

The 100,000 Genomes Project

MSc in Genomic Medicine

Lucy Crooks

25/1/2016

<http://www.genomicsengland.co.uk>

Genomics England

- Genomics England is a company set up in 2013
- Owned by the Department of Health
- It is managing the 100,000 Genomes Project
- Not intended to be for profit

People in the study

Rare disease

17,000 mostly children
+ two close relatives

There are 6,000 rare diseases
affecting 3 million people in UK

Cancer

25,000 patients
One sequence from tumour,
One from healthy tissue

There are 300,000 new cancer
cases a year in UK

Goals of the Project

- To start using Genomic Medicine in the NHS
- Create a huge resource for research
- Use sequence data ethically respecting patients' privacy
- Grow UK genomics industry

Sequence Data

- DNA will be extracted from blood, saliva or tissue
- Aimed at 30X coverage for rare diseases and 75X for cancer
- Samples taken in Genomic Medicine Centres (GMCs)
 - 13 including Yorkshire and Humber
- Sequenced on Illumina machines in a purpose-built facility next to the Sanger Institute
- Finished by end of 2017. 6,351 so far.. (Jan 5th)

What patient information will be recorded?

- Considerable detail requested by Genomics England input by the GMC
- Diagnosis and treatment at every hospital visit
 - Hospital Episode Statistics (HES) from the Health and Social Care Information Centre (HSCIC)

How will the data be analysed?

Alignment

Issac Genome Alignment Software

Variant calling

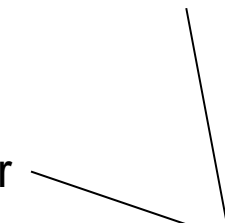
Issac Variant Caller

Illumina

Quality filtering

Identify key variant

WuxiNextcode



Comparison of Issac with other software

	Issac	BWA+GATK
Time taken (hours)	6.8	44.5
% reads mapped	93.4	95.7
% SNPs with Mendelian errors	0.139	0.126
% sites genotyped	94.5	95.8

Raczy *et al.* (2013) Bioinformatics 29:2041-2043

Who will have access to the data?

- Healthcare workers involved in patient care will have access to data on their patients
 - Downloadable clinical report
 - Only people who will see patient identifiable data
- Healthcare workers who belong to GeCIP
- Researchers with an approved project
 - Will have to sign a data access agreement and have identity checked
- Companies
 - Will be charged for access
 - Will have to sign a data access agreement and have identity checked

GeCIP

Genomics England Clinical Interpretation Partnership

- Has domains (topics)
- Healthcare workers in a domain get immediate access to relevant data
- Academic researchers in a domain **with an approved project** get immediate access to relevant data
- Academic researchers with an approved project **outside** their domain get access to relevant data after **six months**
- Industry and academic researchers not part of GeCIP **with an approved project** get access to relevant data **after six months for a fee**

Data security

- Individuals' sequence data will not leave Data Centre behind NHS firewall
- Researchers will be given secure access through a virtual machine
- Machines will have extensive software installed
- Activity will be monitored

Information Governance

- Important element of the project
- Caldicott principles and Data Protection Act apply
- The purpose for using data must be justified
- Data must be processed for the specific purpose for which it was obtained