Introduction to group projects

Bio311

03/21/2017

Yeast as model system

- Well-studied, single cell eukaryote
- Important to humans (beer and wine, bread, bioprocessing, disease)
- Small genome, tractable molecular genetics
- Lots and lots of data available
  - e.g. See NIH Gene Expression Omnibus (GEO)
- We will use yeast data for group projects

THINK / PAIR / SHARE

- “Anything found to be true of E. coli must be true of elephants.”–Jacques Monod
- THINK/PAIR/SHARE: What does Monod mean by this?
- Model systems provide a simpler, more tractable tool to discover general principles of biology

Projects: biological questions

1. What are the molecular functions of TFs in yeast?
   - What are the functions of genes the TF binds?
   - What genes are co-expressed with the TF?
   - What genes are most affected by a TF knockout?

1. How do networks function dynamically to enable physiological and metabolic adjustment in response to environmental cues (stress, nutrients)?
   - What other TFs also bind the genes controlled by my TF?
   - How do genes controlled by the network respond to other stresses?

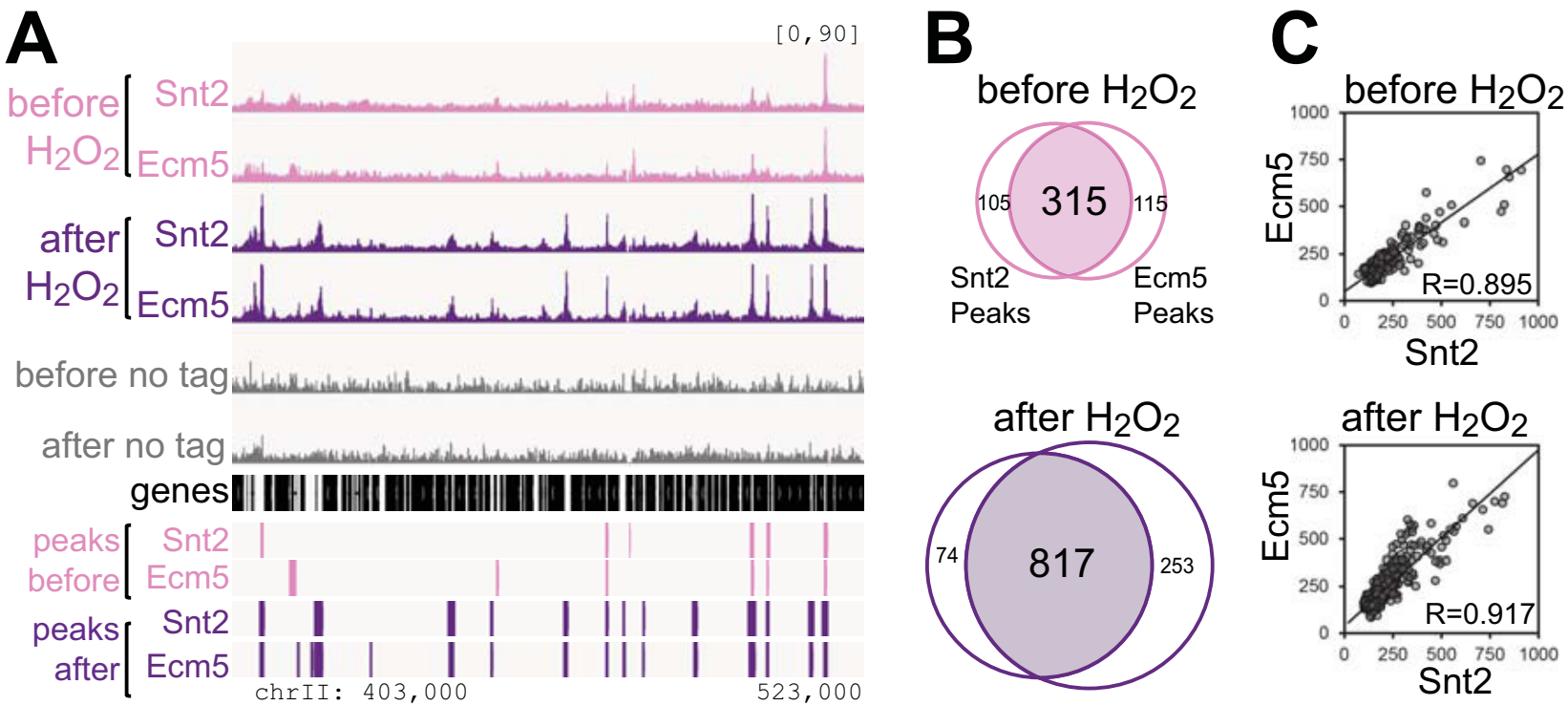
Example from the literature

- How will we apply what we learned so far to answer the biological questions?
- How will I use the literature to help answer the questions?
- How do I find out what the dataset means and how the data were collected?

Example data, exercise 1: Understanding the biological question

- Open the Baker et al. 2013 paper
- Read the abstract and introduction
- Skim the result and look at the figures
- Answer these questions
  - What was the purpose of the study?
  - What were the main conclusions?
  - What evidence supports these conclusions?

Example data: ChIP-seq



Baker et al. 2013, Mol Cell Bio, 33:

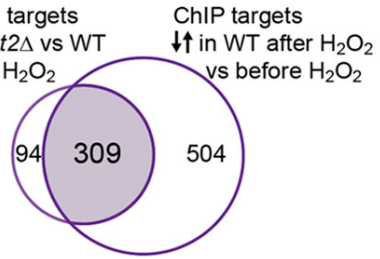
Example data: Gene expression

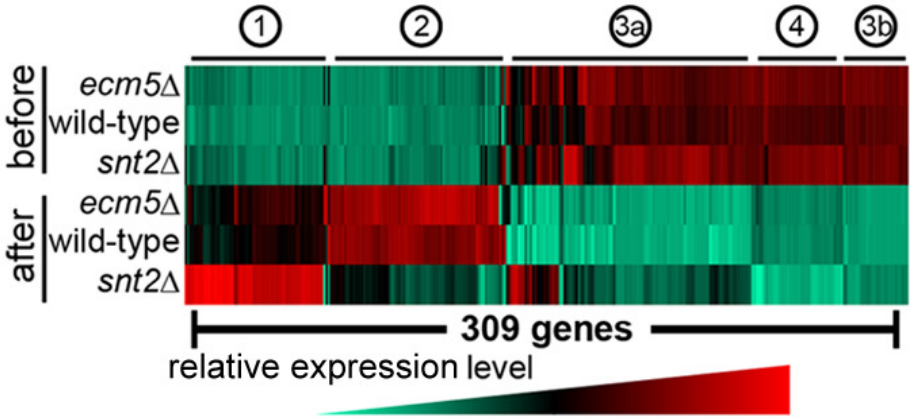
A

	down-regulated		
	snt2Δ	both	*ecm5Δ*
before H_2O_2	38	5	19
after H_2O_2	262	0	6

	up-regulated		
	snt2Δ	both	*ecm5Δ*
before H_2O_2	134	3	14
after H_2O_2	475	1	1

D





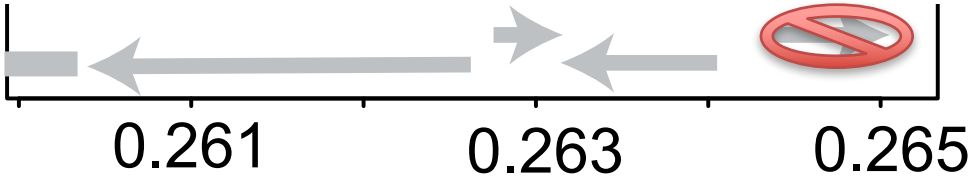
Baker et al. 2013, Mol Cell Bio, 33:

Interlude: What is a mutant?

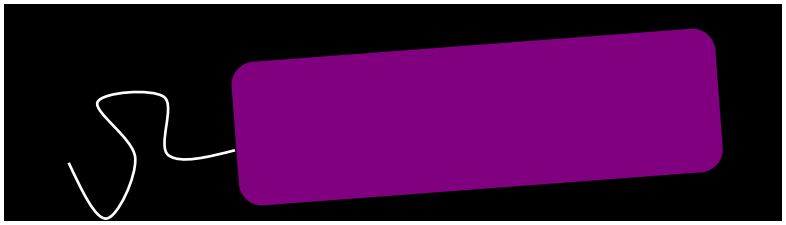
THINK / PAIR / SHARE: Come up with two definitions of “mutant”

1. General or colloquial (i.e. how your grandmother may define it)
2. Specific to our projects – Why are they useful?

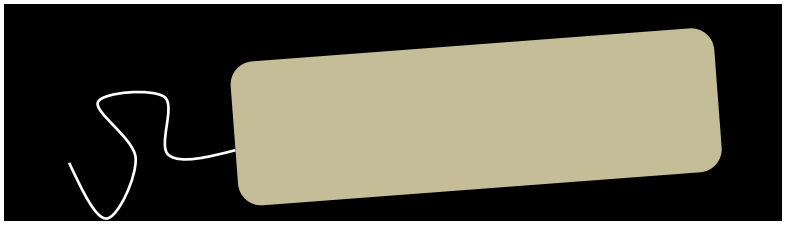
What is a mutant?



"WILD TYPE"



"TF MUTANT"



GENOTYPE --→ PHENOTYPE

What is a mutant?

- An individual that is genetically different from “wild-type” background (the majority of individuals of that organism in its natural environment or a commonly accepted reference)
- Changes in DNA sequence (heritable)
- Integral to the process of evolution and the study of genetics

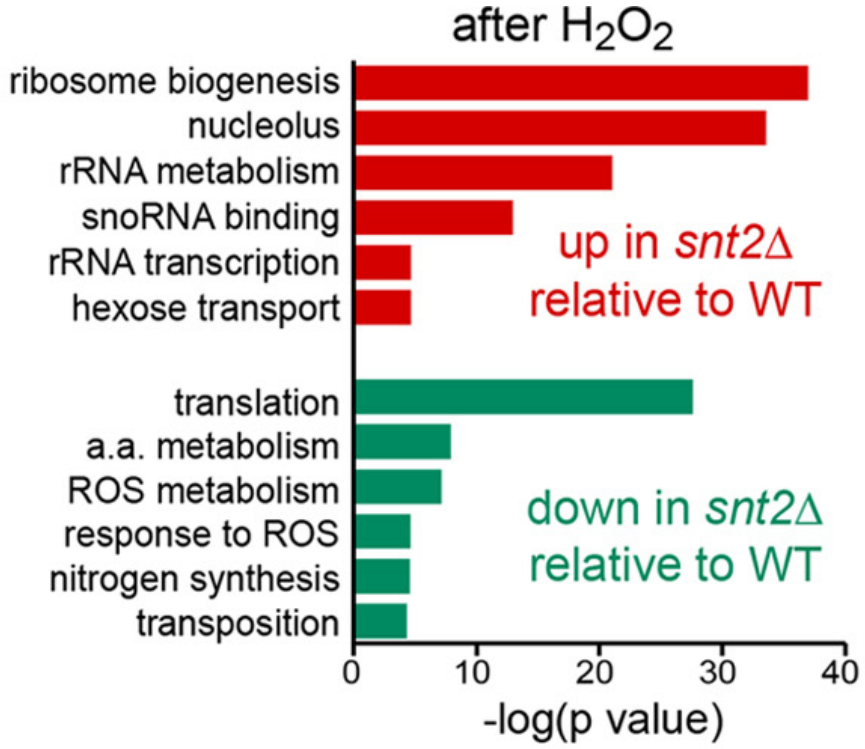
Types of Mutations

- Point mutations (single nucleotide)
  - Nonsense
    - Nucleotide change → stop codon
  - Missense
    - 1 nucleotide → 1 a.a. change
  - Silent mutation
    - Same a.a. coded by mutated DNA (i.e. CGG and CGC)
    - Mutation affects non-coding DNA (introns, intergenic regions)
- Insertions
  - frameshifts
- Deletions
  - Whole gene= knock-out (or KO)

Think/pair/share
Why make a mutant?

- Study the function of a gene and its product
- Efficient detection of the expression of a gene or a protein (“knock-in” reporter gene)
- Bioengineering and synthetic circuits

Example data: Gene Ontology



Baker et al. 2013, Mol Cell Bio, 33:

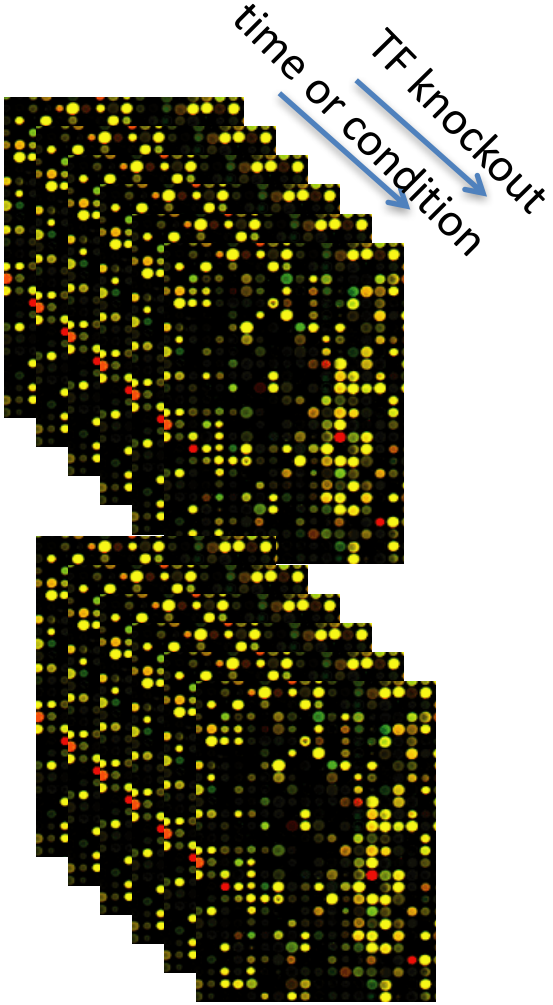
Example data: exercise 3

- Take a few minutes and write a list of data analysis methods you would use to answer these questions:
  - What sets of genes change expression in response to the environmental perturbation?
  - What patterns of expression do you observe for these genes?
  - What genes change in the TF knockout relative to the wild type?
  - What genes are directly regulated by the TF?

PROJECT EXPERIMENTS AND DATA

RNA

ChIP-chip



Project data: Each group will be assigned a ChIP and a gene expression dataset

ChIP-chip or -seq: binding data for 1 TF across genome

genes	Peak location	positions	intensity	p.value
VNG0019H	15773	15752	2.12159812	0.00421941
VNG0037H	32224	32222	6.69E+00	0.00125313
VNG0050C	43223	43222	1.03703311	0.01674641
VNG0050C	43791	43792	4.36955675	0.00601504
VNG0051G	44855	44852	4.08208823	0.00676692
VNG0052H	48961	48962	4.26655536	0.00676692
VNG0057H	53231	53232	3.59549585	0.01052632

Where does the TF bind?

Gene expression: Wild type vs mutant of same TF
OR Wild type over time in response to a stress

GENE	dtrmB_+glu_a	dtrmB_+glu_a	dtrmB_+glu_a	dtrmB_-glu_a	dtrmB_-glu_a
VNG0001H	0.053	0.0525	0.021	-0.151	-0.048
VNG0002G	0.11	0.08	0.0555	-0.085	-0.084
VNG0003C	0.1315	0.1335	0.082	0.005	0.062
VNG0005H	0.1575	0.037	-0.055	0.031	0.0525
VNG0006G	0.054	0.0955	0.03	0.005	-0.054
VNG0008G	0.035	0.013	-0.004	-0.039	-0.0315

How are genes differentially expressed in TF knockout vs wild type?

What genes are co-expressed with the TF over time?

Project data: each group also has access to ChIP-chip data for all TFs

- What other TFs control the genes that are controlled by my TF of interest? (What is the network controlling the stress response of interest)?
- How does the expression of genes controlled by this network change over time or in response to a TF knockout?

EXAMPLE PROJECT WORKFLOW

1. Wild type data over time during stress.
   a) Cluster genes according to common patterns
   b) Functions of genes in each cluster
   c) Which of these clusters contain TFs of interest?
   d) Given correlations between TF of interest and genes in that cluster, formulate hypothesis for which genes are regulated by the TF.
2. Use ChIP-chip data to test the hypothesis
3. Make gene regulatory network for your TF
4. Ask which other TFs (from the larger ChIP dataset) regulate genes in your network.
5. Compare what you got with the findings of the paper(s) that first reported the dataset you are working with

THERE ARE MANY “RIGHT” ANSWERS

- There are p-values, ratios, statistical confidence levels.
- In scientific research, you make decisions and conclusions based on whether you are convinced by the evidence, then argue your points to the community
- There are, however, “WRONG” answers
  - Conclusions that do not logically fit the data
  - Conclusions that violate known facts

Group membership

- Choose your own group with the following stipulations:
  - Each group must contain 3-4 people, 10 groups total
  - Each group must contain at least 2 people who did not know each other outside of this class
  - The instructors reserve the right to rearrange some groups for balance of expertise and numbers

Group datasets

- By Thurs. 3/23 at noon, email your choices for group membership to all three instructors.
- On Thursday, work with your group to read the paper associated with your data & familiarize yourselves with data / experiment and come up with a plan.
- By the end of class Thursday, hand in a bullet-point list of your group's plan and/or preliminary figure for credit/no credit participation points.

PROJECT EXPECTATIONS

- Be rigorous, methodical, skeptical.
- Read the literature – find papers in addition to the one assigned on the topic of your project
- Explore – click on things in databases for more information.
- Meet with your group at least 2 hrs/week outside of class
- ASK!!!! It's ok not to know something.

PROJECT EXPECTATIONS

- SAVE YOUR WORK (figures, notes, ideas, etc). Using Markdown will assist you!
- Homework: Progress reports due periodically. The due dates, homework format, and conceptual overview will be posted on GitHub.
- Present your results in a poster at end of semester.
- Everyone in the group MUST PARTICIPATE. Don't let your team down.

QUESTIONS?