How to convert your data into QTL viewer Data:

Step 1: Generate the following files in TAB separated format (no headers)

features.txt

	NAME	DATA TYPE	EXAMPLE	NOTE
			LAYOUT	
1	feature_id	STRING	ENMUSG00000019966	
2	group_id	STRING	ENMUSG00000019966	
3	chrom	CHAR (2 characters)	10	
4	location	FLOAT	100.01563	location is in Mb
5	name	STRING	Kitl	
6	description	STRING	Kit ligand	

Example:

ENSMUSG00000000001	ENSMUSG00000000001	3	108.126713	Gnai3	guanine nucleotide binding protein
ENSMUSG00000000049	ENSMUSG00000000049	11	108.378875	Apoh	apolipoprotein H
ENSMUSG00000000088	ENSMUSG00000000088	9	57.5268828	Cox5a	cytochrome c oxidase subunit Va

NOTE: if genes, group_id = feature_id,
   if pQTL, the group_id will be gene_id while the feature_id will be protein id

markers.txt

	NAME	DATA TYPE	EXAMPLE	NOTE
			LAYOUT	
1	marker_id	STRING	rs6412653	
2	chrom	CHAR (2 characters)	X	
3	location	FLOAT	11.232	location is in Mb

Example:

1_677642	1	0.677642	
rs48774772	9	57.522172	
1_32749486	1	32.7494861	

lod.txt - *dataset of LOD scores in FLOATING point, features (rows) x markers (columns)*
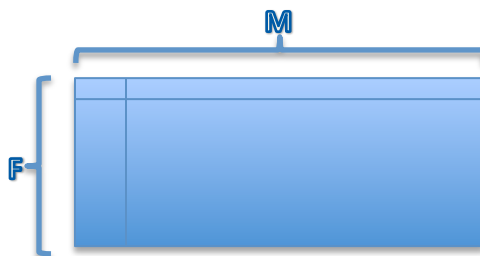
M is the number of markers
F is the number of features

NOTE: ALL values should be of type FLOAT, if there is no value, leave that cell empty

strains.txt

LAYOUT			
NAME	**DATA TYPE**	**EXAMPLE**	**NOTE**
1 strain_id	STRING	A	
2 name	STRING	A/J	
3 description	STRING		

Example:

A	A/J
B	C57BL/6J
C	129S1/SvImJ

coef_[strain_id].txt

The easiest way to import the coefficient data is to create one file per strain (N). Rows would be features (F) and columns would be markers (M).
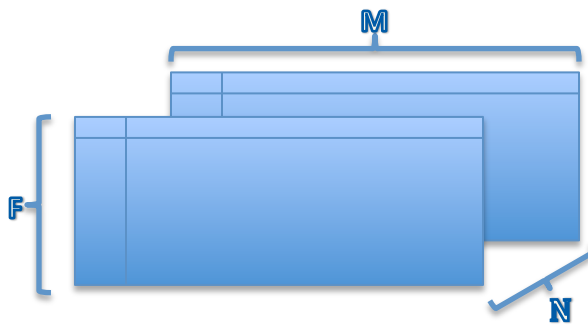
NOTE: ALL values should be of type FLOAT, if there is no value, leave that row empty

F is the number of features
M is the number of markers
N in the number of strains

Example: If you have 2 strains with IDs of 'A' and 'B', you would have 2 files called 'coef_A.txt' and 'coef_B.txt'. The reason for this format is for making the importing from the scripts easier.

samples.txt

	NAME	DATA TYPE	EXAMPLE	NOTE
1	sample_id	STRING	Mouse142	
2	name	STRING	Mouse 142	
3	description	STRING		

LAYOUT

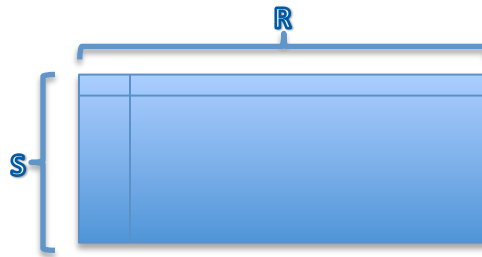## Example:

Mouse142	Mouse 142	
4sd56	Mouse #1	My favorite mouse
abds	ABDS	Unique mouse in the experiment

factors.txt - *factors (examples are sex, diet, tissue, etc)*

	NAME	DATA TYPE	EXAMPLE	NOTE
1	factor_id	STRING	tissue	
2	Name	STRING	Tissue	
3	description	STRING	The type of tissue	

LAYOUT

Example:

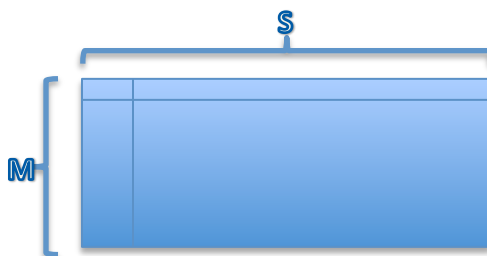|		
Sex	Sex	M,F are present
Diet	Diet	Low fat, high fat, chow diets are used

phenotypes.txt – *phenotypes with samples (rows) and factors (columns)*

R is the number of factors
S in the number of samples

Example: 2 factors (sex, diet), 6 samples

F	LowFat
F	HighFat
F	Chow
M	Chow
M	HighFat
M	LowFat

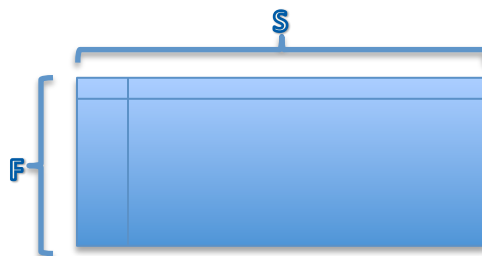## genotypes.txt – *genotypes, markers (rows) and samples(columns)*

R is the number of factors
S in the number of samples

Example: 6 samples, many markers (omitted for brevity)

CC	AA	BB	AB	CC	AB
CB	AA	CB		CB	BB
CC	AA	CB	AB	CC	BB
BC	AC	CB	AB	CC	AB
CC	AA	CB	AB	CC	BB

expression.txt – *expression values, features (rows) and samples(columns)*

F is the number of features
S in the number of samples

Example: 6 samples, many features (omitted for brevity)

-0.26623	1.7633	1.23753	0.21211	-0.12122	0.78667
0.876322	1.12122	1.87212		1.32322	1.82162
1.98872	0.29903	-0.672833	0.21642	1.24422	-1.93722
0.23232	-0.233223	0.323232	0.217652	-1.232332	0.828821
1.7832786	1.872563	-1.23222	-0.81221	0.165212	1.34651

NOTE: ALL values should be of type FLOAT, if there is no value, leave that cell empty

Step 2: Run the following Python script:

qtl_viewer/utils/txt_to_hdf5.py

PARAMETER	REQUIRED	DESCRIPTION	NOTE
-o, --out	Yes	Output file	
-d, --dataset	Yes	Name of the dataset	Example: adipose
-f, --features	Yes	Name of features file	
-l, --lod	Yes	Name of lod scores file	
-m, --markers	Yes	Name of markers file	
--dfA	Yes	Degrees of freedom autosomal	
--dfX	Yes	Degrees of freedom, X chrom	
-c, --csv	No	Use csv file instead of TSV	Not fully tested
--datasetname	No	The display name of the dataset	Adipose Tissue instead of adipose
--coef	No	Name of the ALL coefficient files	*_[strain_id].txt works nice
--strains	No	Name of the strains file	
--samples	No	Name of the samples file	
--factors	No	Name of the factors file	
--phenotypes	No	Name of phenotypes file	
--genotypes	No	Name of genotypes file	
--expression	No	Name of expression data file	

Example:

```
python qtl_viewer/utils/txt_to_hdf5.py
    --out proteomics.h5
    --dataset proteomics_do192
    --datasetname "Proteomics 192"
    --dfA 7
    --dfX 14
    --features proteomics_features.txt
    --lod proteomics_lod.txt
    --markers proteomics_markers.txt
    --strain proteomics_strains.txt
    --coef proteomics_coef_*.txt
    --factors proteomics_factors.txt
    --phenotypes proteomics_phenotypes.txt
    --genotypes proteomics_genotypes.txt
    --samples proteomics_samples.txt
    --expression proteomics_expression.txt
```