Sequencing Guideline

for

MH Guide,

MH Guide/Mendel,

MH Guide/BRCA

Manufacturer:

Molecular Health GmbH
Kurfuersten-Anlage 21
69115 Heidelberg
Germany

Contents

1. About this document

This document describes the necessary process requirements to enable data analyses using MH Guide, MH Guide/Mendel, and MH Guide/BRCA. It includes a description of the quality of the starting materials and additional technical requirements. If not stated otherwise, the notion of MH Guide covers MH Guide, MH Guide/Mendel, and MH Guide/BRCA.

The sequencing laboratory is a qualified laboratory capable of providing input data for MH Guide. Qualified laboratories may at the same time also be MH Guide, MH Guide/Mendel or MH Guide/BRCA customers.

This sequencing guideline must be followed to ensure analytical product performance and the Service Level Agreement (SLA) of MH Guide.

2. Molecular Health and laboratories: General information

Molecular Health (MH) provides MH Guide to healthcare professionals incl. clinical laboratories via a web application as a Software as a Service (SaaS). The underlying infrastructure (e.g., servers and databases) is operated at a data center on servers managed by MH. Since the final variant detection including their interpretation by MH Guide strongly depends on sample quality and consequently on the Next Generations Sequencing (NGS) data quality provided by the lab, the following document lists the sequencing guidelines, including sample criteria, that must be fulfilled by the laboratory and the hospital/physician providing the sample to the laboratory. This sequencing guideline lists the requirements for the various molecular alteration data that are supported as input data for MH Guide analyses, in particular:

- Data in FASTQ format
- Data in Variant Call File (VCF) format
- Data from additional tests

Data assessed are either from an IVD or a validated laboratory procedure of a lab fulfilling the requirements detailed in section “Requirements for qualified laboratories”.

For information on MH Guide, including web browser access, SaaS, SLA and Support Services, see the master service agreement and the Instructions for Use.

Requirements for qualified laboratories

To ensure the quality and reproducibility of the processes, molecular alteration data shall be obtained from laboratories that are qualified to conduct medical genetic testing fulfilling the regulatory and statutory requirements from the regulatory environment MH Guide is marketed in. Dependent on the regulatory environment, laboratories should be accredited according to pertinent standards and regulations, such as e.g., ISO 15189 or the Clinical Laboratory Improvement Amendments (CLIA).

Somatic WES assays in MH Guide must be performed with a control (normal) sample to ensure the quality of the overall interpretation of somatic and germline variants. For somatic panel assays in MH Guide a somatic sample needs to be provided. For MH Guide/Mendel and MH Guide/BRCA, a germline sample is required.

Sample material that has not been used for sequencing is stored for two months by the sequencing lab, unless stipulated otherwise. Upon request of the ordering entity, the material must be destroyed or returned to the ordering entity or the patient.

If the laboratory uses its own Variant Detection Pipeline (VDP) and delivers a VCF file, the laboratory is obliged to store the FASTQ, the BAM files and the VCF according to the applicable regulations in the territory.

If the laboratory provides FASTQ and uses MH Guide for variant detection, the laboratory is obliged to store the FASTQ file according to the applicable regulations in the territory. If other intermediate files need to be retained by applicable regulations, these can be downloaded from MH Guide for a period of 30 days after analysis completion.

Recommended processing time

MH strongly recommends lab processing times (from sample to FASTQ/VCF) of maximum 10-15 business days to allow for a timely availability of analysis results for clinical decision making.

Sequencing and data transfer guidelines

Sensitive patient data

MH Guide processes the samples and data associated with the patient exclusively in pseudonymized form, e.g., by using a bar code assigned to the patient or a patient ID. The physician/laboratory guarantees that MH cannot identify the patient.

Identification of samples

The **MH Guide order number** is a unique identifier (**Case ID**). Samples associated with this ID can be identified at every step of the process. This number is transmitted to MH using either the MH Order Portal or the SFTP process.

Scenario 1: The order is placed via the MH Order Portal and the Case ID is generated by MH. This Case ID must be used at every step of the process, including file upload via SFTP.

Scenario 2: All files including the CaseID.xml are transferred via SFTP, without using MH Order Portal. In this case, the MH Guide client can provide their own Case ID, which must be used by the lab when uploading data following the specifications in the MH Guide Instructions for Use.

Data structure

The MH Guide, MH Guide/Mendel, MH Guide/BRCA, and MH Order Portal Instructions for Use provide a detailed description of the supported data formats and naming conventions to be used.

Transfer of the data to MH

Molecular alteration data in FASTQ or VCF format can be uploaded via the MH Order Portal or, for high throughput, via SFTP. The runtime for the data upload depends on the network bandwidth, on the assay type, coverage, and other parameters.

The MH Guide, MH Guide/Mendel, MH Guide/BRCA, and MH Order Portal Instructions for Use provide a detailed description of the data transfer process.

All raw files (FASTQ files) must be in gzip format.

For good performance of the data upload an internet connectivity of 100 Mbit/s should be available on the site from which data are uploaded.

Better would be **1 Gbit/s** (the MH data center provides this bandwidth), since the data transfer of larger FASTQ files (size depends on target region and on coverage) might take a long time. So, for WES and FASTQ transfer MH strongly suggests 1Gbit/s internet connectivity to ensure upload performance.

The **network transfer time**, if the line is completely reserved for the transfer, is:

- MH Panel (500x average coverage, 1 GB gzipped) using a 100 Mbit/s upload line: **5 min (best case)**
- MH Panel (500x average coverage, 1 GB gzipped) using a 1 Gbit/s upload line: **1 min (best case)**
- WES (200x average coverage, 20 GB gzipped) using a 100 Mbit/s upload line: **2 hours (best case)**
  WES (200x average coverage, 20 GB gzipped) using a 1 Gbit/s upload line: **10 min (best case)**

3. Next-generation sequencing (NGS) data or genetic and molecular alteration data requirements

FASTQ data requirements

Sample and sequencing requirements

The determination of the **tumor content** of tumor samples is typically performed by a pathologist and the result for each sample is transmitted to MH via the MH Order Portal or in an xml file via SFTP. A tumor cellularity of $\geq$ **20%** is required for somatic analyses with MH Guide. CNA detection requires a tumor cellularity $\geq$ **50%**. Analytical performance for CNA detection in MH Guide cannot be ensured with lower tumor content.

For FASTQ input: During the onboarding process, the laboratory must transmit the BED file matching the WES, Panel Kit used for sequencing. If another BED file is used in subsequent sequencing runs, then the new BED file must be provided to MH. The MH Guide customer is responsible for transmitting this information in order to ensure proper processing of the files.

DNA sequencing specifications for somatic panel analyses using MH Guide (FASTQ)

DNA enrichment/Kits: hybrid capture based

Read Type: Illumina PairedEnd

Quality Values:

- Error rate of read #1 and #2: $\leq$1% (error rate of the reads matching the PhiX Spike-In)
- Bases exceeding Q30: $\geq$ 90%
- QC values as listed in section 4 must be fulfilled

Adapter Trimming: Adapter trimming (e.g., needed for FFPE samples) by the laboratory is **mandatory**

Real coverage (diagnostic coverage)* of the **panel**:

Somatic **Panel** analyses with MH Guide require an average real coverage* of 500x or a minimum average real coverage of $\geq$ **300x**, in which at least 96% of the target region has $\geq$ **100x** average real coverage. Higher coverages (>500x) may generate substantially bigger files and need high computing resources, depending on the target region size. Coverages up to 2000x are supported by MH on request and are associated with higher prices. Details are defined in the Master Service Agreement (MSA).

* The **average real coverage** is defined as on-target coverage after removal of duplicate read pairs. The target region is the enriched region for sequencing, which is defined via a BED file. The value can be seen in the quality analysis section in MH Guide as "Average real coverage" (see also section 4).

Other Panel specifics:

(optional) RNA sequencing for detection of fusions: > 10 Mio Illumina Paired End reads, minimum read length 50 bp; Supported assay types: Whole Transcriptome, hybrid capture. Datasets with up to 262 million Paired End Reads are supported by MH.

DNA sequencing specifications for paired somatic/germline whole exome sequencing (WES) using MH Guide VDP (FASTQ)

DNA enrichment/Kits: hybrid capture based, e.g., Agilent SureSelectXT Human All Exon, V6 or V7.

Read Type: Illumina Paired End

Quality Values:

- Error rate of read #1 and #2 ≤1% (error rate of the reads matching the PhiX Spike-In)
- Bases exceeding Q30: ≥ 90%
- QC values as listed in section 4 must be fulfilled

Read Length: max 400 bases

Adapter Trimming: Adapter trimming (e.g., needed for FFPE samples) by the laboratory is **mandatory**

Real coverage (diagnostic coverage)* of the **exome**:

Paired **Exome** analyses with MH Guide require for the somatic sample a minimum average real coverage* of ≥ **200x**, in which at least 80% of the target region has ≥ **100x** average real coverage in the tumor sample. MH strongly recommends to sequence **both** tumor and control (germline) samples for the WES experiment to reach a comprehensive variant interpretation. Variant calling accuracy is described in the MH Guide Instructions for Use. Average real coverage for the control sample can be reduced to **40x** at the minimum. To improve precision and sensitivity and to ensure accurate identification of somatic SNVs and Indels that are also present at low frequency in the control sample, a coverage of 100x for this sample is recommended. Higher coverages (>200x) of the whole exome for the tumor sample generate substantially bigger files and need high computing resources, depending on the target region. Coverages up to 600x for the somatic sample and 160x for the germline control sample are supported by MH on request and are associated with higher prices. Details are defined in the Master Service Agreement (MSA).

* The **average real coverage** is defined as on-target coverage after removal of duplicate read pairs. The target region is the enriched region for sequencing, which is defined via a BED file. The value can be seen in the quality analysis section in MH Guide as "Average real coverage" (see also section 4).

Other WES specifics:

CNA detection requires a tumor cellularity > 50%.

(optional) RNA sequencing for detection of fusions: > 10 Mio Illumina Paired End reads, minimum read length 50 bp; Supported assay types: Whole Transcriptome, hybrid capture. Datasets with up to 262 million Paired End Reads are supported by MH.

DNA sequencing specifications using MH Guide/Mendel and MH Guide/BRCA VDP (FASTQ)

Germline analyses with MH Guide/Mendel or MH Guide/BRCA require a minimum average diagnostic (real) coverage* of ≥ **100x**, in which at least **96%** of the target region has ≥ **40x** average real coverage.

DNA enrichment/Kits: hybrid capture based

Read Type: Illumina Paired End

Read Length: max 400 bases

Quality Values:

- Error rate of read #1 and #2: ≤1% (error rate of the reads matching the PhiX Spike-In)
- Bases exceeding Q30: ≥ 90%
- QC values as listed in section 4 must be fulfilled

* The **average real coverage** is defined as on-target coverage after deleting duplicate read pairs. The target region is the enriched region for sequencing, which is defined via a BED file. The value can be seen in the quality analysis section in MH Guide as "Average real coverage" (see also section 4).

VCF data requirements

VCF data provided as molecular alteration input data must adhere to the MH VCF specification as detailed in the Instructions for Use. If support for VCF format adaption to the MH VCF format is required, MH provides adapters for range of pipelines or can implement a customized VCF adapter as a consulting service. Before analyzing VCF data with MH Guide, MH needs to set up a labtest in MH Guide that reflects the validated performance of the underlying LDT or IVD. For the setup of the labtest, please contact the MH customer support.

Additional Test Results requirements

Additional test results that can be added to individual MH Guide and MH Guide/Mendel cases, must originate from validated LDTs or IVDs to ensure only validated molecular alteration data is reported in MH Guide and MH Guide/Mendel.

4. Input data quality criteria (QC) and error handling

Warnings and errors

Warnings and errors are threshold values that highlight potential data issues and may require extra technical QC validation by the customer. MH reserves the right to inform the laboratory and/or the customer about input data QC issues.

An error is a threshold value that will result in a rejection of the data by MH.

Warnings and errors are triggered based on various coverage-based thresholds such as average real coverage or a higher fraction of targeted bases with lower than expected coverage.

Errors that lead to a rejection (failure) of the analysis typically occur when files do not match (e.g., different number of reads or non-matching IDs in paired-end FASTQ files), when the files are corrupted, when they are syntactically wrong, or when files uploaded to the MH SFTP server do not comply with MH file naming policies as defined in the MH Guide, MH Guide/Mendel, and MH Guide/BRCA Instructions for Use.

For analyses based on VCF input data, the analysis also fails, if no valid variants were found in the VCF. This may indicate that the format does not conform to the MH VCF format.

Warnings and errors are indicated for each case with a yellow case data icon. Detailed QC statistics are available for the user in the QC log files for a case in MH Guide and MH Order Portal. Additional details are provided in the MH Guide, MH Guide/Mendel, and MH Guide/BRCA Instructions for Use.

Values triggering molecular input data warnings and errors in MH Guide

FASTQ Data

MH Guide Whole Exome (WES) paired analyses:

- Number of **read pairs (FASTQ)**:
  - Warning threshold < 50.000
- Fraction of **targeted bases with minimum real coverage** (100x) (after aligning):
  - Warning threshold Exome (tumor): < 80%
- Average **real coverage** (excluding duplicates):
  - Warning threshold (tumor) < 200x
  - Warning threshold (control) < 40x
- Fraction of **mapped reads** (after aligning):
  - Warning threshold < 95%
  - Error threshold: < 85%
- Fraction of **discordant read pairs** (after aligning):
  - Warning threshold: > 5%

MH Guide unpaired (panel) analyses:

- Number of **read pairs (FASTQ)**:
  - Warning threshold < 50000
- Fraction of **targeted bases with minimum real coverage** (100x), (after aligning):
  - Warning threshold: < 96%
- Average **real coverage** (excluding duplicates):
  - Warning threshold: < 300x
- Fraction of **mapped reads** (after aligning):
  - Warning threshold < 95%
  - Error threshold: < 85%
- Fraction of **discordant read pairs** (after aligning):
  - Warning threshold: > 5%

MH Guide/Mendel and MH Guide/BRCA unpaired analyses:

- Number of **read pairs (FASTQ)**:
  - Warning threshold < 50000
- Fraction of **targeted bases with minimum real coverage** (20x), (after aligning):
  - Warning threshold: < 96%
  - **Error** threshold: < 80%
- Average **real coverage** (excluding duplicates):
  - Warning threshold: < 100x
- Fraction of **mapped reads** (after aligning):
  - Warning threshold < 95%
  - Error threshold: < 85%
- Fraction of **discordant read pairs** (after aligning):
  - Warning threshold: > 5%

Cases with one or more of the above error or warning conditions are shown in MH Guide with the status “completed with quality warnings” or “completed discordant read pairs warning”. For more details, see the QC section and files for each case that you can download from the **Dashboard** or from within the individual case.

VCF data

Quality warnings and the internal status: “completed with quality warnings” are highlighted by a yellow case data icon in MH Guide. The case status is shown in the **Case history**:

- **Completed (without errors):** although no errors are reported, the analysis may contain variants with mapping warnings (e.g., uniprot-ensembl inconsistencies), these errors are typically of low impact. The mapping warnings are listed in the VCF error and warning log.
- **Completed with a quality warning:** there is at least one variant for which an interpretation and protein mapping error was triggered (e.g., the variant definition in the VCF is incorrect)

Additional Test Result Data

Additional Test Results can be entered by the user using the Additional Test Results feature of MH Guide and MH Guide/Mendel. The corresponding dialog is guiding the user through the input of the test results such as a genetic variant or biomarker. Syntactically accurate entries are processed by MH Guide and MH Guide/Mendel and no warnings and errors are triggered.

Analysis Feedback on completion and QC warnings

Analysis feedback when using the MH Order Portal

Success of the data analysis is shown in the MH Order Portal by the status **“Data analysis complete”**. Failure of a data analysis is reflected in the MH Order Portal by the status “**Data analysis failed**”. For each case, you can download lists of mapped variants and variants with errors from the **Dashboard** in MH Guide and MH Order Portal. Cases that are completed with quality warnings are highlighted by a yellow case data icon on the **Dashboard**.

Analysis Feedback when using SFTP

Information about the order is provided in the customer’s download folder on the MH SFTP in the **“case_start_information”** log file. This log shows various events with a time stamp, for example, confirmation that the internal transfer from the sftp server to the import server has occurred, and that the CaseID.xml was successfully read and interpreted. If there are no errors, this indicates that the CaseID.xml is syntactically right and the order was created successfully.

Analysis Feedback in MH Guide

Cases that are completed with quality warnings are highlighted by a yellow case data icon on the **Dashboard**.

Consequences of limited data quality

If the data quality is below the standards and the requirements as listed in this document, MH reserves the right to:

- Still charge the (patient) case
- Not process the case through MH Guide, MH Guide/Mendel or MH Guide/BRCA
- Re-charge the (patient) case after receiving improved data

If not defined differently in the applicable MSA.