2024 External Quality Assessment for

Next Generation Sequencing Somatic

Scheme Summary Report

EMQN CIC
Dr Simon Patton
EMQN CIC Office
3rd Floor, ICE Building
3 Exchange Quay
Salford, M5 3ED, UK
Tel: +44 (0)161 757 1591
Email: office@emqn.org
Website: www.emqn.org

Genomics Quality Assessment (GenQA)
Professor Sandi Deans
Laboratory Medicine, NHS Lothian
NINE, Edinburgh BioQuarter
Little France Road
Edinburgh, EH16 4UX
Tel: +44 (0)131 242 6898
Email: info@genqa.org
Website: www.genqa.org

EMQN is a community interest company (CIC) registered in England and Wales (Companies House Registration: 12020789).

GenQA is operated under the legal entity of OUH NHS Foundation Trust and is provided from two sites

Dear Colleague, 30/06/2025

The 2024 Next Generation Sequencing (NGS) Somatic pilot external quality assessments (EQAs) are run as a collaboration between EMQN and GenQA. This Summary report includes assessment data from all participants for both EQAs. Your EQA provider is responsible for these EQAs, and all correspondence related to them should be directed to either EMQN or GenQA at the relevant address.

The assessment is now complete and your individual laboratory scores have been agreed. Please go to your EQA provider's website to download your Individual Laboratory Report (ILR), Data Quality Report and Variant Call Analysis Report.

EQA DESIGN AND PURPOSE

The aim of these EQAs is to assess data quality and accuracy of NGS analysis for somatic SNV (single nucleotide variant) and indel (insertion/deletions <50 base pair) variants

Please note these are pilot EQAs and as such Performance Criteria DO NOT apply.

These EQAs were designed to be **platform and gene-target independent**. Participating laboratories were provided with the following samples, with full testing consent (please note: the same tumour sample was provided for both EQAs):

- Tumour testing only (no matched sample provided) (for the purpose of this report we will call it '**Tumour-Only'**): one DNA sample extracted from fresh frozen Lung squamous cell carcinoma. This EQA was designed for laboratories performing somatic variant testing not requiring germline variant subtraction.

- Tumour with germline subtraction analysis (matched germline sample provided) (for the purpose of this report we will call it '**Somatic-Matched'**): two DNA samples, one extracted from fresh frozen Lung squamous cell carcinoma and a matched germline DNA sample extracted from fresh frozen normal lung tissue. This EQA was designed for laboratories performing analysis which involves the subtraction of germline variants from the variant analysis.

Participants were asked to process the sample(s) and subsequent generated data using their normal NGS procedure(s). Participants were asked to use their 'in-house' testing strategy which could include analysis of a single gene, gene panel and/or whole exome or genome. In addition, participants were encouraged to submit up to three different sets of results/data.

Data collection, quality control (QC), storage and analysis to EMQN / GenQA defined standards and requirements was subcontracted to a commercial company, Euformatics (http://euformatics.com/). Working with Euformatics has enabled the EQA providers to assess data quality and provide direct comparison of different methodologies.

Participants were required to submit a list of identified variants as VCF (Variant Call Format) file(s) and a BED (Browser Extensible Data) file. In addition, submission of raw and processed sequence data (e.g. BAM/CRAM, FASTQ) was encouraged. All submitted data was assessed and quality metrics for each sample were calculated and analysed, as well as the ability to identify appropriate variants. The cumulative data from all participants was used to provide averages for reference.

The EQA is designed and overseen by members of the NGS EQA specialist advisory group (SAG) set up by both organisations (Table 1).

Table 1: *SAG members*

Name	Role	EQA Affiliation
Dr Joo Wook Ahn	Chair	None
Dr Jonathan Coxhead	Member	None
Dr Bauke Ylstra	Member	None
Dr Paul Westwood	Member	None
Dr Chris Boustred	Member	None
Dr Erika Souche	Member	None
Dr Kevin Balbi	Member	None
Dr Joseph Halstead	Member	None
Dr Helena Ahlfors	Member	None
Ms Farrah Khawaja	Deputy Director	GenQA
Dr Dave Cregeen	Deputy Director	GenQA
Prof Sandi Deans	Director	GenQA
Dr Simon Patton	CEO	EMQN
Dr Weronika Gutowska-Ding	Scheme Organiser	EMQN

If you would like to become involved in this advisory committee, please contact the EMQN (office@emqn.org) or GenQA (info@genqa.org) offices for further information.

REPORTING REQUIREMENTS

This was a technical EQA scheme, and participants were not expected to submit variant classification or a clinical report on the results of their analyses. The following were requested:

- A **technical survey** describing the sequencing approach, bioinformatics pipeline, and internally defined quality thresholds (submitted online to the EQA Euformatics website).
- **VCF file** of detected variants (SNVs and small indels) mapped to **hg38/GRCh38 or hg19/GRCh37** - this list should be compiled after QC and region of interest (ROI) filtering (and after germline subtraction for the 'tumour with germline subtraction analysis' EQA).
  - **Version 4.x variant calling format (VCF) files should be submitted (VCF standard format defined at https://samtools.github.io/hts-specs/).**

  Additional requirements on top of the basic VCF format requirements:
  - ONLY variants with the filter set to '.' or to 'PASS' in the 7th column were assessed. Any other text in the 7th column excluded the variant from assessment.
  - **Please note that any variants filtered out were not assessed.**
- **BED file** defining (fully matching) the genomic co-ordinates of the ROI analysed**.** An optional second BED file could be uploaded to further limit the regions where variant calls should be assessed (the "clinical target"). BED files should have a minimum of: **Chr, Start, End**. A fourth name field was optional. Any overlapping regions were merged. Please see the BED standard format defined at Ensemble https://www.ensembl.org/info/website/upload/bed.html for format specifications. Illumina manifest files needed to be converted to the BED format.
- The corresponding **FASTQ file** for the data analysed. **All FASTQ files must have been compressed in the GNU zip format**, an open source file compression program (see http://www.gzip.org/). This is indicated by the .gz file extension.
- The corresponding **BAM or CRAM file** for the data analysed. **CRAM files** are now supported for submission.

RESULTS SUMMARY

Four documents[1] have been generated for EACH laboratory and these reports are accessible from your EQA provider's website account. The documents are as follows:

- **EQA Summary report** (this document) summarising all the results.
- An individual **Data quality report** containing selected quality metrics from the submitted FASTQ, BAM, and VCF files, benchmarked against the distribution of the same metric from other laboratories.
- An Individual **Variant consensus analysis report** containing a comparison of the variants reported by your laboratory against a list of consensus variants. Variants are classified as: concordant with the consensus ("Agree"); not concordant with the consensus ("Disagree"), not reported by your laboratory ("Missing"); reported by your laboratory but not present in the consensus ("Extra"); reported by your laboratory but could not be assessed against the consensus e.g. uncertain consensus, or length is >50bp ("Not Assessed").
- **Individual Laboratory Report (ILR)** – summarising the overall Performance status of your laboratory for this EQA.

2024 EQA Participation

This year, 97 laboratories successfully submitted 153 different datasets for the 'tumour-only' EQA (an increase of 10.1% in participating laboratories and 16.8% increase in submissions since 2023). For the 'somatic-matched' EQA 37 laboratories submitted 64 datasets (an increase of 8.8% in participating laboratories and 12.3% in number of submissions since 2023).

Submission Issues

The following tumour-only EQA submissions were excluded from the consensus set: five submissions which did not include a BAM file, three duplicate submissions, eight submissions which did not contain SNVs, 53 submissions which did not contain indels.

The following somatic-matched EQA submissions were excluded from the consensus set : 21 submissions which did not include a BAM file, seven submissions which included germline variants, one duplicate submission, one submission which did not contain SNVs, 13 submissions which did not include indels.

In some cases, the target region was not reported correctly, with the submitted BED file listing regions not adequately covered by the sequencing. For the participant consensus creation, this was addressed by admitting only those submissions into the consensus set where the corresponding BAM file was available and considering only those regions within the target where the BAM file indicated >100X read depth.

Additionally, issues with VCF file validity were common. According to the EBI VCF validator (https://github.com/EBIvariation/vcf-validator), only 31% of the submitted VCF files were considered valid according to the VCF specification (version 4.1, 4.2, or 4.3). It was possible to work around the file validity issues which did not impact the variant concordance analysis.

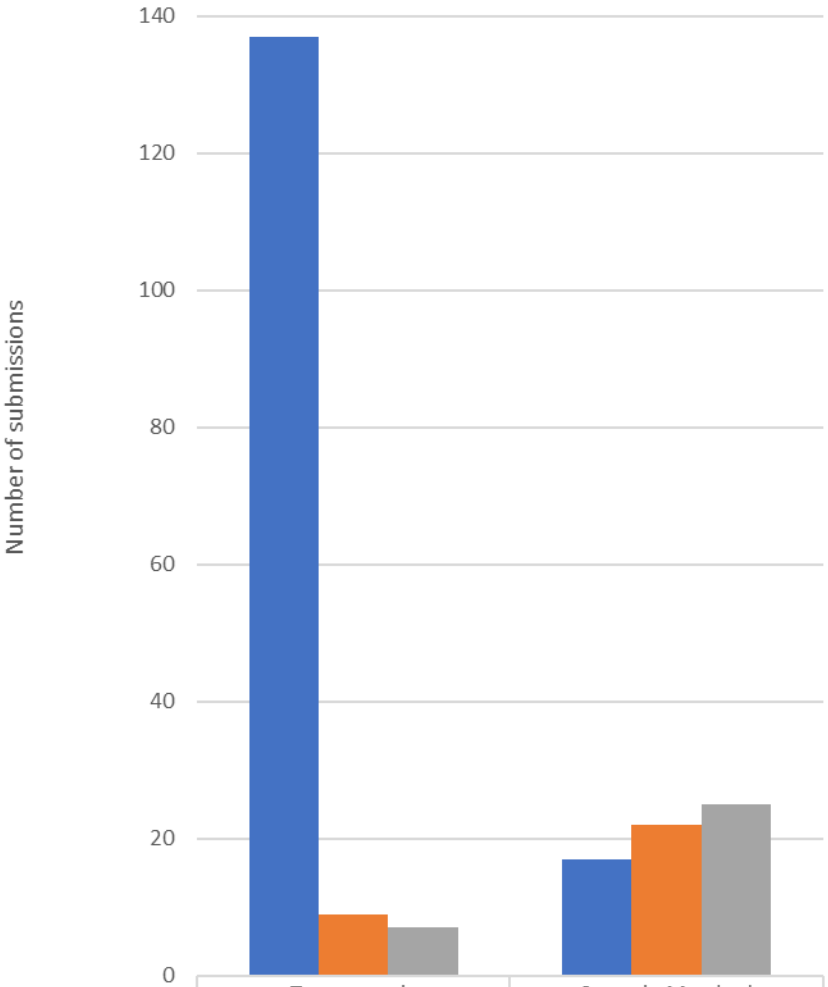
Variant calling and genes analysed

Most submissions for the 'tumour-only' (no matched sample) EQA were targeted sequencing panels (137/153, 89.5%). There was an increase in submissions for whole genome sequencing (from 14/57, 24.6% in 2023 to 25/64, 39.1%) in the 'somatic-matched' EQA (Figure 1).

[1] One report per submitted data set.

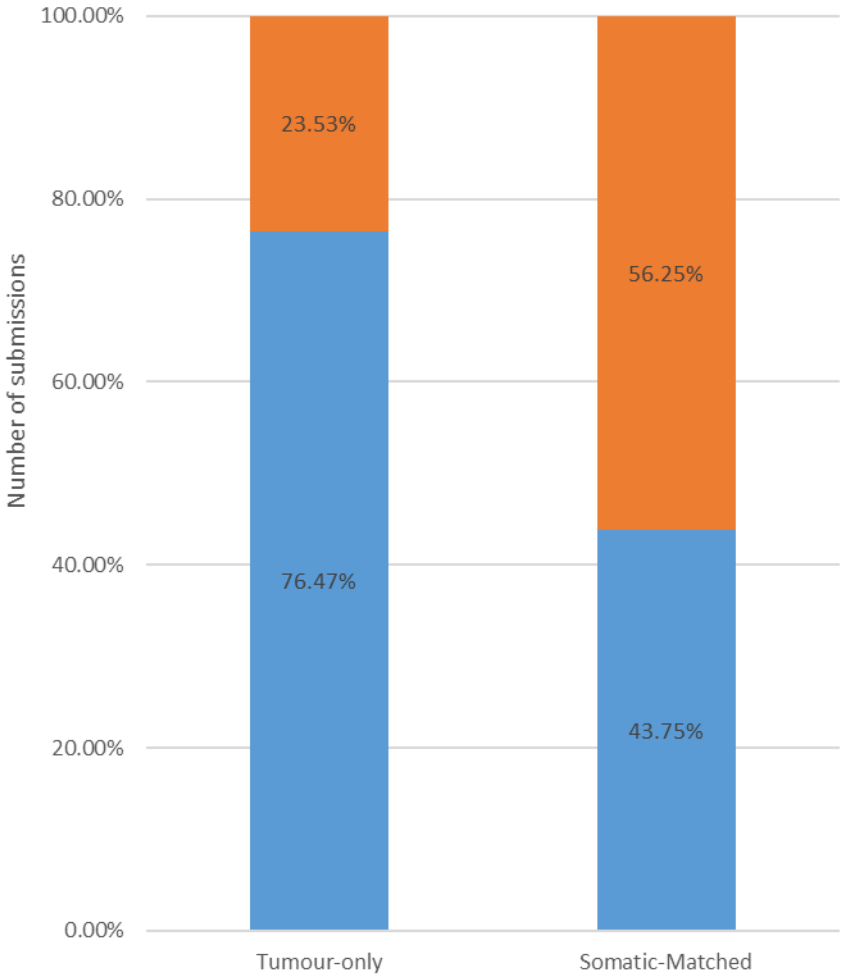
In both EQAs, up to three separate submissions were permitted from participants with results mapped either to hg19/GRCh37 or hg38/GRCh38 (Figure 2). A further increase was observed in laboratories using GRCh38 in the 'tumour-only' EQA – 23.5% (36/153) submissions were mapped to GRCh38 this year (an increase from 20.6% in 2023). The percentage of submissions mapping to GRCh38 in the 'somatic-matched' EQA remains similar at 56.3% (36/64), compared with 56.1% (32/57) in 2023.

Figure 1: *Sequencing approaches for submissions in the NGS somatic 2024 EQAs.*



	Tumour-only	Somatic-Matched
Targeted	137	17
Whole Exome	9	22
Whole Genome	7	25

Figure 2: *Reference genome used for submissions to the NGS Somatic 2024 EQAs.*

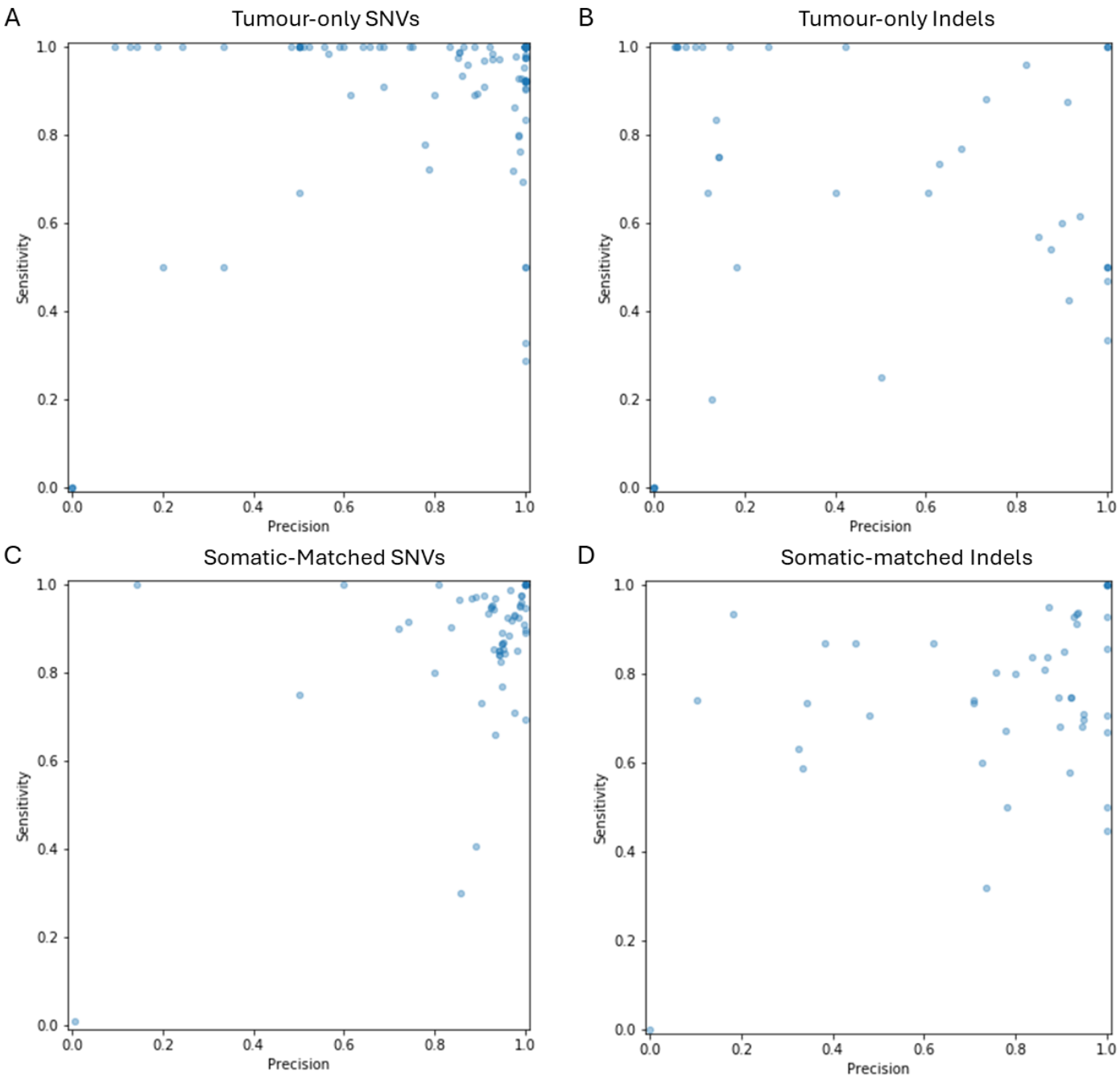


As for previous EQAs, submissions were mapped to GRCh37 which was used by the majority of participants. For the variant consensus analysis report, the variants were remapped back to the original genome build used by the participant.

Please NOTE: to ensure fairness for the laboratories that filter out germline variants, for the 'tumour-only' EQA we have based the results on the likely-somatic variants (variants with frequency in common population below 1%) only (see Appendix A for further explanation).

Figure 3 shows the sensitivity and precision for both EQAs submissions separated into SNVs and Indels.

Figure 3: *Sensitivity and precision of variant calling A) Tumour-only SNVs, B) Tumour-only Indels, C) Somatic-matched SNVs and D) Somatic-matched Indel submissions.*



Laboratory practice

NGS tumour-only EQA

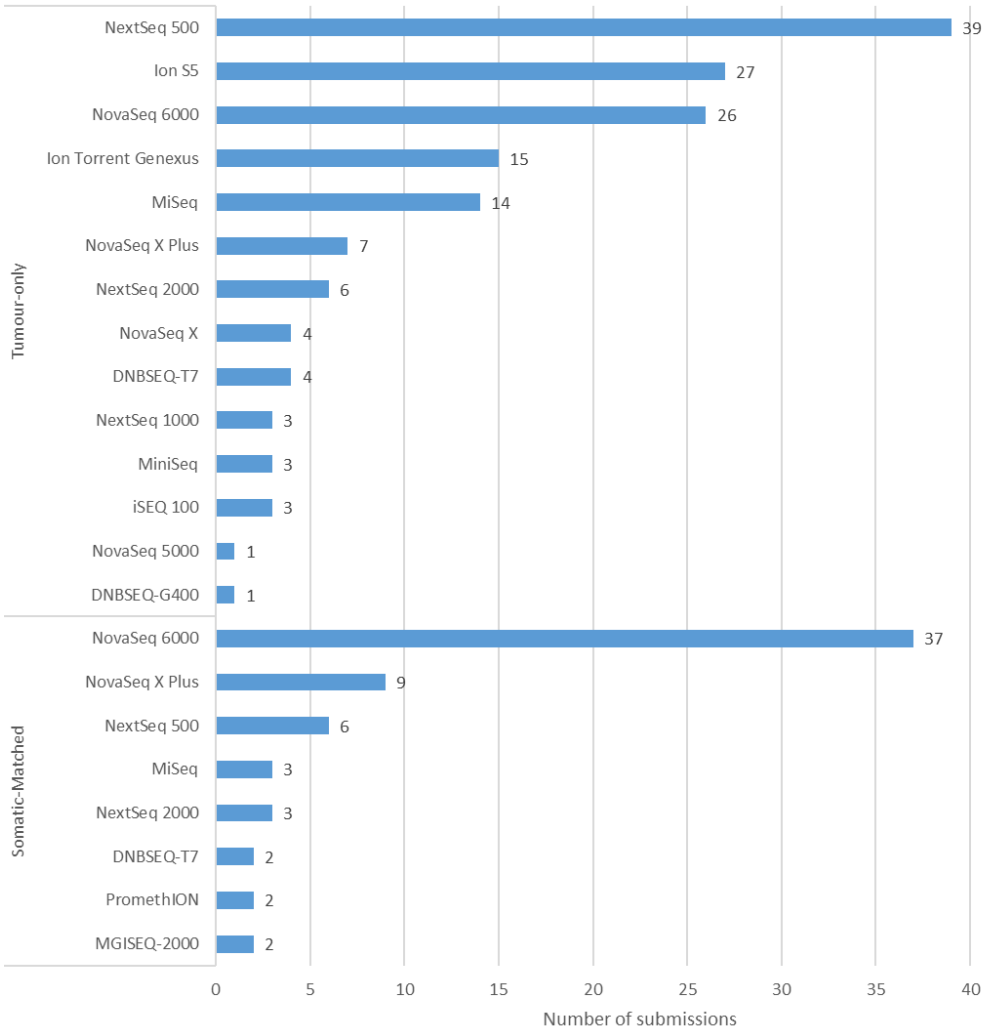
Most laboratories (69.3%) continue to use one of the Illumina NGS platforms, with NextSeq 500 remaining the most common (25.5%, a small decrease from 27.4% of all submissions in 2023). Life Technologies Ion Torrent platforms were selected by 27.5% of participants.

NGS somatic-matched EQA

As seen in the 'tumour-only' (no matched sample) EQA, the majority of laboratories (90.6%) continue to rely on Illumina NGS platforms. While the NovaSeq 6000 remains the most commonly used system, its usage has declined to 57.8% of submissions, down from 70.2% in 2023. The NovaSeq X Plus is the second most utilized platform, accounting for 14.1% of submissions. Additionally, there were two submissions using the Oxford Nanopore PromethION.

For the detailed breakdown of platforms used please see Figure 4.

Figure 4. *Sequencing platforms used by the participants.*



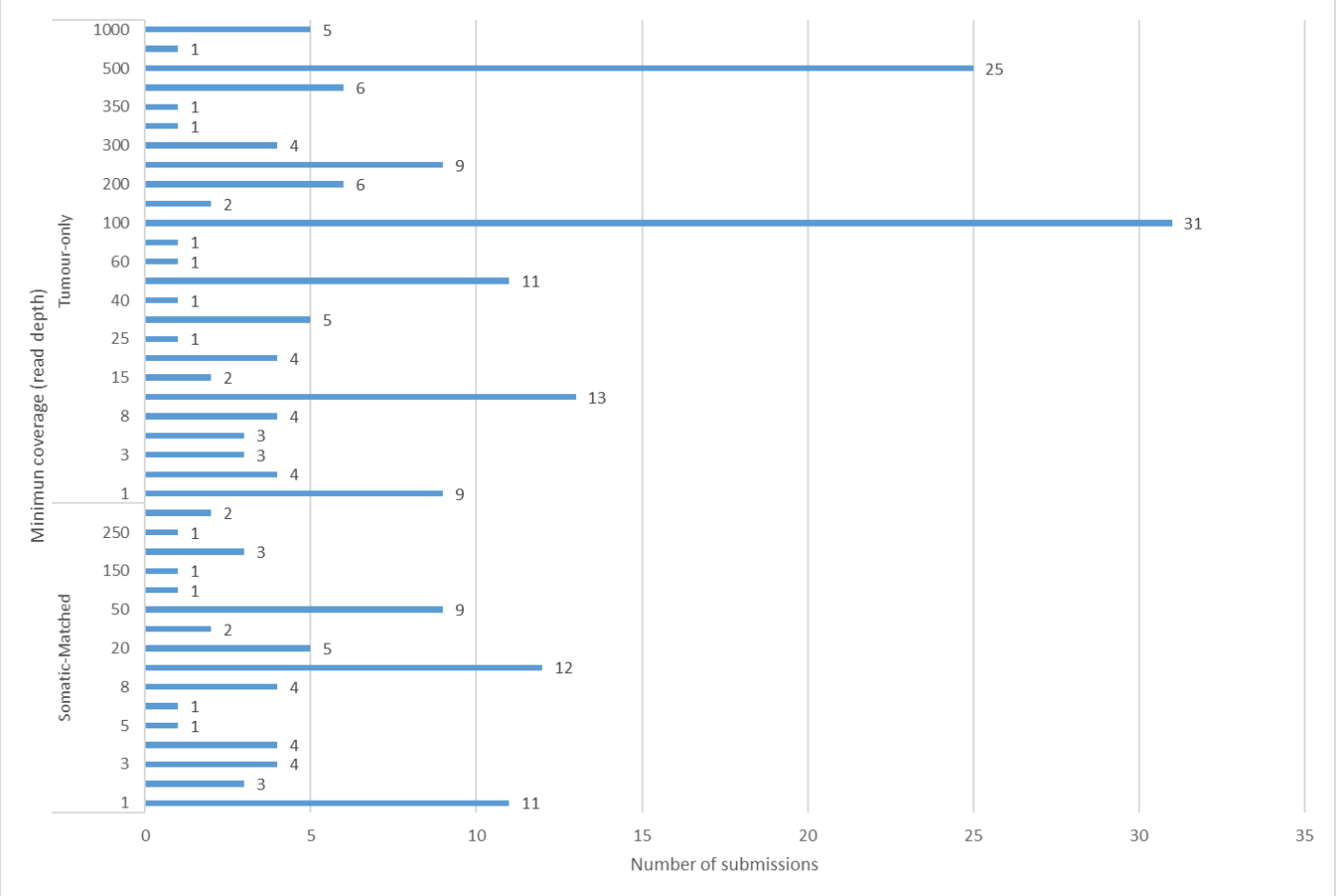
Commercial kits for library preparation were used by 83.7% of 'tumour-only' (no matched sample) EQA and 76.6% of 'somatic-matched' EQA participants. This is an increase from 2023 where they were used by 76.3% and 71.9% of participants respectively.

A range of capture kits were employed by participants, with ThermoFisher kits slightly surpassing Illumina in usage (26.2% vs. 25.5%) among those in the 'tumour-only' EQA. In contrast, for the 'somatic-matched' EQA, Illumina kits were the most commonly used, accounting for 31.3% of submissions and surpassing Twist Bioscience (26.6%) for the first time. No other brands exceeded 10% of submissions in either EQA.

The confidence in reporting diagnostic test results from NGS increased for the 'tumour-only' EQA, with 75.4% participants saying the results would be reported without confirming by an orthogonal method if the minimum coverage threshold was met (46.6% in 2023). In contrast, in the 'somatic-matched' EQA only 24.6% of participants said the same (decrease from 29.8% in 2023).

Figure 5 shows the distribution of minimum reportable coverage used to call a variant.

Figure 5: *Distribution of minimum sequencing coverage for reporting variants. Coverage is an average number of reads that align to known reference bases (also known as read depth).*



TP53 variant

TP53 variants are frequently associated with non-small cell lung cancer and specifically lung squamous cell carcinoma, being detected in up to 50% of patients. Consensus analysis identified a *TP53* variant in this tumour DNA sample: NM_000546.6:c.743G>A p.(Arg248Gln)

GRCh38 NC_000017.11:g.7674220C>T GCRh37 NC_000017.10:g.7577538C>T

Tumour-only EQA: 120/153 submissions (78%) targeted this position, with 117/120 (98%) calling the variant.

Somatic-matched EQA: 64/64 submissions (100%) targeted this position, with 60/64 (94%) calling the variant.

The average VAF of the called variant in the submission was 51%.

Bioinformatics analysis

Various in-house, commercial, platform provided and outsourced bioinformatics pipelines were used by participants of both EQAs (Table 2).

Table 2. *Bioinformatics strategies used by participants.*

	NGS (tumour-only)	NGS (somatic-matched)
Commercial	13.5%	8.6%
In-house Pipeline	50.1%	82%
Outsourced	7.8%	1.3%
Platform Provided	21.5%	8.1%
Other	7.1%	N/A

The proportion of 'somatic-matched EQA submissions using in-house pipelines remains high (82%), with the proportion of submissions for the 'tumour-only' EQA also increasing to just over 50% (from 38.2% in 2023).

Participants continue using Burrows-Wheeler Aligners (BWA) (85.5% 'tumour-only' and 74.6% 'somatic-matched' submissions). Interestingly, while only 2.1% of 'tumour-only' participants use DRAGEN, it was used by 17.1% of 'somatic-matched' submissions.

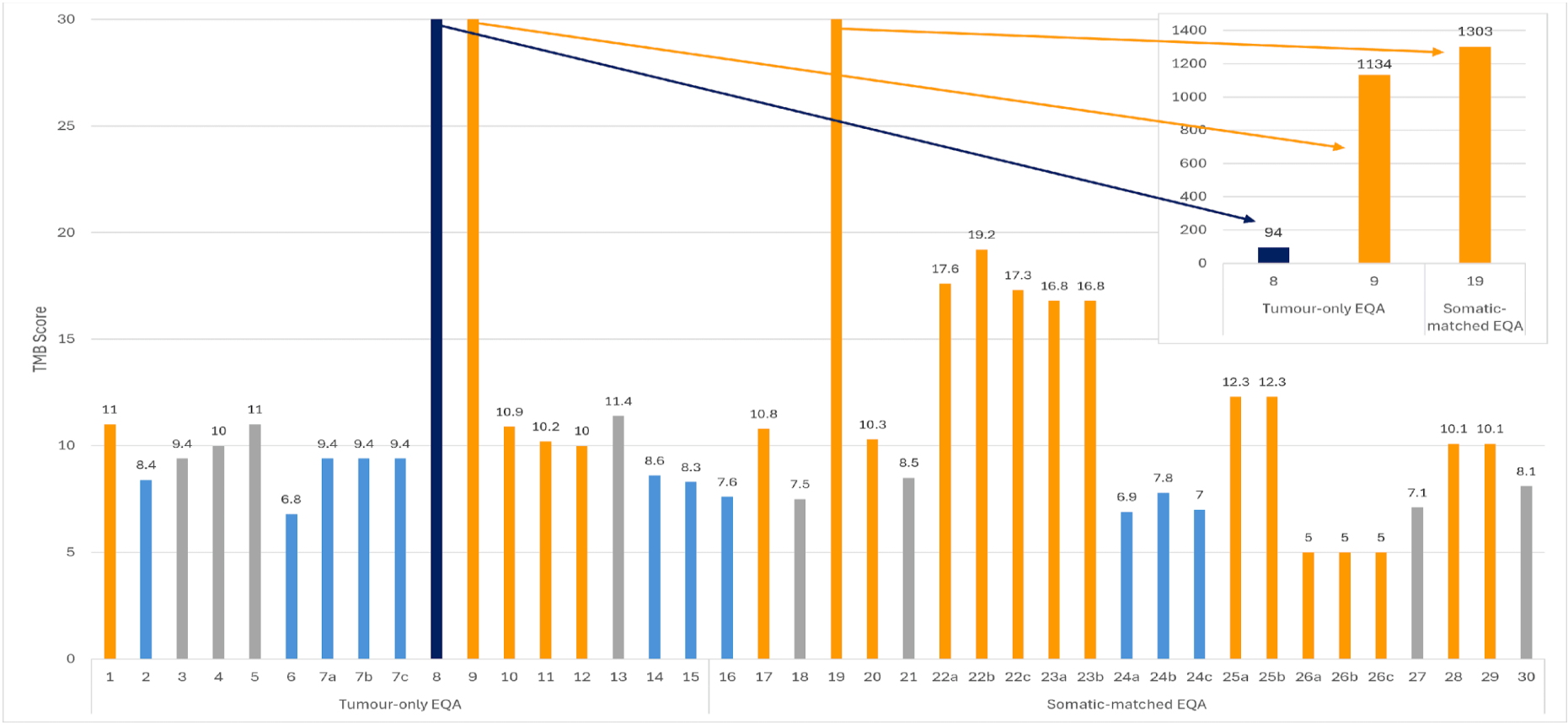
GATK tools were the most used choice for variant callers in the 'tumour-only' EQA at 42.6% and 'somatic-matched' EQA at 32.1%.

Tumour mutational burden (TMB)

The figure below (Figure 6) presents the results of Tumour mutational burden (TMB) reported for both EQAs. Thirty laboratories submitted forty TMB results (and increase from last year's 29 submissions by 19 laboratories). Twenty results (20/40, 50%) were classed by laboratories as 'high' TMB, eleven (27.5%) were classed as 'low', one (2.5%) was classed as 'intermediate', and the remaining eight (20%) were not classified. Cut off values varied but the most widely used appeared to be '10' for the threshold of 'high' TMB which was stated by 19 participants 19/30, 63%). Three participants reported extremely high TMB scores (94, 1134 and 1303) however the reason for this was unclear.

It should be noted that these results have not been verified by EMQN/GenQA and are for benchmarking purposes only.

Figure 6: *TMB results submitted by participants. Several participants submitted TMB results for more than one submission, designated a, b, c.*



Key:

High
Intermediate
Low
Not stated

CONFIDENTIALITY

Details of the confidentiality policies of each provider can be found:

- EMQN: https://www.emqn.org/participating-in-eqa/terms-conditions/ - In section 4.6 Performance evaluation.
- GenQA: https://genqa.org/confidentiality.php

SUBCONTRACTED ACTIVITIES

Your EQA provider does not subcontract activities such as EQA planning, evaluation of performance or the authorization of reports. However, some activities are subcontracted, for example the preparation of materials may be performed by suitably accredited providers.

FINAL COMMENTS

The assessment team would like to thank all participants for their hard work, prompt return of results and their co-operation during this exercise.

In 2025, separate somatic EQAs will continue to be offered for participants that require germline samples and for those that do not. Please make sure that you **only** participate in the 'tumour with germline subtraction analysis' EQA if your usual practices require subtraction of germline variants.

The purpose of the EQA service is to educate and facilitate the raising of standards. Thank you for participating in this EQA and we hope you have found it a useful EQA exercise. We look forward to your participation in the 2025 EQAs.

Kind regards,

Dr Simon Patton
CEO
EMQN

Professor Sandi Deans
Director
GenQA

REFERENCES

1. McGowan-Jordan, J., Simons, A., Schmid, M. (2016), ISCN 2016: An International System for Human Cytogenomic Nomenclature. S. Karger, Basel.
2. http://onlinelibrary.wiley.com/doi/10.1002/humu.22981/abstract https://varnomen.hgvs.org/
3 ISO15189:2022(en), Medical laboratories — Requirements for quality and competence. https://www.iso.org/obp/ui/#iso:std:iso:15189:ed-4:v1:en4.
4. Deans ZC, *et al*. Recommendations for reporting results of diagnostic genomic testing. Eur J Hum Genet. 2022 Apr 1. doi: 10.1038/s41431-022-01091-0. Epub ahead of print. PMID: 35361922.

AUTHORISATION/APPROVAL

This document has been authorised/approved on behalf of EMQN by: Dr Simon Patton on 30/06/2024

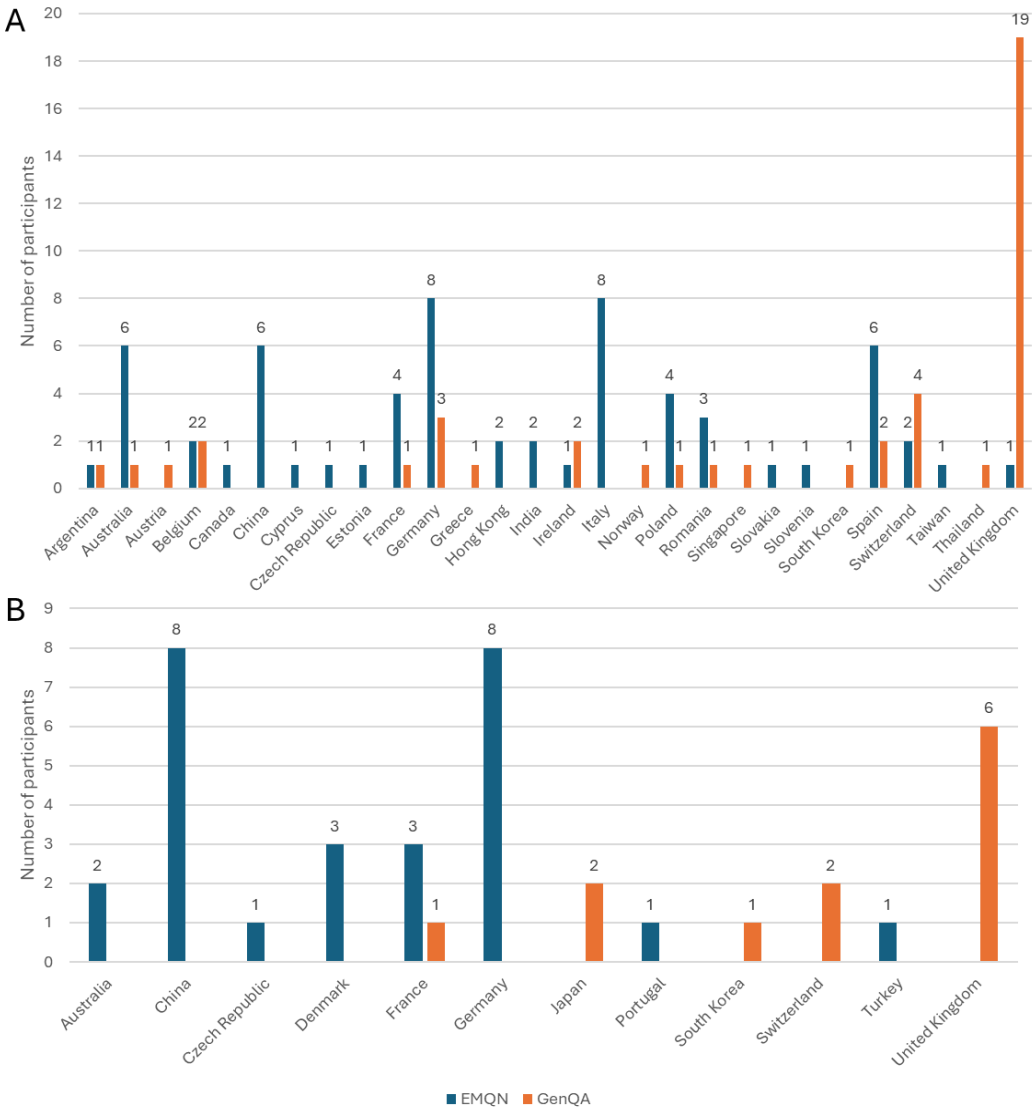
CEO of EMQN

APPENDIX A – PARTICIPATION

Table 3: *Number of participating laboratories*

	NGS (Somatic)	NGS (Somatic-matched)
Number of registrations	**106**	**39**
Number of withdrawals	**2**	**1**
Number of laboratories that did not submit results	**7**	**1**
Total number participating laboratories	**97**	**37**
Number of laboratories submitting more than one set of results	**40**	**17**

Figure 7: *Participating countries for both EQAs: A. NGS 'tumour-only' and B. NGS 'somatic-matched'*



APPENDIX B - SAMPLES PROVIDED AND GENERATING PARTICIPANT CONSENSUS

DNA extracted from the same tumour sample was provided for both EQAs i.e., fresh frozen squamous cell carcinoma of the lung. The additional germline sample for the 'somatic-matched' EQA was DNA extracted from normal fresh frozen lung tissue from the same patient. A participant consensus variant set was established separately for both EQAs.

The participant consensus required at least five submissions to target each variant position at >100X read depth with >2/3 agreeing on the called alternative allele. Where there was less than 1/3 support for a variant call, the consensus was taken to be the absence of a variant; where there was more than 2/3 support for a variant call, the consensus was taken to be the presence of a variant. For positions where the agreement was between 1/3 and 2/3, the consensus was considered to be inconclusive. Only positions within exons expanded +/- 30 bp in both directions were considered.

- The participant consensus variant set was computed separately for SNVs and Indels.
- Any duplicate submissions were removed from the consensus.
- As there were not enough NGS submissions that used GRCh38 to create a participant consensus, these submissions were remapped to GRCh37 to create a consensus sequence for all submissions. However, for the variant consensus analysis report the variants were remapped back to the original genome build used by the participant.
- For both EQAs we have NOT assessed variants common in the population (allele frequency over 1%).

The individual participant submissions were normalised by splitting multi-nucleotide variants (MNV) into SNVs and left-aligning indels and evaluated against the participant consensus following the Global Alliance for Genomics and Health's (GA4GH) guidelines (https://www.ga4gh.org/).

Only variants passing filters were assessed. Variants where the variant allele fraction was below the level of detection reported by the participant or 5% (whichever is higher) were not assessed.

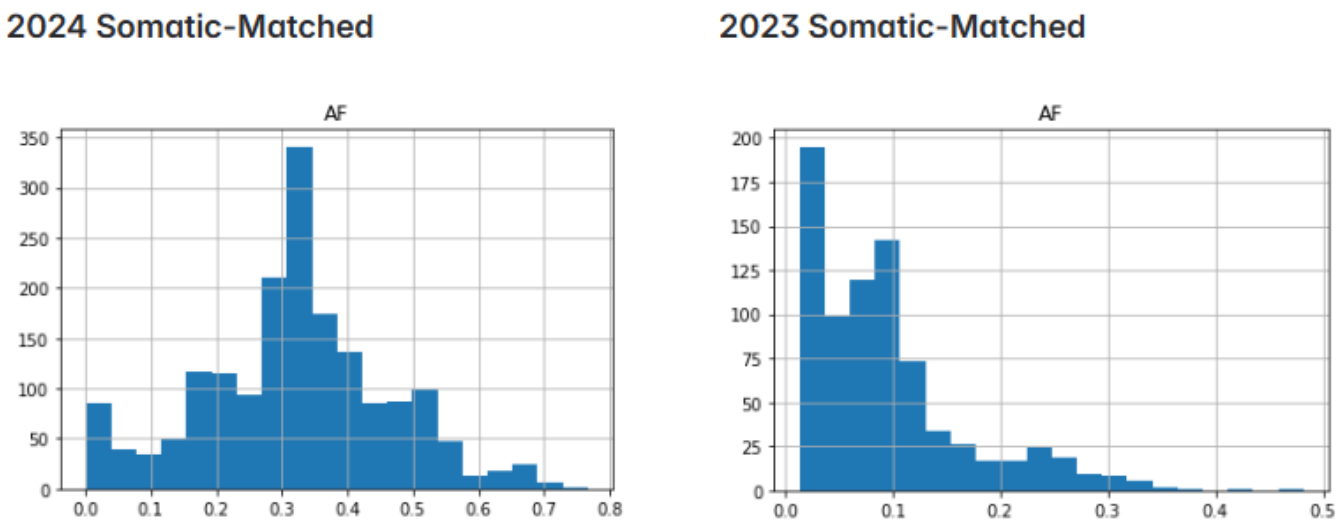
Observations with generating the consensus for the 'tumour-only' EQA

This year, among SNV calls made by at least one-third of participants at a given position, only about 10% were found to have strong consensus support. This represents a significant difference from the 2023 Somatic scheme, where approximately 90% of such SNV calls achieved strong consensus.

As the same tumour sample was used for both NGS somatic EQAs it was possible to compare variant calling across EQAs. Using data from the 'tumour with germline subtraction analysis' EQA a generally higher VAF (around 35%) was observed for the 2024 tumour sample compared to that used in 2023 (less than 10%) (Figure 8).

Using the NGS 'somatic-matched' consensus data, it became evident that true somatic variants were consistently supported by strong consensus in the 'tumour-only' scheme. Variants lacking strong consensus support may be germline variants and many could be excluded by filtering out variants present at higher than 1% in gnomdAD.

Figure 8. *Variant allele fraction of the somatic-matched consensus variants 2024 (left) and 2023 (right)*



APPENDIX C – INDIVIDUAL LABORATORY FEEDBACK

Data quality and variant analysis are reported separately in the following documents: (a) **Data Quality Report** (PDF), and (b) **Variant Consensus Analysis Report** (Excel). These reports, plus the **Individual Laboratory Report** (ILR) are specific to YOUR laboratory and are designed to give feedback which you can use to help improve your processes. The following pages provide an explanation of the content of each report, both of which can be downloaded from your respective EQA organiser's scheme webpage.

Data Quality Report [2]

The Data Quality Report contains our assessment of a range of selected quality metrics from the data submitted by your lab for the three file types: FASTQ, BAM/CRAM and VCF. We have also provided a brief glossary describing the applied quality metrics and an explanation about why it is important to consider them for validation and ongoing QC of your NGS processes.

Variant Consensus Analysis Report [3]

Reported variants were normalised and compared against the consensus variants computed from participant submissions (EQA Participant consensus variant set). The number of participating laboratories agreeing with the EQA participant consensus genotype (participant consensus ratio) is shown (Figure 9). Following GA4GH's recommendations we have stratified the results by SNVs and Indels.

Classification:

- **Agree** – participant's variant that matched the participants' consensus (True positive)
- **Extra** – participant's variant which was not present in the participants' consensus (False positive)
- **Missing** – participant has missed a variant present in the participants' consensus genotype (False negative)
- **Not Assessed** – participant's variant could not be assessed against the consensus (e.g. uncertain consensus)

[2] Each laboratory's report will be called Data quality report (NGS pilot 2024) Somatic / Data quality report (NGS pilot 2024) Somatic-Matched

[3] Each laboratory's report will be called Variant consensus analysis report (NGS pilot 2024) Somatic / Variant consensus analysis report (NGS pilot 2024) Somatic-Matched

Figure 9: *An extract from the Variant consensus analysis report. This data describes each laboratory's submitted variants (VCF file) cross-referenced against the consensus variants that should have been detected based on each laboratory's region of interest (BED file).*

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Analysis set:	2023 Somatic (GRCh37)											
2	Reference genome:	GRCh37											
3	AF threshold submitted:	2.00%											
4	AF threshold applied:	5.00%											
5	Stated VCF content:	Germline and somatic variants											
6													
7	Region:	all		high confidence									
8	Type:	snp	indel	snp	indel								
9	True positives:	3	0	2	0								
10	False positives:	1732	1435	1536	1140								
11	False negatives:	0	0	0	0								
12	Sensitivity:	100.00%		100.00%									
13	Precision:	0.17%	0.00%	0.13%	0.00%								
14	F-Score:	0.35%		0.26%									
15													
16	Variant position	Type	Gene	Submitted variation	Submitted AF	EQA variation	EQA AF	EQA consensus rat	Population frequency	Classification	Notes	Region	
17	1:36932017	snp	CSF3R	T/C	0.57%					Not Assessed	Below AF threshold	high confidence	
18	1:115258744	snp	NRAS	C/G						Extra		high confidence	
19	1:115258744	snp	NRAS	C/T						Extra		high confidence	
20	1:120457886	snp	NOTCH2	T/C	0.35%				0.00%	Not Assessed	Below AF threshold	high confidence	
21	10:89692897	indel	PTEN	AAAGGGACGAACTGG/A						Extra		high confidence	
22	11:32413561	snp	WT1	G/A	0.16%				0.00%	Not Assessed	Below AF threshold	high confidence	
23	11:32413565	snp	WT1	C/A						Extra		high confidence	
24	11:32413565	snp	WT1	C/G						Extra		high confidence	
25	11:32414254	snp	WT1	A/G	0.52%					Not Assessed	Below AF threshold	high confidence	
26	11:32414256	snp	WT1	A/G	0.39%					Not Assessed	Below AF threshold	high confidence	
27	11:32414263	snp	WT1	G/A						Extra		high confidence	

APPENDIX D – GLOSSARY OF QUALITY METRICS TERMS

Variant call assessment

- **Sensitivity**

Proportion of actual positives that are correctly identified as such. Defined as Sensitivity = TP / (TP + FN), where TP is the number of true positives and FN is the number of false negatives. Also known as recall.

- **Precision**

Proportion of actual positives among all reported positives. Defined as Precision = TP / (TP + FP), where TP is the number of true positives and FP is the number of false positives.

- **F-Score**

Harmonic mean of sensitivity and precision. Defined as F-Score = 2 × Sensitivity × Precision / (Sensitivity + Precision). Also known as the F_1 score.

FASTQ file

- **Base quality (Phred quality score)**

Base quality score (Phred quality score) is a measure of the quality of the identification of the nucleotide bases generated by automated DNA sequencing. Phred quality scores are assigned to each nucleotide base call in automated sequencer traces and are used for assessment of sequence quality, recognition and removal of low-quality sequence, and determination of accurate consensus sequences.

  - **Consideration for validation:** An acceptable raw base call quality score threshold should be established during validation. Filters should be established to eliminate base calls lower than the established quality score. The tolerance for low base quality is higher in long-read than in short read technologies because the sequence length and accuracy at the base level is less critical for alignment.
  - **Consideration for ongoing quality control:** The quality of the base calling signal should be monitored over time and among used instruments by examining the base quality across reads for each sequencing run.

BAM file

- **Uniformity (%)**

  The percentage of bases on target covered at 0.1 x median coverage.

- **Reads on/off target (%)**

  The fraction of reads mapped to the target region as a function of the total amount of (mapped and unmapped) reads. This is a raw, or minimum value for the amount of informative reads, while the read enrichment disregards unmapped reads and provides a filtered, or maximum value for the amount of informative reads. The fraction of reads on the target region is equal to 100 minus the fraction of off target reads.

  - **Consideration for validation:** Following the on or off target fraction will give an idea about the target specificity. A high off target percentage means low specificity of the run. The off target depicts the non reliable reads and should be established during validation.
  - **Consideration for ongoing quality control:** It is important to keep a record of the fraction of reads on the target region because it provides confidence in the validity of results in the region of study and will make sure that no regions have been omitted in the test.

- **Error rate on target**

  The proportion of mismatched bases on target.

- **Insert size**

  Number of bases between paired-end reads.

VCF file

- **Ti/Tv ratio**

  This is the ratio of the number of transitions (substitutions between purines or between pyrimidines) to the number of transversions (substitutions between a purine and a pyrimidine) in the variants called in an experiment. The Ti/Tv ratio is close to 2.1 for the whole human genome and closer to 2.9 in exons, and depends among other on the GC content of the DNA fragment considered. It is therefore specific to the target area in a normal genome. Cancer genomes show aberrant evolution and can have quite different Ti/Tv ratios.

  - **Consideration for validation:** The ratio of transitions to transversions (Ti/Tv) should be established separately for each target capture protocol and compared to published values.
  - **Consideration for ongoing quality control:** The Ti/Tv ratio should be monitored with every sample to detect a change in test performance. When the Ti/Tv ratio is lower or higher than expected, this is an indication that the quality of base calls was low, and potentially contains errors.