

HDF5 Data Structure in QTL Viewer

/[dataset]

Attributes of dataset

ATTRIBUTE	DATA TYPE	EXAMPLE	NOTE
name	STRING	Our Big Experiment	Name of the experiment
dfA	NUMERIC	7	Degrees of freedom, autosomal
dfX	NUMERIC	14	Degrees of freedom, X chromosome

/features – dataset of features (gene, transcript, etc) ids (in our case ensembl ids),

	NAME	DATA TYPE	EXAMPLE	NOTE
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	

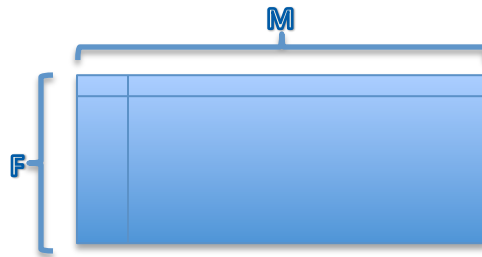
NOTE: if genes, group_id = feature_id,
if pQTL, the group_id will be gene_id while the feature_id will be protein id

/markers – dataset of markers (SNPs)

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

/lod/lod – dataset of LOD scores in FLOATING point, features (rows) x markers (columns)

Example: **F** is the number of features, **M** is the number of markers



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

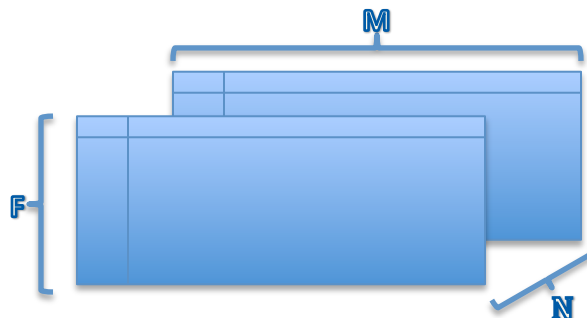
FOR THE COEFFICIENT EFFECT PLOT (not required)

/coef/strains – dataset of strains (example would be CC founders)

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

/coef/coef – dataset of coef scores in FLOATING point, features x strains x markers

Example: **F** is the number of features, **M** is the number of markers, **N** in the number of strains



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

FOR THE FACTORIAL VIEWER (NOT REQUIRED)

/samples - dataset of samples (mice)

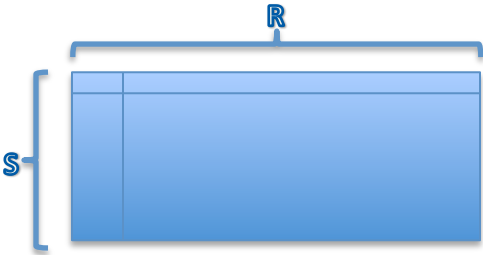
	NAME	DATA TYPE	EXAMPLE	NOTE
1	sample_id	STRING	Mouse142	
2	name	STRING	Mouse 142	
3	description	STRING		

/phenotypes/factors - dataset of 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	

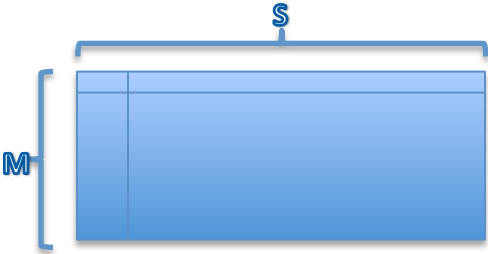
/phenotypes/phenotypes - dataset of phenotypes samples (rows) x factors (columns)

Example: **R** is the number of factors, **S** in the number of samples



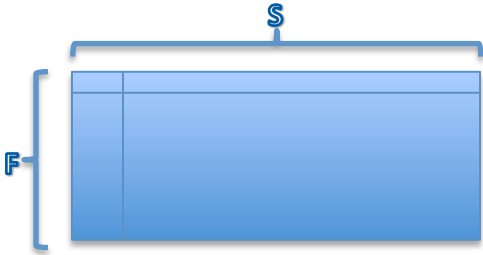
/genotypes/genotypes - dataset of genotypes markers (rows) x samples (columns)

Example: **R** is the number of factors, **S** in the number of samples



/expression/expression - dataset of expression values, features (rows) x samples (columns)

Example: **F** is the number of features, **S** in the number of samples



Example structure of HDF5

```
/proteomics_dol92 Item(path='/proteomics_dol92', shape=())
/proteomics_dol92/coef Item(path='/proteomics_dol92/coef', shape=())
/proteomics_dol92/coef/coef Item(path='/proteomics_dol92/coef/coef', shape=(6716, 8, 2146))
/proteomics_dol92/coef/strains Item(path='/proteomics_dol92/coef/strains', shape=(8,))
/proteomics_dol92/expression Item(path='/proteomics_dol92/expression', shape=())
/proteomics_dol92/expression/expression Item(path='/proteomics_dol92/expression/expression', shape=(6716, 192))
/proteomics_dol92/features Item(path='/proteomics_dol92/features', shape=(6716,))
/proteomics_dol92/genotypes Item(path='/proteomics_dol92/genotypes', shape=())
/proteomics_dol92/genotypes/genotypes Item(path='/proteomics_dol92/genotypes/genotypes', shape=(2146, 192))
/proteomics_dol92/lod Item(path='/proteomics_dol92/lod', shape=())
/proteomics_dol92/lod/lod Item(path='/proteomics_dol92/lod/lod', shape=(6716, 2146))
/proteomics_dol92/markers Item(path='/proteomics_dol92/markers', shape=(2146,))
/proteomics_dol92/phenotypes Item(path='/proteomics_dol92/phenotypes', shape=())
/proteomics_dol92/phenotypes/factors Item(path='/proteomics_dol92/phenotypes/factors', shape=(2,))
/proteomics_dol92/phenotypes/phenotypes Item(path='/proteomics_dol92/phenotypes/phenotypes', shape=(192, 2))
/proteomics_dol92/samples Item(path='/proteomics_dol92/samples', shape=(192,))
```