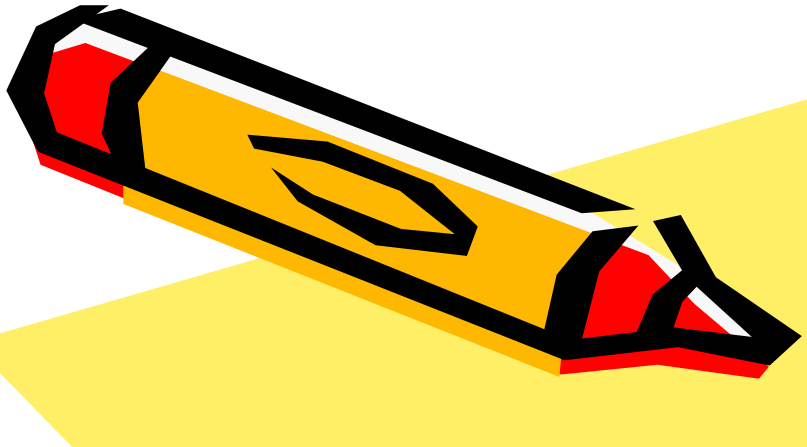
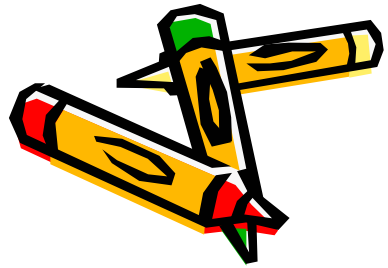


Genomics, Proteomics and Bioinformatics

蛋白(質)體學

Syllabus

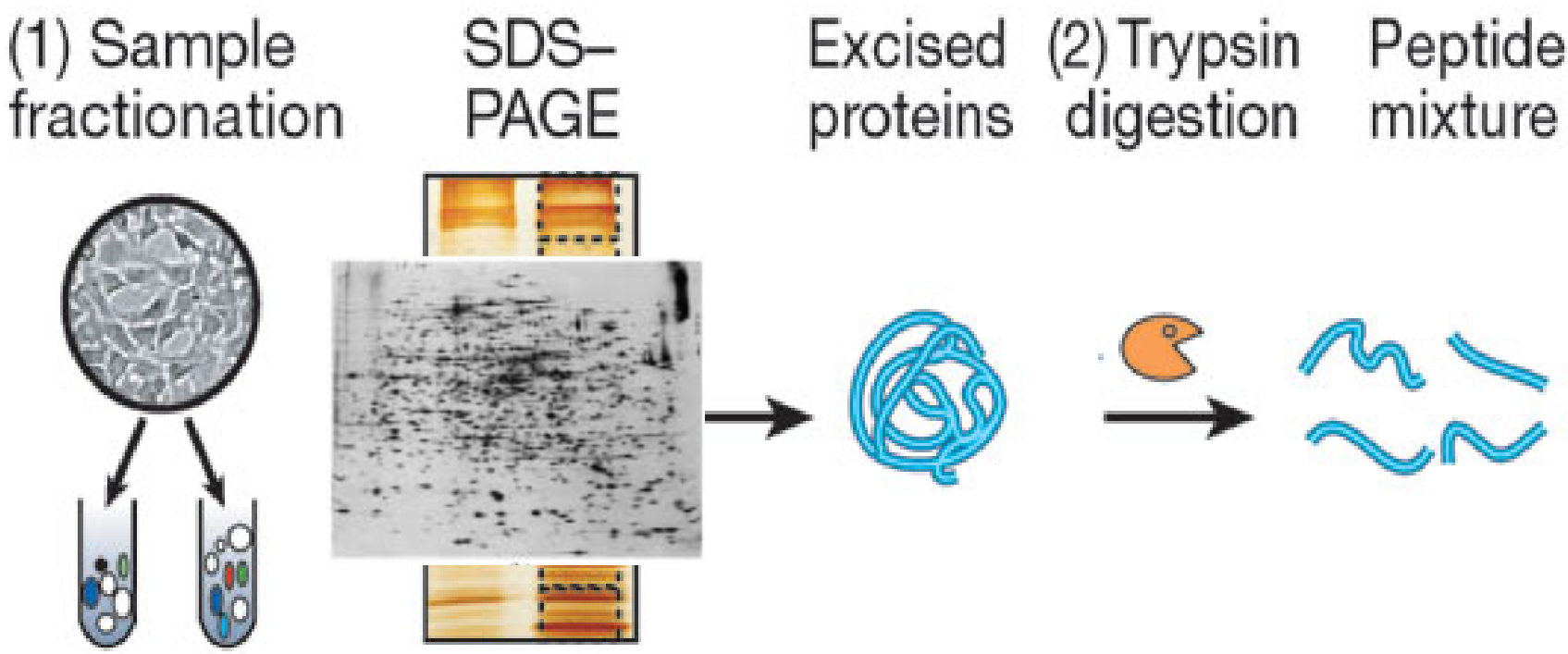


- Schedule
  - 3/10 Introduction, Proteomic strategies
  - 3/17 Strategies, Functional analysis
  - 3/24 Protein-protein interaction
  - 3/31 Case studies
- Grading
  - 50% Attendance and class performance
    - Be prepared to talk in class and read after class
  - 50% Oral presentation

Slides and reading assignments will be available at www.huichun.tcu.edu.tw

MS-based Proteomics

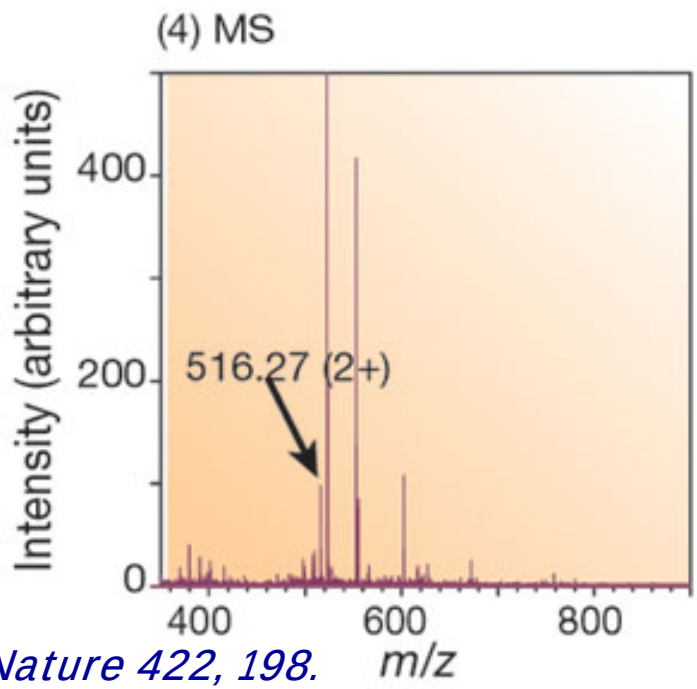
- Sample preparation for MS analysis

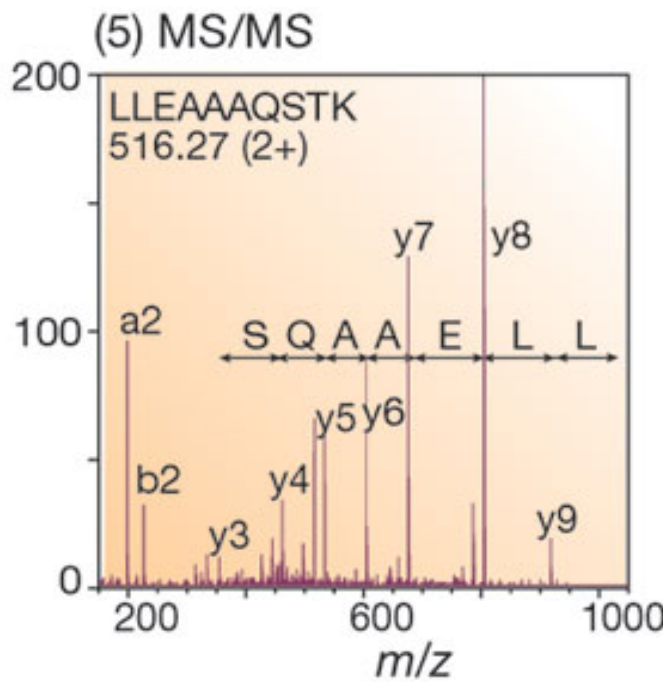


Aebersold & Mann, 2003 Nature 422, 198.

Mass Spectroscopy

- Ionization
  - MALDI: matrix-assisted laser desorption/ionization
  - ESI: electrospray ionization
- Mass determination: mass/charge (m/z) ratio
  - MS: peptide identification
  - MS/MS: sequencing

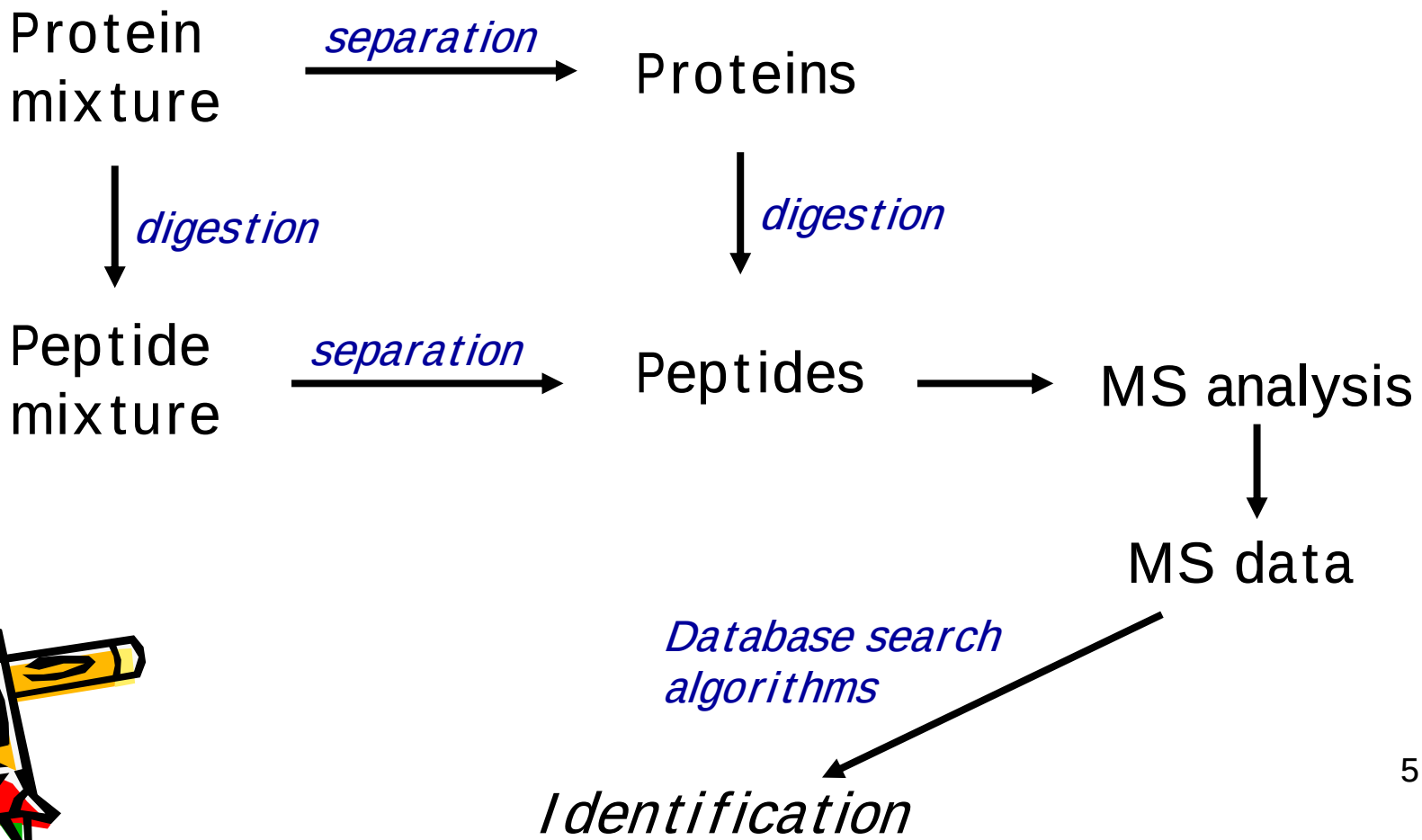


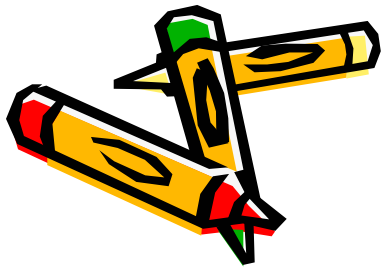


Aebersold & Mann, 2003 Nature 422, 198.

Proteomics, "the original view"

- 2D gel separation → MS analysis → Protein identification
  - Originated around 1993





“System biology”

DNA
(a gene)

↓

mRNA
(transcript)

↓

Protein

↓

Cellular function(s)

Genome

↓

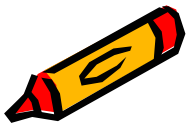
Transcriptome

↓

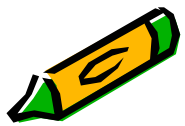
Proteome

↓

Protein chemistry vs. Proteomics

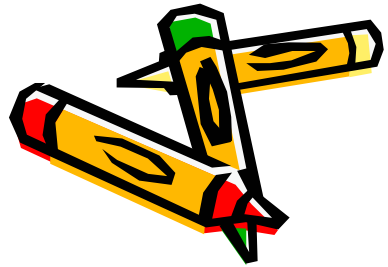


- Individual proteins
- Complete sequence analysis
- Emphasis on structure and function
- Structural biology

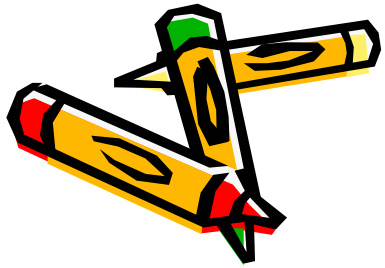


- Complex mixtures
- Partial sequence analysis
- Emphasis on identification by database matching
- Systems biology

The “-omics”



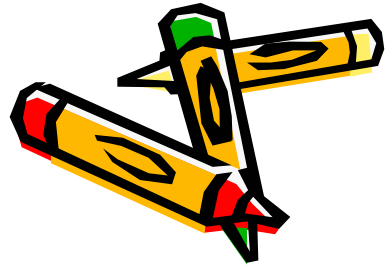
- Redefine how we think about biology and the workings of living systems
  - DNA → genome
  - mRNA → transcriptome
  - Metabolites → metabolome
  - All interactions between biomolecules → interactome
- System biology



Genomics

- Advances in 1990's
- Gene
  - Expressed sequence tag (EST)
  - Sequence database
- Information
  - Public accessible
  - Browser-based, user-friendly bioinformatics tools
- Oligonucleotide microarray (DNA chip)
  - PCR
  - Hybridization of oligonucleotides to complementary sequences

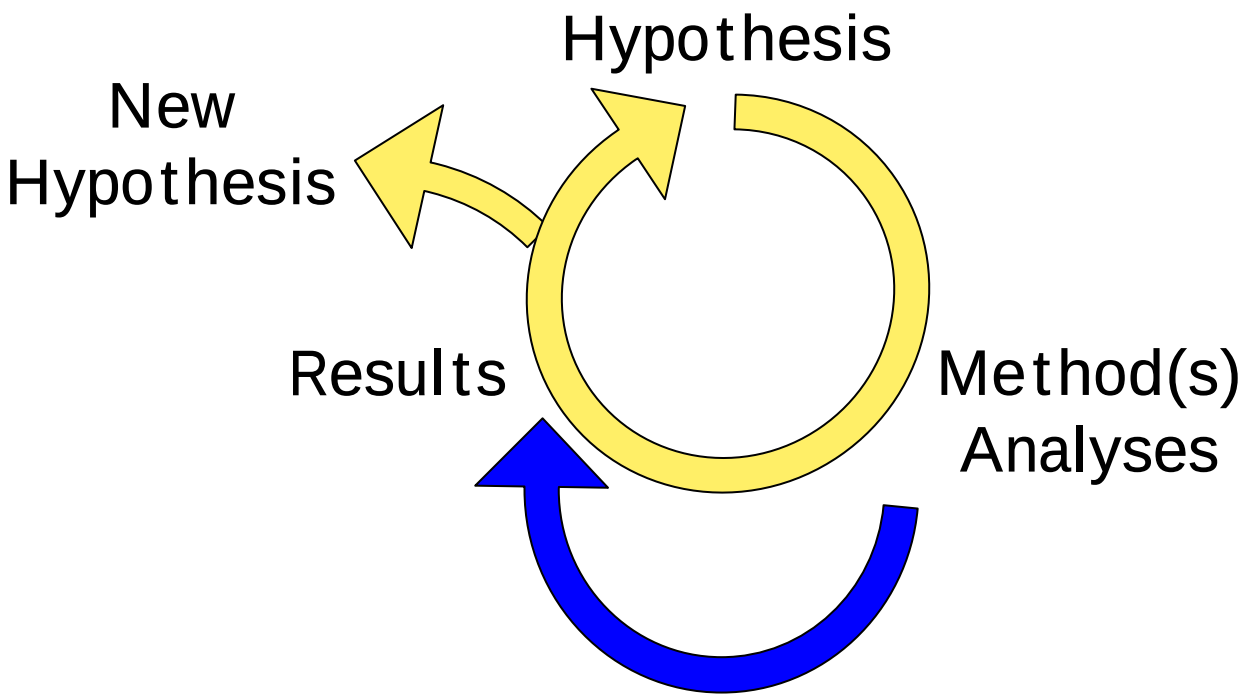
Proteomics



- An analytical challenge !!
- One genome ➔ many proteomes
  - Stability of mRNA
  - Posttranslational modification
  - Turnover rate
  - Regulation
- No protein-equivalent PCR
  - Protein does not replicate
- Proteins do not hybridize to complementary a.a. sequence
  - Ab-Ag

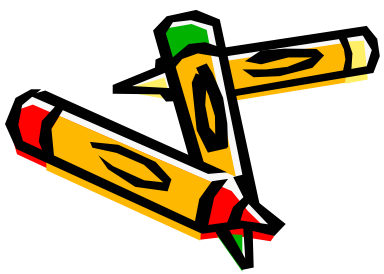
The Research Cycle

- Hypothesis driven research
- Method driven research
  - The “omic science”



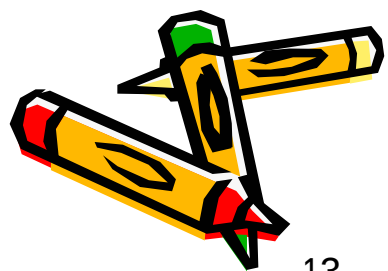
Questions to ask

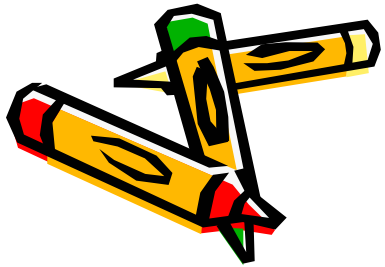
- What it is ?
  - Molecular function
  - Knockout
- Why is this being done ?
  - Biological process
  - Y2H
- Where is this ?
  - Cellular compartment
  - Immunofluorescence



New high-throughput strategies

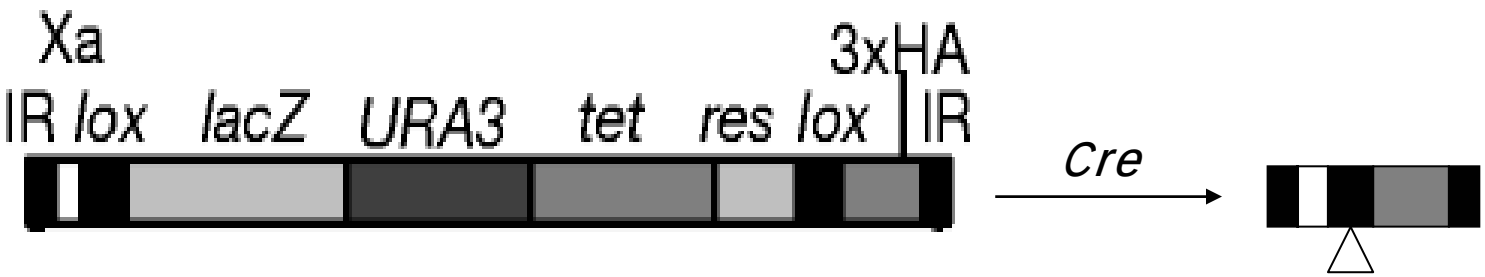
- What it is ?
  - Molecular function
  - *Genomic Knockout*
    - Random transposon tagging (yeast)
      - Michael Snyder at Yale
    - PCR based (bar code) mutagenesis (yeast)
      - Ron Davis at Standford
    - RNAi (*C. elegans*)
    - ???





Transposon

- Mobile pieces of DNA that can hop from one location in the genome to another.
- Jumping gene
- Tn3 derived from *E. coli* used in *Saccharomyces cerevisiae*

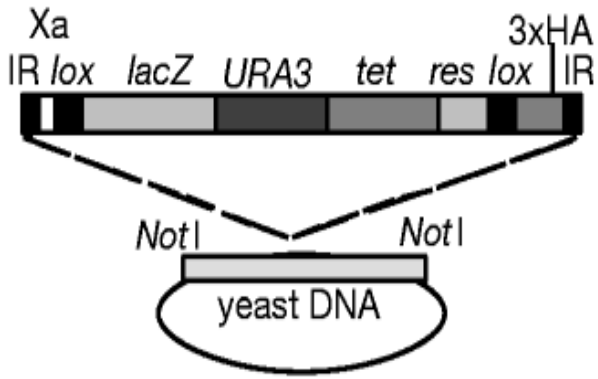


- A modified minitransposon (mTn3) by Michael Snyder at Yale
  - Why a *lac*Z without a promoter and start codon ?
  - Why *URA3* and *tet* gene in mTn3?
  - Why *lox* ? (Cre recombinase)
  - Significance of homologous recombination ?

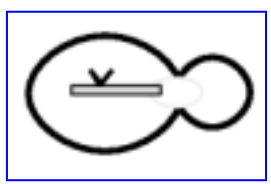
Ross-Macdonald et al., 1999 Nature 402, 413.

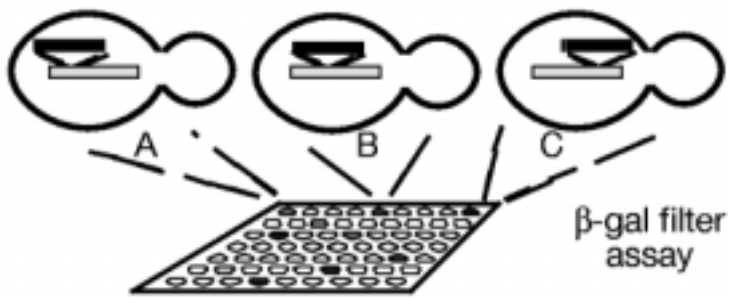
The mTn insertion project

- To create mutations:
  - A yeast genomic plasmid library in *E. coli* was randomly mutagenized by mTn insertion
    - Transformants were mated to initiate transposition
    - Transconjugates (tet^R) were mated again to resolve cointegrate structure.
  - The mutated gene is reintroduced to yeast by homologous recombination
    - Replace mTn-mutated gene with *wt* gene
    - In *URA*3-lacking strain



- Mutated yeast strains were analyzed:
  - Phenotypes
  - Structure-function relationships
  - Differential gene expression
  - Protein localization
- Results:
  - 11,232 strains turned blue



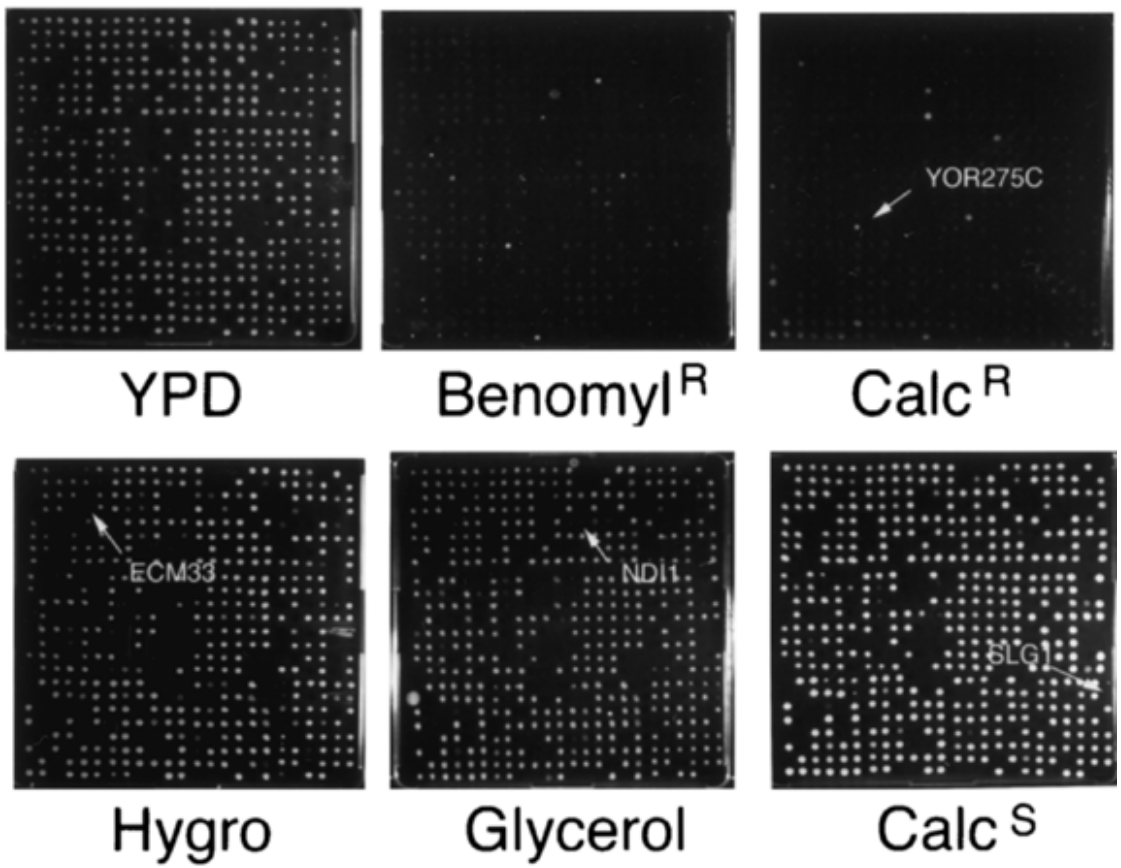


mTn approach to yeast genome

- 92,544 plasmid prep. & yeast transformation
- 11,232 colonies (strains) turned blue
- 6,358 strains sequenced
  - 1,917 different annotated ORFs
  - 328 non-annotated ORFs
    - “gene” = ORFs > 100 codons
- What’s next ?

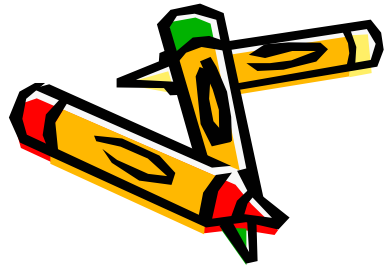
Phenotype macroarrays

- 7,680 mTn-insertion alleles were transformed into a haploid strain
  - 1,082 (14%) inviable → essential for viability
  - Phenotype screen: 96 strains x 6 = 576 strains

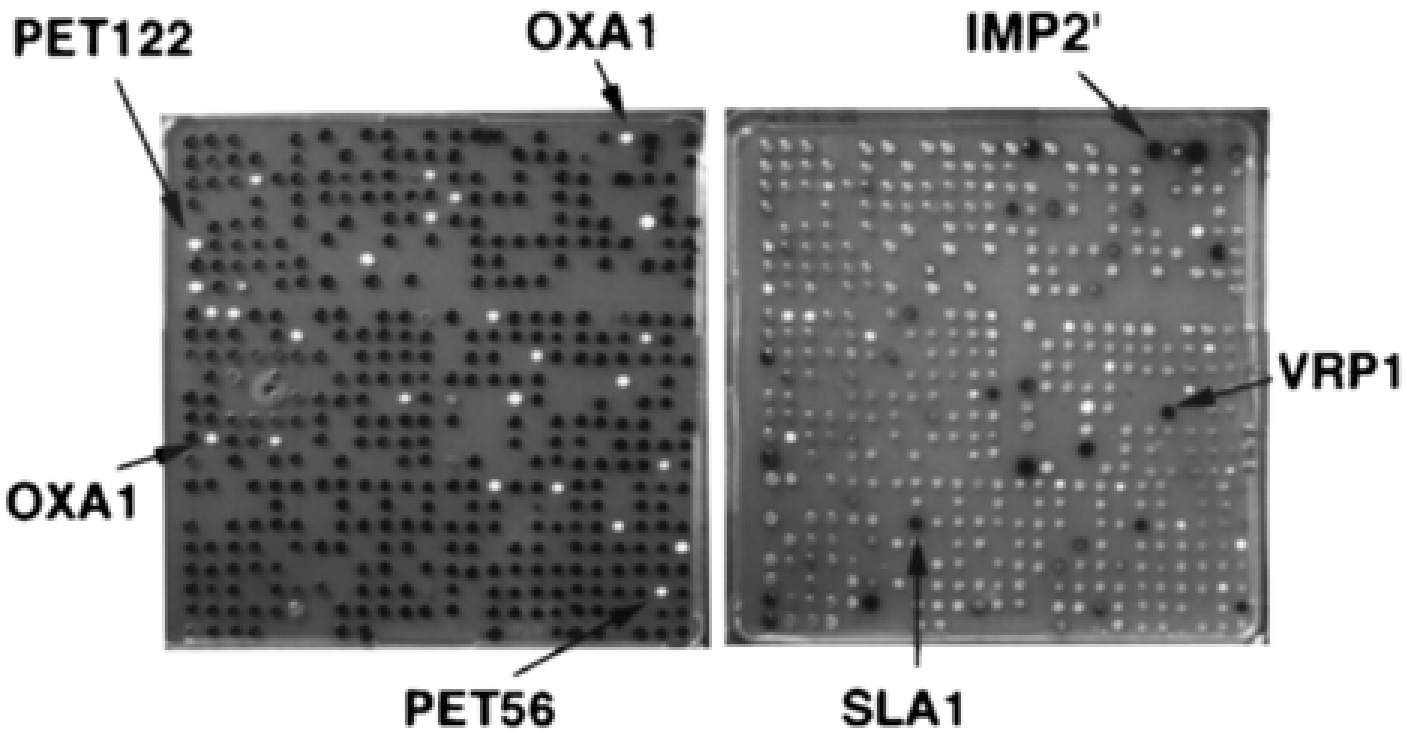


- YPD (normal growth medium)
- Benomyl (microtubule formation)
- Calcofluor (cell wall binding dye)
- Hygromycin (cell wall synthesis)
- Glycerol (a C source)

Array analysis



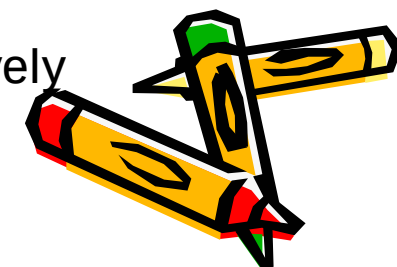
- Metabolic pathway
  - Oxidative phosphorylation → red vs. white (normal)
  - Alkaline phosphatase + BCIP → blue vs. white (normal)



Now what ?

- 576 strains, 20 growth conditions
- Data, data, data….
- Look for extremes ?
- Cluster
  - Group together similar patterns
  - Mathematical description
    - Co-expression
    - Standard correlation coefficient
  - Graphical representation
    - Original experimental observation
    - Color: dark and light
    - Visualize and understand the relationships intuitively

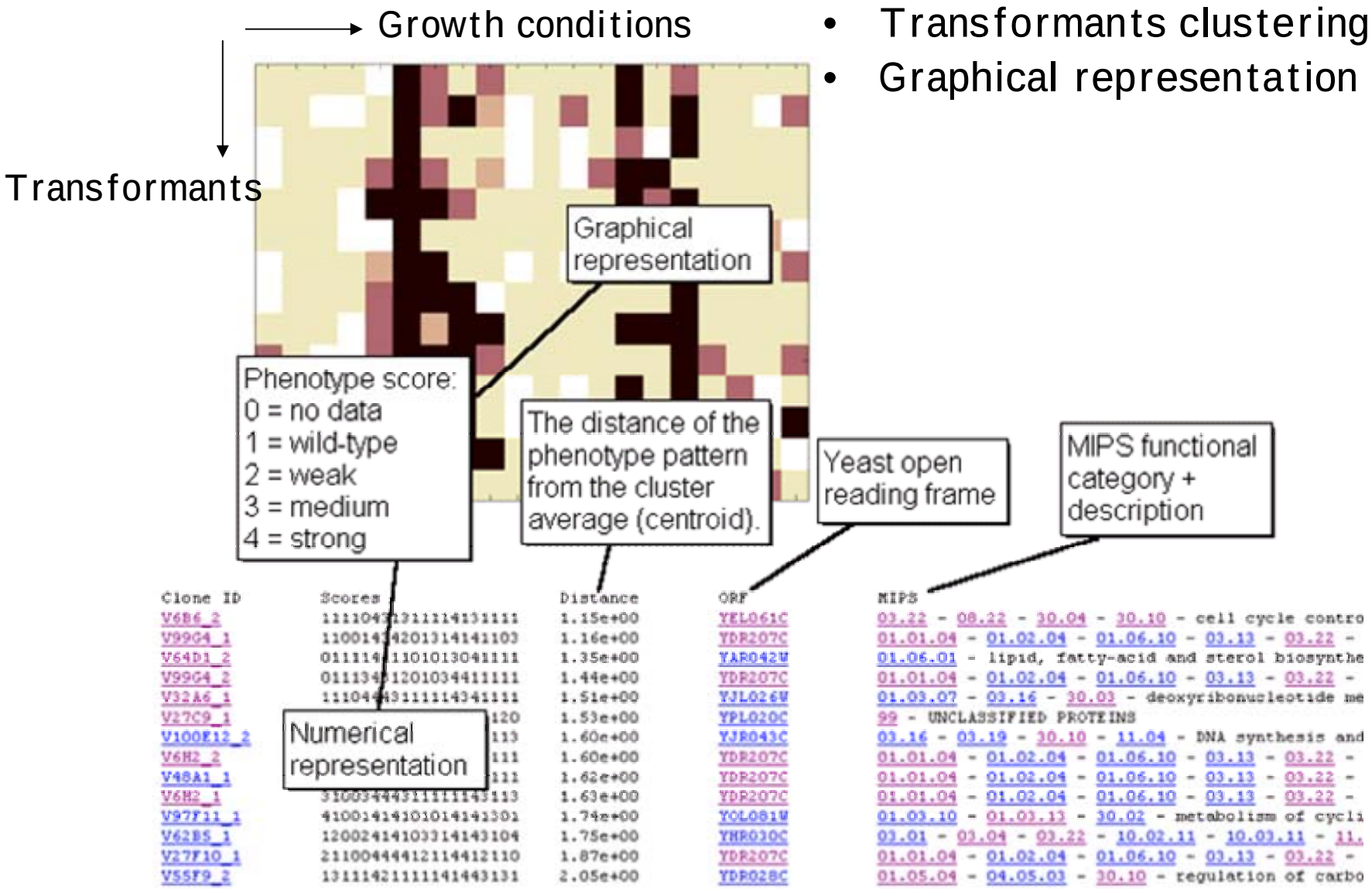
1,2,3,4,……,8,9,10



http://bioinfo.mbb.yale.edu/genome/phenotypes/

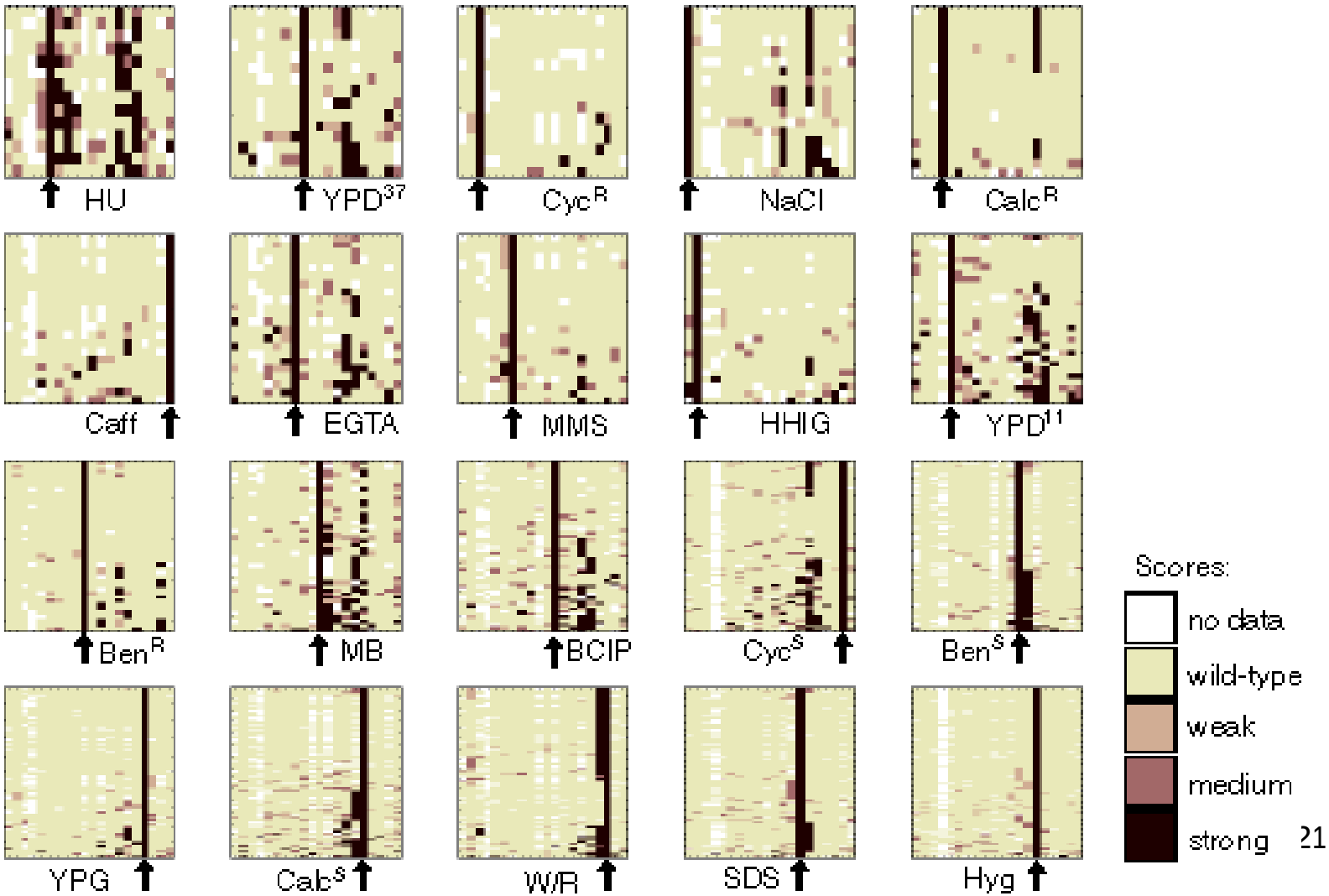
Data analysis

- Transformants clustering
- Graphical representation



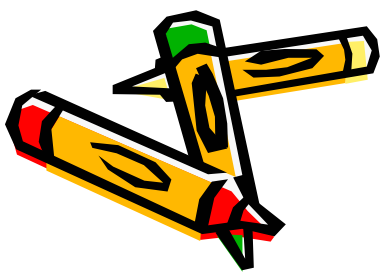
Data clusters

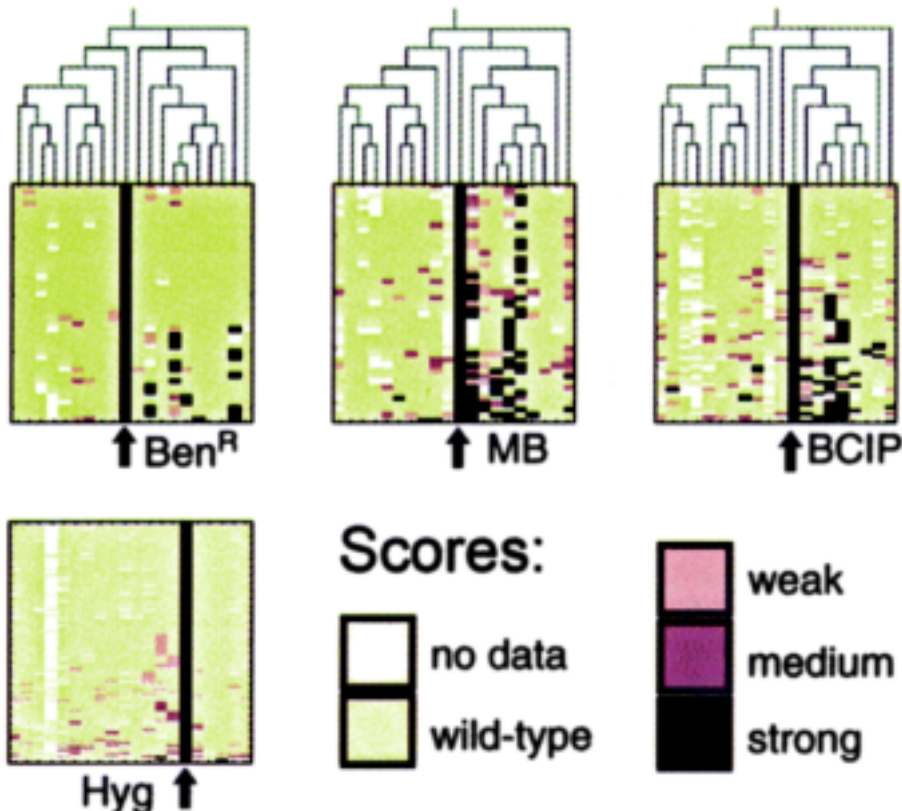
- 20 Growth conditions



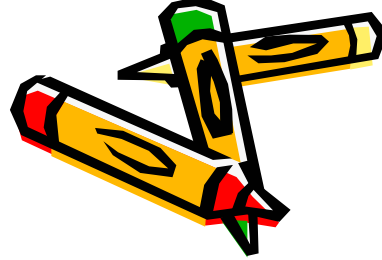
Double cluster

- Horizontal cluster: transformants
- Vertical cluster: growth conditions
- Identify assays for functionally related proteins





Growth condition cluster



- Clustering growth conditions that result in similar phenotypes
- More effective screening functionally related proteins

Table 1 Phenotypes scored by macroarray analysis

Assay/growth conditions	Number*	Colour code‡	Abbreviation§
YPD + 8 mM caffeine	27	Purple	Caff
Cycloheximide hypersensitivity: YPD + 0.08 μg ml^{-1} cycloheximide at 30 °C	28†	Pink	Cyc^{S}
White/red colour on YPD	39	Yellow	W/R
YPGlycerol	54	Yellow	YPG
Calcofluor hypersensitivity: YPD + 12 μg ml^{-1} calcofluor at 30 °C	[illegible]	Purple	Calc^{S}
YPD + 46 μg ml^{-1} hygromycin at 30 °C	136	Purple	Hyg
YPD + 0.003% SDS	109	Purple	SDS
Benomyl hypersensitivity: YPD + 10 μg ml^{-1} benomyl	67	Green	Ben^{S}
YPD + 5-bromo-4-chloro-3-indolyl phosphate at 37 °C	35	Purple	BCIP
YPD + 0.001% methylene blue at 30 °C	12	Purple	MB
Benomyl resistance: YPD + 20 μg ml^{-1} benomyl	11†	Green	Ben^{R}
YPD at 37 °C	29	Cyan	YPD^{37}
YPD + 2 mM EGTA	30	Black	EGTA
YPD + 0.008% MMS	16	Pink	MMS
YPD + 75 mM hydroxyurea	21	Pink	HU
YPD at 11 °C	20	Cyan	YPD^{11}
Calcofluor resistance: YPD + 0.3 μg ml^{-1} calcofluor at 30 °C	4†	Purple	Calc^{R}
Cycloheximide resistance: YPD + 0.3 μg ml^{-1} cycloheximide	2†	Pink	Cyc^{R}
Hyperhaploid invasive growth mutants	25	Orange	HHIG
YPD + 0.9 M NaCl	13	Black	NaCl

Cell wall biogenesis and maintenance

DNA metabolism

Discovery Questions

- What advantage is there to clustering the phenotypes in this manner?

- Some of the genes identified in this analysis had no known function. How can clustering these data help us predict possible functions?

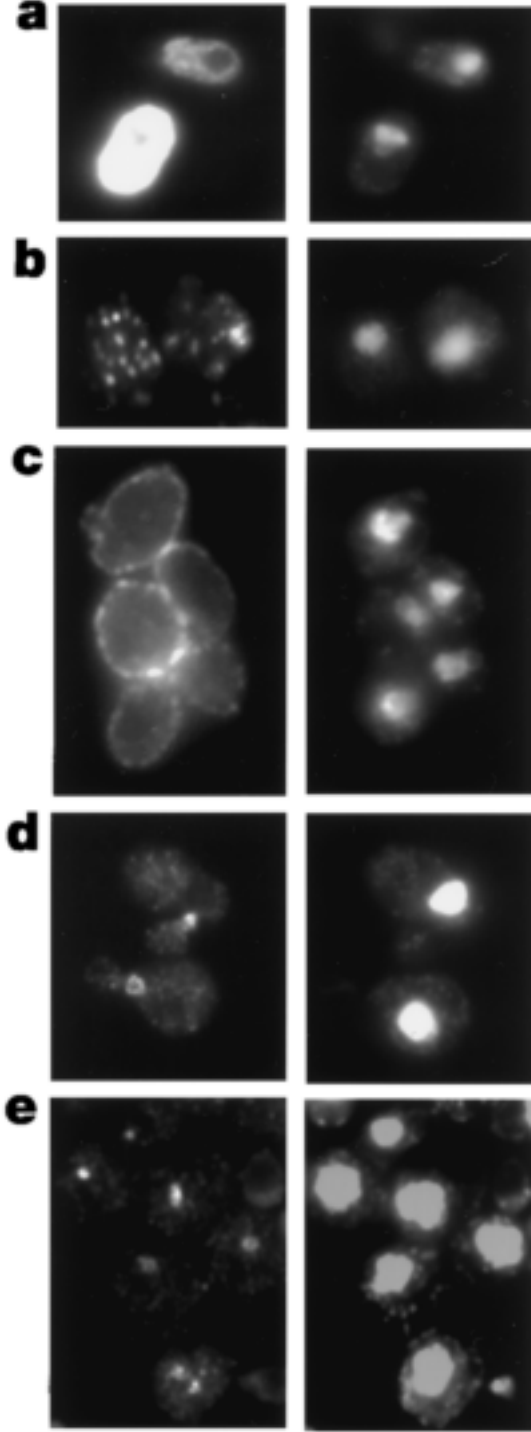
Cellular location

- HA-tag fusion protein and HA monoclonal Ab
- Immunofluorescence photomicrographs
  - Left: epitope tagged proteins
  - Right: DNA stained with DAPI

Table 2 Observed immunofluorescence patterns for in-frame HAT fusions

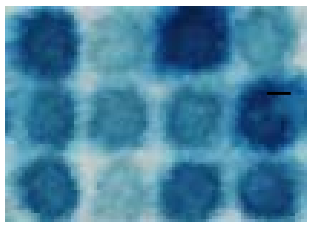
Pattern	Number*
Discrete	
Nuclear	
general	41
granular	64
nucleolus	22
Nuclear rim/endoplasmic reticulum	29
Mitochondrial	37
Spindle pole body/microtubules	5
Cell periphery	11
Cytoplasmic patch/dots	10
Cell neck	2
Vacuole	2
General cytoplasmic	
Uniform, finely speckled	3
Granular, fibrous	189
Background	925
Total	1,340

*All strains exhibiting discrete cellular or cytoplasmic localizations were tested at least twice. Complete data sets for all tested clones can be accessed at http://ycmi.med.yale.edu/ygac/triples.htm.

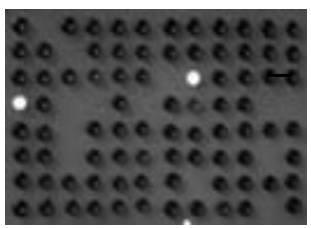


TRIPLES DATABASE YGAC

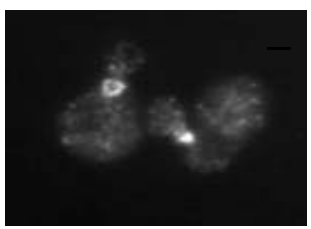
- TRIPLES
  - A database of TRansposon-Insertion Phenotypes, Localization, and Expression in Saccharomyces.
  - From a single transposon insertion event, three types of data concerning gene function are generated:



Using transposon-encoded *lacZ* as a reporter, we have performed β-gal filter assays to determine when a given gene is expressed during the yeast life cycle.



Transposon insertion results in truncation of the host gene, enabling us to determine disruption phenotypes on a large-scale by macroarray analysis.



The subcellular localization of transposon-tagged proteins has been analyzed using monoclonal antibodies directed against the mTn-encoded HA epitope.