The Belgian next generation sequencing guidelines for haematological and solid tumours

PARTICIPANTS

Adamis	Stavroula	Université catholique de Louvain
Aftimos	Philippe	Institut Jules Bordet Brussels
Antoine-Poirel	Hélène	Fondation Registre du Cancer
Antoniou	Aline	Sciensano Brussels
Bakkus	Marleen	Universitair Ziekenhuis Brussel
Bols	Alain	AZ Sint-Jan Brugge-Oostende
Boone	Elke	AZ Delta Roeselare
Bourgain	Claire	Imelda Bonheiden
Breems	Dimitri	Ziekenhuis Netwerk Antwerpen
Brouwers	Barbara	AZ Sint-Jan Brugge-Oostende
Caers	Joseph	Centre hospitalier universitaire de Liège
Caljon	Ben	Universitair ziekenhuis Brussel
Canon	Jean-Luc	Grand Hopital de Charleroi
Carestia	Luciano	Clinique Saint-Pierre Ottignies
Chintinne	Marie	Institut Jules Bordet Brussels
Claes	Kathleen	Universitair Ziekenhuis Gent
Cleemput	Hugo	Emmaus hospital
Collignon	Joelle	Centre hospitalier universitaire de Liège
Criel	Marieke	Imelda Bonheiden
Cuppens	Kristof	Jessa Ziekenhuis Hasselt
De Grève	Jacques	Universitair Ziekenhuis Brussel
De Schouwer	Pieter	Ziekenhuis Netwerk Antwerpen
Debaere	Elfride	Universiteit Ziekenhuis Gent
Deeren	Dries	AZ Delta Roeselare
Defour	Jean-Philippe	Université catholique de Louvain
Dekairelle	Anne-France	Université catholique de Louvain
De Keersmaecker	Sigrid	Sciensano Brussels
Dequeker	Els	Universitair Ziekenhuis Leuven
Demanet	Christian	Universitair ziekenhuis Brussel
Demarchin	Jerome	Centre hospitalier régional de la Citadelle
Demedts	Ingel	AZ Delta Roeselare
Denys	Barbara	Universitair Ziekenhuis Gent
Devos	Helena	AZ Sint-Jan Brugge-Oostende

Dewaele	Stefanie	AZ Sint-Lucas Gent
Dewispelaere	Laurent	Hopital Erasme Brussels
D'haene	Karl	Algemeen Stedelijk Ziekenhuis Aalst
El Housni	Hakim	Hopital Erasme Brussels
Frans	Johan	Imelda Bonheiden
Froyen	Guy	Jessa Ziekenhuis Hasselt
Govaerts	Elke	AZ Sint-Lucas Gent
Ghislain	Vanessa	Sciensano Brussels
Havelange	Violaine	Université catholique de Louvain
Hébrant	Aline	Sciensano Brussels
Heimann	Pierre	Hopital Erasme Brussels
Heylen	Evelien	Ziekenhuis Netwerk Antwerpen
Jacobs	Koen	AZ Sint-Lucas Gent
Kock	Marc	Histogenex Anwerpen
Koopmansch	Benjamin	Centre hospitalier universitaire de Liège
Kruse	Vibeke	Universitair Ziekenhuis Gent
Laffut	Wim	H. Hartziekenhuis Lier
Lambert	Frederic	Centre hospitalier universitaire de Liège
Lambin	Suzan	Universitair Ziekenhuis Antwerpen
Lefesvre	Pierre	Universitair ziekenhuis Brussel
Le Mercier	Marie	Universitair Ziekenhuis Antwerpen
Lenie	Sandy	Universitair Ziekenhuis Brussel
Lierman	Els	Universitair Ziekenhuis Leuven
Louagie	Henk	AZ Sint-Lucas Gent
Lybaert	Willem	AZ Nikolaas
Maes	Brigitte	Jessa Ziekenhuis Hasselt
Maes	Marie-Berthe	Universitair Ziekenhuis Antwerpen
Martens	Geert	AZ delta Roeselare
Meers	Stef	Universitair Ziekenhuis Antwerpen
Metsu	Sofie	Histogenex Antwerpen
Nollet	Friedel	AZ Sint-Jan Brugge-Oostende
Obbels	Dagmar	Imelda Bonheiden
Prenen	Hans	Universitair Ziekenhuis Leuven
Raicevic	Gordana	Sciensano Brussels
Roland	Isabelle	Hopital Erasme Brussels
Rutten	Annemie	GZA Hospital Sint-Augustinus
Salgado	Roberto	Universitair Ziekenhuis Antwerpen
Salmon	Jean-Paul	Centre hospitalier régional de la Citadelle
Saussoy	Pascale	Université catholique de Louvain
Sautois	Brieuc	Centre hospitalier universitaire de Liège
Selleslag	Dominik	AZ Sint-Jan Brugge-Oostende
Senterre	Jean-Marc	Centre hospitalier régional de la Citadelle
Sibille	Catherine	Centre hospitalier régional de la Citadelle
Sibille	Anne	Centre hospitalier universitaire de Liège

Sticca	Tiberio	Institut Jules Bordet Brussels
Spans	Lien	Universitair Ziekenhuis Leuven
Thibaut	Kristin	H.-Hartziekenhuis Lier
Vael	Carl	Algemeen Ziekenhuis Klina Brasschaat
Vaeyens	Freya	OLV Ziekenhuis Aalst
Van Campenhout	Christelle	Sciensano Brussels
van Dam	Pieter-Jan	Histogenex Antwerpen
Van den broecke	Caroline	AZ Sint-Lucas Gent
Van den Bulcke	Marc	Sciensano Brussels
Van der Meulen	Joni	Universitair Ziekenhuis Gent
Van Dooren	Sonia	Universitair Ziekenhuis Brussel
Vandepoele	Karl	Universitair Ziekenhuis Gent
Vander Borght	Sara	Universitair Ziekenhuis Leuven
Van Houdt	Jeroen	Universitair Ziekenhuis Leuven
Van Huysse	Jacques	AZ Sint-Jan Brugge-Oostende
van Laere	Steven	Universitair Ziekenhuis Antwerpen
Vanneste	Kevin	Sciensano Brussels
Vannuffel	Pascal	Institut de Pathologie et de Génétique Gosselies
Van Ongeval	Johan	AZ Sint-Lucas Gent
Van Valckenborgh	Els	Sciensano Brussels
Venken	Tine	Ziekenhuis Netwerk Antwerpen
Vermeij	Joanna	Ziekenhuis Netwerk Antwerpen
Vermeire	Stefanie	AZ Sint-Jan Brugge-Oostende
Vermeulen	Katrien	Universitair Ziekenhuis Antwerpen
Weyn	Christine	Universitair Ziekenhuis Antwerpen
Wu	Ka Lung	Ziekenhuis Netwerk Antwerpen

1. DEFINITIONS

Accreditation: procedure by which an authoritative body gives formal recognition that an organization is competent to carry out specific tasks.[1]

Allelic read percentage or allelic frequency: is the percentage of variant reads in a background of normal reads.

Analytical accuracy: measurement of the discrepancy between the measured value and the true value. Can be established by analysing well-characterised reference samples with known confirmed sequence variants.[2-4]

Analytical sensitivity: likelihood that the assay will detect the targeted sequence variations if present (true positive rate).[5]

Analytical specificity: probability that the assay will not detect a sequence variation when none are present (true negative rate).[5]

DNA library: collection of DNA fragments.

FASTQ format: is a text-based format for storing nucleotide sequence and its corresponding quality scores (encoded with a single ASCII character).

Limit of detection: is the lowest actual percentage of variants that can be consistently detected.

Log file: is a file that records events that occur in an operating system or other software runs, or messages between different users of communication software.

Pipeline: is a bioinformatics workflow management system which executes a series of computational or data manipulation steps that relate to bioinformatics and is organized so that the output of one is the input of the following.

Precision: degree of agreement between replicate measurements of the same material that can be determined by assessing the reproducibility (between-run precision, the consistency of results from the same sample under different conditions) and repeatability (within-run precision, the consistency of results from the same sample under the same condition).[5]

Referral laboratory: 'external laboratory to which a sample is submitted for examination'.[1]

Reference Materials: are well-characterised, homogeneous, stable samples with certified properties for their intended purpose.[6]

Reportable range: region of the genome for which the sequence derived by the NGS test meets the quality determined during the validation process.[5]

Region of interest (ROI): region of the genome that the NGS test claims to assess.

Targeted NGS: the sequencing of a subset of genes or regions of the genome.

Template: is the DNA strand that serves as a pattern for the generation of other molecule.

Turnaround time: 'elapsed time between two specified points through pre-examination, examination and post-examination processes'.[1]

Validation: 'confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled'.[1]

Variant: a called nucleotide that differs from the reference sequence. Verification: 'confirmation, through provision of objective evidence, that specified requirements have been fulfilled'.[1]

2. INTRODUCTION

Targeted next generation sequencing (NGS) analysis is a complex procedure including two main parts, the 'wet bench' part or lab experimental test and the 'dry bench' or bioinformatics part. Both parts are composed of many steps and for each of these steps optimal assay conditions and analysis settings must be determined.[7]

The aim of this document is to provide guidelines to facilitate and harmonise implementation, verification and validation of targeted NGS tests to detect acquired somatic mutations in DNA, in (haemato)-oncology. These guidelines aim to harmonise clinical conclusions if a sample is processed through different pipelines by different operators at different times and at different sites.[8] Actual platforms are based on different chemistries and each of them has specific parameters and test requirements. These guidelines aim to provide generic recommendations to all stakeholders (laboratories, BELAC-auditors, experts, etc.) valid independently of the platform used. They are to be considered as complementary to the International Organization for Standardization (ISO)15189 standard (medical laboratories) as well as to other inter- national NGS guidelines, for example those from EuroGentest, the Association for Clinical Genetic Science (ACGS), the Dutch Society for Clinical Genetic Laboratory Diagnostics (VKGL), the College of American Pathologists (CAP), IQN Path ASBL and the Centers for Disease Control and Prevention (CDC).[3,5,7-10]

Validation or verification, and internal quality control (IQC) procedures at the different steps in the workflow (by defining quality parameters and by measuring quality metrics) and on the whole process (by determining performance characteristics), must be performed to assure and maintain accurate test results. Moreover, periodic external quality assessment (EQA) should be performed in order to ensure that performance complies with (inter)nationally accepted performance criteria.

3. GENERAL REQUIREMENTS FOR TARGETED NGS TESTS TRACEABILITY

3.1 General documentation

Laboratories should have Standard Operating Procedures (SOPs) for all steps involved in sample preparation, wet and dry bench parts of the NGS process, as well as in the review and reporting of results according to the requirements of ISO 15189.[2,4,8] Accurate systems for tracking the software version used should be available.

3.2 Logistical documentation

As required by the ISO 15189 standard, for each test, information regarding the operation of instruments (e.g. calibration and maintenance records, log files, etc.) as well as any essential information on critical reagents (e.g. lot number, expiration data, etc.) should be recorded.

3.3 Targeted gene panel description documentation

During the test development, a precise description (at the genome and transcript level) of all specific gene hotspots that can be analysed should be available. The minimally required variants to be analysed for a specific tumour type are those which have a proven clinical utility (to define diagnosis and/or to predict response or resistance to specific cancer therapies (therapeutic) and/or to determine prognosis (patient outcome) for that specific tumour type. These variants have been established within the expert group of the 'Commission of Personalized Medicine' (ComPerMed). In addition, the NGS gene panel may also contain a limited number of gene targets with hotspots that are expected to have clinical relevance in the future, in a specific or in another tumour type. All variants which are part of the validation should be listed with the transcript accession (NM-reference) number of the gene, all exons and the specific, delineated regions that are targeted. Targeted NGS gene panels should generally contain only single nucleotide variants (SNVs) and small indels (Insertion or Deletion of bases in the DNA). At present other types of alterations (large indels, amplifications, translocations) are preferentially detected by other techniques.[8] The size up to which indels can be detected should be carefully determined and reported in the validation dossier. If alterations other than SNVs and small indels are included, a thorough validation and quality assurance should be established for each type of these reported molecular alterations.

3.4 Validation/verification report

The experimental approach, results, conclusions and any other relevant details of the validation or verification process (validation and verification chapters) should be recorded in a validation/verification report. The validation report should contain the empirically determined performance characteristics of the test (e.g. sensitivity, specificity, precision, accuracy and limit of detection) as well as critically relevant quality metrics (test development and validation chapters). The validation/verification criteria can be reported in peer-reviewed publications.[11] Any deviation from the developed and validated test should be recorded and documented according to ISO 15189 standards and revalidation/reverification (validation chapter) should be considered.[2,3,5,12]

3.5 Test run documentation

For each run, a test report should contain the values of relevant quality metrics in order to demonstrate that the reported sequence meets the quality criteria set in the validation report. A test report should also contain the sequences which are in the reportable ranges.

3.6 Data storage

The FASTQ (in a compressed manner) or BAM files (Binary version of a SAM file), the Variant Call Format (VCF) (>v4.0) and the final clinical report which interprets the clinically relevant somatic variants should be stored in the laboratory for a time period as legally required.[2,7-9,12,13] These files should be stored together with the traceability documents, as in some circumstances it may help in explaining the results.[7,9,14]

3.7 Data transfer

All data transfers should use secure network connections that allow verification of the data transfers. An external hard disk between the various components of the computing hardware i.e. from sequencer to the analytical computer and/or to storage location can also be used.[2] The policy and procedure should be adequately documented.

Appropriate and validated measures should be taken to avoid data corruption during transfer (e.g. by using checksum generation during file transfer, management of data permissions, secured backup of copies of FASTQ files maintained elsewhere).[2] Appropriate error messages should be generated where case corrupted files are detected.

3.8 Reference material

Reference materials can exist or variant-engineered human cell lines (Reference sample) as well as informatics data files (Reference informatics data file). Reference materials should be used for test validation/verification prior to implementation and for continued quality assessment of the validated NGS pipeline (validation, verification and quality control chapters).[5]
Reference samples should have well-documented sequencing data, should ideally be available on a continuous basis and should resemble as much as possible the patient specimens in order to accurately reflect the testing conditions.[2,5,12]
Reference informatics data files are files created by computational methods simulating patient sample sequences or by sequencing biological samples with well-documented variants (SNVs and indels) and allelic frequency close to real data.[2,5,12] They should be compatible with the sequencing platform's output taking into account the used platform characteristics such as read length, read number, etc. These reference informatics data files can be used for the validation of the dry bench part.[5]
A combination of reference informatics data files and reference samples is recommended to provide a robust framework for test validation/verification (validation, verification and quality control chapters).[5]

3.9 Risk analysis

The ISO 15189 standard requires that the laboratory evaluate the impact of work processes and potential failures on examination results as they affect patient safety, and that the laboratory modify processes to reduce or eliminate the identified risks, and document decisions and actions taken.[1] As NGS workflows are

complex and consist of many different steps (from pre-analytical step to tertiary interpretation), risk analysis is particularly appropriate to reduce potential erroneous results and should be performed prior to implementation, e.g. as part of the validation process. Any identified risk should be included in the validation report (e.g. in a Fishbone diagram), and addressed appropriately within the validation of the test.

3.10 Outsourcing

If tests are outsourced, the ISO 15189 requires that the referring laboratory has a procedure for the selection and evaluation of the referral laboratory. The referral laboratory should be accredited according to ISO 15189 for the NGS test in (haemato)-oncology and licensed by the Minister of Public Health. In addition, the quality of the referral laboratory should be continuously monitored by the referring laboratory. Responsibilities to- wards the interpretation and reporting of the results stay with the referring laboratory.[1]

Specifically for NGS tests, there is a tendency to outsource only parts of the tests that may not necessarily comply with present license requirements. In any case, however, in agreement with the ISO 15189 standard, for any outsourced parts of the activity, the referring laboratory should be able to monitor the quality of the subcontractor, and demonstrate that outsourcing does not negatively influence the reliability of the final results. The outsourcing of parts of the NGS pipeline should be subject to the required risk analysis.

4. TECHNICAL REQUIREMENTS AND PERFORMANCE SPECIFICATION

It is necessary to prove that the test performs to the highest achievable level of performance required for answering to a particular clinical question and that this level of performance is maintained in all routine analyses.[11] Desired performance characteristics for performing any NGS in (haemato)-oncology should be defined in advance and integrated in the validation plan.

The level of validation/verification depends on the availability of acceptable performance specifications (*Figure 1*):

- in case of a new in-house or modified IVD CE-marked test or technology, the entire process should be validated (validation chapter) for meeting the a priori defined performance specifications.
- in case of the implementation of a IVD-CE marked test or technology with documented performance specifications or of a validated test with minor modification of the experimental protocol or of the composition of the gene panel (e.g. when adding a new gene), only a verification (verification chapter) is required.

FIGURE 1 Workflow for the implementation of a NGS test.

5. SAMPLE ACCEPTANCE CRITERIA AND SAMPLE PREPARATION

5.1 Sampling and fixation

Sample preparation is a crucial step for all high-quality molecular analyses. Samples of poor quality (e.g. due to fixation artefacts) or with insufficient quantity of amplifiable DNA can significantly affect the sensitivity and specificity of the test (validation chapter) and can lead to false negative or false positive results.[2]

For tumour tissue, the sample fixation after surgical removal should be validated for the purpose of NGS sequence analysis. It is recommended to proceed to the fixation within 1h after surgical removal preferably in 10% neutral buffered formalin during a specific timelapse (generally between 6-72h). Delayed or suboptimal fixation may result in DNA degradation due to apoptosis and/or necrosis.[15] For blood and bone marrow samples, recipients with common anticoagulants such as EDTA

should be used.[16] NGS on decalcified bone samples is possible if a weak acid- or EDTA-based decalcification protocol has been properly administered, though it may affect the success rate of the NGS test.[17,18]

5.2 Specimen selection quality

The sample (tissue from primary tumour or metastasis, blood, bone marrow, etc.) should contain sufficient neoplastic cells. The minimally required percentage of neoplastic cells in a background of normal cells should be defined during the validation process, taking into account that the assessment of the tumour cell content by different pathologists can be imprecise and inaccurate.[19] The latter will be estimated on Haematoxylin and Eosin (HE) stained adjacent slides for biopsies and by cytological and/or flow cytometric examination for blood/bone marrow by adequately trained staff with documented competence. For formalin fixed paraffin embedded (FFPE) material, macrodissection or (manual) micro- dissection may be performed to enrich the neoplastic cell proportion.[1,7] The estimated neoplastic cell content of the material used for the DNA extraction should be reported and taken into account in the technical validation of the (negative) results.

In case neoplastic cell content is below the minimally required percentage as determined in the validation re- port, laboratories should state in the clinical report that negative results (e.g. the absence of any mutation) may not be reliable and that repeated analysis on material with a sufficiently high neoplastic cell content is advised. The treating physician should receive this information as soon as possible in order to allow for repeated sampling with as little delay as possible. A registration of transfer/receipt of this information should be kept.

5.3 DNA quality and quantity

After DNA extraction, DNA quality (e.g. purity, degradation) and/or quantity can be assessed by fluorescence, by optical density or by qPCR amplification. The laboratories should determine a measurement method for DNA quality and/or quantity and should evaluate how DNA quantity and/or quality influence the reliability of the results during the validation process (validation chapter). The minimal amount of specimen required and the success rate of the pipeline considering the DNA quantity/quality of the representative clinical samples should be evaluated during the validation process.

6. TEST DEVELOPMENT

6.1 General

For each NGS test or technology not yet implemented, usually no performance specifications are available and a protocol (SOP) should be established. Desired performance characteristics should be adequately defined in advance and integrated in the development and validation plan. A formally validated NGS-assay should be finalised before the implementation in routine practice. The laboratories should determine the optimal conditions for the test in order to meet the predefined minimal performance requirements (validation chapter). Iterative cycles should be performed until all assay conditions and analysis settings (test development chapter) meet the minimal predefined performance requirements. In case the latter cannot be reached, the lab should redetermine the minimal performance requirements and restart the procedure.

At each step of the test, critical parameters, quality metrics, their thresholds and their acceptable ranges should be defined. This development step will allow:

1. gaining the necessary experience with the test by identifying any critical step, parameter and quality metrics that may affect performance of the test,
2. interrupting the run prematurely before completion if significant deviations from the acceptance criteria are detected or if quality metric thresholds are not achieved, and
3. ensuring reliable NGS test results.[5,11] As quality metrics may vary between or even within laboratories, depending on the different platforms applied, each laboratory should establish its own quality metric thresholds.[2]

6.2 Wet bench part of the NGS process: DNA library preparation

There are different methods to prepare the DNA library
for a targeted NGS analysis.[20] Each of them is a succession of a number of the following critical steps, depending on the assay used:

• Fragmentation

At first, DNA will be fragmented into DNA fragments of an optimal length determined by the downstream platform. Input mass of DNA and fragmentation conditions should be determined.
At the end of the fragmentation, two main quality metrics have to be measured and documented: (a) the size distribution of fragmented DNA samples and (b) the amount of fragmented DNA sample.[2,3] Both can be assessed by fragment analyser, spectrophotometric readings, gel image or real-time qPCR or similar instruments.

• Target enrichment
   a. Hybridization target capture allows the enrichment of the library with targeted regions. Biased capture can occur (e.g. in poor DNA quality sample) which is especially critical for equal representation of the different barcoded samples if pooled during DNA library preparation.[2,12]
   b. Enrichment can also be achieved by amplification-based methods. Fragments are generated by PCR with primers targeting specific regions.

In general, hybridisation-based target capture is generally less sensitive but generates less false positives than PCR-based enrichments.[21]

- Adapter and barcoding ligation

Platform-specific adapter sequences and sequencing primers will be ligated to both ends of the DNA. If different DNA samples are pooled, barcodes are added to enable individual sample identification and to extract sequences obtained from a particular patient sample from the total data set. These barcodes consist of a unique DNA sequence having at least 1 bp difference between each other, typically three or more.[9] If every run contains the same targets, different barcode indexes should be used between consecutive runs in order to avoid sample leakage.[2] The number of samples that can be pooled should be determined during validation (validation chapter) and depends on the desired coverage read depth of the regions to be sequenced.[5,12]

- Tagmentation

The tagmentation method prepares genomic DNA libraries by using a transposase enzyme to simultaneously fragment and tag DNA by adding specific adapters to both ends of the fragments. These adapter sequences will amplify the insert DNA by PCR which adds index barcodes sequences.
Different parameters are critical in the tagmentation method:
- The amount of DNA input: ratio of transposase complexes to sample DNA is critical in order to obtain transposition events separated by the appropriate distances.
- All reaction parameters, such as temperatures and reaction time, must be tightly controlled as the fragment size depends on the reaction efficiency.20

6.3 Normalisation

If DNA samples are pooled, normalisation should be performed in order to have equal representation of each sample. This can be achieved by diluting the DNA libraries to equal molarities based on fluorescence analysis, optical density measurement or qPCR. Alternatively, normalisation can be performed directly on beads.

6.4 Clonal amplification

Before the clonal amplification step, an optimal quantity of DNA should be determined during the test development. An accurate estimation of the purified library quantity (e.g. DNA fragments with proper ligated adaptors and indexes) is crucial to obtain the optimal clonal amplification. Depending on the method used, optional quality and quantity controls can be performed at the end of the clonal amplification.

6.5 Sequencing

During this step, the clonal amplicons or the DNA fragments are sequenced in a flow cell by generation of luminescence or fluorescence images, which are then converted into sequences. Actual commercial platforms are based on different chemistries and each of them have specific parameters and test requirements.

6.6 Dry bench part of the NGS process or bioinformatics

The bioinformatics pipeline can be divided into three analytical steps: Primary analysis, secondary analysis

and tertiary analysis. Tertiary analysis is discussed in chapter tertiary analysis.
Several bioinformatics pipelines may be evaluated during the development plan, separately and/or combined. Combination of two or more pipelines may result in a higher sensitivity and specificity than with the use of just one. This analysis should adequately be documented in the validation report.

6.7 Primary analysis

Primary analysis consists of different steps but not all of those are mandatory in this phase, since they can al- so be done later in the process; the selection of those should be defined during the validation process together with their corresponding quality metrics and thresholds:

- Base calling: the raw electronic information from the sequencer is converted into nucleotide positions, and quality scores are assigned to each base. This is performed by the instrument's algorithms. The laboratory has relatively limited control in this phase.
- Demultiplexing: when samples are pooled before the sequencing, the data must be assigned in silico to the sample of origin by comparing the index barcodes and the reference index set.
- Primer and adapter trimming: primers and adapters have to be trimmed from the obtained sequences in order to align them properly to the reference sequence and call variants.
- Low-quality base trimming: a base quality score (Q-score or Phred score) is assigned to each base that estimates the error probability for each base.[3,4,9] This is platform dependent and should be monitored during the run.[3]
- Read quality control: allows for checking whether the sequencing data is of sufficiently good quality to ensure variant calling analysis. Reads containing bases with many too low Q-scores should be removed by informatics filters before aligning to the reference sequence. If only the 3rd end of the read has low Q-scores, only this part of the read can be trimmed before alignment.[5] Too short reads after trimming should also be removed as it might cause problems during the mapping.

The outputs of the primary analysis phase are FASTQ files, which contain the succession of nucleotides corresponding to all the reads produced by the sequencer.[8,12]

6.8 Secondary analysis

This phase contains different steps. Again, not all of them are mandatory; the selection should be defined during the validation process together with their corresponding quality metrics and thresholds:

- Reads contained in the FASTQ files (often 50–400 base pairs) are aligned to the reference sequence (read mapping), with software of choice which depends on local preferences and platforms. Mapping can be performed onto the target sequences or onto the full reference genome. Mapping to the whole reference sequence is preferred as it considerably reduces false-positive variant calls despite the fact that more computation time and space are required.[3]
- When a fragmentation or tagmentation step or amplicon-based technology is performed, duplicate reads resulting from clonal amplification should be removed by using informatics filters, as their inclusion generates a risk of skewing the allelic fractions. Keeping only the one with the highest quality score is recommended.[5]
- Indels should be evaluated on sufficient samples with (large > 15 bp) insertions/deletions and a local realignment should be evaluated to check if this additional step can improve the detection of indels.
- Base quality recalibration algorithms might be used to generate more accurate Q-scores.[5]

After these different steps, the output is the SAM (Sequence Alignment/Map) file, which is a tab-delimited text file that contains sequence alignment data, or the BAM file. It includes several types of

information such as the mapped read sequences, base quality scores, mapping quality scores, and the position of insertions/ deletions/matches in the alignment.[8]

- Variant calling: once the reads are mapped, differences with the reference sequence are identified as SNVs or indels.

The output of this last part of the secondary analysis is the VCF file, which contains for each variant the chromosomal position, type of variant, coverage, allelic frequency, gene name and the quality scores. These secondary analysis steps can be performed on or off the NGS instrument.[12]

During the test development, settings and quality metrics are determined and optimised in order to increase performance characteristics (validation chapter). Acceptable ranges and thresholds for each of these quality metrics should be defined and documented.[2,3,8,9] Settings are:

- Alignment settings (seed length, mismatch tolerance, mismatch penalties, gap penalties and gap extension penalties).
- Informatics filter settings which allow ignoring any read that map to non-targeted regions and analysing only reads mapping to the specific regions targeted.[5]

They are determined and optimised in order to increase the specificity (validation chapter) of the test. For example, by sequencing at the minimal coverage read depth (validation subchapter) a sufficient number of normal samples in which no variants are present, including low quality samples, and then adjusting the settings in order that no false-positives are detected.

Quality metrics are:

- Mapping quality scores which measure the uncertainty that a read is mapped properly to the reference sequence / genome.
- Proportion of duplicated reads (if appropriate).
- Coverage read depth of the region(s) of interest, which is the number of independent overlapping base calls. Coverage read depth threshold should be established during the validation to ensure adequate sensitivity (validation chapter) in the region(s) of interest.[2,3,5,9,12] Reference materials (general requirements for targeted NGS tests chapter) are preferably used to define the minimum coverage read depth for which additional coverage does not significantly improve the accuracy of the sequence (e.g. plotting the number of false positive and negative results as a function of coverage).[5] Variants not meeting the minimum coverage read depth should be tested by other methods or reported as not informative. Table or graph (e.g. histogram or box-and-whisker plot) of coverage read depths for each target area should be provided.[4]
- Average read coverage depth which is the average number of overlapping reads within the total sequenced area.[4]
- Allelic read percentage (also called variant allelic frequency (VAF)) which defines the percentage of variant reads in a background of normal reads. A minimum allelic read percentage should be equal to or higher than the limit of detection, which is determined during the validation process (validation chapter).[4,5]

In order to assess that the sequencing run has performed correctly, some quality metrics can also be measured and their corresponding threshold should be determined during the validation plan. These metrics are platform specific and should be determined during the validation process.

- For example, for the Illumina platform, the optimal range of the cluster density (CD), which is the number of clones (or clusters) per mm2 can be defined. Cluster density mostly depends on the concentration of the DNA library pool and should reach a minimal threshold in order to obtain a sufficient number of reads.
- If applicable, a threshold for the error rate, which is based on the sequence of a known spiked-in control (e.g. PhiX), can be determined. The error rate is directly related to the Q-scores.

7. VALIDATION

7.1 General

Test validation is necessary to ensure that a new test is performing properly as intended for its clinical use. Desired performance characteristics should be defined in advance and integrated in the validation plan. The level of validation is function of the type of test.

Following the ISO 15189 norm, a validation is required for any new in-house (NGS) technology. The validation should apply to all intermediate steps of the entire (NGS) process and should include a deep investigation of the critical parameters defined in the test development, in order to detect any source of variation and interference and to verify that the desired performance criteria and requirements for process control are met.

7.1 Performance characteristics determination

Performance characteristics include limit of detection, analytical sensitivity, specificity, precision and accuracy (see definitions). These characteristics should be empirically established and validated separately for each type of variant.[4,5]

These performance characteristics depend on several quality metrics described in the chapter on test development such as coverage read depth, allelic read percentage and base quality scores (Q-score) and on pipeline settings. The influence of the sample types on the performance characteristics should be considered.

7.2 Limit of detection

A crucial step in every validation plan is the establishment of the limit of detection (LOD). LOD can be assessed, for example, by dilution series of well-characterised DNA samples with known mutations in wild-type DNA, or by using reference samples (for example HorizonDx samples in which different mutation types, at various VAF's ranging from 30% to less than 2%, in various genes are engineered).[12] This way of testing has the advantage that it may incorporate several confounding factors that may impede the LOD in daily practice. The dilution should be performed to the point that the variant of interest can no longer be detected. From these dilution series, the minimal required coverage read depth to detect a variant at desired VAF can be determined.

The limit of detection is usually around 5% of allelic frequency. Minimum five variants (SNVs and indels) should be analysed.

7.3 Analytical sensitivity and specificity

Analytical sensitivity and specificity are related to different quality metrics described in chapter 6. For example, the desired sensitivity and specificity may not be achieved when coverage read depth and base quality scores (Q-score) are below the threshold.[5]

Analytical sensitivity and specificity should be empirically established separately for each type of variant (SNVs and indels), using samples that are representative for the intended clinical sample type. Well-characterised reference materials (such as HorizonDx references) or clinical samples already analysed by another independent method such as Sanger sequencing, qPCR, or NGS by another lab may be used.[4,5,8]

A sufficiently high number of variants and samples, adequately defined and representative for the clinical purpose, should be investigated in the validation process.[6] The numbers of tested variants will be smaller for smaller gene panels and higher for bigger ones.[22,23]

A sensitivity of at least 95% and a specificity of at least 99% should be pursued.

7.4 Analytical precision

Repeatability can be established by sequencing the same sample (minimum three different samples) using different barcodes in triplicate at least under the same conditions in the same run.[4,5,7,8]

Reproducibility can be established by sequencing the same sample (minimum two different samples per variant type (indels and SNVs)), in three different runs on the same instrument, or on different instruments if applicable (instrument variability), and by different technicians (inter operator variability).[3,4,7,8,12]

The inter-operator reproducibility for the classification of variants should also be assessed.

A repeatability and a reproducibility of at least 95% should be pursued.[12]

7.5 Analytical accuracy

Analytical accuracy should be established by sequencing well-characterised reference materials with multiple variants (that are representative for the intended clinical sample type) including those with allelic frequencies close to the established detection limits. In addition, the data obtained from the analytical sensitivity and specificity assays should be included. Analytical accuracy should be ≥99%.

7.6 Validation of changes in the bioinformatics part only

Change in a part of the process, for example software updates or software changes, requires a validation of the particular bioinformatics part. In-house available data files or files from other NGS-accredited labs may be used.2,4

The validation can be achieved by using existing data, which are representative of the analysed tumour samples from at least 50 variants and 30 previously analysed samples (depending on the size of the panel, see above) with known mutations that include SNVs and indels to verify that all the variants are still detected with the same analytical sensitivity, preferentially across a wide range of coverage levels.

Assessment of the quality metrics should be done to ensure that no significant differences exist between the different software versions to enable the detection of all relevant variants.[2-4,9,12]

The software update release notes describing the modifications should be logged.

8. VERIFICATION

If performance specifications are available (for CE- marked IVD-compliant kits), the NGS test should be verified in their own laboratory in order to establish that specifications are met, in other words that the test is performing correctly as stated by the manufacturer. Moreover, critical quality metrics and parameters (test development chapter) should be measured.

The verification procedure is also applicable when minor modifications to the experimental protocol or to the composition of the gene panel (e.g. when adding a new gene) of a validated workflow are performed.

For verification, at least 10 retrospective samples with known variants (including SNV and indels) should be tested and at least 10 variants detected in prospective samples should be confirmed by an independent reference method, which may be the original NGS method.[4]

9. QUALITY CONTROL

9.1 General

To ensure and maintain accurate test results, quality controls should be performed periodically at different levels:

- Internal Quality Control (IQC) should ensure that the process (instrument-reagents-operators) is working properly each time samples are processed. Procedures for IQC should cover checkpoints at different critical steps (by controlling quality metrics and quality parameters) and on the whole process (by determining performance characteristics).
- External Quality Assessment (EQA), with unknown material provided by a third party, should ensure that the performance of the laboratory itself and of the method used, complies with (inter)nationally accepted performance criteria.

9.2 Internal quality control

Procedures on internal quality control should be implemented by the laboratory to monitor the performance of the entire analytical process for each NGS test and its reproducibility over time. This should also allow for detecting errors or nonconformities during the process and eventually will indicate the need to interrupt the process if necessary.[5] The performance specifications and quality metric thresholds derived from the validation/verification process or from the manufacturer will be used to assess the validity of each test run.

9.3 Quality control materials

- Positive controls

A positive control, such as an engineered DNA reference material, should be included to assess the NGS test on a regular basis and at critical steps (for example when starting a new lot of critical reagents) and should

contain multiple known somatic variants of different types, preferably near the limit of detection of the assay in order to assess that low percentage variants can reproducibly be identified.[2,4,5]

The frequency of analysing positive controls should be based on the stability of the procedure and the risk of harm to the patient from an erroneous result.[1] Particularly for sequencing, evaluating predefined run quality metrics may be adequate to assure the validity of each single run, making the analysis of a positive control in each run superfluous.

Laboratories should document the use of positive controls and monitor the results over time.

- Negative controls

It is advised to include a no-template control during the PCR steps within the template preparation to check for sample contamination.[4,9]

Moreover, data analysis can be performed to check if reads are generated from a barcode used in the previous run and not in the current run and if reads are generated for targets not included in the current run.

Different mutational profile for each sample is a strong indication that there is no sample contamination, making the use of negative control superfluous.

- Performance monitoring

The performance measures determined in the validation process (validation chapter) should be recorded in the validation/verification report (validation/verification report subchapter) and in subsequent routine diagnostic runs. Comparison to those of an optimal validated run can be used to monitor the reproducibility and the overall quality.[2]

- Quality metrics monitoring

Quality metrics should be monitored at each run and routinely collected and compared to those of an optimal validated run.[2]

Any significant deviations should be investigated and may require repeating the test.[2,5] It can also help in de- fining the source of the problem in an underperforming test.[2,4]

- External quality control

Proficiency testing (PT) and EQA should be performed periodically at least once a year to monitor the test performance, by analysing well-defined materials provided by an independent third party but unknown to the laboratory.[2] Laboratories should share with each other well-characterised samples and data files to collaboratively improve and standardise NGS testing.[8]

10. TERTIARY ANALYSIS

Based upon secondary analysis data, tertiary analysis is mainly composed of two different steps: (1) the annotation and the biologic classification of the identified sequence variants and (2) their clinical classification and their clinical utilities annotations. This part is performed off-instrument.

1. Each variant should be annotated with dedicated software that annotates each variant in relation to its position in the gene (exonic, coding, amino acid change, etc.), classified into 5 biologic classes (cfr 11.2), following a systematic and documented procedure which should be described in the traceability documents and be part of the quality system.

2. Secondly, each variant should be classified into 4 clinical classes (cfr 11.3) and annotated with their clinical utilities (diagnostic, prognostic or therapeutic). Clinical classification and clinical utilities annotations are based on literature search and screening into different database (such as Cosmic, dbSNP, My Cancer Genome, ClinVar, Civic, MD Anderson) and both should be discussed within a post-analytical discussion forum, also called Molecular Advisory Board (MAB) (national or local) (composed of clinicians, pathologists, molecular geneticists, etc.). The functionality of the MAB should be adequately documented.

Figure 2a describes the workflow for the biological classification of somatic variants (v1). Figure 2b gives an overview of the variant classification scoring table with exceptions.

Workflow for the biological classification of somatic variants (v1)

Exonic and splice variant fulfill **technical** validation (Lab-specific thresholds for coverage and VAF, artefact, etc....) — No → Ignore or repeat

↓ Yes

Common variants in the population — Any etnic-based MAF >0.1% (GnomAD, ESP, 1000 genome) — Yes → Benign or Probably Benign

↓ No

In Consensus Pathogenic Variants (CPV) list* — Yes → Pathogenic

No ↓   No ↓

No clear LoF mutation (missense, in-frame indel)   Clear LoF mutation (fs, stop, AG/GT splice)

↓   ↓   ↓

'Scoring system' *(see scoring table)*   In Oncogene**   In Ts gene**

    ↓   ↓

    VUS   Probably Pathogenic

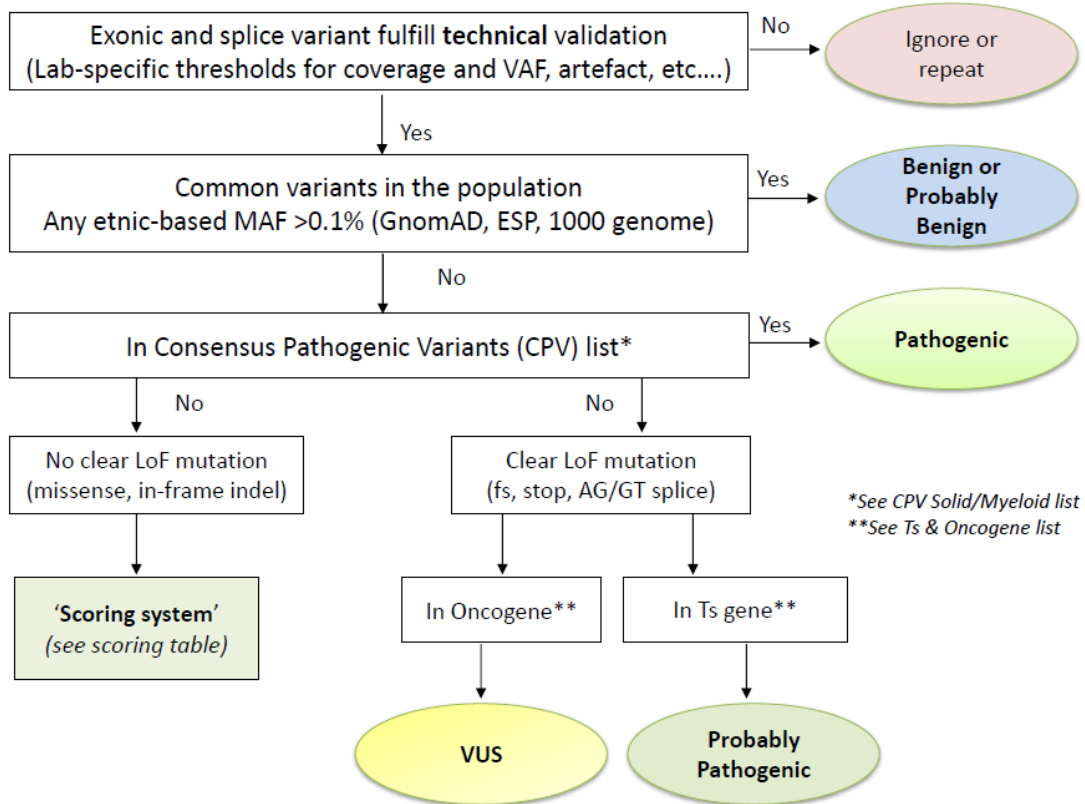*See CPV Solid/Myeloid list
**See Ts & Oncogene list

Figure 2a: Workflow for the biological classification of somatic variants (v1).
Consensus Pathogenic Variants (CPV) and Tumor Suppressor (Ts) & Oncogene list are in Annexes.

Variant Classification Scoring Table

Parameter	Score 2	Score 1	Score 0.5	Score 0	Score -1
Total # of entries of that particular AA change at that position in COSMIC	Solid: ≥50 *Hemato: ≥10*	50 > x > 10 *10 > x > 5*	/ */*	≤10 *≤5*	/ */*
In silico prediction tools SIFT and MutationTaster	/	/	Both damaging or deleterious	Other	/
Harmful in functional studies (PubMed, JAX-CKB, MDA, MCG)	/	/	Yes	Not reported	No
Described in at least one genomic db (CIVIC, ClinVar, OncoKb, VarSome)	/	/	As (Probably) Pathogenic	Not described /unknown	As (Probably) Benign

Variants with a **score ≥2** will be classified as '**Probably Pathogenic**'
Variants with a **score < 2** are classified as '**VUS**'

https://cancer.sanger.ac.uk/cosmic/
http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html
http://mutationtaster.org/   http://oncokb.org/#/
https://civicdb.org/home   http://www.clinvar.com/
https://varsome.com/   https://pct.mdanderson.org
https://ckb.jax.org/   https://www.mycancergenome.org/

#: number (counts)
AA: amino acid
db: database

EXCEPTIONS:

➢ Variants in **TP53** have to be checked in - the IARC db (http://p53.iarc.fr/)
or - the Seshat db (http://vps338341.ovh.net/)

➢ The c.1934dupG in the polyG stretch of **ASXL1** is 'Probably Pathogenic' only if VAF is significantly higher than that obtained for the polyG stutter error (is method specific).

➢ Splice site variants are restricted to the -2, -1 and +1, +2 exon-flanking intronic positions.

➢ **MET** exon 14 splice site variants have to be seen as a CPV and not as splice variants.

➢ **CALR** out-of-frame indels have to be seen as CPVs and not as frameshift variants.

➢ **FLT3** ex14 ITD and **NPM1** ex11 delins have to be seen as CPVs and not as in-frame or frameshift mutations, respectively.

➢ A variant that occurs multiple times in the internal db at germline VAF should be evaluated for being a rare SNP

CPV: Consensus Pathogenic Variant
db: database
ITD: Internal Tandem Duplication
SNP: Single Nucleotide Polymorphism
VAF: Variant Allele Frequency

Figure 2b: Variant classification scoring table with exceptions.

11. REPORTING

11.1 Variant annotation

It was decided to use the HGVS annotation (http://www.hgvs.org/) published by 'den Dunnen et al. Hum Mutat 2016'.

Special attention: the protein annotation has to be between brackets e.g. p.(Val600Glu) to indicate that the variant was found at DNA level only.

The 3-letter code for the amino acid has to be used in the results section (recommendation of HGVS) but the one letter code can also be added between brackets e.g. p.(Val600Glu) (V600E).

Use '*' instead of 'Ter'. HGVS allows both but we reached a consensus terminology. E.g. p.(Cys102*) to describe a stop codon, or p.(Cys102Tyrfs*13) to describe a frameshift.

Positions are based on the NM_ number of the main transcript.

11.2 Biological classes

There is a general consensus to use these 5 biological classes:

Biological class	Reporting
Pathogenic	must be reported
Probably pathogenic	must be reported
VUS	must be reported clearly separated from pathogenic and probably pathogenic variants
Probably benign	should not be reported
Benign	should not be reported

This classification is based on the ACMG and AMP Standards and Guidelines publication of Richards et al. Genet Med 2015, even though these guidelines are meant for germline variants.

Remarks:

Pathogenic and Probably pathogenic biological classes are present in the clinical report in the section 'Test results'. VUS variants can be in the 'Test results' section, clearly separated from the other variants or at the end of the report.

As the clinical report is written in the hospital's language, it was decided to translate 'probably pathogenic' in:

French	→ Probablement pathogénique
Dutch	→ Vermoedelijk pathogeen
German	→ Wahrscheinlich pathogen

11.3 Clinical classes

There is a general consensus to use 4 clinical classes:

Clinical class	Reporting
Tier I: Strong clinical significance	must be reported
Tier II: Potential clinical significance	must be reported
Tier III: Unknown clinical significance	must be reported
Tier IV: variants deemed benign or likely benign	should not be reported

This classification is based on the ACMG and AMP Standards and Guidelines publication of Li et al. J Mol Diagn 2017.
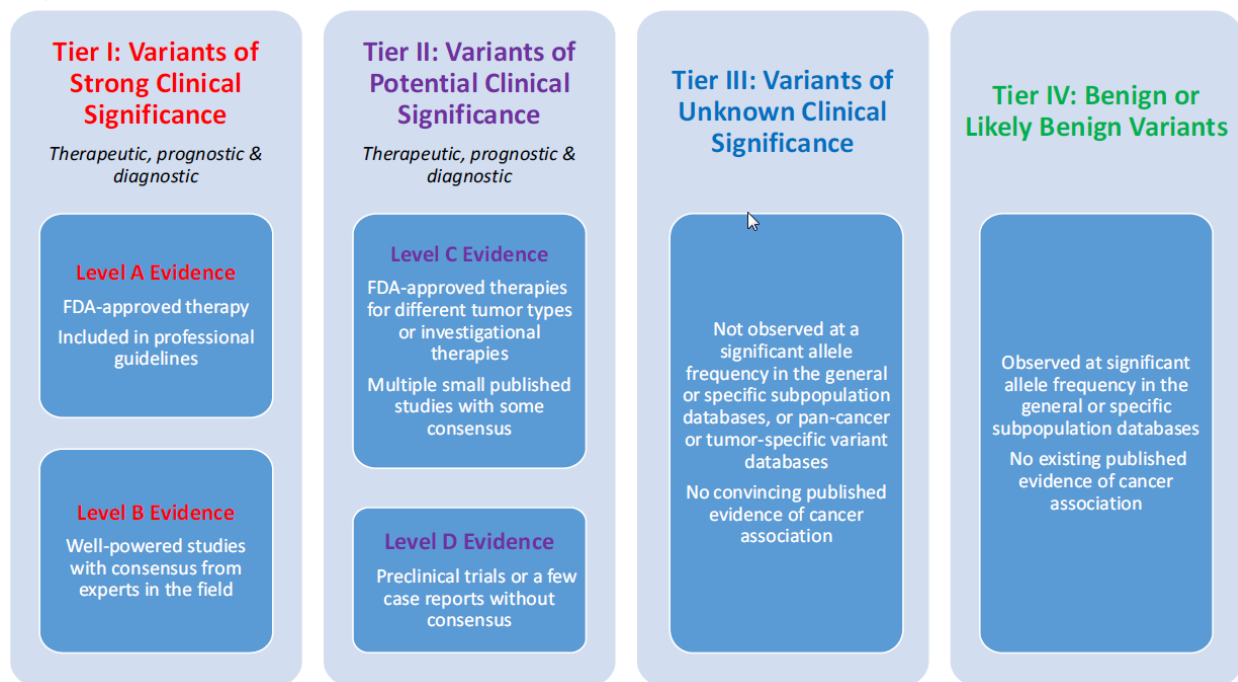
Clinical classes (Strong clinical significance (Tier I), Potential clinical significance (Tier II) and Unknown clinical significance (Tier III)) are present in the clinical report in the section 'Conclusions and interpretation of the report' (see below).

Remark: VUS variants are always considered as a Tier III class.

French translation:

Clinical class	Reporting
Tier I: Strong clinical significance	Impact clinique avéré
Tier II: Potential clinical significance	Impact clinique potentiel
Tier III: Unknown clinical significance	Impact clinique indeterminé

Dutch translation:

Clinical class	Reporting
Tier I: Strong clinical significance	Significant klinisch belang
Tier II: Potential clinical significance	Mogelijk klinisch belang
Tier III: Unknown clinical significance	Ongekende klinische betekenis

German translation :

Clinical class	Reporting
Tier I: Strong clinical significance	Relevante klinische Bedeutung
Tier II: Potential clinical significance	Mögliche klinische Bedeutung
Tier III: Unknown clinical significance	Unbekannte klinische Bedeutung

11.4 Clinical report

The clinical report should contain the following information, in the hospital's language:

11.4.1 Laboratory, patient and physician identification

As required by the ISO15189 norm, laboratory, patient and physician identification should be reported.

11.4.2 Medical information

The following information is required, if applicable for the tumor type, but can be presented in a different format.

Primary tumor type and histology	e.g. NSCLC, adenocarcinoma e.g. Acute myeloid leukemia
Clinical Information and Request	e. g. stage of tumor, treatment phase, therapy considered, …

11.4.3 Sample information

The following information is required, if applicable for the tumor type, but can be presented in a different format.

Sample ID (primary lab)	XXXX
Sampling date	e.g. 24-09-2018
Date of sample received	e.g. 01-10-2018
Sample tumoral stage	Primary/metastasis
Sample anatomic site	e.g. colon, lung, liver, blood, pleural cavity, bone marrow, axillary lymph node …
Sample type	eg. Resection, (endoscopic) biopsy, fluid, aspirate, trephine biopsy …
Sample procedure	e.g. FFPE, frozen tissue, fresh tissue;…
Neoplastic cells (%)	e.g. 10
Sample quality	Disclaimer if sample does not fulfill pre-analytical requirements

11.4.4 Test results

A table format is preferred but is not mandatory.

Gene name	at coding DNA level	at protein level	Biological classes	Allelic frequency (%)
EGFR	c.2573T>G	p.(Leu858Arg)*	Pathogenic	5

Non interpretable results :

List them (only the hotspot, or specify that gene xxx was not totally covered) and explain the reason (due to insufficient coverage, below the validated threshold/….)

Analysis not possible :

explain the reason
(no more available material / necrosis/insufficient material / DNA quality too low/ …)

*(L858R) may be added

Remarks:
- NM_refs (versioned) have to be specified either in the Method (see part 8) or here in this table.
- Genomic position should not be included.
- The use of the three-letter AA code is obligatory. The one-letter AA code may also be included.

11.4.5 Conclusions and interpretation of the report

This chapter should contain the conclusions and clinical interpretation of each pathogenic and probably pathogenic variant.
Each variant has to be annotated with their clinical significance (strong clinical significance, potential clinical significance and unknown clinical significance) and their clinical impact (e.g. sensitivity/resistance to a therapy, involved in the diagnosis, involved in the prognosis).
In case of the absence of mutation in genes for which a mutation confers a resistance to a treatment (e.g. CRC without KRAS mutation), the absence of the mutation has to be indicated.
References should be added and provided in an unequivocal manner (e.g. Li et al 2016 J Mol Diagn; or, PMID 27993330).

11.4.6 Identification of the person(s) who interpret and/or validate the results of the analysis

As required by the ISO 15189 norm.

11.4.7 Date of the report

As required by the ISO 15189 norm.

11.4.8 Methods

The following information is required either in this part of the clinical report or on the laboratories web site with a referral in the report to the website. The referral should mention a version number of the information (e.g. 'Information for patients and users- version XX').

- Brief method for the wet and dry lab analyses (kit and software names and versions)
- Sequencer reference (e.g. MiSeq, Ion GeneStudio S5, ...)
- Limit of detection (VAF threshold, ...)
- List of genes and exons analyzed (e.g. BRAF exons 11 and 15; DNMT3A all coding exons; ...) with NM_ accession numbers with version (e.g. NM_00522.3)
- List of all exons or regions having insufficient coverage in most samples (e.g. DNMT3A exon 6 AA165-184; TET2 exon 4; ...) .
- Reference genome used (e.g. Hg19).
- A disclaimer that states that this NGS test cannot differentiate between somatic and germline variants.

12. TURNAROUND TIME

The turnaround time for the entire NGS analysis from biopsy to reporting should be appropriate for the intended clinical purpose and in agreement with the tumour specific guidelines, if available, though in general a turnaround time of ≤15 working days is highly recommended.

13. REFERENCES

1. ISO15189: Medical laboratories - Requirements for quality and competence. 1-11- 2012.

2. RCPA. Implementation of Massively Parallel Sequencing. 1-3-2014.

3. Weiss MM, Van der Zwaag B, Jongbloed JD, et al. Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. Hum Mutat. 2013,34:1313-21.

4. Department of health NY. Next Generation" Sequencing (NGS) guidelines for somatic genetic variant detection. 1-3-2015.

5. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012,30:10336.

6. May W. Standard Reference Materials®, definitions of terms and modes used at NIST for value-assignment of reference materials for chemical measurements. 2000.

7. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med. 2015,139:481-93.

8. Gert Matthijs. Guidelines for diagnostic next generation sequencing. 1-12-2014.

9. ACGS. Practice guidelines for targeted Next Generation Sequencing analysis and interpretation. 2014.

10. Deans ZC, Costa JL, Cree I, et al. Integration of next-generation sequencing in clinical diagnostic molecular pathology laboratories for analysis of solid tumours; an expert opinion on behalf of IQN Path ASBL. Virchows

Arch. 2017;470(1):5-20.

11. Mattocks CJ, Morris MA, Matthijs G, et al. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet. 2010,18:1276-88.

12. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013,15:733-47.

13. KB: Medisch dossier. 3-5-1999.

14. European Medicines Agency (EMA). ICH guideline E18 on genomic sampling and management 4 of genomic data. 22-1-2016.

15. Jouret Mourin A, Cuvelier C, Demetter P, et al. RAS-testing in colorectal cancer: Belgian guidelines. Belg J Med Oncol. 2015;9:183-90.

16. Cree IA, Deans Z, Ligtenberg MJ, et al. Guidance for laboratories performing molecular pathology for cancer patients. J Clin Pathol. 2014,67:923-31.

17. Chen H, Luthra R, Goswami RS, et al. Analysis of Pre-Analytic Factors Affecting the Success of Clinical Next-Generation Sequencing of Solid Organ Malignan- cies. Cancers (Basel). 2015,7:1699-1715.

18. Goswami RS, Luthra R, Singh RR, et al. Identification of Factors Affecting the Success of Next-Generation Sequencing Testing in Solid Tumors. Am J Clin Pathol. 2016,145:222-37.

19. Smits AJ, Kummer JA, de Bruin PC, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. Mod Pathol. 2014,27:168- 74.

20. Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges. Biotechniques. 2014,56:61-4,66,68, passim.

21. Samorodnitsky E, Jewell BM, Hagopian R, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Hum Mutat. 2015,36:903-14.

22. Froyen G, Broekmans A, Hillen F, et al. Validation and Application of a Custom-Designed Targeted Next-Generation Sequencing Panel for the Diagnostic Mu- tational Profiling of Solid Tumors. PLoS One. 2016,11:e0154038.

23. D'Haene N, Le MM, De NN, et al. Clinical Validation of Targeted Next Generation Sequencing for Colon and Lung Cancers. PLoS One. 2015,10:e0138245

ANNEXES

Consensus pathogenic variants (CPV) Solid / Myeloid list

Consensus pathogenic variants (CPV) in the Convention genes of Solid tumors v1

Gene	Transcript ID	Hs1	Hs2	Hs3	Hs4	Hs5	Hs6	Hs7	Hs8	Hs9	Hs10	Hs11	Hs12
ALK	NM_004304.4	F1174L	R1275Q										
BRAF	NM_004333.5	G469A/E/R/V	D594G	T599-K601if-del/ins	V600E/K/M/R	K601E							
BRCA1	NM_007294.3	none											
BRCA2	NM_000059.3	none											
EGFR	NM_005228.4	G719A/C/S	ex19if-del/ins	ex20if-ins	T790M	C797S	L858R	L861Q					
ESR1	NM_000125.3	K303R	E380Q	V392I	S463P	V533M	V534E	P535H	L536P/Q/R	Y537C/N/S	D538G		
GNAS	NM_000516.5	R201C/H											
H3F3A	NM_002107.4	K28M	G35R/W										
HRAS	NM_005343.3	G12C/D/S/V	G13C/D/R/S/V	Q61H/K/L/R									
IDH1	NM_005896.3	R132C/G/L/H/S											
IDH2	NM_002168.3	R140L/Q/W	R172K/S										
KIT	NM_000222.2	D419if-del	S501-F504if-ins	K550-V560if-indel	W557G/R	V559A/D	V560D	L576P	K642E	V654A	D816H/V/Y	N822K	Y823D
KRAS	NM_004985.4	G12A/C/D/F/R/S/V	G13C/D/R/S/V	A59T	Q61H/K/L/R	K117N	A146T						
MET	NM_001127500.2	ex14 skipping											
NRAS	NM_002524.4	G12A/C/D/R/S/V	G13C/D/R/S/V	A59T	Q61H/K/L/R	K117N	A146T						
PDGFRA	NM_006206.5	S566_E577if-del	D842V	D842_I843if-del	V561D								
PIK3CA	NM_006218.3	R88Q	N345K	C420R	E542K	E545A/D/G/K/Q/V	Q546K/P/R	M1043I	H1047L/R				

if-del: inframe deletion

if-ins: inframe insertion

if-dup: inframe duplication

of-del: out of frame deletion

of-ins: out of frame insertion

_ : denotes the exact positions of that change

- : denotes a region in which the change has to be located

Consensus pathogenic variants (CPV) in the Convention genes of Myeloid tumors v1

Gene	Transcript ID	Hs1	Hs2	Hs3	Hs4	Hs5	Hs6
ASXL1	NM_015338.5	none					
CALR	NM_004343.3	ex9of-del	ex9of-ins				
CEBPA	NM_004364.3	none					
CSF3R	NM_156039.3	T618I					
DNMT3A	NM_175629.2	R882C/H					
EZH2	NM_004456.4	Y646F/H/N/S					
FLT3	NM_004119.2	ex14if-dup	D835A/E/H/V/Y				
IDH1	NM_005896.3	R132C/G/H/L/S					
IDH2	NM_002168.3	R140L/Q/W	R172K/M/S				
JAK2	NM_004972.3	ex12 if-del/if-dup	V617F				
KIT	NM_000222.2	D419if-del	D816H/V/Y	N822K			
MPL	NM_005373.2	S505N	W515any ms				
NPM1	NM_002520.6	ex11of-ins					
RUNX1	NM_001754.4	none					
SETBP1	NM_015559.3	D868N	G870S				
SF3B1	NM_012433.3	E622D	R625C/H	H662Q	K666N/R/T	K700E	G742D
SRSF2	NM_003016.4	P95H/L/R	P95_R102del				
TET2	NM_001127208.2	none					
TP53	NM_000546.5	R175H	Y220C	G245S	R248Q/W	R273C/H	R282W
U2AF1	NM_006758.2	S34F/Y	Q157P/R				
WT1	NM_024426.5	none					

if-del: inframe deletion

if-ins: inframe insertion

if-dup: inframe duplication

of-del: out of frame deletion

of-ins: out of frame insertion

any ms: any missense variant

Tumor suppressor (Ts) & Oncogene List

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge
ABL1	Oncogene		Oncogene	Oncogene
ACVR1	Oncogene			Oncogene
ACVR1B	TSG			TSG
AKT1	Oncogene		Oncogene	Oncogene
AKT2	Oncogene			Oncogene
AKT3	Oncogene			Oncogene
ALK	Oncogene		Oncogene	Oncogene
AMER1	TSG			TSG
ANKRD11	TSG			TSG
APC	TSG	TSG	TSG	TSG
AR	Oncogene		Oncogene	Oncogene
ARAF	Oncogene			Oncogene
ARID1A	TSG	TSG	TSG	TSG
ARID1B	TSG		TSG	TSG
ARID2	TSG	TSG	TSG	TSG
ARID5B	TSG			TSG
ASXL1	TSG	TSG	TSG	TSG
ASXL2	TSG			TSG
ATM	TSG	TSG	TSG	TSG
ATR	TSG	TSG		TSG
ATRX	TSG		TSG	TSG
AURKA	Oncogene			Oncogene
AURKB	Oncogene			Oncogene
AXIN1	TSG	TSG	TSG	TSG
AXIN2	TSG	TSG		TSG
AXL	Oncogene			Oncogene
B2M	TSG		TSG	TSG
BAP1	TSG	TSG	TSG	TSG
BARD1	TSG	TSG		TSG
BCL10	TSG	TSG		TSG
BCL2	Oncogene		Oncogene	Oncogene
BCL2L11	TSG			TSG
BCL6	Oncogene			Oncogene
BCOR	TSG		TSG	TSG
BLM	TSG	TSG		TSG
BRAF	Oncogene		Oncogene	Oncogene
BRCA1	TSG	TSG	TSG	TSG

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge	
BRCA2	TSG	TSG	TSG	TSG	
BRD4	Oncogene			Oncogene	
BTK	Oncogene			Oncogene	
CALR	Oncogene			Oncogene	
CARD11	Oncogene			Oncogene	
CASP8	TSG	TSG	TSG	TSG	
CBL	TSG		Oncogene	Oncogene/TSG	
CCND1	Oncogene			Oncogene	
CCND2	Oncogene			Oncogene	
CCND3	Oncogene			Oncogene	
CCNE1	Oncogene			Oncogene	
CDC73	TSG	TSG	TSG	TSG	
CDH1	TSG	TSG	TSG	TSG	
CDK4	Oncogene			Oncogene	
CDK6	Oncogene			Oncogene	
CDKN1A	TSG	TSG		TSG	
CDKN1B	TSG			TSG	
CDKN2A	TSG	TSG	TSG	TSG	
CDKN2B	TSG	TSG		TSG	
CDKN2C	TSG	TSG		TSG	
CEBPA	TSG	TSG	TSG	TSG	
CHEK1	TSG	TSG		TSG	
CHEK2	TSG	TSG		TSG	
CIC	TSG		TSG	TSG	
CREBBP	TSG	TSG	TSG	TSG	
CRKL	Oncogene			Oncogene	
CRLF2	Oncogene			Oncogene	
CSF1R	Oncogene			Oncogene	
CSF3R	Oncogene			Oncogene	
CTCF	TSG	TSG		TSG	
CTLA4	Oncogene			Oncogene	
CTNNB1	Oncogene		Oncogene	Oncogene	
CYLD	TSG			TSG	
DAXX	TSG		TSG	TSG	
DDR2	Oncogene			Oncogene	
DNMT1	Oncogene			Oncogene	
DNMT3A	TSG		Oncogene	TSG/Oncogene	DNMT3A has to be considered as a Ts gene in **acute myeloid leukemia** (AML) and in **myelodysplastic syndromes**

E2F3	*Oncogene*			Oncogene
EGFR	*Oncogene*		*Oncogene*	Oncogene
EP300	*TSG*		*TSG*	TSG
ERBB2	*Oncogene*		*Oncogene*	Oncogene
ERBB3	*Oncogene*			Oncogene
ERBB4	*Oncogene*	*TSG*		Oncogene/TSG
ERCC2	*TSG*			TSG
ERCC3	*TSG*			TSG
ERCC4	*TSG*			TSG
ERF	*TSG*	*TSG*		TSG
ERG	*Oncogene*			Oncogene
ERRFI1	*TSG*	*TSG*		TSG
ESR1	*Oncogene*			Oncogene
ETV1	*Oncogene*			Oncogene
ETV6	*Oncogene*	*TSG*		TSG/Oncogene
EWSR1	*Oncogene*			Oncogene
EZH2	*Oncogene*		*Oncogene*	Oncogene
FAM123B	*TSG*			TSG
FANCA	*TSG*			TSG
FANCC	*TSG*			TSG
FAT1	*TSG*	*TSG*		TSG
FBXW7	*TSG*	*TSG*	*TSG*	TSG
FGF19	*Oncogene*			Oncogene
FGF3	*Oncogene*			Oncogene
FGF4	*Oncogene*			Oncogene
FGFR1	*Oncogene*			Oncogene
FGFR2	*Oncogene*		*Oncogene*	Oncogene
FGFR3	*Oncogene*		*Oncogene*	Oncogene
FGFR4	*Oncogene*			Oncogene
FH	*TSG*	*TSG*		TSG
FLT1	*Oncogene*			Oncogene
FLT3	*Oncogene*		*Oncogene*	Oncogene
FLT4	*Oncogene*			Oncogene
FOXA1	*Oncogene*			Oncogene
FOXL2	*Oncogene*			Oncogene
FUBP1	*TSG*		*TSG*	TSG
GATA1	*TSG*			TSG
GATA2	*Oncogene*			Oncogene
GATA3	*TSG*			TSG
GLI1	*Oncogene*			Oncogene
GNA11	*Oncogene*		*Oncogene*	Oncogene
GNAQ	*Oncogene*		*Oncogene*	Oncogene

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge
GNAS	Oncogene		Oncogene	Oncogene
H3F3A	Oncogene			Oncogene
HGF	Oncogene			Oncogene
HIST1H3B	Oncogene			Oncogene
HLA-A	TSG			TSG
HLA-B	TSG			TSG
HNF1A	TSG		TSG	TSG
HRAS	Oncogene		Oncogene	Oncogene
IDH1	Oncogene	TSG	Oncogene	Oncogene/TSG
IDH2	Oncogene		Oncogene	Oncogene
IGF1	Oncogene			Oncogene
IGF1R	Oncogene			Oncogene
IGF2	Oncogene			Oncogene
INPP4B	TSG	TSG		TSG
IRS2	Oncogene			Oncogene
JAK1	Oncogene			Oncogene
JAK2	Oncogene		Oncogene	Oncogene
JAK3	Oncogene			Oncogene
JUN	Oncogene			Oncogene
KDM5C	TSG		TSG	TSG
KDM6A	TSG		TSG	TSG
KDR	Oncogene			Oncogene
KEAP1	TSG			TSG
KIT	Oncogene		Oncogene	Oncogene
KLF4	Oncogene			Oncogene
KMT2A	TSG			TSG
KMT2B	TSG			TSG
KMT2C	TSG			TSG
KMT2D	TSG			TSG
KRAS	Oncogene		Oncogene	Oncogene
MAP2K1	Oncogene		Oncogene	Oncogene
MAP2K4	TSG	TSG		TSG
MAP3K1	TSG		TSG	TSG
MAPK1	Oncogene			Oncogene
MDM2	Oncogene			Oncogene
MDM4	Oncogene			Oncogene
MED12	Oncogene		Oncogene	Oncogene
MEF2B	Oncogene			Oncogene

IDH1 has to be considered as an oncogene in glioma, **acute myeloid leukemia (AML) and myeloproliferative neoplasms (MPN)**

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge
MEN1	TSG	TSG	TSG	TSG
MET	Oncogene		Oncogene	Oncogene
MGA	TSG			TSG
MLH1	TSG	TSG	TSG	TSG
MLL2	TSG			TSG
MLL3	TSG			TSG
MPL	Oncogene			Oncogene
MSH2	TSG	TSG	TSG	TSG
MSH6	TSG		TSG	TSG
MTOR	Oncogene			Oncogene
MYC	Oncogene			Oncogene
MYCN	Oncogene			Oncogene
MYD88	Oncogene		Oncogene	Oncogene
NCOR1	TSG		TSG	TSG
NF1	TSG	TSG	TSG	TSG
NF2	TSG	TSG	TSG	TSG
NFE2L2	Oncogene		Oncogene	Oncogene
NFKBIA	Oncogene			Oncogene
NOTCH1	Oncogene/TSG	TSG	TSG	TSG/Oncogene
NOTCH2	Oncogene/TSG		TSG	TSG/Oncogene
NPM1	TSG		TSG	TSG
NRAS	Oncogene		Oncogene	Oncogene
NTRK1	Oncogene			Oncogene
NTRK2	Oncogene			Oncogene
NTRK3	Oncogene			Oncogene
PARP1	Oncogene			Oncogene
PAX5	TSG			TSG
PAX8	Oncogene			Oncogene
PBRM1	TSG	TSG	TSG	TSG
PDGFRA	Oncogene		Oncogene	Oncogene
PDGFRB	Oncogene			Oncogene
PHF6	TSG			TSG
PIK3C3	Oncogene			Oncogene
PIK3CA	Oncogene		Oncogene	Oncogene
PIK3CB	Oncogene			Oncogene
PIK3CD	Oncogene			Oncogene
PIK3CG	Oncogene			Oncogene
PIK3R1	TSG		TSG	TSG
PIK3R3	TSG			TSG

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge
PMS1	TSG	TSG		TSG
POLD1	TSG			TSG
POLE	TSG			TSG
PPM1D	Oncogene			Oncogene
PPP2R1A	Oncogene			Oncogene
PPP6C	Oncogene			Oncogene
PRDM1	TSG			TSG
PRKCI	Oncogene	TSG		Oncogene/TSG
PTCH1	TSG	TSG	TSG	TSG
PTEN	TSG	TSG	TSG	TSG
PTPN11	Oncogene		Oncogene	Oncogene
RAC1	Oncogene			Oncogene
RAD21	TSG			TSG
RAD50	TSG			TSG
RAD51	Oncogene			Oncogene
RAD51C	TSG	TSG		TSG
RAD51D	TSG			TSG
RAF1	Oncogene			Oncogene
RASA1	TSG			TSG
RB1	TSG	TSG	TSG	TSG
RBM10	TSG			TSG
RET	Oncogene		Oncogene	Oncogene
RHEB	Oncogene			Oncogene
RHOA	Oncogene	TSG		Oncogene/TSG
RICTOR	Oncogene			Oncogene
RIT1	Oncogene			Oncogene
RNF43	TSG		TSG	TSG
ROS1	Oncogene			Oncogene
RPTOR	Oncogene			Oncogene
RRAS2	Oncogene			Oncogene
RUNX1	TSG		TSG	TSG
RYBP	TSG			TSG
SDHA	TSG	TSG		TSG
SDHB	TSG	TSG		TSG
SDHC	TSG			TSG
SDHD	TSG	TSG		TSG
SETBP1	Oncogene			Oncogene
SETD2	TSG		TSG	TSG
SF3B1	Oncogene		Oncogene	Oncogene

HugoSymbol	OncoKB	TSGene 2.0 (Vanderbilt)	Vogelstein and al., Science, 2013	Merge
SHQ1	TSG			TSG
SMAD2	TSG	TSG	TSG	TSG
SMAD3	TSG	TSG		TSG
SMAD4	TSG	TSG	TSG	TSG
SMARCA4	TSG	TSG	TSG	TSG
SMARCB1	TSG	TSG	TSG	TSG
SMO	Oncogene	TSG	Oncogene	Oncogene/TSG
SOCS1	TSG			TSG
SOX17	TSG			TSG
SOX9	TSG			TSG
SPEN	TSG			TSG
SPOP	TSG		Oncogene	Oncogene/TSG
SRC	Oncogene			Oncogene
SRSF2	Oncogene			Oncogene
STAG2	TSG		TSG	TSG
STAT3	Oncogene			Oncogene
STK11	TSG	TSG	TSG	TSG
TCF7L2	TSG	TSG		TSG
TERT	Oncogene			Oncogene
TET2	TSG	TSG	TSG	TSG
TGFBR1	TSG			TSG
TGFBR2	TSG	TSG		TSG
TMPRSS2	Oncogene			Oncogene
TNFAIP3	TSG			TSG
TOP1	TSG			TSG
TP53	TSG	TSG	TSG	TSG
TP53BP1	TSG			TSG
TRAF7	TSG			TSG
TSC1	TSG	TSG	TSG	TSG
TSC2	TSG	TSG		TSG
TSHR	Oncogene			Oncogene
U2AF1	Oncogene		Oncogene	Oncogene
VHL	TSG	TSG	TSG	TSG
WT1	TSG			TSG
XRCC2	TSG			TSG
YAP1	Oncogene	TSG		Oncogene/TSG