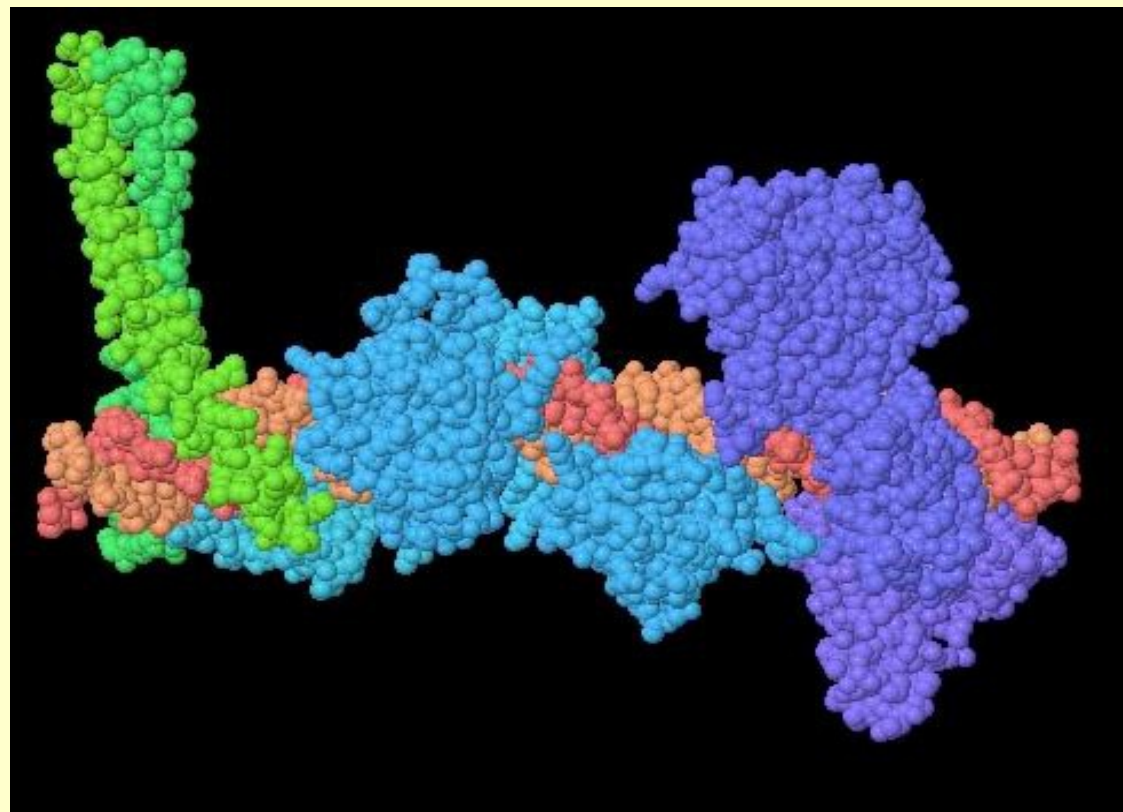


Genomics, Bioinformatics & Medicine

<http://biochem118.stanford.edu/>

Gene Expression and Cancer

<http://biochem118.stanford.edu/14%20Gene%20Expression.html>



Doug Brutlag

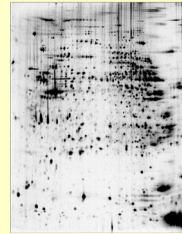
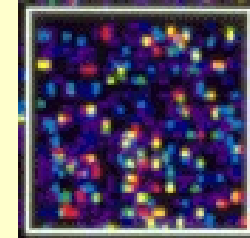
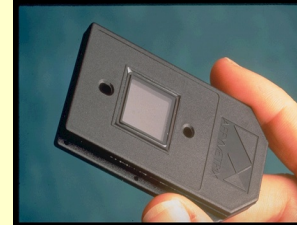
Professor Emeritus of Biochemistry & Medicine
Stanford University School of Medicine

© Doug Brutlag 2015

Leveraging Genomic Information

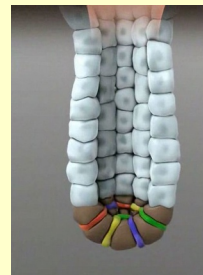
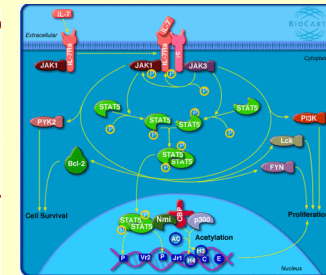
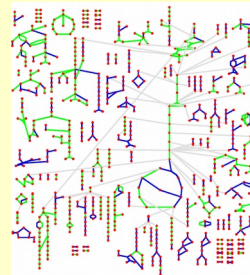
Novel Diagnostics

DNA Microchips & Microarrays
Gene Expression - RNA
Proteomics - Protein



Understanding

Metabolism
Cell Signaling
Differentiation
Disease



OMIM
Online Mendelian Inheritance in Man

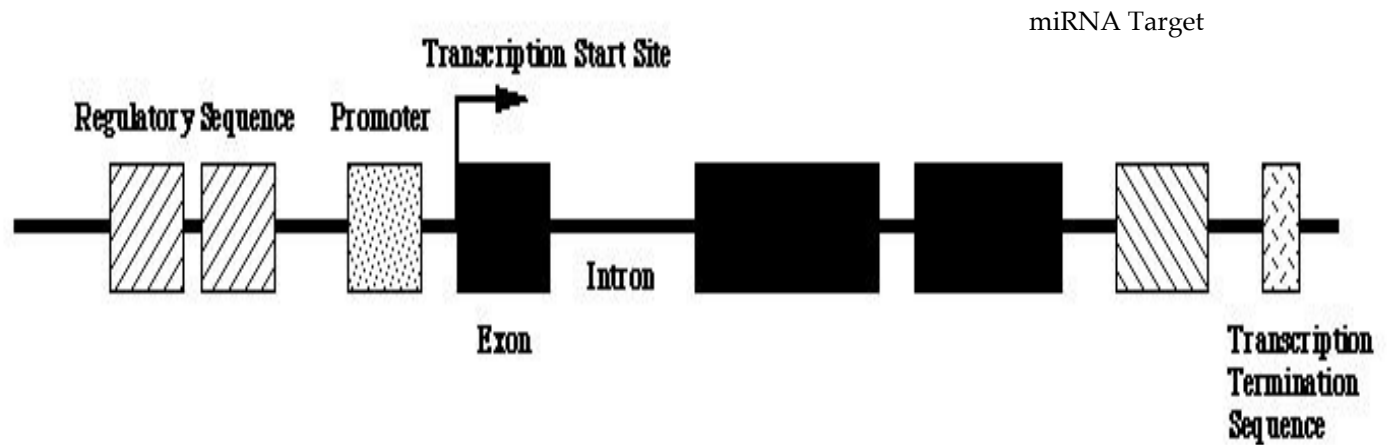


Johns
Hopkins
University

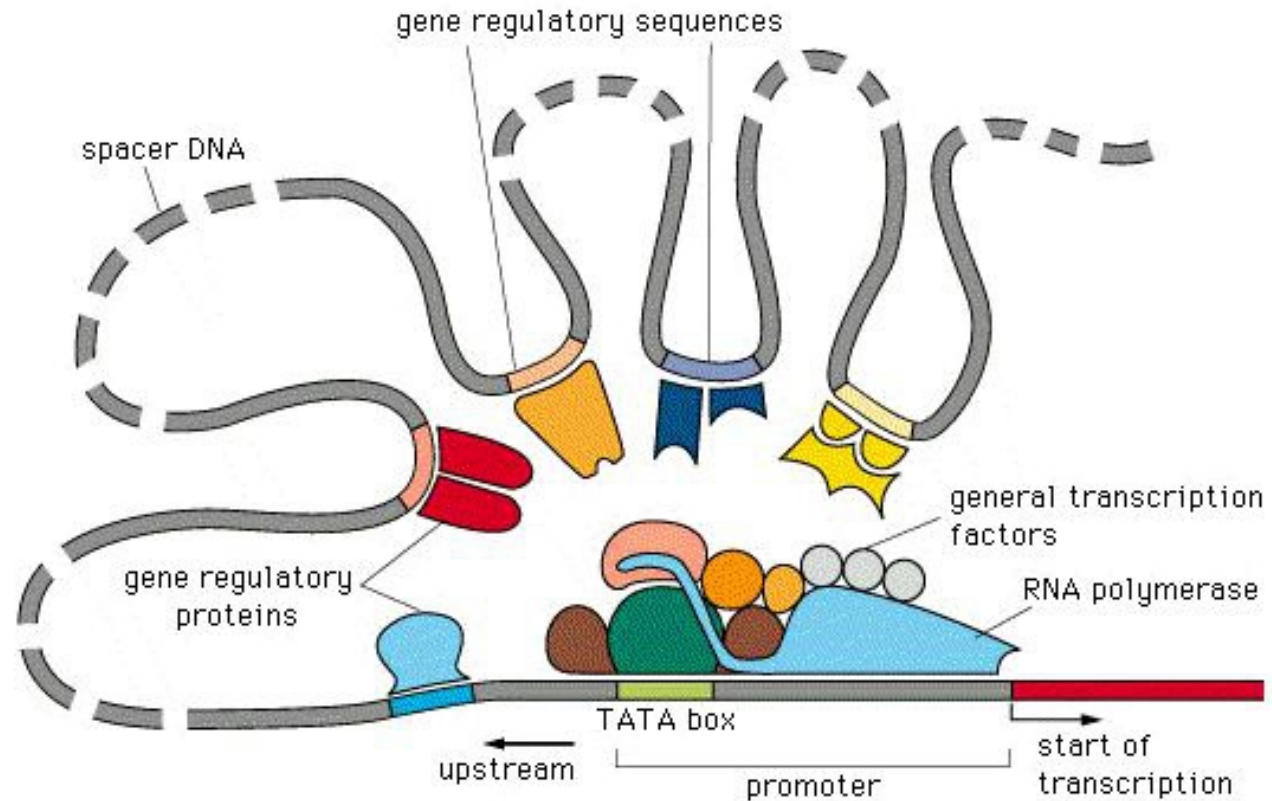
Gene Regulatory Mechanisms

- Transcriptional Mechanisms
 - Type of promoters & RNA polymerase
 - Control of Transcription
 - Transcription Factors and transcription factor binding sites
- RNA processing
 - Capping
 - Splicing and Alternative Splicing
 - Poly-Adenylation
 - RNA export to cytoplasm
 - RNA degradation rates
 - Micro RNAs (miRNAs) inhibit translation and degrade mRNA
 - Silencer RNAs (siRNAs or RNAi) degrading mRNA
- Epigenetic Mechanisms
 - DNA methylation
 - Histone modifications
 - Acetylation
 - Methylation
 - Phosphorylation, etc.
 - Chromatin remodeling

Eukaryotic Gene Structure



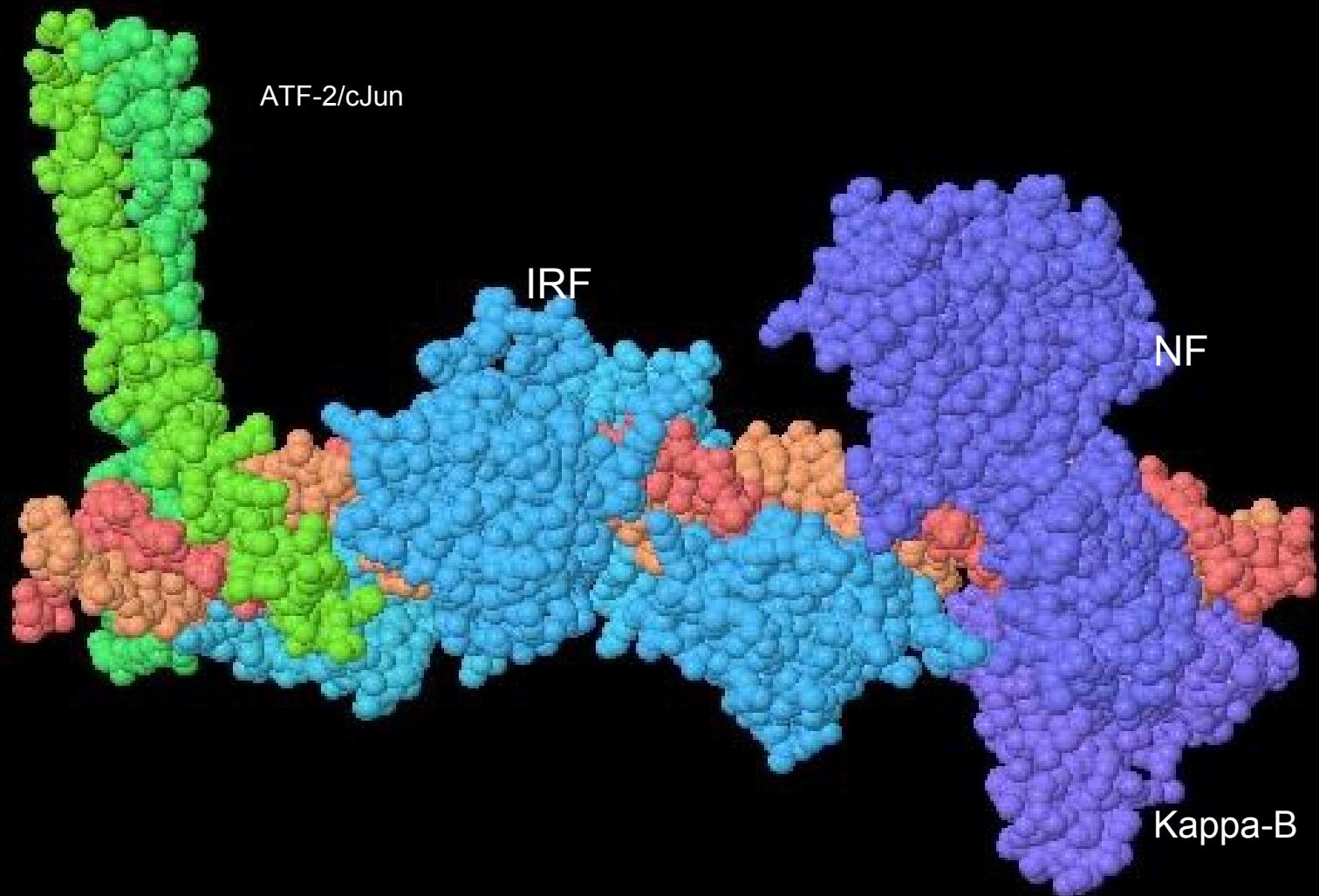
Multiple Enhancer Sequences



©1996 GARLAND PUBLISHING

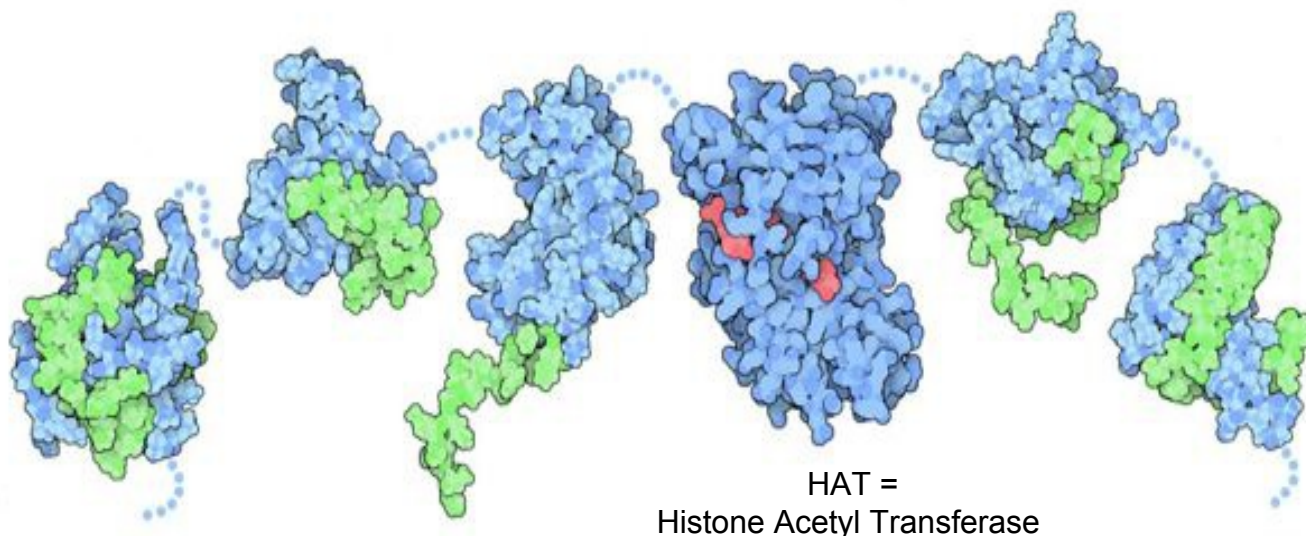
Enhanceosome: Exploring the Structure

http://www.rcsb.org/pdb/education_discussion/molecule_of_the_month/download/Enhanceosome.pdf



Enhanceosome: Integrating the Signal

http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb122_2.html



HAT =
Histone Acetyl Transferase

Integrating the Signal

Once the transcription factors bind to the different sites in the enhancer DNA sequence, the signal must somehow be sensed and used to activate the gene. In many cases, this is performed using CREB-binding protein or the similar protein p300. This protein is composed of many connected domains, (PDB entries **1l8c**, **1kdx**, **1jsp**, **3biy**, **2ka6** and **1kbh**), which bind to different proteins in the assembled enhanceosome. Then, a large domain in the center acts as a histone acetyltransferase, modifying histones in nucleosomes and causing them to disassemble and reveal the gene. In the interferon- β gene, a nucleosome normally covers the start site of transcription, blocking transcription. Assembly of the enhanceosome leads to removal of this nucleosome, allowing the gene to be expressed.

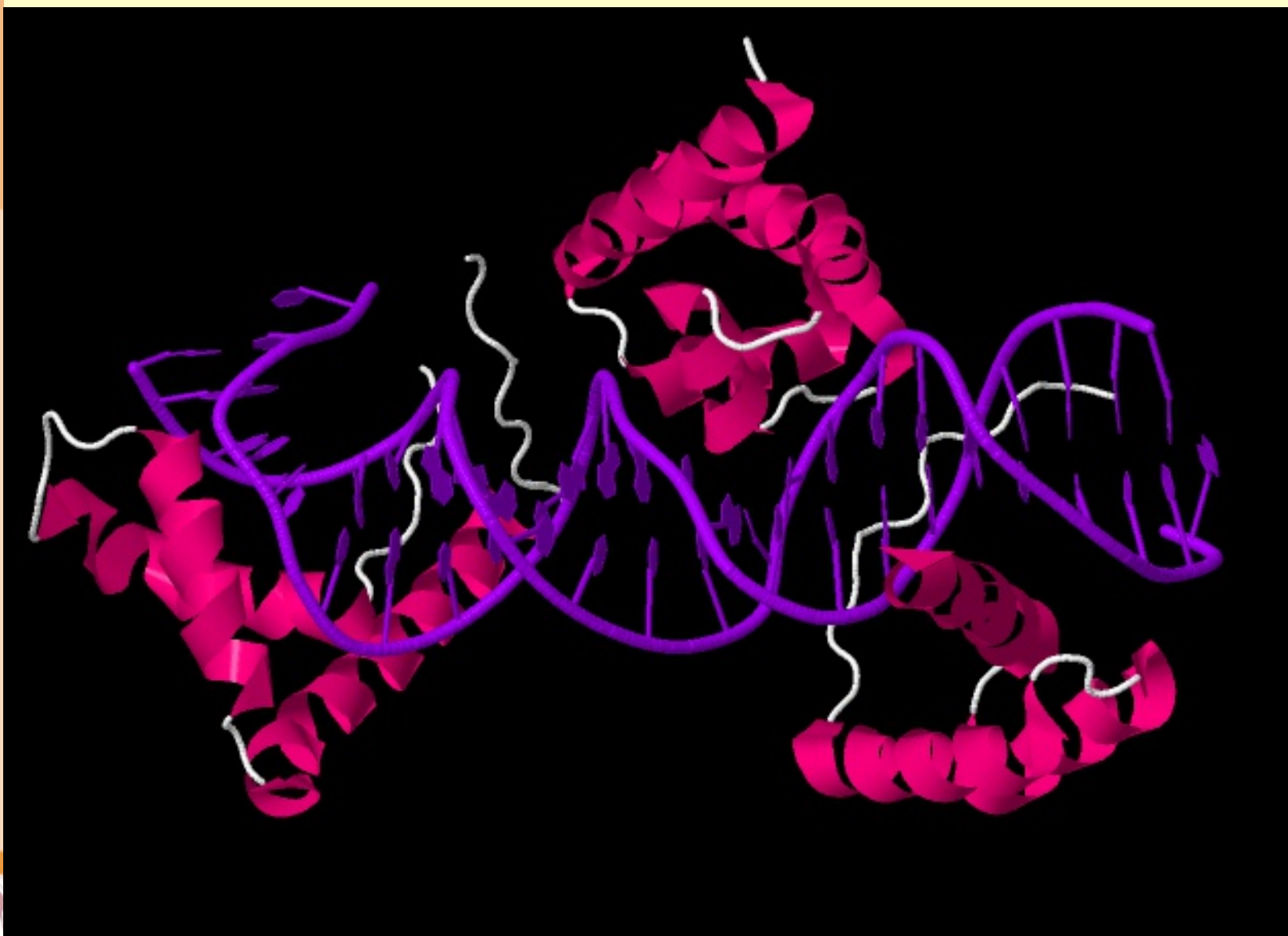
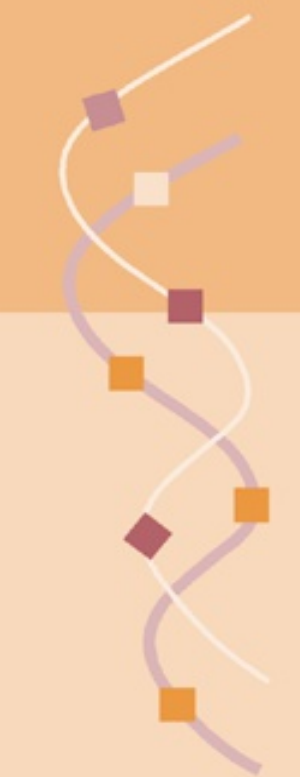
Previous: [Enhanceosome](#)

Home: [Enhanceosome](#)

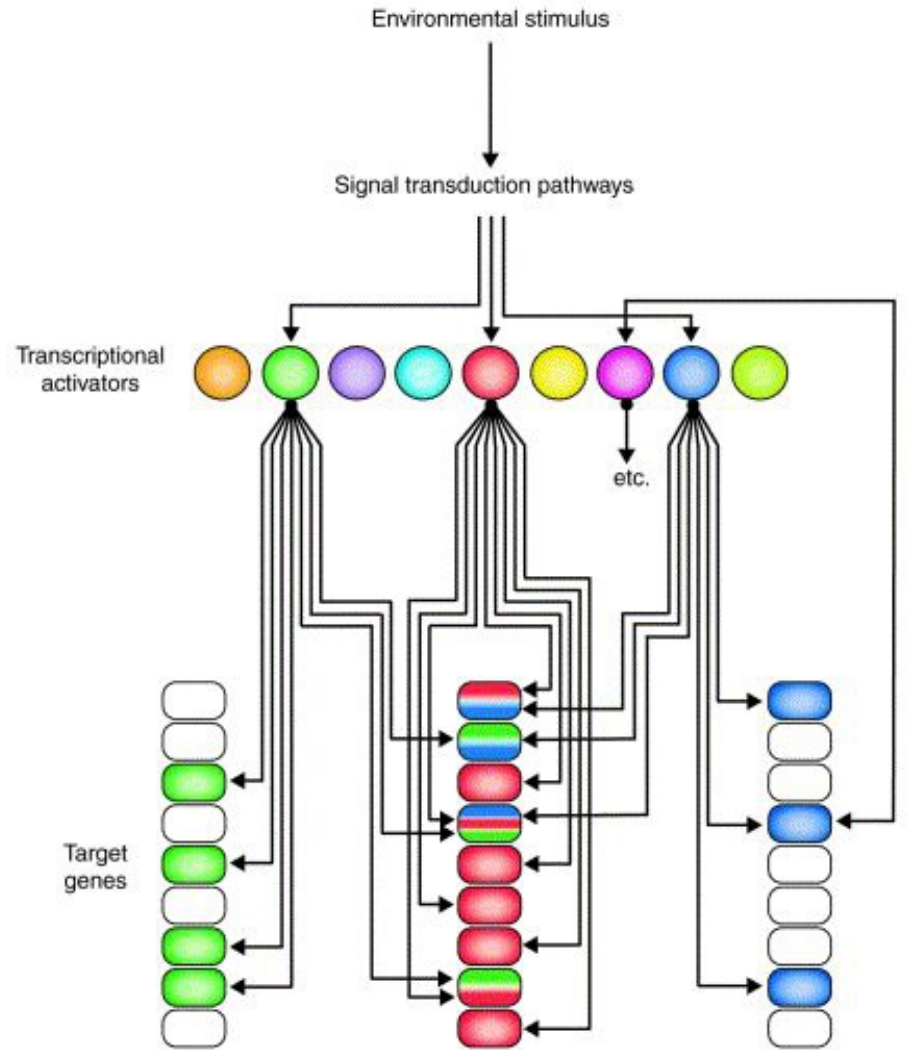
Next: [Exploring the Structure](#)

Oct3 / 4-Sox2 bound to FGF-4 Enhancer

<http://www.pdb.org/pdb/explore/explore.do?structureId=1GT0>

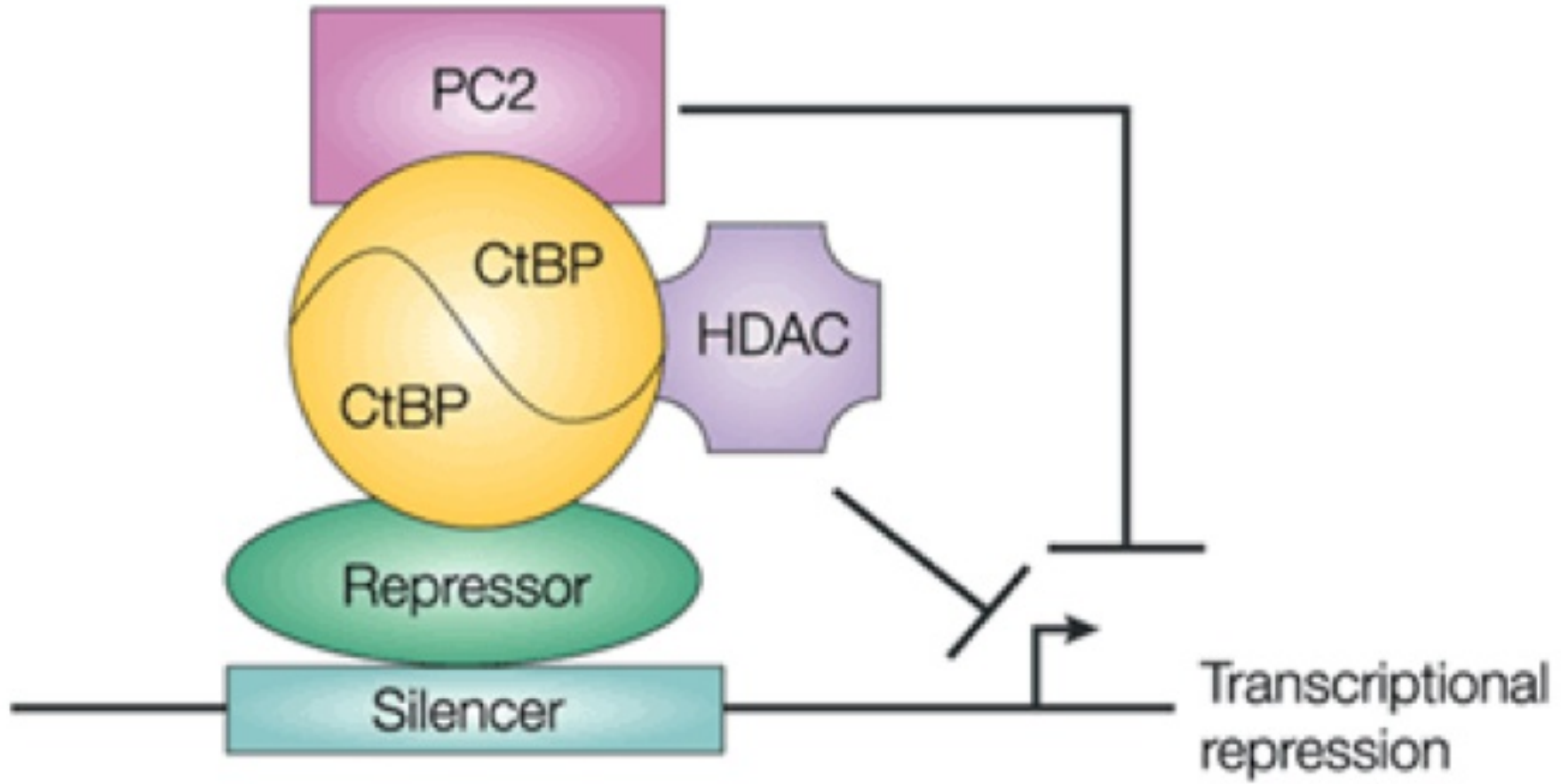


Gene Expression Regulatory Network



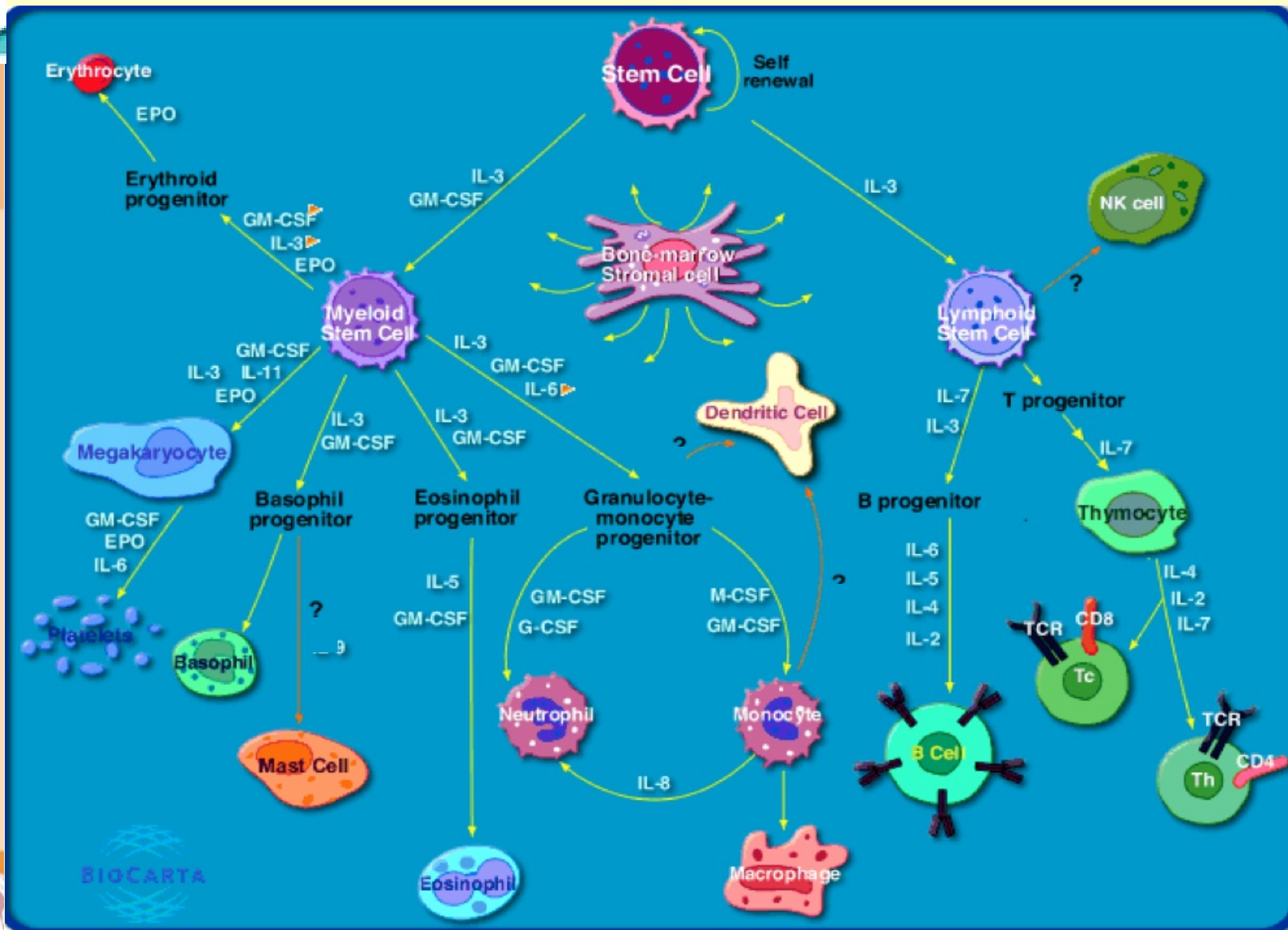
Current Opinion in Genetics & Development

Repressor Bound to Silencer Sites



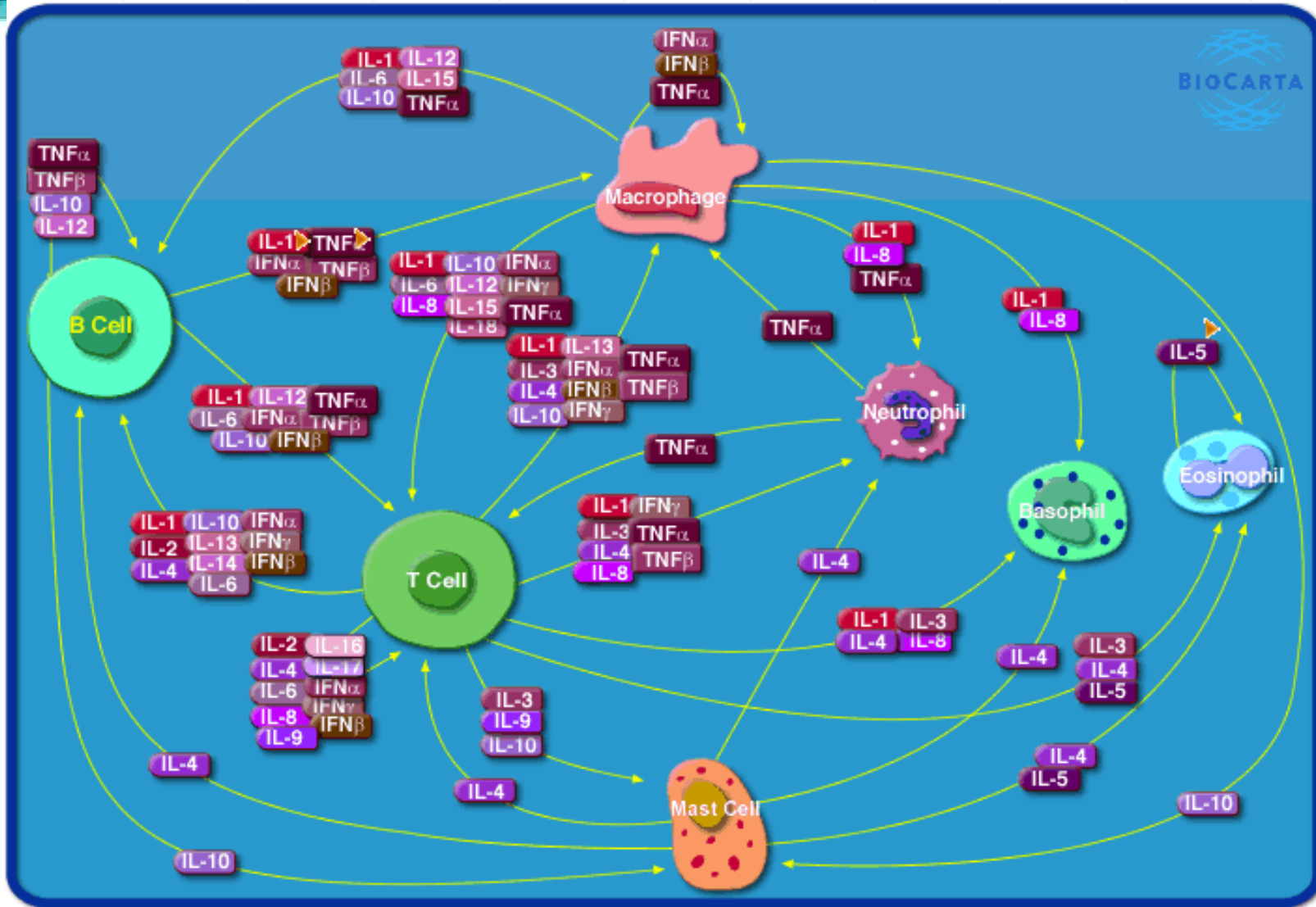
Hematopoiesis

http://cgap.nci.nih.gov/Pathways/BioCarta/h_stemPathway



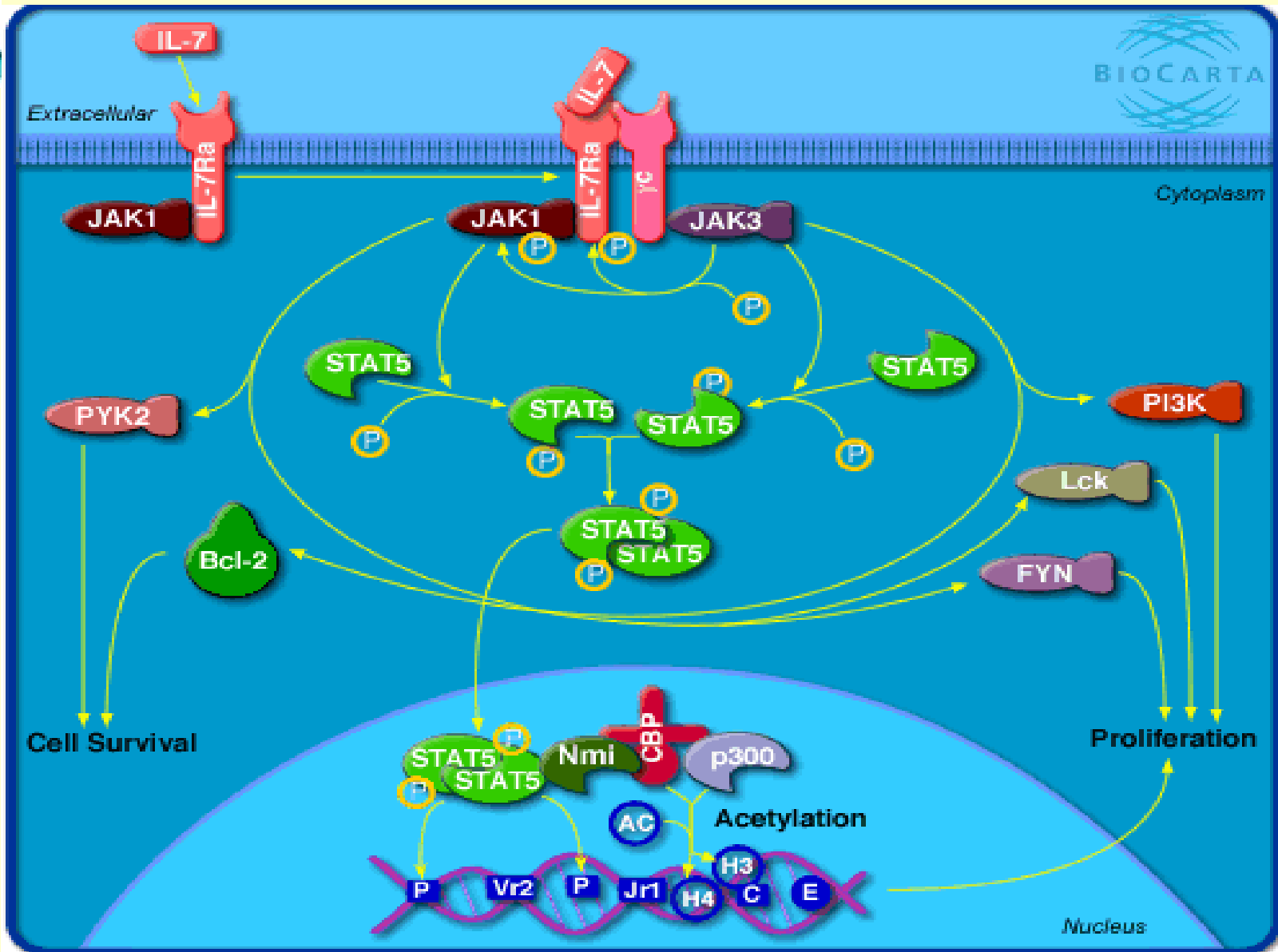
Cytokine Network

http://cgap.nci.nih.gov/Pathways/BioCarta/h_cytokinePathway



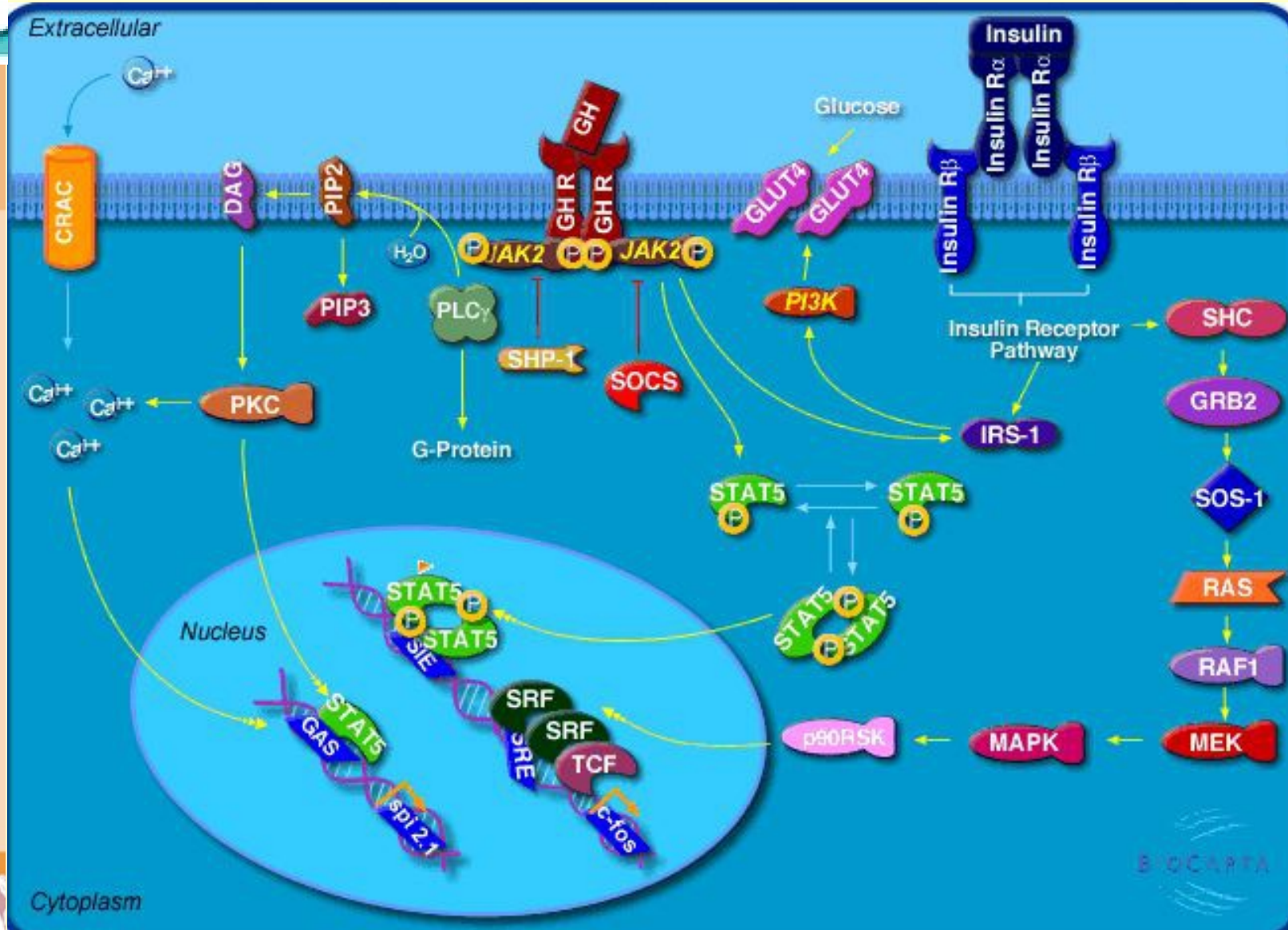
IL7 Regulatory Pathway

http://cgap.nci.nih.gov/Pathways/BioCarta/h_il7Pathway



Growth Hormone Receptor Pathway

http://cgap.nci.nih.gov/Pathways/BioCarta/h_ghPathway

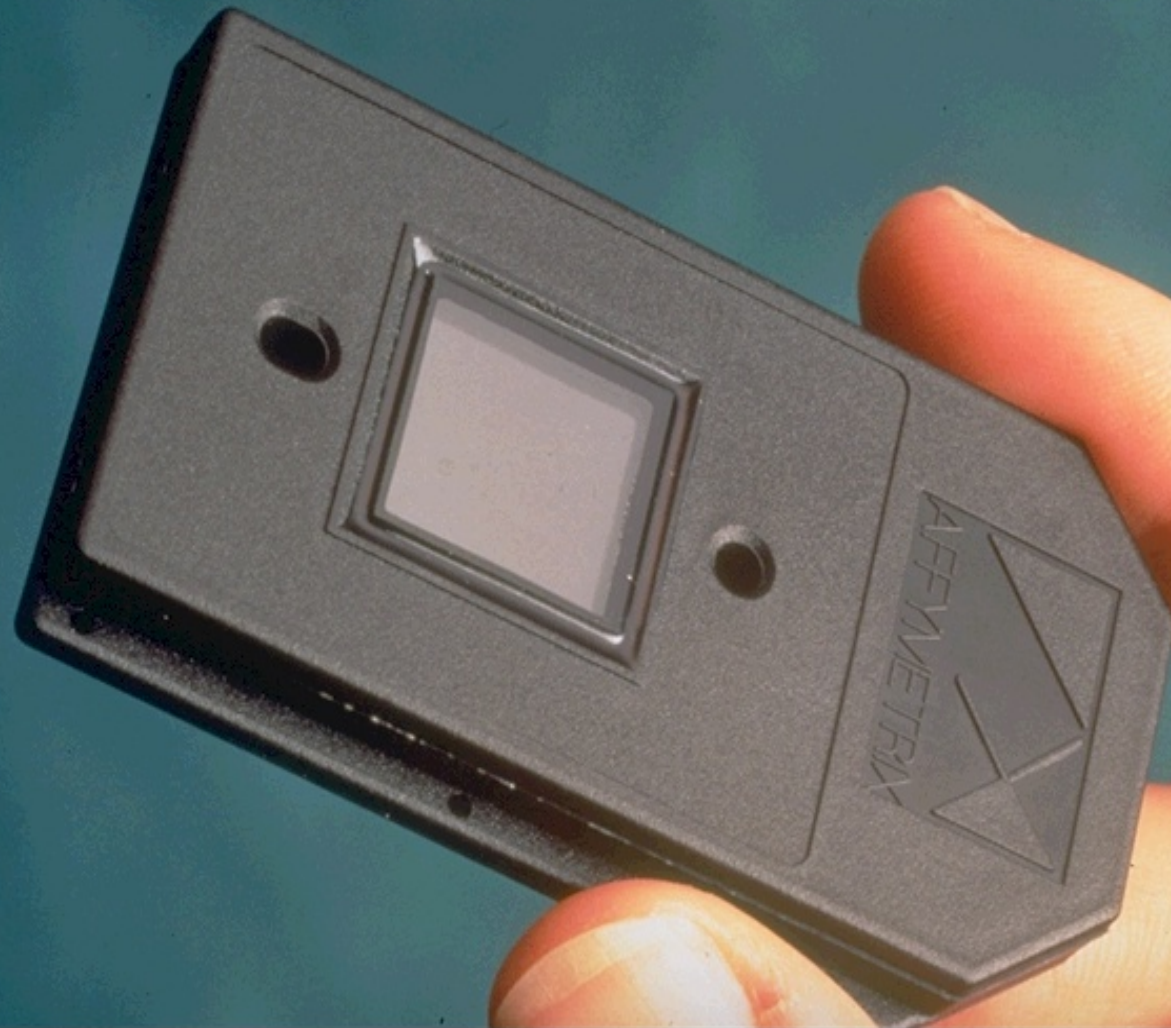


DNA Microarrays & DNA Chips Accelerate Gene Expression Analysis

- Parallel Analyses
 - Analyze entire genomes instead of single genes
 - Analyze expression of entire genome
 - Analyze genetic polymorphisms (SNPs)
- Miniaturization
- Automation

Diagnosis Using DNA Arrays

<http://www.affymetrix.com/>



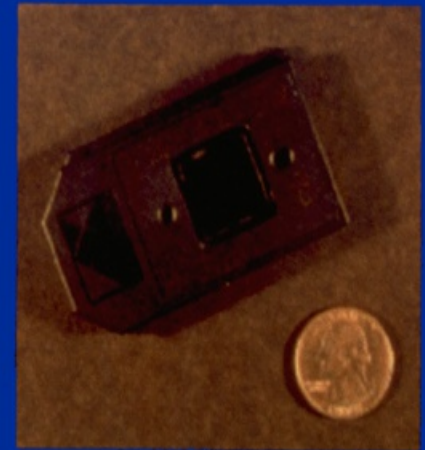
DNA Micro Arrays

- High-density grid of DNA sequences.



- Any collection of 25mers (1,200,000) can be synthesized in 100 steps

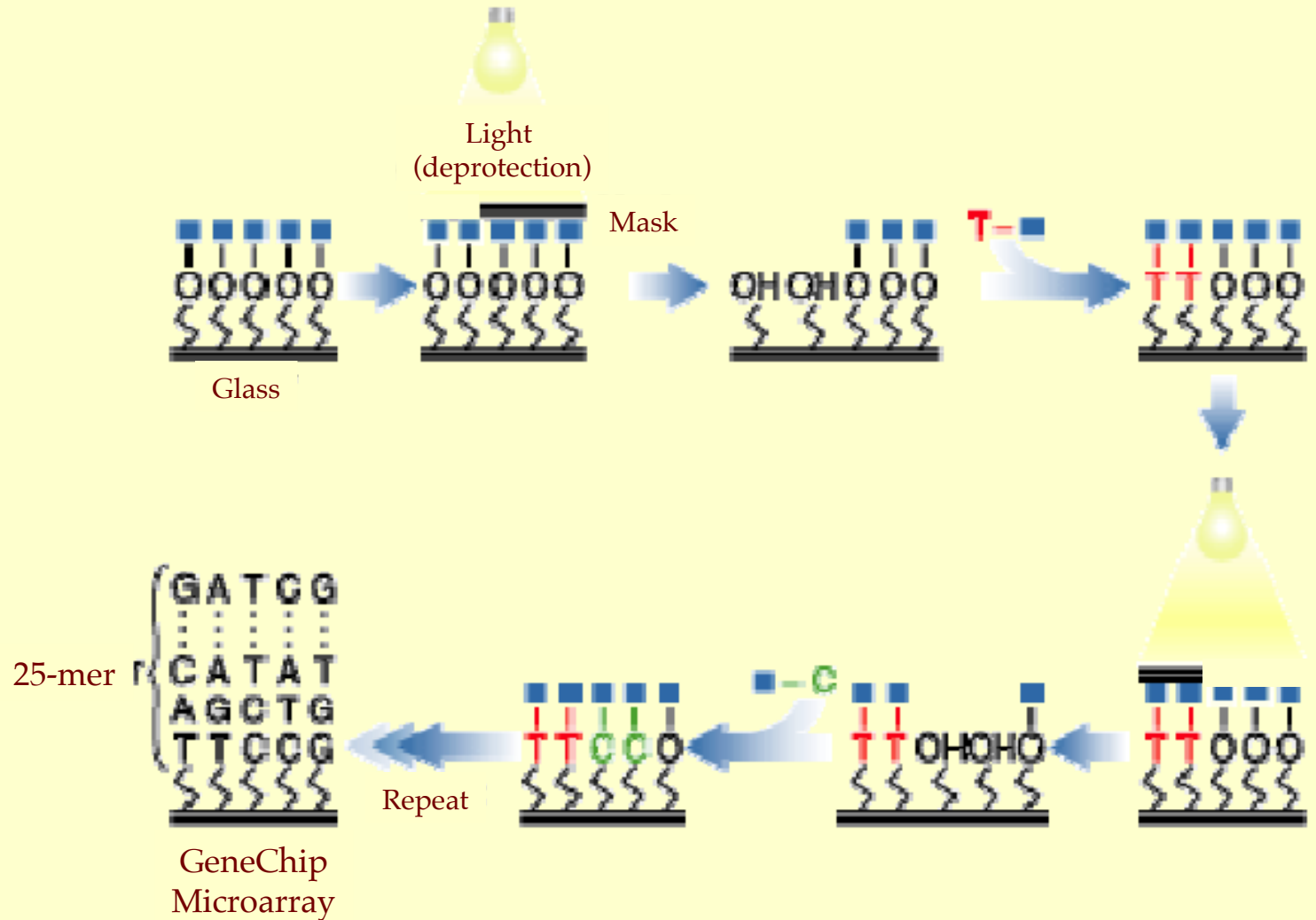
- The location and identity of each sequence on the glass surface is known.



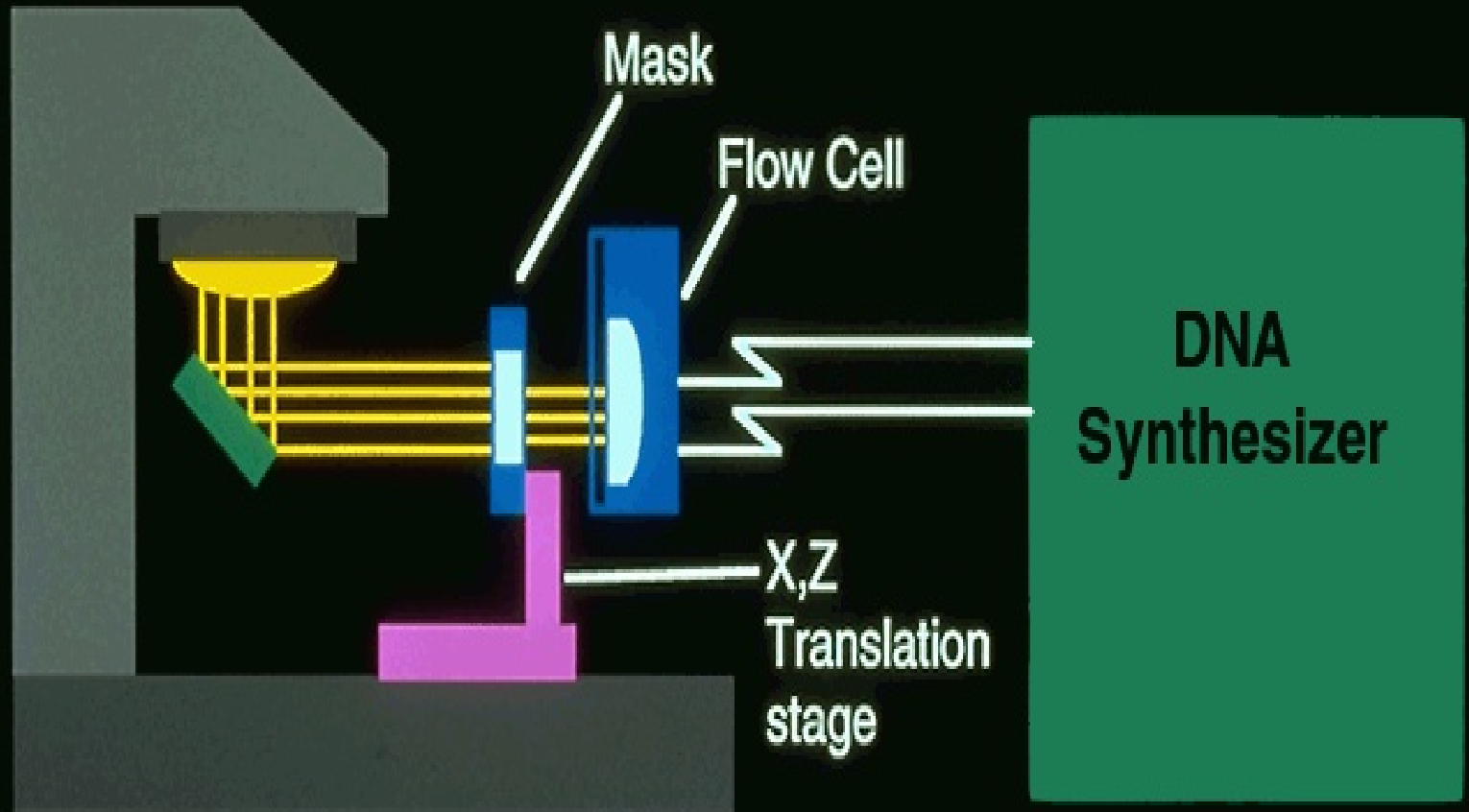
Light Directed Oligonucleotide Synthesis

<http://www.youtube.com/watch?v=ui4BOtwJEXs&feature=related>

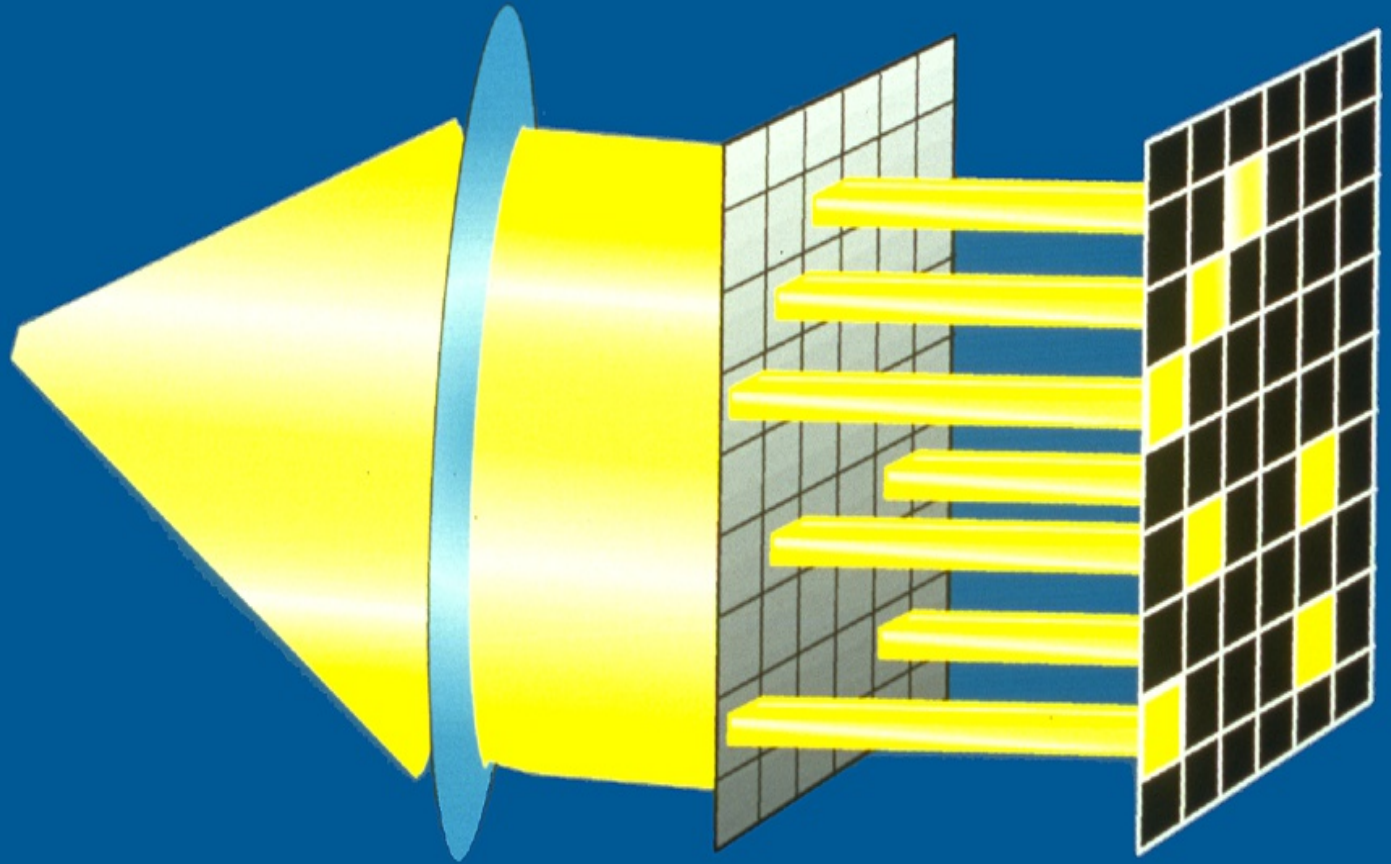
Photolithography



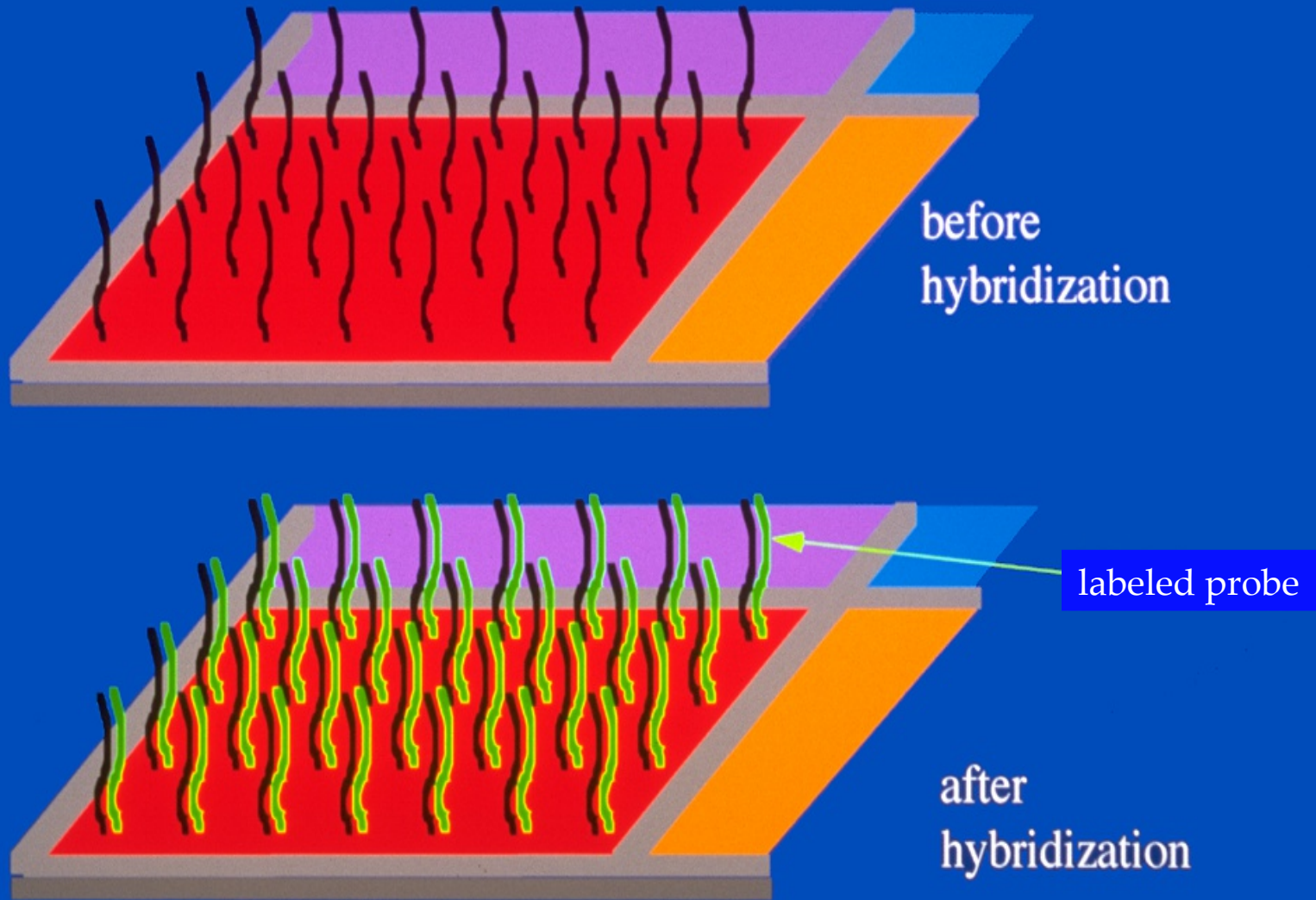
Automated DNA Chip Synthesis



Photolithography Masks



Hybridization & Detection

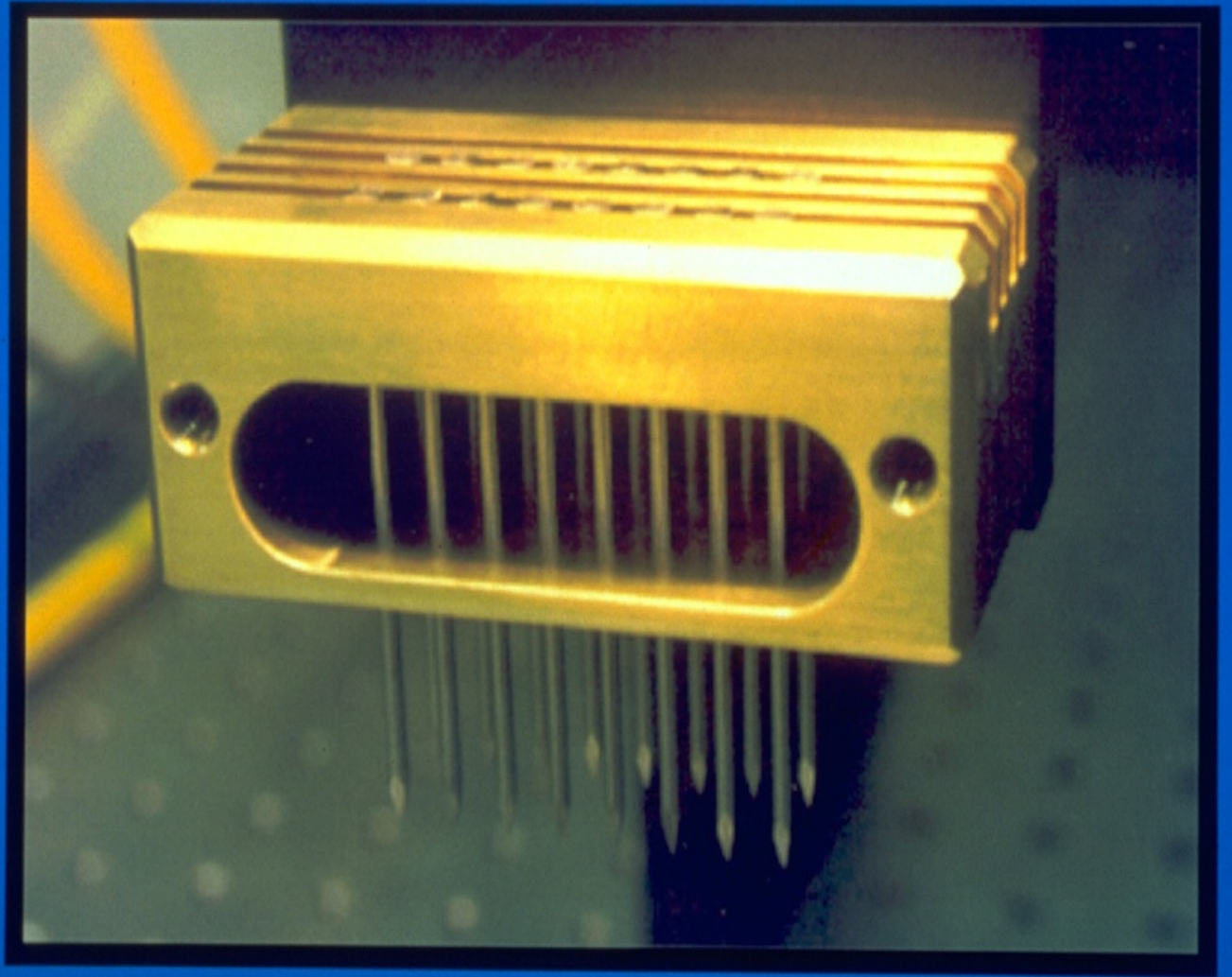


Microarrayer in Pat Brown's Lab

<http://cmgm.stanford.edu/pbrown/>



High Precision DNA Printing



Mechanical Spotting Microarrays

<http://www.arrayit.com/>

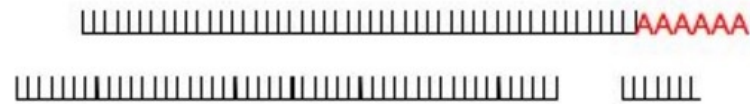
Bubble Pin Technology



RNA Seq for mRNA Profiling

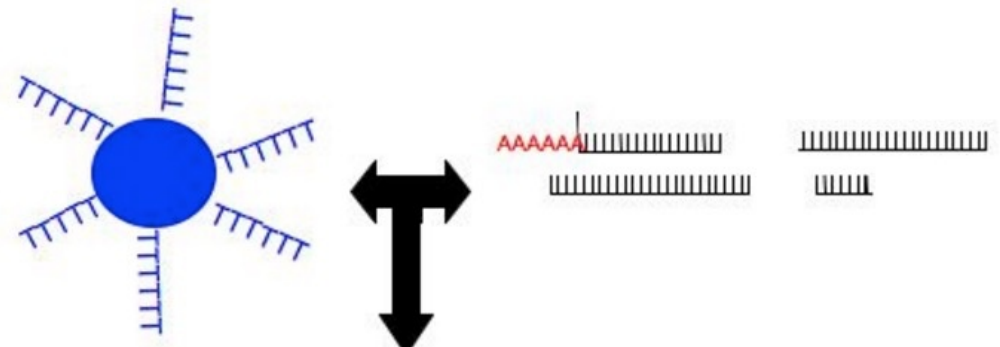
<https://en.wikipedia.org/wiki/RNA-Seq>

Isolate Total RNA

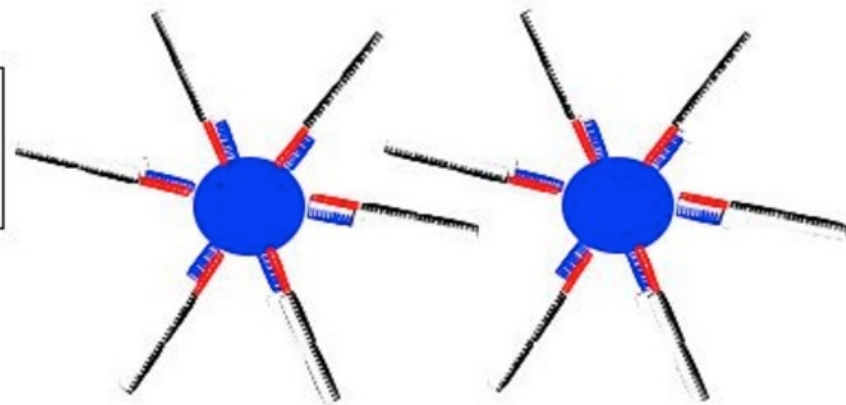


Fragmentation
and/or Isolation

In this case, isolation via Poly(T)
coated magnetic beads



Poly(A) RNA molecules
bind to the Poly(T)
magnetic beads



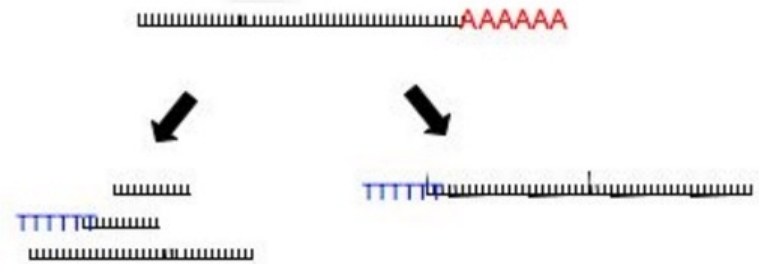
RNA Seq for mRNA Profiling

<https://en.wikipedia.org/wiki/RNA-Seq>

Magnetically isolate and wash beads



Fragment and/or Reverse Transcribe



Fragmentation (if not done already), size selection, and sequence



Illumina Solexa, Roche 454, or ABI SOLiD
Graphic shown here is Illumina

DNA Chips are used to Measure Gene Expression

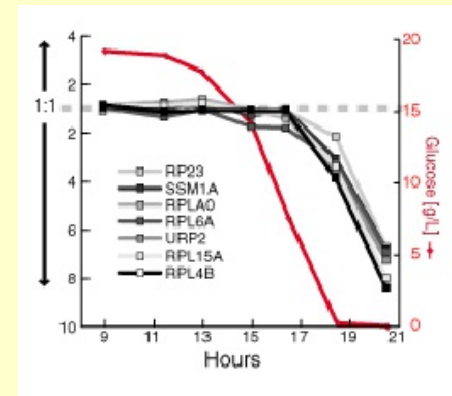
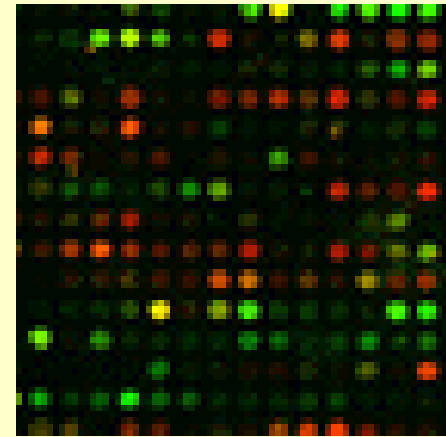
MicroArray analysis of whole genome gene expression



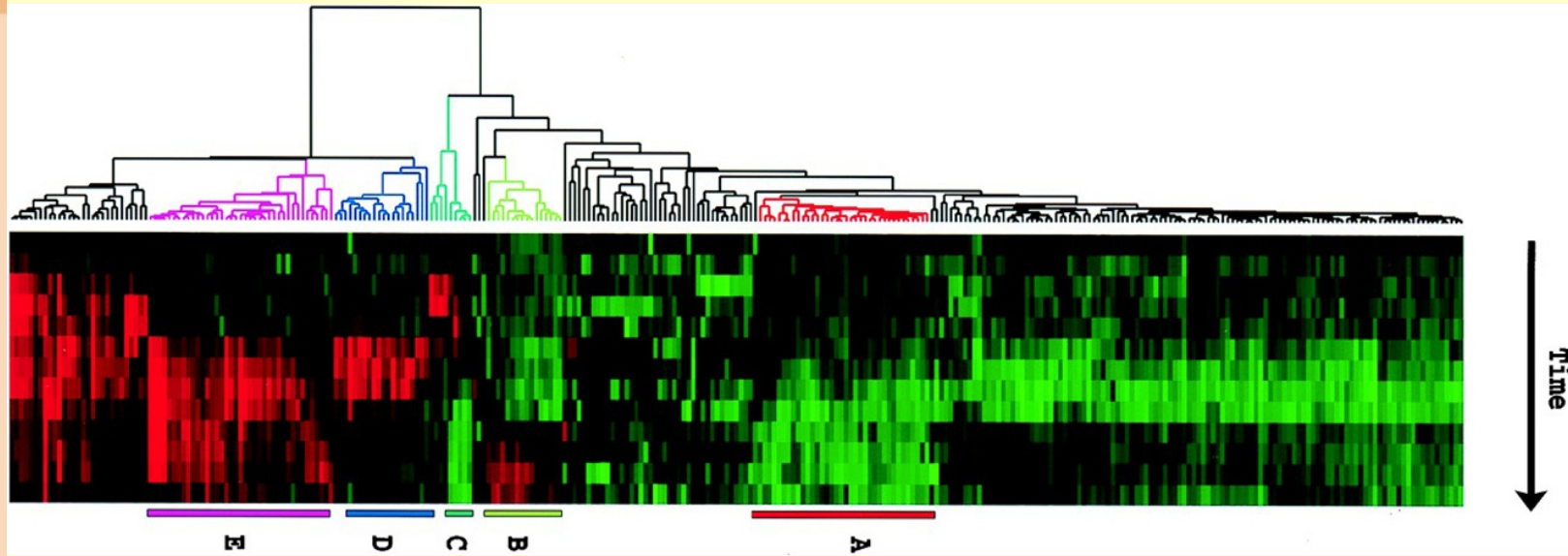
Clustering of genes based on their expression pattern



Searching for conserved sequence motifs regulating the expression



Megacluster of Yeast Gene Expression



Human Gene Expression Signatures



Finding Transcription Factor Binding Sites

Upstream Regions

Co-expressed
Genes

GATGGCTGCACCACGTGTATGC . . . ACG
CACATCGCATCACGTGACCAGT . . . GAC
GCCTCGCACGTGGTGGTACAGT . . . AAC
TCTCGTTAGGACCATCACGTGA . . . ACA
CGCTAGCCCACGTGGATCTTGA . . . AGA

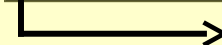
Pho 5

Pho 8

Pho 81

Pho 84

Pho ...



Finding Transcription Factor Binding Sites

Upstream Regions

Co-expressed
Genes

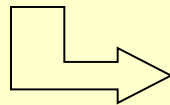
GATGGCTGCAC**CACGTG**TATGC . . . ACG**ATGTCTCGC**
CACATCGCAT**CACGTG**ACCAGT . . . GAC**ATGGACGGC**
GCCTCG**CACGTG**GTGGTACAGT . . . AAC**ATGACTAAA**
TCTCGTTAGGACCAT**CACGTGA** . . . ACA**ATGAGAGCG**
CGCTAGCC**CACGTG**GATCTTGT . . . AGA**ATGGCCTAT**

Finding Transcription Factor Binding Sites

Upstream Regions

Co-expressed
Genes

```
ATGGCTGCACCACGTTTATGC . . . ACGATGTCTCGC
CACATCGCATCACGTGACCAGT . . . GACATGGACGGC
  GCCTCGCACGTGGTGGTACAGT . . . AACATGACTAAA
TTAGGACCATCACGTGA . . . ACAATGAGAGCG
CGCTAGCCCACGTTGATCTTGT . . . AGAATGGCCTAT
```



Pho4 binding

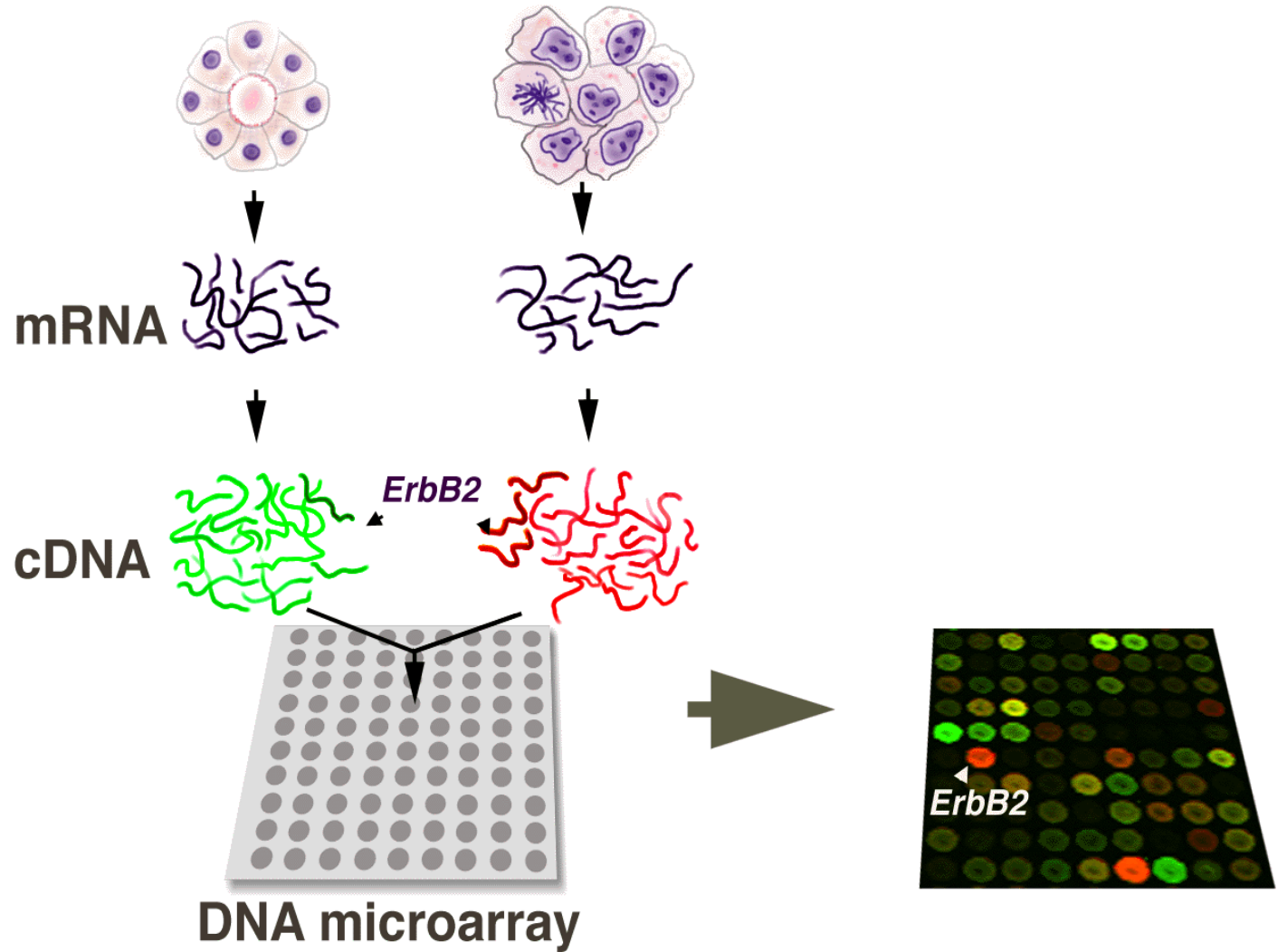
Discovering Transcription Factor Binding Sites is Difficult

- Binding sites are short (5-15 base pairs)
- Not highly conserved (as little as 50%)
- Located in long intergenic regions (>10kb)
- Not always present (false positive genes)

Three Algorithms

- **BioProspector**
 - Presented in 2000
 - Extends Gibb's sampling (stochastic method)
 - For any cluster of sequences
- **MDscan**
 - Deterministic approach
 - Enumerative
 - Very fast
 - For sequences with some ranking information
- **MotifCut and MotifScan**
 - Graph-based
 - Does not use PSSMs
 - Novel and sensitive

cDNA Diagnostics Using DNA Microarrays

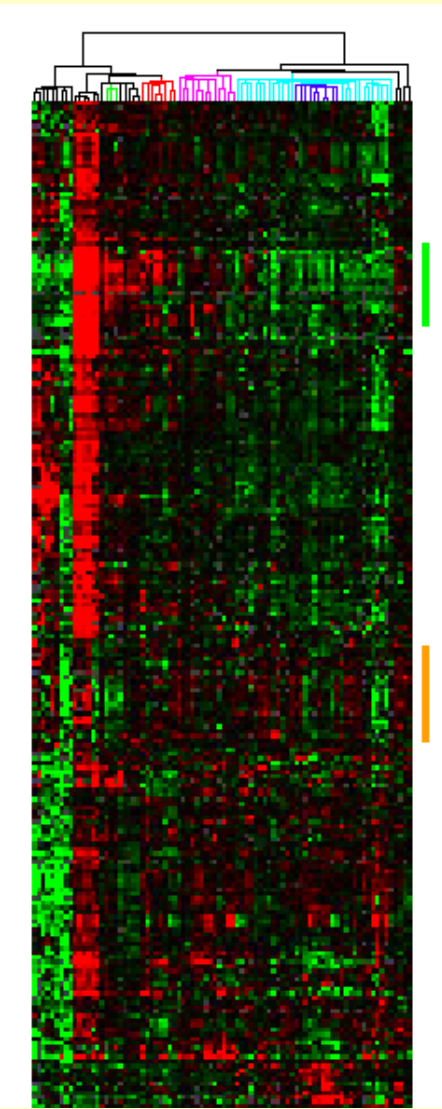


72 Breast Cancers Classified by 451 Gene Expression Assays

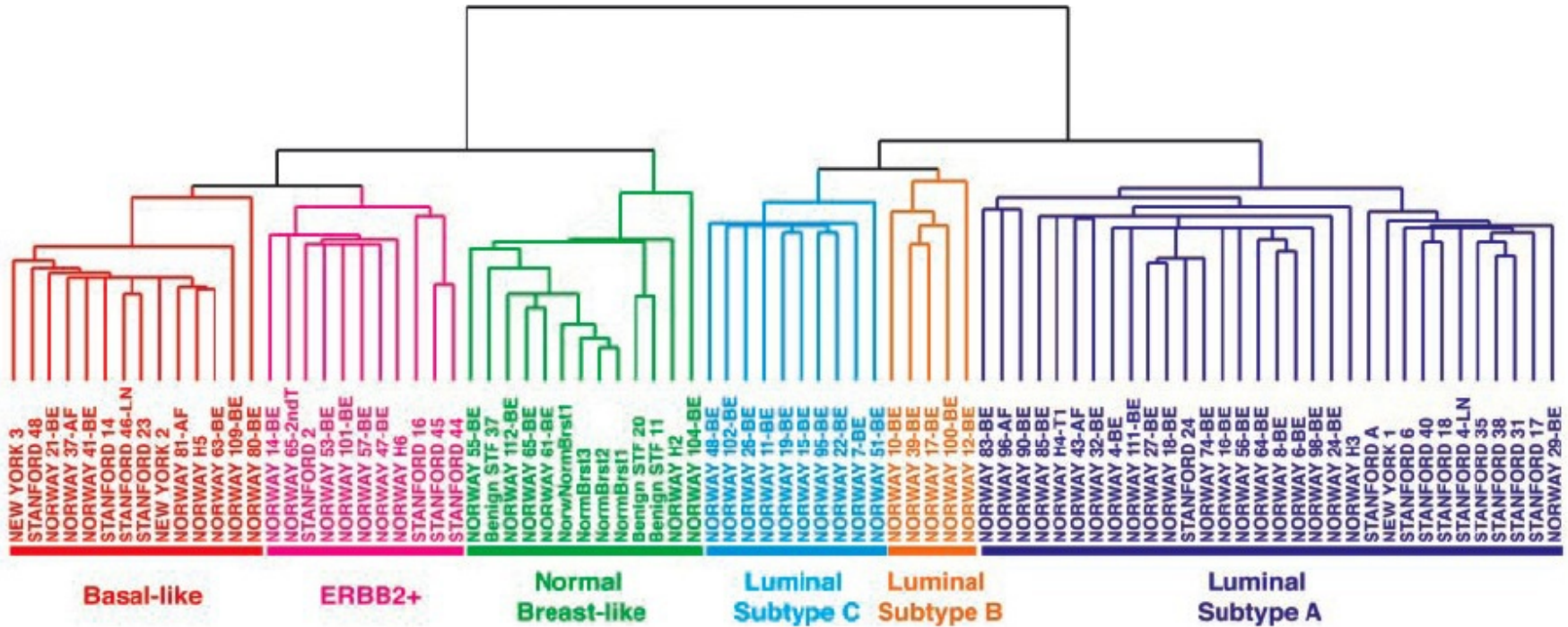


72 Patients →

↑
451 Genes



Breast Cancers Classified by 451 Gene Expression Assays



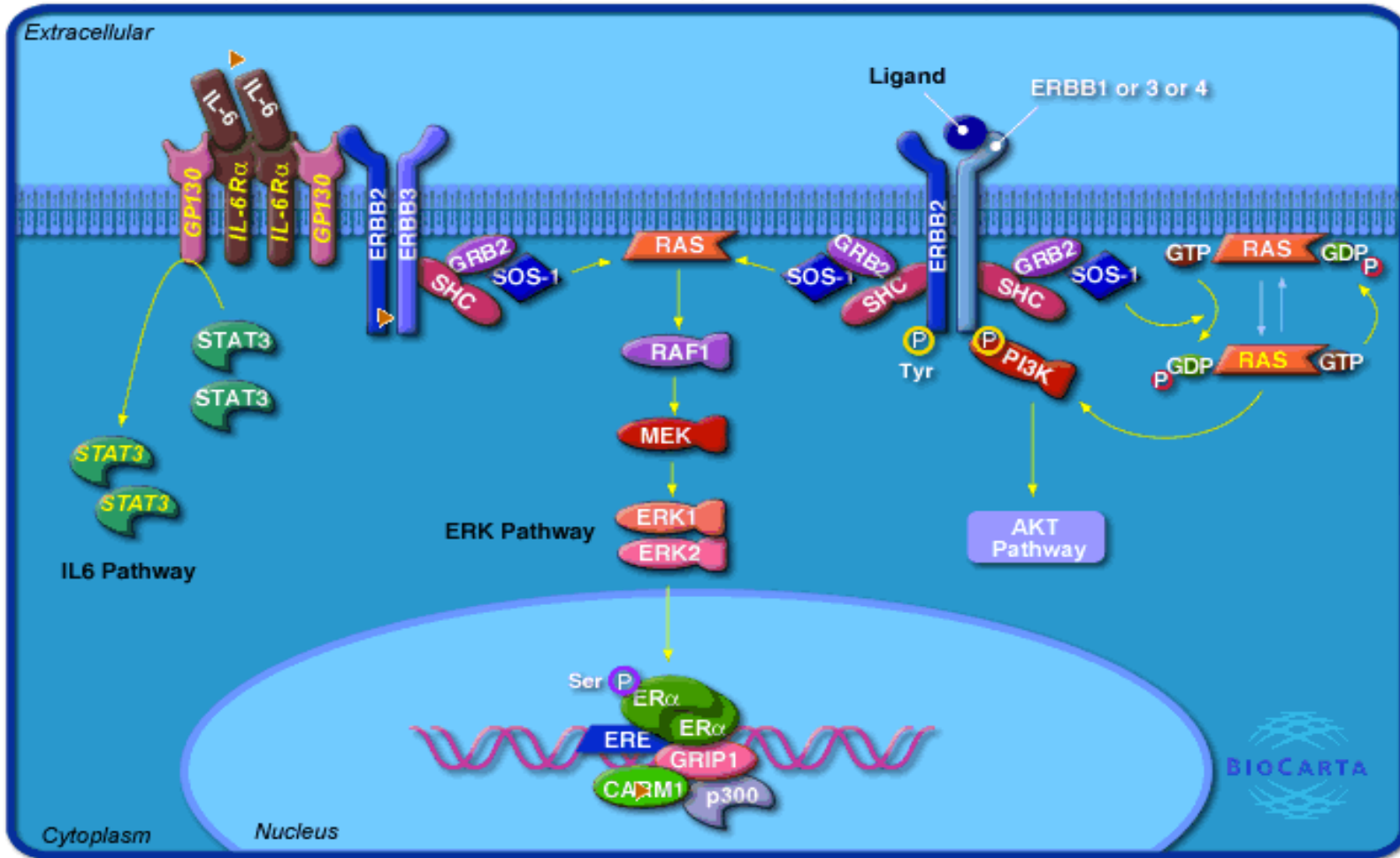
ERB-B2 in Signal Transduction & Oncology

http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways

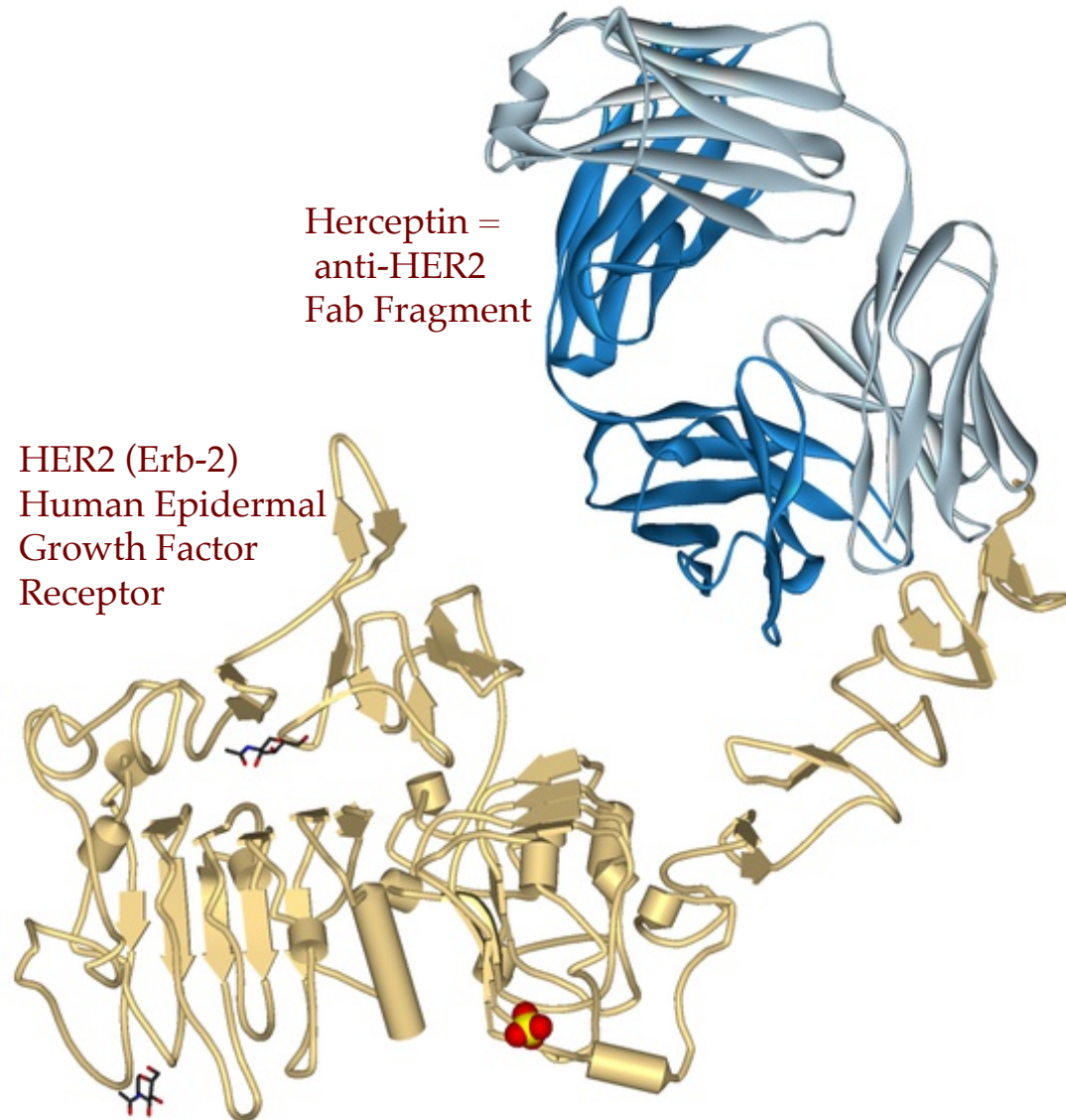
5 ▶ Role of ERBB2 in Signal Transduction and Oncology

Submitted by: ✉ Guru: ✉

[COMMENT ON THIS PATHWAY](#) | [DESCRIPTION](#) | [CONTRIBUTORS](#) | [SAVE THIS LINK](#) | [SUBMIT](#) | [LEGEND](#)

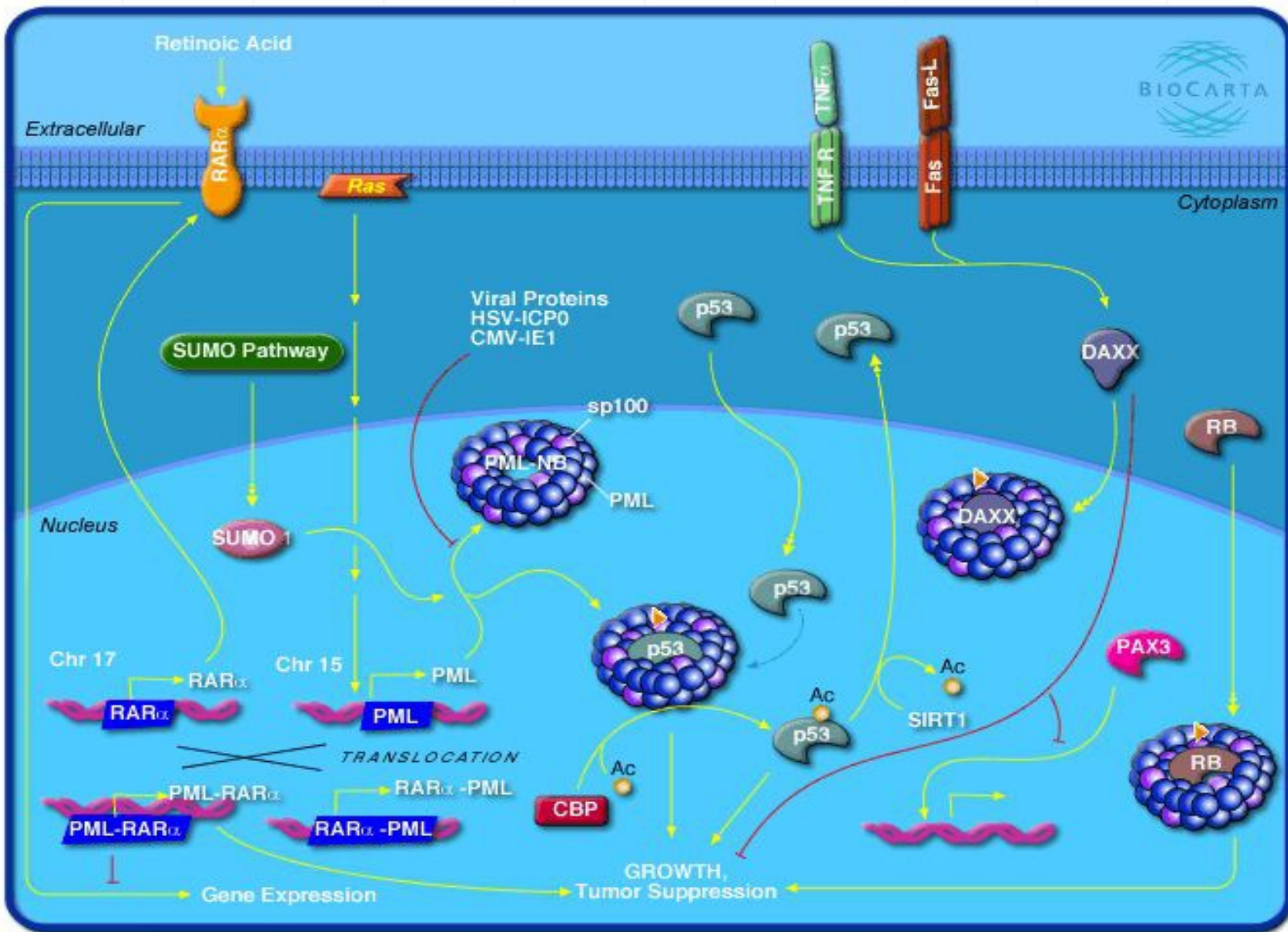


Herceptin binds to ERB-B2 (HER-2) Blocking Cell Growth



Regulation of transcriptional activity by PML

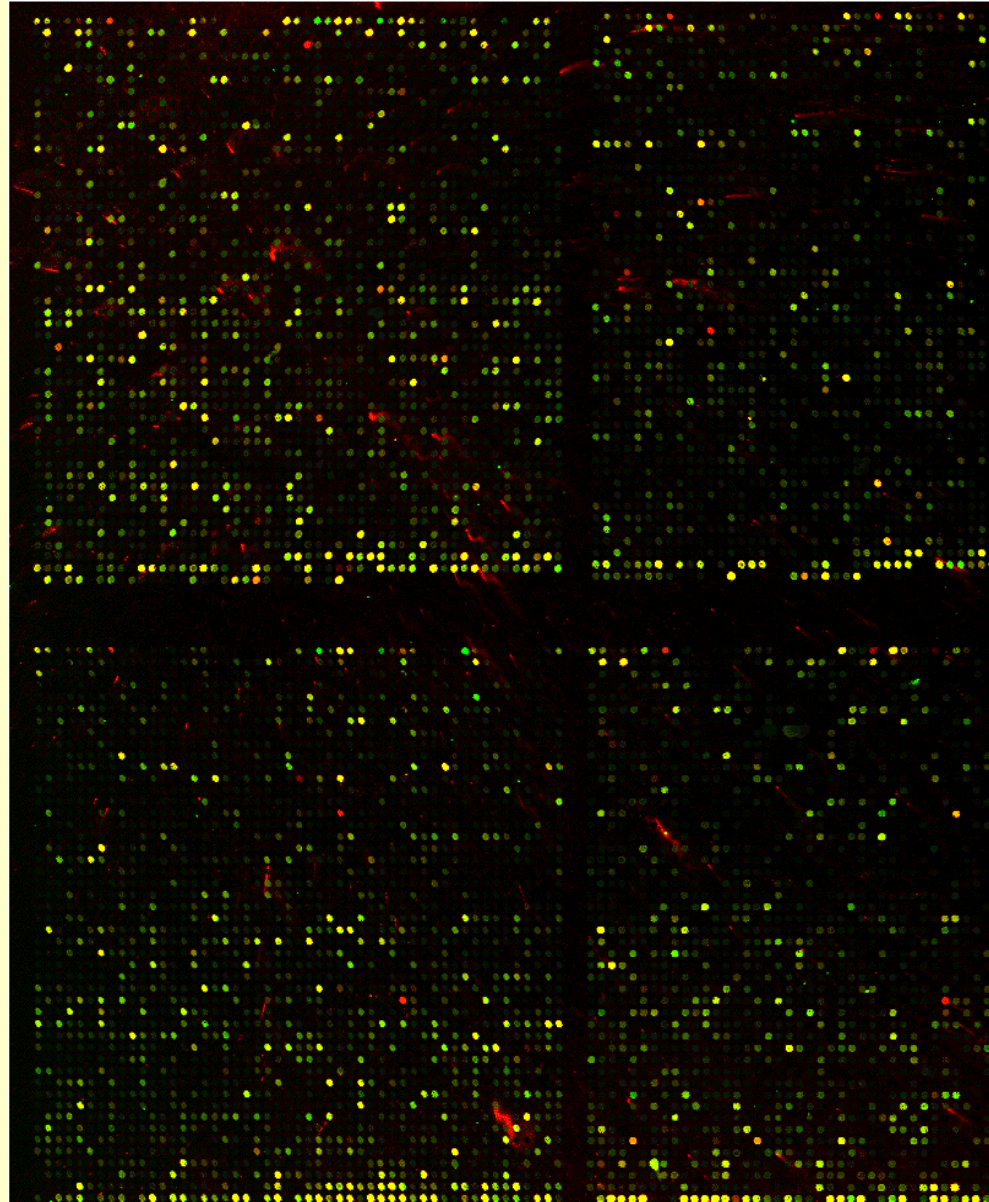
http://cgap.nci.nih.gov/Pathways/BioCarta/h_pmlPathway



Acute Promyelocytic Leukemia

Tumor cDNA + Retinoic Acid (24 hr)

(Doug Ross & Pat Brown)

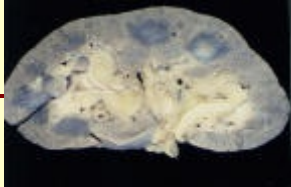


Predicting Site of Origin for Cancers of Unknown Primary

- 1844 tumors tested one at a time versus all 18 tissues of origin
- Retrospective study on well characterized patient samples
- Uses PathChip (functionally similar to Affymetrix HU-133A GeneChip)
 - 604 specimens used for training
 - 636 specimens used for test
 - 604 specimens in reserve for final validation
- Reproducibility from lab to lab
- Performance based on sensitivity (> 70%) & accuracy (> 95%)

Predicting Site of Origin for Cancers of Unknown Primary

Sample

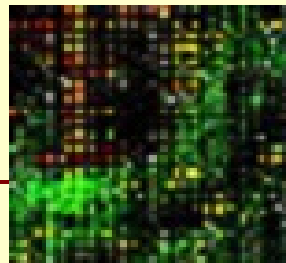


Clinical Laboratory

- ❖ Receive sample
- ❖ Process specimen, mRNA extraction and microarray hybridization
- ❖ Submit data file to PathWork for processing

- ❖ Deliver report to oncologist

Microarray data file



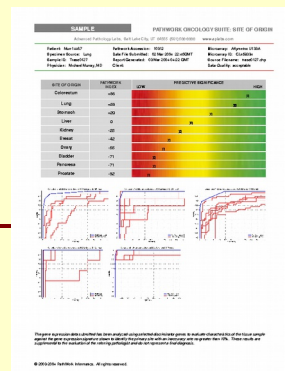
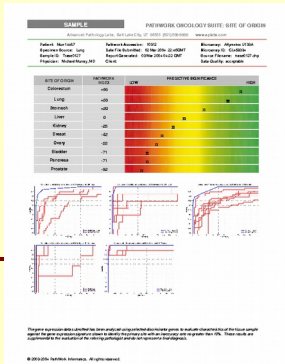
I
N
T
E
R
N
E
T

- ❖ Receive microarray data file
- ❖ Check file and data quality

- ❖ Run model against microarray data file

- ❖ Generate report file
- ❖ Deliver report file back to clinical lab

PDF Report

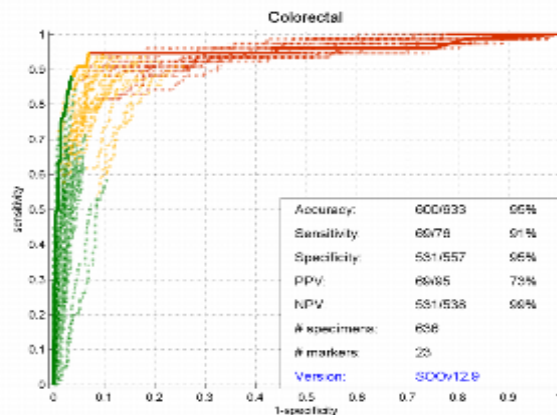


**DEMONSTRATION
EXAMPLE REPORT**

**PATHWORK ONCOLOGY SUITE: SITE OF ORIGIN
GENE EXPRESSION-BASED MOLECULAR DIAGNOSTIC**

Patient ID: Liv18350	Client Name:	Data Quality: Acceptable
Specimen Source: Liver	Client ID:	Microarray: Affymetrix U133A
Medical Record:	Order Pathologist:	Microarray ID: CL18350hg133a11
Case Number:	Treating Physician:	Source Filename: 18350hg133a11.cel
PathWork Accession: SOO7982	Date/Time Processed: 29.Jul.2005 11:54PDT	Version: SOOv12.9

SITE OF ORIGIN	PPV	NPV	SIMILARITY SCORE	NEGATIVE	SIMILARITY SCORE	POSITIVE
Colorectal	73%	99%	92	[Color scale bar from red to green]		
Hepatocellular	100%	100%	57	[Color scale bar from red to green]		
Gastroesophageal	48%	98%	-37	[Color scale bar from red to green]		
Pancreatic	88%	99%	-40	[Color scale bar from red to green]		
Non-Small Cell Lung Cancer	85%	98%	-48	[Color scale bar from red to green]		
Ovarian	98%	98%	-72	[Color scale bar from red to green]		
Breast	93%	98%	-74	[Color scale bar from red to green]		
Malignant Pleural Mesothelioma	100%	100%	-80	[Color scale bar from red to green]		
Kidney	97%	99%	-81	[Color scale bar from red to green]		
Squamous	48%	100%	-86	[Color scale bar from red to green]		
Thyroid	95%	100%	-88	[Color scale bar from red to green]		
Bladder	67%	99%	-89	[Color scale bar from red to green]		
Prostate	91%	100%	-89	[Color scale bar from red to green]		
Central Nervous System	100%	100%	-95	[Color scale bar from red to green]		
Germ Line	88%	100%	-96	[Color scale bar from red to green]		
Non Hodgkin's Lymphoma	91%	100%	-100	[Color scale bar from red to green]		
Melanoma	91%	100%	-100	[Color scale bar from red to green]		
Soft Tissue-Sarcoma	39%	99%	-100	[Color scale bar from red to green]		



KEY

Sample Site: Liver
 Primary Site: CO
 Percent Tumor: 50-75
 Pathology Morphology: Adenocarcinoma
 Sample Description: PARTIAL
 HEPATECTOMY; METASTATIC
 ADENOCARCINOMA PRIMARY IN
 COLON, STATUS POST
 CHEMOTHERAPY.

For use by CLIA-certified clinical laboratories only. The clinical interpretation of the results should be made in context of the patient's clinical history and other diagnostic tests performed by a qualified individual.

Advanced Pathology Labs
 123 Diagnostic Drive, Multiplex, CA 99212 (800) 555-1212

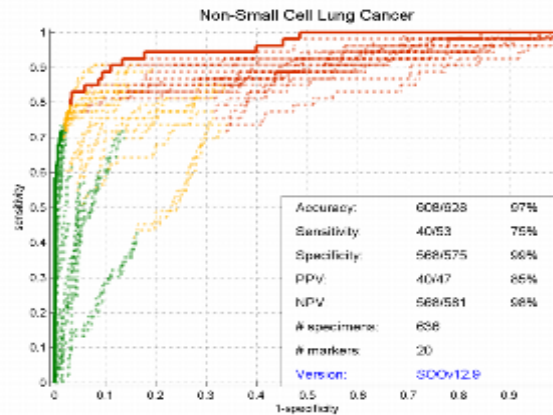
©2003-2005 PathWork Informatics. All rights reserved.

**DEMONSTRATION
EXAMPLE REPORT**

**PATHWORK ONCOLOGY SUITE: SITE OF ORIGIN
GENE EXPRESSION-BASED MOLECULAR DIAGNOSTIC**

Patient ID: Lun31878	Client Name:	Data Quality: Acceptable
Specimen Source: Lung	Client ID:	Microarray: Affymetrix U133A
Medical Record:	Order Pathologist:	Microarray ID: CL31878hg133a11
Case Number:	Treating Physician:	Source Filename: 31878hg133a11.cel
PathWork Accession: SOO6030	Date/Time Processed: 29.JUL.2005 11:54PDT	Version: SOOV12.9

SITE OF ORIGIN	PPV	NPV	SIMILARITY SCORE	NEGATIVE	SIMILARITY SCORE	POSITIVE
Non-Small Cell Lung Cancer	85%	98%	98			
Colorectal	73%	99%	2			
Gastroesophageal	48%	98%	-43			
Non Hodgkin's Lymphoma	91%	100%	-56			
Pancreatic	86%	99%	-57			
Kidney	97%	99%	-65			
Squamous	48%	100%	-68			
Melanoma	91%	100%	-70			
Bladder	67%	99%	-71			
Ovarian	96%	98%	-71			
Malignant Pleural Mesothelioma	100%	100%	-73			
Thyroid	95%	100%	-74			
Breast	93%	98%	-75			
Central Nervous System	100%	100%	-79			
Prostate	91%	100%	-89			
Hepatocellular	100%	100%	-90			
Germ Line	88%	100%	-91			
Soft Tissue-Sarcoma	39%	99%	-93			



KEY

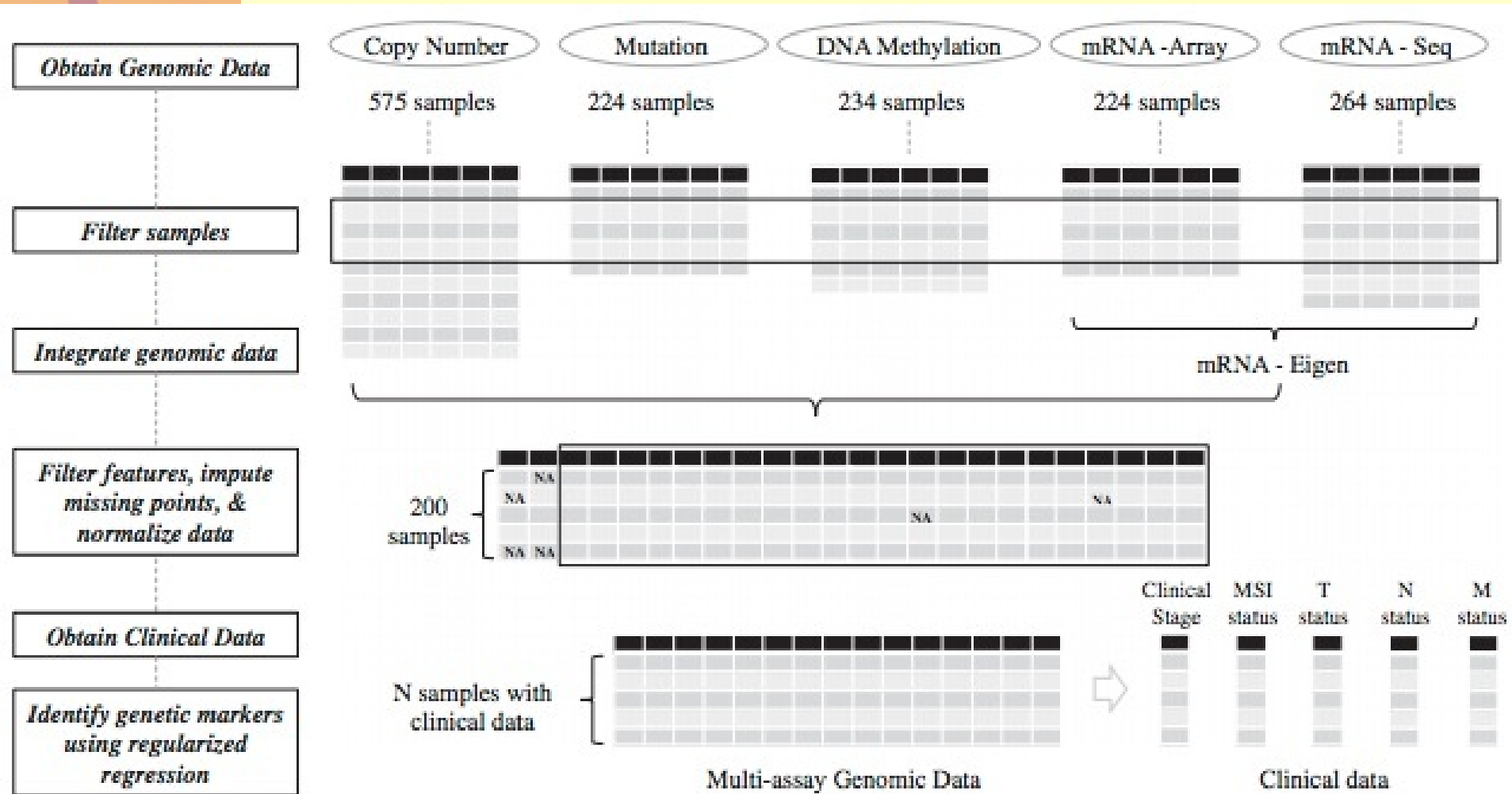
Sample Site: Lung
 Primary Site: CO
 Percent Tumor: 25-50
 Pathology Morphology: Adenocarcinoma
 Sample Description: LULLUNG WEDGE BIOPSY; METASTATIC ADENOCARCINOMA C/W COLONIC PRIMARY DIAGNOSED 7 YEARS PRIOR; STAGING AT TIME OF ORIGINAL DIAGNOSIS UNKNOWN.

For use by CLIA-certified clinical laboratories only. The clinical interpretation of the results should be made in context of the patient's clinical history and other diagnostic tests performed by a qualified individual.

Advanced Pathology Labs
 123 Diagnostic Drive, Multiplex, CA 99212 (800) 555-1212

Genes Associated with Colorectal Cancer

<http://www.biomedcentral.com/content/pdf/1755-8794-6-54.pdf>



Genes Associated with Colorectal Cancer

<http://www.biomedcentral.com/content/pdf/1755-8794-6-54.pdf>

Genomic assay	Direct association	Inverse association
Copy number	Increase with advanced stage	Decrease with advanced stage
Mutation	More frequent in advanced stage	Less frequent in advanced stage
Gene expression	Higher expression in advanced stage	Lower expression in advanced stage
Methylation	Hyper-methylation in advanced stage	Hypo-methylation in advanced stage
Examples	<p>A B C</p>	<p>D E F</p>

Genes Associated with Colorectal Cancer

<http://www.biomedcentral.com/content/pdf/1755-8794-6-54.pdf>

Copy Number			
Rank	Gene	Chr	Sign
1	<i>FUBP1*</i>	1p31.1	↓
2	<i>CUX1</i>	7q22.1	↑
3	<i>WRN*</i>	8p12	↓
4	<i>MALT1*</i>	18q21.32	↓
5	<i>ERCC5*</i>	13q33.1	↑
6	<i>HOXA13*</i>	7p15.2	↑
7	<i>HOXA11*</i>	7p15.2	↑
8	<i>HOXA9*</i>	7p15.2	↑
9	<i>IPO5</i>	13q32.2	↑

9 out of 9 genes

Mutations		
Rank	Gene	Sign
1	<i>GNAS*</i>	↑
2	<i>ESRRA</i>	↑
3	<i>RUNX1T1*</i>	↓
4	<i>KAT6B*</i>	↓
5	<i>GPHN*</i>	↓
6	<i>STIL*</i>	↓
7	<i>PALB2*</i>	↓

7 out of 7 genes

Methylation		
Rank	Gene	Sign
1	<i>WRN*</i>	↑
2	<i>FOXP4</i>	↓
3	<i>HSD3B7</i>	↓
4	<i>SYK*</i>	↓
5	<i>PRELP</i>	↓
6	<i>MUC1*</i>	↑
7	<i>NTRK1*</i>	↑
8	<i>RNF113A</i>	↑
9	<i>GPR125</i>	↓
10	<i>MGMT</i>	↓

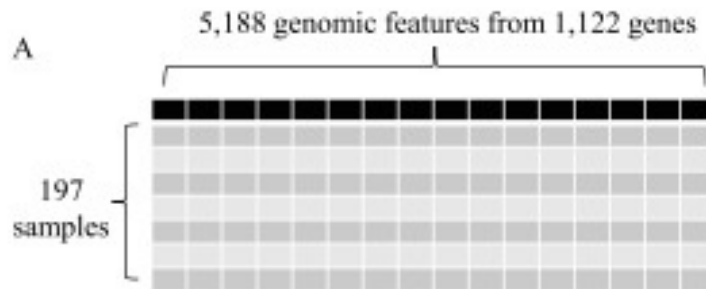
10 out of 33 genes

mRNA		
Rank	Gene	Sign
1	<i>PDK4</i>	↑
2	<i>REEP5</i>	↑
3	<i>TH</i>	↑
4	<i>DDX5*</i>	↓
5	<i>CYP26B1</i>	↓
6	<i>SLC22A11</i>	↓
7	<i>WRN*</i>	↓
8	<i>OR51E2</i>	↑
9	<i>TNFAIP3*</i>	↓
10	<i>DDIT3*</i>	↑

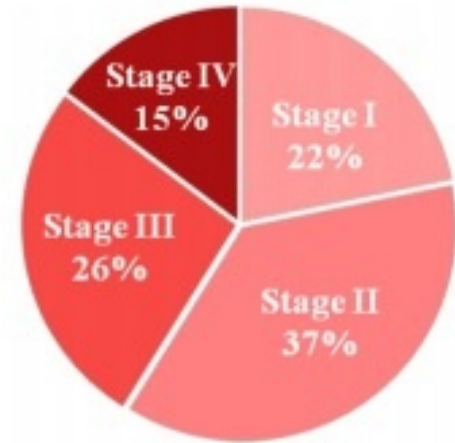
10 out of 65 genes

Genes Associated with Colorectal Cancer

<http://www.biomedcentral.com/content/pdf/1755-8794-6-54.pdf>



Clinical stage

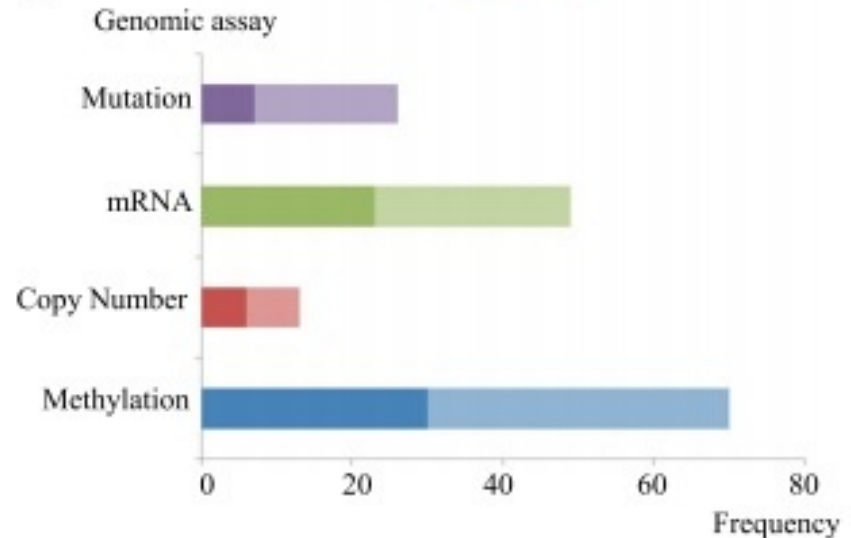


B

Rank	Gene	Assay	Chr	Sign
1	<i>FOXP4</i>	Methyl	6p21.1	↓
2	<i>REEP5</i>	mRNA	5q22.2	↑
3	<i>PDK4</i>	mRNA	7q21.3	↑
4	<i>OR51E2</i>	mRNA	11p15.4	↑
5	<i>WRN</i>	CN	8p12	↓
6	<i>SYK</i>	Methyl	9q22.2	↓
7	<i>S100P</i>	mRNA	4p16.1	↓
8	<i>HIP1</i>	mRNA	7q11.23	↑
9	<i>SDHC</i>	mRNA	1q23.3	↑
10	<i>DDIT3</i>	mRNA	12q13.3	↑
11	<i>ADRA2C</i>	mRNA	4p16.3	↑
12	<i>DDX5</i>	mRNA	17q23.3	↓
13	<i>SYK</i>	mRNA	9q22.2	↓
14	<i>FCRL4</i>	mRNA	1q23.1	↓
15	<i>SLC22A11</i>	mRNA	11q13.1	↓

Top 15 out of 158 selected features

C



Genes Associated with Colorectal Cancer

Table 3 Top 25 candidates associated with advanced clinical stage

Gene	Chr	Score	Copy Number	Elastic-net Feature Rank		
				Gene Expression	Methylation	Mutation
<i>WRN*</i>	8p12	27.03	5 ↓	106 ↓	29 ↑	
<i>SYK*</i>	9q22.2	24.09		13 ↓	6 ↓	
<i>DDX5*</i>	17q23.3	18.94	70 ↓	12 ↓		
<i>ADRA2C</i>	4p16.3	18.13		11 ↑	81 ↓	
<i>GNAS*</i>	20q13.32	16.9			39 ↓	68 ↑
<i>SEMA3B</i>	3p21.31	16.66		82 ↓	28 ↓	
<i>HSD17B2</i>	16q23.3	14.45	17 ↑	120 ↑		
<i>TTN</i>	2q31.2	13.72		80 ↓		66 ↓
<i>FHIT*</i>	3p14.2	13.56			71 ↑	77 ↓
<i>HIST1H4I*</i>	6p22.1	12.82		35 ↑	122 ↓	
<i>FOXP4</i>	6p21.1	12.74			1 ↓	
<i>REEP5</i>	5q22.2	12.66		2 ↑		
<i>PDK4</i>	7q21.3	12.58		3 ↑		
<i>OR51E2</i>	11p15.4	12.49		4 ↑		
<i>S100P</i>	4p16.1	12.25		7 ↓		
<i>HIP1*</i>	7q11.23	12.17		8 ↑		
<i>ZNF570</i>	19q13.12	12.09		50 ↑		116 ↓
<i>SDHC*</i>	1q23.3	12.09		9 ↑		
<i>DDIT3*</i>	12q13.3	12		10 ↑		
<i>CRTC1*</i>	19p13.11	11.92			130 ↑	38 ↓
<i>FCRL4*</i>	1q23.1	11.84		14 ↓	155 ↓	
<i>SLC22A11</i>	11q13.1	11.6		15 ↓		
<i>FLT1</i>	13q12.2	11.51		16 ↑		
<i>CYP26B1</i>	2p13.2	11.35		18 ↓		
<i>RNF113A</i>	Xq24	11.27			19 ↑	

The Genetics of Personalized Cancer Medicine

DEC
05
2012

The Genetics of Personalized Cancer Medicine

Quicktime

Hanlee Ji, MD

Assistant Professor of Medicine - Oncology, Stanford School of Medicine

Cystic Fibrosis Array

