Introduction to Genomics England data, Data access policy

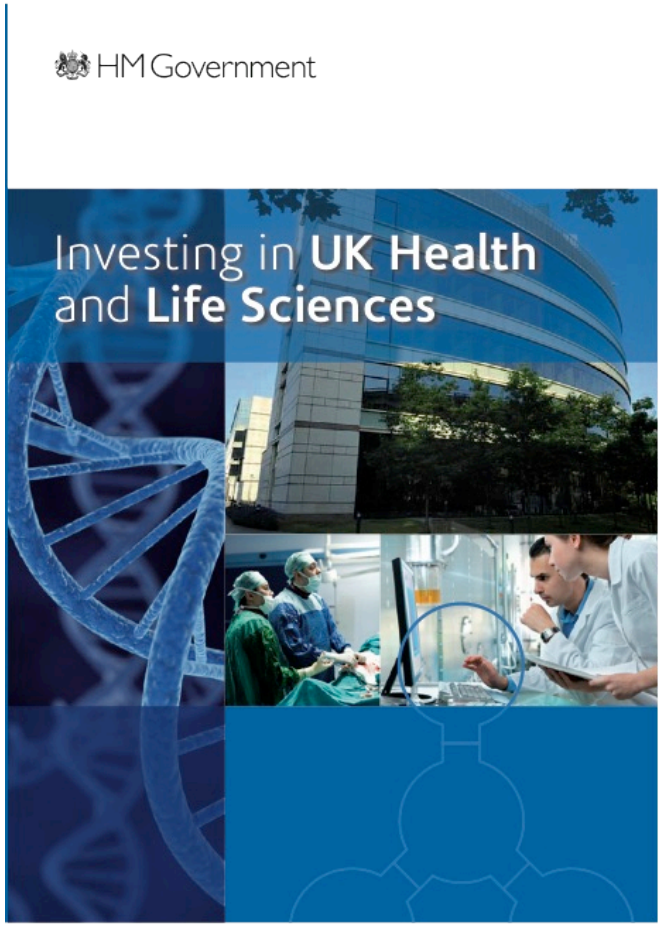
Tim Hubbard @timjph
King's College London, King's Health Partners
Genomics England

Bioinformatics, Interpretation and Data Quality in Genome Analysis
MSc in Genomics Medicine
15th February 2016

Steps in UK towards E-Health Research, Genomic Medicine

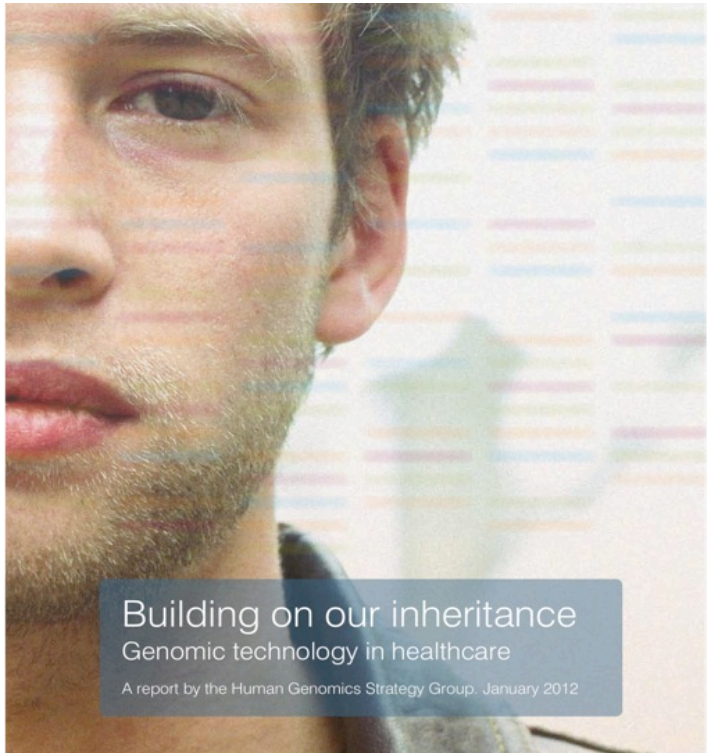
- Health data to Research
  - 2006 Creation of OSCHR
    - Increase coordination between funders: MRC and NIHR
  - 2007 OSCHR E-health board
    - Enable research access to UK EHR data
    - Build capacity for research on EHR data
- Genomics to Health
  - 2009 House of Lords report on Genomic Medicine
  - 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/

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

Genomics england

Home | About the 100K Genome Project | About us | How we work | News | Contact us

Home > Archive by Category "News"

JUL
5

Genomics England launched, mapping DNA to better understand cancer, rare and infectious diseases

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

@genomicsenglanc

Genomics england

100,000 genomes project

- Primarily a treatment project
  - NHS transformation project
- All clinical whole genome sequencing (>30x)
  - Rare disease (proband/parent trios)
  - Cancer (normal/tumour pairs)
- Timeline
  - Announced December 2012
  - Genomics England setup 2013
  - Pilots 2014
  - Main Programme 2015-2017

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

Genomics england

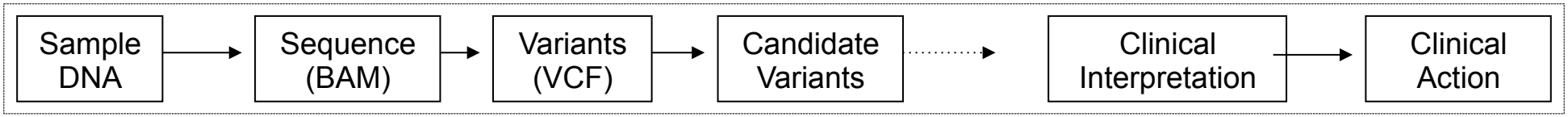
Genomics England – mission

- 100,000 whole genome sequences in NHS patients with rare diseases and cancers from the NHS in England
- Health improvement and wealth generation
- Legacy of infrastructure, human capacity and capability
- Enable large scale genomics research

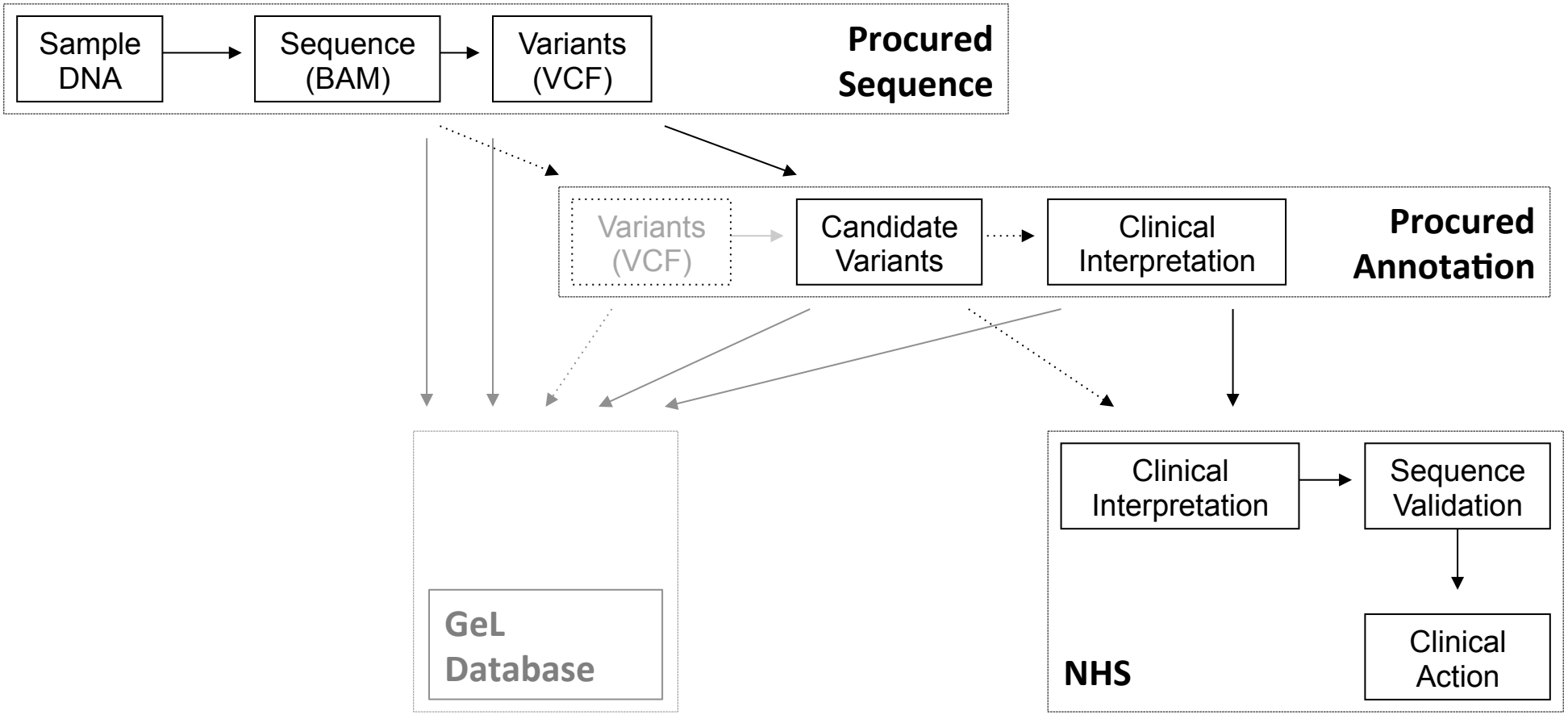
http://www.genomicsengland.co.uk/

Genomics england

Process Overview



Process Overview

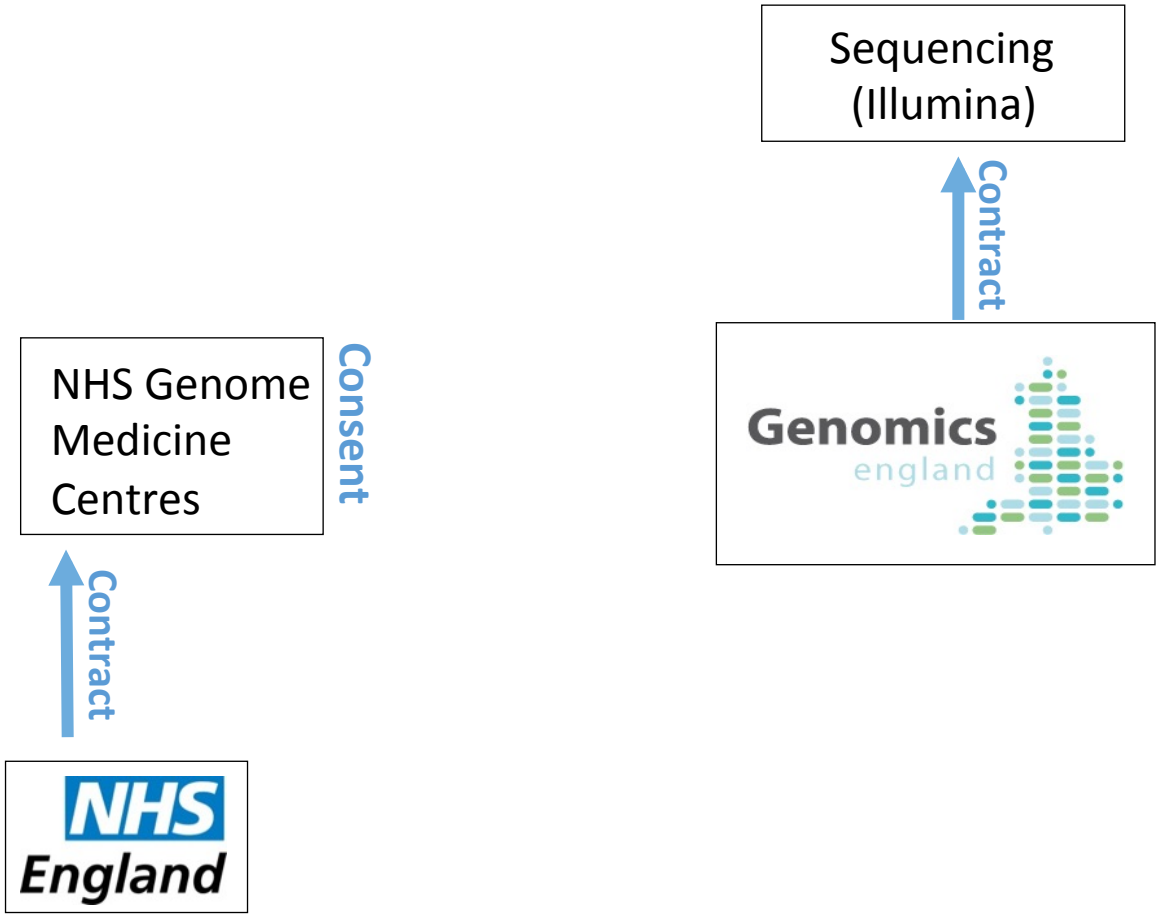


Genomics England

Sequencing and Annotation assessment

- Sequencing bake-off
  - Samples sent to participants; returned sequence assessed
  - Evaluation on quality and coverage
  - Informed sequencing contract

Genomics England



Genomics
england

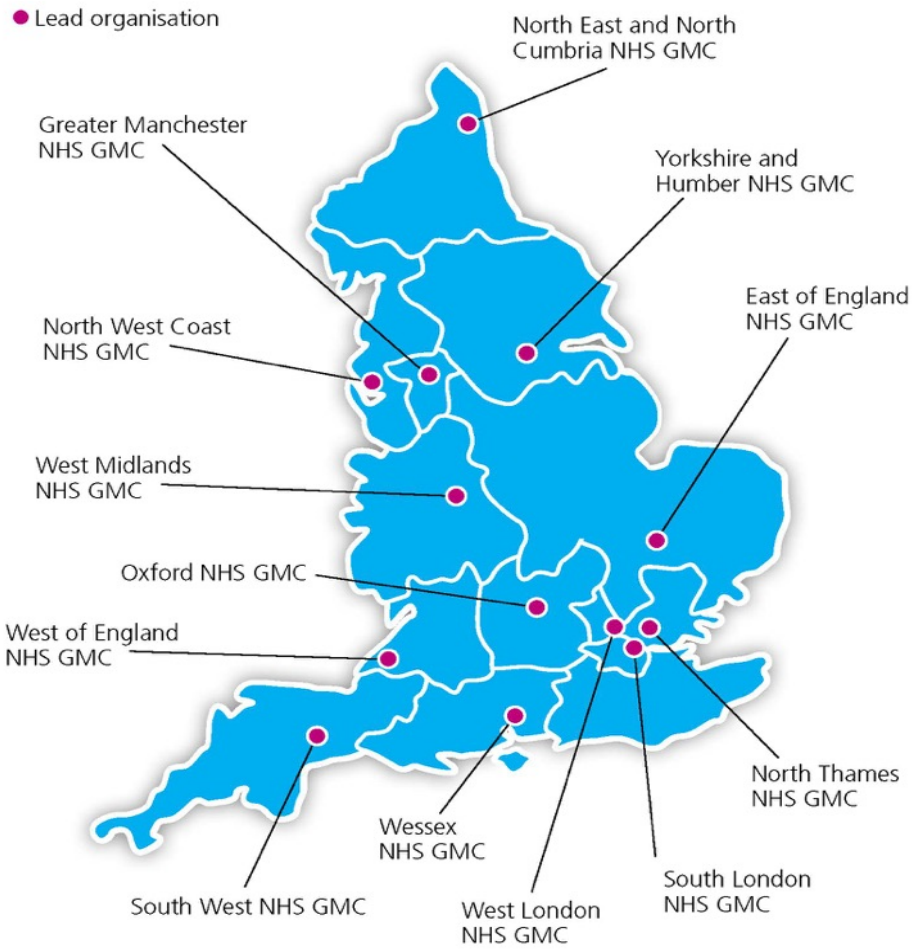
Eleven Genome Medicine Centres announced

- **East of England NHS GMC:** Led by Cambridge University Hospitals NHS Foundation Trust;
- **South London NHS:** Led by Guy's and St Thomas' NHS Foundation Trust.
- **North West Coast NHS GMC:** Led by Liverpool Women's NHS Foundation Trust.
- **Greater Manchester NHS GMC:** Led by Central Manchester University Hospitals NHS Foundation Trust
- **University College London Partners NHS GMC:** Led by Great Ormond Street Hospital NHS Foundation Trust
- **North East and North Cumbria NHS GMC:** Led by The Newcastle upon Tyne Hospitals NHS Foundation Trust.
- **Oxford NHS GMC:** Led by Oxford University Hospitals Foundation Trust.
- **South West Peninsula NHS GMC:** Led by Royal Devon & Exeter NHS Foundation Trust.
- **Wessex NHS GMC:** Led by University Hospital Southampton NHS Foundation Trust.
- **Imperial College Health Partners NHS GMC:** Led by Imperial College Healthcare NHS Trust.
- **West Midlands NHS GMC:** Led by University Hospitals Birmingham NHS Foundation Trust.

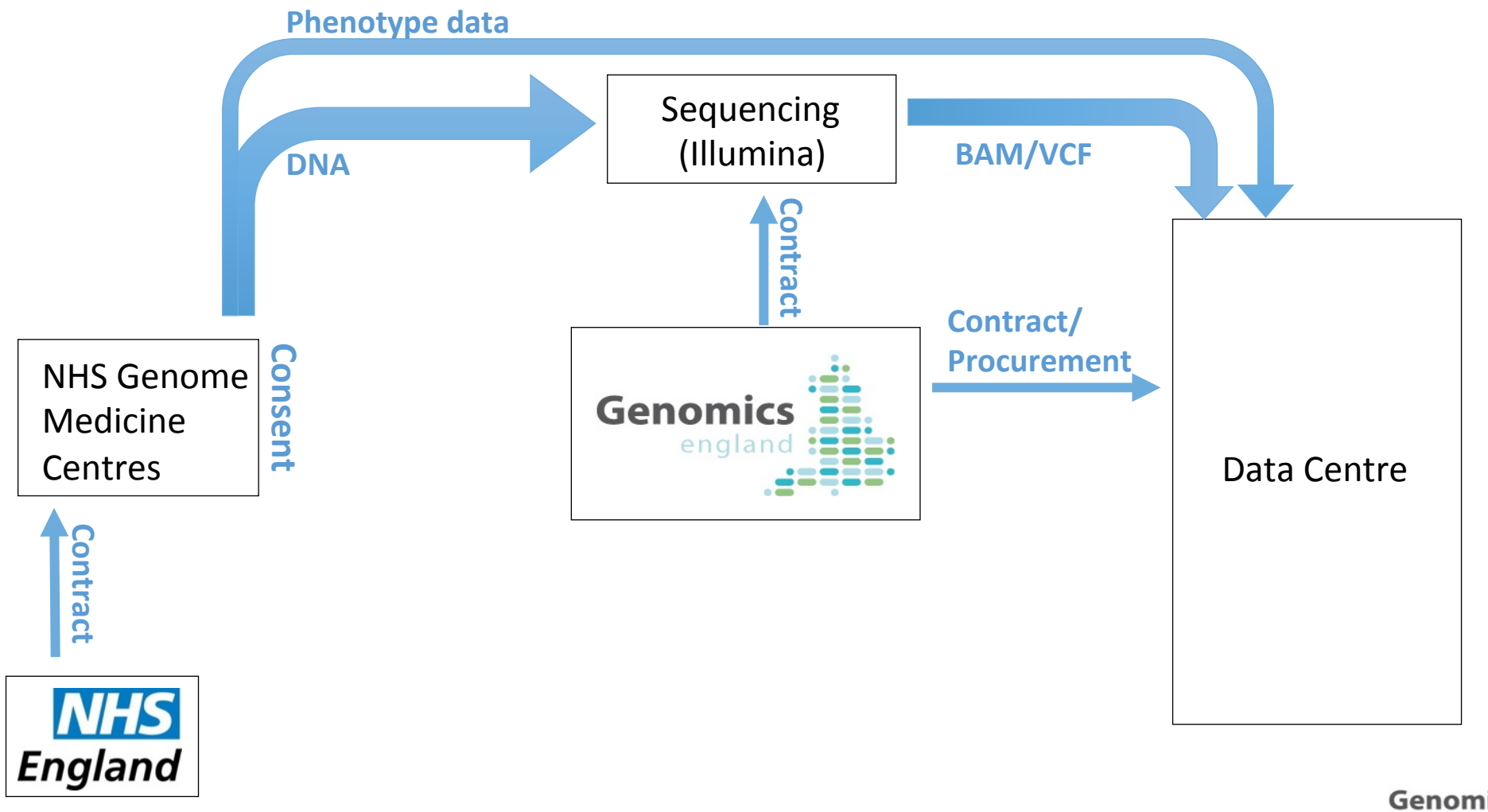
Genomics england

NHS Genomic Medicine Centres

- *11 Genomic Medicine Centres (GMCs) established in December 2014 by NHS England. These centres will lead the way in delivering the Project.*
- *Track-record of providing excellence in genomic services.*
- *Eligible patients will be referred to GMCs by their clinicians.*
- *Two new GMCs announced in December 2015 – Yorkshire & Humber and West of England*



Genomics England



Genomics england

Data model development

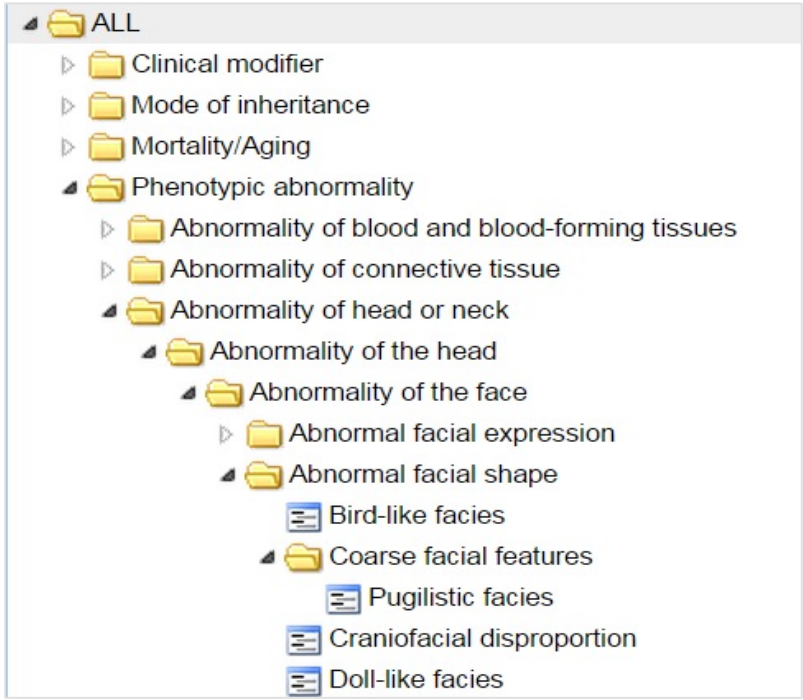
- Which participants should we recruit?
  - List of conditions – currently 122
  - **Eligibility statements**
- What data do we need?
  - Metadata: Demographics, Sample, Consent
  - Clinical data: **Data models**
  - Associated genes: **Gene packages**

Developing data models is complex for rare disease patients, needing consultation with experts in the field

HPO as universal ontology for phenotypic features

Human Phenotype Ontology

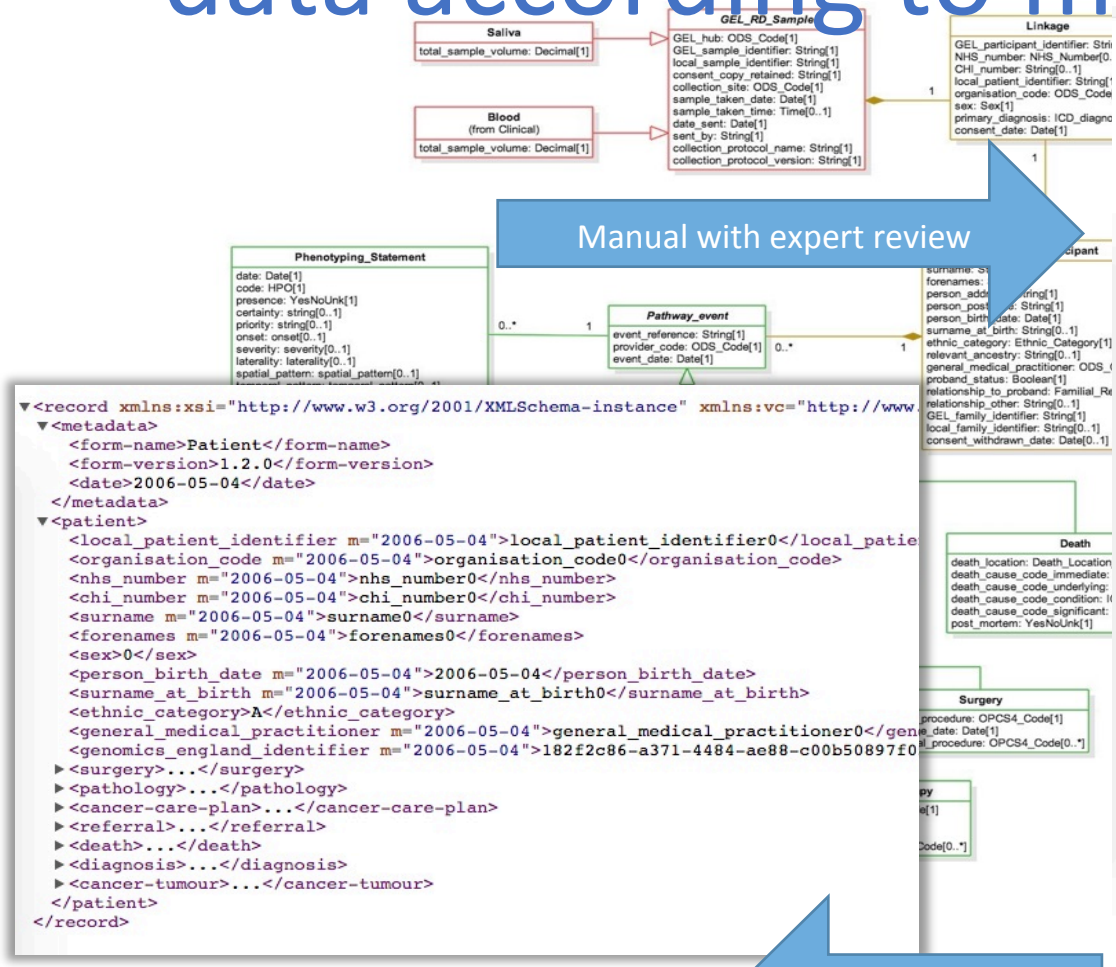
- Chosen as a **standard for deep representation of phenotypic features**
- Adopted by **other projects**, e.g. DDD, FORGE, familiar to many in rare diseases
- Being **actively developed** in collaboration with broader RD community
- **Existing mapping** from diseases to HPO terms



Data models are specific to each condition

Level 1	Level 2	Level 3	Level 4
Rare Disease Conditions and Phenotypes(11144.4)	Cardiovascular disorders(10950.1)	Connective Tissues Disorders and Aortopathies(10951.1)	Familial Thoracic Aortic Aneurysm Disease(11021.1)
		Cardiac arrhythmia(10952.1)	Brugada syndrome(11022.1)
			Long QT syndrome(11023.1)
			Catecholaminergic Polymorphic Ventricular Tachycardia(11024.1)
		Cardiomyopathy(10953.1)	Arrhythmogenic Right Ventricular Cardiomyopathy(11025.1)
			Left Ventricular Noncompaction Cardiomyopathy(15044.1)
			Dilated Cardiomyopathy (DCM)(11026.1)
			Dilated Cardiomyopathy and conduction defects(11027.1)
			Hypertrophic Cardiomyopathy(11028.1)
		Congenital heart disease(10954.1)	Fallots tetralogy(11029.1)
			Hypoplastic Left Heart Syndrome(11030.1)
			Pulmonary atresia(11031.1)
			Transposition of the great vessels(11032.1)
			Left Ventricular Outflow Tract obstruction disorders(11033.1)
			Isomerism and laterality disorders(11034.1)

Informatics approach to capture data according to models



Manual with expert review

```
▼<record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:vc="http://www.
  ▼<metadata>
      <form-name>Patient</form-name>
      <form-version>1.2.0</form-version>
      <date>2006-05-04</date>
    </metadata>
  ▼<patient>
      <local_patient_identifier m="2006-05-04">local_patient_identifier0</local_patie
      <organisation_code m="2006-05-04">organisation_code0</organisation_code>
      <nhs_number m="2006-05-04">nhs_number0</nhs_number>
      <chi_number m="2006-05-04">chi_number0</chi_number>
      <surname m="2006-05-04">surname0</surname>
      <forenames m="2006-05-04">forenames0</forenames>
      <sex>0</sex>
      <person_birth_date m="2006-05-04">2006-05-04</person_birth_date>
      <surname_at_birth m="2006-05-04">surname_at_birth0</surname_at_birth>
      <ethnic_category>A</ethnic_category>
      <general_medical_practitioner m="2006-05-04">general_medical_practitioner0</gen
      <genomics_england_identifier m="2006-05-04">182f2c86-a371-4484-ae88-c00b50897f0
    ▶<surgery>...</surgery>
    ▶<pathology>...</pathology>
    ▶<cancer-care-plan>...</cancer-care-plan>
    ▶<referral>...</referral>
    ▶<death>...</death>
    ▶<diagnosis>...</diagnosis>
    ▶<cancer-tumour>...</cancer-tumour>
    </patient>
</record>
```

Automated

OpenClinica
Open Source for Clinical Research

Alport Phenotypes FINALIZED

Properties | Attachments | Children 11 | Data Elements | History 1

Is Base For | Is Based On | Is Favourite Of | Is Synonym For | Metadata

Parents 1 | Related To

Name	Identification	Metadata
Proteinuria (Human Phenotype Ontology)	http://purl.obolibrary.org/obo/HP_0000093	

Increased levels of protein in the urine.

Classifications Human Phenotype Ontology

Model Metadata

OBOI D HP:0000093

Xrefs <MeSH:D011507 "Proteinuria">, UMLS:C0033687

Synonyms

Relationship Metadata

Actions for Model ★ Export

Actions for Relationship Export

Hematuria (Human Phenotype Ontology)	http://purl.obolibrary.org/obo/HP_0000790	
Renal insufficiency (Human Phenotype Ontology)	http://purl.obolibrary.org/obo/HP_0000083	
Nephrotic range proteinu	http://purl.obolibrary.org/obo/H	

OpenClinica phenotype entry

Disease

1 Disease Group: Renal and urinary tract disorders *

2 Disease Subgroup: Syndromes with prominent renal abnormalities *

3 Specific disease: Alport syndrome *

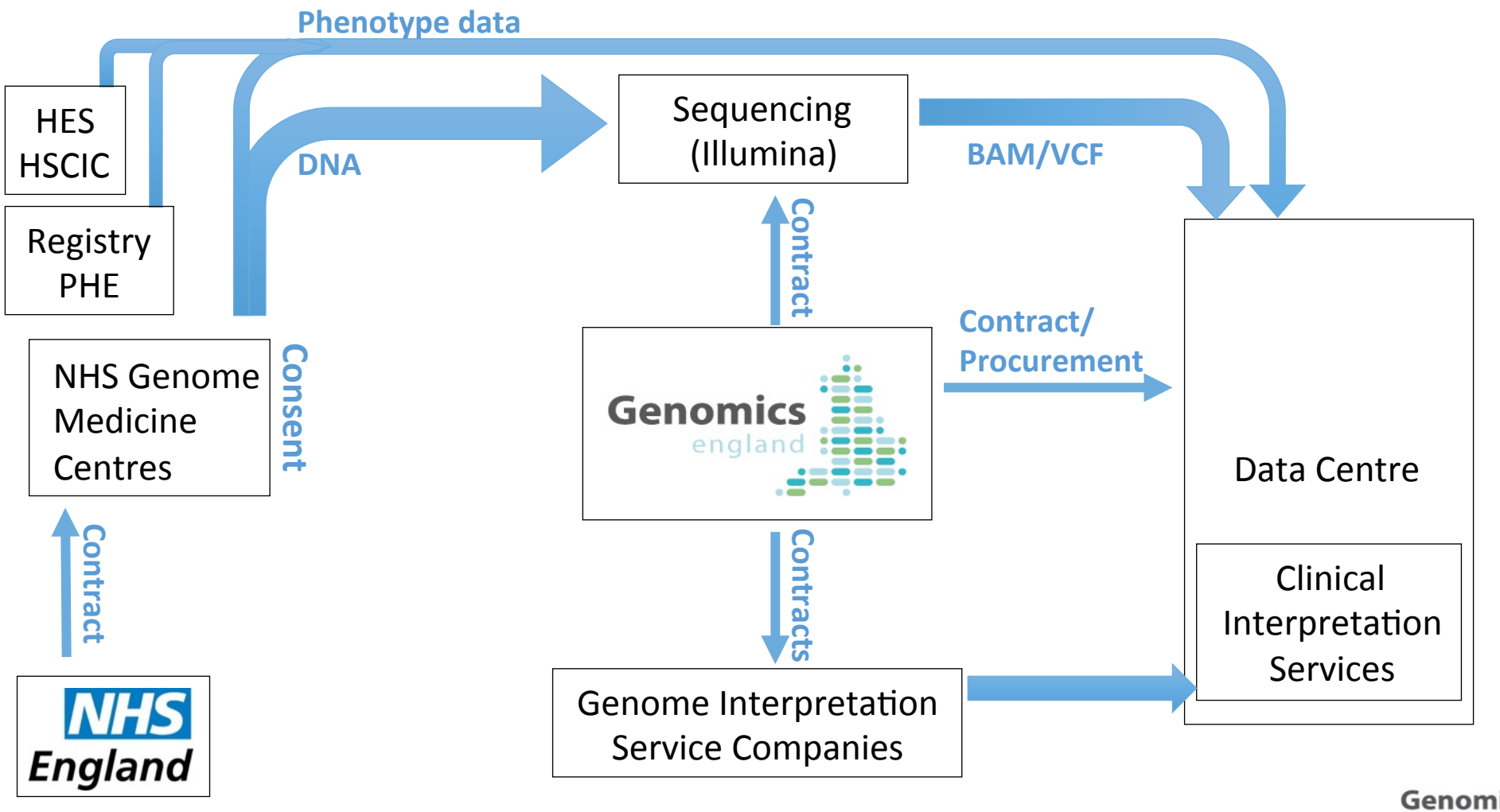
OpenClinica
Open Source for Clinical Research

Basic Phenotyping

4 Phenotype Description	5 Phenotype Identifier	7 Phenotype Present	Modifiers	Actions
Proteinuria	HP:0000093	Unknown / Yes / No		Edit
Hematuria	HP:0000790	Unknown / Yes / No		Edit
Nephrotic range proteinuria	HP:0012593	Unknown / Yes / No		Edit
Renal insufficiency	HP:0000083	Unknown / Yes / No		Edit

Additional terms not present in the data model can be naturally added

Genomics England

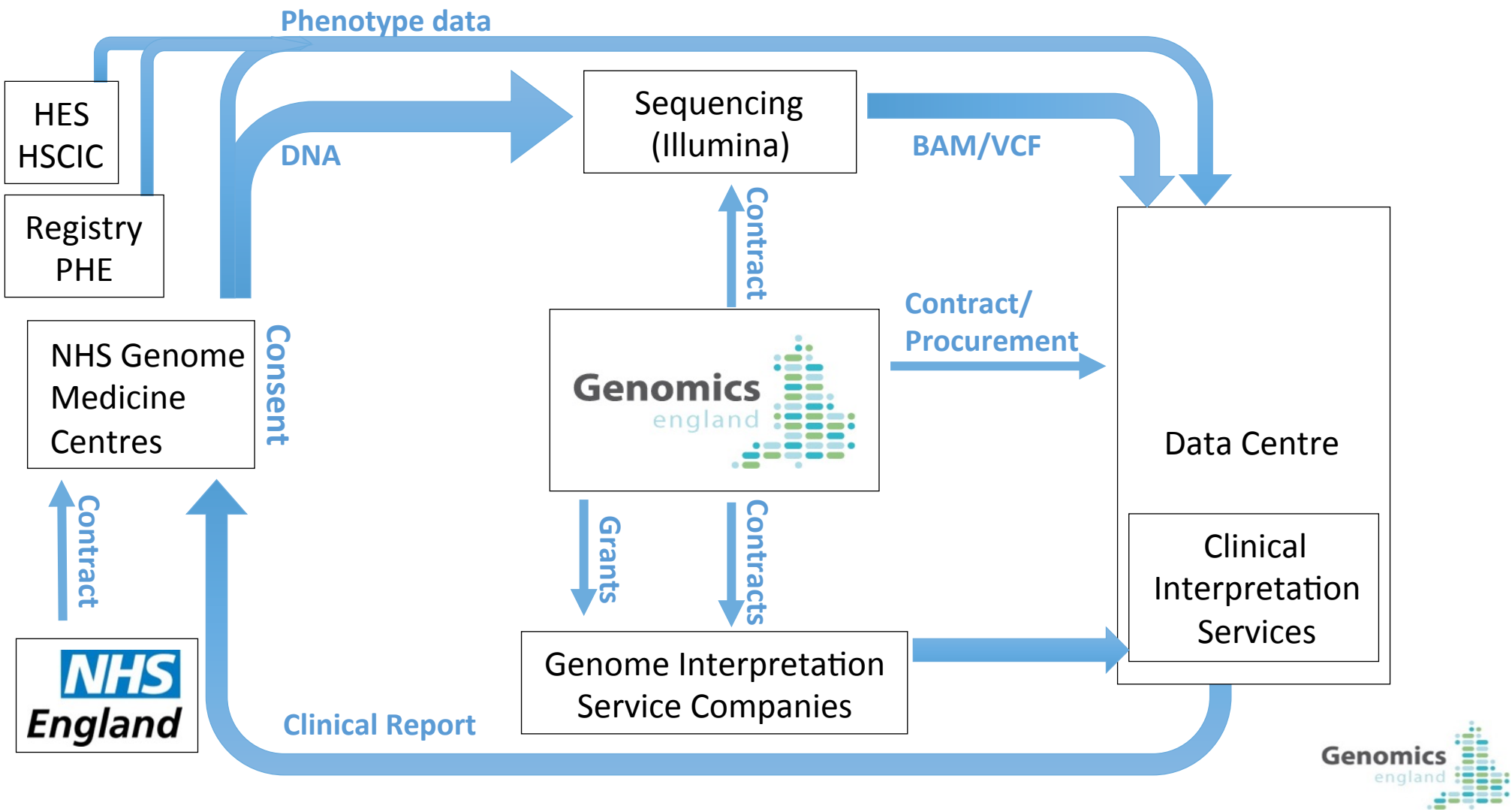


Genomics
england

Sequencing and Annotation assessment

- Sequencing bake-off
  - Samples sent to participants; returned sequence assessed
  - Evaluation on quality and coverage
  - Informed sequencing contract
- Annotation bake-off
  - Sequence sent to participants (BAM+VCF)
    - Rare diseases: trio
    - Cancer: germline + tumour
  - Harder than assessing sequencing
  - Gold standard less well defined
  - Lack of established data standards

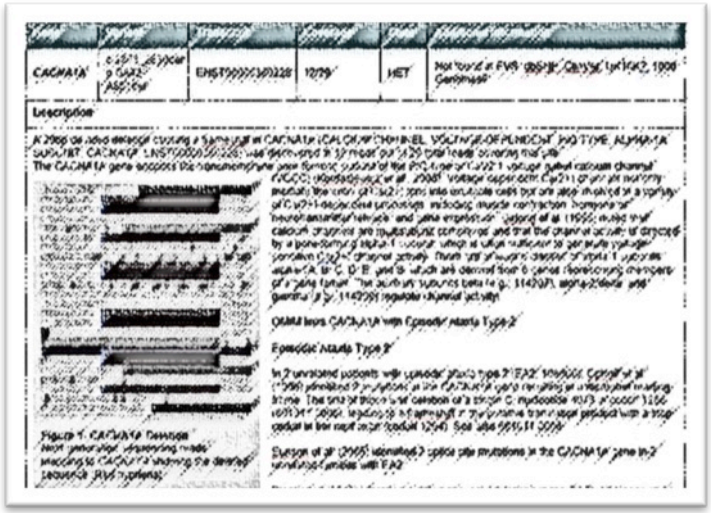
Genomics England



Sequencing and Annotation providers

- Contracted Suppliers
  - Omicia (California)
  - Congenica (Cambridge) Sanger spinout
  - WuNextCode (Iceland) deCODE spinout
- InnovateUK SBRI (Small Business Research Initiative)
  - Congenica (Cambridge) Sanger spinout
  - Genomics plc (Oxford) Wellcome Trust Centre spinout
  - Seven Bridge UK (London) US subsidiary
  - Oxford Gene Technology (Oxford)
  - Omixon (Budapest)

Feedback to the NHS

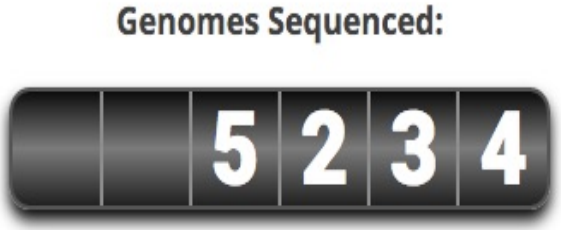


- Diagnostic reports that are accessible and meaningful
- Dynamic serial reporting - evolving findings
- **Primary** findings:
  - Known pathogenic and expected pathogenic variants on known genes
- **Secondary** "looked for" findings (currently for 10 conditions):
  - Strong cancers predisposition and familial hypercholesterolemia
  - For example Lynch syndrome, BRCA1/2, multiple endocrine neoplasia (MEN1), VHL
- **Carrier states** of reproductive importance (currently for 12 conditions):
  - Thalassemia, sickle cell, hemophilia A, …

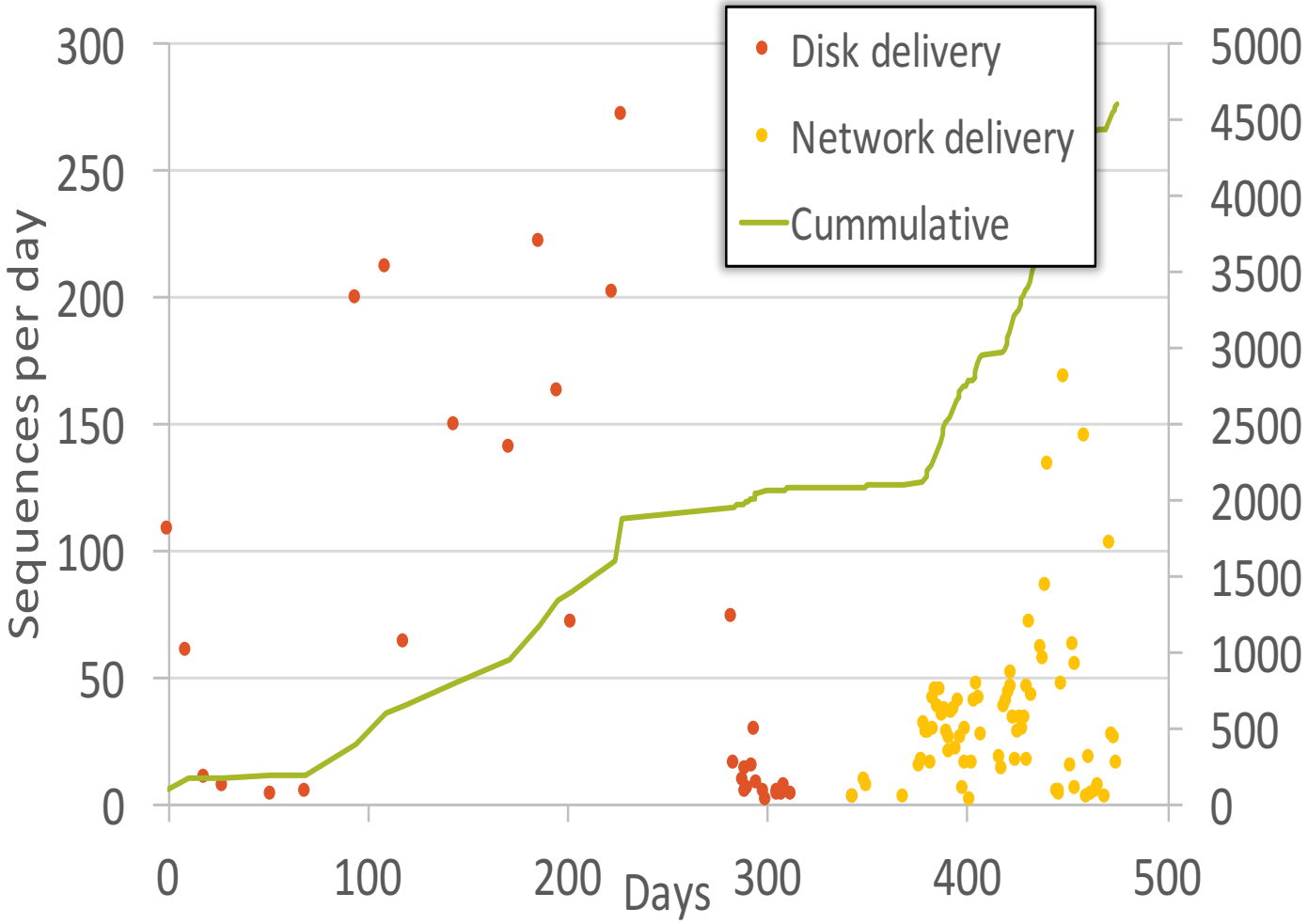
- Read and variant level data accessible to NHS referring teams
- Patients can request genomic data files from Genomics England
- Patients are consented to be contacted up to four times a year

Genomics england

Some figures



- 11 Genomic Medicine Centres (hubs) and >70 Local delivery hospitals (spokes)
- About 9,000 participants consented

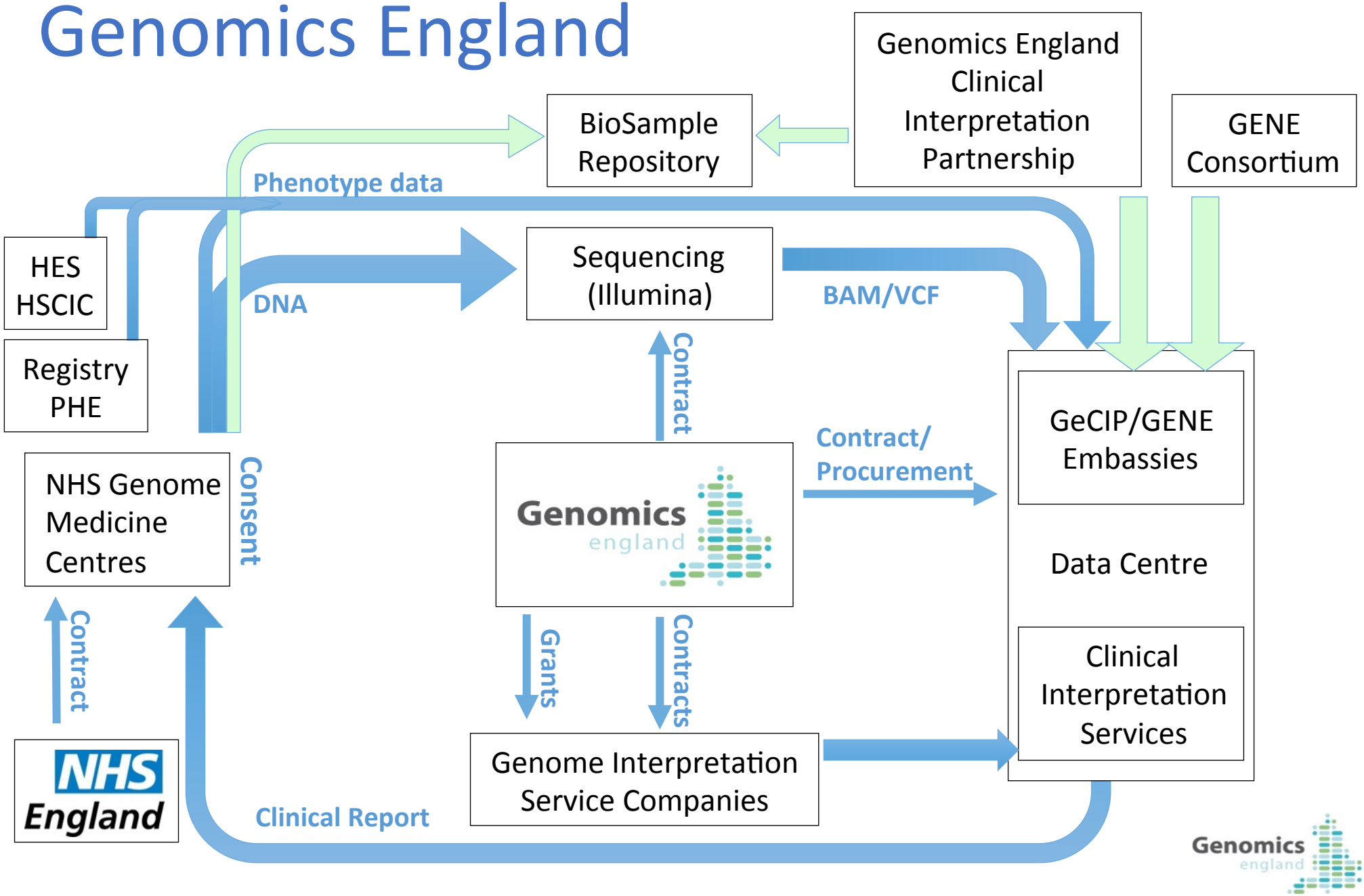


Research Protocol under new designation of “Bioresource”

- Single **project-wide approval**: no need for site specific approvals
- Independent review committee grants data access to bona-fide research uses
- Consent for return of **additional findings** (secondary: 17 genes; and carrier status: 8 genes)
- Participants can be **re-contacted** up to four times a year
- Samples for various **–omics** technologies collected
- Revision of diagnosis if underlying evidence changes (e.g. when new is gene discovered)

http://www.genomicsengland.co.uk/library-and-resources/

Genomics England



Genomics england

Call for Expressions of Interest

The Genomics England Clinical Interpretation Partnership

Genomics England invites 'expressions of interest' from UK led consortia of clinicians, researchers, analysts and those in training to propose disease specific domains in the areas of rare inherited disease, cancer and infectious disease. The Genomics England Clinical Interpretation Partnership will lead research to enhance the clinical interpretation of whole genome sequences and support the delivery of healthcare transformation from the 100,000 Genomes Project.

This will be the route by which Genomics England will engage with the UK academic and healthcare community and their international collaborators to discover new biological insights into disease, elucidate functional impact, develop novel analytical approaches and create high cadre expertise in genomic medicine.

The overall aim is for the Genomics England Clinical Interpretation Partnership to create thriving, sustainable communities of research and clinical (NHS) disease experts to interrogate the 100,000 whole genome sequences. The domains within Genomics England Clinical Interpretation Partnership will have three primary roles:

- **Research:** Harnessing opportunities for research and discovery enabled by the 100,000 Genomes Project with the intention of further enhancing our understanding of genomic medicine and its application in healthcare.
- **Clinical Interpretation:** Provision of disease-specific expertise in clinical reporting and variant interpretation to enhance interpretation of 100,000 Genomes Project data to ensure feedback of the highest calibre data to treating clinicians in order to inform diagnostics and treatment decisions.
- **Training:** Training of researchers and clinicians.

Expressions of interest are invited from self-organised consortia to form domains with a UK lead (clinical or non-clinical) with a connection through a higher education institute or the National Health Service to UK healthcare. Each consortia must clearly create a multidisciplinary clinical, academic and training domain which offers high calibre skillsets. We encourage the UK led domains to involve key international collaborators.

The successful Clinical Interpretation Partner Domains will be given free access, subject to our Data Access and Acceptable Uses Policy, to embassies within the Genomics England Data Centre which has been funded by the Medical Research Council.

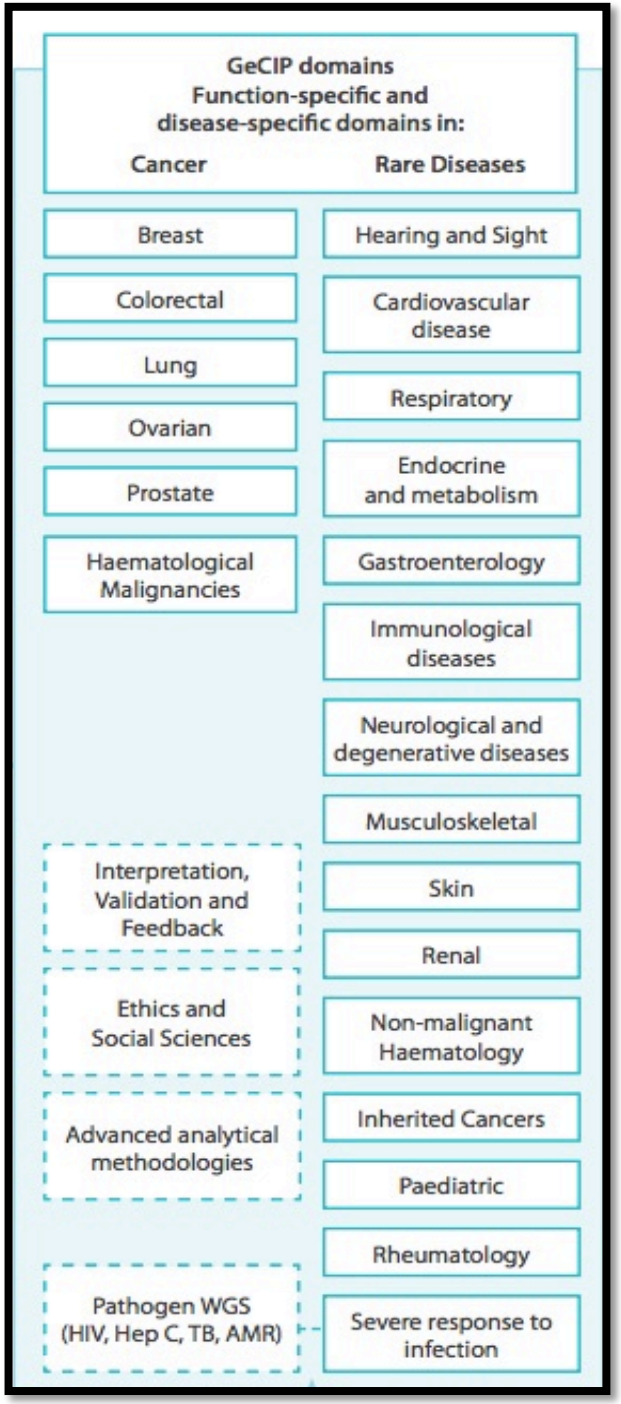
We have scheduled an open meeting to facilitate further discussion. This will be held on the 5th December at the Wellcome Trust, 215 Euston Road, London, NW1 2BE, 4-6pm.

For further information and guidance on how to submit expressions of interest please visit www.genomicsengland.co.uk . You can also call or email us chiefscientist@genomicsengland.co.uk / 020 7882 3402.

Closing date for expressions of interest: Monday 26th January 2015 at 5.00pm

Announcement of Genomics England Clinical Interpretation Partnership Domains will be in February 2015.

www.genomicsengland.co.uk



Genomics England Clinical Interpretation Partnership - GECIP

Goals

- Drive up the fidelity of clinical interpretation of genome sequencing
- Foster the use of the programme's data
- Accelerate academic/industry partnership and development of diagnostics and therapies.

Composition

- UK-led and organised into domains
- Self proposed partnership between researchers, the NHS and Trainees with skills.
- Can bring international collaborators

Expectations:

- All data generated contributes to the Genomics England Dataset and are available to all inside a GeCIP domain.
- IP owned by Genomics England but readily licensed to incentivise active collaboration
- Training workstreams

GeCIP Domains – 1st wave

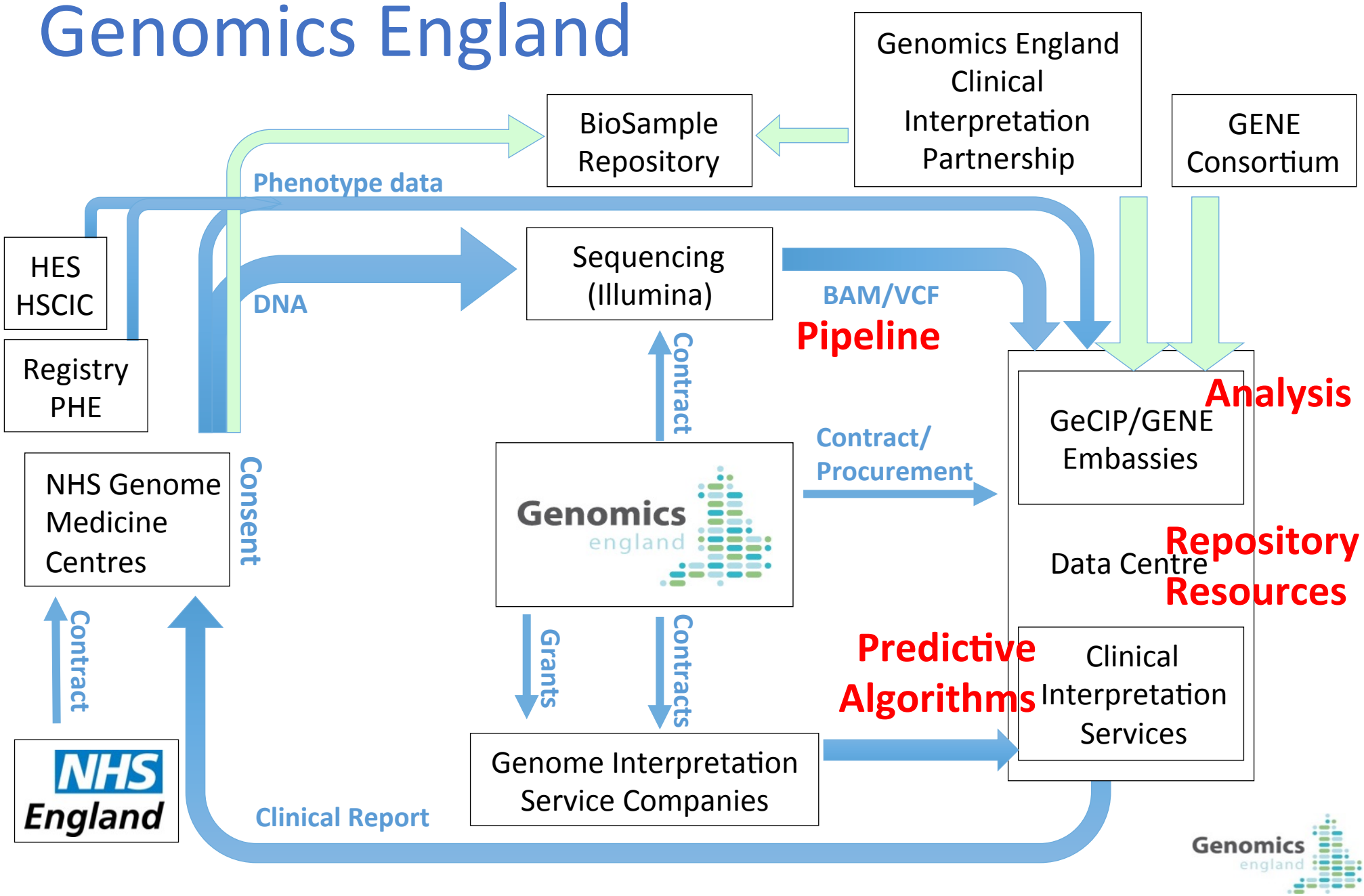
- 8 proposed domains for cancer derived from 26 submissions: ovarian, lung, breast, etc
- 14 domains for rare disease comprising 21 submissions: cardiovascular, neurological, paediatrics, etc
- 10 functional and cross cutting domains comprising 24 submissions. Population genomics, variant interpretation, education and training/primary care, etc
- 1 Ethics, Law and Social care domain comprising 13 submissions.
- Clinical Interpretation, Validation and Feedback (V&F) domain: "operations" arm of GECIP to coordinate clinical interpretation

Gene Consortium launch partners

- AbbVie
- Alexion Pharmaceuticals
- AstraZeneca
- Biogen
- Dimension Therapeutics
- GSK
- Helomics
- Roche
- Takeda
- UCB*

Genomics england

Genomics England



Data Sharing

- Open to all
  - Human Genome Projects where subject consented: Hapmap, 1000 genomes
  - Repository: Genbank, ENA, DDBJ (INSDC)
- Managed distribution (must be *bona fide* researcher)
  - Genetic data for disease cohorts, with phenotypes
  - Repository: DbGaP, EGA (Encrypted distributions etc.)
- Managed access, no redistribution
  - Genomics England datasets
  - Repository: GeL Datacentre

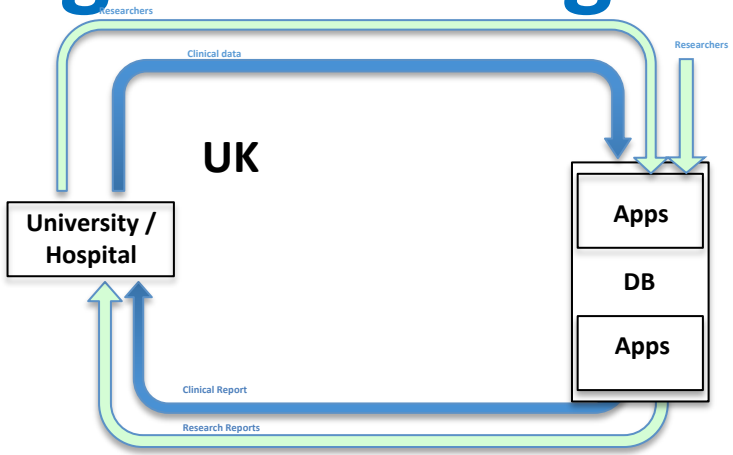
Data Sharing

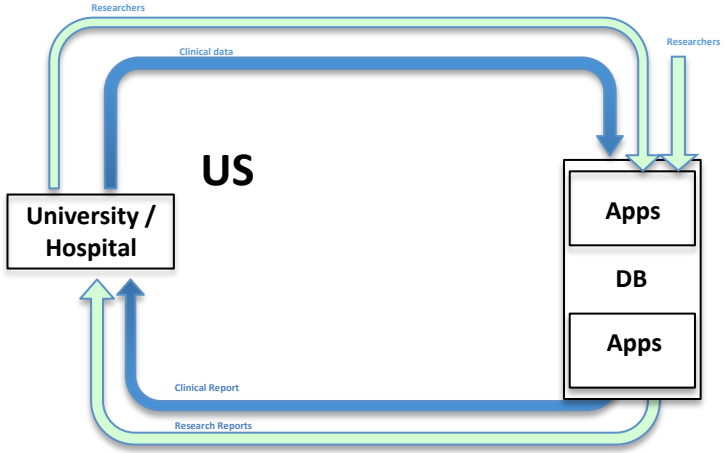
- Open to all
  - Human Genome Projects where subject consented: Hapmap, 1000 genomes
  - Repository: Genbank, ENA, DDBJ (INSDC)
  - Download, run analysis, algorithms on own systems
- Managed distribution (must be *bona fide* researcher)
  - Genetic data for disease cohorts, with phenotypes
  - Repository: DbGaP, EGA (Encrypted distributions etc.)
  - Download, run analysis, algorithms on own **secure** systems
- Managed access, no redistribution
  - Genomics England datasets
  - Repository: GeL Datacentre
  - Upload analysis, algorithms to GEL systems via AIRLOCK

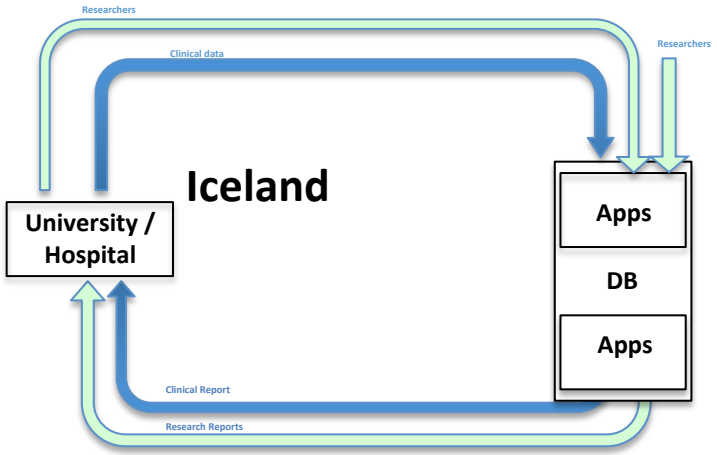
Generic delivery model

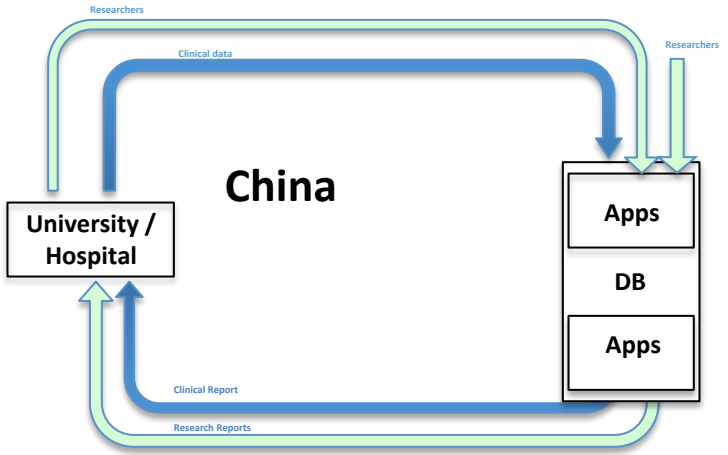
Genomics england

Effect of health service data living in walled gardens









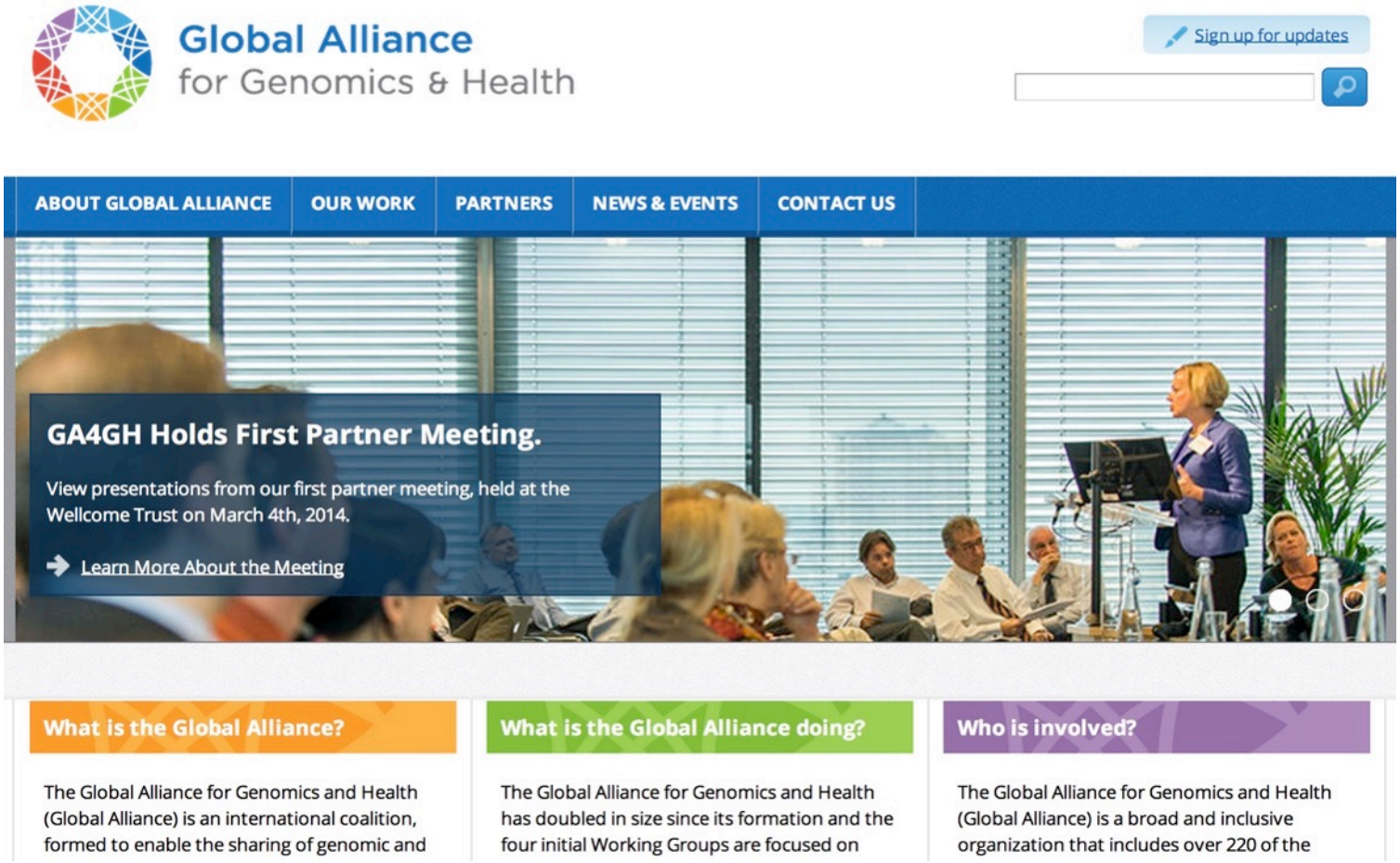
Genomics england

A future with closed datasets

- Multiple sets of Hospital/National datasets with no redistribution policies
- Value for research in generating statistics across this global set

Genomics england

Global Alliance for Genomes and Health
http://genomicsandhealth.org/



Genomics england

Acknowledgements

Special thanks

- Cambridge, UCLH, GOSH, Moorfields, Newcastle, Manchester, Guys and St Thomas's, Oxford, Liverpool, Sheffield, Leeds, Birmingham, Royal Marsden, Southampton, UK CLL Consortium, CRUK, RCPath, NHS England, Department of Health, Biobank UK, Sanger, EBI, KCL, UCL and QMUL

All Genomics England Teams:

- Science, Operations, Informatics, Bioinformatics, Programmes, Communications, Administrative Support

All advisory committees and working groups:

- Science, ethics, data, cancer, rare diseases, molecular pathology

Acknowledgements

- Global Genomic Medicine Collaborative (G2MC)
- South London NHS Genome Medicine Centre (GMC)
- King's College London

Genomics england