



An International Centre for Mouse Genetics



Mammalian Genetics Unit

EuroPhenome and (mouse) phenome data resources

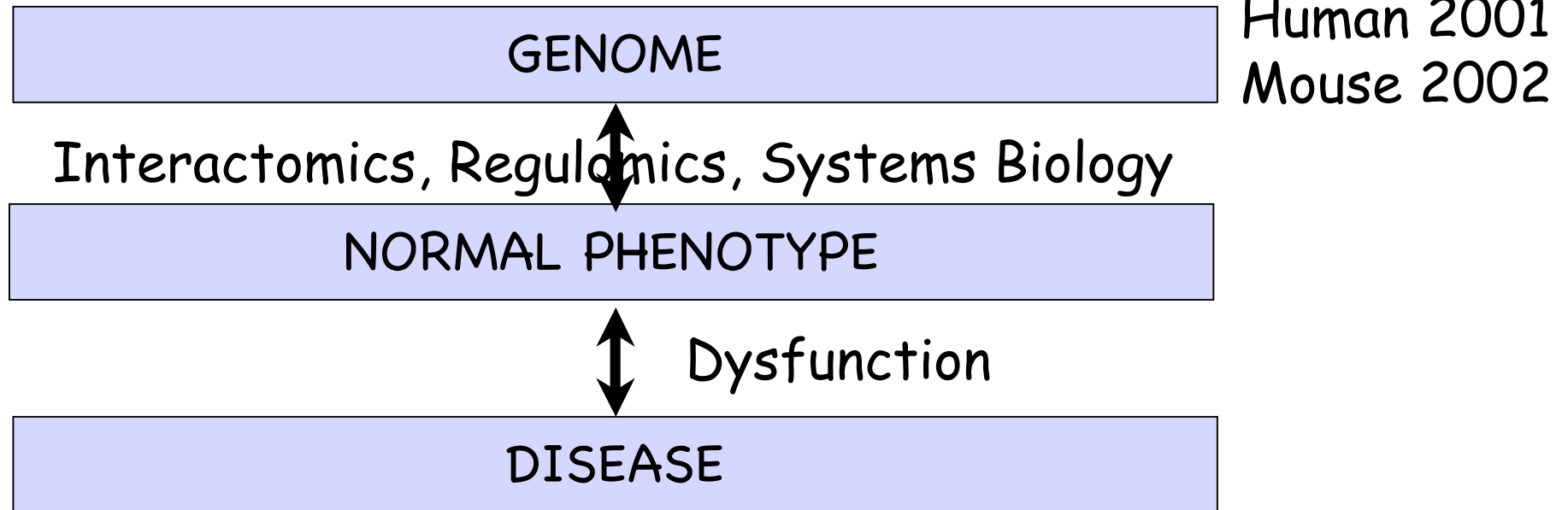
John Hancock

MRC Harwell

Plan

- The context – high throughput genomics and phenomics
- History of mouse mass phenomics
- EuroPhenome
- EuroPhenome and data policy

From Genome to Phenotype to Disease

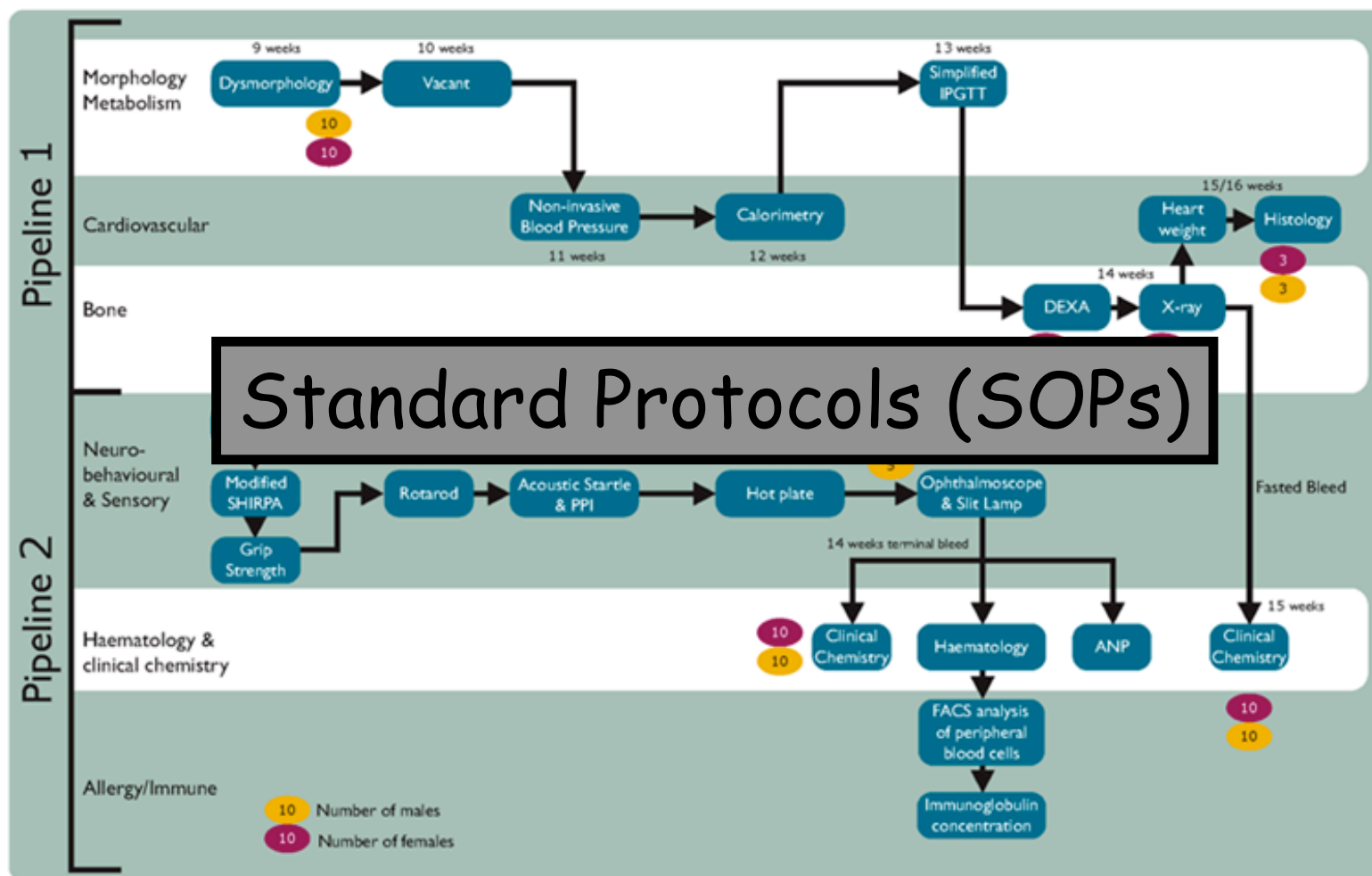


Mouse Phenomics: Systematic Gathering of Phenotype Data on the Laboratory Mouse

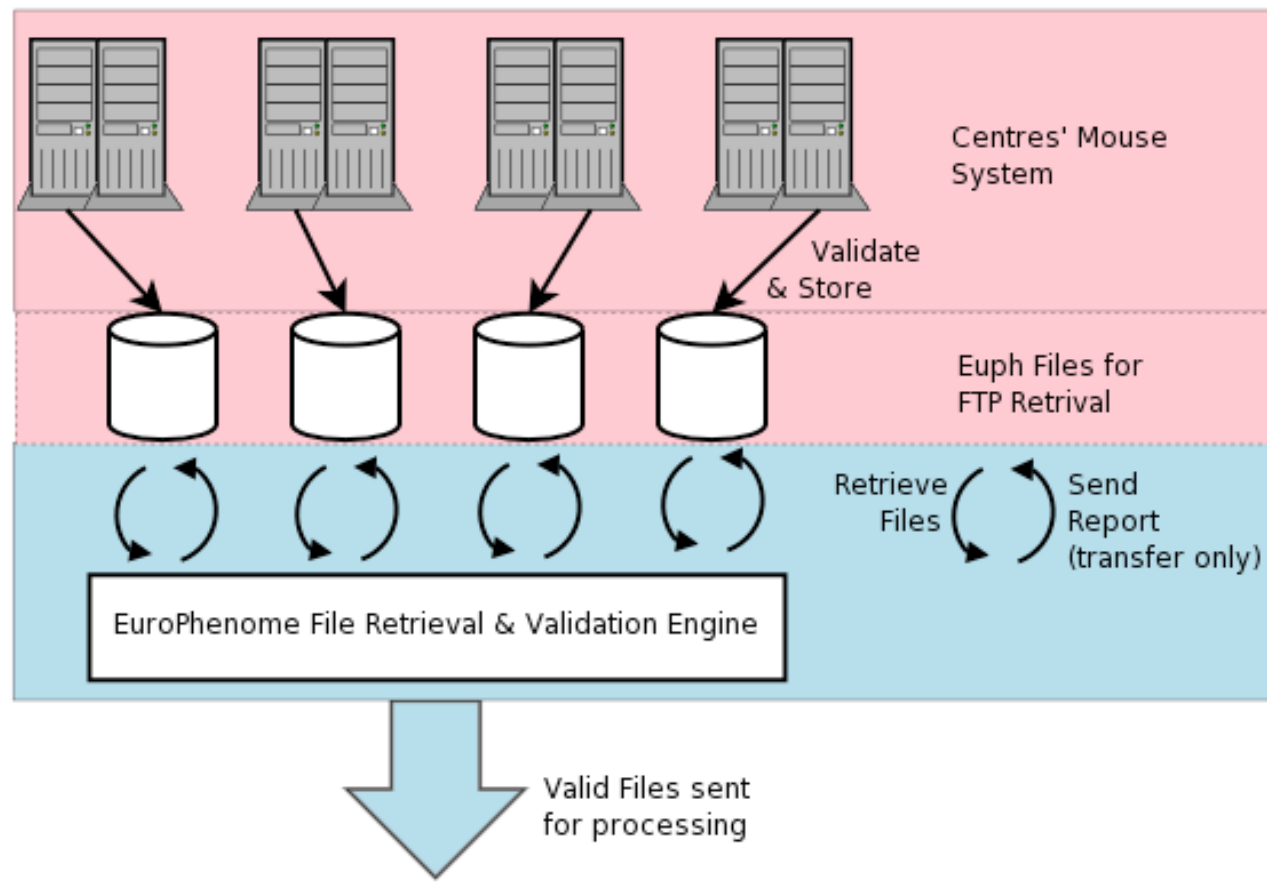
- Mouse Phenome Project (Jax) - May 1999-Present
 - Collect phenotype data on inbred mouse lines
 - Many labs, each collecting a particular measurement
- EUMORPHIA - 2002-2006
 - Standardise methods for collecting same measurement in different labs
 - EMPReSS Database of SOPs
 - EuroPhenome database of inbred strain phenotypes
- EUMODIC - 2007-2011
 - Collect data on 500 EUCOMM knockout mouse lines
 - Data in EuroPhenome
- IMPC (International Mouse Phenotyping Consortium)
 - Two meetings so far - 2007, 2008



EMPreSS slim primary phenotyping screen



Data Capture in EUMODIC



EuroPhenome Features

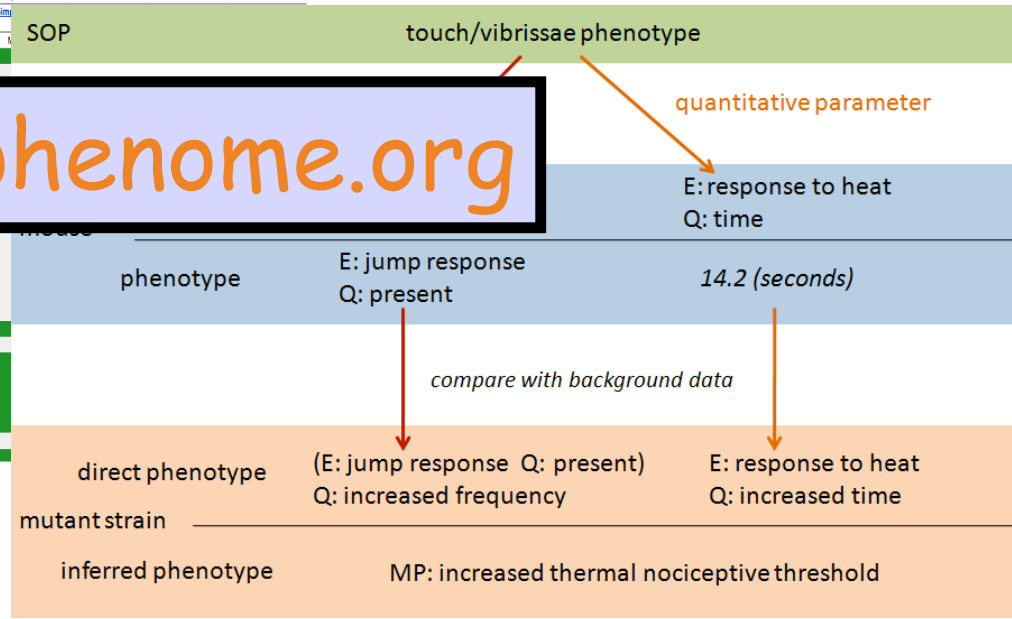
Annotations

Gene	Internal Id	Procedure	Parameter	Term	Data
A42	A42m1Wtsi	Clinical Chemistry	Glucose	increased circulating glucose level	Graph
A42	A42m1Wtsi	Simplified	Blood glucose concentration	impaired glucose tolerance	Graph
A42b	A42m1Wtsi	Clinical Chemistry	Glucose	decreased circulating glucose level	Graph

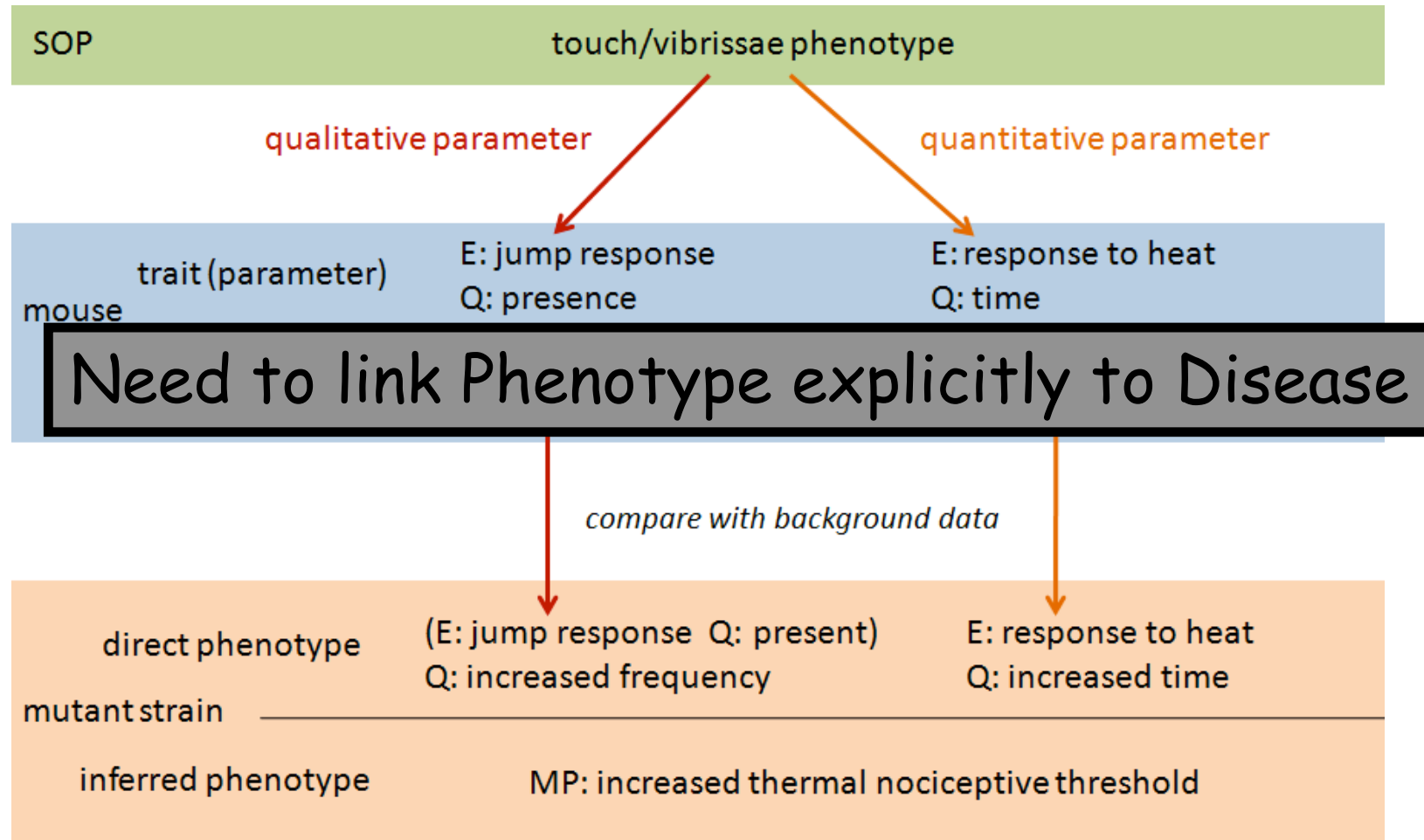
Controls

Choose Pipeline of Interest: 1 2
 Sex: M+F
 P Range: 0.0001
 Effect Size +/-: 10
 Legend: [Green] [Red]

www.europhenome.org



Coding data using EQ & MP in EuroPhenome



EUMODIC Data Release

- All data generated using EUMODIC funding must be submitted to EuroPhenome
- The data that is submitted is defined
- It will be released publicly as soon as it has been processed into the database
- Centres can update (“finish”) their data by uploading a modified version
- Data can only be excluded if they fail project-wide QC
- Data cannot be excluded because they “don’t look right”
- NB: This is not DNA sequence!!



EuroPhenome in Future

- Aim is to become a community resource
 - Data exchange with RIKEN (July 09 Meeting)
 - Feed MP annotations to Jax
 - Accept data from other providers (TCP)
 - Local installations (standard data format & content)
- Need for a formal data release policy
- Take account of EUMODIC policy and build on it
 - Compatible with Fort Lauderdale principles?
- Database will need a SAB
 - Advice
 - Criticism
 - Meeting before end 2009 with main stakeholders



Acknowledgements

- EUMORPHIA, EUMODIC
 - Steve Brown
 - Hilary Gates
 - Scientific Partners

- EuroPhenome, EMPReSS
 - Ann-Marie Mallon
 - Hugh Morgan
 - Tim Beck
 - Andy Blake
 - Simon Greenaway
 - All who have and are contributing data
 - All who have contributed to SOP development

Funders:

EU (FP5, FP6)
MRC



Sanger Institute Sequence Data Release Policy

- * Completed sequences are submitted to the EMBL/Genbank/DDBJ database as soon as they are finished and annotated. The finished sequences and current copies of the corresponding EMBL files are available from our ftp site. These can also be accessed via the appropriate project web pages.
- * Unfinished, assembled sequences are released on to our ftp site as they become available, automatically every night. In order to screen out the most common artefacts we only output assembled contigs greater than 1 kb. This data is a pre-release and is clearly marked on our site as incomplete data, subject to error and possibly even to misidentification. As such it is used at your own risk. This instant release in no way affects the quality of our finished data submitted as in 1.

Unfinished sequences from our ftp site are also collected regularly by the EBI and entered into the HTG division of the EMBL/Genbank/DDBJ database.

Unfinished sequences should be regarded as unpublished data. If you make use of unfinished Sanger Institute sequences, we request you to please read our Guidelines on use of data in publications.

We also provide BLAST servers for both finished and unfinished sequence data.

- * Maps of clones chosen for sequencing are displayed on our Web pages. These displays are updated every night and show the current progress of each clone in the sequencing process. Currently only human maps are available, accessible via the appropriate chromosome page under our Human Genome Project web page.

Again, this information should be regarded as unpublished data, and users should refer to our Guidelines on use of data in publications.

<http://www.sanger.ac.uk/notices/release-policy.shtml>



An International Centre for Mouse Genetics

Sharing Data and Resources for Functional Genomics
Hotel Ponte Sisto Rome 20-22nd May 2009



Points to Consider

- Finished vs unfinished – what is the equivalent for phenotype?
- EuroPhenome, MPD etc \approx GenBank. Local release?
- Guidelines for use of data in publications:

"As well as finished sequences, which are freely available through EMBL/Genbank, we also make unfinished sequence data available. Users of unfinished sequence data from the Sanger Institute should not publish or otherwise disseminate the information without appropriate acknowledgement. The appropriate format is as follows:

"These sequence data were produced by the xxx Sequencing Group at the Sanger Institute and can be obtained from <ftp://ftp.sanger.ac.uk/pub/yyy>"

where "xxx" is the name of the relevant sequencing project, such as "Human Chromosome 22" or "P. falciparum", and "yyy" is the appropriate ftp directory from which the data has been obtained. An http URL may be given in place of the ftp directory if that was the source of the data.

Unfinished sequences should all have been assigned an EMBL/Genbank accession number (in the HTG division). This should be quoted when available.

Please ensure that you use the latest available version of unfinished sequences. In the interval between first unfinished release and finishing, errors are continuously being discovered and corrected."

