

GUIDELINE

Open Access



# Recommendations for bioinformatics in clinical practice

Ksenia Lavrichenko<sup>1</sup>, Emilie Sofie Engdal<sup>2</sup>, Rasmus L. Marvig<sup>2</sup>, Anders Jemt<sup>3,4</sup>, Jone Marius Vignes<sup>5</sup>, Henriikki Almusa<sup>6</sup>, Kristine Bilgrav Saether<sup>4</sup>, Eiríkur Briem<sup>7</sup>, Eva Caceres<sup>3,4</sup>, Edda María Elvarsdóttir<sup>7</sup>, Magnús Halldór Gíslason<sup>2</sup>, Maria K. Haanpää<sup>8</sup>, Viktor Henmyr<sup>9</sup>, Ronja Hotakainen<sup>10,11</sup>, Eevi Kaasinen<sup>10,11</sup>, Roan Kanninga<sup>12</sup>, Sofia Khan<sup>10,11</sup>, Mary Gertrude Lie-Nielsen<sup>5</sup>, Majbritt Busk Madsen<sup>2</sup>, Niklas Mähler<sup>13</sup>, Khurram Maqbool<sup>3</sup>, Ramprasad Neethiraj<sup>14</sup>, Karl Nyrén<sup>4</sup>, Minna Paavola<sup>8</sup>, Peter Pruißscher<sup>3,4</sup>, Ying Sheng<sup>1</sup>, Ashish Kumar Singh<sup>15</sup>, Aashish Srivastava<sup>5</sup>, Thomas K. Stautland<sup>5</sup>, Daniel T. Andreasen<sup>16</sup>, Esmee ten Berk de Boer<sup>3,4</sup>, Søren Vang<sup>16</sup>, Valter Wirta<sup>3,4</sup> and Frederik Otzen Bagger<sup>2\*</sup>

## Abstract

**Background** Next-generation sequencing (NGS) is well established in clinical diagnostics, and whole-genome sequencing (WGS) is increasingly becoming the method of choice, as a result of lower prices and robust comprehensive data. While guidelines exist for variant interpretation and laboratory quality considerations, there remains a need for standardised bioinformatics practices to ensure clinical consensus, accuracy, reproducibility and comparability.

**Methods** This article presents consensus recommendations developed by 13 clinical bioinformatics units participating in the Nordic Alliance for Clinical Genomics (NACG) by expert bioinformaticians working in clinical production. The recommendations are based on clinical practice and focus on analysis types, test and validation, standardisation and accreditation, as well as core competencies and technical management required for clinical bioinformatics operations.

**Results** Key recommendations include adopting the hg38 genome build as reference, and a standard set of recommended analyses, including the use of multiple tools for structural variant (SV) calling and in-house data sets for filtering recurrent calls. Clinical bioinformatics in production should operate at standards similar to ISO 15189, utilising off-grid clinical-grade high-performance computing systems, standardised file formats and strict version control. Reproducibility should be ensured through containerised software environments. Pipelines must be documented and tested for accuracy and reproducibility, minimally covering unit, integration and end-to-end testing. Standard truth sets such as GIAB and SEQC2 for germline and somatic variant calling, respectively, should be supplemented by recall testing of real human samples that have been previously tested using a validated method. Data integrity must be verified using file hashing, while sample identity must be confirmed through fingerprinting and genetically inferred identification markers such as sex and relatedness. Finally, clinical bioinformatics should encompass diverse skills, including software development, data management, quality assurance and domain expertise in human genetics.

\*Correspondence:

Frederik Otzen Bagger  
frederik.otzen.bagger@regionh.dk

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

**Conclusions** These recommendations provide a consensus framework for standardising bioinformatics practices across clinical WGS applications and can serve as a practical guide to facilities that are new to large-scale sequencing-based diagnostics, or as a reference for those who already run high-volume clinical production using NGS.

## Background

In the field of clinical variant interpretation and classification, there are best practice recommendations available, such as the ACMG guidelines [1] and the ACGS guidelines [2]. However, these guidelines rely heavily on the accuracy of the bioinformatics pipelines used to call the variants. In 2018, Roy et al. [3] presented consensus recommendations for the validation of clinical NGS bioinformatics pipelines. Since then, the requirements in the field have grown, notably in the throughput of samples and the size of the data. Both parameters are largely driven by the success of whole-genome sequencing (WGS) which has proven a major advantage over targeted gene panels or exome sequencing, both of germline cells for diagnosing hereditary diseases [4–7] and for identification of treatment targets in somatic mutations for cancer diagnostics [8, 9]. Therefore, sequencing-based diagnostics units now handle a growing number of samples, and for each of these, more data is generated. Moreover, there has been an increase in expectations regarding turnaround time (TAT), quality standards and performance evaluation within an operating clinical bioinformatics unit. An increase in knowledge on the optimal analysis pipeline, as well as new diagnostic scores and markers, also means a higher number of expected analysis outputs from the same sample, and it also calls for individually tailored analysis targeted towards specific patient groups [10]. Therefore, a classic bioinformatics core facility providing support for medical research [11] is now a very different entity compared to units supporting production-scale sequencing analysis for diagnostics, even if there is some overlap in competencies. The development towards the professional organisation of large-scale automated continuous production in the field of clinical bioinformatics has been accelerated as a result of increasingly larger capacity sequencers that require a high sample influx for cost-effective production and a low per-sample price [12] and acceptable TAT for urgent samples. The dynamics being that running large sequencing batches of samples is considerably cheaper, but high sample flow is needed to be able to start a large sequencing batch often.

Here, we report on current practice in 13 clinical bioinformatics units in the healthcare systems of the Nordic (and Northern Europe) countries and provide a set of general recommendations for clinical bioinformatics at scale. We will cover the processes from raw data output

from the sequencer to data processing and analysis steps until they result in interpretable information, usually a file with variants, or a clinical score.

The recommendations are based on a survey, a workshop and final approval by the members of Nordic Alliance for Clinical Genomics present at the Clinical Bioinformatics workshop in Helsinki 2023 [13]. By design, only recommendations that could achieve unanimous votes have passed, and as such, all authors have the right to veto any recommendation. In brief, a summary of the survey provided the starting point for a presentation at a plenary session where first the most radical interpretation of a recommendation was presented, and subsequently, the language was softened or the statements were moderated until a full agreement could be reached, or the recommendation was rejected. Here we give recommendations that are based on clinical practice at at least two sites. In practice, we found that most sites follow most of the recommendations.

The scope of our survey and workshop was primarily next-generation sequencing (NGS), acknowledging that there may be differences between tools and pipelines for targeted panels, exome and whole-genome sequencing (WGS). We had as an aim to present a perspective that can be generalised to all these settings, and when differences were incompatible we would prioritise emphasis on WGS.

We acknowledge the relevance of sequencing platforms like long-/short-read and different vendors and technologies; however, we decided a priori not to specifically address platform-specific recommendations but merely record the platforms in use.

All authors and participants in the survey are engaged in clinical diagnostics, operating bioinformatics pipelines in production environments. Our recommendations aim to serve as a practical guide for hospitals or facilities new to large-scale sequencing-based diagnostics, or as a reference for inspiration or discussion for those who already run high-volume clinical production using NGS. The recommendations provide a basis for both clinical genomic operation for diagnostics use, as well as for medical research depending upon it.

## Methods

Recommendations were based on a survey (Additional file 1: Table S1) sent to NACG members participating in the workshop for Bioinformatics in Clinical

**Table 1** Recommendations for sequencing-based clinical bioinformatics**1. Genome build Hg38 is the reference for alignment**

2. A recommended standard set of analyses: SNV, CNV, SV, STR, LOH, variant annotation, PRS (optional), and for cancer, TMB, HRD, MSI
3. Several tools are needed in combination for calling structural variants
4. Structural variants must be filtered using a tool-specific matched in-house dataset, to filter common variants and false positives variant calls
5. Use reliable air-gapped clinical production-grade HPC and IT systems
6. Clinical bioinformatics in production should operate under ISO15189 or similar
7. Standardised file formats and terminologies should be used
8. Pipelines should be well documented, and tested for both accuracy and reproducibility against predefined acceptance criteria
9. Production code must be subjected to manual review and testing
10. Computer code and documentation must be developed and managed under strict version control in a git-tracked system
11. Validation using standard truth sets, GIAB (germline) and SEQC2 (somatic), should be accompanied by a recall-test of previous real human clinical cases from validated—preferably from orthogonal—methods
12. Pipelines should be tested at the levels of unit, integration, system, IT performance and end-to-end tests
13. Integrity of data must be verified using file hashing (e.g. MD5 or sha1)
14. Identity of sample should be verified via inference of identifying traits and checks for relatedness between samples (sample fingerprinting tests)
15. Software should be encapsulated in containers or conda environments
16. Clinical bioinformatics teams must be able to attract a combined skill set covering software development, data management, quality assurance and domain knowledge in human genetics

SNV single nucleotide variant, CNV copy number variant, SV structural variants, STR short tandem repeat expansions, LOH, loss of heterozygosity, PRS Polygenic Risk Score, TMB tumour mutational burden, HRD homologous recombination deficiency, MSI microsatellite instability, HPC high-performance computing, GIAB Genome in a Bottle, SOP standard operating procedure

Practice held in Helsinki from 28 to 29 September 2023. In the invitation, participants were specifically asked to compile answers from their unit, representing the opinions of the team. All statistics are based on the survey, and members have been asked to confirm that their views are represented whenever free-text information had to be summarised. The most extreme statement that could be devised from the questionnaire was put forth first, and if any concerns were raised, a less extreme alternative was proposed until a consensus could be found. Every member had veto rights on any statement. Three topics were planned for more detailed active discussions; participants were divided into three teams that were rotated three times to visit all topics, with a static session chair. The session chair summarised the discussions into recommendations that were put to vote in a final plenary session.

Participants were asked to support recommendations that they, based on their expertise, would recommend for clinical bioinformatics in production. Therefore, not all sites have a production environment that follows all of the recommendations, but by design, no recommendation is given that is not in production at at least two sites. All sites have been promised anonymity in their answers in this publication, to ensure that we can maintain the honest and open discussion that is at the heart of NACG.

**Results and discussion**

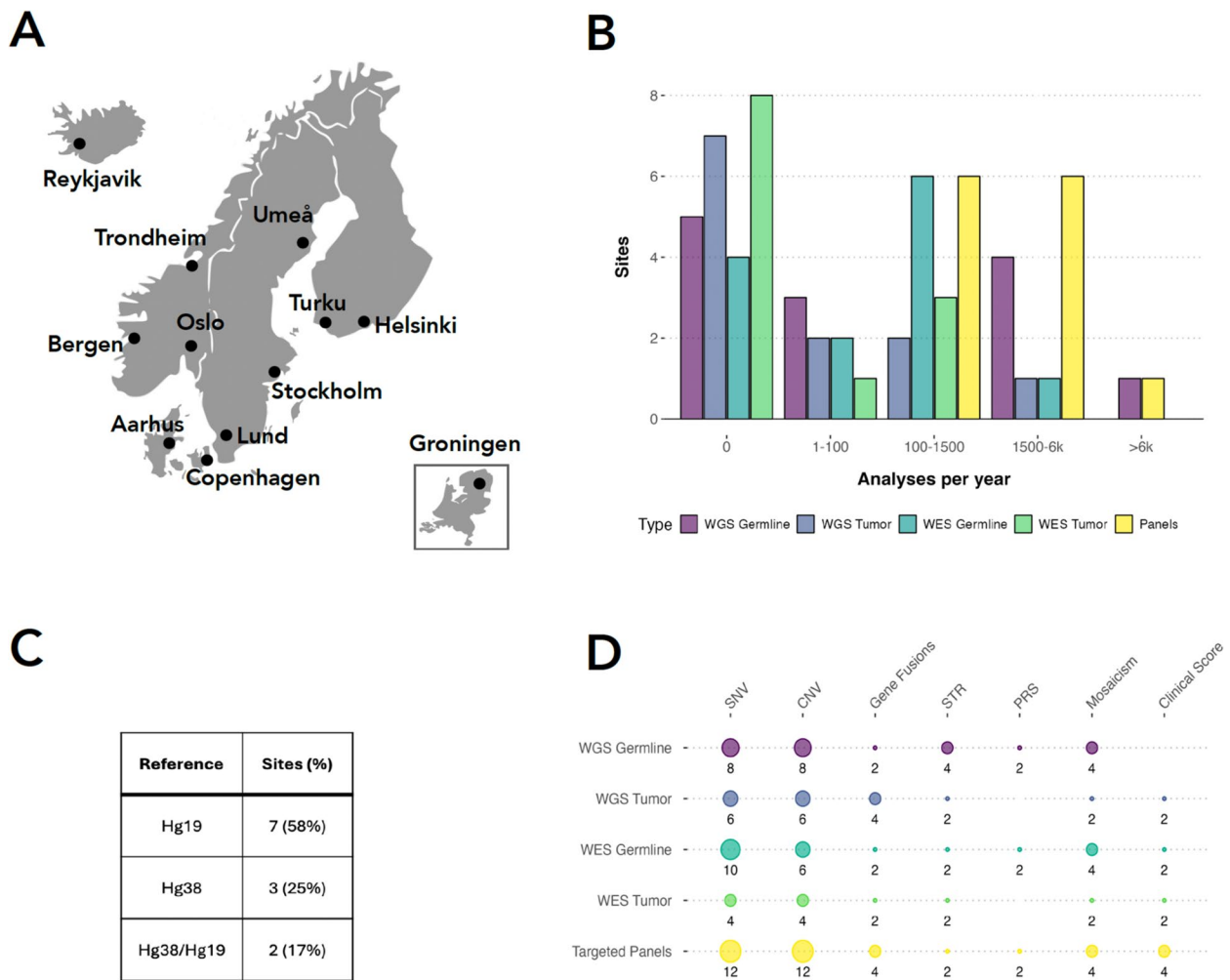
Based on a survey and a workshop in Helsinki in September 2023, with members of the NACG we arrived at 16 recommendations for clinical bioinformatics in production (Table 1), as well as summary statistics on sample numbers, methods and implemented standards at each site (Fig. 1 and Additional file 1: Table S1).

**A recommended set of analyses for clinical NGS**

The set of recommended analyses for clinical production of WGS that emerged in this study is overall in line with previous reports [7, 10, 14] with some optional points specific to whether the facility is performing cancer analysis or not. Automated quality assurance is usually handled partially or fully within the analysis pipeline; it is a multifaceted task influencing all aspects of production, which will be handled in later sections.

Whereas there is a core set of operations that all facilities have to perform, there are—even when using the same software or tools—a multitude of parameters that should be set and optimised differently depending on the purpose and nature of sequencing data (Additional file 3: Table S1). In the following, we will stay at the level of describing a core set of analyses that we recommend for NGS-based diagnostics:

- De-multiplexing of raw sequencing output to disentangle pooled samples (BCL to FASTQ),



**Fig. 1** Overview of the NACG community. **A** Geographic distribution of NACG sites: a map illustrating the 12 member sites of NACG network and one guest site from the Netherlands. **B** Annual analysis volume per site: the volume of analyses performed per year across NACG sites. The x-axis shows four categories of analysis volume, while the y-axis indicates the number of sites (out of 13) in each category. Different colours represent various types of analyses, as indicated by the legend. **C** Genome reference build usage: a summary of genome reference build statistics across the 13 participant sites. **D** Details of pipelines in clinical production: the y-axis shows the type of analysis, and the x-axis shows the type of variant or metric reported. Bubble size reflects the number of sites in each intersection category. The colour legend is the same as in panel B

- Alignment of sequencing reads to a reference genome (FASTQ to BAM):
- Variant calling (BAM to VCF)
  - SNVs and small insertions and deletions (indels)
  - CNVs (deletions and duplications)
  - SVs including insertions, inversions, translocations and complex rearrangements.
  - STRs
  - LOH: loss of heterozygosity regions (indication of UPD, uniparental disomy)
  - Mitochondrial SNVs and indels
- Variant annotation (VCF to annotated VCF).

Of note, SNVs and indels (<50 base pairs) are often handled and conceptualised together because the same tools can output both. Inversely, copy number variants (CNV) and short tandem repeat expansions (STR) are often handled separately from other structural variations (SV) because tools and methods, at least historically, were dedicated to either one type or the other. Mitochondrial variant calling also benefits from a tailored approach.

Variant annotation was not part of the scope for these recommendations. It is a large and complex task and requires significant maintenance to keep the annotation data sources relevant and updated. For some units, the choice of annotation is considered the domain of the

clinical interpretation, and by others, it is handled in collaboration with the bioinformatics unit—also depending on the graphical user interface software used for variant interpretation and classification. There are also differences as to whether to use soft-filtering (just add information in the form of tags to the VCF-file) or to hard-filter (remove variants from the VCF-file) that do not pass a given criterion. Similarly, some units approach the variant annotation task as a variant prioritisation problem (ranking of variants from likely clinically relevant to likely irrelevant), and others by providing a fixed list for interpretation (often in the form of predefined gene panels).

### Optional analyses

- Microsatellite instability (MSI): MSI analysis assesses mutations in microsatellite regions to identify DNA mismatch repair defects, used for guiding immunotherapy treatment in cancers.
- Homologous recombination deficiency (HRD): HRD testing evaluates the integrity of homologous recombination repair pathways predicting response to PARP inhibitors in cancers, particularly ovarian and breast cancer.
- Tumour mutational burden (TMB): TMB quantifies the number of somatic tumour mutations. This is used as a proxy for producing neoantigens that could trigger an immune response, thus identifying patients likely to benefit from immunotherapy.
- Polygenic risk scores (PRS): PRS estimates an individual's genetic predisposition to complex diseases by aggregating effects of multiple genetic variants. Only one site reports to have this in production and it may be pending further validation and standardisation.

### Genome build hg38 as a reference

Not all sites have yet switched from GRCh37/hg19 to GRCh38/hg38 (Fig. 1C), but this transition is a recommendation, despite the significant transition cost. The hg38 build has reduced gaps, improved mappability and variant calling and corrected several clinically relevant errors present in the hg19 reference [15]. In addition, some of the important public reference data repositories continue their releases only in GRCh38/hg38 build—such as gnomAD v4 [16]. Despite these improvements, there are still several false duplications in several medically relevant genes [17], which should be addressed specifically, e.g. by masking those regions.

The transition to hg38, however, touches upon a large number of dependencies both in the bioinformatics processing, variant annotation, interpretation, reporting and legacy data that potentially need to be re-analysed. Using

automated tools to lift local accumulated variant classification and other manual annotation is not optimal, and usually means that any accumulated manual knowledge is not reliable for automated translation between the two different genome references [18]. The public annotation databases, tools and pipelines relevant for clinical production now all support hg38, with the exception of the panel-based TSO500 Illumina DRAGEN pipeline [19] that is in production at several sites.

### Several tools are required for structural variant calling

There are several strategies for calling structural variants, including models based on read depth quantification, assessment of split read mapping, distant or mis-oriented mapping of read-pairs and de novo assembly [20]. All sites experience and agree that all tools for calling structural variants have their caveats and that a combination of tools and strategies is needed, which is also what most sites have in production. There are different approaches to how to integrate the output from several callers, which could be either a consensus of calls (an intersect), a merge of calls (a union) or a weighted output that includes uncertainty measures, either at the variant level or the level of the tool (e.g. we always include calls from X, but only from Y, if there is agreement with Z).

We have seen that the implementation of long-read sequencing technology will aid precision in variant calling, but standards for CNV and other structural variant calling methods greatly need community standards for benchmarking and calling [21]. The latter has recently been aided by [22].

### False positive filter

To control for the large number of false positives, all sites use an in-house database of previously called variants, containing the frequency of how many times a certain structural variant has been called before. This database, sometimes referred to as a false-positive-filter (FPF) will consist of a mix of false positive calls and real but frequent variants, both of which are uninteresting for identification of clinically relevant rare variants [23]. Given the high number of false positives from any SV calling tool [20] determined by the specific strategy for calling the variants, the background database must be reconstructed at every update of the pipeline that significantly affects the calls, be it from a new version of the tool or a change in upstream processing. The practical experience is also that changes in wet-lab library preparation kit, the sequencing chemistry, and sequencer also have a sizable impact on the nature of false calls, just like the extent of familial ancestry between the patient and the samples in the FPF will impact the effectiveness of the filter [24].

### **Comparing structural variants**

To use FPF, it is necessary to be able to compare variants from a patient to a database. Comparing two or more structural variants is a non-trivial task because the precise break-end position or structural variant class is often not called with precision. Therefore, tools like SVDB [25] or Truvari [26] can be used to set criteria for overlap both in terms of distance between the break-ends and the required percentage overlap.

### **Accreditation and quality standards**

It is of importance for clinical laboratories to ensure trust in the quality of the performed work. A way to achieve this is to obtain accreditation by authoritative bodies to certify that the laboratory operates according to a certain scheme, e.g. a standard defined by a non-governmental international organisation or a national governmental body. All participants were accredited and performed analyses according to the requirements of either ISO 15189 [27] or ISO 17025 [28] standards (12 and 1, respectively) defined by the International Organization for Standardization. The difference between them is that ISO 17025 is a general standard for testing and calibration laboratories, while ISO 15189 is specific to laboratories that perform medical tests based on patient material, e.g. extensive genetic analyses. ISO 17025 serves as a normative reference for ISO 15189.

The accreditation is obtained for specific analyses, and it may not be feasible for a laboratory to have all analyses accredited, e.g. due to costs associated with accreditation, low volume of samples for the given analysis, frequent changes in procedures, or large heterogeneity in the procedures required for different sample types.

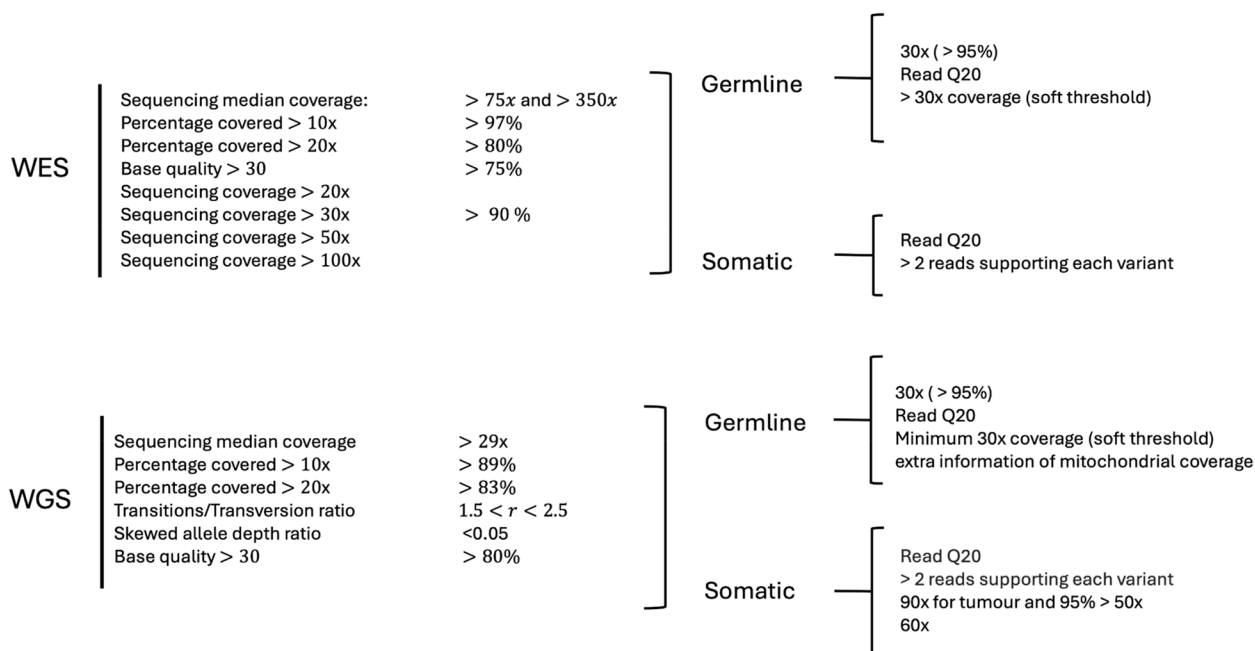
The scope of an accredited analysis is defined by the accredited laboratory, and as such, accredited analyses are, while they may be similar, often not equivalent across accredited laboratories. For example, two separate analyses in one laboratory may correspond to a single analysis in another laboratory. Accordingly, the mere count of analyses each laboratory has accredited does not translate into to which extent the accredited analyses cover the possible spectrum of genetic tests. Instead, we asked the participants for the aims of the bioinformatics pipelines that were used in accredited analyses. The accredited analyses across the 13 participants included bioinformatics pipelines aimed at one or more of the following analyses: calling germline small variants, copy number variants and structural variants in whole-genome sequencing data from DNA; somatic small variants in whole-genome sequencing data from DNA; gene fusion variants in sequencing data from total RNA; and small germline variants and copy number variants in sequencing data from DNA gene panels (Additional file 1: Table S1).

Quality schemes, such as the ISO 15189 or ISO 17025 standards, set requirements to cover a wide range of aspects that influence the quality of the work. While we do not make recommendations for all such aspects here, the participants agreed that methods within clinical bioinformatics are often not standardised, highly customised per laboratory and used for a scope that is not covered by the validation performed by the provider. Also, there is a lack of relevant external quality assessment programmes. Accordingly, all laboratories recommended special attention to, that bioinformatics pipelines are rarely externally validated for scopes relevant to clinical usage but instead requires validation by the user. We acknowledge that there is a need for standards that specifically address clinical bioinformatics in operation, a work that is currently being intensively progressed by the Global Alliance for Genomics and Health (GA4GH) [29], which has already published important standards for specific operations, like benchmarking of small variants [30] and data sharing.

All participants recommended to work according to, and ideally to be accredited according to, an ISO 15189-equivalent quality management system. Furthermore, a future implication is that work performed according to ISO 15189 standards fulfils European Union regulations 2017/745 (Medical Devices Regulation) and 2017/746 (In Vitro Diagnostic Medical Devices Regulation [31]; requirements for laboratory-developed devices (e.g. clinical bioinformatics pipelines) for which an equivalent CE-marked device is not available [32].

### **Safety, quality and compliance**

In clinical bioinformatics, a robust quality management system is a requirement, fully in line with any other diagnostics procedure and very different from research- or project-driven bioinformatics. Importantly, clinical diagnostic production relies on the ability to process samples continuously, rather than by manual handling of batches. This entails the need for a high level of automation and systems integration, and both data and operations must be designed and conceptualised as flows, rather than units handled successively. This additionally removes many sources of human errors and allows for scalable production. Automated quality procedures must balance the need for every single patient to receive an answer (it is not a viable solution to simply discard a large fraction of the samples to get a high quality dataset) and the detrimental effects that a wrong answer could have for a patient. The latter is further accentuated by the fact that germline genomic data is a lasting resource for later diagnostics and general reuse by the patient, immediate family and society at large.



**Fig. 2** Summary of reported coverage and quality threshold used in clinical production. Consensus of minimal thresholds and quality metrics remains limited for targeted panels and WES, whereas a minimum sequence coverage depth is agreed upon across all units for both somatic and germline testing

**Personal identification**

Patient information must be stored safely and be correct. Genomic information is, as per definition, both personal and sensitive in the European Union (GDPR recital 34 [33]), and additionally, healthcare information—by bioinformaticians sometimes referred to as clinical meta-data—is needed to run bioinformatics pipelines and to analyse the results. In some instances, it becomes a balancing act between data security, which often presents itself as a form of inaccessibility and data barriers, and patient safety that will suffer if the healthcare professionals do not have easy and operational access to the right information at the right time, or are forced to adopt either manual or unintended procedures to circumvent rigidly designed safety structures. In addition to high requirements for data handling and data flow, verification and validation must be in place to confirm data integrity in data flows; all sites use and recommend automated checks of hashes, e.g. MD5 or SHA-1, for data verification. Additionally, sample identity checks is a recommendation to detect swaps or mixing, and can be in the form of dual sequencing of samples in two different laboratory and data processing flows—one of which only needs to contain a small number of polymorphic sites—and subsequent relatedness checks [34], or relatedness checks can be performed within samples from the same sequencing flowcell, or within a time window, and obviously between family members when a full trio

is sequenced for diagnostics of inherited diseases. Other sample identity checks include assessment and control of data from sex chromosomes X and Y [35] and ancestry inference.

**Quality control**

There was no consensus on which parameters to use for automated quality control (QC). All units aim for either a mean or median sequencing coverage depth of >30-fold for germline WGS and 60–90-fold for somatic WGS, but little consensus exists regarding minimal coverage depth for panels or WES (Fig. 2), also previously discussed elsewhere [3, 36–38]. Other scores used for quality assessment are Phred scores, estimated sequencing error rate (Q20 or Q30), allele depths ratio, Q pred scores, GC ratio, AT dropout and skewed allele depth ratio (<0.05). Thresholds for most of these quality scores are used in combination with a definition of a fraction of the genome that should pass the threshold in order to account for hard-to-sequence regions of the genome while still applying a stringent threshold. Several units report using soft thresholds or have intervals that do not necessarily stop the sample from being sent to the interpreting personnel but flag it with a warning. The tools reported to calculate and collect automated QC in production are: FastQC [39], MultiQC [40], Picard [41], GATK [42], Illumina SAV [43], Qualimap [44], OmnomicsQ [45], Mosdepth [46], SamBamBa [47] and in-house tools. All units report

that manual visual inspection of variants is performed during the variant interpretation by a person skilled in the art (e.g. geneticist, biologist, molecular pathologists, variant review scientists, or medical specialist) prior to drafting the clinical report.

Coordination of quality management varies depending on the organisation or hospital that the unit operates in. In some cases, a dedicated team handles all quality-related tasks, including accreditation and audits. Such dedicated teams may not include competences in bioinformatics or computational engineering, and since the available standards have been devised for medical laboratories, this can lead the bioinformatics team to adapt to practices more suited for wet lab. Alternatively, quality tasks may be managed by bioinformatics team members who also work in the field, but may have less experience in broad quality practices. Both approaches have their caveats, and the choice is often determined more by the available resources and competences at the site, rather than design. All sites agreed that accreditation and quality work is valuable, also because it helps in maintaining a quality and safety aware culture.

#### **Version control**

A fundamental part of accreditation according to the standard is documentation of procedures. To most bioinformatics or general software developers, this is a natural part of good coding practices and rarely needs fundamental changes in working routines. Use of Git is recommended for collaborating on code as it makes it both easy and natural to enforce version control and support release management. The choice of platform is even between GitHub (5/13) and GitLab (5/13).

A broader description of the pipelines, the tools and the components, as well as test results, is necessary in addition to the code and user documentation. Versions of the software and pipelines, for example, must be clearly documented and accessible, not obscured within git tags, and available to anyone receiving output from the pipeline.

#### **Benchmarking, validation and testing of bioinformatics tools and pipelines**

##### **Test data**

An important aspect of ensuring high quality in production is the ability to quantify the results. Truth sets for test and validation comprise a set of sequencing data with the corresponding target—e.g. variants or diagnostics score—that is the desired result of an analysis. These should be designed to be representative of clinical samples and variants to be analyzed, and data should preferably have been sequenced in-house, to account for local technical biases. For common truth sets, the biological

material can be acquired for this purpose. There is recognition of the over-reliance on the major public truth sets from the Genome in a Bottle Consortium (GIAB) [17, 30, 48–51] and the Sequencing Quality Control 2 (SEQC2) [52], which are often used both for training tools and for validation. This dual use introduces double dipping and highlights the need for additional local truth sets to avoid overfitting and ensure robust performance, even if these additions are more difficult to work with. A dedicated test set to identify variants in hard-to-call clinically relevant genes [17] can make the test more relevant and realistic and add variants not present in the standard GIAB truth set. There are also multiple External Quality Assessment (EQA) programmes [53] with the same aim, albeit often in a blinded validation that does not lend itself to further testing. All units use, in some form, previous samples to test and validate new pipeline releases, knowing that it carries a clear bias towards recall at the expense of precision. The most treasured recall sets are resequencings of samples where the target has been identified with orthogonal methods like SNP arrays, Sanger sequencing, or Multiplex Ligation-dependent probe amplification (MLPA) or even manual “eye-balling” of sequencing reads because these cases alleviate the known and potentially monumental recall bias towards the approaches or algorithms used for generating the target in the truth.

Use of synthetically introduced variants using BAMsurgeon [54] has previously been tested within NACG as a communal quality assessment effort [55] but no unit uses this strategy routinely. One concern that arose from the BAMsurgeon test by NACG is that synthetically introduced variants fail to trigger thresholds that depend on the neighbourhood of the variant, e.g. identification of active regions, local re-alignment, or scores that comprise a genomic-spatial effect including read- and mapping-quality, as well as unknown technical artefacts including editing the distance between the sample and the reference genome. Nonetheless, such tools could be promising for cost-effective and robust validation of bioinformatics pipelines, as also highlighted in recent guidelines [56].

For testing purposes several facilities use reduced datasets, e.g. data from only select chromosomes or otherwise truncated raw data, to enable quick, complete pipeline runs during development (Additional file 1: Table S2, Additional file 3: Fig. S2). Tools and pipelines for WGS can have a total CPU cost of several days or weeks, which makes it cost prohibitive to run structured testing of many tools using many samples. Towards the same goal one unit has constructed a “super-sick” sample by constructing a genome from several clinical samples, in which a large number of variants can be tested in a single run.

**Table 2** Nomenclature of quality assessment and optimisation in development of tools and pipelines

---

**Benchmarking** systematically evaluates tools and methods against standardised test sets and criteria, focusing on performance metrics and suitability. **Verification** ensures that methods and tools meet design specifications. **Validation** confirms their reliability and accuracy in practical applications, addressing challenges such as genetic variation complexity and sequencing biases. **Testing** involves applying criteria to assess functionality, while **success criteria** define the acceptable standards for performance, accuracy and reliability

---

During benchmarking and validation of new releases of a bioinformatics pipeline, natural attention is given to the detection of clinically relevant variants. It is acknowledged that a pipeline release may fail to capture clinically important variants in a recall test of previously detected variants, and the decision to accept this depends on the overall improvement in pipeline performance. Additionally, overall recall rates for variant detection may decline upon updating the truth sets to newer and more comprehensive versions. Most bioinformatics pipelines running on a compute cluster will yield minor differences in repeated runs due to stochastic variability, and this should be accounted for when assessing pipeline performance, just as robustness tests should be performed (Table 2).

#### **Release management**

Tools should be evaluated based on speed, cost, user acceptance and performance metrics like *F*-scores, precision and accuracy. Success criteria for pipeline changes need to be defined in advance to ensure changes achieve intended results. Version control and software release management are pivotal to ensure transparency and reproducibility of validation results. There are different structures of management of these processes in the different teams, but all teams report to work in close collaboration with clinical personnel to devise and improve the pipeline. Changes to the pipeline are thus requested, scoped and devised in a multidisciplinary dialogue with inclusion of all relevant competences and professions, and neither the users of the results nor the teams responsible for producing them are passive or act merely on external requests.

#### **Pipeline testing**

After scoping a change request or addition to the pipeline and benchmarking a broad selection of tools and solutions, including one or more rounds of test and validation, a release candidate can be devised. Since bioinformatic pipelines consist of a number of successive processing steps with each step depending on the previous one, errors and imprecisions will propagate. Additionally, many practical aspects including data annotation, formatting conventions, file naming, or directory hierarchy may greatly influence the compatibility of tools within a pipeline. Therefore, testing of pipelines involves

a combination of testing of individual components, integration and end-to-end testing of the full pipeline. Below is an outline of typical steps that are needed, alongside biological and clinical understanding and evaluation.

*Unit testing* typically involves use of minimal datasets to allow rapid assessment of specific components within the pipeline. Synthetic data may be used to test targeted functionalities, ensuring that each component operates as expected. This approach is particularly useful when integrating new tools, as it allows for focused evaluation before full pipeline deployment and a way to perform structured benchmarking of relevant tools. At this point in the development process, there would often exist a “code freeze” or “release candidate” with a named git commit to be further evaluated. Any new code in the release candidate should be subject to manual code review.

*Integration testing* involves running the entire pipeline with more extensive datasets to ensure comprehensive functionality and performance. It is recommended to use both minimal and larger datasets during this phase. For instance, a typical practice is to run patient samples through the pipeline to verify it meets predefined success criteria. Some groups also employ test-driven development methodologies, where tests are defined before making changes to the pipeline, ensuring that modifications achieve the desired outcomes.

*System testing and end-to-end testing* is performed at most sites (10/13) and includes verification and validation of the consequences on overall performance of the pipeline using representative datasets.

*IT performance testing* is the evaluation of the system performance and responsiveness under different workloads and scenarios. It is often performed as stress tests, e.g. by concurrently starting many analyses pipelines. Additionally, security impact should be assessed and tested if specifically relevant to a release.

When to perform which test depends on the scope of the addition or change. As a general rule, we suggest at major releases to perform end-to-end testing, system testing, performance and usability testing and at minor releases to perform unit and integration testing, followed by regression and end-to-end testing. It will, however, depend on the change scenario and the internal definition of major or minor releases. The latter is often influenced by the assessment of the needed scale of testing.

**Table 3** Key quality aspects in clinical bioinformatics in production

---

**Use of established tools:** Utilise well-established and validated tools and algorithms  
**Accuracy:** Validate pipelines and implement quality control measures to detect and correct errors and assess the performance  
**Reproducibility:** Ensure pipelines produce consistent results when run on the same data  
**Documentation and transparency:** Provide thorough documentation and transparency for understanding and troubleshooting  
**Specific considerations for different pipelines:** Address unique safety and quality requirements that are different e.g. for personalised treatment identification in cancer and rare disease diagnostics  
**Competence:** Employ experienced bioinformaticians to develop, review, upgrade and maintain pipelines  
**Infrastructure:** Use suitable computational infrastructure to ensure scalability and reliability  
**External controls:** Participate in external quality assessment programs to benchmark pipeline performance

---

**Transition from validation to production**

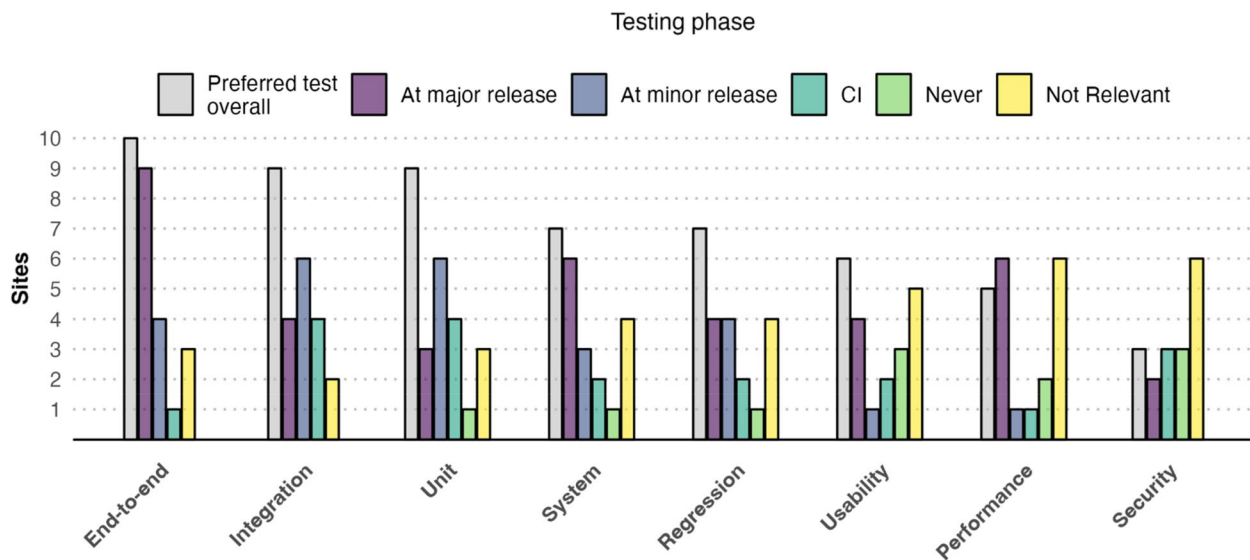
Transitioning from validation to clinical production requires careful management to ensure that the pipeline is ready for routine use. There is no consensus on whether to continue using minimal datasets or to employ larger, more representative datasets for this transition. One approach involves running actual patient samples and confirming that results align with expected results from previous pipeline versions or orthogonal testing. Manual interpretation is often still required in many cases, although efforts are being made to develop automated validation processes to reduce the workload on clinical staff. The workload associated with performing the manual validation is recognised as substantial—also to personnel outside of the bioinformatics team—and there

is a clear need to automate as much of the process as possible to enhance efficiency. Nonetheless, manual review of output by interpretation specialists remains a critical component, particularly in cases where automated validation does not cover all aspects of pipeline performance. One unit reports the use of Continuous Integration (CI) testing, where automated test scenarios are run upon each commit, merge and release (Table 3, Fig. 3).

**IT management**

**High performance cluster**

All teams agree that direct access to a high performance cluster (HPC) is recommended. Local HPCs can be preferable for security and availability and ease of systems integration. Administration of such a cluster is



**Fig. 3** Reported use of test concepts to ensure quality of pipelines. The typical types of tests and test phases are shown on a polar barplot with the following notation for testing: Unit, testing individual components or functions in isolation; Integration, testing the interaction and integration between multiple components or modules; System, testing the entire system as a whole to ensure it meets the specified requirements; Regression, re-testing previously working functionality to ensure that changes have not introduced new issues; End-to-end, verifying the flow and functionality of the entire system from start to finish; Performance, evaluating the system’s performance and responsiveness under different workloads and scenarios; Security, identifying vulnerabilities and ensuring the system’s resistance against potential security threats; Usability, assessing the system’s user-friendliness and ease of use. The preferred test type overall is shown in grey. Counts on the barplot indicate the number of sites using each test type at each stage

a specialised task that requires IT competencies that are not a priori part of a bioinformatician's training. All teams have some tasks in terms of IT management, but there are large differences on whether the main responsibility and workload are inside or outside of the team. Genomic information is, per definition, person identifiable and sensitive information, and nearly all sites have a closed infrastructure not connected to the internet, exactly for security reasons.

#### **Software containerisation and collaboration**

One recommended route to robust production is to encapsulate software in containers (in production at respondents are Docker [57] and Apptainer/Singularity [58]), or alternatively, to predefine environments (in production at respondents are, Conda [59] and EasyBuild [60]) that the code should run in. This approach greatly reduces a number of risks arising from coexisting on shared IT infrastructures, and it makes code sharing and co-development less dependent on where the code needs to run. Some have experienced sizable resource overhead from containerisation, which can be costly in a high-throughput production. The major challenge is the magnitude of data, requiring code to be tightly optimised to the hardware for efficient and cost-effective use of the cluster. This means that there are many parameters to balance at the expense of others, such as limitations on RAM, disk speed, local node disk size and read—write capacity (IO) to central storage, number of cores, and job queuing time—each of which could prove a bottleneck in one cluster, but not in another. Containerisation is not a direct solution for data security, as reviewed here [61], and does not replace the need for benchmarking [62], which is in line with our recommendations.

There are, however, very successful community developments of bioinformatics pipelines that are containerised with a high level of platform adaptability in order to be less sensitive to the computing environment. Notable is the popular nf-core [63] to which several of the units are contributing. Nf-core features an organised release structure, code review and continuous integration tests. Strategic use of multicenter collaborations helps prevent loss of competence when staff leave and enhances safety and quality by involving multiple sites, each contributing distinct competencies and experiences.

#### **IT security**

Security is often not specific to a pipeline release but is a continuous effort supported by dedicated IT-security teams in the organisation, where there are also competences to perform Data Protection Impact Assessment (DPIA) when needed. Two-factor authentication (2FA)

for server/GitHub access is now considered a minimum standard to improve security and prevent unauthorised access. Most sites use an air-gapped HPC (without direct access to the internet) which is a general recommendation, as it eliminates a sizable risk factor.

#### **Pipeline orchestration**

Most teams use a workflow manager; most popular are Nextflow [64] (4/8) or Snakemake [65] (3/8), but also Cromwell [66] and StackStorm [67], and job schedulers like Crontab and Dagu [68] are in production. A workflow manager allows for orchestrating parallelised workflows on a computer cluster where a bioinformatics pipeline can consist of multiple parts, each with different resource requirements and dependencies. Exceptions to this are two teams running commercial pipelines: one unit is using Nextflow to orchestrate jobs to Illumina DRAGEN hardware [69] that can only process serially, and two teams running the cloud-based analysis service SOPHiA DDM [70] by SOPHiA GENETICS.

#### **Data and metadata management**

Field standards for managing sequencing data and metadata in genomics informatics include ISO/TS 23357:2023 [71], which outlines data fields from sequence reads to variant evaluation, and ISO/TS 20428:2024 [72], which provides a template for clinical sequencing reports in electronic health record (EHR) systems. The HL7 FHIR [73] standard facilitates effective data integration and transmission. Best practices involve using standardised formats, comprehensive metadata, and systems that enable smooth data exchange. Adhering to FAIR principles (Findable, Accessible, Interoperable, Reusable) and structured EHR templates is recommended for maintaining data quality and interoperability [74].

Within NACG there is a high diversity in methods for approaching data and metadata management, including flat files, sample sheets, gzip-compressed FASTQ files, various database technologies utilising some kind of common lab identifiers as well as clinical decision support software (Additional file 3: Fig. S3). Common practice is to use Laboratory Information Management Systems (LIMS) for these purposes. Integrations using the HL7 messaging standard are also worth mentioning (1/13).

In terms of propagation of the necessary metadata through the analytical pipelines, JSON and filenames are the most used methods (9/13 for both), closely followed by flat files (logs and other files; 7/13), database technology (e.g. SQL, PostgreSQL; 6/13), YAML (5/13), environmental variables (4/13) and Nextflow channels (1/13).

Most units use pseudonymisation during data handling to minimise availability of personal information.

There is no consensus on which files can be deleted from the production output, aside from SAM and BAM, that are now replaced by the compressed format CRAM. Some facilities save raw output from the machine (from Illumina these can have the file extension.bcl), whereas others discard these files after processing.

### Staff and competences for clinical production

The success of a clinical bioinformatics unit is highly dependent on the staff and their competencies. Key tasks include building data flows, setting up bioinformatic pipelines, developing front and back-end applications, troubleshooting production runs, interpreting result quality and ensuring regulatory compliance [75].

The roles required vary significantly depending on the operation's size and services offered. Smaller sites need staff with a broad skill set in order to fill the roles required, while larger facilities can afford more specialised positions. Essential competencies include software development, data management, quality assurance and domain knowledge in genetics. At larger sites with separate teams dedicated to system development, bioinformatic workflows or quality assurance, deep domain knowledge may be less critical for some positions, as teams can rely on each other's expertise to fill the gap (Additional file 3: Fig. S4).

All sites agree that a highly automated workflow is preferable. Ideally, bioinformaticians should only need to troubleshoot results, with the system handling standard cases autonomously, from initiating workflows to delivering results. Automation enables staff to focus on more complex tasks and further refine the workflows and systems. Minimising repetitive tasks also helps keep staff engaged and motivated, which is crucial for retention.

Recruitment of new personnel is a challenge for most sites (12/13). The skills needed are often in high demand, making it difficult to find the right candidates. Traditionally, recruitment has been done through academic networks. However, as more specialised roles are required, it is necessary to look beyond these networks. Skilled front and backend developers, as well as system architects, are often not found in academic settings, making it necessary to advertise more broadly and possibly engage external recruiters. While most sites cannot match the salaries of the private sector, NACG sites offer the chance to be part of a larger mission to the benefit of society, work with cutting-edge technology and participate in innovative research. Additionally, offering remote work options, flexible hours and opportunities for internal training and career progression are strong motivators for many employees.

## Conclusions

Here we present 16 consensus recommendations for clinical bioinformatics in production, developed by bioinformaticians actively working in development and operation. The recommendations are specific and practical, and in production at least two sites. We furthermore report on the current practice at the 13 participating clinical bioinformatics units, members of The Nordic Alliance for Clinical Genomics (NACG).

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01543-4>.

Additional file 1. Table S1. Tools used in clinical production.

Additional file 2. Table S2. Datasets used to validate and benchmark (Summary list across all types of analysis).

Additional file 3. Figure S1. Summary of provided analyses across units. Figure S2. Data types used to validate and benchmark performance across tools. Figure S3. Data and meta-data management. Figure S4. Competences across NACG sites.

### Acknowledgements

We thank the Nordic Alliance for Clinical Genomics (NACG) members and workshop participants for their valuable contributions and insights.

### Authors' contributions

F.O.B. conceived the study; F.O.B., K.L. and E.S.E. designed the study. F.O.B., K.L., E.S.E., R.L.M., A.J. and J.M.V. drafted the manuscript. All other authors appear in the author list ordered according to the alphabet. All authors participated in data collection and data analysis. All authors actively read, edited and approved the manuscript.

### Funding

Open access funding provided by Copenhagen University Not applicable. This work received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Data availability

The survey questions and summary of results are provided in Additional file 1: Table S1. All other data generated or analysed during this study are included in the published article (and its supplementary information files).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

In the past 3 years, F.O.B. has received compensation from AstraZeneca and is a founder and real owner of Fobinf ApS, advising Imunitrack ApS owned by Eli Lilly, HERvolution ApS, Novo Nordisk A/S and Aida Oncology ApS. E.K. was at the time of submission employed at Blueprint Genetics Oy.

#### Author details

<sup>1</sup>Oslo University Hospital, Oslo, Norway. <sup>2</sup>Rigshospitalet, Copenhagen, Denmark. <sup>3</sup>Karolinska University Hospital, Stockholm, Sweden. <sup>4</sup>Karolinska Institutet, Stockholm, Sweden. <sup>5</sup>Haukeland University Hospital, Bergen, Norway. <sup>6</sup>Institute for Molecular Medicine Finland, Helsinki, Finland. <sup>7</sup>Landspítali University Hospital, Reykjavik, Iceland. <sup>8</sup>Turku University Hospital, Turku, Finland. <sup>9</sup>Skåne University Hospital, Malmö, Sweden. <sup>10</sup>Helsinki University Hospital,

Helsinki, Finland. <sup>11</sup>University of Helsinki, Helsinki, Finland. <sup>12</sup>University Medical Center Groningen, Groningen, Netherlands. <sup>13</sup>University Hospital of Umeå, Umeå, Sweden. <sup>14</sup>KTH Royal Institute of Technology, Stockholm, Sweden. <sup>15</sup>Olav's University Hospital, Trondheim, Norway. <sup>16</sup>Aarhus University Hospital, Aarhus, Denmark.

Received: 16 October 2024 Accepted: 29 August 2025

Published online: 17 October 2025

## References

- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–24.
- Association for Clinical Genomic Science. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020. 2020.
- Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn*. 2018;20(1):4–27.
- Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheatle-Jarvela AM, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N Engl J Med*. 2018;379(22):2131–9.
- Lindstrand A, Eisfeldt J, Pettersson M, Carvalho CMB, Kvarnung M, Grigelioniene G, et al. From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med*. 2019;11(1):68.
- Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med*. 2021;385(20):1868–80.
- Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med*. 2021;13(1):40.
- Sosinsky A, Ambrose J, Cross W, Turnbull C, Henderson S, Jones L, et al. Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 genomes cancer programme. *Nat Med*. 2024;30(1):279–89.
- Hodder A, Leiter SM, Kennedy J, Addy D, Ahmed M, Ajithkumar T, et al. Benefits for children with suspected cancer from routine whole-genome sequencing. *Nat Med*. 2024;30(7):1905–12.
- Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodama M, et al. Whole genome sequencing in clinical practice. *BMC Med Genomics*. 2024;17(1):39.
- Chicco D, Jurman G. Ten simple rules for providing bioinformatics support within a hospital. *Biodata Min*. 2023;16(1):6.
- Eisenstein M. Illumina faces short-read rivals. *Nat Biotechnol*. 2023;41(1):3–5.
- The Nordic Alliance for Clinical Genomics. NACG 14th workshop agenda. 2023. [https://nordicclinicalgenomics.org/assets/main/resources/nacg\\_ws14\\_agenda\\_vers140923-1694979857.pdf](https://nordicclinicalgenomics.org/assets/main/resources/nacg_ws14_agenda_vers140923-1694979857.pdf).
- Kobren SN, Baldrige D, Velinder M, Krier JB, LeBlanc K, Esteves C, et al. Commonalities across computational workflows for uncovering explanatory variants in undiagnosed cases. *Genet Med*. 2021;23(6):1075–85.
- Li H, Dawood M, Khayat MM, Farek JR, Jhangiani SN, Khan ZM, et al. Exome variant discrepancies due to reference-genome differences. *Am J Hum Genet*. 2021;108(7):1239–50.
- gnomAD v4.0. 2023. <https://gnomad.broadinstitute.org/downloads#v4>.
- Wagner J, Olson ND, Harris L, McDaniel J, Cheng H, Functammanan A, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol*. 2022;40(5):672–80.
- Pan B, Kusko R, Xiao W, Zheng Y, Liu Z, Xiao C, et al. Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinformatics*. 2019;20(Suppl 2):101.
- Illumina Inc. DRAGEN TruSight Oncology 500 Analysis Software v2.1 2023. <https://support.illumina.com/downloads/trusight-oncology-500-analysis-software-v2-1-documentation.html>.
- Gabrielaite M, Torp MH, Rasmussen MS, Andreu-Sanchez S, Vieira FG, Pedersen CB, et al. A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers (Basel)*. 2021. <https://doi.org/10.3390/cancers13246283>.
- Yuan N, Jia P. Comprehensive assessment of long-read sequencing platforms and calling algorithms for detection of copy number variation. *Brief Bioinform*. 2024. <https://doi.org/10.1093/bib/bbae441>.
- Popic V, Rohlicek C, Cunial F, Hajirasouliha I, Meleshko D, Garimella K, et al. Cue: a deep-learning framework for structural variant discovery and genotyping. *Nat Methods*. 2023;20(4):559–68.
- Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Res*. 2017;6:664.
- Eisfeldt J, Martensson G, Ameer A, Nilsson D, Lindstrand A. Discovery of novel sequences in 1,000 Swedish genomes. *Mol Biol Evol*. 2020;37(1):18–30.
- Eisfeldt J. SVDB: Structural Variant DataBase software. 2019.
- English AC, Menon VK, Gibbs RA, Metcalf GA, Sedlazeck FJ. Truvari: refined structural variant comparison preserves allelic diversity. *Genome Biol*. 2022;23(1):271.
- Standardization IOF. Medical laboratories - Requirements for quality and competence. Geneva: International Organization for Standardization; 2012.
- Standardization IOF. General requirements for the competence of testing and calibration laboratories. Geneva: International Organization for Standardization; 2017.
- Rehm HL, Page AJH, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. *Cell Genom*. 2021;1(2).
- Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nat Biotechnol*. 2019;37(5):555–60.
- Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, (2017).
- MDCG 2023–1 Guidance on the health institution exemption under Article 5(5) of Regulation (EU) 2017/746 on in vitro diagnostic medical devices, (2023).
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), Recital 34, (2016).
- Pedersen BS, Bhetariya PJ, Brown J, Kravitz SN, Marth G, Jensen RL, et al. Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. *Genome Med*. 2020;12(1):62.
- Liu S, Zeng Y, Wang C, Zhang Q, Chen M, Wang X, et al. seGMM: a new tool for gender determination from massively parallel sequencing data. *Front Genet*. 2022;13:850804.
- Rehder C, Bean LJH, Bick D, Chao E, Chung W, Das S, et al. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23(8):1399–415.
- Marshall CR, Chowdhury S, Taft RJ, Lebo MS, Buchan JG, Harrison SM, et al. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *NPJ Genom Med*. 2020;5:47.
- Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2017;19(3):341–65.
- Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Institute / Bioinformatics: University of Cambridge; 2010.
- Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
- Broad Institute. Picard Toolkit. Broad Institute; 2019.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernyt-sky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
- Illumina Inc. Illumina Sequencing Analysis Viewer (SAV). Illumina; n.d.

44. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics*. 2012;28(20):2678–9.
45. Oy E. OmnomicsQ: NGS Data Quality Management and Validation Software. Euformatics Oy; n.d.
46. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34(5):867–8.
47. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032–4.
48. Jarvis ED, Formenti G, Rhie A, Guaracino A, Yang C, Wood J, et al. Semi-automated assembly of high-quality diploid human reference genomes. *Nature*. 2022;611(7936):519–31.
49. Zook JM, Hansen NF, Olson ND, Chapman L, Mullikin JC, Xiao C, et al. A robust benchmark for detection of germline large deletions and insertions. *Nat Biotechnol*. 2020;38(11):1347–55.
50. Zook JM, McDaniel J, Olson ND, Wagner J, Parikh H, Heaton H, et al. An open resource for accurately benchmarking small variant and reference calls. *Nat Biotechnol*. 2019;37(5):561–6.
51. Berner LT, Law BE. Plant traits, productivity, biomass and soil properties from forest sites in the Pacific Northwest, 1999–2014. *Sci Data*. 2016;3:160002.
52. Fang LT, Zhu B, Zhao Y, Chen W, Yang Z, Kerrigan L, et al. Establishing community reference samples, data and call sets for benchmarking cancer mutation detection using whole-genome sequencing. *Nat Biotechnol*. 2021;39(9):1151–60.
53. Hastings RJ, Howell RT. The importance and value of EQA for diagnostic genetic laboratories. *J Community Genet*. 2010;1(1):11–7.
54. Ewing AD, Houlihan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods*. 2015;12(7):623–30.
55. The Nordic Alliance for Clinical Genomics. NACG 10th Workshop report. 2021.
56. Duncavage EJ, Coleman JF, de Baca ME, Kadri S, Leon A, Routbort M, et al. Recommendations for the use of in silico approaches for next-generation sequencing bioinformatic pipeline validation: a joint report of the Association for Molecular Pathology, Association for Pathology Informatics, and College of American Pathologists. *J Mol Diagn*. 2023;25(1):3–16.
57. Docker DM. Lightweight Linux containers for consistent development and deployment. *Linux J*. 2014;239:2.
58. Kurtzer GM, Sochat V, Bauer MW. Singularity: scientific containers for mobility of compute. *PLoS ONE*. 2017;12(5):e0177459.
59. Anaconda Inc. Anaconda Software Distribution. Anaconda Inc.; n.d.
60. EasyBuild community. EasyBuild - building software with ease. EasyBuild community; n.d.
61. Kadri S, Sboner A, Sigaras A, Roy S. Containers in bioinformatics: applications, practical considerations, and best practices in molecular pathology. *J Mol Diagn*. 2022;24(5):442–54.
62. Brooks TG, Lahens NF, Mrcela A, Grant GR. Challenges and best practices in omics benchmarking. *Nat Rev Genet*. 2024;25(5):326–39.
63. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol*. 2020;38(3):276–8.
64. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35(4):316–9.
65. Koster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520–2.
66. Broad Institute. Cromwell: Scientific workflow engine for genomics. Broad Institute; n.d.
67. StackStorm Community. StackStorm: Event-driven automation platform. Linux Foundation; n.d.
68. dagu-org (community). Dagu: Lightweight and Powerful Workflow Engine. GitHub; n.d.
69. Illumina Inc. DRAGEN secondary analysis. Illumina; n.d.
70. SOPHiA GENETICS. SOPHiA DDM: Data-Driven Medicine Platform. SOPHiA GENETICS; n.d.
71. International Organization for Standardization. Clinical genomics data sharing specification for next-generation sequencing. *Genomics informatics*. Geneva, Switzerland. 2023.
72. International Organization for Standardization. Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records. *Genomics Informatics*. 2024.
73. HL7 FHIR Foundation. HL7 Standards - FHIR Specification. 2016.
74. Ryu B, Shin SY, Baek RM, Kim JW, Heo E, Kang I, et al. Clinical genomic sequencing reports in electronic health record systems based on international standards: implementation study. *J Med Internet Res*. 2020;22(8):e15040.
75. Competency Hub: Clinical bioinformatics professionals. 2022. <https://competency.ebi.ac.uk/framework/nhs/1.0>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.