Cutting-Edge Technologies: Landscape & Perspectives for Health Outcomes: Genomics and Health Data

Prof Tim Hubbard, King’s College London
Head of Genome Analysis, Genomics England
Associated Director, Health Data Research UK London

Cutting-Edge Health Technologies: Opportunities and Challenges
A joint technical symposium by WHO, WIPO and WTO

Geneva October 2019
Overview

1. Genomic medicine as a Cutting-Edge Technology
2. Power of data reuse for clinical care; data use for research
3. Complexity of human biology; progress in understanding it
4. Other health care data types
5. Data infrastructure: need for trust; need for regulation
1. Potential to **start** implementing genome medicine from whole genome sequence (WGS)
Fall in price of sequencing beats Moore’s law
Price low enough for healthcare application

![Graph showing the fall in price of sequencing and comparison to Moore's Law over time. The y-axis represents cost per genome in dollars, and the x-axis represents years from 2001 to 2012.]
Subset of medical conditions with limited complexity where whole genome is useful

- Limit objective to explaining an existing medical condition
  - Easier than predicting if a person will suffer from an illness in the future
- Limit to diseases where number of causal genetic variants is small
  - Rare disease diagnostics (single variant)
  - Cancer (small number of variants)

- 100,000 genomes project has focused on these cases
  - Participants have to have eligible condition
  - Clinical reporting limited to the affected condition
100,000 genomes project has been successfully delivered

**Samples**
- **122,650**
  - Samples collected and received at the UK Biocentre

**Genomes**
- **112,198**
  - Genomes sequenced

**Analysis and Results**
- Results for **86,944** genomes sent to NHS
- Equivalent to **21,551** cancer genomes and **65,393** rare disease
- **85,954**
  - Cancer
  - Rare disease

**28,508**
- Cancer
- Rare disease

~50% cancer reports contain potential for a therapy or a clinical trial

20-25% actionable findings in rare disease reports

**Figures as at 01/06/2019**
As 100,000 genomes project ends, WGS continues in NHS Genome Medicine Service

• Implementation of NHS Genome Medicine Service
  • Single NHS genetic testing directory from single gene test to WGS
  • Pharmacogenomics tests incorporated for first time

• Commitment to 1m whole genomes sequenced in UK over next 5 years
  • 500,000 from continuation of 100,000 genomes project in NHS
  • 500,000 from sequencing of research cohort UK Biobank
  • Aspiration for 5m genomic tests over next 5 years

• 100,000 genomes project legacy
  • Existing infrastructure to support WGS sequencing and data handling within NHS
Jessica Wright

• Jessica, aged 4
• Rare condition that causes epilepsy and affects her movement and general development.
• Took part in the 100,000 Genomes Project rare disease programme with parents at Great Ormond Street Hospital.
• Found that she had a genetic variant in the SLC2A1 gene - makes a protein that transports a certain type of sugar into the brain. Mistakes in the SLC2A1 gene can cause ‘Glut1 deficiency syndrome’ – Jessica’s diagnosis.
• In some patients who have Glut1 deficiency syndrome a very low-carbohydrate diet (ketogenic) can help reduce the number of seizures.
• Thanks to WGS analysis, Jessica’s clinician was able to recommend this diet for her, which helped with her seizures.
Clinical use of whole genome sequence

- Cumulative prevalence of "rare diseases" is significant (~5%)
- Precise Genomic diagnosis (causal variant)
  - Ends diagnostic odyssey
  - May suggest alternative treatment with existing medications
  - Provides molecular target for development of new therapies
  - May be amenable to gene therapy (e.g. CRISPR)
- An individuals genome is unchanging (except genome of cancer cells)
  - If stored, can be analysed for different conditions over an individuals lifetime
  - Analysis cost is near instant, nearly zero additional cost
  - Can be an low cost alternative to companion diagnostic
- Organisations such as Global Genome Medicine Collaborative (G2MC) sharing expertise on implementation [https://g2mc.org/](https://g2mc.org/)
2. Power of infrastructure enabling data reuse for clinical care; data use for research
Data design for 100,000 genomes project

- Clinical Reports
- Data Centre
- Clinical App
- Patients
- Consent
- Sample / Clinical Data
- Genome Medicine Centres (GMCs)
- Genome Laboratory Hubs (GLHs)
- Firewall
Data design for 100,000 genomes project +Reuse

- Pharmacogenomics?
- Risk scores?
- Clinical Reports

- Genes
- Phenotypes
- Environment

- Patients
- Clinicians

- Consent

- Sample / Clinical Data

- Data Centre
  - Clinical Apps

- Genome Medicine Centres (GMCs)
- Genome Laboratory Hubs (GLHs)

- Firewall
Data design for 100,000 genomes project + Reuse + Trusted Research Environment (TRE)

- Patients
- Clinicians
- Consent
- Sample / Clinical Data
- Genome Medicine Centres (GMCs)
- Genome Laboratory Hubs (GLHs)

- Firewall
- Clinical Reports
- Data Centre
  - Clinical Apps
  - Research Apps (Di-identified)
- Access Review Committee
- Airlock
- Export: summary only
- Import: Tools External data

Researchers
3. Complexity of human biology and progress in understanding it
Biological complexity - Human Biology

• Human body has ~37 trillion cells (37,000,000,000,000)
  • 200+ types of cell: Blood, muscle, nerve, immune system etc.

• Each cell has 2 complete copies of entire instructions to make a human body
  • The genome sequence - 3 billion letters of DNA (3,000,000,000)

• Differences between Individuals:
  • On average 4 million differences between genomes of any two individuals
Biological complexity - Human Biology

**Organism**, system of interacting organs

**Organ**, system of interacting cells

**Cell**, living system built from DNA, RNA, Proteins and the products of their interactions

**Protein** (translation of gene coding regions)

**RNA** (transcripts of expressed genes)

**DNA** (copy of genetic material in every cell)
Medical Information

**Organism**, system of interacting organs

**Organ**, system of interacting cells

**Cell**, living system built from DNA, RNA, Proteins and the products of their interactions

**Protein** (translation of gene coding regions)

**RNA** (transcripts of expressed genes)

**DNA** (copy of genetic material in every cell)

Clinical Medicine
### Medical Information + Biological Information

| **Organism**, system of interacting organs |
| **Organ**, system of interacting cells |
| **Cell**, living system built from DNA, RNA, Proteins and the products of their interactions |
| **Protein** (translation of gene coding regions) |
| **RNA** (transcripts of expressed genes) |
| **DNA** (copy of genetic material in every cell) |

- [Clinical Medicine](#)
- [Human Genome Project](#)
Organism, system of interacting organs

Organ, system of interacting cells

Cell, living system built from DNA, RNA, Proteins and the products of their interactions

Protein (translation of gene coding regions)

RNA (transcripts of expressed genes)

DNA (copy of genetic material in every cell)
Medical Information + Biological Information - modelling in reality

**Organism**, system of interacting organs

**Organ**, system of interacting cells

**Cell**, living system built from DNA, RNA, Proteins and the products of their interactions

**Protein** (translation of gene coding regions)

**RNA** (transcripts of expressed genes)

**DNA** (copy of genetic material in every cell)

- Clinical Medicine
- iPS cells, organoids
- Human Cell Atlas
- Expression, other omics data
- ENCODE Project; Gene set
- Human Genome Project
Medical Information + Biological Information + Individual differences – *modelling effects*

**Organism**, system of interacting organs
- predict consequences for disease

**Organ**, system of interacting cells
- predict consequences for disease

**Cell**, living system built from DNA, RNA, Proteins and the products of their interactions
- predict consequences for disease

**Protein** (translation of gene coding regions)
**RNA** (transcripts of expressed genes)

**DNA** (copy of genetic material in every cell)

**Clinical Effect on Individual**

**Clinical Medicine**

**iPS cells, organoids**

*Human Cell Atlas*

**Expression, other omics data**

*ENCODE Project; Gene set*

**Sequence Variants of Individual**

**Human Genome Project**
• Human body has ~37 trillion cells (37,000,000,000,000)
  • 200+ types of cell: Blood, muscle, nerve, immune system etc.
• Each cell has 2 complete copies of entire instructions to make a human body
  • The genome sequence - 3 billion letters of DNA (3,000,000,000)
• Differences between Individuals:
  • On average 4 million differences between genomes of any two individuals

• Need to study millions of humans to increase understanding of Human Biology and to better predict response of individuals to healthcare
  • Only health systems have adequate scale
  • Requires analysis across national systems – Federated analysis
  • Organizations such as Global Alliance for Genomes and Health (GA4GH) [https://www.ga4gh.org/](https://www.ga4gh.org/) leading on policies and implementation for federated data analysis
• Need to better represent all human populations to ensure algorithms are not biased
4. Other health care data types
Growth in health data not limited to genomics

- WGS
- 100,000 Genomes Project
- Genome Medicine Service
- Images
- Devices
- PROMS

Data per individual

- 10K
- 100K
- 1M
- 10M
- 100M
Coordinated UK initiatives to improve health data landscape for both care and for research

Electronic Health Records
- Genomics England: Research Environment
- InnovateUK Centres
- NHS Digital: Digital Services Platform
- Longitudinal Patient Records
- Imaging Pathology & AI

10,000 Genomes Project
- WGS
- Genome Medicine Service

Data per individual

10K 100K 1M 10M 100M
• Launched in April 2018
• Partnership with academia, industry, NHS, charities and patients.
• Science, Training, Infrastructure, Engagement
• Not a data controller
UK Health Data Research Alliance: Uniting the UK’s health data

Improved coordination of different Trusted Research Environments

Single gateway improve discovery for research users
5. Data infrastructure: need for trust; need for scale; need for regulation
What are the requirements to build big health data infrastructure?
What are the requirements to build big health data infrastructure?

• Public Trust
  • Engagement

• Contractual requirements for EHR, Cloud providers
  • Data Security
  • Data Use
  • Data Standards
  • Data Interoperability
  • Data Control
100,000 genome project example of participant engagement

Ensures the interests of participants are always at the centre of the 100,000 Genomes Project. They do this by:

- Making sure experiences of participants are at the heart of the project
- Responding to feedback.
- Overseeing who should have access to participant data

Access Review Committee
GeCIP Board
Ethics Advisory Committee

Participants

The 100,000 Genomes Project
Joining the National Participant Panel

Are you taking part in the 100,000 Genomes Project?

Genomics England is looking for participants to be part of the national 100,000 Genomes Project Participant Panel.

The role of the Panel is to ensure that the interests of participants are always at the centre of the 100,000 Genomes Project. They will make sure that the experiences of participants are improved, respond to feedback and oversee who should have access to participant data.
Summary

1. Genomic medicine as a Cutting-Edge Technology
2. Power of data reuse for clinical care; data use for research
3. Complexity of human biology; progress in understanding it
4. Other health care data types
5. Data infrastructure: need for trust; need for regulation
Acknowledgements

• 100,000 genomes project patients and their families, Genomics England Team, NHS England, Health Education England, Public Health England

• Genomic Medicine Centres in England, Scotland, Wales and Northern Ireland, UK CLL Consortium, CRUK, RCPath, NHSE, DoH, Biobank UK, NIHR BioResource Rare Disease, DDD, NIHR Translational Research Collaborative

• King’s College London, Wellcome Trust Sanger Institute, European Bioinformatics Institute, University College London and Queen Mary’s University London, Health Data Research UK.