The 100,000 Genomes Project: Infrastructure, Security, Ethics

Professor Dame Sue Hill  @CSOsue
Chief Scientific Officer for England

May 2018

NHS England and NHS Improvement
Harnessing and securing the power to transform healthcare

Outcomes: Genomics and other data-driven medicine have the power to significantly improve human health and to inform more effective and efficient health systems.

Comparison: At the heart of this is a recognition that the ability to provide the best care for patients will be significantly enhanced by comparing their data with others – with the infrastructure, ethical & cybersecurity issues that this brings.

Discovery: Alongside immediate care, a crucial role for this compiled and compared data is the ability to inform discovery of new treatments & care approaches across academia and industry – bringing a longer-term dimension to the use and security of data.
Challenges in delivery

• Delivering the full power of this new technology raises two important challenges

  – Ensuring the technical – and human – infrastructure is there to derive, process and analyse results to a clinically useful timescale, and that the system is accurate and reliable

  – Recognising the power of the data generated raises significant ethical issues for individuals (who may find out things about themselves or their family they may not expect to know) and more broadly around the access, ownership, control, security and monetarisation of sensitive personal information
The 100,000 Genomes Project: How it worked

Set up by the UK Government in 2012 to sequence 100,000 genomes from patients with rare and inherited disease

**KEY PRINCIPLES**

1. WGS extends current diagnostic scope
2. Recruitment from routine care, treated through routine channels
3. Participants consent to sharing of de-identified data for multiple purposes (R&D & industry use & longitudinal access)
4. Establishes model for transformational change
Genomics England Managed Infrastructure

- Procured NIHR Biosample Centre – new protocols with requirement for QC and plating of samples
- Procured NHS Genomic Sequencing Centre- Hinxton - 15 of the highest throughput and calibre sequencing machines working for the NHS
- Inward investment by Illumina in a new European HQ- £60m
- Created Genomics England data handling, processing & research environment (investment from multiple sources)
- Developed panel app – curating the world’s best panels for NHS Care and allowing clinician choice of genes to be applied
- Developed a semi-automated WGS analytical pipeline
- Established the Genomics England Clinical Interpretation Partnership – over 3000 researchers from 24 countries – drive up diagnoses for the NHS
- Established pre-competitive discovery forum helping to shape opportunities to bring new medicines and diagnostics to the NHS
- Scotland, Northern Ireland and Wales joined with MRC funds
NHS England service infrastructure - 13 NHS Genomic Medicine Centres

- NHS England procured **13 NHS Genomic Medicine Centres** to provide population-based networks to bring together local hospitals and outreach services as delivery partners with each GMC covering 3-7m people and delivering against defined service specification for the project contribution.

- NHS England established a **Genomics Implementation Unit** to oversee performance and to provide improvement support for embedding changes.

- **13 Lead GMC organisations** were responsible across the network and **delivery partners** for clinical leadership, informatics developments & data collation, sample collection and processing, tissue handling, DNA extraction, establishing MDTs and local arrangements for WGS validation and feedback to patients.

- **Proactive patient & public involvement & awareness raising initiatives**

- **Workforce development & service development** for mainstreaming.

- **Identification of suitable patients from routine care**

- **Involvement of patients in ethics, data & consent issues**

- **Supply of high-quality processed samples**

- **Collection of linked phenotypic and clinical data**

- **Validation of WGS findings and feedback to patients**
Critical developments in data management

- Established systems for national **sample tracking & tracing** (e.g., GS1 barcoding)
- Developed a consent **model** for processing of information for multiple purposes
- Established a **WGS analytical pipeline** to interface with NHS (provided by Genomics England)
- Developed a rich clinical **dataset** with standard nomenclature and **data models** (HPO, SNOMED CT)
- Created **NHS informatics and data infrastructure** for genomics - bringing multiple systems together to **collate and curate** individual clinical data
- Built understanding of fundamental need for data sharing – public & professional
- Established national **database** of consented and deidentified genomic (WGS) and clinical data
- Established national database of consented and deidentified genomic (WGS) and clinical data
How the Project fits together – sample & data infrastructure

- **Registry (PHE)**
- **HES (HSCIC)**
- **Other NHS Clinical Data**

- **Biobank sample**
  - *Inc Omics samples*

- **Sequencing (Illumina)**

- **DNA**
  - **BAM/VCF**

- **Phenotype**

- **NHS Genomic Medicine Centres**
- **Genomics England**
- **Genome Interpretation Service Companies**
- **GeCIP/GENE Embassies**
- **Data Centre**
  - **Clinical Interpretation Services**

- **NHS Firewall**
- **Discovery Forum**

- **Controlled access – a reading library not a lending library**

- **System has ongoing security assessment inc regular penetration testing**

- **Longitudinal clinical data**

- **Patients & frontline services**
Genomics England Analysis pipeline

DNA

Phenotypes & Pedigree

Gene Panels

Variant filtering

GeCIP(s)

Clinical assessment

Validation Outcomes

Patient/family

Gene Panels

Report QC

Reporting tool

Review

Annotation Companies

Genome sequence

Annotated VCFs

Tiered variants

QC

Patient/family
Reporting back to the NHS

1. View family pedigree
2. Review variants and close case
3. Download
Multi-sided platform

Genomics platform
- Scalability
- Governance
- Regulation
- Openness
- Collaboration

Precision medicines

Real world data

Delivery (health system)
- Delivery networks
  - GLHs

Industry networks
- Discovery Forum

Development (industry)

Research (academia)
- Research insight
- Impact, funding

Academic networks
- (GeCIPs)

Primary data

Discovery Forum

Impact, funding

Precision medicines

Real world data

Delivery (health system)
- Delivery networks
  - GLHs
Progress to date

**Samples**

- 120,500 Samples collected from NHS GMCs and sent to biorepository
  - Cancer: 36,859
  - Rare disease: 83,641

**Genomes**

- 109,072 Genomes sequenced
  - Cancer: 24,782
  - Rare disease: 84,290

**Analysis and Results**

- Results for 76,527 genomes sent to NHS GMCs

**Research**

- Data release of genomes in RE: 91,271
- GeCIP members with access to RE: 2,424
- Over 175 Discovery Forum members
Key considerations: Data - Consent, control & cybersecurity

100,000 Genomes Project established clear principles around scope & operation of consent - Individuals can withdraw their consent (& data) at any time. Created an environment that protects patients' data – addressing both infrastructure & human factors with physical, electronic & human measures on data security.

Access and security – getting the balance right

Focus on maintaining privacy, confidentiality and security of data

Enabling appropriate and timely access for clinical teams, allowing data sharing and data comparison

Key structures and governance help ensure the security & integrity of data

• A reading library not a lending library – data manipulation done on Genomics England systems with strict controls on what can be exported.

• Scientific & Data Access committees provide prior scrutiny to ensure research projects are an appropriate use of data – with patients involved throughout.
Ensuring robust and ongoing cybersecurity

- **Data centres**: tested to accredited standards with robust ongoing scrutiny
- **Disaster Recovery and Business Continuity** built in throughout
- **Regular testing** as well as at upgrade and key releases, **penetration tests** from a variety of suppliers (using key standards such as OWASP)
- Incorporating **International standards** of CyberSecurity and the **latest advice** from key UK government agencies
- Continually **evolving and changing** approach to account to check for the latest vulnerabilities and protections
- Robust programme of **access control** for users with **ongoing checks** and challenges

Main data centre housed in former cold war nuclear bunker
Ethics Integral and integrated throughout

• Project design embraced the importance of thinking carefully and in an inclusive way about the ethical issues from the outset - to ensure ‘permission to operate’ in areas such as additional looked for findings, consent arrangements and patient choice, reporting plans

• Established independent Ethics Advisory Group to advise the Genomics England & recruited ethicist to provide Board level input & assurance

• Recognising the critical nature of protecting the data, through security and very careful control over who can access and use the data

• Recognising importance of establishing and maintaining public (& professional) trust and confidence – esp around involvement/ oversight/ high ethical standards

• Requires support and advice for health professionals, managers and health providing institutions e.g. hospitals
Ethics in mainstream genomics: the big issues

- **Consent**
  - Need a model of consent to support patients in making a decision for WGS – for healthcare & research
  - Needs to deal with uncertainties & openness in a WGS decision – ie broad, but genuinely informed

- **Feedback of findings**
  - What to feed back about presenting condition (*given uncertainty around VUS*)
  - Managing information beyond presenting condition (*additional findings*) – what obligation to go looking?
  - Sensible clinical approach, given resource restrictions

- **Data sharing**
  - Recognition of fundamental need for sharing for clinical care
  - Also need for sharing for research/ generating new knowledge
  - Consent can’t do all the work – need oversight mechanisms too
  - Public (& professional) trust and confidence requires appropriate governance for data access – with participant involvement

- **Equity of access**
  - Ensuring benefits are available to all
Patient involvement in the Project – National Participant Panel

Role of the Panel is to ensure the interests of participants are always at the centre of the 100,000 Genomes Project. They do this by:

- Making sure experiences of participants are at the heart of the project
- Responding to feedback
- Overseeing who should have access to participant data
Public views on use of genomic data

• Genomics England & UK Research & Innovation have carried out a detailed public dialogue on issues round genomic medicine, building on the issues highlighted in the CMO report *Generation Genome*

• Key findings:
  – **Public optimistic** about the potential of genomics, but delivery will require widespread support and engagement
  – Public clear about core principles of the NHS ‘contract’ – Reciprocity; Altruism; & solidarity (*though not familiar with that language*)
  – Genomic medicine **may change expectations** around donation of data
  – Clinicians and researchers will need ‘genomic literacy’ to support patients/donors and explain the **ever-closer relationship between research & clinical care**
  – Public have **clear limits about** how far they think **genomic data** – and information derived from it – should be used
  – Want assurances that there is a **robust governance** framework & consent process
  – **Key exclusions:** Genetic Engineering; differentiating societal groups; predictive insurance tests; targeted marketing

Data lessons from the Project for future delivery of genomic medicine

- Need to continue to populate a **UK Genomic Database** of deidentified data
- Underpin the new **Genomic Medicine Service** introducing WGS into routine care
- Need to **inform and support clinicians**’ use of genomic testing including support around selection & ordering
- Facilitating **return of results** to clinicians from sequencing/clinical interpretation
- Extraction and integration of **clinical phenotypic data** from multiplicity of sources to support genomic analysis
- Need for clear **management information** about use of testing correlated with outcomes
- **Integration** with other major systems eg EPR, LIMSx

... a new software platform to sit above NHS organisations’ systems
The new national genomic infrastructure

System Contracts & budgets, Strategic Planning, Assurance, Delivery

Engagement & networks through NHS Genomic Med Centres

- Engaged & informed patients
- Clinical genetics & other services
- Ordering clinicians

National Testing System for **all** genomics (single gene – WGS) delivered through 7 Genomic Lab Hubs

- National Genomic Test Directory – specifies tests & approach

Clinical Interpretation & decision support

- National Genomic Informatics System (NGIS)

Integrated & coordinated workforce development – HEE Genomics Education Programme

Underpinning ongoing NHS/ NIHR clinical research initiatives

Informed and shaped by patient participation & societal engagement

Controlled access to inform research & discovery

Elements delivered in partnership with Genomics England

- New patient choice & consent model – data for multiple purposes
- Ordering clinicians

Engaged & informed patients
Data flows in the Genomic Medicine Service

Test Order (with clinical data, HPO, pedigree, consent) → Sample → WGS Test

Ordering Clinician

GLH

DNA Extraction
Non-WGS Test
Sequencing and Analysis
Interpretation and Genomic MDT

Illumina Sequencing

GLH

DNA Sample Plating

NGIS

Test Order and Sample Tracking
De-identified

CVA, Variant Store, Knowledge Bases

Bio-informatics pipeline and Analysis

Outcomes

Secondary Data (NHSD, PHE)

Research Environment

Genomic Lab Hub Report

• Cases (WGS data)
• BAM and VCF files
• Variant data via CIPAPI

GLH Tools

DSS

Interpretation Portal
NHS service local integration

Integration Roadmap

Phase 1 – Test Order
- **Version 1** - Minimal dataset standard (NPEx/LIMS/NGIS)
- **Version 2** - Full (interoperability) dataset standard for WGS and non-WGS tests (includes HPO/SNOMED mapped)

Phase 2 – Test Report
- **Version 1** – Electronic flowing of standardised report format (PDF) with minimal outcomes dataset
- **Version 2** – Full standard / LHCRE & NRLS integration

Phase 3 – PGx etc…
- Electronic flowing of data to local, regional (LHCRE) and national systems - variant, polygenic risk scores, screening analysis, drug response / dosage – alerts etc…
National Genomics Informatics Service (NGIS) benefits

- Provides a framework to identify and recruit cohorts and trial-eligible patients
- Ongoing record of research and diagnostic genomic analyses of each individual
- Longitudinal life course data from national data sources held alongside data provided by researchers, clinicians & patients
- Integration with external systems including hospital LIMS via APIs and standardized message formats e.g. FHIR
- Readily configurable clinical data capture for rapid adaptation for new cohorts or new data
- Genomic data generated initially for research can be used for clinical care, where indicated and agreed
- Creates a shared national genomic knowledgebase where new data from research can be rapidly translated into the clinic
Real World Dataset

- Mortality data
- Hospital Episode Statistics
- Digital pathology data
- Genome data
- Enrolment data
- Biobank data
- Genetic diagnosis data
- Biobank diagnosis data
- Hospital Episode data
- Digital pathology data
- Mortality data
- Wearables? Social media?
- Primary health records...
- Secondary health records...
- Primary prescribing data...
- Secondary prescribing data...
- Enrolment data
- Biobank data
- Genetic diagnosis data
- Biobank diagnosis data
- Hospital Episode data
- Digital pathology data
- Mortality data
- Wearables? Social media?
- Primary health records...
- Secondary health records...
- Primary prescribing data...
- Secondary prescribing data...
The pros and cons of the genomic revolution

GA4GH: “We estimate that over 60 million patients will have their genome sequenced in a healthcare context by 2025”

- More tests
- More and speedier diagnoses
- More choice for patients and their families
- More decisions about management and treatment
- More decisions about research participation
- More standardisation
- More information
- More complexity
- More change
- More uncertainty

The Global Alliance for Genomics and Health aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic & health-related data sharing.
