Genomics England

Biomedical Research in Cancer & Rare Disease

Parker Moss

Manchester, 6 July 2022

Our history

December 2012

Announced by former Prime Minister David Cameron – an Olympic Legacy

July 2013

Genomics England formally launched by then Secretary of State for Health during NHS 65th Anniversary Celebrations

November 2016

Former Prime Minister Theresa May opens a new Sequencing Centre

July 2017

Chief Medical Officer launches Generation Genome and the Life Sciences report

December 2018

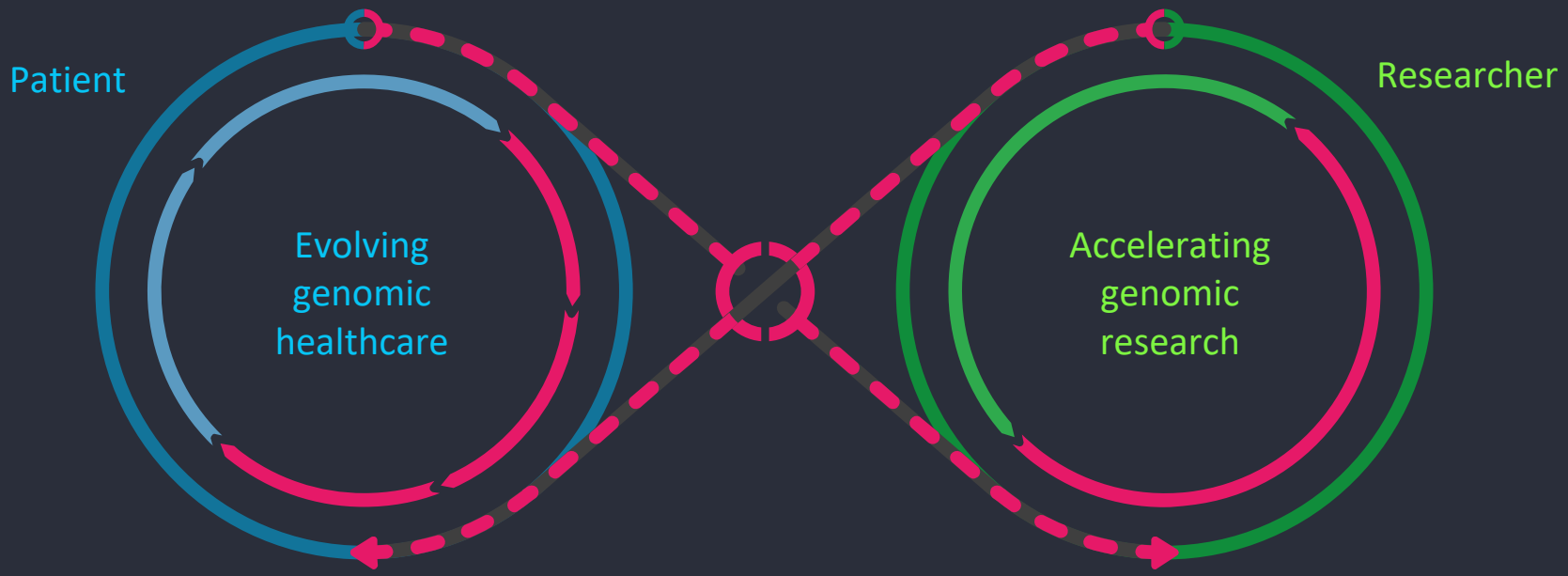
Genomics England reaches goal of sequencing 100,000 genomes

January 2019

Long Term Plan “an NHS where access to secure linked clinical, genomic and other data will support new medical breakthroughs and consistent quality of care”


Our vision is a world
where everyone benefits
from genomic healthcare

GEL Strategy: An Infinity Loop



Patients

Healthcare teams

Researchers

The UK ecosystem

Academia ~4000 researchers

BioPharma & Start-ups

Tech
(e.g., AWS, Nvidia, Lifebit, Congenica)

Lab Tech
(e.g., Illumina, Oxford Nanopore)

Funders
(e.g., Innovate, BEIS, MRC, CRUK, Wellcome, Lifearc)

8 out of 10

Tier 1 pharma access our data

100,000 Genomes Project Data

Release v.15

Genomics

	Cancer	Rare Disease	Total
Participants	17,243	75,948	93,191 + 35K COVID
Genomes	31,208 Germline + Tumour 30x 100x	75,894 Germline <20% Singleton	107,102 + 35K COVID

100,000 Genomes Project Data

Genomics

Clinical Data

- Tumour staging
- Tumour location
- Histological subtype
- Treatment regimen
- Pathology full-text
- Radiology full-text

- Hospital Episode Statistics
- Mental Health Services Data Set

- Mortality data ONS

- COVID-19 status

- Exit questionnaire

- Primary Care Data (coming soon)

100,000 Genomes Project Data

Genomics

Clinical Data

Consent

Clinically accredited pipelines

for diagnostics

Lifetime follow-up

+ full retrospective data

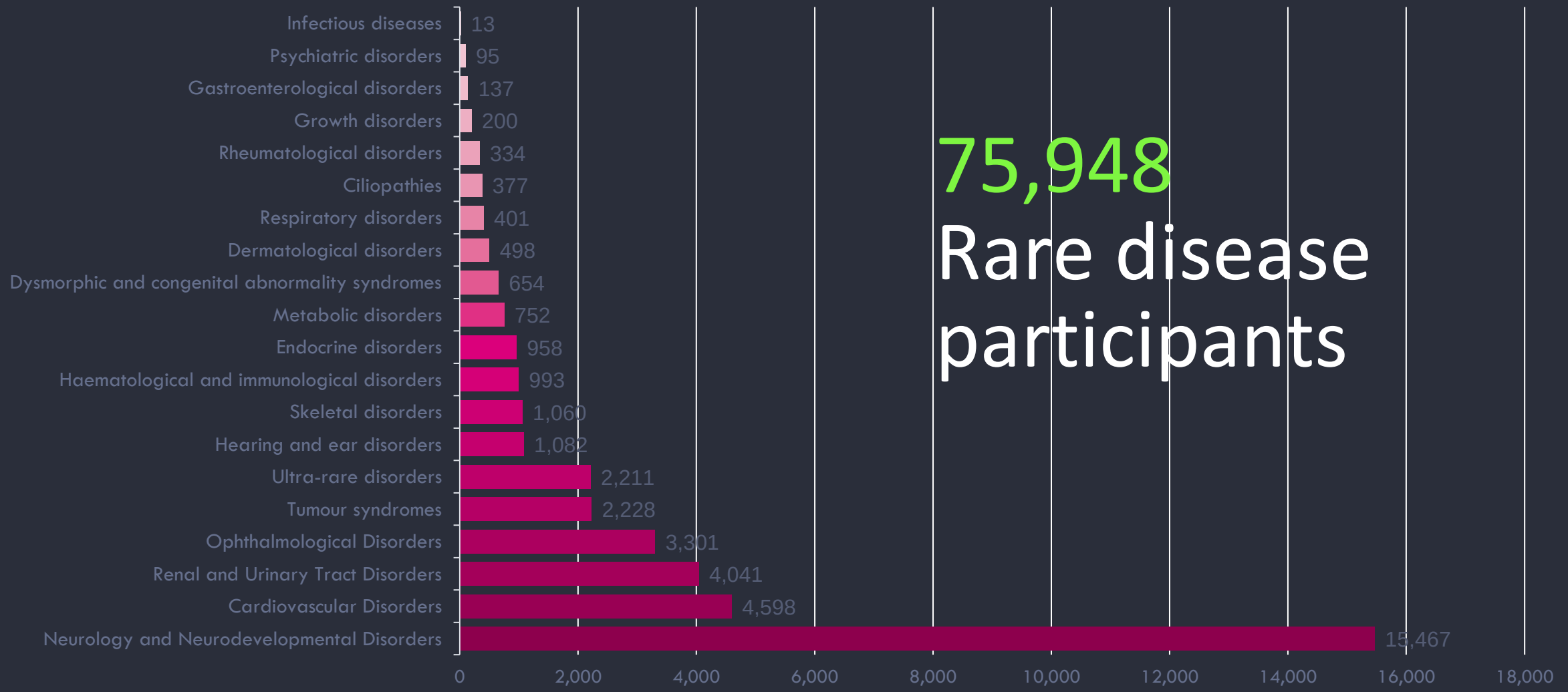
Re-engagement

re-phenotyping

re-sampling

re-cruiting

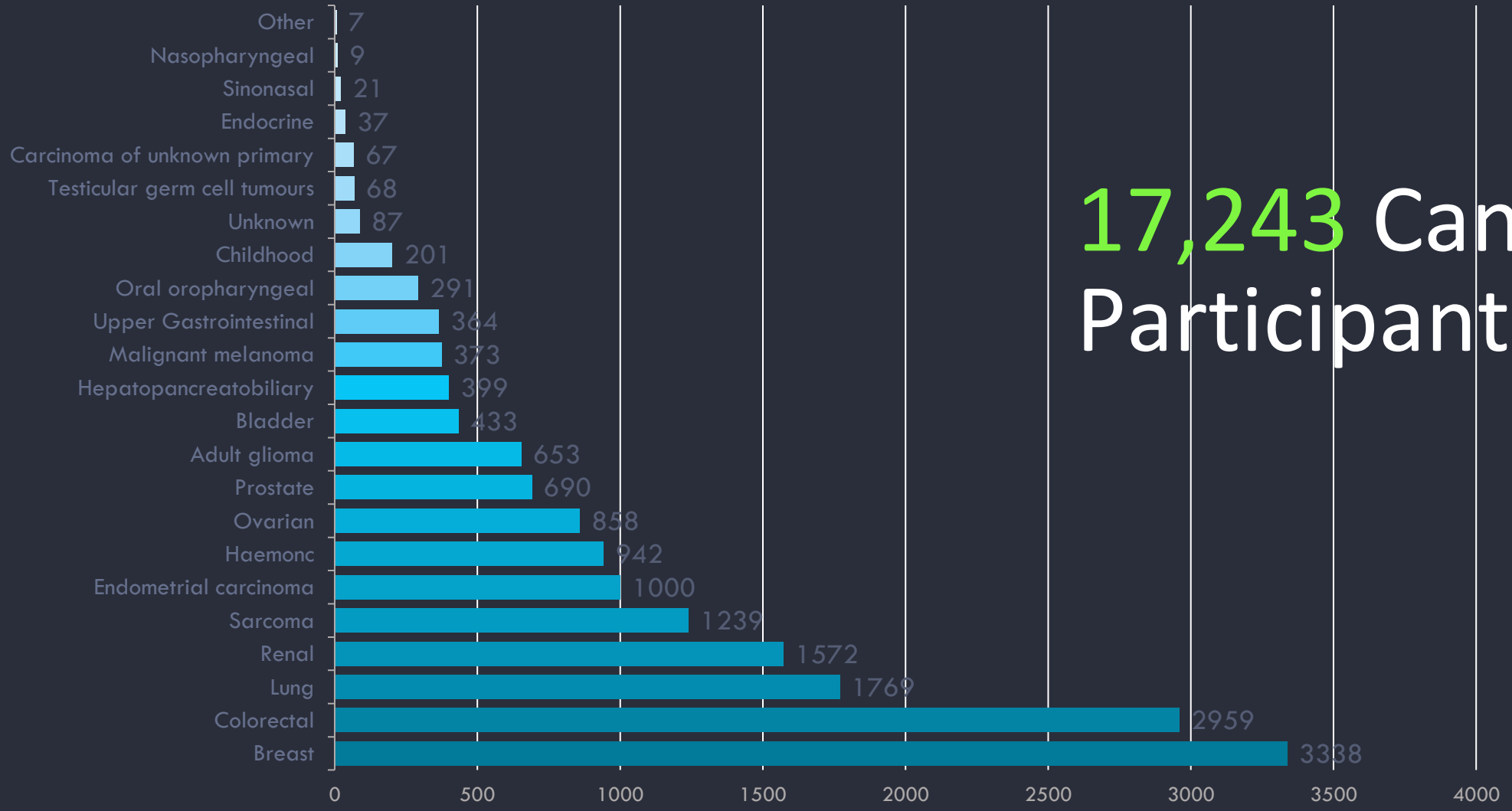
100KGP Rare Diseases Participants



75,948
Rare disease participants

See Cohort browser for genomes count detail in research environment

100KGP Cancer participants



17,243 Cancer Participants

See Cohort browser for genomes count detail in research environment

NHS Genomic Medicine Service

Beyond 100,000

WGS Cancer indications

- Wave 1: Acute Leukemias, Paediatric Tumors, Sarcomas
- Wave 2: Ovarian HGS, Triple Negative Breast, Glioma, Other Heam Onc, Various relapse & refractory

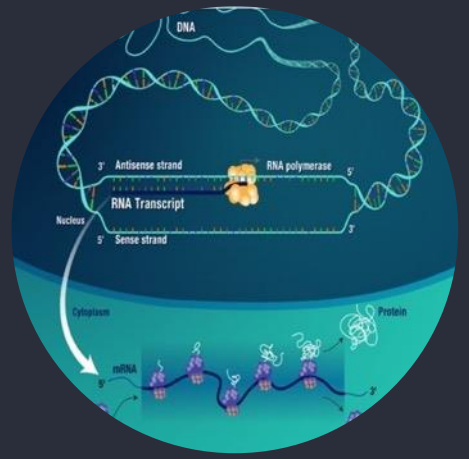
WGS Rare Disease indications

- Wave 1: 20 rare conditions
- Wave 2: +10 rare conditions

~ 10,000 participants will enter the dataset in next data release v.16

NHS Genomic Medicine Service

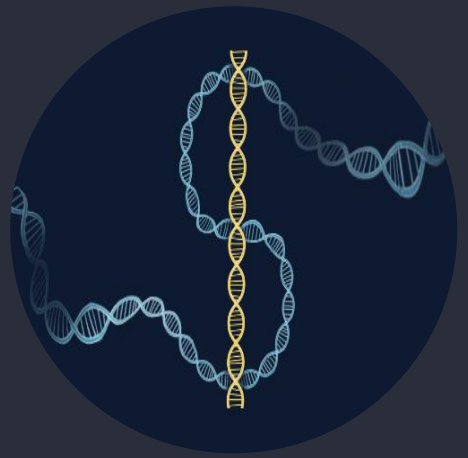
New initiatives



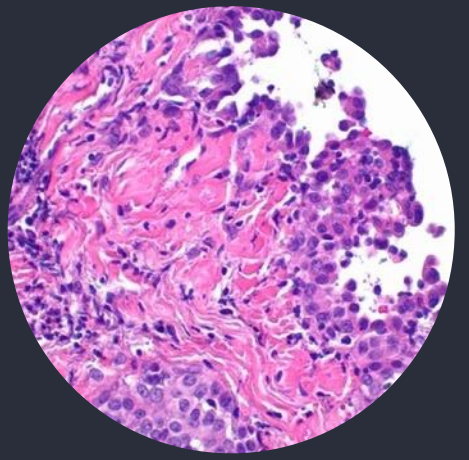
Transcriptomics & Proteomics in Rare Disease

Newborn sequencing

Diversity



Nanopore Sequencing in Cancer

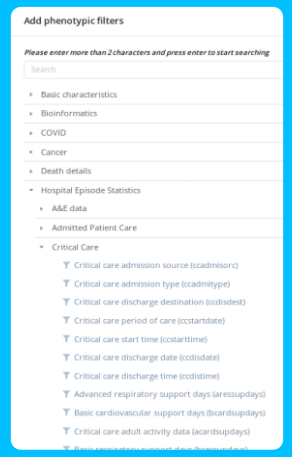


Radiogenomics in Cancer

GEL's Trusted Research Environment: AWS & Lifebit

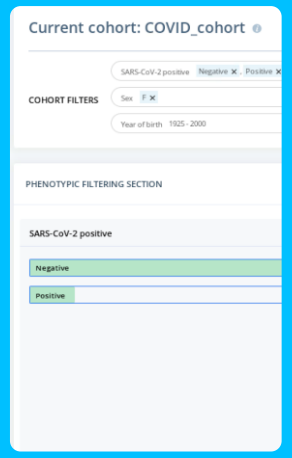
GEL CloudOS

Explore Data



Using Intuitive data browser

Build Cohort



Flexible patient stratification

Analyse Data

Tool repository & Bring Your Own

GWAS & PheWAS

Seamless collaboration in Private Work Spaces

Command Line Interface & APIs

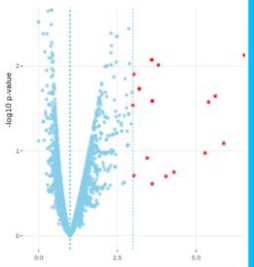
Workflow Automation

Interactive Analysis

jupyter Shiny R python

Visualise results

igv



docker GitHub Bitbucket nextflow CWL bcbio Beagle BOW TIE

Lifebit AI Engine

Advanced Analytics & Deep Learning

Data Science | AI | ML | NLP

Target Generation

Automate targets generation & novel drug discovery from multi-omic and other data

Validation Automation

NLP & Knowledge Graph to validate findings & accelerate approval

Data Federation

Genomics england

Pharma Data

aws

View demo here
Password:
GEL_Lifebit_Demo

Rare Disease Analysis with Aggregate VCFs

Data Products

gVCF: ~78K rare disease and cancer germline

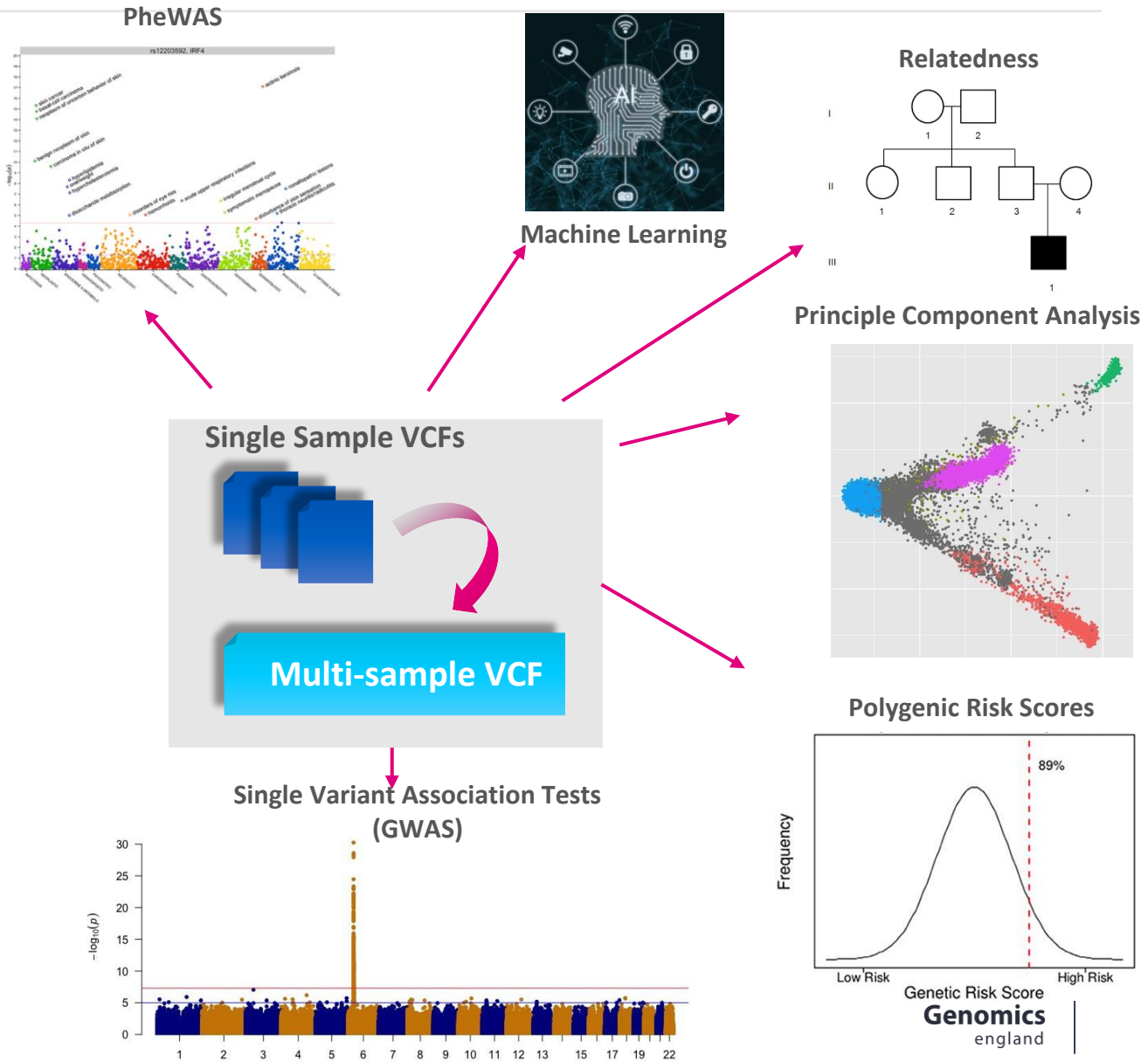
Tight sample QC with PCA and detailed functional annotation files (**Ensembl Variant Effect Predictor**, **ClinVar**, loss of function from **loftee**, population specific allele frequencies from **gnomAD** and **ExAC**, non-coding epigenetic markers from **ENCODE**, **Roadmap** and **spliceAI**)

Clustering of filtered results into groups such as:

- Chromosome
- Variant type
- Consequence type
- Deleteriousness
- Pathogenicity

Accelerates:

- Derived Ancestry, Allele frequencies
- GWAS, PheWAS, PRS
- AI/ML validation

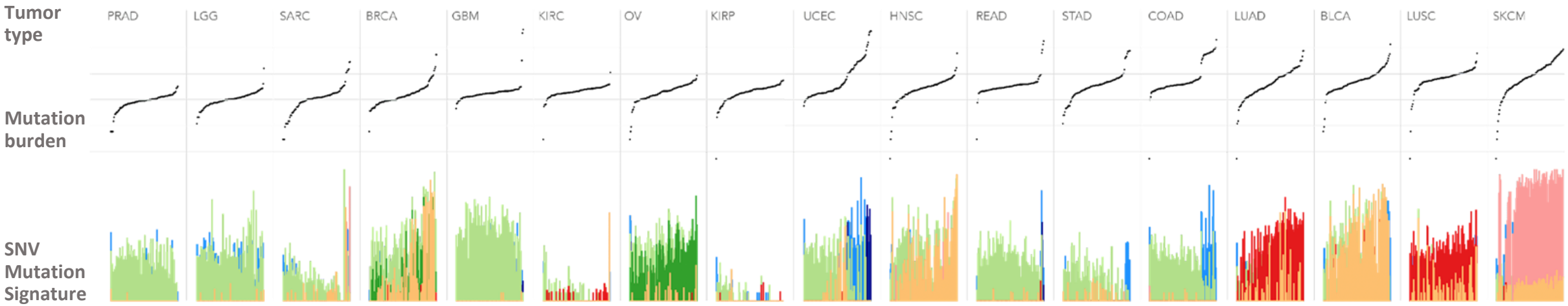


Genomics england

Pan genomics markers in cancer

Data Products

Tumour mutation burden and signatures pre-calculated in the GEL research environment



Tumor mutation burden (TMB)

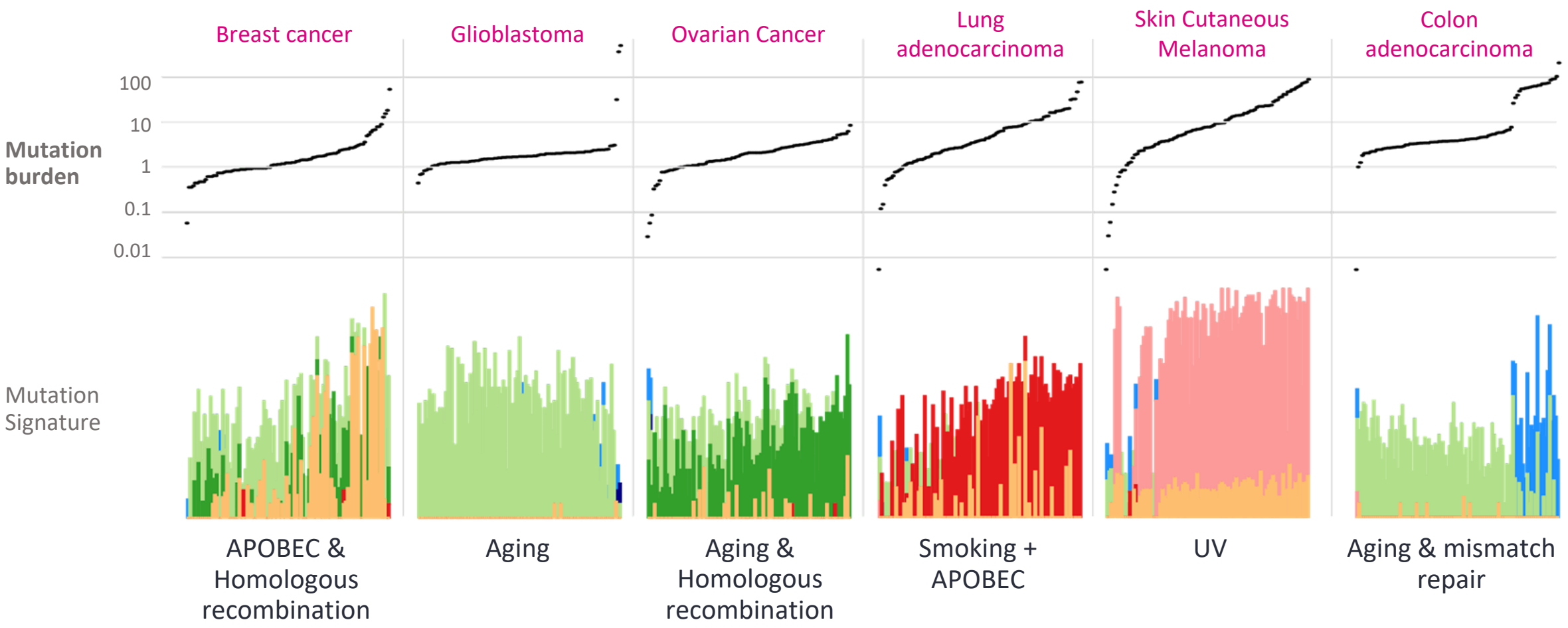
Calculated as the number of somatic non-synonymous small variants per Mb of coding sequences. High TMB can indicate patient's suitability for immunotherapy

Mutation signature

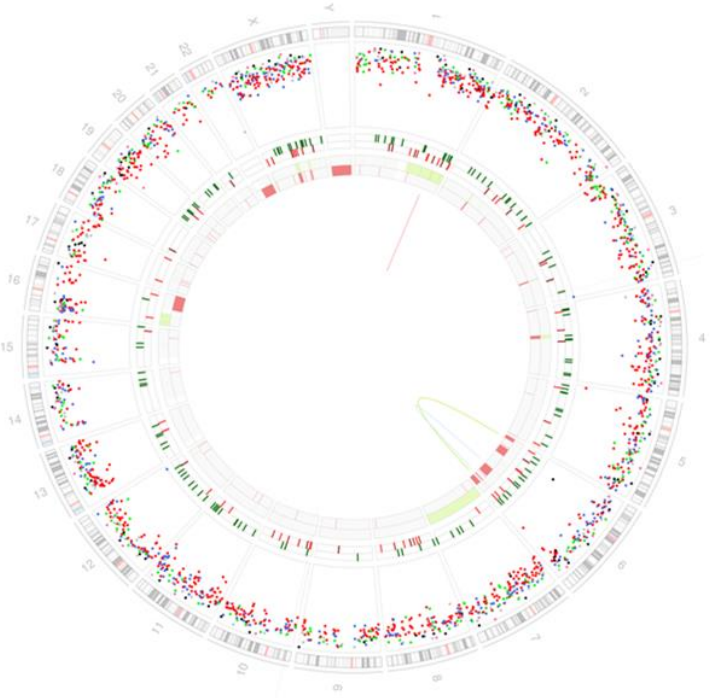
Characteristic combinations of mutation types arising from specific mutagenesis processes. For example, smoking causes C>A transversion signature

Pan genomics markers in cancer

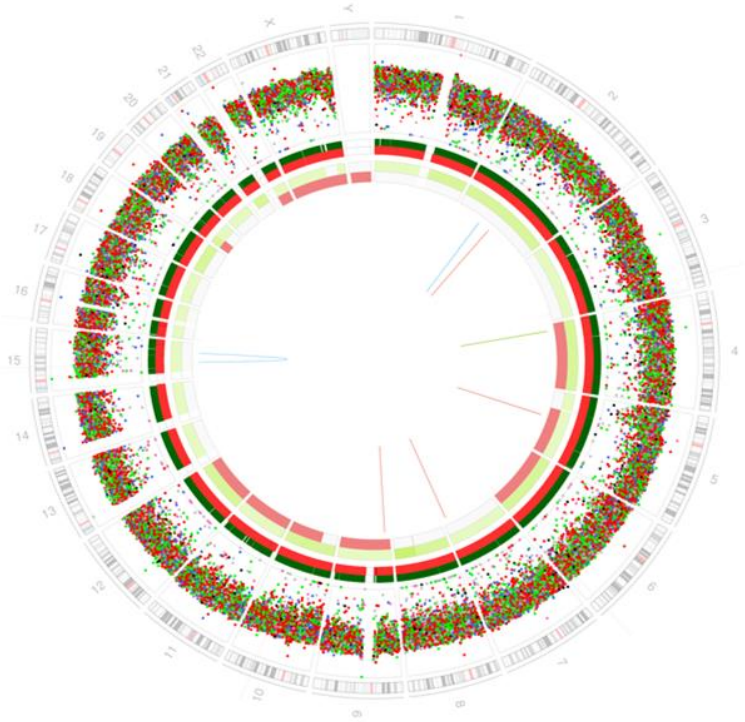
Data Products



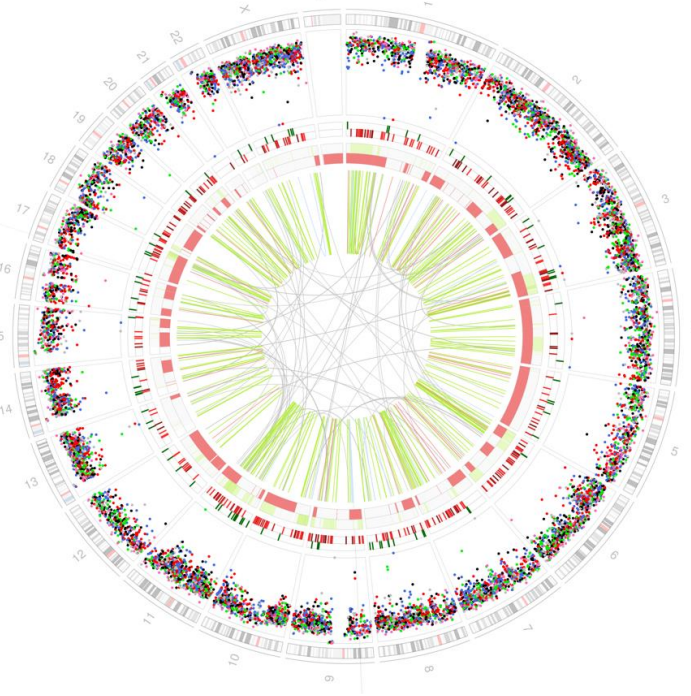
Feature Extraction from the Whole Genome



ER positive breast cancer
Good outcome



ER positive breast cancer
Poor outcome
Mismatch repair deficient

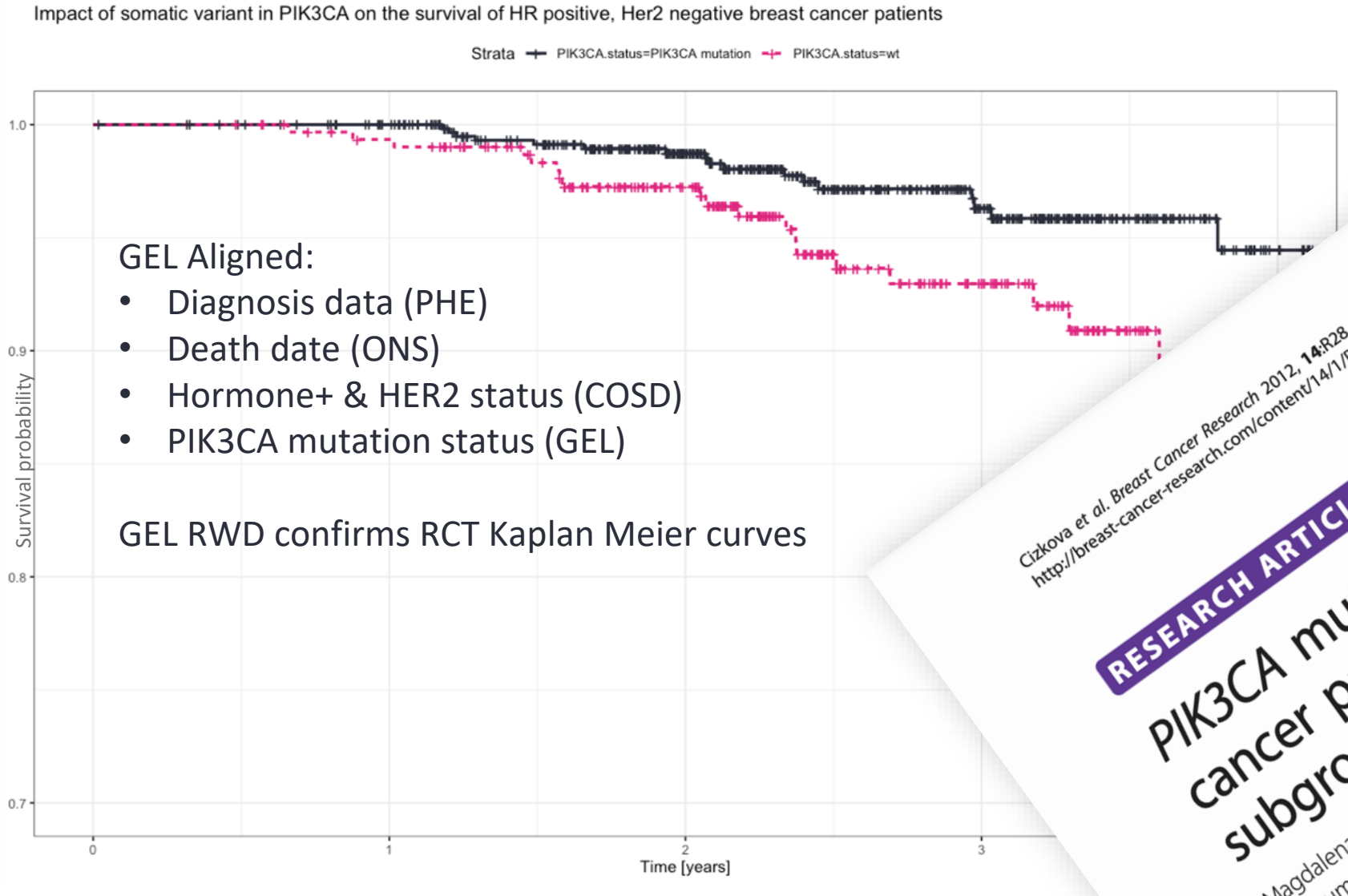


BRCA1 defective
Poor Outcome
Homologous Recombination deficiency

Nik-Zainal et al, Nature, 2016

Outcomes analysis from GEL RWD

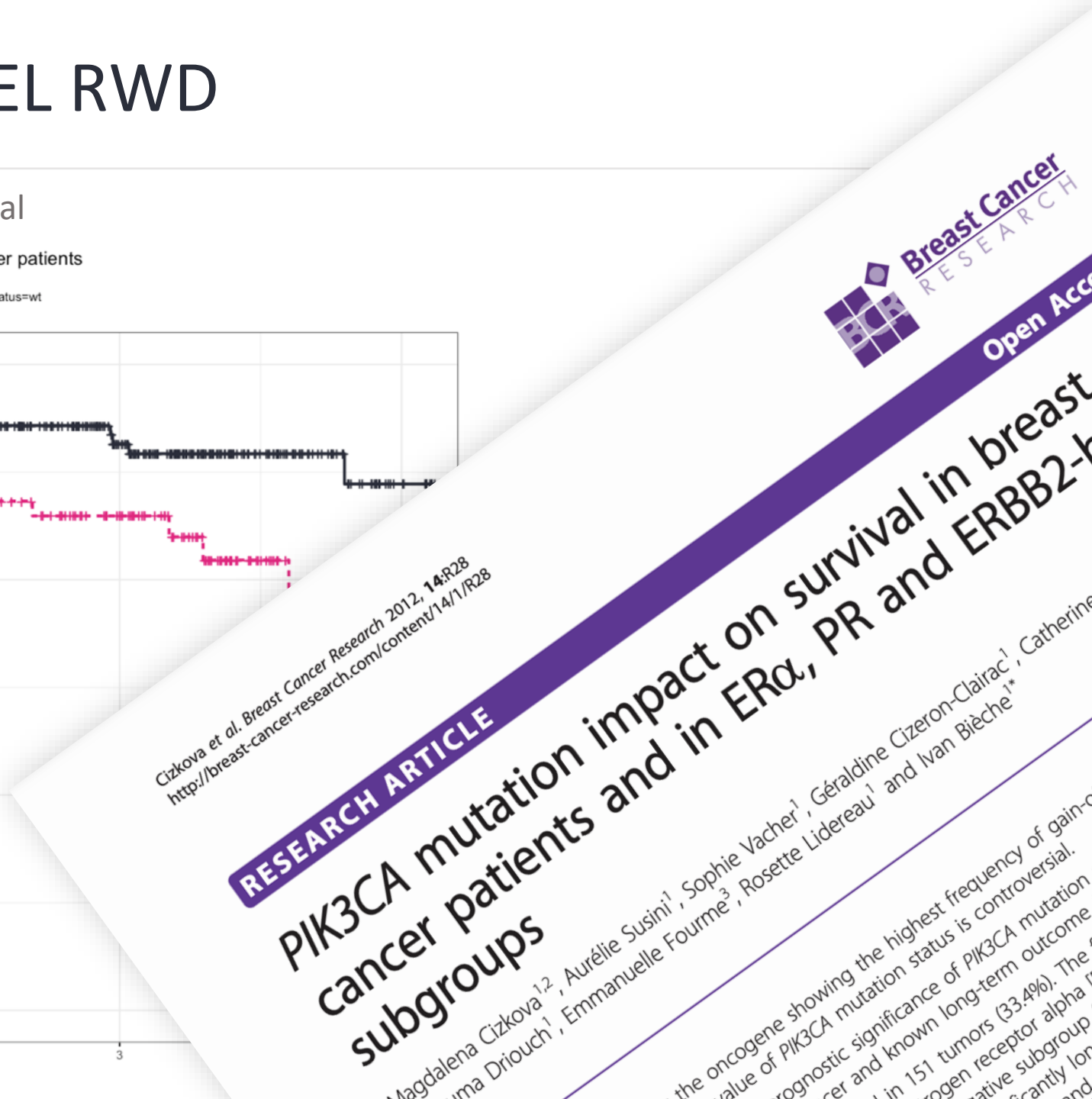
PIK3CA from GEL RWD predicts Breast Cancer Survival



GEL Aligned:

- Diagnosis data (PHE)
- Death date (ONS)
- Hormone+ & HER2 status (COSD)
- PIK3CA mutation status (GEL)

GEL RWD confirms RCT Kaplan Meier curves



Clinically relevant findings by patient

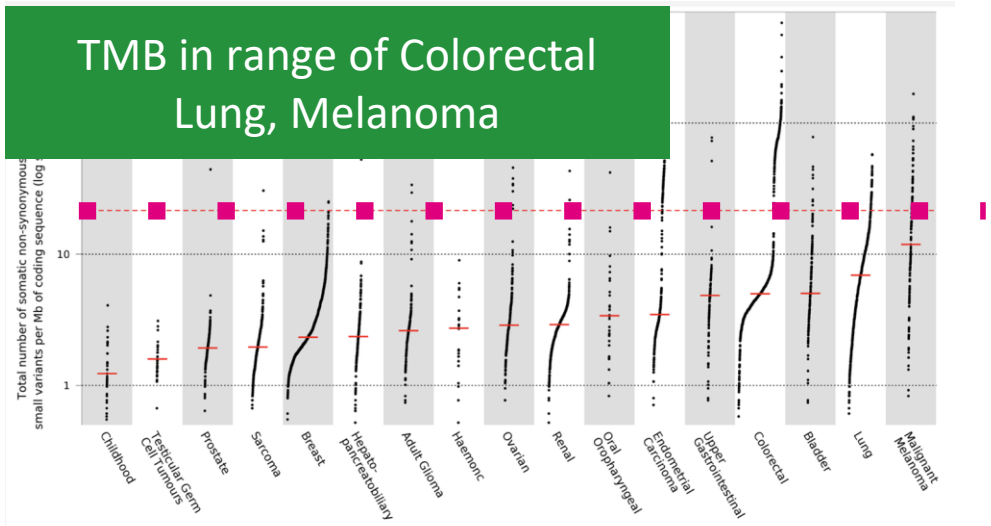
Data Products

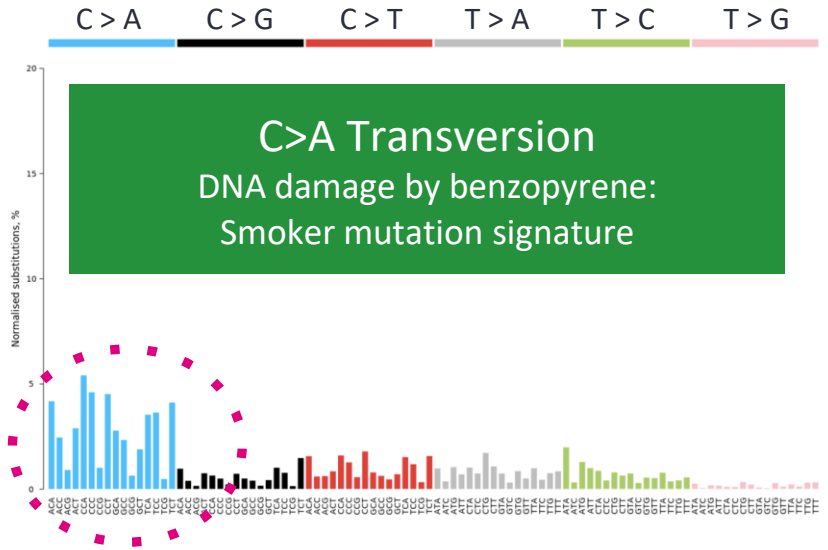
Disease type	Disease subtype	Annotation group	Tumor type	Topography	Morphology
Carcinoma of unknown primary	Pathology indeterminate cancer NOS non carcinoma lymphoma sarcoma etc	SOLID	Metastases	C7.3,TX2000 Brain Tumor	M80106

Gene	GRCh38 coordinates ref/alt allele	Transcript	CDS change and protein change	Population germline allele frequency (1KG | gnomAD)	VAF	Alt allele/total read depth	Gene-level actionability	Gene mo... ac...
TP53	17:7675077G >A	ENTST00000269 305	c.535C>Tp. (His179Tyr)	- | -	0.78 (LOH)	57/73	Trial (NSC lung ca, ca, colorectal ca, head neck SCC, ovarian ca, prostate ca Trial (NSC lung ca, breast ca, colorectal ca, esophageal SCC, head neck SCC, ovarian ca, pancreatic ca, urothelial ca) Trial (ovarian serous ca)	

Potential lung cancer drivers: RB1 & NF1 Frame Shifts

TMB in range of Colorectal Lung, Melanoma





C>A Transversion
DNA damage by benzopyrene:
Smoker mutation signature

Lung is the likely primary tumor

Working with Genomics England

- A recallable resource of exquisitely annotated rare disease and cancer data
- A world class bioinformatics team available for collaboration
- A community of 4000 clinical-academic collaboration partners, deeply embedded in GEL data
- A gateway to the NHS ecosystem

Thank you

Parker.moss@genomicsengland.co.uk