

DNA Microarrays in Medicine

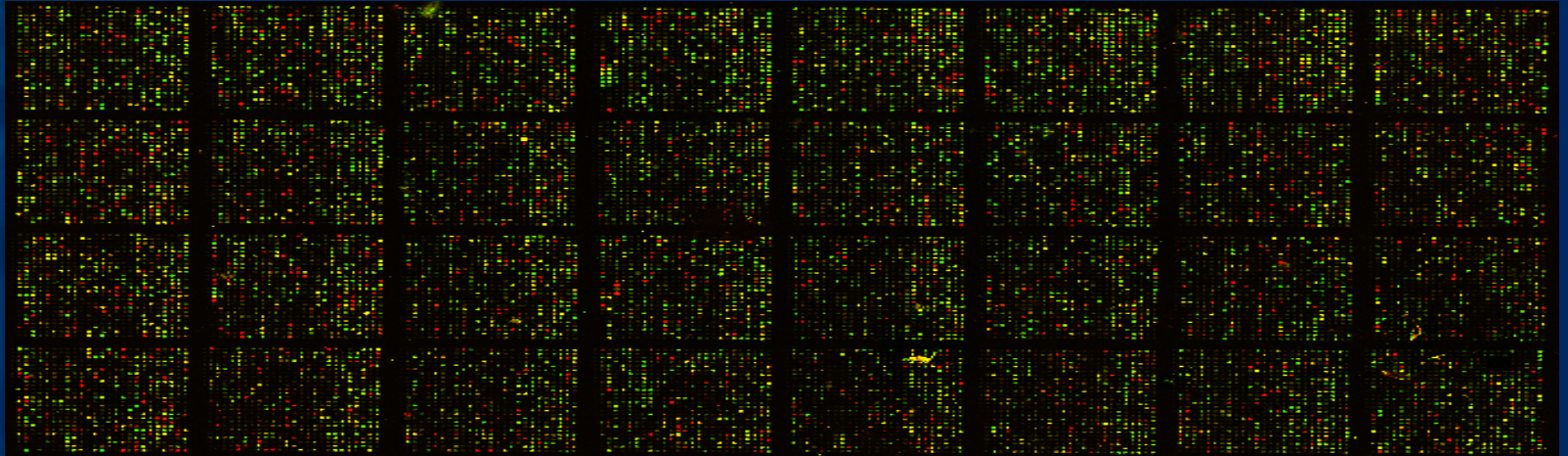
Expressing yourself one pixel at a time

J. Patrick Vandersluis, Ph.D.
HealthRx Corporation
www.HealthRx.com

Patrick@HealthRx.com

Agenda

- The Data of Biological Systems
- Review of Life Sciences Notions
- DNA Microarrays:
- The Bioinformatics Approach



The Data of Biological Systems

- Genomic-centric view of bioinformatics
 - Sequence data are signals
 - Proteins as Expression
 - Metabolic Pathways
- Biomedical signals
 - Signals data are sequences
 - ECG, EEG, EMG
- Both gene expression data and signals have significant temporal thread
- In combination, huge discoveries are waiting to be made

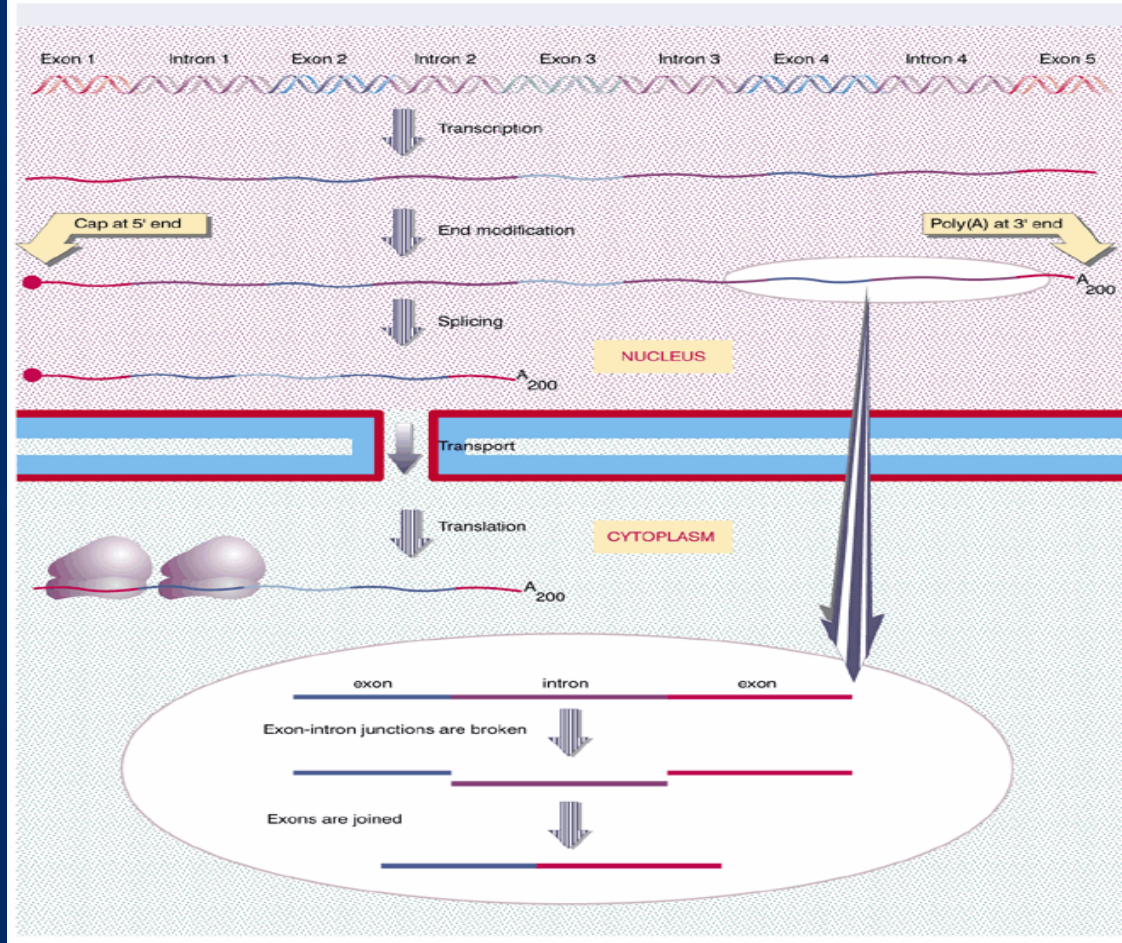
Review of Life Sciences Notions

- Sequences don't tell the whole story
- In same genome, expression is cell-type specific, tissue specific, and environment specific
- Disease, stress, metabolism cause variability of gene expression
- There is some “distance” between sequence and protein
- Just how do we get from genes to function?
- DNA -> mRNA -> Protein -> Function

RNA Processing

- Occurs In nucleus
- Pre-mRNA → mRNA
- Introns removed
- Mature mRNA transported through nuclear membrane
- Translated in cytoplasm
- Why look at mRNA instead of protein?

Figure 22.2 Overview: RNA is modified in the nucleus by additions to the 5' and 3' ends and by splicing to remove the introns. The splicing event requires breakage of the exon-intron junctions and joining of the ends of the exons; the expanded illustration shows the principle schematically, but not the actual order of events. Mature mRNA is transported through nuclear pores to the cytoplasm, where it is translated.



Protein, mRNA Profiles Differ

- Temporal differences between gene expression and protein accumulation
- Differential stability of mRNA and protein
 - Difference between rate of synthesis and amount of product
- Spatial differences, compartments, transport of mRNA and protein
- Differential processing of mRNA yielding various protein products
- Protein post-translational modifications

Protein Post-Translational Modifications: Examples

- Phosphorylation
- Glycosylation
- Proteolytic cleavage
- Disulfide bond formation
- Methylation
- Prenylation
- Adenylation
- Association with cofactors
- Etc.



One Gene - Multiple Proteins

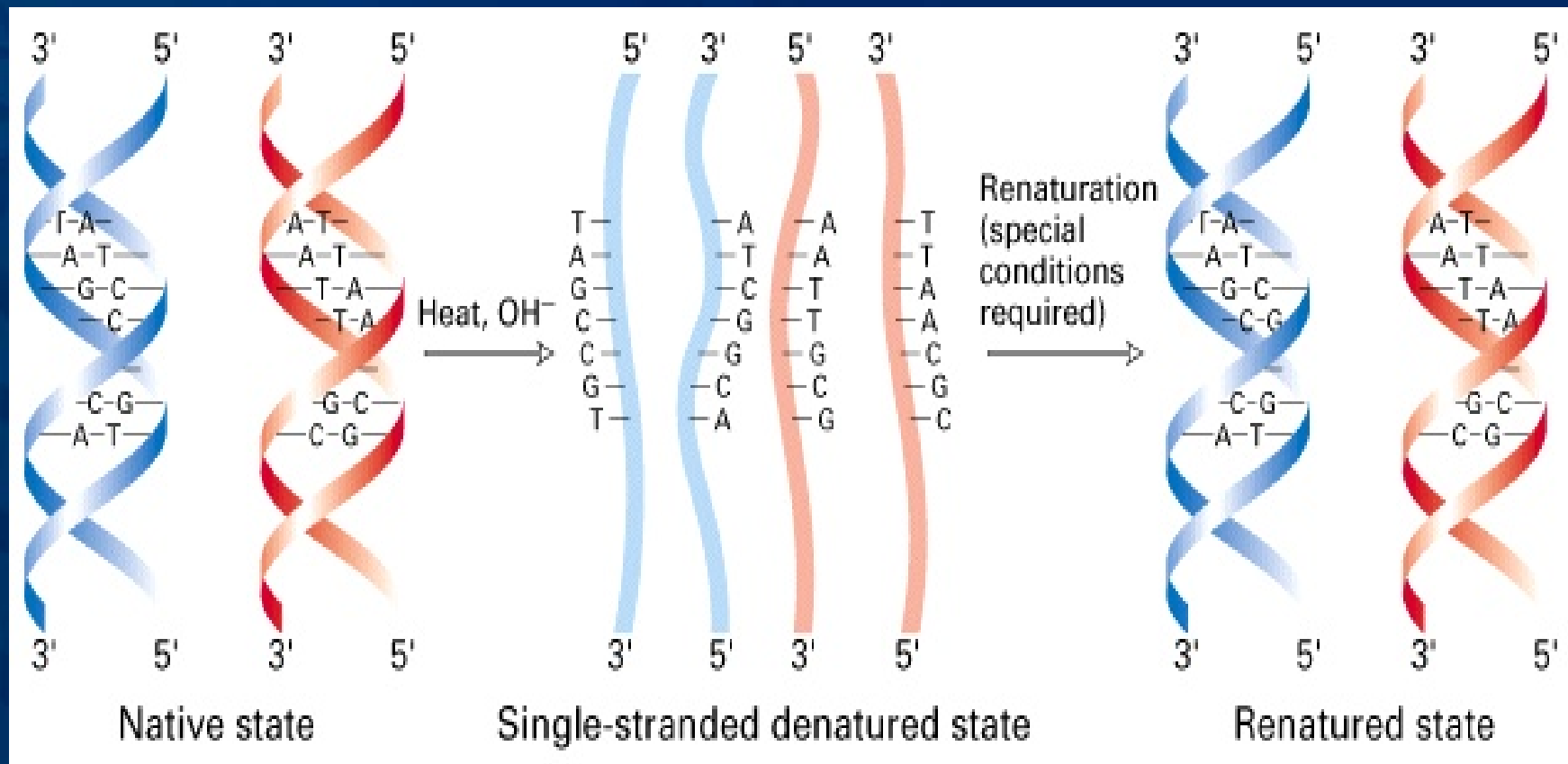
Insulin as Example

- Translated as preproinsulin, followed by membrane transport and processing.
- Cleavage yielding proinsulin and peptide C, then cleavage of proinsulin into 2 peptides.
- Ultimately three peptides: two (A, B) combine via disulfide bonds to form insulin.
- The third peptide C has unique biological activity of its own: cardiovascular effects.
- Summary: 1 gene, 3 polypeptides, 2 proteins, with 2 functions

What is the Problem to be Solved?

- Need strategy for understanding why certain proteins are produced under specific conditions (function)
- Understanding that mRNA is a measure of gene activity, need effective technique for measuring it through time
- Need tool to compare entire experimental transcriptome with a reference genome through time and varying environmental conditions to give greater understanding of complex gene interactions

Reminder - Nucleic Acid Hybridization is Specific



DNA Microarrays

- DNA Chips
- Massively parallel measurements
- Allow simultaneous measurement of the level of transcription for every gene over time
- Provide means for collecting data about differential gene expression over time in changing environment
- Robotic manufacture, standardized
- Automated readers, computer interpreted

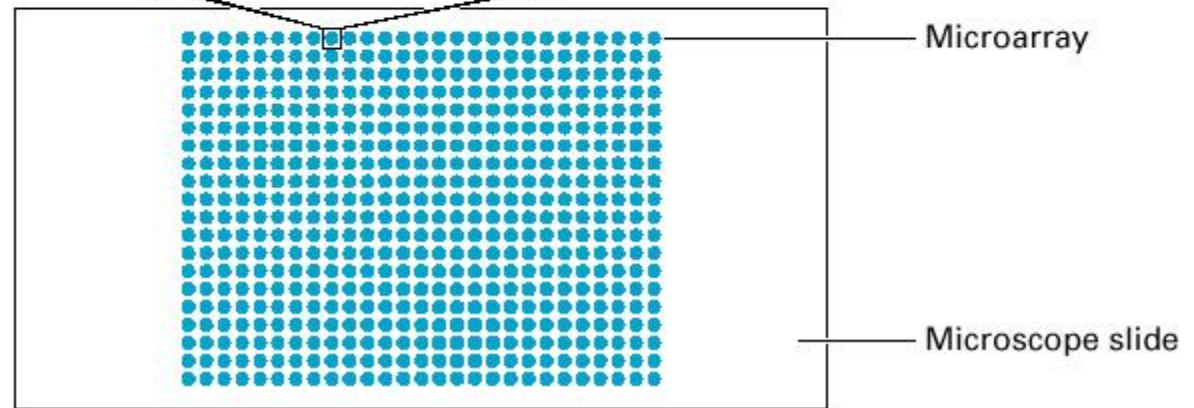
Making DNA Microarrays

- Known genes are amplified using PCR
- PCR products are verified, purified, and spotted onto ordinary glass slide
- Spot is about the size of a printed period
- Typical size is 6,200 spots, made robotically at about 12 slides per hour
- Once spotted, DNA is denatured, linked to slide with covalent bonds
- Stored at room temperature, very stable

GeneChip® in Schematic View

Sequence of one gene

```
TCCTTTCCGG AACGGTTGGC GTCTGCGCAC GCGGGTGTGG GGCATGACAT
GCCGCCCCAG GAACAACCCC GACACGGCTT TAAGCCTCTC AAATCGCTGT
AGACATCATC TTTACGTGCT TGCCACCATT TGCCACCATT AGGGCTGTTC
CCGCGACGAC TCGCCATTCA ACCTCAGTCC TTCGGGTTGA GCGAGTGGGT
CGCGCGCAAG GTGCGAATGG GTCGCGCGCA AAGTGTTGCG CTGGCTGTAT
TATATGCTGC CTATAGCGAG ACTAACGACC CACACTTTCA CACAAGGATT
TCCCGCTAAT GGGTACCTCG CGTCAGGACC TTGACGCAAG CGCGCCTTCG
GTTGGCCCA AGCTTGCTAG GACTACTTAT CTTGAGCTCA TTTAACATCC
CGGCGCCTCT CCGGGAGCGG TCGTCGCGAA GAAGTCAAAC CCGGAACGGC
GTTGACAAAG CGTGGAGACA TCGATACCTC TGTGTCAGCG GCCACAAATC
```



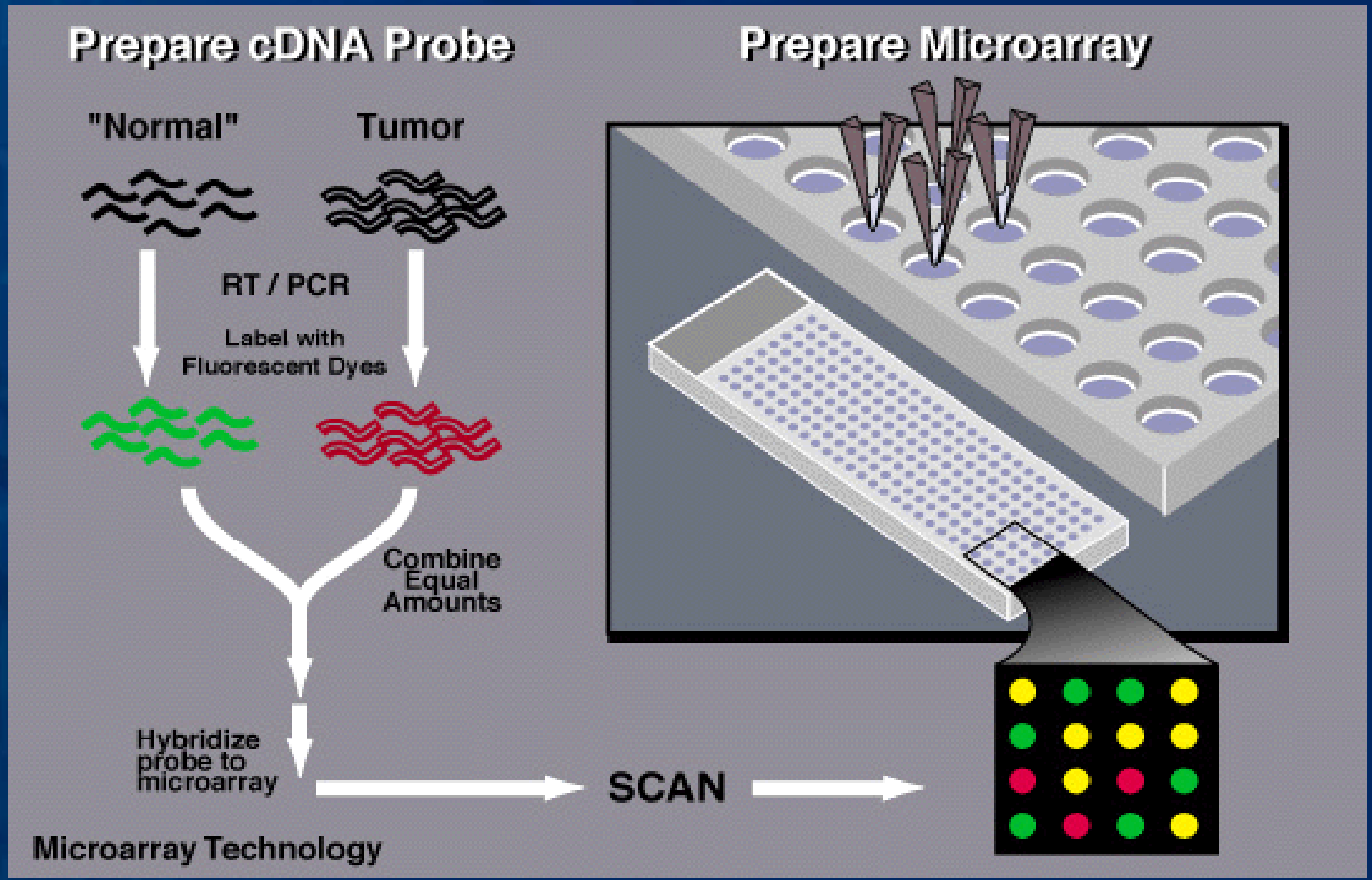
GeneChip® Probe Array



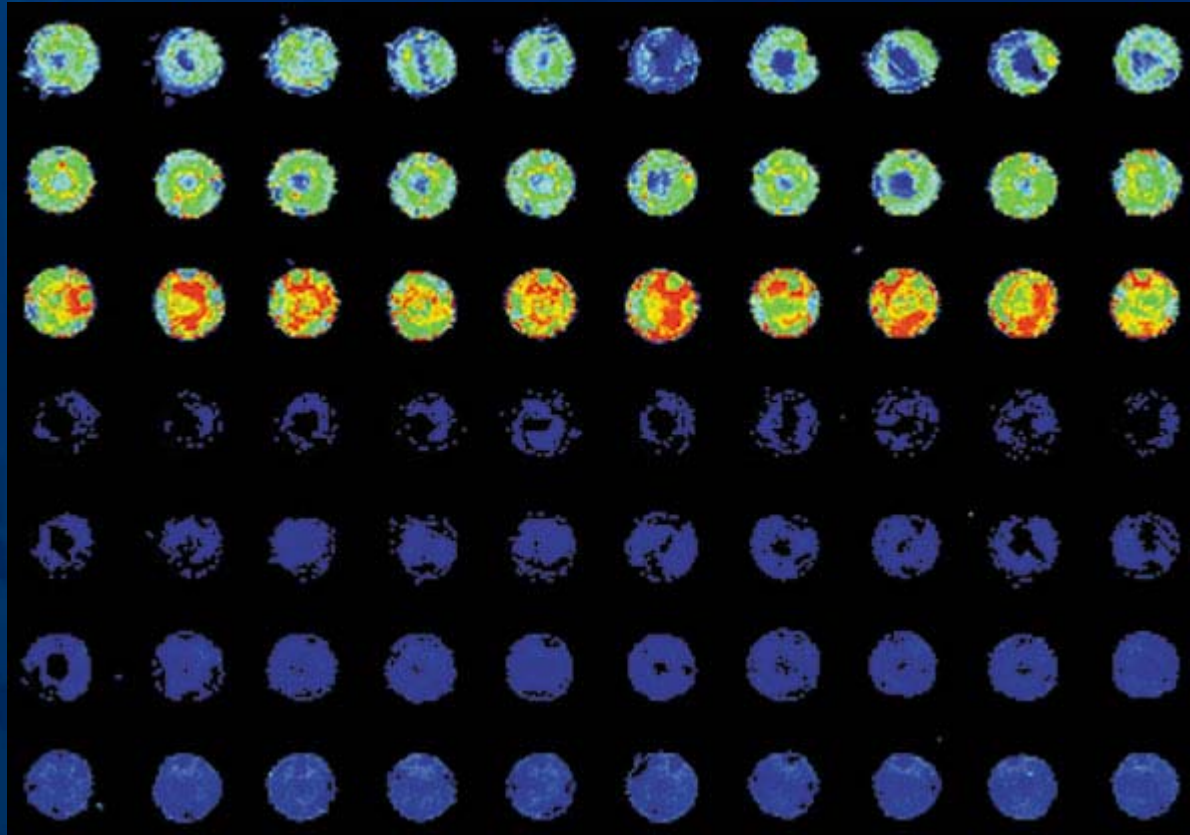
DNA Microarray Processing

- mRNAs harvested from two populations of cells
- For each population, mRNA copied to cDNA with reverse transcriptase, using nucleotides labeled with fluorescent dye
- Resulting cDNA mixed in equal amounts, microarray flooded and incubated overnight and allowed to hybridize
- Microarray washed, scanned with laser, and images captured
- Green and red scans are done separately, merged computationally. Yellow represents exactly 1:1 ratio of control/reference and experimental – an unusual situation
- Image processed to record color and intensity of spots

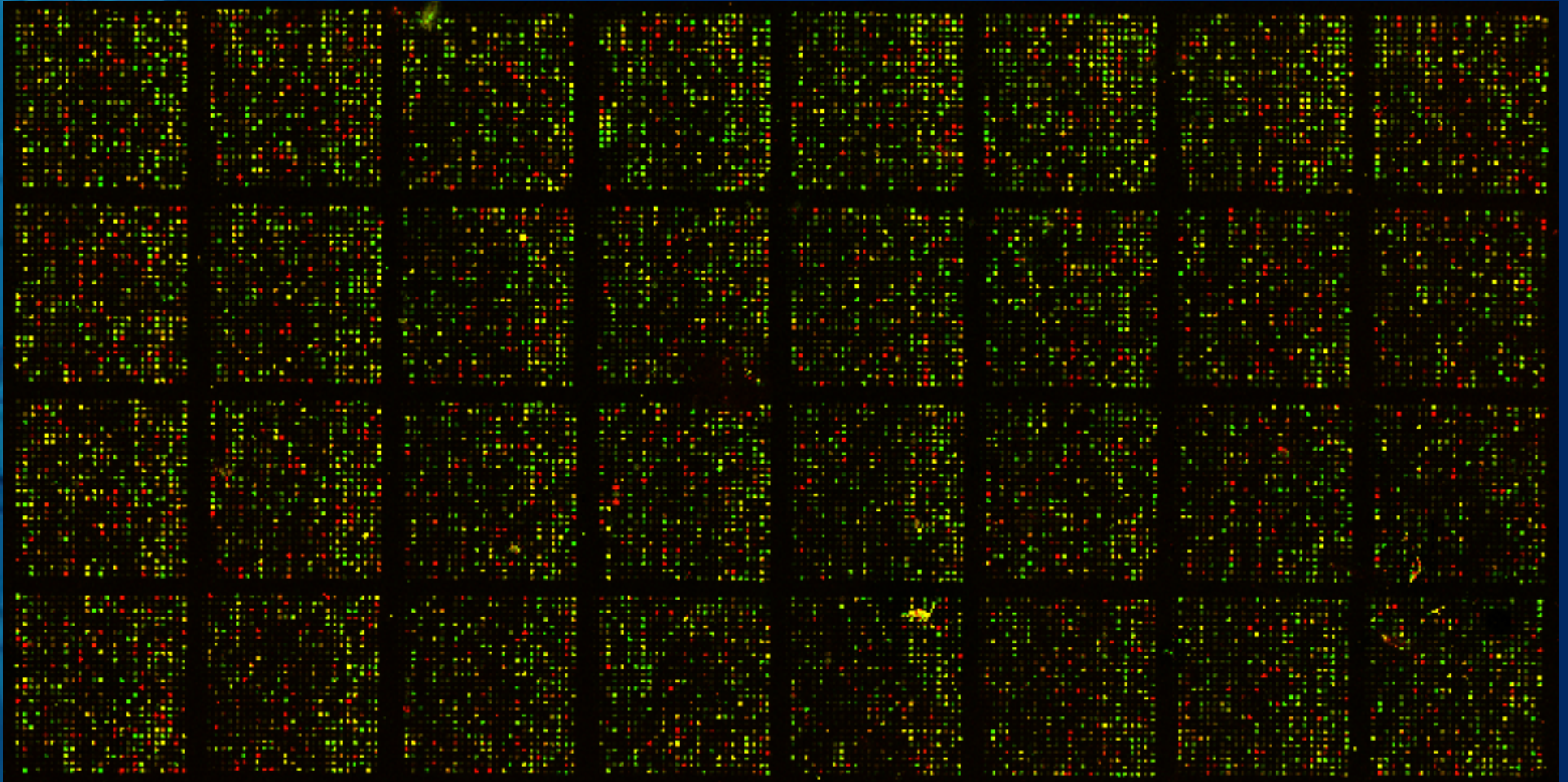
Labeled Cell cDNA Binds to EST or Gene DNA on Microarray



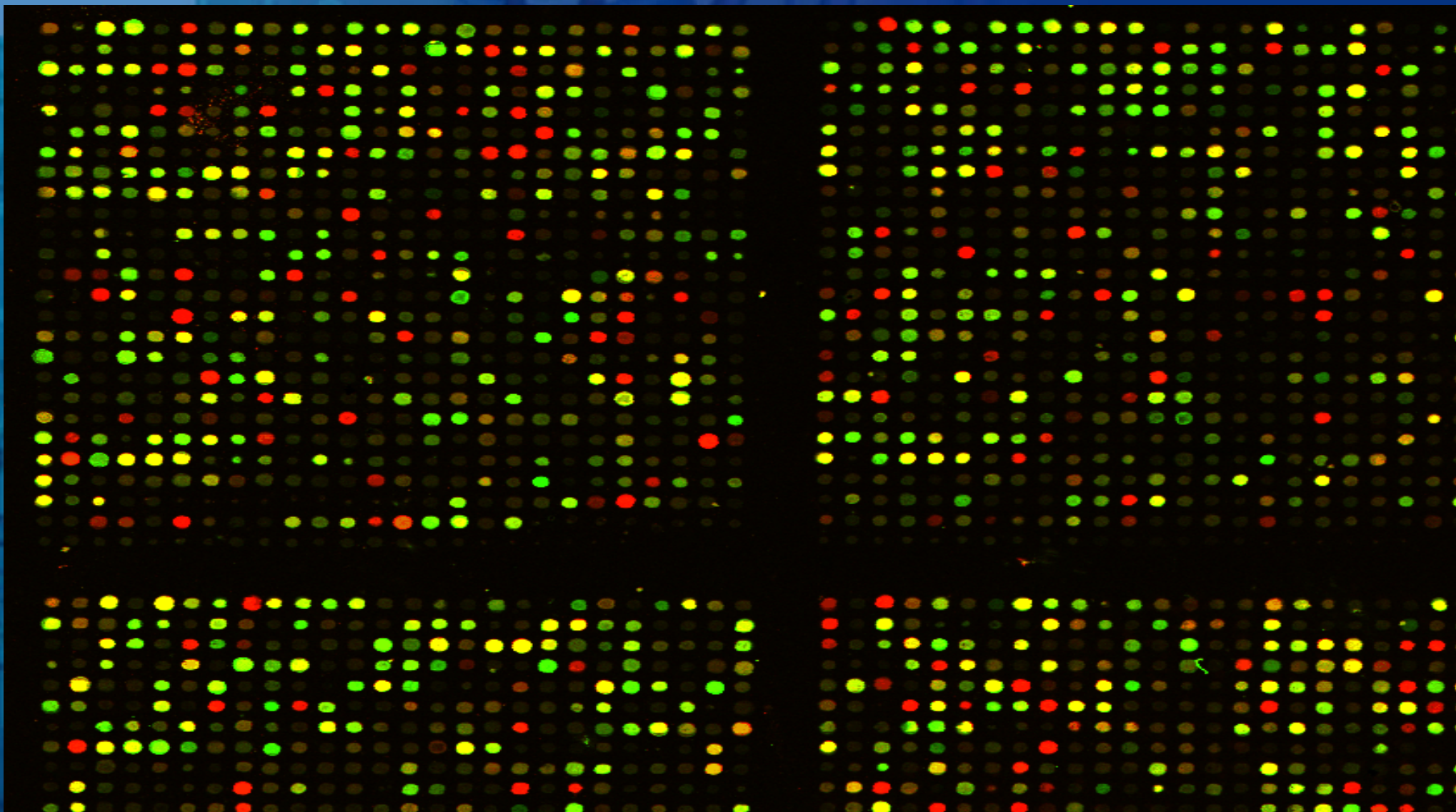
Sample of Affymetrix Arrayer Output



“All” 19,353 Genes of *C. elegans*

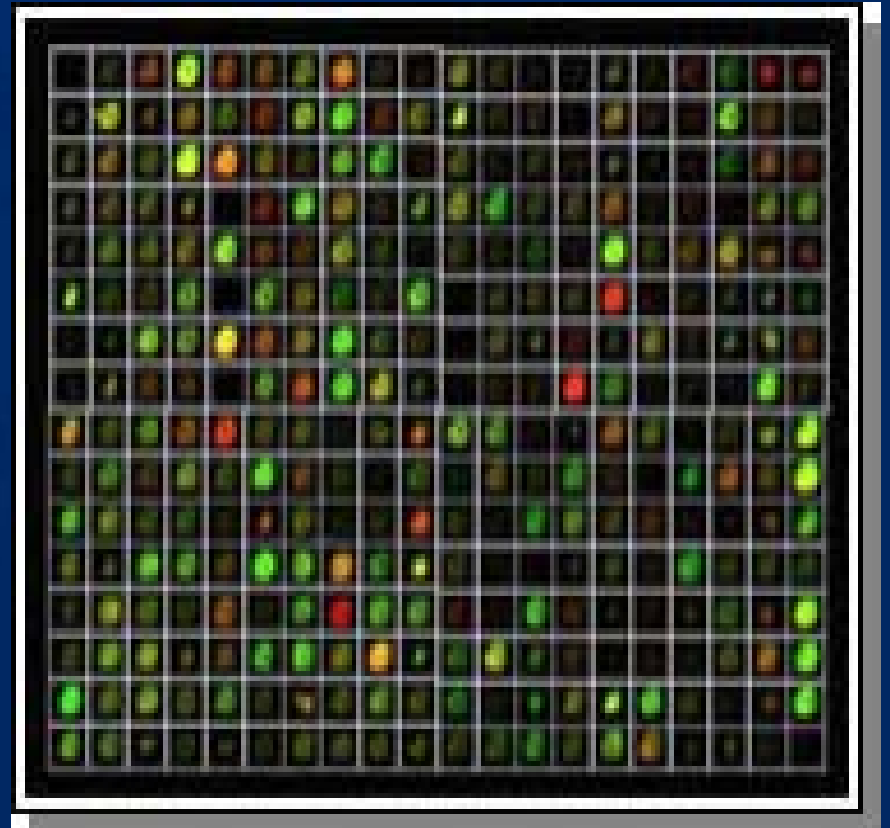


Zoomed View



Fluorescent Microarray Data

- Microarray raw data: Color and intensity correlate with expression level
- Each spot represents hybridization of probe to a different defined DNA: EST or cDNA representing a gene
- Absolute intensity has no meaning, only relative intensity



cDNA Arrays

- In a population of cDNAs, each cDNA sequence and number of copies tell us which gene was expressed by how much, but exon regions only
- Compare with genomic (reference) DNA to find introns, promoters, intergenic DNA
- cDNAs represent the entire set of genes expressed in a particular cell or tissue type at a particular time
 - Compare normal vs. disease, tissue types, metabolic stress, differing environment
- Progress from structural genomics to functional genomics

Now What?

- We now know how to collect plenty of expression data
- We can create a a large database with quantitative information
- How do we go from data to understanding?

The Bioinformatics Approach

- Deals with biological data as collections, in very large numbers
- Typically sequence-oriented data
- Algorithms seek to parallelize
- Computationally intensive
- Born of readily-available massive computing resources
- Data leads to hypothesis rather than reverse

Summary

- Understanding biological systems requires more than defining sequence data
- From sequence to function is a long jump
- Microarray data are fodder for tools that give insight into expression under varying conditions
- Combining microarray data with other clinical data along common temporal thread will provide insight into both normal and pathophysiology
- With your help, great discoveries will be made in this space!