Module 5: Mining phenotype databases, to identify mouse models of clinical relevance.

Aims

- Introduction to the International Mouse Phenotyping Consortium (IMPC).
- Introduction to the statistical analysis and annotation of high-throughput phenotyping data.
- Introduction to the description of mouse phenotypes using the Mammalian Phenotype Ontology.
- To query, interpret and analyse mouse phenotypic data within EuroPhenome to identify phenotypes associated with interesting genes.
- To introduce plans for the IMPC web portal and the MPI2 consortium.
- To identify mouse models with interesting and rare phenotypes and correlate these with human disease.

Introduction

Phenotype data on mouse mutants is primarily captured from publications by curators into the MGI database at the Jackson Laboratory. This data is primarily comprised of post composed phenotype ontology annotations to mutant alleles. In recent years, the generation of knockouts from the International knockout mouse consortium (IKMC) has enabled the community to perform high throughout standardised phenotyping. The major pilot project to do this is called EUMODIC (www.eumodic.org) and the primary data is captured and analysed post capture in the EuroPhenome (www.europhenome.org) database. This data also gets ontological annotations, but primarily on the basis of automatic analysis rather than scientifically curated peer reviewed results. In addition two other portals exist with phenotyping data; The Sanger Mouse Genetics Portal (www.sanger.ac.uk/mouseportal) and the KOMP K312 portal (www.kompphenotype.org). Under the umbrella of IMPC (described below),
all of these data portals will be coordinated into one portal – www.mousephenotype.org.

**IMPC -** [http://www.mousephenotype.org](http://www.mousephenotype.org)

**International Mouse Phenotyping Consortium**

The International Mouse Phenotyping Consortium (IMPC) comprises a group of major mouse genetics research institutions along with national funding organisations formed to address the challenge of developing an encyclopedia of mammalian gene function.

The IMPC envisages a ten year programme to undertake a broad-based, systematic genome-wide phenotyping project of knockout mice generated from the embryonic stem cell mutant resources developed by the International Knock-out Mouse Consortium (IKMC). Each mutant line will undergo a broad suite of high-throughput tests to identify developmental, anatomical, physiological, behavioural and pathological phenotypes.

It is anticipated that this landmark programme will produce a paradigm shift in our understanding of basic molecular, cellular and systems biology, as well as feed the biopharmaceutical pipeline by enhancing our understanding of the genetic bases for disease.
Goals

- Establish a world-wide consortium of mouse centres with capacity and expertise for large-scale primary phenotyping

- Establish a world-wide consortium of mouse production centres to generate germ line transmission of targeted knockout mutations in embryonic stem cells for all known and predicted mouse genes

- Test each mutant mouse line (5,000 mouse lines in the first 5 years, and ultimately up to 20,000) through a broad based primary phenotyping pipeline in all the major adult organ systems and most areas of major human disease

- Through this activity and employing data annotation tools, systematically aim to discover and ascribe biological function to each gene, driving new ideas and underpinning future research into biological systems

- Establish collaborative "networks" with specialist phenotyping consortia or laboratories, providing standardized secondary phenotyping that enriches the primary dataset, and end-user, project specific tertiary level phenotyping that adds value to the mammalian gene functional annotation and fosters hypothesis driven research

- Provide a centralized data centre and portal for free, unrestricted access to primary and secondary data by the scientific community, promoting sharing of data, genotype-phenotype annotation, and the development of open source data analysis tools;
EuroPhenome - (http://www.EuroPhenome.org/)
Repository of high-throughput phenotyping data

The EuroPhenome project (http://www.europhenome.org) is a comprehensive resource for raw and annotated high throughput phenotyping data arising from projects such as EUMODIC. EUMODIC is gathering data from the EMPReSSslim pipeline (http://www.empress.har.mrc.ac.uk/) which is performed on inbred mouse strains and knock-out lines. The data is generated in four clinics across Europe (WTSI; ICS; HMGU and MRC Harwell). In addition EuroPhenome is capturing data from the CMHD in Canada.

The EuroPhenome interface allows the user to access the data via the phenotype or genotype. It also allows the user to access the data in a variety of ways, including graphical display, statistical analysis and access to the data via web services. The raw phenotyping data captured in EuroPhenome is annotated by an annotation pipeline which automatically identifies statistically different mutants from the appropriate baseline and assigns ontology terms for that specific test. Mutant phenotypes can be quickly identified using two EuroPhenome tools: PhenoMap, a graphical representation of statistically relevant phenotypes and mining for a mutant using ontology terms. To assist with data definition and cross-database comparisons, phenotype data is annotated using combinations of terms from biological ontologies.
Current Data Status

<table>
<thead>
<tr>
<th>Mutant Strains</th>
<th>502</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbred Strains</td>
<td>44</td>
</tr>
<tr>
<td>Mice</td>
<td>30,276</td>
</tr>
<tr>
<td>Data Points</td>
<td>9,643,237</td>
</tr>
<tr>
<td>Annotations</td>
<td>2,505</td>
</tr>
</tbody>
</table>

Other phenotyping data portals

Sanger Mouse Genetics Portal

The approaches that are being developed in IMPC build on the efforts of a number of pilot programmes around the world such as the EUMODIC programme and the MGP programme. Projects such as the MGP programme at the WTSI also have web portals (http://www.sanger.ac.uk/mouseportal/).
KOMP K312 Portal (www.kompphenotype.org)

The initial data are coming from a Recovery Act (ARRA) supported project to produce and characterize 312 mutant knockout lines produced primarily from KOMP targeted ES cells. The phenotype screen in adult mice includes LacZ reporter gene tissue distribution in all mutant lines, as well as fertility and viability, necropsy, behavioral phenotypes and a comprehensive analysis of transcriptome changes in homozygous mutant mice for 100 of the mutant lines.
Worked example: Gene Search

1. From EuroPhenome home page search for the Akt2 gene. What procedures have significant annotations in pipeline 2?

---

**STEP 1:**
Load EuroPhenome: http://www.europhenome.org

**STEP 2:**
In the 'Find Gene' Box start typing Akt2 and select that gene.

**STEP 3:**
Hover over the procedures (red boxes) to see their significant parameters and annotations.
2. What is the Standard Deviation of Akt2 Het mice?

3. Click on the ‘Procedure’ pull down list, go to Calorimetry. What parameters show highly significant T-test results?
4: What OMIN diseases are associated with Akt2?

**STEP 7:**
Go back to gene summary page for Akt2.

**STEP 8:**
Click on Associated Human Diseases.
Worked Example: Phenotype Search

EuroPhenome also provides a number of tools to search the dataset for phenotypes of interest.

**STEP 1:**
Type a phenotype of interest e.g. abnormal glucose homeostasis.

**STEP 2:**
View the data for a result of interest by clicking 'data'.
1: For your phenotype of interest can you find a gene with an annotation, if so what is the parameter and procedure measured which resulted in the phenotype?

2. What is the difference in the Phenomap between steps 4 and 5?

3. If you change the P-value of abnormal startle reflex to P= 1.0E-5, how does this change your results?
Worked Example: Data integration and accessibility

EuroPhenome provides the facility to view all the mutant mouse lines associated with procedures in the Phenomap. From EuroPhenome's main page click ‘Phenomap’ (top left). Take some time to explore the information provided by the Phenomap.

All the mouse lines, genes, different pipelines and significant parameters/annotations can be visualised with the Phenomap.

EuroPhenome’s data can be downloaded via their BioMart.

**STEP 1:**
Click on View PhenoMap

**STEP 2:**
Identify a gene and parameter of interest, to download all data click ‘View Data’.

**STEP 1:**
From ‘EuroPhenome Tools’ on the home page click on ‘Access EuroPhenome data with BioMart’
IKMC MartSearch (http://www.knockoutmouse.org/martsearch)

**STEP 1:** From the IKMC martsearch search for the gene Apoe

**STEP 2:** Click on Dataset. Select EuroPhenome Annotations from the pull down.

**STEP 3:** Click on Filters. Select Akt2 from the ‘Line Name’ select box.

**STEP 4:** Click Results
Worked Example: IMPC (http://www.mousephenotype.org/) and IMPRESS (http://www.mousephenotype.org/impress/)

The International Mouse Phenotyping Consortium portal provides links or direct information on knockout mutations, timelines IMPC pipeline and procedures, knockout line status and the gene list. The IMPC pipeline consists of 15 core procedures ranging including Modified SHIRPA, grip strength and DEXA. Each procedure is attributed with a phenotype parameter and a corresponding phenotype annotation. The procedures and parameters are stored in IMPReSS and comprehensive information is provided on the IMPReSS website, linked from the main IMPC home page.
IMPReSS provides an ontology search page where the different pipelines, procedures and parameters are shown for a particular MP term.

**STEP 2:**
From IMPRESS’s home page click on Ontologies.

**STEP 3:**
In the text box type the MP term ‘abnormal circulating glucose’

IMPC holds information on the status of each gene in the gene list. The status is either assigned, genotype confirmed or mouse production in progress.
1. Browse IMPReSS and the gene list for your phenotype and gene of interest.
Tasks

1. Find the significant annotations associated to the gene Akt2. What procedures are significant? Are the majority of the significant procedures associated with male or female mice? Is Akt2 associated with a human disorder?

2. Using the advanced phenotype search. How many genes are significantly associated with abnormal glucose homeostasis and behaviour / neurological phenotype and not abnormal triglyceride level?

3. Using the OMIN Phenotype Mapper search what human gene is associated with Acromegaly? Using the filter by human gene search, what other disorders are associated with this gene? What parameters from pipeline 1 are significantly associated with this gene?

4. Using the BioMart find the difference in the number of parameters in pipeline 2 at a statistical significance score of $10^{-3}$ and $10^{-5}$ for Akt2 (Unclick all attributes apart from Europhenome ID, parameter name and phenotype procedure). What are the procedures associated with these parameters?

5. Using the PhenoMap, how does the result set differ if you change the P range and Effect Size?

6. From the IMPReSS website what procedures test for abnormal body weight in the IMPC Pipeline?
Answers

1. The significant annotations in pipeline 1 are; decreased body weight, decreased total body fat amount, decreased lean body mass, abnormal bone mineralisation, increased lean body mass, decreased total body fat amount and decreased circulating triglyceride level after experiment.

The significant annotations in pipeline 2 are; increased circulating alkaline phosphatase level, increased blood urea nitrogen level, increased circulating amylase level, increased platelet cell number and decreased body weight.

The significant procedures for pipeline 1 are; body weight, DEXA and fasted clinical chemistry.

The significant procedures for pipeline 2 are; clinical chemistry, haematology and body weight.

The significant procedures are seen in both male and female mice equally.

Akt2 is associated with Diabetes mellitus, type II.

2. Thirteen genes are associated with abnormal glucose homeostasis and behaviour / neurological phenotype and not abnormal triglyceride level

3. The human gene is associated with Acromegaly is GNAS.

The other disorders associated with GNAS are; Pseudopseudohypoparathyroidism, ACTH-independent macronodular adrenal
hyperplasia, McCune-Albright syndrome, Prolonged bleeding time, brachydactyly and mental retardation, Pseudohypoparathyroidism Ib, Pseudohypoparathyroidism Ia, Osseous heteroplasia, progressive and Pseudohypoparathyroidism Ic.

The parameters significantly associated with GNAS are; increased body weight, decreased systematic arterial systolic blood pressure, increased lean body mass, increased circulating free fatty acid level, increased circulating glycerol level and increased heart weight.

4. The difference in the number of parameters in pipeline 2 at a statistical significance score of $10^{-3}$ and $10^{-5}$ for Atk2 is 5.

The procedures associated with these parameters are; Clinical Chemistry and Haematology.

5. n/a

6. Grip strength, calorimetry, IPGTT and Body Weight.