Genomics england

# The 100,000 Genomes Project
# Transforming Healthcare

Berlin Institute of Health

Prof Sir Mark Caulfield FMedSci

Chief Scientist

William Harvey Research Institute
Queen Mary University of London

# Disclosures

- Genomics England is a Department of Health Company
- Seconded to Genomics England from Queen Mary/Barts who pay my salary
- Multiple industry partnerships e.g. Illumina, iQVIA
- No shares in anything except failed banks in 2008

# The 100,000 Genomes Project Milestones

Announced by David Cameron, former Prime Minister in December 2012 –An Olympic Legacy

Genomics England launched by then Secretary of State for Health in speech during NHS 65th Anniversary Celebrations, July 2013

Opening of new Sequencing Centre by Theresa May in 2016

illumina

CMO's Generation Genome and the Life Sciences report in 2017

NHS England

Commissioning of new NHS Genomic Medicine Service October 2018

Reached goal of sequencing 100,000 genomes in December 2018



"aspiration to undertake 5 million genome analyses over the next 5 years"

# The 100,000 Genomes Project in numbers

Over **100,000** genomes

Over **97,000** patients and family members


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

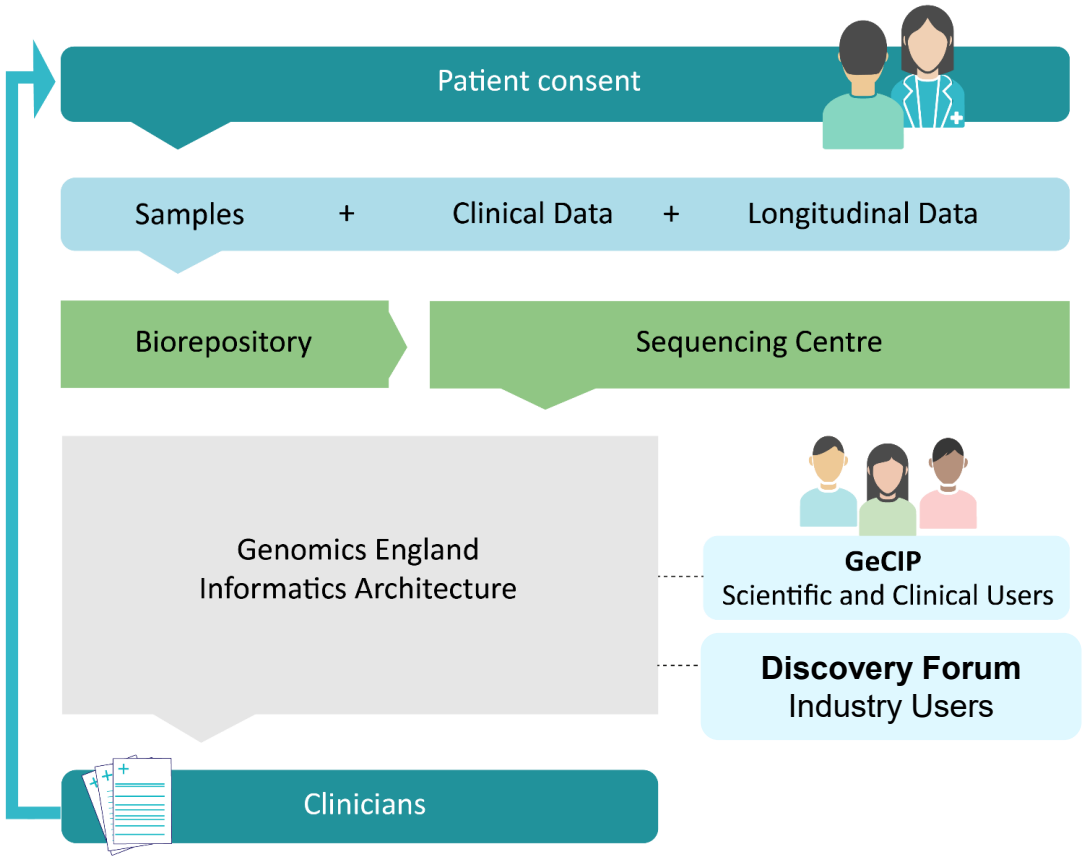
**13** Genomic Medicine Centres, and **98** NHS Trusts within them were involved in recruiting participants

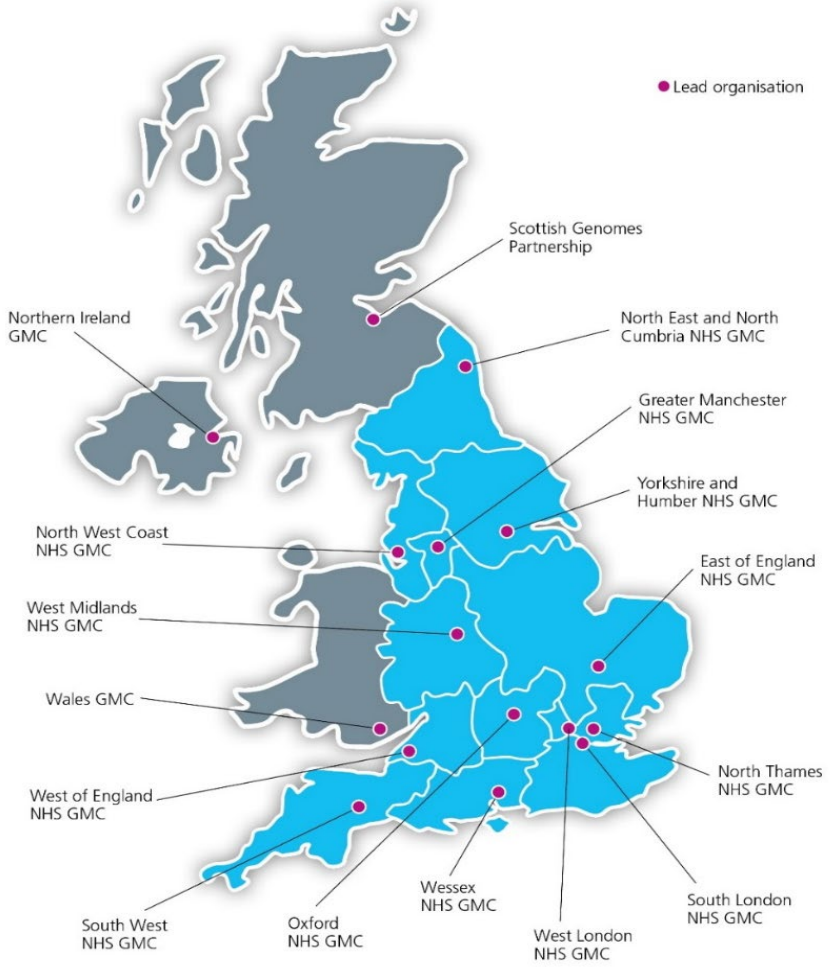
Around **5,000** NHS staff (doctors, nurses, pathologists, laboratory staff, genetic counsellors)

Over **3,000** researchers and trainees

# How did the 100,000 Genomes Project work

- 13 NHS Genomic Medicine Centres covering England, over 98 hospitals
- Responsible for identifying and recruiting participants and for clinical care following results
- Northern Ireland, Scotland and Wales joined

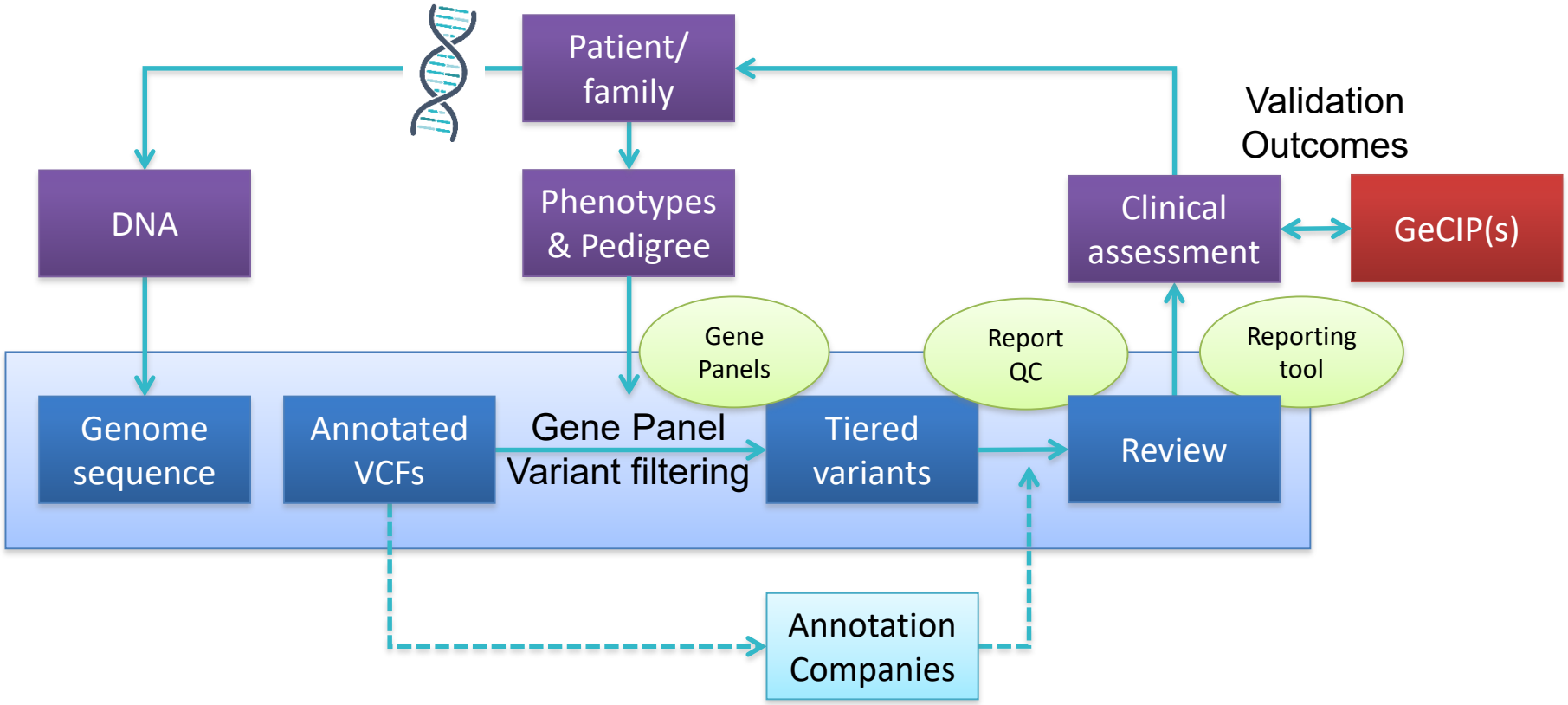


# Scalable disease diagnostics

## Sequence depth germline 36x to 40x

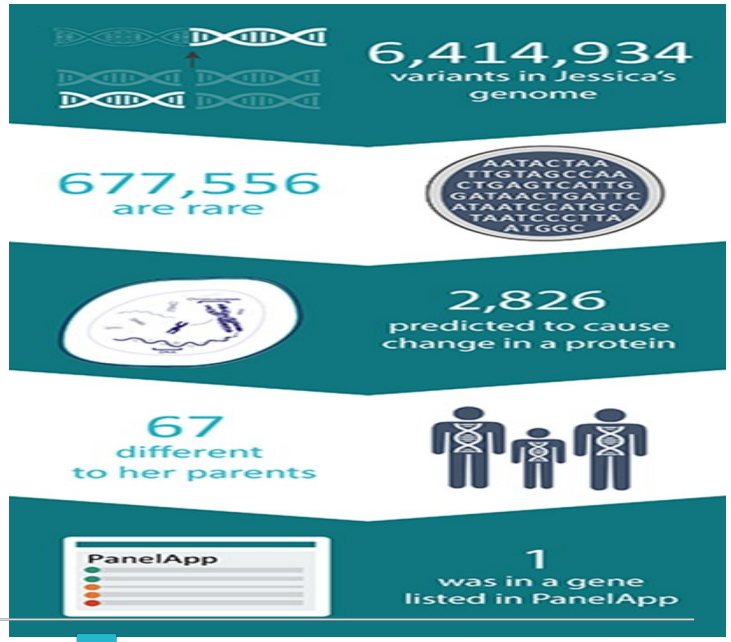
## Somatic 82x to 100x

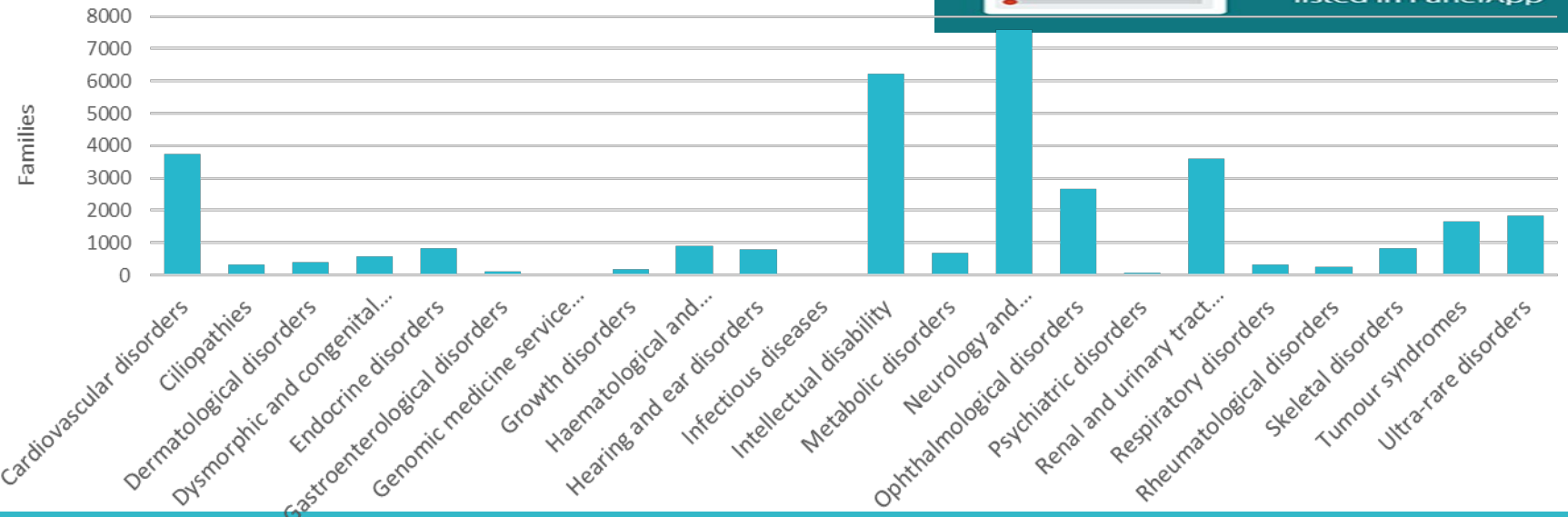


# Rare diseases

# Rare Inherited diseases

- <6% of the UK population
- 1200 disorders unmet need
- Standardised eligibility & phenotyping
- Human Phenotyping Ontology
- Automated analytics
- NHS confirm gene panels & close cases



# Application in the NHS

## 4 month old fluctuating neurology, repeated infections in NICU

- Died with no diagnosis, immune testing negative, enrolled in the project by parents.
- Mother unexpectedly pregnant didn't want to know diagnosis
- We found a pathogenic Transcobalamin 2 mutation
- Brother born and tested – sadly affected BUT high dose B12 given
- J Paediatrics 1974

## 10 year old girl admitted to ITU with life threatening chicken pox

- Prior unusual severe infections.  Detailed immune testing no diagnosis.
- Mutations in *CTP synthase 1* gene affects B and T lymphocyte responses to infection of both capsulated bacterial infection and viruses
- Curative bone marrow transplant- Siblings tested and not at risk of these infections
- *J Allergy Clin Immunol 2016 Vol: 138: 6*

## 5 year old boy unexplained anaemia, developmental delay, short stature and constipation

- ?Diamond-Blackfan anaemia, may have limited lifespan due to cancer risk
- 100,000 Genomes Project Intellectual Disability Panel via Panel app
- Tier 1 *de novo* variant identified in *Thyroid Hormone Receptor Alpha*
- Now receiving high dose thyroid hormone replacement
- *Nature Reviews Endocrinology* **volume 10**, 582–591(2014)

# 5 year old with a rare disease

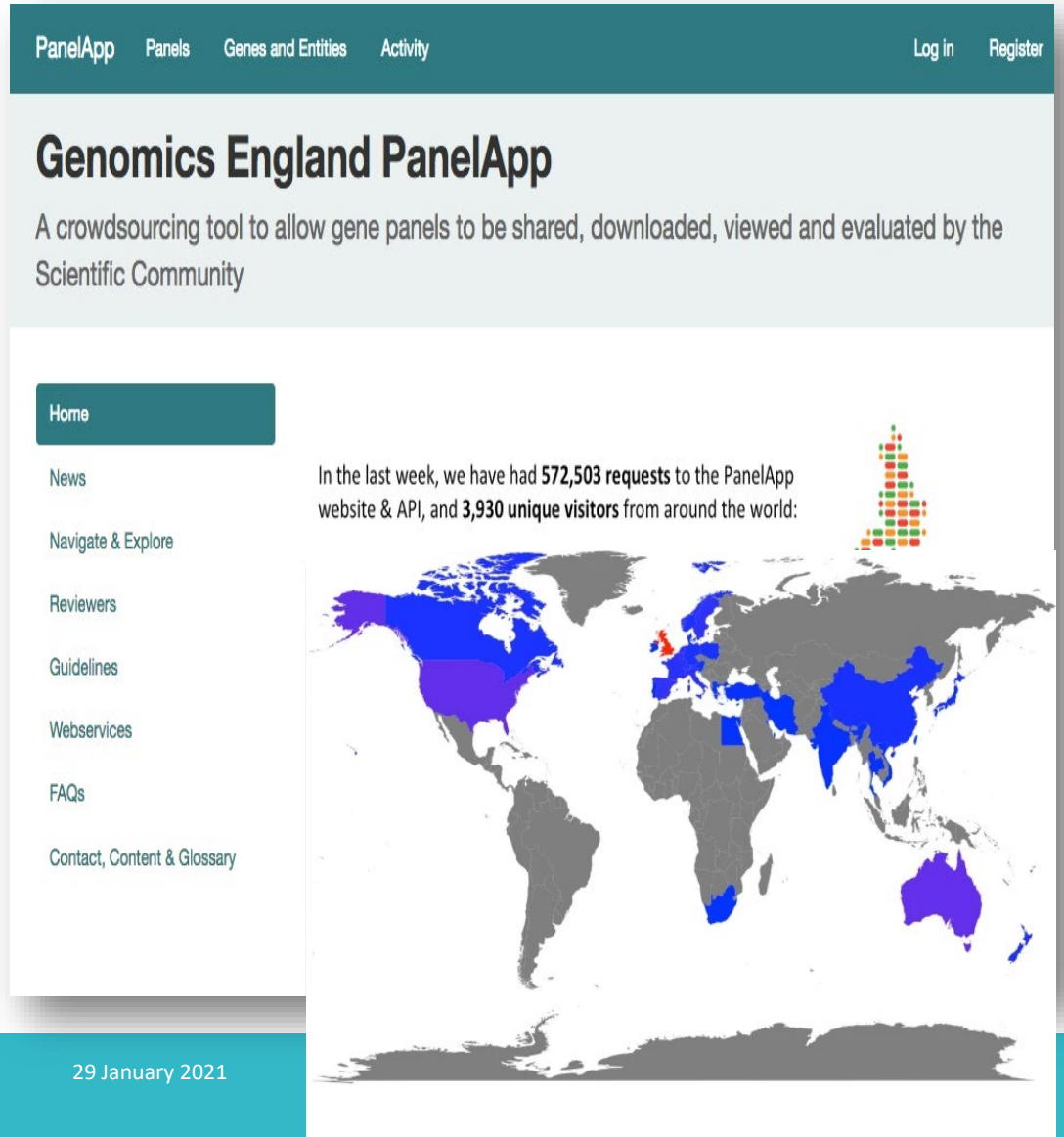
From Mum:

- “Thanks to the 100,000 Genomes Project my 5 yr old has been diagnosed with GAMT deficiency which is a treatable metabolic disorder.
- Before diagnosis it was thought she had a degenerative neurological disorder and I was waiting for her end up in a wheelchair.  She had uncontrollable epilepsy, the developmental age of a 12 month old and couldn’t retain any skills, she was very much locked inside her mind and unable to communicate in any way.
- After 6 months of treatment later with creatine, ornithine and arginine she is a new child, the light is back in her eyes, she’s learning new skills every day and she is free of epilepsy finally.
- It has been nothing short of a miracle, I’ve spent so long waiting for her to get even worse or even die from her epilepsy, it doesn’t feel real that she is improving and thriving.”

https://mamaunexpected.com/2018/02/12/1904-days-d-is-for-diagnosis/

# PanelApp

https://panelapp.genomicsengland.co.uk/




**Step 1**
Creation
of initial gene panel
for each disorder

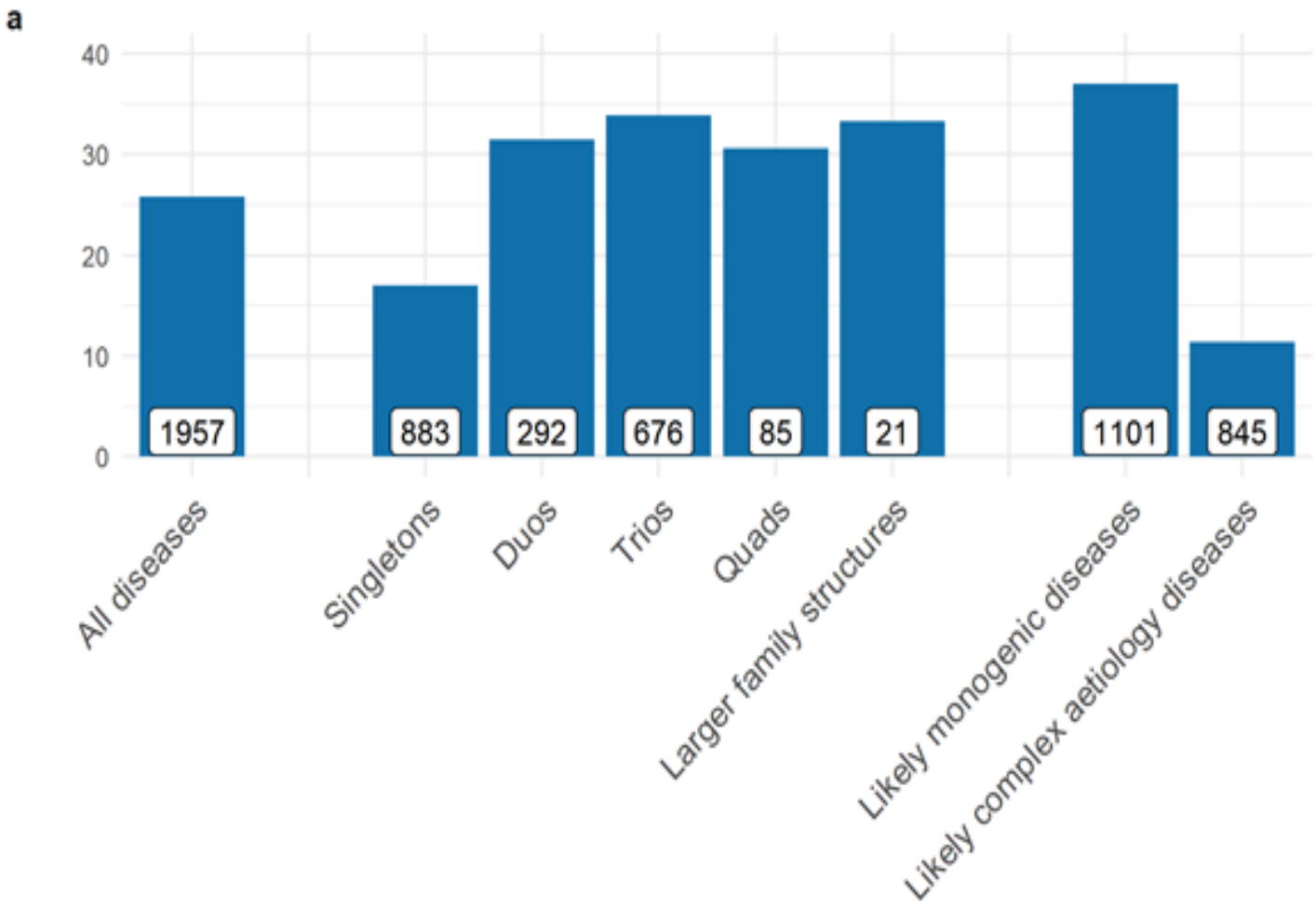
**Version 0 gene panel**

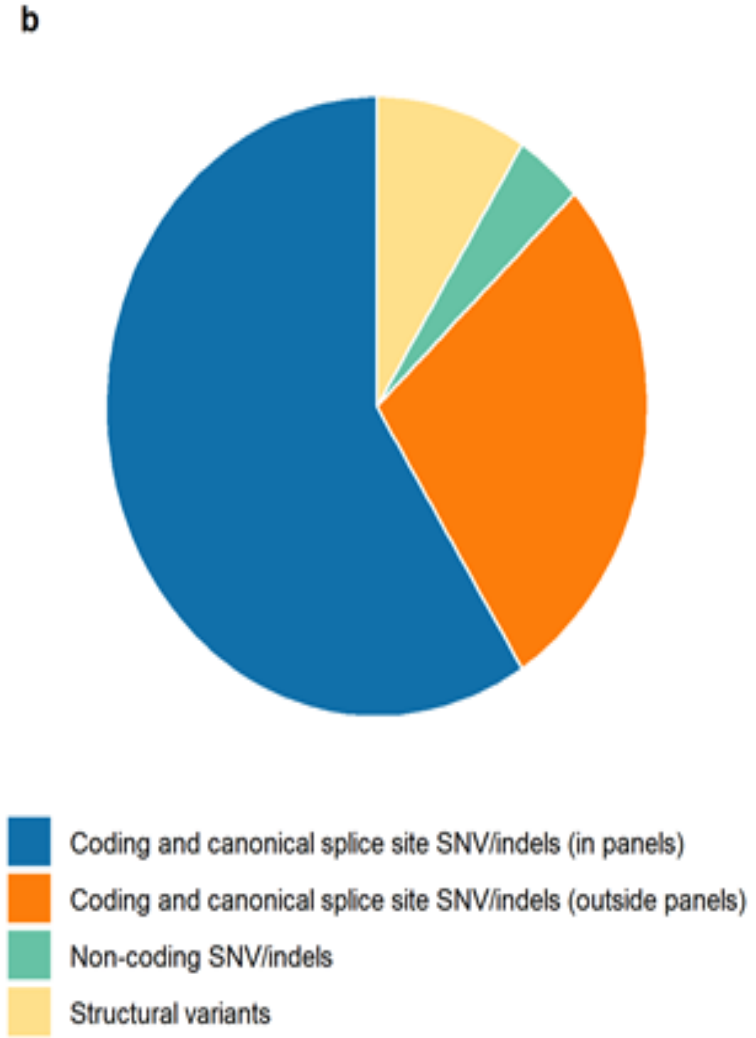

**Step 2**
Crowdsourcing
of
Expert Knowledge

**Step 3**
Consolidation
of reviews
& further curation

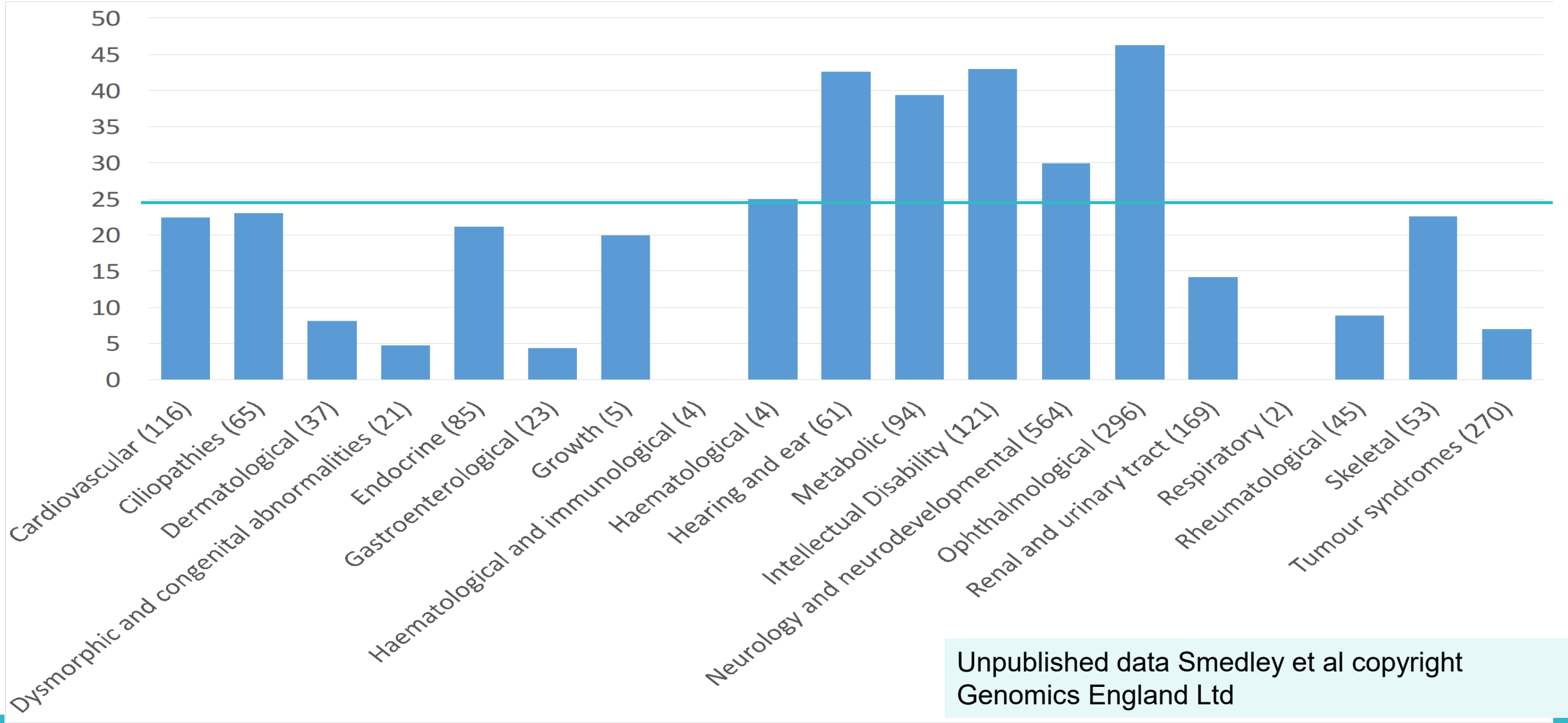
*Use other resources to verify reviews:*
OMIM
Gene2Phenotype
Orphanet
ClinGen

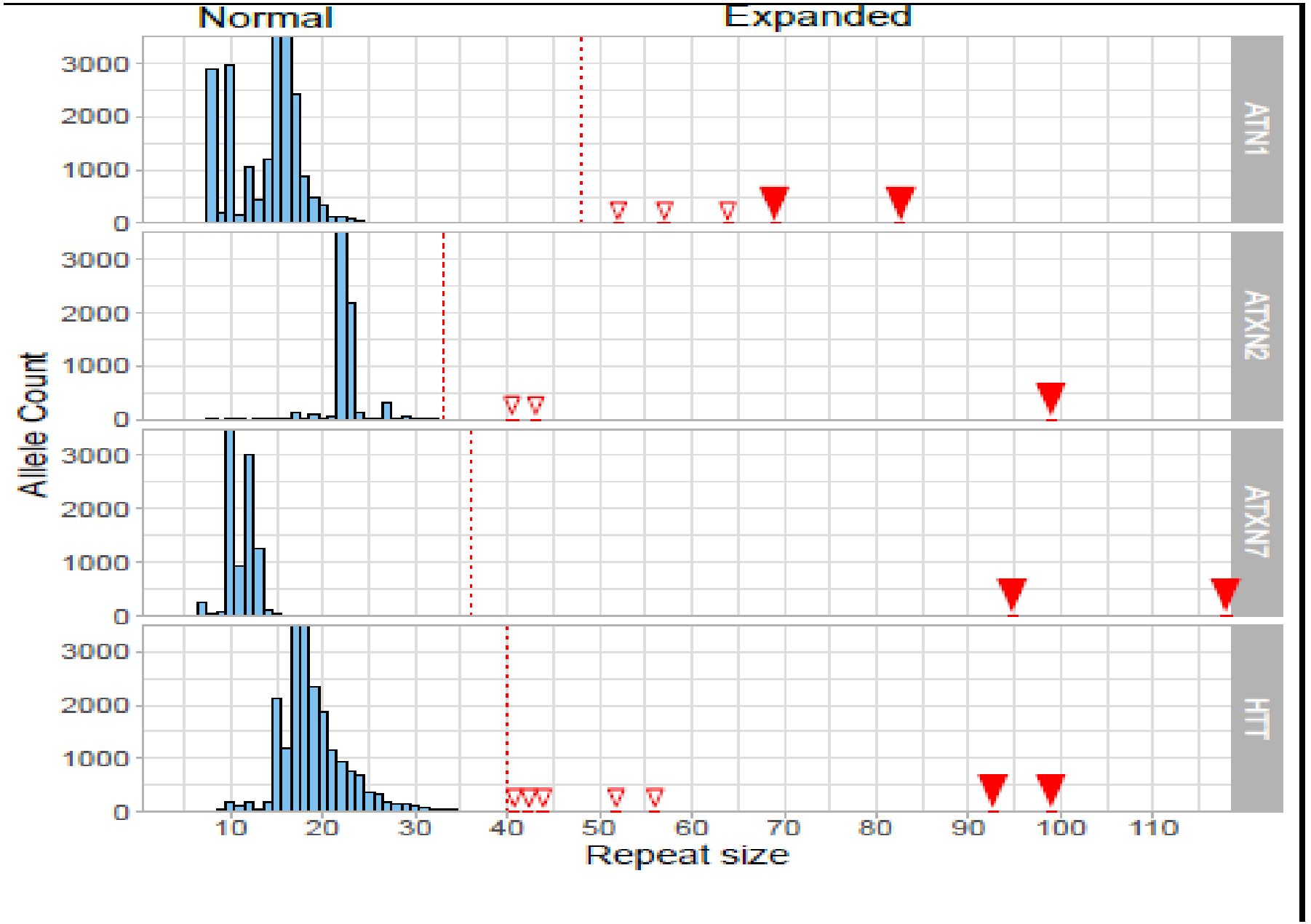
# Rare Disease Diagnoses and Family Size
# 2183 families from 160 rare disease categories in 4660 people



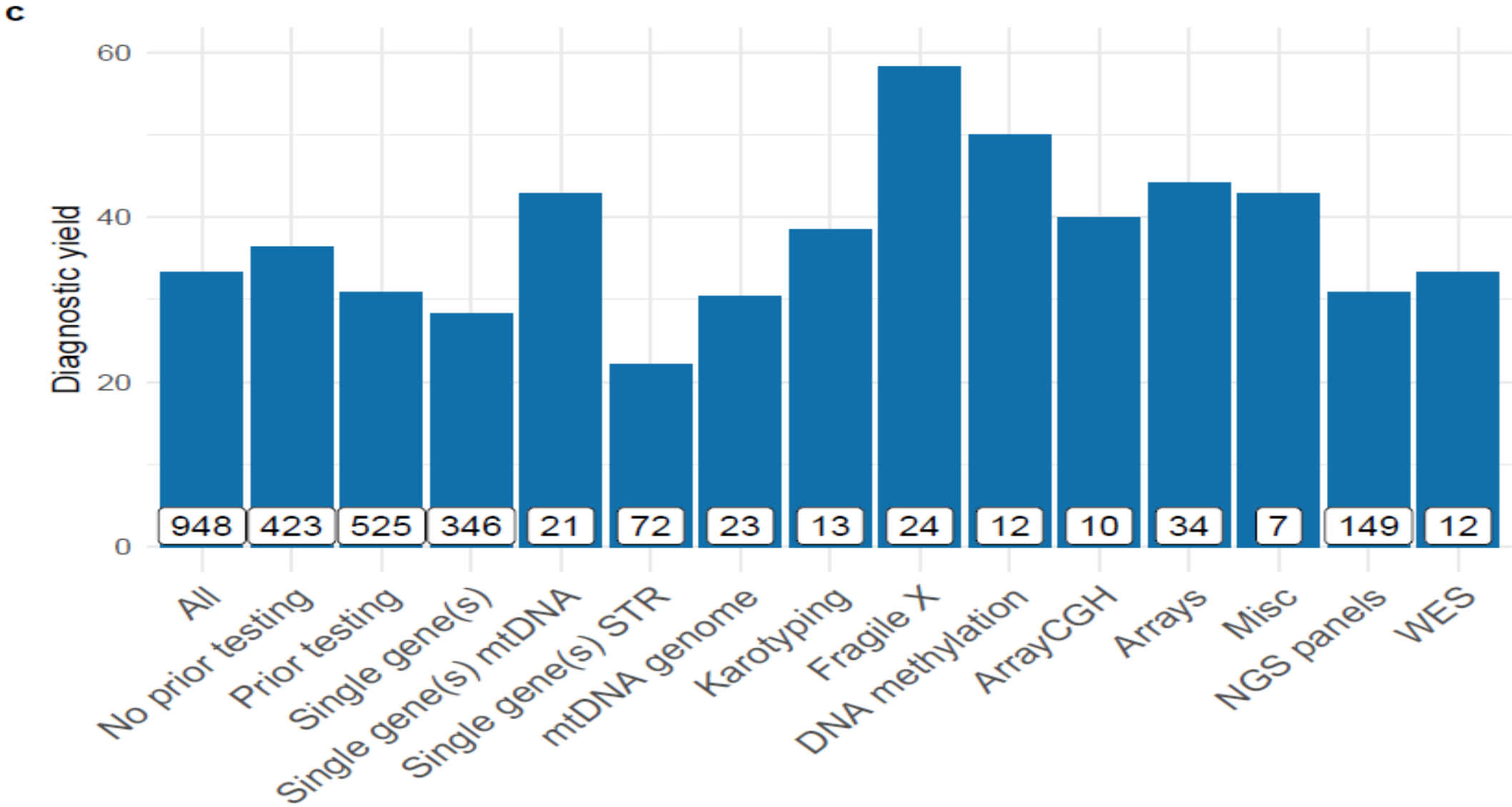



# Confirmed Diagnostic Yield by Category

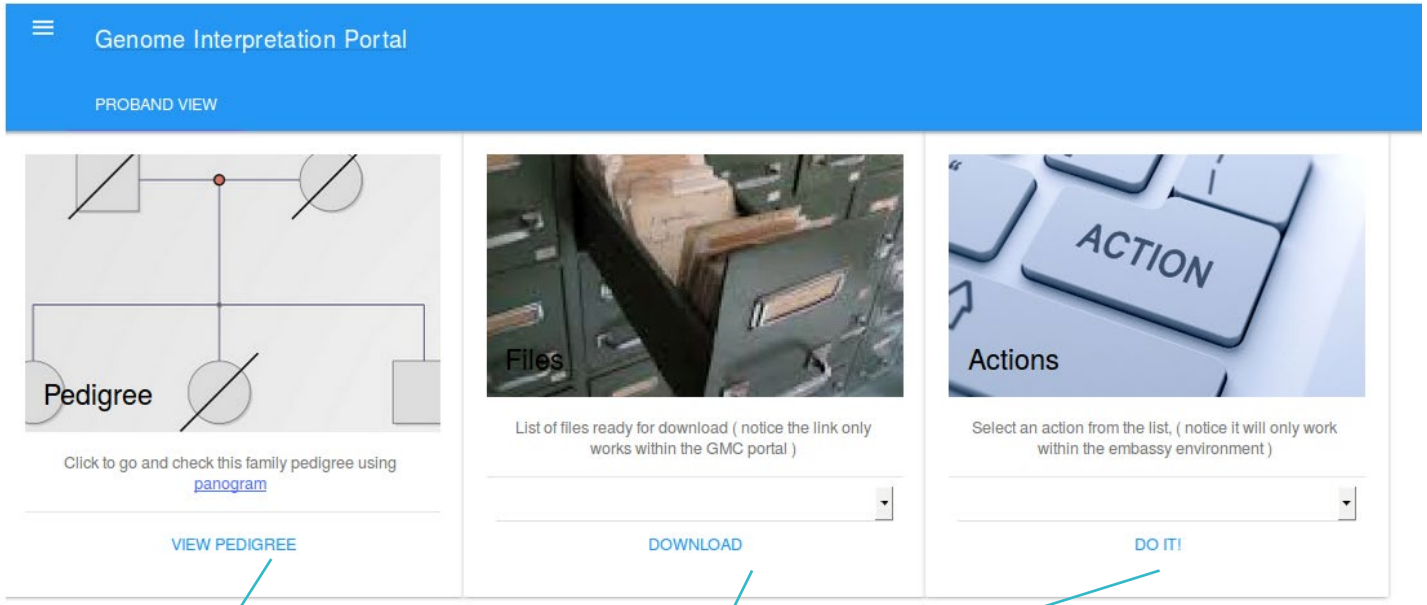



Normal
Expanded
ATN1
ATXN2
ATXN7
HTT
Allele Count
3000
2000
1000
0
10
20
30
40
50
60
70
80
90
100
110
Repeat size

# Diagnostic uplift from whole genomes over usual care



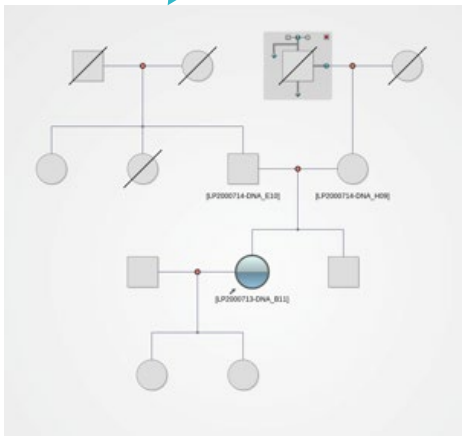
# NHS Reporting portal

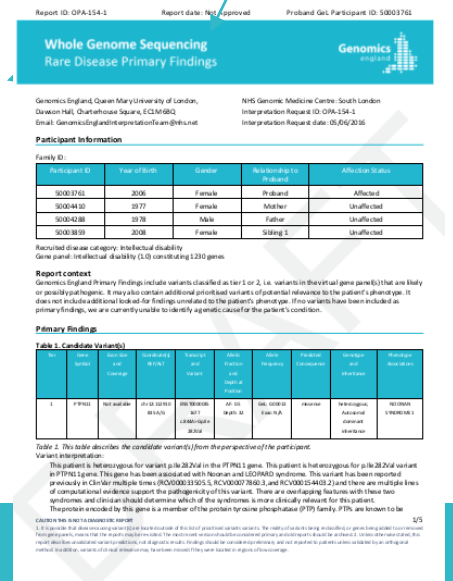


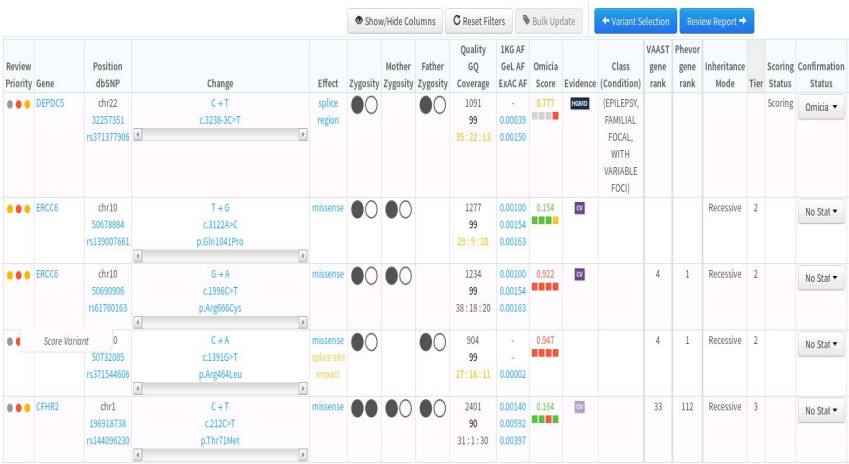

View family pedigree

Download the report

Review variants and close case





# Healthcare Benefits – NHS feedback & druggability tests

- 24% of diagnoses had immediate clinical utility
- 0.2% no current benefit
- The remainder had unknown utility at this early stage.

## Headline clinical utility

- 3 probands changed medication
- 20 diagnoses leading to additional surveillance for the proband or relatives,
- 13 eligible for clinical trials,
- 52 - the diagnosis could inform future reproductive choices
- 32 - other benefits e.g. dietary change or vitamins

## Druggability assessments

- 33% of the diagnoses have an existing approved drug that targets the protein

## Cohort-wide burden tests

- 500 genomewide significant signals – potential diagnoses

# Diagnostic odyssey of children born 2003 onward

- **Families spent 6 years** (median 75 months) attended a median of 68 hospital appointments prior to diagnosis
- Unaffected relatives attended a median of 18 appointments over 120 months from birth.
- Post-diagnosis over 18 months, fewer focused clinical episodes
- **Affected participants used 183,273** episodes of hospital care via the emergency department, outpatients, inpatients and critical care,
- **Cost £87 million** (median cost of £15,310 per participant)
- **Compared to 53,706 episodes at a cost of £21 million** (median cost of £4,285/participant) for the unaffected participants
- Not including visits to the family physician, or disease treatment costs.

# Rare Disease Pilot – 240,000 Hospital Episodes

## Total Cost of Hospital Episodes 1CD10 coded since 2007

| Dataset | Number of episodes | Total cost |
|---|---|---|
| Critical care | 347 | £387,085 |
| Accident and Emergency care | 16,696 | £2,530,399 |
| Inpatient care | 43,714 | £77,276,982 |
| Outpatient care | 177,125 | £26,243,354 |
| **TOTAL** | **237,882** | **£106,437,820** |

$138,189,286

Unpublished – Buchanan and Wordsworth

# Clinical Variant Ark

## CVA in Numbers

**3,156,215**
Variants

**36,278**
Cases

**6,613**
Phenotypes

## Overall diagnosis

**20.14 %**
Current diagnostic yield*

**5,921 cases**
Positive primary findings

**23,482 cases**
Negative primary findings

**1,814 cases**
Inconclusive findings

*number of positive primary findings against total number of non inconclusive cases

## Pathogenicity of variants

**3,969**
Pathogenic

**2,926**
Likely pathogenic

**4,458**
VUS

**495**
Likely benign

**225**
Benign

# Cancer

# Making Whole Genomes Work for Cancer

Cancer is a disease of disordered genomes

## Tested Vacuum packaging and refrigeration

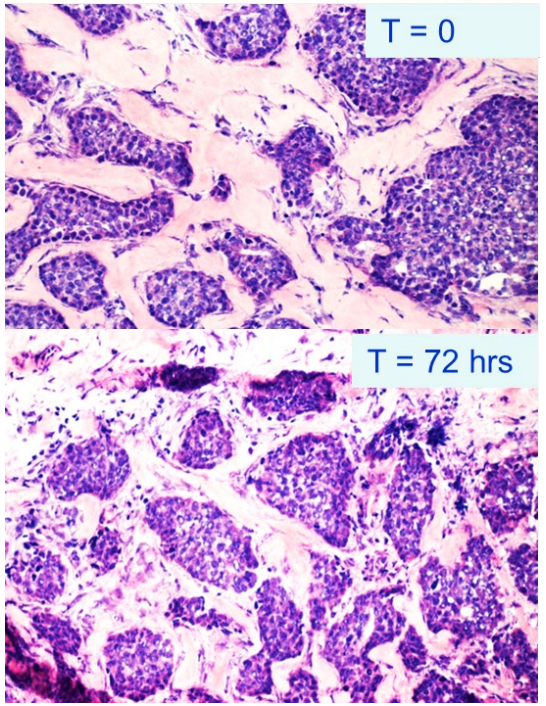
- Conclusion: Histomorphology maintained out to 72 Hrs
  Out of hours storage as fresh tissue perfectly possible

## Compared same cancer/same patient handled differently

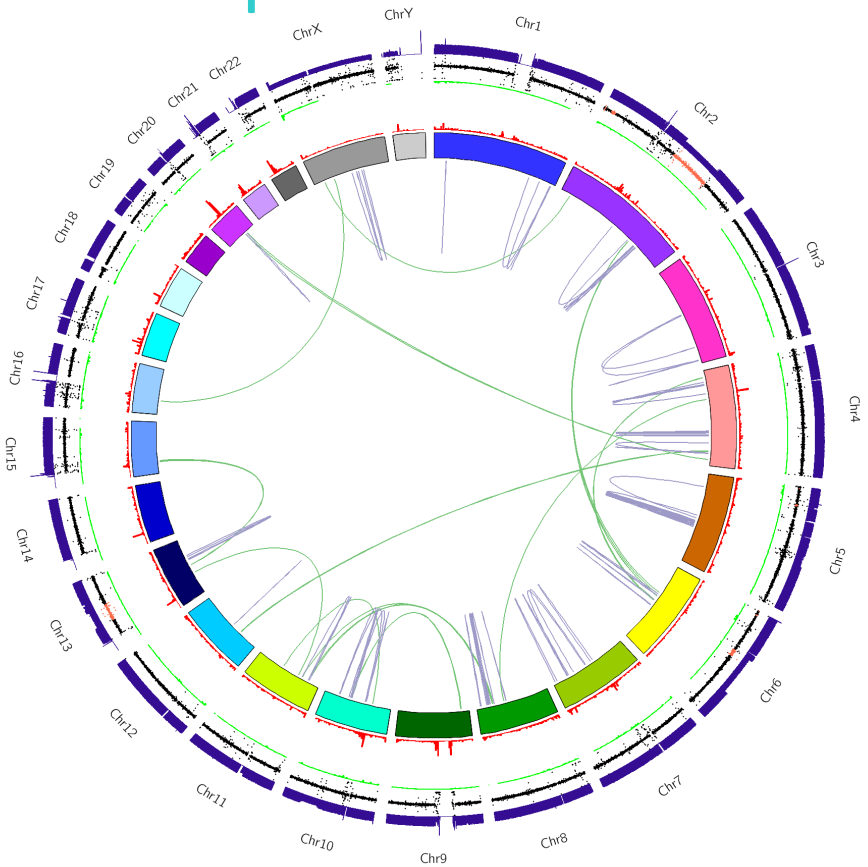
- Even optimised FFPE still worse than fresh tissue for WGS

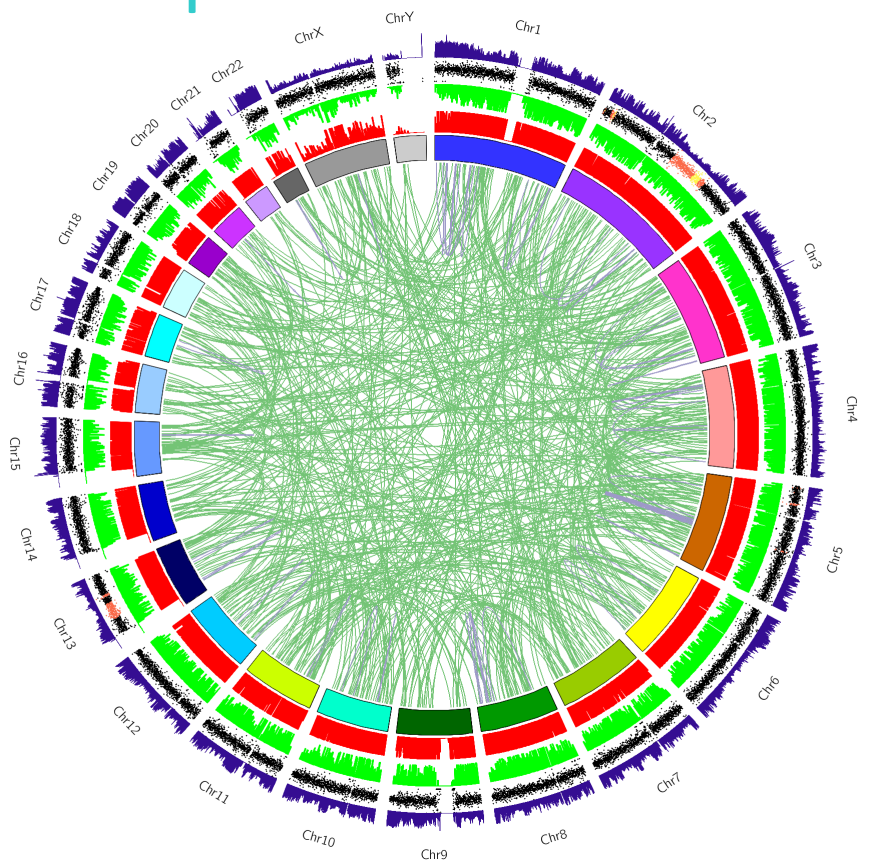
## Tested and mobilised 400 molecular pathology pathways

- Mainstream supply of fresh frozen tissue
- PCR free genomes with 750 ng input DNA
- Alternative fixatives e.g. Paxgene
- Shaken biopsy for EBUS
- Genomic and a Pathology biopsy



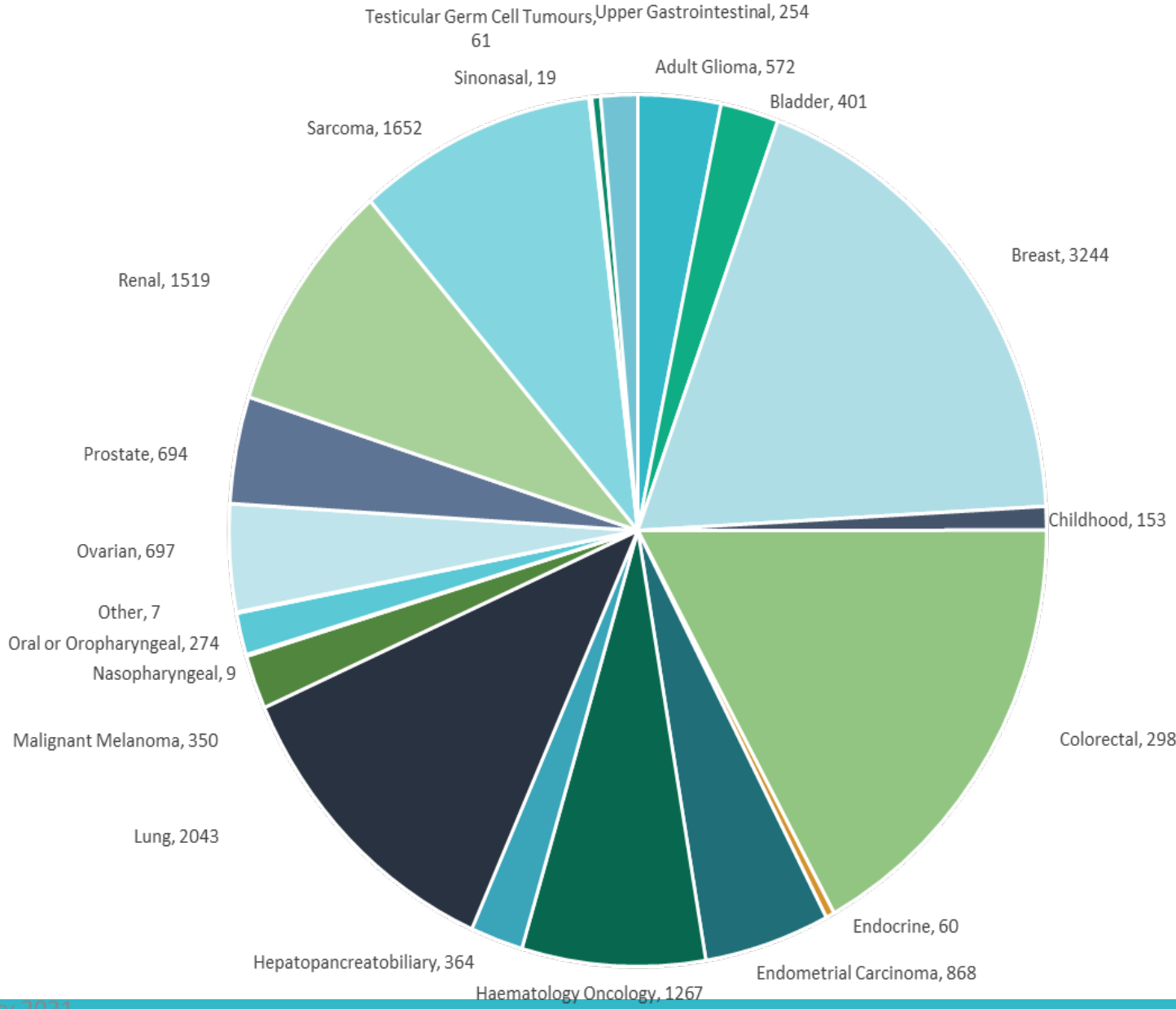
# An overview of somatic changes in FF and FFPE samples taken from the same prostate tumour



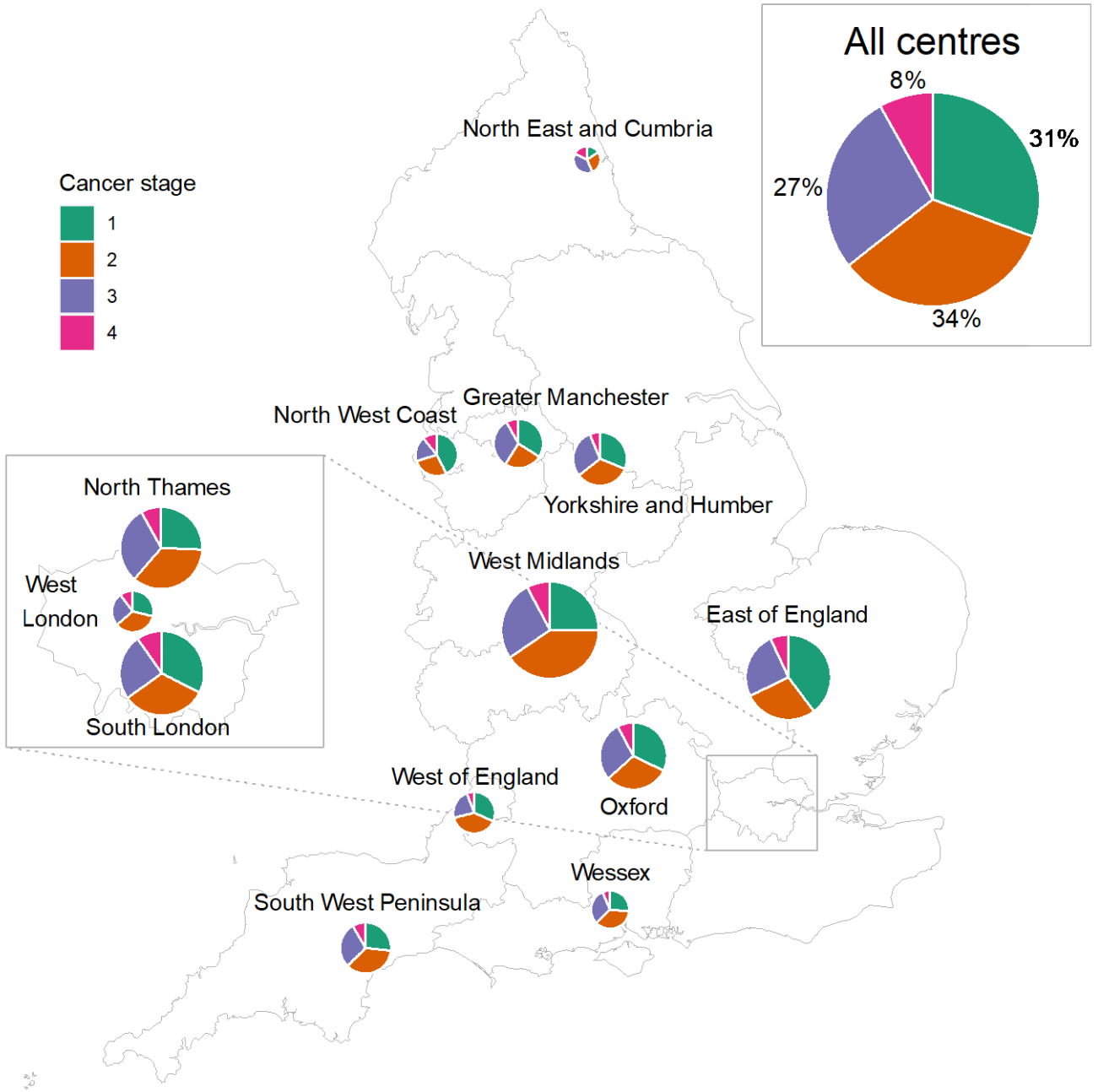

| | AT dropout | CG dropout | Evenness of coverage | Chimeric reads, % | SNVs | Indels |
|---|---|---|---|---|---|---|
| GL | 2.61 | 1.91 | 6.77 | 0.32 | NA | NA |
| FF | 5.22 | 2.48 | 11.56 | 0.65 | 10083 | 1573 |
| FFPE | 17.30 | -17.30 | 41.26 | 1.27 | 698797 | 41645 |

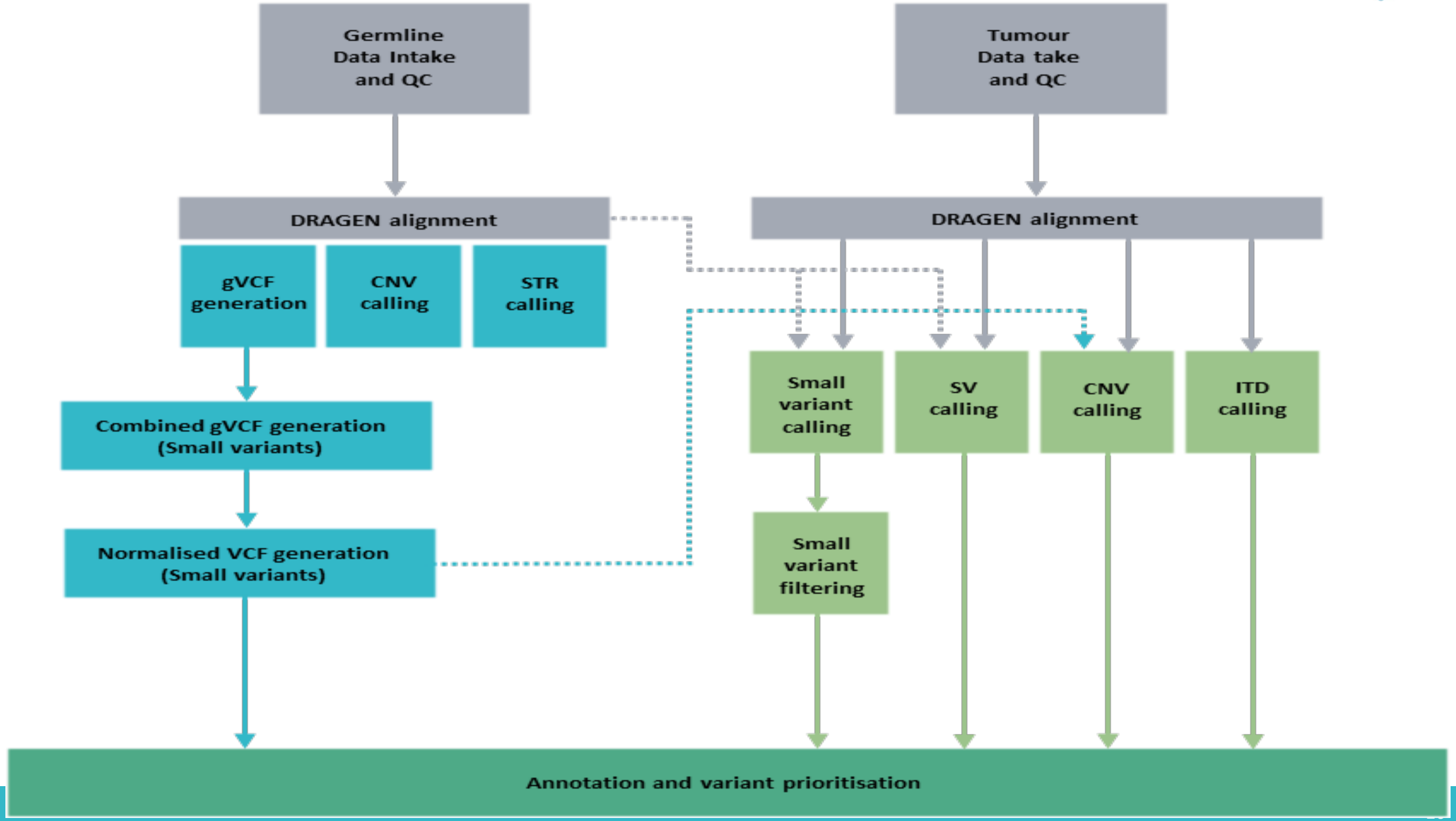
# Cancer Patients by Tumour Type




| Cancer Type | No. of Tumours |
|---|---|
| Adult Glioma | 572 |
| Bladder | 401 |
| Breast | 3244 |
| Childhood | 153 |
| Colorectal | 2980 |
| Endocrine | 60 |
| Endometrial Carcinoma | 868 |
| Haematology Oncology | 1267 |
| Hepatopancreatobiliary | 364 |
| Lung | 2043 |
| Malignant Melanoma | 350 |
| Nasopharyngeal | 9 |
| Oral or Oropharyngeal | 274 |
| Other | 7 |
| Ovarian | 697 |
| Prostate | 694 |
| Renal | 1519 |
| Sarcoma | 1652 |
| Sinonasal | 19 |
| Testicular Germ Cell Tumours | 61 |
| Upper Gastrointestinal | 254 |
| Grand Total | **17488** |

# Tumour stage




Germline
Data Intake
and QC
Tumour
Data take
and QC
DRAGEN alignment
DRAGEN alignment
gVCF
generation
CNV
calling
STR
calling
Combined gVCF generation
(Small variants)
Normalised VCF generation
(Small variants)
Small
variant
calling
SV
calling
CNV
calling
ITD
calling
Small
variant
filtering
Annotation and variant prioritisation

# Pan genomic markers 5700 patients 136 actionable genes (Genome Oncology)

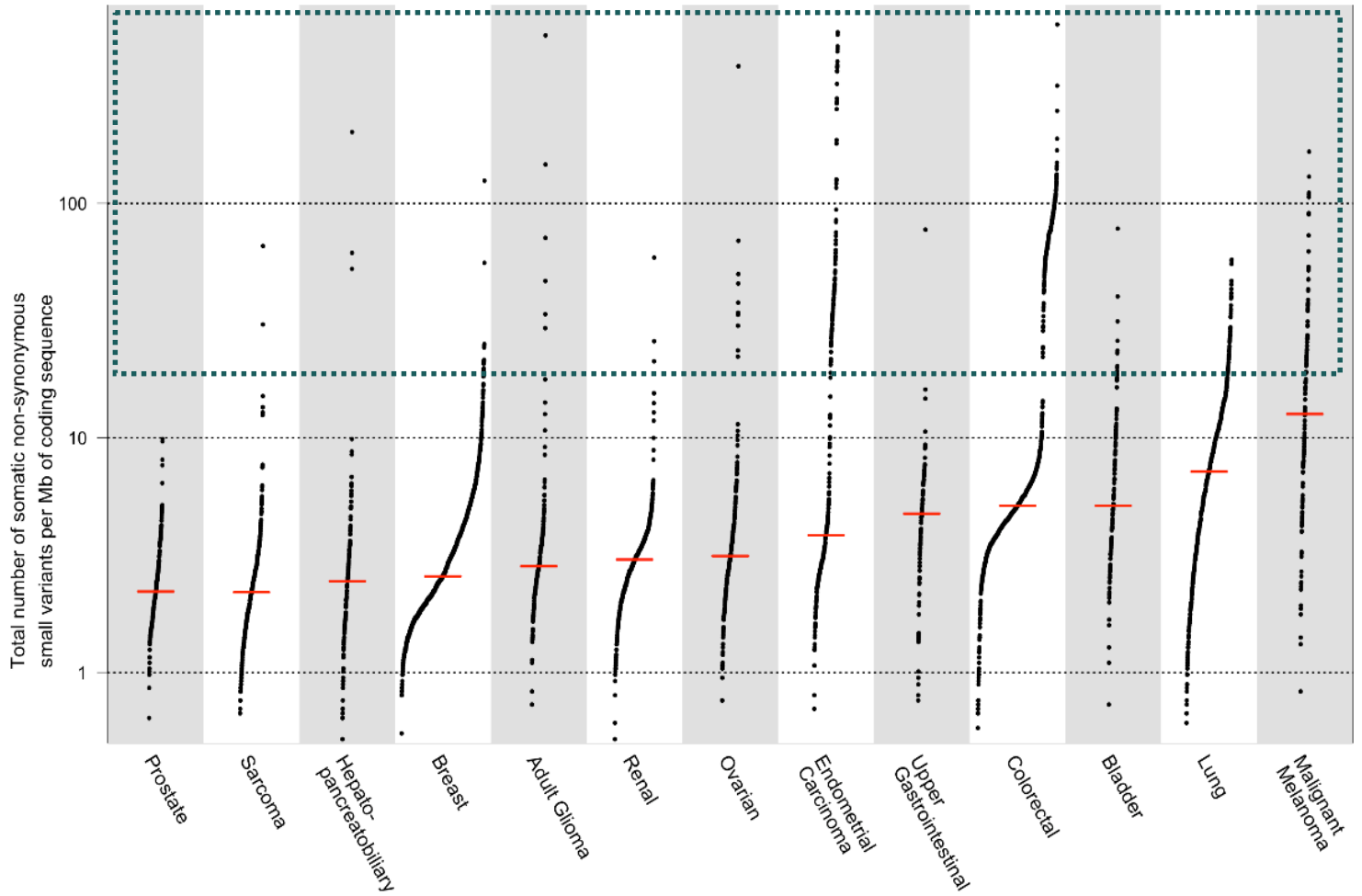
High Mutational Tumour Burden

↓

Higher tumour immunogenicity

↓

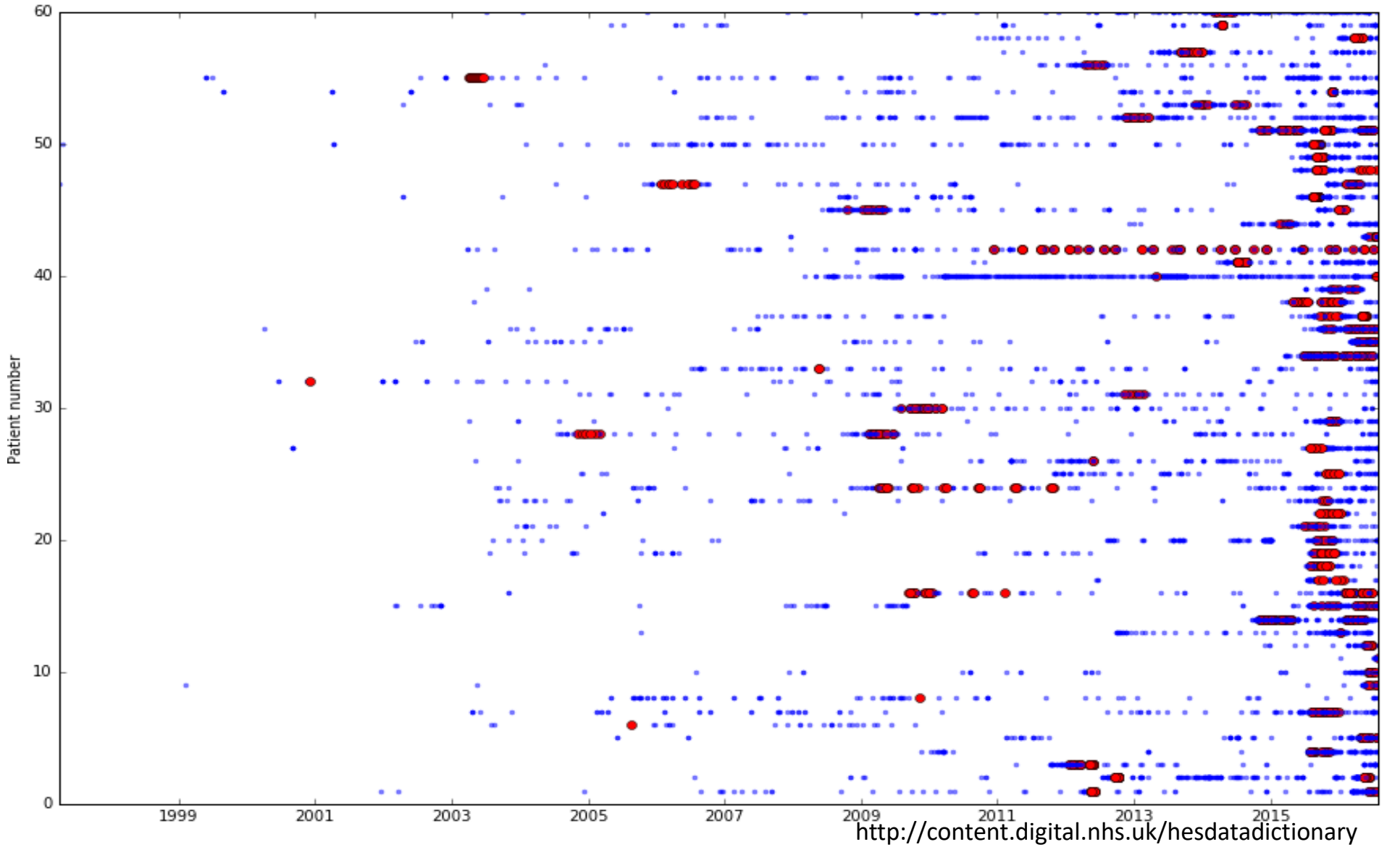
Greater anti-tumour response to immunotherapies






27

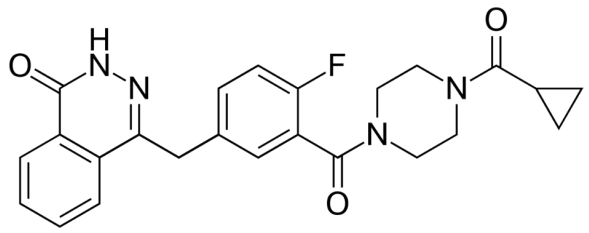
# Hospital Episodes delving deeper 1997-2005
# Previous treatment, 61 patients care pathways




http://content.digital.nhs.uk/hesdatadictionary

# Cancer Case: Implications of result

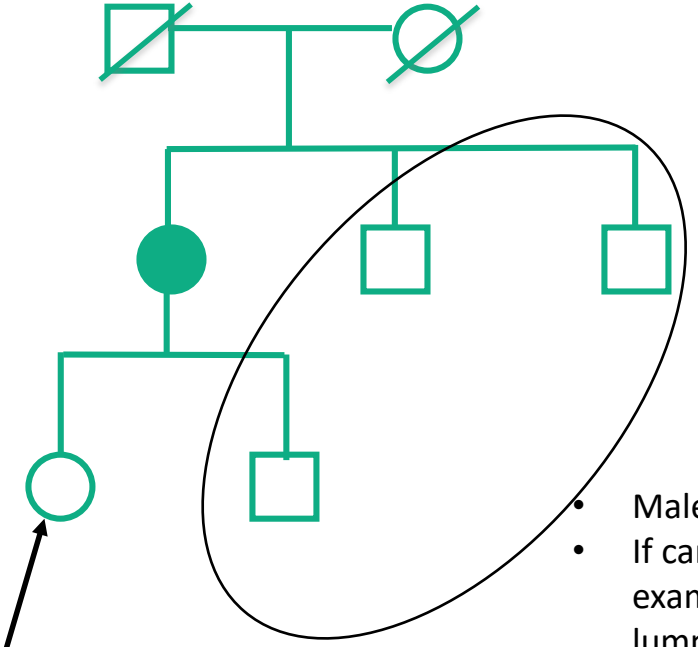
## For the patient




- Targeted therapy with Olaparib through clinical trial (OLYMPIA)

- 1-3/10 women develop ovarian cancer
- Offer risk reducing surgery
- 1 in 2 lifetime chance of left sided breast cancer – requires ongoing screening or consideration of risk reducing surgery

## For her family



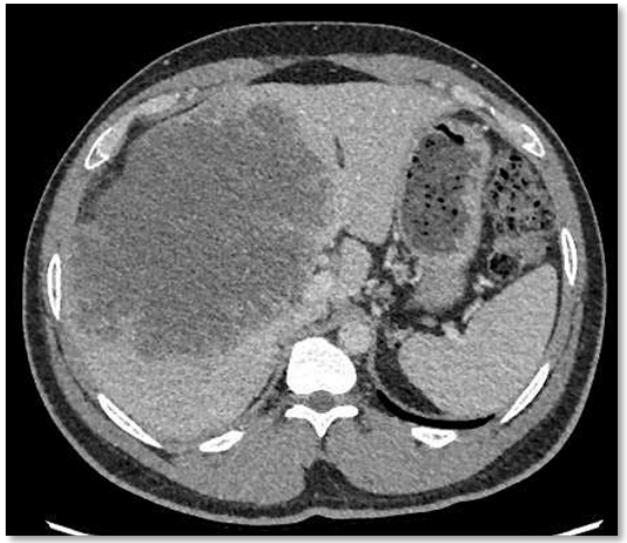
Predictive test
BRCA2 carrier

- Daughter will receive breast screening with scans yearly from age 30
- consider risk reducing surgery
- consider chemoprevention

- Males awaiting testing
- If carriers, self examination for breast lumps
- Eligible for prostate cancer screening from age 40

# Metastatic Colorectal Case – 42 year old male

- Past history of well-controlled HIV (off-therapy) and previous Hepatitis B infection
- Weight loss and jaundice
- CT
  - Circumferential sigmoid thickening
  - Liver lesion & possible satellite
  - Nodule

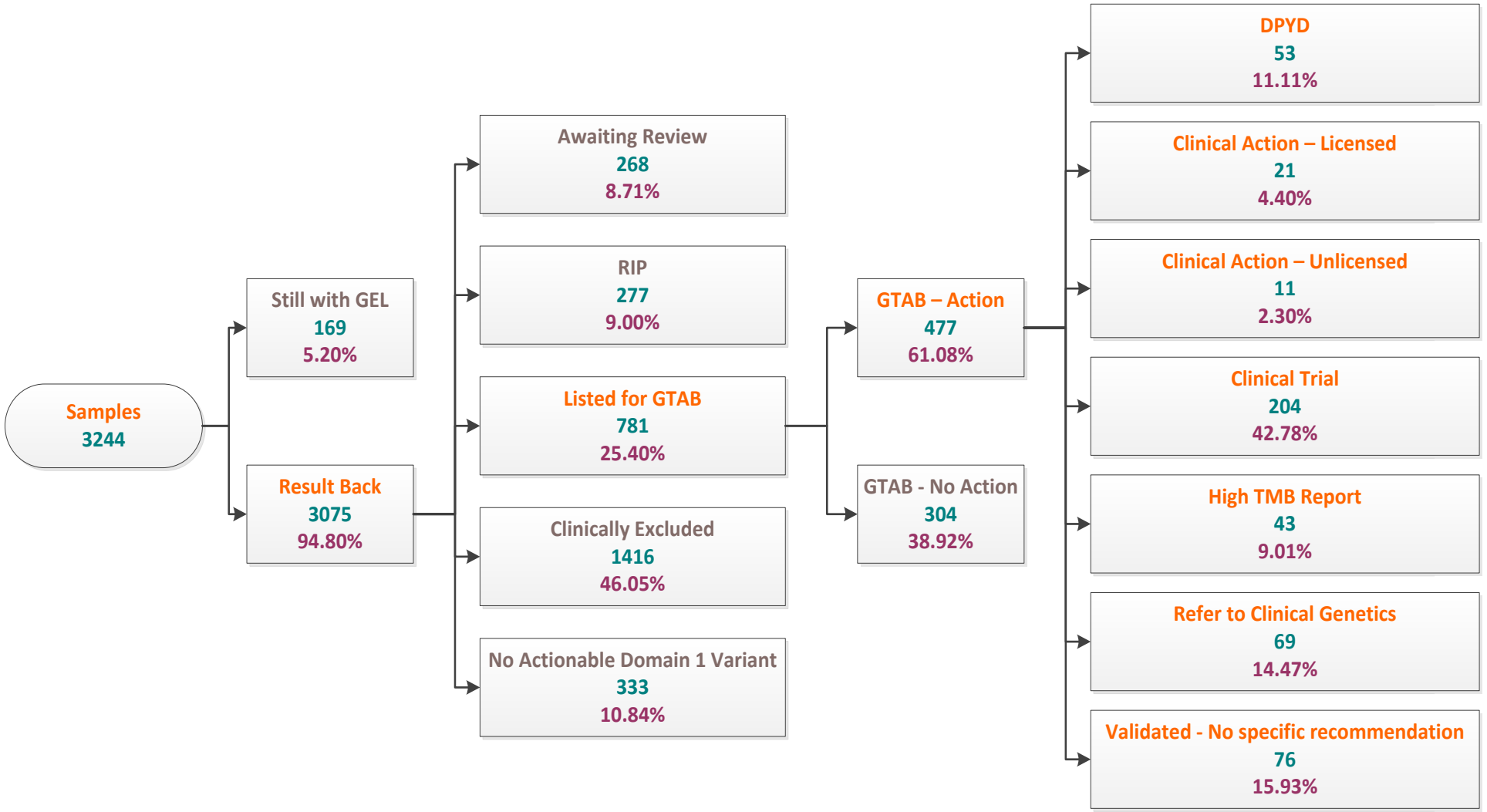


- Sigmoid biopsy
  - Moderately differentiated adenocarcinoma
  - NewGene mutation testing (Usual Care)
  - MSI/MMR not done via NHS (as biopsy)
  - 100K fresh frozen sample sent
- Received 6 cycles of usual chemo but the liver lesion was  enlarging liver
- WGS report
  - 83 domain 1 variants including
    - *NRAS* c.175G>A  VAF 0.13
    - *KRAS* c.38G>A  VAF 0.11
- **Germline *MSH6* mutation**

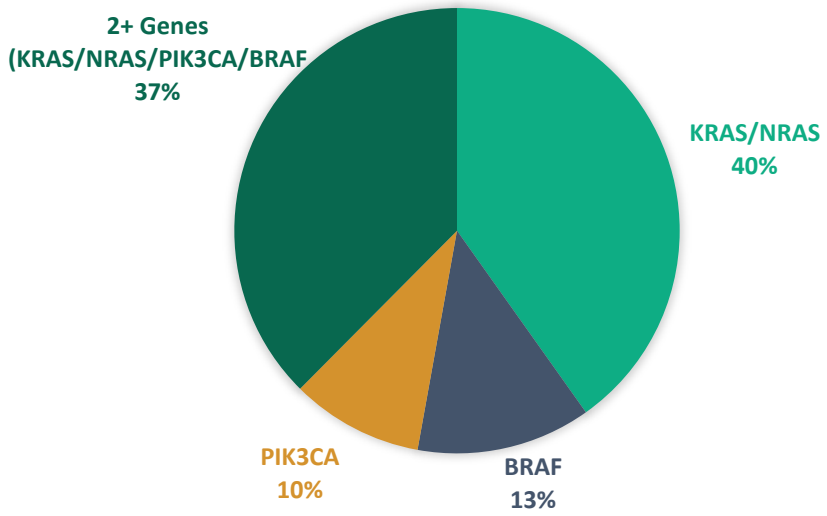
- Next steps
  - Immunotherapy trials

Genomics england

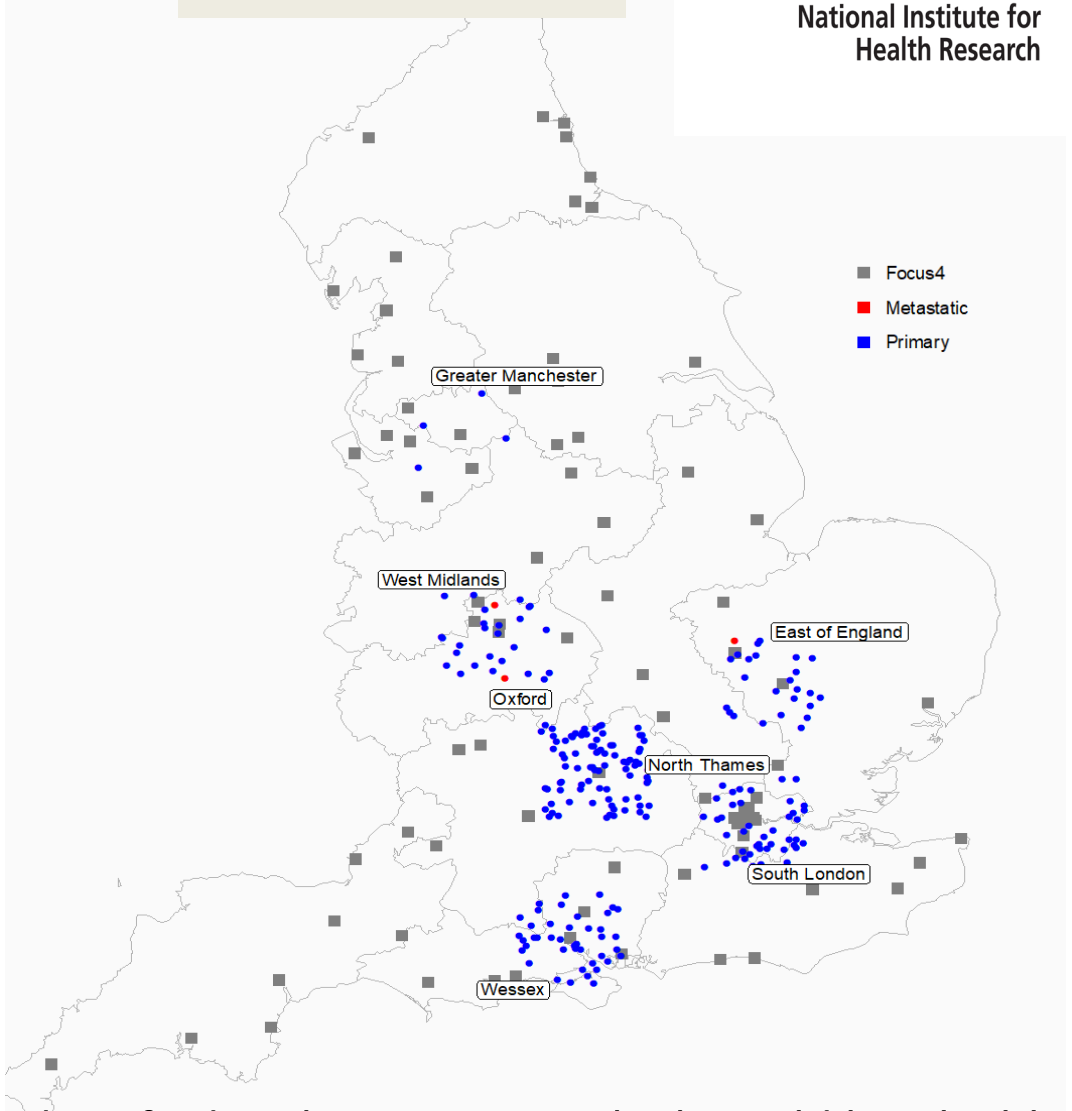
# Genomic Tumour Board – Clinical Utility



# Facilitating recruitment of patients to clinical trials



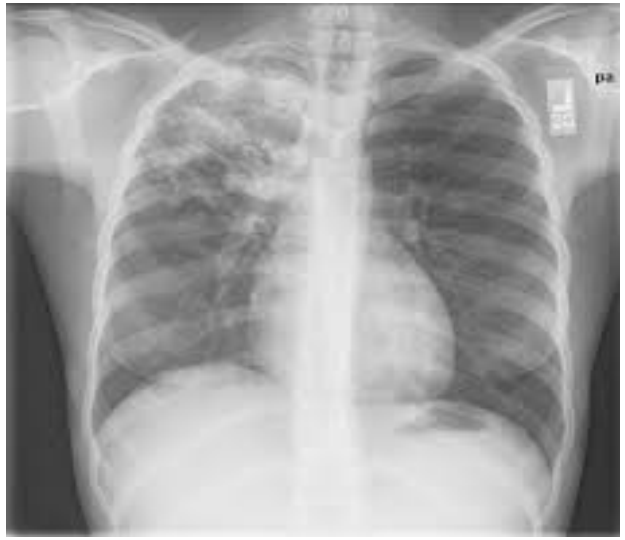

## FOCUS-4 Trial




229 colorectal cancer patients were identified with mutations which could be eligible for FOCUS4 trial, if they were to develop a recurrence

# Infections

# Infections & Pathogens




The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing

The CRyPTIC Consortium and the 100,000 Genomes Project

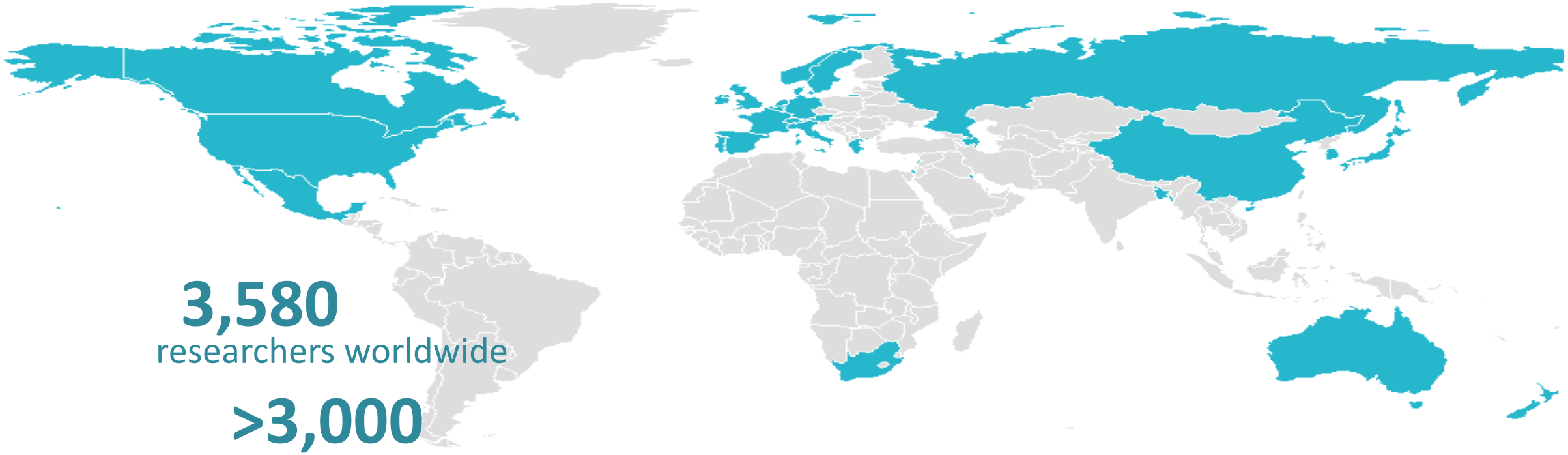
- 10,000 TB strains sequenced
- WGS correctively predicted drug sensitivity enabling precision care for TB
- NHS implemented TB sequencing for diagnosis (1000 organisms/month)
- Global registry of TB resistance

# Clinical Pharmacogenetics International Consortum - actionable allele summary

- **CYP2C9 & phenytoin** – 13% (861) have a genotype that could increase toxicity, reduced starting dosage is recommended.
- **CYP2C19 & clopidogrel** – 30% (1988) have a risk genotype for adverse CV events due to lack of efficacy.
- **Warfarin dosing algorithm** – 27% (1789) had a CYP4F2 genotype including *3 (rs2108622 T), 12.5% (828) have a CYP2C9 genotype affecting dosage.
- **CYP3A5 & tacrolimus** – tacrolimus has a narrow therapeutic window; 2% (132) have a genotype that may require higher dose to achieve target INR.
- **DPYD & fluoropyrimidines** –could be at increased risk for severe or even fatal drug toxicity.
- **Testing algorithms for HLA region and CYP2D6**
- 60,000 whole genomes – 100% possess a CPIC actionable gene-drug pair
- Median of 4 gene drug pairs

# The Genomics England Clinical Interpretation Partnership

# Genomics England Clinical Interpretation Partnership



**3,580**
researchers worldwide

**>3,000**
researchers with data access

**413** academic institutions

**£50 million in grants won**

Discovery Forum of 130 companies partnering us to add value for patients

# Genomics England - Sept 2020 release – 3.8 billion clinical data points alongside 111,000 genomes

| | | | |
|---|---|---|---|
| **Genomes** | **111,232 genomes**<br>• **37,224 Cancer**<br>• **74,008 Rare Disease** | **Primary clinical data** | **89,139 participants**<br>• **17,339 Cancer**<br>• **71,800 Rare Disease** |

| | | |
|---|---|---|
| **Secondary data** | • Hospital Episode Statistics (HES)<br>• Patient Reported Outcome Measures (PROMs)<br>• Mental Health Services Data Set (MHSDS) | • Office for National Statistics (ONS) – mortality data and cancer flagging<br>• ICNARC intensive care dataset for COVID response |

| | | | |
|---|---|---|---|
| **Clinically interpreted data & QC** | • **33,827 families** with Tier 1, 2 and 3 variants from interpretation pipeline<br>• **20,032 families** with GMC exit questionnaires<br>• **61,138** tiered and quality checked rare disease genomes;<br>• **31,590** quality checked cancer genomes | **Quick view tables** | • Final genomes from the 100,000 Genomes Project<br>• A new aggregate VCF file with v8 data<br>• COVID ICNARC data<br>• PHE cancer and SACT data<br>• PHE Mental health data |

Primary care datasets – SNOMED Codes and quantitative data under COPI Notice imminent

# Transforming healthcare

# National Genomic Medicine Service

**National**
**Genomic Medicine Service**

**Genomic Medicine Services Alliances**
(announcement pending)

**National Laboratory Network**
**7** Genomic Laboratory Hubs

**National Genomics Research Library**

**Whole Genome Sequencing Provider**

**Clinical Interpretation Pipeline**

**Workforce development**
upskilling of existing staff

**Industry/ academic/ international partnerships**

NHS Lead

Genomics England Lead

## National Test Directory

300,000 Tests reviewed
25% upgraded to new technologies

21 categories of rare disease

4 Cancers for WGS

More edge cases in cancer

Annual Directory Review

Pharmacogenomics from April 2021

500,000 whole genomes sequenced from the NHS in the next 5 years

- Offered consent for research
- Longitudinal Life Course
- Recall for research
- International researchers and industry

# The range of genomic testing available

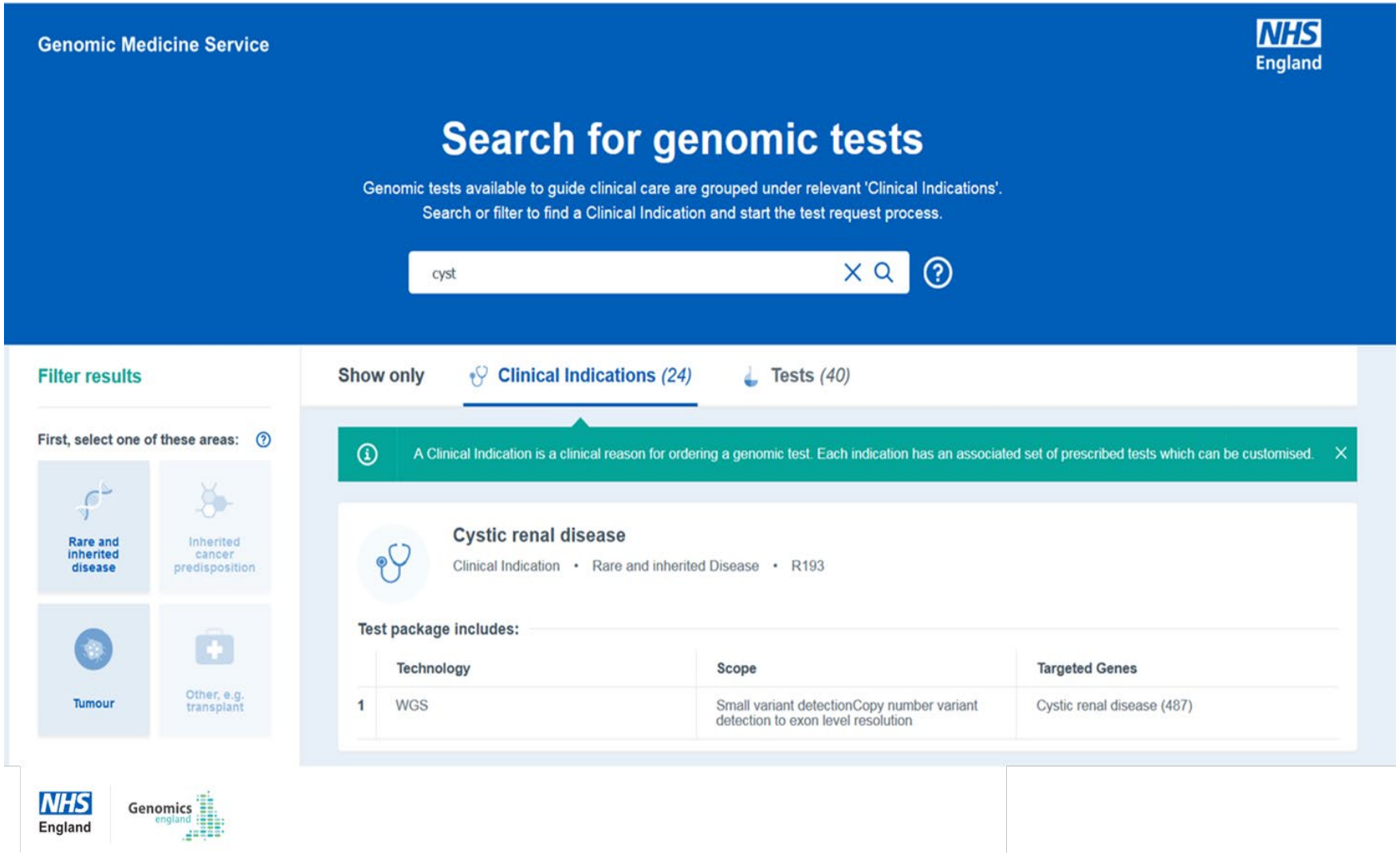
- Test for ~300 rare disease clinical indications and 120 cancer clinical indications identified across 22 test technologies with ~75 panels/subpanels
- Builds upon the substantial evidence base and evaluation by UKGTN since 2003
- The Test Directory identifies core tests provided by all centres together with specialist tests to be provided by a limited number of centres

| Cancer |
|---|
| **Majority of testing** |
| Pan-solid cancer large panel |
| Pane-haematological large panel |
| Paediatircs, Sarcoma - WGS |
| |
| **Smaller volume tests** |
| Single gene tests |
| Karyotype, FISH |
| Methylation tests |

| Rare Disease | Est prop'n of reports |
|---|---|
| Targeted mutation testing | 20-25% |
| Microarray | 10-20% |
| WGS | 10-25% |
| Small panel | 10-15% |
| STR testing | 10-15% |
| WES or large panel | 2-14% |
| MLPA or equivalent | 5-7% |
| Common aneuploidy testing | 5-7% |
| Karyotype | 3-5% |
| Single gene sequencing | 3-5% |
| FISH; DNA repair defect testing; Methylation testing; UPD testing; X-inactivation testing; Identity testing; Microsatellite instability; NIPT; NIPD; PGD | each <2% |
| Other | 2-5% |

# National Genomics Informatics Service

A platform for digital genomic health



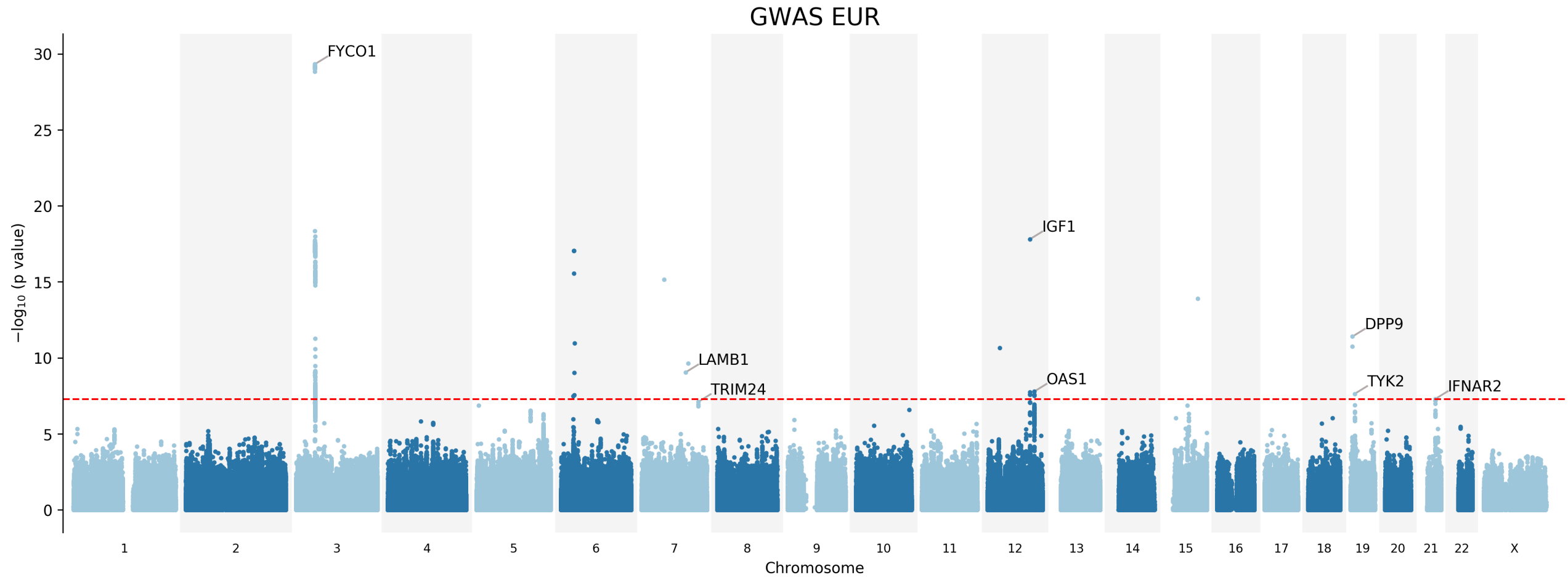
# Genetic mechanisms in severe COVID 19 Illness



Kenneth Baillie

- 2224 severe COVID-19 cases from 209 ITUs genotyped matched controls and cases from Biobank UK, GEL, International collaboration

| Patient Characteristics | GenOMICC (n=2109) | | ISARIC 4C (n=134) | |
|---|---|---|---|---|
| | | *missing data* | | *missing data* |
| Female sex | 624 (30%) | | 46 (34%) | |
| Age (yrs, mean ± SD) | 57.3 ± 12.1 | | 57.3 ± 2.9 | |
| European ancestry | 1573 (75%) | | 103 (76%) | |
| South Asian ancestry | 219 (10%) | | 18 (13%) | |
| African ancestry | 174 (8%) | | 8 (6%) | |
| East Asian ancestry | 143 (7%) | | 6 (4%) | |
| Significant comorbidity | 396 (19%) | 49 (2%) | 42 (28%) | 31 (21%) |
| Invasive ventilation | 1557 (74%) | 35 (2%) | 25 (19%) | 31 (23%) |
| Died (60 days) | 459 (22%) | 338 (16%) | 22 (16%) | 30 (22%) |

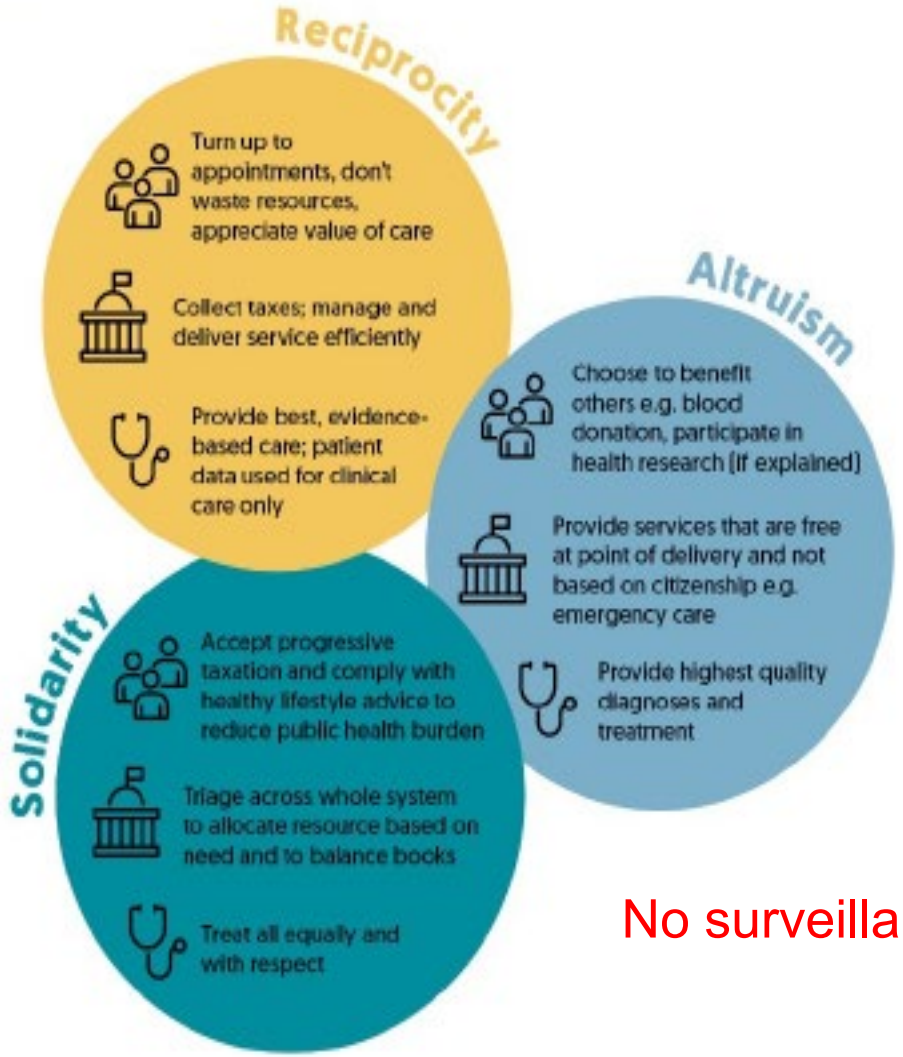

GWAS EUR
FYCO1
IGF1
DPP9
LAMB1
TRIM24
OAS1
TYK2
IFNAR2
−log10 (p value)
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

# GenOMICC first paper published in Nature 2020

- **7 genome wide significant loci detected**
- **3 potential therapies**
- Interferon 2 Receptor – potential target interferon
- Tyrosine kinase 2 – Jak1/Jak 2 inhibitors used in cancer
- e.g. Baracitinib
- CCR2 – biologic tested in rheumatoid arthritis and psoriasis

# The Public and Patients at the heart of the Programme

## Public views of key behaviours in the social contract now




Public don't understand how research ecosystem works / feeds clinical care

Genetic manipulation

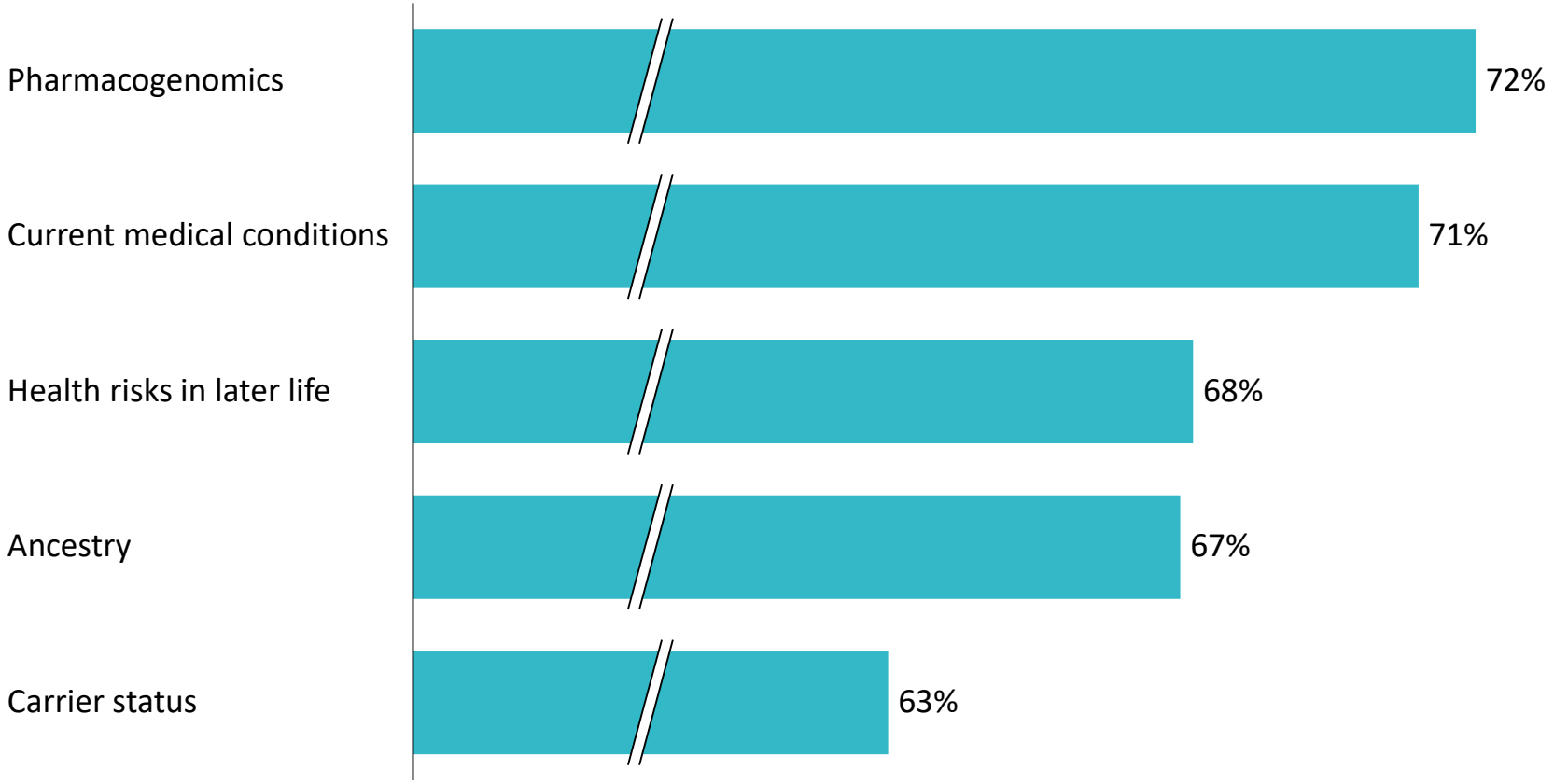
Commercial uses for healthcare

Commercial interests aren't spontaneously seen as part of the system

No surveillance society

# Public survey showed significant interest in receiving personalised genomic results as Genomic Volunteers

## Proportion of respondents interested in personally receiving specific genomic results




Source: Ipsos-Mori survey of the public perspectives of genomic volunteers. Electronic survey of 1,866 people (selected to be a representative sample of adults aged 16 – 75 across England)

HM Government

GENOME UK
The future of healthcare

Not in scope of this SR bid

Diversity Genomics and PGX

Cancer 2.0

PRS Common Diseases

GEL Core Funding inc. NHS Informatics

Rare Disease Diagnoses from WGS

NHSE/I Genomics services

*Funded via core NHS budget*

Newborns "Generation Genome"
Early diagnosis
1 in 190 births
9 children every day

# Future – UK Life Sciences Strategy

- International Partnerships with
- France, Australia, Hong Kong, Qatar, British Columbia, Japan

Multi-omics and new technologies

- Long read technologies
- cftDNA
- Transcriptomics
- Multi-omics
- Standardisation
- Other disease areas
- Population cohorts

# Transforming the future genomic medicine service

The National Health Service will have:

- A national Genomic Medicine Service providing consistent & equitable care for 55 million population
- Operating to common national standards, specifications & protocols
- Standardised genomic consent for NHS care and Research
- Delivering an approved national testing directory covering use of single gene to WGS
- Building a single UK Genomic Knowledgebase
- national NHS database with all tests
  that will enable care, effectiveness, and outcomes
- De-identified data for academic & industry research
- An ambition for 5 Million Genomic Tests & Early Detection Cohorts
- The future is a global coalition of intellects driving genomics into healthcare and our goal is for the UK to be at the heart of that

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

# Thank you to everyone who has taken part in the 100,000 Genomes Project

# Stay in touch

Email Mark Caulfield
m.j.caulfield@qmul.ac.uk

@genomicsengland #genomes100k

Like the ‘Genomics England’ page

Follow ‘Genomics England’

www.genomicsengland.co.uk

# UK Precision Medicine – Life Sciences Strategy

| | Diagnosis and Precision Care | | Prevention | Research |
|---|---|---|---|---|
| Key UK Infrastructure: | **The 100,000 Genomes Project and Genomic Testing in the NHS** | | **Accelerating Detection of Disease** | **UK Biobank** |
| Population | 100,000 participants with rare diseases and cancer | Genomic testing across the UK includes 500,000 WGS by 2024. | 5 million healthy people at the time of recruitment | 500,000 participants healthy at the time of recruitment |
| Genomic data | Whole Genome Sequencing | Whole Genome Sequencing and non-whole genome sequencing | Genotyping – Polygenic Risk Scores | Genotyping<br>Whole Exomes and Whole Genome Sequencing |
| Complementary data | Phenotypic and long-term clinical data | Phenotypic and long-term clinical data collection | Health-related data | Deep phenotyping and health-related data |
| Bio-sampling | ✔ | ✔ | ✔ | ✔ |
| Clinical feedback | ✔ | ✔ | ✔ | ✘ |
| Recontact | ✔ | ✔ | ✔ | ✔ |
| UK Wide | ✔ | ✔ | ✔ | ✔ |