Towards improved data management and sharing in behavioral neuroscience and genomics

Iiris Hovatta, PhD
University of Helsinki
Department of Biosciences / Genetics
http://www.helsinki.fi/neurogenomics, iiris.hovatta@helsinki.fi
Outline

• Background to the genetic and neurobiological basis of anxiety disorders

• Data management and sharing
  – Mouse behavioral data
  – RNA-seq data
  – Genomewide association study (GWAS) data
Research interest: anxiety

**Subjective manifestations:**
- A heightened sense for anxiety, fear, panic, worry

**Objective manifestations:**
- Activation of the autonomous nervous system

Annual prevalence of anxiety disorders in Finland is 5.7%

- Panic disorder: 1.8%
- Generalized anxiety disorder: 1.7%
- Social phobia: 1.1%
- Agoraphobia: 0.8%
- Phobia NOS: 1.0%

Genes + Stressful life events

Pirkola et al. 2005
**Humans**
- Outbred
- Heterogeneous with regard to disease phenotype
- Variable environment
- Large cohorts tedious to collect
- No access to brain tissue

**Inbred mice**
- Inbred
- Controlled environment
- Unlimited numbers can be produced
- Access to any tissue at any time point

---

**Human phenotype**

**Characterization of phenotype in mouse**

**Genetic studies in human population**

**Identification of candidate networks**

**Functional analysis of candidate genes**
Neurobiological Phenotypes in Inbred Mouse Strains

- Low anxiety
- High alcohol preference (alcoholism)
- Low pre-pulse inhibition
- High running wheel activity (attention deficit hyperactivity)
- Resistant to neuronal cell death after seizures (stroke, neuroprotection)

Common environment

- C3H/HeJ
- A/J
- FVB/N
- C57BL/6J
- DBA/2J
- 129S6/SvEvTac
- 129X1/SvJ
- 129S1/SvImJ

- High neurogenesis in enriched environment (neural recovery after injury)
- Low running wheel activity
- Low intrastrain aggression (personality disorders)
- Low alcohol preference
- Susceptible to audiogenic seizures (epilepsy)
- High anxiety (anxiety disorders)
- High pre-pulse inhibition (autism, schizophrenia)

* Sequenced by the public consortium
** Sequenced by Celera

Hovatta and Barlow 2005; Databasing the brain: data to knowledge
Transcriptomic approach to identify anxiety-associated gene networks

Model of stress-induced anxiety

10-day social defeat stress

Behavioral analysis & Gene expression profiling: RNAseq & small RNAseq

Bioinformatic data analysis

Anxiety-associated gene regulatory networks

Data from behavioral tests
Behavioral data

- 4 inbred mouse strains
- N=10-25 per group
- 7 behavioral tests & phenotypes
  - Social aversion
  - Light-dark box
  - Open field
  - Elevated plus maze
  - Novelty induced hypophagia
  - Locomotor activity
  - Weight
- Subset of mice: 2 time points
Data management and sharing issues in behavioral neuroscience

• Data management at individual lab level:
  – Most of the data stored in Excel spreadsheets / SPSS
  – Need for databases: safe storing, search options

• Primary data sharing across labs
  – No open access tradition or requirements from funders and journals
  – Challenges in standardizing behavioral test protocols between labs
    • Equipment
    • Housing conditions
    • Time of testing
    • Age
    • Order of testing if several tests performed on same animals
Genetics of Mouse Behavior: Interactions with Laboratory Environment

John C. Crabbe,¹* Douglas Wahlsten,² Bruce C. Dudek³

Strains of mice that show characteristic patterns of behavior are critical for research in neurobehavioral genetics. Possible confounding influences of the laboratory environment were studied in several inbred strains and one null mutant by simultaneous testing in three laboratories on a battery of six behaviors. Apparatus, test protocols, and many environmental variables were rigorously equated. Strains differed markedly in all behaviors, and despite standardization, there were systematic differences in behavior across labs. For some tests, the magnitude of genetic differences depended upon the specific testing lab. Thus, experiments characterizing mutants may yield results that are idiosyncratic to a particular laboratory.
Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades

Douglas Wahlsten* †, Alexander Bachmanov‡, Deborah A. Finn§, and John C. Crabbe§

*Department of Psychology, University of Alberta, Edmonton, AB, Canada T6G 2E9;
‡Monell Chemical Senses Center, Philadelphia, PA 19104-3308; and
§Department of Behavioral Neuroscience, Oregon Health and Science University, and Portland Research Center and Veterans Affairs Hospital, Portland, OR 97239-3098

LETTERS TO EDITOR

Influences of laboratory environment on behavior

Elissa J. Chesler¹, Sonya G. Wilson¹, William R. Lariviere¹, Sandra L. Rodriguez-Zas² & Jeffrey S. Mogil¹ · ³

¹ Dept. of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, USA
² Dept. of Animal Science, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, USA
³ Dept. of Psychology, McGill University, Montreal, Quebec H3A 1B1, Canada

PERSPECTIVE ABSTRACT

Published online: 30 March 2009 | doi:10.1038/nmeth.1312

Environmental standardization: cure or cause of poor reproducibility in animal experiments?

Sabine Helene Richter³-², Joseph P Garner² & Hanno Würbel²

Standardizing tests of mouse behavior: Reasons, recommendations, and reality☆

Douglas Wahlsten*

Department of Psychology, University of Alberta, P-220 Biological Sciences Building, Edmonton, Alberta, Canada T6G 2E9
Transcriptomic approach to identify anxiety-associated gene networks

Model of stress-induced anxiety

10-day social defeat stress

Behavioral analysis & Gene expression profiling: RNAseq & small RNAseq

Bioinformatic data analysis

Anxiety-associated gene regulatory networks
RNA-seq data

• 2 inbred mouse strains
• 3 groups per strain
  – stress susceptible
  – stress resilient
  – control
• N=6 (RNAseq) or 10 (small RNAseq) per group
• 4 brain regions
  – Medial prefrontal cortex
  – Ventral hippocampus
  – Hypothalamus
  – Amygdala
RNA-seq data management and sharing

• Data storage, management, and analysis at the individual lab level: Well-functioning practices at the UH and CSC – IT Center for Science in Finland
• Many journals require submission of data at the time of publication, this has become a standard procedure
• Well-established procedures for microarray data sharing have been implemented to RNA-seq data
• E.g. ArrayExpress (EMBL-EBI), GEO (NCBI), CIBEX (DDBJ, Japan)
• Data is collected to MIAME and MINSEQE standards (Functional Genomics Data Society)
• Main issues:
  – The usefulness of the data depends on the depth of the metadata information
  – Giving credit to the producers of the data
  – User needs to have some bioinformatic skills to analyze the data
Primary data sharing for neurobiologists

• Integration of behavioral and RNA-seq data in easily accessible format
• How is gene X changing expression due to psychosocial stress?

- Issue: lack of expertise in database construction
- Difficulty in hiring project-based database administrators. Could universities/funding agencies etc. have a pool of administrators that individual investigators could use when needed?
Functional verification of networks to show causality

- Formation of testable hypotheses
  - Pharmacology
  - Existing transgenic models
  - Lentivirus-mediated gene transfer (overexpression/silencing)

Behavioral testing to assess the phenotype
EUOMM - The European Conditional Mouse Mutagenesis Program is founder member and European cornerstone of the International Knockout Mouse Consortium (IKMC).

EUOMM contributes the largest fraction of conditionally trapped and targeted genes in mouse C57BL/6N embryonic stem (ES) cells to the IKMC.

EUOMM vectors, mutant ES cells and mutant mice are distributed worldwide, enabling functional genomics research in a standardized and cost-effective manner by a much wider biomedical research community than has been possible previously.

EUOMM mutant ES cells and vectors can be obtained from the European Mouse Mutant Cell Repository (EuMMCR).

EUOMM mutant mice are archived and distributed by the European Mouse Mutant Archive (EMMA).
International Mouse Phenotyping Consortium pipeline

http://www.mousephenotype.org/impress

How to share deep phenotyping data produced by individual labs?
Do genetic variants in human homologous genes predispose to anxiety disorders?

Health 2000 population-based cohort:
- 6005 Finns interviewed with CIDI
- information on psychological distress and childhood adverse life events

Genome-wide association study

Meta-analysis with other cohorts
GWAS data

• The human genomics community is very open to data sharing
• Straightforward in the case of data submitted to a database, e.g. dbGaP (NCBI)
  – Application to the Data Access Committee
  – The research needs to fit within the original informed consent
  – See NIH GWAS Data Sharing Policy (Nat Genet Sept 2014)
• Most of the GWAS data is not in databases
  – Need for primary data sharing policies
Conclusions

• Sharing primary behavioral data
  – Need for better data management and sharing practices
  – Mouse Phenome Database collects phenotypic data from some strains
  – Sharing of integrated behavioral and RNA-seq data
  – How to share deep phenotyping primary data of transgenic models?

• Gene expression data
  – Tradition for open access
  – Well-established procedures for data management and sharing exist

• GWAS SNP data
  – Need for data sharing policies
  – Informed consent for re-use of data