Genomics of the Murine Immune System

Emphasis on technical challenges

John Hutton, MD
Professor of Pediatrics
University of Cincinnati
Director of Biomedical Informatics
Cincinnati Children's Hospital

Note

- I am in bioinformatics and analyze data, generally I do not create primary data
- Most information has been published with relevant colleagues, but some examples use unpublished data of colleagues and are not to be copied or quoted
- Special Thanks
  - Bruce Aronow
  - Anil Jegga
  - Marsha Wills-Karp

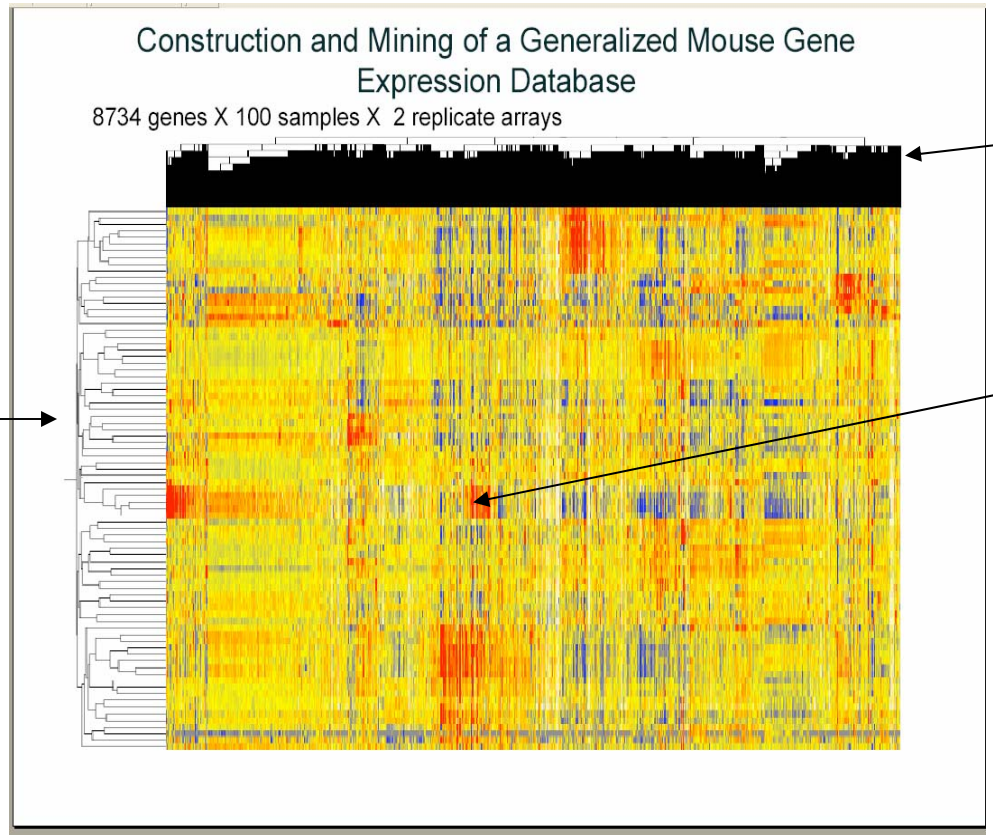
Immune Genes

1. Identify the genes that are essential for the differentiation, maintenance, and function of the immune system
   - What tissues or cell types comprise the immune system
   - We chose thymus, lymph nodes (unstimulated and 10 days after egg white lysozyme), spleen, activated T-cells, PB mononuclear cells
   - We did not monitor changes in cell populations
   - Expression arrays of sorted cells are now available
2. Annotate immune genes with regard to function
3. Identify regulatory elements that govern their expression

cDNA arrays

- Largely of historical interest, first microarrays that were available
- Some cDNA arrays were commercially produced
- Most are custom, locally produced
- Genome projects (sequencing and annotation), shared international databases, and advances in technology permitted switch to more specific oligonucleotide arrays

Hierarchical clustering of 8734 expressed sequences across 65 tissues (cDNA microarray)



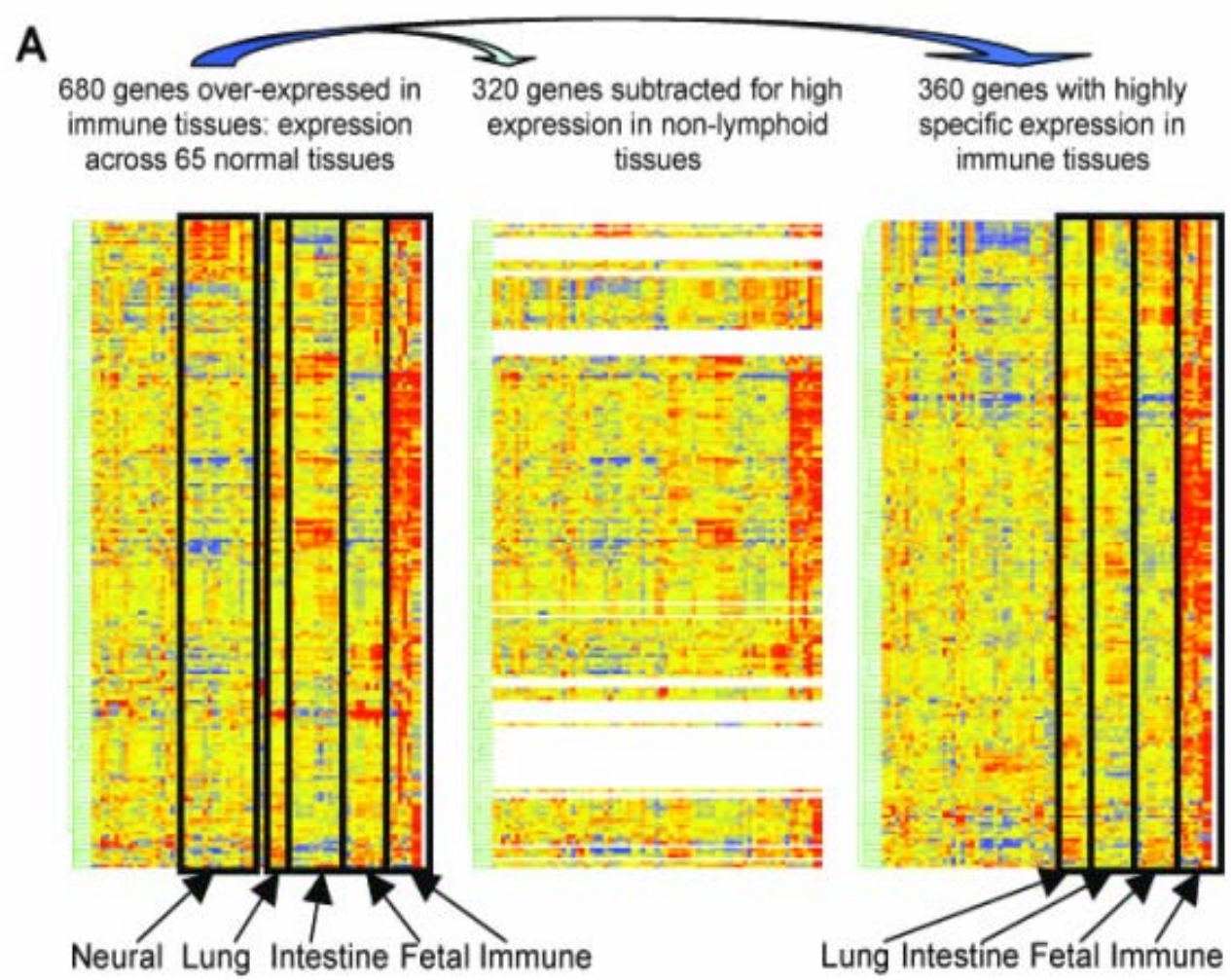
8734 genes

65 tissues

Localized cluster. Subset of genes highly expressed in subset of tissues

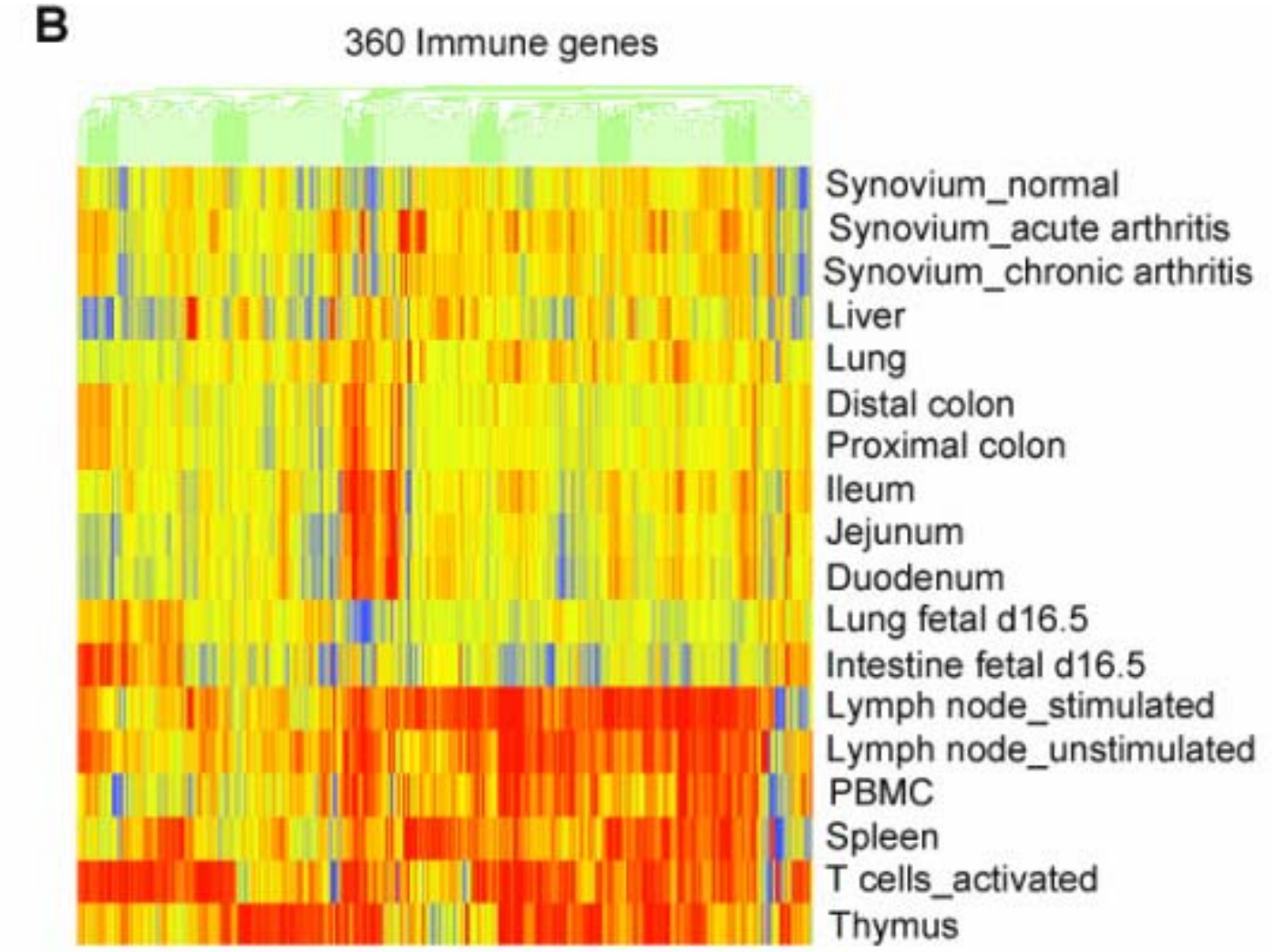
Hutton et al. Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system, BMC Genomics 5:82, 2004

Selection of 360 “Immune Genes”



Hutton et al. BMC Genomics 5:82, 2004

Expression of 360 “Immune Genes”



“Immune genes” selected by preferential expression do not necessarily have a specialized function

- General processes in immune tissues found in all dividing and metabolically active cells
  - Transcription 38 genes
  - Cell cycle or DNA replication 22 genes
  - Protein synthesis 20 genes
- Specialized processes in immune tissues and some other types of cells
  - Immune or defense response 59 genes
  - Receptor or cell signaling 47 genes
  - Apoptosis 14 genes
  - Transport 13 genes
  - Adhesion 10 genes
  - Chemotaxis 8 genes

Examples of Specialized Processes in Immune Tissues

- Presentation of foreign peptides to T-cells
  - H2-Aa, H2-Ab1, H2-DMa, H2-Eb1, H2-K, H2-L, H2-Ob, H2-Q7, B2m
- Transmission of signals: signaling cascade
  - Jak1, Stat1, Stat3, Stat4, Stat10, Rac2, Adcy7, Dgkz, Map3k1
- Production of or response to chemokines
  - Ccl4, Ccl6, Ccl19, Ccl22, Ccr2, Cxcl13, Cxcr4, S100a8
- Most are probably constitutive and not induced

Immune Genes
Special Technical Challenges

- Immune tissues contain heterogeneous cell populations
- Cellular composition changes with stimulation
- Different cell types have different expression profiles
- Immune genes are more than genes expressed in lymphoid cells, e.g. macrophages are “immune cells”
- Chemokines, immunoglobulins, histocompatibility genes are in large families with shared sequences and confusion about identity of a specific gene
- Many immune genes rearrange
  - Very difficult to identify orthologs (human and mouse)
  - Difficult to identify transcription start (i.e. exon 1)

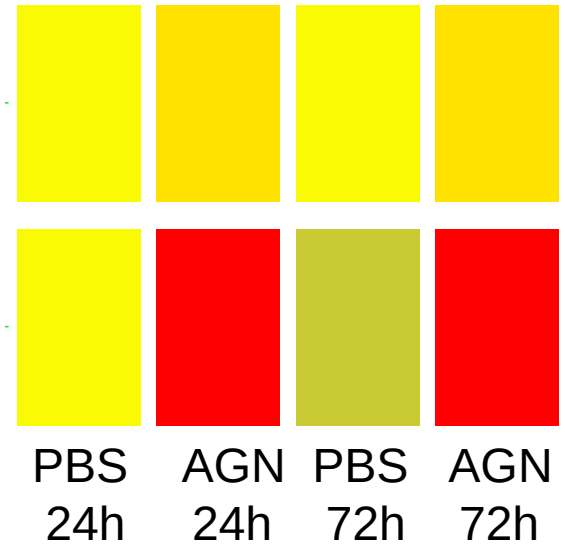
Oligonucleotide Arrays

- More specific than cDNA arrays
- Several technologies
  - Spotted on glass slides
  - Affymetrix chips
  - Illumina beads
- Expensive, both for arrays and equipment to read and process data

Different probes to “same gene” give different results

Clca3 = Gob-5 not Gob5 (Goblet cell, secretory) Affymetrix U74Av2
100623_at 100657_f_at 162287_r_at

Pulmonary response of sensitized mouse to intratracheal allergen



100657_f_at (no signal)

100623_at

162287_r_at

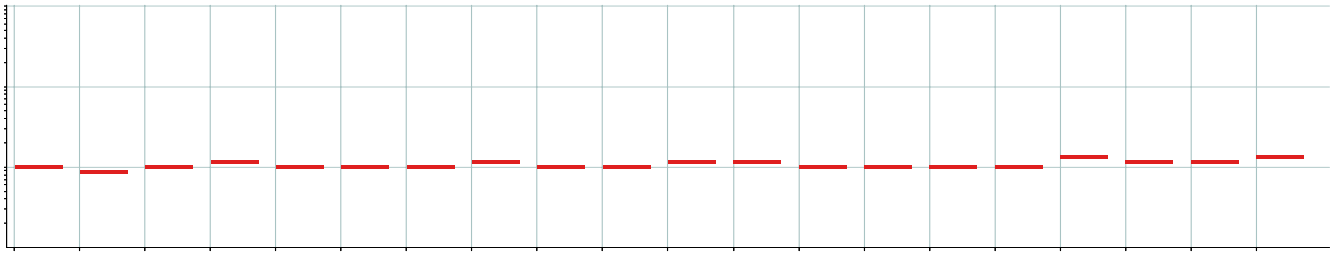
Many genes have several Affymetrix probe sets on same chip
Be aware: Different probe sets may give different results for same gene

What Affymetrix says about probes

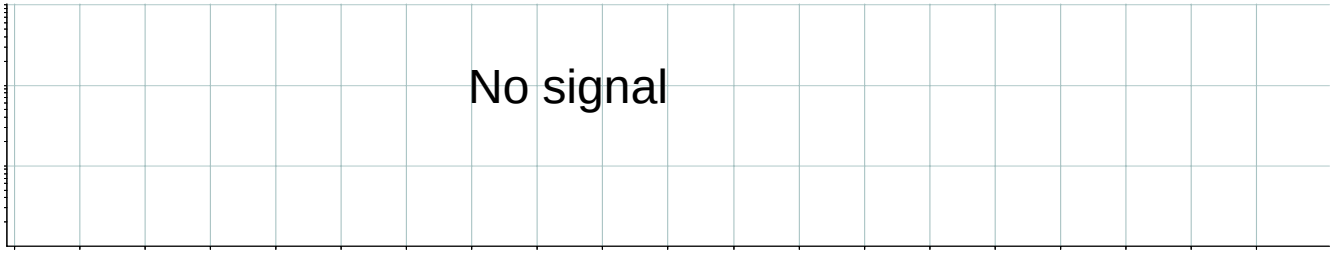
Extensions:

- _at Detects the anti-sense strand of the gene, unique to single transcript or common in transcripts from same gene or same gene family
- _s_at Multiple transcript variants share common sequence, not unique to a single transcript, not necessarily same gene family
- _x_at or _f_at Mixed set, cross hybridizes with at least one transcript outside family

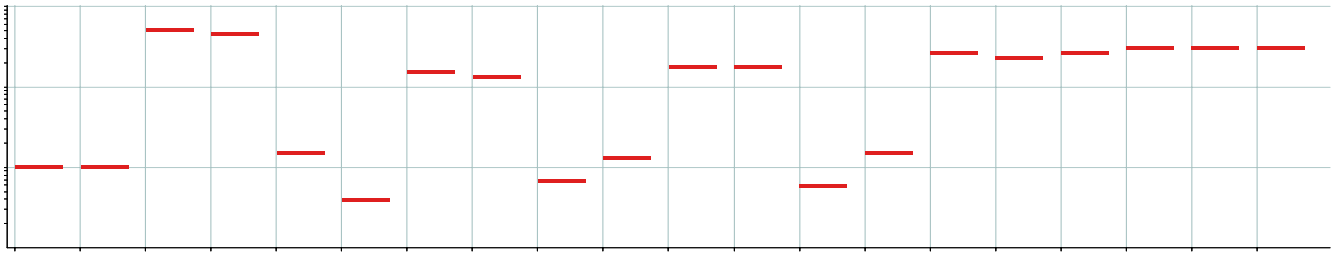
Expression of Clca3 (Gob-5) (U74Av2) in lung after allergen



Clca3 measured by 100623_at



Clca3 measured by 100657_f_at



Clca3 measured by 162287_r_at

Lessons About Aliases

Chi3l3

- MOE 430
- Common name : Chi3l3
- Map : 3 50.5 cM
- GenBank Accession # : NM_009892
- Synonyms: Chi3l3; Ym1
- Description: chitinase 3-like 3
- GO Biological Process: 6954; inflammatory response; traceable author statement
- GO Molecular Function: 4563; beta-N-acetylhexosaminidase activity; inferred from direct assay; 4568; NOT chitinase activity; inferred from direct assay

- U74Av2
- Common name : Chi3l3
- Map : 3 50.5 cM
- GenBank Accession # : NM_009892
- Synonyms: Chi3l3; Ym1; eosinophil chemotactic factor-L; ECF-L
- Description: chitinase 3-like 3
- GO BiologicalProcess: 6954; inflammatory response; traceable author statement
- GO Molecular Function: 4563; beta-N-acetylhexosaminidase; inferred from direct assay; 4568; chitinase activity; inferred from direct assay

Affymetrix

- GeneChip arrays quantitate known and annotated transcripts
- Over 39,000 transcripts measured per chip, 11 pairs of probes per transcript (GeneChip Mouse Expression Array MOE430), 25-mers,
- “GeneChip Tiling Arrays interrogate genomes at regular intervals, including both annotated and “junk” regions. Using this neutral approach, tiling arrays have been used to discover new transcripts”
- CustomExpress Arrays are available in a variety of formats to accommodate content requirements, ranging from 520 to over 61,000 sequences per array

Illumina BeadChips

- Technology very different from Affymetrix, not clear who will win on cost, ease of use and data analyses, speed, specificity
- Full-length 50-mer probes
- The Sentrix Mouse-6 BeadChip is designed to analyze six discrete mouse RNA samples on one chip, interrogating in each sample nearly 48,000 sequences from the mouse transcriptome

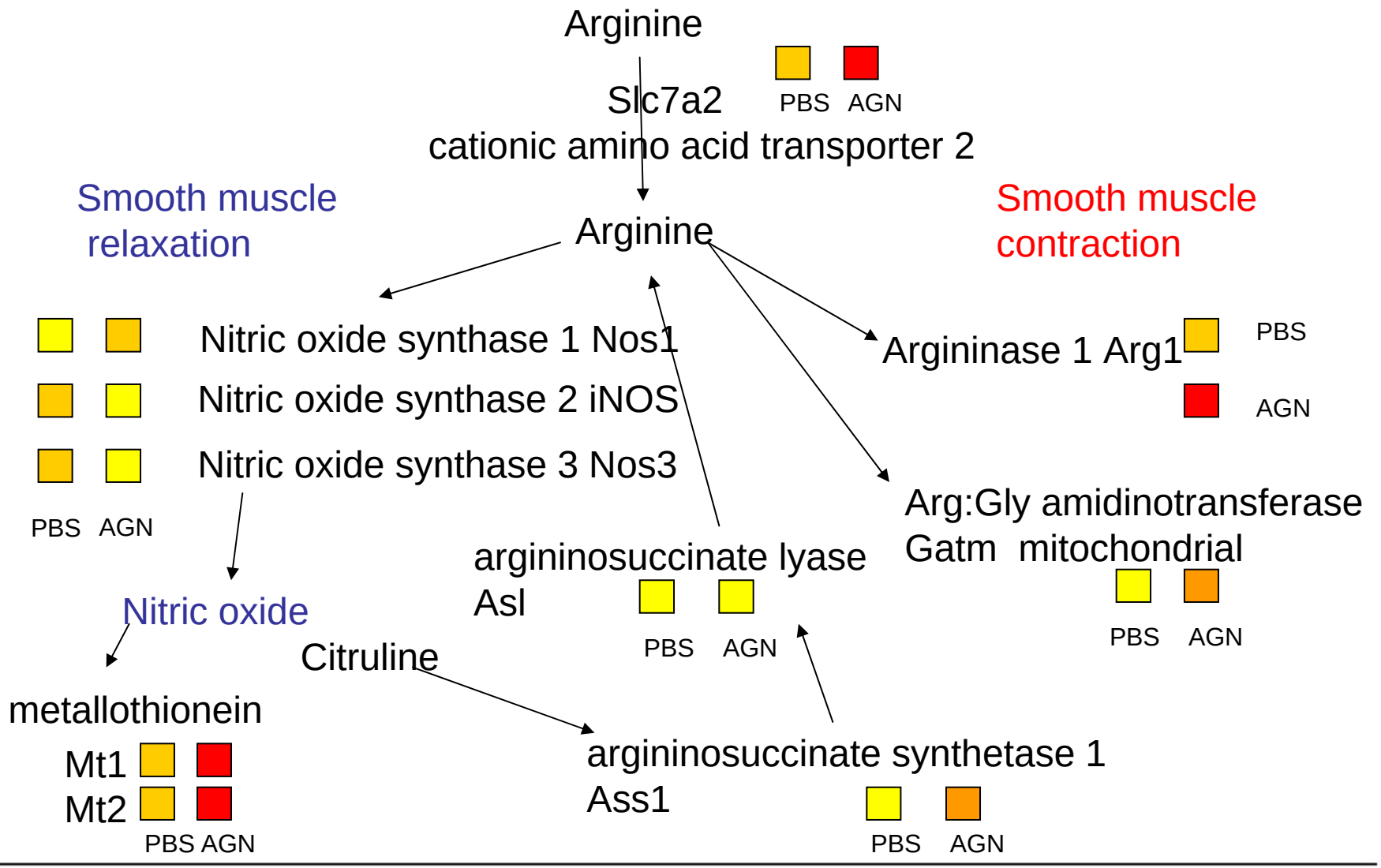
Illumina Toxicogenomics

- **Human Tox Gene Set**
  - 622 human genes of interest to toxicological screening. The set contains genes involved in genotoxicity, oxidative stress, inflammatory, and Phase 1 and 2 metabolism responses
  - Interleukins, BCL2, complement, chemokines, caspases, heat shock, interleukins, TNFs, nitric oxide synthase, no arginase, no chitinase, no Igs, no HLA
- **Orthologous Mouse Tox Gene Set**
  - 503 Genes, orthologs of the 622 Human Tox Gene Set
  - Housekeeping 7; Apoptosis 55; Drug metabolism 38; GPCR 82; Cancer 312; Immunology/Inflammation 137

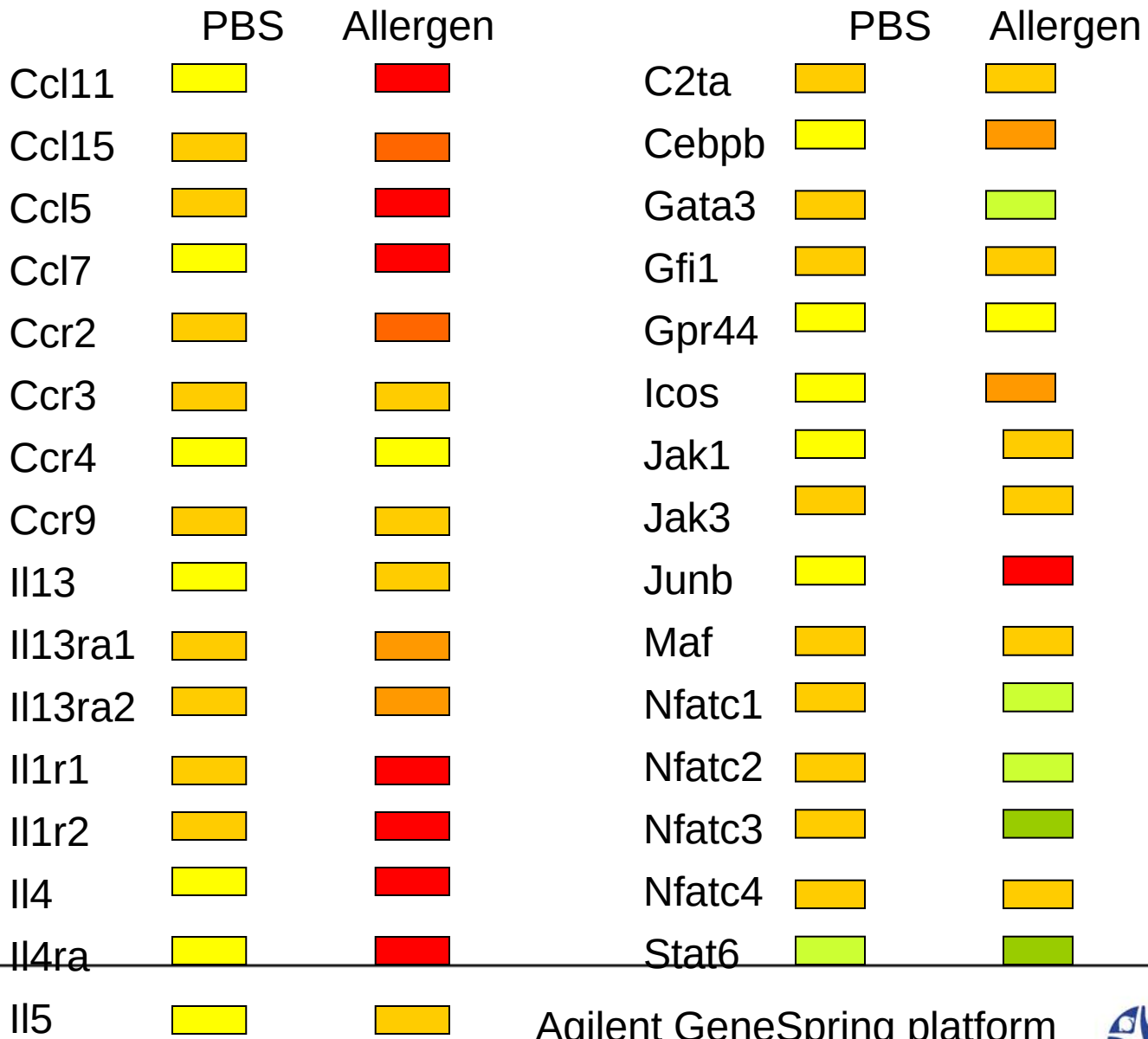
Pathway Analyses

- Pathway must have a well defined sequence of steps
- Must know names of genes encoding enzymes and substrates in the pathway
- Must have a file with picture of pathway or items to be displayed, jpg or similar format, to download
- Must have probes corresponding to gene names are on the microarray
- Must select the proper probes for measurement and display of expression
- There are tools to group genes by expression and suggest possible pathway

Arginine and the Hyper-responsive Airway
6 hrs after allergen to sensitized mouse (Wills-Karp data)



Expression of Th2 Genes in Lung after Allergen (MOE430 6hr, Wills-Karp)



Agilent GeneSpring platform

Probe and chip selection

- Investigators can study the same biological phenomenon, using the same basic experimental design, but get very different results from different microarrays
  - "Genes" with the same name on different microarrays may not be the same "gene"
  - The "probe" on a microarray may not measure what you think it does
- The field is changing rapidly, both annotation and technology
- Pathway analysis as a measure of the biological response to a stimulus is in its infancy and has high potential to improve interpretation of an expression profile

Recommendation for Mouse Immunotoxicology

- Choice of platform (Affymetrix vs Illumina) probably a matter of cost, speed of processing, and preferred technology, not specificity or number of genes
- Select several well accepted and documented mouse models used to assess xenobiotics for immunotoxic effects and create an annotated reference resource of RNAs for investigators
- Do not assume that you know the time course of gene expression in relation to toxicity

Reference Resource for Immunotoxicology

- Collect many tissues (vary dose, time, strain, gender) from treated and control mice - at least 3 biological replicates
- Relevant biological parameters on each tissue (histology, cell markers)
- Prepare a lot of high quality mRNA from each and freeze in small aliquots
- Annotate each RNA with experimental and biological parameters and make available to investigators

Reference Resource - 2

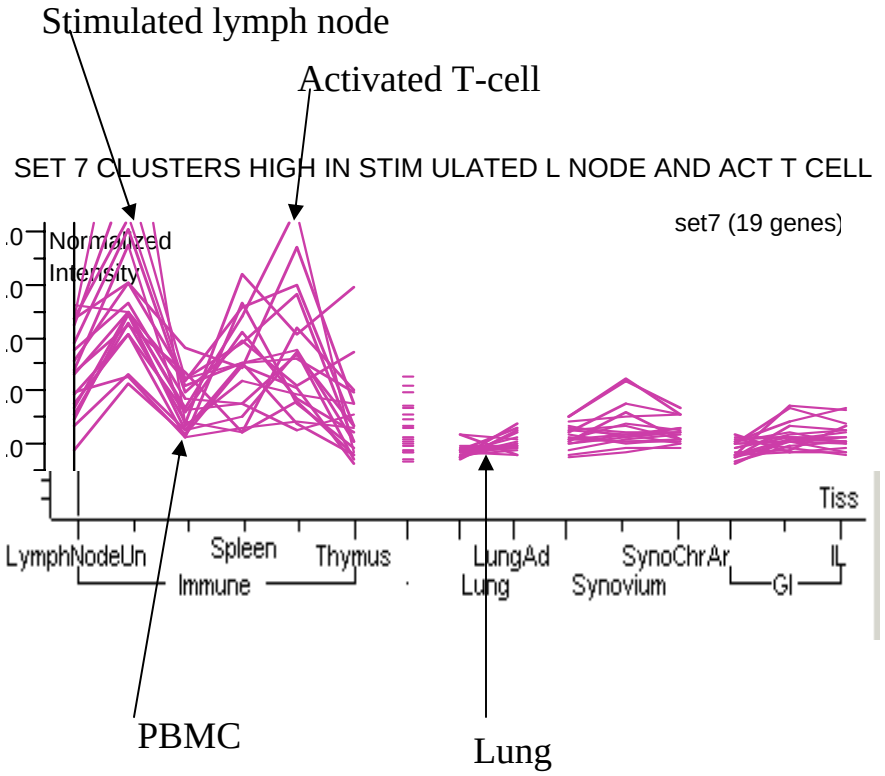
- Do not assume that populations of cells in immune tissues are the same before and after treatment
  - Selective cell death, measuring residual cell sub-population
  - Cell migration, prominent in immune system
  - Cell differentiation vs migration
- Do not assume that immunotoxic effects are limited to so called immune cells or tissues
- Do not assume that you can identify a unique set of immune genes that accurately measure immunotoxicity of every substance
- Do not assume that genomics is cheaper or faster or better than anatomic methods for detection of toxicity, as opposed to understanding toxicity

Grouping of Immune Genes by Expression

- Identify genes that are similarly expressed across tissues or after treatment
- Unproved hypothesis, mammalian genes that are expressed together share regulatory elements
- Conceivably could help identify genes that are induced together as part of induction of a coordinated pathway

Example of clustering to identify a set of genes that share a pattern of expression across tissues

"Immune genes" are differently expressed in different "immune tissues"



The 19 gene set

Cd72	NM_007654	CD72 antigen	4
B3gnt5	NM_054052	UDP-GlcNAc:betaG	16
Irf5	NM_012057	interferon regulator	6
Hck	NM_010407	hemopoietic cell kir	2
Stk10	NM_009288	serine/threonine kir	11
Tnfrsf13b	AK004668	tumor necrosis fact	11
Abca7	NM_013850	ATP-binding casse	10
Ncf4	NM_008677	neutrophil cytosolic	15
Lyn	BC031547	Yamaguchi sarcom	4
Map3k1	AF117340	mitogen activated p	13
Tap1	NM_013683	transporter 1, ATP-	17
Dock2	NM_033374	dedicator of cyto-ki	11
Sema4d	NM_013660	sema domain, imm	13
Map4k1	BC005433	mitogen activated p	7
Cdc6	NM_011799	cell division cycle 6	11
Ly86	NM_010745	lymphocyte antigen	13
Icsbp	NM_008320	interferon concensu	8
Serpina3g	BC002065	serine (or cysteine)	12

Regulatory Modules

Assumptions Underlying Computational Search for Regulatory Modules

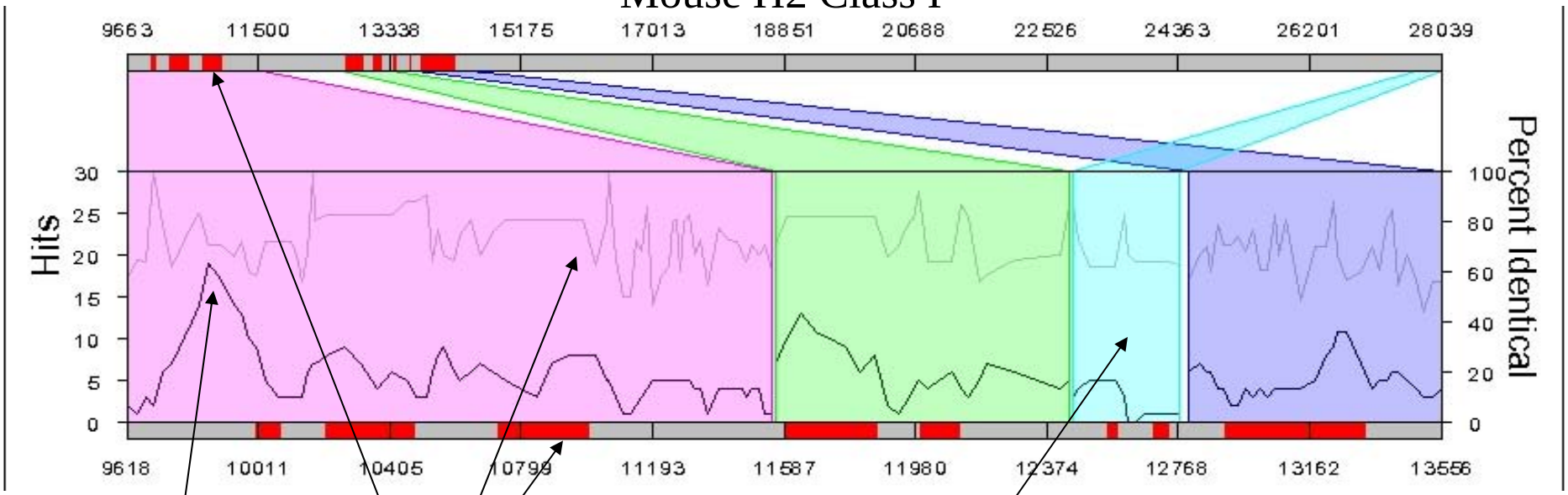
- Shared regulatory modules underlie coordinate expression of at least some genes
- Clusters of TF binding sites within regulatory modules are phylogenetically conserved
- Sequences of binding sites for transcription factors are phylogenetically conserved
- Transcription factor specificities are phylogenetically conserved

Present Computational Analyses of Gene Regulation Require

- Sets of “coordinately regulated genes”
- Mouse and human orthologs
- Genes, not pseudogenes
- Orthologs, not paralogs
- Unequivocal identification of exon 1

A Test: Class I Histocompatibility Genes
Exon 1 begins at bp 10,001

Mouse H2 Class I



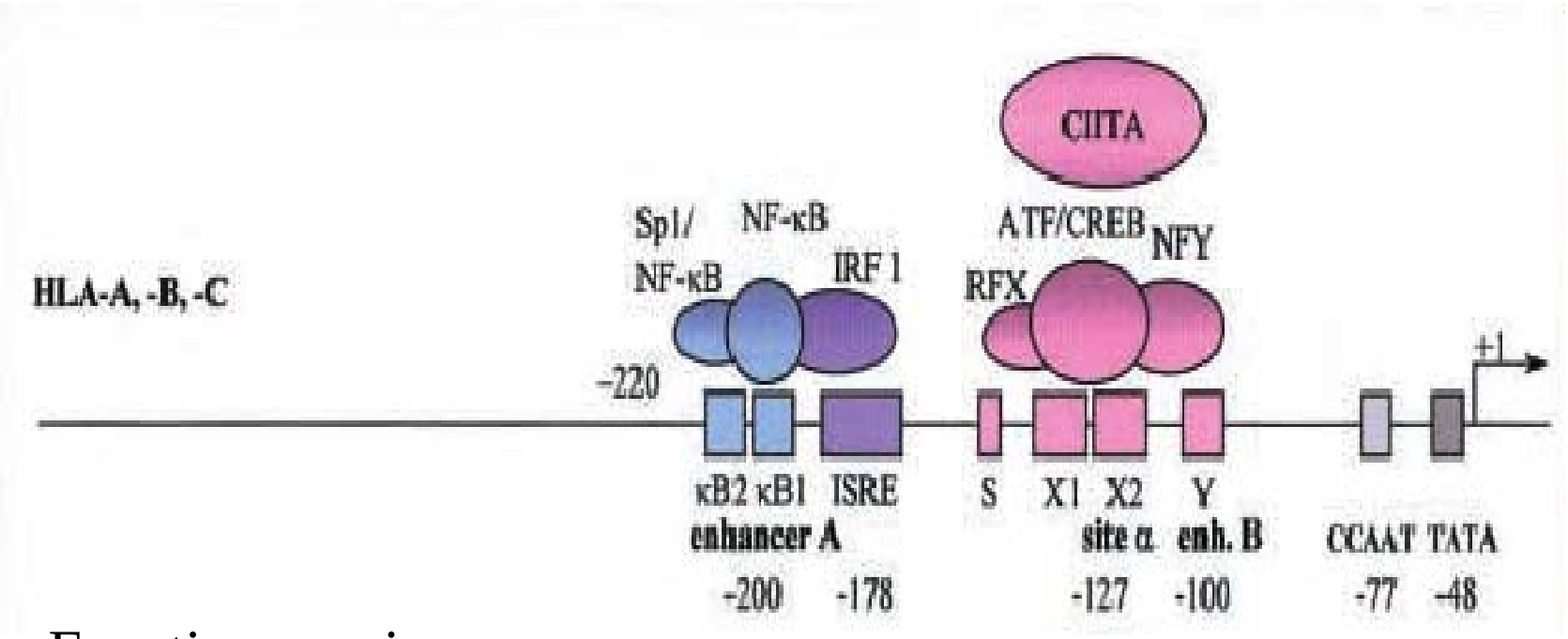
Human HLA Class I

Rearrangement of sequences

Good conservation of layout and sequence of exons

Cluster of TF binding sites in promoter – We examine this region

TF Binding Sites in HLA Class I Promoter Identified by Experiments

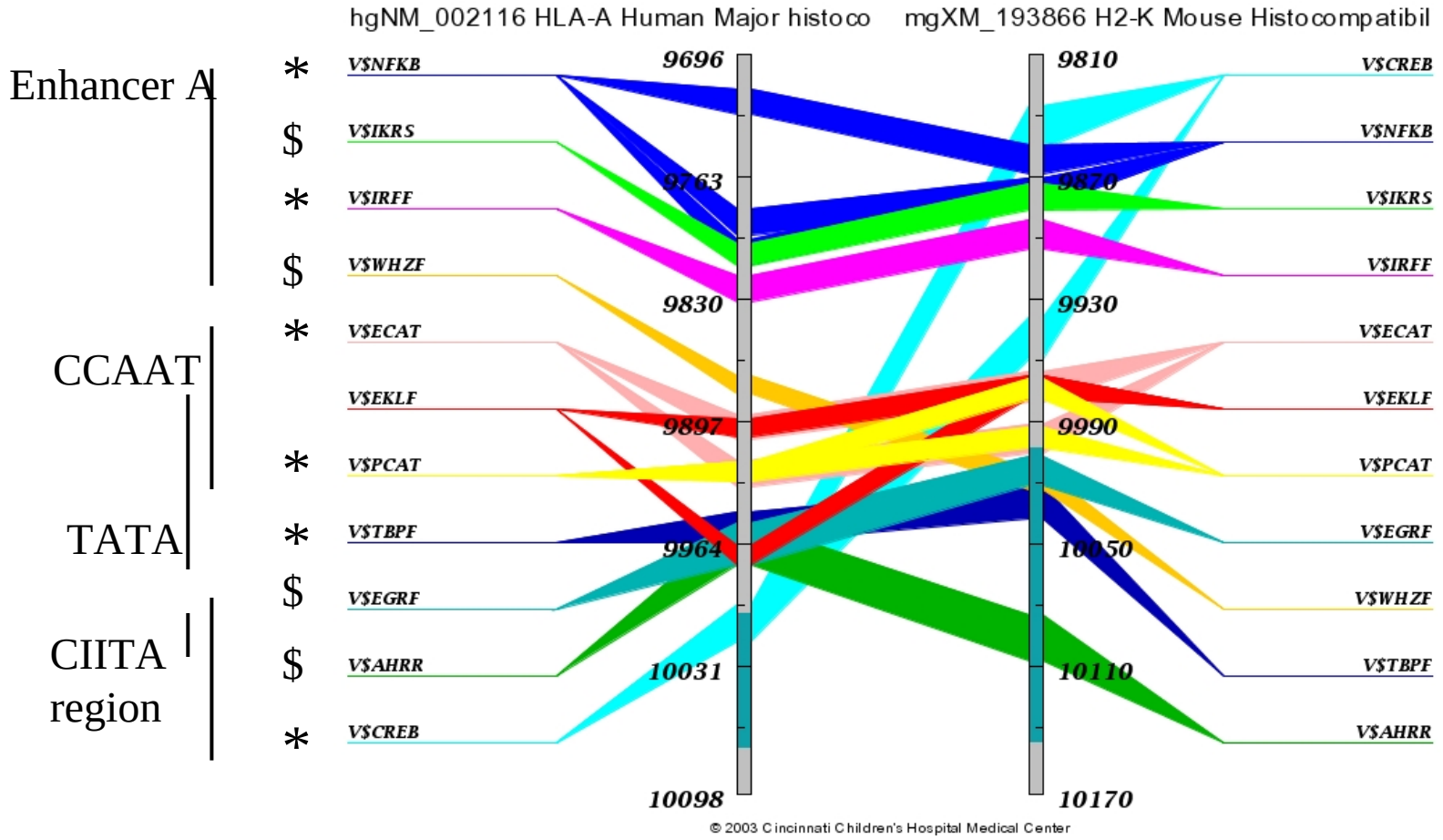


Function requires:

- Open chromatin to expose the TF binding sites to TFs
- Presence of the TF proteins in the cell

Comparison of Experimentation and Computation

Promoters of Class I Histocompatibility Genes



* Experimentation and computation $ Computation only

What Computation Can Add to Studies of Gene Regulatory Elements

- Predicts new TF binding sites within established modules to guide additional biological experiments
- New sites in class I
  - IKRS role in lymphocyte differentiation
  - WHZF family includes “nude” gene critical for development of thymus
  - EGR role during positive selection in thymus
  - AhR role in death signal in lymphocytes
- Predicts location and composition of new regulatory modules to guide difficult biological experiments
- Not likely to help define causes of immunotoxicity

End