## 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

**M**

**F**

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

**FOR THE COEEFICIENT 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

**M**

**F**

**N**

***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

**R**

**S**

**/genotypes/genotypes –** dataset of genotypes markers (rows) x samples (columns)

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

**S**

**M**

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

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

**S**

**F**

## Example structure of HDF5

/proteomics\_do192 Item(path='/proteomics\_do192', shape=())

/proteomics\_do192/coef Item(path='/proteomics\_do192/coef', shape=())

/proteomics\_do192/coef/coef Item(path='/proteomics\_do192/coef/coef', shape=(6716, 8, 2146))

/proteomics\_do192/coef/strains Item(path='/proteomics\_do192/coef/strains', shape=(8,))

/proteomics\_do192/expression Item(path='/proteomics\_do192/expression', shape=())

/proteomics\_do192/expression/expression Item(path='/proteomics\_do192/expression/expression', shape=(6716, 192))

/proteomics\_do192/features Item(path='/proteomics\_do192/features', shape=(6716,))

/proteomics\_do192/genotypes Item(path='/proteomics\_do192/genotypes', shape=())

/proteomics\_do192/genotypes/genotypes Item(path='/proteomics\_do192/genotypes/genotypes', shape=(2146, 192))

/proteomics\_do192/lod Item(path='/proteomics\_do192/lod', shape=())

/proteomics\_do192/lod/lod Item(path='/proteomics\_do192/lod/lod', shape=(6716, 2146))

/proteomics\_do192/markers Item(path='/proteomics\_do192/markers', shape=(2146,))

/proteomics\_do192/phenotypes Item(path='/proteomics\_do192/phenotypes', shape=())

/proteomics\_do192/phenotypes/factors Item(path='/proteomics\_do192/phenotypes/factors', shape=(2,))

/proteomics\_do192/phenotypes/phenotypes Item(path='/proteomics\_do192/phenotypes/phenotypes', shape=(192, 2))

/proteomics\_do192/samples Item(path='/proteomics\_do192/samples', shape=(192,))