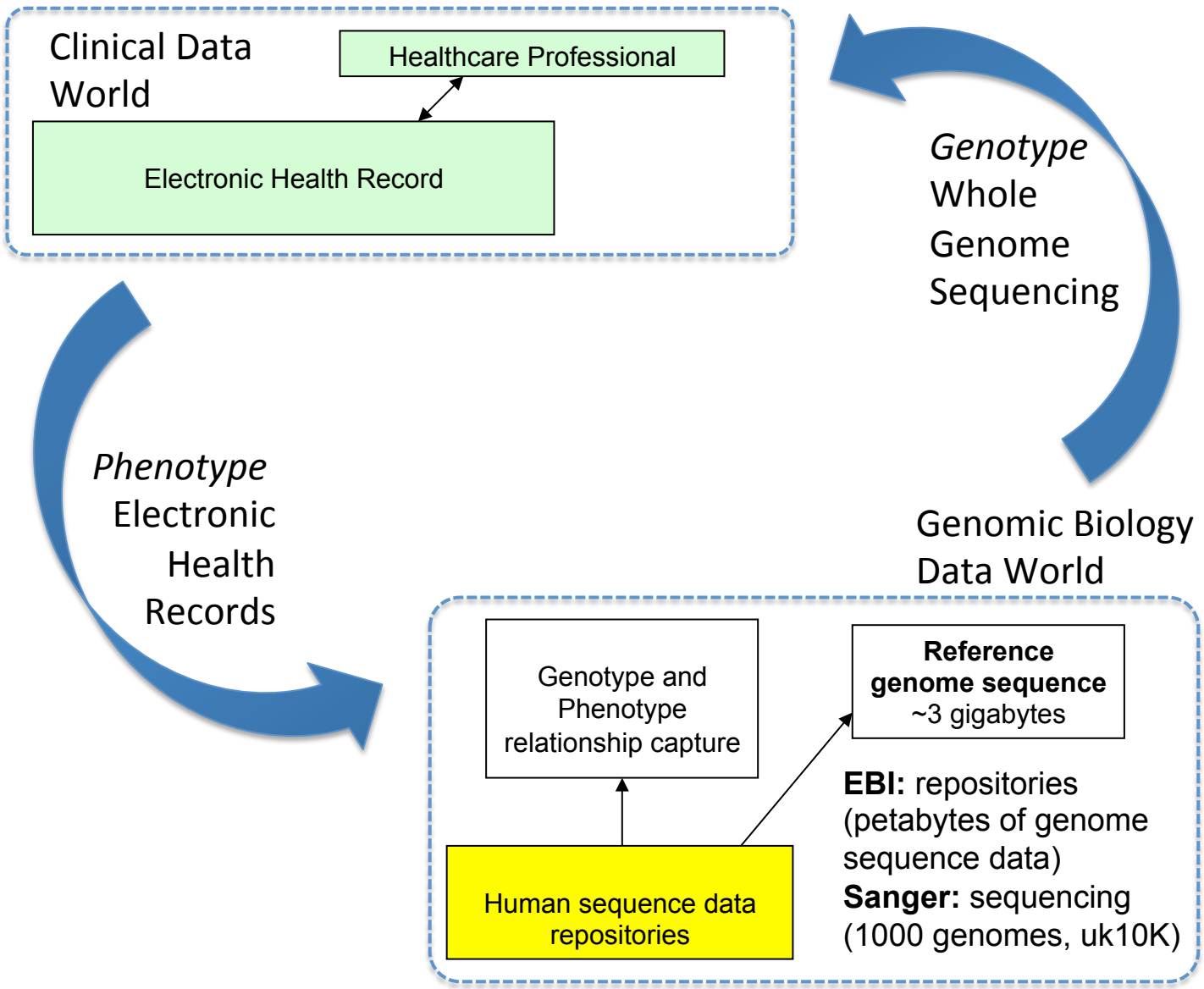
The 100,000 Genomes Project and Genomics England

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

From Systems Medicine to Personalized Health
31st March 2014, Bern

Linking Health data to Research



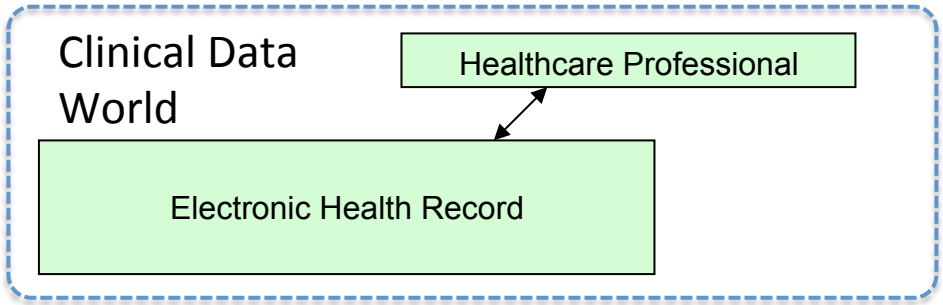
Steps in UK towards E-Health Research, Genomic Medicine

- 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

No10: http://www.number10.gov.uk/news/uk-life-sciences-get-government-cash-boost/
BIS/DH: http://www.dh.gov.uk/health/2011/12/nhs-adopting-innovation/

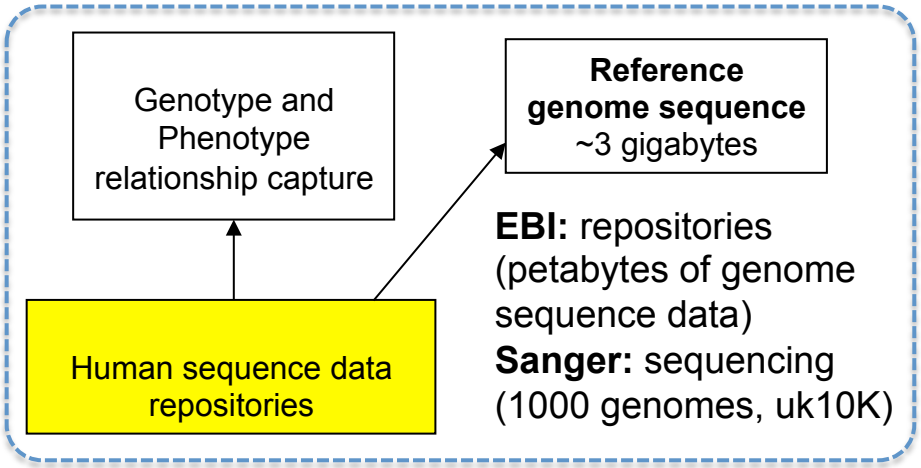
Linking Health data to Research



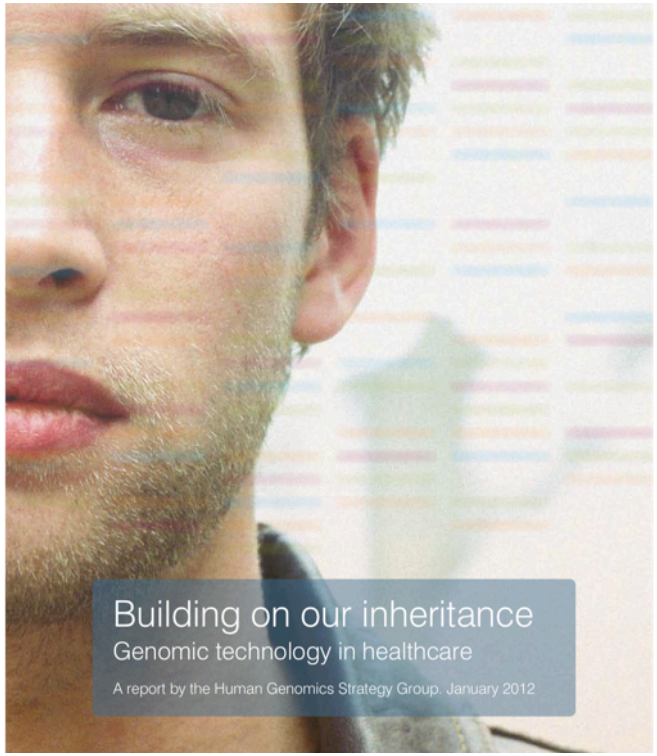
Clinical Practice Research Datalink (CRPD)

Phenotype Electronic Health Records

Farr Institute



2012: Human Genome Strategy Group report
UK Life Science Strategy Update; 100K Genomes



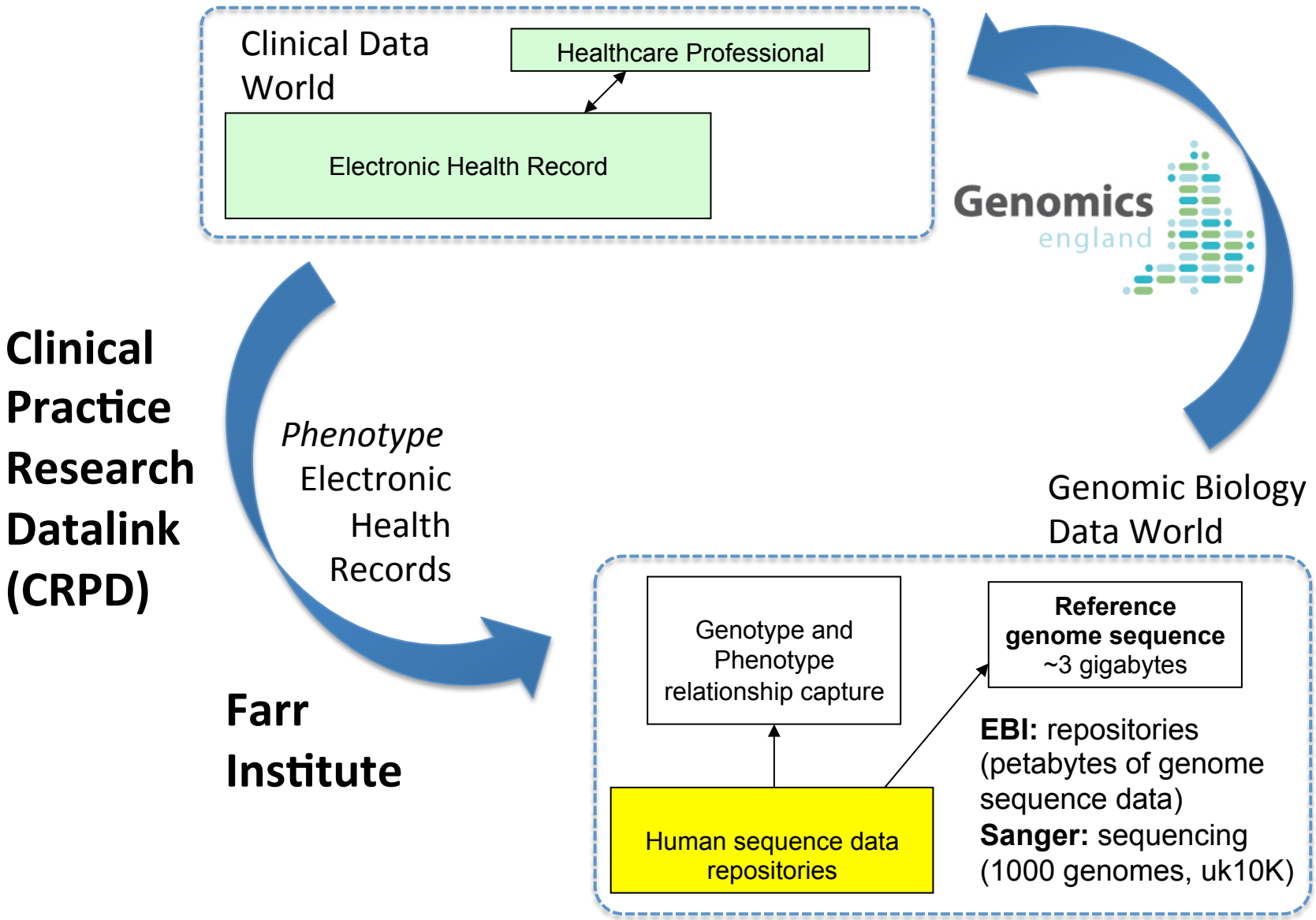
DH: http://www.dh.gov.uk/health/2012/01/genomics/
BIS: http://www.gov.uk/office-for-life-sciences/

Genomics England

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

@genomicsengland

Linking Health data to Research



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

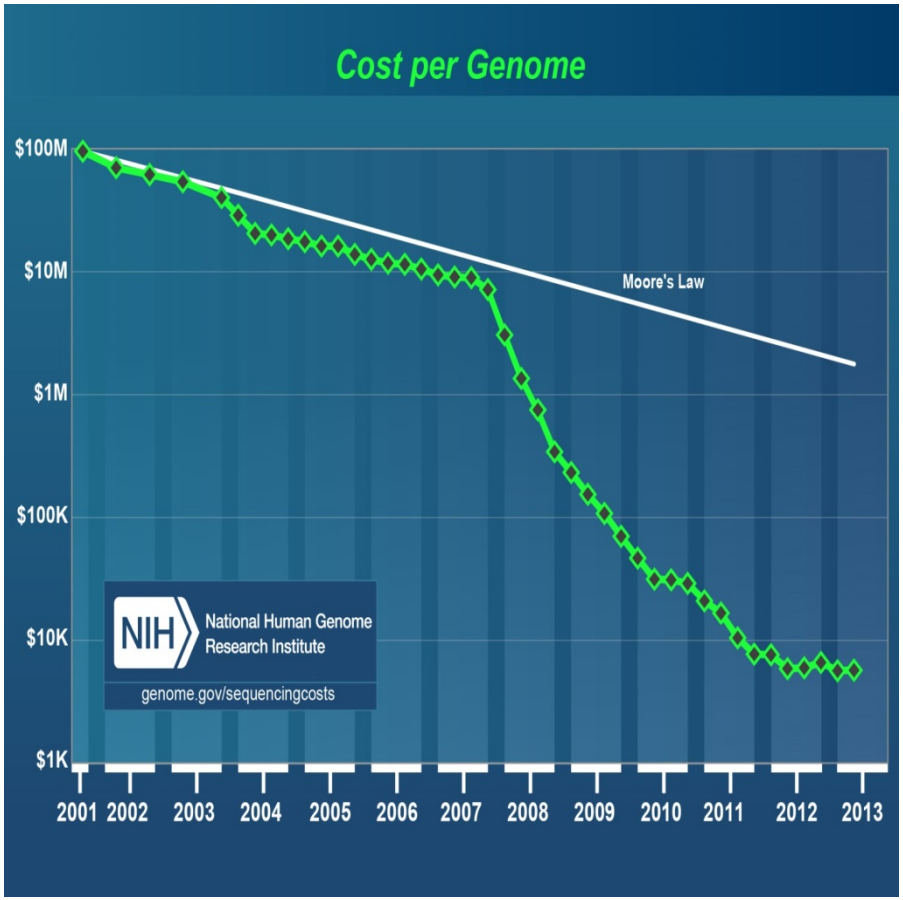
Genomics england

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

Genomics england

Is now the moment to commit to WGS



Data Type	Large-scale structural changes	Balanced translocations	Distant consanguinity	Uniparental disomy	Novel/known coding variants	Novel/known non-coding variants
Targeted gene sequencing	⊗	⊗	⊗	⊗	☑	⊗
SNP arraya	☑	⊗	⊗	☑	☑	⊗
Array CGH	☑	⊗	⊗	⊗	⊗	⊗
Exome	☑	⊗	⊗☑	⊗☑	☑	⊗
Whole Genome	☑	☑	☑	☑	☑	☑

Genomics england

WGS500 Results

- 7 Novel genes for disease
- 6 Novel phenotypes for known genes
- 2 pathogenic regulatory variants in or downstream of known candidate genes
- 6 genes missed by prior Sanger Sequencing

MENDELIAN

Of 95 families, to date

- 23 families have new clinical diagnosis
  - NB pre-screened for known genes
  - result will increase with follow-up
- 74 families in follow up studies
- Over 50% of these have strong lead candidate

Rare inherited diseases

- >5% of the population
- 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

Genomics England – Rare Disease Partnering opportunities

- NIHR Translational Research Collaborative
- NIHR BioResource
- £20m for deeper phenotyping
- Decipher & Deciphering Developmental Disorders
- NHS Clinical Genetics Service
- Farr Institute
- International Rare Diseases Research Consortium
- US and EU programmes

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

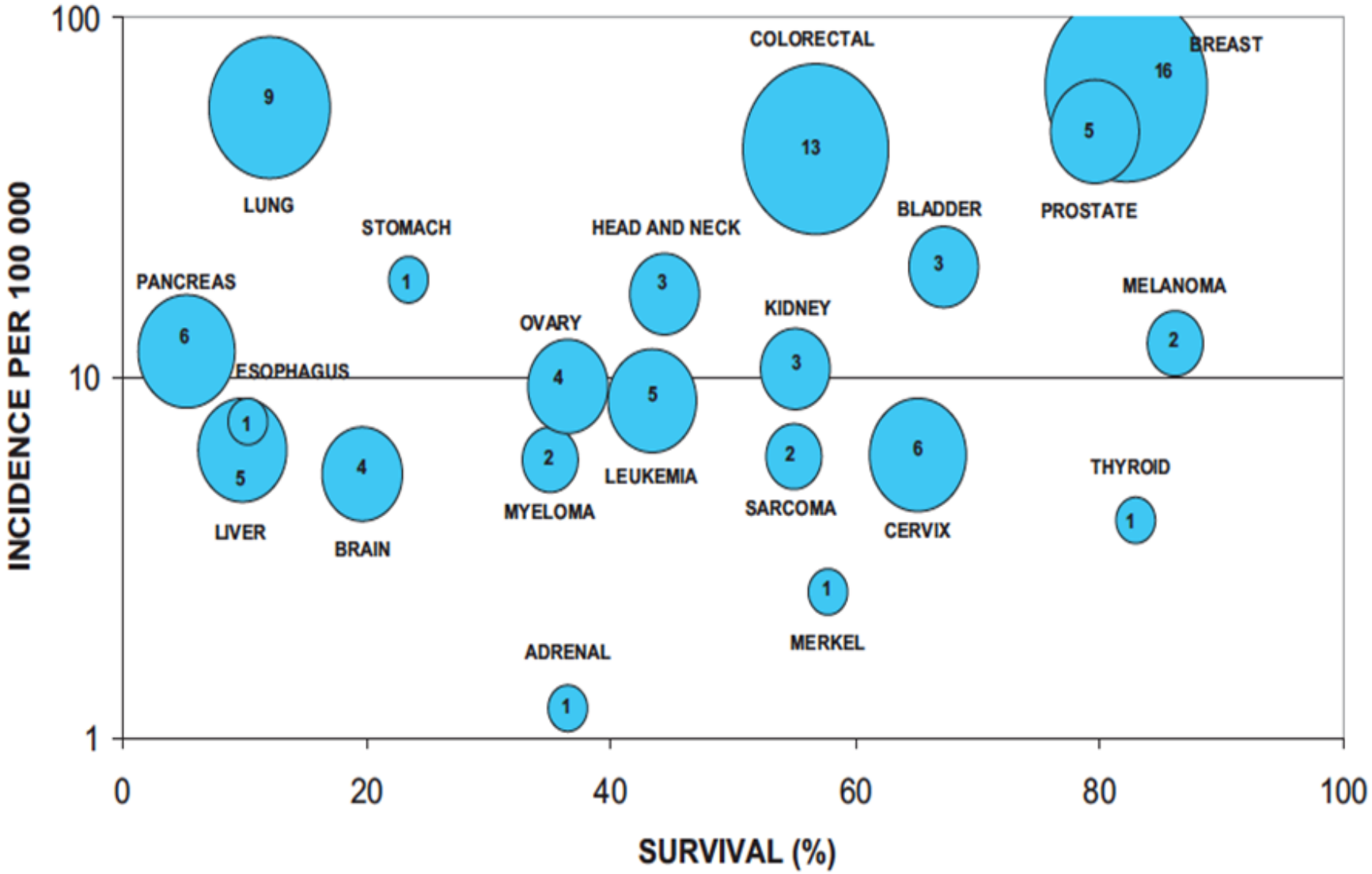
- Other Cancers - Breast, colon, prostate and unknown primary

- Rare and Childhood Cancers

- Drugs target mutations
- Tumour heterogeneity

Genomics england

Incidence v survival at 5 years



Pathogens

- Stratifying response, minimising adverse events and tracking outbreaks
- M. Tuberculosis resistance and epidemiology
- Hepatitis C genotype selects therapy
- HIV –Treatment for life and resistance testing is in the care pathway.
- Extreme human response to sepsis
- International linked datasetss

Public Health England

Genomics England Pilots

- Phase 1- Sequencing and Annotation Competition – now
- 4 providers 15 samples (5 tumour – normal pairs and 5 germline)
- Testing Sequencing QA and annotation

- Phase 2a-2000 Rare Inherited Disease WGS- 30x depth – over 2014
- Partnering NIHR BioResource and Translational Research Collaborative
- 5 centres - 928 samples since end of November- 1st 96 are in sequencing.

- Phase 2b- 3000 Cancer Patients (Lung, Breast, Ovary, Prostate & Colon)
- Somatic (?50-80x) and germline (30-40x) – tendering now
- Optimise Molecular Pathology pipeline
- 11 CRUK Centres and BRCs

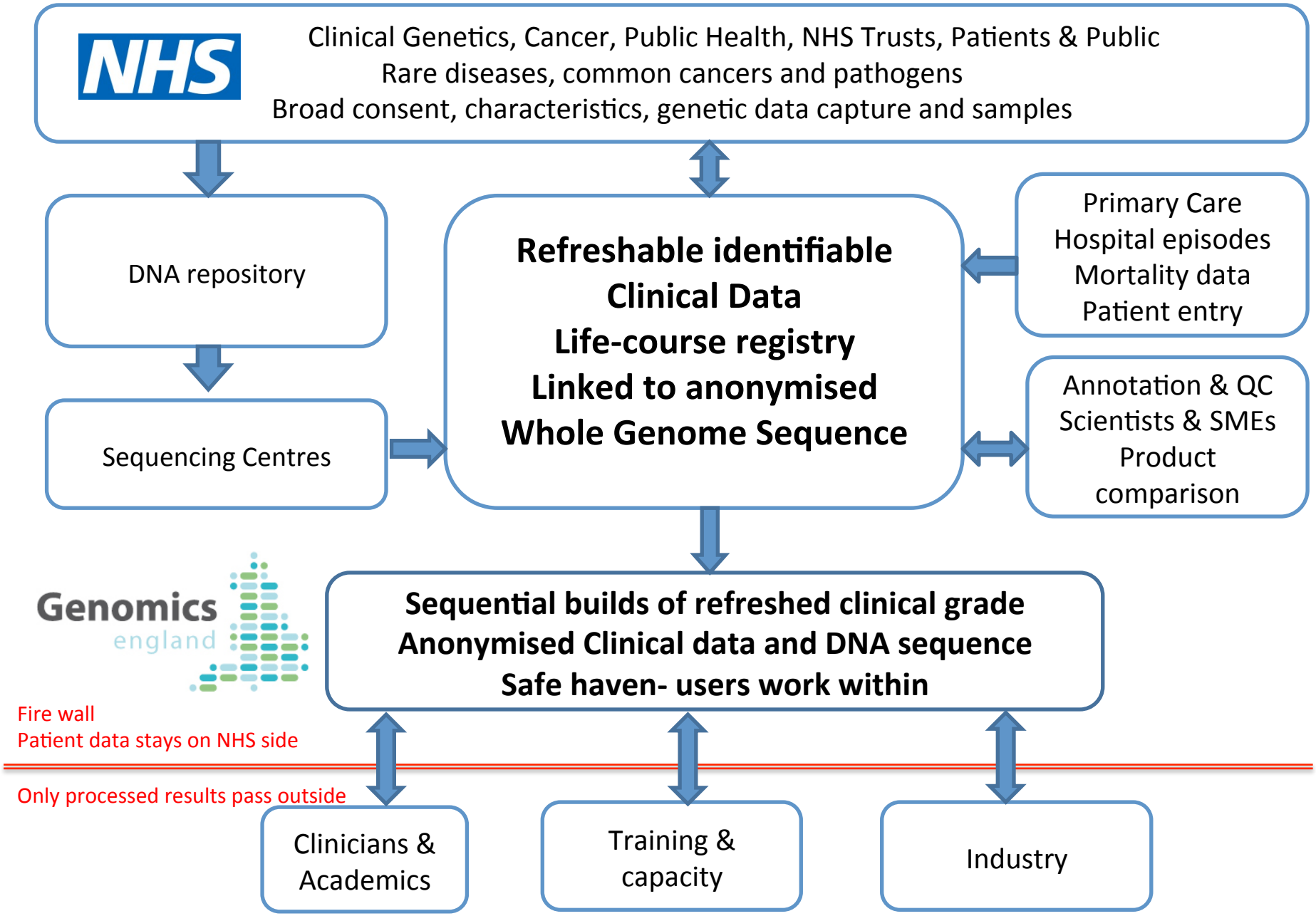
- Pathogens will be with Public Health England

Phase 3- Main Programme

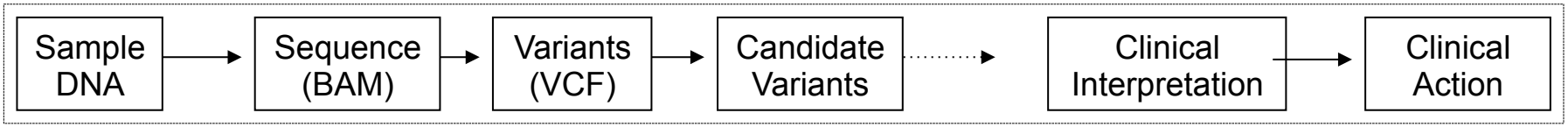
- Preparation for main programme underway
- Formal procurement to take place mid 2014
- Biorepository to be established
- Data architecture – advanced planning
- Envisage 2 or more Sequencing Centres in England
- WGS expected volume:
- 20k in 2015; 30k in 2016; 40k in 2017
- Elasticity in the pipeline

Genomics england

Genomics England – Operational Plan

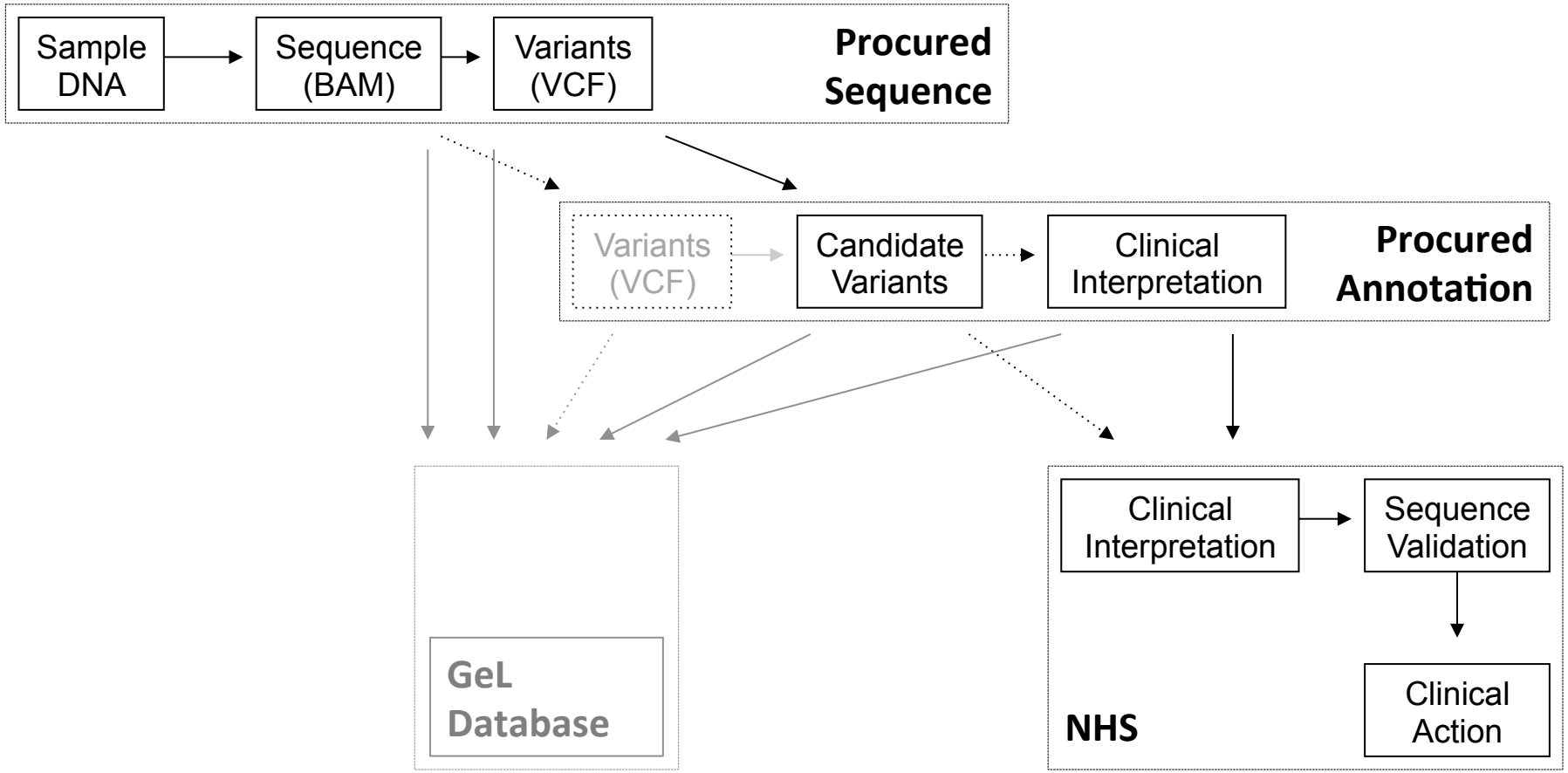


Process Overview

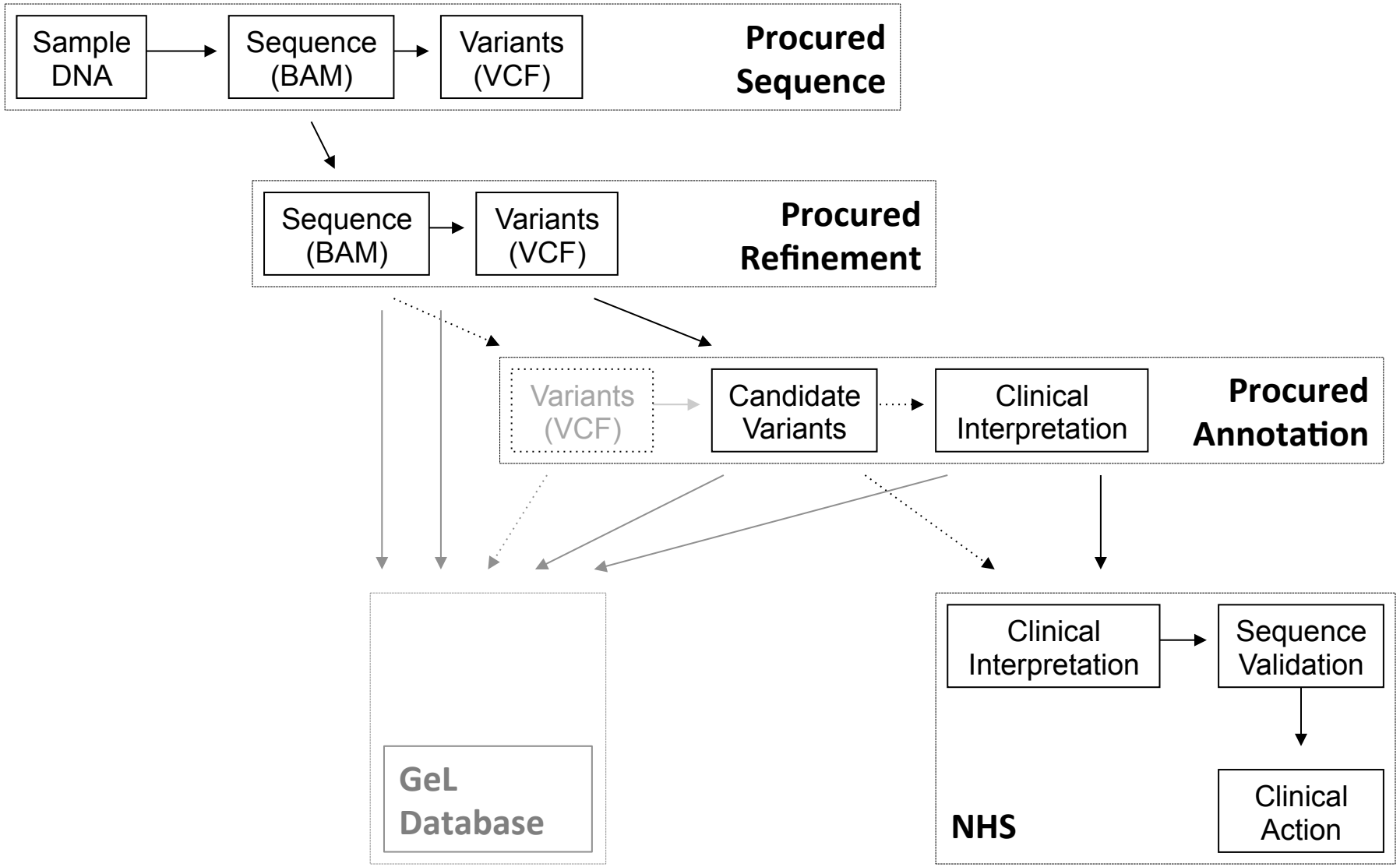


Genomics england

Process Overview



Genomics england



Genomics england

Sequencing assessment

- Data returned by suppliers, being assessed
- Evaluation on quality and coverage

Genomics england

Annotation assessment

- Harder than assessing sequencing
- Gold standard less well defined
- Lack of established data standards

Genomics england

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/
- CAGI – 2010, 2011, 2013
  - https://genomeinterpretation.org/

Genomics england

Data provided by GeL

- Sequence from providers (BAM+VCF)
  - Rare diseases: trio
  - Cancer: germline + tumour
- Phenotype data available to clinicians

Genomics england

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

Genomics england

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

Genomics england

Bake offs (phase 1)
Initial annotation assessment

- Information collection exercise on 5+5 studies
  - Investigate levels of annotation available
- Minimal file format requirements
  - Will inform future specifications for file formats

- Will select multiple suppliers for pilot (phase 2)
  - Ability to deliver timely, consistent data, etc.

Genomics england

Pilot (phase 2)
Ongoing annotation assessment

- Precise file format requirements, with optional sections

- Best suppliers will be invited to tender to provide annotation for main programme (2015-2017)

Genomics england

Main programme (phase 3)
Annotation generates clinical feedback

- 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

UK Genomic Medicine Research Data Infrastructure

the Farr, Oxford, Cambridge, EBI, Sanger, UK Biobank, Newcastle, Dundee, Kings, QMUL (lead)

- MRC Clinical Research Capabilities Call
- International and National Research Platform
- Rich clinical dataset, enriched by e-health (Farr, UK Biobank), high fidelity sequence, QA and annotation
- Inter-operable flexible data centre 60PB and 10,000 cores connected to sequencing centres, partners and users
- Applications e.g. Decipher, Ensemble, ENCODE, user developed software
- Embassies for research and combining datasets
- Cost of Research Data Centre is £24m - timing is key

Engaging with this programme

Clinical Interpretation Partnership
Academics, NHS, Philanthropy and Funders

Research Engagement

- 2000 –extending Rare disease to 6000
- Cancer just starting now 3000 aimed for.
- PHE – Pathogen Pilots

- Designate Genomics England Centres
- Local Lead and extended team
- High fidelity phenotypes and quality DNA
- Genomics England owns the data
- Clinical Interpretation Partnership
- Access to data on the samples
- 6 months protected space to analyse and submit
- Co-authors on papers

Multi-omics Cancer Repository

- RNA transcriptomics, micro RNAs
- Epigenetics, Proteomics and metabolomics
- Cell free circulating DNA (liquid biopsy)
- Sequential biopsies & WGS (trials)
- Immortal Cancer Cell Lines for drug sensitivity, single cell physiomics or functional genomics

- Harness the potential of MRC/NIHR Phenome Centres
- Experimental Cancer Medicine Centres
- International Cancer Genome Consortium and the Cancer Genome Atlas

- Barretina, J. *et al. Nature* **483**, 603-607, (2012).
- Garnett, M. J. *et al. Nature* **483**, 570-575, (2012).

Stratified medicine and therapeutic innovation

- Rare Inherited Disease – sub-phenotypes
- Priming functional studies at single cell, tissue, experimental models and the human Lab
- Therapeutic innovation or repurposing

- Cancer – Molecular pathology
- Priming functional studies at single cell, tissue, experimental models and the human Lab (strategic aim 1)
- Therapeutic innovation, repurposing or stratified medicine

- Pathogen resistance, epidemiology of disease outbreak
- Stratified healthcare link to PHE Reference labs and human data
- Opportunity for pathogen research

International Partnership

- Ethical and communications research - creating and partnering to set global standards

- Competitor or partner programmes
- Veterans Administration
  - Large cohort (1M), contract let for 5000 WGS
- US Integrated Healthcare Systems
  - Scripps – Wellderly Study sequenced 2000 volunteers to study lifestyle
  - Inova Health – aims to provide worlds larges WGS database, 1500 completed thus far
- Clinical Research programmes
  - Oxford 500 WGS, DDD (Sanger) up to 12,000 children (mainly exomes)
  - International Rare Diseases Consortium/Orphanet
  - International Cancer Genome Consortium
  - St Jude and Moffit Cancer Centres
  - Middle Eastern programmes
  - Chan Shoon-Shiong Foundation
  - Pharma and SMEs
  - Personal Genome Project – aims for 100k, all open data

Universities and NHS Education

- Health Education England
- UK Medical Schools

- Genomic medicines transformative potential needs an enabled healthcare team
- Needs a national approach
- Next generation sequencing
- Diagnostic potential in rare disease
- Stratified medicine in cancer
- Pathogen drug resistance
- Bioinformatics

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

NHS England

Public Health England

NHS National Institute for Health Research

CANCER RESEARCH UK

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