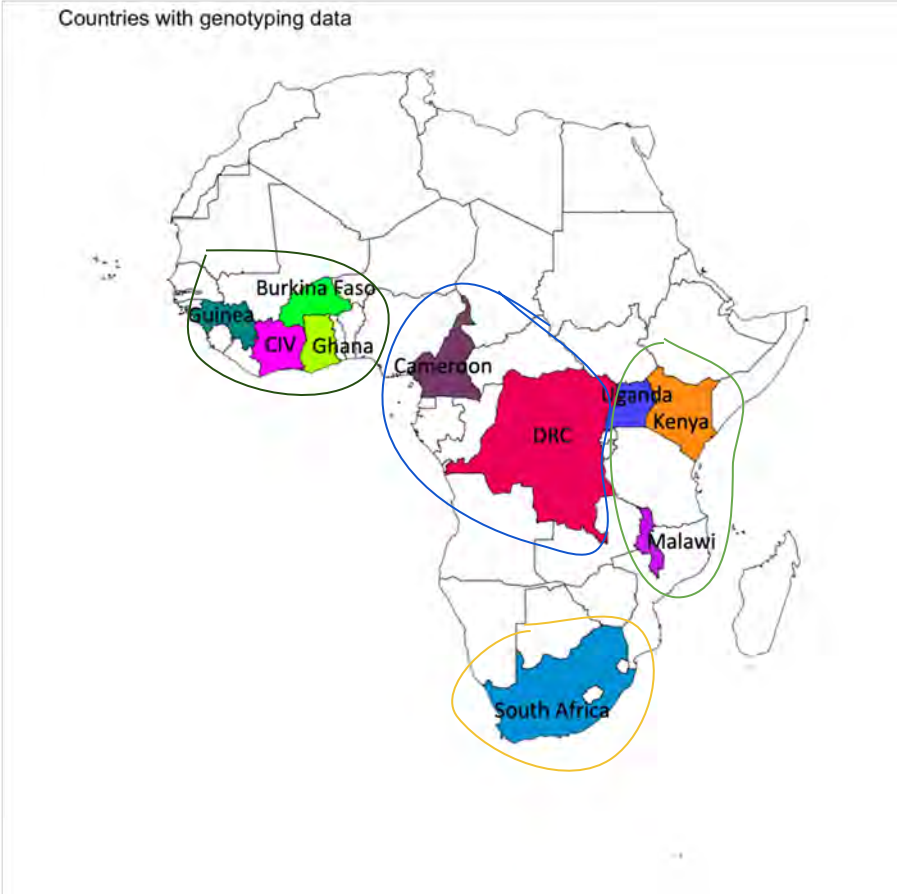


West: n = 4,274
East: n = 2,803
South: n = 4,681
Central: n = 672

Generated in mapchart.net

Population cohorts for mCAs



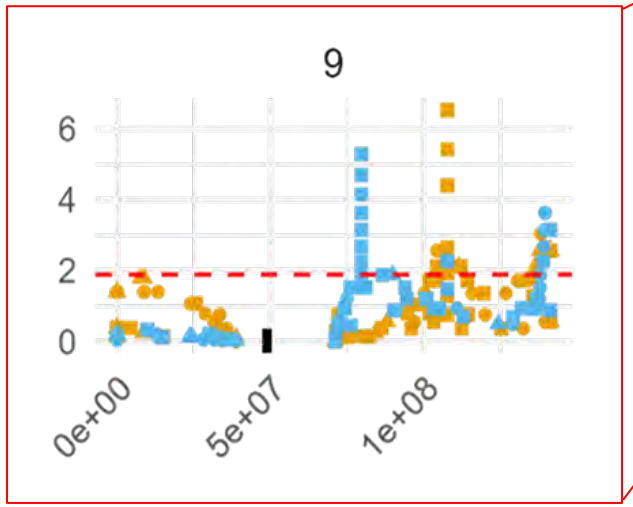


With Michele Ramsay

Geographic Variation in Somatic Mutations

mCA breakpoint hotspots

Binomial test of whether breakpoints are significantly enriched for being overlapped by mCAs



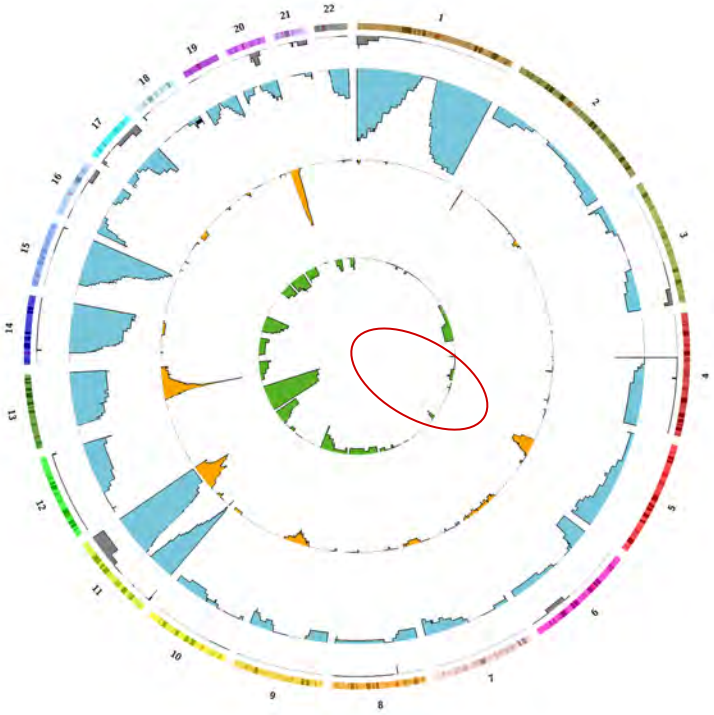
mCA Type
- CN-LOH
- Gain
- Loss

Geographic_Region
- Africa
- North America

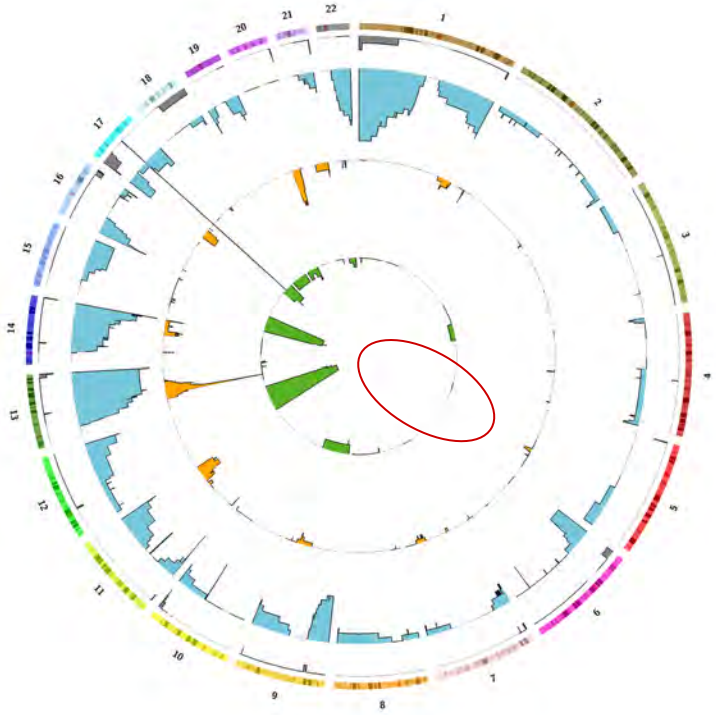
mCA patterns differ across ancestries

European ancestry

CaG-GSA (Quebec)
n = 25,843

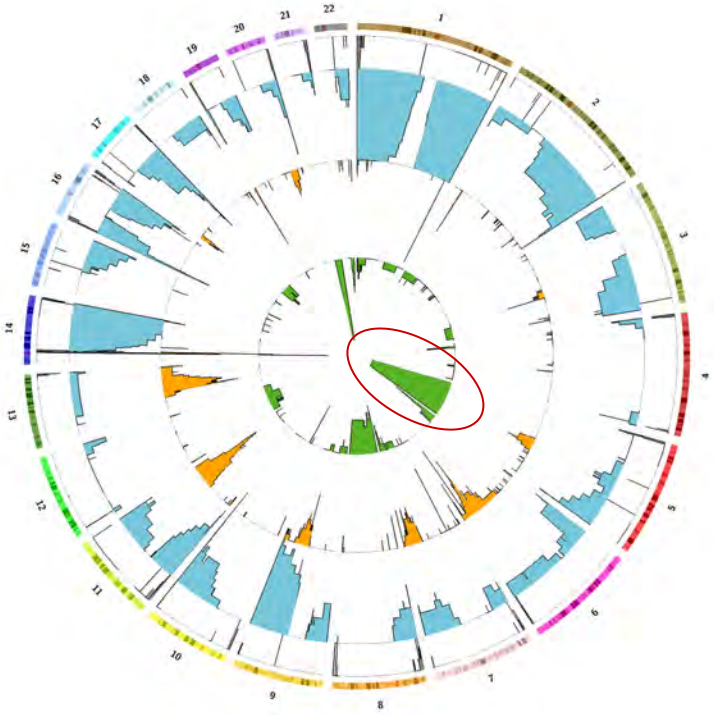


OHS (Ontario)
n = 15,227



African ancestry

AWI-Gen + TrypanoGEN (Africa)
n = 12,841

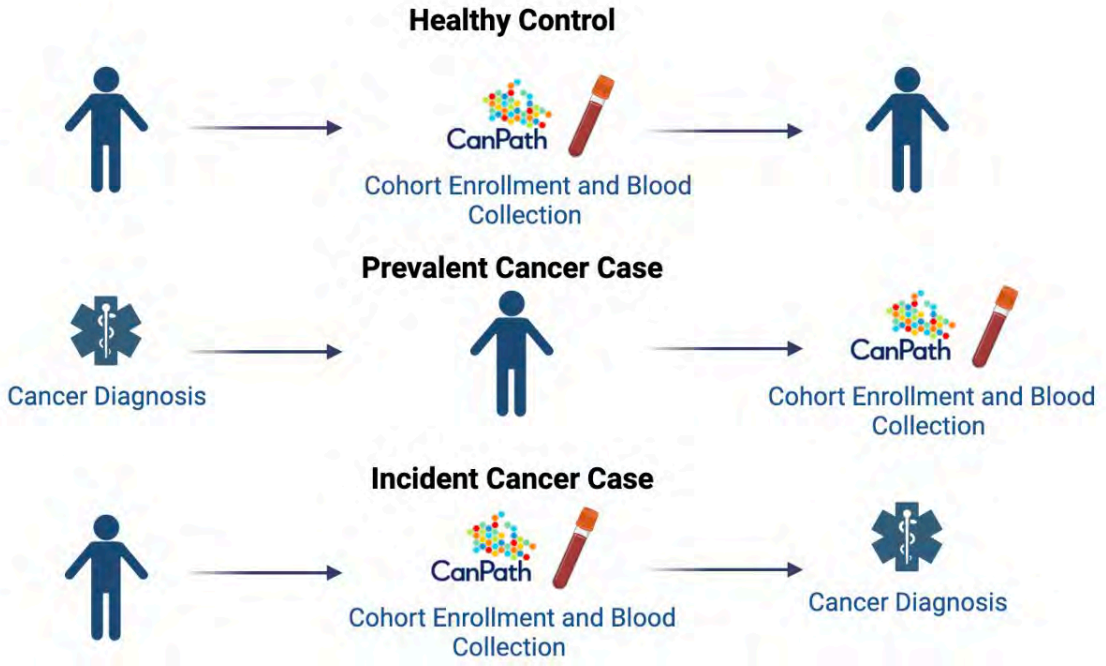


= gain, = loss, = CN-LOH, = undetermined

mCAs accumulate across ARCH- and cancer-associated genes

Individuals with at least one mCA are at significantly greater risk of progressing to blood cancer

Almost all participants in CanPath have consented to administrative health linkages

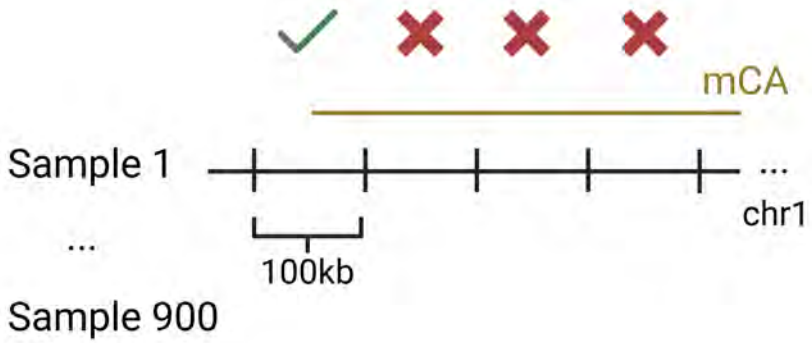


Variable		N	Hazard ratio	p
Number of mCAs	0	7306		Reference
	1-2	272		5.06 (2.47, 10.38) <0.001
	3+	18		26.80 (8.30, 86.56) <0.001
Age		7596		1.08 (1.04, 1.12) <0.001
Sex	FEMALE	4721		Reference
	MALE	2875		2.63 (1.52, 4.56) <0.001

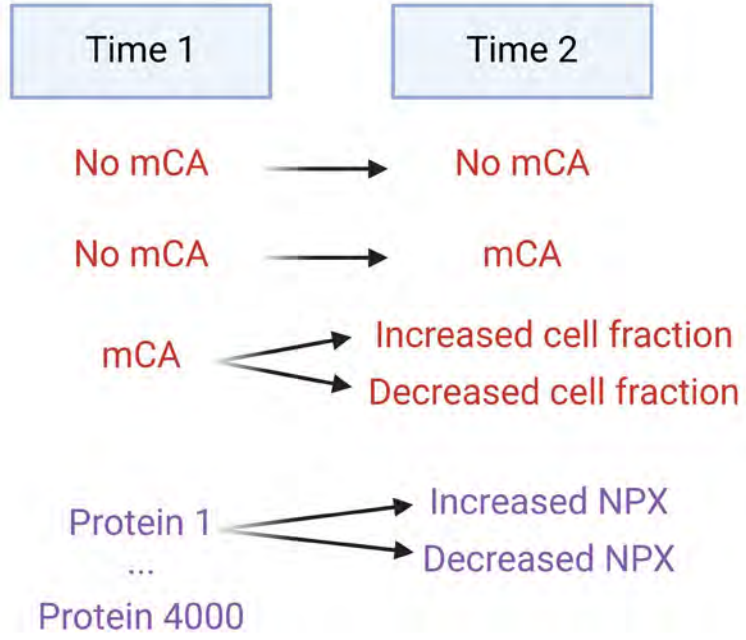
1 2 5 10 20 50

Joint proteomic and CH profiling in the Ontario Health Study to uncover how somatic structural variation impacts the proteome

- pQTL mapping – a type of genome-wide association study (GWAS) to determine what genetic mutations (exposure) are associated with changes in protein levels (outcome)
- Can discover *cis* (<1Mb from gene encoding protein) and *trans* (>1Mb) associations



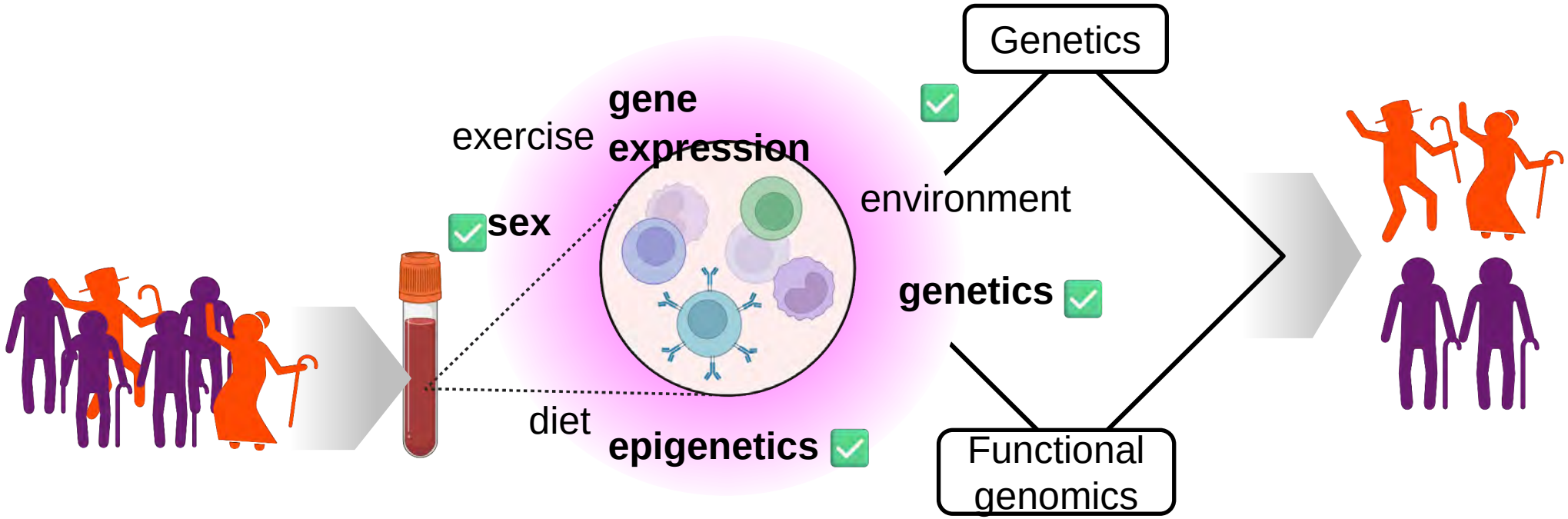
Protein	NPX
IL13	2.43
pro-BNT	0.25
TNFa	-1.17
...	



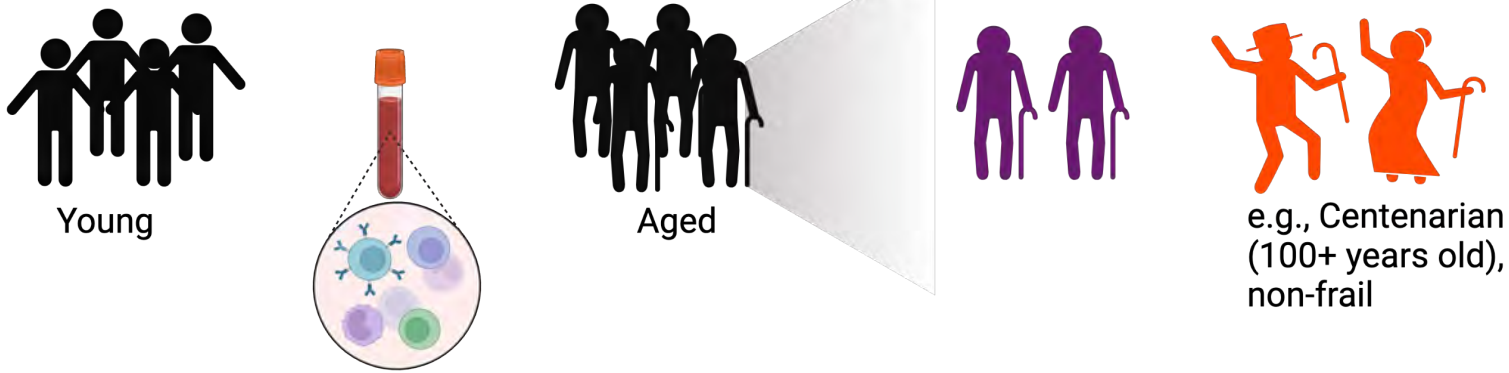
E.g. Do individuals with mCA breakpoint in bin1, chr1 have a higher/lower level of IL13?

Summary and future directions

What factors contribute to healthy aging of blood?



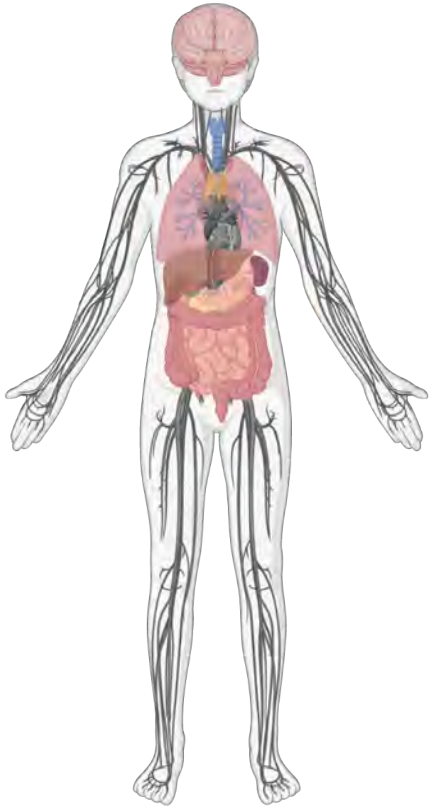
Classical studies of aging don’t capture tissue-specific variation



Aged vs. young blood

- ↑ myeloid cells
- ↑ exhaustion & senescence
- ↑ inflammation
- ↓ naive cells
- ↓ phagocytosis
- ↓ cytotoxicity
- ↓ activation

Centenarian



Health check

- brain ☑
- lungs ☑
- liver ☑
- …etc

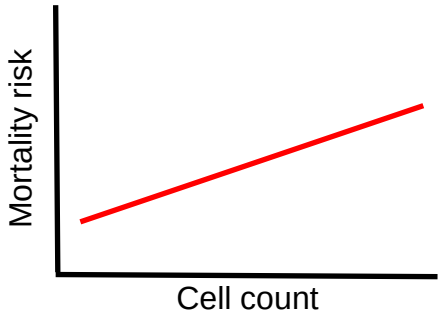
- heart ⊗
- blood ⊗

Intermountain risk score predicts 5-year mortality

CBC data + age
n = 25 000

5 year follow-up

Risk modeling

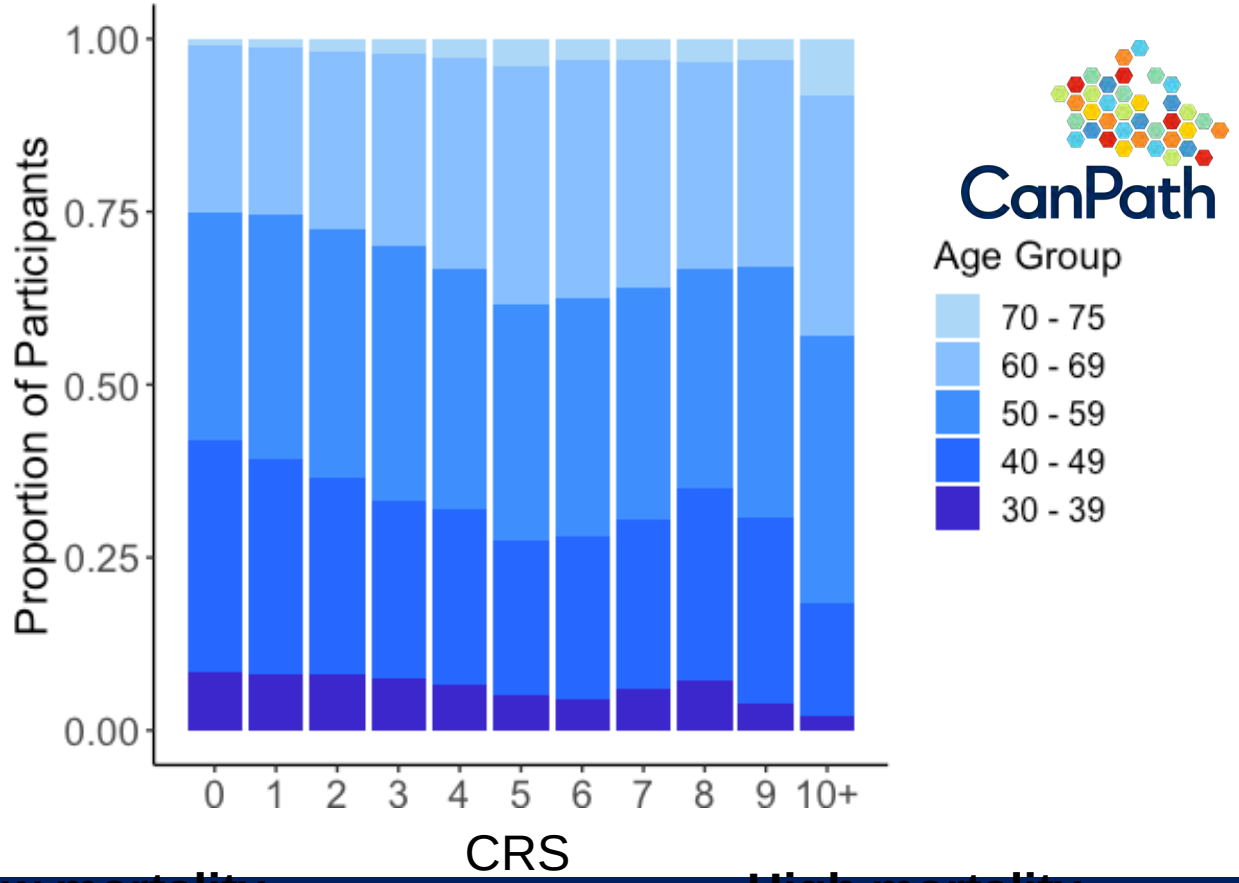


Low IRS = Low mortality risk = Healthy blood

Horne et al. (2009) Am J Med

Complete blood count Risk Score increases with age

CRS is a modified version of Intermountain Risk Score[1] without the age effect
∴ comparable across all ages



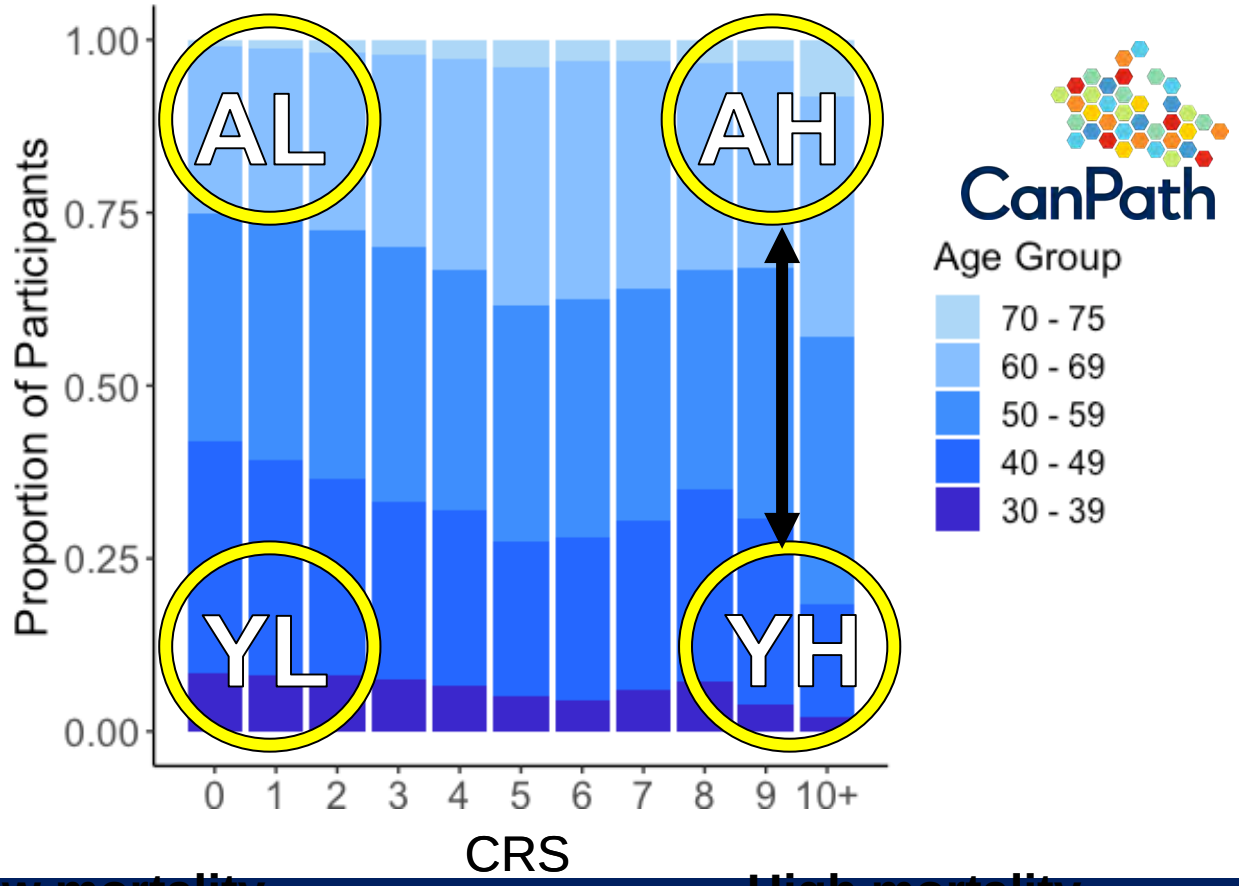
Variables in CRS

- Hematocrit
- White blood cell concentration
- Platelet concentration
- Mean corpuscular volume
- Mean corpuscular hemoglobin concentration
- Red blood cell distribution width

[1]*Horne et al. (2009) Am J Med*

CanPath

Identifying mechanisms of healthy aging in blood



Hypothesis 1

Protective mechanism: AL blood is different from AH blood

Hypothesis 2

Healthy aging mechanism: AL blood is similar to YL blood

Accelerated aging mechanism: YH blood is similar to AH

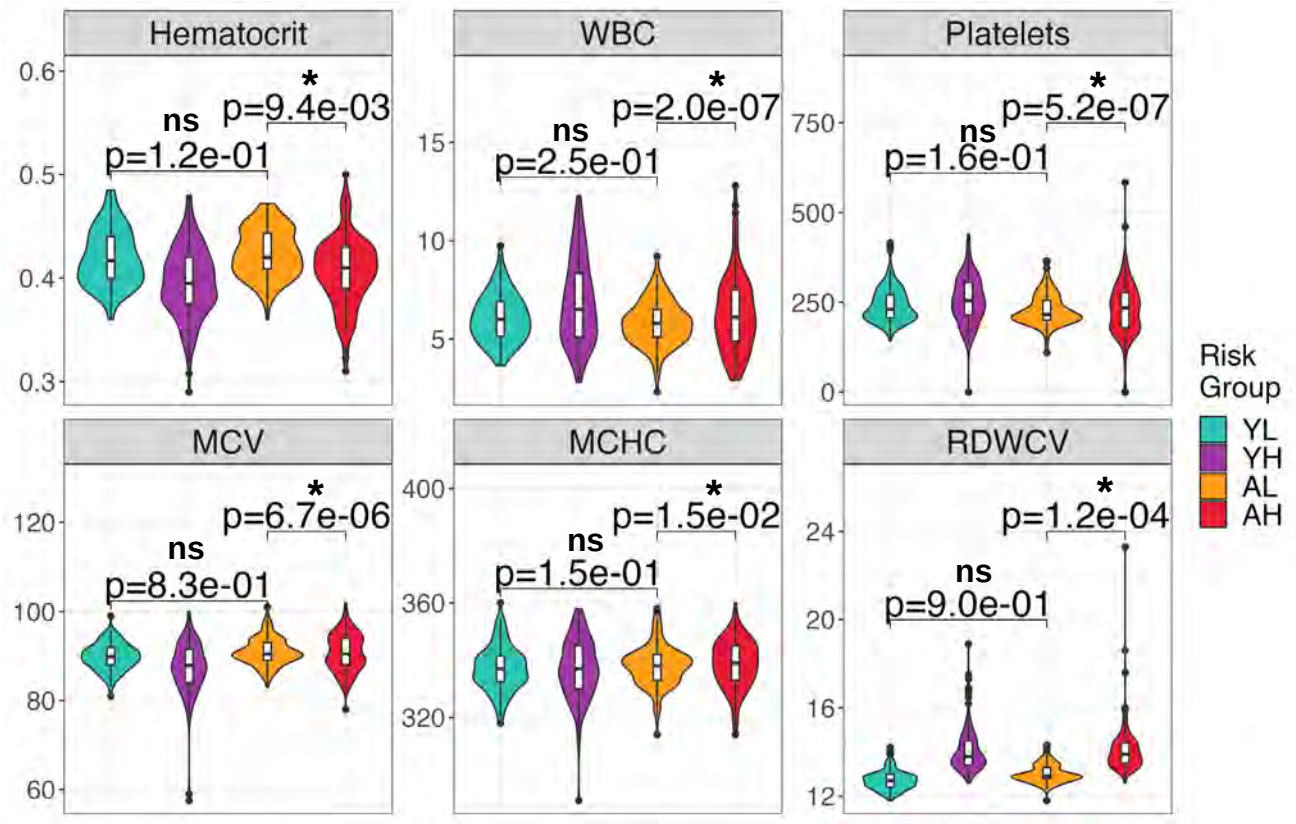
Variance of blood cell phenotypes among aged and young low-risk individuals smaller compared to high-risk

Young: 30 – 45 years old

Aged: 65 – 79 years old

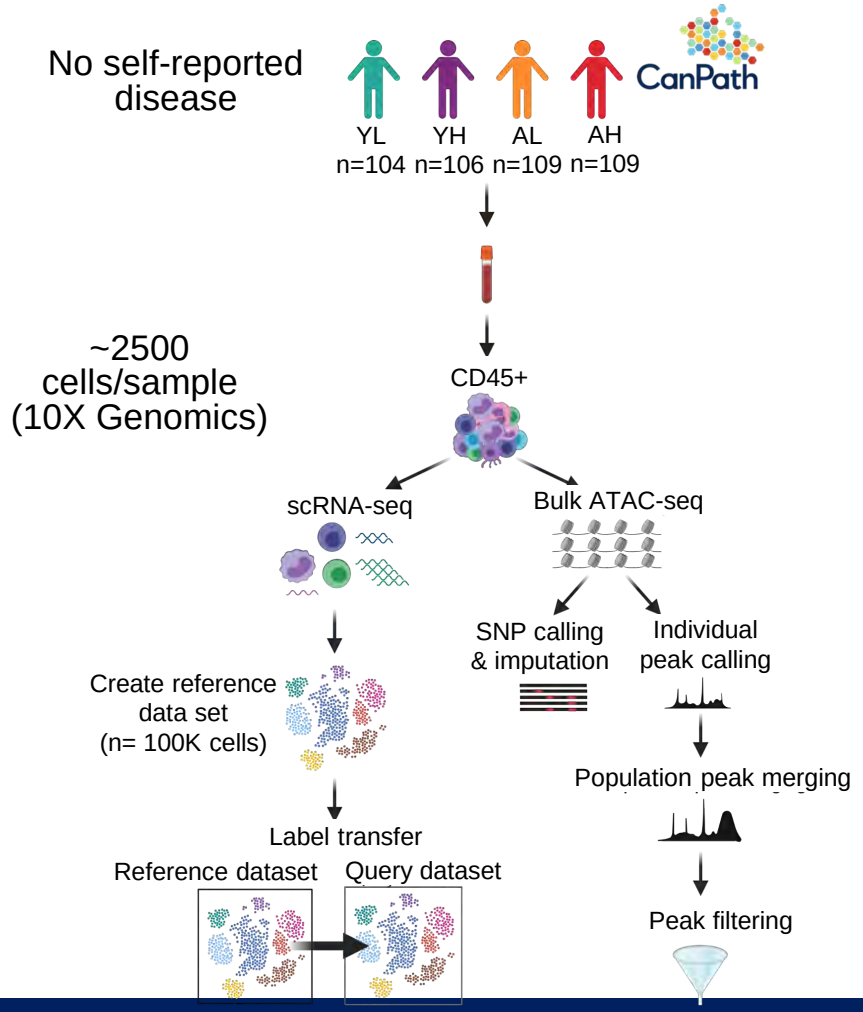
Low-risk: CRS 0 – 3

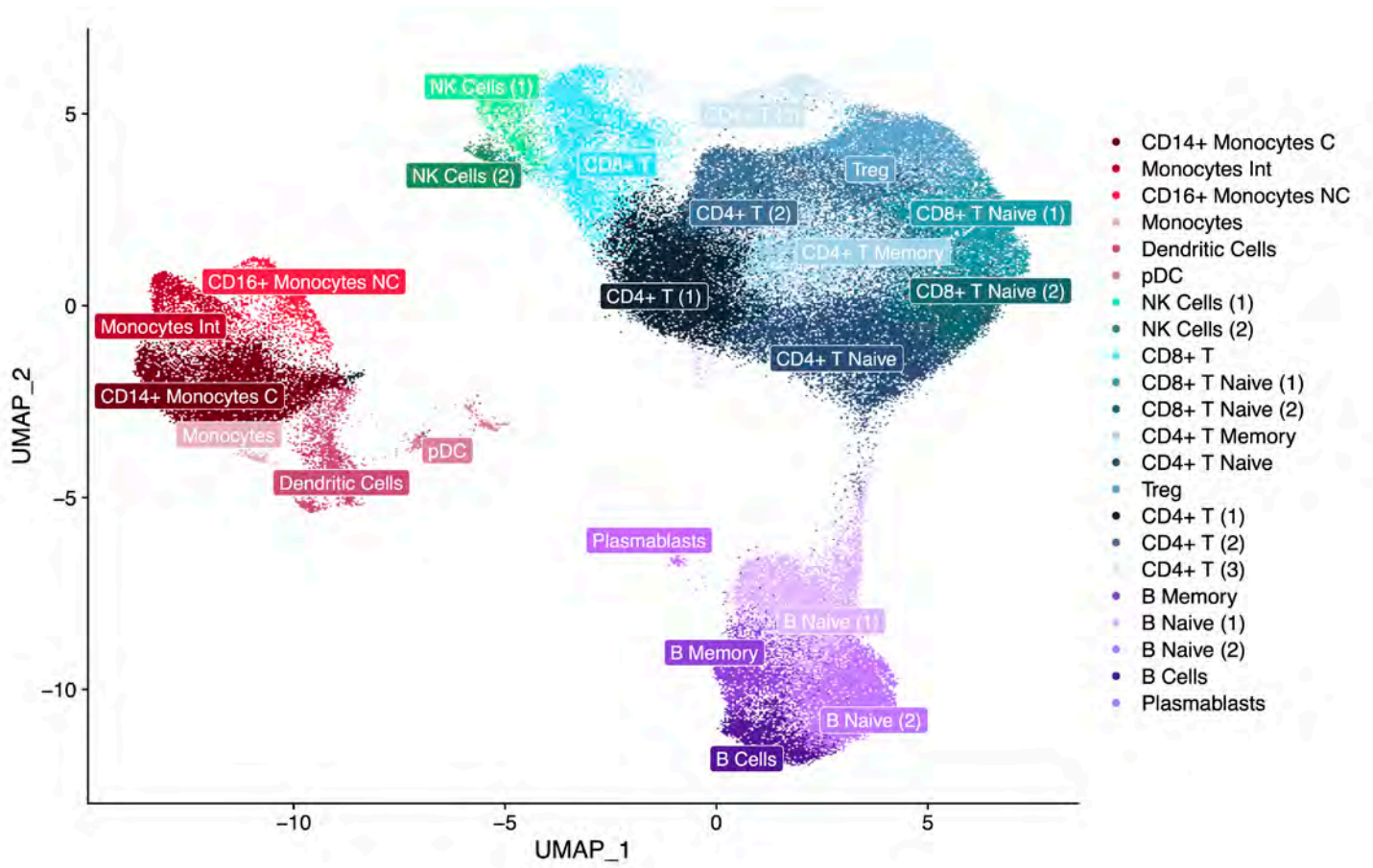
High- risk: CRS 5+



p-values calculated from Levene's test for variance

Single-cell RNA sequencing identifies major blood cell populations from bio-banked blood samples

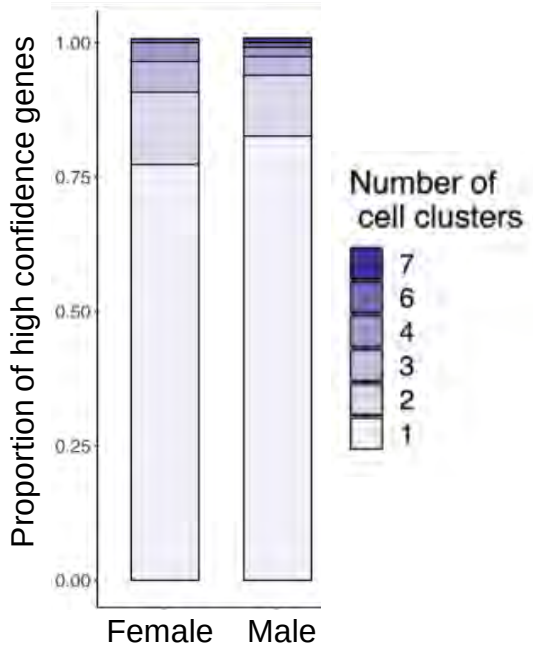




CanPath

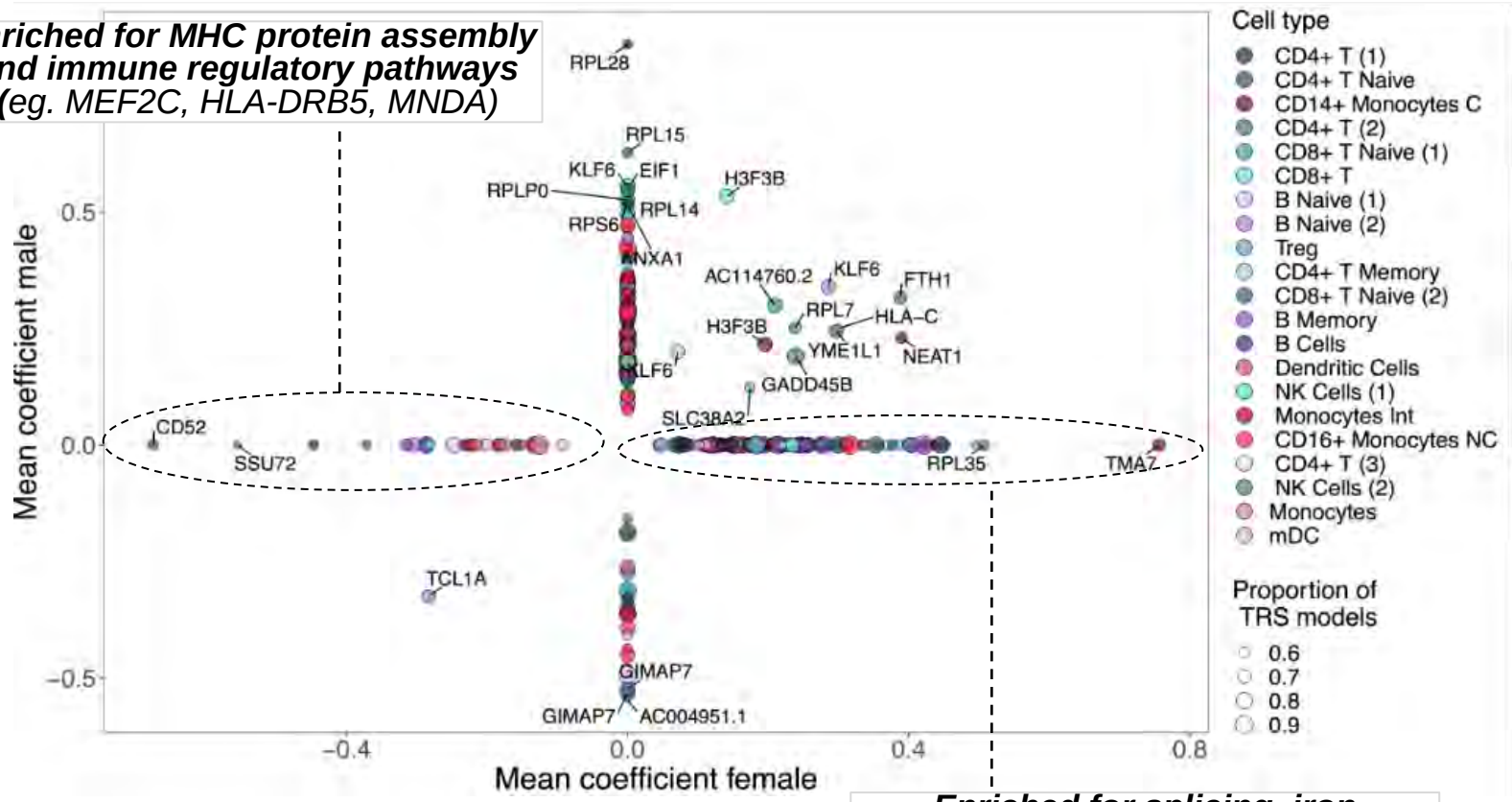
Differential gene expression is sex and cell-type specific

High-confidence genes: non-zero effects in ≥50% of models



77% and **83%** of high confidence genes are cell-type specific in females and males

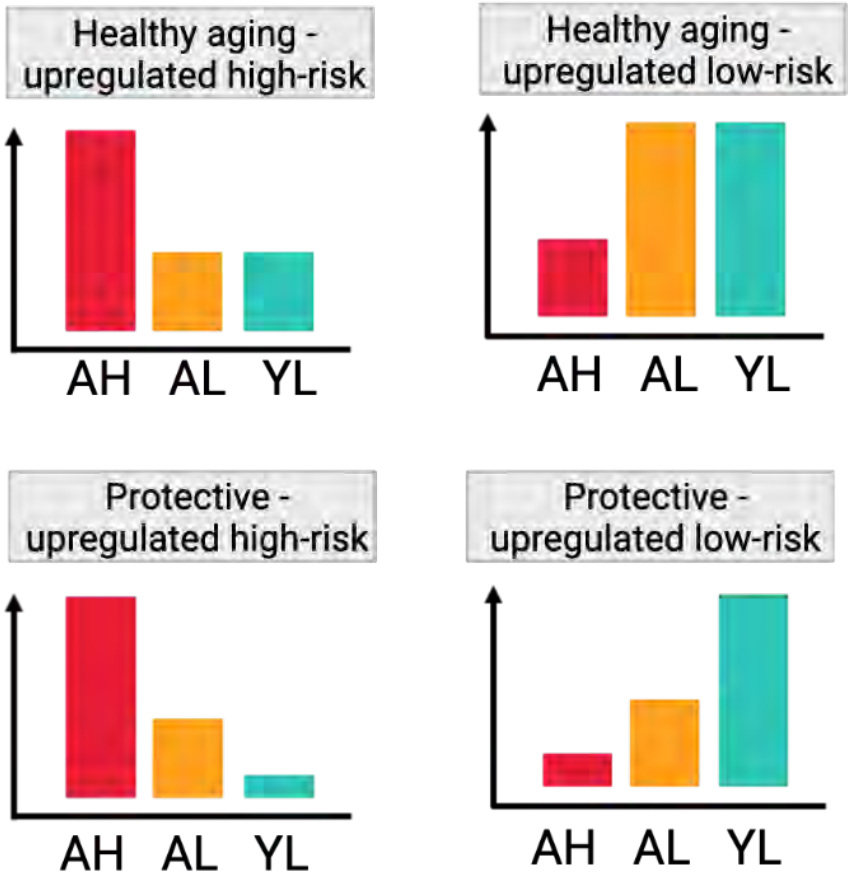
Enriched for MHC protein assembly and immune regulatory pathways *(eg. MEF2C, HLA-DRB5, MNDA)*

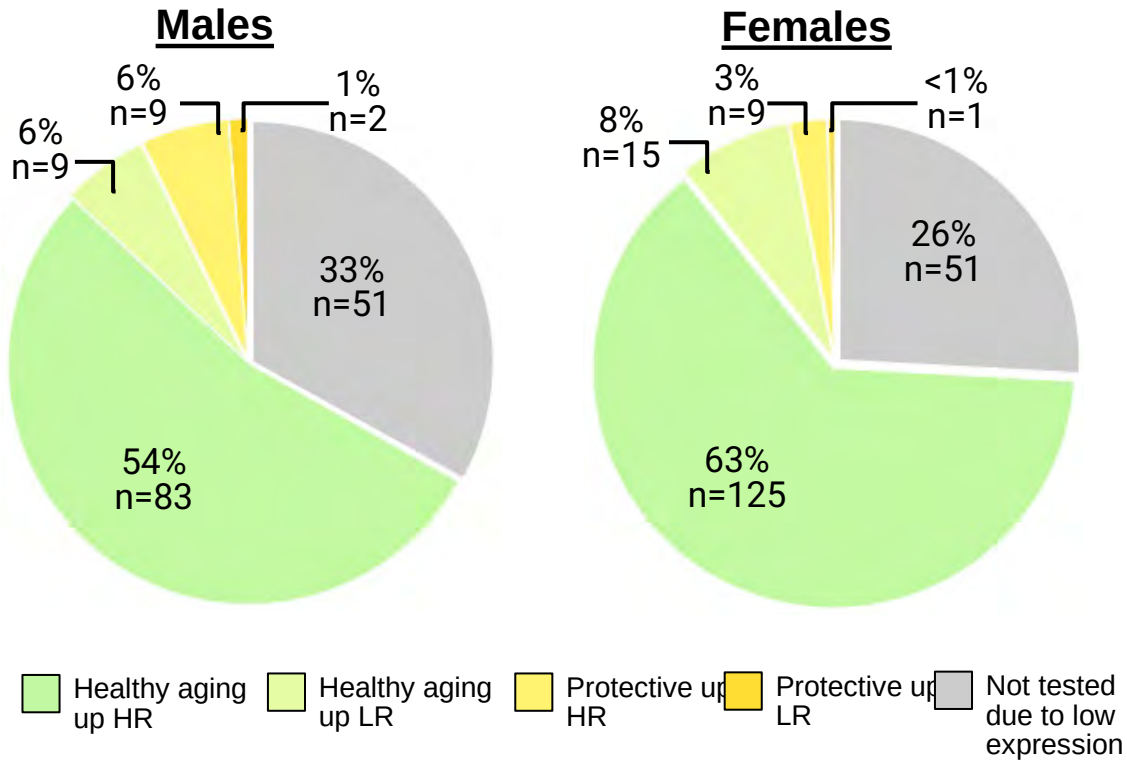


Enriched for splicing, iron homeostasis, and ribosome assembly pathways *(eg. SLC38A2, FTH1, SBDS)*

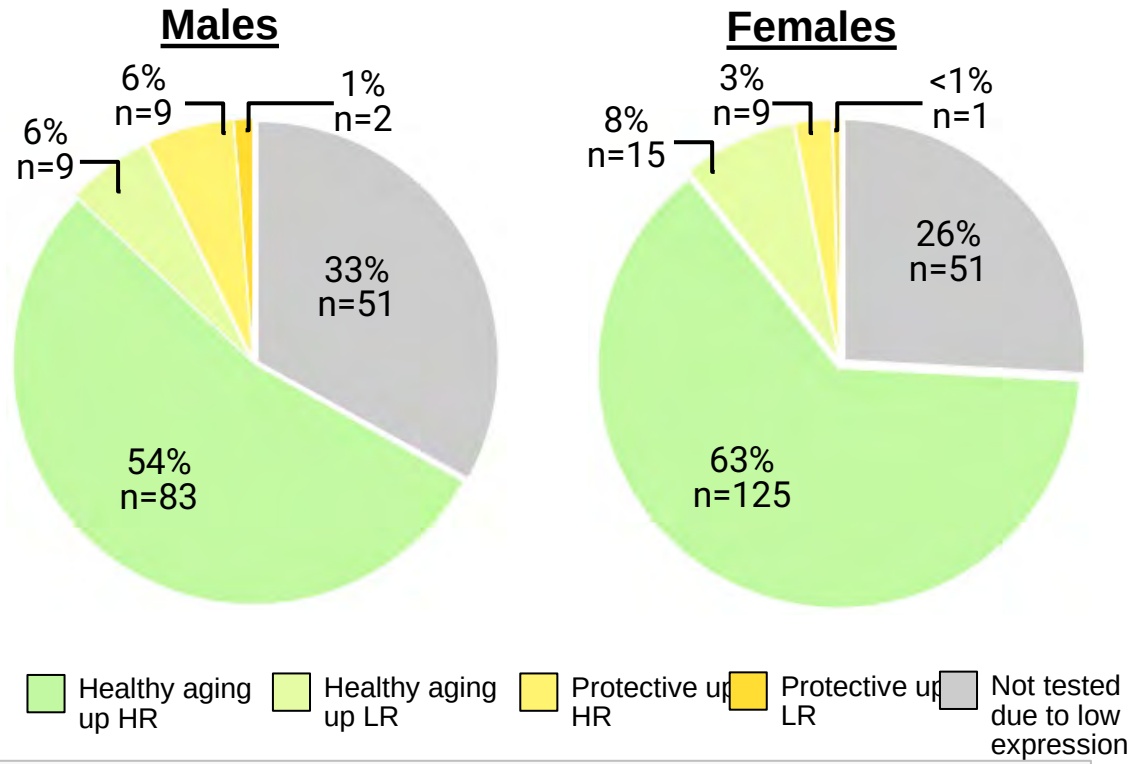
CanPath

Most DGE similarly expressed between YL and AL individuals





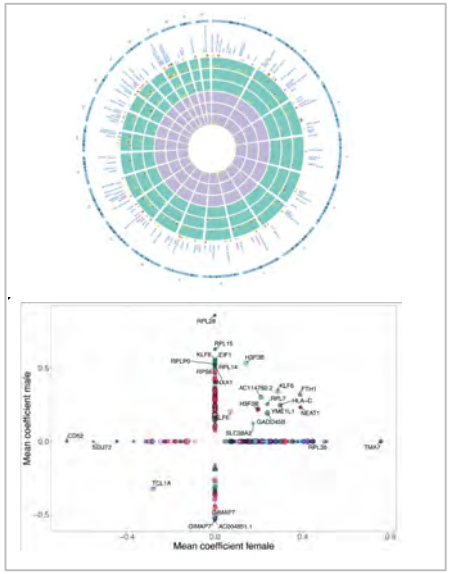
Most DGE similarly expressed between YL and AL individuals



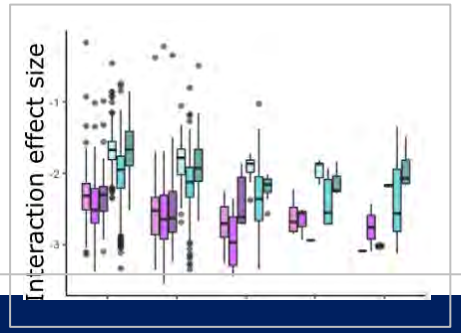
Maintenance of expression levels similar to YL group is the dominant mechanism observed in transcriptional signatures of AL

Summary

Factors contributing to healthy blood aging



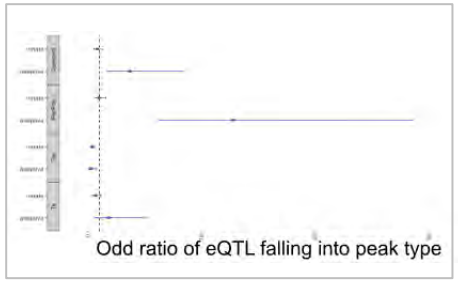
Genetic and **transcriptional** variation associated with healthy blood aging is **sex and cell type specific**



Genetic regulation of gene expression associated with CRS in **innate cells** is stronger and more abundant

Mechanisms of healthy blood aging

Maintenance of gene **expression** similar to **young** individuals



Maintenance of **repressed chromatin**

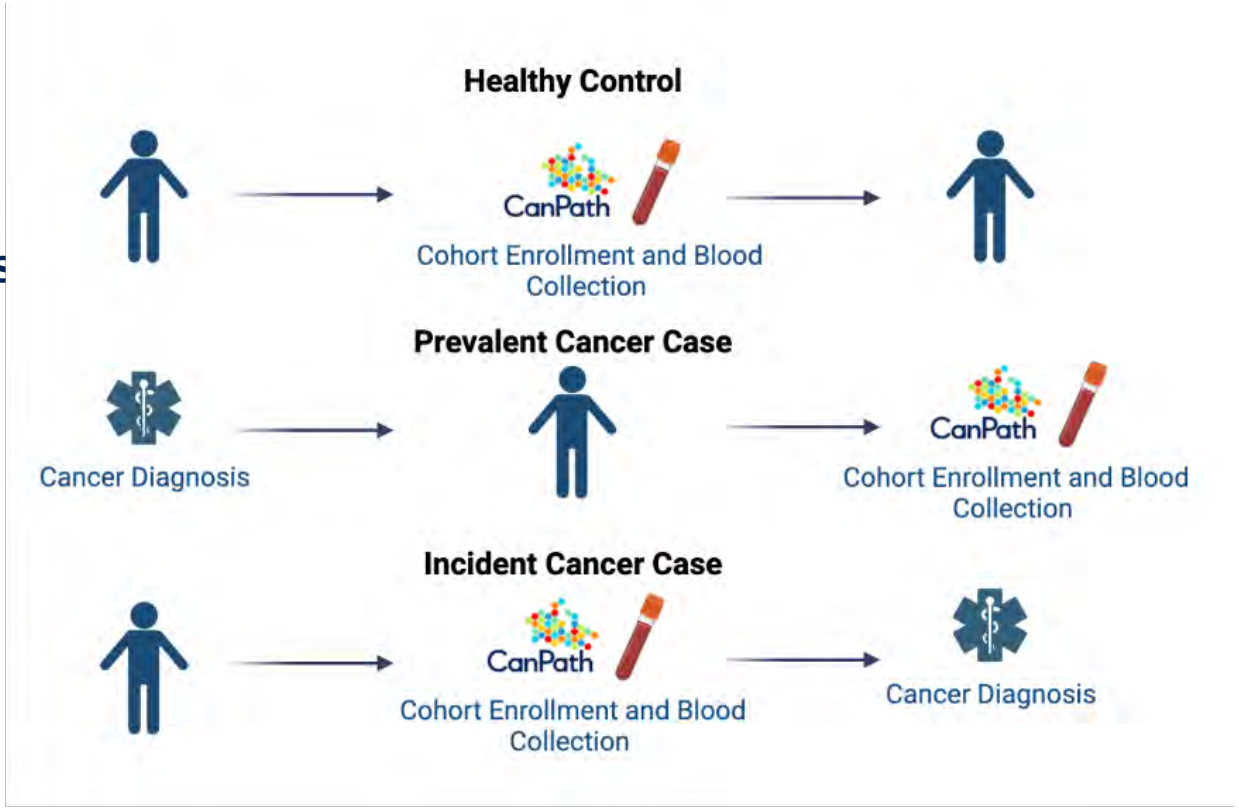
Profiling pre-diagnosis plasma cell-free DNA methylomes up to seven years prior to clinical detection reveals early signatures of cancers

Nicholas Cheng

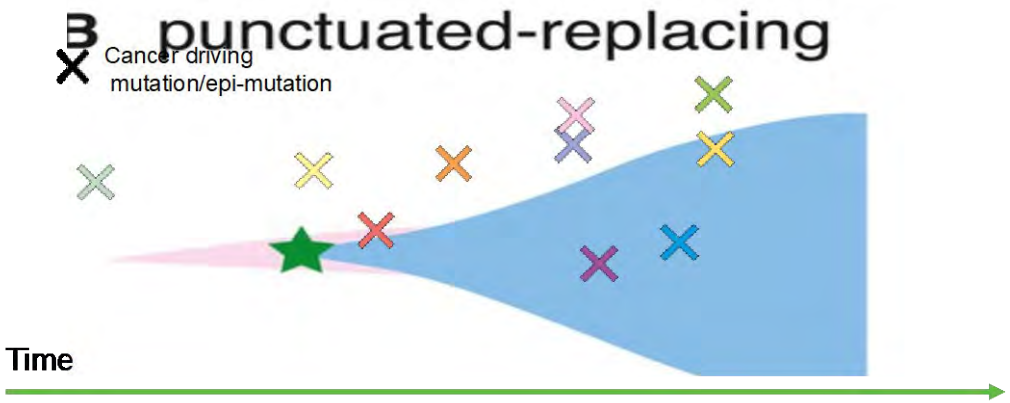
AWADALLA LAB
Pioneering Genomics for Precision Health.

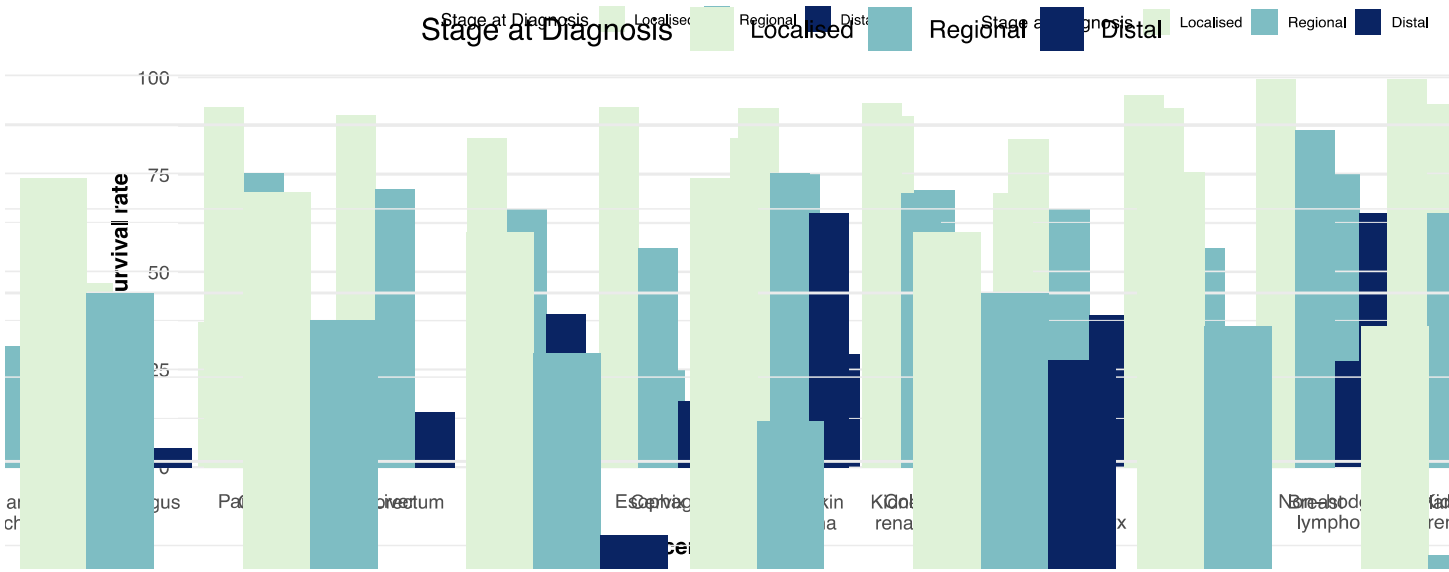
Looking ahead: CanPath is building the Canadian Cancer Study to advance Canadian cancer research and discovery

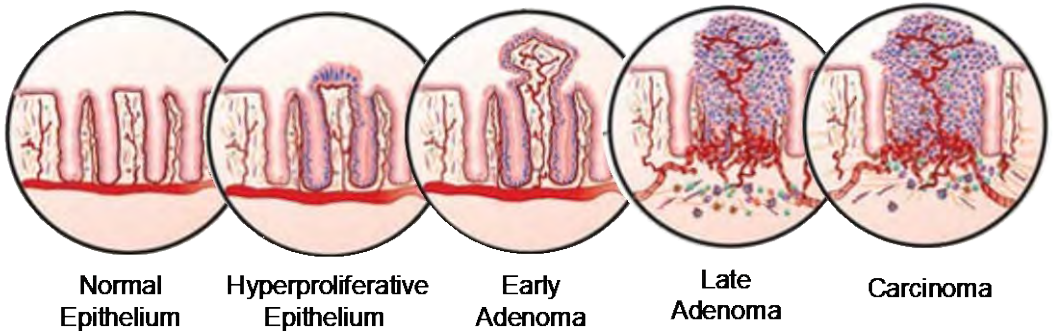
- CanPath is building the Canadian Cancer Study to advance research and discovery for the leading case of death in Canada
- With linked clinical information, we can identify which participants joined the cohort **before developing diseas**
- Using samples collected before disease onset, we are able to develop **novel approaches to detect disease years before current methods**
- We are adopting a three-pronged approach to **build the data resources required to enable early cancer prevention and detection research:**
  - Linking to national administrative data holdings
  - Harmonizing aggregate cancer data reporting nationally
  - Hosting linked individual-level cancer outcomes



Treating cancer early increases survival







Kopetz et al. **Curr Gastroenterol Rep** (2019)

Current effective population-based screening recommendations

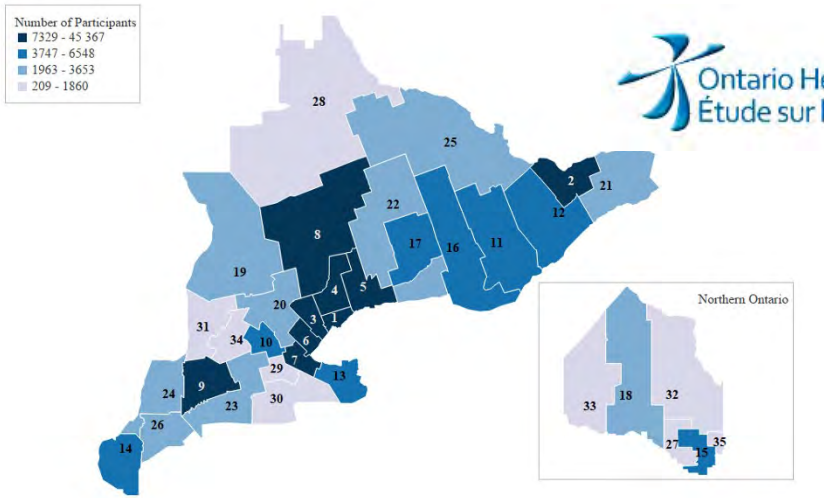
- Breast cancer (ages 50 – 74)
  Regular mammography every 2 to 3 year
- Colorectal cancer (ages 50 – 74)
  FOBT every 2 years or flexible sigmoidoscopy every 10 years
- Cervical cancer (ages 25 – 69)
  PAP smear every 3 years
- Lung cancer (age 55 – 74 smokers)
  Low dose CT

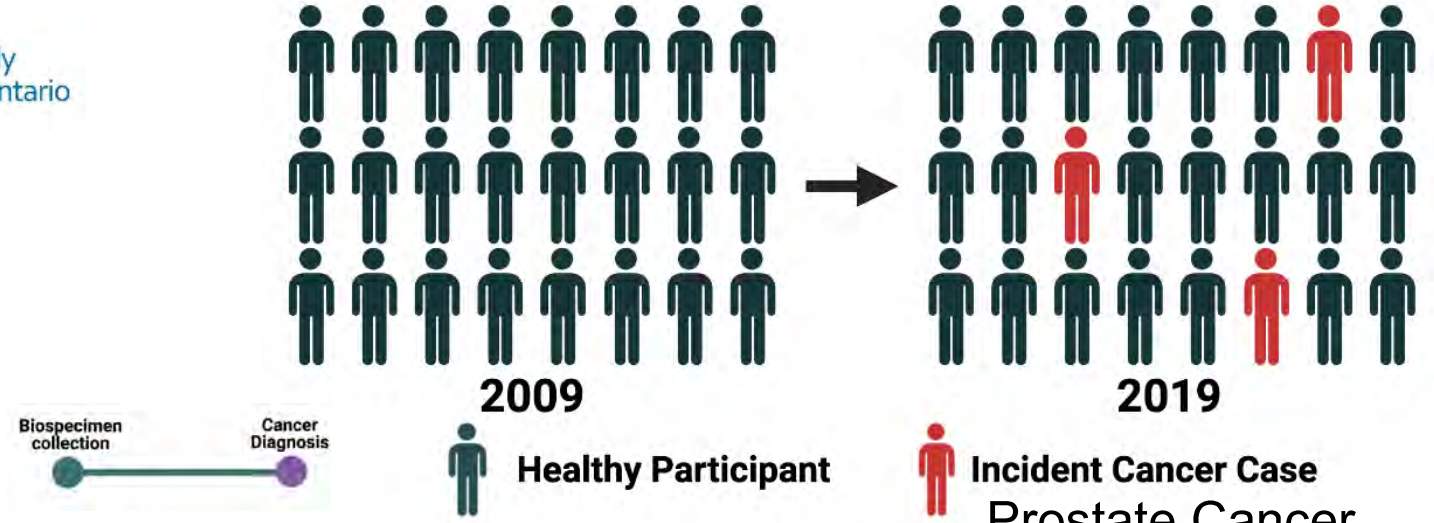
Canadian Cancer Statistics (2018)

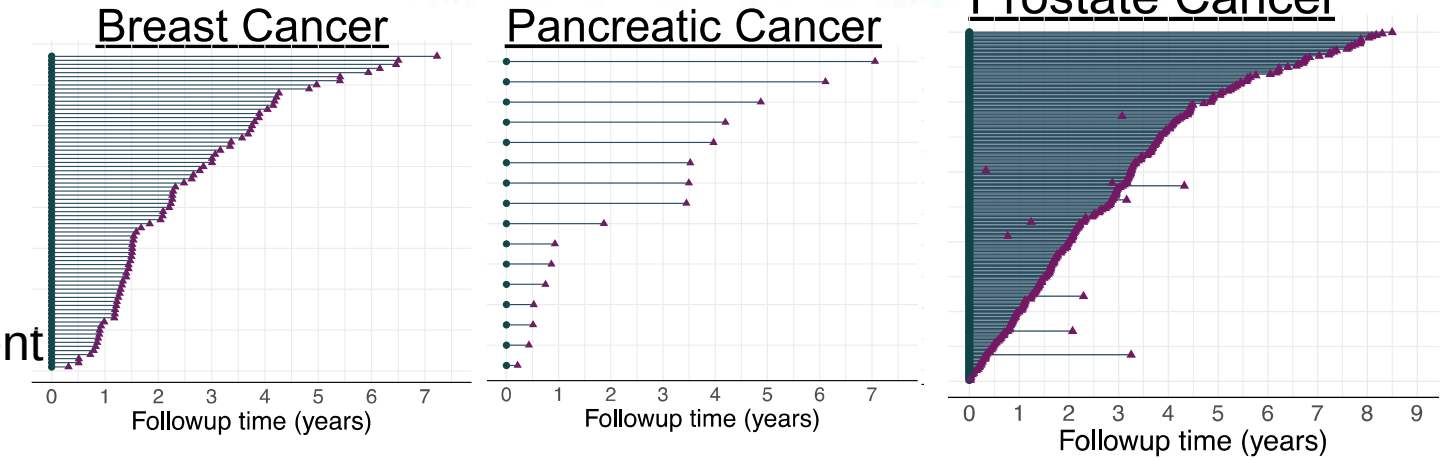
CanPath

Liquid Biopsy Approaches for Early Cancer Detection

Leveraging population cohorts to study early cancer detection prior to clinical detection



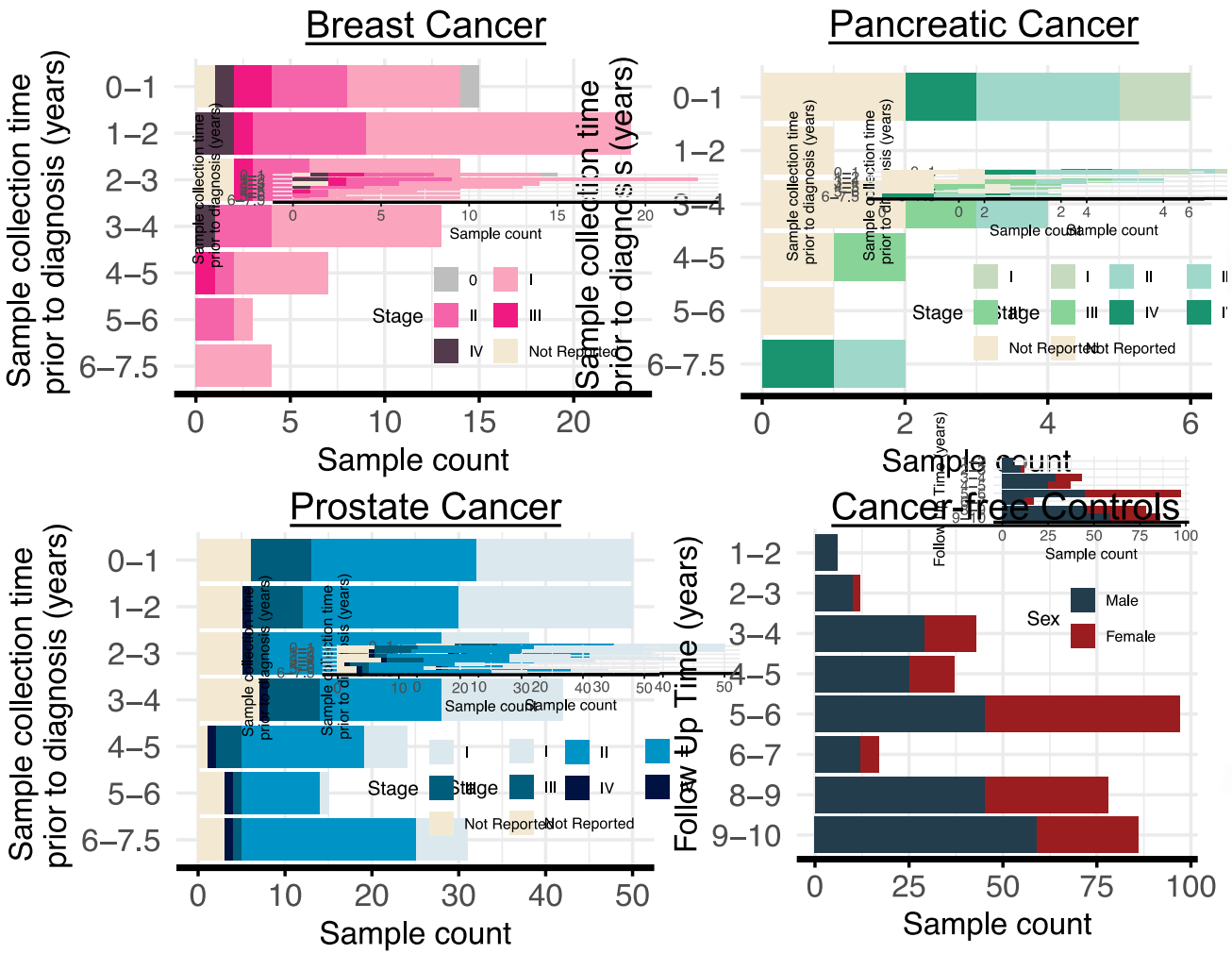


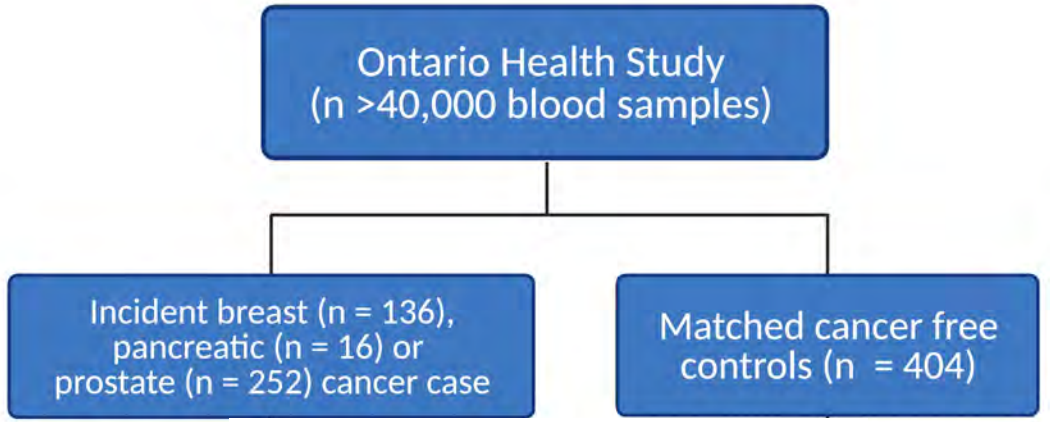


n > 40,000 blood plasma samples from cancer-free participants at time of enrollment

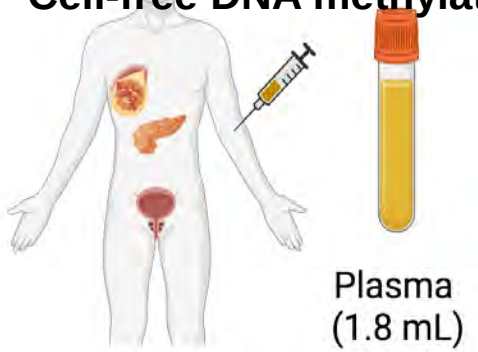
Kirsh et al. **International Journal of Epidemiology** (2022)

Identifying pre-diagnosis cases up to seven years prior to diagnosis within OHS





Cell-free DNA methylation

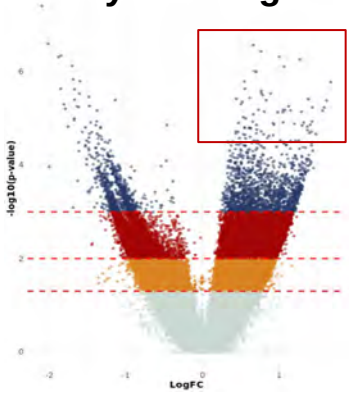


Shen, Singhania, Fehringer et al. 2019 *Nature*

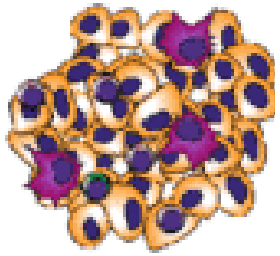
CanPath

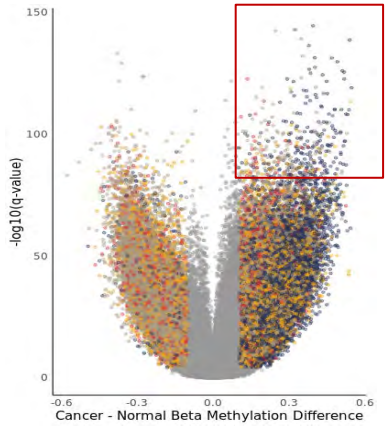
Pre-diagnosis cfDNA methylation signatures share concordant signatures with bulk cancer tissues

OHS Pre-diagnosis Cancer vs Control
cfDNA hypermethylated regions



TCGA Bulk Cancer Tissue vs PBL/Adjacent Normal Hypermethylated regions

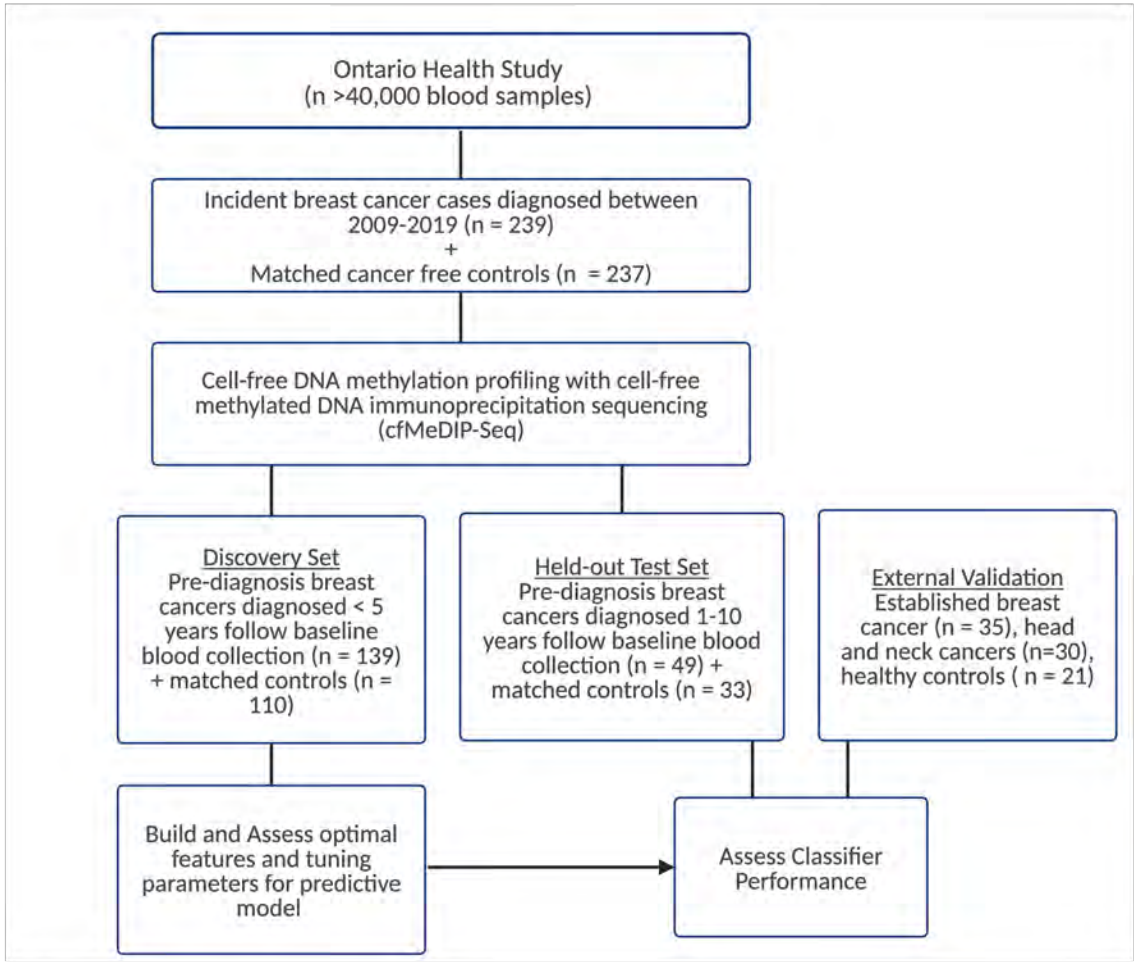




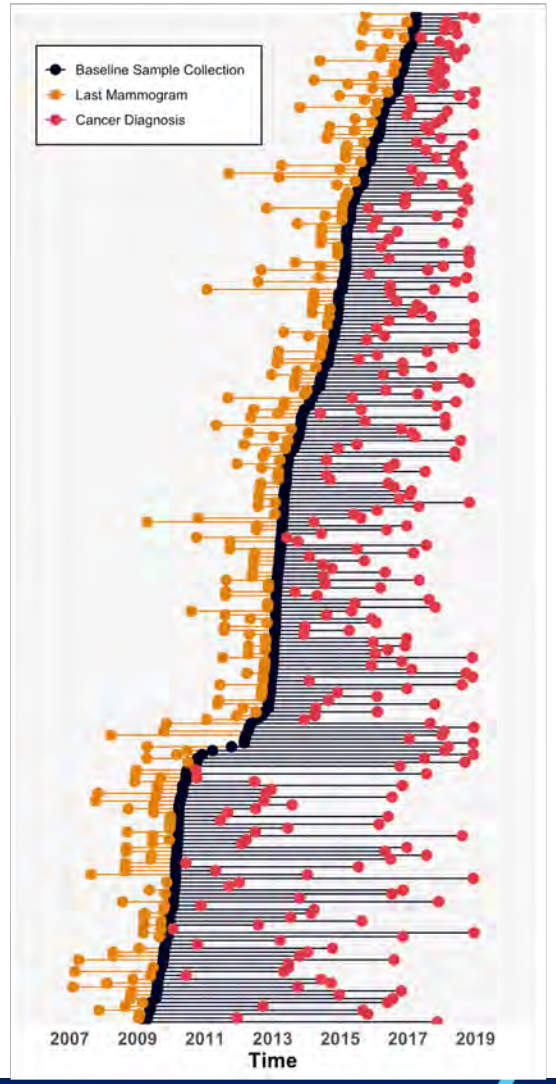
CanPath

Incident Breast Cancer Pre-diagnosis early detection pipeline

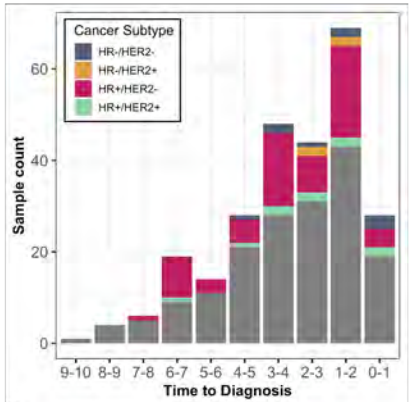
A



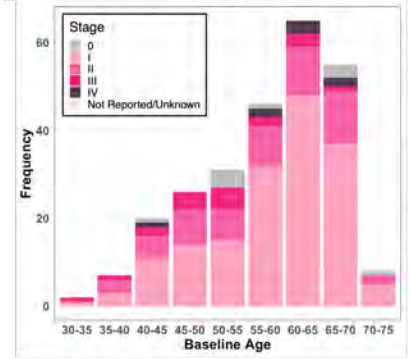
B



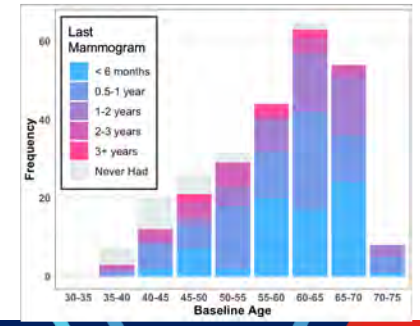
C



D



E

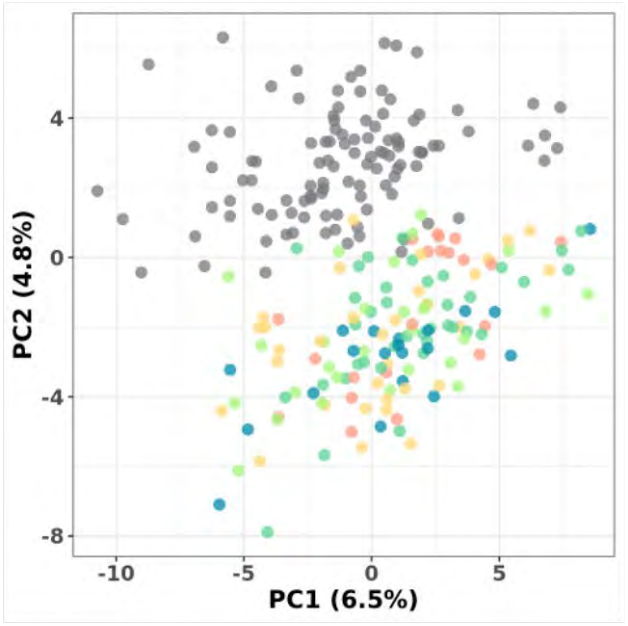


CanPath

Differentially methylated regions discriminates pre-diagnosis breast cancer from control samples

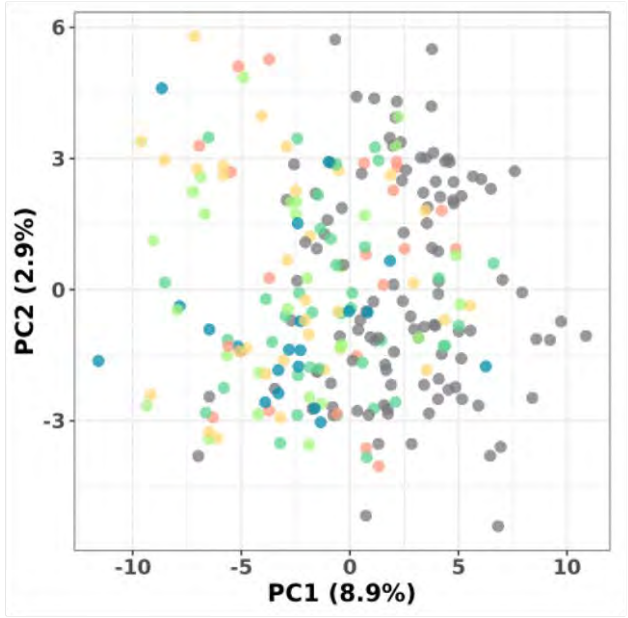
A

Top 150 Hypermethylated +
Top 150 Hypomethylated Regions



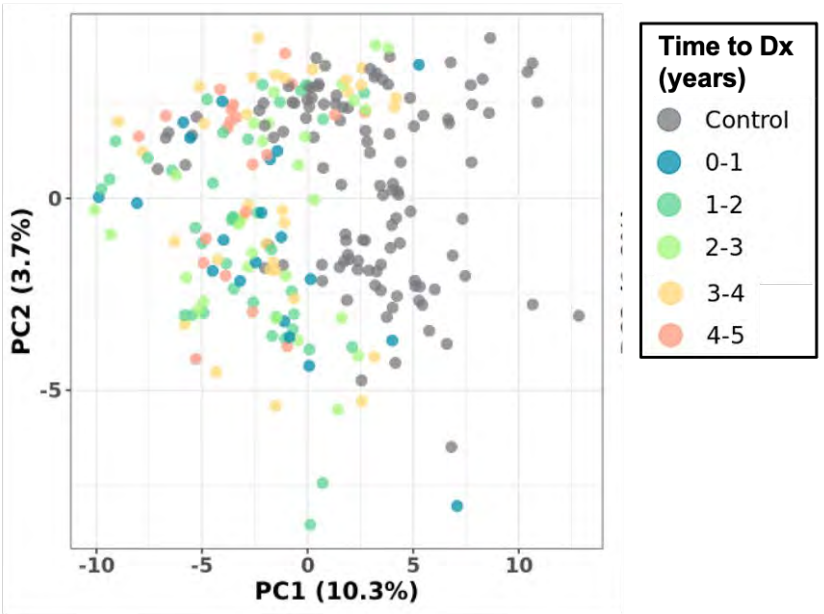
B

Top 300 Hypomethylated Regions



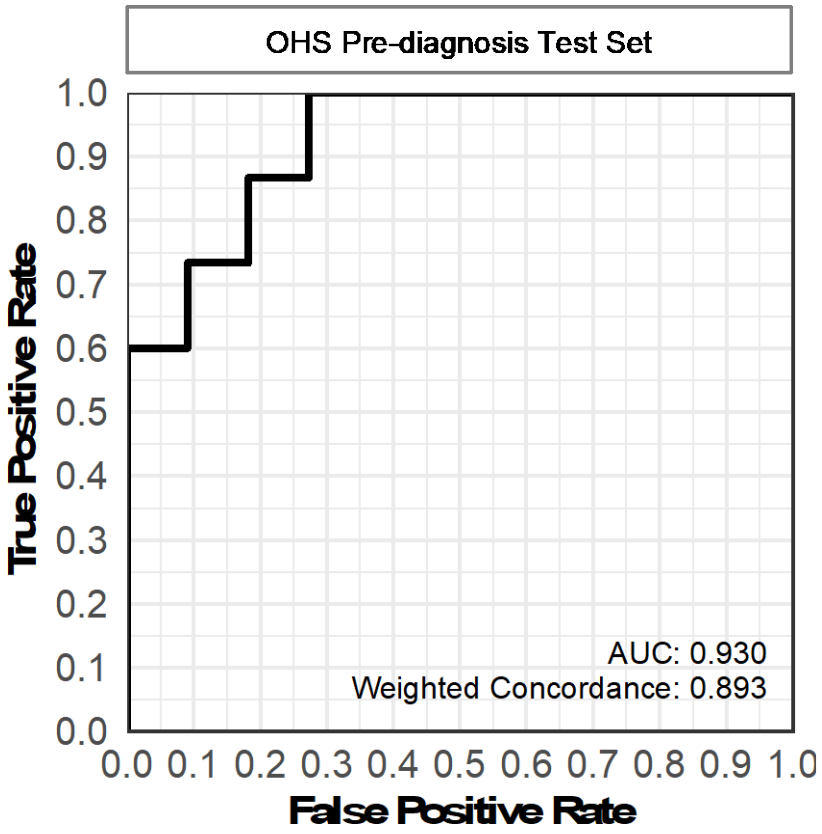
C

Top 300 Hypermethylated Regions

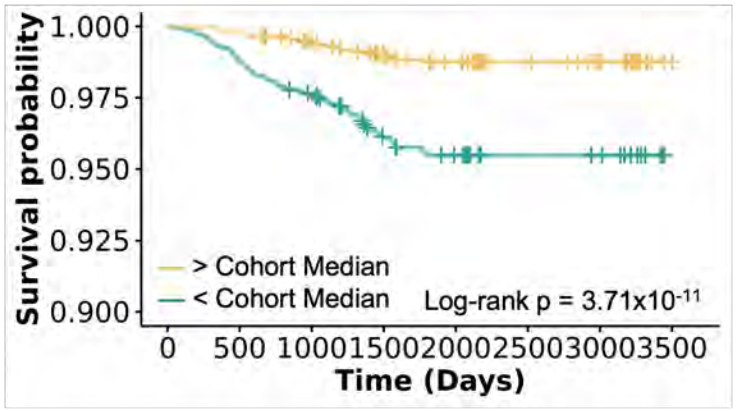


D

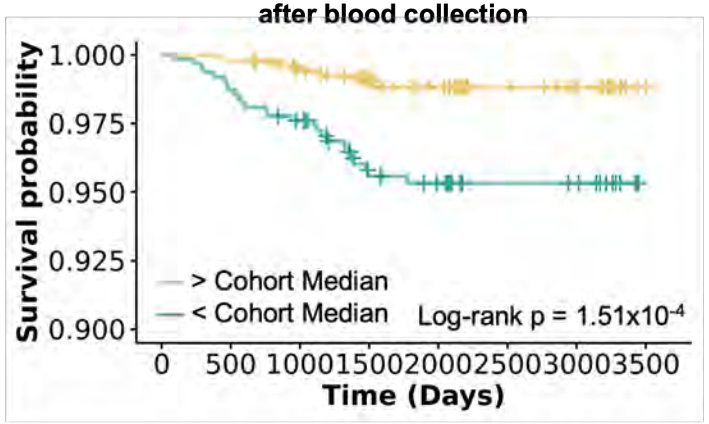
cfDNA methylation signatures detects breast cancers preceding mammogram



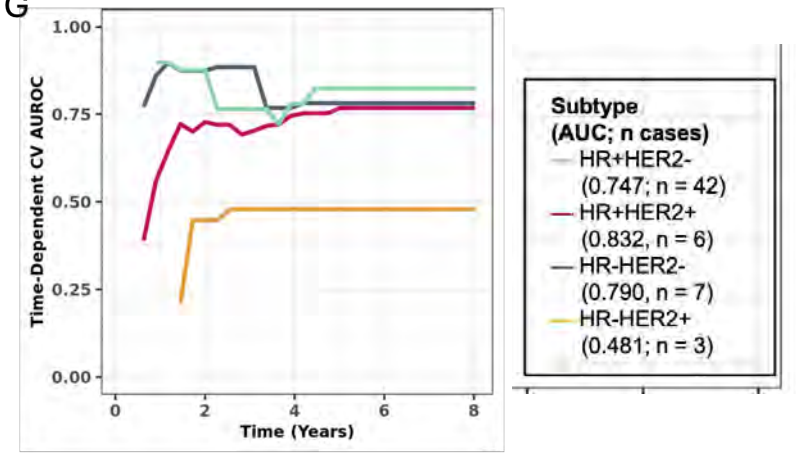
H



I



G

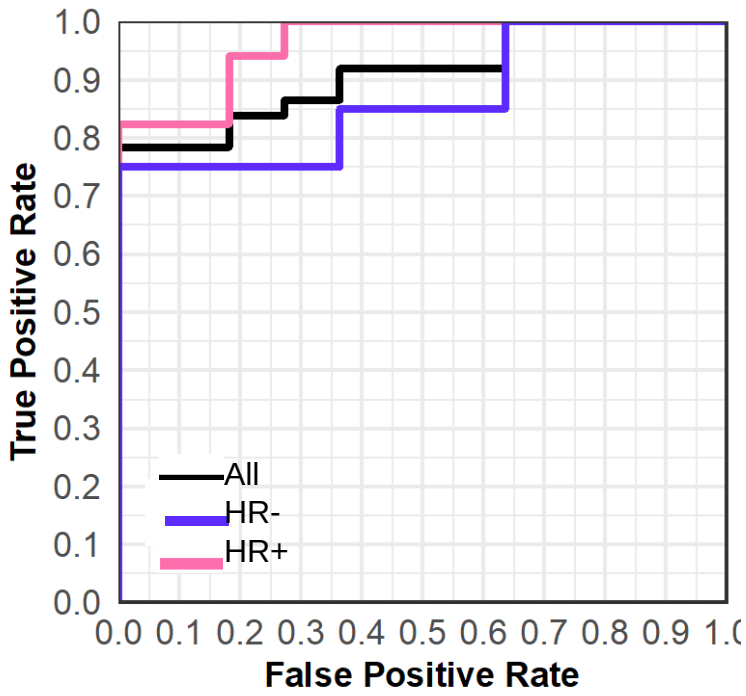


Cheng et al. (in Review)

CanPath

Pre-diagnosis cfDNA signatures are generalizable to other cancers

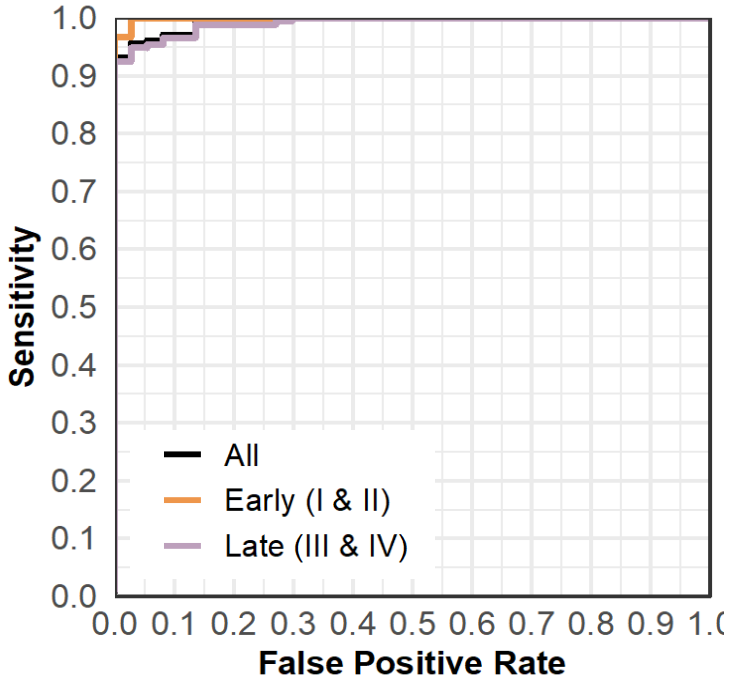
Metastatic Breast Cancer



Breast Cancer Train Set
Pre-dx pancreatic cancer cases (n = 67)
Cancer-free controls (n = 59)

Breast Cancer Test Set
External Ppst-dx pancreatic cancer cases (n = 35)
Non-breast cancer controls (n = 11)

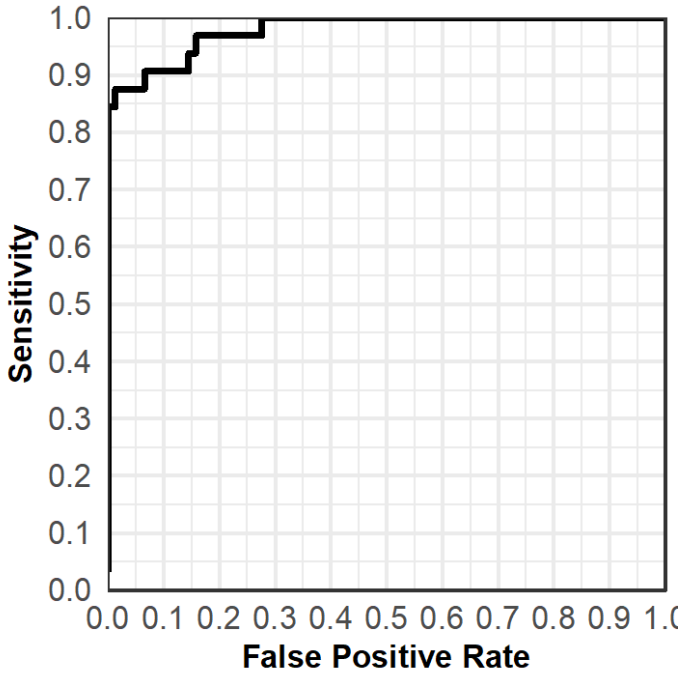
Prostate Cancer



Prostate Cancer Train Set
Pre-dx prostate cancer cases (n = 47)
Cancer-free controls (n = 47)

Prostate Cancer Test Set
External post-dx prostate cancer (n = 102)
Cancer-free controls (n = 58)

Pancreatic Cancer

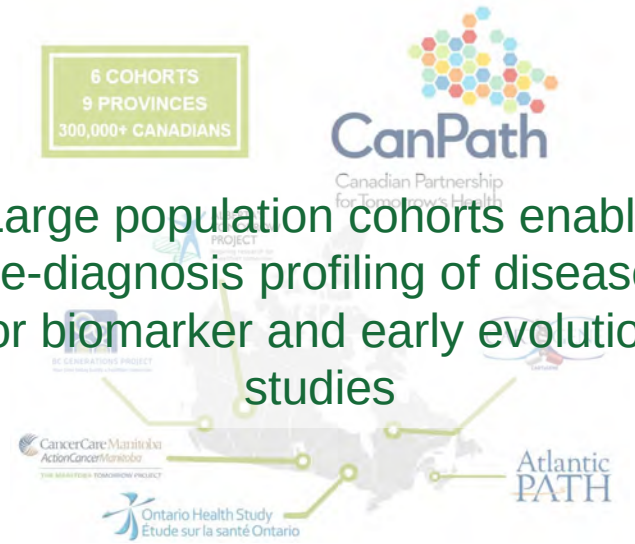


Pancreatic Cancer Train Set
Pre-dx pancreatic cancer cases (n = 16)
Cancer-free controls (n = 50)

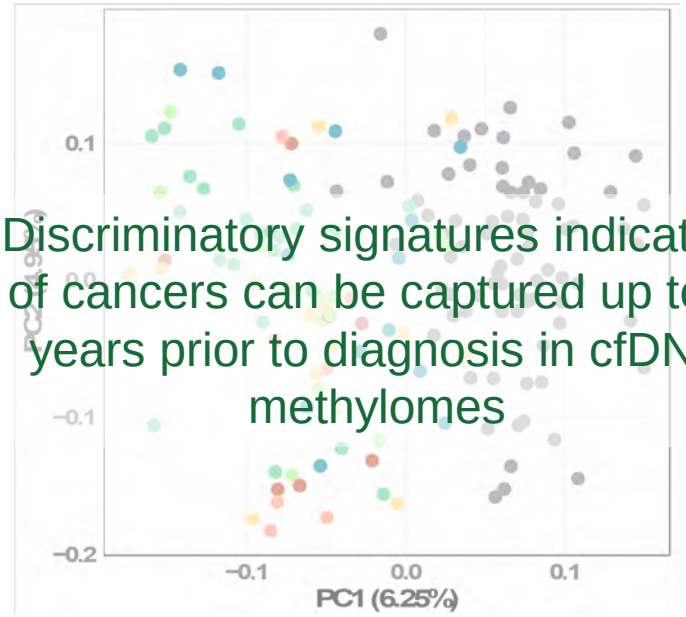
Pancreatic Cancer Test Set
External post-dx pancreatic cancer cases (n = 38)
Cancer-free controls (n = 80)

CanPath

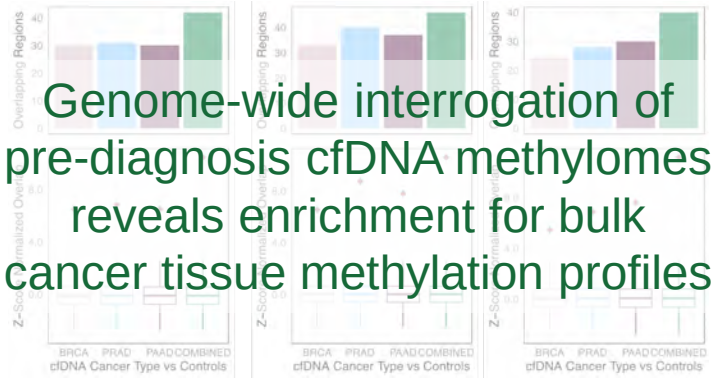
Summary



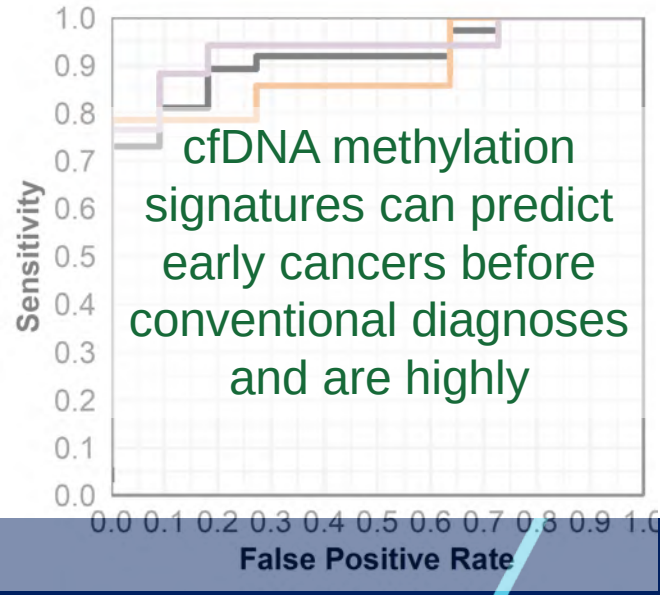
Large population cohorts enable pre-diagnosis profiling of diseases for biomarker and early evolution studies



Discriminatory signatures indicative of cancers can be captured up to 7 years prior to diagnosis in cfDNA methylomes



Genome-wide interrogation of pre-diagnosis cfDNA methylomes reveals enrichment for bulk cancer tissue methylation profiles



cfDNA methylation signatures can predict early cancers before conventional diagnoses and are highly

Acknowledgements

Awadalla Lab

Dr. Philip Awadalla

Mawussé Agbessi

Jarry Barber

Vanessa Bruat

Nicholas Cheng

Dr. Marie-Julie Favé

Elias Gbeha

Heather Gibling

Michelle Harwood

Ido Nofech-Mozes

Tom Ouellette

Kimberly Skead

Committee members

Dr. John Dick

Dr. Rayjean Hung

Past lab members

Dr. Armande Ang Houle

Elizabeth Hall

Jasmina Uzunović

Collaborators

Dr. David Soave

OICR Ontario Institute for Cancer Research Collaborate. Translate. Change lives.

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

GenomeCanada

CanPath

CanPath is a partnership between leading health institutes from coast to coast

CanPath

Hosted by:

University of Toronto Dalla Lana School of Public Health

In partnership with:

OICR Ontario Institute for Cancer Research

National Funder:

Canadian Partnership Against Cancer / Partenariat canadien contre le cancer

Regional cohorts:

BC Generations Project

Alberta's Tomorrow Project

Healthy Future Sask

CancerCare Manitoba / Action Cancer Manitoba / The Manitoba Tomorrow Project

Ontario Health Study / Étude sur la santé Ontario

CARTaGENE

Atlantic PATH

Hosted by:

BC Cancer

Alberta Health Services

Saskatchewan Cancer Agency

CancerCare Manitoba / Action Cancer Manitoba

OICR Ontario Institute for Cancer Research

Centre de recherche CHU Sainte-Justine

Dalhousie University

Regional Funders:

Alberta Government

Alberta Cancer Foundation

OICR Ontario Institute for Cancer Research

GenomeQuébec

GenomeCanada

CARTaGENE: A POPULATION-BASED COHORT TO STUDY THE EPIDEMIOLOGY AND GENETICS OF DISEASES IN THE PROVINCE OF QUEBEC (CANADA)

Guillaume Lettre
guillaume.lettre@umontreal.ca
March 10 2025

INSTITUT DE CARDIOLOGIE DE MONTRÉAL

Université de Montréal

Outline

1. Description of CARTaGENE
2. Genetic structure: past and future
3. Clinical genetics: pathogenic variants, allele frequencies, and rare diseases
4. Common diseases: GWAS and polygenic risk scores

CARTaGENE: a cohort to study chronic diseases in Quebec

Mission

CARTaGENE (CaG) is a publicly funded research platform created to accelerate health research and lower costs.

Mandates

Recruitement and follow-up of a population cohort (43,000 participants).

Creation of a database and biobank for health research available to researchers (public/private).

Collected data

Health questionnaire

- Demographic and socio-economic factors
- Lifestyle
- Mental health
- Psychosocial environment
- Individual and familial history of disease
- Use of health care services
- Prescribed medications and other products
- Women's and men's health

Physical measures

- Anthropometric measures
- Blood pressure
- Bioimpedance
- Grip strength
- Bone density
- Arterial stiffness
- Lung function
- Partial resting electrocardiogram (4 lead ECG)
- Cognitive function

Nutrition

- 165 questions about intake in the past year (Canadian Diet History Questionnaire II)
- Provides nutrient and food group intake estimates

Environment

- Full residential and occupational histories
- Residential and occupational exposures

Other data

- MRI: Heart and brain
- COVID-19: Health follow-up and serology

Collected biospecimens

Biological samples

- Whole blood
- Plasma
- Serum
- Red blood cells
- Whole blood DNA
- Tempus tubes for RNA extraction
- Urine

Biochemical and haematological measures

- Complete blood count
- Lipid profile, electrolytes, glucose, glycated hemoglobin, ALT, AST, GGT, creatinine, uric acid, albumin, Free T4, TSH

Genomics

- Whole-Genome Genotyping Data (Infinium® illumina Global Screening Array, illumina Omni 2.5M, Affymetrix UK biobank Axiom® Array) (30,000 participants)
- Whole Genome Sequencing Data (2000 participants)
- RNA-Seq Data (1000 participants)
- Whole Exome Sequencing (WES) (200 participants)

CARTaGENE unique features

1. **Administrative health data** linkage (e.g. hospitalisation, medical diagnosis, medications, deaths, cancer registries, etc.) from 1998 onward

2. **Broad consent** for use of data & samples + possibility of recontact

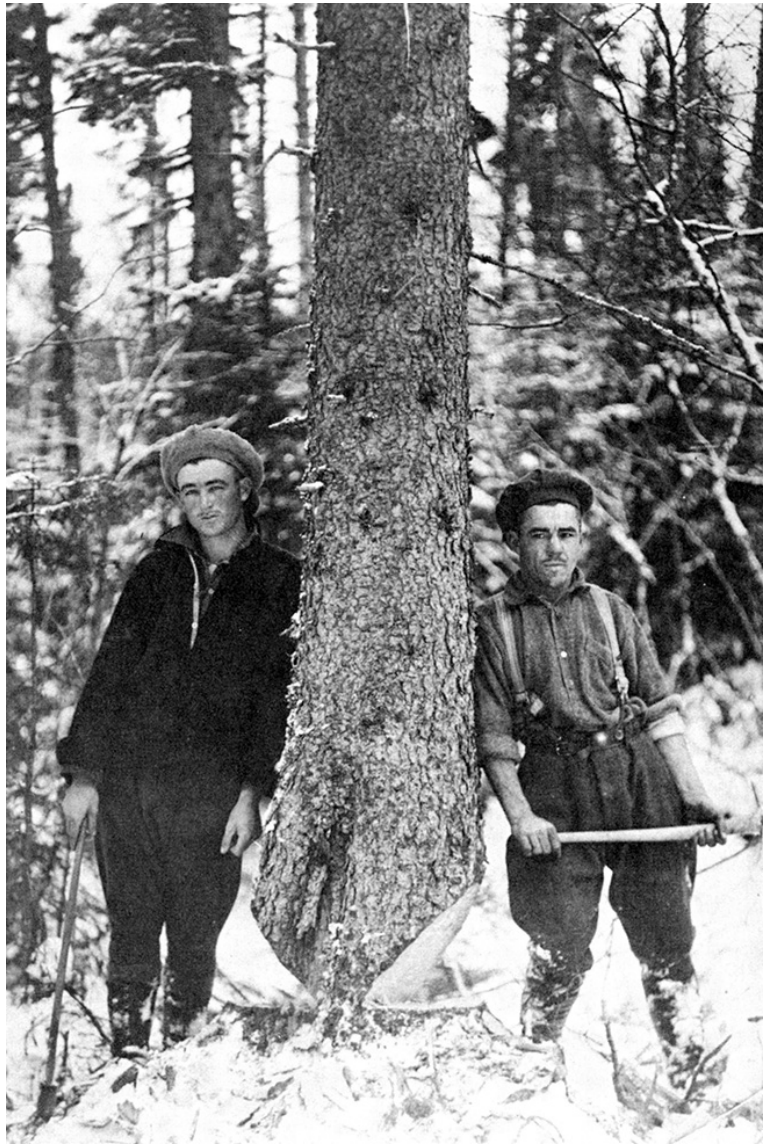
3. **Genealogical** reconstructions – Balsac (UQAC)

French Canadians

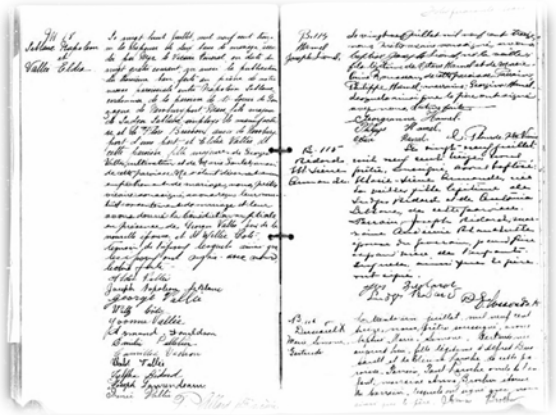
Quebec has a population of 8.6 million individuals, of which approximately 7.3 million are French Canadians.

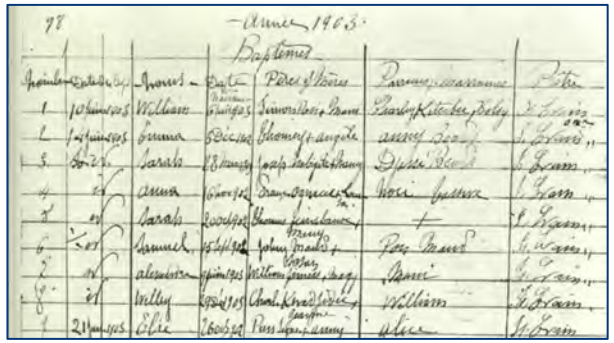
The majority of French-Canadian ancestry is derived from ~8,500 settlers from France in the 17th and 18th centuries.

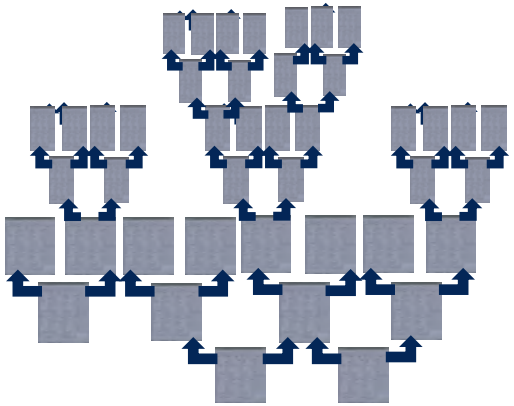
French Canadians carry on average less than 1% of ancestry tracing back to indigenous populations and the rest is mostly attributed to French ancestry.



Balsac: the genealogy of Quebec's French Canadians







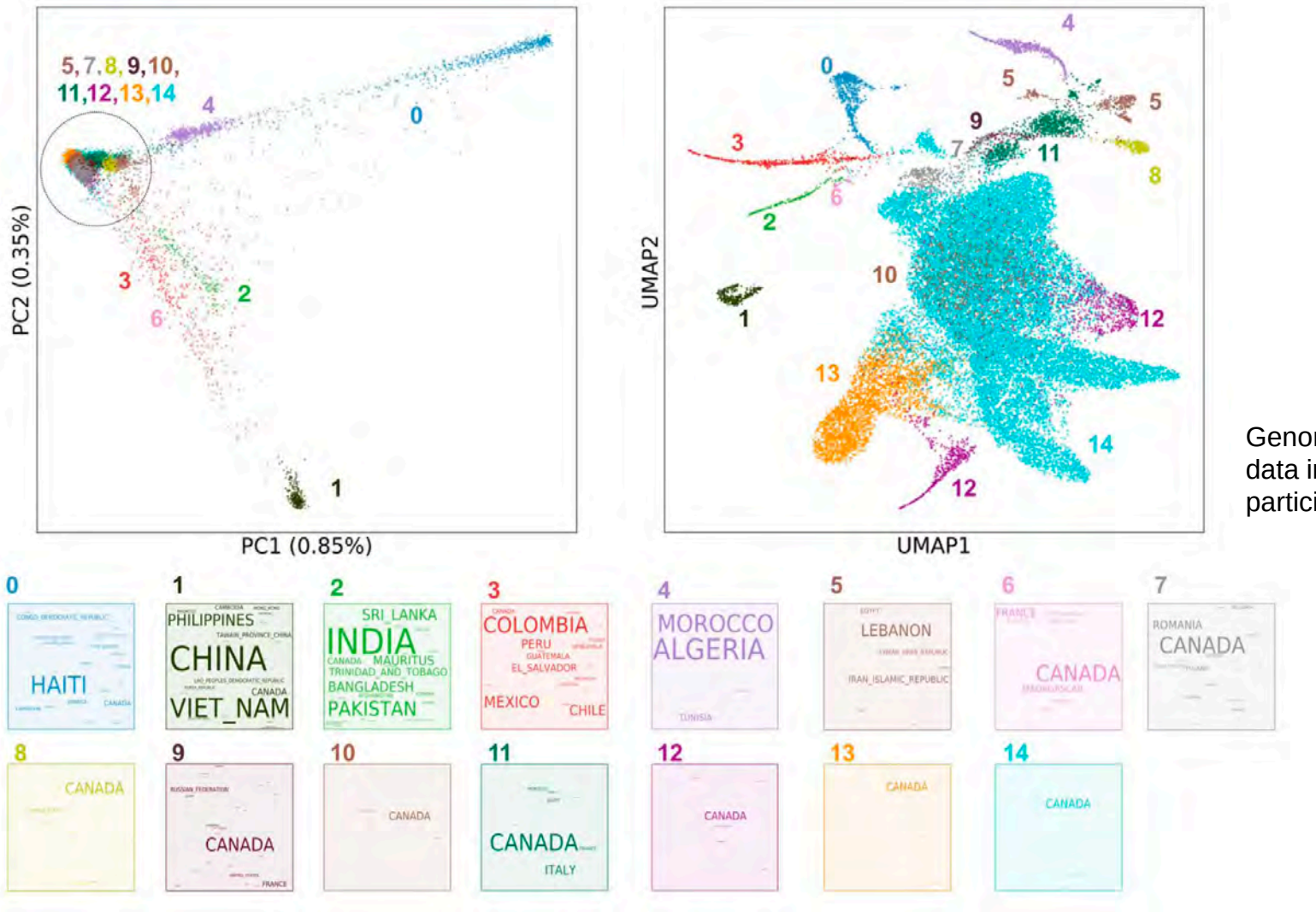
BALSAC

Marriage certificates from the beginning of European settlement in the 17th century to the contemporary period.

Outline

1. Description of CARTaGENE
2. Genetic structure: past and future
3. Clinical genetics: pathogenic variants, allele frequencies, and rare diseases
4. Common diseases: GWAS and polygenic risk scores

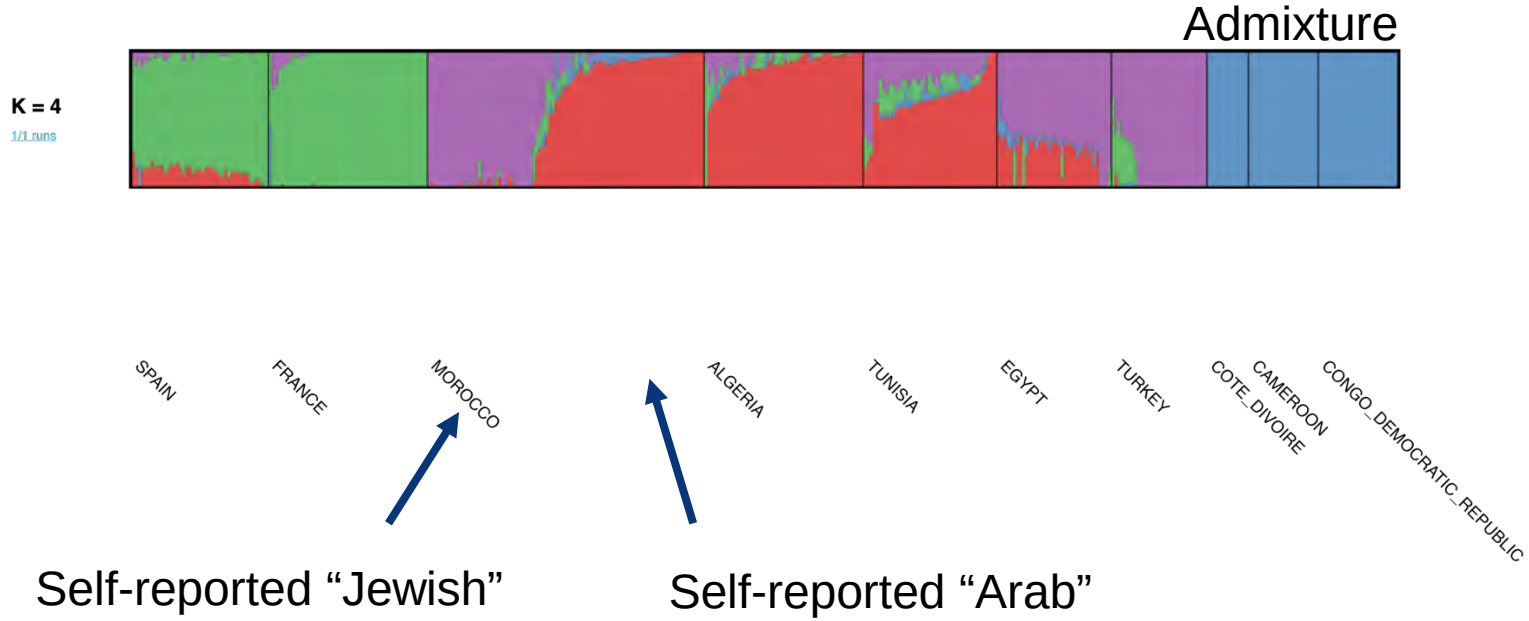
Genetic structure in CARTaGENE

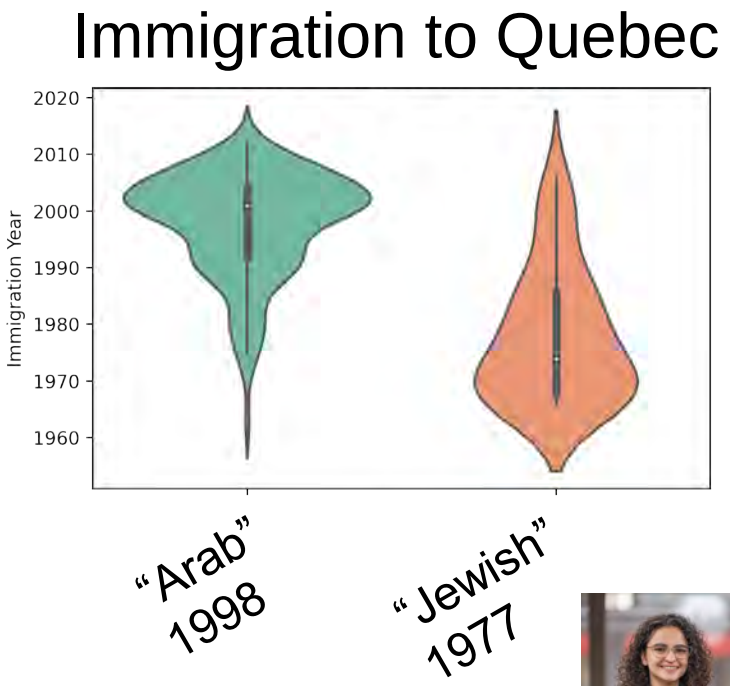


Genome-wide genotyping data in ~30K CaG participants

Alex Diaz-Papkovich

CARTaGENE participants with 4 Moroccan grandparents (MO)

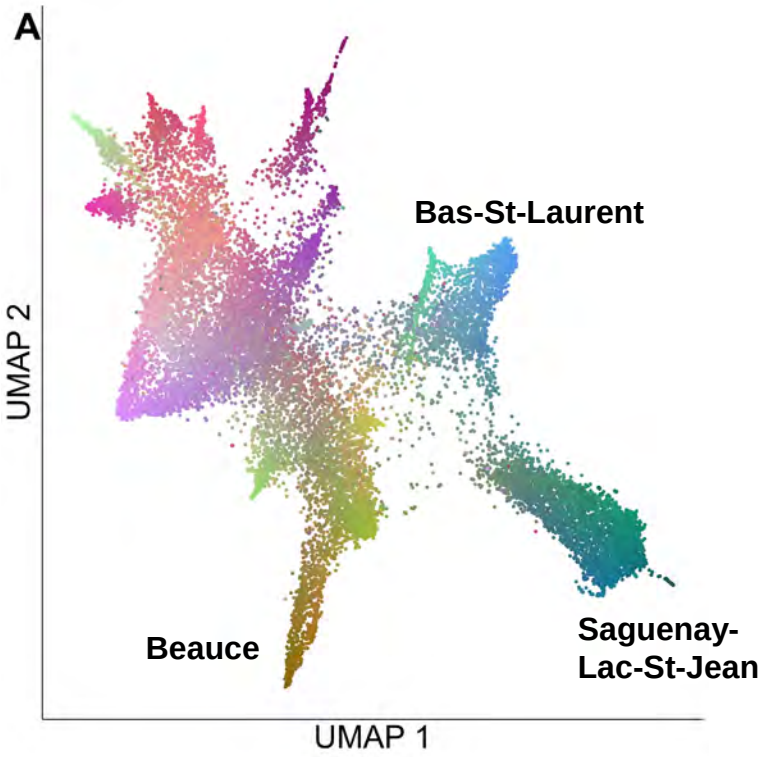


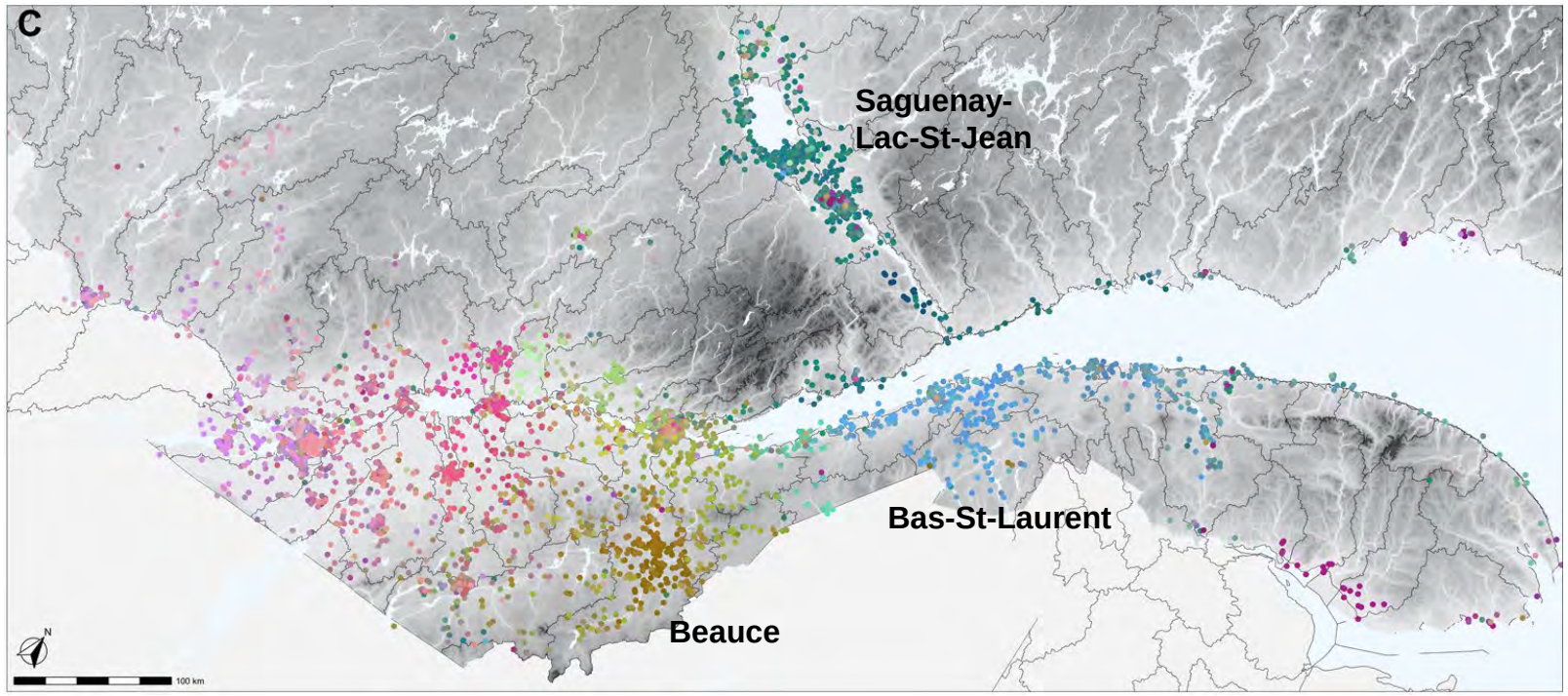


Georgette Femerling Romero

Regional distribution of genetic variation in French Canadians

Luke Anderson-Trocme





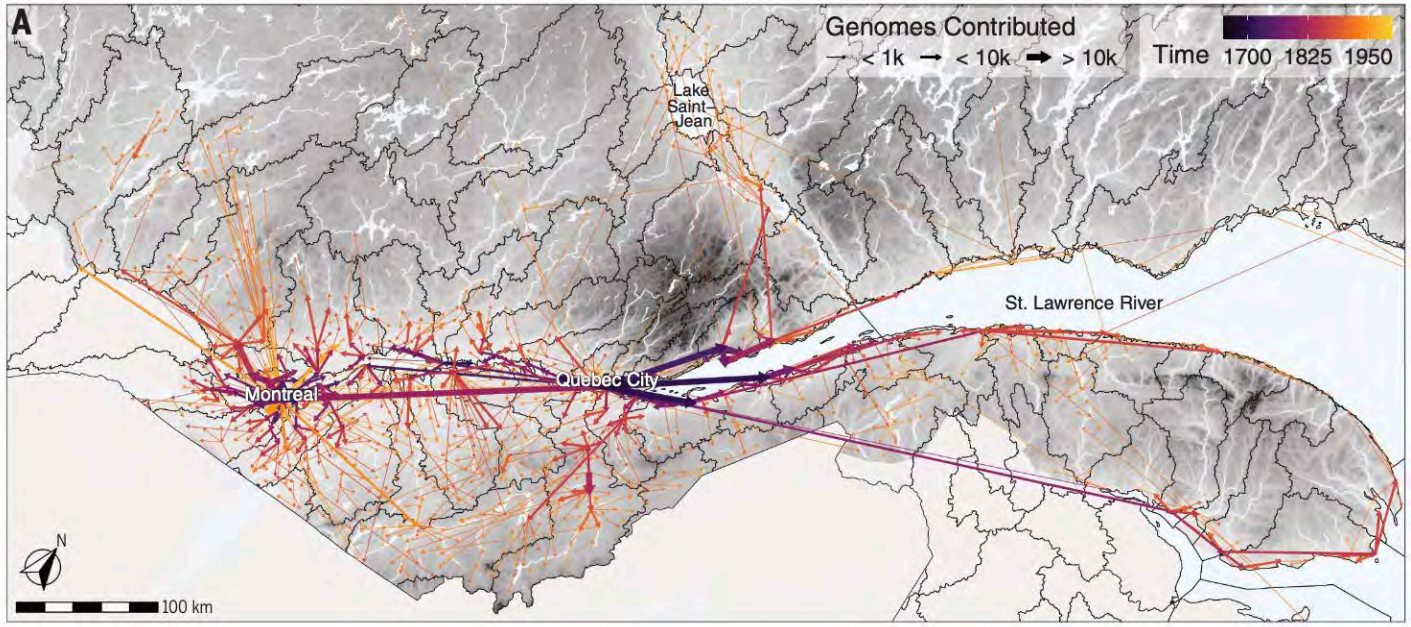
Geography influenced French-Canadian migration

RESEARCH

HUMAN GENETICS

On the genes, genealogies, and geographies of Quebec

Luke Anderson-Trocmé[1,2], Dominic Nelson[1,2], Shadi Zabad[3], Alex Diaz-Papkovich[1,4], Ivan Kryukov[1,2], Nikolas Baya[5], Mathilde Touvier[6], Ben Jeffery[5], Christian Dina[7], Hélène Vézina[8], Jerome Kelleher[5], Simon Gravel[1,2]*



Primary routes of estimated genetic ancestry inferred from genealogy

Anderson-Trocme, …, Gravel. *Science* 2023

Outline

1. Description of CARTaGENE
2. Genetic structure: past and future
3. Clinical genetics: pathogenic variants, allele frequencies, and rare diseases
4. Common diseases: GWAS and polygenic risk scores

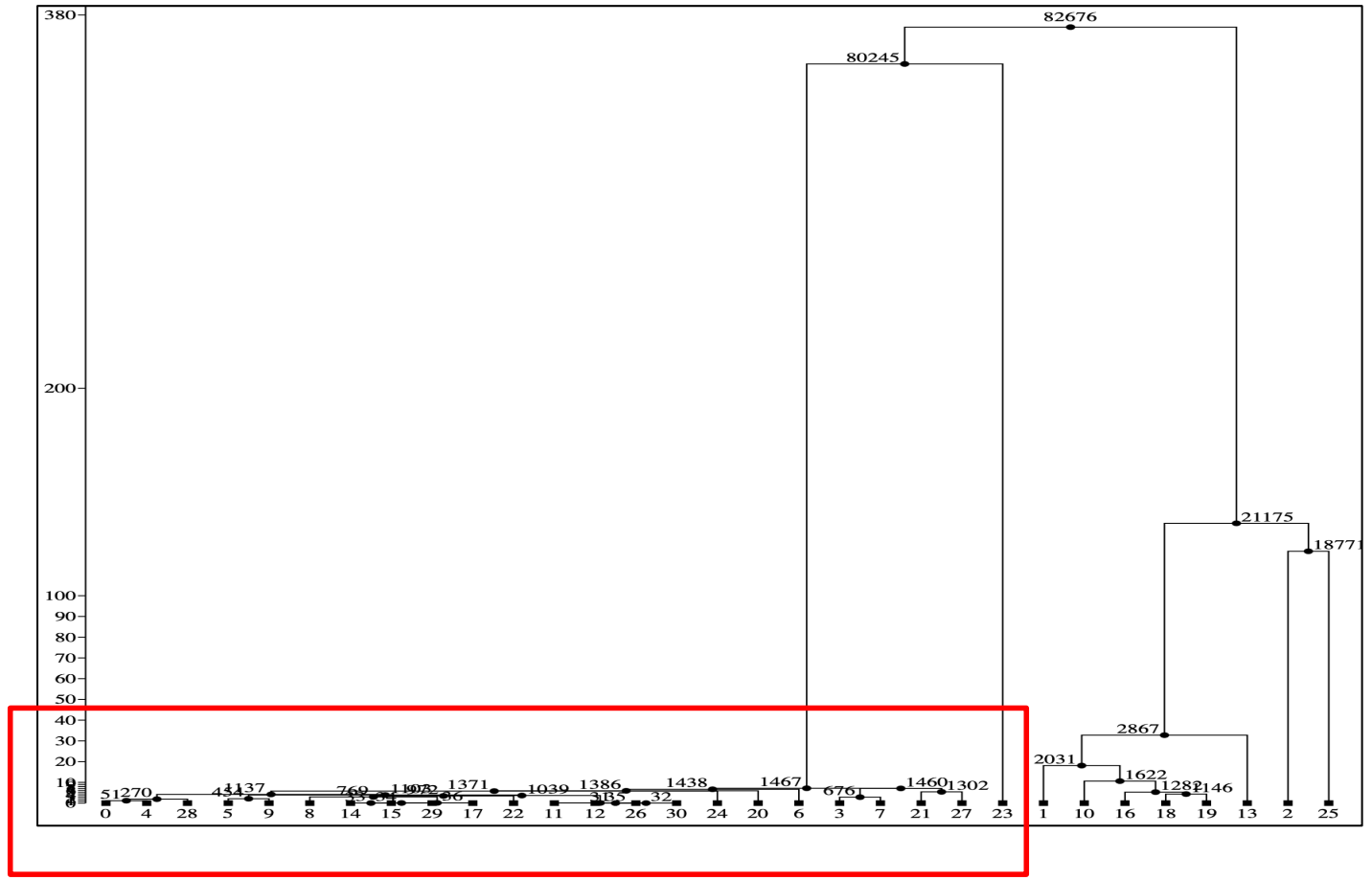
Example of the founder effect in clinical genetics

- **Primary ciliary dyskinesia (PCD)** is an autosomal recessive disorder (prevalence 1/15,000).

- It is characterized by early onset of a progressive decline in lung function due to an inability to clear mucus and particles from the airways.

- Shapiro and colleagues found a mutation in *HYDIN* (p.Arg3476Ter) in several French-Canadian families with PCD.

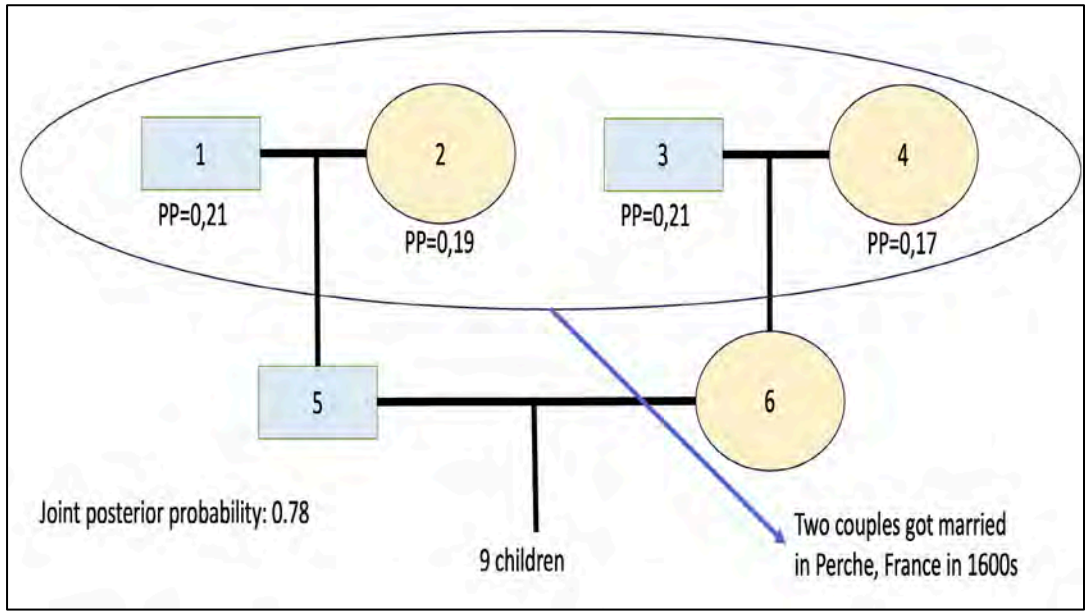
Reconstruction of the ancestral recombination graph for the 31 CARTaGENE carriers of the *HYDIN* mutation

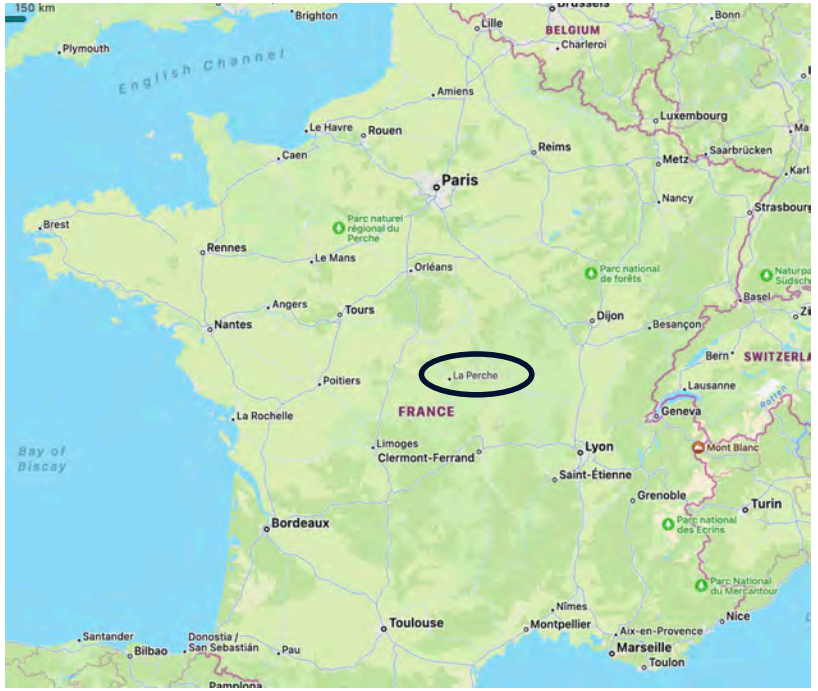


22/31 carriers coalesce within the first 10 generations

Alejandro Mejia Garcia

Genealogy links the *HYDIN* mutation to two couples from Perche, France



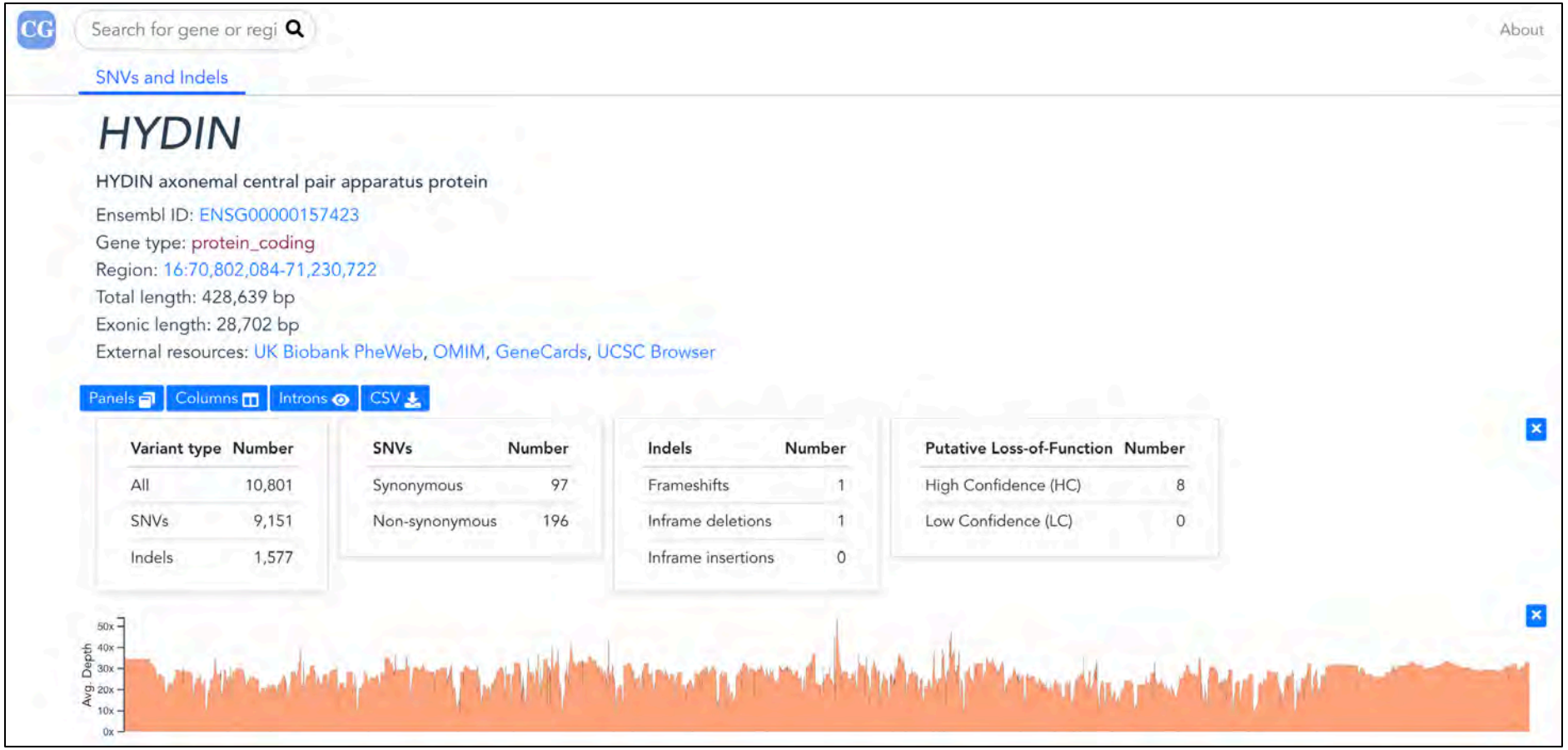


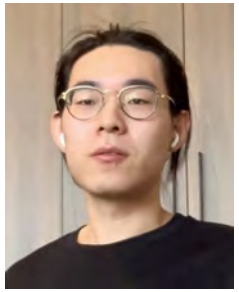
Whole-genome sequencing (WGS)

- **High-coverage (33.4X) short-read WGS of 2,171 CARTaGENE participants:**
  - 1,762 self-identified French Canadians (FC)
  - 163 participants with four Haitian grandparents (HA)
  - 127 participants with four Moroccan grandparents (MO)

- **80,407,530 single-nucleotide variants and insersion-deletions**
  - 16.8% novel variants
  - 38.6% singletons

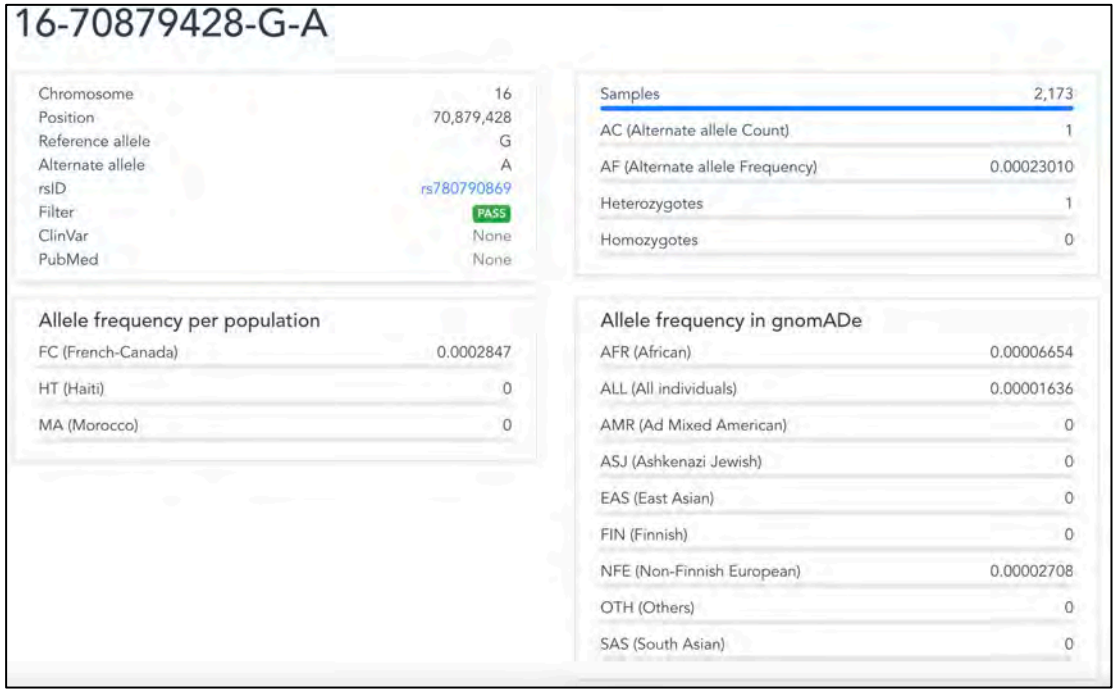
CARTaGENE allele frequency browser





Hongyu Xiao

Pathogenic variants in the Quebec population



16-70879428-G-A

Chromosome	16
Position	70,879,428
Reference allele	G
Alternate allele	A
rsID	rs780790869
Filter	PASS
ClinVar	None
PubMed	None

Samples	2,173
AC (Alternate allele Count)	1
AF (Alternate allele Frequency)	0.00023010
Heterozygotes	1
Homozygotes	0

Allele frequency per population

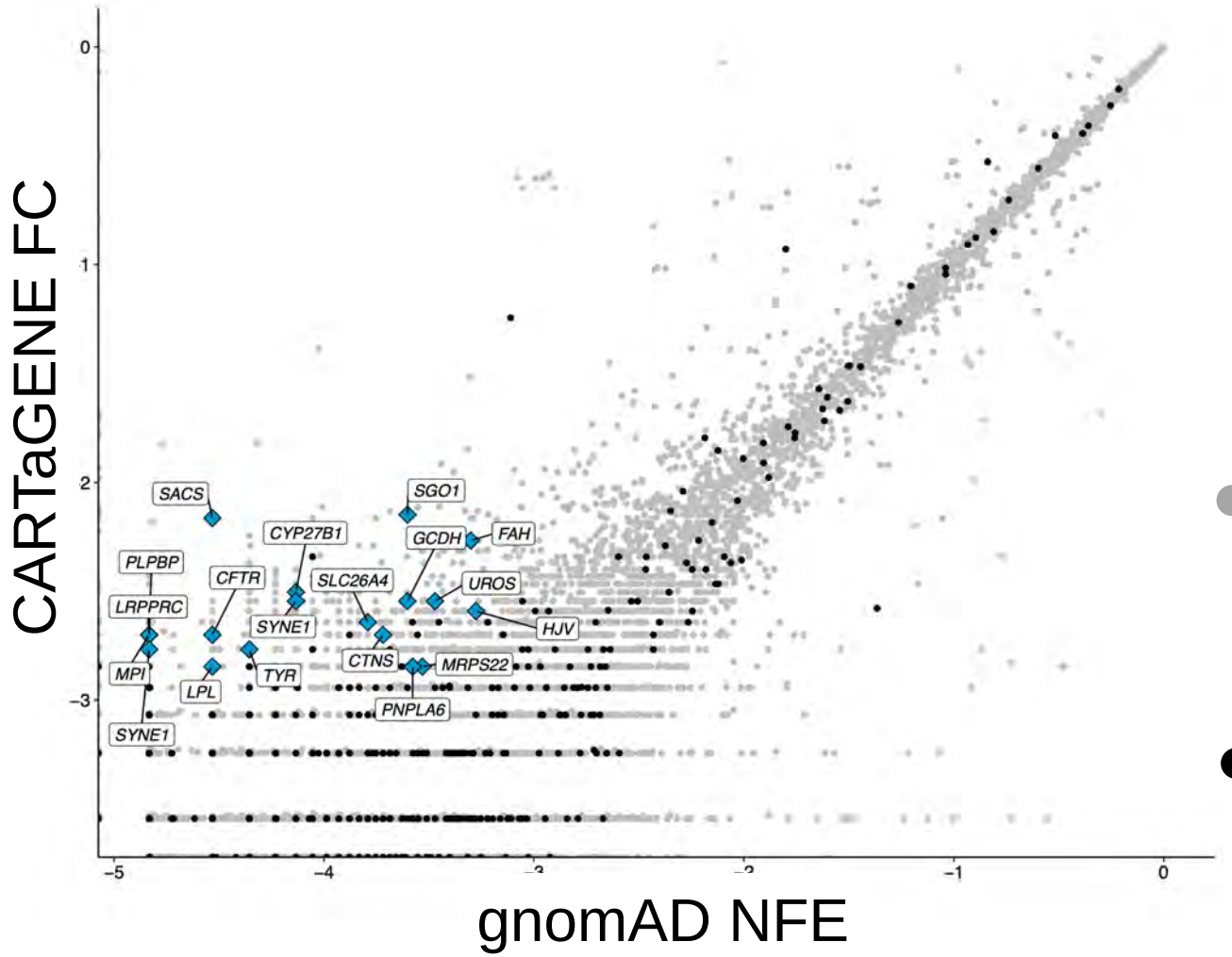
FC (French-Canada)	0.0002847
HT (Haiti)	0
MA (Morocco)	0

Allele frequency in gnomADe

AFR (African)	0.00006654
ALL (All individuals)	0.00001636
AMR (Ad Mixed American)	0
ASJ (Ashkenazi Jewish)	0
EAS (East Asian)	0
FIN (Finnish)	0
NFE (Non-Finnish European)	0.00002708
OTH (Others)	0
SAS (South Asian)	0

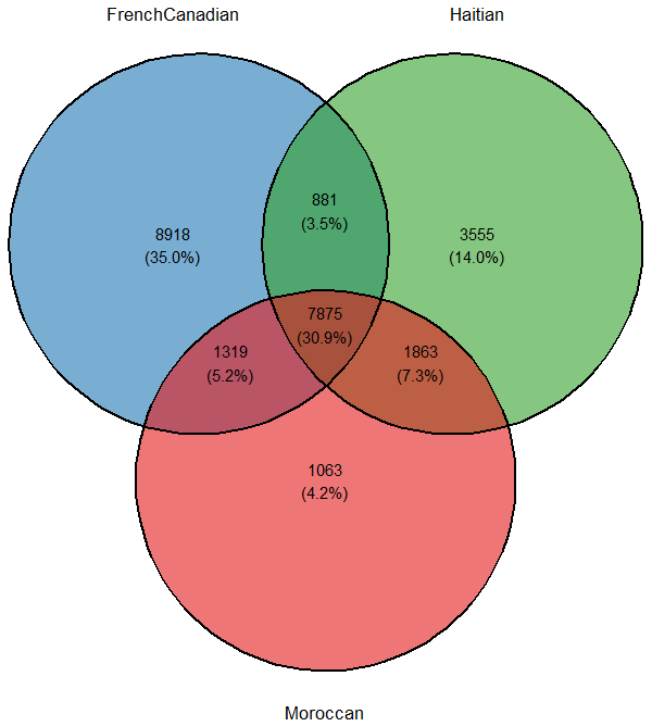
HYDIN p.Arg3476Ter
(PCD)

Allele frequency of founder mutations

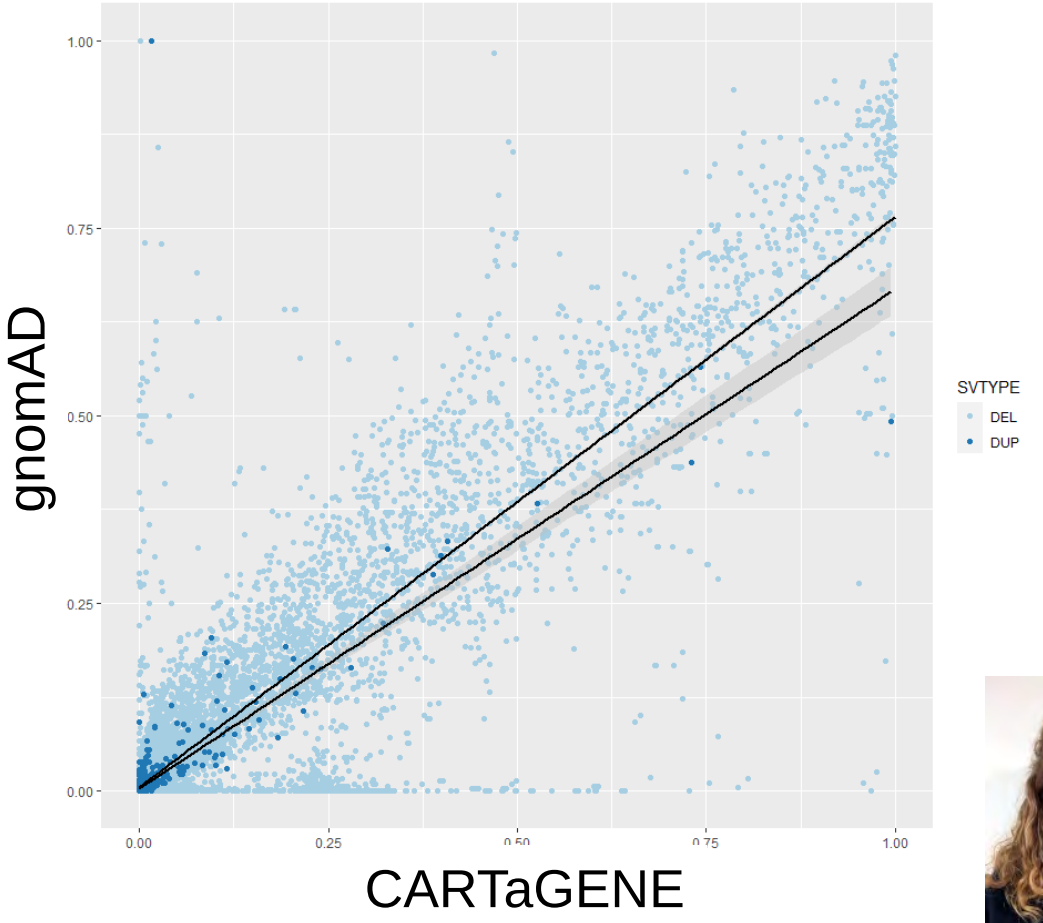


- High-impact variants: nonsense, frameshift indels, essential splice site
- ClinVar pathogenic

Structural variants (called from WGS)



24,022 deletions and 1,798 duplications



Rose Laflamme

LDLR deletion is the main cause of familial hypercholestorelemia in French Canadians

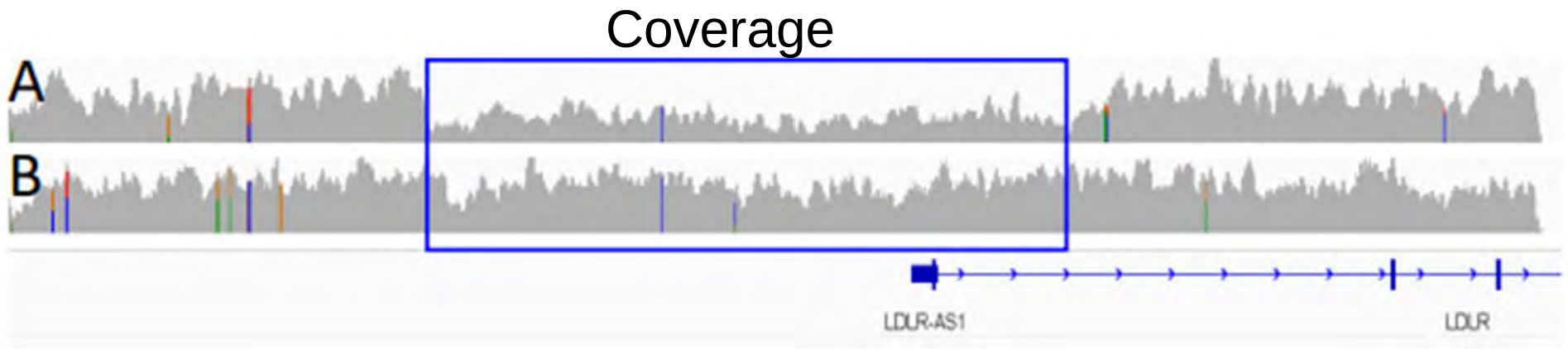
734 THE NEW ENGLAND JOURNAL OF MEDICINE Sept. 17, 1987

DELETION IN THE GENE FOR THE LOW-DENSITY-LIPOPROTEIN RECEPTOR IN A MAJORITY OF FRENCH CANADIANS WITH FAMILIAL HYPERCHOLESTEROLEMIA

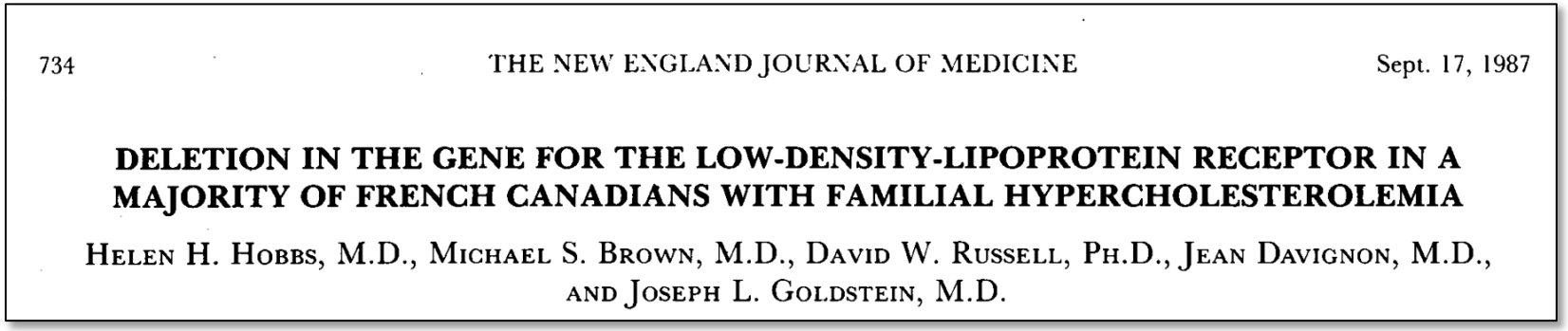
HELEN H. HOBBS, M.D., MICHAEL S. BROWN, M.D., DAVID W. RUSSELL, PH.D., JEAN DAVIGNON, M.D., AND JOSEPH L. GOLDSTEIN, M.D.

Heterozygous

Homozygous for the reference



LDLR deletion is the main cause of familial hypercholestorelemia in French Canadians



734 THE NEW ENGLAND JOURNAL OF MEDICINE Sept. 17, 1987

DELETION IN THE GENE FOR THE LOW-DENSITY-LIPOPROTEIN RECEPTOR IN A MAJORITY OF FRENCH CANADIANS WITH FAMILIAL HYPERCHOLESTEROLEMIA

HELEN H. HOBBS, M.D., MICHAEL S. BROWN, M.D., DAVID W. RUSSELL, PH.D., JEAN DAVIGNON, M.D., AND JOSEPH L. GOLDSTEIN, M.D.

6 carriers of the *LDLR* +15kb deletion
Deletion frequency in French Canadians: 0.14%
Association with LDL-cholesterol $P=4.8 \times 10^{-5}$ (+1.34 mmol/L)

Other pathogenic structural variants

- **Osteogenesis imperfecta**
  - Deletion exon 9 *P3H1* (506 bp)
  - 4 carriers
- **Nephropatic cystinosis**
  - Deletion *SHPK* + exons 1-9 *CTNS* (57.1kb)
  - 2 carriers
- **Tay-Sachs disease**
  - Deletion exon 1 *HEXA* (7.9kb)
  - 1 carrier
  - 10 times more frequent in French Canadians
  - ~80% of this mutation are found in French-Canadian patients

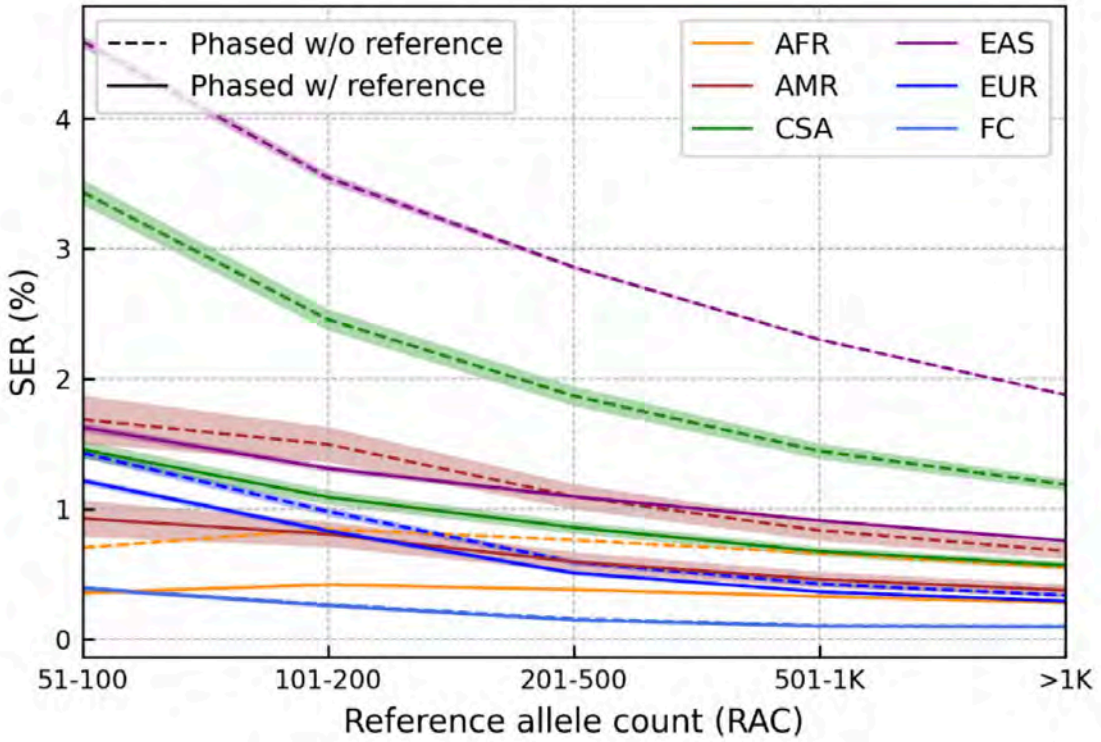
Outline

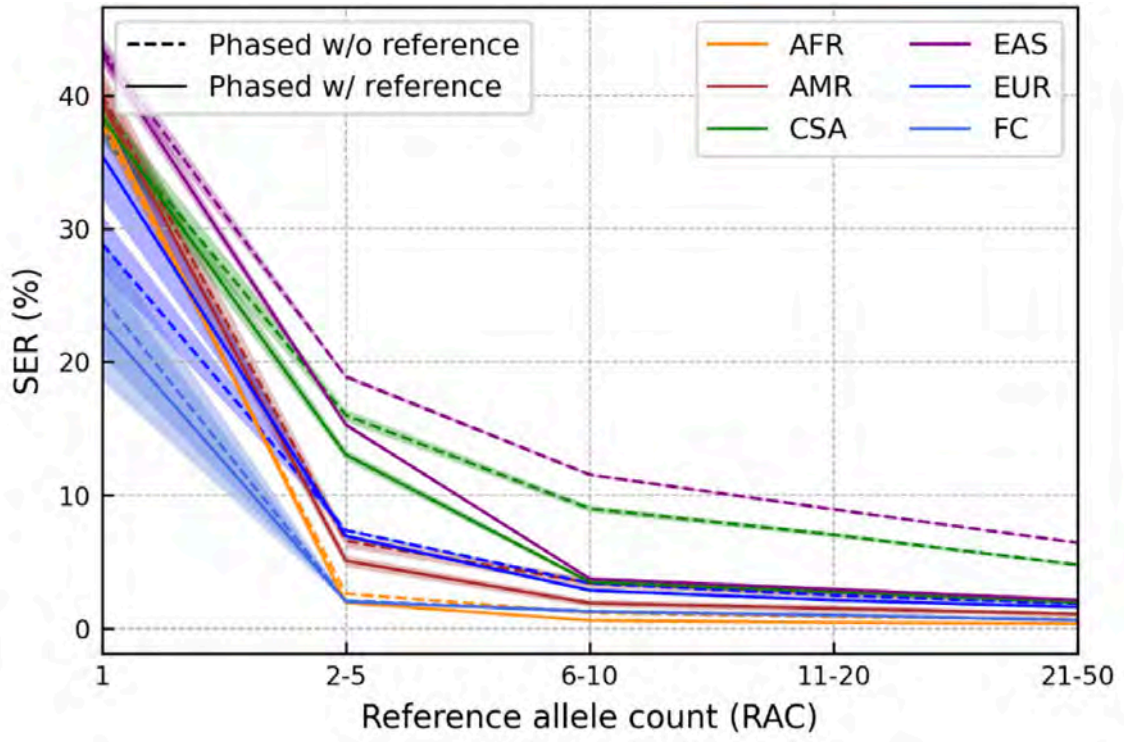
1. Description of CARTaGENE
2. Genetic structure: past and future
3. Clinical genetics: pathogenic variants, allele frequencies, and rare diseases
4. Common diseases: Genotype imputation and GWAS

Building a new panel of reference haplotypes

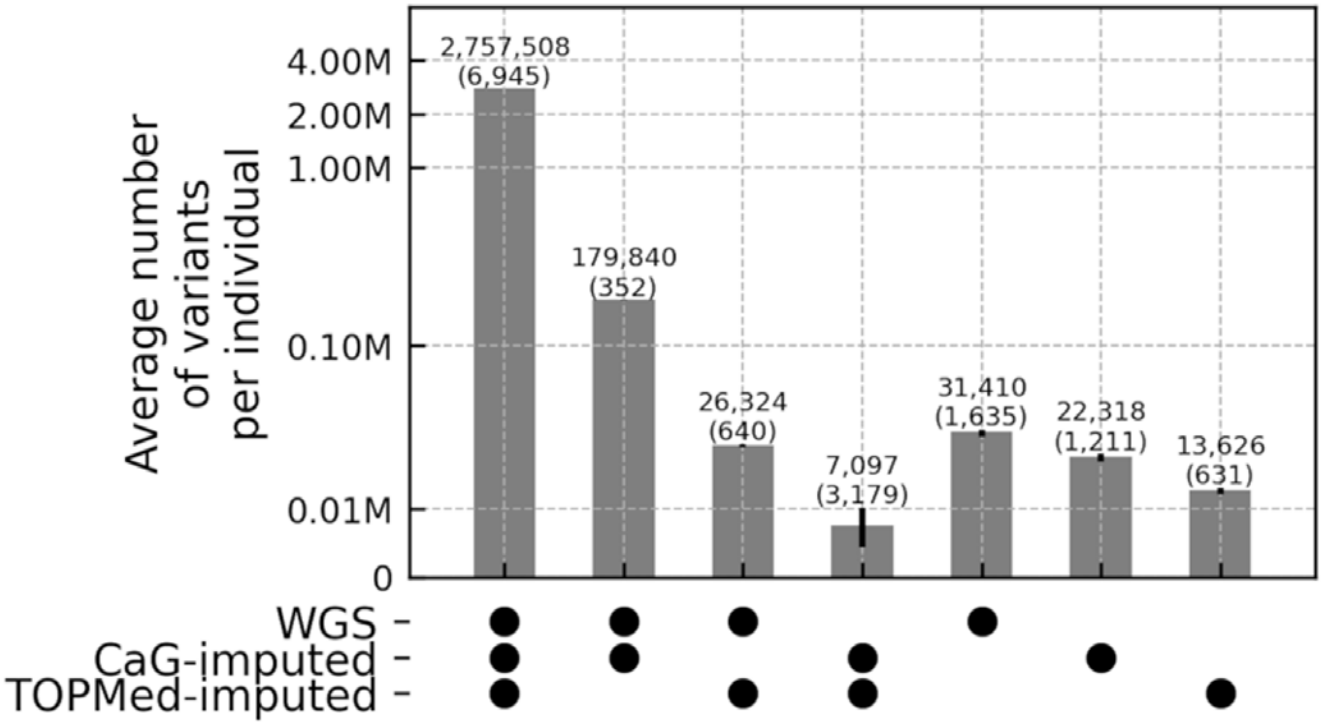
Daniel Taliun

Quality of the phasing data using independent WGS data
(SER=switch error rate)



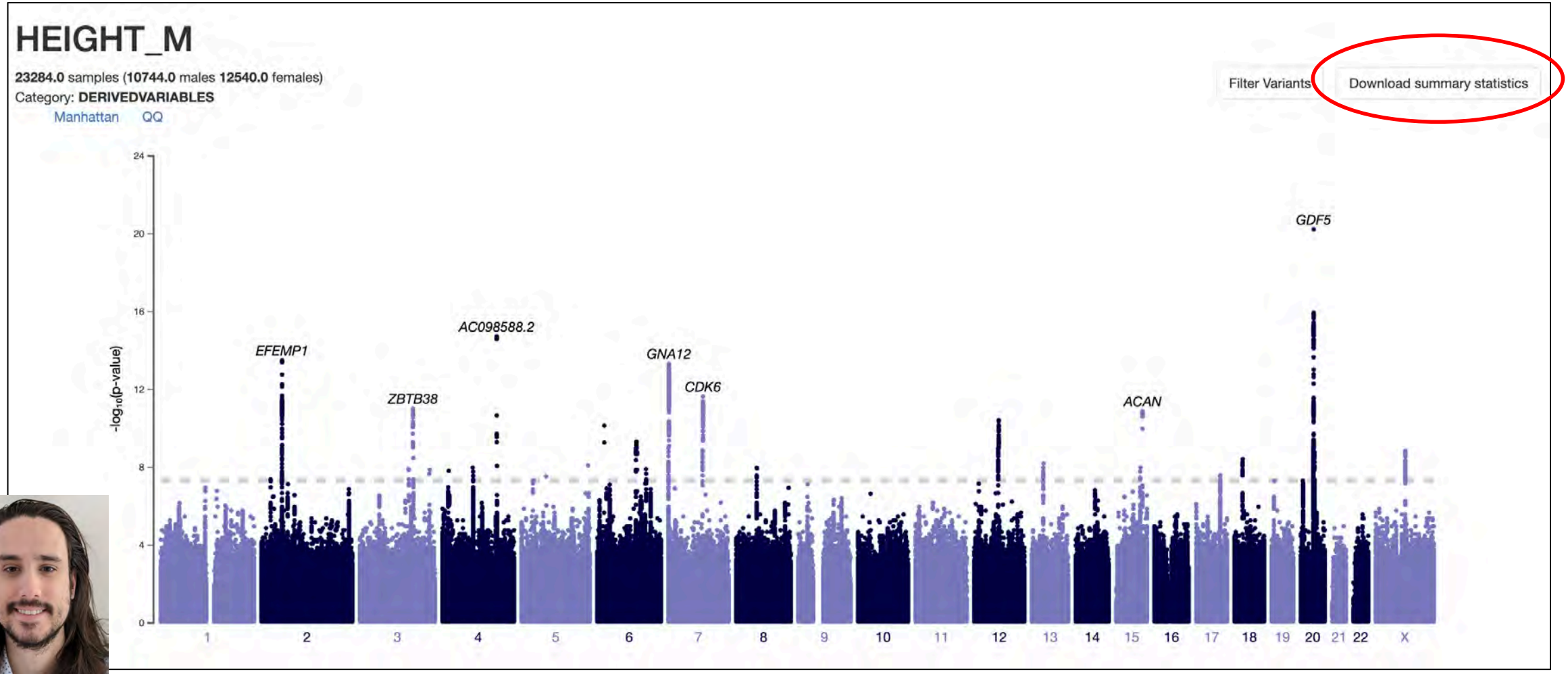


Imputation quality using the CaG panel



Calibration using WGS data from 141 independent non-CaG FC

GWAS results for 100s of phenotypes (PheWeb)



Vincent Chapdelaine

Utility of the CARTaGENE GWAS results

Because the CARTaGENE data is recent, it has not been used in most published GWAS meta-analyses. This opens many **opportunities**:

1. **Replication** of new GWAS hits
2. Derivation of new instruments for **Mendelian randomization** studies
3. Calibration of existing **polygenic risk scores (PRS)**

Strategy to make new GWAS discoveries

Ken Sin Lo

80 M variants identified by WGS

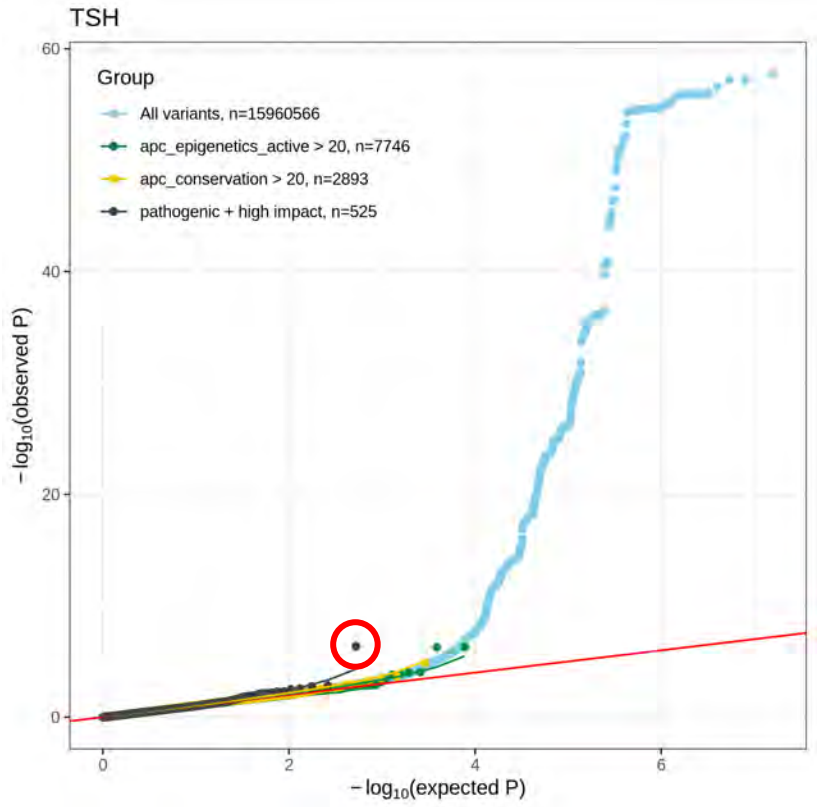
12 M variants enriched in French Canadians (ratio AF_{FC} / AF_{gnomAD_NFE} >4)

530 high-impact variants
(nonsense, frameshift indels, essential splice sites, ClinVar)

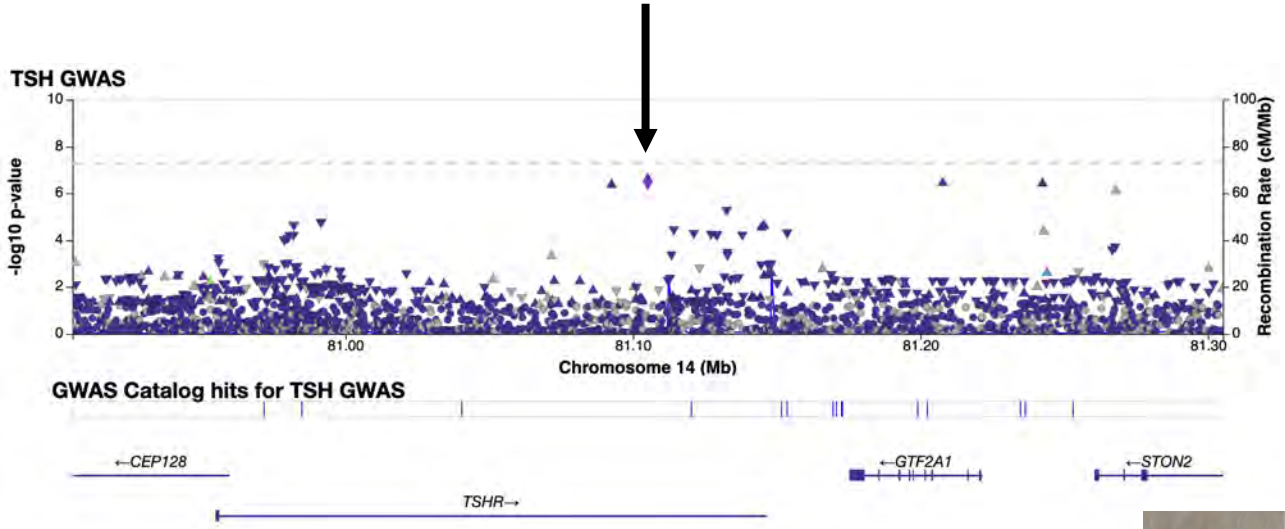
3400 conserved variants
FAVOR $aPC_{conserved}$ score >15

8600 epigenetically active variants
FAVOR $aPC_{epigenetics}$ score >15

New variant associated with thyroid-stimulating hormone (TSH) levels

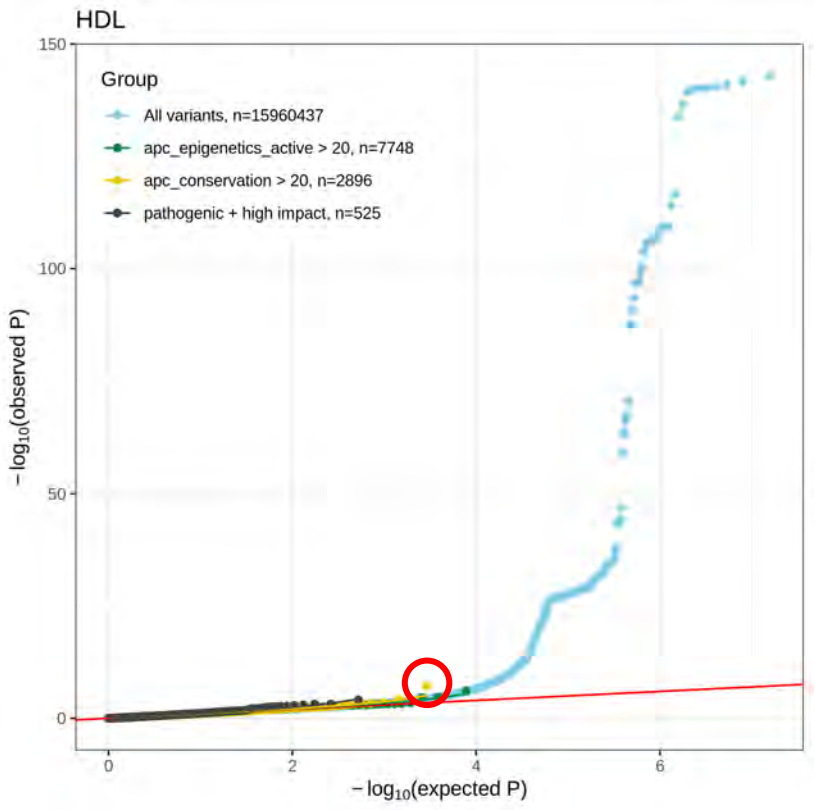


rs121908863 (p.P162A)
MAF_CaG=0.14%; MAF_gnomAD$_{NFE}$=0.029% (4.8X)
P_{CaG}=4.3x10^{-7}
$P_{replication\_CLSA}$=9.8x10^{-6}

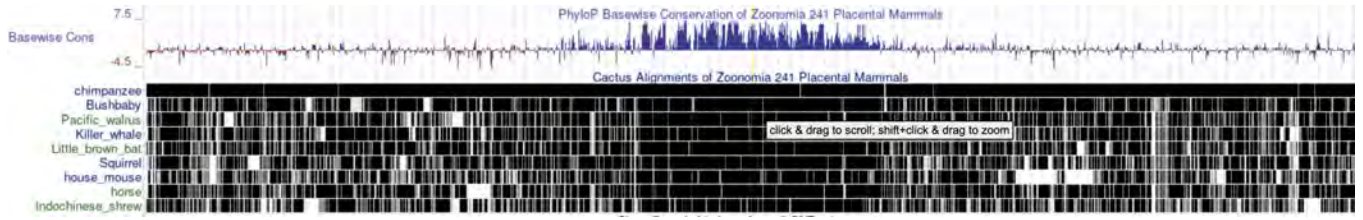


Justin Bellavance

New variants associated with HDL-C



rs548046676
MAF_CaG=0.34%, MAF_NF=0.062% (5.5X)
$P_{CARTaGENE}=5.4x10^{-8}$
$P_{replication\_CLSA}=0.023$



No protein-coding genes in 600-kb window
ABCA1 is 3.1-Mb away

Conclusions

1. Despite existing major players (*e.g.* UKBB, BBJ), **population-specific cohorts** remain relevant and important.
2. CARTaGENE is a critical component to develop **precision medicine initiatives** in Quebec (and Canada) for both rare and common diseases.
3. CARTaGENE can also enable genetic (and epidemiological/public health) discoveries with **global impact**.
4. CARTaGENE is **open** for business!

How to access the CARTaGENE data

Data Access: Rapid, simple, efficient

INFORMATIONS
access@cartagene.qc.ca
514-345-2156

VISIT OUR WEB SITE
www.cartagene.qc.ca

FOLLOW US

Acknowledgments – CARTaGENE genetics

Rose Laflamme

Ken Sin Lo

Simon Gravel

Martine Tétreault

Luke Anderson-Trocme

Jean-Christophe Grenier

Peyton McClelland

Alejandro Mejia Garcia

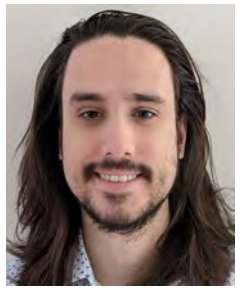
Georgette Femerling

Justin Bellavance

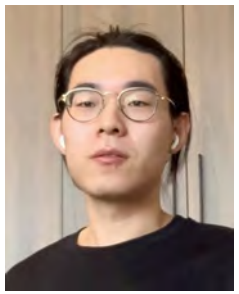
Sarah Gagliano-Taliun

Justin Pelletier

Claude Bhérer



Vincent Chapdelaine



Hongyu Xiao

Mohadese Sayahian Dehkordi

Daniel Taliun

Also: Alex Diaz-Papkovich

CARTaGENE team & funding

Operations

Jean-Baptiste Rivière – Director

Catherine Labbé – Scientific Affairs and Access

Mengting Xu – Epidemiologist

Julie Bergeron – Epidemiologist

Frédéric Latour – Data Manager

Georgette Romero - Bioinformatician

Maude Handsbury – Administrative assistant

Charles Rivard – Finances

Émilie Harvey – Participant Personal information

Scientific leadership

Vikki Ho – Scientific co-director

Simon Gravel – Scientific co-director

Guillaume Lettre – Scientific co-director

Funding

Genome Quebec & Genome Canada

CHU Ste-Justine

Canadian Institutes of Health Research

Canada Research Chair Program

Montreal Heart Institute Foundation

http://www.mhi-humangenetics.org
guillaume.lettre@umontreal.ca

Postdoc positions available to study the genetics and genomics of heart and blood diseases

CanPath.ca

CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain

ACCESS OFFICE HOURS

CanPath 15

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Discover how HDRN Canada and CanPath can help you navigate multi-regional data access through DASH


CanPath 15
lifebit
SYNTHETIC DATASET WORKSHOP
Teaching & Researching with Health Data in a Trusted Research Environment
Tuesday, June 24
University of Toronto

CanPath.ca

CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain