The 100,000 Genomes Project

Jim Davies

Seattle, October 2015
The 100,000 Genomes Project

Project announced by the Prime Minister in December 2012

Genomics England announced by Secretary of State for Health in July 2013
Rare Disease

- incidence for each disease < 0.05%
- > 7000 diseases
- cumulative incidence in population ~5%
- cardiovascular, endocrine and metabolism, gastroenterology and hepatology, hearing and sight, immunology and haematology, inherited cancer predisposition, musculoskeletal, neurological, paediatric sepsis, paediatrics, renal, respiratory, skin
- 50,000 participants
Rare Disease

- 30x coverage whole genome sequences
- Phenotyping using Human Phenotype Ontology

<table>
<thead>
<tr>
<th>Disease Group</th>
<th>Renal and urinary tract disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Subgroup</td>
<td>Syndromes with prominent renal abnormalities</td>
</tr>
<tr>
<td>Specific disease</td>
<td>Alport syndrome</td>
</tr>
</tbody>
</table>

**Basic Phenotyping**

<table>
<thead>
<tr>
<th>Phenotype Description</th>
<th>Phenotype Identifier</th>
<th>Phenotype Present</th>
<th>Modifiers</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteinuria</td>
<td>HP:00000093</td>
<td>Unknown</td>
<td></td>
<td>Edit</td>
</tr>
<tr>
<td>Hematuria</td>
<td>HP:0000790</td>
<td>Unknown</td>
<td></td>
<td>Edit</td>
</tr>
<tr>
<td>Nephrotic range proteinuria</td>
<td>HP:0012593</td>
<td>Unknown</td>
<td></td>
<td>Edit</td>
</tr>
<tr>
<td>Renal insufficiency</td>
<td>HP:0000083</td>
<td>Unknown</td>
<td></td>
<td>Edit</td>
</tr>
</tbody>
</table>
Cancer

- ovarian, breast, colorectal, prostate, lung, leukaemia, sarcoma, renal, rare and childhood cancers, cancer of unknown primary
- 25,000 participants
- 75x coverage somatic, 30x coverage germline
- dataset reflects requirements for genomic medicine
Expectations

- New Diagnoses
- NHS Infrastructure
- Scientific Discovery
- Economic Growth

12 October 2015
Consented participants, samples, phenotyping, clinical and laboratory data, clinical reports, confirmatory testing, and follow-up.

**Identified Data**
Collected against agreed models, linked to secondary data from other sources.

**Other Data**
Hospital Episode Statistics, registry data, mortality, and more.

**Research Data**
De-identified versions of the data made available via a virtual desktop infrastructure.

**Tools and Services**
Quality assurance, clinical interpretation, tools for analysis, reference data.

**Participants**
NHS Genomic Medicine Centres

**Biorepository**
DNA & samples for multi-omics

**Sequencing**
illumina

**Researchers**
Genomics England Clinical Interpretation Partnerships, Industry Consortia, and other approved researchers
Infrastructure

- Genomic Medicine Centres (NHS England)
- NIHR national biosample centre
- Hinxton sequencing centre (Wellcome Trust)
- secure data centre (NIHR, MRC)
Data Acquisition

- registration, consent, family history, phenotyping via web-based interface
- clinical and laboratory data submitted as XML messages, reports upon clinical events
- secondary data from clinical audits and national reporting datasets
- integrated, longitudinal record via LabKey server for review and quality assurance

- patient-reported data
Research Environment

- separate data centre
- a research ‘embassy’ for each group
- access to shared LabKey server
  - explore data within policy
  - take snapshots for analysis
- private cloud
- ‘airlock’ system

- virtual desktop infrastructure
Progress

- 11 Genomic Medicine Centres
- 5000 whole genome sequences
- 125 rare diseases
- 5 cancers
- 1 pathogen
- 4 tool providers
- 10 pharma companies

- 2500 researchers
Contributors (include)

- NIHR Biomedical Research Centres
- NIHR Bioresource for Rare Diseases
- NIHR Health Informatics Collaborative
- Medical Research Council
- Cancer Research UK
- Wellcome Trust
- ...

12 October 2015
Challenges

• earn the trust of:
  • patients
  • clinicians
  • researchers

• help to transform:
  • clinical informatics
  • research informatics
  • culture
WHAT DOES GENOMICS ENGLAND DO WITH YOUR DATA?

GO!

CONSENT

YOU CAN WITHDRAW AT ANY TIME
3

- Your health data
- Your medical record data
- Your genome data
WHAT WILL SCIENTISTS DO WITH THE DATA?

UNDERSTAND WHICH DIFFERENCES ARE IMPORTANT

BEST TREATMENT

- USE CURRENT MEDICINES BETTER
- NEW DRUGS
- NEW DIAGNOSTIC TESTS
Thanks

• The Genomics England team: in particular,
  - Sir John Chisholm (Executive Chair)
  - Professor Mark Caulfield (Chief Scientist)
  - Vivienne Parry (Communications and Outreach)
  - Ed Stafford (Director of Informatics)

• LabKey
  - enjoy the rest of the conference!