Statistics in Systems Genetics

Mission:

Develop and make available data, statistical methodology and analytical tools for applying quantitative genetics and topological analysis of transcript expression to identify systems that produce disease and illuminate the path to repair of damaged systems, with a clear understanding of individual needs.

Laura Saba, PhD
Research Assistant Professor
Department of Pharmaceutical Sciences
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of Colorado Denver

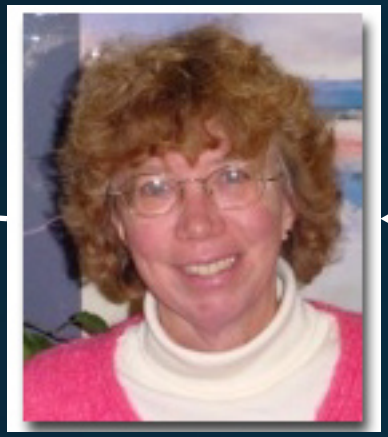
Presentation at http://github.com/LauraSaba

My CV

Mighty Meloneer
Rocky Ford, CO

Oredigger
Chemistry Degree
from CSM

??
PhD in Biostatistics



Diane Fairclough
"Latent Pattern
Mixture Models"

9 years studying
genetics for alcohol-
related traits

Collaborative Science

Dr. Boris Tabakoff
Professor
Dept Pharmaceutical Sciences
School of Pharmacy
Pharmacologist

Dr. Paula Hoffman
Professor
Dept Pharmacology
School of Medicine
Neurophysiologist

Dr. Katerina Kechris
Associate Professor
Dept Biostatistics
School of Public Health
Biostatistician

Mr. Spencer Mahaffey, ME
Computer Programmer

Ms. Lauren Vanderlinden, MS
Biostatistician

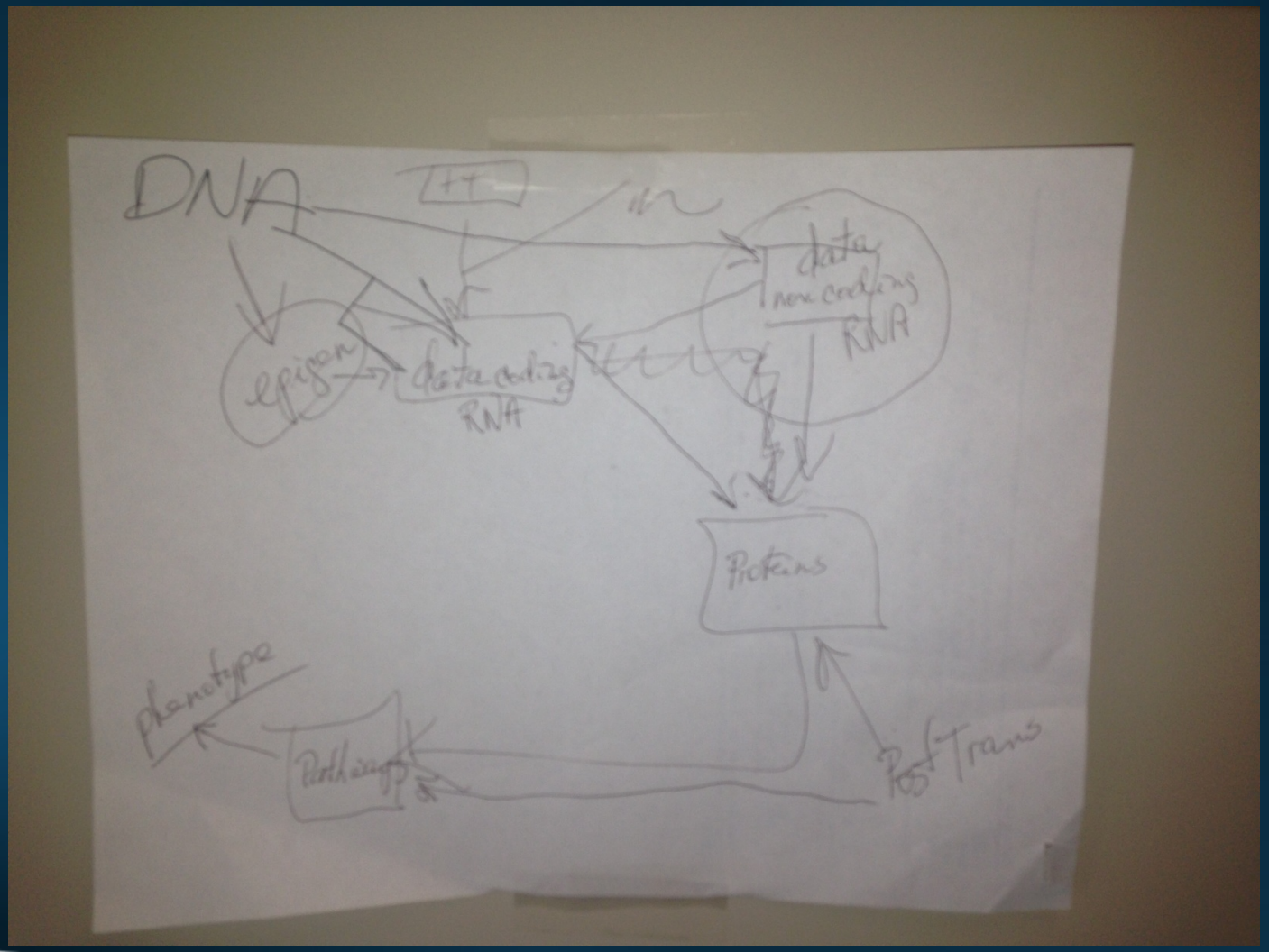
Dr. Stephen Flink
Mathematician

Ms. Yinni Yu, MS
Lab Manager

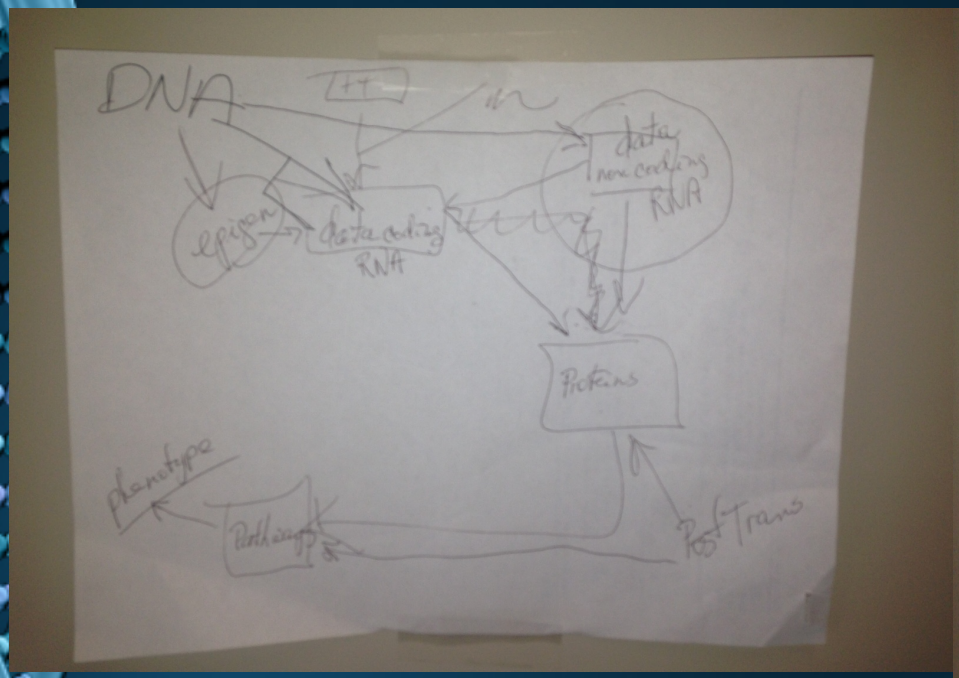
Systems Genetics

"The reductionist approach has successfully identified most of the components and many of the interactions but, unfortunately, offers no convincing concepts or methods to understand how system properties emerge...the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously and by rigorous data integration with mathematical models" (Sauer *et al., Science 2007*).

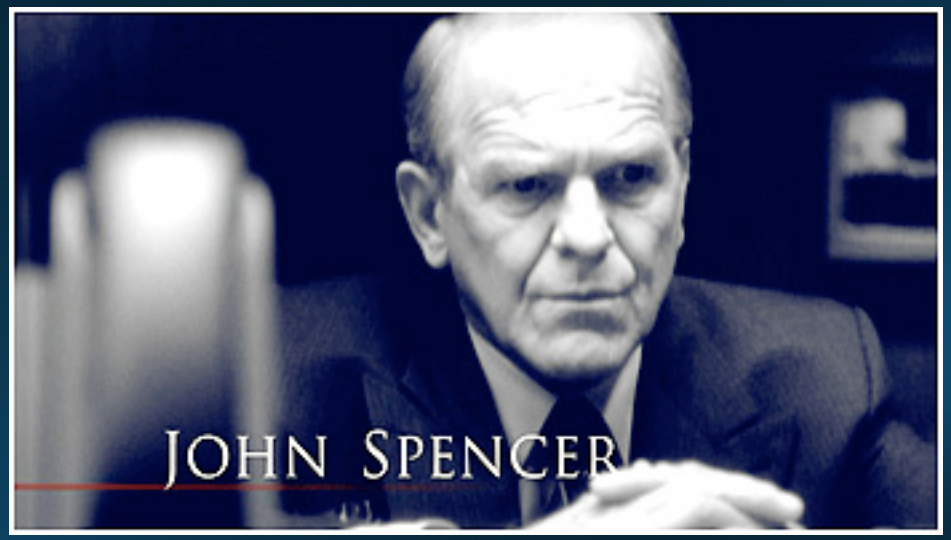
System Genetics



System Genetics



Alcoholism



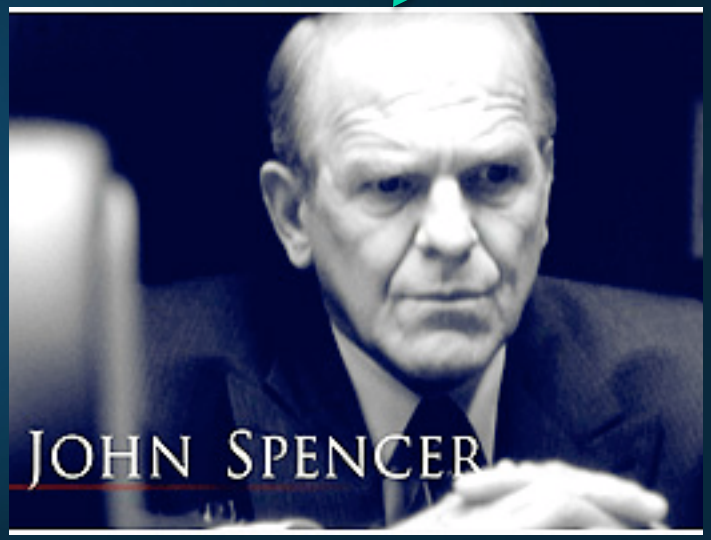
“That's because you think it has something to do with smart and stupid. Do you have any idea how many alcoholics are in Mensa? You think it's a lack of willpower? That's like thinking somebody with anorexia nervosa has an overdeveloped sense of vanity…” – John Spencer as Leo McGarry *The West Wing*

Why Did He Become An Alcoholic?

Genetics.
Leo's Dad was an alcoholic

Past Trauma.
Vietnam War Vet

Daily Stress.
Chief of Staff to US President



Neurological Changes.
Previous excessive drinking leads to biological changes

Animal Models

Can control their environment

Examine biological traits prior to and immediately after initial exposures

Can control alcohol intake

Can control their genetics

Can measure biological traits in any organ at any time point

How do you measure alcoholism in animals?

Types of Data

- Phenotypes
  - free choice consumption
  - deprivation induced drinking
  - alcohol tolerance
  - alcohol sensitivity
- DNA Sequence
  - SNP arrays
  - DNA-Seq
- RNA Expression Levels and Structure
  - array or RNA-Seq
  - different tissues, tissue parts, cell types, etc.
  - alternative splice variants, untranslated regions, etc.

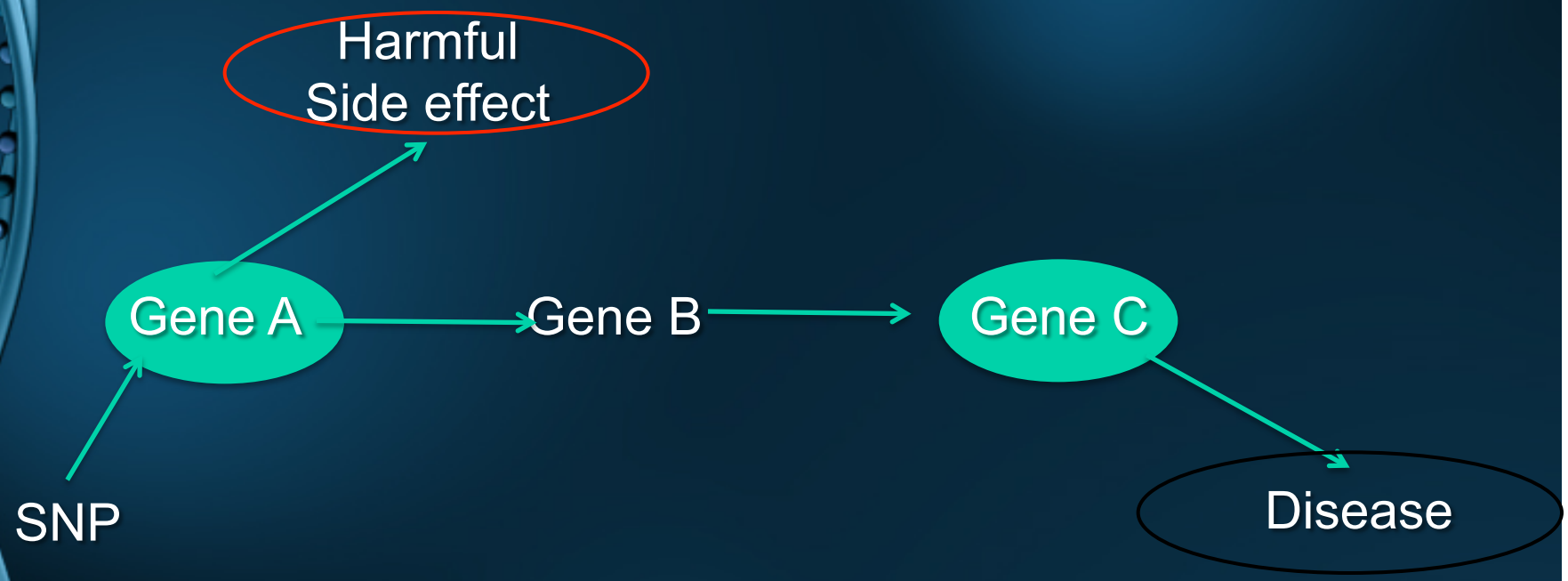
Statistics-Related Questions

Questions:

1. What are the best methods for processing data from these emerging technologies, e.g., high throughput sequencing?
2. Can we integrate data from DNA, RNA, and phenotype to identify a genetic pathway associated with disease?
3. Can we statistically merge data across species?

Drug Development

- Can we use these data to predict changes in a phenotype based on a pharmacologic manipulation of the pathway?



Current Projects

- **Massive RNA-Seq (Tabakoff/Hoffman/Saba)**
  - goal is to generate data on approximately 500 samples
  - integrating spike-in information into quantification and estimation of uncertainty
  - using reads to determine structure of unannotated genes/ isoforms
- **eQTL integration across tissues and species (Kechris)**
  - Use hierarchical Bayesian modeling to leverage the abundance of data on mouse/rat RNA expression in human eQTL and GWAS studies
- **prediction with validation (Saba)**
  - evaluate prediction results/methods applied to large mouse populations by manipulating the expression of MeCP2 in regions of mouse brain and evaluating genome-wide changes in expression
- **systems genetics methods for complex traits**

For Students Interested in Statistical Genetics

- Technology outpaces statistical methods → LOTS OF OPPORTUNITIES
- Lots of publically available array and high throughput sequencing data (e.g., http://phenogen.ucdenver.edu)
- Mostly program in R and command line tools developed for RNA-Seq data