

Mouse Pipelines Phenotyping Data Resource

<ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/>

This document describes the content of each of the directories found at the above location and how to navigate the site in order to locate the data of interest.

Subdirectories

Baselines

Here you will find spreadsheets listing IDs of specimens in the baselines for all pipelines. The files in the *Index* directory will then allow you to determine which file(s) within the *Pipelines* directory contain your measurements of interest.

Please note that the baselines here contain all wildtype animals on a specified genetic background processed through the indicated experimental pipeline. You may need to filter this data further based on relevant metadata.

If you are using these baselines in a comparison with mutant data, the correct baseline to use is indicated in the filename of the population. Examples of constructing a dataset in this way are detailed in the Examples file in the Documentation subdirectory.

Colonies

This directory contains any additional related documentation concerning each colony that has been phenotyped. Generally, these may be one or more of the following:

Datasheet/Technical datasheet – genotyping and QC information.

Mouse Technical datasheet – information on genetic background, breeding and phenotype observations.

Eye histopathology report – a subset of lines will have reports detailing findings following H&E staining and pathology assessment of 9 sections from 3 male mutant pairs of eyes.

CMHD pathology report – a subset of lines will have reports detailing histopathology analysis following H&E staining of sections from full necropsy tissues collected at the end of the pipeline from 2 male and 2 female animals per line analysed.

Note that not all documents are available for every line.

Documentation

These files contain details of the pipelines utilised by the phenotyping programme, descriptions and standard operating procedures (SOPs) for the assays performed and information regarding statistics used to analyse the generated data, as well as a glossary document explaining the terminology used.

Many of these documents are also available in other subdirectories, for example, descriptions of pipelines are also within each pipeline folder in the *Pipelines* subdirectory.

Heat Map

The heatmap is a visualisation of the phenotypic calls of significance made for each assay an allele was tested for.

HeatMap.xlsx vertically lists each allele processed and indicates if the outcome of the test listed horizontally was considered significant or not using a colour scheme and standardised set of rules. The help file within this directory provides details on this.

HeatMapCalls.xlsx lists the specific parameter(s) from a test that were deemed significant and also provides the corresponding Mammalian Phenotype (MP) term.

MP Terms – Mammalian Phenotype Ontology Term

(http://www.informatics.jax.org/vocab/mp_ontology); a method for organizing and classifying phenotyping information from mammalian species. This allows for easy comparison of data from difference centres and sources to facilitate discoveries of genes involved in different phenotypes.

MA Terms – Adult Mouse Anatomy Term (http://www.informatics.jax.org/vocab/gxd/ma_ontology); a method for easily classifying the area of the mouse affected by an observation.

Index

The files within this directory show where to find the measurement file(s) for your specimens of interest.

You may have a list of animals from a baseline file, or a population file. To find the file containing the raw data from the test you are interested in, you will need to look up those animal IDs in the index files, which will point you to the subdirectories and the specific file where data collected from those animals can be found.

To make this more efficient, an 'index of indexes' file is also present which lists the specific pipeline(s) represented in each of the main index files.

Examples of constructing a dataset in this way are detailed in the Examples file in the Documentation subdirectory.

Pipelines

This is where the raw experimental data is found, primarily in csv formatted files. The directory is divided into folders for each pipeline; details of the pipelines can be found within each folder as well as in the Documentation subdirectory.

Once inside a pipeline folder, you will find further folders for each test performed on that pipeline. Within these may be one or more folders, depending on the different data types generated for that test.

Use the index files in the *Index* directory to find out which file will contain the data for the specimens you are interested in.

Examples of how to do this are detailed in the Examples file in the Documentation subdirectory.

Populations

Here you will find files listing the IDs of the groups of animals for each allele tested on the indicated pipeline.

The file for your gene of interest will give you the list of mutant animals tested plus the local controls (the controls processed during the same weeks as the mutants). The unique identifiers for each animal are the Mouse Name or Mouse Barcode.

The raw data for each test can be accessed via the Pipelines directory. The index files will allow you to determine which file(s) contain your measurements of interest. If you would also like to assemble a baseline, the population file should specify a baseline ID for a file that can be found within the Baselines directory.

Examples of constructing a dataset in this way are detailed in the Examples file in the Documentation subdirectory.