100,000 Genomes & Genomics England

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

Global Leaders in Genomic Medicine
Washington 8-9th January 2014
UK Health System 101

• Four separate health services
  – NHS England
  – NHS Wales
  – NHS Scotland
  – Health & Social Care in Northern Ireland (HSC)
• NHS (England)
  – ~1.4 million employees
  – ~£110 billion annual budget
• Structure in England changed 1\textsuperscript{st} April 2013
Report marks progress in first year of Dementia Challenge

The Dementia Challenge champion groups have produced a report on their progress since the challenge was launched in March 2012.

The health and care system helps people lead healthier lives, recover well from illness and live better for longer.

National partners make commitment to join up health and social care

Care and Support Minister Norman Lamb has launched plans to join up health and social care.
Linking Health data to Research

Clinical Data World

Electronic Health Record

Healthcare Professional

Genotype
Whole Genome Sequencing

Genomic Biology Data World

Reference genome sequence
~3 gigabytes

EBI: repositories
(petabytes of genome sequence data)

Sanger: sequencing
(1000 genomes, uk10K)

Human sequence data repositories

Genotype and Phenotype relationship capture

Phenotype
Electronic Health Records

Electronic Health Record
Steps in UK towards E-Health Research, Genomic Medicine

• Health data to Research
  – 2006 Creation of OSCHR
    • Increase coordination between funders: MRC and NIHR
  – 2007 OSCHR E-health board
    • Enable research access to UK EHR data
    • Build capacity for research on EHR data

• Genomics to Health
  – 2009 House of Lords report on Genomic Medicine
  – 2010 Creation of Human Genomic Strategy Group (HGSG)
2011: UK Life Sciences Strategy

Linking Health data to Research

Clinical Data World

Electronic Health Record

Healthcare Professional

Phenotype

Electronic Health Records

Clinical Practice

Research

Datalink (CRPD)

Farr Institute

Genomic Biology Data World

Human sequence data repositories

Genotype and Phenotype relationship capture

EBI: repositories (petabytes of genome sequence data)

Sanger: sequencing (1000 genomes, uk10K)

Reference genome sequence ~3 gigabytes
2012: Human Genome Strategy Group report
UK Life Science Strategy Update; 100K Genomes

DH: http://www.dh.gov.uk/health/2012/01/genomics/
BIS: http://www.gov.uk/office-for-life-sciences/
Genomics England launched, mapping DNA to better understand cancer, rare and infectious diseases
Linking Health data to Research

Clinical Data World

Electronic Health Record

Healthcare Professional

Phenotype

Electronic Health Records

Genomics England

EBI: repositories (petabytes of genome sequence data)

Sanger: sequencing (1000 genomes, uk10K)

Reference genome sequence ~3 gigabytes

Human sequence data repositories

Genotype and Phenotype relationship capture

Clinical Practice Research Datalink (CRPD)

Farr Institute

Clinical Data World

Genomic Biology Data World

Electronic Health Record

Linking Health data to Research

Clinical Data World

Electronic Health Record

Healthcare Professional

Phenotype

Electronic Health Records

Genomics England

EBI: repositories (petabytes of genome sequence data)

Sanger: sequencing (1000 genomes, uk10K)

Reference genome sequence ~3 gigabytes

Human sequence data repositories

Genotype and Phenotype relationship capture

Clinical Practice Research Datalink (CRPD)

Farr Institute

Clinical Data World

Genomic Biology Data World
Genomics England - mission

- **100,000 patients with rare inherited disease, common cancers and pathogens from the NHS in England**
- Whole Genome Sequencing
- Generate improved health and wealth for UK
- Legacy of infrastructure, human capacity and capability
- Become World-leader in Healthcare application of Genomic Medicine
- £100m funding over the next 5 years
Scale compared to existing WGS

- 1000 genomes and UK10K
  - low coverage genomes (~4x illumina)
- Limited number of ‘clinical grade’ WGS
  - TCGA: ~700
  - ICGC: ~700
  - WGS 500: 500
Is now the moment to commit to WGS

Cost per Genome

- Large-scale structural changes
- Balanced translocations
- Distant consanguinity
- Uniparental disomy
- Novel/known coding variants
- Novel/known non-coding variants

Data Type

- Targeted gene sequencing
- SNP array
- Array CGH
- Exome
- Whole Genome

Genomics England
Rare inherited diseases

- 7% of the population or about 5/10,000 people
- 7000 rare disorders - disabling, shorten life, costly
- Circa 85% have a single gene defect
- Early knowledge may avoid disability
- Testing for >700 disorders extant within the NHS diagnostic laboratory network (UKGTN)
- Represents <1/4 of known disease genes.
- Whole Genome Sequencing 25-50% increase in discovery
Genomics England
Will look for mutual wins

• NIHR Translational Research Collaborative
• NHS Clinical Genetics Service & Organ Based Specialists
• WGS500 - Oxford
• International Rare Diseases Consortium
  – Aiming for 200 new treatments
• Deciphering Developmental Disorders (Exome)

• Link to trial opportunities for new therapies
• Increased recognition from industry of the value of niche markets
Cancer

- Lung Cancer - 40,000 cases/year in the UK, (35K die/year)
- Largest cause of cancer death, therapies modestly effective only applicable to 10-15% of patients
- CRUK Stratified Medicine’s initiative

- Other Cancers - Breast, colon, prostate and unknown primary

- Rare and Childhood Cancers

- Drugs target mutations
- Tumour heterogeneity
Incidence v survival at 5 years
Pathogens

- Stratifying response, minimising adverse events and tracking outbreaks
- HIV – Treatment for life and resistance testing is in the care pathway.
- Hepatitis C genotype selects therapy
- M. Tuberculosis resistance and epidemiology
Clinical Genetics, Cancer, Public Health, NHS Trusts, Patients & Public
Rare diseases, common cancers and pathogens
Broad consent, characteristics, genetic data capture and samples

DNA repository

Sequencing Centres

Refreshable identifiable
Clinical Data
Life-course registry
Linked to anonymised
Whole Genome Sequence

Primary Care
Hospital episodes
Mortality data
Patient entry

Annotation & QC
Scientists & SMEs
Product comparison

Sequential builds of refreshed clinical grade
Anonymised Clinical data and DNA sequence
Safe haven- users work within

Clinicians & Academics
Training & capacity
Industry

Fire wall
Patient data stays on NHS side

Only processed results pass outside
Genomics England – Implementation Plan

• Phase 1: bake-offs
  – Sequencing comparison underway
  – Annotation comparison to follow

• Phase 2: Pilots
  – 2000 Rare Inherited Disease WGS- 30x depth –January 2014
  – 3000 Cancer Patients (Lung, Breast & Colon)
  – Each 1000 somatic (50x) and 1000 germline (30x) – tender imminent
    – Pathogens pilot will be planned with Public Health England

• Phase 3: Main study
  – 30,000 WGS per year

• Education
  – Developing a National Programme to transform capacity and capability
  – UK Universities and Medical Schools
Process Overview

Sample DNA → Sequence (BAM) → Variants (VCF) → Candidate Variants → Clinical Interpretation → Clinical Action
Sequencing assessment

• Bake off in progress – samples with suppliers
• Evaluation will be on quality and coverage
Annotation assessment

• Harder than assessing sequencing
• Gold standard less well defined
• Lack of established data standards
Past assessment exercises

- CASP – Critical Assessment of Structure Prediction (since 1994, CASP11 in 2014)
- GASP, RGASP – Gene prediction and RNAseq assessments
- CLARITY Challenge – 2012
  - http://genes.childrenshospital.org/
  - https://genomeinterpretation.org/
Data provided by GeL

• Sequence from providers (BAM+VCF)
  – Rare diseases: trio
  – Cancer: germline + tumour

• Phenotype data available to clinicians
Types of annotation anticipated

- Filtered, ranked lists of variants with estimates of pathogenicity and confidence
- Expected impact at level of genes, pathway
- Tools organising literature around affected genes, pathways
- Clear, simple clinical reports
- Suggested clinical interventions
Assessment criteria

• Accuracy
• Clinically informative
• Rapid turnaround
• Understandable output
• Standardised output data formats
• Ability to operate at scale
  – 2014: ~25 samples/day (pilot)
  – 2015: ~50 samples/day
  – 2017: ~100 samples/day
Initial annotation assessment

- Information collection exercise on 15 samples
  - Investigate levels of annotation available

- Minimal file format requirements
  - Will inform future specifications for file formats

- Will select multiple suppliers for pilot
  - Ability to deliver timely, consistent data, etc.
Ongoing assessment during Phase 2

• Precise file format requirements, with optional sections

• Best suppliers will be invited to tender to provide annotation for main programme (2015-2017)
Annotation expectations for Phase 3

• Software will run as Virtual Machines within GeL datacentre, c.f. Apps

• Software will be subject to evolving compliance requirements, c.f. CLIA dry lab

• No need for ‘Apps’ to be comprehensive: potential for specialist software, e.g. specific diseases, pharmacogenomics etc.
Genomics England

- 100,000 WGS on NHS patients and pathogens
- Aware of the challenges
- 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
Acknowledgements

Genomics England
NHS England Genome Strategy Board
Department of Health Human Genome Strategy Group

Wellcome Trust Sanger Institute
Discussions with many at Wellcome Trust, OSCHR, NIH, ENCODE