in partnership with

Building a Collaborative Clinical Genomics Service

Jon Lockley (Institute of Cancer Research)
Mike Hubank (Royal Marsden Hospital)

Making the discoveries that defeat cancer

ICR
At a glance

Top 4 global cancer research organisation

Top-ranked UK academic institution (REF)

20 drug candidates discovered since 2005

More than 1,000 staff

£161.9m income
£110.0m expenditure

Awarded Athena SWAN Silver

More than 900 scientific papers

Partnerships with 163 different companies

Top UK university for invention income

141 research students
143 MSc students

The Institute of Cancer Research
Major Historical Discoveries

Discovered the carcinogenic compounds in cigarette smoke and coal (1920s)

Discovered many of the first drugs (still) used in chemotherapy treatments (1950s)

Provided the first conclusive evidence that cancer is caused by damage to DNA (50s-60s)

Discovered one of the major ways by which RAS oncogenes cause normal cells to become cancerous (80s)

Discovered the BRCA2 gene (90s)

The Royal Marsden Hospital

The Royal Marsden was the first hospital in the world dedicated to the study and treatment of cancer. It was founded as the Free Cancer Hospital in 1851 by Dr William Marsden

Rated ‘Outstanding’ by the Care Quality Commission (CQC)

Staff and facilities often shared with ICR

ICR & RMH
Joint Research Strategy

1

Unravelling cancer's complexity
We will comprehend the full complexity of cancer by **harnessing the power of new technologies and Big data**

2

Innovative approaches
We will take on the challenge of cancer's complexity, evolution and drug resistance through the discovery of innovative new approaches to cancer treatment

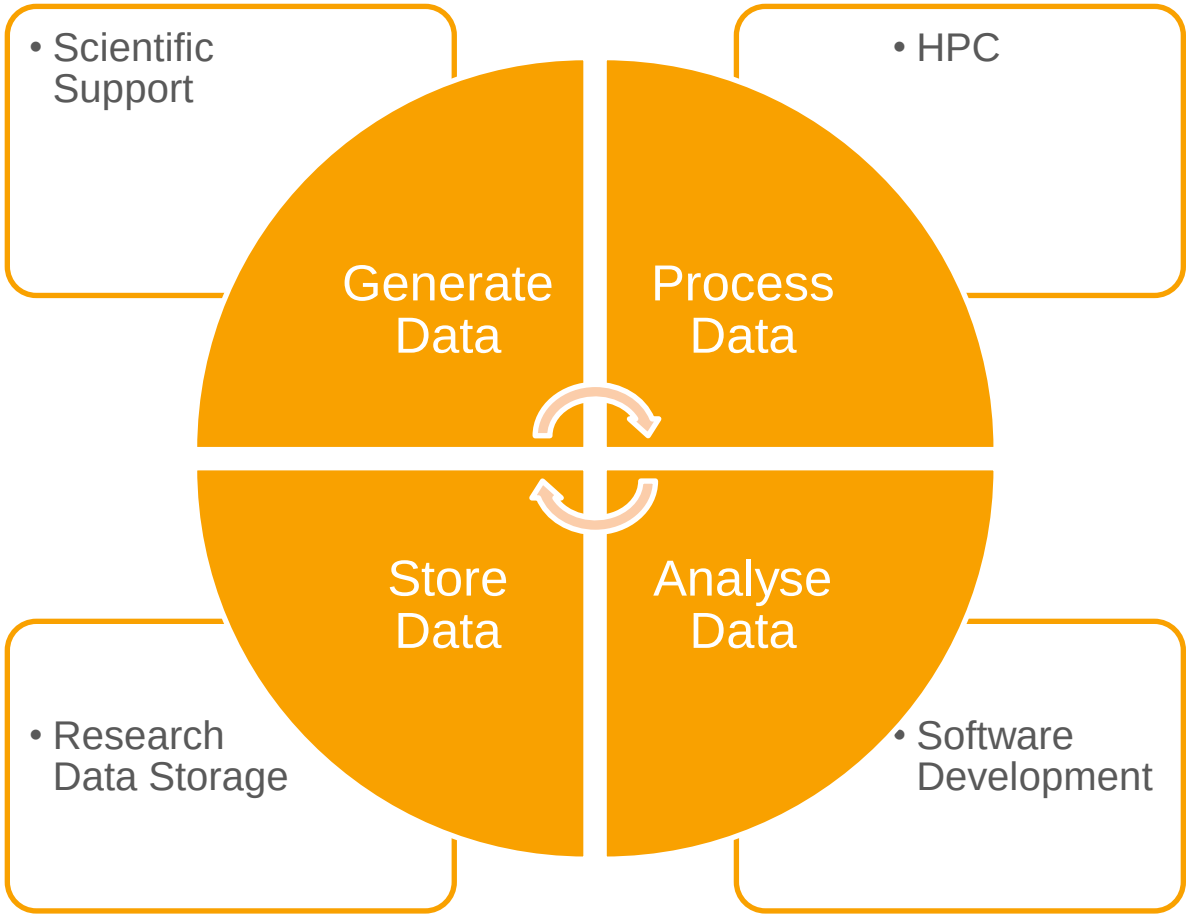
3

Smarter, kinder treatments
We will move a step closer to cure by bringing personalised treatments into the clinic and developing them for patients

4

Making it count
We will deliver better outcomes and improved quality of life for patients by establishing innovative treatments, diagnostics and strategies for prevention as part of routine healthcare

ICR
Scientific Computing Services



Research Computing Infrastructure:

Our HPC infrastructure is housed entirely at Slough/Virtus

Our storage system is geographically distributed between London and Slough

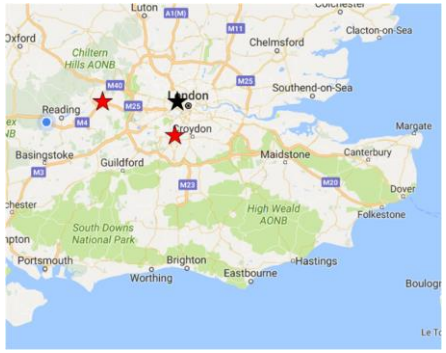
All three sites are connected by a private 10Gb network (with an option to increase b/w)

Traffic can reroute in the event of failure

Each site has a 10Gb internet (JANET) connection

HPC: 2,000 cores, 2PB scratch

Storage: 6.5PB GridScaler + Object Store

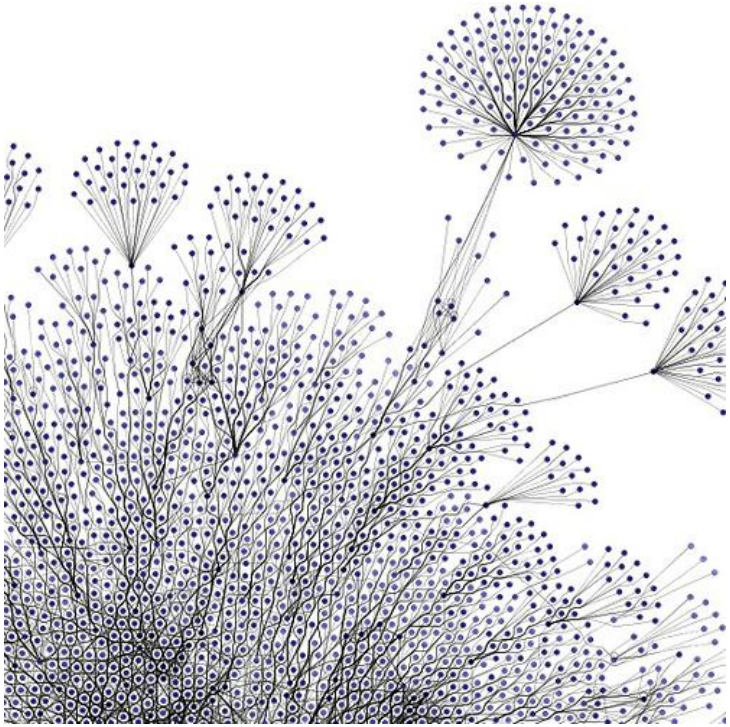


The Complexity of Cancer

Cancer is a highly complex disease with more than 200 types, each caused by a different combination of genetic defects.

Even among patients with a single type of cancer, such as breast or lung cancer, the genetic triggers vary from one person to the next.

It is important to take into account those differences as we discover new cancer treatments – an approach known as personalised medicine



https://www.icr.ac.uk/news-features/latest-features/mel-greaves-science-writer-of-the-year-2013/unravelling-the-complexity-of-cancer

NHS ENGLAND GENOMICS

Seven regional testing hubs
- Germline
- Somatic (Cancer)

London N/W/E (GOSH/RMH+)

London S (Kings/Guys)

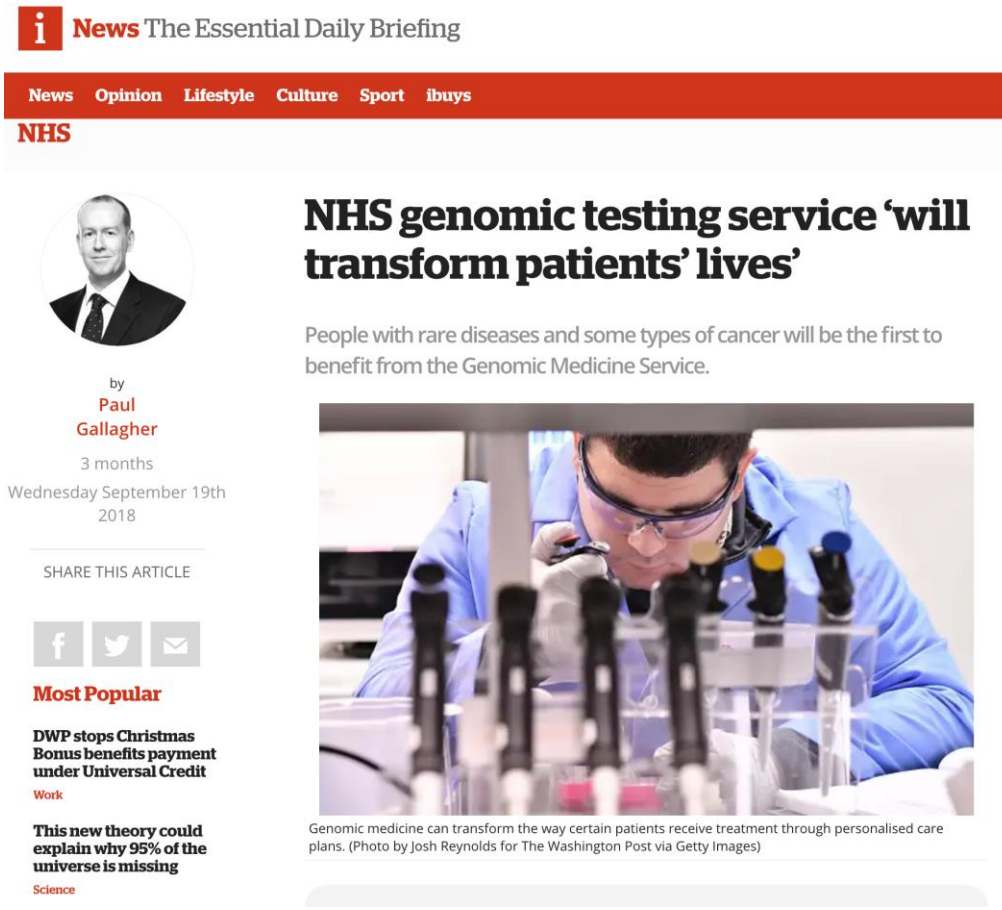
East of England (Cambridge)

Northwest (Manchester)

Southwest (Bristol)

Northeast (Leeds/Newcastle)

Midlands (Birmingham)



i News The Essential Daily Briefing

News Opinion Lifestyle Culture Sport ibuys

NHS

by Paul Gallagher

3 months

Wednesday September 19th 2018

SHARE THIS ARTICLE

Most Popular

DWP stops Christmas Bonus benefits payment under Universal Credit

Work

This new theory could explain why 95% of the universe is missing

Science

NHS genomic testing service 'will transform patients' lives'

People with rare diseases and some types of cancer will be the first to benefit from the Genomic Medicine Service.

Genomic medicine can transform the way certain patients receive treatment through personalised care plans. (Photo by Josh Reynolds for The Washington Post via Getty Images)

How we diagnose: Gene Panels

Sequenced Exomes are compared to a panel of genes of interest to look for variants

Higher-end sequencers allow for bigger panels and deeper (less noisy) reads in less time

More samples can be run more cost effectively

The larger sequencers and panels also have a greater chance of showing variants which are unknown or not understood, allow for research to be carried out on data intended as a clinic test

…but produce *lots* more data.

Why Test by Gene Panel?

Comprehensive Panel
250+ genes
to Whole Genomes

2 weeks

250 samples per run
1.5 Tb per run

NOVASEQ

£150/sample
>1,000,000
0.015p/test

Expanded Panel
25 – 250 genes

25 samples per run
40Gb per run

NEXTSEQ

£200/sample
10,000 – 200,000
2p/test

Focused Panel
1 – 25 genes

5 samples per run
10Gb per run

MISEQ

£230/sample
100 – 20,000
£2.30 per test

Current situation

Single gene test £70-150
+
Single gene test £120
+
Single gene test £120
+
Single gene test £120
+ …

4 days

Months or longer

>£120/sample
1 – 100 targets
£120 per target test

Creating the new service

RMH's challenge:

- Big sequencers and lots of data makes sense but…
- Very limited local computing resources
- Limited HPC expertise
- How do we step up to a much greater numbers of samples in a reasonable amount of time?
- Approached the ICR

ICR's challenge:

- We can't rework our existing research environment
- Have a "typical" academic level of security
- Need to provide a guaranteed level of service
- How can we accommodate a clinical work load alongside research?
- Don't want to run multiple clusters again (again)
- Is the answer technological or is Information Governance sufficient?

Information Governance

Started looking at Information Governance issues around hosting clinical work…

…then GDPR landed

What we had previously been referring to as anonymised data now became pseudonymised data under GDPR

ICR became the “data processor”

RMH became the “data controller”

(So the party who processes the data is the controller whereas the party who controls the system is the data processor – everyone clear?)

Information Governance: Key Points

We are only dealing with exomic data

- at most, 2% of genome

The data is only stored on HPC temporarily

No data which could be used to identify the patient needs to be held at ICR

- we can mandate that it is never stored by us

ICR policy says genomic data may be stored, it is the act of trying to obtain the identity of the individual from that data which is prohibited

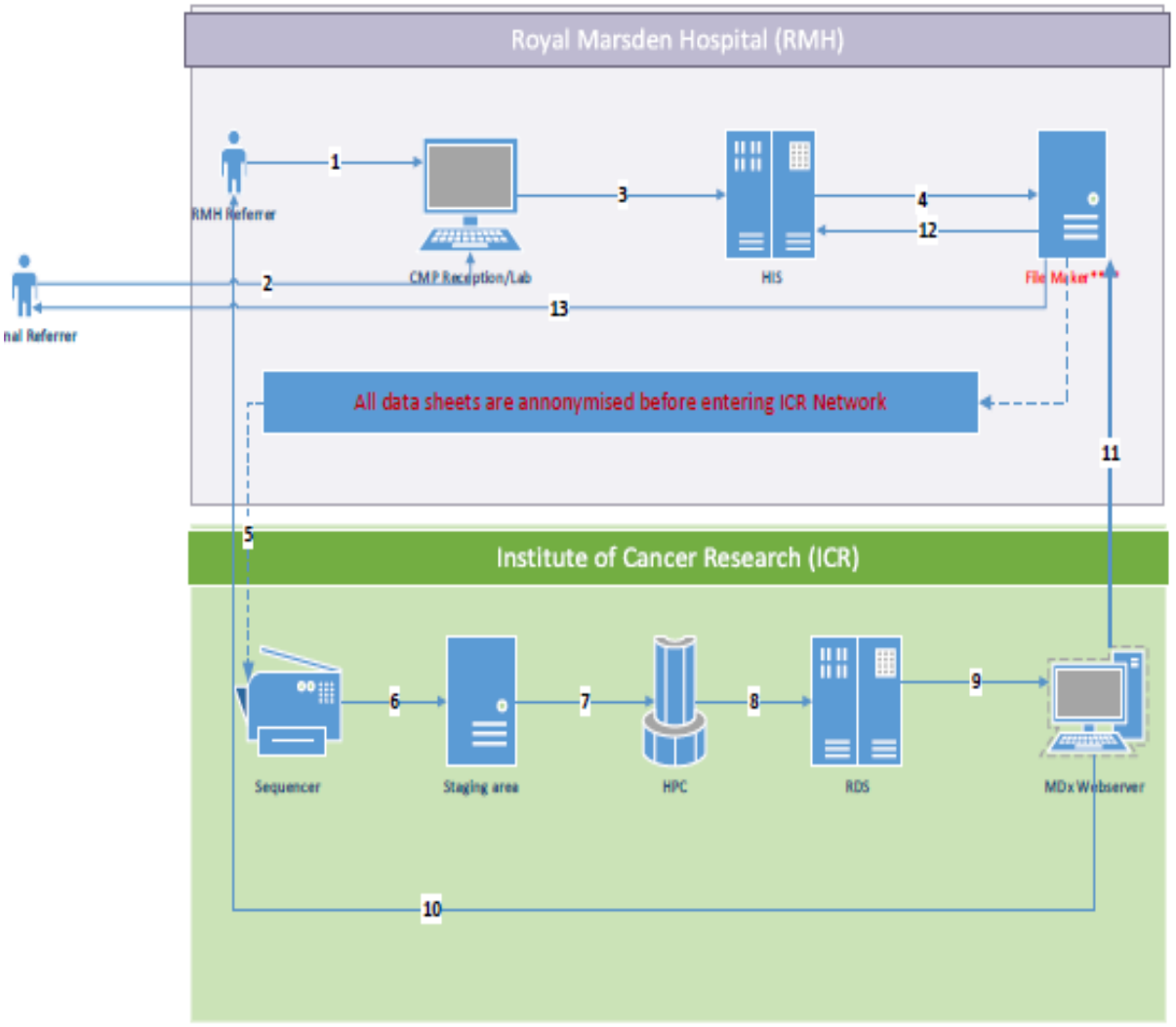
The RMH staff also have “visiting worker” status at ICR

- Manages identities, puts them under all IT and SciComp policies

RMH was required to conduct a Data Protection Impact Assessment (DPIA)

The upshot is that there was no IG reason why, with the correct workflow, we cannot run this type of clinical service on a research HPC system without major adaptations

RMH and ICR Data Flow Diagram



Data flow description

1. Internal referral details with patients information gets sent to CMP reception/Lab.
2. External referral details with or without anonymised ID gets sent to CMP reception/Lab
3. Referral information (internal & external) feeds into HIS
4. Referral information is manually transferred in XML format to File Maker.
5. Datasheet is downloaded and transferred to relevant machine in ICR
6. Sequencing data generated feeds into a temporary staging area.
7. Sequencing data feeds into High Performance Computing (HPC) cluster for analytics.
8. Result is saved to the ICR's Research Data Store (RDS).
9. Sequencing data is viewed and analysed via MDX Webserver
10. Genomics test is shared directly with referrer via email bypassing file maker. No personal information is involved as data is received with anonymised data.
11. Results requiring reporting are manually transferred to File Maker where de-identified patient data is re-united with their identifiers.
12. Genomics test result from internal referrers feed into HIS.
13. Genomics test result from external referrers gets sent via secure NHS.Net email.

**Note: File Maker sits on ICRs Network but can only be accessed by RMH Staff Using RMH Login details. A project to encrypt the back end (database) of File is at the Final stage.

Steps 4 and 12 include are manually transferred process. Clarification of this manually transferred process is as follows: The operator will manually download the data to a shared drive *all* within the same network drive that sits behind the fire wall on the same desk top/pc and only use a desk top/pc for this process.

RM has two ways to analyse sequencing data:
1. For MiSeqs, RM downloads the sample sheet from FileMaker and manually uploads it to the sequencers to start sequencing. The sample sheet is stored in the run folder and transferred to HPC for analysis. No patient data/information is stored.
2. For NextSeq/NovaSeq, the sample sheet is uploaded to MDx webserver after sequencing finishes and then copied to HPC for analysis.

***There is no direct data flow between RMH and ICR. ***

Research Data Store (RDS)

- Resilient: Based over three geographically separate sites
- Replicate storage with snapshots taken daily
- Very fast access (2GB/sec)—to and from various element
- Defined policy for appropriate and secure use

Genomic Analysis: Timeline

VALIDATION

STANDARDISATION

TURNAROUND TIMES

QUALITY

REGULATORY

LAB – 3 days

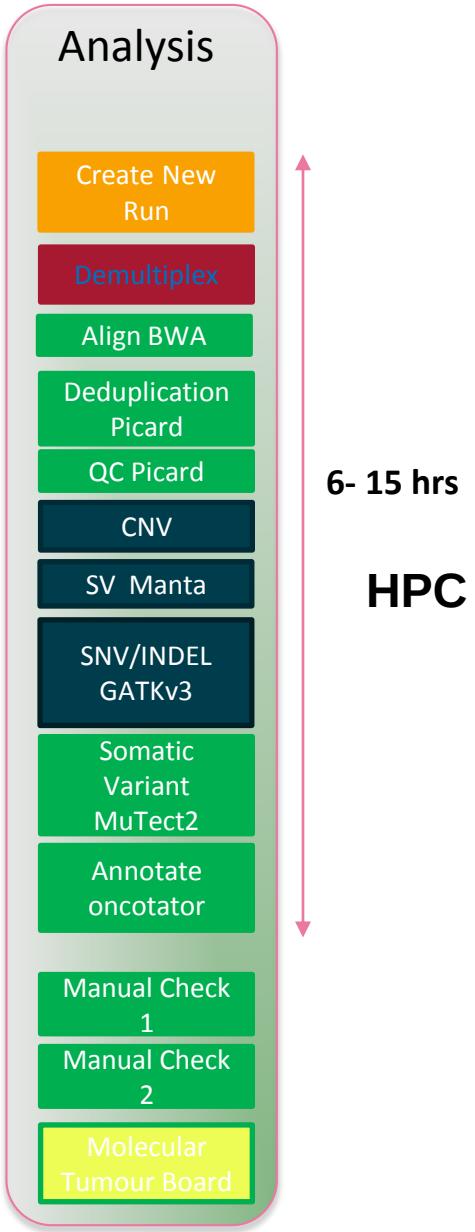
DNA Sequencer

HPC up to 1 day

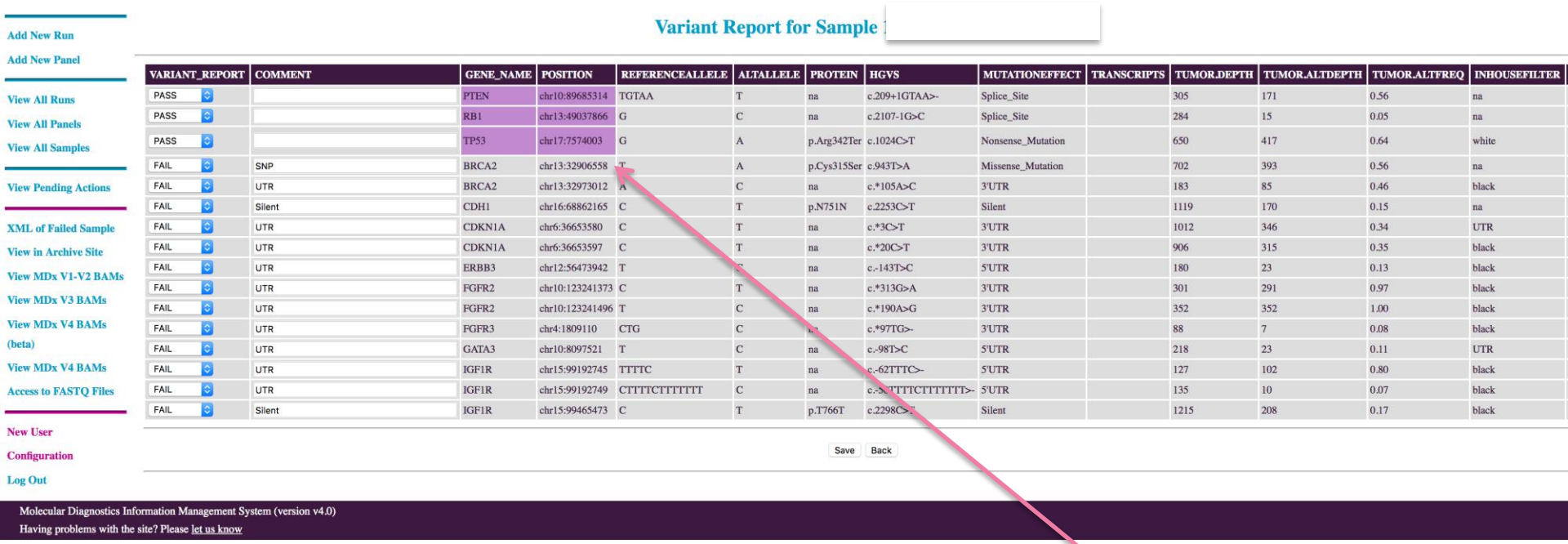
Summarised output

Check

Report for Clinical Action



Webserver V4



VARIANT_REPORT	COMMENT	GENE_NAME	POSITION	REFERENCEALLELE	ALTALLELE	PROTEIN	HGVS	MUTATIONEFFECT	TRANSCRIPTS	TUMOR.DEPTH	TUMOR.ALTDEPTH	TUMOR.ALTFREQ	INHOUSEFILTER
PASS		PTEN	chr10:89685314	TGTAA	T	na	c.209+1GTAA>-	Splice_Site		305	171	0.56	na
PASS		RB1	chr13:49037866	G	C	na	c.2107-1G>C	Splice_Site		284	15	0.05	na
PASS		TP53	chr17:7574003	G	A	p.Arg342Ter	c.1024C>T	Nonsense_Mutation		650	417	0.64	white
FAIL	SNP	BRCA2	chr13:32906558	T	A	p.Cys315Ser	c.943T>A	Missense_Mutation		702	393	0.56	na
FAIL	UTR	BRCA2	chr13:32973012	A	C	na	c.*105A>C	3'UTR		183	85	0.46	black
FAIL	Silent	CDH1	chr16:68862165	C	T	p.N751N	c.2253C>T	Silent		1119	170	0.15	na
FAIL	UTR	CDKN1A	chr6:36653580	C	T	na	c.*3C>T	3'UTR		1012	346	0.34	UTR
FAIL	UTR	CDKN1A	chr6:36653597	C	T	na	c.*20C>T	3'UTR		906	315	0.35	black
FAIL	UTR	ERBB3	chr12:56473942	T	-	na	c.-143T>C	5'UTR		180	23	0.13	black
FAIL	UTR	FGFR2	chr10:123241373	C	T	na	c.*313G>A	3'UTR		301	291	0.97	black
FAIL	UTR	FGFR2	chr10:123241496	T	C	na	c.*190A>G	3'UTR		352	352	1.00	black
FAIL	UTR	FGFR3	chr4:1809110	CTG	C	na	c.*97TG>-	3'UTR		88	7	0.08	black
FAIL	UTR	GATA3	chr10:8097521	T	C	na	c.-98T>C	5'UTR		218	23	0.11	UTR
FAIL	UTR	IGF1R	chr15:99192745	TTTTC	T	na	c.-62TTTC>-	5'UTR		127	102	0.80	black
FAIL	UTR	IGF1R	chr15:99192749	CTTTTCTTTTTTT	C	na	c.-58TTTTCTTTTTTT>-	5'UTR		135	10	0.07	black
FAIL	Silent	IGF1R	chr15:99465473	C	T	p.T766T	c.2298C>T	Silent		1215	208	0.17	black

Variants automatically called by pipeline in under 24hrs
Checked by clinical scientist and reported
Actionable variants identified
Treatments recommended

Up to 384 samples processed simultaneously.

Clinically relevant
Mutations Identified

EXAMPLE: Detection of an Actionable Mutation in Pediatric Cancer

RMH tested one young patient with an unusual cancer

Results showed a break and refusion of a chromosome variants

A literature search subsequently found a overseas case of another child with the same variants which described a treatment regime

The same regime has been applied to the RMH patient with positive results

Mayo Clin Proc Innov Qual Outcomes. 2017 Jul; 1(1): 111–116.
Published online 2017 Apr 27. doi: 10.1016/j.mayocpiqo.2017.04.003
PMCID: PMC6134903
PMID: 30225407

A Rare *STRN-ALK* Fusion in Lung Adenocarcinoma Identified Using Next-Generation Sequencing–Based Circulating Tumor DNA Profiling Exhibits Excellent Response to Crizotinib

Yan Yang, MD, PhD,[a] Shu-Kui Qin, MD, PhD,[d] Jian Zhu, MD,[b] Rui Wang, MD,[a] Yu-Mei Li, MD,[a] Zong-Yu Xie, MD,[c] and Qiong Wu, MD, PhD[a,*]

Next Steps

Current service is implemented with MiSeqs and a NextSeq500

A NovaSeq comes into service in the new year, also use ICR central Storage for larger amounts of data

HPC is the only realistic way to deliver high throughput genomic analysis with a clinically acceptable turnaround time.

Thinking about Hybrid HPC…

Costs, Reliability and compliance with Information Governance policies are important considerations from the beginning

Service has been successfully implemented and is delivering clinically actionable results, benefiting patients without having to make major changes to the ICR's research services.

Acknowledgements

Thanks go to the Clinical Molecular Pathology team at RMH, particularly Lina Yuan and Sabri Jamal

The Scientific Computing team at ICR, particularly Ben Talbot and Dugan Witherick

The ICR Information Governance team