Large scale computing with medical metabolic phenotyping data

Christoph Steinbeck and the PhenoMeNal consortium

European Bioinformatics Institute (EMBL-EBI)
INTERNATIONAL CONSORTIUM ANNOUNCES THE 1000 GENOMES PROJECT

Major Sequencing Effort Will Produce Most Detailed Map Of Human Genetic Variation to Support Disease Studies

An international research consortium has been formed to create the most detailed and medically useful picture to date of human genetic variation. The 1000 Genomes Project will involve sequencing the genomes of at least a thousand people from around the world. The project will receive major support from the Wellcome Trust Sanger Institute in Hinxton, England, the Beijing Genomics Institute Shenzhen in China and the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH).

Drawing on the expertise of multidisciplinary research teams, the 1000 Genomes Project will develop a new map of the human genome that will provide a view of biomedically relevant DNA variations at a resolution unmatched by current resources. As with other major human genome reference projects, data from the 1000 Genomes Project will be made swiftly available to the worldwide scientific community through freely accessible public databases.

LINKS

 Download the meeting report
 View the participants
Individual Human Genomes sequenced

Year


1

1000

10000

100000

1000000
In September 2014, an estimated 228,000 human genomes have been completely sequenced by researchers around the globe.

— Francis de Souza, president of Illumina
[O]ur understanding of the human genome has changed in the most fundamental ways. The small number of genes -- some 30,000 -- supports the notion that we are not hard wired. We now know the notion that one gene leads to one protein, and perhaps one disease, is false.

Craig Venter, June 2001
Genes are not the full story
The Exposome

Nutrition

Disease

Exercise

Age

Environment

Drugs
Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome

Shin Yoshimoto, Tze Mun Loo, Koji Atarashi, Hiroaki Kanda, Seidai Sato, Seiichi Oyadomari, Yoichiro Iwakura, Kenshiro Oshima, Hidetoshi Morita, Masahira Hattori, Kenya Honda, Yuichi Ishikawa, Eiji Hara & Naoko Ohtani

Affiliations  Contributions  Corresponding author

Nature 499, 97–101 (04 July 2013)  doi:10.1038/nature12347
Received 10 December 2012  Accepted 04 June 2013  Published online 26 June 2013
Corrigendum (February, 2014)
Reaction times following external change

- Genetics (decades, centuries…)
- Epigenetics (days, month, years,…)
- Gene Expression (hours)
- Metabolism (seconds)
The Metabolome is the most accessible and dynamically changing Molecular Phenotype
Metabolomics

Measures occurrence and concentrations of many small molecules (metabolites) in an organism at once.
Diagnostic fluids

- Urine (time-averaged data)
- Plasma (snap-shot data)

Other accessible analytical compartments

- Specialized fluids and biopsies (selected fluids)
- Pathological fluids
- Artificial fluids

Organism Parts
Metabolomics uses a wide-range of analytical techniques

Nuclear Magnetic Resonance (NMR)

Mass Spec
Typical 500 MHz $^1$H-NMR spectrum of human cerebrospinal fluid. Numbers indicate the following metabolites:


source: http://www.csfmethabolome.ca
1. DSS
2. L-serine
3. 3-hydroxybutyric acid, 4. acetoacetate, 5. creatinine, 6. pyruvate, 7. L-alanine, 8. 3-phosphoglyceric acid, 23. L-histidine, 24. L-tyrosine, 25. methanol, 26. glycerol

source: http://www.csfmetabolome.ca
Typical 500 MHz $^1$H-NMR spectrum of human cerebrospinal fluid. Numbers indicate the following metabolites:


source: http://www.csfmetabolome.ca
Untargeted Metabolomics of Cerebrospinal Fluid

Alzheimer’s Disease (AD), Mild Cognitive Impairment (MCI), Cognitive Normal (CN)
Untargeted Metabolomics of Cerebrospinal Fluid

Basis for a diagnostic tool

Alzheimer’s Disease (AD), Mild Cognitive Impairment (MCI), Cognitive Normal (CN)
• 500 Million people in European Union
• Full Genomes (soon for less than $1000 per patient)
• Urine Metabolome < 20 Euros per patient
Phenome Centres popping up all over the world

- London
- Birmingham
- Shanghai
- NIH RCMRCs
- ...

...
welcome

MRC and NIHR funded and led by Imperial College London and King's College London, the National Phenome Centre will deliver broad access to a world-class capability in metabolic phenotyping.

What is the Phenome Centre and who is involved?

PROFESSOR JEREMY NICHOLSON
DIRECTOR OF MRC-NIHR PHENOME CENTRE AND HEAD OF DEPARTMENT OF SURGERY AND CANCER, IMPERIAL COLLEGE LONDON

Staff related links

→ Professor Jeremy K Nicholson (Director of the National Phenome Centre)
→ Dr Anthony C Dona (NMR Manager)
→ Mr Matthew R Lewis (Mass Spectrometry Manager)
→ Miss Lynn Maslen (Centre Manager)
→ Dr Jake TM Pearce (Informatics Manager)
→ Dr Rachel J Shaw (Head of Science Operations and Analytics)

News stories

→ New centre will decipher roles of nature and nurture in human health
→ National phenome research facility to open at Imperial

Health legacy for London 2012
> 100,000 patient samples / year
> Several PetaBytes/year
=> ExaBytes of human data at moderate scale-up
How do you make sense of all that data?
Share them
-
Free and Open
Privacy issues with data from human cohorts
Open BioData Sharing

- OpenGenomes.org
- Commercial community efforts (23AndMe …)
- Quantified-Self revolution
  - Individual efforts
  - Apple health kit
- Who will write the FaceBook for Personal Medical Data?
'Smart toilets and sewer sensors are coming'

TECHNOLOGY / 21 MARCH 14 / by YANIV J TURGEMAN AND ERIC ALM AND CARLO RATTI

This article was taken from the March 2014 issue of Wired magazine. Be the first to read Wired's articles in print before they're posted online, and get your hands on loads of additional content by subscribing online.

Advances in genomics are moving the bioengineering laboratory into the city sewers. In 2014, our wastewater systems will become smart, with automated sampling and lab-on-a-chip biosensors enabling the real-time prediction of viral outbreaks, making connections
Advances in genomics are moving the bioengineering laboratory into the city sewers. In 2014, our wastewater systems will become smart, with automated sampling and lab-on-a-chip biosensors enabling the real-time prediction of viral outbreaks, making connections.
The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson  06.23.08
The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson  06.23.08

Data-driven Science
http://www.ebi.ac.uk/metabolights

MetaboLights

open-access, cross-species, cross-application, long-term supported

MetaboLights Database

Reference Layer
- Chemistry
- Spectroscopy
- Biology

Primary Literature
Primary data and Meta-Data, Spectra, Protocols, Synopses, ...

Experimental Repository
Data growth in EBI data repositories
Data growth in EBI data repositories

3-month doubling time for Metabolomics
Metabolic differences in ripening of *Solanum lycopersicum* ‘Ailsa Craig’ and three monogenic mutants

Stephan Beisken, Mark Earl, Charles Baxter, David Portwood, Zsuzsanna Ament, Aniko Kende, Charlie Hodgman, Graham Seymour, Rebecca Smith, Paul Fraser, Mark Seymour, Reza M. Salek & Christoph Steinbeck

**Abstract**

Application of mass spectrometry enables the detection of metabolic differences between groups of related organisms. Differences in the metabolic fingerprints of wild-type *Solanum lycopersicum* and three monogenic mutants, *ripening inhibitor* (*rip*), *non-ripening* (*nop*) and *Colourless non-ripening* (*cnn*), were identified, suggesting potential targets for breeding and crop improvement.
Global Collaboration in Metabolomics and the BioSciences
COSMOS
COOrdination of Standards in MetabolOmicS

- Create missing standards & formats
- Define workflows for dissemination
- Create world-wide data network

European FP7 coordination action coordinated by us at EMBL-EBI, Hinxton, Cambridge
Global network for exchange and discoverability of metabolomics data

Includes study as well as reference data
Latest datasets

A Large-scale Metabolic Profiling Strategy for the Dissection of Plant Defense against Fungal Pathogens
by Konstantinos Aliferis

"Here we present a metabolic profiling strategy employing direct infusion Orbitrap mass spectrometry (MS) and gas chromatography-mass spectrometry (GC/MS) for the monitoring of soybean's (Glycine max L..."

A Metabolic Profiling Strategy for the Dissection of Plant Defense against Fungal Pathogens
by Konstantinos Aliferis

"Here we present a metabolic profiling strategy employing direct infusion Orbitrap mass spectrometry (MS) and gas chromatography-mass spectrometry (GC/MS) for the monitoring of soybean's (Glycine max L..."

PIWI grape leaves
by Panagiotis Arapitsas

"LC-MS untargeted analysis of leaves from two fungus-resistant grape varieties (PIWI) grape varieties (Regent and Phoenix) collected from two countries (Italy and Germany)..."
What’s next?
Contributing to the development of P4 Medicine

• predictive
• preventive
• personalized
• participatory
PhenoMeNal: Large Scale Computing with Medical Metabolic Phenotyping Data

- H2020
- 3 Years
- 13 Partners
- 8 Mio €
- 830 PM
- Kick-off 9/15
PhenoMeNaIal Context

Human factors
- Genomic
- Epigenetic
- SNP variations
- Protein synthesis
- Transporter activity
- Enzyme function
- Immune status

Individual/Population Health Status
- Healthy
- At risk
- Disease

Metabolic/Physiological Phenotypes
- Normal
- Intermediate
- Pathological

Environmental factors
- Dietary inputs
- Microbiome
- Allostatic load
- Stress
- Drug use
- Pollutant exposure
- Disease vectors
- +/- Parasites

Population-level Models
Metabolome-wide Association
- Biomarkers: Disease risk, disease, nutrition, gut functional ecology

Drug or nutritional intervention studies

Personalized Healthcare Models
(safety and efficacy)
PhenoMeNal: Large Scale Computing with Medical Metabolic Phenotyping Data
PhenoMeNal Objectives

1. To use existing open source community standards, integrate tools, resources and methods for the management, dissemination and computational analysis of very large datasets of human metabolic phenotyping and genomic data into a secure and sustainable e-Infrastructure
PhenoMeNal Objectives

2. To operate and consolidate the PhenoMeNal e-infrastructure based on existing internal and external HPC (High-performance computing), cloud, and grid resources, including the EGI and the EGI Federated Cloud, and to extend it to world-wide computational infrastructures
PhenoMeNal Objectives

3. To **improve and scale-up tools** used within the infrastructure to cope with very large datasets

4. To **establish technology for a water-tight audit trail** for the processing of human metabolic phenotyping data from the raw data acquisition all the way to the generation of high-level biomedical insights (such as a medical diagnosis)
5. To establish privacy-protection methods that allow working with highly sensitive molecular phenotype data

6. To foster the worldwide adoption of PhenoMeNal through a wide range of outreach, dissemination, networking and training activities

7. To develop a model to ensure sustainability of the PhenoMeNal network
PhenoMeNal Structure

- WP1: Management (EMBL-EBI)
- WP2: Sustainability of PhenoMeNal (U Leiden)
- WP3: Dissemination and Outreach (U Birmingham)
- WP4: Interfacing with Biomedical European Infrastructures (CIRMMP Florence)
- WP5: Operations and Maintenance of PhenoMeNal GRID/Cloud (U Uppsala)
- WP6: PhenoMeNal Virtual Research Community Gateway (EMBL-EBI)
- WP7: Privacy and Ethics (Imperial College)
- WP8: Data Provenance, Compliance, and Integrity (U Oxford)
- WP9: Tools, Workflows, Audit and Data Management (IPB Halle)
PhenoMeNal: Large Scale Computing with Medical Metabolic Phenotyping Data
PhenoMeNal-EGI cooperation

Phenomics User Community

PhenoMeNal

ELIXIR cloud activities

EGI

Indigo Data Cloud

iSphere

AAI  Federated Cloud  AppDB  ....
Funding and Collaborators

UK Research Councils (BBSRC, MRC)
European Commission
Thanks for your hospitality