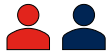
Genomic and Environmental Influences on Canadian Health Phenotypes and Chronic Disease Outcomes

Philip Awadalla
National Scientific Director, Canadian Partnership for Tomorrows Health
Director, Computational Biology, Ontario Institute for Cancer Research
Professor, University of Toronto

July 18th, 2024

CanPath
Canadian Partnership for Tomorrow's Health

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**

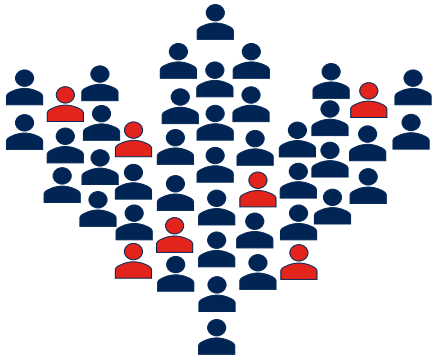
Questions that can be answered:

- 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 is Canada's largest population health research platform

CanPath provides a national platform for population-level health research in Canada and globally.



- CanPath collects **real-world data from one in every 100 Canadians** to enable discovery and innovation in **disease detection, treatment, control** and **prevention**
- Over the past decade, **CanPath has brought together scientists across Canada** and **leveraged over $208 million in investments** to create the nation's largest population cohort and biobank
- Canada has a unique opportunity to leverage the CanPath platform to **advance government priorities** and **build a healthier Canada**

CanPath is following the health of over 330,000 adult Canadians for decades

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

CanPath

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. Riaz Alvi
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

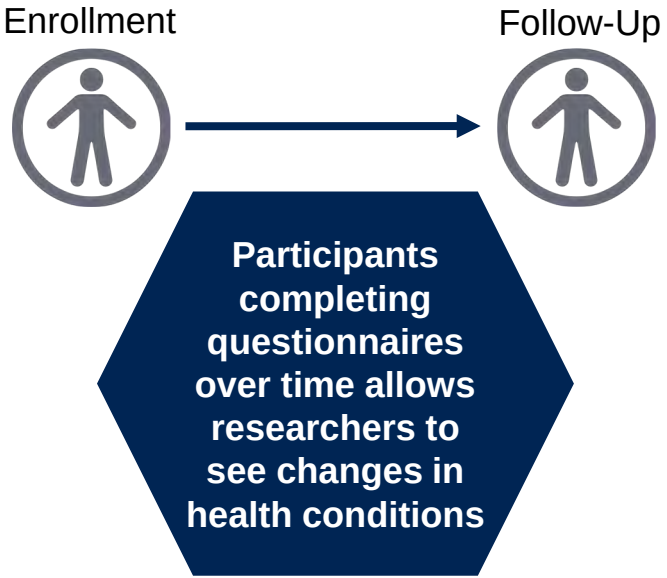
Dr. Guillaume Lettre
Scientific Co-Director, CARTaGENE

Dr. Robin Urquhart
Scientific Director, Atlantic PATH

Mr. Jason Hicks
Executive Director, Atlantic PATH

Over 330,000 participants have completed detailed questionnaires over the years

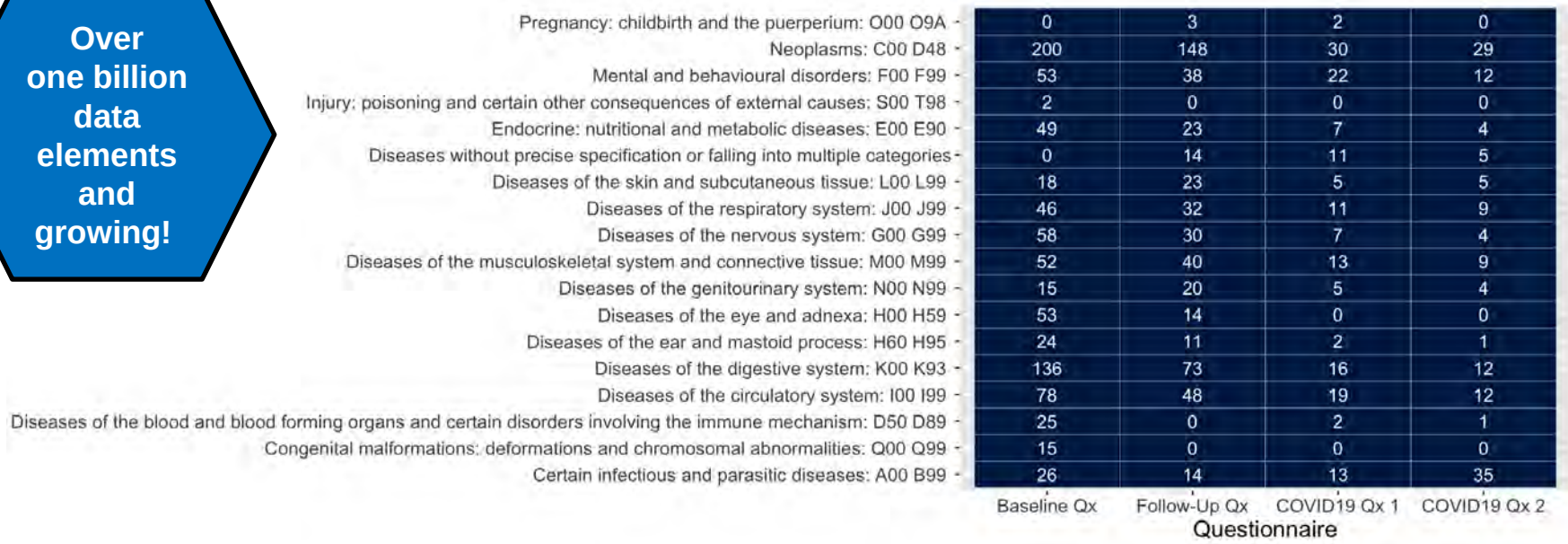
- Participant demographics
- Changes in health status
- Mental Health
- Medical history
- Prescribed medication
- Family health history
- Anthropometric measurements
- Working status
- Household income
- Behaviours (sleep, alcohol, tobacco, marijuana use, and e-cigarette use)



CanPath has collected over 1,600 variables on disease outcomes across multiple timepoints

Over one billion data elements and growing!

Number of variables collected across disease categories



- Collected detailed information on vascular disease, cardiac disease and cognitive function using MRI scans
- Data collected from 10,000 Canadians through existing cohorts, including **1,500 First Nations people** living in Canada
- These unique data are being used to evaluate the impact of different **environmental determinants on cardiovascular health**

With **Dr. Sonia Anand, Dr. Matthias Friedrich and the late Dr. Jack Tu.**

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 requesting administrative health data along with cohort data and/or samples.

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 location.

Canadian Institute for Health Information

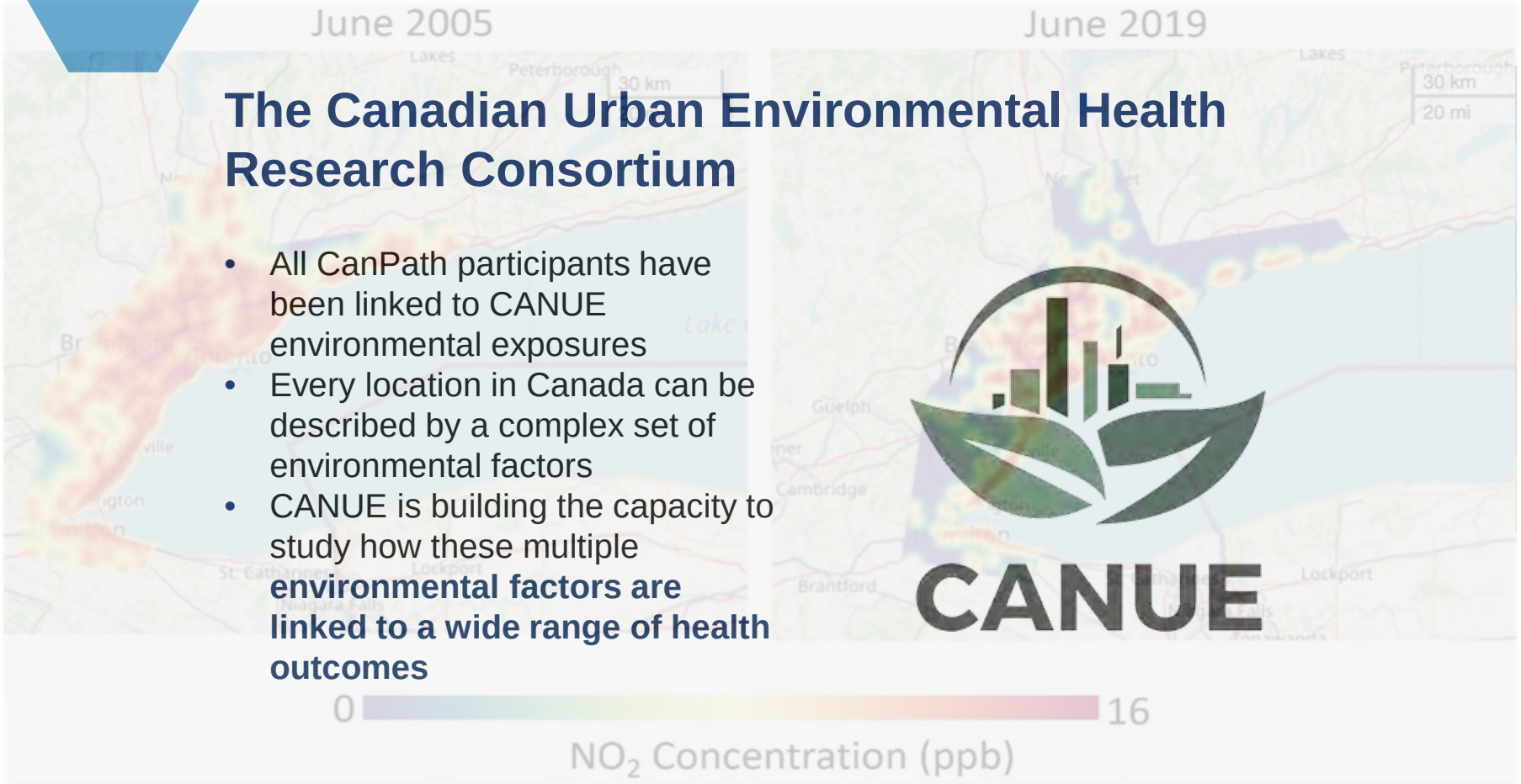
Institut canadien d'information sur la santé

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

The Canadian Urban Environmental Health Research Consortium

- All CanPath participants have been linked to CANUE environmental exposures
- Every location in Canada can be described by a complex set of environmental factors
- CANUE is building the capacity to study how these multiple **environmental factors are linked to a wide range of health outcomes**



CanPath

Volume 616 Issue 7955, 6 April 2023

Subscribe

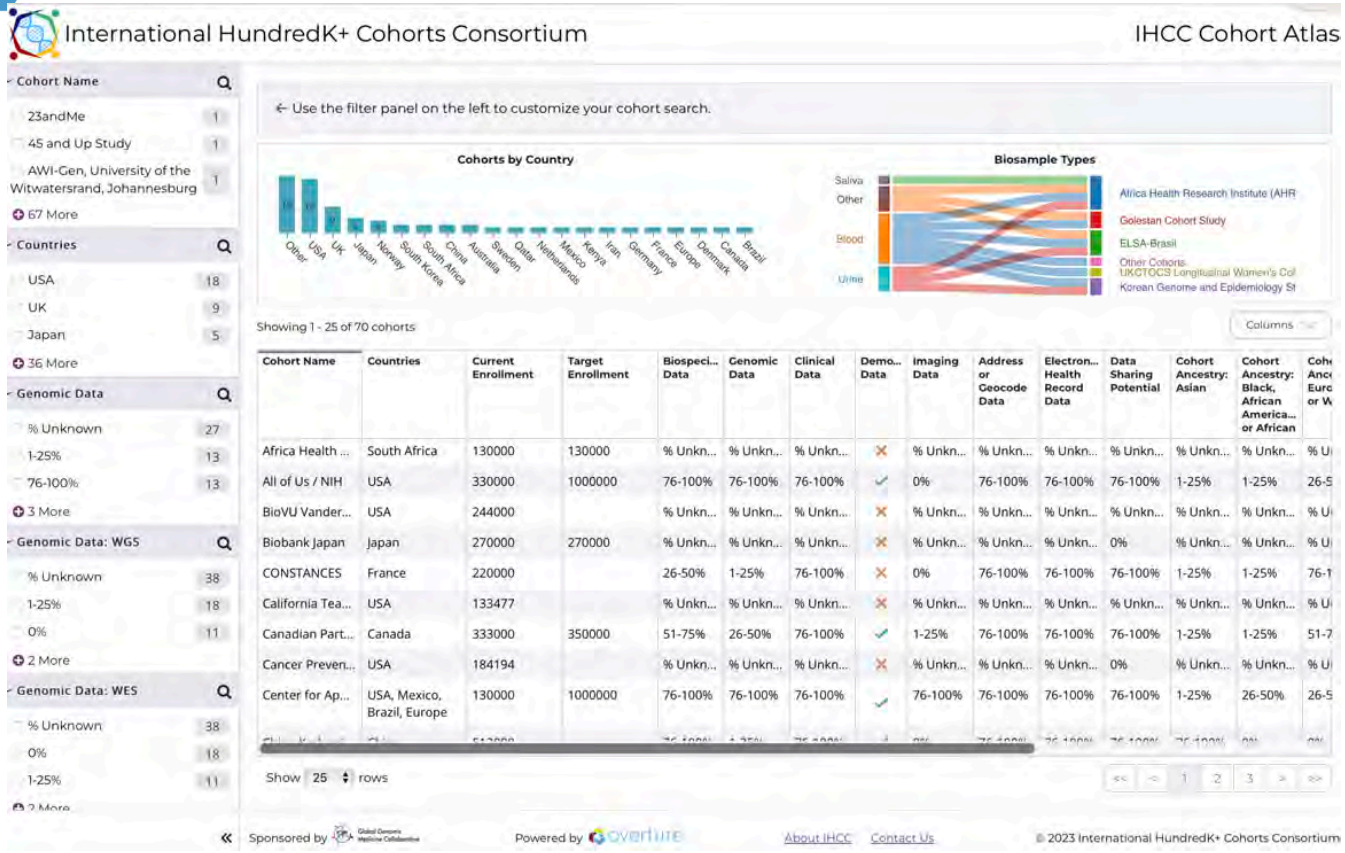
Tumour promotion

In 1947, Isaac Berenblum proposed that the development of cancer was a two-stage process: the first step introduces mutations into healthy cells, the second then promotes tumour growth through tissue inflammation. In this week's issue, Charles Swanton and his colleagues investigate the role of particulate matter in prompting the development of non-small-cell lung cancers and find that cancer initiation in response to pollution conforms to Berenblum's model. The researchers investigated especially fine particles called $PM_{2.5}$, which are smaller than 2.5 micrometres and are typically found in smoke and vehicle emissions. Looking at nearly 33,000 people from four countries, they found a clear link between prolonged exposure to PM

— show all

Cover image: Amal KS/Hindustan Times via Getty Images

International Activities



CanPath

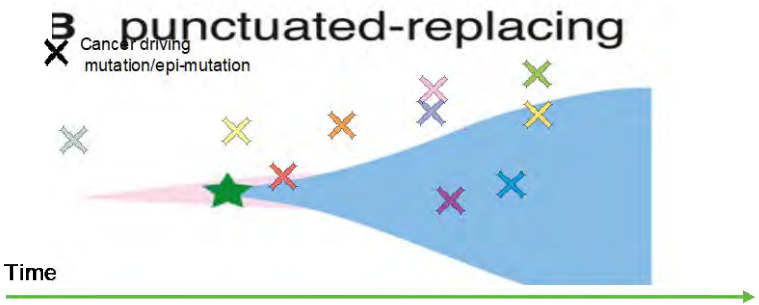
Profiling pre-diagnosis plasma cell-free DNA methylomes up to seven years prior to clinical detection reveals early signatures of cancers

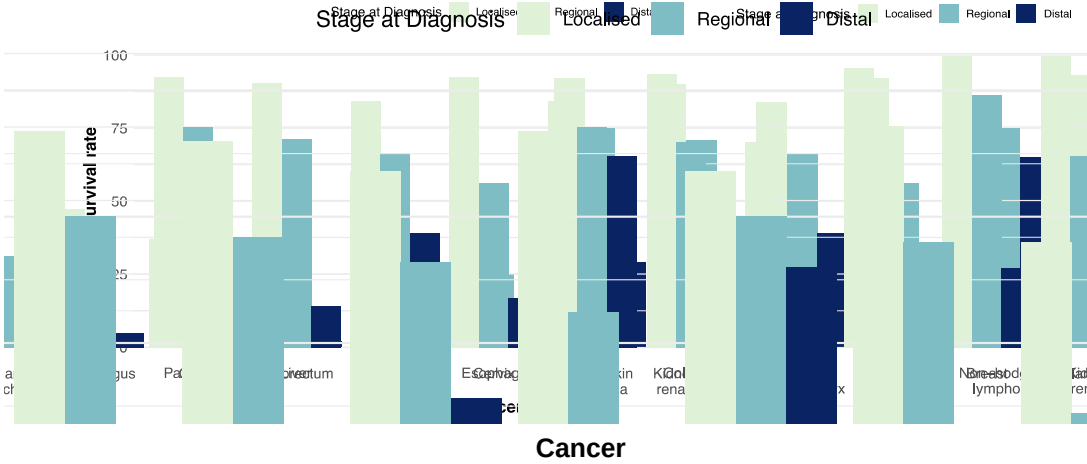
Nicholas Cheng

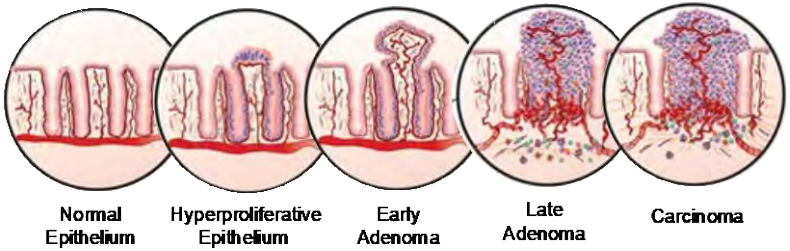
AWADALLA LAB

Pioneering Genomics for Precision Health.

Treating cancer early increases survival







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

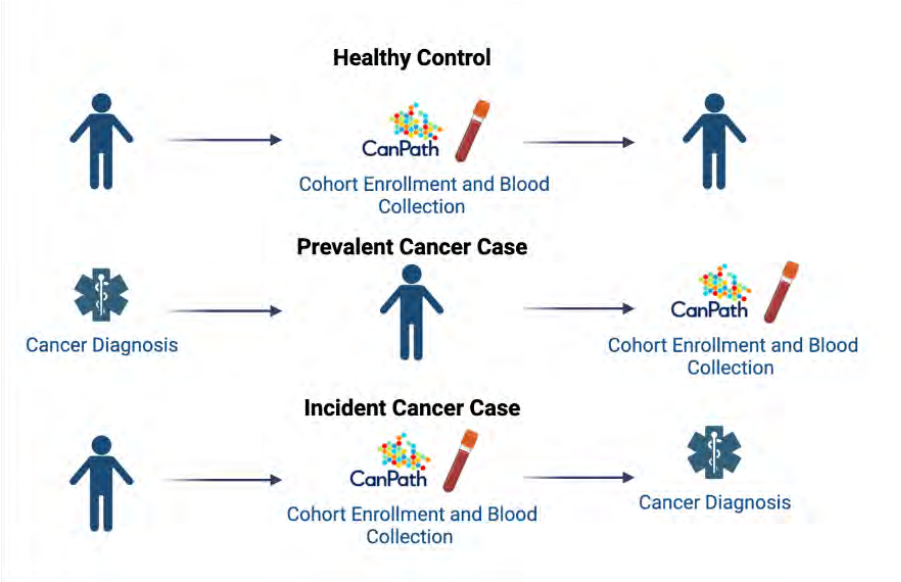
Current effective population-based screening recommendations

- 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

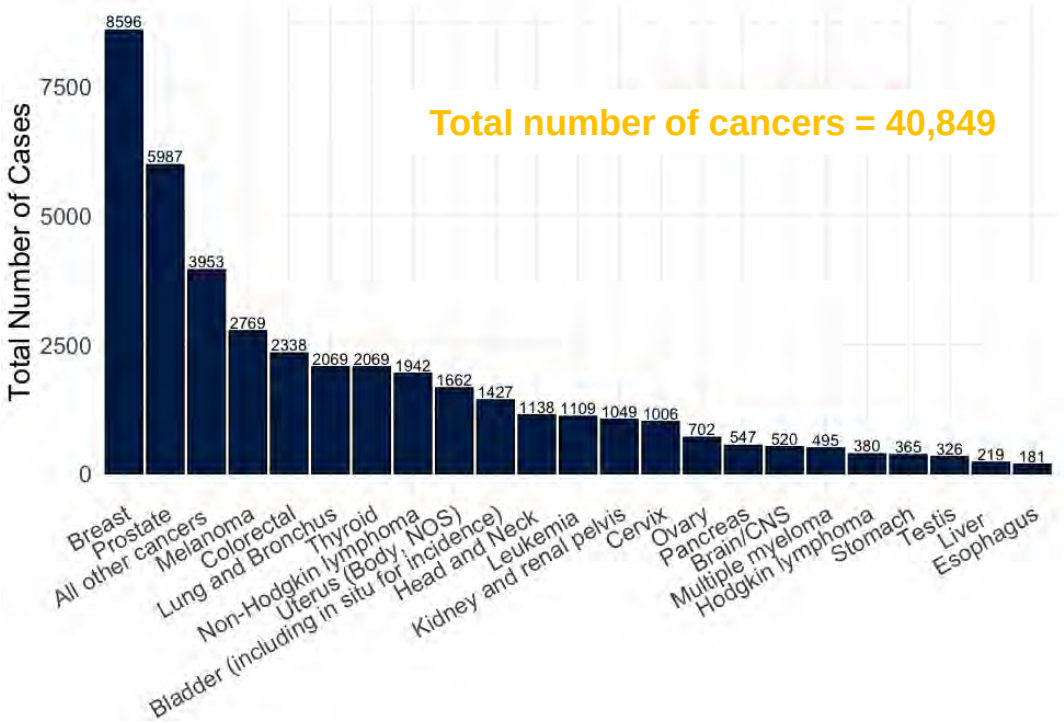
Canadian Cancer Statistics (2018)

CanPath 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 disease**
- Using samples collected before disease onset, we are able to develop **novel approaches to detect disease years before current methods**
- We are adopting a multi-stage approach to **build the data resources required to enable early cancer prevention and detection research:**
  - Harmonizing aggregate cancer data reporting nationally
  - Hosting linked individual-level cancer outcomes



Mapping cancer cases in the CanPath cohort



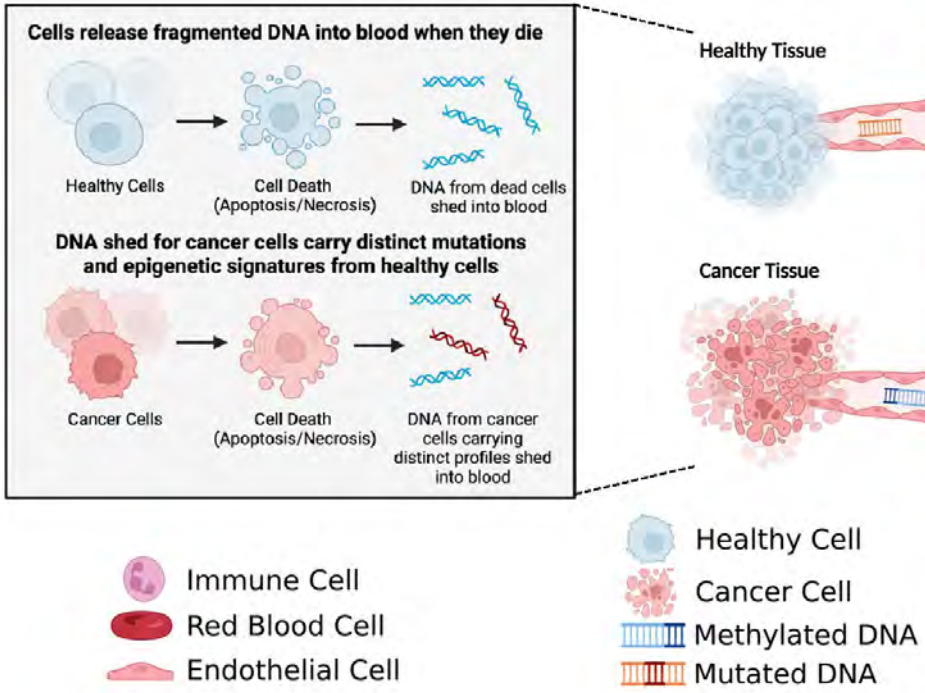
We are leveraging provincial linkages to map map CanPath cancer data and biosample holdings

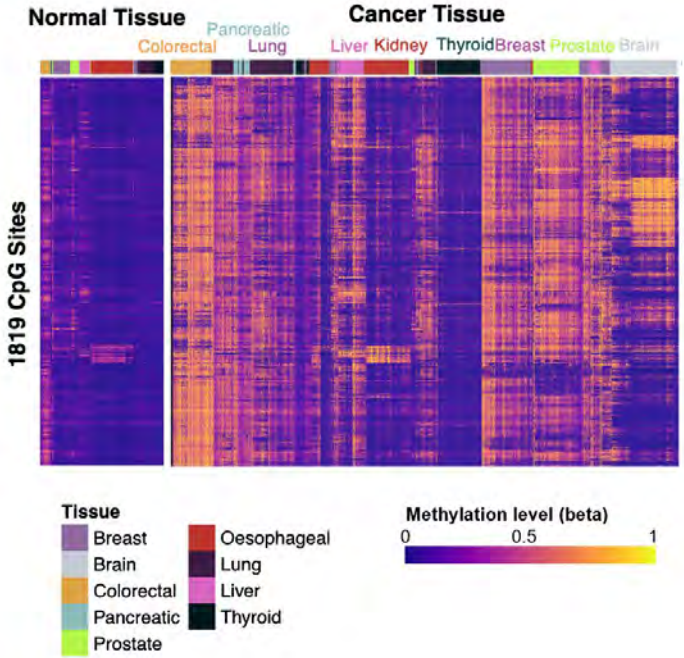
All cancer data is collected and grouped according to Canadian Cancer Statistic guidelines

Regions included:

- **Atlantic Path**
- **Alberta for Tomorrow Project**
- **Ontario Health Study**
- **BC Generations Project**

Liquid Biopsy Approaches for Early Cancer Detection



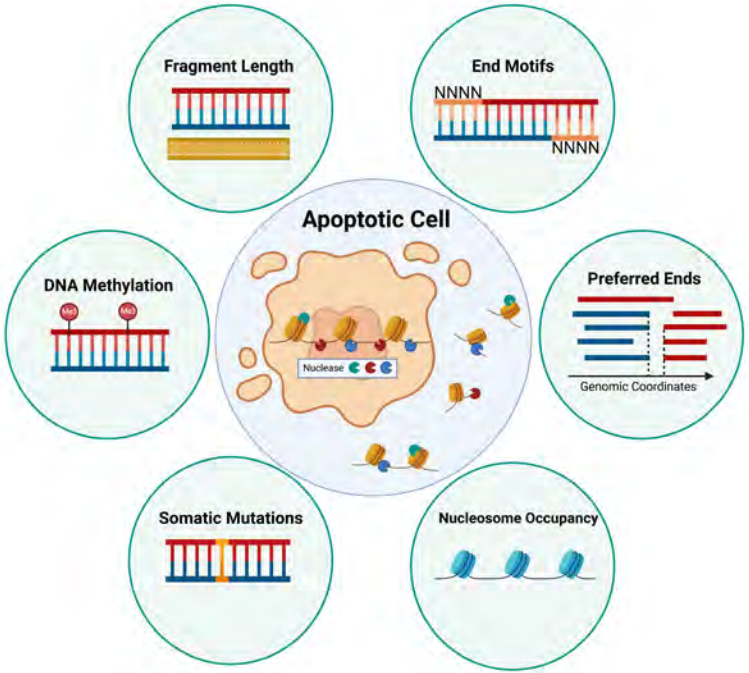


Cell-free DNA as a biomarker of disease and tissue damage

Cell-free DNA (cfDNA) are typically shed from dying cells.

While most cfDNA circulating in blood are derived from leukocytes, increased shedding from other tissues can be indicative of tissue damage and diseases.

Cell-free DNA genomic, epigenomic and fragmentomic signatures can be utilized as biomarkers for early disease detection.



DNA Methylation Patterns Discriminates between Cancer/Normal tissue and Cancer Tissue of Origin

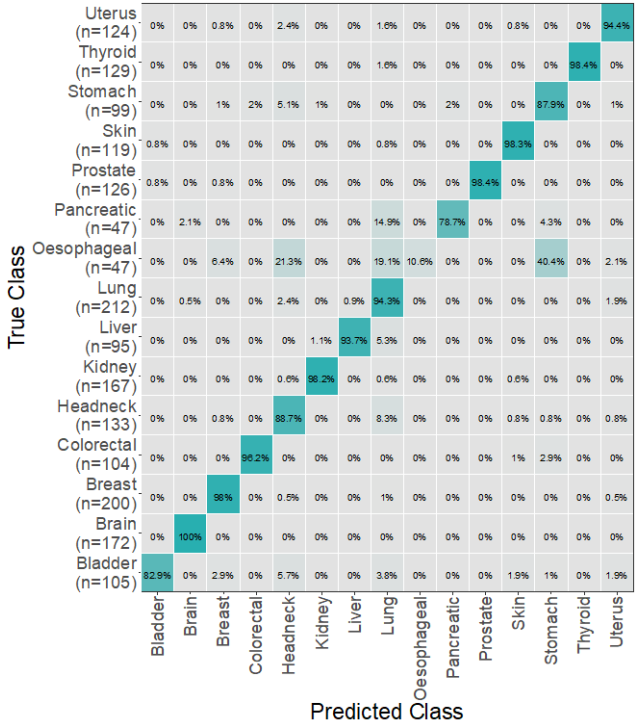
Using DNA methylation signatures to classifying solid tissue biopsies as cancer or healthy using DNA methylation signatures

		Actual Class		CpG Sites: 500
		Tumour	Healthy	Total (n)
Predicted Class	Tumour	6545	7	Precision: 99.9% 4958
	Healthy	454	217	370
		Sensitivity: 95.6%	Specificity: 96.6%	5328

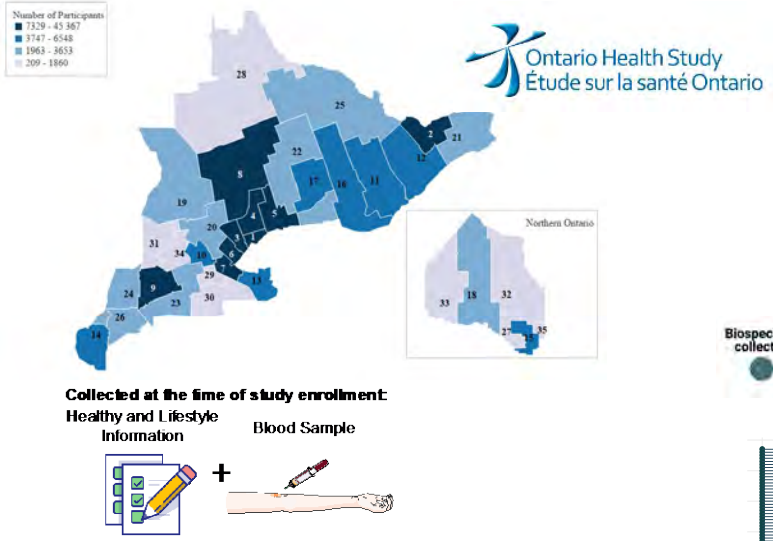
F1 Score: 98%

TCGA

Predictive models trained using 450k DNA methylation array data from paired TCGA cancer and normal tissue across of over 15 cancer types

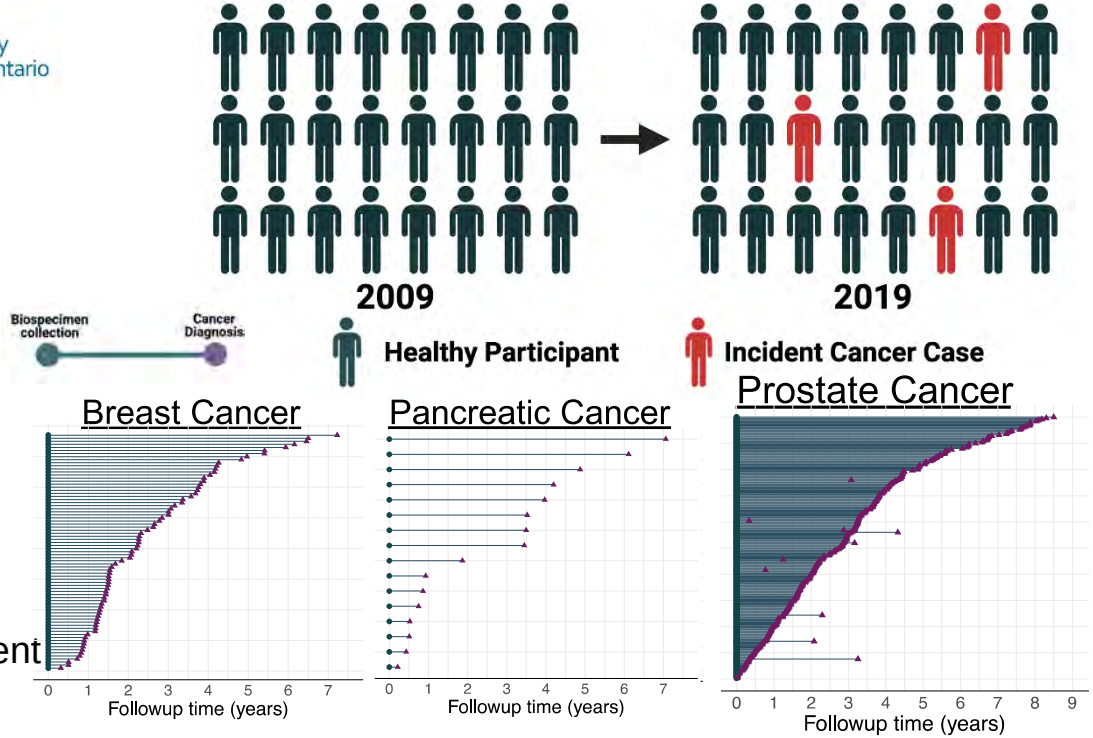


Leveraging population cohorts to study early cancer detection prior to clinical detection

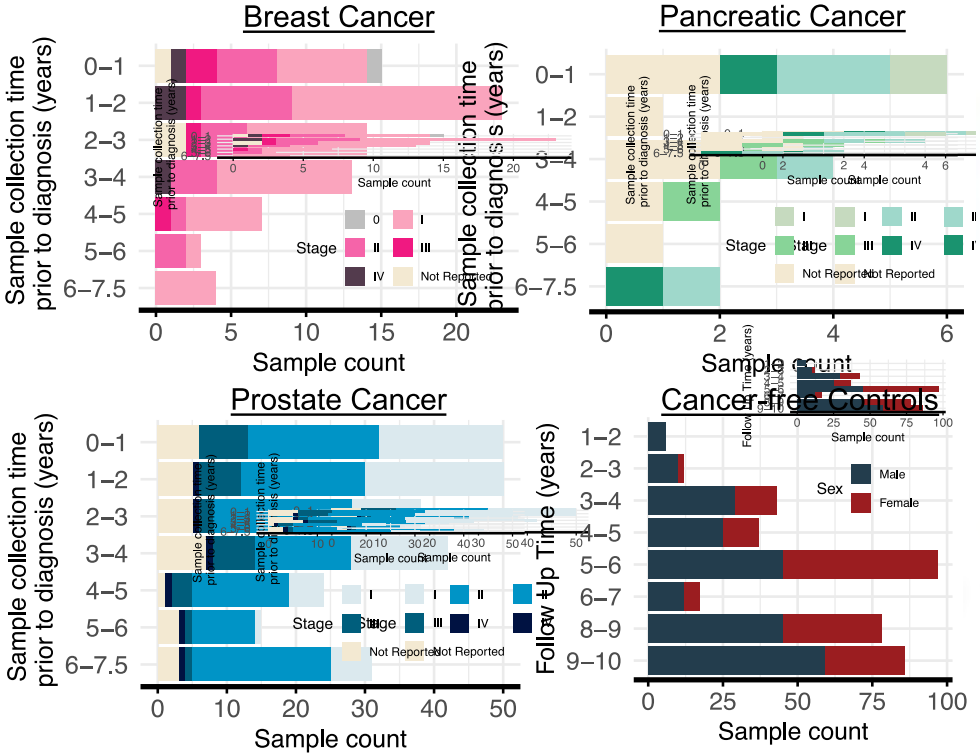


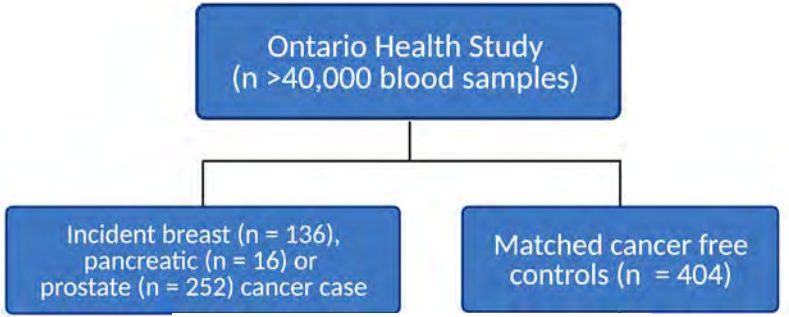
n > 40,000 blood plasma samples from cancer-free participants at time of enrollment

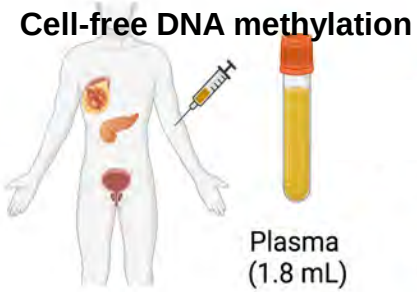
Kirsh et al. **International Journal of Epidemiology** (2022)



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



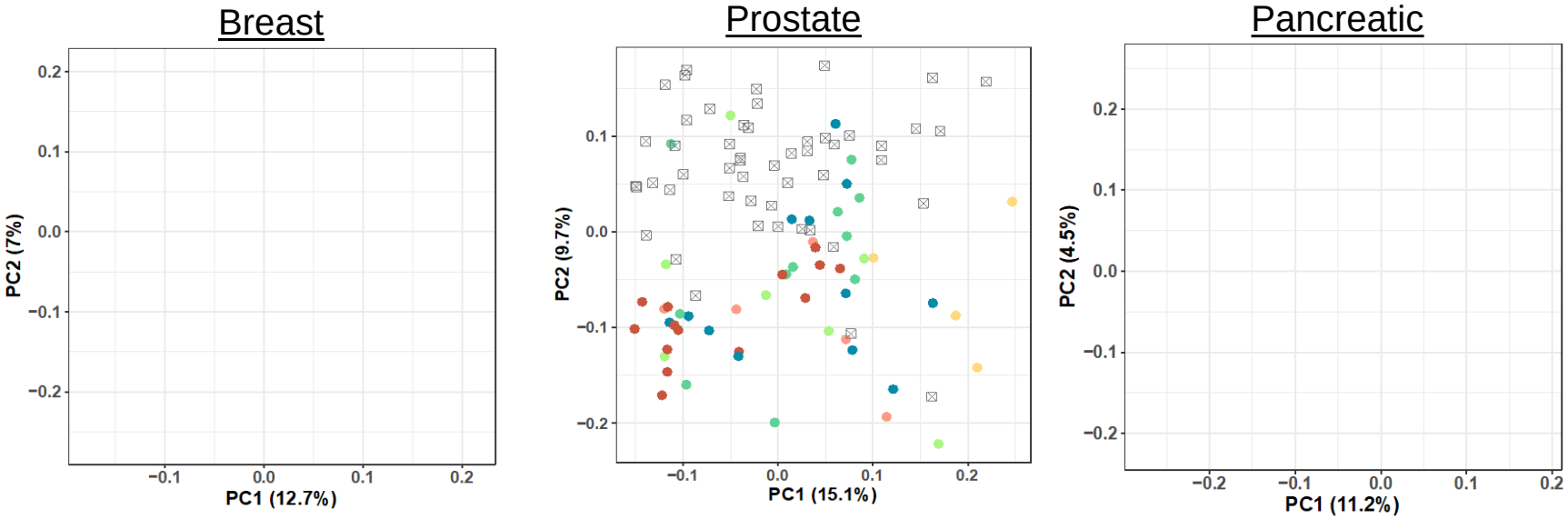




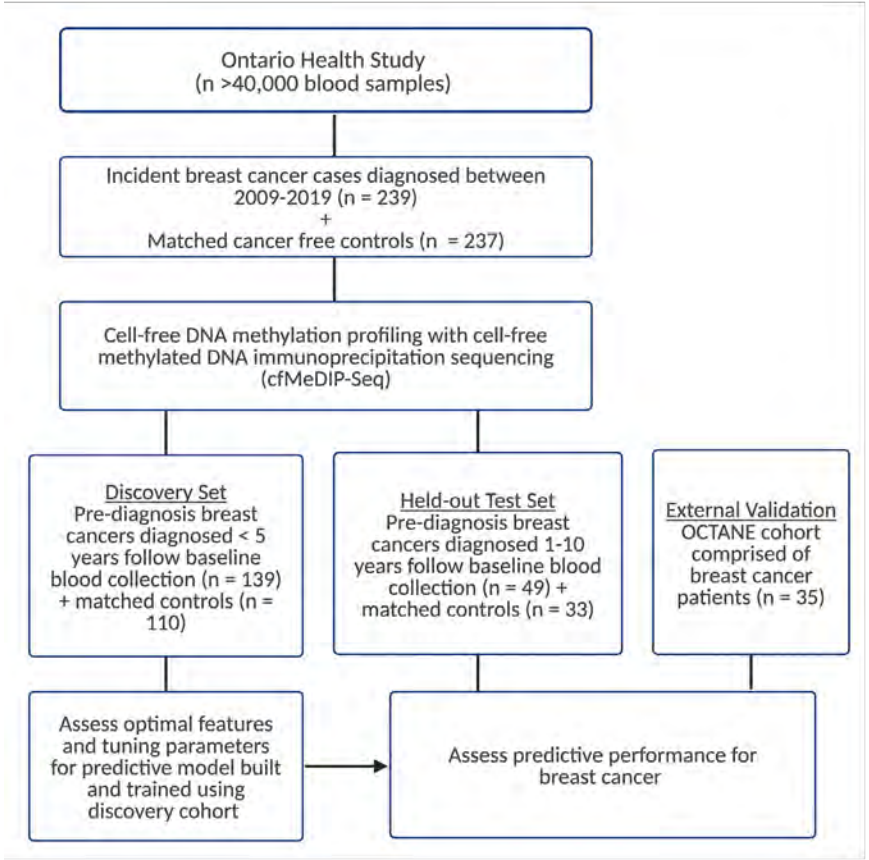
Shen, Singhania, Fehringer et al. 2019 *Nature*

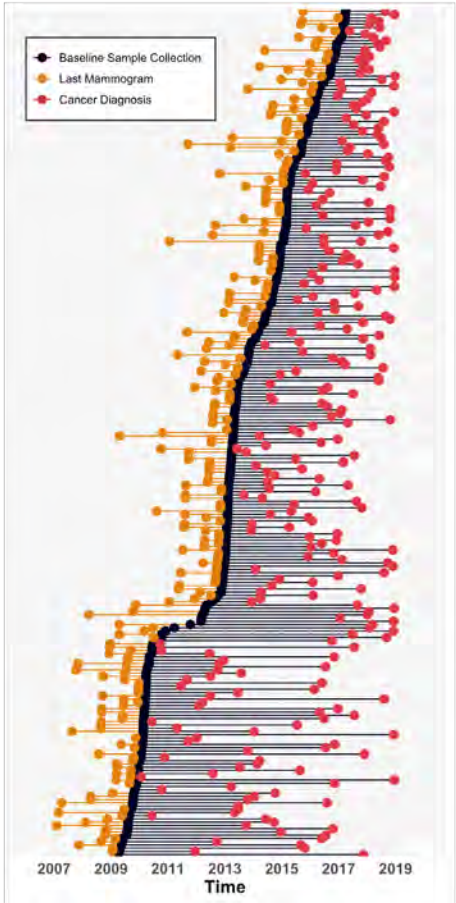
CanPath

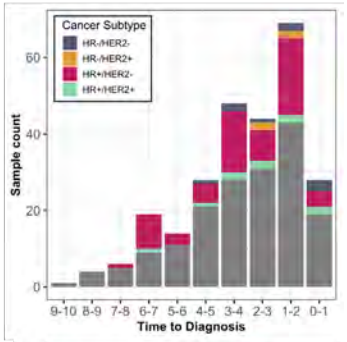
Discriminating individuals that will develop cancers using pre-diagnosis cfDNA methylation signatures

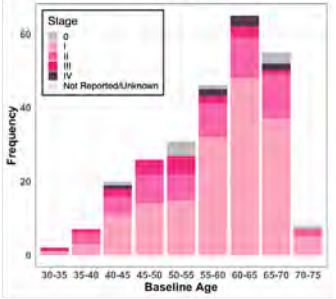


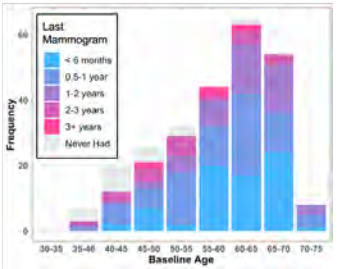
CanPath


Ontario Health Study
(n >40,000 blood samples)
Incident breast cancer cases diagnosed between 2009-2019 (n = 239)
+
Matched cancer free controls (n = 237)
Cell-free DNA methylation profiling with cell-free methylated DNA immunoprecipitation sequencing (cfMeDIP-Seq)
Discovery Set
Pre-diagnosis breast cancers diagnosed < 5 years follow baseline blood collection (n = 139) + matched controls (n = 110)
Held-out Test Set
Pre-diagnosis breast cancers diagnosed 1-10 years follow baseline blood collection (n = 49) + matched controls (n = 33)
External Validation
OCTANE cohort comprised of breast cancer patients (n = 35)
Assess optimal features and tuning parameters for predictive model built and trained using discovery cohort
Assess predictive performance for breast cancer


Baseline Sample Collection
Last Mammogram
Cancer Diagnosis
2007
2009
2011
2013
2015
2017
2019
Time

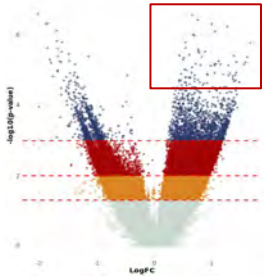

Cancer Subtype
HR-/HER2-
HR-/HER2+
HR+/HER2-
HR+/HER2+
Sample count
0
20
40
60
9-10 8-9 7-8 6-7 5-6 4-5 3-4 2-3 1-2 0-1
Time to Diagnosis


Stage
0
I
II
III
IV
Not Reported/Unknown
Frequency
0
20
40
60
30-35 35-40 40-45 45-50 50-55 55-60 60-65 65-70 70-75
Baseline Age

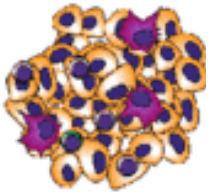

Last Mammogram
< 6 months
0.5-1 year
1-2 years
2-3 years
3+ years
Never Had
Frequency
0
20
40
60
30-35 35-40 40-45 45-50 50-55 55-60 60-65 65-70 70-75
Baseline Age

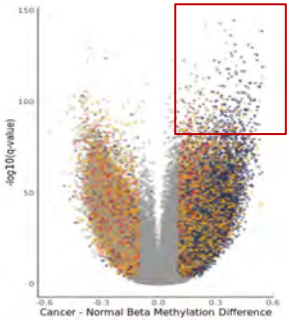
Pre-diagnosis cfDNA methylation signatures share concordant signatures with bulk cancer tissues

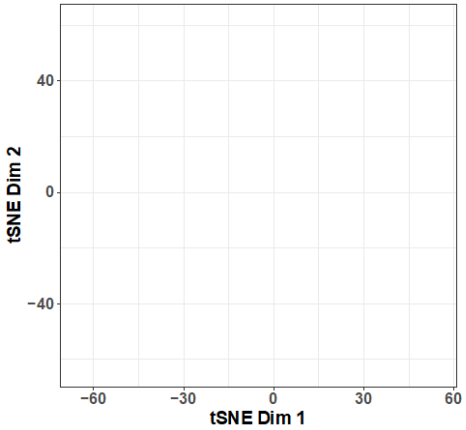
OHS Pre-diagnosis Cancer vs Control cfDNA hypermethylated regions



TCGA Bulk Cancer Tissue vs PBL/Adjacent Normal Hypermethylated regions





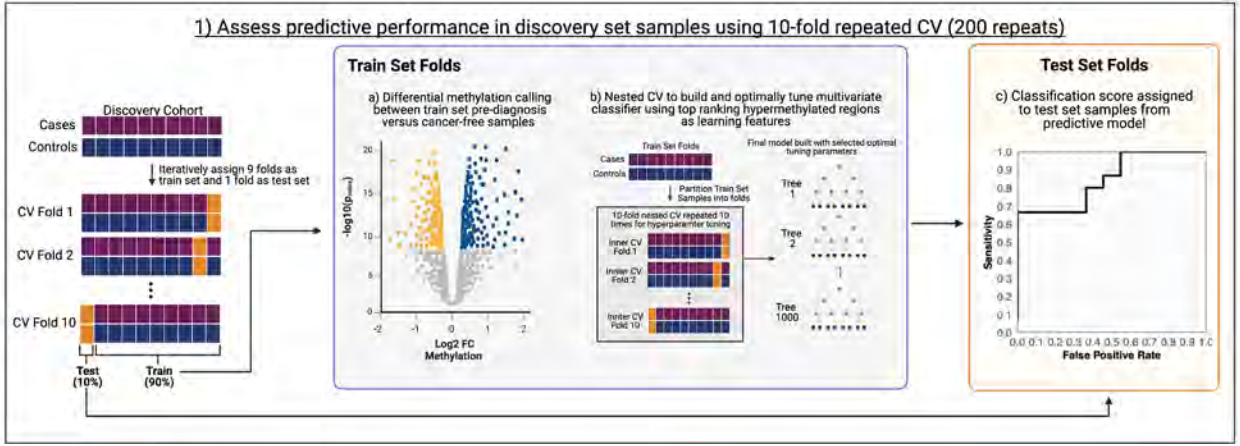


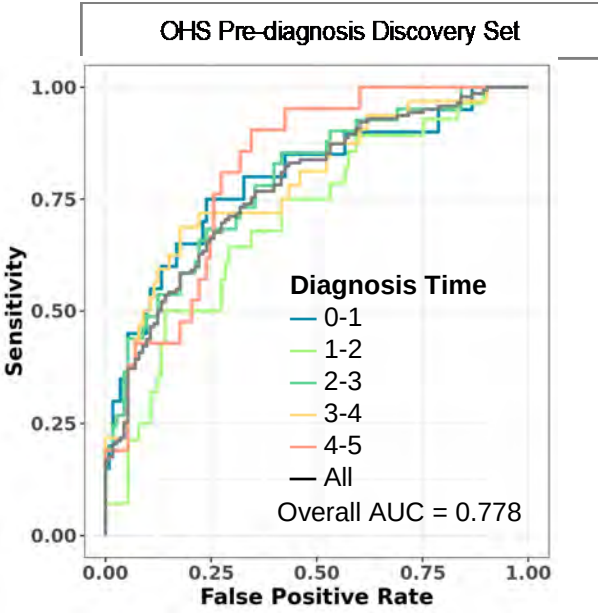
Pancreat
Breast
Prostate
Blood

Tumour
Normal

CanPath

cfDNA methylation signatures predicts breast cancer in pre-diagnosis blood samples



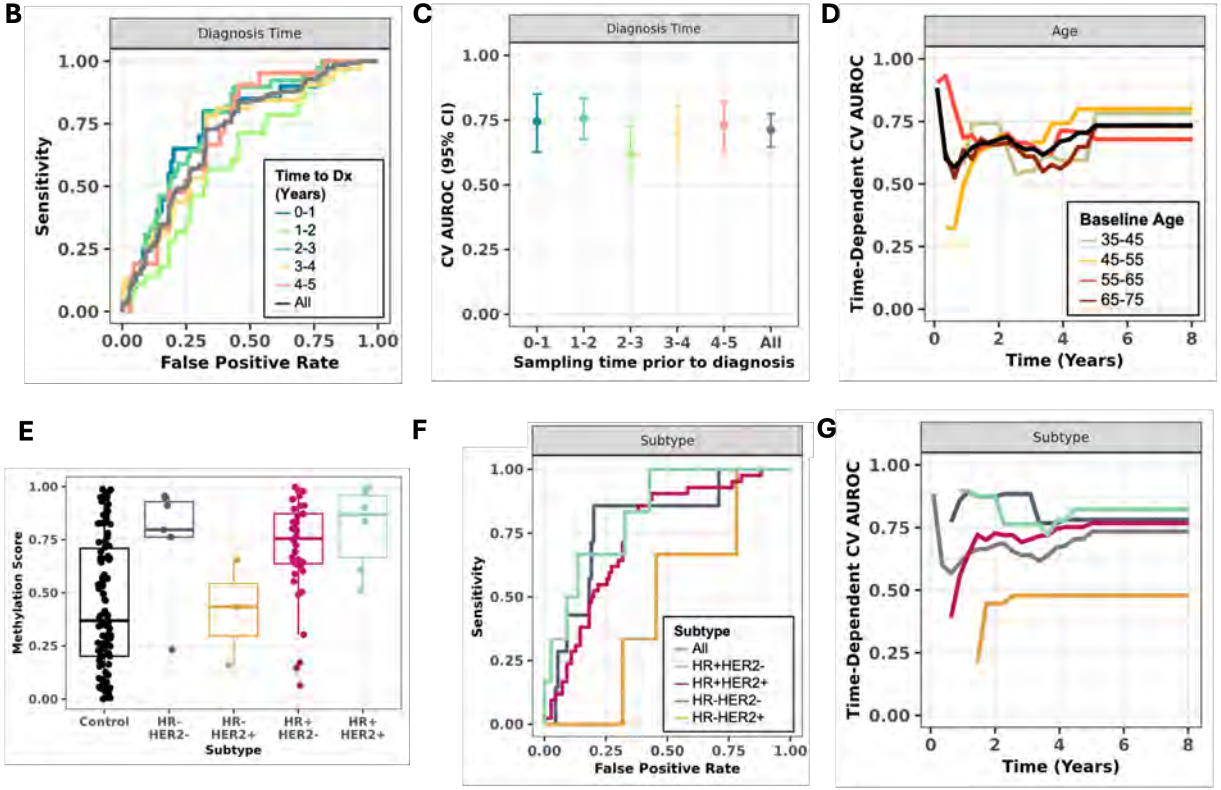


Discovery Cohort:
Pre-diagnosis breast cancer cases (n = 67)
Matched controls (n = 59)

Random forest model per iteration:
1000 Trees
150 Hypermethylated Regions
Tuning with 10-fold CV with 10 repeats

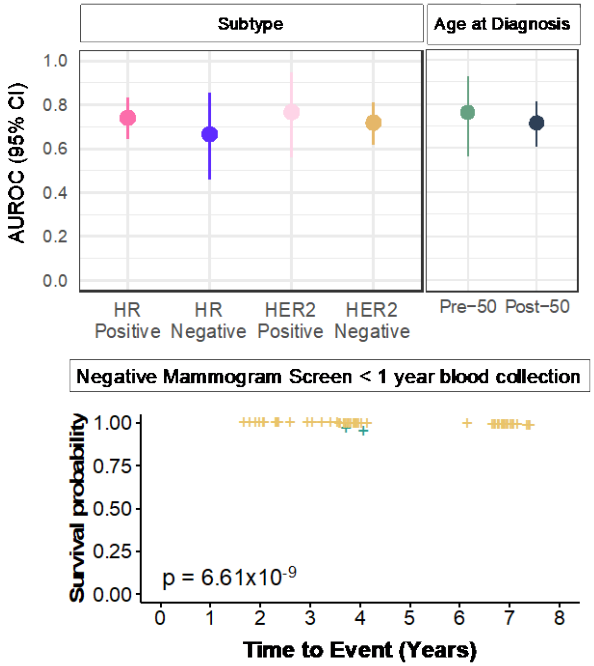
Cheng et al. (in Review)

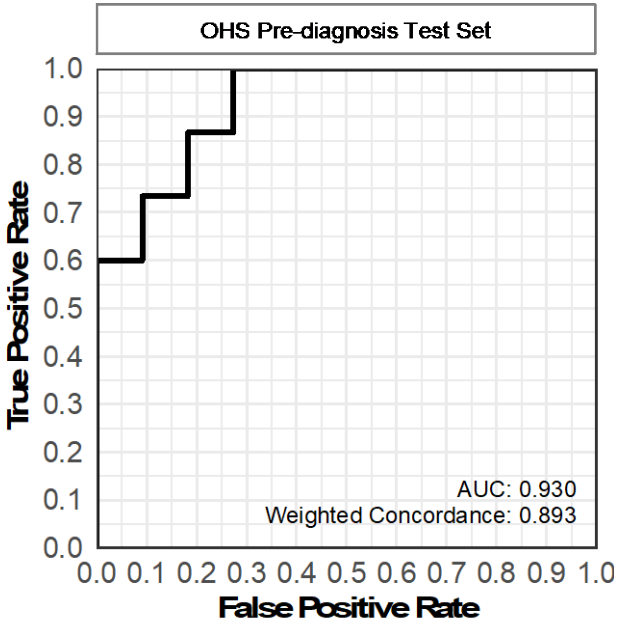
Performance varies for cancer subtypes and age of diagnosis

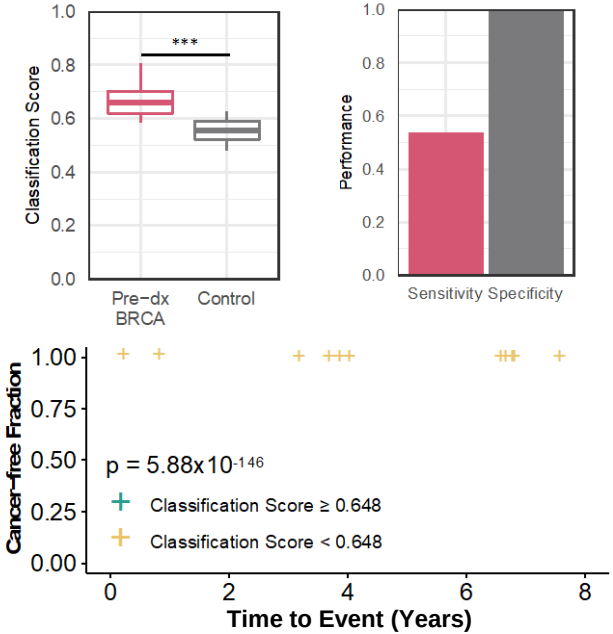


CanPath

cfDNA methylation signatures detects breast cancers prior preceding mammogram

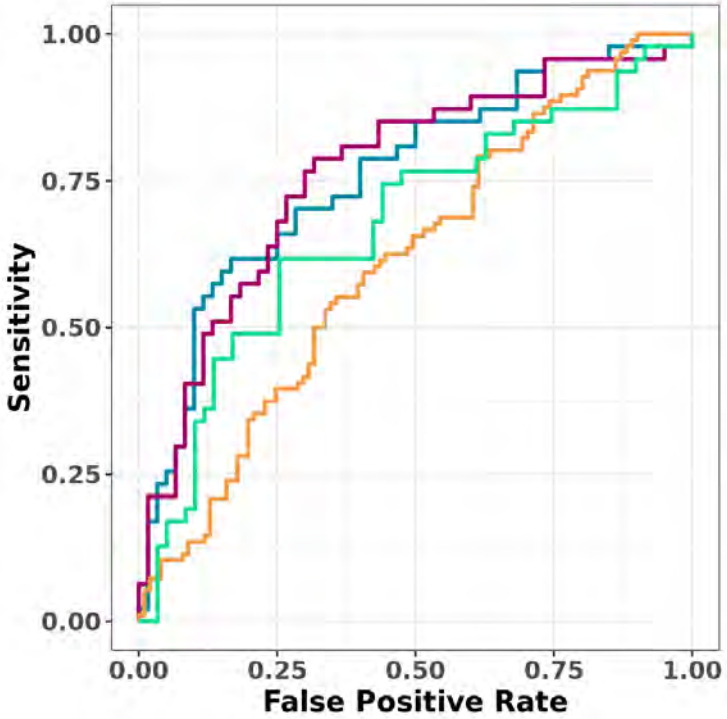


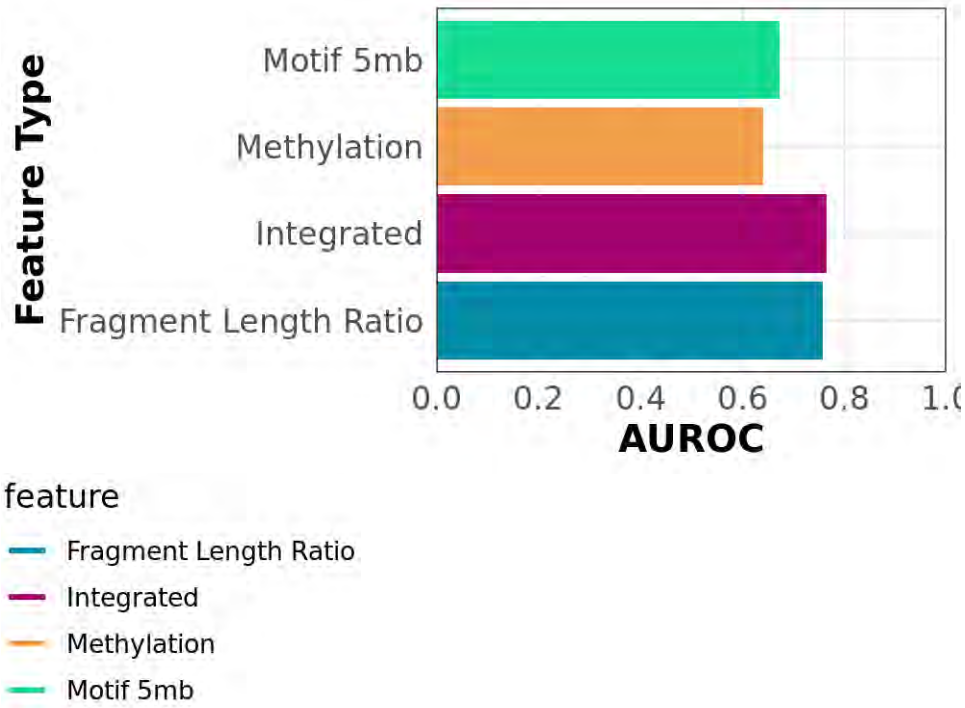




Cheng et al. (in Review)

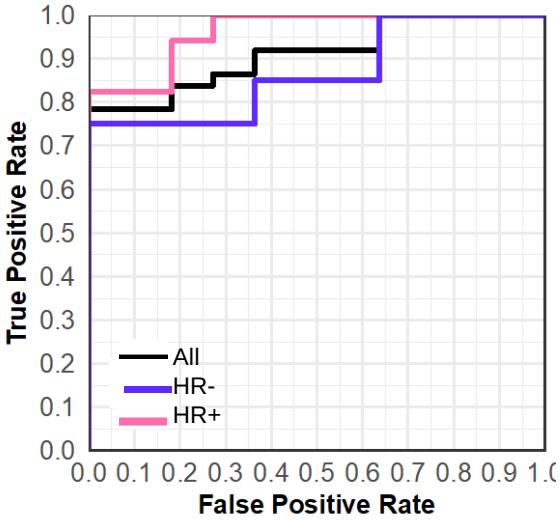
Prostate Cancer Integrated feature score performance





Future work: Pre-diagnosis cfDNA signatures are generalizable to other cancers

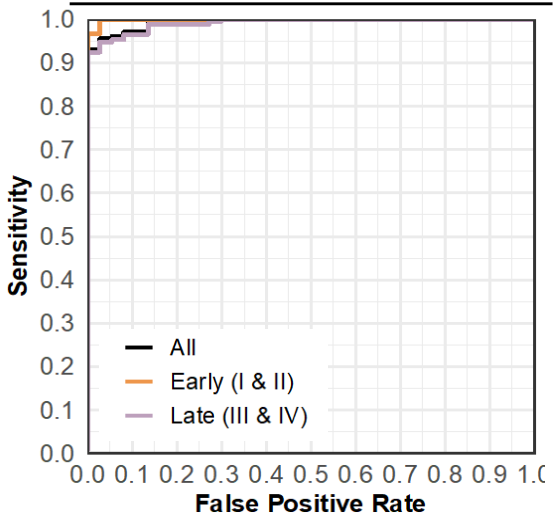
Metastatic Breast Cancer



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)

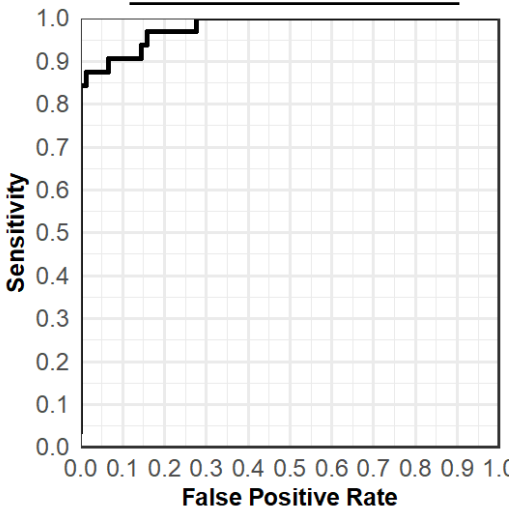
Prostate Metastatic Cancer



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

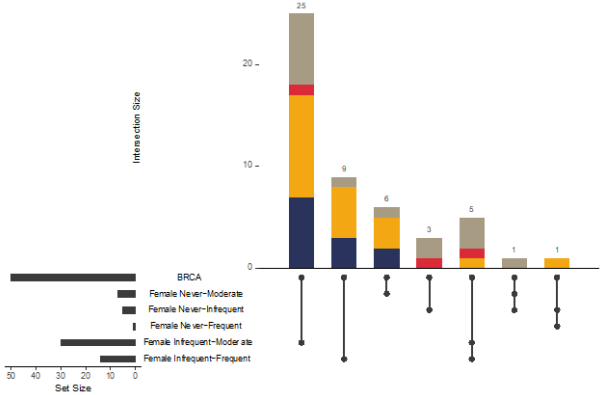


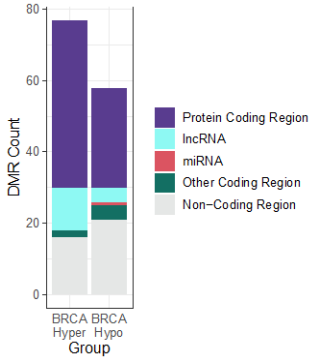
Pancreatic Cancer Train Set
Pre-dx pancreatic cancer cases (n = 16)
Cancer-free controls (n = 50)

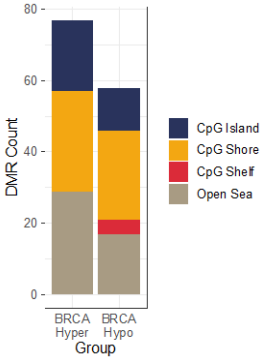
Pancreatic Cancer Test Set
External post-dx pancreatic cancer cases (n = 38)
Cancer-free controls (n = 80)

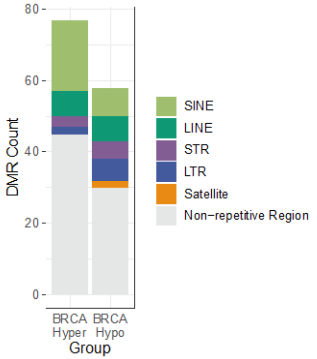
CanPath

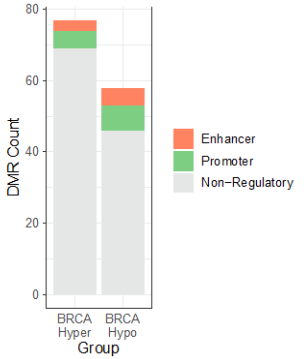
Alcohol associated DMRs that overlap breast cancer DMRs





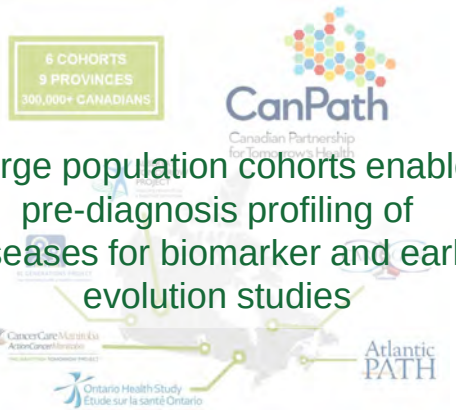




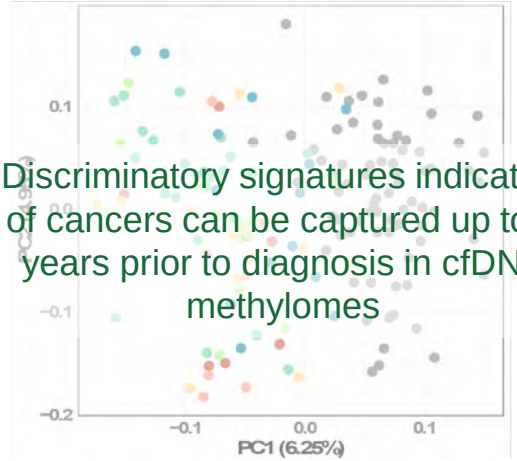


Future work: Can we see signatures of specific risk factors associated with cancer or disease?

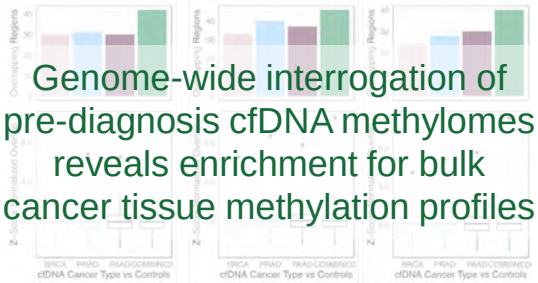
Summary



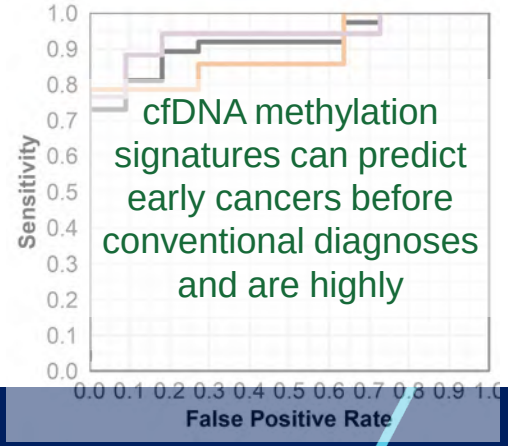
Large population cohorts enable pre-diagnosis profiling of diseases for biomarker and early evolution studies



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

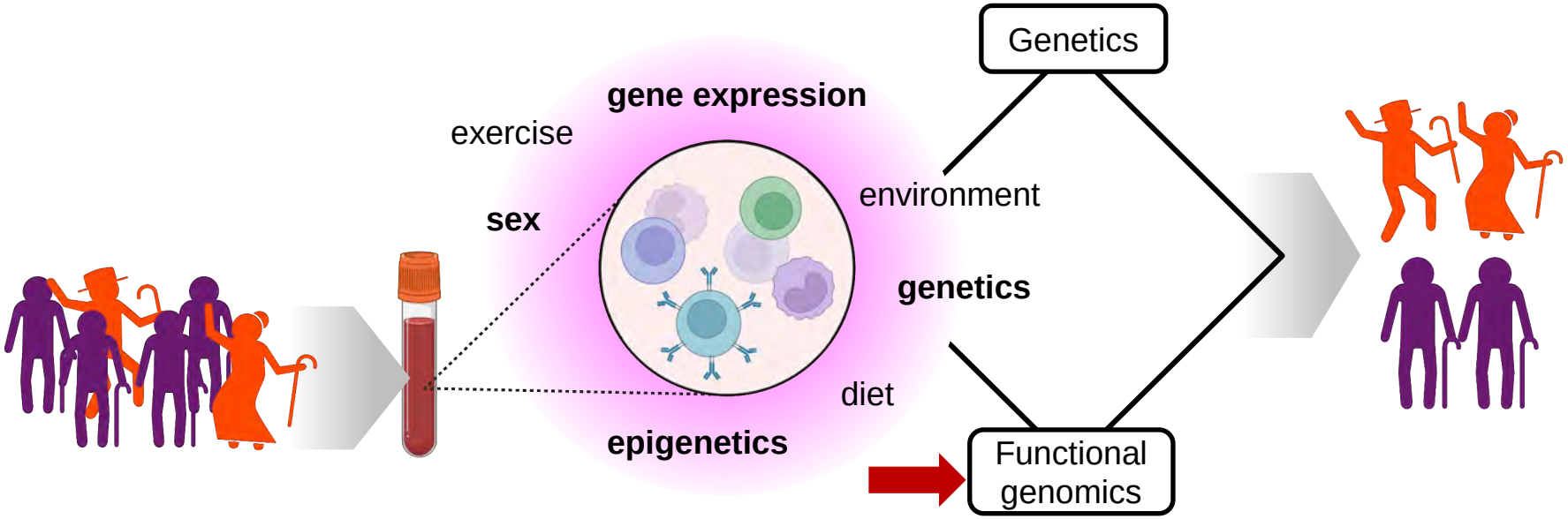


Genome-wide interrogation of pre-diagnosis cfDNA methylomes reveals enrichment for bulk cancer tissue methylation profiles



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

What factors contribute to healthy aging of blood?

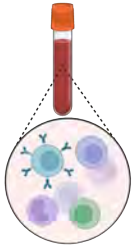


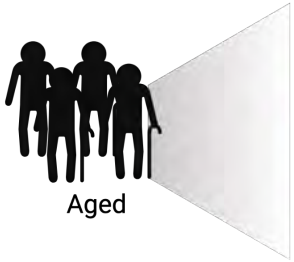
Mechanisms of healthy aging in blood cells using single-cell -omics

Elyssa Bader, PhD

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

Young





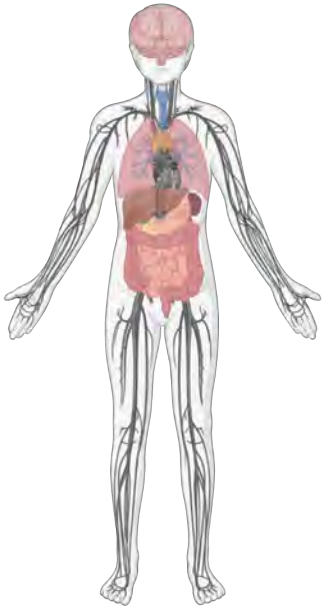
Aged

e.g., Centenarian (100+ years old), non-frail

Aged vs. young blood

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

Centenarian



Health check

- brain ☑
- lungs ☑
- liver ☑
- …etc

- heart ⊗
- blood ⊗

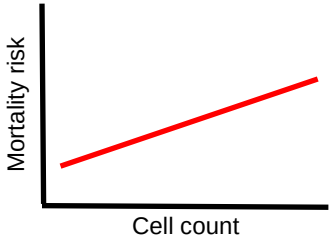
Intermountain risk score predicts 5-year mortality

CanPath
Canadian Partnership for Tomorrow's Health

CBC data + age
n = 25 000

5 year follow-up

Risk modeling



Horne et al. (2009) Am J Med

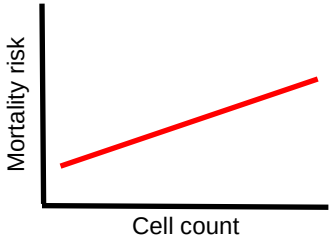
Intermountain risk score predicts 5-year mortality

CanPath
Canadian Partnership for Tomorrow's Health

CBC data + age
n = 25 000

5 year follow-up

Risk modeling

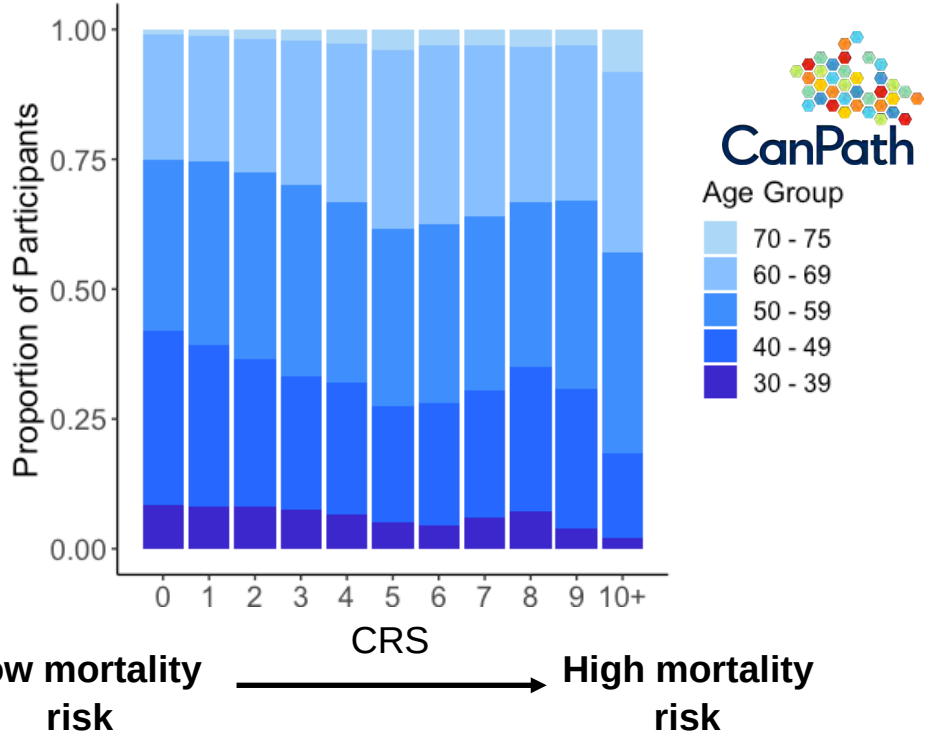


Low IRS = Low mortality risk = Healthy blood

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



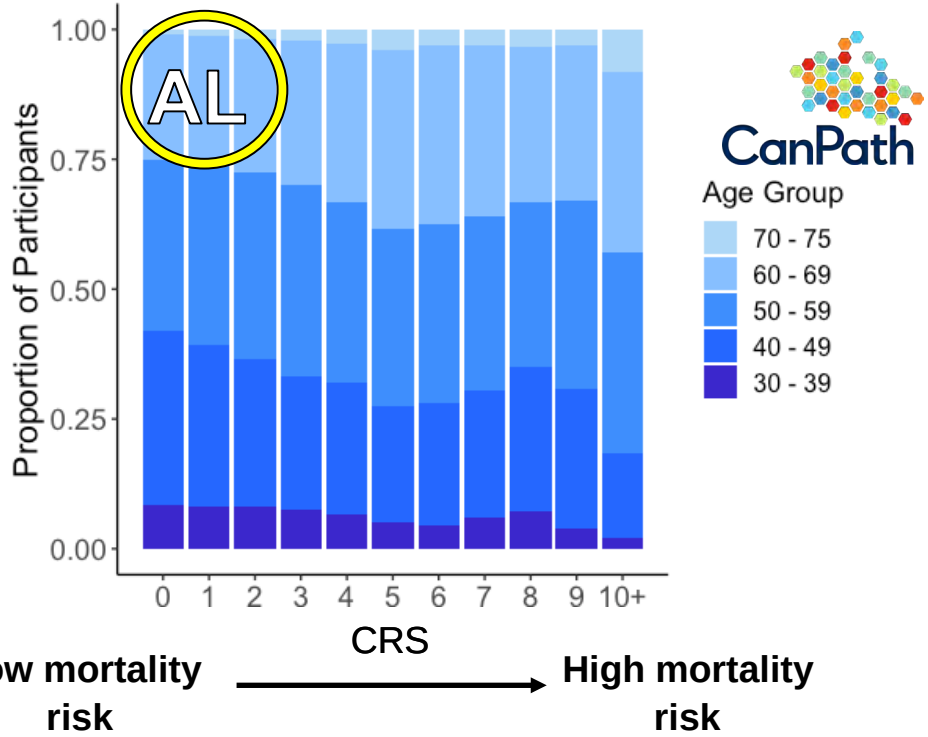
Variables in CRS

- 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*

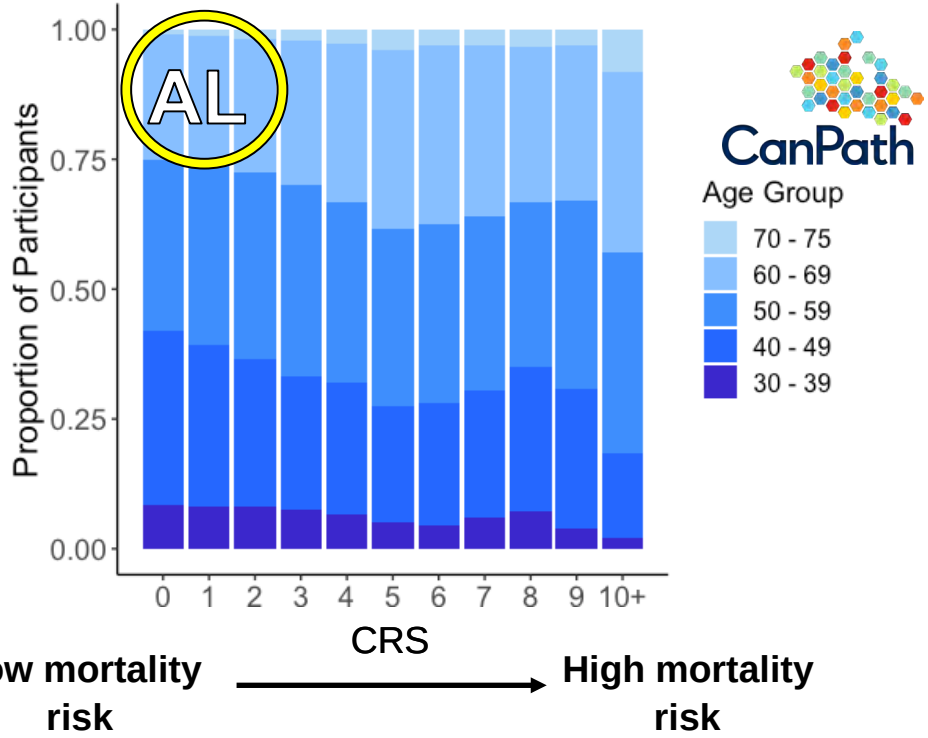
Identifying mechanisms of healthy aging in blood

CanPath
Canadian Partnership for Tomorrow's Health



Low mortality risk → **High mortality risk**

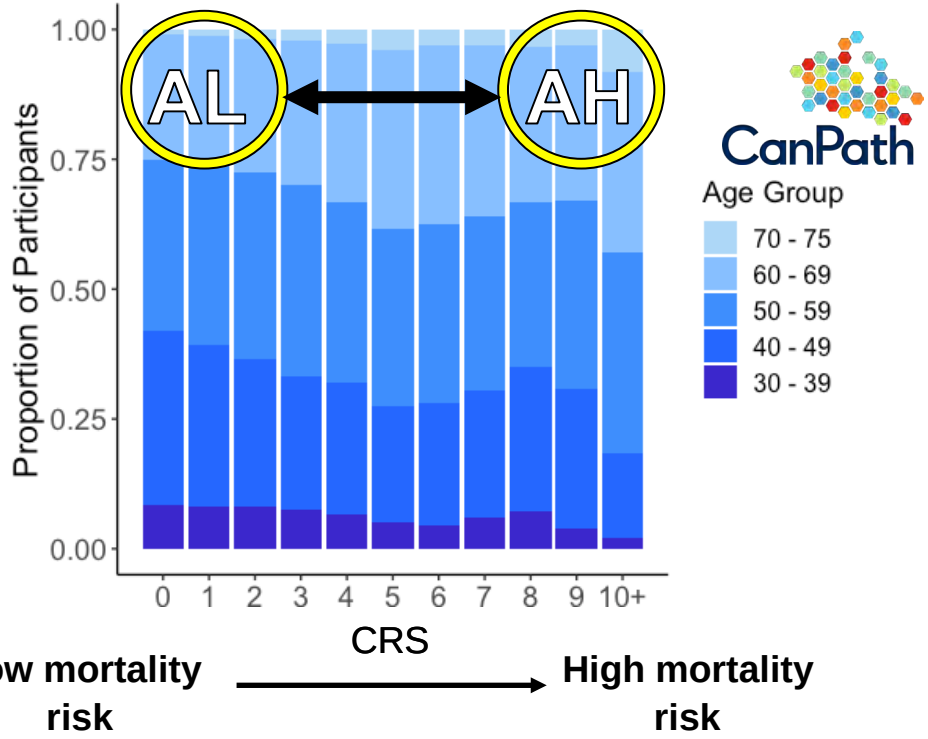
Identifying mechanisms of healthy aging in blood



Hypothesis 1

Protective mechanism: AL blood is different from AH blood

Identifying mechanisms of healthy aging in blood

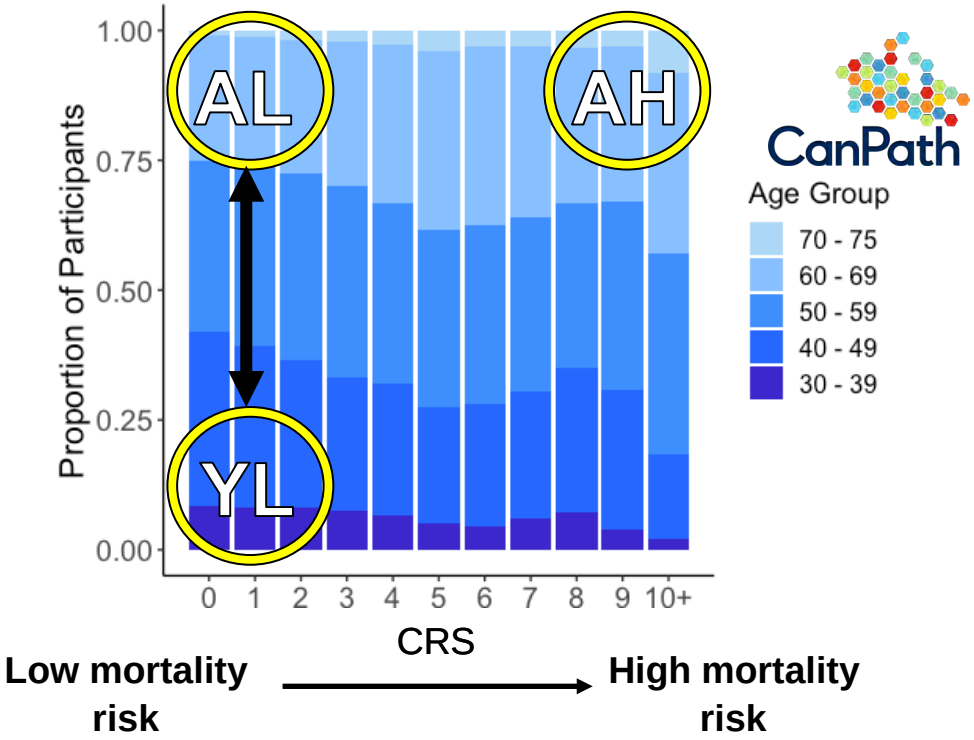


Hypothesis 1

Protective mechanism: AL blood is different from AH blood

Identifying mechanisms of healthy aging in blood

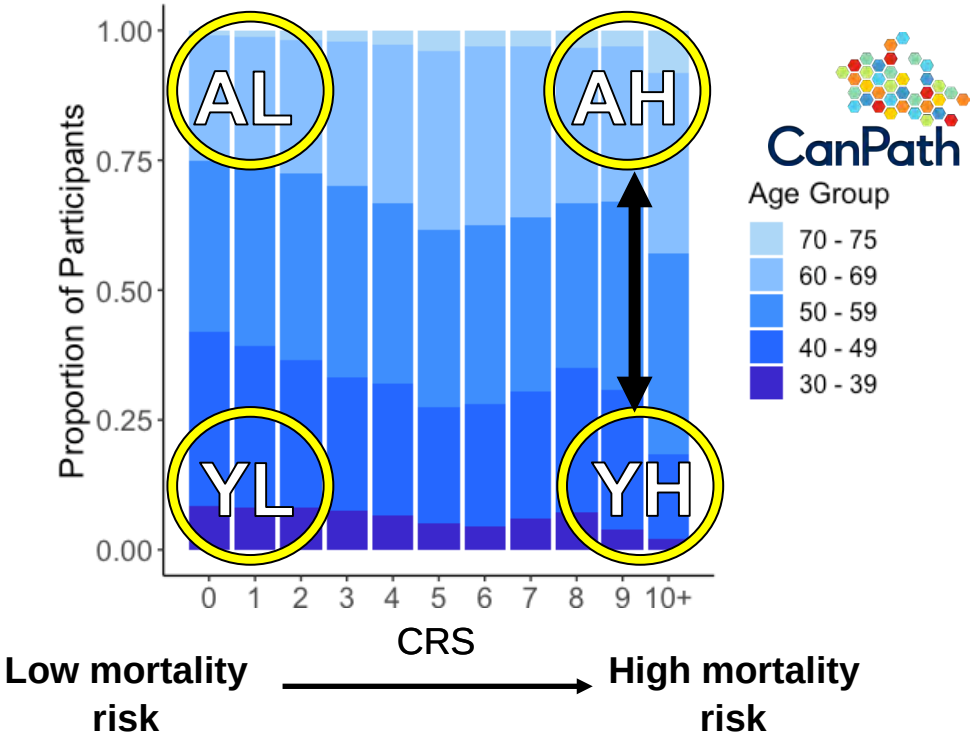
CanPath
Canadian Partnership for Tomorrow's Health



Hypothesis 1
Protective mechanism: AL blood is different from AH blood

Hypothesis 2
Healthy aging mechanism: AL blood is similar to YL blood

Identifying mechanisms of healthy aging in blood



Hypothesis 1

Protective mechanism: AL blood is different from AH blood

Hypothesis 2

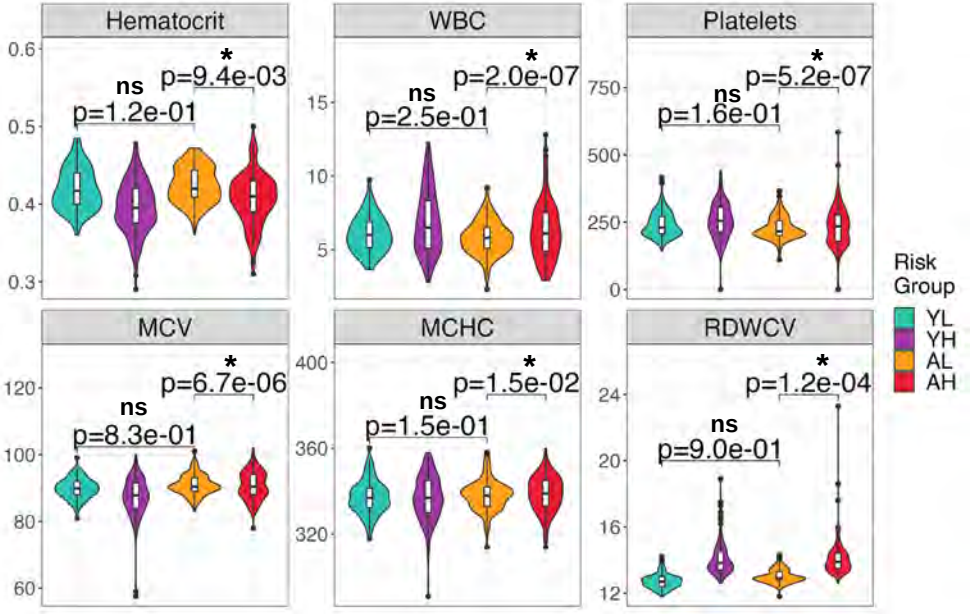
Healthy aging mechanism: AL blood is similar to YL blood

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

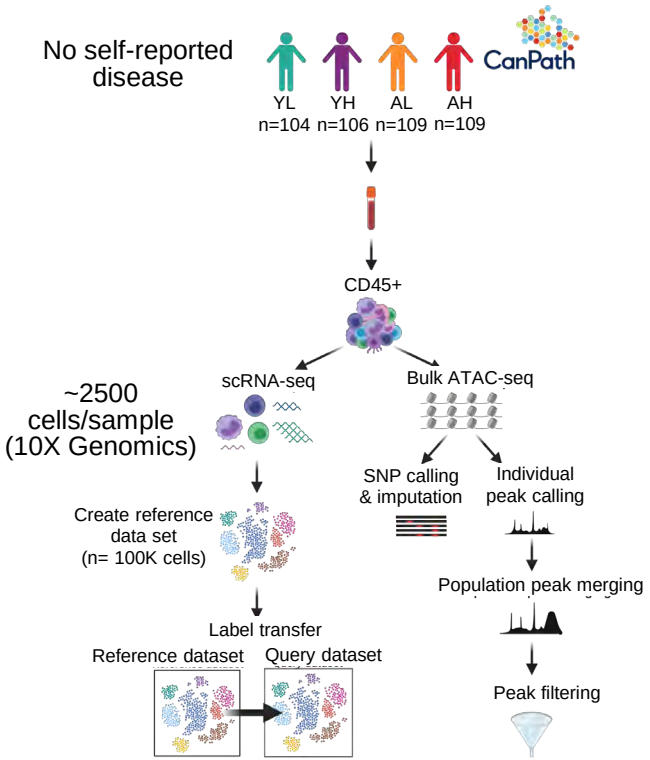
Low-risk: CRS 0 – 3
High- risk: CRS 5+

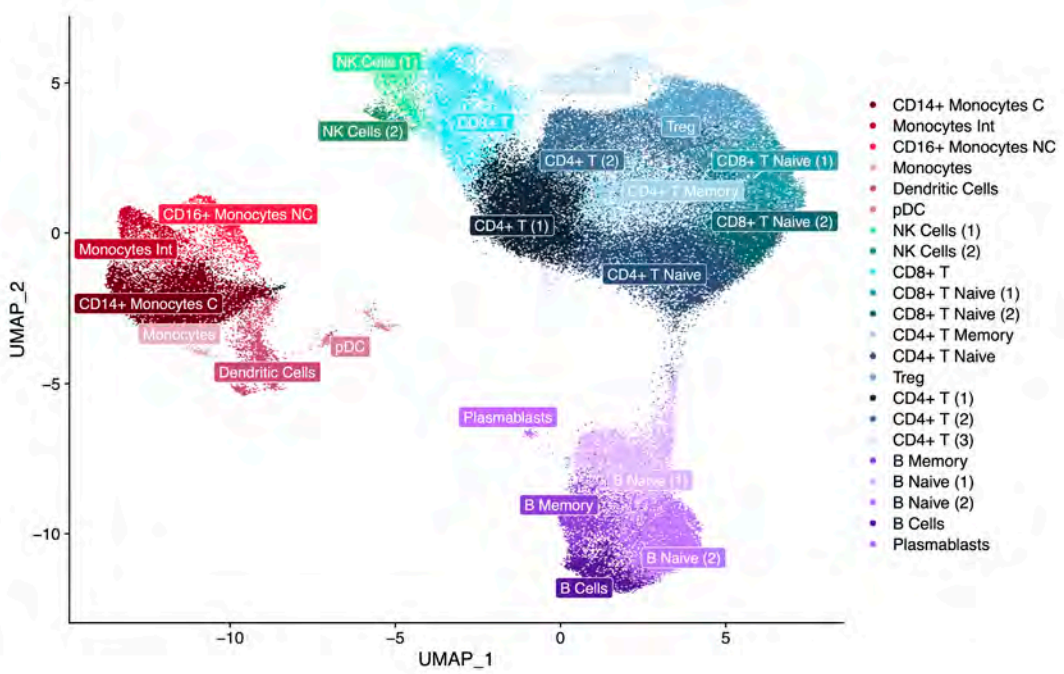


p-values calculated from Levene's test for variance

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

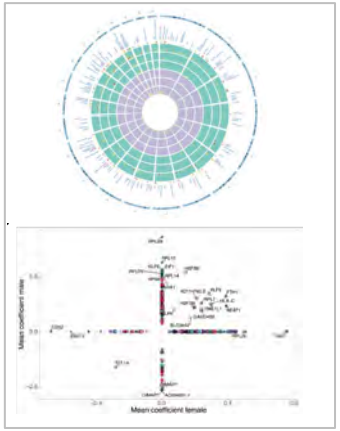
CanPath
Canadian Partnership for Tomorrow's Health



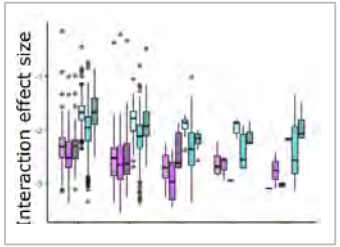


Summary

Factors contributing to healthy blood aging



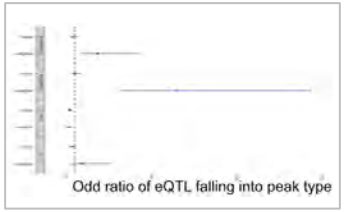
Genetic and **transcriptional** variation associated with healthy blood aging is **sex and cell type specific**



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**

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

Acknowledgements

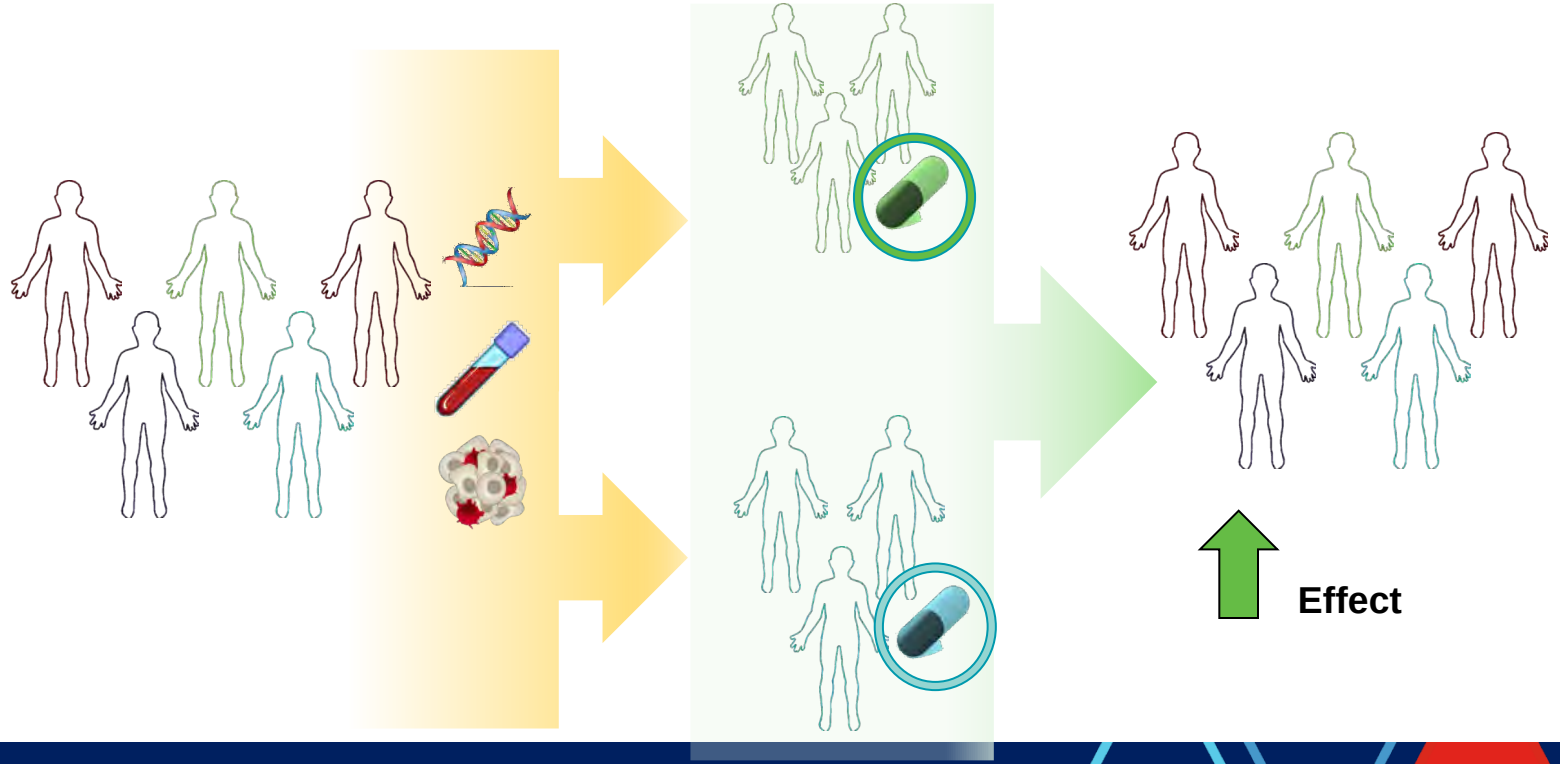
Awadalla Lab
Dr. Philip Awadalla
Mawussé Agbessi
Jarry Barber
Vanessa Bruat
Nicholas Cheng
Dr. Marie-Julie Favé
Elias Gbeha
Heather Gibling
Michelle Harwood
Ido Nofech-Mozes
Tom Ouellette
Kimberly Skead

Committee members
Dr. John Dick
Dr. Rayjean Hung

Past lab members
Dr. Armande Ang Houle
Elizabeth Hall
Jasmina Uzunović

Collaborators
Dr. David Soave

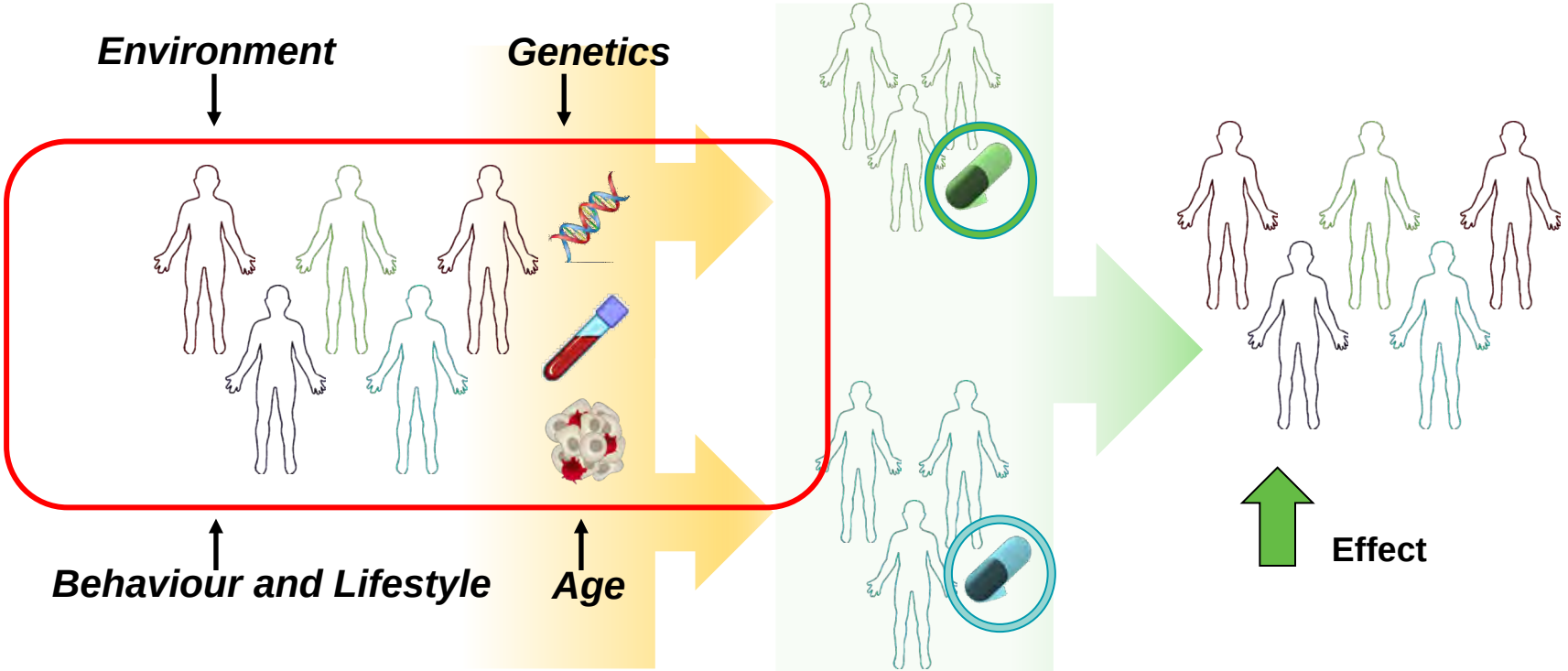
Personalized approaches to improving health outcomes



CanPath

Adapted from Barbeau, J. *Crown Bioscience Blog (2018)*

Personalized approaches to improving health outcomes



CanPath

Adapted from Barbeau, J. *Crown Bioscience Blog (2018)*

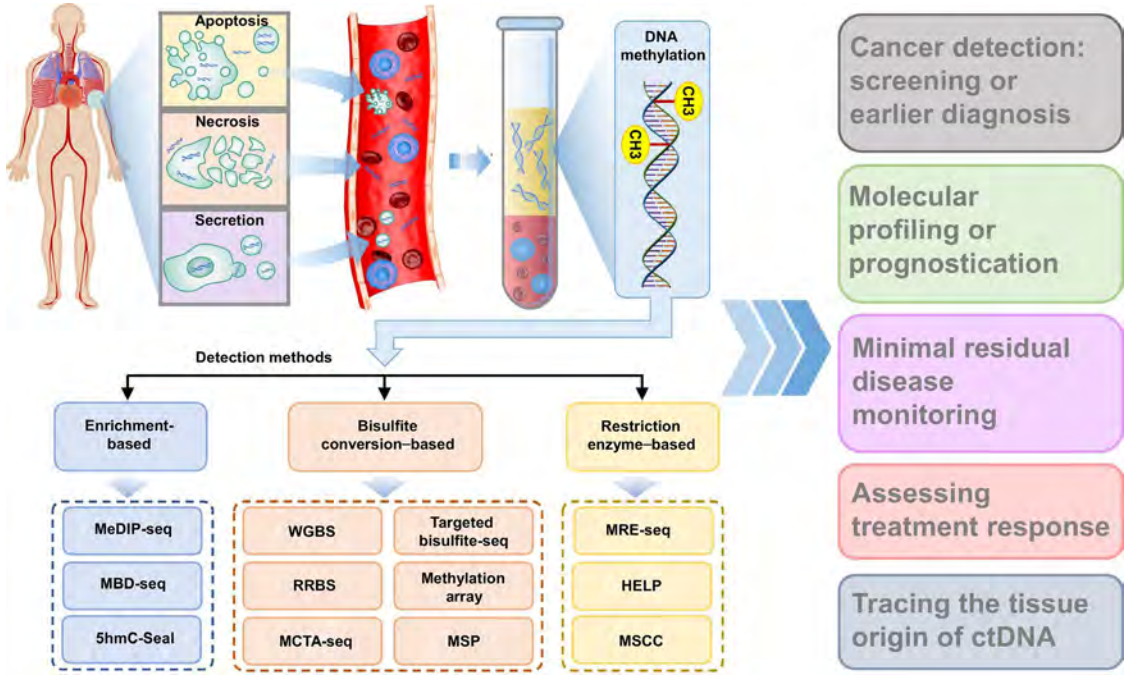
Differentially methylated regions in cell free DNA

Sources of DMRs in cfDNA

1) Disease process altering tissue or immune methylation profiles (eg cancer)

2) Increased cfDNA shedding from damaged tissues (eg organ transplant)

3) Consequence of environmental exposure (eg alcohol)



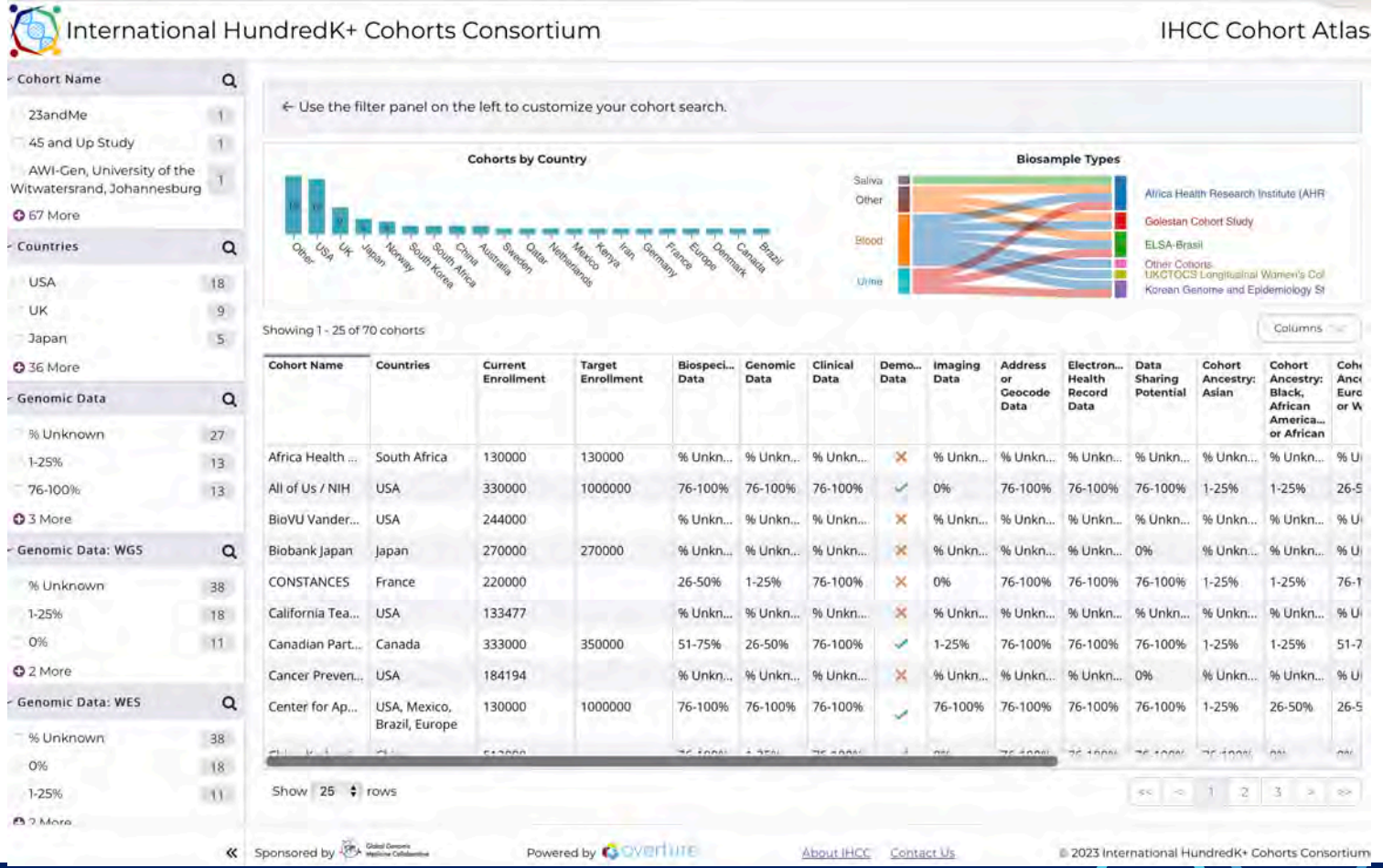
Luo et al. (2021)

Integration with international efforts

For Canada to be competitive in health research, it is crucial to have a large population cohort.

CanPath is Canada's only initiative working with other large cohorts around the world through IHCC (International Hundred Thousand Cohort Consortium)

23andMe
Biobank Japan
China Kadoorie Biobank
Canadian Partnership for Tomorrow's Health (CanPath)
EPIC
Kaiser Permanente Research Program
LifeGene
Million Veteran Program
Million Women Study
Multiethnic Cohort Study
MyCode Community Health Initiative
Nurses' Health Study (NHS/NHSII)
US Precision Medicine Initiative/All of Us
Tohoku Medical Megabank Project
UK Biobank


International HundredK+ Cohorts Consortium
IHCC Cohort Atlas
Cohort Name
23andMe 1
45 and Up Study 1
AWI-Gen, University of the Witwatersrand, Johannesburg 1
67 More
Countries
USA 18
UK 9
Japan 5
36 More
Genomic Data
% Unknown 27
1-25% 13
76-100% 13
3 More
Genomic Data: WGS
% Unknown 38
1-25% 18
0% 11
2 More
Genomic Data: WES
% Unknown 38
0% 18
1-25% 11
2 More
← Use the filter panel on the left to customize your cohort search.
Cohorts by Country
Other
USA
UK
Japan
Norway
South Korea
South Africa
China
Australia
Sweden
Qatar
Netherlands
Mexico
Kenya
Iran
Germany
France
Europe
Denmark
Canada
Brazil
Biosample Types
Saliva
Other
Blood
Urine
Africa Health Research Institute (AHR
Golestan Cohort Study
ELSA-Brasil
Other Cohorts
UKCTOCS Longitudinal Women's Col
Korean Genome and Epidemiology St
Showing 1 - 25 of 70 cohorts
Columns
Cohort Name
Countries
Current Enrollment
Target Enrollment
Biospeci... Data
Genomic Data
Clinical Data
Demo... Data
Imaging Data
Address or Geocode Data
Electron... Health Record Data
Data Sharing Potential
Cohort Ancestry: Asian
Cohort Ancestry: Black, African America... or African
Africa Health ...
South Africa
130000
130000
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
All of Us / NIH
USA
330000
1000000
76-100%
76-100%
76-100%
0%
76-100%
76-100%
76-100%
1-25%
1-25%
BioVU Vander...
USA
244000
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
Biobank Japan
Japan
270000
270000
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
0%
% Unkn...
% Unkn...
CONSTANCES
France
220000
26-50%
1-25%
76-100%
0%
76-100%
76-100%
76-100%
1-25%
1-25%
California Tea...
USA
133477
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
Canadian Part...
Canada
333000
350000
51-75%
26-50%
76-100%
1-25%
76-100%
76-100%
76-100%
1-25%
1-25%
Cancer Preven...
USA
184194
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
% Unkn...
0%
% Unkn...
% Unkn...
Center for Ap...
USA, Mexico, Brazil, Europe
130000
1000000
76-100%
76-100%
76-100%
76-100%
76-100%
76-100%
76-100%
1-25%
26-50%
Show 25 rows
Sponsored by
Powered by overture
About IHCC
Contact Us
© 2023 International HundredK+ Cohorts Consortium

Predicting health outcomes from hematopoietic evolution

KIMBERLY SKEAD,

DEPT. MOLECULAR GENETICS, UNIVERSITY OF TORONTO AND THE ONTARIO INSTITUTE FOR CANCER RESEARCH

NATIONAL SCIENTIFIC COORDINATOR, CANADIAN PARTNERSHIP FOR TOMORROW'S HEALTH (CANPATH)

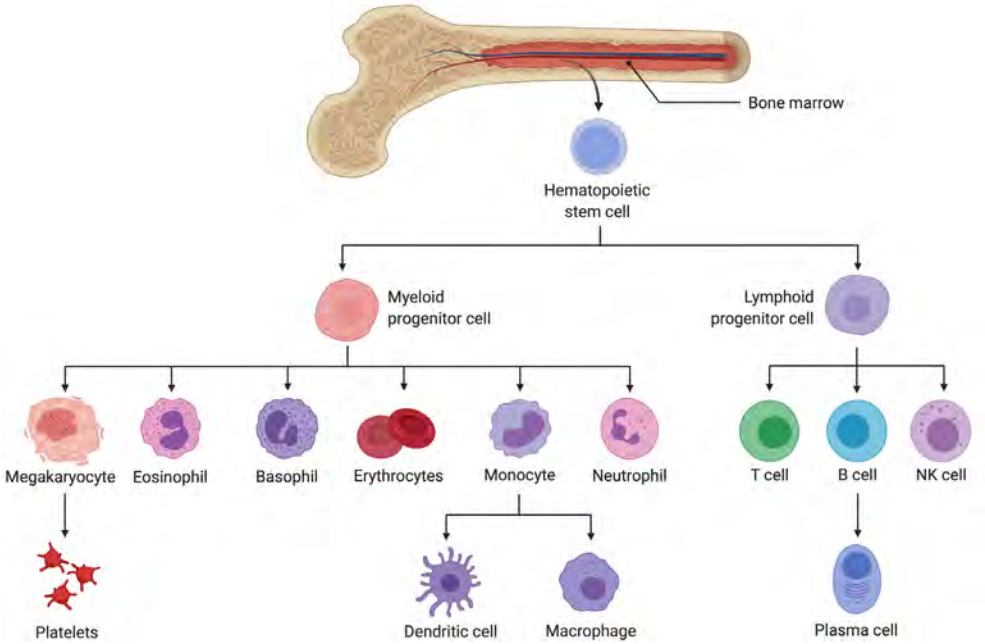
AWADALLA LAB
Pioneering Genomics for Precision Health.

UNIVERSITY OF TORONTO

CanPath
for Tomorrow's Health

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 potentially other chronic diseases**



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**

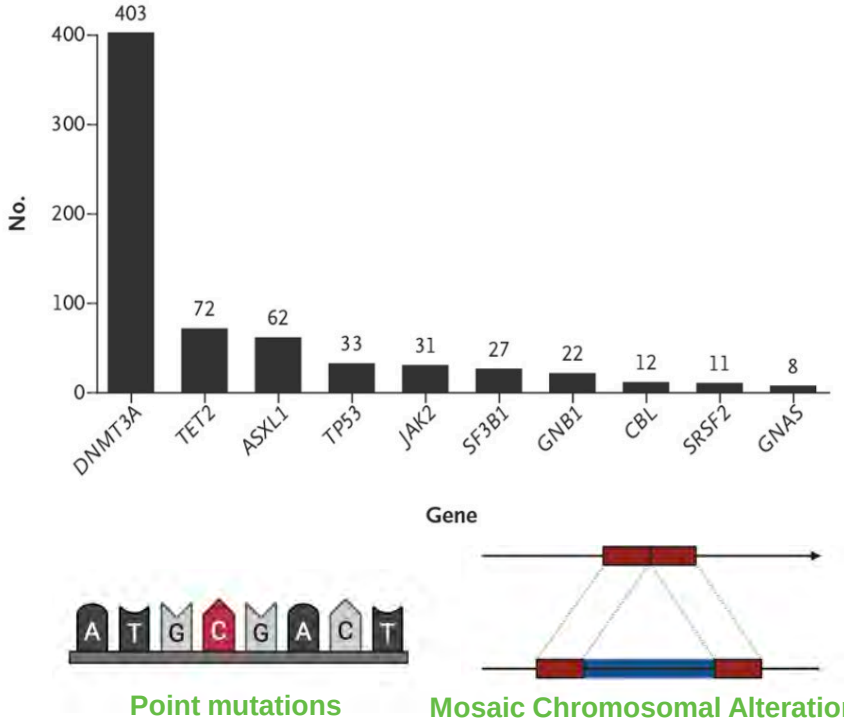
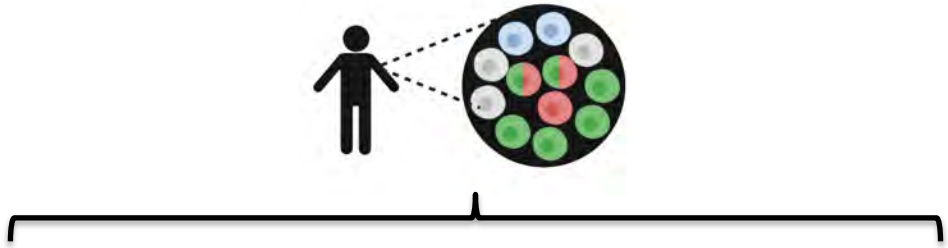


Figure from Jaiswal et al. NEJM (2014), Jaiswal et al. 2014, Genovese et al. 2014, Xie et al. 2014, Loh et al. 2018

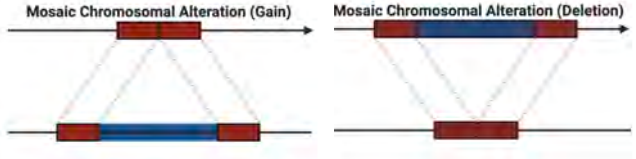
Interacting evolutionary pressures in blood shape health outcomes as we age



How does our blood evolve as we progress to disease?

Skead, K., et al. *Nature Communications* (2021)

Why are large mutations tolerated in our blood?



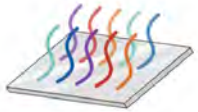
Skead, K., et al. In review at *Nature Genetics*

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



Genotype Array

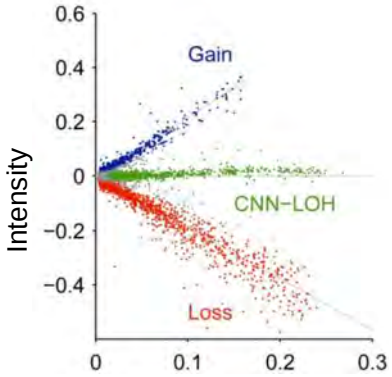


Figure adapted from Loh, P. et al. 2018

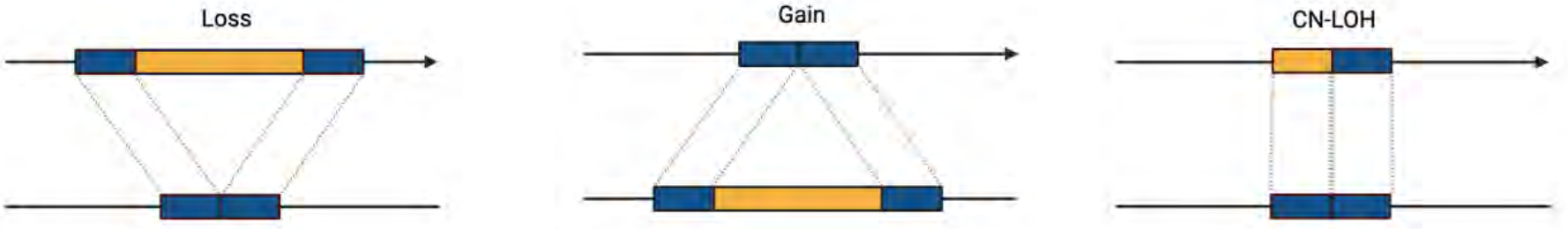
Why are large mutations tolerated in our blood?

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**



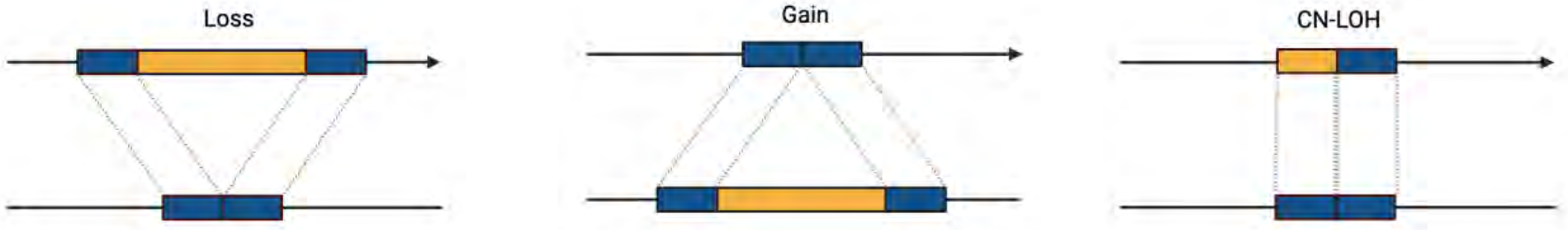
Why are large mutations tolerated in our blood?

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

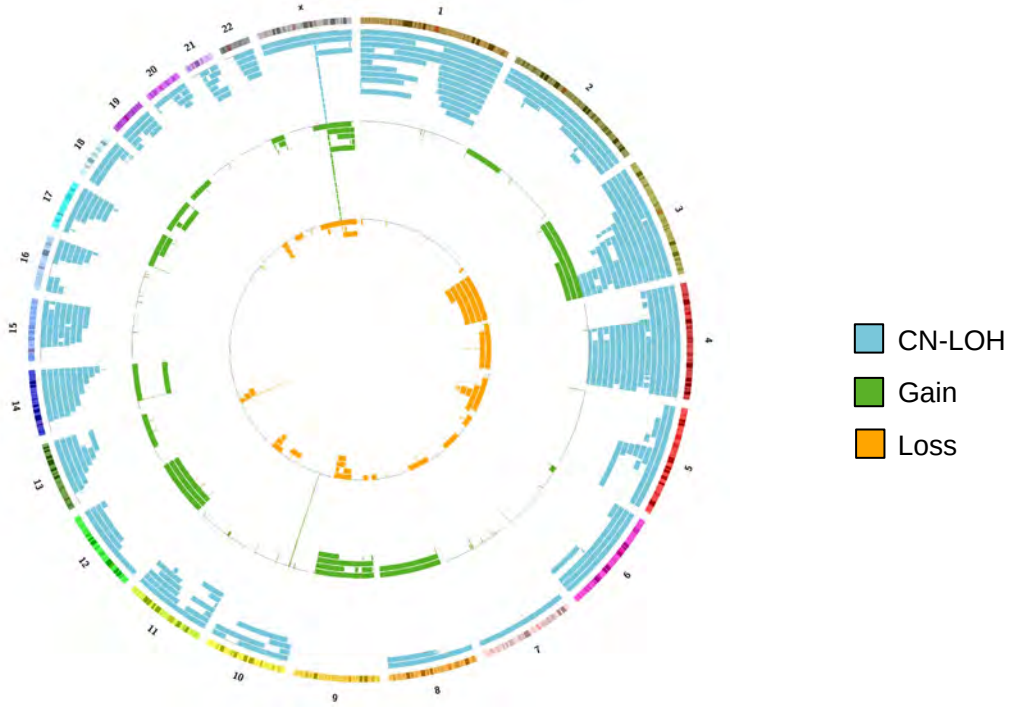
Nature **559**, 350–355 (2018) | Cite this article

- 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?**



ARCH attributable to mosaic chromosomal alterations is three times more common than previously estimated

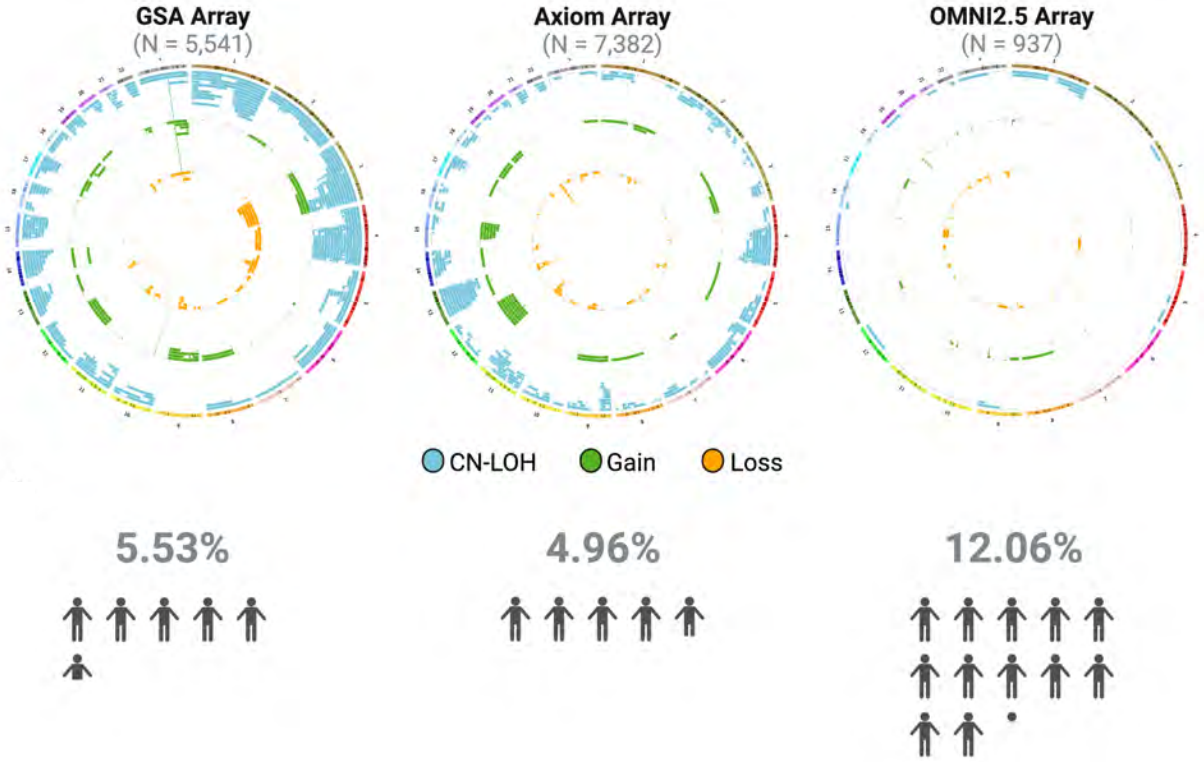
Mosaic chromosomal alterations were called from genotype array data across ~14,000 individuals



ARCH attributable to mosaic chromosomal alterations is three times more common than previously estimated

Mosaic chromosomal alterations were called from genotype array data across ~14,000 individuals

We capture a higher prevalence of mCAs (2.5x) than previously reported using denser sequencing arrays

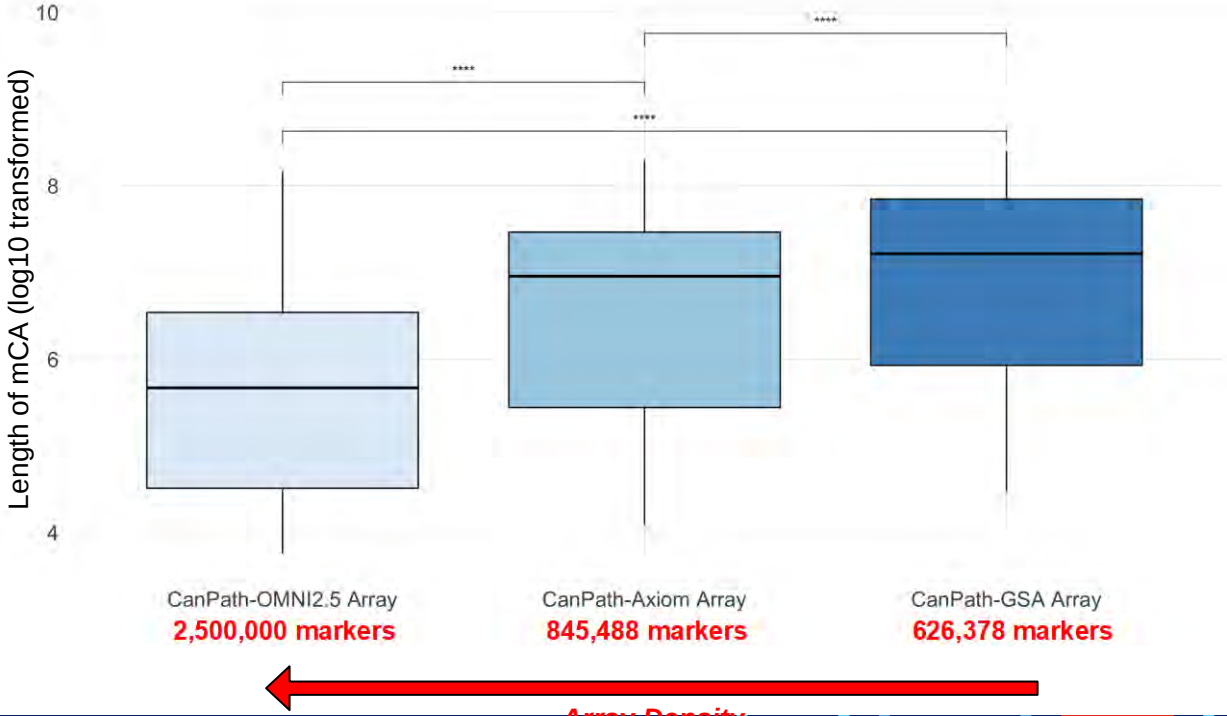


ARCH attributable to mosaic chromosomal alterations is three times more common than previously estimated

Mosaic chromosomal alterations were called from genotype array data across ~14,000 individuals

We capture a higher prevalence of mCAs (3x) than previously reported using denser sequencing arrays

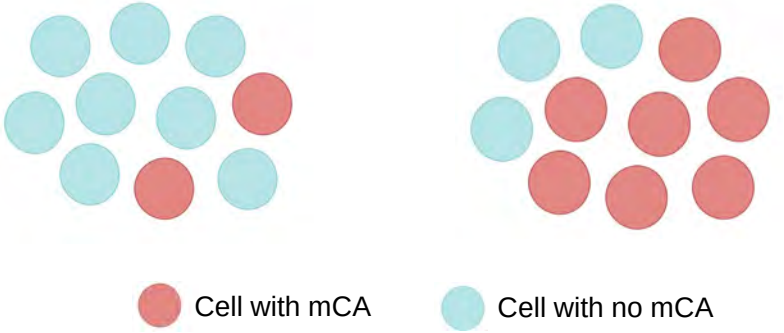
Higher density arrays enable us to detect smaller mCAs that were previously missed



Determining the impact of selection on shaping mCA accumulation in blood

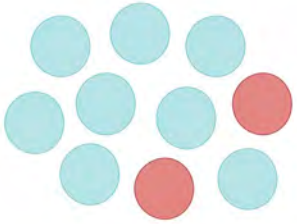
Low cell fraction

High cell fraction

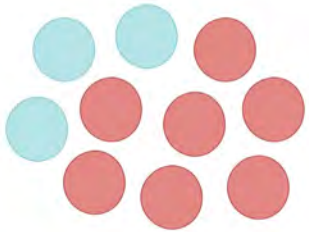


Determining the impact of selection on shaping mCA accumulation in blood

Low cell fraction



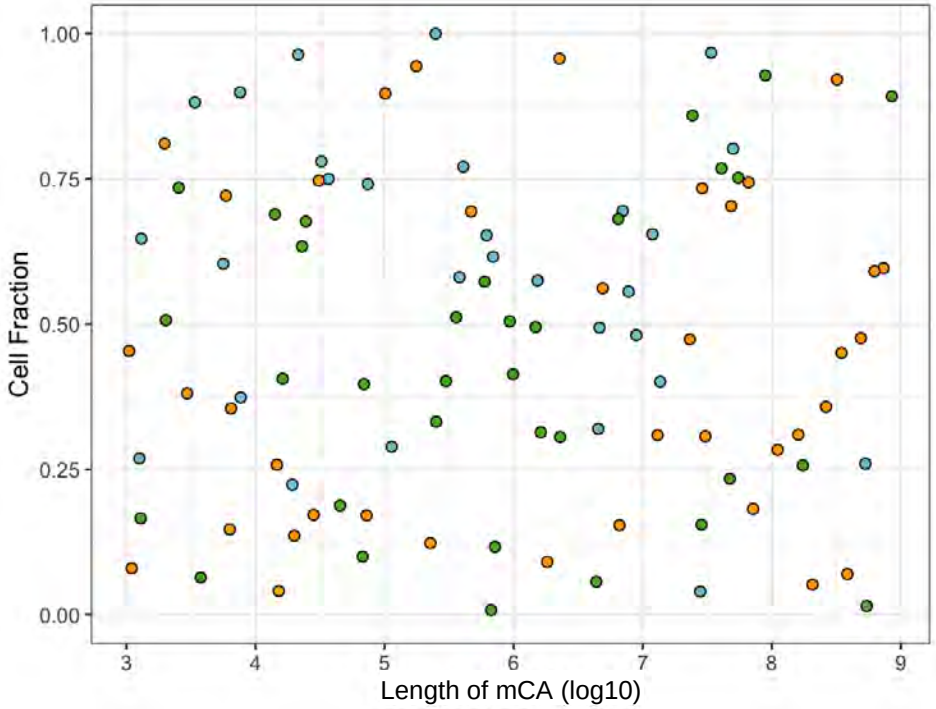
High cell fraction



Cell with mCA

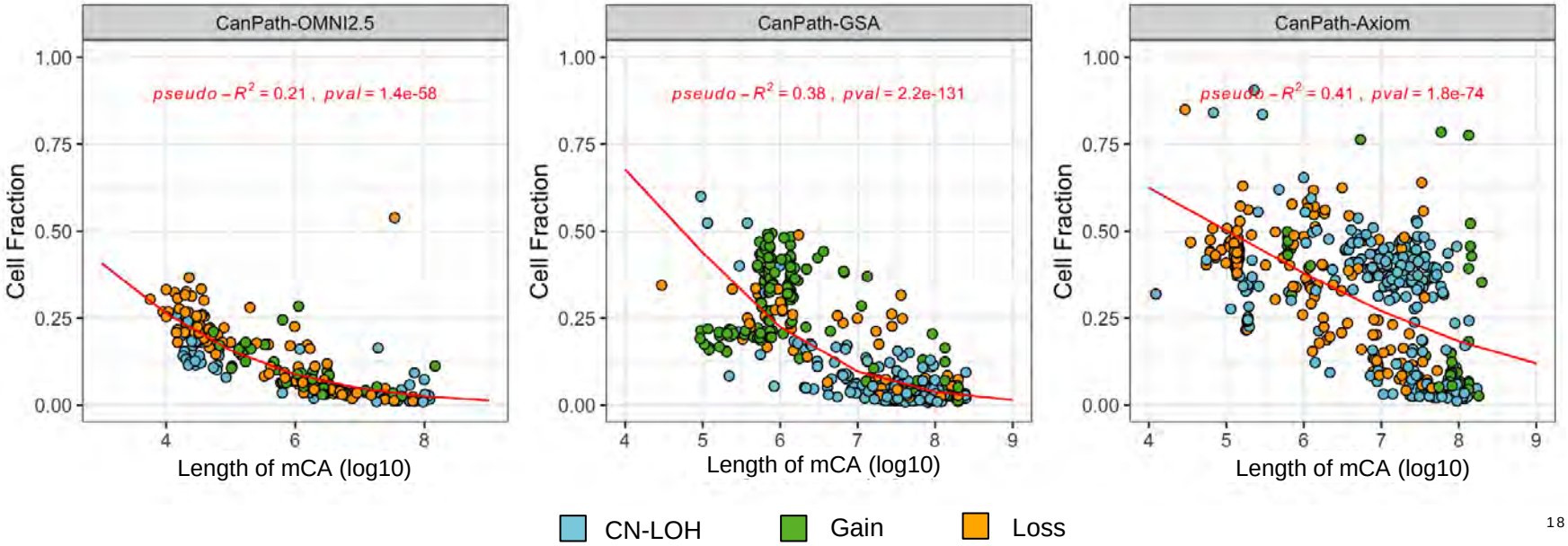
Cell with no mCA

Under a **neutral model of evolution, we would not expect to see an association** between the frequency of a mCA and the size of a mCA



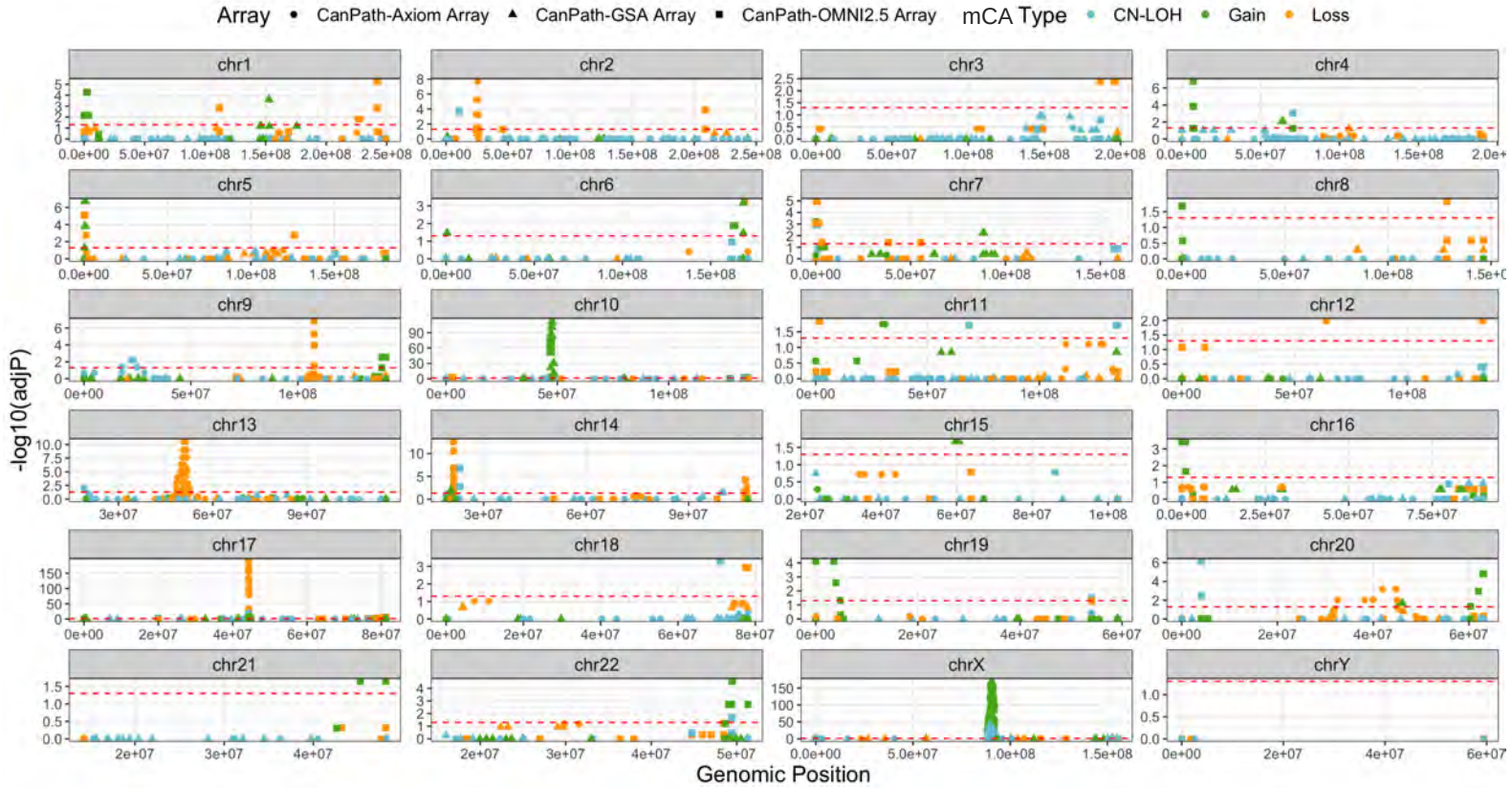
Larger mosaic chromosomal alterations are observed at low frequencies in the hematopoietic pool

- The size of mCAs impact the frequency at which they segregate in our blood
- **Negative selection plays a role in removing large mCAs** from the hematopoietic population

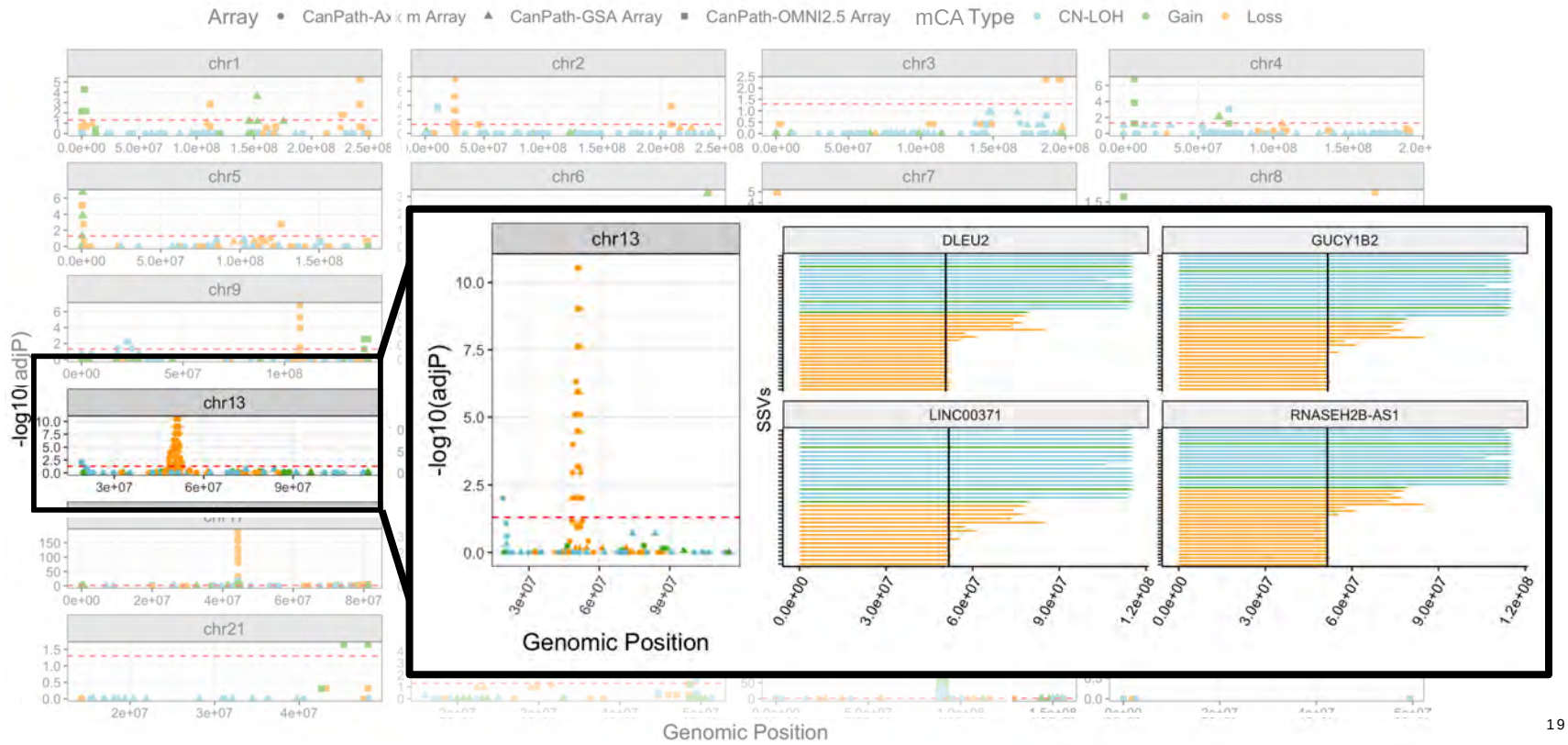


CanPath

Genome-wide hotspotting approach detects regions which harbour a high burden of mosaic chromosomal alterations



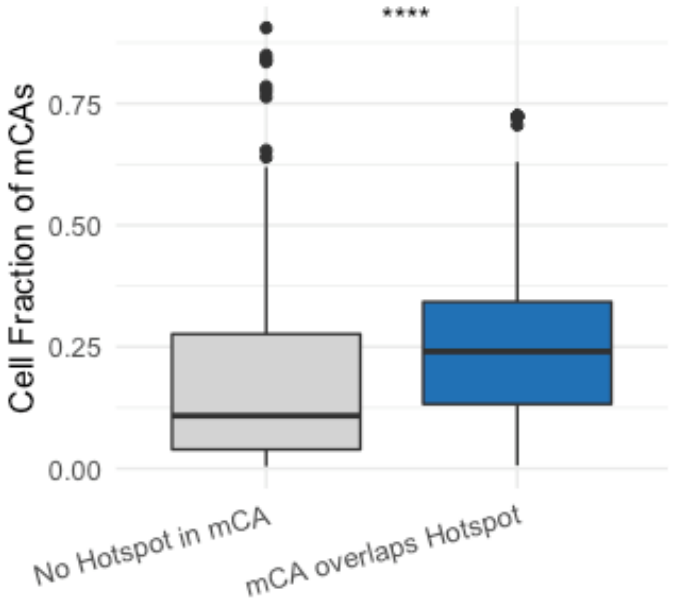
mCAs accumulate across ARCH- and cancer-associated genes



mCAs accumulate across ARCH- and cancer-associated genes

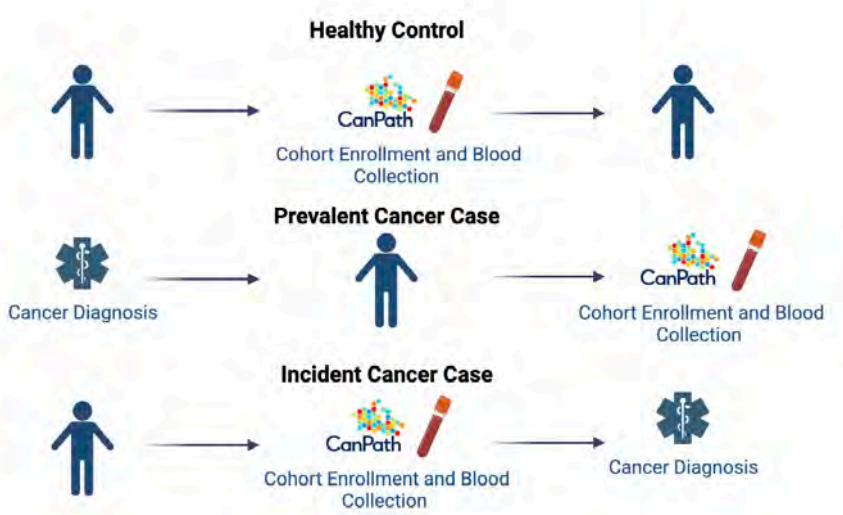
mCA hotspots suggest positive selection is shaping mCA retention and frequencies across the genome

- mCAs accumulate across ARCH- and cancer-associated genes
- mCAs which overlap at least one hotspot are at a significantly higher cell fraction than mCAs which do not overlap hotspots
- Positive selection may be retaining cell fractions at higher frequencies at select regions of the genome



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

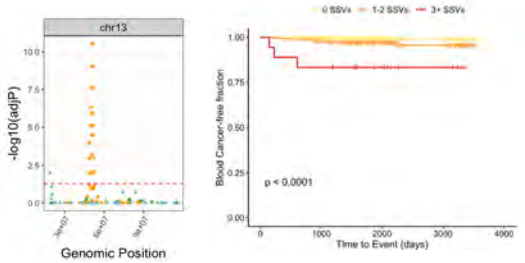


Variable		N	Hazard ratio	p
Number of mCAs	0	7306		Reference	
	1-2	272		5.06 (2.47, 10.38)	<0.001
	3+	18		26.80 (8.30, 86.56)	<0.001
Age		7596		1.08 (1.04, 1.12)	<0.001
Sex	FEMALE	4721		Reference	
	MALE	2875		2.63 (1.52, 4.56)	<0.001

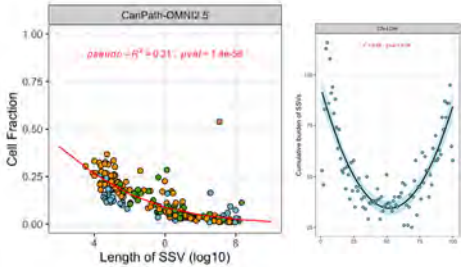
1 2 5 10 20 50

Key take-aways

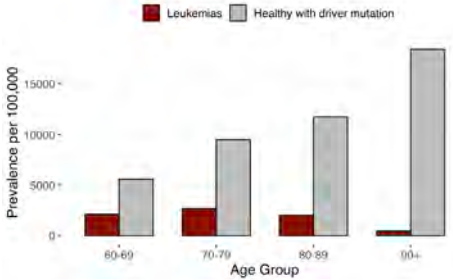
mCA-associated ARCH is much more common in the population than previously estimated



mCAs are preferentially retained at regions implicated in cancer and ARCH and are associated with an increase in cancer risk



Selection plays a role in maintaining structural variability in the population and on shaping the impact of mCAs on gene expression

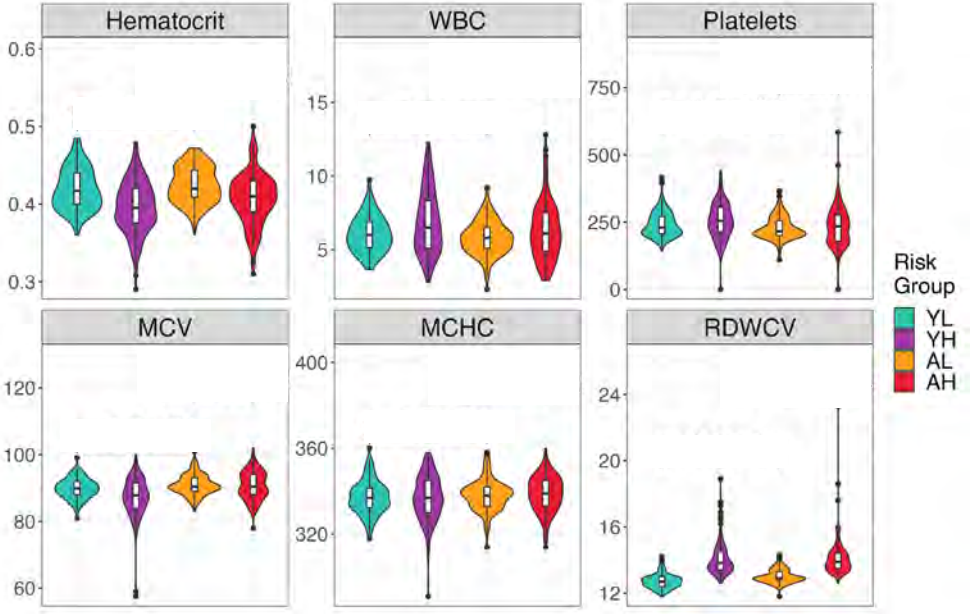


Interrogating mutational spectra through a multi-omics lens can shed light on how somatic mutations impact cancer risk

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+

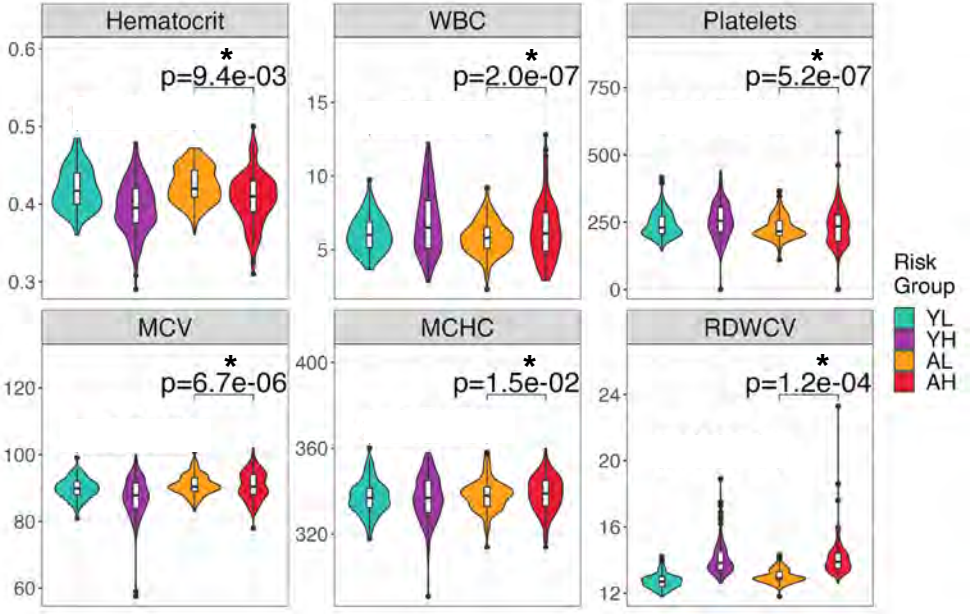


p-values calculated from Levene's test for variance

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