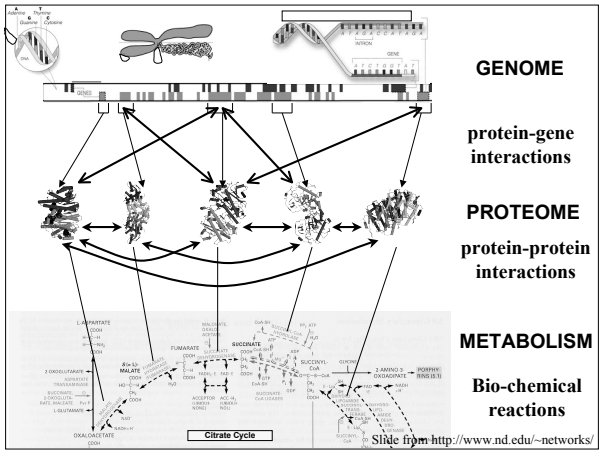
Bioinformatics 2

From genomics & proteomics to biological networks



Slide from http://www.nd.edu/~networks/

Biological Profiling

- Microarrays
  - cDNA arrays
  - oligonucleotide arrays
  - whole genome arrays
- Gene Networks and Inference
- Prediction methods

Why microarrays?

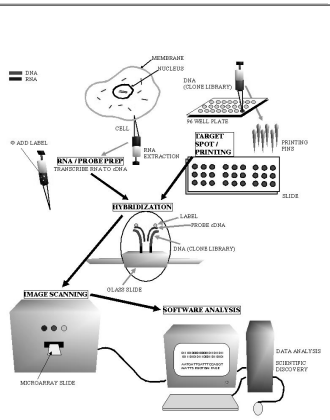
- What genes are expressed in a tissue and how does that tissue respond to one of a number of factors:
  - change in physical environment
  - experience
  - pharmacological manipulation
  - influence of specific mutations

What do we actually get?

- A snap-shot of the mRNA profile in a biological sample
- With the correct experimental conditions we can compare two situations
- Not all biological processes are regulated through mRNA expression levels

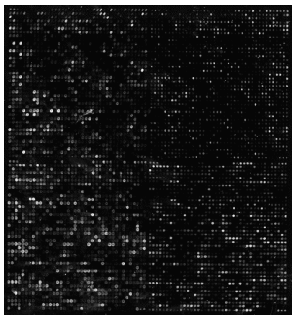
What can we learn?

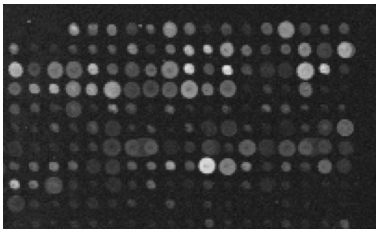
- Identify functionally related genes
- Find promoter regions (common regulation)
- Predict genetic interactions
- If we change one variable a network of gene responses should compensate
- Homeostasis is a fundamental principle of biology - almost all biological systems exist in a controlled state of negative feedback.



From medir.ohsu.edu/~geneview/education/ Normalization%20Presentation%2011-7-2001.ppt

Robot spotted array.





http://experimental.act.cmis.csiro.au/Spot/index.php

Spot detection software review:
http://ihome.cuhk.edu.hk/~b400559/arraysoft_image.html

The Transcriptome

- Microarrays work by revealing DNA-DNA binding.
- Transcriptional activators also bind DNA
- Spot genomic DNA onto glass slides
- Label protein extracts
- Hybridise to the genomic probes
- Reveals domains that include promoter regions

Dealing with microarray data

- Many experiments focus on a few samples over a few conditions to examine gene expression changes in response to an environmental change.
- Can we reuse this data to learn about networks of genes?

Predicting gene networks

- Very active area of research.
- Multiple microarrays contain information that we can use to predict networks
- The data from microarray experiments was not collected for that express purpose.
- The Microarray Gene Expression Database group are proposing management systems to help

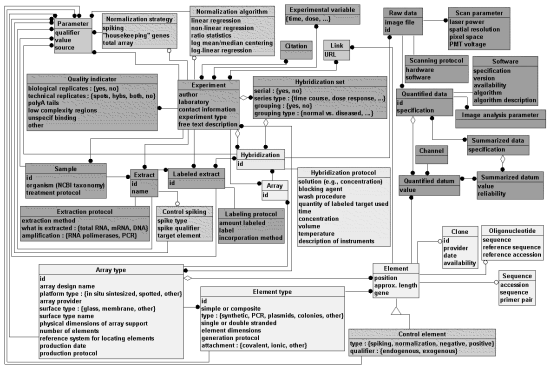
MGED

- Microarray Gene Expression Database Group
- MIAME
- MAGE
- Ontologies
- Normalization

- http://www.mged.org/

MIAME

- Minimum amount of information required to unambiguously describe a microarray experiment.
  - relevant details of the experiment.
  - biological repetition
  - reuse of the data
  - comparison across experiments.



MIAME Express

- Under development at EBI
- GUI interface to the MIAME database system

- http://www.ebi.ac.uk/microarray/MIAMExpress/miamexpress.html

MAGE

- Microarray And Gene Expression
- Developing standards for representing and exchanging microarray data.

- MAGE-OM Object Model for microarray data developed using Unified Modelling Language (UML)

MAGE

- MAGE-ML - XML based model for communication
- MAGE-STK - toolkit for handling and converting between MAGE-OM and MAGE-XML on a variety of platforms
  - Java and PERL APIs available

Excerpts from a Sample Description

courtesy of M. Hoffman, S. Schmidtke, Lion BioSciences

Organism: mus musculus [NCBI taxonomy browser]
Cell source: in-house bred mice (contact: norma.howells@itg.fzk.de)
Sex: female [MGED]
Age: 3 - 4 weeks after birth [MGED]
Growth conditions: normal
controlled environment
20 - 22 ºC average temperature
housed in cages according to German and EU legislation
specified pathogen free conditions (SPF)
14 hours light cycle
10 hours dark cycle
Developmental stage: stage 28 (juvenile (young) mice) [GXD "Mouse Anatomical Dictionary"]
Organism part: thymus [GXD "Mouse Anatomical Dictionary"]
Strain or line: C57BL/6 [International Committee on Standardized Genetic Nomenclature for Mice]
Genetic Variation: Inbr (J) 150. Origin: substrains 6 and 10 were separated prior to 1937. This substrain is now probably the most widely used of all inbred strains. Substrain 6 and 10 differ at the H9, Igh2 and Lv loci. Maint. by J,N, Ola. [International Committee on Standardized Genetic Nomenclature for Mice]
Treatment: in vivo [MGED] intraperitoneal injection of Dexamethasone into mice, 10 microgram per 25 g bodyweight of the mouse
Compound: drug [MGED] synthetic glucocorticoid, Dexamethasone, dissolved in PBS

Biomaterial Concepts

- Environmental or experimental history: A description of the conditions the organism has been exposed to that are not one of the variables under study.
  - Culture conditions: A description of the isolated environment used to grow organisms or parts of the organism.
    - atmosphere, humidity, temperature
    - light: The photoperiod and type (e.g., natural, restricted wavelength) of light exposure.
    - nutrients: The food provided to the organism (e.g., chow, fertilizer, DEMM 10%FBS, etc.).
    - medium: The physical state or matrix used to provide nutrients to the organism (e.g., liquid, agar, soil)
    - density range: The concentration range of the organism.
    - contaminant organisms: Organisms present that were not planned as part of the study (e.g., mycoplasma).
    - removal of contaminants: Steps taken to eliminate contaminant organisms.
    - host organism or organism parts: Organisms or organism parts used as a designed part of the culture (e.g., red blood cells, stromal cells).
  - Generations: The number of cell divisions if the organism or organism part that is cultured is unicellular otherwise the number of breedings.
  - Clinical history: The organism's (i.e., the patient's) medical record.
  - Husbandry: water, bedding, barrier facility, pathogen test results
  - Preservation: seed dormancy, frozen storage

Biomaterial Concepts

- Treatment: The manipulation of the biomaterial for the purposes of generating one of the variables under study.
  - somatic modification: The organism has had parts removed, added,or rearranged.
  - genetic modification: The organism has had genes removed, added, or rearranged.
  - starvation: The organism (or organism part) has been deprived of nutrients.
  - infection: The organism (or organism part) has been exposed to a virus or pathogen.
  - behavioral stimulus: The organism is forced to respond to a stimulus with some behavior (e.g., avoidance, obtaining a reward, etc.)
  - agent-based treatment: The treatment is effected by a defined chemical, biological, or physical agent.
  - agent type: chemical (drugs), biological (macromolecule), physical (stress from light, temperature, etc.)
    - agent application: In vivo, in vitro, in situ; qualitative or quantitative
  - treatment protocol: method of treatment
  - treatment parameters: constant, variable
  - treatment duration: length of treatment

Biomaterial Concepts

- Biomaterial preparation: A description of the state and condition of the biomaterial.
  - Time of day when the biomaterial was generated (i.e., sampled). Pathological staging: pre or post mortem at sampling
  - state at start of treatment (age, time of day)
  - physio-chemical composition of the sample: amount of material, number of cells, purity.
  - Extraction: Chemical extraction, Physical extraction
  - protocol: method used.
  - Pool types:
    - Multiple: Biomaterial prepared from multiple specimens, but same Organism, Genotype, Phenotype and treatment.
    - Individually: Biomaterial prepared from individually specimen, but same Organism, Genotype, Phenotype and treatment

MGED Ontologies

- Develop standard ontologies for describing experimental procedures associated with microarray data.
- Ontologies specific for:
  - sample (e.g. species, anatomical location etc)
  - by concept
  - array design

MGED Normalization

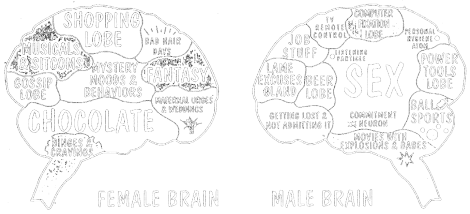
- Working group to discus standards for normalization in microarray experiments
  - Differences in labeling efficiency between dyes
  - Differences in the power of the two lasers.
  - Differing amounts of RNA labeled between the 2 channels.
  - Spatial biases in ratios across the surface of the microarray.

Gene networks and network inference

What is a gene network

- Genes do not act alone.
- Gene products interact with other genes
  - Inhibitors
  - Promoters
- The nature of genetic interactions in complex
  - Takes time
  - Can be binary, linear, stochastic etc
  - Can involve many different genes

What makes boys boys and girls girls?



Sugar, Spice and synthetic Oestrogens?

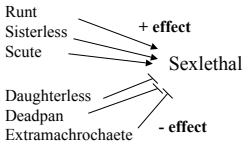
Sex determination: a gene cascade (in flies…)

6 Genes detect X:A ratio

	Females	Males
Runt Sisterless Scute	↑	↓
Daughterless Deadpan Extramachrochaete	↓	↑

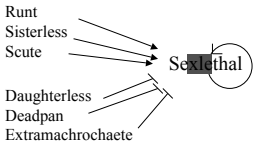
Sex determination (in flies…)

6 Genes regulate 'Sexlethal'



Sex determination (in flies…)

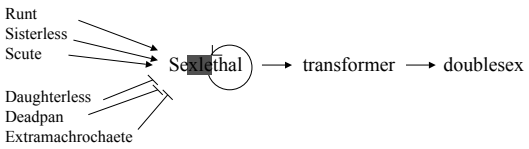
Sexlethal can then regulate itself...



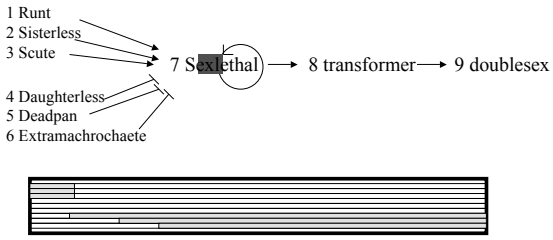
Sex determination

(in flies…)

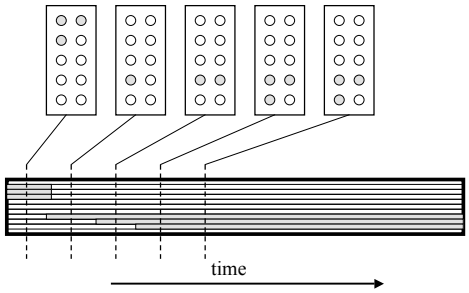
Downstream cascade builds...



Gene expression and time



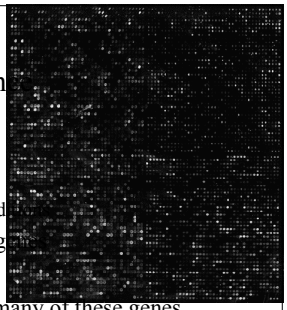
Gene microarrays



Gene Network Inference

- Gene micro-array data
- Learning from micro-array data
- Unsupervised Methods
- Supervised Methods
- Edinburgh Methods

Gene Network Inferen



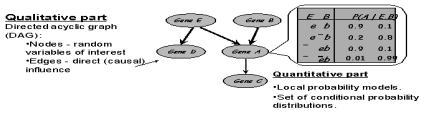
- Gene micro-array data
  - Time Series array data
  - Tests under ranges of cond
- Unlike example - 1000s g
- Lots of noise
- Clustering would group many of these genes together
- Aim: To infer as much of the network as possible

Learning from Gene arrays

- Big growth industry but difficult problem
- Initial attempts based on unsupervised methods:
  - Basic clustering analysis - related genes
  - Principal Component Analysis
  - Self Organising Maps
  - Bayesian Networks

Bayesian 'gene' networks

- Developed by Nir Friedman and Dana Pe'er
- Can be easily adapted to a supervised method



Learning Gene Networks

- The field is generally moving towards more supervised methods:
  - Bayesian networks can use priors
  - Support Vector machines
  - Neural Networks
  - Decision Trees

Can we combine network knowledge with gene inference?

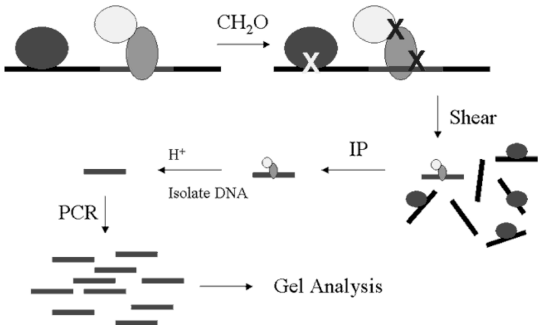
- Scale free architecture
  - Chance of new edges is proportional to existing ones
  - Highly connected nodes may well be known to be lethal
- Network motifs
  - Constrain the types of sub networks
- Prior Knowledge
  - Many sub networks already known

How can we find or test potential gene interactions?

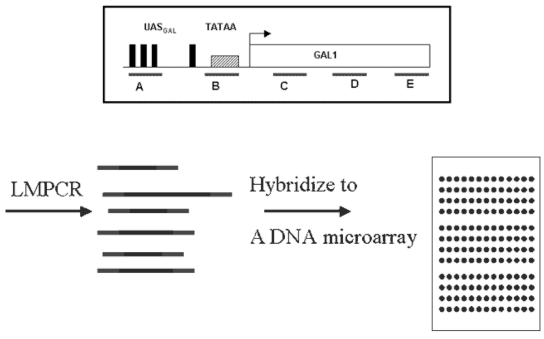
ChIP to Chip

Chromatin Immunoprecipitation to Microarray (Chip)

Protein-DNA interactions
de-novo prediction has many false positives
Which DNA sites do actually bind a specific TF?
Requires an antibody to the protein



http://proteomics.swmed.edu/chiptochip.htm



http://proteomics.swmed.edu/chiptochip.htm

Conclusions

- Gene network analysis is a big growth area
- Several promising fields starting to converge
  - Complex systems analysis
  - Using prior knowledge
  - Application of advance machine learning algorithms
  - AI approaches show promise