Manual for Demo Data

SEQUENCE Pilot module SeqPatient

Version 5.0.0 (July 31st 2018)



PREMIUM ANALYSIS TOOLS FOR GENETIC PROFESSIONALS



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 web: www.jsi-medisys.com

(for research use only)

Table of Contents

1 Introduction to SEQUENCE Pilot – SeqPatient	2
2 Installation	2
2.1 SEQUENCE Pilot	2
3 Installation of demo data	2
4 What is configured	2
5 The first login	3
6 SEQUENCE Pilot screen	3
7 Create your own user	4
8 Configuring of the base caller	5
9 Load sequencing result files	5
10 Joining	6
11 Worklist	9
12 Sequence	9
12.1 Order/Statistic/Protocol/Family	.10
12.2 Reference	.11
12.3 Genes	.11
12.4 Sources	. 12
12.5 Positions/Resultfiles	. 12
12.6 Variation/Mutation Table	.13
12.7 Electropherogram	. 14
12.8 Show	. 16
12.9 Functions	. 17
13 Analysis of the demo data	. 18
13.1 COL1A1	. 18
13.1.1 Mutation in E13 position -14	.19
13.1.2 Mutation in E22 position 38	.22
13.1.3 Mutation in E43 position 34	.23
13.1.3.1 Edit Bases	.23
13.1.3.2 Undo Function	.25
13.1.4 Mutation in E43 position 37	.25
13.1.5 Mutation in E43 position +19	.26
13.1.6 Mutation in E49 position 19	.27
13.1.7 Mutation in E49 position 58	.28
13.1.8 Mutation in E52 position -12	.29
13.1.9 Result of the COL1A1 analysis	.30
13.1.10 Validation	. 30
13.1.11 Printing of a report	.31
13.1.12 Basecaller and Auto Edit	.33
13.1.13 How to analyse the order again	.35
13.2 Mareno	. 36
13.2.1 Sample with DNA number 159	.36
13.2.2 Sample with DNA number 160 – Statistic Warnings	.37
13.3 Discussion of the analysis	. 39
14 Archiving, Statistics and Mutation database	.40
15 Shortcuts	.40

1 Introduction to **SEQUENCE** Pilot – SeqPatient

The module **SeqPatient** is made for bioinformatical analysis of Sanger sequencing data and is compatible with all sequencers. It uses an unique learning basecaller based on statistic data and has its own auto edit function. Moreover an internal mutation database is integrated. After the analysis, a diagnostic report for each sample is received.

SeqPatient is a simple, easy to use desktop application including a variety of functionalities to fit the program to many different specific needs.

2 Installation

2.1 SEQUENCE Pilot

In case you have done the Installation of SeqPilot already, please proceed with chapter 3.

For the installation of **SEQUENCE Pilot**, please do the following:

- Go to our website <u>http://www.jsi-medisys.com/free-trial-license</u> to obtain a free trial license.
- You will receive installation instructions for our latest version of SeqPilot.

3 Installation of demo data

The demo data package <code>SeqPatient_DemoData.exe</code> includes the complete configuration and example files for the genes *COL1A1* and *Mareno* and shows step by step how to load and analyse the data.

Please note, that this package shouldn't be installed to systems already including data for genes *COL1A1* or *Mareno*. This might overwrite your local configuration and statistic values. In this case, please contact our support associates (*support@jsi-medisys.de*) for a separated installation. For the installation of the demo data, do the following:

- Stop SEQUENCE Pilot. The program should not run during installation !!!
- Go to our website <u>http://www.jsi-medisys.de/DownLoadFiles/DemoData/SeqPatient_DemoData.exe</u>
- Press [Execute] in the dialogue file download security warning.
- Press [Execute] again in case of another security warning.
- Press [Next] in the dialogue Setup SeqPatient Demo Data.
- Check the box I accept the agreement after reading the Software License Agreement and press [Next].
- Press [Next] after reading the Info and Installation Instructions for Sequence Pilot demo data.
- Select the destination folder, where the SEQUENCE Pilot demo data should be installed. By default this is c:\SeqPilot. Then press [Next].
- Press [Install].
- Press [Finish] when the SEQUENCE Pilot installation is completed.

4 What is configured

All sequencing result files and configuration data are extracted and deposited within the Demodirectory of your installation. The COLIA1.txt and Mareno.txt gene files – previously downloaded from <u>www.ensembl.org</u> are copied to the GeneFiles-directory of the installation. View them with any text editor or check the contents with **SEQUENCE Pilot** (menu item *SeqPatient/Gene Admin*). For each gene, information such as gene name, loaded files, information for each sequence version, coding sequence (CDS), mRNA, translation, cDNA and known variations are listed.

Several master files for *Amp. Modules*, *Sequencing Primer* and *Thresholds* are added to the system. They are shown within the corresponding operations in the category *SeqPatient* on the left side of the screen. Please refer to the **Manual SEQUENCE Pilot** - **module SeqPatient** for a detailed description.

Moreover reference sequences, statistic values and found mutations from previously analysed samples for *COL1A1* and *Mareno* are loaded.

5 The first login

Once **SEQUENCE Pilot** is installed and licensed, login using the username *jsi* (leave the password field empty).



In **SEQUENCE Pilot** a **menu** is available which consists of the items *System, SeqPatient* and *Help.* Furthermore several **categories** and **operations** are shown on the left side of the screen. Open a category or an operation by clicking on the corresponding icon. Then different screens with different functionalities are opened.

6 **SEQUENCE** Pilot screen

7 Create your own user

How to get there: Category Administration Operation Users

After the first login, create a new user. After the new user has been created, the user *jsi* is deleted automatically.

Click on the category *Administration* on the left side of the screen and then click on the operation *Users*.

Image: Service Serv	Image: Service Serv	-	Administration	For re	search	use o	nly		
ers User: test Password: Confirm Password: Name First: Last: Function Address Phone numbers Phone numbers Phone is authorised to: E-Mail Address Send User is authorised to:	ers Use: lest ractive Password: Confirm Password: ractive User First: Last: Function Function Phone numbers Phone Fax: Mobile: E-Mail Address Send User is authorised to: Send		Logout						
Password: Name First Last: Function Function Function Address Phone numbers Pinone: Fax: Send User is authorised to: V edit masterfiles V edit masterfiles V delete orders	Password: Confirm Password: Name First Last: Function Function Address Phone numbers Phones Fax: Mobile: E-Mail Address ✓ edit masterfiles	ers User:	Test	✓ active					
Name First Last: Function Address Phone numbers Phone: Fax: Phone: Fax: E-Mail Address send User is authorised to: Ver is authorised to: Send	Name First Last Function Address Phone numbers Phone Fax Phone send User is authorised to: Image: Send User is authorised to: Image: Send User is authorised to: Image: Send	Password:		Confirm Passwor	± [
Function Address Address Phone numbers Phone Phone Fax: Mobile: E-Mail Address Send	Function	Name First:					Active	User test	
Function Address Address Phone numbers Phone: Fax: Phone: Fax: Mobile: Be-Mail Address E-Mail Address Send User is authorised to: send	Function Address Phone numbers Phone: Fax: Mobile: Mobile: E-Mail Address Send User is authorised to: Image: Send Image: Send Image: Send Image: Send								
Address Address Phone numbers Phone immers Phone Fax Mobile: E-Mail Address User is authorised to: Send User is authorised to: Send	Address Address Phone numbers Phone is authorised to: User is authorised to: Verify edit users Verif	Function				-	1		
Address Phone numbers Phone: Phone: Fax: Phone: Fax: Send User is authorised to: Send User is authorised to: Send	Address Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address E-Mail Address Send User is authorised to: Send User is authorised to: Send								
Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address User is authorised to: Green edit masterfiles Green edit m	Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address User is authorised to: Send User is authorised to: Send	Address					3		
Phone numbers Phone Fax: Phone Fax: Be-Mail Address E-Mail Address User is authorised to: Gr edit users Gr edit users Gr edit masterfiles Gr make medical validation Gr delete orders	Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address E-Mail Address Send User is authorised to: Great users G	1							
Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address E-Mail Address User is authorised to: Green distance of the server of th	Phone numbers Phone: Phone: Fax: Mobile: E-Mail Address E-Mail Address Gradie users	-							
Phone numbers Phone: Phone: Fax: Mobile: Fax: Send User is authorised to: Greedit masterfiles Freedit	Phone numbers Phone: Phone: Fax: Mobile: Fax: Send Fax: Send User is authorised to: Gr edit users Gr edit users Gr edit users Gr edit masterfiles Gr delete orders Fr delete orders								
Phone numbers Phone Phon	Phone numbers Phone: Phone: Fax: Mobile: Fax: E-Mail Address E-Mail Address E-Mail Address Fax: Send	[
E-Mail Address	E-Mail Address								
E-Mail Address User is authorised to: Vertice dit users Vertice dite dite dite dite dite dite dite dit	E-Mail Address	Phone numbers		Eav.					
E-Mail Address	E-Mail Addresssend	Phone numbers Phone:		Fax:					
E-Mail Address	E-Mail Address	Phone numbers Phone: Mobile:		Fax:					
User is authorised to: Vertice edit users Vertice edit masterfiles Vertice edit masterfiles Vertice edit masterfiles Vertice edit masterfiles Vertice edit masterfiles Vertice edit masterfiles	User is authorised to:	Phone numbers Phone: Mobile:		Fax:					
User is authorised to:	User is authorised to: ✓ edit users ✓ edit masterfiles ✓ make medical validation ✓ delete orders	Phone numbers Phone: Mobile: E-Mail Address		Fax:		send			
User is authorised to:	User is authorised to:	Phone numbers Phone: Mobile: E-Mail Address		Fax:		send			
 ✓ edit masterfiles ✓ make medical validation ✓ delete orders 	 ✓ edit masterfiles ✓ make medical validation ✓ delete orders 	Phone numbers Phone: Mobile: E-Mail Address		Fax:		send			
✓ make medical validation ✓ delete orders	I♥ make medical validation ↓ delete orders	Phone numbers Phone: Mobile: E-Mail Address User is authorised I	to;	Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised 1 I v edit u V edit u	to: sers nasterfiles	Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised t Viser is authorised t v edit u v edit u v edit u v edit u v delete	to: isers nasterfiles e medical validatio e orders	Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised I I edit u I edit u I edit u I edit u I edit u I edit u I edit u	to: sers nasterfiles medical validatio e orders	Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised f I v edit u V edit u V edit m V edit u V edit n	to: isers nasterfiles e orders	- Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised f I v edit u I v edit u I v edit u I v deleta	to: isers masterfiles medical validation e orders	Fax:		send			
		Phone numbers Phone: Mobile: E-Mail Address User is authorised t IV edit u IV edit u IV edit u IV edit u IV edit u IV edit u	to: sers nasterfiles medical validation e orders			send			
Previous	President New	Phone numbers Phone: Mobile: E-Mail Address User is authorised t I I edit u I I edit u I I edit u I I delete	to: 	Fax:		send			
Erevious Negt	Previous Negt	Phone numbers Phone: Mobile: E-Mail Address User is authorised 1 I I edit u I I edit u I I edit u I I delete	to: isers nasterfiles e orders Brevious	Pax:		send			

- Fill out the field *User* (max. 4 letters).
- Mark active and all options user is authorized to....
- Press [Save].
- Select the operation Logout and then login with the new user.

8 Configuring of the base caller

How to get there: Category SeqPatient Operation Thresholds [master file]

Before the files can be loaded, configure the base caller to ensure that the same results as described on the following pages are shown. Ensure that there are the same settings as in the picture below. If not, change the values and press [Save].

Thresholds	
Directory: ABI	Name: default
Base: *	Sequencer Type:
Leading Seq.:	
Absolut Count: 0	
Percent Count: 40 %	
Warning Height: 0	
Warning Phred: 0	

9 Load sequencing result files

How to get there: Menu SeqPatient Item Load Sequencing Resultfiles...



If this item is selected, the following dialogue is opened:

Load Sequer	ncing Resultfiles	×
Destination:	ABI	•
	Files	Cancel

Within the field *Destination*, select the sequencer being worked with. By default the sequencer type *ABI* is already preselected. Please do not change this entry, because the demo files are made with ABI sequencers.

Loading of the sequencing result files for COL1A1:

- Press [Files...].
- Open the directory SeqPilot/Demo/COL1A1.
- Select all result files with ctrl+A.
- Press [Open] to load the selected result files.

Load the Mareno files in the same way. They are located in the folder Mareno of the Demo-directory.

Follow the process of loading by looking at the status bar, shown at the lower right of the **SEQUENCE Pilot** screen:



The dot will be **red**, if the loaded files are still being imported. The number of remaining files to import is shown after the dot. The dot will be **green**, if all loaded files are imported. The number after the dot is *0* then.

Depending on the hardware used, the loading of all demo result files needs approximately two to three minutes.

10 Joining

How to get there: Category SeqPatient Operation Joining

By default this operation lists all result files or orders loaded at the current date. (To show orders from other dates, remove the mark in *Date/Select Orders* and press [Search]).

All result files that have the same DNA number are joined to one order. Result files that could not be joined to an order are shown in the *Upper Table* whereas result files that could be joined to orders and are ready for analysis are shown in the *Lower Table*.

	Upper 7	Table		Se	leçt Or	ders			Low	er Tab	ble	
Sequence Pilot												- 0
stem Seq <u>P</u> atient	Help											
Administration Joi	ining											
LIS									21.965		-W.	T
SenPatient	No DNA No	Source	Gene Reference	E Location	Dir Position	AmpModule 5	SeqPrimer	RunDat	e Alleles	/Variations	Filename	Hint
		gene file	Mareno	E10	TWO .	Mar-E10 I	MAR-EIOF	06/04/2	2005 C		AU/_mar_E10F_159.ab1 B06 mar E10R 159.ab1	no DNA No
3	3	gene file	Mareno	E10	rev	Mar-E10	MAR-E10R	06/01/2	2005 C		B06_mar_E10R_160.ab1	no DNA No
Joining 4	ł	gene file	Mareno	E10	fwd	Mar-E10 1	MAR-E10F	06/01/2	2005 C		B07_mar_E10F_160.ab1	no DNA No
Worklist												
					1							
2000 S	Select Orders					n.			27		2.0	2414
Sequence [2ate: 🔽 31.07.3	2018 💌	Dna No.:		ample S.		-	Gene:		•	Search	
Town in the second seco					— V				S.,		Class	Autotoin
	Date Range: 10 131.07.	2018	Order No.:		Project:	0.00					Glear	
Archiving	State: any (not a	rchived) 🔻	Pat. Name:			(• act. (ina	ct, (all					
receiving	· · · · ·				_							
TTTT	/alidation: any	*	Patient ID:									Liodate
<u> </u>												gradie
mp. Modules		- Norma		DNIA No.	Due file (Due e	Data	1	Direc	Annabladida	CarDeiner	Allalas (Mailallana	
		INdine	- State	I DNA NO	Pronie/Proc.	Date	Location	Direc	AmpMouule	Segenmer	Alleles / Variauoris	Filename
A	COL 101	P*12	compt.	12	SEQUOLIA	05/27/2005	F1	fud	COL 141-E01	CIAIEIE		(12 COL161-E01 C161E1E 2222) ab1
	COLIAI		~	12		05/27/2005	F1	rev	COL 1A1-E01	CIAIEIP		(12_COLIA1E01_CIAIE1E_2222).ab1
Seq. Primer	and COLIAI			12		05/07/2005	F2	find	COL 141-E02	C1A1E2E		(12_COL1A1-E02_C1A1E2E_2222) ab1
master file]	COLIA1		~	12		06/07/2005	E2	rev	COL 1A1-E02	C1A1E2B		(12_COLIA1-E02_CIA1E2B_2222).ab1
	ad COLIA1		~	12		07/08/2005	E3	fwd	COL1A1-E03	C1A1E3F		(12 COL1A1-E03 C1A1E3F 2222).ab1
ALGO	COLIA1		~	12		06/08/2005	E3	rev	COL1A1-E03	C1A1E3R		(12 COL1A1-E03 C1A1E3R 2222).ab1
Statistic	COL1A1		~	12		05/27/2005	E4	fwd	COL1A1-E0	C1A1E4F		(12_COL1A1-E04_C1A1E4F_2222).ab1
master file]	COL1A1		~	12		05/27/2005	E4	rev	COL1A1-E04	C1A1E4R		(12_COL1A1-E04_C1A1E4R_2222).ab1
	COL1A1		~	12		06/07/2005	E5	fwd	COL1A1-E05	C1A1E5F		(12_COL1A1-E05_C1A1E5F_2222).ab1
0000	COL1A1		N	12		06/07/2005	E5	rev	COL1A1-E05	C1A1E5R		(12_COL1A1-E05_C1A1E5R_2222).ab1
LL_	and COLIA1		~,~	12		05/27/2005	E6,E7	fwd	COL1A1-E0	C1A1E6-7F		(12_COL1A1-E06-07_C1A1E6-7F_2222).ab1
master file]	COL1A1		~,~	12		06/20/2005	E6,E7	rev	COL1A1-E0	C1A1E6-7R		(12_COL1A1-E06-07_C1A1E6-7R_2222).ab1
	COL1A1		~,~	12		05/27/2005	E8,E9	fwd	COL1A1-E0	C1A1E8-9F		(12_COL1A1-E08-09_C1A1E8-9F_2222).ab1
	-add COLIA1		~,~	12		05/27/2005	E8,E9	rev	COL1A1-E0	C1A1E8-9R	2	(12_COL1A1-E08-09_C1A1E8-9R_2222).ab1
•••	COLIA1		~,~	12		07/01/2005	E10,E11	fwd	COL 1A1-E1	C1A1E10-11F		(12_COL1A1-E10-11_C1A1E10-11F_2222).ab1
Vew RF Dirs	COLIAI		~,~	12		05/27/2005	E10,E11	rev	COL 141 E12	CIA1E10-11	٩	(12_COLIA1-E10-11_CIA1E10-11R_2222).ab1
fugarer me1	COLIAI		~	12		05/27/2005	E12 E17	1WU	COLIAI-EI2	CIAIE12P		(12_C0L1A1=12_C1A1E12F_2222).db1 (12_C0L1A1=E12_C1A1E12F_2222).db1
	COLIAI		01.07	12		05/08/2005	E13 E14	fud	COL1A1-E1	C1A1E13-14	- C	(12_COLIATE12_14_CIATE12_14E_2222).ab1
****	COLIA1		~~~~	12		06/08/2005	E13.E14	rev	COLIAI-E1	C1A1E13-14	2 C	(12 COLIA1-E13-14 CIA1E13-14R 2222).ab1
Thresholds	COLIA1		~,~	12		05/27/2005	E15,E16	fwd	COL1A1-E1	C1A1E15-16	5	(12 COL1A1-E15-16 C1A1E15-16F 2222).ab1
[master file]	COLIA1		~,~	12		05/27/2005	E15,E16	rev	COLIA1-E1	C1A1E15-16	2	(12_COL1A1-E15-16_C1A1E15-16R_2222).ab1
	COL1A1		~,~	12		05/27/2005	E17,E18	fwd	COL1A1-E1	C1A1E17-18	-	(12_COL1A1-E17-18_C1A1E17-18F_2222).ab1
	COL1A1		~,~	12		05/27/2005	E17,E18	rev	COL1A1-E1	C1A1E17-18	ર	(12_COL1A1-E17-18_C1A1E17-18R_2222).ab1
	COL1A1		~,~	12		06/08/2005	E19,E20	fwd	COL1A1-E1	C1A1E19-20F		(12_COL 1A 1-E19-20_C1A 1E19-20F_2222).ab1
	- OF COLIAI		~,~	12		06/08/2005	E19,E20	rev	COL1A1-E1	C1A1E19-20F	2	(12_COL1A1-E19-20_C1A1E19-20R_2222).ab1
	and GOLDIE					or too looor	F21.F22	fixed	COL 1A1-F2	C1A1E21-22E	= C	(12 COL1A1-E21-22 C1A1E21-22E 2222) ab1
	COL1A1		~,~	12		06/20/2005		1114				(15_COCIMICEI-22_CIMICEI-22 _2222).001
	COL1A1		~,~ ~,~	12 12		06/08/2005	E21,E22	rev	COL1A1-E2	C1A1E21-22	R C	(12_COLIA1-E21-22_CIA1E21-22R_2222).ab1
	and COLIA1 and COLIA1 and COLIA1		~,~ ~,~ ~	12 12 12		06/08/2005 05/27/2005	E21,E22 E23	rev fwd	COL1A1-E2 COL1A1-E23	C1A1E21-22F	R C	(12_COL1A1-E21-22_C1A1E21-22R_2222).ab1 (12_COL1A1-E21-22_C1A1E21-22R_2222).ab1 (12_COL1A1-E23_C1A1E23F_2222).ab1
	ed COLIA1 ed COLIA1 ed COLIA1 ed COLIA1		~,~ ~,~ ~ ~	12 12 12 12		06/08/2005 06/08/2005 05/27/2005	E21,E22 E23 E23	rev fwd rev	COL1A1-E2 COL1A1-E23 COL1A1-E23	C1A1E21-22F C1A1E23F C1A1E23R	ξ C	(12_COL1A1-E21-22_C1A1E23E_2222).ab1 (12_COL1A1-E23_C1A1E23E_22222).ab1 (12_COL1A1-E23_C1A1E23E_2222).ab1 (12_COL1A1-E23_C1A1E23E_2222).ab1

For automatic generation of orders, special *sample naming conventions* have to be used, to ensure that the program gets all necessary data. This information has to be entered within the sample naming field of the sequencer in the following way:



Any information can be entered before and after the parentheses. The **DNA number is obligatory** for automatic joining of result files to an order.

So what can we see in our demo data?

The *COL1A1* files could automatically be joined to an order due to correct sample naming conventions. The order is found in the *Lower Table*, and is ready for analysis. All result files joined to this order are listed if you press the "+" in the column *Order*. To see the used sample naming, right click on one of the result files and select the item *show annotations...* from the context menu.



The marked line shows the used *sample naming conventions*:

Annotation		
Filename:/N GELP GELN comment signal factor User TUBE WODL	<pre>WewResultFiles/ABI/(12_COL1A1-E01_C1A1E1F_2222).ab1 : : : : : 3871, 2628, 2359, 3311 : Molek : B1 . 3220</pre>	
SMPL machine number lane number runstart runstop runstart time runstop time FW0_	: (12 COLIAI-EO1 CIAIEIF 2222) : Flip-16114-008 : 1792 : 05/27/2005 : 05/27/2005 : 17:17:01 : 18:13:33 : GATC	
	Save as	ОК

The four *Mareno* files are listed in the *Upper Table*. They could not be joined to an order automatically because the DNA No is missing. Right click on one of the result files and select the item *show annotations* from the context menu to open the following dialogue. The marked line shows the entry that is not compliant with the desired *sample naming conventions*.

Annotation		
Filename:/N GELP GELN comment signal factor User THEF	<pre>WewResultFiles/ABI/A07_mar_E10F_159.ab1 : : : : : : 1389, 1500, 1211, 845 : .</pre>	
MODI	. 3100	
SMPL	: A07 mar E10F 159	
machine number lane number runstart runstop runstart time runstop time FWO_	:: 3100 : 1280 : 06/04/2005 : 06/04/2005 : 00:50:09 : 03:15:24 : GATC	
		Save as OK

Files in the *Upper Table* have to be joined to an order manually before analysing. To join this result files to an order, we have to enter a DNA number for each file. Therefore, right click on all *Mareno* files in the *Upper Table* that show the number **159** in the file name and select the item *edit DNA No....*



Enter the DNA number **159** and then press *[OK]*. For the other two *Mareno* files please do the same, but enter DNA number **160**.

To join the files to an order, click on [Autojoin] in the dialogue part Select Orders.

Select Orders					
Date: 27.06.2011 -	Dna <u>N</u> o.:	Sample S.:	Gene:	Search	Join
Date Range: 27.06.2011 -	Order No.:	Project:		Clear	Autojoin
State: any (not archived) 💌	Pat. Name:	☞ act. ○ inact. ○ all			
Validation: any 💌	Patient ID:				Update
					opuate

The files are then moved to the Lower Table.

Now look at the column *Alleles/Variations* in the *Lower Table*. First press the + in front of the order line in the *Lower Table* to expand the order and see all joined result files. The column *Alleles/Variations* shows the mutations/variations, that were detected in each result file (*C*: base change; *I*: insertion, *D*: deletion; *M*: missing bases).

11 Worklist

How to get there: Category SeqPatient Operation Worklist

This operation lists all existing orders and their joined result files. The dialogue part *Select Orders* and the *Worklist Table* are the same as *Select Orders* and the *Lower Table* in *Joining* (chapter 10). To proceed with the analysis select the first order in the *Lower Table* of *Joining* or in the *Worklist Table* and switch to the operation *Sequence*.

12 Sequence

How to get there: Category SeqPatient Operation Sequence

The operation *Sequence* is the main operation for the analysis. Before analysing our demo data, important features of this operation are explained below. Double click the first order in the *Joining* or *Worklist* to open operation *Sequence*.



12.1 Order/Statistic/Protocol/Family

- Order: This tab shows all important information about the order like order number, patient, DNA number, date etc.
- Statistic:



Sequences of the same chemistry mostly show the same electropherogram characteristics, that is background, peak compression, curve height etc.. Same chemistry means same gene, exon, sequencer, sequencing direction, amp. module and seq. primer. Therefore, they can be used for statistic evaluation. For each selected base position of a result file, this tab shows the

corresponding statistic values (statistic value of the peak area, statistic value of the background, standard deviation). These were calculated over all existing result files of the same chemistry already archived (see chapter 14), where the selected base was detected homozygously. Heterozygous positions are not used to calculate the statistic. The present statistics are listed in the operation *Statistic [master file]* (see also chapter 14). The advantage is, that all analyses can be compared with previous analyses. Therefore the statistic helps to decide whether a postion is heterozygous.

- Protocol: This tab shows a list of all base changes as well as ignored bases to left and to right of the currently selected gene within the dialogue part Genes for each result file. By selecting an entry, the cursor jumps to the corresponding base position within the Electropherogram.
- Family: This tab shows a list of all orders belonging to the same family ID. The current order is always shown in the first line. Click on one of the other orders to open it.

12.2 Reference

A reference sequence is shown that corresponds to the sequence of the result file in the electropherogram. A reference sequence is a sequence of the same chemistry (gene, exon, sequencer, direction, amp module and seq. primer) which was analysed previously and defined as a reference. Only one reference sequence can be saved for each chemistry.

To save a result file sequence as a reference, right-click on a base of the result file sequence (see 13.6 *Electropherogram* for result file sequence) and select the item *reference* > *save as* from the context menu.

If a reference sequence with the same chemistry has been saved already, there is the option to overwrite it. All reference sequences are listed in the lower table of the operation *Statistic [master file]* (see chapter 14 for *Statistic [master file]*).

12.3 Genes

This dialogue part shows all examined genes of an order. To open a gene map which displays the analysed parts of the gene and the existing result files, right-click on the gene entry *COL1A1* and select the context menu item *show* > *gene map....*



The dialogue Gene Map opens:

еМар	• 🔀
Order 110390001 Gene: COL1A1	
5 (12 C 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	>- ■ 子
	Š
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	52 3
	ОК

The shown gene parts are highlighted with different colours:

- Green highlighted exons show that these exons are sequenced (at least one result file exists).
- Exons that are not sequenced are highlighted orange.
- Light green parts at the beginning and/or end of a sequenced exon are analysed intron parts.
- Introns parts that are not sequenced are highlighted pink.
- Red highlighted parts within an exon or intron mark detected mutations/variations.

The arrows above and below the gene parts show the existing result files in forward and reverse direction respectively. The arrows can be coloured differently:

- $\blacksquare \quad \underbrace{>1_{138...}} \qquad \text{Result file exists.}$
- Statistics and the second s
- Figure 1 138-14...> Result file is required (files not loaded yet).

So what is shown in this case? The complete *COL1A1* gene has been sequenced in both directions. We have two result files for all exons, one forward and one reverse direction.

12.4 Sources

Lists the sources that are used as reference sequences (as set up in operation *SeqPrimer [master file]*). Possible sources are *genome, genefile, fasta file* or *chromosome range*. The Reference e.g. genome *hg19* is listed in column *Reference*.

12.5 Positions/Resultfiles

This dialogue part shows all loaded result files belonging to the selected gene in the dialogue part *Genes*. To display the result file base sequences of an exon within the electropherogram, click on it once.

12.6 Variation/Mutation Table

This table shows all detected variations/mutations.

P m	Index	Gene	Transcript	Location	Pos.	T	Nuc Change	AA Change	S	Hint	web Ref.	c. HGVS	p. HGVS	mut Entry	mut Effect
0 🗹	1	COL1A1	ENST00000225964	E13	-14	С	t -> g (het)				rs176394	c.850-14		0/4 of 13	intronic v
	2	COL1A1	ENST00000225964	E22	38 (1490)	С	G -> A (het)	G -> D (497)				c.1490G>A	p.Gly497	0/1 of 15	patho. m
	3	COL1A1	ENST00000225964	E43	34 (3070)	С	G -> C (het)	D -> H (1024)				c.3070G>C	p.Asp102		
	4	COL1A1	ENST00000225964	E43	37 (3073)	С	G -> T (het)	G -> C (1025)				c.3073G>T	p.Gly102		
	5	COL1A1	ENST00000225964	E43	+19	С	a -> c (het)					c.3090+			
	6	COL 1A1	ENST00000225964	E49	19 (3541) /	I	C (homo)	[STOP] AA		R		c.3540_3	p.Gly118		
	7	COL1A1	ENST00000225964	E49	58 (3580)	С	G -> C (het)	D -> H (1194)				c.3580G>C	p.Asp119		
	8	COL1A1	ENST00000225964	E52	-12	С	g -> a (het)				rs224949	c.4240-1		5/7 of 15	intronic v
<													-		3

The table is divided into the tabs *all, distinct, other* and *filter*. The number of mutations present on each tab is shown as well:

- *all:* lists all mutations
- *distinct:* lists mutations where the base is mutated in both sequencing directions
- other: lists mutations where one sequence (forward or reverse) shows the reference sequence
- *filter*: lists mutations that were filtered using the SNP database filter (in the Demo Data no SNP databases are installed for filtering. For more information, please have a look at our User Manual SeqPatient (Menu Help > Manuals) or contact our support team).

If an entry is selected within this table, the corresponding position will be shown in the *Electropherogram*. For the analysis, it is usually enough to check the here listed entries.

The Variation/Mutation Table is explained in the following table:

Column	Description
Print	The mutation is not printed in the report.
	The mutation is printed in the report table.
mouse-click	The mutation is printed in the report table and the corresponding electropherogram is also printed.
mut DB	Mark the box if the mutation should be added to the mutation database during the archiving process (chapter 14).
Index	Here the table entries are indexed.
Gene	Name of the gene
Transcript	Name of the transcript
Location	Location (e.g. exon) of the detected variation/mutation.
Pos.	Nucleotide position of the detected variation/mutation. The first number shows the position in exon/intron count mode, the number in parenthesis shows the position in cDNA count mode. The later is only shown if the mutation is localized within an exon. In case of insertions or deletions only the forward sequence position is shown.
Туре	Variation/Mutation Type: <i>C</i> (change of a nucleotide (SNP)), <i>D</i> (deletion), <i>I</i> (insertion), <i>M</i> (missing bases)
Nuc Change	The changed bases.
AA Change	The resulting changes in the amino acid sequence.
State	State of the loaded result file: V (technical validated), S (medical validated)
Hint	The entry is <i>RF changed</i> if the reading frame is changed by the variation/mutation. (Reading frame has to be available!).

web Ref.	Reference index of the detected variation, if it is described in the downloaded gene file.
HGVS nomenclature	Here the HGVS nomenclature for each table entry is listed.
mut Entry	In case the mutation is present in the internal mutation database the column <i>mut Entry</i> shows the frequency (homo/hetero) of the mutation.
mut Effect	If the mutation is in the mutation database and a information is entered, it will be shown in this column.
TValidation	User and date of technical validation.
MValidation	User and date of medical validation.

Note: In case SNP databases are installed, information from the SNP database for a mutation can be shown in additional columns (in the Demo Data no SNP databases are installed for filtering. For more information, please have a look at our **User Manual SeqPatient** (Menu *Help > Manuals*) or contact our support team).

12.7 Electropherogram

In this dialogue part the electropherogram curves of the result files are shown. The reference sequence, the result file sequences and the combined sequences (of both result files) for bases and amino acids, are also shown. The location in the result file (exon/intron) is also displayed.



There are different count modes possible. They can be changed with the combo box (count mode):

- *Exon/Intron*: the first position of every exon/intron is 1.
- *cDNA:* refers to the absolute position of the base in the gene, excluding introns.

- AA: every amino acid is counted.
- File: first base of the file is 1.
- Gene abs: absolute position in the gene.
- Genome Position: genomic position.

Above the electropherogram the <u>location overview</u> is shown, giving an overview of the selected location in the dialogue part *Position/Resultfiles*.



Introns are marked yellow whereas exons are marked blue. Entries of the mutation database (see chapter 14 *Mutation [master file]*) are indicated as bars below the overview (here in grey, color is user defined). *Web Refs* and Local Infos are indicated as bars above the location overview (here in light blue and blue respectively). Web Refs are known variations described in the downloaded gene file (in our case *COL1A1* from **Ensembl**). Local Infos is information deposited at a base position by the user (color is user defined). You can jump to a location within the electropherogram by clicking on the overview.

The <u>peak height ratio diagram</u> is shown graphically for each position below the location overview. The bars show positions with a high background.

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:



typical heterozygous position (bars present for forward and reverse bases)

Calculation of the bars:

peak height ratio = peak area highest not reference bases/(peak area highest not reference base + peak area reference base)

e.g. reference base is T, highest not reference base is G

peak height ratio= peak area G/ (peak area G + peak area T) The peak height ratio is shown as a bar in the graph for each base position.

It is possible to show the electropherogram data in a separate window. Therefore right-click one of the result file sequences and select *result file view* from the context menu:



The result file view gives a better overview in case many sequences are present for one location.



12.8 Show

This dialogue part offers different functions, which enable to jump quickly to the corresponding positions within the electropherogram.



- *reading*: the selected reading frame is shown in the *electropherogram*.
- check: there are different items that can be checked. The jumper aside serves to move forward or backward to the corresponding positions within the electropherogram.
- *edited:* jump to edited base positions (auto edit function or user).
- het. pos.: jump to heterozygous positions.
- mism.: jump to mismatches.

12.9 Functions

This dialogue part comprises different functions to edit the orders.

Functions	
Previous	Next
T.V. M.V.	print
Extras ->	Protocol

- [Previous] and [Next]: jump to the previous or next file in the Worklist or Lower table of Joining respectively.
- [T.V.] and [M.V.]: to technically and medically validate the order.
- [print...]: to print a report.
- *Extras ->]*: undo-, redo-function, enter a comment or change the order state.
- *Protocol*: to open the protocol which shows all recent modifications of the order.

13 Analysis of the demo data

13.1 COL1A1

Now look at the first order with gene COL1A1 (DNA No 12). Look at the result files in the dialogue part Positions/Resultfiles/column Var.:

#	Loc.	Dir.	Var.	w.	Con.	
	E1	fwd		1	1	
	E1	rev			1	
	E2	fwd			A/	
	E2	rev			1	H
	E3	fwd			1	
	E3	rev			1	
	E4	fwd			1	
	E4	rev			1	
	E5	fwd			1	
	E5	rev			1	
	E6	fwd			1	
	E6	rev			1	
	E7	fwd			1	
	E7	rev			A/	
	E8	fwd			A/	
	E8	rev			i	
	E9	fwd			1	
	E9	rev			1	
	E10	fwd			1	
	E10	rev			1	
	E11	fwd			1	
	E11	rev			1	
	E12	fwd			1	
	E12	rev			1	
	E13	fwd	C		1	
	E13	rev	С		1	
	E14	fwd			1	
	E14				i.	1

There are no variations/mutations found in **exons 1** to **12**, **14** to **21**, **23** to **40**, **42**, **44** to **48** and **50** to **51** (no entry in the column *Var*.). These sequences do not need to be checked.

The result files of exons 13, 22, 43, 49 and 52 show at least one mismatch (entry in the column *Var.*).

There are **eight** variations/mutations detected which are listed in the *Variation/Mutation Table* (tab *all*). These positions need to be checked.

all 8	dis	stinct 4	other 4	filter 0												Ca	ll Genes 👎 se	ected Gene
P	m	In	Gene	Transcript	Location	Pos.	T	Nuc Change	AA Change	S	Hint	web Ref.	c. HGVS	p. HGVS	mut Entry	mut Effect	TValidation	MValidati
Ø	1	1	COL1	ENST0	E13	-14	C	t->g (het)				rs176394	c.850-14		0/0 of 13	intronic v		
0	1	2	COL1	ENST0	E22	38 (1490)	С	G -> A (het)	G -> D (497)				c.1490G>A	p.Gly497	0/0 of 15	patho. m		
0		3	COL1	ENST0	E43	34 (3070)	С	G -> C (het)	D -> H (1024)				c.3070G>C	p.Asp102				
0		4	COL1	ENST0	E43	37 (3073)	С	G -> T (het)	G -> C (1025)				c.3073G>T	p.Gly102				
	1	5	COL1	ENST0	E43	+19	С	a -> c (het)					c.3090+					
		6	COL1	ENST0	E49	19 (3541) /	I	C (homo)	[STOP] AA		R		c.3540_3	p.Gly118				
		7	COL1	ENST0	E49	58 (3580)	С	G -> C (het)	D -> H (1194)				c.3580G>C	p.Asp119				
	1	8	COL1	ENST0	E52	-12	С	g -> a (het)				rs224949	c.4240-1		0/0 of 15	intronic v		
<																		>

Four mutations are listed on tab *distinct*, for those position both sequences (forward and reverse sense) do not show the wildtype.

NST0 E13	-14	с						and the second second second		more carro y	THUR BIT SET	1 Tunoraon	, , , unuu
NSTO E22			t -> g (het)				rs176394	c.850-14		0/0 of 13	intronic v		
LEL	38 (1490)	С	G -> A (het)	G -> D (497)				c.1490G>A	p.Gly497	0/0 of 15	patho.m		
NST0 E43	+19	С	a -> c (het)					c.3090+					
NST0 E52	-12	С	g -> a (het)				rs224949	c.4240-1		0/0 of 15	intronic v		
	NSTO E43 NSTO E52	NSTO E43 +19 NSTO E52 -12	NSTO E43 +19 C NSTO E52 -12 C	NSTO E43 +19 C a -> c (het) NSTO E52 -12 C g -> a (het)	NSTO E43 +19 C a -> c (het) NSTO E52 -12 C g -> a (het)	NSTO E43 +19 C a -> c (het) NSTO E52 -12 C g -> a (het)	NSTO E43 +19 C a -> c (het) NSTO E52 -12 C g -> a (het)	NST0 E43 +19 C a -> c (het) NST0 E52 -12 C g -> a (het) rs224949	NSTO E43 +19 C a -> c (het) c.3090 + NSTO E52 -12 C g -> a (het) rs224949 c.4240-1	NST0 E43 +19 C a -> c (het) c.3090+ NST0 E52 -12 C g -> a (het) rs224949 c.4240-1	NSTO E43 +19 C a -> c (het) c.3090+ NSTO E52 -12 C g -> a (het) rs224949 c.4240-1 0/0 of 15	NSTO E43 +19 C a -> c (het) c.3090+ NSTO E52 -12 C g -> a (het) rs224949 c.4240-1 0/0 of 15 intronic v	VSTO E43 +19 C a -> c (het) c.3090+ NSTO E52 -12 C g -> a (het) rs224949 c.4240-1 0/0 of 15 intronic v

The other **four** mutations are listed on tab *other*. For these mutations only one sequence displays a mutation whereas the other sequence shows the reference sequence. Mutations listed on tab *other* are therefore not as likely to be real mutations as distinct mutations.

D. I.m.	Tre	Come	Transmirt	1 and the se	Dee	1 T	No. Change	AA Change	10	Linet	und Daf	- UCUC	- HOVE	and Calma	-	Thelidation	Malidati
P 10	of the	Gene	Transcript	LOCAUON	POS.	1	Nuc change	AA Change	5	Film	web ker.	C. HGVS	p. novs	muteniny	I muteriett	Ivalidation	Mvalluau
	1	COL1	ENST0	E43	34 (3070)	С	G -> C (het)	D -> H (1024)				c.3070G>C	p.Asp102				
	2	COL1	ENST0	E43	37 (3073)	С	G -> T (het)	G -> C (1025)				c.3073G>T	p.Gly102				
	3	COL1	ENST0	E49	19 (3541) /	I	C (homo)	[STOP] AA		R		c.3540_3	p.Gly118				
	4	COL1	ENST0	E49	58 (3580)	С	G -> C (het)	D -> H (1194)				c.3580G>C	p.Asp119				

13.1.1 Mutation in E13 position -14

Please click on the first entry (**exon 13**) of the Variation/Mutation Table (tab all). The electropherogram view changes to this position.

There is a k (g+t) in the forward and reverse sequence at base position -14. The EMBL reference sequence (green line) shows a t.



Now look at the statistics. Select the tab Statistic on the upper right part of the screen.



For each base position of a result file this tab shows the corresponding statistic values. The statistic values are calculated over all existing result files of the same chemistry which are already archived (*Archiving* and *Statistics* is explained in chapter 14). The number below the bases shows the number of archived orders where the base was called <u>homozygous</u>. Those are used to calculate the statistic peak area (SPA, shown in <u>blue</u>), the standard deviation (SD) and the relative standard deviation (RSD). The statistic peak area (SPA) is a measure for the average peak area. The result file peak area (RPA) of the result file being analysed at the moment is shown in green. The relative standard deviation (RSD) shows to which extent the result file peak area exceeds the statistic peak area.

For heterozygous positions, peak areas are smaller than for homozygous positions. That is, heterozygous positions can be detected not only by the existence of a second peak, but also by comparison of the peak area with the statistic peak area.

The statistic shows, that the result file peak area for base t at base position **-14** in the forward sequence is just half of the statistic peak area, calculated over **18** archived result files. Furthermore the result file peak area for base g is as high as the result file peak area for base t. This is typical for heterozygous postions.

The reverse sequence shows the same.

The peak height ratio diagram (below the location overview) also shows a typical heterozygous position: there are background bars present for forward and reverse bases.

	Intron 12	
	1	
-		

In the electropherogram the selected mutation has a grey background. This means it is a known mutations that is present in the internal *Mutation database* (see chapter 14). It therefore also has an entry in the *Variations/Mutations Table* in the column *mut Entry*.

P m.	. In	Location	Pos.	Type	Nuc Change	AA Change	S	Hint	web Ref.	HGVS nome	mut Entry	mut Effect	TValidation	MValidation
0	1	E13	-14	С	t -> g (het)				rs17639446	c.850-14T>G	0/4 of 13	intronic var		
0 🗹	2	E22	38 (1490)	С	G -> A (het)	G -> D (497)				c.1490G>A	0/1 of 15	patho. mut		
0 🗹	3	E43	+19	С	a -> c (het)					c.3090+19				
0	4	E52	-12	С	g -> a (het)				rs2249492	c.4240-12G	5/7 of 15	intronic var		

Moreover there is an entry in the column *Web Ref* of the *Variations/Mutations Table*. This means that the mutation is also described in the downloaded gene file (from **EMBL** or **Genbank**, in our case *COL1A1* from **EMBL**).

Move the mouse to a base within the **EMBL** reference sequence (green line). A tool tip window shows information about this position. This information is only shown if there is a web Ref and/or a mut Ref at this position!



If the computer is connected to the internet, the web reference can be looked at by right-clicking on base **a** within the **EMBL** reference sequence or on the corresponding entry in the *Variation/Mutation Table* and select the context menu item *web Ref.* > *show*....

Nuc Change	AA Change	State	Hint	we	1.0		t 1	0	t
web Ref	> sl	now rs17639446		rs:		local Info	>		-
c mut Ref						web Ref	>	show rs17	639446
c macrice					5	mut Ref	>	0	1
show missing ba	ises				_	add mut to DB			- 10-
c ignore (toggle)	.6 (E4	19/124)	R		-	сору	>	0	
move to other									
setting	>			rsa					
export	>								
show all transcri	pt info	Exon 13							
filter	>								
	× –								

If the context menu item *mut Ref.* > *show* ... is selected, the dialogue *Mutation* is opened. This dialogue gives information about mutations that were detected in earlier orders and are stored in the *Mutation database* (see chapter 14).

Mutation		×
Gene: COL1A1 Transcript: ENST00000	225964	Type: C Location: E13 Position: -14
Nuc Change: T -> G	AA Change:	g./cDNA: 0/0 of 0 Reload
Nuc Name: c.850-14T>G Set	AA Name:	Set Frequency: 0/4 of 13 Reset
Classification: undefined Color	Info Intern:	Show exist Info Extern: Show exist
Disease No: Show	Mut Effect:	intronic variation Color p. mut Effect: undefined 💌 %
Web Ref:	Ethnicity:	▼ p. mut Causality: undefined ▼ %
Print: Show:	Color:	▼ show Default: Changed: 02/15/2007
Genome Position	Weblinks:	Index Link Comment
		1 http://www.ncbi.nim.nih.gov/ NCBI entry Add
< >>		Remove
_ Orders		Organ / Phenotype
Index Order No. Zygosity genomic/cD.	Module	DNA No Etr Organ Phenotype
1 70460015		
2 70460011		
3 70460009		Add
70700003		
		Remove
<		>
Delete		OK Cancel

13.1.2 Mutation in E22 position 38

The next entry in the Variation/Mutation Table – exon 22 – shows a R (A+G) at base position 38 in the forward and reverse sequence. The EMBL reference sequence displays a G.



The statistic shows that the result file peak area for base **G** in the forward sequence is just half of the statistic peak area, calculated over **27** archived result files. Furthermore the result file peak area for base **A** is as high as the result file base area for base **G**. The reverse sequence shows the same. This is a true heterozygous position (**R**).

In the electropherogram the mutation has a grey background. This shows, that there is a mutation reference at that position. This can also be seen in the column *mut Ref.* of the *Variation/Mutation Table*. For detailed descriptions see chapter 13.1.1.

13.1.3 Mutation in E43 position 34

The next entry shows an **S** (C+G) at base position **34** in the forward sequence. The **EMBL** reference sequence and the reverse sequence display a **G** at this position. This is a not distinct mutation, since it occurs in one sequencing direction only.



The statistic of the forward sequence does not look like the statistics of the heterozyous positions we have seen before (chapter 13.1.1 and 13.1.2). Here, the result file peak area for base **G** is close to the statistic peak area, calculated over **20** archived result files. Moreover the peak area for base **C** is much smaller than the peak area for base **G** and also close to the statistic peak area. This means that the archived result files - where a homozygous **G** was called - have a similar background for base **C** then our present result file.

That is, we have no heterozygous base position (S) in the forward sequence and have to change the S into a G.

13.1.3.1 Edit Bases

There are several options to edit the base:

- Left-click on the S in the forward sequence. Left-clicking a base sets the base of the reference sequence. A G is set automatically.
- Right-click the S and open the context menu shown below:.



- Here you can use the items *edit* (1) or *edit bases...*(2) to do your modifications.
 - (1) *edit*: Select the base that should be set, in our case *edit* > G.
 - (2) edit bases... : Opens the dialogue Edit Bases.



Manual for demo data SEQUENCE Pilot – module SeqPatient

To set a **G** at **base position 34** right-click on it within the *edited line* and select *G* from the context menu. Then press *[OK]*.

The Mutation in E43 at position 34 disappears in the Variation/Mutation Table.

13.1.3.2 Undo Function

There is a function available to undo manual edits in the electropherogram. Therefore press the button *[Extras->]* in the dialogue part Functions and select *undo....*

comm	ient	
undo.		Next
redo		print
order	state 🕨	
1/1 >	Extras ->	Protoco

The following dialogue opens:

Action		
hange base E43, 279 S>G		

The latest edit in the electropherogram is listed on top. If you select it and press *[OK]* the edit will be removed. In our case the listed base change is all right. Therefore we do not use the undo-function.

13.1.4 Mutation in E43 position 37

The next entry in the *Variation/Mutation Table* shows a K (G+T) in exon 43 at base position 37 in the forward sequence.



The **EMBL** reference sequence and the reverse sequence have a **G**. As the mutation we just edited (chapter 13.1.3), this is also a not distinct mutation.

The statistic shows that we have the same situation here, as described in chapter 13.1.3. This is not a real heterzogous position.

So the **K** in the forward sequence needs to be replaced by **G**. Therfore you can left-click or right-click on the **K** as done before (chapter 13.1.3.1):

- Left-clicking sets a G automatically (base of the reference sequence is set).
- You can right-click on the base and select edit > G.
- You can right-click on the base and select *edit bases…* to open the dialogue *Edit Bases*: Right-click on the K within the *edited line* and select G from the context menu. Then press [OK].

13.1.5 Mutation in E43 position +19

The next entry in the Variation/Mutation Table shows a **m** (**c**+**a**) in **exon 43** at base position **+19** in the forward sequence. The **EMBL** reference sequence has an **a** at this position. This is a distinct mutation because no reverse sequence is present.



If you have a look at the statistic you can see, that we have a similar situation as described in the chapters 13.1.3 and 13.1.4. In **11** archived result files, where a homozygous **A** was called, there is always a certain background for base **C** present. Moreover the peak area of base **C** is much smaller than the peak area of base **A**.

So probably base position +19 is not heterozygous. We remove the **m** and set an **a** as done before (chapter 13.1.3.1). Therfore you can left-click or right-click on the **m**:

- Left-clicking sets an a automatically (base of the reference sequence is set).
- You can right-click on the base and select *edit* > A.
- You can right-click on the base and select *edit bases…* to open the dialogue *Edit Bases*: Right-click on the **M** within the *edited line* and select the item A from the context menu. Then press [OK].

13.1.6 Mutation in E49 position 19

The next entry in the Variation/Mutation Table is exon 49. At base position 19 there is an insertion of a **C** in the forward sequence. The **EMBL** reference sequence and the reverse sequence show no base. This is again a not distinct mutation.



The statistic also shows that there are no archived result files, where a C or any other base at this base position has been detected before. That is, the insertion (C) in the forward sequence isn't one.

To remove the insertion, you can left-click or right-click on the **C** as done before (chapter 13.1.3.1):

- Left-clicking removes the C automatically (base of the reference sequence is set).
- You can right-click on the base and select *edit* > *None*.
- You can right-click on the base and select *edit bases…* to open the dialogue *Edit Bases*: To remove the C at **base position 19** right-click on it within the *edited line* and select the item *None* from the context menu. After the C is removed, press [OK].

13.1.7 Mutation in E49 position 58

The next entry in the Variation/Mutation Table is exon 49 position 58. There is a S(C+G) in the forward sequence The EMBL reference sequence and the reverse sequence show a G. This is again a not distinct mutation.



The statistic shows that the result file peak area for base **G** in the forward sequence is close to the statistic peak area, calculated over **17** archived result files, that showed a **G**. The archived result files also show a certain background for base **C**. The reverse sequence doesn't show a heterozygous base position.

That is, the heterozygous base position (**S**) in the forward sequence isn't one.

To remove the insertion, you can left-click or right-click on the **S** as done before (chapter 13.1.3.1):

- Left-clicking sets a **G** automatically (base of the reference sequence is set).
- You can right-click on the base and select *edit* > *G*.
- You can right-click on the base and select *edit bases…* to open the dialogue *Edit Bases*: To set the G right-click on the S within the *edited line* and select the item *None* from the context menu. After the G is set, press [OK].

13.1.8 Mutation in E52 position -12

The last entry in the Variation/Mutation Table – exon 52 – shows a r (a+g) at base position -12 in the forward and the reverse sequence. The EMBL reference sequence shows a g.

The statistic shows, that there are **10** archived result files where **a** was detected homozygously (upper statistic). Compared to the statistic values the result file peak area for base **a** is only half of the statistic peak area. Moreover there is a peak for **g**.



A homozygous **g** was detected in **6** archived result files (lower statistic). To see the statistic click on the *G* in the tab *Statistic*. Here, the result file peak area for base **g** in the forward sequence is just half of the statistic peak area. Furthermore the result file peak area for base **a** is as high as the result file base area for base **g**. The reverse sequence shows the same heterozygous base position.

This is a true heterozygous base position (r).

There is also a web Ref and a mut Ref entered which shows that it is a known mutation that was detected before. For detailed descriptions see 14.1.1.

13.1.9 Result of the COL1A1 analysis

The analysis is finished and there are **three** remaining entries in the *Variation/Mutation Table* (tab *all)*. All of them listed on tab *distinct* initially. The mutations listed on tab *other* were no real mutations.

all 3	3 distinct 3 other 0 filter 0																	
P	m	In	Gene	Transcript	Location	Pos.	Туре	Nuc Change	AA Change	State Hin	web Ref.	c. HGVS	p. HGVS	mut Entry	mut Effect	TValidation	MValidation	W
٥	V	1	COL1	ENST0	E13	-14	С	T -> G (het)			rs17639446 (Ge	c.850-14T>G		0/4 of 13	intronic variation			d
٥	V	2	COL1	ENST0	E22	38 (1490)	С	G -> A (het)	G -> D (497)			c.1490G>A	p.Gly497Asp	0/1 of 15	patho. mutation			d
0	V	3	COL1	ENST0	E52	-12	С	G -> A (het)			rs2249492 (Gen	c.4240-12G>A		5/7 of 15	intronic variation			d

That's the result of your COL1A1 analysis!

13.1.10 Validation

It is possible to technically and medically validate all result files of the order. The technical and medical validation shows that all validated variations/mutations have been looked at and confirmed.

For the validation of single exons mark the corresponding exon within the dialogue part *Positions/Resultfiles* and use the buttons [*T.V.*] and [*M.V.*] respectively below this dialogue part. To validate all exons at one time, use the buttons [*T.V.*] and [*M.V.*] respectively shown in the dialogue part *Functions*.

	111			P.	Extras -> Protoco
2	E52	rev	С	/SV	
5	E52	fwd	C	/SV	
	E51	rev	-	/S	T.V. M.V. print
	E51	fwd		/S	
	E50	rev		N	
	E50	fwd		N	Previous Next
	E49	rev		N	
	E49	fwd		M/V	Functions
	E48	rev		1	
	E48	fwd		1	
	E47	rev		1	

After the technical validation, the field *Condition* in the dialogues *Gene* and *Position/Resultfiles* is highlighted **green** and shows the entry *V*. After the medical validation the field *Condition* is highlighted **yellow** and shows the entry *S*. The user who did the validation and the date are shown in the columns *TValidation* and *MValidation*. Moreover after medical validation all edits are locked, no further modifications can be made unless the medical validation is removed by clicking [*M.V.*] again.

13.1.11 Printing of a report

After finishing the analyses the report can be printed. Which mutations/variations are printed and if the electropherogram data is printed depends on the settings in the *Variation/Mutation Table* in the column *Print* (see chapter 12.6).

To print a report click on [print...] in the dialogue part Functions.

gene COL1A1	result compl.	printed yes
< <p>Variation Image: P Variation Image: P Variat</p>	options Selection only select all Select all not printed report SeqPatient-Family Report SeqPatient-Report SeqPatient-Report short	

Here you can choose what to print:

- Check genes to print. The options selection only, select all and select all not printed helps checking the items.
- In the field Variation you can choose which tab of the Variation/Mutation Table to print (all, distinct or other). By default the tab all is selected.
- In the field *report* two report forms can be selected by default. The report Seq Variation short contains all information of the report Seq Variation except the resultfile information and the gene map.

Open a print preview using [Preview] and use [Print] to print your results.



With a little usage you will see, that complete analysis of *COL1A1* gene can be done in less than 5 minutes.

13.1.12 Basecaller and Auto Edit

Even though there are many exons in our *COL1A1* sample, only **eight** entries were found in our *Variation/Mutation Table*. There are two reasons for this:

(1) SEQUENCE Pilot contains its own basecaller. This means it does not take the bases set by the sequencers Basecaller. During the import of a result file it calculates the bases by curve values. The rules for calculation can be defined in the thresholds (operation *Thresholds [master file]*). There is one exception: heterozygous base positions set by the sequencers base caller are never replaced.

In the dialogue *Edit Bases* the bases calculated by the internal basecaller are shown in the <u>calculated line</u>, whereas the bases set by the sequencers basecaller are shown in the <u>original line</u>.

To see the following example select **E3** in the *Position/Resultfile table,* select **base position 4** in the electropherogram. Then right click on the **C** in the forward sequence and open *Edit Bases.*



To get an impression how many original base positions within a result file have been changed by the internal Basecaller of **SEQUENCE Pilot**, use the jumper *change* within this dialogue (*Edit Bases*). (Alternatively go to the dialogue part *Show* and select the item *change* of the jumper *check*). In the result file above **20** bases were changed.

- (2) SEQUENCE Pilot contains an own Auto Edit function. This function enables the program to
 - delete insertions when there are none.
 - replace "N" base positions by a base.
 - extend compressions (W ==> A + T).
 - remove heterozygous base positions, based on the already existing statistic data (operation Statistic [master file], see chapter 14). The statistic data will only be factored into the auto edit function if at least five result files of the same chemistry exist.

All base positions changed by the Auto Edit function are shown in the *edited line* within the dialogue *Edit Bases*.

Here are some examples:

<u>extend compressions</u> (exon 8 fwd base position -4): A **N** called by the sequencer was replaced by an **M** by the internal base caller. The **M** was extended to **C+A** by the Auto Edit.



To get an impression how many base positions within a result file have been edited by the Auto Edit function of **SEQUENCE Pilot**, use the jumper *edited* within the dialogue *Edit Bases*. In the first example below **2** bases were edited in the second example **3** bases were edited. (Alternatively the jumper *edited* of the dialogue part *Show* shows all edited positions).

<u>remove heterozygous base positions</u>, based on the already existing statistic data (exon 43 fwd-base position -2



Manual for demo data SEQUENCE Pilot – module SeqPatient

By default the auto edit is activated. If desired it can be deactivated.

13.1.13 How to analyse the order again

You have two possibilities to remove modifications, if you want to repeat your analysis. First remove the validation by pressing [*T*.*V*.] and [*M*.*V*.] in the dialogue part *Functions*.

Recalculate-Function: Recalculates already loaded files. Additionally to removal of manual edits, thresholds can be modified and changed settings (such as an updated software version, updated gene files (reference sequence) or a modified lis.ini file) are automatically used to recalculate the result. Edits done in the Variation/Mutation table (column Print/Mut DB) and ignores to left or to right are not deleted!

To recalculate all result files for a gene right click on the gene in the *Genes* table and select *editing* > *recalculate....* (Single result files can be recalculated with the same context menu items in the dialogue part *Positions/Resultfiles*).



The following dialogue opens. Here the *Threshold* for the internal Basecaller and Auto Edit can be modified. Basecaller, Auto Edit or single Auto Edit functions can be deactivated by removing the mark. Press *[OK]* to recalculate the file.

Thresholds ×									
~	Basecaller Absolut Count:	Genome set: Diploid 💌							
	Percent Count: 40 %	Recalculate heterozygous bases							
	Slope 1: 20 %	Warning Height: 0							
	Slope 2: 10 %	Warning Phred: 0							
	Auto Edit Bases Compression Decompression N's	Statistic Mosaic Score: 6 RSD: 5 Statistic Min # statistics: 5							
	L	OK Cancel							

Delete the order and load the files again: With this option all edits and settings are deleted. To delete the order select the category *LIS* and the operation *Orderlist*. Then right-click on the first order (DNA No 12) and select the context menu item *delete order*. Now you can load the order again as described in chapter 9.

LIS Select Orders Date: Image: Im	Administration	Orderfist								
Orderlist Order Date DNA No Imaster file] Order 07/31/2018 10:02:43 12 Jump to Order Input 07/31/2018 10:26:07 159 Or/31/2018 10:26:07 160 Profiles [master file] export order show protocol show patient 07/31/2018 10:26:07 160	LIS Order Input	Select Orders Date: 31.07.2018 Date Range: 31.07.2018 State: any active	Dna <u>N</u> o.: <u>O</u> rder No.: <u>P</u> at. Name: Patient ID:							
Image: master file] Image: master fi	Procedures	Orderlist Order	Date	DNA No						
Profiles [master file] export order show protocol show patient	[master file]	Jump to Order Input	07/31/2018 10:02:43 07/31/2018 10:26:07 07/31/2018 10:26:07	12 159 160						
	Profiles [master file]	export order show protocol show patient								

13.2 Mareno

13.2.1 Sample with DNA number 159

Now go to the second order (first *Mareno* sample). To do this click on *[Next]* in the dialogue part *Functions*. The buttons *[Next]* and *[Prev]* will jump to the next or previous entry of the *Lower Table* of *Joining* or the *Worklist* respectively. Alternatively you can switch to the operation *Joining* or *Worklist*, select the second entry (DNA No **159**) and return to the operation *Sequence*.

There is only **1** mismatch at position **288** which has to be checked. But this position really is a **R** and is not changed. So there is one remaining entry in the *Variation/Mutation Table*.

P, In Gene Transcript Location Pos. Type Nuc Change AA Change State Hint web Ref. c. HGVS p. HGVS mut Entry mut Effect TValdation MValdation W	all 1 distinct 1	other 0	filter 0															C all Genes ⊙ selected Gene
	P m In	Gene	Transcript	Location	Pos.	Туре	Nuc Change	AA Change	State H	int web Ref.	c. HGVS	p. HGVS	mut Entry	mut Effect	TValidation	MValidation	W	
X Y 1 Materio Enstru E10 266 (2000) C A -> G (Tet) M -> V (094) C.2000A>G P.Meto94Val 0	☑ ☑ 1	Mareno	ENST0	E10	288 (2080)	С	A -> G (het)	M -> V (694)			c.2080A>G	p.Met694Val					d	

That's the result of this Mareno analysis !

Technically and medically validate all result files of this order and print the report like described in chapters 13.1.8 and 13.1.9.

13.2.2 Sample with DNA number 160 – Statistic Warnings

Now go to the last order (second *Mareno* sample). To do this, click on *[Next]* in the dialogue part *Functions*.

There is only **1** mismatch at position **490** which has to be checked. But this position is really an **A** and is not changed. So there is one remaining entry in the *Variation/Mutation Table*.

all 1 district 1 other 0 fitter 0																			
P	P m	. In	Gene	Transcript	Location	Pos.	Туре	Nuc Change	AA Change	State	Hint	web Ref.	c. HGVS	p. HGVS	mut Entry	mut Effect	TValidation	MValidation	W
K	3 🗹	1	Mareno	ENST0	E10	490 (2282)	С	G -> A (homo)	R -> H (761)				c.2282G>A	p.Arg761His					d

No, we are not finished. There is a statistic warning for this file !!!

Below the dialogue part *Positions/Resultfiles* there are the two warnings *bases* and *statistic*.



(1) *bases*: There is no warning if the dot is highlighted **grey** and a warning if the dot is highlighted **red**. The warning shows that there are missing bases. Missing bases result if there are amp. modules and/or seq. primers defined and some of the expected bases are missing in the result file.

(2) *statistic*: There is no warning if the dot is highlighted green and a warning if the dot is highlighted or ange or red.

What do statistic warnings mean ?

SEQUENCE Pilot calculates a <u>dissimilarity score</u> for each base position of the result files. The score value depends on the deviation of the peak areas and background areas from the average statistic values and is a measure of the difference between the peak areas and the background to the statistic average. Normal fluctuations of the peak areas and backgrounds result in *dissimilarity scores* between 0 and 10.

By default we have defined two limits **20** and **40** for the *dissimilarity score*. If the *dissimilarity score* passes the first limit (**20**), the program gives a warning by showing an *s* within the corresponding field of the column *Warning* of the *Positions/Resultfiles table* and colours it orange. The corresponding position in the electropherogram is <u>not</u> highlighted. If the dissimilarity score passes the second limit (**40**), the program gives a warning by showing an *S* within the corresponding field of the column *Warning* and colours it **red**. The corresponding position in the electropherogram is <u>not</u> highlighted in the electropherogram is highlighted **green**.



What can be seen in our demo data?

There is a statistic warning for the *Mareno* sample as it is shown by the **red** highlighted dot *statistic* and the entry **S** in the dialogue part *Positions/Resultfiles,* column *Warning*. So there is at least one position with a *dissimilarity score* above **40**.

To move to the corresponding positions within the electropherogram use the dialogue part *Show*. Select *statistic* in the field *check* and use the jumper behind the field to move to the position in the electropherogram. The jumper statistic goes through the base positions sorted by *dissimilarity score* (highest score first) as shown above.

The base at position **248** in **exon 10** has a dissimilarity score of **54**. The value is shown in the dialogue part *Show* below the jumper *statistic*. The position in the electropherogram is therefore highlighted with a **green** background, which indicates a score above the threshold of **40**.

This position does not look like a heterozygous position. There are well defined peaks of **G** in the forward and reverse sequence. Moreover there is a background of **C** in both sequences. **But in the statistics there is a significant deviation from the normal peak areas which leads to the statistic warning.**

Moreover in the background bar diagram (below the location overview) you can see that there is a <u>forward and reverse bar</u> present, which usually is the case for heterozygous positions.



In the bar diagram you can also see that there are multiple base positions showing background. You can also use the jumper *statistic* in the dialogue part *Show* to find base positions with a dissimilarity score above **20**. This is very unusual.

What happened to this Mareno sample ?

There was an allelic drop-out, this means a mismatch at a PCR-Primer position for one allele. This allele was not amplified with the PCR as expected and is only present with a small copy number. Therefore several peaks do not match to the normal relative peak areas and we get high *dissimilarity scores*.

At position **248** the **G** peak is too small and we have a **C** background. Knowing the allelic drop-out, we see that the **C** is no background but a real peak and the small peak area is a result of the small copy number.

SEQUENCE Pilot can not define positions like **248** to be heterozygous, but it will give a warning if something differs from the statistic average.

Knowing this, position **490** can also be defined to be heterozygous, which was previously assumed to be homozygous.

So the warning means – do not trust the results and think about what could have happened.

13.3 Discussion of the analysis

All variations/mutations which appeared in the *Variations/Mutations Table* were checked. **Three** mutations/variation in the *COL1A1* sample and **one** mutation in the first *Mareno* sample (DNA number 159) were found. All of them were distinct mutations. The second *Mareno* sample (DNA number 160) could not really be analysed because there was a problem with the PCR.

There were many exons in our *COL1A1* sample that were not checked, because there was no mismatch found. As described for the *Mareno* sample with DNA number 160, all positions are checked against the average statistic peak areas and there was no warning for any *COL1A1* or *Mareno* (DNA number 159) result file.

Check the *dissimilarity scores* in the *Mareno* sample (DNA number 159) or the *COL1A1* sample to see if there are any statistic aberrations. Therefore press [*Prev*] in the dialogue part *Functions* and to switch to the *Mareno* sample (DNA number 159) or the *COL1A1* sample. Select any result file in the *Positions/Resultfiles table.* Then check the dissimilarity scores with the jumper *statistic*. In most cases there will be "no entry" - this means there is no peak with a *dissimilarity score* above **1**. For some exons there will be *dissimilarity scores* in the range of 1 to 3. The highest *dissimilarity score* is in exon 4, position +15 of the *COL1A1* sample with a value of **9.05**.

Another test for the dissimilarity scores (remove the medical validation before you do your edits!):

Move to exon 10 of the *Mareno* sample (DNA number 159), edit the heterozygous position 288 in the forward, reverse or both sequences to '**A**' (this simulates a not recognized heterozygous position) and check the *dissimilarity scores*. Try the same with the found heterozygous base position -12 in exon 52 of the *COL1A1* sample. Set a '**g**' here. Statistic warnings appear with dissimilarity scores of **325** and **170** respectively:



Mareno DNA No 159, E10, postion 288

Col1A1, E52, postion -12

Now it can be seen why result files do not have to be checked when no variations are found and no statistic warnings exist. **SEQUENCE Pilot** checked the files for us already by calculating *dissimilarity scores*!

14 Archiving, Statistics and Mutation database

How to get there: category SeqPatient operation Archiving, operation Statistic [master file] and operation Mutation [master file] respectively

Medically validated orders can be archived with the operation *Archiving*. The electropherogram data is moved out of the database during archiving. All other information is still stored in the database. The advantage to this is that the database is much smaller and faster.

During archiving the statistics is created automatically which can help to detect mutations/variations. Moreover detected mutations/variations are added to the internal *Mutation database*.

The operation *Archiving* is divided into two parts. With *Select Orders* it is possible to search for orders to archive using different criteria. Only processed orders with the state *compl. MV* (complete medical validation) are automatically listed. Then select the orders to be archived in the list below and press [archive].

The present statistics for all genes can be viewed in the operation *Statistic [master file]*. One statistic is created out of all archived orders of the same chemistry. Same chemistry means same gene, exon, sequencer, sequencing direction, amp. module and seq. primer. Here also the reference sequences are listed.

The *Mutation database* can be viewed in the operation *Mutation [master file]*. Here all detected mutations are stored. For stored mutations, information will also be obtained in further analyses. There will be an entry in the column *mut ref* in the *Variation/Mutation Table* (chapter 12.6) if the mutation was detected before and added to the mutation database. If detected mutations are added to the *Mutation database* during archiving will depend on the settings in the *Mutations/Variations Table* in the column *mut db* (chapter 12.6).

15 Shortcuts

There are shortcuts available to make the program more convenient. Some of the short cuts can be used in the operation *Sequence* others in *Edit Bases* and some in both.

Short cut	Description	Operation Sequence	Edit bases
Ctrl + arrow key right	Cursor jumps one base to the right.	Х	
Ctrl + arrow key left	Cursor jumps one base to the left.	Х	
Ctrl + arrow key right	Cursor jumps one curve point to the right.		Х
Ctrl + arrow key left	Cursor jumps within one curve point to the left.		Х
Ctrl + Shift + arrow key right	Cursor jumps one base to the right.		Х
Ctrl + Shift + arrow key left	Cursor jumps one base to the left.		Х
Ctrl + arrow key up	Cursor jumps one sequence up.	Х	Х
Ctrl + arrow key down	Cursor jumps one sequence down.	Х	Х
E	Cursor jumps to the next edited position.	Х	Х
Shift + E	Cursor jumps to the previous edited position.	Х	Х
М	Cursor jumps to the next mismatching position.	Х	Х
Shift + M	Cursor jumps to the previous mismatching position.	Х	Х
Н	Cursor jumps to the next heterozygous position.	х	Х
Shift + H	Cursor jumps to the previous heterozygous position.	х	Х
С	Depending on the selected entry in the jumper <i>check</i> (dialogue part <i>Show</i>) the cursor jumps to the next position.	Х	
Shift + C	Depending on the selected entry in the jumper <i>check</i> the cursor jumps to the previous position.	Х	
Ctrl + Shift + C	Changes the entry in the jumper check (dialogue part Show).	Х	
Ctrl + A, C, G, T, M, R, W, S, Y or K	Replaces the selected base in the electropherogram by A, C, G, T, M, R, W, S, Y or K.	Х	Х
Ctrl + Space	Replaces the selected base in the electropherogram to the one in the reference sequence.	Х	Х
Ctrl + Del	Deletes the selected base in the electropherogram.	Х	Х
G	Opens the dialogue Gene map (dialogue part Gene).	Х	
Ν	Opens the next order.	Х	
Р	Opens the previous order.	Х	
Τ	Technical validation of the location.	Х	
V	Medical validation of the whole order.	Х	
Ctrl. + Shift + L	Sequence is ignored from the selected position to the beginning.	Х	Х
Ctrl. + Shift + R	Sequence is ignored from the selected position to the end.	Х	Х
Pos1	Cursor jumps to the first base.	Х	Х
End	Cursor jumps to the last base.	Х	Х
Page up	Cursor jumps one page backward.	Х	Х
Page down	Cursor jumps one page forward.	Х	

For a detailed description have a look at the **Manual SEQUENCE Pilot** – module **SeqPatient**.