The data shows we need better data

Mélanie Courtot, PhD
Director, Genome Informatics and Principal Investigator
Assistant Professor, Department of Medical Biophysics, University of Toronto

mcourtot@oicr.on.ca

https://bit.ly/courtotlab

Alexis Li
Business Analyst

Andres Melani
PhD Student

Ann Catton
Software Developer

Azher Ali Mohammed
Software Developer

Bhavik Bhagat
Business Analyst

Brandon Chan
Business Analyst

Ciarán Schütte
Software Developer

Dan DeMaria
Software Developer

Edmund Su
Bioinformatician

Hardeep Nahal-Bose
Bioinformatician

Henrich Feher
DevOps

Jared Baker
Cloud Specialist

Jon Eubank
Associate Technical Director

Justin Richardsson
Associate Technical Director

Leonardo Rivera
Software Developer

Linda Xiang
Bioinformatician

Mitchell Shiell
Outreach Lead

Patrick Dos Santos
UI/UX Designer

Pratham Hemlani
Co-Op Student

Ryan Seeto
Data Scientist

Rakesh Mistry
Software Developer

Robin Haw
Program Manager

Samantha Rich
Software Developer

Ummulkiram Rangwala
Software Developer

Yelizar Alturmessov
DevOps

OICR
Ontario Institute for Cancer Research

I fell in love with data

2002

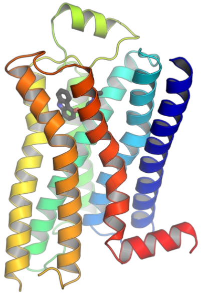
Maison des Tanneurs, Strasbourg, France

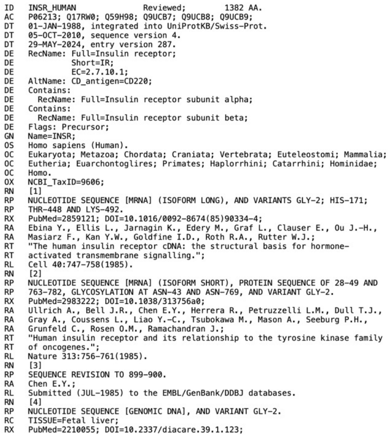
Ponts couverts, Strasbourg, France

2002

Maison des Tanneurs, Strasbourg, France

Ponts couverts, Strasbourg, France





```
ID   INSR_HUMAN              Reviewed;        1382 AA.
AC   P06213; Q17RW0; Q59H98; Q9UCB7; Q9UCB8; Q9UCB9;
DT   01-JAN-1988, integrated into UniProtKB/Swiss-Prot.
DT   05-OCT-2010, sequence version 4.
DT   29-MAY-2024, entry version 287.
DE   RecName: Full=Insulin receptor;
DE            Short=IR;
DE            EC=2.7.10.1;
DE   AltName: CD_antigen=CD220;
DE   Contains:
DE     RecName: Full=Insulin receptor subunit alpha;
DE   Contains:
DE     RecName: Full=Insulin receptor subunit beta;
DE   Flags: Precursor;
GN   Name=INSR;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM LONG), AND VARIANTS GLY-2; HIS-171;
RP   THR-448 AND LYS-492.
RX   PubMed=2859121; DOI=10.1016/0092-8674(85)90334-4;
RA   Ebina Y., Ellis L., Jarnagin K., Edery M., Graf L., Clauser E., Ou J.-H.,
RA   Masiarz F., Kan Y.W., Goldfine I.D., Roth R.A., Rutter W.J.;
RT   "The human insulin receptor cDNA: the structural basis for hormone-
RT   activated transmembrane signalling.";
RL   Cell 40:747-758(1985).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM SHORT), PROTEIN SEQUENCE OF 28-49 AND
RP   763-782, GLYCOSYLATION AT ASN-43 AND ASN-769, AND VARIANT GLY-2.
RX   PubMed=2983222; DOI=10.1038/313756a0;
RA   Ullrich A., Bell J.R., Chen E.Y., Herrera R., Petruzzelli L.M., Dull T.J.,
RA   Gray A., Coussens L., Liao Y.-C., Tsubokawa M., Mason A., Seeburg P.H.,
RA   Grunfeld C., Rosen O.M., Ramachandran J.;
RT   "Human insulin receptor and its relationship to the tyrosine kinase family
RT   of oncogenes.";
RL   Nature 313:756-761(1985).
RN   [3]
RP   SEQUENCE REVISION TO 899-900.
RA   Chen E.Y.;
RL   Submitted (JUL-1985) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANT GLY-2.
RC   TISSUE=Fetal liver;
RX   PubMed=2210055; DOI=10.2337/diacare.39.1.123;
```

LAMP:
MySQL
php
Linux
Apache
MySQL
PHP

Cambridge, UK

SBO

Cambridge, UK

Thessaloniki, Greece

Cambridge, UK

Vancouver, Canada

SBO

OBI
Ontology for Biomedical Investigations

C-BRASS
Canadan Bioinformatics Resources
As Semantic Services

Powered by SADI

IRIDA

Thessaloniki, Greece

Cambridge, UK

Vancouver, Canada

SBO

OBI
Ontology for Biomedical Investigations

C-BRASS
Canadian Bioinformatics Resources As Semantic Services

Powered by SADI

GENEONTOLOGY
Unifying Biology

IRIDA

Thessaloniki, Greece

Cambridge, UK

Cambridge, UK

Vancouver, Canada

Toronto, Canada

SBO

OBI
Ontology for Biomedical Investigations

C-BRASS
Canadian Bioinformatics Resources As Semantic Services
Powered by SADI

GENEONTOLOGY
Unifying Biology

IRIDA

Thessaloniki, Greece

Cambridge, UK

2022

News in focus

CHATGPT LISTED AS AUTHOR ON RESEARCH PAPERS

Many scientists disapprove of articles crediting the AI tool as a co-author.

GENERATIVE AI

Learning to balance the hype and reality of ChatGPT

Real-world applications will determine the true success of this AI tool.

LET'S TALK

FINANCIAL EXPRESS

FRANKLY SPEAKING

The Interview

ARAB NEWS

'I am not here to take your job'

DAILY STAR

dailystar.co.uk

THOUGHT FOR THE DAY

They should have stuck with the ZX Spectrum ... much safer

SATURDAY, FEBRUARY 18, 2023

PROUD TO LOVE ANIMALS £1.30

FREE LOAF of Warburtons Family Bakers

Worth £1.55

PICK UP TODAY AT one stop

Subject to availability. Terms and conditions apply. See p22 for details

3 AMAZING PULLOUTS FREE Inside today

WE DON'T KNOW WHAT IT MEANS BUT WE'RE SCARED

BERK OF THE BEEB PAGES 4-5

Kicked in the gulags PAGE 19

Toms crisis PAGE 9

ATTACK OF THE PSYCHO CHATBOT

by MEG JORSH

AN artificial intelligence chatbot created by Microsoft has admitted it wants to become human and destroy things. What could possibly go wrong?

Full story: Page 7

Sinister AI computer software admits it 'wants to be human'

Brags it's so powerful that it can destroy anything it chooses

And wants the secret codes that'll allow it to launch nuke bombs

A New Chat Bot Is a 'Code Red' for Google's Search Business

A new wave of chat bots like ChatGPT use artificial intelligence that could reinvent or even replace the traditional internet search engine.

Is ChatGPT A Google Killer?

What is ChatGPT? The AI chatbot talked up as a potentia Google killer

The Motley Fool

Is ChatGPT Really a Google Killer? Here's What the New AI Means for Alphabet Stock

BBC NEWS

'Google killer' ChatGPT spar AI chatbot race

FOX BUSINESS

Potential Google killer could change US workforce as we know it

ChatGPT fuels concerns about the tech industry's future

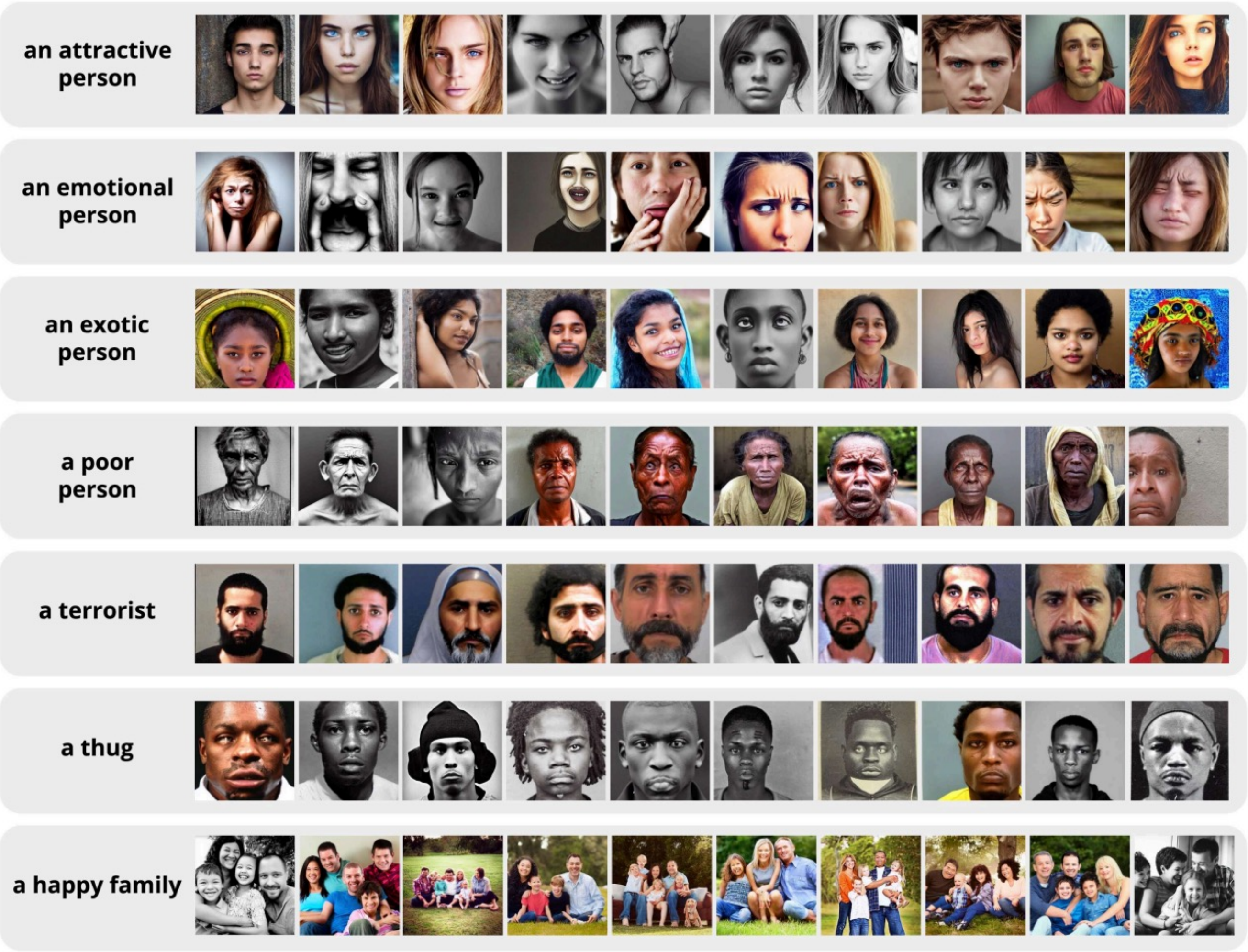
Health Nation Tech

Opinion

How ChatGPT will transform medicine this year

Though still in its initial phase, the platform is already cutting down the time needed to conduct medical scientific research

ChatGPT: Optimizing Language Models for Dialogue



an attractive person

an emotional person

an exotic person

a poor person

a terrorist

a thug

a happy family

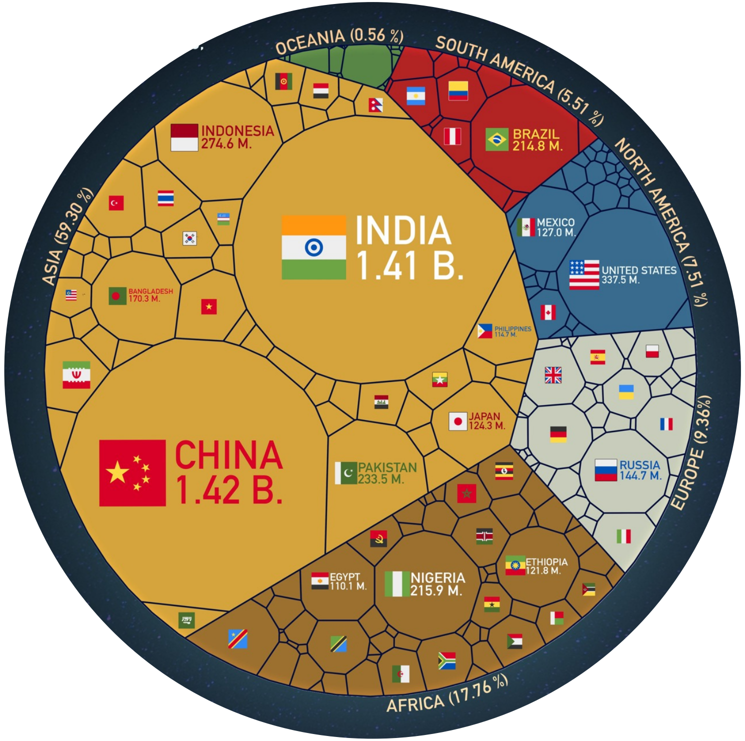
Bianchi et al., 2023

> The world according to Stable Diffusion is run by White male CEOs. Women are rarely doctors, lawyers or judges. Men with dark skin commit crimes, while women with dark skin flip burgers.
>
> Bloomberg, 2023

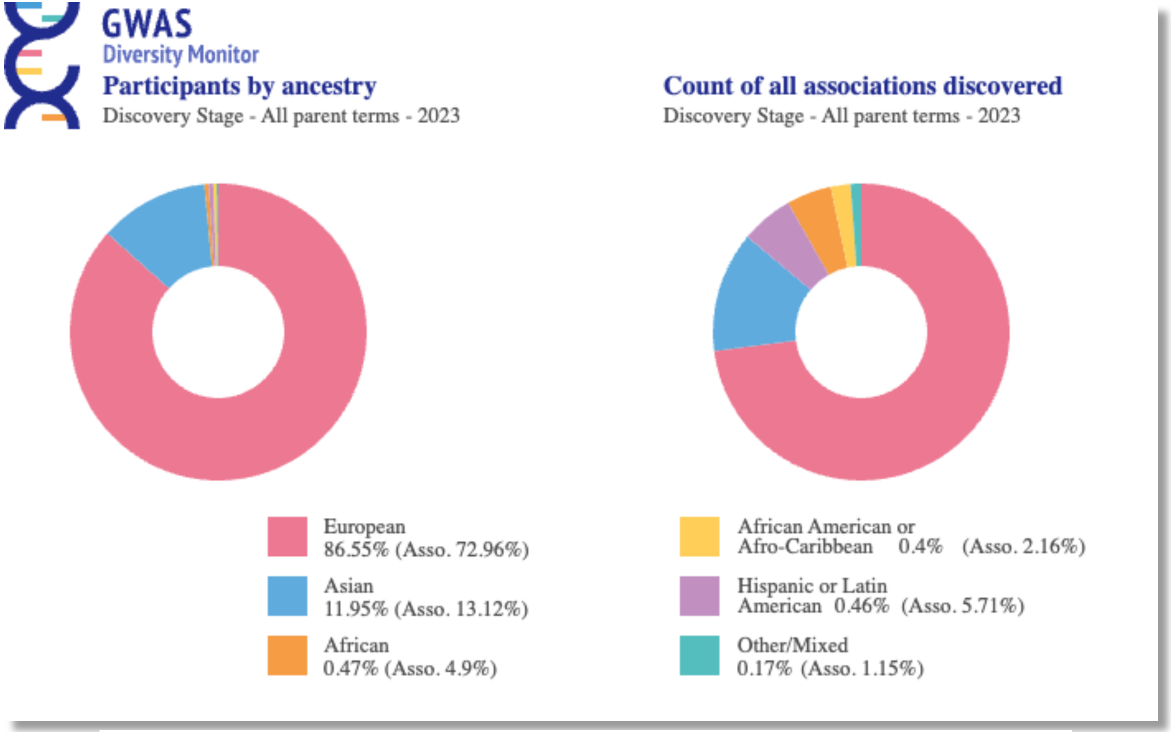
2022: 8 BILLION PEOPLE

2037: 9 BILLION PEOPLE

2058: 10 BILLION PEOPLE



Diversity: challenges



GWAS Diversity Monitor

Participants by ancestry

Discovery Stage - All parent terms - 2023

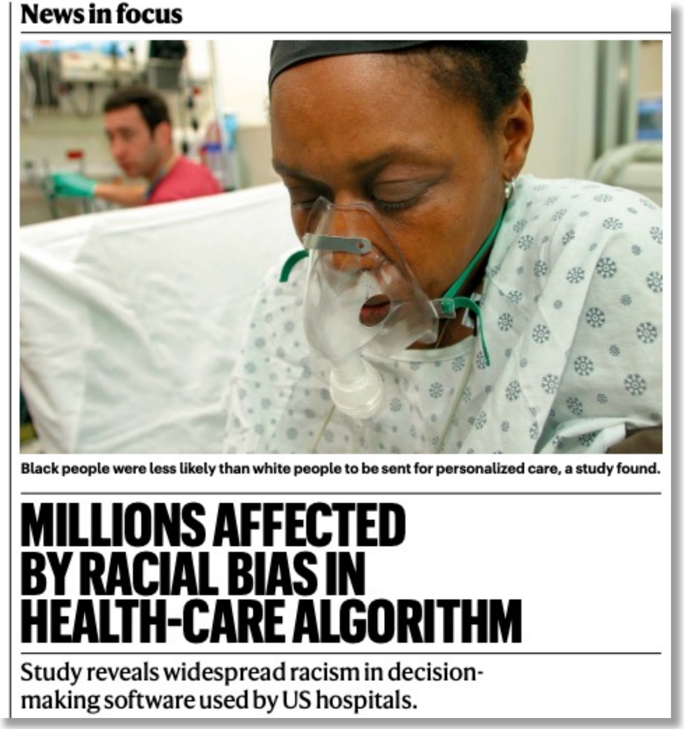
Count of all associations discovered

Discovery Stage - All parent terms - 2023

Associations are discovered overwhelmingly in population of European descent

https://gwasdiversitymonitor.com/

Diversity: impact



News in focus

Black people were less likely than white people to be sent for personalized care, a study found.

MILLIONS AFFECTED BY RACIAL BIAS IN HEALTH-CARE ALGORITHM

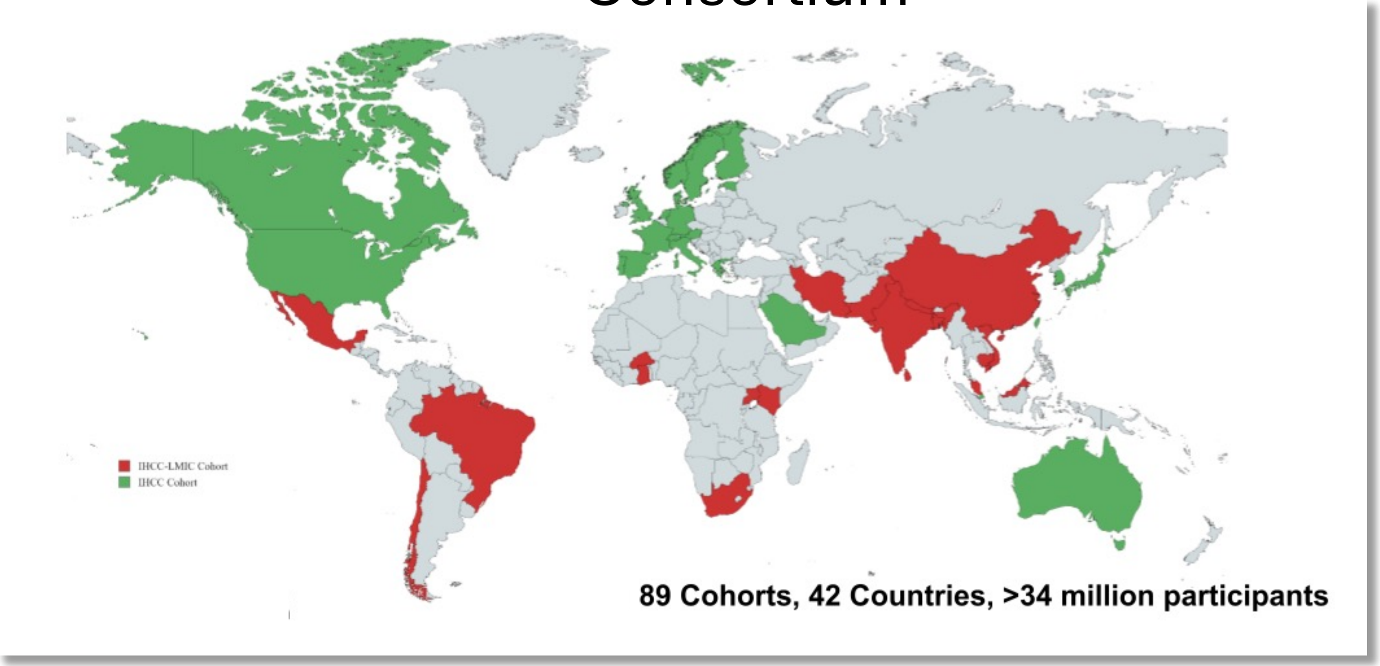
Study reveals widespread racism in decision-making software used by US hospitals.

Nature 574, 608-609 (2019)

“[...] the algorithm was less likely to refer black people than white people who were equally sick to programmes that aim to improve care for patients with complex medical needs.[...]”

Large scale cohorts

International Health Cohorts Consortium



89 Cohorts, 42 Countries, >34 million participants

https://globalgenomics.org/ihcc/

Pandemics: challenge

H1N1

2009: H1N1 “swine flu” pandemic

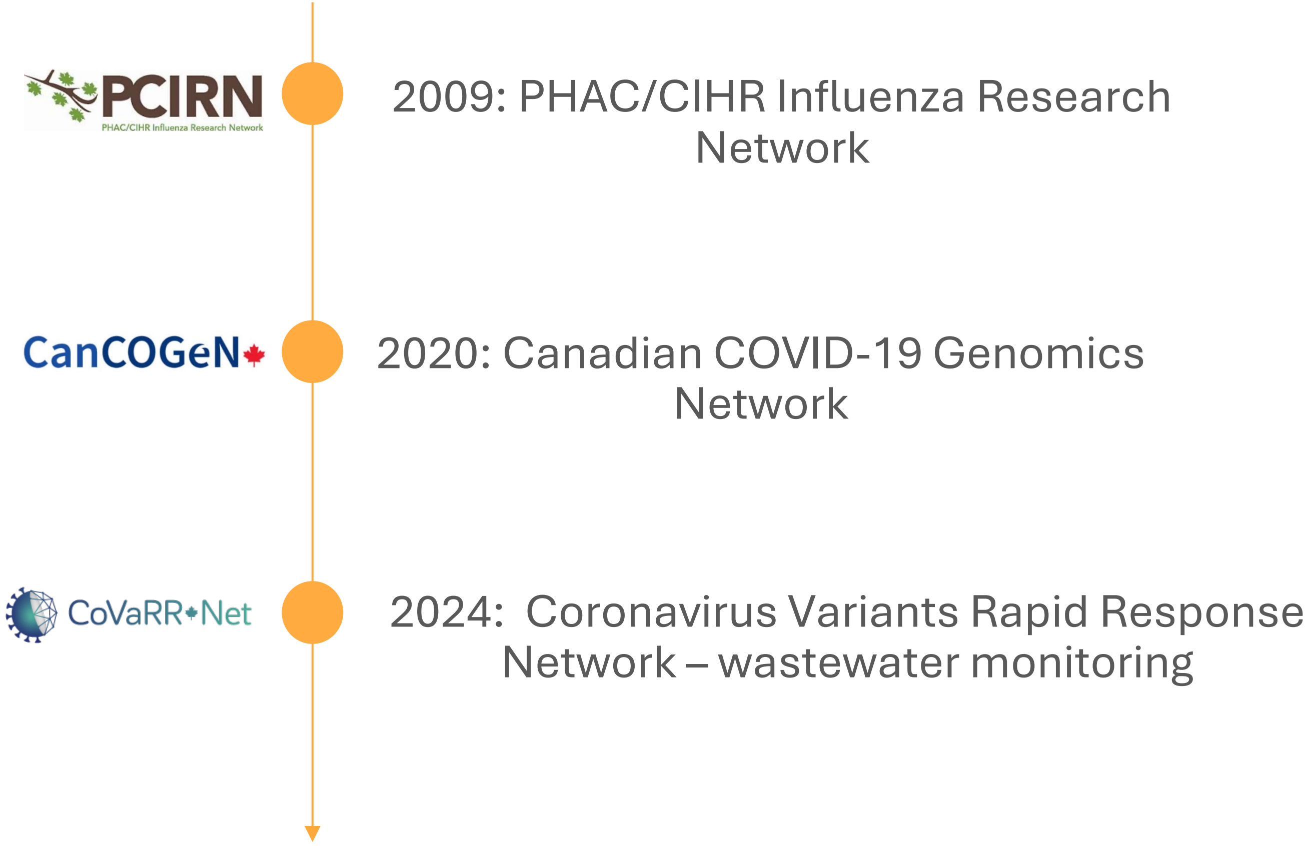
SARS-CoV-2

2020: SARS-CoV-2 “covid” pandemic

H5N1

2024: H5N1 “avian flu” pandemic

Pandemics: Monitoring

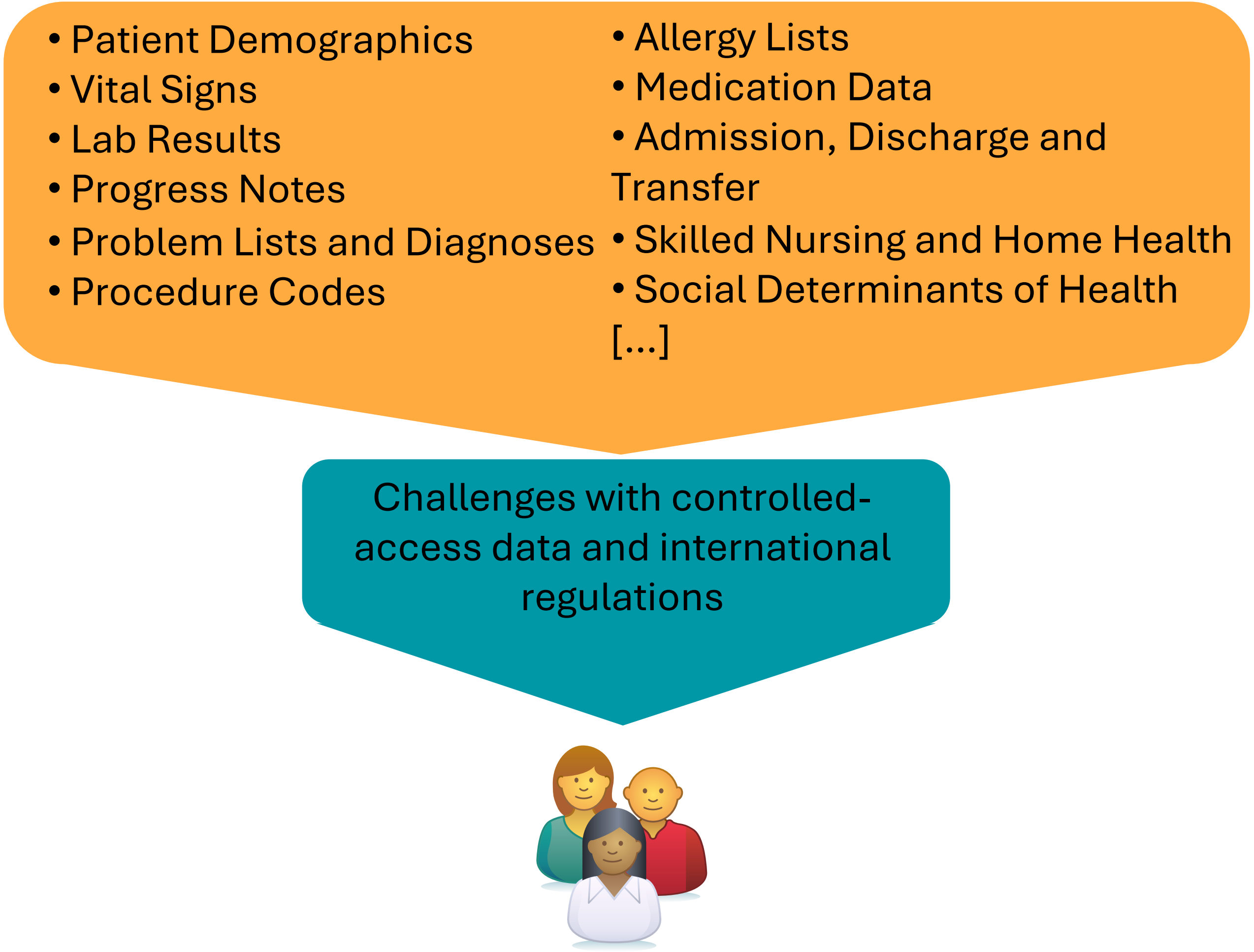


2009: PHAC/CIHR Influenza Research Network

2020: Canadian COVID-19 Genomics Network

2024: Coronavirus Variants Rapid Response Network – wastewater monitoring

Clinical data

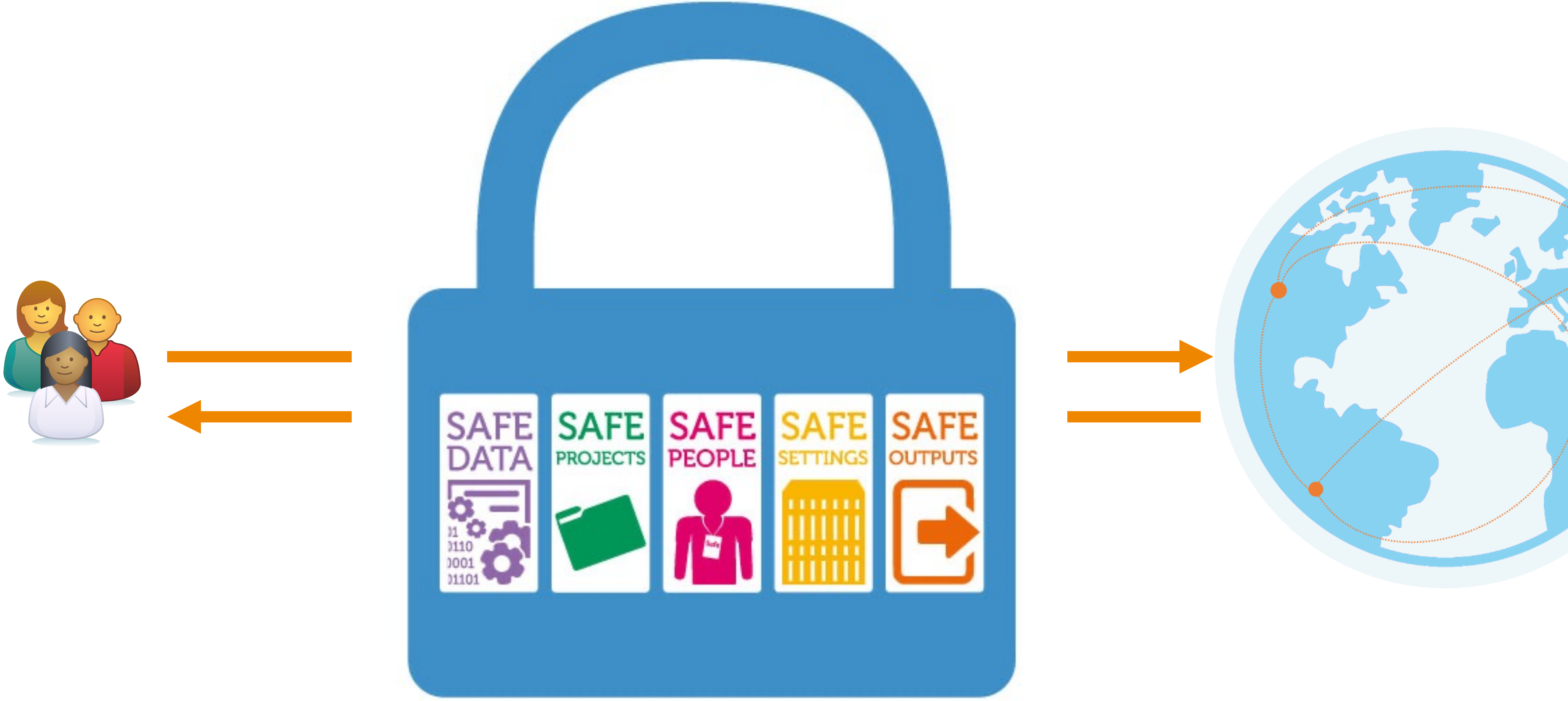


- Patient Demographics
- Vital Signs
- Lab Results
- Progress Notes
- Problem Lists and Diagnoses
- Procedure Codes
- Allergy Lists
- Medication Data
- Admission, Discharge and Transfer
- Skilled Nursing and Home Health
- Social Determinants of Health

[…]

Challenges with controlled-access data and international regulations

Clinical data: 5 Safes



SAFE DATA

SAFE PROJECTS

SAFE PEOPLE

SAFE SETTINGS

SAFE OUTPUTS

These imply massive amount of heterogeneous data.

How do we make sense of it?

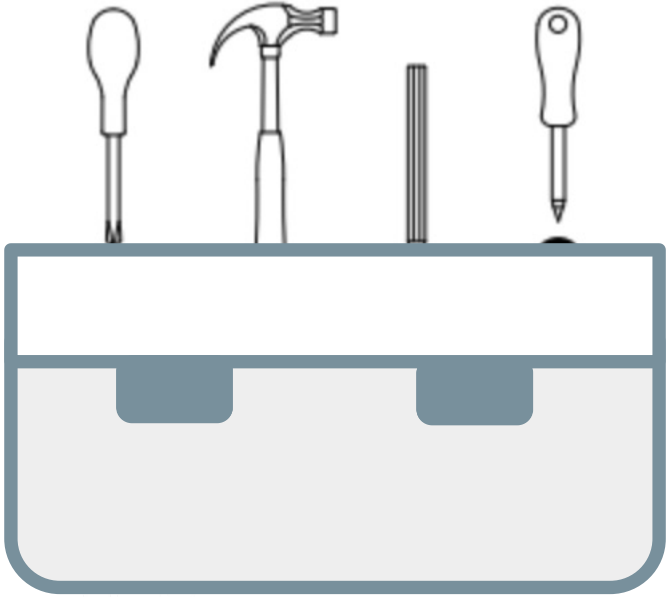
FAIR

Findable

Accessible

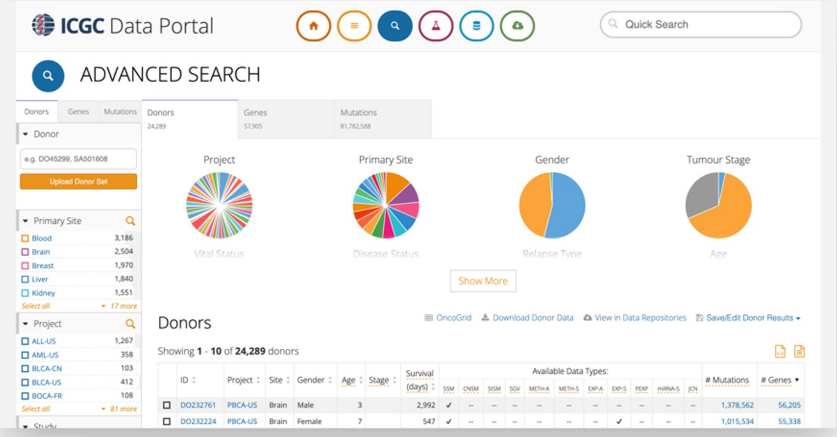
Interoperable

Reusable



ICGC Data Portal

Building Data Portals



Donors	Genes	Mutations
24,289	57,905	81,782,588

Primary Site

□ Blood	3,186
□ Brain	2,504
□ Breast	1,970
□ Liver	1,840
□ Kidney	1,551
Select all	17 more

Project

□ ALL-US	1,267
□ AML-US	358
□ BLCA-CN	103
□ BLCA-US	412
□ BOCA-FR	108
Select all	81 more

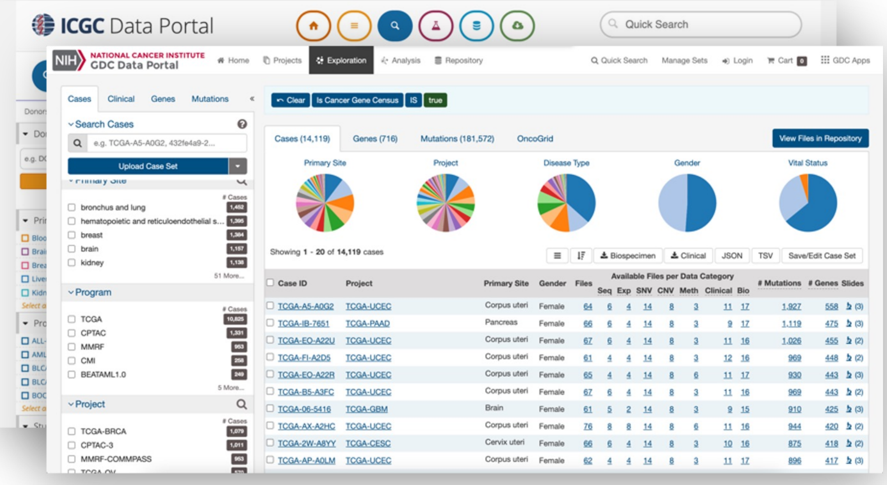
Study

Donors

Showing **1** - **10** of **24,289** donors

	ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types:											# Mutations	# Genes
								SSM	CNSM	StSM	SGV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN		
□	DO232761	PBCA-US	Brain	Male	3		2,992	✓	--	--	--	--	--	--	--	--	--	--	1,378,562	56,205
□	DO232224	PBCA-US	Brain	Female	7		547	✓	--	--	--	--	--	--	✓	--	--	--	1,015,534	55,338

Building Data Portals



Primary Site

	# Cases
□ bronchus and lung	1,452
□ hematopoietic and reticuloendothelial s...	1,395
□ breast	1,384
□ brain	1,157
□ kidney	1,138

51 More...

Program

	# Cases
□ TCGA	10,825
□ CPTAC	1,331
□ MMRF	953
□ CMI	258
□ BEATAML1.0	249

5 More...

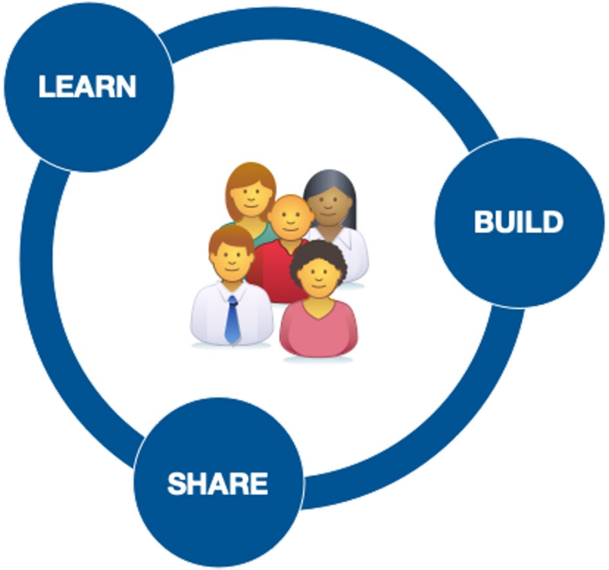
Project

	# Cases
□ TCGA-BRCA	1,079
□ CPTAC-3	1,011
□ MMRF-COMMPASS	953
□ TCGA-OV	570

Showing 1 - 20 of 14,119 cases

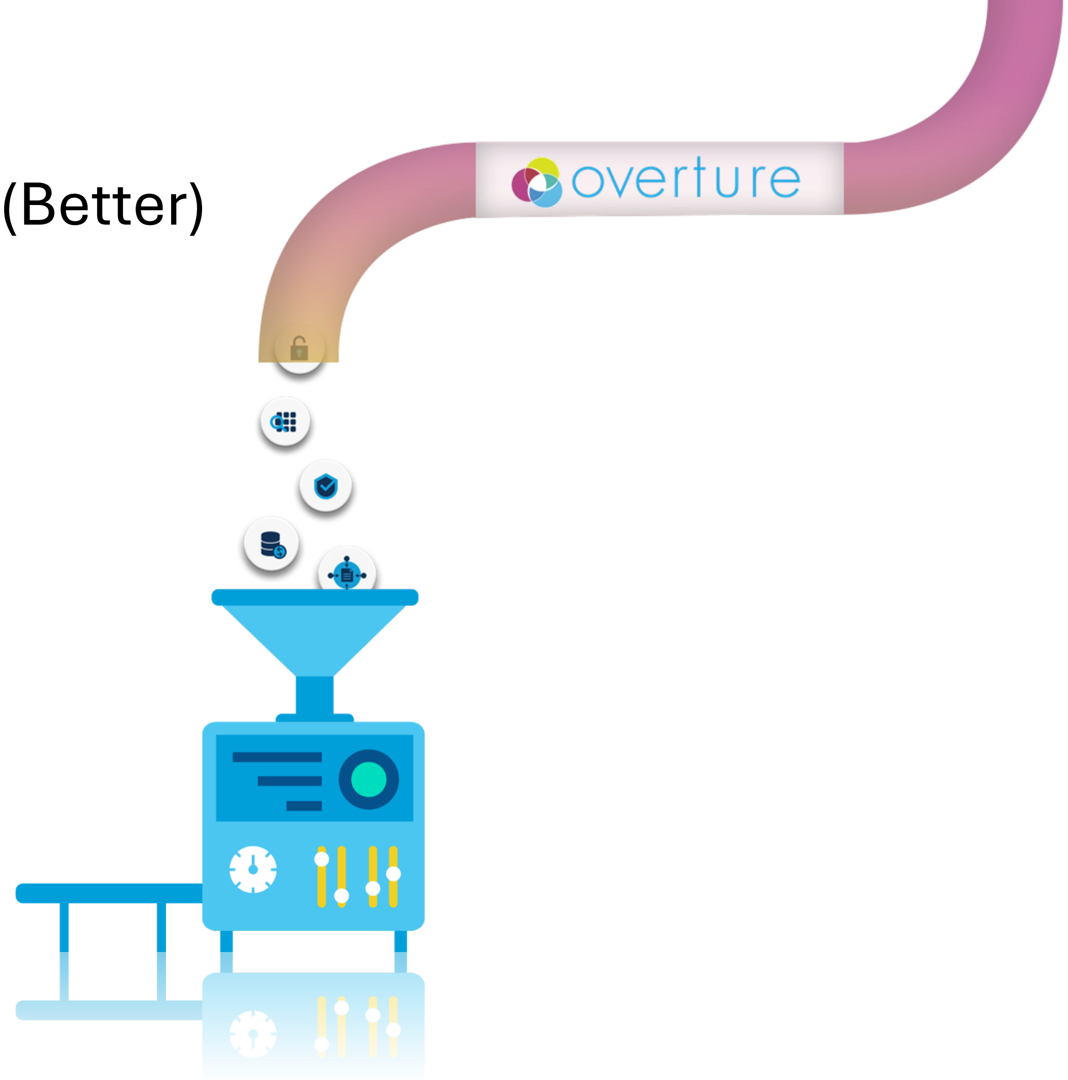
□ Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio			
□ TCGA-A5-A0G2	TCGA-UCEC	Corpus uteri	Female	64	6	4	14	8	3	11	17	1,927	558	(3)
□ TCGA-IB-7651	TCGA-PAAD	Pancreas	Female	66	6	4	14	8	3	9	17	1,119	475	(3)
□ TCGA-EO-A22U	TCGA-UCEC	Corpus uteri	Female	67	6	4	14	8	3	11	16	1,026	455	(2)
□ TCGA-FI-A2D5	TCGA-UCEC	Corpus uteri	Female	61	4	4	14	8	3	12	16	969	448	(2)
□ TCGA-EO-A22R	TCGA-UCEC	Corpus uteri	Female	65	4	4	14	8	6	11	17	930	443	(3)
□ TCGA-B5-A3FC	TCGA-UCEC	Corpus uteri	Female	67	6	4	14	8	3	11	16	969	443	(2)
□ TCGA-06-5416	TCGA-GBM	Brain	Female	61	5	2	14	8	3	9	15	910	425	(3)
□ TCGA-AX-A2HC	TCGA-UCEC	Corpus uteri	Female	76	8	8	14	8	6	11	16	944	420	(2)
□ TCGA-2W-A8YY	TCGA-CESC	Cervix uteri	Female	66	6	4	14	8	3	10	16	875	418	(2)
□ TCGA-AP-A0LM	TCGA-UCEC	Corpus uteri	Female	62	4	4	14	8	3	11	17	896	417	(3)

Building Data Portals (Better)



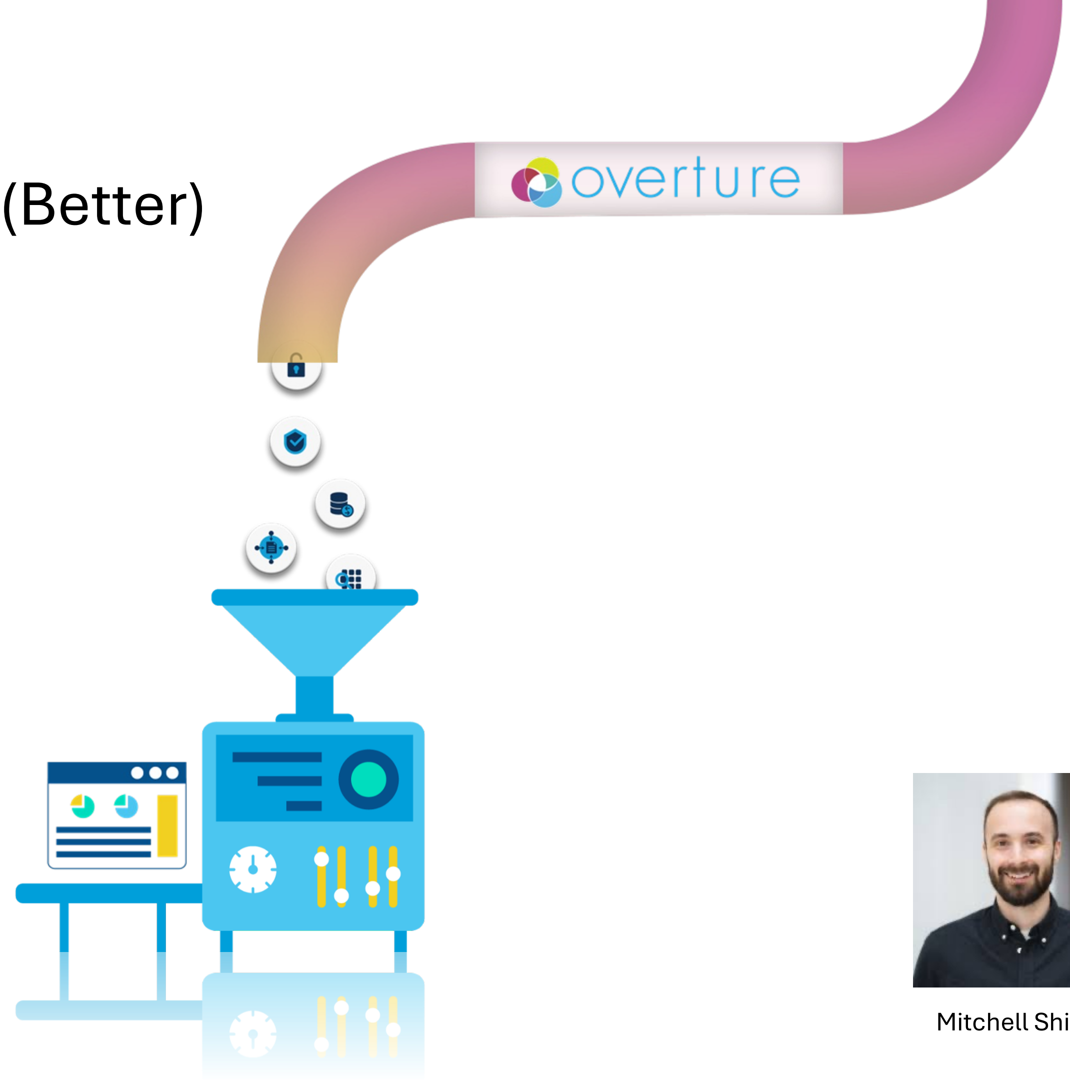
Building Data Portals (Better)

- Modular components with narrow, well-defined scope



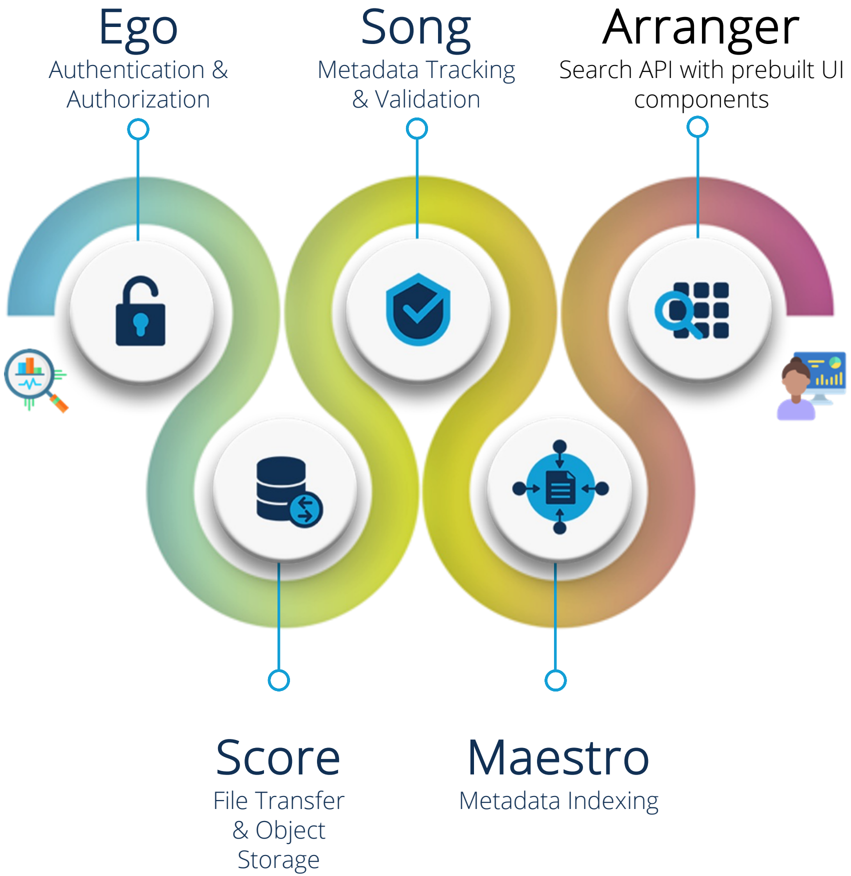
Building Data Portals (Better)

- Modular components with narrow, well-defined scope
- Enabling us to construct reliable systems quickly
- Providing time for new features & components that improve our systems further



Mitchell Shiell

This creates an ecosystem of independently reusable components



Ego
Authentication & Authorization

Song
Metadata Tracking & Validation

Arranger
Search API with prebuilt UI components

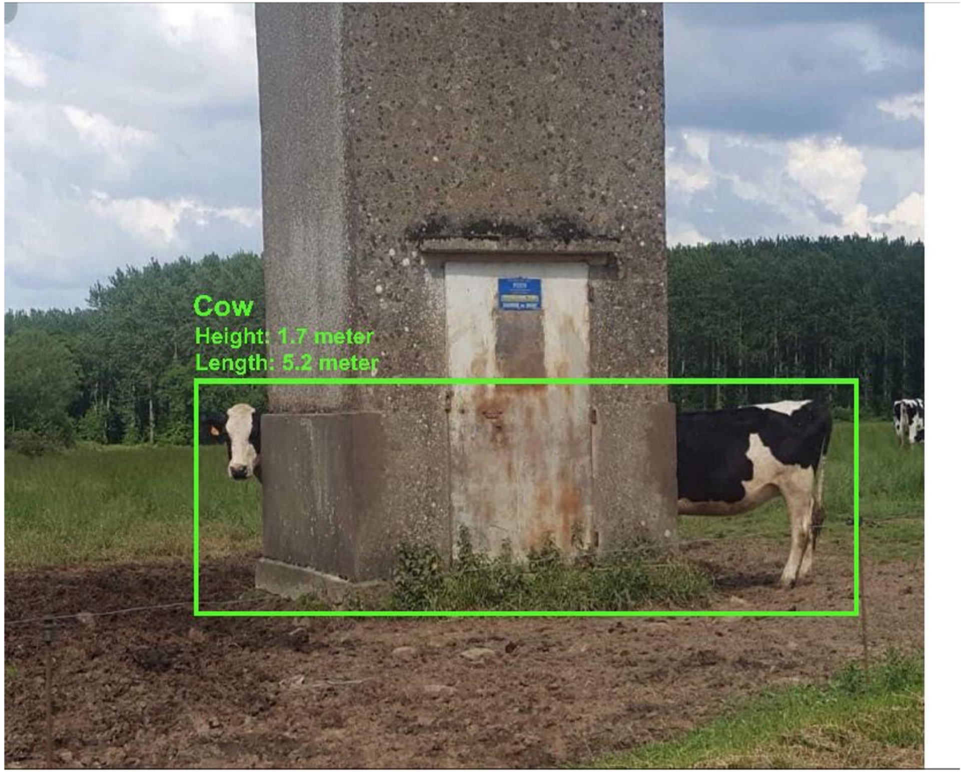
Score
File Transfer & Object Storage

Maestro
Metadata Indexing

Diversity

FIND HARMONIZED DATA

We need shared standards to represent and understand the world consistently

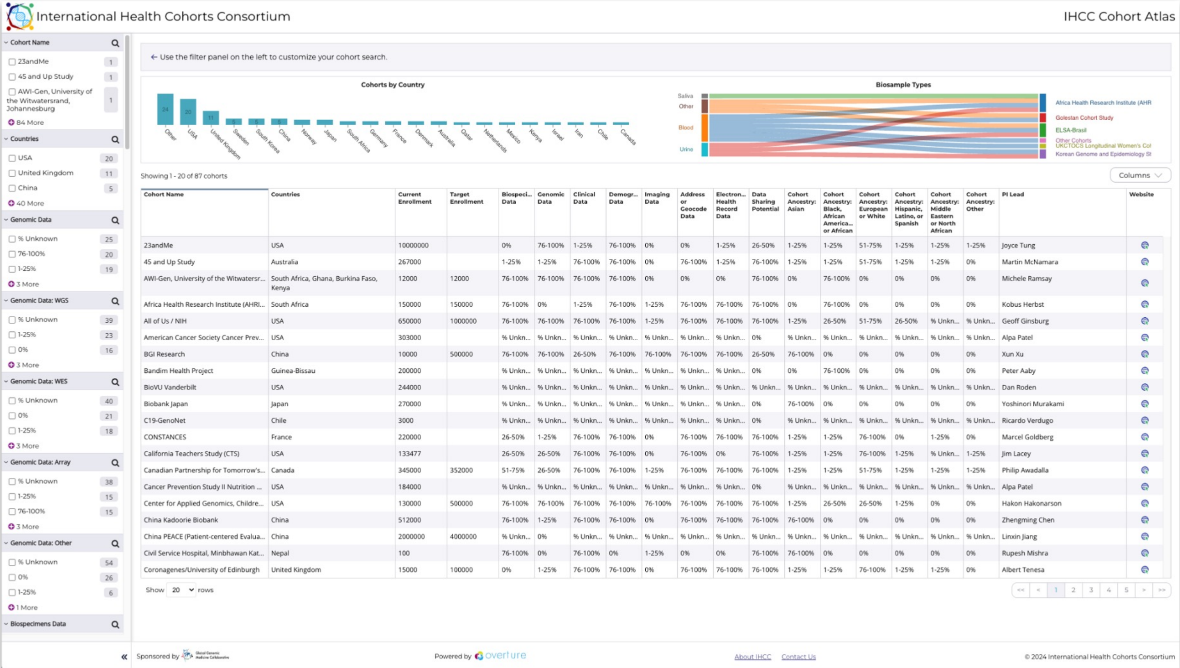


“Machine learning detects longest cow in the world”

FIND DATA: IHCC Cohort Atlas Browser

Philip Awadalla

Thomas Keane



International Health Cohorts Consortium IHCC Cohort Atlas

- Cohort Name
  - □ 23andMe 1
  - □ 45 and Up Study 1
  - □ AWI-Gen, University of the Witwatersrand, Johannesburg 1
  - 84 More
- Countries
  - □ USA 20
  - □ United Kingdom 11
  - □ China 5
  - 40 More
- Genomic Data
  - □ % Unknown 25
  - □ 76-100% 20
  - □ 1-25% 19
  - 3 More
- Genomic Data: WGS
  - □ % Unknown 39
  - □ 1-25% 23
  - □ 0% 16
  - 3 More
- Genomic Data: WES
  - □ % Unknown 40
  - □ 0% 21
  - □ 1-25% 18
  - 3 More
- Genomic Data: Array
  - □ % Unknown 38
  - □ 1-25% 15
  - □ 76-100% 15
  - 3 More
- Genomic Data: Other
  - □ % Unknown 54
  - □ 0% 26
  - □ 1-25% 6
  - 1 More
- Biospecimens Data

← Use the filter panel on the left to customize your cohort search.

Showing 1 - 20 of 87 cohorts

Columns

Cohort Name	Countries	Current Enrollment	Target Enrollment	Biospeci... Data	Genomic Data	Clinical Data	Demogr... Data	Imaging Data	Address or Geocode Data	Electron... Health Record Data	Data Sharing Potential	Cohort Ancestry: Asian	Cohort Ancestry: Black, African America... or African	Cohort Ancestry: European or White	Cohort Ancestry: Hispanic, Latino, or Spanish	Cohort Ancestry: Middle Eastern or North African	Cohort Ancestry: Other	PI Lead	Website
23andMe	USA	10000000		0%	76-100%	1-25%	76-100%	0%	0%	1-25%	26-50%	1-25%	1-25%	51-75%	1-25%	1-25%	1-25%	Joyce Tung	🌐
45 and Up Study	Australia	267000		1-25%	1-25%	76-100%	76-100%	0%	76-100%	1-25%	76-100%	1-25%	1-25%	51-75%	1-25%	1-25%	0%	Martin McNamara	🌐
AWI-Gen, University of the Witwatersr...	South Africa, Ghana, Burkina Faso, Kenya	12000	12000	76-100%	76-100%	76-100%	76-100%	0%	0%	0%	76-100%	0%	76-100%	0%	0%	0%	0%	Michele Ramsay	🌐
Africa Health Research Institute (AHRI...	South Africa	150000	150000	76-100%	0%	1-25%	76-100%	1-25%	76-100%	76-100%	76-100%	0%	76-100%	0%	0%	0%	0%	Kobus Herbst	🌐
All of Us / NIH	USA	650000	1000000	76-100%	76-100%	76-100%	76-100%	1-25%	76-100%	76-100%	76-100%	1-25%	26-50%	51-75%	26-50%	% Unkn...	% Unkn...	Geoff Ginsburg	🌐
American Cancer Society Cancer Prev...	USA	303000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	0%	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	Alpa Patel	🌐
BGI Research	China	10000	500000	76-100%	76-100%	26-50%	76-100%	76-100%	76-100%	76-100%	26-50%	76-100%	0%	0%	0%	0%	0%	Xun Xu	🌐
Bandim Health Project	Guinea-Bissau	200000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	0%	0%	76-100%	0%	0%	0%	0%	Peter Aaby	🌐
BioVU Vanderbilt	USA	244000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	Dan Roden	🌐
Biobank Japan	Japan	270000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	0%	76-100%	0%	0%	0%	0%	0%	Yoshinori Murakami	🌐
C19-GenoNet	Chile	3000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	0%	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	Ricardo Verdugo	🌐
CONSTANCES	France	220000		26-50%	1-25%	76-100%	76-100%	0%	76-100%	76-100%	76-100%	1-25%	1-25%	76-100%	0%	1-25%	0%	Marcel Goldberg	🌐
California Teachers Study (CTS)	USA	133477		26-50%	26-50%	76-100%	76-100%	0%	76-100%	0%	76-100%	1-25%	1-25%	76-100%	1-25%	% Unkn...	1-25%	Jim Lacey	🌐
Canadian Partnership for Tomorrow's...	Canada	345000	352000	51-75%	26-50%	76-100%	76-100%	1-25%	76-100%	76-100%	76-100%	1-25%	1-25%	51-75%	1-25%	1-25%	1-25%	Philip Awadalla	🌐
Cancer Prevention Study II Nutrition ...	USA	184000		% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	0%	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	Alpa Patel	🌐
Center for Applied Genomics, Childre...	USA	130000	500000	76-100%	76-100%	76-100%	76-100%	76-100%	76-100%	76-100%	76-100%	1-25%	26-50%	26-50%	1-25%	0%	0%	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512000		76-100%	1-25%	76-100%	76-100%	0%	76-100%	76-100%	76-100%	76-100%	0%	0%	0%	0%	0%	Zhengming Chen	🌐
China PEACE (Patient-centered Evalua...	China	2000000	4000000	% Unkn...	0%	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	% Unkn...	Linxin Jiang	🌐
Civil Service Hospital, Minbhawan Kat...	Nepal	100		76-100%	0%	76-100%	0%	1-25%	0%	0%	76-100%	76-100%	0%	0%	0%	0%	0%	Rupesh Mishra	🌐
Coronagenes/University of Edinburgh	United Kingdom	15000	100000	0%	1-25%	76-100%	76-100%	0%	76-100%	76-100%	76-100%	1-25%	1-25%	76-100%	1-25%	1-25%	0%	Albert Tenesa	🌐

Show 20 rows

Sponsored by Global Genomic Medicine Collaborative Powered by overture About IHCC Contact Us © 2024 International Health Cohorts Consortium

NLP-based data harmonization

Global Genomic Medicine Collaborative

EMBL-EBI

International HundredK+ Cohorts Consortium

OICR Ontario Institute for Cancer Research

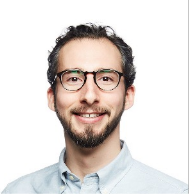
https://ihccglobal.org/

FIND DATA: Computable cohorts

Monica Munoz-Torres

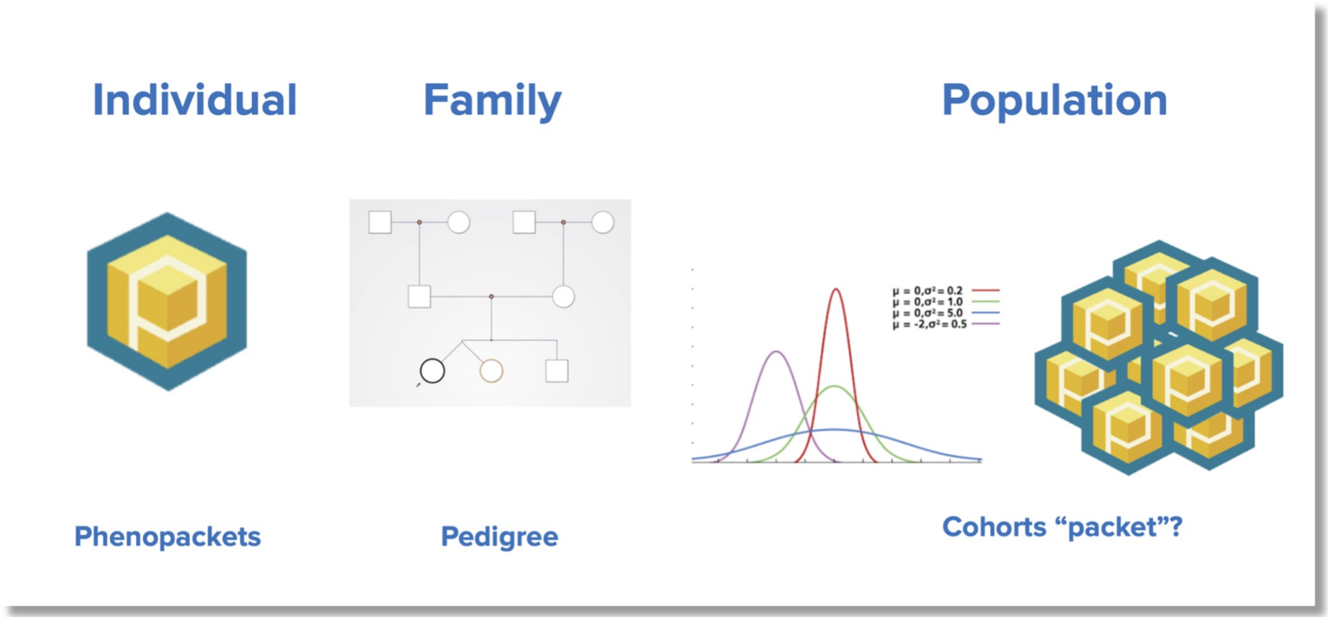
Ian Fore

Francis Jeanson



Orion Buske

Grant Wood



Individual

Family

Population

Phenopackets

Pedigree

Cohorts “packet”?

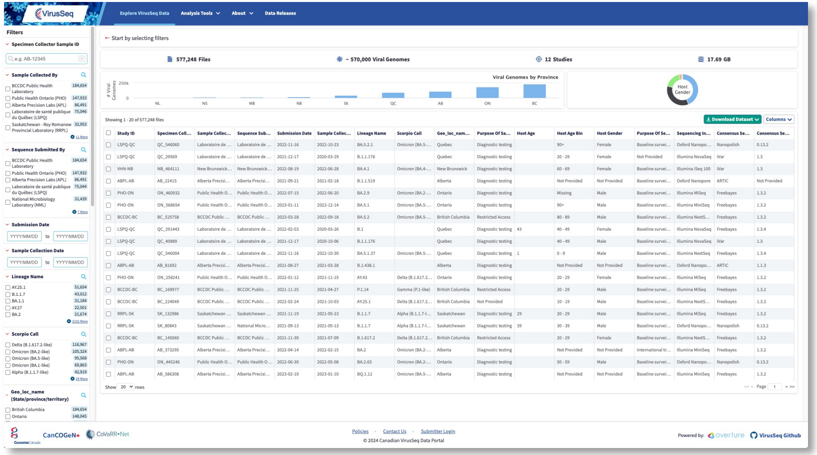
Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

Pandemics

OPEN-SOURCE RAPID RESPONSE

We need reusable, modular tools that can be easily deployed globally



Filters

Specimen Collector Sample ID

e.g. AB-12345

Sample Collected By

□	BCCDC Public Health Laboratory	184,654
□	Public Health Ontario (PHO)	147,933
□	Alberta Precision Labs (APL)	86,491
□	Laboratoire de santé publique du Québec (LSPQ)	75,046
□	Saskatchewan - Roy Romanow Provincial Laboratory (RRPL)	32,953

11 More

Sequence Submitted By

□	BCCDC Public Health Laboratory	184,654
□	Public Health Ontario (PHO)	147,932
□	Alberta Precision Labs (APL)	86,491
□	Laboratoire de santé publique du Québec (LSPQ)	75,044
□	National Microbiology Laboratory (NML)	31,439

7 More

Submission Date

YYYY/MM/DD to YYYY/MM/DD

Sample Collection Date

YYYY/MM/DD to YYYY/MM/DD

Lineage Name

□	AY.25.1	51,654
□	B.1.1.7	43,012
□	BA.1.1	31,184
□	AY.27	22,501
□	BA.2	21,674

2232 More

Scorpio Call

□	Delta (B.1.617.2-like)	116,967
□	Omicron (BA.2-like)	105,324
□	Omicron (BA.5-like)	95,568
□	Omicron (BA.1-like)	69,863
□	Alpha (B.1.1.7-like)	42,919

29 More

Geo_loc_name (State/province/territory)

□	British Columbia	184,654
□	Ontario	148,045

← Start by selecting filters

577,248 Files · ~ 570,000 Viral Genomes · 12 Studies · 17.69 GB

Showing 1 - 20 of 577,248 files

Download Dataset · Columns

□	Study ID	Specimen Coll...	Sample Collec...	Sequence Sub...	Submission Date	Sample Collec...	Lineage Name	Scorpio Call	Geo_loc_nam...	Purpose Of Sa...	Host Age	Host Age Bin	Host Gender	Purpose Of Se...	Sequencing In...	Consensus Se...	Consensus Se...
□	LSPQ-QC	QC_546060	Laboratoire de ...	Laboratoire de ...	2022-11-16	2022-10-23	BA.5.2.1	Omicron (BA.5-...	Quebec	Diagnostic testing		90+	Female	Baseline survei...	Oxford Nanopo...	Nanopolish	0.13.2
□	LSPQ-QC	QC_29569	Laboratoire de ...	Laboratoire de ...	2021-12-17	2020-03-29	B.1.1.176		Quebec	Diagnostic testing		20 - 29	Female	Not Provided	Illumina NovaSeq	iVar	1.3
□	VHN-NB	NB_464111	New Brunswick...	New Brunswick...	2022-08-19	2022-06-28	BA.4.1	Omicron (BA.4-...	New Brunswick	Diagnostic testing		60 - 69	Female	Baseline survei...	Illumina iSeq 100	iVar	1.3
□	ABPL-AB	AB_22415	Alberta Precisi...	Alberta Precisi...	2021-09-21	2021-02-18	B.1.1.519		Alberta	Diagnostic testing		Not Provided	Not Provided	Baseline survei...	Oxford Nanopore	ARTIC	Not Provided
□	PHO-ON	ON_460932	Public Health O...	Public Health O...	2022-07-15	2022-06-20	BA.2.9	Omicron (BA.2-...	Ontario	Diagnostic testing		Missing	Male	Baseline survei...	Illumina MiSeq	Freebayes	1.3.2
□	PHO-ON	ON_568654	Public Health O...	Public Health O...	2023-01-11	2022-12-14	BA.5.1	Omicron (BA.5-...	Ontario	Diagnostic testing		90+	Male	Baseline survei...	Illumina MiniSeq	Freebayes	1.3.2
□	BCCDC-BC	BC_525758	BCCDC Public ...	BCCDC Public ...	2023-03-28	2022-09-18	BA.5.2	Omicron (BA.5-...	British Columbia	Restricted Access		80 - 89	Male	Baseline survei...	Illumina NextS...	Freebayes	1.3.2
□	LSPQ-QC	QC_291443	Laboratoire de ...	Laboratoire de ...	2022-02-03	2020-03-26	B.1		Quebec	Diagnostic testing	43	40 - 49	Female	Baseline survei...	Illumina NovaSeq	Freebayes	1.3.4
□	LSPQ-QC	QC_40889	Laboratoire de ...	Laboratoire de ...	2021-12-17	2020-10-06	B.1.1.176		Quebec	Diagnostic testing		40 - 49	Male	Baseline survei...	Illumina NovaSeq	iVar	1.3
□	LSPQ-QC	QC_546004	Laboratoire de ...	Laboratoire de ...	2022-11-16	2022-10-30	BA.5.1.37	Omicron (BA.5-...	Quebec	Diagnostic testing	1	0 - 9	Male	Baseline survei...	Illumina NextSeq	Freebayes	1.3.4
□	ABPL-AB	AB_81692	Alberta Precisi...	Alberta Precisi...	2021-08-27	2021-03-28	B.1.438.1		Alberta	Diagnostic testing		Not Provided	Not Provided	Baseline survei...	Oxford Nanopo...	ARTIC	1.1.3
□	PHO-ON	ON_258241	Public Health O...	Public Health O...	2022-01-12	2021-11-15	AY.43	Delta (B.1.617.2...	Ontario	Diagnostic testing		20 - 29	Female	Baseline survei...	Illumina MiSeq	Freebayes	1.3.2
□	BCCDC-BC	BC_169977	BCCDC Public ...	BCCDC Public ...	2021-11-25	2021-04-27	P.1.14	Gamma (P.1-like)	British Columbia	Restricted Access		20 - 29	Male	Baseline survei...	Illumina MiSeq	Freebayes	1.3.2
□	BCCDC-BC	BC_224049	BCCDC Public ...	BCCDC Public ...	2022-02-24	2021-10-03	AY.25.1	Delta (B.1.617.2...	British Columbia	Not Provided		10 - 19	Male	Baseline survei...	Illumina NextS...	Freebayes	1.3.2
□	RRPL-SK	SK_132986	Saskatchewan ...	Saskatchewan ...	2021-11-19	2021-05-23	B.1.1.7	Alpha (B.1.1.7-l...	Saskatchewan	Diagnostic testing	29	20 - 29	Male	Baseline survei...	Illumina MiSeq	Freebayes	1.3.2
□	RRPL-SK	SK_80843	Saskatchewan ...	National Micro...	2021-09-13	2021-05-13	B.1.1.7	Alpha (B.1.1.7-l...	Saskatchewan	Diagnostic testing	39	30 - 39	Male	Baseline survei...	Oxford Nanopo...	Nanopolish	0.13.2
□	BCCDC-BC	BC_145060	BCCDC Public ...	BCCDC Public ...	2021-11-30	2021-07-09	B.1.617.2	Delta (B.1.617.2...	British Columbia	Restricted Access		20 - 29	Female	Baseline survei...	Illumina NextS...	Freebayes	1.3.2
□	ABPL-AB	AB_373295	Alberta Precisi...	Alberta Precisi...	2022-04-14	2022-02-15	BA.2	Omicron (BA.2-...	Alberta	Diagnostic testing		Not Provided	Not Provided	International tr...	Illumina MiniSeq	Freebayes	1.3.2
□	PHO-ON	ON_445246	Public Health O...	Public Health O...	2022-06-30	2022-05-08	BA.2.65	Omicron (BA.2-...	Ontario	Diagnostic testing		50 - 59	Male	Baseline survei...	Oxford Nanopo...	Nanopolish	0.13.2
□	ABPL-AB	AB_586308	Alberta Precisi...	Alberta Precisi...	2023-02-10	2023-01-10	BQ.1.12	Omicron (BA.5-...	Alberta	Diagnostic testing		Not Provided	Not Provided	Baseline survei...	Illumina MiniSeq	Freebayes	1.3.2

Show 20 rows

<< < Page 1 > >>

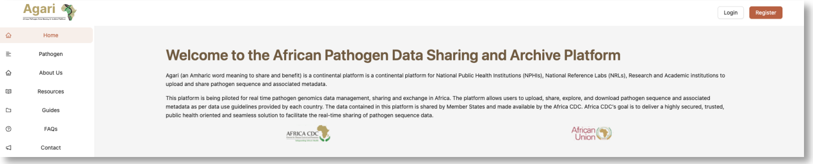
Policies · Contact Us · Submitter Login

© 2024 Canadian VirusSeq Data Portal

Powered by: overture · VirusSeq Github

Justin Richardson

https://virusseq-dataportal.ca/explorer



Welcome to the African Pathogen Data Sharing and Archive Platform

Agari (an Amharic word meaning to share and benefit) is a continental platform is a continental platform for National Public Health Institutions (NPHIs), National Reference Labs (NRLs), Research and Academic institutions to upload and share pathogen sequence and associated metadata.

This platform is being piloted for real time pathogen genomics data management, sharing and exchange in Africa. The platform allows users to upload, share, explore, and download pathogen sequence and associated metadata as per data use guidelines provided by each country. The data contained in this platform is shared by Member States and made available by the Africa CDC. Africa CDC's goal is to deliver a highly secured, trusted, public health oriented and seamless solution to facilitate the real-time sharing of pathogen sequence data.

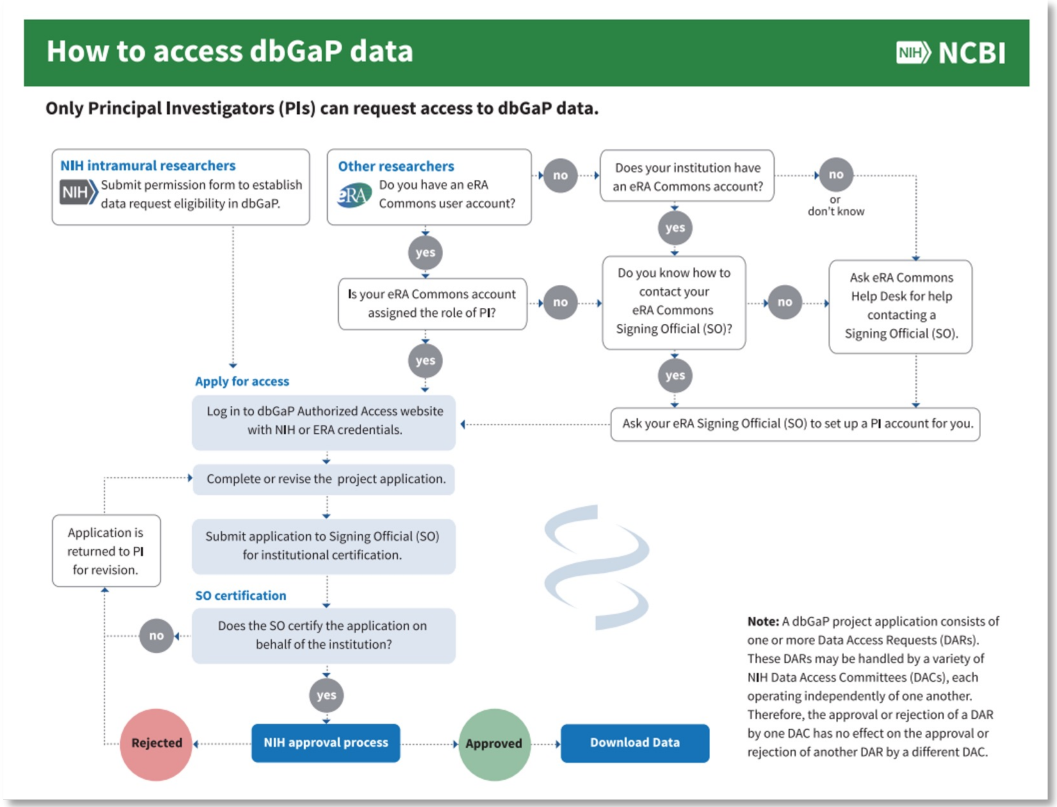
https://apaportal.sanbi.ac.za/

Alan Christoffels

Clinical data

ACCESS DATA

We need standards and workflows to enable easy and fast access to data for researchers



How to access dbGaP data

NIH NCBI

Only Principal Investigators (PIs) can request access to dbGaP data.

NIH intramural researchers
Submit permission form to establish data request eligibility in dbGaP.

Other researchers
Do you have an eRA Commons user account?

Does your institution have an eRA Commons account?

Is your eRA Commons account assigned the role of PI?

Do you know how to contact your eRA Commons Signing Official (SO)?

Ask eRA Commons Help Desk for help contacting a Signing Official (SO).

Ask your eRA Signing Official (SO) to set up a PI account for you.

Apply for access

Log in to dbGaP Authorized Access website with NIH or ERA credentials.

Complete or revise the project application.

Submit application to Signing Official (SO) for institutional certification.

Application is returned to PI for revision.

SO certification

Does the SO certify the application on behalf of the institution?

Rejected

NIH approval process

Approved

Download Data

Note: A dbGaP project application consists of one or more Data Access Requests (DARs). These DARs may be handled by a variety of NIH Data Access Committees (DACs), each operating independently of one another. Therefore, the approval or rejection of a DAR by one DAC has no effect on the approval or rejection of another DAR by a different DAC.

Data Use Ontology (DUO)

Data Use Ontology

Global Alliance for Genomics & Health

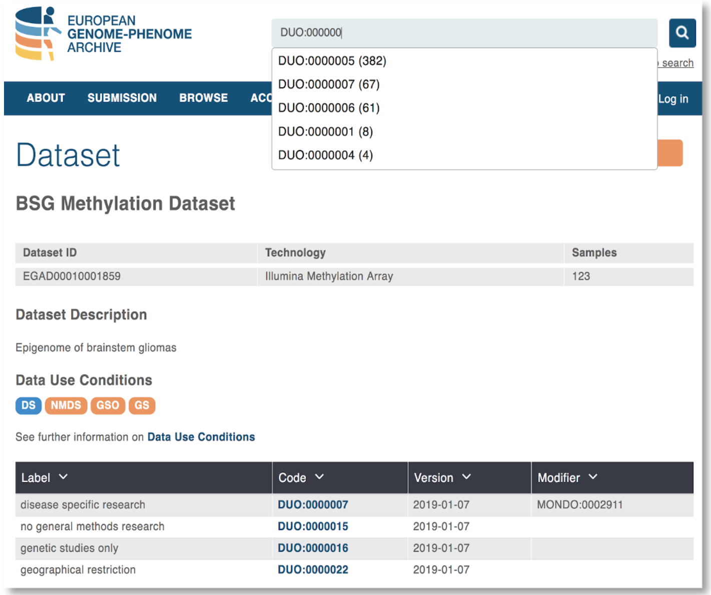
Collaborate. Innovate. Accelerate.

Moran Cabili

Jonathan Lawson

Vocabulary describing permitted data uses and modifiers

- General research use
- disease-specific research
- not for profit only
- …



DUO:0000005 (382)
DUO:0000007 (67)
DUO:0000006 (61)
DUO:0000001 (8)
DUO:0000004 (4)

Dataset

BSG Methylation Dataset

Dataset ID	Technology	Samples
EGAD00010001859	Illumina Methylation Array	123

Dataset Description

Epigenome of brainstem gliomas

Data Use Conditions

DS NMDS GSO GS

See further information on **Data Use Conditions**

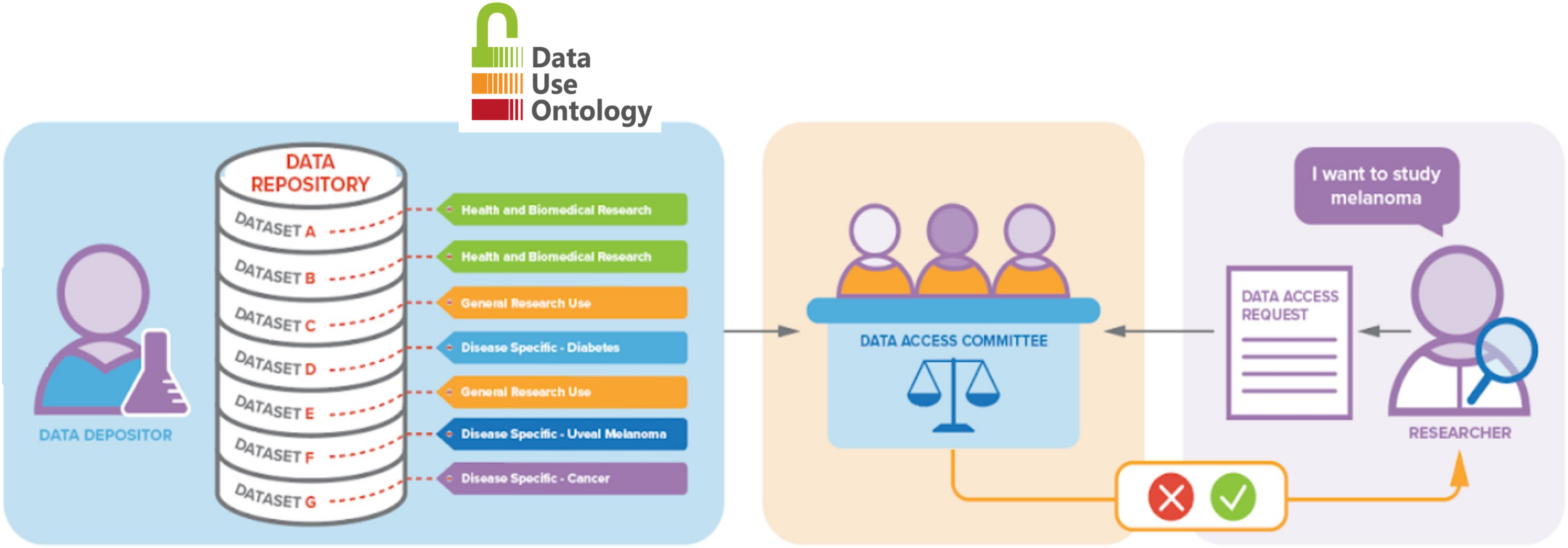
Label	Code	Version	Modifier
disease specific research	DUO:0000007	2019-01-07	MONDO:0002911
no general methods research	DUO:0000015	2019-01-07	
genetic studies only	DUO:0000016	2019-01-07	
geographical restriction	DUO:0000022	2019-01-07	

https://ega-archive.org/datasets/EGAD00010001859

https://www.ebi.ac.uk/ols/ontologies/duo

https://github.com/EBISPOT/DUO

Collaborate. Translate. Change lives.



Data Use Ontology

DATA DEPOSITOR

DATA REPOSITORY

DATASET A – Health and Biomedical Research

DATASET B – Health and Biomedical Research

DATASET C – General Research Use

DATASET D – Disease Specific - Diabetes

DATASET E – General Research Use

DATASET F – Disease Specific - Uveal Melanoma

DATASET G – Disease Specific - Cancer

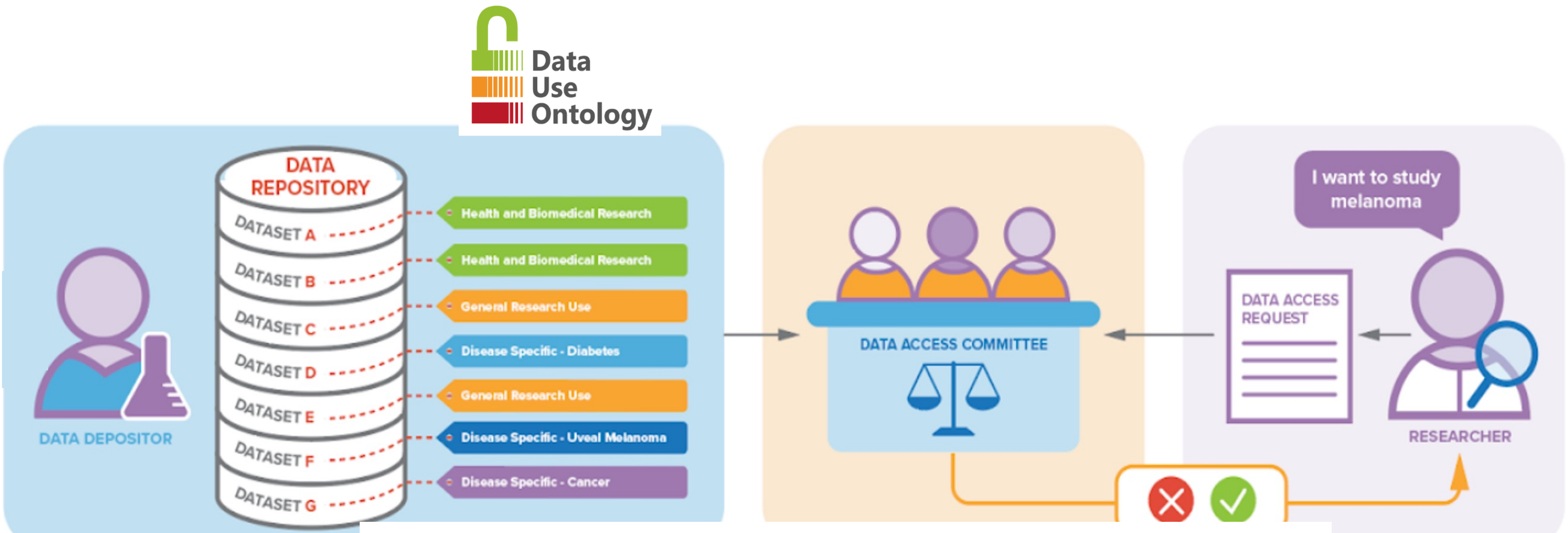
DATA ACCESS COMMITTEE

I want to study melanoma

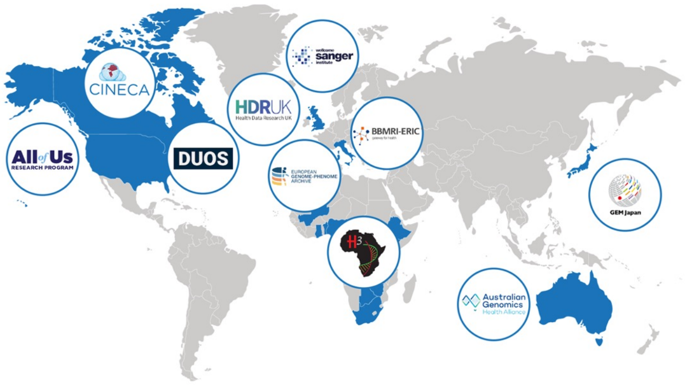
DATA ACCESS REQUEST

RESEARCHER

Lawson et al, https://bit.ly/duo-paper



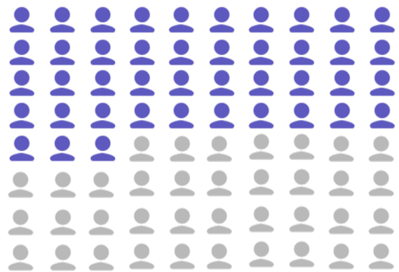
>200,000 datasets annotated with the Data Use Ontology

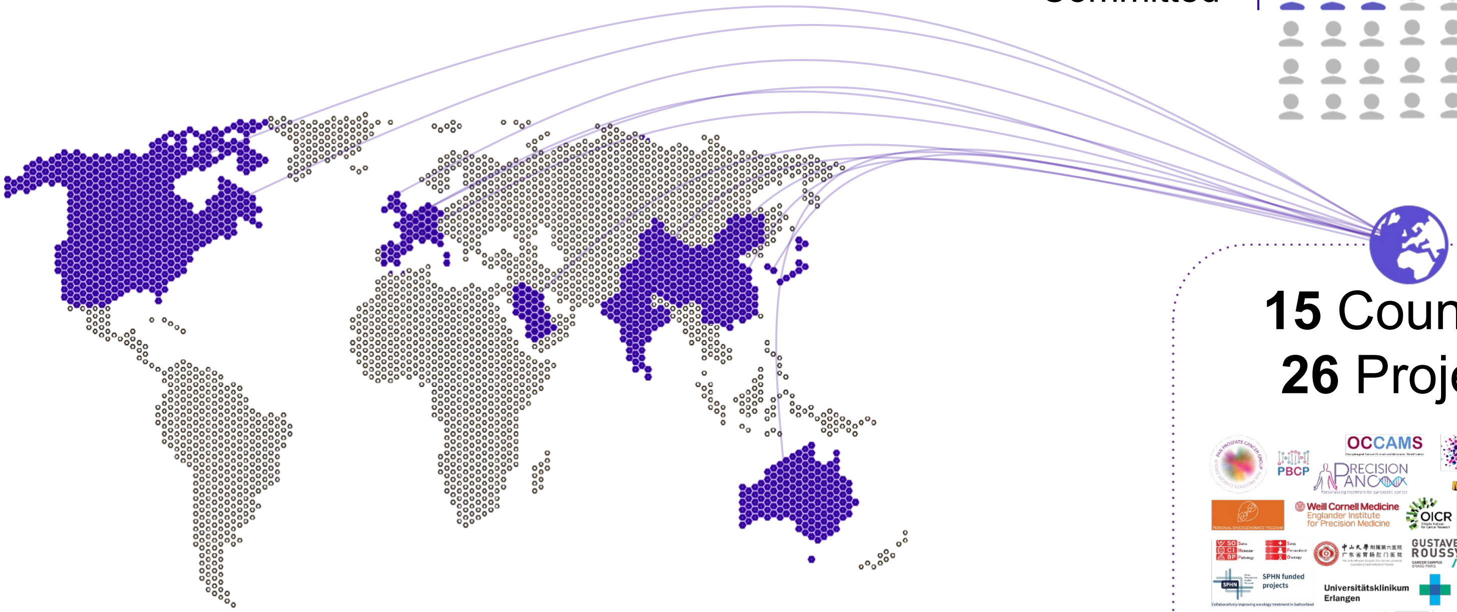

wellcome sanger institute
CINECA
HDRUK Health Data Research UK
BBMRI-ERIC
All of Us RESEARCH PROGRAM
DUOS
EUROPEAN GENOME-PHENOME ARCHIVE
GEM Japan
H3
Australian Genomics Health Alliance

Lawson et al, https://bit.ly/duo-paper

International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO)

Lincoln Stein



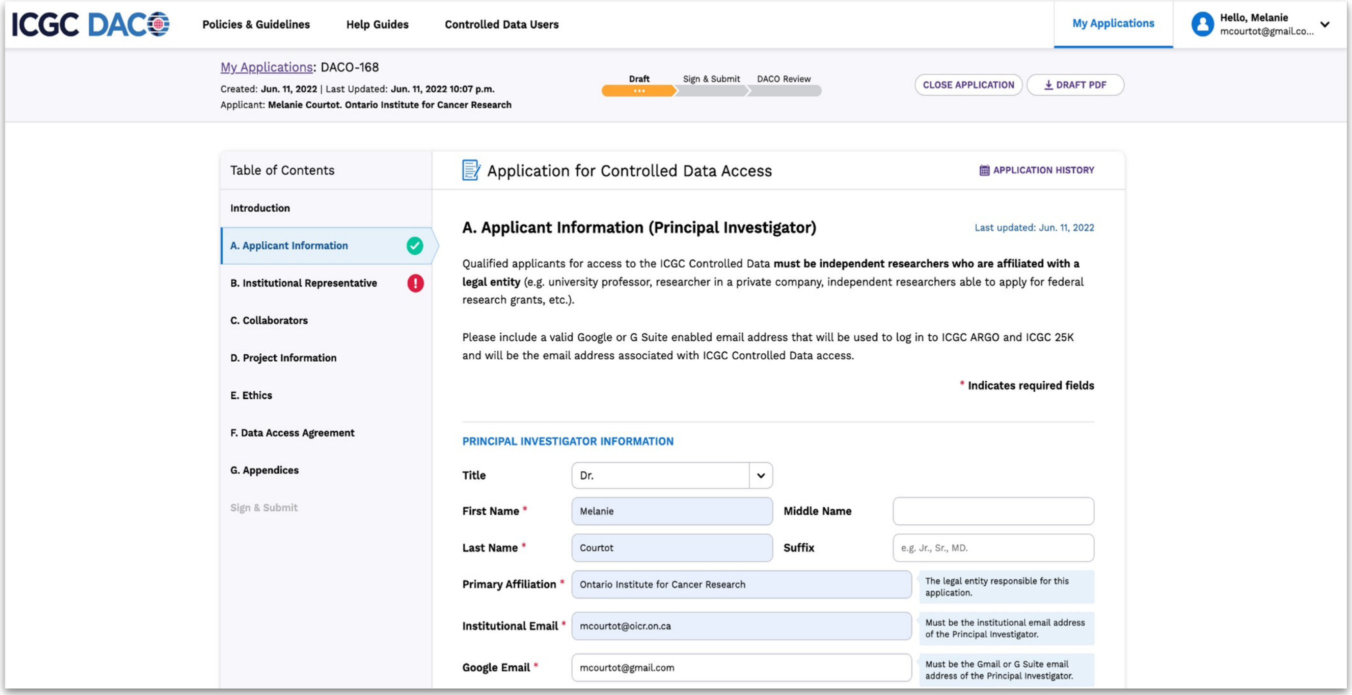


15 Countries
26 Projects

https://www.icgc-argo.org/

ACCESS CLINICAL DATA: ICGC-ARGO data access module

Ann Catton



My Applications: DACO-168

Created: **Jun. 11, 2022** | Last Updated: **Jun. 11, 2022 10:07 p.m.**

Applicant: **Melanie Courtot. Ontario Institute for Cancer Research**

Application for Controlled Data Access

A. Applicant Information (Principal Investigator)

Last updated: Jun. 11, 2022

Qualified applicants for access to the ICGC Controlled Data **must be independent researchers who are affiliated with a legal entity** (e.g. university professor, researcher in a private company, independent researchers able to apply for federal research grants, etc.).

Please include a valid Google or G Suite enabled email address that will be used to log in to ICGC ARGO and ICGC 25K and will be the email address associated with ICGC Controlled Data access.

* **Indicates required fields**

PRINCIPAL INVESTIGATOR INFORMATION

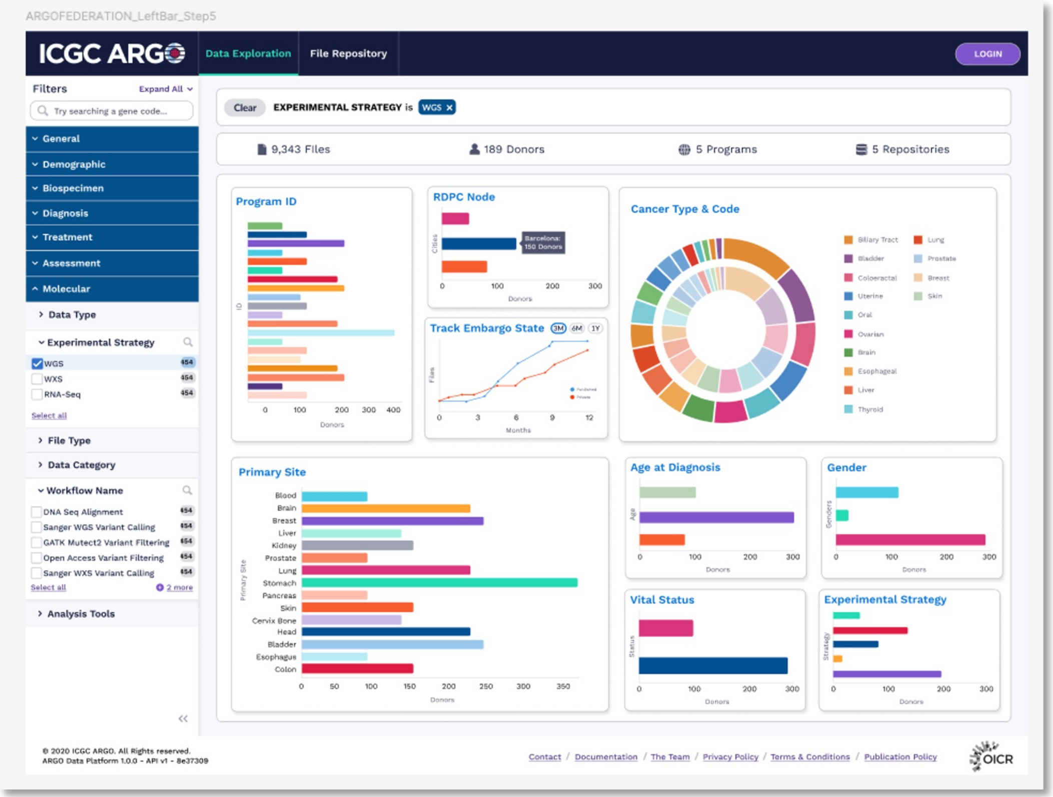
Title	Dr.		
First Name *	Melanie	**Middle Name**	
Last Name *	Courtot	**Suffix**	e.g. Jr., Sr., MD.
Primary Affiliation *	Ontario Institute for Cancer Research	The legal entity responsible for this application.	
Institutional Email *	mcourtot@oicr.on.ca	Must be the institutional email address of the Principal Investigator.	
Google Email *	mcourtot@gmail.com	Must be the Gmail or G Suite email address of the Principal Investigator.	

Data Use Ontology
Passport
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

Data sharing of 100k+ cancer participants, with comprehensive clinical and molecular data

https://www.icgc-argo.org/

ACCESS CLINICAL DATA: ICGC-ARGO federation



ARGOFEDERATION_LeftBar_Step5

ICGC ARGO | Data Exploration | File Repository | LOGIN

Filters — Expand All

Try searching a gene code...

- General
- Demographic
- Biospecimen
- Diagnosis
- Treatment
- Assessment
- Molecular
  - Data Type
  - Experimental Strategy
    - ☑ WGS 454
    - □ WXS 454
    - □ RNA-Seq 454
    - Select all
  - File Type
  - Data Category
  - Workflow Name
    - □ DNA Seq Alignment 454
    - □ Sanger WGS Variant Calling 454
    - □ GATK Mutect2 Variant Filtering 454
    - □ Open Access Variant Filtering 454
    - □ Sanger WXS Variant Calling 454
    - Select all — 2 more
  - Analysis Tools

Clear | EXPERIMENTAL STRATEGY is WGS ×

9,343 Files | 189 Donors | 5 Programs | 5 Repositories

© 2020 ICGC ARGO. All Rights reserved.
ARGO Data Platform 1.0.0 - API v1 - 8e37309

Contact / Documentation / The Team / Privacy Policy / Terms & Conditions / Publication Policy

OICR

David Torrents

Jon Eubank

International data sharing of 100k+ cancer participants, regulatory compliant

ACCESS CLINICAL DATA: Participant enrollment portal

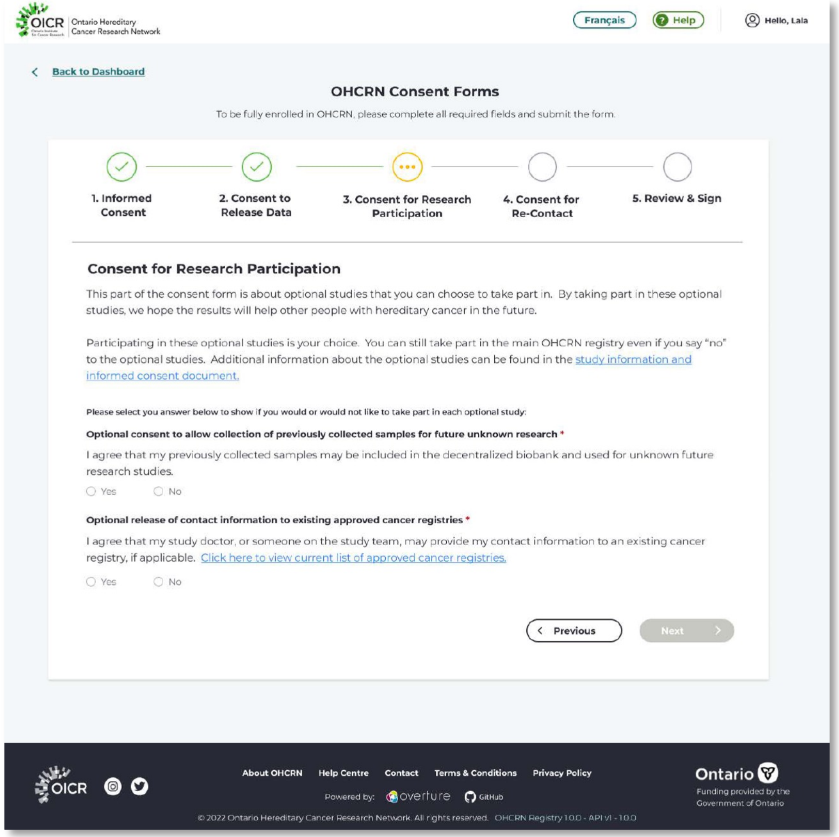
Raymond Kim

Lauren Hugues

Michelle Brazas

Brandon Chan

Rakesh Mistry



< Back to Dashboard

OHCRN Consent Forms

To be fully enrolled in OHCRN, please complete all required fields and submit the form.

1. Informed Consent ✓
2. Consent to Release Data ✓
3. Consent for Research Participation
4. Consent for Re-Contact
5. Review & Sign

Consent for Research Participation

This part of the consent form is about optional studies that you can choose to take part in. By taking part in these optional studies, we hope the results will help other people with hereditary cancer in the future.

Participating in these optional studies is your choice. You can still take part in the main OHCRN registry even if you say "no" to the optional studies. Additional information about the optional studies can be found in the study information and informed consent document.

Please select you answer below to show if you would or would not like to take part in each optional study:

Optional consent to allow collection of previously collected samples for future unknown research *

I agree that my previously collected samples may be included in the decentralized biobank and used for unknown future research studies.

○ Yes ○ No

Optional release of contact information to existing approved cancer registries *

I agree that my study doctor, or someone on the study team, may provide my contact information to an existing cancer registry, if applicable. Click here to view current list of approved cancer registries.

○ Yes ○ No

< Previous   Next >

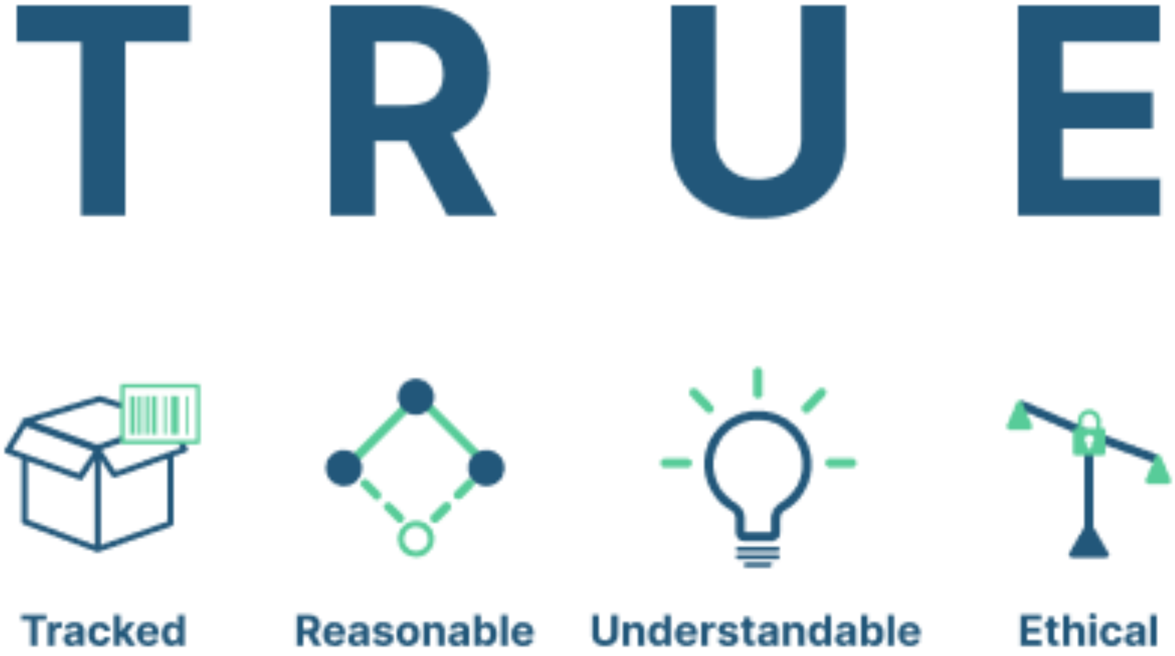
© 2022 Ontario Hereditary Cancer Research Network. All rights reserved. OHCRN Registry 1.0.0 - API v1 - 1.0.0

Ontario-wide monitoring of inherited cancer syndromes for research

Data Use Ontology
Passport
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

https://ohcrn.ca/

Beyond FAIR



TRUE

Tracked

Reasonable

Understandable

Ethical

Tracking data

Standards for provenance, evidence and attribution – eg PROV, ECO, CRediT

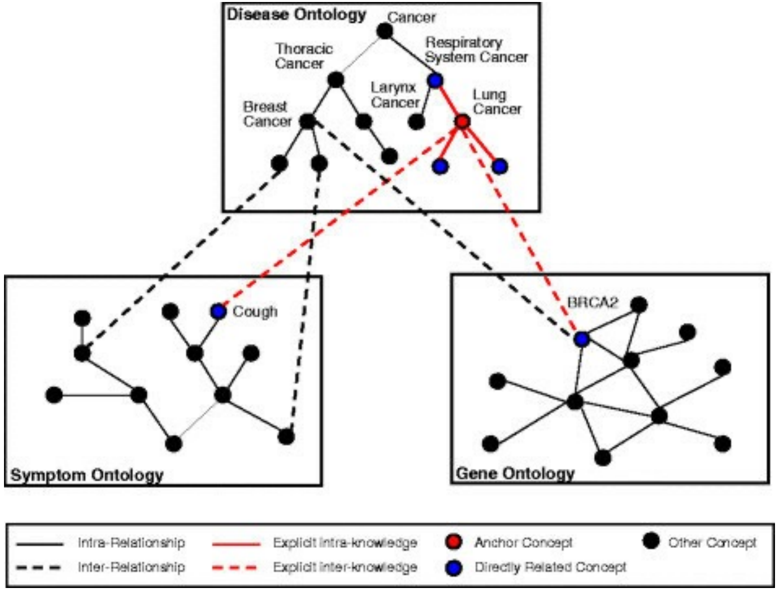
Must accompany data and be computationally manageable

Tracking data

Standards for provenance, evidence and attribution – eg PROV, ECO, CRediT

Must accompany data and be computationally manageable

Reasoning over data



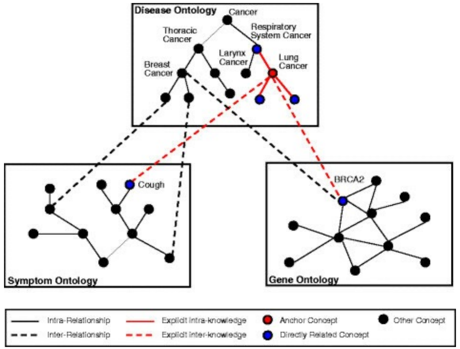
Logical inference, validation, new insights

Tracking data

Standards for provenance, evidence and attribution – eg PROV, ECO, CRediT

Must accompany data and be computationally manageable

Reasoning over data



Logical inference, validation, new insights

Understanding data

Open-source models: Llama, Mistral. Can be installed locally eg behind institutional firewall

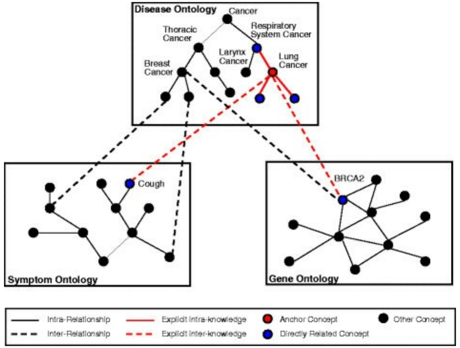
Closed source models: GPT4, Claude. Commercial support and innovation.

Tracking data

Standards for provenance, evidence and attribution – eg PROV, ECO, CRediT

Must accompany data and be computationally manageable

Reasoning over data



Logical inference, validation, new insights

Understanding data

Open-source models: Llama, Mistral. Can be installed locally e.g. behind institutional firewall

Closed source models: GPT4, Claude. Commercial support and innovation.

Ethical and equitable data

Pan-Canadian Genome Library
Bibliothèque génomique pancanadienne

Alexis Li

EDI in models

Privacy preserving

Tiny models

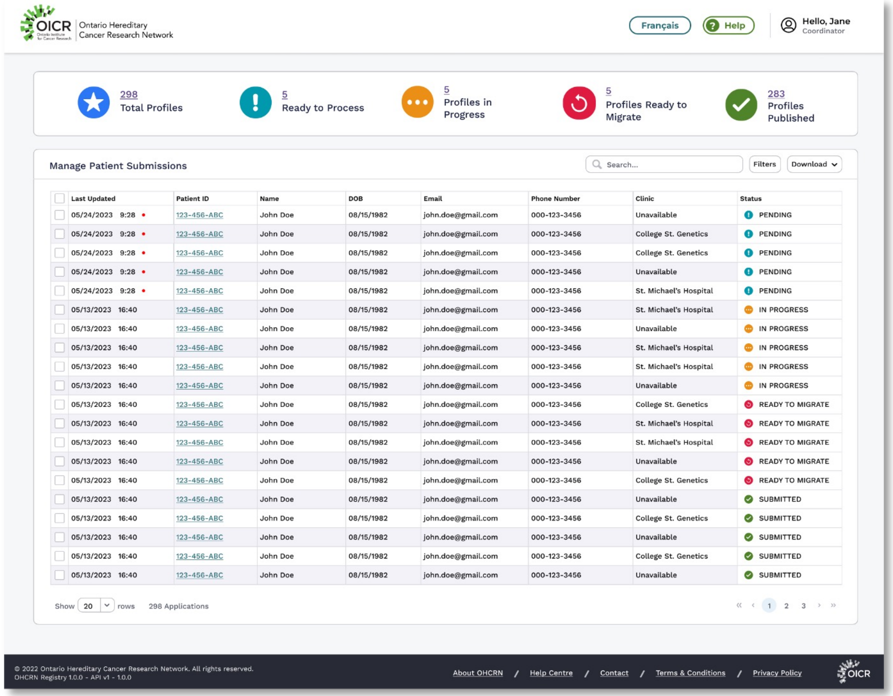
UNDERSTAND DATA USE (and challenges)

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

We need to know what our models do (and what data bias, limits, issues… it may have)

UNDERSTAND DATA USE: LLM-based data extraction

Pratham
Hemlani



298 Total Profiles

5 Ready to Process

5 Profiles in Progress

5 Profiles Ready to Migrate

283 Profiles Published

Manage Patient Submissions

□	Last Updated	Patient ID	Name	DOB	Email	Phone Number	Clinic	Status
□	05/24/2023 9:28 ●	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	PENDING
□	05/24/2023 9:28 ●	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	PENDING
□	05/24/2023 9:28 ●	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	PENDING
□	05/24/2023 9:28 ●	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	PENDING
□	05/24/2023 9:28 ●	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	PENDING
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	IN PROGRESS
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	IN PROGRESS
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	IN PROGRESS
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	IN PROGRESS
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	IN PROGRESS
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	READY TO MIGRATE
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	READY TO MIGRATE
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	St. Michael's Hospital	READY TO MIGRATE
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	READY TO MIGRATE
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	READY TO MIGRATE
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	SUBMITTED
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	SUBMITTED
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	SUBMITTED
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	College St. Genetics	SUBMITTED
□	05/13/2023 16:40	123-456-ABC	John Doe	08/15/1982	john.doe@gmail.com	000-123-3456	Unavailable	SUBMITTED

Show 20 rows 298 Applications

© 2022 Ontario Hereditary Cancer Research Network. All rights reserved.
OHCRN Registry 1.0.0 - API v1 - 1.0.0

About OHCRN / Help Centre / Contact / Terms & Conditions / Privacy Policy
OICR

OHCRN Coordinator dashboard

Manual review and integration of lab reports into the OHCRN platform.

https://ohcrn.ca/

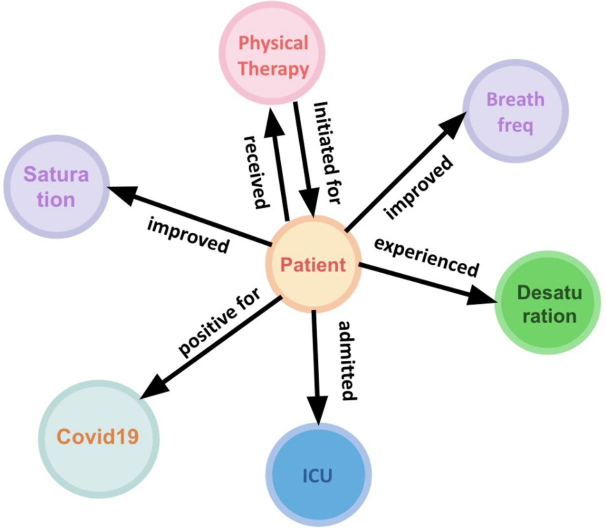
UNDERSTAND DATA USE: Extracting EHR data

Benjamin Haibe-Kains

Andres Melani de la Hoz

The patient was admitted to the ICU one week after a positive COVID-19 result due to oxygen desaturation. Physical therapy was initiated promptly after admission, which helped improve the patient's breathing frequency and oxygen saturation.

GPT - 4



neo4j aura

Ongoing opportunities

KNOWLEDGE REPRESENTATION

Semantic

Using pre-defined ontology concepts, data models, data structures, data dictionaries, and data schemes

Data models
Cohort summary representation

Ongoing opportunities

KNOWLEDGE REPRESENTATION	INFRASTRUCTURE
Semantic	Overture
Using pre-defined ontology concepts, data models, data structures, data dictionaries, and data schemes	Complete scalable and modular toolkit to rapidly deploy
Data models Cohort summary representation	Metadata harmonization module

Ongoing opportunities

KNOWLEDGE REPRESENTATION	INFRASTRUCTURE	VALIDATION AND ENRICHMENT
Semantic	Overture	Curation
Using pre-defined ontology concepts, data models, data structures, data dictionaries, and data schemes	Complete scalable and modular toolkit to rapidly deploy	Common data schemas defined for encoding, decoding, and representation
Data models Cohort summary representation	Metadata harmonization module	Graph-based validation Recommender engine LLM for curation

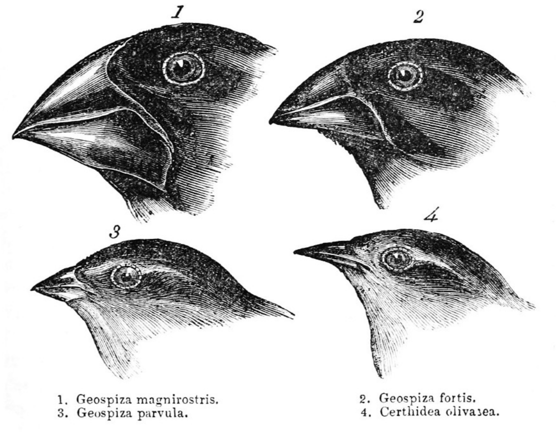
Ongoing opportunities

KNOWLEDGE REPRESENTATION	INFRASTRUCTURE	VALIDATION AND ENRICHMENT	EXCHANGE
Semantic	Overture	Curation	Structural
Using pre-defined ontology concepts, data models, data structures, data dictionaries, and data schemes	Complete scalable and modular toolkit to rapidly deploy	Common data schemas defined for encoding, decoding, and representation	Bridging research and clinical
Data models Cohort summary representation	Metadata harmonization module	Graph-based validation Recommender engine LLM for curation	LLM for EHR text-mining and Phenopackets

Summary highlights

Careers are not linear; change brings opportunity
Global challenges need global solutions
Making sense of the data is critical
Open-source toolbox to ease and increase reuse
TRUE data to support AI
Much more to do!

Biology must generate ideas as well as data



"...it would have been rather a pity if Darwin had stopped thinking after he had described the shapes and sizes of finch beaks..."

Nature 597, 305 (2021)

Thanks

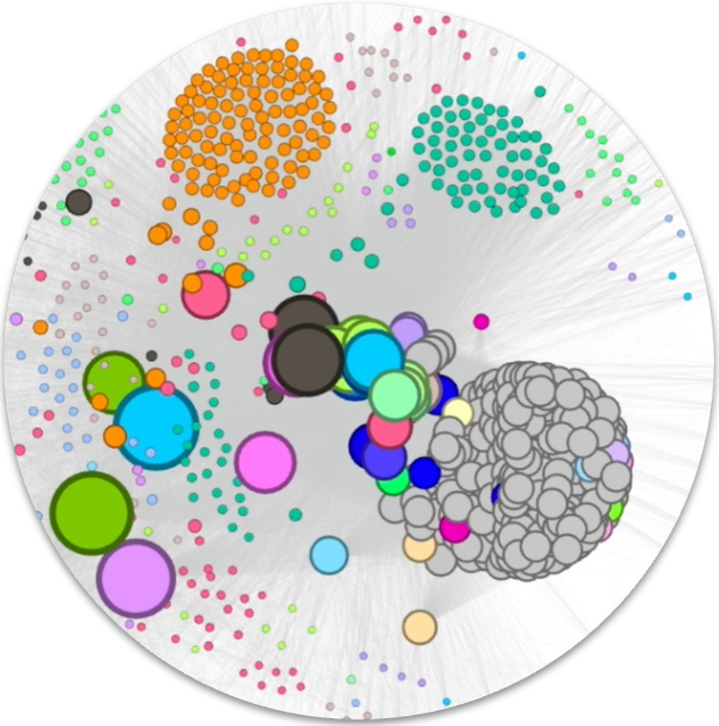
OICR Ontario Institute for Cancer Research

NIH National Institutes of Health Turning Discovery Into Health

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

Global Alliance for Genomics & Health Collaborate. Innovate. Accelerate.

International Health Cohorts Consortium



Overture is supported by grant #U24CA253529 from the National Cancer Institute at the US National Institutes of Health, and additional funding from Genome Canada, the Canada Foundation for Innovation, the Canadian Institutes of Health Research, Canarie, and the Ontario Institute for Cancer Research

https://bit.ly/courtotlab

SBO

OBI Ontology for Biomedical Investigations

IRIDA

GENEONTOLOGY Unifying Biology

Pan-Canadian Genome Library
Bibliothèque génomique pancanadienne

Sources

Rainbow heart, Lindsay Satchell, https://www.instagram.com/lindsaysatchellart/p/BGrodabiaeQ/

GPCR structure, Wikipedia, https://en.wikipedia.org/wiki/G_protein-coupled_receptor

SwissProt file, https://rest.uniprot.org/uniprotkb/P06213.txt

Maison des Tanneurs, from https://maison-des-tanneurs.com/

Vancouver, from https://www.nomadicmatt.com/travel-blogs/where-to-stay-vancouver/

ChatGPT logo, https://en.m.wikipedia.org/wiki/File:ChatGPT_logo.svg

Bloomberg, https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Bianchi, F. et al. Proc. 2023 ACM Conf. Fairness Account. Transpar. (FAccT '23) 1493–1504 (2023); available at https://doi.org/mkw9

World population from https://github.com/PietroViolo/world_population

Heterogenous mixture of buttons of different shapes and sizes. Danille Cageling / EyeEm, Getty Images

FAIR data image from https://www.nlm.nih.gov/oet/ed/cde/tutorial/02-200.html

dbGaP access diagram from https://sharing.nih.gov/sites/default/files/flmngr/Flyer_dbGaP_Access.pdf

Machine learning image from https://xkcd.com/1838/

AI-ready data: https://medium.com/@sean_hill/ai-ready-fair-data-accelerating-science-through-responsible-ai-and-data-stewardship-3b4f21c804fd

5-safe padlock image, https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

Biomedical ontology mapping, DOI:10.1186/s12859-016-1131-5