Our history

- **December 2012**: Announced by former Prime Minister David Cameron – an Olympic Legacy
- **July 2013**: Genomics England formally launched by then Secretary of State for Health during NHS 65th Anniversary Celebrations
- **November 2016**: Former Prime Minister Theresa May opens a new Sequencing Centre
- **July 2017**: Chief Medical Officer launches Generation Genome and the Life Sciences report
- **December 2018**: Genomics England reaches goal of sequencing 100,000 genomes
- **January 2019**: Long Term Plan “an NHS where access to secure linked clinical, genomic and other data will support new medical breakthroughs and consistent quality of care”
Our vision is a world where everyone benefits from genomic healthcare.
GEL Strategy: An Infinity Loop

- Patient
  - Evolving genomic healthcare
- Researcher
  - Accelerating genomic research

Patients  Healthcare teams  Researchers
The UK ecosystem

- Academia ~4000 researchers
- BioPharma & Start-ups
- Tech (e.g., AWS, Nvidia, Lifebit, Congenica)
- Lab Tech (e.g., Illumina, Oxford Nanopore)
- Funders (e.g., Innovate, BEIS, MRC, CRUK, Wellcome, Lifearc)

8 out of 10 Tier 1 pharma access our data
## 100,000 Genomes Project Data

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Cancer</th>
<th>Rare Disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17,243</td>
<td>75,948</td>
<td>93,191</td>
</tr>
<tr>
<td><strong>Genomes</strong></td>
<td>31,208</td>
<td>75,894</td>
<td>107,102</td>
</tr>
<tr>
<td>Germline + Tumour</td>
<td>Germline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30x</td>
<td>100x</td>
<td>&lt;20% Singleton</td>
<td></td>
</tr>
<tr>
<td>+ 35K COVID</td>
<td>+ 35K COVID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Release v.15
100,000 Genomes Project Data

Genomics

- Tumour staging
- Tumour location
- Histological subtype
- Treatment regimen
- Pathology full-text
- Radiology full-text

Clinical Data

- Hospital Episode Statistics
- Mental Health Services Data Set
- Mortality data ONS

- COVID-19 status
- Exit questionnaire
- Primary Care Data (coming soon)
100,000 Genomes Project Data

Genomics

Clinical Data

Consent

Clinically accredited pipelines
for diagnostics

Lifetime follow-up
+ full retrospective data

Re-engagement
re-phenotyping
re-sampling
re-recruiting
100KGP Rare Diseases Participants

See [Cohort browser](#) for genomes count detail in research environment
100KGP Cancer participants

See Cohort browser for genomes count detail in research environment
NHS Genomic Medicine Service

WGS Cancer indications

• Wave 1: Acute Leukemias, Paediatric Tumors, Sarcomas
• Wave 2: Ovarian HGS, Triple Negative Breast, Glioma, Other Heam Onc, Various relapse & refractory

WGS Rare Disease indications

• Wave 1: 20 rare conditions
• Wave 2: +10 rare conditions

~ 10,000 participants will enter the dataset in next data release v.16
New initiatives

Transcriptomics & Proteomics in Rare Disease

Newborn sequencing

Diversity

Nanopore Sequencing in Cancer

Radiogenomics in Cancer
Rare Disease Analysis with Aggregate VCFs

Data Products

- gVCF: ~78K rare disease and cancer germline
- Tight sample QC with PCA and detailed functional annotation files (Ensembl Variant Effect Predictor, ClinVar, loss of function from loftee, population specific allele frequencies from gnomAD and ExAC, non-coding epigenetic markers from ENCODE, Roadmap and spliceAI)

Clustering of filtered results into groups such as:
- Chromosome
- Variant type
- Consequence type
- Deleteriousness
- Pathogenicity

Accelerates:
- Derived Ancestry, Allele frequencies
- GWAS, PheWAS, PRS
- AI/ML validation
Pan genomics markers in cancer

Tumour mutation burden and signatures pre-calculated in the GEL research environment

**Tumor mutation burden (TMB)**
Calculated as the number of somatic non-synonymous small variants per Mb of coding sequences. High TMB can indicate patient’s suitability for immunotherapy.

**Mutation signature**
Characteristic combinations of mutation types arising from specific mutagenesis processes. For example, smoking causes C>A transversion signature.
Pan genomics markers in cancer

Mutation burden

Breast cancer | Glioblastoma | Ovarian Cancer | Lung adenocarcinoma | Skin Cutaneous Melanoma | Colon adenocarcinoma

Mutation Signature

APOBEC & Homologous recombination | Aging | Aging & Homologous recombination | Smoking + APOBEC | UV | Aging & mismatch repair

Data Products

SNV signatures

Mismatch repair | Smoking | APOBEC | Homologous recombination | Ageing | UV

16
Feature Extraction from the Whole Genome

ER positive breast cancer
Good outcome

ER positive breast cancer
Poor outcome
Mismatch repair deficient

BRCA1 defective
Poor Outcome
Homologous Recombination deficiency

Nik-Zainal et al, Nature; 2016
Outcomes analysis from GEL RWD

PIK3CA from GEL RWD predicts Breast Cancer Survival

Impact of somatic variant in PIK3CA on the survival of HR positive, Her2 negative breast cancer patients

GEL Aligned:
• Diagnosis data (PHE)
• Death date (ONS)
• Hormone+ & HER2 status (COSD)
• PIK3CA mutation status (GEL)

GEL RWD confirms RCT Kaplan Meier curves

https://breast-cancer-research.com/content/15/4/R38

RESEARCH ARTICLE

PIK3CA mutation impact on survival in breast cancer patients and in ERα, PR and ERBB2-positive subgroups

Natalie Czakova1, Audelle Soubri2, Sophie Vacher3, Géraldine Czerny-Claria, Catherine Deroubaix, Emmanuelle Tourne, Rosette Lefevre and Ivan Bachelier*
Clinically relevant findings by patient

<table>
<thead>
<tr>
<th>Disease type</th>
<th>Disease subtype</th>
<th>Annotation group</th>
<th>Tumor type</th>
<th>Topography</th>
<th>Morphology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carcinoma of unknown primary</td>
<td>Pathology indeterminate cancer NOS non carcinoma lymphoma sarcoma etc</td>
<td>SOLID</td>
<td>Metastases</td>
<td>C7.3,TX2000</td>
<td>M80106</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>GRCh38 coordinates ref/alt allele</th>
<th>Transcript</th>
<th>CDS change and protein change</th>
<th>Population germline allele frequency (1KG</th>
<th>gnomAD)</th>
<th>VAF</th>
<th>Alt allele/total read depth</th>
<th>Gene-level actionability</th>
<th>Gene mode of action</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>17:7675077G&gt;A</td>
<td>ENTST00000269 305</td>
<td>c.535C&gt;Tp. (His179Tyr)</td>
<td>-</td>
<td>-</td>
<td>0.78 (LOH)</td>
<td>57/73</td>
<td>Trial (NSC lung ca, ca, colorectal ca, head neck SCC, ovarian ca, prostate ca)</td>
<td>Trial (NSC lung ca, breast ca, colorectal ca, esophageal SCC, head neck SCC, ovarian ca, pancreatic ca, urothelial ca)</td>
</tr>
</tbody>
</table>

**TMB in range of Colorectal Lung, Melanoma**

**C>A Transversion**
DNA damage by benzopyrene:
Smoker mutation signature

**Lung is the likely primary tumor**

**Potential lung cancer drivers:**
RB1 & NF1
Frame Shifts
Working with Genomics England

- A recallable resource of exquisitely annotated rare disease and cancer data
- A world class bioinformatics team available for collaboration
- A community of 4000 clinical-academic collaboration partners, deeply embedded in GEL data
- A gateway to the NHS ecosystem
Thank you

Parker.moss@genomicsengland.co.uk