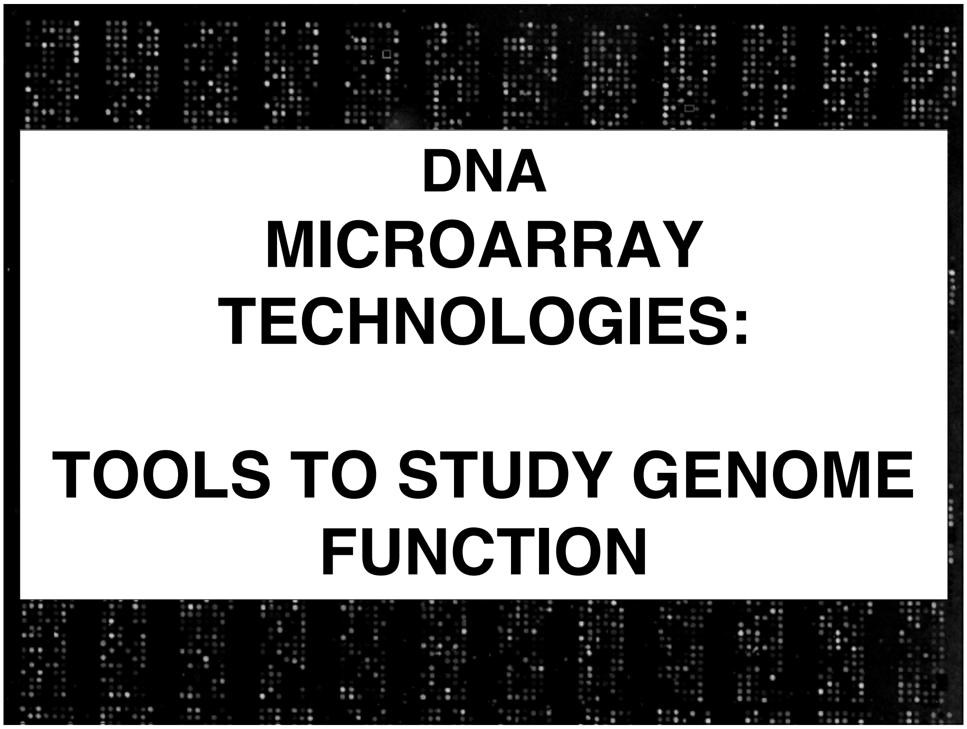
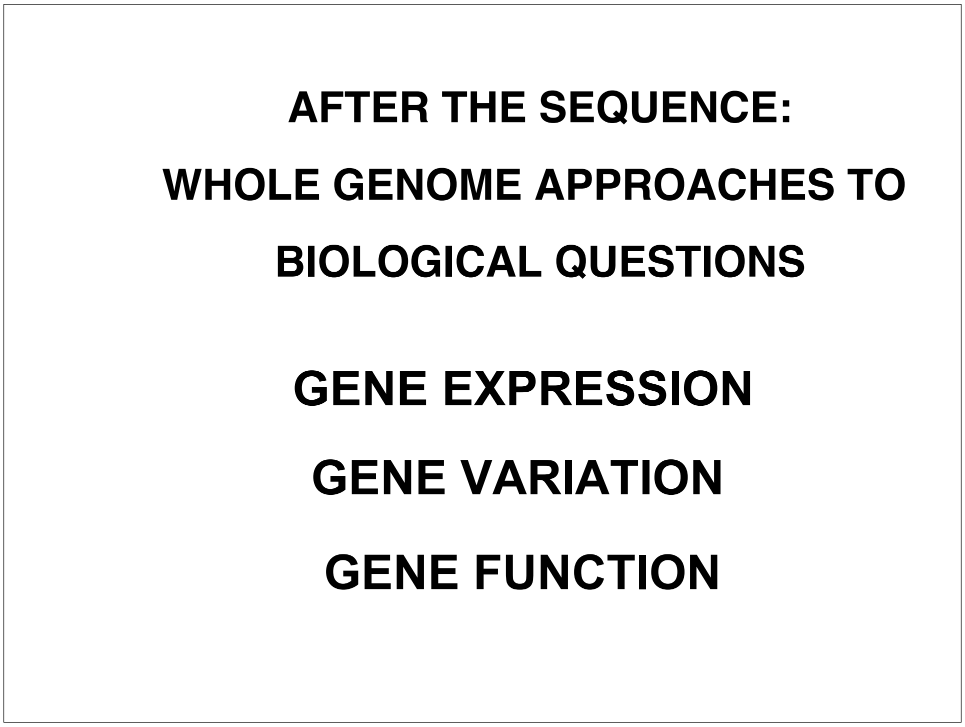


DNA MICROARRAY TECHNOLOGIES:

TOOLS TO STUDY GENOME FUNCTION



AFTER THE SEQUENCE:

WHOLE GENOME APPROACHES TO BIOLOGICAL QUESTIONS

GENE EXPRESSION

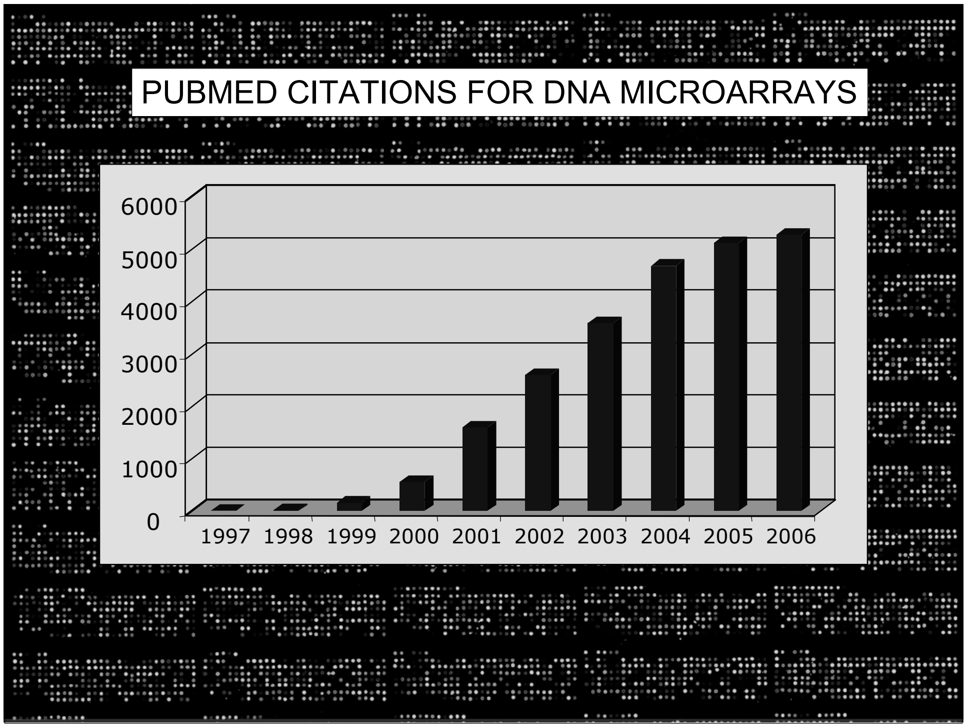
GENE VARIATION

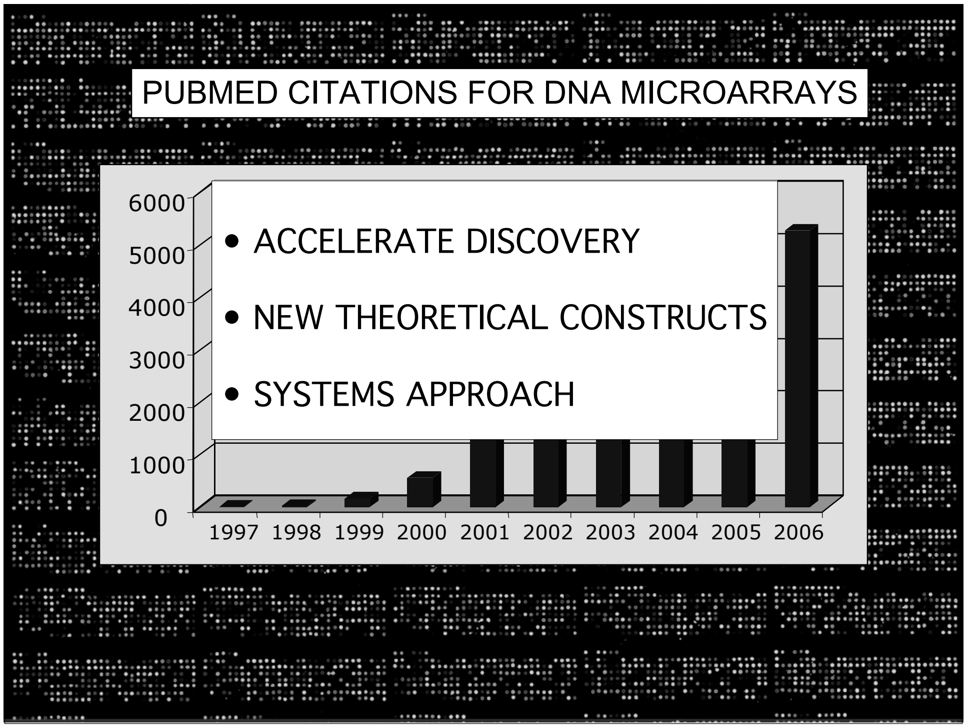
GENE FUNCTION

MICROARRAYS PROVIDE A TOOL FOR WHOLE GENOME ANALYSIS

PRIMARY IMPACT:

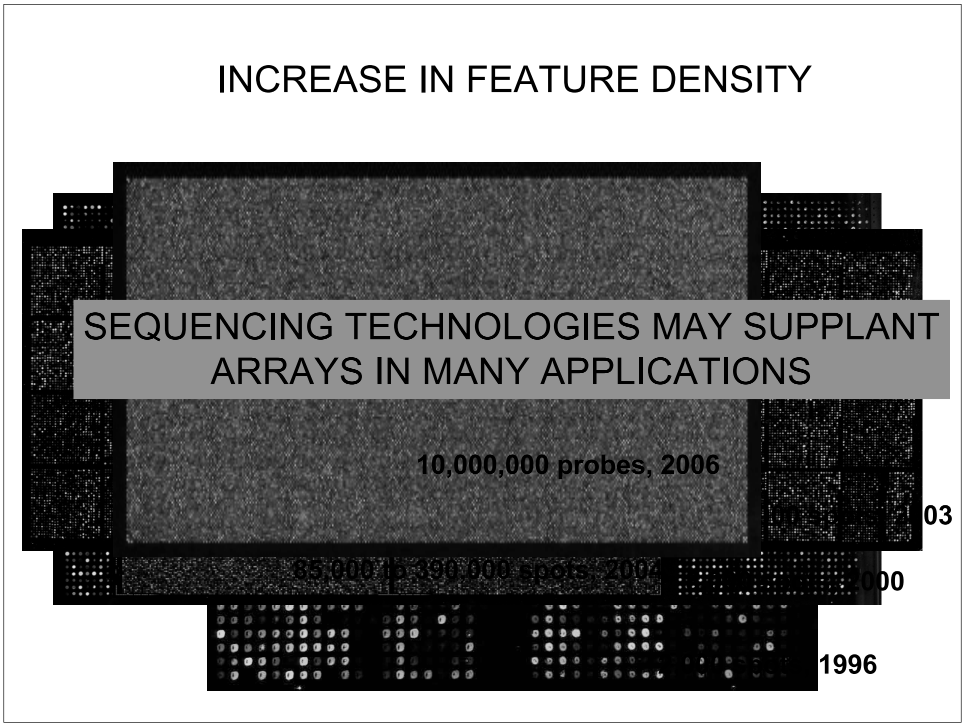
ACCELERATED DISCOVERY AND HYPOTHESIS GENERATION


PUBMED CITATIONS FOR DNA MICROARRAYS



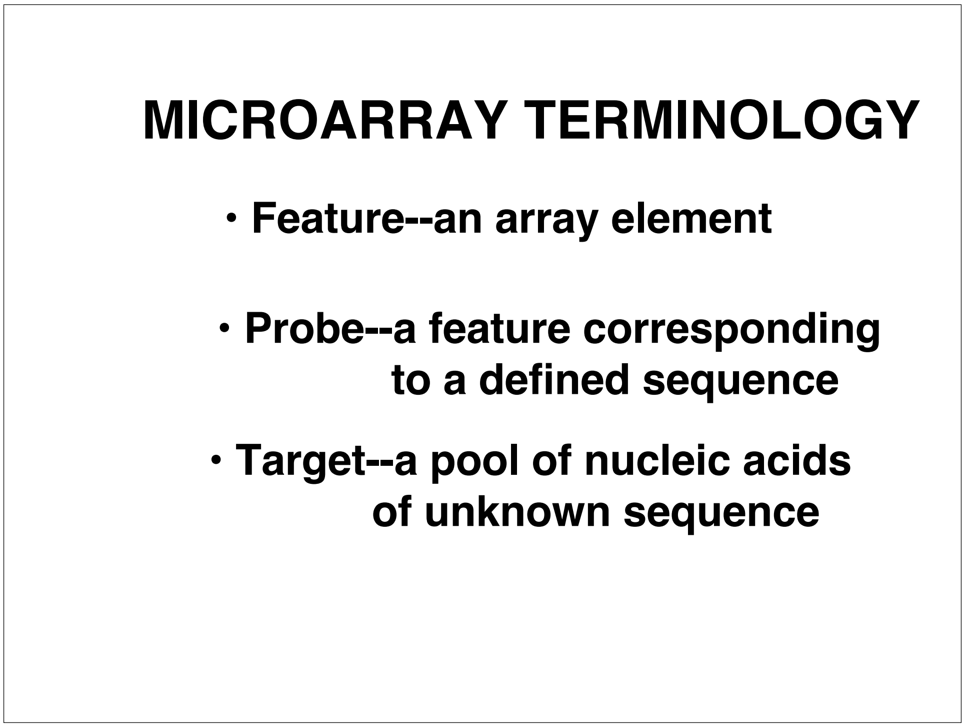
PUBMED CITATIONS FOR DNA MICROARRAYS

- ACCELERATE DISCOVERY
- NEW THEORETICAL CONSTRUCTS
- SYSTEMS APPROACH



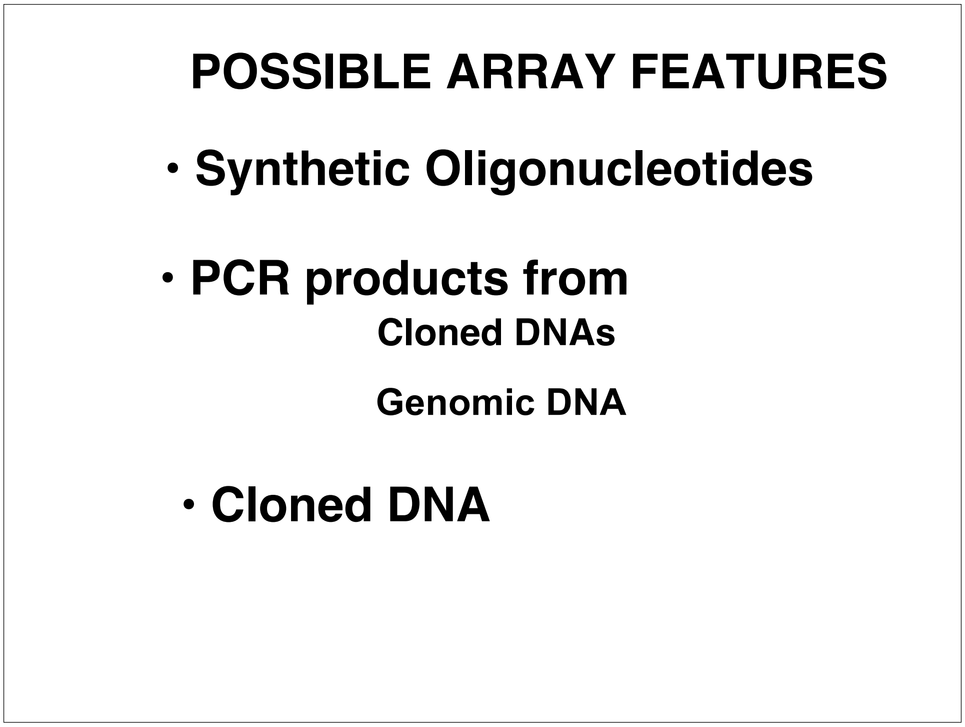
INCREASE IN FEATURE DENSITY

SEQUENCING TECHNOLOGIES MAY SUPPLANT ARRAYS IN MANY APPLICATIONS



MICROARRAY TERMINOLOGY

- Feature--an array element
- Probe--a feature corresponding to a defined sequence
- Target--a pool of nucleic acids of unknown sequence

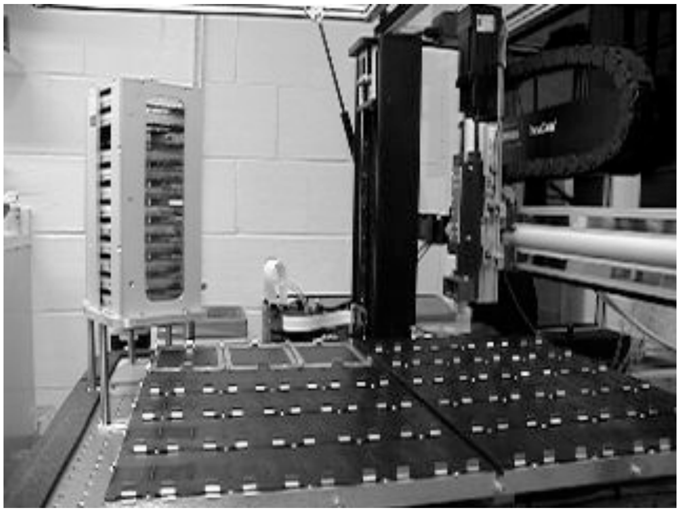


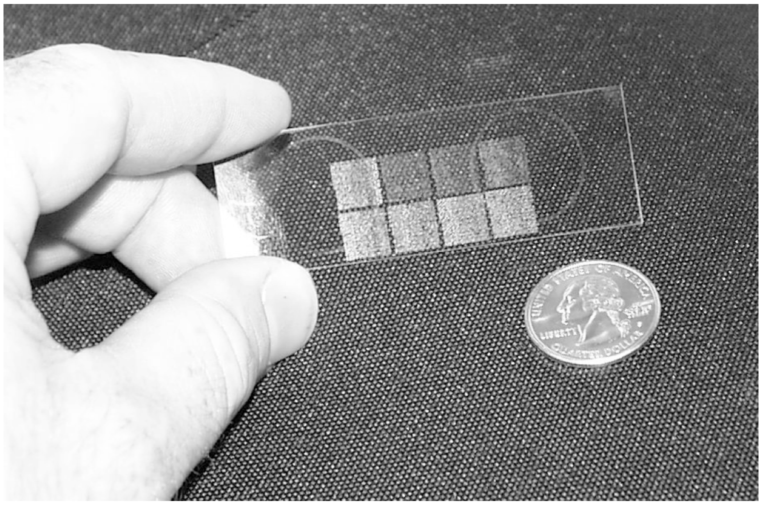
POSSIBLE ARRAY FEATURES

- Synthetic Oligonucleotides
- PCR products from
  - Cloned DNAs
  - Genomic DNA
- Cloned DNA

Microarray Manufacture

- **Printing**



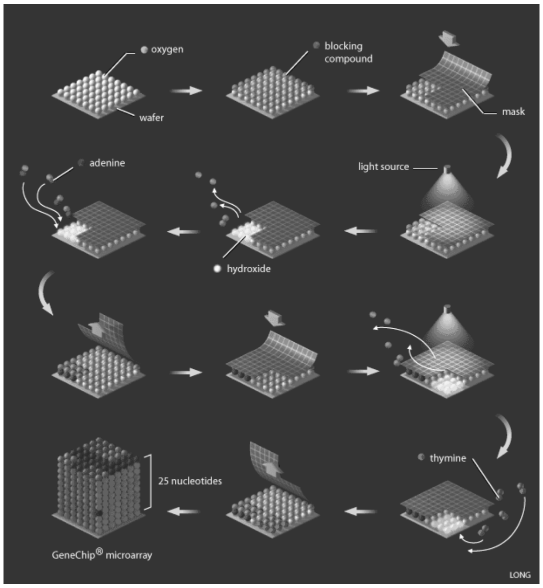


Microarray Manufacture

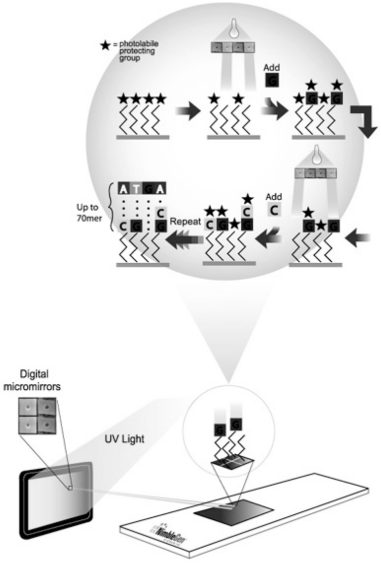
- Printing
- Synthesis *in situ*

  light directed
  mechanically directed

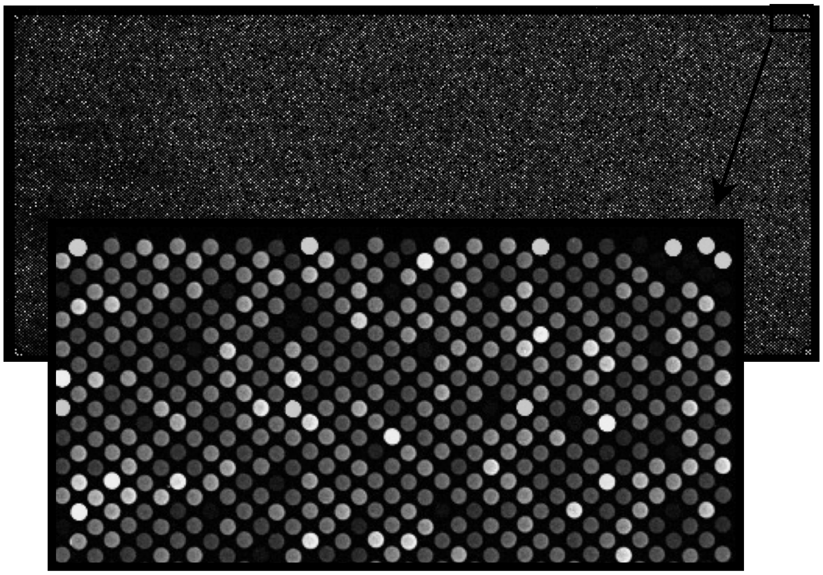
LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

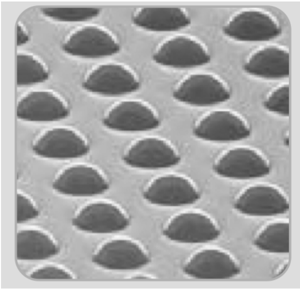


LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

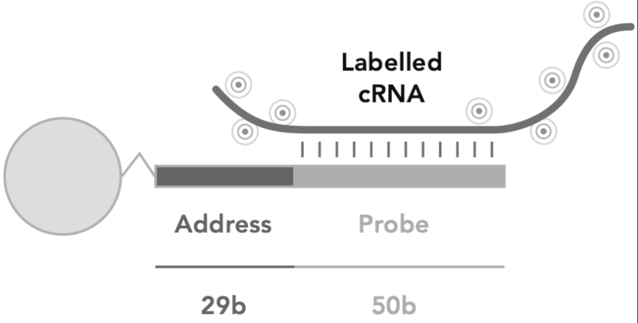


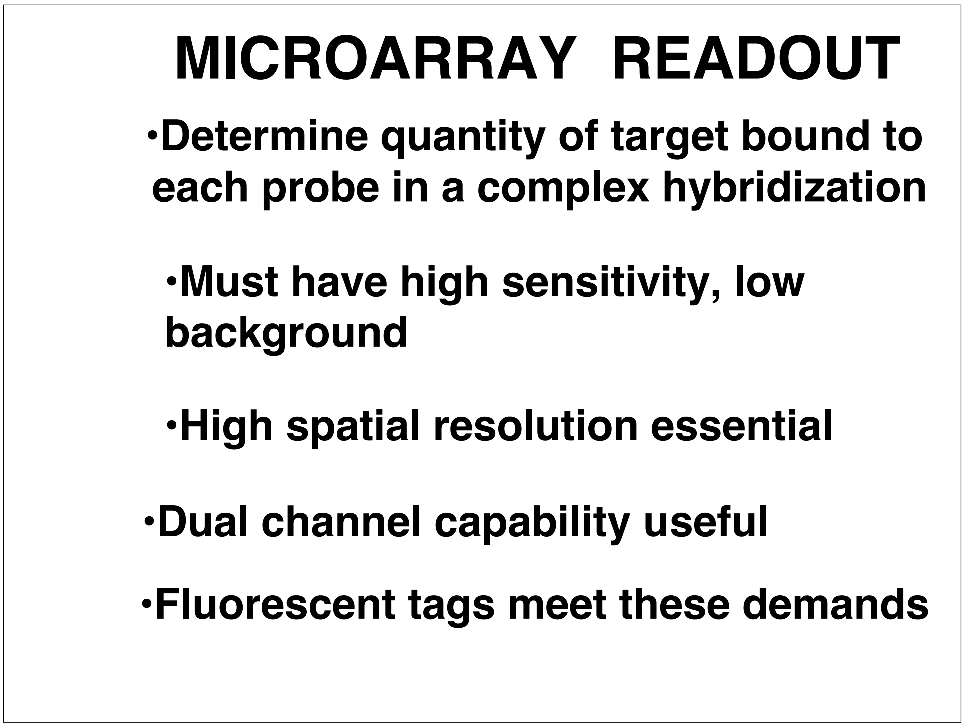
INK JET DIRECTED SYNTHESIS





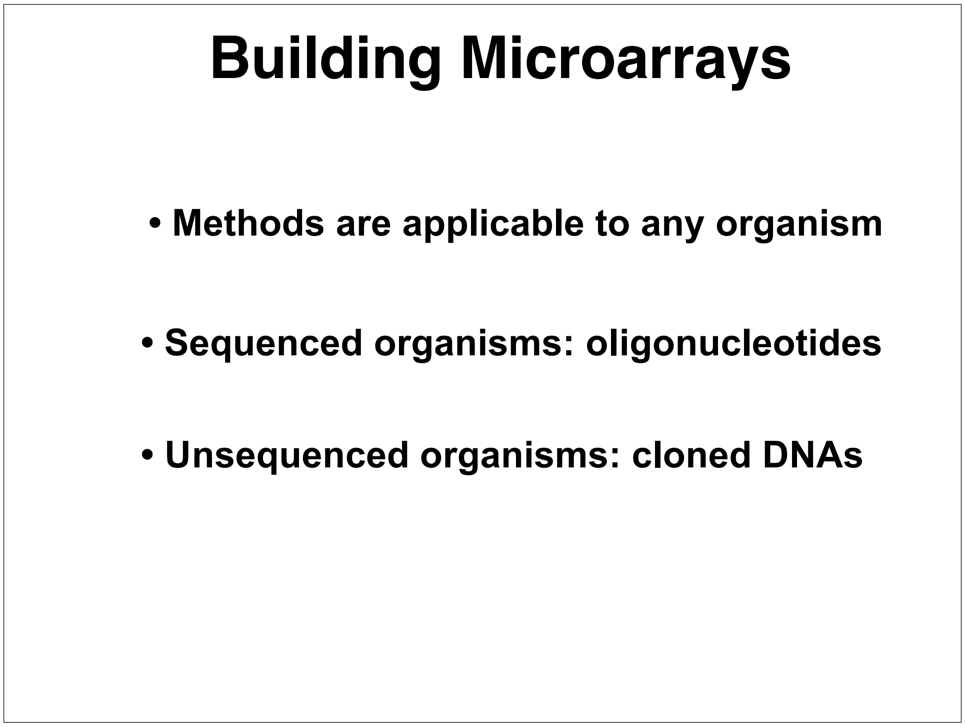
RANDOMLY POSITIONED HIGH DENSITY
ARRAYS OF ADDRESSABLE OLIGONUCLEOTIDES
COUPLED TO BEADS





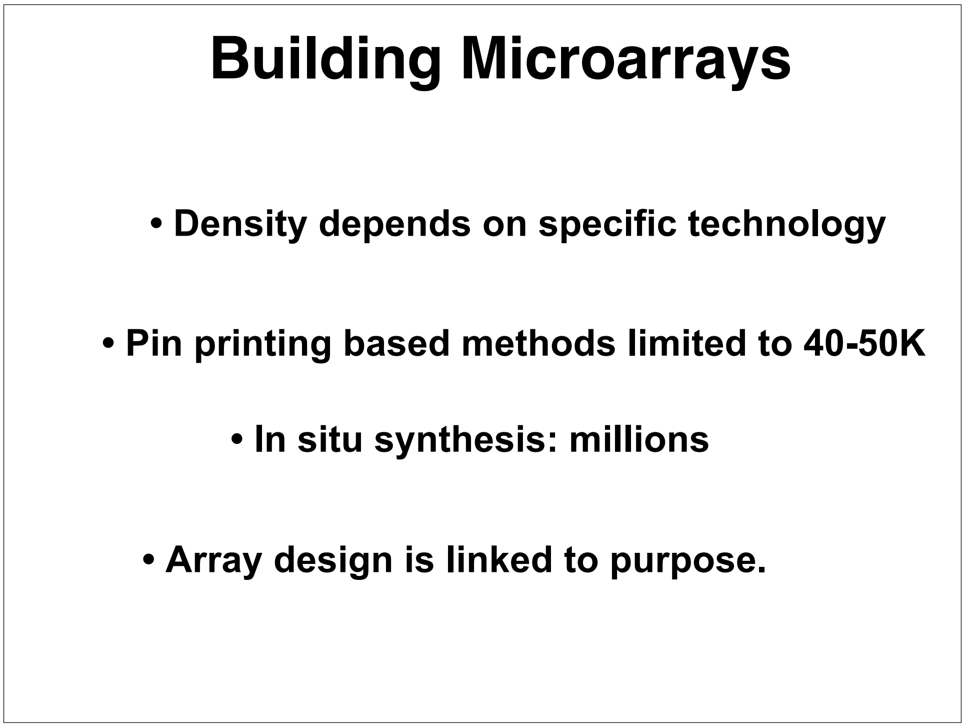
MICROARRAY READOUT

- Determine quantity of target bound to each probe in a complex hybridization
  - Must have high sensitivity, low background
  - High spatial resolution essential
- Dual channel capability useful
- Fluorescent tags meet these demands



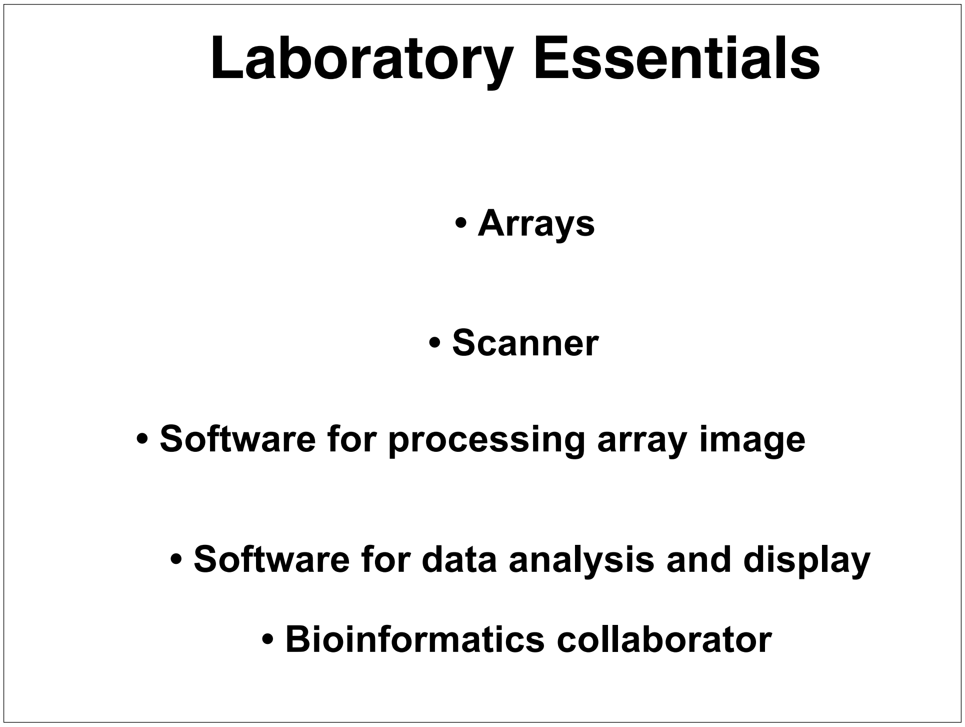
Building Microarrays

- Methods are applicable to any organism
- Sequenced organisms: oligonucleotides
- Unsequenced organisms: cloned DNAs



Building Microarrays

- Density depends on specific technology
- Pin printing based methods limited to 40-50K
- In situ synthesis: millions
- Array design is linked to purpose.



Laboratory Essentials

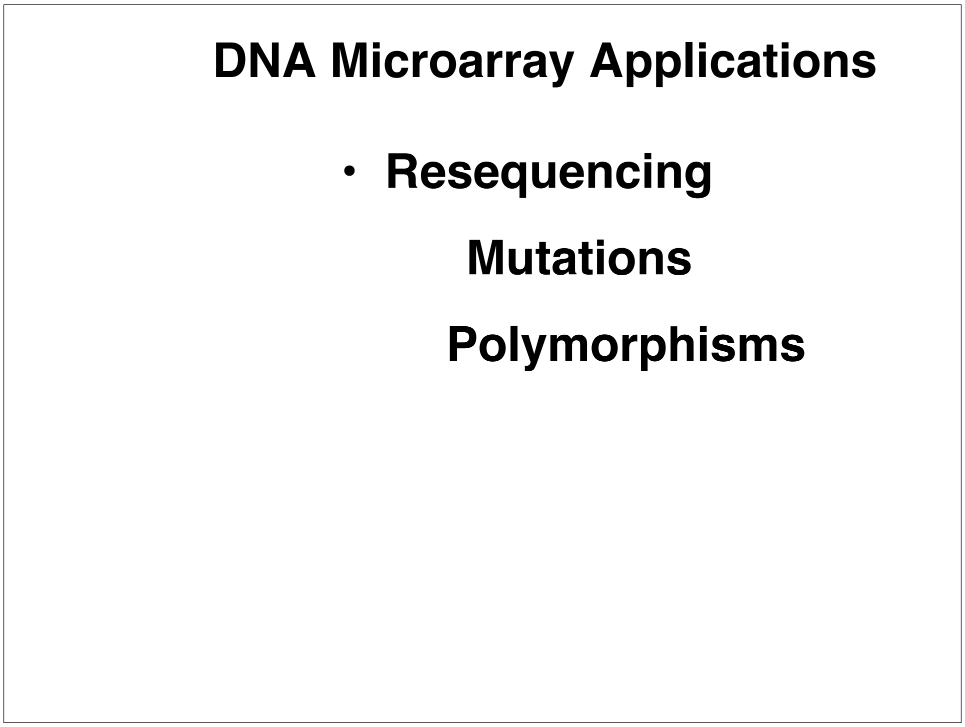
- Arrays
- Scanner
- Software for processing array image
- Software for data analysis and display
- Bioinformatics collaborator

DNA Microarray Applications

- Resequencing
- Comparative Genomic Hybridization
- Gene Expression
- Transcription factor localization
- Chromatin/DNA modification

DNA Microarray Applications

- Resequencing
- Comparative Genomic Hybridization
- Gene Expression
- Transcription factor localization
- Chromatin/DNA modification

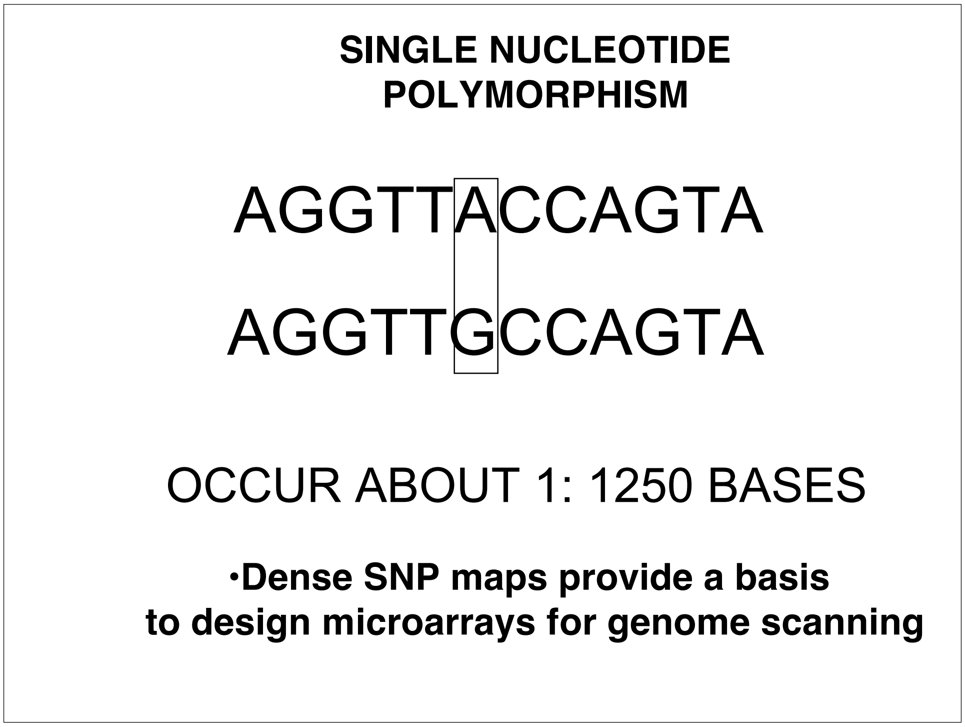


DNA Microarray Applications

- **Resequencing**

Mutations

Polymorphisms



SINGLE NUCLEOTIDE POLYMORPHISM

AGGTTACCAGTA

AGGTTGCCAGTA

OCCUR ABOUT 1: 1250 BASES

•Dense SNP maps provide a basis to design microarrays for genome scanning

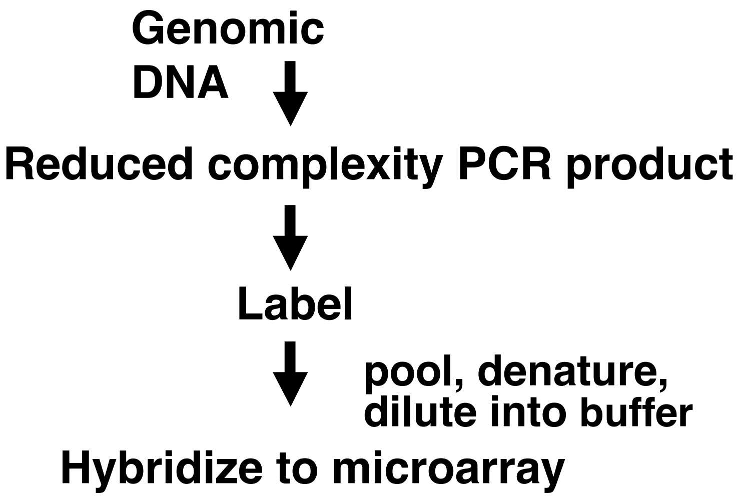
DNA Microarray Applications

- SNP detection

Differential hybridization

Extension/ligation strategies

LABELLING SNPs



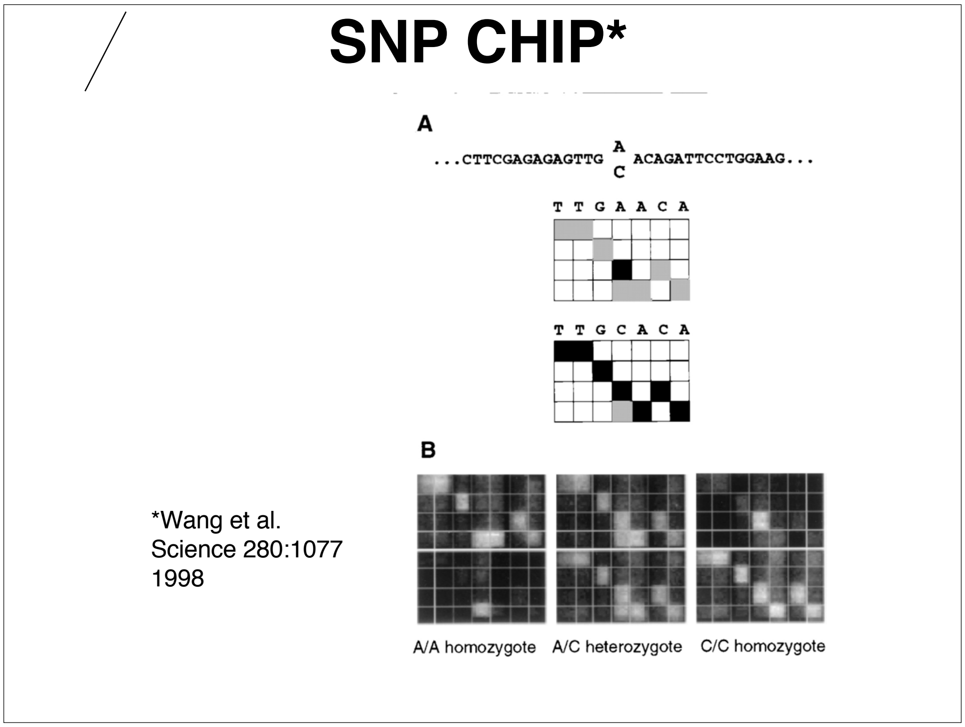
Genomic DNA

Reduced complexity PCR product

Label

pool, denature, dilute into buffer

Hybridize to microarray



SNP CHIP*

*Wang et al.
Science 280:1077
1998

SNP CHIPS

HAVE ACHIEVED HIGH DENSITY

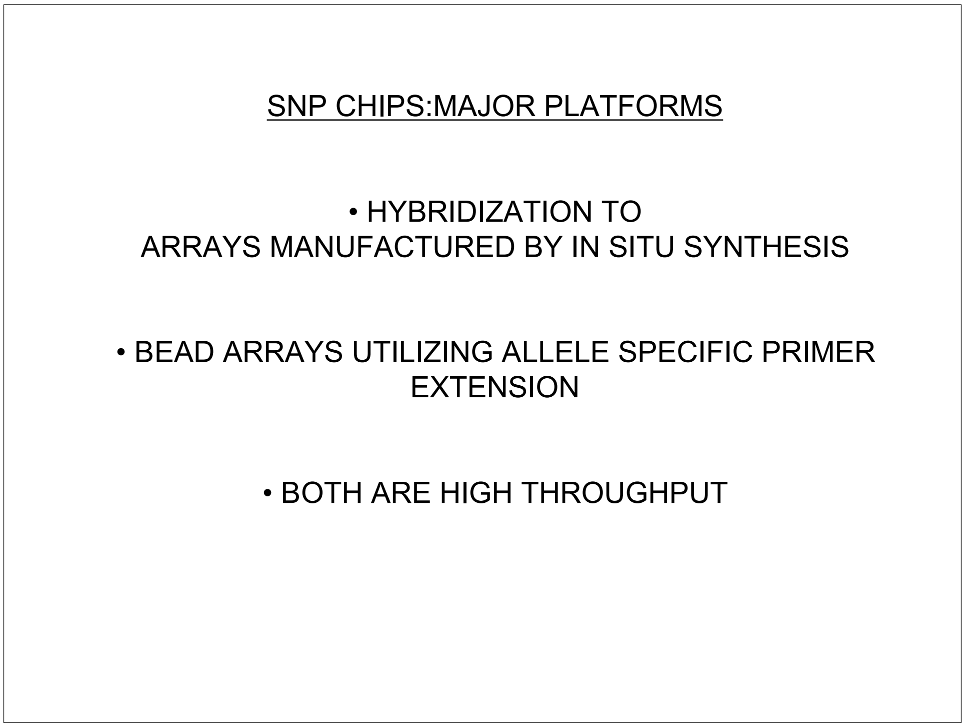
1,586,383 SNPS

HINDS ET AL. SCIENCE 307:1072 (2005)

COMMERCIAL CHIPS AVAILABLE: ≈1,000,000 SNPS

THIS WILL INCREASE

VIABLE OPTION FOR:
SNP GENOTYPING
CNV’S
CANCER ALLELIC IMBALANCE
AND COPY NUMBER.

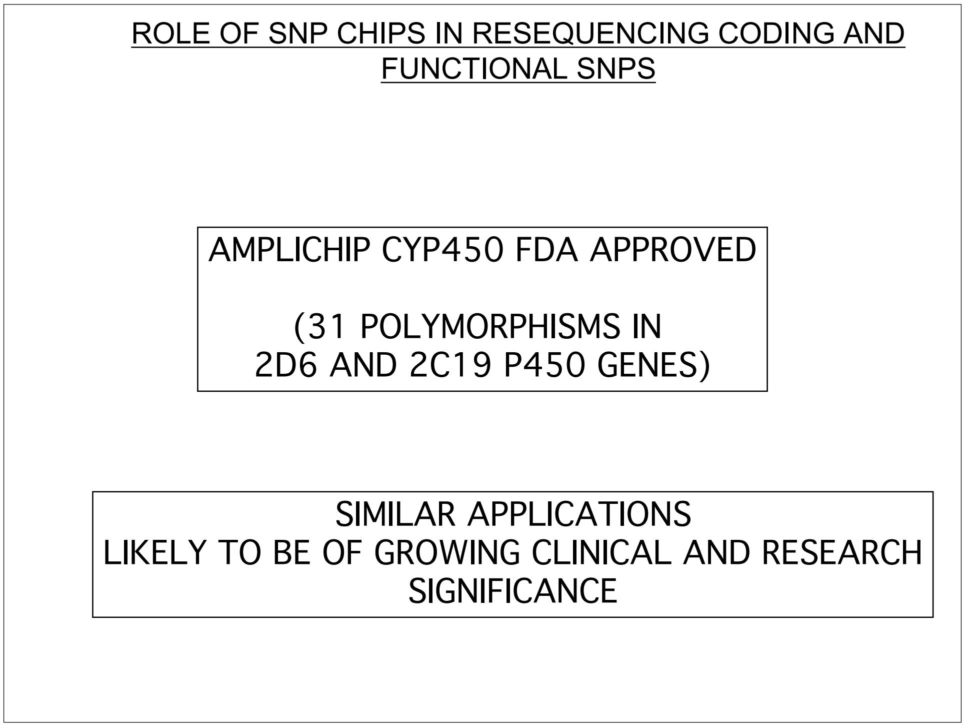


SNP CHIPS:MAJOR PLATFORMS

- HYBRIDIZATION TO ARRAYS MANUFACTURED BY IN SITU SYNTHESIS

- BEAD ARRAYS UTILIZING ALLELE SPECIFIC PRIMER EXTENSION

- BOTH ARE HIGH THROUGHPUT

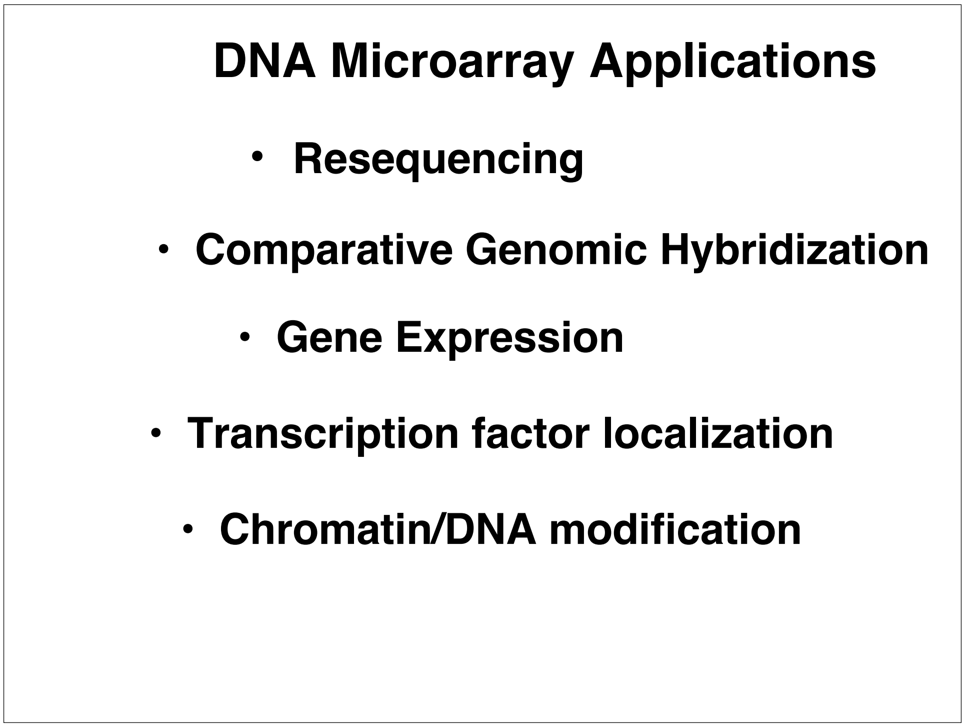


ROLE OF SNP CHIPS IN RESEQUENCING CODING AND FUNCTIONAL SNPS

AMPLICHIP CYP450 FDA APPROVED

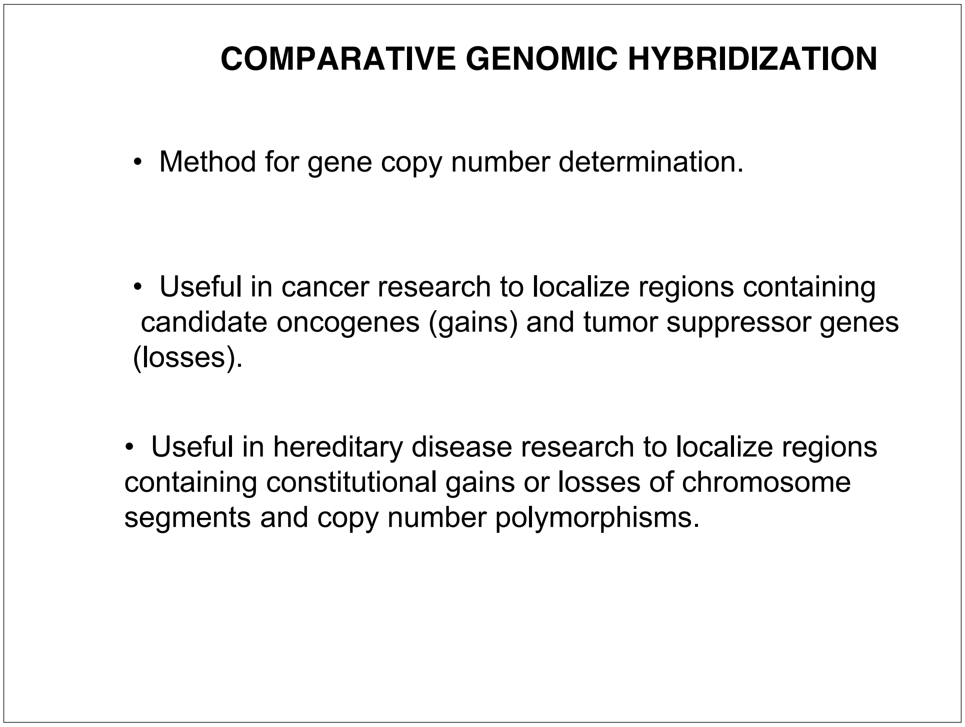
(31 POLYMORPHISMS IN
2D6 AND 2C19 P450 GENES)

SIMILAR APPLICATIONS
LIKELY TO BE OF GROWING CLINICAL AND RESEARCH
SIGNIFICANCE



DNA Microarray Applications

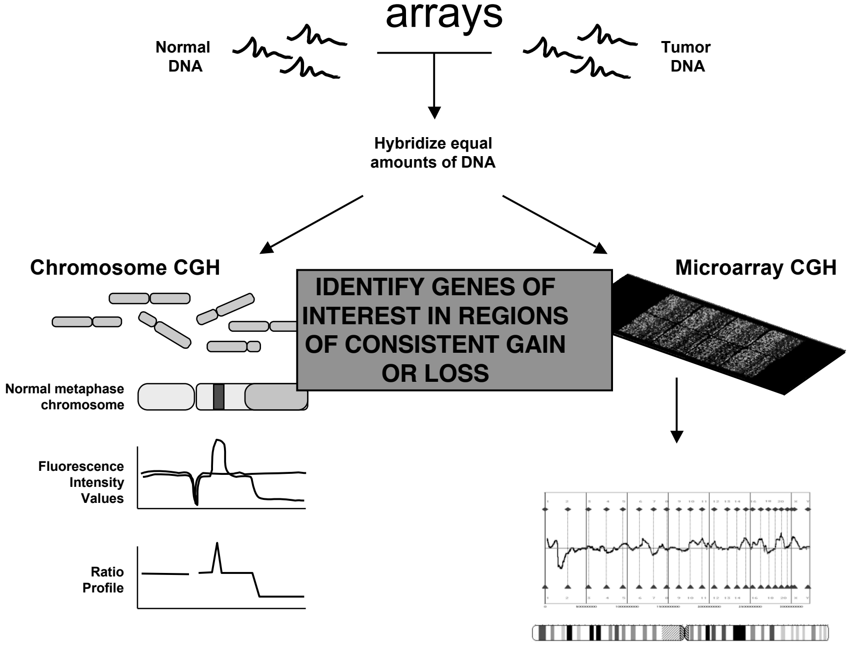
- Resequencing
- Comparative Genomic Hybridization
- Gene Expression
- Transcription factor localization
- Chromatin/DNA modification



COMPARATIVE GENOMIC HYBRIDIZATION

- Method for gene copy number determination.
- Useful in cancer research to localize regions containing candidate oncogenes (gains) and tumor suppressor genes (losses).
- Useful in hereditary disease research to localize regions containing constitutional gains or losses of chromosome segments and copy number polymorphisms.

Comparative Genomic Hybridization (CGH) arrays



Normal DNA

Tumor DNA

Hybridize equal amounts of DNA

Chromosome CGH

Microarray CGH

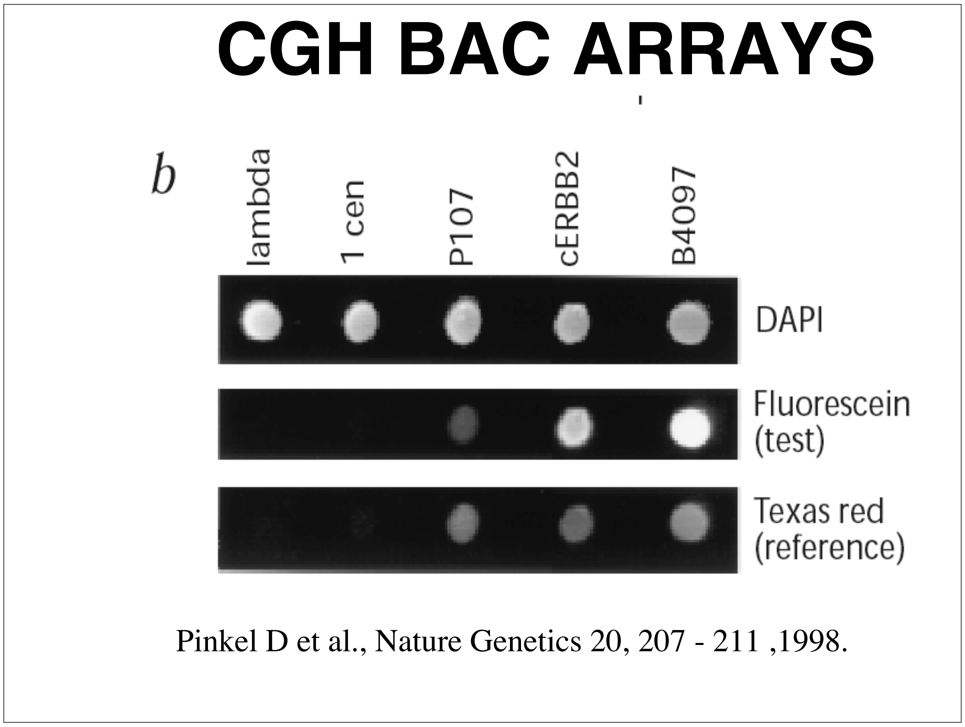
IDENTIFY GENES OF INTEREST IN REGIONS OF CONSISTENT GAIN OR LOSS

PLATFORMS FOR ARRAY BASED COMPARATIVE GENOMIC HYBRIDIZATION (CGH)

- BACs
- cDNAs
- Oligonucleotides

ARRAY CGH

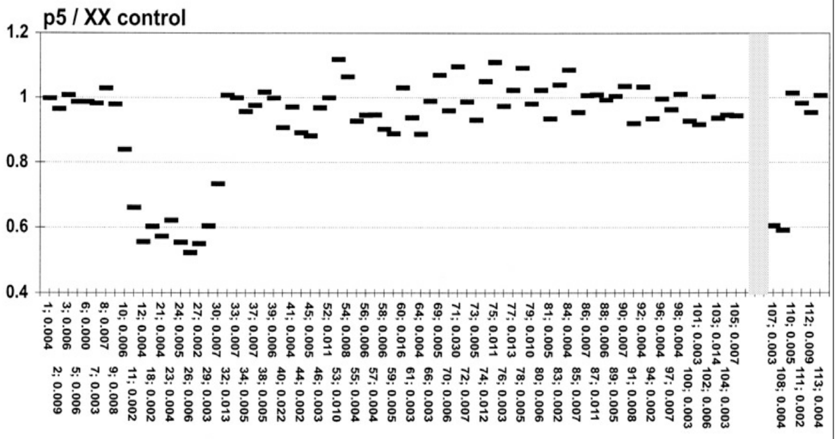
- HIGH RESOLUTION.
- SIMPLIFIED IMAGE ANALYSIS.
- HIGH THROUGHPUT.
- OLIGO STRATEGY ALLOWS GENOME BASED DESIGN.



CGH BAC ARRAYS

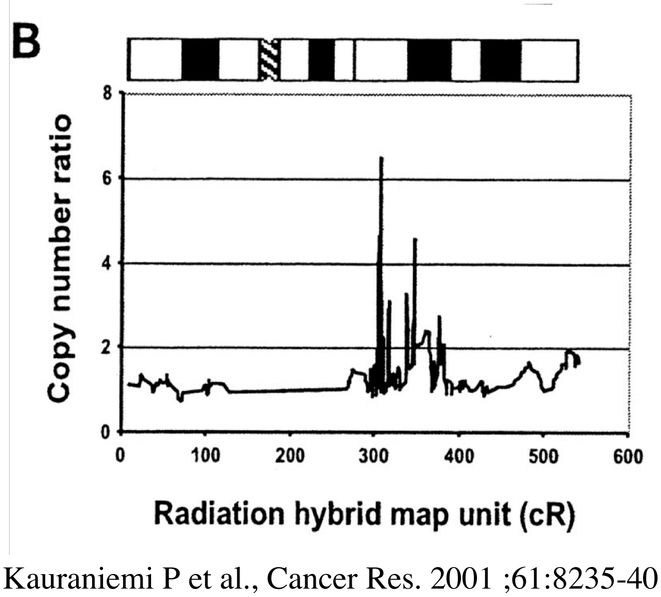
Pinkel D et al., Nature Genetics 20, 207 - 211 ,1998.

CGH BAC ARRAYS

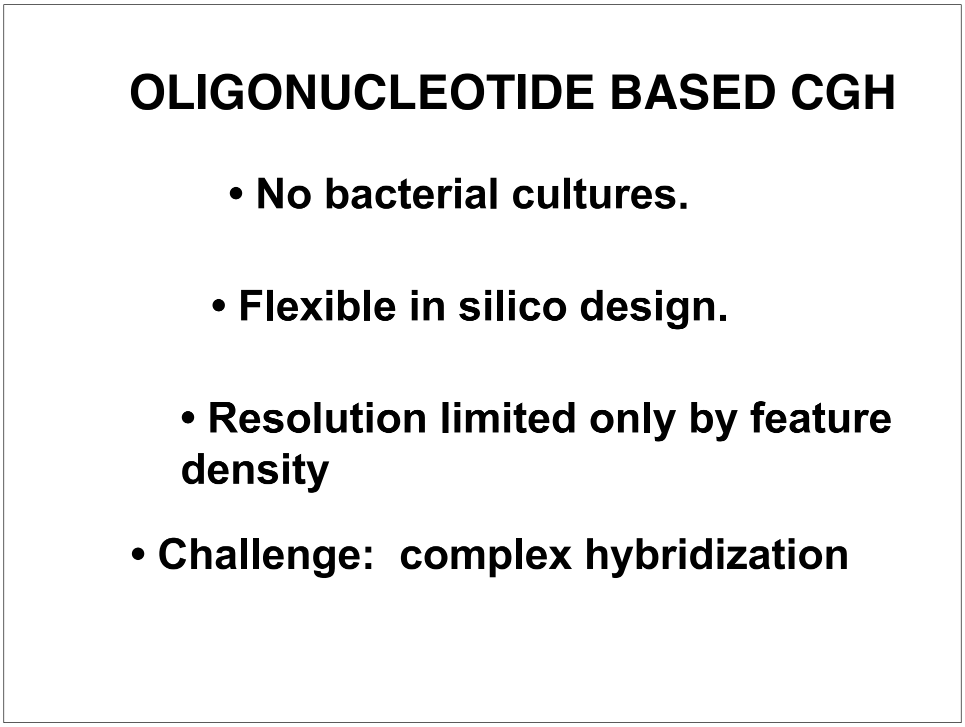


Bruder CE et al., Hum Mol Genet. 2001;10:271-82.

CGH cDNA

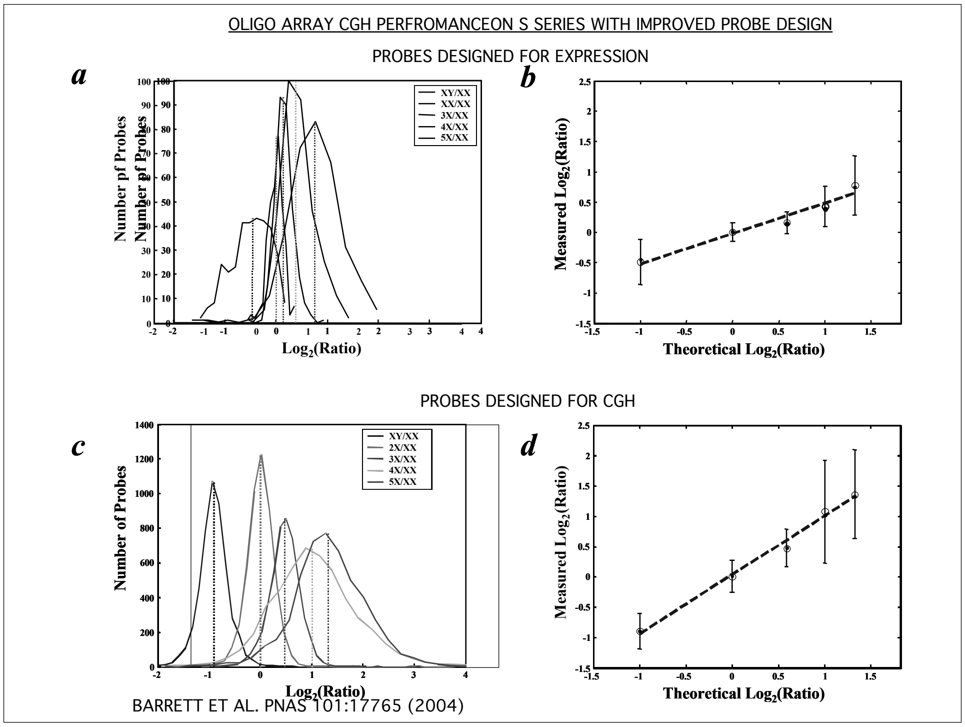


Kauraniemi P et al., Cancer Res. 2001 ;61:8235-40.



OLIGONUCLEOTIDE BASED CGH

- No bacterial cultures.
- Flexible in silico design.
- Resolution limited only by feature density
- Challenge: complex hybridization



OLIGO ARRAY CGH PERFROMANCEON S SERIES WITH IMPROVED PROBE DESIGN

PROBES DESIGNED FOR EXPRESSION

PROBES DESIGNED FOR CGH

BARRETT ET AL. PNAS 101:17765 (2004)

COMPARISON OF cDNA AND OLIGONUCLEOTIDE CGH

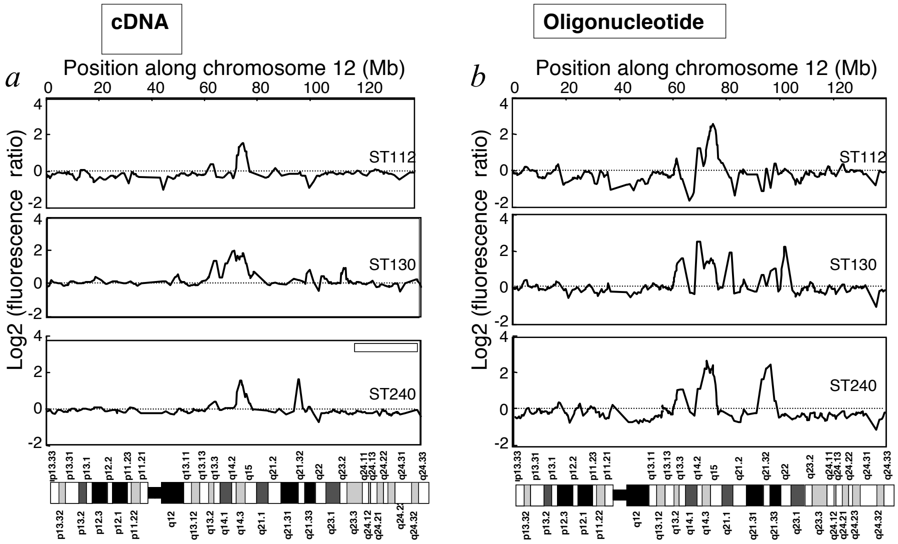
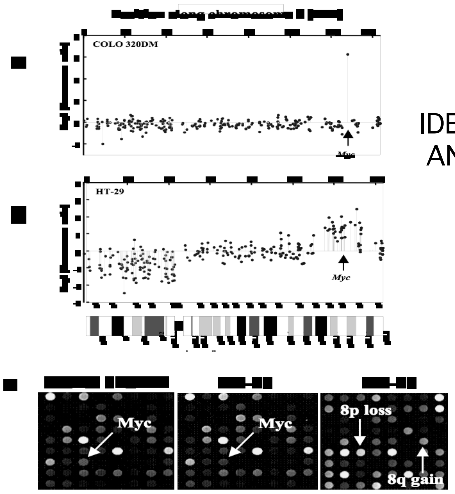
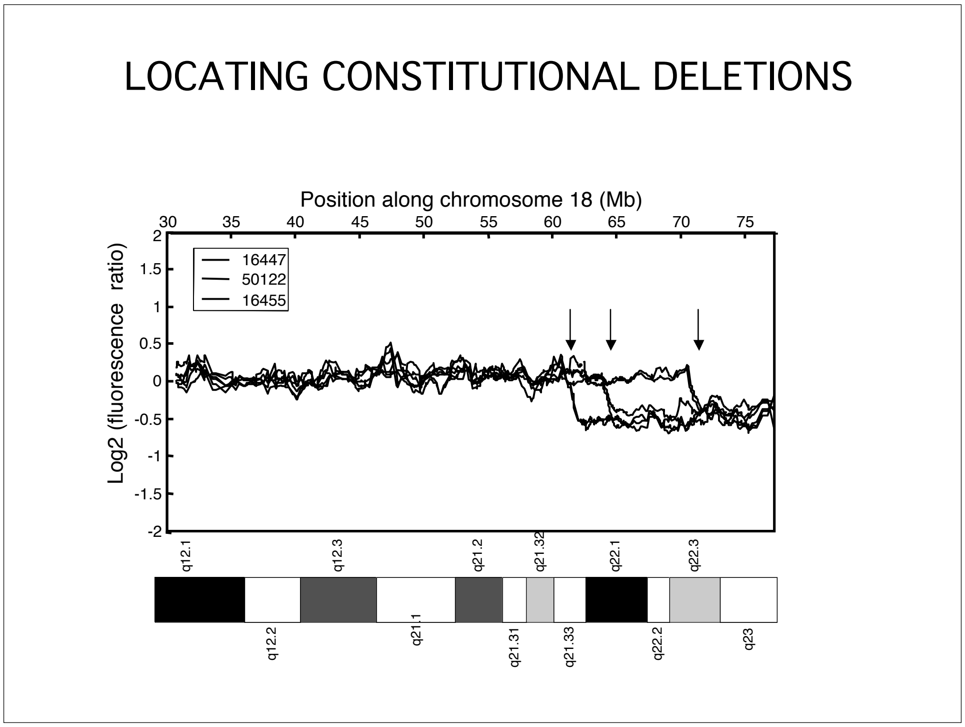


Figure 4

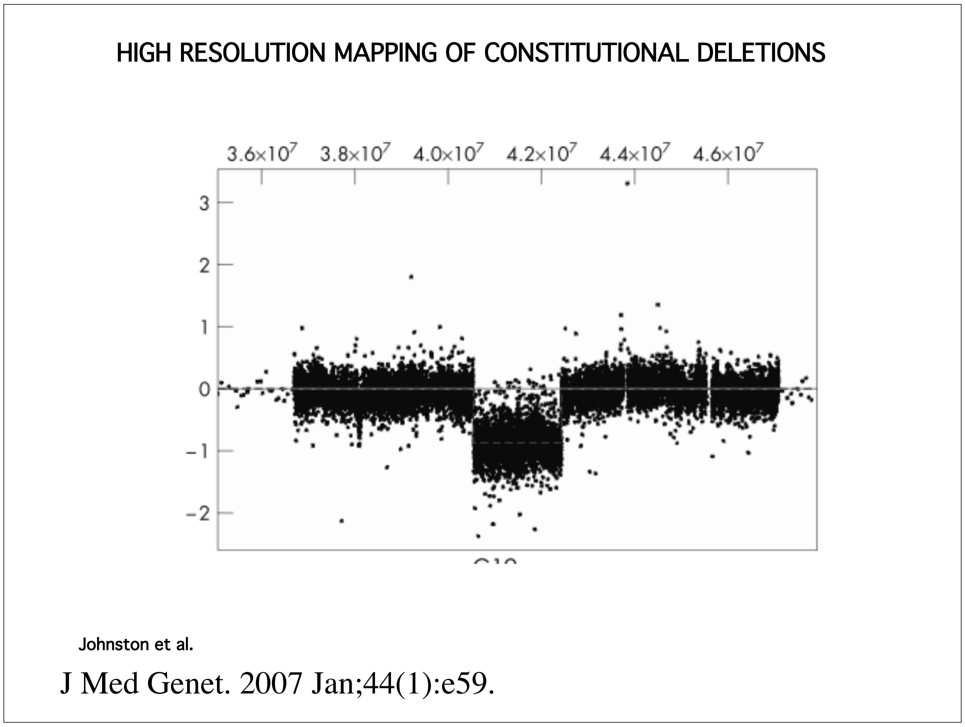


IDENTIFICATION OF GAINS AND LOSSES IN CANCER CELLS

[illegible]



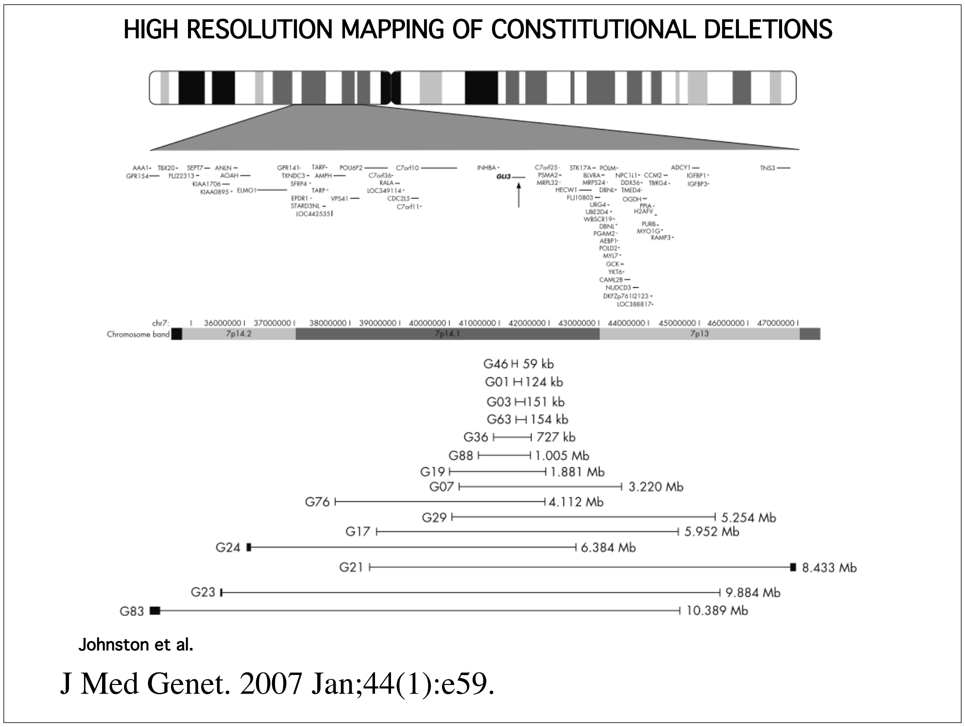
LOCATING CONSTITUTIONAL DELETIONS



HIGH RESOLUTION MAPPING OF CONSTITUTIONAL DELETIONS

Johnston et al.

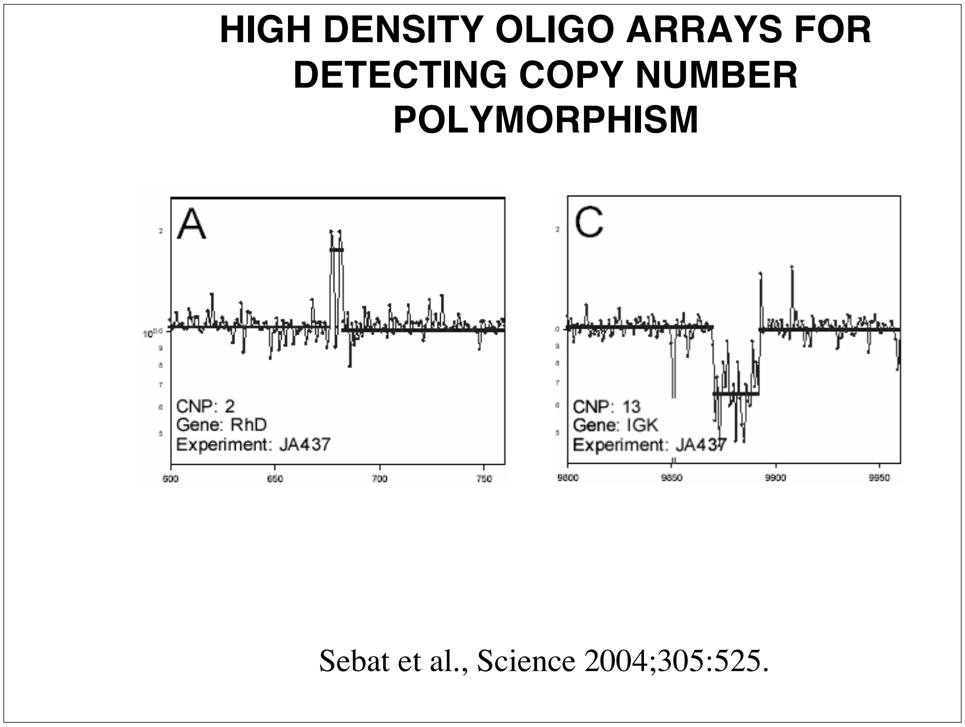
J Med Genet. 2007 Jan;44(1):e59.



HIGH RESOLUTION MAPPING OF CONSTITUTIONAL DELETIONS

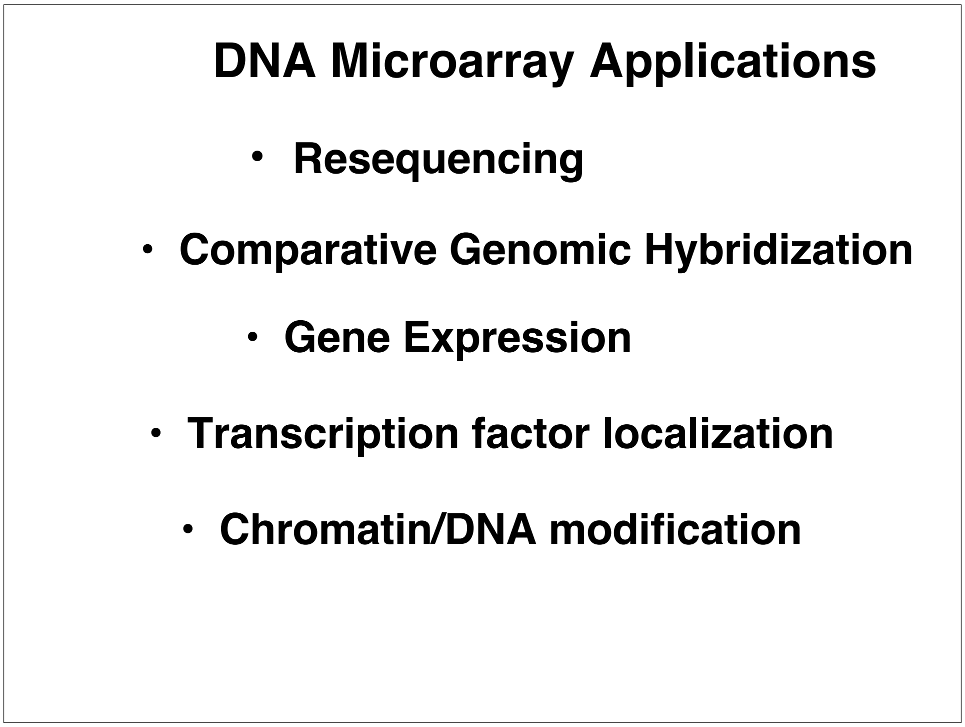
Johnston et al.

J Med Genet. 2007 Jan;44(1):e59.



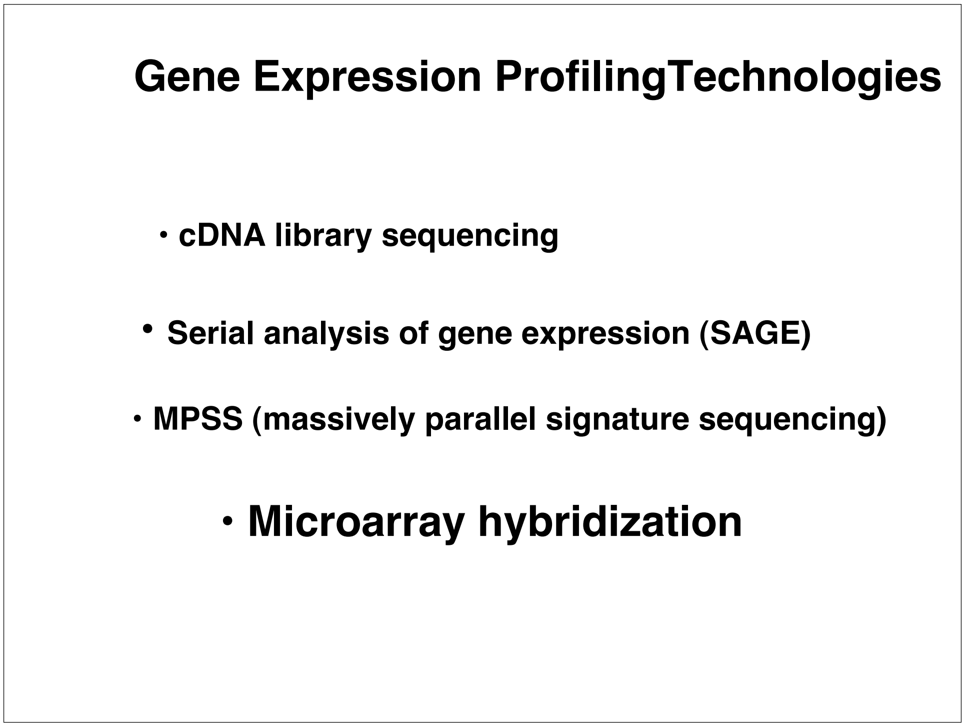
HIGH DENSITY OLIGO ARRAYS FOR DETECTING COPY NUMBER POLYMORPHISM

Sebat et al., Science 2004;305:525.



DNA Microarray Applications

- Resequencing
- Comparative Genomic Hybridization
- Gene Expression
- Transcription factor localization
- Chromatin/DNA modification



Gene Expression ProfilingTechnologies

- cDNA library sequencing
- Serial analysis of gene expression (SAGE)
- MPSS (massively parallel signature sequencing)
- Microarray hybridization

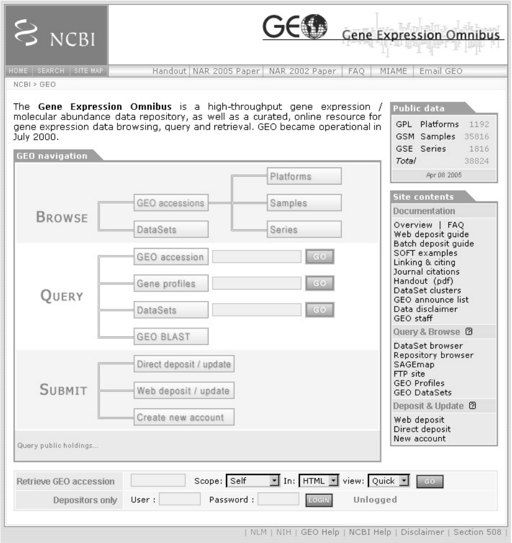
Reports on Microarray Data Quality

Nature Biotechnology

September 2006

Accessing Expression Data

- Individual Lab and Journal Sites; public databases



The **Gene Expression Omnibus** is a high-throughput gene expression / molecular abundance data repository, as well as a curated, online resource for gene expression data browsing, query and retrieval. GEO became operational in July 2000.

Public data		
GPL	Platforms	1192
GSM	Samples	35816
GSE	Series	1816
Total		38824

Apr 08 2005

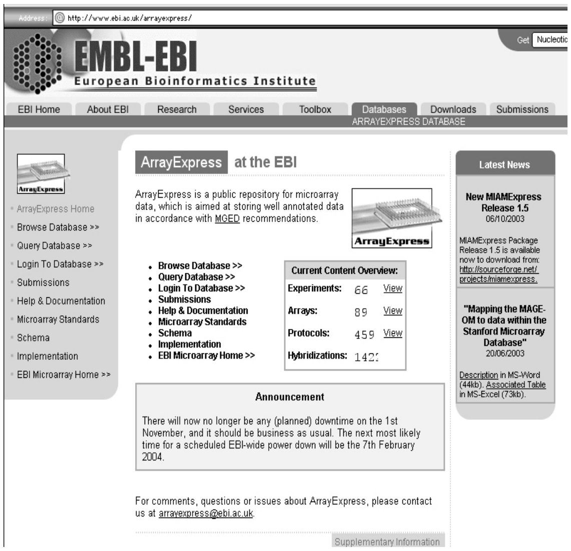
Retrieve GEO accession Scope: Self In: HTML view: Quick GO

Depositors only User : Password : Unlogged

GEO

http://www.ncbi.nlm.nih.gov/geo/

Accessing Expression Data



ArrayExpress at the EBI

ArrayExpress is a public repository for microarray data, which is aimed at storing well annotated data in accordance with MGED recommendations.

- Browse Database >>
- Query Database >>
- Login To Database >>
- Submissions
- Help & Documentation
- Microarray Standards
- Schema
- Implementation
- EBI Microarray Home >>

Current Content Overview:		
Experiments:	66	View
Arrays:	89	View
Protocols:	459	View
Hybridizations:	142:	

Announcement

There will now no longer be any (planned) downtime on the 1st November, and it should be business as usual. The next most likely time for a scheduled EBI-wide power down will be the 7th February 2004.

For comments, questions or issues about ArrayExpress, please contact us at arrayexpress@ebi.ac.uk.

Latest News

New MIAMExpress Release 1.5
06/10/2003

MIAMExpress Package Release 1.5 is available now to download from: http://sourceforge.net/projects/miamexpress.

"Mapping the MAGE-OM to data within the Stanford Microarray Database"
20/06/2003

Description in MS-Word (44kb). Associated Table in MS-Excel (73kb).

Publishing Expression Data

•MIAME standard

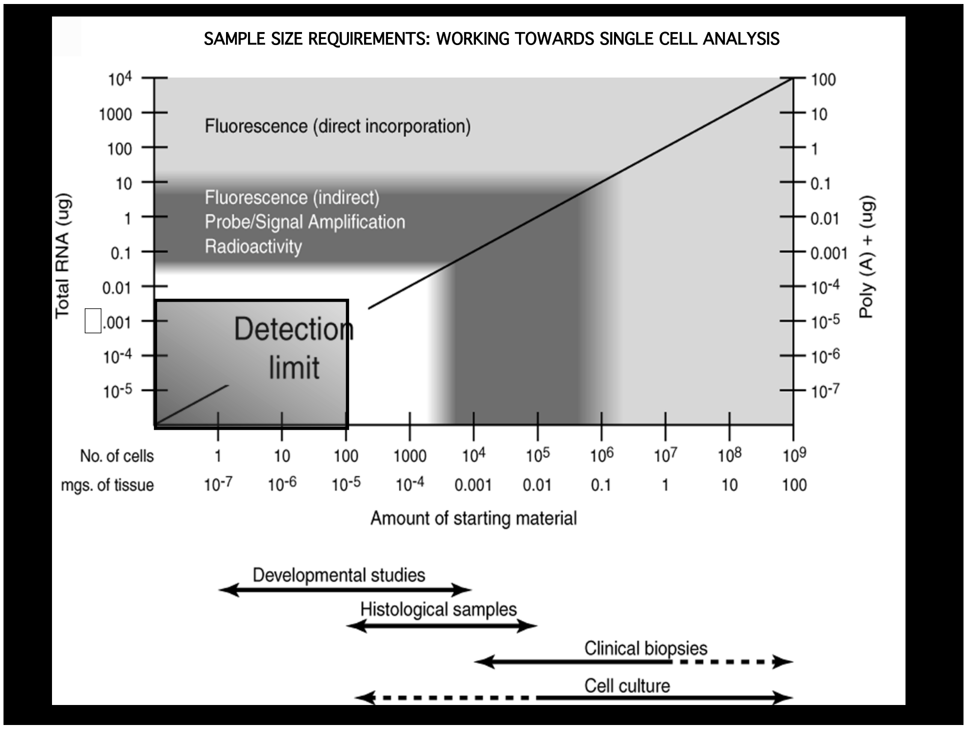
Minimum Information about a Microarray Experiment

- Format required by many journals
- Essential for database submissions

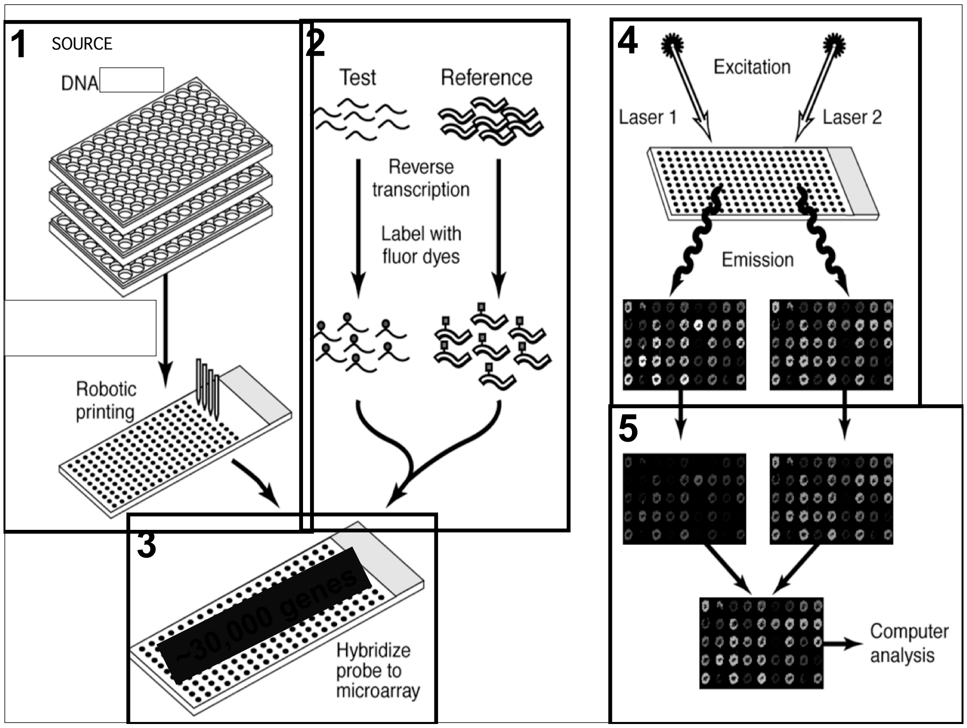
http://www.mged.org/Workgroups/MIAME/miame.html

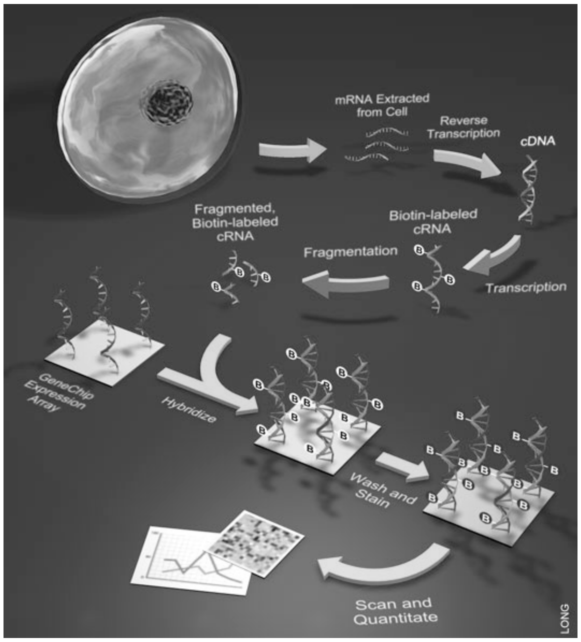
STRATEGIES FOR SIGNAL GENERATION FROM mRNA

- **Fluorochrome conjugated cDNA**
- **Ligand substituted nucleotides with secondary detection (e.g. biotin-streptavidin)**
- **Radioactivity**
- **RNA amplification**



SAMPLE SIZE REQUIREMENTS: WORKING TOWARDS SINGLE CELL ANALYSIS



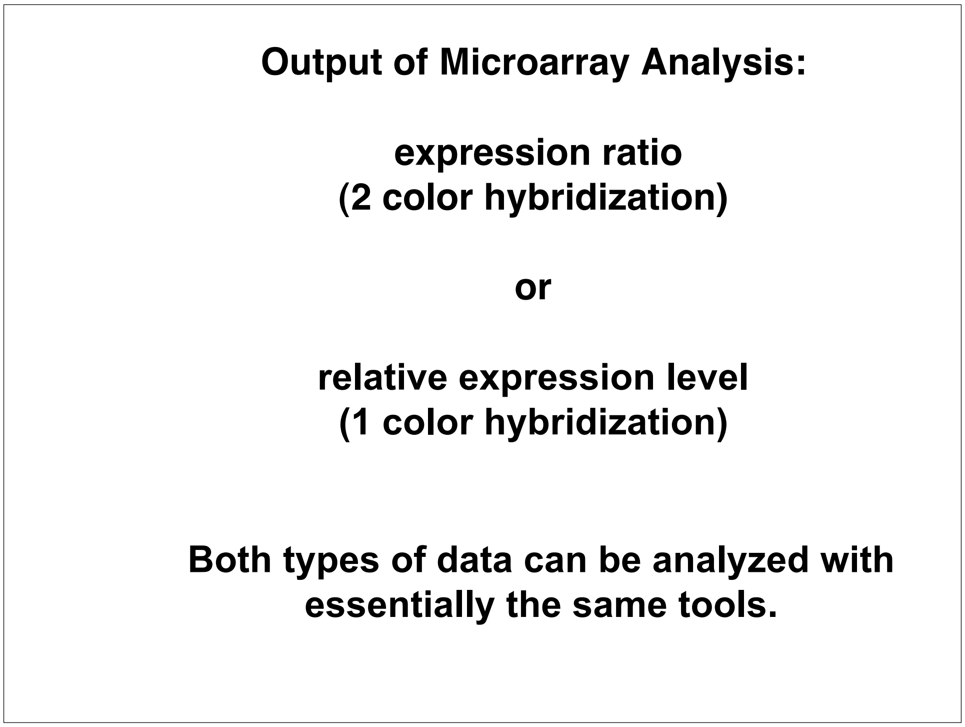


ONE COLOR

HYBRIDIZATION

ON AN OLIGO

ARRAY



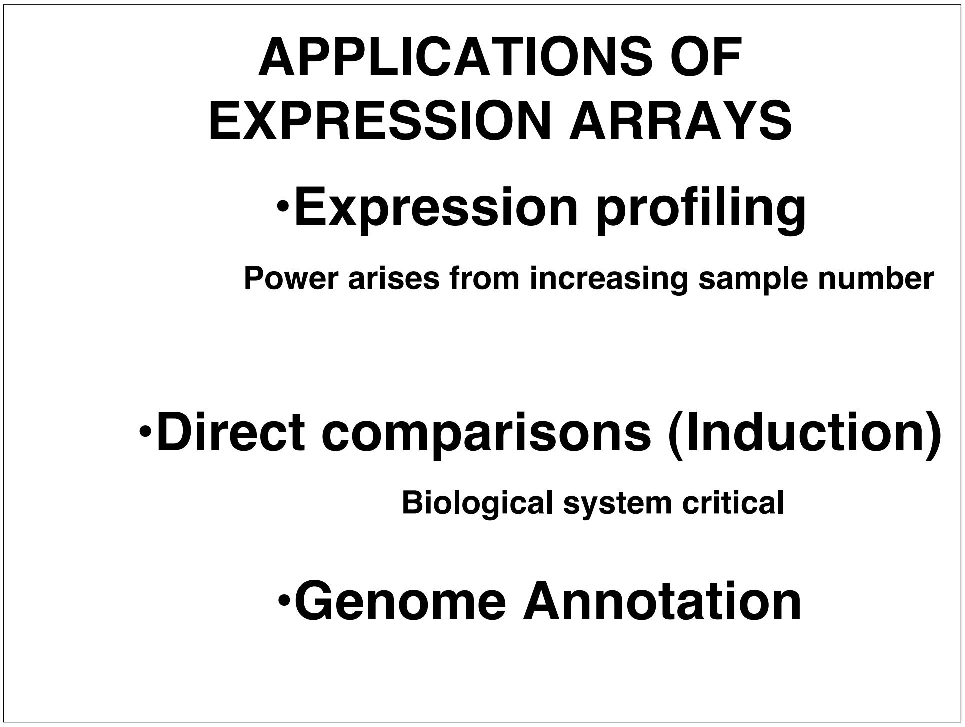
Output of Microarray Analysis:

expression ratio
(2 color hybridization)

or

relative expression level
(1 color hybridization)

Both types of data can be analyzed with essentially the same tools.



APPLICATIONS OF EXPRESSION ARRAYS

•Expression profiling

Power arises from increasing sample number

•Direct comparisons (Induction)

Biological system critical

•Genome Annotation

A RECURRING PROBLEM

Disease Genes

Transcription factors

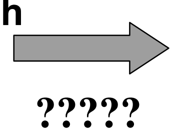
Hormones/growth factors

Drugs

Toxins

Infectious agents

Physical agents



<u>Downstream Genes</u>

•Direct targets

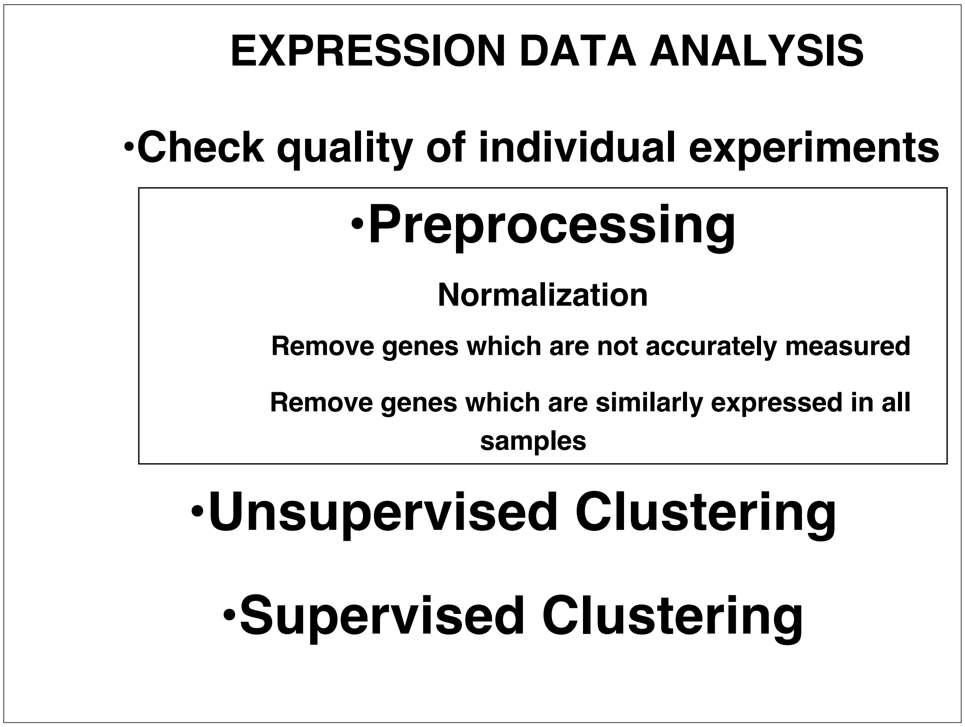
•Indirect targets

EXPRESSION DATA ANALYSIS

•Large amount of data

•Requires visualization and analysis tools

Recent overview of microarray bioinformatics:
Simon R, Curr Opin Biotechnol. 2008 Feb;19(1):26-9.



EXPRESSION DATA ANALYSIS

•Check quality of individual experiments

•Preprocessing

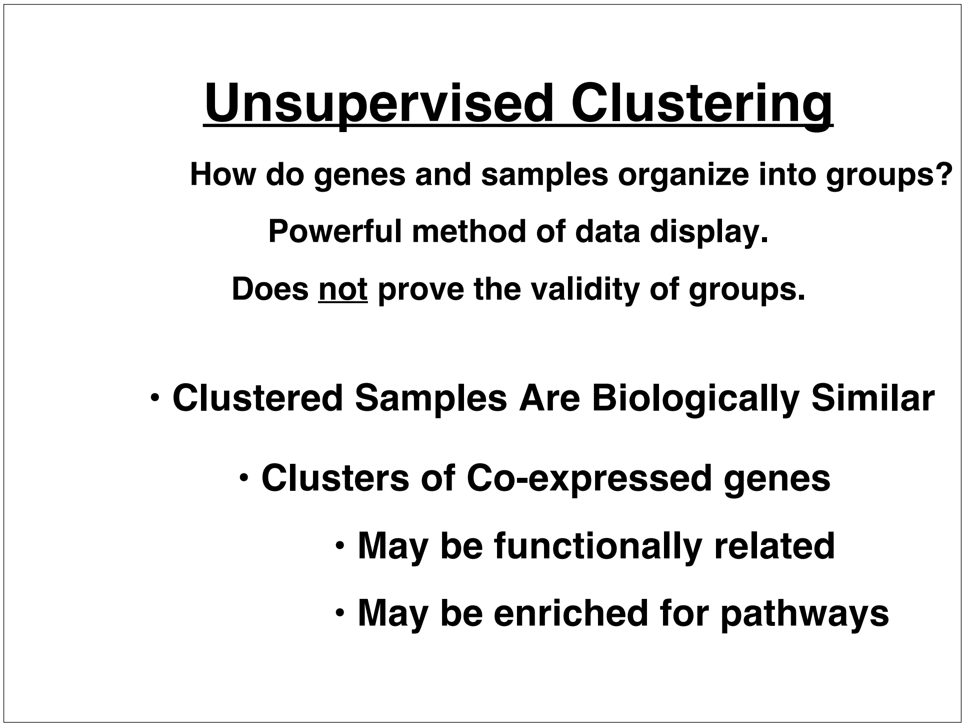
Normalization

Remove genes which are not accurately measured

Remove genes which are similarly expressed in all samples

•Unsupervised Clustering

•Supervised Clustering



Unsupervised Clustering

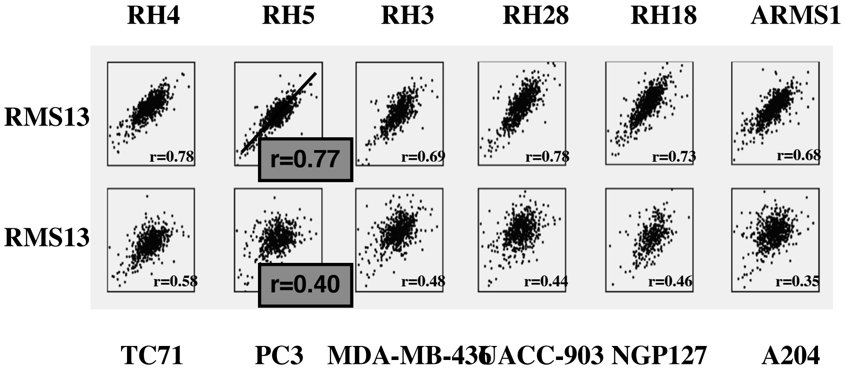
How do genes and samples organize into groups?

Powerful method of data display.

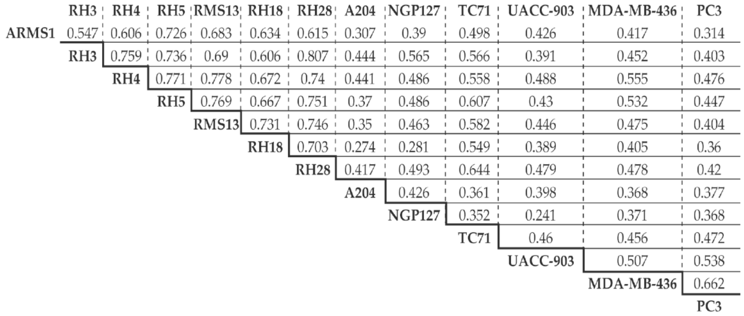
Does not prove the validity of groups.

- Clustered Samples Are Biologically Similar
- Clusters of Co-expressed genes
  - May be functionally related
  - May be enriched for pathways

UNSUPERVISED CLUSTERING IS BASED ON A GLOBAL SIMILARITY METRIC

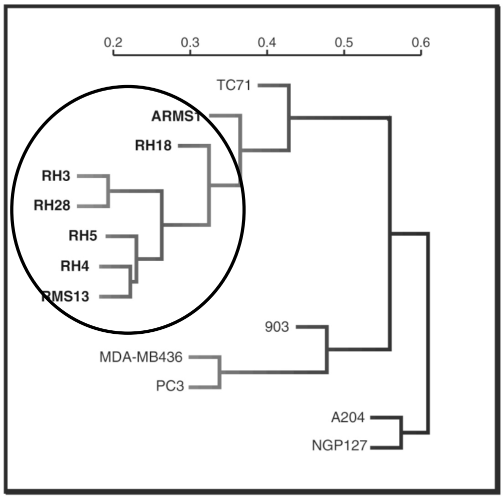


Matrix of Pearson Correlation Coefficients Distance Map

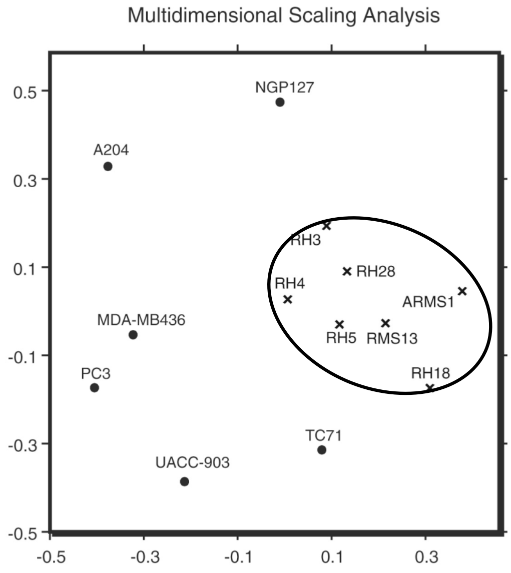


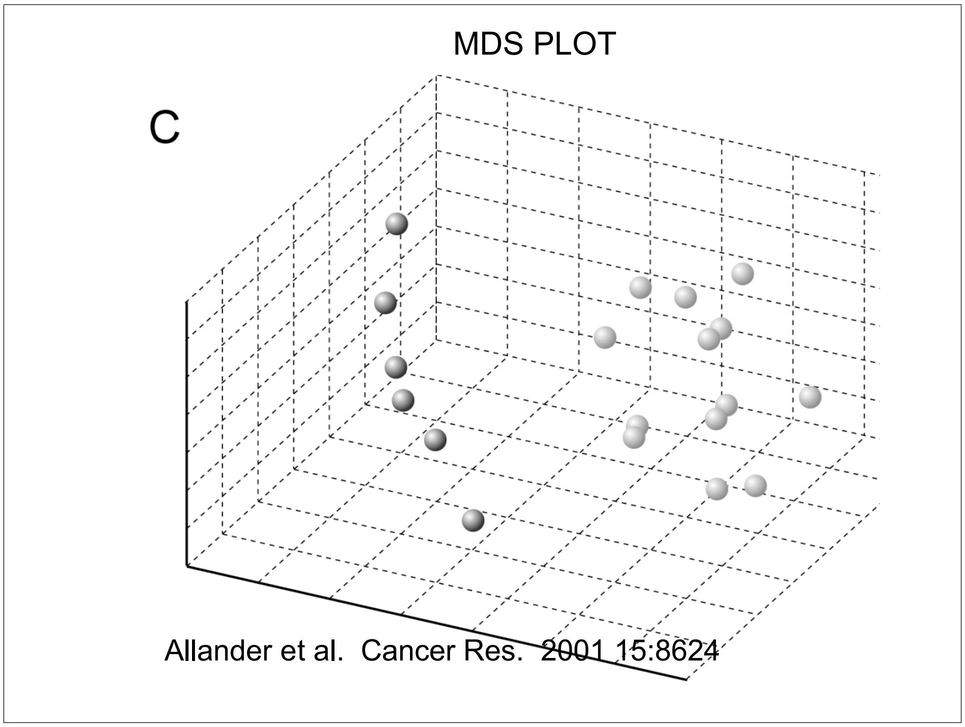
	RH3	RH4	RH5	RMS13	RH18	RH28	A204	NGP127	TC71	UACC-903	MDA-MB-436	PC3
ARMS1	0.547	0.606	0.726	0.683	0.634	0.615	0.307	0.39	0.498	0.426	0.417	0.314
RH3		0.759	0.736	0.69	0.606	0.807	0.444	0.565	0.566	0.391	0.452	0.403
RH4			0.771	0.778	0.672	0.74	0.441	0.486	0.558	0.488	0.555	0.476
RH5				0.769	0.667	0.751	0.37	0.486	0.607	0.43	0.532	0.447
RMS13					0.731	0.746	0.35	0.463	0.582	0.446	0.475	0.404
RH18						0.703	0.274	0.281	0.549	0.389	0.405	0.36
RH28							0.417	0.493	0.644	0.479	0.478	0.42
A204								0.426	0.361	0.398	0.368	0.377
NGP127									0.352	0.241	0.371	0.368
TC71										0.46	0.456	0.472
UACC-903											0.507	0.538
MDA-MB-436												0.662
PC3												

Hierarchical Clustering Dendrogram



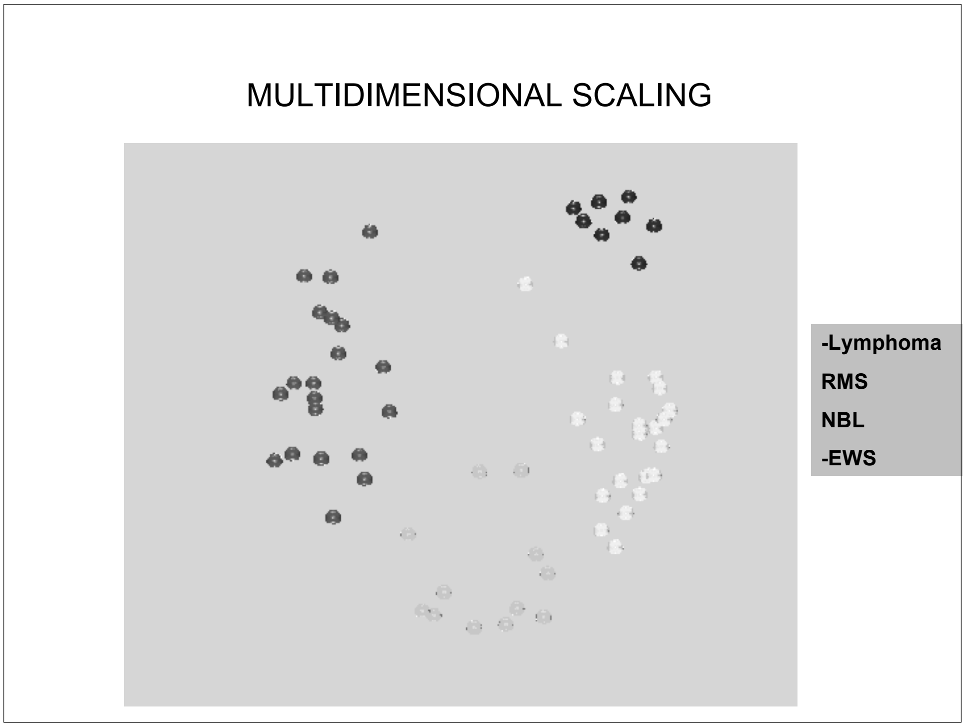
DATA DISPLAY BY MULTIDIMENSIONAL SCALING





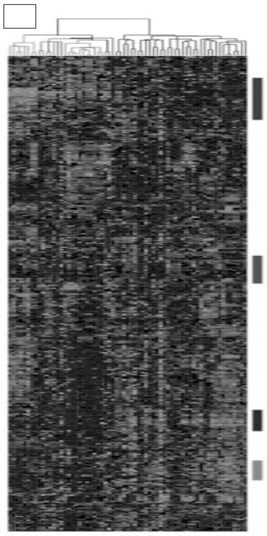
MDS PLOT

Allander et al. Cancer Res. 2001 15:8624



MULTIDIMENSIONAL SCALING

CLUSTERING GENES AND SAMPLES



Perou et al. Nature 2000 406:747

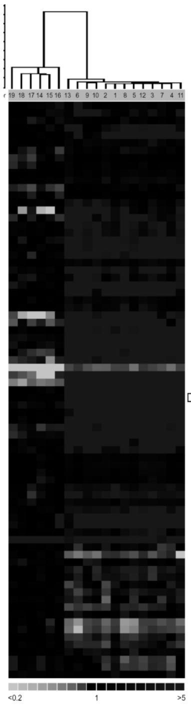
Supervised Clustering

What genes distinguish samples in selected groups from each other?

- **Choice of groups can be based on any known property of the samples.**
- **Many possible underlying methods: t-test or F-statistic frequently used.**
- **Output includes ranked gene list.**

- **Leads to the development of classifiers which can be applied to unknown samples.**

- **Must address the problem of false discovery due to multiple comparisons and discrepancy between sample/gene numbers.**

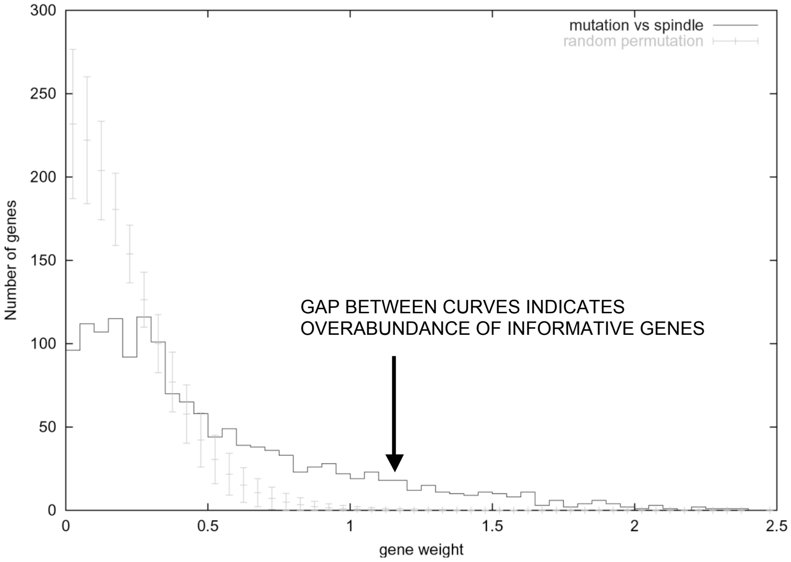


cloneID	Gene Symbol	Weight	α-value
378565	GHR	1	0
795309	SOD3	1.4	0
486641	MOCS2	1.5	0.00003
629896	MAP1B	1.8	0.00008
67769	KCNK3	1.7	0.00003
40304	SDHA	1.6	0.00006
2800976	CDX1	1.6	0
2556956	GOT2	1.1	0.0001
841370	GOT2	1.2	0.00003
42059	DLD	1.9	0.00001
41940	CXORF1	1.5	0.00006
878815	ARF3	1.4	0.00004
47202	ARF1	1.4	0
171912	EST	1.6	0.00008
347268	EST	1.5	0
1957165	RAB-R	2	0
2510105	PHF1	1.5	0.0001
450386	PHF1	1.6	0.00006
1461664	BCHE	2.8	0
2113077	KCNH2	1.9	0
50266	EST	1.7	0
824792	BM045	1.4	0.00009
292939	EST	1.3	0.00006
77533	INPP5A	1.9	0.00006
1049346	EST	1.5	0.00004
970684	MAP2	2.3	0.00008
789014	SLC12A4	2.3	0
789011	AAMP	1.7	0.00004
82225	SFRP1	1.9	0.00006
1435103	ATP2A3	2.1	0.00002
22493	RPL26	1.8	0.00004
52741	ABCA3	2.3	0
773421	EST	2.6	0
840575	EST	3	0.00008
2920924	ANXA3	2.7	0
489626	PFKM	3.3	0
1161564	DMN	3.3	0
2469213	ANXA3	4.5	0
2568905	GPR20	6.5	0
269806	**KIT**	7.6	0
375827	PTP4A3	2.6	0
174627	SCG2	2.6	0
205239	PRKCQ	2.8	0
2832322	TNFRSF6B	2	0
2164126	PRKCQ	2.6	0
303139	EST	2.4	0
2442397	ALDH2	1.6	0.00008
40026	SLC25A4	2.1	0
37196	EST	2.2	0.00008
39993	SOD1	1.6	0
950489	SOD1	1.4	0.00008
212165	PRDX2	1.6	0
2506776	PRDX2	1.3	0.00003
563444	FOXF1	1.6	0
753420	EST	1.9	0
2551468	GUCY1A3	1.6	0.00004
160485	GUCY1A3	1.5	0.00008
264951	NBS1	1.5	0.00008
839991	COL1A2	1.5	0.00001
1471829	DDXL	1.1	0.00004
2541394	EIF3S7	1.6	0
2576914	AKR1A1	1.5	0
44307	H2BFQ	1.5	0
745496	EST	1.3	0.00008
1594019	SUPT4H1	1.7	0
1492412	EST	1.6	0.00008
949940	RPL27A	1.5	0.00005
971367	RPS8	1.9	0
2467016	RPS27	1.3	0.00007
2548661	TIM44	1.7	0
2548854	RPL12	1.6	0.00008
2577093	MST1	1.7	0.0001
33632	GOT1	1.4	0
2578421	IGFBP4	1.3	0.00009
2545705	CNN3	1.6	0
486787	CNN3	1.5	0
755762	GRN	1.3	0.00003

HIERARCHICAL CLUSTERING OF SAMPLES/GENES USING THE GENES SELECTED BY SUPERVISED ANALYSIS

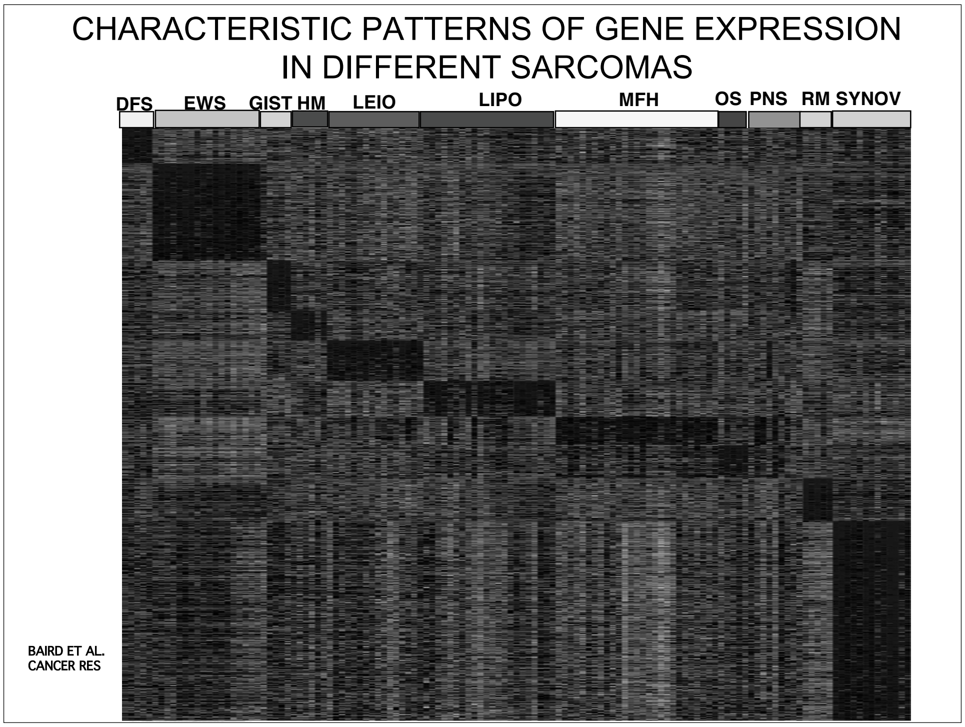
Allander et al. Cancer Res. 2001 15:8624

OVERABUNDANCE OF INFORMATIVE GENES DEMONSTRATED BY RANDOM PERMUTATION TEST



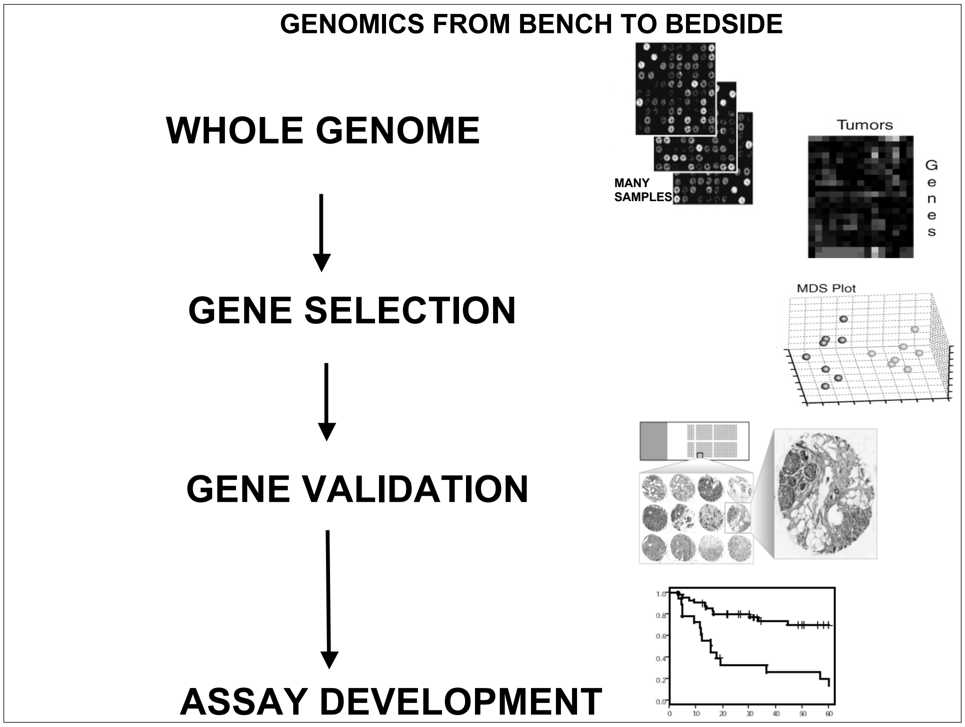
GAP BETWEEN CURVES INDICATES OVERABUNDANCE OF INFORMATIVE GENES

Allander et al. Cancer Res. 2001 15:8624



CHARACTERISTIC PATTERNS OF GENE EXPRESSION IN DIFFERENT SARCOMAS

BAIRD ET AL.
CANCER RES



GENOMICS FROM BENCH TO BEDSIDE

WHOLE GENOME

GENE SELECTION

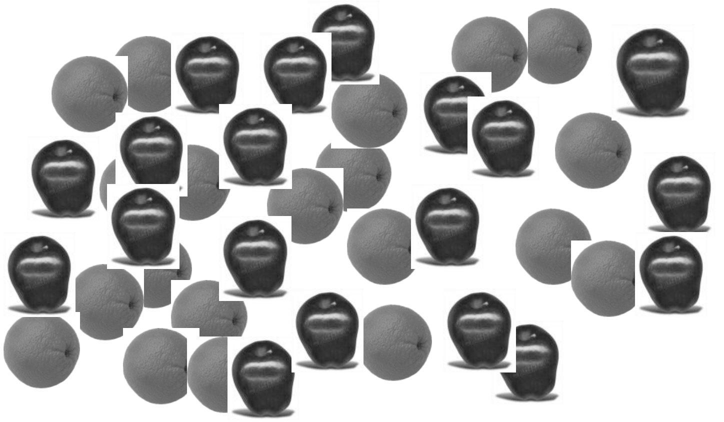
GENE VALIDATION

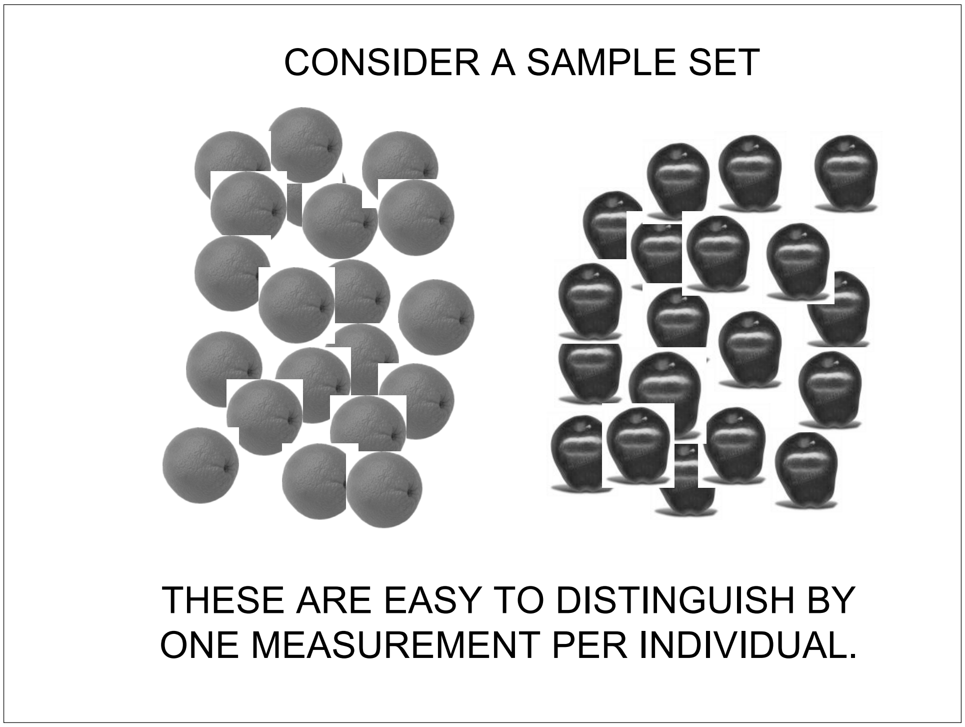
ASSAY DEVELOPMENT

SIGNAL STRENGTH VARIES IN TISSUE PROFILING EXPERIMENTS

THE MOST INTERESTING QUESTIONS TEND TO BE ASSOCIATED WITH WEAKER SIGNAL.

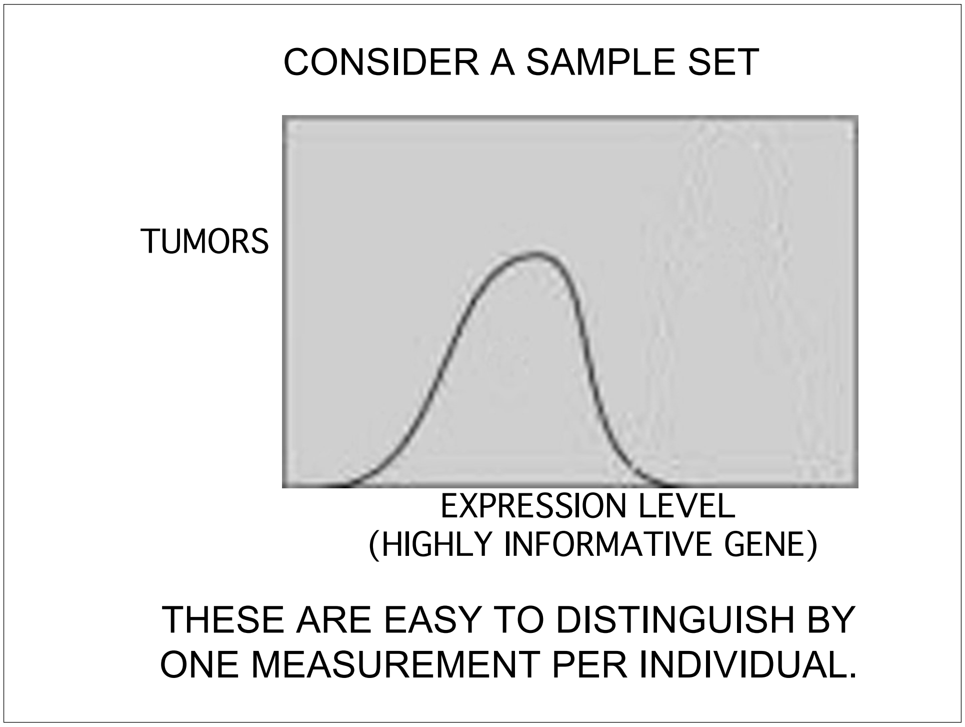
CONSIDER A SAMPLE SET





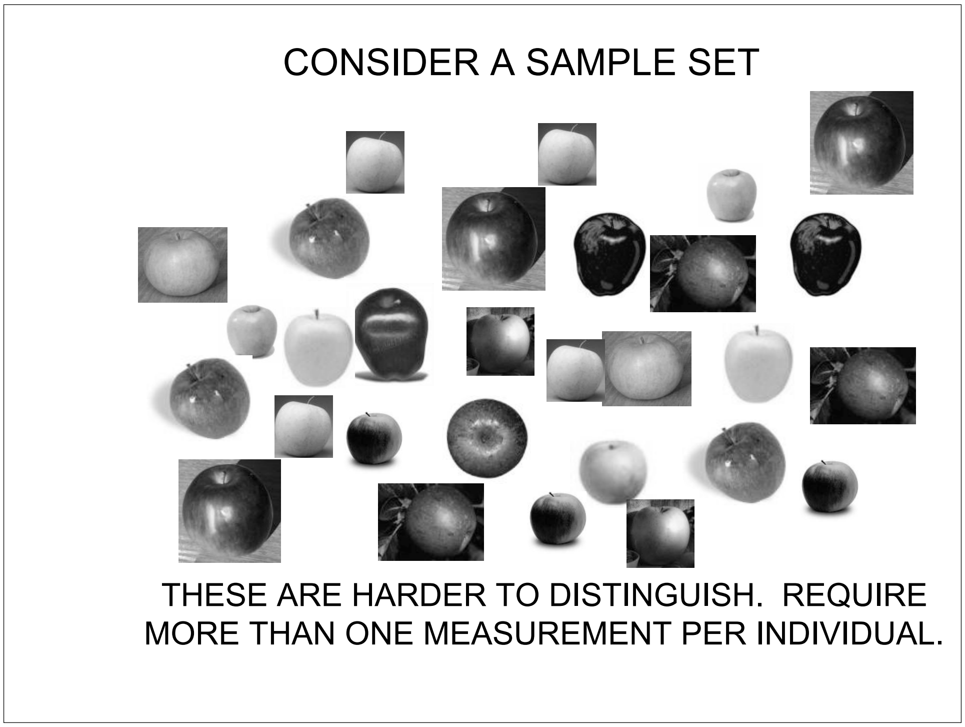
CONSIDER A SAMPLE SET

THESE ARE EASY TO DISTINGUISH BY ONE MEASUREMENT PER INDIVIDUAL.



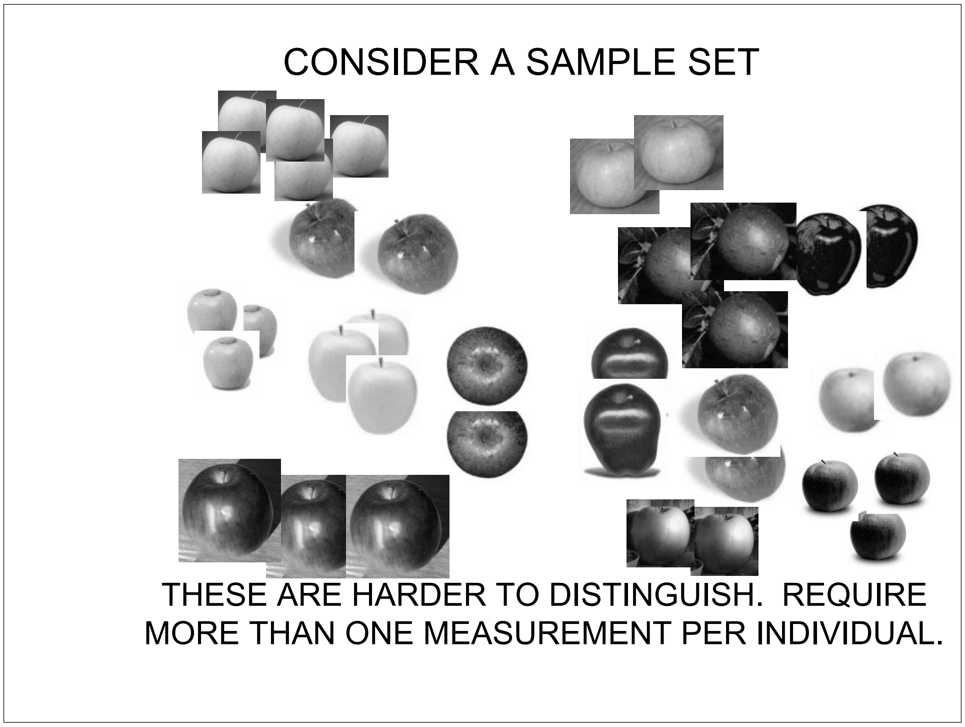
CONSIDER A SAMPLE SET

THESE ARE EASY TO DISTINGUISH BY ONE MEASUREMENT PER INDIVIDUAL.



CONSIDER A SAMPLE SET

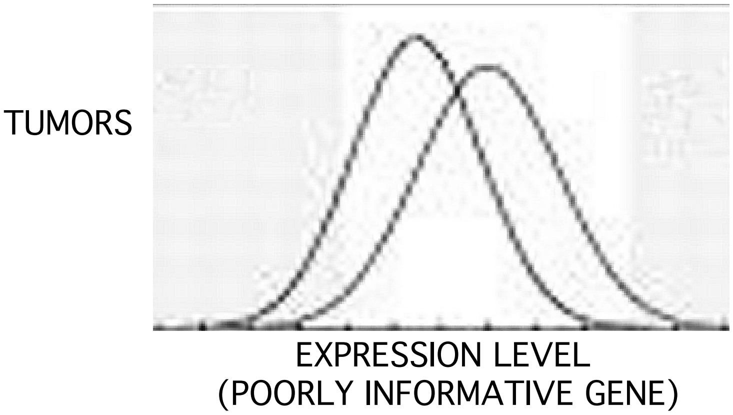
THESE ARE HARDER TO DISTINGUISH. REQUIRE MORE THAN ONE MEASUREMENT PER INDIVIDUAL.



CONSIDER A SAMPLE SET

THESE ARE HARDER TO DISTINGUISH. REQUIRE MORE THAN ONE MEASUREMENT PER INDIVIDUAL.

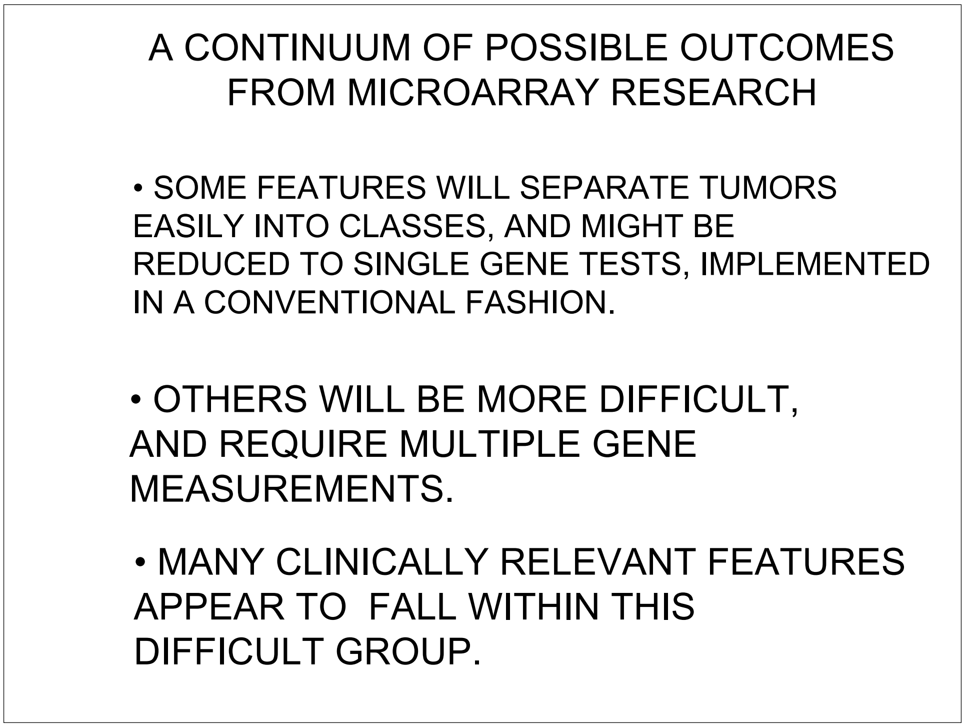
CONSIDER A SAMPLE SET



THESE ARE HARDER TO DISTINGUISH. REQUIRE MORE THAN ONE MEASUREMENT PER INDIVIDUAL.

WE CAN TELL APPLES FROM ORANGES.

CAN WE DISTINGUISH DIFFERENT KINDS OF APPLES?

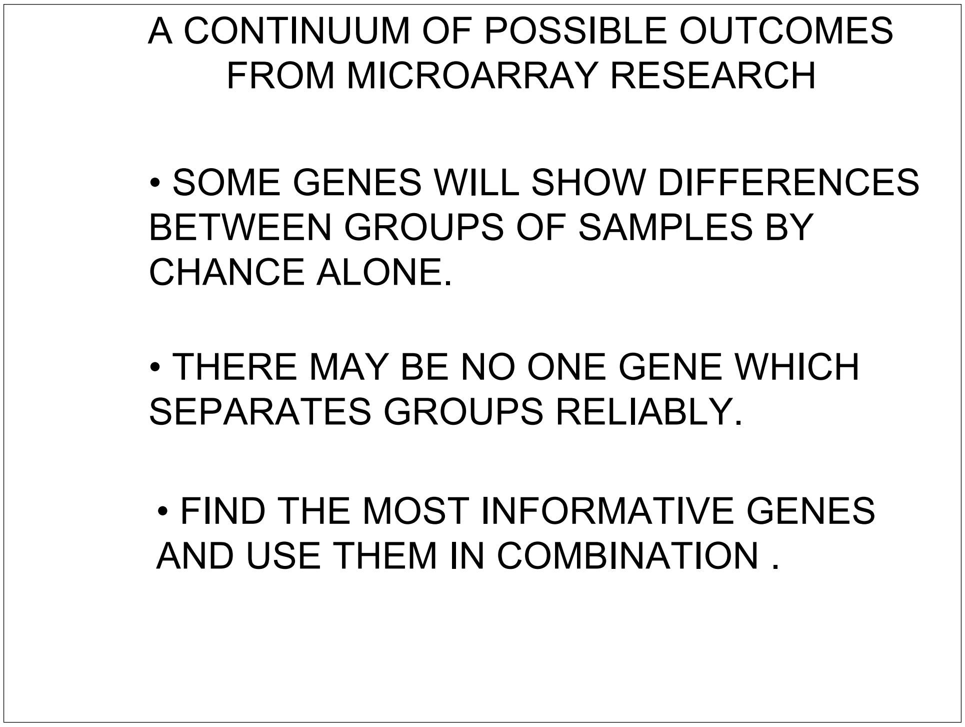


A CONTINUUM OF POSSIBLE OUTCOMES FROM MICROARRAY RESEARCH

- SOME FEATURES WILL SEPARATE TUMORS EASILY INTO CLASSES, AND MIGHT BE REDUCED TO SINGLE GENE TESTS, IMPLEMENTED IN A CONVENTIONAL FASHION.

- OTHERS WILL BE MORE DIFFICULT, AND REQUIRE MULTIPLE GENE MEASUREMENTS.

- MANY CLINICALLY RELEVANT FEATURES APPEAR TO FALL WITHIN THIS DIFFICULT GROUP.

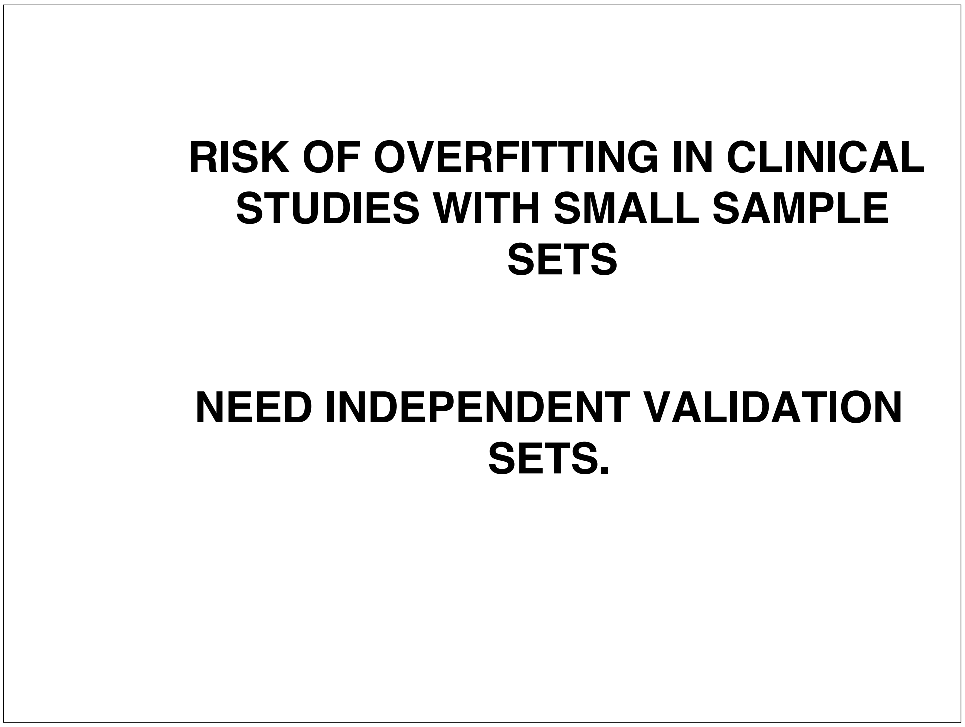


A CONTINUUM OF POSSIBLE OUTCOMES FROM MICROARRAY RESEARCH

- SOME GENES WILL SHOW DIFFERENCES BETWEEN GROUPS OF SAMPLES BY CHANCE ALONE.

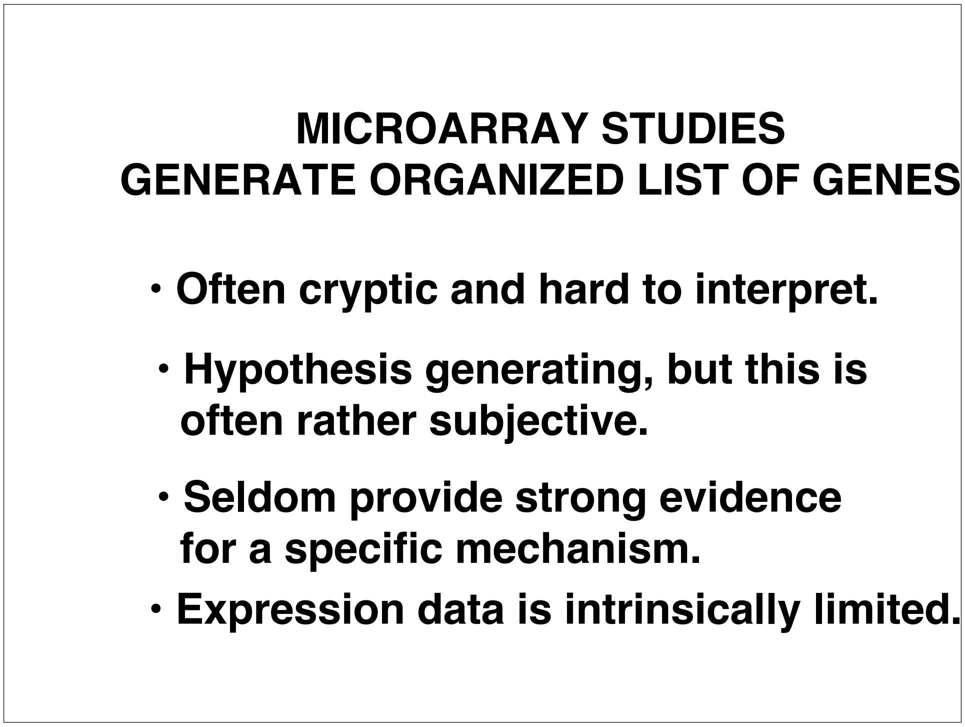
- THERE MAY BE NO ONE GENE WHICH SEPARATES GROUPS RELIABLY.

- FIND THE MOST INFORMATIVE GENES AND USE THEM IN COMBINATION .



RISK OF OVERFITTING IN CLINICAL STUDIES WITH SMALL SAMPLE SETS

NEED INDEPENDENT VALIDATION SETS.

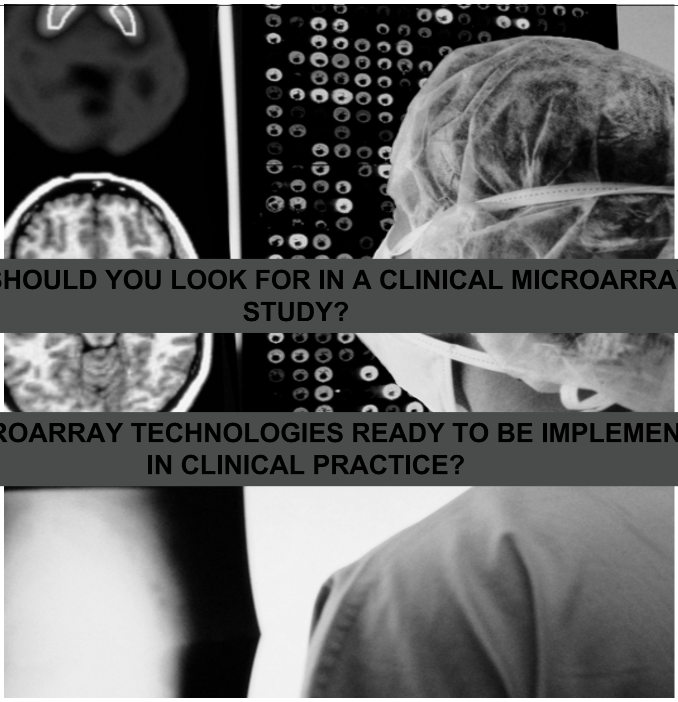


MICROARRAY STUDIES GENERATE ORGANIZED LIST OF GENES

- **Often cryptic and hard to interpret.**
- **Hypothesis generating, but this is often rather subjective.**
- **Seldom provide strong evidence for a specific mechanism.**
- **Expression data is intrinsically limited.**

GETTING BEYOND GENE LISTS

- Optimal use of gene annotations.
- Optimizing use of public data.
- Incorporating data from model systems.
- Linking expression data to sequence.
- Adding other types of genome scale data.



WHAT SHOULD YOU LOOK FOR IN A CLINICAL MICROARRAY STUDY?

ARE MICROARRAY TECHNOLOGIES READY TO BE IMPLEMENTED IN CLINICAL PRACTICE?

WHAT TO LOOK FOR IN CLINICAL CORRELATIVE STUDIES USING MICROARRAYS

- WELL DEFINED QUESTION AND PATIENT SAMPLE.
- HIGH QUALITY ARRAY MEASUREMENTS (HARD TO ASSESS WITHOUT REFERENCE TO PRIMARY DATA---SHOULD BE MADE PUBLIC).
- APPROPRIATE AND RIGOROUS STATISTICAL ANALYSIS OF ARRAY DATA.
- FORMAL CLASSIFIER THAT CAN BE APPLIED TO NEW SAMPLES.
- VALIDATION SAMPLE SET.

WHAT TO LOOK FOR IN CLINICAL CORRELATIVE STUDIES USING MICROARRAYS

- **GOAL SHOULD BE TO SEEK AND VALIDATE CLINICALLY RELEVANT SIGNATURES WITHIN DEFINED PATIENT GROUPS FOR WHICH NO CURRENT FEATURES ADEQUATELY ANSWER THE CLINICAL QUESTION POSED.**

EXPRESSION PROFILING IN THE CLINIC?

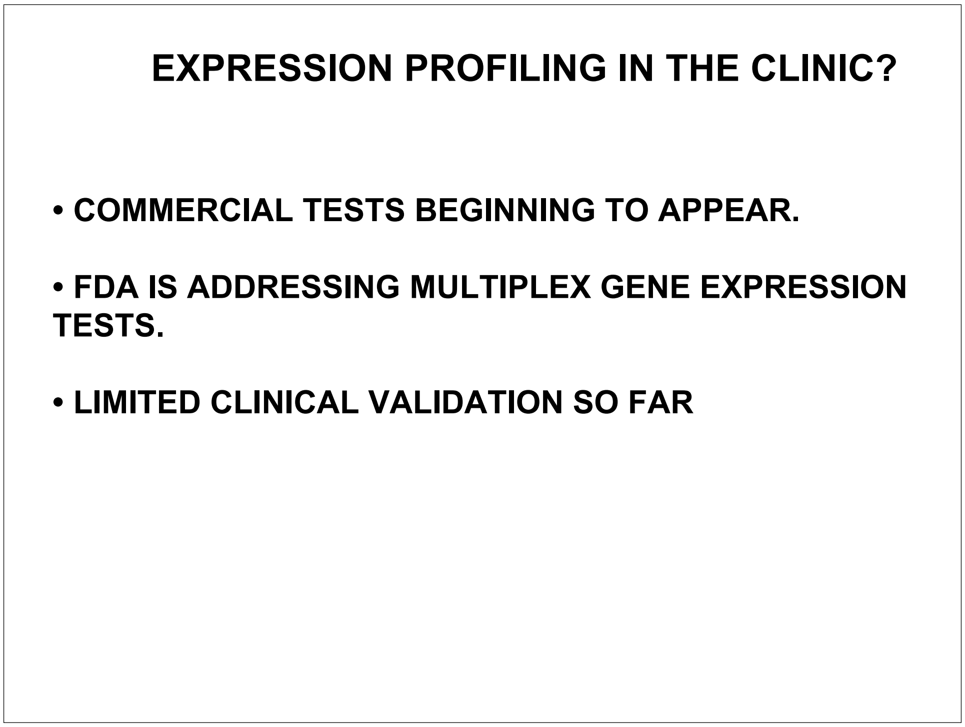
PROBLEMS:

- **SPECIALIZED TECHNOLOGY**
- **RNA IS UNSTABLE**
- **FROZEN TISSUE NOT PART OF USUAL OR SAMPLE FLOW**

EXPRESSION PROFILING IN THE CLINIC?

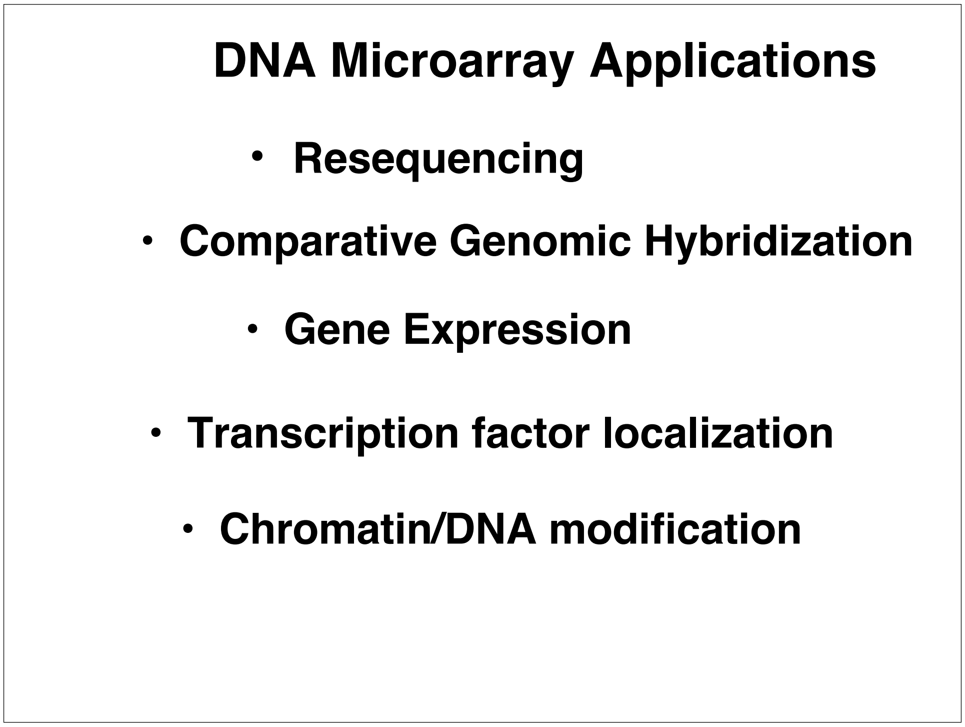
OPTIONS:

- **REFERENCE LABORATORIES**
- **RNA PRESERVATIVES**
- **USE OF PARAFFIN EMBEDDED MATERIALS.**



EXPRESSION PROFILING IN THE CLINIC?

- **COMMERCIAL TESTS BEGINNING TO APPEAR.**
- **FDA IS ADDRESSING MULTIPLEX GENE EXPRESSION TESTS.**
- **LIMITED CLINICAL VALIDATION SO FAR**



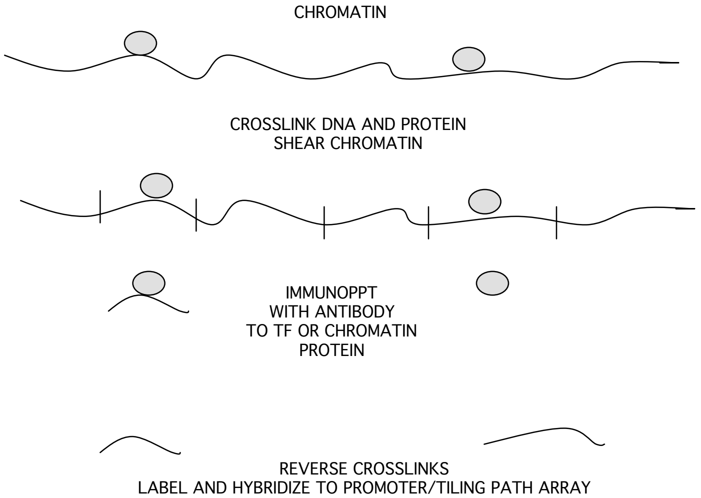
DNA Microarray Applications

- **Resequencing**
- **Comparative Genomic Hybridization**
- **Gene Expression**
- **Transcription factor localization**
- **Chromatin/DNA modification**

APPLICATIONS OF TILING PATH ARRAYS

- CGH
- EXPRESSION
- ChIP CHIP
- DNAse HYPESENSITIVE SITES
- ANY ENRICHED PREPARATION OF INTERESTING SEQUENCES

TRANSCRIPTION FACTOR LOCALIZATION ON ARRAYS

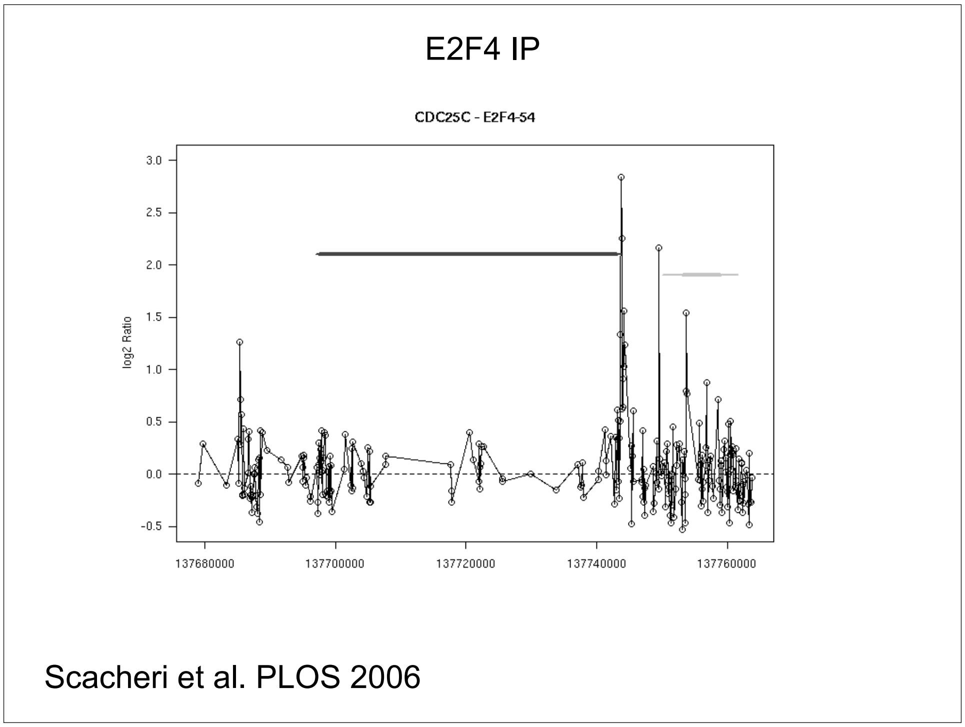


CHROMATIN

CROSSLINK DNA AND PROTEIN
SHEAR CHROMATIN

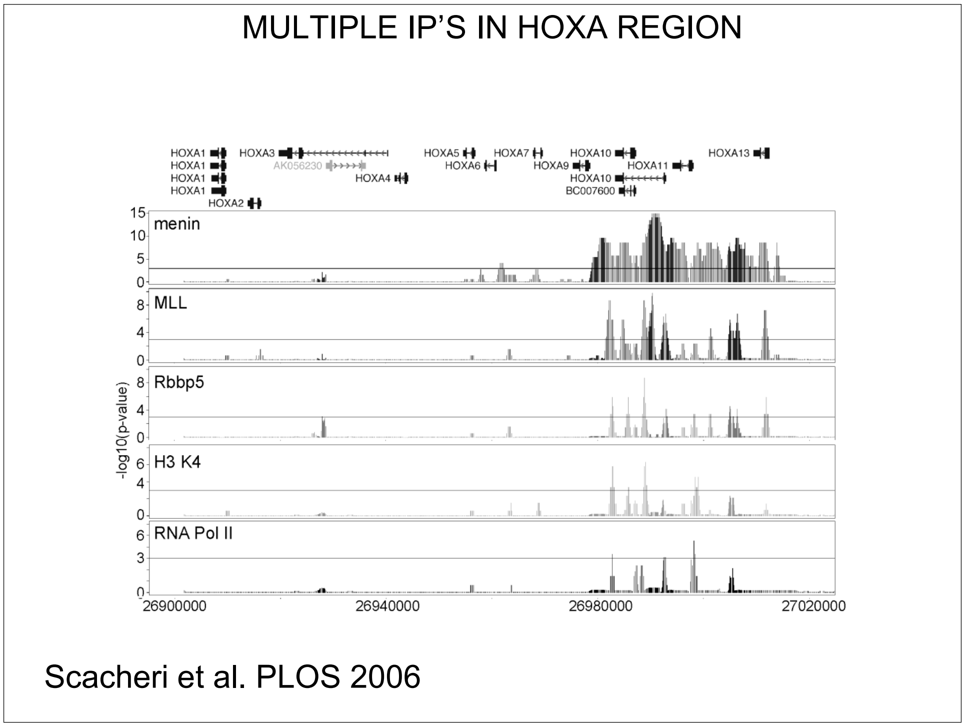
IMMUNOPPT
WITH ANTIBODY
TO TF OR CHROMATIN
PROTEIN

REVERSE CROSSLINKS
LABEL AND HYBRIDIZE TO PROMOTER/TILING PATH ARRAY



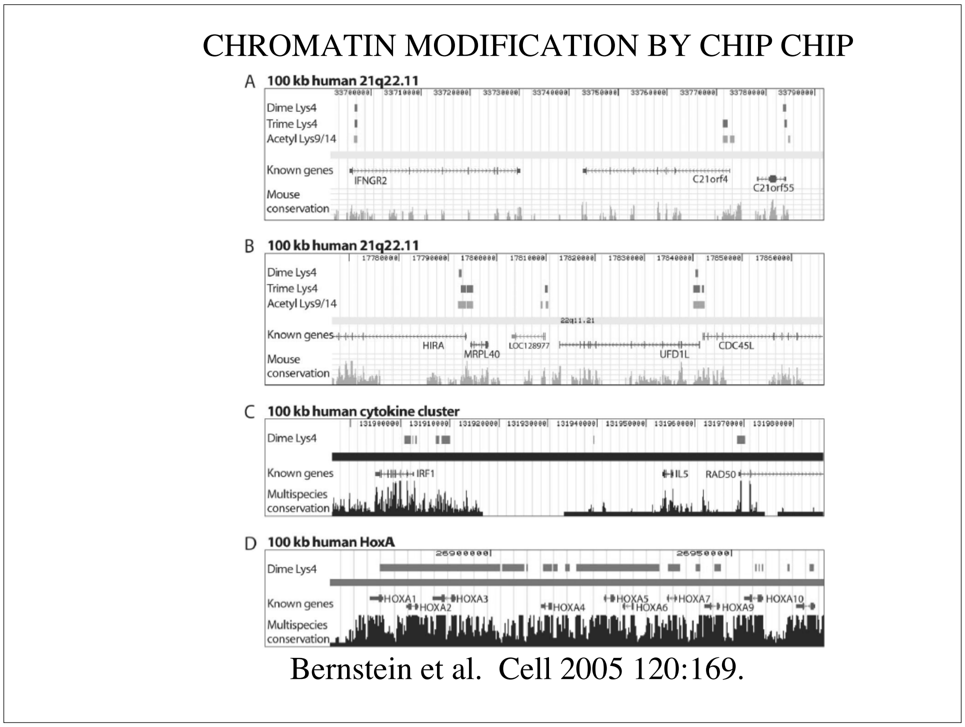
E2F4 IP

Scacheri et al. PLOS 2006



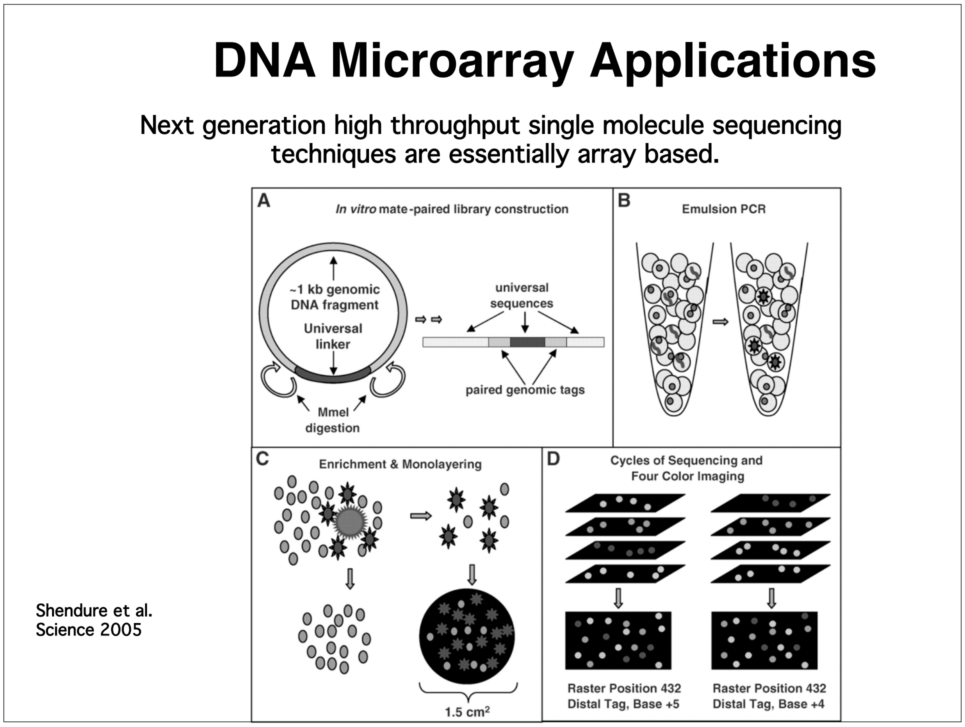
MULTIPLE IP'S IN HOXA REGION

Scacheri et al. PLOS 2006



CHROMATIN MODIFICATION BY CHIP CHIP

Bernstein et al. Cell 2005 120:169.

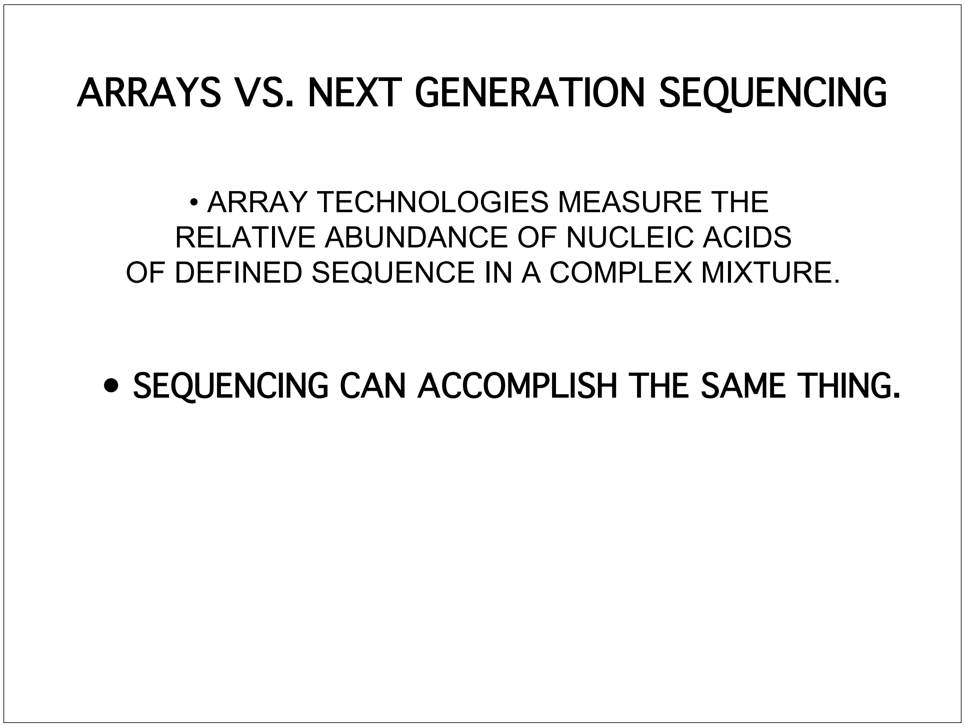


DNA Microarray Applications

Next generation high throughput single molecule sequencing techniques are essentially array based.

A
In vitro mate-paired library construction
~1 kb genomic DNA fragment
Universal linker
MmeI digestion
universal sequences
paired genomic tags
B
Emulsion PCR
C
Enrichment & Monolayering
1.5 cm^2
D
Cycles of Sequencing and Four Color Imaging
Raster Position 432 Distal Tag, Base +5
Raster Position 432 Distal Tag, Base +4

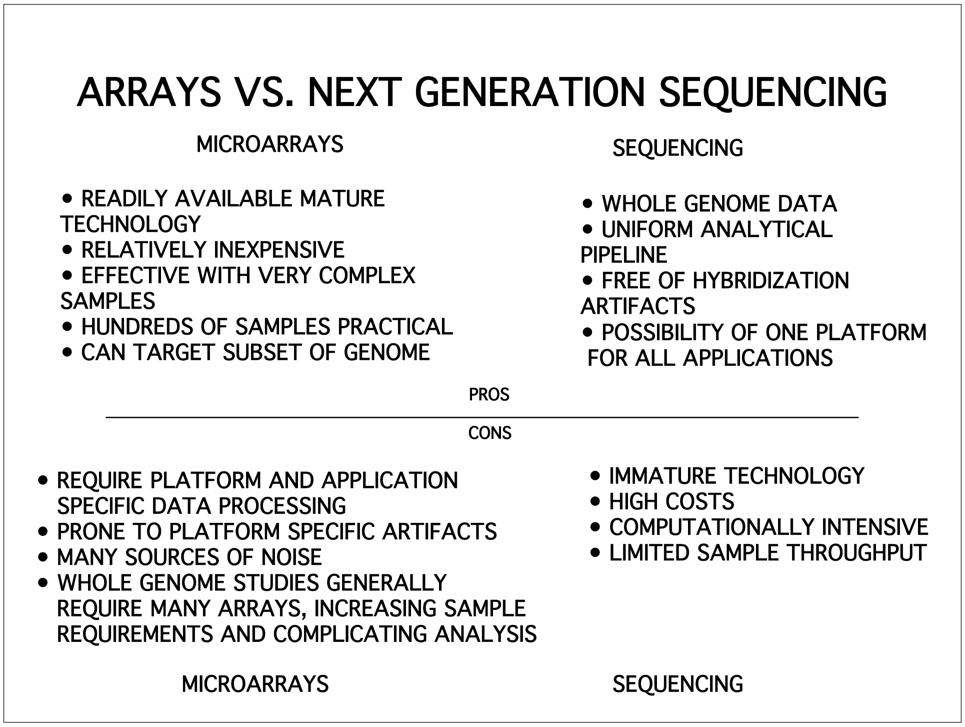
Shendure et al.
Science 2005



ARRAYS VS. NEXT GENERATION SEQUENCING

- ARRAY TECHNOLOGIES MEASURE THE RELATIVE ABUNDANCE OF NUCLEIC ACIDS OF DEFINED SEQUENCE IN A COMPLEX MIXTURE.

- SEQUENCING CAN ACCOMPLISH THE SAME THING.



ARRAYS VS. NEXT GENERATION SEQUENCING

MICROARRAYS

- READILY AVAILABLE MATURE TECHNOLOGY
- RELATIVELY INEXPENSIVE
- EFFECTIVE WITH VERY COMPLEX SAMPLES
- HUNDREDS OF SAMPLES PRACTICAL
- CAN TARGET SUBSET OF GENOME

SEQUENCING

- WHOLE GENOME DATA
- UNIFORM ANALYTICAL PIPELINE
- FREE OF HYBRIDIZATION ARTIFACTS
- POSSIBILITY OF ONE PLATFORM FOR ALL APPLICATIONS

PROS

CONS

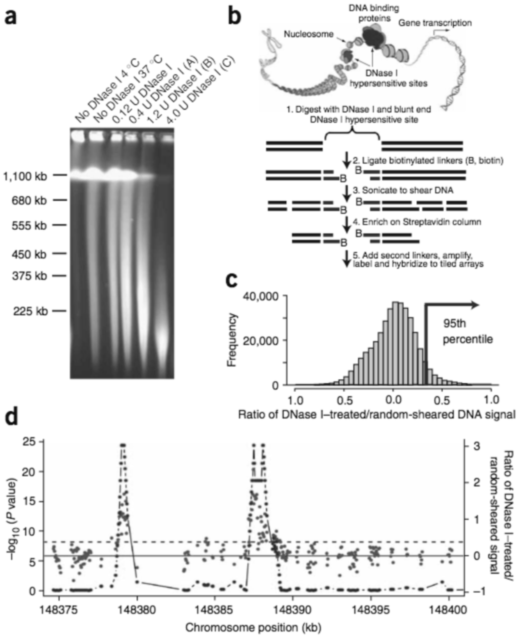
- REQUIRE PLATFORM AND APPLICATION SPECIFIC DATA PROCESSING
- PRONE TO PLATFORM SPECIFIC ARTIFACTS
- MANY SOURCES OF NOISE
- WHOLE GENOME STUDIES GENERALLY REQUIRE MANY ARRAYS, INCREASING SAMPLE REQUIREMENTS AND COMPLICATING ANALYSIS

MICROARRAYS

- IMMATURE TECHNOLOGY
- HIGH COSTS
- COMPUTATIONALLY INTENSIVE
- LIMITED SAMPLE THROUGHPUT

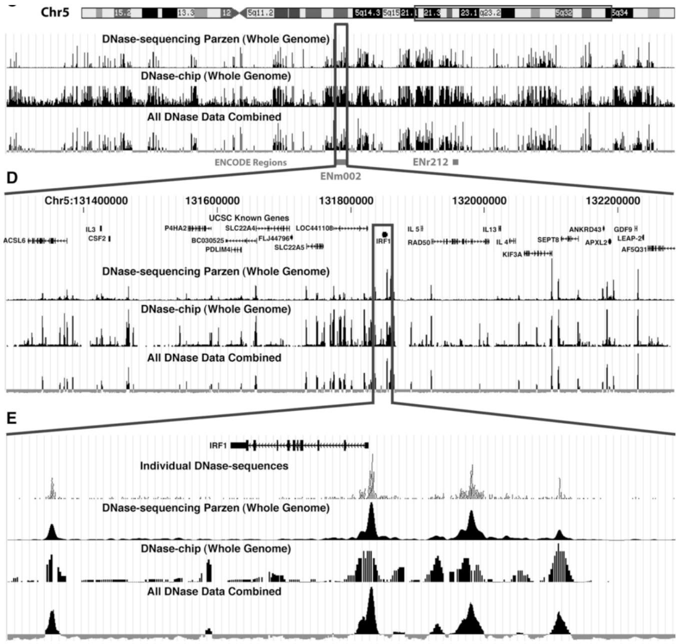
SEQUENCING

DNase HS Mapping by Microarray

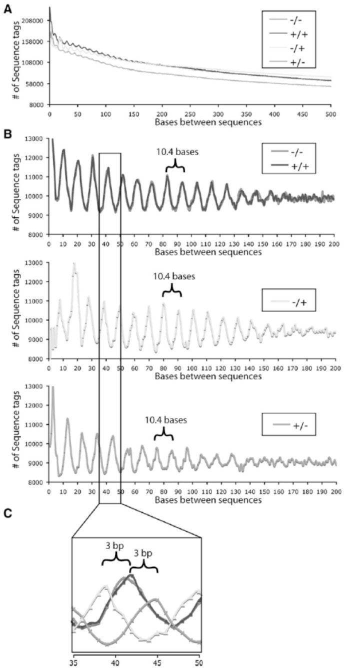


Crawford et al.
Nature Methods 2006

DNase HS Mapping by Microarray and Sequencing



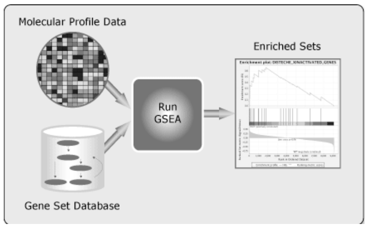
Boyle et al. Cell 2008



Boyle et al. Cell 2008

FRONTIERS OF INTEGRATED GENOMICS

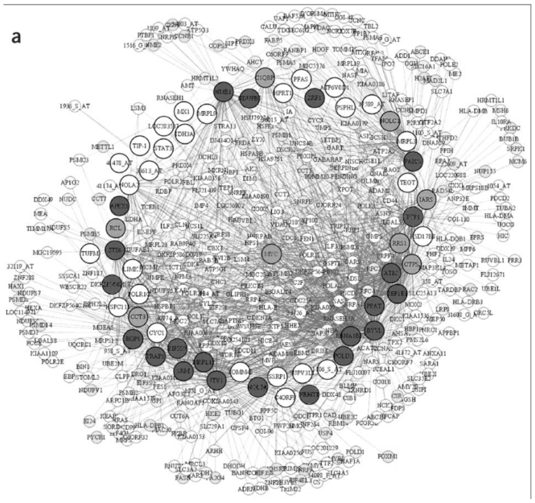
- DEVELOPING SPECIFIC SIGNATURES FOR GENES, PATHWAYS, COMPOUNDS
- REQUIRES LARGE AMOUNTS OF DATA
- GENE SET ENRICHMENT ANALYSIS (GSEA)



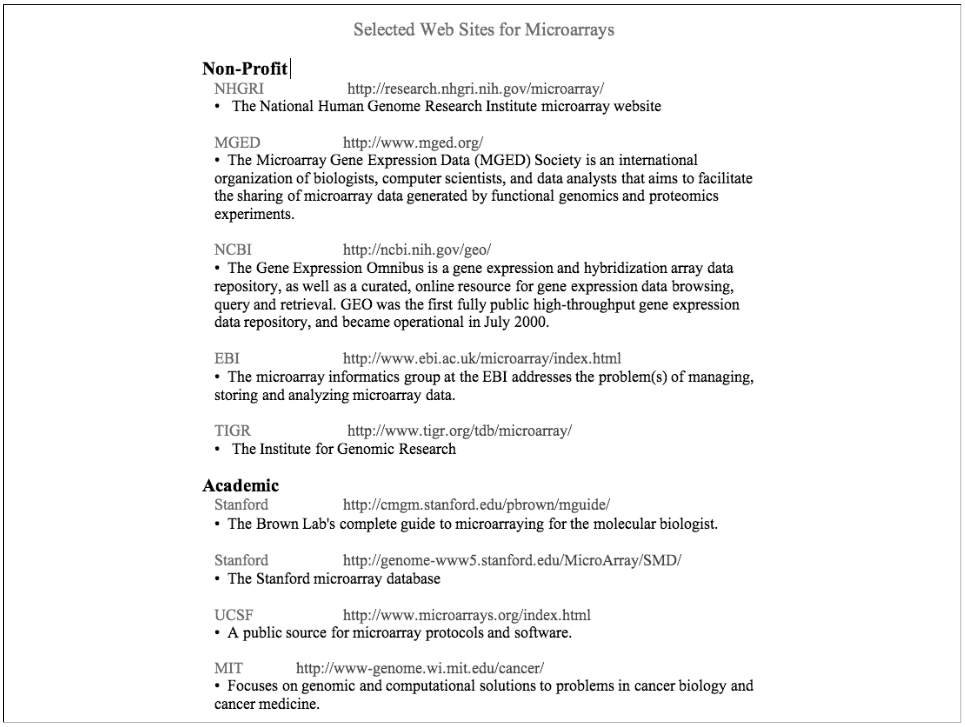
http://www.broad.mit.edu/gsea/

FRONTIERS OF INTEGRATED GENOMICS

CONSTRUCTING CELLULAR NETWORKS FROM GENOMIC DATA THROUGH DATA AND DATABASE INTEGRATION



Basso et al. Nat Genet. 2005 Apr;37(4):382-90.



Selected Web Sites for Microarrays

Non-Profit

NHGRI http://research.nhgri.nih.gov/microarray/

- The National Human Genome Research Institute microarray website

MGED http://www.mged.org/

- The Microarray Gene Expression Data (MGED) Society is an international organization of biologists, computer scientists, and data analysts that aims to facilitate the sharing of microarray data generated by functional genomics and proteomics experiments.

NCBI http://ncbi.nih.gov/geo/

- The Gene Expression Omnibus is a gene expression and hybridization array data repository, as well as a curated, online resource for gene expression data browsing, query and retrieval. GEO was the first fully public high-throughput gene expression data repository, and became operational in July 2000.

EBI http://www.ebi.ac.uk/microarray/index.html

- The microarray informatics group at the EBI addresses the problem(s) of managing, storing and analyzing microarray data.

TIGR http://www.tigr.org/tdb/microarray/

- The Institute for Genomic Research

Academic

Stanford http://cmgm.stanford.edu/pbrown/mguide/

- The Brown Lab's complete guide to microarraying for the molecular biologist.

Stanford http://genome-www5.stanford.edu/MicroArray/SMD/

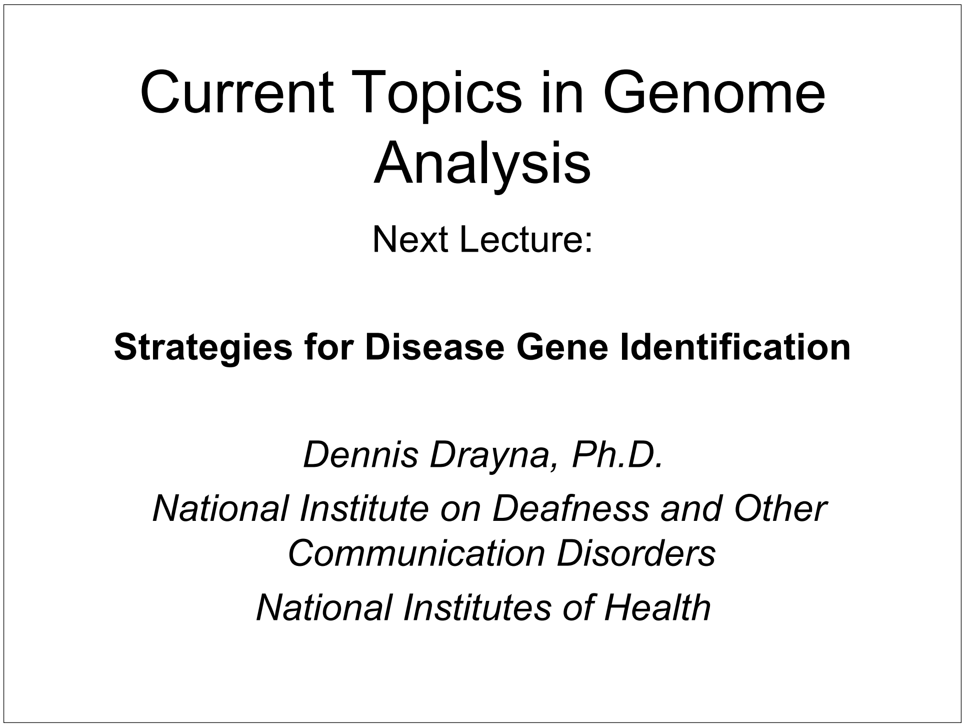
- The Stanford microarray database

UCSF http://www.microarrays.org/index.html

- A public source for microarray protocols and software.

MIT http://www-genome.wi.mit.edu/cancer/

- Focuses on genomic and computational solutions to problems in cancer biology and cancer medicine.



Current Topics in Genome Analysis

Next Lecture:

Strategies for Disease Gene Identification

Dennis Drayna, Ph.D.
National Institute on Deafness and Other Communication Disorders
National Institutes of Health