The 100,000 Genomes Project

Dr Tom Fowler, Director of Public Health

Personalised Medicine Policy Forum – 26th October
The 100,000 Genomes Project

Background

Announced by the Prime Minister in December 2012

An Olympic Legacy

Genomics England announced by Secretary of State for Health in speech during NHS 65th Anniversary Celebrations, July 2013
The first human genome sequence

26th June 2000 - Cost $3.2 billion

100,000 Genomes at Millennium Prices - Cost $320 trillion
Four main aims

1. To bring benefit to NHS patients
2. To create an ethical and transparent programme based on consent
3. To enable new scientific discovery and medical insights
4. To kickstart the development of a UK genomics industry
Rare disease
The scale of rare diseases

1 in 17 people will suffer from a rare disease at some point in their lives.

In the UK alone that equates to approximately 3.5 million people.

Only a quarter of rare diseases have had their molecular basis defined, meaning many risk being undiagnosed and therefore untreated.

There are at least 6,000 rare diseases.

Many rare diseases (approximately 80%) are of genetic origin.

Seventy-five per cent of rare diseases affect children.

30% of rare disease patients die before their fifth birthday.
The case for deep phenotypic data in rare disease

Luke Jostins and Gil McVean

29 October 2015
Phenotyping in Rare Disease

• Working on 110 phenotypes
• Detailed and genomically primed clinical phenotyping
• Diagnostic tests- genomic, imaging, pathology
• Human Phenotype Ontology
• Disease progression
• Additional phenotypes guided by clinicians
• Building a National Rare Disease Registry
• Linked to Primary Care, Hospital Episodes and Outcomes
First families diagnosed from the Newcastle BioResource pilot

1. Leslie Hedley, 57
WGS revealed Mr Hedley’s kidney failure was caused by a particular genetic variant (INF2 mutation). His family is also being tested and their blood pressure can now be effectively controlled by drugs available on the NHS.

2. William and Allan Carpenter aged 79 and 69
Suffered from muscle loss and weakness. They have now been diagnosed with inherited nerve damage, known as peripheral neuropathy.
The brothers may be joining a treatment trial which, if successful, could prevent family members developing the same condition.
Father and daughter: inherited kidney disease
# Diagnostic Odyssey

## Distal sensory-motor neuropathy

<table>
<thead>
<tr>
<th>Previous molecular investigations</th>
<th>Additional investigations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEFL gene seq (2010) (*£400)</td>
<td>Nerve conduction studies (x 4) - £1600</td>
</tr>
<tr>
<td>PMP22 gene seq (2010) (*£300)</td>
<td></td>
</tr>
<tr>
<td>MPZ seq (2010) £200</td>
<td></td>
</tr>
<tr>
<td>CX32 seq (2009) (*£250)</td>
<td></td>
</tr>
<tr>
<td>Dosage PMP22 (Fl PCR dosage test 2002) (*£155)</td>
<td></td>
</tr>
</tbody>
</table>

**Total cost:** £2905 (excluding routine biochemical investigations)
Cancer
Cancer

• Disease of disordered genomes – over 200 drivers known
• Drugs target mutations. Tumour heterogeneity/ evolution of cancer/ Stratified medicine

• Lung, breast, colon, prostate, ovary and, Leukaemia and
• Rare and Childhood Cancers, unknown primary

• Diagnosis and therapy is evolving

• International Cancer Genomes Consortium- the Cancer Genome Atlas
• Cancer Research UK Stratified Medicine Programme
• UK Leukaemia consortia
• NIHR Biomedical Research Centres
Lung cancer before and after Gefitinib
Phenotyping in Cancer

- Phenotype models - commonality ICGC
- Diagnostics and Imaging
- Molecular Pathology & genomic tests
- Linked to NHS disease specific cancer registries
- MDT Information- Chemo & Radiotherapy
- Outcomes
Pathology Processing: Tissue

1. Cutting the tissue
2. Fixation & Processing
3. Tissue embedding in paraffin
4. Microscopic evaluation
5. Slide staining
6. Block cut to make slides
7. Boundary & % tumor purity or cellularity
8. TissueMark® provides boundary and % tumor
9. Digital review by one or more pathologists
10. MACRODISSECTION & MOLECULAR ANALYSIS
Pathogens
Infections
Pathogens

• Stratifying response, minimising adverse events and tracking outbreaks
• M. Tuberculosis resistance and epidemiology
• Hepatitis C genotype selects therapy
• HIV – Treatment for life and resistance testing is in the care pathway.
• Extreme human response to sepsis
• International linked datasets
Establishment Phase

- Illumina - NHS Genomic Medicine Sequencing Centre in Hinxton
- UK Data Infrastructure for Genomic Medicine (with MRC)
- NIHR National Biosample Centre - £24 million state-of-the-art facility to store the samples
- 11 NHS Genomic Medicine Centres in England to enrol, validate and feedback to patients
Genomics England specific Responsibilities

- Sequencing Centre
- Biorepository
- Consent
- Clinical Data
- Samples
- Biorepository
- Genomics England Informatics Infrastructure
- Scientific + Clinical Users
- Commercial Users
- Sequencing Centre

Clinician -> Consent
Clinical Data -> Consent
Samples -> Biorepository
Biorepository -> Genomics England Informatics Infrastructure
Genomics England Informatics Infrastructure -> Scientific + Clinical Users
Genomics England Informatics Infrastructure -> Commercial Users
Genomics England Informatics Infrastructure -> Sequencing Centre
Pipelines from sequence to patient

**Sequence Data**
- Data Validation
- MD5
- BAM and VCF
- QC

**Orchestration**
- Orchestrator Engine
- Sample ID Tracking

**Participant data from LabKey**
- Archiving on tape
- StateDB

**Participant Catalogue**
- Consistent?
- Reconciliation Checks: Sex, Ethnicity, Tumour-Normal pairing, Family relationships, Contamination (normal, non-human, other human), LOH checks, Heterozygocity estimation

**Data Verification**
- Export

**Illumina redelivery**
- Delivery API

**Clinical Interpretation**
- Report Request
  - BAMs
  - VCFs
  - Sex
  - Pedigree
  - Phenotypes
  - Annotation
  - Presumed MOI/Family history/Pedigree?
  - Consent status
  - Cancer type
  - Primary vs met
  - Stage

**Secondary Analysis**
- Alignment and Variant calling
- Variant score calibration and filtering

**Variant STORE**
- Genotypes and Annotation
- Annotation store
- CellBase

**Gene discovery**
- Ranked Candidate Variants
- RARE DISEASES
- Karyotype
- Ranked Candidate Variants on drivers
- Fusion genes
- Plots
- Other relevant info

**Cancer Gene Lists**
- KP variants
- BAM statistics
- Aggregate statistics
- Calculation

**Report Generation**
- Ranked Candidate Variants
- GEL sample tracking
- Report DB

**Report distribution**
- Report review
- Report triaging
- and distribution

**Variant Review**
- Variant Justification

**Clinical Decision and Reporting Support Tool**
- Feedback to participant
- Variant technical validation and positive patient identification
- Clinical Scientist Signoff
- Further functional validation, cascade testing, transmission testing

**GMC Validated Report**
- Feedback to patient
- Referral to GECIP

**GECIP**
- Multidisciplinary team support
- Application
- Further functional validation, cascade testing, transmission testing

**Clinical Reporting Pipeline**
- Randomized file paths?

**Legend**
- Tool: process or service
- External process
- Subprocess
- Database
- Data record (and schema)
- Process to patient

29 October 2015
Eleven Genomic Medicine Centres (GMCs) established in December 2014 by NHS England. These centres will lead the way in delivering the 100,000 Genomes Project.

- Track-record of providing excellence in genomic services.
- Eligible patients will be referred to GMCs by their clinicians.
- First patients recruited by Manchester GMC in March.
What will we be telling participants?

• Information about a patient’s main condition

• Information about ‘serious and actionable’ conditions (optional)

• Carrier status for non affected parents of children with rare disease (optional)
Clinical Interpretation Partnerships

- We ran a world-wide competition to test the state-of-the-art in annotation and interpretation of genomic data
- It is clear that scientific and clinical understanding is far from mature
- So we have initiated the creation of a network to draw in the best clinical and scientific minds from the UK and around the world to access our data and interpret the results

The Genomics England Clinical Interpretation Partnerships are

- Defined by diseases or cross cutting themes. We expect between 30-50 GeCIP Domains in all
- Each must have a clear and unique focus and be open to all experts in that field. Many to number more than 100 members
- Once designated GeCIPs have free access to the data infrastructure and the dataset
- All IP held by Genomics England but available for licence

Work of ~30 disease groups involving upwards of 2500 Clinicians and Scientists
## GeCIP Domains

### Rare
- Cardiovascular
- Endocrine and Metabolism
- Gastroenterology and Hepatology
- Hearing and Sight
- Immunology and Haematology
- Inherited Cancer Predisposition
- Musculoskeletal
- Neurological
- Paediatric Sepsis
- Paediatrics
- Renal
- Respiratory
- Skin

### Cancer
- Breast
- Colorectal
- Lung
- Ovarian
- Prostate
- Childhood Solid Cancers
- Haematological Malignancy
- Pan Cancer
- Renal
- Sarcoma

### Functional
- Electronic Records
- Validation and Feedback
- Ethics and Social Science
- Functional Effects
- Health Economics
- Machine Learning, Quantitative Methods and Functional Genomics
- Population Genomics
- Translational Research
- Functional Cross Cutting
- Stratified healthcare
Health Education England

Upskilling the workforce

• 9 University providers of MSc in Genomic Medicine – aimed at NHS healthcare professionals working in England

• HEE - Genomics Education Programme, online training courses and resources
  • University of Birmingham
  • Newcastle University
  • University of Manchester
  • University of Sheffield
  • Imperial College London
  • Queen Marys University of London
  • St Georges, University of London
  • University of Cambridge
  • University of Southampton
GENE Consortium
Working with industry

• Ten companies have come together to create the Genomics Expert Network for Enterprises (GENE) Consortium to oversee a year-long Industry Trial

• Aims to identify most effective and secure way of bringing industry expertise into the 100,000 Genomes Project in order to realise the potential benefits for patients.

• AbbVie, Alexion, AstraZeneca, Berg, Biogen, Dimension Therapeutics, GSK, Helomics, Roche, Takeda
Progress so far

- Over 5,000 participants have already agreed to take part as part of the pilots: 3,500 in rare diseases and almost 2,000 in cancer.
- We have already delivered 5,000 whole genome sequences, and we are starting to interpret these to help patients.
- These first 5,000 sequences are from the pilot phase and our GeCIP domains and GENE Consortium partners will shortly begin working with this data.
- [http://www.genomicsengland.co.uk/100000-genomes-project-update/](http://www.genomicsengland.co.uk/100000-genomes-project-update/)
PanelApp

A crowdsourcing tool for gene panels

- A publically-available resource that allows gene panels to be viewed, downloaded and evaluated by the Scientific Community.

- Initial gene panels have been established for all the approved rare diseases (Version 0), and graded using a traffic light system to indicate the number of sources.

- We are seeking expert review of these panels.

Aims:

- Source expert knowledge to establish a **final diagnostic grade gene panel** (or “green list”) for each disorder that will be used in the classification of genetic variants to aid clinical interpretation of rare disease genomes (Version 1).

- Engage the Scientific Community, encourage open debate, and begin to establish consensus on gene panels for rare diseases.

- A mechanism to allow access to the panels, standardisation of terms and collection of gene-disease related information, accumulation of reviews over time, and updated releases (Version 2...).
PanelApp

https://bioinfo.extge.co.uk/crowdsourcing/PanelApp/

**Public access**
- View and download gene panels.
- View Reviewers’ comments.

**Register to be a reviewer**
- View and download gene panels.
- View Reviewers’ comments.
  - Evaluate genes and make comments.
  - Add genes to a gene panel.
PanelApp information

https://bioinfo.extge.co.uk/crowdsourcing/PanelApp/

Welcome to the Genomics England PanelApp
A crowdsourcing tool to allow gene panels to be shared, downloaded, viewed and evaluated by the Scientific Community

Introduction

Download the PanelApp Handbook PDF here (document version 3.3)
Watch a quick 3 minute introductory video to Panel App
Watch a video providing instructions for Reviewers of PanelApp gene panels

Gene panels for all approved rare disease categories in the Genomics England 100,000 Genomes Project have been initially established using the criteria outlined in Figure 1 from four sources. Each source was manually searched for key phenotypic words (e.g. disease category and/or key phenotypes in the eligibility statement), and genes tested for under these phenotypes were included. Gene lists from experts in particular disease areas have also been added to the panels, some of which include those submitted during the establishment of eligibility criteria and data models for the 100,000 Genomes Project rare disease programme.

Each gene in a panel has been assigned an initial level of confidence based on the number of the 4 sources that the gene was collected from, as indicated by a traffic light system:

- Green = highest level of confidence; a gene from 3 or 4 sources.
- Amber = intermediate; a gene from 2 sources.
- Red = lowest level of confidence; a gene from 1 of the 4 sources or from an expert list.

It should be noted that no additional curation of the literature was carried out to establish these lists.
## Example: Familial Diabetes panel

### Genes

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Mode of Inheritance</th>
<th>Rating</th>
<th>Gel Status</th>
<th>OMIM</th>
<th>Source of Evidence</th>
<th>Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCK</td>
<td>BOTH monoallelic and biallelic, autosomal or pseudoautosomal</td>
<td>Green List (high evidence)</td>
<td>4</td>
<td>138079</td>
<td>Emory Genetics Laboratory, Illumina TruGenome Clinical Sequencing Services UKGTN Radboud University Medical Center, Nijmegen</td>
<td>Maturity-onset diabetes of the young (MODY); Maturity Onset Diabetes of the Young; MODY, type II, 125851; Transient Neonatal Diabetes, Recessive; Permanent Neonatal Diabetes Mellitus (recessive); Maturity Onset Diabetes of the Young (Dominant)</td>
</tr>
<tr>
<td>HNF1A</td>
<td>MONOALLELIC, autosomal or pseudoautosomal, imprinted status unknown</td>
<td>Green List (high evidence)</td>
<td>4</td>
<td>142410</td>
<td>Emory Genetics Laboratory, Illumina TruGenome Clinical Sequencing Services UKGTN</td>
<td>MODY, type III, 600496 (Diabetes mellitus, noninsulin-dependent, 2), 125853 (Diabetes mellitus, insulin-dependent), 222100 (Hepatic adenoma, somatic, 142310; Renal cell carcinoma, 144700; Diabetes mellitus, insulin-dependent, 20, 612520; Maturity Onset Diabetes of the Young)</td>
</tr>
<tr>
<td>HNF4A</td>
<td>MONOALLELIC,</td>
<td>Green List (high evidence)</td>
<td>4</td>
<td>600281</td>
<td>Emory Genetics Laboratory, Radboud University Medical Center, Nijmegen</td>
<td>MODY, type I, 125850; Maturity Onset Diabetes of the Young</td>
</tr>
</tbody>
</table>

*High evidence Ratings: 2
Low evidence Ratings: 0
Number of comments: 2*
In 10 years - Genomics England

• 100,000 WGS on NHS patients and pathogens
• WGS deployed routinely- also in other diseases
• Harnessing electronic health records
• Patient reported outcomes and more remote monitoring
• New diagnostics and therapies and opportunities for patients
• By end of 2017
Stay in touch

www.genomicsengland.co.uk

info@Genomicsengland.co.uk

@genomicsengland
#genomes100k

Follow ‘Genomics England’