

# Instructions for training RH-MLPA-Analysis to your own data

## Contents

Instructions for training RH-MLPA-Analysis to your own data .....	1
1 Training the analysis program with your own data.....	2
2 File formats .....	6
2.1 The format of a “MLPA-training-list file”.....	6

We must admit that training is not made userfriendly, it is only the use of RH-MLPA-Analysis after training that is easy to perform.

Training is based on the same kind of files that are used for analysis, i.e. GeneMapper or Peak Scanner data have to be converted first. Se details for this in “instructions-demo.pdf”.

You can train one of the supplied MLPA probe sets if the probes of your actual lot. number fit the probes of the supplied RH-MLPA-Analysis probeset. I.e. if you have a later lot. number then check the documentation that came with the kit to check whether or not some probes have been replaced by MRC-Holland.

Sometimes we have added e.g. more than one MLPA probeset for a kit to RH-MLPA-Analysis, even when the probes have not changed, but where we have found some significant differences in the produced probe areas for one or more probes. In this way we can use the program to analyse new samples and then still have the possibility to reanalyse old samples of the previous lot. number.

This can be seen in e.g. the two versions of P095 and the two versions of P069:

a) P095-vs02 Aneuploidy MLPA kit (lot 0505)

b) P095-vs03 Aneuploidy MLPA kit (lot 1106)

The Y-probe peak at 160 nt is about 50% higher than for our old lot. 0505

a) P069-vs01 Human Telomer MLPA Kit (lot 0205, 0505)

b) P069-vs02 Human Telomer MLPA Kit (lot 0606)

The 1q peak is lower and the 10p peak is higher than for our old lot. 0505

Alternatively you can make a new copy of the RH-MLPA-Analysis program before you train to a new lot. number, and then use the old program for analysing old samples, and the new program for analysing new samples of the new lot. number.

By the v4.87 version of the program you can type in the actual lot. number and a comment to each probeset you train.

# 1 Training the analysis program with your own data

See also the paragraphs 2.4 and 2.5 in the document (instructions-demo.pdf), because they make a good explanation of the training process for the SALSA P023-vs01 DiGeorge Syndrome/VCFS MLPA Kit, so here we just mention what might be missing in the training demo example.

Current MLPA probe set: P023-vs01  
Current MLPA-training-list file name: P023-vs01-t02.txt  
No. of samples used for training: 39  
Program version: 4.83

Path to the main folder containing MLPA-data folders: C:\RH-MLPA-v483\P023-vs01-Training\Data  
Path to the main folder containing MLPA-traininglist folders: C:\RH-MLPA-v483\P023-vs01-Training  
Folder that contains the MLPA-traininglists: Norm-lists  
New MLPA-training-list file name: P023-vs01-t02.txt  
No. of samples: [ ]

Last analysed sample (updated currently during training etc.)  
MLPA-data folder: 2005-06-16\_509 P023  
MLPA-data file: P023-N39  
Number of Y chromosomes: [ ]  
Reference data name: t-ref

Current option settings for the actual probe set  
Scale "normal" ratios to 1.0. Use this for probe sets like P095 where e.g. a trisomy 21 would otherwise during normalization scale normal ratios below ratio 1.0. (Only used during analysis, i.e. not during training.)   
Extra distance allowed to reference peaksize during classification (e.g. 2) Set it to 0 for well trained probe sets. 0.50  
Normalization method: Separate groups: PeakArea/(mean group PeakArea)

Analyse the samples listed in the MLPA-training-list file. Print Report and update the logfile statistics for each sample.  
Train the program by the samples listed in the MLPA-training-list file. Can only do full training, i.e. not add a few new cases to the previous training!  
Analyse the samples listed in the MLPA-training-list file. But do only update the Logfile statistics for each sample.

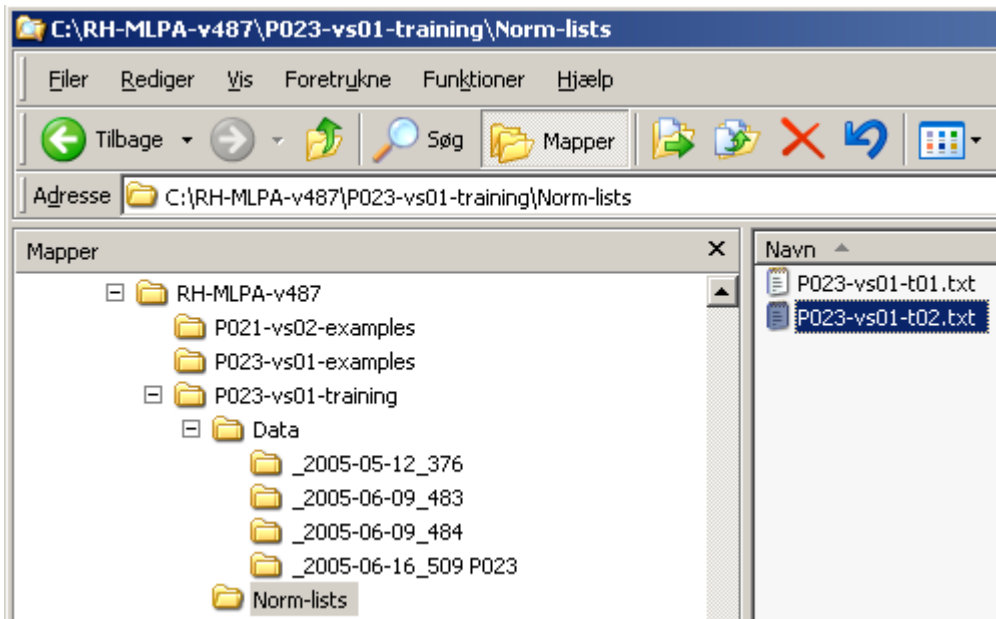
As seen in the training form above a single file controls the training. The file name has to be typed into the field “New MLPA-training-list name”. The fields above this field describe the folder path to the file and the folder path to all sample files mentioned in the “MLPA-training-list file”.

For each MLPA data file that is going to contribute to the training, the “New MLPA-training-list file” contains a line describing the name of the file. The format and contents of the supplied demo file “P023-vs01-t02.txt” is shown below and in “2.1 The format of a “MLPA-training-list file”.

The trainig form can also be used to create a summary of sample analysis results for abnormal cases or for large collections of routine data by the use of the buttons “Print Report and update ...” or “Do only update Logfile ...”, a few more details about this is mentioned in “instructions-demo.pdf”.

Befor doing training we normally make a backup copy of the current program, because you might totally spoil the reference data if the supplied cases are of too bad quality or you by mistake have added wrong files or probe sets for training.

**That's how the demo files are placed on the harddisk of the computer that runs Access 2000:**



**And that's how the training form tells where to read "P023-vs01-t02.txt" on the hard disk**

Path to the main folder containing MLPA-traininglist folders
C:\RH-MLPA-v483\P023-vs01-Training
Folder that contains the MLPA-traininglists
Norm-lists
New MLPA-training-list file name
P023-vs01-t02.txt

**And that's how the training form tells where to read the individual MLPA data files for training**

Path to the main folder containing MLPA-data folders
C:\RH-MLPA-v483\P023-vs01-Training\Data

As seen by the below lines that are cut outs of the whole MLPA-training list file ("P023-vs01-t02.txt" shown in 2.1) each filename can be prefixed by a folder name, because we prefer to let MLPA data files stay in a separate folder inside "C:\RH-MLPA\P023-vs01-training\Data" for each run:

```
_2005-05-12_376\P23-N1
_2005-05-12_376\P23-N10
.
.
_2005-06-09_483\P023-N15
_2005-06-09_483\P023-N16
.
.
_2005-06-09_484\P023-N23
_2005-06-09_484\P023-N24
.
.
```

When running the training for the first time for one of the supplied MLPA probe sets, it might be a good idea to type in 1, 2 or 3 (nt) in this field:

Extra distance allowed to reference peaksize during classification (e.g. 2)

Because the peak recognition part of the program otherwise only recognizes peaks close to the fragment sizes of the reference data (defined by the previous training). After the first training with an allowed extra distance we usually does a second training , but this time with Extra distance set to 0.0.

When running the training for the first time for data from another system than ABI 3100 for one of the supplied MLPA then it might be a good idea first to click at the button

Edit or set reference fragment sizes (lengths) manually

on the main menu of the prgram, and for each peak enter a fragment size value close to the fragment size you have for your system.

Having done that you should experiment and see how the program acts on your data when you just tries to analyse them, both to check that you have entered in some useful values so that the peaks of your system can be recognized, but also because we have found that the program after such a simple manual training has been able to analyse normal and abnormal P095 samples supplied from other laboratories in the World.

f-ManuallySetFragmentSize - formular

SALSA P023-vs01 DiGeorge Syndrome/VCFB MLPA Kit

**Edit or set reference fragment sizes (lengths) manually**

Chromosome position	Size (nt) by MRC	Size (nt) used by the program	Size (nt) generated by training	Difference between used and MRC	Difference between used and trained
cm64	64	61,29	61,29	-2,71	0,00
cm70	70	66,88	66,88	-3,12	0,00
cm76	76	72,75	72,75	-3,25	0,00
cm82	82	78,68	78,68	-3,32	0,00
2q14 synt.	94	90,35	90,35	-3,65	0,00
10p12	139	135,77	135,77	-3,23	0,00
4q34	148	144,30	144,30	-3,70	0,00
22q11	157	155,80	155,80	-1,20	0,00
4q35	166	165,67	165,67	-0,33	0,00
7p15	175	173,63	173,63	-1,37	0,00
4q34	184	181,87	181,87	-2,13	0,00
22q11	193	192,36	192,36	-0,64	0,00
4q22	202	201,14	201,14	-0,86	0,00
10p15	211	208,71	208,71	-2,29	0,00
4q22	220	218,14	218,14	-1,86	0,00
10p14	229	227,99	227,99	-1,01	0,00

When you start training then the fragment size values you have entered are replaced by the mean fragment peak size of the data used for training, and if your data are of a too bad quality you might end up by having spoiled the reference data statistics. It is therefore always a good idea to make a backup copy of the program before you train it.

## 2 File formats

### 2.1 The format of a “MLPA-training-list file”

Training of the reference data is based on normal samples, and the training system has to know which normal samples to include and where each the MLPA data file is places.

The “MLPA-training-list file” that is required when doing the training (see “1 Training the analysis program with your own data”) contains this information in relation to the general folder where you have placed your MLPA data files.

The supplied file “P023-vs01-t02.txt” for training the P023 probe set is based on normal samples from three runs on the ABI and each set or run data has been placed to a separate folder (run 476, 483 and 484), and the file looks like this:

! Anything that follows an exclamation is regarded as being a comment

```
_2005-05-12_376\P23-N0 ! Example of a comment
_2005-05-12_376\P23-N1
_2005-05-12_376\P23-N10
_2005-05-12_376\P23-N11
_2005-05-12_376\P23-N12
_2005-05-12_376\P23-N13
_2005-05-12_376\P23-N14
_2005-05-12_376\P23-N2
_2005-05-12_376\P23-N3
_2005-05-12_376\P23-N4
_2005-05-12_376\P23-N5
_2005-05-12_376\P23-N6
_2005-05-12_376\P23-N7
_2005-05-12_376\P23-N8
_2005-05-12_376\P23-N9
```

```
_2005-06-09_483\P023-N15
_2005-06-09_483\P023-N16
_2005-06-09_483\P023-N17
_2005-06-09_483\P023-N18
_2005-06-09_483\P023-N19
_2005-06-09_483\P023-N20
_2005-06-09_483\P023-N21
```

!\_2005-06-09\_483\P023-N22 Example of excluding a sample of bad quality from training

```
_2005-06-09_484\P023-N23
_2005-06-09_484\P023-N24
_2005-06-09_484\P023-N25
_2005-06-09_484\P023-N26
_2005-06-09_484\P023-N27
_2005-06-09_484\P023-N28
_2005-06-09_484\P023-N29
_2005-06-09_484\P023-N30
```

```
_2005-06-09_484\P023-N32
_2005-06-09_484\P023-N33
_2005-06-09_484\P023-N34
_2005-06-09_484\P023-N35
_2005-06-09_484\P023-N36
_2005-06-09_484\P023-N37
```

```
!_2005-06-16_509 P023\P023-N22
_2005-06-16_509 P023\P023-N31
_2005-06-16_509 P023\P023-N38
_2005-06-16_509 P023\P023-N39
```