Genomics England and the 100,000 Genomes Cancer Project

J Louise Jones
Lead for Molecular Pathology, Genomics England
Professor of Pathology, Barts
About the 100K Genome Project

We are a new company set up by the Department of Health to help deliver the 100k Genome Project first announced by the Prime Minister David Cameron in December 2012.

This project will sequence the personal DNA code – known as a genome – of up to 100,000 patients over the next five years. This unrivalled knowledge will help doctors’ understanding, leading to better and earlier diagnosis and personalised care. Based on expert scientific advice, we will start by tackling cancer, rare diseases and infectious diseases.

The company will manage contracts for sequencing, data linkage and analysis, and set standards for patient consent.

“**The UK will become the first ever country to introduce this technology in its mainstream health system.**”

Genomics England was announced by Jeremy Hunt, Secretary of State for Health, as part of the NHS 65th birthday celebrations on 5 July 2013.

He said: “The NHS has a long track record as a leader in medical science advances and it must continue to push the boundaries by unlocking the power of DNA data.”

“The UK will become the first ever country to introduce this technology in its mainstream health system – leading the global race for better tests, better drugs and above all better, more personalised care to save lives.

“Genomics England will provide the investment and leadership needed to dramatically increase the use of this technology and drive down costs.”
Genomics England – the mission

• **100,000 WGS** on NHS patients
  - Rare diseases
  - Cancer
  - [and pathogens]

• Working with NHS, academics and industry to make the UK a *world leader in Genomic Medicine*

• **Transformation** of NHS so that NGS can become routine investigation

• Generate **health and wealth**

• Leave a **legacy** of infrastructure, human capacity and capability
Rationale

For Whole Genome Sequencing
<table>
<thead>
<tr>
<th>Data Type</th>
<th>Coding variants</th>
<th>Non-coding variants</th>
<th>Copy Number variants</th>
<th>Large Structural variants</th>
<th>Balanced Translocations</th>
<th>Uniparental disomy</th>
<th>Consanguinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted gene sequencing</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Whole Genome</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SNP arrays</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Array CGH</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

The whole genome is the best way at getting at all the different types of mutations and changes in the DNA that can cause disease.
Genomics England and 100,000 Genomes programme

Overview, establishing the structures, challenges and progress
Delivering Sequencing at scale

- Delivery at scale: 700 WGS per week
- **Wellcome Trust** awarded £27m to establish NHS Genome Sequencing Centre at Hinxton
- Partnership with **Illumina**
- Installing huge structural capacity (**X10 sequencers**) for sequencing at scale
- Sequencing **protocols** for different materials eg FFPE-derived DNA vs FF tissue vs blood
Collecting and handling samples

• Processing of multiple sample types from routine NHS clinics
• Collection of multiple samples from families (disparate locations)
• Collection of tumour material
• Collection (processing) of different –omics samples
• DNA extraction
• Barcoding, handling, identification
• Secure transport and transfer
• NIHR National Biosample Centre (Milton Keynes) to store samples
Massive data requirements

- **MRC** award to establish the infrastructure for UK Genomic Medicine Data Architecture (£24m)
- Data collection
- Data storage, transfer
- Data processing
- Data analysis, interpretation
- Data Clinical reporting
- Return of clinical-grade results
- Data security and Information Governance
Managing sample and data flow

Clinical Genetics, Oncology services, Public Health, NHS Trusts, Patients & Public

→ Broad consent, clinical data, molecular pathology data and samples

Clinical Data Infrastructure
Identifiable Clinical Data
Whole Genome Sequence
Life-course data

Clinical Data Infrastructure

Identifiable Clinical Data
Whole Genome Sequence
Life-course data

Research Data Infrastructure
Sequential builds of pseudonymised data and WGS
Safe haven reading library of data

• Primary Care
• Hospital episodes
• Cancer Registries
• Rare Disease registries
• Infectious Disease
• Mortality data
• Patient entry

Genomics England

Fire wall
Patient data stays in safe haven

Only processed results pass outside

Clinicians & Academics
Training & capacity
Industry

Annotation & QC Scientists & SMEs
Product comparison
Complicated Ethical considerations

Ethics Advisory Group

Protocol: Research Ethics Committee approval

• NHS Genomic Medicine Consent: multiple types/ages
• Life Course Refreshable Dataset
• Return of secondary findings
• Release of data to third parties – academics, clinicians and industry
• Recall of patients for additional data and samples
• Key focus on Patient protection issues: insurance, mental capacity
Central Partners in the Program
June 2014: Invitation to tender as Genomics Medicine Centres

Delivery of:
1. Appropriate patient groups
2. High quality phenotype data
3. Quality DNA
4. Consent
5. Validation of findings
6. Return of results
Genomics England Genomic Medicine Centres

Announced December 2014

1. **East of England NHS GMC** Led by Cambridge University Hospitals NHS Foundation Trust;
2. **South London NHS GMC** Led by Guy’s and St Thomas’ NHS Foundation Trust.
3. **North West Coast NHS GMC** Led by Liverpool Women’s NHS Foundation Trust.
4. **Greater Manchester NHS GMC** Led by Central Manchester University Hospitals NHS Foundation Trust
5. **University College London Partners NHS GMC** Led by Great Ormond Street Hospital NHS Foundation Trust
6. **North East and North Cumbria NHS GMC** Led by The Newcastle upon Tyne Hospitals NHS Foundation Trust.
7. **Oxford NHS GMC** Led by Oxford University Hospitals Foundation Trust.
8. **South West Peninsula NHS GMC** Led by Royal Devon & Exeter NHS Foundation Trust.
9. **Wessex NHS GMC**. Led by University Hospital Southampton NHS Foundation Trust.
10. **Imperial College Health Partners NHS GMC**. Led by Imperial College Healthcare NHS Trust.
11. **West Midlands NHS GMC**. Led by University Hospitals Birmingham NHS Foundation Trust.
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.
100KGP– involvement of GeCIP

NHS clinical patients
Management: Standard Care/Trials

Samples Clinical Data Consent

NHS clinical patients
Management: Standard Care/Trials

Samples Clinical Data Consent

Established Scientific Corpus (Literature, databases)

Interpretation

Called variants

GEL Sequencing/Annotation

Multi-omics

Linked patient data

Data repository

Bio-repository

Academic

Industry

GeCIP adds value
GeCIP Domains for cancers: initial overview

Six main tumour groups

- **Prostate**: Prof Johan De Bono
- **Lung**: Prof Charlie Swanton
- **Breast**: Dr Nick Turner
- **Ovarian**: Dr James Brenton
- **Colorectal**: Prof Ian Tomlinson
- **Haematological Malignancies**: Prof Anna Schuh

- Embedded in NCRI CSG
- Strong clinical representation: including (11) GMCs
- Strong academic representation
HEE: Education in Genomic Medicine

- **Workforce development in Genomic Medicine:**
  - specialist genetics, pathology and specialist clinical workforce
- **Specialized scientific training:** fellowships funded over 3/5 years:
  - Molecular Pathology including Infections and Pathogens
  - Genetics / genomics
  - Bioinformatics
- **MSc in Genomic Medicine**
  - CPD access to MSc modules for specialist practitioners
- **Specialist on-line learning**
- **Bioinformatics workshops**
Partners from Industry: GENE consortium

• Partners in GENE consortium may access 100KGP data
• Partners engage in development of 100KGP program
• Develop interfaces with clinical academic community
• Contact: Paul Jones
Genomics England
Cancer Programme
Cancer tumour sequencing: Background progress to date

• Comprehensive catalogue of all somatic mutations in all tumour types with clinical data
  • TCGA, ICGC, MSK bioportal

• Tumours are genomically heterogeneous

• Genomic architecture: multiple genes with low frequency mutations

• “Noise” from passenger mutations

• Exome vs Genome (total WGS in TCGA: 803, ICGC: 1649)

• Analysis required within and between tumour groups: new molecular based taxonomies

• Analysis against clinical parameters/outcomes required

• Larger series are required, with enhanced clinical characterisation
100,000 Genomes Project: real opportunities

• Routine collection of samples through NHS clinics
• **Single consent** for clinical diagnostics and research
• Collection at same time of **additional samples** for:
  • Transcriptomics, proteomics
  • cfDNA
• **Serial samples in time**, Multiple samples in space
• Collection of **high quality clinical data**
  • Clinical follow-up data
  • Data linkage: life-course data
• **Re-contact patients**
Objectives of the Cancer program

1. Embedding within the NHS structures, process, expertise and skills for routine capture of high quality cancer samples and clinical data and return of genomic results
2. High quality research
3. Improve management of individual patients

Many challenges

1. Representation across tumour types
2. Sample acquisition: molecular pathology
3. Clinical data capture
Representation across tumour types

- 25,000 Tumour Normal pairs
- Common Solid Cancers:
  - Breast Cancer
  - Colorectal Cancer
  - Lung Cancer
  - Prostate Cancer
  - Ovarian Cancer
- Other solid tumours:
  - Childhood solid tumours
  - Rarer cancers eg renal cell carcinoma, sarcoma
- Haematological Malignancies (CLL program within trials from CLL consortium)
Sample acquisition

• Fresh Frozen vs Formalin-fixed Paraffin-Embedded (FFPE)

• Minimal data exist on WGS from FFPE samples

• Cancer Pilots (multi-site)

• Illumina

• Commercial partners

• NeQAS (National Quality Assurance Scheme)

• Joint Molecular Pathology working group

• Development of optimised protocols for FFPE and FF: application likely tumour-specific
Clinical data capture

• Rich dataset capturing life-course data of patient and tumour

• Predominantly via common formats/data items already collected for routine cancer data collection (COSD, SACT, RTDS)

• Working with NCRS/NCIN to optimise data capture
Clinical data capture
Opportunities: large-scale sequencing in cancer patients as a routine test in the NHS

• Genomics as routine test in cancer:
  • massive expansion achievable through scale

• Clinical infrastructure: working with the NHS
  • High quality structures for data capture
  • Driving advances in molecular pathology: cancer sample processing
  • Returning and applying results: Tumour Sequencing Boards

• Data unification (multi-source)

• Efficiency between clinical/research/industry with streamlining of discovery/translation
  • Universal structure for consent, samples, clinical data
  • Cross-talk between clinicians and academics
  • Shared methodologies, collaborative analytical infrastructure
Put genomic medicine right into the National Health Service; have the National Health Service as the central hub for genomic research
Genomics England – who are we?

**Officers of Genomics England**
- Sir John Chisholm (Executive Chair)
- Prof Mark Caulfield (Chief Scientist)
- Nick Maltby (Company Secretary)
- Prof Jim Davies (Informatics)
- Viv Parry (Outreach)

**Board of Genomics England**
- Prof Dame Sally Davies (CMO)
- Kevin Dean
- Prof Sir John Bell
- Jon Symonds (Audit)
- Prof Sir Malcom Grant (NHSE)

**Science and Clinical Team**
- Prof Mark Caulfield
- Dr Tom Fowler
- Dr Clare Turnbull
- Dr Jeanna Mahon Pearson
- Prof Louise Jones
- Dr Ellen Thomas
- Ms Jo Whittaker
- Dr Richard Scott

**Advisory Committees**
- **Science**: Sir John Bell
- **Informatics**: Kevin Dean
- **Ethics**: Prof Mike Parker