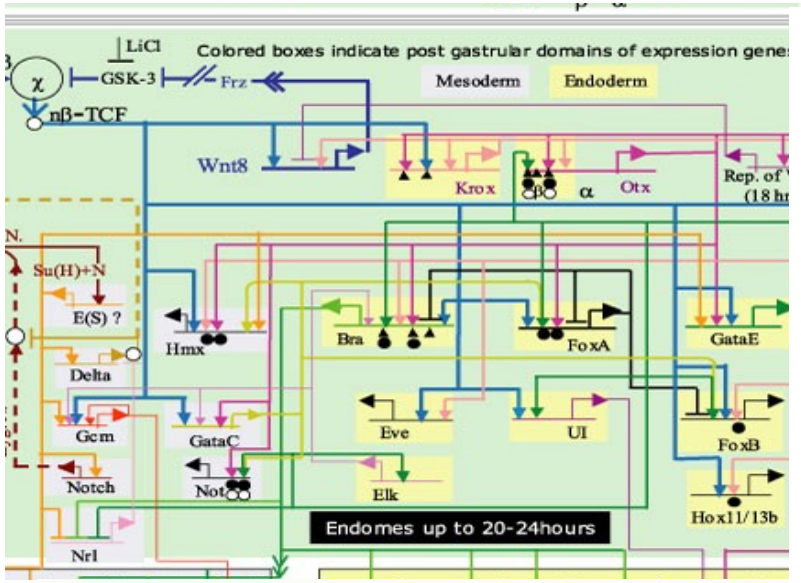
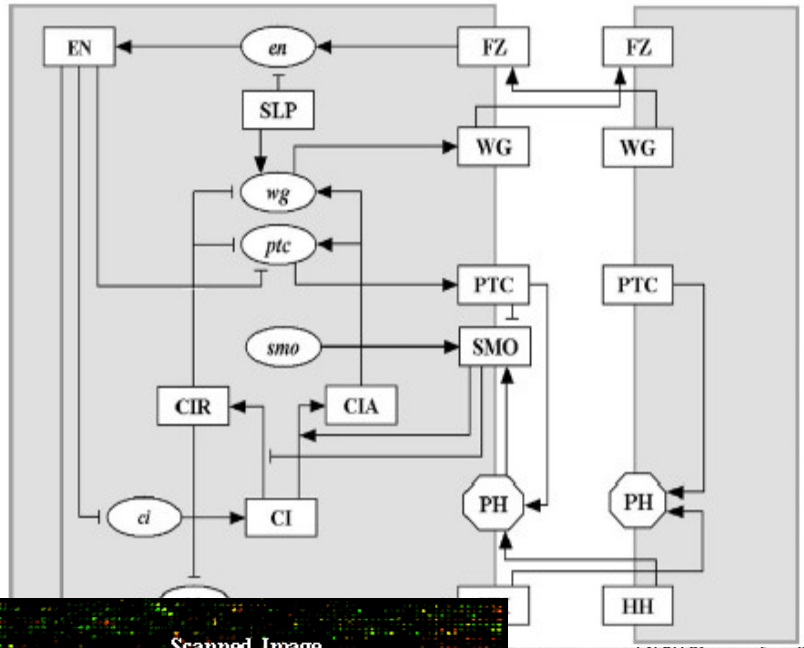
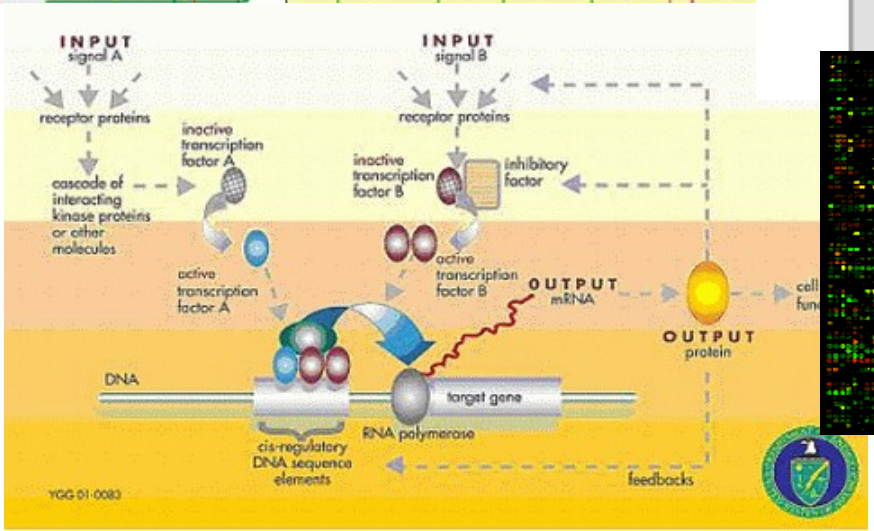
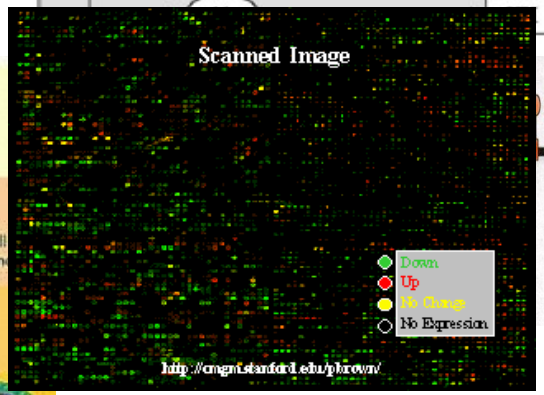
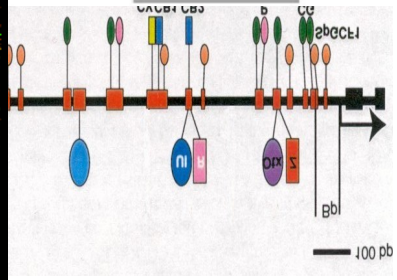
ECS 234 Lecture:
The Genome and Biotechnologies











Today:

- Basic Molecular Biology
  - Genes, Proteins, etc.
  - Genomes
  - Central Dogma
- Observing the Central Dogma
  - PCR - amplification
  - DNA Sequencing - genome
  - Microarrays - transcriptome
  - ChIP – Protein DNA interactions
  - Yeast Two-Hybrid – Protein Protein Interacts.

Biological Preliminaries

- Life is survival of information
- Properties of life:
  - Information exchange (communication)
  - Procreation (passing on information)
  - Evolution (change)

- A machine that's set in motion and never stops

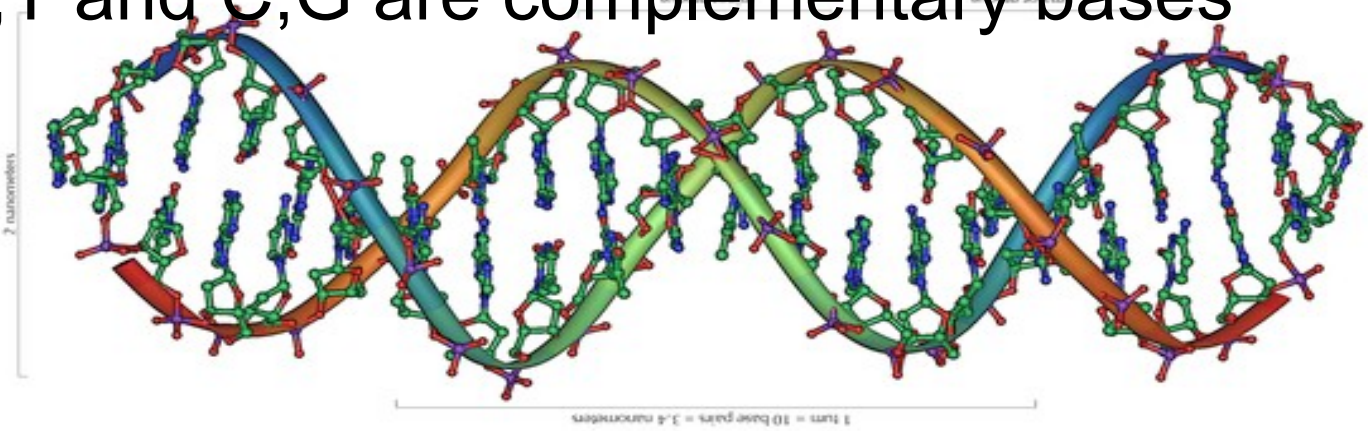
Preliminaries

- Top-down Organization of life
  - Social groups etc.
  - Organisms
  - Species, etc.
  - Organs, Tissues
  - Cells: units of life
  - Organelles etc.
  - Molecules of life: DNA, RNA, proteins

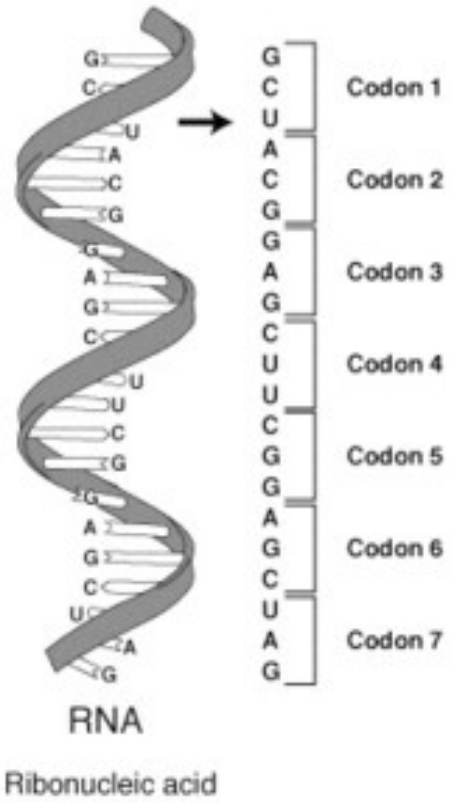
Preliminaries

- DNA: Inheritable information
- Units of Inheritance: genes, regulatory regions, ?
- Proteins: Day-to-day footwork
- Both are complex polymer molecules

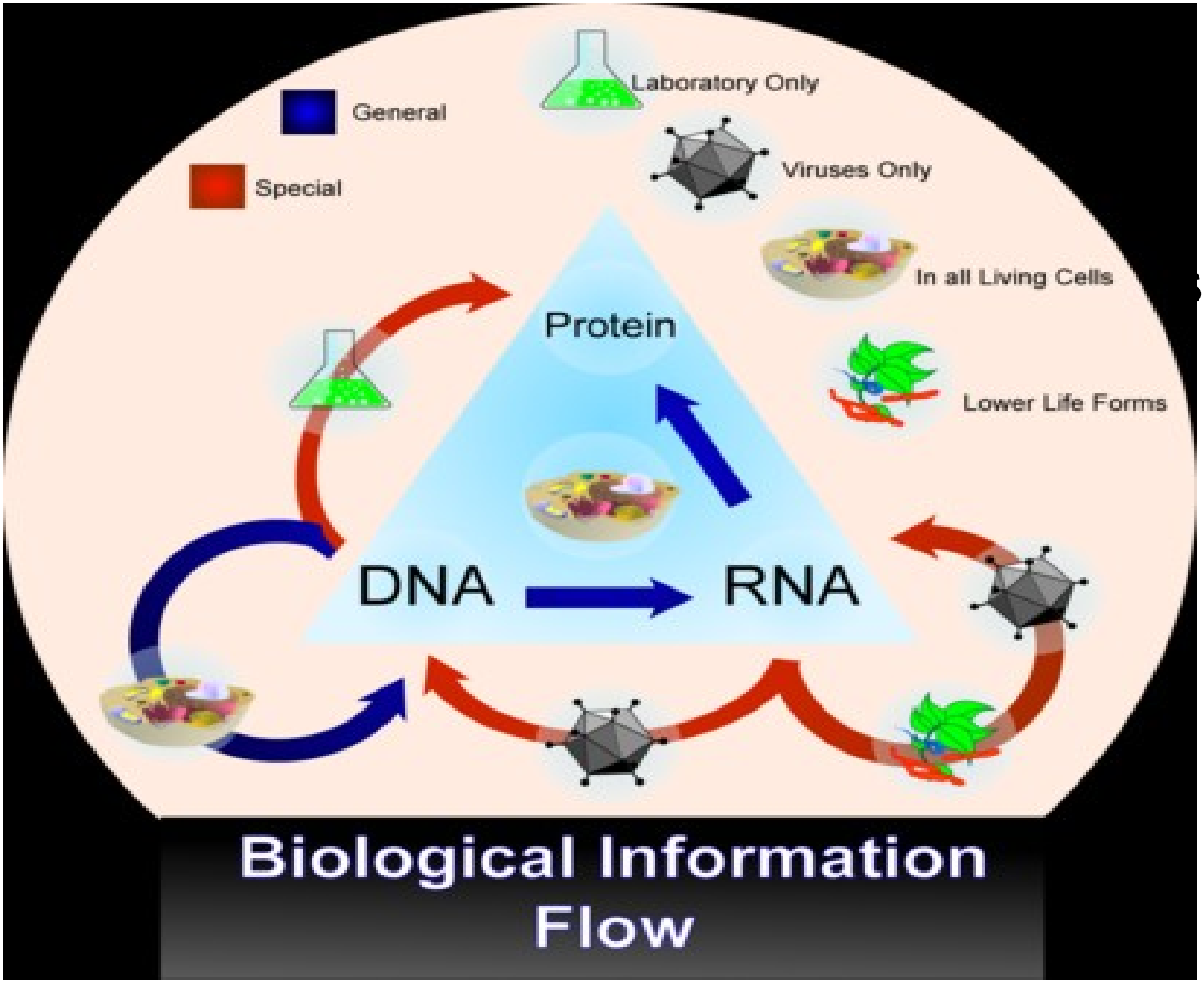
- DNA: String over the alphabet {A,C,G,T}
- A,T and C,G are complementary bases



Genes to Proteins



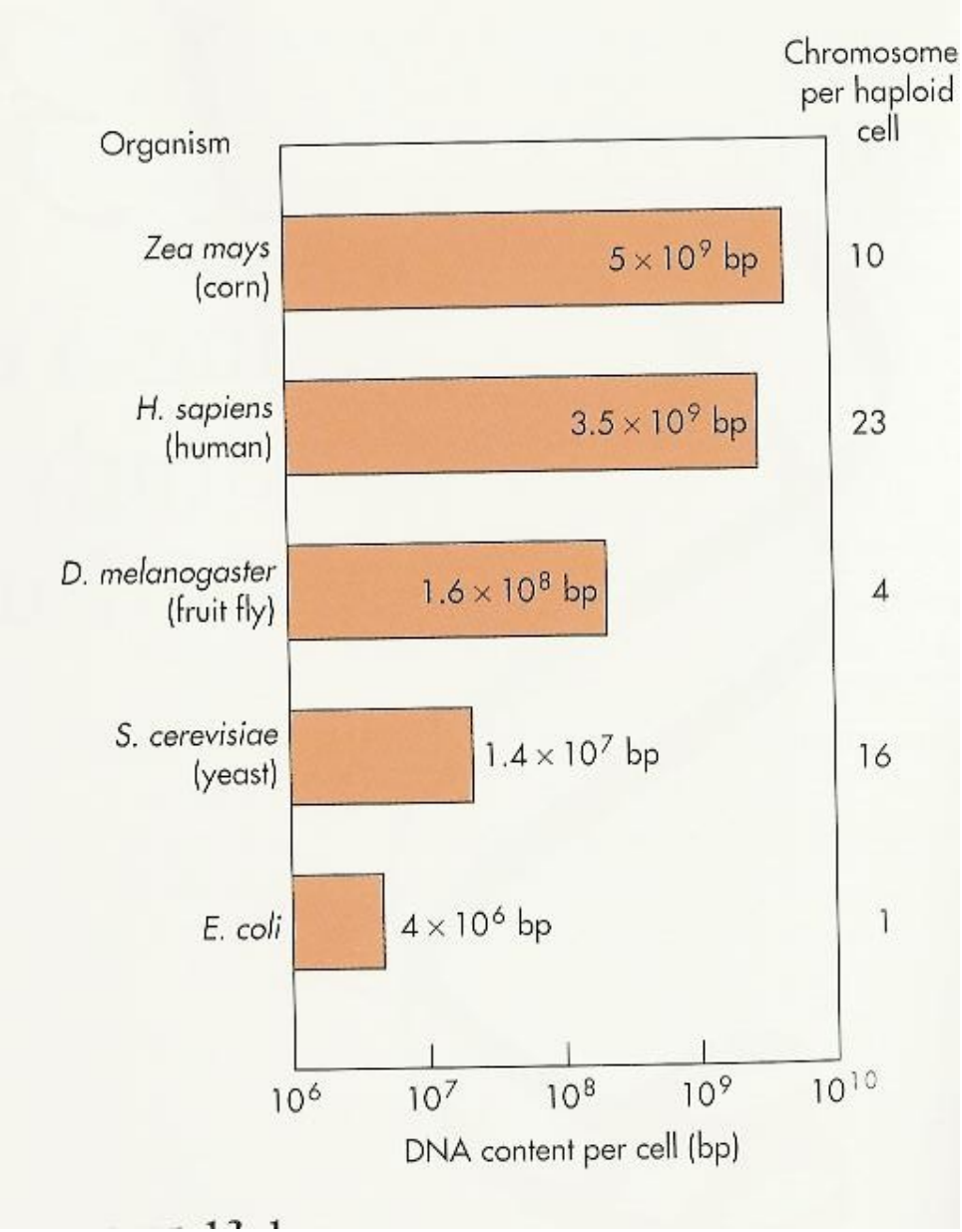
First Base of Codon	Second Base of Codon: U		C		A		G		Third Base of Codon
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC	phe	UCC	ser	UAC	tyr	UGC	cys	C
	UUA	leu	UCA	ser	UAA	STOP	UGA	STOP	A
	UUG	leu	UCG	ser	UAG	STOP	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC	leu	CCC	pro	CAC	his	CGC	arg	C
	CUA	leu	CCA	pro	CAA	gln	CGA	arg	A
	CUG	leu	CCG	pro	CAG	gln	CGG	arg	G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC	ile	ACC	thr	AAC	asn	AGC	ser	C
	AUA	ile	ACA	thr	AAA	lys	AGA	arg	A
	AUG	met*	ACG	thr	AAG	lys	AGG	arg	G
	* = START								
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC	val	GCC	ala	GAC	asp	GGC	gly	C
	GUA	val	GCA	ala	GAA	glu	GGA	gly	A
	GUG	val	GCG	ala	GAG	glu	GGG	gly	G


Laboratory Only
General
Viruses Only
Special
In all Living Cells
Protein
Lower Life Forms
DNA
RNA
Biological Information Flow

Genomes

Organization and complexity

- Genomes are the union of all DNA in an organism (there are different types of DNA: nuclear and mitochondrial)
- Only small % (2%) of the human genome is genes. The rest contains various promoter regions and “junk?” (>50%)
- Genome sizes vary among organisms, shortest for Phages and Viruses, longest for mammals and some plants (figure from Baldi)


Chromosome per haploid cell
Organism
Zea mays (corn)
5 × 10^9 bp
10
H. sapiens (human)
3.5 × 10^9 bp
23
D. melanogaster (fruit fly)
1.6 × 10^8 bp
4
S. cerevisiae (yeast)
1.4 × 10^7 bp
16
E. coli
4 × 10^6 bp
1
10^6
10^7
10^8
10^9
10^10
DNA content per cell (bp)

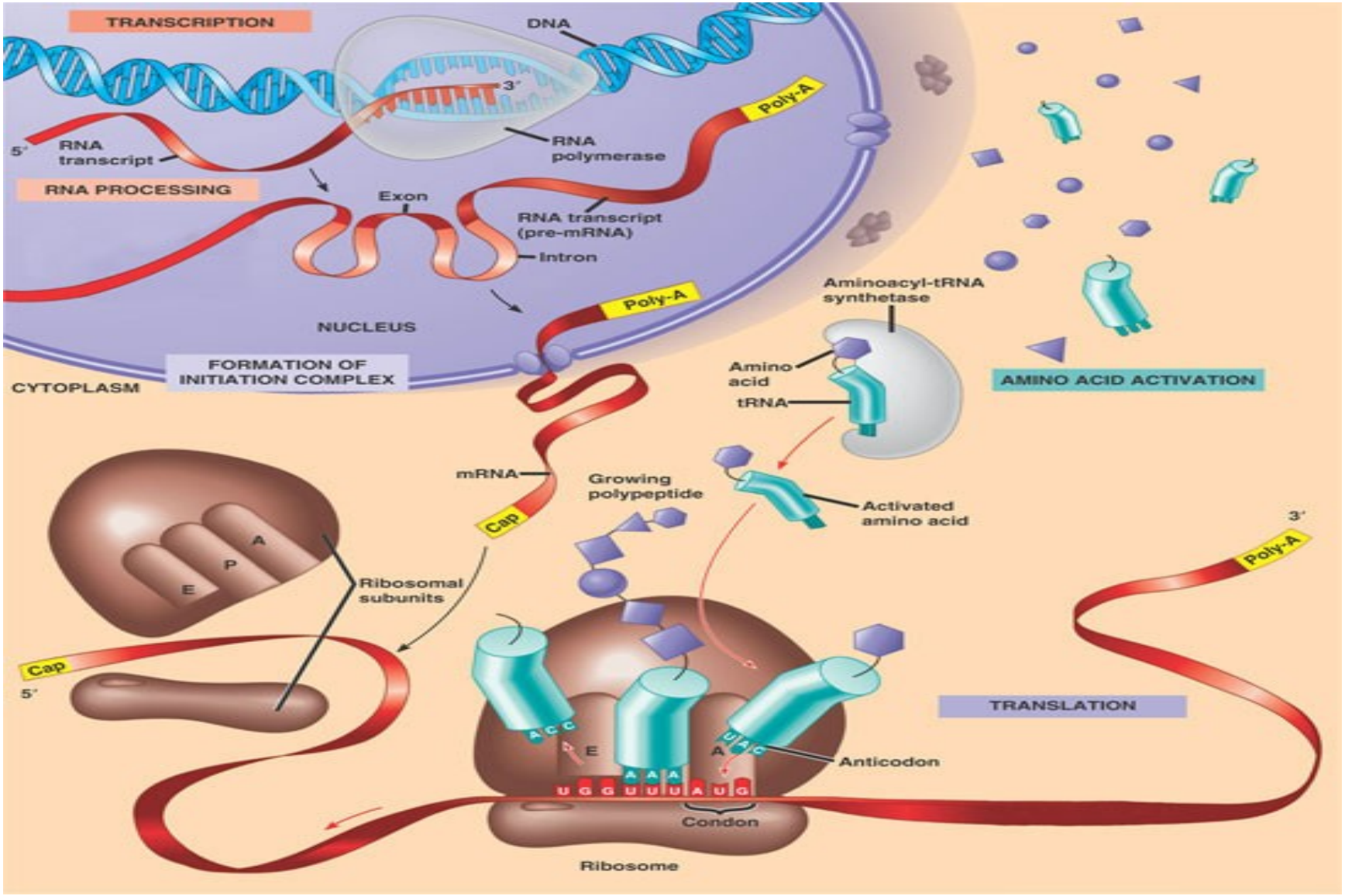
Evolution

- Changes in the genomes
- Mutations: changes in genome driven by random or particular events. Can be single base change or larger events.
- Recombination: mixing of genomes to produce a new one
- Natural selection: beneficial changes are passed on

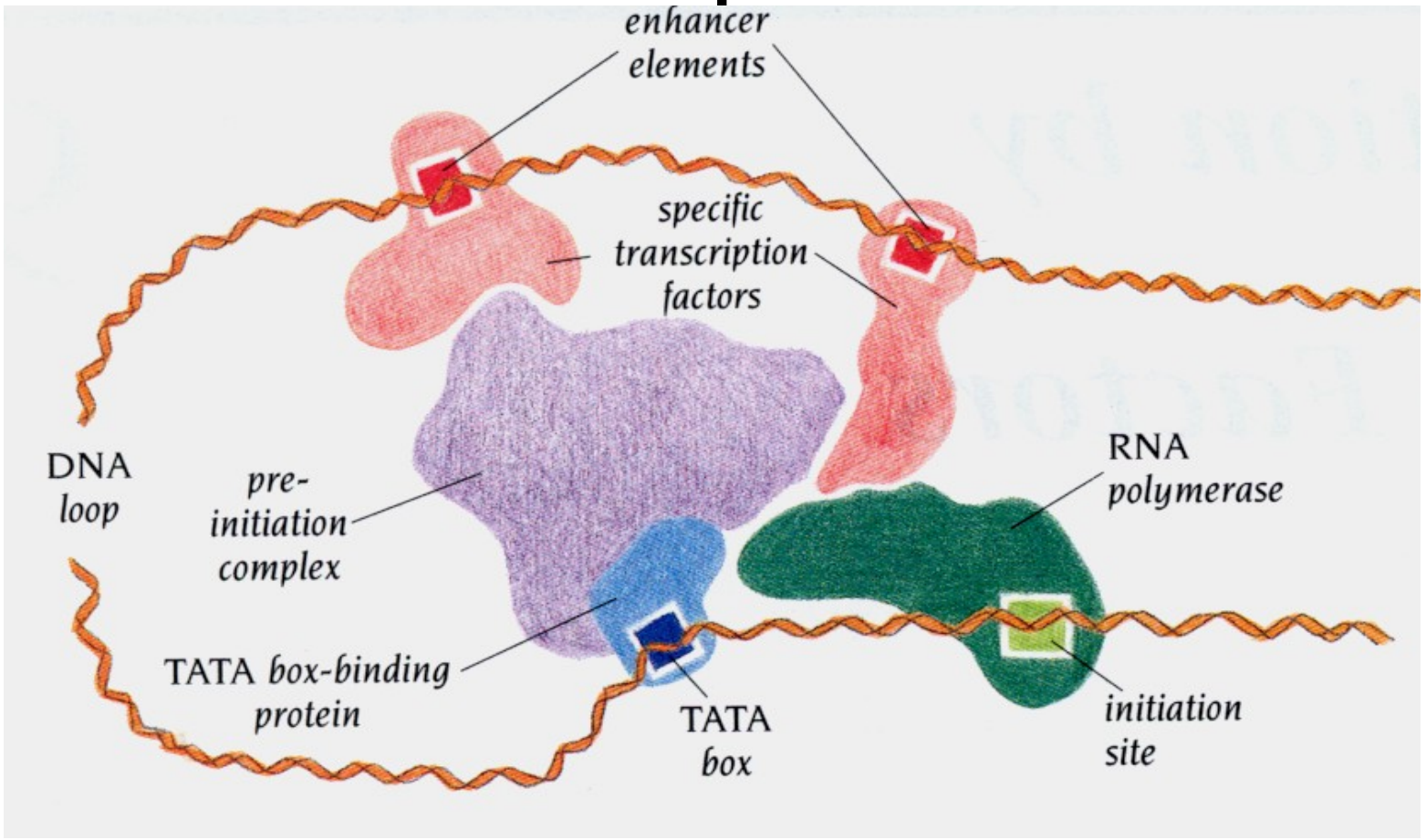
Similarity of genomes (i.e. organisms)

- Evolution implies that different organisms would have common ancestors
- Thus similarity comparisons (homology searches) provide clues to evolutionary ancestry (mention phylogeny)

Central Dogma of Molecular Biology

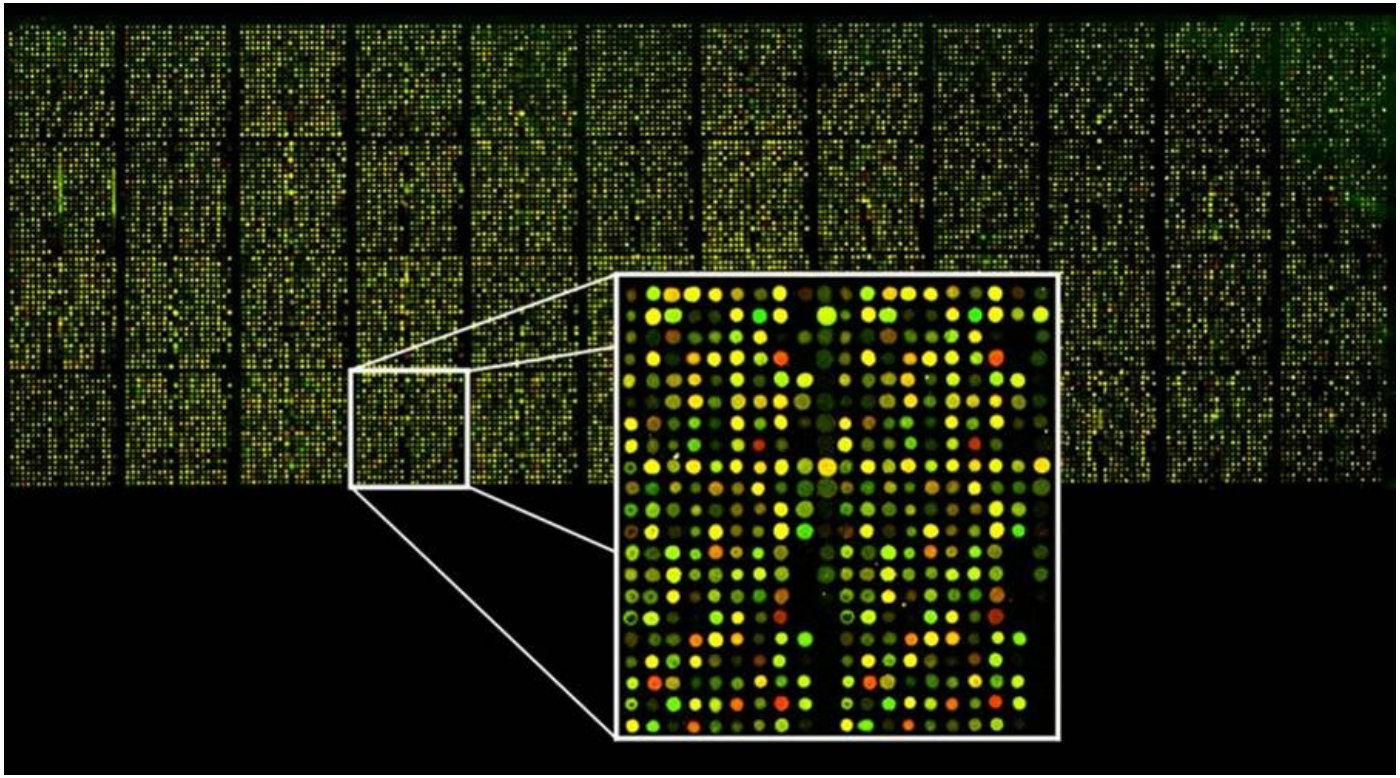


Transcriptional Regulation of Gene Expression



Transcriptome

All possible gene expressions in the organism



Organization and Complexity

- Transcriptome is the measurable level of all different mRNA's in an organism
- One DNA template multiple mRNAs: alternative splicing
- DNA to mRNA: one way street because of alternative splicing
- The "when and where" of mRNA concentration is coded in the promoter regions, and possibly elsewhere

Evolution

- Evolution of gene expression under emergent properties like network organization

Similarity of Organisms

- Comparison of gene expression from a "system's perspective"

Proteome

Localization, abundance, and interaction of all proteins in an organism

- Structure: Amino acid sequence, 3D crystal structure
- Structure => Function?
- Sequence homology not always good indicator of functional similarity
- Study of protein expression

Other -omes

- Interactome
- Cellome
- Biome
- Metabolome
- Envirome
- ?

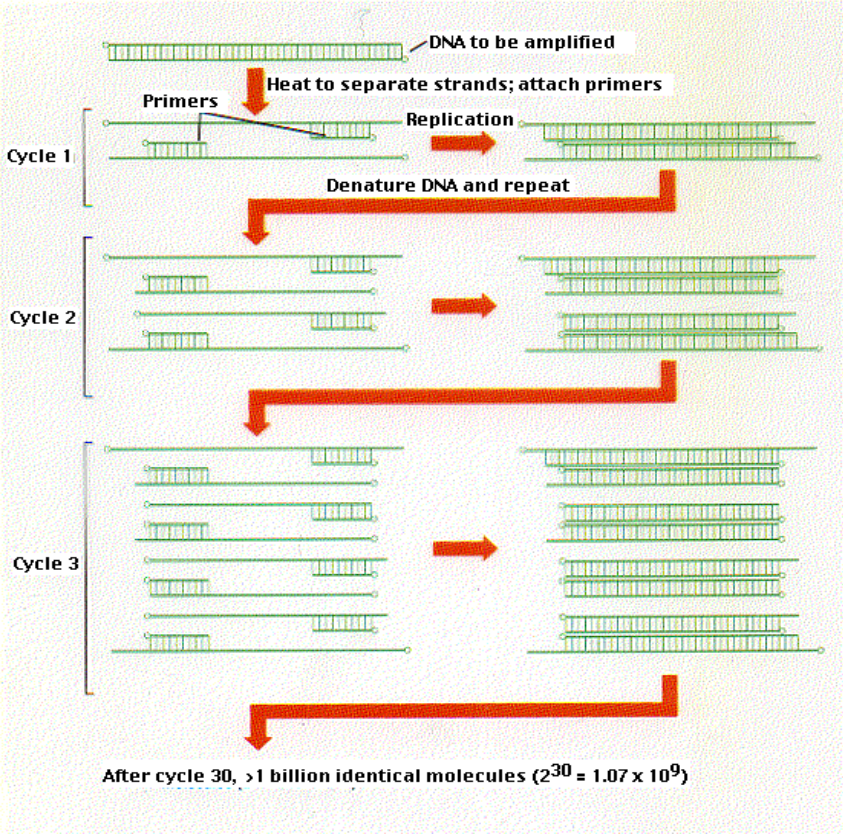
BioTechnologies

- Observing the Central Dogma
  - seeing the minute: PCR
  - DNA growing and sequencing
  - gene expression: microarrays
  - DNA-protein interactions: immunoprecipitation
  - protein-protein interactions: yeast 2 hybrid
- Large-Scale Technologies:
  - Thousands of measured variables
  - Require computational processing

1. Seeing the minute: PCR

Producing multiple copies of given DNA fragment (amplification)

1. Start with a double stranded DNA molecule
2. Separate strands into templates by heating the mixture
3. Cool to allow “primers” to attach to single strands
4. The primers identify the starting points for DNA synthesis
5. DNA synthesis of strands complementary to the templates
6. Go to 1.



PCR properties

- The primers can determine the amplified DNA fragment if chosen to flank that region
- n steps of the above produce 2^n copies of the intended DNA fragment

  2^30 ~ 10^9

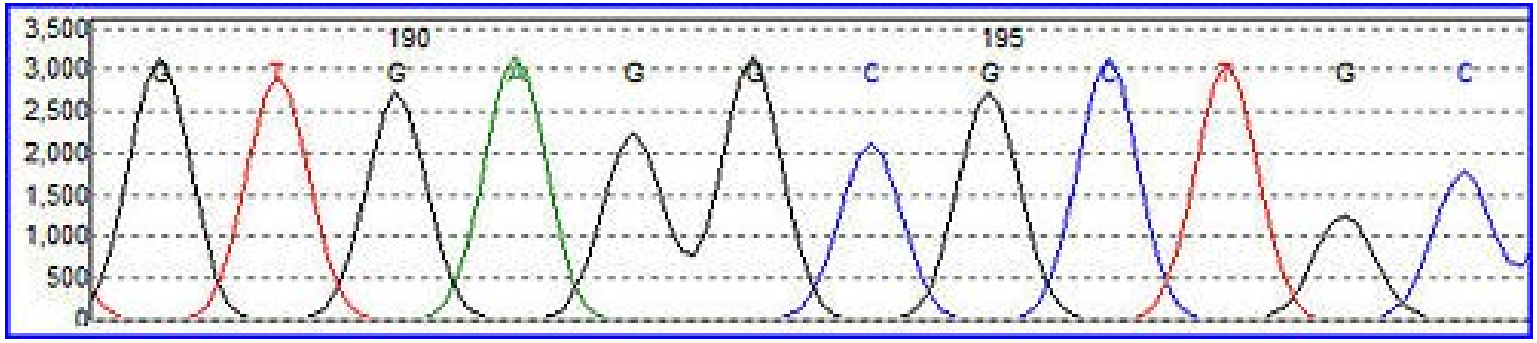
2. Growing DNA: Synthesizers

ABI 3900 High-Throughput DNA Synthesizer

3. DNA Sequencing

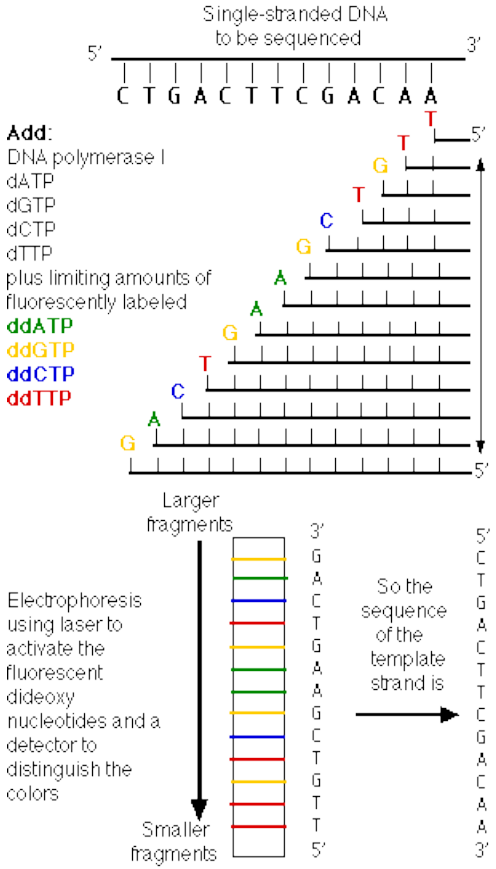
Reading the string: determining the exact positions of the base pairs A, C, G, T

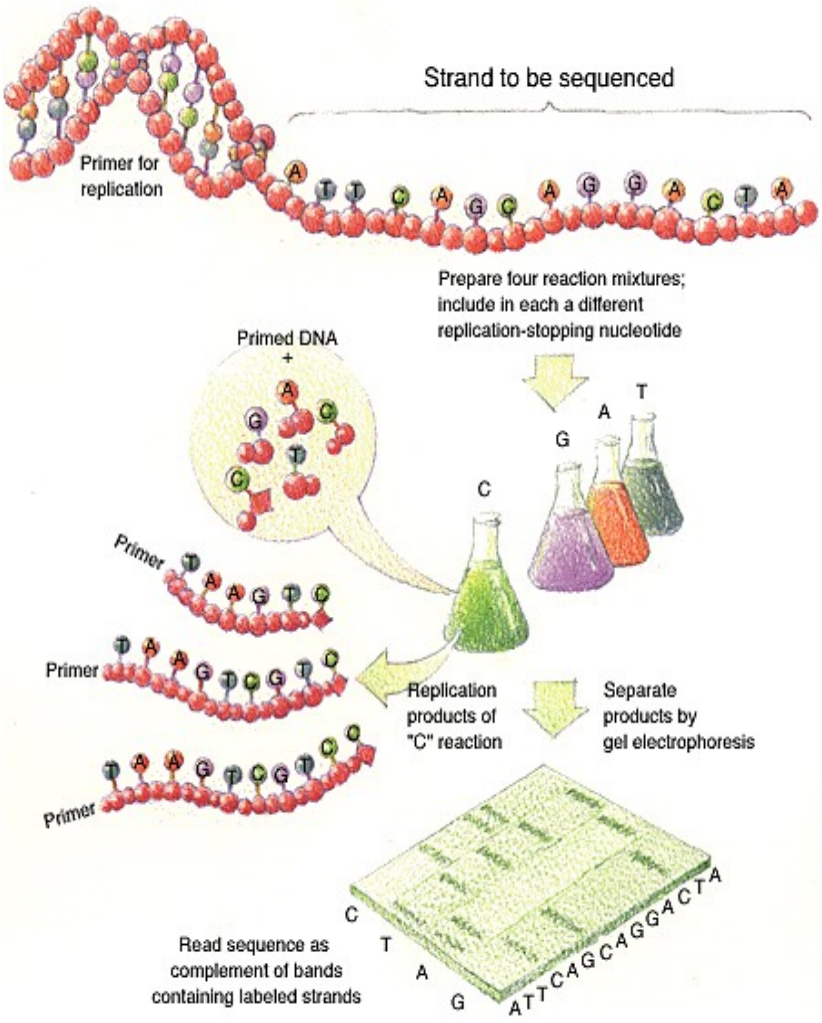
- Break DNA into manageable fragments (500 – 700 bp)
- Sequence the fragments
- Assemble the fragments

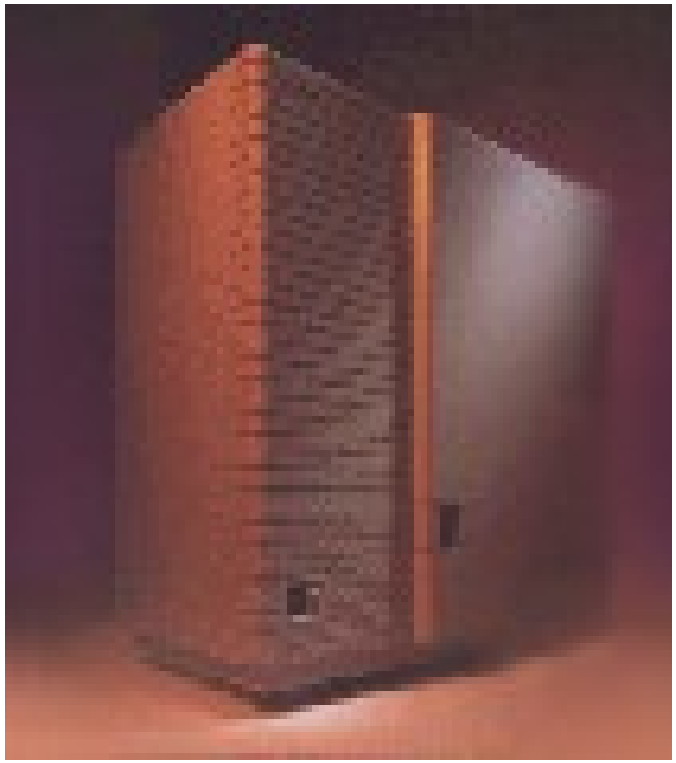


Sequencing Fragments

- Digest the DNA to be sequenced into small, 500 -700 bp fragments
- Replicate sample (fragment) into four bins
- Each bin has a sufficient amount of all four bases and Polymerase
- Bin associated with base x has in addition a special version of x, a stopping version, which stops replication
- The stopping bases are also fluorescently labeled
- DNA replication creates fragments of different lengths in the bins, but all fragments in a bin end in the same labeled base
- Using Gel electrophoresis the fragments are separated by length, thus identifying the base at any given length.





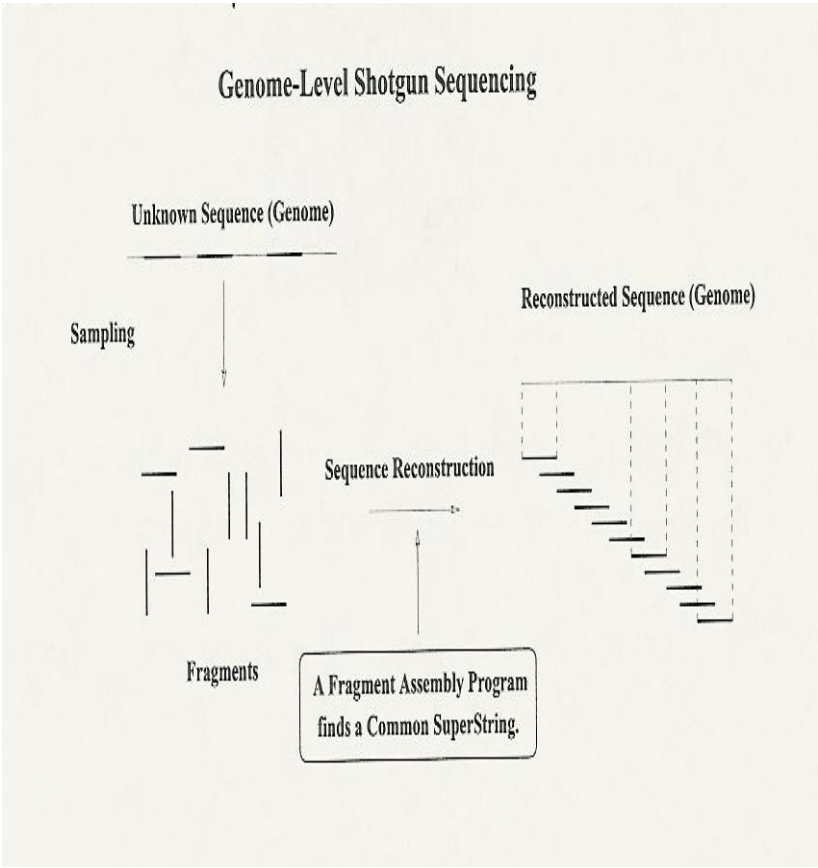


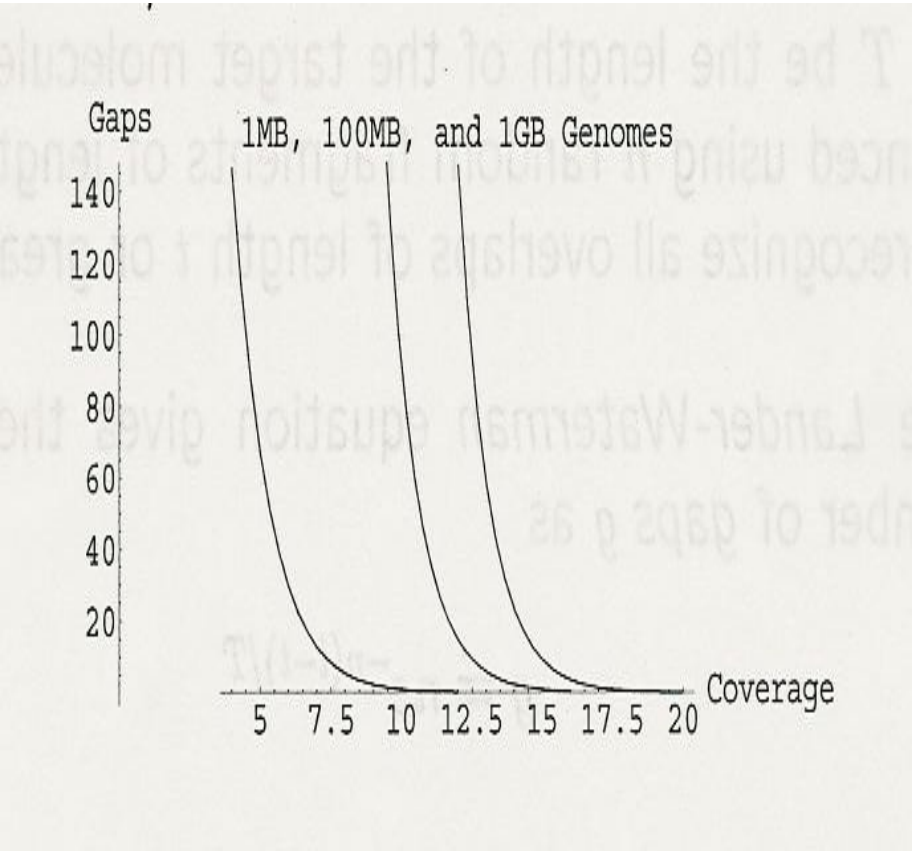
ABI Prism 377, modern capilary sequencer

Sequencing approaches

- Shotgun sequencing: “random” overlapping fragments (Celera)
- Mapped sequencing: shorter sequences are anchored (Human Genome Consortium)
- New techniques: 454, Illumina, Polonies

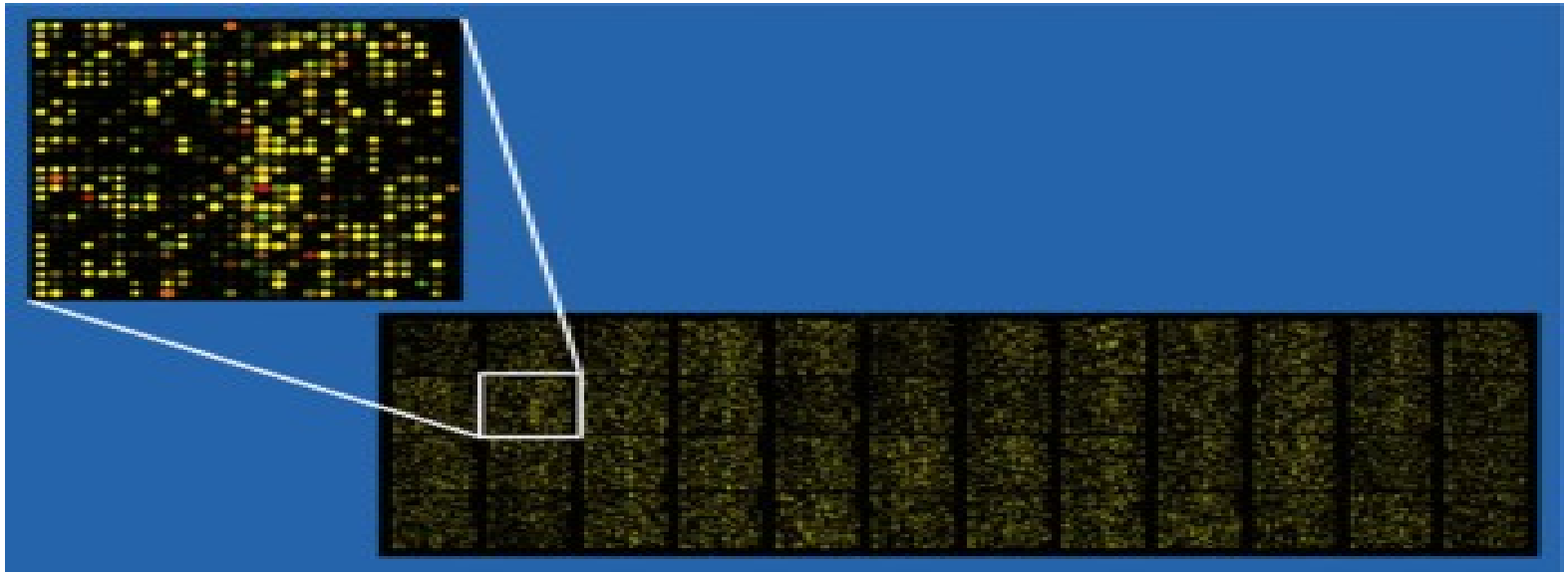
Fragment assembly and coverage





4. Microarrays

- Testing for the presence and concentration of mRNAs, Proteins, or Protein-DNA interactions
- Can do 10^6 experiments at a time!
- Hypothesis generation vs. promise of complete description on a large scale



What are Microarrays good for?

- Gene Expression Profiling

  - Genes that behave differently to treatments in same organisms

  - Different organisms
- Identifying naturally oscillating genes in the cell: example cell cycling genes in yeast
- Identifying SNPs
- Tumor vs normal cells
- Transcription Factor Regulation

Microarray Formats

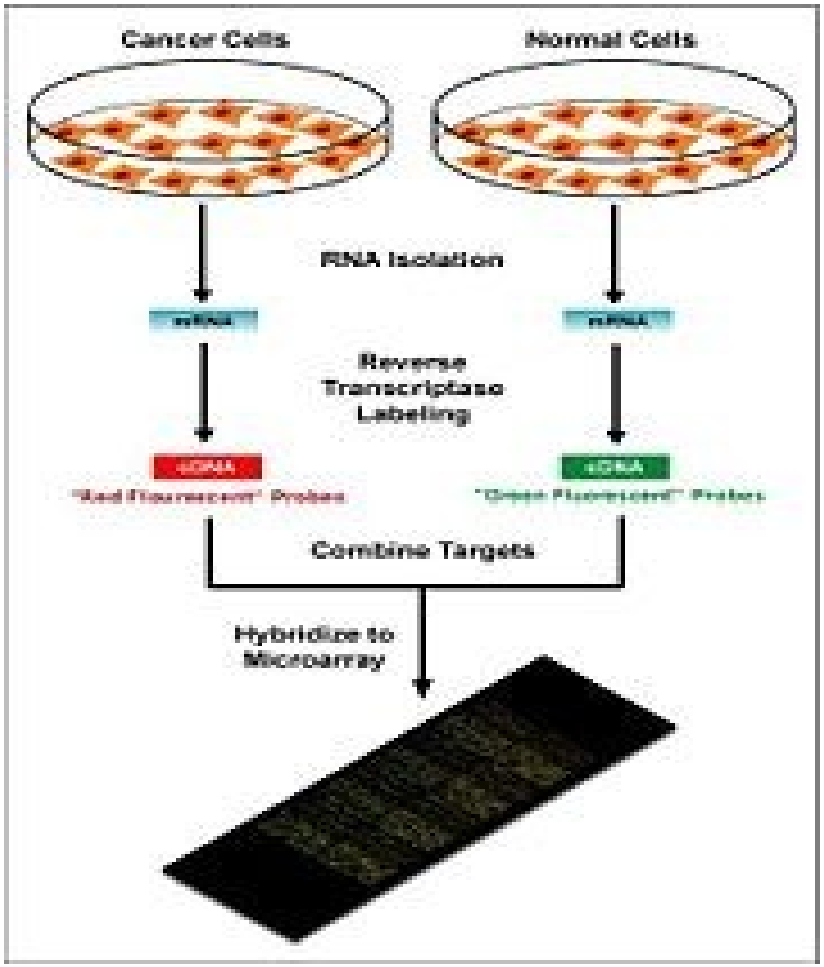
Spotted Microarrays

- Ed Southern 25 years ago, Patrick Brown recently
- Glass slide DNA arrays: 100,000 sites per 1cm^2

Oligonucleotide arrays

- Photolithographic method (Affymetrix Inc.) just like computer chips
- Masks used to synthesize oligonucleotides to a chip 1,000,000 sites per 1cm^2.

Two-color arrays vs one



Other Microarray Technologies

- Photolytography
- Ink jet (Agilent),
- Addressable beads (Lynx),
- etc.

Sources of Error in Microarrays

- Length of probes
- Cross and self hybridization
- Environmental conditions
- Array (plate) and dye idiosyncrasies
- other?

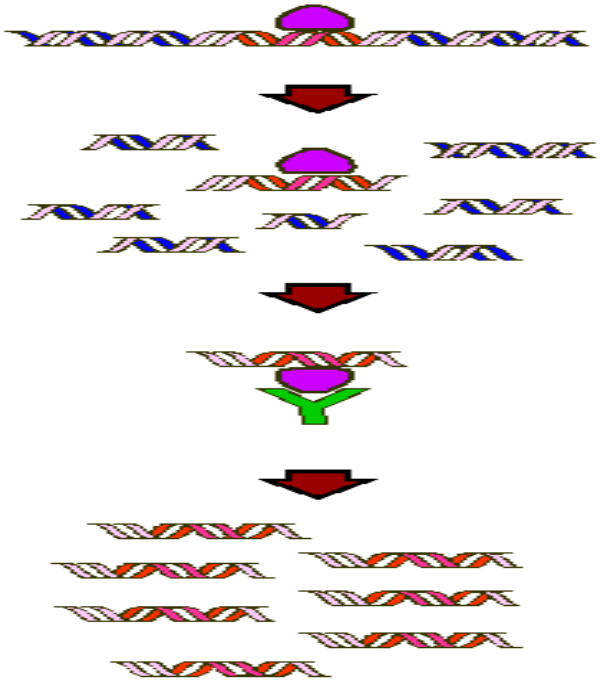
Algorithmic Problems

- Probe design
- Plate design
- Data Analysis:
  - over/under expression
  - classification,
  - clustering,
  - regulation inference,
  - gene networks

5. ChIP

ChIP: Chromatin Immuno-Precipitation

Resolves: DNA-Protein Interactions



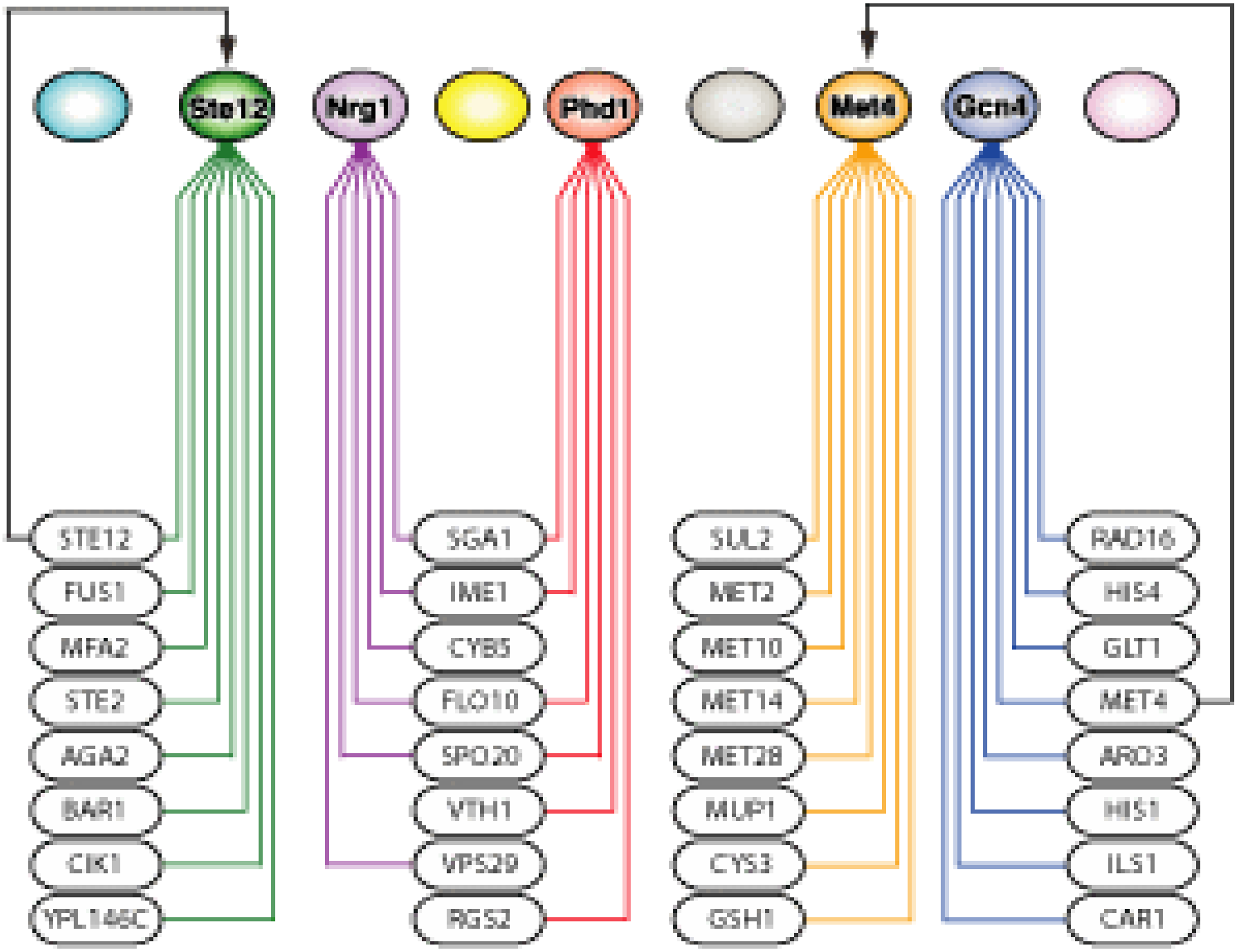
Protein bound to DNA

Sheared DNA

Antibody binding and protein release

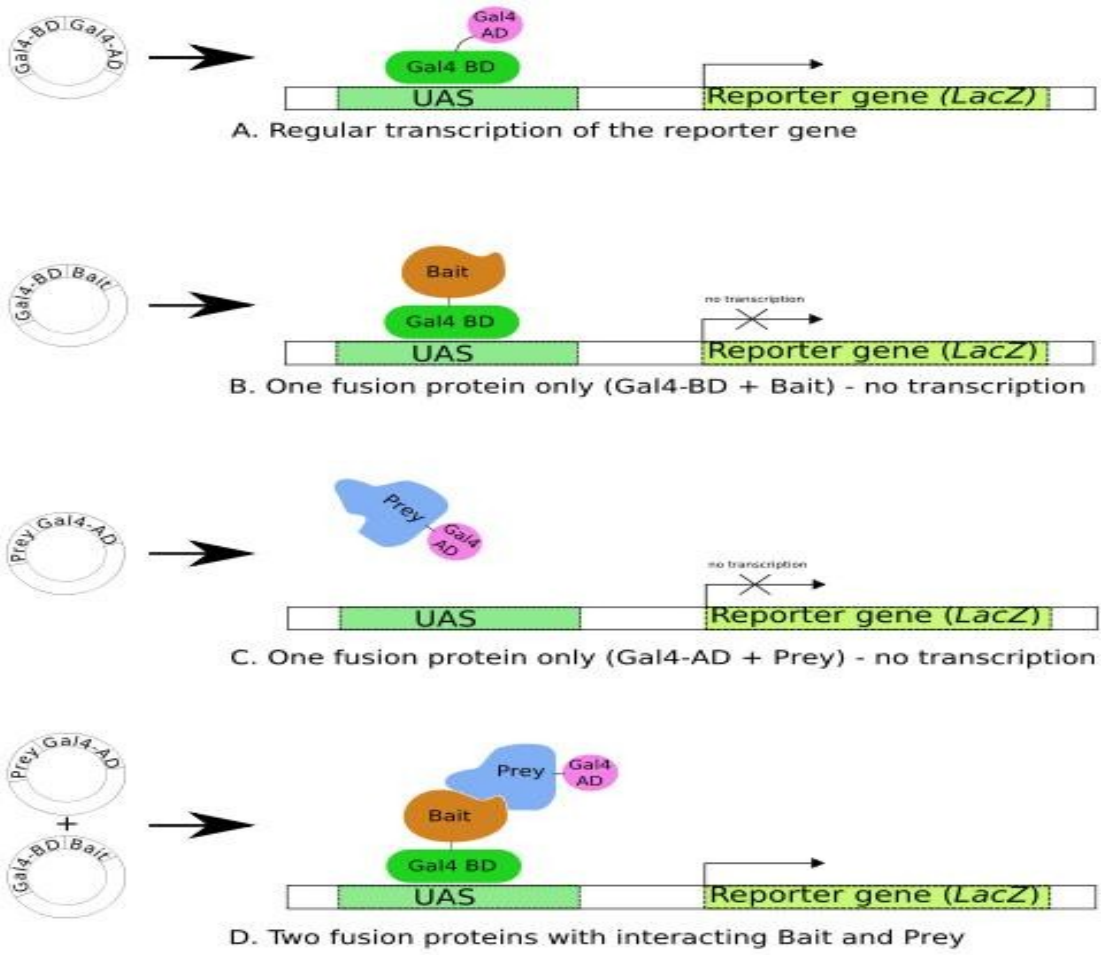
PCR the isolated DNA

Sequence the fragments



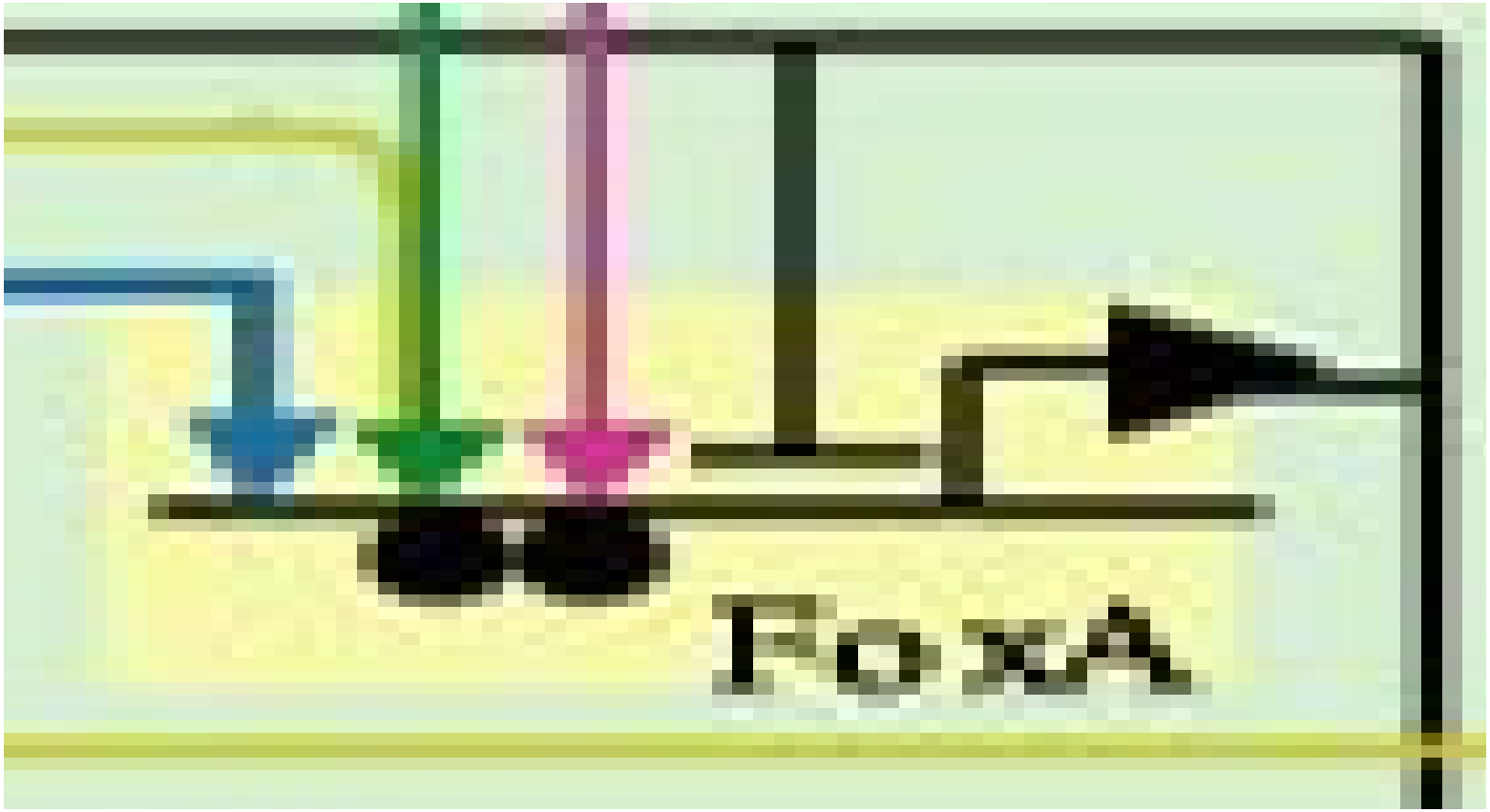
http://web.wi.mit.edu/young/regulator_network/

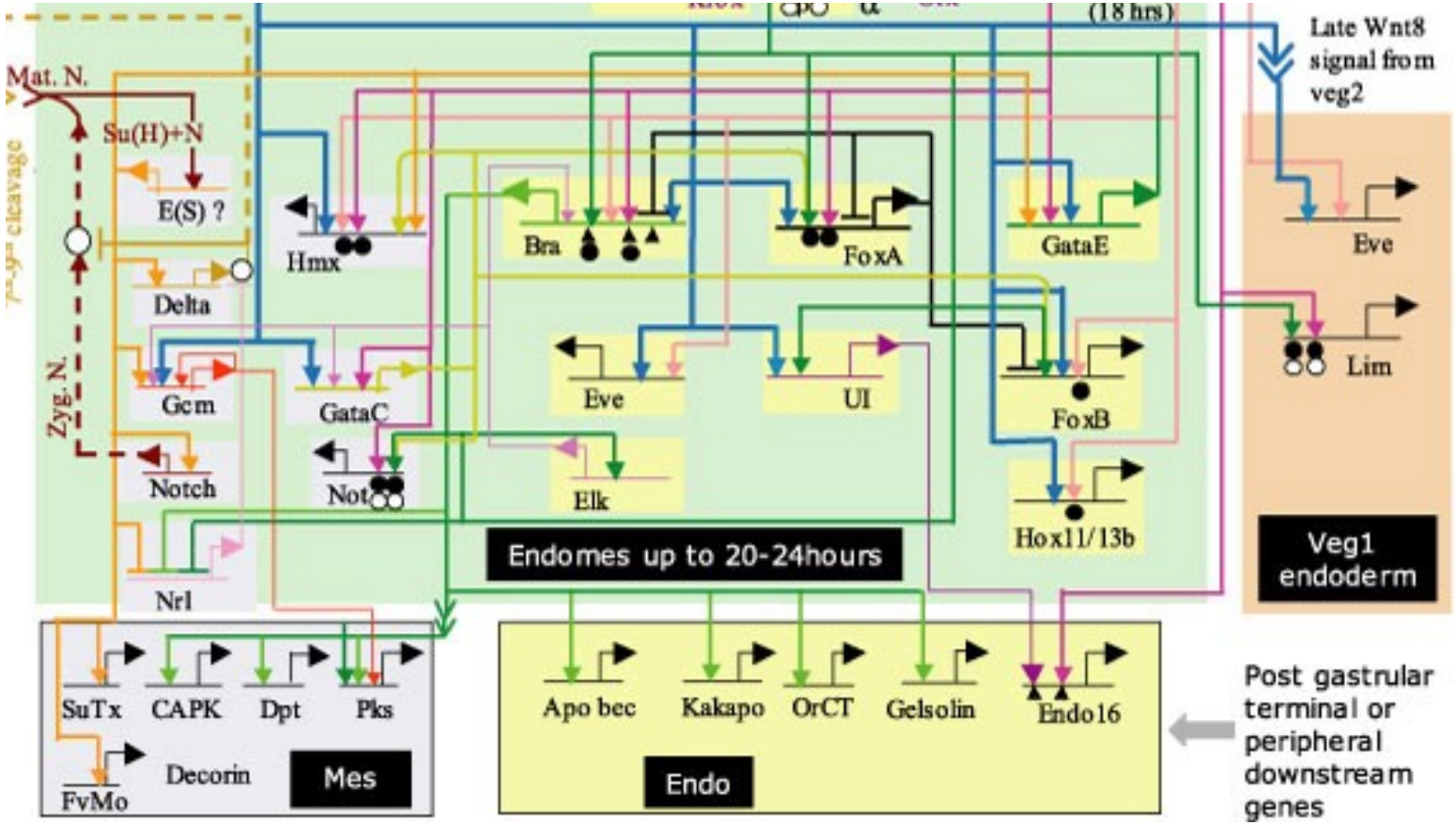
6. Two-Hybrid Screens

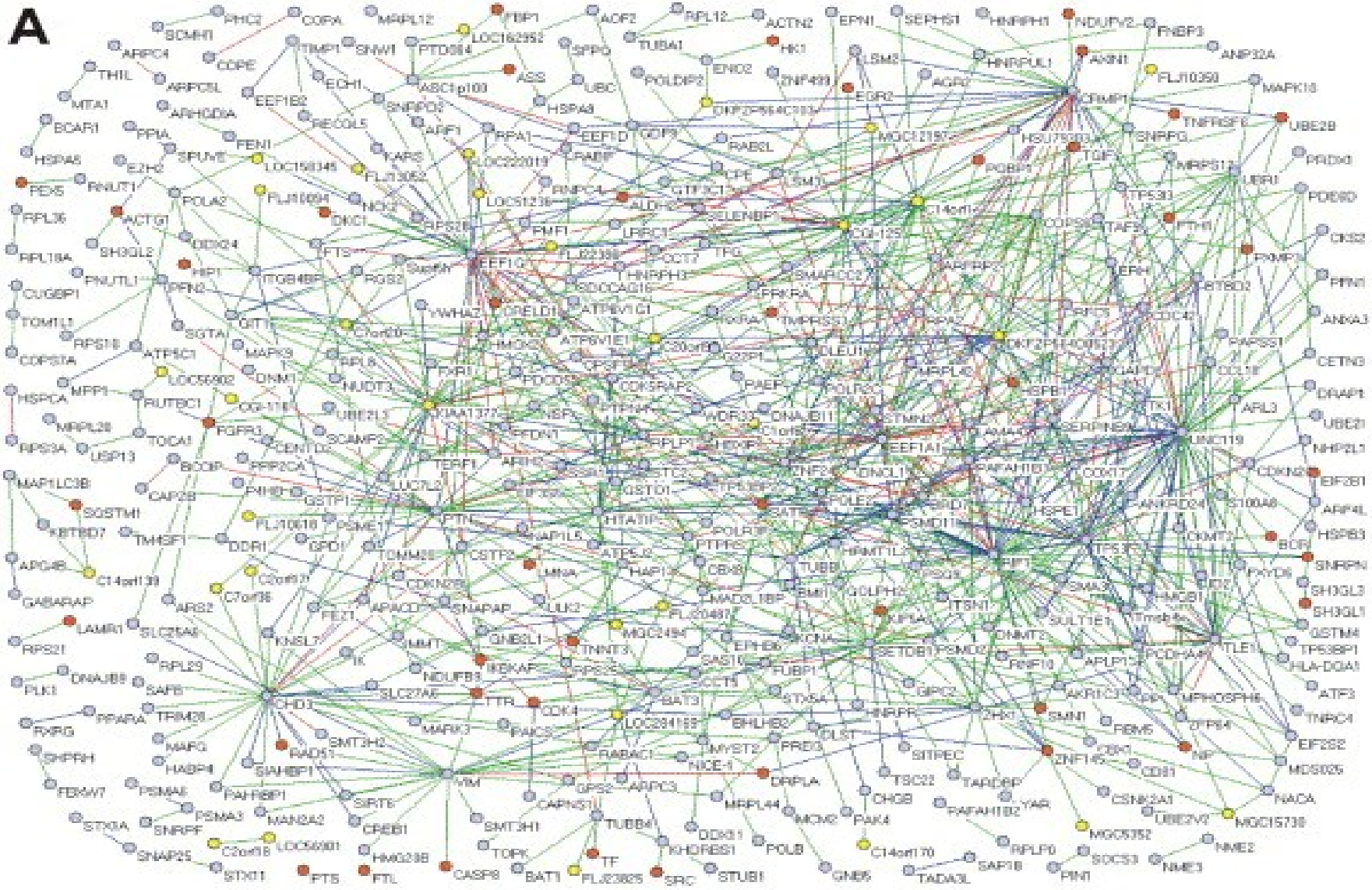


Used to elucidate Protein-Protein Interactions

Gene Regulation Simplified




(18 hrs)
Late Wnt8 signal from veg2
Mat. N.
Su(H)+N
E(S) ?
Delta
Zyg. N.
Gcm
Notch
Nrl
Hmx
GataC
Not
Bra
FoxA
GataE
Eve
Ui
Elk
FoxB
Hox11/13b
Eve
Lim
Veg1 endoderm
Endomes up to 20-24hours
SuTx
CAPK
Dpt
Pks
Decorin
FvMo
Mes
Apo bec
Kakapo
OrCT
Gelsolin
Endo16
Endo
Post gastrular terminal or peripheral downstream genes



Human Protein-Protein Interaction Network, via Y2H screens (Stelzl et al 2005)