Strategies to work with HLA data in human populations: a tutorial to format HLA data and make basic analyses

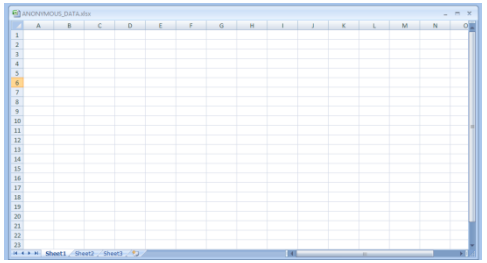
Prof. Alicia Sanchez-Mazas & Dr José Manuel Nunes

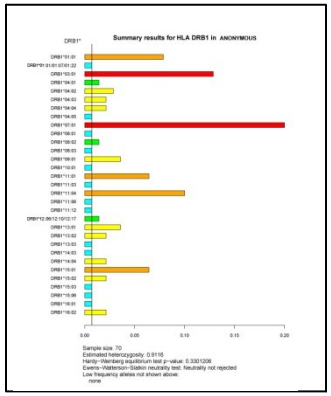
Dpt Genetics & Evolution – Anthropology Unit

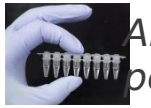
University of Geneva

Anonymous population sample

How to format my data and compute basic statistics?





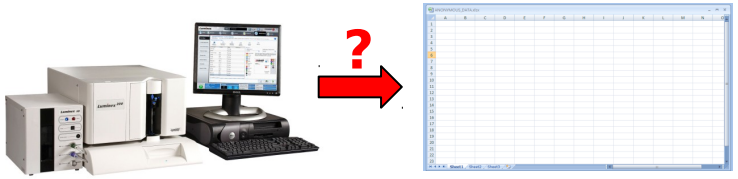


Anonymous population sample

The data

Most of the times, your original table is an Excel file with genotypes, let's work with it

http://geneva.../Anonymous .xls



Data formating

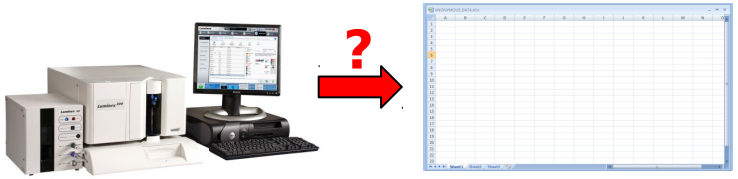
This is an example of original table of HLA genotypes for one locus:It contains a sample number, the population name, and pairs of possible genotypes

Sample Number

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_16	Anonymous	not done					
18	ANON_17	Anonymous	*07010101/07010102	*130201				
19	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
20	ANON_19	Anonymous	*040301	*160201				
21	ANON_20	Anonymous	*010101	*110101/110108				
22	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
23	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
24	ANON_23	Anonymous	*090102	*110401				
25	ANON_24	Anonymous	*07010101/07010102	*1506				

detail | detail (3) | detail (2)

Ready 100%



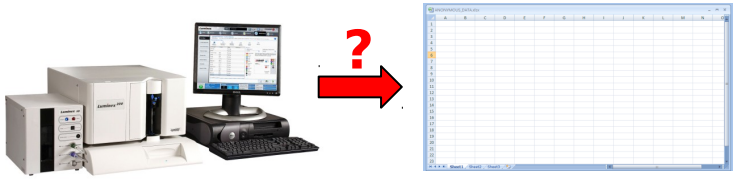
Data formating

The data may include untyped samples (here for sample number ANON_16, in green): These should be removed. Select the line(s) and remove it (them).

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_16	Anonymous	not done					
18	ANON_17	Anonymous	*07010101/07010102	*130201				
19	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
20	ANON_19	Anonymous	*040301	*160201				
21	ANON_20	Anonymous	*010101	*110101/110108				
22	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
23	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
24	ANON_23	Anonymous	*090102	*110401				
25	ANON_24	Anonymous	*07010101/07010102	*1506				

detail | detail (3) | detail (2)

Ready 100%



Data formating

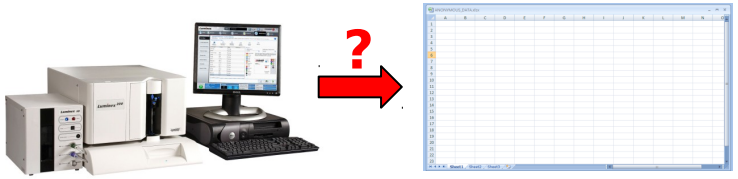
Done: untyped sample(s) have been removed (here ANON_16).

ANON_15

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
22	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | detail (2)

Ready | Count: 8 | 100%



Data formating

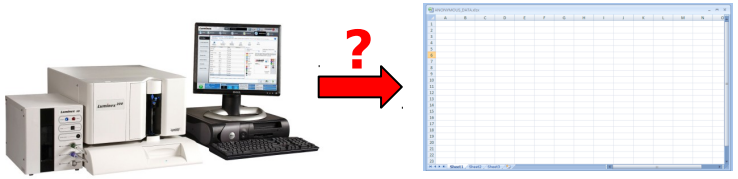
The data may include ambiguous typings (here in green for ANON_21 and ANON_22)..

ANON_21

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
22	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | detail (2)

Ready | Count: 4 | 100%



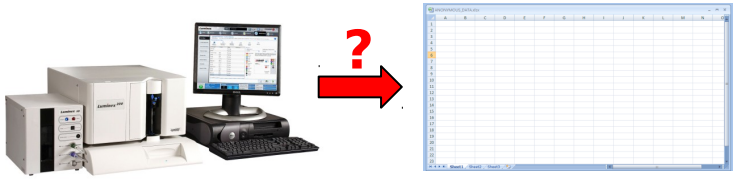
Data formating

..and/or an old nomenclature (here in green)

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
22	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | detail (2)

Ready 100%



Data formating

..ambiguous data and data coded with an old nomenclature can be recoded with the transliterate tool of Gene[rate]..

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108					
4	ANON_3	Anonymous	*03010101/03010102/030108					
5	ANON_4	Anonymous	*15010101/15010102					
6	ANON_5	Anonymous	*07010101/07010102					
7	ANON_6	Anonymous	*010101					
8	ANON_7	Anonymous	*010101					
9	ANON_8	Anonymous	*010101					
10	ANON_9	Anonymous	*010101/0107/0122					
11	ANON_10	Anonymous	*010101					
12	ANON_11	Anonymous	*07010101/07010102					
13	ANON_12	Anonymous	*07010101/07010102					
14	ANON_13	Anonymous	*07010101/07010102					
15	ANON_14	Anonymous	*0402					
16	ANON_15	Anonymous	*110401					
17	ANON_17	Anonymous	*07010101/07010102					
18	ANON_18	Anonymous	*03010101/03010102/030108					
19	ANON_19	Anonymous	*040301					
20	ANON_20	Anonymous	*010101					
21	ANON_21	Anonymous	*07010101/07010102					
22	ANON_22	Anonymous	*040501					
23	ANON_23	Anonymous	*090102					
24	ANON_24	Anonymous	*07010101/07010102					
25	ANON_25	Anonymous	*03010101/03010102/030108					

detail | detail (3) | detail (2)

Ready

Accueil Gene[VA] - Mozilla Firefox

geneva.unige.ch/generate/main.php?tool=transl

Gene[RATE]

Main

Tools

- Phenotype
- Haplotype
- Transliterate
- Uniformate
- File conversions
- Frequency estimation

Navigation:

AGP lab
Laboratory of Anthropology, Genetics and Peopling History

Generate
Tools for manipulation of data with ambiguities

Documentation
Howtos, manuals, questionnaires and other documents

AHPD
Project and documents

Search geneva.unige.ch:

transliterate UNIFORMAT datafiles

Perform allele(s) substitutions in UNIFORMAT files

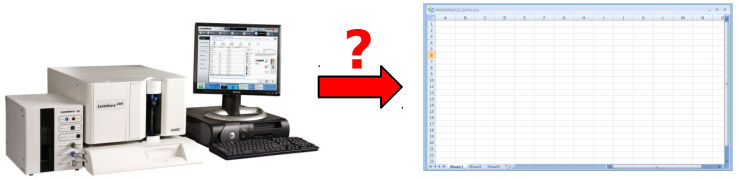
Substitution files are one column lists of **old_allele - new_allele**,
where new_allele is a abridgedUNIFORMAT allele expression: *allele or allele:allele or ....*
Substitutions related to different loci are grouped under **LOCUS** keyword, and the keyword for each locus is mandatory even if a given locus is not being changed (see example below where LOCUS 2 does not contain substitution rules).
It is recommended that uniformate be used to validate the new data obtained.

```
# Example substitution file
LOCUS 1
A*010101 - A*0101
A*02 - A*0201:A*0202:A*0203

# note no substitution data below
LOCUS 2

LOCUS 3
C*0101 - C*01
C*0102 - C*01
null - C*0101:null
```

Substitution description file: Browse...

UNIFORMAT data file: Browse...

Reset | Calculate



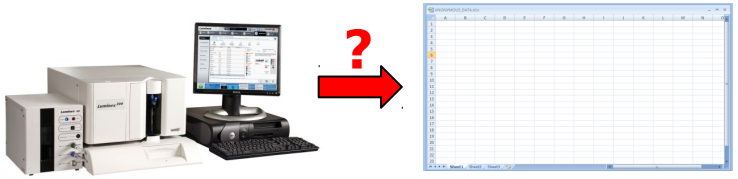
Data formating

But to use the Gene[rate] tools, your data have first to be converted into a text file containing your data in the UNIFORMAT data format. Let's do it step by step

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
22	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | detail (2)

Ready 100%



Data formating

UNIFORMAT data format:

Bar (|)

1st possible allelic combination

2nd possible allelic combination

DRB1*07010101/07010102,DRB1*1116|DRB1*07010101/07010102,DRB1*130101

1st allele (ambiguous) of possible genotype 1

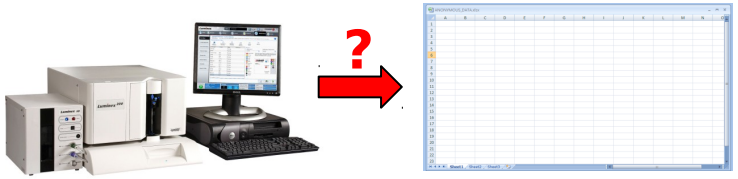
2nd allele of possible genotype 1

1st possible allele of ambiguous allele 1

2nd possible allele of ambiguous allele 1

Comma (,)

Slach (/)



Data formating

Step 1: Insert a column after each pair of columns defining a genotype. Start with the first genotype pair.

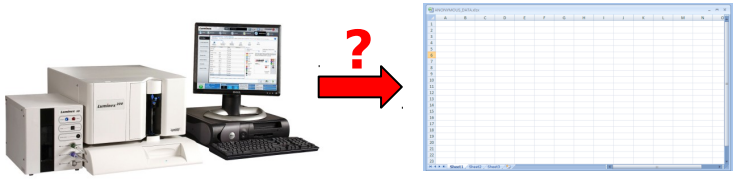
DRB1-allele1

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2	DRB1-a		DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
22	ANON_22	Anonymous	*040501	*100101	*040503	*100101	*040504	*100101
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

Cut
Copy
Paste
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Column Width...
Hide
Unhide

detail | detail (3) | Anonymous | Sheet2

Ready | Count: 6 | 100%



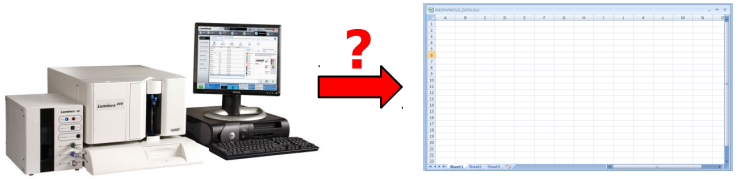
Data formating

Step 1: done for the first genotype pair!

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2		DRB1-allele1	DRB1-allele2	DRB1-allele1
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	*07010101/07010102
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	*040504
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | Anonymous | Sheet2

Ready 100%



Data formating

Step 1: continue for the second genotype pair.

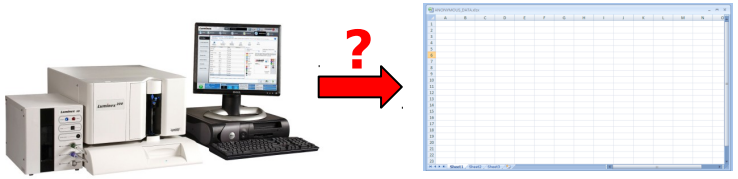
DRB1-allele1

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2		DRB1-allele1	DRB1-allele2	
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	*07010101/07010102
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	*040504
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

Cut
Copy
Paste
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Column Width...
Hide
Unhide

detail | detail (3) | Anonymous | Sheet2

Ready | Count: 5 | 100%



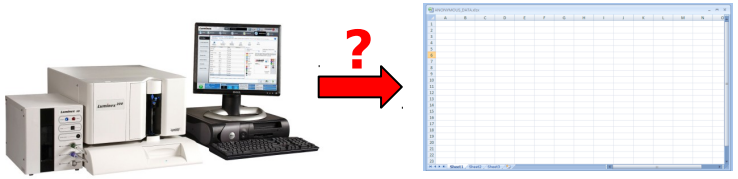
Data formating

Step 1: done for the second genotype pair!

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2		DRB1-allele1	DRB1-allele2	
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | Anonymous | Sheet2

Ready 100%



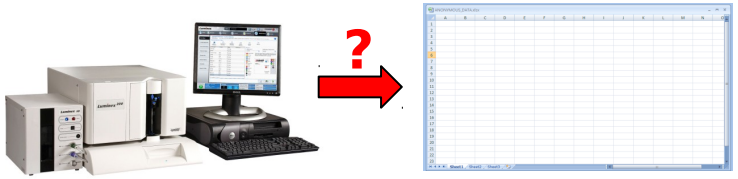
Data formating

Step 2: Fill the new column swith a bar (|)
The bar (|) can be written by using <Alt Gr> and « / »
Start with the first column to fill.

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2		DRB1-allele1	DRB1-allele2	
2	ANON_1	Anonymous	*1404	*1404	|			
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401	|			
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401	|			
5	ANON_4	Anonymous	*15010101/15010102	*150201	|			
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102	|			
7	ANON_6	Anonymous	*010101	*160201	|			
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108	|			
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready | Count: 7 | 100%



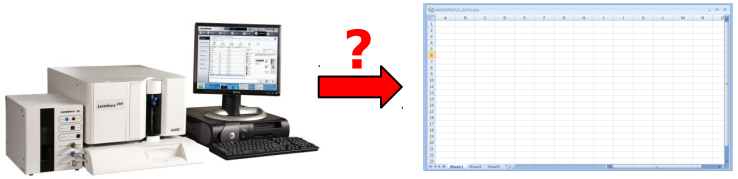
Data formating

Step 2: done for the first column to fill!
Start with the second column to fill.

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-allele2		DRB1-allele1	DRB1-allele2	
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*07010101/07010102				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*03010101/03010102/030108				
9	ANON_8	Anonymous	*010101	*03010101/03010102/030108				
10	ANON_9	Anonymous	*010101/0107/0122	*07010101/07010102/070102				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready | Count: 10 | 100%



Data formating

Step 3: Insert a semi-column (;) between the two alleles of each allelic pair
First insert a column.

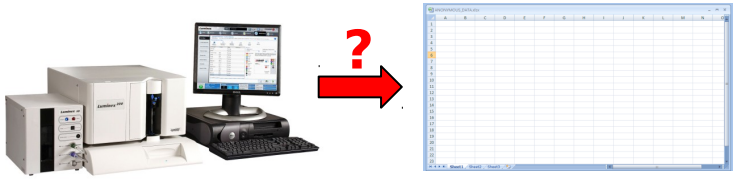
DRB1-allele2

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1	DRB1-a		DRB1-allele1	DRB1-allele2	
2	ANON_1	Anonymous	*1404	*1404				
3	ANON_2	Anonymous	*03010101/03010102/030108	*110401				
4	ANON_3	Anonymous	*03010101/03010102/030108	*110401				
5	ANON_4	Anonymous	*15010101/15010102	*150201				
6	ANON_5	Anonymous	*07010101/07010102	*070101				
7	ANON_6	Anonymous	*010101	*160201				
8	ANON_7	Anonymous	*010101	*030101				
9	ANON_8	Anonymous	*010101	*030101				
10	ANON_9	Anonymous	*010101/0107/0122	*070101				
11	ANON_10	Anonymous	*010101	*110101/110108				
12	ANON_11	Anonymous	*07010101/07010102	*07010101/07010102				
13	ANON_12	Anonymous	*07010101/07010102	*111201				
14	ANON_13	Anonymous	*07010101/07010102	*160201				
15	ANON_14	Anonymous	*0402	*07010101/07010102				
16	ANON_15	Anonymous	*110401	*15010101/15010102				
17	ANON_17	Anonymous	*07010101/07010102	*130201				
18	ANON_18	Anonymous	*03010101/03010102/030108	*130201				
19	ANON_19	Anonymous	*040301	*160201				
20	ANON_20	Anonymous	*010101	*110101/110108				
21	ANON_21	Anonymous	*07010101/07010102	*090102		*07010101/07010102	*0906	
22	ANON_22	Anonymous	*040501	*100101		*040503	*100101	
23	ANON_23	Anonymous	*090102	*110401				
24	ANON_24	Anonymous	*07010101/07010102	*1506				
25	ANON_25	Anonymous	*03010101/03010102/030108	*15010101/15010102				

Cut
Copy
Paste
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Column Width...
Hide
Unhide

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready | Count: 71 | 100%



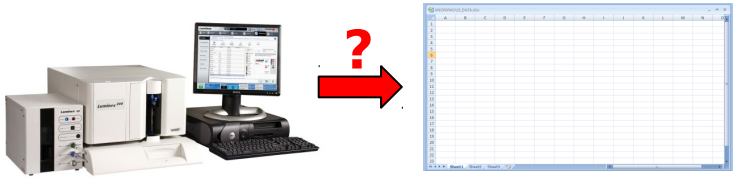
Data formating

Step 3: column inserted for the first genotype (allelic pair)!

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	DRB1-allele2
2	ANON_1	Anonymous	*1404		*1404			
3	ANON_2	Anonymous	*03010101/03010102/030108		*110401			
4	ANON_3	Anonymous	*03010101/03010102/030108		*110401			
5	ANON_4	Anonymous	*15010101/15010102		*150201			
6	ANON_5	Anonymous	*07010101/07010102		*07010101/07010102			
7	ANON_6	Anonymous	*010101		*160201			
8	ANON_7	Anonymous	*010101		*03010101/03010102/030108			
9	ANON_8	Anonymous	*010101		*03010101/03010102/030108			
10	ANON_9	Anonymous	*010101/0107/0122		*07010101/07010102/070102			
11	ANON_10	Anonymous	*010101		*110101/110108			
12	ANON_11	Anonymous	*07010101/07010102		*07010101/07010102			
13	ANON_12	Anonymous	*07010101/07010102		*111201			
14	ANON_13	Anonymous	*07010101/07010102		*160201			
15	ANON_14	Anonymous	*0402		*07010101/07010102			
16	ANON_15	Anonymous	*110401		*15010101/15010102			
17	ANON_17	Anonymous	*07010101/07010102		*130201			
18	ANON_18	Anonymous	*03010101/03010102/030108		*130201			
19	ANON_19	Anonymous	*040301		*160201			
20	ANON_20	Anonymous	*010101		*110101/110108			
21	ANON_21	Anonymous	*07010101/07010102		*090102		*07010101/07010102	*0906
22	ANON_22	Anonymous	*040501		*100101		*040503	*100101
23	ANON_23	Anonymous	*090102		*110401			
24	ANON_24	Anonymous	*07010101/07010102		*1506			
25	ANON_25	Anonymous	*03010101/03010102/030108		*15010101/15010102			

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready 100%



Data formating

Step 3: insert a column for the second genotype (allelic pair).

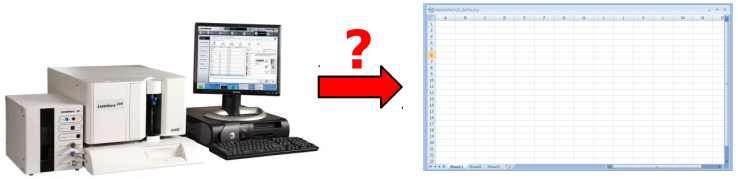
DRB1-allele2

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1	
2	ANON_1	Anonymous	*1404		*1404			
3	ANON_2	Anonymous	*03010101/03010102/030108		*110401			
4	ANON_3	Anonymous	*03010101/03010102/030108		*110401			
5	ANON_4	Anonymous	*15010101/15010102		*150201			
6	ANON_5	Anonymous	*07010101/07010102		*07010101/07010102			
7	ANON_6	Anonymous	*010101		*160201			
8	ANON_7	Anonymous	*010101		*03010101/03010102/030108			
9	ANON_8	Anonymous	*010101		*03010101/03010102/030108			
10	ANON_9	Anonymous	*010101/0107/0122		*07010101/07010102/070102			
11	ANON_10	Anonymous	*010101		*110101/110108			
12	ANON_11	Anonymous	*07010101/07010102		*07010101/07010102			
13	ANON_12	Anonymous	*07010101/07010102		*111201			
14	ANON_13	Anonymous	*07010101/07010102		*160201			
15	ANON_14	Anonymous	*0402		*07010101/07010102			
16	ANON_15	Anonymous	*110401		*15010101/15010102			
17	ANON_17	Anonymous	*07010101/07010102		*130201			
18	ANON_18	Anonymous	*03010101/03010102/030108		*130201			
19	ANON_19	Anonymous	*040301		*160201			
20	ANON_20	Anonymous	*010101		*110101/110108			
21	ANON_21	Anonymous	*07010101/07010102		*090102		*07010101/07010102	*0906
22	ANON_22	Anonymous	*040501		*100101		*040503	*100101
23	ANON_23	Anonymous	*090102		*110401			
24	ANON_24	Anonymous	*07010101/07010102		*1506			
25	ANON_25	Anonymous	*03010101/03010102/030108		*15010101/15010102			

Cut
Copy
Paste
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Column Width...
Hide
Unhide

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready | Count: 6 | 100%



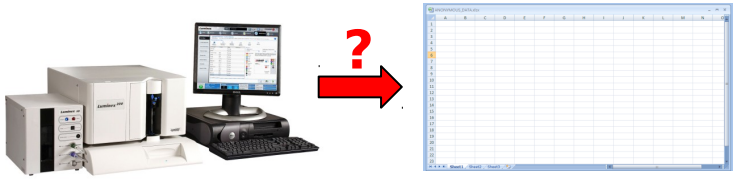
Data formating

Step 3: done (column inserted for the second genotype)!

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous	*1404		*1404			
3	ANON_2	Anonymous	*03010101/03010102/030108		*110401			
4	ANON_3	Anonymous	*03010101/03010102/030108		*110401			
5	ANON_4	Anonymous	*15010101/15010102		*150201			
6	ANON_5	Anonymous	*07010101/07010102		*07010101/07010102			
7	ANON_6	Anonymous	*010101		*160201			
8	ANON_7	Anonymous	*010101		*03010101/03010102/030108			
9	ANON_8	Anonymous	*010101		*03010101/03010102/030108			
10	ANON_9	Anonymous	*010101/0107/0122		*07010101/07010102/070102			
11	ANON_10	Anonymous	*010101		*110101/110108			
12	ANON_11	Anonymous	*07010101/07010102		*07010101/07010102			
13	ANON_12	Anonymous	*07010101/07010102		*111201			
14	ANON_13	Anonymous	*07010101/07010102		*160201			
15	ANON_14	Anonymous	*0402		*07010101/07010102			
16	ANON_15	Anonymous	*110401		*15010101/15010102			
17	ANON_17	Anonymous	*07010101/07010102		*130201			
18	ANON_18	Anonymous	*03010101/03010102/030108		*130201			
19	ANON_19	Anonymous	*040301		*160201			
20	ANON_20	Anonymous	*010101		*110101/110108			
21	ANON_21	Anonymous	*07010101/07010102		*090102		*07010101/07010102	
22	ANON_22	Anonymous	*040501		*100101		*040503	
23	ANON_23	Anonymous	*090102		*110401			
24	ANON_24	Anonymous	*07010101/07010102		*1506			
25	ANON_25	Anonymous	*03010101/03010102/030108		*15010101/15010102			

detail | detail (3) | Anonymous | Anonymous (2) | Sheet2

Ready 100%



Data formating

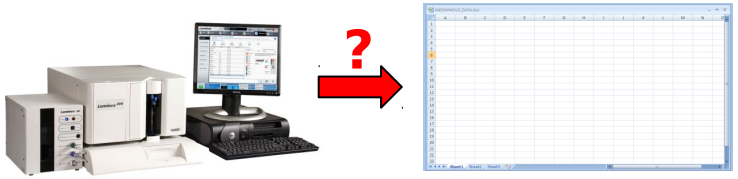
Step 3: fill the first new column with a semi-column (;)

fx ;

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous	*1404	;	*1404			
3	ANON_2	Anonymous	*03010101/03010102/030108	;	*110401			
4	ANON_3	Anonymous	*03010101/03010102/030108	;	*110401			
5	ANON_4	Anonymous	*15010101/15010102	;	*150201			
6	ANON_5	Anonymous	*07010101/07010102	;	*07010101/07010102			
7	ANON_6	Anonymous	*010101	;	*160201			
8	ANON_7	Anonymous	*010101	;	*03010101/03010102/030108			
9	ANON_8	Anonymous	*010101	;	*03010101/03010102/030108			
10	ANON_9	Anonymous	*010101/0107/0122		010101/07010102/070102			
11	ANON_10	Anonymous	*010101		*110101/110108			
12	ANON_11	Anonymous	*07010101/07010102		*07010101/07010102			
13	ANON_12	Anonymous	*07010101/07010102		*111201			
14	ANON_13	Anonymous	*07010101/07010102		*160201			
15	ANON_14	Anonymous	*0402		*07010101/07010102			
16	ANON_15	Anonymous	*110401		*15010101/15010102			
17	ANON_17	Anonymous	*07010101/07010102		*130201			
18	ANON_18	Anonymous	*03010101/03010102/030108		*130201			
19	ANON_19	Anonymous	*040301		*160201			
20	ANON_20	Anonymous	*010101		*110101/110108			
21	ANON_21	Anonymous	*07010101/07010102		*090102		*07010101/07010102	
22	ANON_22	Anonymous	*040501		*100101		*040503	
23	ANON_23	Anonymous	*090102		*110401			
24	ANON_24	Anonymous	*07010101/07010102		*1506			
25	ANON_25	Anonymous	*03010101/03010102/030108		*15010101/15010102			

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 8 | 100%



Data formating

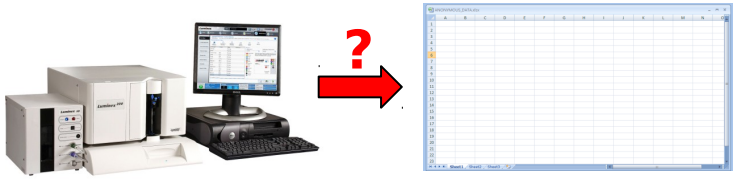
Step 3: done (first new column filled with a semi-column)!

fx ;

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous	*1404	;	*1404	|		
3	ANON_2	Anonymous	*03010101/03010102/030108	;	*110401	|		
4	ANON_3	Anonymous	*03010101/03010102/030108	;	*110401	|		
5	ANON_4	Anonymous	*15010101/15010102	;	*150201	|		
6	ANON_5	Anonymous	*07010101/07010102	;	*07010101/07010102	|		
7	ANON_6	Anonymous	*010101	;	*160201	|		
8	ANON_7	Anonymous	*010101	;	*03010101/03010102/030108	|		
9	ANON_8	Anonymous	*010101	;	*03010101/03010102/030108	|		
10	ANON_9	Anonymous	*010101/0107/0122	;	*07010101/07010102/070102	|		
11	ANON_10	Anonymous	*010101	;	*110101/110108	|		
12	ANON_11	Anonymous	*07010101/07010102	;	*07010101/07010102	|		
13	ANON_12	Anonymous	*07010101/07010102	;	*111201	|		
14	ANON_13	Anonymous	*07010101/07010102	;	*160201	|		
15	ANON_14	Anonymous	*0402	;	*07010101/07010102	|		
16	ANON_15	Anonymous	*110401	;	*15010101/15010102	|		
17	ANON_17	Anonymous	*07010101/07010102	;	*130201	|		
18	ANON_18	Anonymous	*03010101/03010102/030108	;	*130201	|		
19	ANON_19	Anonymous	*040301	;	*160201	|		
20	ANON_20	Anonymous	*010101	;	*110101/110108	|		
21	ANON_21	Anonymous	*07010101/07010102	;	*090102	|	*07010101/07010102	
22	ANON_22	Anonymous	*040501	;	*100101	|	*040503	
23	ANON_23	Anonymous	*090102	;	*110401	|		
24	ANON_24	Anonymous	*07010101/07010102	;	*1506	|		
25	ANON_25	Anonymous	*03010101/03010102/030108	;	010101/15010102	|		

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 24 | 100%



Data formating

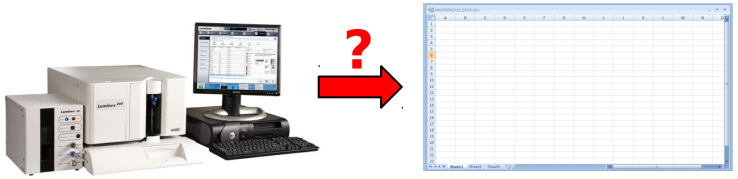
Step 3: fill the second new column with a semi-column (;)

fx ;

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous	*1404	;	*1404	|		;
3	ANON_2	Anonymous	*03010101/03010102/030108	;	*110401	|		;
4	ANON_3	Anonymous	*03010101/03010102/030108	;	*110401	|		;
5	ANON_4	Anonymous	*15010101/15010102	;	*150201	|		;
6	ANON_5	Anonymous	*07010101/07010102	;	*07010101/07010102	|		;
7	ANON_6	Anonymous	*010101	;	*160201	|		;
8	ANON_7	Anonymous	*010101	;	*03010101/03010102/030108	|		;
9	ANON_8	Anonymous	*010101	;	*03010101/03010102/030108	|		;
10	ANON_9	Anonymous	*010101/0107/0122	;	*07010101/07010102/070102	|		;
11	ANON_10	Anonymous	*010101	;	*110101/110108	|		
12	ANON_11	Anonymous	*07010101/07010102	;	*07010101/07010102	|		
13	ANON_12	Anonymous	*07010101/07010102	;	*111201	|		
14	ANON_13	Anonymous	*07010101/07010102	;	*160201	|		
15	ANON_14	Anonymous	*0402	;	*07010101/07010102	|		
16	ANON_15	Anonymous	*110401	;	*15010101/15010102	|		
17	ANON_17	Anonymous	*07010101/07010102	;	*130201	|		
18	ANON_18	Anonymous	*03010101/03010102/030108	;	*130201	|		
19	ANON_19	Anonymous	*040301	;	*160201	|		
20	ANON_20	Anonymous	*010101	;	*110101/110108	|		
21	ANON_21	Anonymous	*07010101/07010102	;	*090102	|	*07010101/07010102	
22	ANON_22	Anonymous	*040501	;	*100101	|	*040503	
23	ANON_23	Anonymous	*090102	;	*110401	|		
24	ANON_24	Anonymous	*07010101/07010102	;	*1506	|		
25	ANON_25	Anonymous	*03010101/03010102/030108	;	*15010101/15010102	|		

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 9 | 100%



Data formating

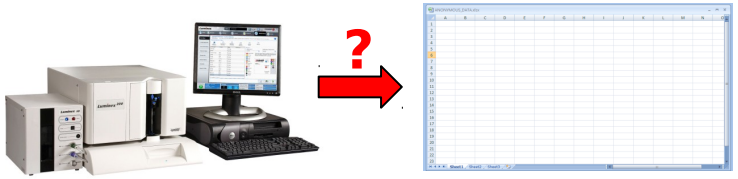
Step 3: done (second new column filled with a semi-column)!
The two alleles of each genotype has to be separated in the same way.

fx ;

	A	B	C	D	E	F	G	H
1	Sample Number	Population	DRB1-allele1		DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous	*1404	;	*1404	|		;
3	ANON_2	Anonymous	*03010101/03010102/030108	;	*110401	|		;
4	ANON_3	Anonymous	*03010101/03010102/030108	;	*110401	|		;
5	ANON_4	Anonymous	*15010101/15010102	;	*150201	|		;
6	ANON_5	Anonymous	*07010101/07010102	;	*07010101/07010102	|		;
7	ANON_6	Anonymous	*010101	;	*160201	|		;
8	ANON_7	Anonymous	*010101	;	*03010101/03010102/030108	|		;
9	ANON_8	Anonymous	*010101	;	*03010101/03010102/030108	|		;
10	ANON_9	Anonymous	*010101/0107/0122	;	*07010101/07010102/070102	|		;
11	ANON_10	Anonymous	*010101	;	*110101/110108	|		;
12	ANON_11	Anonymous	*07010101/07010102	;	*07010101/07010102	|		;
13	ANON_12	Anonymous	*07010101/07010102	;	*111201	|		;
14	ANON_13	Anonymous	*07010101/07010102	;	*160201	|		;
15	ANON_14	Anonymous	*0402	;	*07010101/07010102	|		;
16	ANON_15	Anonymous	*110401	;	*15010101/15010102	|		;
17	ANON_17	Anonymous	*07010101/07010102	;	*130201	|		;
18	ANON_18	Anonymous	*03010101/03010102/030108	;	*130201	|		;
19	ANON_19	Anonymous	*040301	;	*160201	|		;
20	ANON_20	Anonymous	*010101	;	*110101/110108	|		;
21	ANON_21	Anonymous	*07010101/07010102	;	*090102	|	*07010101/07010102	;
22	ANON_22	Anonymous	*040501	;	*100101	|	*040503	;
23	ANON_23	Anonymous	*090102	;	*110401	|		;
24	ANON_24	Anonymous	*07010101/07010102	;	*1506	|		;
25	ANON_25	Anonymous	*03010101/03010102/030108	;	*15010101/15010102	|		;

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 24 | 100%



Data formating

Step 4: delete useless columns such as to keep only one identifier column (here the column with the population name has to be removed)

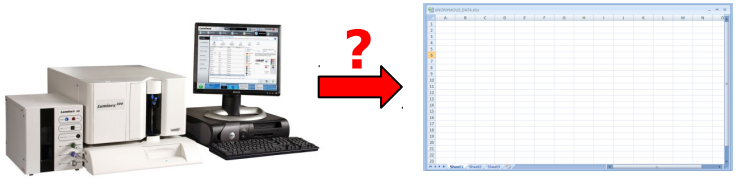
Population

	A	B	C	D	E	F	G	H
1	Sample Number	Population			DRB1-allele2		DRB1-allele1	
2	ANON_1	Anonymous			*1404			
3	ANON_2	Anonymous			*110401			
4	ANON_3	Anonymous			*110401			
5	ANON_4	Anonymous			*150201			
6	ANON_5	Anonymous			*07010101/07010102			
7	ANON_6	Anonymous			*160201			
8	ANON_7	Anonymous			*03010101/03010102/030108			
9	ANON_8	Anonymous			*03010101/03010102/030108			
10	ANON_9	Anonymous			*07010101/07010102/070102			
11	ANON_10	Anonymous	*010101		*110101/110108			
12	ANON_11	Anonymous	*07010101/07010102		*07010101/07010102			
13	ANON_12	Anonymous	*07010101/07010102		*111201			
14	ANON_13	Anonymous	*07010101/07010102		*160201			
15	ANON_14	Anonymous	*0402		*07010101/07010102			
16	ANON_15	Anonymous	*110401		*15010101/15010102			
17	ANON_17	Anonymous	*07010101/07010102		*130201			
18	ANON_18	Anonymous	*03010101/03010102/030108		*130201			
19	ANON_19	Anonymous	*040301		*160201			
20	ANON_20	Anonymous	*010101		*110101/110108			
21	ANON_21	Anonymous	*07010101/07010102		*090102		*07010101/07010102	
22	ANON_22	Anonymous	*040501		*100101		*040503	
23	ANON_23	Anonymous	*090102		*110401			
24	ANON_24	Anonymous	*07010101/07010102		*1506			
25	ANON_25	Anonymous	*03010101/03010102/030108		*15010101/15010102			

Cut
Copy
Paste
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Column Width...
Hide
Unhide

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 71 | 100%



Data formating

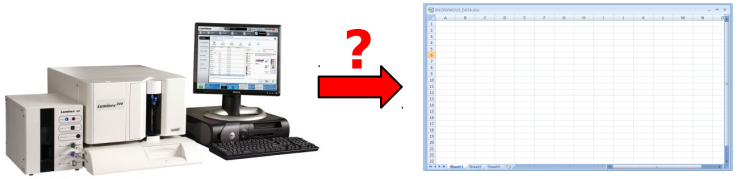
Step 4: done (useless column removed)!

DRB1-allele1

	A	B	C	D	E	F	G	H
1	Sample Number	DRB1-allele1		DRB1-allele2		DRB1-allele1		DRB1-allele2
2	ANON_1	*1404	;	*1404	|		;	
3	ANON_2	*03010101/03010102/030108	;	*110401	|		;	
4	ANON_3	*03010101/03010102/030108	;	*110401	|		;	
5	ANON_4	*15010101/15010102	;	*150201	|		;	
6	ANON_5	*07010101/07010102	;	*07010101/07010102	|		;	
7	ANON_6	*010101	;	*160201	|		;	
8	ANON_7	*010101	;	*03010101/03010102/030108	|		;	
9	ANON_8	*010101	;	*03010101/03010102/030108	|		;	
10	ANON_9	*010101/0107/0122	;	*07010101/07010102/070102	|		;	
11	ANON_10	*010101	;	*110101/110108	|		;	
12	ANON_11	*07010101/07010102	;	*07010101/07010102	|		;	
13	ANON_12	*07010101/07010102	;	*111201	|		;	
14	ANON_13	*07010101/07010102	;	*160201	|		;	
15	ANON_14	*0402	;	*07010101/07010102	|		;	
16	ANON_15	*110401	;	*15010101/15010102	|		;	
17	ANON_17	*07010101/07010102	;	*130201	|		;	
18	ANON_18	*03010101/03010102/030108	;	*130201	|		;	
19	ANON_19	*040301	;	*160201	|		;	
20	ANON_20	*010101	;	*110101/110108	|		;	
21	ANON_21	*07010101/07010102	;	*090102	|	*07010101/07010102	;	*0906
22	ANON_22	*040501	;	*100101	|	*040503	;	*100101
23	ANON_23	*090102	;	*110401	|		;	
24	ANON_24	*07010101/07010102	;	*1506	|		;	
25	ANON_25	*03010101/03010102/030108	;	*15010101/15010102	|		;	

detail | detail (3) | Anonymous | Anonymous-3 | Sheet2

Ready | Count: 71 | 100%



Data formating

Steps 5: save your table as a Text (Tab delimited) file and close it.
Here the file has been named « Anonymous.txt ». But a better filename should include both the population name and the locus name,
e.g. « Anonymous_DRB1.txt »

Sample Number

	A	B					G	H
1	Sample Number	DRB1-allele1						DRB1-allele2
2	ANON_1	*1404						
3	ANON_2	*03010101/03010102/						
4	ANON_3	*03010101/03010102/						
5	ANON_4	*15010101/15010102						
6	ANON_5	*07010101/07010102						
7	ANON_6	*010101						
8	ANON_7	*010101						
9	ANON_8	*010101						
10	ANON_9	*010101/0107/0122						
11	ANON_10	*010101						
12	ANON_11	*07010101/07010102						
13	ANON_12	*07010101/07010102						
14	ANON_13	*07010101/07010102						
15	ANON_14	*0402						
16	ANON_15	*110401						
17	ANON_17	*07010101/07010102						
18	ANON_18	*03010101/03010102/						
19	ANON_19	*040301						
20	ANON_20	*010101	;	*110101/110108	|		;	
21	ANON_21	*07010101/07010102	;	*090102	|	*07010101/07010102	;	*0906
22	ANON_22	*040501	;	*100101	|	*040503	;	*100101
23	ANON_23	*090102	;	*110401	|		;	
24	ANON_24	*07010101/07010102	;	*1506	|		;	
25	ANON_25	*03010101/03010102/030108	;	*15010101/15010102	|		;	

detail | detail (3) | detail (2)

Save As

Save in: Teaching_Session

My Recent Documents
Desktop
My Documents
My Computer
My Network Places

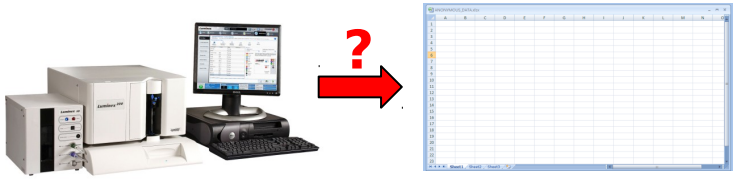
Name | Size | Type | Date

File name: Anonymous.txt

Save as type: Text (Tab delimited) (*.txt)

Tools | Save | Cancel

Ready 100%



Data formating

Step 6: open your new .txt file with a <u>text editor</u> (e.g. Textpad) (but NOT with Word!)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt

```
Sample Number   DRB1-allele1          DRB1-allele2          DRB1-allele1          DRB1-allele2          DRB1-allele1          DRB1-alle
ANON_1  *1404   ;       *1404   |               ;
ANON_2  *03010101/03010102/030108       ;       *110401 |
ANON_3  *03010101/03010102/030108       ;       *110401 |
ANON_4  *15010101/15010102      ;       *150201 |
ANON_5  *07010101/07010102      ;       *07010101/07010102
ANON_6  *010101 ;       *160201 |               ;
ANON_7  *010101 ;       *03010101/03010102/030108       |
ANON_8  *010101 ;       *03010101/03010102/030108       |
ANON_9  *010101/0107/0122       ;       *07010101/07010102/0701
ANON_10 *010101 ;       *110101/110108  |               ;
ANON_11 *07010101/07010102      ;       *07010101/07010102
ANON_12 *07010101/07010102      ;       *111201 |
ANON_13 *07010101/07010102      ;       *160201 |
ANON_14 *0402   ;       *07010101/07010102      |
ANON_15 *110401 ;       *15010101/15010102      |
ANON_17 *07010101/07010102      ;       *130201 |
ANON_18 *03010101/03010102/030108       ;       *130201 |
ANON_19 *040301 ;       *160201 |               ;
ANON_20 *010101 ;       *110101/110108  |
ANON_21 *07010101/07010102      ;       *090102 |       *07010101/07010102      ;       *0906   |       *07010101/07010102      ;       *0909
ANON_22 *040501 ;       *100101 |
ANON_23 *090102 ;       *110401 |
ANON_24 *07010101/07010102      ;
ANON_25 *03010101/03010102/030108
ANON_27 *110401 ;       *110401 |
ANON_28 *110401 ;       *110401 |
ANON_29 *110401 ;       *110401 |
ANON_30 *07010101/07010102      ;
ANON_31 *010101 ;       *130201 |
ANON_32 *07010101/07010102      ;       *130101 |
ANON_33 *040401 ;       *130101 |
ANON_35 *03010101/03010102/030108       ;       *0402   |
ANON_36 *07010101/07010102      ;       *15010101/15010102      |
ANON_37 *080302 ;       *110401 |
ANON_38 *010101 ;       *150201 |
ANON_39 *110601 ;       *120101/1206/1210/1217  |
ANON_40 *07010101/07010102      ;       *080101/080103  |
ANON_41 *080201 ;       *130101 |
ANON_42 *010101 ;       *07010101/07010102      |
ANON_43 *03010101/03010102/030108       ;       *07010101/07010102      |
ANON_44 *07010101/07010102      ;       *090102 |
ANON_45 *03010101/03010102/030108       ;       *160101 |
ANON_46 *010101 ;       *03010101/03010102/030108       |
ANON_47 *07010101/07010102      ;       *1116   |       *07010101/07010102      ;       *130101 |       *07010101/07010102      ;       *130104
```

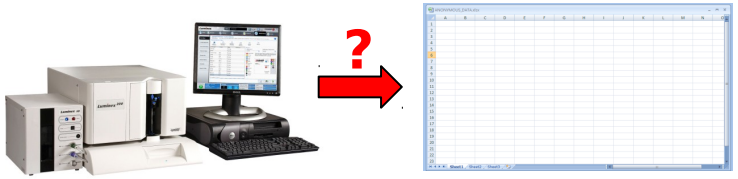
ANSI Characters

File: Anonymous.txt, 4482 bytes, 72 lines, PC, ANSI

As a text editor, use for example:

- **Notepad, Textpad (Windows)**
- **Wordpad, Gedit (Mac)**
- **Gedit (Linux)**

Tip: at any time, use the « visible » icon to check the format



Data formating

Step 7: use the *Replace* command to perform additional changes
- remove tabulations before and after the bars (|)
(remember that the bar can be written by using <Alt Gr> and « / »)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt

ANSI Characters

Sample Number	DRB1-allele1	DRB1-allele2	DRB1-allele1	DRB1-allele2
ANON_1	*1404	*1404		
ANON_2	*03010101/03010102/030108	*110401		
ANON_3	*03010101/03010102/030108	*110401		
ANON_4	*15010101/15010102	*150201		
ANON_5	*07010101/07010102	*07010101/07010102		
ANON_6	*010101	*160201		
ANON_7	*010101	*03010101/03010102/030108		
ANON_8	*010101	*03010101/03010102/030108		
ANON_9	*010101/0107/0122	*07010101/07010102/070102		
ANON_10	*010101	*110101/110108		
ANON_11	*07010101/07010102	*07010101/07010102		
ANON_12	*07010101/07010102	*111201		
ANON_13	*07010101/07010102	*160201		
ANON_14	*0402	*07010101/07010102		
ANON_15	*110401	*15010101/15010102		
ANON_17	*07010101/07010102	*130201		
ANON_18	*03010101/03010102/030108	*130201		
ANON_19	*040301	*160201		
ANON_20	*010101	*110101/110108		
ANON_21	*07010101/07010102	*090102	*07010101/07010102	*0906	*07010101/07010102	*0909
ANON_22	*040501	*100101		
ANON_23	*090102	*110401		
ANON_24	*07010101/07010102			
ANON_25	*03010101/03010102/030108			
ANON_27	*110401	*110401		
ANON_28	*110401	*110401		
ANON_29	*110401	*110401		
ANON_30	*07010101/07010102			
ANON_31	*010101	*130201		
ANON_32	*07010101/07010102			
ANON_33	*040401	*130101		
ANON_35	*03010101/03010102/030108			
ANON_36	*07010101/07010102			
ANON_37	*080302	*110401		
ANON_38	*010101	*150201		
ANON_39	*110601	*120101/1206		
ANON_40	*07010101/07010102			
ANON_41	*080201	*130101		
ANON_42	*010101	*07010101/07010102		
ANON_43	*03010101/03010102/030108	*07010101/07010102		
ANON_44	*07010101/07010102	*090102		
ANON_45	*03010101/03010102/030108	*160101		
ANON_46	*010101	*03010101/03010102/030108		
ANON_47	*07010101/07010102	*1116	*07010101/07010102	*130101	*07010101/07010102	*130104

Replace

Find what: \t|\t
Replace with: |

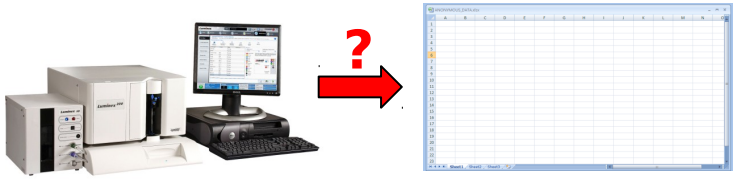
Find Next
Replace
Replace Next
Replace All
Close
Help

Conditions:
- Text / Hex
- Match whole words
- Match case
- Regular expression

Scope:
- Active document
- Selected text
- All documents

In the Replace box:
- select « regular expression »
- replace \t|\t with | (\t is a tabulation)
- repeat this replacement until no more changes are done

For Help, press F1



Data formating

Step 7: use the *Replace* command to perform additional changes - tabulations before and after the bars (|) have been removed!

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt *]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt *

```
Sample Number   DRB1-allele1          DRB1-allele2          DRB1-allele1          DRB1-allele2
ANON_1  *1404   ;       *1404|  ;|;
ANON_2  *03010101/03010102/030108       ;       *110401|        ;|;
ANON_3  *03010101/03010102/030108       ;       *110401|        ;|;
ANON_4  *15010101/15010102      ;       *150201|        ;|;
ANON_5  *07010101/07010102      ;       *07010101/07010102|     ;|;
ANON_6  *010101 ;       *160201|        ;|;
ANON_7  *010101 ;       *03010101/03010102/030108|      ;|;
ANON_8  *010101 ;       *03010101/03010102/030108|      ;|;
ANON_9  *010101/0107/0122       ;       *07010101/07010102/070102|      ;|;
ANON_10 *010101 ;       *110101/110108| ;|;
ANON_11 *07010101/07010102      ;       *07010101/07010102|     ;|;
ANON_12 *07010101/07010102      ;       *111201|        ;|;
ANON_13 *07010101/07010102      ;       *160201|        ;|;
ANON_14 *0402   ;       *07010101/07010102|     ;|;
ANON_15 *110401 ;       *15010101/15010102|     ;|;
ANON_17 *07010101/07010102      ;       *130201|        ;|;
ANON_18 *03010101/03010102/030108       ;       *130201|        ;|;
ANON_19 *040301 ;       *160201|        ;|;
ANON_20 *010101 ;       *110101/110108| ;|;
ANON_21 *07010101/07010102      ;       *090102|*07010101/07010102      ;       *0906|*07010101/07010102        ;       *0909
ANON_22 *040501 ;       *100101|*040503 ;       *100101|*040504 ;       *100101
ANON_23 *090102 ;       *110401|        ;|;
ANON_24 *07010101/07010102      ;       *1506|  ;|;
ANON_25 *03010101/03010102/030108       ;       *15010101/15010102|     ;|;
ANON_27 *110401 ;       *110401|        ;|;
ANON_28 *110401 ;       *110401|        ;|;
ANON_29 *110401 ;       *110401|*110401 ;       *110402|        ;
ANON_30 *07010101/07010102      ;       *110101/110108| ;|;
ANON_31 *010101 ;       *130201|        ;|;
ANON_32 *07010101/07010102      ;       *130101|        ;|;
ANON_33 *040401 ;       *130101|        ;|;
ANON_35 *03010101/03010102/030108       ;       *0402|  ;|;
ANON_36 *07010101/07010102      ;       *15010101/15010102|     ;|;
ANON_37 *080302 ;       *110401|        ;|;
ANON_38 *010101 ;       *150201|        ;|;
ANON_39 *110601 ;       *120101/1206/1210/1217| ;|;
ANON_40 *07010101/07010102      ;       *080101/080103| ;|;
ANON_41 *080201 ;       *130101|        ;|;
ANON_42 *010101 ;       *07010101/07010102|     ;|;
ANON_43 *03010101/03010102/030108       ;       *07010101/07010102|     ;|;
ANON_44 *07010101/07010102      ;       *090102|        ;|;
ANON_45 *03010101/03010102/030108       ;       *160101|        ;|;
ANON_46 *010101 ;       *03010101/03010102/030108|      ;|;
ANON_47 *07010101/07010102      ;       *1116|*07010101/07010102        ;       *130101|*07010101/07010102      ;       *130104
```

Replace

Find what: \t|\t

Replace with: |

Find Next | Replace | Replace Next | Replace All | Close | Help

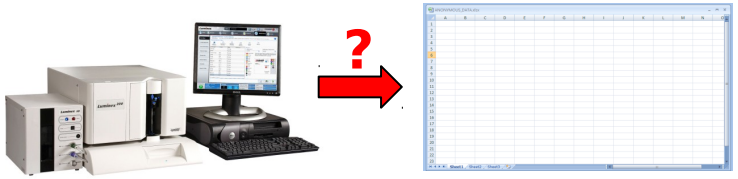
Conditions: Text / Hex; Match whole words; Match case; Regular expression

Scope: Active document; Selected text; All documents

ANSI Characters

```
33  !
34  "
35  #
36  $
37  %
38  &
39  '
40  (
41  )
42  *
43  +
44  ,
45  -
46  .
47  /
48  0
49  1
50  2
51  3
52  4
```

1 1 Read Ovr Block Sync Rec Caps



Data formating

Step 7: use the *Replace* command to perform additional changes
- Perform other changes as indicated below (1)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt *]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt *

```
Sample Number   DRB1-allele1          DRB1-allele2          DRB1-allele1          DRB1-allele2
ANON_1   *1404    ;        *1404|  ;|;
ANON_2   *03010101/03010102/030108        ;        *110401|        ;|;
ANON_3   *03010101/03010102/030108        ;        *110401|        ;|;
ANON_4   *15010101/15010102       ;        *150201|        ;|;
ANON_5   *07010101/07010102       ;        *07010101/07010102|      ;|;
ANON_6   *010101 ;        *160201|        ;|;
ANON_7   *010101 ;        *03010101/03010102/030108|       ;|;
ANON_8   *010101 ;        *03010101/03010102/030108|       ;|;
ANON_9   *010101/0107/0122        ;        *07010101/07010102/070102|       ;|;
ANON_10  *010101 ;        *110101/110108| ;|;
ANON_11  *07010101/07010102       ;        *07010101/07010102|      ;|;
ANON_12  *07010101/07010102       ;        *111201|        ;|;
ANON_13  *07010101/07010102       ;        *160201|        ;|;
ANON_14  *0402    ;        *07010101/07010102|      ;|;
ANON_15  *110401 ;        *15010101/15010102|      ;|;
ANON_17  *07010101/07010102       ;        *130201|        ;|;
ANON_18  *03010101/03010102/030108        ;        *130201|        ;|;
ANON_19  *040301 ;        *160201|        ;|;
ANON_38  *010101 ;        *150201|        ;|;
ANON_39  *110601 ;        *120101/1206/1210/1217| ;|;
ANON_40  *07010101/07010102       ;        *080101/080103| ;|;
ANON_41  *080201 ;        *130101|        ;|;
ANON_42  *010101 ;        *07010101/07010102|      ;|;
ANON_43  *03010101/03010102/030108        ;        *07010101/07010102|      ;|;
ANON_44  *07010101/07010102       ;        *090102|        ;|;
ANON_45  *03010101/03010102/030108        ;        *160101|        ;|;
ANON_46  *010101 ;        *03010101/03010102/030108|       ;|;
ANON_47  *07010101/07010102       ;        *1116|*07010101/07010102        ;        *130101|*07010101/07010102       ;        *130104
```

Replace

Find what: \t;\t
Replace with: ,

Find Next
Replace
Replace Next
Replace All
Close
Help

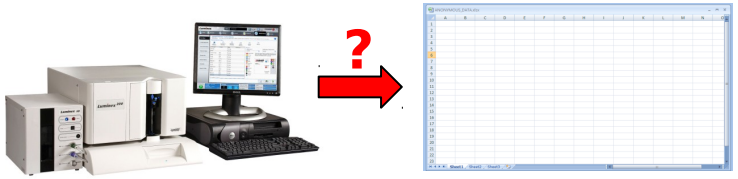
Conditions:
- Text / Hex
- Match whole words
- Match case
- Regular expression

Scope:
- Active document
- Selected text
- All documents

ANSI Characters

```
33  !
34  "
35  #
36  $
37  %
38  &
39  '
40  (
41  )
42  *
43  +
44  ,
45  -
46  .
47  /
48  0
49  1
50  2
51  3
52  4
```

In the Replace box:
- select « regular expression »
- replace \t;\t with ,
- repeat this replacement until no more changes are done



Data formating

Step 7: use the *Replace* command to perform additional changes - Perform other changes as indicated below (2)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt *]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt *

```
Sample Number   DRB1-allele1          DRB1-allele2          DRB1-allele1          DRB1-allele2
ANON_1   *1404    ;        *1404|  ;|;
ANON_2   *03010101/03010102/030108        ;        *110401|        ;|;
ANON_3   *03010101/03010102/030108        ;        *110401|        ;|;
ANON_4   *15010101/15010102       ;        *150201|        ;|;
ANON_5   *07010101/07010102       ;        *07010101/07010102|      ;|;
ANON_6   *010101 ;        *160201|        ;|;
ANON_7   *010101 ;        *03010101/03010102/030108|       ;|;
ANON_8   *010101 ;        *03010101/03010102/030108|       ;|;
ANON_9   *010101/0107/0122        ;        *07010101/07010102/070102|       ;|;
ANON_10  *010101 ;        *110101/110108| ;|;
ANON_11  *07010101/07010102       ;        *07010101/07010102|      ;|;
ANON_12  *07010101/07010102       ;        *111201|        ;|;
ANON_13  *07010101/07010102       ;        *160201|        ;|;
ANON_14  *0402    ;        *07010101/07010102|      ;|;
ANON_15  *110401 ;        *15010101/15010102|      ;|;
ANON_17  *07010101/07010102       ;        *130201|        ;|;
ANON_35  *03010101/03010102/030108        ;        *0402|  ;|;
ANON_36  *07010101/07010102       ;        *15010101/15010102|      ;|;
ANON_37  *080302 ;        *110401|        ;|;
ANON_38  *010101 ;        *150201|        ;|;
ANON_39  *110601 ;        *120101/1206/1210/1217| ;|;
ANON_40  *07010101/07010102       ;        *080101/080103| ;|;
ANON_41  *080201 ;        *130101|        ;|;
ANON_42  *010101 ;        *07010101/07010102|      ;|;
ANON_43  *03010101/03010102/030108        ;        *07010101/07010102|      ;|;
ANON_44  *07010101/07010102       ;        *090102|        ;|;
ANON_45  *03010101/03010102/030108        ;        *160101|        ;|;
ANON_46  *010101 ;        *03010101/03010102/030108|       ;|;
ANON_47  *07010101/07010102       ;        *1116|*07010101/07010102        ;        *130101|*07010101/07010102        ;        *130104
```

Replace

Find what: \t;\t

Replace with: ,

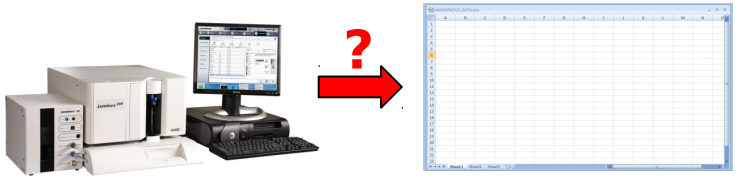
Find Next | Replace | Replace Next | Replace All | Close | Help

Conditions: Text / Hex; Match whole words; Match case; Regular expression

Scope: Active document; Selected text; All documents

ANSI Characters

Then in the Replace box:
- **replace \t$ with nothing ($ is the end of ligne)**
- **replace |\t with nothing**
- **replace | with nothing**



Data formating

Step 7: use the *Replace* command to perform additional changes
- Perform other changes as indicated below (3)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt *]

File Edit Search View Tools Macros Configure Window Help

Anonymous.txt *

```
Sample Number	DRB1-allele1DRB1-allele2DRB1-allele1DRB1-allele2|DRB1-allele1DRB1-allele2$
ANON_1	DRB1*1404,DRB1*1404
ANON_2	DRB1*03010101/03010102/030108,DRB1*110401
ANON_3	DRB1*03010101/03010102/030108,DRB1*110401
ANON_4	DRB1*15010101/15010102,DRB1*150201
ANON_5	DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_6	DRB1*010101,DRB1*160201
ANON_7	DRB1*010101,DRB1*03010101/03010102/030108
ANON_8	DRB1*010101,DRB1*03010101/03010102/030108
ANON_9	DRB1*010101/0107/0122,DRB1*07010101/07010102/070102
ANON_10	DRB1*010101,DRB1*110101/110108
ANON_11	DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_12	DRB1*07010101/07010102,DRB1*111201
ANON_13	DRB1*07010101/07010102,DRB1*160201
ANON_14	DRB1*0402,DRB1*07010101/07010102
ANON_15	DRB1*110401,DRB1*15010101/15010102
ANON_17	DRB1*07010101/07010102,DRB1*130201
ANON_18	DRB1*03010101/03010102/030108,DRB1*130201
ANON_37	DRB1*080302,DRB1*110401
ANON_38	DRB1*010101,DRB1*150201
ANON_39	DRB1*110601,DRB1*120101/1206/1210/1217
ANON_40	DRB1*07010101/07010102,DRB1*080101/080103
ANON_41	DRB1*080201,DRB1*130101
ANON_42	DRB1*010101,DRB1*07010101/07010102
ANON_43	DRB1*03010101/03010102/030108,DRB1*07010101/07010102
ANON_44	DRB1*07010101/07010102,DRB1*090102
ANON_45	DRB1*03010101/03010102/030108,DRB1*160101
ANON_46	DRB1*010101,DRB1*03010101/03010102/030108
ANON_47	DRB1*07010101/07010102,DRB1*1116|DRB1*07010101/07010102,DRB1*130101|DRB1*07010101/07010102,DRB1*130104$
```

Replace

Find what: *

Replace with: DRB1*

Find Next | Replace | Replace Next | Replace All | Close | Help

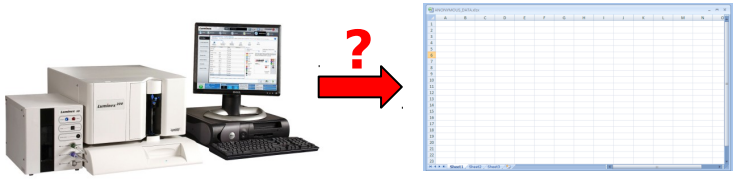
Conditions: Text / Hex / Match whole words / Match case / Regular expression

Scope: Active document / Selected text / All documents

ANSI Characters

Replaced 158 occurrence(s)

As allele names should start with locus name,
In the Replace box:
- unselect « Regular expression »
- replace * with DRB1* (i.e. the locus you are working with)



Data formating

Step 7: use the *Replace* command to perform additional changes
- Perform other changes as indicated below (4)

TextPad - [C:\Documents and Settings\Alicia Sanchez-Mazas\My Documents\Congres&Conferences\2012_16IHIW_Liverpool\Teaching_Session\Anonymous.txt *]

File Edit Search View Tools Macros Configure Window Help

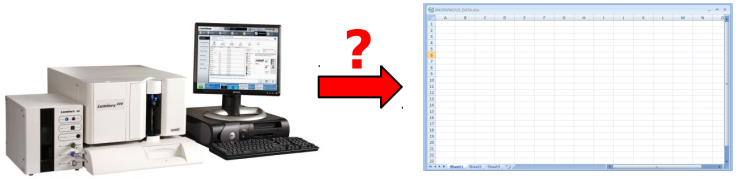
New Ctrl+N
Open... Ctrl+O
Close
Close All
Save Ctrl+S
Save As... F12
Save All Ctrl+Maj+S
Revert
Rename...
Workspace
Manage Files... F3
Print... Ctrl+P
Print Preview Ctrl+Maj+P
Page Setup...
Send...
1 Anonymous.txt
2 Anonymous.unif
3 test.unif
4 ANONYMOUS_DRB1.unif
Exit

```
B1*1404,DRB1*1404
B1*03010101/03010102/030108,DRB1*110401
B1*03010101/03010102/030108,DRB1*110401
B1*15010101/15010102,DRB1*150201
B1*07010101/07010102,DRB1*07010101/07010102
B1*010101,DRB1*160201
B1*010101,DRB1*03010101/03010102/030108
B1*010101,DRB1*03010101/03010102/030108
B1*010101/0107/0122,DRB1*07010101/07010102/070102
B1*010101,DRB1*110101/110108
B1*07010101/07010102,DRB1*07010101/07010102
B1*07010101/07010102,DRB1*111201
B1*07010101/07010102,DRB1*160201
B1*0402,DRB1*07010101/07010102
B1*110401,DRB1*15010101/15010102
B1*07010101/07010102,DRB
B1*03010101/03010102/030
B1*040301,DRB1*160201
B1*010101,DRB1*110101/11
B1*07010101/07010102,DRB
B1*040501,DRB1*100101|DR
B1*090102,DRB1*110401
B1*07010101/07010102,DRB1*1506
B1*03010101/03010102/030108,DRB1*15010101/15010102
B1*110401,DRB1*110401
B1*110401,DRB1*110401
B1*110401,DRB1*110401|DRB1*1
B1*07010101/07010102,DRB1*11
B1*010101,DRB1*130201
ANON_32 DRB1*07010101/07010102,DRB1*13
ANON_33 DRB1*040401,DRB1*130101
ANON_35 DRB1*03010101/03010102/030108,
ANON_36 DRB1*07010101/07010102,DRB1*15
ANON_37 DRB1*080302,DRB1*110401
ANON_38 DRB1*010101,DRB1*150201
ANON_39 DRB1*110601,DRB1*120101/1206/1
ANON_40 DRB1*07010101/07010102,DRB1*08
ANON_41 DRB1*080201,DRB1*130101
ANON_42 DRB1*010101,DRB1*07010101/0701
ANON_43 DRB1*03010101/03010102/030108,
ANON_44 DRB1*07010101/07010102,DRB1*09
ANON_45 DRB1*03010101/03010102/030108,
ANON_46 DRB1*010101,DRB1*03010101/0301
ANON_47 DRB1*07010101/07010102,DRB1*11
ANON_48 DRB1*07010101/07010102,DRB1*07
```

- **Remove first line to start with genotypes**

Your file is ready!
Save your file by giving it the extension .unif instead of .txt and close it

Save the active document



Data formating

Gene[RATE]

Main

Tools

- Phenotype
- Haplotype
- Transliterate
- Uniformate
- File conversions
- Frequency estimation
- One-locus summary
- Regional analysis

Navigation:

AGP lab
Laboratory of Anthropology, Genetics and Peopling History

HLA data analysis
Tests for HWE, LD, Disease-association

Generate
Tools for manipulation of data with ambiguities

Documentation
Howtos, manuals, questionnaires and other documents

AHPD
Project and documents

validate UNIFORMAT syntaxe

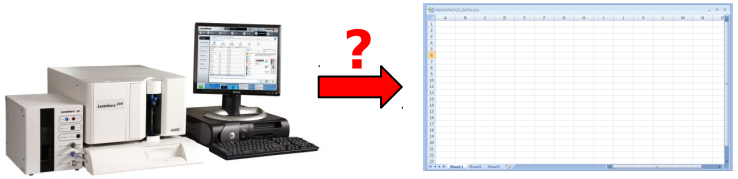
Check UNIFORMAT files for errors and performs expansion of valid abreviations (@, allele:allele).
Should be used after hand modification or *transliterate* of datafiles.

UNIFORMAT data file: C:\Documents and Settii Browse...

Reset Check

Use the Gene[rate] tools at:
http://geneva.unige.ch/generate

Steps to format your data for Transliterate: Validate the format with Uniformate



Data formating

Gene[RATE]

Main

Tools

- Phenotype
- Haplotype
- Transliterate
- Uniformate
- File conversions
- Frequency estimation
- One-locus summary
- Regional analysis

Navigation:

AGP lab
Laboratory of Anthropology, Genetics and Peopling History

HLA data analysis
Tests for HWE, LD, Disease-association

Generate
Tools for manipulation of data with ambiguities

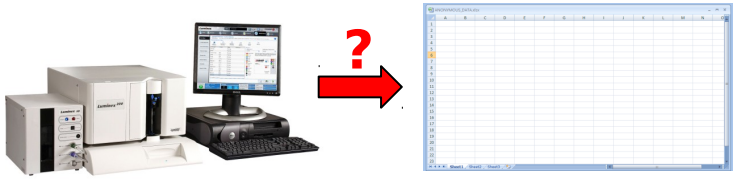
Documentation
Howtos, manuals, questionnaires and other documents

AHPD
Project and documents

Validated expanded UNIFORMAT data

Download the results file

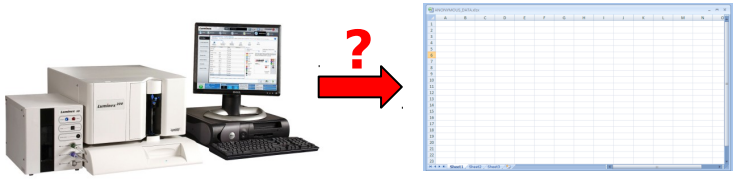
Steps to format your data for Transliterate:
Validate the format with Uniformate: done!



Data formating

```
ANON_1  DRB1*1404,DRB1*1404
ANON_2  DRB1*03010101/03010102/030108,DRB1*110401
ANON_3  DRB1*03010101/03010102/030108,DRB1*110401
ANON_4  DRB1*15010101/15010102,DRB1*150201
ANON_5  DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_6  DRB1*010101,DRB1*160201
ANON_7  DRB1*03010101/03010102/030108,DRB1*010101
ANON_8  DRB1*03010101/03010102/030108,DRB1*010101
ANON_9  DRB1*010101/0107/0122,DRB1*07010101/07010102/070102
ANON_10 DRB1*010101,DRB1*110101/110108
ANON_11 DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_12 DRB1*07010101/07010102,DRB1*111201
ANON_13 DRB1*07010101/07010102,DRB1*160201
ANON_14 DRB1*07010101/07010102,DRB1*0402
ANON_15 DRB1*110401,DRB1*15010101/15010102
ANON_17 DRB1*07010101/07010102,DRB1*130201
ANON_18 DRB1*03010101/03010102/030108,DRB1*130201
ANON_19 DRB1*160201,DRB1*040301
ANON_20 DRB1*010101,DRB1*110101/110108
ANON_21 DRB1*07010101/07010102,DRB1*090102|DRB1*07010101/07010102,DRB1*0906|DRB1*07010101/07010102,DRB1*0909
ANON_22 DRB1*040501,DRB1*100101|DRB1*100101,DRB1*040503|DRB1*100101,DRB1*040504
ANON_23 DRB1*110401,DRB1*090102
ANON_24 DRB1*07010101/07010102,DRB1*1506
ANON_25 DRB1*03010101/03010102/030108,DRB1*15010101/15010102
ANON_27 DRB1*110401,DRB1*110401
ANON_28 DRB1*110401,DRB1*110401
ANON_29 DRB1*110401,DRB1*110401|DRB1*110401,DRB1*110402
ANON_30 DRB1*07010101/07010102,DRB1*110101/110108
ANON_31 DRB1*010101,DRB1*130201
ANON_32 DRB1*07010101/07010102,DRB1*130101
ANON_33 DRB1*130101,DRB1*040401
ANON_35 DRB1*03010101/03010102/030108,DRB1*0402
ANON_36 DRB1*15010101/15010102,DRB1*07010101/07010102
ANON_37 DRB1*110401,DRB1*080302
ANON_38 DRB1*150201,DRB1*010101
ANON_39 DRB1*110601,DRB1*120101/1206/1210/1217
ANON_40 DRB1*07010101/07010102,DRB1*080101/080103
ANON_41 DRB1*130101,DRB1*080201
ANON_42 DRB1*07010101/07010102,DRB1*010101
ANON_43 DRB1*03010101/03010102/030108,DRB1*07010101/07010102
ANON_44 DRB1*07010101/07010102,DRB1*090102
ANON_45 DRB1*03010101/03010102/030108,DRB1*160101
ANON_46 DRB1*03010101/03010102/030108,DRB1*010101
ANON_47 DRB1*07010101/07010102,DRB1*130101|DRB1*07010101/07010102,DRB1*1116|DRB1*07010101/07010102,DRB1*130104
ANON_48 DRB1*07010101/07010102,DRB1*07010101/07010102
```

Steps to format your data for Transliterate:
Validate the format with Uniformate: file downloaded



Data formating

Gene[RATE]

Main

Tools

- Phenotype
- Haplotype
- Transliterate
- Uniformate
- File conversions
- Frequency estimation
- One-locus summary
- Regional analysis

Navigation:

AGP lab
Laboratory of Anthropology, Genetics and Peopling History

HLA data analysis
Tests for HWE, LD, Disease-association

Generate
Tools for manipulation of data with ambiguities

Documentation
Howtos, manuals, questionnaires and other documents

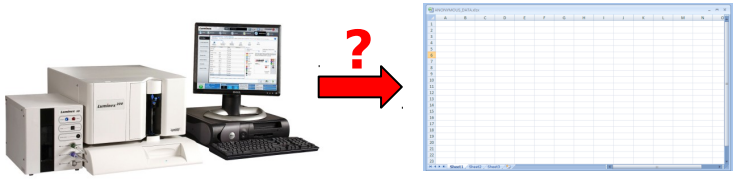
AHPD
Project and documents

Validated expanded UNIFORMAT data

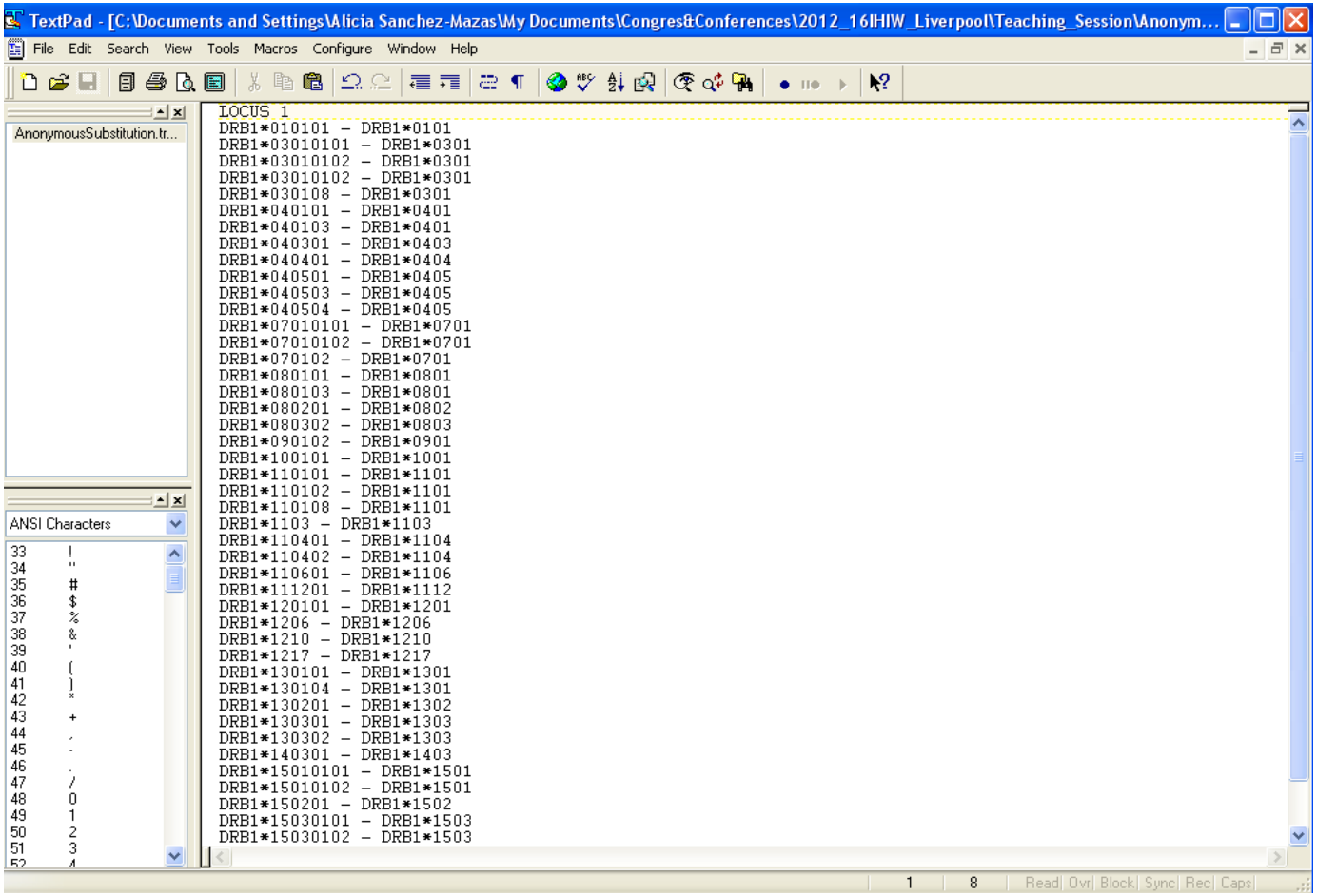
An error occured, please try again.

If the error persists contact us.

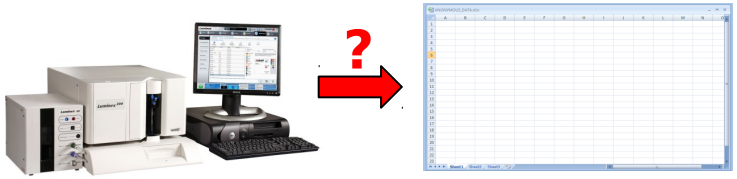
Steps to format your data for Transliterate:
Validate the format with Uniformate. If not good, an error message appears: correct the file.



Data formating



Prepare a substitution file for Transliterate to reduce high resolution to 2nd level (4 digits) or old nomenclature into new



Data formating

Gene[RATE]

Main

Tools

- Phenotype
- Haplotype
- Transliterate
- Uniformate
- File conversions
- Frequency estimation
- One-locus summary
- Regional analysis

Navigation:

AGP lab
Laboratory of Anthropology, Genetics and Peopling History

HLA data analysis
Tests for HWE, LD, Disease-association

Generate
Tools for manipulation of data with ambiguities

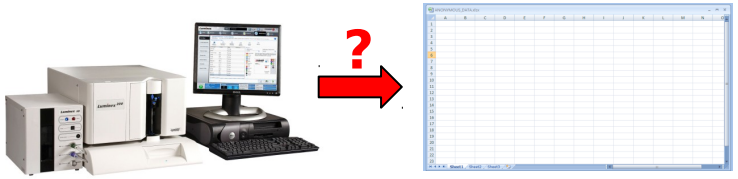
Documentation
Howtos, manuals, questionnaires and other documents

AHPD
Project and documents

Transliterated file

Download the results file

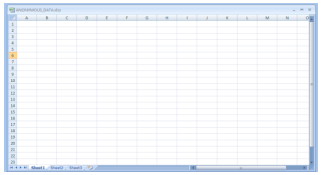
Make substitutions to reduce high resolution to 2nd level (4 digits): done!



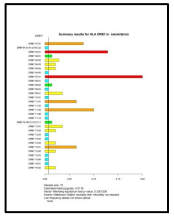
Data formating

```
ANON_1  DRB1*1404,DRB1*1404
ANON_2  DRB1*03010101/03010102/030108,DRB1*1104
ANON_3  DRB1*03010101/03010102/030108,DRB1*1104
ANON_4  DRB1*15010101/15010102,DRB1*1502
ANON_5  DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_6  DRB1*0101,DRB1*1602
ANON_7  DRB1*03010101/03010102/030108,DRB1*0101
ANON_8  DRB1*03010101/03010102/030108,DRB1*0101
ANON_9  DRB1*010101/0107/0122,DRB1*07010101/07010102/070102
ANON_10 DRB1*0101,DRB1*110101/110108
ANON_11 DRB1*07010101/07010102,DRB1*07010101/07010102
ANON_12 DRB1*07010101/07010102,DRB1*1112
ANON_13 DRB1*07010101/07010102,DRB1*1602
ANON_14 DRB1*07010101/07010102,DRB1*0402
ANON_15 DRB1*1104,DRB1*15010101/15010102
ANON_17 DRB1*07010101/07010102,DRB1*1302
ANON_18 DRB1*03010101/03010102/030108,DRB1*1302
ANON_19 DRB1*1602,DRB1*0403
ANON_20 DRB1*0101,DRB1*110101/110108
ANON_21 DRB1*07010101/07010102,DRB1*0901|DRB1*07010101/07010102,DRB1*0906|DRB1*07010101/07010102,DRB1*0909
ANON_22 DRB1*0405,DRB1*1001
ANON_23 DRB1*1104,DRB1*0901
ANON_24 DRB1*07010101/07010102,DRB1*1506
ANON_25 DRB1*03010101/03010102/030108,DRB1*15010101/15010102
ANON_27 DRB1*1104,DRB1*1104
ANON_28 DRB1*1104,DRB1*1104
ANON_29 DRB1*1104,DRB1*1104
ANON_30 DRB1*07010101/07010102,DRB1*110101/110108
ANON_31 DRB1*0101,DRB1*1302
ANON_32 DRB1*07010101/07010102,DRB1*1301
ANON_33 DRB1*1301,DRB1*0404
ANON_35 DRB1*03010101/03010102/030108,DRB1*0402
ANON_36 DRB1*15010101/15010102,DRB1*07010101/07010102
ANON_37 DRB1*1104,DRB1*0803
ANON_38 DRB1*1502,DRB1*0101
ANON_39 DRB1*1106,DRB1*120101/1206/1210/1217
ANON_40 DRB1*07010101/07010102,DRB1*080101/080103
ANON_41 DRB1*1301,DRB1*0802
ANON_42 DRB1*07010101/07010102,DRB1*0101
ANON_43 DRB1*03010101/03010102/030108,DRB1*07010101/07010102
ANON_44 DRB1*07010101/07010102,DRB1*0901
ANON_45 DRB1*03010101/03010102/030108,DRB1*1601
ANON_46 DRB1*03010101/03010102/030108,DRB1*0101
ANON_47 DRB1*07010101/07010102,DRB1*1301|DRB1*07010101/07010102,DRB1*1116
ANON_48 DRB1*07010101/07010102,DRB1*07010101/07010102
```

Make substitutions to reduce resolution to 2nd level (4 digits): file downloaded.

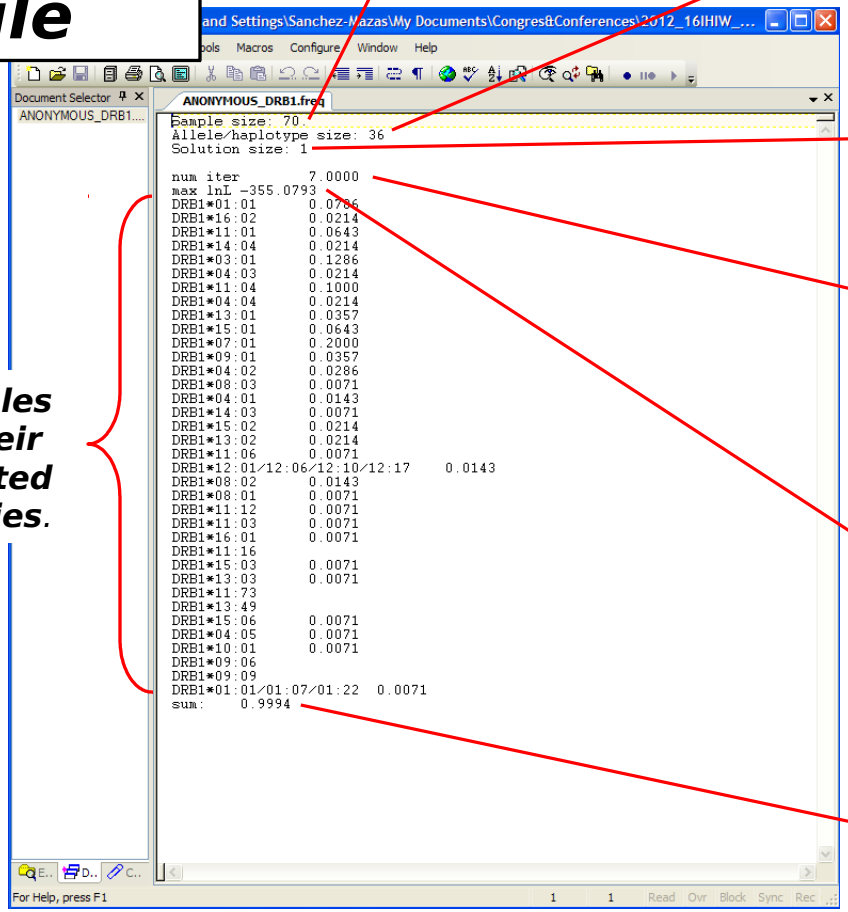


?



Estimating frequencies

Results file



Number of individuals sampled

Number of different alleles or haplotypes in sample

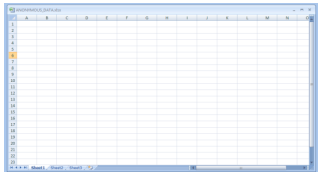
Number of solutions with maximum likelihhoods obtained to estimate frequencies: *should ideally be 1. If greater than 1, may indicate deviation to HWE (to check).*

Number of iterations of the algorithm to reach maximum likelihood: *ideally low number (less than 10 for one-locus allele frequencyies estimation, but greater for ambiguous or multi-locus haplotype frequencies estimation).*

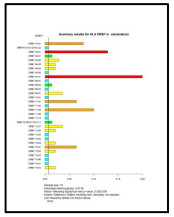
List of alleles with their estimated frequencies.

Value of the maximum likelihood: reached: *always negative.*

Sum of allele frequencies: *should be 1 but may be slightly lower at the 4th and following decimals.*

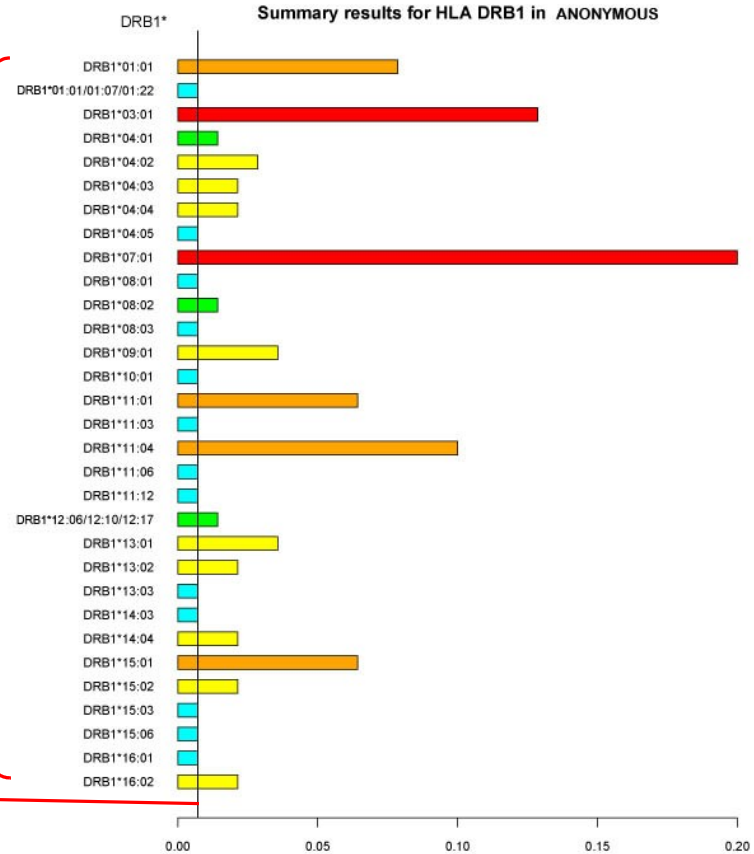


?



Estimating frequencies

Generated graph

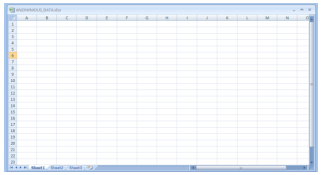


List of alleles with their estimated frequencies.

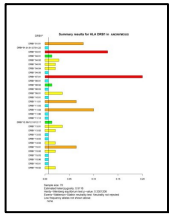
Frequency threshold: allele frequencies below this line represent less than 1 allele copy in sample (see next slide)

Summary statistics.

Sample size: 70
Estimated heterozygosity: 0.9116
Hardy–Weinberg equilibrium test p–value: 0.3301206
Ewens–Watterson–Slatkin neutrality test: Neutrality not rejected
Low frequency alleles not shown above:
none



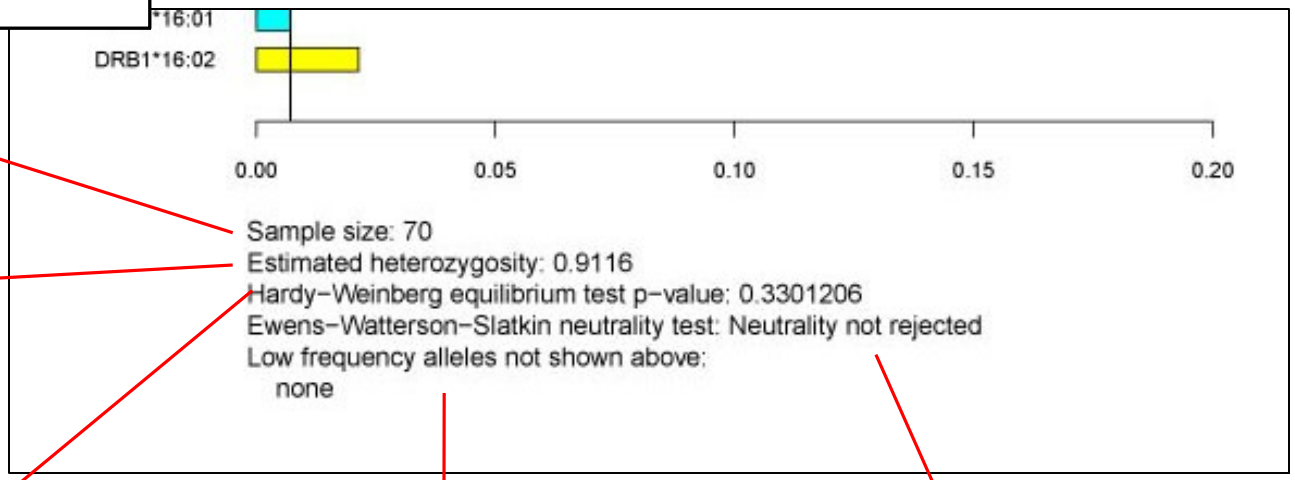
?



Estimating frequencies

Generated statistics

Summary statistics.



Number of individuals sampled (original data)

Estimated heterozygosity (requires HWE)

Result of Hardy-Weinberg equilibrium test

Low frequency alleles: *alleles with frequencies below threshold (of 1 copy in the sample) are listed here. Such alleles may occurr with ambiguous or multi-locus haplotypic data.*

Result of the neutrality test