

I. Instructions for filenames:

1. Files can be on:

- USB Hard drive (NTFS/extFAT)
- FTP/sFTP server
- SSH share

For other sharing platform please contact us.

2. File organization into the drive or shared folder should be as follows:

`/ {SAMPLENAME} / {SAMPLENAME}_[*]_[FLOWCELL]_[LANE]_{READ} . {EXT}`

- Fields in {} are mandatory, fields in [] are not. The underscore "_" delimits fields, it can be skipped if some fields are missing: e.g:
`{SAMPLENAME}_[LANE]_{READ} . {EXTENSION}` or `{SAMPLENAME}_{READ} . {EXTENSION}`
- No space, no punctuation or special characters (, . : ; ! " £ \$ % & / () [] ? \ < > { } + * # etc) no non-latin characters, no lexicographic signs on letters (like à â ç etc) are allowed in filenames. Only latin characters and "_", "-" are allowed.
- `{SAMPLENAME}` must match exactly the field Foreign ID on the sample-sheet*. Name is case sensitive, thus capital and small letters should be used in the same way in the Foreign ID and `{SAMPLENAME}`
- `{SAMPLENAME}` should not contain underscores, use "-" instead. Underscore "_" is a delimiter.
- `[FLOWCELL]` flowcell name can be placed here if necessary
- `[LANE]` number can be in the format: 1, 01, 001, L1, L01, L001
- `{READ}` can be in the format 1, 2 or R1, R2
- `{EXT}` extension must be either *fastq.gz* or *bam*. If *fq.gz* please rename it to *fastq.gz*, if *fastq* please compress it with the tool *bgzip*
<http://www.htslib.org/doc/bgzip.html>
- Filenames for the pairs Read 1 and Read 2 must be identical except the `{READ}` field: eg:
`SAMPLE1234_H2LBAABXY_L001_R1.fastq.gz` and
`SAMPLE1234_H2LBAABXY_L001_R2.fastq.gz`
- Every sample should be into a separate folder. Not more than one folder per sample.

3. Please be aware that custom filename formats or unsupported file formats cannot be processed.

4. The SampleSheet `IHG_External_Sample_Template_v2018.11.xlsx` must be prepared *

* see next page for instructions

II. Instructions for Sample Sheet

IHG_External_Sample_Template_v2018.11.xlsx

The field are defined as follows. You don't have to fill the Sample ID (we'll assign it), but the Foreign ID field with the sample name.

Filenames related to the sample should start with Foreign ID followed by and underscore “_”

Field	Description	Mandatory	Field Type
Sample ID	LEAVE IT BLANK	No	freetext
Foreign ID	Your sample ID; The name of the external sequence file (FASTQ/BAM) MUST start with this ID	Yes	freetext
Pedigree	Pedigree identifier	Yes	freetext
Comment	Comment	No	freetext
Sex	Sex of the sample	Yes	defined values only
Affected	Is the sample affected by the disease (Yes/No)	Yes	defined values only
Organism	Organism of the sample	Yes	defined values only
Tissue	Tissue of the sample	Yes	defined values only
Disease	Disease the sample has been diagnosed with (see “Disease” tab for possible values)	Yes	defined values only
Library Type	Type of the library, i.e. genomic, exomic,...	Yes	defined values only
Read Type	Type of the reads, i.e. paired-end, single-end, ...	Yes	defined values only
Exome Assay	If library type is exomic: type of the used exome enrichment kit	Yes if libtype exomic	defined values only

For samples in which a pedigree was sequenced, please assign a pedigree to the family and specify parent relations in the separate table

IHG_External_Sample_Trios_v2018.11.xls** (only for samples with family relationships)

** see next page for instructions

III. Instructions for Trio table

IHG_External_Sample_Trios_v2018.11.xls

Please fill this table only for the samples that have parents included in the dataset. If both parents are included only trios specified here can have trio analysis enabled.

Field	Description	Mandatory	Field Type
Pedigree	Pedigree as in Sampleheet	Yes	text
Foreign ID Child	Foreign ID Child as in Samplesheet	Yes	text
Foreign ID Father	Foreign ID Father as in Samplesheet	Yes (if father present)	text
Foreign ID Mother	Foreign ID Mother as in Samplesheet	Yes (if mother present)	text