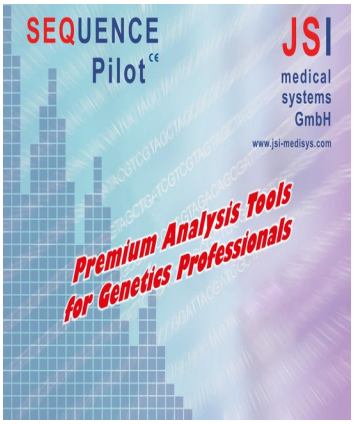
What's new

SEQUENCE Pilot CE

Version 4.2.0
(02/05/2015)



developed by

JSI medical systems GmbH
Tullastr. 18
77975 Ettenheim
GERMANY
phone: +49-7822/440150-0
Fax: +49-7822/440150-20
email: *mail@jsi-medisys.com*

JSI medical systems Corp.
1215 W Imperial Hwy, Suite 205
Brea, CA 92821
USA
phone: +1-714-332-0139
fax: +1-714-332-0131
email: *mail-us@jsi-medisys.com*

web: *www.jsi-medisys.com*

(for research use only)

Table of Contents

1 All Modules....4
1.1 64 bit-database....4
1.2 Automatic Timeout....4
1.3 Lis.ini....4
1.4 Menu System....4
1.5 Menu Help....5
1.6 Operation Users [master file]....5
1.7 Operation Orderlist....5
1.8 Operation Projects [master file]....5
1.9 Operation Joining....5
1.10 Operation Sequence....6
1.11 Reports....6
1.12 Jsi Service....6
2 Modules SeqPatient and SeqNext....7
2.1 Gene Admin....7
2.2 Operation Sequence....7
2.3 Operation Mutation [master file]....7
2.3.1 Transcript ID....7
2.3.2 SeqType genomic or cDNA....8
2.3.3 Select Mutations....9
2.3.4 Other new features....10
3 Modules SeqPatient and SeqHLA....11
3.1 Operation Joining....11
4 Module SeqPatient....11
4.1 Import of Variation DB files....11
4.2 Gene Admin....12
4.3 Operation Joining....12
4.4 Operation Sequence....12
4.4.1 Electropherogram....12
4.4.2 Genes....13
4.4.3 Positions/Resultfiles....14
4.4.4 Variation/Mutation table....14
4.5 Operation Mutation [master file]....15
5 Modules SeqNext and SeqNext-HLA....15
5.1 Automatic zipping of result data....15
5.2 Operation Users [master file]....15
5.3 Operation Run....15
6 Module SeqNext....15
6.1 Genomes....15
6.2 Import of Variation DB files....15
6.3 File UnusedReads.txt....17
6.4 smMIP Processing....17
6.5 Menu Help....18
6.6 Operation ROI [master file]....18
6.6.1 All tabs....18
6.6.2 Filters for background reads....19
6.6.3 Tab add PCR and tab add Panel....20
6.6.4 Tab add PCR/tab Gene....20
6.6.5 Tab add PCR/tab Fasta....21
6.6.6 Tab Add Panel....21
6.7 Set up of Multiplicom MASTR assays....22
6.8 Operation ROI Groups [master file]....23
6.8.1 Import of Pseudogenes from Pseudogenes.org....24
6.9 Operation Run....25
6.9.1 Section Settings....25
6.9.2 Patient table....27
6.9.3 Autorun....27

6.10 Operation Joining/Poollist....28
6.11 Operation Joining....28
6.12 Operation Sequence....29
6.12.1 Order/Protocol/Family/Pool....29
6.12.2 Files, ROI Groups, Genes/Chromosomes and ROIs/Locations....29
6.12.3 Files,Genes/Chromosomes and ROIs/Locations....30
6.12.4 Files....30
6.12.5 ROIs/Location....31
6.12.6 Summary....33
6.12.7 Variation/Mutation table....33
6.12.7.1 Mutation Calling of Insertions, Deletions and Indels overlapping an ROI....33
6.12.7.2 Tabs....34
6.12.7.3 Variation DB display and filter function....35
6.12.7.4 Context menu....36
6.12.7.5 Table headers....36
6.12.8 Electropherogram/Sequences....36
6.12.9 Show....38
6.12.10 Report....38
6.13 Operation Archiving....38
6.14 CNV analysis....38
6.14.1 Operation ROI [master file]....38
6.14.2 Operation Analysis mode CNV [master file]....39
6.14.3 Operation Joining....39
6.14.4 Operation Sequence/CNV window....39
6.15 Operation Pool....40
7 Modules SeqHLA and SeqNext-HLA....40
7.1 Importer....40
7.2 Menu SeqHLA / SeqNext-HLA....40
7.3 Operation Archiving....41
8 Module SeqHLA....41
8.1 Operations Amp modules [master file], SeqPrimer [master file], Sequence....41
8.2 Operation Amp modules [master file]....41
8.3 Operation SeqPrimer [master file]....41
8.4 Operation Joining....42
8.5 Operation Sequence....42
9 Module SeqNext-HLA....43
9.1 Operation ROIs [master file]....43
9.2 Operation ROIs Group [master file]....45
9.3 Operation Run....45
9.3.1 Data....45
9.3.2 Multiple Processing Cores....45
9.3.3 Importer....45
9.3.4 Settings....46
9.3.5 Start a run....49
9.4 Operation Joining....49
9.5 Operation Sequence....49
9.5.1 Files/Genes/ROIs and Locations....49
9.5.2 Location overview....51
9.5.3 Combined, HT1 and HT2 sequence....51
9.5.4 Reads sequences....52
9.5.5 Reads view....52
9.5.6 Matching table....52
9.5.7 Show....53
9.5.8 Validation....53
10 Module MLPA....53
10.1 Operation Joining....53
10.2 Operation Sequence....53
11 Talkmaster....54

1 All Modules

1.1 64 bit-database

All **SEQUENCE Pilot** modules and processes can be provided as 64 bit version now.

Note: For server installations, please contact our **support team before updating your system** to use the 64 bit version.

For single user installations: To use the 64 bit version, you have to change the link of the **SEQUENCE Pilot** desktop icon. Therefore go to the `bin`-directory of your **SEQUENCE Pilot** installation (by default this is `C:\SeqPilot\bin`). Create a desktop link for the file `SeqPilot64.exe`. When the new desktop icon is used, the 64 bit version starts automatically. Otherwise (with the default desktop icon) the 32 bit version is started. In case you do not want to use the 32 bit version anymore, please remove the default desktop icon.

1.2 Automatic Timeout

For server installations (Standalone=0) there is an automatic timeout after 15 minutes (in case no other timeout is defined in the `lis.ini`-file).

1.3 Lis.ini

The `lis.ini`-entry "`ProjectSelect=`" can not only be set in section `[LIS]` but for all modules separated now. With this entry a project can be pre-selected in the section *Select Orders* of *Joining, Worklist* and *Orderlist*. Therefore make the entry into the section of the module in the `lis.ini`-file. This is:

- `[SeqPilot]` for modules **SeqPatient** and **SeqHLA**
- `[SeqNext]` for module **SeqNext**
- `[SeqNextHLA]` for module **SeqNext-HLA**

1.4 Menu System

Protocol Manager:

Using the Protocol Manager you can select, what is shown in the Protocol.
All events that can be shown are listed in the column *Events*. There are several columns that can be changed for each event:

- *Protocol: on/off:* event is listed/not listed in the Protocol respectively.
- *Show default: on/off*: All events that are default events can be searched for in the Protocol. When the Protocol is opened in the section *Select Protocol* the event *default* can be searched for. When this is set, only events that are set to *Show default* "on", are listed in the Protocol.
- *Show protocol tab: on/off*: event is listed on the tab *Protocol* (operation*Sequence*).
- *Colour R,G,B*: the event can be highlighted in color. By default the events "set/delete order state MV", "set/delete gene state MV" and "change base" are highlighted.
- *Tooltip*: A tooltip for the event can be shown in the tab *Protocol* (operation *Sequence*). By default a tooltip is shown for the event "change base".
- *Shortcut*: In this column a shortcut for an event can be entered. In case a shortcut is entered it is shown on the tab *Protocol* (operation *Sequence*) in the column event (instead of the complete event)
- *Modules*: This column can not be changed. In case an event is not available for all modules, the module name is listed here.

Order Info:

Order Info gives a statistic/overview about order states.
You can search for a Date/Date range and Module(s) in the section Select Orders. It is then shown how many orders were loaded, archived or extracted for the selected date range and module. For active orders the number of complete, incomplete, technical and medical validated orders is shown.

Task Scheduler (only for Client/Server installations):

With the *Task Scheduler*, you can set time intervals to automatically archive and/or extract your orders regularily. The tabs *Archive* and *Extract* are available to specify the time intervals and orders that should be archived/extracted.

Note: This feature can only be used if the User Right *edit Scheduler* is active.

For further information please have a look at the **User Manual**.

1.5 Menu Help

There are several new items available, which explain possible Warnings for **SeqNext-HLA** and/or **SeqNext:**

- *SeqNext_MutationWarnings*: Warnings that occur in column *Mutations* of the sections *Files, ROI Groups, Genes/Chromosomes* and *ROIs/Locations*.
- *SeqNext_SeqNext-HLA_CoverageWarnings*: Warnings that occur in column *Coverage* in sections *Files, ROI Groups, Genes/Chromosomes* and *ROIs/Locations*.
- *SeqNext-HLA_Warnings*: Warnings that occur in column *Warnings* in sections *Files, ROI Groups, Genes/Chromosomes* and *ROIs/Locations*.
- *SeqNext_UnusedReadsErrorCodes*: Explains possible error codes for unused reads.

1.6 Operation Users [master file]

For Client/Server installation the new User Right *edit scheduler* is available. If active, the *Task Scheduler* (Menu *System*/item *Task Scheduler*) can be used.

1.7 Operation Orderlist

- Extract archived orders: When the button *[Extract]* is pressed, you can decide if you want to extract all orders or the selected orders by pressing *all* or *select*.
- In the search field *Module* the new option “undefined module” is available.

1.8 Operation Projects [master file]

A search function is available for projects. You can search for projects with a certain *Name, Comment* or *State (active, inactive)*.

1.9 Operation Joining

The new `lis.ini` entry `JoinRFValidatedOrder=yes` is available for all modules now. For module **SeqPatient** and **SeqHLA** the entry has to be set in section `[SeqPilot]`, for all other modules in the section of the module, e.g. for module **SeqNext** use the section `[SeqNext]`.

The following cases can be present:

- a medically validated order is present and additional result files are loaded for the patient: the result files are joined to the order and the order state is reset to *complete*. The state of result

files that are not affected remains *MV*.

- an archived order is present and additional result files are loaded for the patient: the order is reactivated. The same procedure as described for medically validated orders is done.
- an extracted order is present and additional result files are loaded for the patient: the order is reextracted and reactivated. The same procedure as described for medically validated orders is done.

1.10 Operation Sequence

After medical validation (button *[MV]*), the button *[TV]* for technical validation is inactive.
When removing medical validation, you can optionally be asked, if you are sure to remove the medical validation. For this option make the following entry in the `lis.ini`-file in the section of the module: `CheckMVReset=yes`.
This is:

- `[SeqPilot]` for modules **SeqPatient** and **SeqHLA**
- `[SeqNext]` for module **SeqNext**
- `[SeqNextHLA]` for module **SeqNext-HLA**
- `[MLPA]` for module **MLPA**

1.11 Reports

Note: All Reports have been adapted. In case you want to use the new layout, please contact our support team.

Reports can be saved as `pdf`-files automatically instead of printed, when the button *[Print]* is pressed. (If this option is set, you can print your Report using *[Preview...]*).

Therefore make the following entries in the `lis.ini`-file/section `[Print]`, located in the `bin`-directory of your **SEQUENCE Pilot** installation:

A default folder must be defined to save the all Report-files:

```
ExportDir=../Reports
```

The file name of the `pdf`-file must be defined as follows:

The filename can be created out of DNANo or OrderNo. Furthermore a free string can be added to the DNA/OrderNo. Therefore make the following entries:

```
PrintExportFile=
```
possible entris are: `DNANo` or `OrderNo`

```
ExportType=PDF
```

```
ExportSuffix + strModule=_Text
```
e.g. entry for module MLPA to add the suffix "_MLPA" to the DNA/OrderNo: `ExportSuffixMLPA=_MLPA`

1.12 Jsi Service

The processes from the `jsiSevice.ini` file (e.g. `SeqResultfiles.exe`) can now be started and ended separately in the Jsi Service window "Status".

2 Modules SeqPatient and SeqNext

2.1 Gene Admin

- In the field *Gene* you can search for genes by entering letters.
- In case you use splitted genes, please set the `lis.ini` entry `[SeqPilot] LoadAlternativeVersion=yes`.
- It is possible to edit stop codons and to switch off the translation in downloaded gene files. Therefore open the gene file (located in the folder `GeneFiles` of your **SEQUENCE Pilot** installation):
  - Switch off the translation (e.g. for untranslated genes): Enter the following into the section of the untranslated transcript(s): `/UNTRANSLATED="yes"`
  - Edit stop codons: To overwrite a stop codon, make the following entry: `/STOP_CODON="ExonIndex,Pos in the Exon"`, e.g. `/STOP_CODON="27,607"`.

    This new line has to be entered in the section `mRNA`. Make sure to enter space characters or tabs identical to the other entries in this section.

2.2 Operation Sequence

Variation/Mutation table:

- The new column *Transcript* is available which lists the used transcript.
- For the `VCF`-export (context menu entry *export>tab to VCF*) you can select, if you want to export the file for CARTAGENIA or for other platforms. Therefore a new window opens, where you can select a path for the output-file (field *Output*) and the types "*General"* or "*CARTAGENIA"* (field *Type*).
- For big insertions/deletions (>30 bp) all mutation data is available. The mutations can be added to the mutation database.
- HGVS nomenclature:
  - The column *HGVS nomenclature* was renamed into *c.HGVS*.
  - The HGVS p.-nomenclature is shown in the new column *p.HGVS*.
- In case the *NucName* and the *AAName* given in the Mutation database differs from the entries in column *c. HGVS* and *p. HGVS*, the entries from the mutation database are listed in the field *mut Entry* in parenthesis.
- For mutations with the *Hint* "not detected" no zygosity is shown in column *Nuc change*.

Report:

For calculation of the position of a mutation/variation, previous mutations (e.g. insertions, deletions) are not regarded any more. Each mutation/variation is regarded "standalone" (in relation to reference sequence).

2.3 Operation Mutation [master file]

2.3.1 Transcript ID

The transcript ID is now available for entries of the *Mutation database*. For all newly added mutations the transcript ID is listed in the column *Transcript* automatically. Moreover the field *Transcript* is available in the dialogue *Mutation.*

Note: For all mutations present in the mutation database already, the transcript is "unknown". It is obligatory to set transcript IDs for all entries in the mutation database before analysing new orders or adding new mutations to the database.

Otherwise:

- **for newly analysed orders mutation database information will be missing.**
- **equal mutations will be present with known transcripts and unknown transcripts and the mutation will be listed twice.**

To set transcripts you have the following options:

- You can set transcripts for all mutations of one or several gene(s): Therefore select one or several gene(s) in the *Gene* table, open the context menu and select *set transcript*.
- You can set a transcript for one or several mutations: Therefore select the mutation(s) in the *Mutation* table, open the context menu and select *set transcript.*

For both options the new dialogue *Transcript* opens:

- In the field *Operation* you can choose to "*automatically set unknown transcripts*", or to "*manually set selected transcript*". For the first option unknown transcripts are set automatically. Note: No transcript can be set in case the mutation was found in orders with different transcripts, in case the gene file was deleted or mutations were not found in real orders (e.g. imported mutations).

  For the second option you can select a transcript that should be set from the genome or gene file.

In case new mutations were added already before setting transcripts for the old entries mutations can be merged after transcript IDs are set.

To merge all mutations for one or several selected open the context menu in the *Gene* table and select *merge...*

The following options are available in the field *Operation*:

- *merge unknown transcript*: in case equal mutations are present, one with a transcript and one with an unknown transcript, the mutations are merged. Only one entry remains, inclunding all information (such as *Orders, Frequency...* of both previous entries).
- *merge equal transcripts*: mutations are only merged in case they are equal (also the transcript has to be equal).
- By pressing *[OK]* transcripts are set and depending on your settings mutations are merged.
- After merging you will get a message showing how many mutations were merged.

Moreover you have the option to merge two mutations in the *Mutation table*. Therefore select the mutation(s) in the *Mutation table*, open the context menu and select *merge mutations >.* You can select which mutation should be added to the other one (Index context menu). This will only make a difference in case columns like *AA Name* are not equal. In case the column *Transcript* is unknown or empty for one mutation and the other mutation has a known transcript the transcript will always be set during merging.

2.3.2 SeqType genomic or cDNA

For each mutation it is listed in which SeqType (genomic or cDNA) the mutation was detected. Therefore the new column *genomic/cDNA of total* is present in the *Mutation table*. For all newly added mutations the count in this column is increased automatically. Moreover the SeqType information is available in in dialogue *Mutation* in the field *g./cDNA*.

For mutations that are present in the database already, the following options are available to set the

count in column *genomic/cDNA of total* for already existing orders:

Note: It is recommended to set the *SeqType* for all entries in the mutation database.

- You can set the *SeqType* for all mutations of one or several gene(s): Therefore select one or several gene(s) in the *Gene* table, open the context menu and select *set genomic/cDNA....*
- You can set the *SeqType* for one or several mutations: Therefore select the mutation(s) in the *Mutation* table, open the context menu and select *set genomic/cDNA....*

For both options the following operations are available:

- *automatically set genomic/cDNA*: the program checks which orders are present for genomic and cDNA sequences and sets the count automatically.
- *manually set all to genomic*: all orders are expected to be genomic data, the genomic count is increased depending on the number of orders.
- *manually set all to cDNA*: all orders are expected to be cDNA data, the cDNA count is increased depending on the number of orders.

2.3.3 Select Mutations

When one or several genes are selected in the *Gene table*, by default all mutations for these are shown in the *Mutation table*. This new section is available to search for mutations using special search items. If *[Search]* is pressed, only mutations that fullfill the search items are listed in the *Mutation table*.

The following search items are available:

- *Transcript*: Only mutations of the selected transcript are listed.
- *Location*: Only mutations of the entered location are listed.
- *Mut Effect*: Only mutations with the entered *Mut Effect* are listed.
- *Frequency*: Values can be entered to search for mutations that are equal, above or below a certain *Frequency*. The values are entered with “>” or “<” or no sign (equal) in front in the hierarchy homo”/”hetero “of” total.

  Examples:

  - >5/: search for homozygous mutations with a frequency > 5
  - /10: search for heterozygous mutations with a frequency equal 10
  - / of <20: search for mutations with a total number of archived orders < 20
  - >10/<5 of >30: search for mutations, which have a frequency homozygous > 10, heterozygous < 5 and a total number of archived orders > 30
  - >20 of >100: search for mutations whith a sum of heterozygous and homomzygous frequency > 20 and a total number of archived orders > 100.
  - 10 of >40: search for mutations whith a sum of heterozygous and homomzygous frequency equal 10 and a total number of archived orders > 40.
- *Date*: List mutations with the selected “Changed date” only. The *Changed date* is listed in the dialogue *Mutation*.
- *Date Range:* List mutations with the selected “Changed date” range only. The *Changed date* is listed in the dialogue *Mutation*.
- *Type*: List mutations of a special type only (*C*: base change, *D*: deletion, *I*: insertion, *Indel*: indel)
- *g./cDNA*: Search for mutations found in a special *Seq Type* (genomic or cDNA), above or below a certain value. The values are entered with “>” or “<” or no sign (equal) in front in the hierarchy genomic DNA”/”cDNA “of” total.

Examples:

- >5/: search for mutations found in genomic DNA in more than 5 orders
- /10: search for mutations found in cDNA in 10 orders
- >5/>6: search for mutations found in genomic DNA in more than 5 orders and in cDNA in more than 6 orders.
- / of <20: search for mutations found in less than 20 orders (cDNA and/or genomic DNA)

- *Disease No*: Lists mutations with a special *disease number* only. The *disease number* is shown in the dialogue *Mutation*.
- *Ethnicity*: Lists mutations with a special *Ethnicity* only. The *Ethnicity* is shown in the dialogue *Mutation*.
- *Organ*: Lists mutations with a special *Organ* only. The *Organ* is shown in the dialogue *Mutation*.
- *Phenotype*: Lists mutations with a special *Phenotype* only. The *Phenotype* is shown in the dialogue *Mutation*.

2.3.4 Other new features

- Big insertions/deletions (>30 bp) can be added to the mutation database. All sequence information is available now in the fields *Nuc Change, Nuc Name AA Change* and *AA Name*

  Note: In case entries from older SeqPilot versions exist for big deletions/insertions not all sequence information is available. When new mutations are added they are listed as a second mutation, because the fields *Nuc Change, Nuc Name, AA Change* and *AA Name* differ. To have one entry again for equal mutations select both mutations in the *Mutation table* and use the context menu item *merge*.
- You can search for a gene by entering letters using the search field *Gene*.
- The following new items are available in the context menu of the *Gene* table for one or several selected gene(s):
  - *delete mutations without orders*: Mutations with no Orders listed in section *Orders* (e.g. imported mutations) are deleted.
  - *set transcript*: a transcript can be set (see chapter 2.3.1)
  - *set genomic/cDNA*: the *SeqType genomic/cDNA* can be set (see chapter 2.3.2)
- The following new items are available in the context menu of the *Mutation* table for one or several selected mutation(s):
  - *set transcript*: a transcript can be set (see chapter 2.3.1)
  - *set genomic/cDNA*: the *SeqType genomic/cDNA* can be set (see chapter 2.3.2)
- The new columns *Changed Date* and *User* are available in the *Mutation table:*
  - *Changed Date*: date when the mutation was changed last (in the dialogue mutation).
  - *User*: User ID of the user who added or changed the mutation.
- Dialogue *Mutation (context menu show mutation)*:
  - The fields *NucName* and *AAName* are in read only mode. They can be changed using the button *[Set]* behind the corresponding field.
  - Button *[Reset] Frequency*: When this button is pressed you are asked, if you are sure that you want to reset the frequency. If *[Yes]* is pressed, the frequency is reset.
- *Orders table*: new column *Module* is available which lists the module that detected the mutation (*SeqNext* and/or *SeqPatient*). The information is only present for newly added orders (not for existing orders).
- In the *Mutation dialogue* as well as in the *Orders table*, the sorting of orders begins with the

latest order.

- Several genes can be selected simultaneously for export (button *[Export]*).

3 Modules SeqPatient and SeqHLA

3.1 Operation Joining

The search for orders is quicker (HLA database/genefiles are not loaded).

4 Module SeqPatient

4.1 Import of Variation DB files

Note: DB SNP imported in the previous SeqPatient versions, does not work any more. To have SNP/known mutation information available again, new databases have to be installed as described below.

A database package is available to install the following databases:

- **dbSNP**: Short Genetic Variations (http://www.ncbi.nlm.nih.gov/SNP/)
- **1000 Genomes**: Catalog of Human Genetic Variation (http://www.1000genomes.org/)
- **ClinVar**: Sequence Variation and its relationship to human health (http://www.ncbi.nlm.nih.gov/clinvar/)
- **CLINVITAE**: Clinically-observed genetic variants (http://clinvitae.invitae.com/)
- **COSMIC**: Catalogue of somatic mutations in cancer (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/)

These can be used to show known SNP information and for filter options (see operation *Sequence*).

For the genomes available on our homepage we offer an `exe`-file, for easy installation of all Variation databases. There are two installation packages we offer for download, one referring to **hg19** the other one referring to **hg38**; please select the correct one for your installation (depends on the installed genome - **hg19** or **hg38**); there is also the possibility to install both packages at the same time; please keep in mind, that in this case both genomes (**hg19** and **hg38**) have to be installed.
The database package can either be downloaded from our homepage or alternatively from our ftp-server. We recommend to use the ftp server for the download of big files, because the download can be restarted in case there is an interruption.

Download from our homepage:

- Go to http://www.jsi-medisys.de/genomes-snp-dbs
- Download the file *hg19-GenomeVarDB* and/or *hg38-GenomeVarDB.*
- After download, please verify the integrity of the downloaded file, i.e. whether it is complete and uncorrupted. To do so, you may use the md5 checksum provided to the right of the respective link. Search for 'md5 checksum' on the web to find suitable verification tools. Within the verification tool, browse and select the downloaded file and let the tool compute the checksum. Compare it to the one provided on our homepage. If the two checksums differ, you have to download the file again.

Download from our ftp server:

- You can reach the file server using the following link: ftp://ftpsrv.jsi-medisys.de

- Please use the following login:

  Username: `SeqPilotData`

  Password: `SeqPilot`
- Download the file *hg19-GenomeVarDB.exe* from folder *GenomeVarDB/hg19* and/or file *hg38-GenomeVarDB.exe* from folder *GenomeVarDB/hg38.*
- The md5 checksums provided with the download links on the website may be used for the respective file on the file server, too (see above).

To install the databases:

- Execute the file *hg19-GenomeVarDB.exe* and/or *hg38-GenomeVarDB.exe* and follow the installation instructions. As destination directory `C:/SeqPilot/GeneFiles` is proposed automatically. In case this is not the folder of your **SeqPilot** installation, please change the link to the `GeneFiles` folder of your **SeqPilot** installation.
- Note: In case you do not want to create the database folder in your **SeqPilot** installation but in another directory on your PC you have to enter the path of the directory in the `lis.ini`-file, section `[SeqPilot]`:

  ```
  GenomeVarDir=Path (e.g. C:\GenomeVar)
  GenomeDBVarDir=Path (e.g. C:\GenomeDBVar)
  ```

4.2 Gene Admin

In case a gene file with an entry in *Gene Admin* can not be found in the `GeneFiles` folder of your installation (no `txt`-file available or `txt`-file is corrupted) and the genome *hg19* is installed, *hg19* sequences are present in *Gene Admin*. With these no SeqPrimers and AmpModules should be defined, since all transcripts are active for *hg19*. The first transcript listed in the *Isoforms table* is always used.

4.3 Operation Joining

The context menu item *settings...* (*Upper table*) is available for several selected files. The settings (*Gene, Amp Module, Seq Primer* and *DNA No.*) can be changed for all selected files simultaneously.

4.4 Operation Sequence

4.4.1 Electropherogram

- Below the location overview a "*peak height ratio diagram*" for each base position is available. The bars show the background for each position. In case there are no or only small bars above/below the middle line, there is no high background. In case a background is present, there is a blue bar above the line for forward bases and a purple bar below the line for reverse bases:

  The bars are calculated as follows:

  peak height ratio = peak area highest not reference bases/(peak area highest not reference base + peak area reference base)
- There is a context menu available in the *peak height ratio diagram* to show the length of the result files below the location overview:
  - *show peak height ratio* (default setting): the *peak height ratio diagram* is shown below the location overview.
  - *show result files*: the length of the result files is shown below the location overview.

Reverse sequences are highlighted in a darker color.

- *show original result files*: the length of the original result files (not shortened due to quality, SeqPrimers...) are shown. Sequence parts that are not used for analysis are highlighted grey.

- The detection of frameshift mutations was improved, especially for more complex cases. It is now possible to detect SNPs after/before a frameshift. The new mutation type *Indel* is therefore available.
  The new context menu item *split sequences* is available in the combined sequence. Using this, the combined sequence is split into two homozygous sequences. Moreover each result file sequence is split into two homozygous sequences. The two homozygous sequences are virtual sequences, representing the expected sequences, if the two alleles would have been sequenced separately. This gives a better overview to resolve more complex sequences. In this view only mutated positions are marked red.

  You can display the original view again by selecting the context menu item original *sequences* in the second combined sequence.

  The split sequence view can also be shown in a separate window. Therefore select the context menu item *result file view* > *split alignment* for one of the result file sequences.

- The electropherograms can be shown in a separate window. This gives a better overview in case many sequences are present for one location. Therefore the new context menu item "*result file view"* is available for the *result file sequences.* You have the option to either show the *"original alignment"* or the *"split alignment"* in the new window. Therefore use the combo box "*original/split*".

- Edit Bases (context menu *edit bases...* in the result file sequence):
  - The context menu in the *gene line* changed to make editing of frameshift mutations more convenient. There is the context menu item *2nd allele...* available. A new window opens to select different types:
    - By default the type *Auto* is selected. When *[OK]* is pressed a frameshift mutations is searched automatically and shown in a new line below the *gene line*.
    - *Deletion*: Here the base sequence of a deletion or the number of deleted bases can be entered. The resulting sequence including the mutation is shown in a new line below the *gene line* after *[OK]* is pressed.
    - *Insertion*: Here the base sequence of an insertion or the number of inserted bases can be entered. The resulting sequence including the mutation is shown in a new line below the *gene line* after *[OK]* is pressed.
    - *Indel*: Here the base sequence of an Indel or the number of inserted/deleted bases can be entered. The resulting sequence including the mutation is shown in a new line below the *gene line* after *[OK]* is pressed.
  - The context menu item *Heterozygote start (toggle)* is not available anymore.

- New context menu item *show family* for the result file sequences: shows all result files belonging to one family in the new *result file view*.

- The following new count modes are available in the combo box count mode:
  - *Gene abs*: absolute position in the gene
  - *Genome Position*: genomic position

4.4.2 Genes

The following new context menu items are available:

- *editing* >
  - *frame shift analysis on/off:* Use this to switch the frameshift analysis on or off. If switched off in case of a frameshift each position is regarded separated, many heteroygous

positions will appear.
Note: In case the frameshift analysis algorithm was used automatically, there is the condition *F* in the column *Condition*/section *Position/Resultfiles.*

- *Indel Gaps...*: The new mutation type *Indel* is present in the *Variation/Mutation table*. By default the maximum number of bases between two base changes or a base change and an insertion/deletion to consider them as an *Indel* is “3”. This number can be changed using this context menu.

4.4.3 Positions/Resultfiles

- The new condition “F” is available in the column *Condition*. This condition is present in case the new frameshift analysis algorithm was used automatically.

  Note: in case a Resultfile is set to “ignore” automatically and the *Condition* “F” is present, the exon might have been ignored due to a frameshift.
- The following new context menus are available:
  - *editing > frame shift analysis on/off:* using this the new frame shift analysis algorithm can be switched on or off.
  - *export > 2nd allele sequence...*: The sequence of the second allele can be exported as a `seq`-file. The second allele is the non wildtype allele. This is helpful to copy sequences with frameshift mutations.
  - *copy > 2nd allele sequence...*: The sequence of the second allele can be copied.
  - *requirenment > cancel*: To cancel requirements all open requirements can be selected simultaneously now. The second allele is the non wildtype allele. This is helpful to copy sequences with frameshift mutations.

4.4.4 Variation/Mutation table

- The table can either be shown for the gene selected in section *Genes* or for all analysed genes. Therefore use the radio button *all Genes/selected Gene* above the *Variation/Mutation table*.
- The new mutation type "*Indel*" is available. This defines several mutations lying very close to each other: This can either be two/several base changes or base change and deletion/insertion. By default the maximum number of bases between two base changes or a base change and an insertion/deletion to consider them as an *Indel* is 3. This gap can be changed in the `lis.ini`-file, located in the `bin`-directory of your installation: Therefore enter the number of bases behind the following entries in section `[SeqPilot]`:
  - for base changes: `DelInsGapSNPToSNP=`
  - for base change/insertion/deletion: `DelInsGapSNPToInDel=`

  Note: This setting can also be changed for each gene separated. Therefore use the context menu *editing > InDel gaps...* in the section *Genes*.
- Imported SNP DBs (ClinVitae, COSMIC, ClinVar, 1000 Genomes, dbSNP) can be used for display and filtering.

  The following context menus are available for filtering and display:
  - *show table > DBs...:* Information for mutations present in the databases can be shown in additional columns of the *Variation/Mutation table*.
  - *filter > DB...:* Mutations can be filtered. For filtering, special internal filters can be used as as well as imported databases. The new tab *filter* is available: it lists mutations that were filtered using the SNP database filter. Mutations on this tab are not listed on tab *all*. It is possible to print this tab on the Report by selecting the tab filter in the *Print/Preview*.

  Please have a look at our **User Manual SeqPatient** for detailed instructions.

- The following new context menu items are available:
  - *ignore (toggle)*: If selected the mutation will be ignored and not printed in the report. The line will be highlighted grey. You can remove the ignore by selecting the item again.
  - *move to other:* Is available for distinct mutations only. Moves the mutation to tab *other*.
  - *move to distinct:* Is available for other mutations only. Moves the mutation to tab *distinct*.
  - *export > tab to VCF*: **Note**: This only works in case the gene files *(Gene Admin*) were mapped to a genome prior to your analysis.

    The selected tab is exported as a `VCF`-file. You can select, if you want to export the file for **CARTAGENIA** or for other platforms. Therefore a new window opens:
    - *Output*: With the button *[…]* you can select a path for the output-file.
    - *Type*: select the types "*General"* or "*CARTAGENIA"*. Choose *"CARTAGENIA"* in case you want to import the `VCF`-file into CARTAGENIA.

4.5 Operation Mutation [master file]

In case several isoforms are active and a sequence is present, which does not belong to the main isoform, the cDNA nomenclature is calculated based on the another active isoform.

5 Modules SeqNext and SeqNext-HLA

5.1 Automatic zipping of result data

After starting a Run the data is zipped automatically. This makes the result data much smaller (approximately 10 % of the unzipped result data).

5.2 Operation Users [master file]

The new user right *User is authorised to edit profile settings* is available. This right is active by default. In case it is inactive, the user can not edit the *Settings/Profiles* in the operation *Run* and in operation *ROI [master file]*.

5.3 Operation Run

- `gz`-files (Illumina) can be loaded now without unzipping.
- Loading of files: Button *[…]*: For Client/Server installations several files can be selected simultaneously now.

6 Module SeqNext

6.1 Genomes

On our homepage we offer *hg38* for installation. Therefore go to http://www.jsi-medisys.de/genomes-snp-dbs.

6.2 Import of Variation DB files

Note: The dbSNP imported in the previous SeqNext versions, does not work anymore. To have dbSNP information available again the new `zip`-files for SNP databases has to be installed as

described below.

An installation package including the following databases is available for genome *hg19*:

- **dbSNP**: Short Genetic Variations (http://www.ncbi.nlm.nih.gov/SNP/)
- **1000 Genomes**: Catalog of Human Genetic Variation (http://www.1000genomes.org/)
- **ClinVar**: Sequence Variation and its relationship to human health (http://www.ncbi.nlm.nih.gov/clinvar/)
- **CLINVITAE**: Clinically-observed genetic variants (http://clinvitae.invitae.com/)
- **COSMIC**: Catalogue of somatic mutations in cancer (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/)

These can be used to show known SNP information and for filter options (see operation *Sequence/Variation/Mutation table*).

For the genomes available on our homepage we offer an `exe`-file, for easy installation of all Variation databases. There are two installation packages we offer for download, one referring to **hg19** the other one referring to **hg38**; please select the correct one for your installation (depends on the installed genome - **hg19** or **hg38**); there is also the possibility to install both packages at the same time; please keep in mind, that in this case both genomes (**hg19** and **hg38**) have to be installed.
The database package can either be downloaded from our homepage or alternatively from our ftp-server. We recommend to use the ftp server for the download of big files, because the download can be restarted in case there is an interruption.

Download from our homepage:

- Go to http://www.jsi-medisys.de/genomes-snp-dbs
- Download the file *hg19-GenomeVarDB* and/or *hg38-GenomeVarDB.*
- After download, please verify the integrity of the downloaded file, i.e. whether it is complete and uncorrupted. To do so, you may use the md5 checksum provided to the right of the respective link. Search for 'md5 checksum' on the web to find suitable verification tools. Within the verification tool, browse and select the downloaded file and let the tool compute the checksum. Compare it to the one provided on our homepage. If the two checksums differ, you have to download the file again.

Download from our ftp server:

- You can reach the file server using the following link: ftp://ftpsrv.jsi-medisys.de
- Please use the following login:

  Username: `SeqPilotData`

  Password: `SeqPilot`
- Download the file *hg19-GenomeVarDB.exe* from folder *GenomeVarDB/hg19* and/or file *hg38-GenomeVarDB.exe* from folder *GenomeVarDB/hg38.*
- The md5 checksums provided with the download links on the website may be used for the respective file on the file server, too (see above).

To install the databases:

- Execute the file *hg19-GenomeVarDB.exe* and/or *hg38-GenomeVarDB.exe* and follow the installation instructions. As destination directory `C:/SeqPilot/GeneFiles` is proposed automatically. In case this is not the folder of your **SeqPilot** installation, please change the link to the `GeneFiles` folder of your **SeqPilot** installation.

- Note: In case you do not want to create the database folder in your **SeqPilot** installation but in another directory on your PC you have to enter the path of the directory in the `lis.ini`-file, section `[SeqPilot]`:

  ```
  GenomeVarDir=Path (e.g. C:\GenomeVar)
  GenomeDBVarDir=Path (e.g. C:\GenomeDBVar)
  ```

6.3 File UnusedReads.txt

All reads that were neither mapped nor aligned are written to a file. The files "`UnusedReads.txt`" are generated for the complete File and for each ROI.

- *Complete File*: The `UnusedReads.txt` locates within the folder *SeqNResults* in the corresponding year.month and Run ID folder.

  The same information can be opened in operation *Sequence*. Therefore use the new context menu item *show > unused reads* in section *Files*.

- ROI: For each ROI all reads that were not aligned are written to the `Unsused Reads.txt` file. It locates within the folder *SeqNResults* in the corresponding year.month, Run ID and ROI folder.

  The same information can be opened in operation *Sequence*. Therefore use the new context menu item *show > unused reads* in section *ROIs*.

Note: The Run ID can be found in the dialogue *show Info* (operation *Joining*/context menu of the Lower table or operation *Sequence*/context menu of section *File*) behind the entry ID.

6.4 smMIP Processing

Method:

In smMIP (single molecule Molecular Inversion Probes), a molecular tag is added to one of the reads of a read pair to be able to assign them to an individual capture event. If a tag occurs more than once for a specific library (amplicon), either a consensus can be built to remove (random) sequencing errors, or only one of the reads is considered for the read depth (coverage) within this amplicon. Thus, it is possible to detect low-frequency or subclonal variation.

Method in **SeqPilot**:

a) When starting a run in operation *Run*, smMIP processing has to be enabled via three variables within the settings.

1. On tab *Trimming*, the length of the molecular tag (e.g. 10) must be given in the field *Remove Bases 5'* (*Remove Bases 3*' must stay empty).
2. On tab *Expert Settings* the boxes *Unique reads only* and *Compl. reads only* must be checked.
3. On tab *Expert Settings*, the boxes *Ignore paired-end info* and *Allow unique paired-end reads* must not be checked.

Given this configuration, a note should occur on tab *Expert Settings*: "smMIP processing enabled with tag length: 10".

Note: Exactely two files have to be loaded in operation *Run* (R1 and R2-files). The R2-file (the file with the reads including the molecular tag) has to be the second one listed in section *Files*/operation *Run*.

b) Processing:

1. Reads are mapped to ROIs, aligned to their respective reference sequences and then assigned to amplicons within the ROIs.

2. Single reads (their paired read is missing, i.e., has not been mapped to the same ROI or amplicon, or does not fullfill the quality filters, are discarded. The discarded reads are added to the number of ignored reads (see Error code 55 in the context of *Unused Reads*)
3. The molecular tag is identified within the read pair and all duplicate read pairs (with respect to an amplicon) are discarded (keeping those base sequence which occurred the most within the group of duplicates). These reads are added to the number of duplicate reads (see Error code 58 in the context of *Unused Reads*).

c) Summary window -> tab *Amplicon*

An additional column *Dupl. Reads* has been added on tab *Amplicon,* showing the amount of reads that have been discarded in processing step 3. The absolute value is followed by a percentage, indicating the amount of duplicates within the set of "complete" reads (which is the sum of "Aligned Reads" and "Dupl. Reads").

6.5 Menu Help

New item *UnusedReadsErrorCodes:* You can open a list of all possible “error codes” for unused reads (present in `UnusedReads.txt` and operation *Sequence* sections *Files* and *ROIs/Locations*/context menu item *show* > *unused reads*).

6.6 Operation ROI [master file]

6.6.1 All tabs

- New column *Restrict* in the *ROI List:*

  Note: the column *Restrict* is inactive by default. Therefore the ROIs (also old ROIs) can be shown in the extended view in operation *Sequence* for new Runs.

  There are three differences present between restricted and extended (not restricted) ROIs.

  1. Not restricted ROIs can be shown in the extended view in operation *Sequence*: ROIs are shown in an extended view, when the *combo box combined* (*Sequences/Electropherogram*) is switched to *combined ext* or *fwd+rev ext.* The ROI is extended at the 5' and 3' end. The extension is as long as the longest reads, that overlaps with the ROI. Moreover the Coverage in the extended area must reach the *Required coverage/Min abs coverage* (set in the *Settings* in operation *Run*).
  2. In the extended area no mutation calling is done. Exception are indels/dels that partly lie within the ROI but reach over the ROI beginning/end (see chapter 6.12.7.1. for detailed information):
     - for extended ROIs (no amplicons defined) the complete mutation is called, also the part that lies outside the ROI.
     - for restricted ROIs the deletion/indel is not called completely. Only the part that lies within the ROI is called. In this case there is a “*” sign bin front of the deletion. No amino acid change is calculated.
  3. The copy number of the reads (number in front of each read in operation *Sequence* and the *Reads view*) is also influenced: Identical reads are listed only once, the copy number is listed in front of the reads. Identical means the following:
     - for restricted ROIs the reads must be identical within the ROI or amplicon borders.
     - for extended ROIs the complete read sequence that can be aligned to the extended ROI must be identical. This is the sequence that is shown in the *Reads view* when o*rg. Reads* is active. For this reason for extended ROIs reads can have a lower copy number as for restricted ROIs.

  The column *Restrict* can be changed for all or selected ROIs simultaneously by using the

context menu in the *ROI list*. The new entry *Restrict* > *set/remove* > *all/selected* is therefore available.

- Settings can be selected for each ROI. In the new column *Settings Profile*, an existing Profile, that should be used for this ROI can be selected. Alternatively the item *settings...* can be selected in the context menu for the ROIs. When a Profile is created or selected, the selected profile is set in the column *Settings Profile* automatically. The selected Settings are always used for the analysis of the ROI, even if another setting is selected in operation *Run*.

  A profile can be entered for several ROIs simultaneously. Therefore select several ROIs in the ROI List and use the context menu item *settings...* to set a profile.

- The following new search fields are available:

  1. *Suffix*: here you can search ROIs with a certain *Suffix*. Therefore press *[Search]* behind *Suffix*.
  2. *Source*: Here you can search for ROIs with a certain *Source*. The following sources are possible: *gene file*, *genome, chromosome range* and *fasta file*.

- The column *Active* can be changed for all or selected ROIs simultaneously with the context menu in the *ROI List*. Use *Active* > *set/remove* > *all/selected*.
- The field/column *Category* was renamed into *Panel*.
- The new column *CNV probe type* lists the CNV probe type. For control ROIs this entry can be set to *Control* (blanc means Target ROI).
- An ROI (including all information such as *Skipped parts, Skipped Sequences, Homologies (only mapping)* and *Amplicons/PCR Primer*) can be copied: Use the context menu in the *ROI List* and press *copy...*. The dialogue *New Panel/Suffix* opens, where you can select a *Panel* and a *Suffix* (added to the ROI Name).
- The section *Ignored Parts* was renamed into *Skipped Parts*.
- The section *Ignored Sequences/Pseudogenes* was renamed into *Skipped Sequences*.
- Section *Amplicons/PCR Primers:* new column *Multiplex No*: here optionally a PCR Multiplex number can be entered for each amplicon.
- In section *Sequence* the chromosomal location, the length of the sequence and the strand is displayed in front of the ROI sequence.

  In case an ROI was defined wrong so that the location is not the defined exon anymore , you will get the warning: “WARNING incorrect ROI”. E.g. BRCA1-E2 [-20..-2000] is defind but exon 2 length is only 99 bp, so the defined ROI is not BRCA1-E2.

- Button *[Import Primer]*: The following is checked when a list of primer pairs is imported:
  1. 1. case: no amplicons are present: all primer pairs, that can be mapped to the extended ROI are added as amplicon inlcuding primer sequences.
  2. 2. case: amplicons are present already for the ROI, but without primer sequences: The primer sequences are only saved in case they enclose an existing amplicon. Otherwise the primers are not imported (you will get a list of invalid primers and possible ROIs). No new (additional) amplicons are added.

6.6.2 Filters for background reads

There are two new options available to filter background reads (e.g. pseudogene sequences) in operation *ROI [master file].*

- Button *[Find Homologies]*: An automatic search for homologous regions in the genome can be done for ROIs to exclude “background reads”: Therefore select one or several ROIs in the *ROI List* and press *[Find Homologies]*. The genome is searched for homologous regions for the selected ROIs. In case they are present, the chromosomal positions are entered in the new section “*Homologies (only mapping)”* for each ROI. Reads that are mapped to the ROI and match to active homologous sequences are discarded.

  Note: When this button is used for the first time, `bwa`-files are created in the `Genome/hg19`

folder of your **SeqPilot** installation (default: `C:/SeqPilot/GeneFiles/Genome/hg19`). This first time creation takes around 1-2 hours. To skip this process, the bwa-files can also be downloaded from our ftp-server:

1. Therefore go to: ftp://ftpsrv.jsi-medisys.de
2. Use the following login:

   Username: *SeqPilotData*

   Password: *SeqPilot*
3. Download the file *Genome/hg19/hg19IndexFiles.exe.*
4. Execute the file and follow the installation instructions; as destination directory `C:\SeqPilot\GeneFiles\Genome` is proposed automatically; in case this isn't the folder of your **SeqPilot** installation, please change the link to the *GeneFiles* folder of your **SeqPilot** installation.

Note: This has only to be done for installations with existing genomes. For newly downloaded genomes from our homepage the `bwa`-files are included already.

- New column *Analysis Mode* in the *ROI List*: ROIs that should only be used for mapping can be defined, to filter "background reads". If “only mapping” is checked, the ROI is not shown in operation *Sequence*.

  The *Analysis Mode* can be changed for all or selected ROIs simulatenously with the context menu in the *ROI List*: Use the entry *Analysis Mode > all/selected> blanc/only mapping*.

Note: A third option is available in *ROI Groups [master file].* There known pseudogene sequences from Pseudogenes.org can be imported.

All read that are filtered because of these three options are listed in the file `UnusedReads.txt` that locates in the `SeqNResults`-folder of your **SEQUENCE Pilot** installation, in the folder for the order: The reads have the flag “4” and the name of the ROI in front (for Homologies: name of the ROI + chromosome range).

Moreover a list of unused reads can easily be opened in operation *Sequence*. Therefore use the context menu item *show > unused reads* in section *Files*.

6.6.3 Tab add PCR and tab add Panel

In the field “*Suffix*” special characters can be entered to automatically define the ROI name:

- The default “*ROI name*” is GENE-LOC-suffix
  - GENE=Gene name
  - LOC=Location (Exon name)
  - additional characters are: TID or ISO=transcript ID
- Example for BRCA1, ENST00000001, E10:
  - In the “*Suffix*” field the following entry is listed: Test-GENE-TID-location-LOC-end
  - The following “*ROI name*” is created: Test-BRCA1-ENST00000001-location-E10-end

6.6.4 Tab add PCR/tab Gene

The following entries are available in the field *Organism*:

- *gene file+genome*: the gene will be searched for in all loaded genomes unless the gene file was loaded (*Gene Admin*). In this case, only the loaded gene will be shown. The column *Source* shows the source of the gene (*genome* or *gene file*).
- *gene file*: the gene from gene files (*Gene Admin*) will be used.

- *genome,* e.g. hg19: the gene will be searched for in the selected genome.

The new field *Chromosome* is available. When a chromosome is selected here, only genes located on this chromosome are available to create ROIs.

6.6.5 Tab add PCR/tab Fasta

In case `fasta`-files are used for ROI set-up, they have to locate in the directory `.../Genefiles/Fasta` of your **SEQUENCE Pilot** installation (by default this is `C:/SeqPilot/Genefiles/Fasta`). This was changed because for Client/Server installations no absolute path for the location of the `fasta`-file was allowed.

6.6.6 Tab Add Panel

The tab *add Enrichment/Kit* was renamed into tab *add Panel*.

New option to save a panel during set up:

During creation of the ROIs, the entries of the table on the left side can be saved. This allows to interrupt the ROI set up (e.g. selection of the correct transcript) before the ROIs are added to the *ROI List*. To save the table on the left side enter a name into the field *Panel.* Then press *[Save]* behind the field. To edit a saved panel press *[Change...]* behind the field *Panel*. After the set up is complete (ROIs are added to the *ROI List* and saved) the saved Panel on the left side can be deleted. Therefore select the corresponding entry in the field *Panel Name* and press *[Delete]* behind the field.

Window *Import Panel File*:

After *[Build...]* is pressed the window *Import panel files* opens automatically. This window also can be opened by pressing *[Columns]* in the section *Settings*. Entries in the window *Import panel file* changed. **Note:** the entries in this window are filled out automatically depending on the used panel file. You can edit columns containing special information such as gene name, transcript ID and information about controls for CNV analysis. Please have a look at the **User Manual SeqNext** for further details.

The following *Settings* changed:

- *Map to gene files (Gene Admin)...:* activate this options in case gene files (from *Gene Admin*) should be used to set up the panel. In case a gene is not present in *Gene Admin*, the gene from the genome selected in the field *Organism* is used.
- *Cut/Expand* is inactive by default (field is empty). The ROI sequences can be cut/expanded to the values set in the fields *5'* and *3'*. The entries in field *Location* change. The following options can be selected:
  - *cut exons:* Exons are cut to the values entered in the field *5'* and *3'* for each ROI: e.g. Setting is 5' "-20" and 3' "30". Exon is cut to "20" bases before and "30" bases after the exon. **Note**: ROIs are only cut and not expanded, therefore this setting can be used for amplicon based sequencing.
  - *cut/expand exons*: Exons are cut or expanded to the values entered in the field *5'* and *3'* for each ROI.
  - *cut coding exons*: refers to the coding sequence (from start to stop codon): e.g. setting is *5'* "-20" and *3'* "30": all ROIs are cut to "20" bases before and "30" bases after the exon. In case the ROI includes a start codon/stop codon it is cut to "20" bases before the start codon and "30" bases after the stop codon. In case non coding exons are present they are set up as ROIs but set to "only mapping" (in column *Analysis Mode*). **Note**: ROIs are only cut and not expanded, therefore this setting can be used for amplicon based sequencing.
  - *cut/expand coding exons*: refers to the coding sequence (from start to stop codon): e.g. setting is 5' "-20" and 3' "30": all ROIs are cut/expanded to "20" bases before and "30"

bases after the exon. In case the ROI includes a start codon/stop codon it is cut/expanded to “20” bases before the start codon and “30” bases after the stop codon. In case non coding exons are present they are set up as ROIs but set to “only mapping” (in column *Analysis Mode*).

- *ROI*: refers to the ROI sequence (genomic position): e.g. setting is *5'* “-20” and *3'* “30”: All ROIs are expanded to “20” bases before and “30” bases after the ROI. In case the setting is *5'* “20” and *3*' “-30”, the first “20” bases and the last “30” bases of the ROI are cut. **Note:** can be used for sheared data.

In case `Manifest`-files are loaded which include pseudogenes sequences, these are set up as ROIs with the *Analysis Mode* “only mapping” automatically.

In case an error is present in the enrichment file, you will get the following message after *[Build]* is pressed: “Not all lines are imported correctely! Please look into the log file!”

You can check the log file using the new button *[Log]*. It opens a dialogue showing the following information:

- *Line*: line in the enrichment file.
- *Log*: In case there is an error the chromosomal position in that line was not translated into an ROI. An according entry is present, e.g chromosomal start/end position is wrong or identical chromosomal positions are present in several lines.
- *Hint*: In case the chromsomal position in the line was translated into an ROI, the Gene, Exon, Location, genome and transcript is listed.

6.7 Set up of Multiplicom MASTR assays

Multiplicom MASTR assays are not set up using `tsv`-files any more. This does not affect ROIs created with `tsv`-files in previous versions!

Multiplicom offers special `bed`-files for SeqNext, that can be used for import. These can be imported in *ROI [master file]*/ tab *add Panel*.
Note: Please make sure to download the special bed-files for the software SeqNext from the Multiplicom homepage. Otherwise the primer information will be missing!

In operation *ROI [master file]*/tab *add Panel*, the following is recommended for Multiplicom `bed`-files:

- Dialouge *Import Panel File*.

  For MASTR assays without copy number variation analysis, please change the following fields:

  - *Sense Primer*: 7
  - *Antisense Primer*: 8

  For MASTR assays including controls for copy number variation analysis, please change the following fields:

  - *Comment*: 4
  - *Sense Primer*: 7
  - *Antisense Primer*: 8
  - *CNV Control:* 4
  - *Key:* Control
- Section *Settings:*
  Make sure to check the box *build amplicons*. In case gene files (from *Gene Admin*) should be used for the setup, activate the option *map to gene files (Gene Admin).* **Note**: The genefile has to be mapped to a genome!

6.8 Operation ROI Groups [master file]

- Button *[Export...]*: The following new option is available when an ROI Group is exported:
  - The Group name/lot number can be replaced. Therefore enter the new names into the fields *Group Name* and *Group Lot* respectively.
  - The panel name can be replaced. Therefore enter the new panel in the field *Panel*.
  - The suffix name can be replaced. Therefore enter the new suffix in the field *Suffix*.
- Button *[Import...]*: the following new options are available when an ROI Group is imported:
  - The Group name/lot number can be replaced. Therefore enter the new names into the fields *Group Name* and *Group Lot* respectively.
  - The panel name can be replaced. Therefore enter the new panel in the field *Panel*.
  - The suffix name can be replaced. Therefore enter the new suffix in the field *Suffix*.
  - In case gene files (from *Gene Admin*) should be used as references, activate the box *map to gene files (Gene Admin)*. In this case the dialogue *Change Gene/Transcript opens....* Here you can adapt the transcript, that should be used as reference. After pressing *[OK]*, you get a message showing how many ROIs were mapped to a gene file (*Gene Admin*). For ROIs that can not be mapped to a gene, the gemome is used as reference.

    Note: The mapping to a gene file only works in case the ROIs were set up with a genome reference. In case a the ROIs were set up with a gene file reference already, no mapping to another gene file can be done. In this case the dialogue *Change Gene/Transcript* is empty!
- The new button *[Copy...]* is available: a copy of the selected *ROI Group* is saved. You can enter a new ROI Group name and lot number.
- The following new search fields are available:
  - *Suffix*: here you can search ROIs with a certain *Suffix*. Therefore press *[Search]* behind *Suffix*.
  - *Source*: Here you can search for ROIs with a certain *Source*. The following sources are possible: *gene file*, *genome, chromosome range* and *fasta file*.
- The following new table columns are available:
  - *Index*: the table entries are numbered.
  - *Source*: Here the source of the sequence is defined. The following sources are possible: *gene file*, *genome, chr. range* and *fasta file*.
  - *Analysis Mode*: in case the entry is “only mapping”, the ROI is not shown in operation *Sequence* (can be set in operation *ROI [master file]*).
  - *Settings Profile:* special Profiles that are used for analysis of the ROI are listed (can be set in operation *ROI [master file]*).
- The context menu entry *show > ROI info...* is available in the *ROI Group* table to open the *ROI info window*.
- Button *[Import Primer]*: The following is checked when a list of primer pairs is imported:
  - 1. case: no amplicons are present: all primer pairs that can be mapped to the extended ROI are added as amplicon inlcuding primer sequences.
  - 2. case: amplicons are present already for the ROI, but without primer sequences: The primer sequences are only saved in case they enclose an existing amplicon. Otherwise the primers are not imported (you will get a list of invalid primers and possible ROIs). No new (additional) amplicons are added.
- When an ROI Group is deleted, you can optionally delete the ROIs belonging to the Group.
- The section *CNV Group* to select control samples for CNV analysis is not available in this operation any more. It was moved to operation *Analysis Mode [master file].*

6.8.1 Import of Pseudogenes from Pseudogenes.org

Pseudogene lists from “Pseudogenes.org” (http://pseudogene.org/) can be imported to filter pseudogene reads. A list of human pseudogenes for import can be downloaded.

Please note: the latest pseudogene files correspond to genome hg38 (NCBI38). In case you use genome hg19 (NCBI37), please use Pseudogene release 74 or older!

To find all Human Pseudogene Releases do the following:

- Go to http://pseudogene.org/Human/
- Press the link behind “**Other Human Pseudogene Sets**”
- A list of Human Pseudogenes opens
- Scroll down to find “Human Pseudogenes (Build 74)”
- Press “download” behind this entry

Note: **For server installations the `Pseudogene.org` file must be placed in the directory defined as “NewResultFiles” directory for module SeqNext**. This folder is specified in the `lis.ini`-file (`bin`-directory of your SeqPilot installation) behind the entry `NewResultFiles=` in section `[SeqNext]`.
For single user installation the downloaded `Pseudogene.org` file can locate in any folder on your PC.

To create an ROI Group containing pseudogenes from Pseudogene.org:

- Enter a Group name in the field *Name* and press *[Save]*
- Press the button *[Import Pseudogene.org file...].*
- A new dialogue opens. Here you can select:
  - The Group Name
  - Optionally a Lot Number
  - An ROI Name Prefix: In case a Prefix is added in front of the ROI name
  - Organism: Select the genome that should be used for mapping of the pseudogenes.
- Select the downloaded file `Pseudogene.org` file.
- Please note, that the import takes several minutes.

Pseudogene ROIs are created automatically. The Pseudogene ROIs are named “ROI Name Prefix_Pseudogene.org_identifier”, the *Panel* is “Pseudogenes.org”. They are set to *Analysis Mode* “only mapping”, therefore pseudogene reads are filtered and not shown as ROIs in operation *Sequence*.

Note: In case the pseudogenes can be mapped to the default genome the chromosomal location is shown, the *SeqType* is “genomic”. Sequences that can not be mapped are stored as `fasta`-sequences. They are also used for filtering.

For pseudogene filtering the ROI Group has to be selected in operation *Run,* in addition to the “real” ROI Group. All reads that can be mapped to known pseudogenes are filtered.
Filtered reads are listed in the file `UnusedReads.txt` that locates in the `SeqNResults`-folder of your SeqPilot installation, in the folder for the order: The reads have the flag “4” and the name of the Pseudogene in front.

Filtered reads can also be viewed with the following context menu in operation *Sequence*: section*Files / show> ususued reads...*.

6.9 Operation Run

6.9.1 Section Settings

Note: In case you still have Profiles that were created with our software version 3.5, please check them carefully. In special cases settings that should not be used might be set.

- New tab *Filters*: here special Filters can be defined. For filtering, special internal filters (JSI) can be used as as well as imported databases (ClinVitae, COSMIC, ClinVar, 1000 Genomes, dbSNP). Mutations that do not pass the filter settings are listed on tab *Filter* in the *Variation/Mutation table*. A filter profile can be saved.

  Please have a look at our **User Manual SeqNext** for detailed instructions.

- New tab *Show Var Dbs*: Information for mutations/variations present in the gene files (*Gene Admin*) and in the SNP databases can be shown in the *Variation/Mutation table* and *sequences*. A profile can be saved.

  One tab is present for the gene file (in case a gene file from *Gene Admin* is used as reference sequence) and for each imported SNP database. The number of variations/mutations described in the databases is listed on each tab as well.

  The *Profile* "default" is present already. The first two entries (*Variation* and *Overview*) are selected in column *Show* on all tabs (*Genefile (Gene Admin)* and SNP databases). Therefore the following is displayed by default:

  1. entry *Variation*: Reference IDs are shown in the *Variation/Mutation table* (column *web Ref.*) for detected mutations. The database is listed in paranthesis behind the ID.

  2. entry *Overview*: Positions with WebRefs *(DB entry)* are shown in the location overview and electropherogram (highlighted grey). Information about the WebRef is shown in a tooltip, when moving over the *gene reference sequence.*

  Moreover information for variations/mutations present in the SNP databases can be shown in additional columns of the *Variation/Mutation table*.

  Please have a look at our **User Manual SeqNext** for detailed instructions.

- The following items on tab *Settings* changed:
  - The Profile/Settings selected in operation *Run* are only applied to the ROIs without profile settings assigned to them (column *Settings Profile* is blanc in operation *ROI [master file]*).
  - The name of the used Profile is available in operation *Sequence* (tab *Orders*) and on the *Report*.
  - The tab *Hide Mutations* is not available any more. The options that were present are available on tab *Filters*/tab *JSI*.
  - The following settings changed:
    - tab *Settings*
      - section *[2] Analyse/ignore region* was renamed into *Required Coverage*
      - section *[3] Coverage Warning* was renamed into *Expected Coverage Warning.* The *Low abs coverage warning* was renamed into *Min abs coverage.*
      - *Randomly sheared*: this setting can be activated in two cases:
        - Reads are amplicon based (amplicons defined) and sheared afterwards. This setting is **obligatory** in case amplicons are defined. Primers are removed.
        - Randomly sheared DNA (no amplicons defined) in combination with paired end reads: This setting is **optional**. If active the fragment size is analysed, which gives a hint for big insertions/deletions.
      - *Force combined (% coverage):* This setting is useful when

- “per dir” settings and/or “ratio read direction” settings are used in sections *[2] Required coverage* or *[4] Mutations* of tab *Settings.* (In case these settings are not used, the *force combined (% coverage)* is greyed out).
- mutations with a high coverage are present in any sequencing direction, but are not called, because s*ettings* in sections *[2]* or *[4]* are not fullfilled.
- In case the coverage of the mutation reaches the *Force combined* value in at least one sequencing direction, the following settings are changed automatically:

  In section *[2] Required Coverage* and in section *[4] Mutations* all “per dir” settings are switched to “combined” and all “ratio read direction” settings are switched to “off”.
- The mutation is called in case all settings are fullfilled now (mutation calling is done using the new settings). It is marked pink in the *Variation/Mutation table* to indicate, that the combined mode was used automatically.

In addition the following options can be activated:

- *only [4]*: For the new calculations the combined mode settings are only used in section *[4] Mutations.* The settings in section *[2] Required coverage* are not changed.
- *no WT:* the *Force combined* setting is only used in case no wildtype is called in both sequencing directions. The wildtype is missing if it does not reach the *Min % coverage*. With this setting the option *only [4]* is set automatically.

- tab *Trimming* > *Remove Ends:* is not available any more.

  Note: In case the options *Adaptor* and *remove bases* are used, adaptors are removed first.
- tab *BAM/SAM*: the following options are available to decide which information is used from the `bam`-file and what is calculated by SeqPilot:
  - section *none: perform mapping, alignment, quality filtering and variation calling with SeqPilot*: mapping, alignment, qualtiy filtering and variation calling is done with **SeqPilot**. The corresponding information from the `bam`-file is not used.
  - section *mixed*:
    - *utilize mapping; perform alignment, quality filtering and variation calling with SeqPilot*: alignment (pairwise alignment to the ROI to find the exact position of deletions, insertions and SNPs), quality filtering and variation calling is done by **SeqPilot**. The mapping information is used from the `bam`-file.
    - *utilize mapping and alignment; perform quality filtering and variation calling with SeqPilot*: quality filtering and variation calling is done by **SeqPilot**. The mapping and alignment information is used from the `bam`-file.
  - section *full*: *utilize mapping and alignment; skip quality filters; perform variation calling with SeqPilot*: Only the variation calling is done by **SeqPilot**. **SeqPilot** does not apply the quality filters. The mapping and alignment information is used from the `bam`-file.
- tab *Expert Settings*
  - *Min % Homozygous*: the percentage coverage a mutation has to reach to be called homozygous instead of heterozygous can be set. In the previous version the minimum coverage for homozygous calling was “100 %– Ignore coverage” automatically.
  - *Allow unique paired end reads*: By default this setting is not active. In case a read pair does not align to the same ROI/amplicon both reads are discarded. If this setting is active the unique reads are aligned to different ROIs.

6.9.2 Patient table

- For one patient per gene analyses it is now possible to enter several patients without barcode in the section *Patient*. In this case different ROIs have to be defined for each patient. Instead of joining the reads to a patient by MID, the reads are then joined to a patient by ROI.
- Import of a patient list (Button *[Import]*): Instead of a barcode an MID-number can be imported.

  Note: In this case the MID-list (`MIDs.txt`) has to be present in the folder `Settings/SeqNext` of your installation. In the `MID.txt`-file the MIDs have to be entered in the following way:

  ```
  MID-1=ACGAGTGCGT
  MID-2=ACGCTCGACA
  MID-3=AGACGCACTC
  MID-4=AGCACTGTAG
  …
  ```

6.9.3 Autorun

Instead of a barcode an MID can be entered into the corresponding field of the Autorun `txt`-file.

Note: In this case the MID-list (`MIDs.txt`) has to be present in the folder `Settings/SeqNext` of your installation. In the `MID.txt`-file the MIDs have to be entered in the following way:

```
MID-1=ACGAGTGCGT
MID-2=ACGCTCGACA
MID-3=AGACGCACTC
MID-4=AGCACTGTAG
…
```

For the Autorun `txt`-file three new fields (9, 10 and 11) are available. The fields have to be entered in the following order:

1. DNA No
2. Barcode (optional) or MID (MID is only possible if the MID is present in the MID list)
3. ROI (Enter the name of the ROI(s) and/or ROI Group(s). an ROI Group with a Lot has to be entered as follows: "`ROI Group Name / Lot`". There has to be a blank before and after the slash.
4. Path of the file(s).
5. Organ/Phenotype (optional): has to be entered in the following way: *Organ / Phenotype*
6. Type (optional, in case no type is entered the type is patient automatically)
7. Settings (optional, in case no settings are available the default settings saved in the `lis.ini`-file are used).

   If not the default settings are used the settings have to be saved in profiles. If you want to use a saved profile, give the profile's name as settings entry, e.g., "MyProfile". There must not be a "FieldSeparator" (default: ";") in the profile's name or before or after it.
8. Project (optional)
9. Organism (optional)
10. Filter: (optional) enter a Filter profile. If nothing is entered the default profile is used.
11. ShowVarDBs: (optional) enter a ShowVarDBs profile. If nothing is entered the default profile is used.

6.10 Operation Joining/Poollist

In the *Lower table/Runs table* the patient type (patient/control) can be edited: therefore select one or several files and use the context menu *edit Patient Type...*: A list opens where the type “patient/control” can be changed. In case a control, that is already joined to patients (*CNV window,* operation *Sequence*) is set as patient, the control gets a pink background in the *Controls list* (*CNV window*).

6.11 Operation Joining

New column *Warning* in the *Upper Table.* Here a warning can appear in case the transcripts, that were used to create the ROIs can not be found in the downloaded genome. This can only be possible in case:

- ROIs/ROI Groups from another **SEQUENCE Pilot** installation were imported (*[Import]* button in *ROI/ROI Groups [master file]*)
- the genome was replaced by a new genome containing other transcripts (e.g. *hg19* with ENSEMBL references was replaced by *hg19* with Genbank references).

In this case, please correct your ROIs in operation *ROI [master file]*.

The following context menus changed in the *Upper table*:

- *show>info...*: In the Info window the following new information is shown:
  - In section *[Settings]* the name of the Profile is listed.
  - In section *[Filter]* the name of the filter and filter settings are listed.
  - In section *[Trimming information]* the number of processed reads, trimmed reads and discarded reads is listed.

- *settings...*:
  - In the field *MIDs/Barcode* the MID-list is available. This requires that the MID list is stored in the **SEQUENCE Pilot** installation (see User Manual).
  - The new tabs *Filters* and *Show Var Dbs* are available to edit these settings for the Run.

New context menu items in the *Lower Table* for one ore several selected orders:

- *expand/collapse order*: Here all orders or the selected order can be expanded or collapsed to show or hide the joined files respectively.
- *export* > (also works for several selected runs)

  For all exports the following is done*:* A directory to save the data can be selected. Alternatively you can create a default directory, that always should be used. To use this option enter the path of the export folder in the `lis.ini` file, section `[SeqPilot]` behind the entry "`ExportDir=`". When the export is finished the number of generated files and the path of the export folder is shown. The file-name is created using the DNA number.
  - *variations (.vcf file)*: The variations of all selected run are exported as `vcf`-file (one `vcf`-file per run).
  - *reads (.bam file)*: exports the reads of all selected runs as `bam`-file (one `bam`-file per run). This file contains already trimmed but not aligned reads. Moreover the used Profile, software version and the Input file link is exported as well.

    The `bam`-file can also be exported in operation *Sequence*/context menu in section *Files*.
  - *reads (.fastq)*: exports the reads of all selected runs as `fastq`-file (one `fastq`-file per run).

During export a status bar shows, which run is exported at the moment. In case the mapping to the genome is missing for a run or another error occurs an according message is shown (for *vcf* and *bam export*).

6.12 Operation Sequence

6.12.1 Order/Protocol/Family/Pool

- On tab *Order* the name of the Profile used for the Run is shown.
- The new tab *Pool* is available in operation *Sequence*. In case the file is included in a patient pool the pool name and all other files included are listed. In case the file is included in several patient pools several tabs are available on tab *Pool*.

6.12.2 Files, ROI Groups, Genes/Chromosomes and ROIs/Locations

- Column *Bases* is not available anymore. The new columns *Req. Cov* and *Exp. Cov* are present now. *Req.Cov* refers to all bases that fullfill the settings *Required Coverage (Settings* section in operation *Run*). *Exp. Cov* lists bases whose coverage fullfill the setting *Min abs coverage*/*Expected Coverage Warning*. The percentage of bases with a sufficient coverage compared to the total number of bases is listed. The number of bases with a sufficient coverage and the total number of bases are shown in parenthesis.
- Column *Coverage* (was available in section *ROIs/Location* only) is now also available in section *Files, ROI Groups* and *Genes/Chromosomes*.

  It lists the following coverage warnings:

  Note: Only one coverage warning is listed, the priority of the warning is as listed below (*not analysed* has the highest priority whereas *expected* has the lowest priority).

  - *not analysed*: An error occurred. Please recalculate your file.
  - *dropout*: No sequences are available for more than 90% of the ROI. (No read are also present in case the Required Coverage Settings (*Min abs coverage and/or Ratio read directions*; defined in operation *Run, section [2]*) are not fullfilled.
  - *nocall*: There is no coverage at one or more positions of the ROI (Only occurs in case no *Required Coverage Settings* are set).
  - *required:* This warning shows, that positions in the file are ignored (not analysed). Positions are ignored in case the Settings for the *Required Coverage* (*Min abs coverage and/or Ratio read directions*; defined in operation *Run, section [2]*) are not fulfilled.

    The ignored positions are greyed out in the sequences (forward, reverse and combined sequence).
  - *quality*: This is shown in case the setting *Low Quality score coverage warning* (operation *Run*) is exceeded.
  - *expected* (this warning was named *low* in the previous version): The coverage of one or several positions is below the *Expected coverage warning/Min abs coverage* (default: 100). The *Min abs coverage* is represented as a red dotted line in the electropherogram. Moreover the *Min abs coveage* is shown in the graphical overview below the location overview. The *Expected Coverage warning/Min abs coverage* can be set in your *Settings* when starting the Run.
- New column *Mutation*: Here warnings can be present in case possible mutations are listed on tab *Warning* of the *Variation/Mutation table*. The following warnings can be present:
  - *cutoff*: Reads are present where a part fits perfectly to the reference and starting from a certain position the bases can not be aligned anymore (e.g as expected for transversion, translocations). A warning is only shown in case the frequency of the sequences, that can not be aligned reaches the *Warning* value defined in the *Settings* (operation *Run, Expert Setting: Warning*, default: 50%). *Hint* in the *Variation/Mutation table* is *cutoff left/right*.

- *Indel*: There are possible deletions/insertions in the ROI that were not called. The warning is present in case there are different overlapping insertion/deletions present for at least one position of the ROI. Together the coverage of these insertions/deletions at a position must fullfill the *Settings Profile* selected in the operation *Run. Hint* in the *Variation/Mutation table* is *possilbe del./possible ins.*
- *fragment size*: **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***. Deviation from standard library size might hint at a large insertion or deletion. Too short fragments give a hint for an insertion, too long fragments a hint for a deletion. In the*Variation/Mutation table* the *Hint fragment size: possible deletion /fragment size: possible insertion* is shown. The fragment size can be checked using the context menu show fragment size in the *coverage graph (Location overview).*
- *orientation:* **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***. This warning shown that anomalous pair orientations are present. This could hint at structural events (like inversions, translocations etc). *Hint* in the *Variation/Mutation table* is *same orientation pair* or *swapped pair orientation*.
  - *same dir. pair:* One read of the read pair has changed the direction. This might be a hint for a possible inversion.
  - *swapped pair orientation*: The reverse read lies in front of the forward reads.

6.12.3 Files,Genes/Chromosomes and ROIs/Locations

The context menu item *editing* > *original* is not available any more.

6.12.4 Files

New context menu items *show* >

- *sequence*: Here a sequence can be searched for in the file. Therefore:
  - Enter the sequence in the field *Seq.*
  - Check *both directions* in case the sequence should be checked in forward and reverse reads.
  - Optionally enter a gene name in the field *Gene*. If this is done the sequence is only searched for in that gene.
  - Optionally enter a ROI name in the field *ROI*. If this is done the sequnece is only searched for in that ROI.
  - Check *only mapping* to get the result for mapped reads only (not for aligned).
  - Press *[Search]*.
  - In the result table it is now listed in which genes/ROIs the sequence was found. Moreover the number of mapped, aligned (fwd/rev) is shown in the corresponding columns.
- *unused reads...*: A new window opens that lists the reads that were unused (not mapped or aligned to an ROI). The same information is also present in the `UnusedReads.txt-file`.

  In the section *Overall*, all errors that occurred are explained, e.g. “Error 51: reads have a quality score not satisfying the user-specified threshold”. The number of reads filtered for each error (mapping/alignment) is listed.

  You can find a list of all possible errors by pressig the button *[?]* in this window, or in the Menu *Help*/item *UnusedReadsErrorCodes.*

  In the middle a search for unused reads of certain *Genes, ROIs*, *Error codes* or *Sequences* can be done. If the button *[Search]* is pressed:
  - the section *Searched* lists the information about unused reads depending on the search

criteria.

- the table below displays all reads that fullfill the search criteria. The following information is given:
  - *ROI*: name of the ROI or Homology region the read could be mapped to. If the read could not be mapped to an ROI this field stays empty.
  - *Error*: Error code number.
  - *Base seq*.: Base sequence of the read. One or several selected base sequences can be copied. Therefore do a right click on the sequence and select *copy selected base sequence(s)* from the context menu.

All ROIs included in the file can be exported as `BAM`- or `FASTQ`-files. Therefore use the context menu item *export>*

For both exports the following is done*:* A directory to save the data can be selected. Alternatively you can create a default directory, that always should be used. To use this option enter the path of the export folder in the `lis.ini` file, section `[SeqPilot]` behind the entry "`ExportDir=`". When the export is finished the number of generated files and the path of the export folder is shown. The file-name is created out of the DNA number.

Note: The export also works for ROIs/Locations (context menu in the section *ROIs/Location*) and in operation *Joining* (context menu in the *Lower Table*) for one or several selected Runs.

- *reads (BAM)*: creates one `BAM`-file and one `BAM.BAI`-file for the complete file.

  The following additional information is exported as a comment in the `bam`-file:

  - SeqPilot version: *@CO BAM file generated by SeqPilot...*
  - path and name of the imported raw data file(s)*: @CO InputFiles=*
  - settings: *@CO ProfileId=*

  Each reads gets a tag (behind the quality with the profile name of the ROI, the read was assigned to, e.g. "PF:Z:CNV":

  - "PF" stands for "Profile"
  - "Z": date
  - "CNV" is the profile name
  - “PM”: additional tag profile modification can be present in case the profile for one ROI was changed.

  Note: Reads that are aligned to several ROIs can be present in the `bam`-file several times. In case the mapping to the genome is missing or another error occurs an according message is shown.
- *reads (FASTQ): reads (FASTQ):* Creates one `FASTQ`-file for the comple file.

The *Info* window (context menu *show>info*) was adapted:

- In section *[Settings]* the name of the Profile is listed.
- In section *[Filter]* the name of the filter and filter settings are listed.
- In section *[Trimming information]* the number of processed reads, trimmed reads and discarded reads is listed.

6.12.5 ROIs/Location

- The Selected ROI can be exported as `BAM`- or `FASTQ`-files. Therefore use the context menu item *export>*

For both exports the following is done*:* A directory to save the data can be selected. Alternatively you can create a default directory that always should be used. To use this option enter the path of the export folder in the `lis.ini` file, section `[SeqPilot]` behind the entry "`ExportDir=`". When the export is finished the number of generated files and the path of the export folder is shown. The file-name includes DNA number.

- *reads (BAM)*: creates a `BAM`-file and a `BAM.BAI`-file for the selected ROI. The same information is present as for the BAM export in section *Files*.
- *export > reads (FASTQ):* creates a `FASTQ`-file for the selected ROI.

- The context menu entry *show > info* changed.
  - In the section *[General]* the following additional information is shown:
  - *CoverageStats:* shows the number of positions with a coverage warning: *expected ^ quality ^ no call ^ required.*
  - *AvgCoverage*: average coverage fwd/rev
  - *AvgQuality score*: average quality score fwd/rev
  - The section *[Adaptors]* is not available anymore. Adaptor trimming is done before reads are mapped to ROIs. Therefore this information is now available in the *Info Window* (operation *Sequence/Files/*context menu *show Info*).
  - In section *[Amplicons]* all amplicons are listed with the following information: location I direction I average coverage fwd/rev I average quality score fwd/rev I coverage warning I number of aligned reads I number of positions with a coverage warning: *expected ^ quality ^ no call ^ required.*
  - In section *[Unused Reads]* the number of unused reads is listed. These are all reads that were mapped to the ROI but not aligned. The error code number is written in front. You can open a list of all possible error codes in the Menu *Help* item *UnusedReadsErrorCodes*.
- New context menu s*how > unused reads:* A new window opens, that lists the reads that were not aligned. The same information is also present in the `UnusedReads.txt-file`.

  Note: Reads that were not mapped to the ROIs are not listed here. They can be displayed using the context menu item *show > unused reads* in section *Files.*

  Note: Reads that were mapped to an homology region of the ROI are not shown by default. But if you do a search for Error code: Error4, the corresponding reads are diplayed.

  In the section *Overall*, all errors that occurred are explained, e.g. “Error 51: reads have a quality score not satisfying the user-specified threshold”. The number of reads filtered for each error is listed.

  You can find a list of all possible errors by pressig the button *[?]*.

  In the middle a search for unused reads of certain *Genes, ROIs*, *Error codes* or *Sequences* can be done. If the button *[Search]* is pressed:
  - the section *Searched* lists the information about unused reads depending on the search criteria.
  - the table below displays all reads that fullfill the search criteria. The following information is given:
    - *ROI*: name of the ROI or Homology region the read could be mapped to. If the read could not be mapped to an ROI this field stays empty.
    - *Error*: Error code number.
    - *Base seq*.: Base sequence of the read. One or several selected base sequences can be copied. Therefore do a right click on the sequence and select *copy selected base sequence(s)* from the context menu.

6.12.6 Summary

All tabs:

- Column *Bases* is not available anymore
- new column *Cov. Info*: Here coverage warnings (*not analysed, dropout, nocall*, *required, quality, expected*) for the Gene are listed. **Note**: Only one coverage warning is listed, the priority of the warning is as listed below (n*ot analysed* has the highest priority whereas *expected* has the lowest priority).
- new column *Mut Info*: Here warnings can be present in case possible mutations are listed on tab *Warning* of the *Variation/Mutation table* (*cutoff, Indel, Fragment size, orientation*).

Tab *Gene/ROI*

- new column *CNV* that lists distinct mutations detected by CNV analysis.

Tab *ROI/Amplicon:*

- The *average coverages fwd/rev* (on tab *ROI* and tab *Amplicon*) are listed in two separate columns now.
- Average quality scores for fwd/rev are listed in the column *QS fwd (avg)* and *QS rev (avg)* on the tabs *ROI* and *Amplicon*.
- Information about coverage warnings is displayed: percentage of bases with coverage warning (number of bases with coverage warning / total number of bases). The following columns are present:
  - *No call*: Positions with coverage warning *no call.*
  - *Required*: Positions with coverage warning *required*.
  - *Quality*: Positions with coverage warning *quality.*
  - *Expected:* Positions with coverage warning expected.
- New column *Called bp*: shows the number of called bases: percentage of called bases (number of called bases/number of ROI/amplicon bases)
- New column *Mean RD*: shows the mean read depth
- The columns *Multiplex No* and *Comment* can be shown. Therefore open the context menu *Mangage table columns* (click on the header) and increase the *Width* for the corresponding columns.

Tab *Amplicon*:

- New column *Dupl. Reads:* shows entries in case SmMIP Processing is used only (see chapter 6.3)
- New columns *Multiplex No* and *Comment*: show the Multiplex number and Comment respectively, that was entered in operation *ROI [master file]* for the amplicon. These columns are not shown by default. To show them right click the table header and select *Manage table columns*. Increase the column *Width* e.g. 20.

6.12.7 Variation/Mutation table

6.12.7.1 Mutation Calling of Insertions, Deletions and Indels overlapping an ROI

Rules for calling of insertions, indels and deletions that reach over the beginning/end of an ROI/amplicon:

An insertion is called:

- in case both flanking bases lie within the ROI/amplicon

- **exception**: in case only one flanking base lies withing the ROI/amplicon mutations are only called when
  - the ROI is extended
  - primers are defined for amplicons (Note: the mutation must start after the primer and reach into the amplicon)

A deletion/indel is called:

- in case it completely lies in the ROI/amplicon
- **exception**: the mutation begins/ends in the ROI/amplicon but reaches over the border. Mutations are only called when
  - the ROI is extended
  - primers are defined for amplicons (Note: The mutation must start after the primer and reach into the amplicon)

Note: For restricted ROIs (column "*Restrict*" in *ROI [master file]* was activated manually) and amplicons without primers the following is called:

Deletions/indels, that reach over the ROI/amplicon beginning/end are not called completely. Only the part that completely lies in the ROI/amplicon is called. Example: A 5bp deletion is present, only 2 bp lie withing the ROI → a 2bp deletion is called.

In the *Variation/Mutation table* there is a "*" in front of the deletion. No amino acid change is calculated (no entry in column *AA change*).

6.12.7.2 Tabs

The following new tabs are available:

- *filter*:: lists mutations that were filtered permanently using the SNP database filter (operation *Run,* operation *Joining,* operation *Sequence/ROI table*).
- *temp. filter:* lists mutations that were filtered temporarily using the SNP database filter or the *settings* filter in the *Variation/Mutation table* (context menu *filter>DB* or *settings...* respectively). The tab is cleared when the order is left.
- *warning*: This tab lists entries of the Type *W*. These are warnings, that give a hint for possible mutations. By default the entries listed on this tab are not listed on tab *all*. You can change this using the context menu in the *Variation/Mutation table settings > add warnings to tab all*. To remove warnings from tab *all* use *setting > remove warnings from tab all*.

  In case entries are present, there is also an corresponding entry in the field *Mutation* in sections *Files, Genes, ROI Groups* and *ROIs/Locations*. In case there are mutations called in an area with a Warning, these mutations are marked with a *W* in the *Location overview*.

  The warnings are further defined in column *Hint*, the following entries are possible:

  - *cutoff left/right:* This warning is shown in case reads are present where a part fits perfectly to the reference and starting from a certain position the bases can not be aligned anymore (e.g as expected for transversion, translocations). The position where the not aligned part of the read starts is listed in the table. In the sequences the not aligned part is cut off, there is a red arrowhead present, showing in the direction of the cut off. To see the complete reads click on the arrowhead and select *show mutation*.

    A warning is only shown in case the frequency of the sequences, that can not be aligned reaches the *Warning* value defined in the *Settings* (operation *Run, Expert Setting: Warning*, default: 50%).
  - *possible del/possible ins*: This warning is shown, in case there are possible deletions/insertions in the ROI that were not called. The warning is present in case there

are different overlapping insertion/deletions present for at least one position of the ROI. Together the coverage of these insertions/deletions at a position must fullfill the *Settings Profile* selected in the operation *Run*.

- *fragment size: possible deletion/insertion*: **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***. For these data the fragment size is checked. This gives a hint for deletions/insertions. Too short fragments give a hint for an insertion, too long fragments a hint for a deletion. The fragment size can be checked using the context menu *show fragment size* in the coverage graph (*Location overview*).
- *same dir. pair:* **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***.This warning is shown in case one read of the read pair has changed the direction. This might be a hint for a possible inversion. When you select the mutation in the *Variation table* it is jumped to the position where the read pair starts in the sequences. In column *Position* the sequence range showing mismatching positions is listed.
- *swapped pair orientation*: **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***.This warning is shown in case the reverse read lies in front of the forward reads. This is a hint for possible translocations.

6.12.7.3 Variation DB display and filter function

Imported SNP DBs can be used for display and filtering. Therefore the context menus *show table>DBs...* and *filter>DB...*are available in the *Variation/Mutation table*.

Show table>DBs:
Information for mutations present in the gene files and mutation databases can be shown in the *Variation/Mutation table* and *sequences*.
One tab is present for the gene file (in case a gene file from *Gene Admin* is used as reference sequence) and for each imported SNP database. The number of Variations/Mutations described in the databases is listed on each tab as well.

The *Profile* “default” is present already. The first two entries (*Variation* and *Overview*) are selected in column *Show* on all tabs (*Gene file* and databases). Therefore the following is displayed by default:

1. entry *Variation*: Reference IDs are shown in the *Variation/Mutation table* (column *web Ref.*) for detected mutations. The database is listed in paranthesis behind the ID.

2. entry *Overview*: Positions with WebRefs *(DB entry)* are shown in the location overview and electropherogram (highlighted grey). Information about the WebRef is shown in a tooltip, when moving over the *gene reference sequence.*

Moreover information for variations/mutations present in the SNP databases can be shown in additional columns of the *Variation/Mutation table*. Therefore please have a look at the User Manual **SeqNext**.

Filter>DB:
Here mutations can be filtered. For filtering, special internal filters (JSI) can be used as as well as imported databases (ClinVitae, COSMIC, ClinVar, 1000 Genomes, dbSNP).
When a filter is applied, all filtered mutations are listed on tab *temp. Filter* (not permanent, tab is cleared when the order is left).

To get a permanent filter, a filter profile can be saved and applied in:

- operation *Sequence* for selected ROIs (*ROIs/Location,* context menu *editing* > *filters...*)
- operation *Joining* for the complete order (context menu *editing* > *settings,* tab *Filter*)
- operation *Run* for the complete order *(tab Filter*)

Please have a look at our **User Manual SeqNext** for detailed instructions.

6.12.7.4 Context menu

- The context menu item *filter* was renamed into *filter>settings...*.
- New context menu item *settings>zygosity>set to homozygous/heterozygous*: Using this a mutation can be set homozygous/heterozygous.
- VCF-export (context menu *export* > *tab to VCF...*): additional information is exported as comment ##:
  - file date: date of the export
  - source: SeqPilot version
  - ImputFileList: The path and name of the imported raw data file(s)
  - INFO<ID=profile name: profile name, in case the settings were modified, several entries can be present

  Note: The VCF-export also works in operation *Joining* (context menu in the *Lower Table*) for one or several selected runs.
- Tab *Warning*: the context menu entry *setting* > *add tab filter to tab all* is available. By default the entries listed on tab *Warning* are not listed on tab *all*. You can change this using this item. To remove warnings from tab *all* use *setting* > *remove warnings from tab all*.

6.12.7.5 Table headers

The following new table headers are available:

- *Weighting:* Tab the mutation is sorted to (*distinct, other, homopolymer, warning*). Moreover there is the entry *forced* in case the forced combined mode was used. The forced combined mode can be used automatically in case the settings *Single/double direction analysis* or *forced combined (% coverage)* are set (operation *Run*). The according mutations are also marked pink in the *Variation/Mutation tabl*e.
- *Filter Step:* In case mutations were filtered (filter DB or settings) the filter step where the mutation was filtered is listed.

6.12.8 Electropherogram/Sequences

- In the tooltip of forward, reverse and combined sequence the average quality score value for each base position is shown in brackets behind the base.
- ROIs can be shown in an extended view:
  Therefore switch the combo box *combined* (in the *Electropherogram/Sequences* part) to *combined ext*. In case you use the *fwd+rev* setting, select *fwd+rev ext*.
  The ROI is extended at the 5' and 3' end, the extension is as long as the longest reads that were aligned. The copy number of the longest reads have to exceed the setting *Min abs coverage*. In the extended area the following sequences are shown:
  - In case amplicons are defined, in the extended area no sequences and coverages are shown, except for the gene reference sequence.
  - In case no amplicons are defined, the pseudo-electropherogram, coverage graph, forward and reverse sequence are additionally shown (sequences are only shown in case they can be aligned to the reference). Variations/Mutations are shown but not called! Reads sequences are not displayed.

  Note:
  - The reads sequences in the extended area can be shown in the *Reads view* (check the box *org. Reads*). Here only reads bases that can be aligned to the reference are shown.

Parts that can not be aligned are shown as grey bars. The complete read sequence can be shown with the context menu item "*show origninal reads...*" for a selected read.

- The extended view is only available in case the ROIs are not restricted: To switch off the extended view function completely for one or several ROIs, activate the column *Restrict* in the operation *ROI [master file]*. In case you activate the column *Restrict* deletions/indels, that partly lie within the ROI/amplicon, but reach over the ROI/amplicon beginning/end are not called completely. Only the part that lies within the ROI/amplicon is called.

- Identical reads: In the old version identical reads were listed only once. There copy number is shown at the left end of each read. This changed: Now reads are only identical in case the complete alignment is identical. To see the complete alignment open the *Reads view* and check the box *org. Reads.* Here all bases must be identical, also bases that are present in the grey areas (outside the ROI/amplicons).

- Location overview:
  - The following additional information is shown in the tooltip in the *Location overview*: Genome, Chromosome, TranscriptID, average coverage and average quality for each amplicons and Warnings.
  - There is a context menu available*:*
    - *show read coverage*: shows the read coverage (not the coverage for each position). For deletions the drop in coverage is not shown anymore in case the deletion is covered by reads.
    - *show fragment size*: **Only present for sheared paired end-data without amplicons (e.g Haloplex data). The setting *randomly sheared* has to be set in the *Settings*/operation *Run***. For these data the fragment size is checked. This gives a hint for deletions/insertions. Too short fragments give a hint for an insertion, too long fragments a hint for a deletion.

      When the item is used the fragment size is shown graphically below the location overview. Moreover the following values are listed: *Avg* (Average fragment size for the location + Standart deviation), *Max* (Maximum fragment size in the file), *Min* (Minimum fragment size in the file).

      In case a peak is present that exceeds the doubled standard deviation, there is also a *Warning* in the *Variation/Mutation table*, tab *Warning*. The warning is defined in column *Hint*, there is the entry *fragment size: possible deletion* (for a deletion) and *fragment size: possible insertion* (for an insertion).

- *Reads view*/check box *original reads*:
  There are differences in the reads view depending on, if the normal view (combo box *combined* or *fwd+rev*) or extended view (combo box *combined ext.* or *fwd+rev ext.*) is selected in the electropherogram.
  - normal view: Before/after the ROI a maximum number of seven bases is highlighted. Bases are not shown.
  - extended view: The reads sequences in the extended area are shown automatically, in case they can be aligned to the gene reference sequence. After the extended area, a maximum number of seven bases is shown.

- Basecalling:
  The basecalling for positions with different bases called in forward and reverse direction changed:
  Example: reference base is C, forward base is C and reverse base is T. If the setting "per dir" is used no mutation can be called because the T was not found in forward direction. No wildtype can be called, because the C was not found in reverse direction. For those positions no combined sequence is called. In the Location overview positions where no combined sequence can be called are greyed out.

6.12.9 Show

New/renmamed *check* jumpers are available:

- *required*: moves to positions that are ignored because they do not fullfill the *Required Coverage settings* (operation *Run*/tab *Settings*).
- *coverage*: here you can jump through positions in the selected *ROI/Location* sorted by coverage. The position with the lowest coverage is jumped to first.
- *web+mut* was renamed in *web/mut Ref*.
- *low* was renamed in *expected*.
- *low quality* was renamed in *quality*.

6.12.10 Report

- In the report the different tabs of the *Variation/Mutation table* can be printed separated now. In the *Print/Preview* the new table *Variation* is present, to select which tab should be printed. You can select several tabs, the tabs will be printed as separated tables. With *[Save as default]* your settings are saved.
- In the dialogue *Print/Preview* all loaded files for a run are shown in the column *Files*.
- The coverage warnings (*expected, no call, drop out, required, quality*) for an ROI can be shown on the Report. Therefore please contact our support team.
- The *mean read depth* (Mean RD) can be printed on the Report. Therefore please contact our support team.

6.13 Operation Archiving

Archiving makes the result data smaller. In the `SeqNResults` folder only the file "`ROI-name.txt`" (e.g. `BRCA1-E02-txt`) remains. All other files, such as `Frags.txt` and `UnusedReads.txt` are deleted. Therefore, after archiving all data can be viewed but no recalculate/edit can be done any more.

The following data can not be viewed anymore:

- UnusedReads (section Files or ROIs/Location > context menu show unused reads).
- Summary: on tab *ROI* and *Amplicon* not all information is present anymore. These are columns *Called bp , Required, Quality* and *Expected, Assigned Reads* and *Aligned Reads* on tab *ROI* and all columns on tab *Amplicon*.

6.14 CNV analysis

6.14.1 Operation ROI [master file]

- Multiplicom MASTR assays including controls for CNV analysis are now imported on tab *Add Panel*. The following fields have to be changed in the window *Import Panel file*:
  - *CNV Control:* 4
  - *Key:* Control
  - *Plex No:* (this is not present in the file yet)

  Moreover in the section *Settings* activate the options *build amplicons*.
- The new column *CNV probe type* lists the CNV probe type. For control ROIs this entry can be set to *Control* (blanc means Target ROI).

  In case a panel file is imported (*Add Panel file*) and the control ROIs have an common

identifier, this information can be entered in the *Import Panel file* dialogue in the section *CNV controls*: enter the column number into the field *CNV control* and the identifier in the field *Key*. If the identifier is present for an ROI the CNV probe type is set to *control* automatically.

- The Plex number for a Multiplex PCR can be saved now. Therefore the Plex numbers can be added in the section *Amplicons/PCR Primers*/column *Multiplex No*.

  In case a panel file is imported and the plex number is present in the file, the column number can be entered in the field *Plex No*. of the *Import Panel file* dialogue for automatic import.

6.14.2 Operation Analysis mode CNV [master file]

- The new field *Multiplex No.* is available in the *Group ROIs* list. Entries into this field can be done in operation *ROI [master file]*. It can be used for sorting the ROIs and easy definition of the analysis modes (in case several Multiplex PCRs were done).
- The controls that are joined to patients automatically can not be set in operation *ROI Groups [master files]* any more. This option is now availabe in operation *Analysis mode CNV [master file*]. The section *CNV Group* has the following new options:
  - activate the check box *TV controls only* to join only controls that are technically validated
  - new control setting *Autorun/RunID*: This option is useful when many patients are loaded in one Autorun file and no validated controls are present. All samples are regarded as control, each sample is compared to all other samples.
- New buttons *[Import]* and *[Export]*: Analysis modes CNV can be imported and exported respectively (`sae`-file).
- In case several analysis modes are defined for one ROI Group, they can be grouped. For grouped analysis modes the results (*CNV table* and diagram) can be shown together in the *CNV window* (operation *Sequence*). To group several analysis modes the button *[Grouping]* is available. If this is pressed a new window is opened to group analysis modes.
- In case an ROI Group with a defined *Analysis Mode CNV* is exported (in operation *ROI Groups [master file]*, button *[Export]*), the grouping information is exported as well.

6.14.3 Operation Joining

With the context menu item *edit Patient Type...* in the *Lower table* the type “patient/control” can be changed for one or several files. In case a control, that is already joined to patients (*CNV window*) is set as patient, the control gets a pink background in the *Controls list* (*CNV window*).

6.14.4 Operation Sequence/CNV window

- *Controls section:* already joined controls, that were set to type *patient* afterwards are marked pink. To remove them press *[Control setting...]* and then *[OK]*.
- Button *[Control Settings...]*: New tab *Run...:* Here all control that have the same *Autorun*/*RunID* are listed. The same RundID is present for:
  - all samples that were loaded in one Autorun File
  - all samples that were loaded in one Run
- The result table in the upper right part can be exported. Therefore right-click into the table and select *export* > *table* from the context menu.
- New section *Analysis Modes Grouping*: In case several analysis modes are defined for one ROI Group, they can be grouped. For grouped analysis modes the results (*CNV table* and diagram) can be shown together in the *CNV window* (operation *Sequence*).

  To group several analysis modes in operation *Sequence* the button *[Grouping]* is available. If this is pressed a new window is opened to group analysis modes. The corresponding entries are then made in operation *Analysis modes CNV [master file]* automatically.

- The new table *Analysis Modes Grouping* is available in the *CNV window*: To see the grouped view select the group name in this table. To see the single view for each analysis mode select the entry *Analysis mode CNV*.

6.15 Operation Pool

- *Variation/Mutation table*:
  - New column *Coverage*: Here the coverage % forward/reverse is listed. In case the mutation was detected in several orders the average coverage is listed.
  - Context menu entry *move to...*: In case a mutation is moved from one tab to another (e.g. move to tab *distinct*) in the pool view, you can decide, if the mutation should be moved to the corresponding tab in the patient orders as well. With the option *pool only*, the mutation is only moved in the pool view, with the option *pool+patients* the mutation is moved in the pool view and for each order.
- Technical and medical validation, buttons *[TV]* and *[MV]*: the dialogue where the user can decide if "TV/MV" is set for the pool only or for "pool and all Patient orders changed. You have the options *pool only, pool+patients* and *patients only* to define what should be validated.

7 Modules SeqHLA and SeqNext-HLA

7.1 Importer

Identical file names can be used now. The files are identified by the file date.

7.2 Menu SeqHLA / SeqNext-HLA

Note: `hlu`-files for database updates are not available anymore. Please use the `exe`-file provided on our homepage http://www.jsi-medisys.com/hla-database for database updates.
This database includes exon as well as intron sequences. Intron sequences can only be visualized in software version 4.2.0, for previous versions they have no function.

- *HLA DB Admin*:
  - An HLA database including intron sequences is now available for installation.

    Note: For intronic regions the sequences are only visualized, but not used for result calculation. For result calculation only exons are used. Introns can only be defined together with exons.

    The database is available on our homepage http://www.jsi-medisys.com/hla-database. Download the `exe`-file that includes intron sequences. To install the database, close your SeqPilot installation. Start the `exe`-file and follow the installation instruction. Make sure to enter the directory of your SeqPilot installation during installation (by default `C:\SeqPilot` is used).
  - In the dialogue *HLA DB Admin* the new field *Show* is available. Here you can select, if *Exon* or *Intron* sequences should be displayed, when button *[show alleles...]* is pressed. Intron sequences are only available in case a database including introns is installed.
- *HLA DB Update*: This menu does not exist anymore. `hlu`-files for database updates are not available anymore. Please use the exe-file provided on our homepage http://www.jsi-medisys.com/hla-database for database updates.
- New item *HLA allele comparison...:* compares alleles of a certain gene. The sequences of both alleles are shown as alignment. Heterozygous positions are shown below.

7.3 Operation Archiving

When orders are archived, entries with mismatches > 0 are deleted in the tables “RFmatch” and “RFmatchDetails”.

8 Module SeqHLA

8.1 Operations Amp modules [master file], SeqPrimer [master file], Sequence

Intronic regions can be shown.

Note: For intronic regions the sequences are only visualized, but not used for result calculation. For result calculation only exons are used. Introns can only be defined together with exons.

The only positions that are evaluated in the intron are the splice sites (first two bases and last two bases of the intron). If a mutation is found here, the warning “X” is shown in the column “Warning” of the “Positions/Resultfiles” section. The splice site mutation is not used for result calculation.

8.2 Operation Amp modules [master file]

The following fields changed:

- *Amplified Geneparts*: A new table is available, that lists all exons of the gene selected in the field *Gene*. Moreover the exon length is shown in the column *Exon length* and the columns *Prev. Intron length* and *Next intron length* show the number of intron bases flanking each exon. In the table one or several exons can be selected. The selected exon(s) are listed below and set automatically.
- New field *SeqType*: here you can select, if you analyse genomic or *cDNA*. For *cDNA* no introns are included.

8.3 Operation SeqPrimer [master file]

The following fields changed:

- *SeqPrimer gene parts*: A new table is available, that lists all exons of the gene selected in the field *Gene*. The table lists all exons of the gene selected in the field *Gene*. Moreover the exon length is shown in the column *Exon length* and the columns *Prev. Intron length* and *Next Intron length* show the number of intron bases flanking each exon.

  The field below shows 0..0 by default. This defines the number of intron bases, that should be included in the SeqPrimer on each side of the exon. With the default setting no intron bases are included. Gene parts are defined as follows:

  - *E1 or E1[0..0]:* the complete exon 1 is amplified.
  - *E2[20..0]:* exon 2 is amplified without the first 20 bases
  - *E2[-20..20]*: exon 2 with the last 20 bases of intron 1 and the first 20 bases of intron 2.
  - *E2[0..-20*]: exon 2 without the last 20 bases
  - *E1,E2;E3 or E1[0..],E2,E3[..0]:* exon 2 to 4 are amplified (for *SeqType “genomic”* intron sequences in between are included, for *SeqType “cDNA”* only exon sequences are analysed).

  When an exon is selected in the table the location is entered into the field below automatically. When several exons are selected in the table, both exons are used, including the intron in between.
- New field *SeqType*: here you can select, if you analyse genomic or *cDNA*. For *cDNA* no introns are included.

Grouping of SeqPrimers is possible:

For grouped primers results can be shown as one sequence in operation *Sequence* (select the PrimerGroup in the section *Group*) .

To group SeqPrimers first add a PrimerGroup in the section *SeqPrimer group*. Then select the group in the field *Primer Group* of the SeqPrimer table for several SeqPrimers.

8.4 Operation Joining

A faster algorithm is used, therefore the processing is faster.

8.5 Operation Sequence

- A location overview is shown above the electropherogram: introns are marked yellow, exons are marked blue. Below the overview, the length of the each resultfile is indicated. Forward sequences are shown in a lighter color as reverse sequences.

  There is a context menu available in the *Location overview*:

  - *show peak height ratio*: Below the location overview a peak height ratio diagram for each base position is shown. The bars show positions with a high background. In case a background is present, there is a bar above the line for forward bases and a bar below the line for reverse bases. The higher the bar is, the bigger is the background. For typical heterozygous positions bars are present for forward and reverse bases.

    The bars are calculated as follows:

    peak height ratio = peak area highest not reference bases/(peak area highest not reference base + peak area reference base)
  - *show result files (default setting)*: the length of the result files is shown below the location overview. Reverse sequences are highlighted in a darker color.
  - *show original result files*: the length of the original result files (not shortened due to quality, SeqPrimers...) are shown. Sequence parts that are not used for analysis are highlighted grey.
- New section *Groups*: Here an entry is available in case SeqPrimer groups exist for the used SeqPrimers. When the entry is selected the electropherogram and sequence data is shown for all grouped SeqPrimers as one sequence. The length of the resultfiles are indicated as bars in the location overview.
- The new warning “X” is available in the *Positions/Resultfiles tabl*e/column *Warning*. This warning shows that there is a mutation present at the first or last two bases of the intron (splice site).
  Note: This function only works in case intronic regions are defined in operation *SeqPrimers [master file]*.
- Section *Show*: the new field *show* is available. Here you can select:
  - *Exon:* it is only jumped to exon positions when the jumpers in section *Show* are used (*check, edited, het. pos.* and *mism.*)
  - *all:* it is jumped to exon and intron positions.
- In case introns are sequenced, they can be saved in the HLA DB for result alleles with unknown intron sequences. Therefore use the new context menu item *intron* > *save as* in the Haplotype sequence.
- *Matching table*/tab*Total Result*/context menu *show sequence specific primers*: primer with lot numbers can be required.

9 Module SeqNext-HLA

Note: Old orders (analysed with previous version SeqHLA-454) can only be viewed with the new SeqNext-HLA. No edits can be done. To analyse old orders with the new SeqNext-HLA a new Run has to be started!

The module has been completely worked over, the functions are similar to the corresponding functions in the module **SeqNext**.

With the new version **Ion Torrent** as well as paired end sequencing data (e.g. **MiSeq**) can be analysed.

9.1 Operation ROIs [master file]

Note: HLA-Kits from the previous versions are not valid anymore. ROIs have to be newly defined, using the procedure described below.

Intronic regions can be defined.

Note: For intronic regions the sequences are only visualized, but not used for result calculation. For result calculation only exons are used. Introns can only be defined together with exons. Moreover only reads, that cover a part of the exons are aligned. Reads that cover intron areas only, are not aligned. Therefore, for big introns there might be a drop in coverage in the middle of the intron.

The only positions that are evaluated in the intron are the splice sites (first two bases and last two bases of the intron). If a mutation is found here, the warning “X” is be shown in the column “Warning” of the “Positions/Resultfiles” section. But the mutation is not used for result calculation (alleles are not separated or excluded due to splice site mutations).

To create an ROI do the following:

- Select the gene in the box *Gene*.
- Select one or several exons, that should be added as ROI, the sequence is shown in the field *Sequence*.
- The field *Location* shows “E..[0..0]” and the field *Seq Range* shows “0 and 0” by default (only exonic region). Here are some examples to explain which sequence is used:
  - E2[0..0]: exon 2
  - E2[-40..0]: last 40 bases of intron 1, exon 2
  - E2[10..0]: exon 2 without the first 10 bases
  - E2[0..-20]: exon 2 without the last 20 bases
  - E2[0..35]: exon 2, first 35 bases of intron 3
  - E2[-20..5]: last 20 bases of intron 1, exon 2, first 5 bases of intron 2
  - E2,E3: exon 2 and exon 3 including intron 2 in between.
- The sequence range can be changed manually for each location in the field *Location* or for all locations in the field *Seq Range*.
- For selected locations an ROI name is proposed in the field *Name*. By default it consists of gene name and exon number, e.g. *A-E01*. This name can be changed. Moreover a suffix can be added to the ROI name, therefore just enter it in the field *Suffix.*
- An entry can be made in the field *Panel*. The search field *Panel* is available in the operations *ROI [master file], ROI Groups [master file*] and *Run*. Therefore all ROIs belonging to one panel can be filtered and selected more easily.
- Press *[Add X->]* to add an ROI for each selected location to the *ROI list*: If for example A-E02

and A-E03 are selected, two ROIs are added, one for E02 and one for E03.

- Alternatively press *[Add 1->]* to add all selected locations as one ROI to the *ROI list*: If for example A-E02 and A-E03 are selected, they are added as one ROI, including the intron sequence in between.
- Press *[Save]*. (The *ROI List* will be empty then, because by default *not saved ROIs* is selected in the field *State*. To see all saved ROIs select the blanc in the field *State*).

In the field "*Suffix*" special characters can be entered to automatically define the ROI name:

- The default "*ROI name*" is GENE-LOC-suffix
  - GENE=Gene name
  - LOC=Location (Exon name)
- Example for gene A, E02:
  - In the "*Suffix*" field the following entry is listed: Test-GENE-location-LOC-ende
  - The following "*ROI name*" is created: Test-A-location-E02-end

Moreover the following columns can be edited in the *ROI List:*

- In the column *SeqType* the entry is *genomic* by default. This can not be changed.
- ROIs that should only be used for mapping can be defined, to filter "background reads". Therefore the option *only mapping* can be activated in the field *Analysis Mode* in the *ROI list*. If "only mapping" is checked, all reads will be aligned to the ROI but it will not be shown in operation *Sequence*.
  This option is e.g. useful for *DRB*. If only *DRB1* should be analysed, you can define *DRB1-DRB9* as ROIs. Then select *only mapping* for *DRB2-DRB9* in the *ROI List.* The pseudogene sequences are filtered and not mapped to *DRB1*. In operation *Sequence* only results for *DRB1* are shown.

  Several ROIs can be set to "only mapping" in the ROI List by selecting the ROIs and using the context menu *Analysis mode...*.
- *Settings Profiles* to be used for the analysis can be selected for ROIs. This Profile is always used for the ROI, even if another Profile is set in operation *Run*.
  In the column *Settings Profile*, an existing Profile can be selected for an ROI. Moreover the item *settings...* can be used in the context menu for one or several selected ROIs. When a Profile is created/selected, it is set in the column *Settings Profile* automatically.

Moreover *Amplicons/PCR Primers*, *Skipped parts* and *Skipped Sequences* can be defined for each ROI:

- Amplicons/PCR Primers are added in the same section. Primers have to be added as primer pairs, amplicons (excluding the primer sequences) are then created automatically. Note: In case primer sequences are entered all reads detected with these primer sequences are joined to the ROI (also background). For cases with background it is recommended to define amplicons only.
- Skipped parts/Skipped sequences can be defined in operation *Sequence* and are entered in the corresponding fields here automatically.

A context menu for one or several selected ROIs is available to

- remove ROIs
- copy ROIs
- select Profile settings
- set the Analysis mode "only mapping"

- active/inactive ROIs

9.2 Operation ROIs Group [master file]

ROIs can be grouped. This has the following advantages:

- Groups can be selected more quickly in the operation *Run*.
- Groups can be exported/imported for exchange with other institutes or **JSI**.
- For a Group a list of primer pairs can be added (button *[Import Primer]*). The primers are joined to the corresponding ROI and amplicons are generated automatically (entered in the section *Amplicons/PCR Primers* in *ROIs [master file]*).

To create a Group:

- Enter a group name in the field *Name.*
- Move all ROIs to group to the *ROI Group List* using *[Add ->].*
- Press *[Save]*.

9.3 Operation Run

9.3.1 Data

Ion Torrent as well as paired end sequencing data (e.g. **MiSeq**) can be analysed now.

9.3.2 Multiple Processing Cores

To make analysis faster multiple processing cores can be used to compute the result files of a run.

Several worker processes can be started, which process several files in parallel (In case the *Run* is started for several files, each worker processes one file). The number of worker processes should be related to the number of cores available on the server.

To define several worker processes:

- Create the file `RemoteComputer.txt` in the "bin"-folder of your **SEQUENCE Pilot** installation.
- Write into the file on which IP port to reach the remote computers, e.g.

  127.0.0.1:7301
  127.0.0.1:7302
  127.0.0.1:7303
  127.0.0.1:7304

Save the file and start **SEQUENCE Pilot** as usual. Now several worker processes are available automatically (in the example above 4 worker processes are started). They can be seen in the task manager (4 entries for *SeqNResultfilesWorker.exe*). They all quit automatically as soon as one quits **SEQUENCE Pilot**.

9.3.3 Importer

By default the Importer uses the maximum number of cores. Therefore the computer might be very slow during data is imported, which might be a problem in case other programs are used as well.
The number of cores used during import can be restricted in the `lis.ini`-file, located in the `bin`-directory of your SeqPilot installation. Therefore make the following entry in the section `[SeqNextHLA]`:

```
MaxImporterThreadCores=1
```

The number behind is the number of cores used for import.

9.3.4 Settings

Adjustable settings for each analysis are available: Settings can be saved as "Profiles". If the *as default* option is activated for a "Profile", the profile is used automatically when a *Run* is started (unless another *Profile/Settings* are selected manually).

The following settings can be adapted:

- Tab *Settings*:
  - *Reads:*
    - *[1] HT Basecalling*
      - *Basecalling coverage:* Decides, if a second base is called or regarded as background. The second bases has to reach this percentage value compared to the first base. **This value is very important and needs to be adapted according to your data!**

        Example: *Basecalling coverage* 20 %
        - Coverage: A: 100, C: 20: Allele 1: A is set; Allele 2: C is set.
        - Coverage: A: 100, C: 19: Allele 1: A is set; Allele 2: A is set. C is regarded as background and not called.
    - *[2] Required Coverage:* If the settings are not fulfilled the positions are greyed out and not used for result calculation.
      - *Min abs. coverage:* Minimum absolute coverage at each position.

        You can choose, if the value has to be reached in both sequencing directions together (select *combined*), or in forward and reverse separated (select *per dir*).
      - *Ratio read directions:* This is the radio between the forward and reverse coverage.
    - *[3] Expected Coverage warning:*
      - *Min abs coverage*: The value entered here is shown as a red dotted line in the electropherogram/coverage overview. In case there are positions with a coverage below this value you get a warning: There is the hint *expected* in the column *Coverage* of sections *ROIs* and *Locations* and the graph color of the *coverage graph* changes from grey to pink.
  - *Warning*: Warnings are shown in operation *Sequence* in the column *Warning* of the *Files, Genes, ROIs* and *Locations* table:
    - *Min reads per Allele*: Warning *F* in case the number of reads joined to allele 1 or 2 is below this value.

      Two options can be additionally set for this setting:
      - *perfect matches only*: Warning *F* in case the number of reads with a perfect match is below this value.
      - *+ complete seq.*: Warning *F* in case the number of reads with a perfect match and that cover the ROI completely is below this value. Note: This option does not work in case amplicons are defined.
    - *Allele 1/Allele 2 proportion*: Warning *A* in case the ratio of allele 1 to allele 2 coverage is below this value.

      Two options can be additionally set for this setting:
      - *perfect matches only*: only the reads with a perfect match to the called alleles are regarded for calculation of the ratio.
      - *+ complete seq.*: only reads with a perfect match to the called alleles and that

cover the ROI completely are regarded for calculation of the ratio. Note: This option does not work in case amplicons are defined.

- *Basecalling % coverage background*: Warning *B* in case a third allele (contamination or background) exceeds this value for at least one base position. You can choose if this warning should be calculated combined or per Haplotype. Use the jumper *bg BC* in the dialogue *Show* to jump to the corresponding positions.
- *Basecalling % Indel background*: Warning *N* in case the coverage of an insertion/deletion exceeds this value. You can choose if this warning should be calculated combined or per Haplotype. Use the jumper *bg Indel BC* in the dialogue *Show* to jump to the corresponding positions.
- *DRB pseudogene plausibility check*: Warning *P* in case *DRB1* expected pseudogenes are wrong.
- *Homo/Hetero check*: Warning *H* in case homozygous as well as heterozygous exons are present for one gene.
- *Splice site check*: Warning *X* in case there is a mutation at the splice site (first or last two bases of the intron). **Note**: This function will only work in case intron sequences are included in the ROI.

◦ *auto TV*: If active technical validation is done automatically in case no warnings exist.

◦ *Reads include PCR primers:* Please fill out the field (choose *yes* or *no*) in case you have amplicons or PCR primers defined, since it improves the assignment to the amplicons.

- *auto*: please change this setting in case you have amplicons defined!
- *yes*: select *yes* in case the reads include PCR primers.
- *no*: select *no* in case the reads do not include PCR primers.

◦ *Mark perfect matches count*: Reads that have a perfect match to an allele can be marked with a "*" in front of the read in operation *Sequence*. The calculation of the "*" is very time consuming, therefore you can enter a read count number into this field.

By default "2" is entered. This means only read that have a count of "2" or higher are marked. Reads that have a count of only "1" or not marked, even though they have a perfect match to an allele. The number can be adapted.

Note: for paired end data all pairs (forward and reverse read together), that show a perfect match to an allele are marked with a star. Here it is possible that the forward read has a match to another allele then the reverse read.

• *Tab Quality Score:* Here settings to exclude bases with bad quality from analysis can be set.

◦ *Quality Score Threshold*:

- In case there is a value entered (e.g. 15) only bases with a *Quality Score* above this value are counted to the coverage! Positions with bad quality are greyed out and not used for analysis.
- The Quality Score filter can be switched off by selecting *Quality Score threshold "off"*.
- In operation *Sequence* base positions, that were not called due to bad quality are shown as bases but are greyed out. In the tool tip of the forward, reverse and combined sequence, there is a new entry *Ns / Ignored,* showing how many bases at a position were ignored due to bad quality. In the coverage graph the coverage of bases that were not called due to bad quality is shown in a lighter grey.

◦ *Ignore Reads Threshold*: If the bases with bad quality in the complete read (including primers/adapters) exceeds this value, the read is ignored. Bad quality means, that the quality is below or equal the *Quality Score Threshold* setting, or that Ns were called by the sequencer.

◦ *Low Quality Score coverage warning*: Here a threshold to get a warning for ROIs with a bad Quality Score can be set: In case bases with low quality exceed the here entered percentage value at a position in an ROI, the warning *quality* is shown in the column

Coverage of section *ROIs* and *Locations* (operation *Sequence*). You can jump to positions with low quality using the jumper *quality* in the section *Show*.

- Tab *Trimming:* settings to trim adaptors or to automatically remove sequences at the ends of the reads can be entered. Note: In case the options *Adaptor* and *remove bases* are used, adaptors are removed first.
  - *Adaptor*: Here adaptor sequences can be entered to trim or discard reads. In your analysis you can get an overview about trimmed reads in section *Files*/context menu *show>Info*.

    The sequence entered here is searched for (also reverse complement in all reads): Enter an adapter sequence in 5'->3' direction. The following fields can be edited for each entered adapter:
    - *Position*
      - *auto*: It is decided automatically, if the adaptor locates at the 5' or 3' end.
      - *5'*: Adaptor can be found at the 5' end only.
      - *3'*: Adaptor can be found at the 3' end only.
    - *Error rate*: here a percentage value can be entered as error rate (wrong bases, that the adapter can contain).
    - *Overlap*: Here the minimum number of adapter bases that must overlap with the read can be entered. Example: overlap is *3*. There must be at least *3* adapter bases found in the read.
  - *Remove bases*: Removes bases at the beginning and/or end of each read. Adaptors are removed prior to that.
- Tab *Expert settings*:
  - *Basecalling:*
    - *Unique reads only:* If checked, the coverage of identical reads is set to 1.
  - *Read Processing*:
    - *Alignment evaluation*:
      - *Skip evaluation*: If checked, no alignment evaluation is done. The filters below are not used.
      - *Max mismatches*: Filter for mismatches a read can contain compared to the reference. The higher the number entered here is, the more mismatches are accepted. In case there are too many mismatches the read is discarded.
      - *Min matching bases:* Percentage of read bases, that have to match to the reference. In case less bases match, the read is discarded.
      - *Keep strong consensus*: The percentage of consecutive bases, that have to match to the reference without a mismatch between them. If this value is reached, the settings *Max mismatches* and *Min Matching bases* are overruled and the read is aligned. This filter is only applied if the read length is above 100.
    - *Compl. reads only*: If checked, reads that do not cover the complete amplicon are discarded.
    - *Barcode 5'+3'*: Choose this setting in case barcodes have to be present at both ends of the reads.
    - *Ignore paired end info*: If checked, paired end information is not used.
    - *Skip reads:* If active, only reads with a copy number that is higher than the entered *Skip count original* value take part in the analysis. Here the original reads from the sequencer (not the reads mapped to the ROI) are used. **Note**: This will reduce your coverage.

9.3.5 Start a run

To start a Run manually:

- Press *[...]* in the section *File* to select your file(s)
- Optionally change the *Settings*
- Add DNA number and optionally a barcode and project in the dialogue *Patients*
- Select ROIs or an ROI group for one or several selected patients
- Press *[Analyse]*

Automatic Start of a Run:
You also have the option to start a Run automatically. With this option all parameters needed for the Run have to be written in a `txt`-file. This `txt`-file is used to start the run instead of entering all parameters in the operation *Run* manually. The file can be created using any text editor and has the following format:
All fields have to be separated by "tab", several entries in one field by "semicolon". (In case one field is not filled out, it has to be separated by tabs anyway):

1. DNA No
2. Barcode (optional)
3. ROI or ROI Group
4. Path of the file(s)
5. Profile
6. Project
7. Resolution (4 digits, 2 digits, max)

In case several files/patients are analysed, copy the first entry into the next line and adapt it. Each line stands for on Run to be started.

Copy the `txt`-file into the folder `bin\Autorun\SeqNextHLA` of your `SeqPilot` installation. The runs are started automatically and the `txt`-file is moved to the backup folder.

9.4 Operation Joining

- The new search field *HLA Group* is available to search for orders using a certain HLA Group.
- With the context menu entry s*ettings...* in the *Upper table*, the *Settings/Profile* (adjusted in the operation *Run*) can be changed. In case the settings are changed the file is recalculated automatically using the new settings. After analysis is completed the file can be joined to the order again using the button *[Autojoin*].
- The dialogue *Info* (can be opened with the context menu entry *show Info* in the *Upper* and *Lower table*) was adapted.

9.5 Operation Sequence

9.5.1 Files/Genes/ROIs and Locations

All sections:

Several warnings can be present in the column *Warning.* If a warning is shown depends on your settings in the operation *Run.*

Note*:* In case the Warnings are present in the intron sequences they are only shown in sections that show intron data. These are *Files, ROIs* and the first entry of *Locations*.

- F: warning *Min reads per allele*: number of reads joined to allele 1 or allele 2 is below the corresponding setting.
- A: *Warning Allele 1/Allele 2 Proportion*: ratio of allele1 to allele 2 is below the corresponding setting.
- B: *Warning Basecalling % coverage* background: a third allele exceeds the settings. Use the jumper *bg BC* in the dialogue *Show* to jump to the corresponding positions.
- N: Warning Basecalling % Indel background: Warning *N* in case the coverage of an insertion/deletion exceeds the settings. Use the jumper *bg Indel BC* in the dialogue *Show* to jump to the corresponding positions.
- P: *DRB pseudogene plausibility check: DRB1* expected pseudogenes are wrong.
- H: *Homo/Hetero check*: Homozygous as well as heterozygous exons are present for one gene.
- X: This warning is shown when a mutation is present at the first or last two bases in the intron. Note: This function only works in case intronic regions are included in the ROI.

Sections ROIs and Locations:

Section *ROIs* lists all defined ROIs used for the Run.

Section *Locations* lists all exons that are covered by the selected ROI. For the location selected here, sequences are shown in the electropherogram. Introns are only displayed if the first entry (*ROI*) is selected.

Several *Coverage warnings* can be present in the column *Coverage* of section *ROIs* and *Locations*:

Note: Only one coverage warning is listed, the priority of the warning is as listed below (*not analysed* has the highest priority whereas *expected* has the lowest priority).

- *not analysed*: an error occurred. Please recalculate your file.
- *dropout*: No sequences are available for more than 90% of the ROI. (No sequences are also present in case the Required Coverage Settings (*Min abs coverage and/or Ratio read directions*; defined in operation *Run, section [2]*) are not fullfilled.
- *nocall*: There is no coverage at one or more positions of the ROI (Only occurs in case no Required Coverage Settings are set).
- *required:* This warning shows, that positions in the file are ignored (not analysed). Positions are ignored in case the Settings for the Required Coverage (*Min abs coverage and/or Ratio read directions*; defined in operation *Run, section [2]*) are not fulfilled.

  The ignored positions are greyed out in the sequences (forward, reverse and combined sequence).
- *quality*: This is shown in case the setting *Low Quality score coverage warning* (operation *Run*) is exceeded.
- *expected*: The coverage of one or several positions is below the *Expected coverage warning/Min abs coverage* (default: 100). The *Min abs coverage* is represented as a red dotted line in the electropherogram. Moreover the *Min abs coveage* is shown in the graphical overview below the location overview. The *Expected Coverage warning/Min abs coverage* can be set in your *Settings* when starting the Run.

You can jump to the corresponding positions with coverage warnings using the jumper *check required*, expected, *no call* and *quality* in the section *Show*.

Several new context menu entries are available in the *ROIs* section:

- *editing* >
  - *reanalyse:* reanalyses the ROI (result is reloaded)
  - *recalculate*: recalculates the ROI (using new settings or changed ROI)

- *settings...*: opens the dialogue *Settings* to change the settings for the selected ROI only. After changing the settings the ROI is recalculated automatically. Adaptors can not be changed for a single ROI, only for the Run. Therefore these fields are greyed out.

- *show*
  - *ROI Info:* Information about the ROI, such as amplicons, primers, skipped sequences and parts are shown. The same fields can be edited in the operation *ROI [master file].*
    The *ROI Info window* has two tabs*: Future analysis* and *Current analysis.* On tab *Current analysis* information about the ROI used for the current analysis are shown.
    ROIs can be changed on tab *Future analysis* only. In case the ROIs are changed, the ROIs that will be used for future runs are shown on tab *Future analysis*. Changes, such as new *Amplicons, Skipped Parts* or *Sequences* are highlighted. After doing a recalculate the highlighting disappears and the new ROI information are also shown on tab *Current analysis*.
  - *Info:* In the section *[General]* the number of assigned and aligned reads for the selected ROI are shown.

9.5.2 Location overview

A location overview is available above the electropherogram: It gives an overview of the selected location in the dialogue part *Locations*. You can jump to a location within the electropherogram by clicking on the overview.
Moreover, in case several amplicons are defined for the ROI, they are shown as red lines below the location overview.
The coverage of the location is shown graphically below the location overview. You can jump to the positions by clicking on the graph. The graph color changes from grey to pink if the coverage is below the *Expected Coverage*. (The Expected Coverage can be set in the *Settings* of the operation *Run* for each run. In case there a bases with a coverage below, there is also the hint *expected* in section *ROIs* and *Locations*).
Moreover the coverage of bases, that were not called due to bad quality, is shown in a lighter grey.

9.5.3 Combined, HT1 and HT2 sequence

There is several new information available in the tool tip of HT1, HT2 and combined sequence:

- *Coverage*: This is the number of called bases at a position.
- *Reads*: This is the number of reads covering a position.
- *Ns/Ignored*: This is the number of Ns and ignored bases (due to bad quality) at a position.
- Moreover quality scores are shown as a color code behind each base. The following colours are used by default:
  - quality score 1-10=Dark Red
  - quality score 11-20=Red
  - quality score 21-30=Yellow
  - quality score 31-40=Light Green
  - quality score 41-99=Dark Green

HT1 and HT2: Another allele can be entered into the haplotype field. When the [→] button is pressed the green HT sequence changes to the entered allele sequence and the alignment to this allele is shown.

There is the new count mode *ROI abs* available. In this count mode the first position of the ROI is 1.

9.5.4 Reads sequences

The "*" marking all reads with a perfect match to an allele is not available any longer by default. To get the mark select *Mark perfect matches count* in the S*ettings* when you start the Run.

Context menu entry *set fragment as HT1/ set fragment as HT2*: For paired end data you can decide if you want to set the selected read only as haplotype or the complete *pair*. Therefore you can select *read* or *pair*. **Note**: This option is only available for pairs.

9.5.5 Reads view

New field *Quality*: If this field is activated, quality scores are shown as a color code below the bases. The same color code as in the tooltip of the allele sequences is used.

New check box *Original reads*:

- The original length of the reads is shown, when the field *org. reads* is checked. The minimum and maximum number of bases, that were removed left and right is indicated in color.
- To see the sequence of the removed bases, the new context menu item *show original reads...* is available. When this item is selected a new *reads view* opens, showing original reads. Bases that were removed at the left and the right a highlighted.
- The read identifier is shown in the tool tip of the read, when the box *org. reads* is checked. Here the file and the line that lists the read is shown as read identifier.
- The read identifier is also shown as a tooltip when the new context menu *show original reads...* is used. For paired end sequencing data the paired end key is shown in this tooltip as well.

New button *[Search]* is available: Using this, you can either search for a sequence in the reads or for a read by giving the line number of the Next Generation Sequencing file. The *Search* window opens when *[Search]* is pressed:

- To search for a sequence:
  - Enter a sequence to be searched for in the field *Sequence*
  - After pressing *[Search]* all reads containing this sequence are listed in a new window. The sequence that was searched for is highlighted.
- To search for a read listed in the Next Generation Sequencing File:
  - Select the file in the first field of the section *Line*.
  - Enter the line number in the second field of the section *Line*.
  - When *Show read group* is active not only the read present in the entered line, but all identical reads are listed.
  - After pressing *[Search]* the read(s) are listed in a new window. The line information is available in the tooltip of the read, when *org. read* is *active*.

New button *[Filter]*: Here you can filter for reads that fulfil special criteria. When *[OK]* is pressed only those reads will be displayed in the reads view.

Context menu entry *set fragment as HT1/set fragment as HT2*: For paired end data you can decide, if you want to set the selected read only as haplotype or the complete *pair*. Therefore you can select *read* or *pair*. **Note**: This option is only available for pairs.

9.5.6 Matching table

There is the new column *S.Factor (Sorting Factor)* available. This factor sorts the alleles by their probability. The smaller the number is, the higher the allele combination is listed. It is calculated the

following way: number of mismatches divided by number of known exons for the allele

Example: Two possible allele combinations show 1 mismatches

- For allele combination 1: allele sequences are known for 4 exons (S.Factor is 0.25)
- For allele combination 2: allele sequences are known for 2 exons (S. Factor is 0.5)

The allele combination 1 is listed first in the *Matching table*.

9.5.7 Show

New jumpers in the *check* box are available:

- *local info:* jumps to positions with a local info deposited.
- *no call:* jumps to positions that have no coverage.
- *expected*: jumps positions with a coverage below the expected coverage (*Settings*, operation *Run*).
- *quality*: jumps to positions with low quality. Those positions are present in case the setting *Low Quality score coverage warning* (operation *Run*) is exceeded. In this case there is also the warning *quality* in the column *Coverage* of the *ROIs and Locations* sections.
- *required*: jumps to positions with a coverage below the required coverage (*Settings*, operation *Run*).
- *bg BC*: jumps to positions with a background warning for base changes. If those positions are present there is also the Warning *B* in the column *Warning* of the *ROIs/Location table.*
- *bg Indel*: jumps to positions with a background warning for indels. If those positions are present there is also the Warning *N* in the column *Warning* of the *ROIs/Location table.*
- *next seq:* for long ROIs/locations that are not sequenced completely (breaks in sequencing): moves to the next sequenced part of the ROI.

9.5.8 Validation

There is the option to make medical validation only possible in case a preferred result is set. Therefore make the following entry in the `lis.ini` file located in the `bin`-directory of your SeqPilot installation in the section `[SeqNextHLA]`:

```
CheckMVPreferredAllele=yes
```

10 Module MLPA

10.1 Operation Joining

- context menu in the Upper Table:
  - new item *edit* > is available for one or several selected resultfiles: Here *DNA number, type*, *gender* and *digested* can be changed. There is a separated sub menu present for each item.
  - *settings* (*Upper table*) is available for several selected files now. Using this a new window opens to change the *Mix* for all selected files simultaneously.

10.2 Operation Sequence

- *Files* table: the new column *MValidation* is available. Here the user and date of the medical validation is listed.
- Report: The sorting of the analysis mode methylation is now adjustable as in the analyse

diagram (before it was sorted by fragment length).

11 Talkmaster

Modules **SeqPatient**/**SeqNext**:

- For calculation of the position of a mutation/variation, previous mutations (e.g. insertions, deletions) are not regarded any more. Each mutation/variation is regarded “standalone” (in relation to reference sequence).
- The fields *Mark* and *MutDB* in the *Variation/Mutation table* can be exported.
- Transcript information can be exported for each mutation.
- The HGVS c- and p-nomenclature can be exported (HGVSNucName and HGVSAAName).
- As export format `vcf` is available now.

The following new items can be exported for module **SeqNext**:

- Average coverages (fwd/rev) for genes, ROIs and amplicons
- Number of variations for each result file for genes, ROIs and amplicons
- Disease number
- MutID

Module **SeqPatient**:

- All information (e.g. *Filename, Primer, SeqGeneparts*...) can be exported per variation now.
- In case several result files per mutation are listed, comments, TV and MV (User, date etc.) can be exported as well.

The following new items can be exported for module **SeqNext-HLA**:

- Average coverage per ROI
- Projects
- Comments
- Warnings

Module **MLPA**:

The copy number changes are exported in the same sorting as in the analyse diagram (before they were sorted by fragment length).