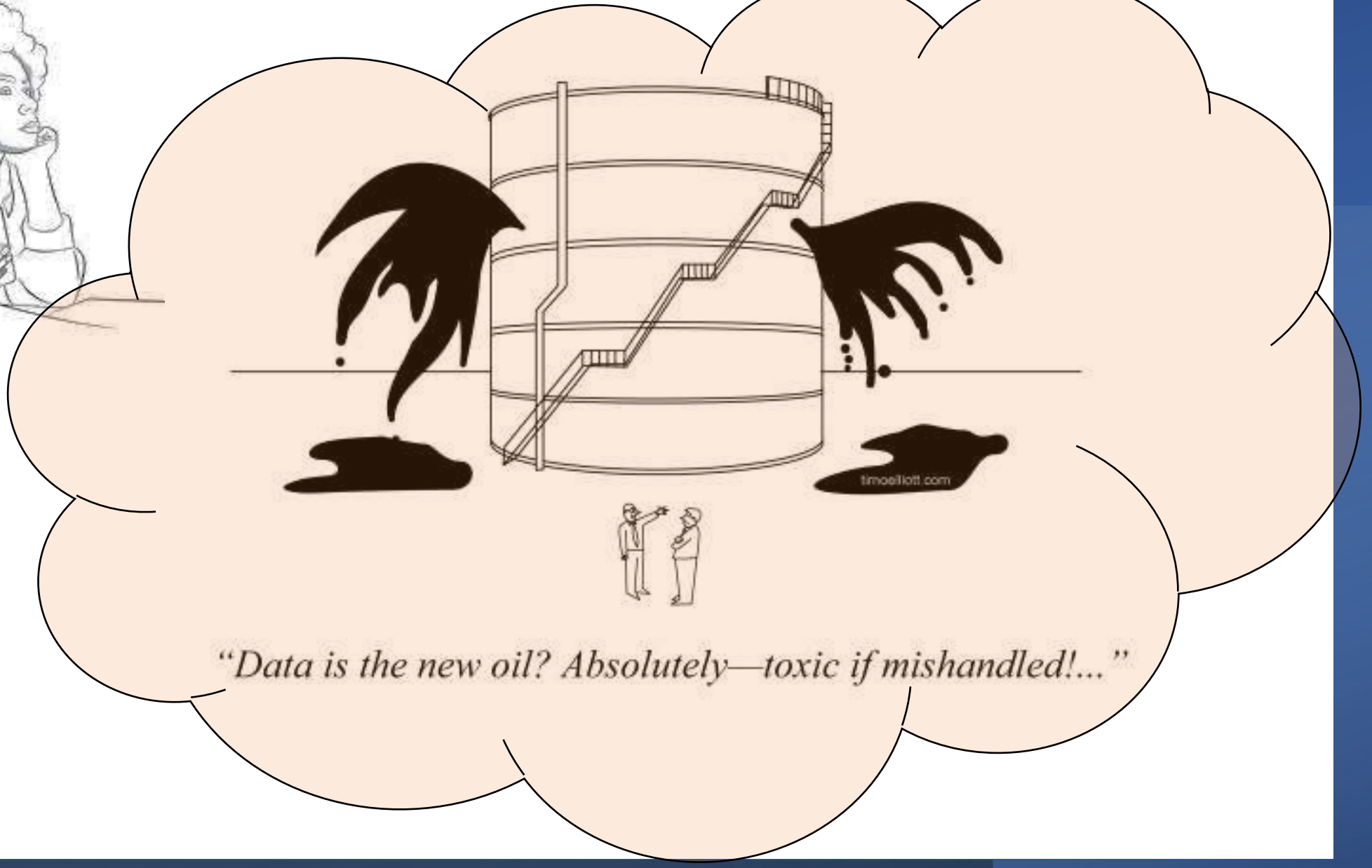
Agency for Science, Technology and Research
SINGAPORE

# **Data to Real World Impact:** A Journey with Purpose

Neerja Karnani
Deputy Director (Clinical)
Bioinformatics Institute, A*STAR
*neerja_karnani@bii.a-star.edu.sg*

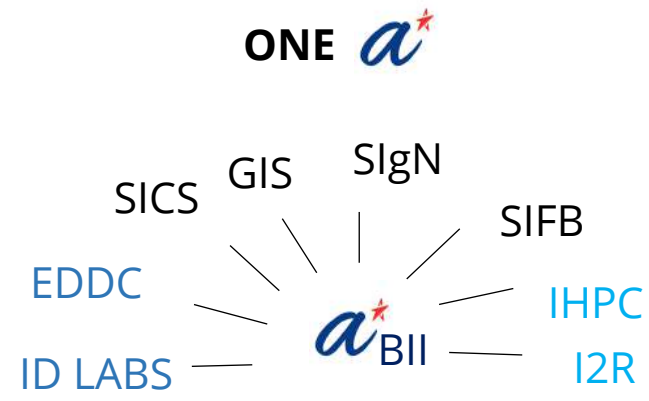



*"Data is the new oil? Absolutely—toxic if mishandled!..."*

# Data to Real World Impact: Our End-to-End Approach

| DATA | INTEGRATION | ANALYZE | REAL WORLD IMPACT |
|---|---|---|---|
| • High quality<br>• Multi-dimensional<br>• Adequate Sample Size<br>• Secure interphase for storage and analysis | • Discovery database and hypothesis playground<br>• Field/Disease specific use cases<br>• Cross-Population/cohort | • Develop problem statements<br>• Use deep domain knowledge to derive meaningful insights | • Disease Risk Prediction and Surveillance<br>• Clinical application<br>• Evidence based intervention |

## Trusted partner in the ecosystem with multiple collaborations with key stakeholders






Research Cohorts

Funding Agencies

Hospitals

International collaborations

# DATA

# **Purpose:** Addressing National health challenges

*Prevention* *to* *Treatment*

## Early life development

GUSTO, S-PRESTO, NiPPeR

## Function and Performance (Adults)

GUSTO, S-PRESTO, NiPPeR, iDAD_SG, TEAMS Study, ATTRaCT, PRISM, HELIOS, SERI, TTSH, MEC, PREPARE

## Healthy longevity

MEMOSA, SG10K Health, SG90, SCHS, DaHA

## Clinical Data

Metabolic Health

Cardiovascular Disease

Mental Wellness


Immune Health

>80 publications

3 patents and Industry licenses

## Multi-Omics Data

Genetics (GWAS & Polygenic Risk Scores)
Epigenetics (EWAS & EpiClocks)
Transcriptomics (Coding & noncoding)
Lipidomics
Metabolomics
Microbiome

National Research Foundation Prime Minister's Office Singapore | NMRC National Medical Research Council Singapore | SG10K Health | PRECISE | Ministry of Health Singapore | National University Hospital | National Centre for Infectious Diseases | SingHealth

Use case 1

# Clinical & Genomic Database Query = Clinical Impact

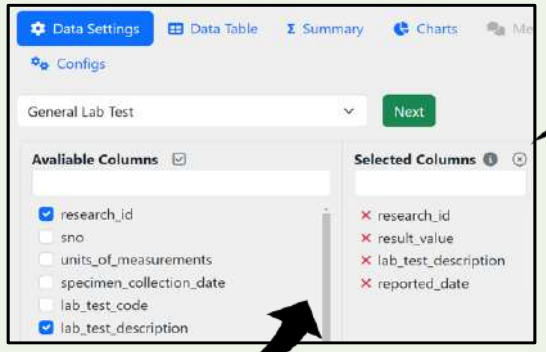

## Clinical SQL Database

Blood Tests, metabolomics

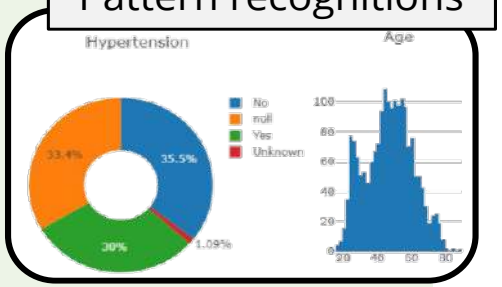
Prescription, Diagnosis

Clinic Visits, Phenotype, Family History

Clinical Filters




Dashboard Stats & Pattern recognitions




Genetic Counsellor App

## Genomic DB Query

WGS

[tile]DB

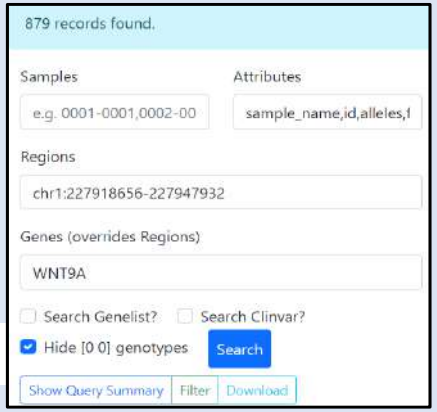
+

External DBs

ClinVar
Clinically relevant variation

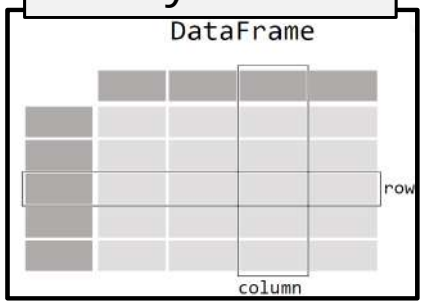
NCBI dbSNP

Fast Querying




Subject
Gene
Chr:Pos

AI & Analysis-ready format




# Use case 2 Multi-platform Engagement and Data Integration for a Unified National Goal


PREPARE


National Centre for Infectious Diseases

Bioinformatics Institute

BII

Demographics Exposures

REDCap
Research Electronic Data Capture

VANDERBILT UNIVERSITY

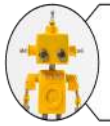
Editing existing Unique patient code SUBJ0001.

| | |
|---|---|
| Unique patient code | SUBJ0001 |
| Month and Year of Birth | 25-02-1979 Today D-M-Y Month, Year (mm-yyyy): |
| Singaporean; Permanent Resident; Long-term pass-holder; Short-term pass-holder | Singaporean Number ;1= Singaporean 2= Permanent Resident 3= Long-term pass-holder 4= Short-term pass-holder 88= Other 99=Unknown |
| Residency if Other | Text: |
| Living situation (majority of time in last 2 weeks) | Others Number ;Home Hotel Visiting friends/relatives home Military base Prison Boarding school/dormitory Nursing home/LTCF Others |
| Other living situation, specify | Text: |

BII

Same format, 9-hospital scale

Enables cross-hospital longitudinal patient tracking

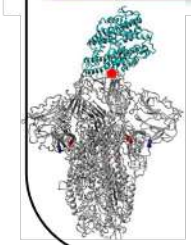


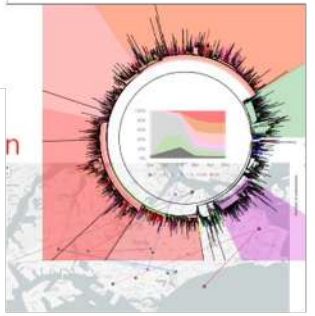
Automated CSV data import

## Virus genome surveillance

BII has global (GISAID, WHO, CEPI) and national role (NCID, MOH)

GISAID





Sebastian Maurer-Stroh

Penny Chan

Vachiranee Limviphuvad

## Inter-individual variation in Infection Severity

SG10K Health

MINISTRY OF HEALTH SINGAPORE

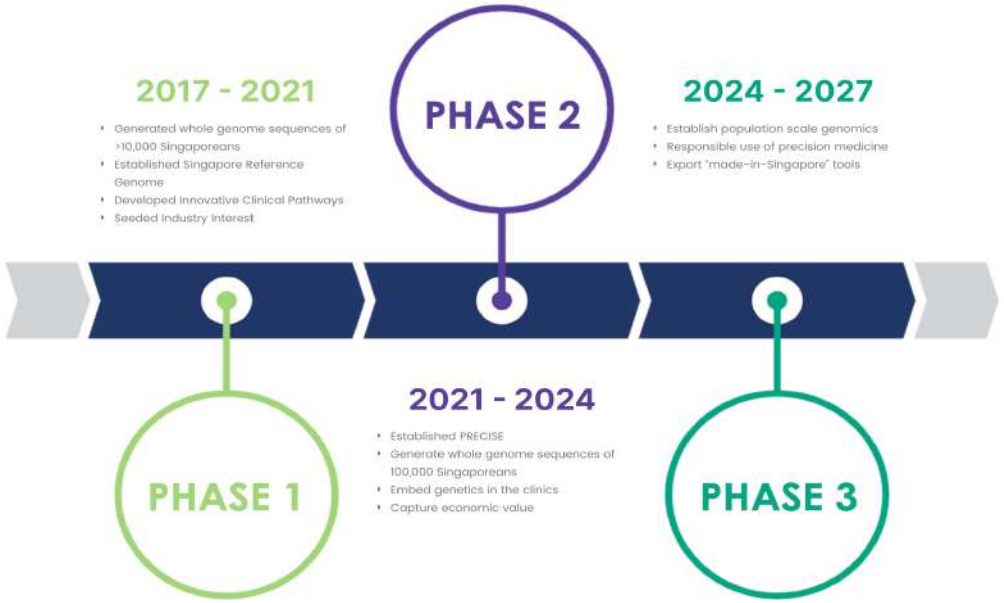
TRUST

Use case 3

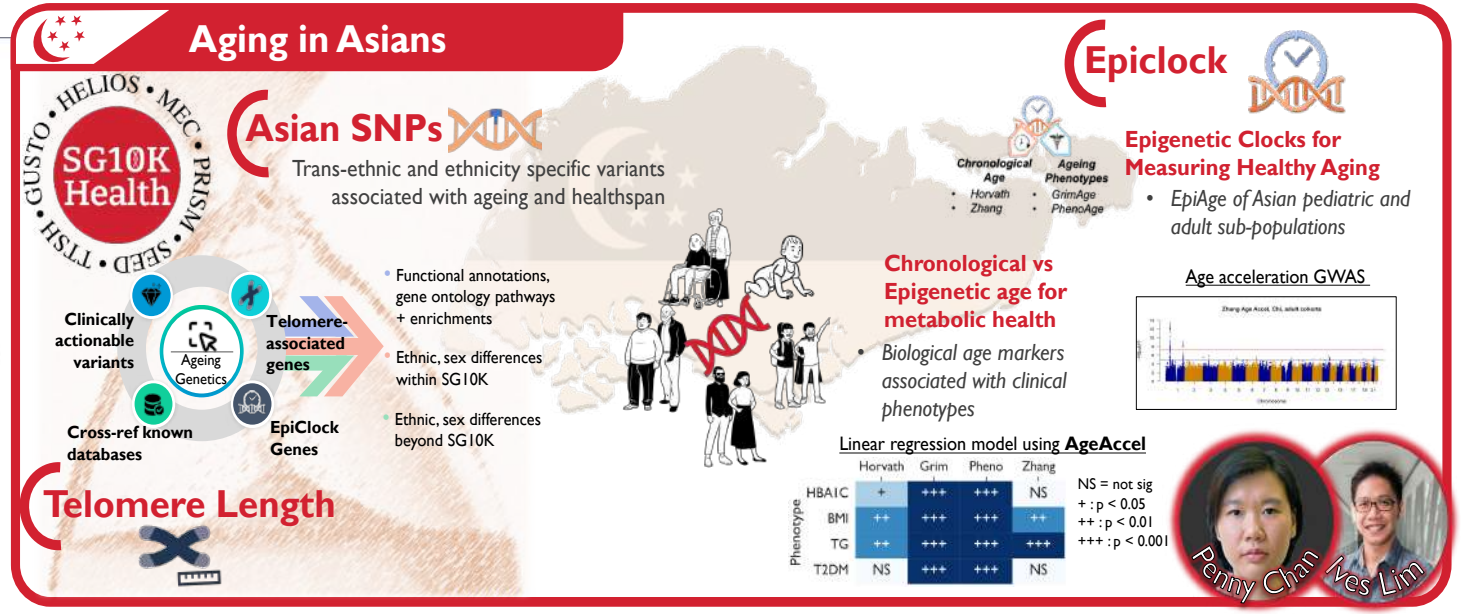
# Clinical and Multi-omics Data Integration to Study Biological Aging in Asians

PRECISE Precision Health Research Singapore

10K SG10K Health 100K IM




SG10K-Health Sentinel Study on Aging




Nestlé Patent

Nature Genetics 2023 Feb;55(2):178-186

Nat Commun. 2022 Nov 5;13(1):6694

Cell 2019 Oct 17;179(3):736-749.e15

iScience. 2023 Mar 31;26(4):106546

Transcriptional signature of Frailty: Nat Commun. 2019 Dec 20;10(1):5808

Abnormal Mitochondrial activity in Muscle wasting. J Cachexia Sarcopenia Muscle. 2023 May 4

Aging in Womb: BMC Med 2022 Jan 25;20(1)

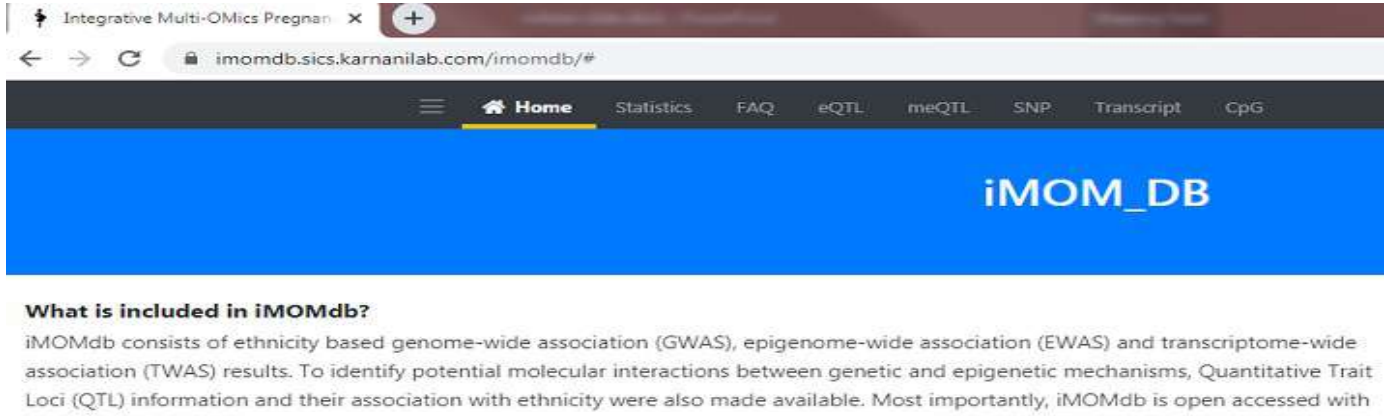
Bioinformatics Institute BII

Genome Institute of Singapore GIS

SG10K Health Consortium

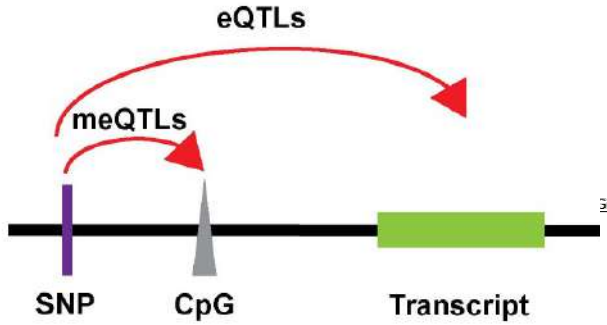
Use case 4

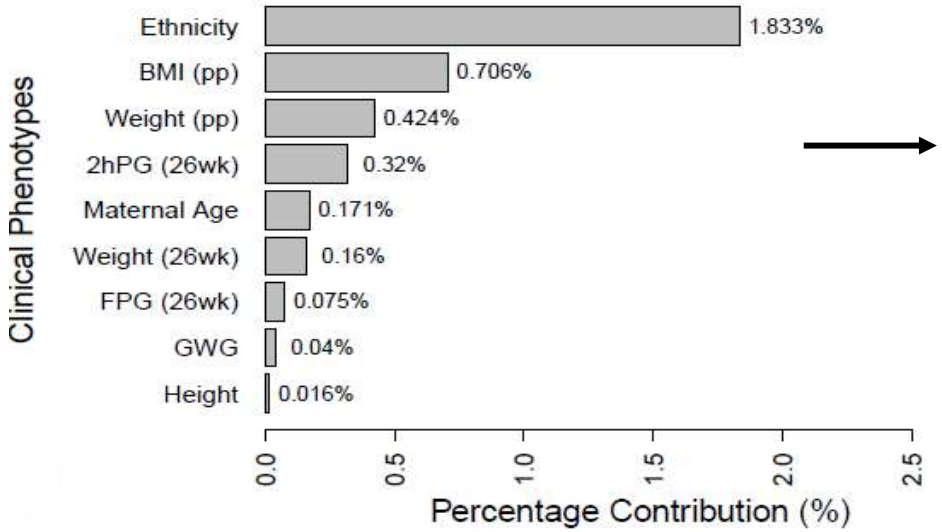
# Integrated Multi-Omics Database (iMOM_db) of Asian Pregnant Women

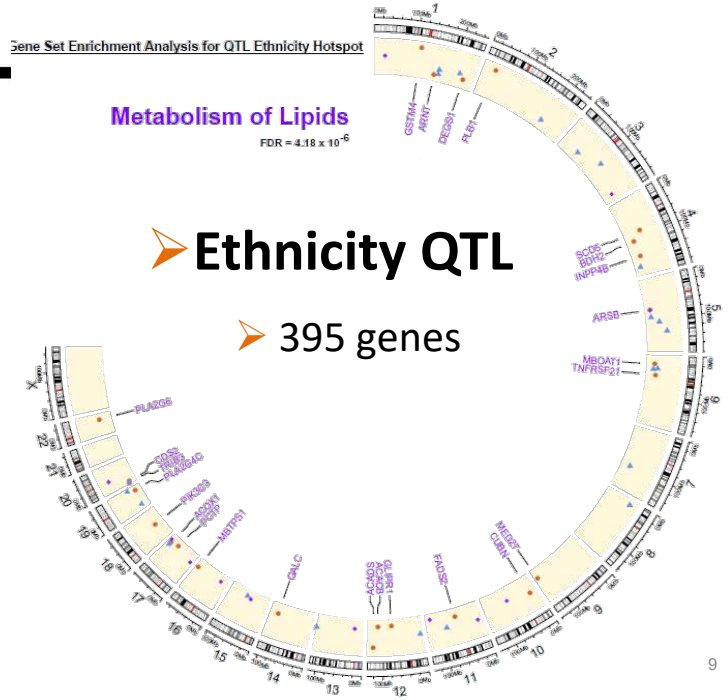



iMOMdb

*Hong, Tan, Lim, Huang and Karnani et al. Human Mol Genet 2022*

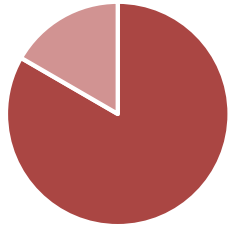




- **Ethnicity QTL**
  - 395 genes

## Plasma lipidome (validation)



- 83.3% of lipids detected sig. assoc. with ethnicity

Use case 5

# Diabetes Risk Prediction Modeling

Mukkesh Kumar
PhD Thesis

## 50% GDM cases routinely missed in SG

> Diabetes Res Clin Pract. 2022 Feb 4;185:109237. doi: 10.1016/j.diabres.2022.109237. Online ahead of print.

### Population-centric risk prediction modeling for gestational diabetes mellitus: A machine learning approach

Mukkesh Kumar [1], Li Chen [2], Karen Tan [2], Li Ting Ang [3], Cindy Ho [3], Gerard Wong [2], Shu E Soh [4], Kok Hian Tan [5], Jerry Kok Yen Chan [6], Keith M Godfrey [7], Shiao-Yng Chan [8], Mary Foong Fong Chong [9], John E Connolly [10], Yap Seng Chong [8], Johan G Eriksson [11], Mengling Feng [12], Neerja Karnani [13]

UK NICE guidelines showed poor predictability in Singaporean women [AUC:0.60 (95% CI 0.51, 0.70)]. The non-invasive predictive model comprising of 4 non-invasive factors: mean arterial blood pressure in first trimester, age, ethnicity and previous history of GDM, greatly outperformed [AUC:0.82 (95% CI 0.71, 0.93)] the UK NICE guidelines.

## Preconception Predictive Risk Modelling for GDM

International Journal of *Environmental Research and Public Health*

MDPI

*Article*

### Automated Machine Learning (AutoML)-Derived Preconception Predictive Risk Model to Guide Early Intervention for Gestational Diabetes Mellitus

Mukkesh Kumar [1,2,3], Li Ting Ang [1,2], Hang Png [1,2], Maisie Ng [1,2], Karen Tan [1], See Ling Loy [4,5], Kok Hian Tan [4,6], Jerry Kok Yen Chan [4,5,7,8], Keith M. Godfrey [9], Shiao-yng Chan [1,8], Yap Seng Chong [1,8], Johan G. Eriksson [1,8,10,11,†], Mengling Feng [3,12,*,†] and Neerja Karnani [1,2,13,*,†]

## GDM to T2D Risk analysis

**Machine Learning Derived Prenatal Predictive Risk Model to Guide Intervention and Prevent the Progression of Gestational Diabetes Mellitus to Type 2 Diabetes.** [Accepted for publication in JMIR Diabetes on 22 March 2022]

Kumar M, Ang LT, Ho C, Soh SE, Tan KH, Chan JK, Godfrey KM, Chan SY, Chong YS, Eriksson JG, Feng M, Karnani N

GUSTO
GROWING UP IN SINGAPORE TOWARDS HEALTHY OUTCOMES

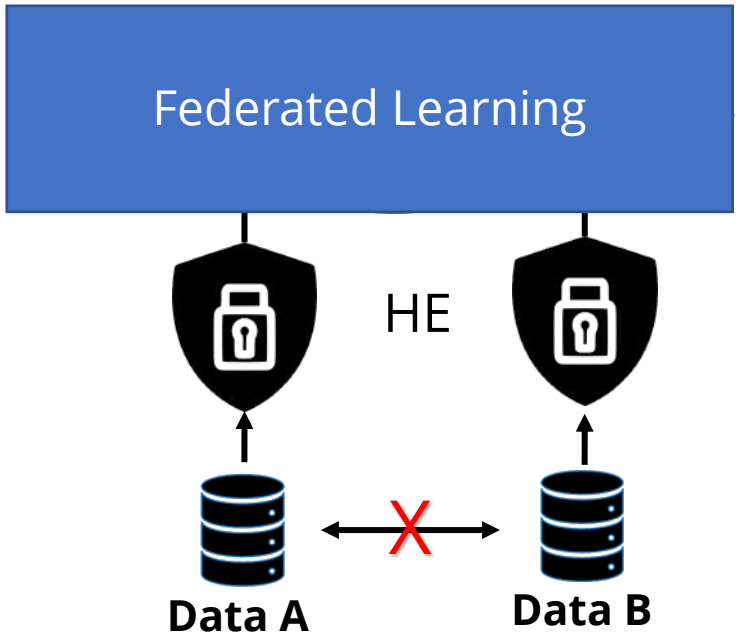
Singapore PREconception Study of long-Term maternal and child Outcomes
S-PRESTO

# Federated Data Learning - Overcoming Data Sharing Obstacles

.

**Homomorphic Encryption (HE)**

Enables computation of encrypted information without need for decryption




**Benefits:**

- ❑ Increased data security owing to HE
- ❑ Reduced regulatory hurdle for local and international cohort data integration

*Example – The Cardiometabolic Health Initiative*

***Polygenic Risk Score calculation***

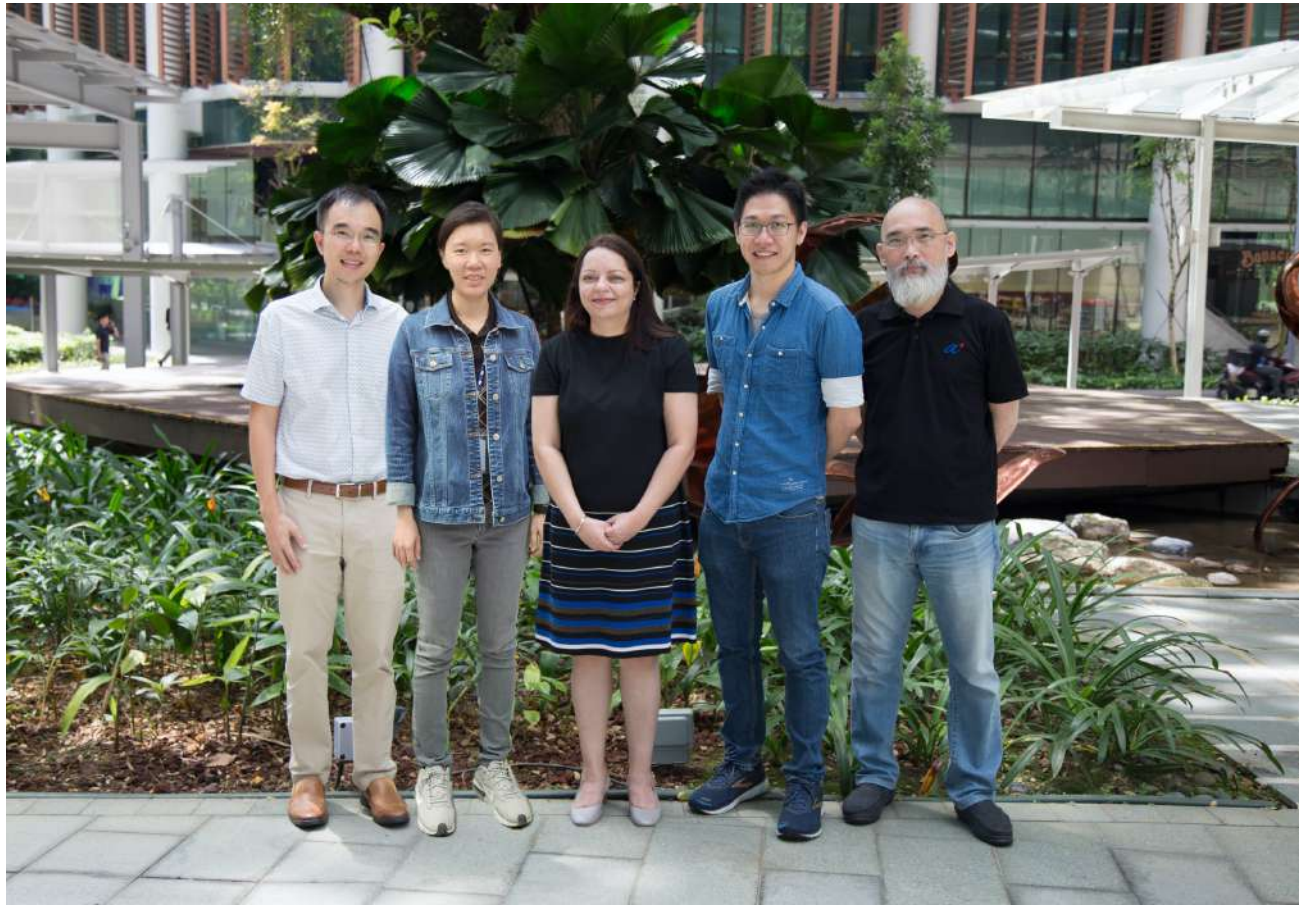
*Two local cohorts (PRISM & ATTRaCT) and >3000 subjects*

Bioinformatics Institute BII

Institute for Infocomm Research I²R

Genome Institute of Singapore GIS

# The Science Mavericks



**Kang Qi**

Developing an LSTM Model for Predicting Nutritional Deficiency Diseases with Medication Administration Records

@ SICS

Jia Xu

Kelly Ong

Gong Min

Felicia Tin

FERRING PHARMACEUTICALS

wilmar

KYOWA

Infinant HEALTH

Nestlé

CREATING GROWTH, ENHANCING LIVES

Agency for Science, Technology and Research
SINGAPORE

# THANK YOU

www.a-star.edu.sg