The 100,000 genomes project

Tim Hubbard @timjph
Genomics England
King’s College London, King’s Health Partners
Wellcome Trust Sanger Institute

ClinGen / Decipher
Washington DC, 26th May 2015
The 100,000 genomes project

• Primarily a treatment project
  • NHS transformation project

• All clinical whole genome sequencing (>30x)
  • Rare disease (proband/parent trios)
  • Cancer (normal/tumour pairs)

• Timeline
  • Announced December 2012
  • Genomics England setup 2013
  • Pilots 2014
  • Main Programme 2015-2017

http://www.genomicsengland.co.uk/
Genomics England – mission

• 100,000 whole genome sequences in NHS patients with rare diseases and cancers from the NHS in England
• Health improvement and wealth generation
• Legacy of infrastructure, human capacity and capability
• Enable large scale genomics research

http://www.genomicsengland.co.uk/
Genomics England

Sequencing
(Illumina)

Contract

Genomics England
Genomics England

NHS Genome Medicine Centres

Sequencing (Illumina)

Contract

Contract

NHS England
Genomic Medicine Centres (GMC)

- Designated local NHS Lead and extended team
- Capacity and capability networks
- High fidelity phenotypes
- Access to data and samples
- February start date

Eleven Genome Medicine Centres announced

- **East of England NHS GMC**: Led by Cambridge University Hospitals NHS Foundation Trust;
- **South London NHS**: Led by Guy’s and St Thomas’ NHS Foundation Trust.
- **North West Coast NHS GMC**: Led by Liverpool Women’s NHS Foundation Trust.
- **Greater Manchester NHS GMC**: Led by Central Manchester University Hospitals NHS Foundation Trust
- **University College London Partners NHS GMC**: Led by Great Ormond Street Hospital NHS Foundation Trust
- **North East and North Cumbria NHS GMC**: Led by The Newcastle upon Tyne Hospitals NHS Foundation Trust.
- **Oxford NHS GMC**: Led by Oxford University Hospitals Foundation Trust.
- **South West Peninsula NHS GMC**: Led by Royal Devon & Exeter NHS Foundation Trust.
- **Wessex NHS GMC**: Led by University Hospital Southampton NHS Foundation Trust.
- **Imperial College Health Partners NHS GMC**: Led by Imperial College Healthcare NHS Trust.
- **West Midlands NHS GMC**: Led by University Hospitals Birmingham NHS Foundation Trust.
Genomics England

Phenotype data

DNA

NHS Genome Medicine Centres

Consent

Sequencing (Illumina)

BAM/VCF

Contract

Contract/Procurement

Data Centre

NHS England

Genomics England
IT infrastructure for genomic medicine: a reading library of clinical and genomic data

- HSCIC / PHE
- NHS - GMC
- Biorepository
- Sequencing Centre
- Clinical Data Infrastructure
  - Identities
  - Relationships
- N3 Network

Data Store
- Research Infrastructure
- Virtual Data Centres
- Analytics services

CROWN HOSTING DATA CENTRES
powered by ARK in partnership with Cabinet Office
Genomics England

DNA → Sequencing (Illumina) → BAM/VCF

Phenotype data

NHS Genome Medicine Centres → Consent → Genomics England

Contract

Genome Interpretation Service Companies

Grants, Contracts

Data Centre → Clinical Interpretation Services

Contract/Procurement

Clinical Report
Research Protocol under new designation of “Bioresource”

• Single **project-wide approval**: no need for site specific approvals

• Independent review committee grants data access to bona-fide research uses

• Consent for return of **additional findings** (secondary: 17 genes; and carrier status: 8 genes)

• Participants can be **re-contacted** up to four times a year

• Samples for various **-omics** technologies collected

• Revision of diagnosis if underlying evidence changes (e.g. when new is gene discovered)

http://www.genomicsengland.co.uk/library-and-resources/
Data model development

• Which participants should we recruit?
  • List of conditions – currently 122
  • Eligibility statements

• What data do we need?
  • Metadata: Demographics, Sample, Consent
  • Clinical data: Data models
  • Associated genes: Gene packages

Developing data models is complex for rare disease patients, needing consultation with experts in the field

To date:

• First release of eligibility statements and data
• First draft (candidates) of gene packages
HPO as universal ontology for phenotypic features

Human Phenotype Ontology

- Chosen as a **standard for deep representation of phenotypic features**
- Adopted by **other projects**, e.g. DDD, FORGE, familiar to many in rare diseases
- Being **actively developed** in collaboration with broader RD community
- **Existing mapping** from diseases to HPO terms
Data models are specific to each condition

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare Disease Conditions and Phenotypes(11144.4)</td>
<td>Cardiovascular disorders(10950.1)</td>
<td>Connective Tissues Disorders and Aortopathies(10951.1)</td>
<td>Familial Thoracic Aortic Aneurysm Disease(11021.1)</td>
</tr>
<tr>
<td></td>
<td>Cardiac arrhythmia(10952.1)</td>
<td></td>
<td>Brugada syndrome(11022.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Long QT syndrome(11023.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Catecholaminergic Polymorphic Ventricular Tachycardia(11024.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cardiomyopathy(10953.1)</td>
<td>Arrhythmogenic Right Ventricular Cardiomyopathy(11025.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Left Ventricular Noncompaction Cardiomyopathy(15044.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dilated Cardiomyopathy (DCM)(11026.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dilated Cardiomyopathy and conduction defects(11027.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hypertrophic Cardiomyopathy(11028.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fallots tetralogy(11029.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hypoplastic Left Heart Syndrome(11030.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pulmonary atresia(11031.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Transposition of the great vessels(11032.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Left Ventricular Outflow Tract obstruction disorders(11033.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Isomerism and laterality disorders(11034.1)</td>
</tr>
</tbody>
</table>
Informatics approach to capture data according to models

Manual with expert review

Automated
OpenClinica phenotype entry

Additional terms not present in the data model can be naturally added.
Call for Expressions of Interest

The Genomics England Clinical Interpretation Partnership

Genomics England invites 'expressions of interest' from UK led consortia of clinicians, researchers, analysts and those in training to propose disease specific domains in the areas of rare inherited disease, cancer and infectious disease. The Genomics England Clinical Interpretation Partnership will lead research to enhance the clinical interpretation of whole genome sequences and support the delivery of healthcare transformation from the 100,000 Genomes Project.

This will be the route by which Genomics England will engage with the UK academic and healthcare community and their international collaborators to discover new biological insights into disease, elucidate functional impact, develop novel analytical approaches and create high cadre expertise in genomic medicine.

The overall aim is for the Genomics England Clinical Interpretation Partnership to create thriving, sustainable communities of research and clinical (NHS) disease experts to interrogate the 100,000 whole genome sequences. The domains within Genomics England Clinical Interpretation Partnership will have three primary roles:

- **Research**: Harnessing opportunities for research and discovery enabled by the 100,000 Genomes Project with the intention of further enhancing our understanding of genomic medicine and its application in healthcare.
- **Clinical Interpretation**: Provision of disease-specific expertise in clinical reporting and variant interpretation to enhance interpretation of 100,000 Genomes Project data to ensure feedback of the highest calibre data to treating clinicians in order to inform diagnostics and treatment decisions.
- **Training**: Training of researchers and clinicians.

Expressions of interest are invited from self-organised consortia to form domains with a UK lead (clinical or non-clinical) with a connection through a higher education institute or the National Health Service to UK healthcare. Each consortium must clearly create a multidisciplinary clinical, academic and training domain which offers high calibre skillsets.

We encourage the UK led domains to involve key international collaborators.

The successful Clinical Interpretation Partner Domains will be given free access, subject to our Data Access and Acceptable Use Policy, to embassies within the Genomics England Data Centre which has been funded by the Medical Research Council.

We have scheduled an open meeting to facilitate further discussion. This will be held on the 5th December at the Wellcome Trust, 215 Euston Road, London, NW1 2BE, 4-6pm.

For further information and guidance on how to submit expressions of interest please visit www.genomicsengland.co.uk.

You can also call or email us chiefscientific@genomicsengland.co.uk / 020 7882 3402.

**Closing date for expressions of interest: Monday 26th January 2015 at 5.00pm**

Announcement of Genomics England Clinical Interpretation Partnership Domains will be in February 2015.

www.genomicsengland.co.uk

---

### GeCIP domains

**Function-specific and disease-specific domains in:**

- Cancer
  - Breast
  - Colorectal
  - Lung
  - Ovarian
  - Prostate
- Rare Diseases
  - Hearing and Sight
  - Cardiovascular disease
  - Respiratory
  - Endocrine and metabolism
  - Haematological Malignancies
  - Gastroenterology
  - Immunological diseases
  - Neurological and degenerative diseases
  - Musculoskeletal
  - Interpretation, Validation and Feedback
  - Skin
  - Renal
- Ethics and Social Sciences
  - Non-malignant Haematology
  - Inherited Cancers
  - Paediatric
- Advanced analytical methodologies
  - Rheumatology
  - Pathogen WGS (HIV, Hep C, TB, AMR)
  - Severe response to infection
Genomics England Clinical Interpretation Partnership - GECIP

Goals
• Drive up the fidelity of clinical interpretation of genome sequencing
• Foster the use of the programme’s data
• Accelerate academic/industry partnership and development of diagnostics and therapies.

Composition
• UK-led and organised into domains
• Self proposed partnership between researchers, the NHS and Trainees with skills.
• Can bring international collaborators

Expectations:
• All data generated contributes to the Genomics England Dataset and are available to all inside a GeCIP domain.
• IP owned by Genomics England but readily licensed to incentivise active collaboration
• Training workstreams
GeCIP Submission metrics

- Over 2000 CVs submitted with final 88 valid applications
- Over 2500 members and associated members credited (CVs not yet submitted)
- Over 4500 including trainees, students potential researchers cited
- Over 500 international collaborators and collaborations declared
- The metrics below show % of domain with the desired characteristics

<table>
<thead>
<tr>
<th>Declared interest</th>
<th>Training</th>
<th>Academic</th>
<th>Clinical</th>
<th>Clinical Scientists</th>
<th>Bioinformatics</th>
<th>International Collaborations</th>
<th>Research</th>
<th>Extant Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of submissions evidencing criteria (out of 78 final valid applications)</td>
<td>89%</td>
<td>94%</td>
<td>78%</td>
<td>83%</td>
<td>74%</td>
<td>62%</td>
<td>96%</td>
<td>71%</td>
</tr>
</tbody>
</table>
GeCIP Domains – 1st wave

- 8 proposed domains for cancer derived from 26 submissions: ovarian, lung, breast, etc
- 14 domains for rare disease comprising 21 submissions: cardiovascular, neurological, paediatrics, etc
- 10 functional and cross cutting domains comprising 24 submissions. Population genomics, variant interpretation, education and training/primary care, etc
- 1 Ethics, Law and Social care domain comprising 13 submissions.
- Clinical Interpretation, Validation and Feedback (V&F) domain: “operations” arm of GECIP to coordinate clinical interpretation
Gene Consortium launch partners

• AbbVie
• Alexion Pharmaceuticals
• AstraZeneca
• Biogen
• Dimension Therapeutics
• GSK
• Helomics
• Roche
• Takeda
• UCB*
Education in Genomic Medicine

• **Workforce development** in Genomic Medicine:
  • specialist genetics, pathology and specialist clinical workforce

• **MSc** in Genomic Medicine
  • CPD access to MSc modules for specialist practitioners

• **Specialized scientific training**: fellowships funded for 3/5 years:
  • Molecular Pathology including Infections and Pathogens
  • Genetics / genomics
  • Bioinformatics

• **Specialist on-line learning**

• **Bioinformatics workshops**
Genomics England

• 100,000 WGS of NHS patients
• Working with NHS, academics and industry to drive Genomic Medicine into the NHS
• Support that with education
• Leave a legacy of NGS Centres, sample pipeline and biorepository, large-scale data store that makes this usable by the NHS
• New diagnostics and therapies and opportunities for patients
• By end of 2017
Genomics England and beyond

• Engine for NHS transformation through GMC contracts

• Rare and cancer data models driving standardisation of secondary care data capture

• Cancer tumour sample collection requirements driving change in pathology services

• Acceptance of dual purpose data centre to securely support clinical interpretation services and research
Acknowledgements

Special thanks

• Cambridge, UCLH, GOSH, Moorfields, Newcastle, Manchester, Guys and St Thomas’s, Oxford, Liverpool, Sheffield, Leeds, Birmingham, Royal Marsden, Southampton, UK CLL Consortium, CRUK, RCPath, NHS England, Department of Health, Biobank UK, Sanger, EBI, KCL, UCL and QMUL

All Genomics England Teams:

• Science, Operations, Informatics, Bioinformatics, Programmes, Communications, Administrative Support

All advisory committees and working groups:

• Science, ethics, data, cancer, rare diseases, molecular pathology