From Molecules to Medical Records – Insights From Large-Scale, Multi-Omic Studies

BIH Lecture Frontiers in Translational Medicine – Scientific and Structural Challenges
11th October 2021

Claudia Langenberg
Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin
MRC Epidemiology Unit, University of Cambridge, UK
Honorary Consultant Physician, Public Health England
‘OMICS?’

Genome
DNA

Transcriptome
RNA

Proteome
Proteins

Metabolome
Metabolites

Phenome
Disease

Extrinsic factors (Age, Diet, Drugs, Lifestyle, …)

- Diabetes
- Insulin resistance
- Obesity
- Fat distribution
The growth of GWAS, 2007–2017

Figure 2. Association of rs10830963 with type 2 diabetes (T2D) in 13 case-control studies.
From association to mechanism

Genome
DNA

Phenome
Disease

- Diabetes
- Insulin resistance
- Obesity
- Fat distribution
Genome to phenome via ‘omics’

Genome
DNA

Transcriptome
RNA

Proteome
Proteins

Metabolome
Metabolites

Phenome
Disease

Extrinsic factors (Age, Diet, Drugs, Lifestyle, …)

Diabetes
Insulin resistance
Obesity
Fat distribution
Blood based ‘omics’ at scale

- Genome
  - DNA

- Transcriptome
  - RNA

- Proteome
  - Proteins

- Metabolome
  - Metabolites

- Phenome
  - Disease
    - Diabetes
    - Insulin resistance
    - Obesity
    - Fat distribution

Extrinsic factors (Age, Diet, Drugs, Lifestyle, …)
Metabolome

Entirety of small molecules (<1kDa) in biospecimens like blood, urine, cells or tissues.

Measurement techniques

Mass spectrometry
(high-sensitivity, 1000s of molecules)

Nuclear magnetic resonance spectroscopy
(high reproducibility)
Metabolome-wide disease associations

EPIC-Norfolk Cohort (n=25,639)
Baseline 1993-97, mean age 60 yrs, 54% women
Untargeted metabolomics (n=987) > 11k participants
Incidence of 27 diseases using record linkage (hospitalisations): 219,415 person years of follow-up

‘Sharedness’ of disease associations

SHARED ASSOCIATIONS ACROSS 9 METABOLITE CLASSES

Antecedents of multimorbidity
Genome to phenome via ‘omics’

Extrinsic factors (Age, Diet, Drugs, Lifestyle,...)

- Genome DNA
- Transcriptome RNA
- Proteome Proteins
- Metabolome Metabolites
- Phenome Disease

- Diabetes
- Insulin resistance
- Obesity
- Fat distribution
‘Mendelian Randomisation’

Assumption 1
The link between the variants and having low cholesterol has to be strong and stable over time.

Assumption 2
The variants must not influence variables that might affect both cholesterol levels and cancer risk.

Assumption 3
The variants must not be associated with cancer risk in any way other than through the relationship to cholesterol.

---

THE CAUSATION DETECTOR
A technique called Mendelian randomization has become the go-to for drawing lessons from epidemiological data. But are scientists overdoing it? By David Adam

---

Epidemiologically, there has been enormous focus on genetic factors over the past 50 years. However, it is now clear that environmental factors, such as diet and exercise, play a significant role in determining health outcomes. Mendelian randomization is a technique that uses genetic variation to infer causality between exposure and outcome, providing insights into the mechanisms underlying disease.
Application: SELECT Trial

Randomized Controlled Trial (SELECT)

- Randomization method
- Exposed: Selenium supplementation
  - Plasma selenium + 114 µg/L
- Control: Placebo
  - Plasma selenium (Baseline)
- Confounders equal between groups
- Prostate cancer risk:
  - HR 1.04 (95% CI 0.91-1.19)

Selenium vs placebo

No. at risk
- Placebo: 8565, 8344, 8081, 7831, 7471, 6399, 4644, 1833, 70
- Selenium: 8600, 8360, 8131, 7826, 7456, 6425, 4075, 1829, 66

Cumulative cases
- Placebo: 31, 123, 202, 293, 384, 474, 532, 563, 575
- Selenium: 31, 123, 202, 293, 384, 474, 532, 563, 575

Klein E.A. et al. JAMA 2011; Yarmolinsky J. et al. JNCI 2018
Application: SELECT Trial

Randomized Controlled Trial (SELECT)

Randomization method

Exposed: Selenium supplementation

Control: Placebo

Plasma selenium + 114 μg/L

Plasma selenium (Baseline)

Confounders equal between groups

Prostate cancer risk: HR 1.04 (95% CI 0.91-1.19)

Mendelian Randomization

Random segregation of alleles

Exposed: Higher selenium alleles

Control: Reference alleles

Plasma selenium + 114 μg/L

Plasma selenium (Baseline)

Confounders equal between groups

Prostate cancer risk: OR 1.01 (95% CI 0.89-1.13)

Yarmolinsky J. et al. JNCI 2018
Maximising power: sample size 10-85k

MRC Fenland Cohort
Baseline 2005-15
N=12,435, mean age 49 years, 54% women
10,708 genotyped using 3 different arrays

Biocrates (AbsoluteIDQ p180)
174 targeted metabolites
- Hexoses
- Amino Acids
- Biogenic Amines
- Acylcarnitines
- Glycerophospholipids
- Sphingolipids
144 regions, 499 locus-metabolite associations

GLP2R and citrulline levels

- GLP2 stimulates intestinal growth > analogues used to treat short bowel syndrome
- Citrulline is a biomarker of intestinal function and target engagement

GLP2R - Glucagon-like peptide 2 receptor
- Expressed in the gut

Glucagon receptor
GLP1R – insulin secretion

GLP2R signaling
GIP levels (chronic)
GIP receptor signaling (beta-cells)
Reduced insulin secretion T2D

Effect of Teduglutide, a Glucagon-like Peptide 2 Analog, on Citrulline Levels in Patients With Short Bowel Syndrome in Two Phase III Randomized Trials

Reduced recruitment of β-arrestin to GLP2R

Genome-wide analyses identify common variants associated with macular telangiectasia type 2

“loci associated with glycine and serine metabolism”

Serine GWAS

Serine-pathway metabolites

In(OR) per genetically-predicted SD-difference in metabolite levels
Dose-response and MacTel2 prediction

Broadening scope: untargeted metabolomics

Surendran P., Stewart I. et al. unpublished
Summary (I of II)

- Increased scale and greater variant diversity increased the number of identified variants
- Cross-platform feasibility
- Specific mQTL characteristics
- Clinical utility and improved understanding of disease mechanisms
- Webserver: a resource for the scientific community (Helmholtz Centre Munich)

https://omicscience.org
https://biocrates.com/2021_cohort_webinar
Genome to phenome via ‘omics’

Genome: DNA

Transcriptome: RNA

Proteome: Proteins

Metabolome: Metabolites

Phenome: Disease

Extrinsic factors (Age, Diet, Drugs, Lifestyle, ...)

Diabetes

Insulin resistance

Obesity

Fat distribution
The plasma proteome

Key roles in diverse biological processes, dysregulated in disease, important drug targets

Concentrations of plasma proteins span almost 10 orders of magnitude

Huge dynamic range

No single technique is currently able to provide reliable measurements for all proteins
From proteins to proteome

Mass spectrometry of protein fragments (peptides)

Antibody-based (similar to an ELISA used in clinical chemistry)

Short oligonucleotides – aptamers – which match the 3D-conformation of the target protein

Population proteomics

12,435 participants born 1950-75 and living in Cambridgeshire

10,708 genotyped

Proteomics data

- **SomaScan v4** (~5,000 aptamers, N=12,435)
- **Olink** (12 panels; ~1,100 proteins, N=485)
- **MS-proteomics** (M. Ralser, Scanning SWATH ~340 proteins, N=485, target >12k)
Actively secreted proteins
  - Coagulation factors
  - Cytokines

Products of cell leakage and turnover

Soluble fragments of membrane proteins

Plasma protein patterns as comprehensive indicators of health

Stephen A. Williams, Mika Kivimaki, Claudia Langenberg, Aroon D. Hingorani, J. P. Casas, Claude Bouchard, Christian Jonasson, Mark A. Sarzynski, Martin J. Shipley, Leigh Alexander, Jessica Ash, Tim Bauer, Jessica Chadwick, Gargi Datta, Robert Kirk DeLisle, Yolanda Hagar, Michael Hinterberg, Rachel Ostroff, Sophie Weiss, Peter Ganz and Nicholas J. Wareham

NATURE MEDICINE | VOL 25 | DECEMBER 2019 | 1851-1857 | www.nature.com/naturemedicine
Isolated post-challenge hyperglycaemia

Elevated glucose 2-hours after an oral glucose load (IGT)
Strongly predictive of cardiometabolic diseases
Rarely measured (complexity, logistics, time)
Very common (globally ~7.5% of adults)

Isolated post-challenge hyperglycaemia is missed by FPG, HbA1c

Question: Is it feasible to identify a (fasting) proteomic signature to design a simple test that predicts isolated IGT?
Predicting post-challenge hyperglycaemia

Carrasco-Zanini J. et al. submitted
Prediction performance

IGT

Isolated IGT

Carrasco-Zanini J. et al. submitted
Top selected proteins from prediction models

• Associated with genetic susceptibility to impaired glucose homeostasis
• Associated with risk to develop T2D in an independent cohort

Carrasco-Zanini J. et al. submitted
Genome to phenome via ‘omics’

- **Genome**
  - DNA

- **Transcriptome**
  - RNA

- **Proteome**
  - Proteins

- **Metabolome**
  - Metabolites

- **Phenome**
  - Disease
    - Diabetes
    - Insulin resistance
    - Obesity
    - Fat distribution

**Extrinsic factors** (Age, Diet, Drugs, Lifestyle,...)
Protein quantitative trait loci

Pietzner M., Wheeler E. et al. under review
Proteo-Genomic Map of the Human Phenome

Fenland Study
(n=10,708)

SomaScan v4 assay
n=4,775 protein targets

Genome-proteome-wide association study

Proteo-genomic map of human health
www.omicscience.org

10,674 pQTLs

Local genetic architecture

Protein-encoding gene

1,548 cis-pQTLs

Pietzner M., Wheeler E. et al. under review
Convergence of soft tissue disorders: EFEMP1

Pietzner M., Wheeler E. et al. under review
Convergence of soft tissue disorders: \textit{EFEMP1}

Pietzner M., Wheeler E. et al. under review
Summary (II of II)

- Value of the plasma proteome for prediction: isolated IGT
- Value of increased breadth and scale
- Integration with phenomic data: genetically anchored disease map
- Next stages
  - ‘Neglected’ phenome
  - Covid19 prognosis
  - Clinical curation

https://omicscience.org
https://lifesciences.somalogic.com/webinar/liquid-health-check/
Computational Medicine, BIH: Maik Pietzner

Diabetes Aetiology Group (MRC Epidemiology)
Nick Wareham, Jian’an Luan, Eleanor Wheeler, Isobel Stewart, Maik Pietzner, Erin Oerton, Rupal Shah, Nicola Kerrison, Julia Carrasco-Zanini Sanchez, Mine Kopru, (graduates Laura Wittemans, Chen Li, Vicky Au Yeung, Nicholas Bowker, Lina Cai)

MRC Epidemiology Colleagues
Fumiaki Imamura, Tom White, Nita Forouhi, John Perry, Felix Day, Ken Ong, Debora Lucarelli

University of Cambridge
Kay-Tee Khaw, Jules Griffin, Albert Koulman

MRC Biostatistics Unit
Paul Newcombe, Stephen Burgess

Institute of Metabolic Science
Fiona Gribble, Frank Reimann, Stephen O’Rahilly, David B. Savage, Sadaf Farooqui

Study PIs and Functional and Laboratory Teams
Fenland, EPIC-Norfolk, InterAct, Interval, UK Biobank

Funding: UK MRC, EU FP7 (EMIF-Metabolic), Pfizer, GSK, NovoNordisk, Somalogic

mGAP (metabolome Genetic Architecture Programme) Team
Isobel Stewart, Luca Lotta, Laura Wittemans, Chen Li, Vicky Au Yeung, Nick Wareham (MRC Epidemiology Unit)
Praveen Surendran, Adam Butterworth, Angela Wood, Joanna Howson, John Danesh (Cardiovascular Epidemiology Unit)
Karsten Suhre (Weill Cornell Medicine, Qatar)
Gabi Kastenmüller, Johannes Raffler, Kieu Do, (Helmholtz, Munich)
Nicole Soranzo, Lorenzo Bomba (Wellcome Sanger Institute)
Cristina Menni, Jonas Zierer, Niccolò Rossi, Alessia Visconti, Mario Falchi, Tim Spector (King’s College London)
Eric Gamazon (Vanderbilt University, Nashville)
Eric Fauman (Pfizer, Boston)

pGAP (proteome Genetic Architecture Programme) Team
Aroon Hingorani, Chris Finan (UCL)

Collaborators
Melanie Bahlo, Roberto Bonelli and the MacTel Consortium
Robert Scott, Adrian Cortez (GSK)
Greg Michelotti, Annie Evans (Metabolon)
Steve Williams, Rachel Ostroff (Somalogic)
Thank you!

www.bihealth.org