

**Analysis of Internal Validation Datasets
Using Open-Source Software *STR-validator***

Sarah Riman, Erica L. Romsos, Lisa Borsuk, and Peter M. Vallone

National Institute of Standards and Technology Gaithersburg,
Maryland, USA



Disclaimer

Points of view in this presentation are mine and do not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Commerce.

NIST Disclaimer Certain commercial products are identified in order to specify experimental procedures as completely as possible. In no case does such an identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of these products are necessarily the best available for the purpose.

Objectives

- ❖ The focus of this workshop is to **introduce the community to the availability of *STR-validator***, an open source software that can be utilized when **analyzing large internal validation data sets**. *STR-validator* was created by Oskar Hanson at the Norwegian Institute of Public Health.
- ❖ Participants will be **trained on how to import data obtained from the internal validation experiments of PowerPlex Fusion 6C into *STR-validator*** and **evaluate parameters** such as: analytical and stochastic thresholds, stutter percentage calculations, peak height ratios, base-pair sizing precision, and sensitivity.

Requirements



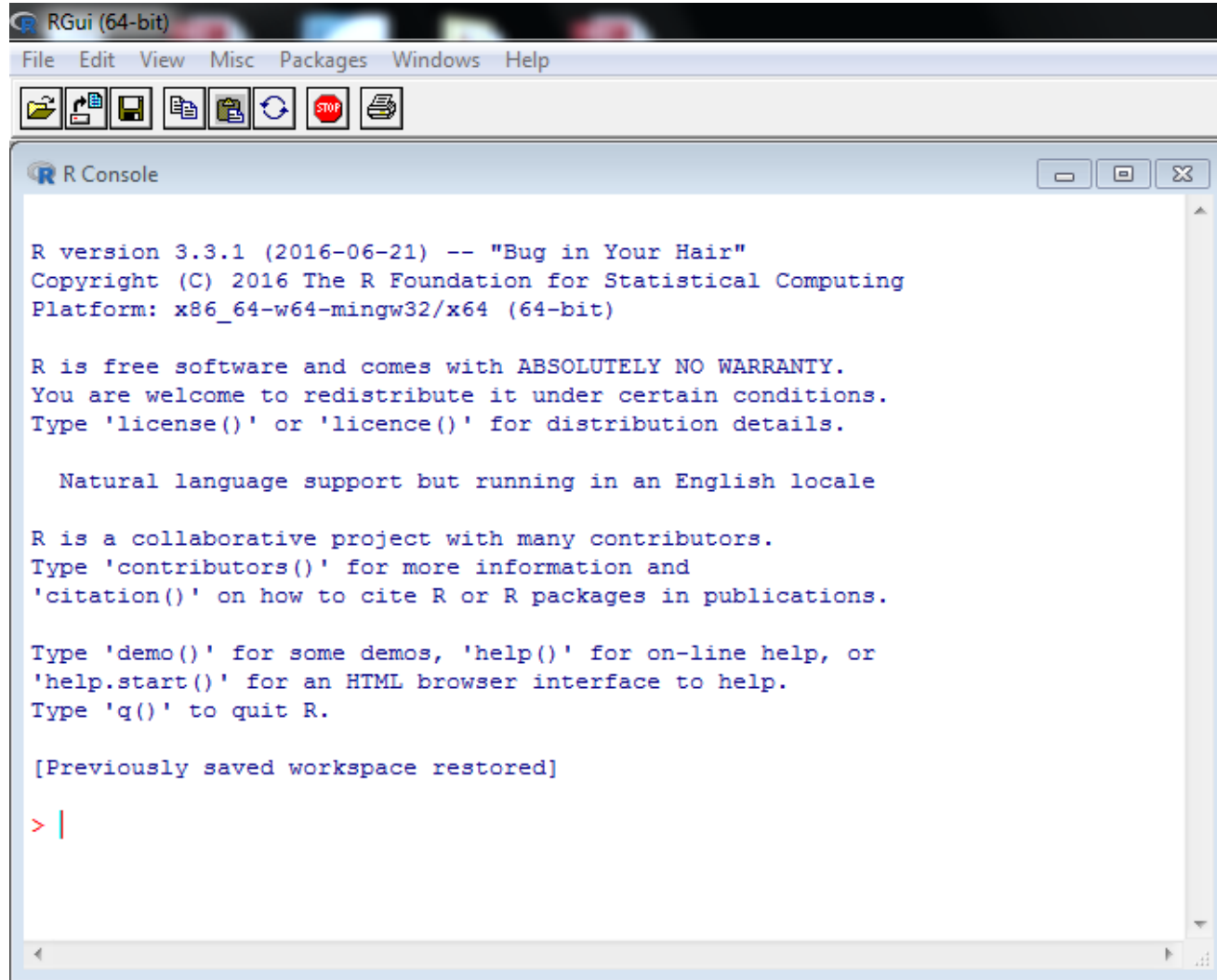
- ✓ Personal computers
- ✓ Installation of the **R Software**
- ✓ Installation of the ***STR-validator* Package**

Workshop Schedule

Time	Topic
9:00 AM-10:00 AM	<ul style="list-style-type: none">❖ Load <i>STR-validator</i> package and launch its GUI❖ Trim and Slim txt.files❖ Check Precision❖ Calculate Stutter Thresholds
10:00 AM – 10:10 AM	Break
10:10 AM-11:00 AM	<ul style="list-style-type: none">❖ Calculate Analytical Thresholds❖ Analyze Peak Height Ratio
11:00 AM -11:10 AM	Break
11:10 AM-12:00 PM	<ul style="list-style-type: none">❖ Calculate Stochastic Thresholds❖ Questions❖ Feedback about the workshop (survey)❖ Workshop ends

Launch R

➤ Launch R by clicking on 



The screenshot shows the RGui (64-bit) application window. The title bar reads "RGui (64-bit)". The menu bar includes "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations (Open, Save, Print, Copy, Paste, Undo, Redo, Stop, Refresh). The main window contains an "R Console" pane with the following text:

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

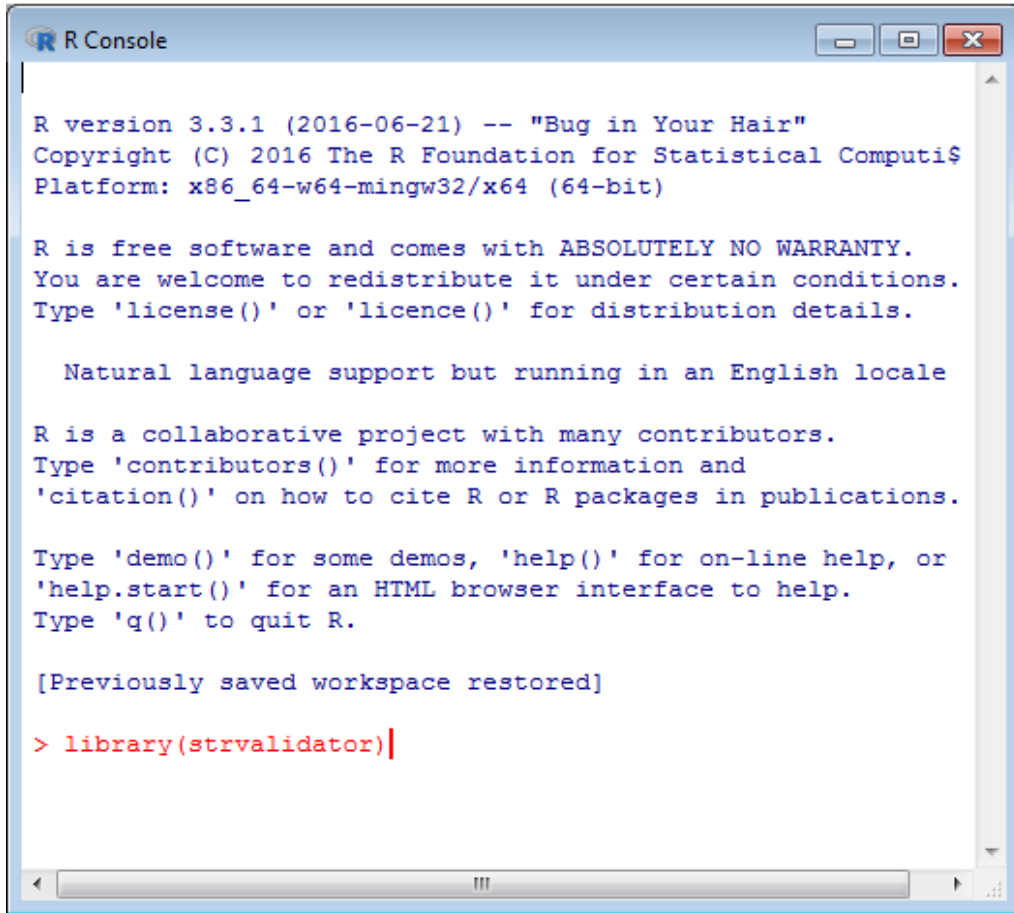
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

Load *STR-validator*

- In the R console, load the *STR-validator* package by typing **library(strvalidator)**



```
R Console
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

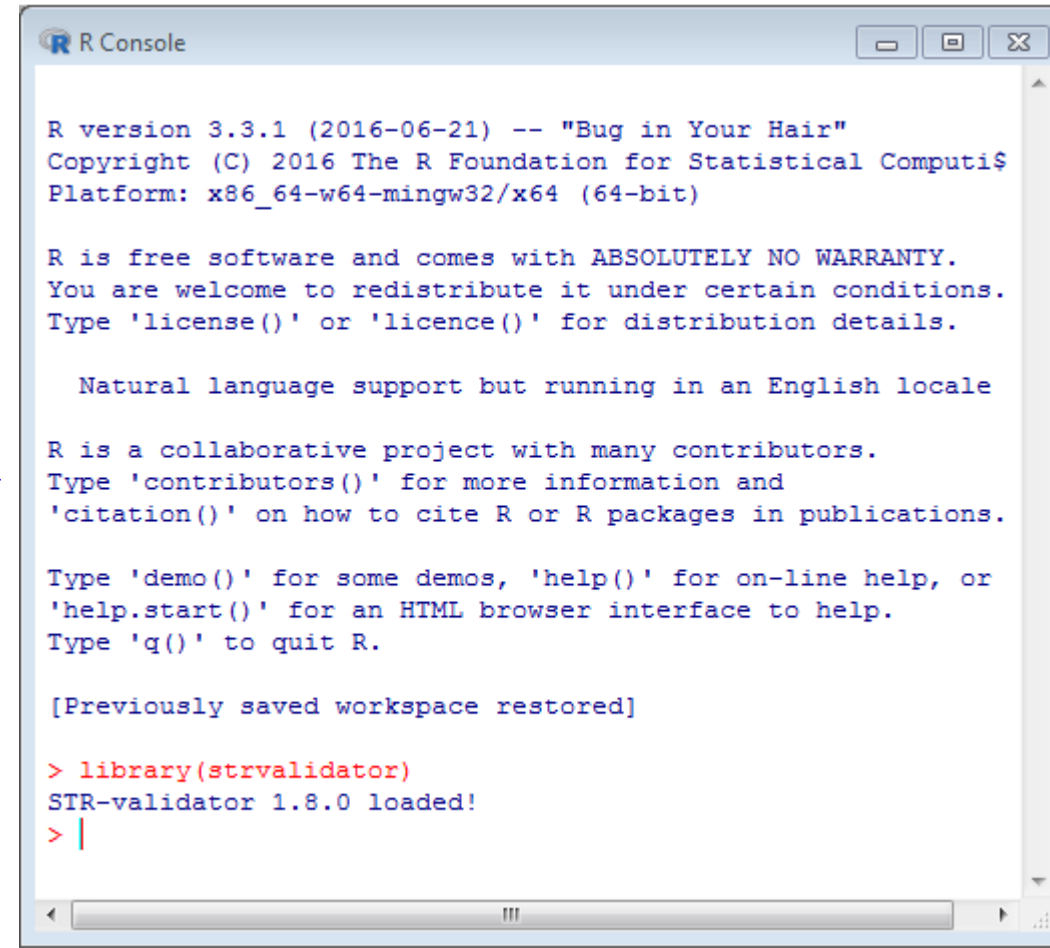
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library(strvalidator)|
```

Press Enter



```
R Console
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

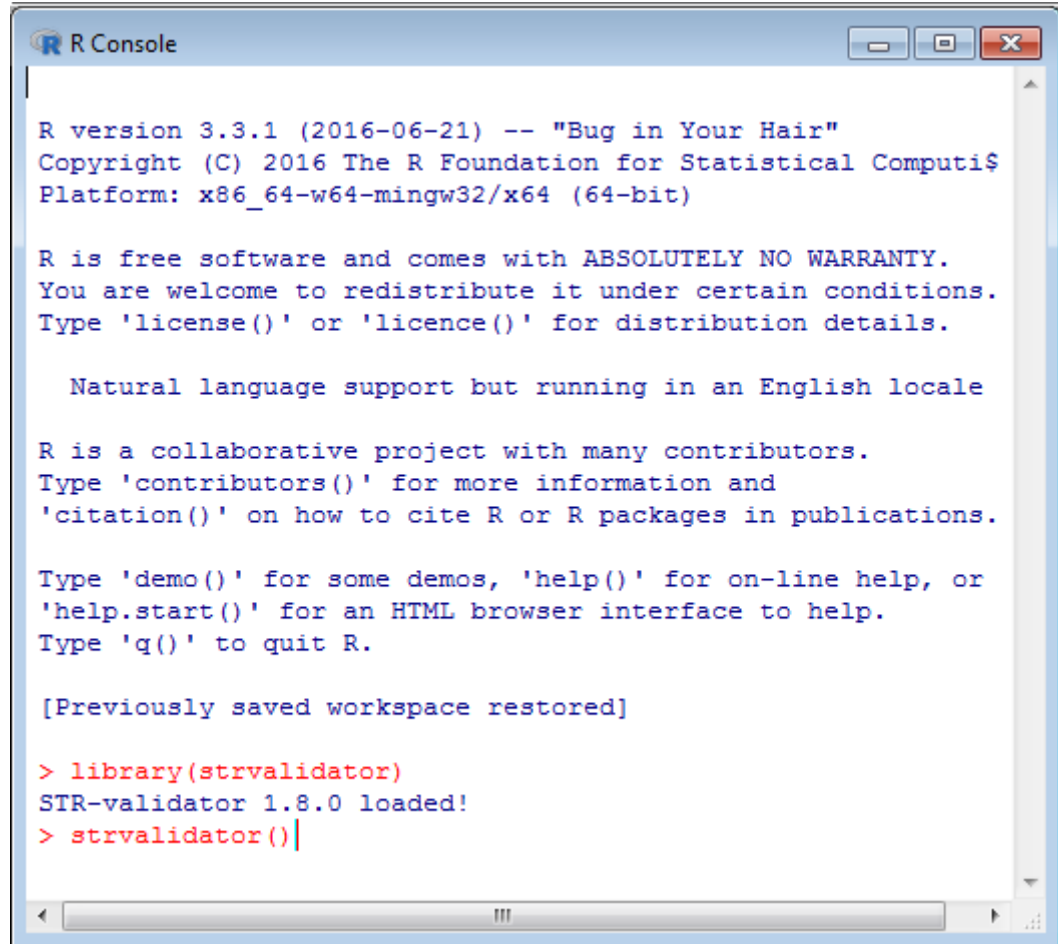
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library(strvalidator)
STR-validator 1.8.0 loaded!
> |
```

Launch *STR-validator* GUI

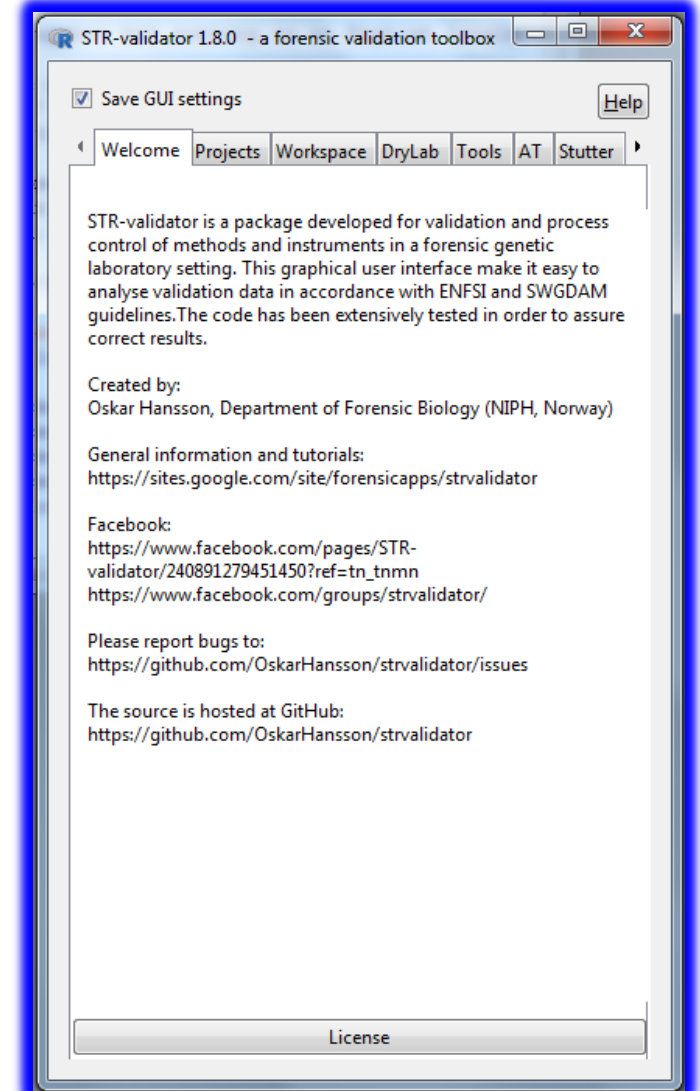
- In the R console, launch the *STR-validator* Graphical User Interphase by typing : **strvalidator()**



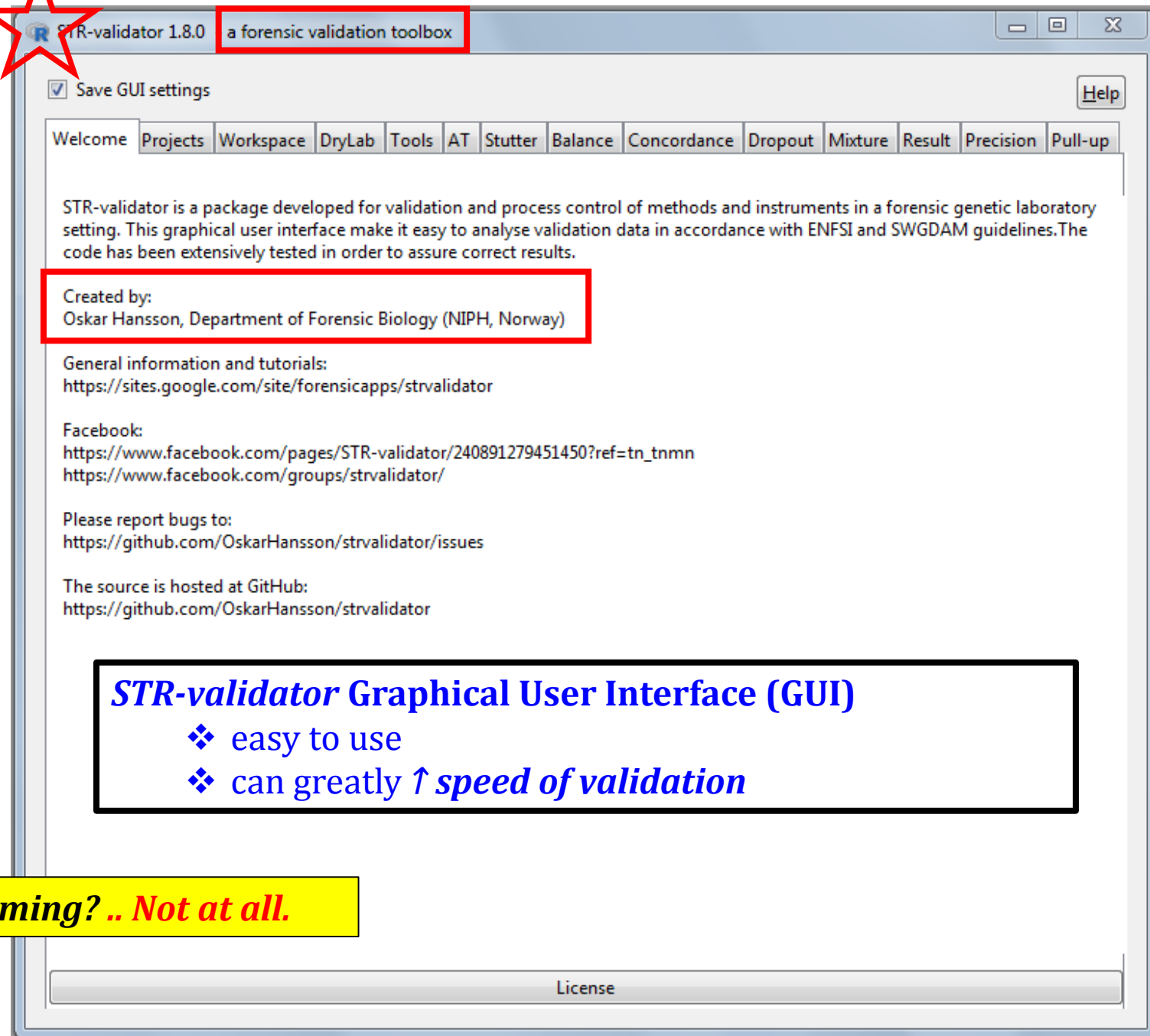
```
R Console  
  
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"  
Copyright (C) 2016 The R Foundation for Statistical Computi$  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
[Previously saved workspace restored]  
  
> library(strvalidator)  
STR-validator 1.8.0 loaded!  
> strvalidator()
```

Press Enter

The *STR-validator* main GUI



What is STR-Validator?



Created by:
Oskar Hansson, Department of Forensic Biology (NIPH, Norway)

STR-validator Graphical User Interface (GUI)

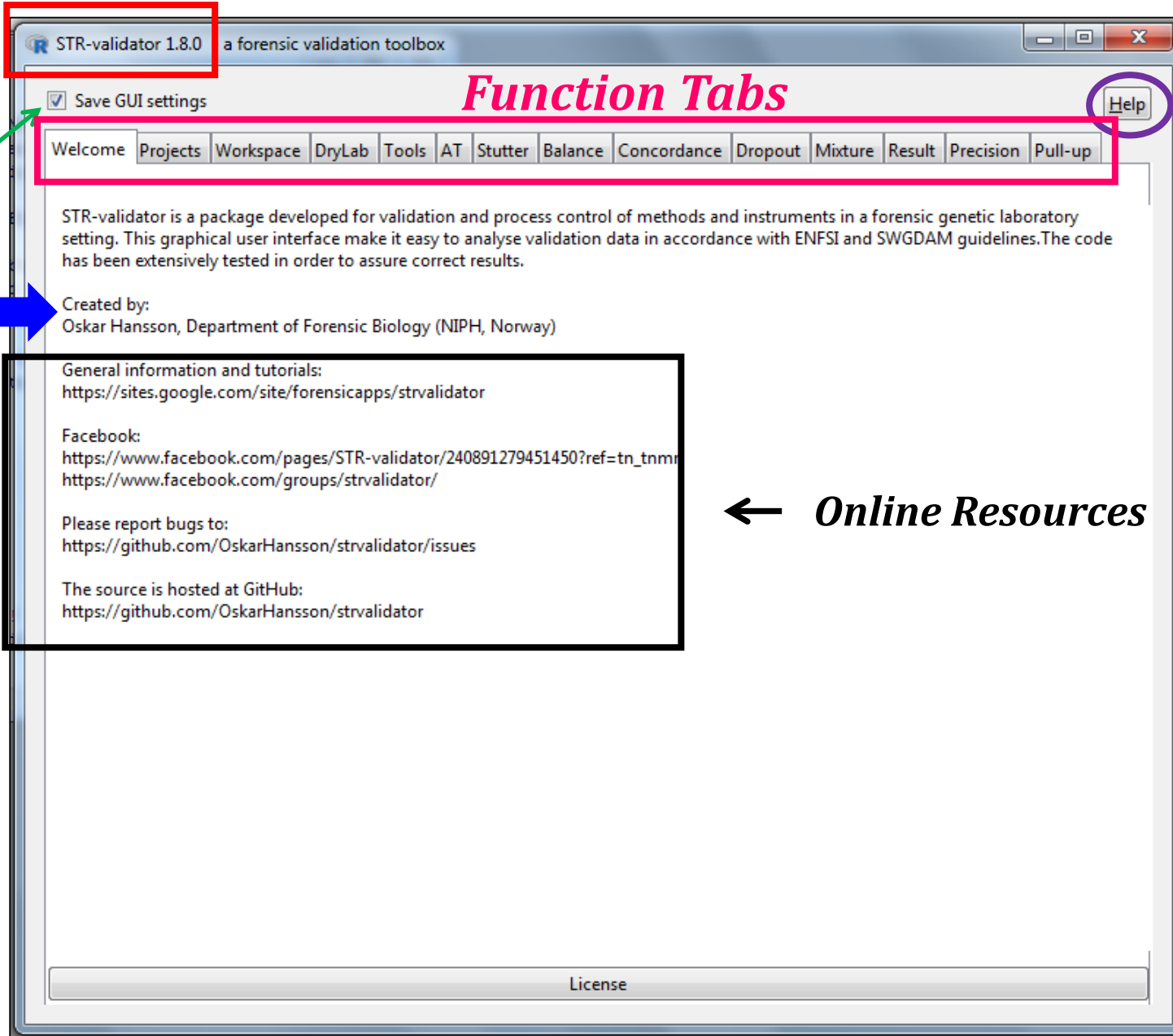
- ❖ easy to use
- ❖ can greatly ↑ *speed of validation*

❖ **Should I be knowledgeable about programming? .. Not at all.**

- ❖ A free and open source R-package
- ❖ Intended for:
 - ❖ Validating STR kits
 - ❖ Processing controls
 - ❖ Comparing methods and instrumentation

STR-Validator GUI Welcome Screen

Current Version



Remember Settings

Function Tabs

Help

Help Page for each Function

Creator of STR-Validator

Created by:
Oskar Hansson, Department of Forensic Biology (NIPH, Norway)

General information and tutorials:
<https://sites.google.com/site/forensicapps/strvalidator>

Facebook:
https://www.facebook.com/pages/STR-validator/240891279451450?ref=tn_tnm
<https://www.facebook.com/groups/strvalidator/>

Please report bugs to:
<https://github.com/OskarHansson/strvalidator/issues>

The source is hosted at GitHub:
<https://github.com/OskarHansson/strvalidator>

Online Resources

License

***Analysis of Internal Validation Study of
PowerPlex Fusion 6C Using STR-Validator***

PowerPlex Fusion 6C

- ❖ The largest commercial STR multiplex kit available for CE use.
- ❖ Has a total of 27 loci including the 20 CODIS core loci.
- ❖ The 27 loci are in 6 dyes and include:
 - SE33, Penta D and Penta E
 - 3 Y-STR markers (DYS391, DYS570, DYS576)
- ❖ A one kit for both direct amplification and casework with a 60 min PCR time capability.
- ❖ It gives ~17 orders of magnitude of improvement using the NIST 1036 data set.

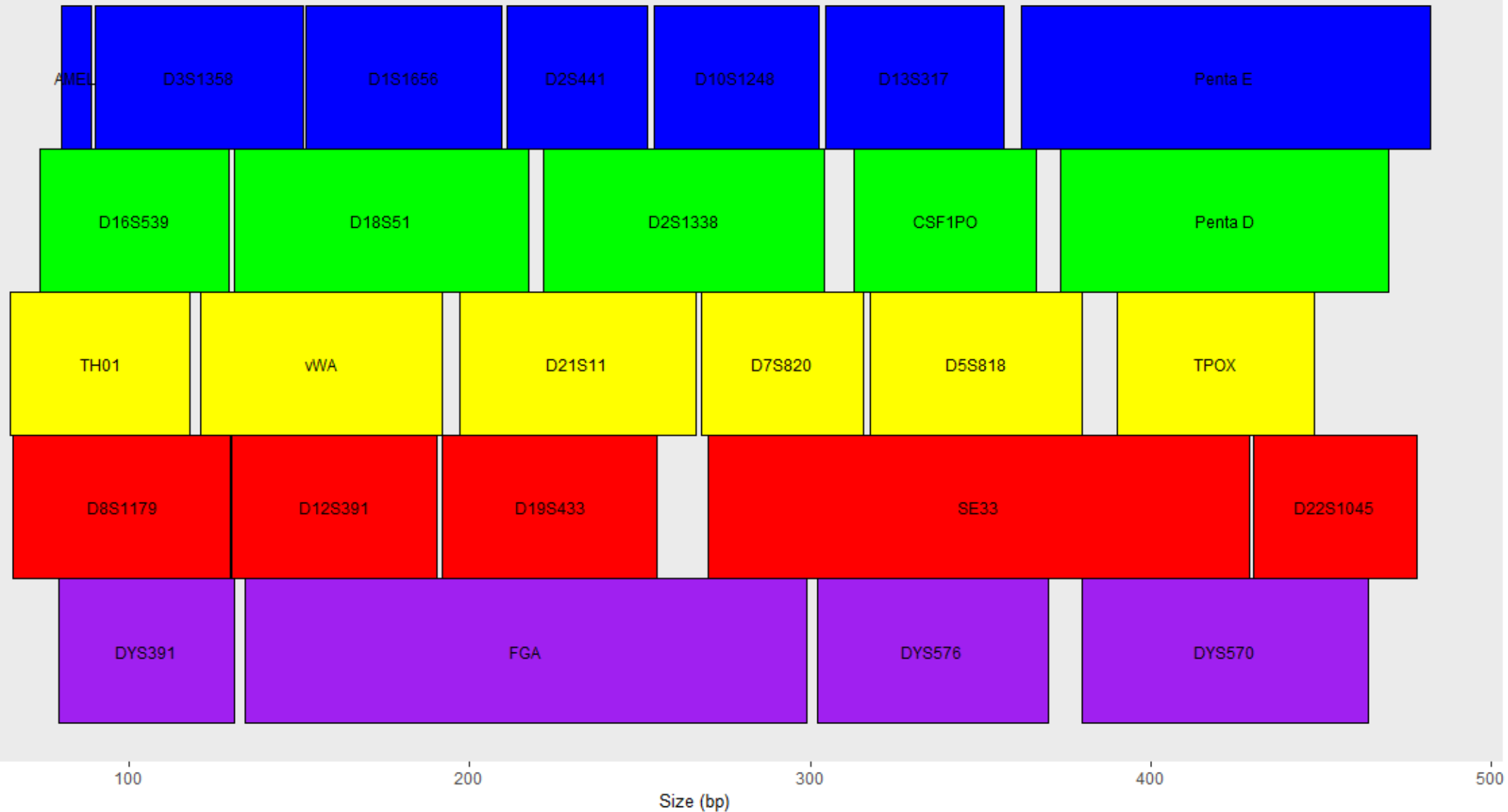


❖ <http://www.promega.com/products/pm/genetic-identity/powerplex-fusion/>

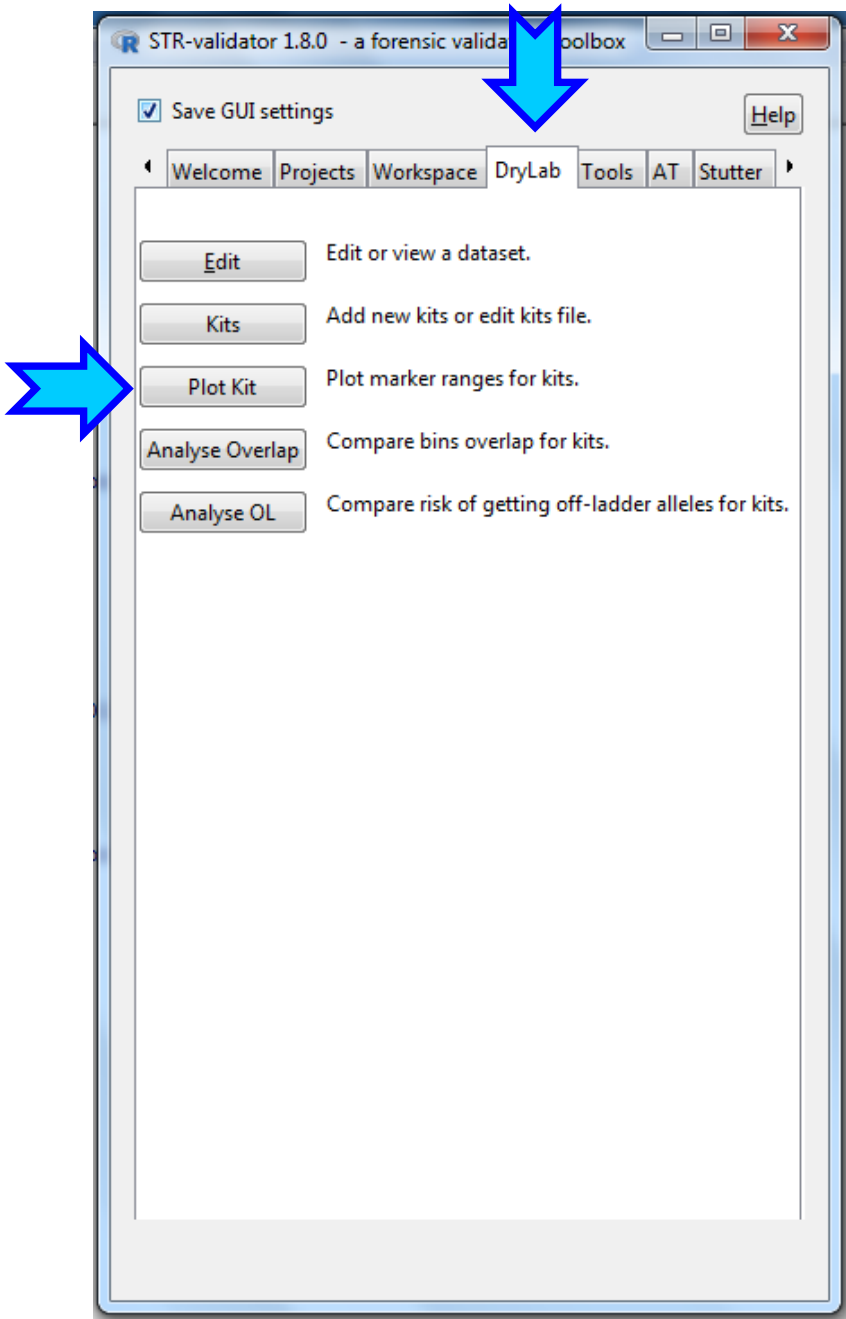
❖ Butler, J.M., Hill, C.R. and Coble, M.D. (2012) Variability of New STR Loci and Kits in US Population Groups. *Profiles in DNA*

Marker size range

PowerPlex Fusion 6C



Plot Kit



Plot PowerPlex Fusion 6C

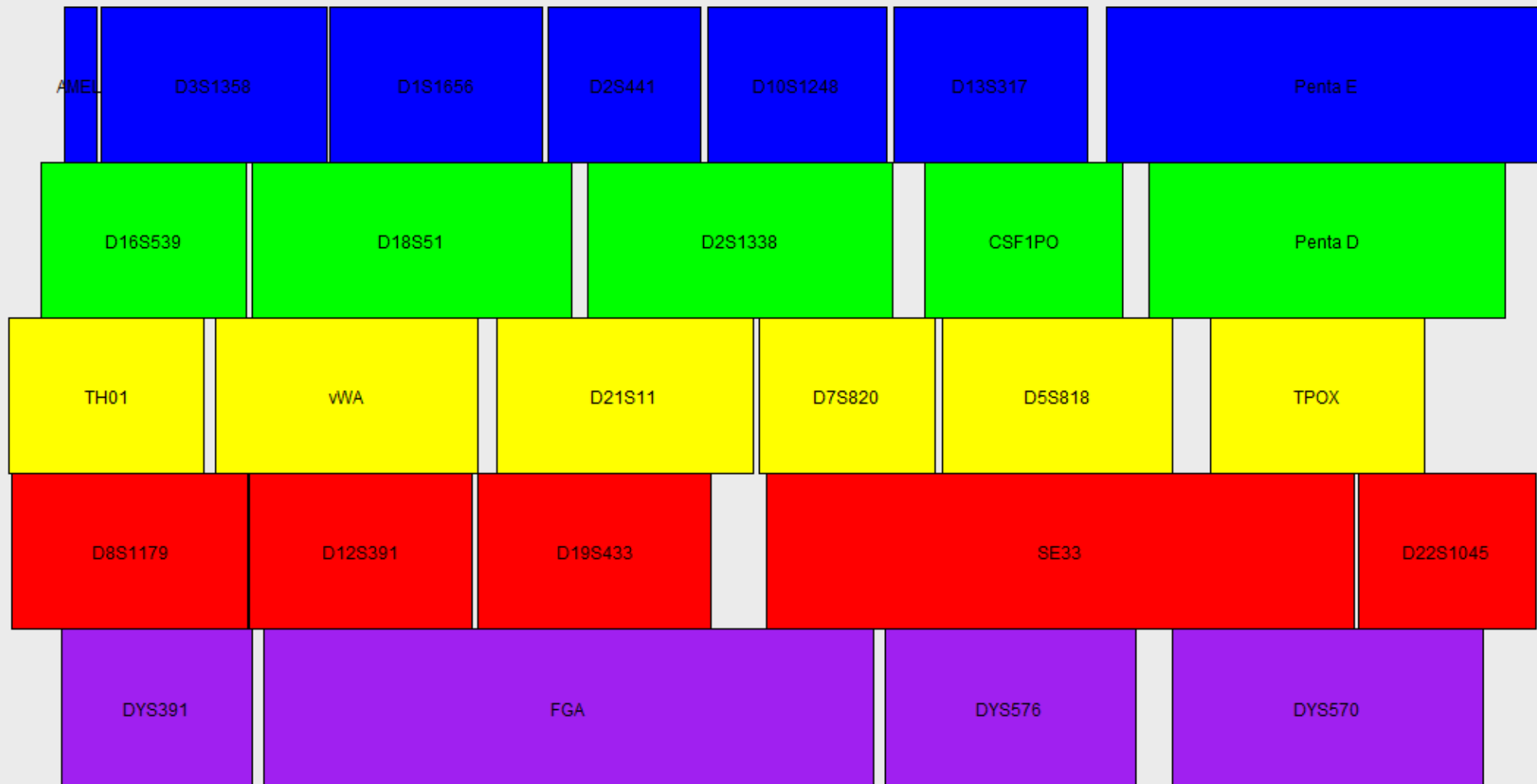
The screenshot shows the 'Plot kit' window with the following settings:

- Save GUI settings
- Select kits:**
 - ESX16
 - ESX17
 - ESX17Fast
 - ES17Fast
 - Fusion
 - Fusion 6C
 - SGMPlus
 - Identifiler
 - NGM
 - NGMSelect
 - GlobalFiler
 - PowerPlex16
 - 24plex
 - ESSPlex
 - ESSplexPlus
 - ESSplexSEPlus
 - ESSplexSEQS
 - Y23
 - YfilerPlus
- Options:**
 - Plot title: Marker size range (Size: 20)
 - X title: Size (bp) (Size: 10)
 - Kit name size: 4
 - Inter kit spacing: 2
 - Marker name size: 3
 - Marker height: 0.5
 - Marker range alpha: 1.0
- Plot kit: plot
- Save as:
 - Name for result: PowerPlexFusion6C_ggplot
 - Buttons: Save as object, Save as image

Three blue arrows point to the 'Fusion 6C' checkbox, the 'plot' button, and the 'Name for result' field.

Marker size range

PowerPlex Fusion 6C



100

200

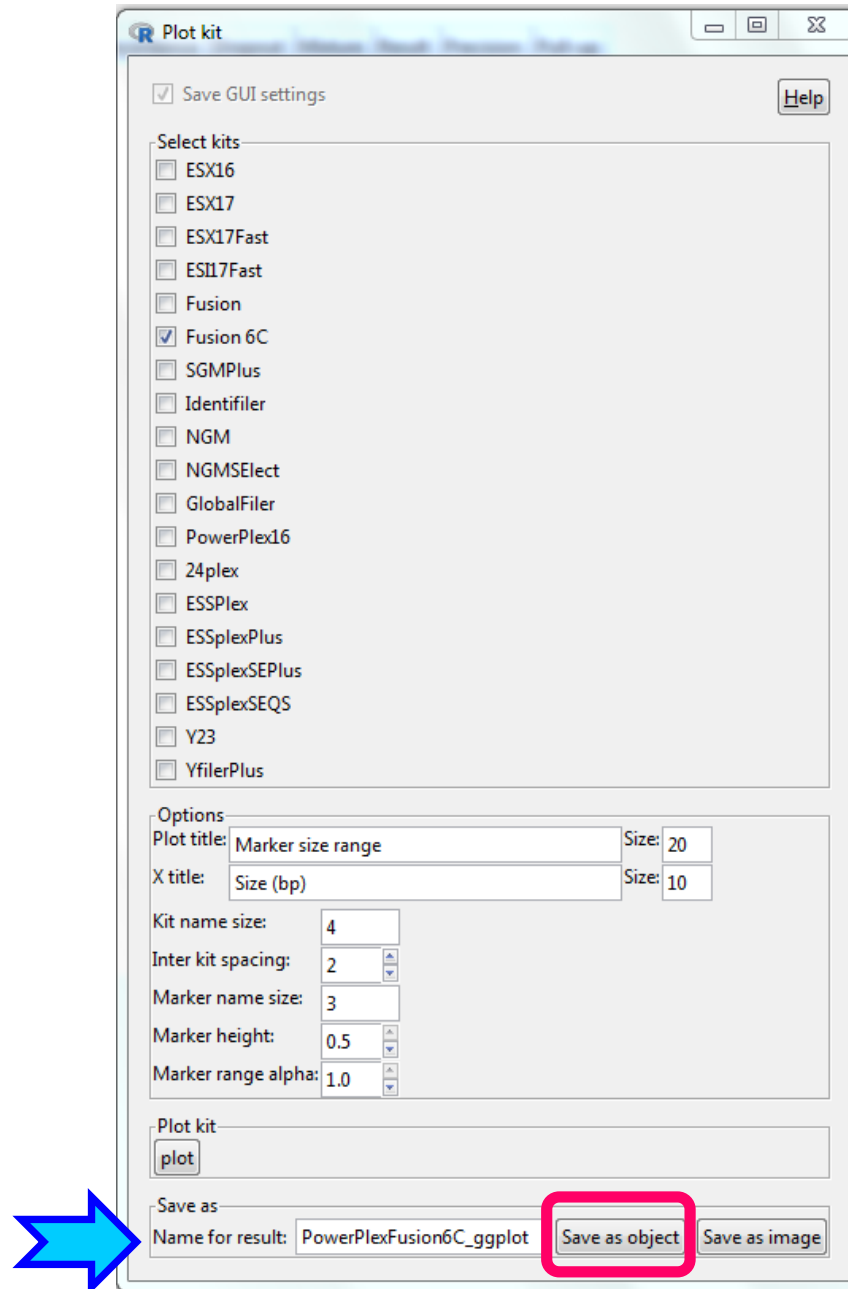
300

400

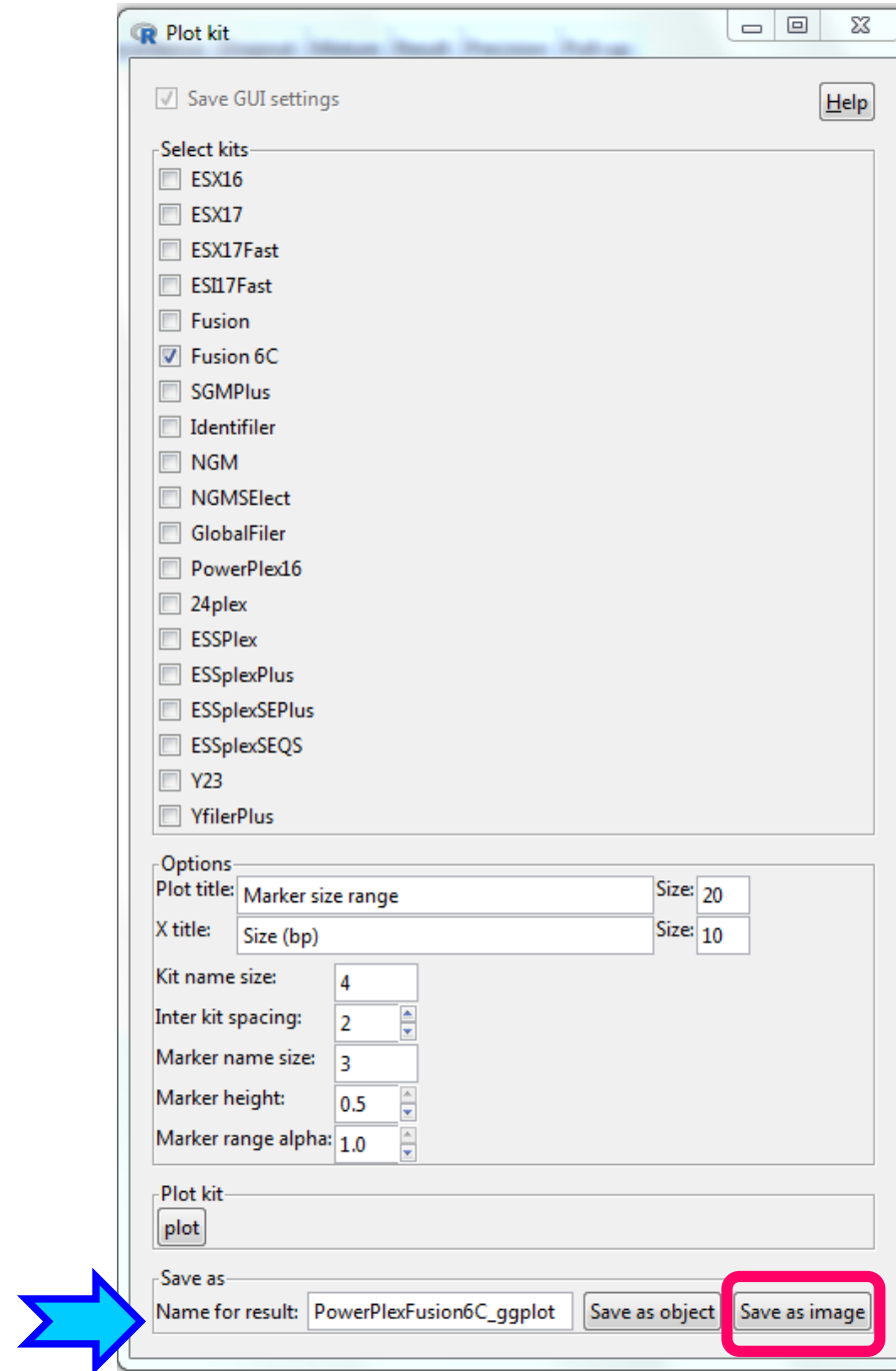
500

Size (bp)

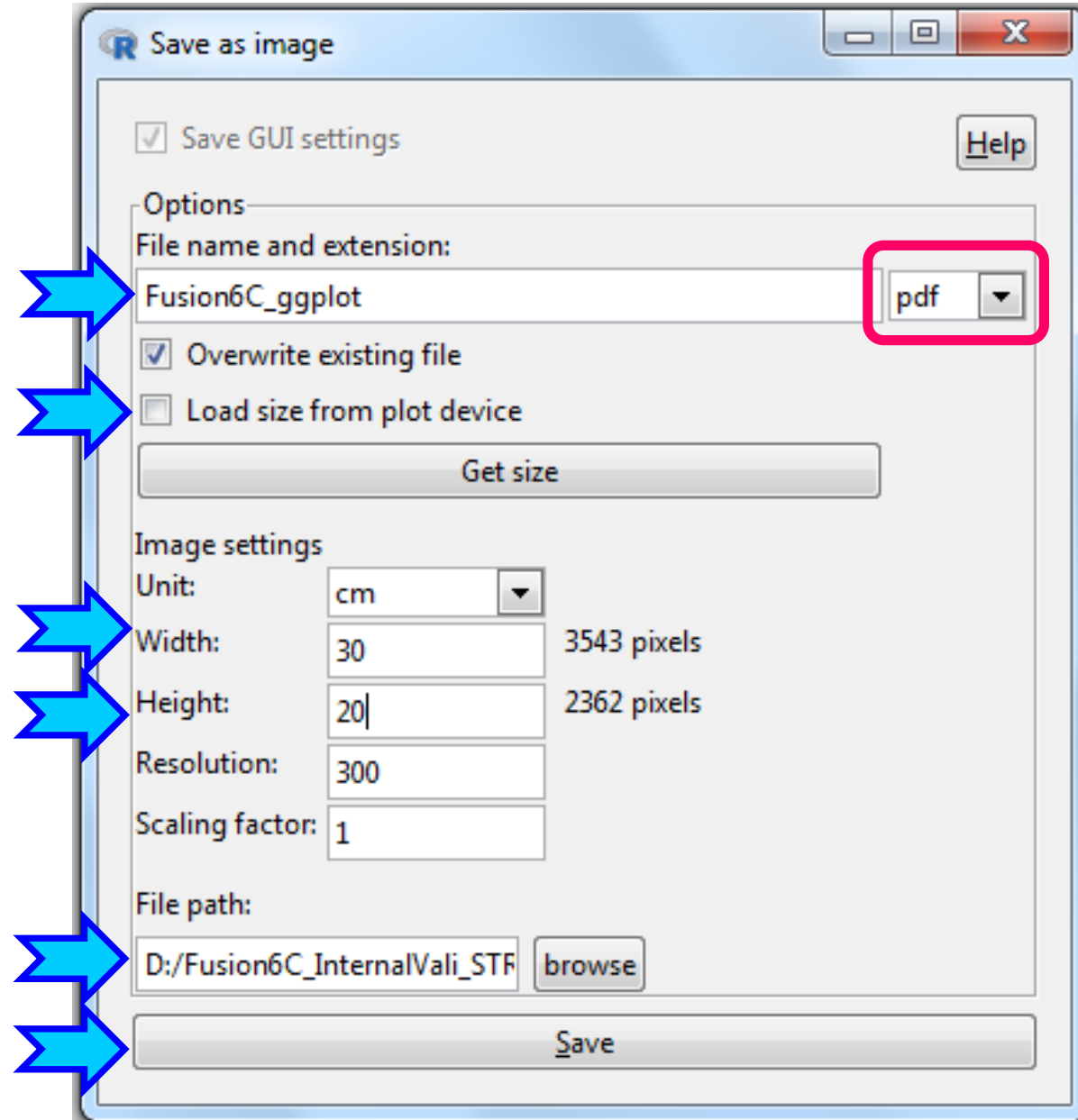
Save PowerPlex Fusion 6C in the workspace



Save the plot as an image

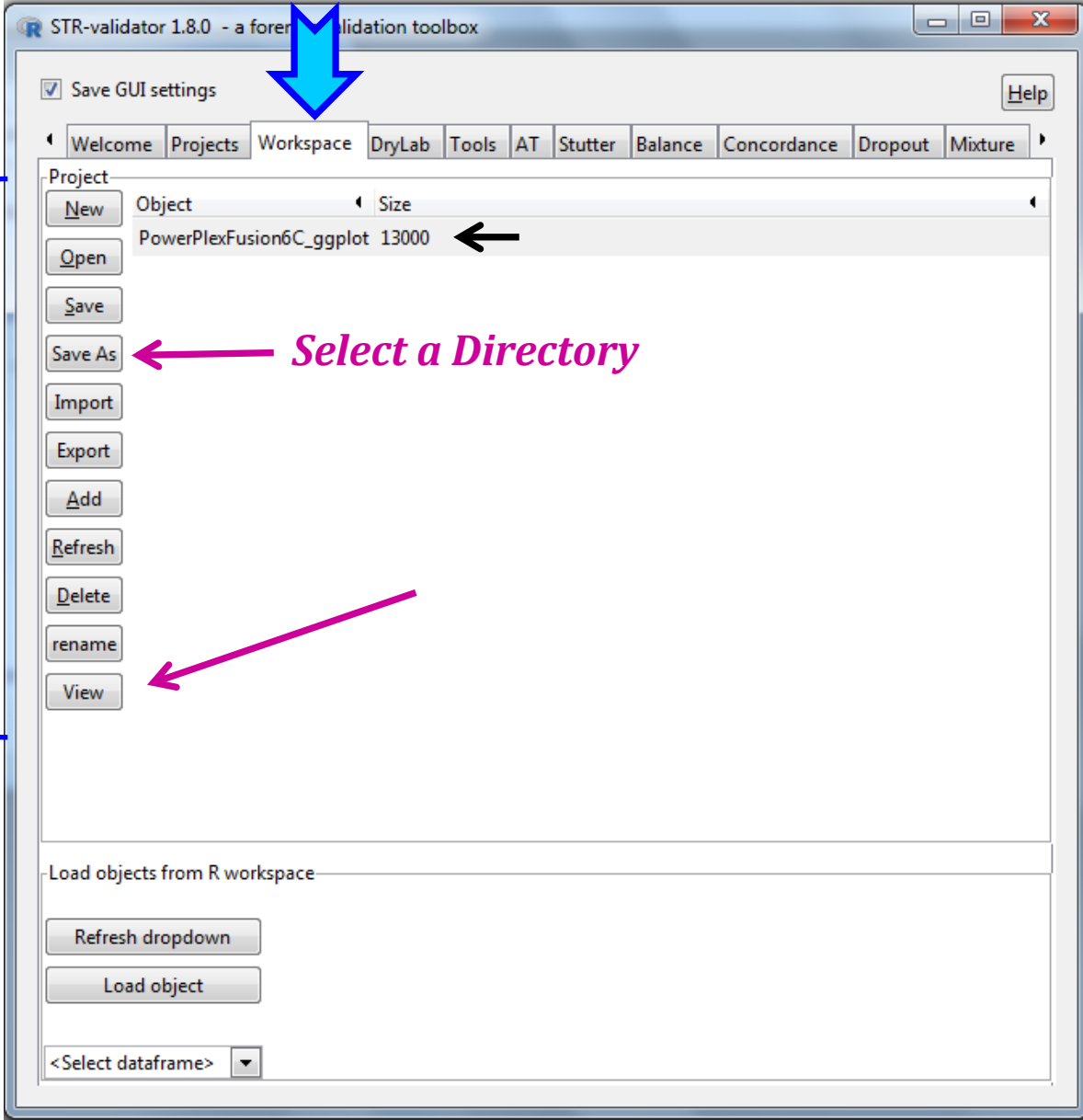
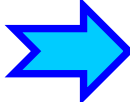


Save the plot as an image

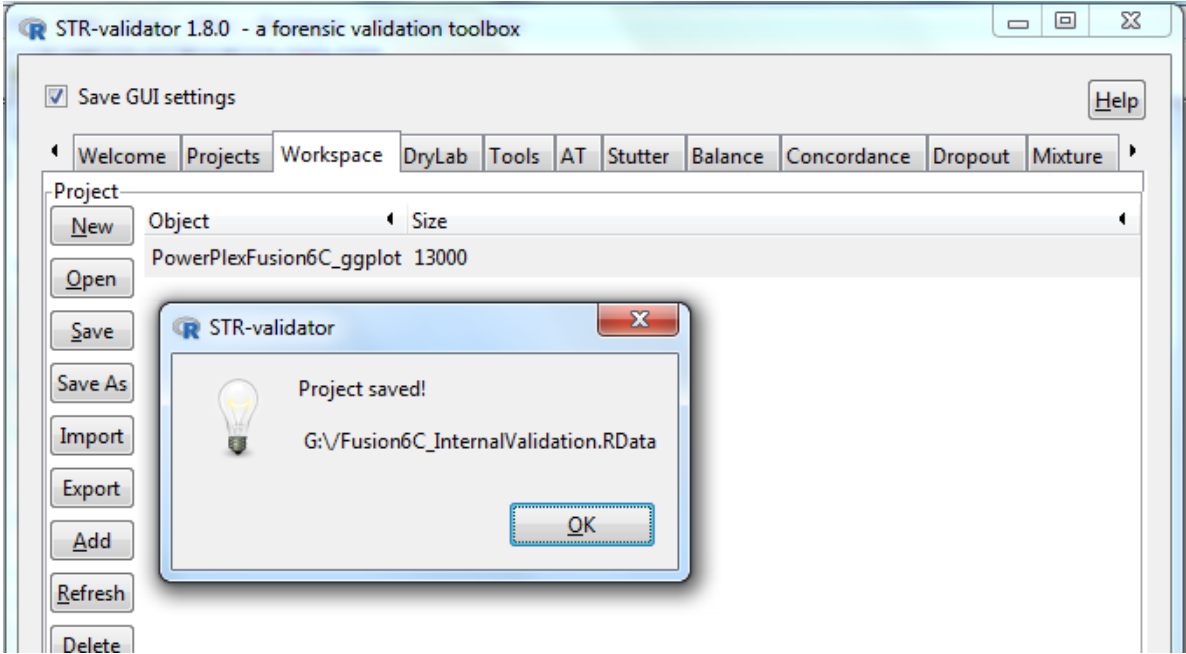
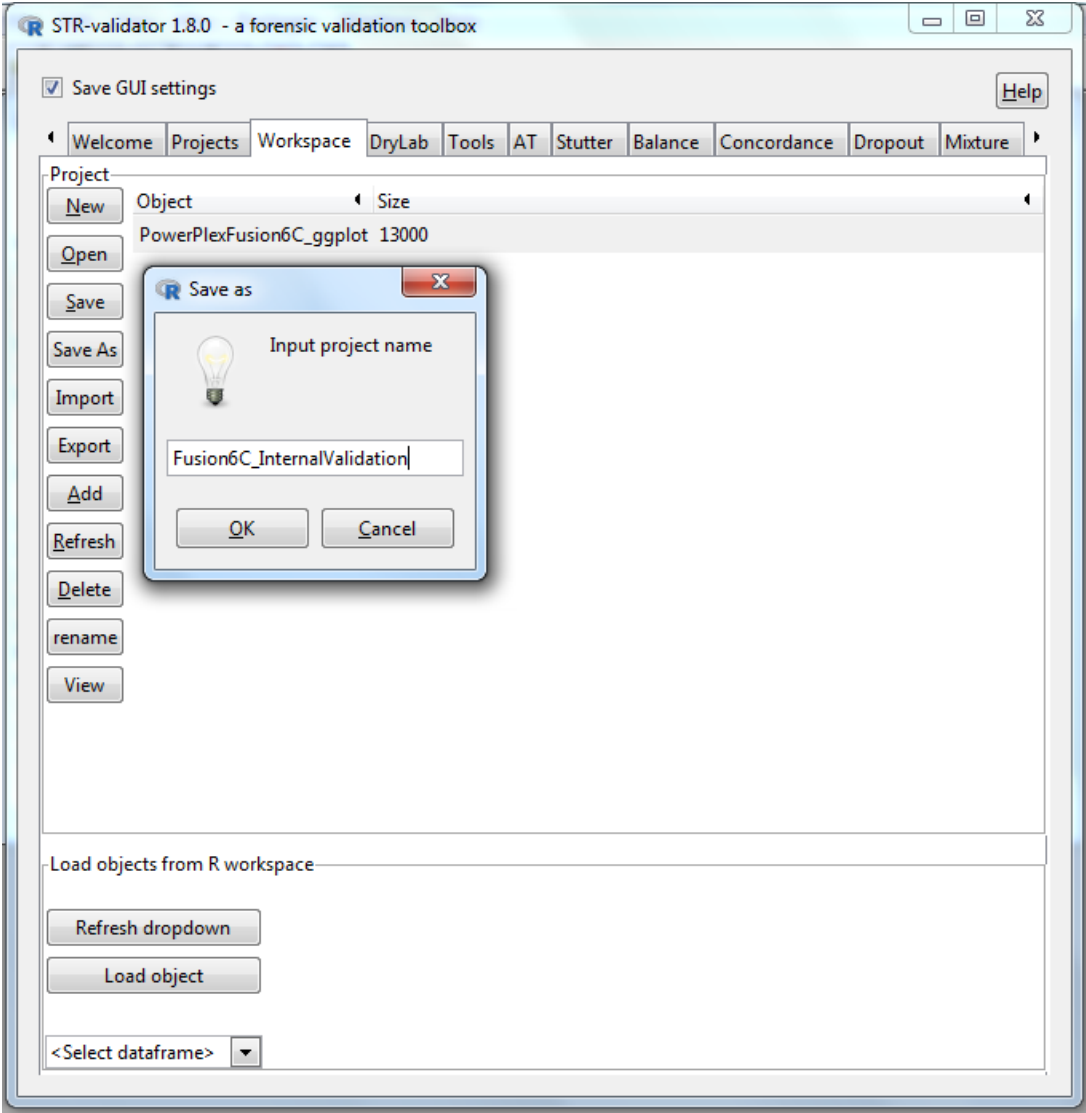


The Workspace Tab

- ❖ *View your plot*
- ❖ *Save your project*



Save Your Project



PowerPlex Fusion 6C



- ❖ This is to remind you that the *STR-validator* will **automatically detect** the Fusion 6C kit.
- ❖ In case the kit of interest is not in the software you can add it to the *STR-validator* through the **DryLab Tab**.

What happens if R quits on you?

```
R Console

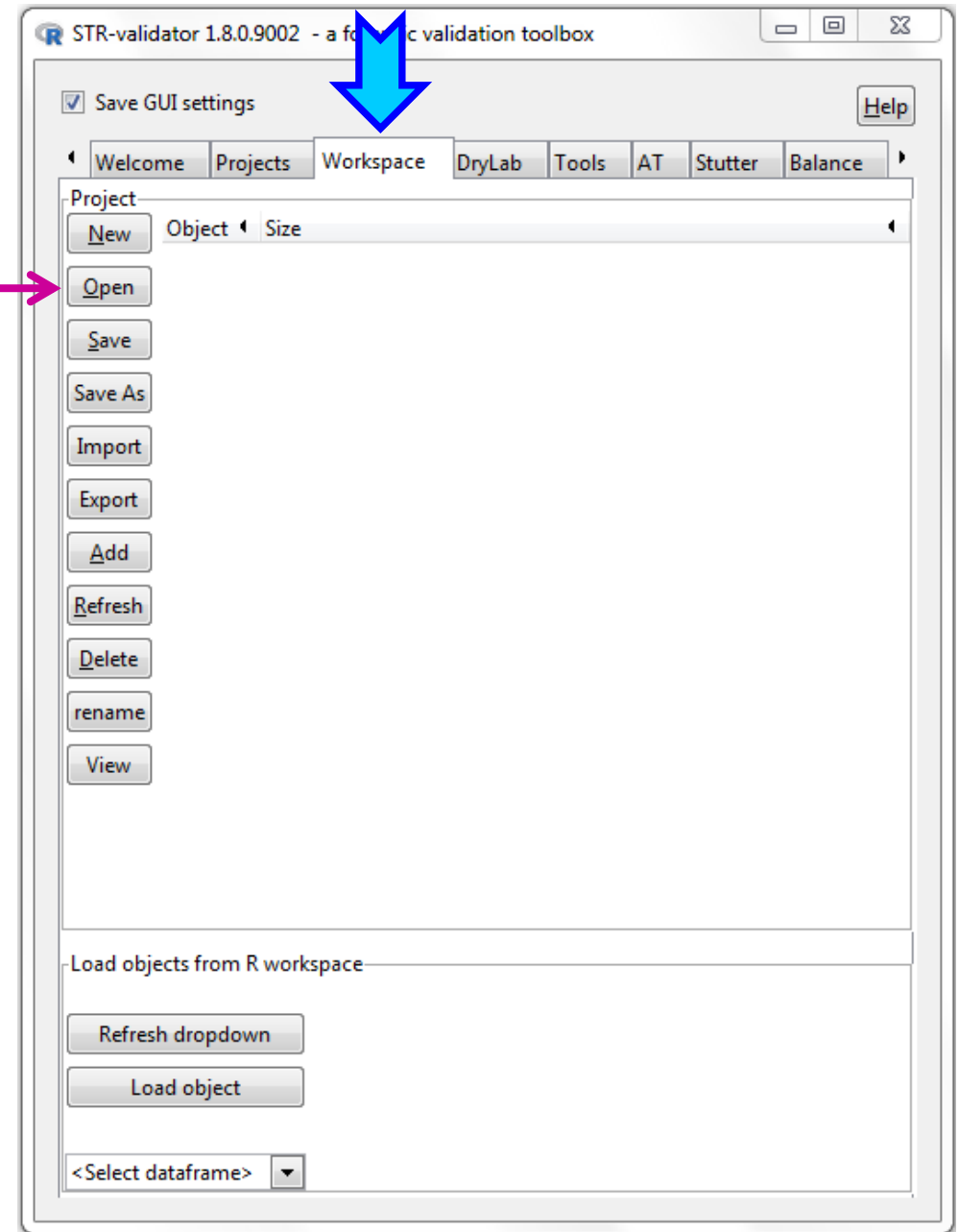
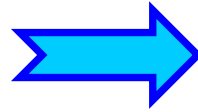
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(strvalidator)
STR-validator 1.8.0 loaded!
> strvalidator()
Loading required package: gWidgetsRGtk2
Loading required package: RGtk2
Loading required package: gWidgets
Loading required package: cairoDevice
STR-validator graphical user interface loaded!
[1] "G:\\\"
> strvalidator()
```



Alternatively, Open a Project from the Project Tab

```
R Console  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> library(strvalidator)  
STR-validator 1.8.0 loaded!  
> strvalidator()  
Loading required package: gWidgetsRGtk2  
Loading required package: RGtk2  
Loading required package: gWidgets  
Loading required package: cairoDevice  
STR-validator graphical user interface loaded!  
[1] "G:\\"  
> strvalidator()
```



STR-validator 1.8.0 - a forensic validation toolbox

Save GUI settings Help

Welcome Projects Workspace DryLab Tools AT Stutter Balance

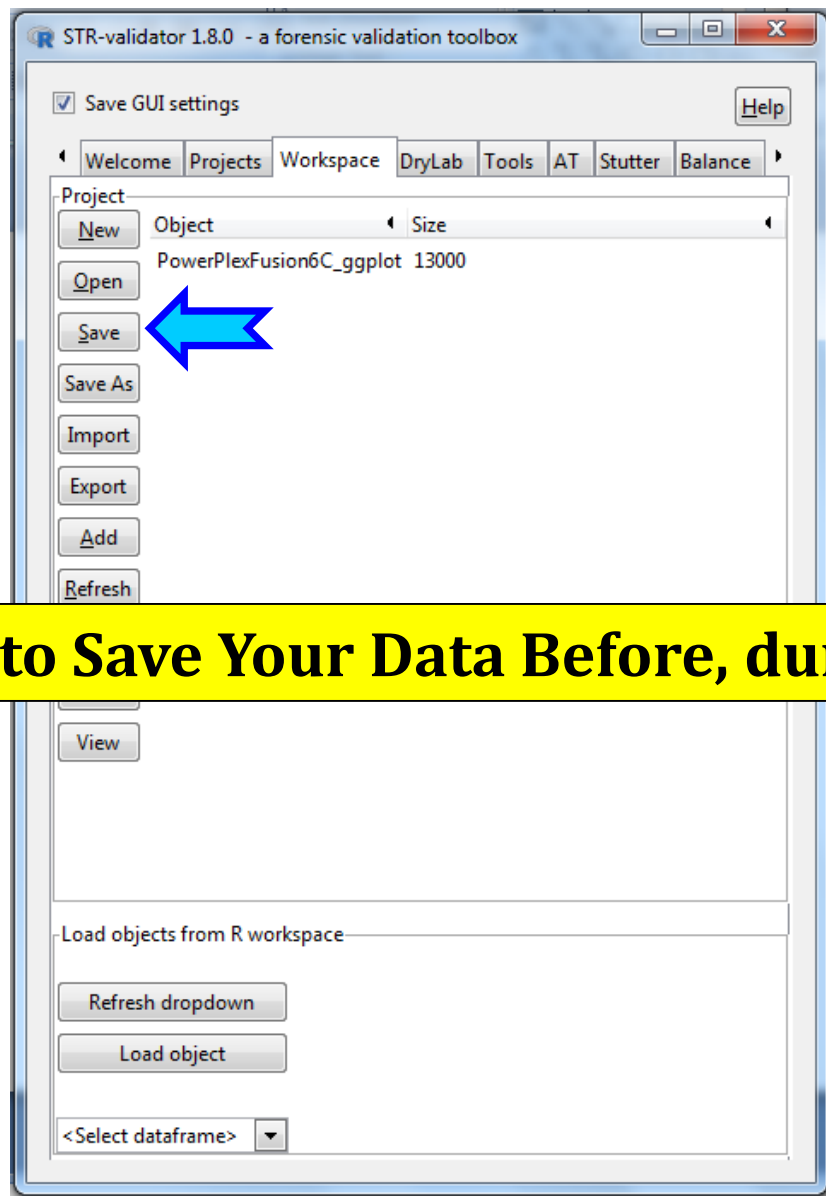
Folder:
G:\ browse

Projects

Name	Date	Size	Id
Fusion6C_InternalValidation.RData	2016-10-22 18:40:52	32087	1

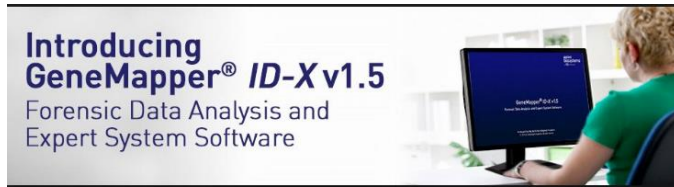
Open Add Delete

Description
Save Project:



Remember to Save Your Data Before, during, and after analysis

How to Prepare the Data for Analysis?

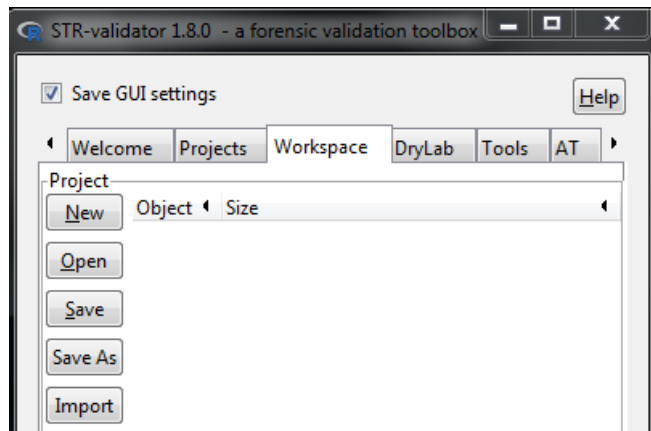


Export (.txt)

Sample Name	Marker	Dye	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10	Allele 11	Allele 12	Allele 13	Allele 14	Allele 15	Allele 16
29_0.03ng	AMEL	B																
29_0.03ng	D3S1358	B	OL	OL		9	OL	OL	OL			17.2						
29_0.03ng	D1S1656	B	OL	OL	OL		OL	OL	OL	OL								
29_0.03ng	D2S441	B	OL	OL		11.3	OL		17									
29_0.03ng	D10S1248	B	OL	OL		16	OL	OL										
29_0.03ng	D13S317	B	OL	OL		8.1	OL	OL										
29_0.03ng	Penta E	B	OL	OL	OL	OL	OL	OL	OL	OL	OL	25	26					
29_0.03ng	D16S539	G	OL	OL			11.3											
29_0.03ng	D18S51	G	OL		9.2	OL		15.2	OL		21.1	OL	OL	OL	OL			
29_0.03ng	D2S1338	G	OL	OL		23	OL	OL	OL									
29_0.03ng	CSF1PO	G																
29_0.03ng	Penta D	G	OL		7.4	OL	OL		17	OL	OL	OL	OL					
29_0.03ng	TH01	Y	OL		7.3		8.3	OL	10	OL	OL	OL	OL					
29_0.03ng	vWA	Y	OL		17	OL	OL	OL										
29_0.03ng	D21S11	Y	OL	OL		30.3	OL		33.1		35							
29_0.03ng	D7S820	Y		6	8.1		9	10.1	12.1		13							
29_0.03ng	D5S818	Y	OL	OL	OL	OL	OL		12	12.1	OL							

Semi-Wide type of table Format = *Unstacked Data*

Import (.txt)



Slim →

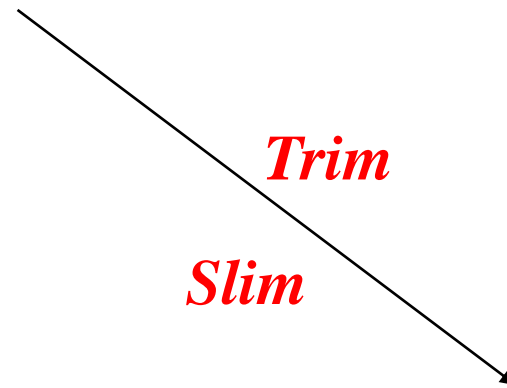
Sample.Name	Marker	Dye	Allele	Size	Height
29_0.03ng	AMEL	B	NA	NA	NA
29_0.03ng	D3S1358	B	OL	92.06	8
29_0.03ng	D3S1358	B	OL	95.36	6
29_0.03ng	D3S1358	B	9	96.68	11
29_0.03ng	D3S1358	B	OL	101.95	9
29_0.03ng	D3S1358	B	OL	113.26	10
29_0.03ng	D3S1358	B	OL	130.44	6
29_0.03ng	D3S1358	B	17.2	133.66	4
29_0.03ng	D1S1656	B	OL	155.95	9
29_0.03ng	D1S1656	B	OL	158.42	8
29_0.03ng	D1S1656	B	OL	162.86	7
29_0.03ng	D1S1656	B	OL	164.83	8
29_0.03ng	D1S1656	B	OL	175.23	7
29_0.03ng	D1S1656	B	OL	179.7	7
29_0.03ng	D1S1656	B	OL	200.71	3
29_0.03ng	D2S441	B	OL	210.62	8
29_0.03ng	D2S441	B	OL	223.02	8
29_0.03ng	D2S441	B	11.3	227.1	8
29_0.03ng	D2S441	B	OL	240.29	8
29_0.03ng	D2S441	B	17	249.25	5
29_0.03ng	D10S1248	B	OL	257.96	6
29_0.03ng	D10S1248	B	OL	261.41	6
29_0.03ng	D10S1248	B	16	287.47	7
29_0.03ng	D10S1248	B	OL	290.03	7
29_0.03ng	D10S1248	B	OL	294.63	7
29_0.03ng	D13S317	B	OL	307.63	8
29_0.03ng	D13S317	B	OL	310.92	8

Semi-long narrow type of table Format = *Slim or Stacked data*

How to Manually Trim/Slim the Data for Analysis in STR-validator?

Sample Name	Marker	Dye	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10	Allele 11	Allele 12	Allele 13	Allele 14	Allele 15	Allele 16
29_0.03ng	AMEL	B																
29_0.03ng	D3S1358	B	OL	OL		9	OL	OL	OL		17.2							
29_0.03ng	D1S1656	B	OL	OL	OL	OL	OL	OL	OL									
29_0.03ng	D2S441	B	OL	OL		11.3	OL		17									
29_0.03ng	D10S1248	B	OL	OL		16	OL	OL										
29_0.03ng	D13S317	B	OL	OL		8.1	OL	OL										
29_0.03ng	Penta E	B	OL	OL	OL	OL	OL	OL	OL	OL	OL	OL	OL	OL	25	26		
29_0.03ng	D16S539	G	OL	OL	OL			11.3										
29_0.03ng	D18S51	G	OL		9.2	OL		15.2	OL		21.1	OL	OL	OL	OL			
29_0.03ng	D2S1338	G	OL	OL		23	OL	OL	OL									
29_0.03ng	CSF1PO	G																
29_0.03ng	Penta D	G	OL		7.4	OL		OL		17	OL	OL	OL					
29_0.03ng	TH01	Y	OL		7.3	8.3	OL		10	OL	OL	OL	OL					
29_0.03ng	vWA	Y	OL		17	OL	OL	OL										
29_0.03ng	D21S11	Y	OL	OL		30.3	OL		33.1		35							
29_0.03ng	D7S820	Y		6	8.1		9	10.1		12.1		13						
29_0.03ng	D5S818	Y	OL	OL	OL	OL		12	12.1	OL								

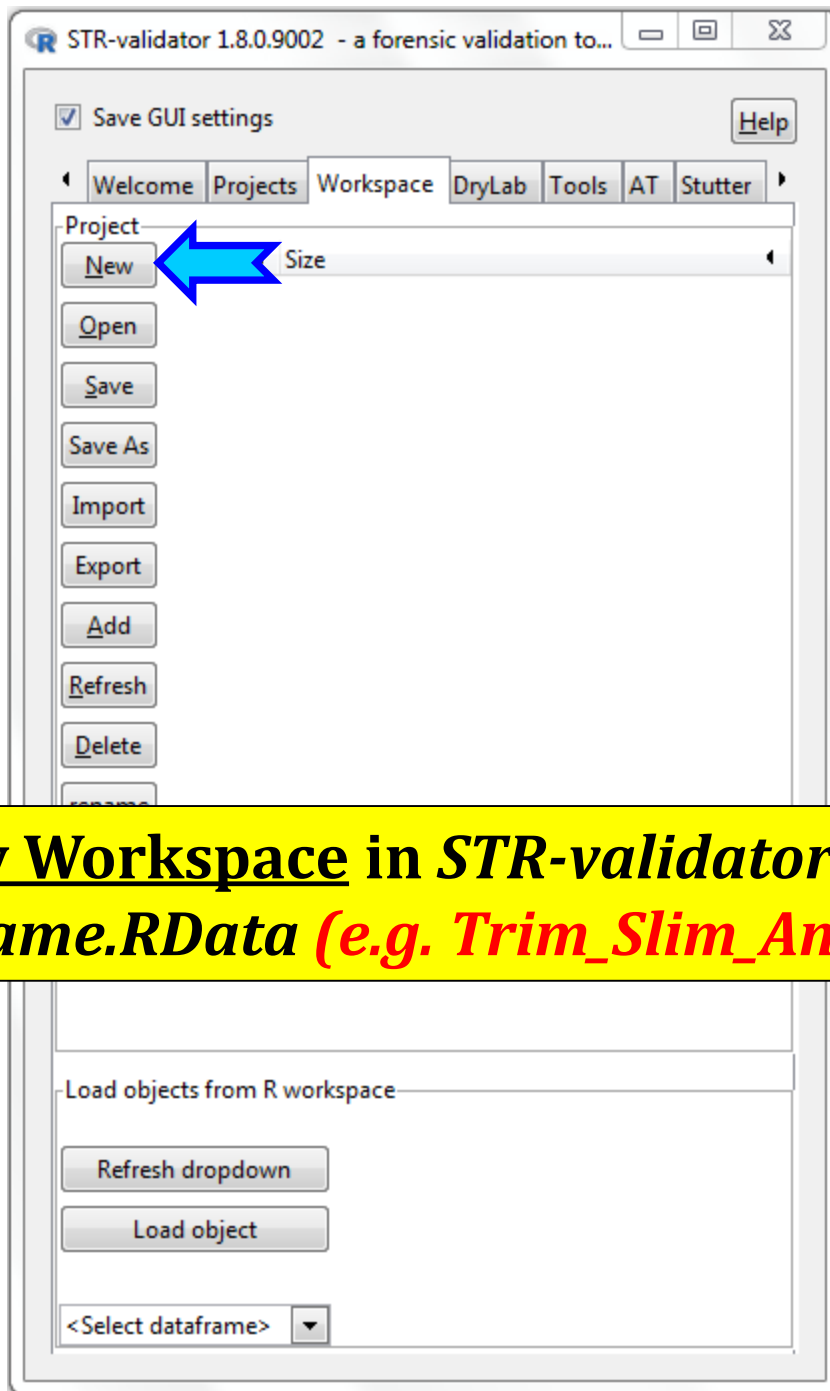
Semi-Wide type of table Format = *Unstacked Data*



Sample.Name	Marker	Dye	Allele	Size	Height
29_0.03ng	AMEL	B	NA	NA	NA
29_0.03ng	D3S1358	B	OL	92.06	8
29_0.03ng	D3S1358	B	OL	95.36	6
29_0.03ng	D3S1358	B	9	96.68	11
29_0.03ng	D3S1358	B	OL	101.95	9
29_0.03ng	D3S1358	B	OL	113.26	10
29_0.03ng	D3S1358	B	OL	130.44	6
29_0.03ng	D3S1358	B	17.2	133.66	4
29_0.03ng	D1S1656	B	OL	155.95	9
29_0.03ng	D1S1656	B	OL	158.42	8
29_0.03ng	D1S1656	B	OL	162.86	7
29_0.03ng	D1S1656	B	OL	164.83	8
29_0.03ng	D1S1656	B	OL	175.23	7
29_0.03ng	D1S1656	B	OL	179.7	7
29_0.03ng	D1S1656	B	OL	200.71	3
29_0.03ng	D2S441	B	OL	210.62	8
29_0.03ng	D2S441	B	OL	222.82	8
29_0.03ng	D2S441	B	OL	227.82	8
29_0.03ng	D2S441	B	OL	232.82	8
29_0.03ng	D2S441	B	OL	237.82	8
29_0.03ng	D10S1248	B	OL	261.41	6
29_0.03ng	D10S1248	B	16	287.47	7
29_0.03ng	D10S1248	B	OL	290.03	7
29_0.03ng	D10S1248	B	OL	294.63	7
29_0.03ng	D13S317	B	OL	307.63	8
29_0.03ng	D13S317	B	OL	310.92	8

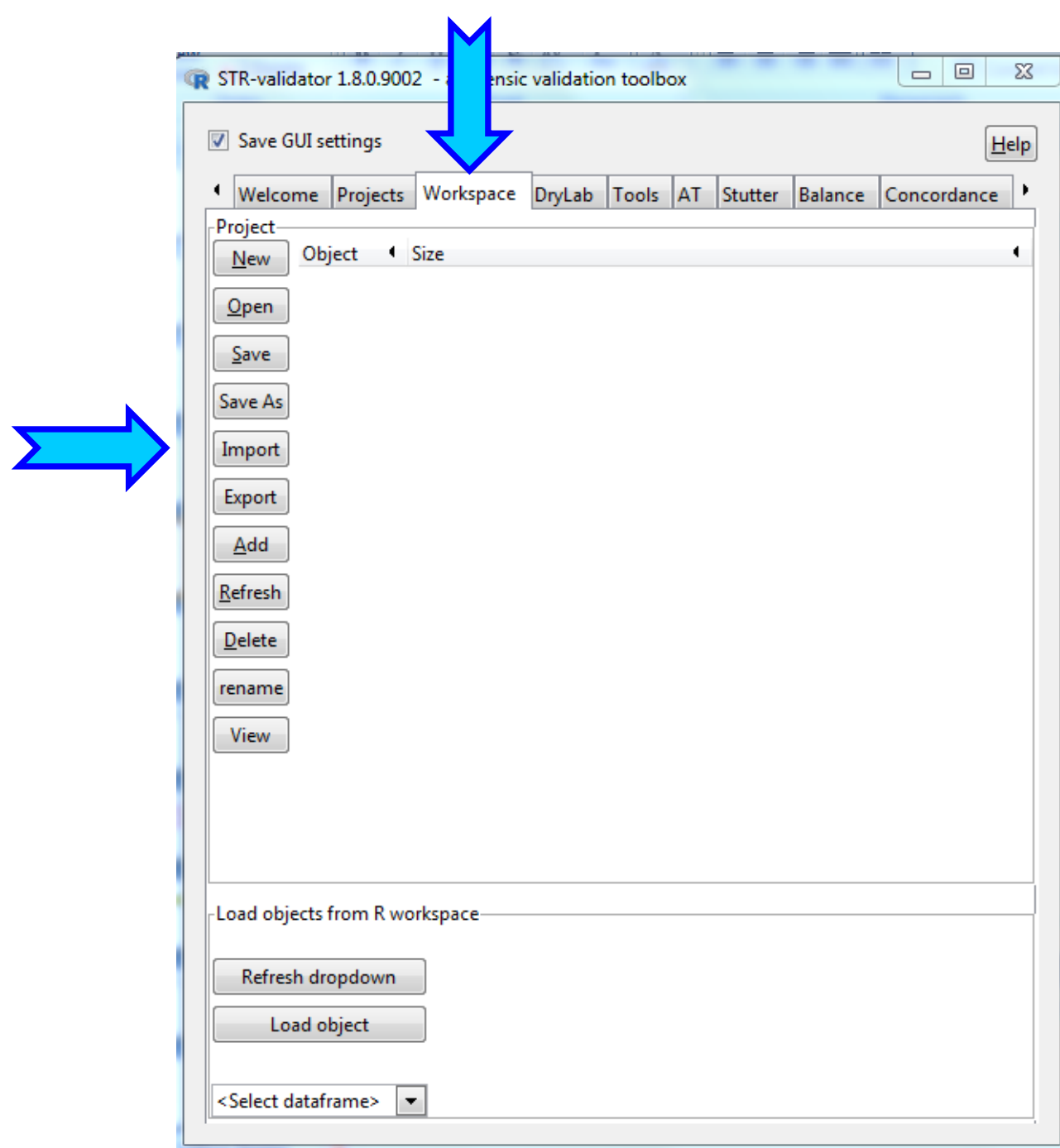
Semi-long narrow type of table Format = *Slim or Stacked data*

- ❖ **Trim**: removing unwanted samples and/or columns
- ❖ **Slim**: transforming files from GeneMapper format into STR-validator format



Open a New Workspace in *STR-validator* GUI and save as *Name.RData* (e.g. *Trim_Slim_Analysis*)

Import DataSet



Import DataSet

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset

Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Precision_Analysis\L... browse

Select a directory... browse

Options

Save file name

Save file time stamp

Delimiter:

TAB

NA strings (separated by comma):

NA,

Auto trim samples

Auto slim repeated columns

+ Multiple files options

+ Trim options

+ Slim options

Save options

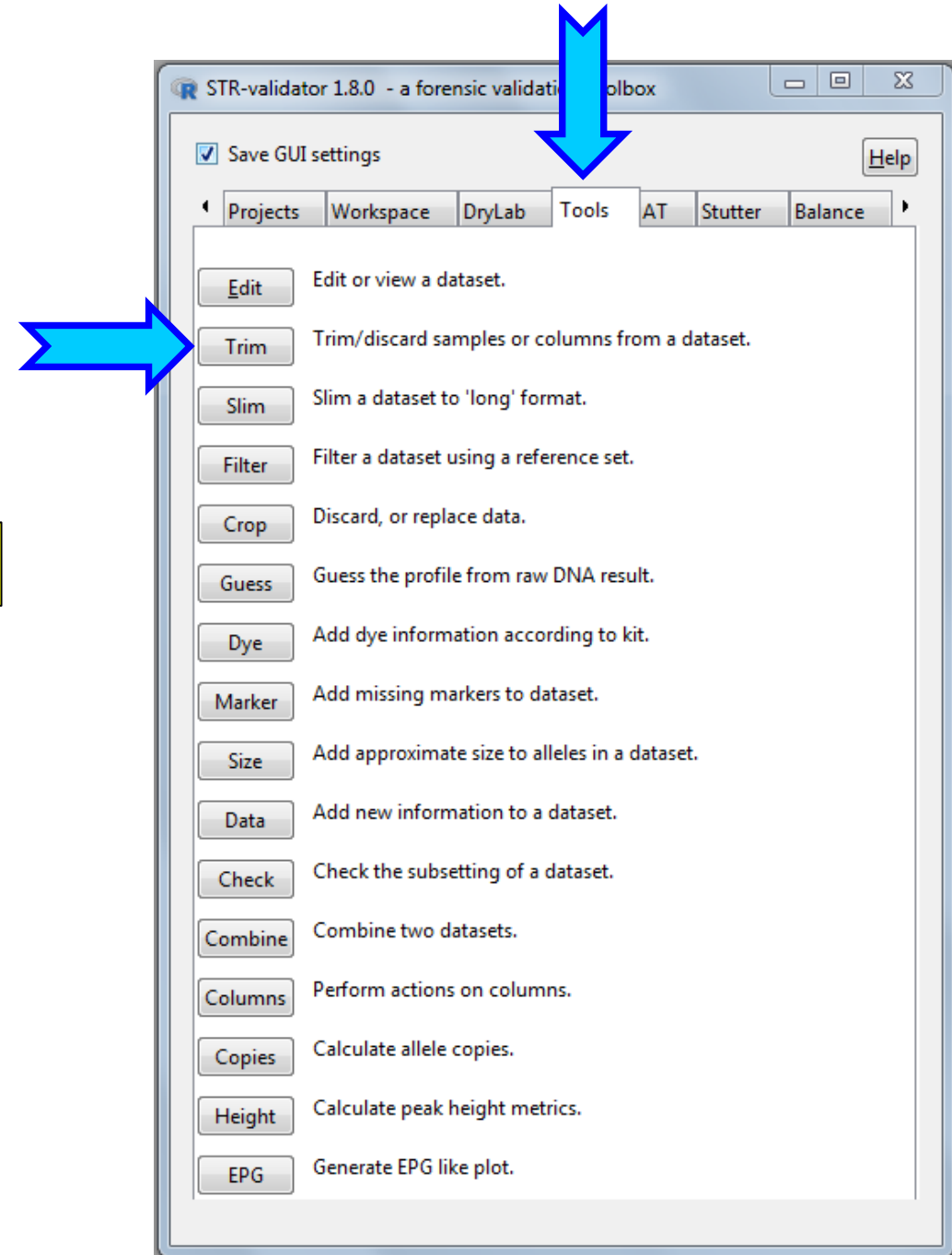
Name:

Set1

Import

Perform Manual Trimming

Trim function removes unwanted samples and columns



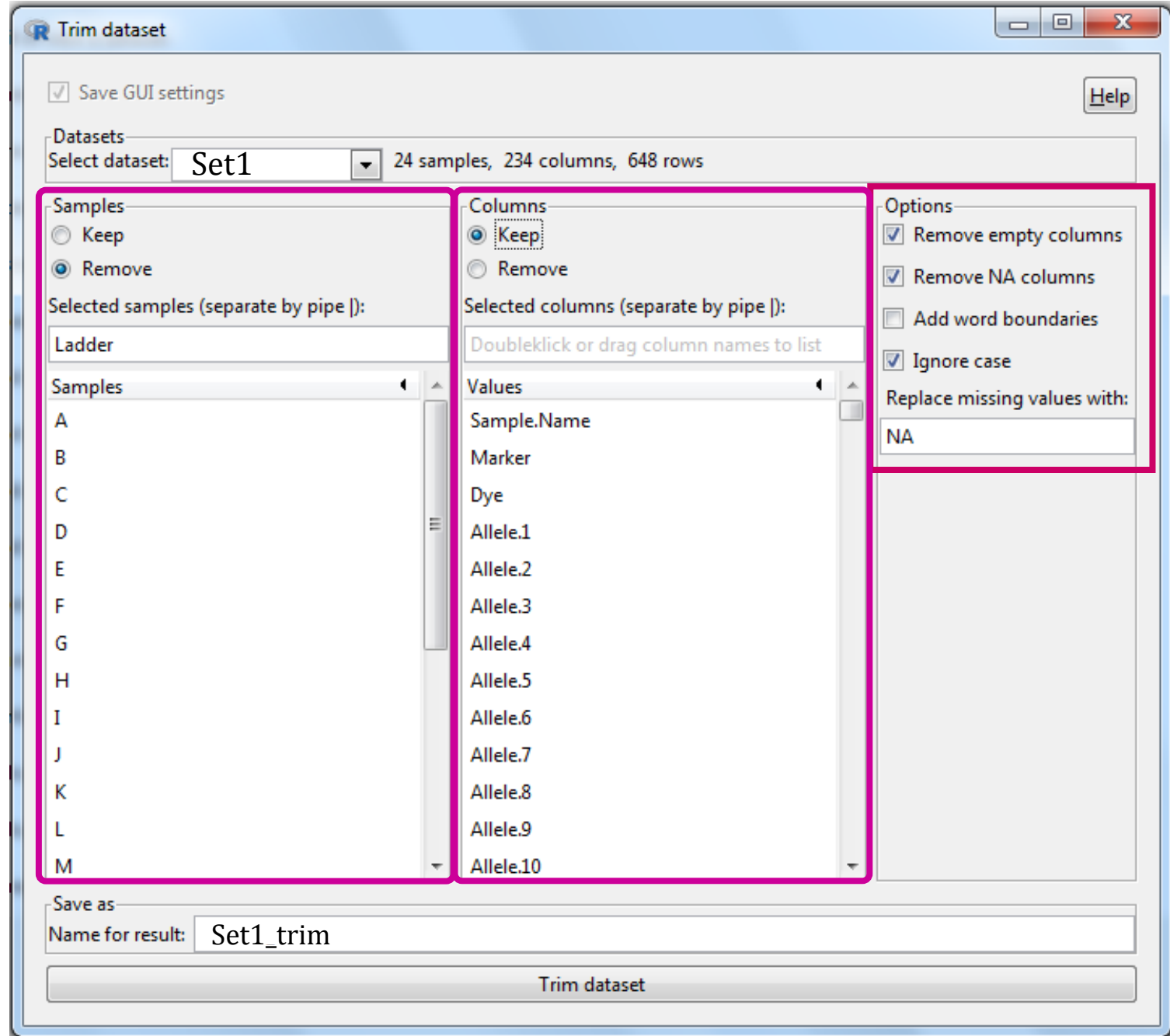
Trim the Ladder

Keep or Remove Sample(s) in the sample frame.

Keep or Remove Column(s) in the Column frame.

A pipe (|) is used for separation.

Double Click on the sample/column you wish to remove or keep.



View the Trimmed Dataset

Save GUI settings

Help

Datasets

Select dataset: Set1_trim 23 samples, 234 columns, 621 rows

Options

Show attributes (separate window)

Limit number of rows to: 100

Copy | Export | Save

Copy Export Save as Set1_trim

Data frame

Sample.Name	Marker	Dye	Allele.1	Allele.2	Allele.3	Allele.4	Allele.5	Allele.6	Allele.7	Allele.8	Allele.9	Allele.10	Allele.11	Allele.12	Allele.13	Allele.14	Allele.15	Allele.16	Allele.17
A	AMEL	B	X	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A	D3S1358	B	9	10	11	12	13	14	15	16	17	18	19	20	21	NA	NA	NA	NA
A	D1S1656	B	8	9	10	11	12	13	OL	14	14.3	15	15.3	16	16.3	17	17.3	18	18.3
A	D2S441	B	8	9	10	11	11.3	12	13	14	15	16	17	NA	NA	NA	NA	NA	NA
A	D10S1248	B	OL	8	9	10	11	12	13	14	15	16	17	18	19	NA	NA	NA	NA
A	D13S317	B	5	6	7	8	9	10	11	12	13	14	15	16	17	NA	NA	NA	NA
A	Penta E	B	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	D16S539	G	4	5	6	7	8	9	9.3	10	11	12	13	14	15	16	NA	NA	NA
A	D18S51	G	OL	7	OL	8	OL	9	9.2	10	10.2	11	11.2	12	12.2	13	13.2	14	14.2
A	D2S1338	G	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	CSF1PO	G	5	6	7	8	9	10	11	12	13	14	15	16	OL	NA	NA	NA	NA
A	Penta D	G	2.2	3.2	5	6	7	8	9	10	11	12	13	14	15	16	17	18	NA
A	TH01	Y	3	4	5	6	7	7.3	8	9	9.3	10	11	OL	13.3	NA	NA	NA	NA
A	vWA	Y	10	11	12	OL	13	OL	14	15	16	17	18	19	20	21	22	23	24
A	D21S11	Y	OL	OL	23.2	24	24.2	25	25.2	26	26.2	27	27.2	28	28.2	29	29.2	30	30.2
A	D7S820	Y	5	6	7	8	9	10	11	12	13	14	15	16	NA	NA	NA	NA	NA
A	D5S818	Y	OL	6	7	8	9	10	11	12	13	14	15	16	17	18	19	NA	NA
A	TPOX	Y	OL	4	OL	5	6	7	8	9	10	11	12	13	14	15	16	OL	NA
A	D8S1179	R	6	7	8	9	10	11	12	13	14	15	16	17	18	19	NA	NA	NA
A	D12S391	R	13	14	15	16	17	17.3	18	18.3	19	20	21	22	23	24	25	26	27
A	D19S433	R	5.2	6.2	7	8	8.2	9	10	10.2	11	11.2	12	12.2	13	13.2	14	14.2	15
A	SE33	R	OL	OL	4.2	5	OL	OL	6.3	OL	8	8.2	9	9.2	10	10.2	11	11.2	12

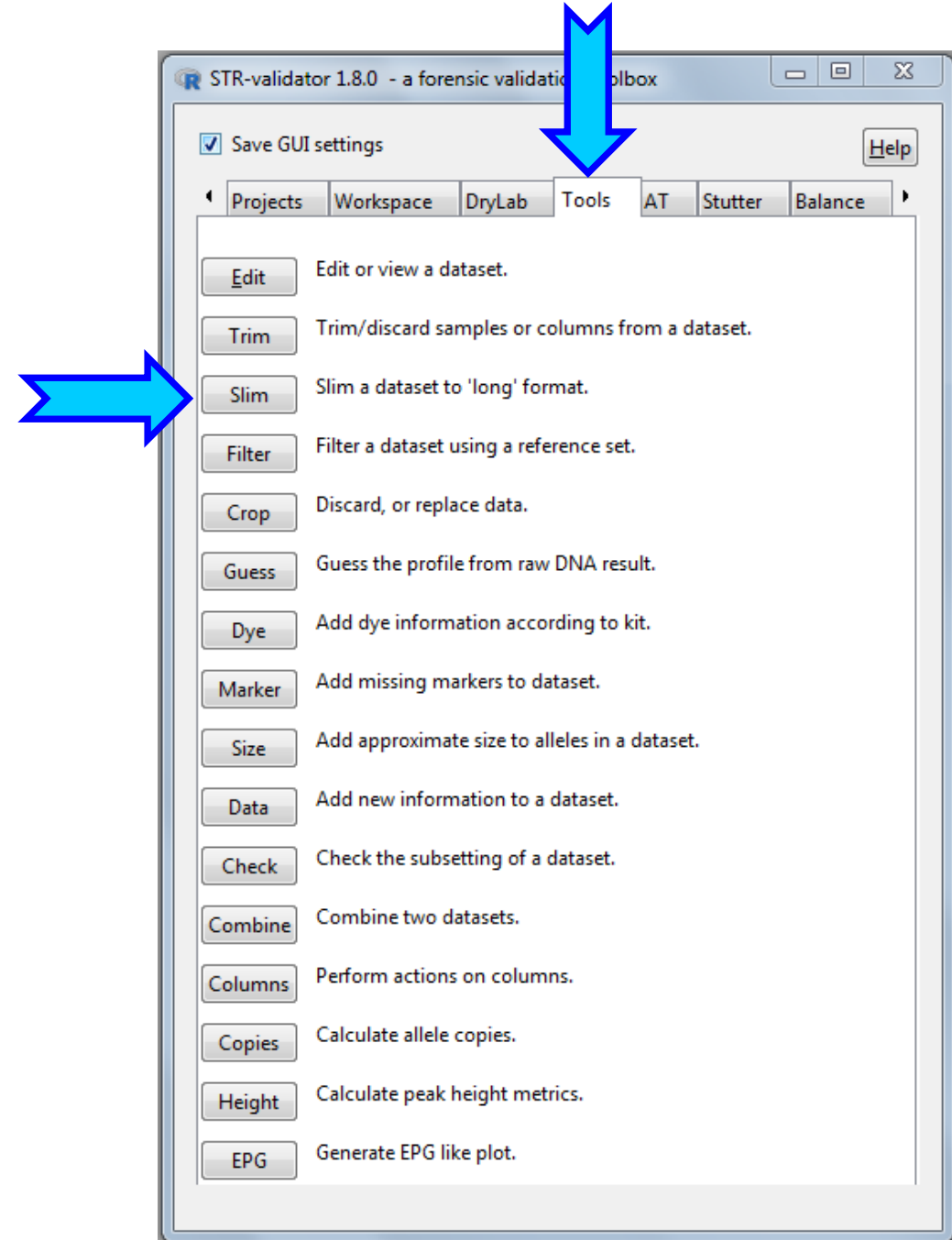
- ❖ Semi-Wide type of table Format = *Unstacked Data*
- ❖ Ladder is removed

Perform Manual Slimming

GeneMapper semi-wide type of table format

Slim function

STR-validator semi-long type of data frame



Slim a Dataset



R Slim dataset

Save GUI settings Help

Datasets
Select dataset: 23 samples, 234 columns, 621 rows

Options
 Keep rows in fixed columns even if no data in stacked columns
(i.e. keep one row per marker for each sample even if no peak)

Tip: Manually edit the columns to fix and stack.
e.g. 'Allele' will stack 'Allele.1', 'Allele.2'...

Fix	Stack
Columns to keep fixed (separate by pipe):	Columns to stack (separate by pipe):
<input type="text" value="Sample.Name Marker Dye"/>	<input type="text" value="Allele Size Height"/>
Values	Values
Sample.Name	Sample.Name
Marker	Marker
Dye	Dye
Allele.1	Allele.1
Allele.2	Allele.2
Allele.3	Allele.3
Allele.4	Allele.4
Allele.5	Allele.5
Allele.6	Allele.6
Allele.7	Allele.7

Save as
Name for result:

Slim dataset

View the Slimmed Dataset

Save GUI settings

Datasets
Select dataset: Ladders_trim_slim 23 samples, 6 columns, 12389 rows

Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
 Ladders_trim_slim

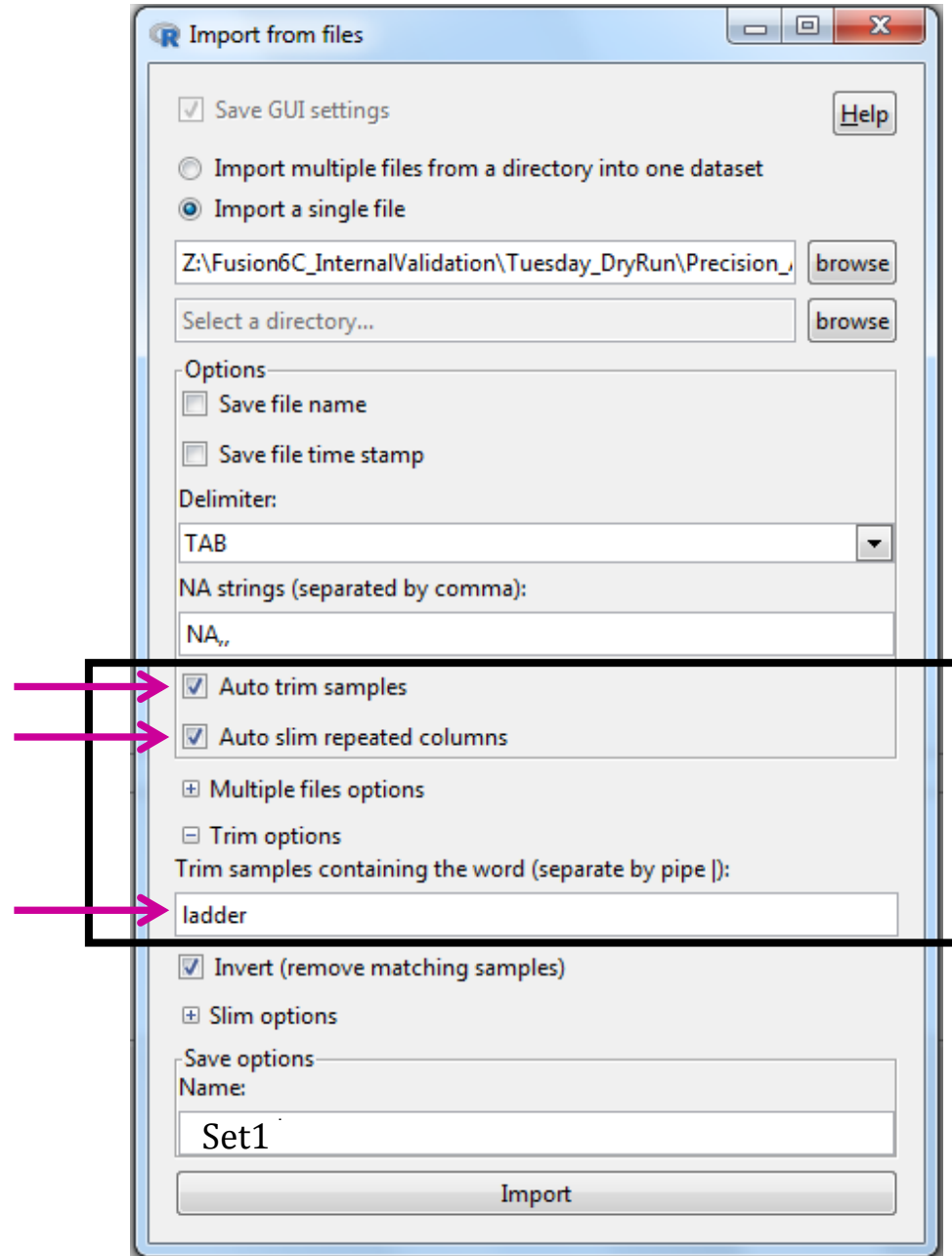
Data frame

Sample.Name	Marker	Dye	Allele	Size	Height
A	AMEL	B	X	81.31	2012
A	AMEL	B	Y	87.44	1898
A	D3S1358	B	9	97.61	1810
A	D3S1358	B	10	101.96	1834
A	D3S1358	B	11	106.27	1847
A	D3S1358	B	12	110.6	1866
A	D3S1358	B	13	114.94	1856
A	D3S1358	B	14	119.3	1860
A	D3S1358	B	15	123.53	1913
A	D3S1358	B	16	127.84	1918
A	D3S1358	B	17	132.05	1894
A	D3S1358	B	18	136.27	1941
A	D3S1358	B	19	140.49	2036
A	D3S1358	B	20	144.64	2190
A	D3S1358	B	21	148.7	75
A	D1S1656	B	8	154.15	101
A	D1S1656	B	9	158.21	1863
A	D1S1656	B	10	162.37	1873
A	D1S1656	B	11	166.42	1859
A	D1S1656	B	12	170.58	1899
A	D1S1656	B	13	174.64	1938
A	D1S1656	B	OL	176.72	54
A	D1S1656	B	14	178.71	1977
A	D1S1656	B	14.3	181.79	1939
A	D1S1656	B	15	182.79	2003
A	D1S1656	B	15.3	185.89	1936
A	D1S1656	B	16	186.79	1949
A	D1S1656	B	16.3	189.9	1963
A	D1S1656	B	17	190.91	2009
A	D1S1656	B	17.3	194.03	1957
A	D1S1656	B	18	194.04	1996

❖ Semi-long narrow type of table Format = *Slim or Stacked data*

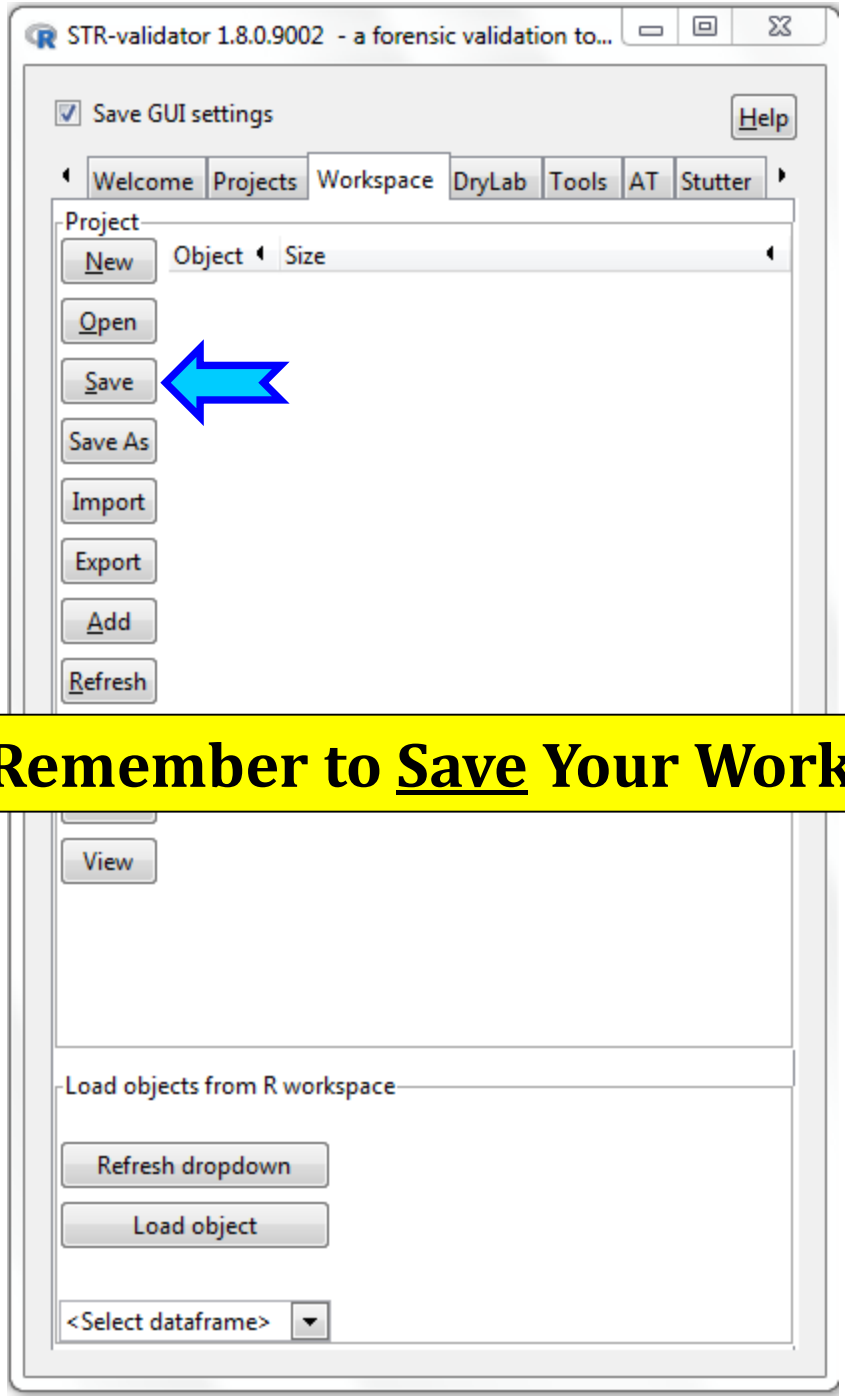
❖ *STR-validator* format

Automatic Trimming and Slimming in STR-validator





Remember to **Slim** your *txt.files* either Manually or Automatically in *STR-validator*



Remember to Save Your Workspace

Precision Analysis

SWGDM Guidelines

Precision

- ❖ Characterizes the degree of mutual agreement among a series of individual measurements/values and results.
- ❖ Depends only on the distribution of random errors and does not relate to the true value or specified value.
- ❖ Is usually expressed in terms of imprecision and computed as a standard deviation of the test results.

How to measure the precision of your instrument?

All measured alleles should fall within a ± 0.5 bp window around the measured size for the corresponding allele in the allelic ladder.

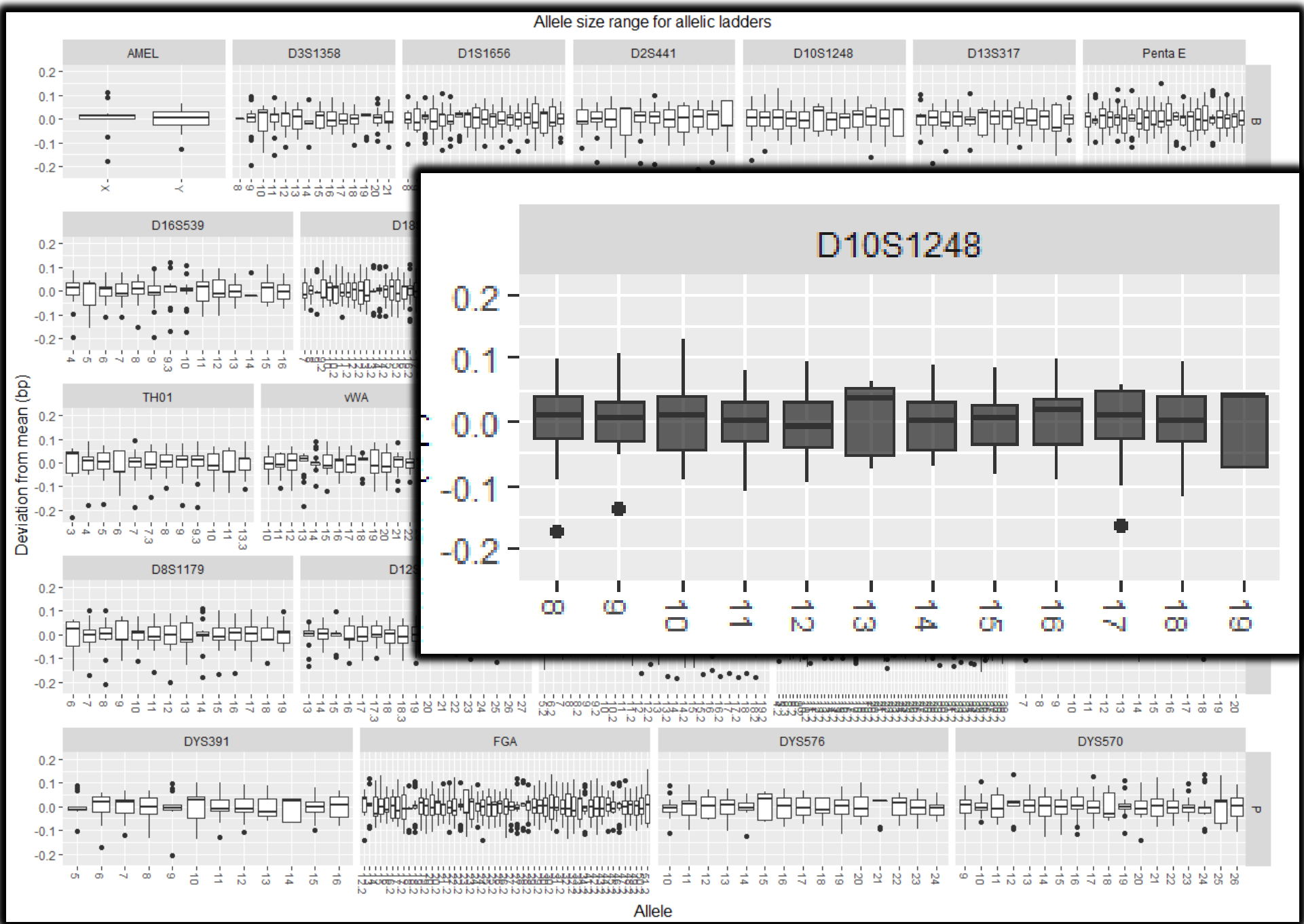
Experimental Procedure for Precision Analysis

One injection of 24 ladders performed

- ❖ 1 ladder assigned as the “ladder”
- ❖ 22 ladders assigned as samples (A-V)
- ❖ Analyzed at your Analytical Threshold (AT)
- ❖ Export **_GenotypeTable.txt** from GeneMapper with at least the following information: **“Sample.Name”, “Marker”, “Allele” and “Size”**.

	10	11	12
A	Ladder	Ladder	Ladder
B	Ladder	Ladder	Ladder
C	Ladder	Ladder	Ladder
D	Ladder	Ladder	Ladder
E	Ladder	Ladder	Ladder
F	Ladder	Ladder	Ladder
G	Ladder	Ladder	Ladder
H	Ladder	Ladder	Ladder

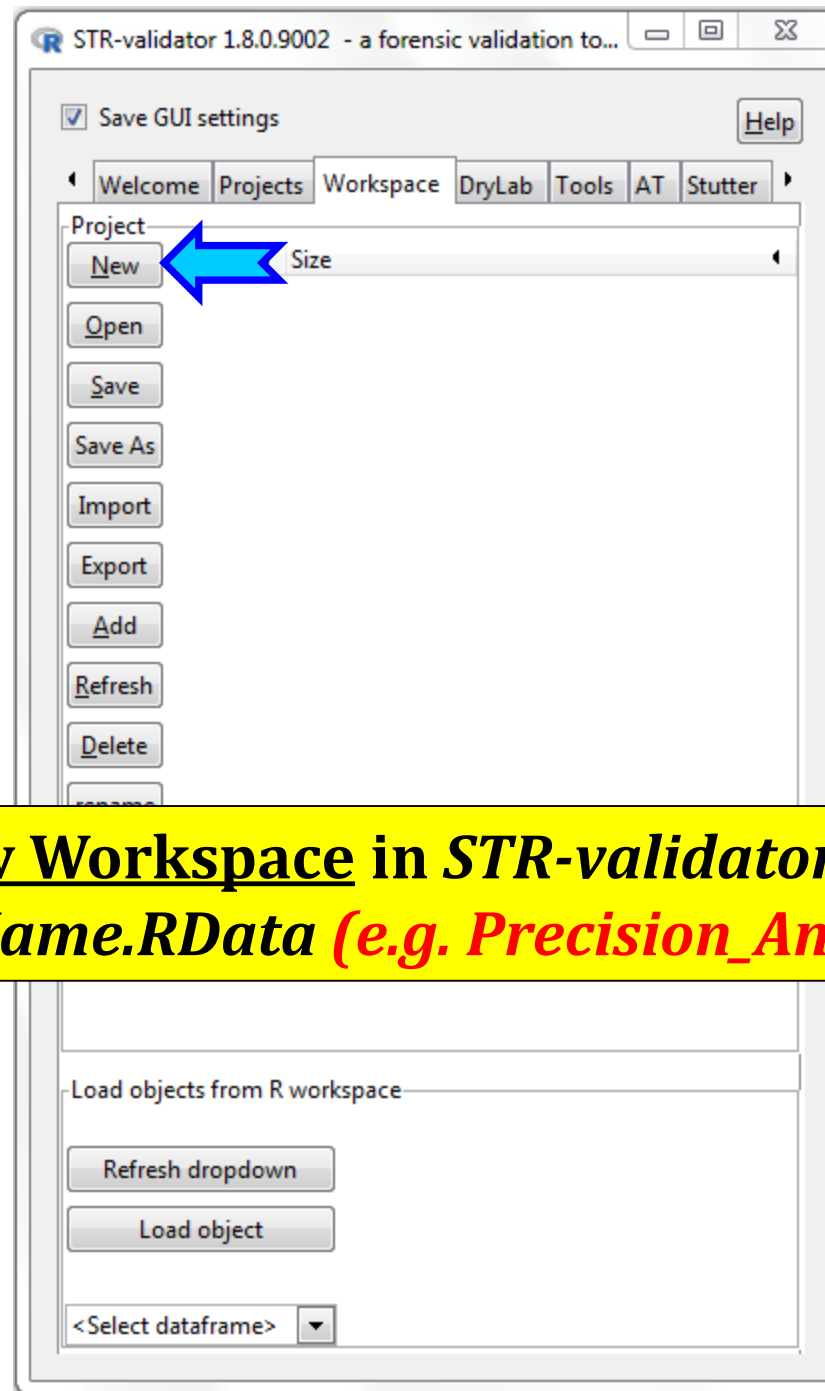
How to Plot Size Precision for the Allelic Ladders?



How to get a Summary on Statistics of Precision?

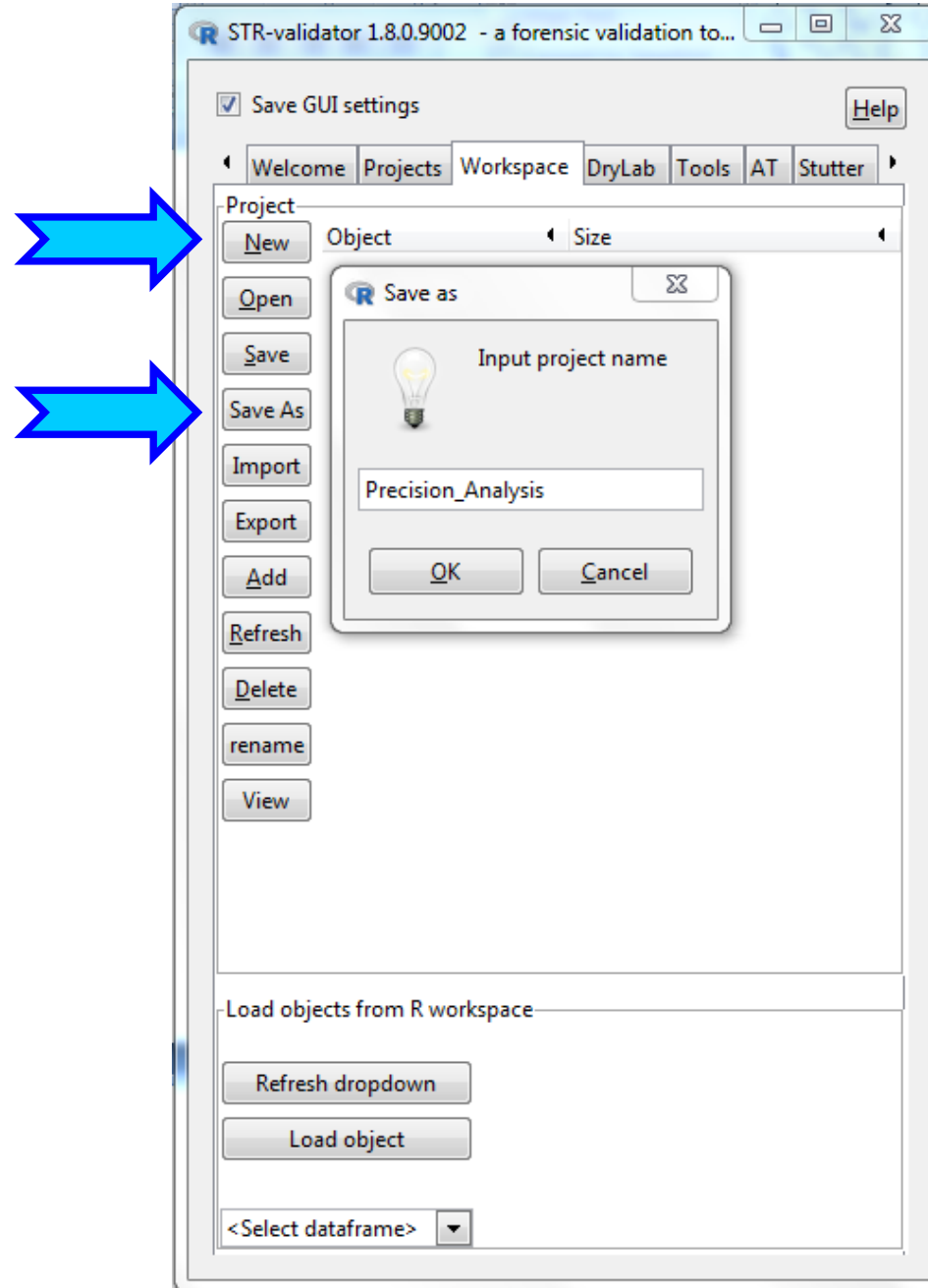
Marker	Allele	Size.Min	Size.Max	Size.Mean	Size.n	Size.Sd
AMEL	X	81.12	81.41	81.29783	23	0.05807178
AMEL	Y	87.34	87.53	87.46565	23	0.04315154
D3S1358	9	97.41	97.7	97.60652	23	0.06012175
D3S1358	10	101.78	101.98	101.9239	23	0.05408236
D3S1358	11	106.13	106.37	106.2813	23	0.04883221
D3S1358	12	110.51	110.7	110.6283	23	0.04858471
D3S1358	13	114.89	115.05	114.9796	23	0.0470472
D3S1358	14	119.2	119.4	119.3191	23	0.04860911
D3S1358	15	123.53	123.65	123.577	23	0.03807108
D3S1358	16	127.74	127.9	127.8091	23	0.04176918
D3S1358	17	131.96	132.1	132.0287	23	0.03696489
D3S1358	18	136.18	136.35	136.2909	23	0.04230979
D3S1358	19	140.39	140.5	140.477	23	0.03993572
D3S1358	20	144.56	144.74	144.6539	23	0.04075862
D1S1656	9	158.2	158.33	158.2578	23	0.0481434
D1S1656	10	162.26	162.45	162.3609	23	0.04198626
D1S1656	11	166.35	166.56	166.46	23	0.05485518
D1S1656	12	170.43	170.67	170.5604	23	0.05182824
D1S1656	13	174.61	174.82	174.7235	23	0.04904624
D1S1656	14	178.69	178.82	178.7835	23	0.04468157
D1S1656	14.3	181.79	181.91	181.8717	23	0.04130289
D1S1656	15	182.71	182.91	182.8465	23	0.05482275
D1S1656	15.3	185.83	186.03	185.9448	23	0.04897769
D1S1656	16	186.79	186.94	186.8796	23	0.04279744
D1S1656	16.3	189.86	190.07	189.9787	23	0.04836456
D1S1656	17	190.87	191.02	190.963	23	0.04016271
D1S1656	17.3	193.91	194.11	194.0439	23	0.04599794
D1S1656	18	194.92	195.1	195.0135	23	0.04029535
D1S1656	18.3	197.97	198.2	198.1057	23	0.06140966

Marker	Allele	Size.Min	Size.Max	Size.Mean	Size.n	Size.Sd
Penta E	7	378.31	378.49	378.4178	23	0.05518206
Penta E	8	383.24	383.4	383.3135	23	0.05131088
Penta E	9	388.1	388.32	388.1939	23	0.04821723
Penta E	10	393	393.18	393.097	23	0.05563148
Penta E	11	397.87	398.11	397.9904	23	0.04733615
Penta E	12	402.73	402.9	402.81	23	0.0533428
Penta E	13	407.48	407.68	407.5852	23	0.05221194
Penta E	14	412.29	412.47	412.3757	23	0.04439759
Penta E	15	417.09	417.3	417.1513	23	0.05198814
Penta E	16	421.8	422.04	421.9465	23	0.06328929
Penta E	17	426.71	426.85	426.8183	23	0.04365695
Penta E	18	431.65	431.87	431.7743	23	0.04294035
Penta E	19	436.61	436.82	436.7287	23	0.0586441
Penta E	20	441.59	441.77	441.6848	23	0.05476143
Penta E	21	446.6	446.78	446.667	23	0.05795255
Penta E	22	451.47	451.7	451.5804	23	0.05716629
Penta E	23	456.29	456.48	456.3865	23	0.05296729
Penta E	24	461.12	461.3	461.1974	23	0.0510986
Penta E	25	465.88	466.07	465.9822	23	0.04738206

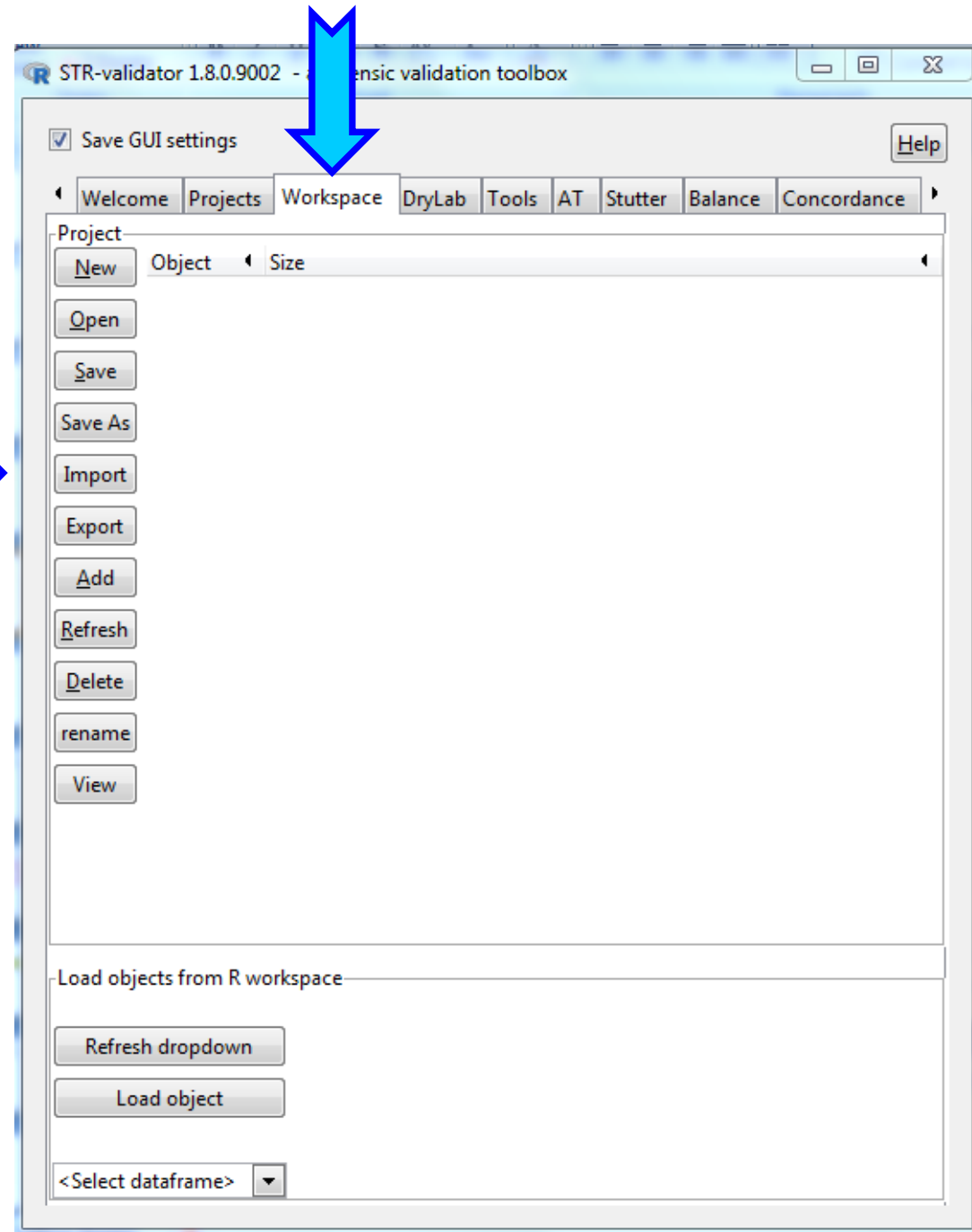
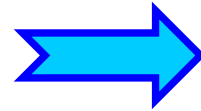


Open a New Workspace in *STR-validator* GUI and save as *Name.RData (e.g. Precision_Analysis)*

Open a New Workspace, Name and Save it



Import Ladder DataSet



Import Ladder DataSet

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset

Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Precision_Analysis\L browse

Select a directory... browse

Options

Save file name

Save file time stamp

Delimiter:

TAB

NA strings (separated by comma):

NA,,

Auto trim samples

Auto slim repeated columns

+ Multiple files options

+ Trim options

+ Slim options

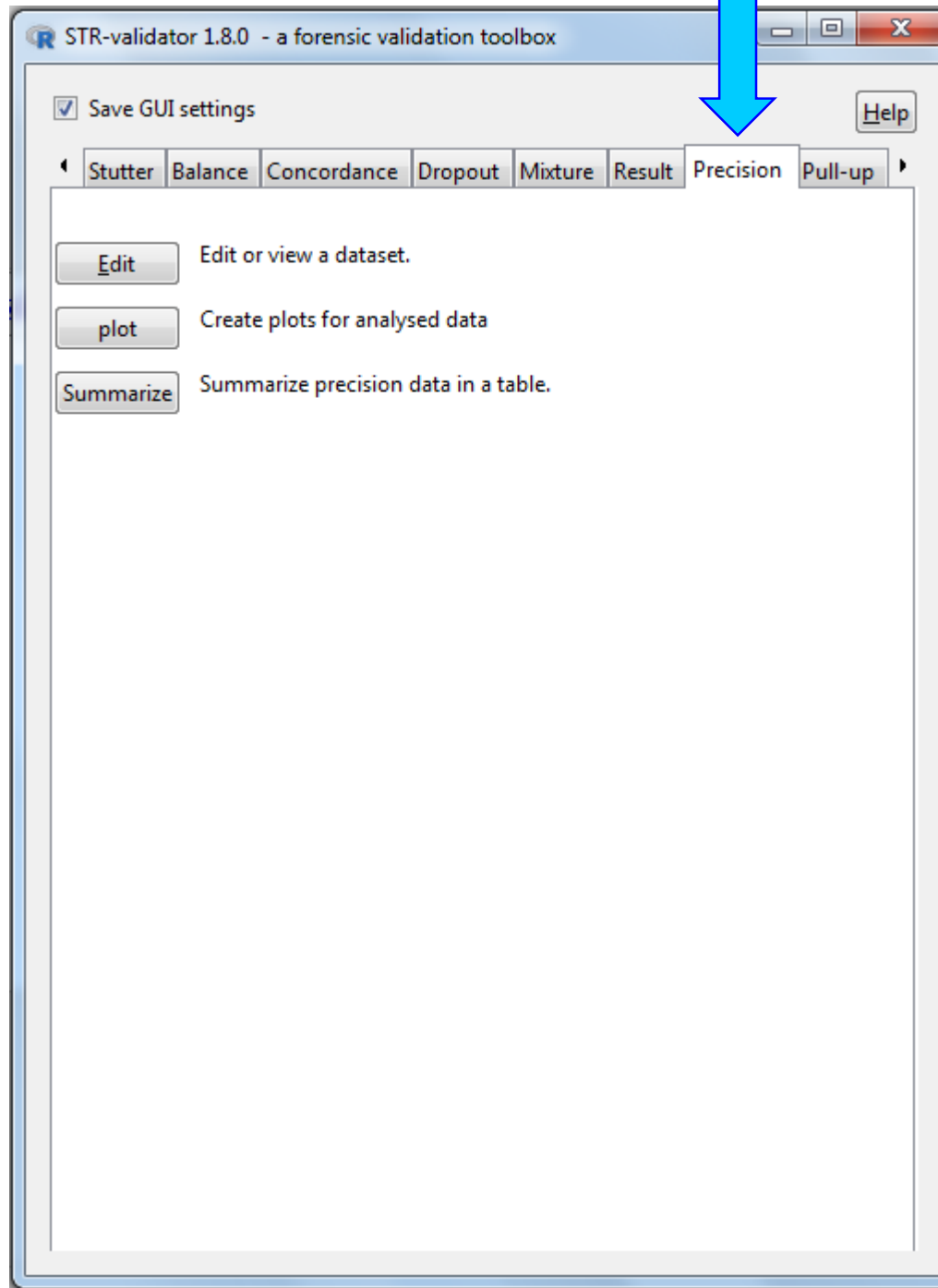
Save options

Name:

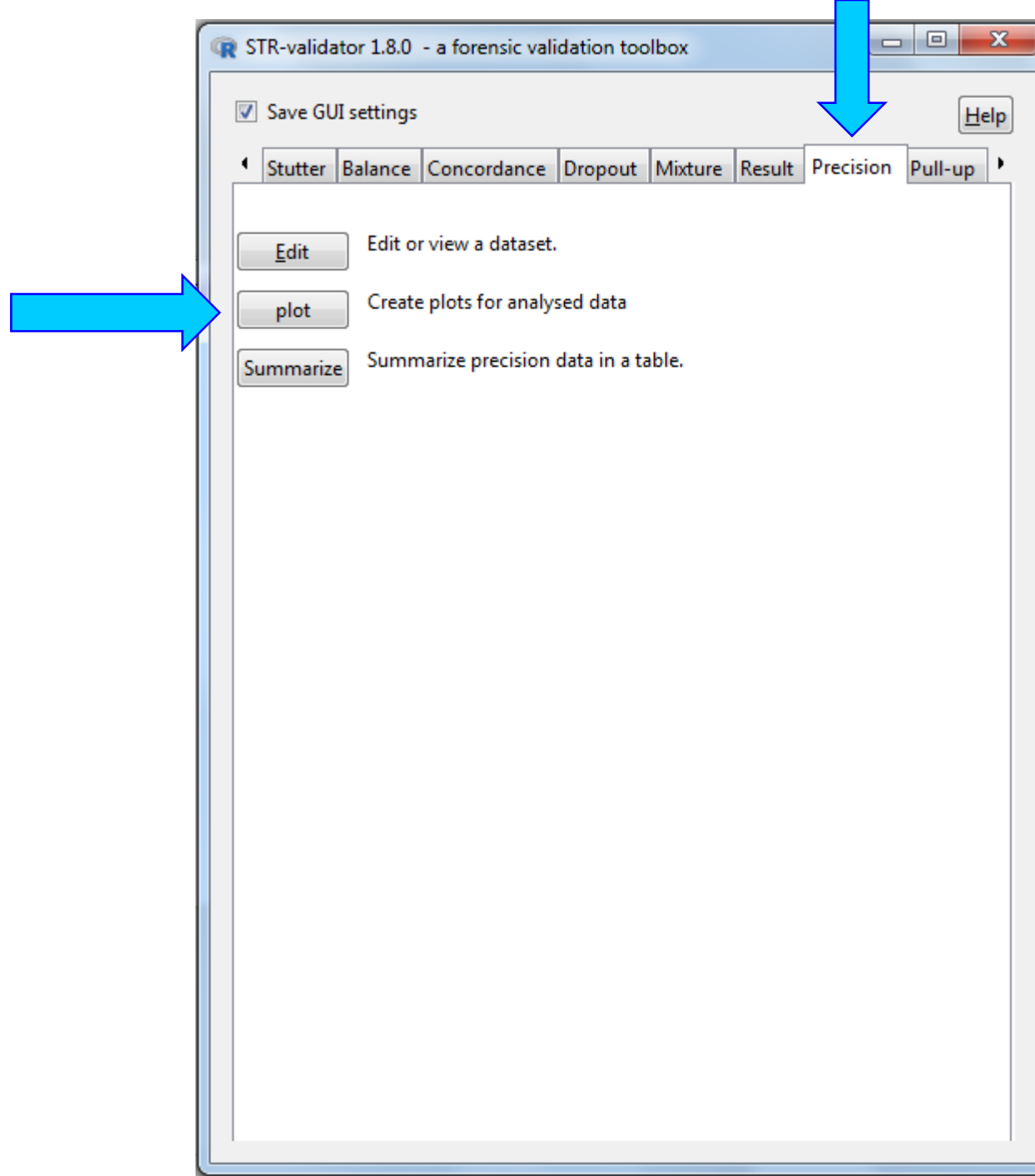
Ladders

Import

Precision Tab



Plot Precision



Plot Precision



Plot precision ⏏ 🖼 🔍

Save GUI settings Help

Dataset and kit
Select dataset: Ladders and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Plot per marker

X axis: Mean Allele

Plot theme: theme_grey0

Data points

Axes

X labels

Plot precision data as dotplot

Size Height Data point

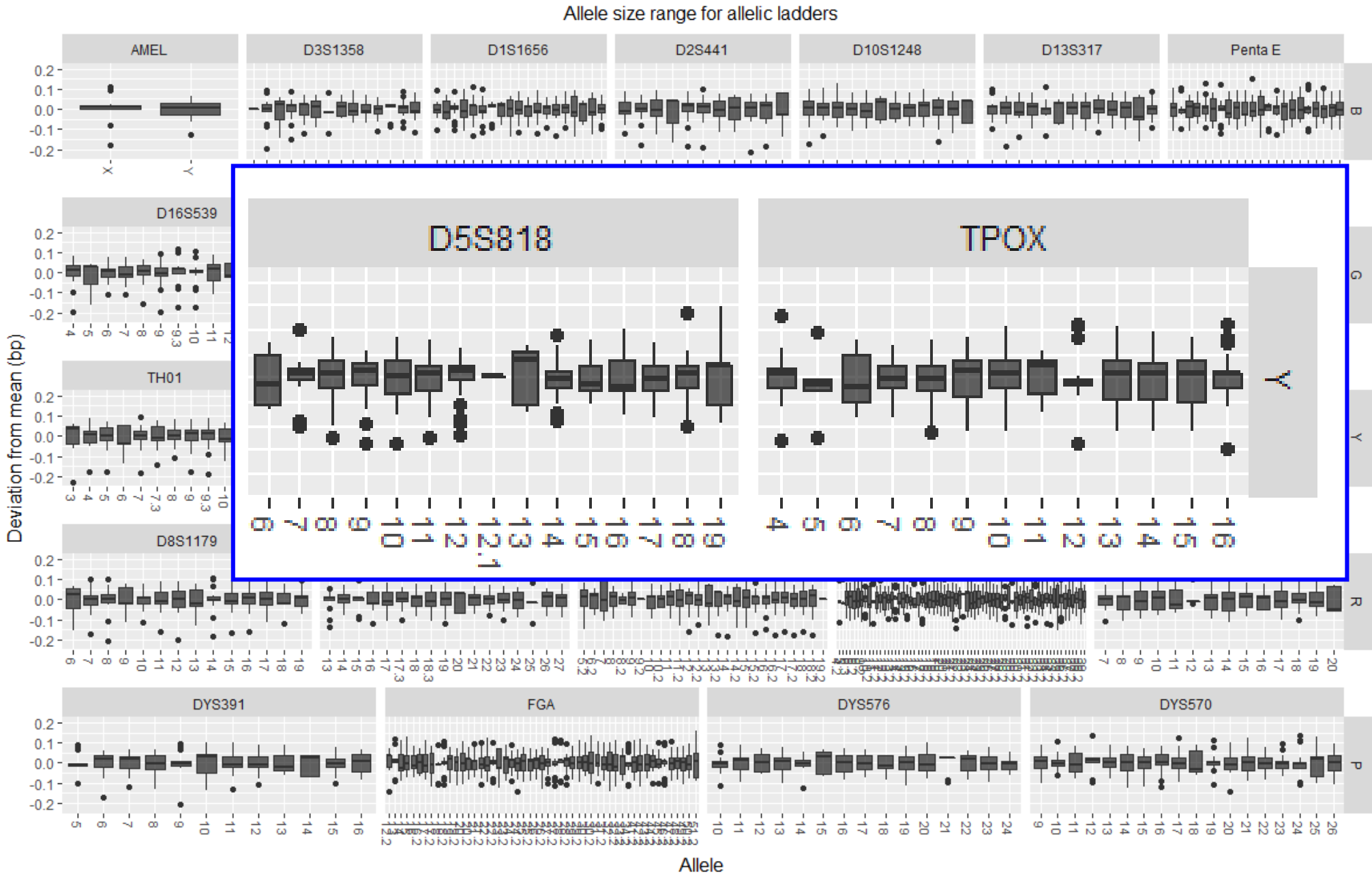
Plot precision data as boxplot

Size Height Data point

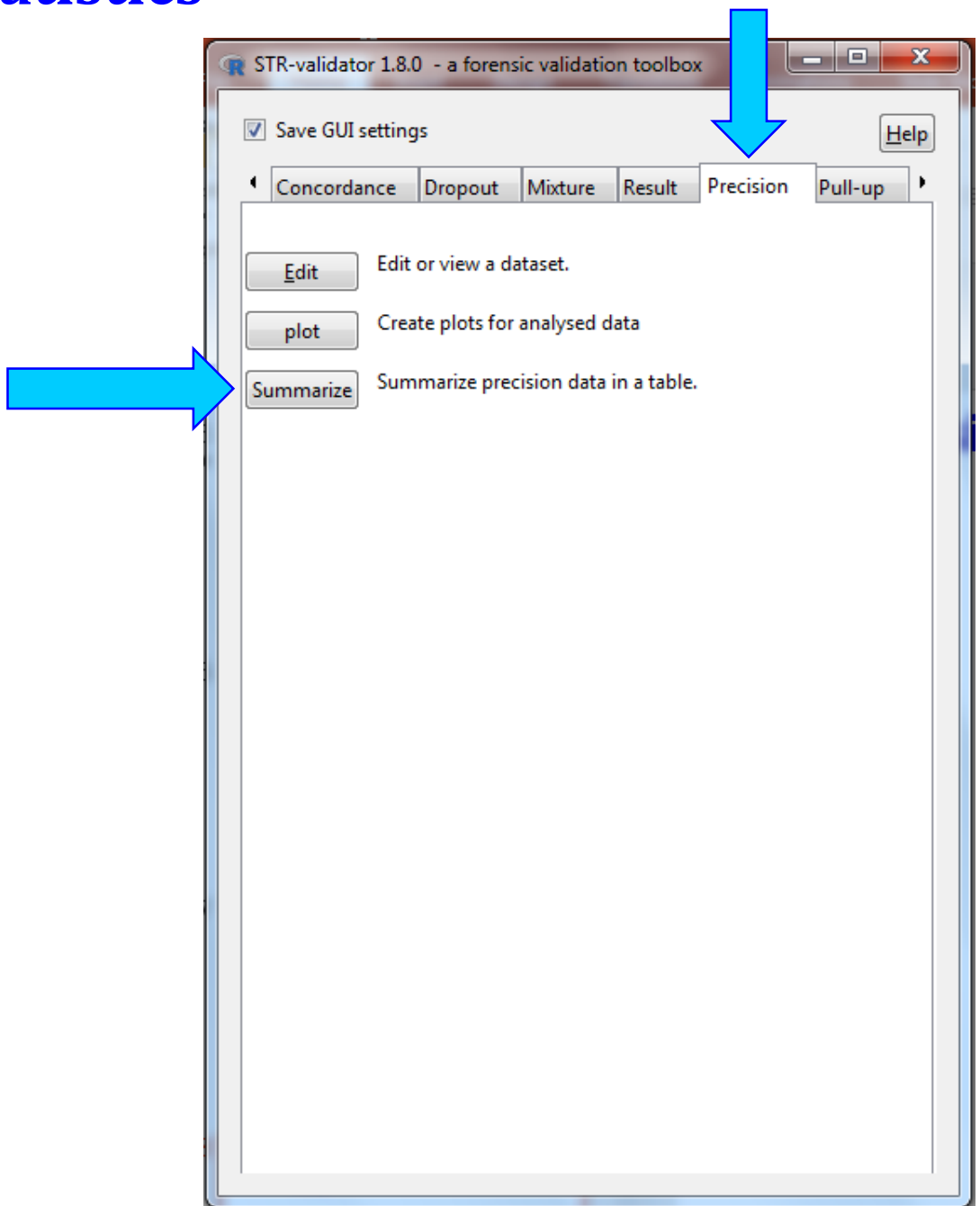
Save as

Name for result: Ladders_ggplot Save as object Save as image

Size Precision Boxplot for the Allelic Ladders by Allele



Calculate Summary Statistics



Calculate Summary Statistics for Precision



Calculate summary statistics for precision

Save GUI settings Help

Datasets
Select dataset: Ladders 23 samples.

Filter
 Filter by reference dataset
 Filter by kit bins
 Do not filter

Select reference dataset:
<Select a dataset> 0 references
Check subsetting

Select kit: Fusion 6C Exclude virtual bins.

Options
 Ignore case

Create key from columns
Marker,Allele

Values
Sample.Name
Dye
Size

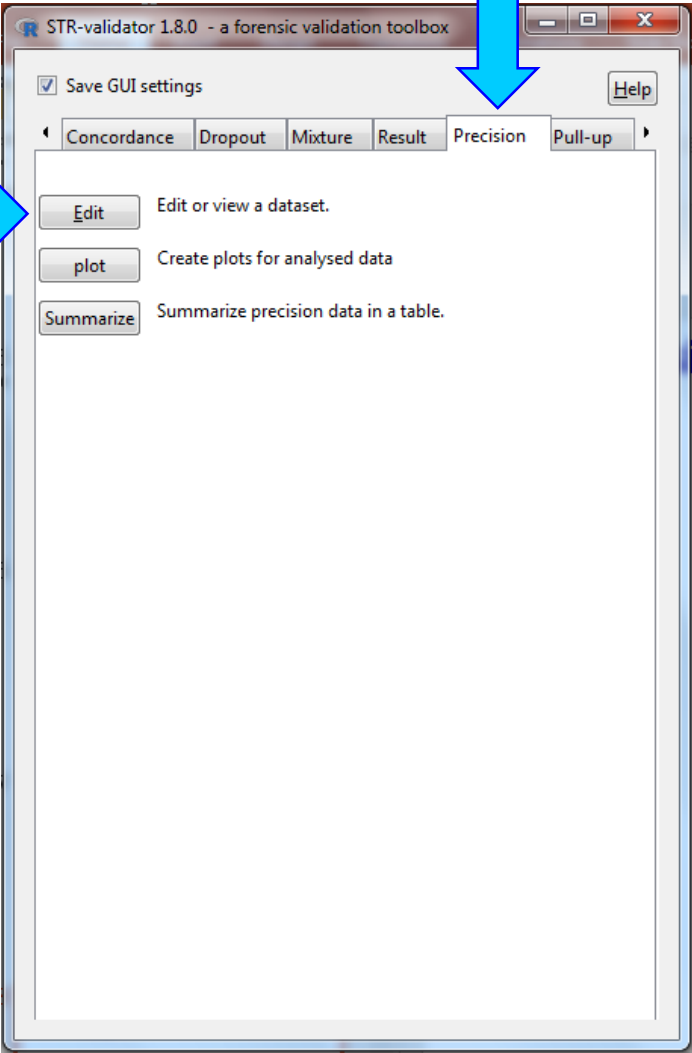
Calculate precision for target columns
Size

Values
Sample.Name
Marker
Dye

Save as
Name for result: Ladders_precision_table

Calculate

Go to Summary Statistics for Precision and Sort "Size.Sd" by Descending Order



Edit or view data frame

Save GUI settings

Datasets
Select dataset: Ladders_precision_table <NA> samples, 7 columns, 433 rows

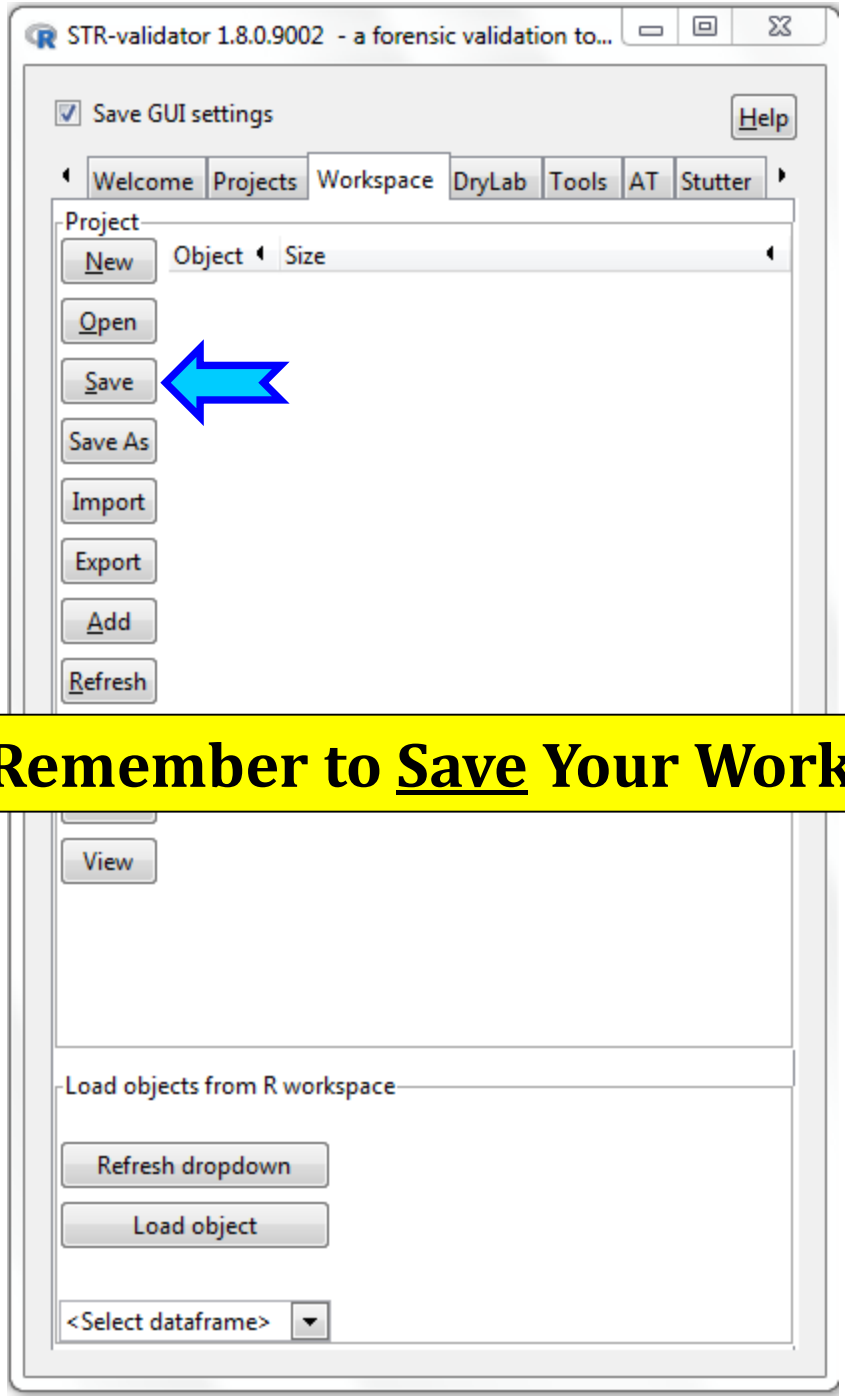
Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
Copy Export Save as Ladders_precision_table_edit

Row.names	Marker	Allele	Size.Min	Size.Max	Size.Mean	Size.n	Size.Sd
168	Penta D	16	448.02	448.27	448.1609	23	0.06374666
170	TH01	3	66.64	66.93	66.8713	23	0.06374666
67	D13S317	16	350.56	350.79	350.7187	23	0.06374666
169	Penta D	17	453.14	453.39	453.24	23	0.06388911
307	D19S433	18.2	246.36	246.61	246.54	23	0.06374666
308	SE33	4.2	274.68	275	274.7935	23	0.06364737
95	D16S539	9	97.7	97.99	97.89696	23	0.06363495
80	Penta E	16	421.8	422.04	421.9465	23	0.06328929
36	D2S441	11	224.69	224.9	224.8574	23	0.06290089
264	D8S1179	12	93.5	93.79	93.7	23	0.06259538
131	D2S1338	17	251.19	251.4	251.3317	23	0.06198623
357	D22S1045	20	474.5	474.63	474.56	22	0.06187545
342	SE33	37	406.39	406.61	406.5004	23	0.06153183
29	D1S1656	18.3	197.97	198.2	198.1057	23	0.06140966
260	D8S1179	8	76.26	76.57	76.46739	23	0.06121626
346	D22S1045	9	441.09	441.28	441.1883	23	0.06102549
331	SE33	26.2	363.91	364.15	364.0104	23	0.06093799
306	D19S433	18	244.33	244.58	244.4904	23	0.06086335
204	D21S11	29	223.85	224.08	223.9974	23	0.06084387
217	D21S11	35.2	250.43	250.65	250.5726	23	0.06046918
44	D10S1248	8	256.16	256.43	256.3335	23	0.06027277
3	D3S1358	9	97.41	97.7	97.60652	23	0.06012175
132	D2S1338	18	255.29	255.57	255.4326	23	0.06009216
177	TH01	9.3	95.69	95.97	95.87913	23	0.05984169
191	vWA	20	166.75	166.95	166.8678	23	0.05976899
41	D2S441	15	241.14	241.43	241.3517	23	0.05974584
37	D2S441	11.3	227.72	227.99	227.9048	23	0.05968627
205	D21S11	29.2	225.84	226.05	225.9939	23	0.05967633
325	SE33	21.2	343.69	343.89	343.8039	23	0.05960012
135	D2S1338	21	267.62	267.82	267.7291	23	0.05946069

Apply function to column
Sort by column (decreasing)
Sort by column (increasing)
Rename column

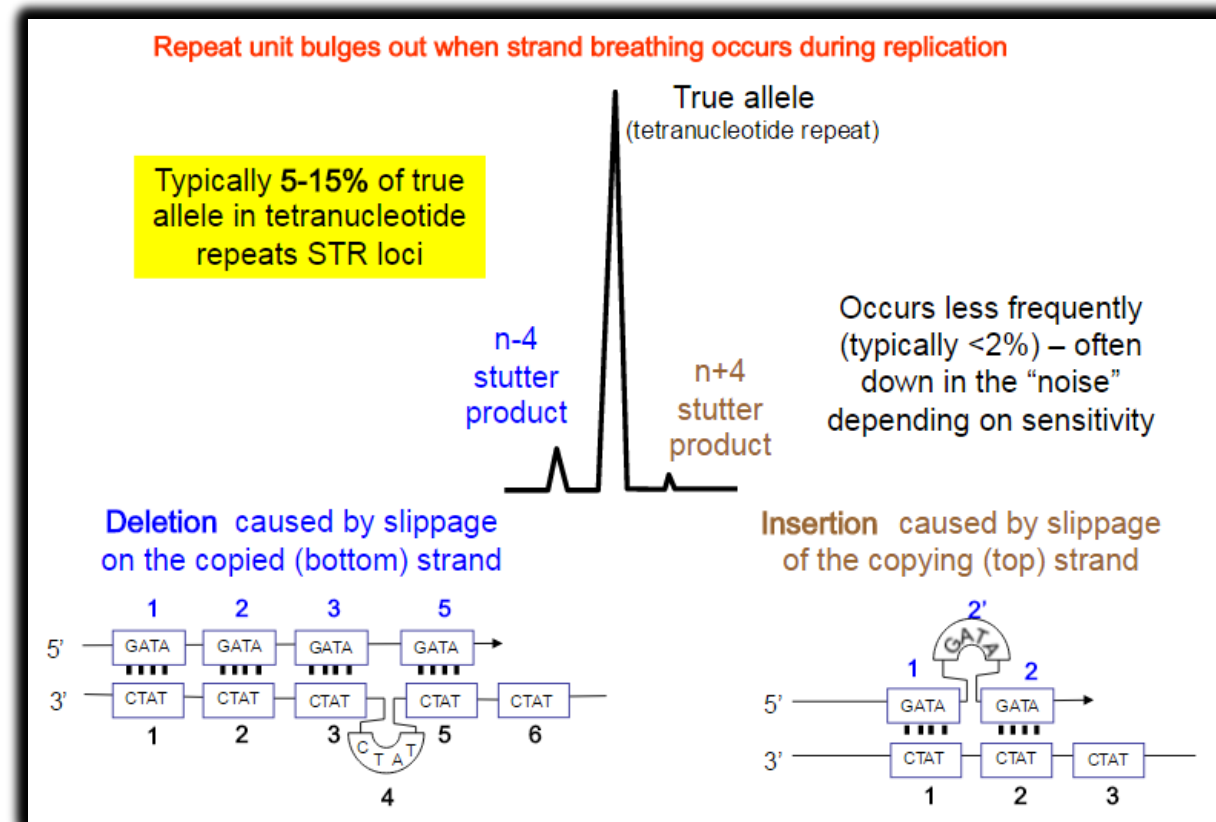
Note that none of the intervals extend near the +/- 0.5 bp range



Remember to Save Your Workspace

Stutter

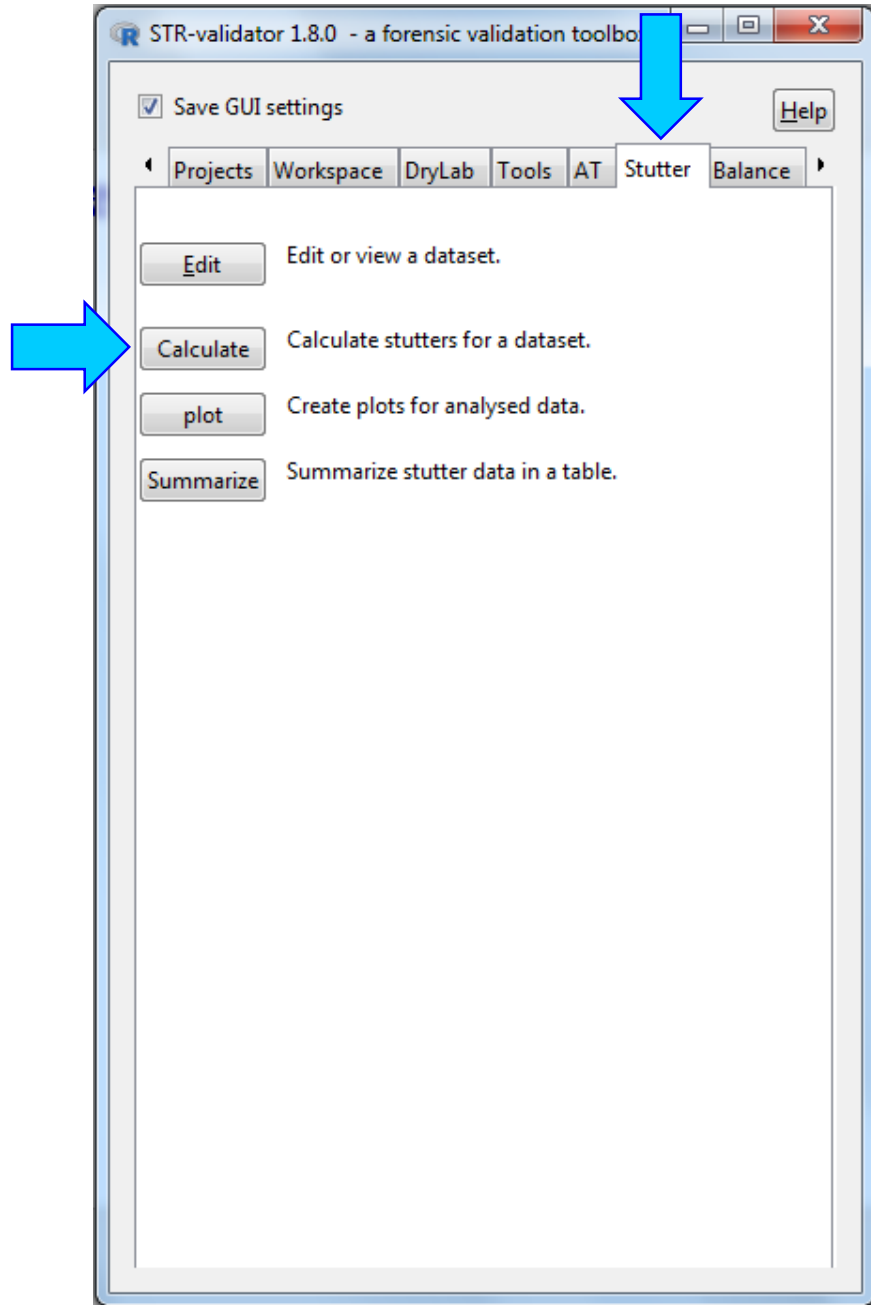
- ❖ Is a well-characterized PCR artifact.
- ❖ Appears as a minor peak one or more repeat units upstream or downstream from a true allele.
- ❖ Results from strand slippage during the amplification process



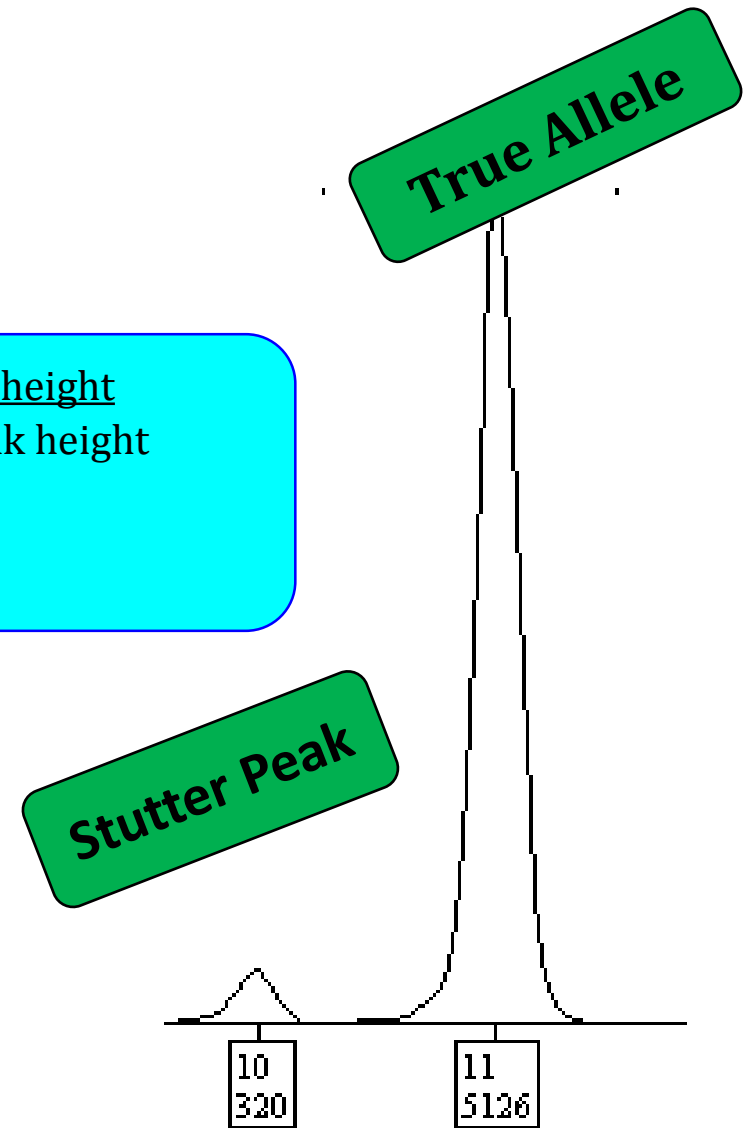
Experimental Procedure for Stutter Ratio

- ❖ 95 single source samples at 1.0 ng of DNA input included in stutter ratio calculation
- ❖ Analyzed at AT=1 in all dye channels with stutter filters turned off
- ❖ Export **_GenotypeTable.txt** from GeneMapper with at least the following information: **"Sample.Name"**, **"Marker"**, **"Allele"**, and **"Height"**.

How are Stutters Calculated in *STR-Validator*?

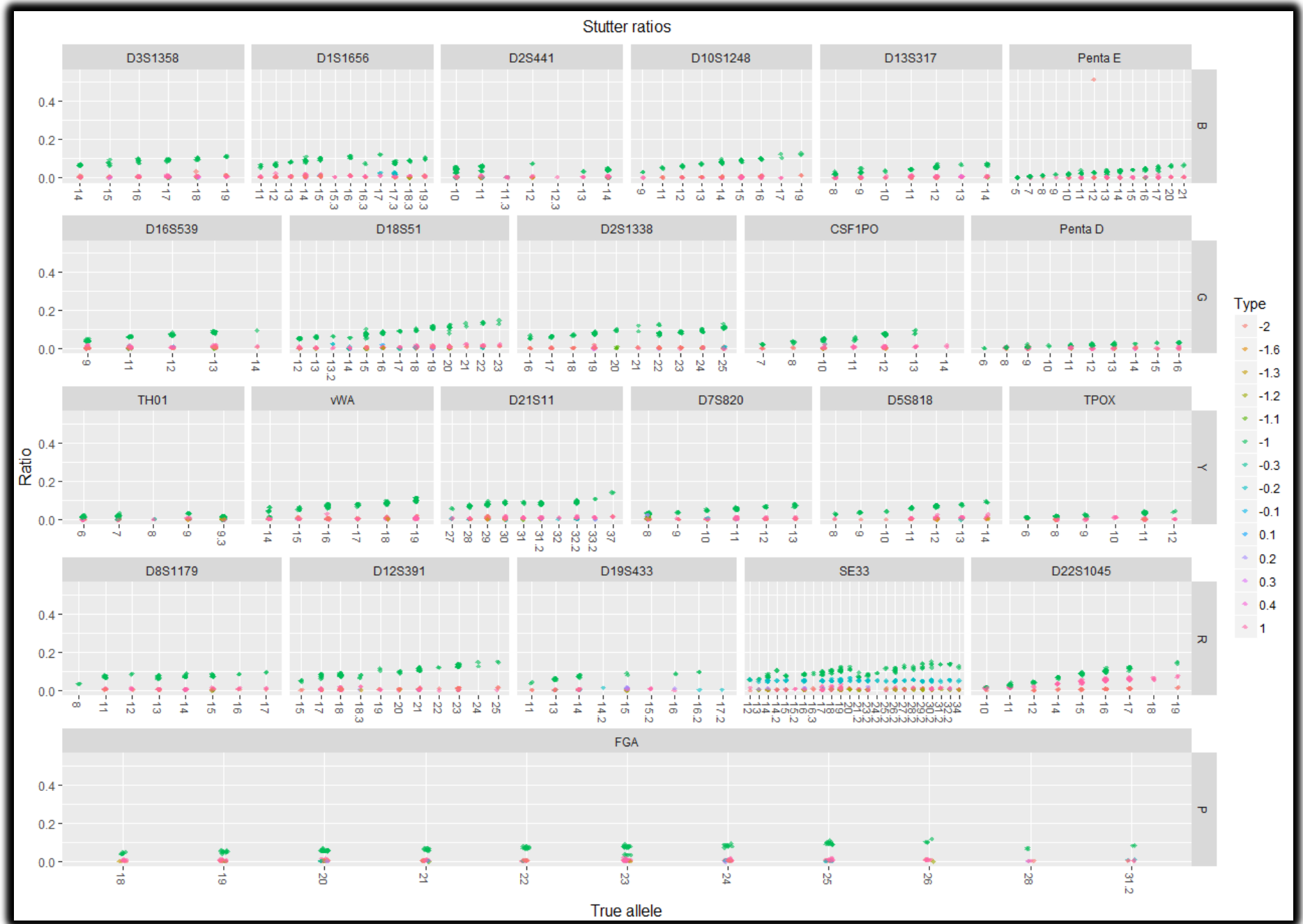


$$\text{Stutter Ratio} = \frac{\text{Stutter peak height}}{\text{True allele peak height}}$$
$$= \frac{320}{5126}$$



$$\text{Stutter peak designation} - \text{True Allele designation}$$
$$= 10 - 11$$
$$= -1 \text{ type of stutter}$$

How to Plot Stutter Ratio as a Function of the True Allele?



How to Calculate Average Stutter Percentage at Each Locus?

Save GUI settings Help

Datasets
 Select dataset: Stutter_NO_OVERLAP_table_locus <NA> samples, 8 columns, 26 rows

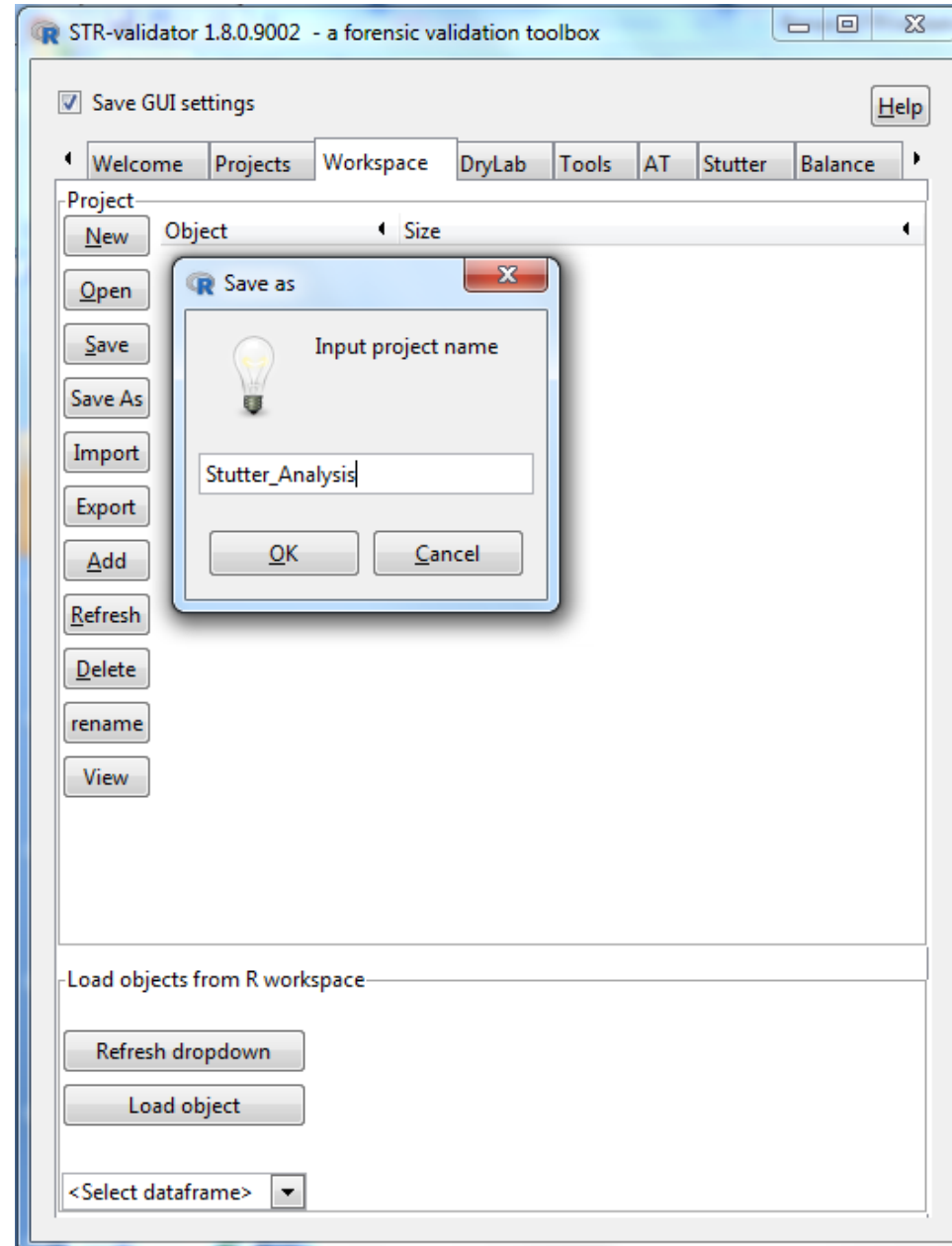
Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
 Stutter_NO_OVERLAP_table_locus_edit

Data frame

Row.names	Marker	Type	n.alleles	n.stutters	Mean	Stdv	Perc.95	Max
12	D12S391	NA	11	243	0.04980856	0.04847917	0.1349447	0.1519692
13	SE33	NA	24	470	0.04328839	0.03980691	0.1197097	0.1529809
14	D22S1045	NA	9	240	0.0492881	0.03602704	0.1155116	0.1497976
4	D18S51	NA	13	345	0.03077113	0.0383081	0.1144051	0.1493128
5	D2S1338	NA	10	216	0.04668302	0.04283466	0.1121561	0.1306066
18	DYS570	NA	6	279	0.03896943	0.04665069	0.1054264	0.4512635
26	D3S1358	NA	6	152	0.03399186	0.03912667	0.1047531	0.1165292
19	D1S1656	NA	12	267	0.03494432	0.03671343	0.1034651	0.1241259
7	vWA	NA	6	231	0.03700235	0.03852297	0.1012851	0.1160602
21	D10S1248	NA	9	138	0.03666577	0.04022816	0.1004014	0.1289291
17	DYS576	NA	7	277	0.03782644	0.03807173	0.1001362	0.1206309
8	D21S11	NA	10	315	0.03642765	0.03805507	0.0933165	0.1425637
16	FGA	NA	11	323	0.03084242	0.0332599	0.09123326	0.1175682
23	D19S433	NA	9	82	0.02826505	0.03191722	0.08694882	0.09970551
25	D8S1179	NA	8	141	0.03309349	0.03351953	0.08690511	0.09765866
3	D16S539	NA	5	218	0.02678956	0.02878527	0.08616047	0.09644599
6	CSF1PO	NA	7	173	0.0269808	0.02846134	0.07948276	0.09817352
10	D5S818	NA	7	194	0.02785527	0.02911381	0.07650658	0.09893651
15	DYS391	NA	3	254	0.02866788	0.02955303	0.0747429	0.09107884
9	D7S820	NA	6	219	0.02144259	0.02254944	0.06668492	0.08316733
1	D13S317	NA	7	172	0.0200418	0.02199814	0.06658742	0.07684729
20	D2S441	NA	7	296	0.01834139	0.0206675	0.05867814	0.07662058
2	Penta E	NA	14	251	0.01928513	0.03561548	0.0505487	0.5127697
11	TPOX	NA	6	167	0.01619873	0.01392732	0.03924471	0.04387755
22	Penta D	NA	10	146	0.01181251	0.009118571	0.02856621	0.03383897
24	TH01	NA	5	241	0.01006018	0.008220588	0.02206194	0.03579098

Open a New Workspace, Name and Save as Stutter_Analysis



Stutter Analysis

STR-validator 1.8.0 forensic validation toolbox

Save GUI settings Help

Projects Workspace DryLab Tools AT **Stutter** Balance

Project

New Object Size

Open

Save

Save As

Import

Export

Add

Refresh

Delete

❖ Import Data Set

❖ Import Reference Set

Load objects from R workspace

Refresh dropdown

Load object

<Select dataframe>

Import Data

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset

Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Stutter_Analysis\Stut browse

Select a directory... browse

Options

Save file name

Save file time stamp

Delimiter:

TAB

NA strings (separated by comma):

NA,

Auto trim samples

Auto slim repeated columns

+ Multiple files options

+ Trim options

+ Slim options

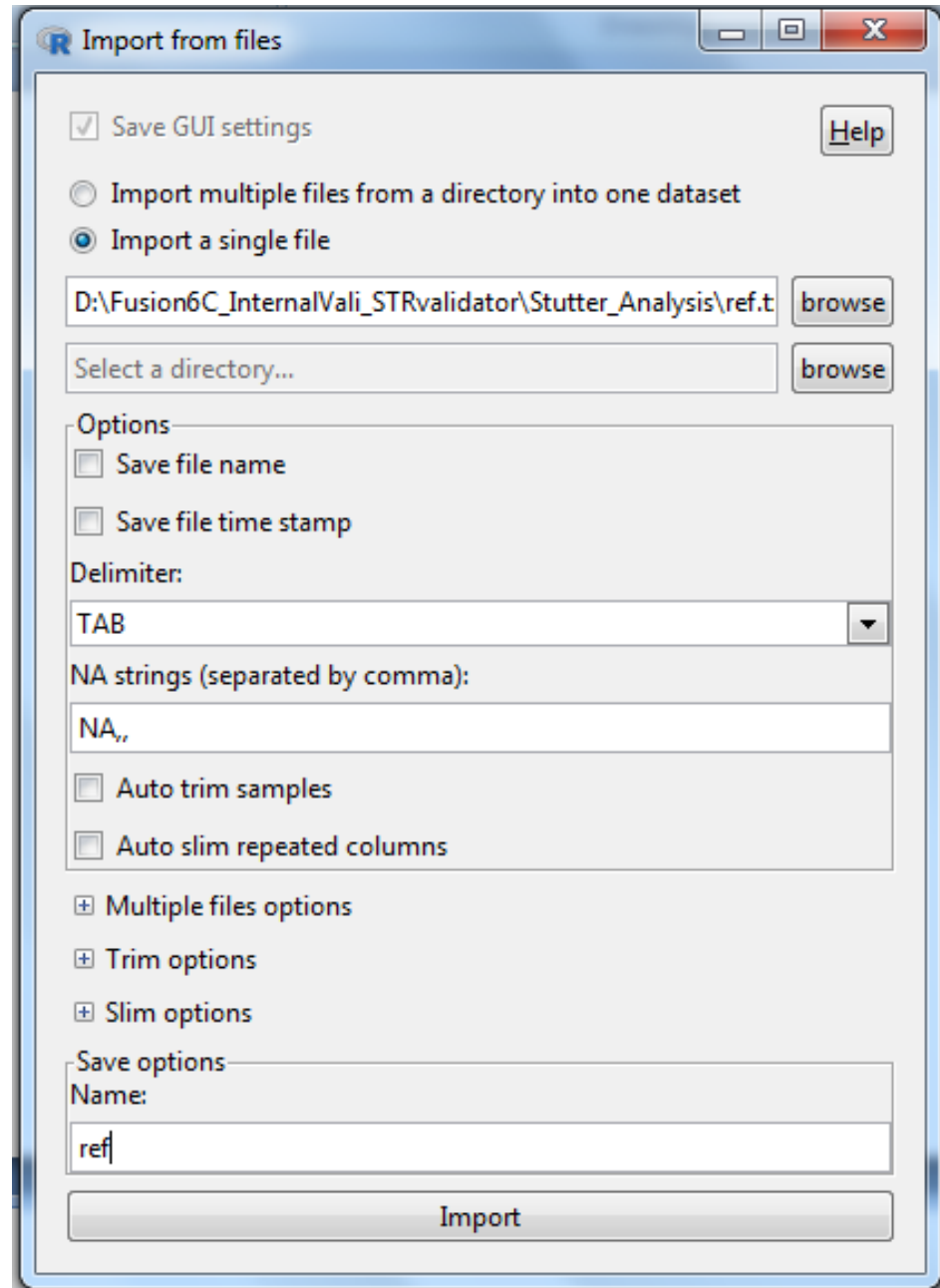
Save options

Name:

Stutter

Import

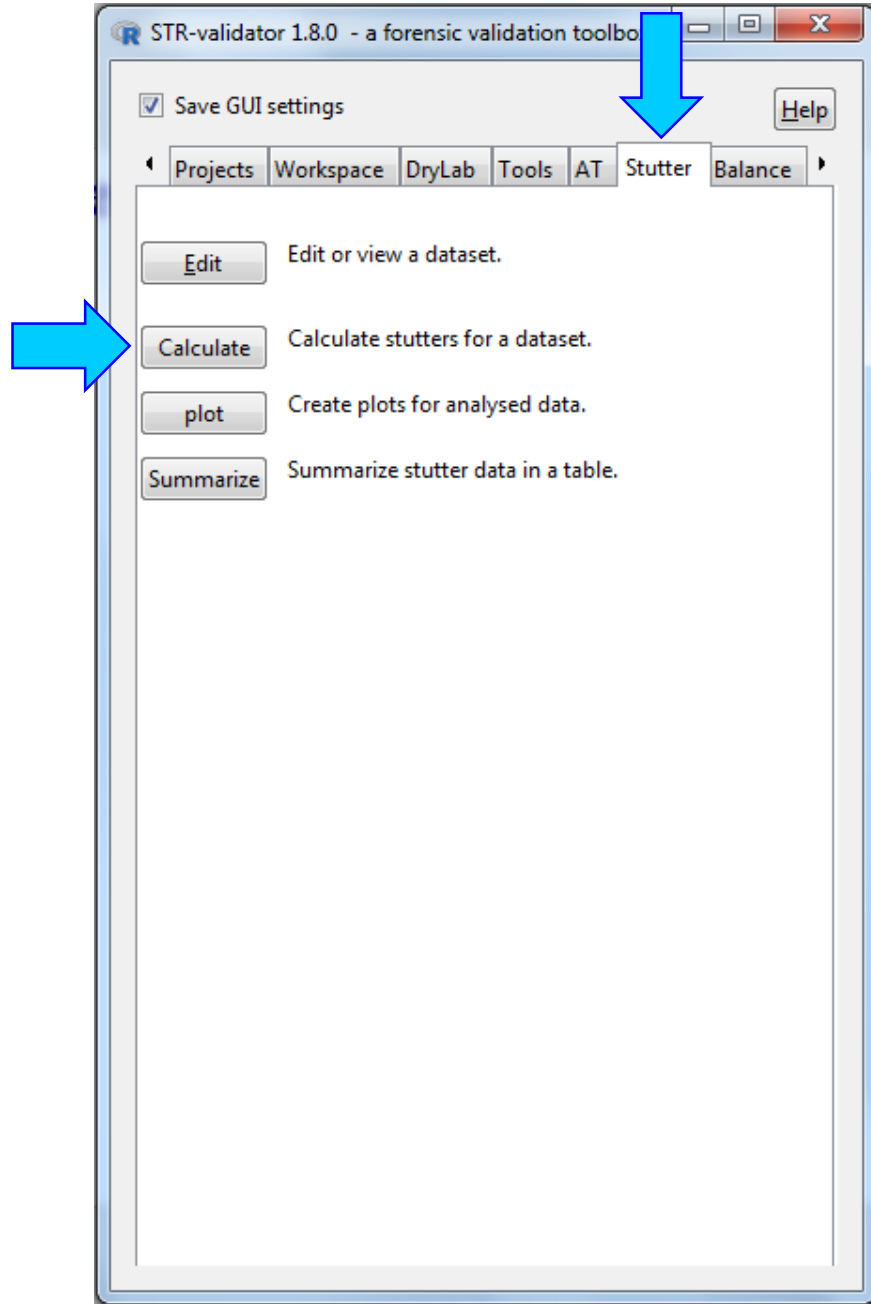
Import Reference





- ❖ Reference set contains the **known profiles** for the dataset samples.
- ❖ Reference set is used to **extract the known alleles** from the dataset.
- ❖ Therefore, it is very **important** to work with **a correct reference set**.
- ❖ Reference dataset requires the following information: **“Sample.Name”, “Marker”, and “Allele”**.

Calculate Stutter



Calculate Stutter Ratio

Calculate stutter ratio

Save GUI settings Help

Datasets
Select dataset: Stutter 95 samples
Select reference dataset: ref 38 references

Check subsetting

Options
Calculate stutter ratio within the the following analysis range:
2 backward stutters to 1 forward stutters.
NB! Additive effects outside the analysis range cannot be controlled.
A narrow range like 0 to +1 can be greatly affected by neighbouring -1 stutters.

Level of interference within the given range
 no overlap between stutters and alleles
 stutter-stutter interference allowed
 stutter-allele interference allowed

Replace 'false' stutters

Row.names	False.Stutter	True.Stutter	Replace
1	-1.9	-1.3	TRUE
2	-1.8	-1.2	TRUE
3	-1.7	-1.1	TRUE
4	-0.9	-0.3	TRUE
5	0.8	0.2	TRUE

Save as
Name for result: Stutter_stutter Kit attribute: Fusion 6C

Calculate

Reference name: A
Subsetted samples: A.1, A.2

Reference name: B
Subsetted samples: B.1, B.2

Reference name: C
Subsetted samples: C.1, C.2

Reference name: D
Subsetted samples: D.1, D.2

Reference name: E
Subsetted samples: E.1, E.2

Reference name: 33
Subsetted samples: 33.1, 33.2

Reference name: 34
Subsetted samples: 34.1, 34.2

Reference name: 35
Subsetted samples: 35.1, 35.2

Reference name: 36
Subsetted samples: 36.2, 36.3, 36.4, 36.5, 36.6, 36.7, 36.8, 36.9, 36.10, 36.11, 36.12, 36.13, 36.1

Reference name: 9947
Subsetted samples: 9947.1, 9947.2

Reference name: 9948
Subsetted samples: 9948.1, 9948.2

Reference name: K
Subsetted samples: K.1, K.2

Reference name: L
Subsetted samples: L.1, L.2

Reference name: M
Subsetted samples: M.1, M.2

Reference name: N
Subsetted samples: N.1, N.2

Reference name: O
Subsetted samples: O.1, O.2

Reference name: P
Subsetted samples: P.1, P.2

Reference name: Q
Subsetted samples: Q.1, Q.2

Reference name: R
Subsetted samples: R.1, R.2

Reference name: S
Subsetted samples: S.1, S.2

Reference name: T

Reference name: 28
Subsetted samples: 28.1, 28.2

Reference name: 29
Subsetted samples: 29.1, 29.2, 29.3, 29.4, 29.5

Reference name: 30
Subsetted samples: 30.1, 30.2

Reference name: 31
Subsetted samples: 31.1, 31.2, 31.3, 31.4, 31.5

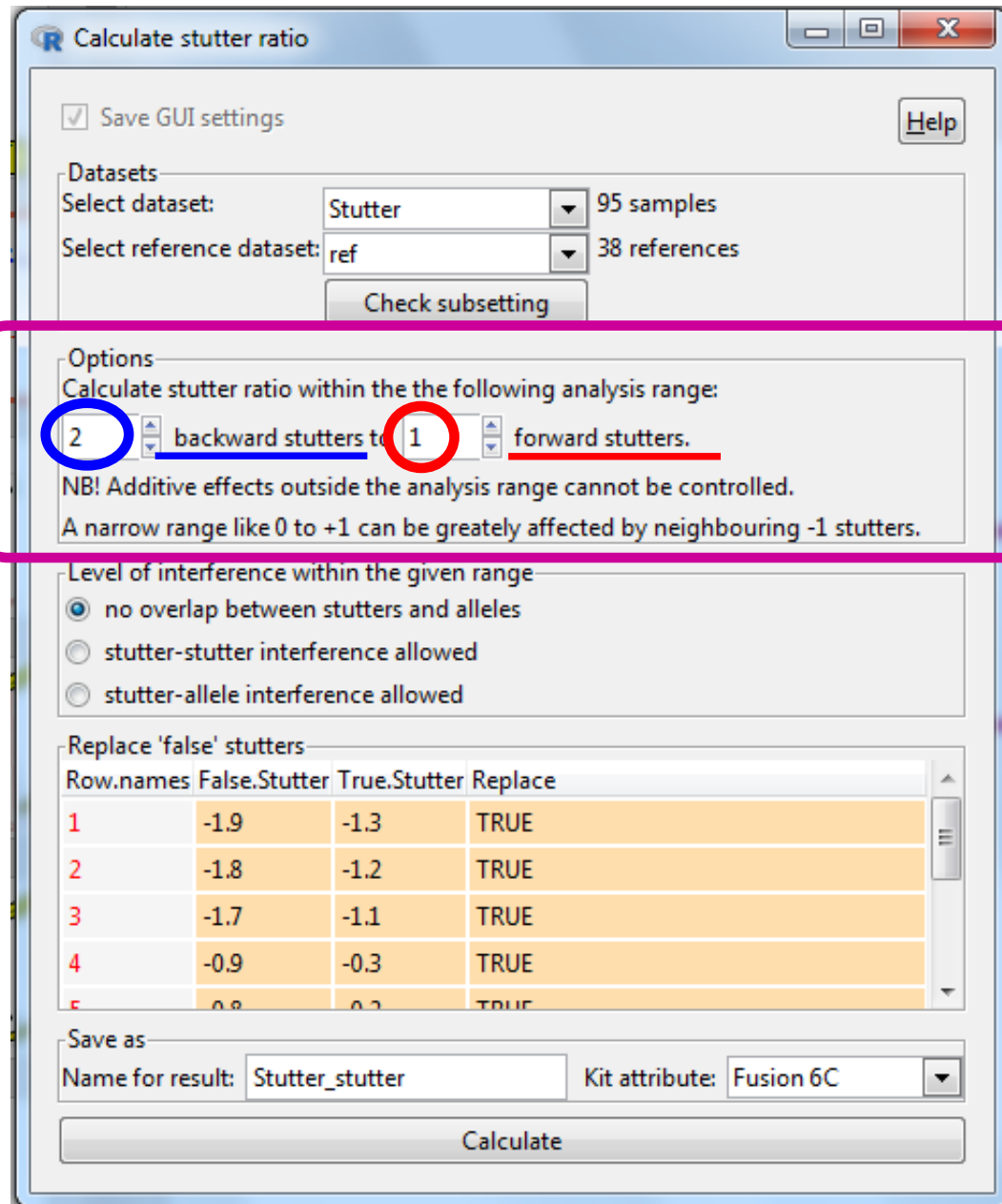
Reference name: 32
Subsetted samples: 32.1, 32.2, 32.3, 32.4, 32.5

Check Subsetting



- ❖ The naming convention for samples is **very important**.
- ❖ To prevent errors, always test the subsetting.

Analysis Range of Stutter Ratio



Calculate stutter ratio

Save GUI settings Help

Datasets
Select dataset: Stutter 95 samples
Select reference dataset: ref 38 references
Check subsetting

Options
Calculate stutter ratio within the the following analysis range:
2 backward stutters to **1** forward stutters.
NB! Additive effects outside the analysis range cannot be controlled.
A narrow range like 0 to +1 can be greatly affected by neighbouring -1 stutters.

Level of interference within the given range
 no overlap between stutters and alleles
 stutter-stutter interference allowed
 stutter-allele interference allowed

Replace 'false' stutters

Row.names	False.Stutter	True.Stutter	Replace
1	-1.9	-1.3	TRUE
2	-1.8	-1.2	TRUE
3	-1.7	-1.1	TRUE
4	-0.9	-0.3	TRUE
5	0.8	0.2	TRUE

Save as
Name for result: Stutter_stutter Kit attribute: Fusion 6C
Calculate

❖ Number of backward stutters = 2
an i.e. max repeat difference $2 = n-2$ repeats

❖ Number of forward stutters = 1
an i.e. max repeat difference $1 = n+1$ repeats

Level of Interference

Calculate stutter ratio

Save GUI settings Help

Datasets

Select dataset: Stutter 95 samples

Select reference dataset: ref 38 references

Options

Calculate stutter ratio within the the following analysis range:

2 backward stutters to 1 forward stutters.

NB! Additive effects outside the analysis range cannot be controlled.
A narrow range like 0 to +1 can be greatly affected by neighbouring -1 stutters.

Level of interference within the given range

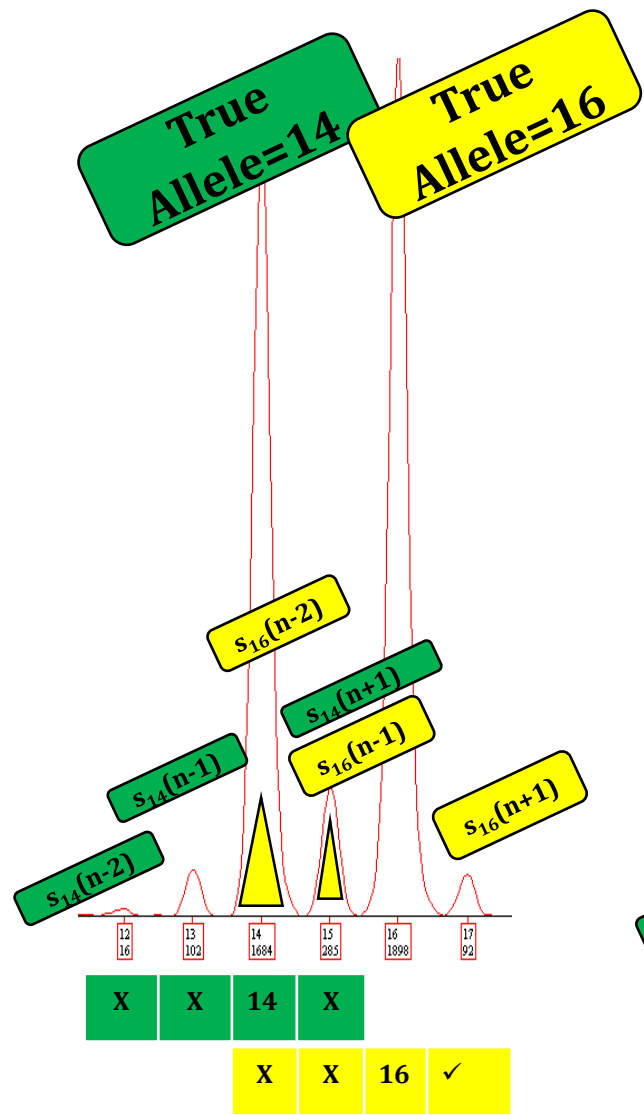
- no overlap between stutters and alleles
- stutter-stutter interference allowed
- stutter-allele interference allowed

Replace 'false' stutters

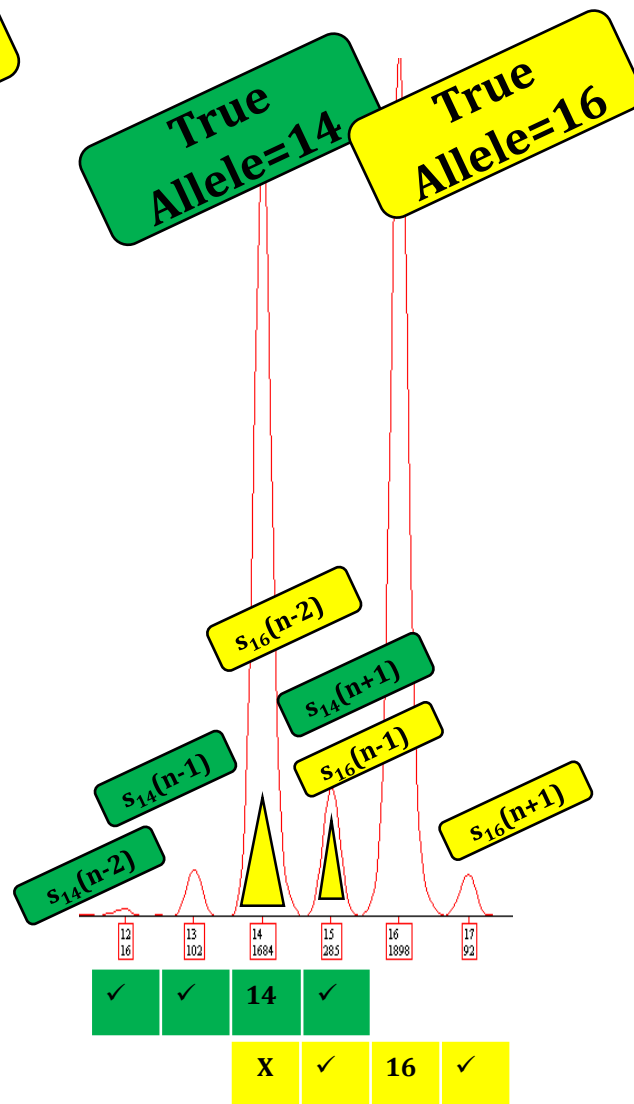
Row.names	False.Stutter	True.Stutter	Replace
1	-1.9	-1.3	TRUE
2	-1.8	-1.2	TRUE
3	-1.7	-1.1	TRUE
4	0.0	0.2	TRUE

Save as

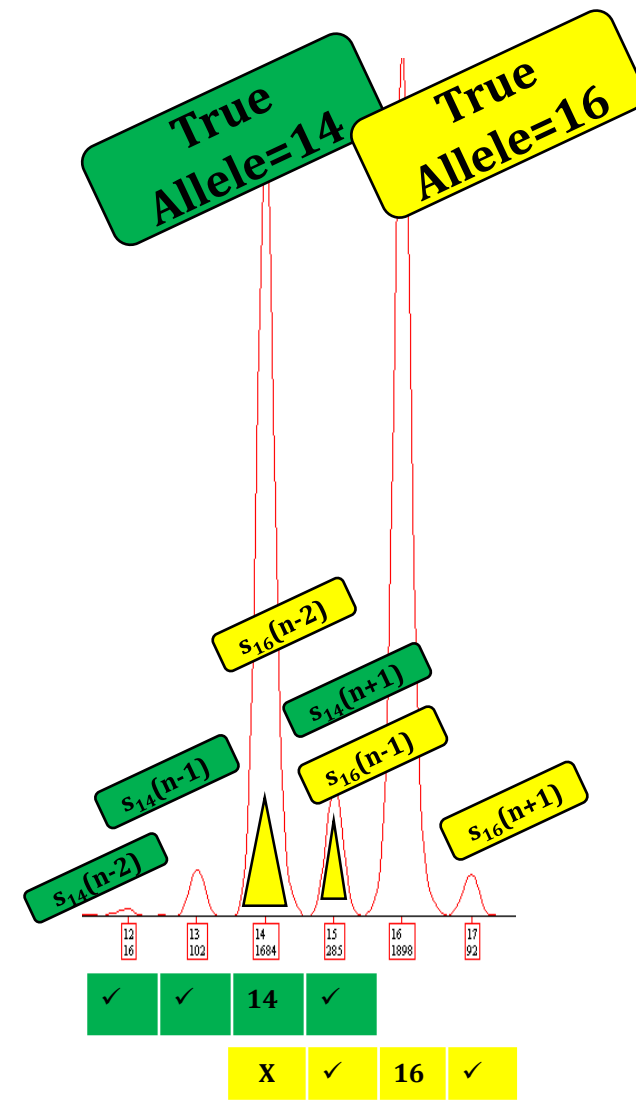
Name for result: Stutter_NO_OVERLAP Kit attribute: Fusion 6C



No Overlap between Stutters and Alleles



Stutter-Stutter Interference Allowed



Stutter-Allele Interference Allowed

Replace "False Stutters"

Calculate stutter ratio

Save GUI settings Help

Datasets

Select dataset: Stutter 95 samples

Select reference dataset: ref 38 references

Check subsetting

Options

Calculate stutter ratio within the the following analysis range:

2 backward stutters to 1 forward stutters.

NB! Additive effects outside the analysis range cannot be controlled.
A narrow range like 0 to +1 can be greatly affected by neighbouring -1 stutters.

Level of interference within the given range

no overlap between stutters and alleles

stutter-stutter interference allowed

stutter-allele interference allowed

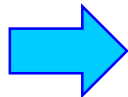
Replace 'false' stutters

Row.names	False.Stutter	True.Stutter	Replace
1	-1.9	-1.3	TRUE
2	-1.8	-1.2	TRUE
3	-1.7	-1.1	TRUE
4	0.0	0.2	TRUE

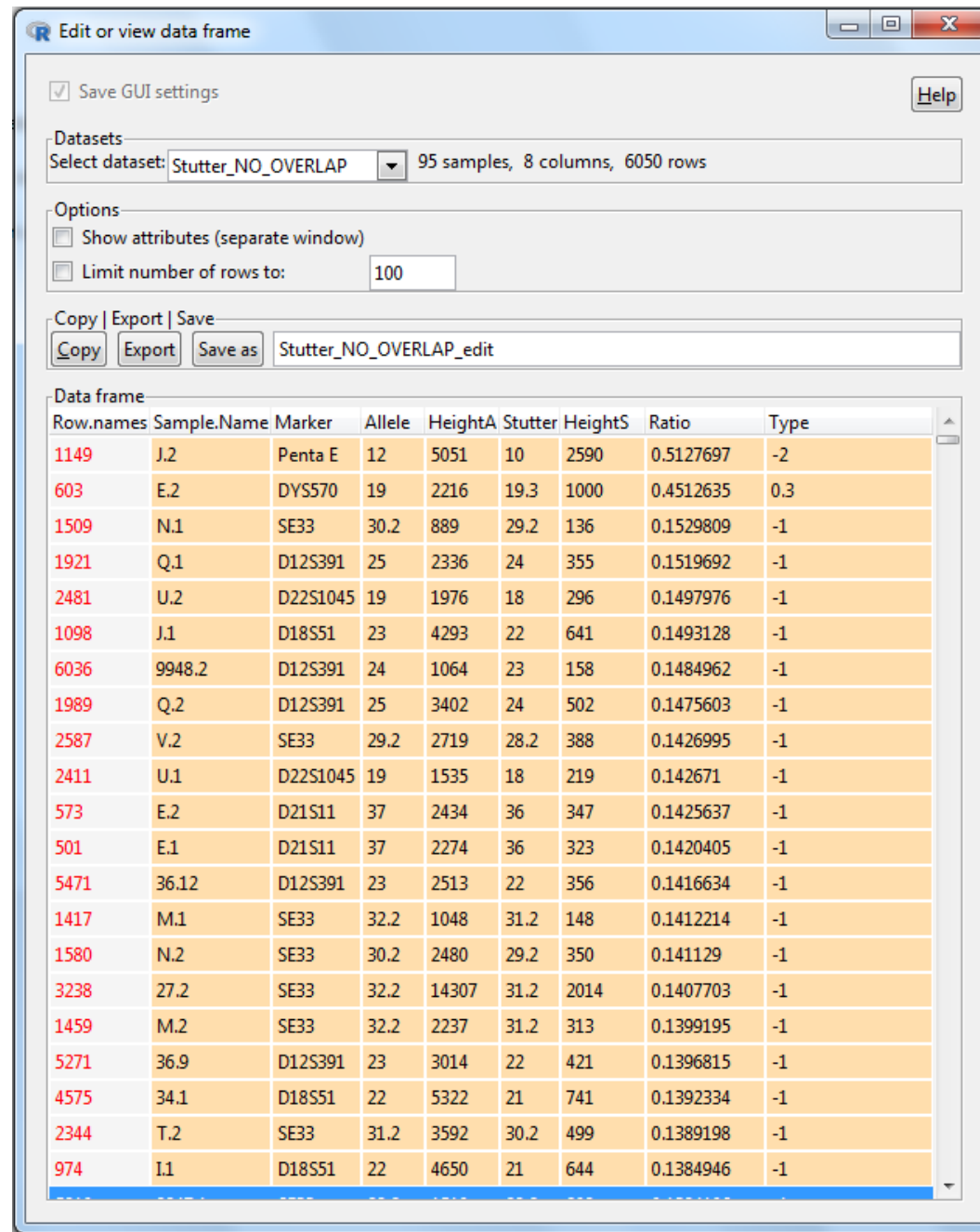
Save as

Name for result: Stutter_NO_OVERLAP Kit attribute: Fusion 6C

Calculate



View the Results and Sort the Column of Ratio



Save GUI settings Help

Datasets
Select dataset: Stutter_NO_OVERLAP 95 samples, 8 columns, 6050 rows

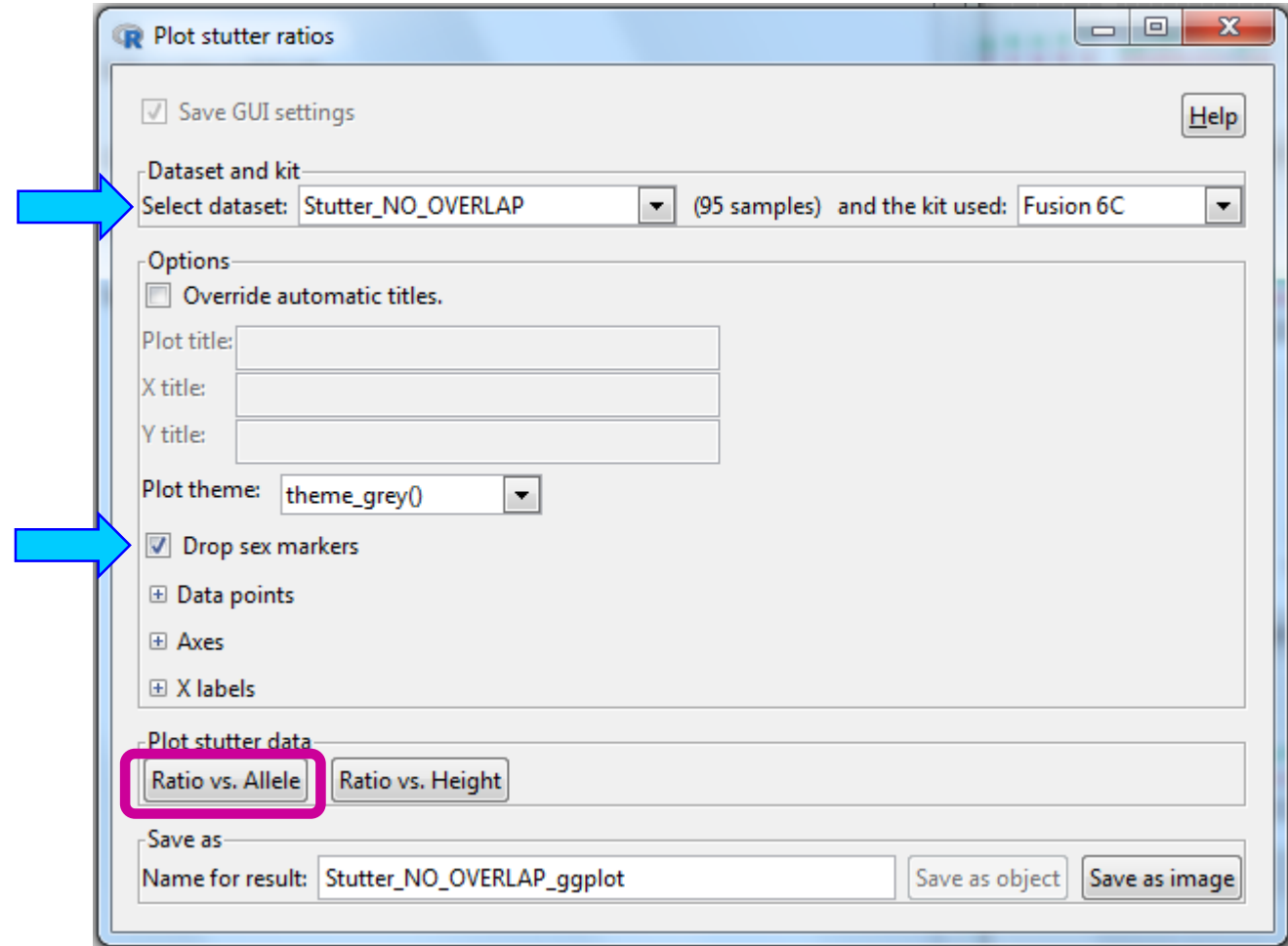
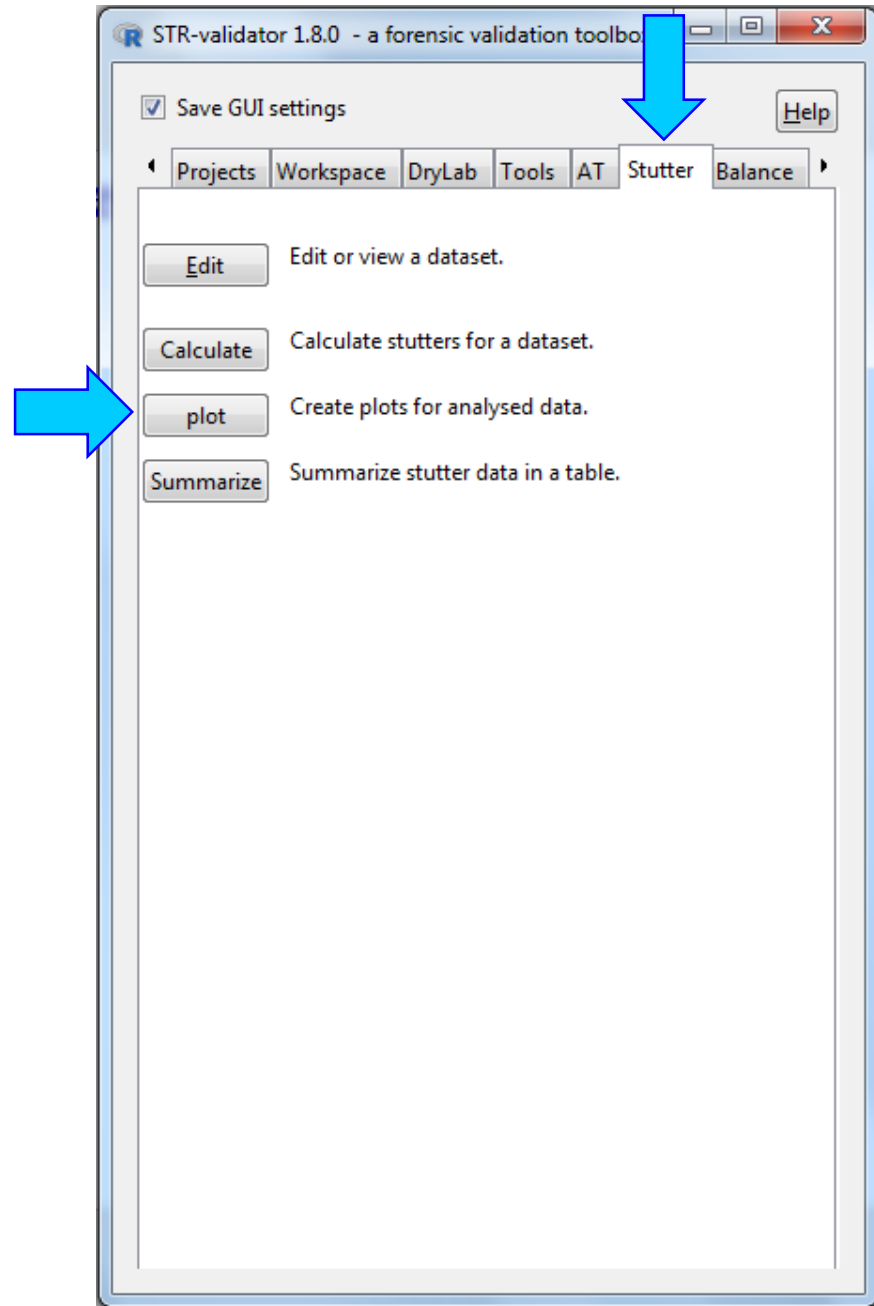
Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
Copy Export Save as Stutter_NO_OVERLAP_edit

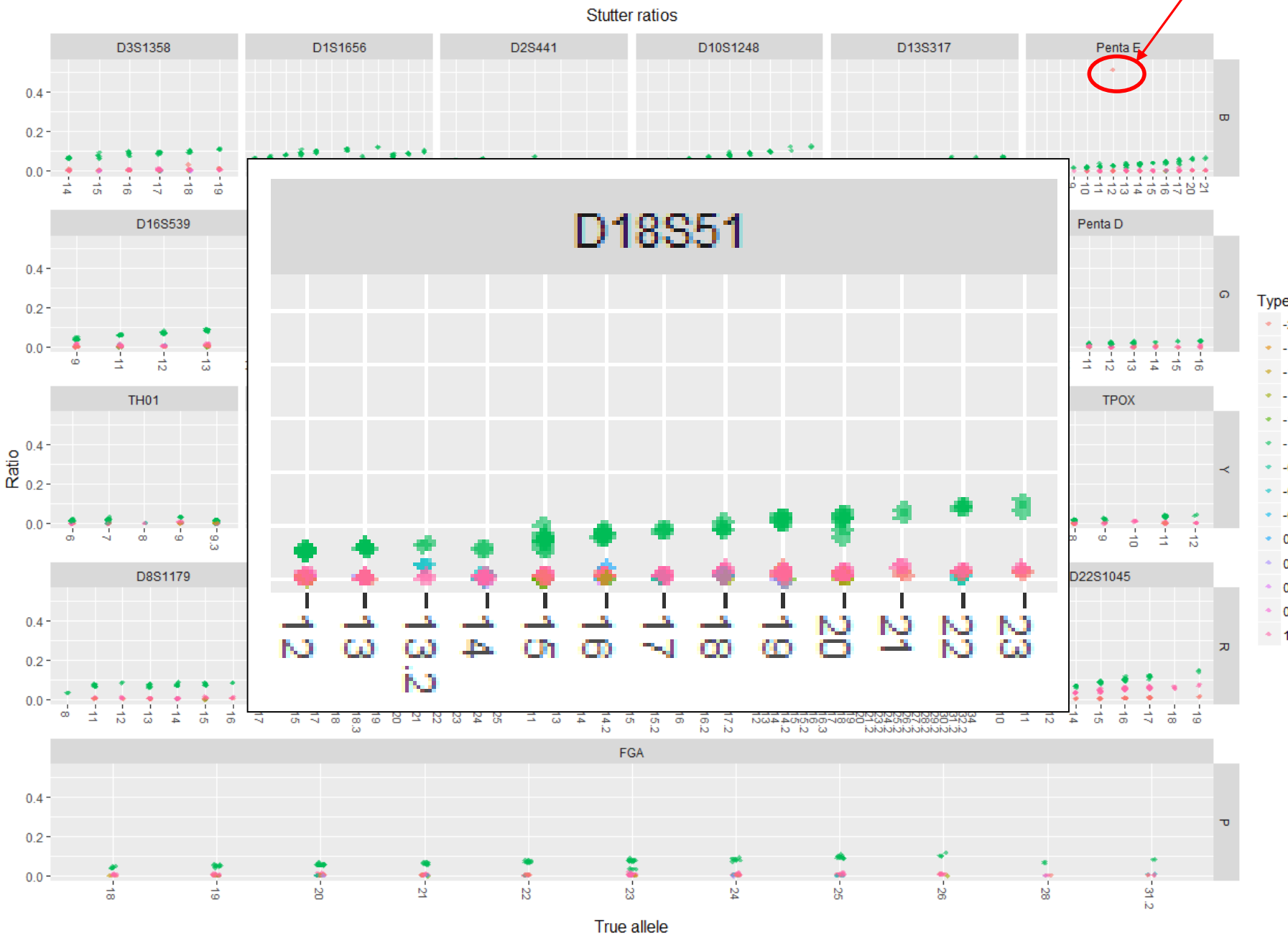
Data frame

Row.names	Sample.Name	Marker	Allele	HeightA	Stutter	HeightS	Ratio	Type
1149	J.2	Penta E	12	5051	10	2590	0.5127697	-2
603	E.2	DYS570	19	2216	19.3	1000	0.4512635	0.3
1509	N.1	SE33	30.2	889	29.2	136	0.1529809	-1
1921	Q.1	D12S391	25	2336	24	355	0.1519692	-1
2481	U.2	D22S1045	19	1976	18	296	0.1497976	-1
1098	J.1	D18S51	23	4293	22	641	0.1493128	-1
6036	9948.2	D12S391	24	1064	23	158	0.1484962	-1
1989	Q.2	D12S391	25	3402	24	502	0.1475603	-1
2587	V.2	SE33	29.2	2719	28.2	388	0.1426995	-1
2411	U.1	D22S1045	19	1535	18	219	0.142671	-1
573	E.2	D21S11	37	2434	36	347	0.1425637	-1
501	E.1	D21S11	37	2274	36	323	0.1420405	-1
5471	36.12	D12S391	23	2513	22	356	0.1416634	-1
1417	M.1	SE33	32.2	1048	31.2	148	0.1412214	-1
1580	N.2	SE33	30.2	2480	29.2	350	0.141129	-1
3238	27.2	SE33	32.2	14307	31.2	2014	0.1407703	-1
1459	M.2	SE33	32.2	2237	31.2	313	0.1399195	-1
5271	36.9	D12S391	23	3014	22	421	0.1396815	-1
4575	34.1	D18S51	22	5322	21	741	0.1392334	-1
2344	T.2	SE33	31.2	3592	30.2	499	0.1389198	-1
974	I.1	D18S51	22	4650	21	644	0.1384946	-1

Plot Stutters

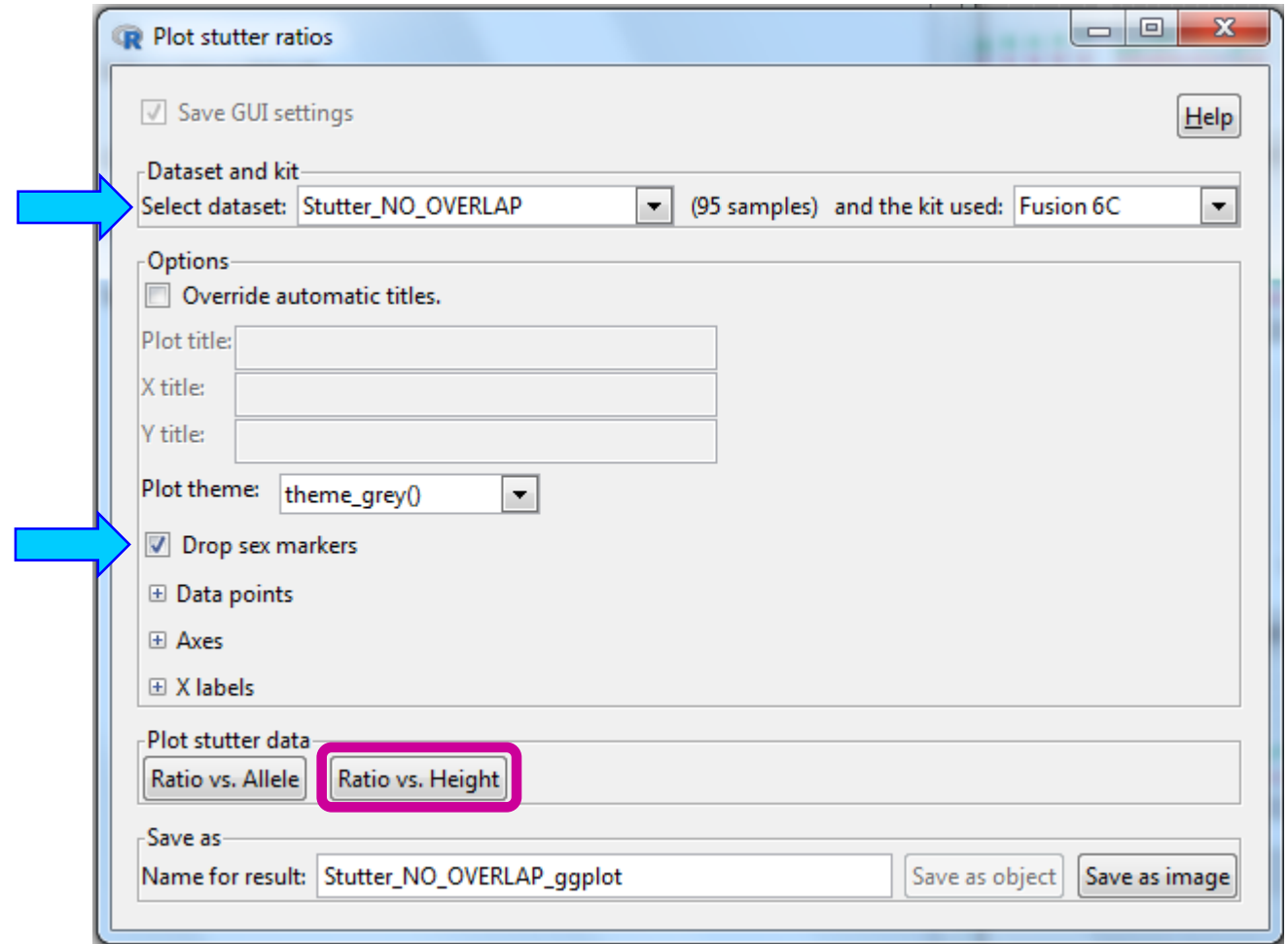
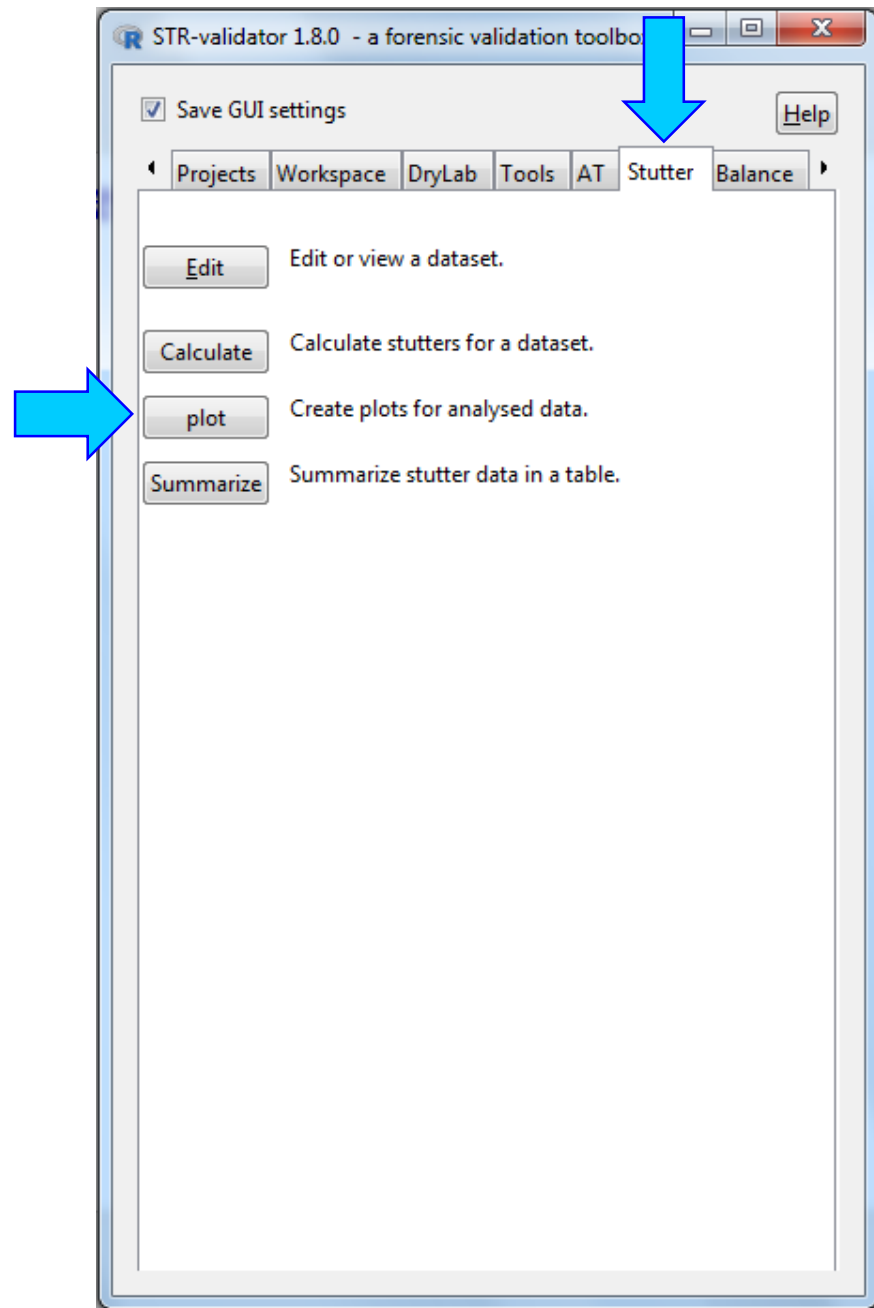


Stutter Ratio as a Function of Parent Allele

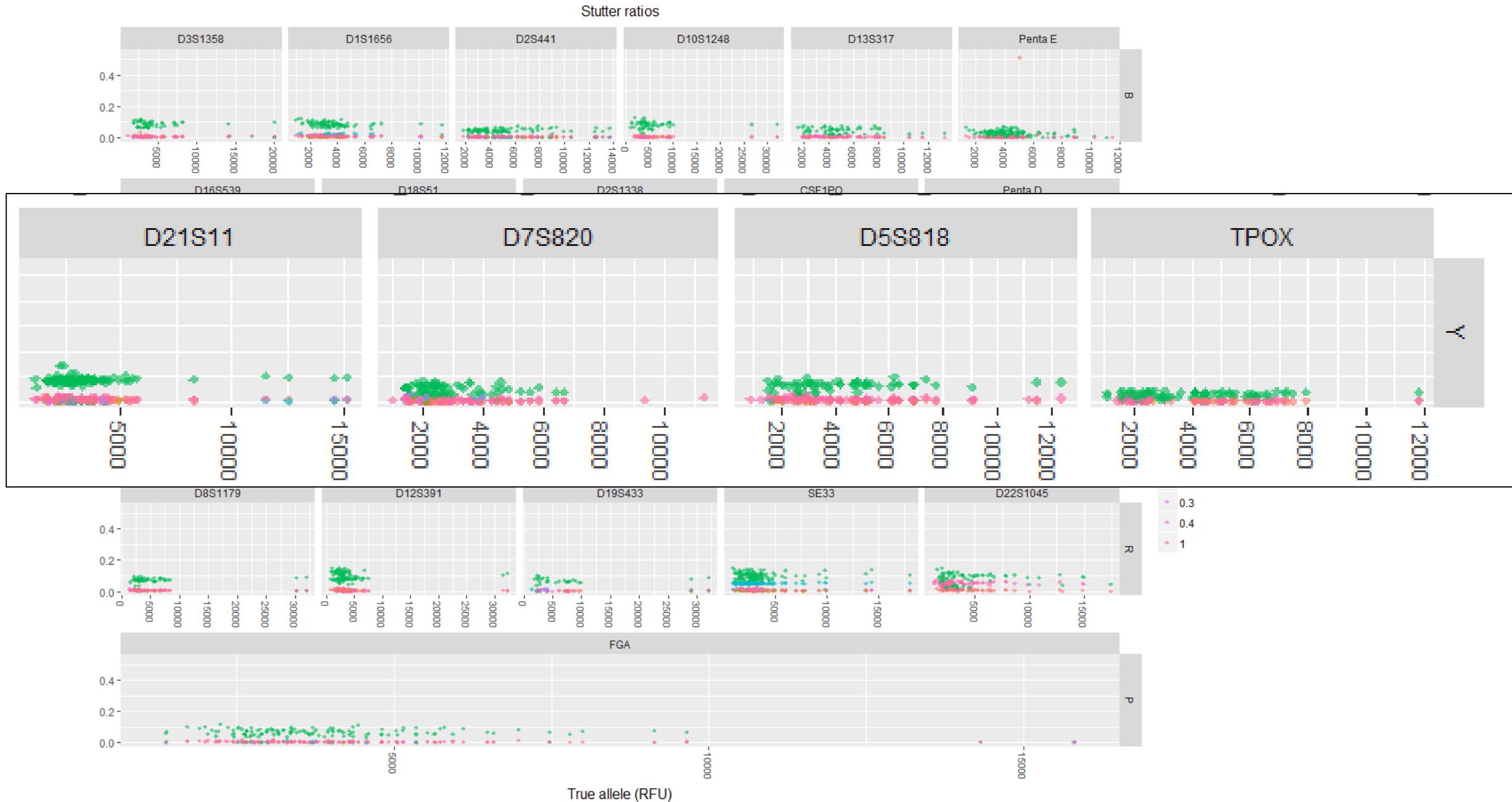


Stutter Ratio increases as the number of repeats increases

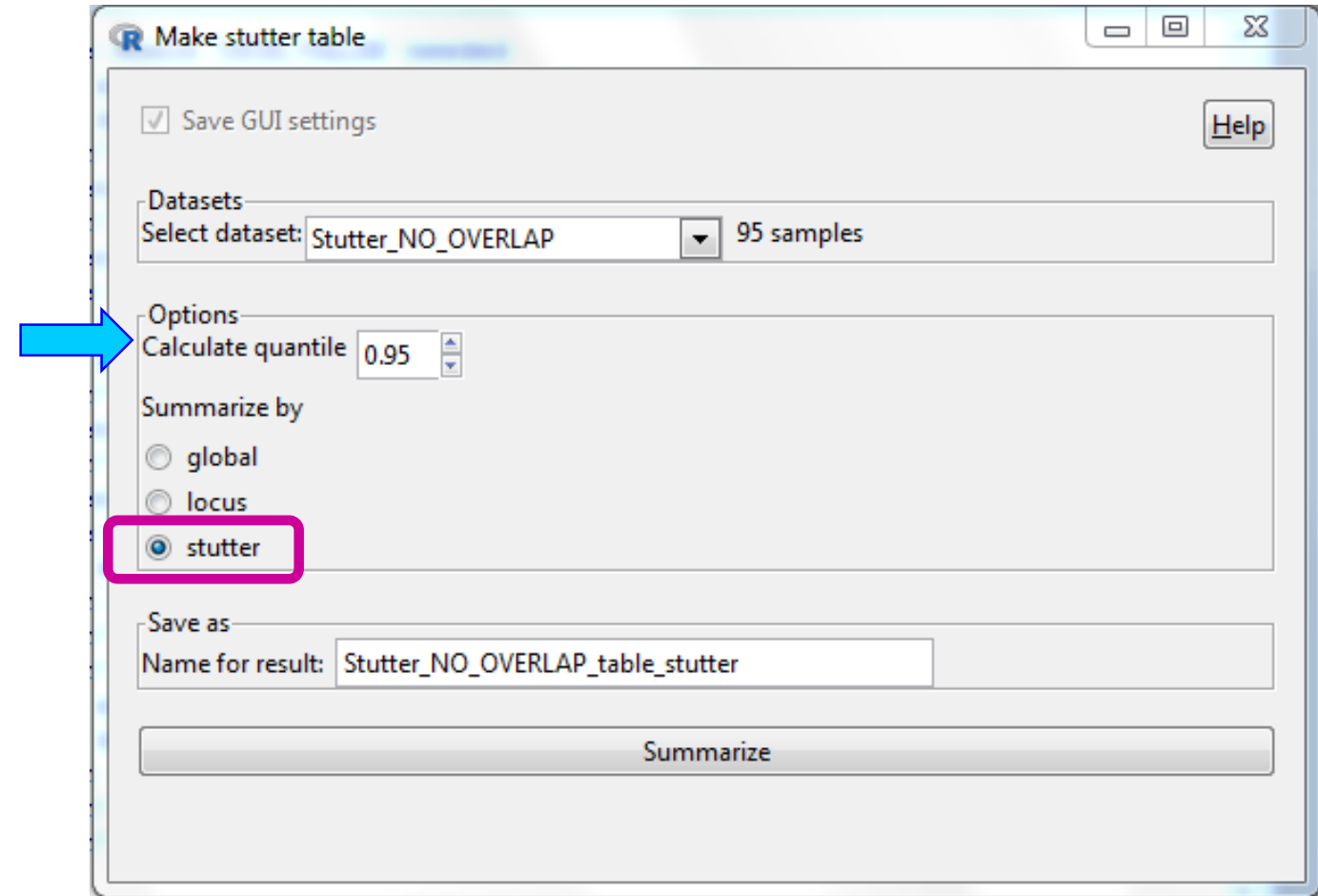
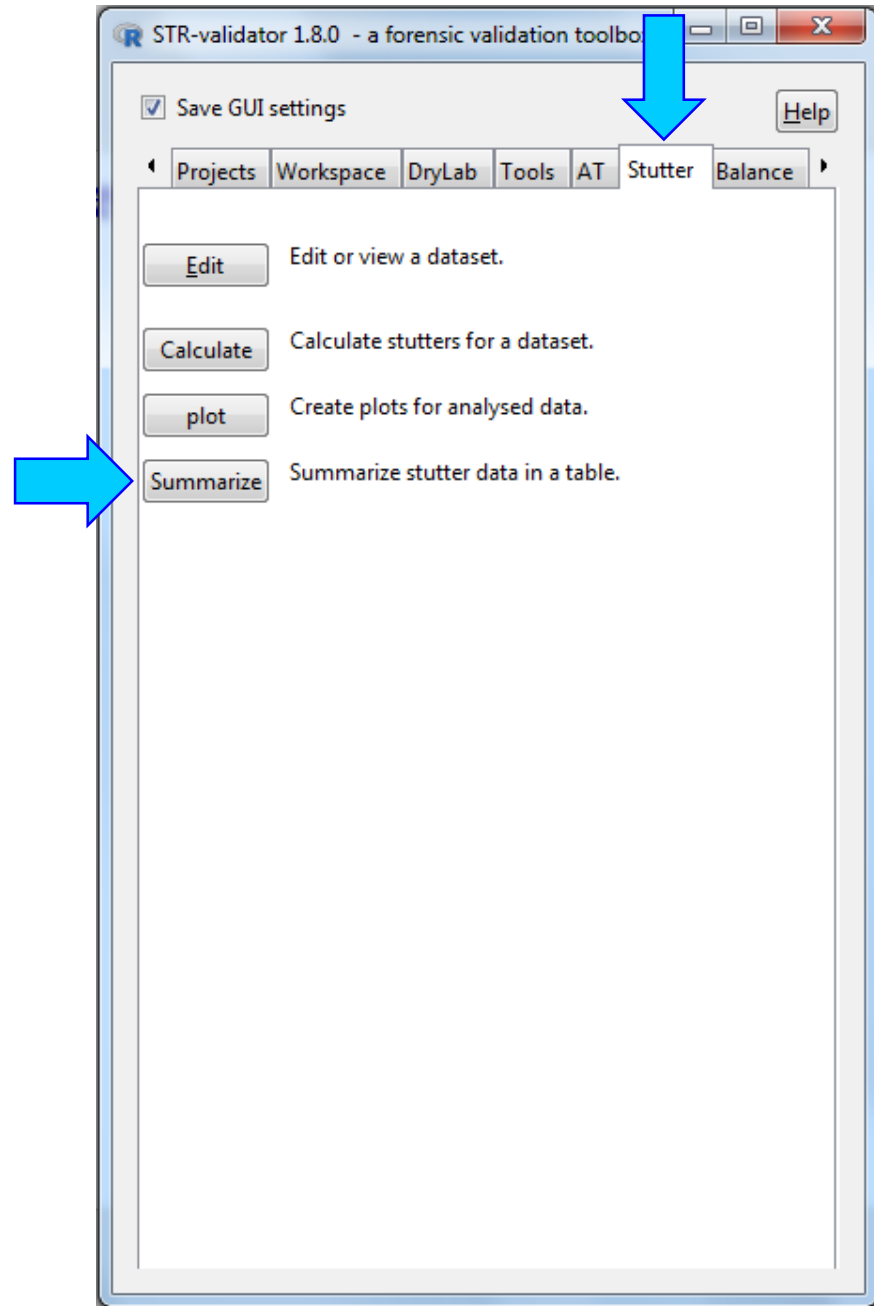
Plot Stutter Ratio as a Function of Peak Height



Stutter Ratio as a Function of Peak Height



Calculate Stutter Statistics by Stutter



View the Results and Sort the Column of Perc.95 (decreasing)

Save GUI settings Help

Datasets
Select dataset: Stutter_NO_OVERLAP_table_stutter <NA> samples, 8 columns, 172 rows

Options
 Show attributes (separate window)
 Limit number of rows to:

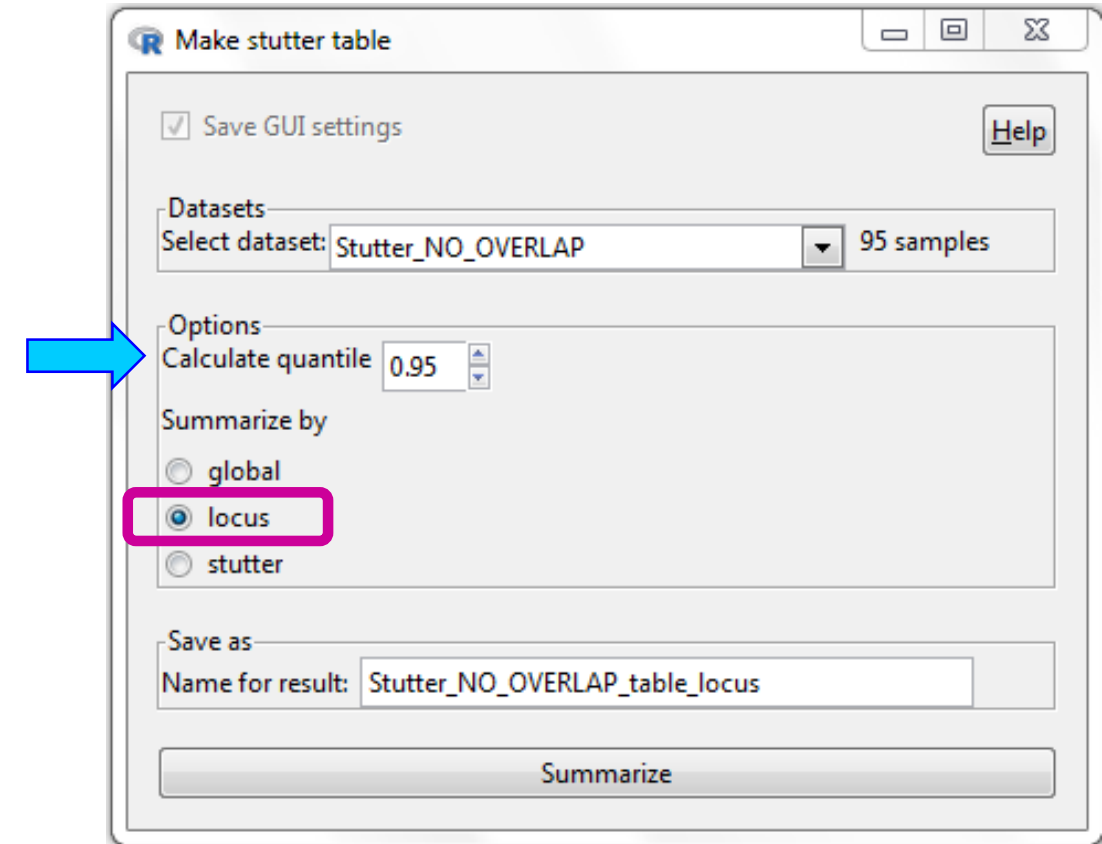
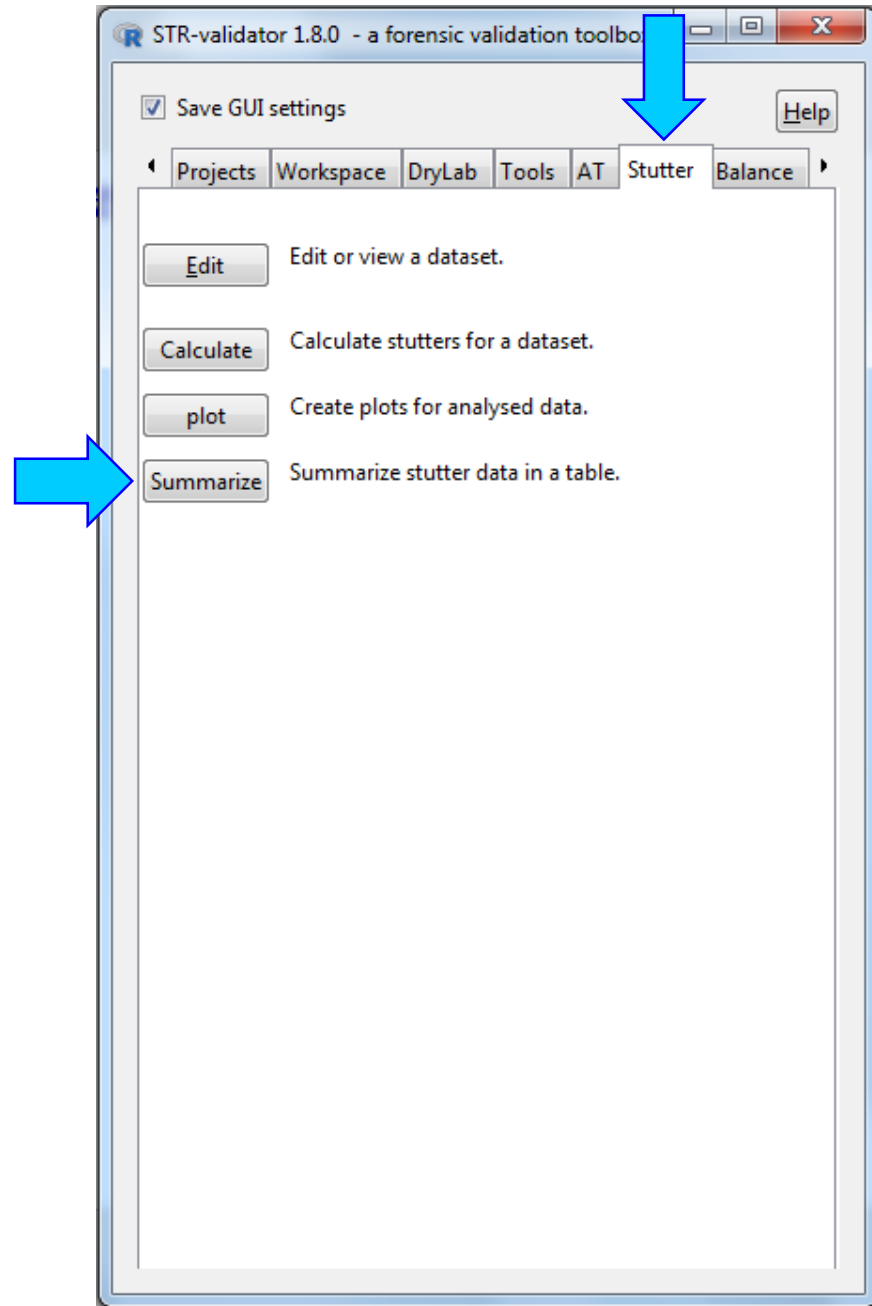
Copy | Export | Save
 Stutter_NO_OVERLAP_table_stutter_edit

Data frame

Row.names	Marker	Type	n.alleles	n.stutters	Mean	Stdv	Perc.95	
120	DYS570	0.3	1	1	0.4512635	NA	0.4512635	
88	SE33	-1	23	118	0.1010629	0.02147765	0.1390698	
80	D12S391	-1	11	112	0.09887554	0.02434941	0.1368667	
19	D18S51	-1	13	102	0.08523912	0.02615641	0.1306905	0.1493128
93	D22S1045	-1	8	92	0.07723934	0.03584264	0.1210751	0.1497976
31	D2S1338	-1	10	108	0.08758822	0.01735611	0.1163259	0.1306066

Apply function to column
Sort by column (decreasing)
Sort by column (increasing)
Rename column

Calculate Stutter Statistics by Locus



View the Results and Sort the Column of Perc.95 (decreasing)

Edit or view data frame

Save GUI settings Help

Datasets
Select dataset: Stutter_NO_OVERLAP_table_locus <NA> samples, 8 columns, 26 rows

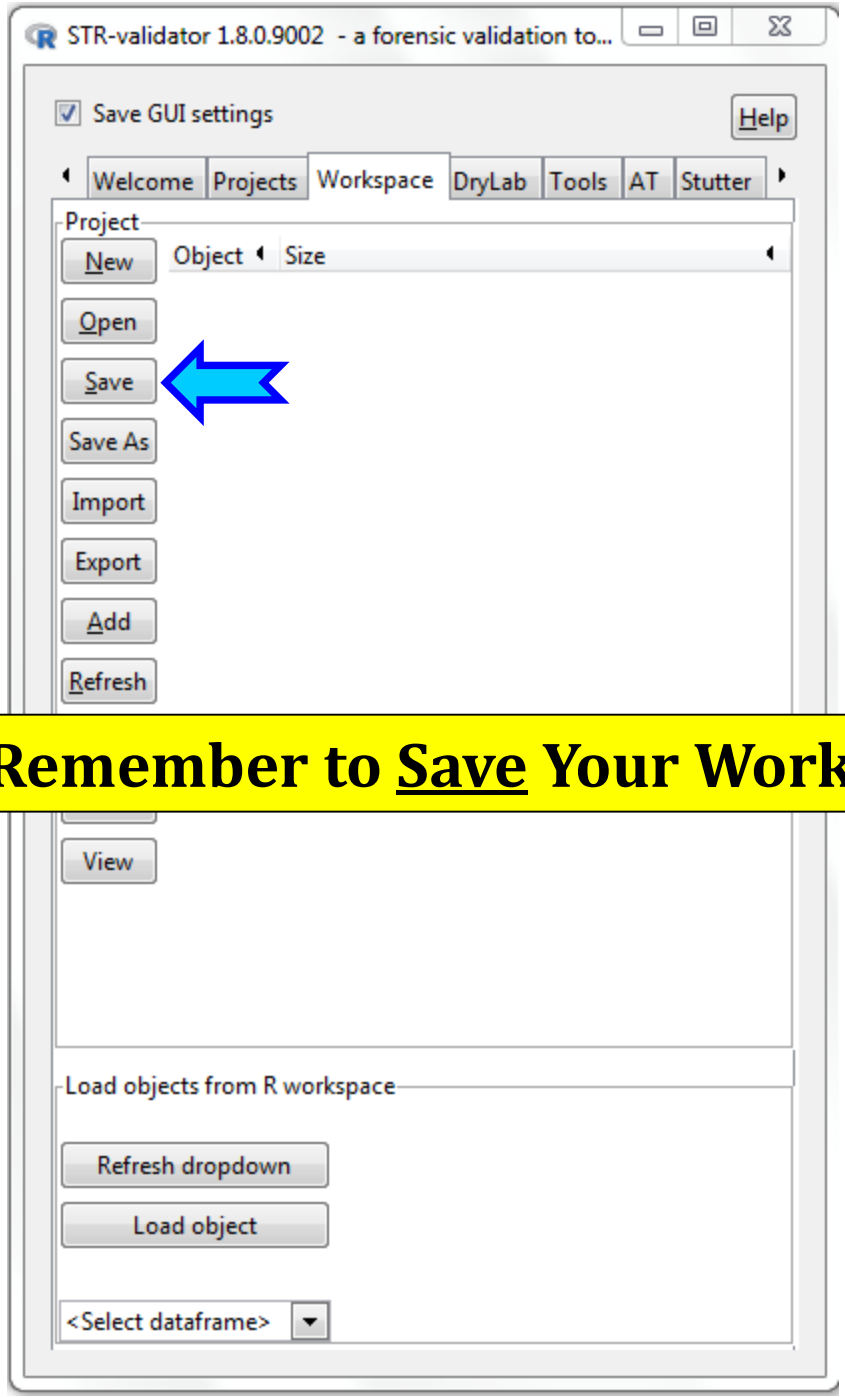
Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
Copy Export Save as Stutter_NO_OVERLAP_table_locus_edit

Data frame

Row.names	Marker	Type	n.alleles	n.stutters	Mean	Stdv	Perc.95	Max
12	D12S391	NA	11	243	0.04980856	0.04847917	0.1349447	0.1519692
13	SE33	NA	24	470	0.04328839	0.03980691	0.1197097	0.1529809
14	D22S1045	NA	9	240	0.0492881	0.03602704	0.1155116	0.1497976
4	D18S51	NA	13	345	0.03077113	0.0383081	0.1144051	0.1493128
5	D2S1338	NA	10	216	0.04668302	0.04283466	0.1121561	0.1306066
18	DYS570	NA	6	279	0.03896943	0.04665069	0.1054264	0.4512635
26	D3S1358	NA	6	152	0.03399186	0.03912667	0.1047531	0.1165292
19	D1S1656	NA	12	267	0.03494432	0.03671343	0.1034651	0.1241259
7	vWA	NA	6	231	0.03700235	0.03852297	0.1012851	0.1160602
21	D10S1248	NA	9	138	0.03666577	0.04022816	0.1004014	0.1289291
17	DYS576	NA	7	277	0.03782644	0.03807173	0.1001362	0.1206309
8	D21S11	NA	10	315	0.03642765	0.03805507	0.0933165	0.1425637
16	FGA	NA	11	323	0.03084242	0.0332599	0.09123326	0.1175682
23	D19S433	NA	9	82	0.02826505	0.03191722	0.08694882	0.09970551
25	D8S1179	NA	8	141	0.03309349	0.03351953	0.08690511	0.09765866
3	D16S539	NA	5	218	0.02678956	0.02878527	0.08616047	0.09644599
6	CSF1PO	NA	7	173	0.0269808	0.02846134	0.07948276	0.09817352
10	D5S818	NA	7	194	0.02785527	0.02911381	0.07650658	0.09893651
15	DYS391	NA	3	254	0.02866788	0.02955303	0.0747429	0.09107884
9	D7S820	NA	6	219	0.02144259	0.02254944	0.06668492	0.08316733
1	D13S317	NA	7	172	0.0200418	0.02199814	0.06658742	0.07684729
20	D2S441	NA	7	296	0.01834139	0.0206675	0.05867814	0.07662058
2	Penta E	NA	14	251	0.01928513	0.03561548	0.0505487	0.5127697
11	TPOX	NA	6	167	0.01619873	0.01392732	0.03924471	0.04387755
22	Penta D	NA	10	146	0.01181251	0.009118571	0.02856621	0.03383897
24	TH01	NA	5	241	0.01006018	0.008220588	0.02206194	0.03579098

The highest stutter ratio is observed in marker (D12S391)



Remember to Save Your Workspace

Workshop Schedule

Time	Topic
9:00 AM-10:00 AM	<ul style="list-style-type: none">❖ Load <i>STR-validator</i> package and launch its GUI❖ Trim and Slim txt.files❖ Check Precision❖ Calculate Stutter Thresholds
10:00 AM – 10:10 AM	Break
10:10 AM-11:00 AM	<ul style="list-style-type: none">❖ Calculate Analytical Thresholds❖ Analyze Peak Height Ratio
11:00 AM -11:10 AM	Break
11:10 AM-12:00 PM	<ul style="list-style-type: none">❖ Calculate Stochastic Thresholds❖ Questions❖ Feedback about the workshop (survey)❖ Workshop ends

The Analytical Threshold

SWGAM Autosomal STR Interpretation Guidelines

Analytical Threshold

- ❖ Peaks at and above this threshold can be **reliably distinguished from background noise** and are generally considered either artifacts or true alleles.

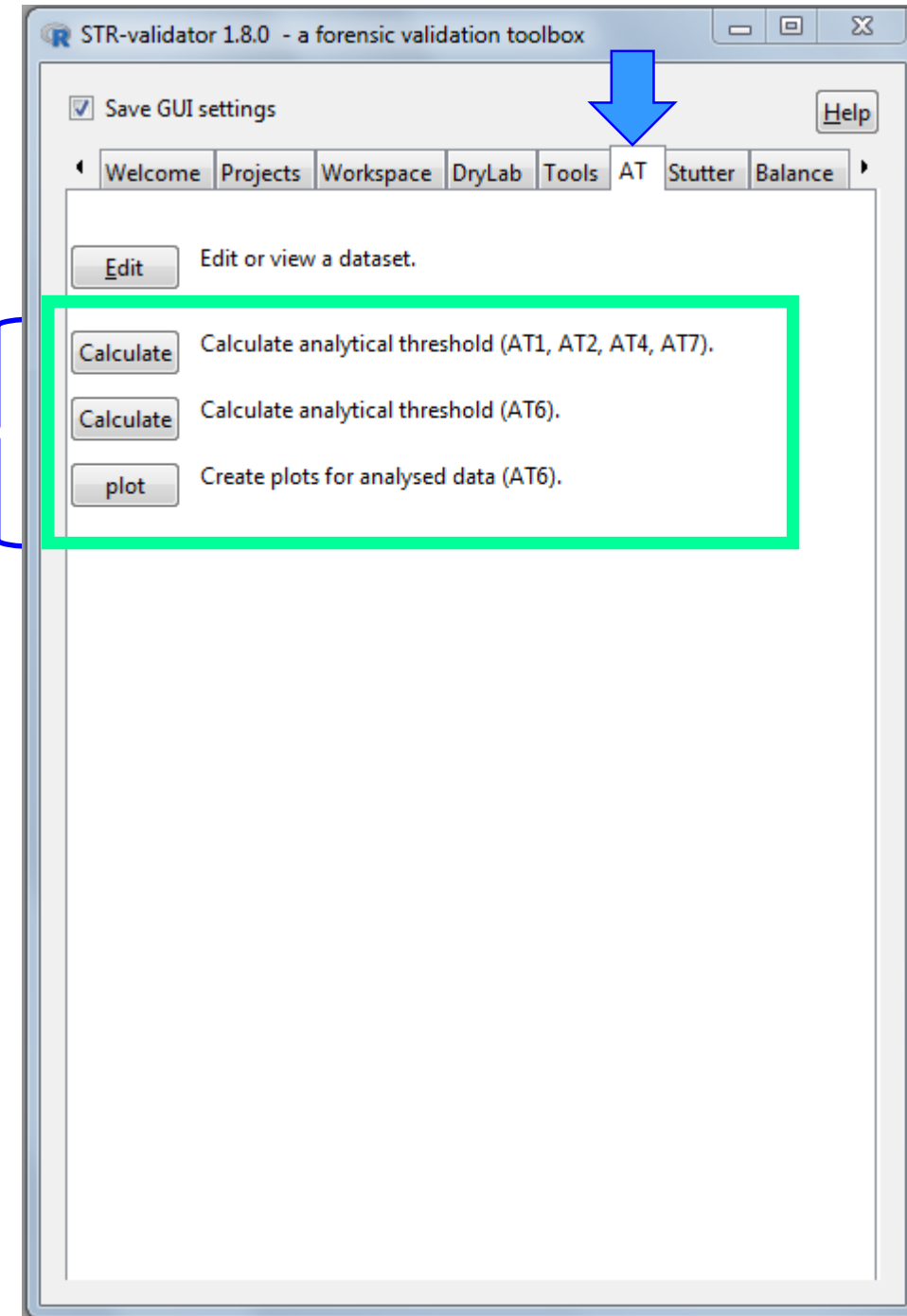
The Analytical Threshold

Experimental Design

- Sensitivity study data
- Three mostly heterozygous samples selected
- DNA input amounts ranged from:
2.0 ng, 1.0 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, and 0.031 ng
- Amplified in triplicate with positive and negative controls
- Analyzed at 1 RFU in all dye channels
- Export the `_SamplePlotSizingTable.txt` from GeneMapper with at least the following information:
“Dye/Sample Peak”, “Sample.FileName”, “Marker”, “Allele”, “Height”, and “Data.Point”.

The Analytical Threshold

- ❖ Different methods for analytical threshold calculations
- ❖ Users can plot the analyzed data
- ❖ Methods 1, 2, 4, and 7 are calculated simultaneously (**except for method 6**)
- ❖ Masked data used to estimate the AT can be exported for manual calculations to confirm the result



TECHNICAL NOTE

CRIMINALISTICS

Joli Bregu,¹ B.S.; Danielle Conklin,¹ M.S.; Elisse Coronado,¹ M.S.; Margaret Terrill,¹ M.S.F.S.; Robin W. Cotton,¹ Ph.D.; and Catherine M. Grgicak,¹ Ph.D.

Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis^{*,†}

ABSTRACT: Determining appropriate analytical thresholds (ATs) for forensic DNA analysis is critical to maximize allele detection. In this study, six methods to determine ATs for forensic DNA purposes were examined and compared. Four of the methods rely on analysis of the baseline noise of a number of negatives, while two utilize the relationship between relative fluorescence unit signal and DNA input in the polymerase chain reaction (PCR) derived from a dilution series ranging from 1 to 0.06 ng. Results showed that when a substantial mass of DNA (i.e., >1 ng) was amplified, the baseline noise increased, suggesting the application of an AT derived from negatives should only be applied to samples with low levels of DNA. Further, the number and intensity of these noise peaks increased with increasing injection times, indicating that to maximize the ability to detect alleles, ATs should be validated for each post-PCR procedure employed.

KEYWORDS: forensic science, minimum distinguishable signal, minimum discernible signal, forensic DNA analysis, analytical threshold, signal to noise

*AT1

*AT2

*AT4

*AT6



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Research paper

Probabilistic characterisation of baseline noise in STR profiles



Ullrich J. Mönich^{a,*}, Ken Duffy^d, Muriel Médard^a, Viveck Cadambe^c, Lauren E. Alfonse^b, Catherine Grgicak^b

^aResearch Laboratory of Electronics, Massachusetts Institute of Technology, United States

^bBiomedical Forensic Sciences, Boston University School of Medicine, United States

^cDepartment of Electrical Engineering, Pennsylvania State University, United States

^dHamilton Institute, Maynooth University – National University of Ireland Maynooth, Ireland

ARTICLE INFO

Article history:

Received 27 February 2015

Received in revised form 22 May 2015

Accepted 2 July 2015

Available online 8 July 2015

Keywords:

Short tandem repeat

Noise

Peak height

Distribution

G-test

Stutter

ABSTRACT

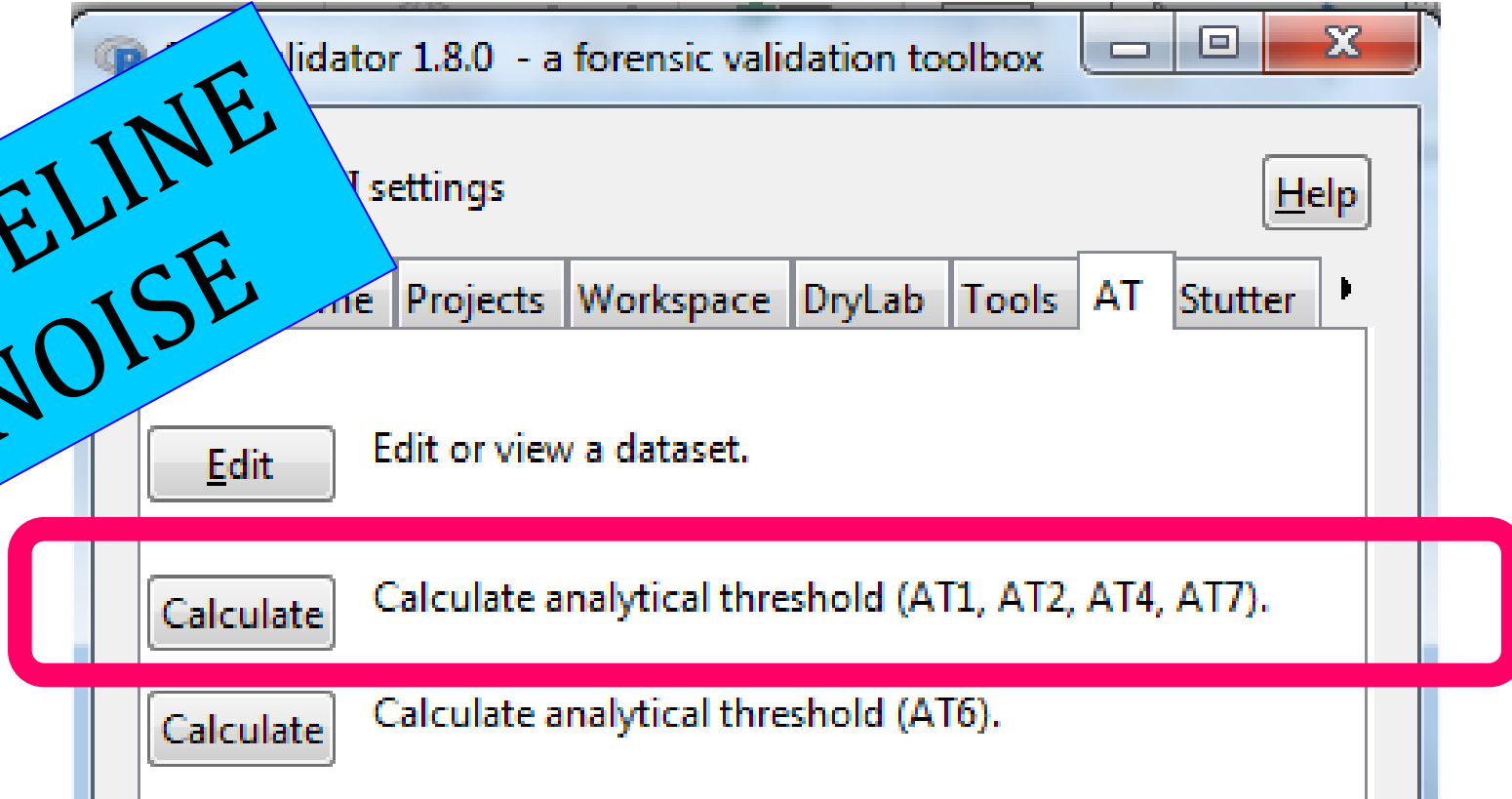
There are three dominant contributing factors that distort short tandem repeat profile measurements, two of which, stutter and variations in the allelic peak heights, have been described extensively. Here we characterise the remaining component, baseline noise. **A probabilistic characterisation of the non-allelic noise peaks is not only inherently useful for statistical inference but is also significant for establishing a detection threshold.** We do this by analysing the data from 643 single person profiles for the Identifiler Plus kit and 303 for the PowerPlex 16 HS kit. This investigation reveals that although the dye colour is a significant factor, it is not sufficient to have a per-dye colour description of the noise. Furthermore, we show that at a per-locus basis, out of the Gaussian, log-normal, and gamma distribution classes, baseline noise is best described by log-normal distributions and provide a methodology for setting an analytical threshold based on that deduction. In the PowerPlex 16 HS kit, we observe evidence of significant stutter at two repeat units shorter than the allelic peak, which has implications for the definition of baseline noise and signal interpretation. **In general, the DNA input mass has an influence on the noise distribution. Thus, it is advisable to study noise and, consequently, to infer quantities like the analytical threshold from data with a DNA input mass comparable to the DNA input mass of the samples to be analysed.**

© 2015 Elsevier Ireland Ltd. All rights reserved.

*AT7

What do these AT methods mean?

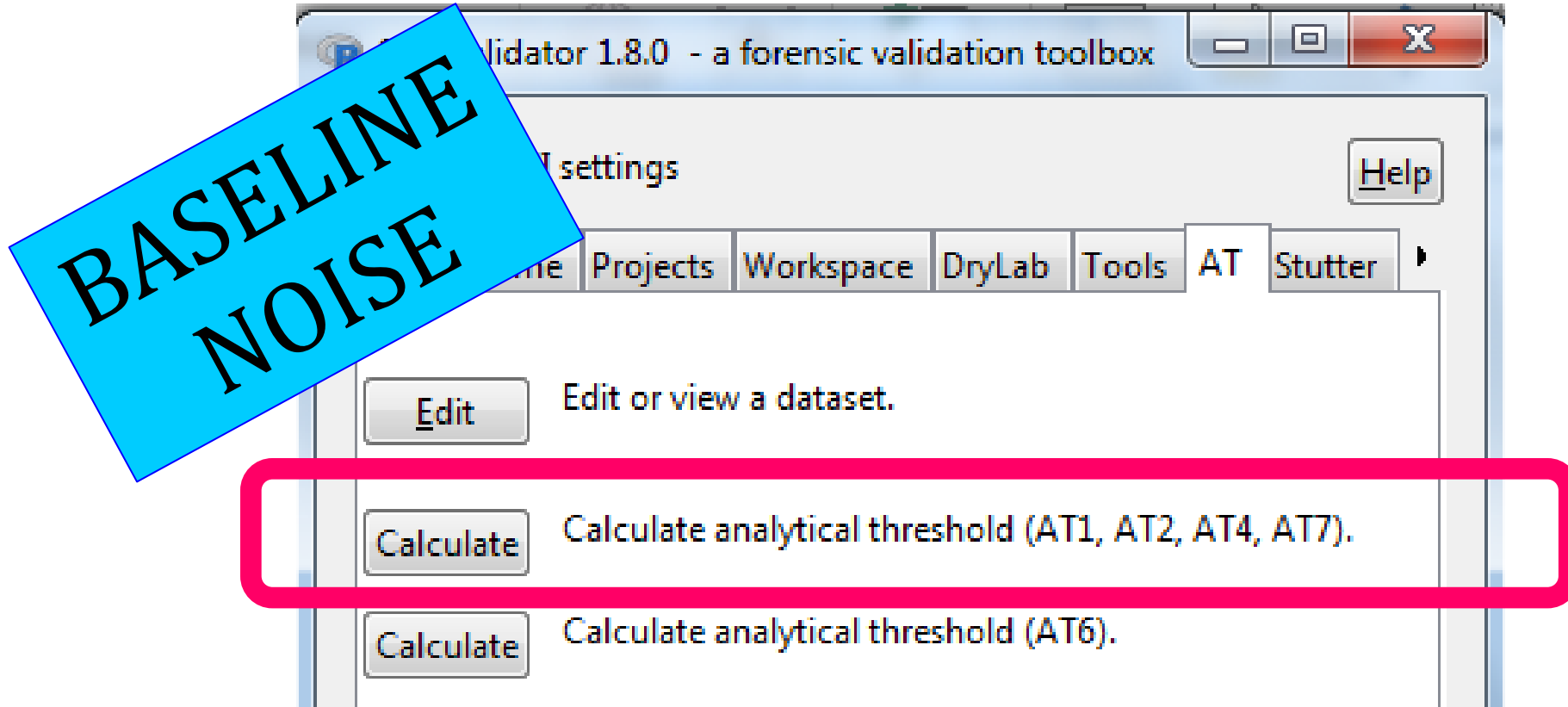
**BASELINE
NOISE**



Analysis of AT1, AT2, AT4, and AT7 in *STR-validator*

DNA Dilution Series Data	AT1	AT2	AT4	AT7
Blue	64	106	69	53
Green	68	137	73	50
Yellow	53	91	58	34
Red	57	107	61	40
Purple	55	107	60	36

The Analytical Threshold (Methods 1, 2, 4, 7)



- ❖ Negative control samples
- ❖ Positive control samples

The Analytical Threshold

Methods 1, 2, 4, and 7



1. Create an Analysis Method with peak amplitude thresholds = 1 RFU in all dye channels

2. Import DNA sensitivity data into GeneMapper

Analysis Method Editor

General Allele **Peak Detector** Peak Quality SQ & GQ Settings

Peak Detection Algorithm: Advanced

Ranges

Analysis: Full Range Sizing: All Sizes

Start Pt: 0 Start Size: 0

Stop Pt: 10000 Stop Size: 1000

Smoothing and Baseline

Smoothing: None Light Heavy

Baseline Window: 51 pts

Size Calling Method

2nd Order Least Squares

3rd Order Least Squares

Cubic Spline Interpolation

Local Southern Method

Global Southern Method

Peak Detection

Peak Amplitude Thresholds:

B: 1 R: 1

G: 1 P: 1

Y: 1 O: 1

Min. Peak Half Width: 2 pts

Polynomial Degree: 3

Peak Window Size: 15 pts

Slope Threshold

Peak Start: 0.0

Peak End: 0.0

Normalization

Use Normalization, if applicable

Factory Defaults

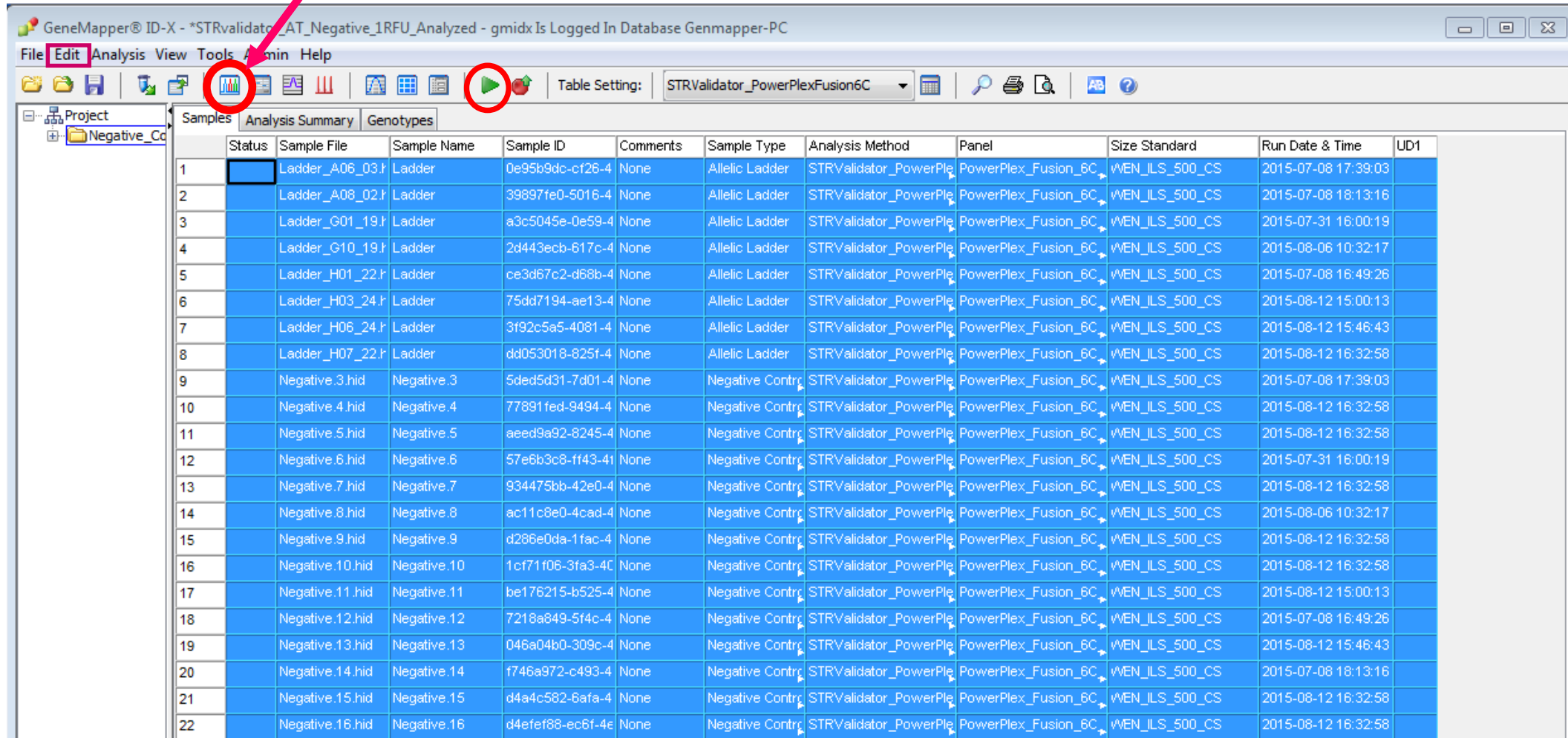
Save As Save Cancel Help

The Analytical Threshold (*Methods 1, 2, 4, & 7*)

3. Analyze the sample

4. Select all samples in the Samples table

5. Open the Samples Plot window



GeneMapper® ID-X - *STRvalidator_AT_Negative_1RFU_Analyzed - gmidx Is Logged In Database Genmapper-PC

File Edit Analysis View Tools Admin Help

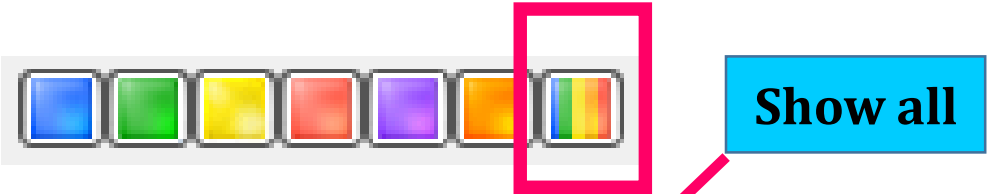
Table Setting: STRValidator_PowerPlexFusion6C

Project: Negative_Cc

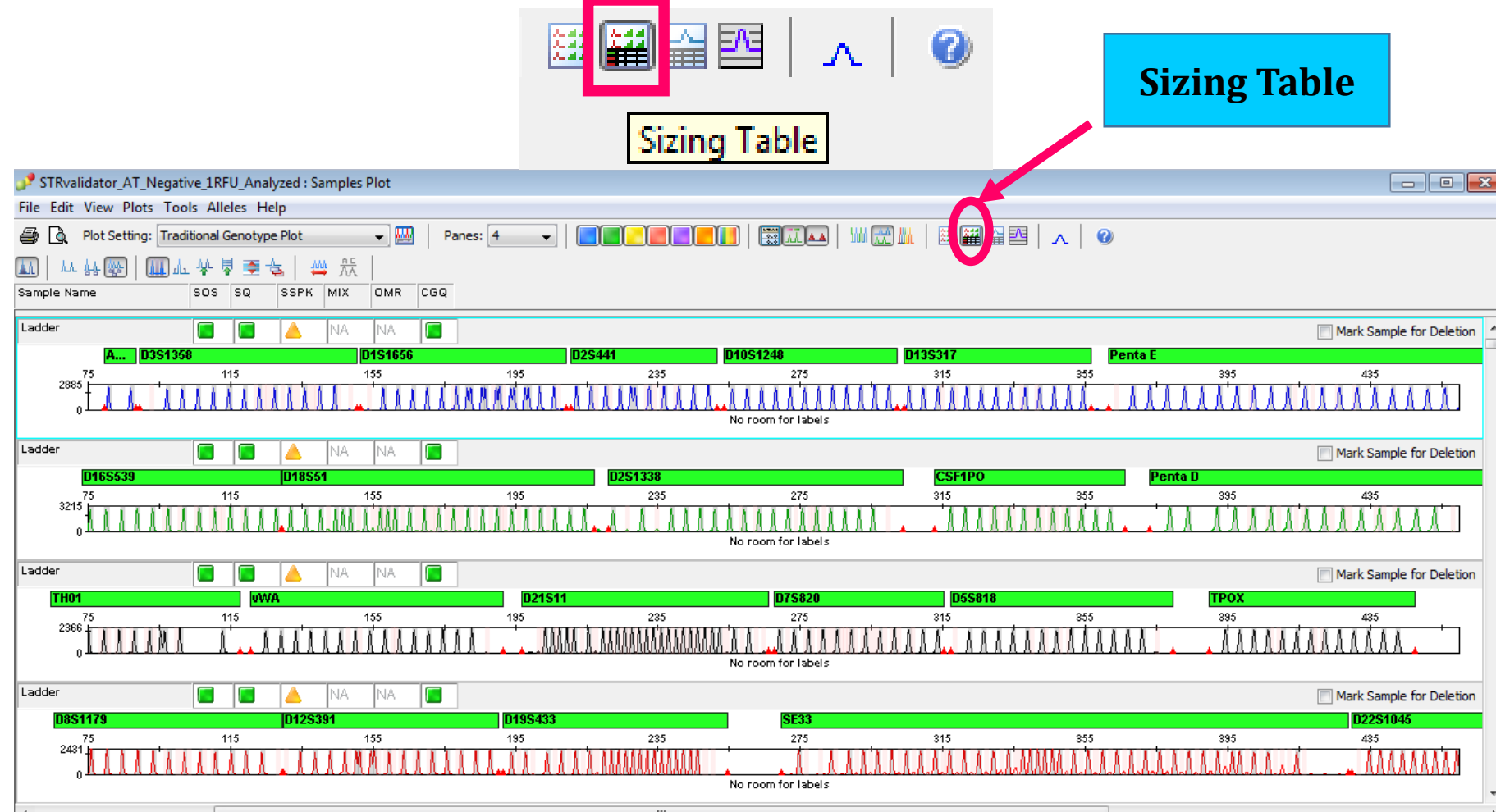
	Status	Sample File	Sample Name	Sample ID	Comments	Sample Type	Analysis Method	Panel	Size Standard	Run Date & Time	UD1
1		Ladder_A06_03.f	Ladder	0e95b9dc-cf26-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 17:39:03	
2		Ladder_A08_02.f	Ladder	39897fe0-5016-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 18:13:16	
3		Ladder_G01_19.f	Ladder	a3c5045e-0e59-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-31 16:00:19	
4		Ladder_G10_19.f	Ladder	2d443ecb-617c-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-06 10:32:17	
5		Ladder_H01_22.f	Ladder	ce3d67c2-d68b-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 16:49:26	
6		Ladder_H03_24.f	Ladder	75dd7194-ae13-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 15:00:13	
7		Ladder_H06_24.f	Ladder	3f92c5a5-4081-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 15:46:43	
8		Ladder_H07_22.f	Ladder	dd053018-825f-4	None	Allelic Ladder	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
9		Negative.3.hid	Negative.3	5ded5d31-7d01-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 17:39:03	
10		Negative.4.hid	Negative.4	77891fed-9494-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
11		Negative.5.hid	Negative.5	aeed9a92-8245-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
12		Negative.6.hid	Negative.6	57e6b3c8-f143-41	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-31 16:00:19	
13		Negative.7.hid	Negative.7	934475bb-42e0-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
14		Negative.8.hid	Negative.8	ac11c8e0-4cad-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-06 10:32:17	
15		Negative.9.hid	Negative.9	d286e0da-1fac-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
16		Negative.10.hid	Negative.10	1cf71f06-3fa3-4C	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
17		Negative.11.hid	Negative.11	be176215-b525-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 15:00:13	
18		Negative.12.hid	Negative.12	7218a849-5f4c-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 16:49:26	
19		Negative.13.hid	Negative.13	046a04b0-309c-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 15:46:43	
20		Negative.14.hid	Negative.14	f746a972-c493-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-07-08 18:13:16	
21		Negative.15.hid	Negative.15	d4a4c582-6afa-4	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	
22		Negative.16.hid	Negative.16	d4efef88-ec61-4e	None	Negative Contr	STRValidator_PowerPlex	PowerPlex_Fusion_6C	WEN_ILS_500_CS	2015-08-12 16:32:58	

The Analytical Threshold *Method 1, 2, 4, and 7*

5. Select to show all dyes



6. Show the Sizing Table



7. The Sizing Table must contain all the following columns

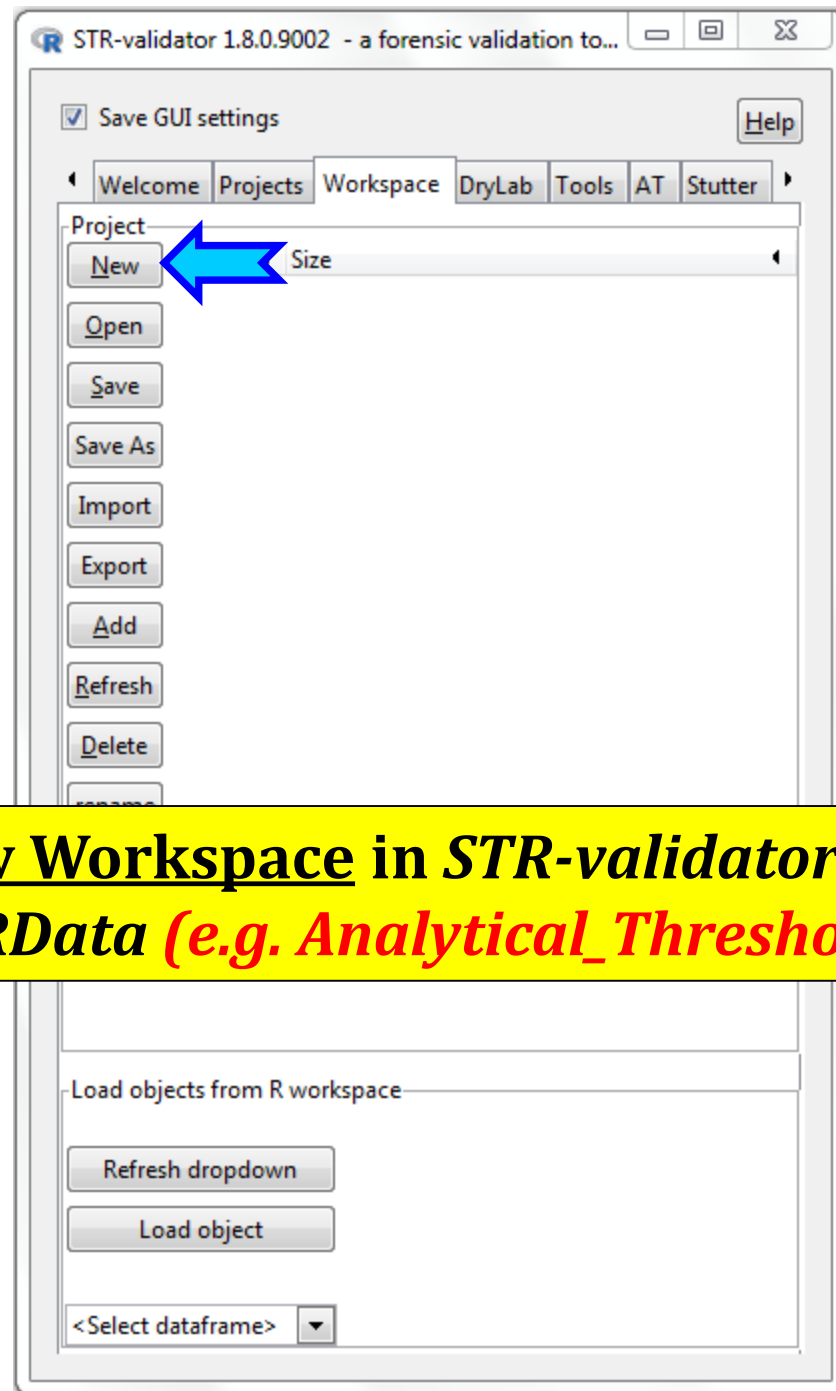
	Dye/Sample Peak	Sample File Name	Marker	Allele	Size	Height	Area	Data Point
■	B,1	Ladder_A06_03.f				4	30	9
■	B,2	Ladder_A06_03.f				4	39	32
■	B,3	Ladder_A06_03.f				3	18	50
■	B,4	Ladder_A06_03.f				4	22	64
■	B,5	Ladder_A06_03.f				2	14	75
■	B,6	Ladder_A06_03.f				3	8	88
■	B,7	Ladder_A06_03.f				2	8	99
■	B,8	Ladder_A06_03.f				3	17	125
■	B,9	Ladder_A06_03.f				3	25	137
■	B,10	Ladder_A06_03.f				3	20	148
■	B,11	Ladder_A06_03.f				4	28	163
■	B,12	Ladder_A06_03.f				4	28	176
■	B,13	Ladder_A06_03.f				5	31	186

- ❖ Dye/Sample Peak
- ❖ Marker
- ❖ Allele
- ❖ Height
- ❖ Data Point

8. Export the Sizing Table

The screenshot shows the 'STRvalidator_AT_Negative_1RFU_Analyzed : Samples Plot' window. The 'File' menu is open, and the 'Export Table' option is highlighted with a blue box and a blue arrow pointing to it from the left. The menu items are: Save Panel (Ctrl+Shift+S), Page Setup, Print (Ctrl+P), Print Preview, Export Table (Ctrl+E), and Close Plot Window (Escape). The background shows a plot with a peak at 155 labeled 'D1S1656'.

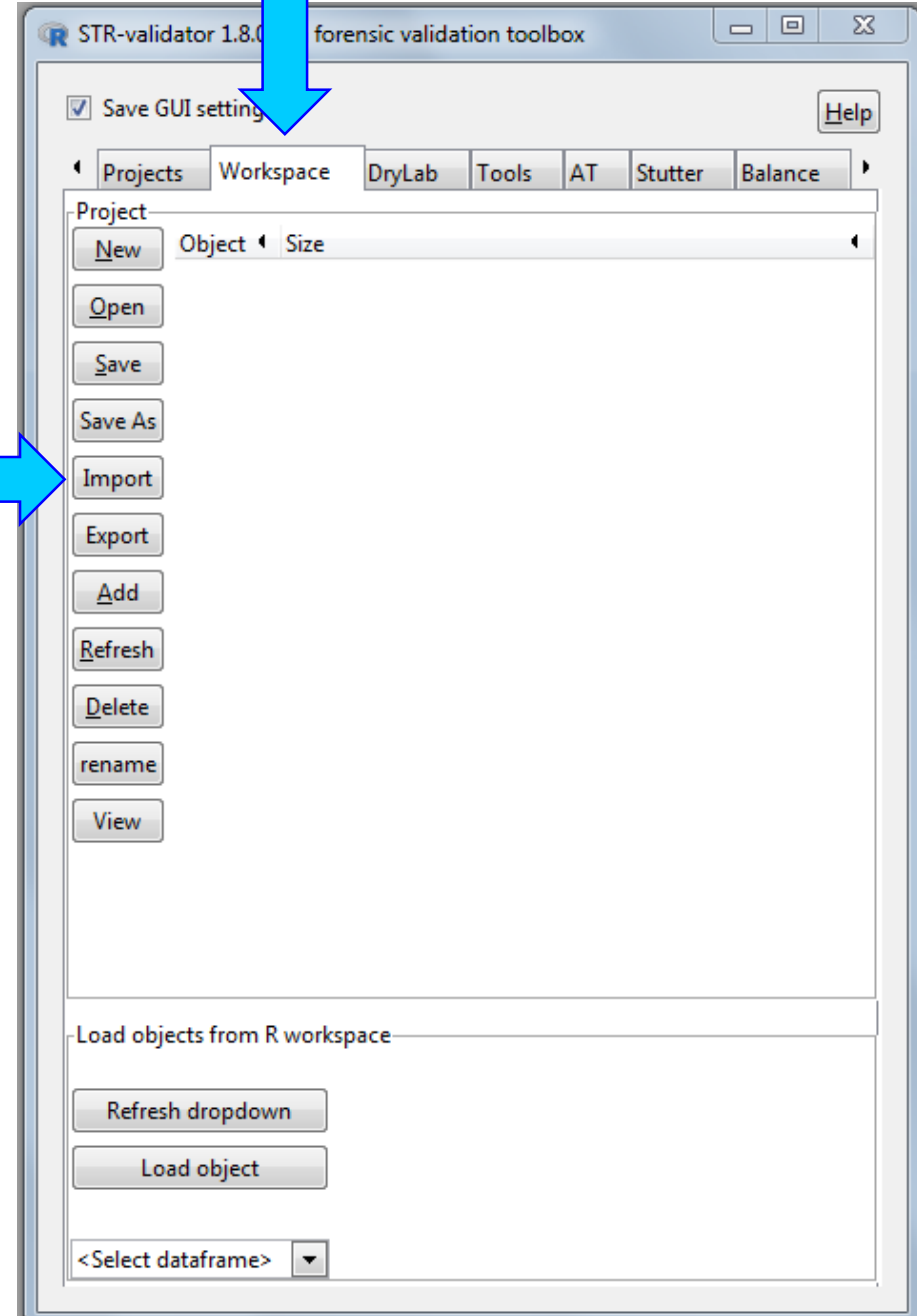
File	Edit	View	Plots	Tools	Alleles	Help
Save Panel				Ctrl+Shift+S	Plot	
Page Setup						
Print				Ctrl+P		
Print Preview						
Export Table				Ctrl+E		
Close Plot Window				Escape		



Open a New Workspace in *STR-validator* GUI and save as *Name.RData (e.g. Analytical_Threshold_Analysis)*

Import DNA Dilution Sizing Table and Reference Data

File_**SamplePlotSizingTable**.txt



Import Data

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset

Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Analytical_Threshold browse

Select a directory... browse

Options

Save file name

Save file time stamp

Delimiter:
TAB

NA strings (separated by comma):
NA,,

Auto trim samples

Auto slim repeated columns

+ Multiple files options

+ Trim options

+ Slim options

Save options

Name:
DNA_Dil_Set

Import

Import Reference

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset

Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Analytical_Threshold browse

Select a directory... browse

Options

Save file name

Save file time stamp

Delimiter:
TAB

NA strings (separated by comma):
NA,,

Auto trim samples

Auto slim repeated columns

+ Multiple files options

+ Trim options

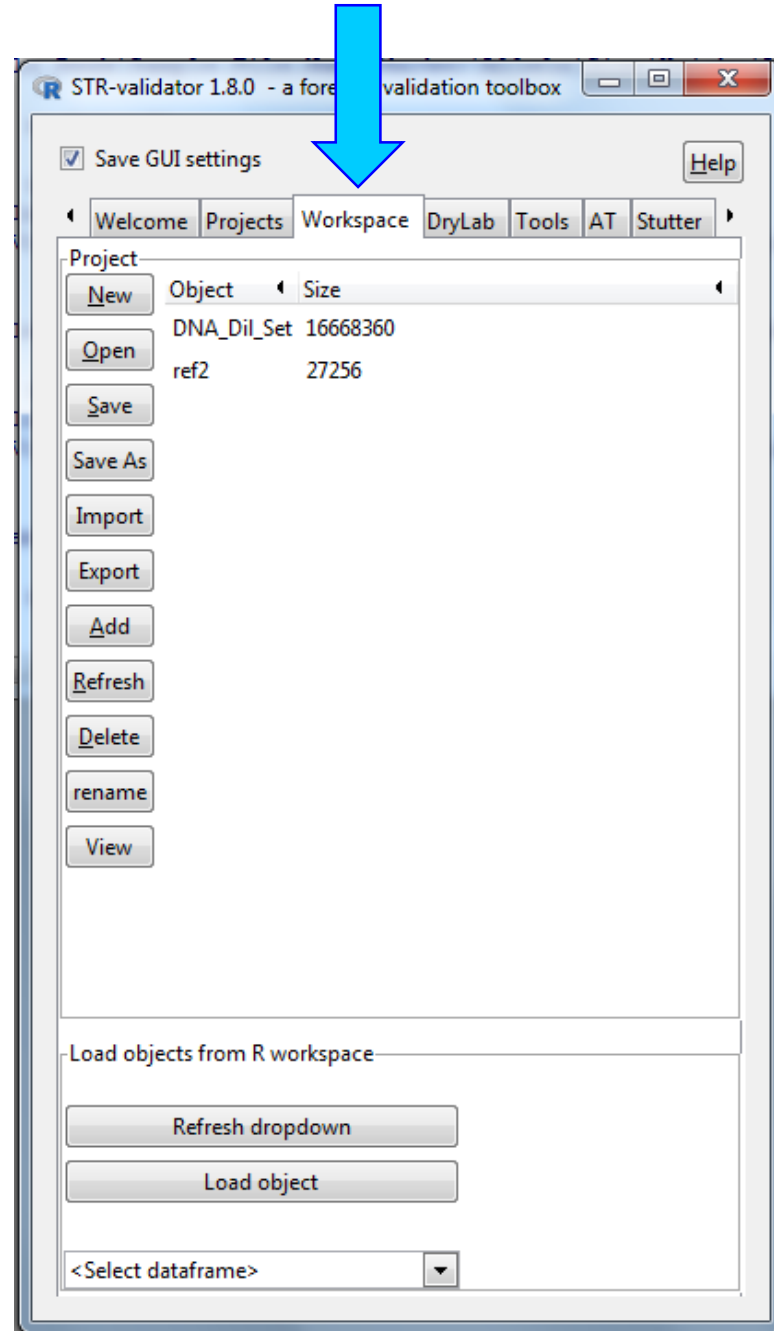
+ Slim options

Save options

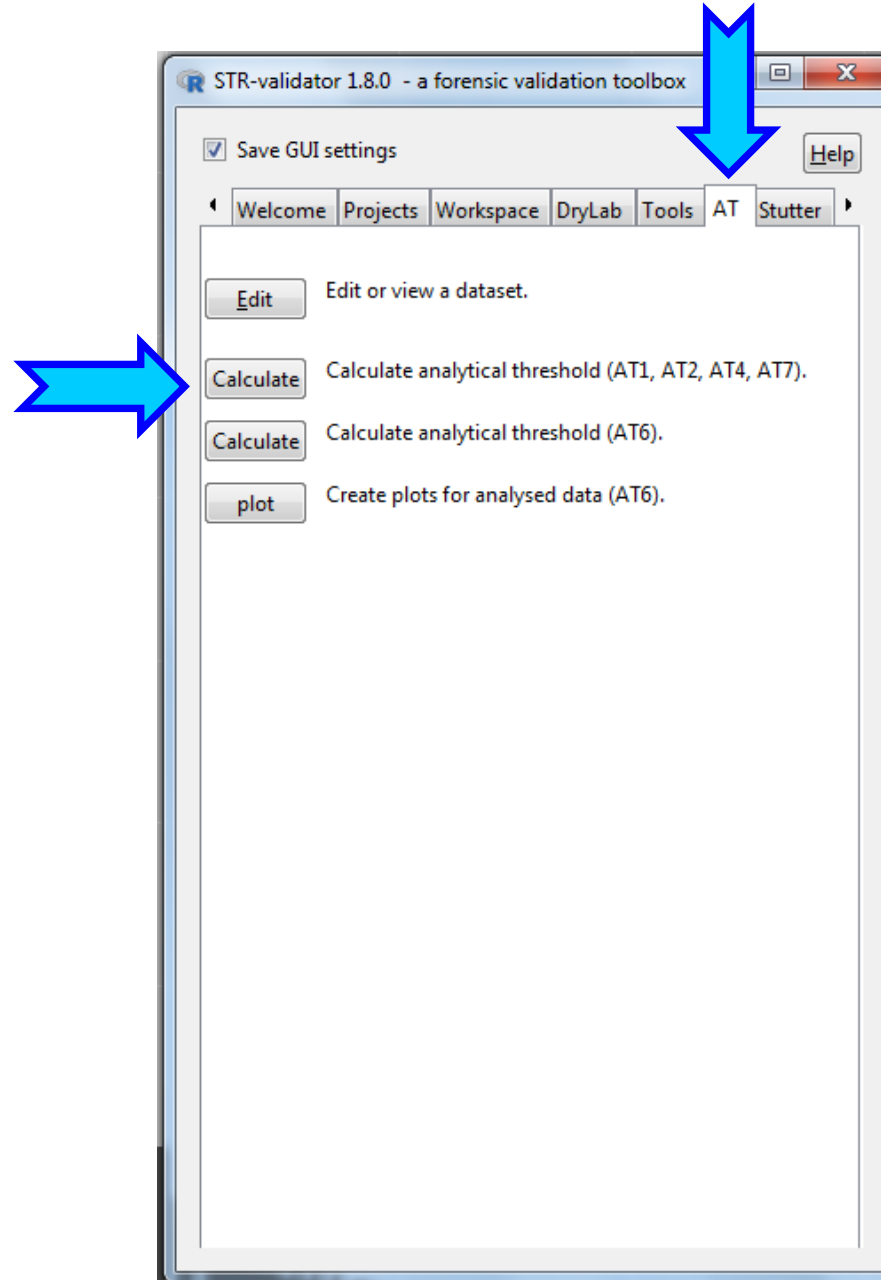
Name:
ref2

Import

Check Your Workspace



Calculate Analytical Thresholds



Calculate ATs

The screenshot shows the 'Calculate analytical threshold' window. Annotations include blue arrows pointing to the 'Select dataset' and 'Select reference dataset' fields, a red oval around the 'Check subsetting' button, and a blue box around the 'Options' section.

Calculate analytical threshold

Save GUI settings Help

Datasets

Select dataset: DNA_Dil_Set 66 samples.

Select reference dataset: ref2 4 samples.

Check subsetting

Select the kit used: Fusion 6C

Options

Ignore case

Add word boundaries

Mask high peaks Mask all peaks above (RFU): 200

Mask sample alleles Range (data points) around known alleles: 45

Mask sample alleles per dye channel

Mask ILS peaks Range (data points) around known peak: 20

Confidence level 'k' (AT1, AT7): 3

Percentile rank threshold (AT2): 0.99

Upper confidence 'alpha' (AT4): 0.01

Prepare data and check masking

Prepare and mask <Select sample> Save plot

Save as

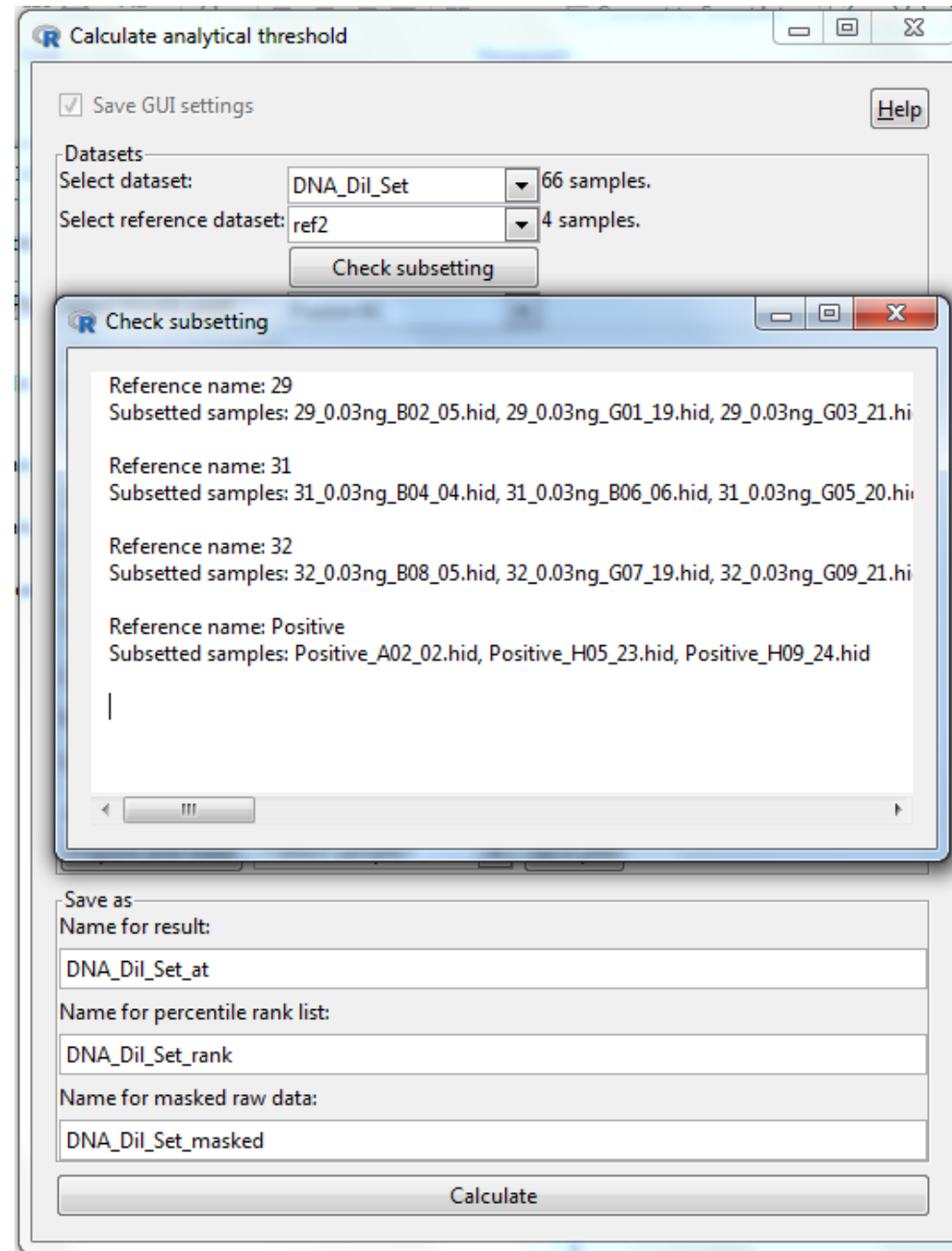
Name for result: DNA_Dil_Set_at

Name for percentile rank list: DNA_Dil_Set_rank

Name for masked raw data: DNA_Dil_Set_masked

Calculate

Check Subsetting



Mask Peaks

- ❖ High peaks
- ❖ Area around samples alleles
- ❖ ILS peaks

Calculate analytical threshold

Save GUI settings Help

Datasets
Select dataset: DNA_Dil_Set 66 samples.
Select reference dataset: ref2 4 samples.
Check subsetting
Select the kit used: Fusion 6C

Options

Ignore case

Add word boundaries

Mask high peaks Mask all peaks above (RFU): 200

Mask sample alleles Range (data points) around known alleles: 45

Mask sample alleles per dye channel

Mask ILS peaks Range (data points) around known peak: 20

Confidence level 'k' (AT1, AT7): 3

Percentile rank threshold (AT2): 0.99

Upper confidence 'alpha' (AT4): 0.01

Prepare data and check masking

Prepare and mask <Select sample> Save plot

Save as

Name for result:
DNA_Dil_Set_at

Name for percentile rank list:
DNA_Dil_Set_rank

Name for masked raw data:
DNA_Dil_Set_masked

Calculate

Options

Ignore case

Add word boundaries

Mask high peaks Mask all peaks above (RFU): 200

Mask sample alleles Range (data points) around known alleles: 45

Mask sample alleles per dye channel

Mask ILS peaks Range (data points) around known peak: 20

Confidence level 'k' (AT1, AT7): 3

Percentile rank threshold (AT2): 0.99

Upper confidence 'alpha' (AT4): 0.01

Manually Inspect the Masking

- ❖ Prepare and Mask
- ❖ Choose a Sample
- ❖ Save Plot



Calculate analytical threshold

Save GUI settings Help

Datasets

Select dataset: DNA_Dil_Set 66 samples.

Select reference dataset: ref2 4 samples.

Select the kit used: Fusion 6C

Options

Ignore case

Add word boundaries

Mask high peaks Mask all peaks above (RFU): 200

Mask sample alleles Range (data points) around known alleles: 45

Mask sample alleles per dye channel

Mask ILS peaks Range (data points) around known peak: 20

Confidence level 'k' (AT1, AT7): 3

Percentile rank threshold (AT2): 0.99

Upper confidence 'alpha' (AT4): 0.01

Prepare data and check masking

32_0.5ng_F08_17.hid

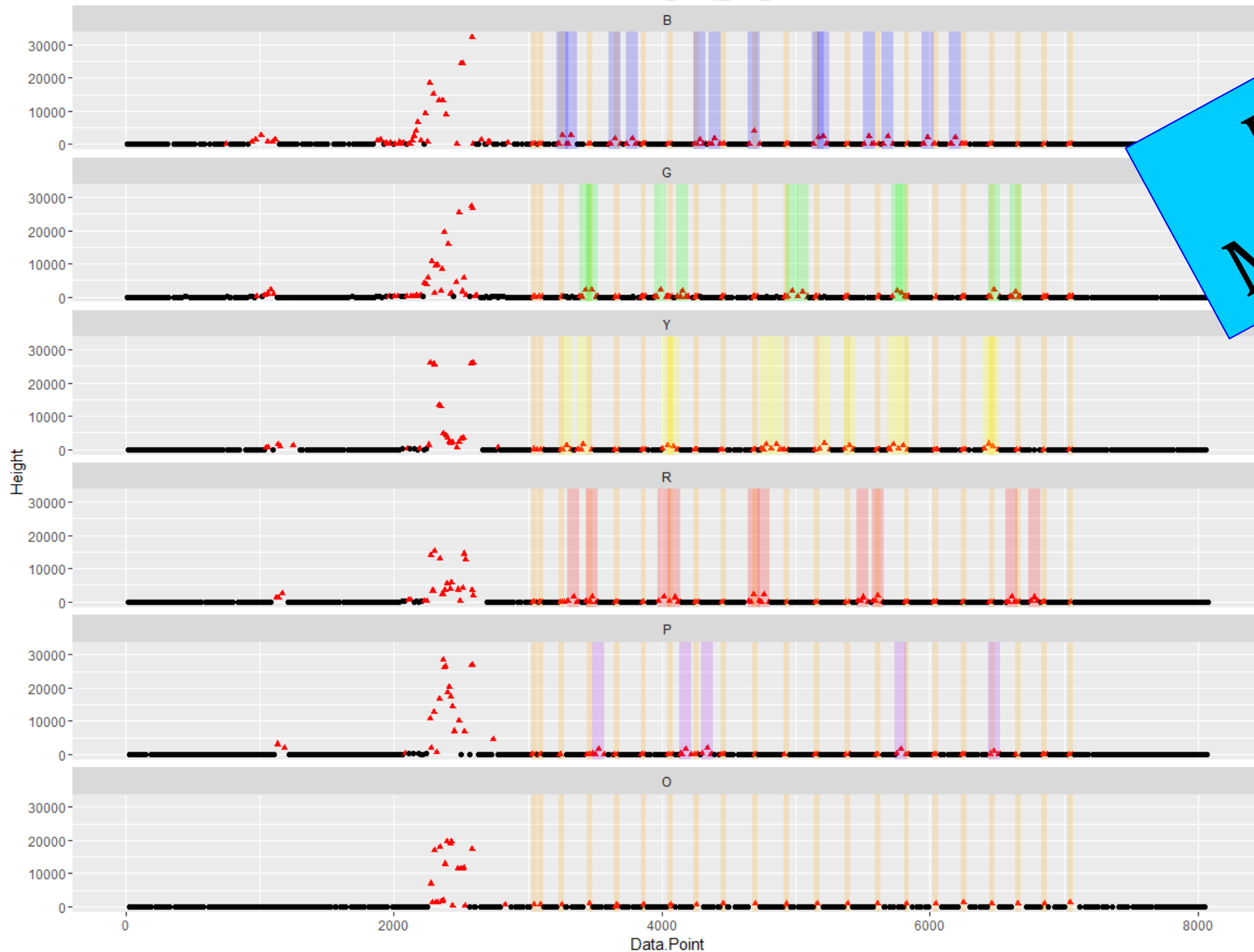
Save as:

Name for result:
DNA_Dil_Set_at

Name for percentile rank list:
DNA_Dil_Set_rank

Name for masked raw data:
DNA_Dil_Set_masked

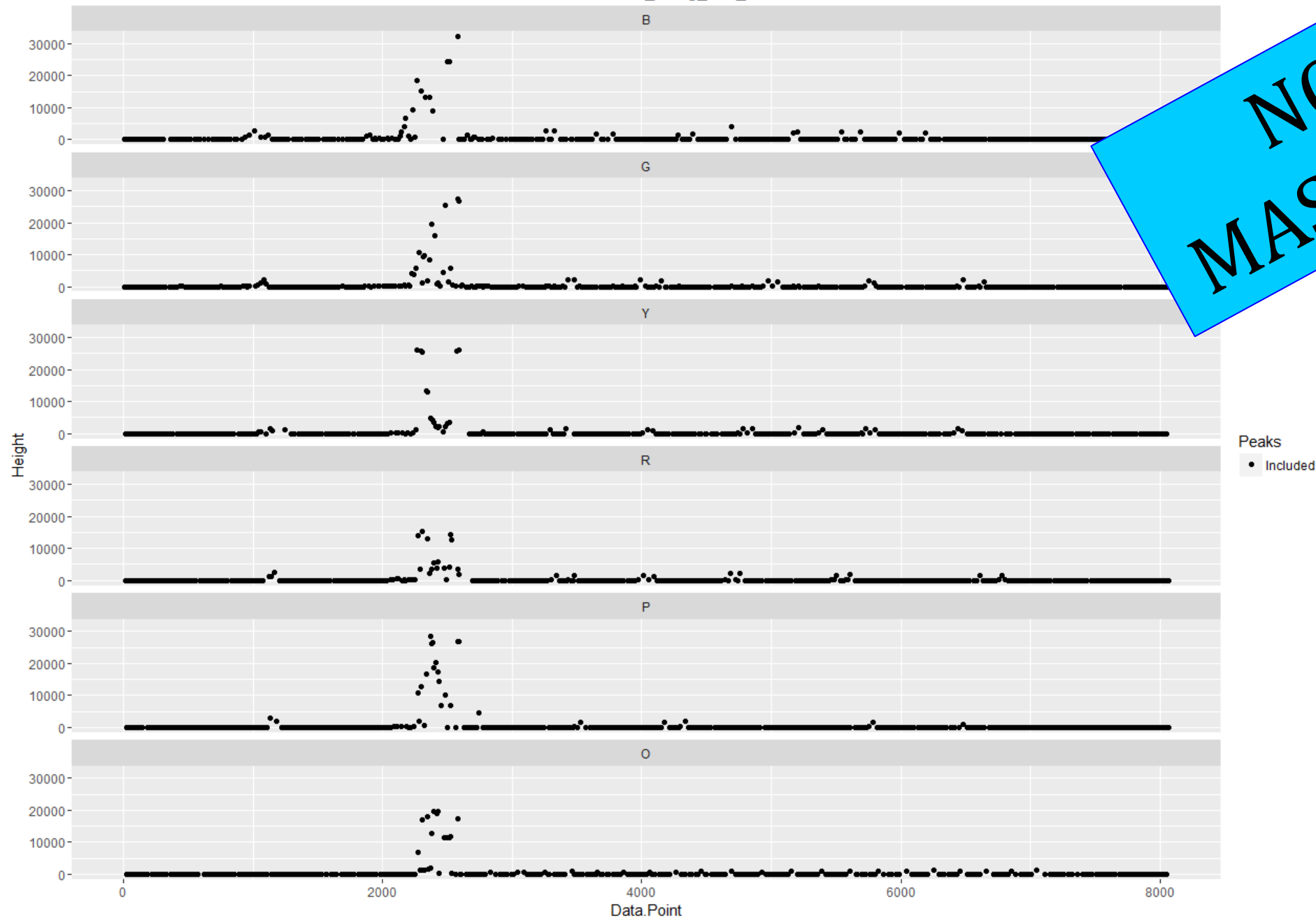
Masked data for 32_0.5ng_F08_17.hid



DATA MASKING

Peaks
● Included
▲ Excluded

Masked data for 32_0.5ng_F08_17.hid



NO MASKING

Calculate analytical threshold

Save GUI settings Help

Datasets

Select dataset: DNA_Dil_Set 66 samples.

Select reference dataset: ref2 4 samples.

Check subsetting

Select the kit used: Fusion 6C

Options

Ignore case

Add word boundaries

Mask high peaks Mask all peaks above (RFU): 200

Mask sample alleles Range (data points) around known alleles: 45

Mask sample alleles per dye channel

Mask ILS peaks Range (data points) around known peak: 20

Confidence level 'k' (AT1, AT7): 3

Percentile rank threshold (AT2): 0.99

Upper confidence 'alpha' (AT4): 0.01

Prepare data and check masking

Prepare and mask 32_0.5ng_F08_17.hid Save plot

Save as

Name for result:
DNA_Dil_Set_at

Name for percentile rank list:
DNA_Dil_Set_rank

Name for masked raw data:
DNA_Dil_Set_masked

Calculate

Saved in the Workspace

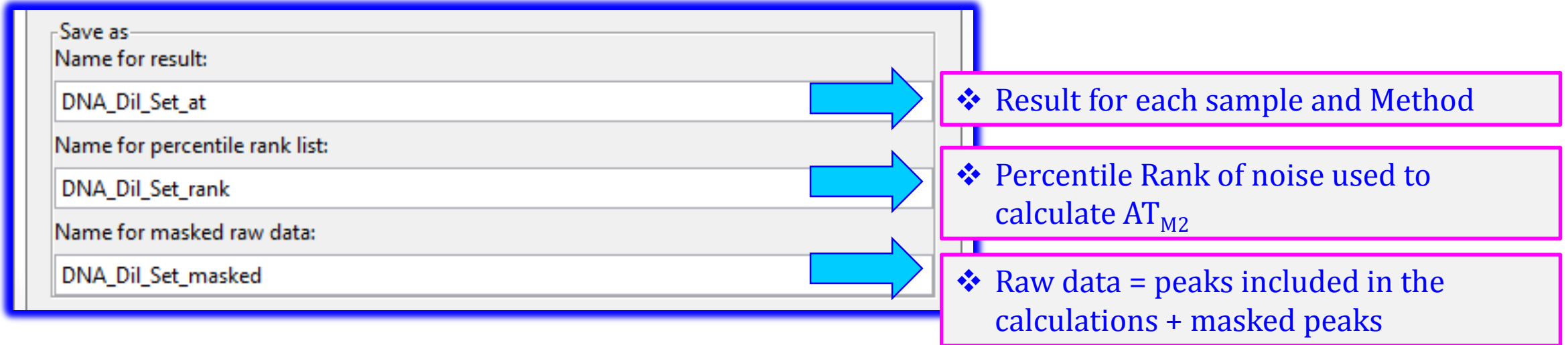
Save as

Name for result:
DNA_Dil_Set_at

Name for percentile rank list:
DNA_Dil_Set_rank

Name for masked raw data:
DNA_Dil_Set_masked

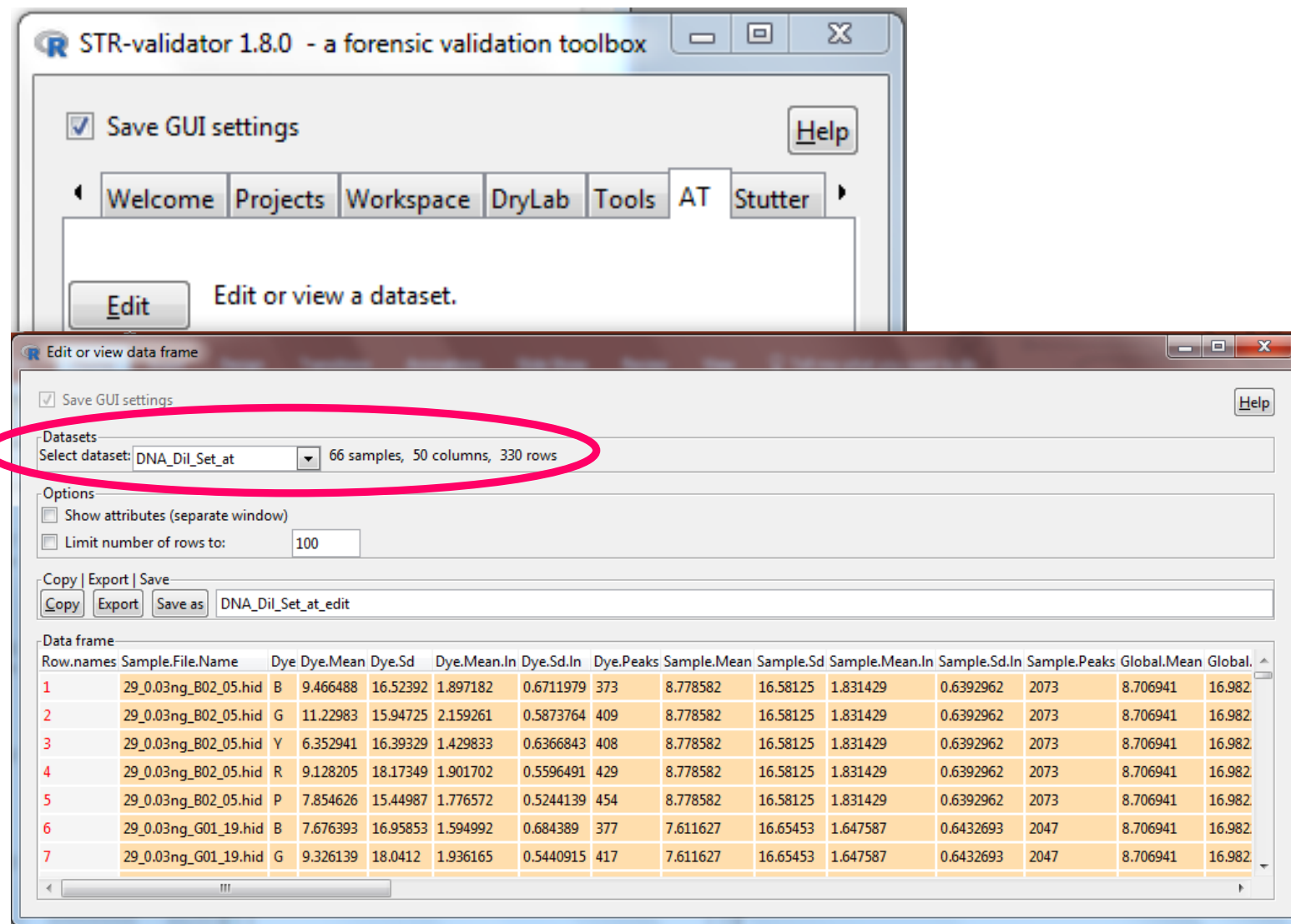
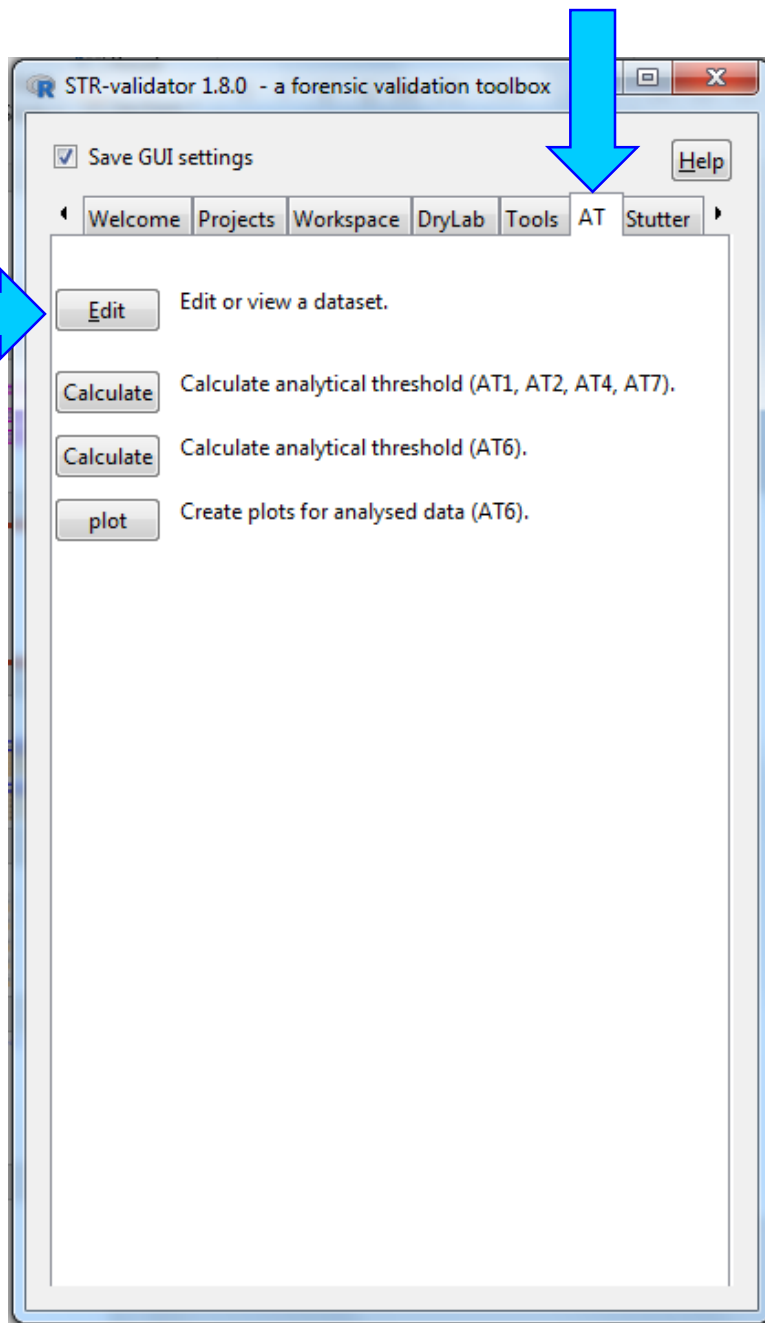
Output of Analysis is a List of Three Data Frames



The image shows a software interface with three input fields for naming data frames. Each field has a cyan arrow pointing to a corresponding description box on the right. The descriptions are enclosed in magenta boxes.

Field Label	Value	Description
Name for result:	DNA_Dil_Set_at	❖ Result for each sample and Method
Name for percentile rank list:	DNA_Dil_Set_rank	❖ Percentile Rank of noise used to calculate AT_{M2}
Name for masked raw data:	DNA_Dil_Set_masked	❖ Raw data = peaks included in the calculations + masked peaks

Results = AT Values for each Sample and Method



AT Results for each sample and Method

The screenshot shows a window titled "Edit or view data frame". At the top, there is a checkbox for "Save GUI settings" and a "Help" button. Below that, the "Datasets" section shows "Select dataset:" with a dropdown menu set to "DNA_Dil_Set_at" and a status indicator "66 samples, 50 columns, 330 rows". The "Options" section includes checkboxes for "Show attributes (separate window)" and "Limit number of rows to:" with a value of "100". A toolbar contains "Copy", "Export", and "Save" buttons, with "Save as" set to "DNA_Dil_Set_at_edit". The "Data frame" section displays a table with columns: Row.names, Sample.File.Name, Dye, Dye.Mean, Dye.Sd, Dye.Mean.In, Dye.Sd.In, Dye.Peaks, Sample.Mean, Sample.Sd, Sample.Mean.In, Sample.Sd.In, Sample.Peaks, Global.Mean, Global.Sd, Global.Mean.In, Global.Sd.In, Global.Peaks. The first two rows are visible, showing data for sample "29_0.03ng_B02_05.hid" with dye "B" and "G".

What do these columns represent?

Row.names	Sample.File.Name	Dye	Dye.Mean	Dye.Sd	Dye.Mean.In	Dye.Sd.In	Dye.Peaks	Sample.Mean	Sample.Sd	Sample.Mean.In	Sample.Sd.In	Sample.Peaks	Global.Mean	Global.Sd	Global.Mean.In	Global.Sd.In	Global.Peaks
8	29_0.03ng_G01_19.hid	Y	5.394402	15.24861	1.234555	0.6504089	393	7.611627	16.65453	1.647587	0.6432693	2047	8.706941	16.98222	1.764262	0.6960548	133403
9	29_0.03ng_G01_19.hid	R	7.861176	14.84295	1.807206	0.4984984	425	7.611627	16.65453	1.647587	0.6432693	2047	8.706941	16.98222	1.764262	0.6960548	133403
10	29_0.03ng_G01_19.hid	P	7.671264	17.71653	1.633735	0.611794	435	7.611627	16.65453	1.647587	0.6432693	2047	8.706941	16.98222	1.764262	0.6960548	133403
11	29_0.03ng_G03_21.hid	B	7.032432	15.85107	1.515458	0.6771645	370	7.135552	15.30042	1.59524	0.6421083	2073	8.706941	16.98222	1.764262	0.6960548	133403
12	29_0.03ng_G03_21.hid	G	9.442353	18.49776	1.921467	0.5745264	425	7.135552	15.30042	1.59524	0.6421083	2073	8.706941	16.98222	1.764262	0.6960548	133403
13	29_0.03ng_G03_21.hid	Y	5.029851	14.16888	1.151514	0.641443	402	7.135552	15.30042	1.59524	0.6421083	2073	8.706941	16.98222	1.764262	0.6960548	133403
14	29_0.03ng_G03_21.hid	R	7.602299	14.45733	1.774031	0.5066734	435	7.135552	15.30042	1.59524	0.6421083	2073	8.706941	16.98222	1.764262	0.6960548	133403
15	29_0.03ng_G03_21.hid	P	6.45805	12.76603	1.57591	0.5391897	441	7.135552	15.30042	1.59524	0.6421083	2073	8.706941	16.98222	1.764262	0.6960548	133403
16	29_0.06ng_C02_08.hid	B	8.052083	17.13569	1.630834	0.7036578	384	7.407837	14.49724	1.673238	0.6162392	2067	8.706941	16.98222	1.764262	0.6960548	133403
17	29_0.06ng_C02_08.hid	G	9.167064	14.70022	1.971133	0.5293034	419	7.407837	14.49724	1.673238	0.6162392	2067	8.706941	16.98222	1.764262	0.6960548	133403
18	29_0.06ng_C02_08.hid	Y	5.25	13.10596	1.273249	0.6138642	404	7.407837	14.49724	1.673238	0.6162392	2067	8.706941	16.98222	1.764262	0.6960548	133403
19	29_0.06ng_C02_08.hid	R	7.494033	12.78332	1.792264	0.4908107	419	7.407837	14.49724	1.673238	0.6162392	2067	8.706941	16.98222	1.764262	0.6960548	133403
20	29_0.06ng_C02_08.hid	P	7.070295	14.34183	1.680471	0.515213	441	7.407837	14.49724	1.673238	0.6162392	2067	8.706941	16.98222	1.764262	0.6960548	133403
21	29_0.06ng_F01_16.hid	B	6.756614	11.97119	1.591708	0.5881045	378	7.292919	14.17878	1.678448	0.5937727	2062	8.706941	16.98222	1.764262	0.6960548	133403
22	29_0.06ng_F01_16.hid	G	9.977833	17.60295	2.032597	0.5308559	406	7.292919	14.17878	1.678448	0.5937727	2062	8.706941	16.98222	1.764262	0.6960548	133403
23	29_0.06ng_F01_16.hid	Y	4.70936	10.7641	1.248168	0.5474188	406	7.292919	14.17878	1.678448	0.5937727	2062	8.706941	16.98222	1.764262	0.6960548	133403
24	29_0.06ng_F01_16.hid	R	7.966746	14.54873	1.837164	0.4993523	421	7.292919	14.17878	1.678448	0.5937727	2062	8.706941	16.98222	1.764262	0.6960548	133403
25	29_0.06ng_F01_16.hid	P	7.022173	14.37759	1.671526	0.5087279	451	7.292919	14.17878	1.678448	0.5937727	2062	8.706941	16.98222	1.764262	0.6960548	133403
26	29_0.06ng_F03_18.hid	B	8.22739	19.4638	1.618652	0.6949557	387	7.652812	15.41487	1.680395	0.6279991	2045	8.706941	16.98222	1.764262	0.6960548	133403
27	29_0.06ng_F03_18.hid	G	10.47666	18.41928	2.027756	0.5926749	407	7.652812	15.41487	1.680395	0.6279991	2045	8.706941	16.98222	1.764262	0.6960548	133403
28	29_0.06ng_F03_18.hid	Y	5.366162	13.18418	1.294675	0.5934581	396	7.652812	15.41487	1.680395	0.6279991	2045	8.706941	16.98222	1.764262	0.6960548	133403
29	29_0.06ng_F03_18.hid	R	7.855037	13.24015	1.835453	0.4938948	407	7.652812	15.41487	1.680395	0.6279991	2045	8.706941	16.98222	1.764262	0.6960548	133403
30	29_0.06na_F03_18.hid	P	6.428571	11.168	1.618244	0.5136017	448	7.652812	15.41487	1.680395	0.6279991	2045	8.706941	16.98222	1.764262	0.6960548	133403

The Analytical Threshold Results

Data frame												
Row.names	Sample.File.Name	Dye	Dye.Mean	Dye.Sd	Dye.Mean.In	Dye.Sd.In	Dye.Peaks	Sample.Mean	Sample.Sd	Sample.Mean.In	Sample.Sd.In	Sample.Peaks
1	29_0.03ng_B02_05.hid	B	9.24359	16.19418	1.87471	0.6724381	390	8.719065	16.19195	1.833805	0.6357667	2182
2	29_0.03ng_B02_05.hid	G	11.10648	15.5694	2.155441	0.5793471	432	8.719065	16.19195	1.833805	0.6357667	2182
3	29_0.03ng_B02_05.hid	Y	6.474299	16.06739	1.45306	0.6510411	428	8.719065	16.19195	1.833805	0.6357667	2182
4	29_0.03ng_B02_05.hid	R	9.006682	17.78084	1.894617	0.5596501	449	8.719065	16.19195	1.833805	0.6357667	2182
5	29_0.03ng_B02_05.hid	P	7.881988	14.99269	1.793957	0.5191003	483	8.719065	16.19195	1.833805	0.6357667	2182

AT for each method per sample

Sample.AT1	Sample.AT2	Sample.AT4	Sample.AT7
57.29491	91	62.02899	42.14452
57.29491	91	62.02899	42.14452
57.29491	91	62.02899	42.14452
57.29491	91	62.02899	42.14452
57.29491	91	62.02899	42.14452

AT for each method per dye per sample

Dye.AT1	Dye.AT2	Dye.AT4	Dye.AT7
57.82614	76	62.7421	49.01
57.81469	117	62.52026	49.08152
54.67646	86	59.5344	30.14997
62.34922	125	67.71485	35.64352
52.86005	89	57.37178	28.53861

AT for each method globally across all samples

AT for each method globally across all samples per dye

Data frame				
Global.Mean	Global.Sd	Global.Mean.In	Global.Sd.In	Global.Peaks
8.679327	16.60921	1.772209	0.6911434	139962

Global.AT1	Global.B.AT1	Global.G.AT1	Global.Y.AT1	Global.R.AT1	Global.P.AT1
58.50696	62.67172	66.48469	52.856	55.30691	54.07239

Global.AT2	Global.B.AT2	Global.G.AT2	Global.Y.AT2	Global.R.AT2	Global.P.AT2
106	102	132	88	102	103

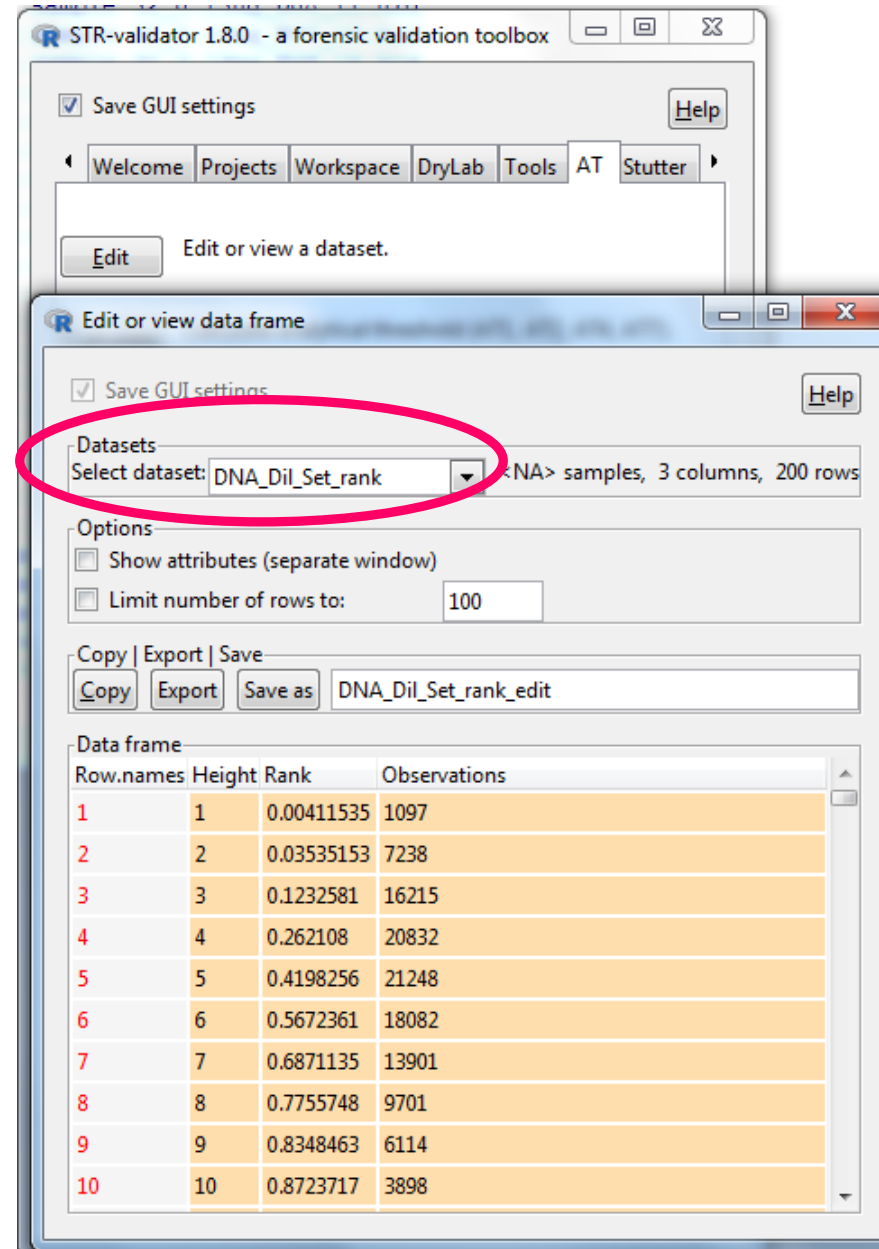
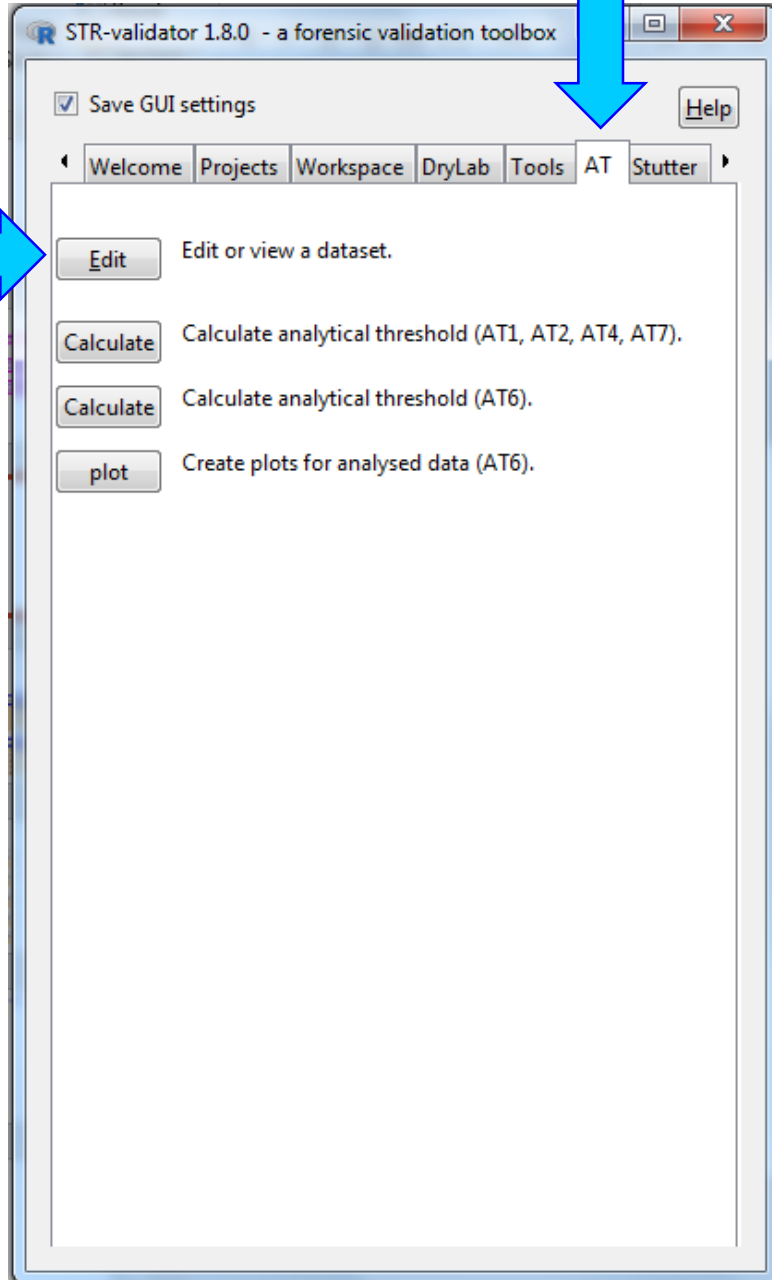
Global.AT4	Global.B.AT4	Global.G.AT4	Global.Y.AT4	Global.R.AT4	Global.P.AT4
47.61019	67.87449	71.85158	57.33405	59.79945	58.5215

Global.AT7	Global.B.AT7	Global.G.AT7	Global.Y.AT7	Global.R.AT7	Global.P.AT7	Total.Samples
46.78859	51.95858	49.77684	36.04511	38.73929	35.49231	66

Summary Statistics after Analyzing 66 Samples at Different DNA input

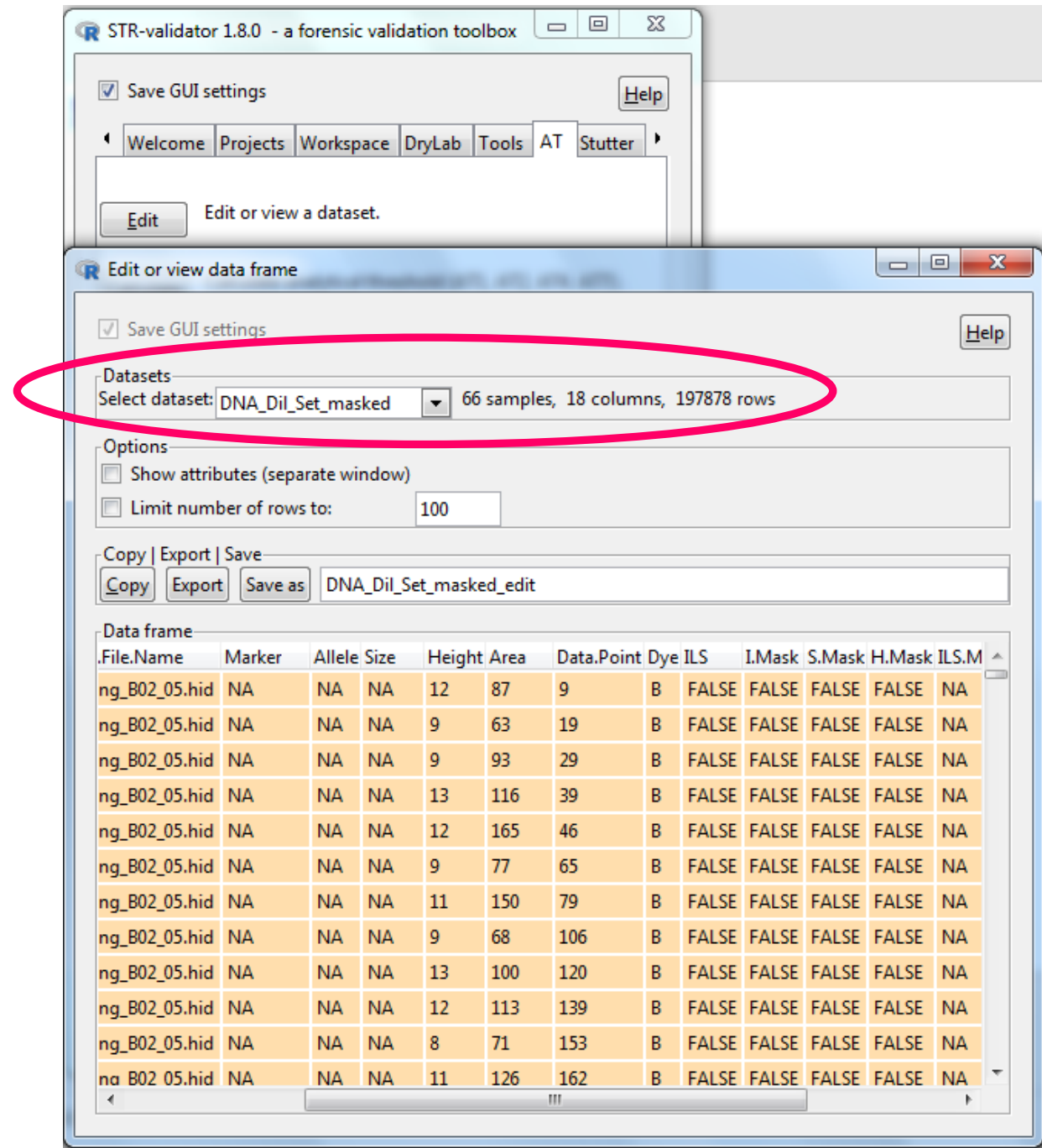
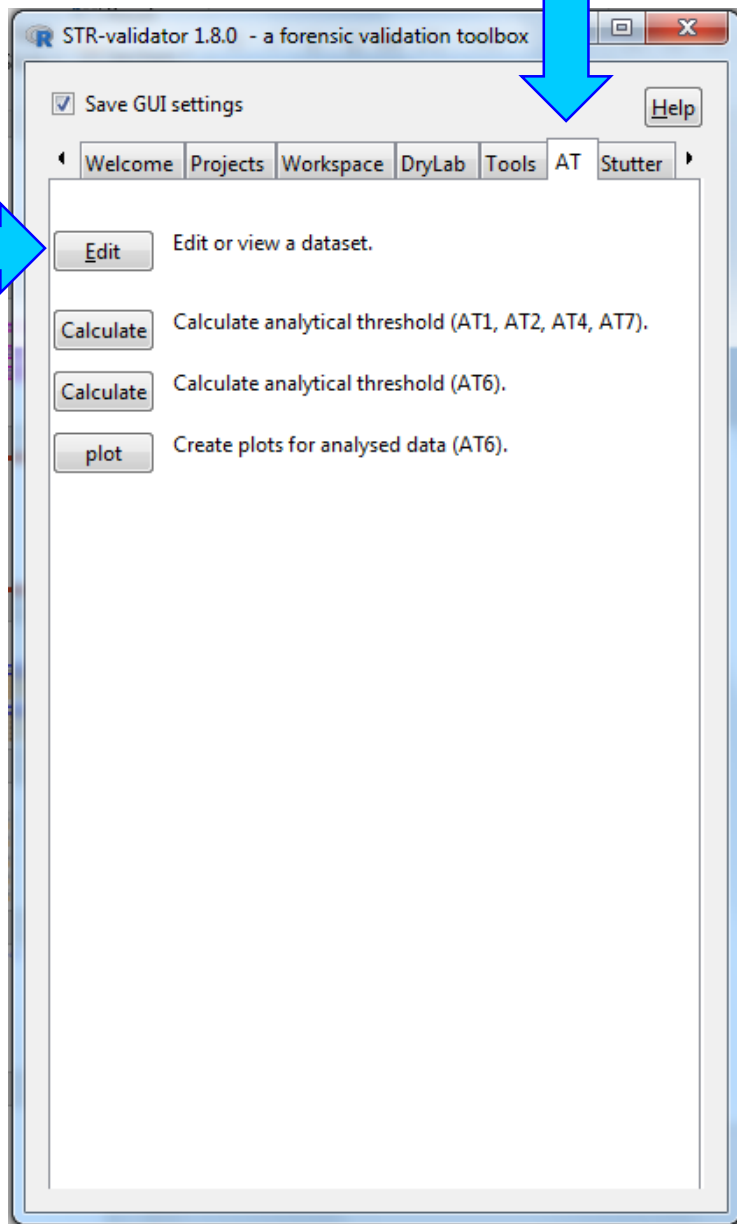
DNA Dilution Series Data	AT1	AT2	AT4	AT7
Blue	64	106	69	53
Green	68	137	73	50
Yellow	53	91	58	34
Red	57	107	61	40
Purple	55	107	60	36

Percentile Rank of noise used to calculate AT_{M2}

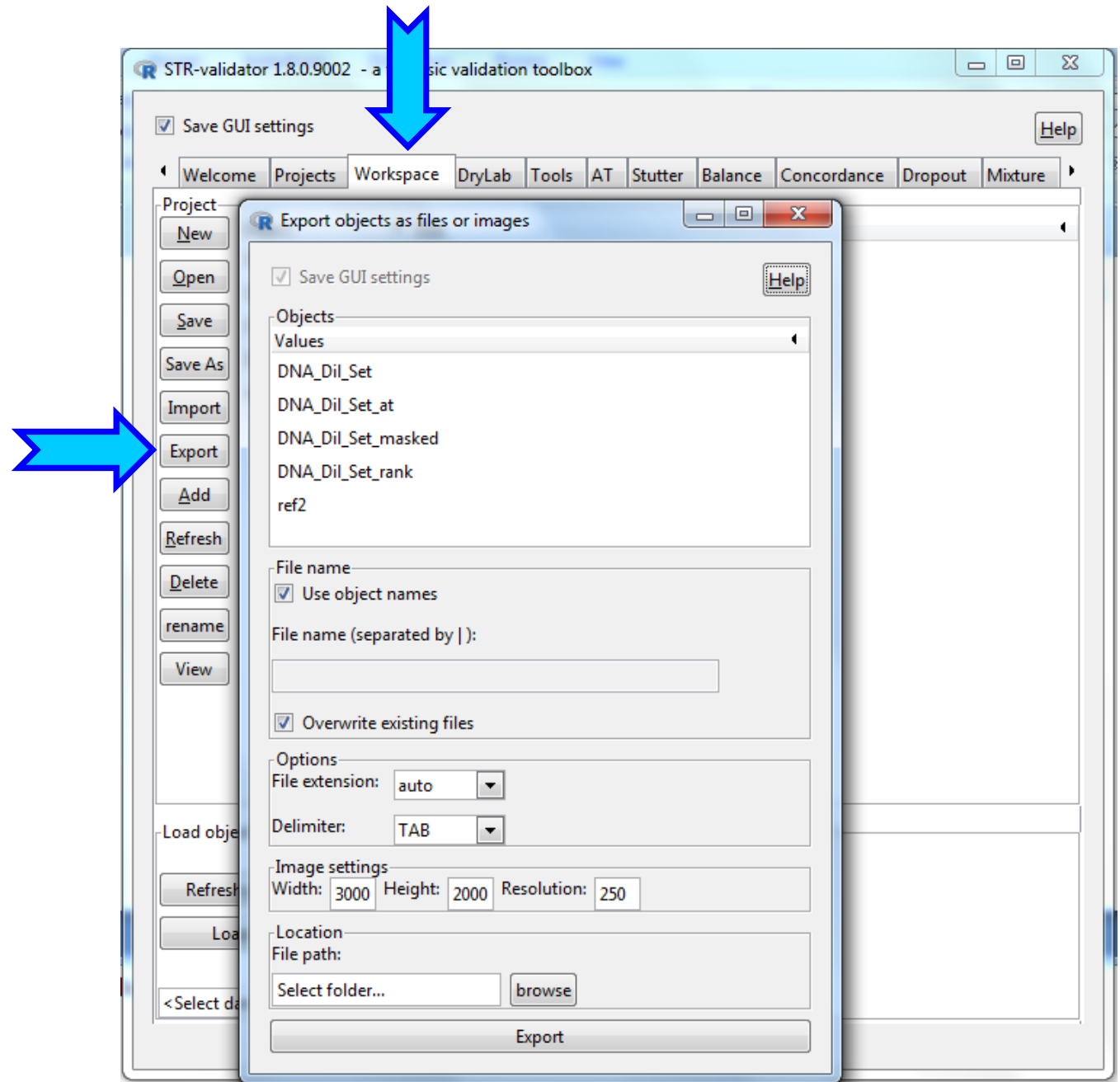


Data frame			
Row.names	Height	Rank	Observations
1	1	0.00411535	1097
2	2	0.03535153	7238
3	3	0.1232581	16215
4	4	0.262108	20832
5	5	0.4198256	21248
6	6	0.5672361	18082
7	7	0.6871135	13901
8	8	0.7755748	9701
9	9	0.8348463	6114
10	10	0.8723717	3898
11	11	0.8962092	2462
12	12	0.9113288	1572
13	13	0.9213736	1108
14	14	0.9285248	799
15	15	0.9338471	621
16	16	0.9381723	533
17	17	0.9420853	512
18	18	0.9456384	435
19	19	0.9488092	411
20	20	0.9516353	343
21	21	0.9541165	320
22	22	0.9564178	294
23	23	0.9585242	267
24	24	0.9604432	245
25	25	0.9622572	240
26	26	0.9639813	219
27	27	0.965458	176
28	28	0.9667549	169
29	29	0.9679393	147
30	30	0.9690562	152

Masked Raw Data

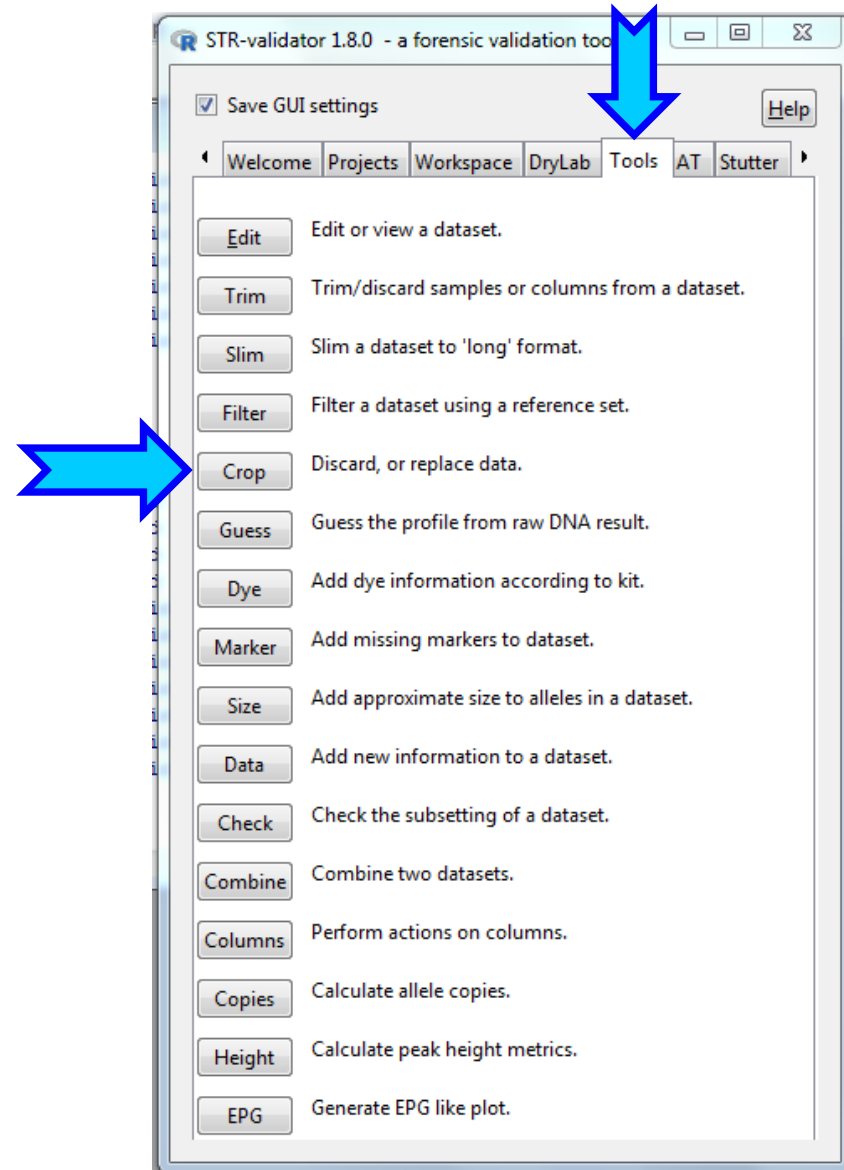


How to Export Masked Data for Manual Check and Calculations ?

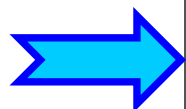
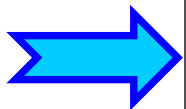
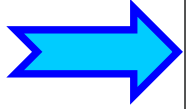
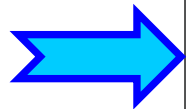


Evaluate the Distribution of Noise

Extract peaks included in the calculation from the masked dataset



Discard Masked Data



Crop or replace values in data frames

Save GUI settings Help

Datasets
Select dataset: DNA_Dil_Set_masked 0 samples, 18 columns, 197878 rows

Column
Select target column: Masked Info: Min: FALSE Max: FALSE Ignore NA for info

Options
 Remove NA

Action:
 Discard values not equal to FALSE with
 Replace values

Target column contain data of type:
 Numeric
 Character

Save as
Name for result: DNA_Dil_Set_masked_FALSE

Hit Apply and Don't Save at this step

Crop Data from ILS

R Crop or replace values in data frames

Save GUI settings Help

Datasets
Select dataset: DNA_Dil_Set_masked 0 samples, 18 columns, 197878 rows

Column
Select target column: Dye Info: Min: B Max: Y Ignore NA for info

Options
 Remove NA

Action:

Discard values equal to 0 with

Replace values

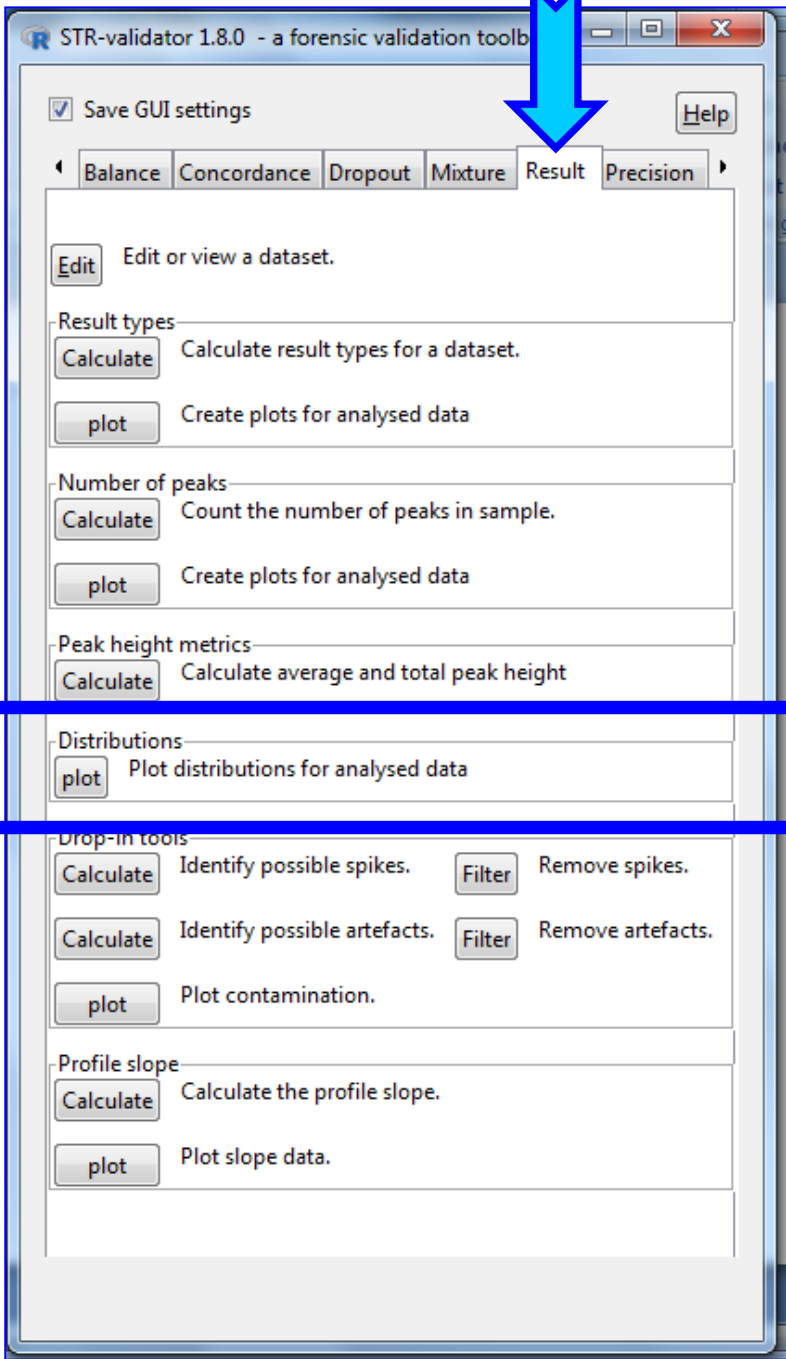
Target column contain data of type:

Numeric
 Character

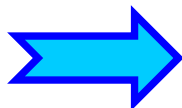
Apply

Save as
Name for result: DNA_Dil_Set_masked_FALSE_0 Save 0 samples, 18 columns, 133403 rows

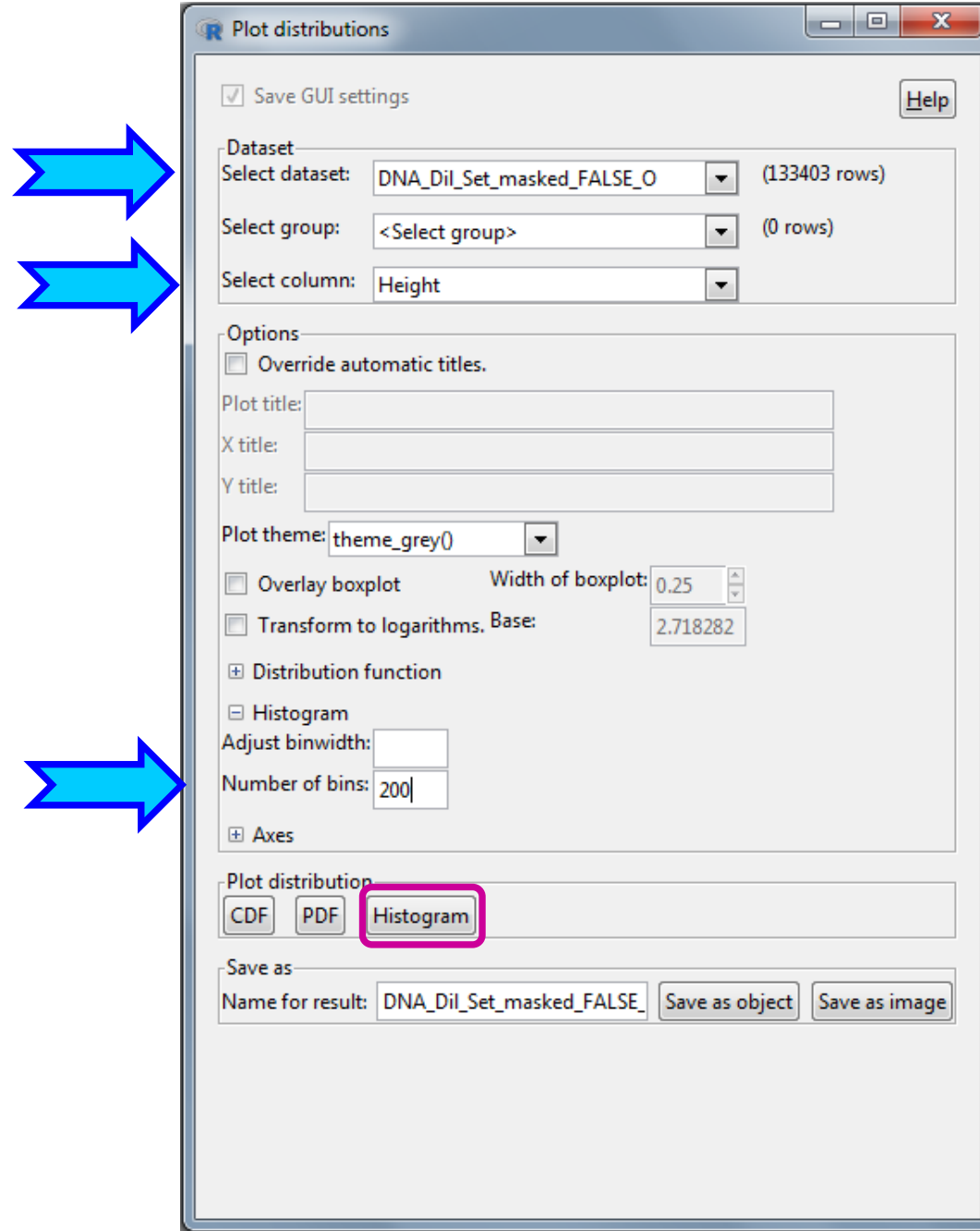
The Result Tab



Check assumptions



Plot Gaussian (Normal) Distribution of Noise



The image shows the 'Plot distributions' dialog box in R. Three blue arrows point to the 'Dataset' section, the 'Options' section, and the 'Plot distribution' section. The 'Dataset' section has 'Select dataset' set to 'DNA_Dil_Set_masked_FALSE_O' (133403 rows), 'Select group' set to '<Select group>' (0 rows), and 'Select column' set to 'Height'. The 'Options' section has 'Override automatic titles' unchecked, 'Plot title', 'X title', and 'Y title' empty, 'Plot theme' set to 'theme_grey()', 'Overlay boxplot' unchecked, 'Width of boxplot' set to 0.25, 'Transform to logarithms' unchecked, 'Base' set to 2.718282, 'Distribution function' checked, 'Histogram' checked, 'Adjust binwidth' empty, and 'Number of bins' set to 200. The 'Plot distribution' section has 'CDF', 'PDF', and 'Histogram' buttons, with 'Histogram' highlighted by a pink box. The 'Save as' section has 'Name for result' set to 'DNA_Dil_Set_masked_FALSE_', 'Save as object', and 'Save as image' buttons.

Plot distributions

Save GUI settings Help

Dataset

Select dataset: DNA_Dil_Set_masked_FALSE_O (133403 rows)

Select group: <Select group> (0 rows)

Select column: Height

Options

Override automatic titles.

Plot title:

X title:

Y title:

Plot theme: theme_grey()

Overlay boxplot Width of boxplot: 0.25

Transform to logarithms. Base: 2.718282

Distribution function

Histogram

Adjust binwidth:

Number of bins: 200

Axes

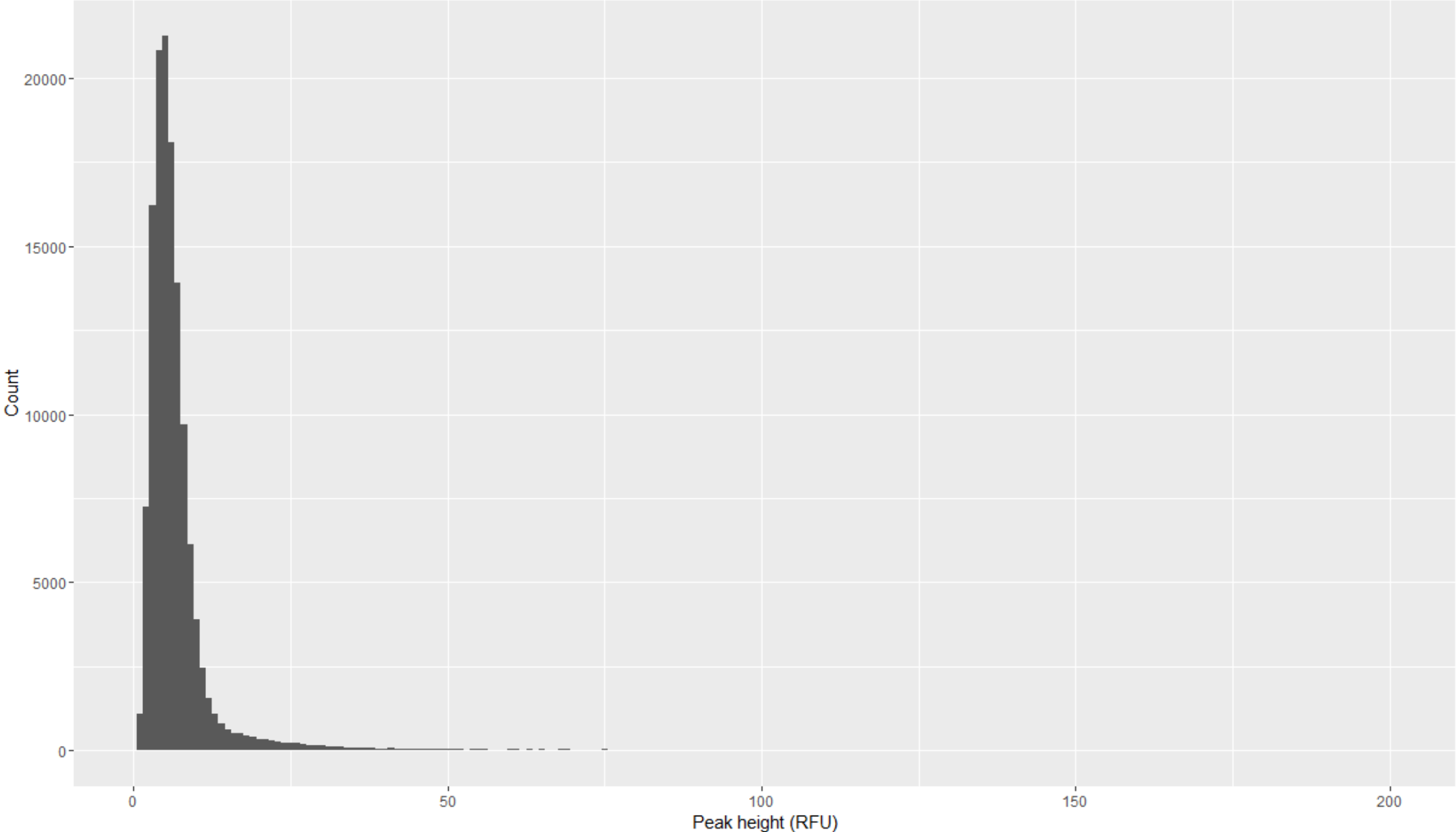
Plot distribution

Save as

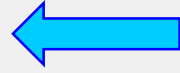
Name for result: DNA_Dil_Set_masked_FALSE_

Gaussian (Normal) Distribution of Noise Signal

Histogram (133403 observations)



Plot Natural Logarithm of Noise



Plot distributions

Save GUI settings Help

Dataset

Select dataset: DNA_Dil_Set_masked_FALSE_O (133403 rows)

Select group: <Select group> (0 rows)

Select column: Height

Options

Override automatic titles.

Plot title:

X title:

Y title:

Plot theme: theme_grey()

Overlay boxplot Width of boxplot: 0.25

Transform to logarithms. Base: 2.718282

Distribution function

Histogram

Adjust binwidth:

Number of bins: 200

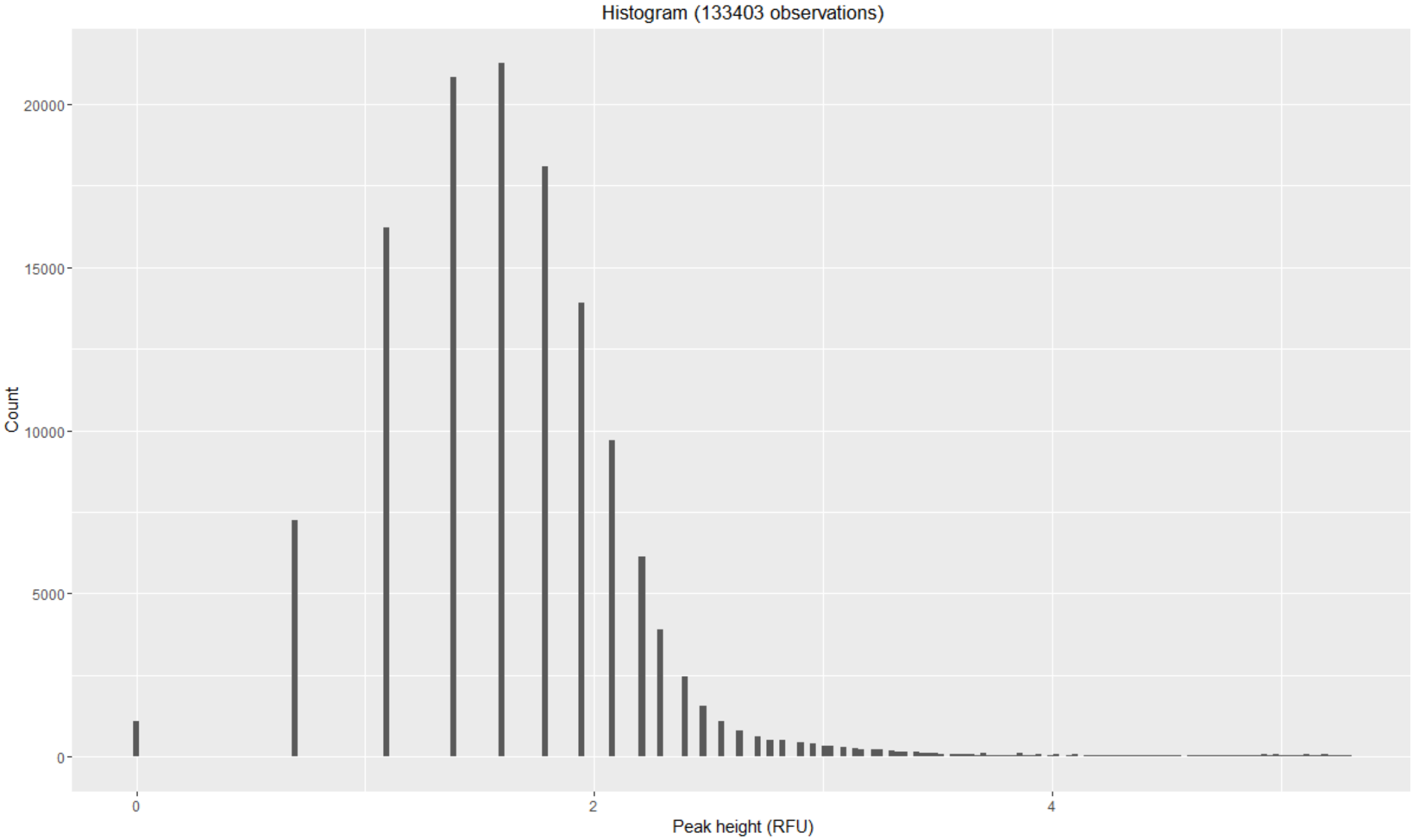
Axes

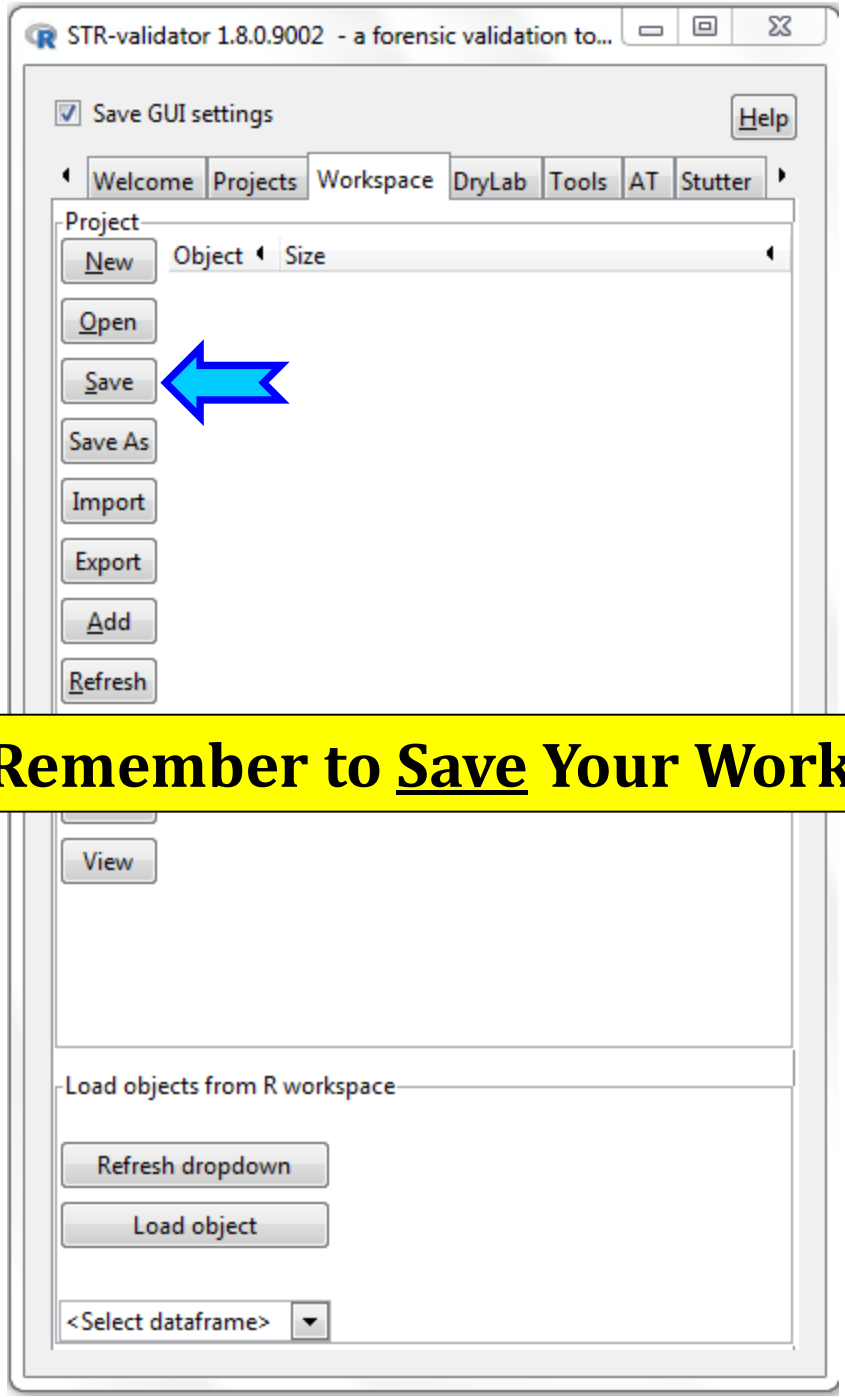
Plot distribution

Save as

Name for result: DNA_Dil_Set_masked_FALSE_O_ggplot Save as object Save as image

Natural Logarithm of Noise Signals





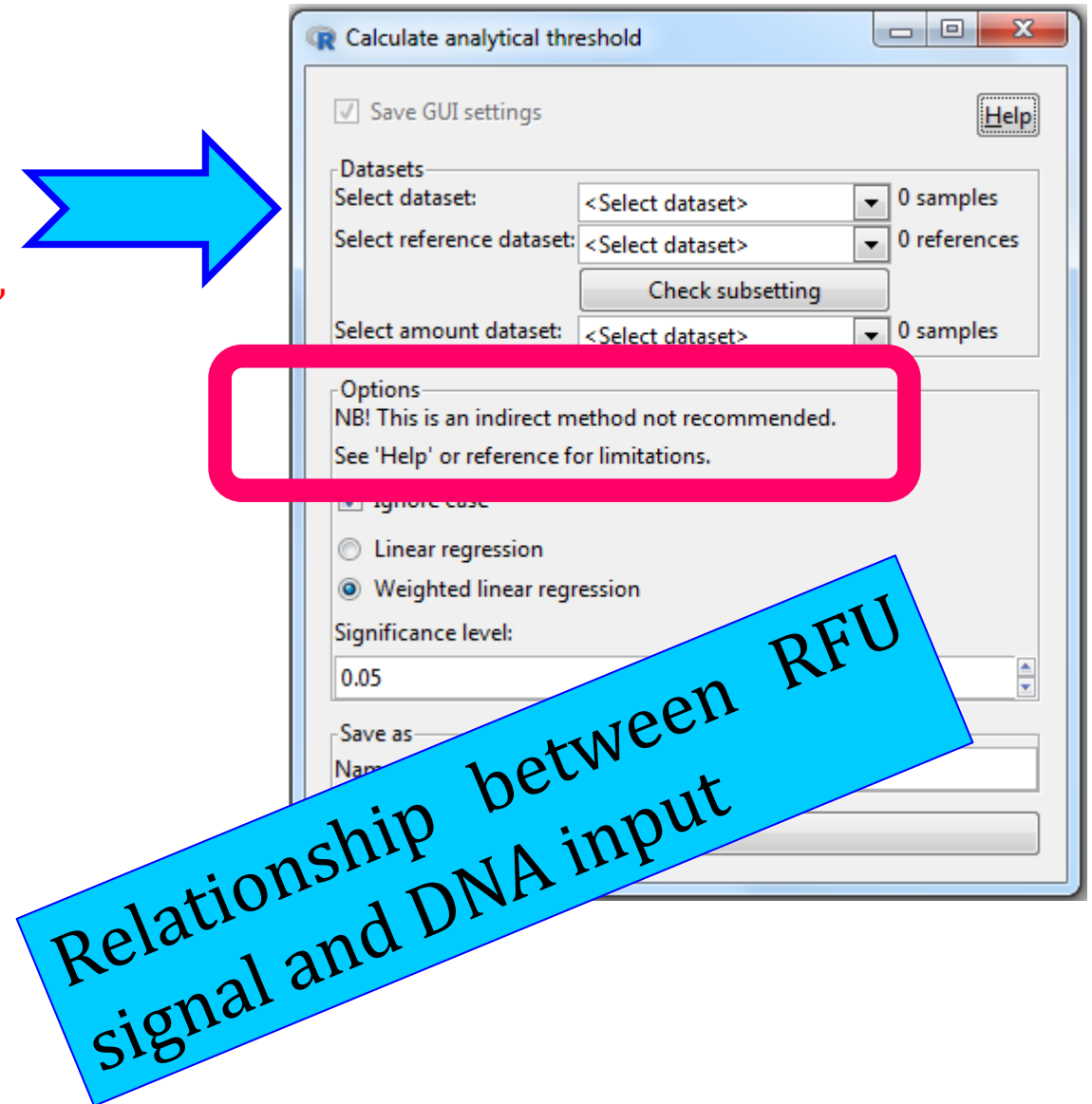
Remember to Save Your Workspace

The Analytical Threshold

Method 6

1. Analyze samples in *GeneMapper* at your AT
2. Export **-GenotypeTable.txt** from GeneMapper with at least the following information: **“Sample.Name”, “Marker”, “Allele” and “Height”**.

Import, from one or several batches of sensitivity studies





- ❖ To Calculate AT6, **a kit must be specified.**
- ❖ However kit is **NOT** an option in the calculateAT6_gui function.
- ✓ **Download the updated STR-validator development version “1.8.0.9002”.**
 - (1) Install devtools by typing or copy/paste the following command in R-console :
`install.packages("devtools", dependencies=TRUE)`
 - (2) Download the updated development version by typing this into the command window
`devtools::install_github("oskarhansson/strvalidator")`

Reference:

<https://github.com/OskarHansson/strvalidator/commit/55aa1e7cb7b257435350cda77b52e1b062c21596>

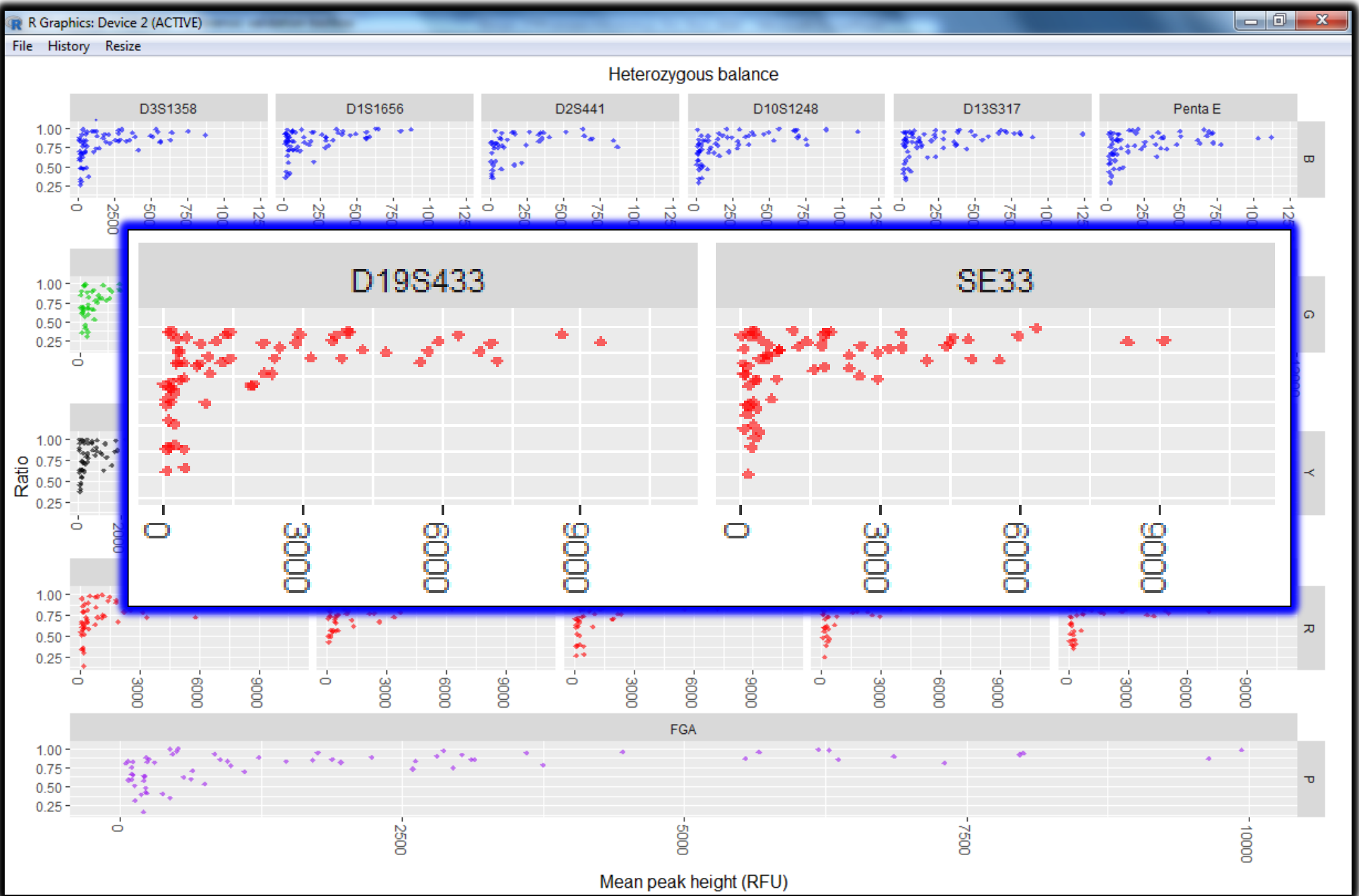
Peak Height Ratio (PHR)

- ❖ Establish potential expectations for allele pairing to define genotypes for mixed samples. It is an indication of which alleles may be heterozygous pairs.
- ❖ To express the PHR as a percentage: **divide the peak height of an allele with a lower relative fluorescence unit (RFU) value by the peak height of an allele with a higher RFU value, and then multiplying this value by 100**

Experimental Design for Peak Height Ratio Analysis

- Sensitivity study data
- Three mostly heterozygous samples selected
- DNA input amounts ranged from:
 - 2.0 ng, 1.0 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, and 0.031 ng
- Amplified in triplicate with positive and negative controls
- Analyzed at your AT
- Export **_GenotypeTable.txt** from GeneMapper with at least the following information: **"Sample.Name"**, **"Marker"**, **"Height"**, and **"Allele"**.

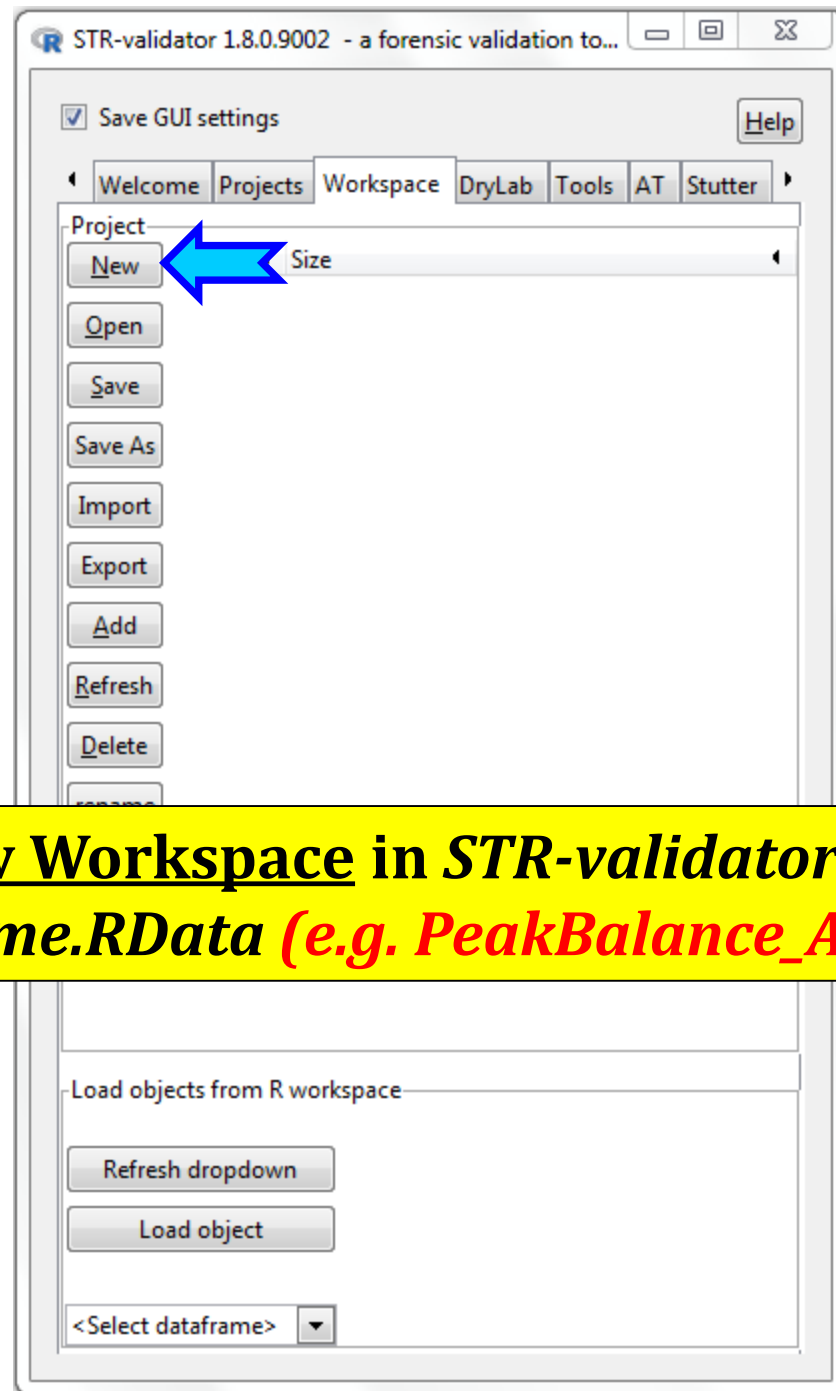
Plot Peak Height Ratio



Summarize Balance at Each Locus

The screenshot shows the 'Edit or view data frame' window in R Studio. At the top, there is a 'Save GUI settings' checkbox and a 'Help' button. Below that, the 'Datasets' section shows 'Select dataset: PeakBalance_hb_table_locus' with a dropdown arrow and the text '<NA> samples, 7 columns, 24 rows'. The 'Options' section has two checkboxes: 'Show attributes (separate window)' and 'Limit number of rows to: 100'. A menu bar with 'Copy | Export | Save' is visible, with buttons for 'Copy', 'Export', and 'Save as'. The filename 'PeakBalance_hb_table_locus_edit' is shown in the text field below the menu bar. The main area displays a data frame table with 24 rows and 8 columns. A context menu is open over the table, listing options: 'Apply function to column', 'Sort by column (decreasing)', 'Sort by column (increasing)', and 'Rename column'. The 'Sort by column (increasing)' option is highlighted.

Row.names	Marker	Hb.n	Hb.Min	Hb.Mean	Hb.Sd	Hb.Max	Hb.Perc.5
20	CSF1PO	61	0.2611276	0.808836	0.1756403	0.9988395	0.3469388
15	D8S1179	64	0.1442006	0.7662984	0.1951862	0.9955556	0.3501812
9	D18S51	66	0.2478873	0.7769547	0.1853633	0.9949166	0.3654003
17	D19S433	64	0.2658228	0.7844988	0.1829694	0.9848334	0.3829258
2	D3S1358	64	0.256917	0.781428	0.1855591	0.9914355	0.3837712
10	D2S1338	61	0.2844311	0.7625711	0.1775064	0.9956921	0.3941606
19	FGA	66	0.1680672	0.7660545	0.1932834	0.9961315	0.3984242
18	D22S1045	61	0.3523035	0.7920664	0.1828193	0.9990485	0.4045369
4	D2S441	44	0.3523132	0.774105	0.1794914	0.9976526	0.4136898
11	Penta D	44	0.376	0.7912205	0.1697334	0.9760568	0.4139601
8	D16S539	64	0.3006431	0.8026873	0.1745262	0.9976437	0.4172329
3	D1S1656	63	0.3618421	0.8333734	0.1533515	0.9996813	0.4437333
6	D13S317	64	0.3341014	0.8135805	0.1657278	0.9856532	0.4466799
5	D10S1248	65	0.2918149	0.7782516	0.1735318	0.999473	0.4679155
7	Penta E	63	0.2926829	0.7792479	0.1657908	0.993295	0.4689736
24	SE33	63	0.2548263	0.8062033	0.1678947	0.9978029	0.4708047
22	D21S11	63	0.2831858	0.8265385	0.1519948	0.9971618	0.4762263
12	TH01	64	0.3769231	0.8242476	0.1598705	1	0.492605
14	TPOX	58	0.3203125	0.7874313	0.1651643	1	0.4938764
21	vWA	64	0.2727273	0.8188568	0.1575281	0.9924906	0.4963417
23	D7S820	60	0.4027149	0.8141149	0.1537968	0.9904153	0.5088509
13	D5S818	60	0.3232323	0.8085893	0.150846	0.9952381	0.5254869
1	AMEL	65	0.2146018	0.8290668	0.1593286	0.9992785	0.5365907
16	D12S391	64	0.4318182	0.8088914	0.1426007	1	0.5524576



Open a New Workspace in *STR-validator* GUI and save as *Name.RData (e.g. PeakBalance_Analysis)*

Import Data

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset
 Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Peak_Balance\Set7.tx browse

Select a directory... browse

Options

Save file name
 Save file time stamp

Delimiter:
TAB

NA strings (separated by comma):
NA,,

Auto trim samples
 Auto slim repeated columns

+ Multiple files options
+ Trim options
+ Slim options

Save options
Name:
PeakBalance

Import

Import Reference

Import from files

Save GUI settings Help

Import multiple files from a directory into one dataset
 Import a single file

D:\Fusion6C_InternalVali_STRvalidator\Peak_Balance\ref.txt browse

Select a directory... browse

Options

Save file name
 Save file time stamp

Delimiter:
TAB

NA strings (separated by comma):
NA,,

Auto trim samples
 Auto slim repeated columns

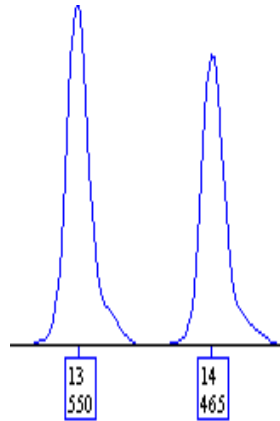
+ Multiple files options
+ Trim options
+ Slim options

Save options
Name:
ref

Import

Intra-locus Peak Balance

D10S1248

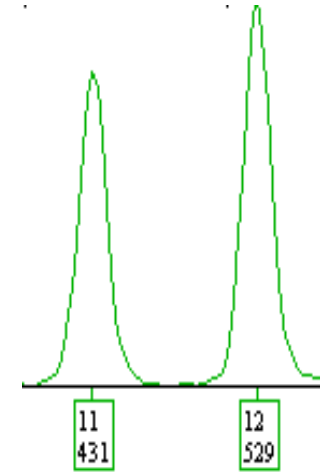


$$Hb = \frac{\text{Peak Height HMW}}{\text{Peak Height LMW}} = \frac{465}{550} = 0.85$$

$$Hb = \frac{\text{Peak Height LMW}}{\text{Peak Height HMW}} = \frac{550}{465} = 1.18$$

$$Hb = \frac{\text{Peak Height smaller}}{\text{Peak Height larger}} = \frac{465}{550} = 0.85$$

CSF1PO



$$= \frac{529}{431} = 1.2$$

$$= \frac{431}{529} = 0.81$$

$$= \frac{431}{529} = 0.81$$

Calculate Balance

STR-validator 1.8.0 - a forensic validation toolbox

Save GUI settings Help

Projects | Workspace | DryLab | Tools | AT | Stutter | **Balance** | Concordance

Edit Edit or view a dataset.

Intralocus and interlocus balance

- Calculate** Calculate intra-locus balance.
- Calculate** Calculate inter-locus balance.
- plot** Create plots for analysed data
- Summarize** Calculate summary statistics for balance data.

Capillary balance

- Calculate** Calculate capillary balance for a dataset.
- plot** Create plots for analysed data
- Summarize** Create summary table for analysed data

Marker peak height ratio

- Calculate** Calculate locus ratio for a dataset.
- plot** Create plots for analysed data

Calculate heterozygote balance

Save GUI settings Help

Datasets

Select dataset: PeakBalance 66 samples.

Select reference dataset: ref 4 samples.

Check subsetting

Options

Pre-processing:

- Remove sex markers
- Remove quality sensors

Define Hb as:

Smaller peak / larger peak

Sample name matching:

- Ignore case
- Add word boundaries
- Exact matching

Post-processing:

- Calculate average peak height

Save as

Name for result: PeakBalance_hb Kit attribute: Fusion 6C

Calculate

Results of Hb Analysis

Edit or view data frame

Save GUI settings Help

Datasets
Select dataset: Set7_hb 66 samples, 13 columns, 1377 rows

Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
 Set7_hb_edit

Data frame

Row.names	Sample.Name	Marker	Dye	Delta	Small	Large	MPH	Hb	TPH	H	Peaks	Expected	Proportion
1	29_0.03ng_A	D3S1358	B	5	111	181	146	0.6132597	4334	117.1351	37	46	0.8043478
2	29_0.03ng_A	D1S1656	B	1	86	107	96.5	0.8037383	4334	117.1351	37	46	0.8043478
3	29_0.03ng_A	D2S441	B	1	99	281	190	0.3523132	4334	117.1351	37	46	0.8043478
4	29_0.03ng_A	D10S1248	B	2	81	138	109.5	0.5869565	4334	117.1351	37	46	0.8043478
5	29_0.03ng_A	D13S317	B	1	75	161	118	0.4658385	4334	117.1351	37	46	0.8043478
6	29_0.03ng_A	Penta E	B	6	81	87	84	0.9310345	4334	117.1351	37	46	0.8043478
7	29_0.03ng_A	D18S51	G	2	135	173	154	0.7803468	4334	117.1351	37	46	0.8043478
8	29_0.03ng_A	D2S1338	G	6	71	190	130.5	0.3736842	4334	117.1351	37	46	0.8043478
9	29_0.03ng_A	Penta D	G	1	93	126	109.5	0.7380952	4334	117.1351	37	46	0.8043478
10	29_0.03ng_A	TH01	Y	2.3	62	62	62	1	4334	117.1351	37	46	0.8043478
11	29_0.03ng_A	D5S818	Y	1	73	81	77	0.9012346	4334	117.1351	37	46	0.8043478
12	29_0.03ng_A	D12S391	R	3	155	155	155	1	4334	117.1351	37	46	0.8043478
13	29_0.03ng_A	D22S1045	R	4	69	94	81.5	0.7340426	4334	117.1351	37	46	0.8043478
14	29_0.03ng_A	FGA	P	3	92	140	116	0.6571429	4334	117.1351	37	46	0.8043478
15	29_0.03ng_B	D3S1358	B	5	108	223	165.5	0.4843049	3485	112.4194	31	46	0.673913
16	29_0.03ng_B	D2S441	B	1	81	213	147	0.3802817	3485	112.4194	31	46	0.673913
17	29_0.03ng_B	D13S317	B	1	69	168	118.5	0.4107143	3485	112.4194	31	46	0.673913
18	29_0.03ng_B	D16S539	G	2	146	233	189.5	0.6266094	3485	112.4194	31	46	0.673913
19	29_0.03ng_B	D18S51	G	2	113	136	124.5	0.8308824	3485	112.4194	31	46	0.673913
20	29_0.03ng_B	CSF1PO	G	1	83	114	98.5	0.7280702	3485	112.4194	31	46	0.673913
21	29_0.03ng_B	Penta D	G	1	79	128	103.5	0.6171875	3485	112.4194	31	46	0.673913
22	29_0.03ng_B	vWA	Y	4	84	97	90.5	0.8659794	3485	112.4194	31	46	0.673913
23	29_0.03ng_B	D21S11	Y	2	118	180	149	0.6555556	3485	112.4194	31	46	0.673913
24	29_0.03ng_B	D7S820	Y	4	73	127	100	0.5748031	3485	112.4194	31	46	0.673913
25	29_0.03ng_B	D19S433	R	1	58	82	70	0.7073171	3485	112.4194	31	46	0.673913
26	29_0.03ng_B	FGA	P	3	64	204	134	0.3137255	3485	112.4194	31	46	0.673913
27	29_0.03ng_C	D3S1358	B	5	113	128	120.5	0.8828125	2323	105.5909	22	46	0.4782609
28	29_0.03ng_C	D1S1656	B	1	126	135	130.5	0.9333333	2323	105.5909	22	46	0.4782609
29	29_0.03ng_C	D10S1248	B	2	116	173	144.5	0.6705303	2323	105.5909	22	46	0.4782609

Plot Balance

STR-validator 1.8.0 - a forensic validation toolbox

Save GUI settings Help

Projects Workspace DryLab Tools AT Stutter **Balance** Concordance

Edit Edit or view a dataset.

Intralocus and interlocus balance

Calculate Calculate intra-locus balance.

Calculate Calculate inter-locus balance.

plot Create plots for analysed data

Summarize Calculate summary statistics for balance data.

Capillary balance

Calculate Calculate capillary balance for a dataset.

plot Create plots for analysed data

Summarize Create summary table for analysed data

Marker peak height ratio

Calculate Calculate locus ratio for a dataset.

plot Create plots for analysed data

Plot balance Help

Save GUI settings

Dataset and kit

Select dataset: PeakBalance_hb (66 samples) and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Plot theme: theme_grey()

Drop sex markers

Plot Log(balance)

Do not facet or wrap

Wrap by Dye

Facet by Marker and wrap by Dye

Data points

Axes

X labels

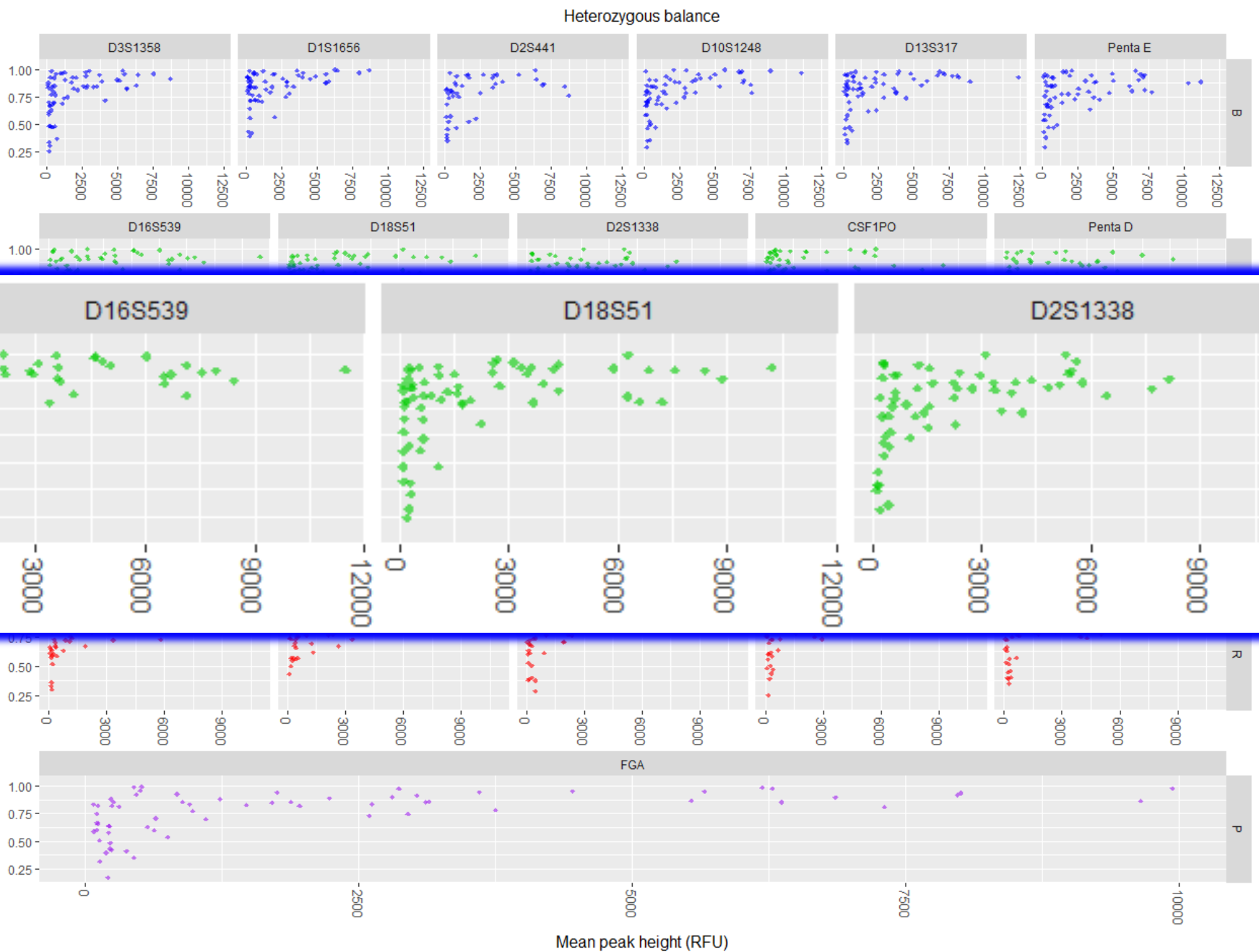
Plot Balance data

Hb vs. Height Hb vs. Delta Hb vs. 'H' Hb vs. Marker Lb vs. Height Lb vs. 'H' Lb vs. Marker

Save as

Name for result: PeakBalance_hb_ggplot Save as object Save as image

Peak Height Ratio plotted by the mean peak height of the locus



Plot Balance

Plot balance

Save GUI settings Help

Dataset and kit
Select dataset: PeakBalance_hb (66 samples) and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Plot theme: theme_grey0

Drop sex markers

Plot Log(balance)

Do not facet or wrap

Wrap by Dye

Facet by Marker and wrap by Dye

Data points

Axes

X labels

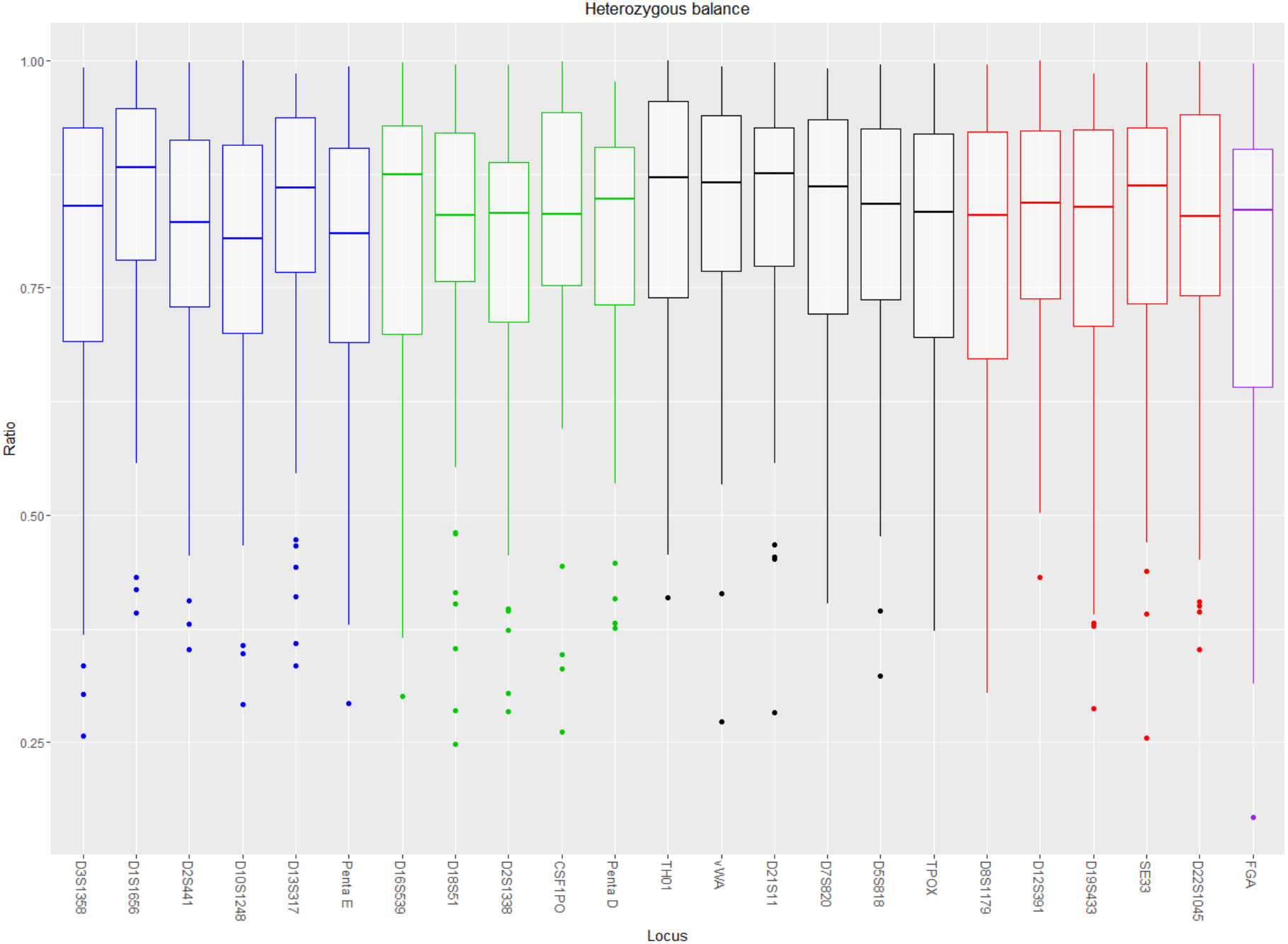
Plot Balance data

Hb vs. Height Hb vs. Delta Hb vs. 'H' **Hb vs. Marker** Lb vs. Height Lb vs. 'H' Lb vs. Marker

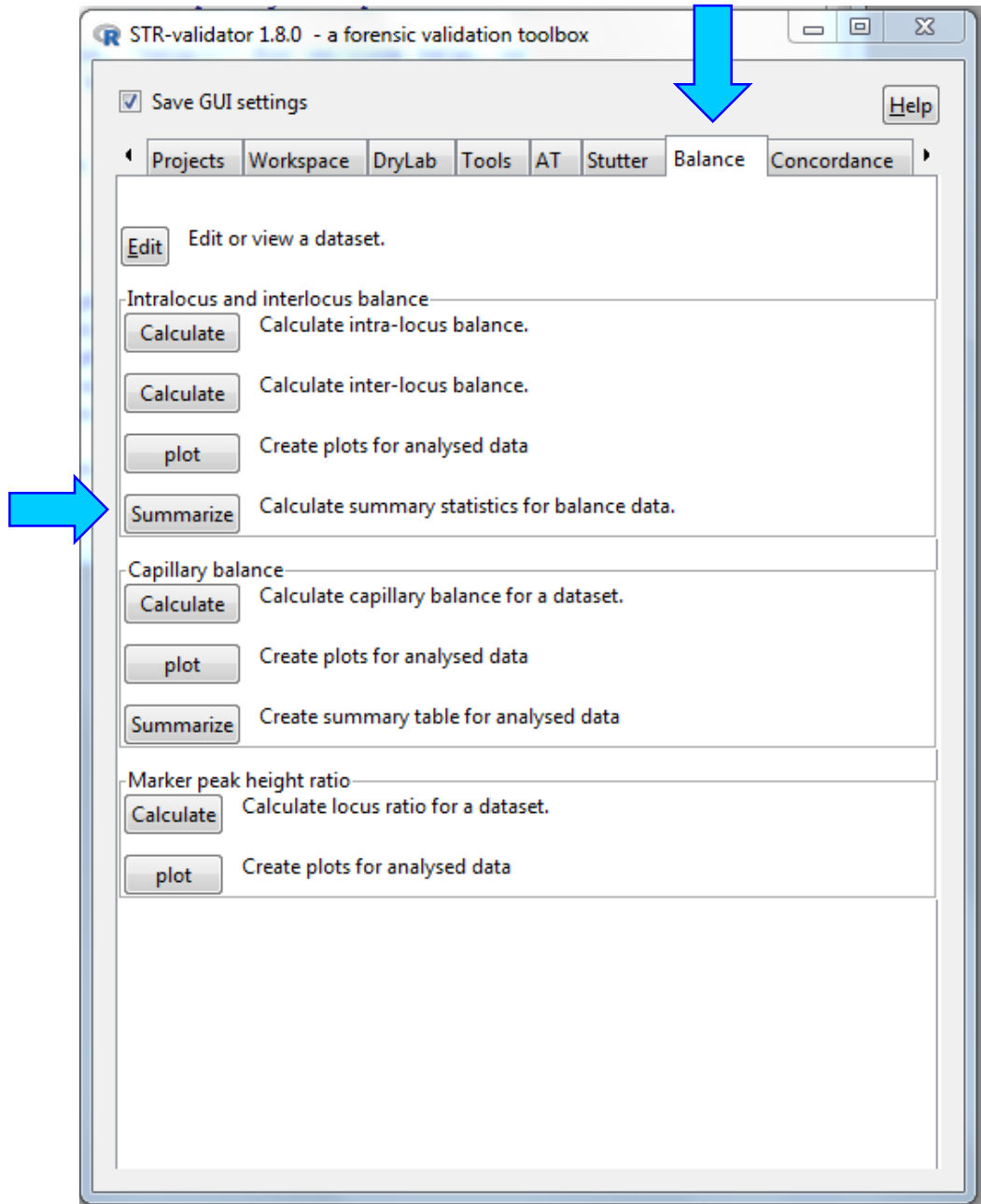
Save as

Name for result: PeakBalance_hb_ggplot Save as object Save as image

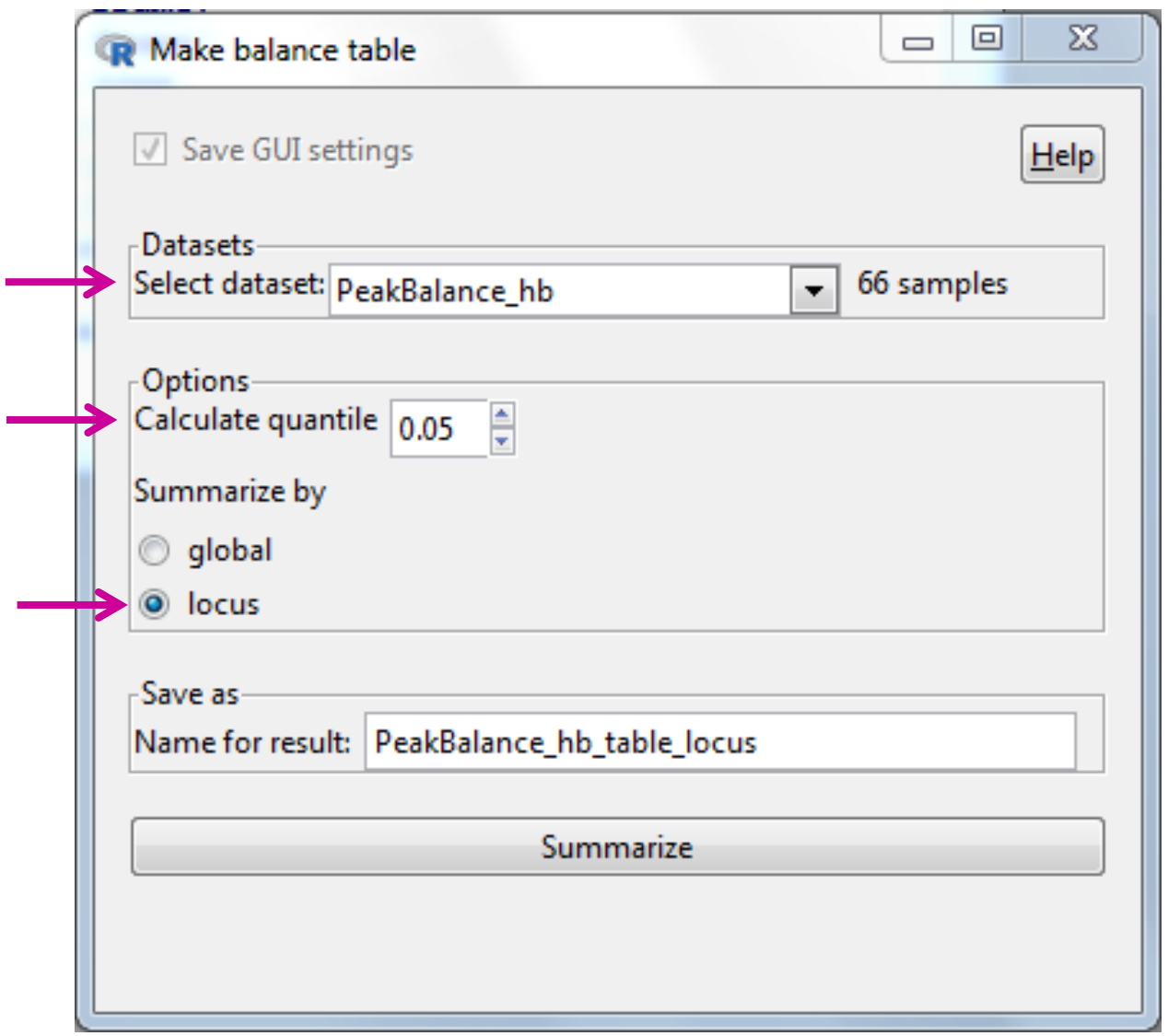
Peak Height Ratio plotted by Locus



Calculate Hb Summary Statistics



Plot Balance Dialogue



View the Results and Sort the Column of Perc.95 (Increasing)

Edit or view data frame

Save GUI settings Help

Datasets
Select dataset: Set7_hb_table_locus <NA> samples, 7 columns, 23 rows

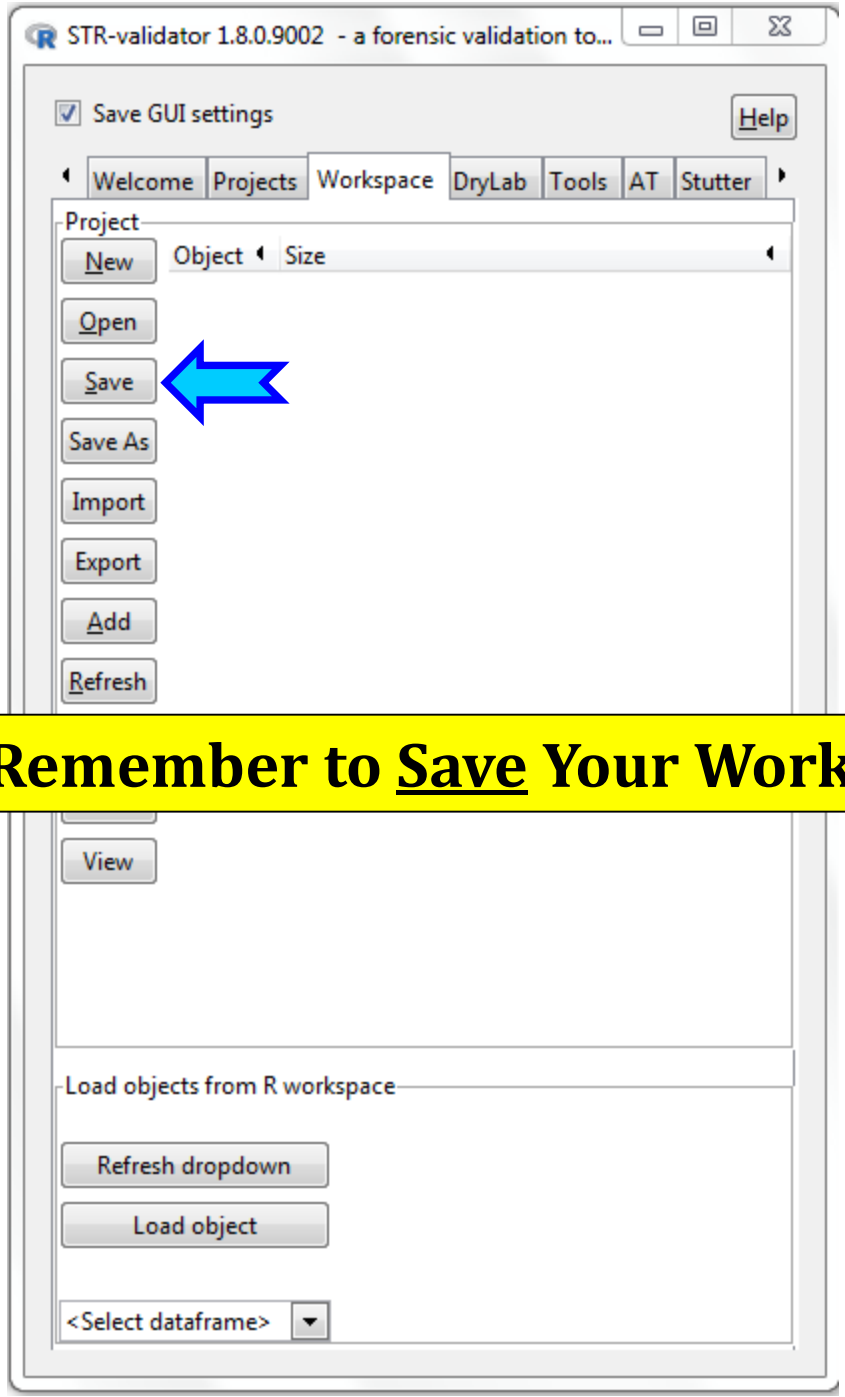
Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
 Set7_hb_table_locus_edit

Data frame

Row.names	Marker	Hb.n	Hb.Min	Hb.Mean	Hb.Sd	Hb.Max	Hb.Perc.5
1	D3S1358	64	0.256917	0.781428	0.1855591	0.9914355	0.3837712
8	D2S1338	58	0.2844311	0.7698788	0.1761992	0.9956921	0.3910891
20	D19S433	63	0.287234	0.7927317	0.172075	0.9848334	0.3916481
14	FGA	65	0.1680672	0.7653731	0.1947077	0.9961315	0.3975715
13							
7							
3							
9	Penta D	44	0.376	0.7912205	0.1697334	0.9760568	0.4139601
16	CSF1PO	60	0.2611276	0.8177378	0.162655	0.9988395	0.4392113
5	D13S317	64	0.3341014	0.8135805	0.1657278	0.9856532	0.4466799
15	D16S539	62	0.3006431	0.8117953	0.1668894	0.9976437	0.4557172
4	D10S1248	64	0.2918149	0.7821397	0.1720263	0.999473	0.4676033
6	Penta E	62	0.2926829	0.7814346	0.1662259	0.993295	0.4687168
22	SE33	61	0.2548263	0.8036358	0.1695414	0.9978029	0.4692443
18	D21S11	63	0.2831858	0.8265385	0.1519948	0.9971618	0.4762263
19	D7S820	58	0.4027149	0.8160377	0.1550674	0.9904153	0.5079193
21	D8S1179	60	0.3035714	0.7915645	0.1690504	0.9955556	0.5118127
11	D5S818	59	0.3232323	0.8104	0.1514817	0.9952381	0.5228839
23	TPOX	55	0.3723197	0.7973925	0.1490738	0.9967345	0.5351691
10	TH01	61	0.4096386	0.8345616	0.1500842	1	0.5411255
2	D1S1656	62	0.3925926	0.8409787	0.1421213	0.9996813	0.5577501
12	D12S391	62	0.4318182	0.8155057	0.1386548	1	0.5625524
17	vWA	61	0.2727273	0.830954	0.1443696	0.9924906	0.62

The worst balance is observed for marker D3S1358



Remember to Save Your Workspace

Workshop Schedule

Time	Topic
9:00 AM-10:00 AM	<ul style="list-style-type: none">❖ Load <i>STR-validator</i> package and launch the GUI❖ Check Precision❖ Calculate Stutter Thresholds
10:00 AM – 10:10 AM	Break
10:10 AM-11:00 AM	<ul style="list-style-type: none">❖ Calculate Analytical Thresholds❖ Analyze Peak Height Ratio
11:00 AM -11:10 AM	Break
11:10 AM-12:00 PM	<ul style="list-style-type: none">❖ Calculate Stochastic Thresholds❖ Questions❖ Feedback about the workshop (survey)❖ Workshop ends

Stochastic Threshold

SWGAM Autosomal STR Interpretation Guidelines

Stochastic Threshold:

- Is the **RFU value above** which it is reasonable to assume that, at a given locus, **allelic dropout of a sister allele has not occurred**.
- Minimizes the chance of wrongly deciding a heterozygous locus as a homozygous one.

Calculating Stochastic Threshold

Experimental Design

- Sensitivity study data
- Three mostly heterozygous samples selected
- DNA input amounts ranged from:
 - 2.0 ng, 1.0 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, and 0.031 ng
- Amplified in triplicate with positive and negative controls
- Analyzed at your AT
- Export **_GenotypeTable.txt** from GeneMapper with at least the following information:
"Sample.Name", "Marker", "Height", and "Allele".

Stochastic Threshold

Forensic Science International: Genetics 6 (2012) 679–688



Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



DNA commission of the International Society of Forensic Genetics:
Recommendations on the **evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods**

P. Gill^{a,b,*}, L. Gusmão^c, H. Haned^d, W.R. Mayr^e, N. Morling^f, W. Parson^g, L. Prieto^h,
M. Prinzⁱ, H. Schneider^j, P.M. Schneider^k, B.S. Weir^l

Forensic Science International: Genetics 3 (2009) 104–111



Contents lists available at ScienceDirect

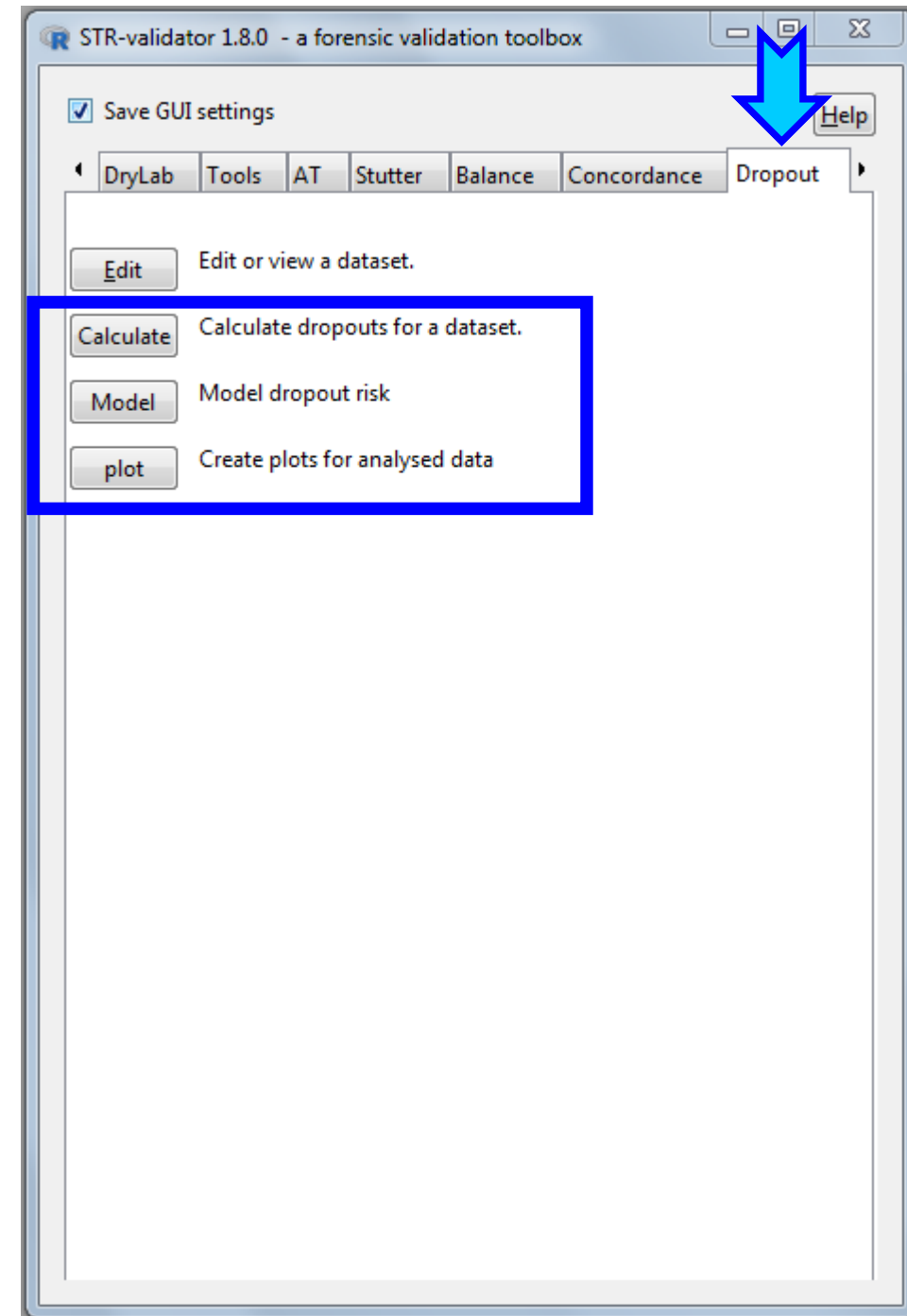
Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

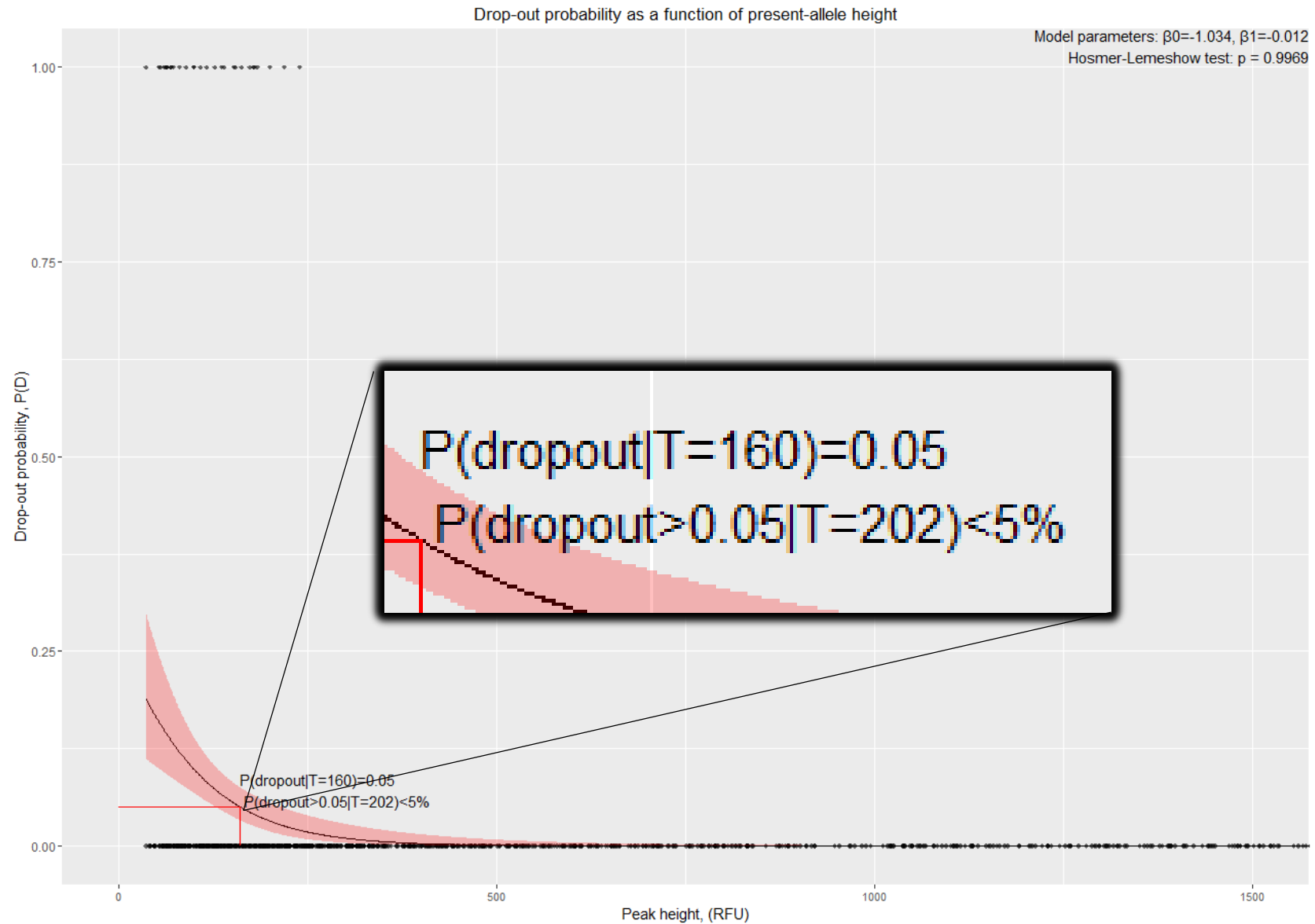


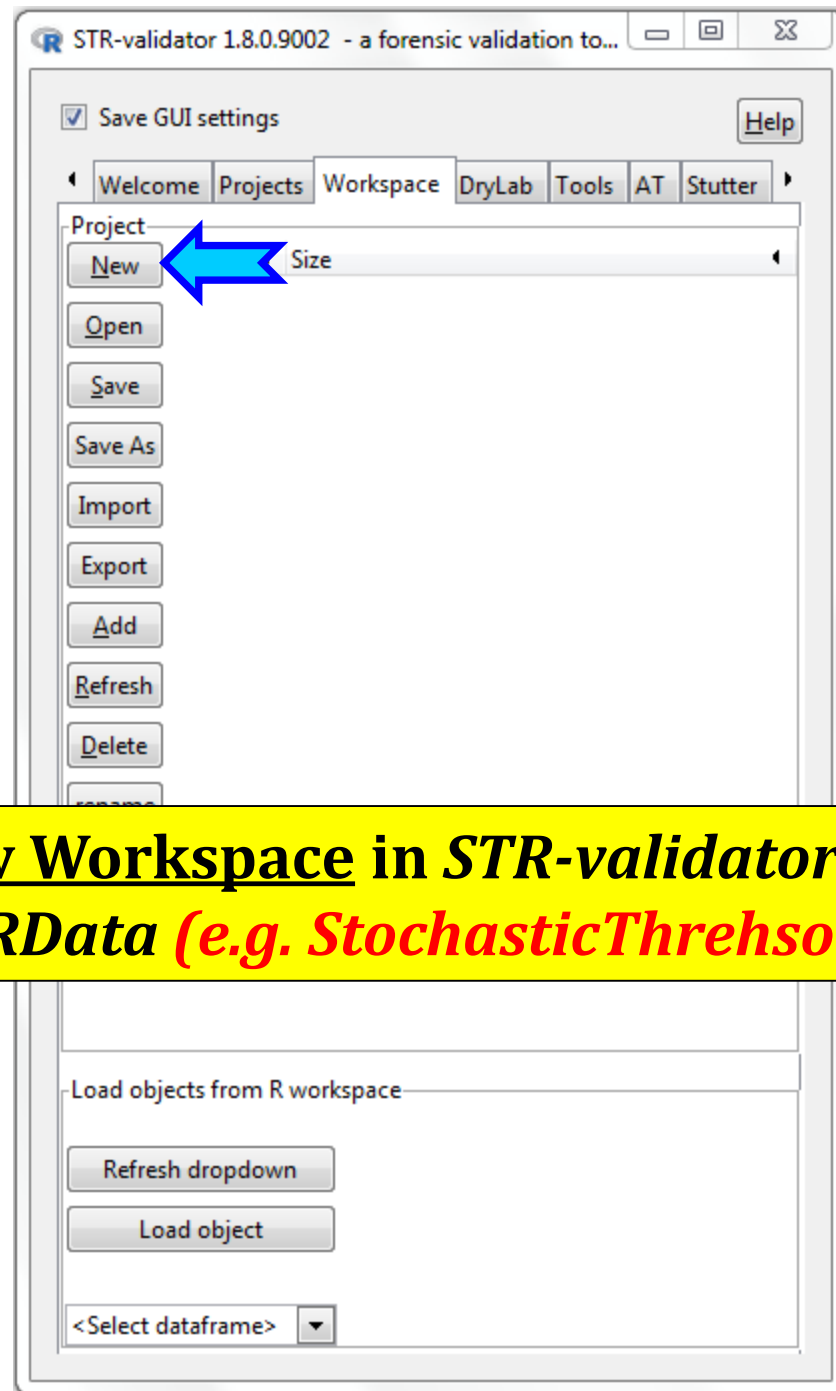
The low-template-DNA (stochastic) threshold—Its determination relative to risk analysis for national DNA databases

Peter Gill^{a,b,*}, Roberto Puch-Solis^c, James Curran^d



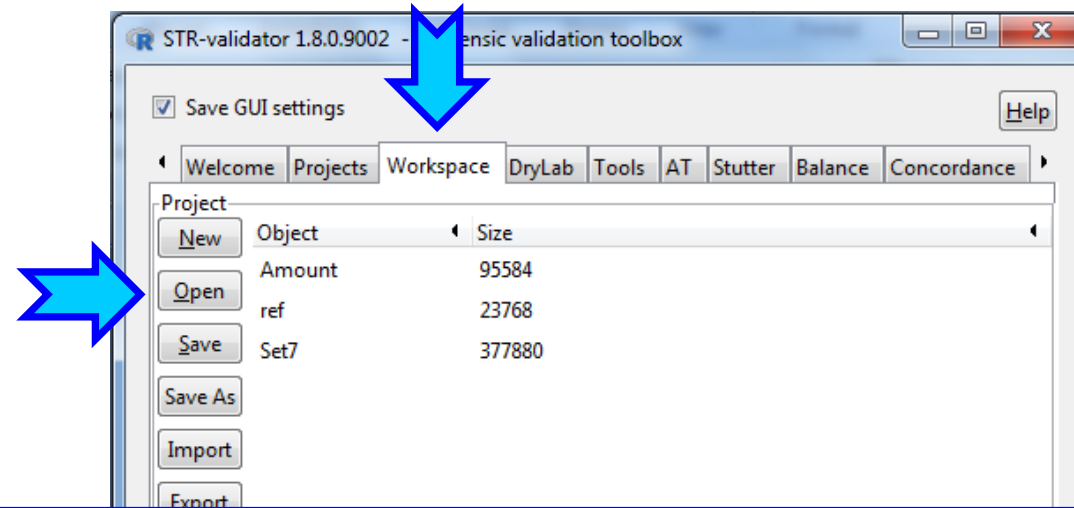
Probability of drop-out modelled by logistic regression



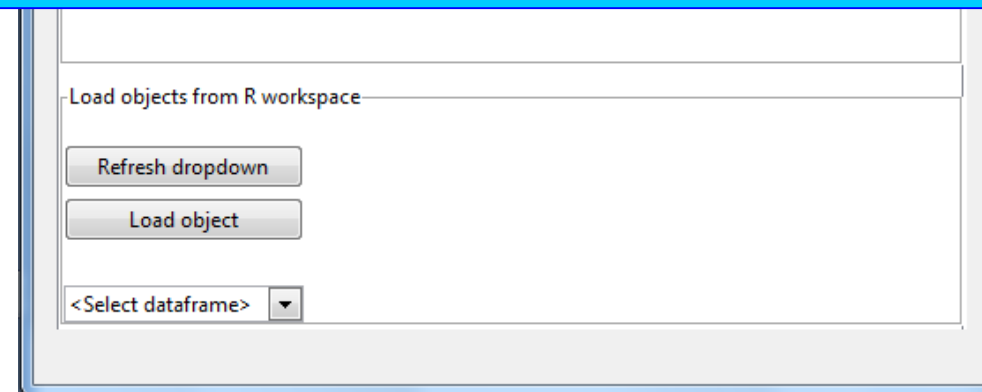


Open a New Workspace in *STR-validator* GUI and save as *Name.RData (e.g. StochasticThreshold_Analysis)*

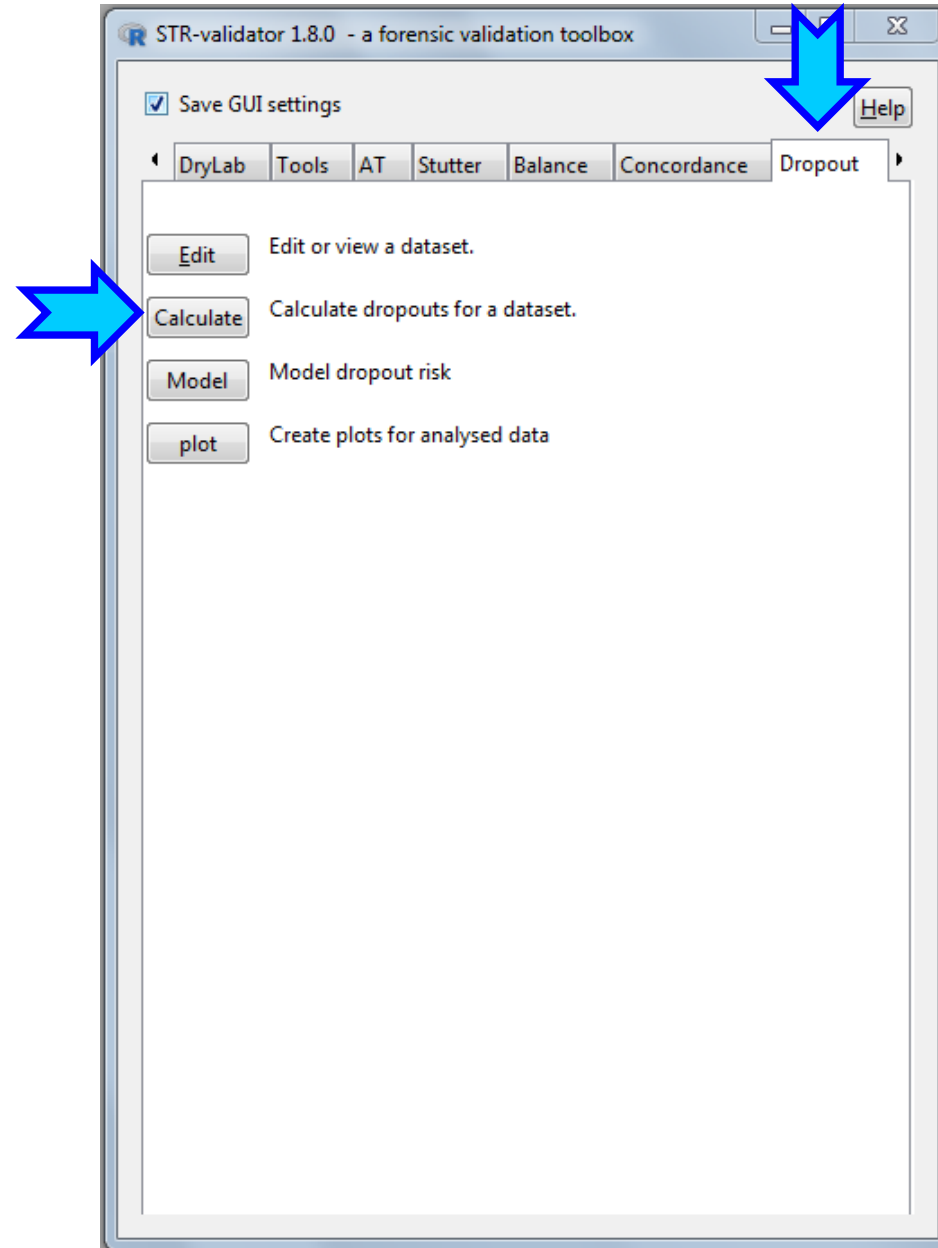
Instead of Import, Click on Open to Import *Stochastic_Threshold_Analysis.RData*



- ❖ Amount
- ❖ Reference Set
- ❖ Data Set



Calculate Dropouts for Set7



Four Methods to Score Drop-out Alleles

Calculate drop-out

Save GUI settings Help

Datasets

Select dataset: Set7 66 samples

Select reference dataset: Ref 4 references

Check subsetting

Select the kit used: Fusion 6C

Options

Ignore case

Remove sex markers

Remove quality sensors

Calculate average peak height

Limit of detection threshold (LDT): 36

Drop-out scoring method for modelling of drop-out probabilities:

Score drop-out relative to the low molecular weight allele

Score drop-out relative to the high molecular weight allele

Score drop-out relative to a random allele

Score drop-out per locus

Save as

Name for result: Set7_dropout

Calculate dropout

- ❖ Drop-Out= Allele with a peak height lower than the limit of detection threshold (LDT).
- ❖ LDT is not the AT. The lowest peak height in the dataset is automatically suggested in the 'Limit of Detection Threshold' field.

Drop out Scoring Result

Edit or view data frame

Save GUI settings Help

Datasets
Select dataset: Set7_dropout 66 samples, 17 columns, 3257 rows

Options
 Show attributes (separate window)
 Limit number of rows to: 100

Copy | Export | Save
 Set7_dropout_edit

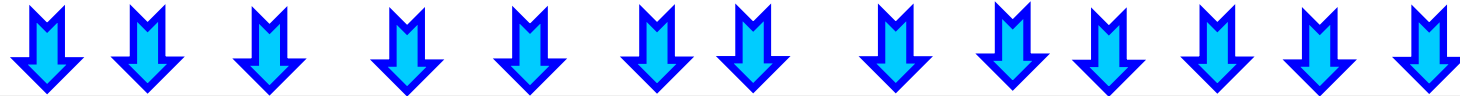
Data frame

Row.names	Sample.Name	Marker	Allele	Height	Dropout	Rf	Method2	MethodL	MethodL.Ph	TPH	H	Peaks	Expected	Proportion
1061	31_0.03ng_A	Penta E	10	239	1	239		1	239	5347	111.3958	44	50	0.88
2297	32_0.06ng_E	D7S820	8	219	1	219		1	219	12645	238.5849	49	50	0.98
1123	31_0.03ng_B	TPOX	6	200	1	200		1	200	5621	117.1042	44	50	0.88
2164	32_0.03ng_B	SE33	17	191	1	191	1	1	191	5881	113.0962	48	50	0.96
21	29_0.03ng_A	CSF1PO	12	183	1	183	1	0	1	4985	101.7347	46	51	0.9019608
1206	31_0.06ng_D	D7S820	10	180	1	180	1	1	180	10505	198.2075	49	50	0.98
1098	31_0.03ng_B	D1S1656	14	179	1	179	1	1	179	5621	117.1042	44	50	0.88
96	29_0.03ng_C	D2S441	10	178	1	178	1	1	178	3489	77.53333	42	51	0.8235294
2188	32_0.03ng_C	D2S1338	20	174	1	174	1	0	1	6356	127.12	46	50	0.92
1088	31_0.03ng_A	D22S1045	17	173	1	173	1	0	1					
1167	31_0.03ng_C	D12S391	17	162	1	162	1	1	1					
1066	31_0.03ng_A	D2S1338	21	159	1	159	1	1	0					
2221	32_0.06ng_D	D3S1358	14	154	1	154	1	1	1					
1079	31_0.03ng_A	TPOX	6	152	1	152	1	0	1	5347	111.3958	44	50	0.88
1168	31_0.03ng_C	D19S433	14	139	1	139	1	1	1	4609	115.225	37	50	0.74
1169	31_0.03ng_C	SE33	20	136	1	136	1	1	0	4609	115.225	37	50	0.74
1307	31_0.06ng_F	D5S818	13	131	1	131	1	0	1	9770	187.8846	48	50	0.96
1152	31_0.03ng_C	D2S1338	19	127	1	127	1	1	0	4609	115.225	37	50	0.74
27	29_0.03ng_A	D21S11	31	124	1	124	1	1	1	4985	101.7347	46	51	0.9019608
157	29_0.06ng_D	TH01	9.3	118	1	118	1	0	1	7653	144.3962	50	51	0.9803922
1124	31_0.03ng_B	D8S1179	12	116	1	116	1	1	0	5621	117.1042	44	50	0.88
2153	32_0.03ng_B	D7S820	13	112	1	112	1	0	1	5881	113.0962	48	50	0.96
1161	31_0.03ng_C	D7S820	10	111	1	111	1	0	1	4609	115.225	37	50	0.74

Sort Column "RFU" (PH of Surviving Allele) by decreasing order

❖ The tallest peak with drop-out of the sister allele is 239 and observed in Penta E.

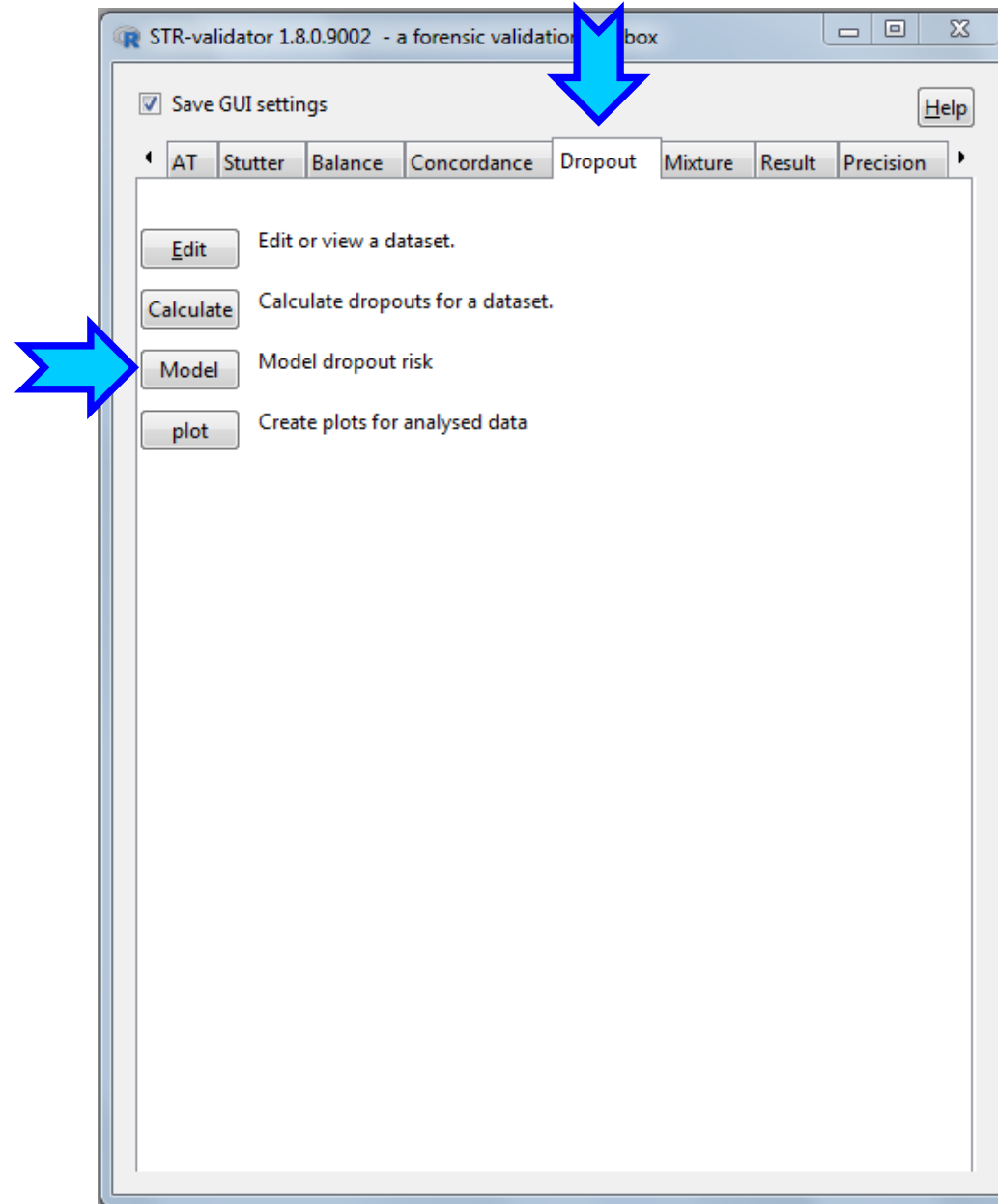
Drop out Scoring Result



Row.names	Sample.Name	Marker	Allele	Height	Dropout	Rfu	Heterozygous	MethodX	Method1	Method2	MethodL	MethodL.Ph	TPH	H	Peaks	Expected	Proportion
1061	31_0.03ng_A	Penta E	10	239	1	239	1	1	1	0	1	239	5347	111.3958	44	50	0.88
2297	32_0.06ng_E	D7S820	8	219	1	219	1	1	1	0	1	219	12645	238.5849	49	50	0.98
1123	31_0.03ng_B	TPOX	6	200	1	200	1	0	1	0	1	200	5621	117.1042	44	50	0.88
2164	32_0.03ng_B	SE33	17	191	1	191	1	0	0	1	1	191	5881	113.0962	48	50	0.96
21	29_0.03ng_A	CSF1PO	12	183	1	183	1	0	1	0	1	183	4985	101.7347	46	51	0.9019608
1206	31_0.06ng_D	D7S820	10	180	1	180	1	1	0	1	1	180	10505	198.2075	49	50	0.98
1098	31_0.03ng_B	D1S1656	14	179	1	179	1	1	1	0	1	179	5621	117.1042	44	50	0.88
96	29_0.03ng_C	D2S441	10	178	1	178	1	1	1	0	1	178	3489	77.53333	42	51	0.8235294

- ❖ Dropout: 0 (no dropout), 1 (allele dropout), and 2 (locus dropout)
- ❖ Rfu: height of surviving allele
- ❖ Heterozygous: 1 for heterozygous and 0 for homozygous
- ❖ Average Peak Height (H) for each sample
- ❖ Total peak Height for each sample
- ❖ Number of Peaks
- ❖ Number of expected peaks
- ❖ Profile Proportion
- ❖ Drop-out is scored: relative to random allele (Method X); if HMW allele is missing (Method 1); if LMW allele is missing (Method 2); if any of the alleles are missing (Method L).

Model Drop-out



Plot Drop-out Prediction

Plot dropout prediction

Save GUI settings Help

Dataset
Select dataset: and the kit used:

Options

Override automatic titles.

Plot title:

X title:

Y title:

Dataset peak height range: 36 - 12813 RFU

Log (Height)

Exclude sex markers

NB! Currently, the recommended methods are the first three options. The fourth alternative has not been evaluated by the DNA Commission. See 'Details' in 'Help' for more information.

Model drop-out from scoring method:

Relative a random allele and peak height of surviving allele

Relative the low molecular weight allele and peak height of surviving allele

Relative the high molecular weight allele and peak height of surviving allele

Relative the locus and peak height of surviving allele, or mean locus peak height

Use average peak height 'H' instead of allele/locus peak height

Print model

Drop-out prediction and threshold

Data points

Axes

NB! Must provide both min and max value.

Limit Y axis (min-max)

Limit X axis (min-max)

X labels

Plot drop-out data

Save as
Name for result: Save as object Save as image

Drop-out prediction and threshold

Mark threshold @ P(D):

Line type Line colour

Print threshold value

Prediction interval:

Print conservative T value

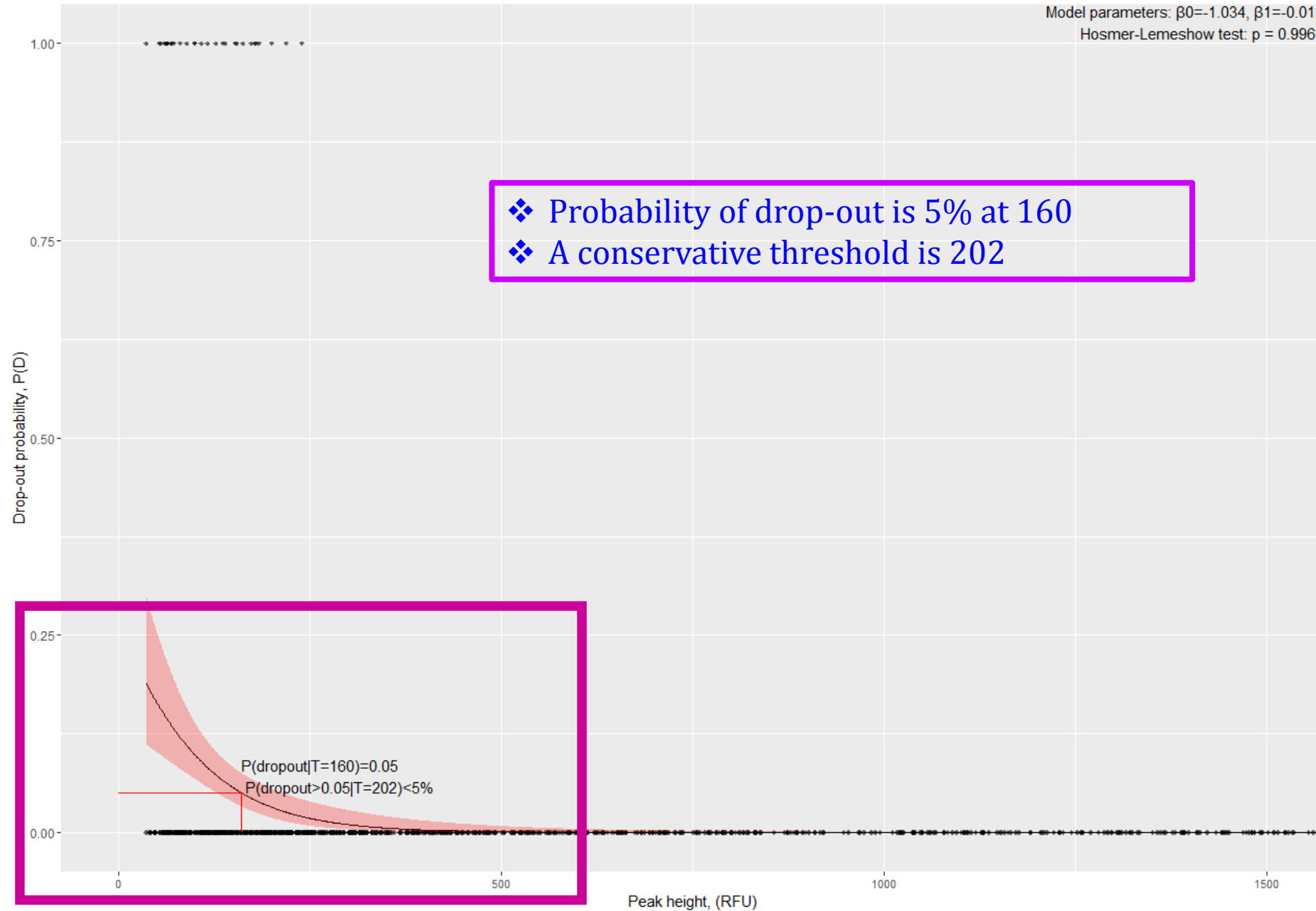
Draw prediction interval: Alpha Fill colour

Probability of drop-out modelled by logistic regression

Drop-out probability as a function of present-allele height

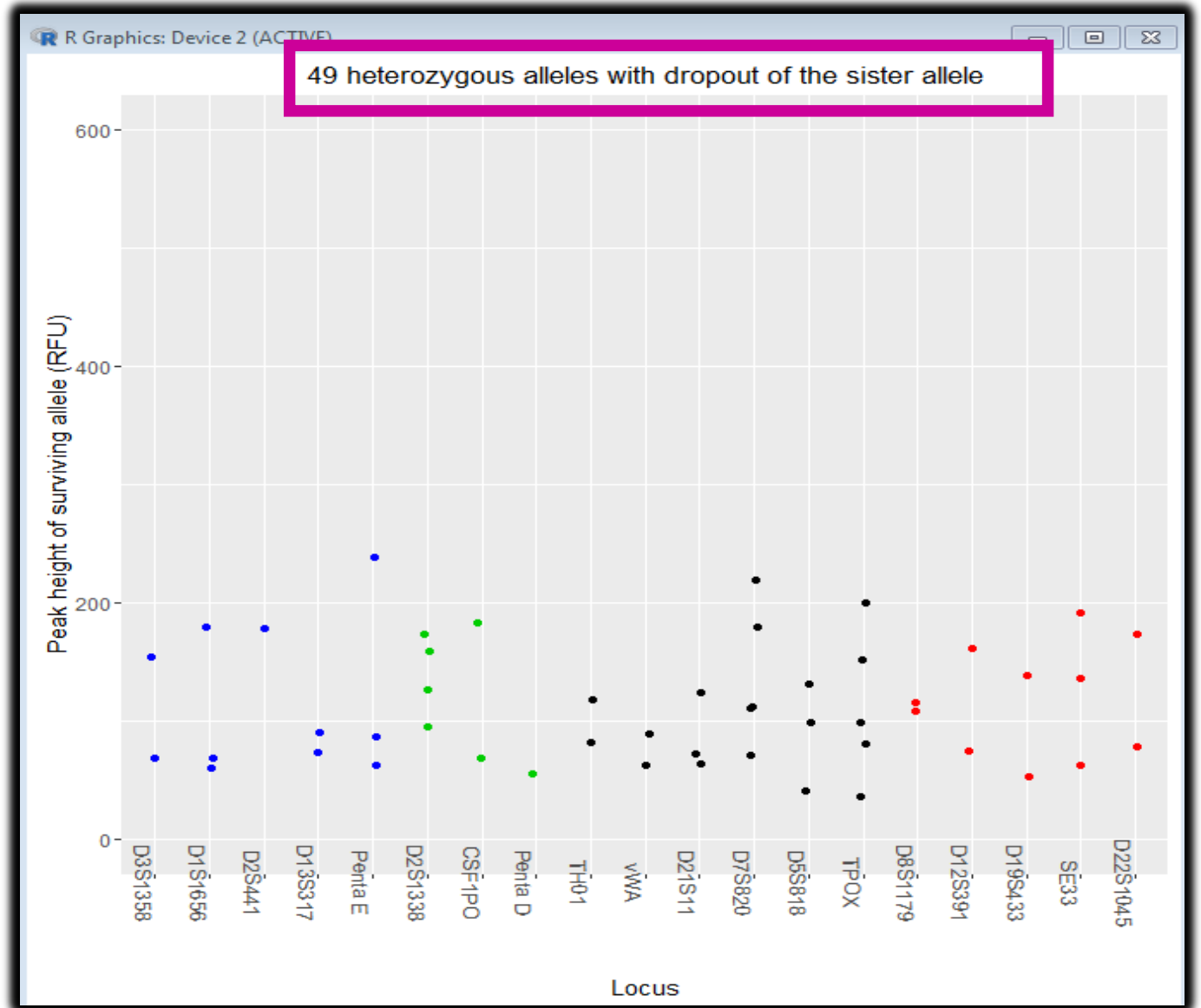
Model parameters: $\beta_0 = -1.034$, $\beta_1 = -0.012$

Hosmer-Lemeshow test: $p = 0.9969$



Plot Drop-Out Data

Dot-plot



Plot Drop out Data

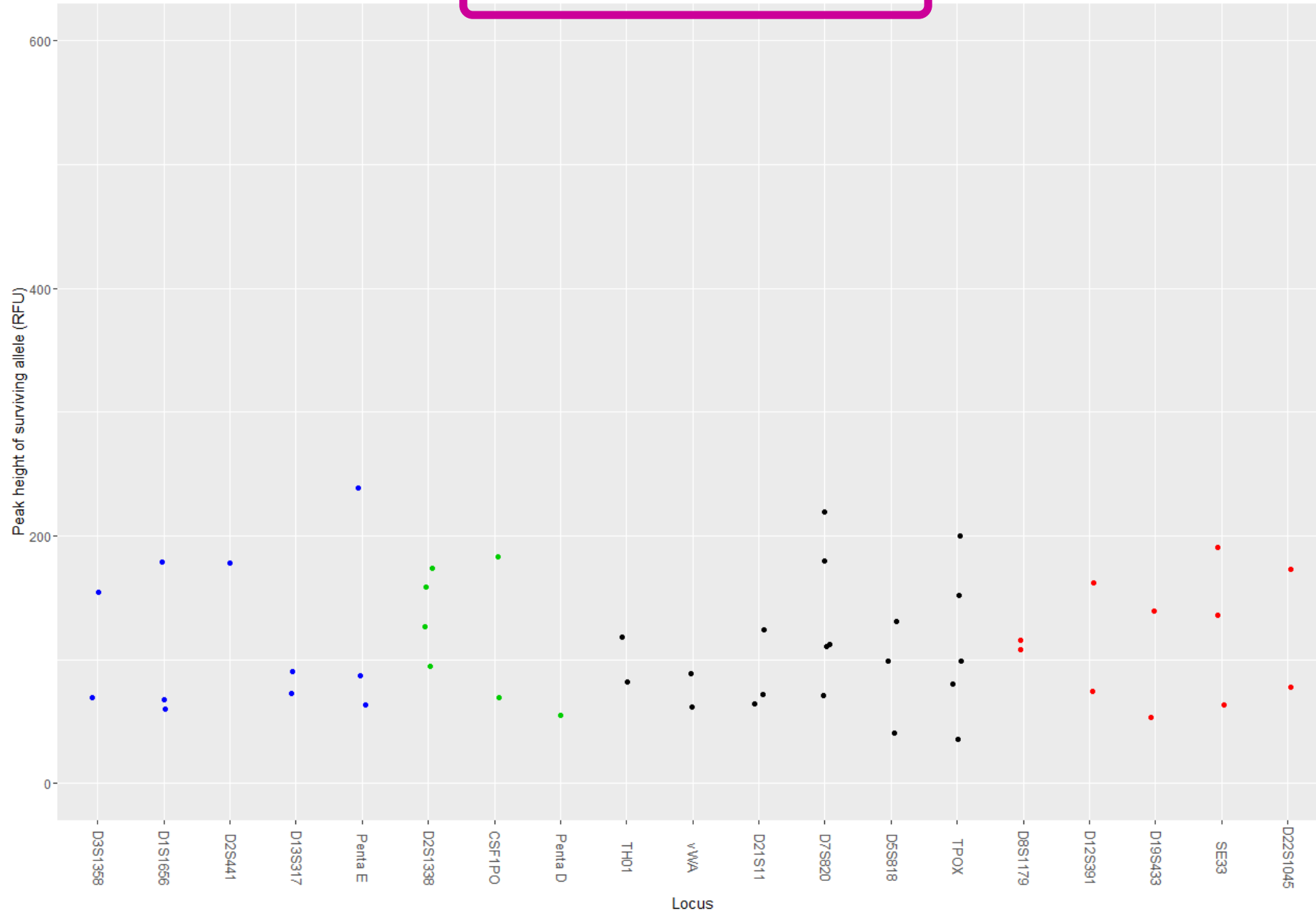
The screenshot shows the 'Plot dropout data' window with the following settings:

- Save GUI settings
- Dataset and kit: Select dataset: **Set7_dropout** and the kit used: **Fusion 6C**
- Options:
 - Override automatic titles.
 - Plot title: []
 - X title: []
 - Y title: []
 - Axes (applies to continuous axes)
 - Limit Y axis (min-max): 0 | 600
 - Limit X axis (min-max): [] | []
 - X labels
- Plot heatmap by: Average peak height | Amount | **Concentration** | Sample
- Other plots: ecdp Plot homozygous peaks. **Dotplot**
- Save as: Name for result: Set7_dropout_ggplot | Save as object | Save as image

Two blue arrows point to the 'Dataset and kit' section and the 'Limit Y axis' section. A pink box highlights the 'Dotplot' button in the 'Other plots' section.

Drop out Events by Marker

49 heterozygous alleles with dropout of the sister allele



Plot Heat-map from the Drop-out Data

The image shows the 'Plot dropout data' dialog box in R. The 'Dataset and kit' section has 'Set7_dropout' selected for the dataset and 'Fusion 6C' for the kit. The 'Options' section includes 'Override automatic titles' (unchecked), 'Axes (applies to continuous axes)' (checked), and 'Limit Y axis (min-max)' set to 0 and 600. The 'Plot heatmap by' section has 'Amount' selected and circled in pink. The 'Other plots' section has 'ecdp' selected. The 'Save as' section has 'Name for result: Set7_dropout_ggplot'. A blue arrow points to the 'Dataset and kit' section, and another blue arrow points to the 'Limit Y axis' section. An error message dialog box is overlaid on the bottom right, displaying a red prohibition sign and the text 'Additional columns required: Amount'. The 'OK' button is visible at the bottom right of the message dialog.

Plot dropout data

Save GUI settings Help

Dataset and kit

Select dataset: Set7_dropout and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Axes (applies to continuous axes)

Limit Y axis (min-max)

0 600

Limit X axis (min-max)

X labels

Plot heatmap by

Average peak height **Amount** Concentration

Other plots

ecdp Plot homozygous peaks.

Save as

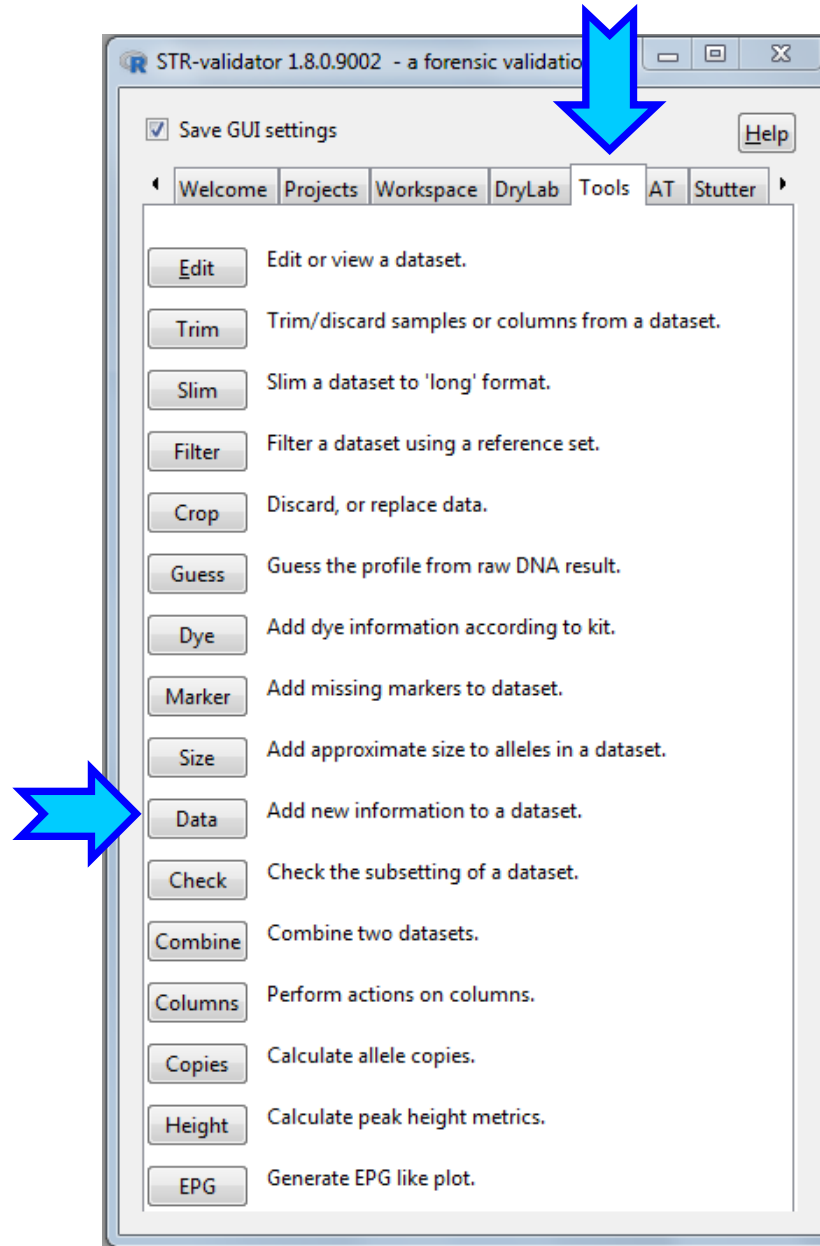
Name for result: Set7_dropout_ggplot

message

Additional columns required:
Amount

OK

Add Amount Information to Set7_Dropout Dataset



Add Amount Information to Set7 Dropout Dataset

Save GUI settings Help

Datasets

Select destination dataset: Set7_dropout 66 samples

Select source dataset: Amount 66 samples

Options

Exact key matching

Ignore case

Select key column: Sample.Name

Select second key column: Marker

Select columns to add to the new dataset: Amount

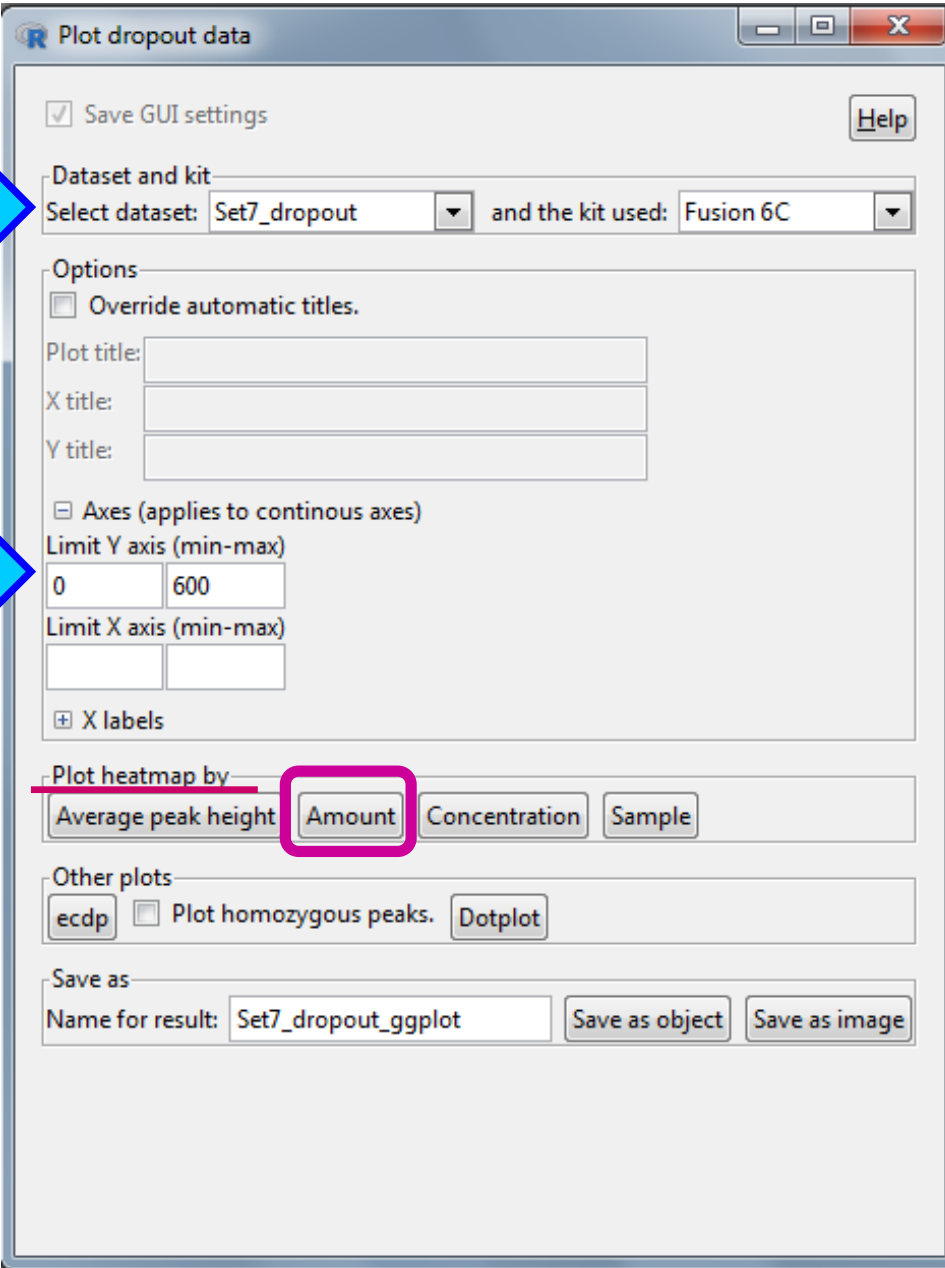
Amount

Save as

Name for result: Set7_dropout_amount

Add new data

Plot Heat-map from the Drop-out Data



The image shows a software window titled "Plot dropout data" with various configuration options. Two blue arrows point to the "Dataset and kit" section and the "Limit Y axis" section. The "Amount" button in the "Plot heatmap by" section is highlighted with a pink circle.

Save GUI settings Help

Dataset and kit
Select dataset: Set7_dropout and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Axes (applies to continuous axes)

Limit Y axis (min-max)

Limit X axis (min-max)

X labels

Plot heatmap by

Other plots
 Plot homozygous peaks.

Save as
Name for result:

Plot Heat-map from the Drop-out Data by Sample Name

Plot dropout data

Save GUI settings Help

Dataset and kit

Select dataset: Set7_dropout and the kit used: Fusion 6C

Options

Override automatic titles.

Plot title:

X title:

Y title:

Axes (applies to continuous axes)

Limit Y axis (min-max)

0 600

Limit X axis (min-max)

X labels

Plot heatmap by

Average peak height Amount Concentration **Sample**

Other plots

ecdp Plot homozygous peaks. Dotplot

Save as

Name for result: Set7_dropout_ggplot Save as object Save as image

Summary of Thresholds

49 Heterozygote allele with a drop-out
of the sister allele

Analysis of Data based on Analytical Method: AT7	Stochastic Threshold	Conservative Stochastic Threshold
Scoring drop-out relative to the LMW allele	160	202
Scoring drop-out relative to the HMW allele	122	157
Scoring drop-out relative to a random allele	138	182
Scoring drop-out per locus	193	227

Acknowledgments:

NIST

❖ *Peter Vallone*

❖ *Erica Romsos*

❖ *Lisa Borsuk*

Norwegian Institute of Public Health

❖ Oskar Hansson

Contact Info:

sarah.riman@nist.gov

(301) 975-4162

References

1. <https://sites.google.com/site/forensicapps/strvalidator>
2. <https://github.com/OskarHansson/strvalidator>
3. <https://cran.r-project.org/web/packages/strvalidator/index.html>
4. O. Hansson, P. Gill, T. Egeland, STR-validator: An open source platform for validation and process control, *Forensic Science International: Genetics* 13 (2014) 154–166.
5. P. Gill, L. Gusmao, H. Haned, W. Mayr, N. Morling, W. Parson, L. Prieto, M. Prinz, H. Schneider, P. Schneider, B. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods, *Forensic Science International: Genetics* 6 (2012) 679–688.
6. Peter Gill, Roberto Puch-Solis, James Curran, The low-template-DNA (stochastic) threshold-Its determination relative to risk analysis for national DNA databases, *Forensic Science International: Genetics*, Volume 3, Issue 2, March 2009, Pages 104-111
7. Torben Tvedebrink, Poul Svante Eriksen, Helle Smidt Mogensen, Niels Morling, Evaluating the weight of evidence by using quantitative short tandem repeat data in DNA mixtures *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Volume 59, Issue 5, 2010, Pages 855-874,
8. J. Bregu et al. Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis. *JFS* (2013) 1 pg 120-129.
9. Ullrich J. Monich, Ken Duy, Muriel Medard, Viveck Cadambe, Lauren E. Alfonse, and Catherine Grgicak. Probabilistic characterisation of baseline noise in STR proles. *Forensic Science International: Genetics*.
10. J.-A. Bright, J. Turkington, J. Buckleton. Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex. *Forensic Sci. Int. Genet.*, 4 (2) (2010), pp. 111–114.