Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

STR validator

Oskar Hansson

Statistical methods in forensic genetics 20-23 May 2013, Copenhagen

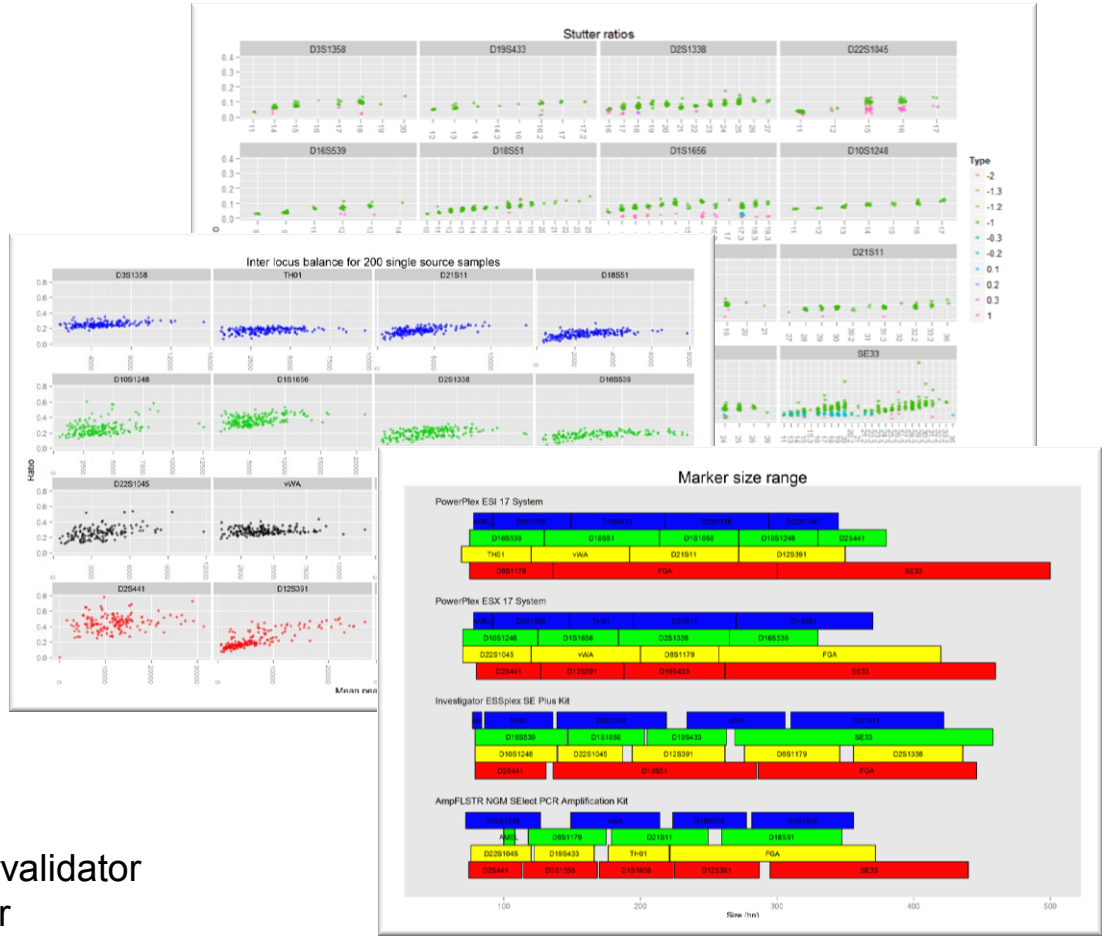
EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

STR validator

- R-package developed by Oskar Hansson at the Norwegian Institute of Public Health (NIPH)
- Validation of forensic STR DNA typing kits
- Process control
- Compare methods and instrumentation



https://sites.google.com/site/forensicapps/strvalidator
https://github.com/OskarHansson/strvalidator
http://cran.r-project.org/web/packages/strvalidator/index.html
https://www.facebook.com/pages/STR-validator/240891279451450?ref=tn_tnmn

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN
Network of Excellence

Process control

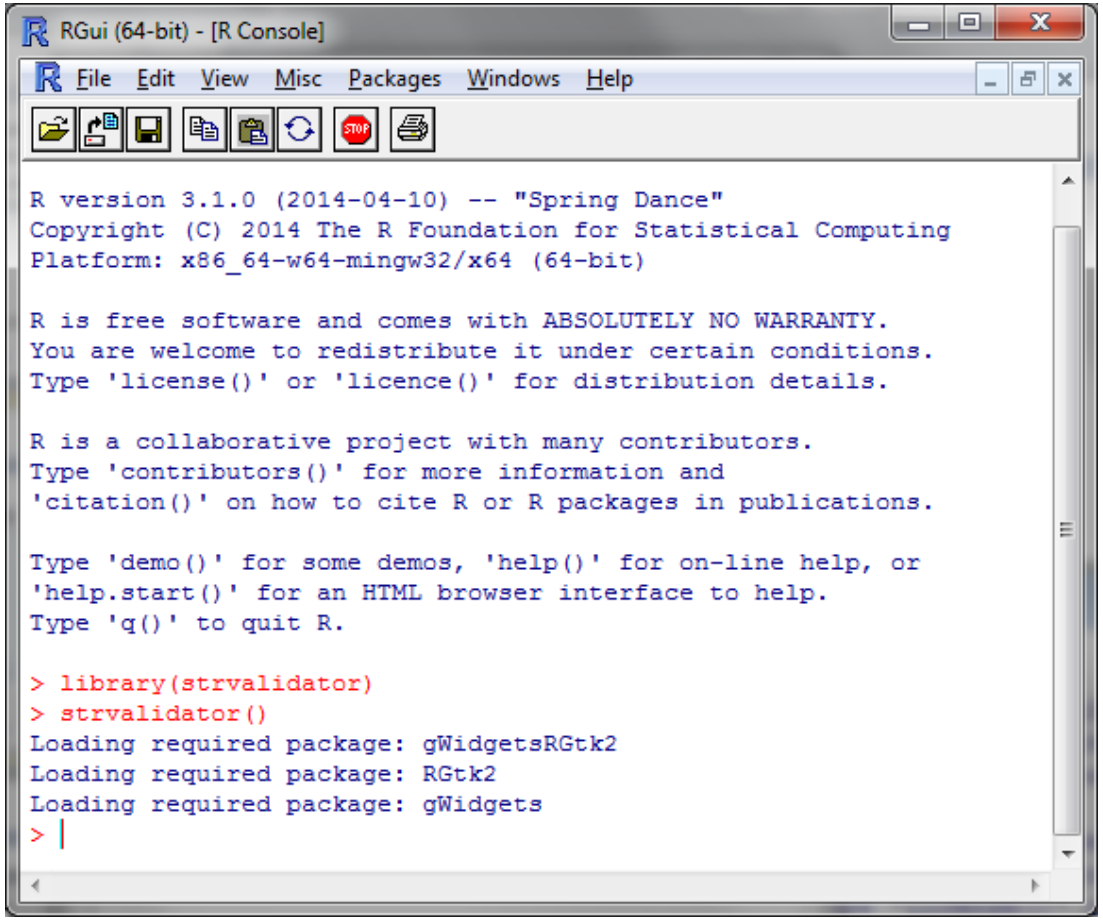
monitor the level of contamination

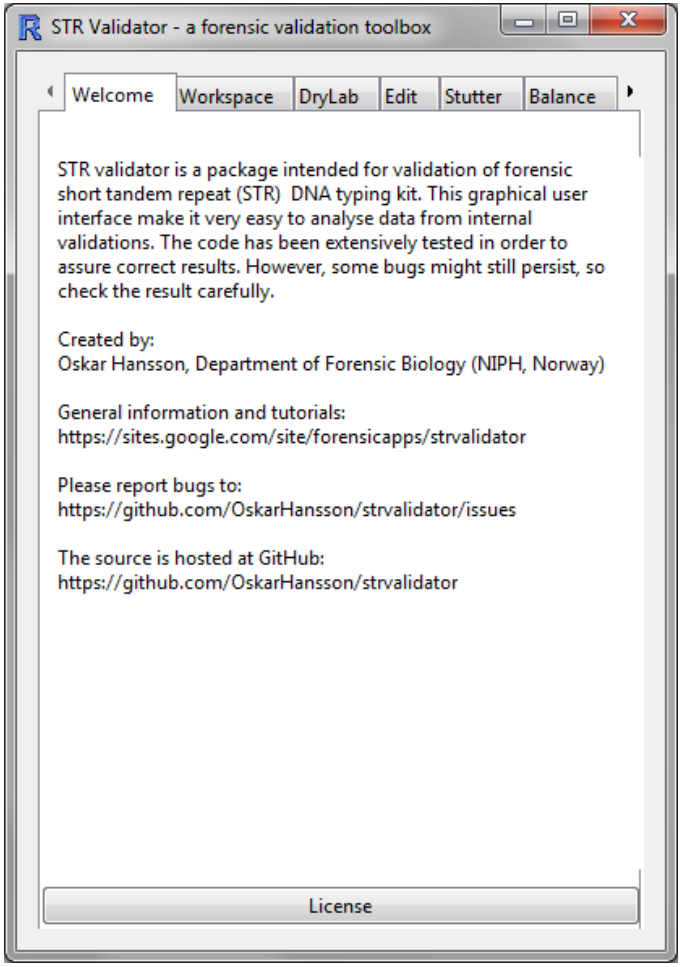
EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

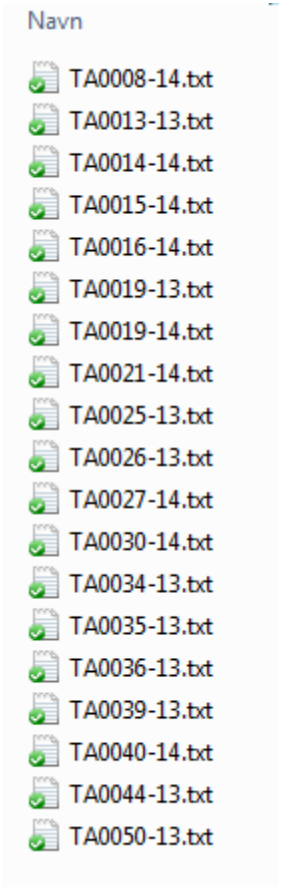
Open the STR validator GUI





EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Data exported from GeneMapper



- Genotypes table exported for all batches/samples
- Plain tab separated text files
- Name consist of batch number and year

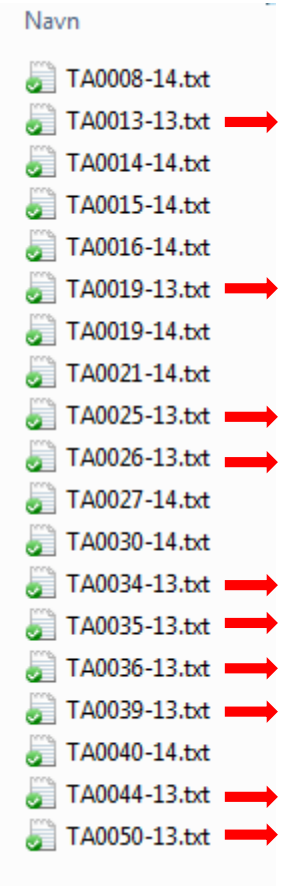
TA0008-14.txt - Notisblokk

Fil Rediger Format Vis Hjelp

Sample Name	Marker	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5
01-Pos.ktr.	AMEL	X	Y		3517	2856
01-Pos.ktr.	D3S1358	17	18		6089	4574
01-Pos.ktr.	TH01	6	9.3		2962	3365
01-Pos.ktr.	D21S11	29	31.2		4301	3806
01-Pos.ktr.	D18S51	16	18		1968	1939
01-Pos.ktr.	D10S1248	13	15		5061	3272
01-Pos.ktr.	D1S1656	12	13		4289	4399
01-Pos.ktr.	D2S1338	22	25		4247	3427
01-Pos.ktr.	D16S539	9	13		4115	3953
01-Pos.ktr.	D22S1045		16			4431
01-Pos.ktr.	VWA	16	19		2614	2933
01-Pos.ktr.	D8S1179	14	15		3402	3043
01-Pos.ktr.	FGA	20	23		1987	2926
01-Pos.ktr.	D2S441	10	14		2694	4064
01-Pos.ktr.	D12S391	18	23		14087	5530
01-Pos.ktr.	D19S433	13	14		2913	2153
01-Pos.ktr.	SE33	15	16		1673	1918
02-Neg.ktr.	AMEL					
02-Neg.ktr.	D3S1358					
02-Neg.ktr.	TH01					

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

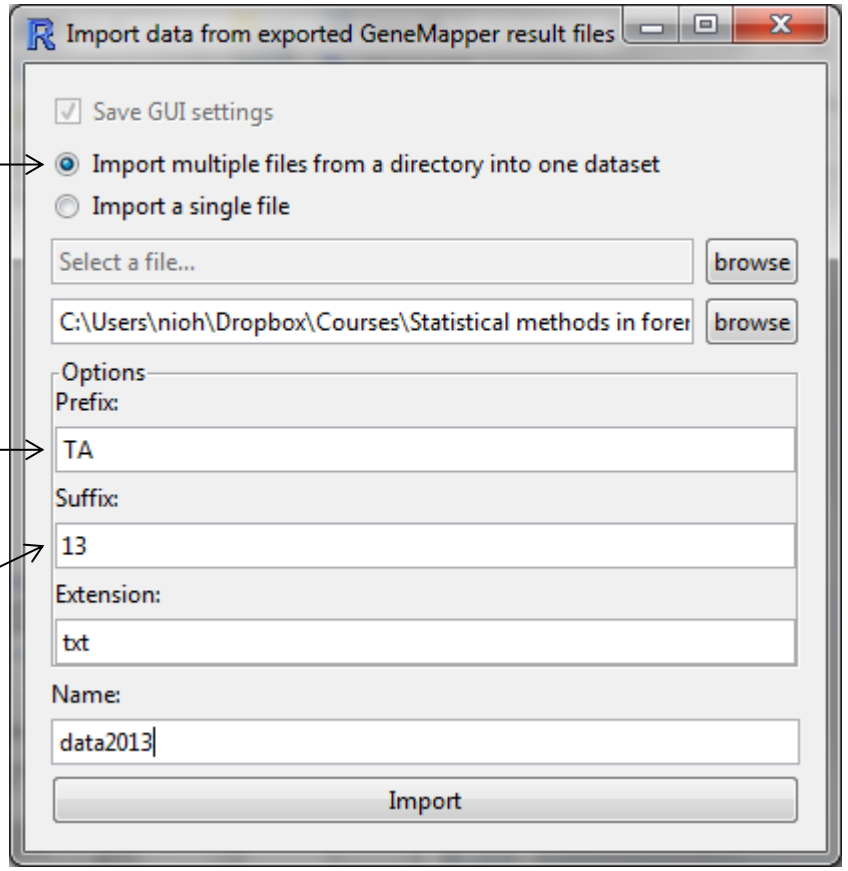
Import data



Import multiple files

Import TA files

Import results from 2013

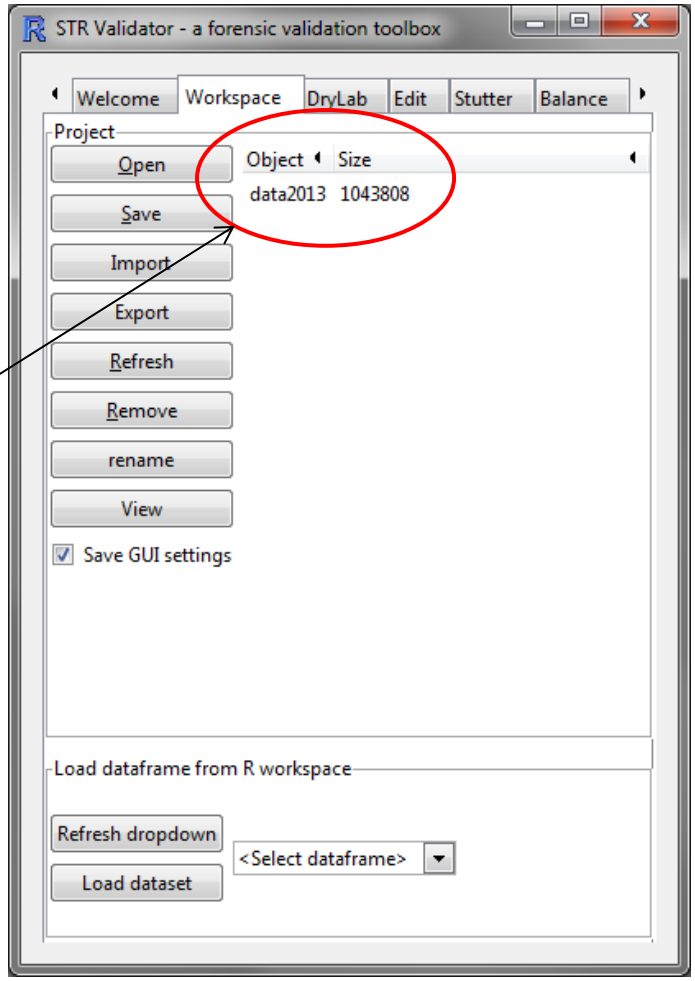


STR validator \\ Tab: Workspace \\ Button: Import

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Project workspace

The dataset is now available in the project workspace



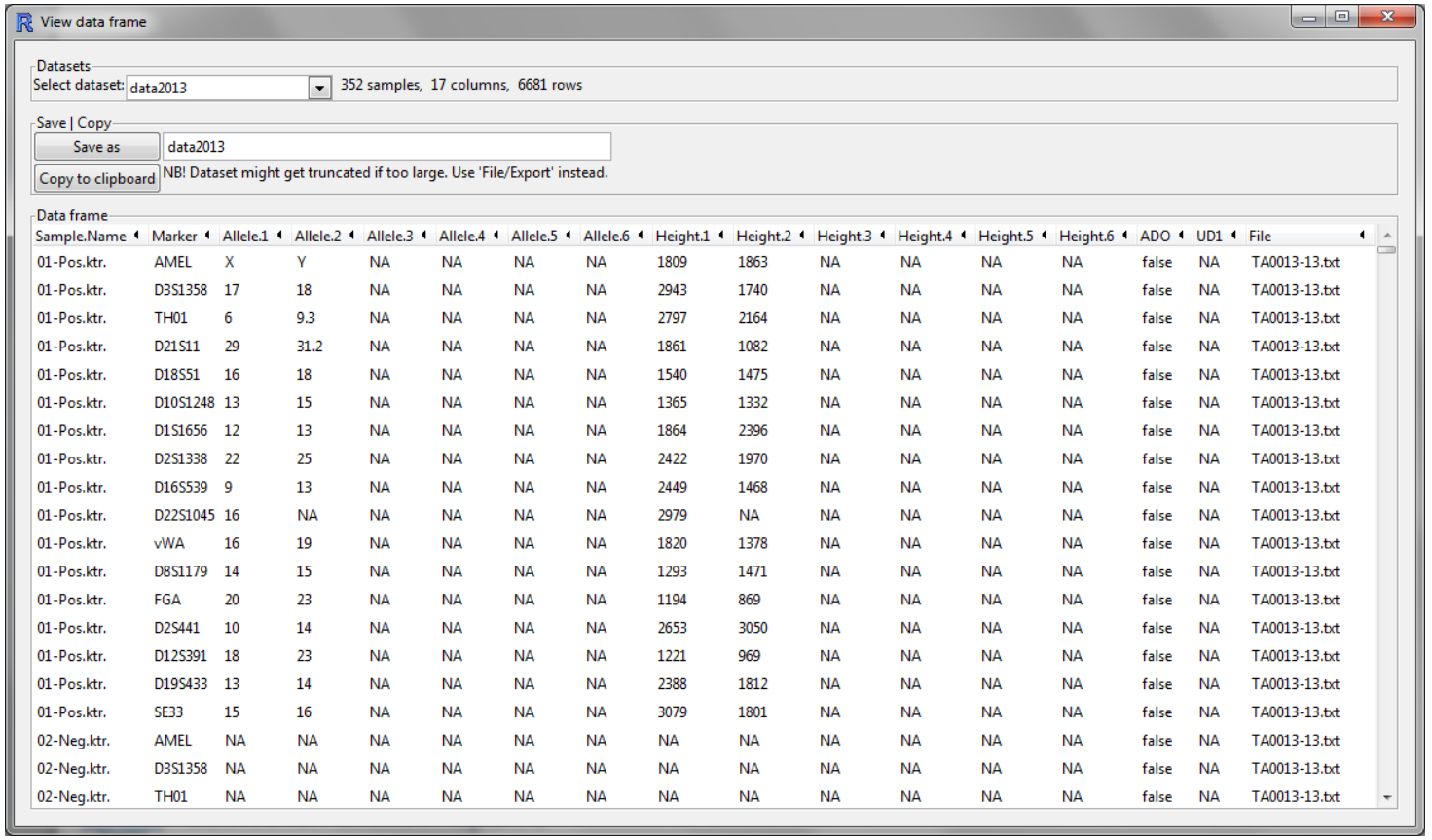
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

View dataset

- 352 unique entries in 'Sample.Name' column
- 17 columns
- 6681 rows
- Recommended to export 'Size' as well



Sample.Name	Marker	Allele.1	Allele.2	Allele.3	Allele.4	Allele.5	Allele.6	Height.1	Height.2	Height.3	Height.4	Height.5	Height.6	ADO	UD1	File
01-Pos.ktr.	AMEL	X	Y	NA	NA	NA	NA	1809	1863	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D3S1358	17	18	NA	NA	NA	NA	2943	1740	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	TH01	6	9.3	NA	NA	NA	NA	2797	2164	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D21S11	29	31.2	NA	NA	NA	NA	1861	1082	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D18S51	16	18	NA	NA	NA	NA	1540	1475	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D10S1248	13	15	NA	NA	NA	NA	1365	1332	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D1S1656	12	13	NA	NA	NA	NA	1864	2396	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D2S1338	22	25	NA	NA	NA	NA	2422	1970	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D16S539	9	13	NA	NA	NA	NA	2449	1468	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D22S1045	16	NA	NA	NA	NA	NA	2979	NA	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	vWA	16	19	NA	NA	NA	NA	1820	1378	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D8S1179	14	15	NA	NA	NA	NA	1293	1471	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	FGA	20	23	NA	NA	NA	NA	1194	869	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D2S441	10	14	NA	NA	NA	NA	2653	3050	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D12S391	18	23	NA	NA	NA	NA	1221	969	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	D19S433	13	14	NA	NA	NA	NA	2388	1812	NA	NA	NA	NA	false	NA	TA0013-13.txt
01-Pos.ktr.	SE33	15	16	NA	NA	NA	NA	3079	1801	NA	NA	NA	NA	false	NA	TA0013-13.txt
02-Neg.ktr.	AMEL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	false	NA	TA0013-13.txt
02-Neg.ktr.	D3S1358	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	false	NA	TA0013-13.txt
02-Neg.ktr.	TH01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	false	NA	TA0013-13.txt

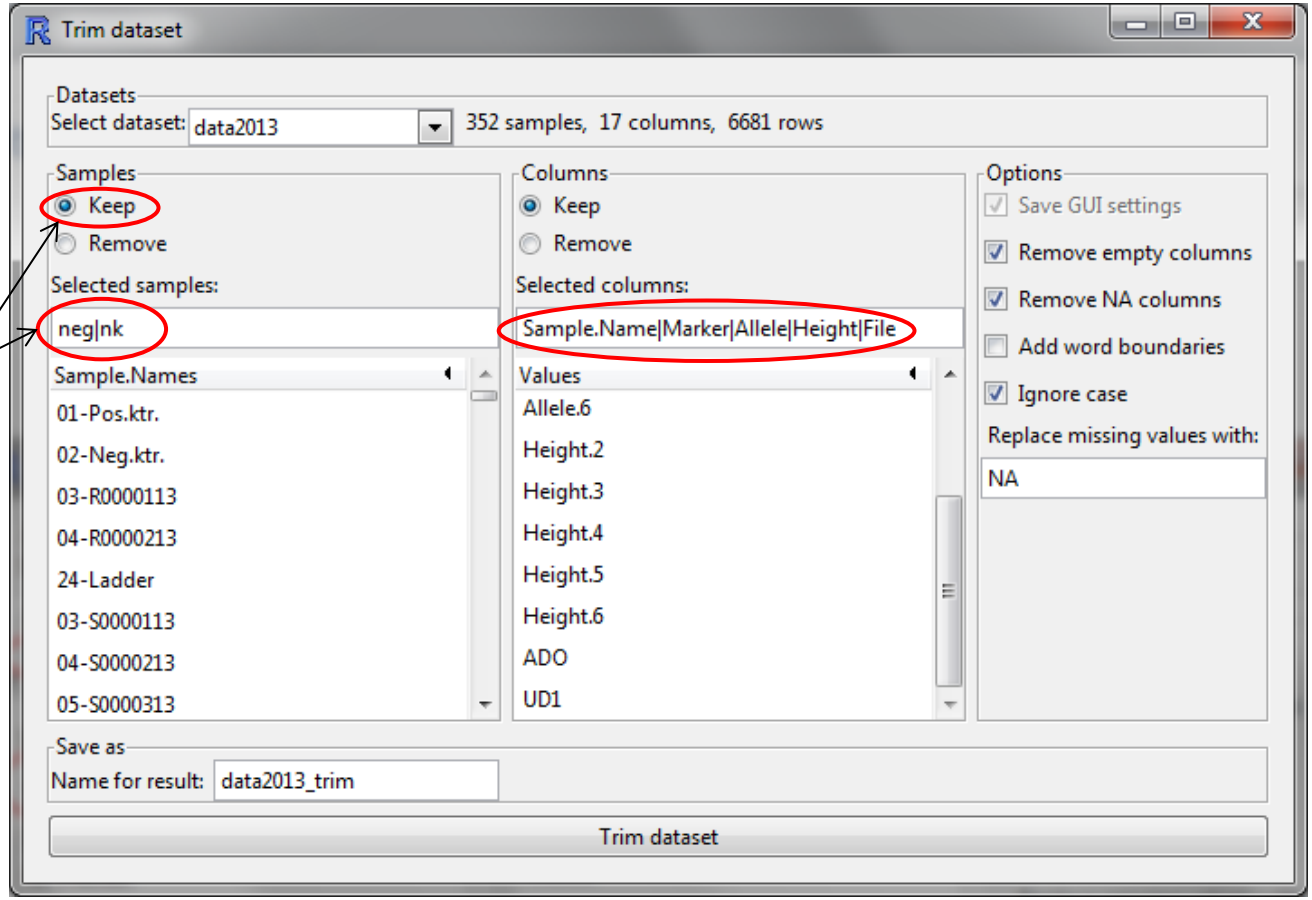
STR validator \\ Tab: Workspace \\ Button: View

SEVENTH FRAMEWORK PROGRAMME

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Trim dataset

- PCR negative control:
  Neg.ktr.
- Extraction negative control:
  NKYY####
- Keep all negative controls i.e. sample names that contain ‘neg’ or ‘nk’
- NB! The ‘File’ column is needed for this analysis

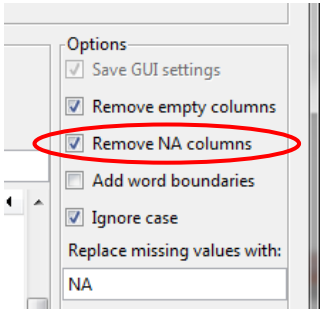


STR validator \\ Tab: Edit \\ Button: Trim

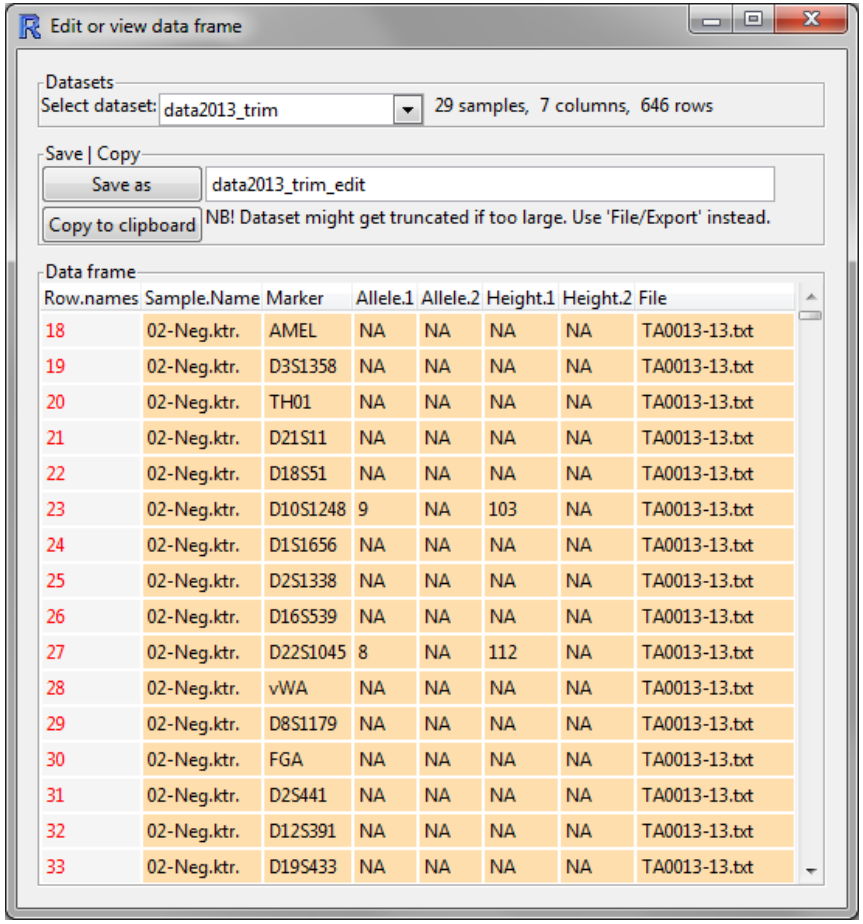
EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

View dataset

- Unwanted samples and columns have been removed
- From the number of allele columns the maximum number of alleles in any sample can be deduced (i.e. 2) because we removed NA columns while trimming:



- NB! number of samples are not correct because our PCR negative controls have identical names (and hence are not unique)



Row.names	Sample.Name	Marker	Allele.1	Allele.2	Height.1	Height.2	File
18	02-Neg.ktr.	AMEL	NA	NA	NA	NA	TA0013-13.txt
19	02-Neg.ktr.	D3S1358	NA	NA	NA	NA	TA0013-13.txt
20	02-Neg.ktr.	TH01	NA	NA	NA	NA	TA0013-13.txt
21	02-Neg.ktr.	D21S11	NA	NA	NA	NA	TA0013-13.txt
22	02-Neg.ktr.	D18S51	NA	NA	NA	NA	TA0013-13.txt
23	02-Neg.ktr.	D10S1248	9	NA	103	NA	TA0013-13.txt
24	02-Neg.ktr.	D1S1656	NA	NA	NA	NA	TA0013-13.txt
25	02-Neg.ktr.	D2S1338	NA	NA	NA	NA	TA0013-13.txt
26	02-Neg.ktr.	D16S539	NA	NA	NA	NA	TA0013-13.txt
27	02-Neg.ktr.	D22S1045	8	NA	112	NA	TA0013-13.txt
28	02-Neg.ktr.	vWA	NA	NA	NA	NA	TA0013-13.txt
29	02-Neg.ktr.	D8S1179	NA	NA	NA	NA	TA0013-13.txt
30	02-Neg.ktr.	FGA	NA	NA	NA	NA	TA0013-13.txt
31	02-Neg.ktr.	D2S441	NA	NA	NA	NA	TA0013-13.txt
32	02-Neg.ktr.	D12S391	NA	NA	NA	NA	TA0013-13.txt
33	02-Neg.ktr.	D19S433	NA	NA	NA	NA	TA0013-13.txt

STR validator \\ Button: Edit

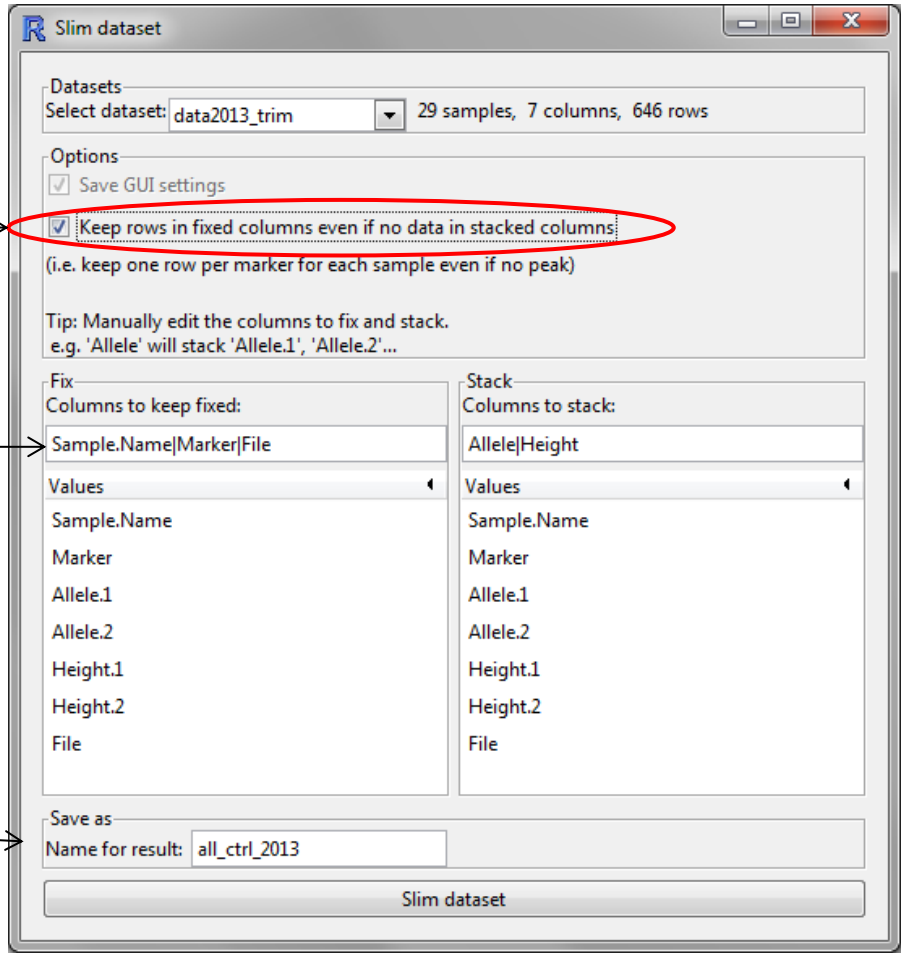
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Slim dataset

Keep rows even if no data i.e. all samples will be in the resulting dataset

Columns are automatically suggested

Change name

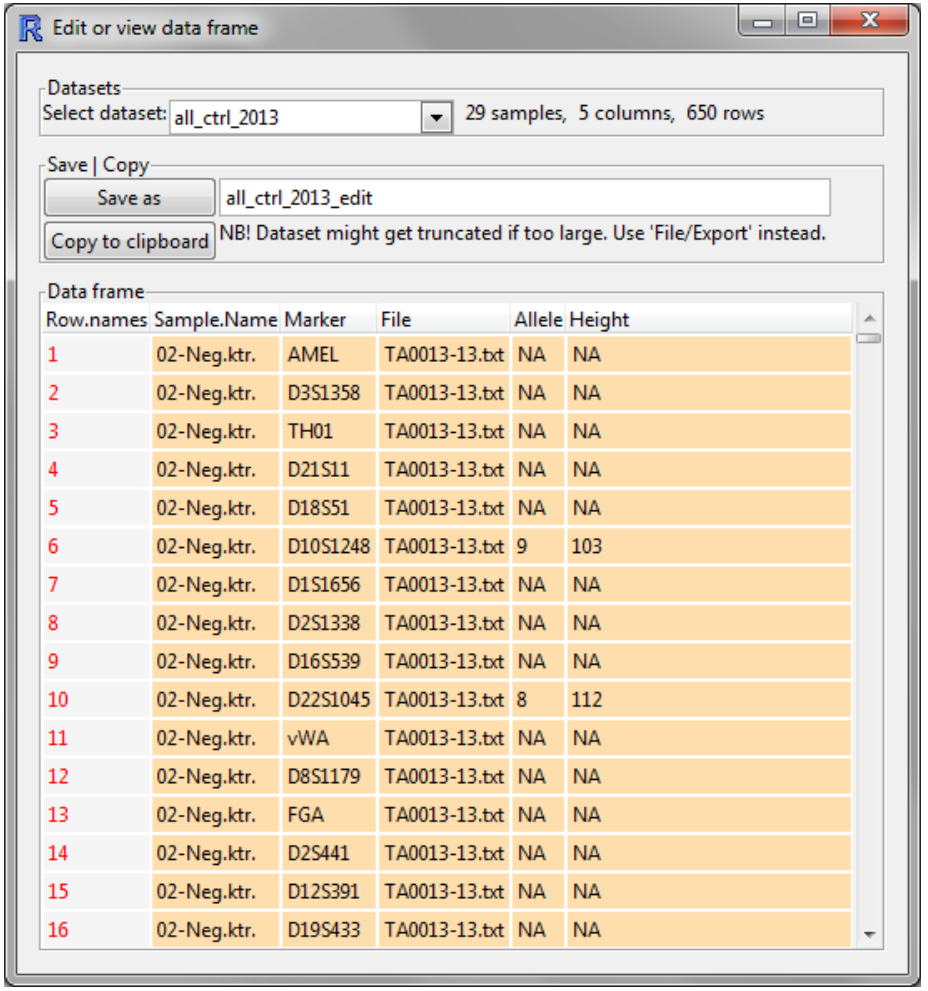


STR validator \\ Tab: Edit \\ Button: Slim

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

View dataset

- 'Allele.1' and 'Allele.2' is now combined into one column
- 'Height.1' and 'Height.2' is now combined into one column



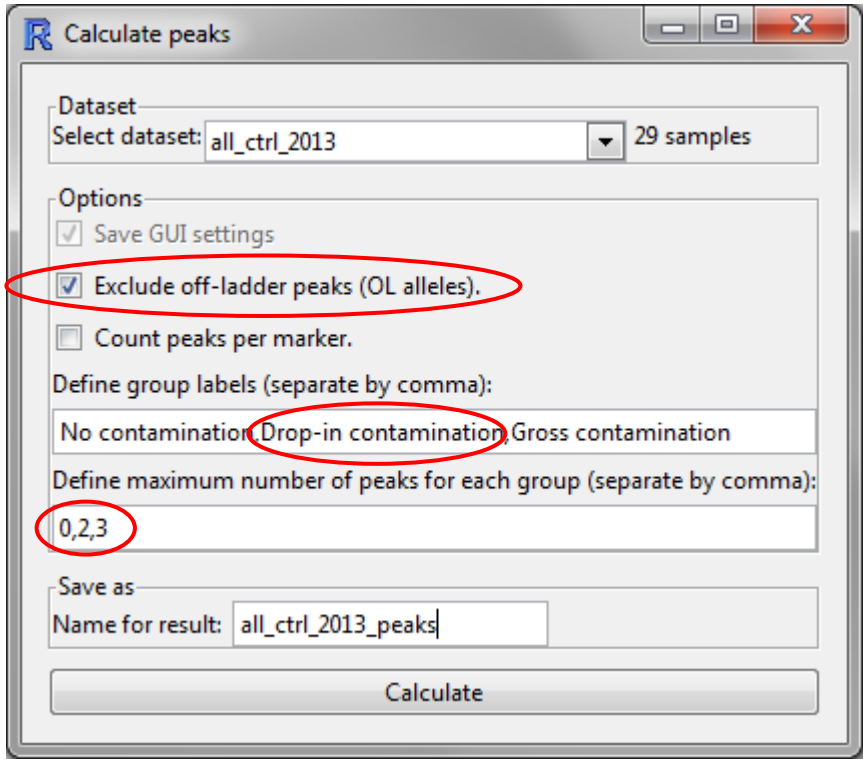
Row.names	Sample.Name	Marker	File	Allele	Height
1	02-Neg.ktr.	AMEL	TA0013-13.txt	NA	NA
2	02-Neg.ktr.	D3S1358	TA0013-13.txt	NA	NA
3	02-Neg.ktr.	TH01	TA0013-13.txt	NA	NA
4	02-Neg.ktr.	D21S11	TA0013-13.txt	NA	NA
5	02-Neg.ktr.	D18S51	TA0013-13.txt	NA	NA
6	02-Neg.ktr.	D10S1248	TA0013-13.txt	9	103
7	02-Neg.ktr.	D1S1656	TA0013-13.txt	NA	NA
8	02-Neg.ktr.	D2S1338	TA0013-13.txt	NA	NA
9	02-Neg.ktr.	D16S539	TA0013-13.txt	NA	NA
10	02-Neg.ktr.	D22S1045	TA0013-13.txt	8	112
11	02-Neg.ktr.	vWA	TA0013-13.txt	NA	NA
12	02-Neg.ktr.	D8S1179	TA0013-13.txt	NA	NA
13	02-Neg.ktr.	FGA	TA0013-13.txt	NA	NA
14	02-Neg.ktr.	D2S441	TA0013-13.txt	NA	NA
15	02-Neg.ktr.	D12S391	TA0013-13.txt	NA	NA
16	02-Neg.ktr.	D19S433	TA0013-13.txt	NA	NA

STR validator \\ Button: Edit

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Calculate peaks

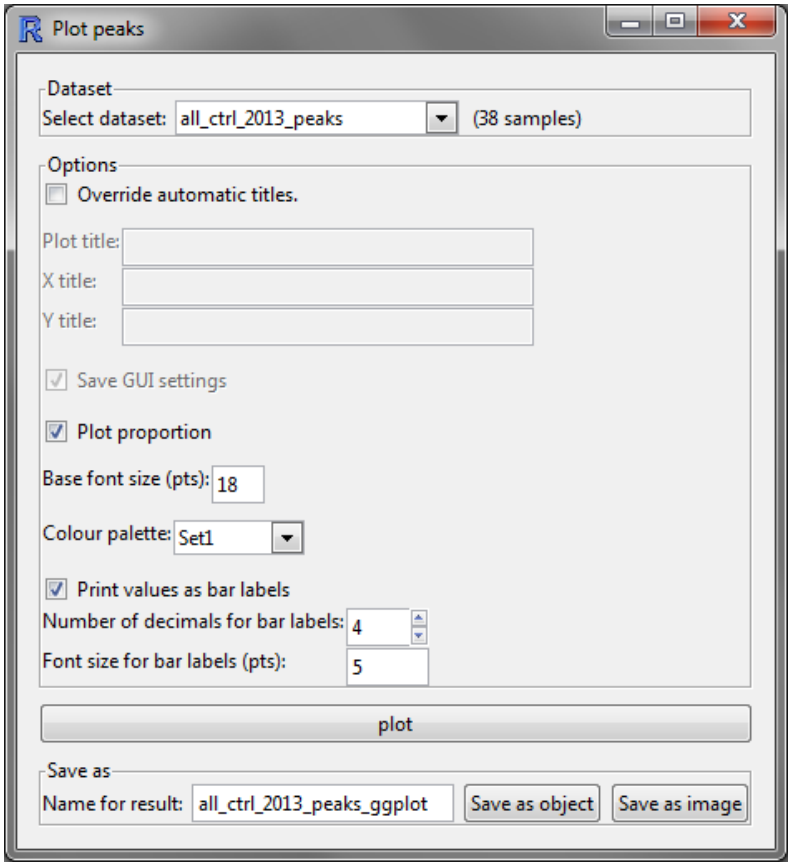
We define drop-in as one or two peaks called as alleles

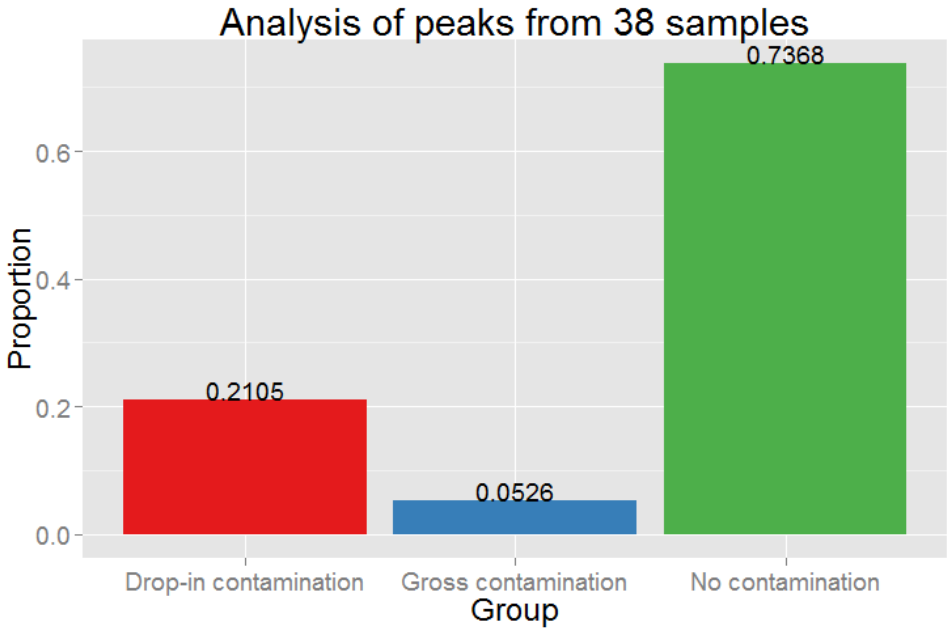


STR validator \\ Tab: Result \\ Group: Number of peaks \\ Button: Calculate

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Plot peaks data





STR validator \\ Tab: Result \\ Group: Number of peaks \\ Button: Plot

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Norwegian Institute of Public Health

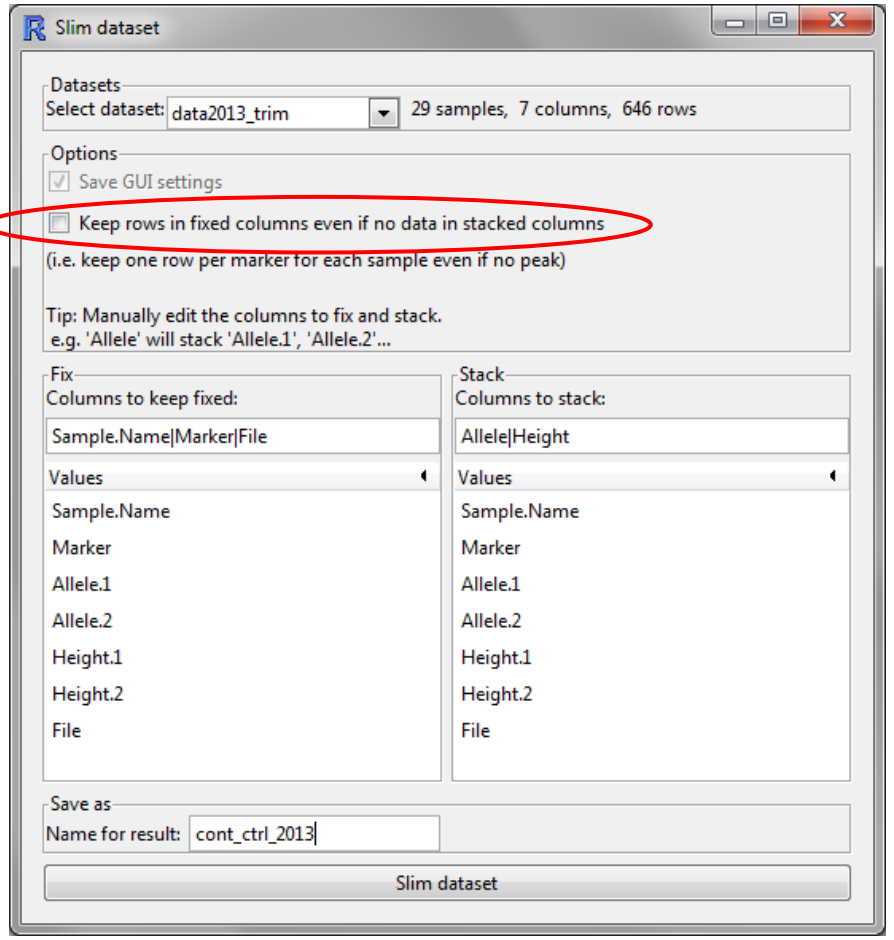
EUROFORGEN Network of Excellence

Estimate the probability of drop-in, Pr(C)

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Slim dataset

Discard rows if no data i.e. only samples with peaks will be in the resulting dataset

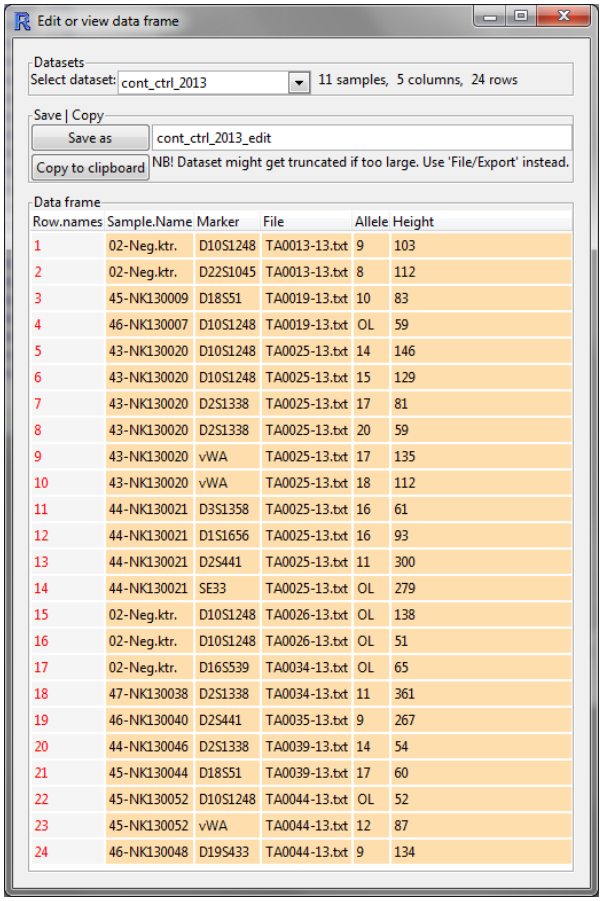


STR validator \\ Tab: Edit \\ Button: Slim

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

View dataset

- Only contaminated controls are in the new dataset
- It is easy to find the batch/file where the contamination occurred



Row.names	Sample.Name	Marker	File	Allele	Height
1	02-Neg.ktr.	D10S1248	TA0013-13.txt	9	103
2	02-Neg.ktr.	D22S1045	TA0013-13.txt	8	112
3	45-NK130009	D18S51	TA0019-13.txt	10	83
4	46-NK130007	D10S1248	TA0019-13.txt	OL	59
5	43-NK130020	D10S1248	TA0025-13.txt	14	146
6	43-NK130020	D10S1248	TA0025-13.txt	15	129
7	43-NK130020	D2S1338	TA0025-13.txt	17	81
8	43-NK130020	D2S1338	TA0025-13.txt	20	59
9	43-NK130020	vWA	TA0025-13.txt	17	135
10	43-NK130020	vWA	TA0025-13.txt	18	112
11	44-NK130021	D3S1358	TA0025-13.txt	16	61
12	44-NK130021	D1S1656	TA0025-13.txt	16	93
13	44-NK130021	D2S441	TA0025-13.txt	11	300
14	44-NK130021	SE33	TA0025-13.txt	OL	279
15	02-Neg.ktr.	D10S1248	TA0026-13.txt	OL	138
16	02-Neg.ktr.	D10S1248	TA0026-13.txt	OL	51
17	02-Neg.ktr.	D16S539	TA0034-13.txt	OL	65
18	47-NK130038	D2S1338	TA0034-13.txt	11	361
19	46-NK130040	D2S441	TA0035-13.txt	9	267
20	44-NK130046	D2S1338	TA0039-13.txt	14	54
21	45-NK130044	D18S51	TA0039-13.txt	17	60
22	45-NK130052	D10S1248	TA0044-13.txt	OL	52
23	45-NK130052	vWA	TA0044-13.txt	12	87
24	46-NK130048	D19S433	TA0044-13.txt	9	134

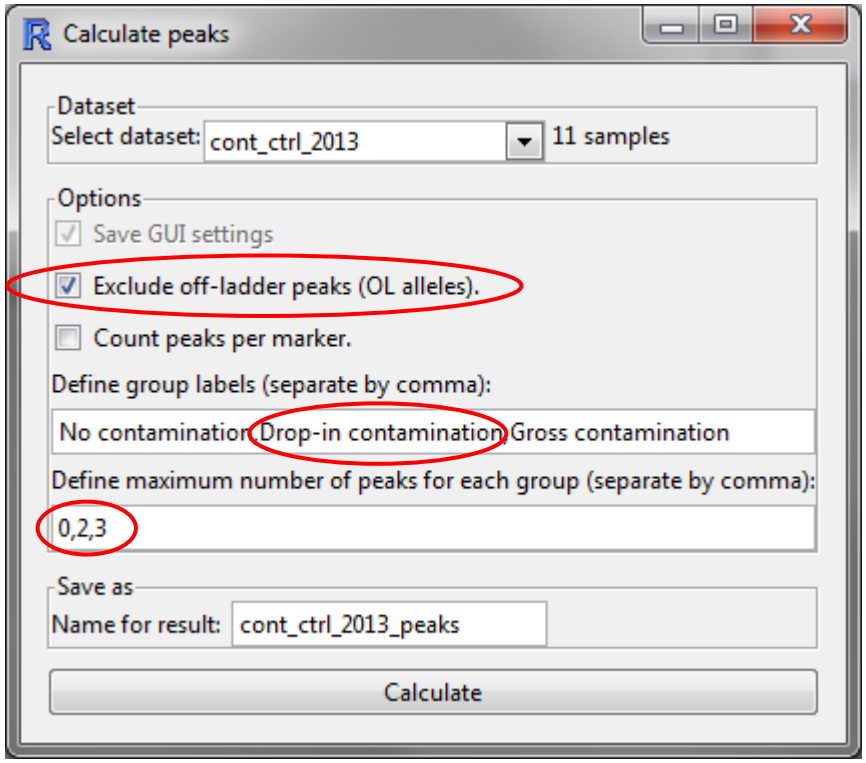
STR validator \\ Button: Edit

SEVENTH FRAMEWORK PROGRAMME

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Calculate peaks

We define drop-in as one or two peaks called as alleles

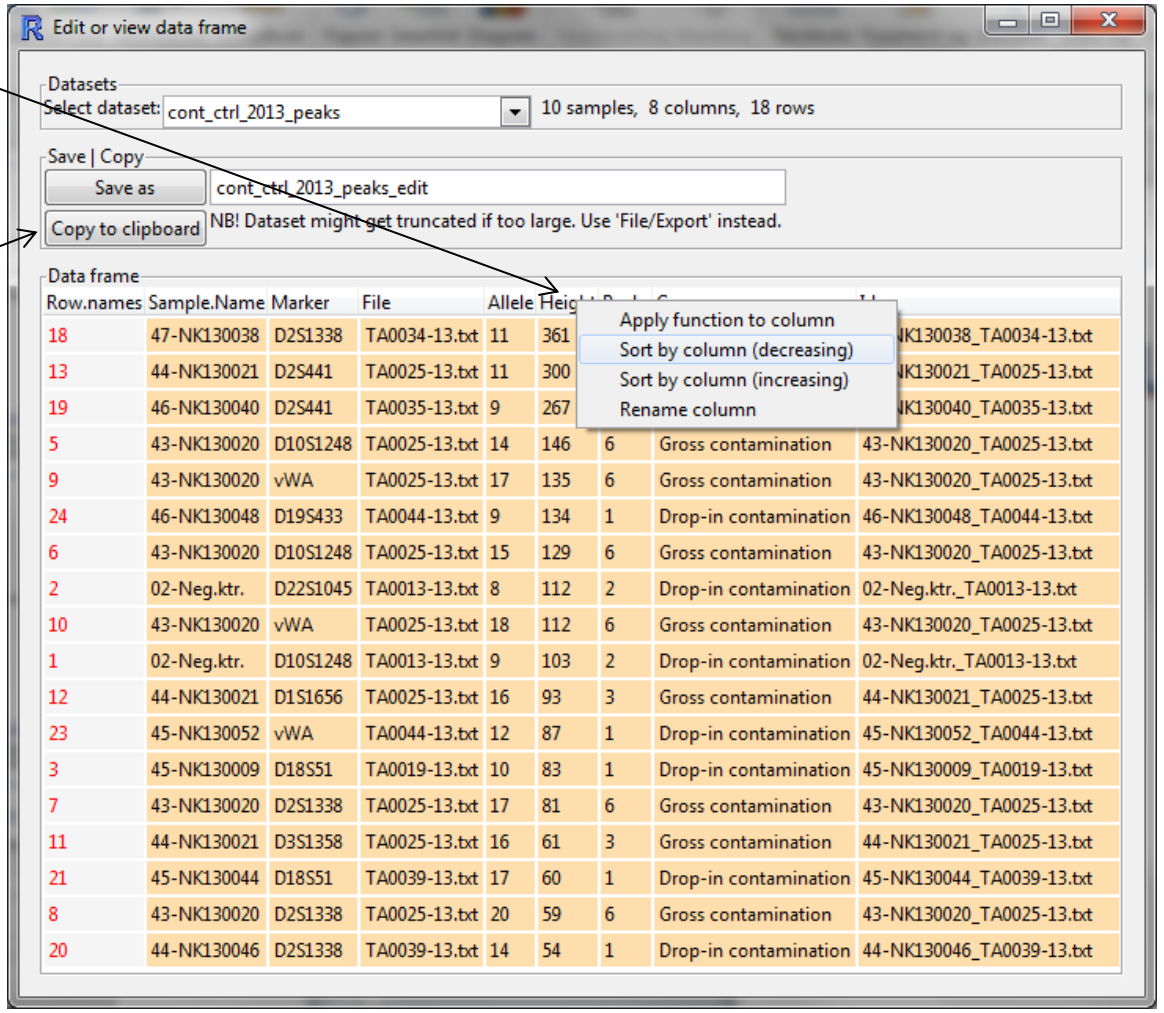


STR validator \\ Tab: Result \\ Group: Number of peaks \\ Button: Calculate

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Calculate probability of drop-in

- Right click on a column name brings up a menu
- Copy to clipboard



Row.names	Sample.Name	Marker	File	Allele	Height			
18	47-NK130038	D2S1338	TA0034-13.txt	11	361			NK130038_TA0034-13.txt
13	44-NK130021	D2S441	TA0025-13.txt	11	300			NK130021_TA0025-13.txt
19	46-NK130040	D2S441	TA0035-13.txt	9	267			NK130040_TA0035-13.txt
5	43-NK130020	D10S1248	TA0025-13.txt	14	146	6	Gross contamination	43-NK130020_TA0025-13.txt
9	43-NK130020	vWA	TA0025-13.txt	17	135	6	Gross contamination	43-NK130020_TA0025-13.txt
24	46-NK130048	D19S433	TA0044-13.txt	9	134	1	Drop-in contamination	46-NK130048_TA0044-13.txt
6	43-NK130020	D10S1248	TA0025-13.txt	15	129	6	Gross contamination	43-NK130020_TA0025-13.txt
2	02-Neg.ktr.	D22S1045	TA0013-13.txt	8	112	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt
10	43-NK130020	vWA	TA0025-13.txt	18	112	6	Gross contamination	43-NK130020_TA0025-13.txt
1	02-Neg.ktr.	D10S1248	TA0013-13.txt	9	103	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt
12	44-NK130021	D1S1656	TA0025-13.txt	16	93	3	Gross contamination	44-NK130021_TA0025-13.txt
23	45-NK130052	vWA	TA0044-13.txt	12	87	1	Drop-in contamination	45-NK130052_TA0044-13.txt
3	45-NK130009	D18S51	TA0019-13.txt	10	83	1	Drop-in contamination	45-NK130009_TA0019-13.txt
7	43-NK130020	D2S1338	TA0025-13.txt	17	81	6	Gross contamination	43-NK130020_TA0025-13.txt
11	44-NK130021	D3S1358	TA0025-13.txt	16	61	3	Gross contamination	44-NK130021_TA0025-13.txt
21	45-NK130044	D18S51	TA0039-13.txt	17	60	1	Drop-in contamination	45-NK130044_TA0039-13.txt
8	43-NK130020	D2S1338	TA0025-13.txt	20	59	6	Gross contamination	43-NK130020_TA0025-13.txt
20	44-NK130046	D2S1338	TA0039-13.txt	14	54	1	Drop-in contamination	44-NK130046_TA0039-13.txt

STR validator \\ Button: Edit

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Calculate probability of drop-in

Paste into a spread-sheet software, add filter and select to show drop-in

	A	B	C	D	E	F	G	H
1	Sample.Name	Marker	File	Allele	Height	Peaks	Group	Id
2	47-NK130038	D2S1338	TA0034-13.txt	11	361	1	Drop-in contamination	47-NK130038_TA0034-13.txt
4	46-NK130040	D2S441	TA0035-13.txt	9	267	1	Drop-in contamination	46-NK130040_TA0035-13.txt
7	46-NK130048	D19S433	TA0044-13.txt	9	134	1	Drop-in contamination	46-NK130048_TA0044-13.txt
9	02-Neg.ktr.	D22S1045	TA0013-13.txt	8	112	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt
11	02-Neg.ktr.	D10S1248	TA0013-13.txt	9	103	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt
13	45-NK130052	vWA	TA0044-13.txt	12	87	1	Drop-in contamination	45-NK130052_TA0044-13.txt
14	45-NK130009	D18S51	TA0019-13.txt	10	83	1	Drop-in contamination	45-NK130009_TA0019-13.txt
17	45-NK130044	D18S51	TA0039-13.txt	17	60	1	Drop-in contamination	45-NK130044_TA0039-13.txt
19	44-NK130046	D2S1338	TA0039-13.txt	14	54	1	Drop-in contamination	44-NK130046_TA0039-13.txt
20								
21								

If x spurious alleles are observed in n controls, Pr(C)=x/n

Pr(C) = 9 / 38 = 0.237

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

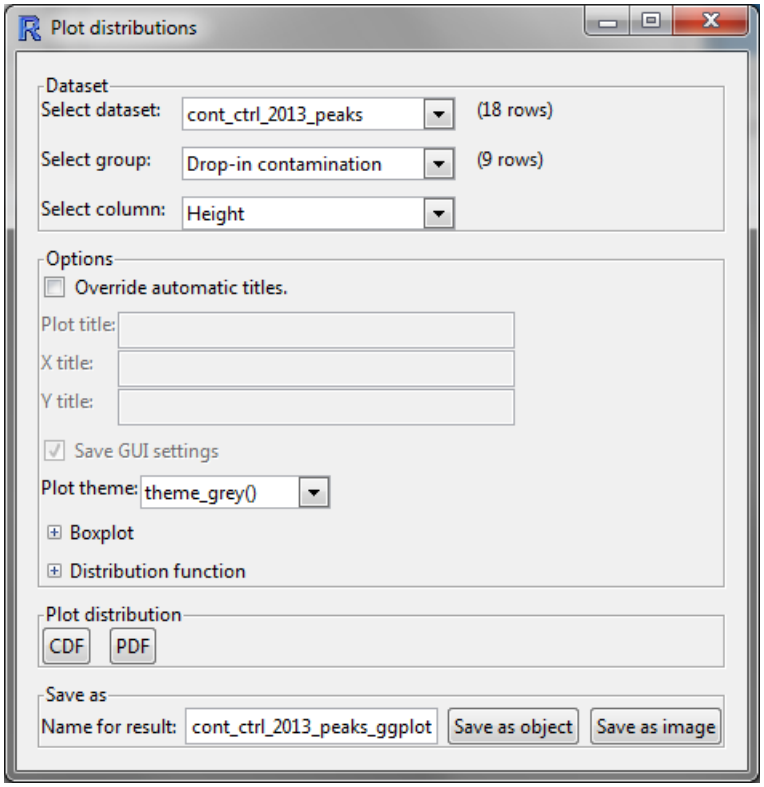
Distribution of height/size for drop-in events

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

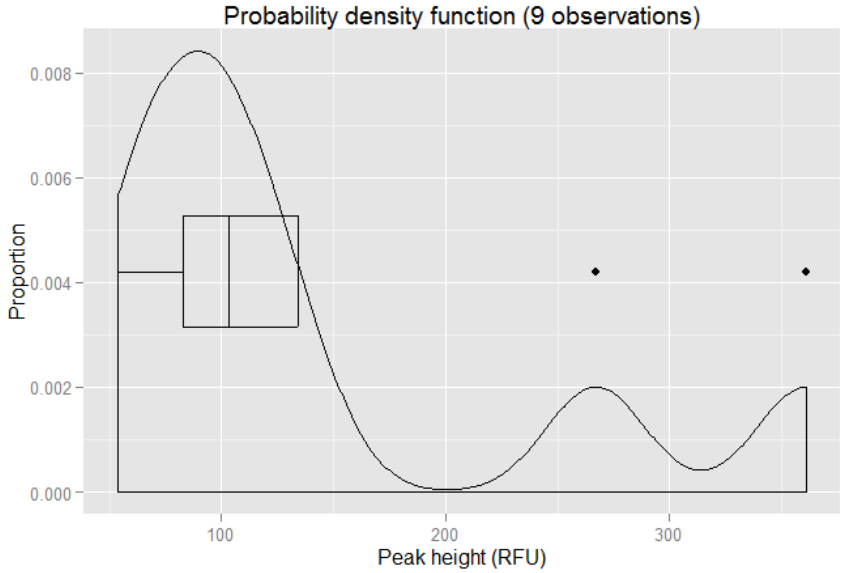
Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Plot distribution of peak heights



STR validator \\ Tab: Result \\ Group: Distributions \\ Button: Plot



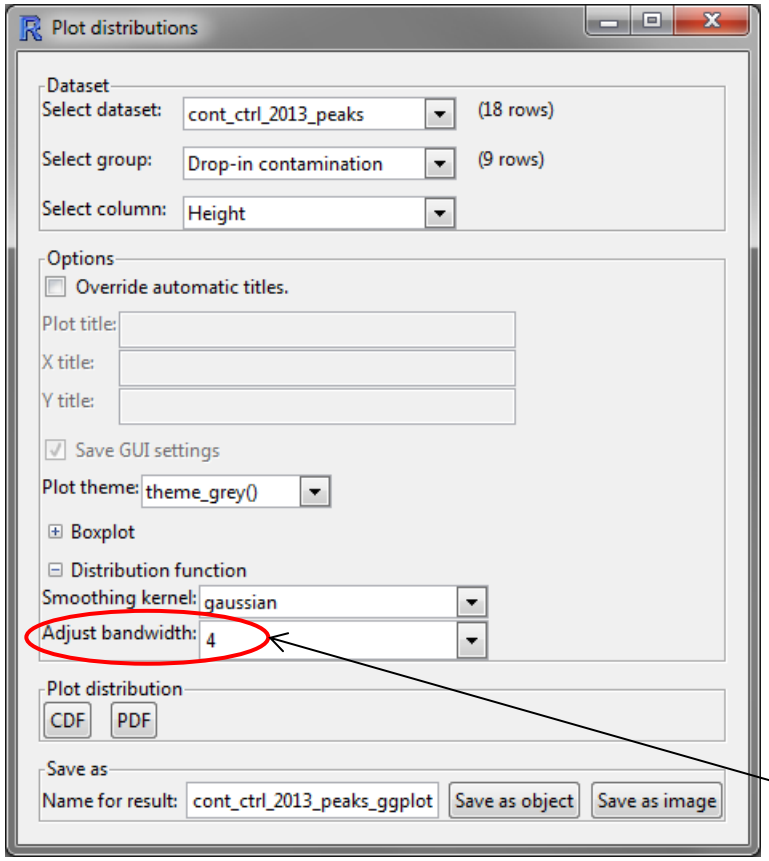
The upper and lower 'hinges' of the boxplot correspond to the 25th and 75th percentiles. The whisker extends from the hinge to the highest/lowest value that is within 1.5 * the inter-quartile range (IQR) of the hinge. Data beyond the end of the whiskers are outliers and plotted as points.

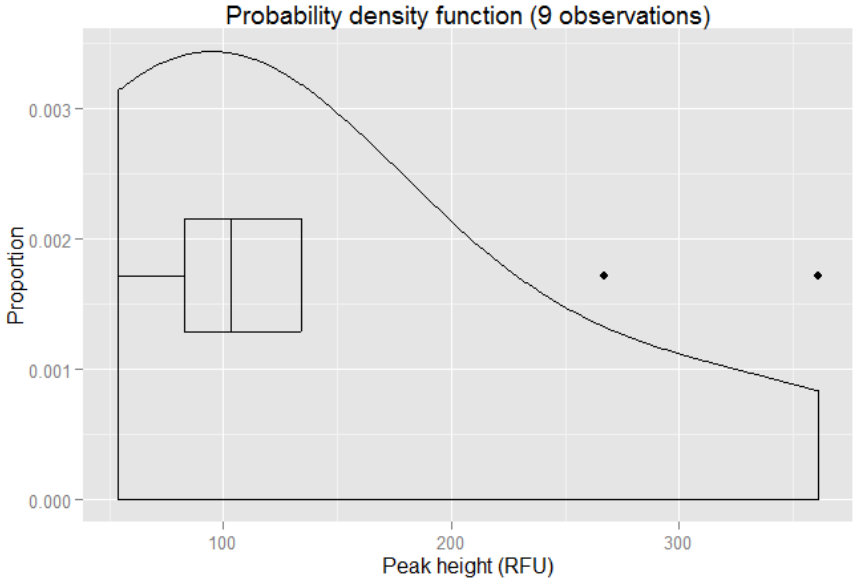
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Plot distribution of peak heights





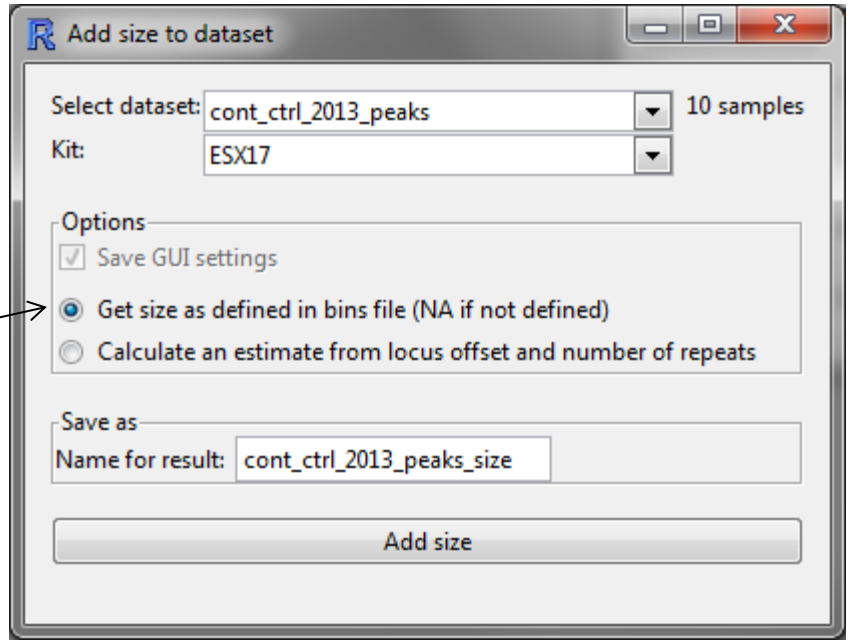
STR validator \\ Tab: Result \\ Group: Distributions \\ Button: Plot

Increase the bandwidth to get a smoother distribution
NB! Can bias the distribution

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Add peak size

- Ideally ‘Size’ is exported from GeneMapper, if not this function can estimate the peak size
- Important to select the correct kit
- Get size from bins file

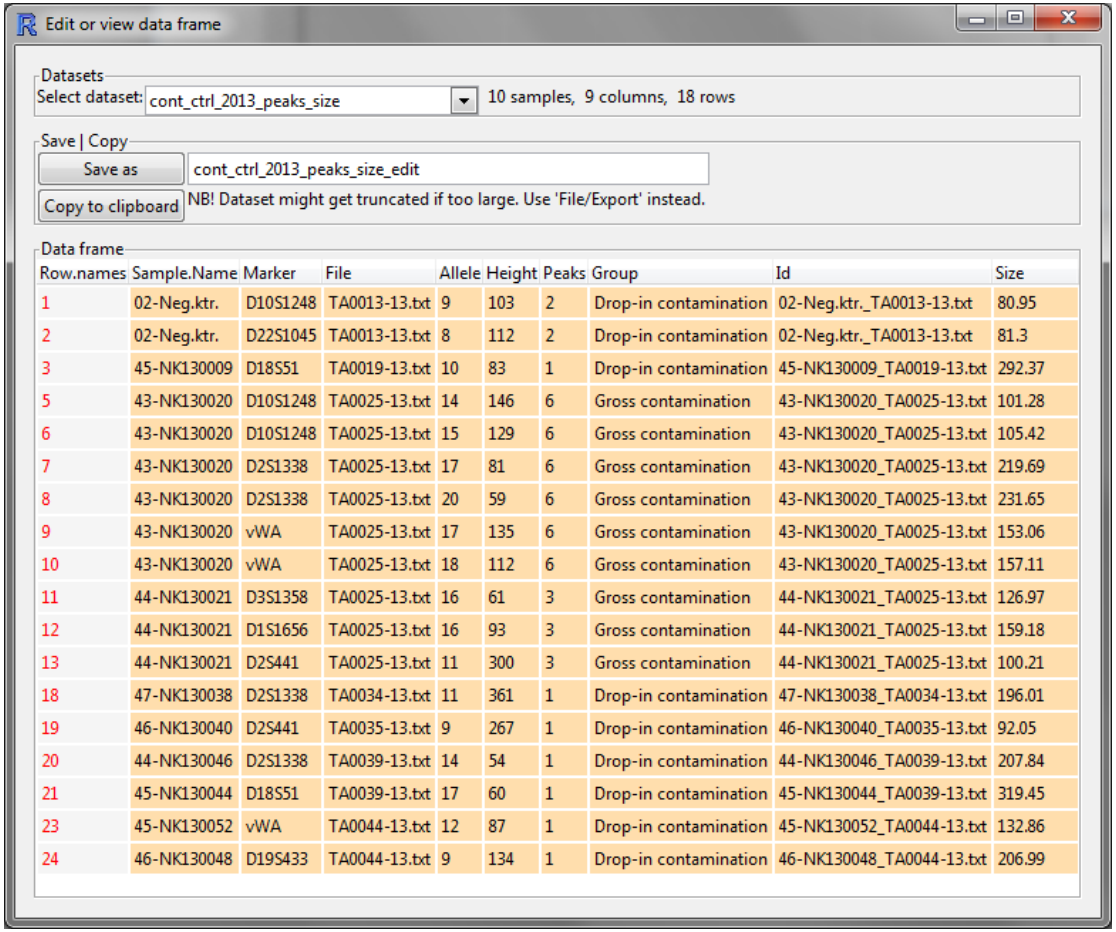


STR validator \\ Tab: Edit \\ Button: Add Size

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

View dataset

‘Size’ column has been added



Row.names	Sample.Name	Marker	File	Allele	Height	Peaks	Group	Id	Size
1	02-Neg.ktr.	D10S1248	TA0013-13.txt	9	103	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt	80.95
2	02-Neg.ktr.	D22S1045	TA0013-13.txt	8	112	2	Drop-in contamination	02-Neg.ktr._TA0013-13.txt	81.3
3	45-NK130009	D18S51	TA0019-13.txt	10	83	1	Drop-in contamination	45-NK130009_TA0019-13.txt	292.37
5	43-NK130020	D10S1248	TA0025-13.txt	14	146	6	Gross contamination	43-NK130020_TA0025-13.txt	101.28
6	43-NK130020	D10S1248	TA0025-13.txt	15	129	6	Gross contamination	43-NK130020_TA0025-13.txt	105.42
7	43-NK130020	D2S1338	TA0025-13.txt	17	81	6	Gross contamination	43-NK130020_TA0025-13.txt	219.69
8	43-NK130020	D2S1338	TA0025-13.txt	20	59	6	Gross contamination	43-NK130020_TA0025-13.txt	231.65
9	43-NK130020	vWA	TA0025-13.txt	17	135	6	Gross contamination	43-NK130020_TA0025-13.txt	153.06
10	43-NK130020	vWA	TA0025-13.txt	18	112	6	Gross contamination	43-NK130020_TA0025-13.txt	157.11
11	44-NK130021	D3S1358	TA0025-13.txt	16	61	3	Gross contamination	44-NK130021_TA0025-13.txt	126.97
12	44-NK130021	D1S1656	TA0025-13.txt	16	93	3	Gross contamination	44-NK130021_TA0025-13.txt	159.18
13	44-NK130021	D2S441	TA0025-13.txt	11	300	3	Gross contamination	44-NK130021_TA0025-13.txt	100.21
18	47-NK130038	D2S1338	TA0034-13.txt	11	361	1	Drop-in contamination	47-NK130038_TA0034-13.txt	196.01
19	46-NK130040	D2S441	TA0035-13.txt	9	267	1	Drop-in contamination	46-NK130040_TA0035-13.txt	92.05
20	44-NK130046	D2S1338	TA0039-13.txt	14	54	1	Drop-in contamination	44-NK130046_TA0039-13.txt	207.84
21	45-NK130044	D18S51	TA0039-13.txt	17	60	1	Drop-in contamination	45-NK130044_TA0039-13.txt	319.45
23	45-NK130052	vWA	TA0044-13.txt	12	87	1	Drop-in contamination	45-NK130052_TA0044-13.txt	132.86
24	46-NK130048	D19S433	TA0044-13.txt	9	134	1	Drop-in contamination	46-NK130048_TA0044-13.txt	206.99

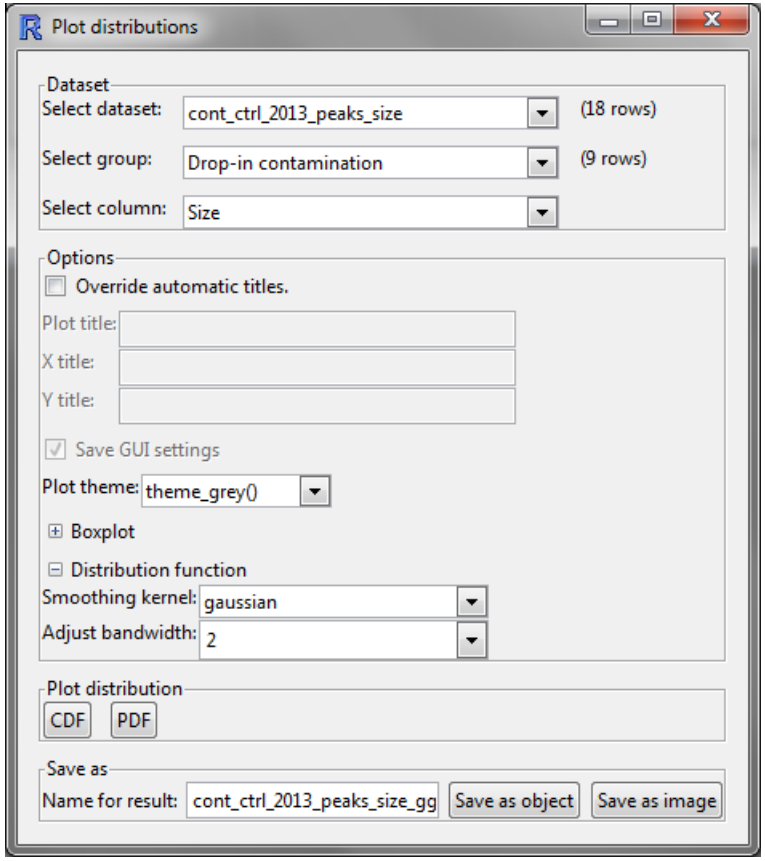
STR validator \\ Button: Edit

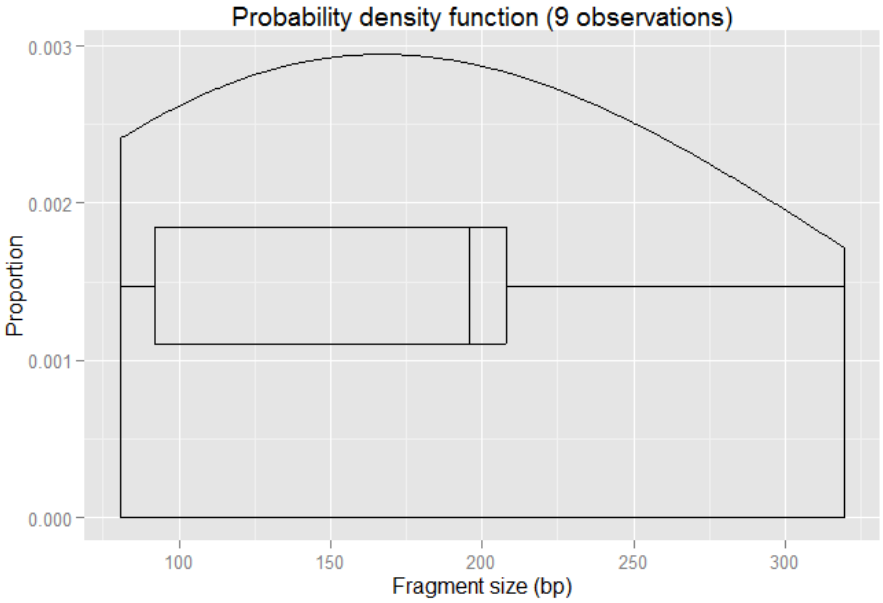
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Plot distribution of peak size





EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Explore the probability of drop-out, Pr(D)

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

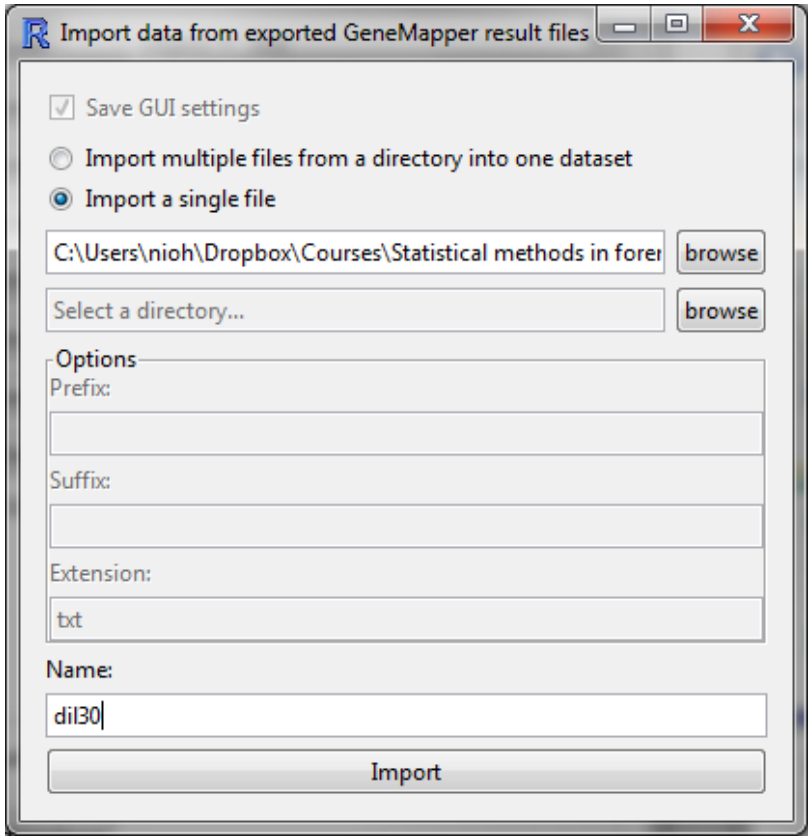
Experimental setup

- 9 single source crime scene samples were selected (labelled S01 – S09)
- Each sample were diluted to give optimal concentration in the PCR reaction
- Two-fold serial dilutions were performed in 9 steps (labelled D1 – D9)

SEVENTH FRAMEWORK PROGRAMME

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Import result

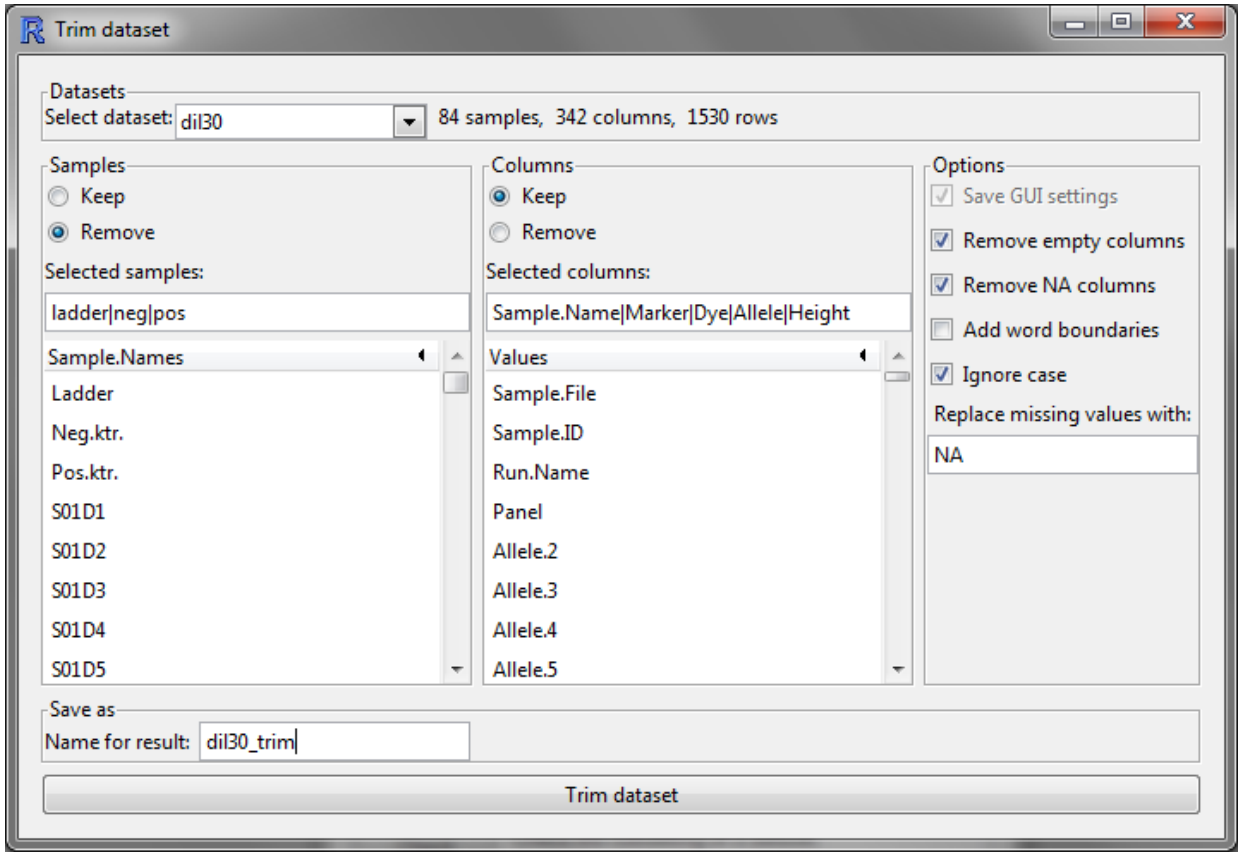


STR validator \\ Tab: Workspace \\ Button: Import

Result file:
DIL30FAST Genotypes Table.txt

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

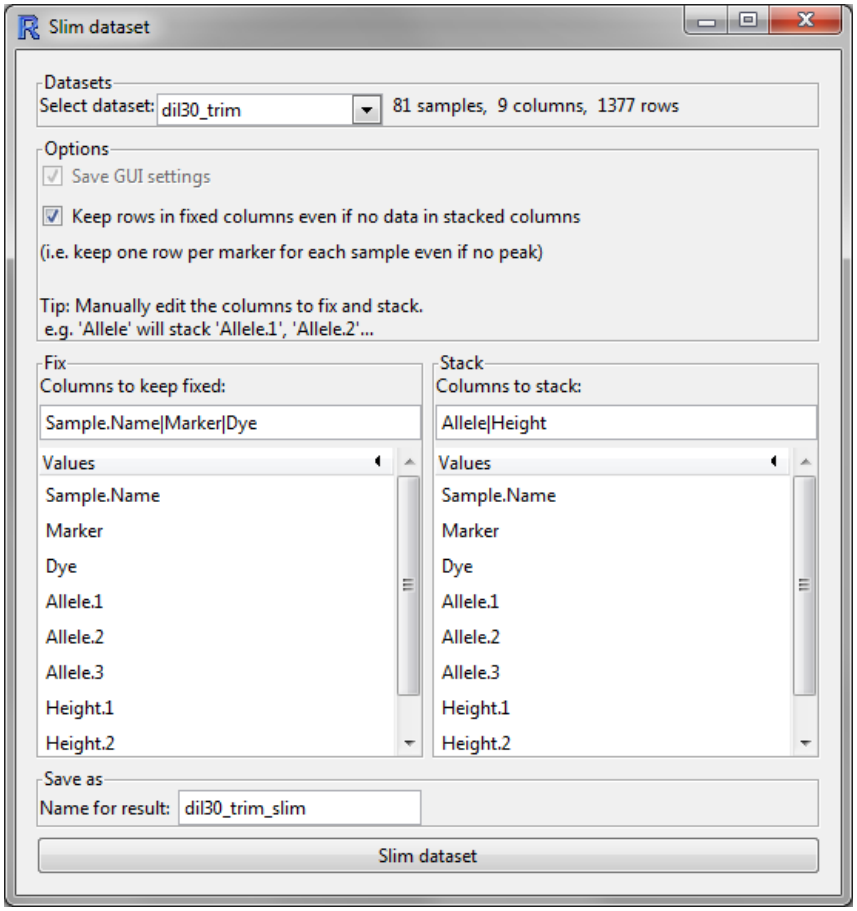
Trim dataset



STR validator \\ Tab: Edit \\ Button: Trim

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Slim dataset



STR validator \\ Tab: Edit \\ Button: Slim

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Norwegian Institute of Public Health

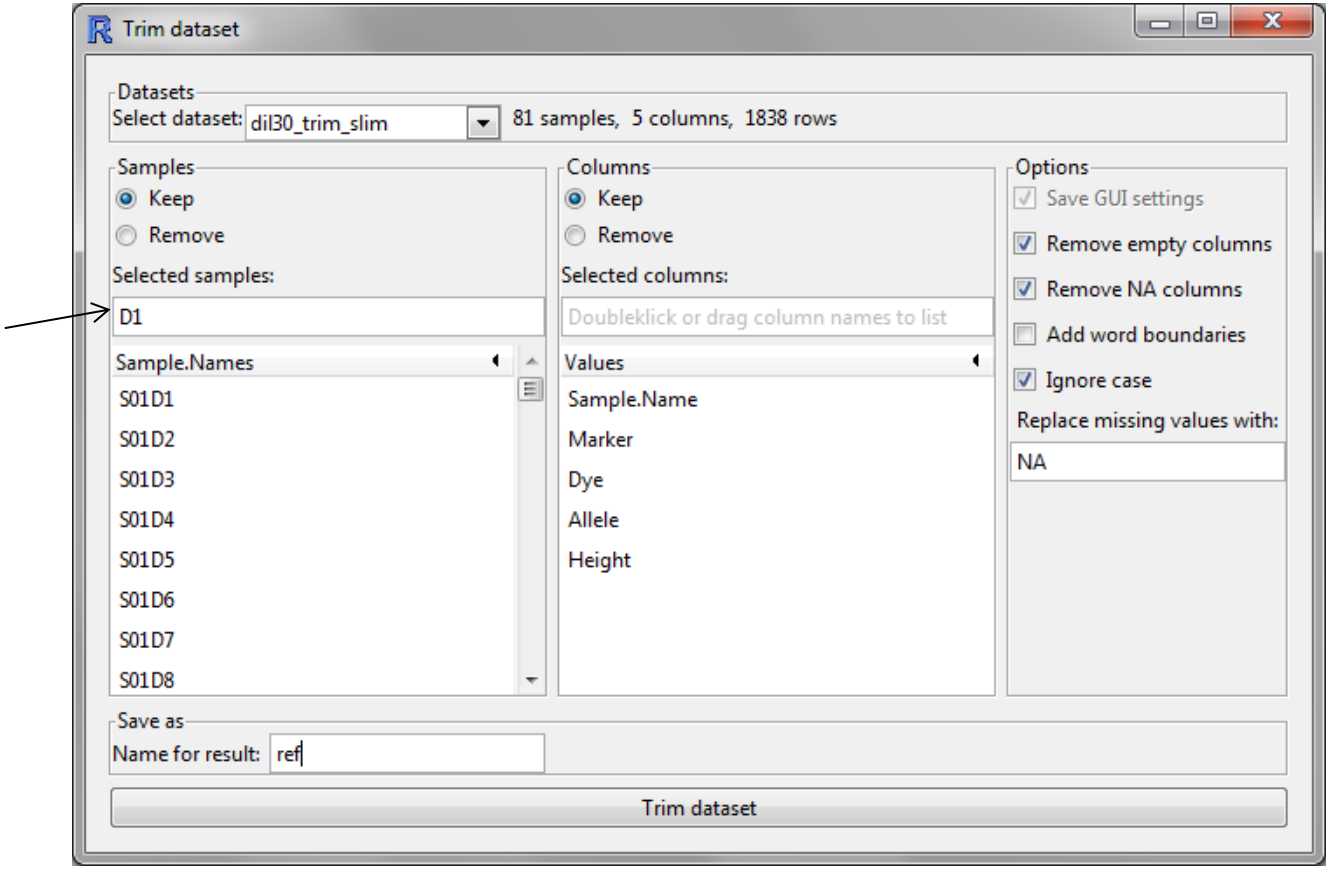
EUROFORGEN Network of Excellence

Create a reference dataset with known profiles

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Create a reference dataset

Keep only samples with optimal amount (i.e. the first dilution from each source sample)



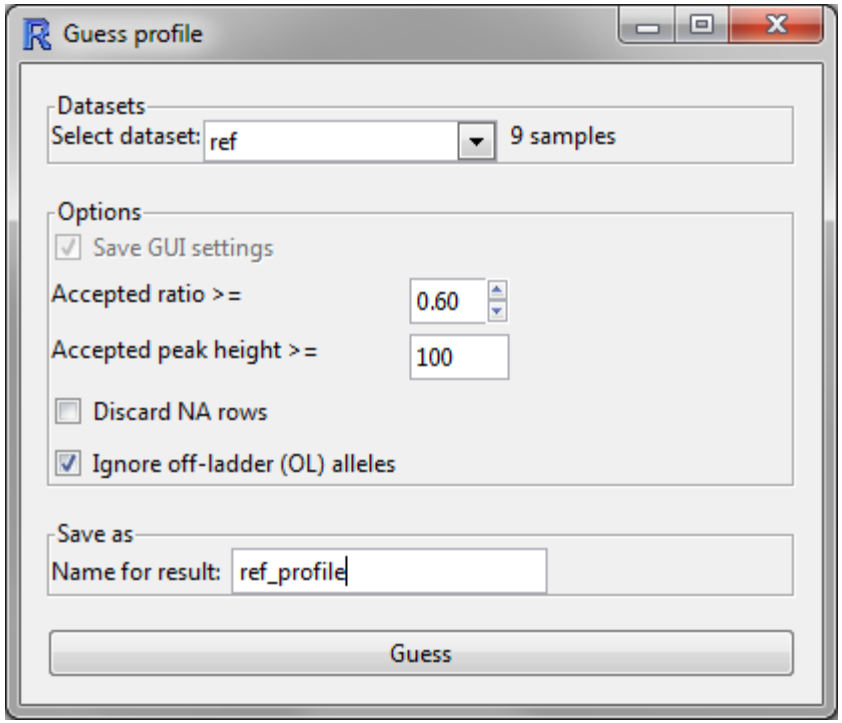
STR validator \\ Tab: Edit \\ Button: Trim

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Guess the correct profile



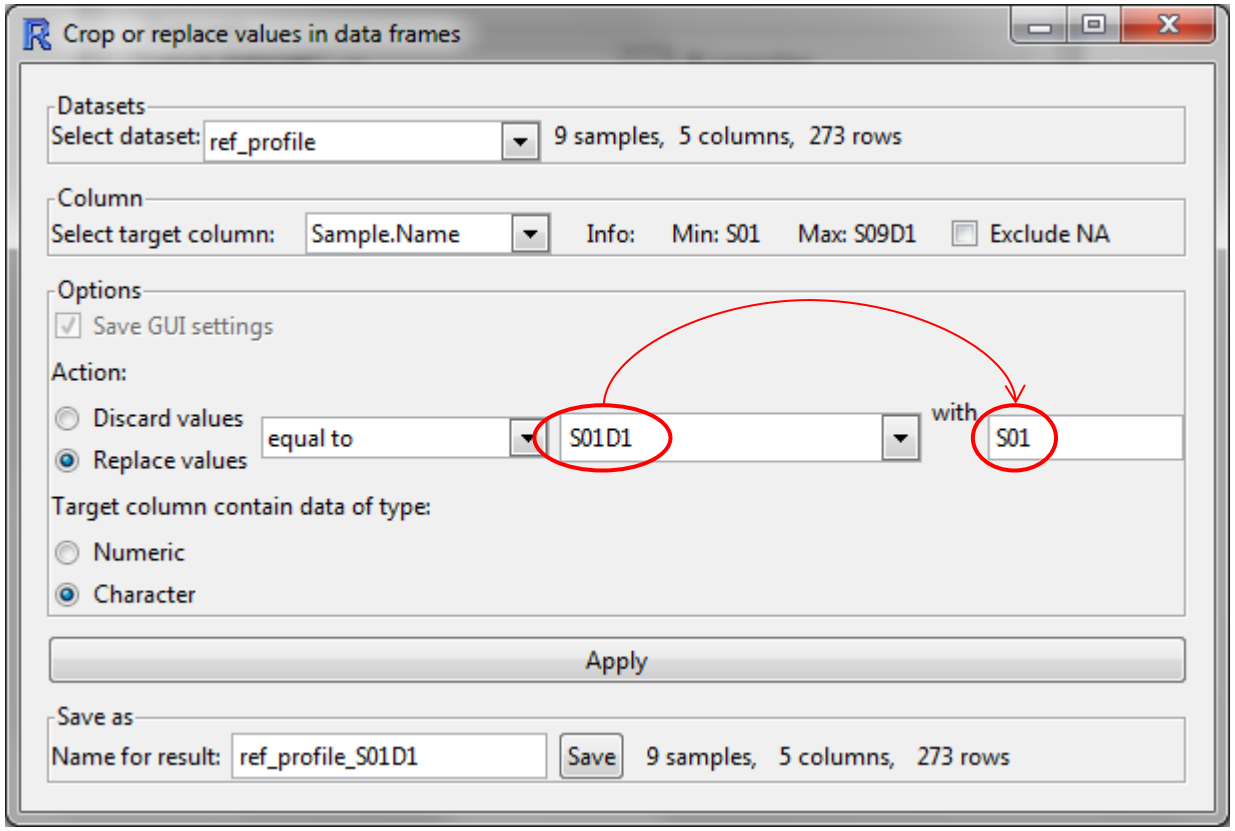
STR validator \\ Tab: Edit \\ Button: Guess

Alternatively/ideally profiles can be manually edited in GeneMapper to contain only the correct profiles and then imported in STR validator

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Replace sample names

- Replace sample names with the source sample name
- Click apply
- Repeat for all samples
- Finally change the name for result to ‘ref’ and click ‘Save’.

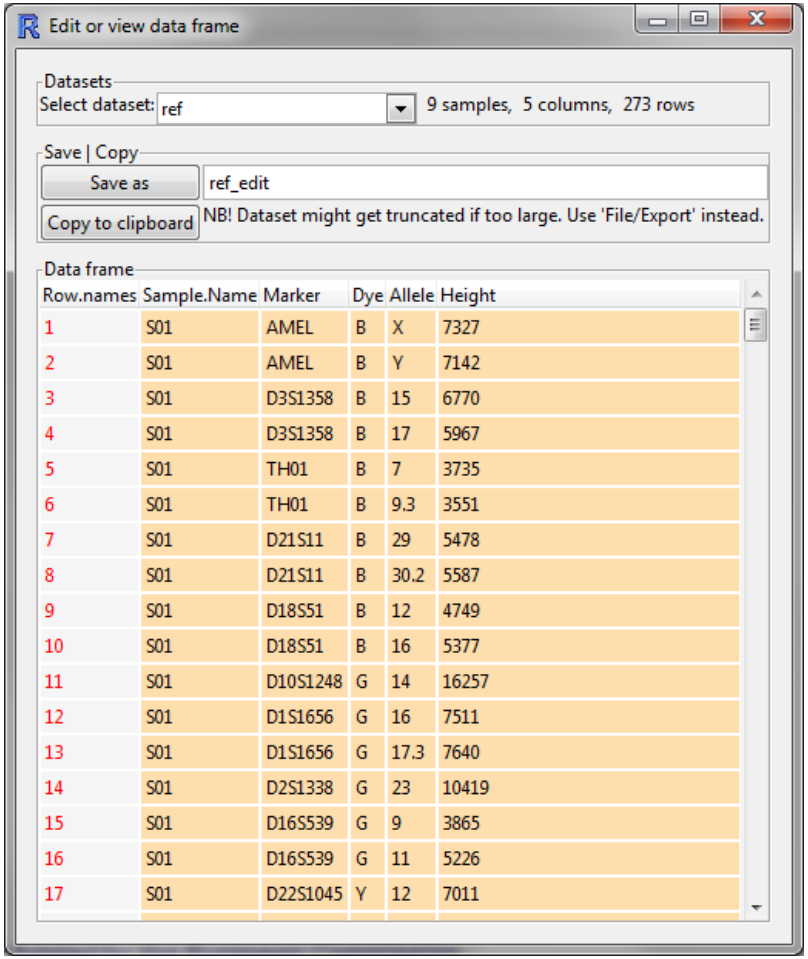


STR validator \\ Tab: Edit \\ Button: Crop

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

View reference dataset

- The reference dataset now contain only the source sample names and the correct(?) profile
- The 'Height' column can be removed, but does no harm
- Confirm the profiles manually against the EPGs or known profiles



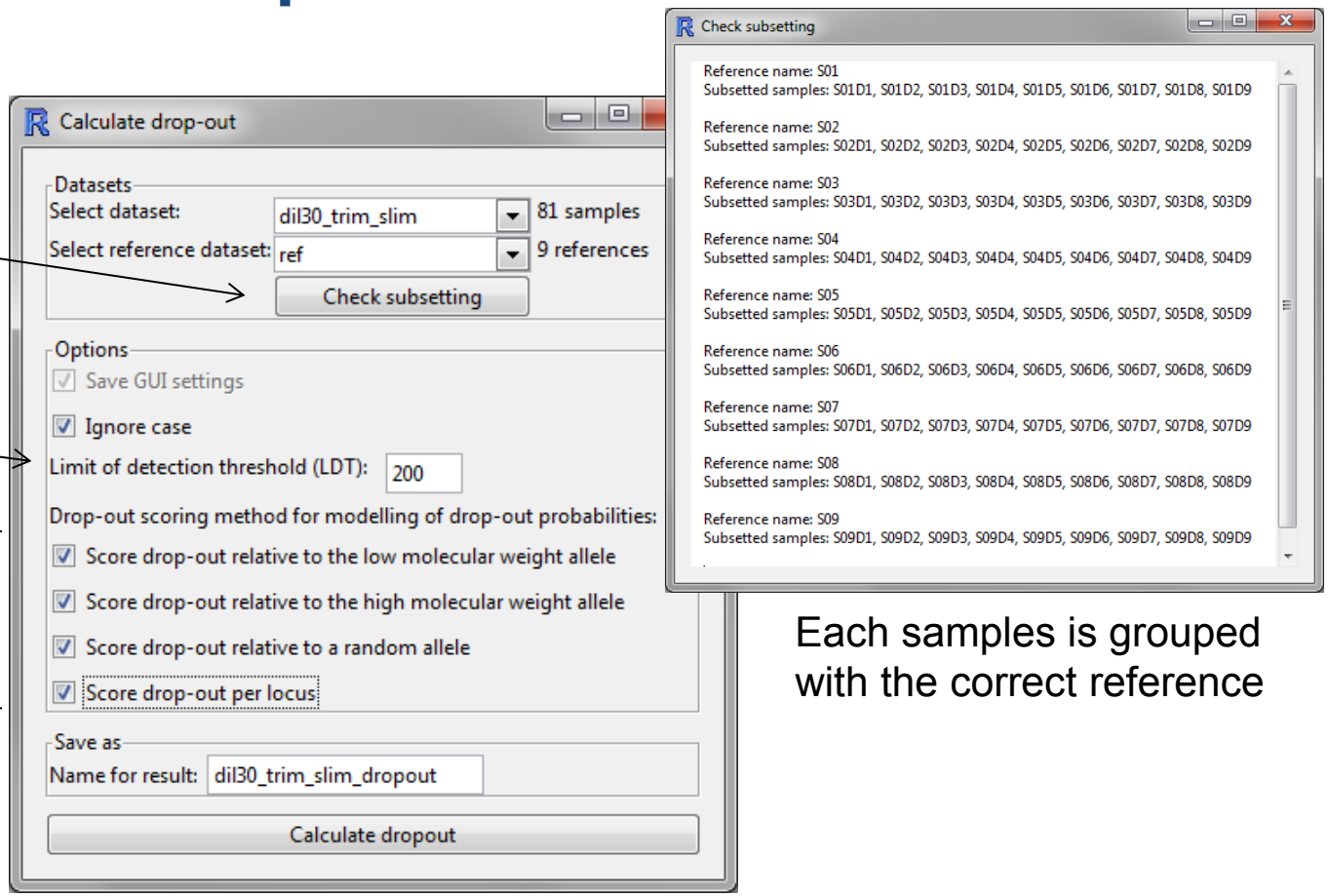
Row.names	Sample.Name	Marker	Dye	Allele	Height
1	S01	AMEL	B	X	7327
2	S01	AMEL	B	Y	7142
3	S01	D3S1358	B	15	6770
4	S01	D3S1358	B	17	5967
5	S01	TH01	B	7	3735
6	S01	TH01	B	9.3	3551
7	S01	D21S11	B	29	5478
8	S01	D21S11	B	30.2	5587
9	S01	D18S51	B	12	4749
10	S01	D18S51	B	16	5377
11	S01	D10S1248	G	14	16257
12	S01	D1S1656	G	16	7511
13	S01	D1S1656	G	17.3	7640
14	S01	D2S1338	G	23	10419
15	S01	D16S539	G	9	3865
16	S01	D16S539	G	11	5226
17	S01	D22S1045	Y	12	7011

STR validator \\ Button: Edit

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Score drop-out events

- Check subsetting
- Drop-out is scored against the LDT
- Four ways to score drop-out



Each samples is grouped with the correct reference

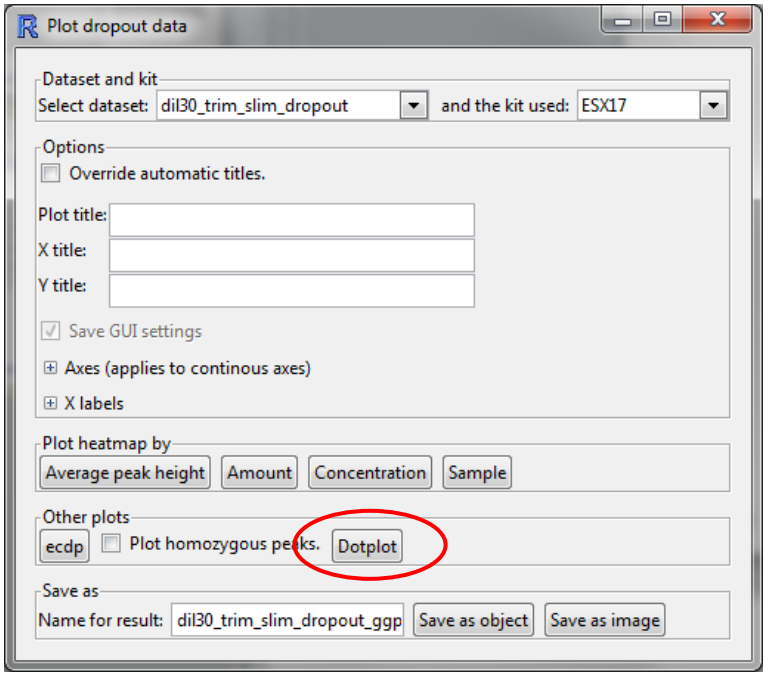
STR validator \\ Tab: Dropout \\ Button: Calculate

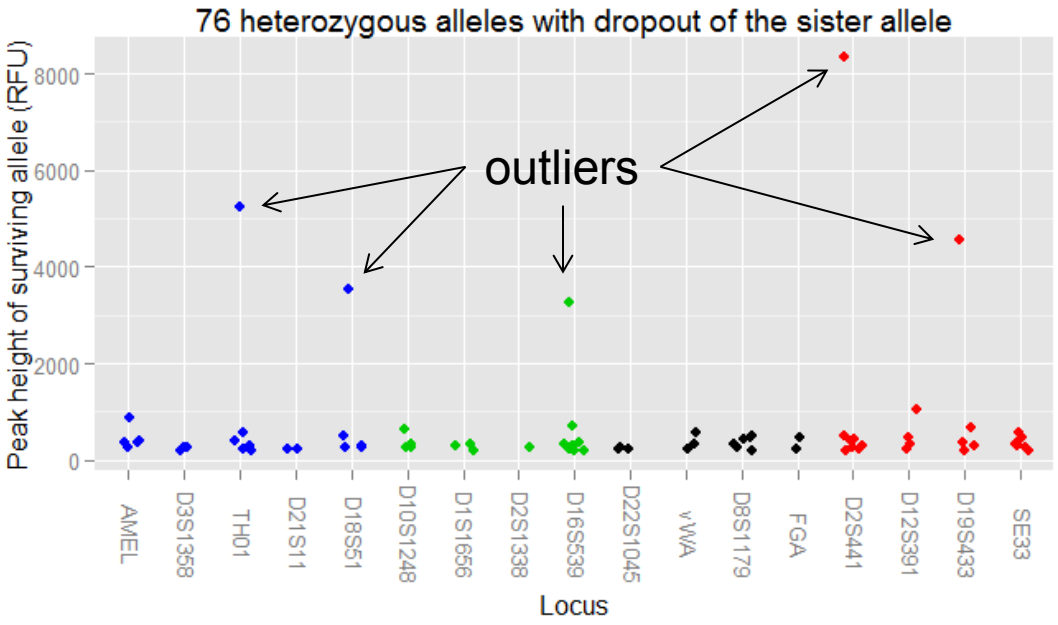
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Quality check the result



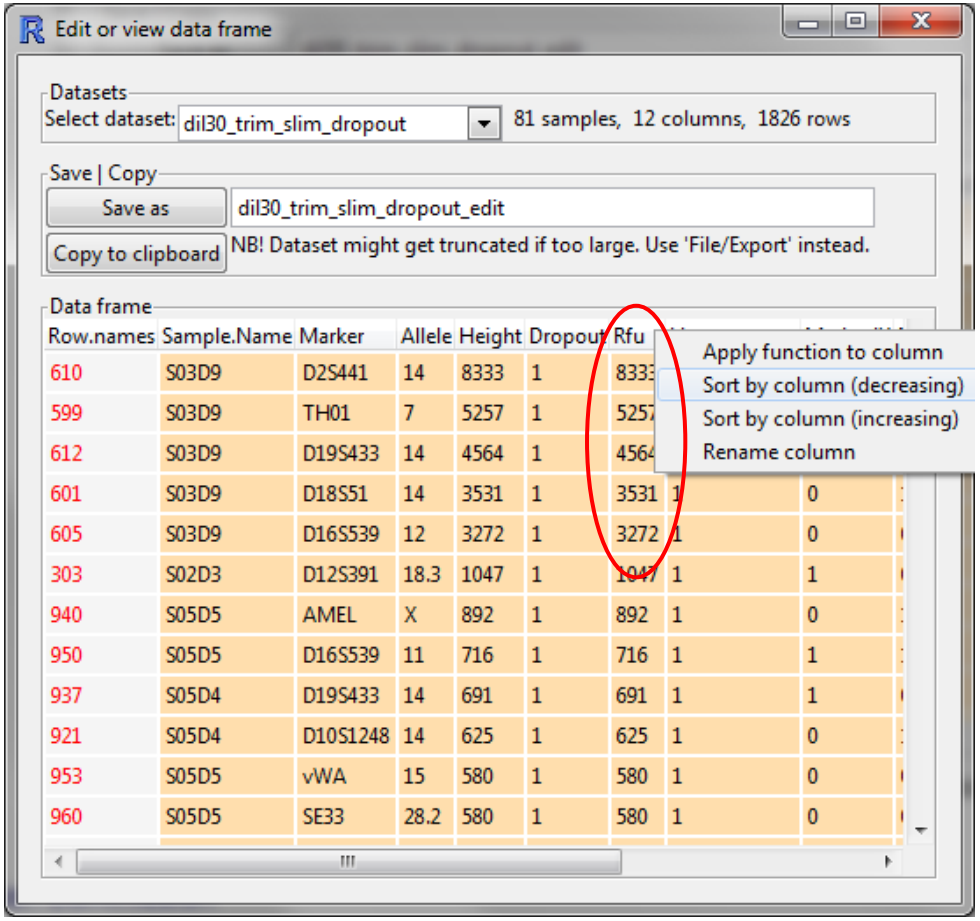


STR validator \\ Tab: Dropout \\ Button: Plot

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Quality check the result

- Sort decreasing in the 'Rfu' column
- All outliers found in sample S03D9



Row.names	Sample.Name	Marker	Allele	Height	Dropout	Rfu		
610	S03D9	D2S441	14	8333	1	8333		
599	S03D9	TH01	7	5257	1	5257		
612	S03D9	D19S433	14	4564	1	4564		
601	S03D9	D18S51	14	3531	1	3531	1	0
605	S03D9	D16S539	12	3272	1	3272	1	0
303	S02D3	D12S391	18.3	1047	1	1047	1	1
940	S05D5	AMEL	X	892	1	892	1	0
950	S05D5	D16S539	11	716	1	716	1	1
937	S05D4	D19S433	14	691	1	691	1	1
921	S05D4	D10S1248	14	625	1	625	1	0
953	S05D5	vWA	15	580	1	580	1	0
960	S05D5	SE33	28.2	580	1	580	1	0

STR validator \\ Button: Edit

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Possible errors

Errors in subsetting

- Sample paired with wrong reference or several references
- → Check subsetting

Errors in reference dataset

- Stutter called as real allele
- Allele not called because too imbalanced
- Off-ladder (OL) allele in reference dataset
- Contamination
- → Then several dilutions from one sample would likely be affected

Errors in sample dataset

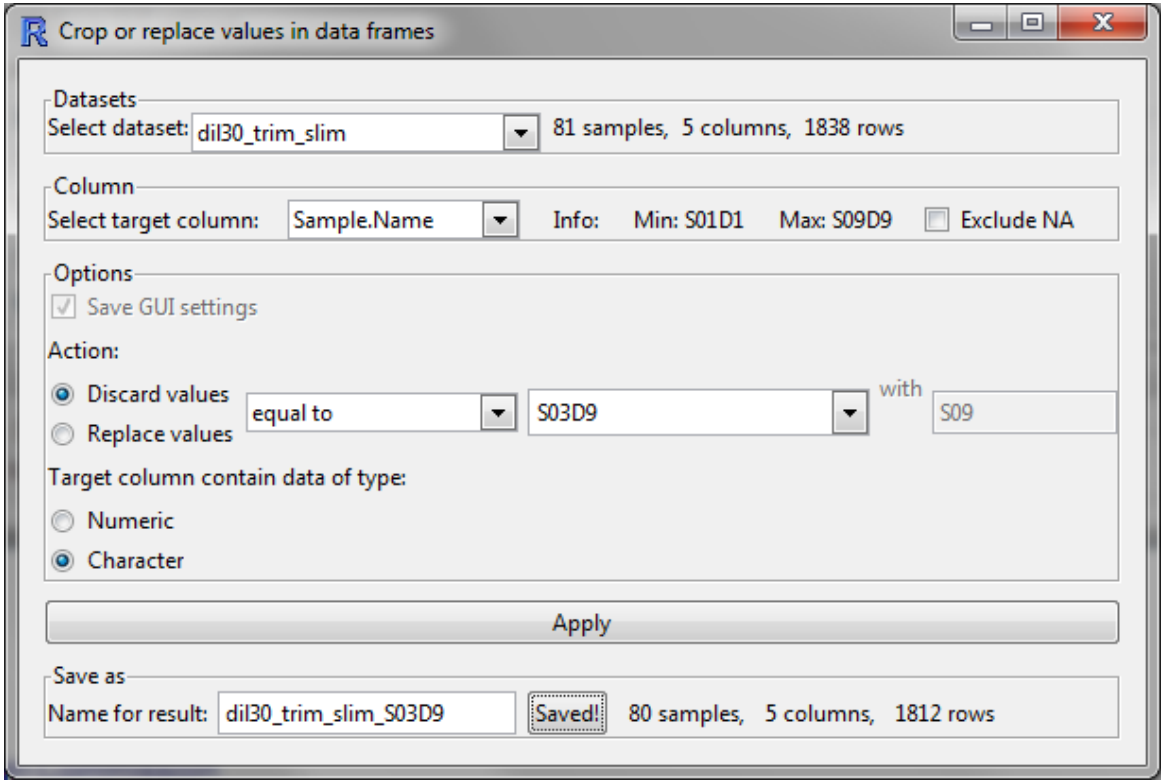
- OL alleles in dataset
- Sample mix-up
- Contamination

SEVENTH FRAMEWORK PROGRAMME

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Fix errors

- The profile in S03D9 did not match any reference profile
- Compared against D1-D8 we expect a blank profile
- The profile was a gross contamination (matched the operator)
- Fixed by removing the sample (use ‘Crop’ or ‘Trim’ function)

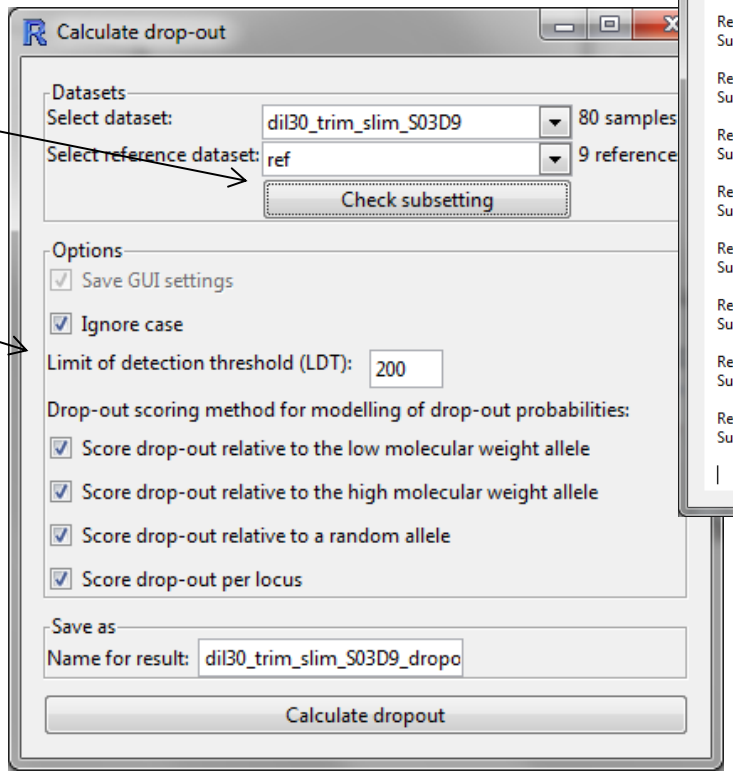


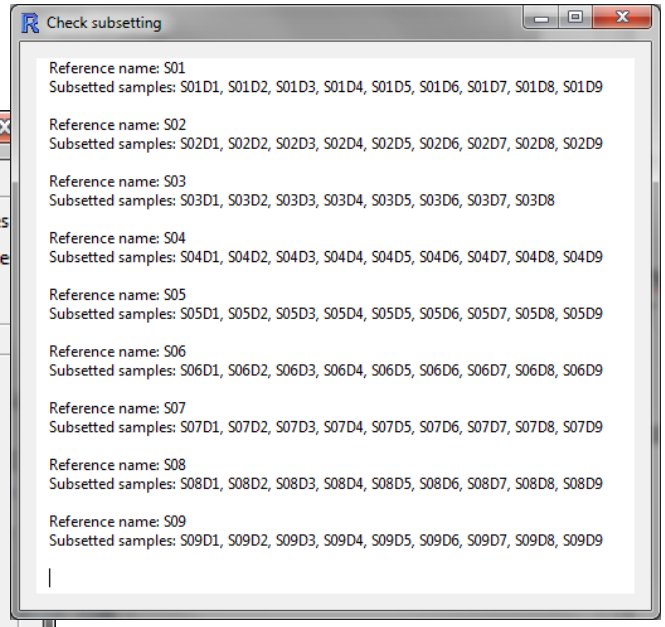
STR validator \\ Tab: Edit \\ Button: Crop

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Score drop-out events

- Check subsetting
- Drop-out is scored against the LDT
- Four ways to score drop-out





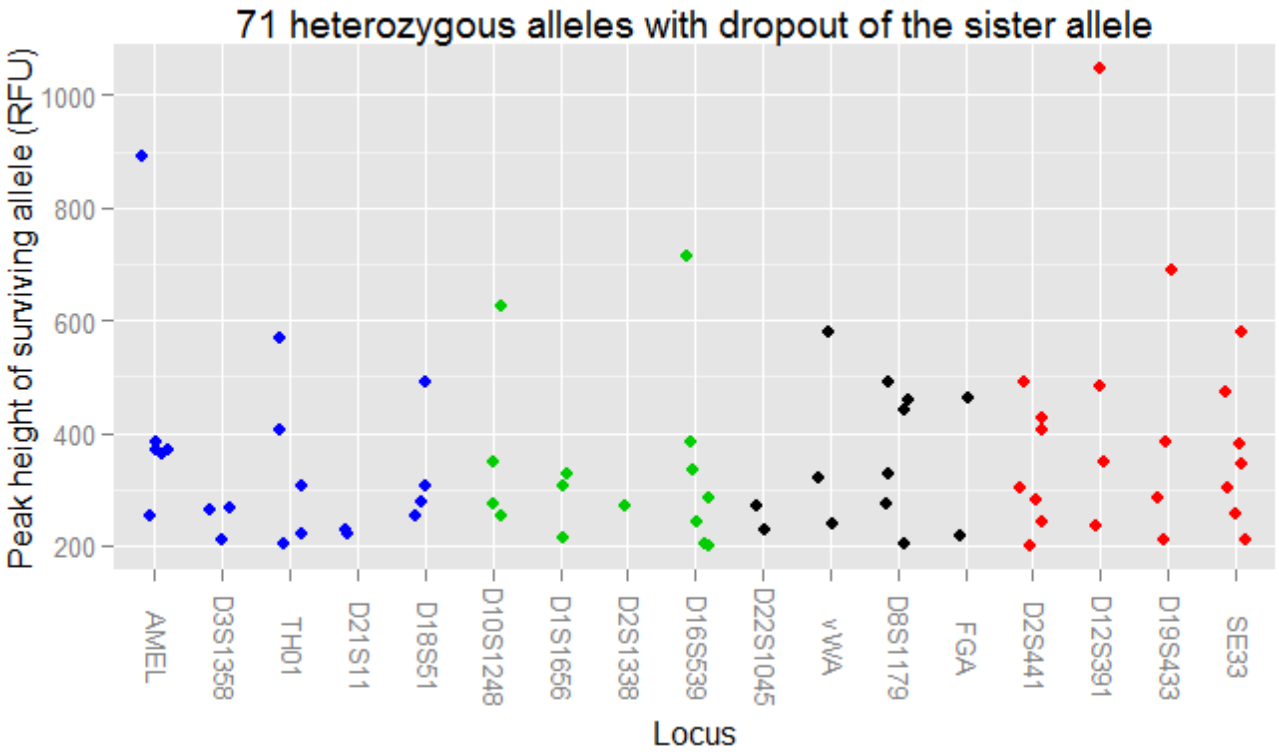
Each samples is grouped with the correct reference and S03D9 is missing

STR validator \\ Tab: Dropout \\ Button: Calculate

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Quality check the result

All extremes are gone, we should continue to verify some of the highest values against the EPGs

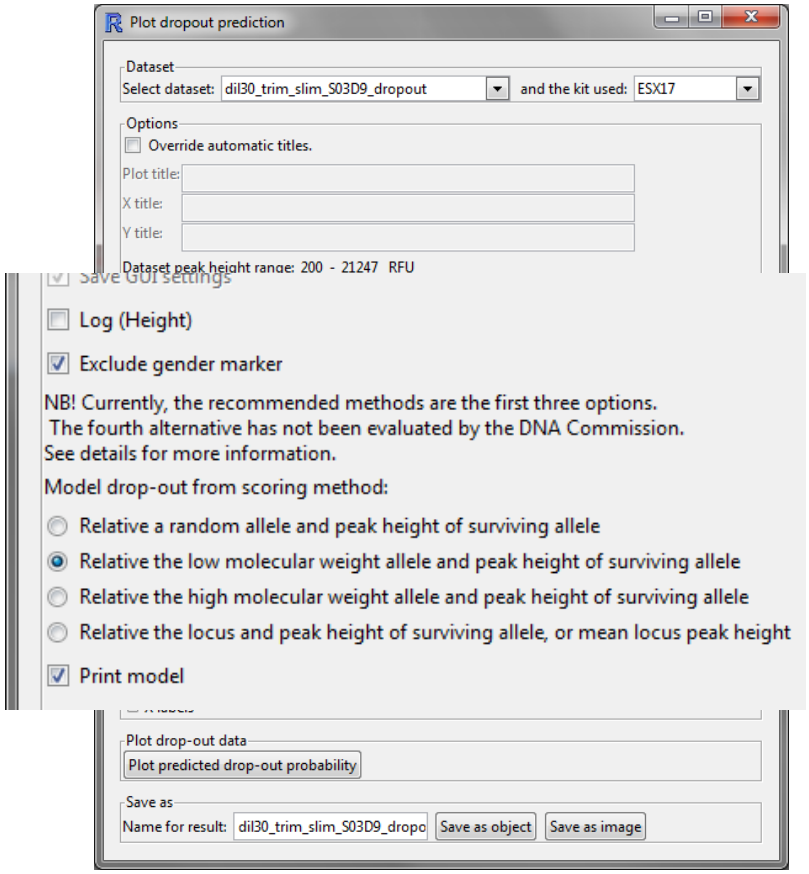


EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

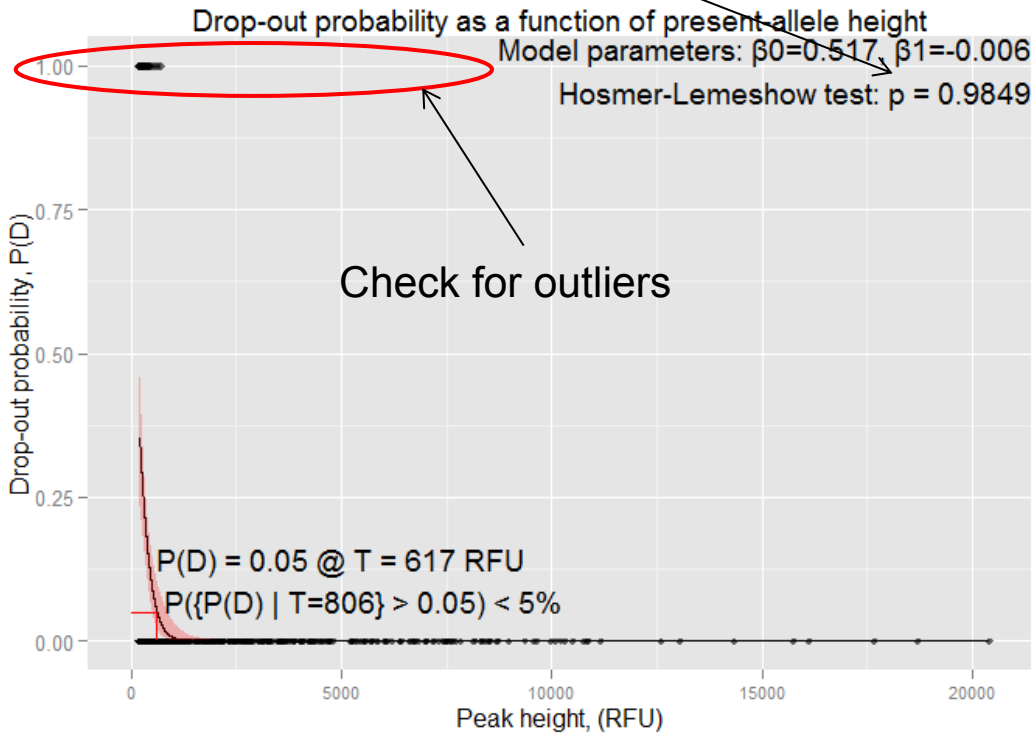
Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Model drop-out



H-L test $p < 0.05$ indicate poor fit between modelled and observed data



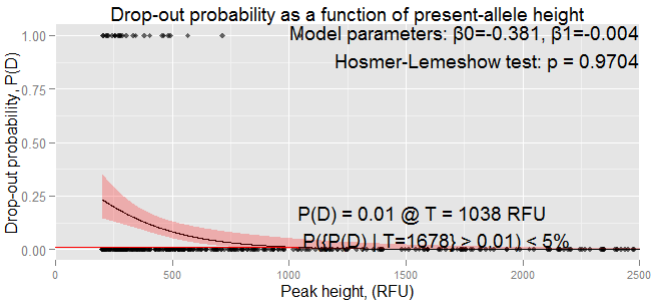
STR validator \\ Tab: Dropout \\ Button: Model

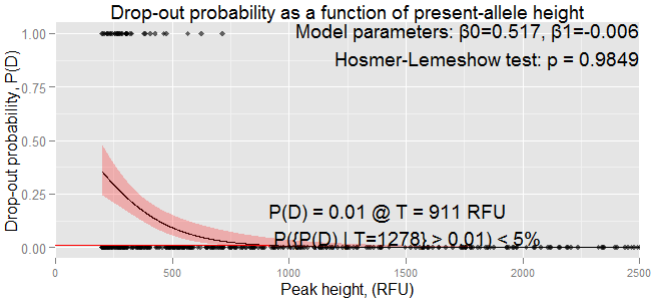
EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Plot dropout prediction

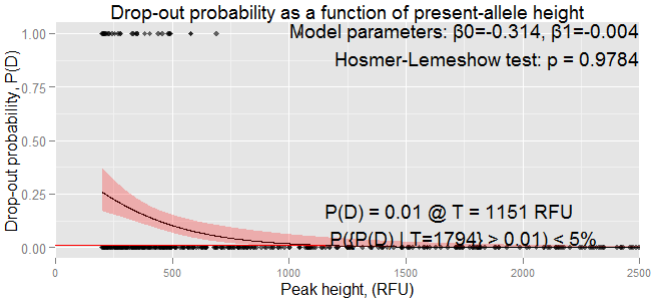
Dataset
Select dataset: <Select dataset> and the kit used: ESI17

Options
Override automatic titles.
Plot title:
X title:
Y title:
Dataset peak height range: - RFU
Save GUI settings
Log (Height)
Exclude gender marker
NB! Currently, the recommended methods are the first three options.
The fourth alternative has not been evaluated by the DNA Commission.
See details 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
Print model
Drop-out prediction and threshold
Mark threshold @ P(D): 0.010 Calculate T for 1%
Line type solid Line colour red
Print threshold value
Prediction interval: 0.950
Print conservative T value
Draw prediction interval: Alpha 0.25 Fill colour red
Data points
Axes
NB! Must provide both min and max value.
Limit Y axis (min-max)
Limit X axis (min-max)
0 2500 Zoom in 0-2500 RFU
X labels
Plot drop-out data
Plot predicted drop-out probability
Save as
Name for result: Save as object Save as image

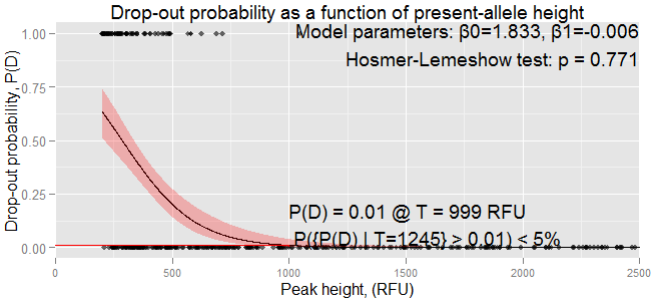




All models fit well with observations ($p > 0.05$)

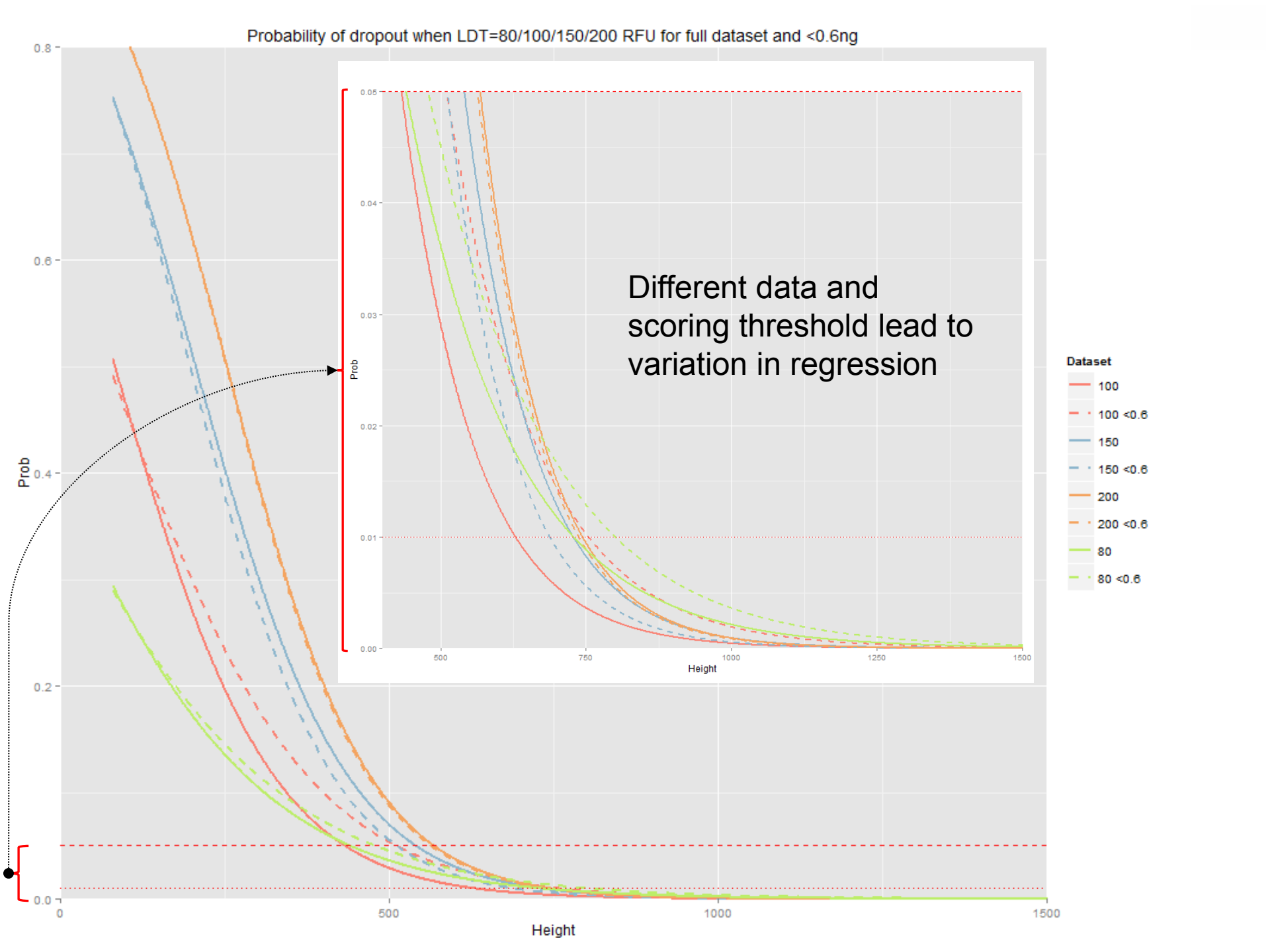


Threshold for 1% risk of drop-out 1000–1500 RFU



NB! The estimated 'T' will vary between models, collected data (LOD), scoring threshold (LDT), methods, and instruments

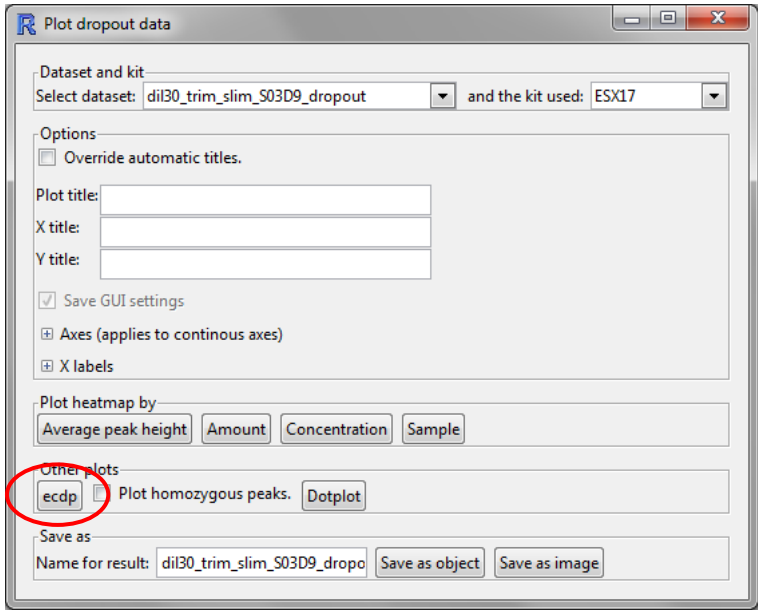
EUROFORGEN Network of Excellence


Probability of dropout when LDT=80/100/150/200 RFU for full dataset and <0.6ng
Prob
Height
0.0
0.2
0.4
0.6
0.8
0
500
1000
1500
Different data and scoring threshold lead to variation in regression
0.00
0.01
0.02
0.03
0.04
0.05
500
750
1000
1250
1500
Dataset
100
100 <0.6
150
150 <0.6
200
200 <0.6
80
80 <0.6

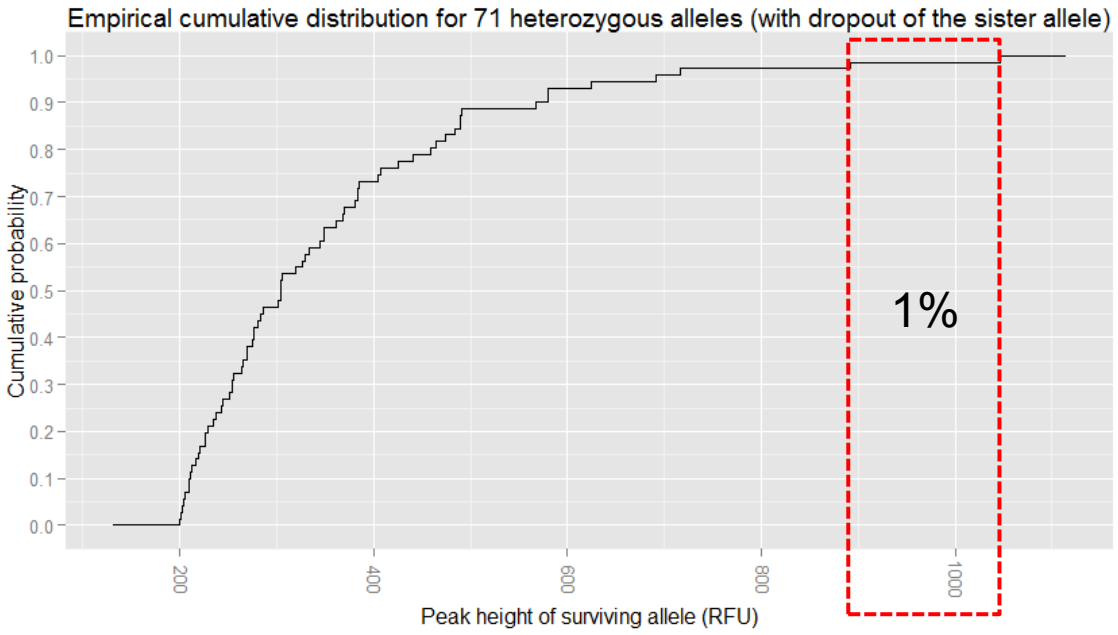
Norwegian Institute of Public Health

EUROFORGEN Network of Excellence

Compare with observations



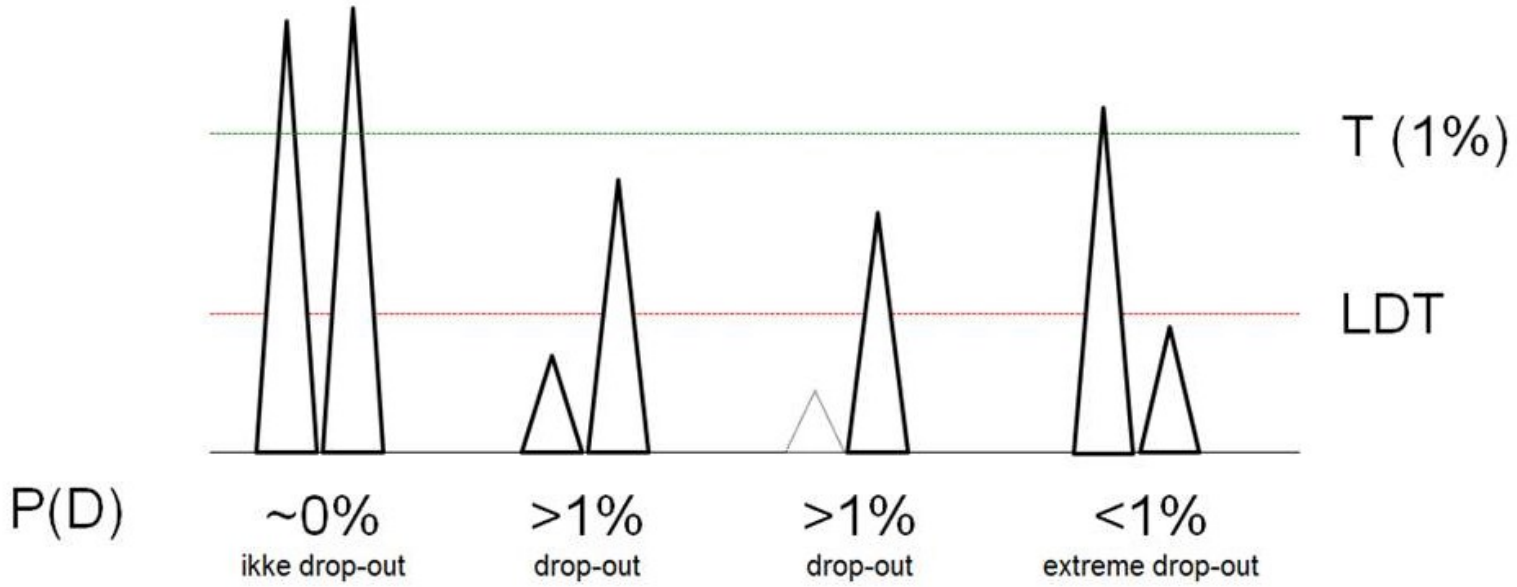
STR validator \\ Tab: Dropout \\ Button: Plot



Observations support that a 1% risk is in the 1000 RFU range
(99% of observed single peaks in heterozygote loci have a lower peak height)

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Conclusion



The risk that a single peak with a height of 'T' RFU is a heterozygote with drop-out of the partner allele (< LDT RFU) is approximately 1% in a single source undegraded sample i.e. it is 'safe' to call it a homozygote (if the partner allele is not distinguishable above the noise and below the LDT)

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

Reference

STR validator — an open source platform for validation and process control

SUBMITTED

Oskar Hansson[a,*], Peter Gill[a,b], Thore Egeland[a,c]

[a]*Norwegian Institute of Public Health, Department of Forensic Biology, Norway*

[b]*University of Oslo, Oslo, Norway*

[c]*Norwegian University of Life Sciences, Oslo, Norway*

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. Prinz[i], H. Schneider[j], P.M. Schneider[k], B.S. Weir[l]

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme