BIOC4004 - Industrial Biochemistry
Lecture 16 - Wed Mar 03, 04

Topics for the Day:

- Annotation (end)
- Functional genomics
- DNA microarrays (intro)

Public Human Genome project:

• Used Ensemble and Genie programs

• Merged that data with known sequences from various Dbases

Ensemble

• uses prediction of Genscan (an HMM program)

• checks these predictions against ESTs, mRNAs and protein motifs in known databases

• Ensemble predicted 35,500 genes

Genie

• tries to match 5' end ESTs with 3' end ESTs to make full-length predictions

In The End:

Came up with total of 31,778 predicted proteins

• 14,882 from known genes

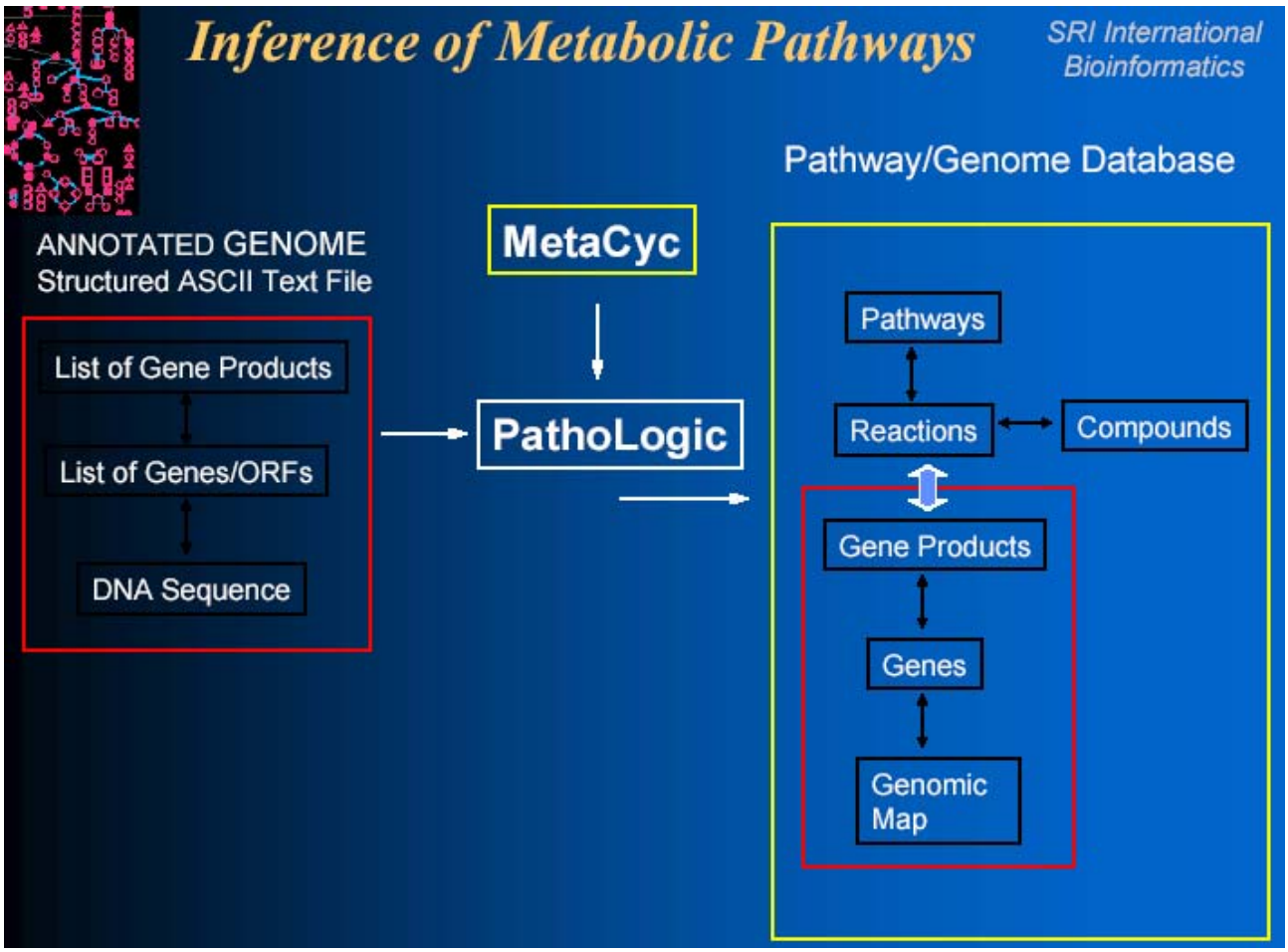
• 12,839 from Ensemble

• 4,057 from Ensemble-Genie

• Avoid "**annotation drift**" from bad GenBank data

Annotation Stage 3. Pathways and Networks

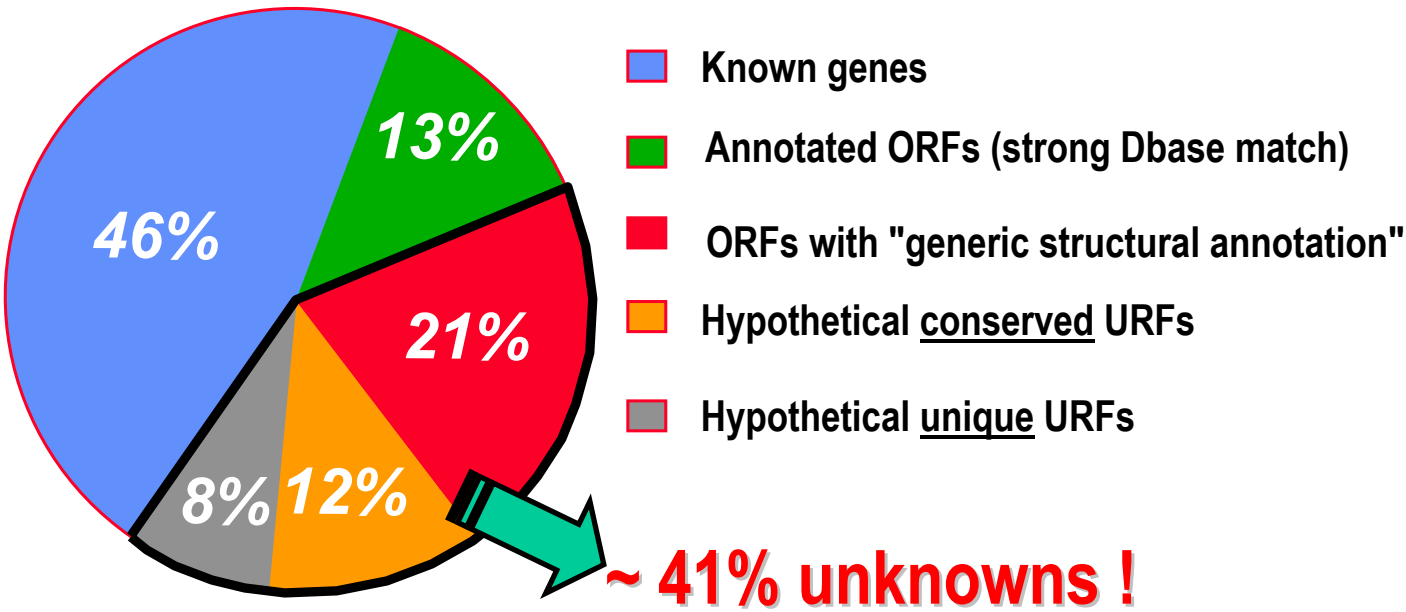
http://MetaCyc.org



- MetaCyc is a comprehensive metabolic pathway DB
- Literature based
- Detailed information on each pathway
- Goal is to contain an example of every different metabolic pathway
- Freely available

The “end-product” of sequencing…

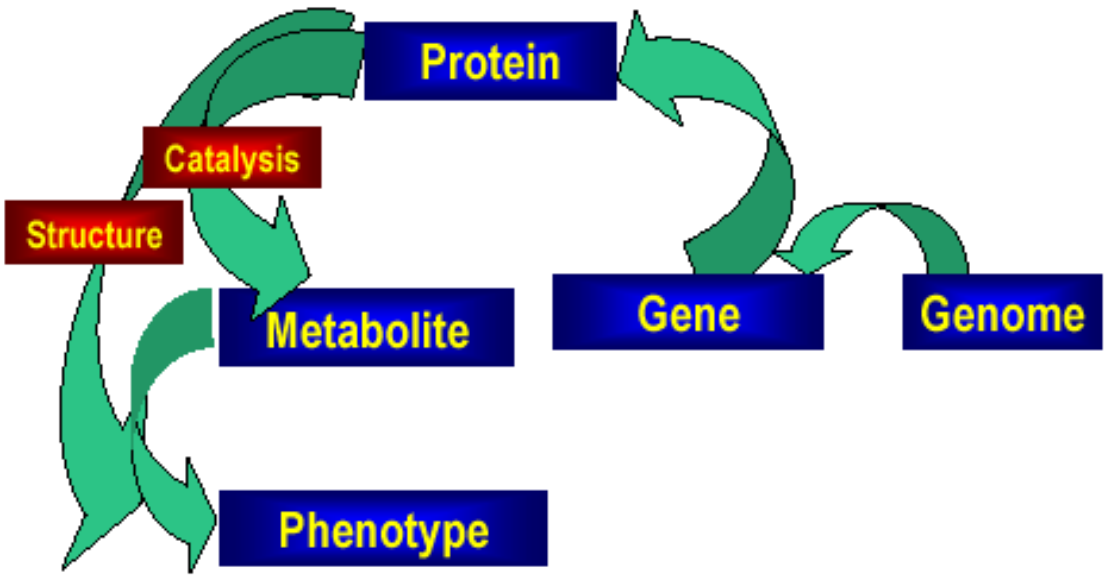
- A fully annotated genome:
  - every predicted gene in the genome
  - a known (or putative) function for as many genes as possible
  - a reconstruction of as many metabolic or regulatory pathways as possible
  - a whole lot of unknowns



- from 40 to 50% of all of the ORFs in virtually every newly sequenced genome are total unknowns (conserved or not)
- lots of “species-specific” stuff !!!

Understanding the biology of an organism

- Phenotype is essentially defined by:
  - coding potential, the genome = DNA
  - regulation of coding potential, the transcriptome = RNA
  - transforming coding potential into effectors:
    - in response to “cell programme”
    - in response to stimuli or environment
      - the proteome = proteins
      - the metabolome = small molecules and/or metabolites

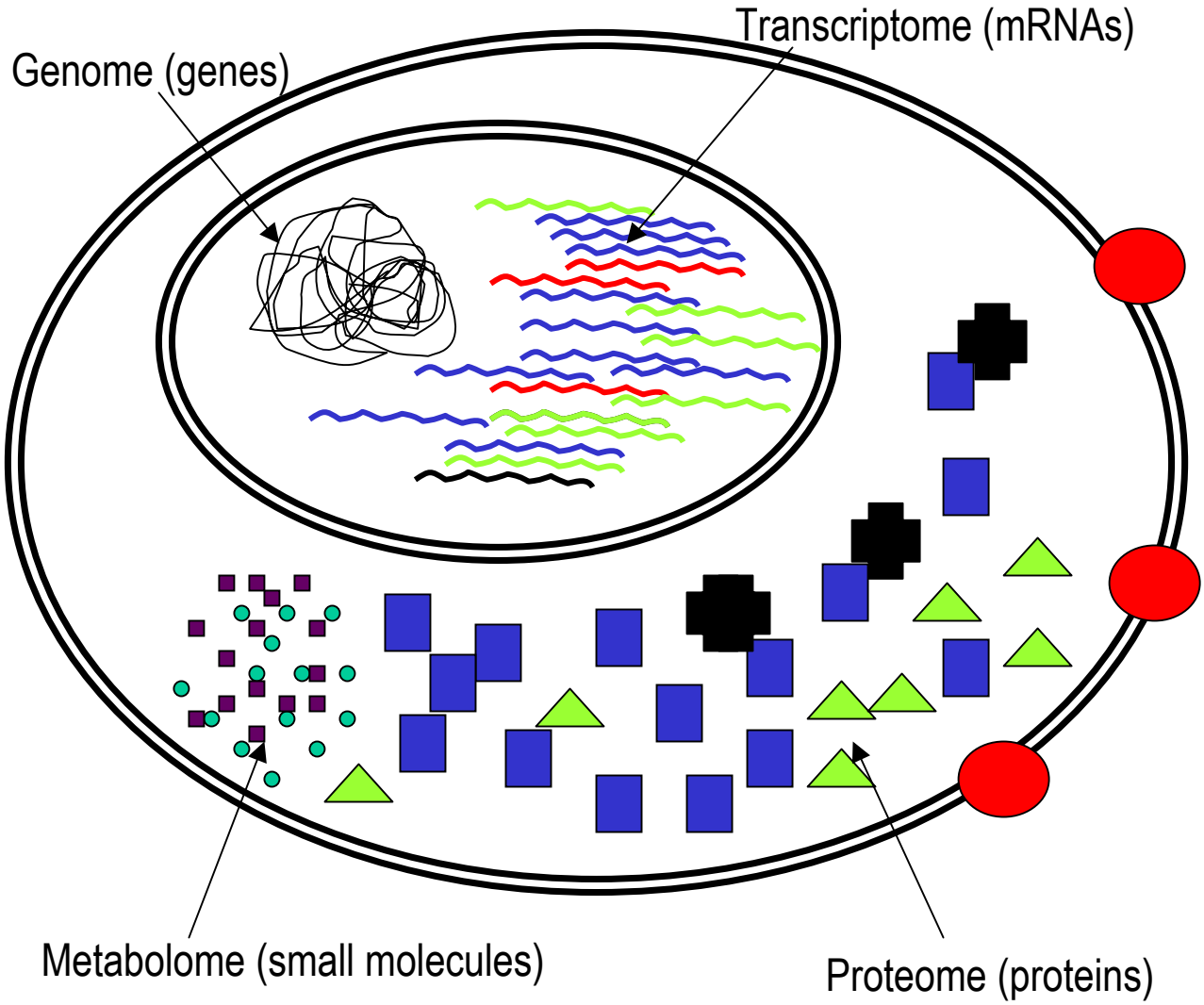


Phenotype dictated by . . .

- **potential contained within the genome**
- **gene expression → protein**
- **protein activity → metabolites**
  - **FEEDBACK LOOP → LIFE**

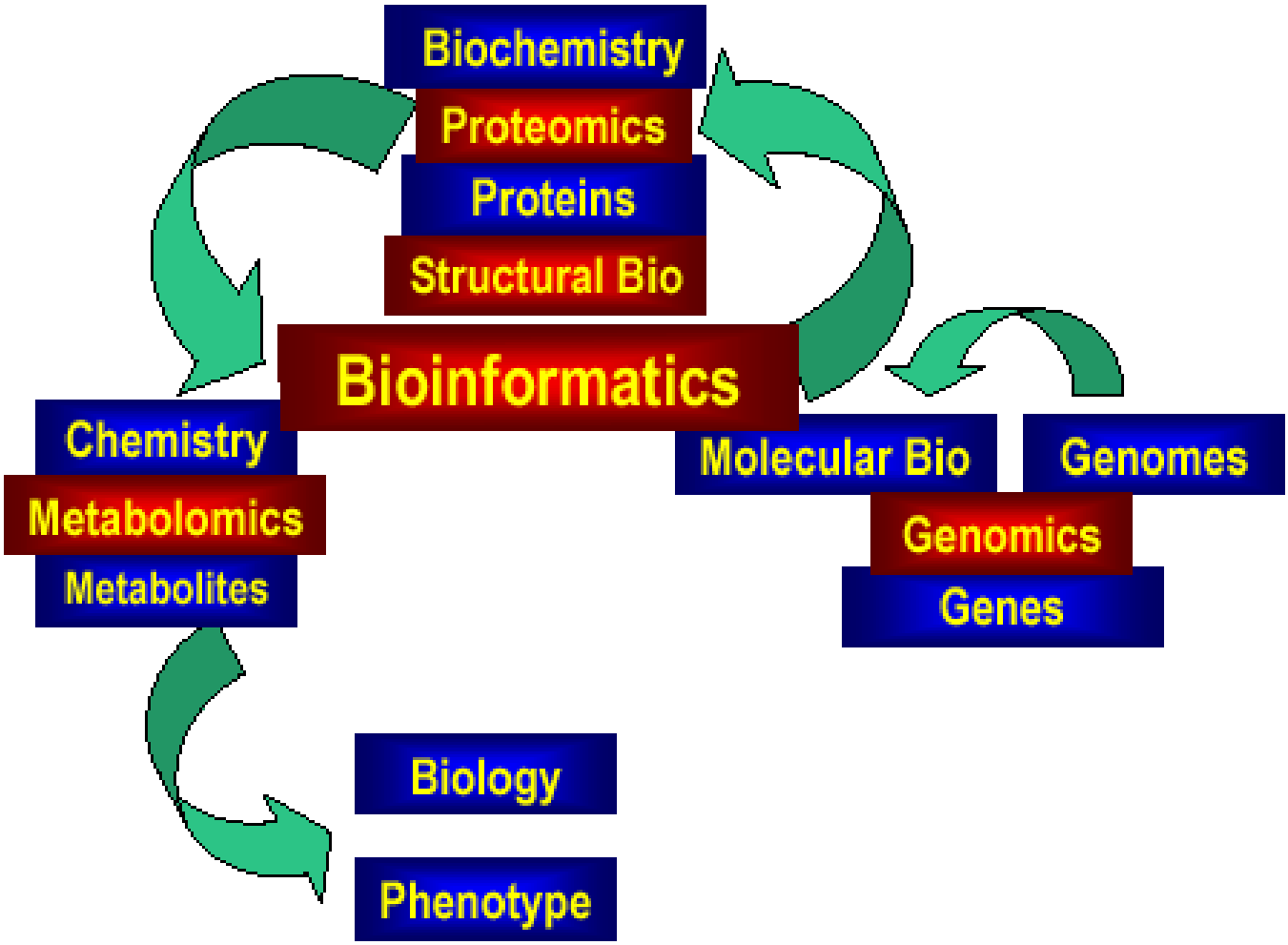
10

What's going on in a (eukaryotic) Cell ?



- A cell is just a bag full of the by-products of "life processes"
- "Life processes" are dynamic; by-products are in state of flux
- Measure them to get (in)direct view of cell processes
- With the advent of genomics, we can tackle this at the “whole-genome level”

Functional Genomics



- What's going on ?
- When is it happening ?
- How much of it is happening ?
- How much of this activity is: cell-type, stimulus, temporally specific ?
- How are these processes regulated ?
- What happens when these processes go awry ?

Functional Genomics

"Biological Relevance"

DNA < RNA < Proteins Metabolites

"Accessibility"

DNA > RNA > proteins > metabolites

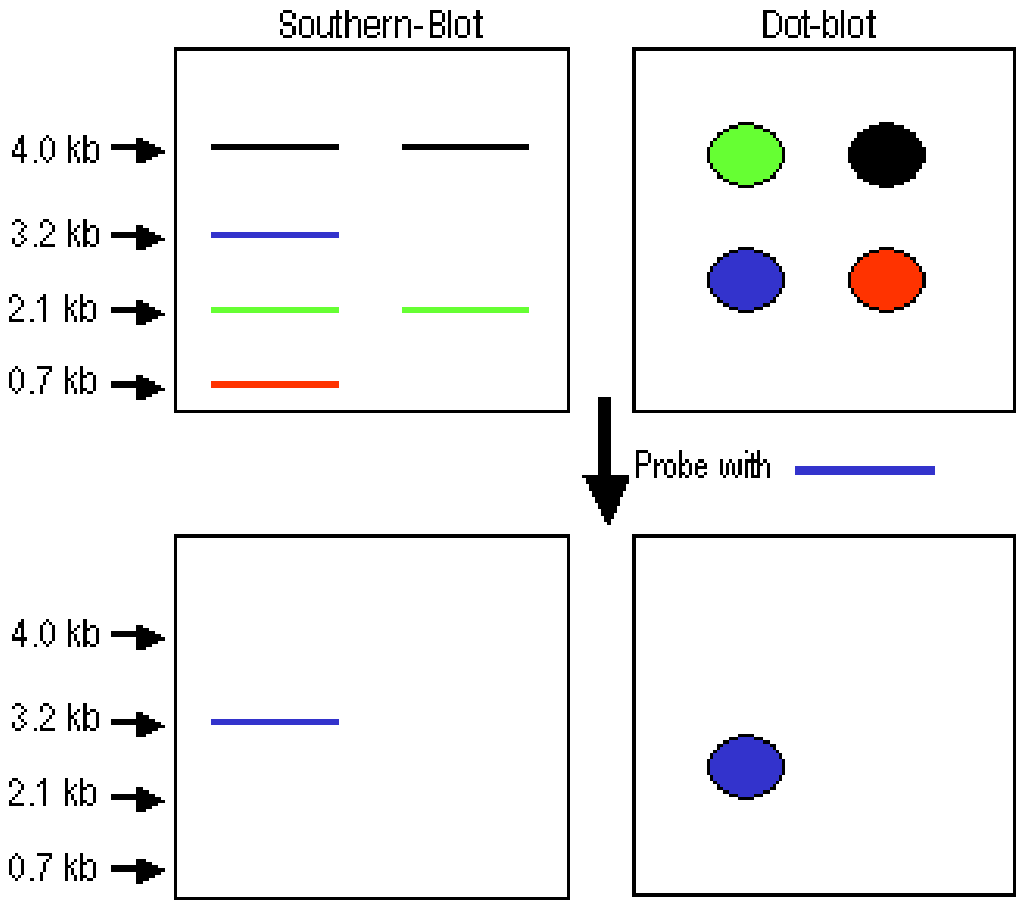
- "biological information" flows down from the genome, to the transcriptome, to the proteome, to the metabolome…
- The ability to measure each is different, and answers slightly different questions
    - every approach has its caveats
    - use every approach, then try to integrate it all later

- having the full-genome sequence of an organism is the first step towards understanding the biology of that organism in a "holistic" level
    - at least the unknowns are all accounted for !
- the aim of functional genomics:
    - "to gain experimental evidence on the activity and interaction between genes using "global methods" (ie. at whole genome level)"
    - to infer the function of genes and to infer how everything comes together (regulatory networks, metabolic pathways)

Functional Genomics at the Nucleic Acid Level: DNA MICROARRAYS

"Dot-Blots" vs. Southern Blots

- Dot-blots are made by spotting various DNA samples on a membrane
- A dot blot can be probed in the same manner as a Southern or Northern blot



- Disadvantages:
  - a dot-blot can only tell you if there is something in your probe that hybridizes to the blot and how much hybridizes. ie. there is no information on fragment size
  - lack of information on fragment size means less (useful) information

  eg. On a Southern: physical size of band that hybridized with probe

  - if spotted sample "contaminated", can't tell if contaminant hybridized to probe

What is a DNA microarray? (part I)

- Nucleic acid samples deposited on a solid matrix, gridded at high density
- A miniaturized DNA dot-blot:
  - **Matrix: derivatized glass, nylon membrane, silicon wafer...**
  - **Nucleic Acid: oligonucleotide, PCR amplicon, plasmid DNA...**
  - **Feature Density:**
    - **spot diameter: 100 - 150 μm**
    - **spot to spot distance: 150 - 200 μm**
    - **~ 40 K spots on a microarray the size of a microscope slide**

- What we can do with microarrays?
  - **A hybridization substrate for assaying "identity" and "quantity" of an nucleic acid analyte**

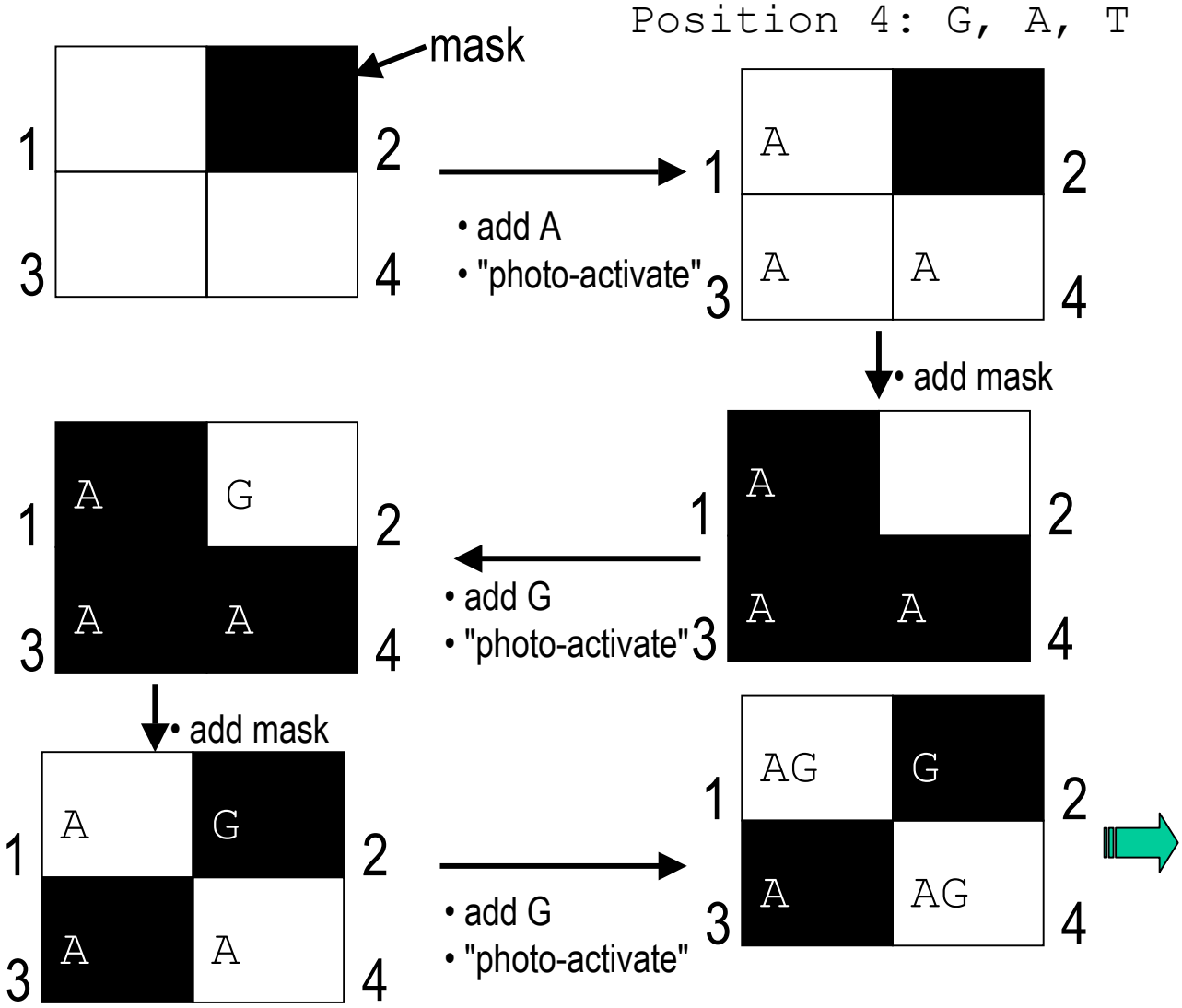
    - **Transcript Profiling (lots of Northern blots)**
      - **how do two samples differ with respect to gene expression**
        - » **how do cancer cells differ from normal cells wrt expression**
    - **Comparative Genomics (lots of Southern blots)**
      - **how do two samples differ with respect to gene content or composition**
        - » **how do cancer cells differ from normal cells wrt gene content**

Affymetrix Genechips: in situ oligonucleotide synthesis

- Photolithography process: uses masks, photo-activated nucleotides, light
- Oligos are "grown" one base at a time

```
Probe 1: AGCTG
Probe 2: GAAAA
Probe 3: ATCAG
Probe 4: AGGCT
```

```
Position 1: A, G
Position 2: G, A, T
Position 3: C, A, G
Position 4: T, A, C
Position 4: G, A, T
```

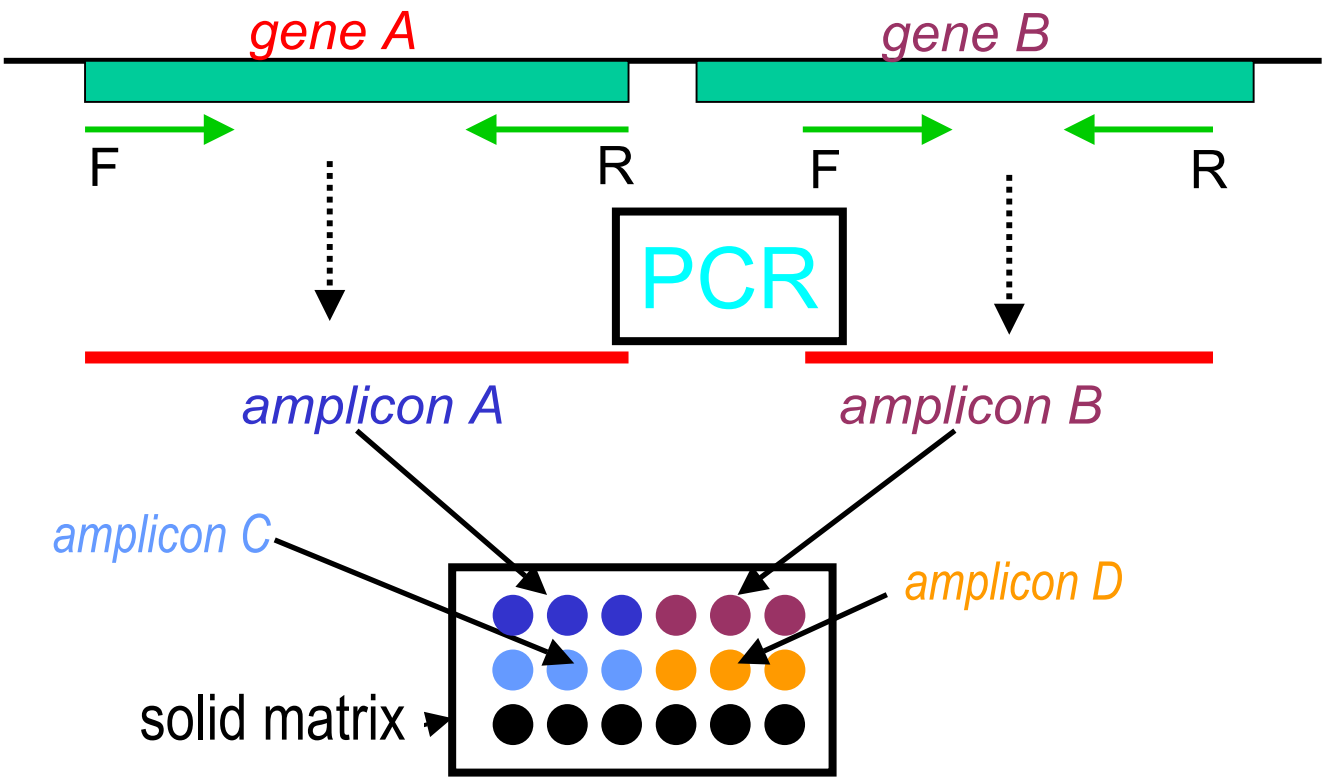


- Technology patented (and protected !!!) by Affymetrix
- You have to buy an expensive deal from them

"Spotted Arrays" The Pat Brown Method (Stanford)

- Relies on deposition of nucleic acid samples on a matrix using a spotting robot
  - robot has μm precision on X-Y axis
  - "spotting tip" dips onto DNA solution
    - picks up 0.5 to 1.0 μl
  - tip "stamps" down onto matrix, spotting some DNA on surface
    - deposits ~ 1-2 nl
    - print many chips from one "dip"
- DNA source:
  - plasmid DNA
  - insert DNA PCR-amplified from clone from a library (shotgun, cDNA…)
  - PCR products amplified from genomic DNA
  - Pre-synthesized oligonucleotides

eg. Example of spotted PCR amplicon chip



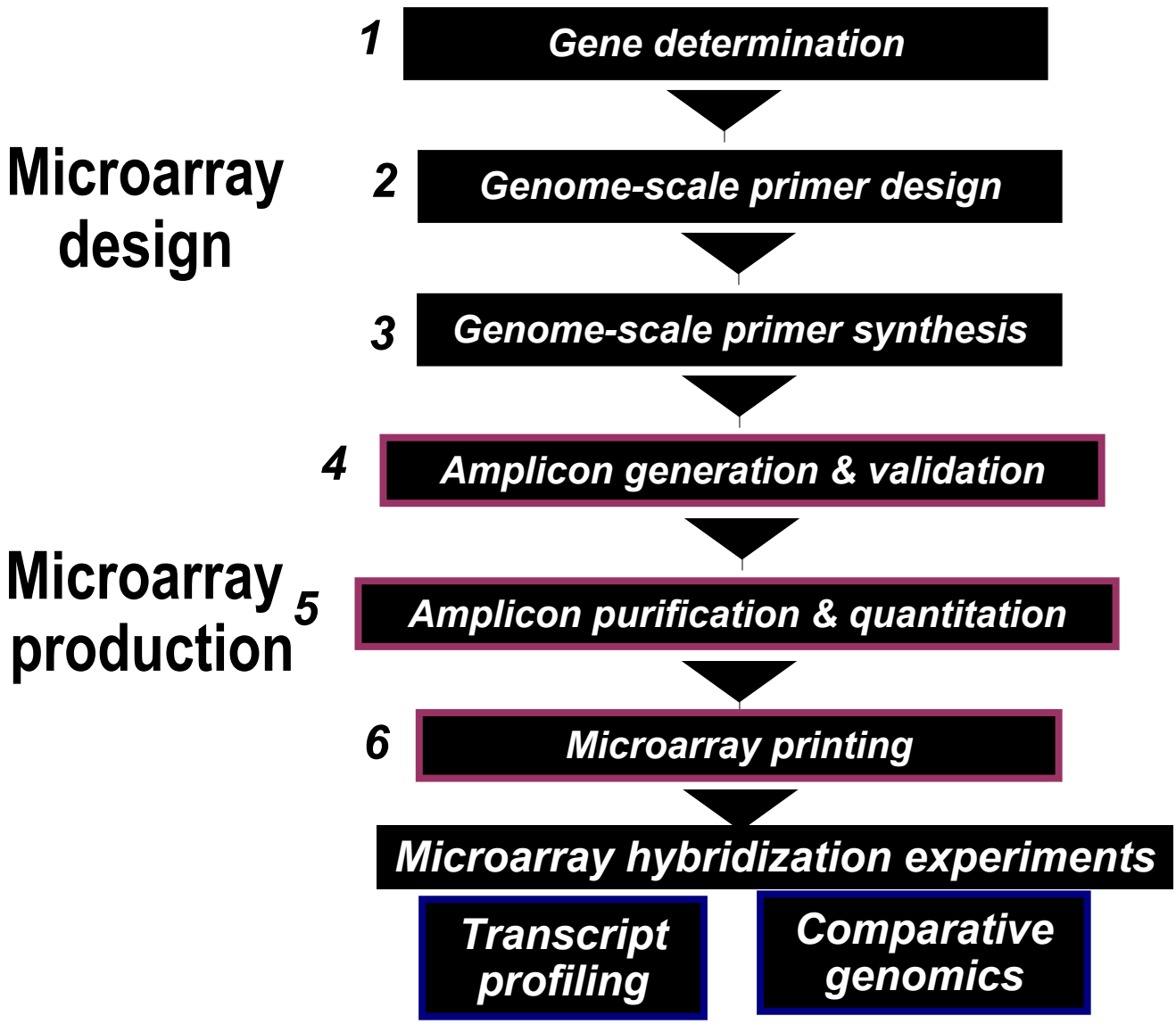
- "No patent" means "universal", less expensive
- once spotting samples ready, easy (and cheap) to make more chips

Microarrays and Inkjet Printing

- Ink-jet technology can be used for making arrays:
  - *in situ* synthesis of Oligonucleotide arrays:
    "activated" nucleotides can be delivered to specific sites on a surface
  - Spotted arrays:
    same a Stanford "contact arrayer" but uses inkjet to deliver sample

Array Design and Production from Genome Data

- You could start from a library
  - how do you ensure full genome coverage ?
- If you have sequence data you may as well use it !!!

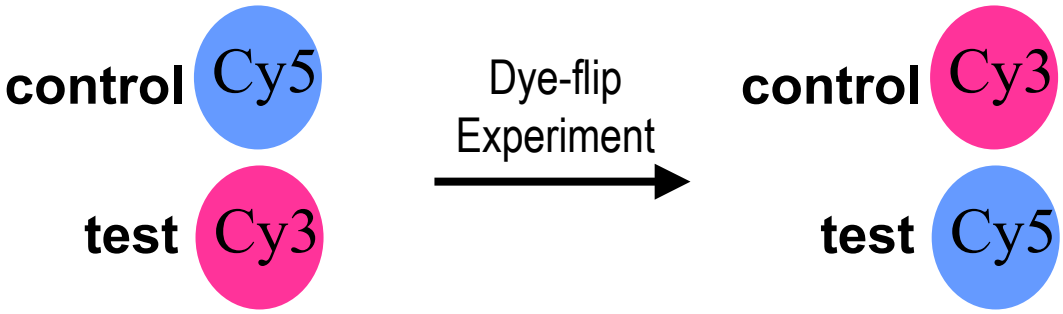


Fluorescent Labeling of DNA samples

Cy-dye Labeled Probes

- Cy3 absorbs at 550 nm, Cy5 absorbs at 650 nm
- can co-hybridize Cy3 and Cy5 labeled probes to the same chip
- You always run "test condition" vs. "control condition"
    - each condition labeled with different dye

- **Enzymatic Labeling:**
    - labeling tag introduced during enzymatic synthesis of probe
        - cDNA: reverse transcriptase
        - DNA: DNA polymerase
    - **Direct Labeling:**
        - fluorescently-tagged nucleotide introduced during enzymatic synthesis of probe
            - original method, dyes are bulky and enzyme doesn't like them
    - **Indirect Labeling:**
        - nucleotide with "amino-allyl" tag introduced during enzymatic synthesis of probe
        - fluorescent label attached by chemical reaction between "amino-allyl tagged" DNA and dye NHS-ester
            - superior method because no dye-effects on enzyme
- **Chemical Labeling:**
    - DNA (or RNA) directly labeled at G residues by activated dye
        - very good but **very** expensive



DNA microarray hybridization experiments

Comparative Genomic Hybridization

Cells

↓ DNA isolation

Genomic DNA

↓ Probe labeling

Genomic DNA probe

Hybridization to microarray

control Cy5

test Cy3

Transcript Profiling

Cells

↓ RNA isolation

RNA

↓ • cDNA synthesis
• Probe labeling

cDNA probe

Hybridization to microarray

control Cy3

test Cy5

↓ Laser Scanning

Cy3 / Cy5 Images

↓ Image Analysis

Spot Intensities

Post Analysis:

- "grouping" genes that behave similarly
- "grouping" samples that behave similarly