How Good is your Next Generation Sequencing? Lessons Learned from Genomics EQA Schemes

Dr Weronika Gutowska-Ding

EMQN CIC

3rd March 2023

Zurich

Content

- EMQN and External Quality Assessment
- EQA in Genomics
- NGS and EQA
- Developing a ‘consensus genome’
- Results of 2022 NGS EQA
- Conclusions

EMQN and External Quality Assessment

What is EMQN?

- International External Quality Assessment (EQA) / Proficiency testing (PT) provider
  - EQA = External audit of laboratory performance
    - Genotyping, interpretation and reporting accuracy
    - Guided by 'best practise'
  - Started in 1997 with 1 scheme (HD) at NHS hospital in Manchester, UK
    - EU FP4 funding
  - Modelled on UK NEQAS (GenQA) schemes (still collaborate on several schemes)
  - Since 2019, EMQN has been a Community Interest Company (CIC) (registered in England and Wales, 12020789)
    - Profit for purpose – 51% invested back in to services to benefit our community (guidelines, new EQA etc)

Mission: helping ensure diagnostic genomic laboratory test results are accurate, reliable and comparable wherever they are produced

- Accredited to ISO 17043

UKAS PROFICIENCY TESTING

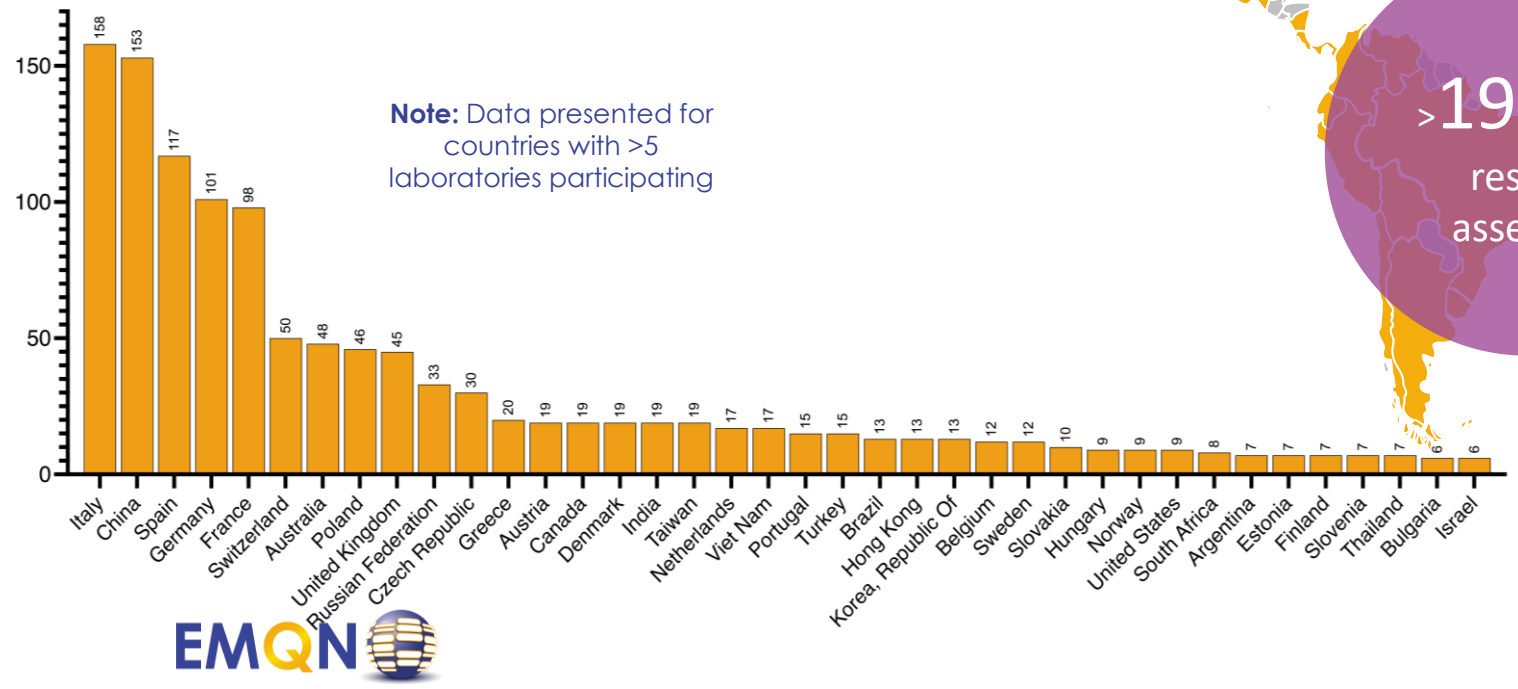
4367

EMQN is an **ISO17043 accredited** provider of EQA schemes

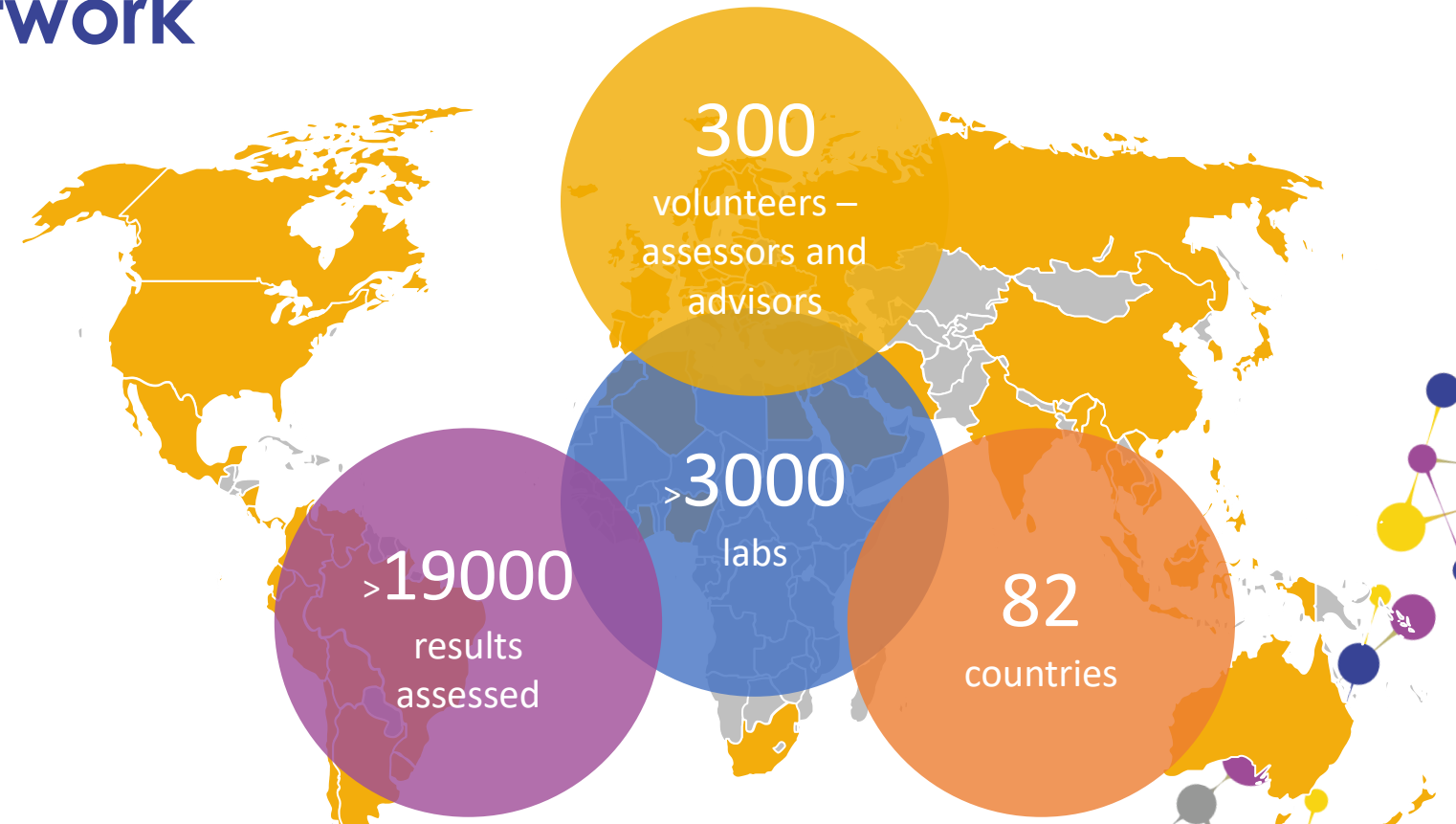
- >60 EQA schemes
- Themed around germline genetics, molecular pathology and technical schemes
  - e.g., NGS, virology, Pharmacogenetics and pre/postnatal genetics
- Plus Inter laboratory comparisons (ILCs) & ring trials

EMQN

EMQN – a global network



Note: Data presented for countries with >5 laboratories participating



EMQN

EQA in Genomics

Implications of Genetic Testing

- Highly predictive for future health
- Carried out at any stage of life and be applied for pre-natal or pre-implantation diagnostics
- Relevant to healthy people as well as to those with an unhealthy condition
- May have implications for the relatives of the person tested

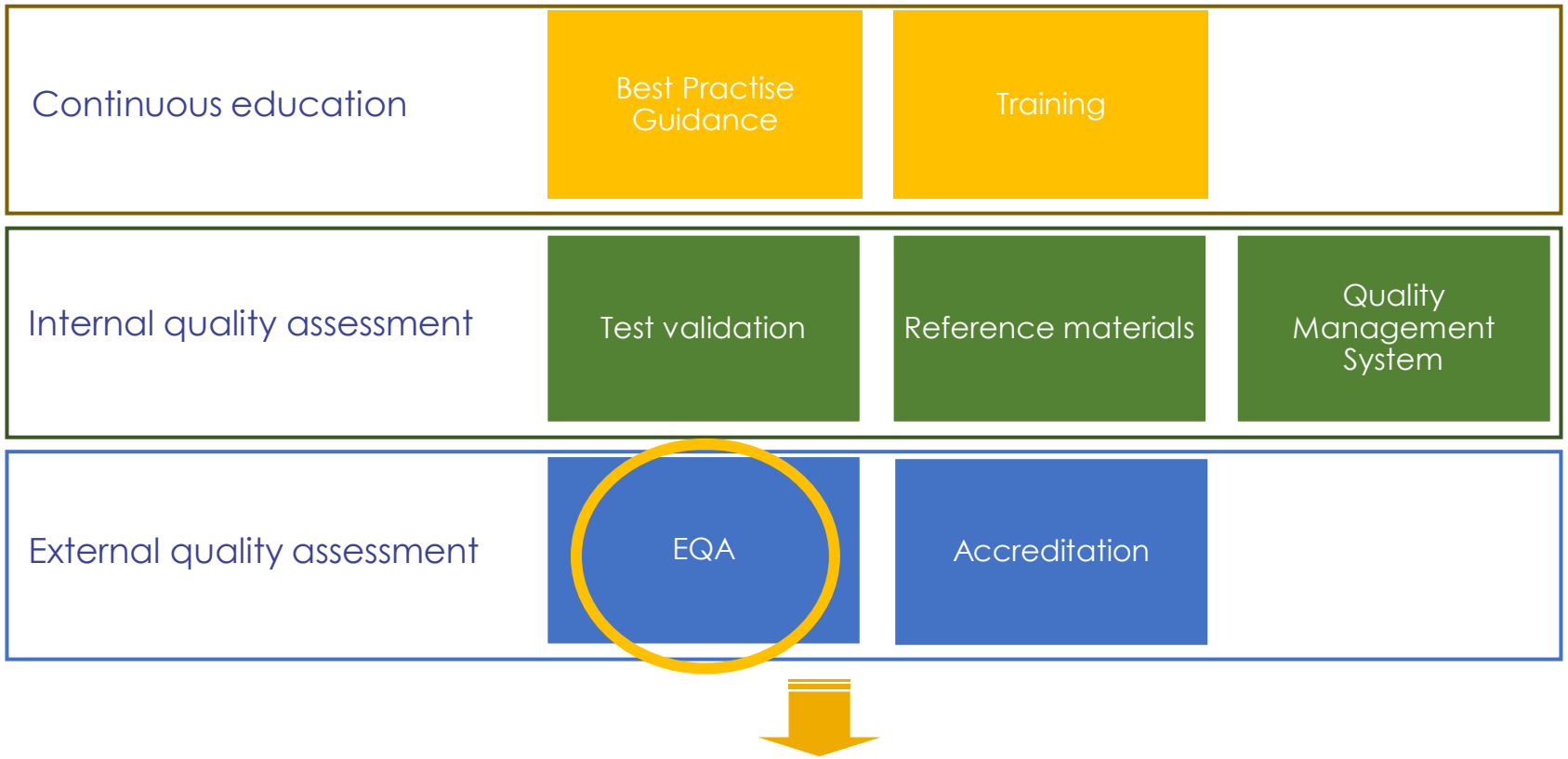
- **The genotype established by a single laboratory test is usually not repeated and forms a permanent part of the medical record of the patient**

EMQN

Why should laboratories participate in QA?

- Test results are potentially life changing
- Errors can result in:
  - Inappropriate reproductive decisions
  - Inappropriate decisions regarding prophylactic surgery/disease monitoring
  - Failure to use the most effective treatment

Quality Assured Genomic Testing



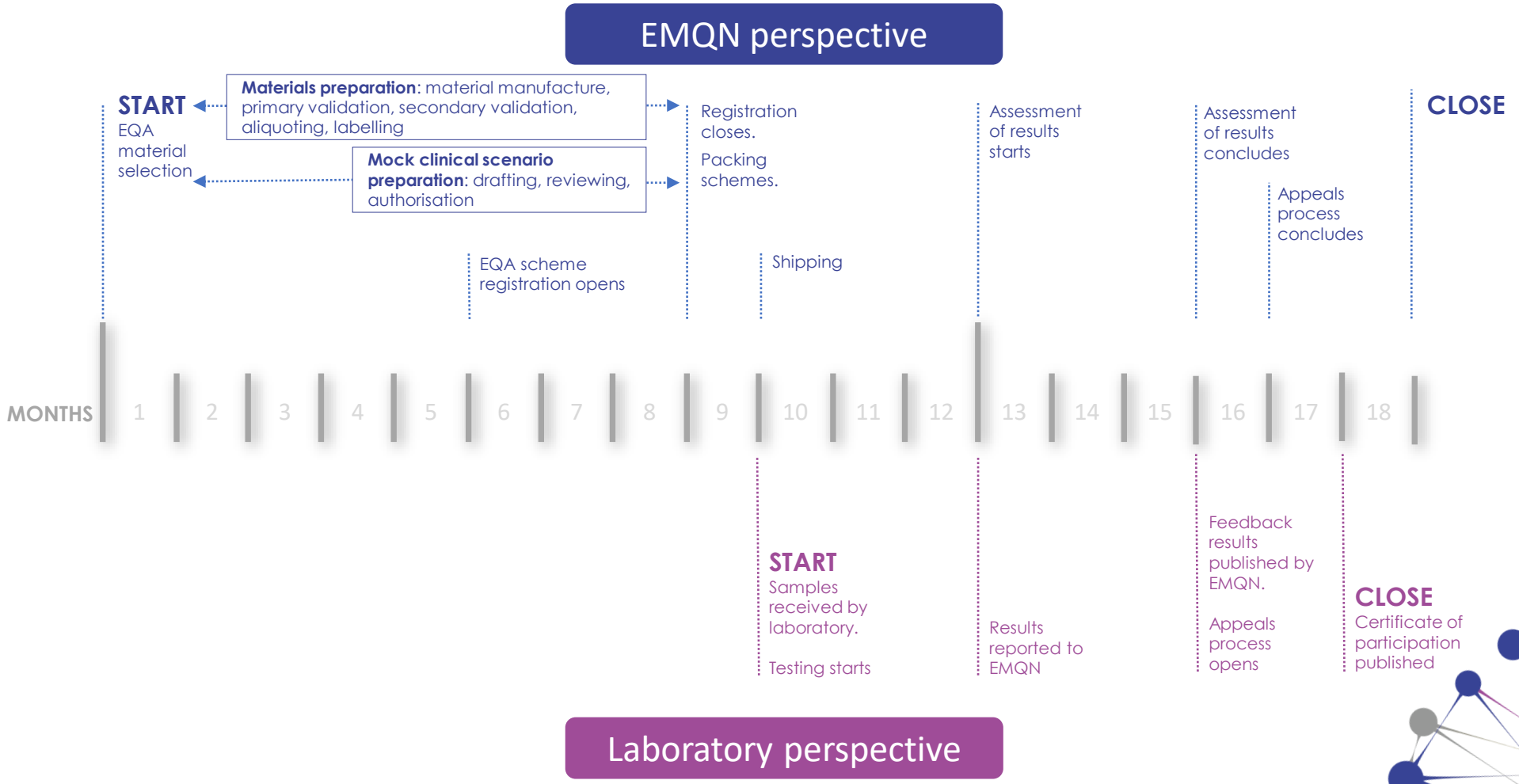
External Quality Assessment (EQA)

- **Main principles**
  - Participants are provided with the same material and return results to a coordinating centre
  - EQA samples are treated exactly like a patient sample
  - The EQA results are compared
  - EQA provides education and training for laboratories

- **Poor performance**
  - Regulators can suspend testing activity if a lab is performing poorly on a number of occasions

EMQN

The EQA lifecycle



NGS and EQA

NGS challenges for EQA

- NGS adoption widespread but no consensus on how to do diagnostic testing
  - Panels vs Exome vs Genome
- Rapidly changing technology
  - 2nd, 3rd generation etc
  - Short reads vs long reads
  - Quick transition from research to diagnostics is difficult
- How to select appropriate materials for somatic testing?
  - gDNA vs FFPE vs fresh frozen

EMQN

NGS challenges for EQA

- EQA scheme needs:
  - Platform, context (Genome, exome, panels), and laboratory setting (Genetics, Oncology etc.) agnostic
  - Test wet analytical process as well as bioinformatics
  - Fit with a laboratory's routine workflows
- How to make it cost effective for
  - For the participant labs
  - For the EQA provider

NGS scheme - bigger than one organisation

- Collaboration
  - EMQN
  - GenQA

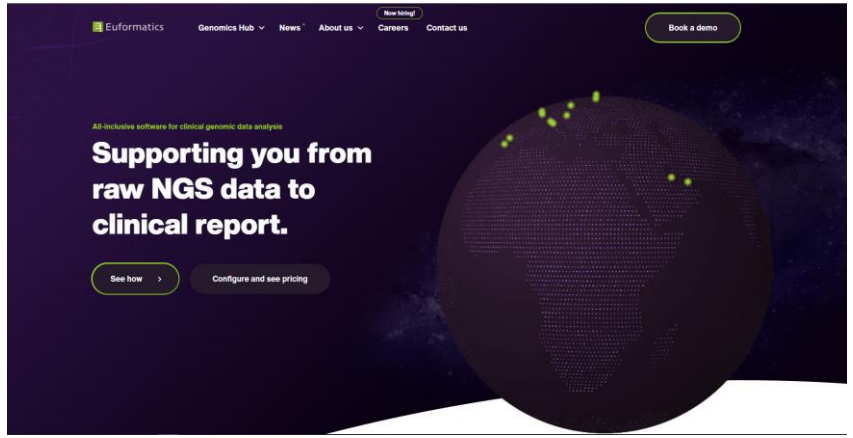
- Scientific oversight
  - Specialist Advisory Group (SAG)

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
Ms Becky Treacy	Scheme Organiser	GenQA
Dr Dave Cregeen	Scheme Organiser	GenQA
Prof Sandi Deans	Director	GenQA
Dr Simon Patton	Director	EMQN
Dr Weronika Gutowska-Ding	Scheme Organiser	EMQN

EMQN

Euformatics – data collection and analysis

- Dedicated online EQA platform accessed directly from the EMQN and GenQA websites
- Big data and Multiple results submissions (up to 3 per scheme type)
- Triage raw data (VCF, BAM, FASTQ, BED)
- Automate data analysis in real time
- Quality metrics and variant consensus analysis



http://euformatics.com

NGS schemes offered

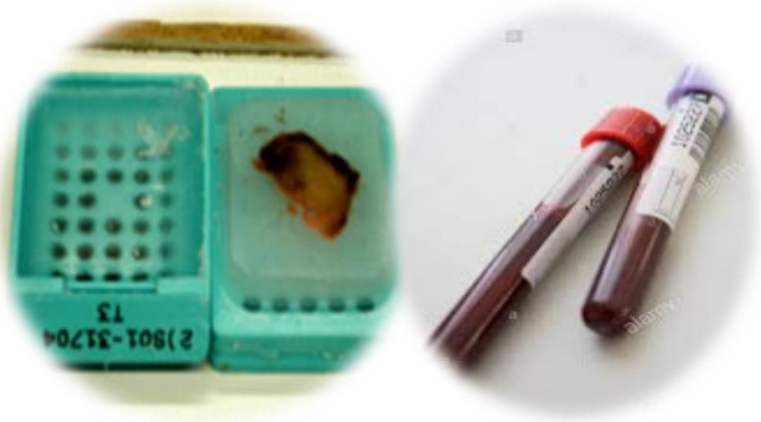
1. Germline

gDNA from cell line

2. Somatic★

gDNA from real fresh frozen/FFPE tumor tissue

3. Somatic- matched★

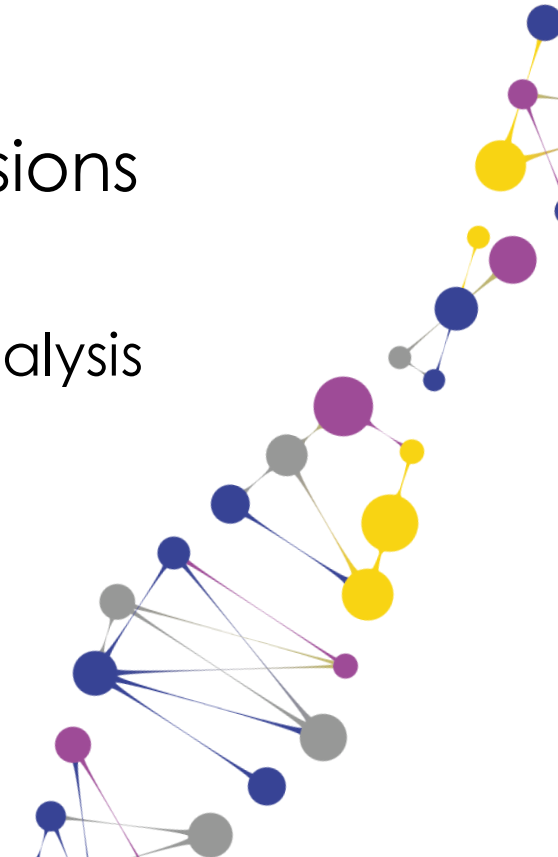


gDNA from real FFPE tumor tissue + matched normal tissue

★ Pilot EQAs – performance criteria do not apply

NGS germline EQA – current design

- The participants receive one genomic DNA sample
  - extracted from a single large homogenized growth of B lymphoblastoid cell lines
- Participants can submit up to three different submissions
  - e.g. gene panel + whole exome + whole genome
  - an optional 4th submission was allowed in 2022 for CNV analysis



EMQN

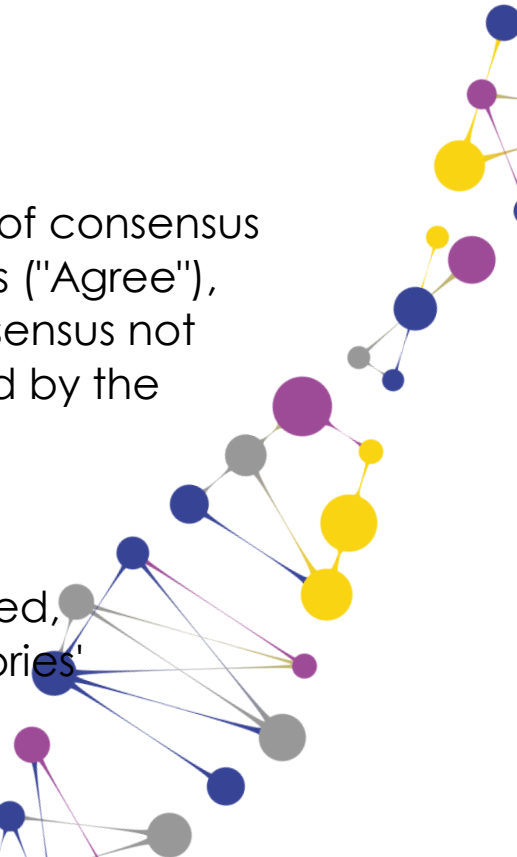
Reporting requirements for the participants

- Technical questionnaire
  - describing the sequencing approach, bioinformatics pipeline, and internally defined quality thresholds
- **VCF file**
  - mapped to GRCh38 or hg19/GRCh37
  - compiled after QC and region of interest (ROI) filtering but before any other interpretational steps e.g. assessing pathogenicity. A minimum of version 4 VCF files is required.
- **BED file**
  - containing the genomic co-ordinates of the ROI analysed. We required a minimum of: Chr, Start, End. Any overlapping regions are merged.
- **FASTQ file**
- **BAM file**

Feedback to participants

- **Aggregated**
  - EQA Summary report summarises all the results

- **Individual**
  - **Variant consensus analysis report**
    - contains a comparison of the variants reported by a laboratory against a list of consensus variants, resulting in classification into variants concordant with the consensus ("Agree"), variants not concordant with the consensus ("Disagree"), variants in the consensus not reported by the laboratory ("Missing" – false negatives), and variants reported by the laboratory which were not in the consensus ("Extra" – false positives)
  - **Data quality report**
    - contains selected quality metrics from the FASTQ, BAM, and VCF files submitted, benchmarked against the distribution of the same metric from other laboratories'



Data quality report

- DQ report shows quality metrics for FASTQ, BAM and VCF files, separately for each submission
- Participants are compared with others using the same platform/kit combination

Submission Germline-2

Lab ID	0056-E
Submission	2022 Germline-2
Platform used	NextSeq 500
Capture kit	TWIST Biosciences TWIST Comprehensive Exome
Reference genome	GRCH37

Number of submissions using the same sequencing platform: 179. Number of submissions using the same sequencing platform and the same capture kit: 24.

Variant call assessment

Region	all		high confidence	
Variant type	snp	indel	snp	indel
True positives	276	11	179	5
False positives	0	7	0	0
False negatives	17	2	7	1
Sensitivity	94.20%	84.62%	96.24%	83.33%
Precision	100.00%	61.11%	100.00%	100.00%
F-Score	97.01%	70.97%	**98.08%**	90.91%

These numbers provide only the summary of the variant assessment. For the detailed results, please see the accompanying file containing the detailed variant-level report.

Quality metrics assessment

FASTQ

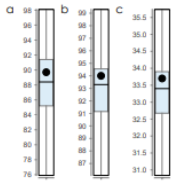


Figure 4. Data obtained from FASTQ file, benchmarking against submissions using the same sequencing platform (N=179): a) **Bases above Q30 (%)**; b) **Bases above Q20 (%)**; c) **Base quality (average)**;

Metric	Value	Explanation
Bases above Q30	89.7%	Bases with Phred score over 30, i.e. 99.9% accuracy
Bases above Q20	94%	Bases with Phred score over 20, i.e. 99% accuracy
Base quality (average)	33.7	Average base quality (Phred scale)

BAM

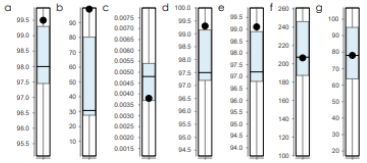


Figure 5. Data obtained from BAM file, benchmarking against submissions using the same sequencing platform and the same capture kit (N=24): a) **Uniformity (%)**; b) **Off target (%)**; c) **Error rate on target**; d) **Coverage at 20X (%)**; e) **Coverage at 30X (%)** ; f) **Insert size (average)**; g) **Insert size (standard deviation)**.

Metric	Value	Explanation
Uniformity	99.5%	Bases on target at >10% of median coverage
Off target	99.1%	Reads mapped outside the target region
Error rate on target	0.0038	Mismatches on target per total bases on target
Coverage at 20X	99.3%	Percentage of bases with read depth 20 on the target
Coverage at 30X	99.1%	Percentage of bases with read depth 30 on the target
Insert size (average)	206.4	Average insert size on target
Insert size (standard deviation)	78	Standard deviation of insert size on target

VCF

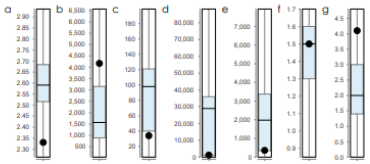


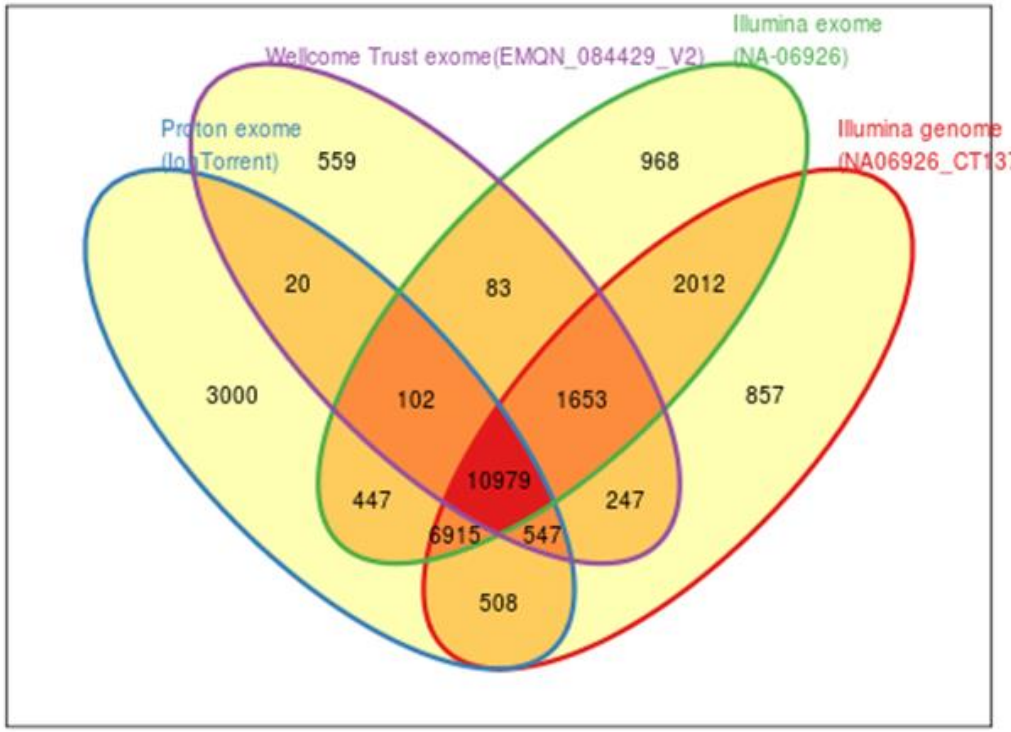
Figure 6. Data obtained from VCF file, benchmarking against submissions using the same sequencing platform and the same capture kit (N=24): a) **Ti/Tv**; b) **Quality of calls (average)**; c) **Depth at calls (median)**; d) **SNPs**; e) **Indels**; f) **Het:hom ratio (SNP)**; g) **Het:hom ratio (Indel)**.

Metric	Value	Explanation
Ti/Tv	2.331	The ratio of transitions vs. transversions in SNPs
Quality of calls (average)	4171	Average quality of SNPs
Depth at calls (median)	34	Median read depth at called variant
SNPs	1219	Number of SNPs
Indels	363	Number of indels
Het:hom ratio (SNP)	1.5	Heterozygous / homozygous SNPs
Het:hom ratio (Indel)	4.1	Heterozygous / homozygous Indels

Developing a consensus genome

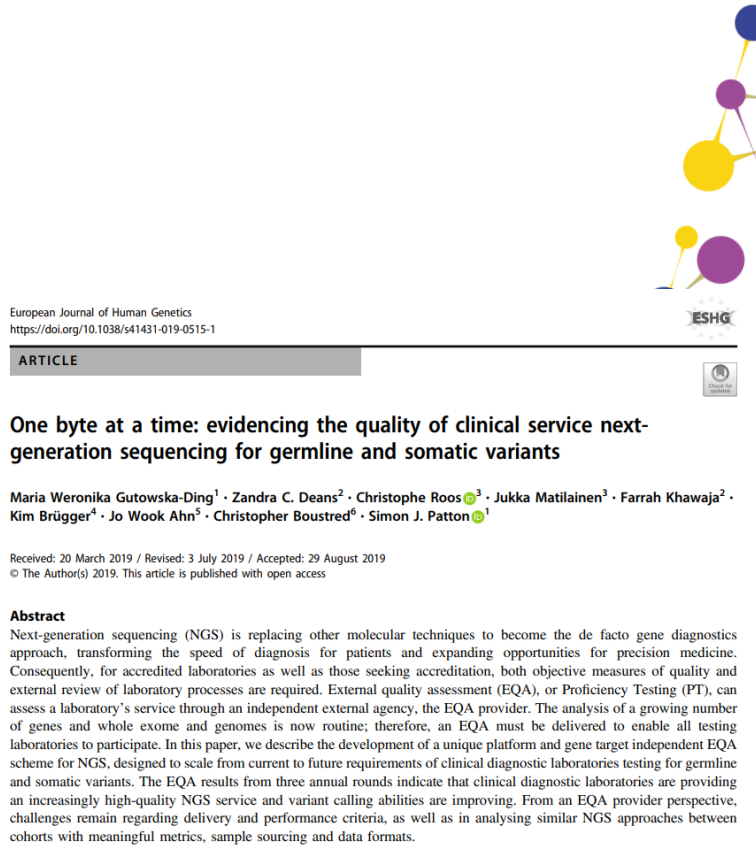
Developing a consensus genome

- **Old approach**
  - DNA samples used in the 2014-2016 schemes underwent validation by four independent testing centres using whole genome and whole exome sequencing on both Illumina and Life Technologies platforms.
  - A "consensus EQA genome" was established comprising variants detected by at least two out of four validation labs. For the purpose of this analysis we have taken into account only the coding regions +/-2bp.



Developing a consensus genome

- **New approach**
  - From 2017 a "participant consensus genome" is established separately for both GRCh37 and GRCh38 submissions. The consensus required at least seven submissions to cover each variant position with at least 75% agreeing on the genotype.
  - In 2017-2019, the participants' consensus genome was also compared to the publicly available data from the Personal Genome Project website.
  - Material was selected because they were one of the genomes used by the Genome in a Bottle Consortium★ (GIAB) to produce and characterise as reference materials
  - there is an extensive publicly available library of reference sequence data related to this material which has allowed us to fine tune and validate our consensus building algorithms – which we published in our 2019 paper in EJHG
  - Since 2020 we solely rely on the participants' consensus genome



European Journal of Human Genetics
https://doi.org/10.1038/s41431-019-0515-1

ARTICLE

One byte at a time: evidencing the quality of clinical service next-generation sequencing for germline and somatic variants

Maria Weronika Gutowska-Ding[1] · Zandra C. Deans[2] · Christophe Roos[3] · Jukka Matilainen[3] · Farrah Khawaja[2] · Kim Brügger[4] · Jo Wook Ahn[5] · Christopher Boustred[6] · Simon J. Patton[1]

Received: 20 March 2019 / Revised: 3 July 2019 / Accepted: 29 August 2019

Abstract
Next-generation sequencing (NGS) is replacing other molecular techniques to become the de facto gene diagnostics approach, transforming the speed of diagnosis for patients and expanding opportunities for precision medicine. Consequently, for accredited laboratories as well as those seeking accreditation, both objective measures of quality and external review of laboratory processes are required. External quality assessment (EQA), or Proficiency Testing (PT), can assess a laboratory's service through an independent external agency, the EQA provider. The analysis of a growing number of genes and whole exome and genomes is now routine; therefore, an EQA must be delivered to enable all testing laboratories to participate. In this paper, we describe the development of a unique platform and gene target independent EQA scheme for NGS, designed to scale from current to future requirements of clinical diagnostic laboratories testing for germline and somatic variants. The EQA results from three annual rounds indicate that clinical diagnostic laboratories are providing an increasingly high-quality NGS service and variant calling abilities are improving. From an EQA provider perspective, challenges remain regarding delivery and performance criteria, as well as in analysing similar NGS approaches between cohorts with meaningful metrics, sample sourcing and data formats.

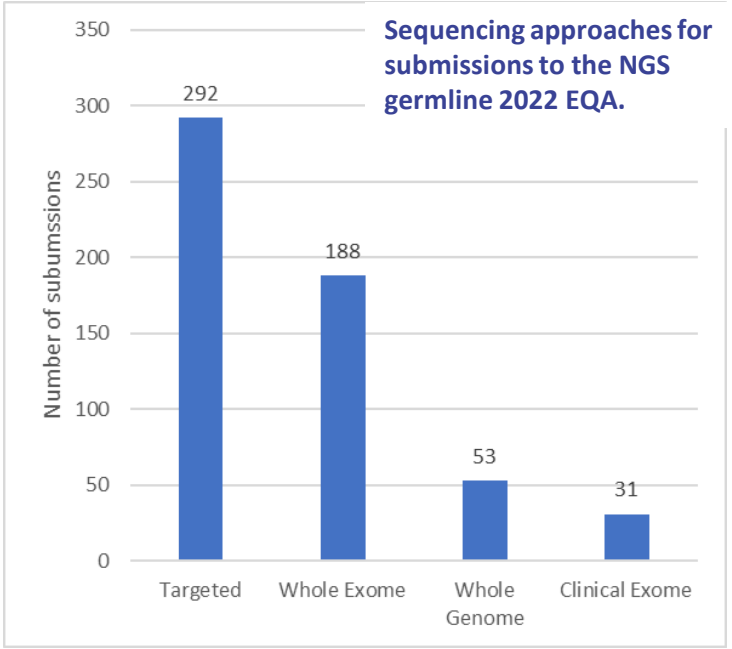
★ https://www.nist.gov/programs-projects/genome-bottle

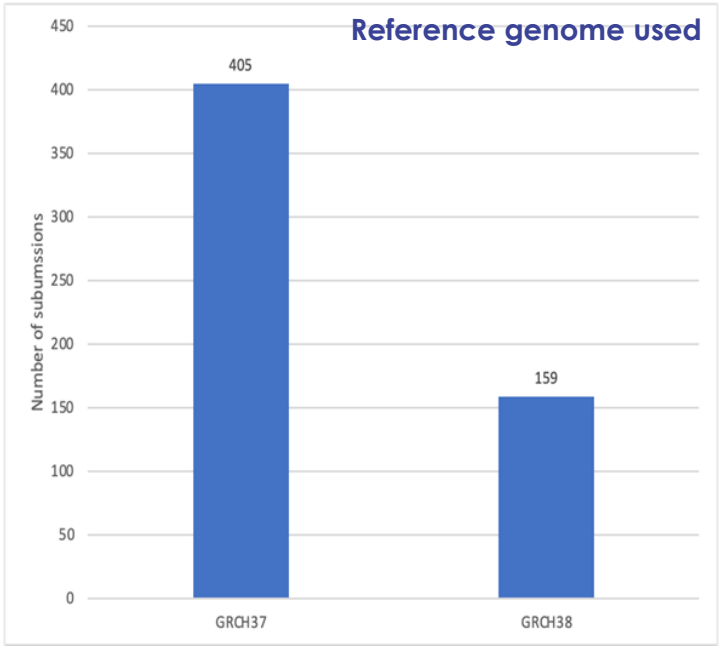
Participants' testing approach

Results

Testing approach

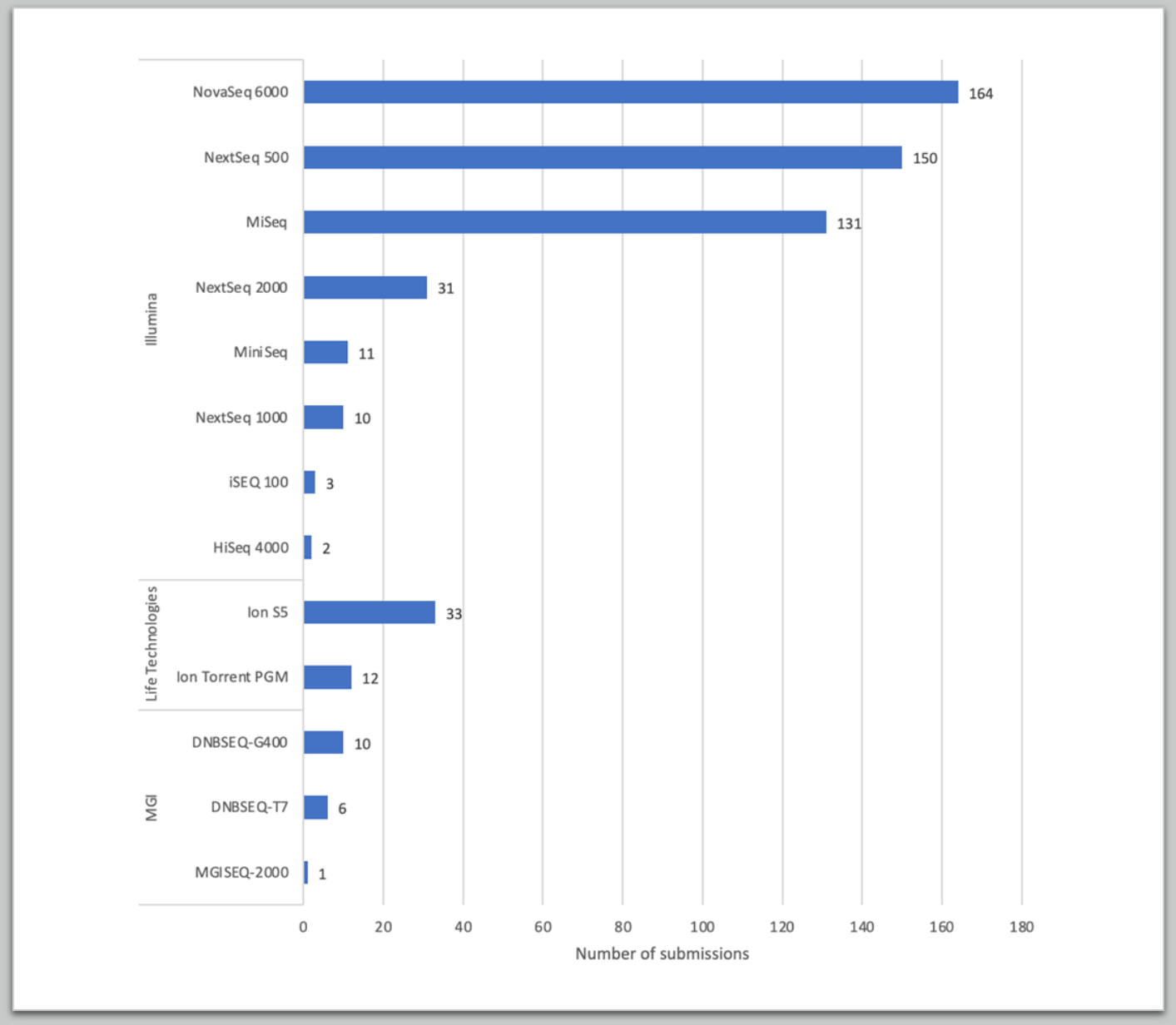
- Participants are allowed to submit up to three sets of results, including whole exome/genome sequencing
- The majority of participants submitted results of targeted panel testing
- They use GRCh37 reference genome
- In all EQA runs *BRCA1* and *BRCA2* were genes tested by the majority of labs





Platform used

- The majority of participants (89%) use one of the Illumina NGS platforms
- Year on year, fewer labs use Roche, Ion Proton and Ion Torrent

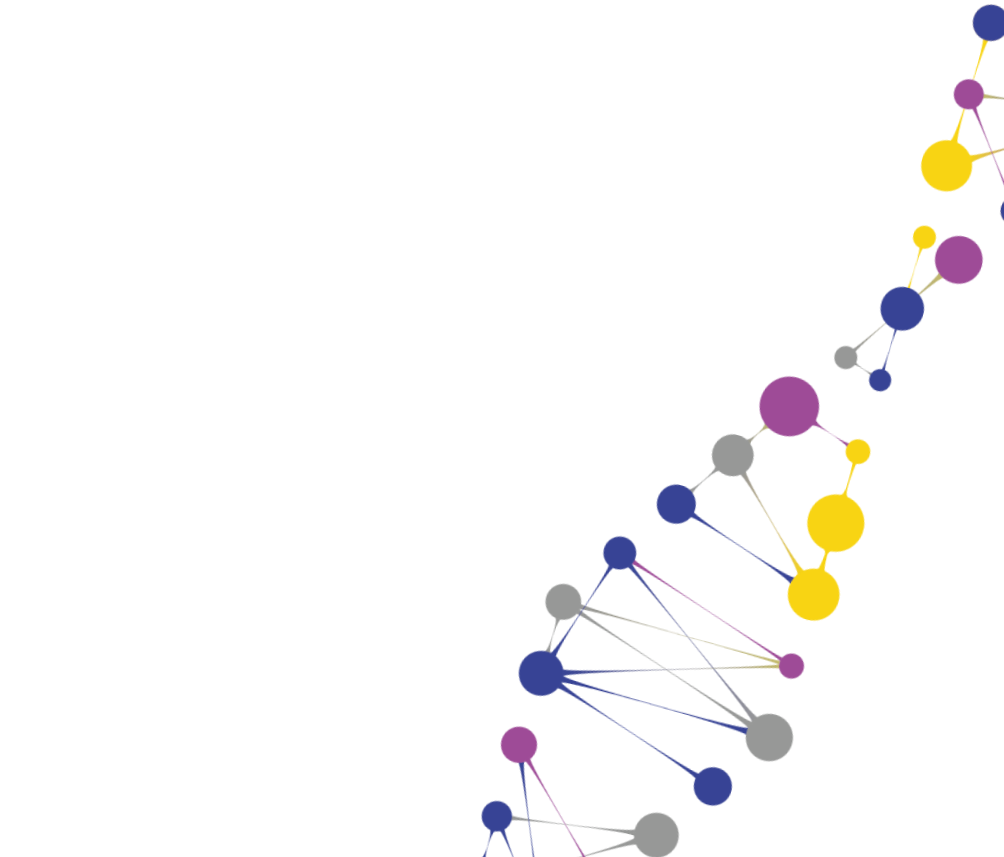


Bioinformatics analysis

- Just over 42% of participants used in-house bioinformatics, with 4.1% outsourcing the analysis
- 23.8% used the platform provided and 29.2% used commercial pipelines (both increase year on year)
- 85% of laboratories use Burrows-Wheeler Aligners (BWA)
- Just over 76% of participants reported using one of the GATK variant callers

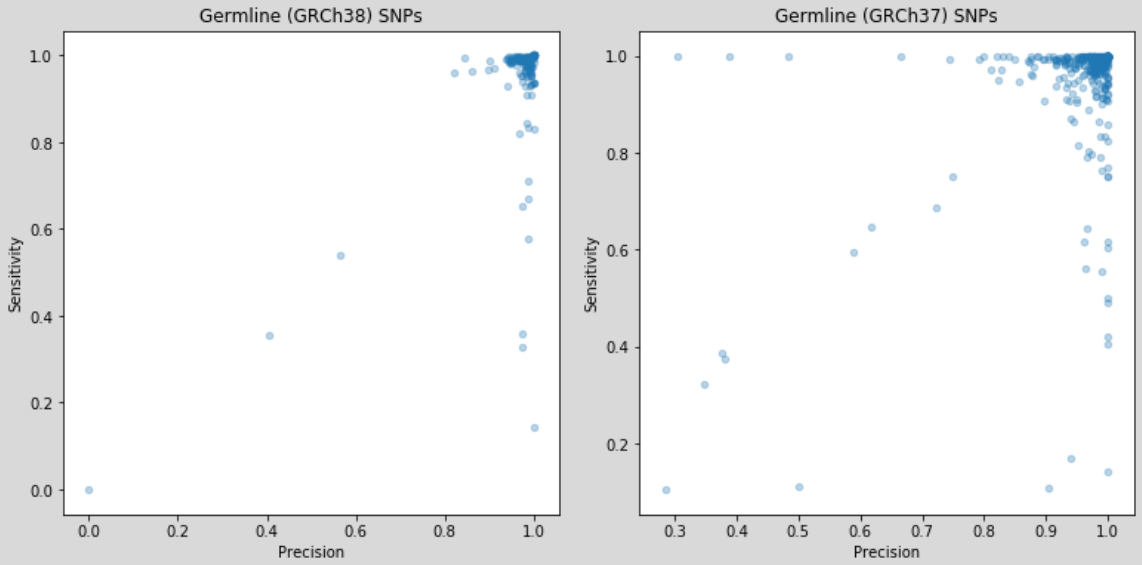
Variant Calling

Results



Variants called – sensitivity and precision of SNPs

- Following the GA4GH's recommendations we have stratified the results by SNPs and Indels
- Generally good for SNPs
- Better in laboratories using GRCh38
- Most prominent outliers with low sensitivity failed to submit a list of gene exclusions making it impossible to narrow analysis to a specific Region of Interest



Variants called – sensitivity and precision of Indels

- calling Indels is much harder; the overall quality of the Indel results is much lower
- Note: if the participants are generally unreliable in calling the indels, then the participant consensus is unreliable too
- participant consensus is missing many true Indels
- we could find out more if we compared the Indel variant concordance results to the average read depth

EMQN

Individual Variant consensus analysis report

- Reported variants were normalised and compared against the consensus variants (EQA genotype)
  - **F-score** was calculated: Harmonic mean of sensitivity and precision. Defined as F-Score = 2 × Sensitivity × Precision / (Sensitivity + Precision)

	A	B	C	D	E	F	G	H	I
1	Analysis set:	2022 Germline (GRCh37)							
2	Reference genome:	GRCh37							
3	Notes:	This submission was excluded from the SNP participant consensus as a submission not submitted for evaluation							
4		This submission was excluded from the Indel participant consensus as a submission not submitted for evaluation							
5									
6									
7	Region:	all		high confidence					
8	Type:	snp	indel	snp	indel				
9	True positives:	30678	1968	20108	667				
10	False positives:	303	526	14	13				
11	False negatives:	654	56	18	0				
12	Sensitivity:	97.91%	97.23%	99.91%	100.00%				
13	Precision:	99.02%	78.91%	99.93%	98.09%				
14	F-Score:	98.46%	87.12%	**99.92%**	99.03%				
15									
16	Variant position	Type	Gene	Submitted genotype	EQA genotype	EQA consensus ratio	Classification	Notes	Region
17	1:69270	snp	OR4F5		G/G	52/99	Not Assessed	Uncertain consensus	low confidence
18	1:69511	snp	OR4F5	G/G	G/G	91/99	Agree		low confidence
19	1:69897	snp	OR4F5		T/C	33/95	Not Assessed	Uncertain consensus	low confidence
20	1:536895	snp	RP5-857K21.4		T/C	26/38	Missing		low confidence
21	1:536961	snp	RP5-857K21.4		T/C	16/22	Missing		low confidence
22	1:877831	snp	SAMD11	C/C	C/C	101/102	Agree		low confidence
23	1:879317	snp	NOC2L	C/T	C/T	104/106	Agree		low confidence
24	1:881627	snp	NOC2L	G/A	G/A	100/101	Agree		high confidence

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)
- **Disagree** – participant's variant did not match the participants' consensus (False positive, false negative)
- **Not Assessed –** participant's variant could not be assessed against the consensus (e.g. uncertain consensus)

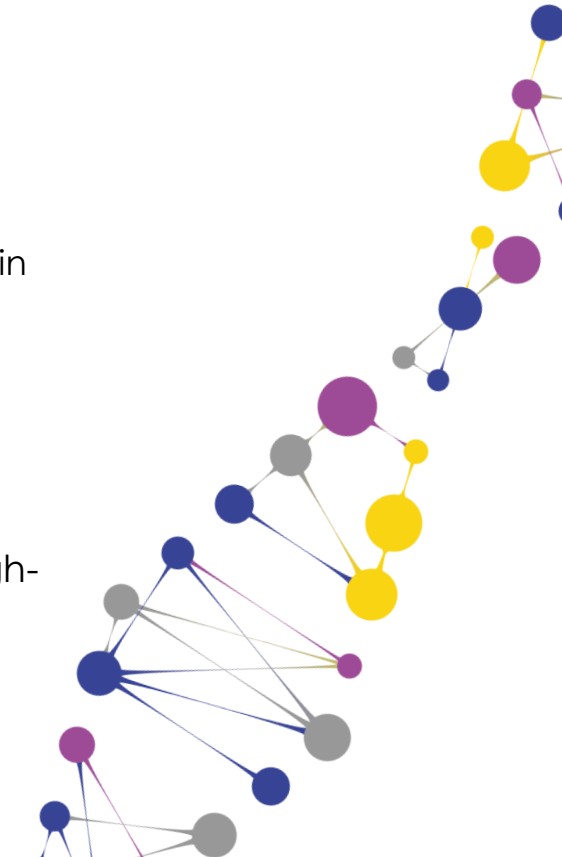
Performance criteria – NGS germline (2021 onwards)

- The marking system includes:
  - NGS variant concordance using the **F-score** for SNPs only, located within the high-confidence (HC) regions of the genome[1].
  - The performance outcome for this EQA is Satisfactory OR Poor.

[1] **Note**:

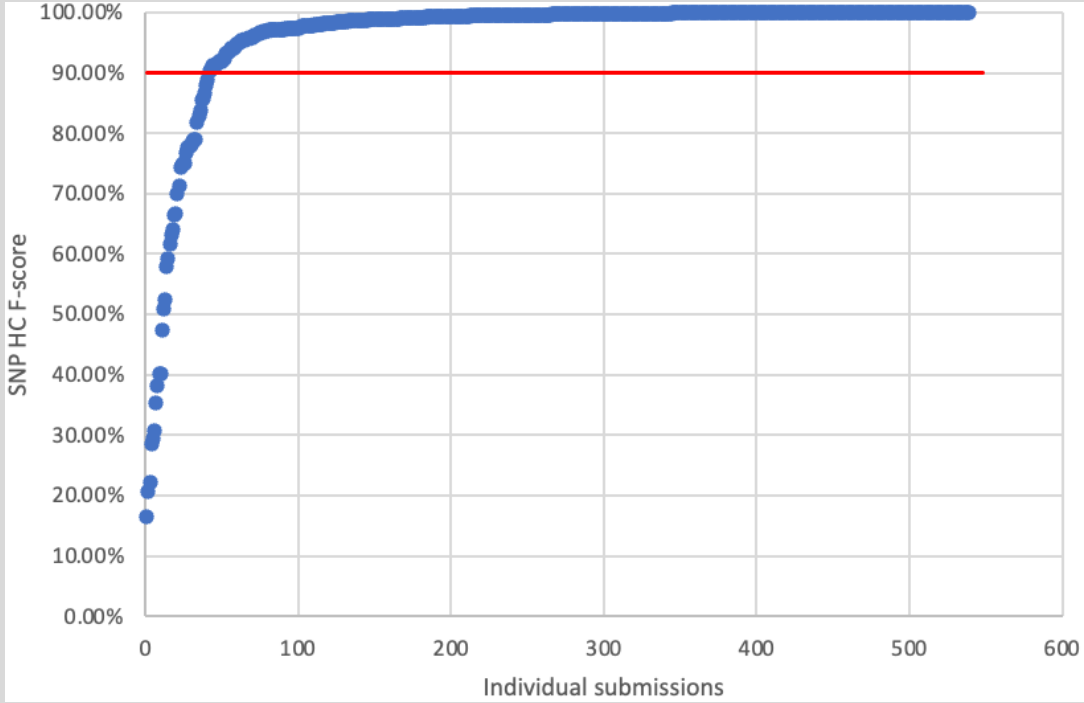
- 'High confidence' is defined as genomic regions exclusive of union of all tandem repeats, all homopolymers >6bp, all imperfect homopolymers >10bp, all difficult to map regions, all segmental duplications, GC <25% or >65%, "Bad Promoters", and "Other Difficult Regions" as published by NIST in Genome In A Bottle - Genome Stratifications (https://doi.org/10.18434/M32190).
- The F-score of indels (<50bp) is excluded from the current Performance Criteria

- Poor performance is defined as:
  - Those participants having **any submission with an F-score below 90%** for SNPs within the high-confidence regions of the genome



2022 EQA performance

- 346 laboratories submitted 564 different datasets
  - 65.96% of all submissions achieved the F-score ≥ 99% (65.95% in 2021)
  - 6.57% scored ≤ F-score 90%; these laboratories were given "Poor Performance" (PP) (decrease from 10% in 2021)



Common issues

- Overall quality of raw data is very good, but…
- Wrong reference genome assembly indicated in the survey
- Invalid characters in the VCF INFO field
- Genotype calls in VCF referring to symbolic (non-specific) alleles
- Incorrect alternate alleles referenced in the VCF genotype field

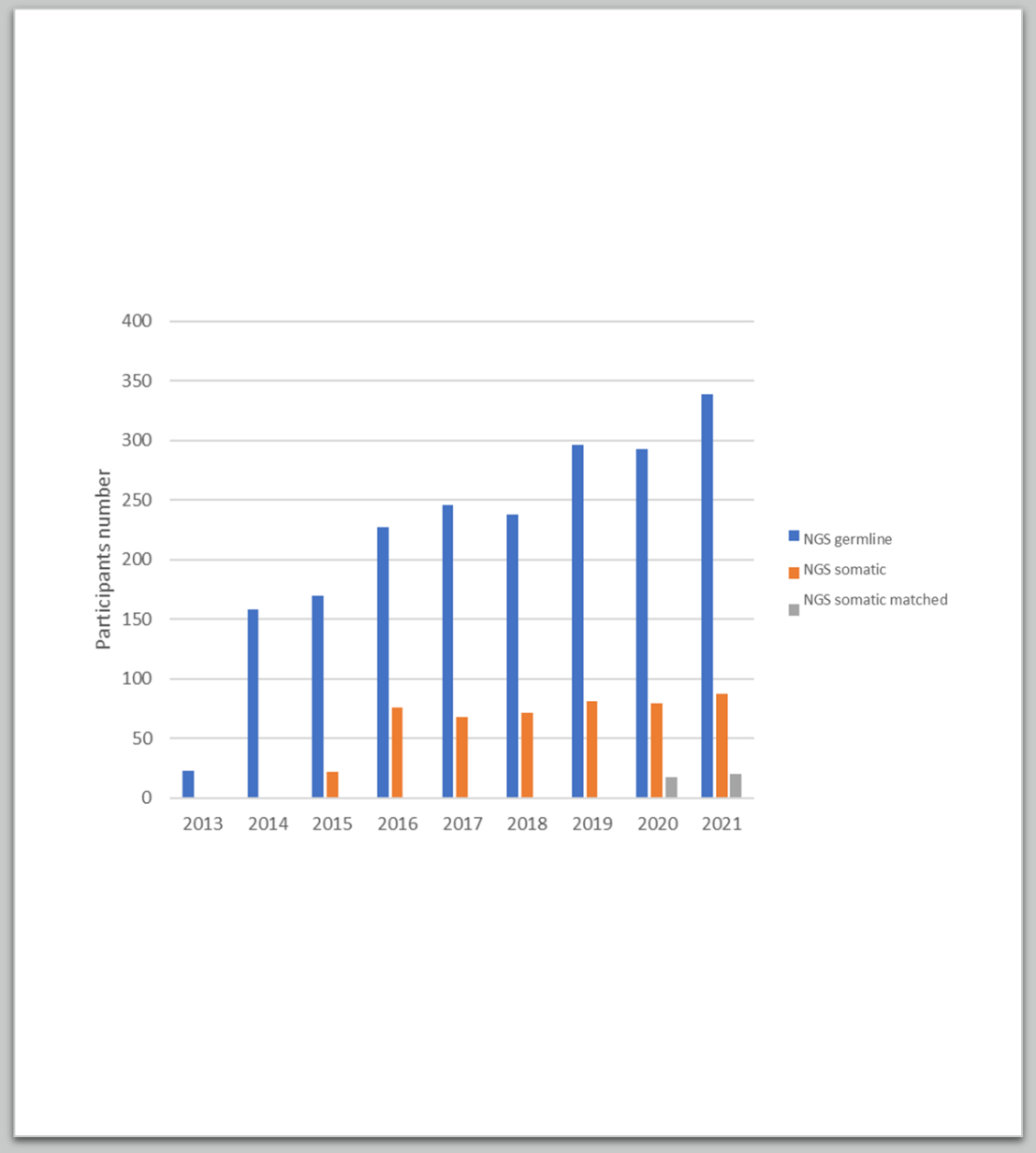
EMQN

Conclusions

Conclusions

- **NGS EQA Achievements**
  - NGS scheme is generic
    - Technology and testing context independent
    - Seamless fit with laboratory processes
  - Analysis of both wet lab and bioinformatics processes
  - Number of participants grows year on year as NGS is becoming more widespread technique in medical diagnostics

- **NSG EQA Challenges**
  - Consensus for indel calling
  - Performance criteria for the NGS somatic EQAs
  - Materials for NGS somatic scheme
  - Getting uniformed results from the participants - data formats not standardised
  - Current analysis skewed towards Illumina platforms and short read sequencing



Final remarks - why participate in EQAs

- EQA provides:
  - an external measure of the quality of YOUR laboratory service
  - may highlight problems with kits and methodology, especially for LDTs
  - may prevent future errors in YOUR diagnostic service
  - provides YOU with continuous education and training
  - performance monitoring, especially for compliance with Laboratory Accreditation standards (ISO15189, 17025)
- Regular EQA participation improves the quality of YOUR testing

EMQN

Acknowledgements

EMQN Team

- Simon Patton, Rebecca Goodall, Michael Woodcock

Euformatics Team

- Jukka Matilainen, Christophe Roos, Rafael Popin

GenQA Team

- Sandi Deans, Becky Treacy, Dave Creegan

Scientific Advisory Group members

Would YOU like to join us?