Use of phenotype data to obtain novel insights into disease causes and mechanisms

Damian Smedley

Director of Genomic Interpretation, Genomics England

Senior Lecturer, Queen Mary University of London

Overview

- Introduction to the 100,000 Genomes Project
- Role of clinical and model organism phenotypes
  - Clinical data collection and panel assignment
  - Automated variant prioritisation
  - Precision mouse models for validation and functional characterisation of the variants

Overview

- Introduction to the 100,000 Genomes Project

- Role of clinical and model organism phenotypes
  - Clinical data collection and panel assignment
  - Automated variant prioritisation
  - Precision mouse models for validation and functional characterisation of the variants

The 100,000 Genomes Project

Genomics england

100,000 genomes

70,000 patients and family members

21 Petabytes of data.
1 Petabyte of music would take 2,000 years to play on an MP3 player.

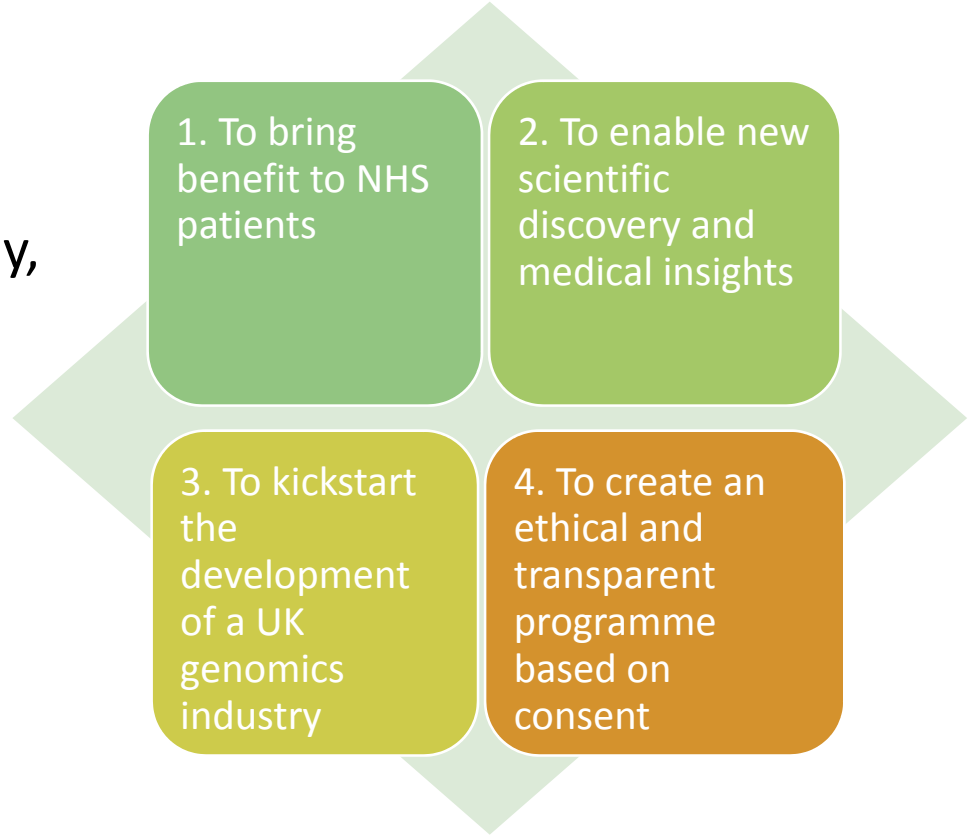
13 Genomic Medicine Centres, and **85** NHS Trusts within them are involved in recruiting participants

1,500 NHS staff
(doctors, nurses, pathologists, laboratory staff, genetic counsellors)

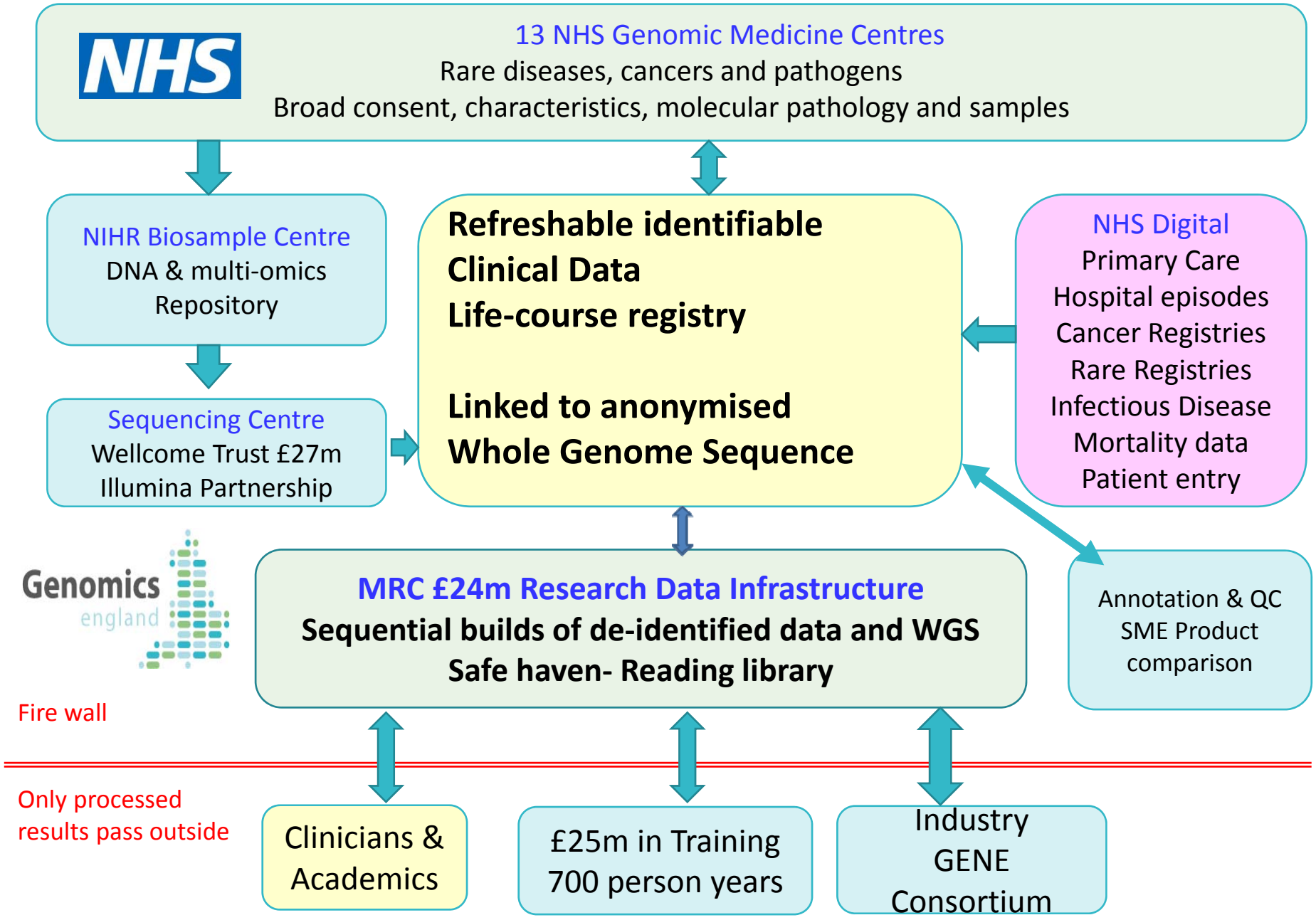
2,500 researchers and trainees from around the world

Goals of the Genomics England project

- Sequence 100,000 genomes
- Cancer and rare genetic disease
- Capture data delivered electronically, store it securely and analyse it within an English data centre (reading library)
- Combine genomes with extracted clinical information for analysis, interpretation, and aggregation
- Create capacity, capability and legacy in personalised medicine for the UK



Organisation of the 100,000 Genomes Project



GeCIP Domains

Rare

- Cardiovascular
- Endocrine and Metabolism
- Gastroenterology and Hepatology
- Hearing and Sight
- Immunology and Haematology
- Inherited Cancer Predisposition
- Musculoskeletal
- Neurological
- Paediatric Sepsis
- Paediatrics
- Renal
- Respiratory
- Skin

Cancer

- Adult Glioma
- Bladder
- Breast
- Colorectal & upper
- Lung
- Melanoma
- Renal Cell
- Sarcoma
- Testis
- Ovarian
- Prostate
- Childhood Solid Cancers
- Haematological Malignancy
- Pan Cancer
- (Ca of) Unknown primary

Functional

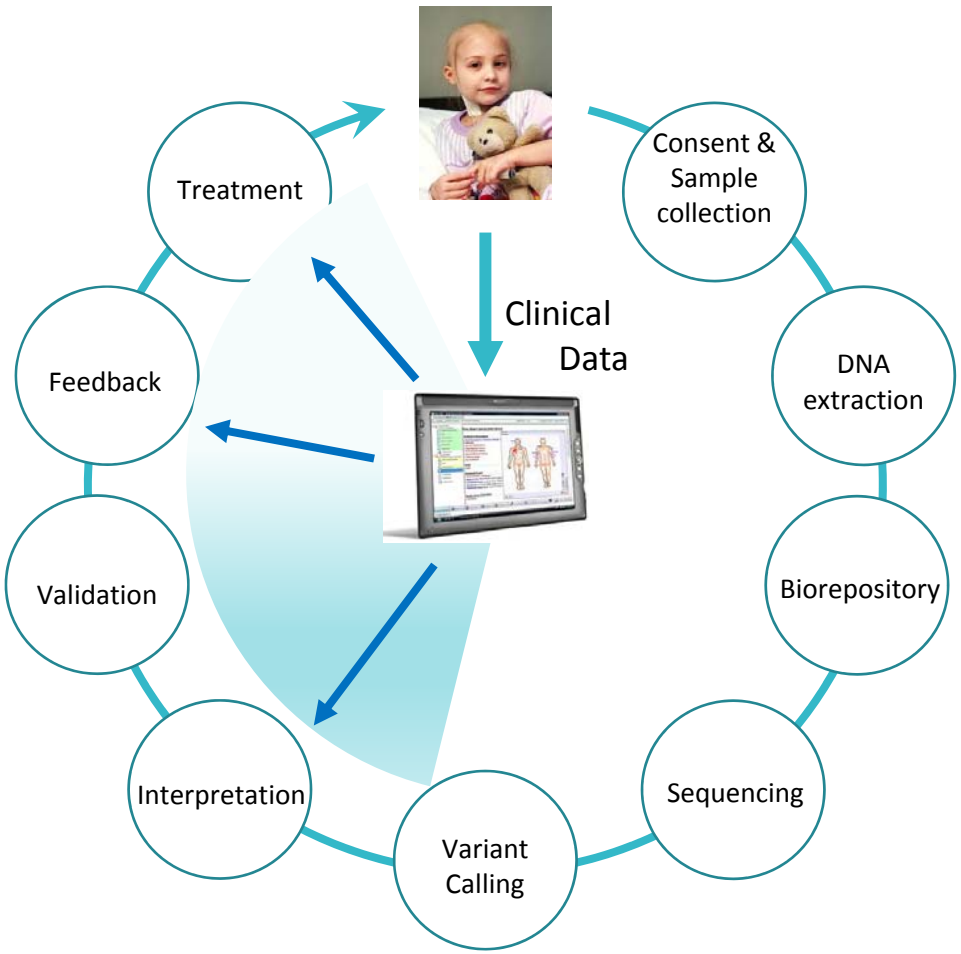
- Electronic Health Records
- Validation and Feedback
- Ethics and Social Science
- Functional Effects
- Health Economics
- Machine Learning, Quantitative Methods and Functional Genomics
- Population Genomics
- Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability
- Functional Cross Cutting
- Education and Training
- Stratified Medicine & Pharmacogenomics

Overview

- Introduction to the 100,000 Genomes Project

- Role of clinical and model organism phenotypes
  - Clinical data collection and panel assignment
  - Automated variant prioritisation
  - Precision mouse models for validation and functional characterisation of the variants

Genomics England is about helping complete the cycle



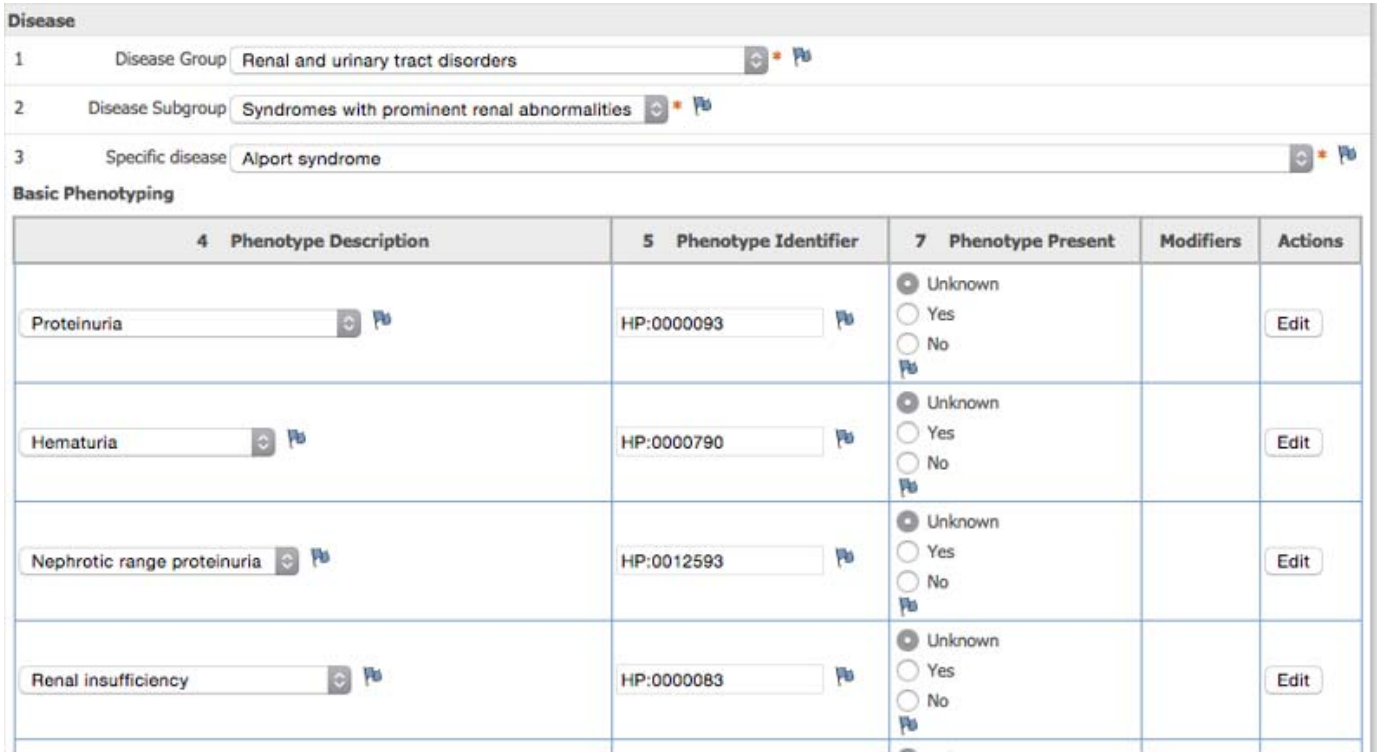
- Treatment cycle for just one patient requires a complex chain of operations
- Most of these operations have not been designed or optimised for the purposes of Genomic Medicine.
- So the task is one of catalysing a **Transformation** in Medical Practice, particularly relating to routine use of coordinated data.
- To achieve this one needs to develop/adopt **standards** across the system

Developing a clinical data capture system for RD genome diagnostics

Genomics england

Standardised clinical data questionnaire forms

- OpenClinica or their own systems exporting in XML schemas
- Defined HPO terms for each disease category; collect positive and negative and additional terms



Disease

1 Disease Group: Renal and urinary tract disorders *

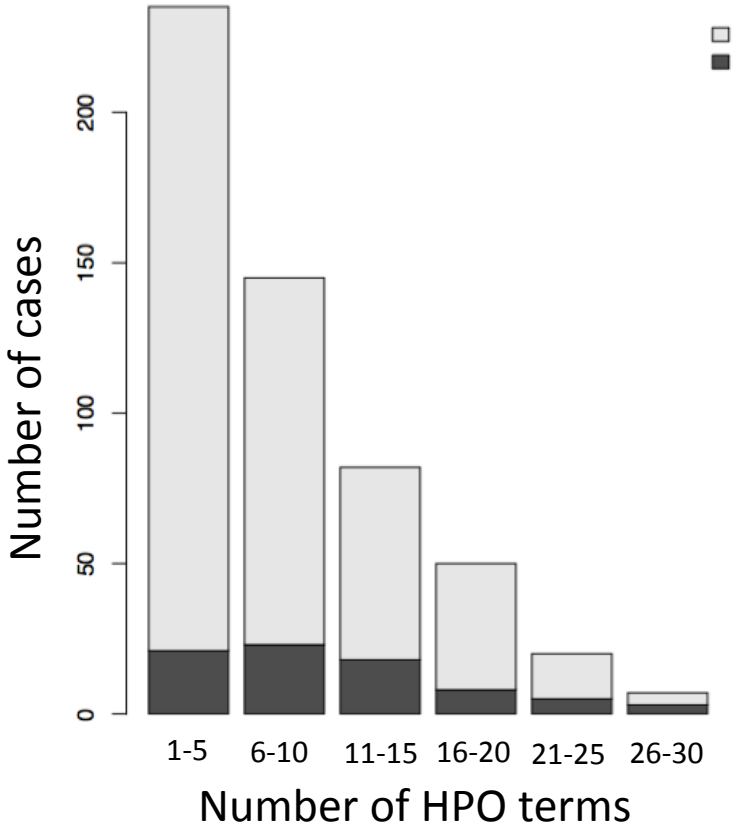
2 Disease Subgroup: Syndromes with prominent renal abnormalities *

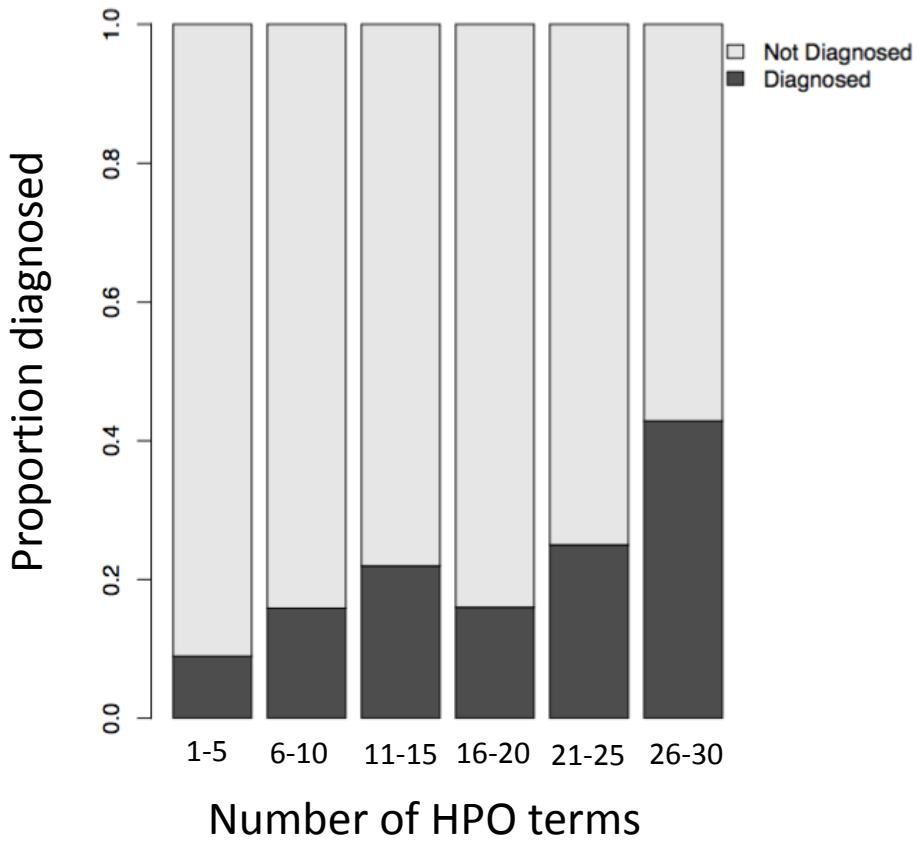
3 Specific disease: Alport syndrome *

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

Diagnostic rate for trios

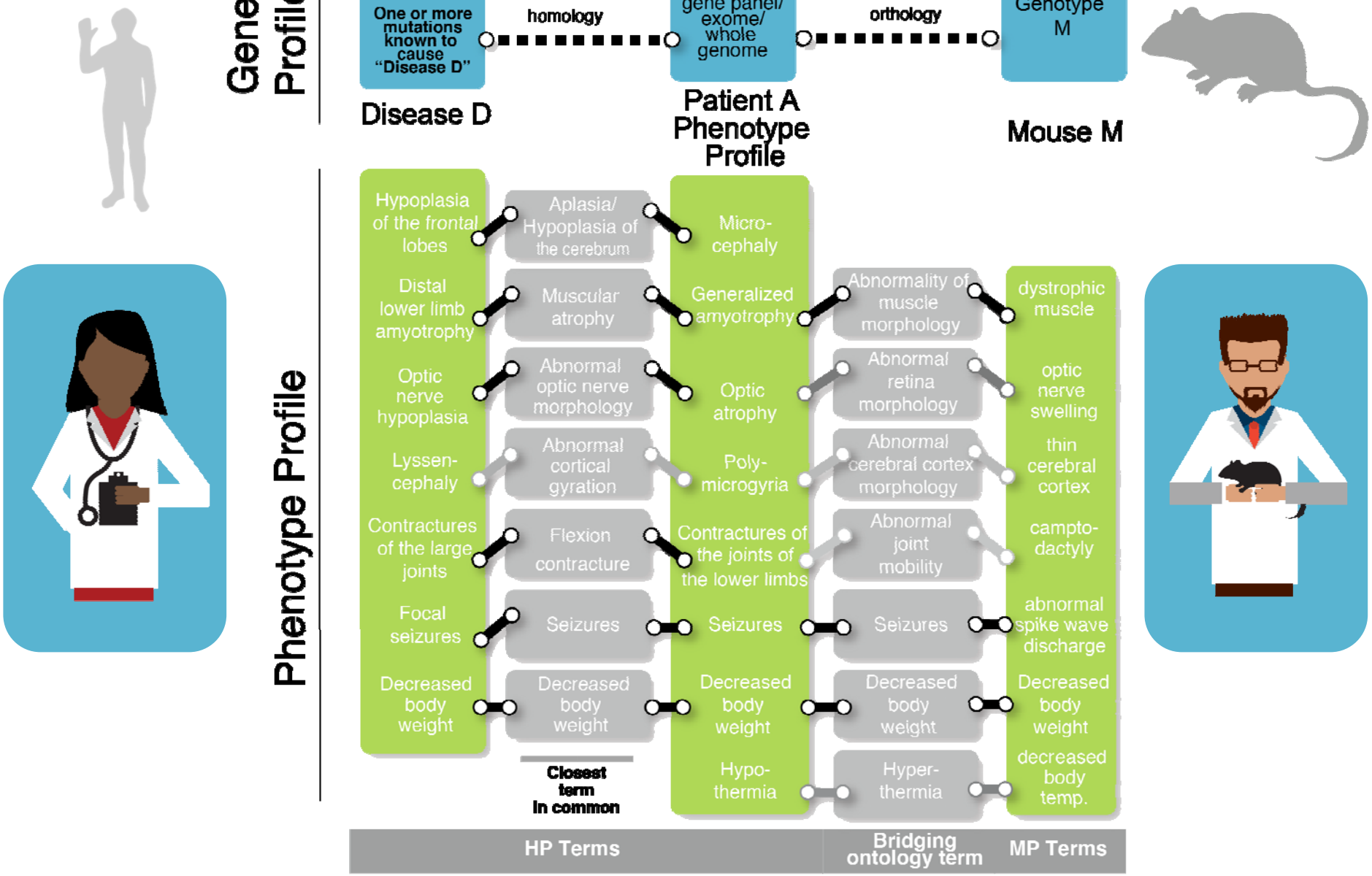




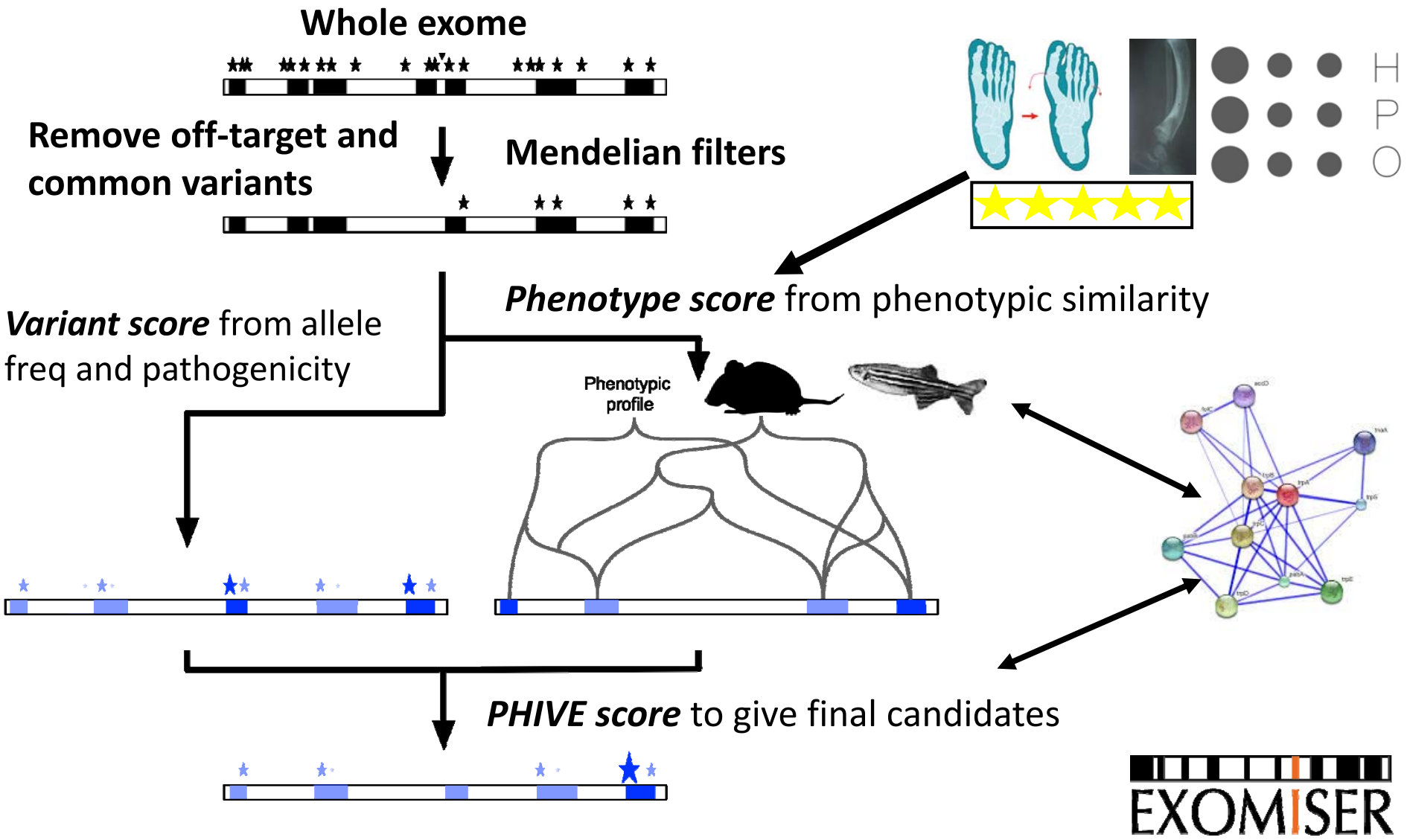
Overview

- Introduction to the 100,000 Genomes Project
- Role of clinical and model organism phenotypes
  - Clinical data collection and panel assignment
  - Automated variant prioritisation
  - Precision mouse models for validation and functional characterisation of the variants

Precision fuzzy phenotype matching

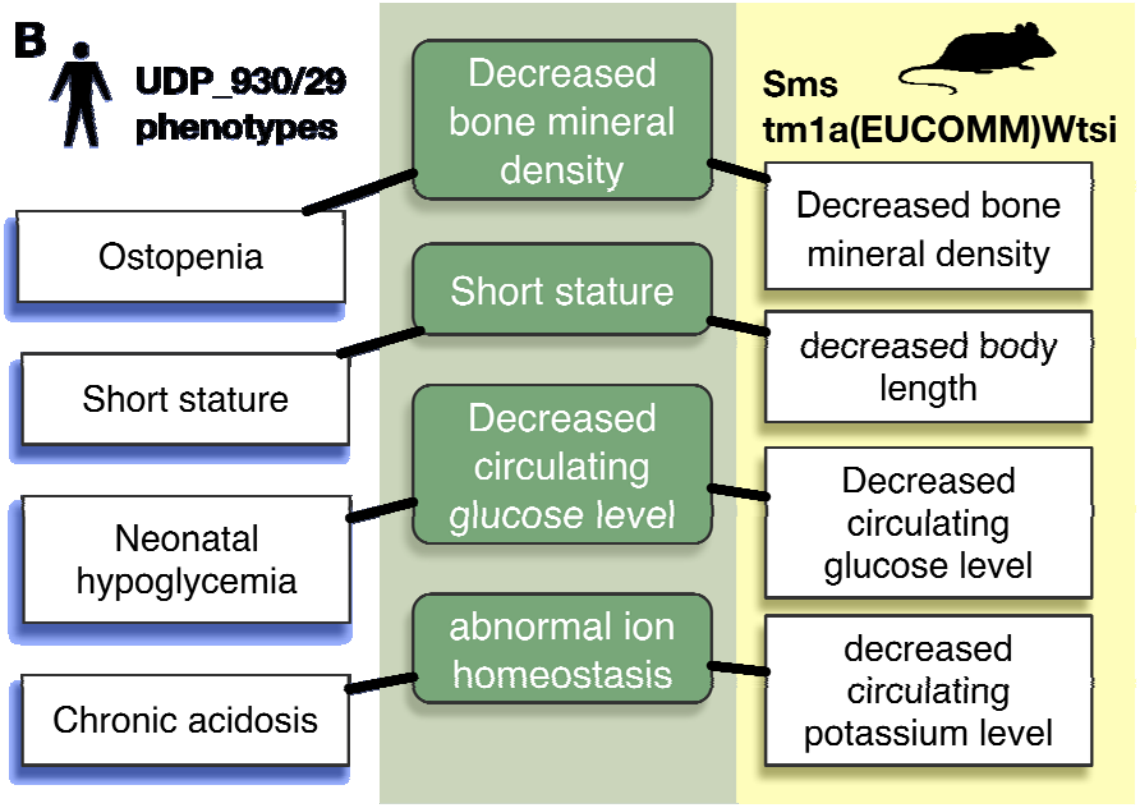


Combining G2P data for variant prioritization

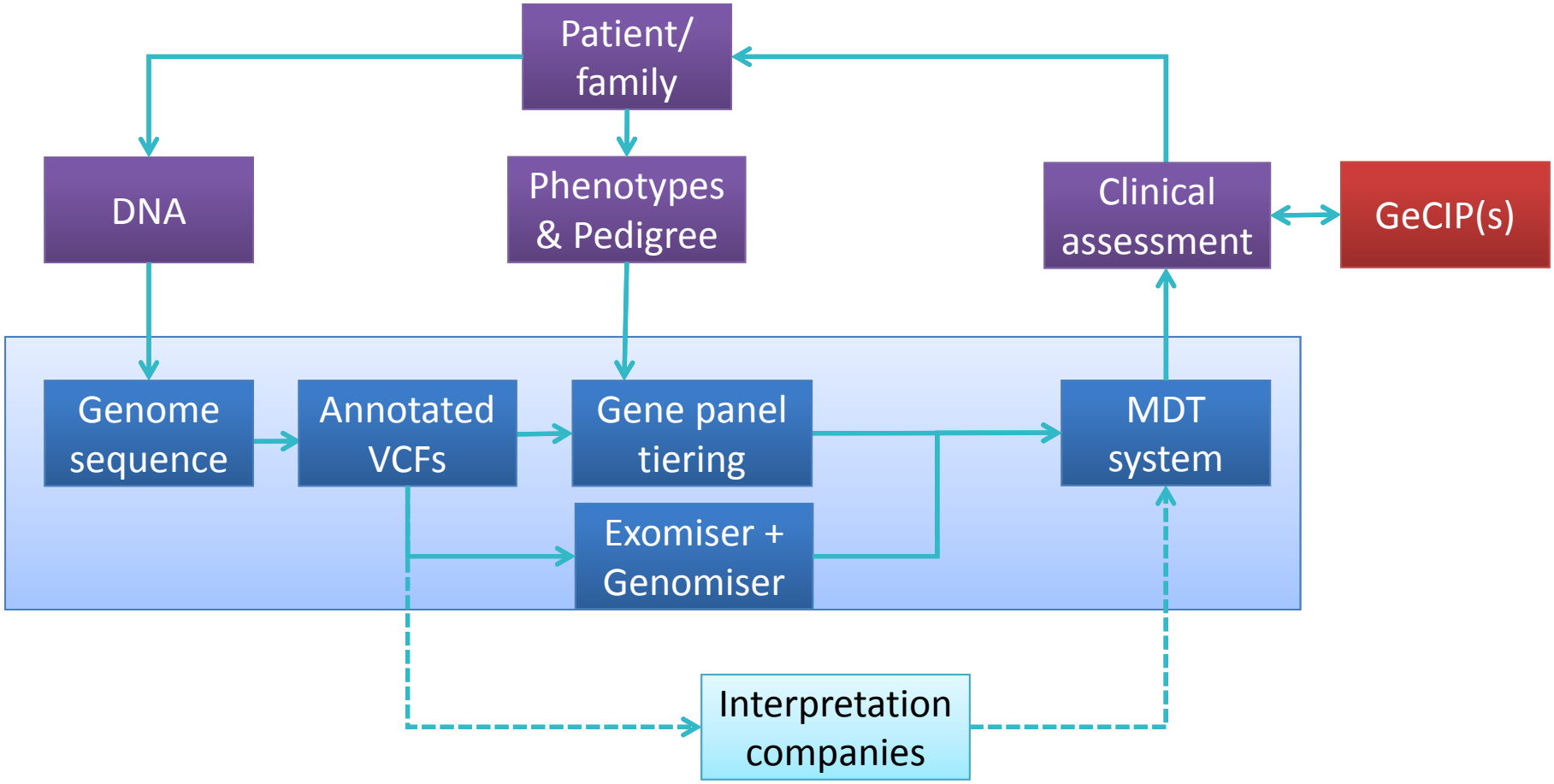


Exomiser software suite

- **How-To guide**: Smedley D et al ***Nature Protocols*** 2015 **10**(12):2004-15.
- **Exomiser**: Robinson PN et al. ***Genome Research*** 2014. **24**(2):340-8.
- **Genomiser**: Smedley D et al. ***Am J Hum Genet.*** 2016. **99**(3):595-606.
- **Undiagnosed Disease Program:** Bone W et al. ***Genetics in Medicine*** 2015.

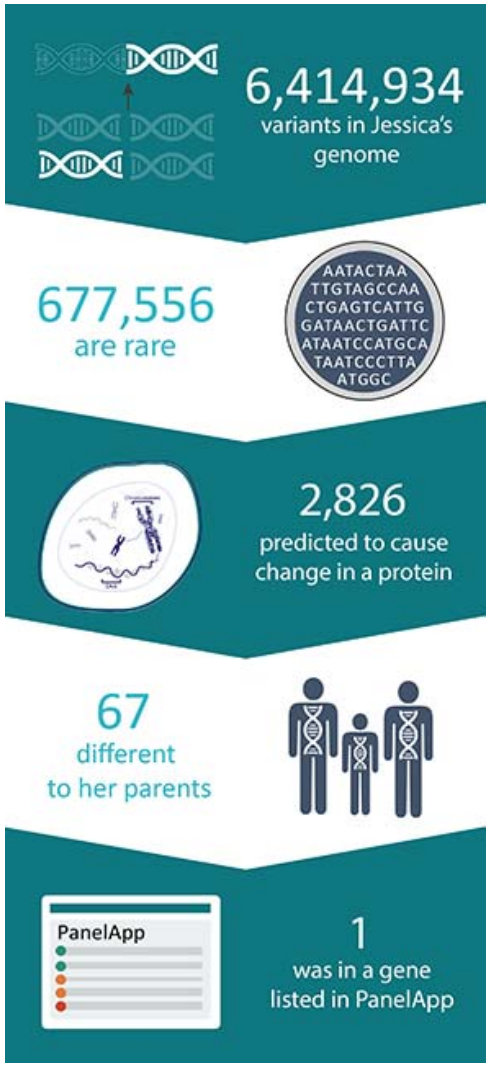


Scalable rare disease diagnostics



omicia CONGENICA WuXiNextCODE

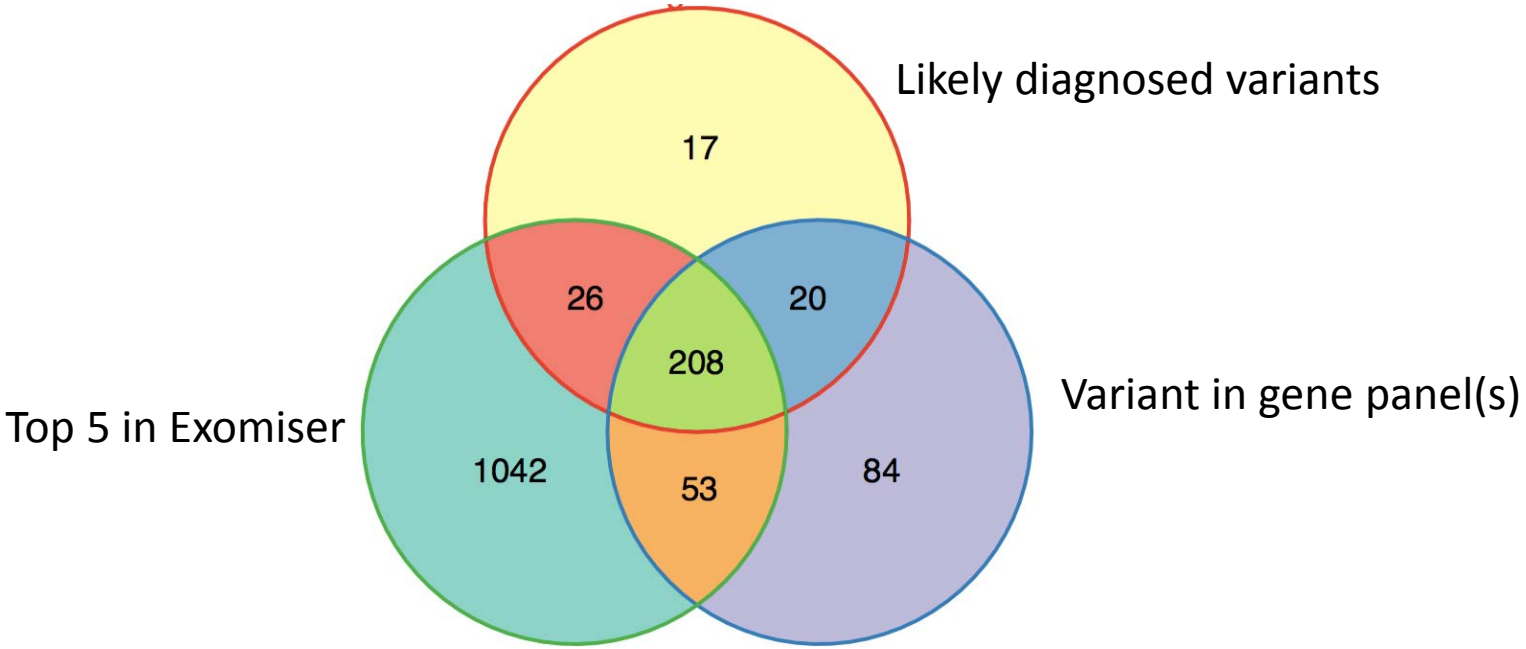
First families diagnosed



- Jessica (aged 4) has a rare condition which causes epilepsy, affects her movement and developmental delay. Standard genetics tests negative.
- De novo deletion in *SLC2A1* identified as the cause of her Glut 1 deficiency syndrome
- Now being successfully treated with a a ketogenic, low-carb diet
- Low risk for future pregnancies

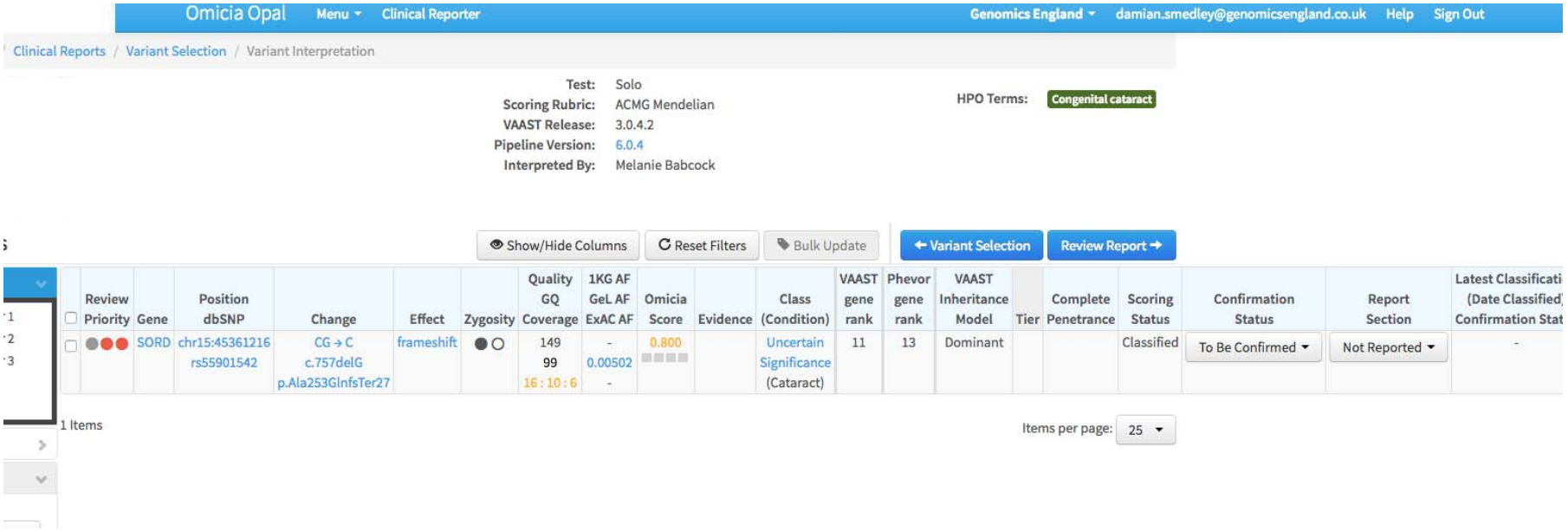
Likely diagnoses in first 1000 reports

- Top Exomiser hit in 59% of cases
- In top 5 Exomiser hits in 88% of cases
- Confirms variants found in gene panel(s)
- Identifies missing genes from applied panels



Addition diagnoses from phenotypes

- Rare, frameshift deletion in *SORD* for a patient with congenital cataracts
- Not in our panel as limited evidence in OMIM and literature
- Highlighted by Omicia's clinical team and top 5 Exomiser match based on existing mouse (spontaneous mutation removing all functional protein) used as a model of cataract development in diabetes



Research candidates

- Likely to have 10-15k rare disease cases without a clear diagnosis from standard pipeline
- Exomiser and Genomiser candidates, especially those based on **model organism phenotypes critical for new disease gene discovery**
- Interactions with GeCIP communities to validate
  - Transcriptomics
  - **Crispr/Cas9 precision animal models for functional validation and ultimately improved treatment**

Overview

- Introduction to the 100,000 Genomes Project
- Role of clinical and model organism phenotypes
  - Clinical data collection and panel assignment
  - Automated variant prioritisation
  - Precision mouse models for validation and functional characterisation of the variants

Aims and Objectives

Use sophisticated tools developed within IMPC to further support genome engineering

Make the technical skills of Harwell available widely

Shorter timescales responsive to scientific and clinical needs

Offer a wider range of genetically quality controlled lines to UK scientists

Nominations from UK Clinicians and academics

Call 1= 87

Call 2= 93

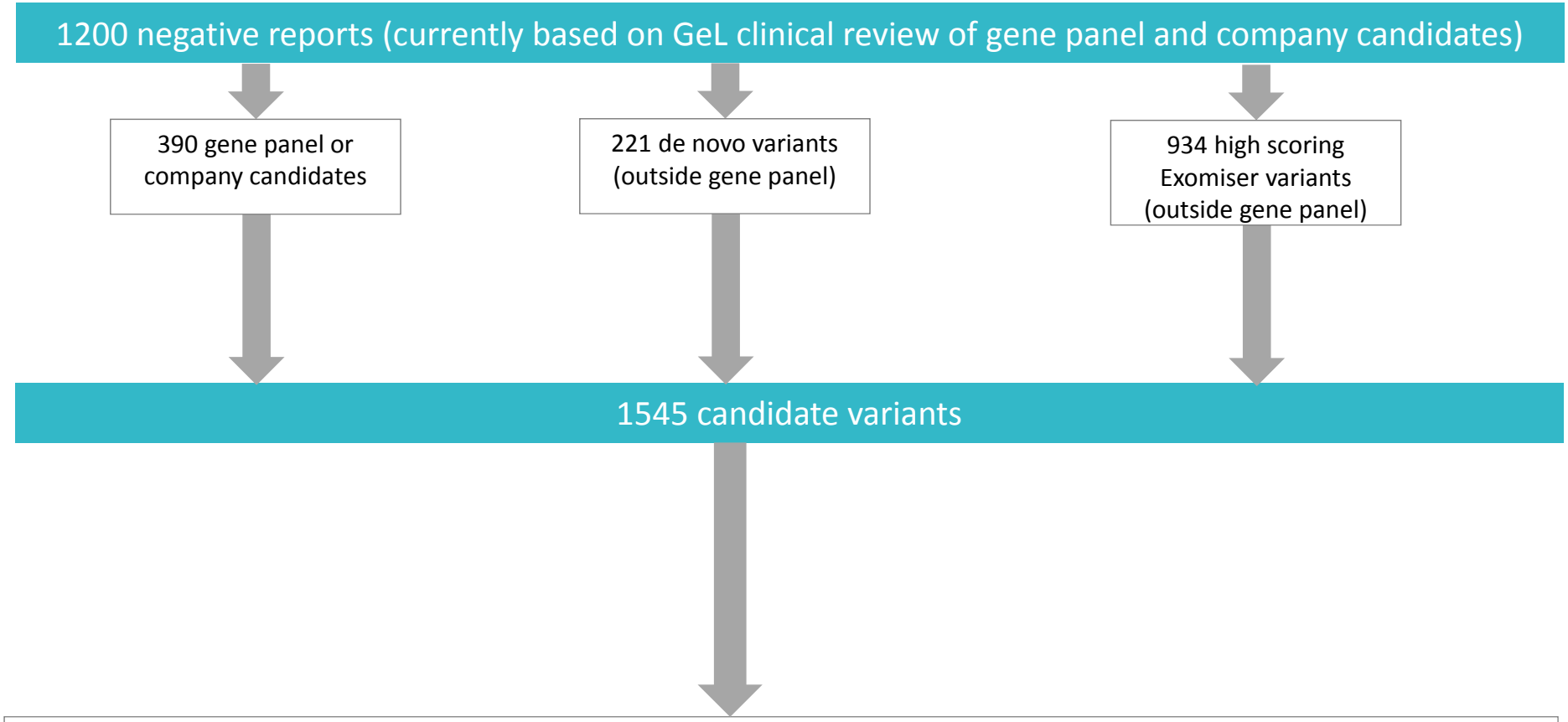
Peer reviewed by a panel of experts

Nominations scored by MRC Harwell for

1) Resource
2) Feasibility

Mouse strains delivered but also available in public archives

Research candidates for functional validation



Call 1- Summer 2016

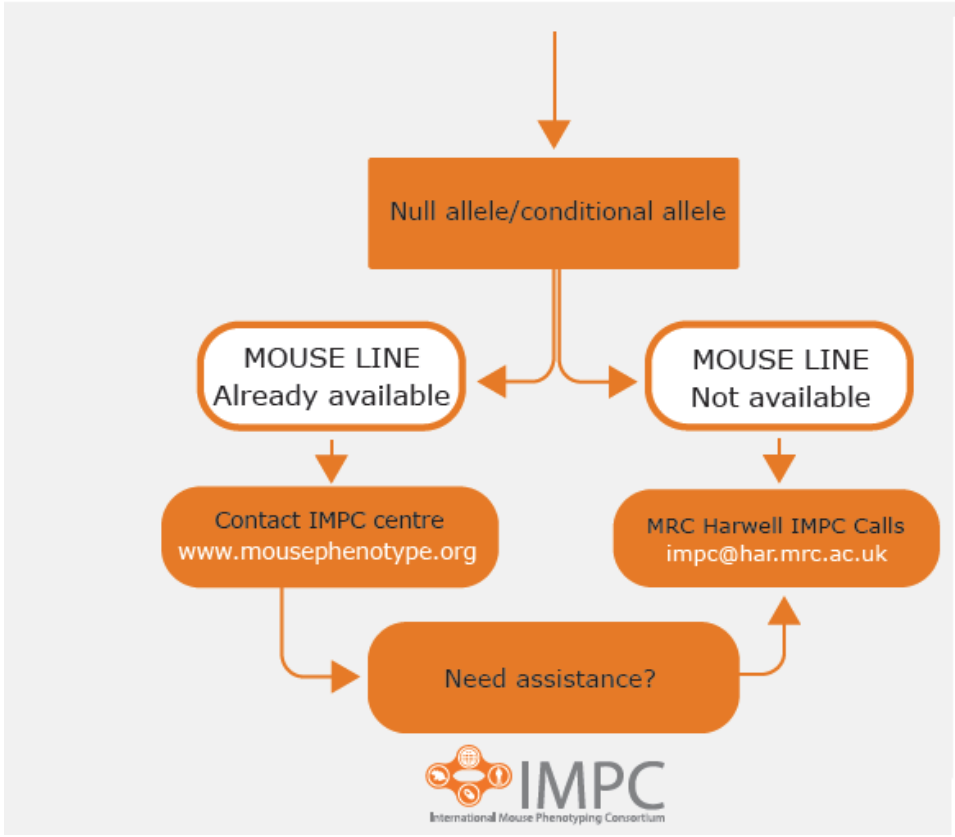
- 87 application
- 23 selected
- 12 point mutations

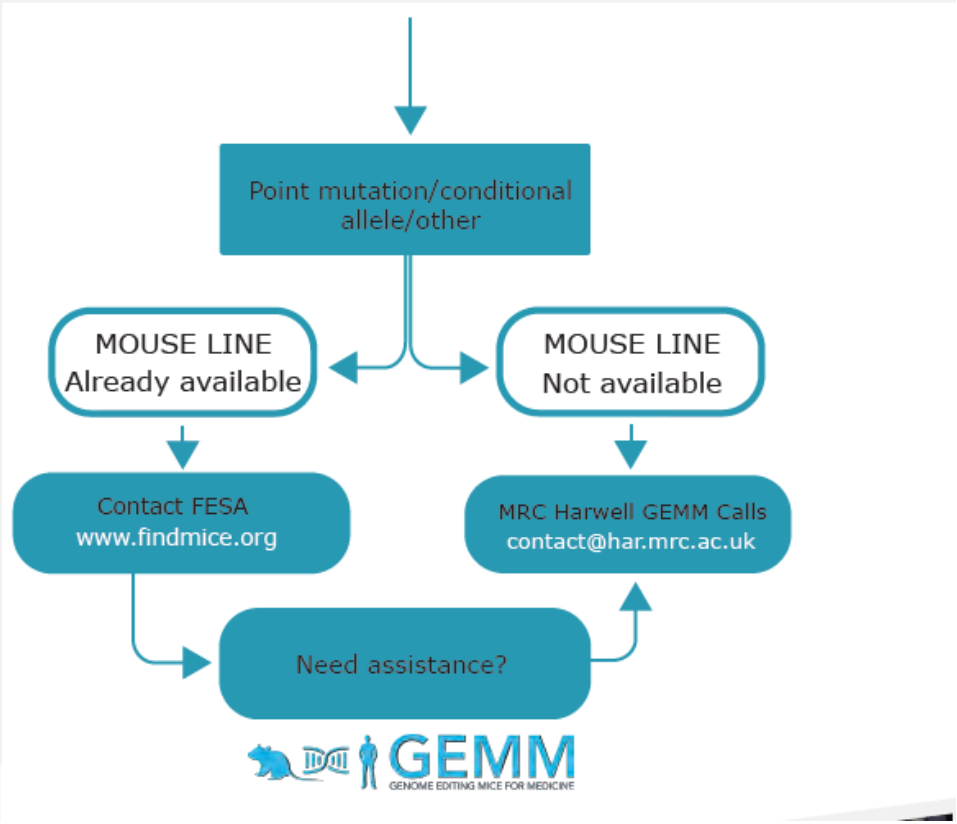
Call 2- Spring 2017

- 93 application
- 19 selected
- 12 point mutations

X-linked intellectual disability
Non-ketotic hyperglycinemia
Thyroid Cancer
Pain
Charcot-Marie-Tooth
Cardiac Myopathy
Epilepsy
infantile-onset dyskinesia and chorea

Opportunities for Mouse Strains





Building the future of genomic medicine

Clinical and model organism phenotypes as well as precision animal models are key to achieving this vision

- 100,000 WGS on NHS patients and pathogens
- The NHS, academics and industry partnerships at the outset to drive Genomic Medicine into the NHS and create wealth
- Building the human capacity and capability
- Key international partnerships to add value
- NHS is preparing to commission WGS from April 2018 as part of a re-procured Genomic Health service
- Concentrating the UK Genomics Knowledgebase in one location
- New diagnostics and therapies and opportunities for patients

Acknowledgements

- 100,000 Genomes Project
  - The patients and their families
  - NHSE staff
  - Genomics England colleagues
  - UK biobank and Illumina
  - Interpretation companies
  - GeCIP members