Linking Disease Model Phenotypes to Human Conditions Workshop

Sept. 10-11, 2015

Session 3:
Large Scale High Throughput Analysis of Disease Model Phenotyping Data and Annotation of Gene Function

Chair: Janan T. Eppig
Jackson Laboratory
Bar Harbor, ME 04609

Factors in Developing Models & Capturing Phenotype Data at Large-Scale

- Effective and defined mutagenesis systems

- Systematic and rigorous phenotyping, including
  - shared SOPs
  - standardized phenotyping pipelines
  - appropriate analysis

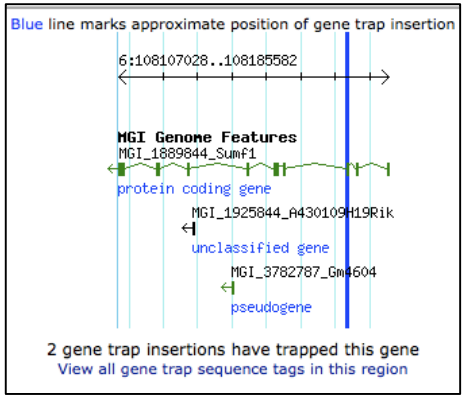
- Integration of data within large-scale efforts

Large-scale mutagenesis for studying gene function & creating disease models in mouse

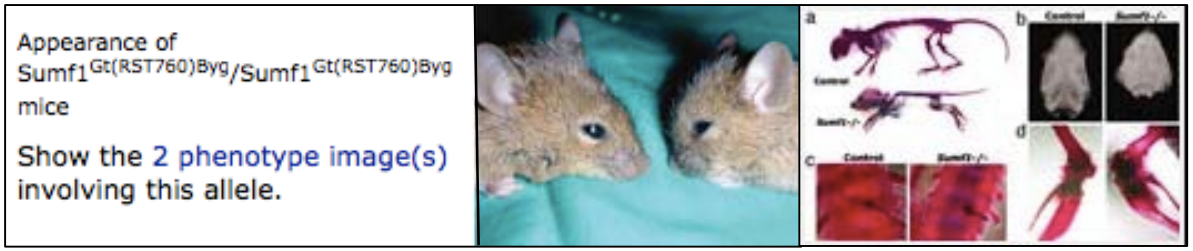
1. International Gene Trap Consortium

(post-project consortium for unified web access to gene traps)

- gene trap insertions in ES cell lines (but varying genetic backgrounds)
- analysis of trap locations not optimal (but now accurate & uniform)
- groups used different trapping strategies, each with their specific biases
- among gene trap groups, created >300,000 gene traps
- scientists continue to obtain mutants of interest from these resources



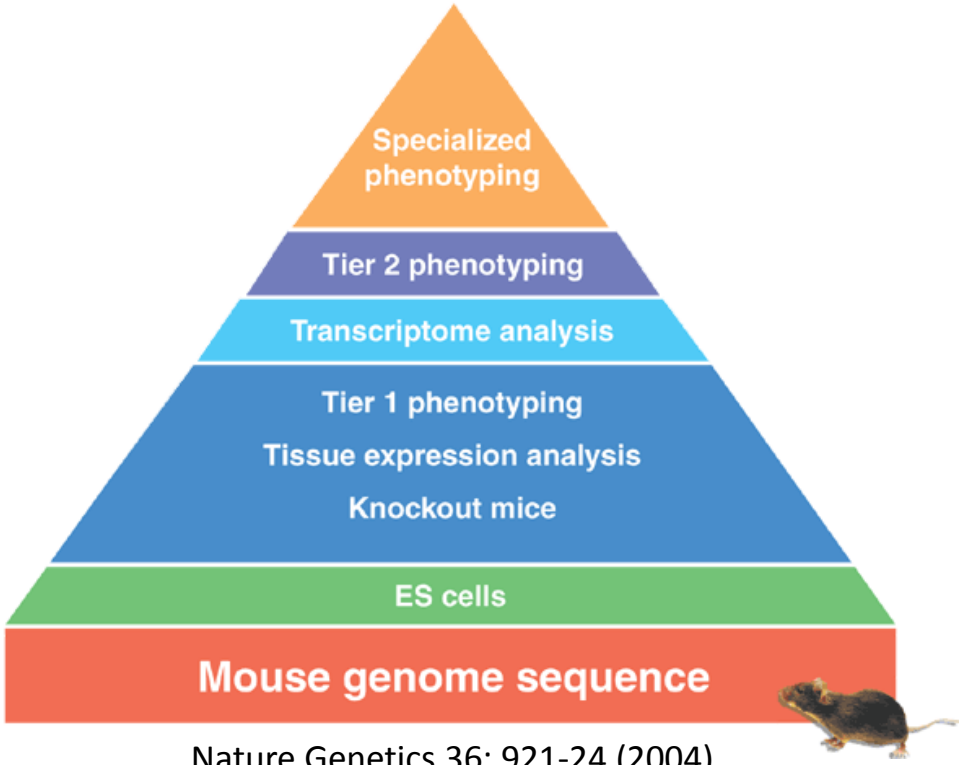
Models Multiple Sulfatase Deficiency (OMIM: 272200)



2. Large-scale ENU mutagenesis (forward genetics)

- phenotype driven screens based on specific systems
- recessive mutants required 3-generation breeding (slow)
- mapping a mutant required crosses & marker analysis*
- >3400 important mutations produced for specific phenotypes
- no joint web presence; but all mutants incorporated into MGI
- ENU mutagenesis is in a revival, aided by sequence-based rapid mapping of new mutations*
- sequencing also identifies “incidental mutations” important for modifier identification & for discovery of new point mutations in other genes.

3. Knocking-out all protein coding genes in the mouse: a systematic approach



Nature Genetics 36: 921-24 (2004)

Ambitious Goals:

- 2006 KOMP (USA) grants for creating kn… N, sim… da (EU…

IKMC: International Knockout Mouse Consortium

- 2011 KOMP2 Phenotyping phase I grants to generate 5000 mouse lines (fr… these thr… , agr…

IMPC: International Mouse Phenotyping Consortium

- 2016 (anticipated) Phenotyping phase II grants to complete the broad-based standard phenotyping on an additional 15,000 mouse lines

Targeted alleles produced by IKMC: 17,351

Mouse lines produced for IMPC phenotyping: 4,843
Mouse lines with IMPC phenotyping complete: 1,384

Other important data exist, though

- From focused laboratory research into specific systems and with a variety of mutation types

- Large-scale projects (at least on the mammalian scale) are, of necessity limited in scope. Focused laboratories will provide needed granular phenotyping

- Integration of these data with large-scale efforts will maximize knowledge

Lessons from these first large-scale phenotype / mutant screens in mice

1. *Organize data with a long-term plan*
   - maintain data access (and continue maintenance of software/hardware and curation/updating of elements (e.g. nomenclature))
   - transfer to a community/infrastructure database

2. *Adopt community standards for nomenclature, strains, IDs of objects screened*
   - makes data accessible and comparable

3. *Annotate/assign phenotype data using standard terms and descriptive metadata.*
   - Use phenotype ontologies
   - Map ontologies to associate terminology among organisms (ex.: MP to HPO to harmonize those phenotype terminologies.