Beyond One Million Genomes

D3.2

Best practices for Next Generation Sequencing

Project Title (grant agreement No)	Beyond One Million Genomes (B1MG) Grant Agreement 951724		
Project Acronym	B1MG		
WP No & Title	WP3 - Standards & Quality Guidelines		
WP Leaders	Ivo Gut (CRG), Jeroen Belien (VUmc)		
Deliverable Lead Beneficiary	3 - CRG		
Deliverable	D3.2 - Best practices for Next Generation Sequencing		
Contractual delivery date	31/05/2022	**Actual delivery date**	09/05/2023
Delayed	Yes		
Authors	Ivo Gut (CRG), Gabriela Aguileta (CRG)		
Contributors	Edwin Cuppen (HMF), Valtteri Wirta (KI), Eivind Hovig (UiO), Gert Matthijs (KU Leuven)		
Acknowledgements (not grant participants)	Lucía Estellés		
Deliverable type	Report		
Dissemination level	Public		

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Document History

Date	Mvm	Who	Description
21/04/2023	0v1	Ivo Gut (CRG)	Initial draft circulated to WP participants for feedback
21/04/2023	0v2	Nikki Coutts (ELIXIR Hub)	Version circulated to B1MG-OG, B1MG-GB and 1+MG Stakeholders for feedback
09/05/2023	1v0	Nikki Coutts (ELIXIR Hub)	Comments addressed. Version uploaded to the EC Portal

Table of Contents

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. Executive Summary

Determining variants in the genome involves several bioinformatic procedures, such as eliminating low-quality sequences, aligning sequencing reads to the human reference genome, and establishing confidence in the presence of a variant based on a threshold. Once a variant is identified, it is annotated to predict its effect. The goal of this task is to establish a best practice protocol for data analysis of whole genome sequencing (WGS) for somatic variants. This protocol will include a recommended suite of software tools with settings that ensure results surpass a required quality threshold.

As part of the 1+MG WG4 project, we are currently benchmarking quality metrics to assess the best standards in the practice of WGS for somatic variants. At the time of writing, we have completed the initial steps of the benchmark and evaluated the quality metrics corresponding to the library preparation and sequencing steps. We compared the performance of sequencing conducted by participating laboratories by examining the value and dispersion of relevant metrics, providing a first assessment of how different sequencing protocols impact the sequences produced. The sequencing laboratories got an overall performance score, and based on the protocols followed by the facilities producing the best results we suggest which are the best practices for library preparation and sequencing.

A general conclusion is that all participants have achieved a good level of quality at the sequencing stage, and the metrics measuring it are largely consistent, as there is very little dispersion. Differences in library preparation and sequencing protocols do not appear to significantly impact the expected quality of results. In the following sections, we explain the performance of different participating laboratories for each one of the relevant sequencing metrics, and suggest general best practices to ensure the best quality.

This deliverable has been significantly delayed due to an important backlog in the preparation, examination, and approval of all ethical requirements to protect sensitive patient data. We made considerable effort in drafting the material transfer agreement (MTA) to guarantee that patient data will be used with all precautions and strict anonymity because genomic data is identifiable.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

2. Contribution towards project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives/key results:

	Key Result No and description	Contributed
Objective 1 Engage local, regional, national and European stakeholders to define the requirements for cross-border access to genomics and personalised medicine data	**1.** B1MG assembles key local, national, European and global actors in the field of Personalised Medicine within a B1MG Stakeholder Coordination Group (WP1) by M6.	Yes
	2. B1MG drives broad engagement around European access to personalised medicine data via the B1MG Stakeholder Coordination Portal (WP1) following the B1MG Communication Strategy (WP6) by M12.	No
	3. B1MG establishes awareness and dialogue with a broad set of societal actors via a continuously monitored and refined communications strategy (WP1, WP6) by M12, M18, M24 & M30.	No
	4. The open B1MG Summit (M18) engages and ensures that the views of all relevant stakeholders are captured in B1MG requirements and guidelines (WP1, WP6).	No
Objective 2 Translate requirements for data quality, standards, technical infrastructure, and ELSI into technical specifications and implementation guidelines that captures European best practice	**Legal & Ethical Key Results**	
	1. Establish relevant best practice in ethics of cross-border access to genome and phenotypic data (WP2) by M36	Yes
	2. Analysis of legal framework and development of common minimum standard (WP2) by M36.	No
	3. Cross-border Data Access and Use Governance Toolkit Framework (WP2) by M36.	No
	Technical Key Results	
	4. Quality metrics for sequencing (WP3) by M12.	Yes
	5. Best practices for Next Generation Sequencing (WP3) by M24.	Yes
	6. Phenotypic and clinical metadata framework (WP3) by M12, M24 & M36.	No
	7. Best practices in sharing and linking phenotypic and genetic data (WP3) by M12 & M24.	No
	8. Data analysis challenge (WP3) by M36.	Yes
	Infrastructure Key Results	
	9. Secure cross-border data access roadmap (WP4) by M12 & M36.	No
	10. Secure cross-border data access demonstrator (WP4) by M24.	No

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Objective 3 Drive adoption and support long-term operation by organisations at local, regional, national and European level by providing guidance on phased development (via the B1MG maturity level model), and a methodology for economic evaluation	**1.** The B1MG maturity level model (WP5) by M24.	No
	2. Roadmap and guidance tools for countries for effective implementation of Personalised Medicine (WP5) by M36.	No
	3. Economic evaluation models for Personalised Medicine and case studies (WP5) by M30.	No
	4. Guidance principles for national mirror groups and cross-border Personalised Medicine governance (WP6) by M30.	Yes
	5. Long-term sustainability design and funding routes for cross-border Personalised Medicine delivery (WP6) by M34.	No

3. Methods

3.1 Somatic Benchmarking Scheme Design

The CNAG organises and takes part in the WGS benchmark for somatic variants. Participant laboratories come from different European countries, and for this initial evaluation there are a total of 7 participant institutes. The overall organisation of the benchmark (Fig. A1) was as follows: 1) The organiser obtained the samples from the Medical University of Graz (MUG), as part of the EASI Genomics initiative. CNAG did the DNA extraction from frozen tissue; 2) A Material Transfer Agreement (MTA) following all the applicable legal and ethical requirements was signed between each participant and the Organiser; 3) CNAG distributed and sent the material (described below) to each participant laboratory; 4) Participants prepared the libraries and sequenced the samples according to their own SOPs; 5) Somatic variant calling was done by participants following their own standard pipelines; 6) Participants sent their results back to the organiser, including a SAV (output file with metrics that is produced by the sequencing instrument), a FASTQ file and the result of variant calling of both small and large variants in the VCF format; 7) CNAG is currently building a gold set using CNAG's and all the participants' results, as well as Nanopore sequencing of the cancer samples to validate structural variants; 8) CNAG will use the gold set to benchmark the results from each participant and will produce reports assessing the performance of all participant laboratories. Figure 1 in the Appendix shows the scheme of the somatic benchmarking.

3.2 Test Items

Participants received DNA from 8 Tumour/Normal pairs extracted from frozen tissue: 2 head and neck squamous cell carcinoma samples, 3 lung squamous cell carcinoma samples, and 3 clear cell renal carcinoma samples, as well as their matching normal samples . Table 1 in the Appendix contains the list and description of each of the Tumour/Normal pairs.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

3.3 Quality Metrics Evaluation

Together with participants, a set of Quality Metrics was defined to evaluate the performance of each participant laboratory for library preparation, sequencing and data analysis (Deliverable D3.1[1]).The assigned value and acceptable range of results were established, according to the guidelines provided in ISO13528 (2015). According to the definition of the guidelines, an assigned value is an estimate of the value of the measure and that is used for calculating scores. Table 2 in the Appendix shows the QC metrics used to compare across laboratories. The chosen QC metrics provide a measure of the quality of the performance of the laboratory at each stage of the benchmark. The organiser will assess QC metrics of sequencing, sample preparation, post alignment, and variant calling of all the participants in order to compare the performance across laboratories.

3.4 Gold Set Generation

A gold set is a curated set of variants that we are highly confident that are true. To obtain a gold set for the somatic benchmark, the FASTQ data will be merged by concatenating the FASTQ files submitted by participants. Collaborators will run their different SOP pipelines to call small and large somatic variants. They will submit the data with an agreed header format to minimise the effort to work with such files. Nanopore data will be used to confirm the large variants only.

3.5 Benchmark variant voter

The variants that are called by all participants will be part of the gold set and in the case of discrepant variants, CNAG has developed a variant voter tool, which is an R shiny app available online to all participants through Shinyapps.io, where they will be presented in genomic viewer screenshots representing discrepant variants and voters will be able to cast their votes more efficiently. The manual curation of difficult variants will help in the construction of the gold set.

3.5 Final reports

ISO 17043 defines the requirements of the participants' reports. We have decided to make two reports: a general report and a participants' report. We will deliver the same general report to all participants. We will deliver a personalised report to each participant separately.

4. Description of work accomplished

4.1 1+MG WP4 Somatic Benchmark

At this stage, CNAG has compared library preparation and sequencing metrics across laboratories to assess the performance of each laboratory, relative to other laboratories and with respect to assigned values. We followed the ISO 13528 (2015) and ISO/IEC 17043 as guidelines for the statistical analysis of QC metrics significance.

[1] https://zenodo.org/record/5018495#.ZFoIuuzMJjc

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

For most of the assessed QC metrics, we do not have a gold standard, therefore we derived the assigned values and uncertainties from the participant consensus. Assigned values corresponded to the mean and uncertainties corresponded to the standard deviation. Extreme outliers were excluded and indicated in the plots where applicable.

4.1.1 Library preparation and sequencing QC metrics

The following are the metrics we have specified to assess the sequencing quality of the participant centres.

1. % bases with Q >= 30 R1 and R2: the percentage of bases where the quality score is equal or higher than 30.
2. PhiX is a control library that is often used in sequencing experiments to monitor sequencing quality and to detect any issues in the sequencing run. The % PhiX alignment R1/R2 is a quality metric that can provide information about the quality of the sequencing run, as a low alignment percentage may indicate issues such as contamination or low sequencing quality. On the other hand, the error rate of the PhiX library reflects the overall error rate of the sequencing run, as the same sequencing chemistry and conditions are used to sequence both the PhiX library and the samples being analysed. A high PhiX error rate may indicate issues such as low quality sequencing reagents, poor instrument calibration, or other issues that may affect the overall quality of the sequencing results. There are two main parameters related to PhiX:
    a. % PhiX alignment R1 and R2: the percentage of reads that align to the PhiX genome in the first read (R1) or the second read (R2) of a paired-end sequencing experiment.
    b. PhiX error rate R1 and R2: The calculated error rate in R1 or R2, as determined by a spiked in PhiX control sample.
3. % passing filter (PF) clusters: The percentage of clusters passing filters after evaluating cycle 1 to 25 as specified by the manufacturer.
4. % phasing R1 and R2: The average percentage of molecules in a cluster per cycle for which sequencing falls behind the current cycle during the read in R1 or R2.
5. % prephasing R1 and R2:The average percentage of molecules in a cluster per cycle for which sequencing jumps ahead the current cycle during the read in R1.

The prephasing percentage refers to the percentage of sequencing reads that are prematurely called during the sequencing process due to the incorporation of incorrect nucleotides. This can occur when the next nucleotide in the sequence is not accurately determined by the sequencing chemistry, leading to an incorrect base being called. High rates of both phasing and prephasing metrics may indicate a defective incorporation of nucleotides. Prephasing can lead to errors in the sequencing data, particularly in regions of the genome with high sequence similarity or repetitive elements.

4.1.2 Participant Laboratory Performance Evaluation

In order to evaluate the quality metrics of a test item, each metric will have an assigned value and uncertainty. An acceptance criteria shall be established based on the specific thresholds for each metric. The outcome of the evaluation will then be decided based on whether the test

item's metric values meet or exceed the established acceptance criteria. By defining the acceptance criteria for each metric, we can ensure that the results of the evaluation are consistent, reliable, and meet the desired quality standards. Overall, establishing an acceptance criteria based on the thresholds for each metric is a critical step in accurately evaluating the quality of a test item. Figure 1 shows how the assigned and indicative values are established and applied to determine whether results are acceptable, questionable or unacceptable.

Figure 1: Example of QC metric evaluation of results with respect to assigned and indicative values

Calculation of thresholds based on the assigned value

- Z-Scores measure how much a Participant deviates from the assigned value given the uncertainty.

Assigned value (mean) x_a — Participant's value x

$$Z = \frac{x_a - x}{\sigma_a}$$

Uncertainty (std deviation) σ_a

- Three outcomes are possible depending on the (absolute) value of the Z-Score.

| Z-score |	Outcome
Below 2	Acceptable
From 2 to 3	Questionable
Above 3	Unacceptable

- Based on the Z-score calculations and the limits for the outcomes, we can establish numeric values for the thresholds that apply.

Thresholds that apply	Acceptable	Questionable	Unacceptable
Upper	(-Inf,]	(,]	(, Inf)
Lower	(, Inf)	(,]	(-Inf,]
Both	(,]	(,] u (,]	(-Inf,] u (, Inf)

Figure 2: In the figure above, a positive z-score indicates that the observation is above the mean, while a negative z-score indicates that the observation is below the mean. A z-score of 0 indicates that the observation is equal to the mean

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

4.1.3 Sequencing and Bioinformatics Pipelines

Sequencing metrics were extracted from SAV and FASTQC files. Illumina's Interop is a software tool used to analyse and interpret data generated by Illumina sequencers. We used Interop to collect data generated during sequencing runs, to obtain quality control metrics that monitor and evaluate sequencing performance. This software allows the detection of any issues that may arise, such as poor cluster density, low quality scores, or poor sequencing performance.

We also used MultiQC, which is a software tool used for the quality control analysis of sequencing data. MultiQC allowed us to summarise and visualise the sequencing metrics across all samples. Conveniently, MultiQC includes a plugin for parsing SAV files that generates a visual report of all samples together.

5. Results

CNAG has compared the sequencing quality metrics of all participants to assess the performance for each one of these parameters. The goal was not to include as many metrics as possible but to select a small subset of them that was meaningful for this purpose. We determined the assigned value by using the mean of all the observations and from the relevant literature and guidelines we obtained an indicative value that serves as an overall reference against which to compare each point. The detail and figure for each comparison is discussed below. The plots show the comparison of the submitted results by laboratory, for normal and tumour samples separately. The solid line represents the assigned (overall mean) value, and the dashed line is the indicative value.

5.1 % Bases with Q >= 30 R1 and R2

For this metric, the indicative value reported in the literature (Marshall et al. 2020), and matches Illumina specifications is ≥ 85%. Results below this value are deemed questionable or unacceptable.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

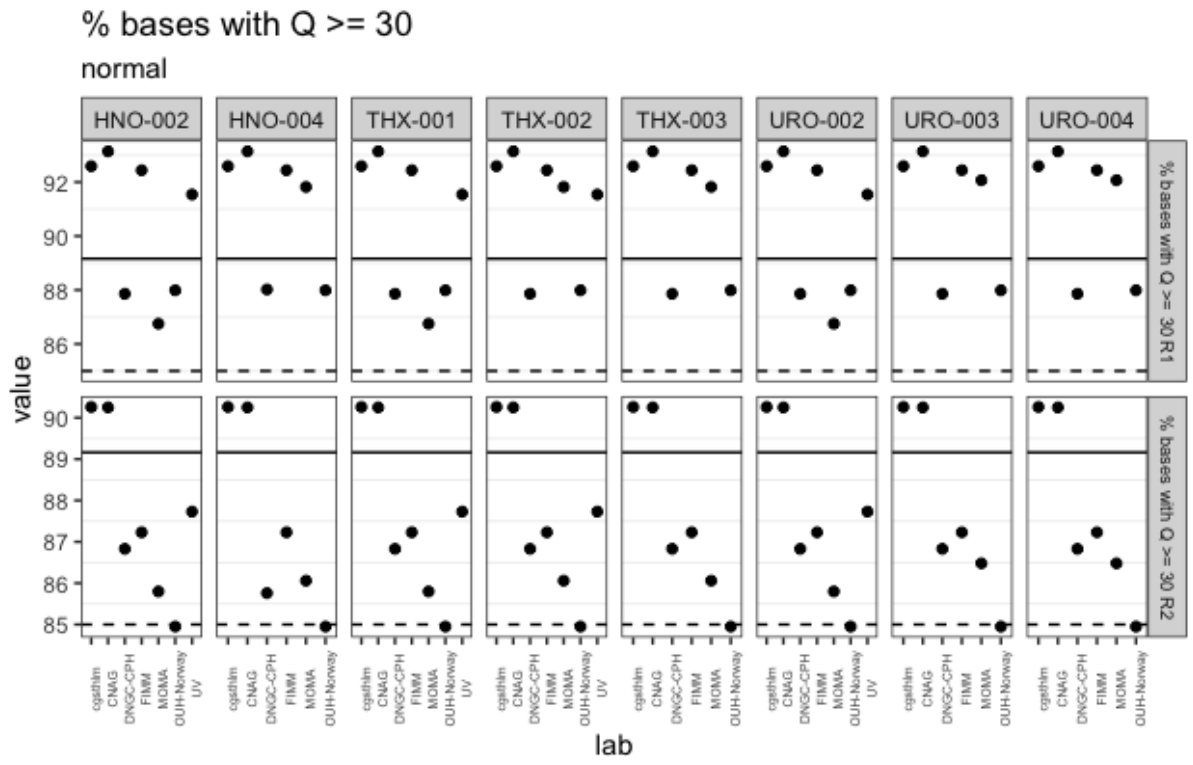
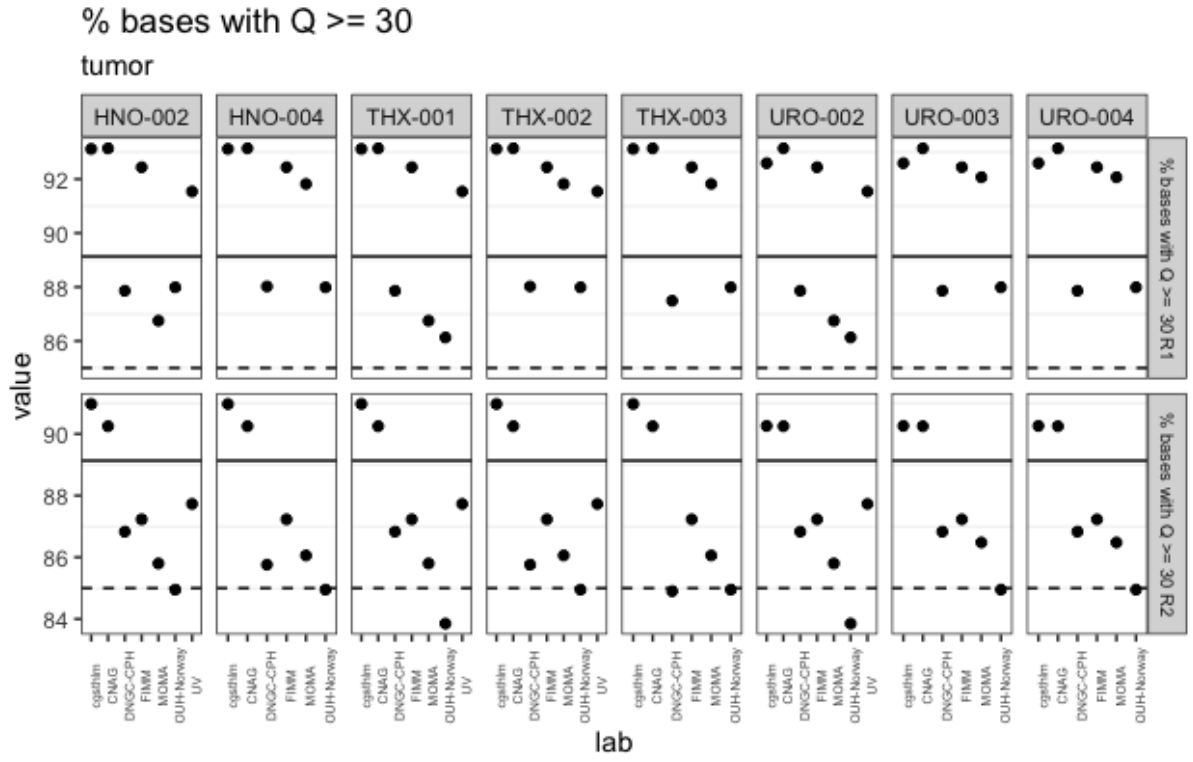


Figure 3: Percentage of bases with Q ≥ 30 for R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is 85% and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples



B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Figure 4: Percentage of bases with Q ≥ 30 for R1(top panel) and R2 (bottom panel). The indicative value (dashed line) is 85% and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

All participant laboratories report % bases with Q ≥ 30 that are above the indicative value (≥ 85) for the R1 strand of the normal sample, except for OUH-Norway that reported values slightly lower than, or right at, 85. In the case of tumour samples, SciLifeLab and CNAG are above average in bases with Q ≥ 30 for all samples, whereas OUH-Norway is at or below the indicative value in all cases. DNGC-CPH has indicative values for samples THX-003 but performs above indicative values for all other cases. Even if some values are slightly below the indicative value, the z-score estimates are all acceptable (all absolute z-scores are < 2, data not shown).

5.2 % PhiX alignment R1 and R2

The indicative value in this case is ≥ 1. Results below this value are deemed questionable or unacceptable. The optimal value is 1 , if 1% Phix is spiked in.

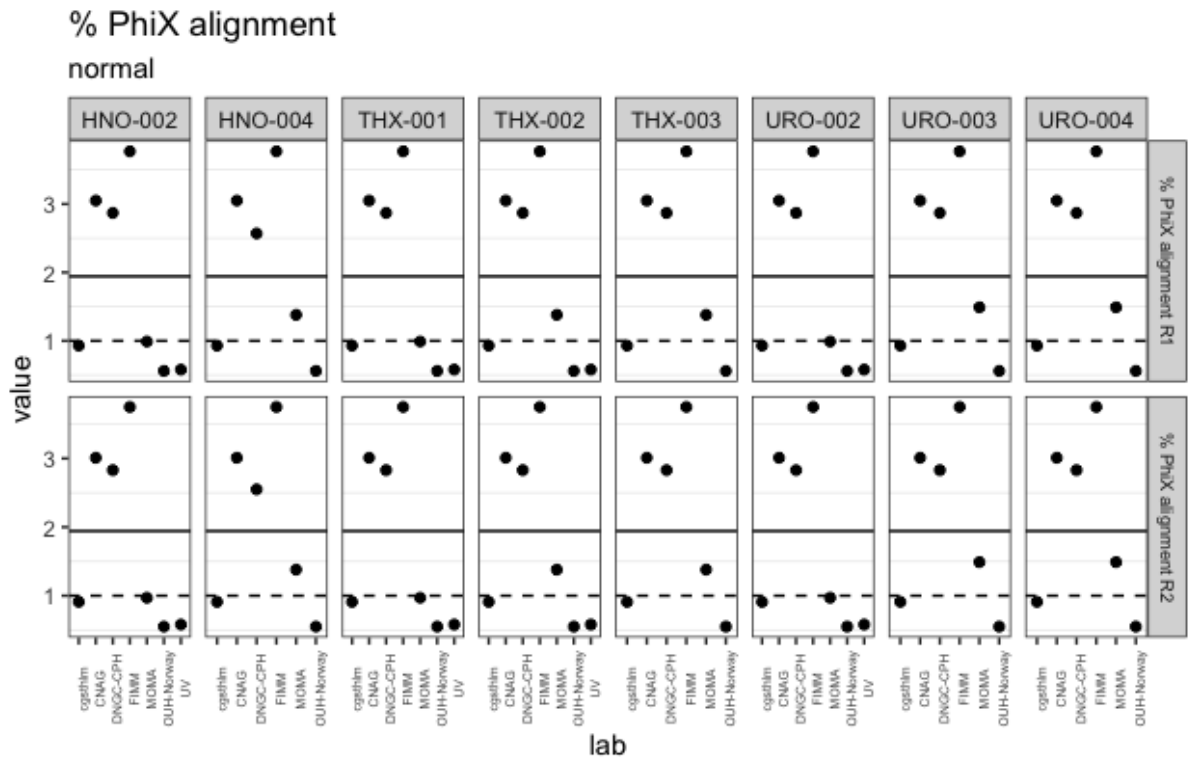


Figure 5: Percentage of the sample that aligned to the PhiX genome in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≥ 1% and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

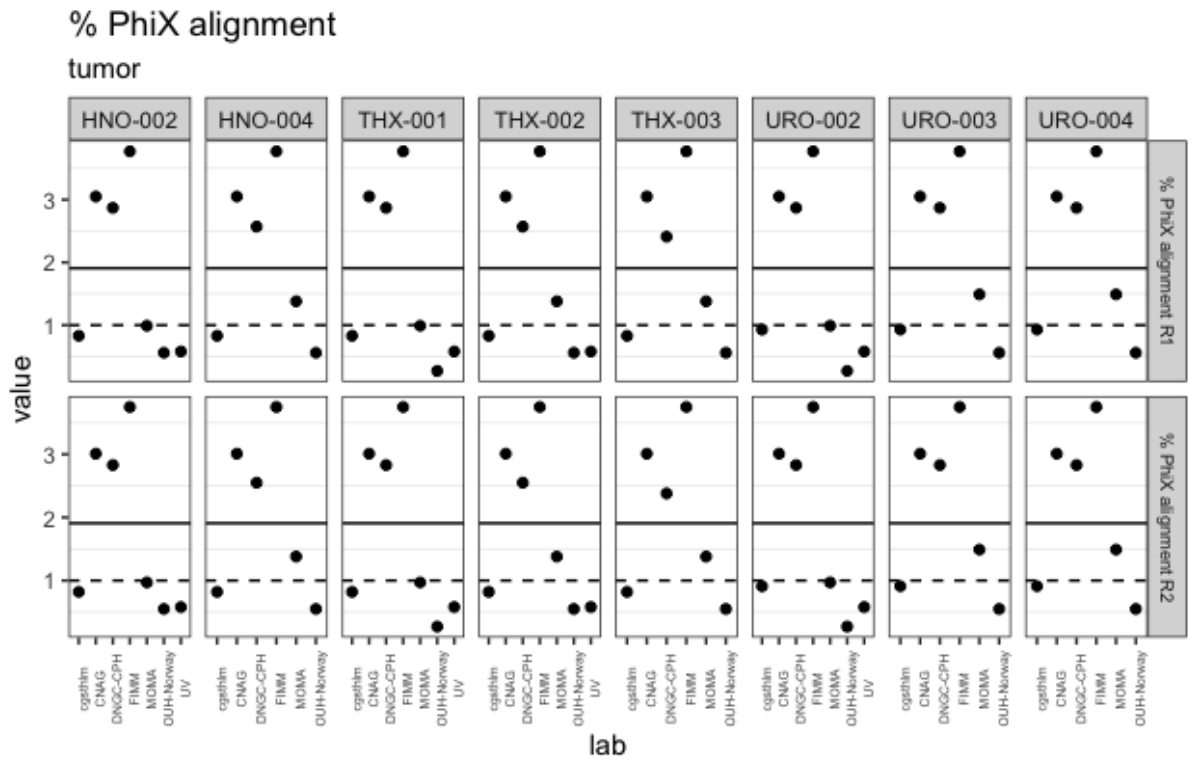


Figure 6: Percentage of the sample that aligned to the PhiX genome in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≥ 1% and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

Although we find many cases where this metric is below 1, the z-scores for all participants across samples are within the acceptable range (below 2). Nevertheless, the laboratories that obtained a value <1 will be advised to check for potential errors, as a low alignment percentage may indicate issues such as contamination or low sequencing quality. Other metrics are also assessed to contrast these results.

5.3 PhiX error rate R1 and R2

The indicative value for this metric is ≈ 0. The optimal is that no error is produced during the sequencing of the spiked in PhiX.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

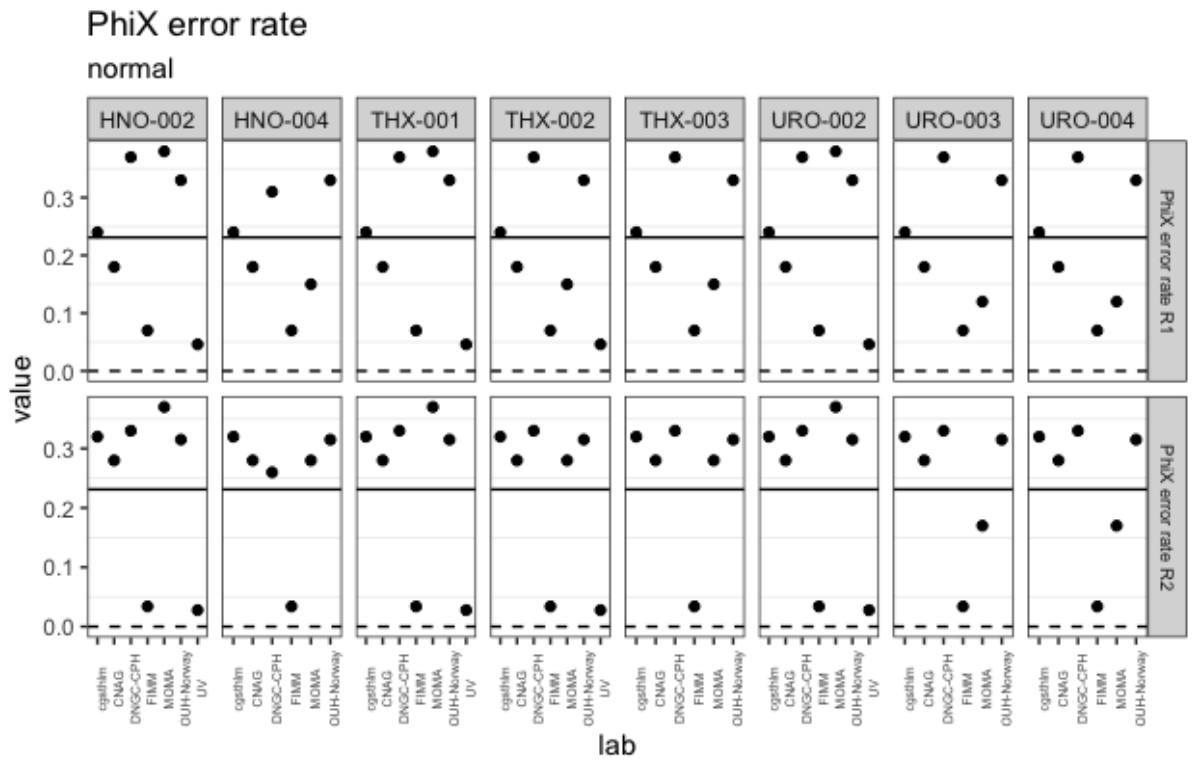
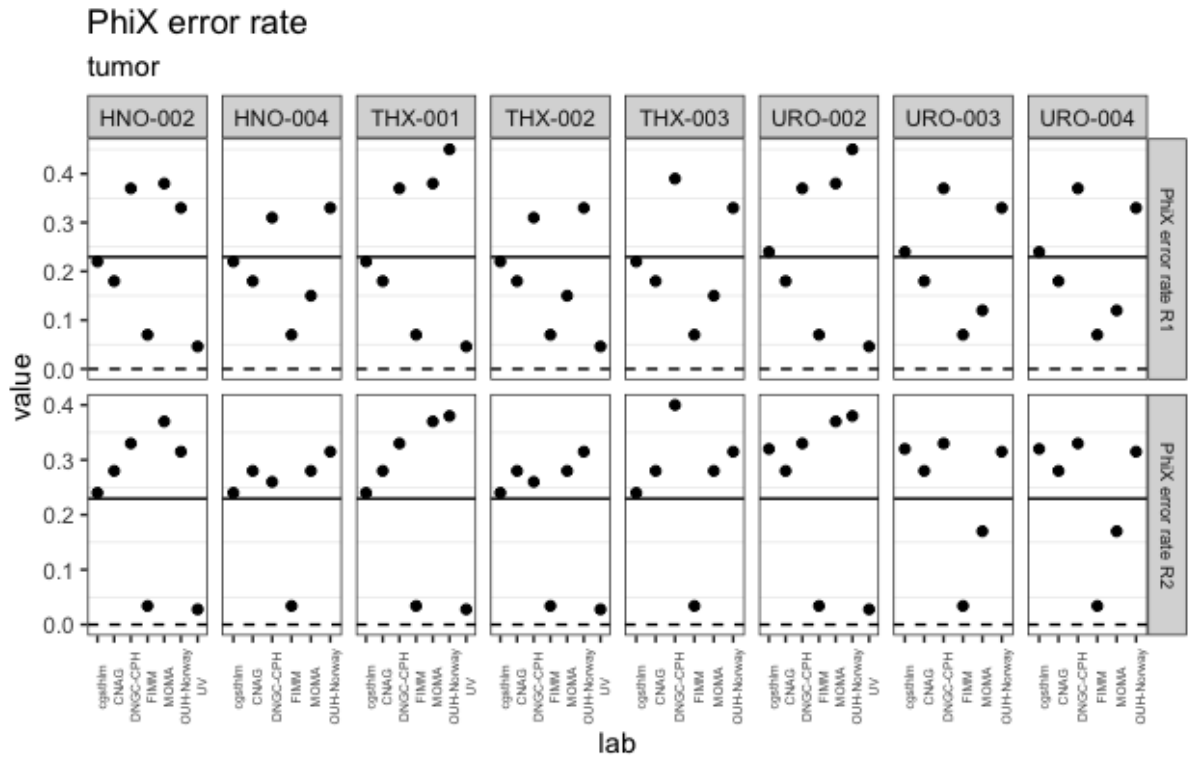


Figure 7: The calculated error rate in R1 (top panel) and R2 (bottom panel), as determined by a spiked in PhiX control sample. The indicative value (dashed line) is ≈ 0, and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples



B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Figure 8: The calculated error rate in R1 (top panel) and R2 (bottom panel), as determined by a spiked in PhiX control sample. The indicative value (dashed line) is ≈ 0, and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

The reported PhiX error rate is below 0.5 in both normal (< 0.4) and tumour (< 0.5) samples across laboratories. All z-scores are also within the acceptable area (<2). The labs that are above the assigned value will be advised to check this parameter, as a high PhiX error rate may indicate issues such as low quality sequencing reagents, poor instrument calibration, or other issues that may affect the overall quality of the sequencing results. Notwithstanding, all z-scores are acceptable and below 2, so the dispersion among sequencing centres is not bad.

5.4 Passing filter (PF) clusters

The indicative value for this metric is > 53, according to Illumina specifications (https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html)

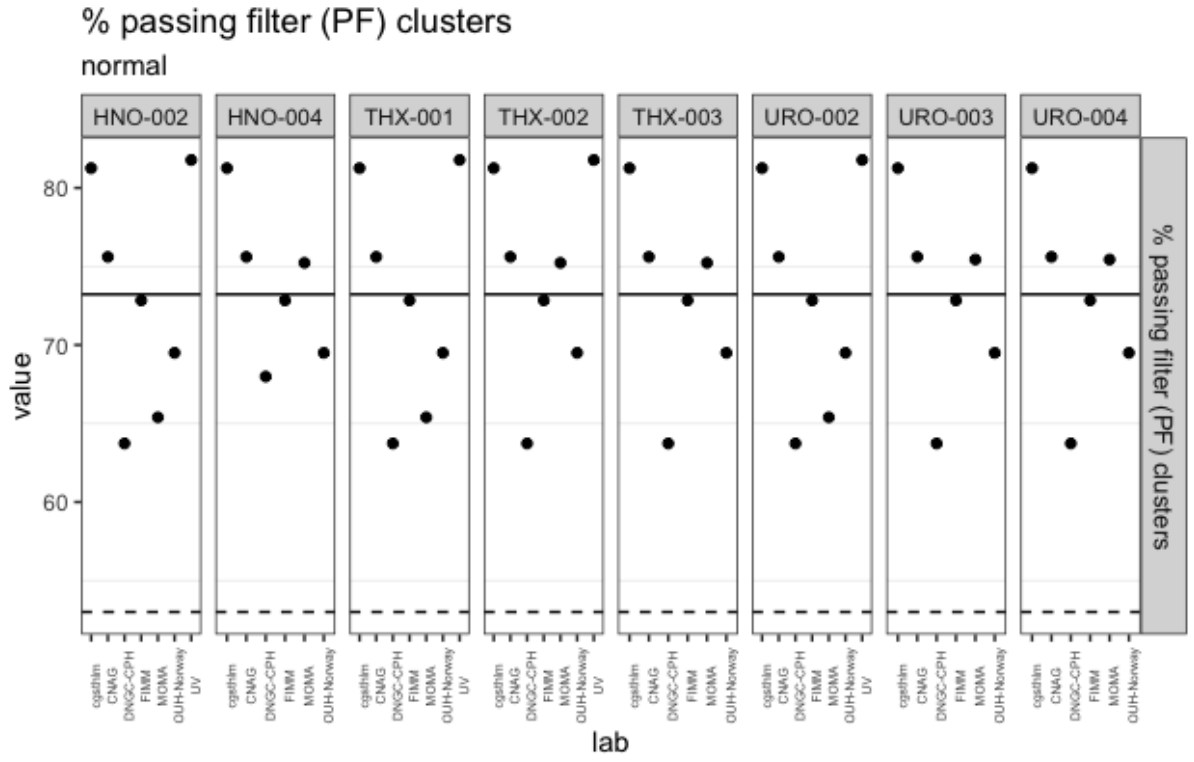


Figure 9: Percentage of clusters passing filters after evaluating cycles 1 to 25, as specified by the manufacturer, in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is > 53% and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

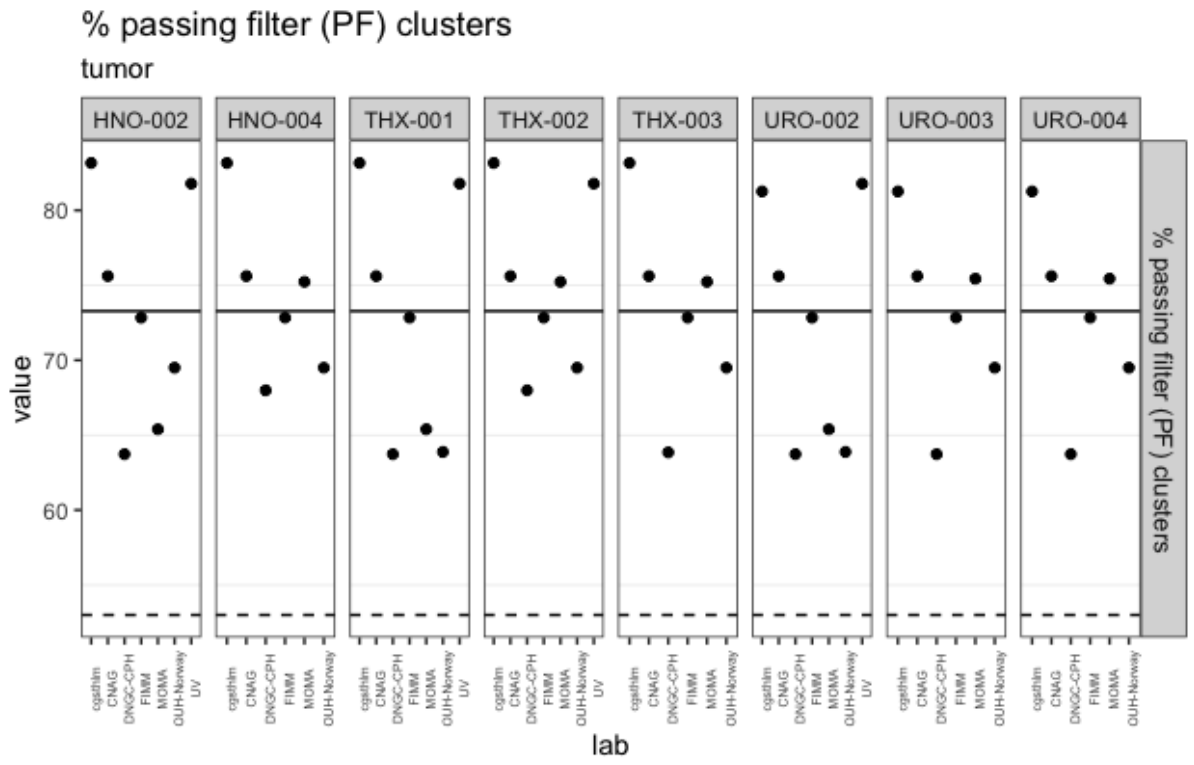


Figure 10: Percentage of clusters passing filters after evaluating cycles 1 to 25, as specified by the manufacturer, in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is > 53% and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

All sequencing centres report % passing filter values well above the indicative value of > 53. All the values give high confidence in the sequencing performed in all participating laboratories.

5.5 % Phasing R1 and R2

The indicative value is ≈ 0. The optimal is that all the molecules in a cluster are synchronised.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

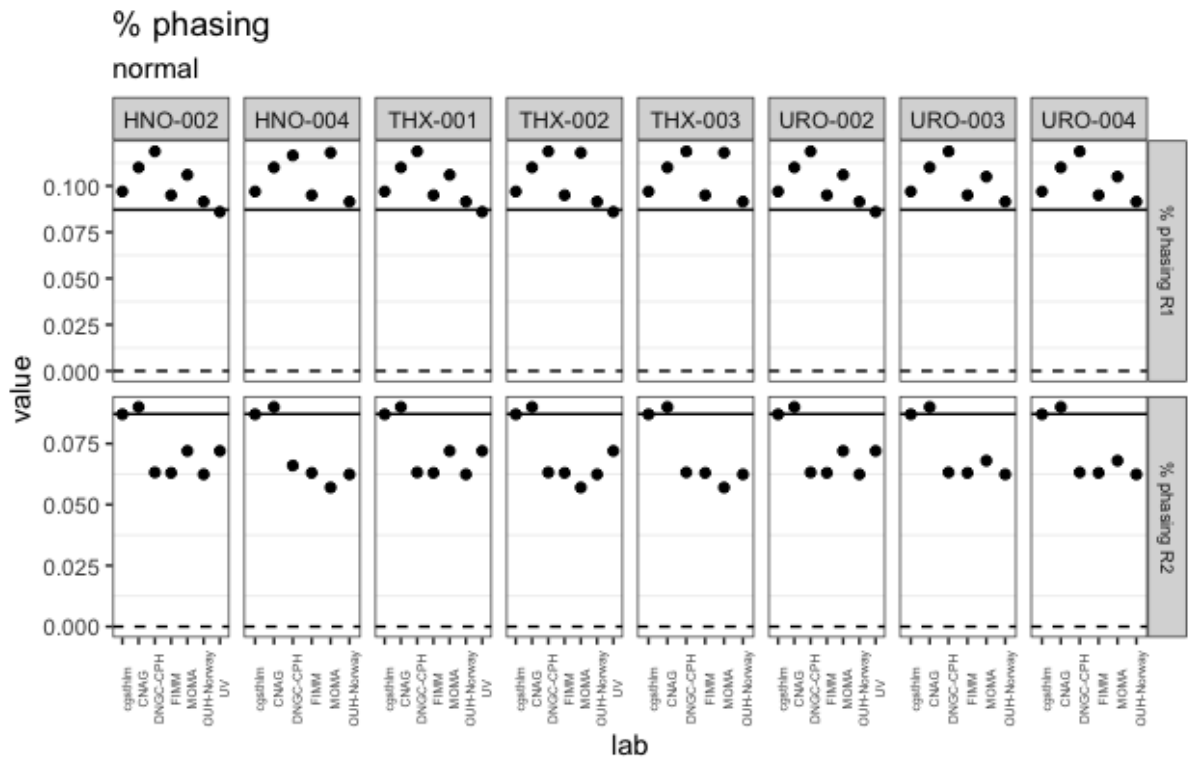
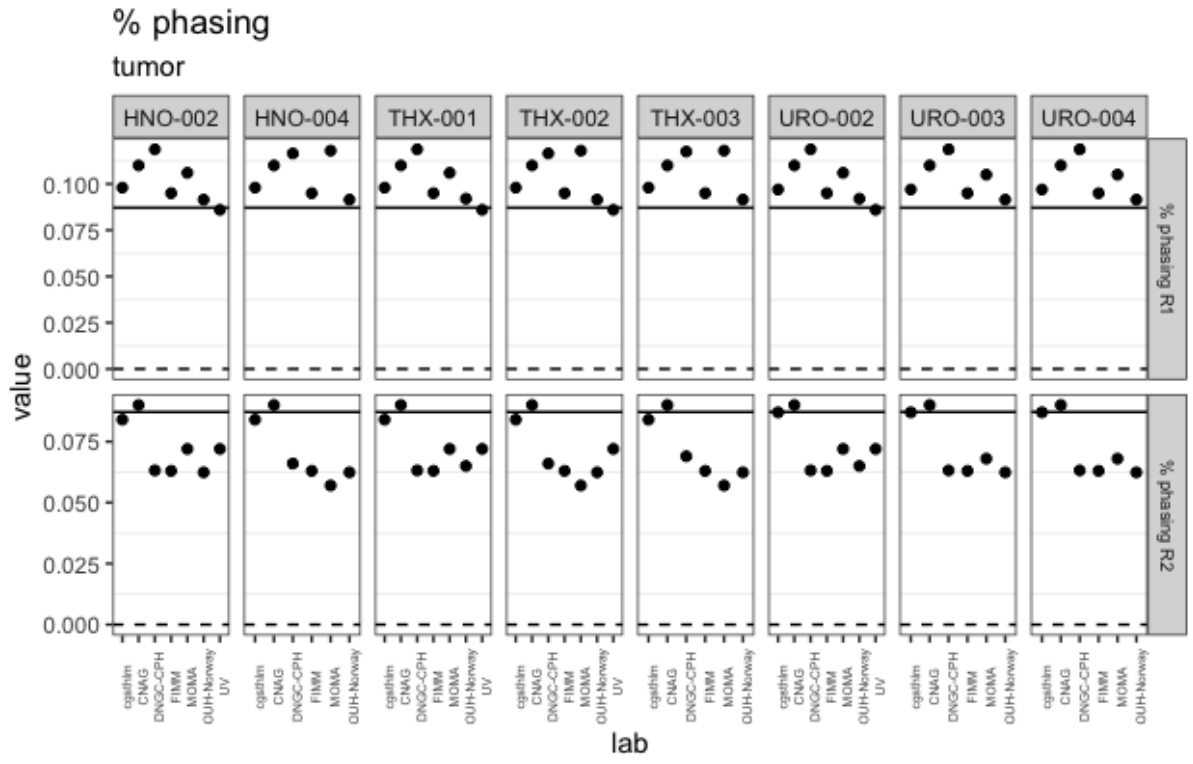


Figure 11: Average percentage of molecules in a cluster, per cycle, for which sequencing falls behind the current cycle during the read in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≈ 0 and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples



B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Figure 12: Average percentage of molecules in a cluster, per cycle, for which sequencing falls behind the current cycle during the read in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≈ 0 and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

The percent of phasing is below 0.2 in all cases, indicating an acceptable level of synchronisation in the clusters. The centres whose reported phasing values are above the assigned level will be advised to check if this parameter can be improved. All z-scores fall in the acceptable area < 2.

5.6 % Prephasing R1 and R2

The indicative value is ≈ 0. The optimal is that all the molecules in a cluster are synchronised.

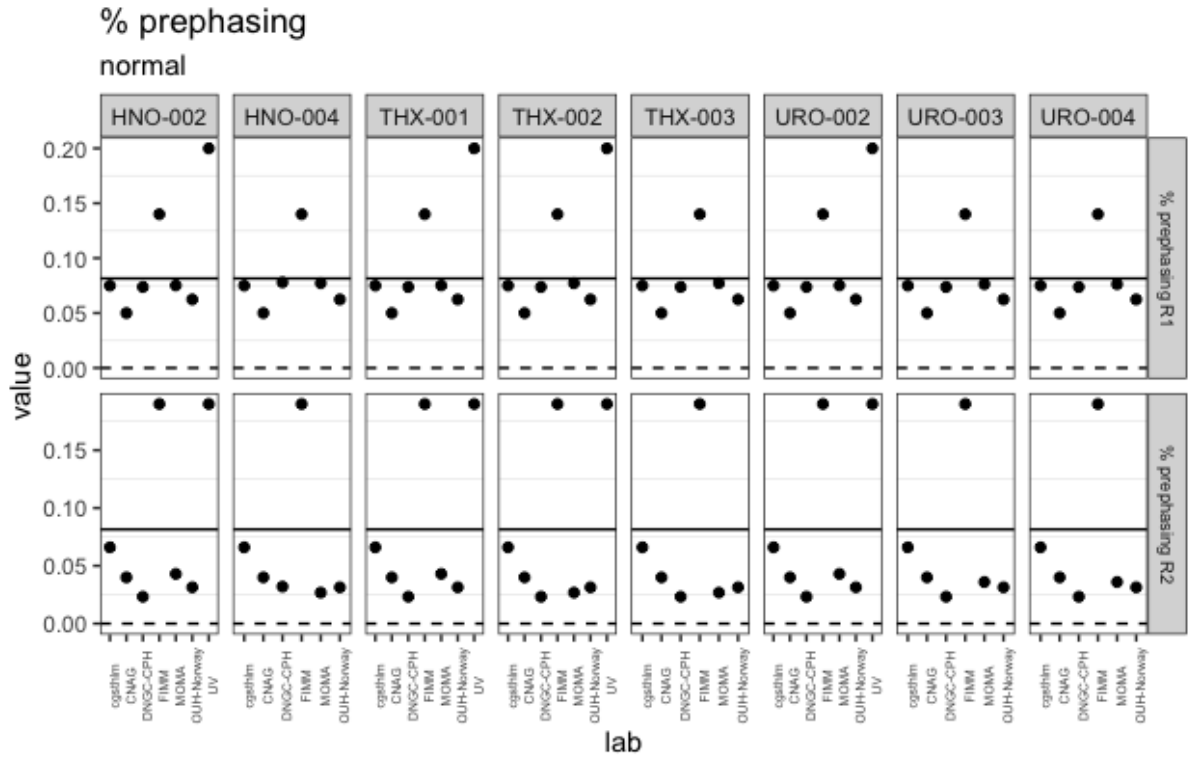


Figure 13: Average percentage of molecules in a cluster, per cycle, for which sequencing jumps ahead the current cycle during the read in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≈ 0 and the assigned value (solid line) corresponds to the overall mean for all observations of normal samples

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

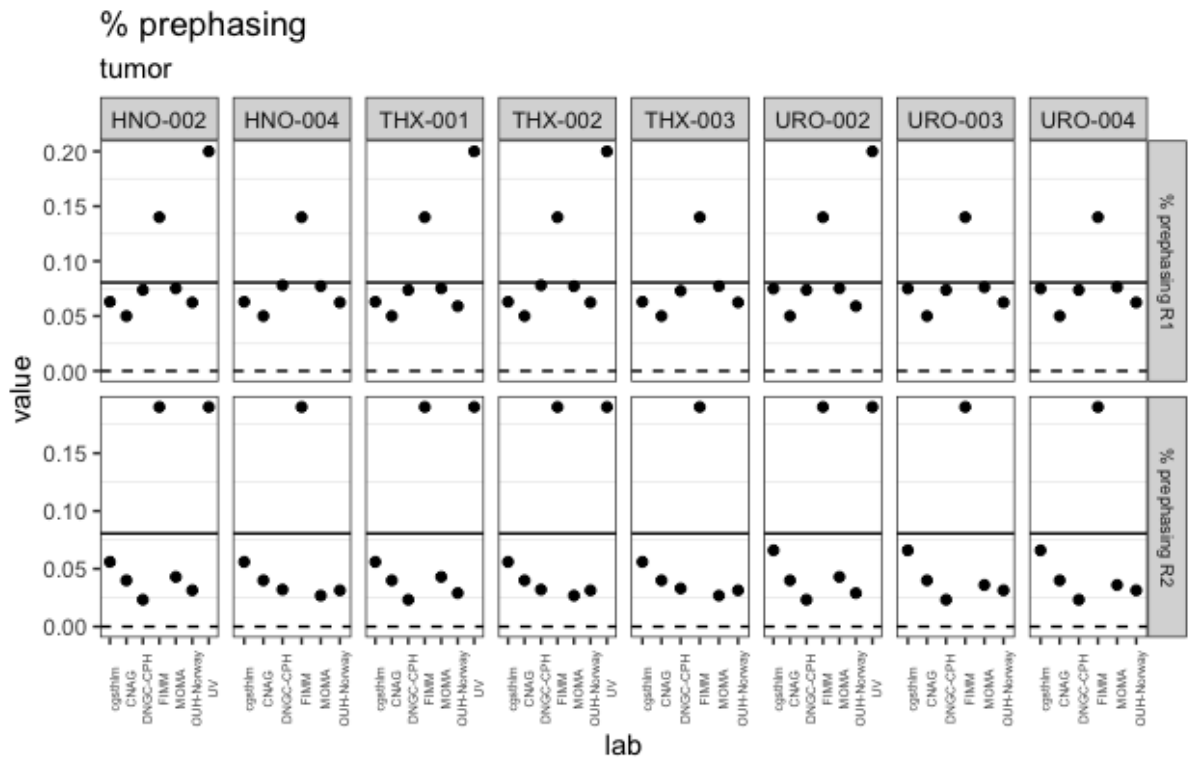


Figure 14: Average percentage of molecules in a cluster, per cycle, for which sequencing jumps ahead the current cycle during the read in R1 (top panel) and R2 (bottom panel). The indicative value (dashed line) is ≈ 0 and the assigned value (solid line) corresponds to the overall mean for all observations of tumour samples

The percent of prephasing is below 0.2 in all cases, for both normal and tumour samples, and across sequencing centres. For this metric there are a few centres with values above the assigned value. This indicates a good performance, as lower prephasing percentage indicates that fewer reads are being prematurely called, which can improve the accuracy and reliability of downstream analyses.

6. Discussion

6.1 Evaluation of the participants performance

Table 1. Summary of the best performing sequencing centres per metric across normal and tumour samples. Note that all participants reported values within the acceptable range and little dispersion (< 2), so the differences in performance are not big

Metric	Normal	Tumour
% bases with Q >= 30 R1	CNAG	CNAG
% bases with Q >= 30 R2	SciLife Lab	SciLifeLab

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

% PhiX alignment R1	FIMM	FIMM
% PhiX alignment R2	FIMM	FIMM
PhiX error rate R1	UV	UV
PhiX error rate R2	UV	UV
% passing filter (PF) clusters	UV	SciLifeLab
% phasing R1	UV	UV
% phasing R2	OUH-Norway	FIMM
% prephasing R1	CNAG	CNAG
% prephasing R2	DNGC-CPH	DNGC-PH

Table 1 shows how the participant laboratories rank according to how close their values are with respect to assigned and indicative values. Participants' values were close to the assigned one in each comparison, as estimated by z-scores. This means that despite small differences in sequencing and followed protocols, the results across centres do not deviate much from the mean, and conform to expectations given by common practice and the literature. Moreover, the indicative values were good indicators of performance and all observations fall near expected values. The differences in sequencing performance were not pronounced. At this stage, differences in sequencing machine and protocol do not appear to have a large impact on the quality achieved.

7. Conclusions

7.1 Best practices for Next Generation Sequencing

From the present benchmarking exercise, we can draw conclusions about the best practices for WGS for somatic variants. The first observation is that all participants have used Illumina sequencers (NovaSeq 6000), and have done PCR-free sequencing. Most participants have employed S4 flowcells and one has also used S1.
What differs among laboratories is the use of automated and manual library preparations. Half have done it manually and half have automated this step. The preparation kits used include TruSeq, MiSeq and one also includes tagmentation.
All of these factors do not seem to have a big impact on the quality of the sequencing, as the results are all satisfactory, fall close to the mean and are well within the acceptable range, as indicated by z-scores.

At this stage, bioinformatic QC pipelines detect significant deviations and all participating laboratories pass all indicative thresholds. For now we only observe the library preparation and sequencing steps. These two processes appear to be followed carefully by all participants and similar sequencing machines, preparation kits and flowcells ensure good quality results. **In conclusion the only used procedure for preparing libraries is PCR-free which from the point of view of sequencing returns sufficient quality.**

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

8. Next steps

The 1+MG WP4 somatic benchmark is an ongoing effort. In the next stages, we will complete the evaluation of the QC metrics listed in Table 2. In parallel, we will continue building a gold set of variants in order to compare the variant calling results from participants. We anticipate that curating the structural and copy number variants will be a significant challenge. To make that step easier, we will employ the variant voter tool to obtain a curated set of variants. After the completion of the benchmark, individual and general reports will be written. One or more publications are also planned in order to reach the largest possible audience that may benefit from having best practice recommendations for NGS.

9. Impact

Despite the explosion of sequencing projects worldwide, there is no standard for the sequencing procedures or the metrics needed to assess their quality. In particular, sequencing of tumour DNA presents specific technical and bioinformatics challenges. Also, new sequencing technologies are often incorporated into analytical pipelines and different laboratories implement them in their own way. Hence, there is an urgent need to establish common standards for the practice of Next Generation Sequencing (NGS), and to require periodic accreditation and validation of laboratories. A first step in this direction is to set up a benchmarking exercise for somatic calling comparing the performance of the methods typically used to process and analyse data by different laboratories. Results from this project suggest best practices that ensure reproducibility and guarantee the highest standards and reliability of the calls, especially in the case of tumour sequencing. Clearly, more QC metrics and the comparison of results with a gold set will give an even more comprehensive picture of the best standards that need to be implemented for each step of the sequencing and variant calling process.

10. Appendix (Supplementary Tables and Figures)

Table 1. The 8 Tumour/Normal pairs used in the somatic benchmarking

Sample Name Tumour	Sample Name N	Cancer Type
URO-002-03	URO-002-17	Clear Cell Renal Carcinoma
URO-003-01	URO-003-03	Clear Cell Renal Carcinoma
URO-004-02	URO-004-06	Clear Cell Renal Carcinoma
HNO-002-07	HNO-002-01	Head & Neck Squamous-Cell Carcinoma
THX-001-06	THX-001-02	Lung Squamous-Cell Carcinoma
THX-002-07	THX-002-01	Lung Squamous-Cell Carcinoma
HNO-004-07	HNO-004-01_02	Head & Neck Squamous-Cell

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

		Carcinoma
THX-003-09	THX-003-04	Lung Squamous-Cell Carcinoma

Table 2. QC metrics for evaluation

	Metric	Method	Threshold	Indicative value
Sequencing metrics (T+N)	% bases with Q ≥ 30	Extract from Illumina SAV	Lower	R1, R2: ≥ 85%
	% PhiX alignment		Lower	R1, R2: 1%
	PhiX error rate		Upper	R1, R2: 0% (optimal)
	% passing filter (PF) clusters		Lower	100% (optimal)
	% phasing		Upper	R1, R2: 0% (optimal)
	% prephasing		Upper	R1, R2: 0% (optimal)
Alignment metrics (T+N)	% duplicate reads	Picard	Upper	< 10%
	Median insert size		Lower	> 300

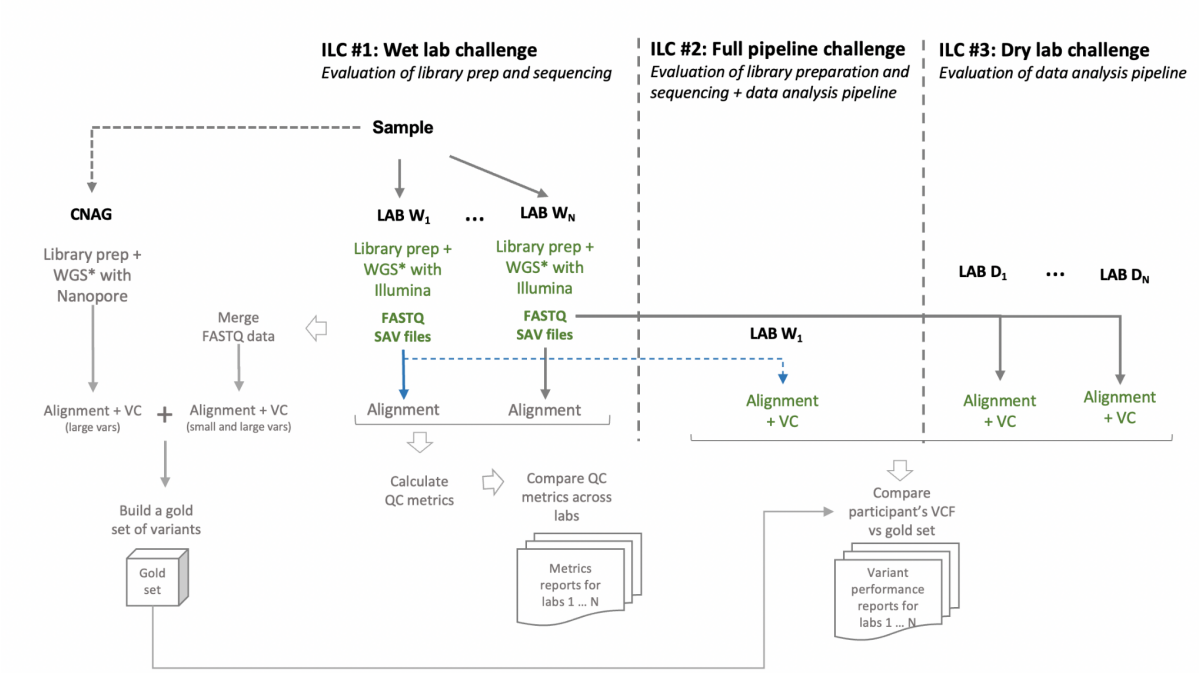
B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	Mean coverage		Lower	Normal: ≥ 30; Tumour: ≥ 70
	Evenness of coverage		Both	1
	% chimeras		Upper	< 1%
Variant calling metrics (T+N)	Ti/Tv ratio	Picard	Both	≈ 2.0
	% callability	Custom script	Lower	> 95%
Somatic metrics	Tumour purity	FACETS	Both	N/A
	Number of somatic SNVs	Picard		
	Number of somatic INDELs			
	Number of somatic SVs	Calculated only for full pipeline challenge using the information from variant callers		
	Number of somatic CNVs			

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Figures

Figure A1. The somatic benchmarking design scheme



B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724