BSC 4934: Q'BIC Capstone Workshop

Giri Narasimhan

ECS 254A; Phone: x3748

giri@cis.fiu.edu

http://www.cis.fiu.edu/~giri/teach/BSC4934_Su09.html

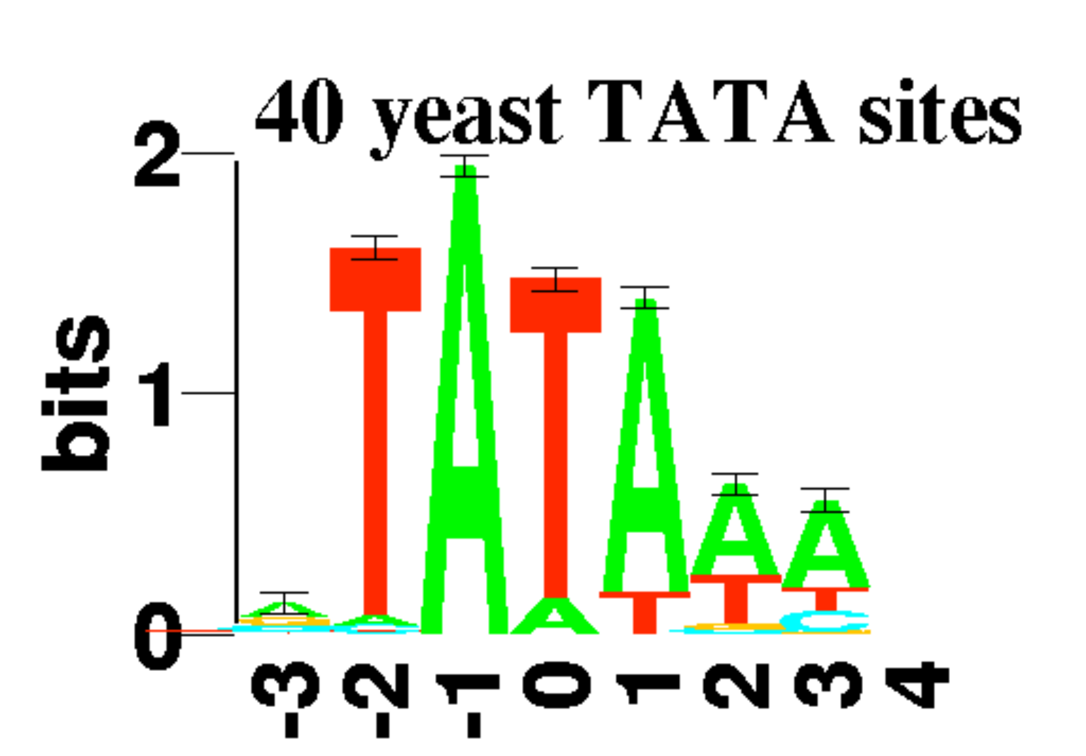
24 June through 7 July, 2009

Patterns in DNA Sequences

- Signals in DNA sequence control events
  - Start and end of genes
  - Start and end of introns
  - Transcription factor binding sites (regulatory elements)
  - Ribosome binding sites
- Detection of these patterns are useful for
  - Understanding gene structure
  - Understanding gene regulation

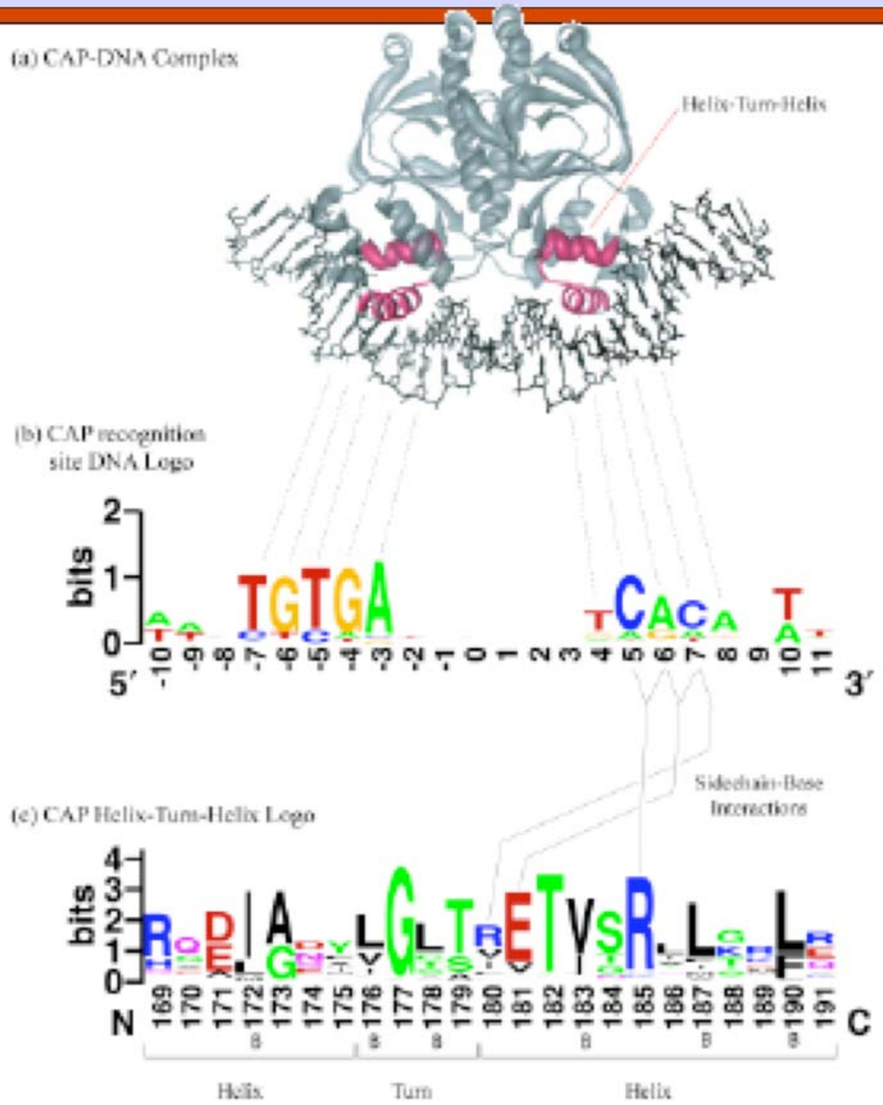
Motifs in DNA Sequences

- Given a collection of DNA sequences of promoter regions, locate the transcription factor binding sites (also called regulatory elements)
  - Example:



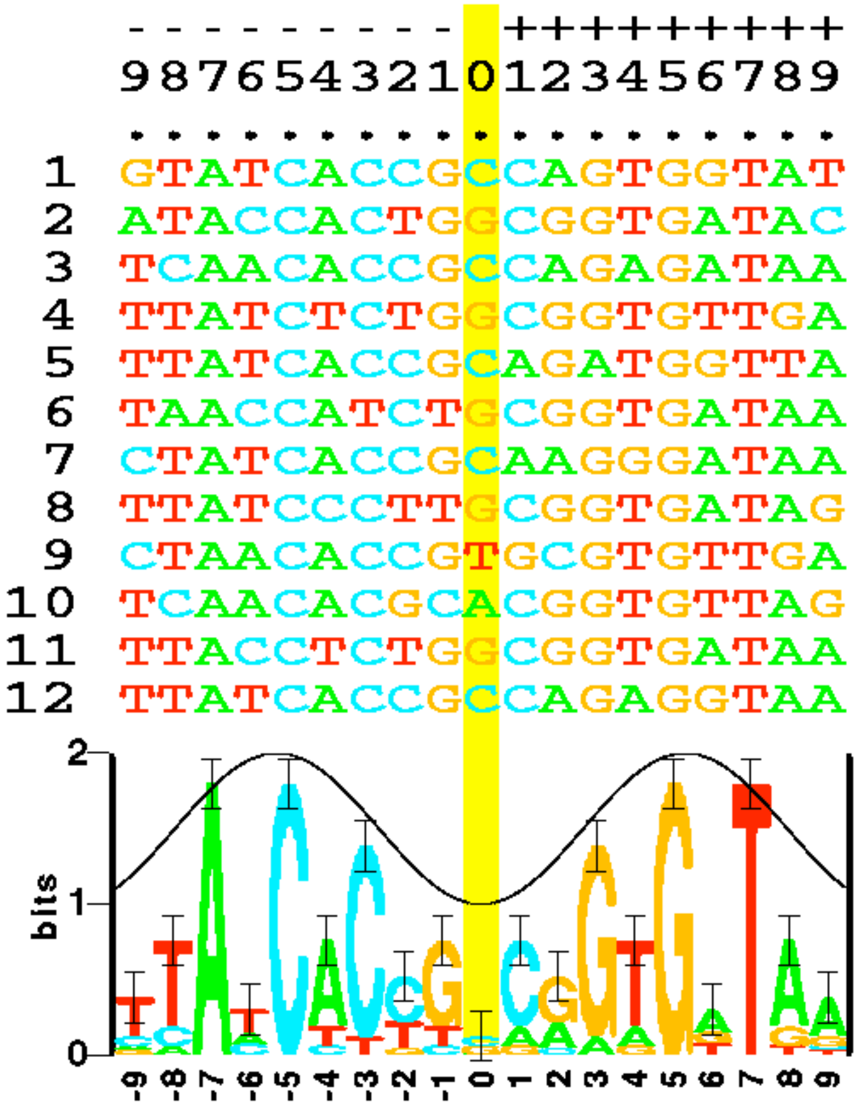
http://www.lecb.ncifcrf.gov/~toms/sequencelogo.html

Motifs



http://weblogo.berkeley.edu/examples.html

Motifs in DNA Sequences



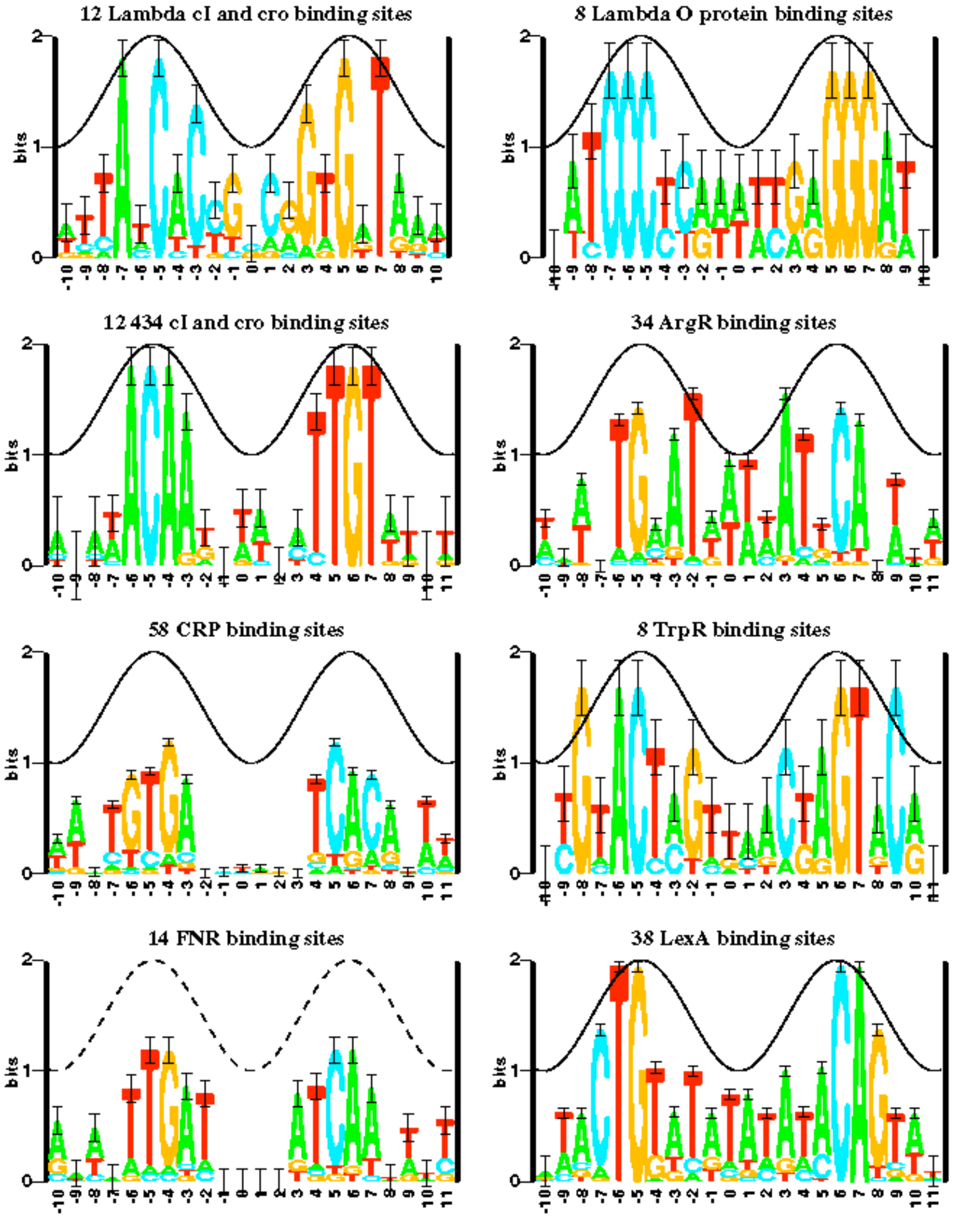
```
    ---------0+++++++++
    9876543210123456789
    ...................
 1  GTATCACCGCCAGTGGTAT
 2  ATACCACTGGCGGTGATAC
 3  TCAACACCGCCAGAGATAA
 4  TTATCTCTGGCGGTGTTGA
 5  TTATCACCGCAGATGGTTA
 6  TAACCATCTGCGGTGATAA
 7  CTATCACCGCAAGGGATAA
 8  TTATCCCTTGCGGTGATAG
 9  CTAACACCGTGCGTGTTGA
10  TCAACACGCACGGTGTTAG
11  TTACCTCTGGCGGTGATAA
12  TTATCACCGCCAGAGGTAA
```

12 Lambda cI and cro binding sites

Fig. 1. Some aligned sequences and their sequence logo. At the top of the figure are listed the 12 DNA sequences from the P_L and P_R control regions in bacteriophage lambda. These are bound by both the cI and cro proteins [16]. Each even numbered sequence is the complement of the preceding odd numbered sequence. The sequence logo, described in detail in the text, is at the bottom of the figure. The cosine wave is positioned to indicate that a minor groove faces the center of each symmetrical protein. Data which support this assignment are given in reference [17].

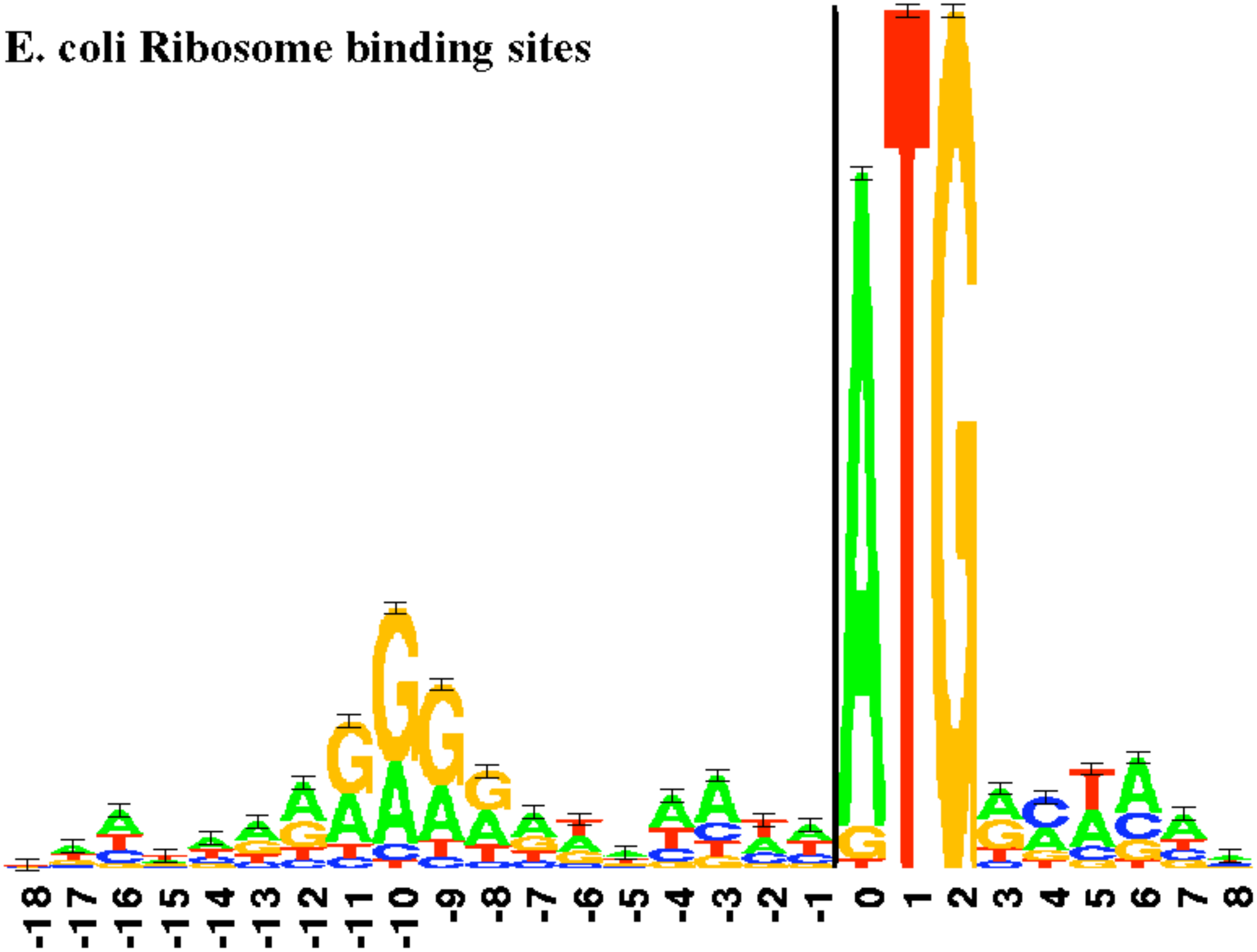
http://www.lecb.ncifcrf.gov/~toms/sequencelogo.html

More Motifs in *E. Coli* DNA Sequences



http://www.lecb.ncifcrf.gov/~toms/sequencelogo.html

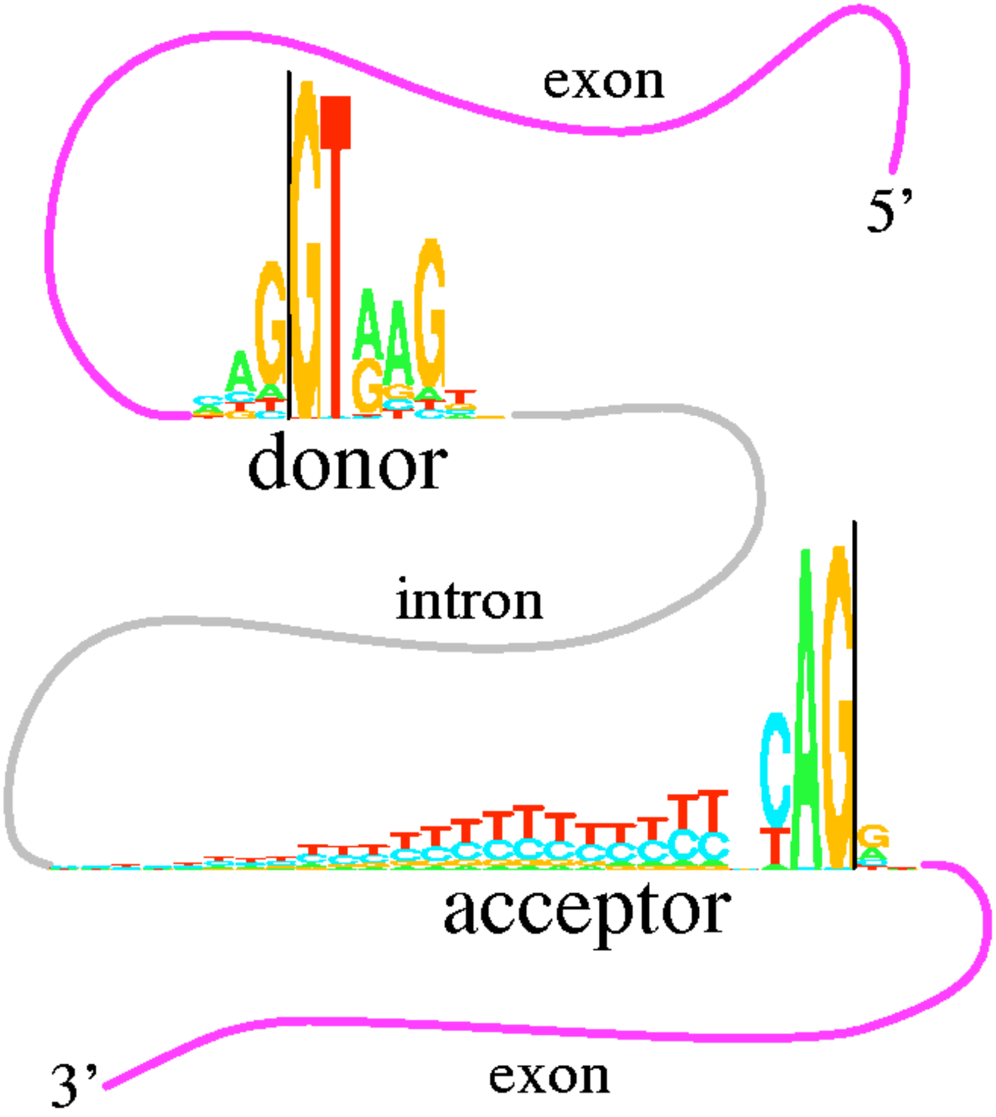
E. coli Ribosome binding sites



http://www.lecb.ncifcrf.gov/~toms/sequencelogo.html

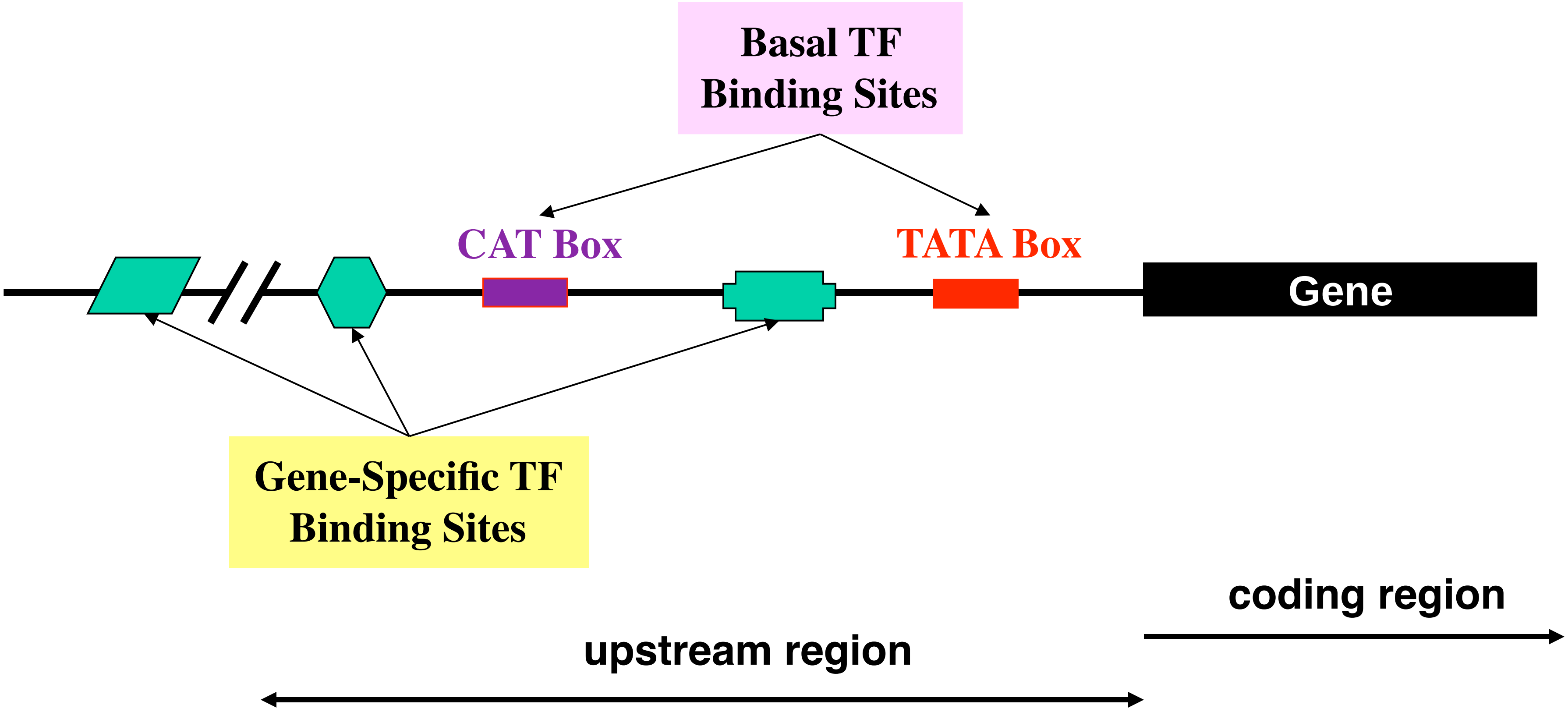
Other Motifs in DNA Sequences: Human Splice Junctions

This figure shows two "sequence logos" which represent sequence conservation at the 5' (donor) and 3' (acceptor) ends of human introns. The region between the black vertical bars is removed during mRNA splicing. The logos graphically demonstrate that most of the pattern for locating the intron ends resides on the intron. This allows more codon choices in the protein-coding exons. The logos also show a common pattern "CAG|GT", which suggests that the mechanisms that recognize the two ends of the intron had a common ancestor. See R. M. Stephens and T. D. Schneider, "Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites", J. Mol. Biol., 228, 1124-1136, (1992)

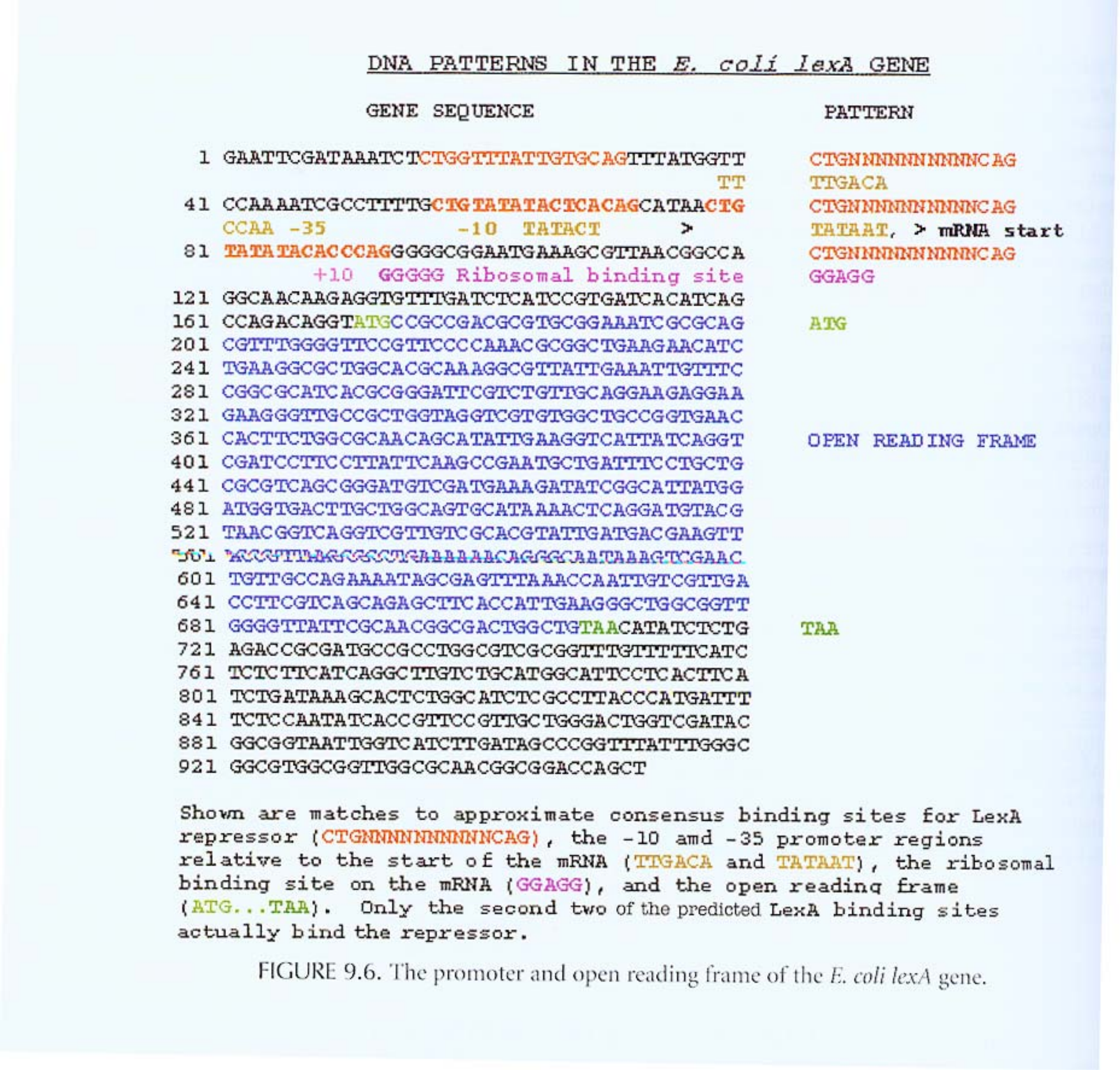


http://www.lecb.ncifcrf.gov/~toms/sequencelogo.html

Transcription Regulation



Prokaryotic Gene Characteristics



DNA PATTERNS IN THE *E. coli lexA* GENE

```
           GENE SEQUENCE                              PATTERN

  1 GAATTCGATAAATCTCTGGTTTATTGTGCAGTTTATGGTT      CTGNNNNNNNNNNCAG
                                          TT      TTGACA
 41 CCAAAATCGCCTTTTGCTGTATATACTCACAGCATAACTG      CTGNNNNNNNNNNCAG
    CCAA -35        -10  TATACT        >          TATAAT, > mRNA start
 81 TATATACACCCAGGGGGCGGAATGAAAGCGTTAACGGCCA      CTGNNNNNNNNNNCAG
          +10  GGGGG Ribosomal binding site       GGAGG
121 GGCAACAAGAGGTGTTTGATCTCATCCGTGATCACATCAG
161 CCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAG      ATG
201 CGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATC
241 TGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTC
281 CGGCGCATCACGCGGGATTCGTCTGTTGCAGGAAGAGGAA
321 GAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAAC
361 CACTTCTGGCGCAACAGCATATTGAAGGTCATTATCAGGT      OPEN READING FRAME
401 CGATCCTTCCTTATTCAAGCCGAATGCTGATTTCCTGCTG
441 CGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGG
481 ATGGTGACTTGCTGGCAGTGCATAAAACTCAGGATGTACG
521 TAACGGTCAGGTCGTTGTCGCACGTATTGATGACGAAGTT
561 ACCGTTAAGGCCCTGAAAAACAGGGCAATAAAGTCGAAC
601 TGTTGCCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGA
641 CCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTGGCGGTT
681 GGGGTTATTCGCAACGGCGACTGGCTGTAACATATCTCTG      TAA
721 AGACCGCGATGCCGCCTGGCGTCGCGGTTTGTTTTTCATC
761 TCTCTTCATCAGGCTTGTCTGCATGGCATTCCTCACTTCA
801 TCTGATAAAGCACTCTGGCATCTCGCCTTACCCATGATTT
841 TCTCCAATATCACCGTTCCGTTGCTGGGACTGGTCGATAC
881 GGCGGTAATTGGTCATCTTGATAGCCCGGTTTATTTGGGC
921 GGCGTGGCGGTTGGCGCAACGGCGGACCAGCT
```

Shown are matches to approximate consensus binding sites for LexA repressor (CTGNNNNNNNNNNCAG), the -10 amd -35 promoter regions relative to the start of the mRNA (TTGACA and TATAAT), the ribosomal binding site on the mRNA (GGAGG), and the open reading frame (ATG...TAA). Only the second two of the predicted LexA binding sites actually bind the repressor.

FIGURE 9.6. The promoter and open reading frame of the *E. coli lexA* gene.

Motifs in DNA Sequences

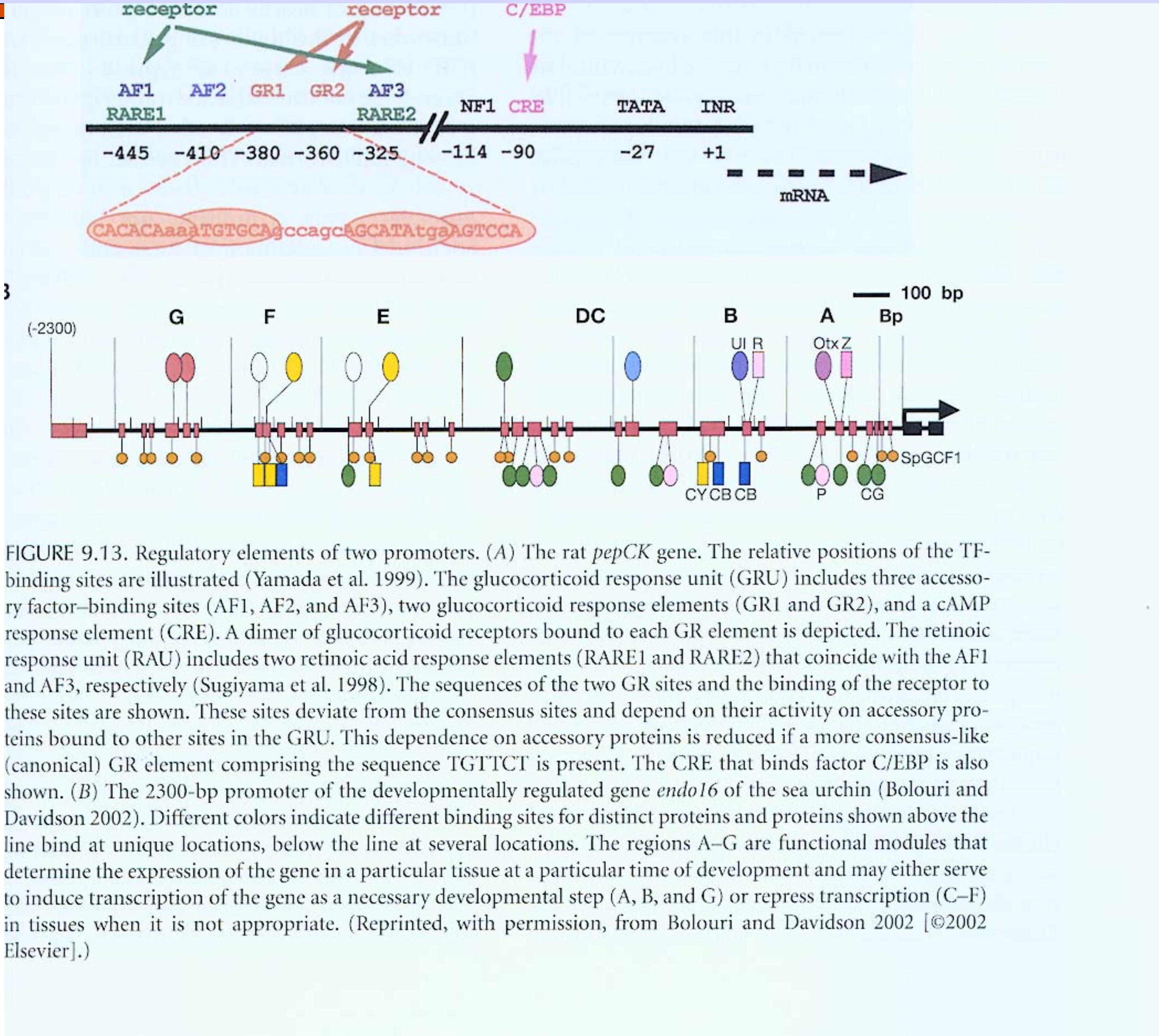
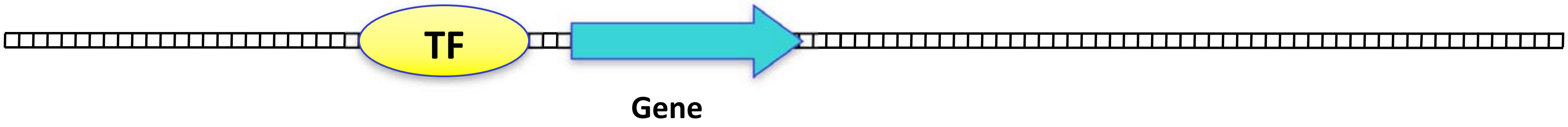


FIGURE 9.13. Regulatory elements of two promoters. (*A*) The rat *pepCK* gene. The relative positions of the TF-binding sites are illustrated (Yamada et al. 1999). The glucocorticoid response unit (GRU) includes three accessory factor–binding sites (AF1, AF2, and AF3), two glucocorticoid response elements (GR1 and GR2), and a cAMP response element (CRE). A dimer of glucocorticoid receptors bound to each GR element is depicted. The retinoic response unit (RAU) includes two retinoic acid response elements (RARE1 and RARE2) that coincide with the AF1 and AF3, respectively (Sugiyama et al. 1998). The sequences of the two GR sites and the binding of the receptor to these sites are shown. These sites deviate from the consensus sites and depend on their activity on accessory proteins bound to other sites in the GRU. This dependence on accessory proteins is reduced if a more consensus-like (canonical) GR element comprising the sequence TGTTCT is present. The CRE that binds factor C/EBP is also shown. (*B*) The 2300-bp promoter of the developmentally regulated gene *endo16* of the sea urchin (Bolouri and Davidson 2002). Different colors indicate different binding sites for distinct proteins and proteins shown above the line bind at unique locations, below the line at several locations. The regions A–G are functional modules that determine the expression of the gene in a particular tissue at a particular time of development and may either serve to induce transcription of the gene as a necessary developmental step (A, B, and G) or repress transcription (C–F) in tissues when it is not appropriate. (Reprinted, with permission, from Bolouri and Davidson 2002 [©2002 Elsevier].)

Single Gene Activation

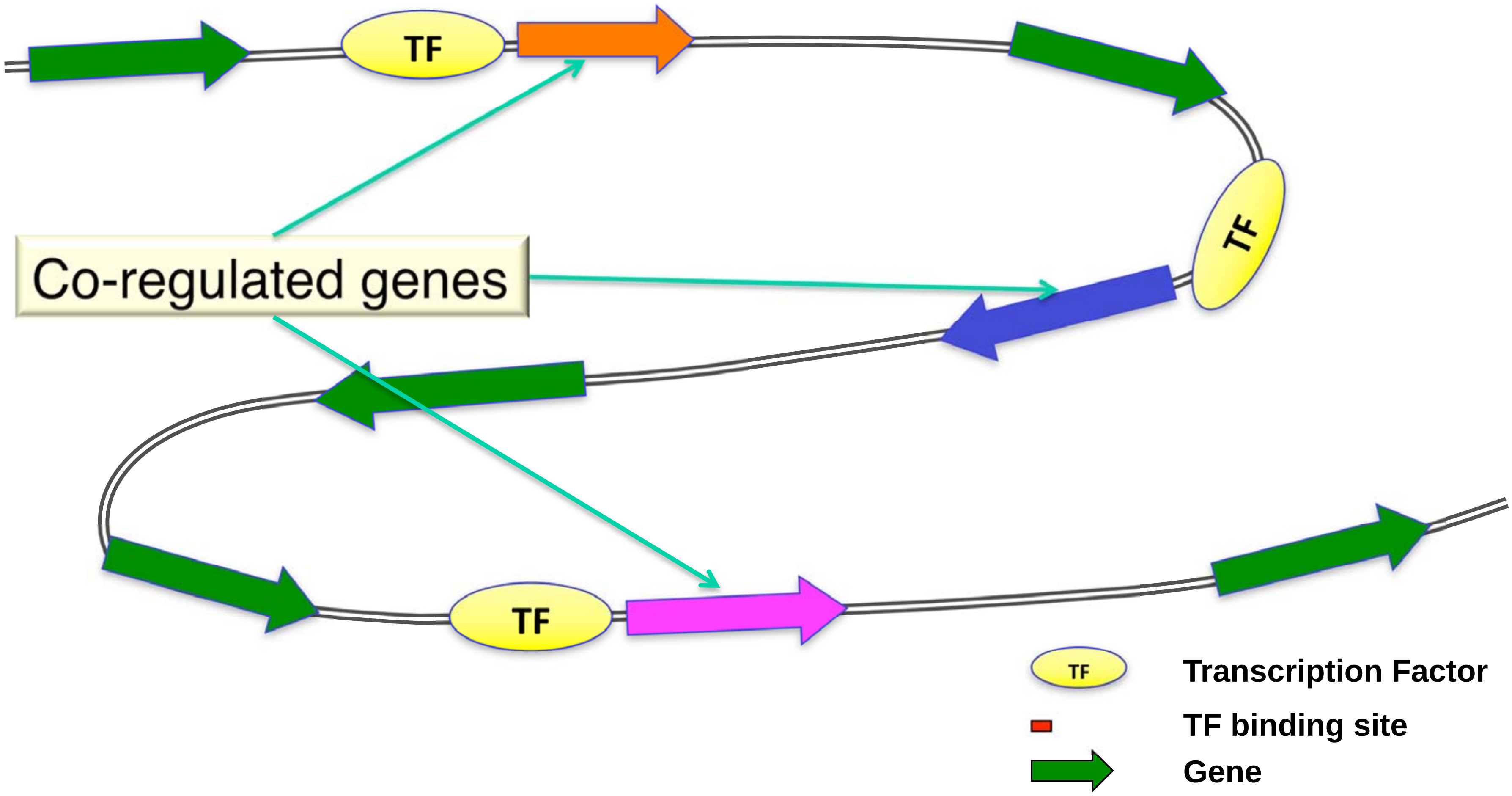


TF Transcription Factor

🟥 TF binding site

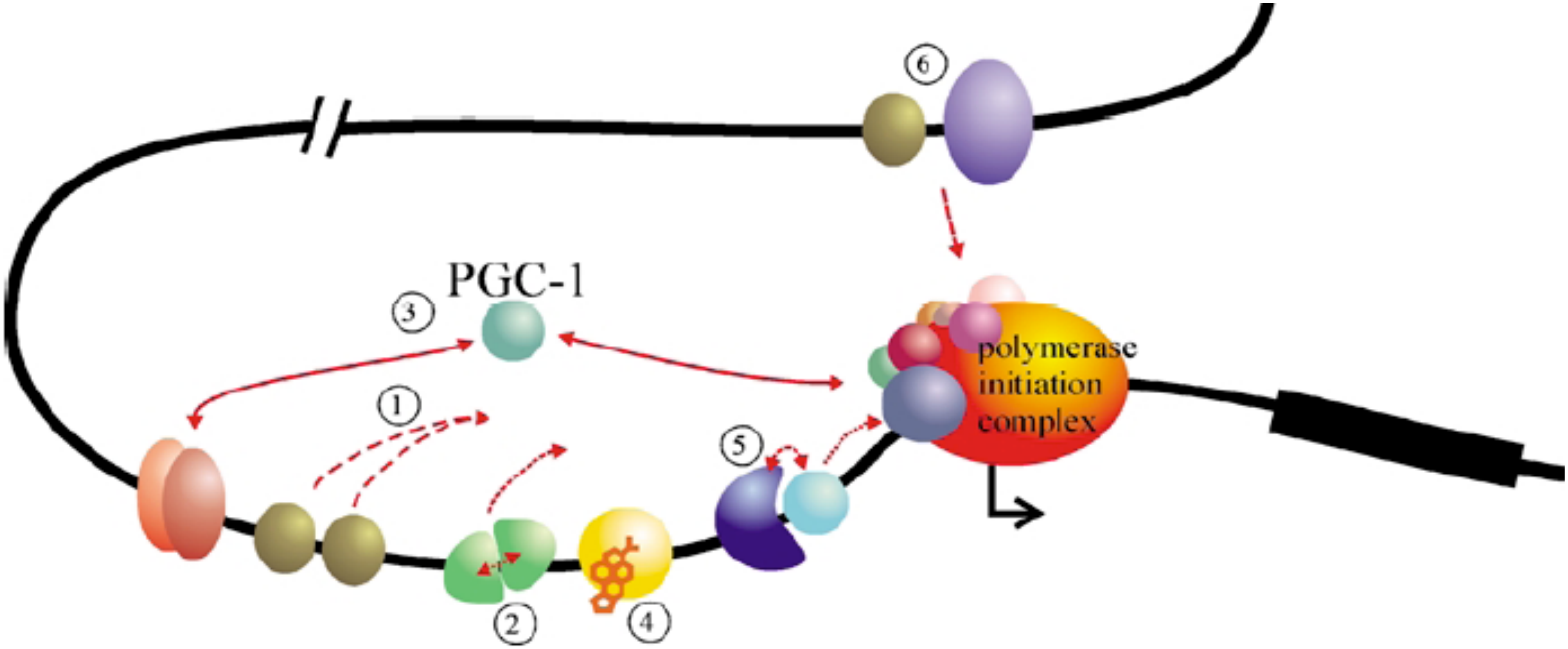
➡ Gene

Multiple Gene Activation



Co-regulated genes

Transcription Regulation

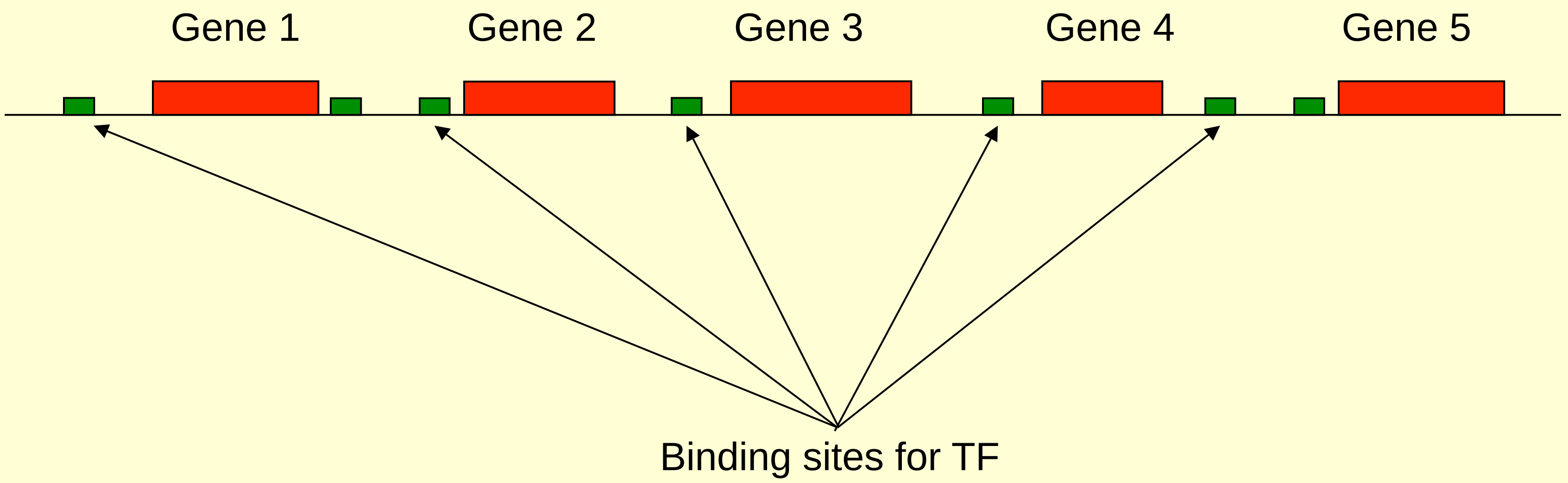


[Goffart *et al. Exp. Physiology* (2003)]

Motif-prediction: Whole genome

Problem: Given the upstream regions of all genes in the genome, find all over-represented sequence signatures.

Basic Principle: If a TF co-regulates many genes, then all these genes should have at least 1 binding site for it in their upstream region.

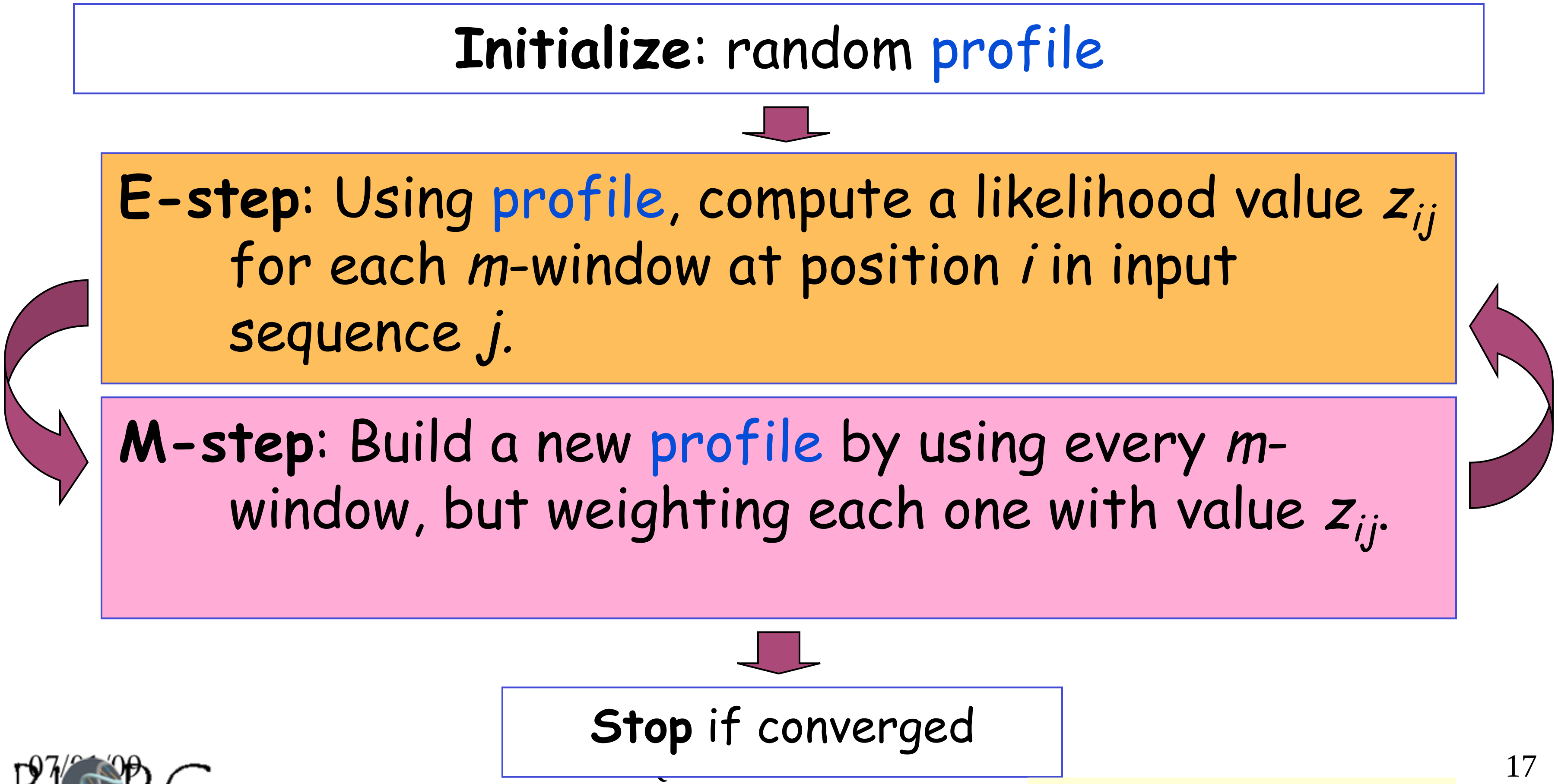


Motif Detection (TFBMs)

- See evaluation by Tompa et al.
  - [bio.cs.washington.edu/assessment]
- Gibbs Sampling Methods: AlignACE, GLAM, SeSiMCMC, MotifSampler
- Weight Matrix Methods: ANN-Spec, Consensus,
- EM: Improbizer, MEME
- Combinatorial & Misc.: MITRA, oligo/dyad, QuickScore, Weeder, YMF

EM Algorithm

Goal: Find θ, Z that maximize Pr (X, Z | θ)



Initialize: random profile

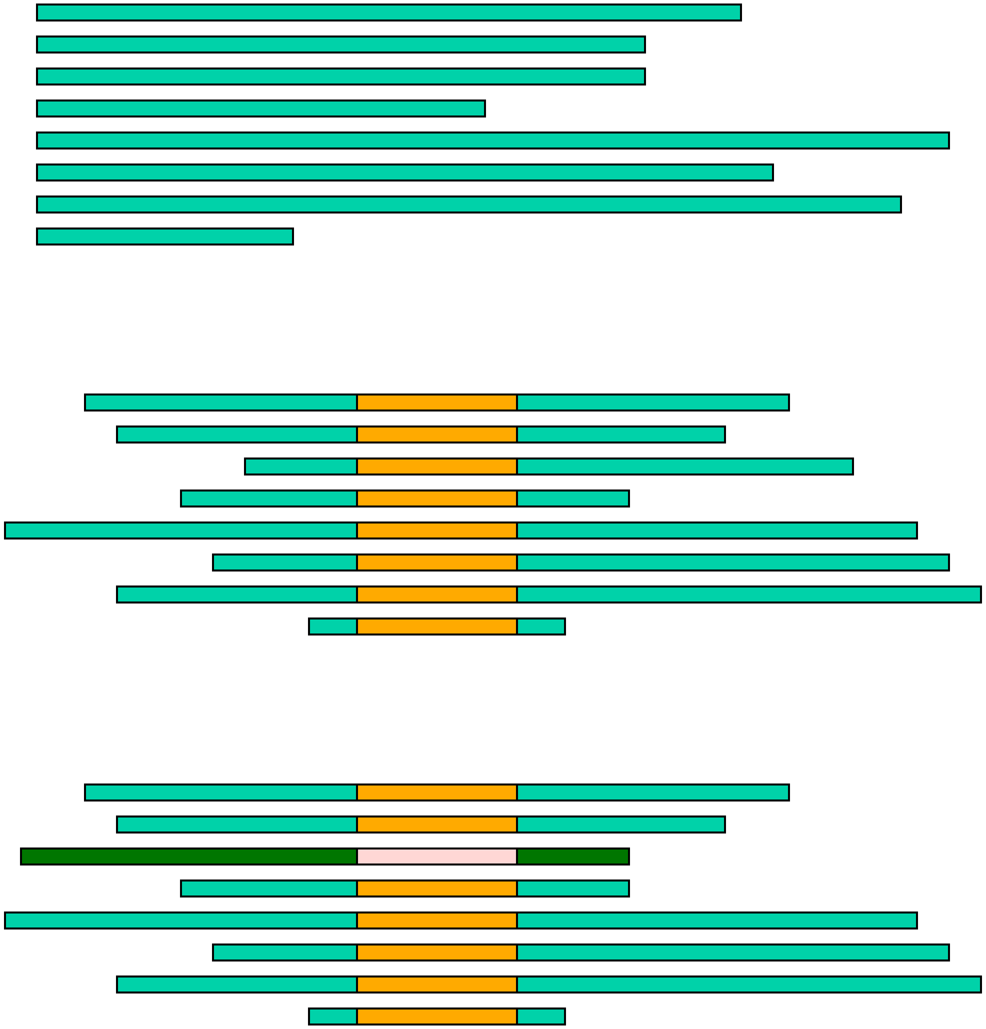
E-step: Using profile, compute a likelihood value z_{ij} for each *m*-window at position *i* in input sequence *j*.

M-step: Build a new profile by using every *m*-window, but weighting each one with value z_{ij}.

Stop if converged

MEME [Bailey, Elkan 1994]

Gibbs Sampling for Motif Detection



Gene Expression

- Process of transcription and/or translation of a gene is called gene expression.
- Every cell of an organism has the same genetic material, but different genes are expressed at different times.
- Patterns of gene expression in a cell is indicative of its state.

Hybridization

- If two complementary strands of DNA or mRNA are brought together under the right experimental conditions they will hybridize.
- A hybridizes to B ⇒
  - A is reverse complementary to B, or
  - A is reverse complementary to a subsequence of B.
- It is possible to experimentally verify whether A hybridizes to B, by labeling A or B with a radioactive or fluorescent tag, followed by excitation by laser.

Measuring gene expression

- Gene expression for a single gene can be measured by extracting mRNA from the cell and doing a simple hybridization experiment.
- Given a sample of cells, gene expression for every gene can be measured using a single microarray experiment.

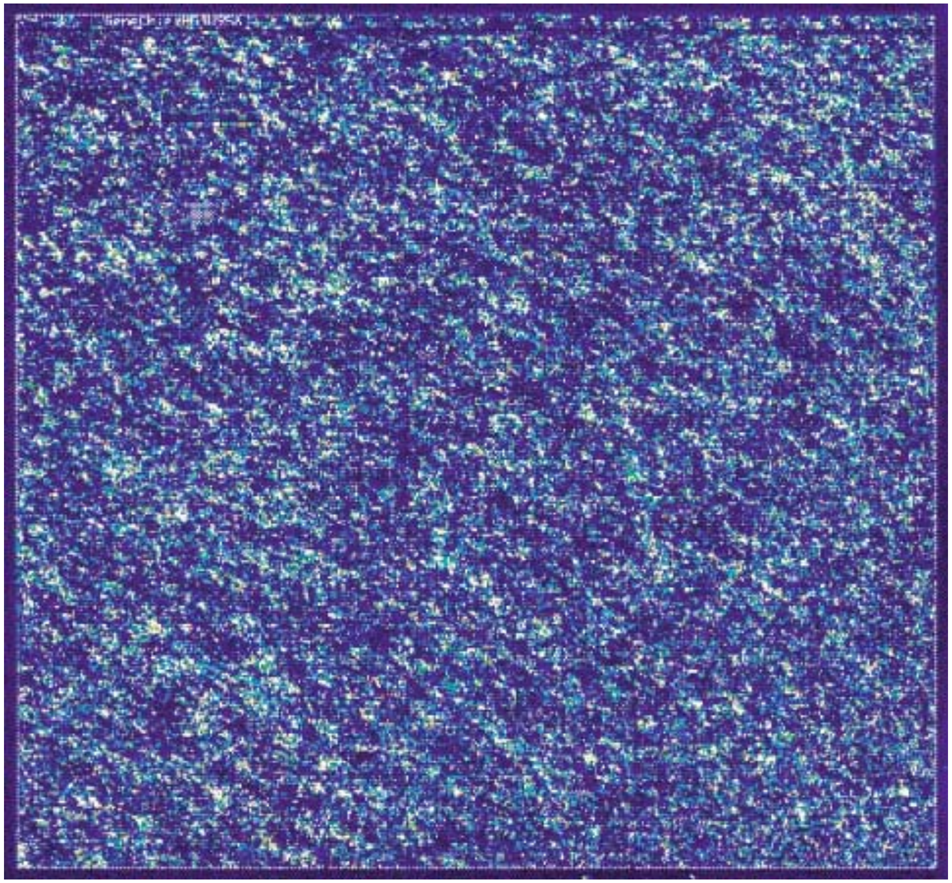
Microarray/DNA chip technology

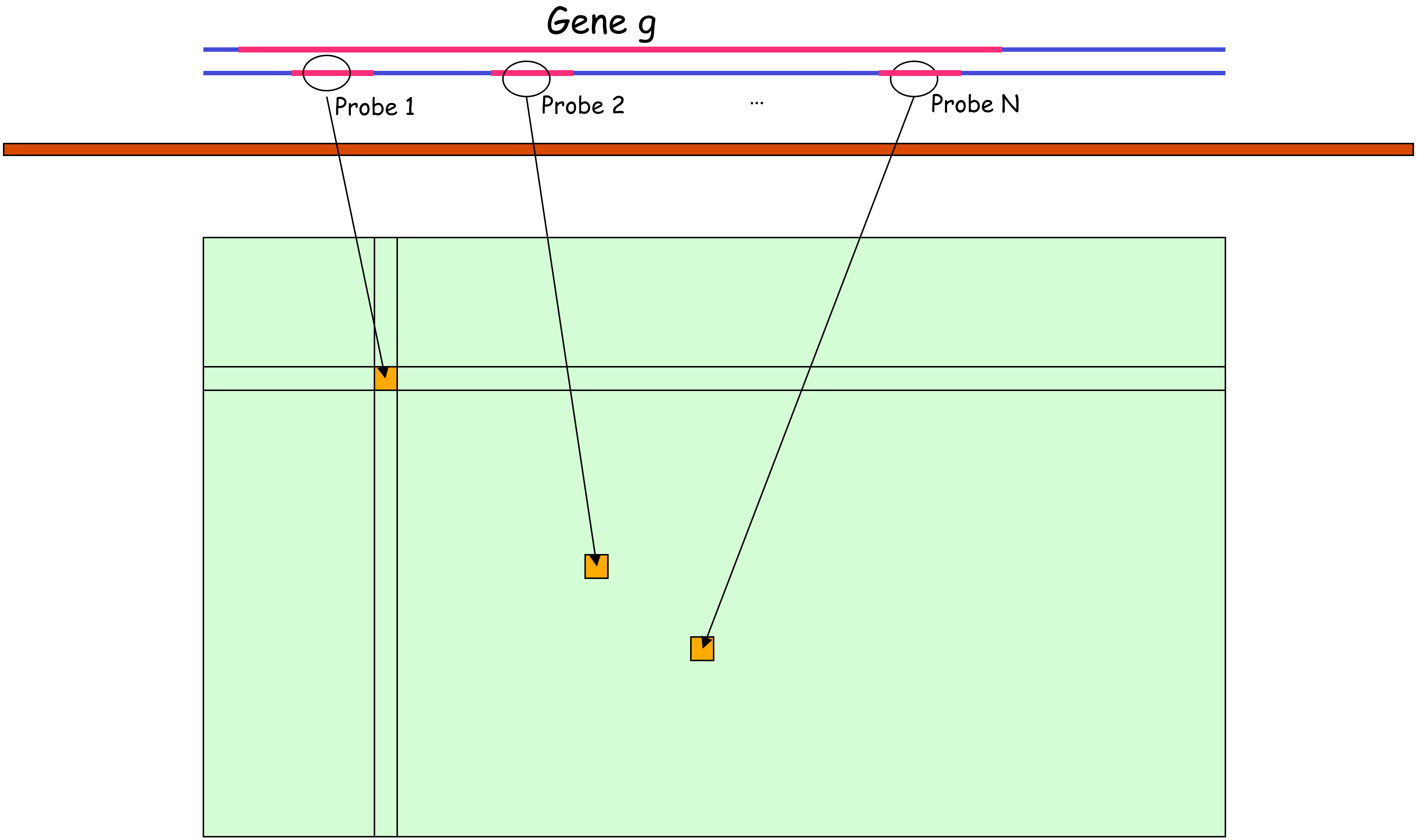
- High-throughput method to study gene expression of thousands of genes simultaneously.
- Many applications:
  - Genetic disorders & Mutation/polymorphism detection
  - Study of disease subtypes
  - Drug discovery & toxicology studies
  - Pathogen analysis
  - Differing expressions over time, between tissues, between drugs, across disease states

Microarray Data

Gene	Expression Level
Gene1	
Gene2	
Gene3	
…	

Gene Chips

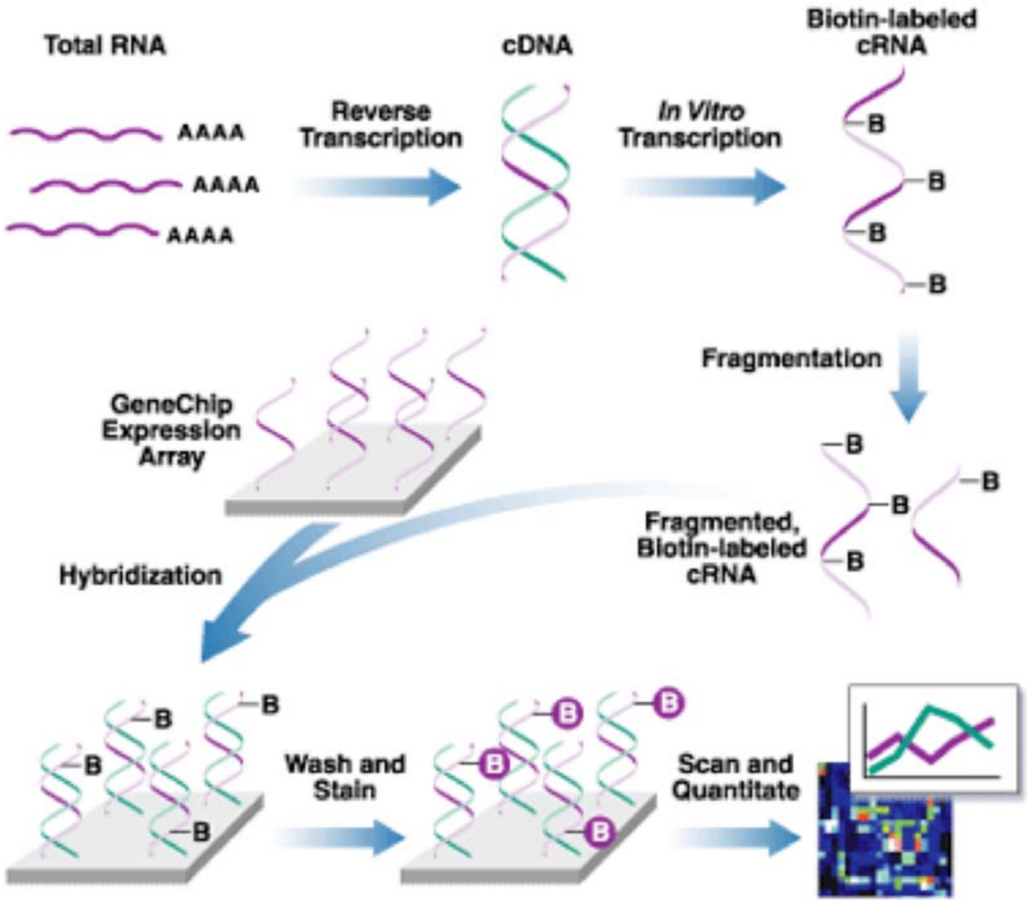



Gene g
Probe 1
Probe 2
...
Probe N

Microarray/DNA chips (Simplified)

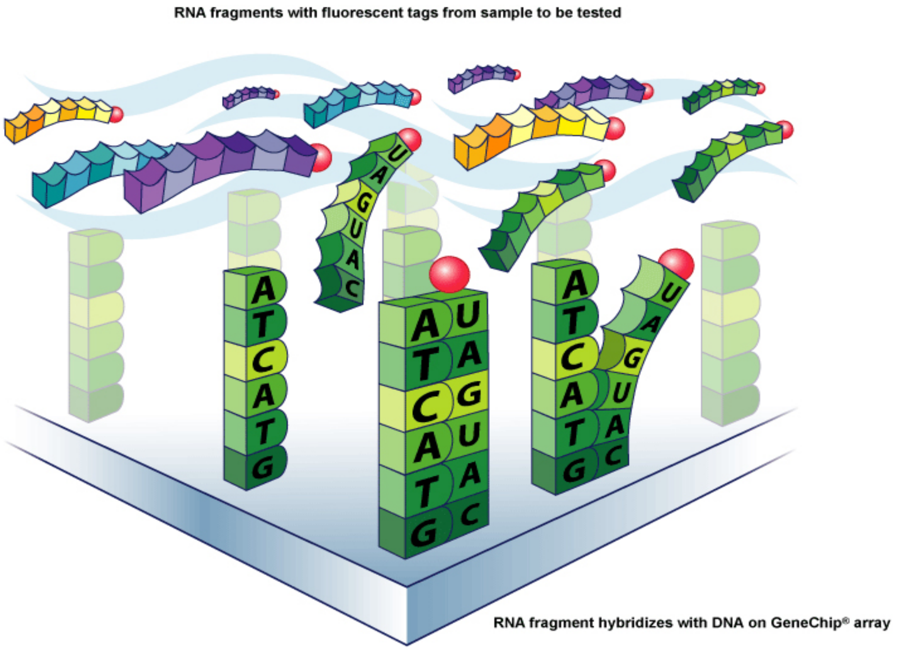
- Construct probes corresponding to reverse complements of genes of interest.
- Microscopic quantities of probes placed on solid surfaces at defined spots on the chip.
- Extract mRNA from sample cells and label them.
- Apply labeled sample (mRNA extracted from cells) to every spot, and allow hybridization.
- Wash off unhybridized material.
- Use optical detector to measure amount of fluorescence from each spot.

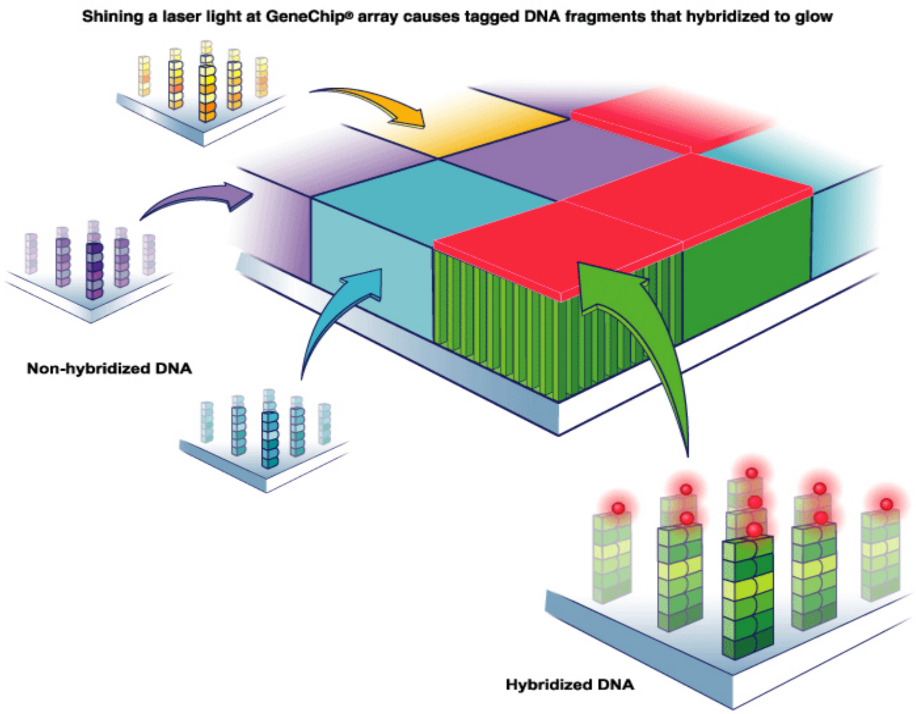
Affymetrix DNA chip schematic



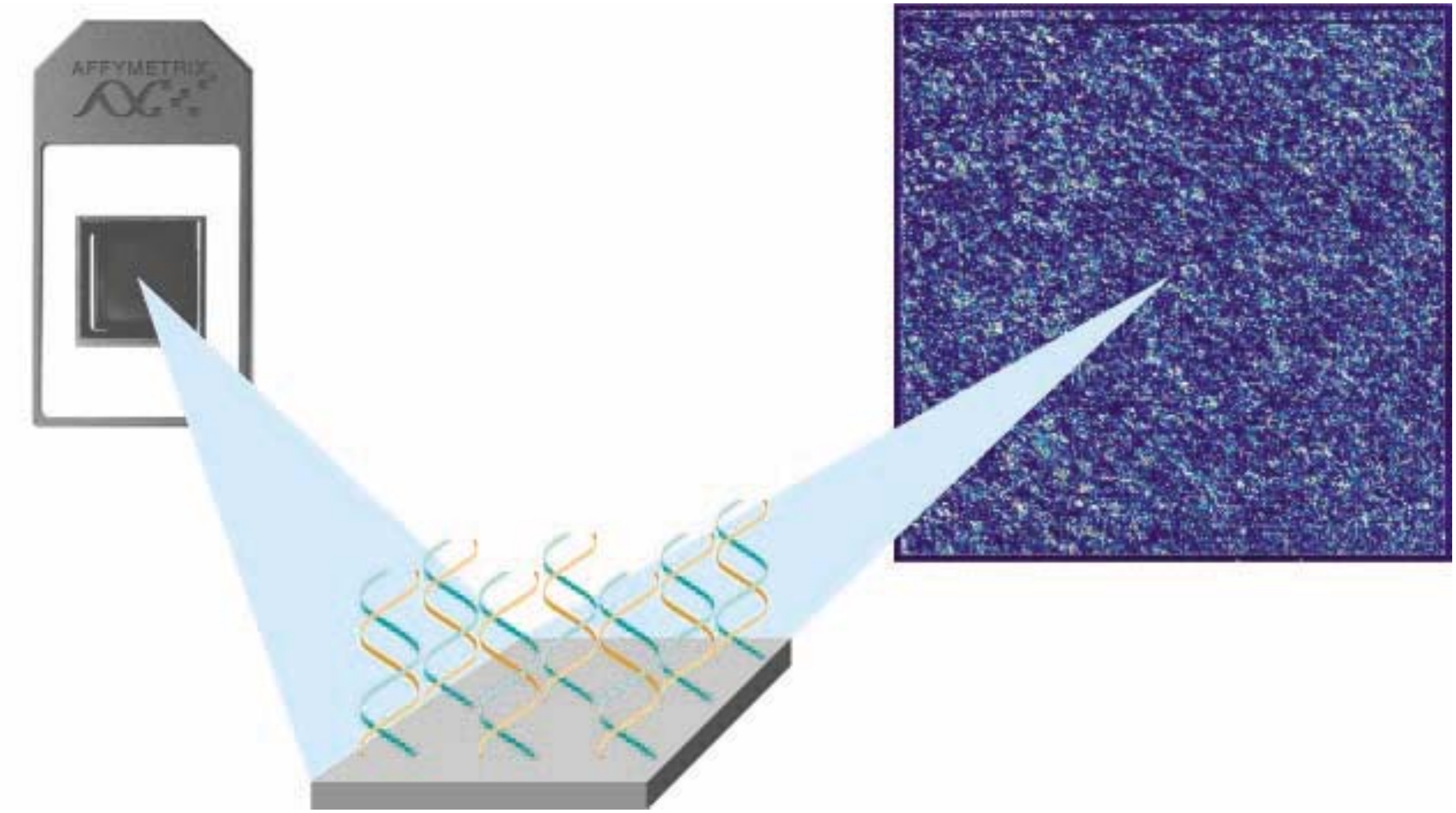
www.affymetrix.com

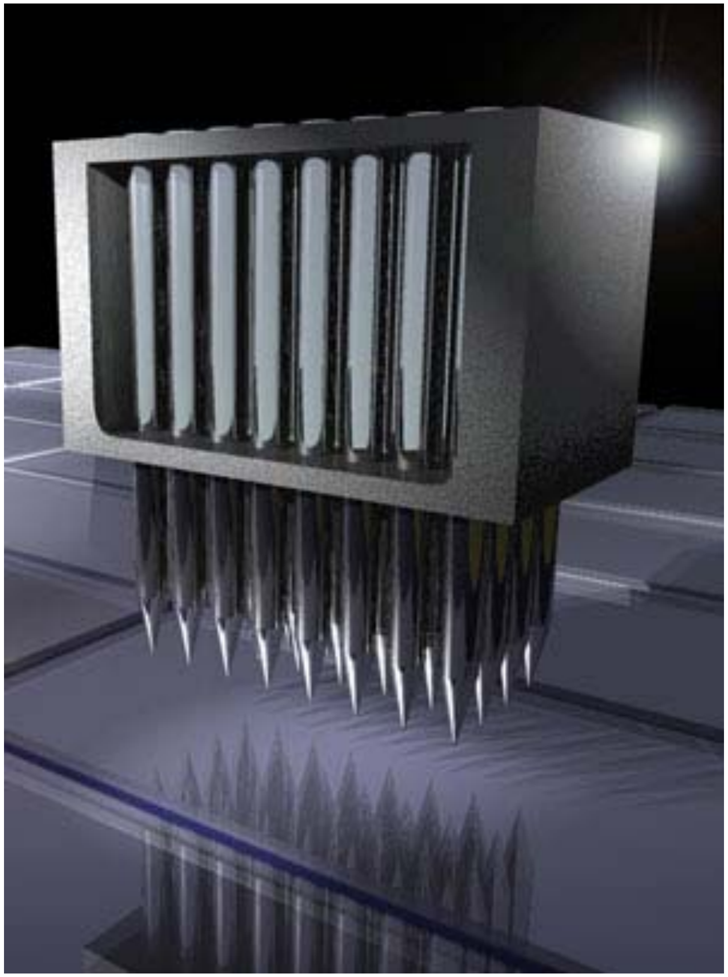
What’s on the slide?


RNA fragments with fluorescent tags from sample to be tested
RNA fragment hybridizes with DNA on GeneChip® array


Shining a laser light at GeneChip® array causes tagged DNA fragments that hybridized to glow
Non-hybridized DNA
Hybridized DNA

DNA Chips & Images

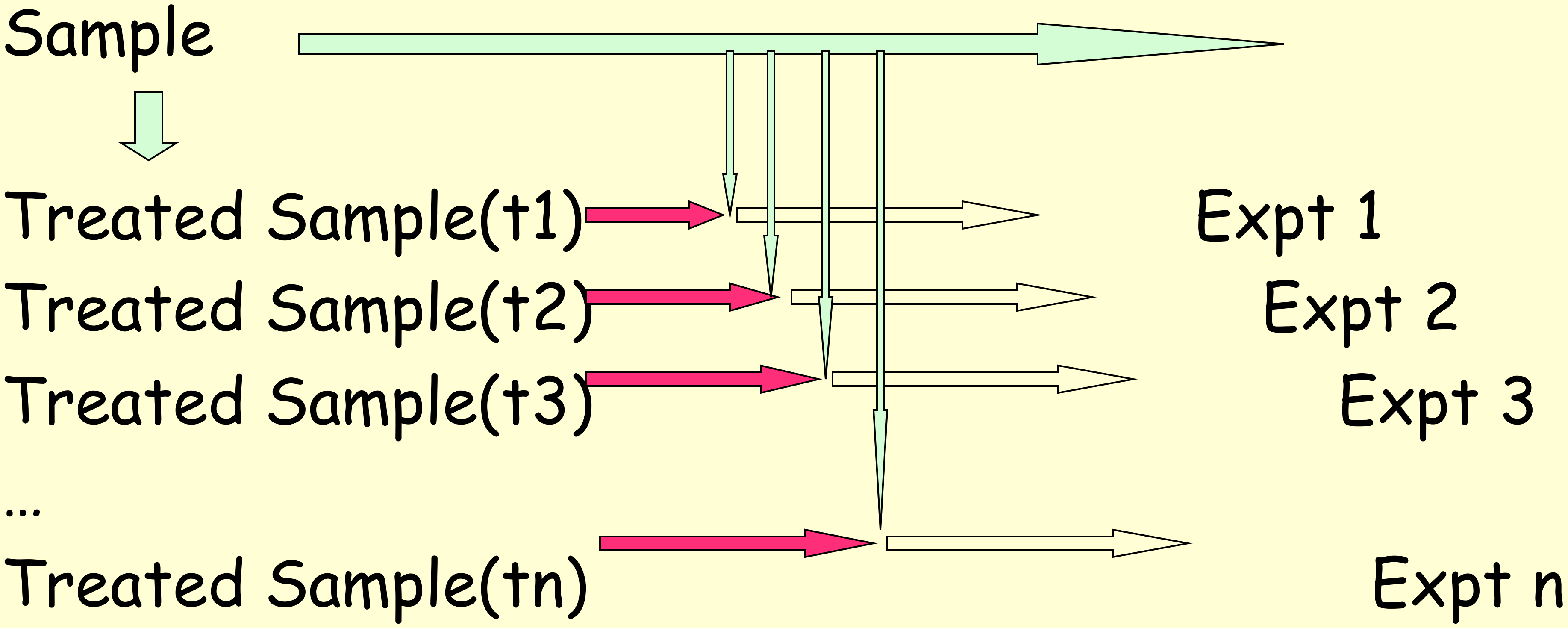




Microarrays: competing technologies

- Affymetrix & Agilent
- Differ in:
  - method to place DNA: Spotting vs. photolithography
  - Length of probe
  - Complete sequence vs. series of fragments

Study effect of treatment over time



Sample

Treated Sample(t1)

Treated Sample(t2)

Treated Sample(t3)

…

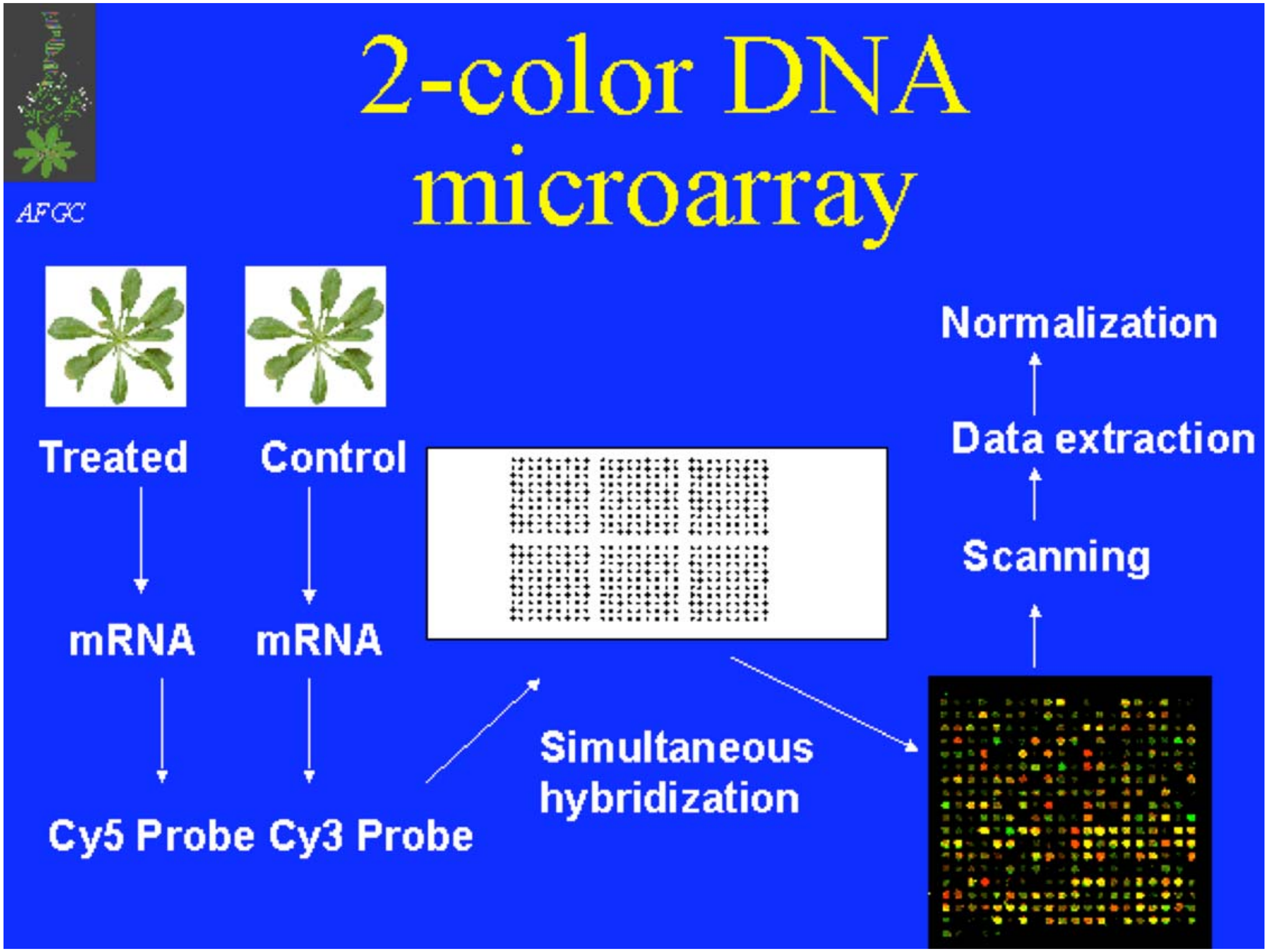
Treated Sample(tn)

Expt 1

Expt 2

Expt 3

Expt n



AFGC

2-color DNA microarray

Treated

Control

mRNA

mRNA

Cy5 Probe Cy3 Probe

Simultaneous hybridization

Scanning

Data extraction

Normalization

http://www.arabidopsis.org/info/2010_projects/comp_proj/AFGC/RevisedAFGC/Friday/

How to compare 2 cell samples with Two-Color Microarrays?

- ❑ mRNA from sample 1 is extracted and labeled with a red fluorescent dye.
- ❑ mRNA from sample 2 is extracted and labeled with a green fluorescent dye.
- ❑ Mix the samples and apply it to every spot on the microarray. Hybridize sample mixture to probes.
- ❑ Use optical detector to measure the amount of green and red fluorescence at each spot.

Sources of Variations & Experimental Errors

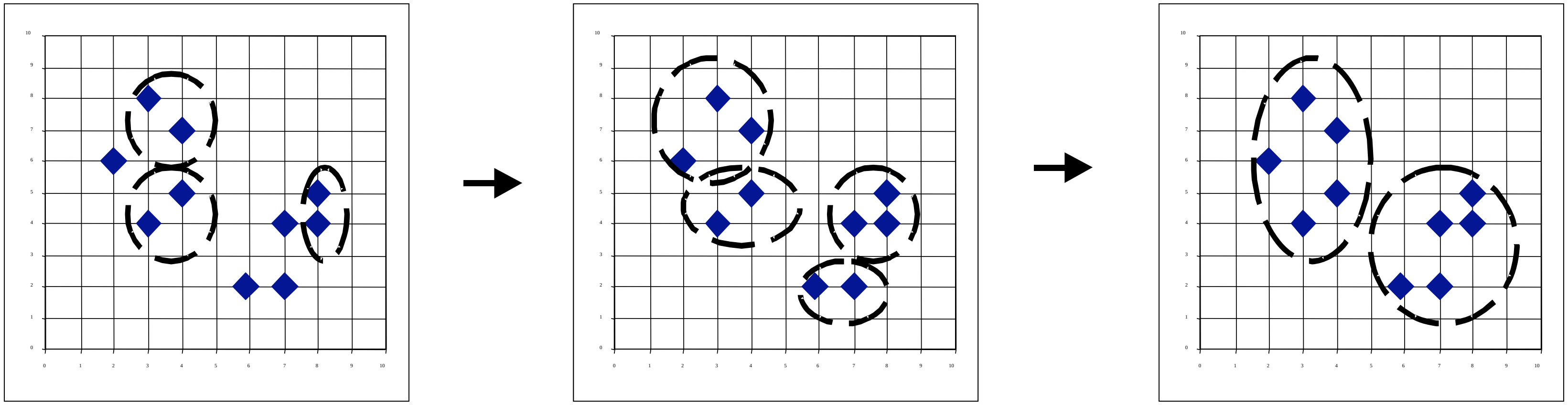
- Variations in cells/individuals
- Variations in mRNA extraction, isolation, introduction of dye, variation in dye incorporation, dye interference
- Variations in probe concentration, probe amounts, substrate surface characteristics
- Variations in hybridization conditions and kinetics
- Variations in optical measurements, spot misalignments, discretization effects, noise due to scanner lens and laser irregularities
- Cross-hybridization of sequences with high sequence identity
- Limit of factor 2 in precision of results
- Variation changes with intensity: larger variation at low or high expression levels

Need to Normalize data

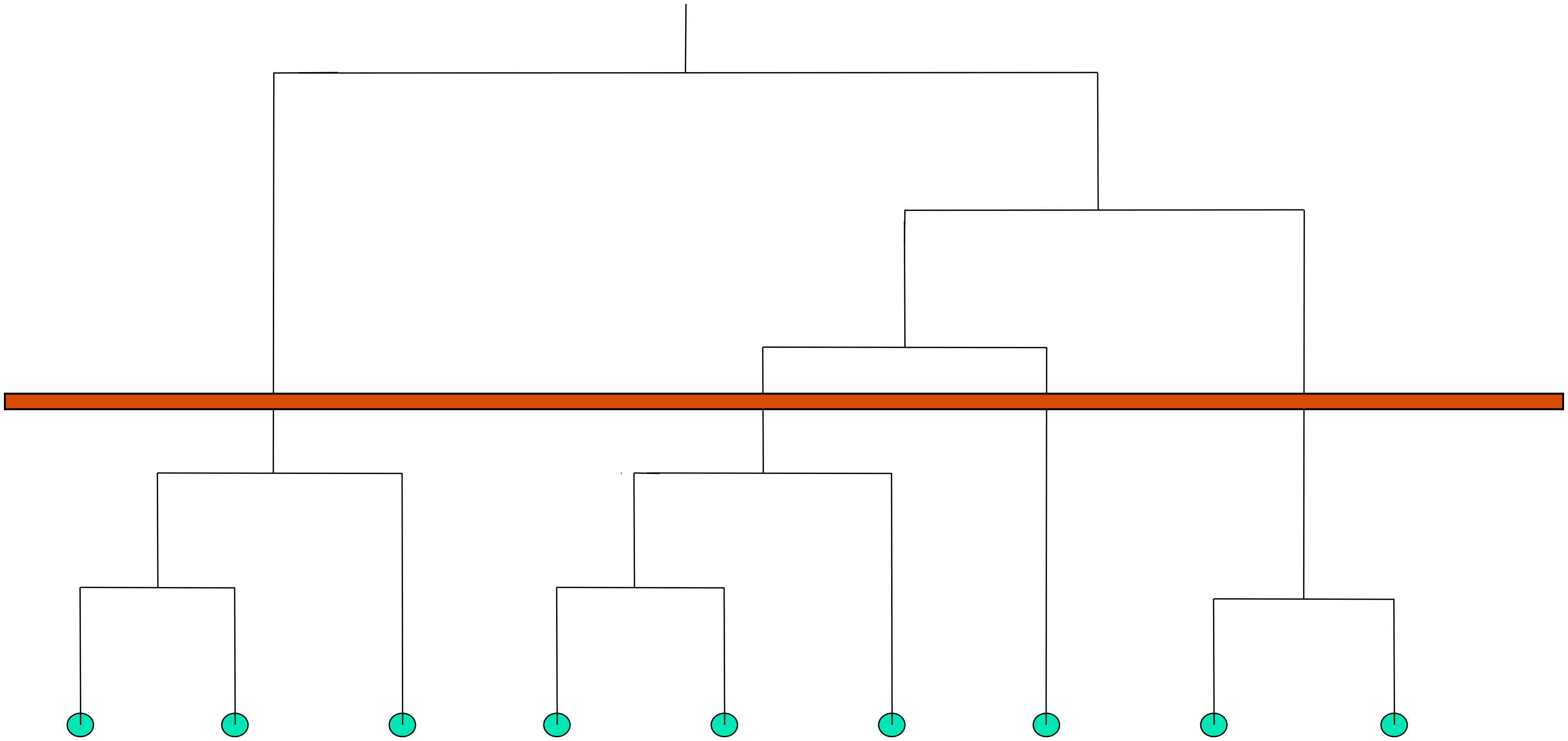
Clustering

- Clustering is a general method to study patterns in gene expressions.
- Several known methods:
  - Hierarchical Clustering (Bottom-Up Approach)
  - K-means Clustering (Top-Down Approach)
  - Self-Organizing Maps (SOM)

Hierarchical Clustering: Example



A Dendrogram



Hierarchical Clustering [Johnson, SC, 1967]

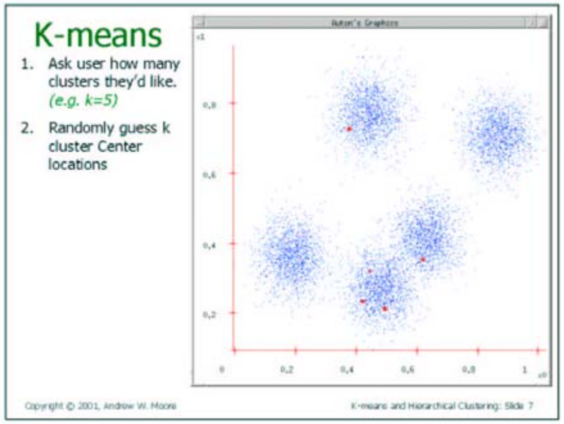
- Given n points in R^d, compute the distance between every pair of points
- While (not done)
  - Pick closest pair of points s_i and s_j and make them part of the same cluster.
  - Replace the pair by an average of the two s_{ij}

Try the applet at: http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

K-Means Clustering: Example

Example from Andrew Moore’s tutorial on Clustering.

Start

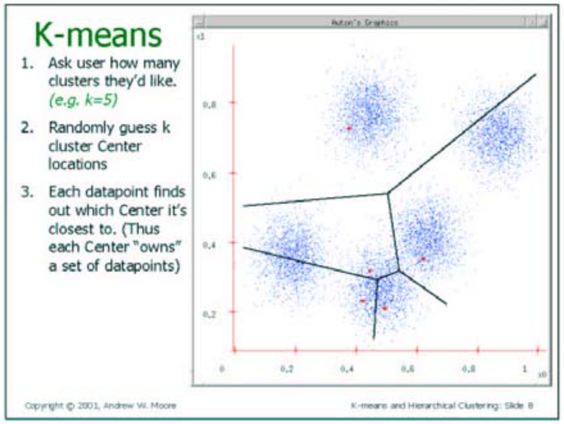


K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 7

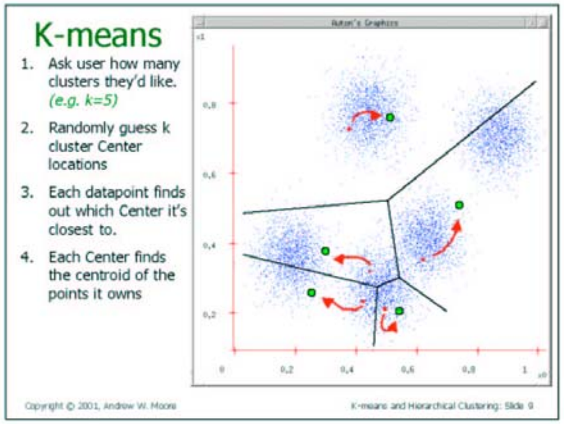


K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to. (Thus each Center "owns" a set of datapoints)

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 8

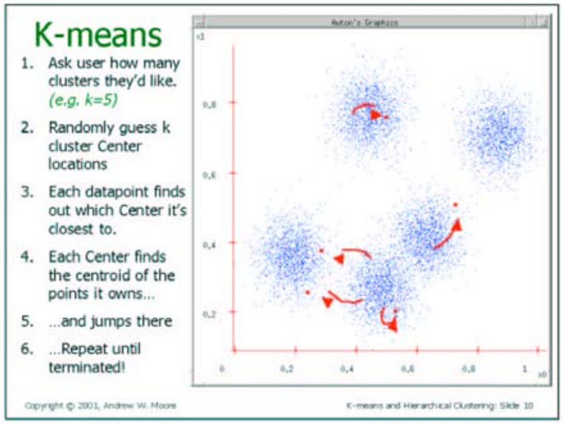


K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 9

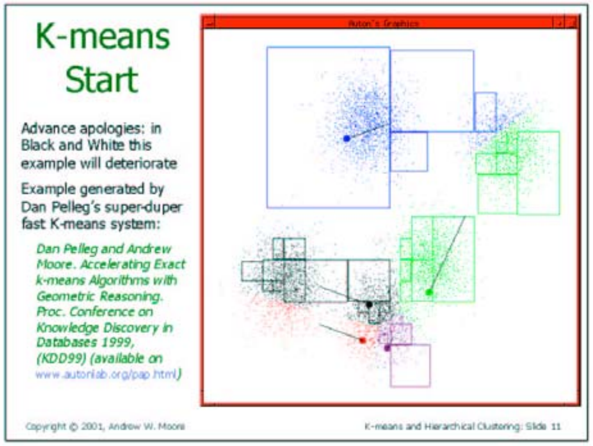


K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns...
5. ...and jumps there
6. ...Repeat until terminated!

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 10



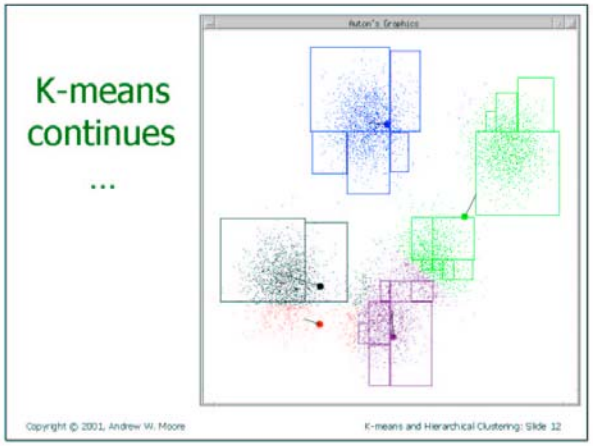
K-means Start

Advance apologies: in Black and White this example will deteriorate

Example generated by Dan Pelleg's super-duper fast K-means system:

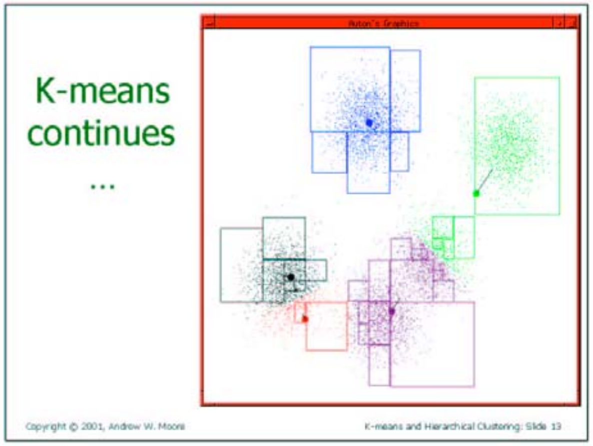
Dan Pelleg and Andrew Moore. Accelerating Exact k-means Algorithms with Geometric Reasoning. Proc. Conference on Knowledge Discovery in Databases 1999, (KDD99) (available on www.autonlab.org/pap.html*)*

K-means and Hierarchical Clustering: Slide 11



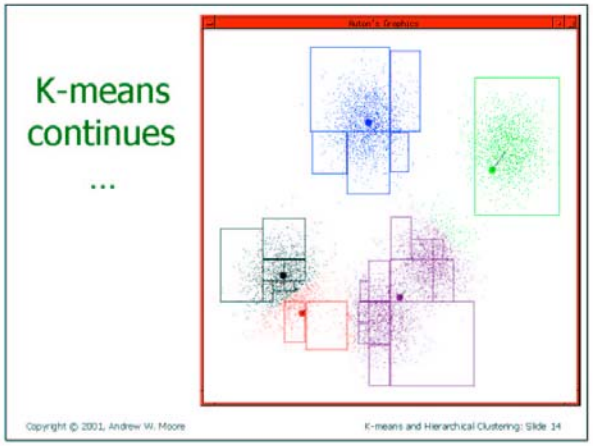
K-means continues …

K-means and Hierarchical Clustering: Slide 12



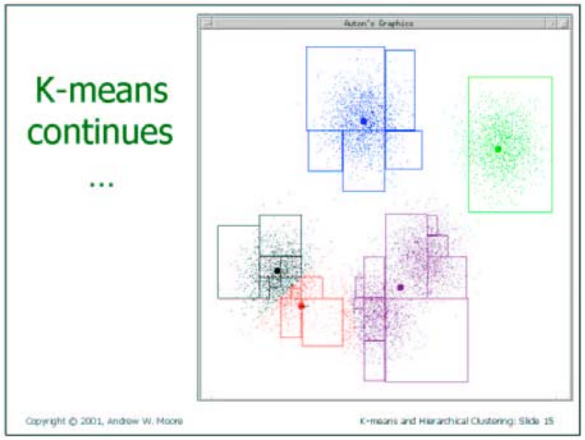
K-means continues …

K-means and Hierarchical Clustering: Slide 13



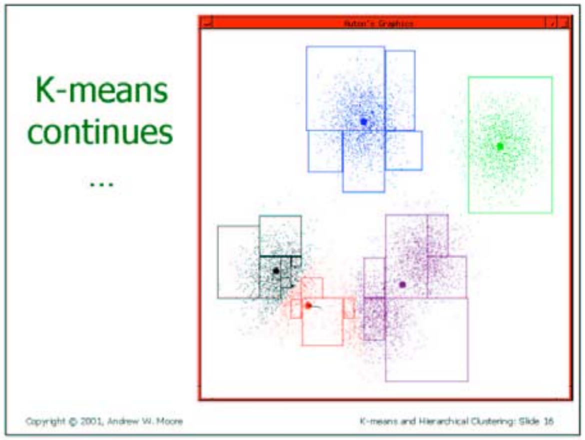
K-means continues …

K-means and Hierarchical Clustering: Slide 14



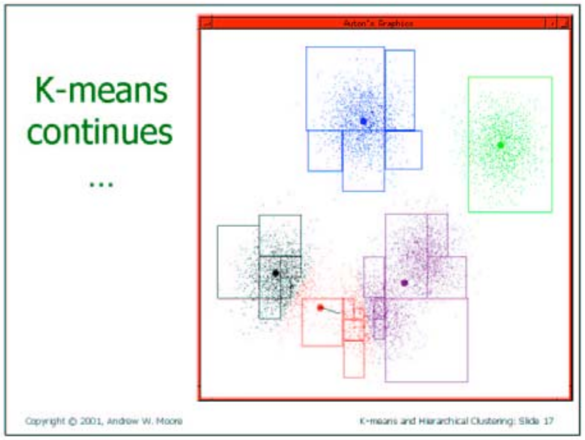
K-means continues …

K-means and Hierarchical Clustering: Slide 15



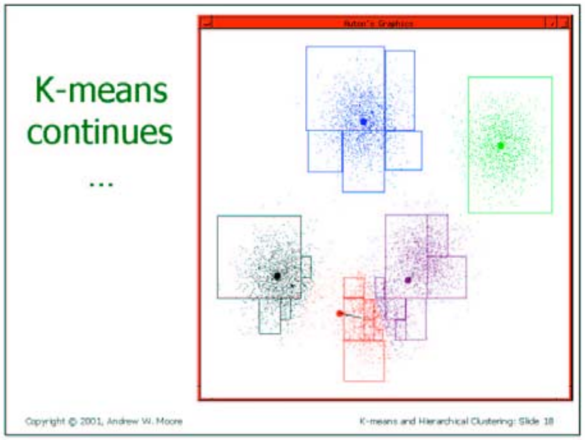
K-means continues …

K-means and Hierarchical Clustering: Slide 16



K-means continues …

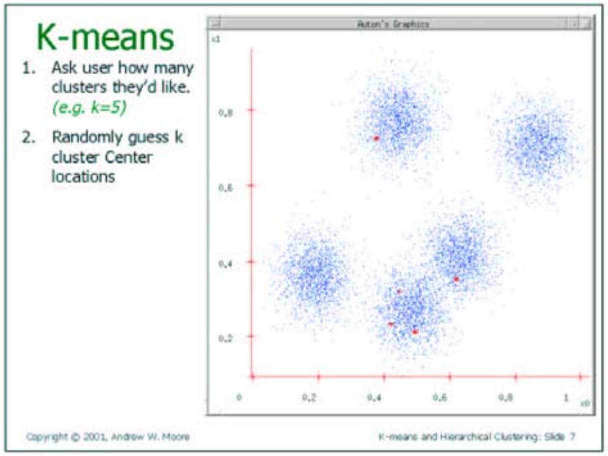
K-means and Hierarchical Clustering: Slide 17



K-means continues …

K-means and Hierarchical Clustering: Slide 18

Start

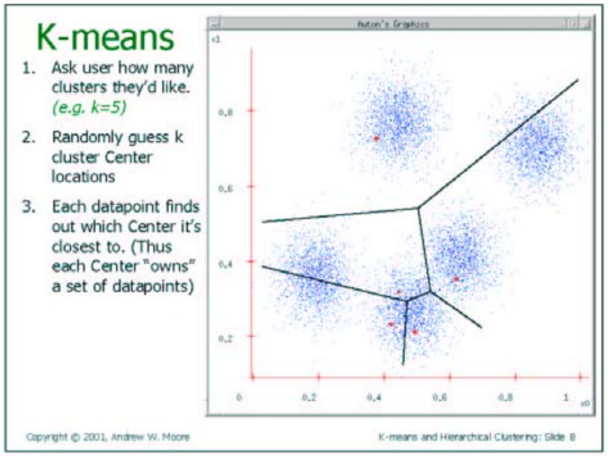


K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 7



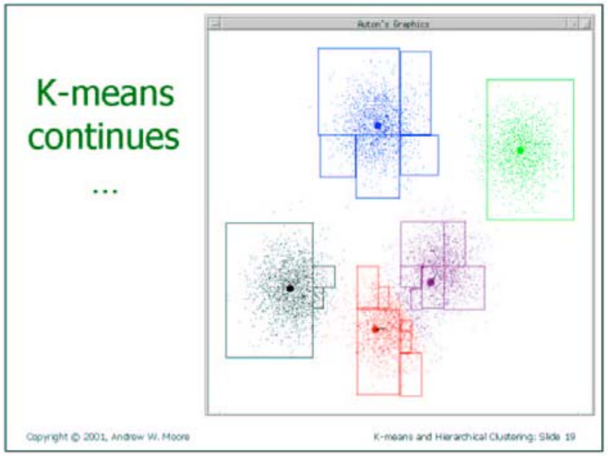
K-means

1. Ask user how many clusters they'd like. *(e.g. k=5)*
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to. (Thus each Center "owns" a set of datapoints)

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 8

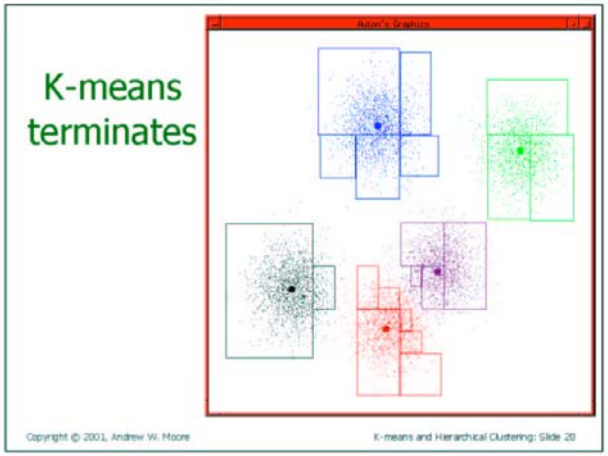
4



K-means continues …

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 19



K-means terminates

Copyright © 2001, Andrew W. Moore

K-means and Hierarchical Clustering: Slide 20

End

10

K-Means Clustering [McQueen ’67]

Repeat

- Start with randomly chosen cluster centers
- Assign points to give greatest increase in score
- Recompute cluster centers
- Reassign points

until (no changes)

Try the applet at: http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

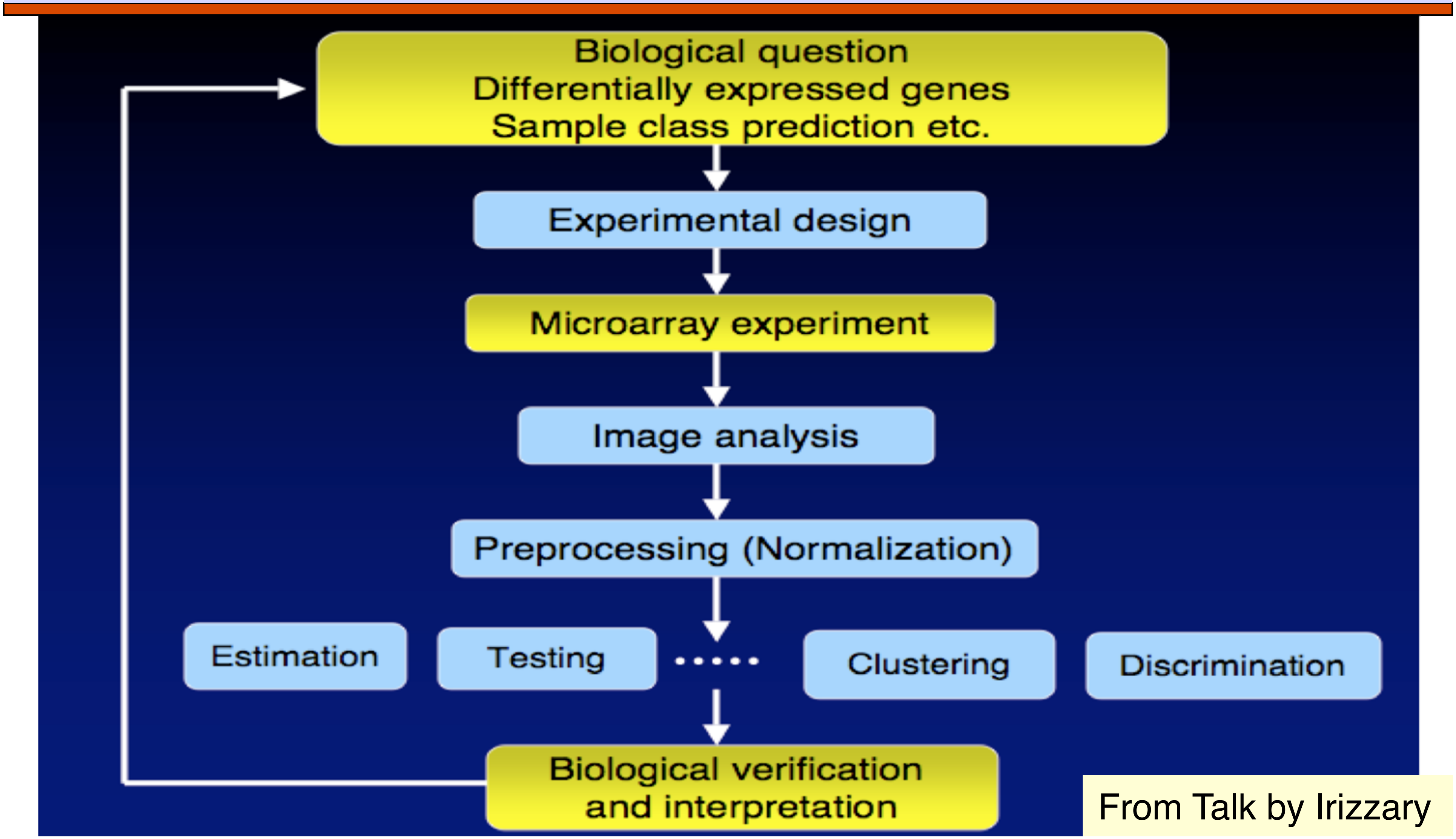
Comparisons

- Hierarchical clustering
  - Number of clusters not preset.
  - Complete hierarchy of clusters
  - Not very robust, not very efficient.
- K-Means
  - Need definition of a mean. Categorical data?
  - More efficient and often finds optimum clustering.

Reading

- The following slides come from a series of talks by Rafael Irizzary from Johns Hopkins
- Much of the material can be found in detail in the following papers from [http://www.biostat.jhsph.edu/~ririzarr/papers/]
  - Irizarry, RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Scherf, U, Speed, TP (2003) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistics. Vol. 4, Number 2: 249-264.
  - Bolstad, B.M., Irizarry RA, Astrand, M, and Speed, TP (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. Bioinformatics. 19(2):185-193.

Inference Process


Biological question
Differentially expressed genes
Sample class prediction etc.

Experimental design

Microarray experiment

Image analysis

Preprocessing (Normalization)

Estimation

Testing

· · · · ·

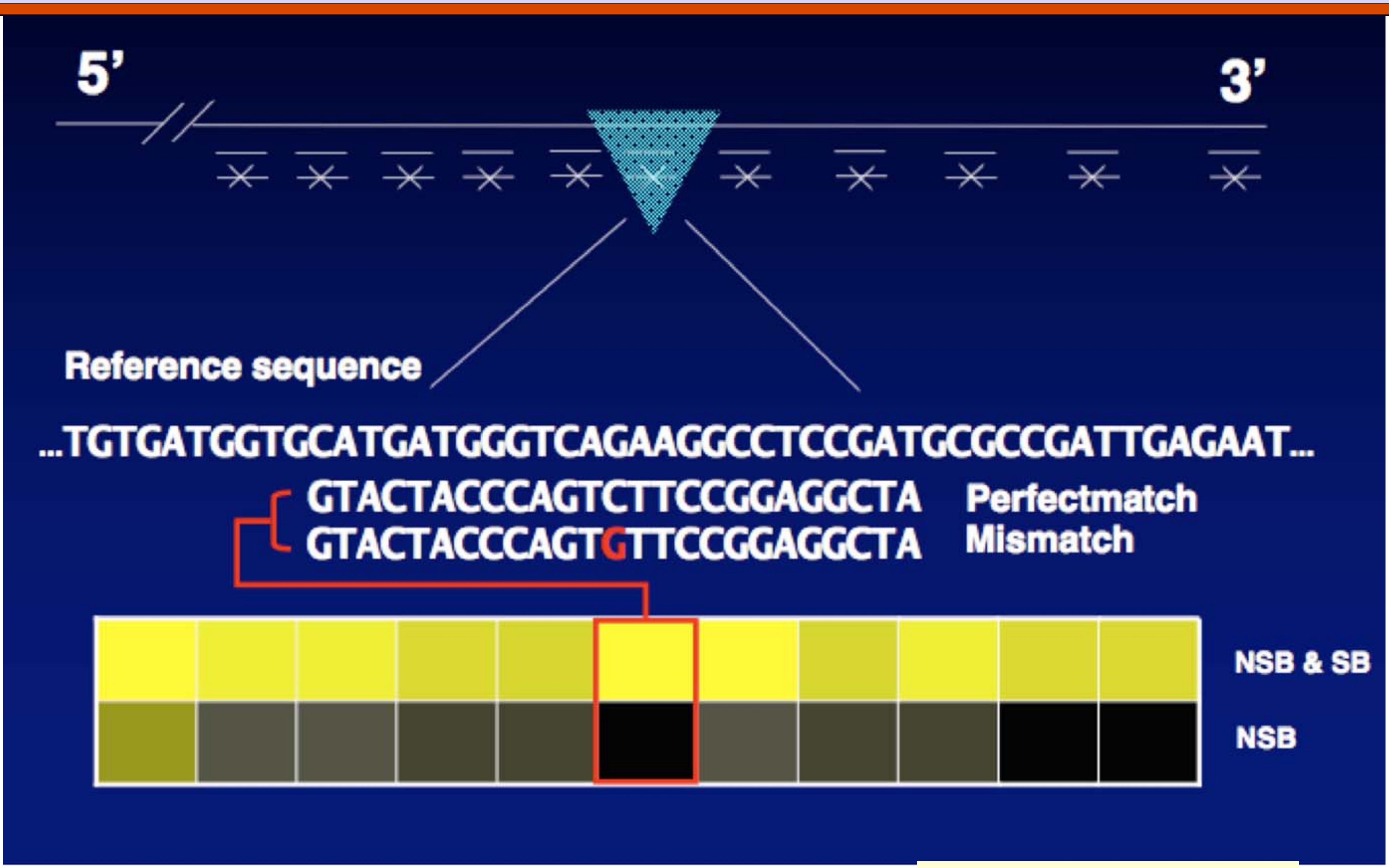
Clustering

Discrimination

Biological verification
and interpretation

From Talk by Irizzary

Affymetrix Genechip Design



Reference sequence

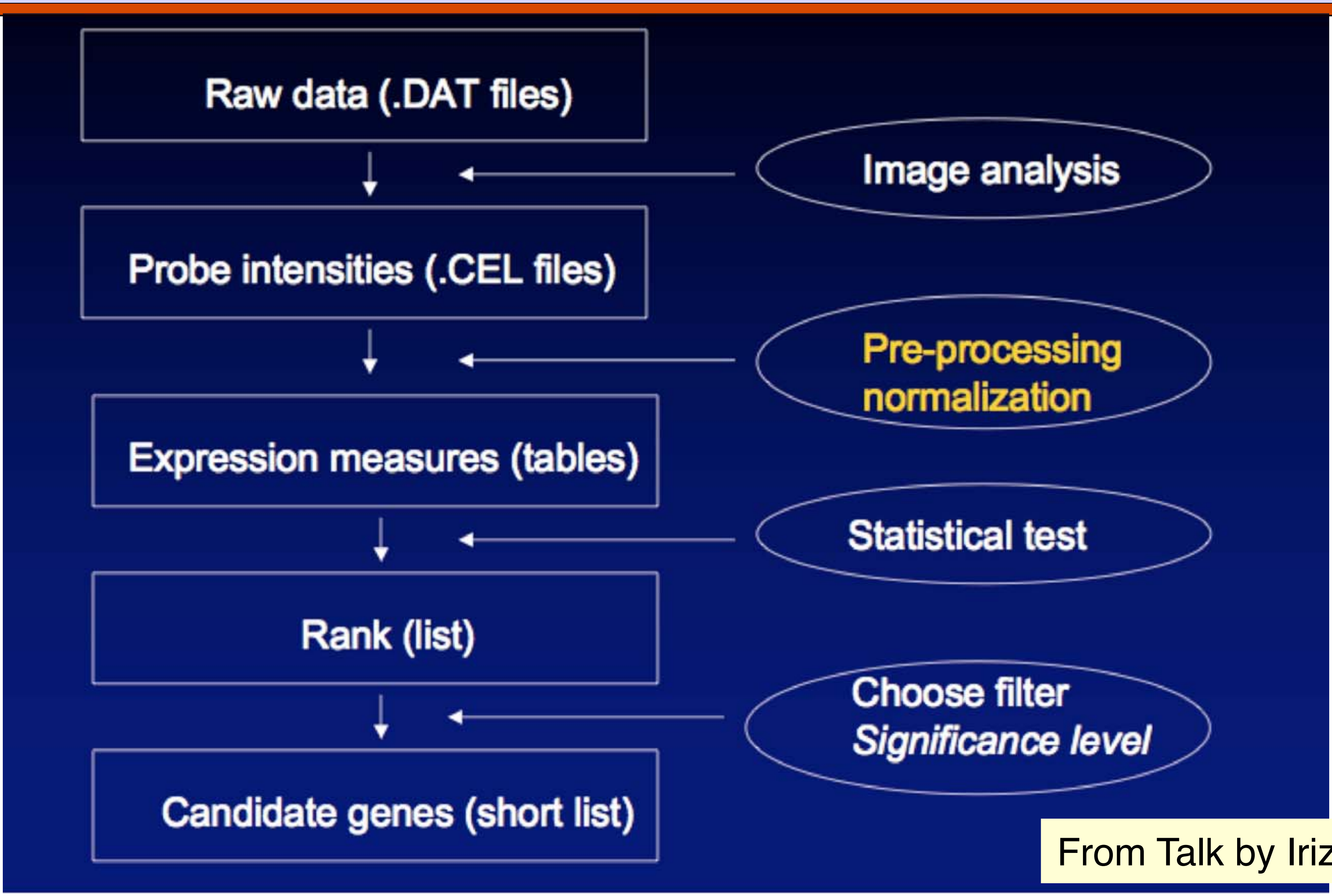
...TGTGATGGTGCATGATGGGTCAGAAGGCCTCCGATGCGCCGATTGAGAAT...

GTACTACCCAGTCTTCCGGAGGCTA Perfectmatch

GTACTACCCAGTGTTCCGGAGGCTA Mismatch

From Talk by Irizzary

Workflow: Analyzing Affy data



Raw data (.DAT files)

Image analysis

Probe intensities (.CEL files)

Pre-processing
normalization

Expression measures (tables)

Statistical test

Rank (list)

Choose filter
Significance level

Candidate genes (short list)

From Talk by Irizzary

Affy Files

- **DAT** file: image file, about 10 million pixels, 30-50 MB
- **CEL** file: cell intensity file with probe level PM and MM values
- **CDF** file: chip description file describing which probes go in which probe sets and the location of probe-pair sets (genes, gene fragments, ESTs)

From Talk by Irizzary

Image analysis & Background Correction

- Each probe cell: 10 X 10 pixels
- Gridding estimates location of probe cell centers
- Signal is computed by
  - Ignoring outer 36 pixels leaving a 8 X 8 pixel area
  - Taking the 75 percentile of the signal from the 8 X 8 pixel area
- Background signal is computed as the average of the lowest 2% probe cell values, which is then subtracted from the individual signals

From Talk by Irizzary

Standard Normalization Procedure

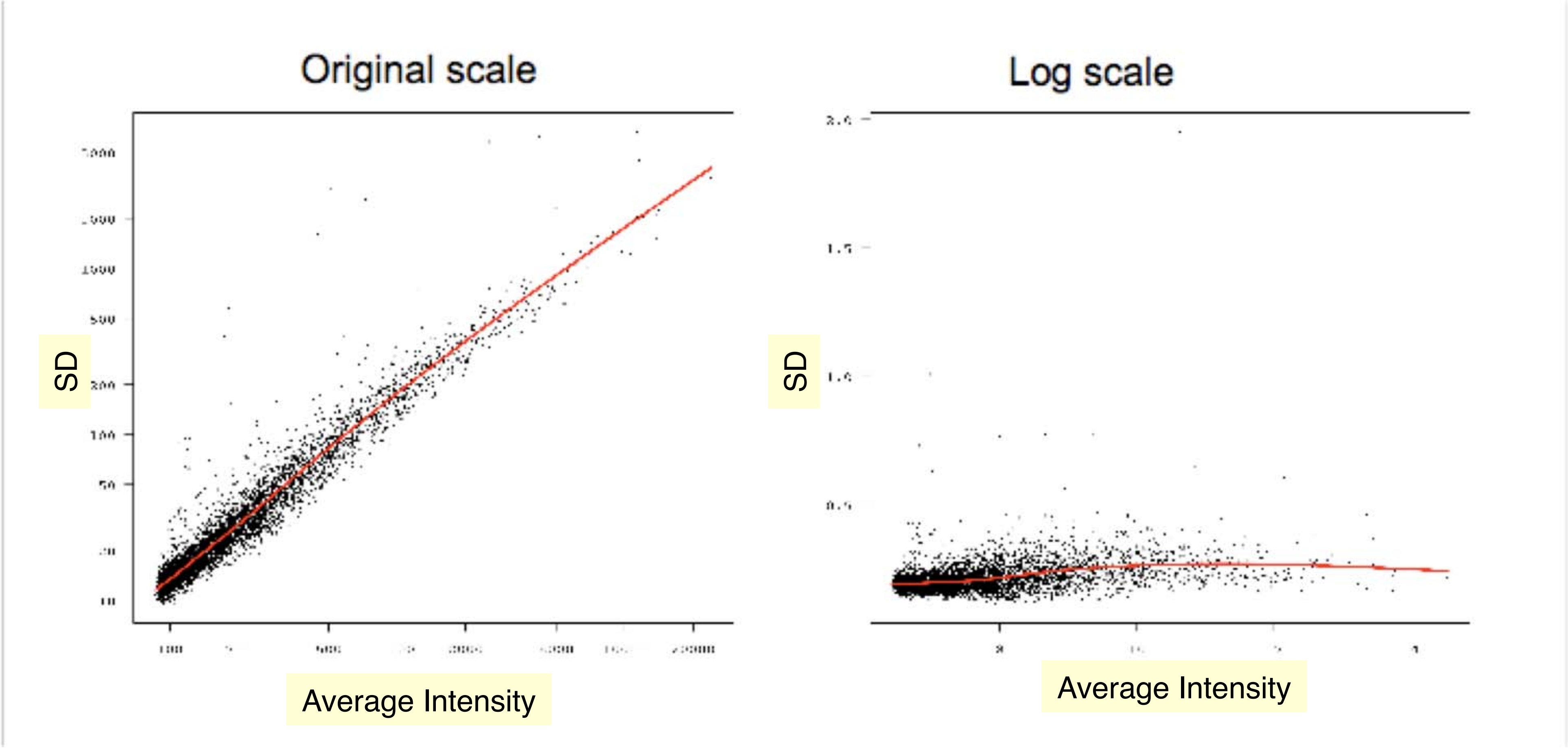
- ❑ Log-transform the data
- ❑ Ensure that the average intensity and the standard deviation are the same across all arrays.
- ❑ This requires the choice of a baseline array, which may or may not be obvious.

Analyzing Affy data

- MAS 4.0
  - Works with PM-MM
  - Negative values result very often
  - Very noisy for low expressed genes
  - Averages without log-transformation
- dChip [Li & Wong, PNAS **98**(1):31-36]
  - Accounts for probe effect
  - Uses non-linear normalization
  - Multi-chip analysis reveals outliers
- MAS 5.0
  - Improves on problems with MAS 4.0

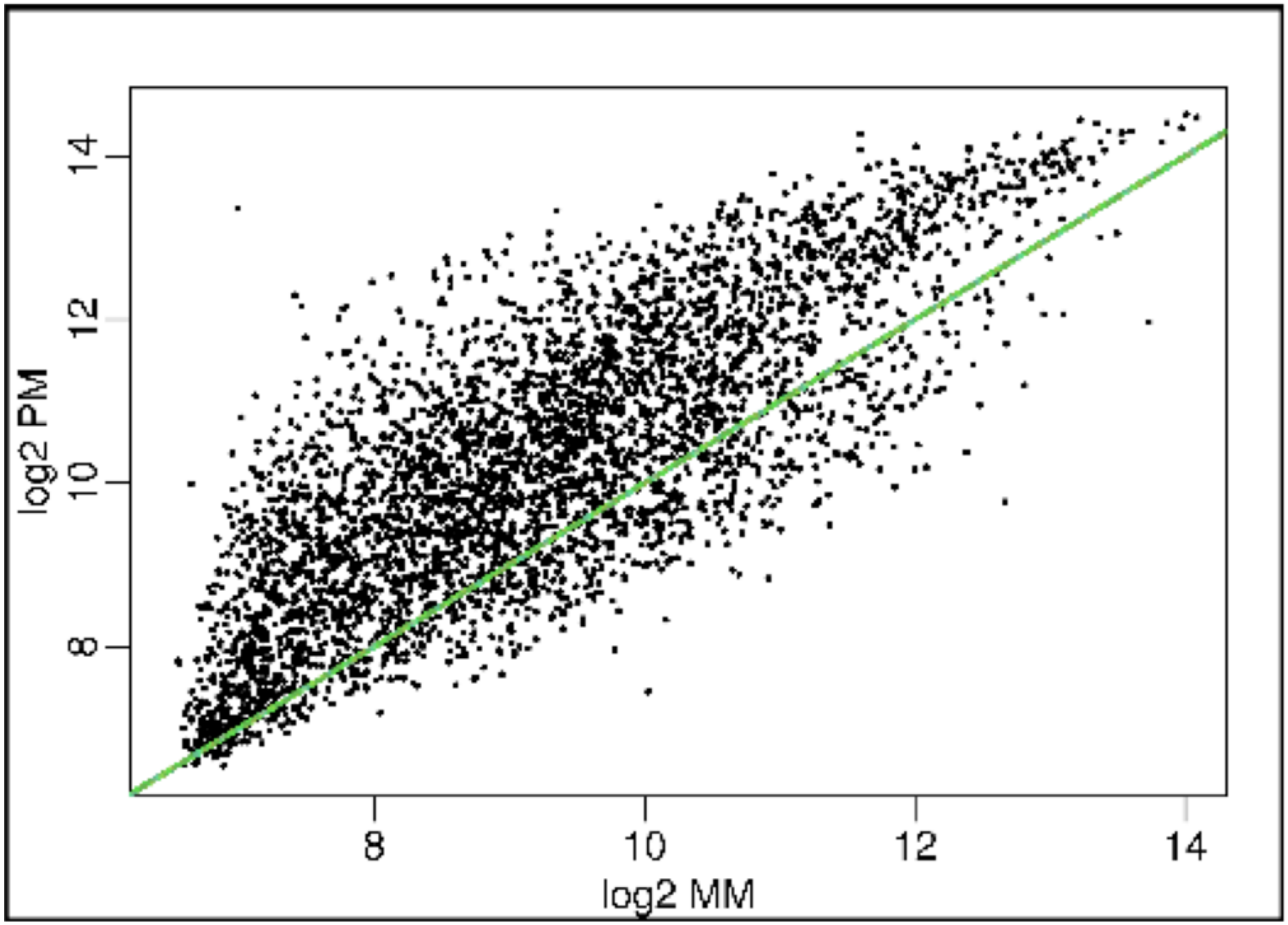
From Talk by Irizzary

Why you use log-transforms?



From Talk by Irizzary

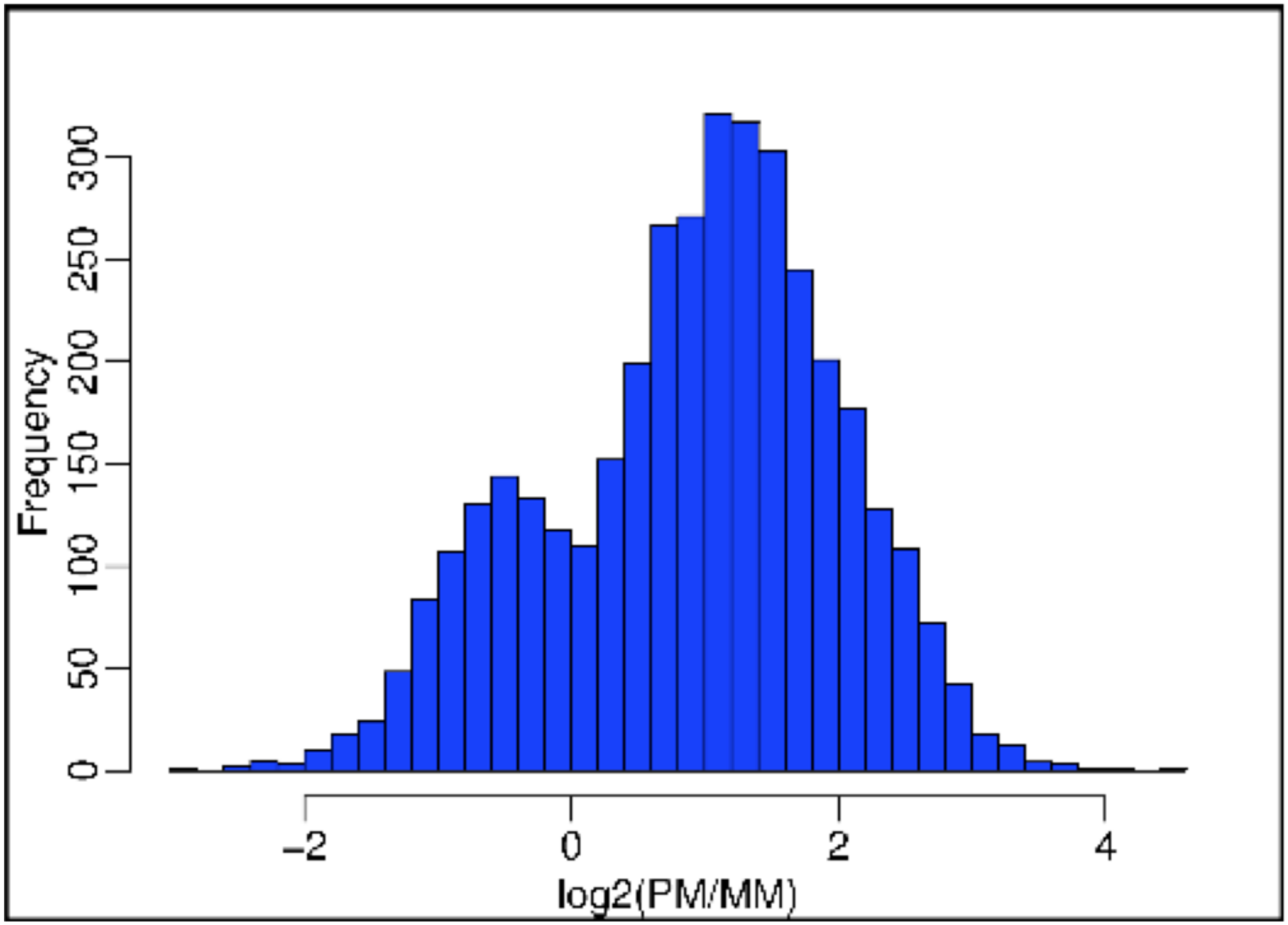
Problem with using (transformed) PM-MM



Sometimes MM is larger than PM!

From Talk by Irizzary

Bimodality for large expression values



From Talk by Irizzary

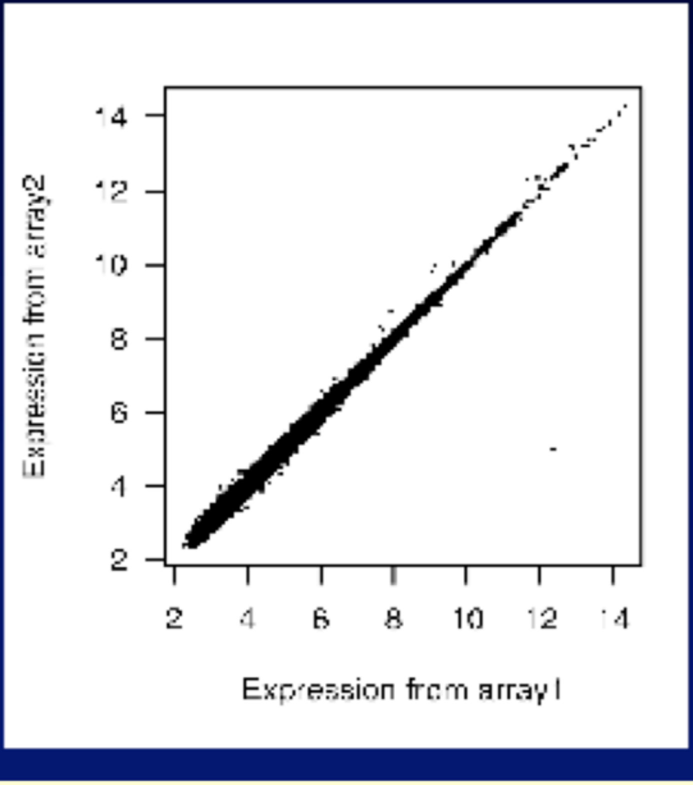
MAS 5.0

- **MAS 5.0** is Affymetrix software for microarray data analysis.
- Ad hoc background procedure used
- For summarization, they use:
  - **Signal = TukeyBiweight{**$\log(PM_j - MM_j^*)$**}**
  - Tukey Biweight: $B(x) = (1 - (x/c)^2)^2$, if $x<c$

    $= 0$ otherwise
- Ad hoc scale normalization used

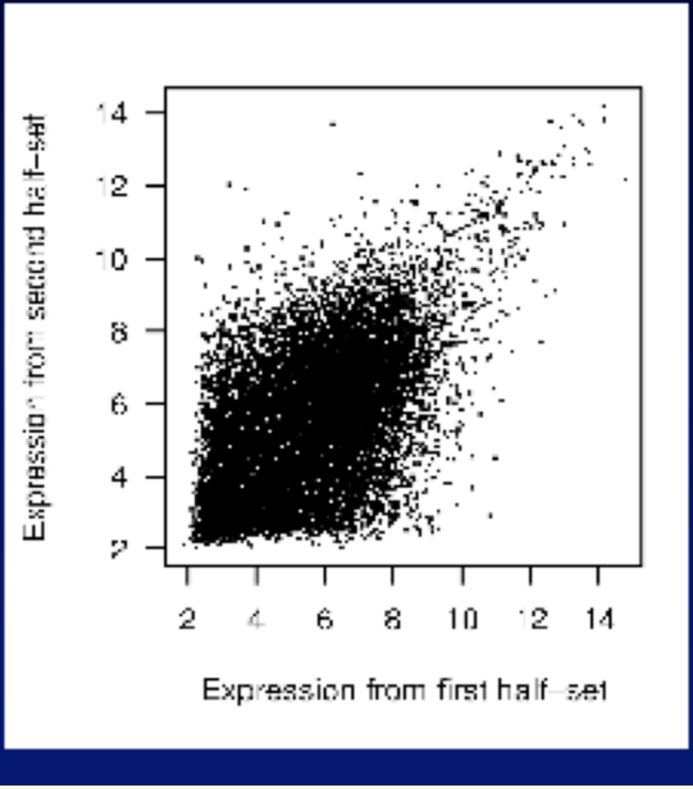
From Talk by Irizzary & PhD thesis by Astrand

2 replicate arrays



Expression from corresponding probes are highly correlated

Correlation is higher than 0.99

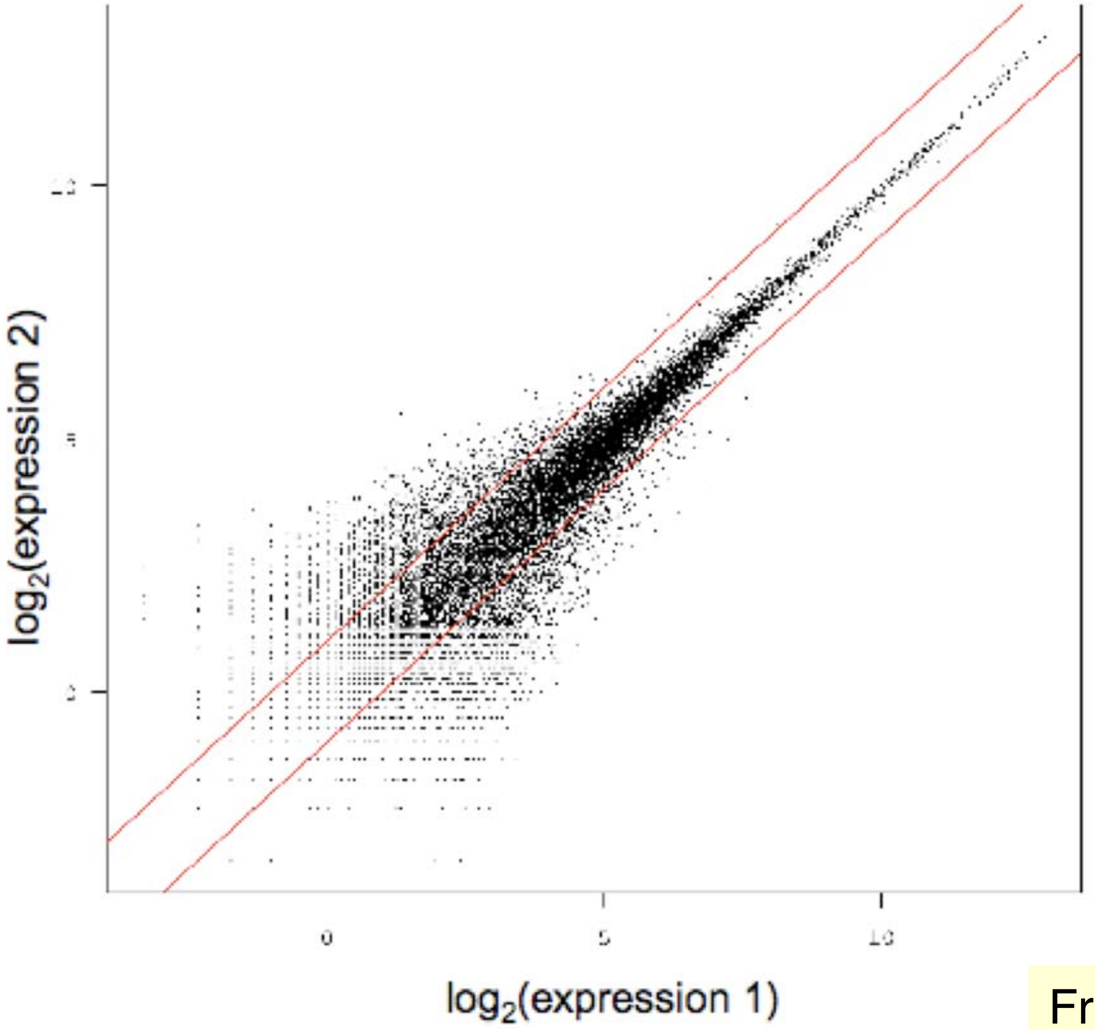


Expression not correlated when probes randomly partitioned

Correlation drops to 0.55

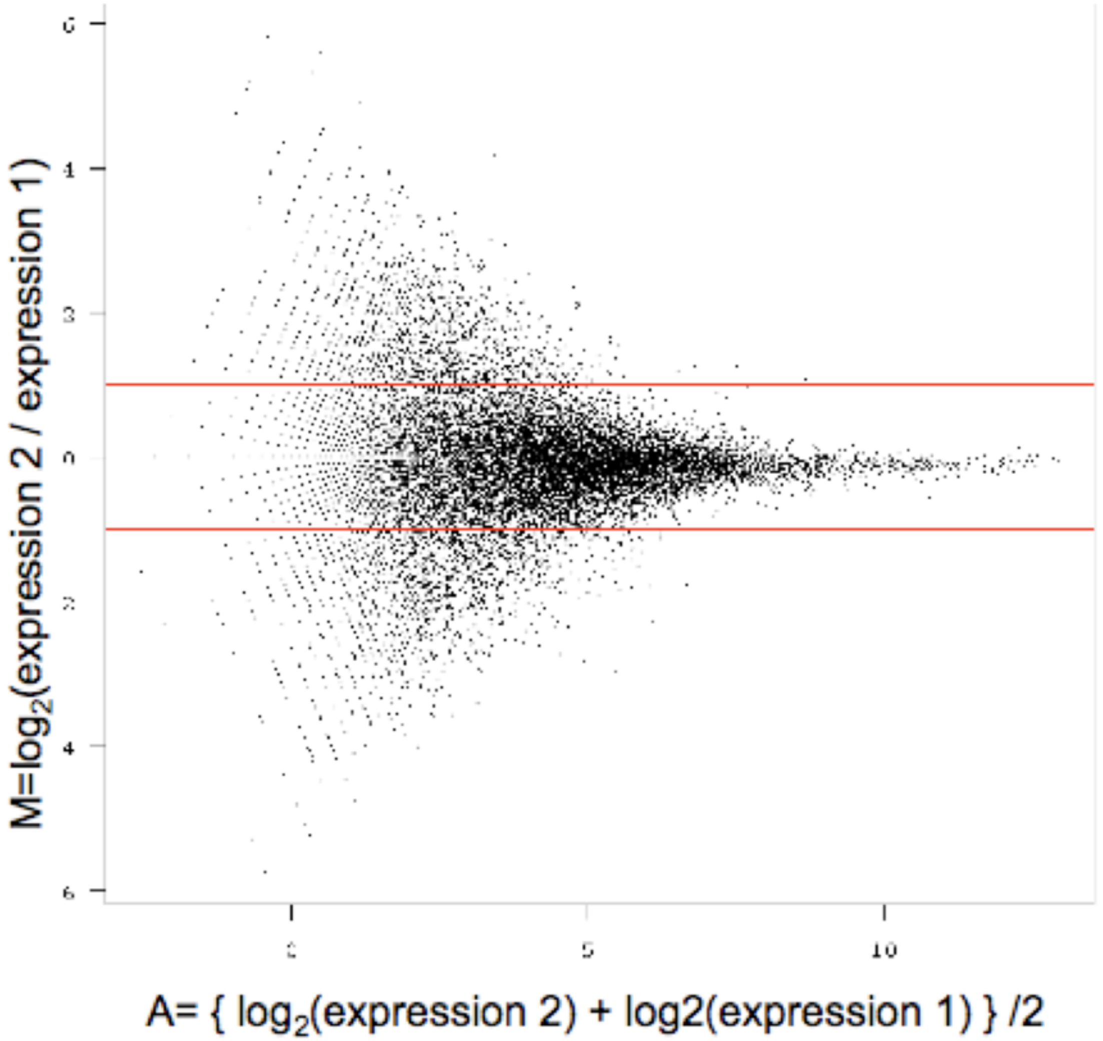
From Talk by Irizzary

We have to deal with **variations**!



From Talk by Irizzary

MvA Plots



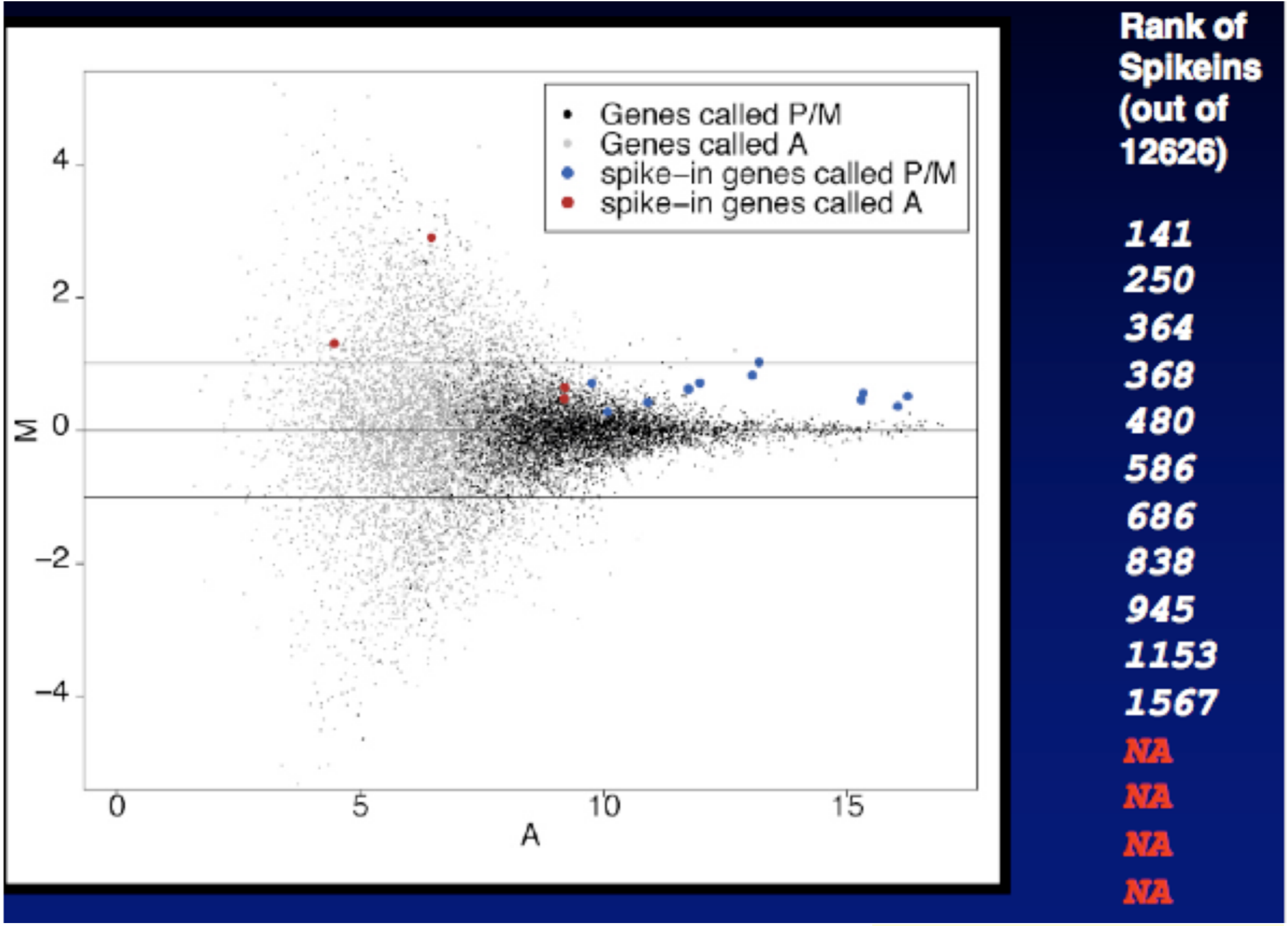
From Talk by Irizzary

Spike-in Experiment

- ❑ Replicate RNA samples were hybridized to various arrays
- ❑ Some probe sets were spiked in at different concentrations across the different arrays
- ❑ Goal was to see if these spiked probe sets “stood out” as differentially expressed

From Talk by Irizzary

Analyzing Spike-in data with MAS 5.0



Rank of Spikeins (out of 12626)

141
250
364
368
480
586
686
838
945
1153
1567
NA
NA
NA
NA

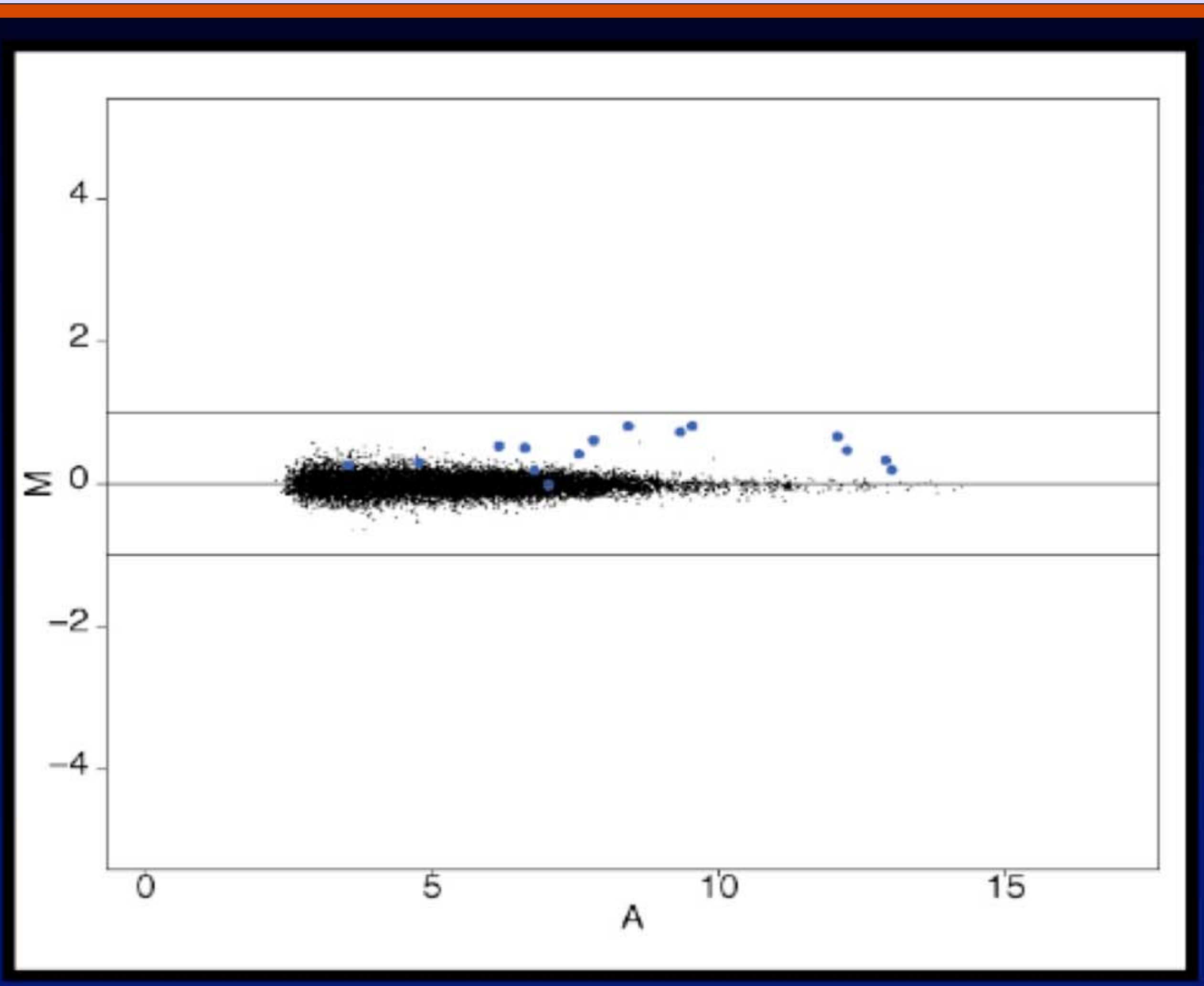
From Talk by Irizzary

Robust Multiarray normalization (RMA)

- Background correction separately for each array
  - Find E{Sig | Sig+Bgd = PM}
  - Bgd is normal and Sig is exponential
- Uses quantile normalization to achieve "identical empirical distributions of intensities" on all arrays
- Summarization: Performed separately for each probe set by fitting probe level additive model
- Uses median polish algorithm to robustly estimate expression on a specific chip
- Also see GCRMA [Wu, Irizzary et al., 2004]

From Talk by Irizzary &
PhD thesis by Astrand

Analyzing Spike-in data with RMA



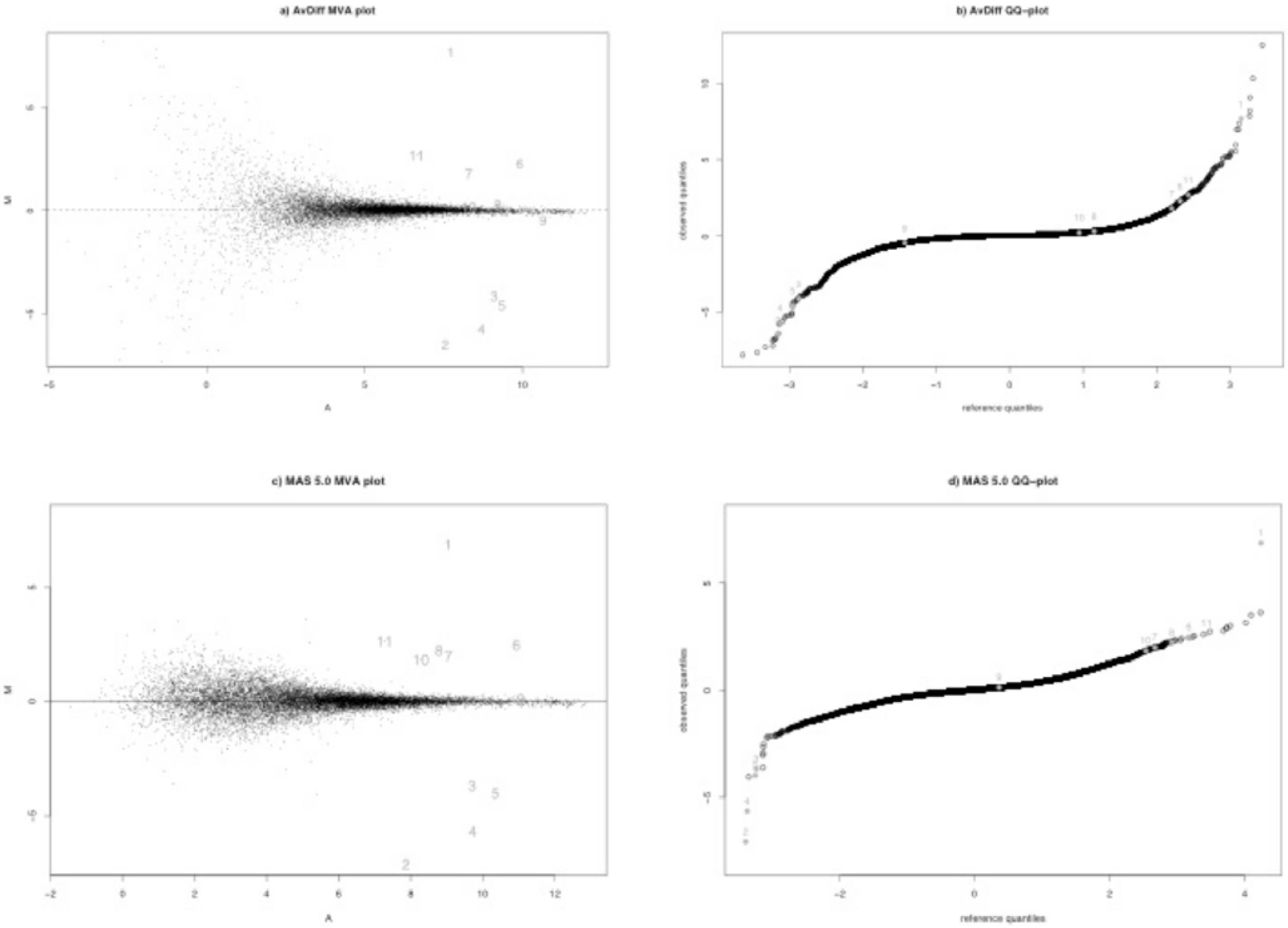
Rank of Spikeins (out of 12626)

1
2
3
4
7
11
15
21
35
122
1182
230
450
1380
11700

Irizarry et al. (2003) *NAR* 31:e15

From Talk by Irizzary

MvA and q-q plots

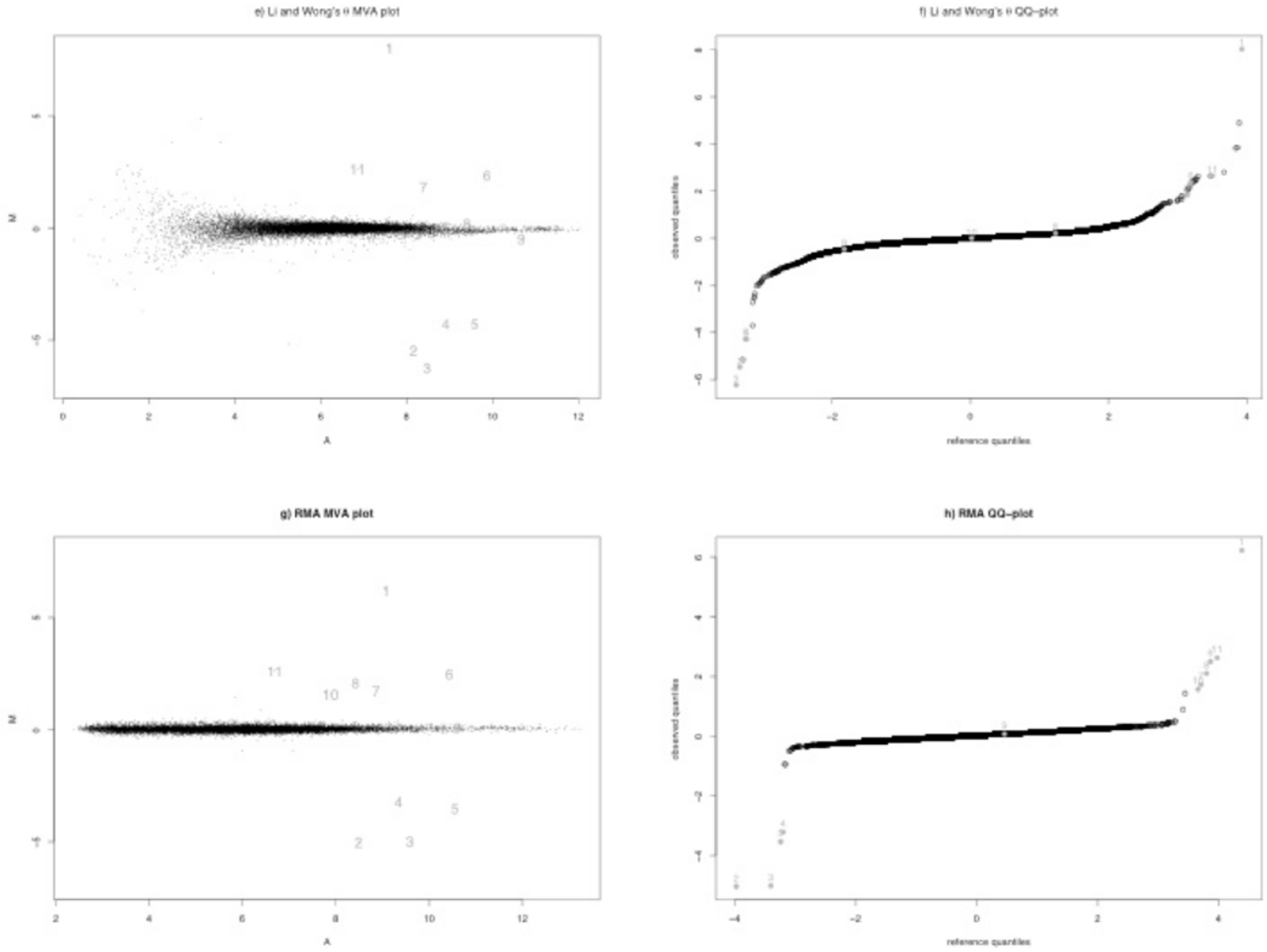


MAS 4.0

MAS 5.0

From Talk by Irizzary

MvA and q-q Plots



MBEI

RMA

From Talk by Irizzary

Before and after quantile normalization

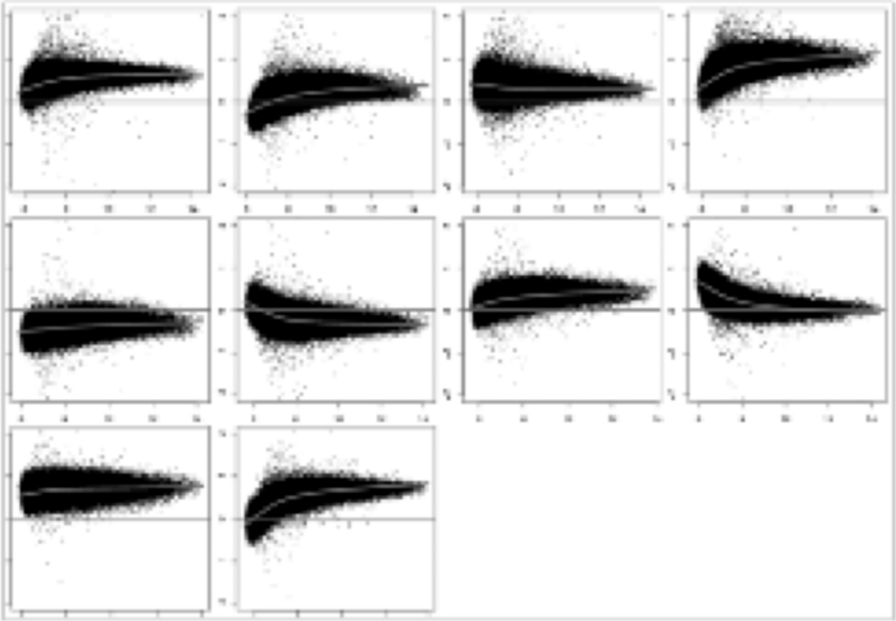


Fig. 2. 10 pairwise *M* versus *A* plots using liver (at concentration 10) dilution series data for unadjusted data.

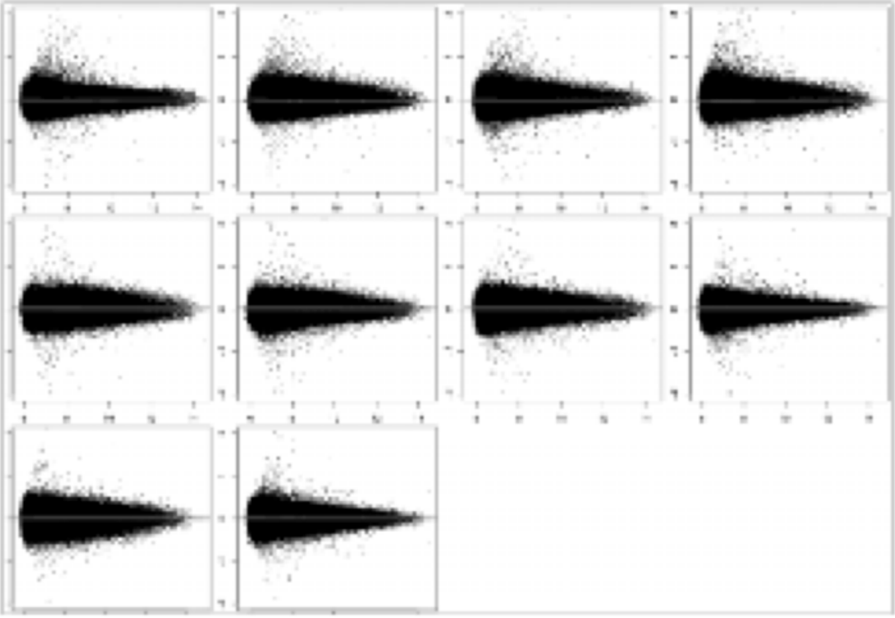


Fig. 3. 10 pairwise *M* versus *A* plots using liver (at concentration 10) dilution series data after quantile normalization.

From Talk by Irizzary

Bioconductor

- **Bioconductor** is an **open source** and open development software project for the analysis of biomedical and genomic data.
- World-wide project started in 2001
- **R** and the **R package system** are used to design and distribute software
- Commercial version of Bioconductor software called **ArrayAnalyzer**

From Talk by Irizzary

R: A Statistical Programming Language

- Try the tutorial at: [http://www.cyclismo.org/tutorial/R/]
- Also at: [http://www.math.ilstu.edu/dhkim/Rstuff/Rtutor.html]