Microarray Technology

(Thinking carefully about data)

Lecture 5

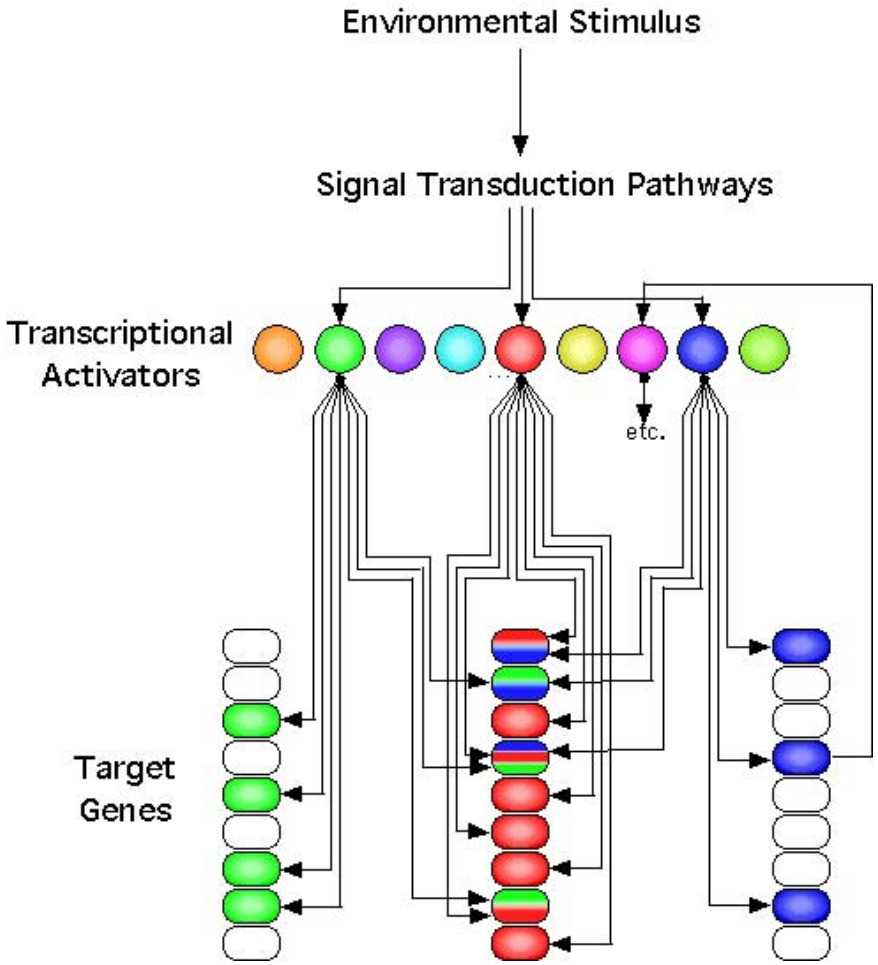
6.874J/7.90J/6.807

David Gifford

Microarrays can access in-vivo data derived from nucleic acid

- What kinds of information would we like to know?
- How do microarrays work?
- What kinds of errors are inherent in microarrays?
- How can we design measurement protocols to derive useful information using microarrays?
  - mRNA expression is just one example of what we can observe
  - Expression arrays are becoming less important over time

Transcriptional Regulatory Pathways



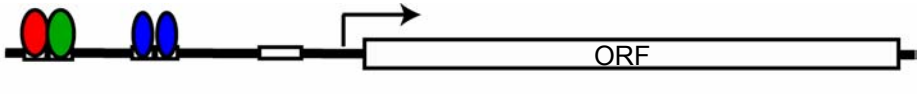
The regulatory pathways that control gene expression programs are uncharted

The mapping of transcriptional regulatory pathways will:

- reveal how cell state, differentiation and response to stimuli are controlled
- suggest new strategies to combat disease

Transcription of Protein-Coding Genes

DNA-binding regulators occupy promoter elements

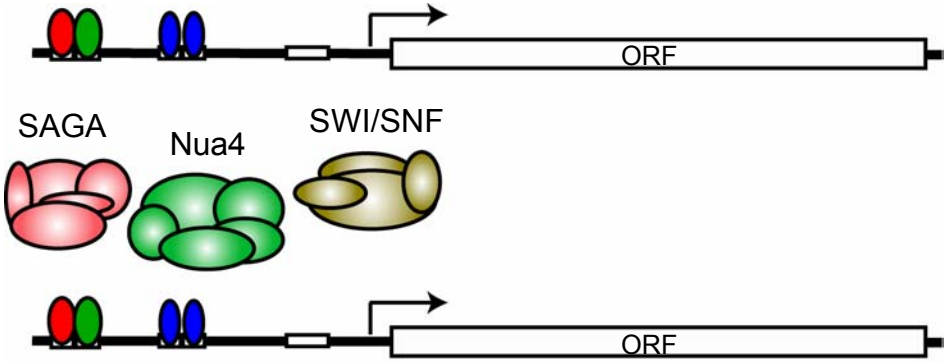


Jacob & Monod, 1961
Gilbert & Muller-Hill, 1967
Ptashne, 1967

Transcription of Protein-Coding Genes

DNA-binding regulators occupy promoter elements

Histone H3 and H4 acetyl transferases and SWI/SNF recruited to promoter



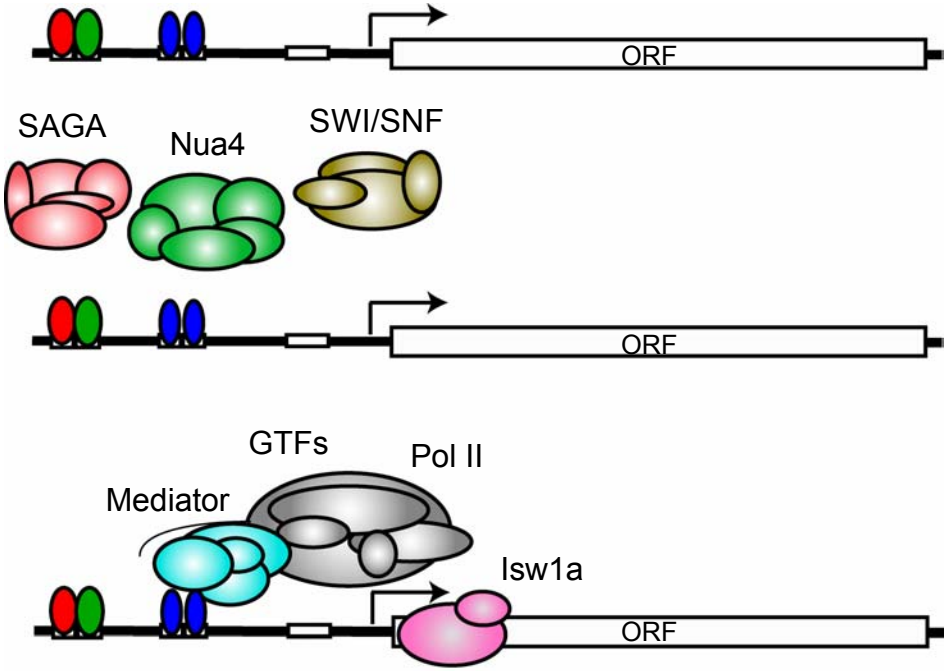
Peterson & Herskowitz, 1992
Brownell et al., 1996
Smith et al., 1998
Cosma et al., 1999
Reid et al., 2000
Bhanmik & Green, 2001
Larschan & Winston, 2001

Transcription of Protein-Coding Genes

DNA-binding regulators occupy promoter elements

Histone H3 and H4 acetyl transferases and SWI/SNF recruited to promoter

GTFs, Mediator, RNA polymerase II, and additional chromatin regulators recruited



Reinberg & Roeder, 1987
Buratowski et al. 1989
Flores et al. 1992
Kolenske & Young, 1994
Kim & Kornberg, 1994
Li et al., 1999
Orphanides & Reinberg, 2002
Pokholok et al., 2002
Morillon et al., 2003

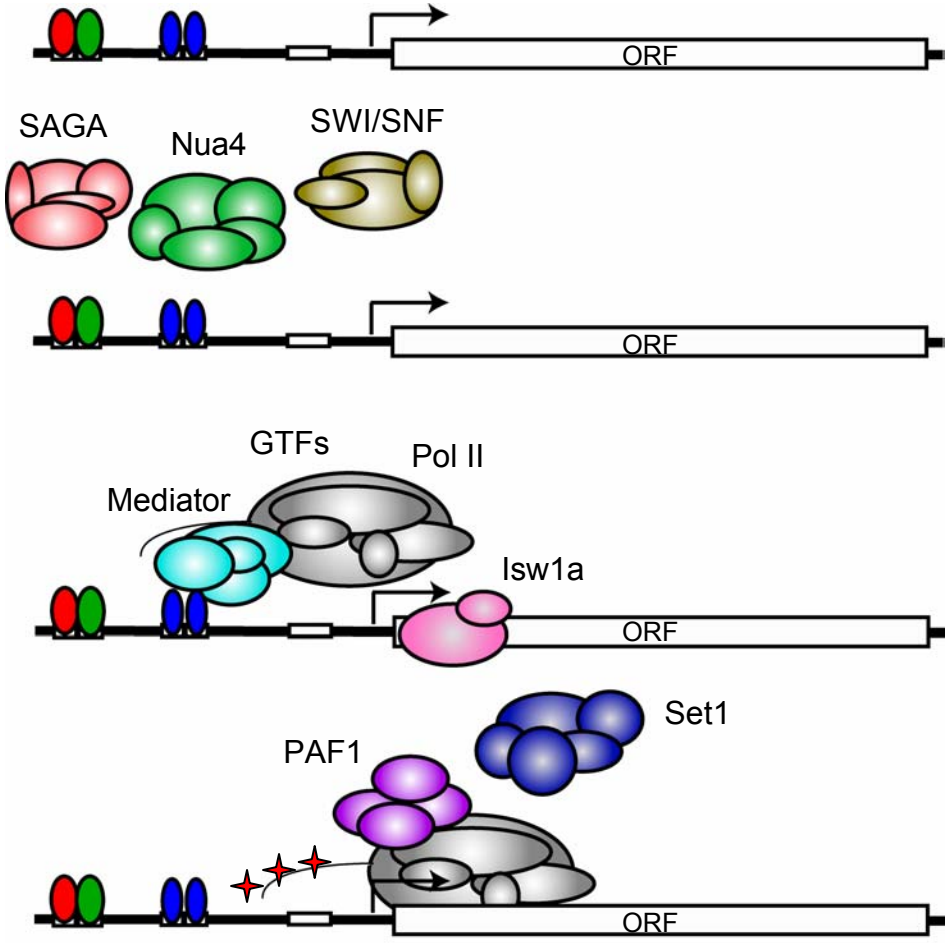
Transcription of Protein-Coding Genes

DNA-binding regulators occupy promoter elements

Histone H3 and H4 acetyl transferases and SWI/SNF recruited to promoter

GTFs, Mediator, RNA polymerase II, and additional chromatin regulators recruited

TFIIH phosphorylates RNA Pol II CTD on Ser5
Recruitment of Paf1 complex and
Set1 histone H3 K4 methyl transferase



Nislow et al., 1997
Briggs et al., 2001
Ng et al., 2003
Krogan et al., 2003

Transcription of Protein-Coding Genes

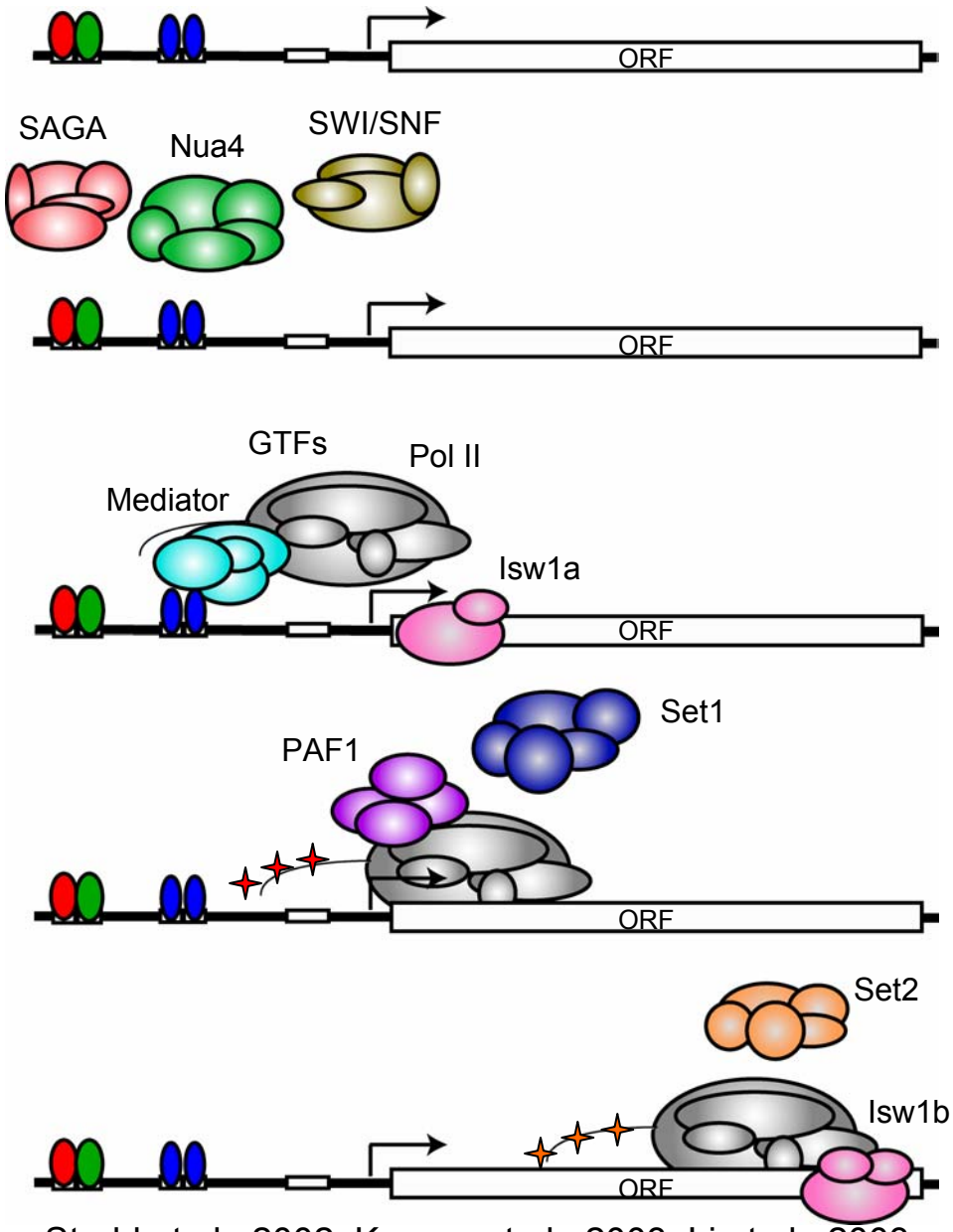
DNA-binding regulators occupy promoter elements

Histone H3 and H4 acetyl transferases and SWI/SNF recruited to promoter

GTFs, Mediator, RNA polymerase II, and additional chromatin regulators recruited

TFIIH phosphorylates RNA Pol II CTD on Ser5
Recruitment of Paf1 complex and
Set1 histone H3 K4 methyl transferase

CTK1 phosphorylates RNA Pol II CTD on Ser2
Recruitment of Set2 H3 K36 methyl transferase and Isw1b



Strahl et al., 2002; Krogan et al., 2003; Li et al., 2003

Microarrays can access in-vivo nucleic acid based information

- What kinds of information would we like to know?
- How do microarrays work?
- What kinds of errors are inherent in microarrays?
- How can we design measurement protocols to derive useful information using microarrays?
  - mRNA expression is just one example of what we can observe
  - Expression arrays are becoming less important over time

DNA Hybridization

Complimentary strands of nucleic acid will bind each other by base pairing

Diagram removed for copyright reasons.

DNA Microarray

By immobilizing different sequences in unique locations on a solid substrate we can measure relative levels of many target DNA or RNA molecules in a complex mixture in parallel

Diagram removed for copyright reasons.

DNA Microarray Content

Oligonucleotides

- Synthesized in situ or printed
- Short Probes (25-70 Bases)
- Multiple probes/gene
- Homemade or Commercial

cDNA

- Prefabricated and printed
- Long Probes (300-1000)
- One probe/gene
- Homemade or Commercial

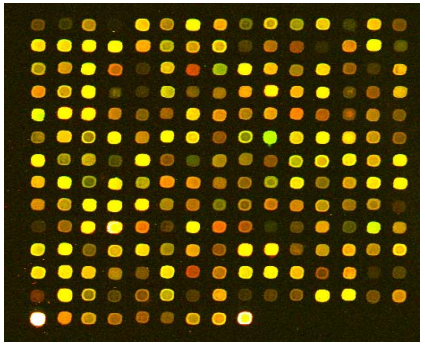
Synthesis of Labeled Target From mRNA

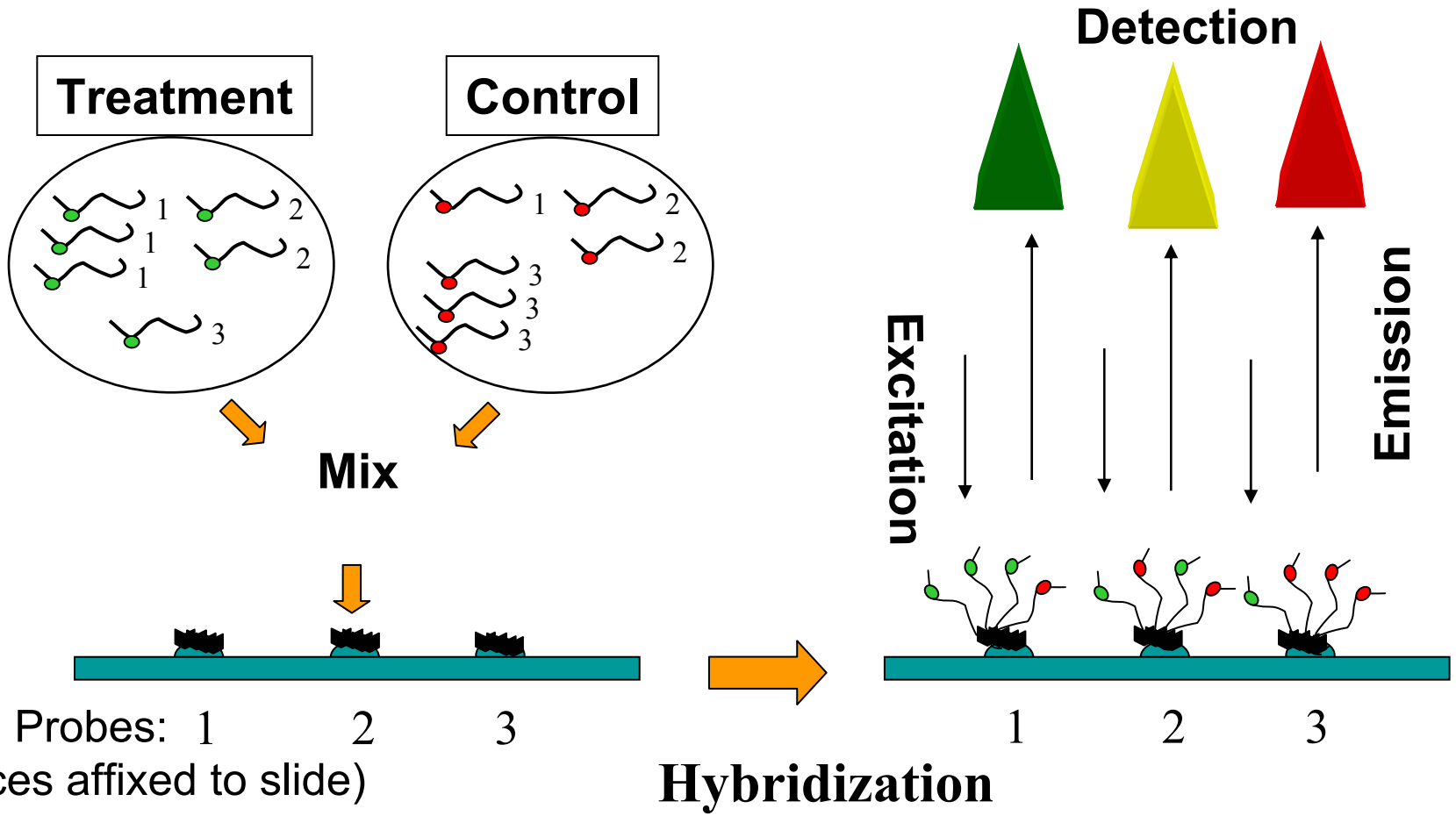
1) Reverse transcription can be used to make double-stranded DNA from RNA, and DNA molecule can be labeled

2) Double stranded DNA molecule can be transcribed by RNA polymerase to produce cRNA

Diagram removed for copyright reasons.

Measuring Relative Target Abundance





Several slides removed for copyright reasons.
Diagrams and descriptions of Affymetrix, Agilent, and MAGE-ML technologies.

Some sources of variation in microarray data

- Sample variation (not noise)
  - Population differences
- Sample preparation (noise)
  - Amplification based artifacts
  - Dye-substrate preferences (dye swap test)
  - Protocol specific issues
- Probe design (noise)
  - Cross-hybridization / non-uniqueness
  - Probe secondary structure
  - Tm issues
- Array based (noise - can be reduced by technical replicates)
  - Hybridization
  - Scanner noise

Is noise additive or multiplicative?

- Examples of additive noise
  - Scanner noise
- Examples of multiplicative noise
  - Probe affinity
  - Amplification non-linearity

$$o_i = n_{1i} d_i + n_{2i}$$

Microarrays can access in-vivo data derived from nucleic acid

- What kinds of information would we like to know?
- How do microarrays work?
- What kinds of errors are inherent in microarrays?
- How can we design measurement protocols to derive useful information using microarrays?
  - mRNA expression is just one example of what we can observe
  - Expression arrays are becoming less important over time

Microarray based measurement protocols

- Genotype
  - Single nucleotide polymorphisms (SNPs)
  - Comparative genomic hybridization (CGH)
- Expression (mRNA)
  - mRNA transcript levels
  - alternative splice variants
  - ncRNAs (miRNAs)
- Protein / RNA interactions
- Genome interaction
- ChIP – Chip (location analysis)
    - TF Binding
    - Chromatin structure
  - Dam Methylase assay

CGH and SNP analysis

- Represent probes that correspond to interesting genetic loci
- Fragment and amplify genome; apply to array
- CGH – probes measure relative copy number of loci of two genomes
  - Can detect annupolidy, certain genetic diseases
- SNP – probes measure presence or absence of SNP at a particular locus
  - Discovering SNPs and linking them to disease is key first step

Image removed for copyright reasons.
See Figure 1 in source cited below.

Pollack,J., Perou,C., Alizadeh,A., Eisen,M., Pergamenschikov,A., Williams,C., Jeffrey,S., Botstein,D. and Brown,P. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, **23**, 41–46.

SNP Array (Affymetrix)

The 40 25-mer probes per locus

= 5 groups of probes with SNP at different position

X 8 probes / position (both strands)

(1) a perfect match for one of the allelic states (allele A);

(2) a single base mismatch for allele A;

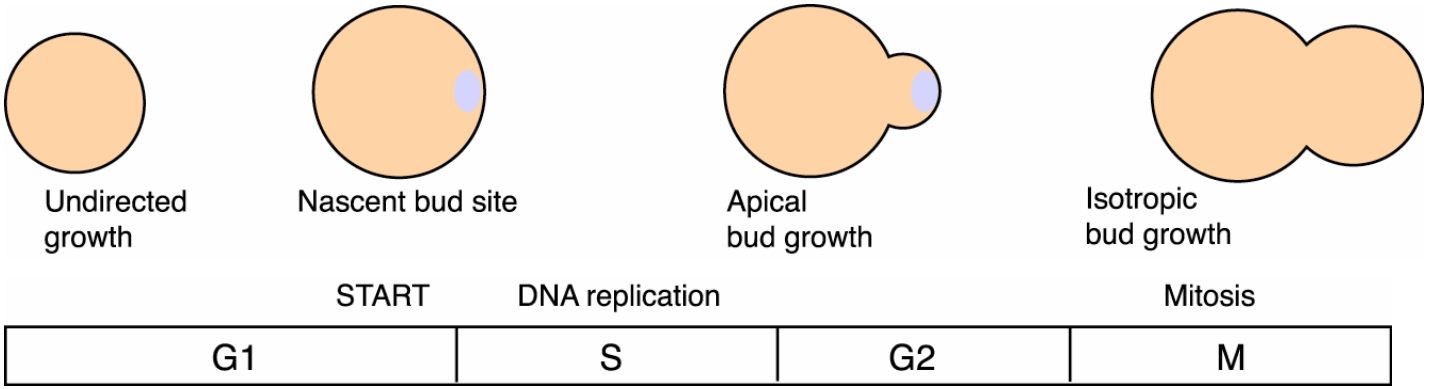
(3) a perfect match for the other allelic state (allele B);

(4) a single base mismatch for allele B

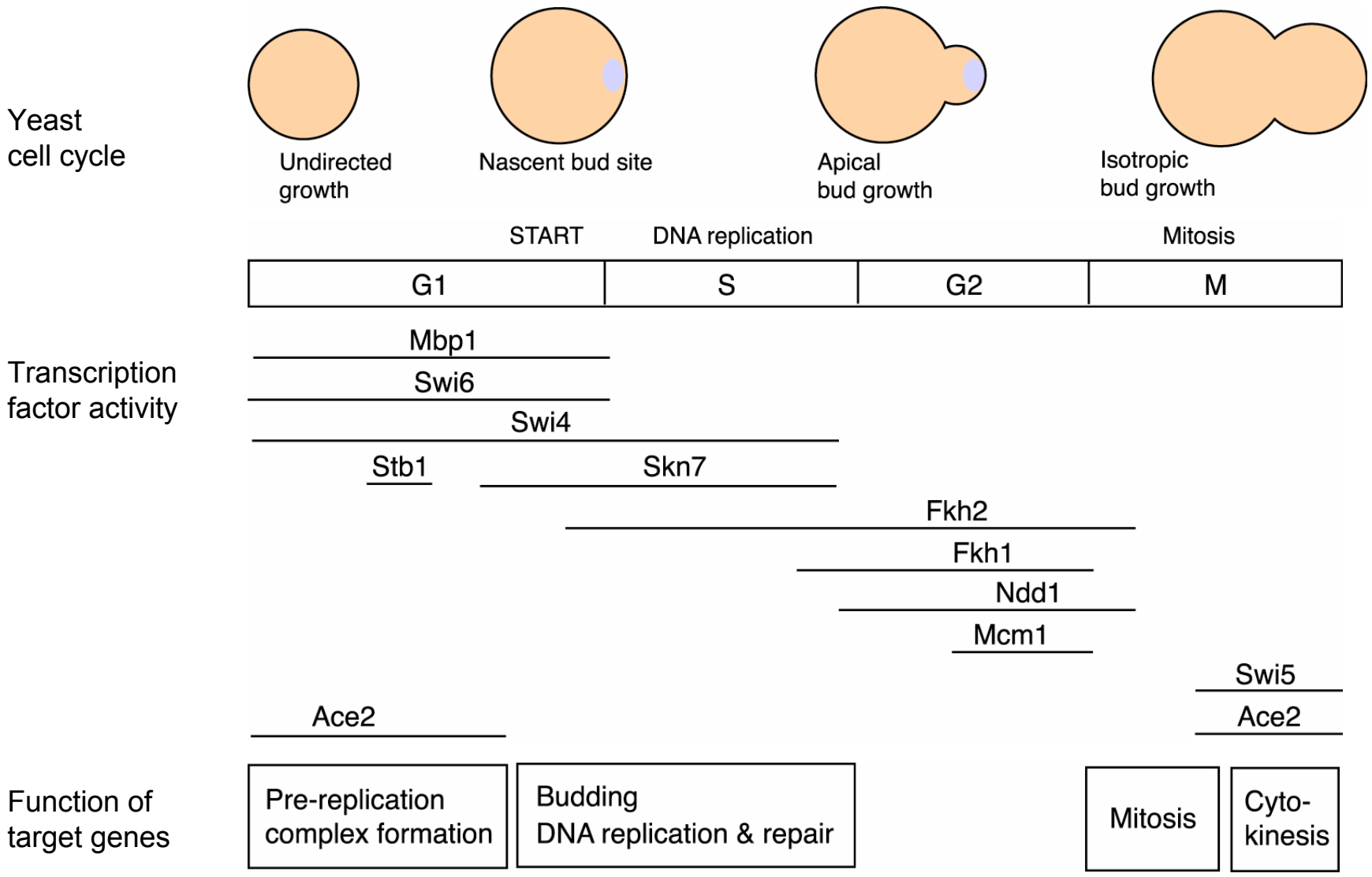
Expression

- Amplify RNA and label with Cy3 / Cy5
  - Can be used with small number of cells (10,000 or so)
  - Select control sample if two color array
- Hybridize to array

Transcriptional Regulation of Yeast Cell Cycle



Transcriptional Regulation of Yeast Cell Cycle



Master Regulators of Human Transcription

Misregulation results in developmental problems and/or adult disease

Brain and Spinal Cord
SOX1-18, OCT6, MeCP2
CBP, NGN, NEUROD
Cerebrum
Cerebellum
Ganglia & nerves

Circulatory System
Myocardin, GATA4, TBX5,
NKX2.5, MEF2, HAND
Heart
Vascular system

Digestive System
HNF1, HNF4, HNF6, CBP,
PGC1, FOXA, PDX1, GATA,
MAFA, NKX2.2
Esophagus
Stomach
Intestines
Liver
Pancreas

Urinary System
HNF1B, HNF4, CDX, FTF
C/EBP, FOXA, GATA
Kidney
Urinary tract

Respiratory System
HNF-3, NKX2.1 and GATA6
Airways
Lungs

Reproductive Organs
ESR1, SERM, C/EBPβ
Ovary
Uterus
Breast
Testis

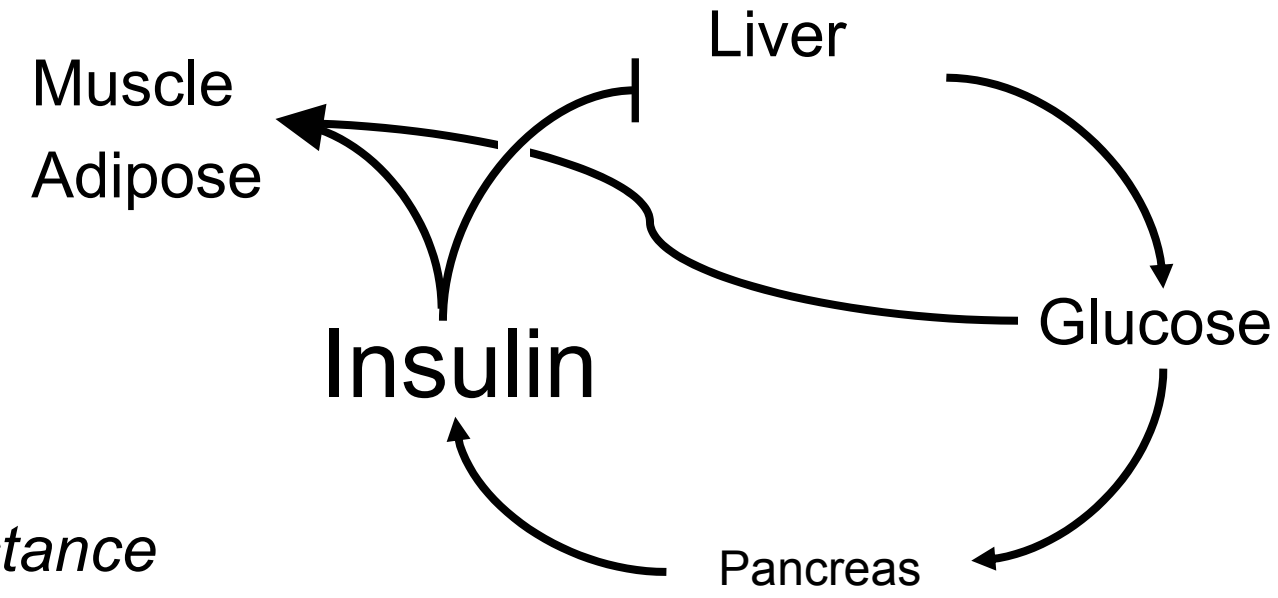
Skeletal and Muscular
MYOD, MEF2, MRF4, MYF5
Bone
Muscle
Cartilage

Hematopoietic System
TAL1, LMO1, LMO2, E2A,
XBP1, AFT6, PAX5, BCL6
Bone marrow
Blood
Embryonic Liver

Immune System
AML1, MLL1, HP1, HOXA7,
HOXA9, HOXC8, C/EBPA,
NFkB family
Thymus
Spleen
Lymph nodes

Sensory Organs
SOX1-18, OCT6, PAX3,
PAX6, NGN, SKIN1
Eye
Ear
Olfactory
Skin
Tongue

Glucose Homeostasis and Type 2 Diabetes



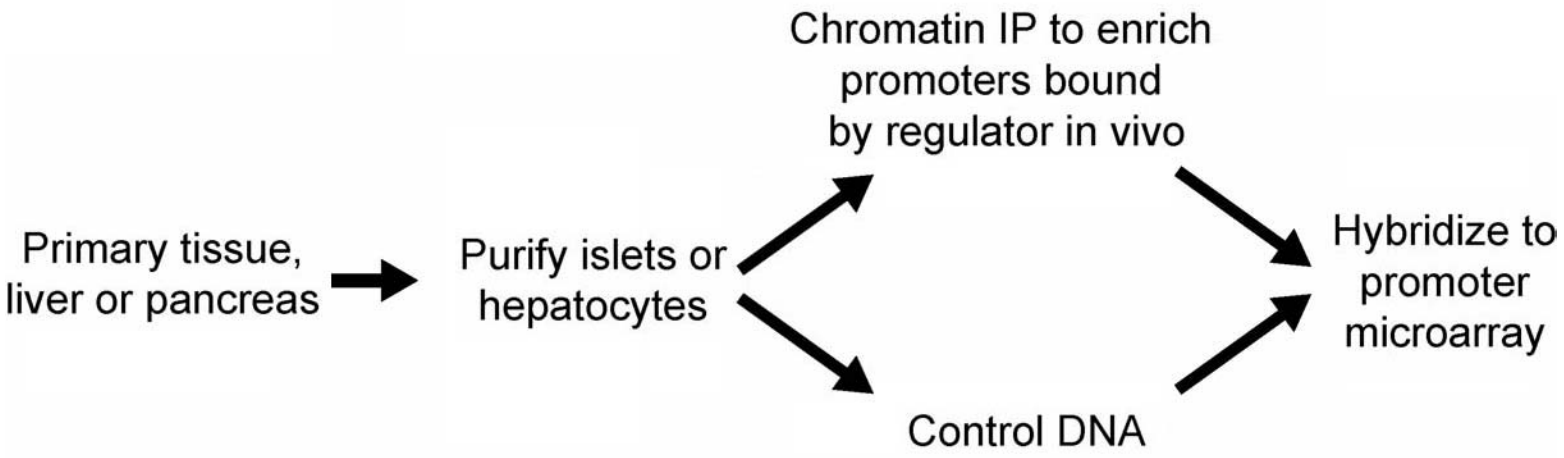
- *Insulin Resistance*
- *Mitochondrial inefficiency*
- *Impaired β cell insulin secretion*

MODY Diabetes

MODY type	Causative Gene	Protein Class	% Cases
MODY 1	***HNF-4α***	Orphan nuclear receptor protein	1
MODY 2	Glucokinase	Key enzyme in glucose sensing	20
MODY 3	***HNF-1α***	POU-homeodomain protein	60
MODY 4	***IPF1/PDX1***	Homeodomain protein	1
MODY 5	***HNF-1β***	POU-homeodomain protein	1
MODY 6	***NeuroD1***	Basic helix-loop-helix protein	1

Fajans et al. *N Engl J Med* (2001) 345 : 971

Genome-Scale Location Analysis in Human Tissues

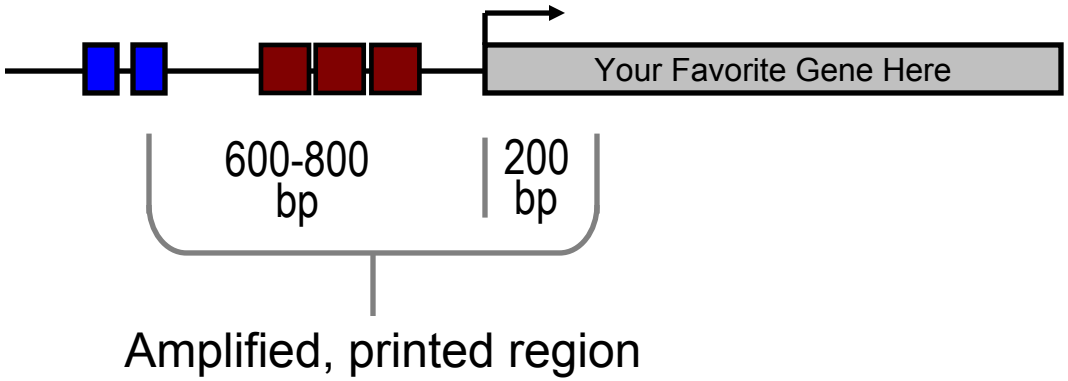


Hu13K Promoter Chip

Image removed for copyright reasons.

13,000 proximal promoter microarray

Targets 600-800 bp upstream and 200 bp downstream of mRNA start site (NCBI).



Hu13K Promoter Chip

1. 80% of binding sequences for known factors are promoter proximal.

Diagram and graph removed for copyright reasons.

2. Long range interactions between distal- and proximal-binding transcription factors can be captured by formaldehyde crosslinking.

   - Carter et al. Nature Genetics 32:623 (2002)
   - Tolhuis et al. Molecular Cell 10:1453 (2002)

Selected HNF1α Targets in Human Hepatocytes

Apolipoproteins

APOA2	Apolipoprotein A-II
APOC3	Apolipoprotein C-III
COPB2	Coatomer protein complex, subunit beta 2

Complement Proteins

C1S	Complement component 1, s subcomponent
C4BPA	Complement component 4 binding protein,a
F10	Coagulation factor X
DAF	Complement decay accelerating factor

Hormones, Cytokines, Growth Factors

ALB	Albumin
IGFBP1	Insulin-like growth factor binding protein 1
MST1	Macrophage stimulating 1 (hepatocyte growth factor-like)
SCYA16	Small cytokine subfamily A, member 16 (Monotactin-1)

Metabolism

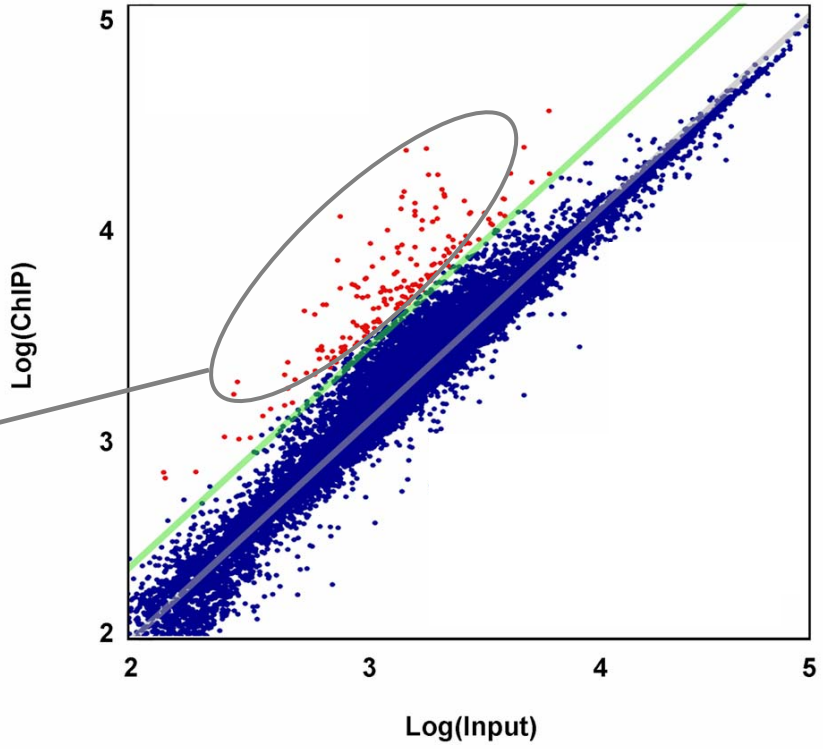
PCK1	Phosphoenolpyruvate carboxykinase 1 (soluble)
G6PT1	Glucose-6-phosphatase, transport protein 1
ADH1A	Alcohol dehydrogenase 1A, alpha polypeptide
XDH	Xanthene dehydrogenase

Receptors, Surface Transmembrane

RBP5	Retinol binding protein 5, cellular
GJB1	Gap junction protein, beta 1 (connexin 32)
RARB	Retinoic acid receptor, beta
SSTR1	Somatostatin receptor 1

Transcription Factors

HNF4A7	Hnf4alpha
ELF3	E74-like factor 3 (ets domain tf, epithelial-specific)
NR0B2	Small heterodimeric protein (Nr0B2)
NR5A2	Alpha-fetoprotein transcription factor



222 Targets