From Genome Function to Biomedical Insight: ENCODE and beyond

Understanding basic biology using functional genomics: solving the genotype-phenotype problem

Brenda Andrews, The Donnelly Centre, University of Toronto

What have we done?

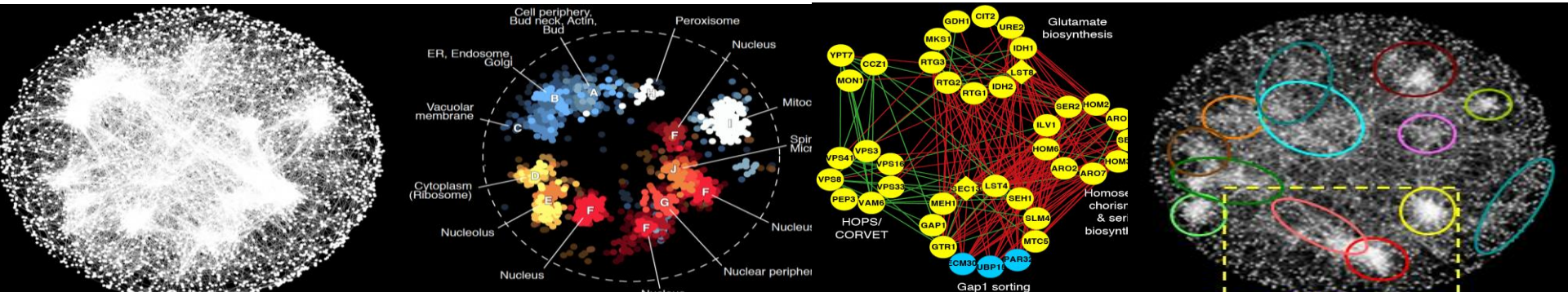
Generated data that predicts functional elements in the human genome (and selected model genomes)

What do we need to do?

- Generate unbiased, genome-scale <u>functional</u> data that will enable useful modeling of biological processes
- Develop methods to examine regulatory events across the genome on timescales relevant to molecular mechanisms – explore combinatorial perturbations
- enable development of quantitative, predictive models

How do we do it?

- Engage model organism and human genetics communities in a joint effort to systematically define the genotype-phenotype relationship



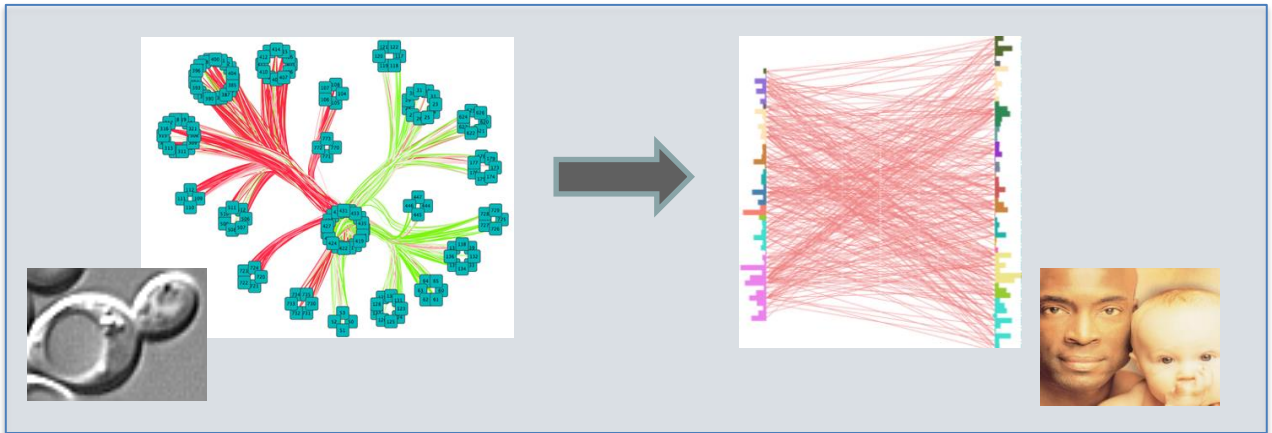
A model project: analysis of genetic interactions in yeast

Reagents for systematic genetics in yeast

- **Deletion array – non-essential genes**
- **Arrays of conditional alleles of essential genes**

- Most single genetic perturbations are of little phenotypic consequence
- The eukaryotic cell is highly genetically buffered
- Need methods for systematic study of genetic interactions
- Need sensitive assays to assess phenotypes in mutant backgrounds

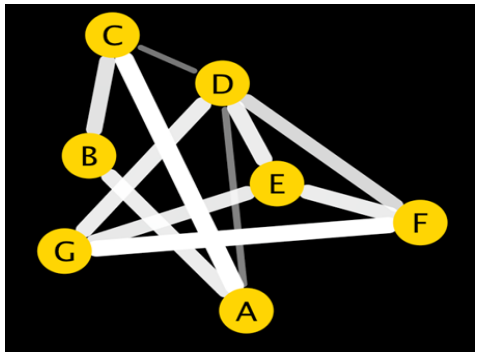
GFP-collection



Assaying Genetic Interactions

<u>Overall Project Goal</u>:

Use yeast mutant arrays to systematically explore genetic interactions



A **Genetic Interaction** occurs if an allele of one gene combines with the allele of another gene to generate an unexpected double mutant phenotype.

Can we use **genetic interactions** to systematically define gene function, biological pathways; connections between bioprocesses; general principles of genetic interactions; explain the missing heritability common in genetic studies of complex diseases

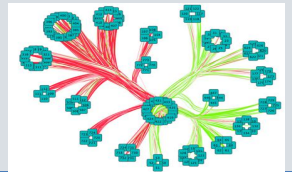
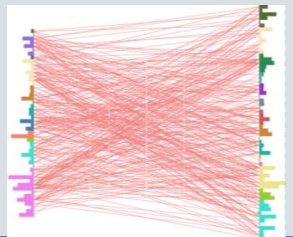
The next phase (now what?):

What areas should NHGRI support to facilitate understanding of basic biological questions/processes? What are the gaps?

"Big Picture' Gap: we need to functionally annotate the human genome through systematic perturbation, which requires considerable investment in unbiased genome-scale functional data generation, in a variety of experimental systems

Why will filling this gap be useful?

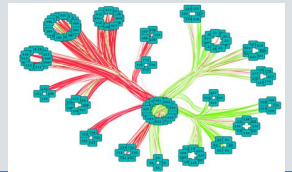
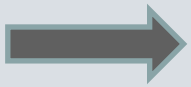
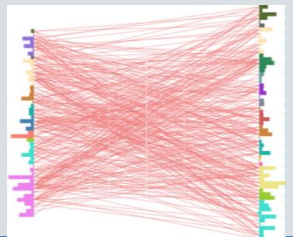
Experience from the yeast community tells us that both systems biologists and more traditional biologists benefit from large-scale, unbiased investigation - to understand the function of biomedically-relevant genes and pathways, they must be considered in their global cellular context.


How to fill the gap – technologies, data generation, projects?

Apply what we've learned about systematic functional annotation of eukaryotic genomes in yeast to human cells

-requires continued support of model system projects to develop computational methods and models and to produce reference maps and 'classifications'
-requires investment in establishing community-wide resources, like genome-scale mutant collections, for efficiently constructing and characterizing genetic perturbations
-requires tools for rapid and quantitative analysis of cell states and phenotypes automated image analysis of cell biological read-outs
-we need to consider genetic interactions, on a large-scale

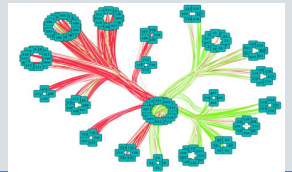
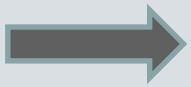
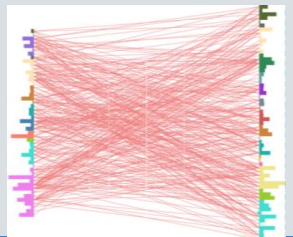






How to fill the gap – technologies, data generation, projects?

Phenotypic readouts of gene perturbation (regulatory element perturbation): Genetic interaction profiles

Genetic interaction mapping in human cells

- Continued technology development for genome engineering
- Map essential genes/measure growth rates across selected cell lines
- Use single perturbation information to choose query genes for full-genome genetic interaction screens (broad biological representation; information-rich queries)
- Functionally rich reference map for interpreting consequences of perturbing putative regulatory elements
- Combinatorial genetic perturbations will be necessary for useful functional annotation


How to fill the gap – technologies, data generation, projects?

Phenotypic readouts of gene perturbation (regulatory element perturbation): "BIG Data" cell biology

- Develop 'tool-kit' of reporters (e.g. fluorescent markers of all subcellular compartments; markers of cell cycle position; reporters of TF activity
- Produce quantitative, single cell data on reporter activity/subcellular compartment phenotypes after genetic/environmental perturbation
- Need a comprehensive set of data from accessible model system to enable intelligent selection of perturbations/conditions to choose for human cell screens
- Need serious effort to improve computational image analysis of cell images
- -move towards 'plug and play' approach (deep learning?) that can be applied to interpret any biological image

Community project (100s of labs)?