Agfa Healthcare Futures Symposium

Shankar Subramaniam

Departments of Bioengineering, Chemistry and Biochemistry

Bioinformatics Program

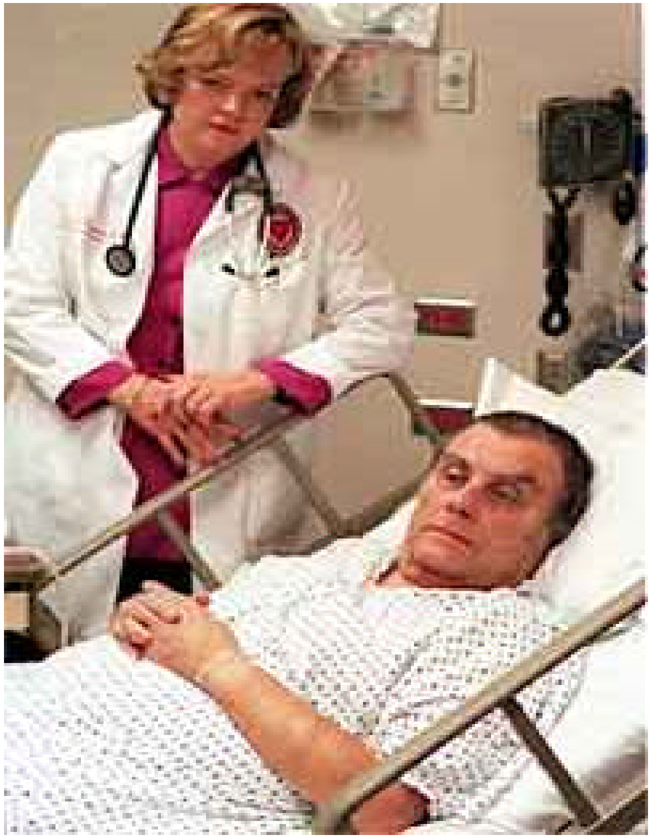
San Diego Supercomputer Center

University of California at San Diego

Bioinformatics

- The impact of genomics
- Systems Approaches to Molecular Medicine
- New Technologies in Molecular Medicine
- Tracking Cells, Tissues and Procedures
- Structuring Knowledge
- Animal Models and Comparative Approaches
- Pharmacogenomics
- In vivo Imaging

IMPACT OF GENOMICS ON HUMAN DISEASES



For Warren Wegele, an alteration in one of his genes caused the treatment that he received for congestive heart failure to be ineffective.

"I've never really gotten ill, and I've always recovered from everything instantly," said Mr. Wegele, who eventually had a heart transplant.

Tiny Changes, Big Effects

Single-molecule variations in genes that have no discernible effect in healthy people can foretell the course of disease or a person's response to drugs when an individual becomes ill

The beta-2 adrenergic receptor gene, which is active in heart and lung cells, illustrates how tiny changes can have enormous medical consequences

IN THE HEART

The beta-2 adrenergic receptor gene directs cells to produce a protein that makes the heart pump

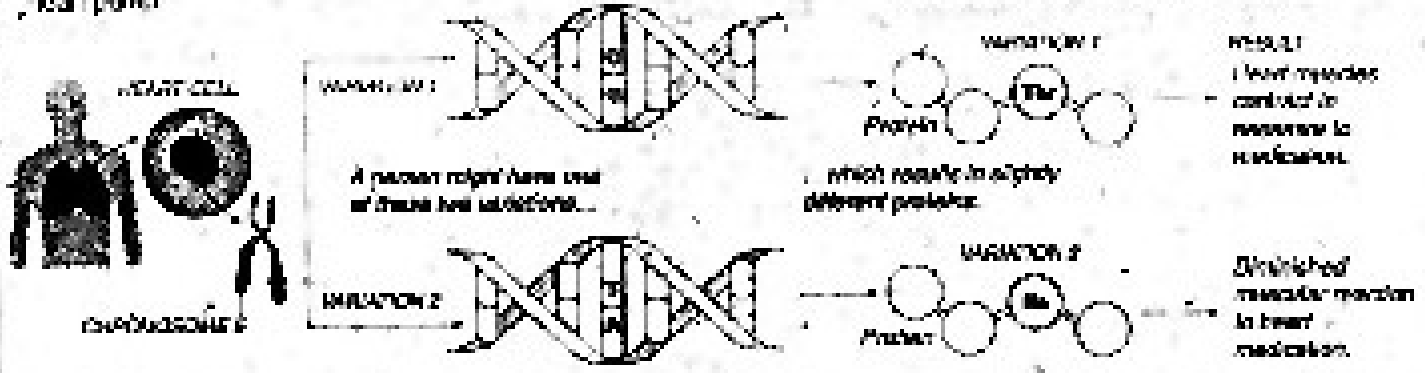
GENETIC VARIATIONS AFFECT THE COURSE OF DISEASE

DNA STRAND (THE GENE)

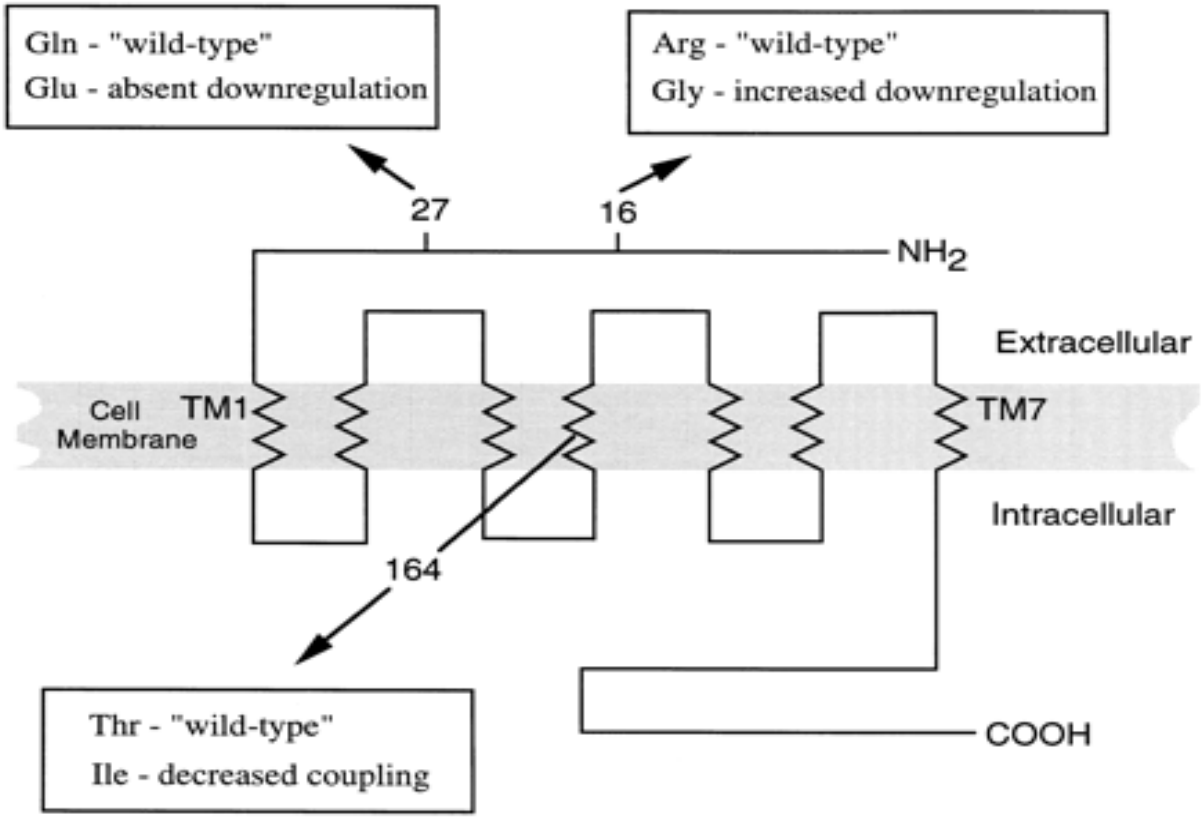
DNA is the code for protein production. Normal variation in the code slightly alter the chemical string of the protein made from that gene.

PROTEIN

The proteins have a variation in their sequence. They function fine unless the person develops congestive heart failure. Then variation 2 adversely affects the patient's prognosis.



Using Gene Testing to Customize a Patient's Medical Treatment



The Ile164 receptor displays a small decrease in binding affinity for catecholamines and certain b-AR antagonists, a substantial decrease in basal and epinephrine-stimulated adenylyl cyclase activities due to defective coupling of the receptor to the stimulatory G protein G_s, and impaired agonist-promoted sequestration.

The Ile164 2-Adrenergic Receptor Polymorphism Adversely Affects the Outcome of Congestive Heart Failure J. Clin. Invest. 1998 102: 1534-1539. Stephen B. Liggett, et al.

Human genetic identity

- Genomic sequence 99.9% identical
- 3,200,000 nucleotides different
- Single base differences in genomes between any two individuals: ca. 3 million
- Amino acid differences in proteomes between any two individuals: ca. 100,000

Variation types

- **Macro:**
  - **Chromosome numbers**
  - **Segmental duplications, rearrangements, and deletions**
- **Medium:**
  - **Sequence Repeats**
  - **Transposable Elements**
  - **Short Deletions, Sequence and Tandem Repeats (including microsatellites)**
- **Micro:**
  - **Single Nucleotide Polymorphisms (SNPs)**
  - **Single Nucleotide Insertions and Deletions (Indels)**

Human Genome and SNPs

- Now that the human genome is (mostly) sequenced, attention turning to the evaluation of variation
- Alterations in DNA involving a single base pair are called single nucleotide polymorphisms, or SNPs
- Map of ~1.4 million SNPs (Feb 2001)
- It is estimated that ~60,000 SNPs occur within exons; 85% of exons within 5 kb of nearest SNP

Number of SNP's

- **When chromosomes from two random people are compared, they differ at about one in 1000 DNA sites**
- **Thus, when two random haploid genomes are compared , there are about 3 million differences**
- **There are probably between 10 and 30 million SNP's in humans, about one every 100 to 300 bases**
- **Of these SNP's, perhaps 4 million are common SNP's, with both alleles of each SNP having a frequency above 20 percent**
- **"2.4 million SNP's have been discovered in the human genome" (National Human Genome Research Institute, July, 2001)**

Single Nucleotide Polymorphism (SNP)

GATTTAGATCGCGATAGAG
GATTTAGATCTCGATAGAG

A SNP is a position in a genome at which two or more different bases occur in the population, each with a frequency >1%.

- **The most abundant type of polymorphism**

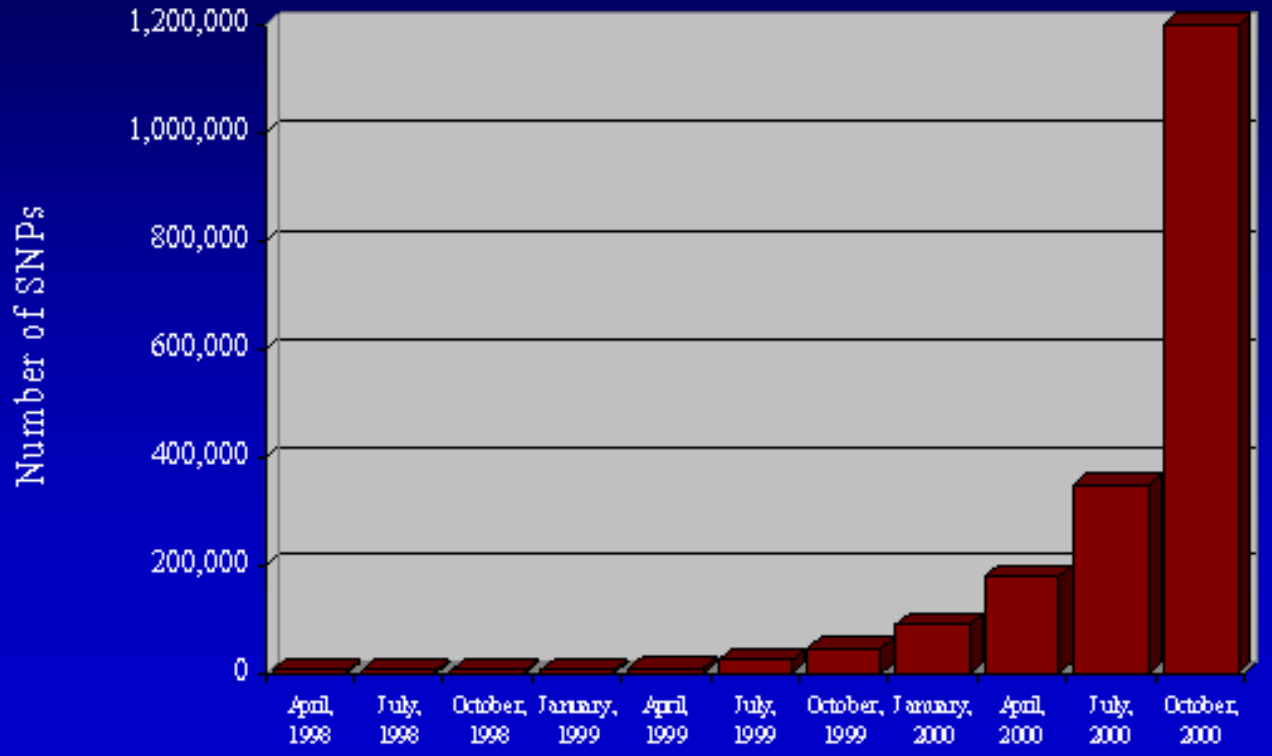
NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy.

The International SNP Working Group (Lander et al)

An explosion in publicly available SNPs

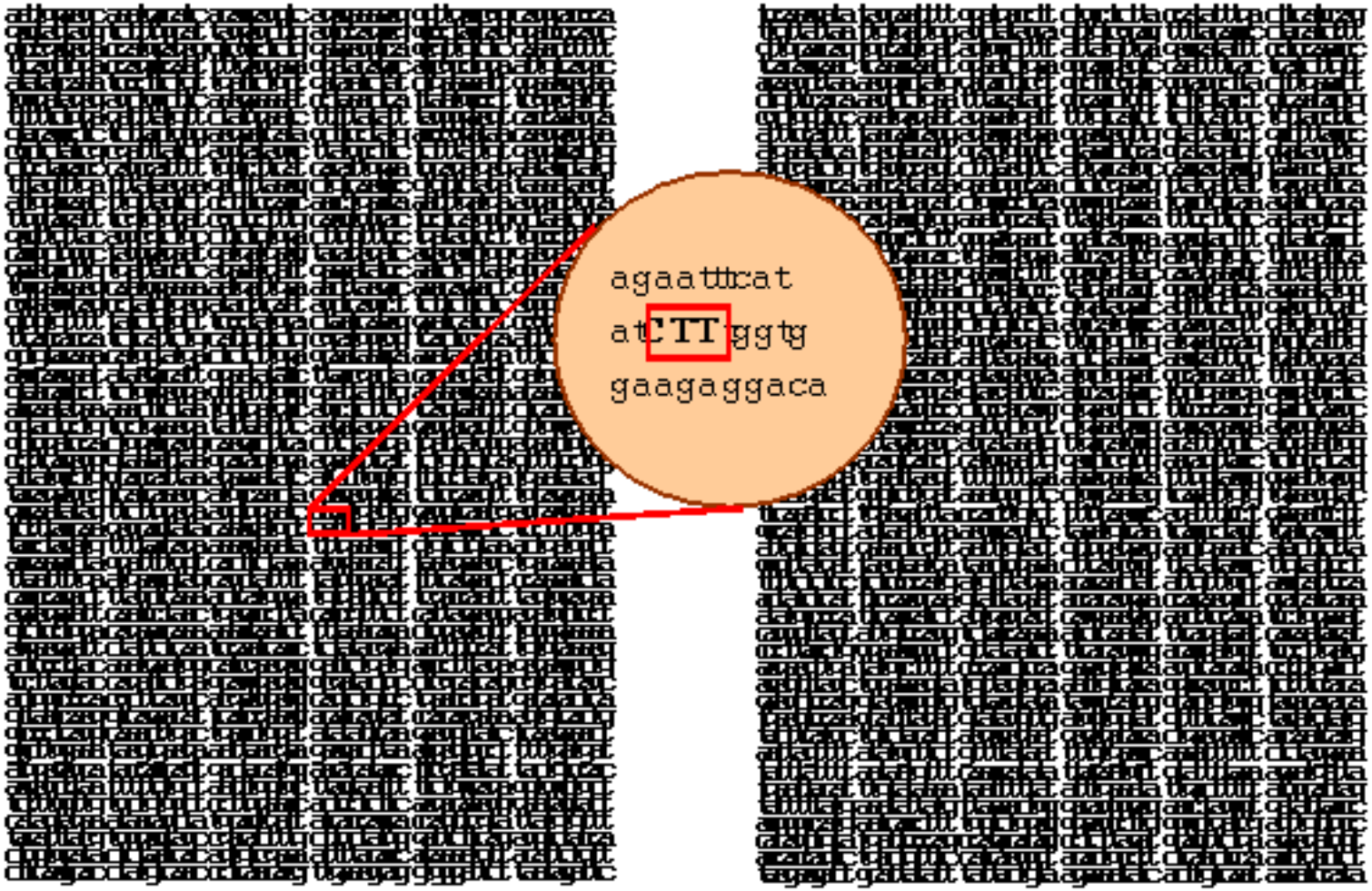


Altshuler et al., Nature, 2000 and http://snp.cshl.org

MONOGENEIC DISEASES

- **The goal of much genetic research is to find genes that contribute to disease**
- **Finding these genes should allow an understanding of the disease process, so that methods for preventing and treating the disease can be developed**
- **For “single-gene disorders”, current methods are usually sufficient**

Cystic Fibrosis Mutation



How SNP's Are Used to Find Genes Contributing to Disease

- To find the regions with the genes that contribute to a disease, the frequency of many SNP alleles are compared in individuals with and without the disease
- When a particular region has SNP alleles that are more frequent in individuals with the disease than in individuals without the disease, those SNP's and their alleles are associated with the disease
- The associations between a SNP and a disease indicate that there may be genes in that region that contribute to the disease

Types of SNPs

- **Genic, coding SNPs**
  - non-synonymous
    - Maintaining vs. altering protein structure/function
  - synonymous
    - Maintaining vs. altering splicing
- **Genic, non-coding SNPs**
  - Regulatory SNPs
    - Maintaining vs. altering gene expression
  - Intronic SNPs
    - Maintaining vs. altering gene expression/splicing
- **Linked SNPs**
  - usually intergenic

Examples of SNPs that cause disease

Primary hypomagnesemia

- patients unable to maintain sufficient levels of Mg^{2+} in their serum
- Affected gene: Na/K ATPase g subunit (123G->A = Gly->Arg in transmembrane region) – mutation *prevents targeting of the protein to cell membrane*

Mitochondrial SNPs

- >50 known disease-causing SNPs in mtDNA
- Most often affect tissues with high energy consumption

BRCA1 (breast cancer-associated antigen 1)

- Silent mutation in codong exon affects splicing (A. Krainer, CSHL)
- Exons contain *exonic spplicing enhancers* and *exonic splicing silencers* – mutations lead to exon skipping and aberrant inclusion, respectively

Examples of SNPs that cause altered non-disease phenotypes

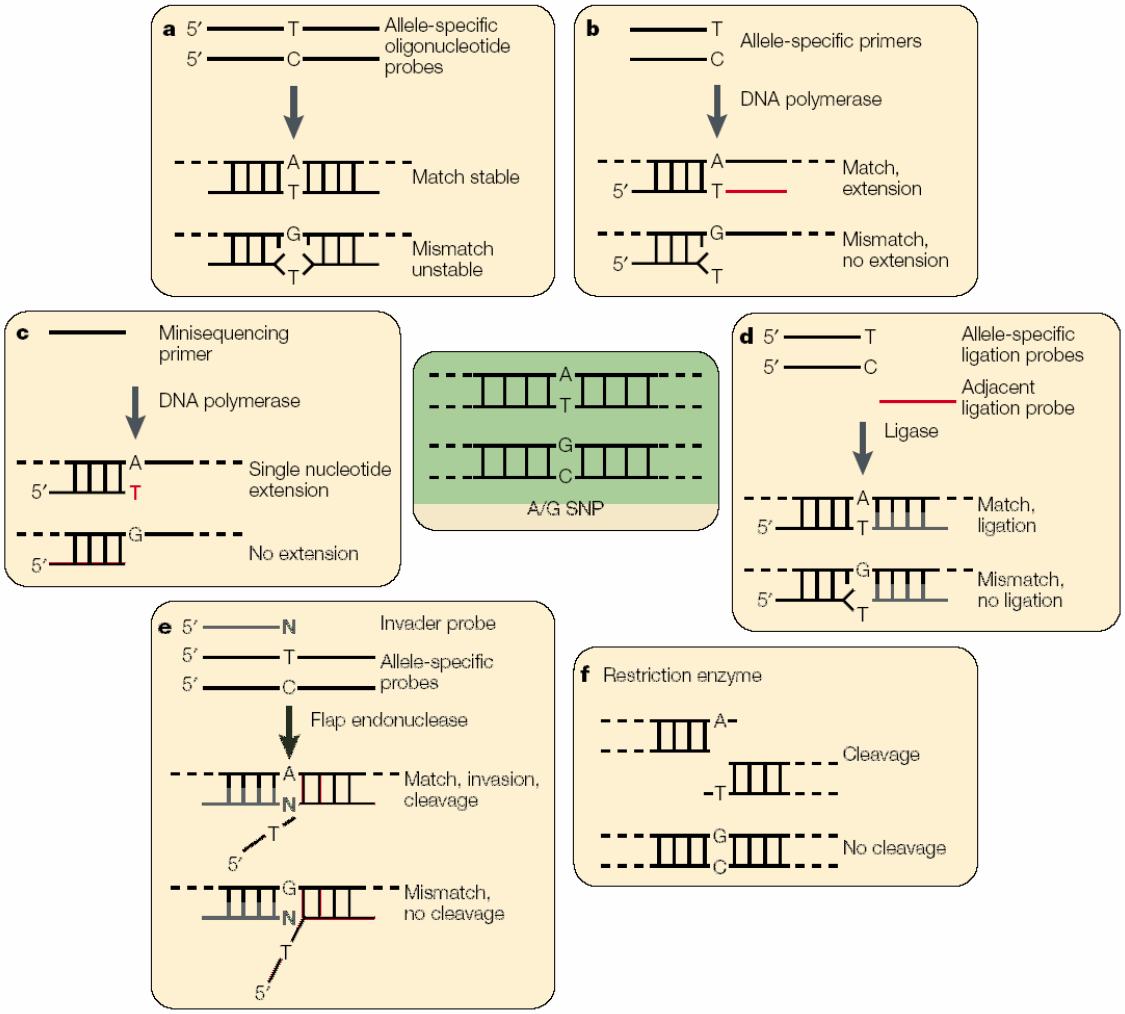
Glucose-6-phosphate dehydrogenase and favism

- First enzyme in the oxidative branch of pentose phosphate pathway, which reduces $NADP^+$ to NADPH
- Different mutations result in partially or completely inactive enzyme
- People (10%) with inactive enzyme experience lysis of red blood cells when consuming fava beans (contains H_2O_2 – NADPH is needed to detoxify it)

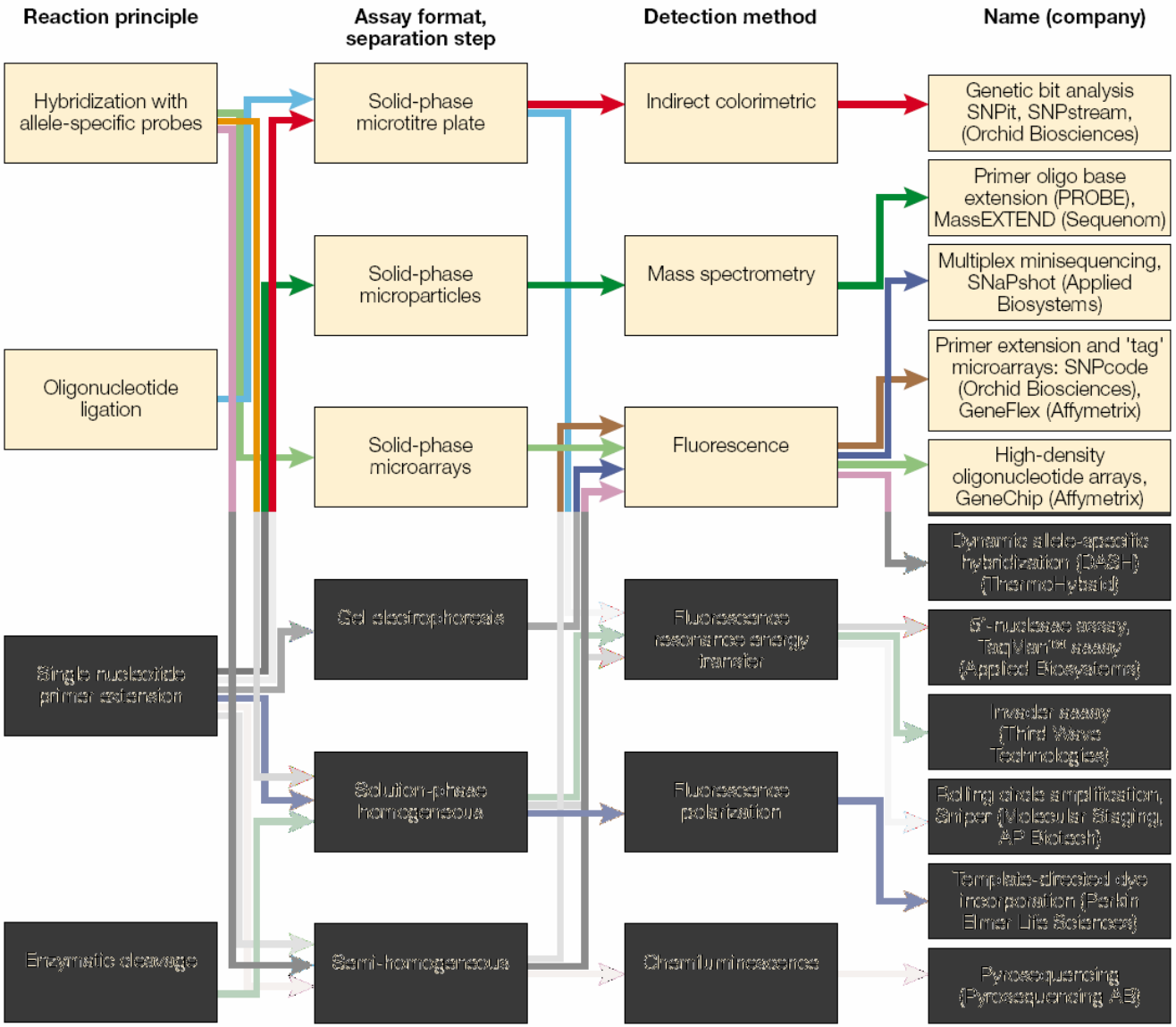
CytP450 mutations and drug responsiveness

- Cytochrome P450 – activates many pre-drugs into active therapeutic compounds
- Different people can be divided into *typical*, *poor*, and *ultra-rapid* metabolizers
- two genes in human:
  - *2D6* – required by more than 40 pre-drugs to for activation; 12 known SNPs altering the gene's activity
  - *2C19* – activates mephentyoin (epilepsy); 2-3% Caucasians and 23% Asians are poor metabolizers

Biochemical principles of various genotyping reactions

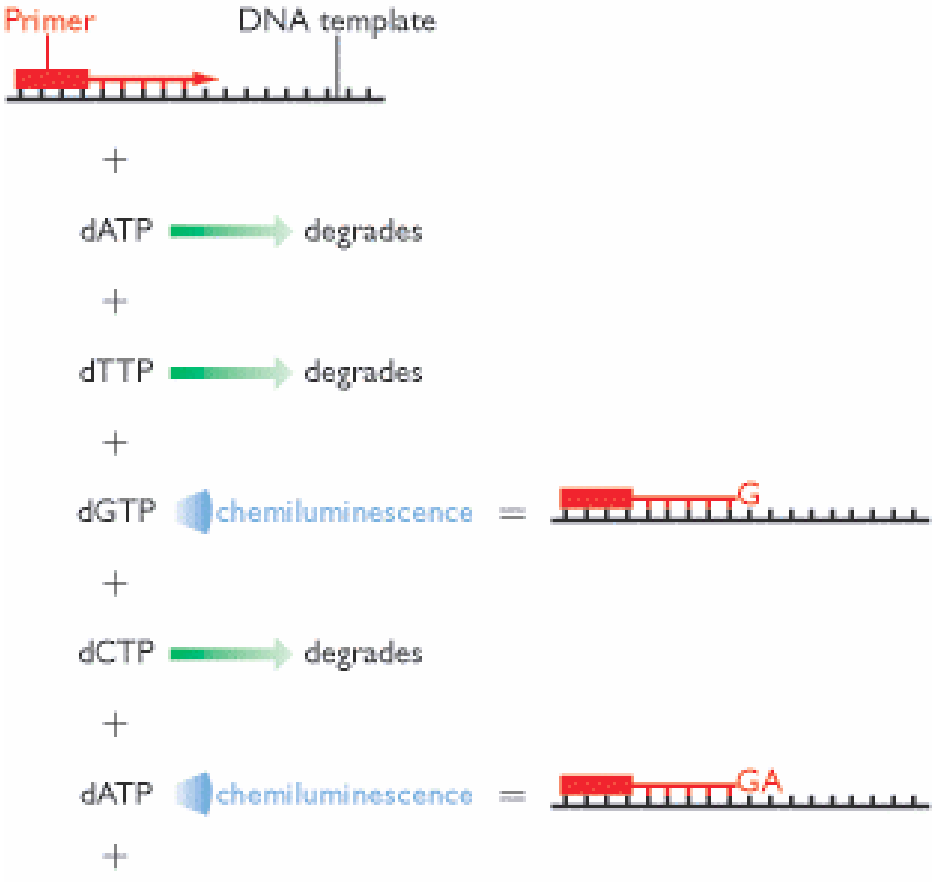


Assays for SNP genotyping

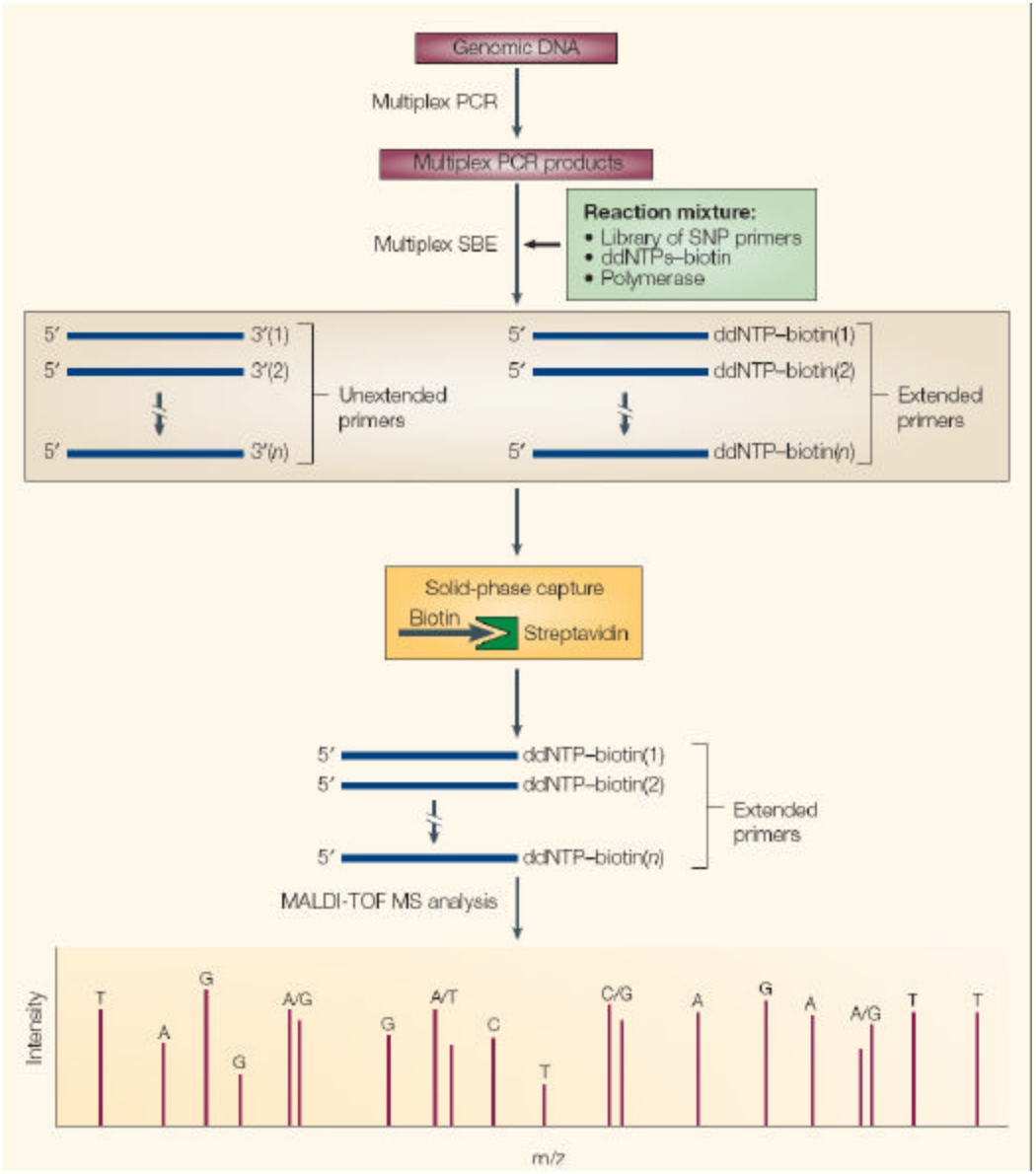


Example - pyrosequencing

- Four enzymes
  - DNA polymerase
  - ATP sulfurylase - converts pyrophosphate to ATP
  - Luciferase – “convert ATP to light”
  - Apyrase - degrades excess nucleotides
- Nucleotides added sequentially



Example – molecular affinity + mass spectrometry



Example – molecular affinity + mass spectrometry

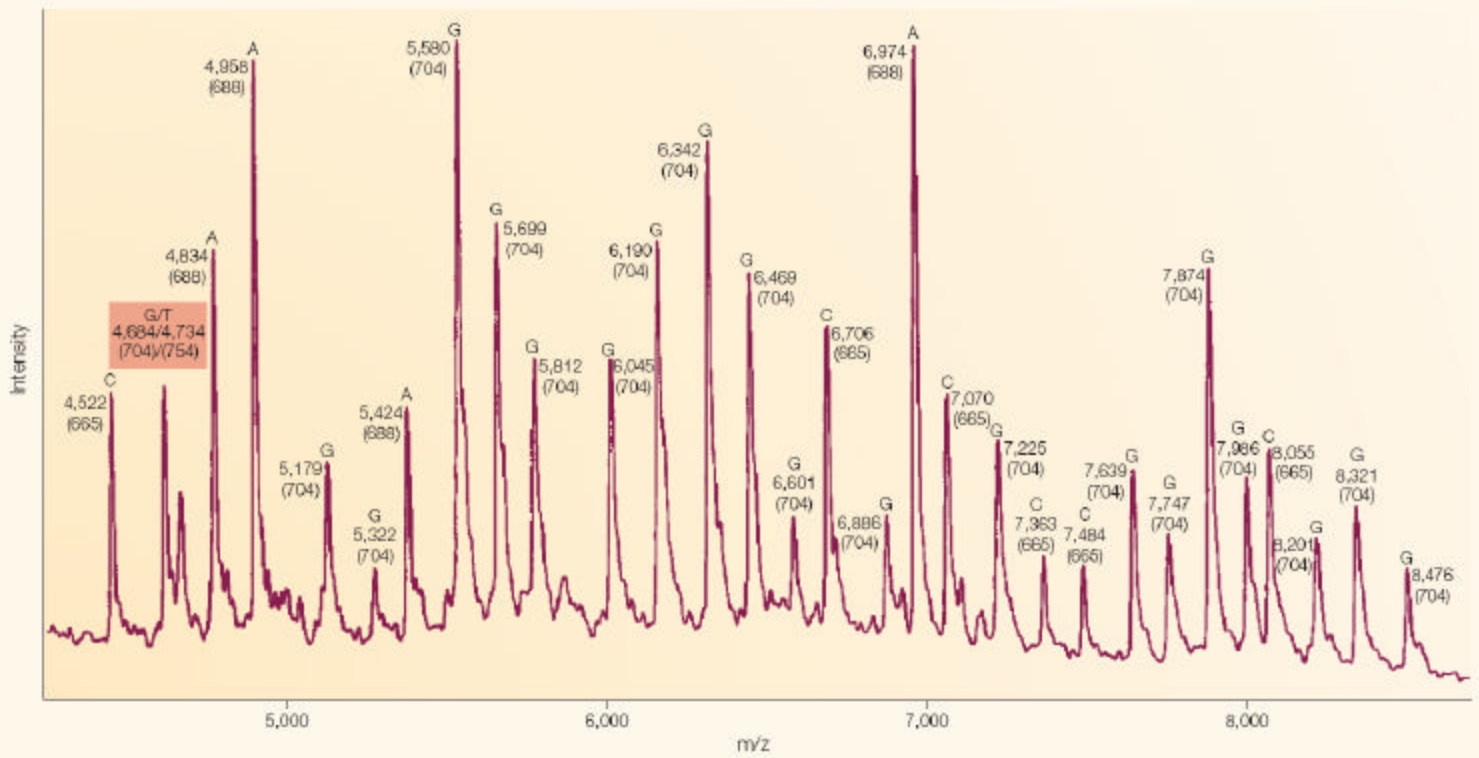
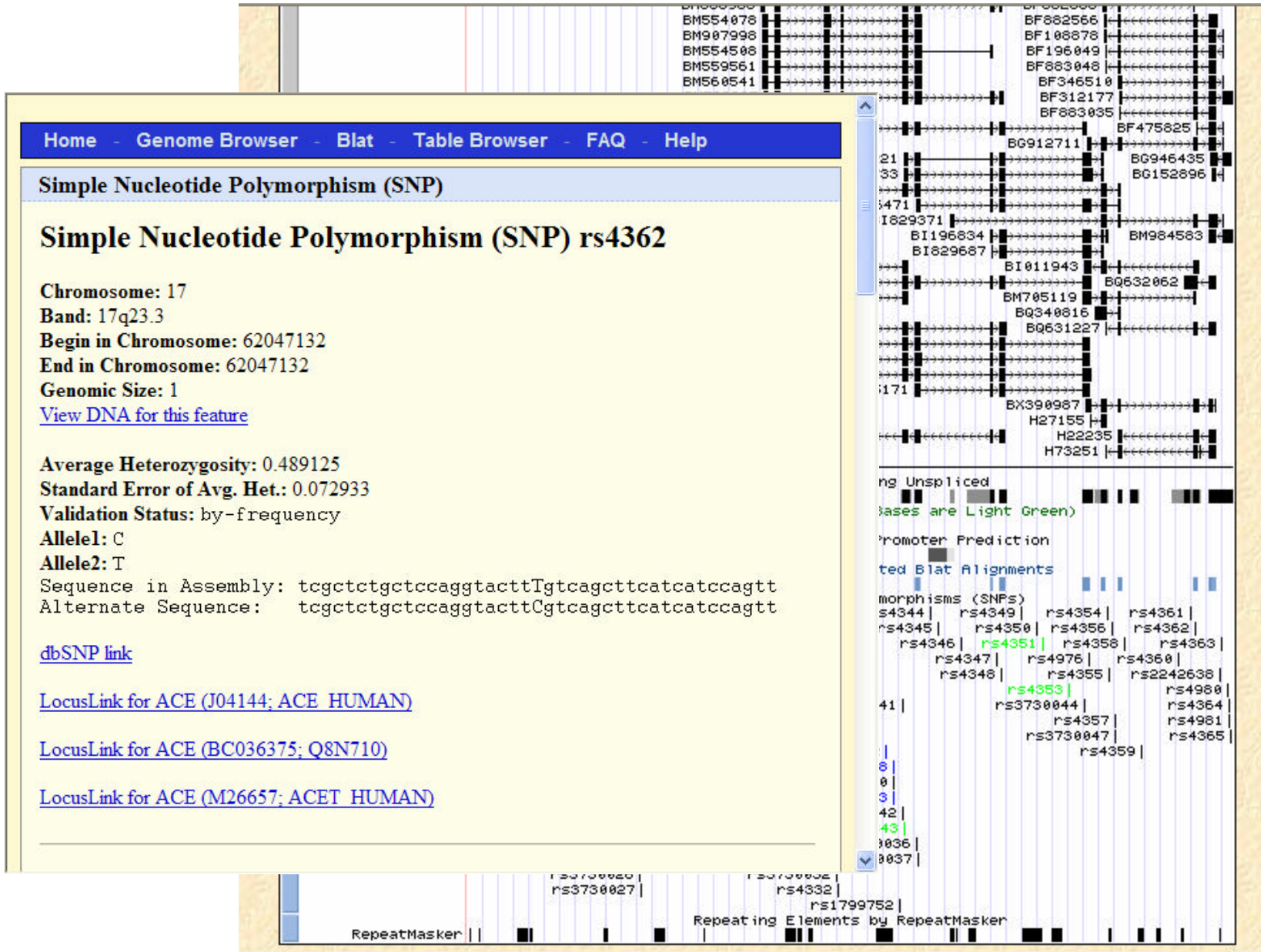
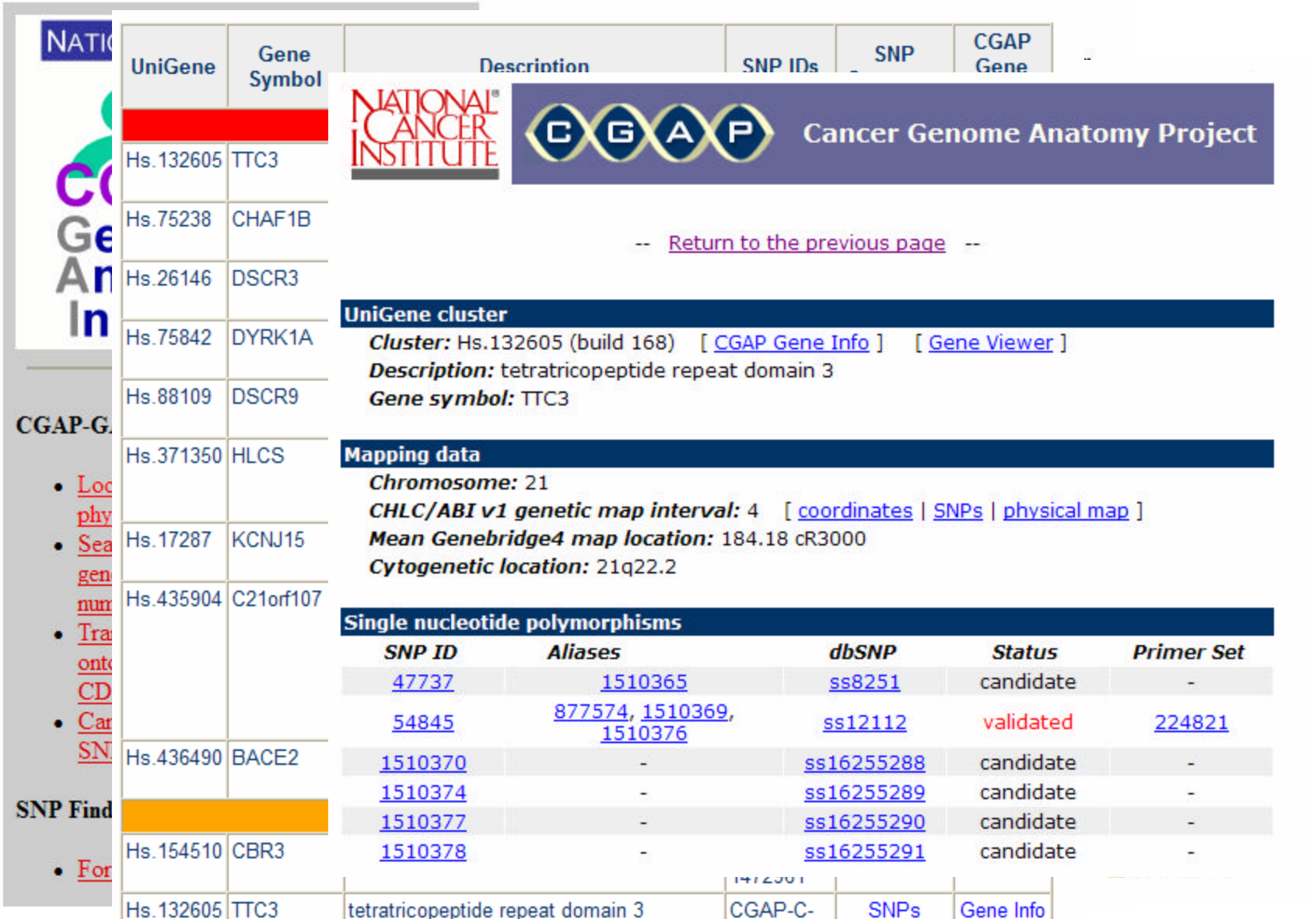


Figure 6 | **Simultaneous detection of nucleotide variations in 30 codons of the *p53* gene using SPC–SBE.** Shown here is a mass spectrum from a head and neck tumour, which contains a heterozygous genotype G/T (4,684/4,734 daltons) in codon 157. Each peak represents a different polymorphism, which is labelled with its nucleotide identity and absolute mass value. The values in parentheses, which denote the mass difference between each DNA extension product and its corresponding primer, are used to determine the nucleotide identity. m/z, mass per charge ratio.

Genome browsers have SNP tracks



CGAP has a gateway to SNP data and a tool for SNP finding in submitted sequences



UniGene	Gene Symbol	Description	SNP IDs	SNP	CGAP Gene
Hs.132605	TTC3				
Hs.75238	CHAF1B				
Hs.26146	DSCR3				
Hs.75842	DYRK1A				
Hs.88109	DSCR9				
Hs.371350	HLCS				
Hs.17287	KCNJ15				
Hs.435904	C21orf107				
Hs.436490	BACE2				
Hs.154510	CBR3				
Hs.132605	TTC3	tetratricopeptide repeat domain 3	CGAP-C-	SNPs	Gene Info

NATIONAL CANCER INSTITUTE

CGAP Cancer Genome Anatomy Project

-- Return to the previous page --

UniGene cluster

Cluster: Hs.132605 (build 168) [CGAP Gene Info] [Gene Viewer]
Description: tetratricopeptide repeat domain 3
Gene symbol: TTC3

Mapping data

Chromosome: 21
CHLC/ABI v1 genetic map interval: 4 [coordinates | SNPs | physical map]
Mean Genebridge4 map location: 184.18 cR3000
Cytogenetic location: 21q22.2

Single nucleotide polymorphisms

SNP ID	Aliases	dbSNP	Status	Primer Set
47737	1510365	ss8251	candidate	-
54845	877574, 1510369, 1510376	ss12112	validated	224821
1510370	-	ss16255288	candidate	-
1510374	-	ss16255289	candidate	-
1510377	-	ss16255290	candidate	-
1510378	-	ss16255291	candidate	-

SNP Consortium has a SNP portal... http://snp.cshl.org/

APBiotech - AstraZeneca - Aventis - Bayer - Bristol-Myers Squib - F.Hoffman-La Roche - Glaxo Wellcome

THE SNP CONSORTIUM LTD

IBM - Motorola - Novartis - Pfizer - Searle - SmithKline Beecham - Wellcome Trust

Home :: Frequency/Genotype :: Linkage Maps :: Protocols
Search :: News :: About :: Help :: Download data :: Feedback

Single Nucleotide Polymorphisms for Biomedical Research

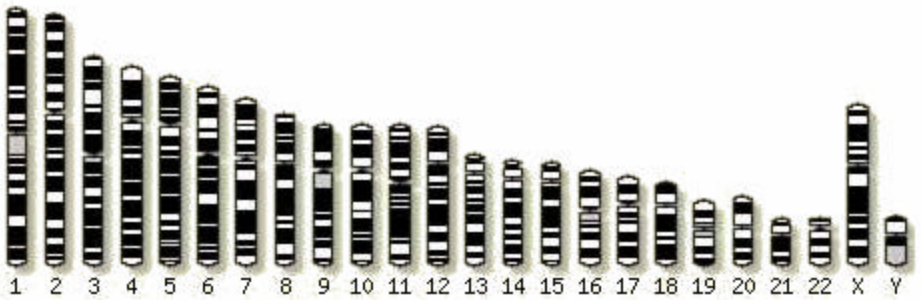


Image courtesy of http://www.ensembl.org, slightly modified from the original

Search: Search (advanced | help)

Single nucleotide polymorphisms (SNPs) are common DNA sequence variations among individuals. They promise to significantly advance our ability to understand and treat human disease. The SNP Consortium (TSC) is a public/private collaboration that has to date discovered and characterized nearly 1.8 million SNPs (more)

International HapMap Project

Home | About the Project | Data

[Edit profile] [Sign out]

Name:	rs4362
Class:	SNP
Type:	snp
	rs4362
Source:	HapMap_imsut-riken
Position:	Chr17:62047132..62047132
Reference strand:	(+) strand relative to the human genome
Alleles:	C/T
MAF:	T 0.49
Panel:	urn:lsid:dcc.hapmap.org:Panel:CEPH-30-trios:1

Assay:

Assay LSID:	urn:lsid:imsut-riken.hapmap.org:Assay:4362:1
pcr_primer_forward:	TGTGGGTGGGAGGCATCTAC
pcr_primer_reverse:	TTAAGACCCAAGGCTGGAGG
invader_probe:	CGGCCTCGCTCTGCTCCAGGTACTTA
allele_probe1:	TGTCAGCTTCATCATCCA
allele_probe2:	CGTCAGCTTCATCATCC
strand:	reverse relative to dbSNP (plus relative to human reference genome)
More info on assay parameters	in '00README.txt' file in the bulk download directory

Protocol:	urn:lsid:imsut-riken.hapmap.org:Protocol:genotyping:1
Platform:	Invader

Genotype Frequencies:

Genotype	Frequency	Number of Individuals
C/C	0.3	18
C/T	0.42	25
T/T	0.28	17
Total Individuals	60	

Genotypes:

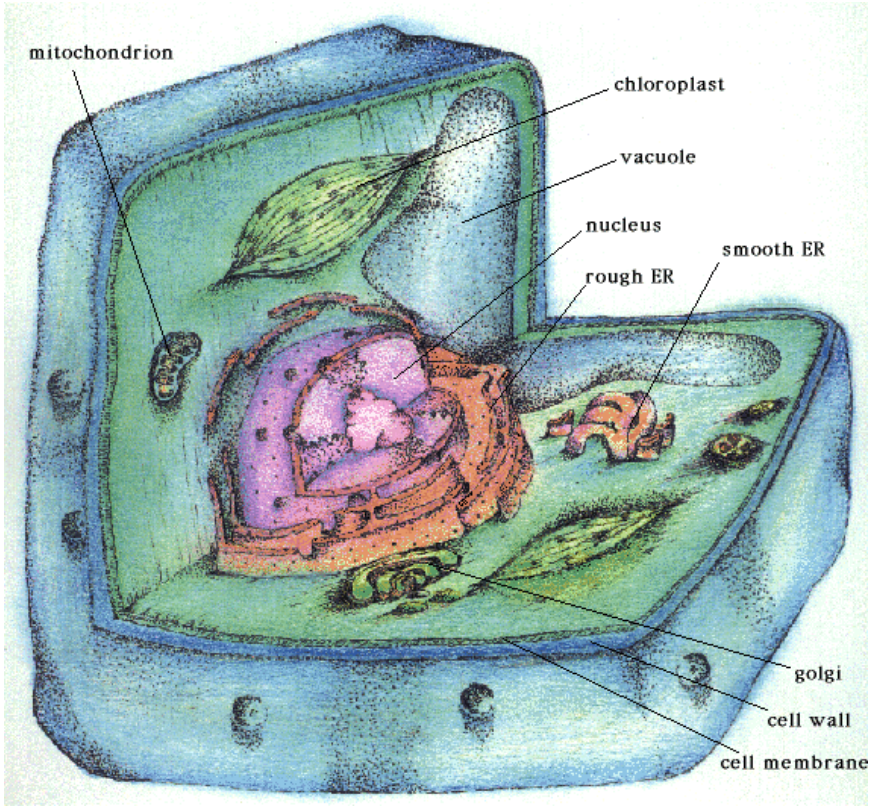
Individual ID	Genotypes	Individual ID	Genotypes	Individual ID	Genotypes	Individual ID	Genotypes
NA06985	TT	NA06991	TT	NA06993	CT	NA06993.DUP	CT
NA06994	CT	NA07000	CC	NA07019	TT	NA07022	TT
NA07029	CT	NA07034	CC	NA07048	CT	NA07055	CT
NA07056	CT	NA07345	CT	NA07348	CT	NA07357	CT
NA10830	CT	NA10831	TT	NA10835	TT	NA10838	TT
NA10839	CT	NA10846	CT	NA10847	CC	NA10851	TT
NA10854	CT	NA10855	CC	NA10856	CC	NA10857	TT
NA10859	CC	NA10860	CT	NA10861	CT	NA10863	TT

IDENTIFYING GENES INVOLVED IN DISEASES
The Mitochondrial Connection

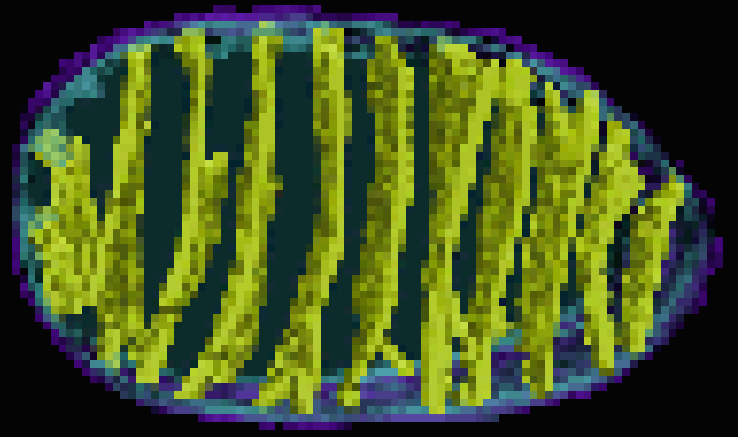
Guda C, Fahy E, Subramaniam S. (2004) *Bioinformatics*, 20:1785-1794

Guda C, Guda P, Fahy E, Subramaniam S. (2004) *Nucleic Acids Res*, 32:W372-W374

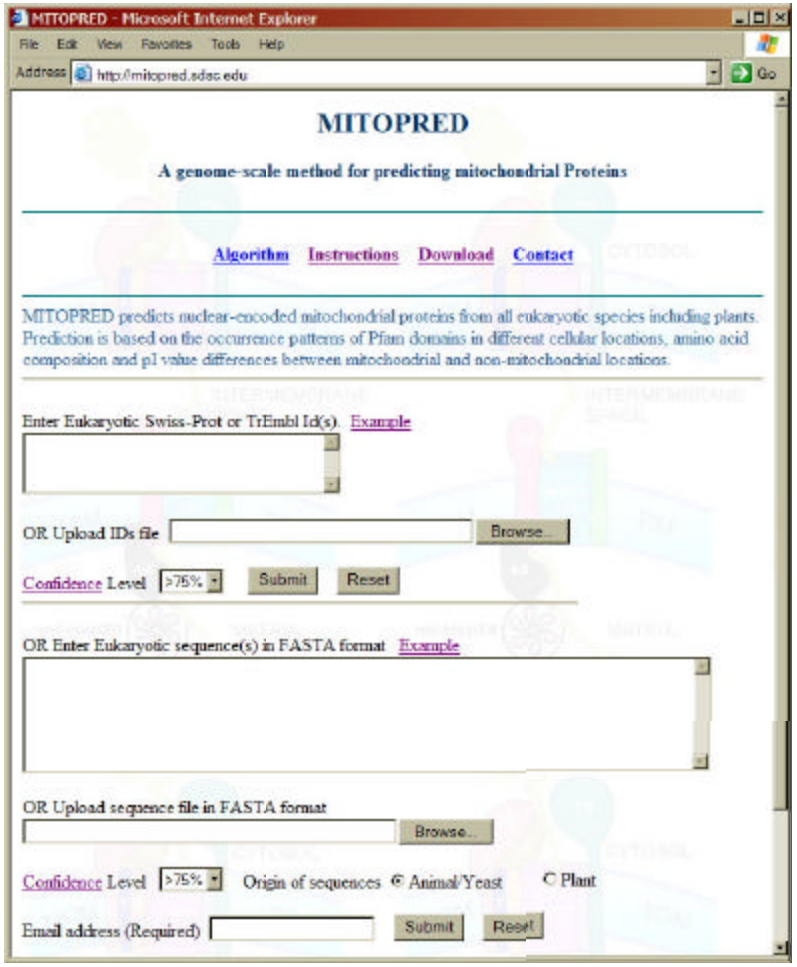
Guda P, Guda C, Subramaniam S. 2004

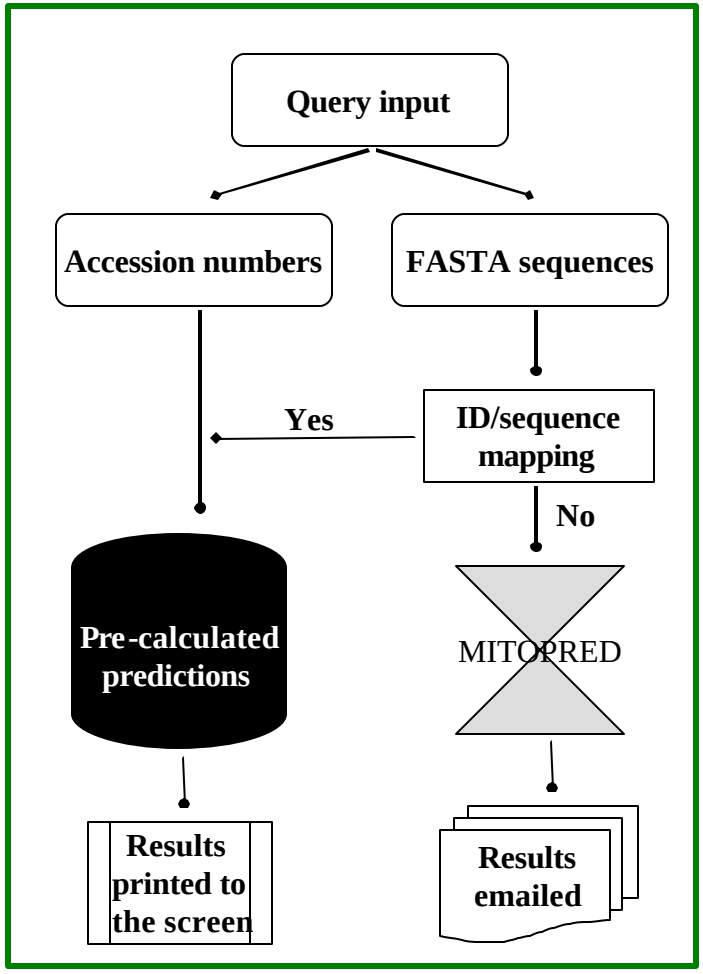


Brown Fat Mitochondrion



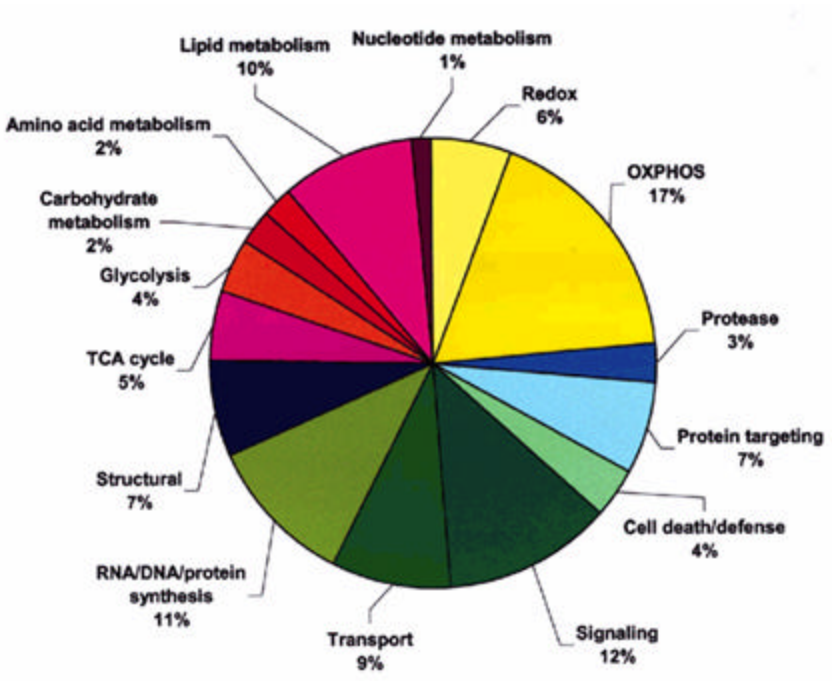
MITOPRED: webserver for genome-scale prediction of mitochondrial proteins





mitochondria proteome

- Genome size: 16,571 bp
- 657 distinct proteins
- 498 (81%) functionally classified into 15 cellular processes.
- 153 unique enzymatic activities



Taylor S. *et al*, Nature (21), 2003

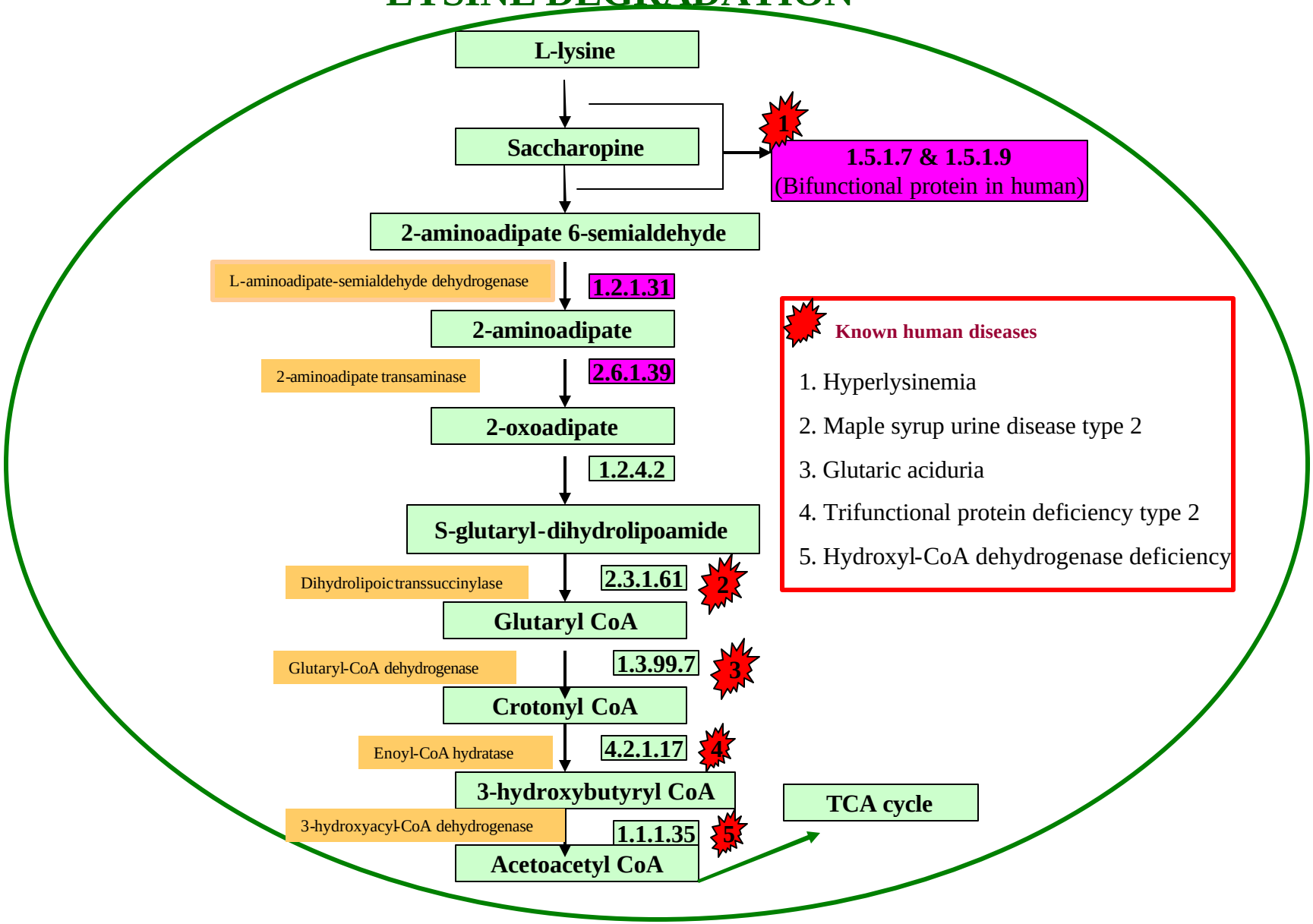
Identification of mitochondria-associated pathways

- About 42 KEGG metabolic pathways are associated with at least one known mitochondrial protein
- Of these, 8 are fully mitochondrial that have been well characterized.
- About 35 pathways span across multiple sub-cellular locations including mitochondria
- The 35 pathways fall under 6 metabolic cores

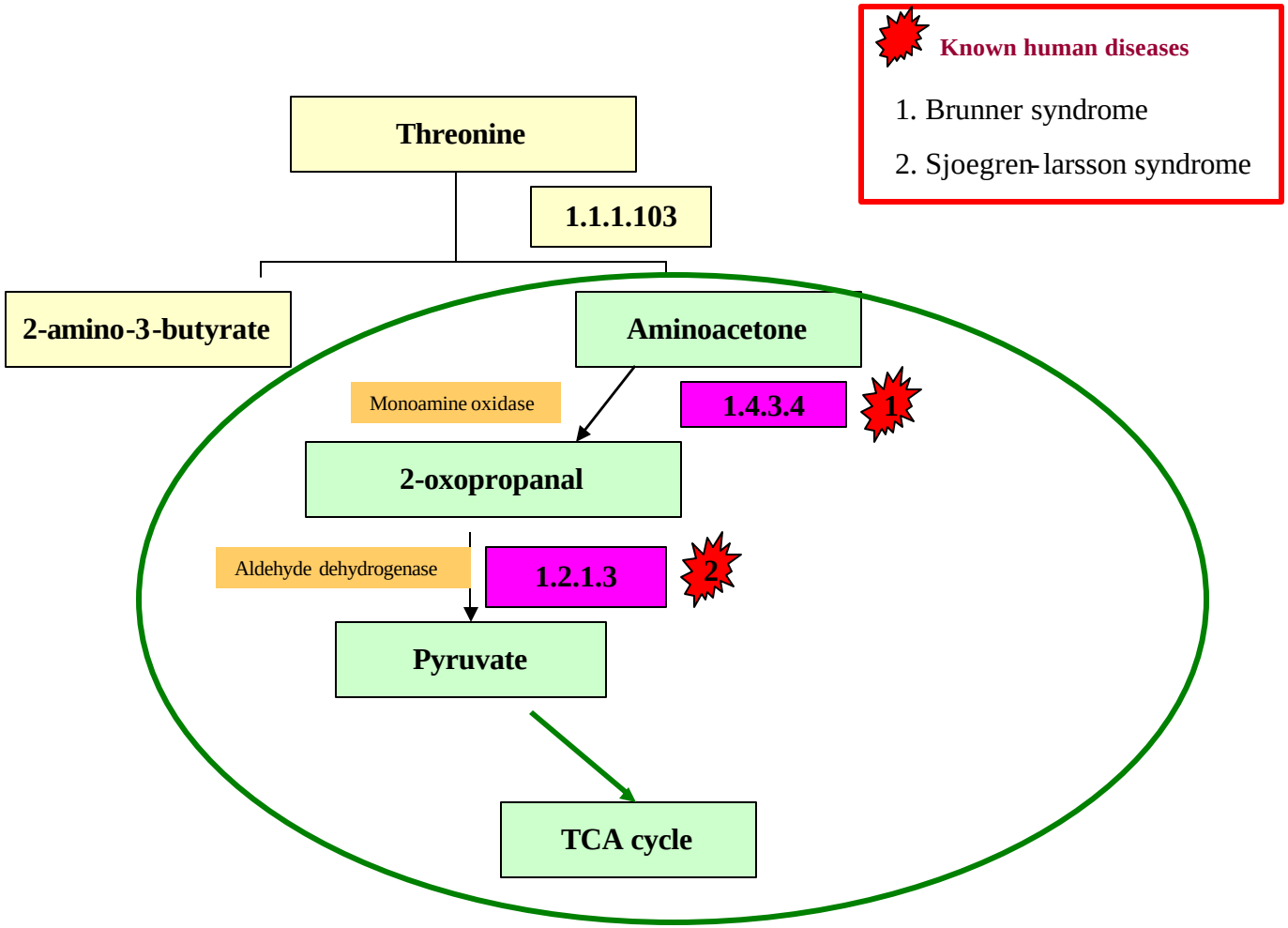
Major metabolic cores

1. Energy metabolism
2. Nucleotide metabolism
3. Amino acid metabolism
4. Carbohydrate metabolism
5. Lipid metabolism
6. Metabolism of Vitamins and co-factors

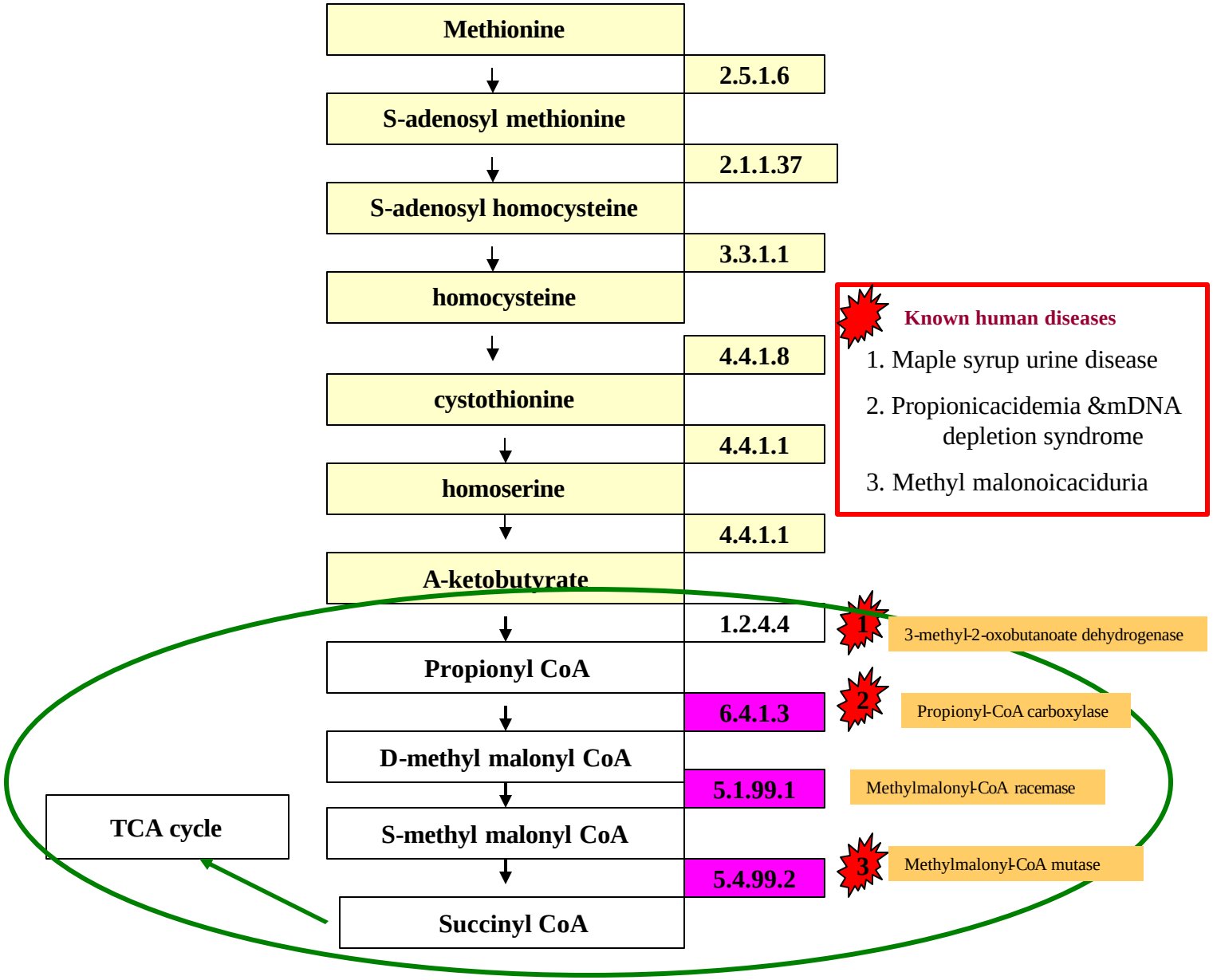
LYSINE DEGRADATION



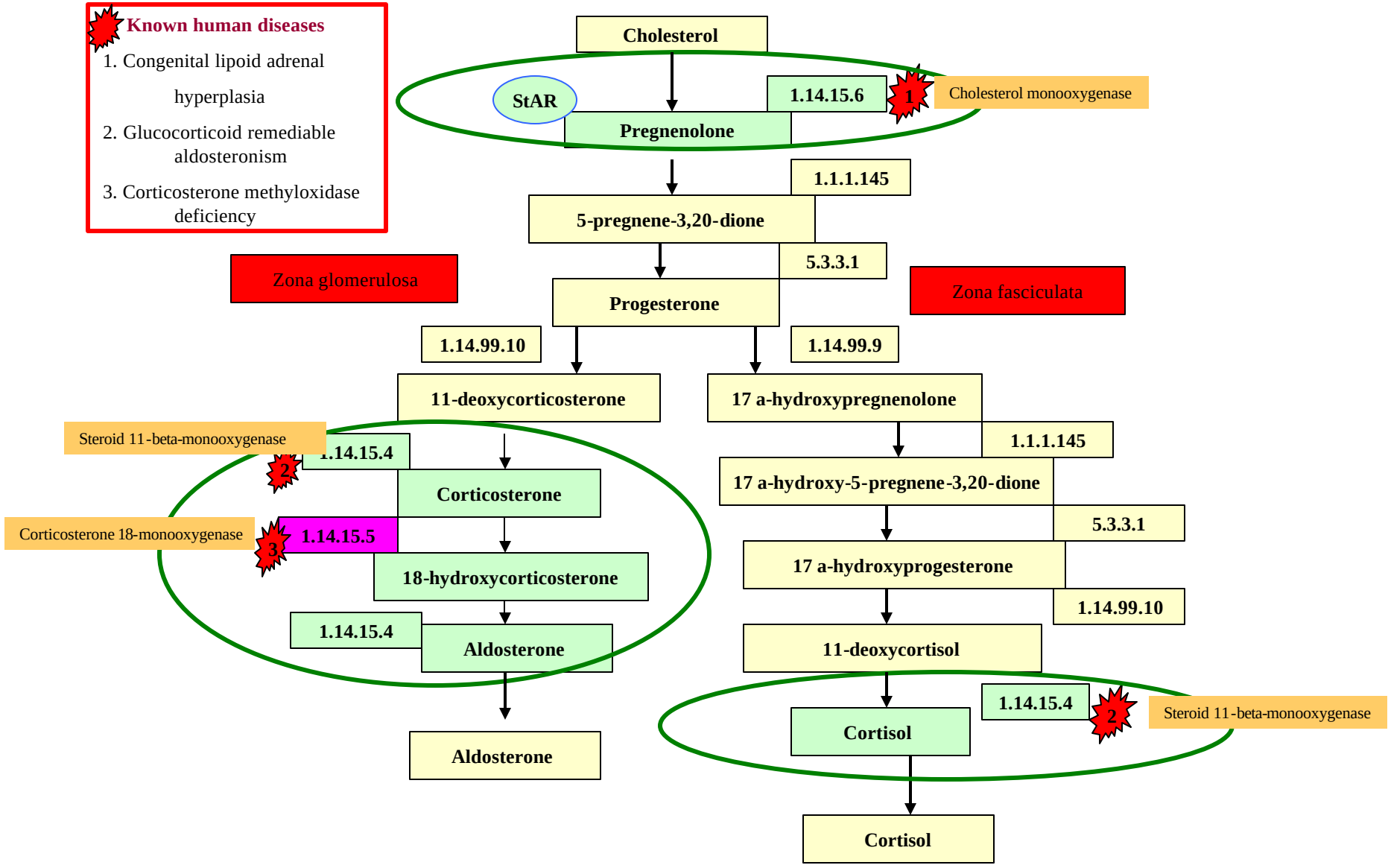
THREONINE DEGRADATION



METHONINE DEGRADATION

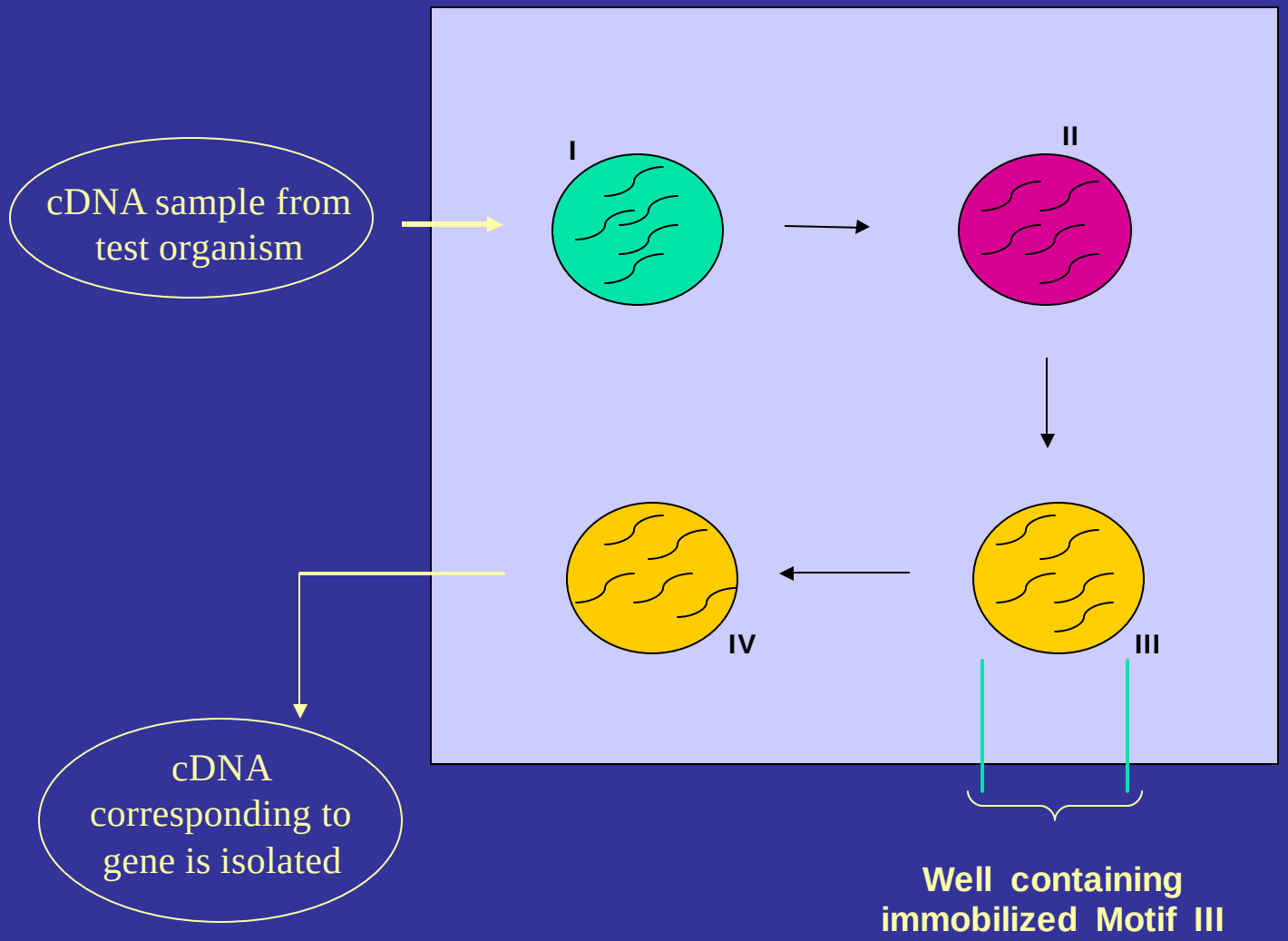


C-21 STEROID HORMONE METABOLISM



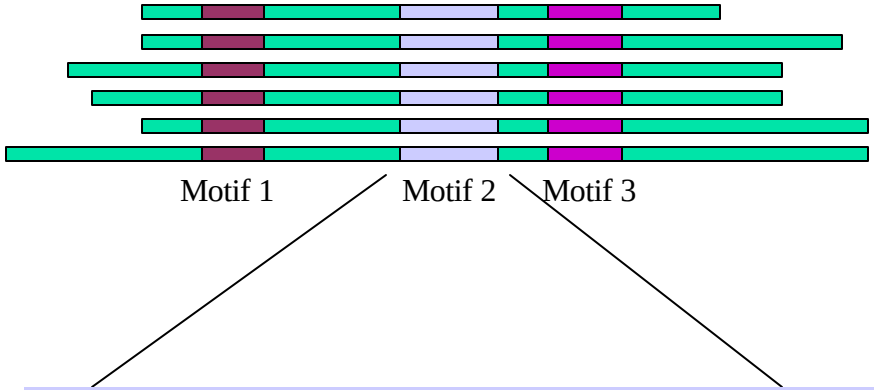
Motif Oligonucleotides Corresponding to the Homologs of a Gene Immobilized on a Microarray

(Assuming 4 sequence motifs in this gene)



Construction of a Motif Oligonucleotide

A hypothetical illustration



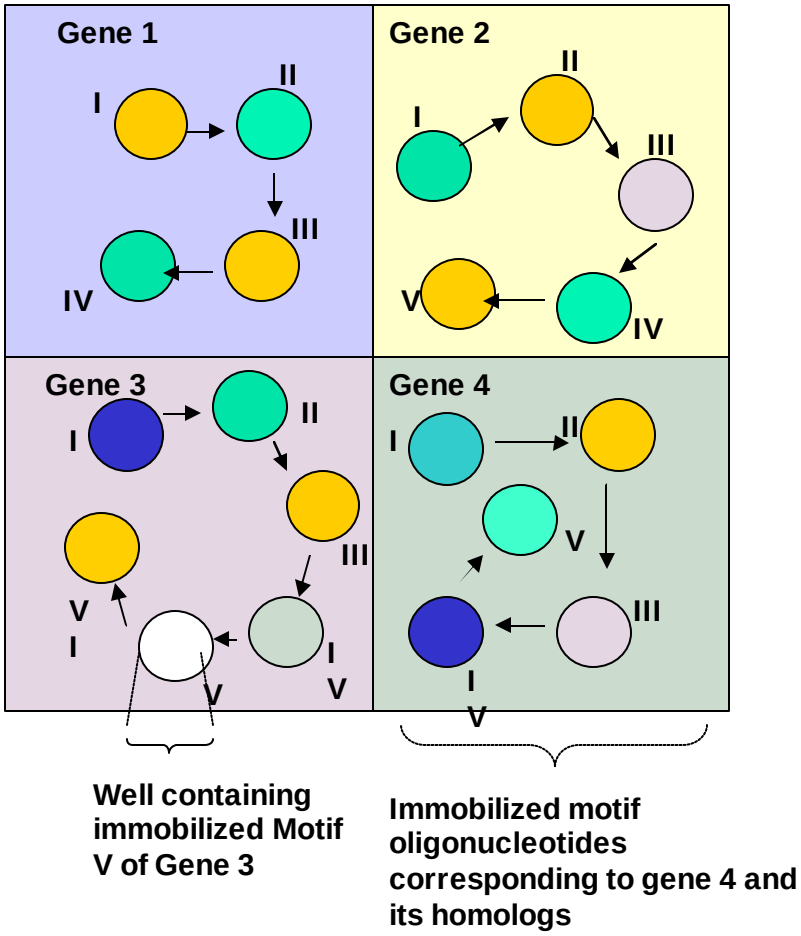
	Weight
E.coli	0.17
B.subtilis	1
M.genitalium	0.92
H.pylori	0.22
A.fulgidus	0.82
T.maritima	0.91

Organism	Motif 2 Sequences	Weight
E.coli	AGGTCCAGGCTAAGTCCGTAGACTGATCGTTAGCTT	0.17
B.subtilis	TGCCTACGGCTAAGTCGGTTCATCCGTCGTTAGCTT	1
M.genitalium	CCGTCCAGGCTAAGTCCGTAGACTGATCAACCGCTT	0.92
H.pylori	AGGTCCAGGCTAAGTCCGTAGAACCTTCGTTAGCTT	0.22
A.fulgidus	AGGTCCAGGGCCGATCCGTAGACTGATCGTTAGCTT	0.82
T.maritima	AGGTCCAGGCTAAGTCCGTATGGTACTCGTTAGGAC	0.91

Weighted consensus: **AGGTCCAGGCTAAGTCCGTAGACTGATCGTTAGCTT**

Experimental Setting

Using the Electronic Semiconductor Microchip from Nanogen™ (www.nanogen.com/tech.htm) to detect the expression of a pathway in an organism



Galactose Metabolism

Vibrio Cholerae

D-Galactose + ATP

↓ **Galactokinase (2.7.1.6)**

a-D-Galactose-1-phosphate + ADP + H^+ + UDP-Glucose

↓ **Galactose-1P uridylyl-transferase (2.7.7.10)**

UDP-Galactose + Glucose 1-phosphate

↓ **UDP-glucose 4-epimerase (5.1.3.3)**

UDP-Glucose

Ramaratnam & Subramaniam. Microb Comp Genomics. 2000;5(3):153-61

And most common diseases are caused by a combination of genes and environment

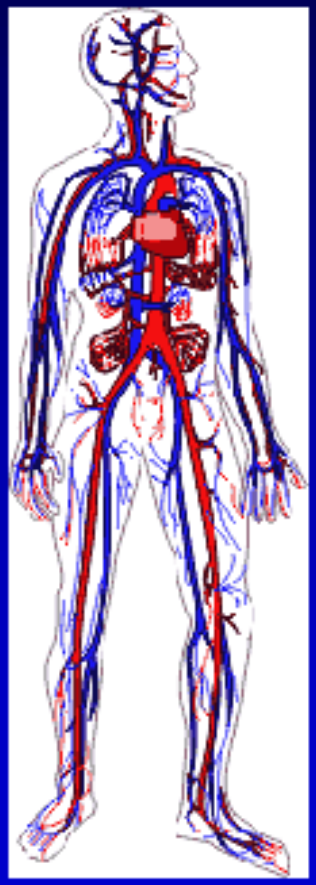
Stroke

Breast cancer

Diabetes

Obesity

Inflammatory Bowel Disease



Manic-depression

Myocardial Infarction

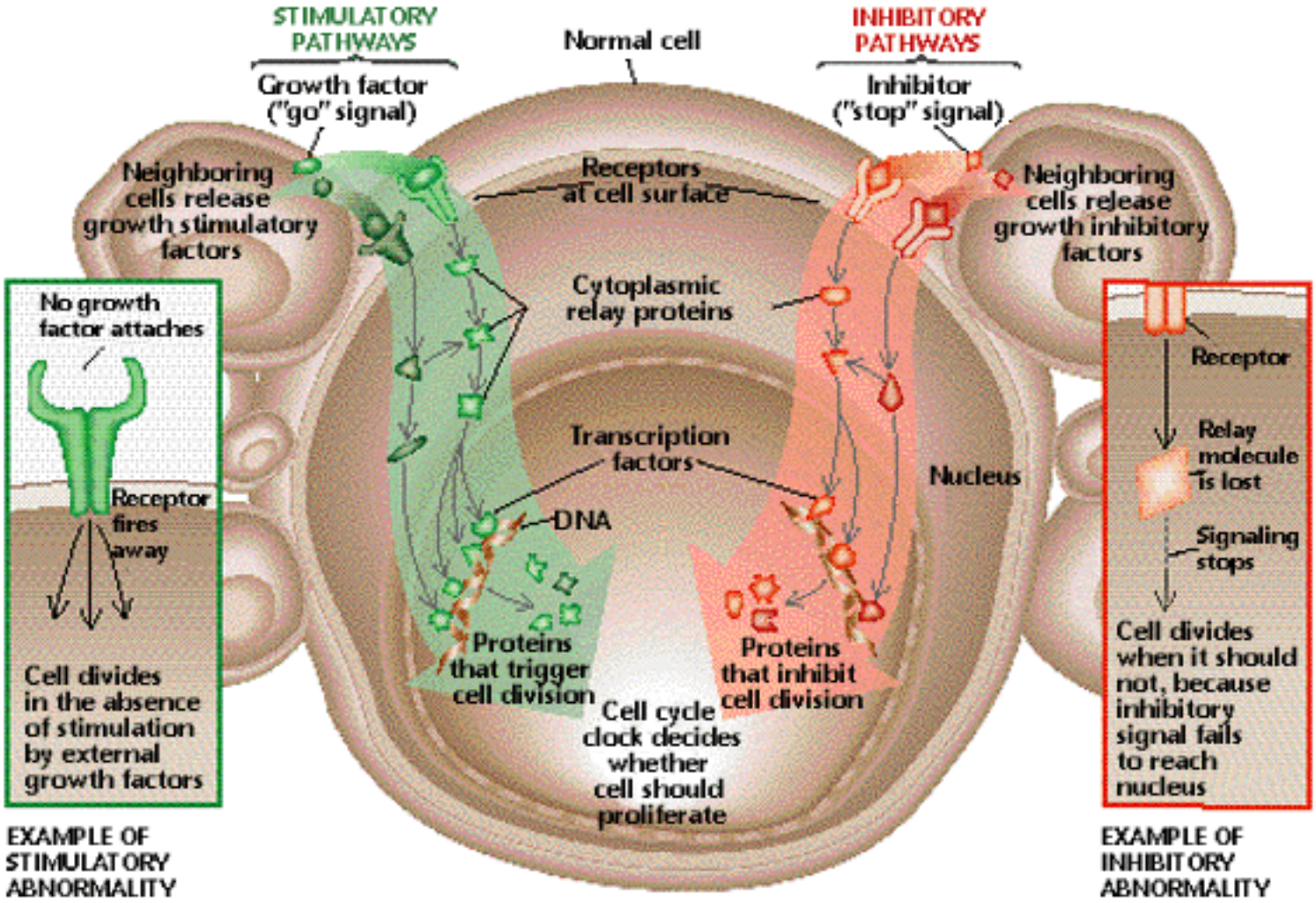
Hypertension

Hyperlipidemia

Schizophrenia

PATHWAYS AND PHENOTYPES

- **CELLULAR NETWORKS AND THE DISEASE CONNECTION**


STIMULATORY PATHWAYS
Growth factor ("go" signal)
Normal cell
INHIBITORY PATHWAYS
Inhibitor ("stop" signal)
Neighboring cells release growth stimulatory factors
Receptors at cell surface
Neighboring cells release growth inhibitory factors
Cytoplasmic relay proteins
No growth factor attaches
Receptor fires away
Transcription factors
Nucleus
DNA
Receptor
Relay molecule is lost
Signaling stops
Proteins that trigger cell division
Proteins that inhibit cell division
Cell cycle clock decides whether cell should proliferate
Cell divides in the absence of stimulation by external growth factors
Cell divides when it should not, because inhibitory signal fails to reach nucleus
EXAMPLE OF STIMULATORY ABNORMALITY
EXAMPLE OF INHIBITORY ABNORMALITY

DISEASE MODELS

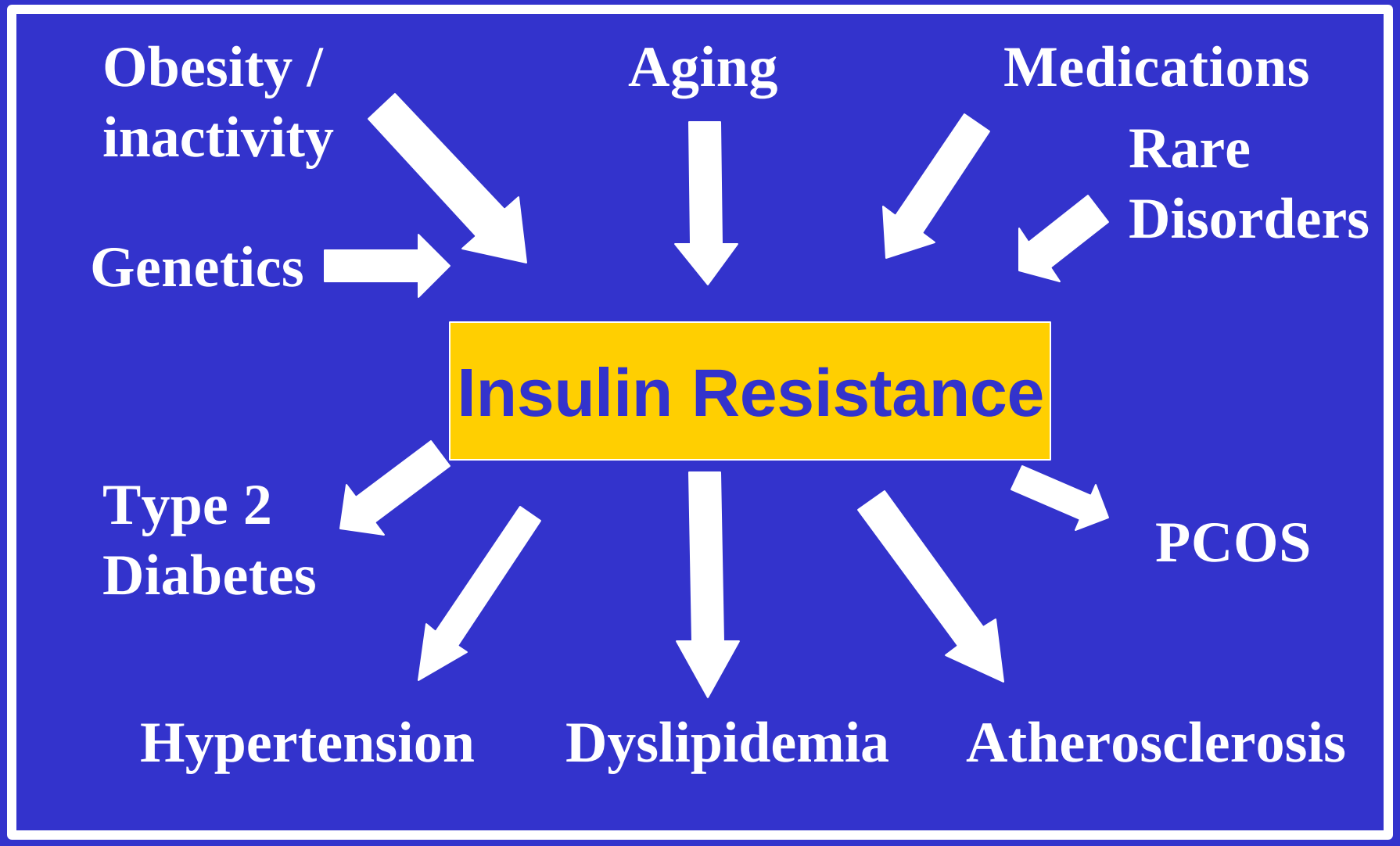
The Systems Biology Perspective

Albert Hsiao, Jerrold Olefsky, Shankar Subramaniam

DIABETES MELLITUS

1. 6-7% OF US POPULATION = 16 MILLION PEOPLE
2. 800,000 NEW CASES/YEAR
3. LEADING CAUSE OF ADULT BLINDNESS
4. LEADING CAUSE OF ADULT RENAL FAILURE
5. LEADING CAUSE OF AMPUTATIONS
6. 2-4X INCREASE CVD INCIDENCE

Insulin Resistance:
Causes and Associated Conditions



Novel Variance-Modeling Approach for Gene Expression Analysis

- Identify the gene expression responses that provide the insulin-sensitizing effect
- Fat, liver, skeletal muscle expression analysis
  - TZD treatment
  - Insulin treatment
  - Normal and diseased tissue

Hsiao, Worrell, Olefsky & Subramaniam, *Bioinformatics* 2004

PPRE and Direct Targets of PPAR-g

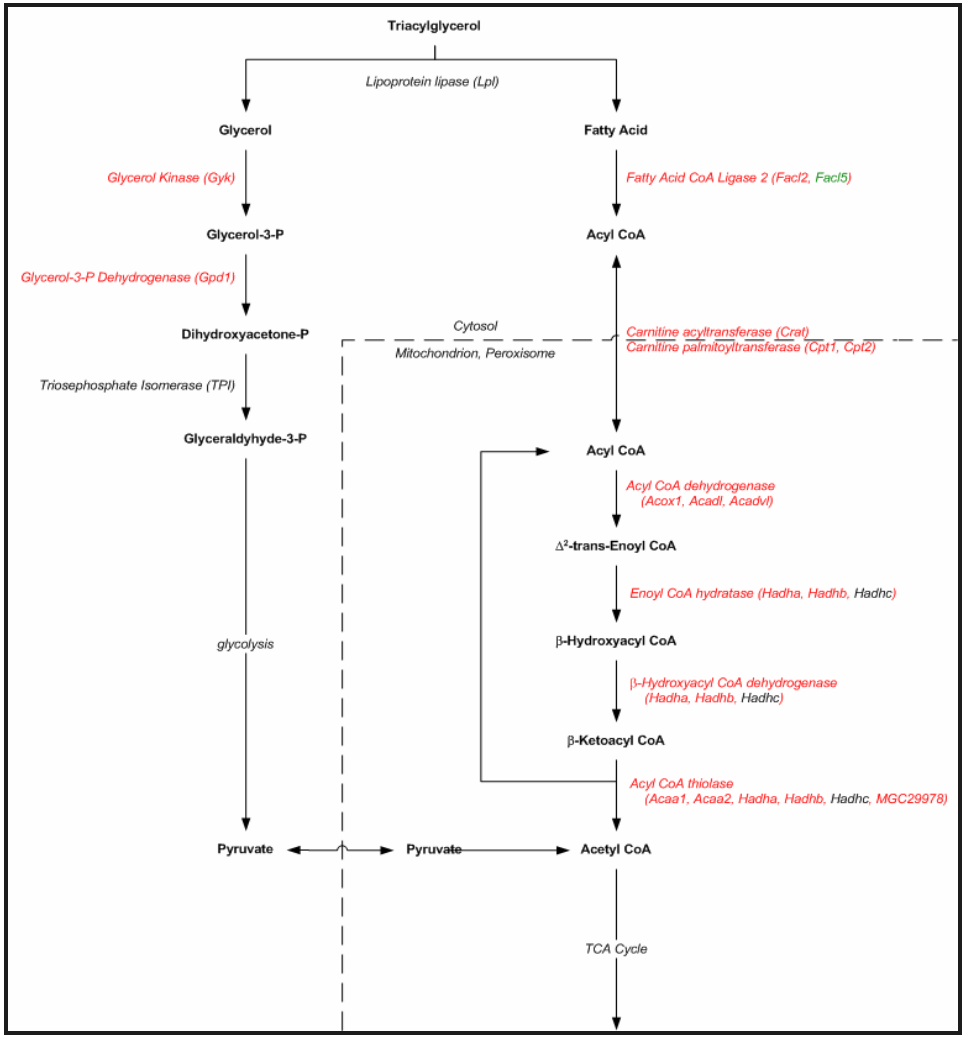
- Many known PPRE experimentally validated
  - EMSA
  - Transient transfection with reporter plasmid
- Most known PPRE are mediators of metabolic state
  - Fatty acid metabolism
  - Cholesterol metabolism
  - Fatty acid, cholesterol transport in blood

Table 1. Locations of functional PPREs from literature.

Gene	Description	Species
Acox1	acyl-Coenzyme A oxidase 1, palmitoyl	rat
Apoc3	apolipoprotein C-III	human
Apoe	apolipoprotein E	human
Aqp7	aquaporin 7	mouse
Cat	catalase	rat
Cd36, FAT	cd36 antigen; fatty acid translocase	mouse
Ces1	carboxylesterase 1, cholesterol ester hydrolase	human
Cpt1a	carnitine palmitoyltransferase 1 a, liver	?
Cpt2	carnitine palmitoyltransferase 2	human
Cyp1 a1	cytochrome p450, family 1, subfamily a, polypeptide 1	rat
Dbi, Acbp	diazepam-binding inhibitor; acyl-CoA binding protein	human, mouse, rat
Fabp1, L-FABP	fatty acid binding protein, liver	
Fabp4, aP2	fatty acid binding protein 4, adipocyte	mouse
Facl2	fatty acid Coenzyme A ligase, long chain 2; acyl CoA synthetase, long chain	rat
Glut2, Slc2a2	solute carrier family 2, member 2	rat
Hmgcs2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2	mouse, rat
Lrp1	low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor)	human
Lpl	lipoprotein lipase	human, rat
Modl	malic enzyme, supernatant	rat
Nr1h3, LXR-a	nuclear receptor subfamily 1, group H, member 3; liver X receptor, alpha	mouse, human
Pck1	phosphoenolpyruvate carboxykinase, cytosolic	human, rat
Ptgs2, COX2	prostaglandin-endoperoxide synthase 2; cyclooxygenase 2	human
Slc27a1, Fatp	solute carrier family 27 (fatty acid transporter), member 1	mouse
Sorbs1, CAP	sorbin and SH3 domain containing 1	mouse
Ucp1	uncoupling protein 1, mitochondrial	mouse
Ucp3	uncoupling protein 3	human

TZD on 3T3-L1 Gene Expression

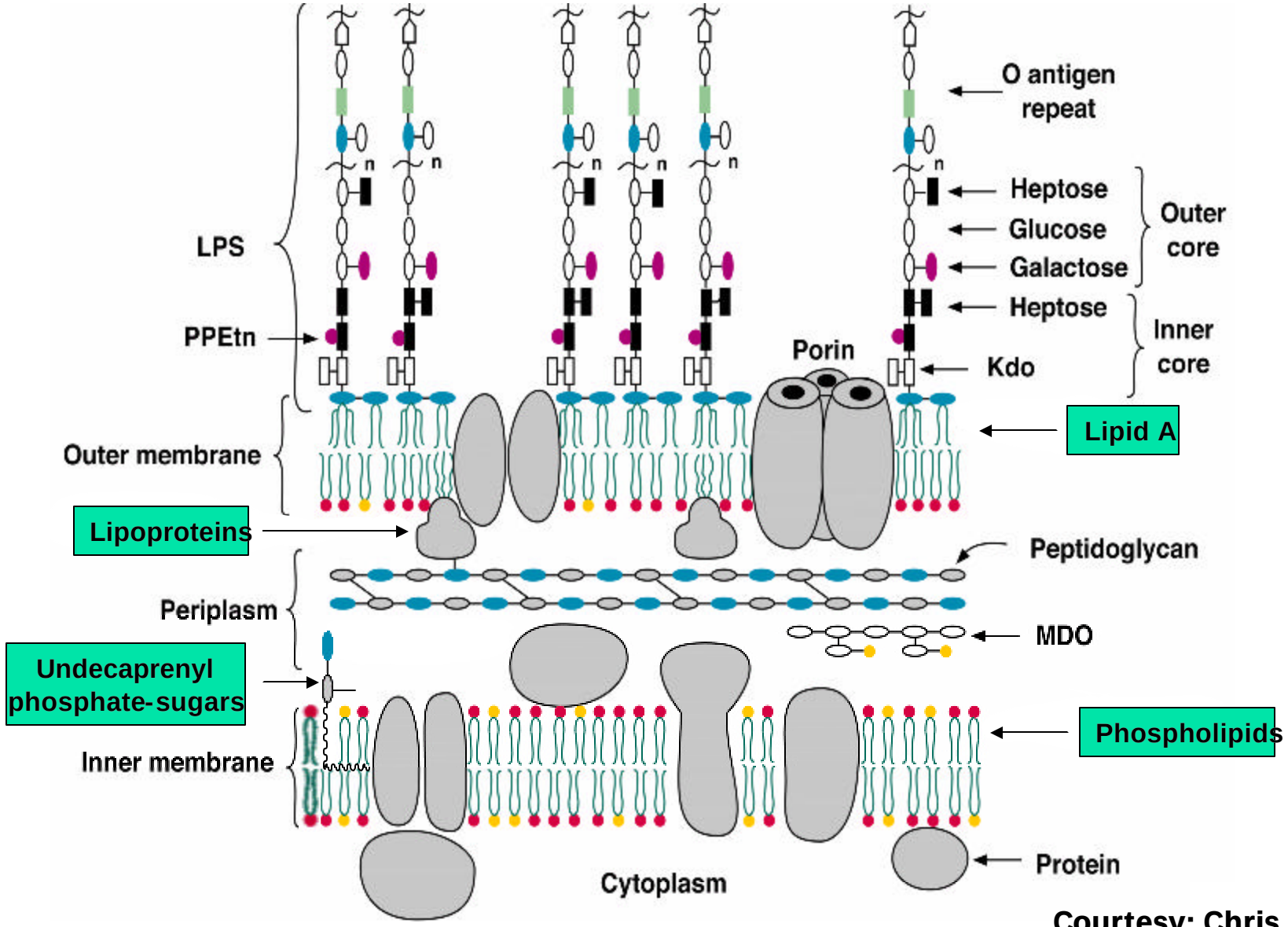
- Rosiglitazone upregulates transcription of almost every enzyme involved in fatty acid oxidation
- Increased fatty acid oxidation may improve lipid profile, as mechanism of improving insulin-resistance
- Results suggest that transcriptional control is important for maintaining cellular metabolic state



DISEASE MODELS
- Infectious Diseases
The Systems Biology Perspective

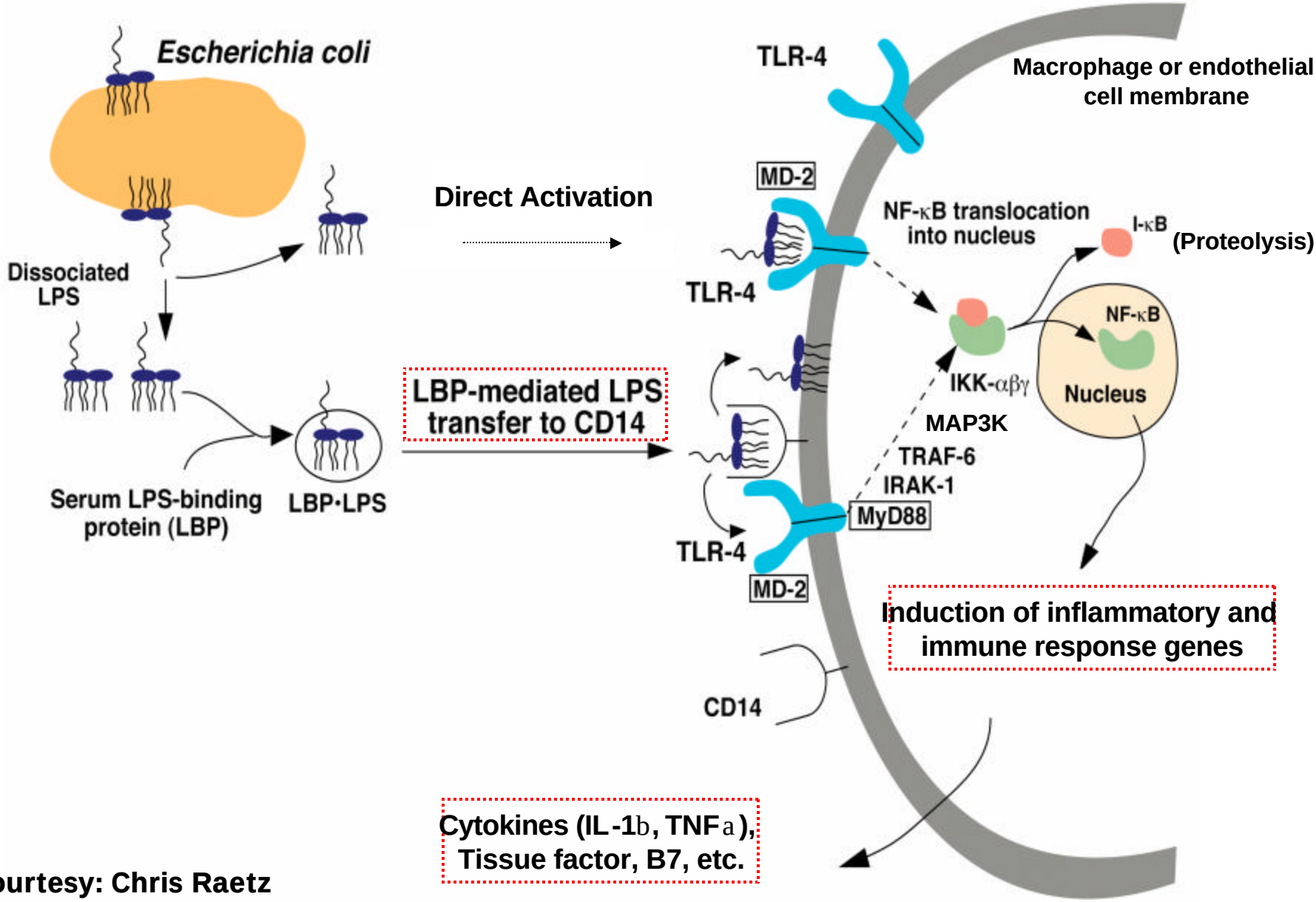
Christopher Benner, Christopher Glass and
Shankar Subramaniam

The *E. coli* or *Salmonella* Cell Envelope



Courtesy: Chris Raetz

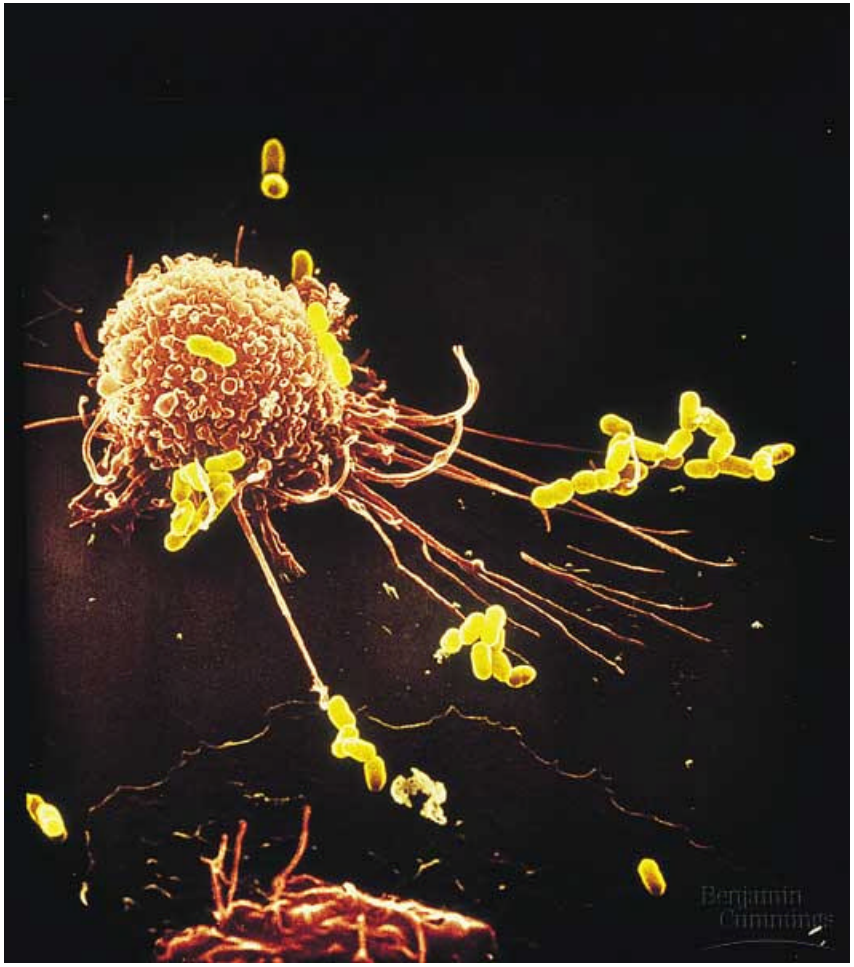
The Lipid A (Endotoxin) Receptor: TLR-4



Courtesy: Chris Raetz

Macrophage Cell Types

- **RAW – Macrophage cell line**
- **TM - Thioglycolate elicited macrophages**
- **BM – Bone Marrow derived macrophages**
- **ES – Embryonic Stem Cells (not macrophage)**

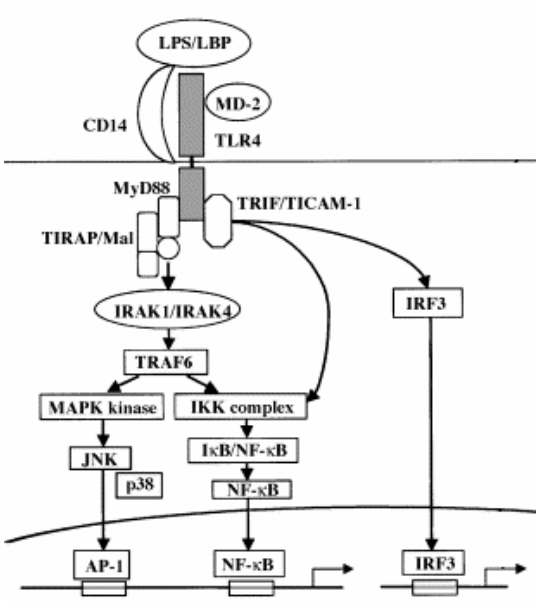


Transcriptome and Network Analysis

Christopher Benner, Christopher Glass and Shankar Subramaniam

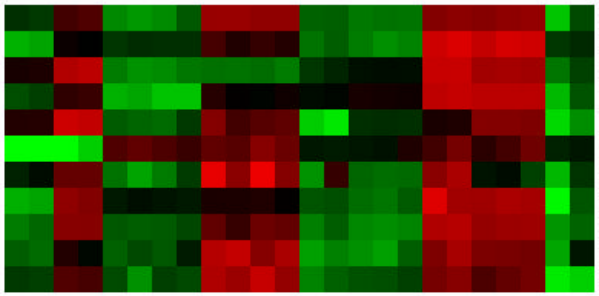
Data from LIPIDMAPS and AfCS Laboratories

Response to LPS



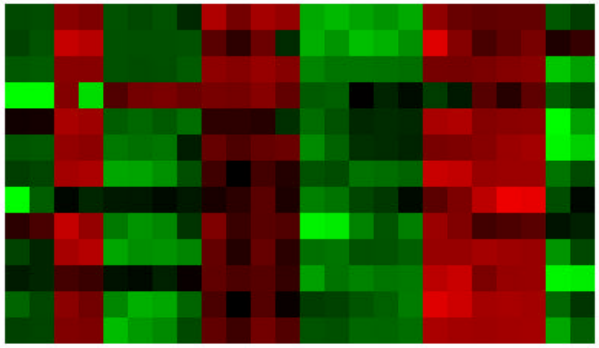
Log_2 Fold Change

Myd88 Dependent Targets



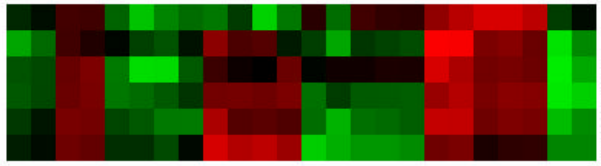
Name	BM	RAW	TM
Il1b	5.19	11.18	10.33
Ptgs2	4.41	2.44	8.22
Gro1	5.72	0.57	7.47
Il1a	4.39	7.17	6.50
Tnf	4.46	5.04	4.85
Serpine1	0.28	0.21	0.66
Il10	3.15	7.93	3.49
Il12b	5.51	0.78	4.92
Il6	8.57	6.17	9.99
Csf2	3.01	6.15	6.85
Scya4	2.04	3.86	3.00

Myd88 Independent Targets

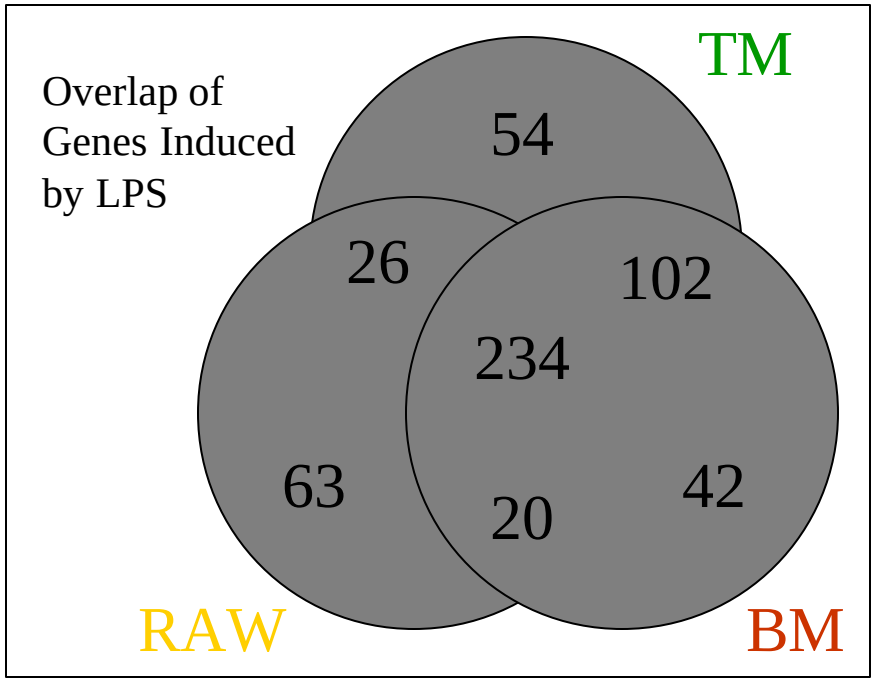


Scyb10	7.78	8.26	10.40
Irf1	3.85	2.10	4.48
Scya5	8.73	8.63	9.47
Icosl	1.46	0.06	0.40
Clecsf9	3.56	3.16	5.18
Traf1	4.65	3.64	4.14
Ifit1	7.43	5.43	8.65
Csf1	1.25	0.52	1.93
Mlp	2.38	3.98	5.08
Vig1	7.86	7.93	8.82
Tlr2	1.62	1.55	5.06
Tyki	6.30	5.66	8.12
Ifit3	6.76	7.30	9.00

NF-kappaB Family



Rela	0.93	0.12	1.67
Relb	1.06	0.67	1.53
Rel	2.34	2.45	1.67
Nfkb1	2.02	2.16	2.46
Nfkb2	1.88	2.55	3.30
Nfkbia	1.89	3.44	3.64



Lipid Genes

Lipid Genes	RAW	TM	BM
Induced in LPS	15	16	15
reduced in LPS	13	33	10
Absent	128	124	150
Total lipid genes	512		

GO enrichment for Absent Genes

Absent All

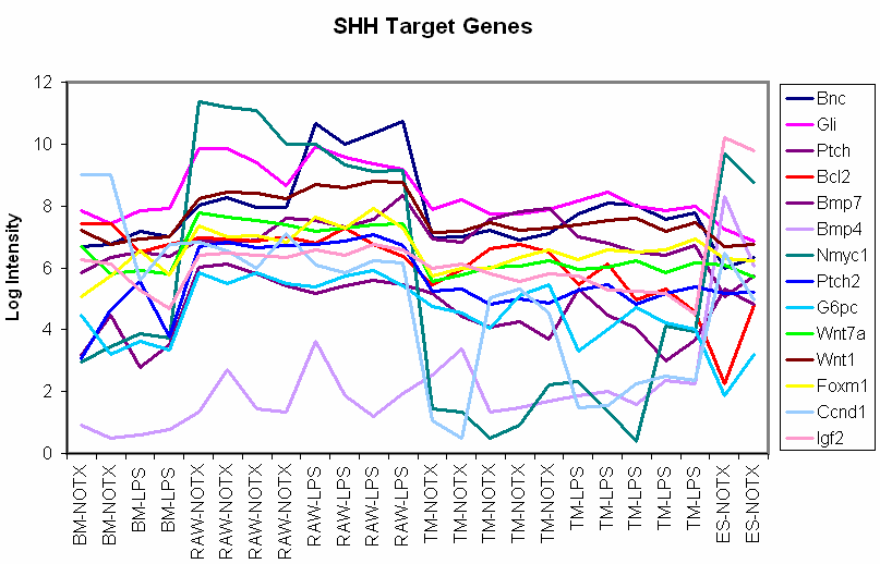
PValue	Group
0.0027899093419244	hormone metabolism (GO:0042445) Genes
0.00283791579678031	hormone biosynthesis (GO:0042446) Genes
0.00283791579678031	C21-steroid hormone biosynthesis (GO:0006700) Genes
0.00536933309022458	C21-steroid hormone metabolism (GO:0008207) Genes

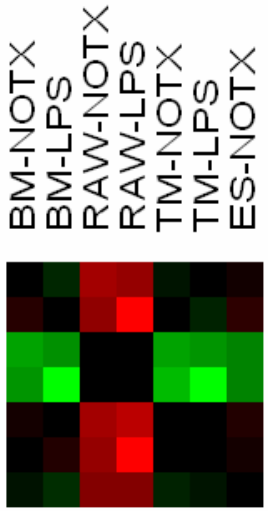
Absent only raw

PValue	Group
0.00560840733250455	cell homeostasis (GO:0019725) Genes
0.00560840733250455	homeostasis (GO:0042592) Genes
0.0194113725493934	p53 modification(30) (p53 modification(30)) Genes
0.0194113725493934	p53 upstream signal/p53 modifiers(26) (p53 upstream signal/p:
0.0267719210457972	p53 Signaling Pathway (p53 Signaling Pathway) Genes

RAW – LPS response

Gene	Fold Change (log2)
Snk	2.2
Pld2	1.0
Ass1	3.0
Gyk	1.4
Ugt1a6	1.2
Hsd3b5	1.7
Cyp2a12	1.1
Gla	2.0
Mthfd2	1.2
Hsd11b2	2.3
Ptgs2	2.4
Pik3ca	1.9
Ggta1	1.0
Agpat4	1.3
Prkr	2.3
Eif2ak3	-1.0
Acadm	-1.1
Mlycd	-1.0
Acad8	-1.2
Adss	-1.2
Lpin1	-1.3
Ptdss2	-1.1
Cte1	-2.0
Ptgds2	-1.6
Slc37a2	-1.3
Alox5ap	-1.3
Glul	-2.1
Acyp1	-1.0



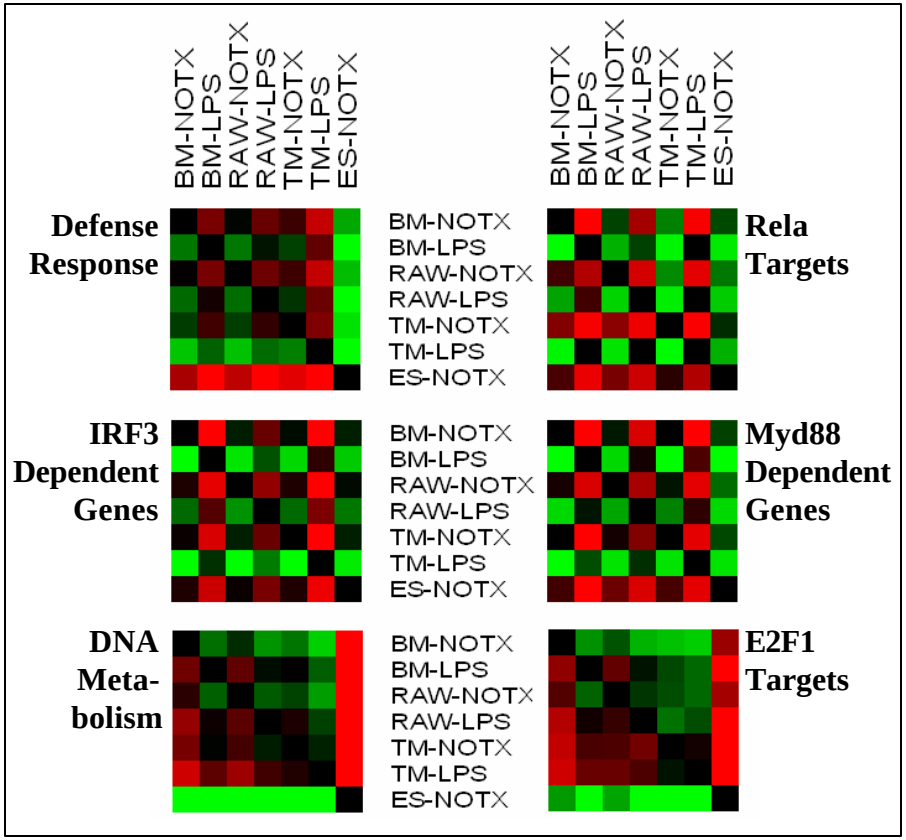


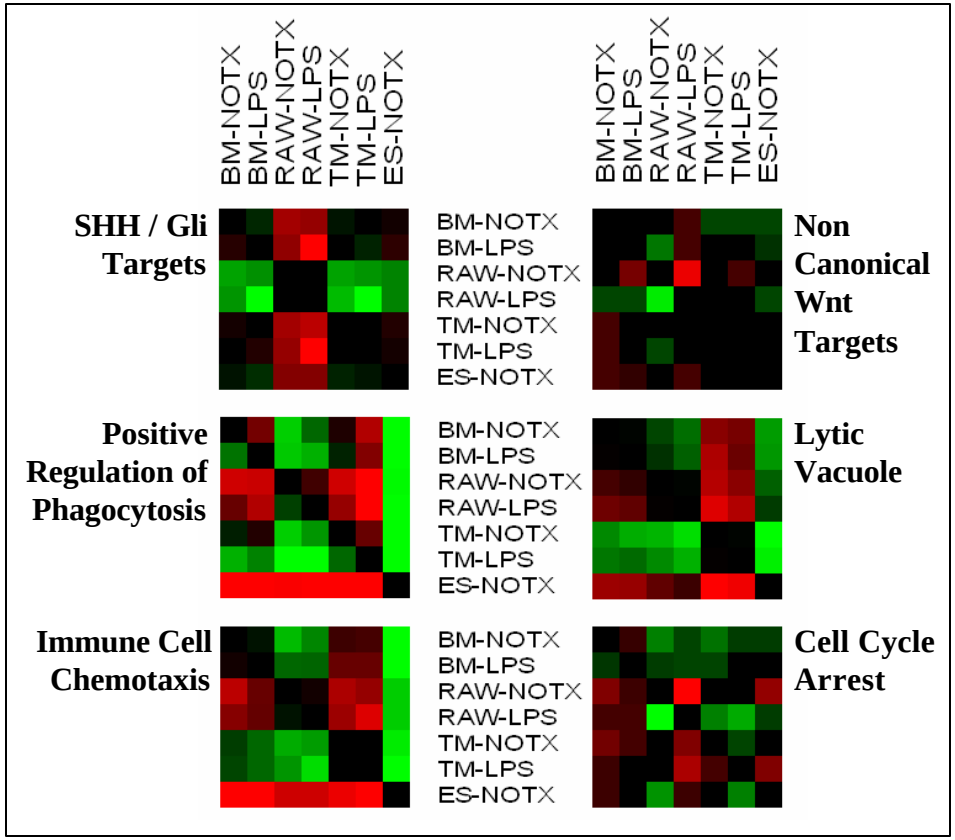
Pathway Comparison

Expression of Column is higher relative to Row

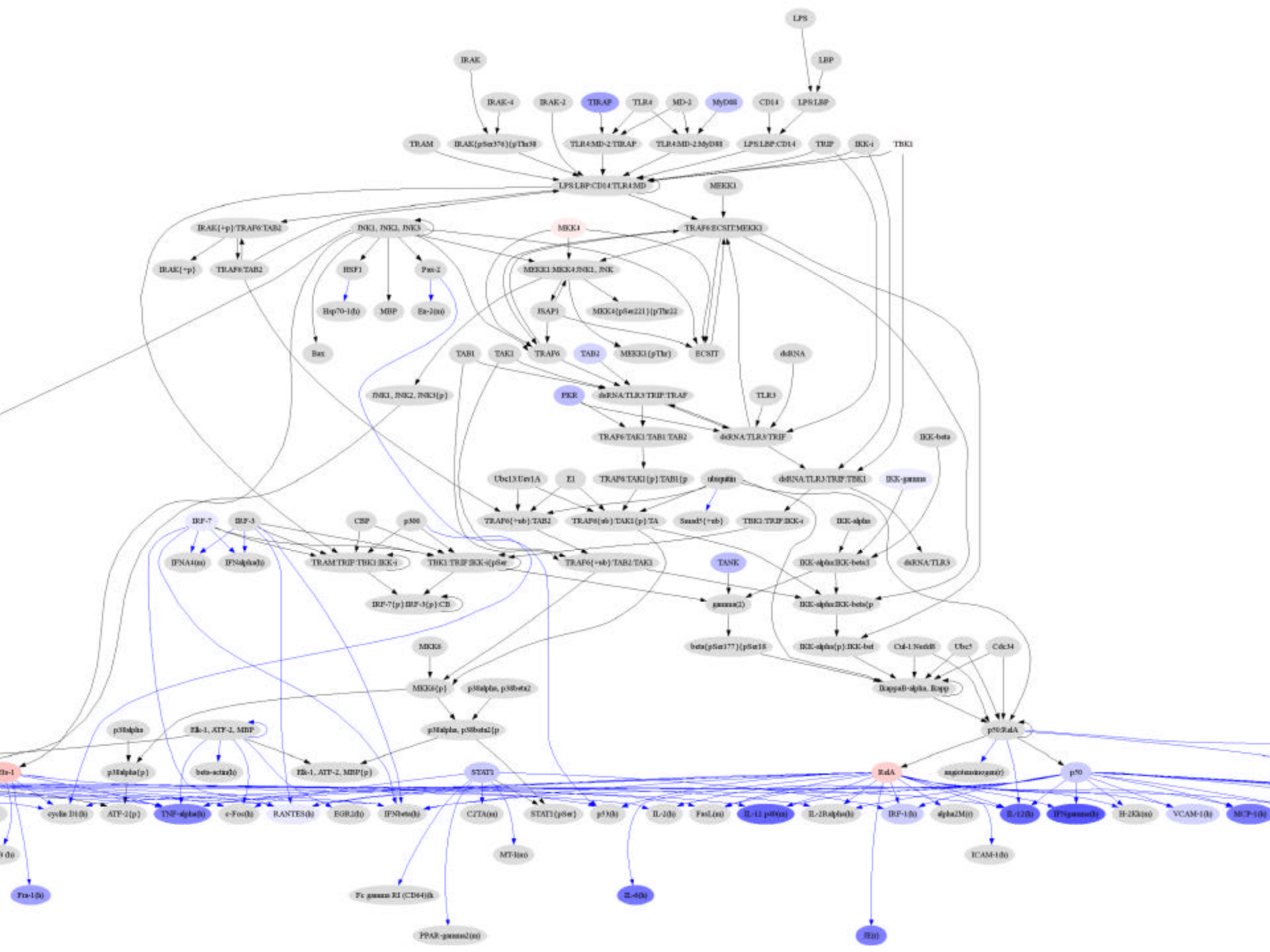
No change in expression between Column and Row

Expression of Column is lower relative to Row

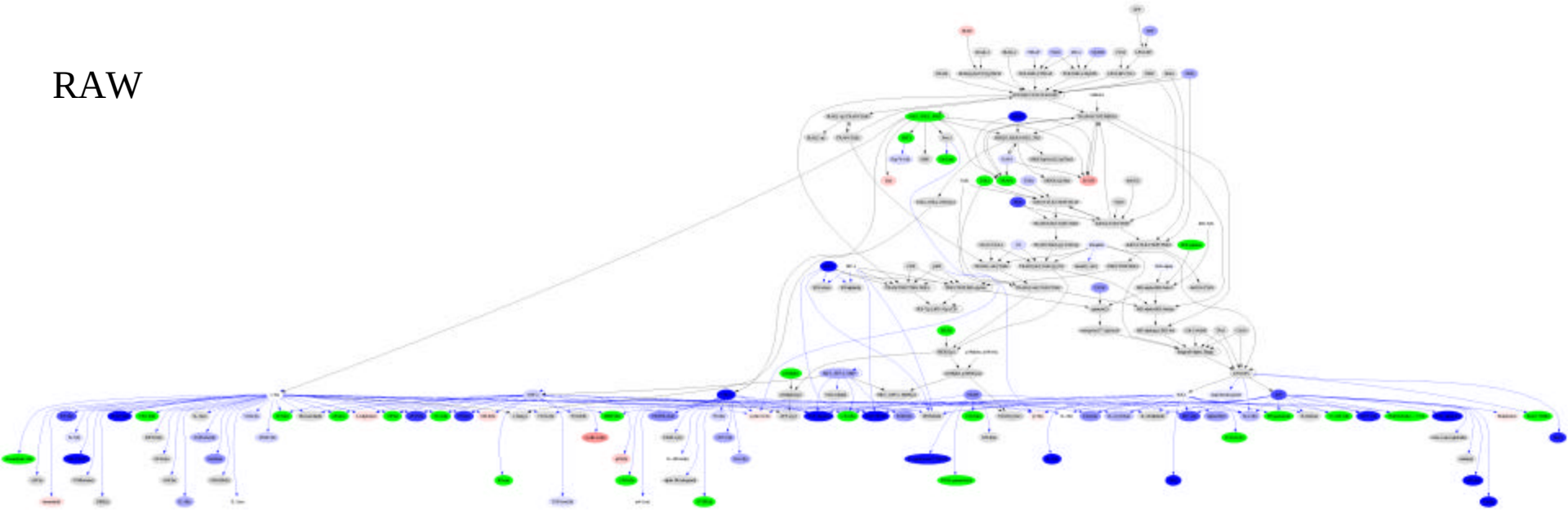




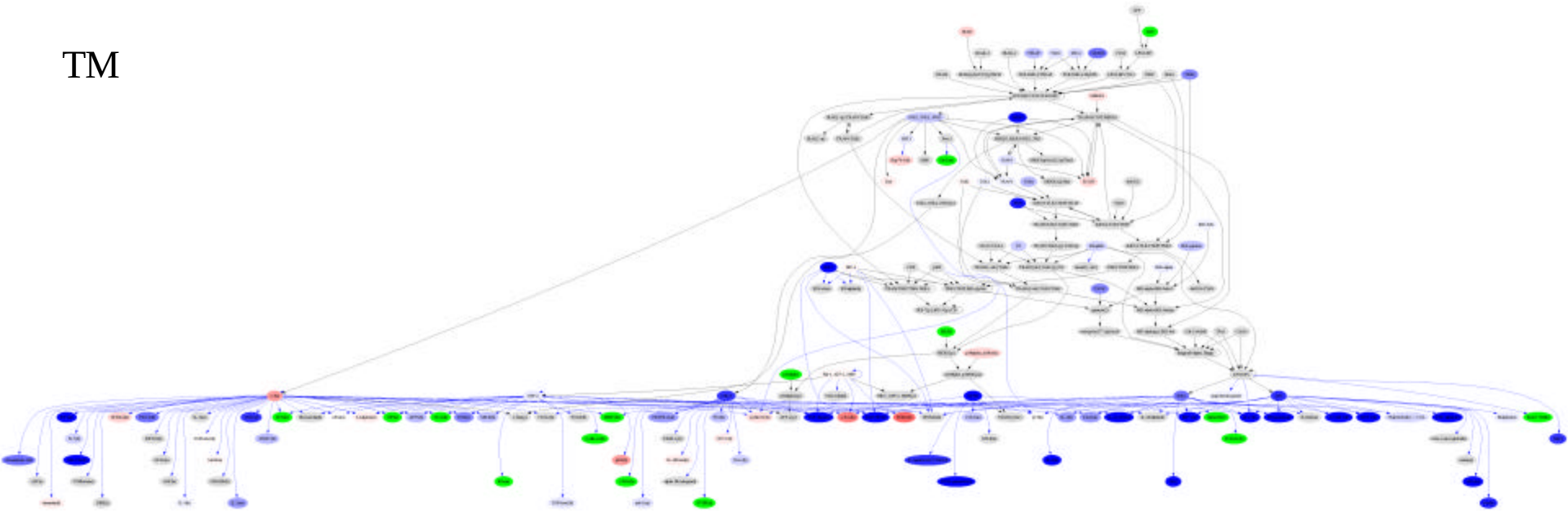
Analysis of Signaling Pathways

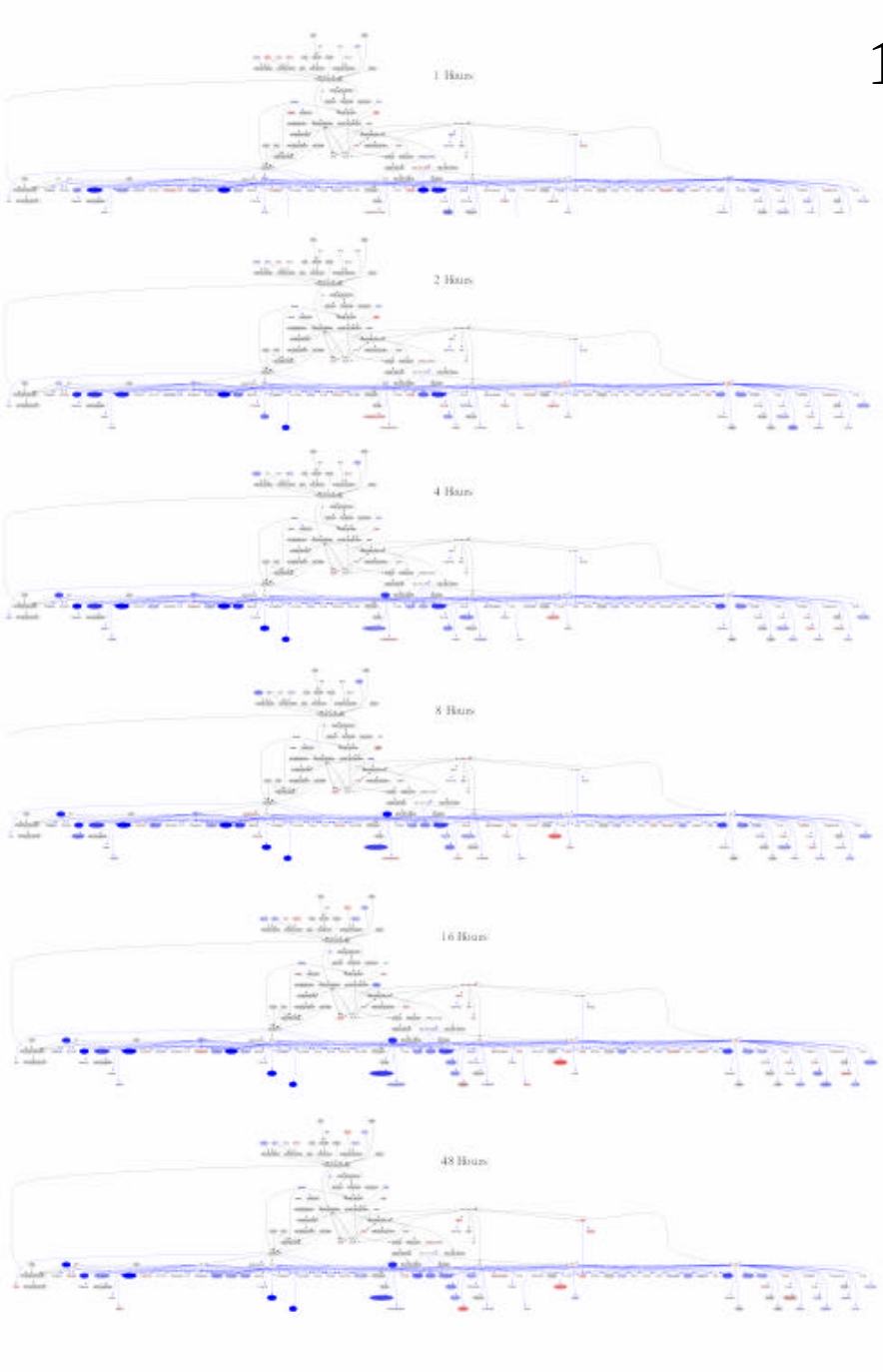

LPS
LBP
IRAK
IRAK-4
IRAK-2
TIRAP
TLR4
MD-1
MyD88
CD14
LPS:LBP
TRAM
IRAK(pSer376)(pThr38
TLR4:MD-2:TIRAP
TLR4:MD-2:MyD88
LPS:LBP:CD14
TRIF
IKK-i
TBK1
LPS:LBP:CD14:TLR4:MD
MEKK1
IRAK(+p):TRAF6:TAB2
JNK1, JNK2, JNK3
MKK4
TRAF6:ECSIT:MEKK1
IRAK(+p)
TRAF6:TAB2
HSF1
Pax-2
MEKK1:MKK4:JNK1, JNK
Hsp70-1(h)
MBP
En-2(m)
JSAP1
MKK4(pSer221)(pThr22
Bax
TAB1
TAK1
TRAF6
TAB2
MEKK1(pThr)
ECSIT
dsRNA
JNK1, JNK2, JNK3(p)
PKR
dsRNA:TLR3:TRIF:TRAF
TLR3
TRAF6:TAK1:TAB1:TAB2
dsRNA:TLR3:TRIF
IKK-beta
Ubc13:Uev1A
E1
TRAF6:TAK1(p):TAB1(p
ubiquitin
dsRNA:TLR3:TRIF:TBK1
IKK-gamma
IRF-7
IRF-3
CBP
p300
TRAF6(+ub):TAB2
TRAF6(ub):TAK1(p):TA
Smad5(+ub)
TBK1:TRIF:IKK-i
IKK-alpha
IFNA1(m)
IFNalpha(h)
TRAM:TRIF:TBK1:IKK-i
TBK1:TRIF:IKK-i(pSer
TRAF6(+ub):TAB2:TAK1
TANK
IKK-alpha:IKK-beta1
dsRNA:TLR3
IRF-7(p):IRF-3(p):CB
gamma(2)
IKK-alpha:IKK-beta(p
MKK6
beta(pSer177)(pSer18
IKK-alpha(p):IKK-bet
Cul-1:Nedd8
Ubc5
Cdc34
MKK6(p)
p38alpha, p38beta2
I(kappa)B-alpha, I(kap
p38alpha
Elk-1, ATF-2, MBP
p38alpha, p38beta2(p
p50:RelA
p38alpha(p)
beta-actin(h)
Elk-1, ATF-2, MBP(p)
STAT3
RelA
angiotensinogen(r)
p50
cyclin D1(h)
ATF-2(p)
TNF-alpha(h)
c-Fos(h)
RANTES(h)
EGR2(h)
IFNbeta(h)
C2TA(m)
STAT1(pSer)
p53(h)
IL-2(h)
FasL(m)
IL-12 p80(m)
IL-2Ralpha(h)
IRF-1(h)
alpha2M(r)
IL-12(h)
IFNgamma(h)
H-2Kk(m)
VCAM-1(h)
MCP-1(h)
ICAM-1(h)
MT-I(m)
Fra-1(h)
Fc gamma RI (CD64)(h
IL-6(h)
PPAR-gamma2(m)
IE(r)

RAW



TM





1h

2h

4h

8h

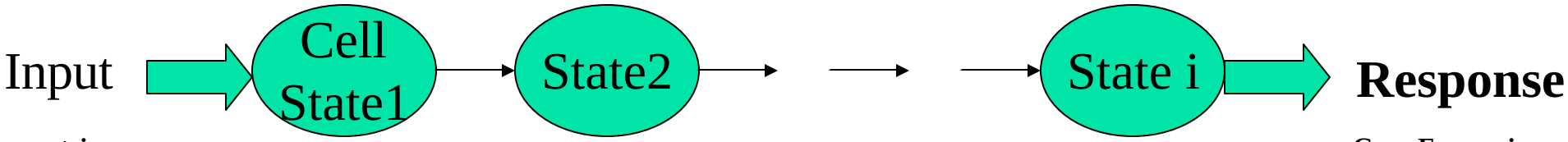
16h

48h

Expression Time Response in the Signaling network

NEW TECHNOLOGIES

CELLULAR RESPONSE TO STIMULUS



- **proteins**
- **peptides**
- **amino acids**
- **nucleotides**
- **retinoids**

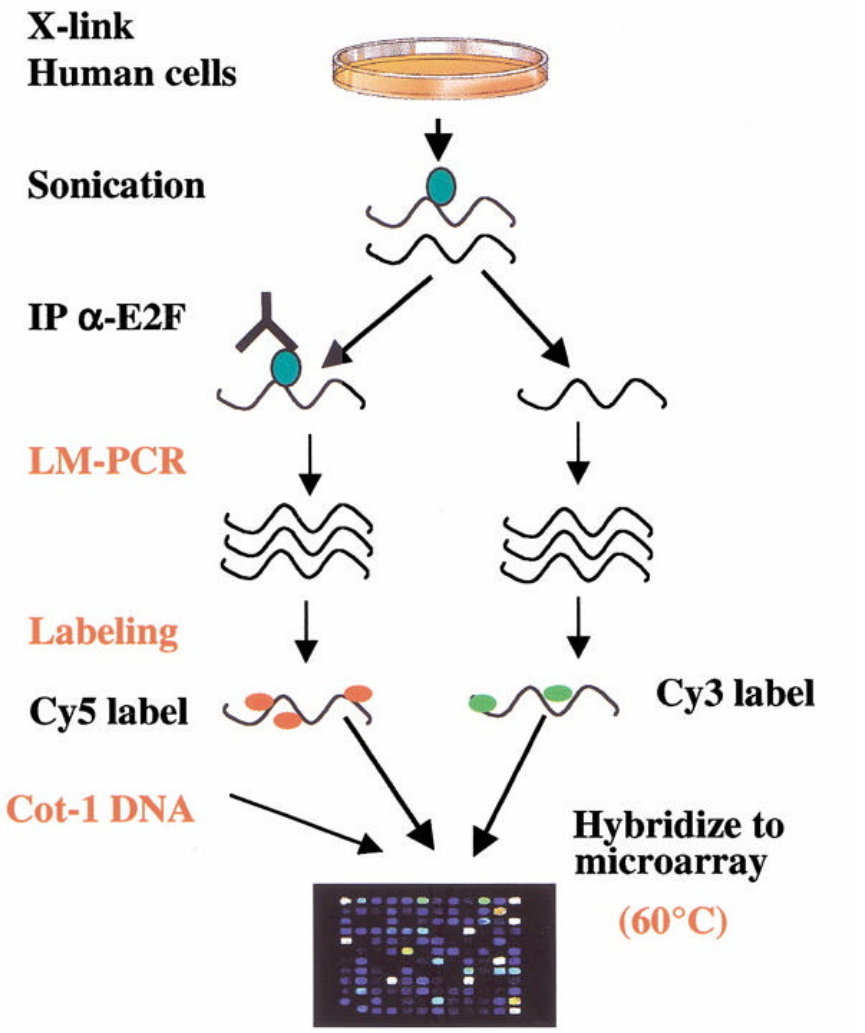
State: genes, proteins, metabolites, ions……

The Parts List Problem!

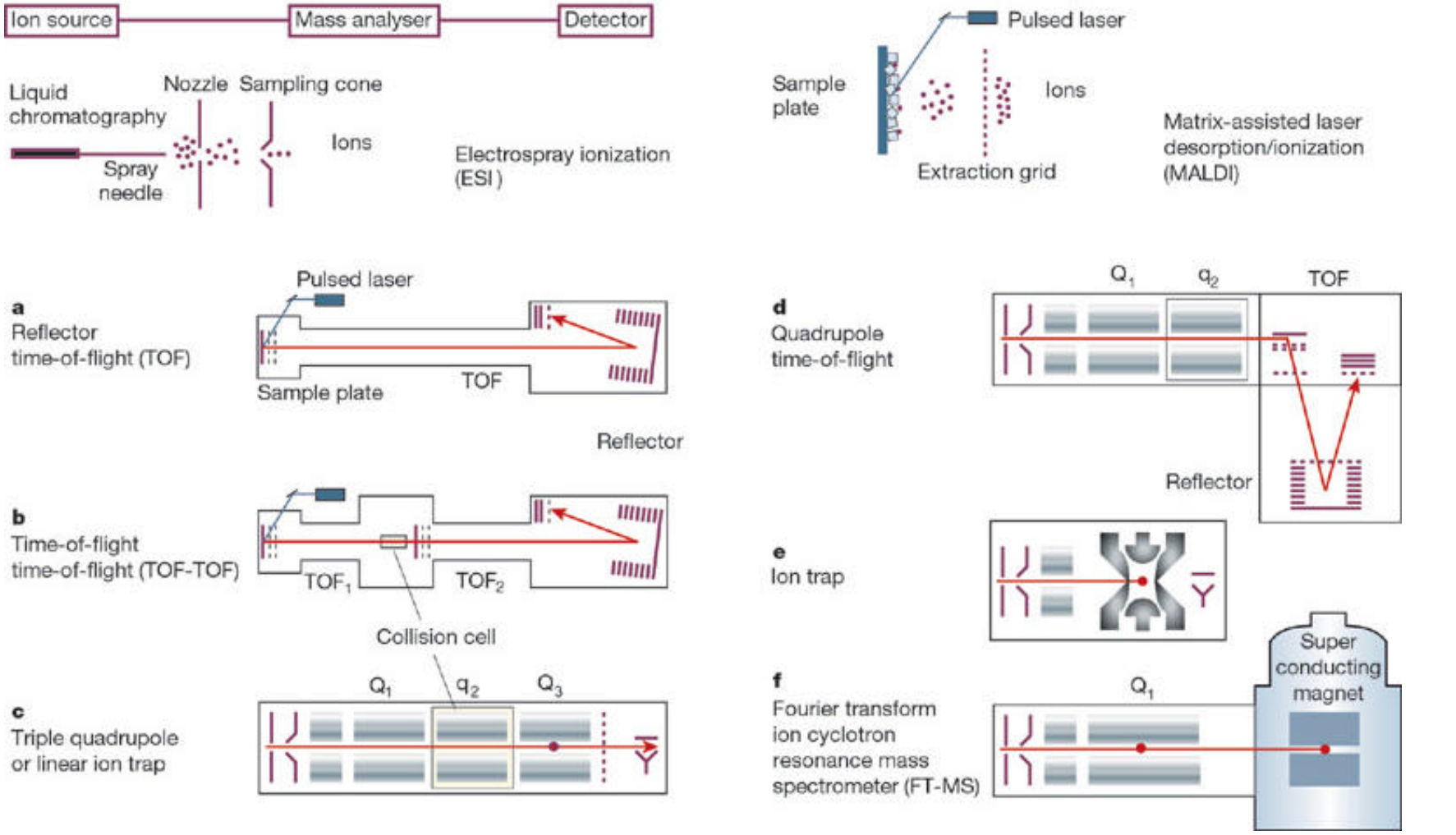
Automated sequencing machines at the Center for Genome Research in the Whitehead Institute

For the future...

- Mouse Promoter Microarray
  - Monitor Protein-DNA interations
  - Perform ChIP on Chip against 40,000 probes

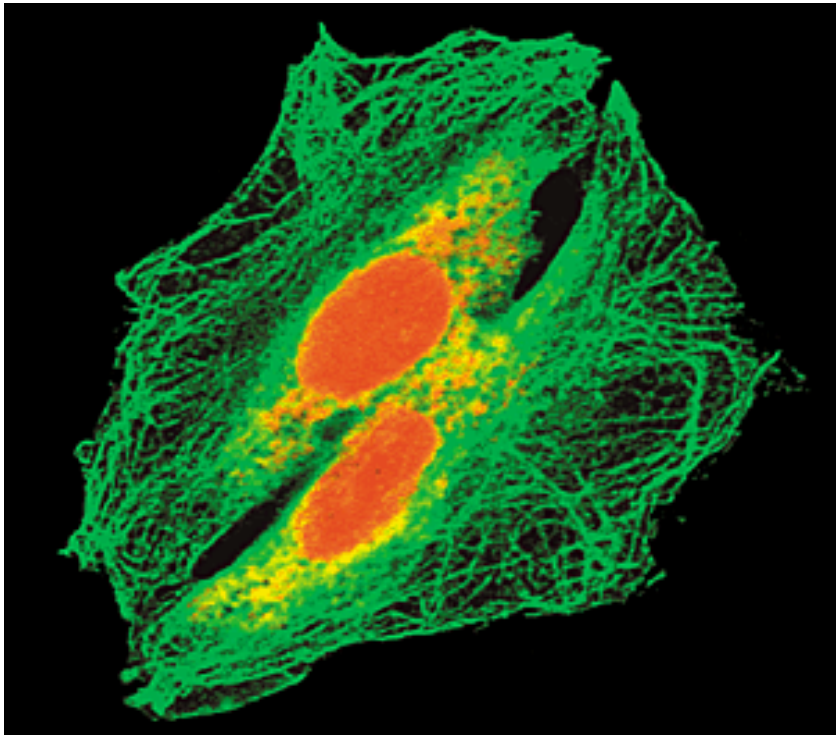


Modern Mass Spectrometry



Quantum Dots probe cell systems

Neither the catalog of the 30,000 or so genes of the human genome, nor that of the proteins that these genes encode, will be sufficient to gain a workable understanding of cell biology. The true complexity in intracellular signaling will only be fully grasped through direct observation of the interactions between key biomolecules. "Which proteins interact with which, and at what time in the cycle? We need intracellular imaging tools, both high-resolution and high sensitivity, to measure this.



A confocal image of fibroblasts using green quantum dots staining tubulin. Nuclei are stained red with ethidium bromide.

Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots
Xingyong Wu *et al.*
Nature Biotechnology 21, 41 - 46 (2003)

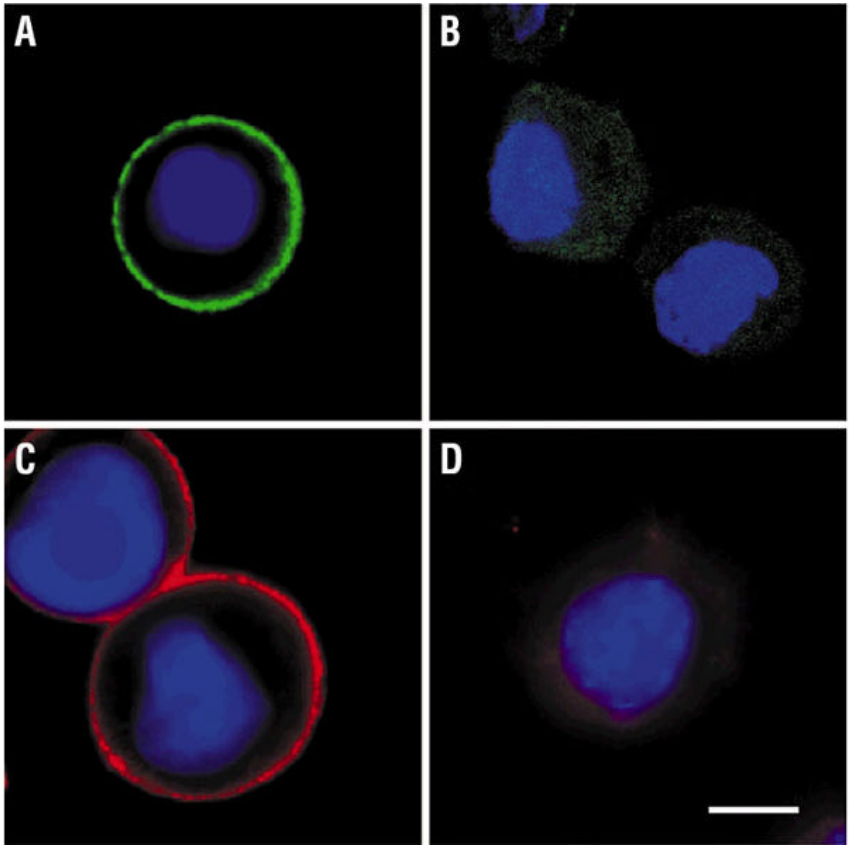
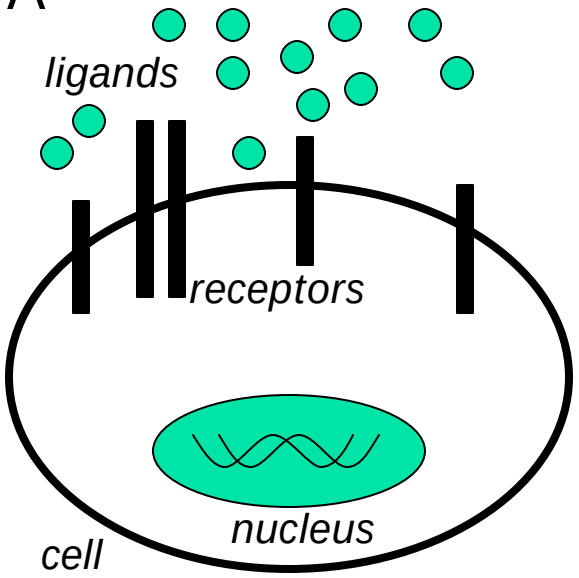


Figure 1: Detection of cancer marker Her2 with QD-IgG.
(A, C) Fixed breast cancer SK-BR-3 cells were incubated with monoclonal anti-Her2 antibody and goat anti-mouse IgG conjugated to QDs. Her2 was clearly labeled with (A) QD 535–IgG and (C) QD 630–IgG (B, D). When cells were incubated with normal mouse IgG and QD-IgG, there were no detectable or very weak nonspecific signals on the cell surface. The nuclei were counterstained with Hoechst 33342 (blue). Filter sets ex 480 20 nm/em 535 10 nm and ex 560 27.5 nm/em 635 10 nm were used for QD 535 and QD 630, respectively. Scale bar, 10 m.

Box 1.

Pappin, Subramaniam, Hunter & Palsson (2005)

A

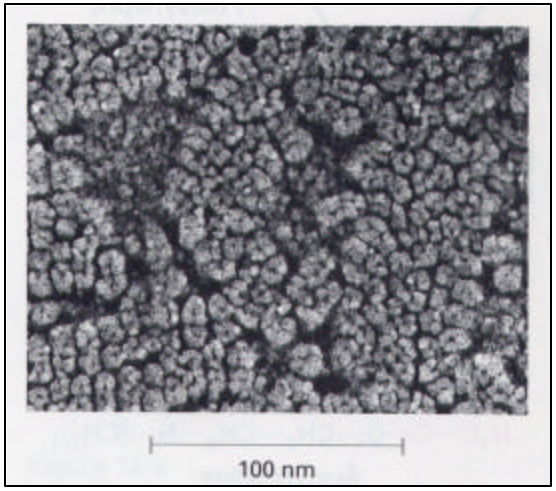


How many ligand molecules might a typical cell be exposed to at a given instant in time?

What is the maximum number of receptors in a cell membrane (see B below)?

How many transcription factors are needed to control 30,000 genes?

B



Axon terminal surface of a neuron. Each protein is an acetylcholine receptor. From Stryer (1995) Biochemistry.

Polymorphic drug-metabolizing enzymes

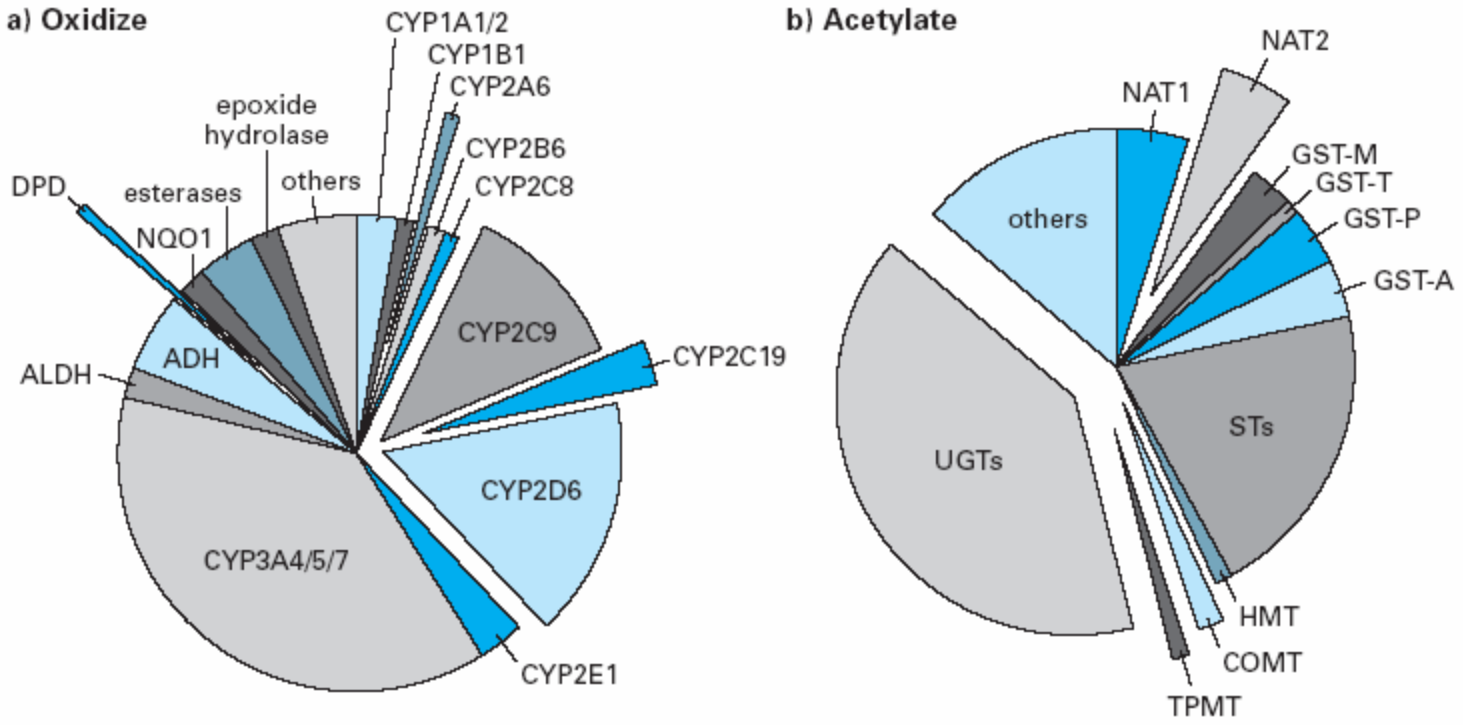
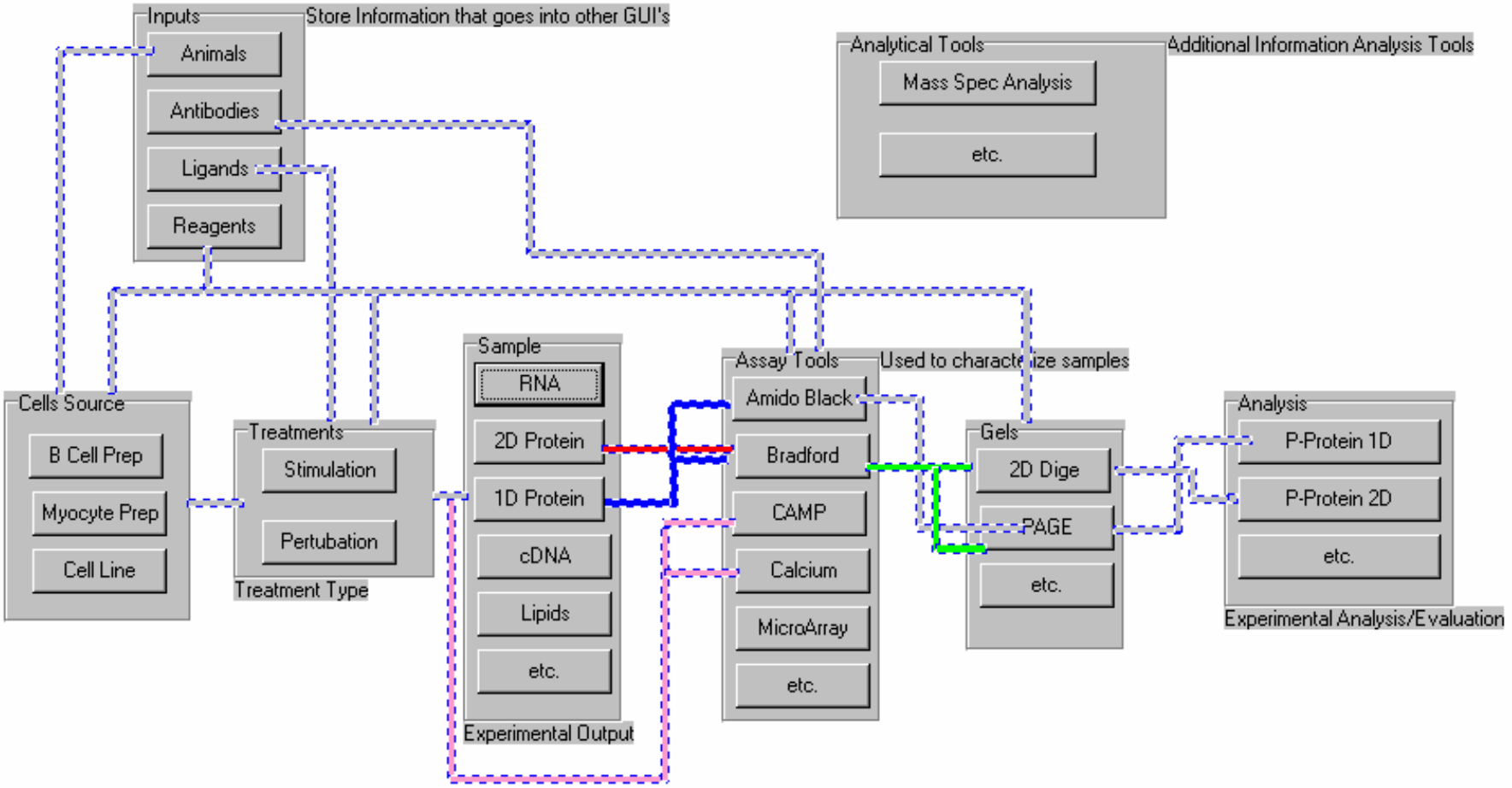
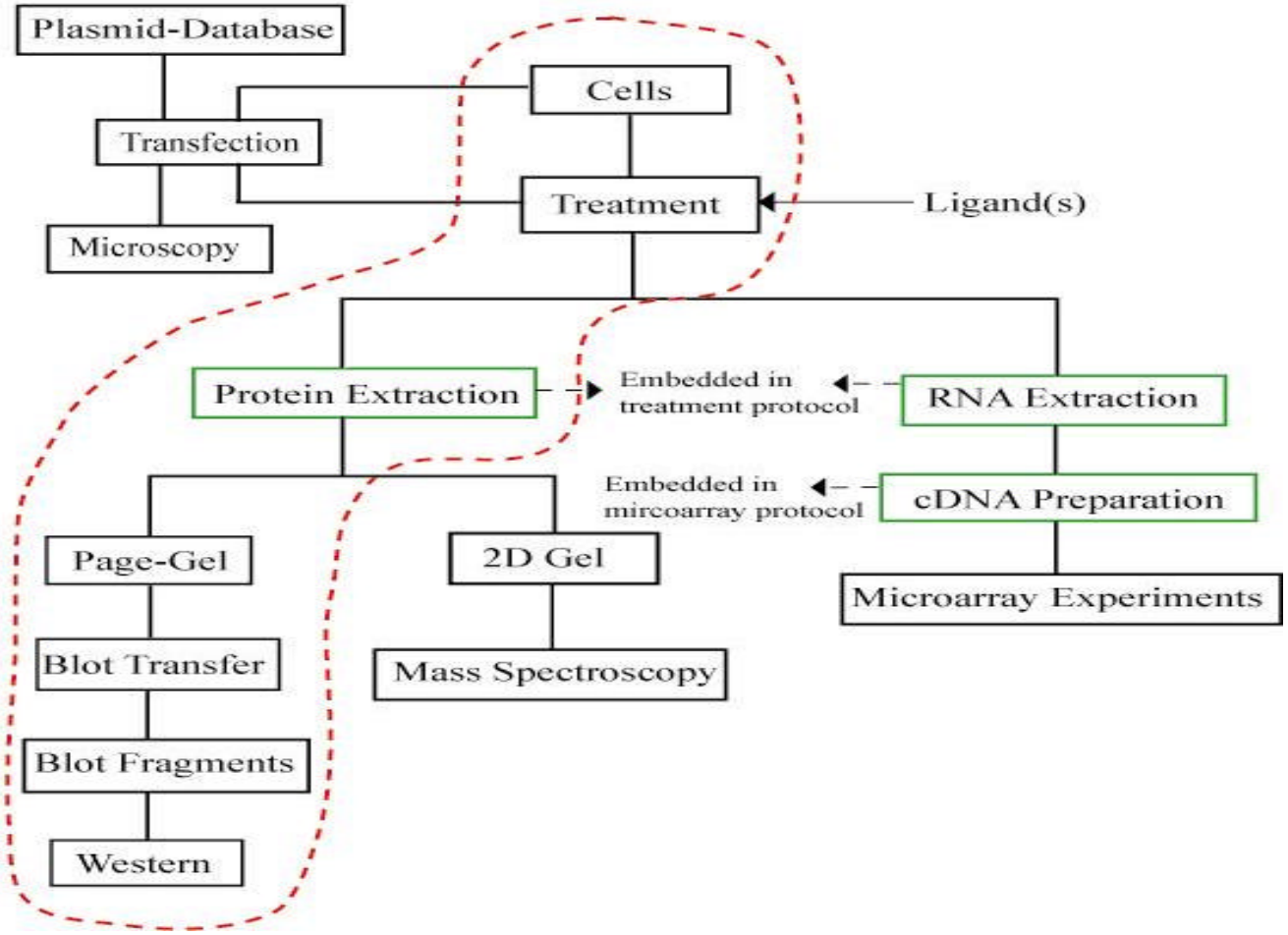


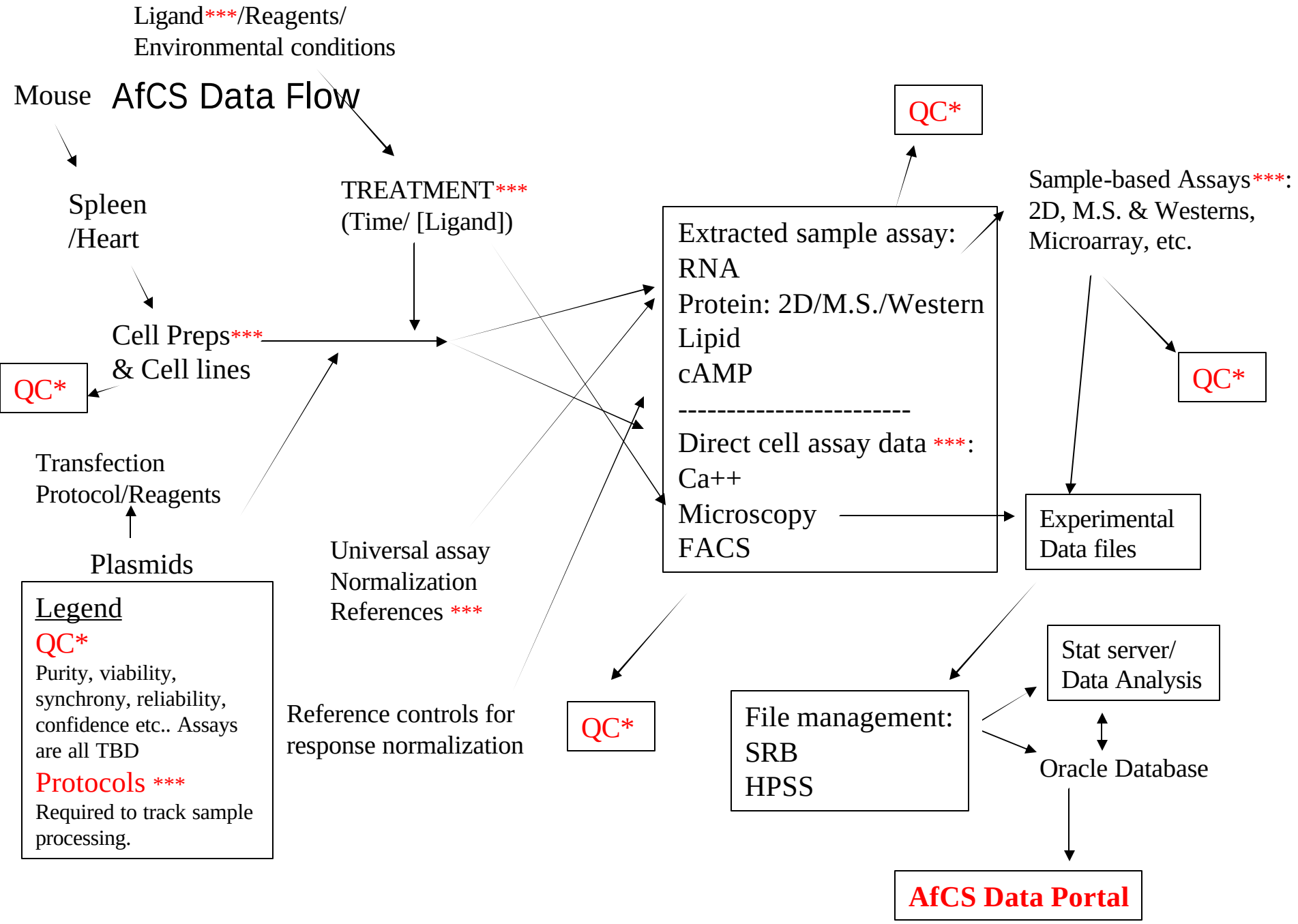
FIGURE 3.9 • Polymorphic drug metabolizing enzymes. **a)** Some enzymes oxidize drugs and make them more reactive. **b)** Other enzymes add acetyl groups onto the most reactive portion of drugs and typically inactivate them. The percentage of oxidization and acetylation of drugs that each enzyme contributes is estimated by the relative size of each section of the corresponding chart.

AfCS Data Flow



A snippet of *AfCS* Experimental Flow




Ligand***/Reagents/
Environmental conditions
Mouse
AfCS Data Flow
Spleen
/Heart
TREATMENT***
(Time/ [Ligand])
QC*
Sample-based Assays***:
2D, M.S. & Westerns,
Microarray, etc.
Extracted sample assay:
RNA
Protein: 2D/M.S./Western
Lipid
cAMP

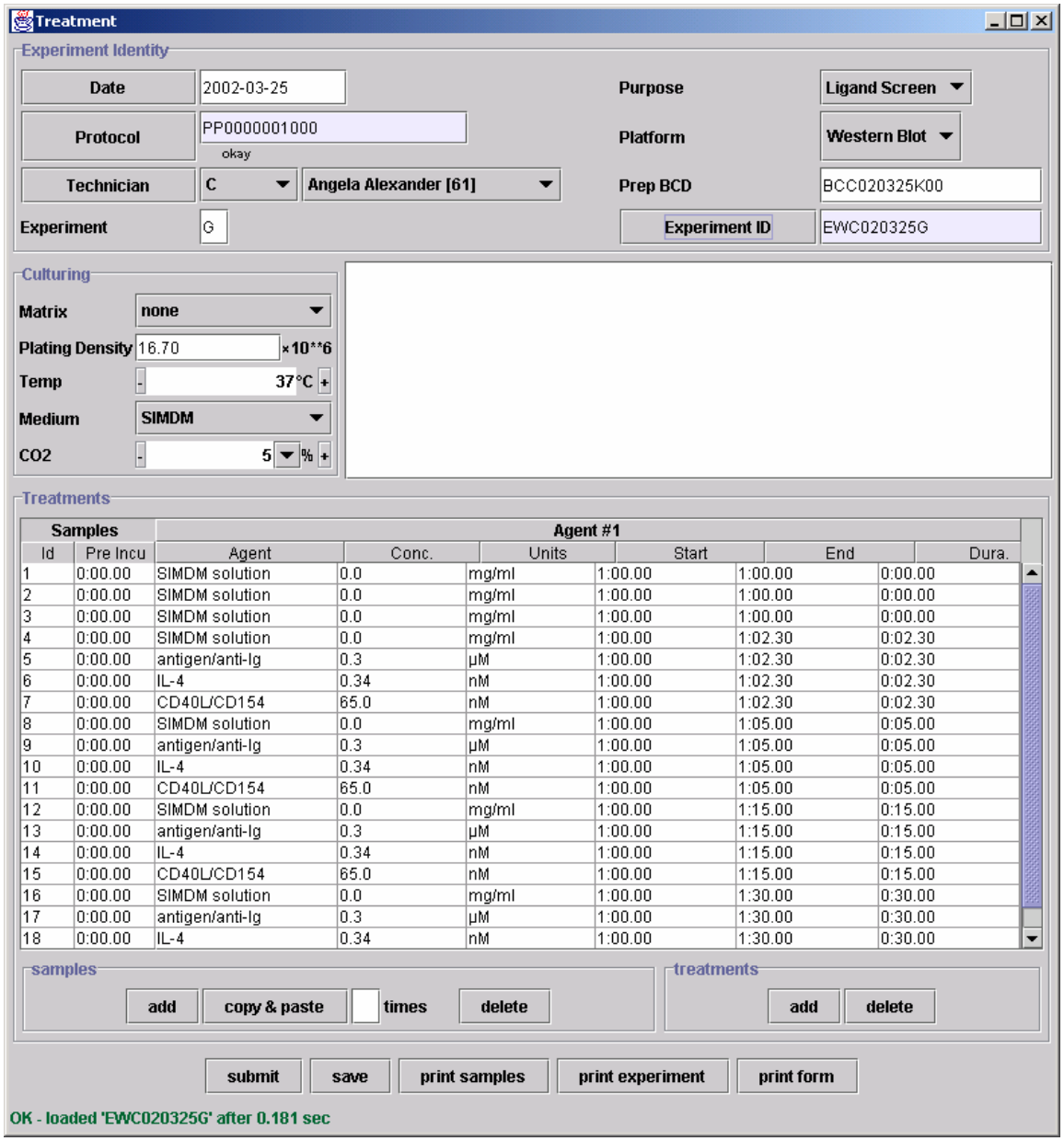
Direct cell assay data ***:
Ca++
Microscopy
FACS
Cell Preps***
& Cell lines
QC*
QC*
Transfection
Protocol/Reagents
Plasmids
Experimental
Data files
Universal assay
Normalization
References ***
Legend
QC*
Purity, viability, synchrony, reliability, confidence etc.. Assays are all TBD
Protocols ***
Required to track sample processing.
Reference controls for
response normalization
QC*
File management:
SRB
HPSS
Stat server/
Data Analysis
Oracle Database
AfCS Data Portal

Treatment GUI

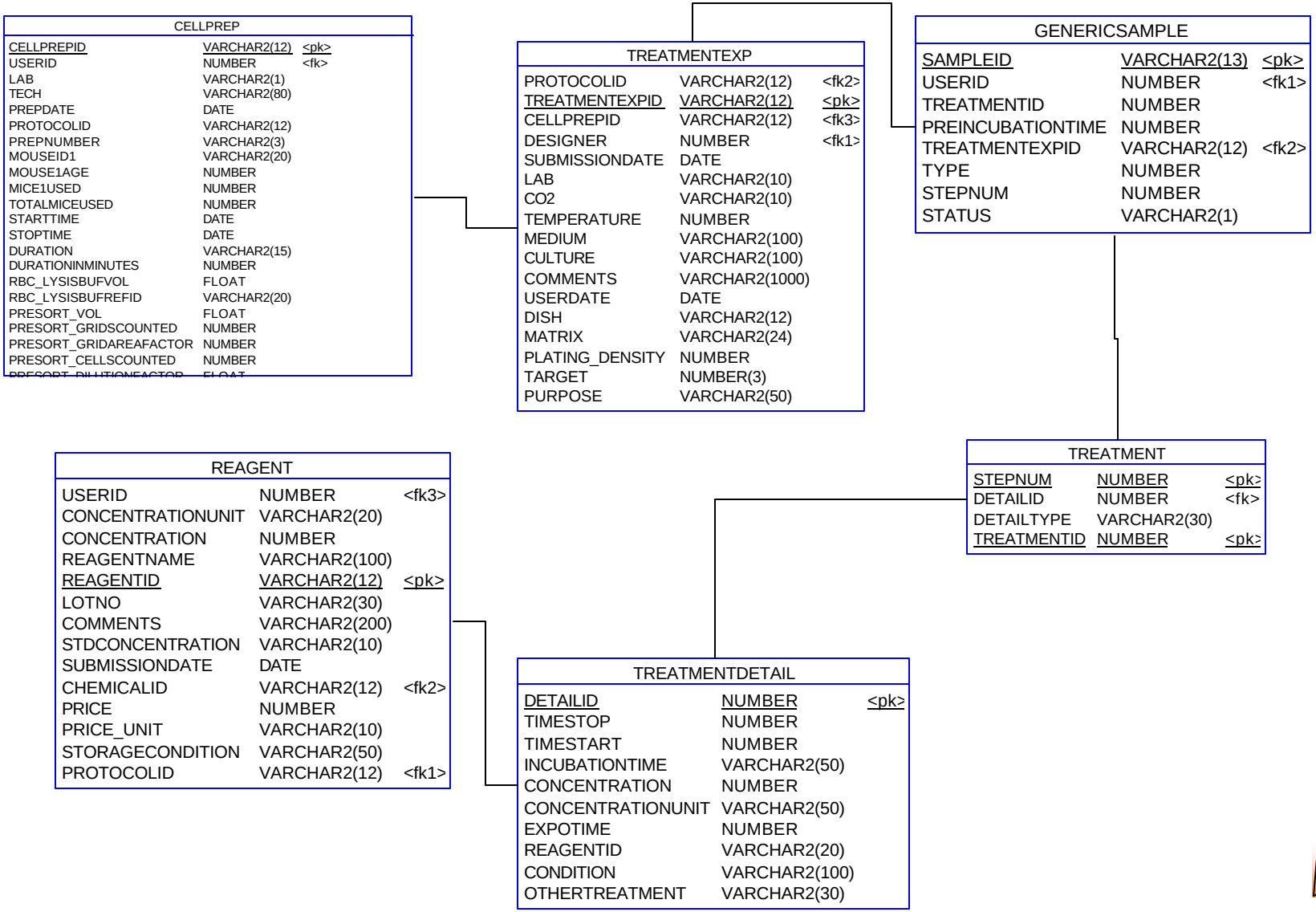
Inputs Cell Prep ID and yields unique treatment and sample IDs. This application also requires a protocol that is tied to the assay for which the sample is created, e.g. Western, Microarray etc. It captures the treatment details including incubation conditions, cell density, ligand and its concentration and time of exposure.

This GUI also functions as a query tool. By entering an ID into the 'Experiment ID' and clicking on 'barcode', it re-loads the data previously submitted for that ID.

Treatment GUI



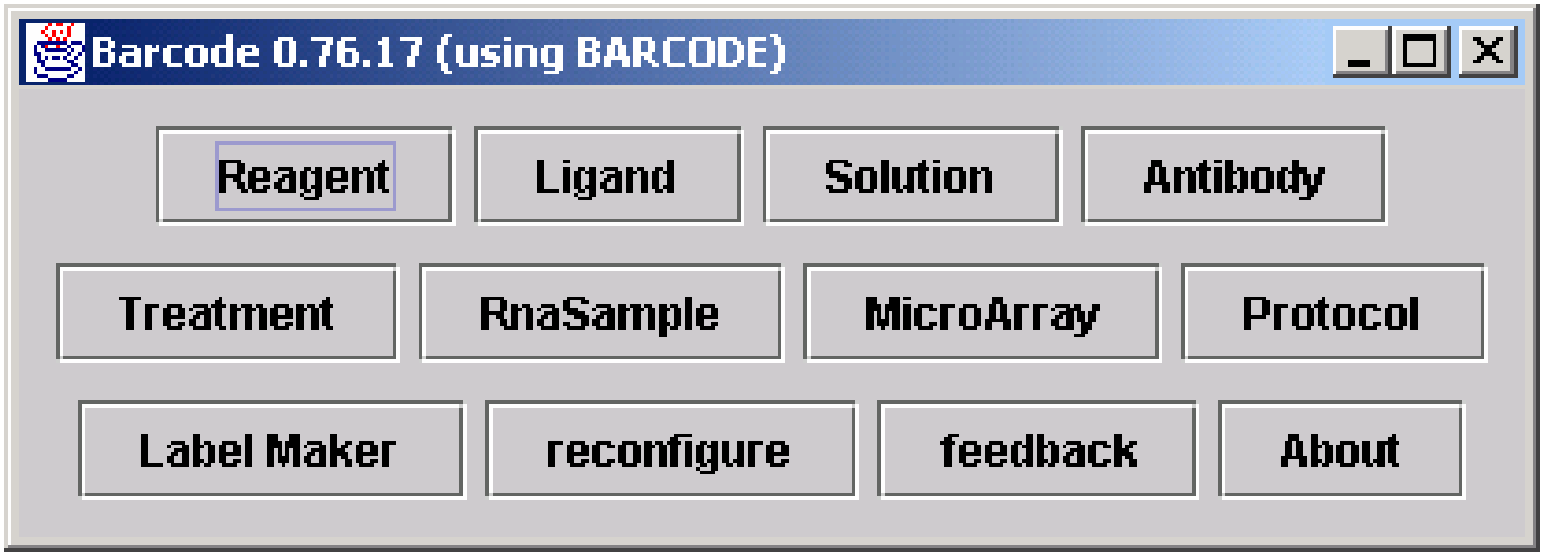
Treatment Tables



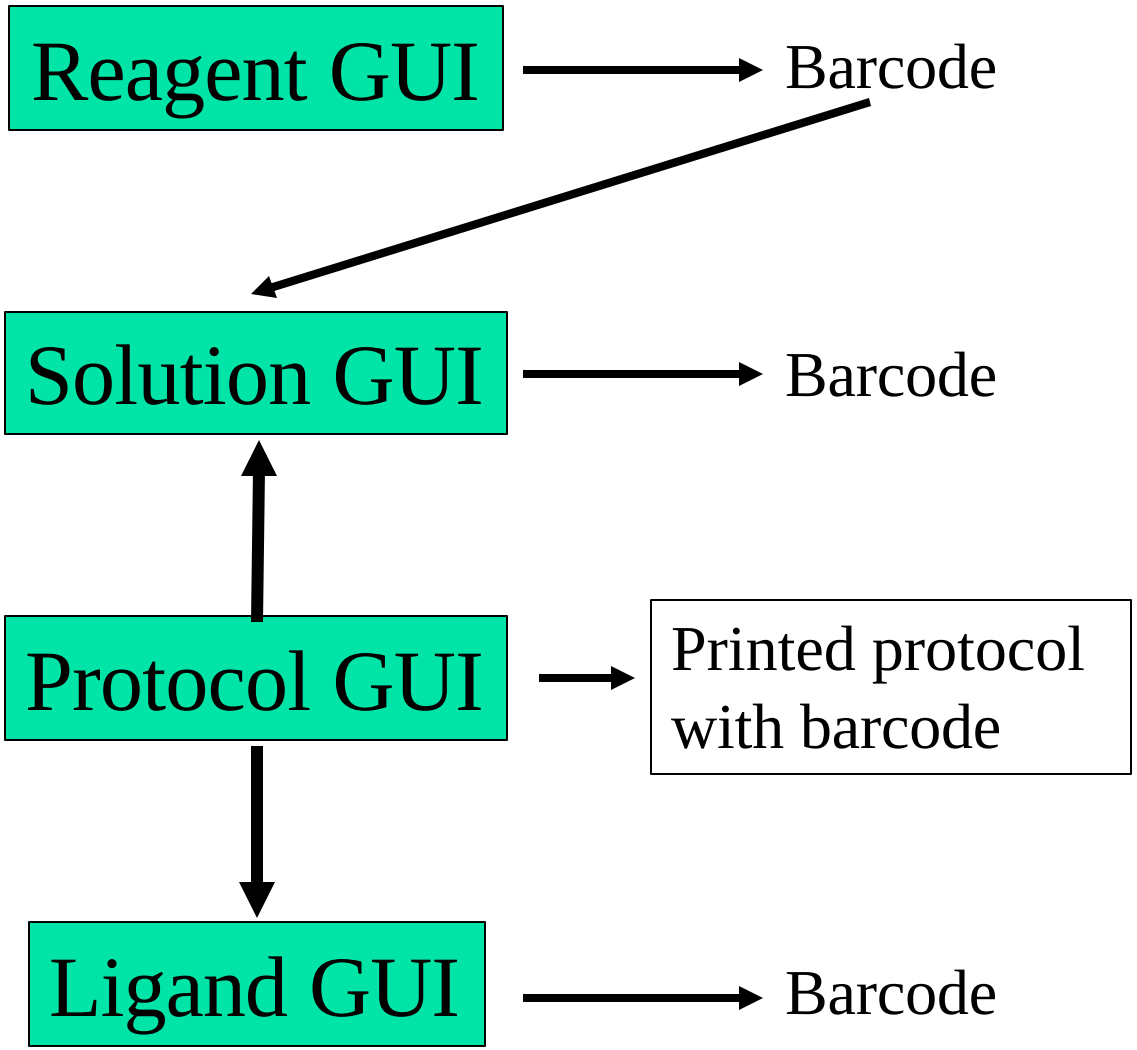
AfCS

2002

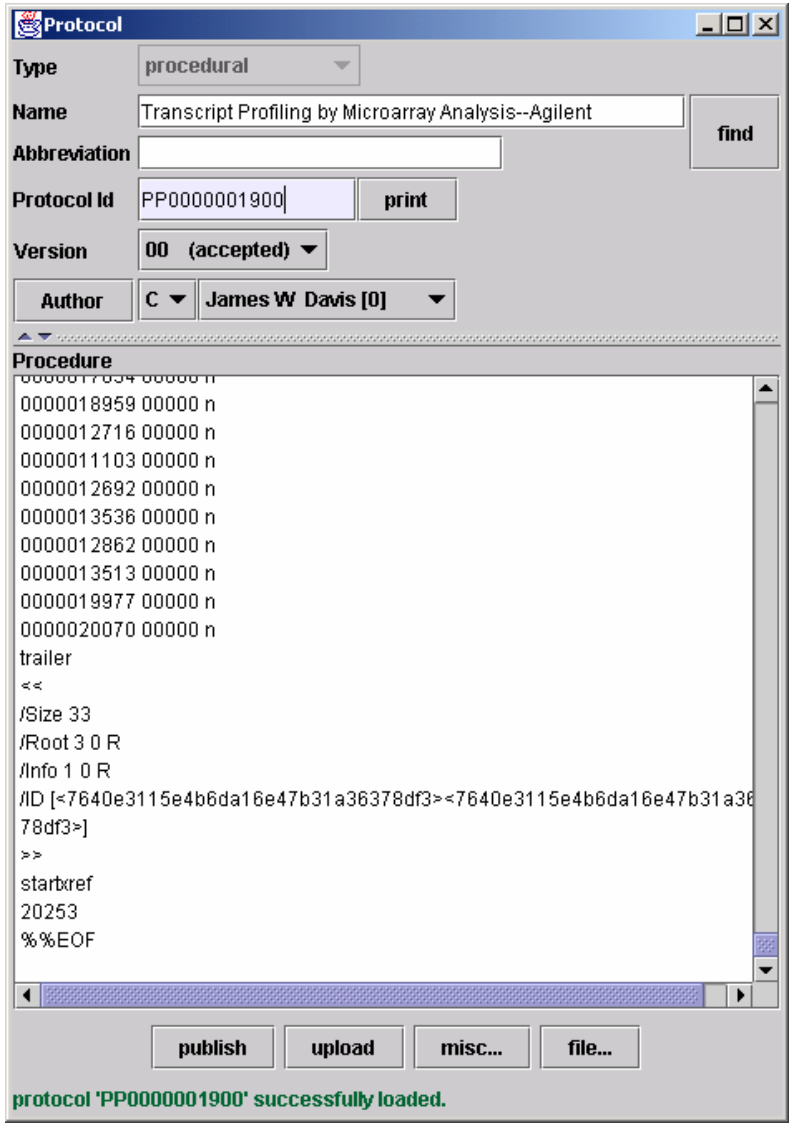
Barcode GUI


Barcode 0.76.17 (using BARCODE)
Reagent
Ligand
Solution
Antibody
Treatment
RnaSample
MicroArray
Protocol
Label Maker
reconfigure
feedback
About

Protocols: Interaction of the GUIs



Protocol GUI



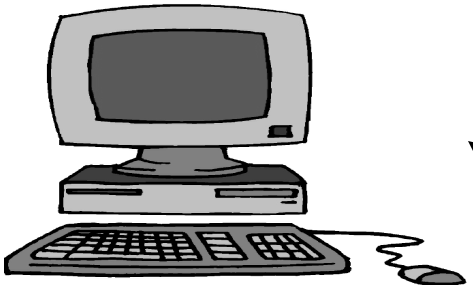
LIMS is based on a client server paradigm comprising of a

- Database server and
- Client (Application)

Applications are developed using java technology, therefore making it easily portable into multiple software platforms.

Data is stored in Oracle 9.2 database – a relative database Management system.

AfCS Laboratory Information Management System (LIMS)



Sample tracking

Barcode Scanner

Data Integrity/ Storage

Oracle
Database

Sample Preparation

Sample Processing

Sample Handling

IP Barcode printer

Barcode

Test tube/
Sample vial

Barcode Database Schema

Data from the aforementioned applications are stored in database tables designed according to data entities.

The basic flow of the schema is as follows:

Harvested or passaged cells are treated under various conditions; then the cells are either assayed directly or extracted components such as RNA, protein and cAMP are prepared for the experiments (i.e. Calcium, cAMP Elisa, Western blot, Microarray etc).

The database schema can be divided into these five sections:

1) Reagent/Ligand/Solution Tables
2) Treatment Tables
3) Microarray Tables
4) Page/Western blot Tables
5) Protocol Tables
6) Transduction/Microscopy Tables

Installation Requirements:

Software:500 MB RAM

Hardware: PC/Linux/Solaris box

Accessory software: Java Runtime Environment(1.2 and later) with java webstart

A barcode printer & a barcode scanner

A barcode printer & a barcode scanner

Zebra:

Zebra Z4M barcode printer - 300DPI - 4MB Total - 2MB Usable
Zebranet Print Server II, External, Parallel

Symbol:

Cyclone Scanner USB & Synapse Adapter

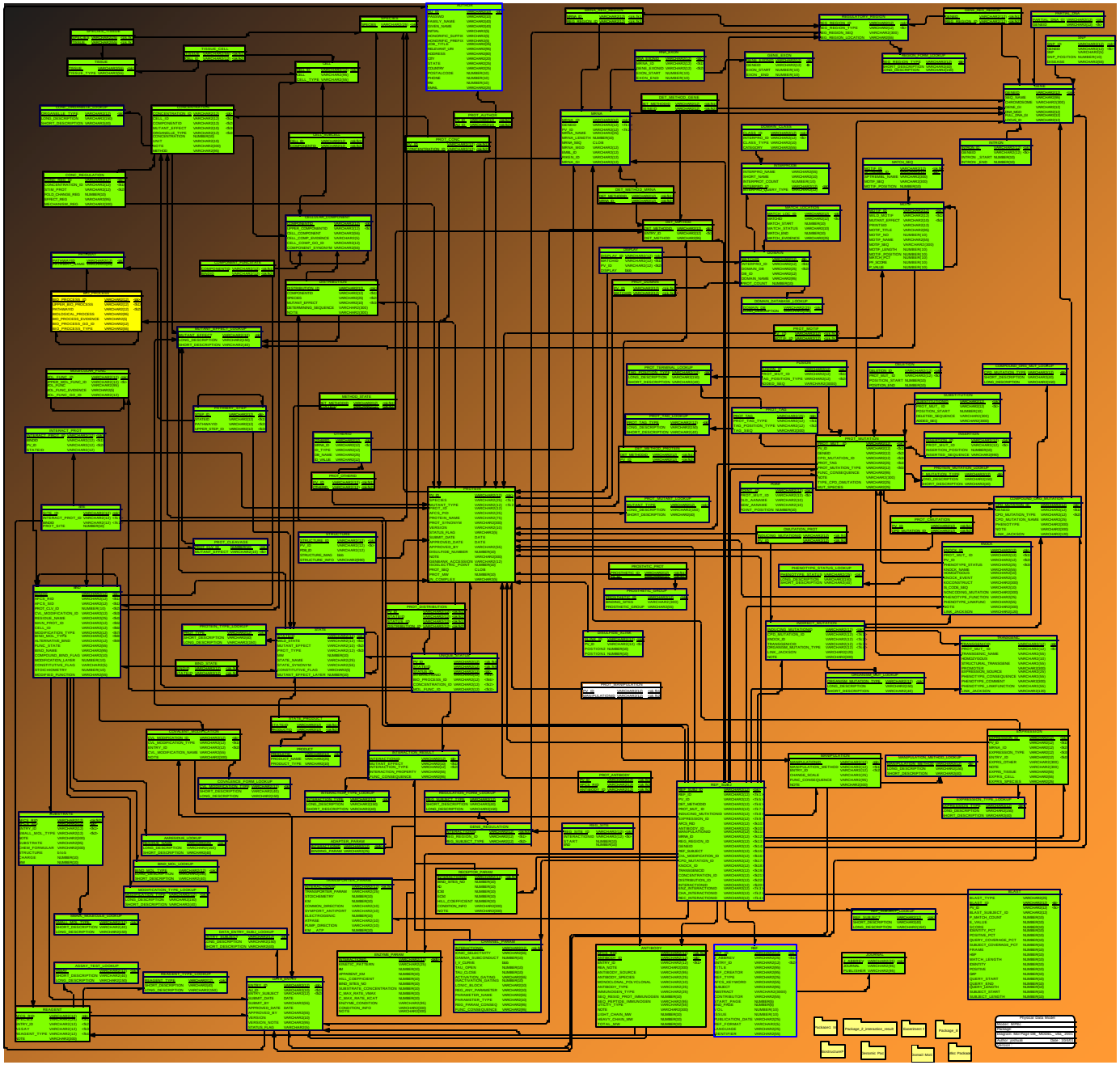
Molecule Pages – Automated Data

- **“Automated Data” for each protein is provided to both the public and the authors, and can be referenced by the author when entering their own data**
- **The types of automated data available will:**
  - **Summaries of and links to external database records that correspond to, or are related to, the author’s protein**
    - **(e.g., Genbank, SwissProt and PDB records)**

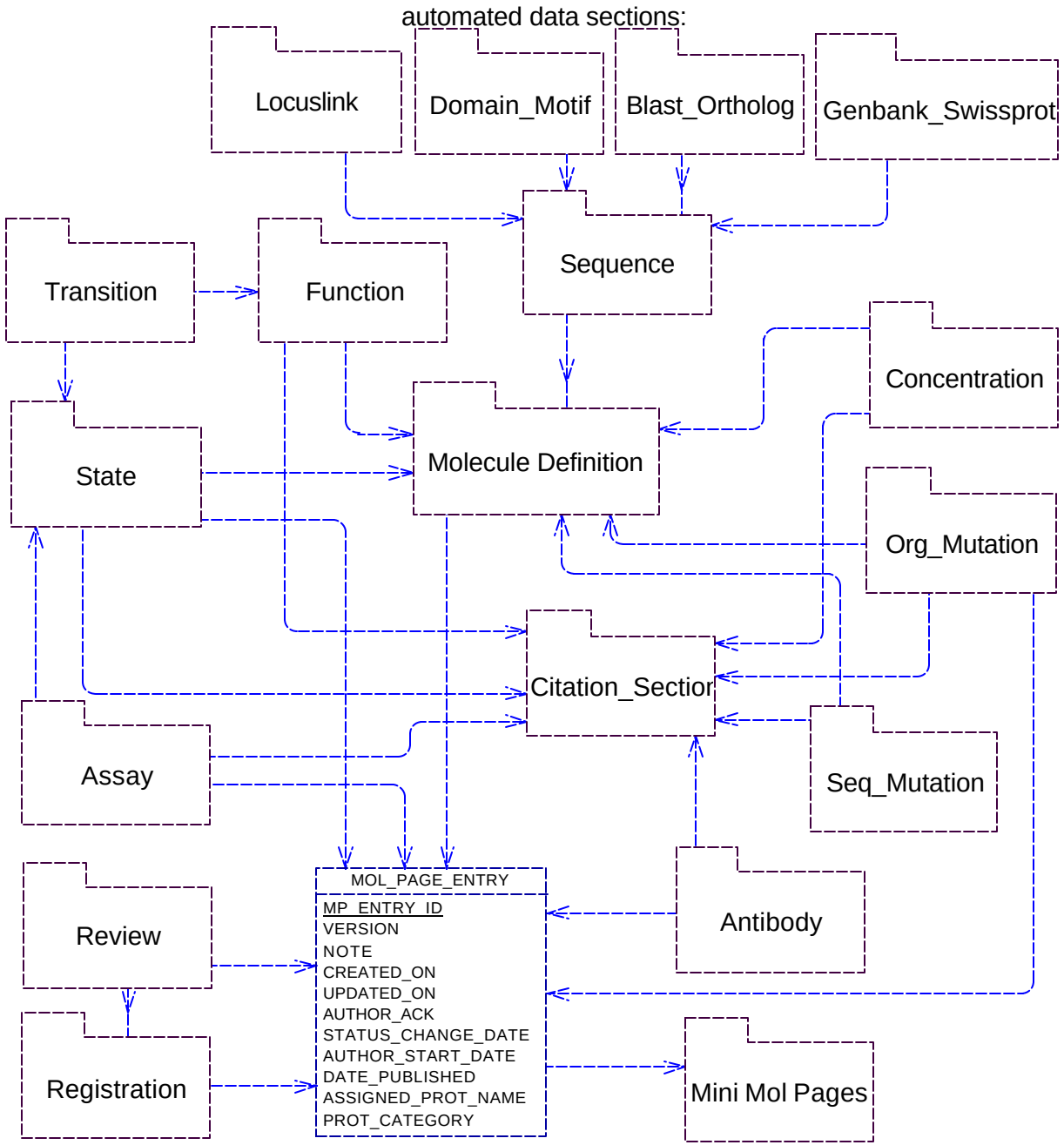
What is needed?

- An Ontology
- A Data Structure
- A Database
- Query Interfaces
- Applications Interfaces
- Data Content

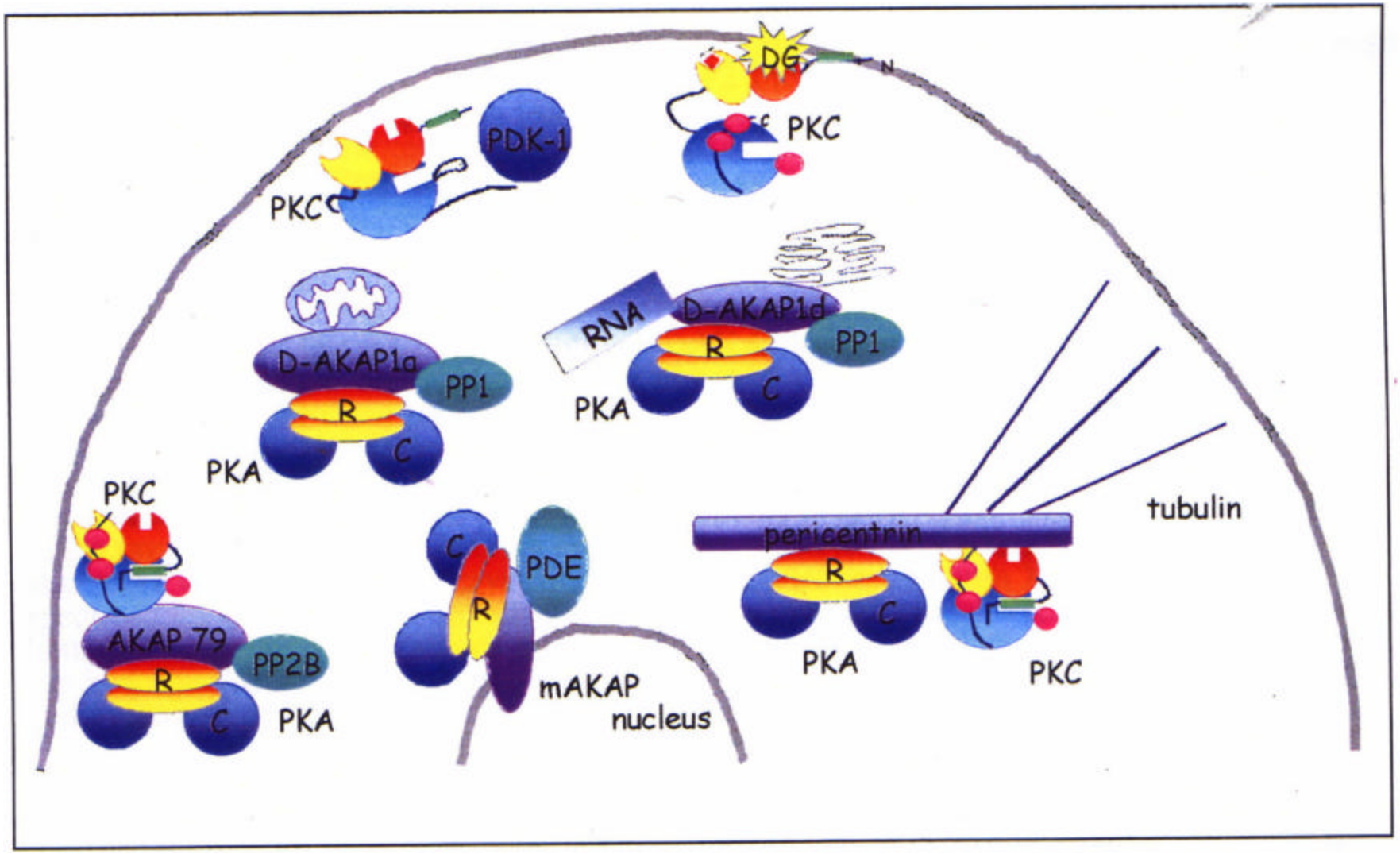
Schema Physical Data Model for Molecule Pages



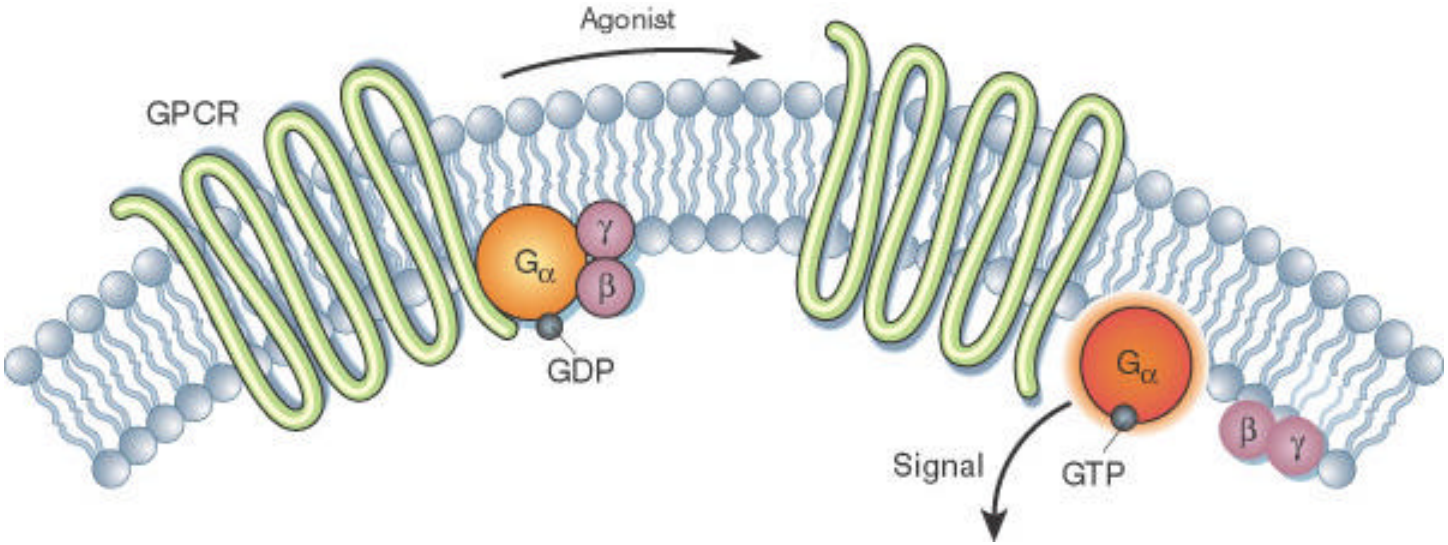
Molecule Page Sections

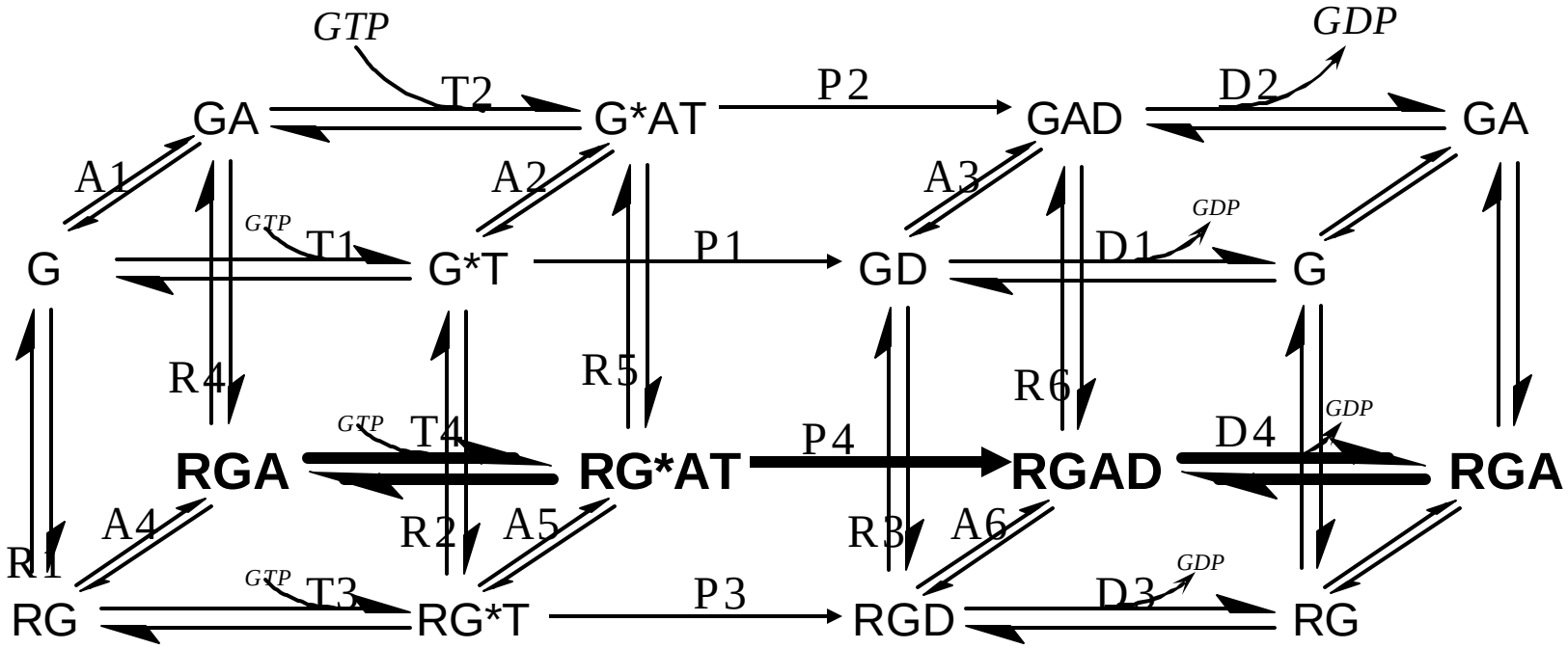


Molecule Pages

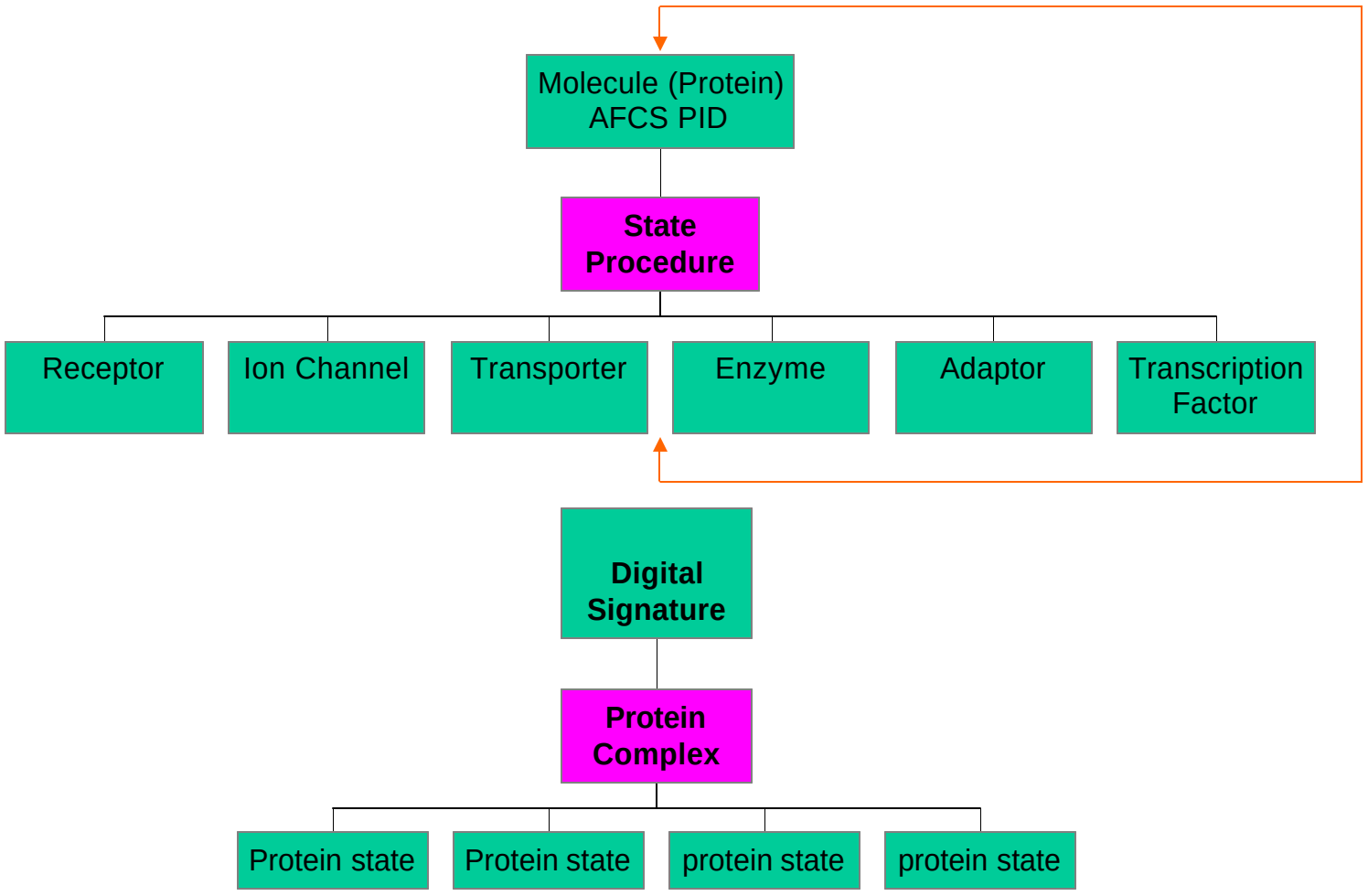


Molecule Pages





Functional State and State Signature



tein A000111
erview
thor Entered Data
1.0, Started 10 Dec
03)
stract
twork Map
ates
ansitions
nctions
tomated Data
cest from 25 Nov 2003)
tabase Links
mains & Motifs
otein Structure
ne Info
thologs & Paralogs
st Data

Adenylyl cyclase type 2

Abstract for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003

Protein Function

All membrane-bound mammalian adenylyl cyclases catalyze the conversion of MgATP to cyclic AMP and pyrophosphate. ACII is the prototype member of the subfamily of adenylyl cyclases that are conditionally activated by G protein beta/gamma subunits.

1312225	Federman AD, Conklin BR, Schrader KA, Reed RR, Bourne HR	Hormonal stimulation of adenylyl cyclase through Gi-protein beta gamma subunits.	Nature, 356, 6365	12 Mar 1992
1719547	Feinstein PG, Schrader KA, Bakalyar HA, Tang WJ, Krupinski J, Gilman AG, Reed RR	Molecular cloning and characterization of a Ca2+/calmodulin-insensitive adenylyl cyclase from rat brain.	Proc Natl Acad Sci U S A, 88, 22	15 Nov 1991
1962211	Tang WJ, Gilman AG	Type-specific regulation of adenylyl cyclase by G protein beta gamma subunits.	Science, 254, 5037	6 Dec 1991
8416978	Taussig R, Quarmby LM, Gilman AG	Regulation of purified type I and type II adenylylcyclases by G protein beta gamma subunits.	J Biol Chem, 268, 1	5 Jan 1993

Regulation of Activity

ACII is stimulated by Gs alpha and forskolin. In the presence of Gs alpha-stimulation, G protein beta/gamma subunits (released from Gi heterotrimers) further stimulate the activity of the enzyme. Gs alpha has been shown to be constitutively palmitoylated at position 2, and reversibly palmitoylated at position 3; there is no reported difference between the singly vs. doubly palmitoylated Gs alpha with respect to the activation of acenylyl cyclases.

9177179	Kleuss C, Gilman AG	Gsalpha contains an unidentified covalent modification that increases its affinity for adenylyl cyclase.	Proc Natl Acad Sci U S A, 94, 12	10 Jun 1997
12574119	Kleuss C, Krause E	Galpha(s) is palmitoylated at the N-terminal glycine.	EMBO J, 22, 4	17 Feb 2003

Interactions with Ligands and Other Proteins

ACII is stimulated by Gs alpha and forskolin. In the presence of Gs alpha-stimulation, G protein beta/gamma subunits (released from Gi

tein A000111
erview
thor Entered Data
1.0, Started 10 Dec
03)
stract
twork Map
ates
ansitions
nctions
tomated Data
cest from 25 Nov 2003)
tabase Links
mains & Motifs
otein Structure
ne Info
thologs & Paralogs
st Data

Adenylyl cyclase type 2

Current state for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003

State summary

State name	(Gnas 2PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR)
State synonym	AC2/Gs2Pa+bg
Computed name	G protein alpha s [Palmitoylation@2, Palmitoylation@3, GTP (guanosine triphosphate)@G_alpha(20-393)]: Adenylyl cyclase type 2 [Glycosylation@unknown]: G protein beta: G protein gamma [Geranylgeranylation@unknown, Methylation@unknown, Proteolysis@unknown]
Cellular compartment	plasma membrane

State constituents

Protein name	Adenylyl cyclase type 2
Covalent modifications	Glycosylation@unknown
Small molecules bound	None
Protein name	G protein alpha s
Covalent modifications	Palmitoylation@2 Palmitoylation@3
Small molecules bound	GTP (guanosine triphosphate)@G_alpha(20-393)
Protein name	G protein beta
Covalent modifications	None
Small molecules bound	None
Protein name	G protein gamma
Covalent modifications	Geranylgeranylation@unknown Methylation@unknown Proteolysis@unknown

tein A000111
erview
thor Entered Data
1.0, Started 10 Dec
03)
stract
twork Map
ates
ansitions
nctions
tomated Data
cest from 25 Nov 2003)
tabase Links
mains & Motifs
otein Structure
ne Info
thologs & Paralogs
st Data

Adenylyl cyclase type 2

State Transitions for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003

Transitions for Adenylyl cyclase type 2

(Gnas PA L82) (Adcy2 G) --> (Adcy2 G)
(Gnas PA L82) (Adcy2 G) --> (Gnas PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR)
(Gnas PA L82) (Adcy2 G) --> (Gnas PA L82) (Adcy2 G L760)[2]
(Adcy2 G) --> (Gnas PA L82) (Adcy2 G)
(Adcy2 G) --> (Gnas 2PA L82) (Adcy2 G)
(Adcy2 G) --> (Adcy2 G 2P)
(Adcy2 G) --> (Adcy2 G L760)[2]
(Gnas 2PA L82) (Adcy2 G) --> (Adcy2 G)
(Gnas 2PA L82) (Adcy2 G) --> (Gnas 2PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR)
(Gnas 2PA L82) (Adcy2 G) --> (Gnas 2PA L82) (Adcy2 G L760)[2]
(Adcy2 G 2P) --> (Adcy2 G)
(Adcy2 G 2P) --> (Gnas PA L82) (Adcy2 G 2P)
(Adcy2 G 2P) --> (Gnas 2PA L82) (Adcy2 G 2P)
(Adcy2 G 2P) --> (Adcy2 G 2P L760)
(Gnas PA L82) (Adcy2 G 2P) --> (Adcy2 G 2P)
(Gnas PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR) --> (Gnas PA L82) (Adcy2 G)
(Gnas 2PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR) --> (Gnas 2PA L82) (Adcy2 G)
(Adcy2 G L760)[2] --> (Adcy2 G)
(Adcy2 G L760)[2] --> (Gnas PA L82) (Adcy2 G L760)[2]

Protein A000111
Overview
Author Entered Data
1.0, Started 10 Dec 03)
Abstract
Network Map
States
Transitions
Functions
Automated Data
Latest from 25 Nov 2003)
Database Links
Domains & Motifs
Protein Structure
Gene Info
Orthologs & Paralogs
Cost Data

Adenylyl cyclase type 2

State Transition for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003

Detailed information for this transition
(Gnas PA L82) (Adcy2 G) -->> (Gnas PA L82) (Adcy2 G L760)[2]

Ligand Association

Inital state	
State name	(Gnas PA L82) (Adcy2 G)
Cellular Localization	plasma membrane
Protein name	Adenylyl cyclase type 2
Covalent modifications	Glycosylation@unknown
Small molecules bound	None
Protein name	G protein alpha s
Covalent modifications	Palmitoylation@2
Small molecules bound	GTP (guanosine triphosphate)@G_alpha(20-393)

Final state	
State name	(Gnas PA L82) (Adcy2 G L760)[2]
Cellular Localization	plasma membrane
Protein name	Adenylyl cyclase type 2
Covalent modifications	Glycosylation@unknown
Small molecules bound	forskolin@unknown
Protein name	G protein alpha s
Covalent modifications	Palmitoylation@2

tein A000111
erview
thor Entered Data
1.0, Started 10 Dec
03)
stract
twork Map
ates
ansitions
nctions
tomated Data
cest from 25 Nov 2003)
tabase Links
mains & Motifs
otein Structure
ne Info
thologs & Paralogs
st Data

Adenylyl cyclase type 2

Transition Data for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003

Enzyme

Reaction catalyzed	States that catalyze this reaction
ATP (adenosine triphosphate) -> cAMP (cyclic AMP) + pyrophosphate	(Adcy2 G)
	(Gnas PA L82) (Adcy2 G)
	(Gnas 2PA L82) (Adcy2 G)
	(Adcy2 G 2P)
	(Gnas PA L82) (Adcy2 G 2P)
	(Gnas PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR)
	(Gnas 2PA L82) (Adcy2 G) (G protein beta*) (G protein gamma* GG ME PR)
	(Gnas PA L82) (Adcy2 G L760)[2]
	(Gnas 2PA L82) (Adcy2 G 2P)
	(Gnas 2PA L82) (Adcy2 G L760)[2]
	(Adcy2 G 2P L760)
	(Adcy2 G L760)[2]

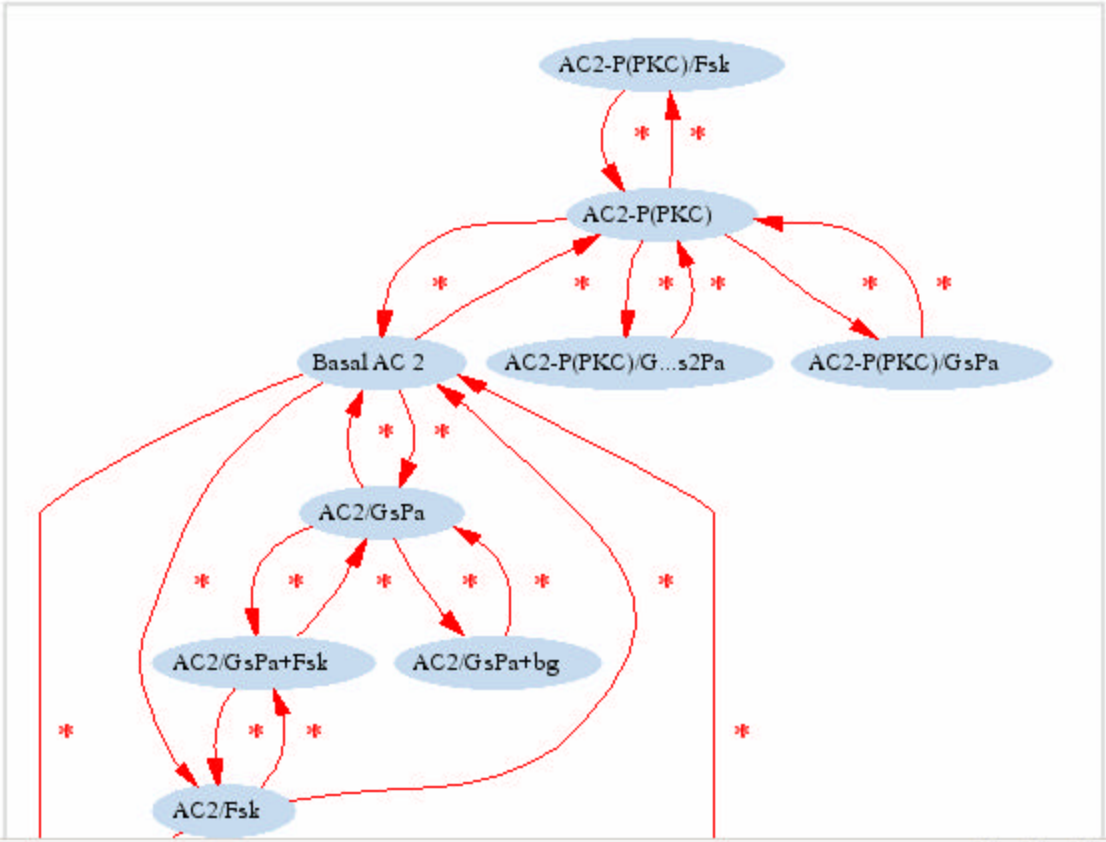
Add new functional data

No states have been created.

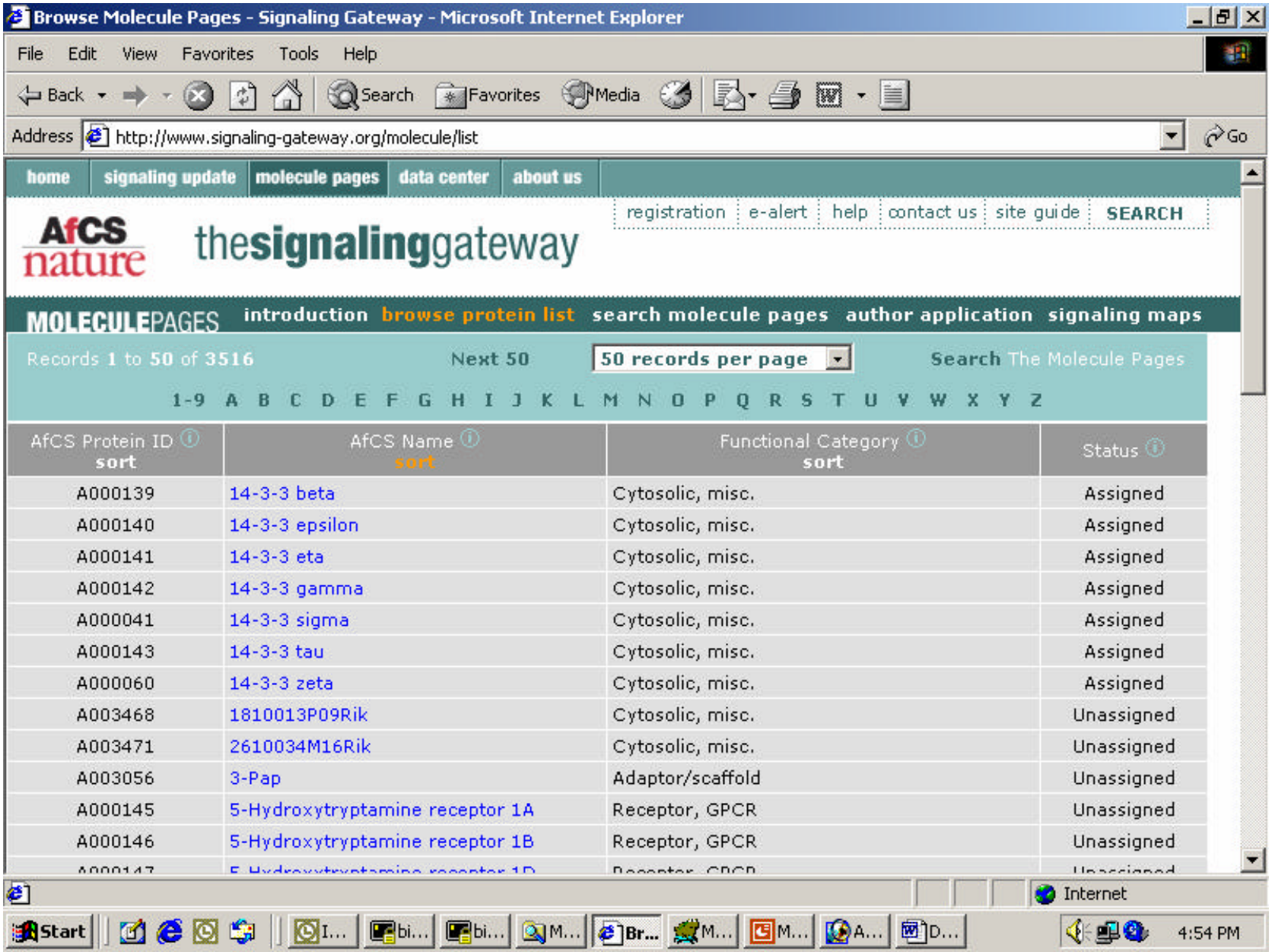
Protein A000111
Overview
Author Entered Data
1.0, Started 10 Dec 03)
Abstract
Network Map
States
Transitions
Functions
Automated Data
Latest from 25 Nov 2003)
Database Links
Domains & Motifs
Protein Structure
Gene Info
Orthologs & Paralogs
Cost Data

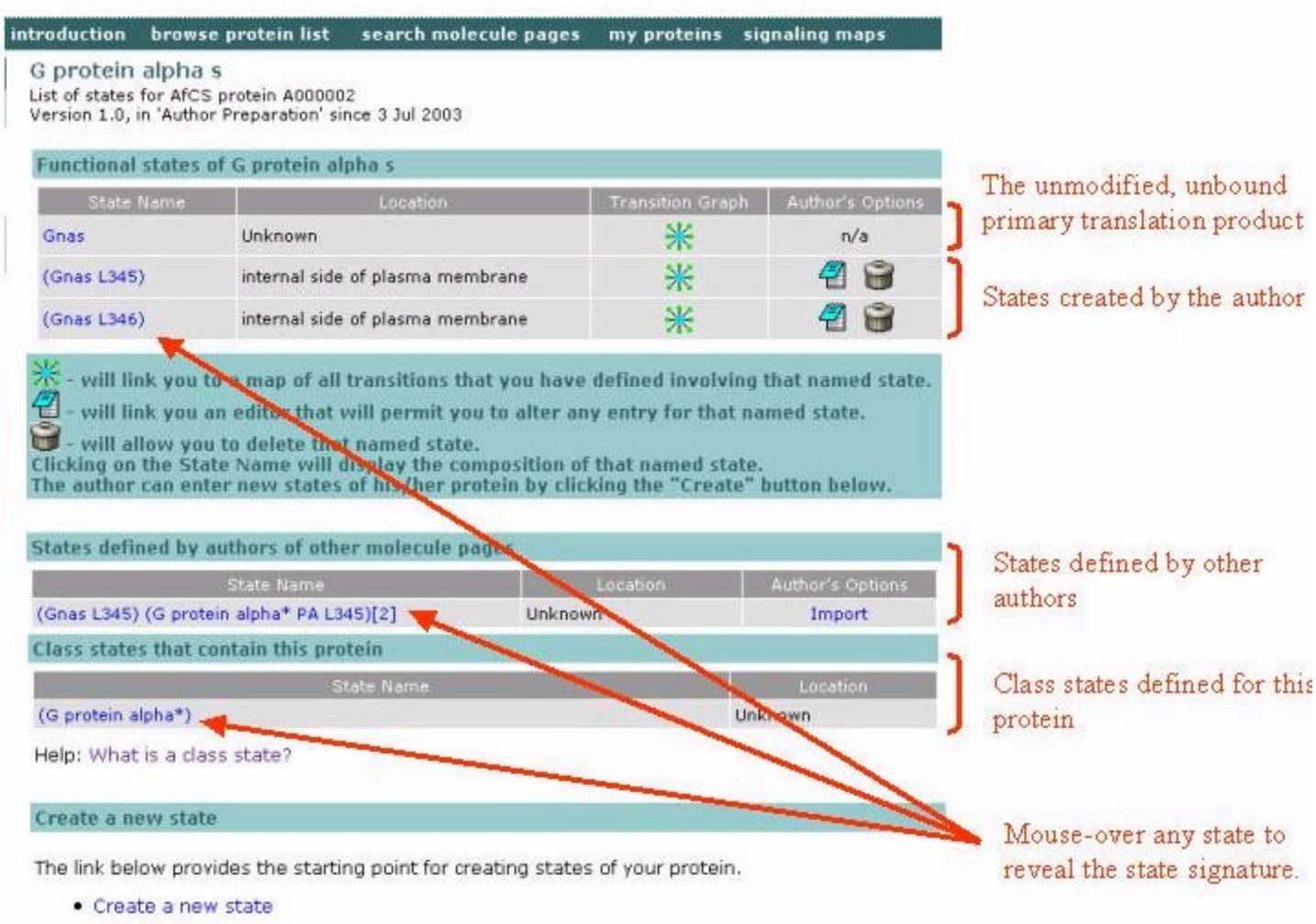
Adenylyl cyclase type 2

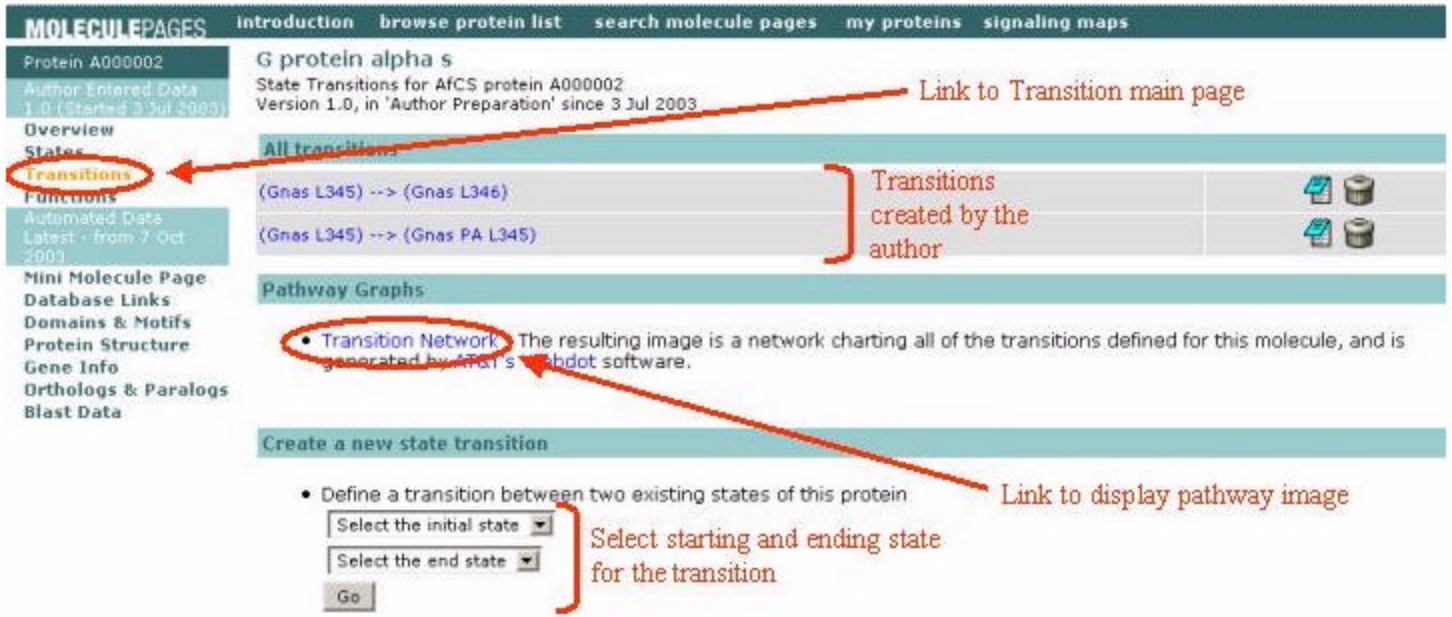
Transition network for AfCS protein A000111
Version 1.0, in 'NPG Editing' since 10 Dec 2003



MP – Protein List



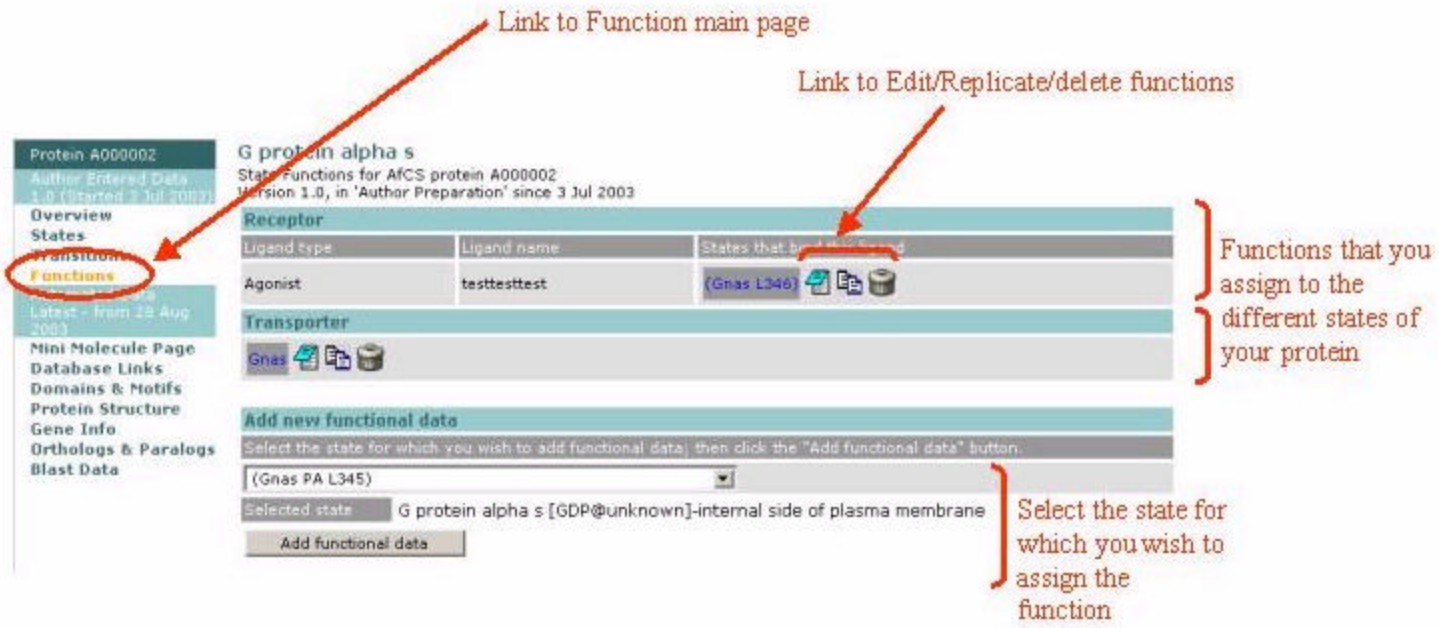

introduction
browse protein list
search molecule pages
my proteins
signaling maps
G protein alpha s
List of states for AfCS protein A000002
Version 1.0, in 'Author Preparation' since 3 Jul 2003
Functional states of G protein alpha s
State Name
Location
Transition Graph
Author's Options
Gnas
Unknown
n/a
(Gnas L345)
internal side of plasma membrane
(Gnas L346)
internal side of plasma membrane
- will link you to a map of all transitions that you have defined involving that named state.
- will link you an editor that will permit you to alter any entry for that named state.
- will allow you to delete that named state.
Clicking on the State Name will display the composition of that named state.
The author can enter new states of his/her protein by clicking the "Create" button below.
States defined by authors of other molecule pages
State Name
Location
Author's Options
(Gnas L345) (G protein alpha* PA L345)[2]
Unknown
Import
Class states that contain this protein
State Name
Location
(G protein alpha*)
Unknown
Help: What is a class state?
Create a new state
The link below provides the starting point for creating states of your protein.
Create a new state
The unmodified, unbound primary translation product
States created by the author
States defined by other authors
Class states defined for this protein
Mouse-over any state to reveal the state signature.



c) Instructions for Creating Transitions.

Transitions are created in the following order:

- *Select the starting and ending states.*
- *Select the appropriate process for the transition.* These can be one of the following:
    - Protein association
    - Ligand association
    - Addition of covalent modification
    - Protein dissociation
    - Ligand dissociation
    - Removal of covalent modification
    - Change of cellular location
    - Intrinsic enzymatic activity
- When applicable, enter the protein that catalyzes this transition. If multiple proteins can catalyze the same



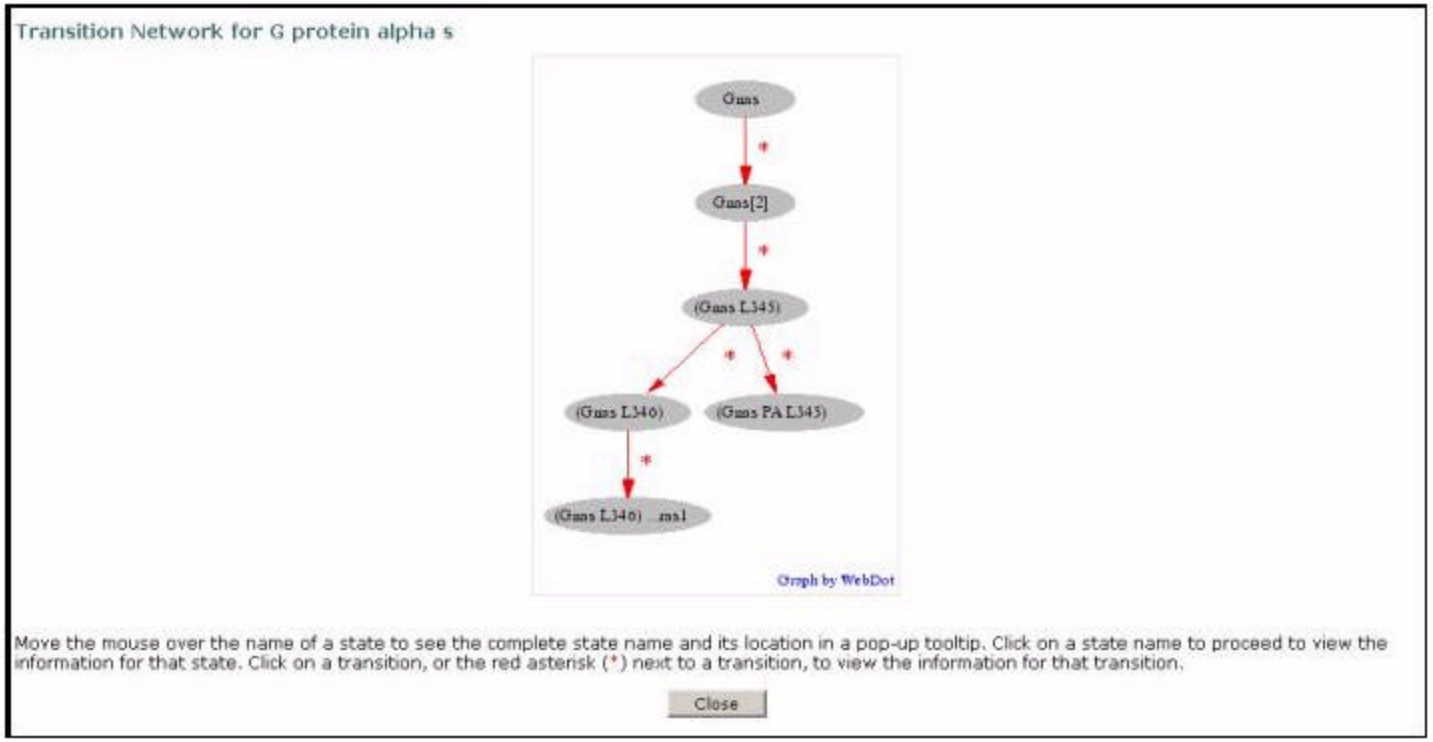
b) Instructions for Entering State Functions.

State functions are entered in the following order:

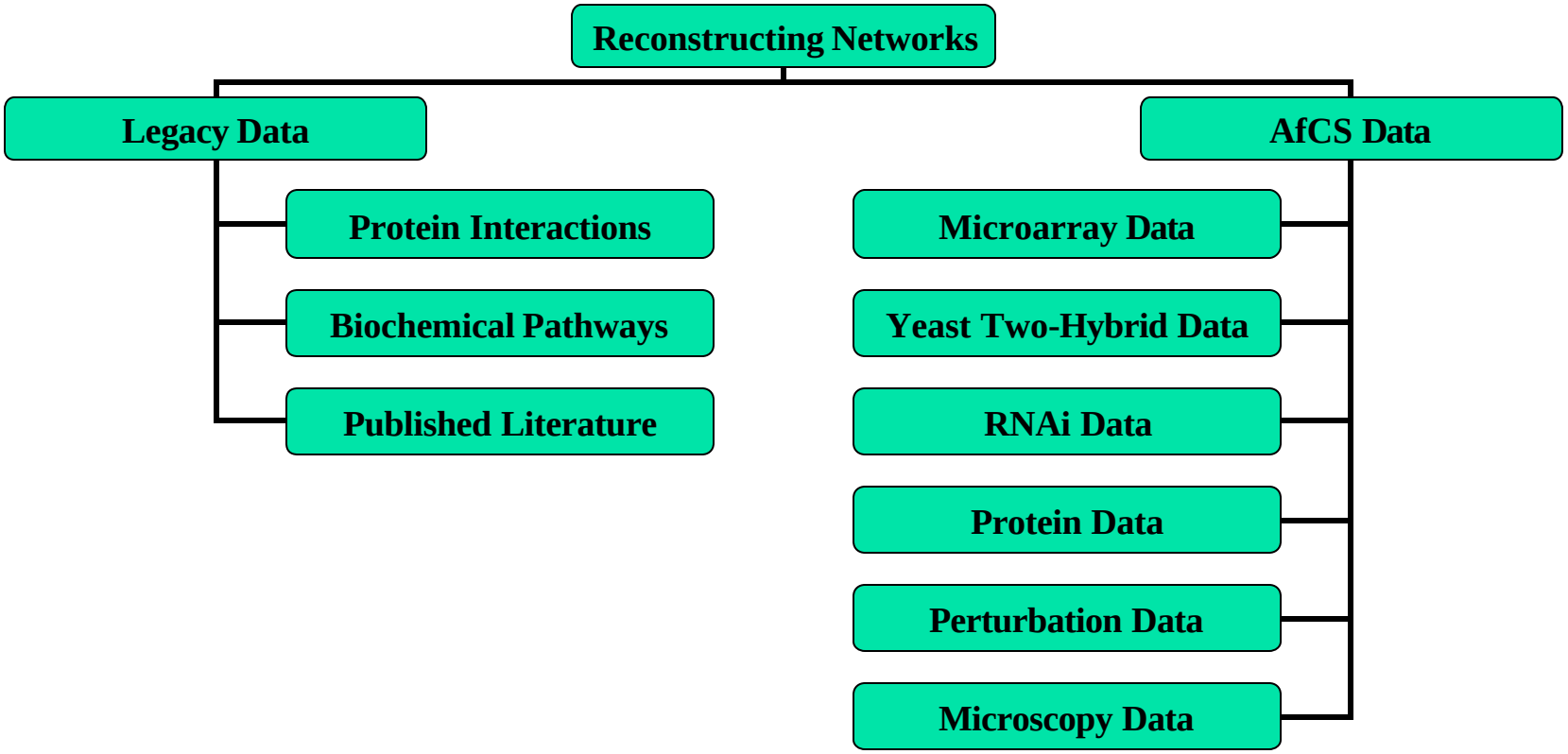
- Select a protein state.
- Select the appropriate function to be ascribed to this state. This can be one of the following (the relevant parameters are entered subsequently and are listed in parentheses):
    - Receptor (Type of ligand - agonist/antagonist/*etc.*, Name of ligand, Binding domain, Stoichiometry, Kd)
    - Enzyme (Substrates, Products, Kd, Vmax, etc.)
    - Channel (Pore selectivity, Conductance, Open and closed times, Activation/inactivation gating, Blockers, etc.)
    - Transporter (Substrate, Direction of transport, Km, Energy requirement, etc.)
    - Transcription Factor (Canonical DNA binding sequence, regulated target genes, etc.). DNA sequences

- *Viewing Transitions*.

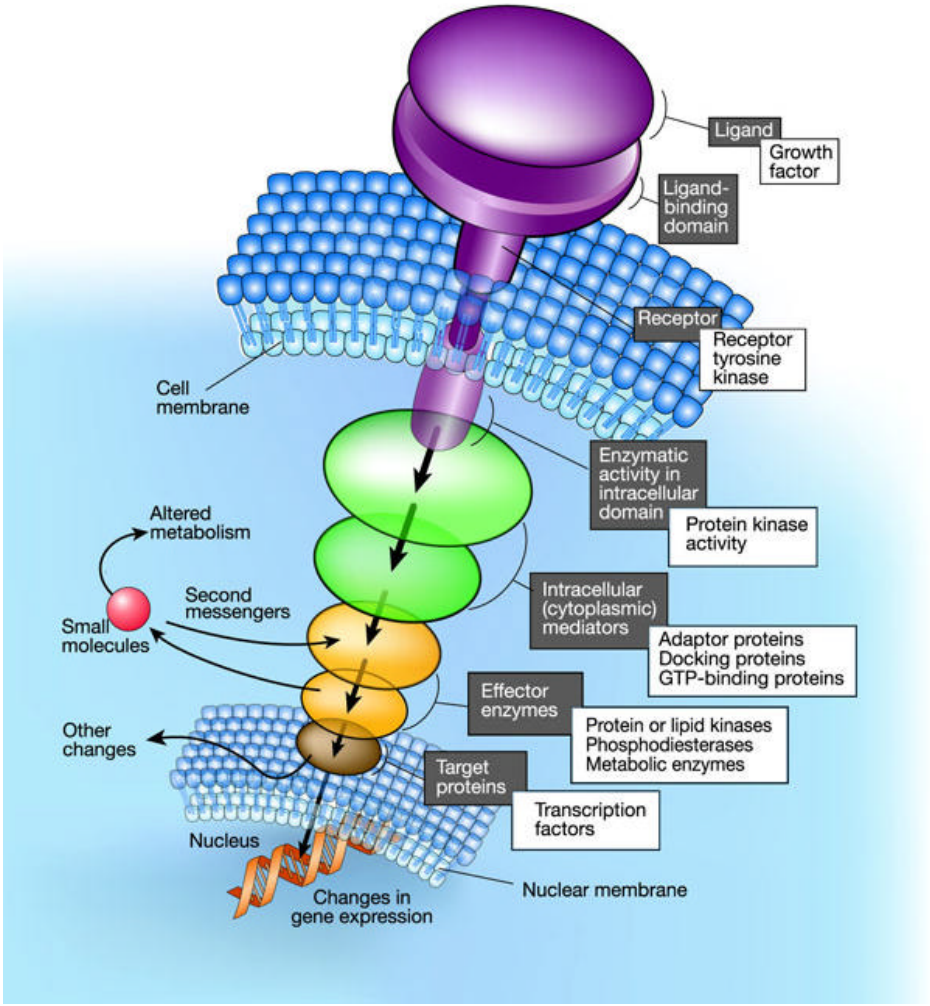
  You can view information that was entered for each transition from two starting points. First, transitions from each of the states can be viewed from the state page by clicking on the relevant \uFFFD. Second, from the Transition page, clicking on Transition Network (located in the Pathway Graphs subsection) will produce a pathway view of all of your transitions.



Reconstructing Networks



Signal Transduction in a Cell

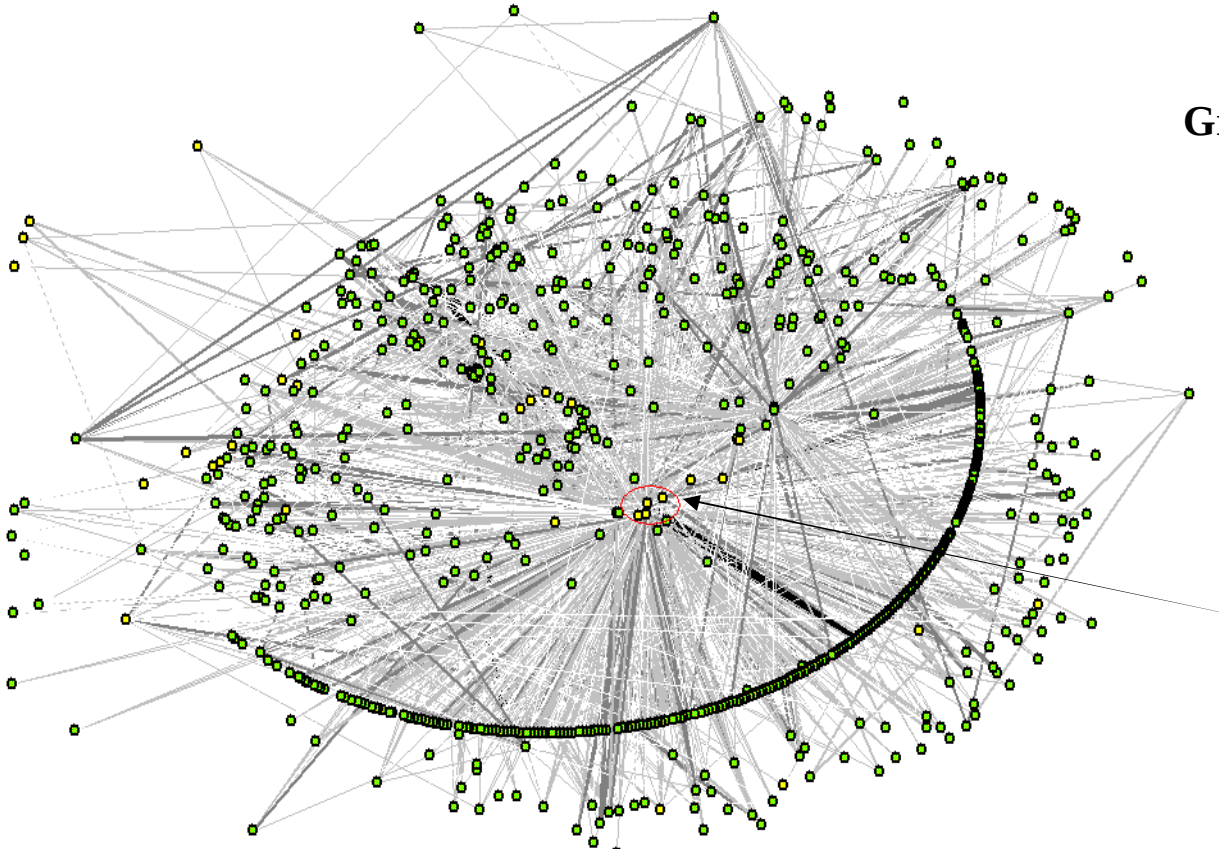


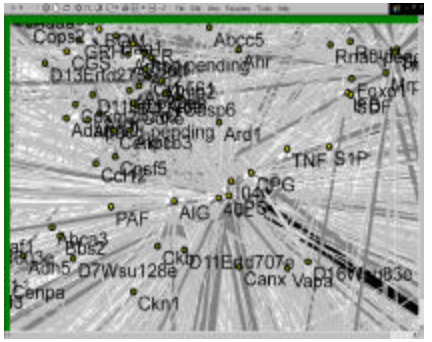
from Downward, **Nature**, August (2001)

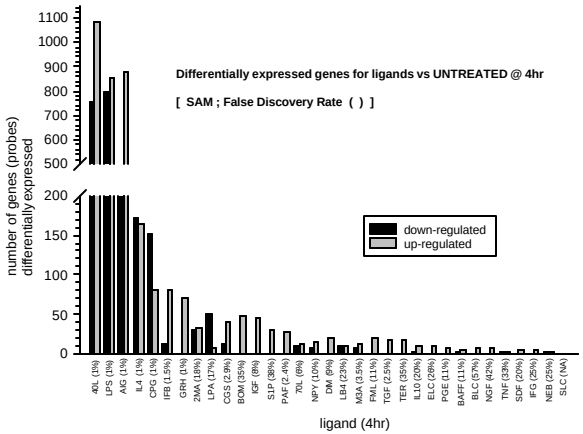
Ligand Screen Transcript Analysis

- B cell samples prepared by Cell Lab (Dallas).
- Cultured for different time periods (.5, 1, 2, and 4 hr) in the presence or absence of ligands before harvesting for total RNA isolation.
- Treated and untreated time-course samples hybridized against a spleen reference.
- After removing the common spleen denominator, comparison to 0 time point data reflects the changes in mRNA levels due to ligand treatment and/or time in culture.
- One of the largest mammalian array sets (33 ligands).
- All of the experiments were done in triplicate. Including in controls >450 arrays (Caltech)

Graph association map (4hr)







The mitogenic response from the ligands AIG, 40L, I04, LPS, CPG dominate at the center of the plot. This is too dense for a clear view (see histogram to the left).

IFβ, GRH, CGS, PAF, TGF, M3A, 2MA also showed a significant gene response.

Similarity measures between genes under different conditions with respect to expression levels for…

… groups of genes Þ **clustering methods**

… pairs of genes Þ **correlation methods**

Linear correlation

$$\frac{\Sigma (x - x_{mean})(y - y_{mean})}{[\Sigma (x - x_{mean})^2 \Sigma (y - y_{mean})^2]^{½}} = r^2_{xy}$$

Partial correlation

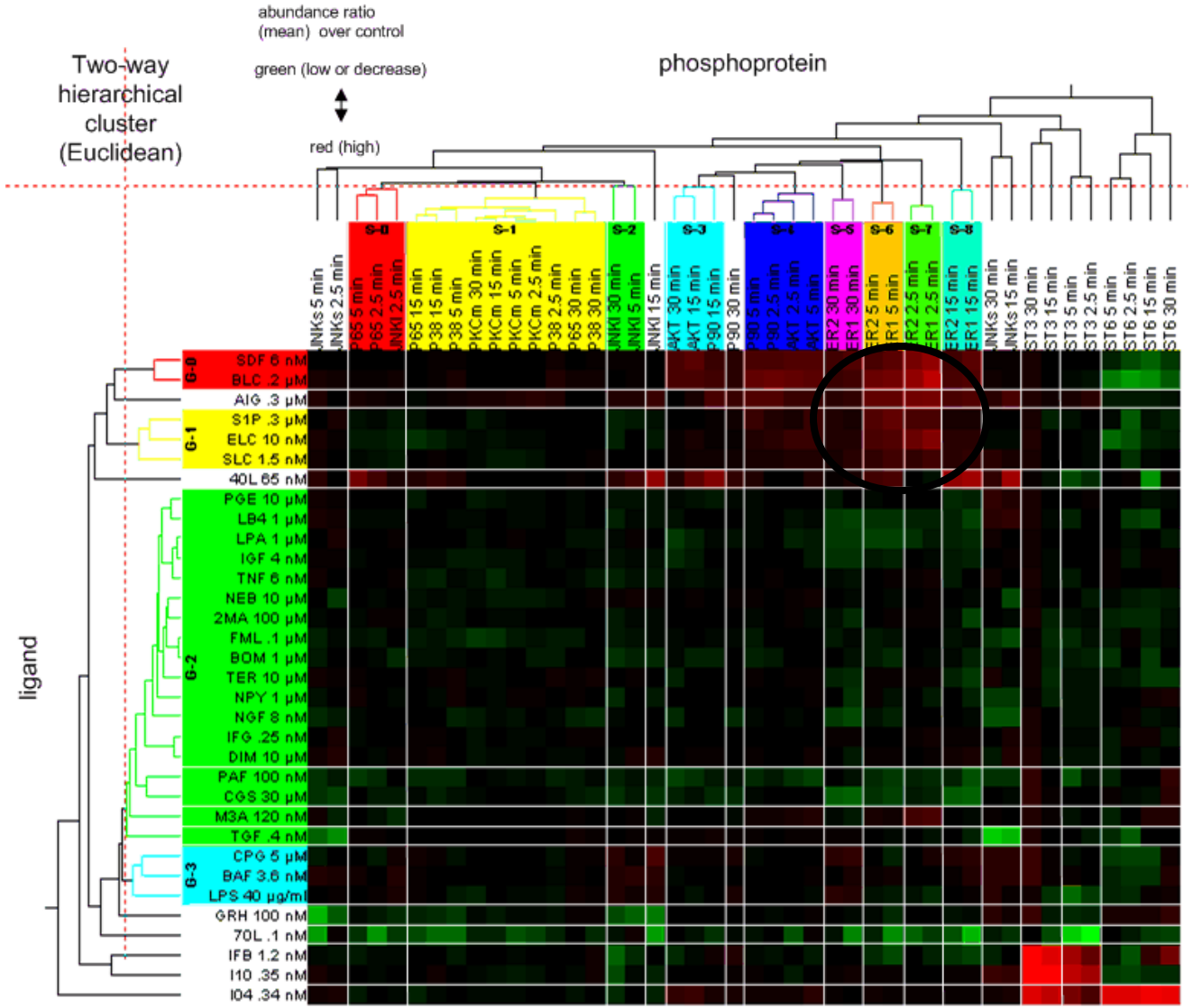
$$\frac{r_{xy} - r_{xz} r_{yz}}{[(1 - r^2_{xz})(1 - r^2_{yz})]^{½}} = r_{xy.z}$$

“marginal” global correlation (for ligand j)

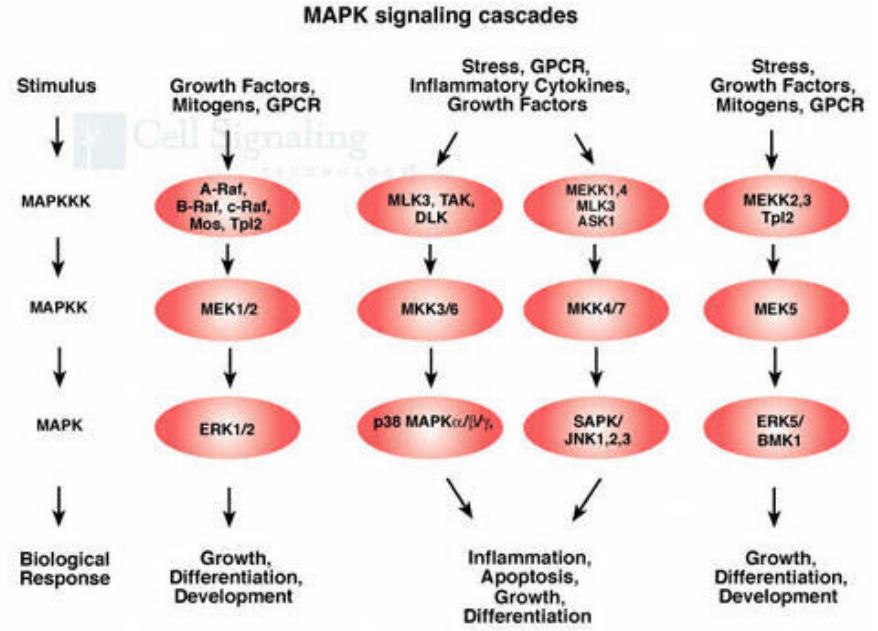
$$r^2_{\text{all xy}} - r^2_{\text{all xy except ligand j}}$$

Two-way hierarchical cluster:

mean ratio (vs control) of phosphoprotein levels and ligand



Several ligands that elicit an ERK response (chemokines + AIG, CD40L) clustered together.



Three main pathways of MAPK and their respective target genes and transcription factors.

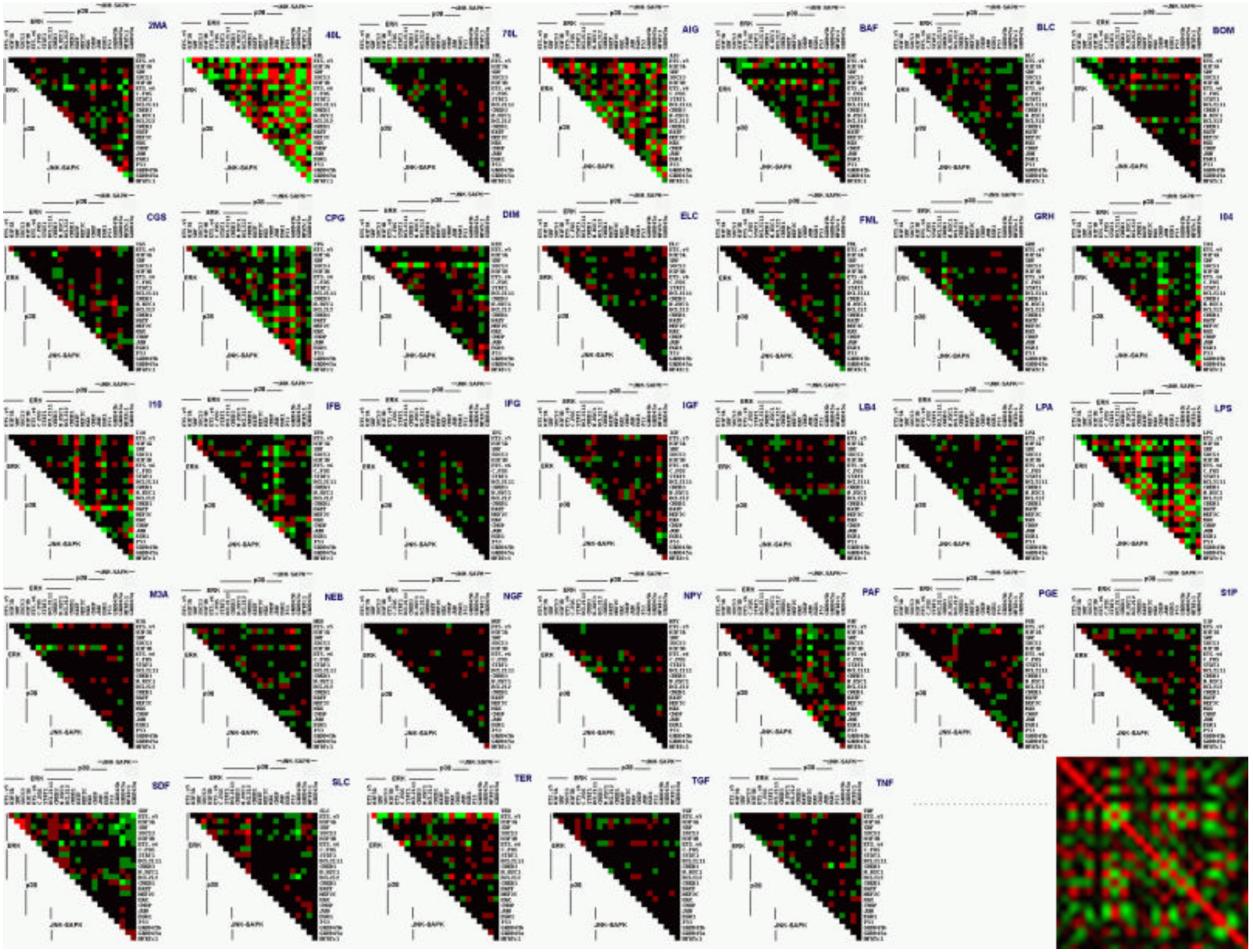
ERK-MAPK	p38	JNK-SAPK
ETS.v6	N.MYC1	NFATC1
H3F3A	MEF2C	Gadd45a
CREB1	CHOP	Gadd45b
C.FOS	Max	Gadd45g
H3F3B	Bcl2l11	Egr1
Socs3	Bcl2l2	CHOP
CREB3	JUN	Max
STAT1	Egr1	JUN
N.MYC1	STAT1	C.FOS
Bcl2l11		P53
Bcl2l2		
SRF		
ETS.v5		

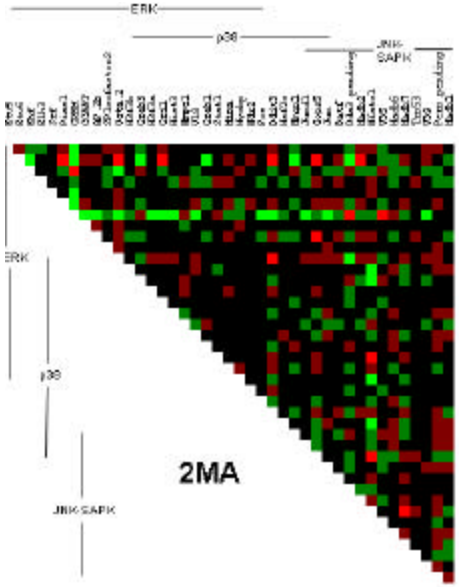
Diagrams are from …

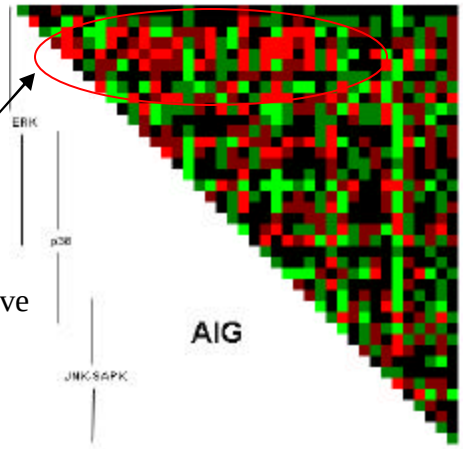
"Mitogen-Activated Protein Kinase Pathways Mediated by ERK, JNK, and p38 Protein Kinases"

G. L. Johnson and R. Lapadat Science 2002 December 6; 298: 1911-1912. (in Review)

Level plots “Marginal” correlation of genes in MAPK pathways







Highly responsive genes from MAPK-ERK pathway

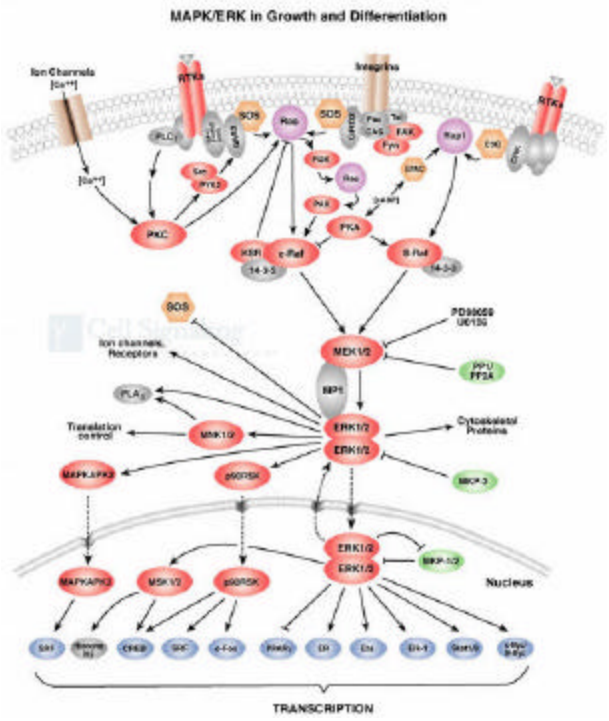
"marginal" global correlation (for ligand j)

difference in correlation =

$r^2_{\text{all xy}} - r^2_{\text{all xy except ligand j}}$

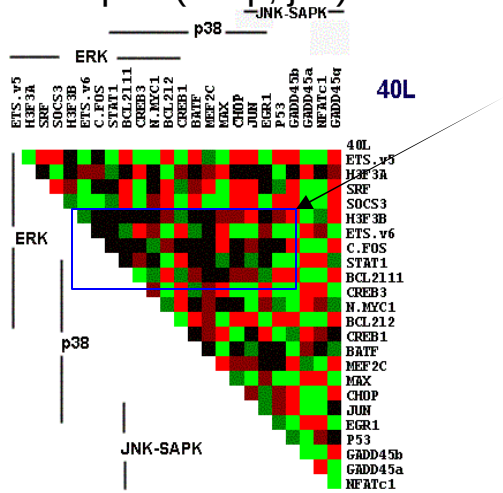
Green indicates negative influence on the gene upon removing ligand j

Red indicates positive influence on the gene upon removing ligand j

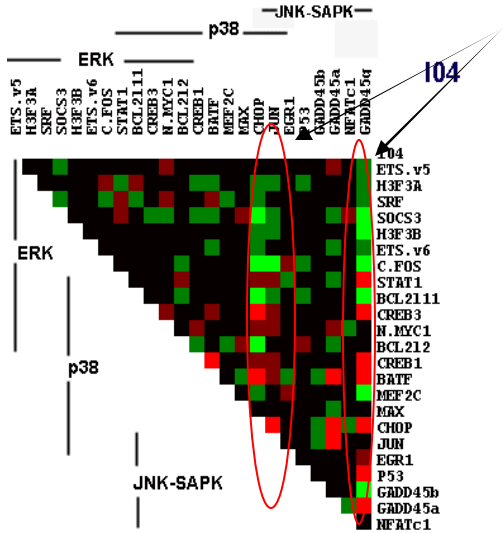


B cells respond to AIG through the MAPK-ERK pathway.

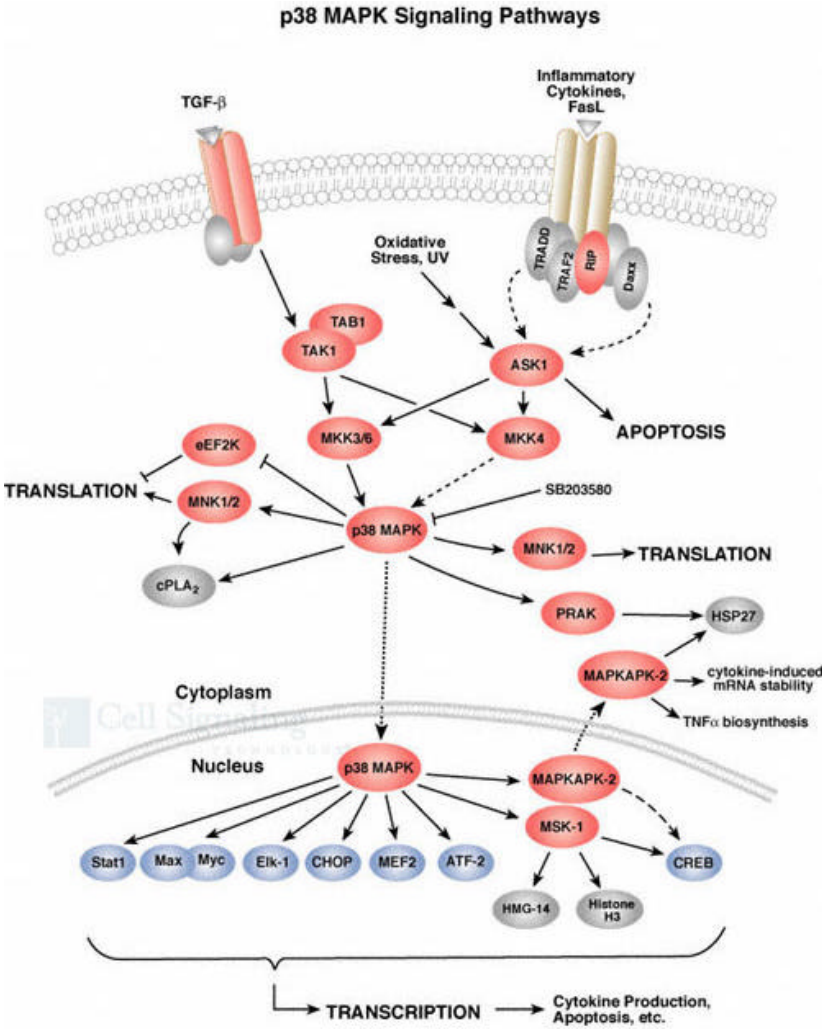
We see the correlation results of removing ligands CD40L (40L) and interleukin 4 (I04) separately from the pool of 33 ligands. The colors red and green refer to decreases/increases in the subsequent correlation similarity matrix respectively. The absolute differential effects are almost uniform across CD40L (with a slightly smaller marginal difference from the ERK related genes h3f3b, ets-v6,c-fos), in contrast to interleukin 4 which shows darker shades, with the color black showing no differences, except for a few p38 (chop, jun) and JNK-SAPK (gadd45q) related genes.



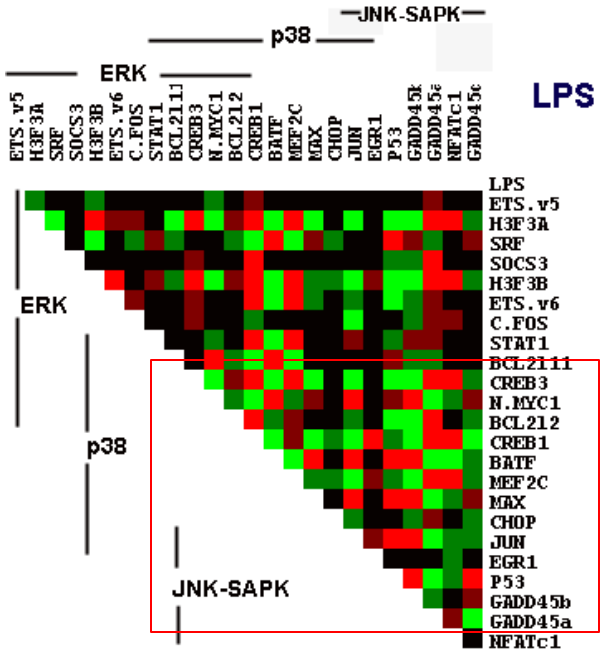
lesser effect in ERK pathway than AIG

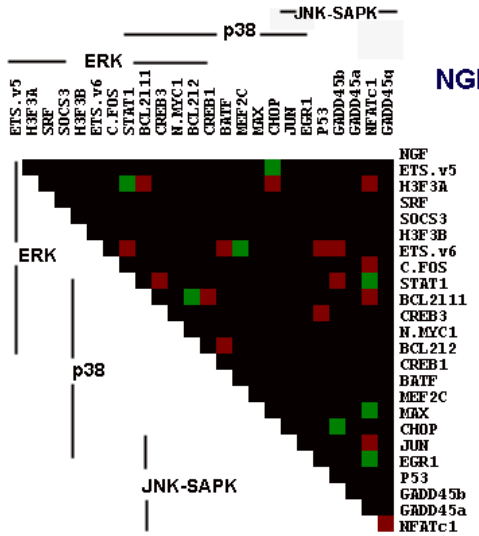


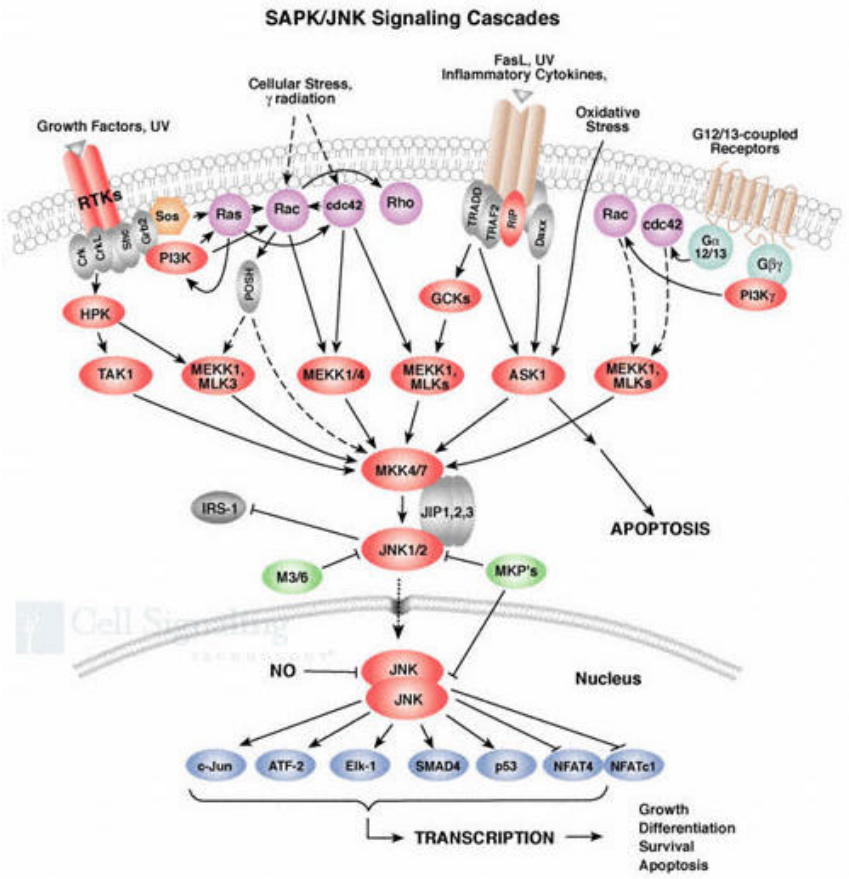
cytokine stress-related genes



B cells do not show any response to NGF but respond to LPS. Note: LPS has more response genes in p38 & JNK-SAPK than ERK.

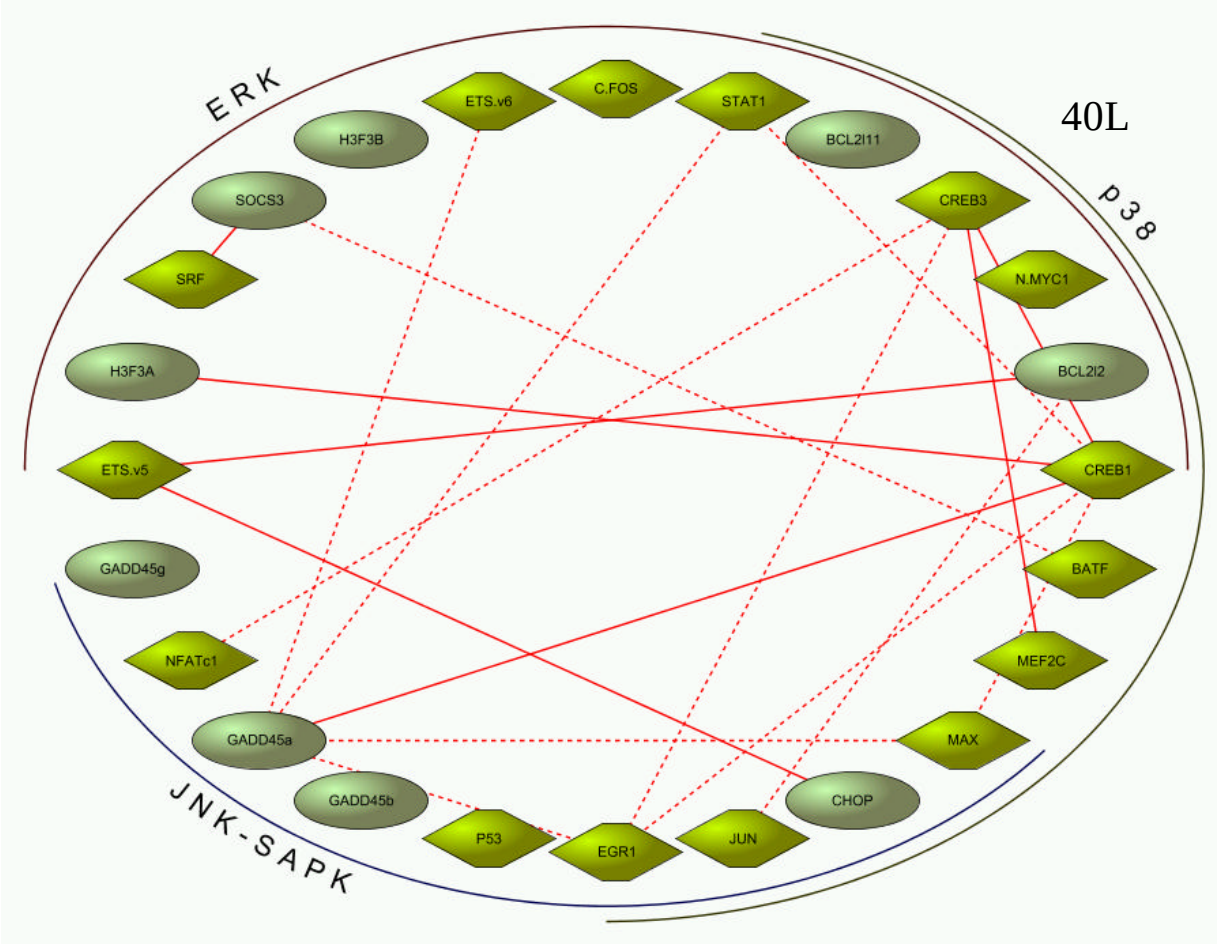




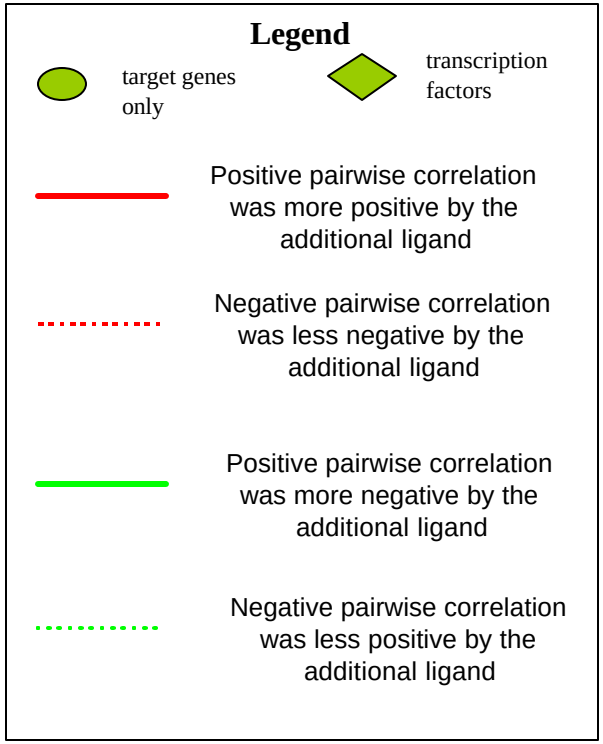


No marginal changes in the pairwise gene correlations in the MAPK pathways from the addition or subtraction of this ligand NGF.

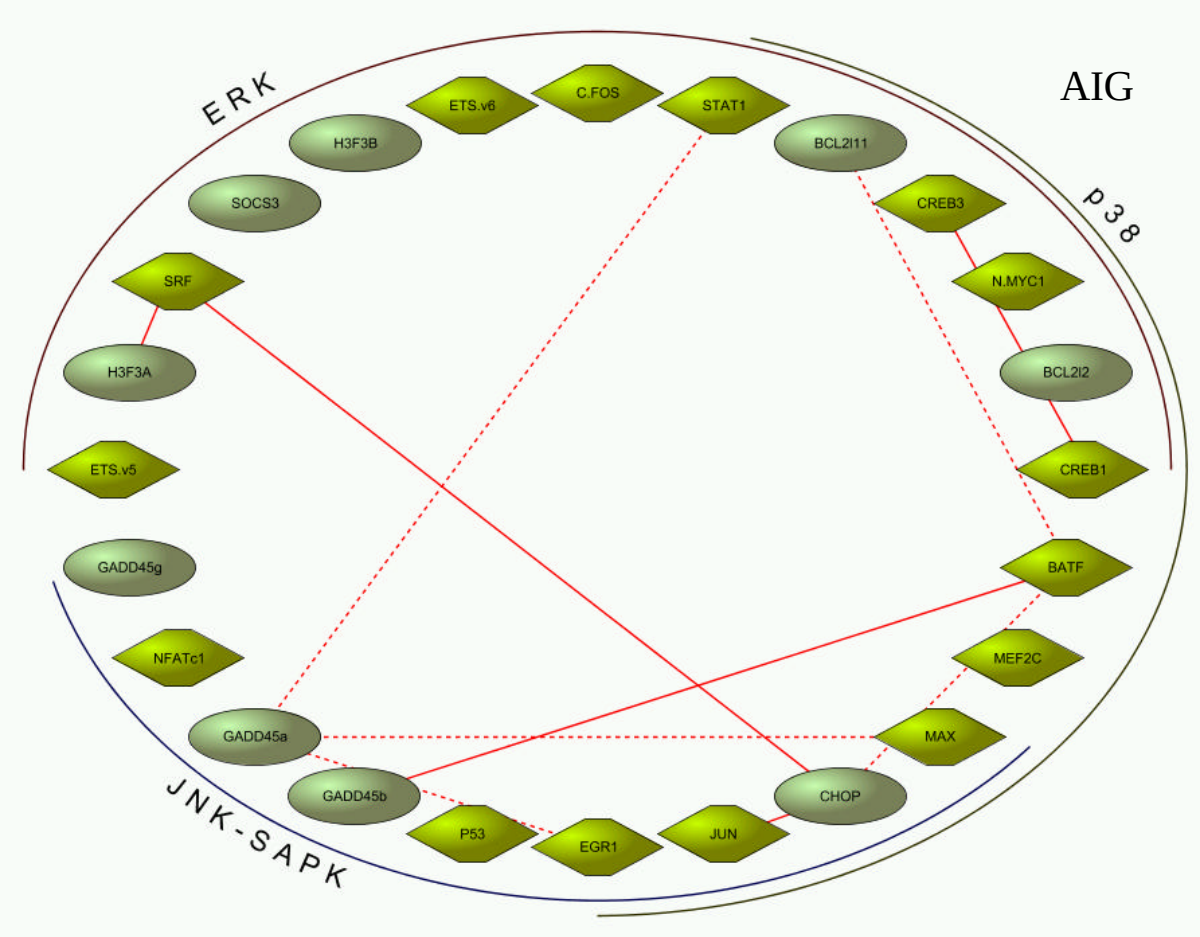
Marginal Correlations Connection Maps for MAPK Pathways



This shows the marginal changes [eg edge threshold Δ=0.1] in the significant pairwise correlation [95% confidence interval for the Fisher transformed distribution] between genes after the addition of the four timepoints of a particular ligand [40L] to the low, intermediate-response ligands (n =112, 28 ligands).



Marginal Correlations Connection Maps for MAPK Pathways



This shows the marginal changes [eg edge threshold Δ=0.1] in the significant pairwise correlation [95% confidence interval for the Fisher transformed distribution] between genes after the addition of the four timepoints of a particular ligand [AIG] to the low, intermediate-response ligands (n =112, 28 ligands).

Legend

target genes only

transcription factors

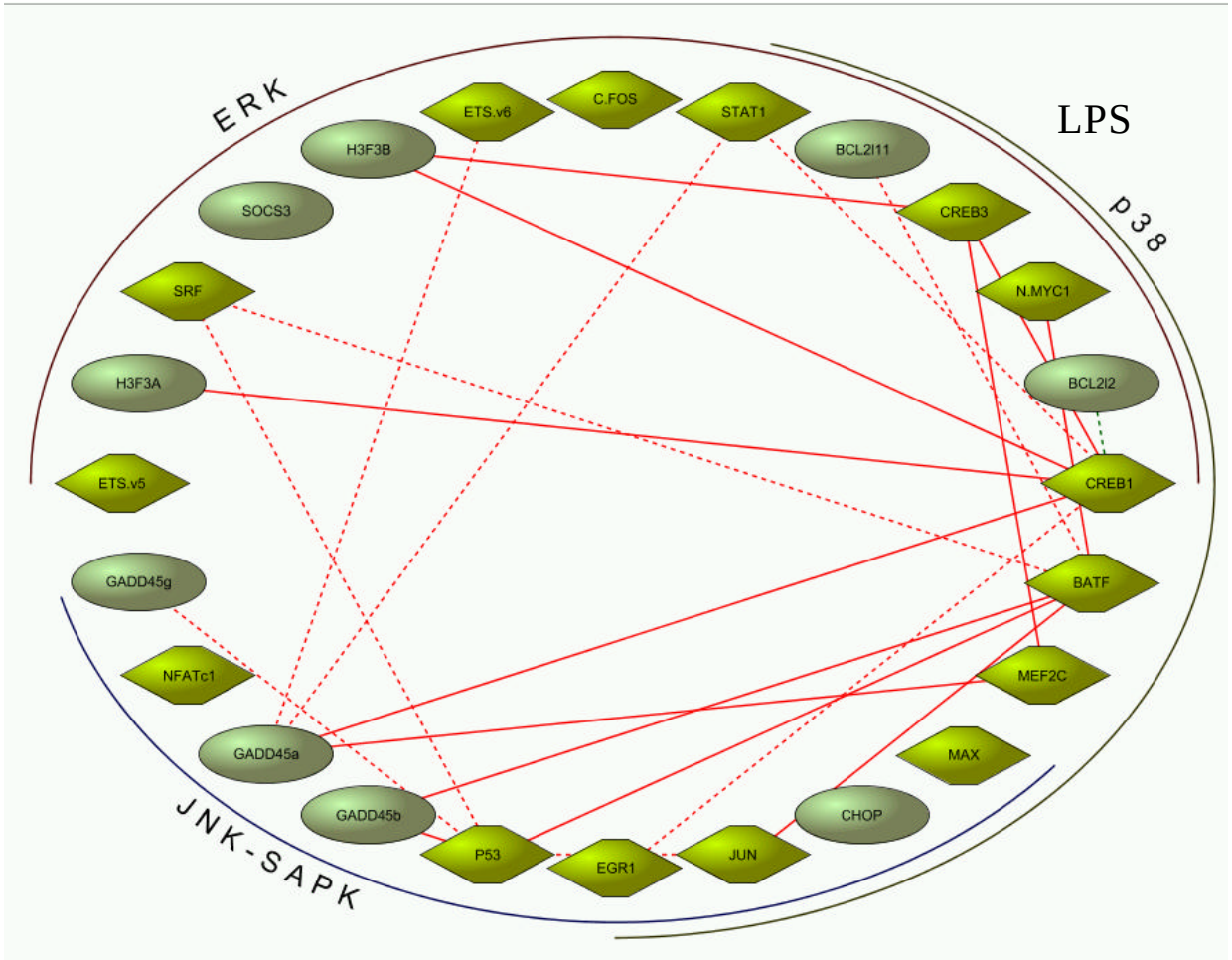
Positive pairwise correlation was more positive by the additional ligand

Negative pairwise correlation was less negative by the additional ligand

Positive pairwise correlation was more negative by the additional ligand

Negative pairwise correlation was less positive by the additional ligand

Marginal Correlations Connection Maps for MAPK Pathways



This shows the marginal changes [eg edge threshold Δ=0.1] in the significant pairwise correlation [95% confidence interval for the Fisher transformed distribution] between genes after the addition of the four timepoints of a particular ligand [LPS] to the low, intermediate-response ligands (n =112, 28 ligands).

Legend

target genes only

transcription factors

Positive pairwise correlation was more positive by the additional ligand

Negative pairwise correlation was less negative by the additional ligand

Positive pairwise correlation was more negative by the additional ligand

Negative pairwise correlation was less positive by the additional ligand

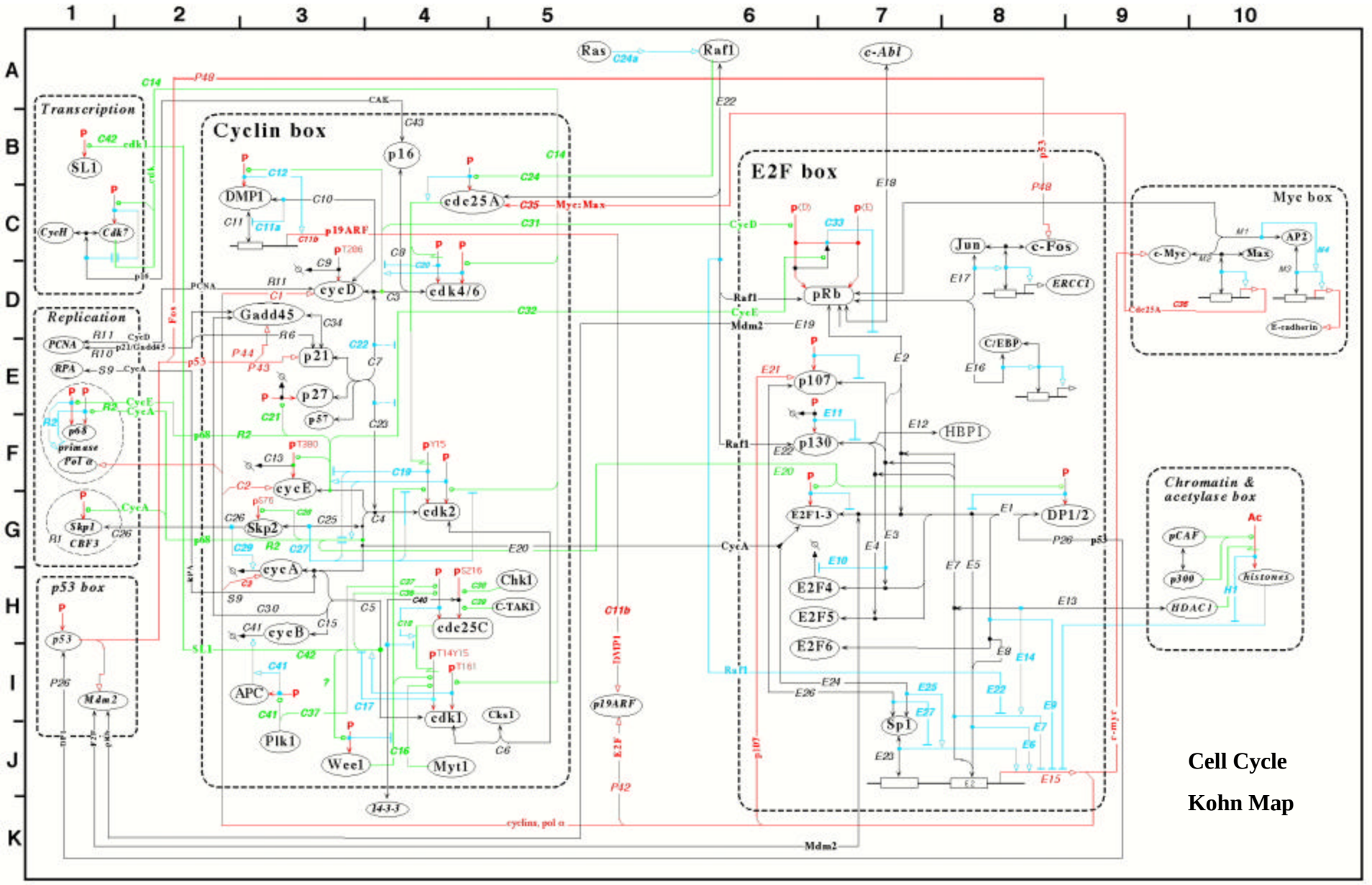
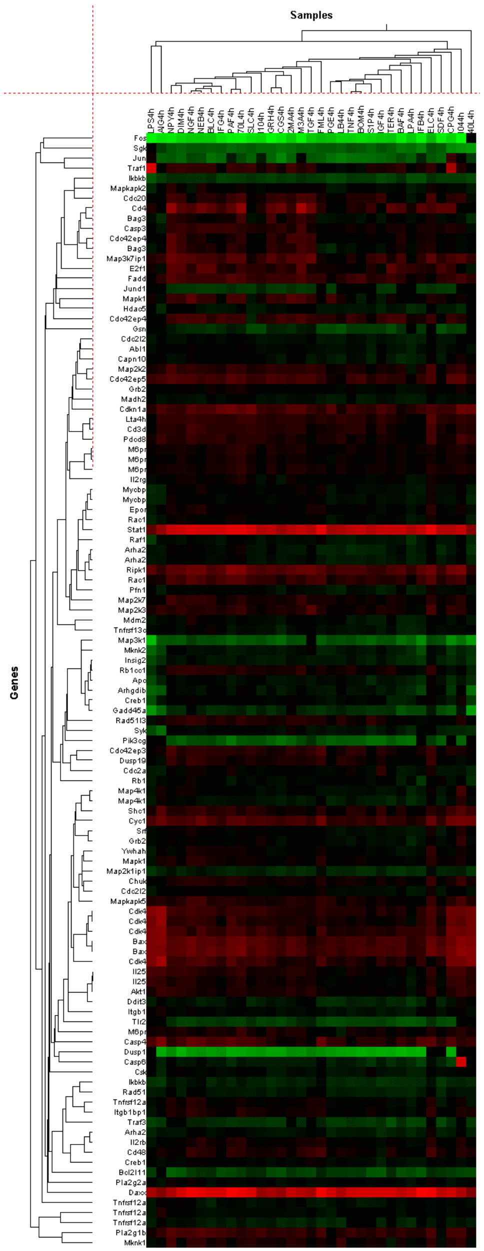
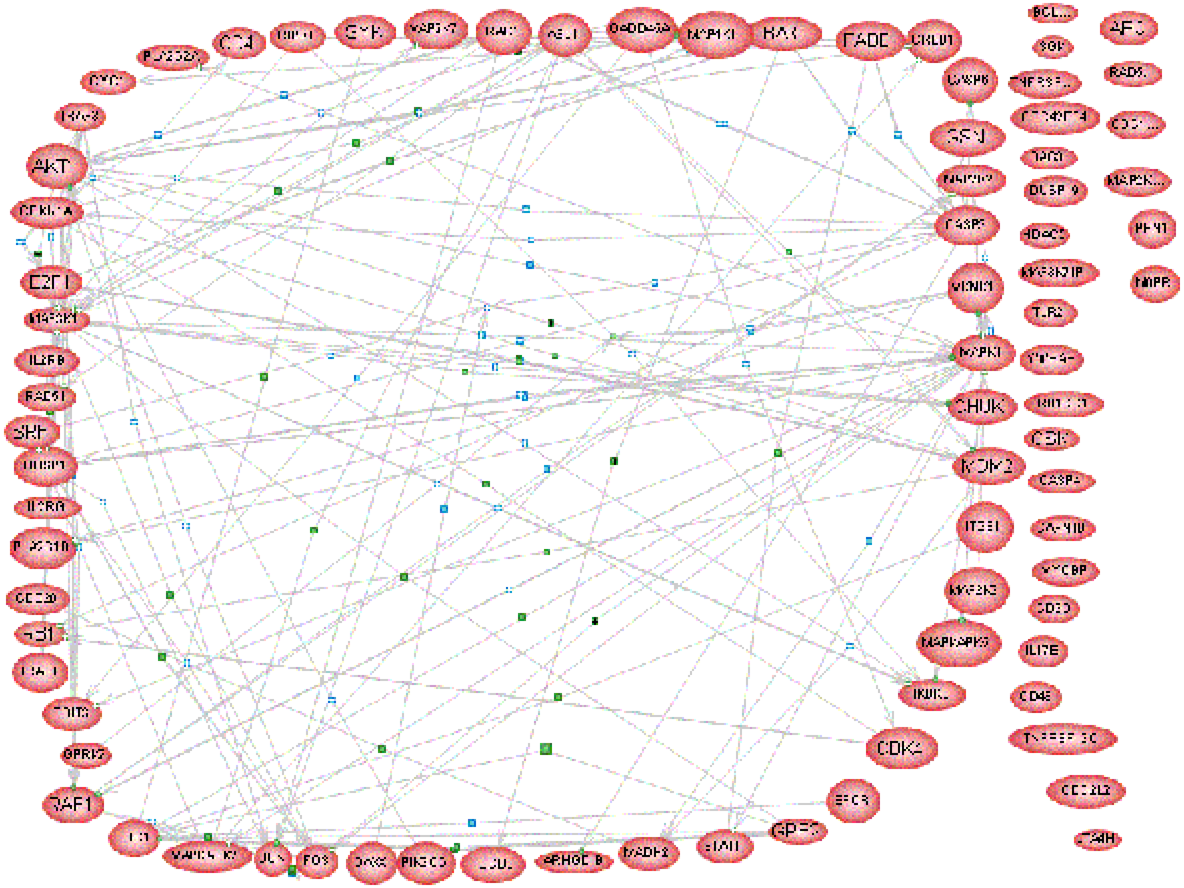
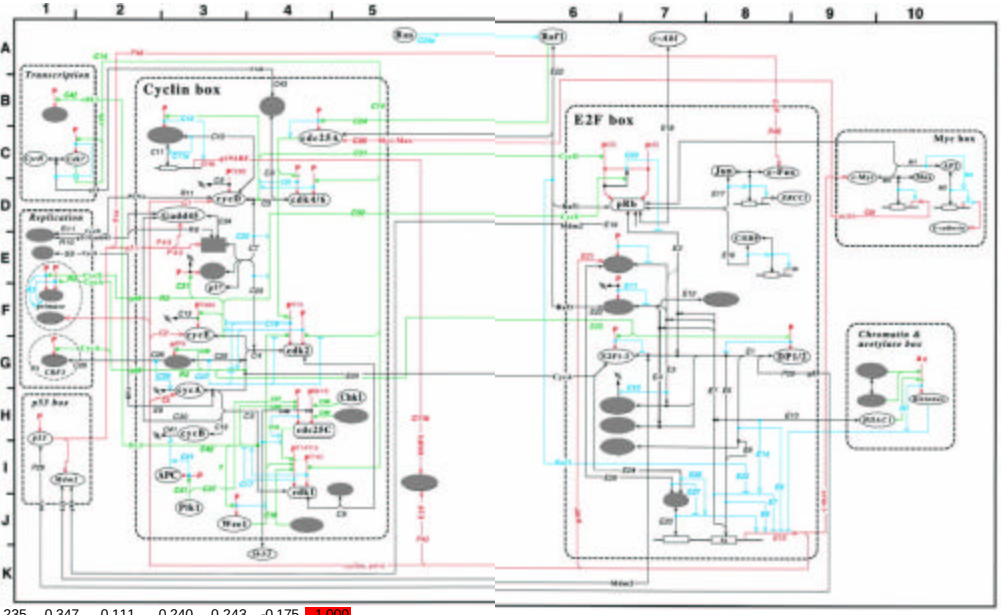


Figure 6A: The Cyclin - E2F cell cycle control system (version 3a - June 8, 1999)


Samples
Genes


AKT
E2F1
RAF1
FOS
CDK2
CHUK
MDM2
FADD
BAX
SRF

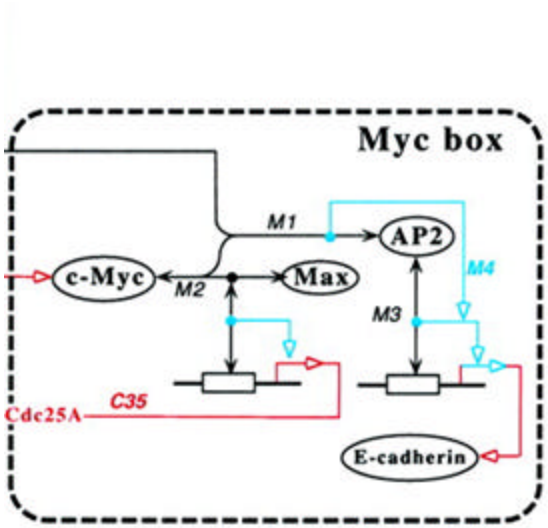
Kohn's Mammalian Cell Cycle Map (with AfCS genes)

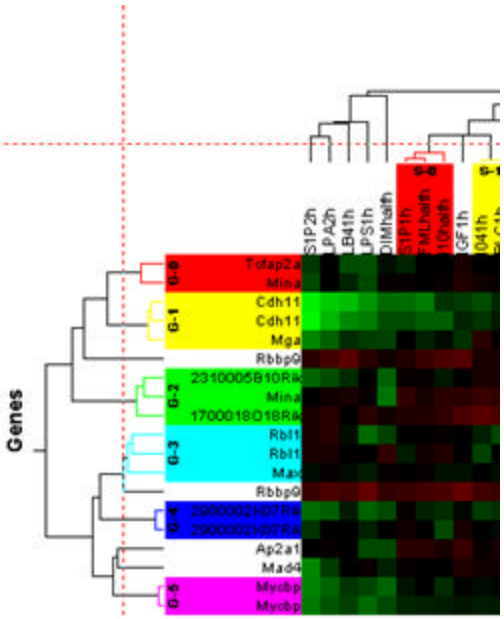


	Mad3	Ndr3	Oazin	Mga	Mina	Ndr1	Cdh2	Rbbp9	Ccnd2	Rb1	Mad	Rb1	0002K07	002K07	M
Mad3	1.000														
Ndr3	0.039	1.000													
Oazin	-0.303	0.080	1.000												
Mga	-0.315	-0.025	0.293	1.000											
Mina	-0.606	-0.004	0.376	0.209	1.000										
Ndr1	-0.177	0.129	0.261	0.002	0.303	1.000									
Cdh2	0.552	0.095	0.017	-0.306	-0.254	-0.022	1.000								
Rbbp9	0.197	0.020	0.017	-0.230	-0.056	0.119	0.243	1.000							
Ccnd2	0.061	0.192	0.178	-0.136	0.152	0.167	0.093	-0.025	1.000						
Rb1	-0.423	0.021	0.171	0.010	0.322	0.012	-0.116	0.115	0.021	1.000					
Mad	-0.151	0.337	0.261	0.023	0.214	0.254	0.033	0.077	0.138	0.194	1.000				
Rb1	0.077	0.186	0.031	-0.094	-0.073	0.126	0.085	0.040	0.120	0.046	0.478	1.000			
2900002K07Rik	-0.086	0.001	-0.045	-0.240	-0.129	0.051	-0.198	0.196	-0.232	0.068	-0.075	0.072	1.000		
2900002K07Rik	-0.122	-0.022	-0.143	-0.165	-0.060	-0.021	-0.247	0.197	-0.393	0.099	-0.104	0.003	0.847	1.000	
Mina	-0.201	0.264	0.240	-0.056	0.175	0.180	0.064	0.117	0.095	0.250	0.586	0.139	0.003	-0.033	-1
Cdh11	-0.460	-0.042	0.254	0.707	0.339	0.138	-0.243	-0.271	-0.088	0.011	0.050	-0.052	-0.204	-0.123	-0
Nmyc1	-0.015	-0.016	-0.069	0.090	-0.079	-0.085	-0.172	-0.179	0.149	0.179	-0.126	0.138	0.062	0.073	-0
2310005B10Rik	-0.247	0.062	0.165	0.194	0.139	0.085	-0.241	-0.067	0.393	0.185	0.032	-0.004	-0.012	-0.057	0
Cdh13	0.237	0.073	0.032	-0.052	-0.336	-0.010	0.258	0.308	0.038	-0.002	0.133	0.142	0.203	0.099	-0
Rbbp4	-0.105	0.062	0.014	-0.110	0.143	0.091	-0.057	-0.239	0.237	0.014	0.088	0.073	-0.129	-0.117	0
Mad4	0.065	0.028	0.013	0.400	-0.265	0.102	-0.226	-0.002	0.094	-0.255	-0.148	-0.045	0.003	-0.041	-0.235
Ap2a2	-0.078	-0.194	-0.143	0.383	-0.058	-0.119	-0.058	-0.333	-0.609	-0.195	-0.121	-0.093	-0.069	0.053	-0.049
Rbbp7	-0.385	0.008	0.255	-0.011	0.314	0.034	-0.209	-0.004	0.211	0.467	0.103	0.025	0.152	0.100	0.144
Rbl1	0.392	0.077	-0.138	-0.608	-0.132	-0.010	0.562	0.409	0.154	0.089	0.114	0.177	0.180	0.166	0.150
Pcdh18	0.540	-0.039	-0.345	-0.026	-0.493	-0.227	0.271	-0.198	-0.358	-0.477	-0.202	-0.103	-0.187	-0.117	-0.169
Rbl1	0.241	0.062	-0.096	-0.646	-0.085	-0.038	0.272	0.464	0.279	0.223	0.023	0.070	0.283	0.238	0.118
Rbbp9	0.724	0.052	-0.137	-0.406	-0.394	-0.087	0.601	0.524	0.057	-0.182	-0.084	0.024	0.047	0.055	-0.084
Mycbp	-0.072	-0.116	-0.009	0.214	-0.040	-0.146	0.126	-0.184	-0.598	0.034	0.108	-0.069	-0.292	-0.138	0.106
1700018O18Rik	-0.032	0.100	0.256	-0.024	0.291	0.181	0.028	0.158	0.777	0.069	0.136	0.081	-0.143	-0.283	0.095
Ccnd2	0.132	-0.230	0.016	0.058	-0.137	-0.048	-0.027	0.029	-0.154	-0.179	-0.157	0.062	0.037	0.090	-0.437
Oaz1	0.002	0.060	-0.073	-0.163	-0.213	-0.057	-0.080	0.068	-0.016	0.044	-0.083	-0.016	0.191	0.145	-0.070
Ahcy	-0.128	-0.003	0.022	-0.110	0.068	-0.020	-0.157	-0.218	0.238	0.184	0.013	0.141	-0.106	-0.125	0.022
Cdh23	-0.151	0.204	0.352	0.054	0.239	0.384	0.031	0.098	0.243	0.076	0.632	0.327	-0.065	-0.124	0.554
Tcfap2b	-0.359	-0.017	0.285	0.013	0.544	0.239	0.151	0.352	-0.047	0.366	0.157	-0.152	-0.071	0.025	0.155
Cdh11	-0.065	-0.005	-0.033	0.371	-0.031	-0.165	-0.167	-0.346	-0.014	-0.086	-0.071	0.021	-0.178	-0.159	-0.217
Cdh11	-0.583	-0.009	0.252	0.569	0.457	0.180	-0.322	-0.363	-0.037	0.053	0.092	-0.057	-0.176	-0.088	0.012
Cdh23	0.548	0.007	-0.359	-0.310	-0.279	-0.283	0.312	0.069	-0.017	-0.308	-0.192	0.023	-0.055	-0.073	-0.221
5730443C07UnK	-0.301	0.045	0.277	0.085	0.321	0.049	-0.140	0.176	0.045	0.350	0.083	-0.047	0.048	0.101	0.103
Ap2a2	-0.294	-0.164	-0.092	0.134	0.077	-0.029	-0.077	-0.063	-0.209	-0.012	-0.042	-0.033	0.038	0.020	-0.104
Pcdhb14	-0.056	-0.149	-0.295	-0.048	-0.294	-0.232	-0.362	-0.254	-0.126	-0.141	-0.210	-0.177	0.205	0.169	-0.157
Cdh11	-0.498	-0.084	0.205	0.586	0.350	0.096	-0.261	-0.508	-0.032	0.003	0.018	-0.092	-0.345	-0.307	-0.046
Mga	-0.235	0.024	0.244	0.153	-0.021	0.002	-0.260	-0.300	0.038	-0.028	0.007	0.029	-0.044	-0.138	-0.046
Tcfap2a	-0.087	0.054	0.286	-0.118	0.548	0.184	0.232	0.108	0.266	0.090	0.152	-0.086	-0.315	-0.277	0.128
Mina	0.122	0.131	0.219	-0.187	0.137	0.068	0.270	0.394	0.500	0.121	0.153	0.073	-0.079	-0.178	0.118
Max	0.507	-0.031	-0.214	-0.485	-0.258	-0.156	0.474	0.301	0.189	-0.073	-0.055	0.077	0.032	-0.004	-0.010
Mycbp	-0.062	-0.057	-0.028	0.304	-0.146	-0.122	-0.049	-0.062	-0.653	-0.039	0.014	-0.173	-0.155	-0.002	-0.007
Cdh13	-0.150	-0.098	0.040	0.163	0.004	-0.135	-0.200	-0.094	-0.018	0.055	-0.008	0.013	-0.068	-0.022	-0.179
Oaz3	0.196	0.086	0.014	-0.389	0.192	-0.008	0.520	0.243	0.185	0.036	0.116	-0.038	-0.123	-0.151	0.126
Oaz1	-0.132	0.025	-0.055	-0.237	-0.216	-0.012	-0.381	-0.035	-0.070	0.025	-0.174	-0.083	0.320	0.278	-0.075
Oaz1	-0.158	-0.038	-0.077	-0.317	-0.090	-0.025	-0.347	0.034	-0.124	0.055	-0.159	-0.081	0.319	0.307	-0.088
Ap2a1	0.257	0.093	0.031	-0.323	-0.081	0.178	0.123	0.333	0.181	0.035	0.054	0.118	0.019	0.004	0.120
Cdh3	0.328	0.107	0.060	-0.323	-0.080	0.030	0.374	0.365	0.126	-0.113	0.007	-0.075	-0.029	-0.077	0.151
4631427C17Rik	-0.032	0.148	0.298	-0.135	0.349	0.162	0.297	0.250	0.275	0.128	0.270	0.203	-0.118	-0.205	0.152
Tcfap2a	0.560	0.161	-0.081	-0.532	-0.124	-0.031	0.612	0.469	0.092	-0.074	0.028	-0.070	-0.018	0.000	0.089
Pcdh8	-0.136	0.111	0.411	-0.026	0.172	0.109	0.023	0.256	0.127	0.227	0.195	0.049	-0.006	-0.062	0.154
2310005B10Rik	0.143	0.103	0.229	0.074	0.088	0.142	0.163	0.120	0.575	-0.182	0.043	0.043	-0.237	-0.376	-0.054
2900002K07Rik	0.125	-0.137	-0.146	-0.139	-0.288	-0.205	-0.135	-0.088	-0.189	-0.160	-0.253	-0.002	0.362	0.446	-0.112
Ndr2	0.123	0.137	0.000	-0.107	-0.138	0.141	0.123	0.221	0.084	-0.167	0.148	0.192	0.059	-0.045	0.178
2310005B10Rik	-0.161	0.068	0.208	0.074	0.273	0.152	-0.103	0.003	0.344	0.062	0.107	-0.012	-0.016	-0.057	0.064

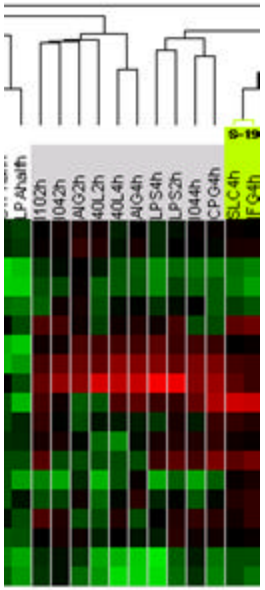
Mad3															
Ndr3															
Oazin															
Mga															
Mina															
Ndr1															
Cdh2															
Rbbp9															
Ccnd2															
Rb1															
Mad															
Rb1															
2900002K07Rik															
2900002K07Rik															
Mina															
Cdh11															
Nmyc1															
2310005B10Rik															
Cdh13															
Rbbp4															
Mad4	1.000														
Ap2a2	-0.039	1.000													
Rbbp7	-0.157	-0.336	1.000												
Rbl1	-0.003	-0.322	-0.346	0.061											
Pcdh18	-0.112	0.069	0.507	-0.613											
Rbl1	0.016	-0.302	-0.667	0.247											
Rbbp9	-0.118	-0.025	-0.299	-0.144											
Mycbp	-0.077	-0.140	0.588	-0.209											
1700018O18Rik	0.253	0.094	-0.603	0.308											
Ccnd2	-0.120	0.137	0.084	-0.205											
Oaz1	0.070	0.010	-0.173	0.050											
Ahcy	0.233	-0.112	-0.186	0.224											
Cdh23	0.088	0.012	-0.177	0.128											
Tcfap2b	-0.170	-0.137	-0.132	0.130											
Cdh11	-0.080	0.321	0.161	0.008											
Cdh11	0.011	0.200	0.380	0.002											
Cdh23	-0.052	-0.115	0.026	-0.357											
5730443C07UnK	0.049	-0.034	-0.245	0.339											
Ap2a2	0.056	0.051	0.299	0.034											
Pcdhb14	-0.012	0.152	0.196	0.027											
Cdh11	0.127	0.122	0.417	-0.011											
Mga	0.142	0.017	0.007	0.162											
Tcfap2a	0.166	-0.308	-0.285	0.120											
Mina	0.098	-0.078	-0.687	0.362											
Max	-0.085	-0.219	-0.290	-0.126											
Mycbp	0.013	-0.237	0.160	0.500	-0.240										
Cdh13	-0.066	-0.094	0.094	0.007	0.025										
Oaz3	0.157	0.024	-0.330	-0.223	0.014										
Oaz1	-0.087	-0.044	0.061	-0.239	0.069										
Oaz1	-0.131	-0.115	-0.075	-0.253	0.049										
Ap2a1	0.208	-0.037	-0.003	-0.353	0.065										
Cdh3	0.061	0.005	-0.109	-0.281	-0.053										
4631427C17Rik	0.070	-0.093	0.214	-0.290	-0.410	0.194									
Tcfap2a	-0.180	0.109	-0.052	-0.244	-0.305	-0.116									
Pcdh8	0.099	0.063	0.004	-0.097	-0.292	0.282									
2310005B10Rik	0.428	-0.048	0.127	0.179	-0.514	0.106									
2900002K07Rik	-0.072	-0.010	-0.008	-0.009	0.095	-0.048									
Ndr2	-0.105	0.186	-0.068	0.181	-0.173	-0.156									
2310005B10Rik	0.499	-0.136	0.255	0.060	-0.289	0.325									

Myc box genes (cell cycle)

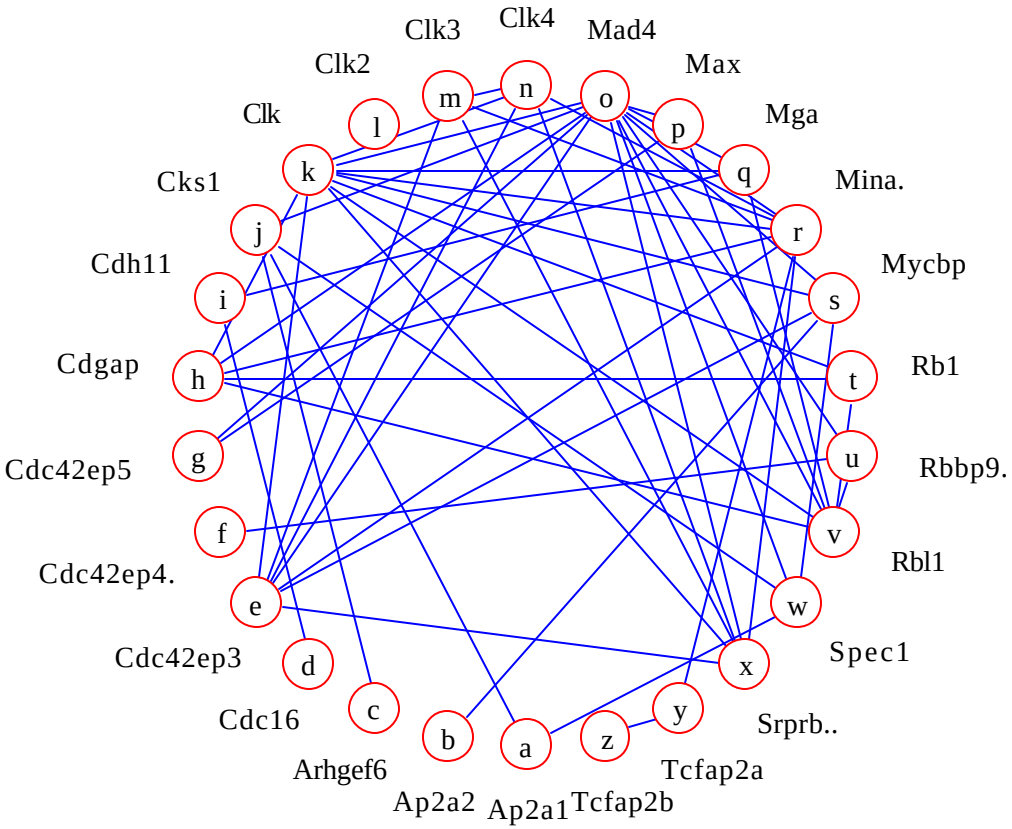




Subcluster of "mitogenic" ligand (late time periods)



MYC Connection Map



Genetic regulatory module generated by partial correlations critical value = 10^{-6}

Connection matrix

Signaling pathways of primary B cell (mouse)

cytosol only