Use of microarrays to study gene expression

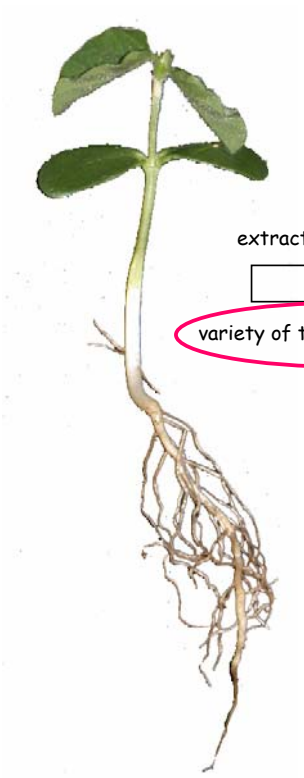
Part I: Developing Arrays

-Steve Clough
sjclough@uiuc.edu

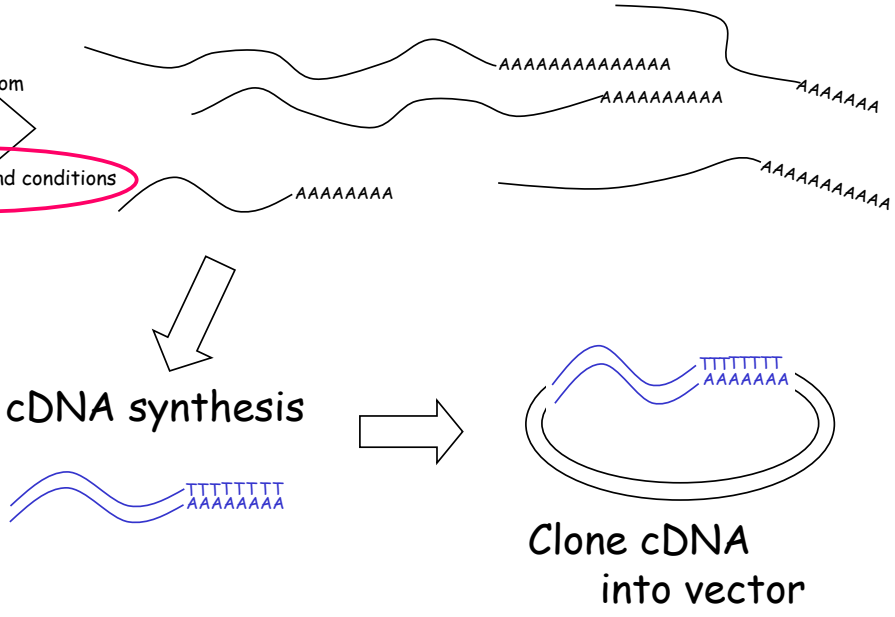
Part I: Developing Arrays

cDNA Arrays: start with cDNA Libraries

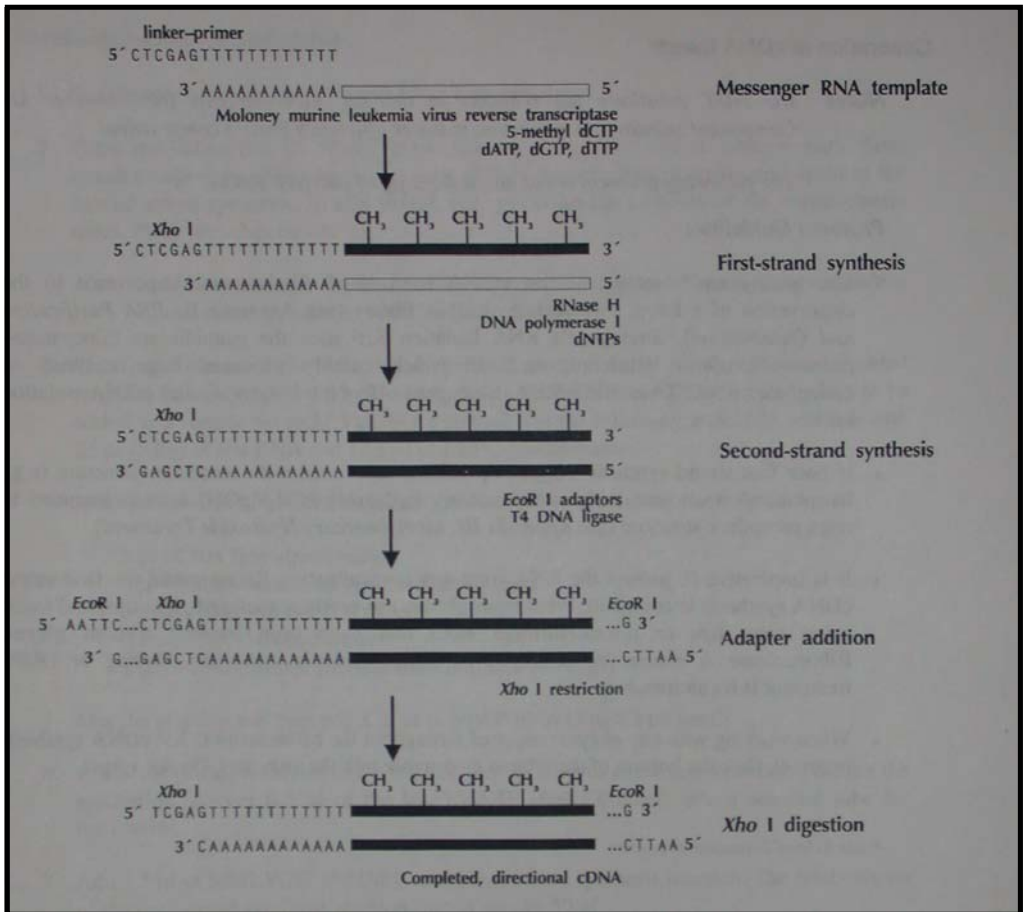
cDNA Library Synthesis
(represents expressed genes)

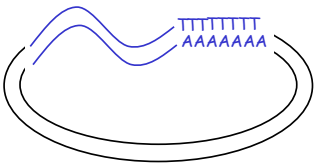


Steve Clough, USDA-ARS
University of Illinois, Urbana

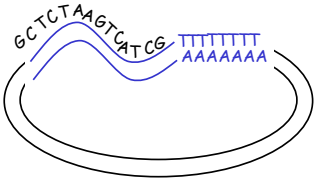


cDNA synthesis by Stratagene





cDNA clone



Sequence cDNA

GCTCTAAGTCATCGTACTAGATCT

= protein kinase

Compare EST sequence to database to identify

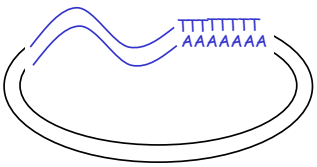
Eliminate duplicates to generate set of unique clones

Contig of overlapping cDNA clones

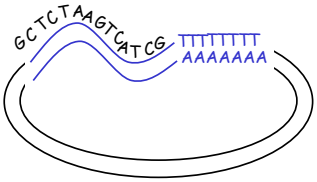
Pick longest one to represent group?

EST sequences usually about 400-600 nts

Steve Clough, USDA-ARS
University of Illinois, Urbana



cDNA clone



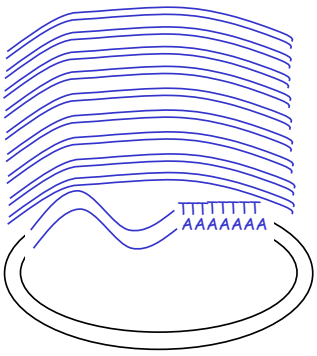
Sequence cDNA

GCTCTAAGTCATCGTACTAGATCT

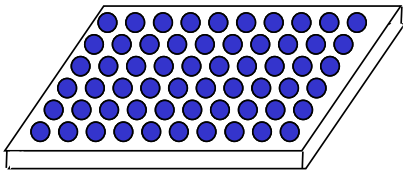
= protein kinase

Compare EST sequence to database to identify

Eliminate duplicates to generate set of unique clones



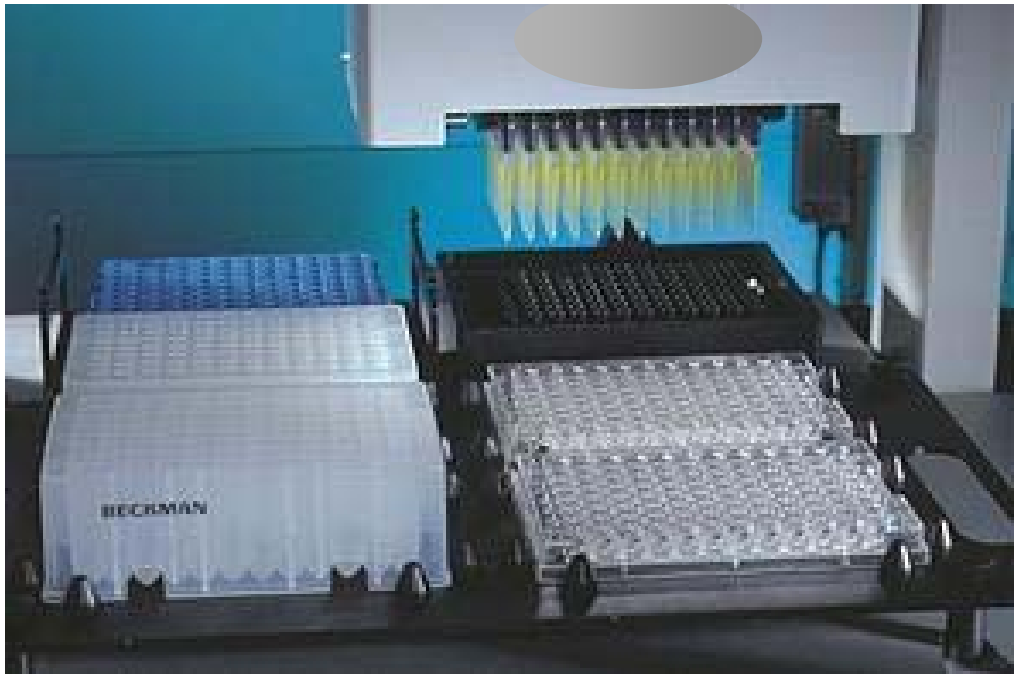
PCR amplify insert of unique clone set



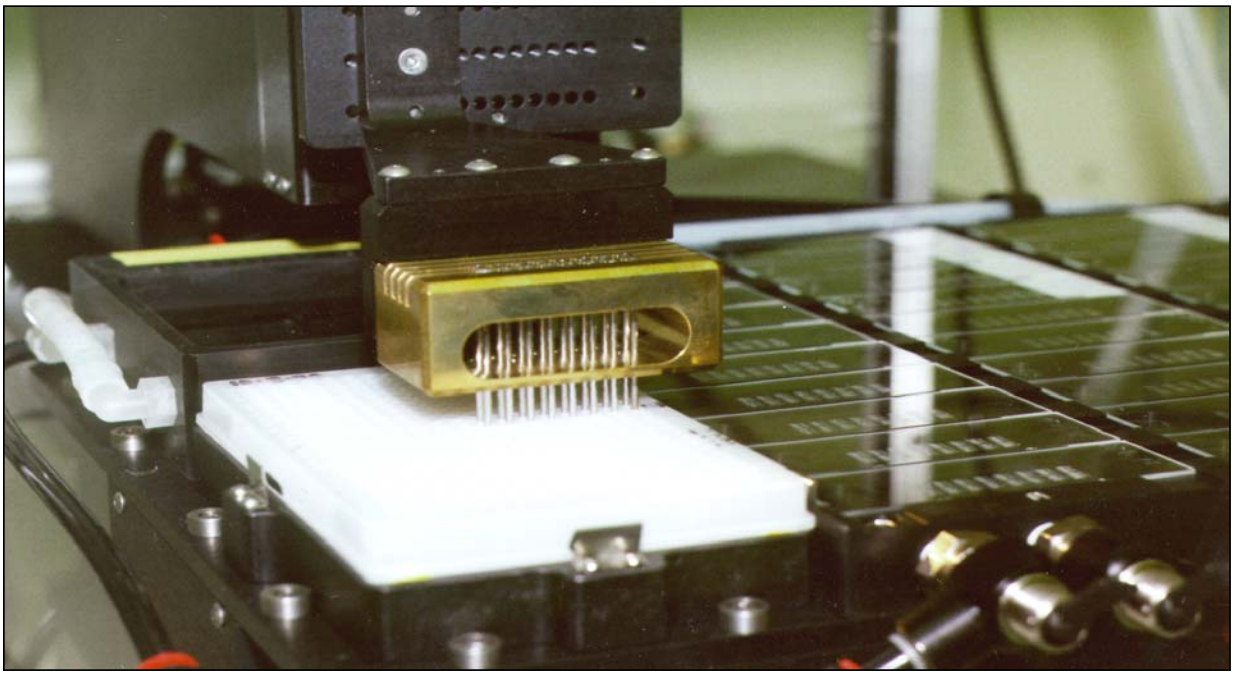
Pipette PCR products into microtiter plates to print onto slides

Steve Clough, USDA-ARS
University of Illinois, Urbana

Multimek™ 96-tip robotic pipettor

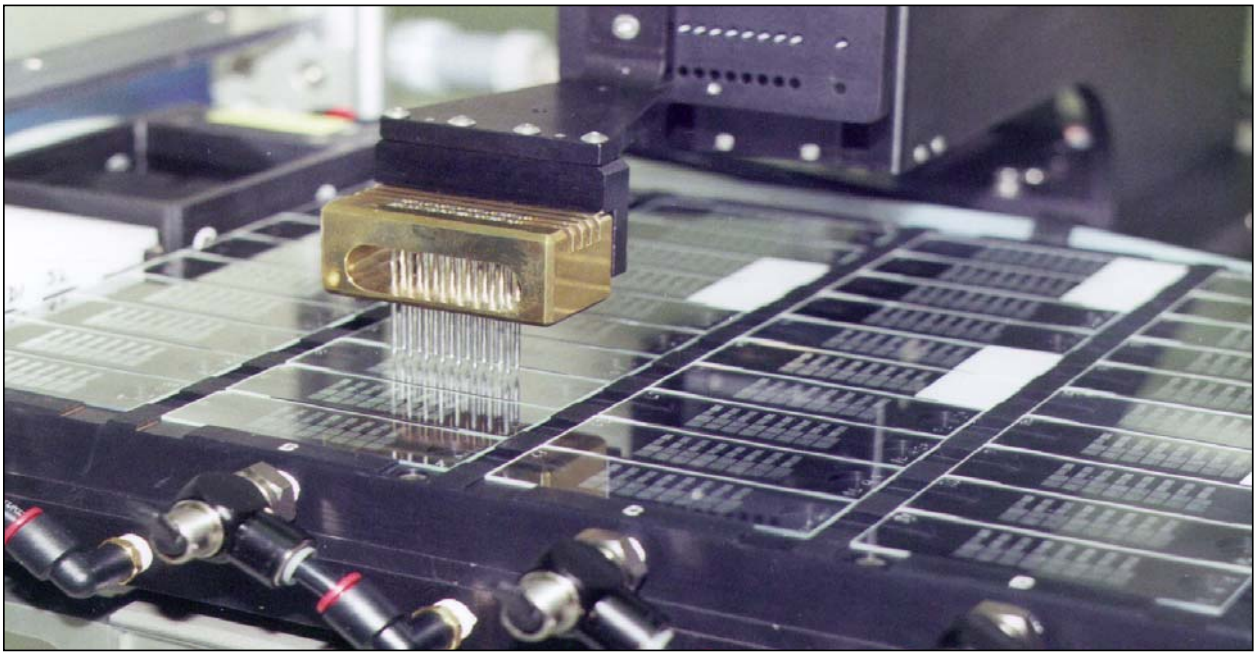


Printing microarrays – picking up PCR samples



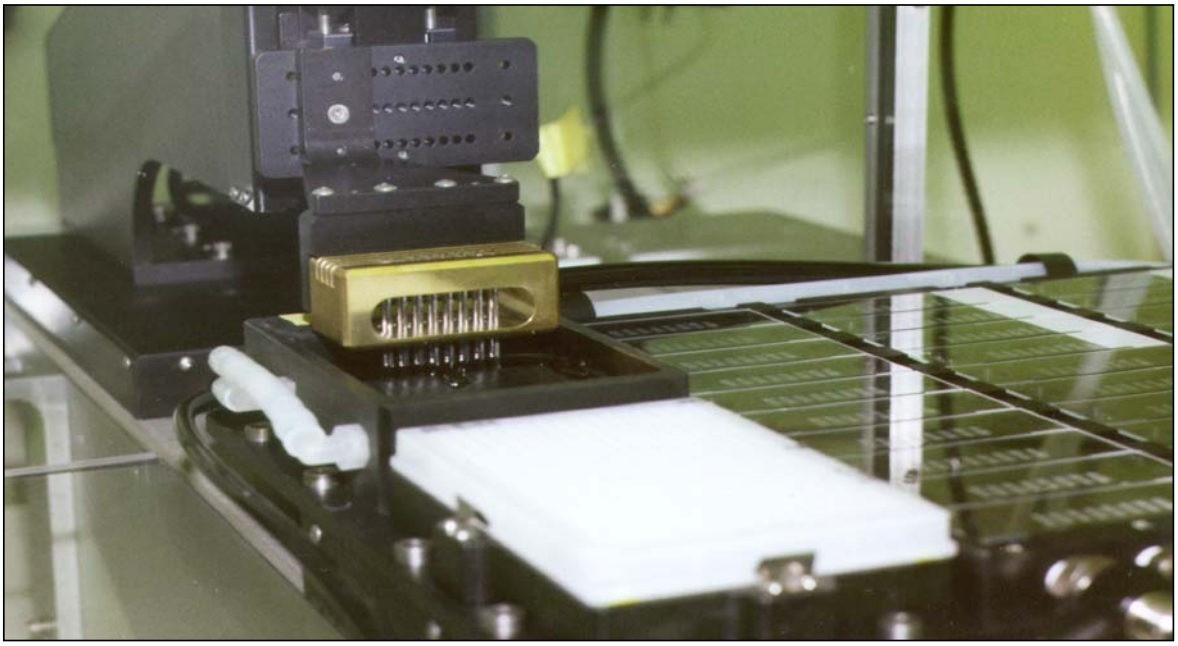
Steve Clough, USDA-ARS
University of Illinois, Urbana

Printing PCR products on glass slides

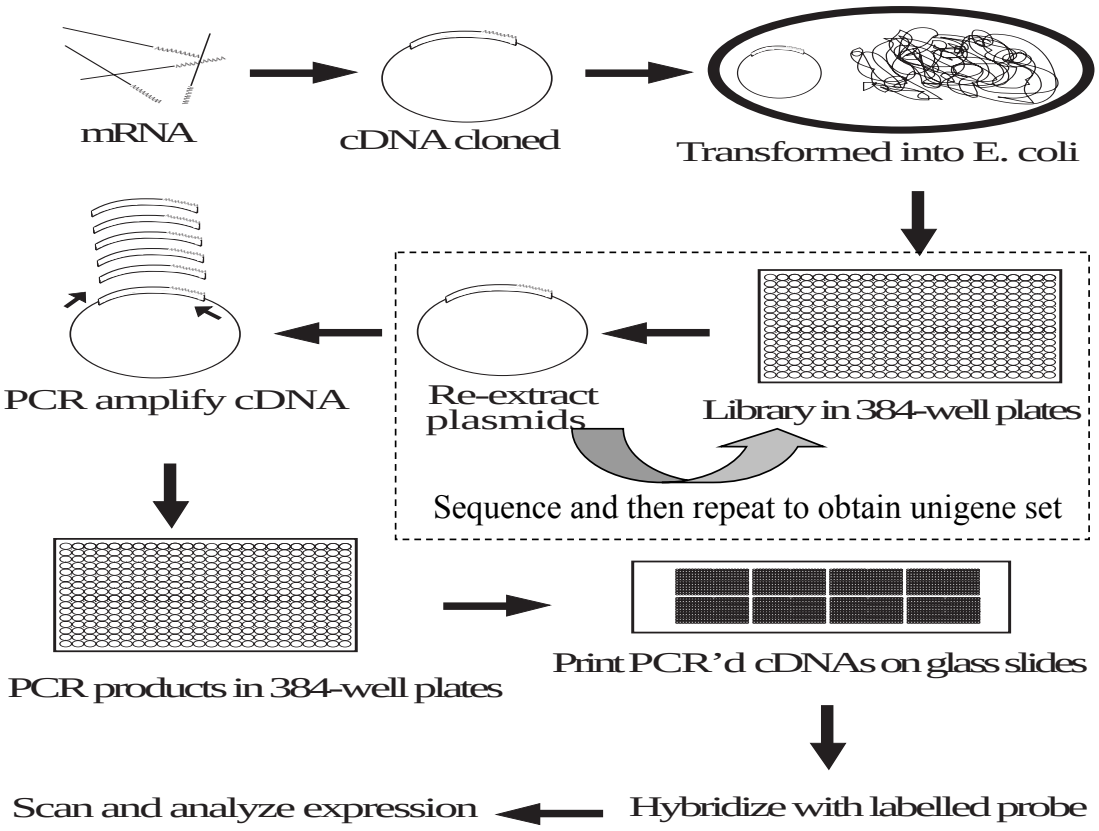


Steve Clough, USDA-ARS
University of Illinois, Urbana

Pin Washing Between PCR Samples

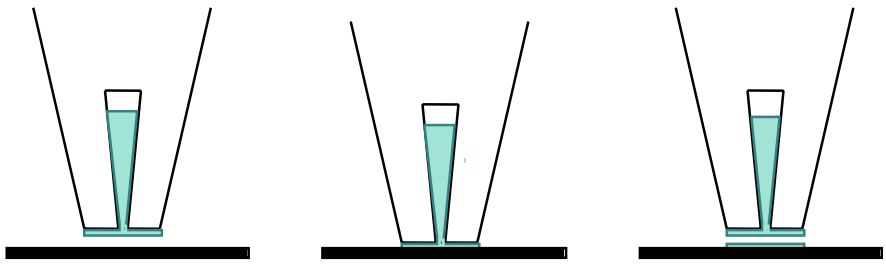


Steve Clough, USDA-ARS
University of Illinois, Urbana



Steve Clough, USDA-ARS
University of Illinois, Urbana

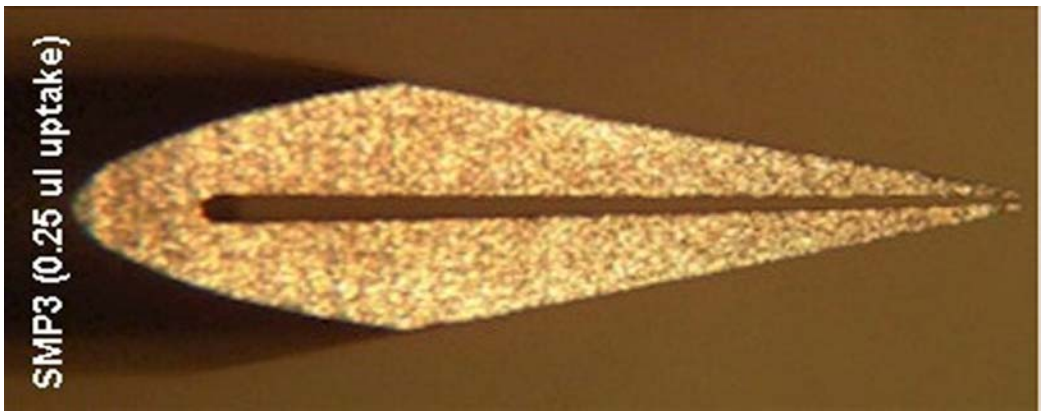
TeleChem ChipMaker2 Pins



Pin pick-up volume	100-250 nl
Spot diameter	75-200 um
Spot volume	0.2-1.0 nl

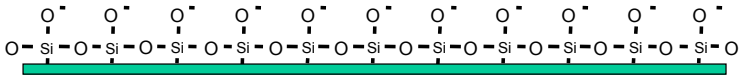
Steve Clough, USDA-ARS
University of Illinois, Urbana

TeleChem's ChipMaker 2 Micro Spotting Pins



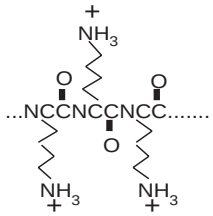
Slide Chemistry

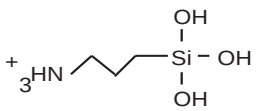
- Glass

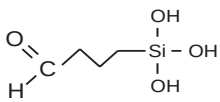


- Coatings

Amine		Aldehyde
Poly-L-lysine	Silanated	Silylated







We use SuperAmine and SuperAldehyde from TeleChem (arrayit.com)

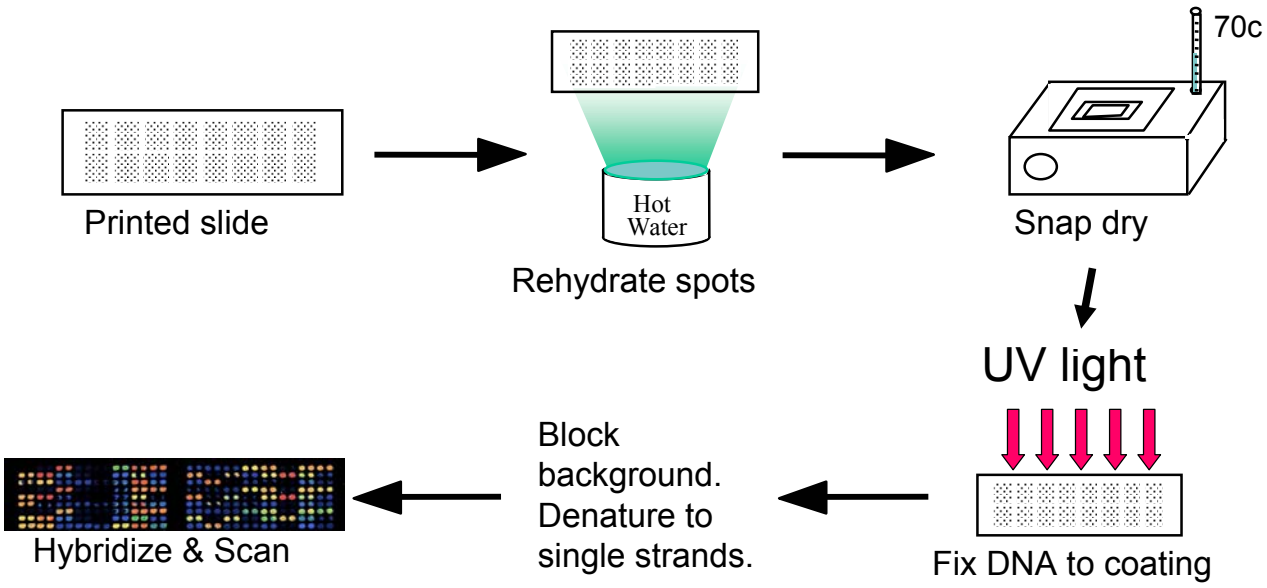
Steve Clough, USDA-ARS
University of Illinois, Urbana

DNA Attachment to slide coating

UV-light induced -- random production of free radicals leading to various C-C, C-O & C-N bonds between DNA and coating

Steve Clough, USDA-ARS
University of Illinois, Urbana

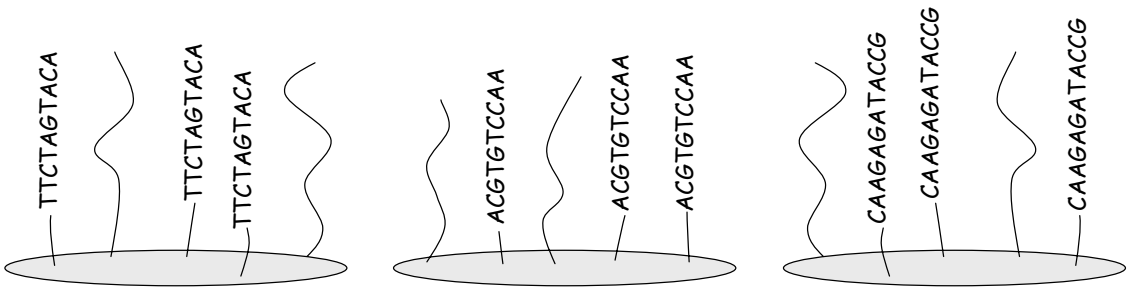
Post-Print Processing



Steve Clough, USDA-ARS
University of Illinois, Urbana

Spots of single-stranded DNA adhered to glass surface

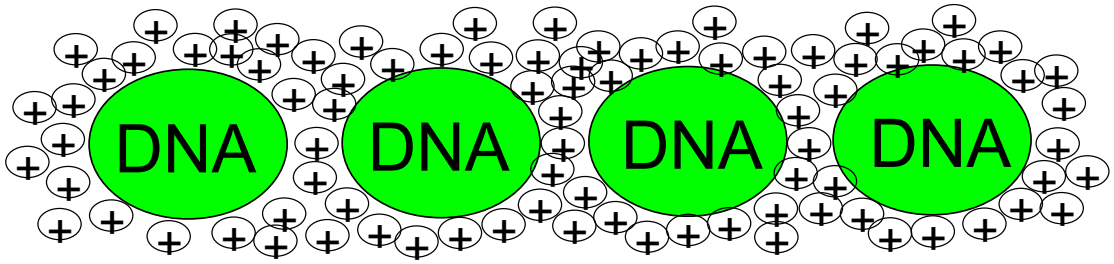
Typically 10-25,000 spots are printed on a standard 1" x 3" microscope slide



Note: DNA does not bind well to glass so glass is specially coated to allow ionic binding (poly-lysine slides) or covalent binding (amine or aldehyde slides)

Steve Clough, USDA-ARS
University of Illinois, Urbana

Blocking slides to reduce background. Example, positively charged amine slides.

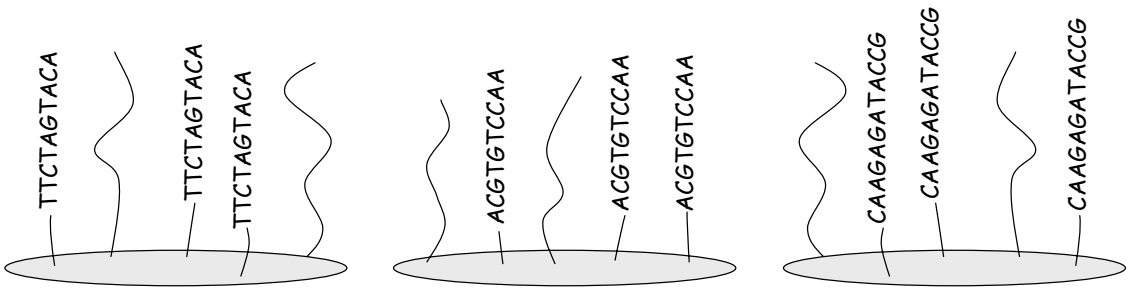


Wash with SDS to block charges and to remove excess DNA.
Then place in hot water to generate single strands.
Repeat SDS wash.

Steve Clough, USDA-ARS
University of Illinois, Urbana

Spots of single-stranded DNA adhered to glass surface

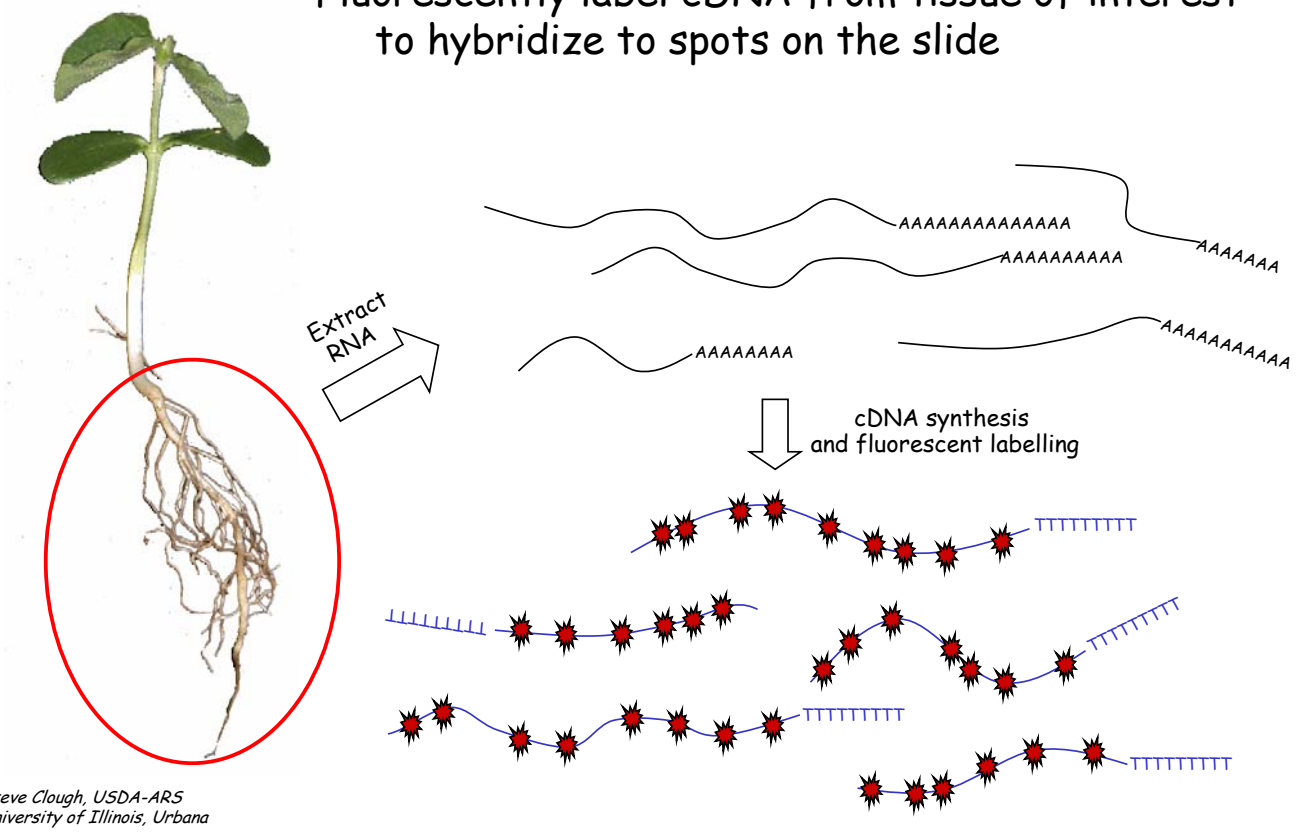
Typically 10-25,000 spots are printed on a standard 1" x 3" microscope slide



Note: DNA does not bind well to glass so glass is specially coated to allow ionic binding (poly-lysine slides) or covalent binding (amine or aldehyde slides)

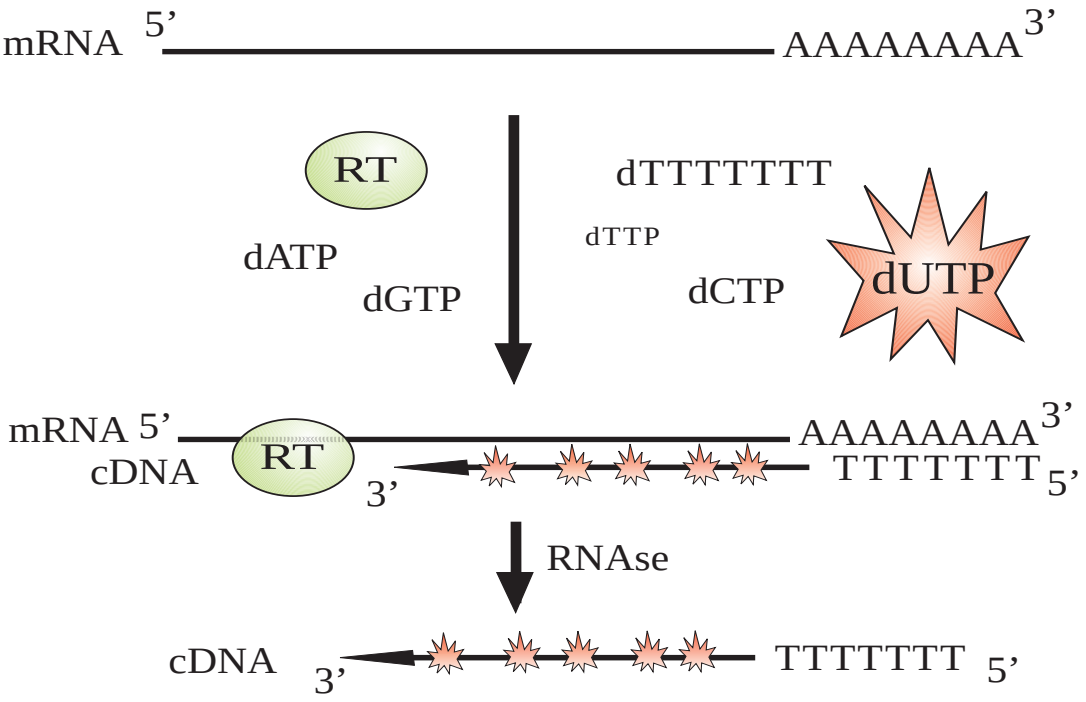
Steve Clough, USDA-ARS
University of Illinois, Urbana

Fluorescently label cDNA from tissue of interest to hybridize to spots on the slide



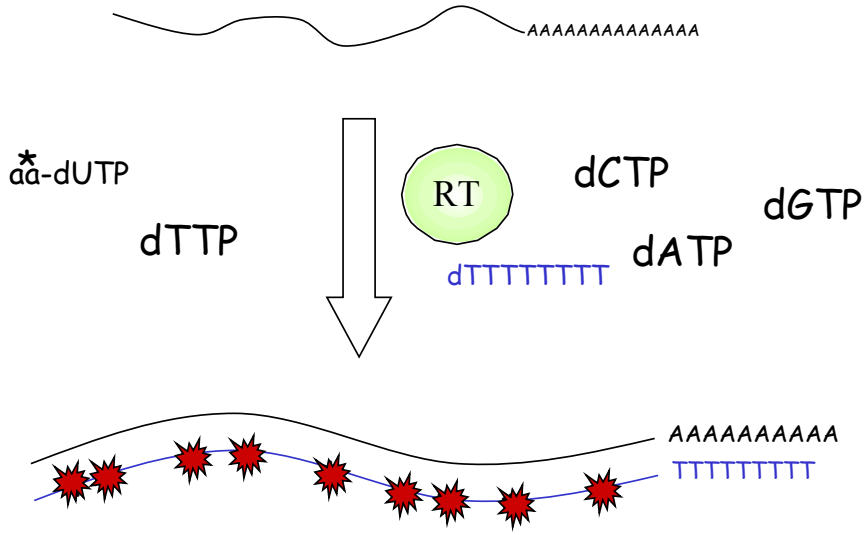
Steve Clough, USDA-ARS
University of Illinois, Urbana

Direct labelling with Reverse Transcriptase



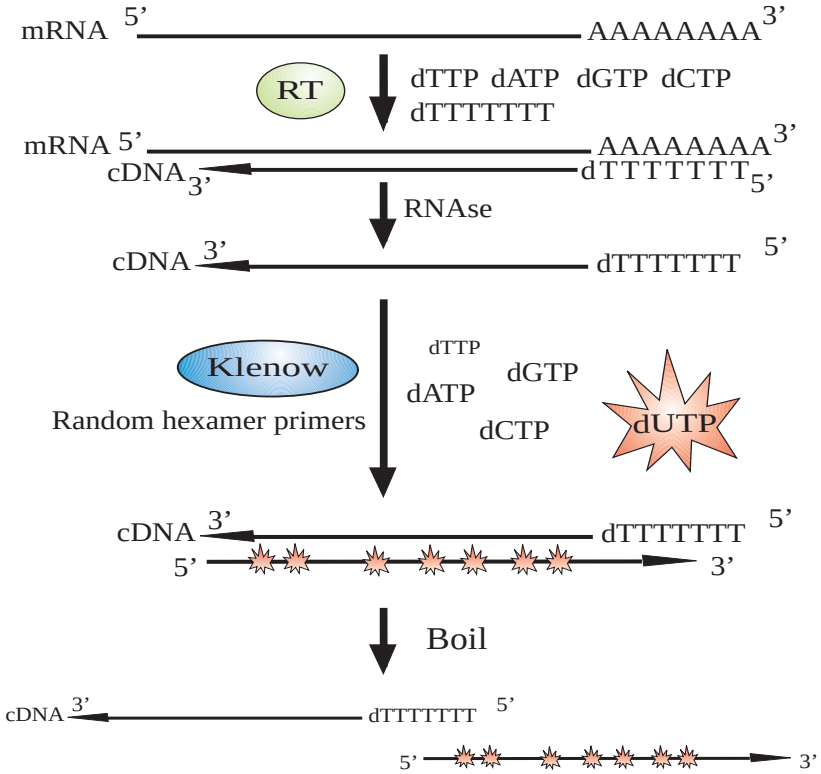
Steve Clough, USDA-ARS
University of Illinois, Urbana

Indirect labelling with aa-dUTP and Reverse Transcriptase



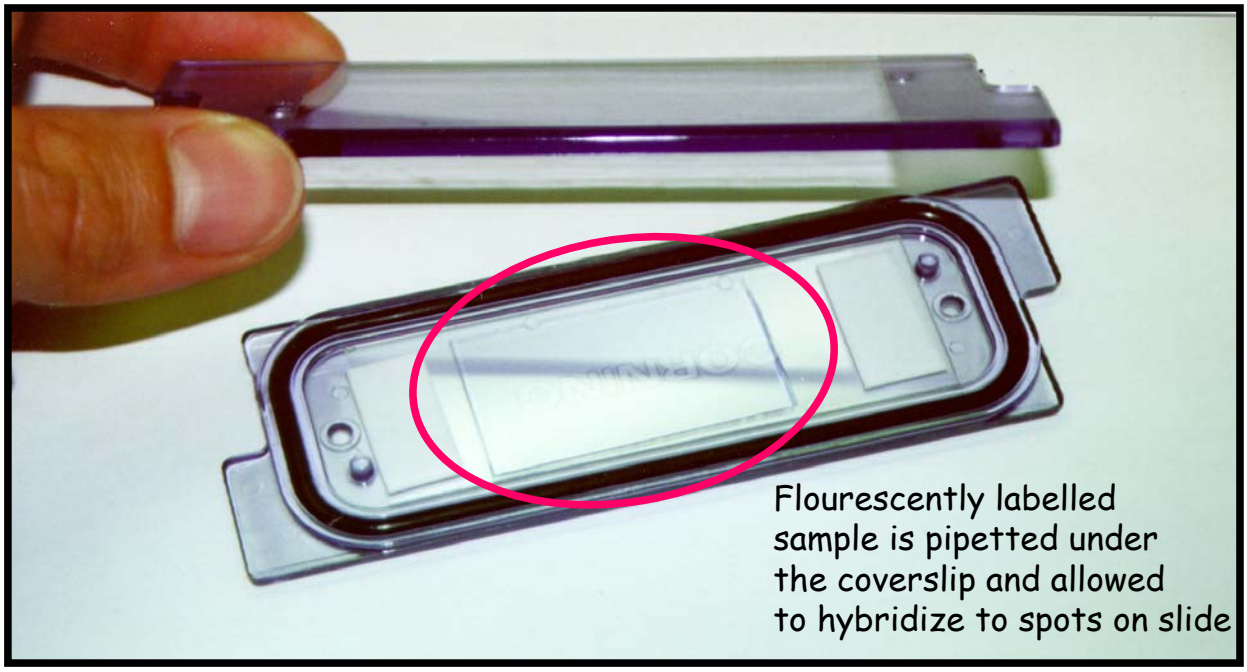
Steve Clough, USDA-ARS
University of Illinois, Urbana

Indirect labelling with Klenow



Steve Clough, USDA-ARS
University of Illinois, Urbana

Hybridization Chamber



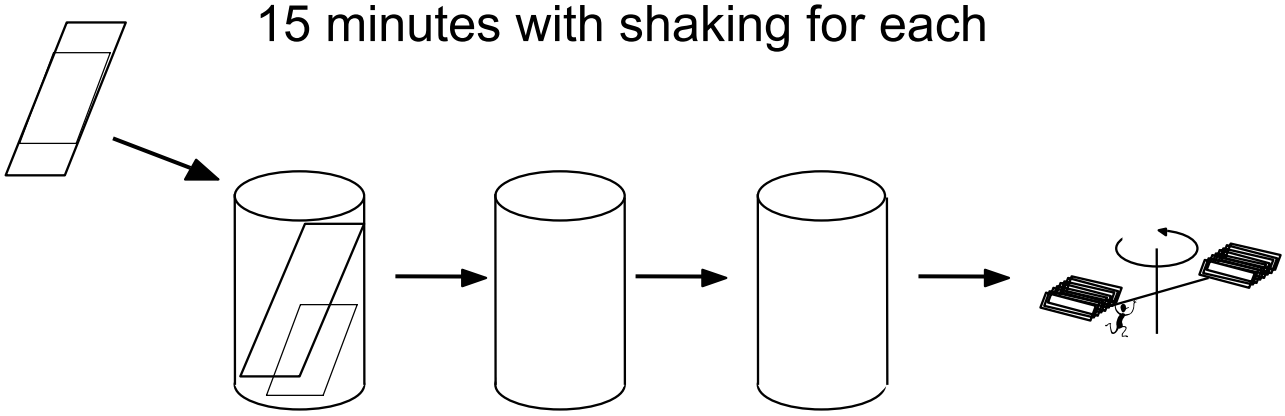
Steve Clough, USDA-ARS
University of Illinois, Urbana

Hybridization in Water Bath

Steve Clough, USDA-ARS
University of Illinois, Urbana

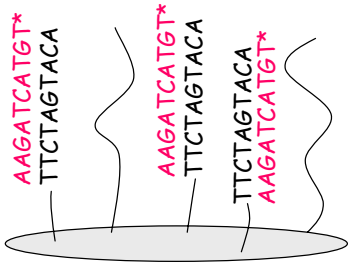
Washing After Hybridization

15 minutes with shaking for each

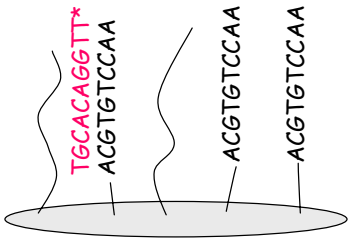


Steve Clough, USDA-ARS
University of Illinois, Urbana

Theory:
Spot A will fluoresce 3 times brighter than Spot B



Spot Gene A



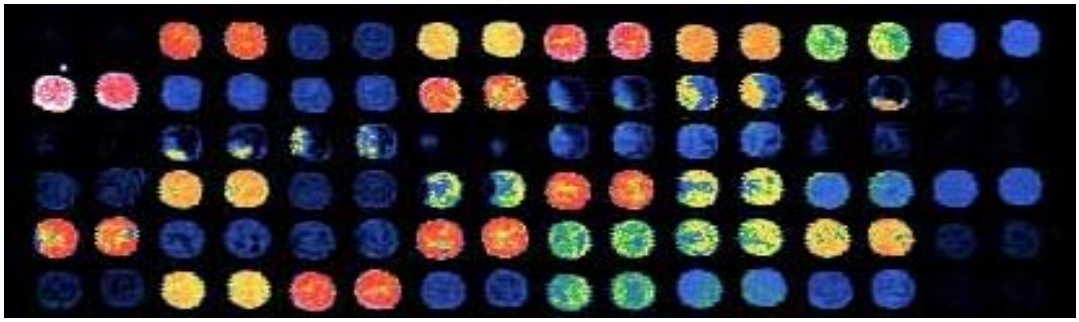
Spot Gene B

Steve Clough, USDA-ARS
University of Illinois, Urbana

We use a ScanArray Express from Perkin Elmer

Steve Clough, USDA-ARS
University of Illinois, Urbana

False Coloring of Fluorescent Signal



Scale of increasing fluorescent intensities

2 65,536

Stronger signal (16 bit image)

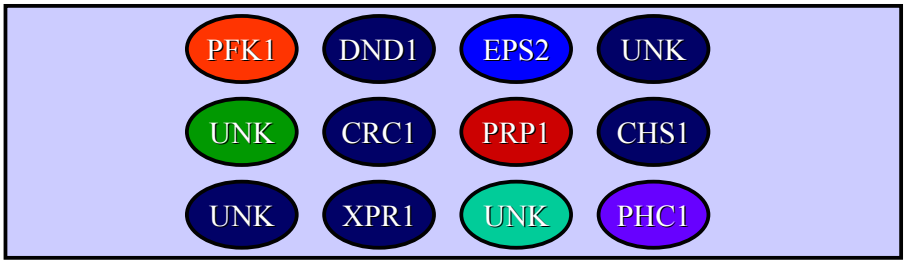
2^1 2^{16}

Steve Clough, USDA-ARS
University of Illinois, Urbana

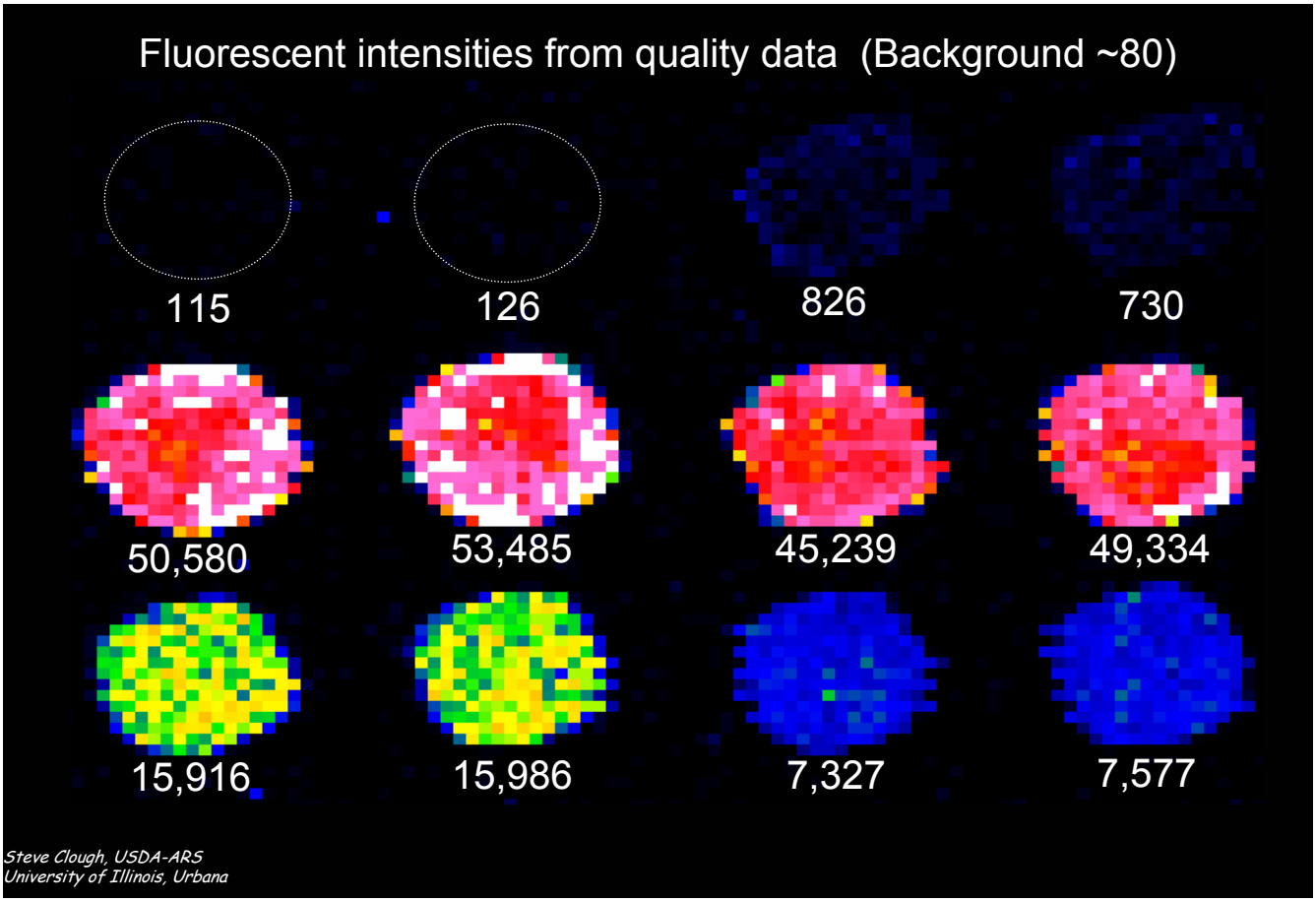
Principles behind gene expression analysis

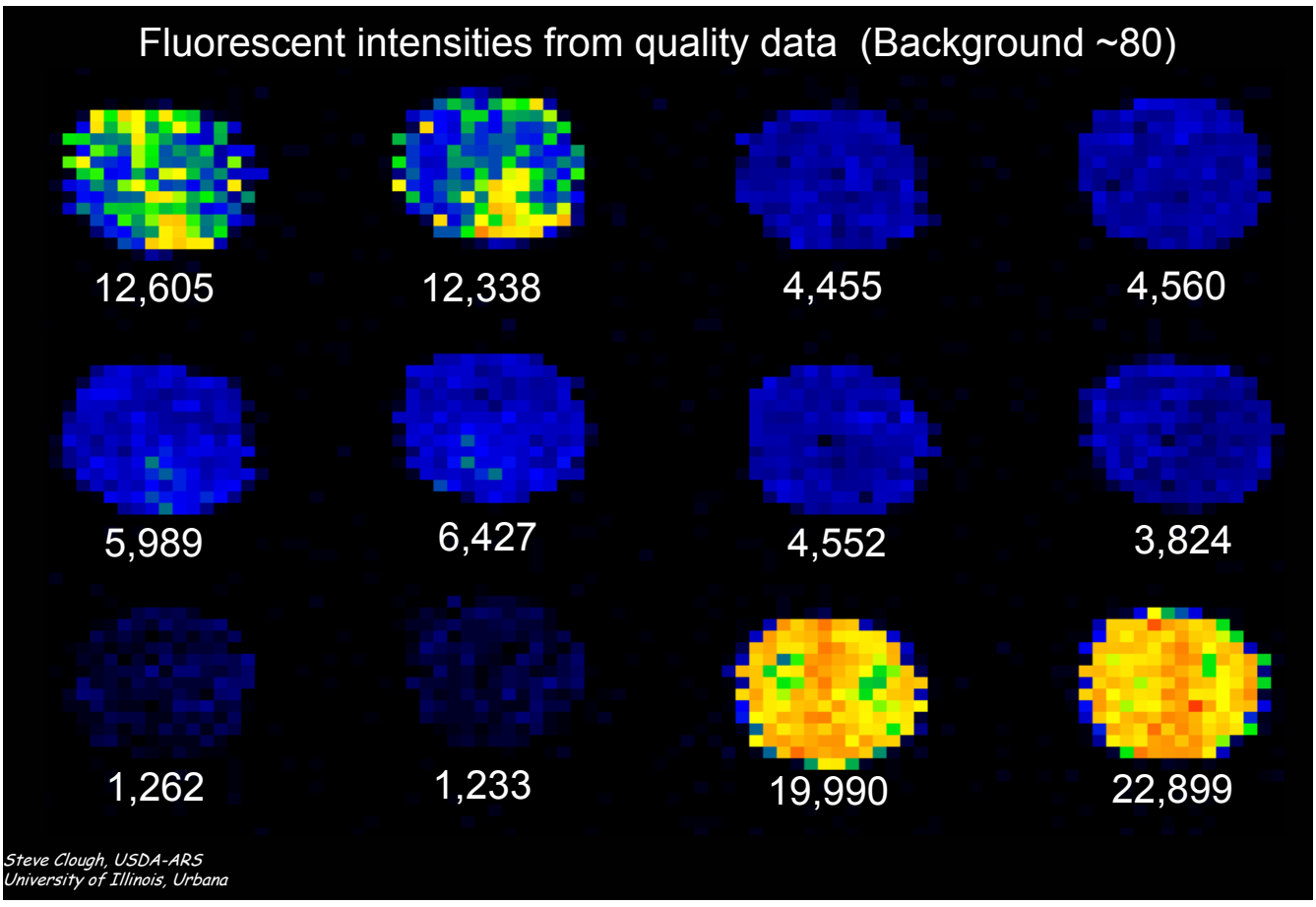
mRNA → Fluorescent cDNA

- Hybridized to array of individual spots of different genes
- Fluorescent intensity of spot is proportional to expression level
- Labelled representation of all recently expressed genes

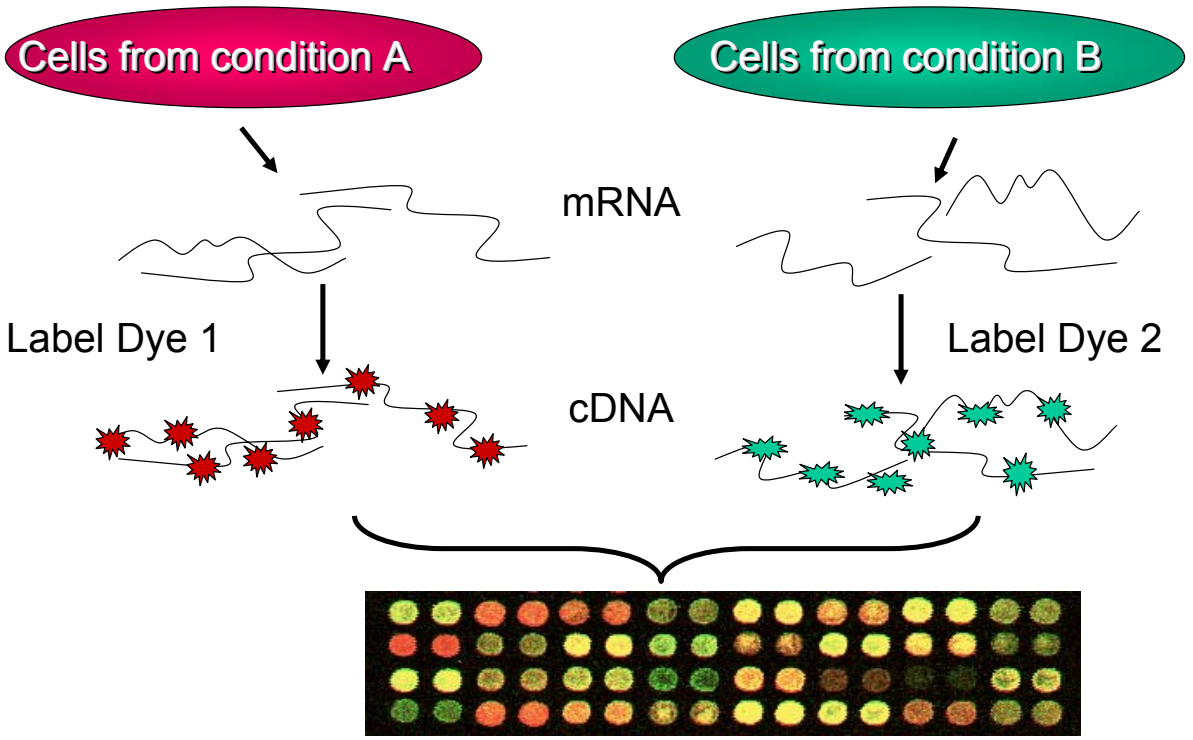


Steve Clough, USDA-ARS
University of Illinois, Urbana


Fluorescent intensities from quality data (Background ~80)
115
126
826
730
50,580
53,485
45,239
49,334
15,916
15,986
7,327
7,577
Steve Clough, USDA-ARS
University of Illinois, Urbana

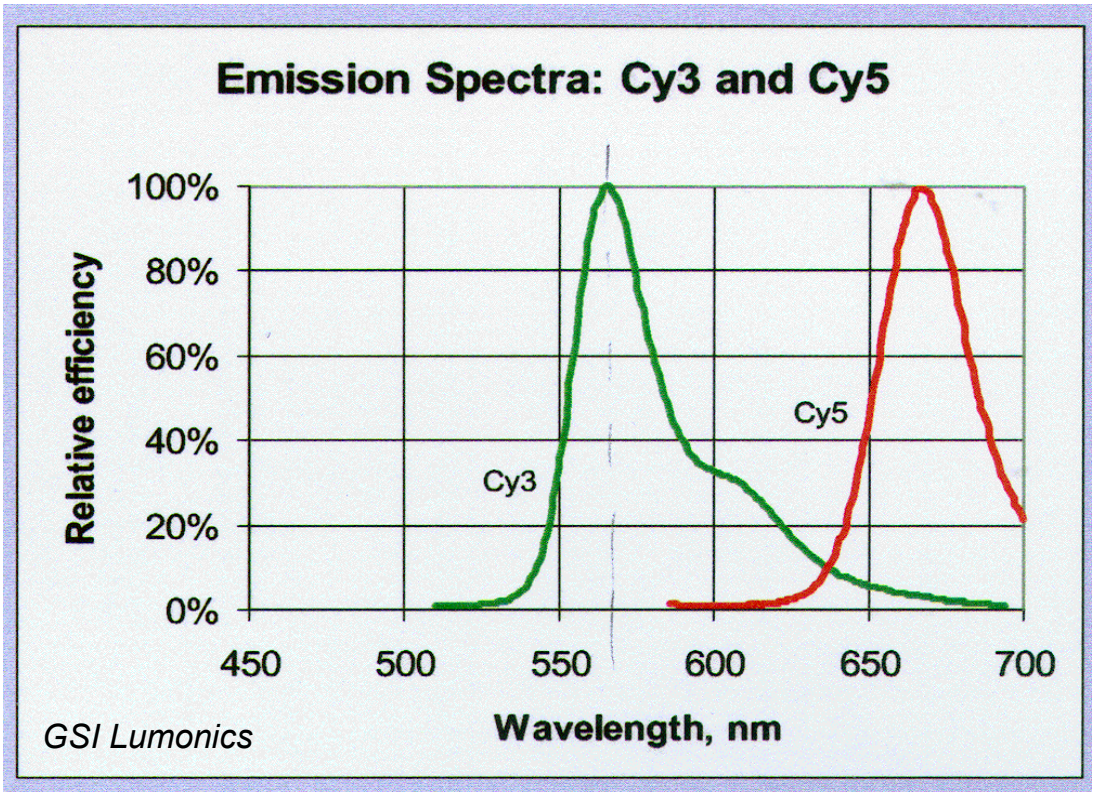

Fluorescent intensities from quality data (Background ~80)
12,605
12,338
4,455
4,560
5,989
6,427
4,552
3,824
1,262
1,233
19,990
22,899
Steve Clough, USDA-ARS
University of Illinois, Urbana

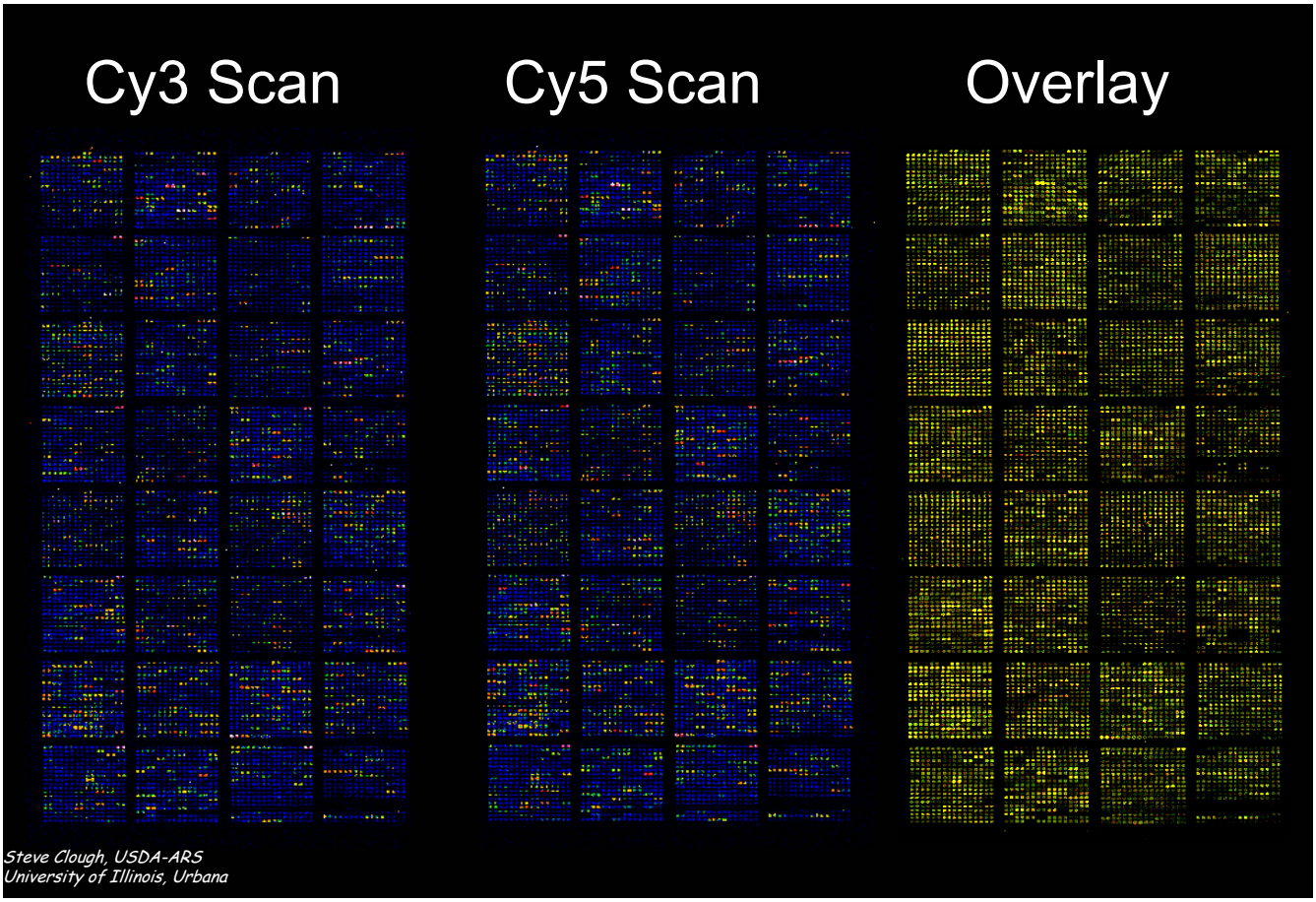
Ratio of Expression of Genes from Two Sources

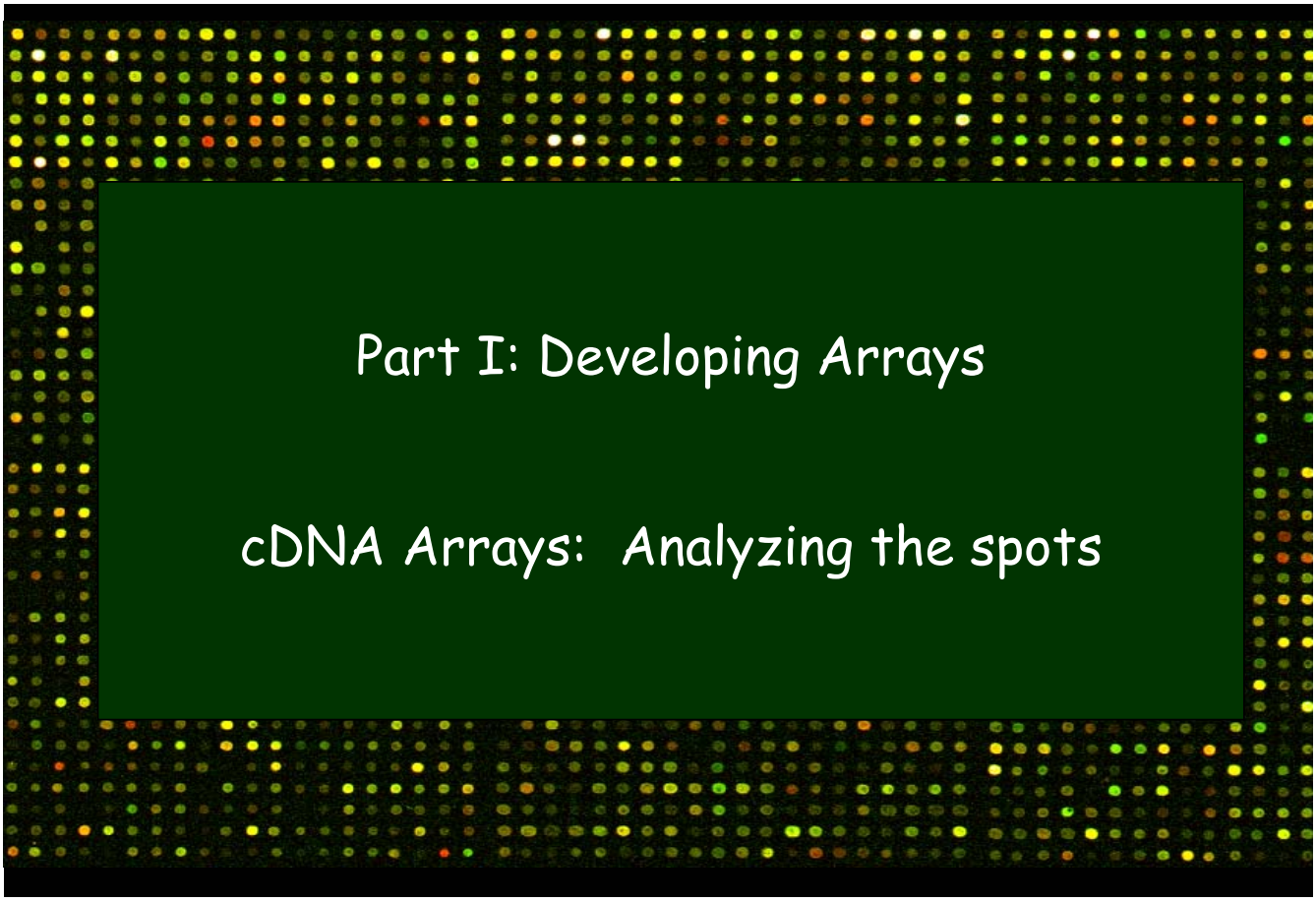


Steve Clough, USDA-ARS
University of Illinois, Urbana

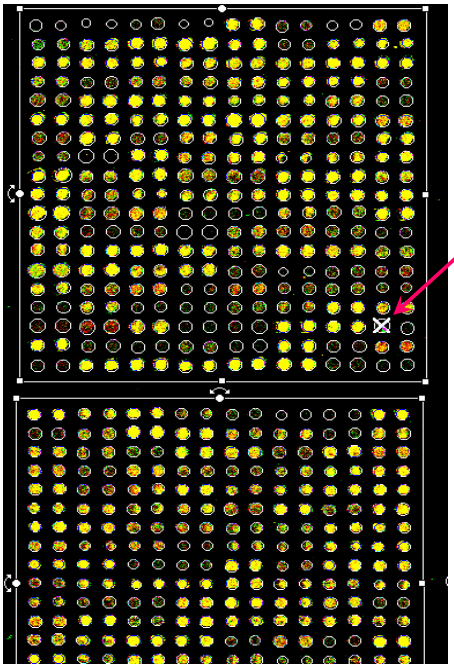
2 dyes with well separated emission spectra allow direct comparison of two biological samples on same slide




Cy3 Scan
Cy5 Scan
Overlay
Steve Clough, USDA-ARS
University of Illinois, Urbana


Part I: Developing Arrays
cDNA Arrays: Analyzing the spots

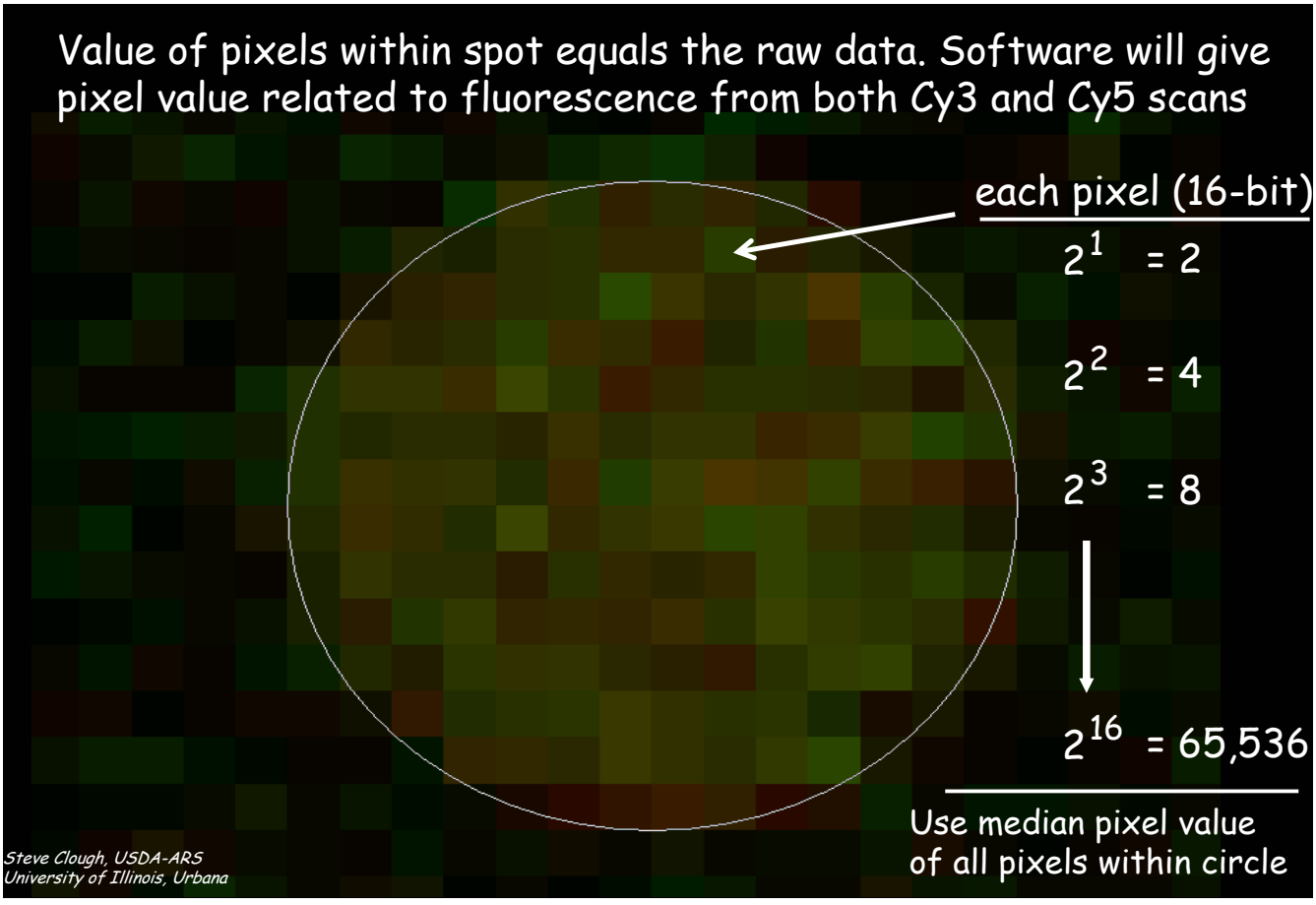
Use software such as GenePix to extract data from image



1. Locate spots, define spot area, collect data from pixels within spots
2. Flags bad spots (ex: dust in spot)
3. Calculates ratio Cy5 fluorescent intensity over Cy3 intensity for each spot
4. Produces tab-delineated tables for import to analysis programs

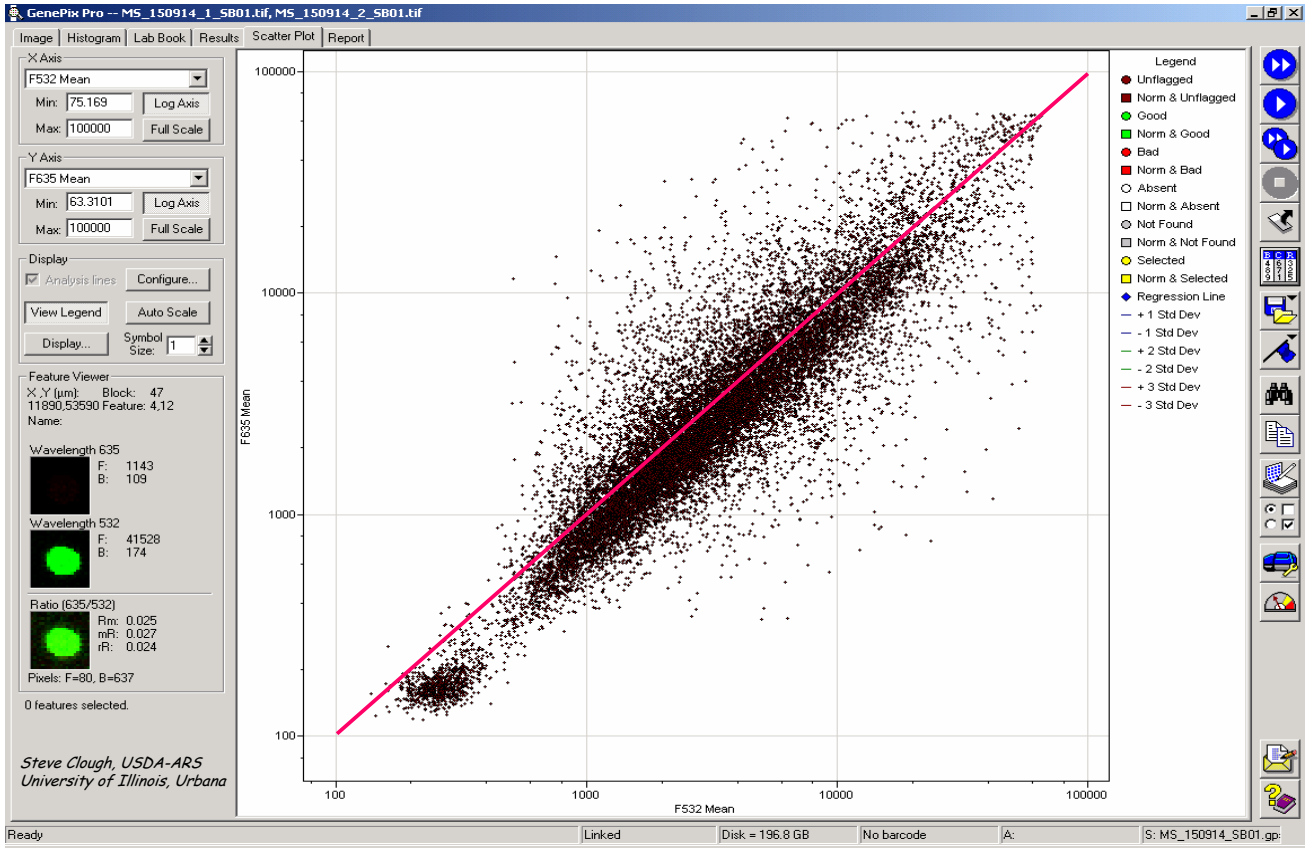
Steve Clough, USDA-ARS
University of Illinois, Urbana

Value of pixels within spot equals the raw data. Software will give pixel value related to fluorescence from both Cy3 and Cy5 scans



Steve Clough, USDA-ARS
University of Illinois, Urbana

Scatter plot of individual slide gives idea of quality and expression

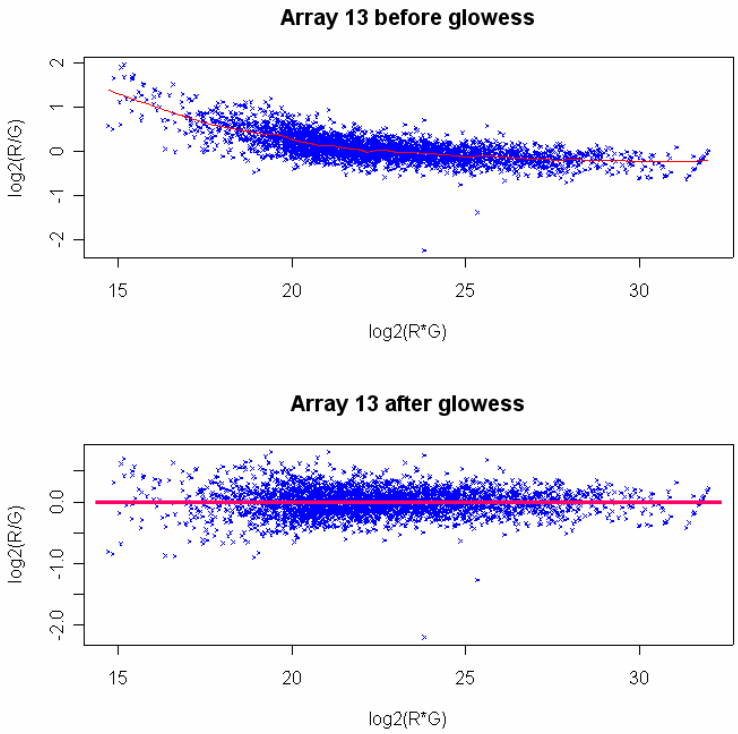


Clean Data and Normalize Intensities

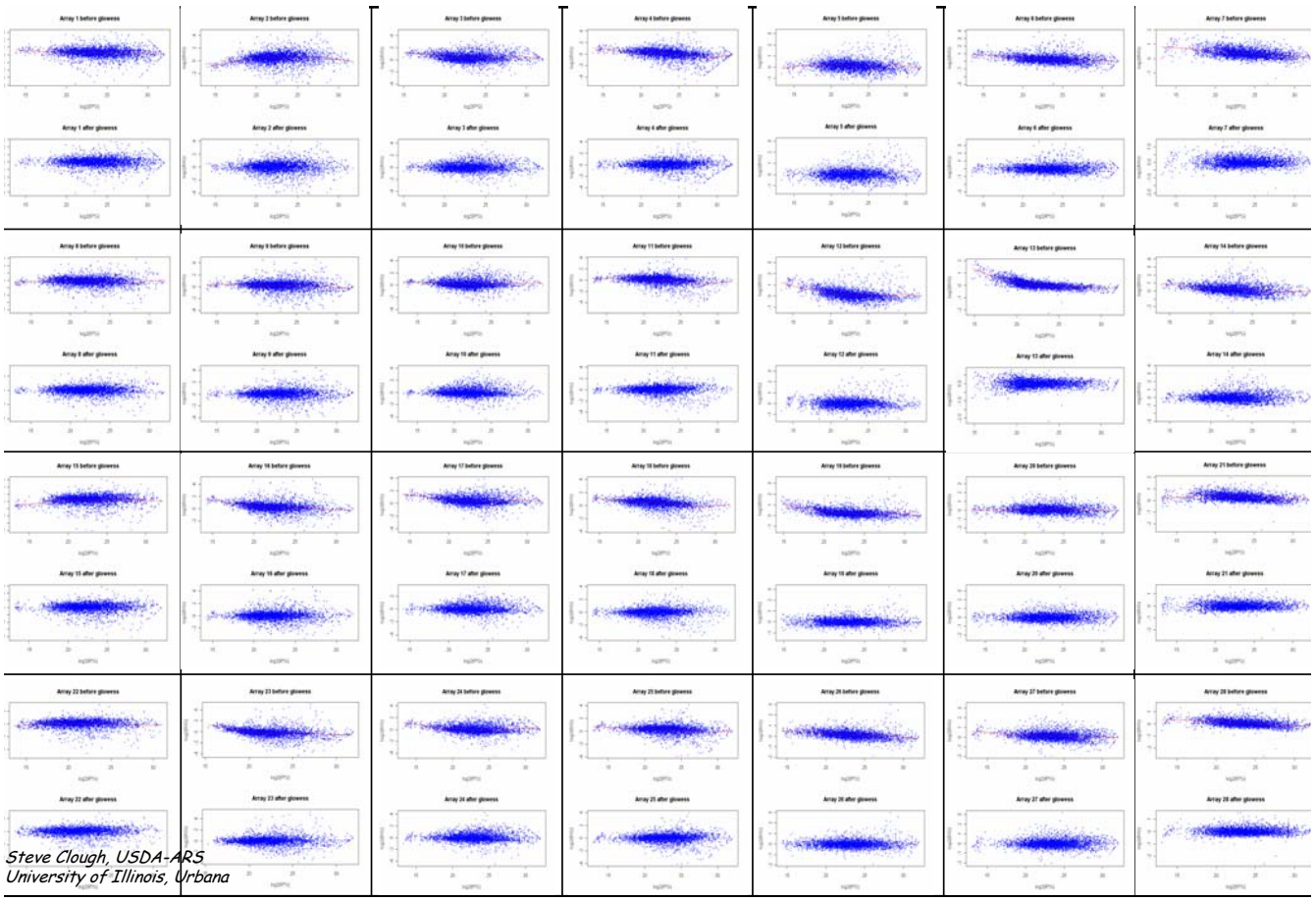
- Removed ‘Flagged’ spots.
- Removed weak spots (high variability).
- Log2 transform.
- Run microarray normalization program such as glowess in R / MAANOVA.
- Use the Log2 normalized intensities values in a statistical package to determine significantly differentially expressed genes in your experiment.

Steve Clough, USDA-ARS
University of Illinois, Urbana

Normalized data shown as M/A plots



Steve Clough, USDA-ARS
University of Illinois, Urbana



Steve Clough, USDA-ARS
University of Illinois, Urbana

Plant Microarray Controls

- Negative: mammalian genes
- Positive: high, medium, low expressers
  tissue specific
  ubiquitous
- Miscellaneous: transgenes
  bacterial
- Labelling efficiency: spiked control mRNA--
  genes that are non-homologous to plants.

Steve Clough, USDA-ARS
University of Illinois, Urbana

Negative Plant Controls

- Spotting solution
- Genes not present in plant
  Mammal-specific:
  antibody / immunoglobulin
  neuro-related
  myosin
  etc.
- Verify 'plant negative' by BLAST against plant databases

Steve Clough, USDA-ARS
University of Illinois, Urbana

Positive Plant Controls

- Tissue specific, high expressers:
  - ex: cotyledon: conglycinin
    roots: auxin down regulated gene 12
    leaves: RUBISCO (small chain)
- Ubiquitous:
  - ex: ubiquitin (med-high)
    EF1 (med-high)
    DAD1 (low-med)
    tubulin (med-high)

Steve Clough, USDA-ARS
University of Illinois, Urbana

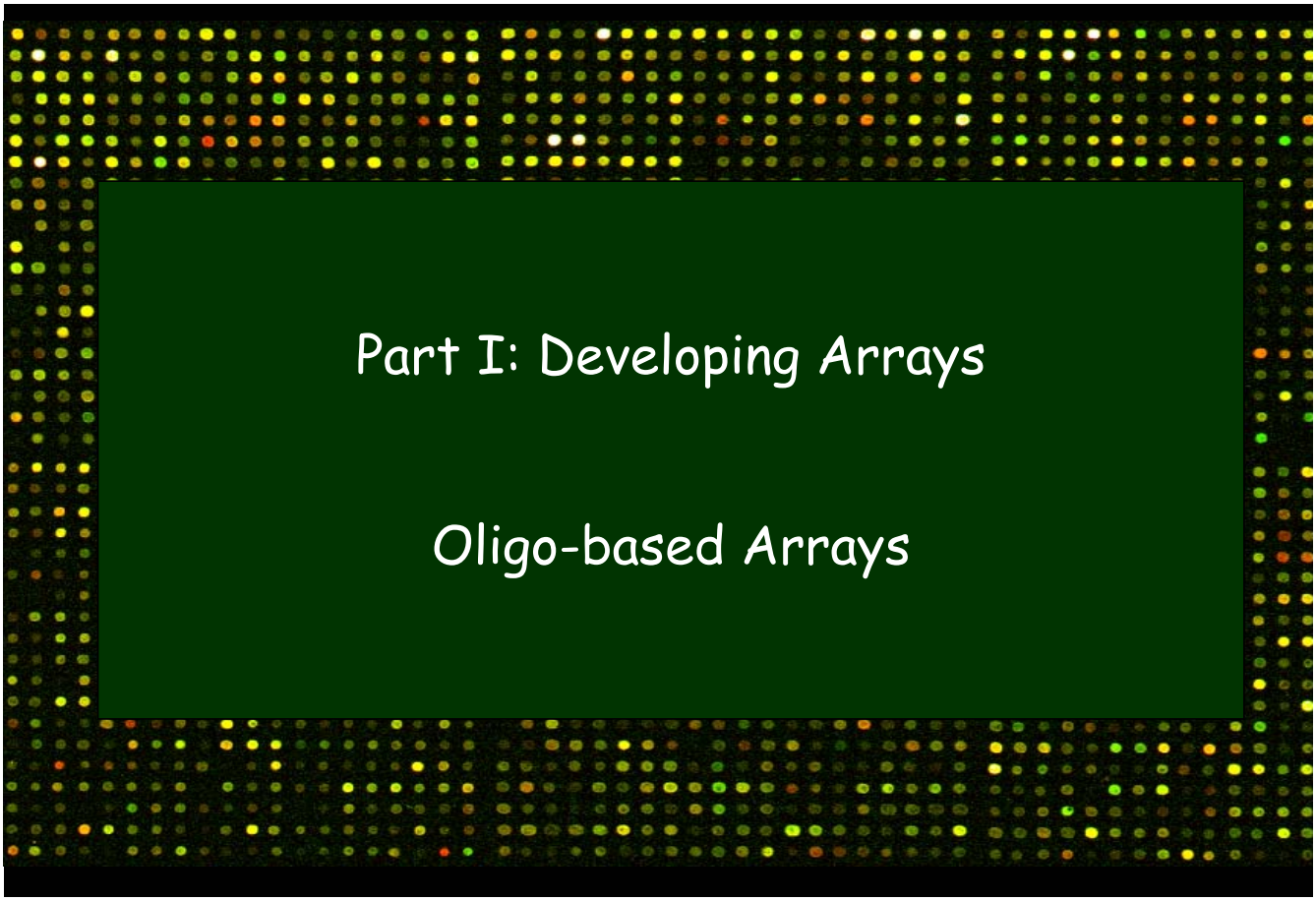
Labelling Efficiency Controls

- Spiked mRNAs:
  - Mammalian genes, non-homologous to plant genome.
    Select several spiking controls (ex: 4).

<u>To use</u>:

Include them on the array.
Clone (with polyT tail) into a T7 or T3 expression vector
Or PCR with T7 or T3 promoter attached to 5' primer
and a poly (dT) to the 3' primer.
Invitro transcribe with T7 or T3 RNA polymerase.
Add this 'mRNA' to your labelling reactions--
each one at a different concentration level
to span the dynamic range of fluorescent intensities

Steve Clough, USDA-ARS
University of Illinois, Urbana



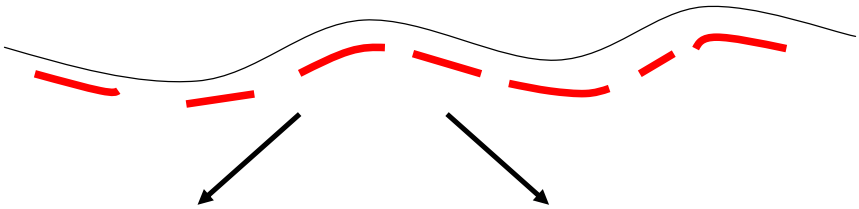
cDNA Arrays vs Oligo Arrays

- cDNA: spot a collection of PCR products from various cDNA clones
- Oligo: spot collections of oligos that span known/predicted ORFs
  - need sequence info
  - only option for prokaryotes
  - 'shagged rug' spots

Steve Clough, USDA-ARS
University of Illinois, Urbana

Oligo-based Microarrays

Design specific oligos for every ORF



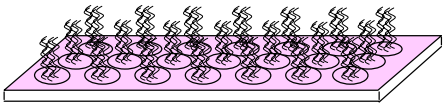
Spotted Microarray

Synthesize with 5'-amino linker

Design one to multiple oligos/ORF

Collect in 384-well plates

Spot on aldehyde coated slides



Affymetrix Gene Chips

Synthesize oligo directly on chip

Proprietary photolithography synthesis

11 oligo/ORF plus mismatches

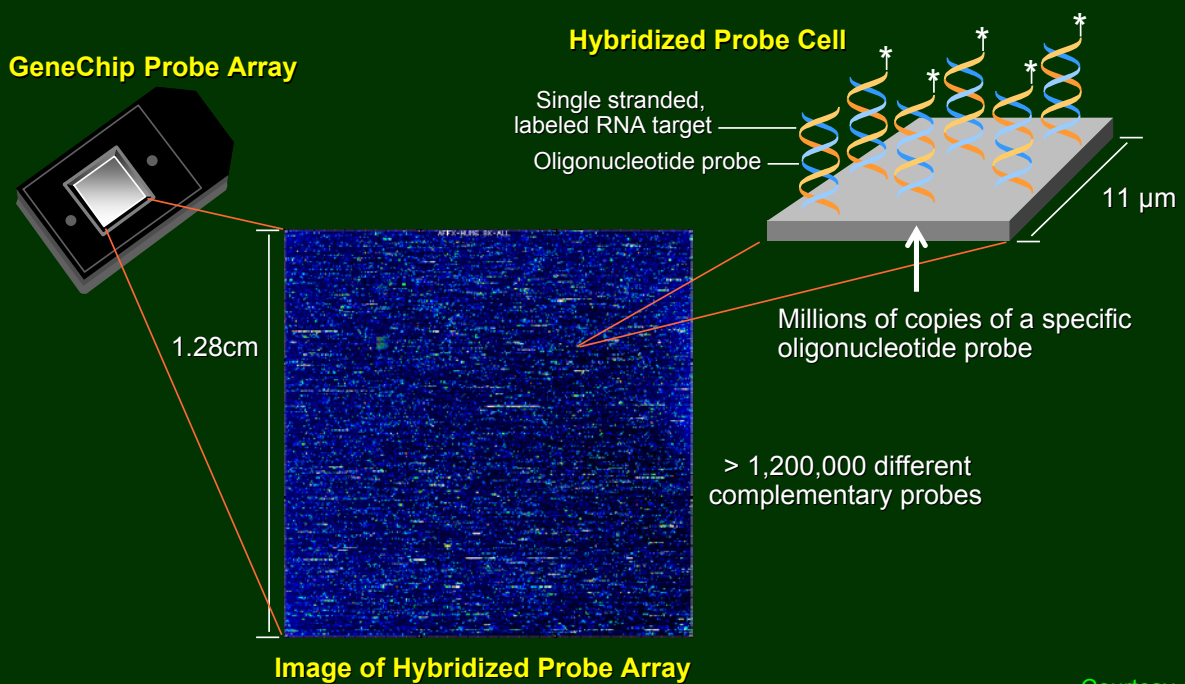
Spotted oligo termed the 'probe'

Perfect match oligos

1-base mismatch oligos

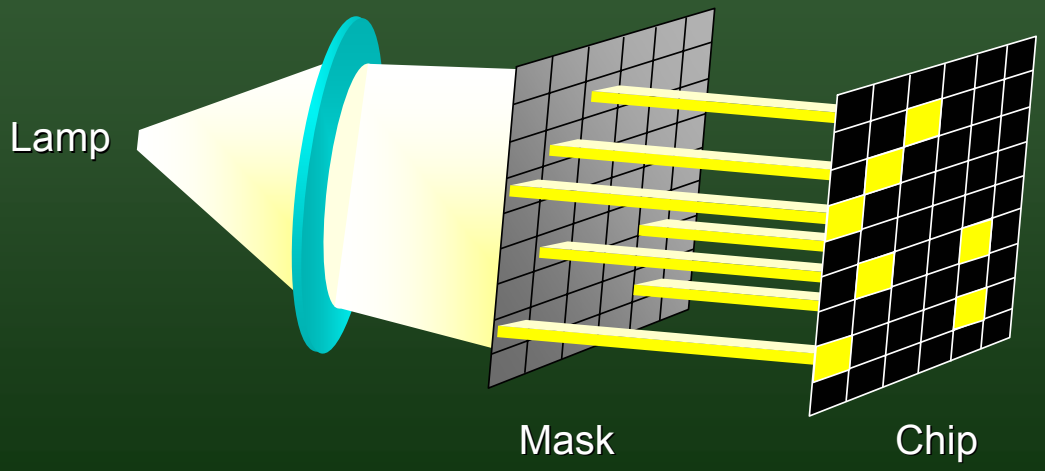
Steve Clough, USDA-ARS
University of Illinois, Urbana

GeneChip® Probe Arrays



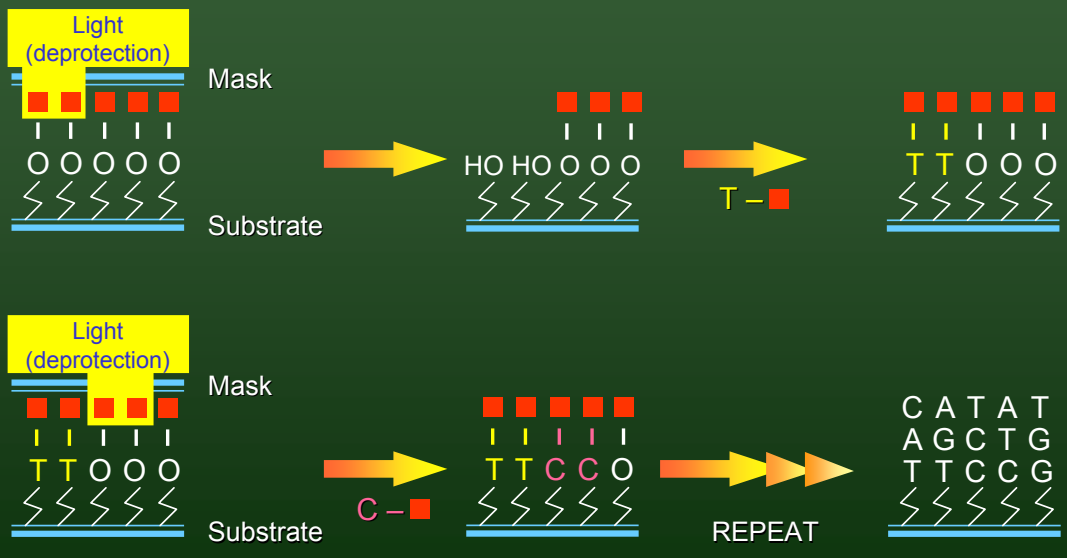
Courtesy of:
Mike Lelivelt

Photolithographic Synthesis



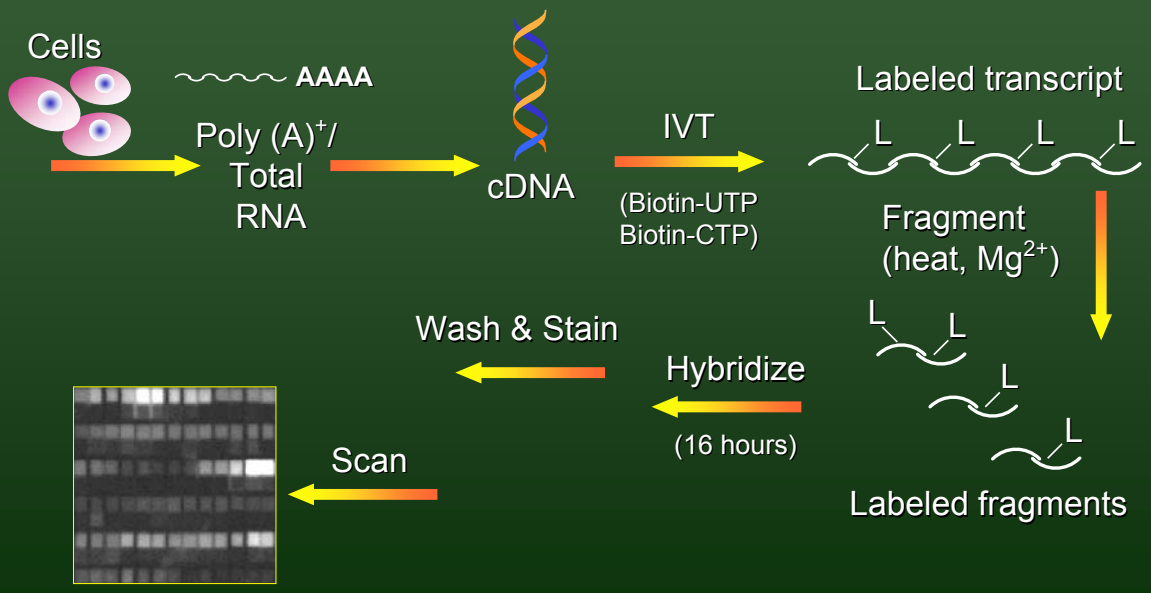
Courtesy of:
Mike Lelivelt

Synthesis of Ordered Oligonucleotide Arrays



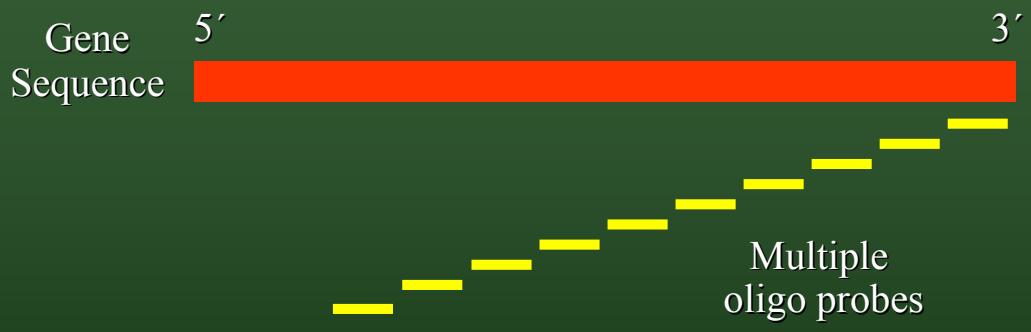
Courtesy of:
Mike Lelivelt

Procedures for Target Preparation



Courtesy of:
Mike Lelivelt

GeneChip® Expression Array Design

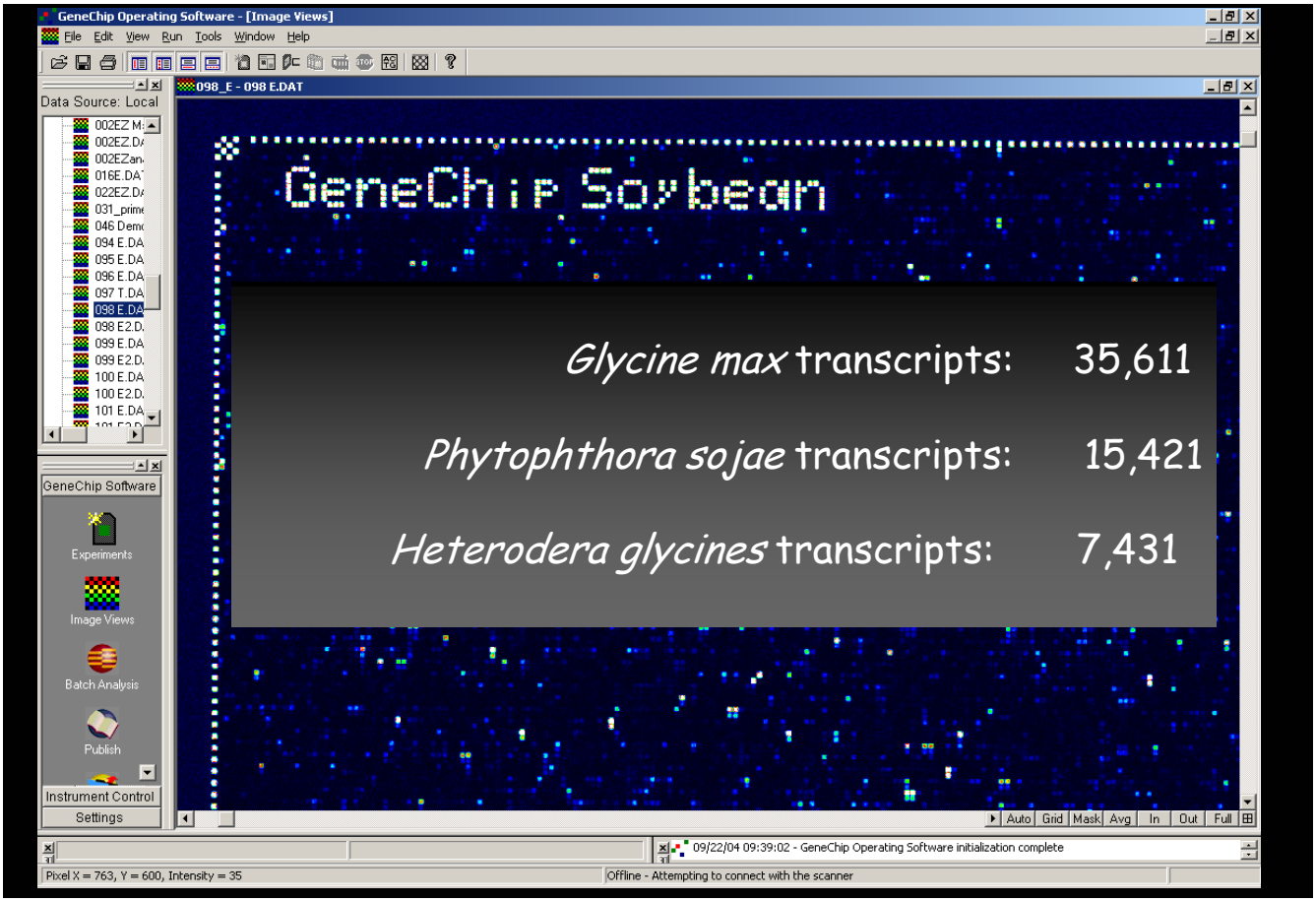


Probes designed to be **Perfect Match**

Probes designed to be **Mismatch**

(On an actual chip, these tiles are randomly dispersed)

Courtesy of:
Mike Lelivelt



Pros and Cons of cDNA platforms

- cDNA: need to construct cDNA libraries from a variety of tissues and conditions and to sequence to verify lack of duplication.
- Cheapest approach. Do not need to have a sequenced genome
- Hybridization involves strands of hundreds of bases, therefore less specificity in binding and cannot differentiate multigene family members. Good if your organism is closely related but not identical to one used to make the cDNA libraries used to make arrays.

Steve Clough, USDA-ARS
University of Illinois, Urbana

Pros and Cons of Oligo-based platforms

- Oligo: spot collections of oligos, usually 50-70 bp long that span known/predicted ORFs. Affymetrix chips use 25mers and 11 or so probes per ORF
- Need lots of sequence information from your organism
- Works best if your organism is same or very closely related to the one used to obtain the sequence information
- More costly than cDNA arrays to manufacture
- Only choice if working with a prokaryote (no polyA tails)

Steve Clough, USDA-ARS
University of Illinois, Urbana