LARGE SCALE ANALYSIS OF GENE EXPRESSION

Evolution and Revolution

AFTER THE SEQUENCE:

WHOLE GENOME APPROACHES TO

BIOLOGICAL QUESTIONS

GENE EXPRESSION

GENE VARIATION

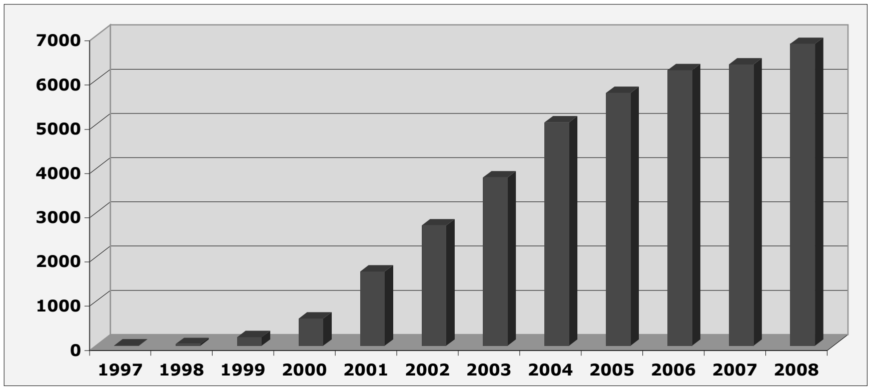
GENE FUNCTION

MICROARRAYS PROVIDE A TOOL FOR WHOLE GENOME ANALYSIS

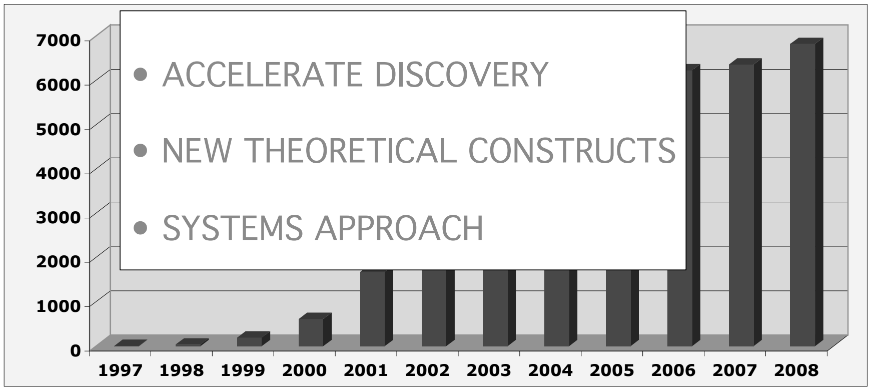
PRIMARY IMPACT:

ACCELERATED DISCOVERY AND HYPOTHESIS GENERATION

PUBMED CITATIONS FOR DNA MICROARRAYS



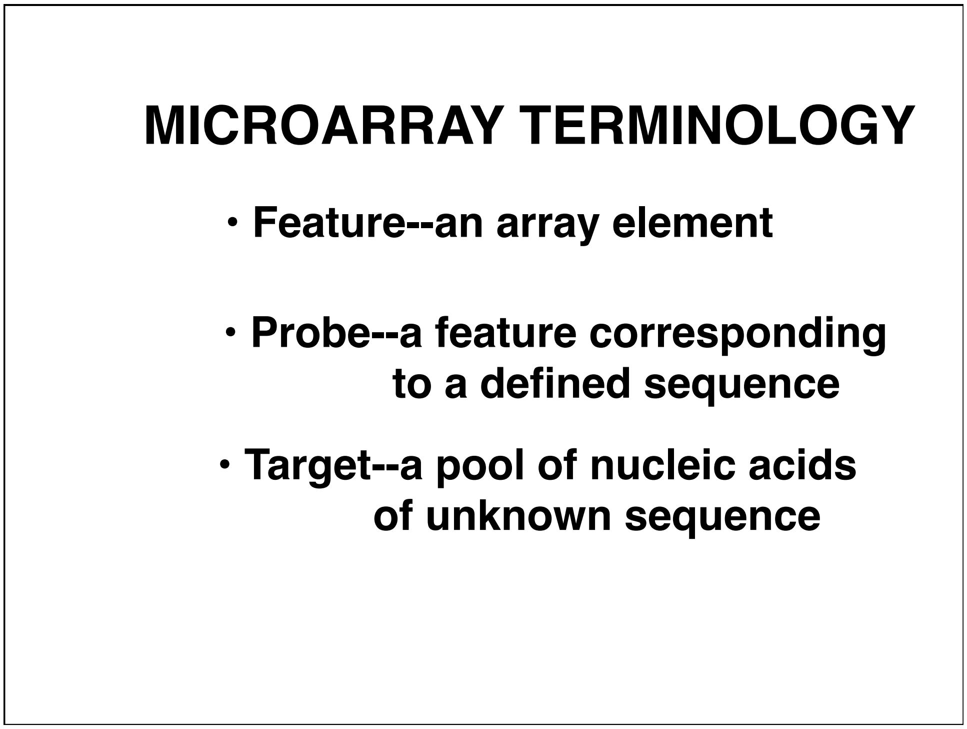
PUBMED CITATIONS FOR DNA MICROARRAYS



- ACCELERATE DISCOVERY
- NEW THEORETICAL CONSTRUCTS
- SYSTEMS APPROACH

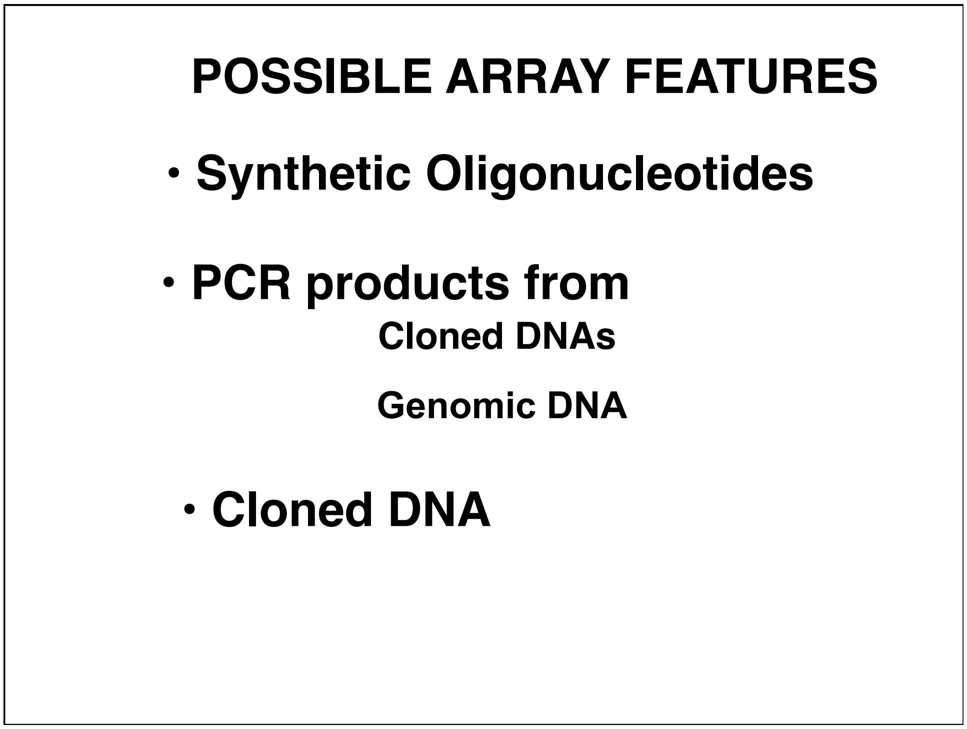
INCREASE IN FEATURE DENSITY

SEQUENCING TECHNOLOGIES MAY SUPPLANT ARRAYS IN MANY APPLICATIONS



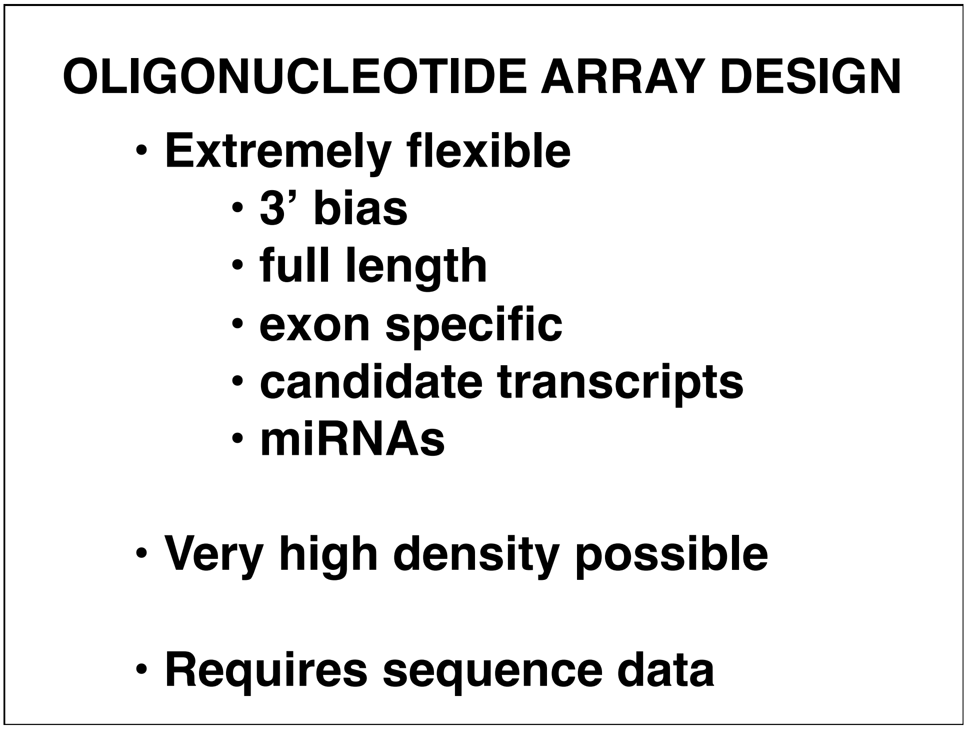
MICROARRAY TERMINOLOGY

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



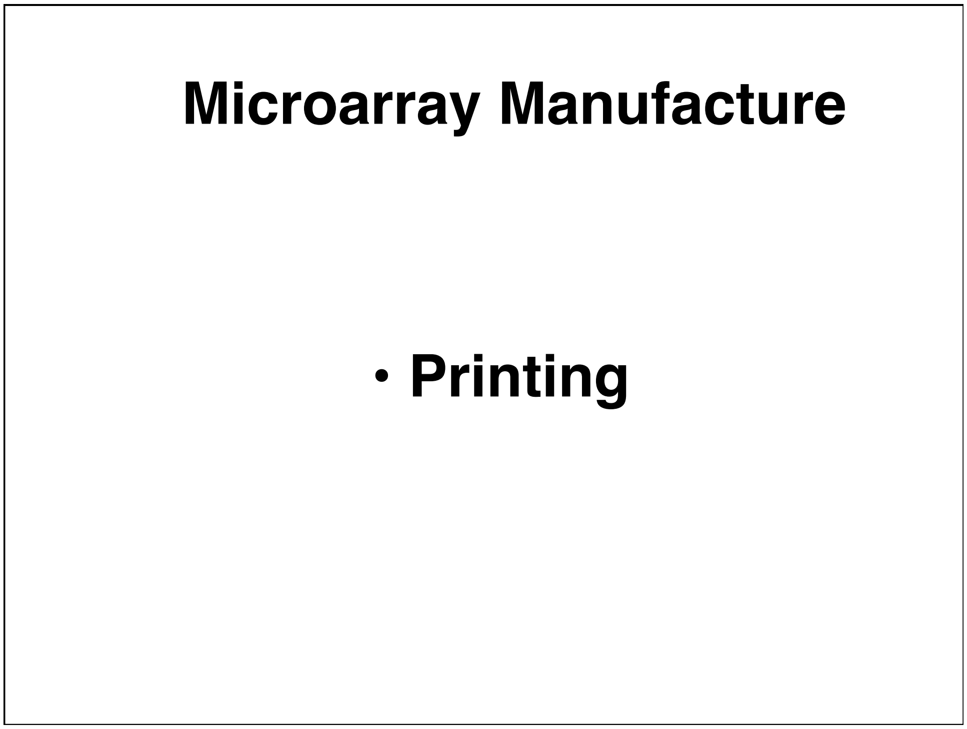
POSSIBLE ARRAY FEATURES

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



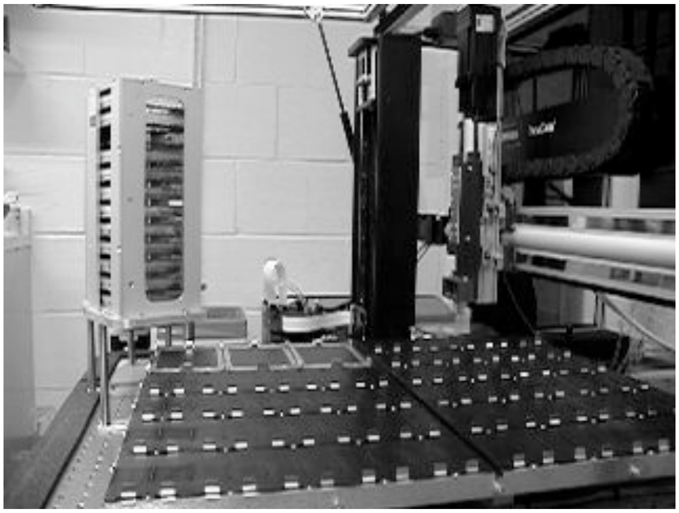
OLIGONUCLEOTIDE ARRAY DESIGN

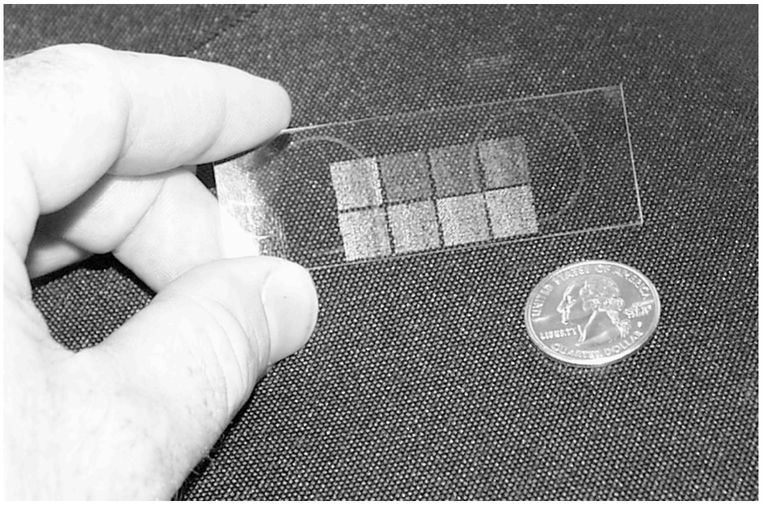
- Extremely flexible
  - 3’ bias
  - full length
  - exon specific
  - candidate transcripts
  - miRNAs
- Very high density possible
- Requires sequence data



Microarray Manufacture

- Printing

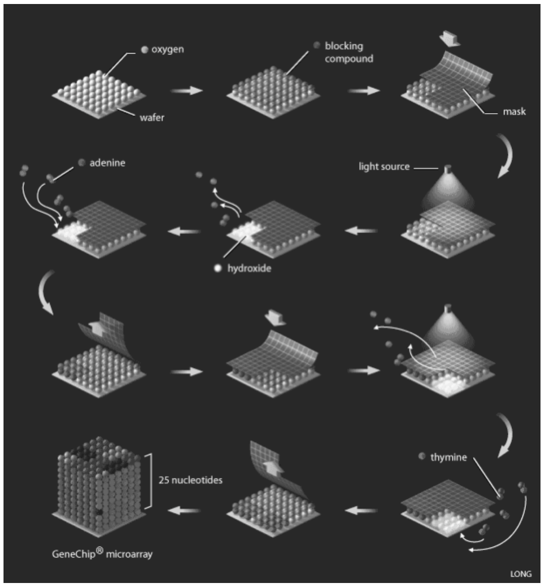



UNITED STATES OF AMERICA
LIBERTY
QUARTER DOLLAR

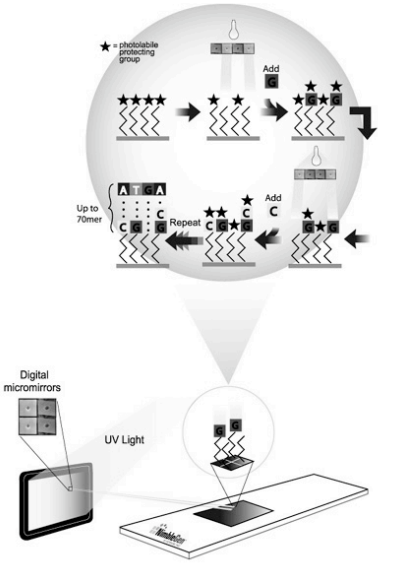
Microarray Manufacture

- **Printing**
- **Synthesis *in situ***
  - light directed
  - mechanically directed

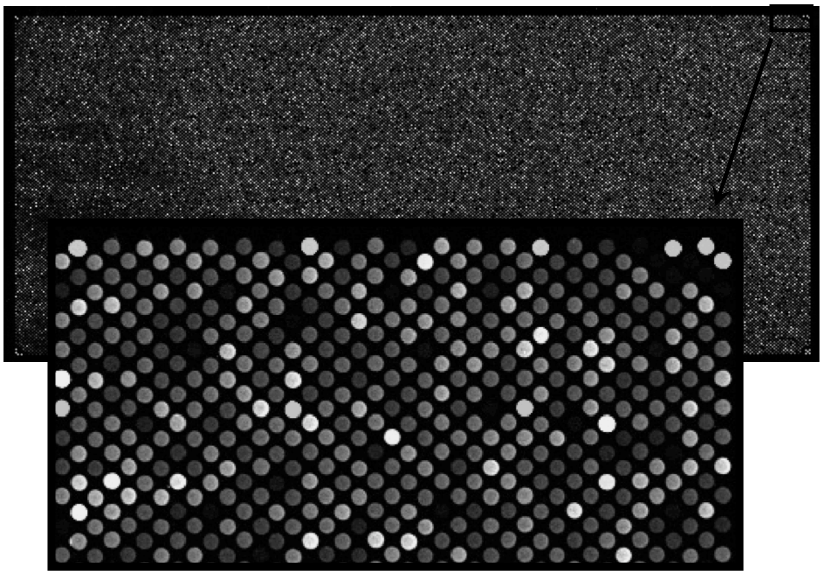
LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

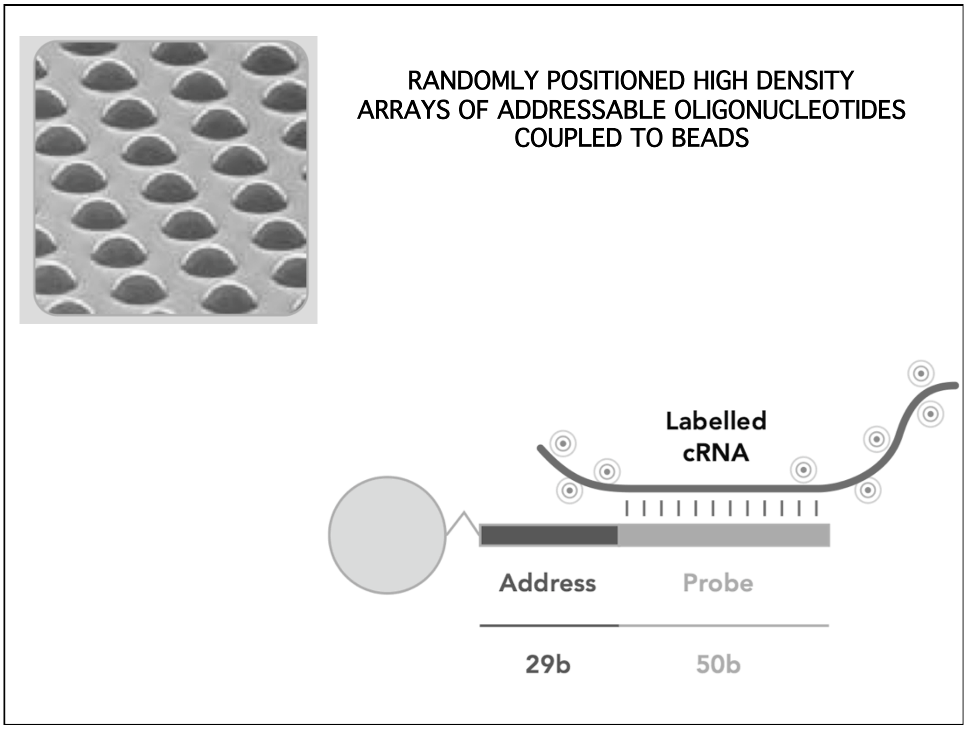


LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

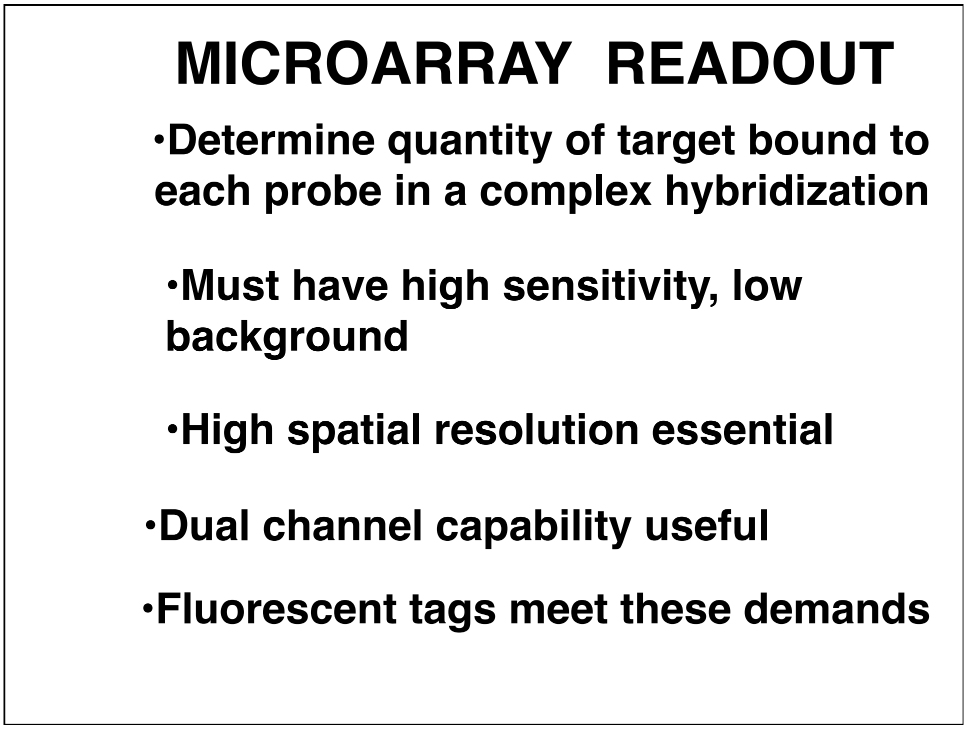


INK JET DIRECTED SYNTHESIS



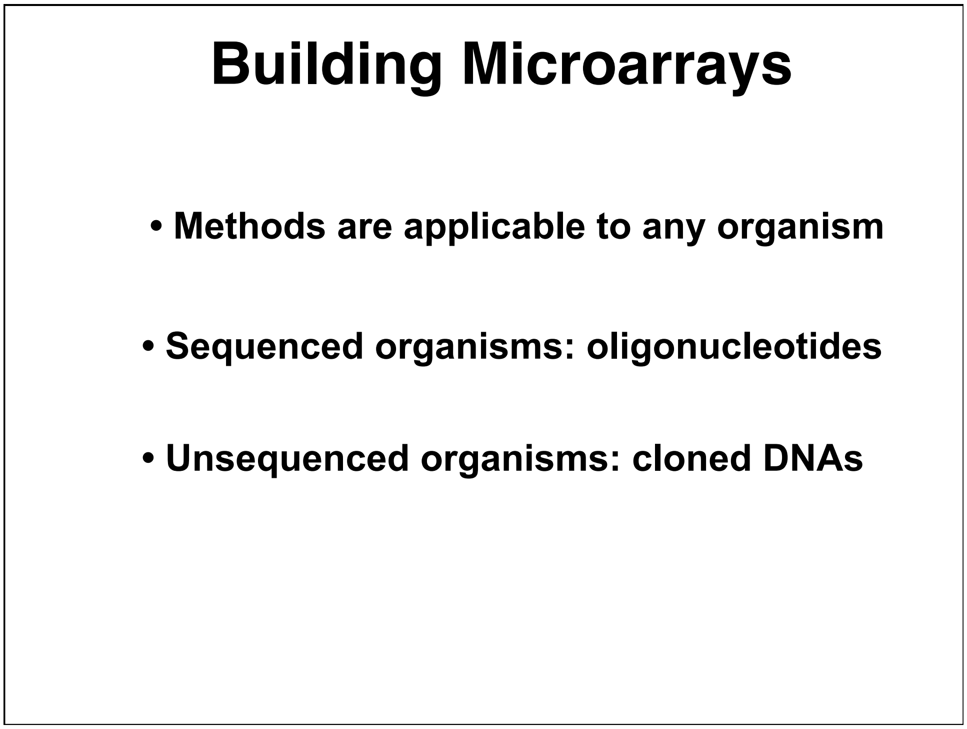


RANDOMLY POSITIONED HIGH DENSITY
ARRAYS OF ADDRESSABLE OLIGONUCLEOTIDES
COUPLED TO BEADS



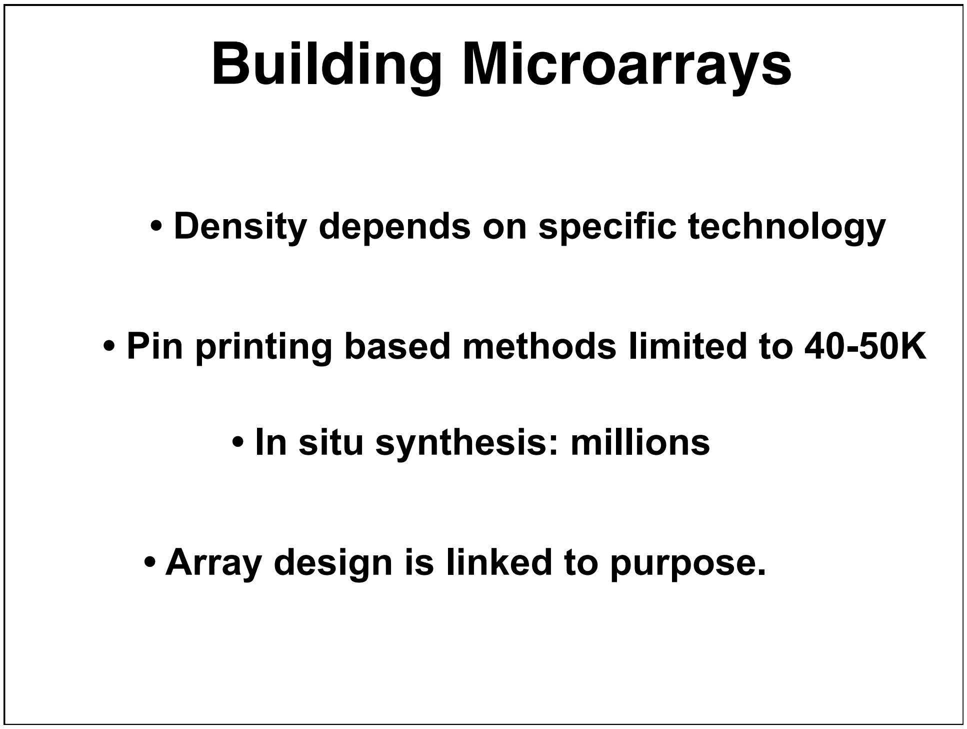
MICROARRAY READOUT

- 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



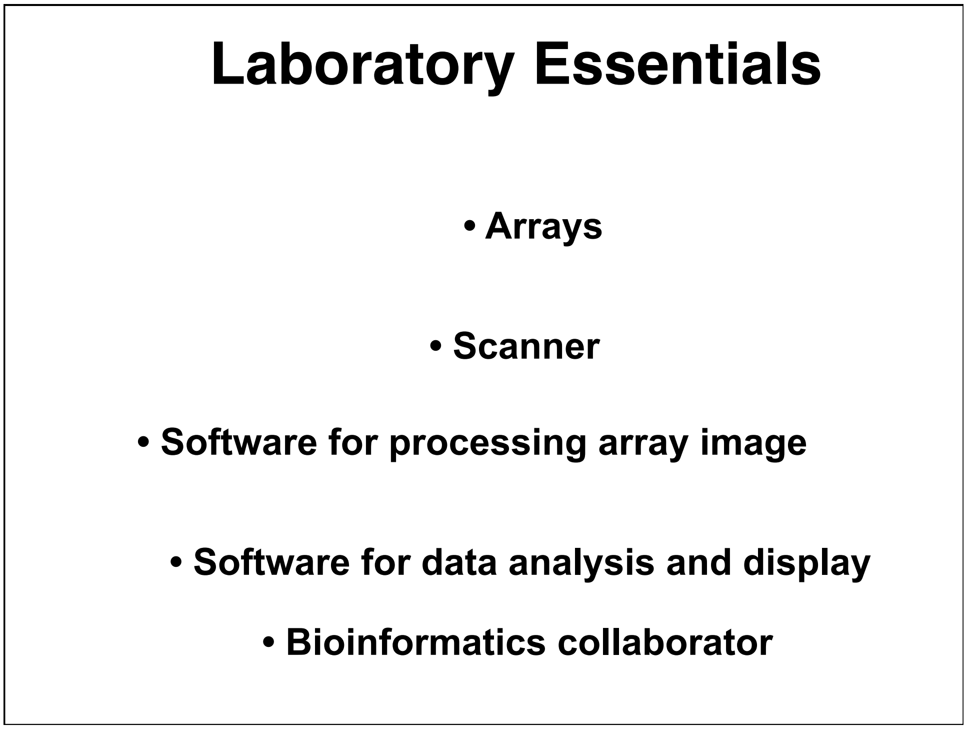
Building Microarrays

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



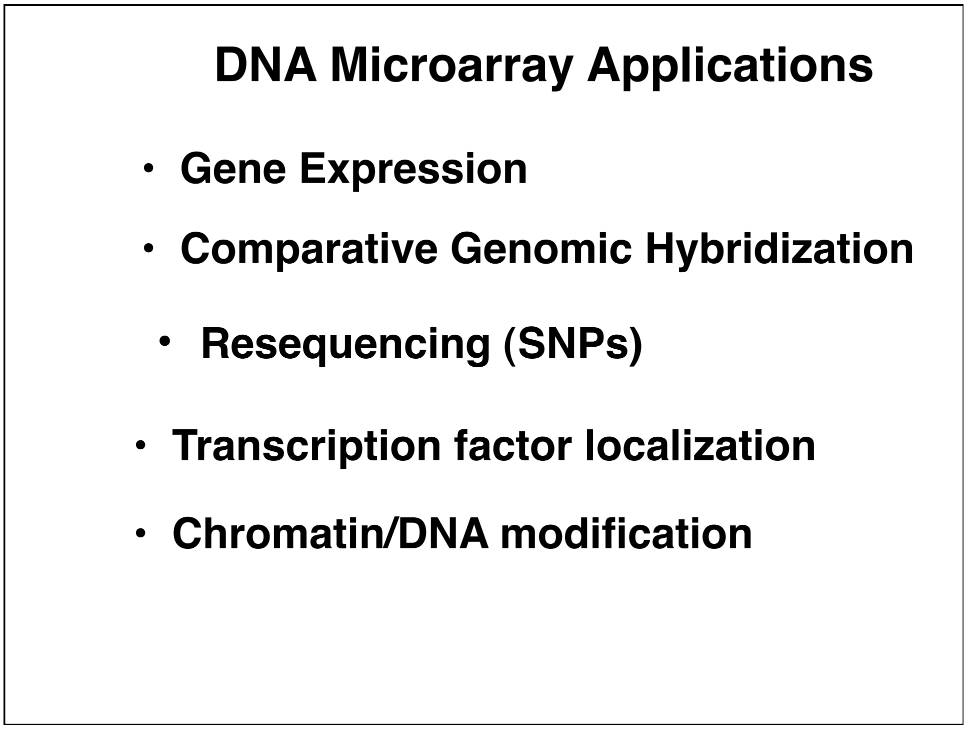
Building Microarrays

- 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

- Gene Expression
- Comparative Genomic Hybridization
- Resequencing (SNPs)
- 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**

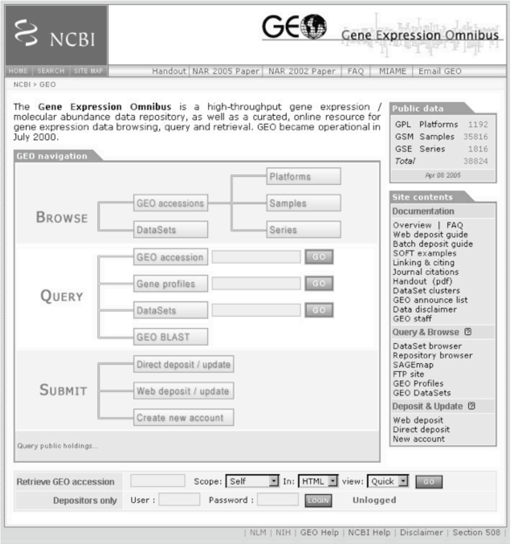
Reports on Microarray Data Quality

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

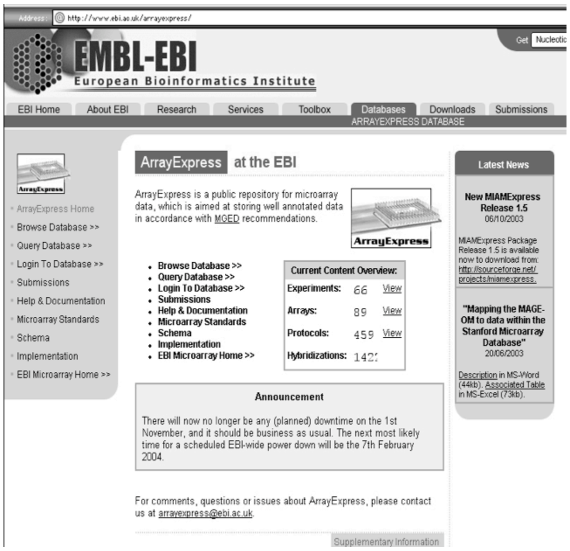
Depositors only User : Password : Unlogged

GEO

Currently contains expression data on 342,783 samples

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

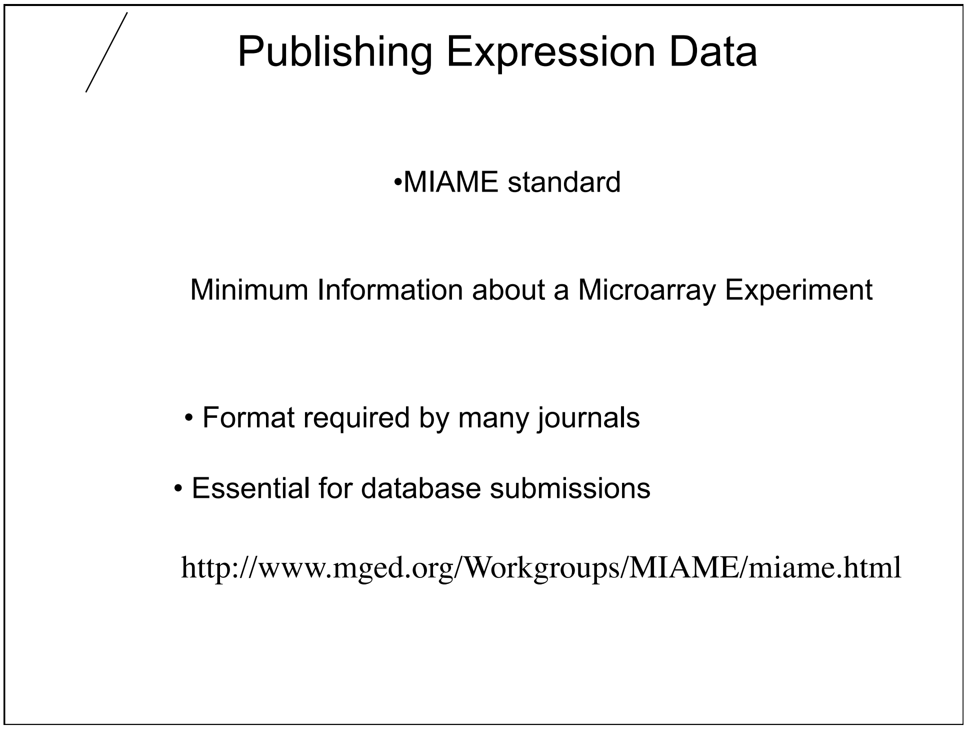
MIAMExpress Package Release 1.5 is available now to download from: http://sourceforge.net/projects/miamexpress.

"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).

288524 assays
Including 34,264
curated, reannotated
Assays

http://www.ebi.ac.uk/microarray-as/ae/



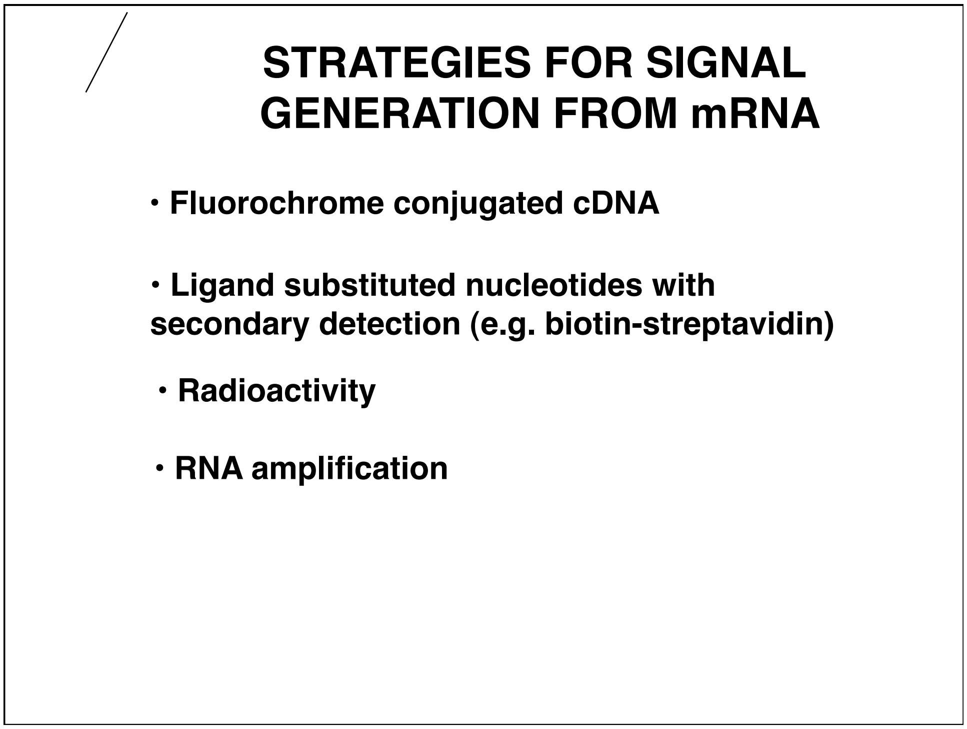
Publishing Expression Data

•MIAME standard

Minimum Information about a Microarray Experiment

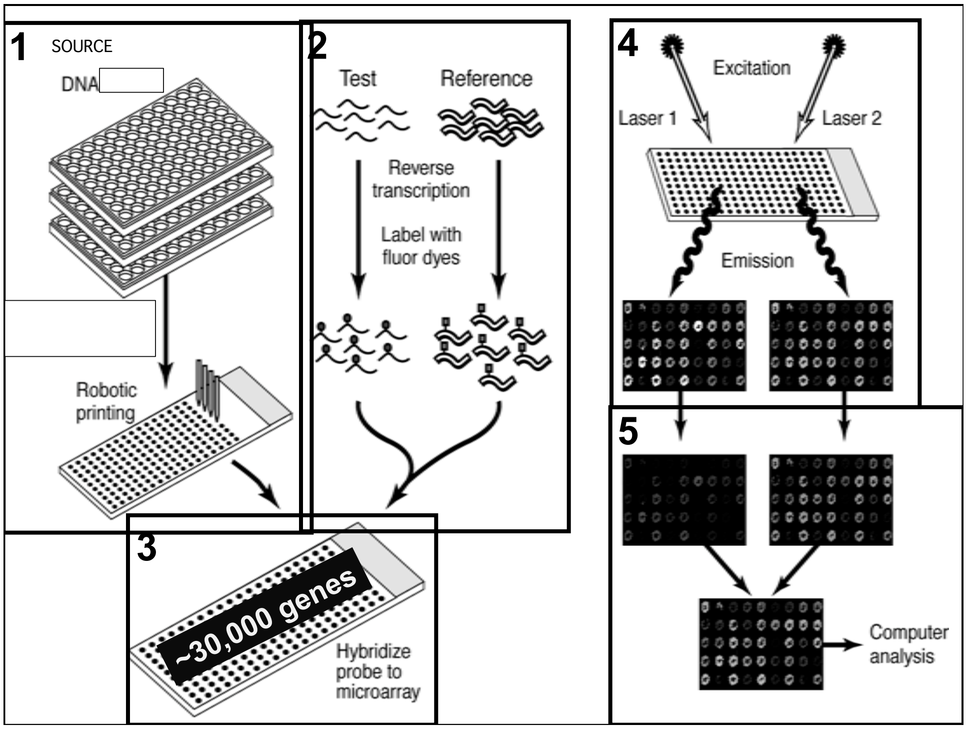
- Format required by many journals
- Essential for database submissions

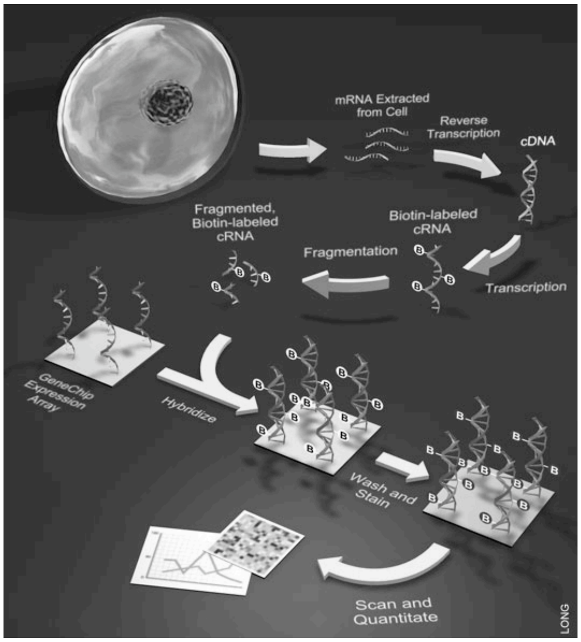
http://www.mged.org/Workgroups/MIAME/miame.html



STRATEGIES FOR SIGNAL GENERATION FROM mRNA

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



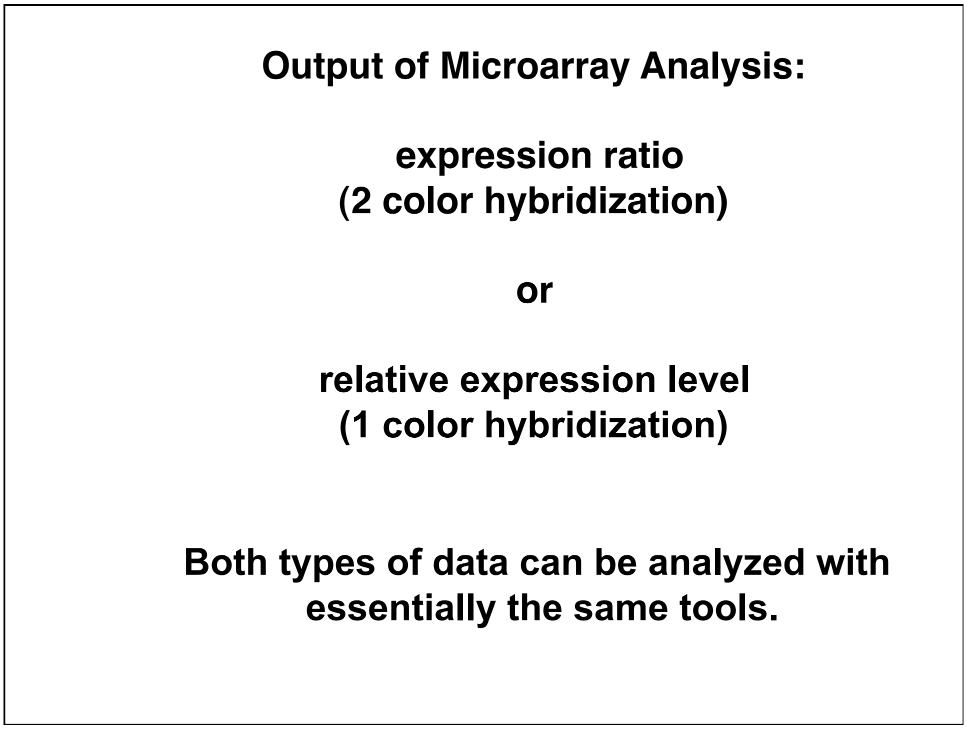


ONE COLOR

HYBRIDIZATION

ON AN OLIGO

ARRAY



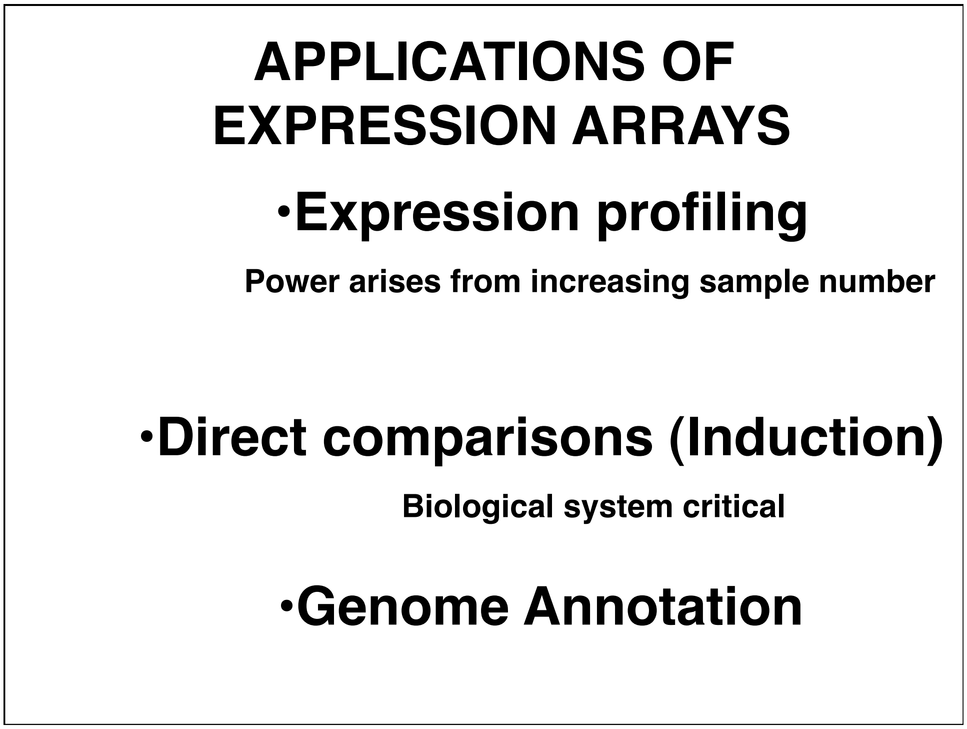
Output of Microarray Analysis:

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

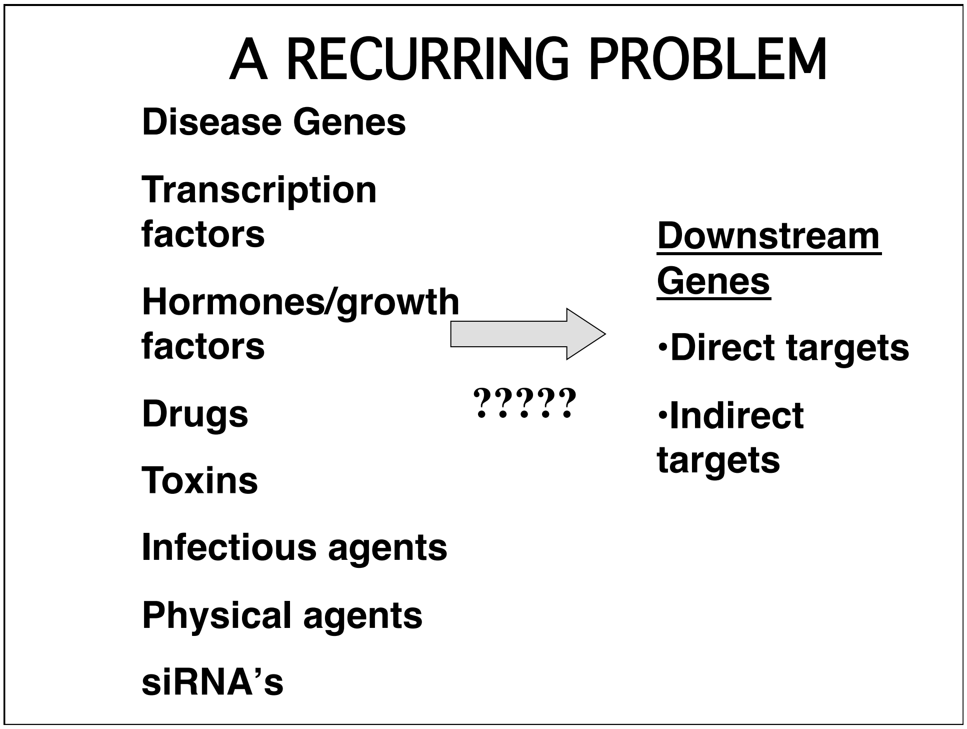
- **Expression profiling**

  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

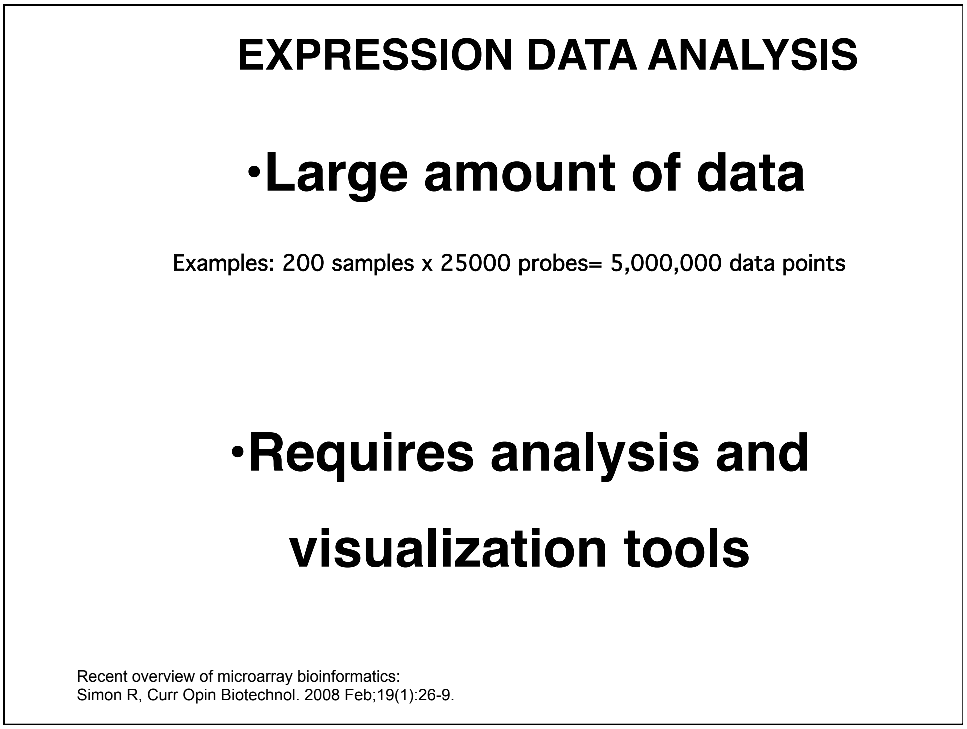
Physical agents

siRNA’s

?????

<u>Downstream Genes</u>

- **Direct targets**
- **Indirect targets**



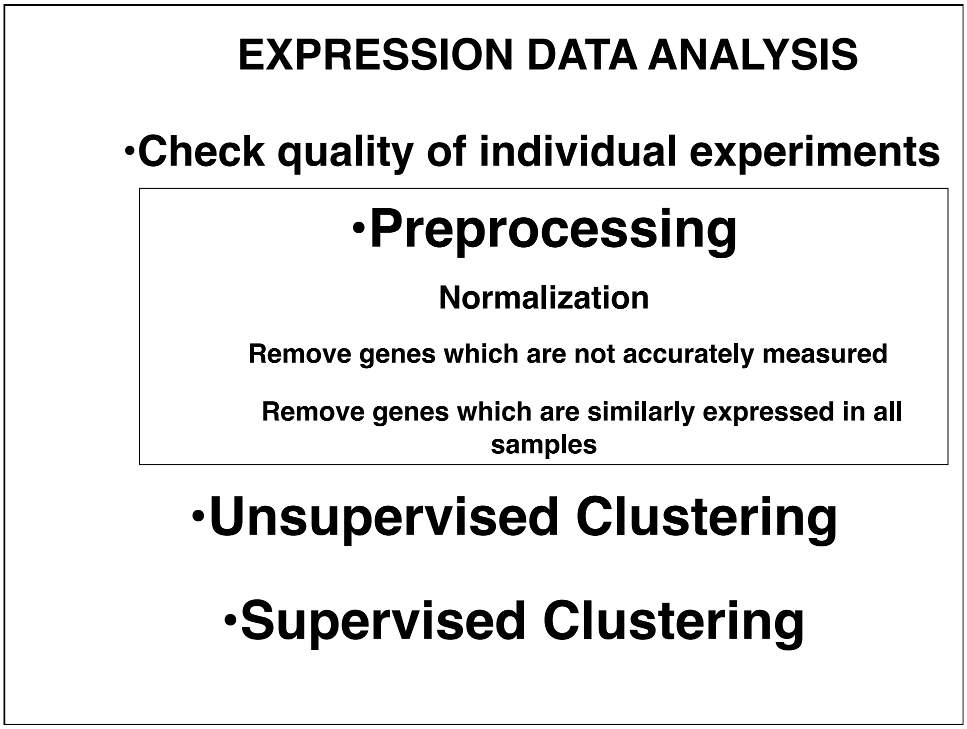
EXPRESSION DATA ANALYSIS

•Large amount of data

Examples: 200 samples x 25000 probes= 5,000,000 data points

•Requires analysis and visualization tools

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



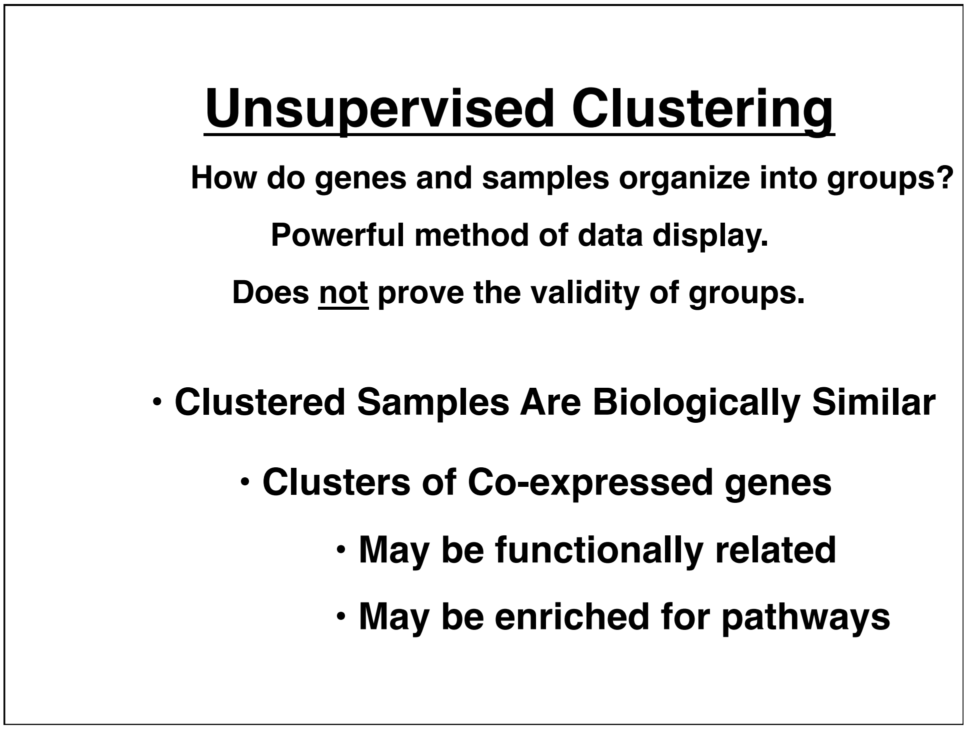
EXPRESSION DATA ANALYSIS

- 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

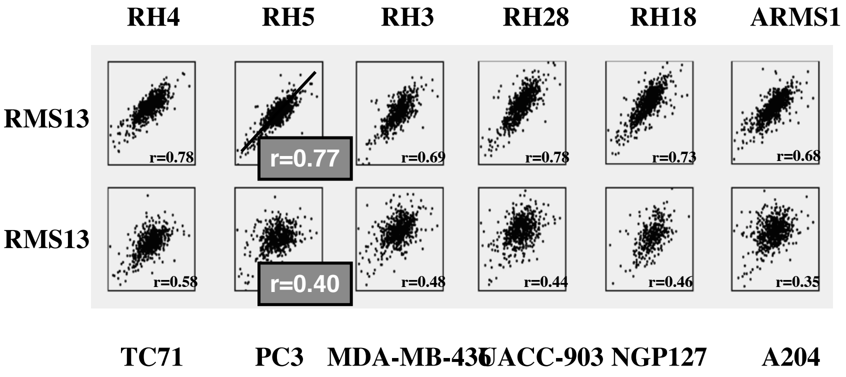
How do genes and samples organize into groups?

Powerful method of data display.

Does <u>not</u> prove the validity of groups.

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

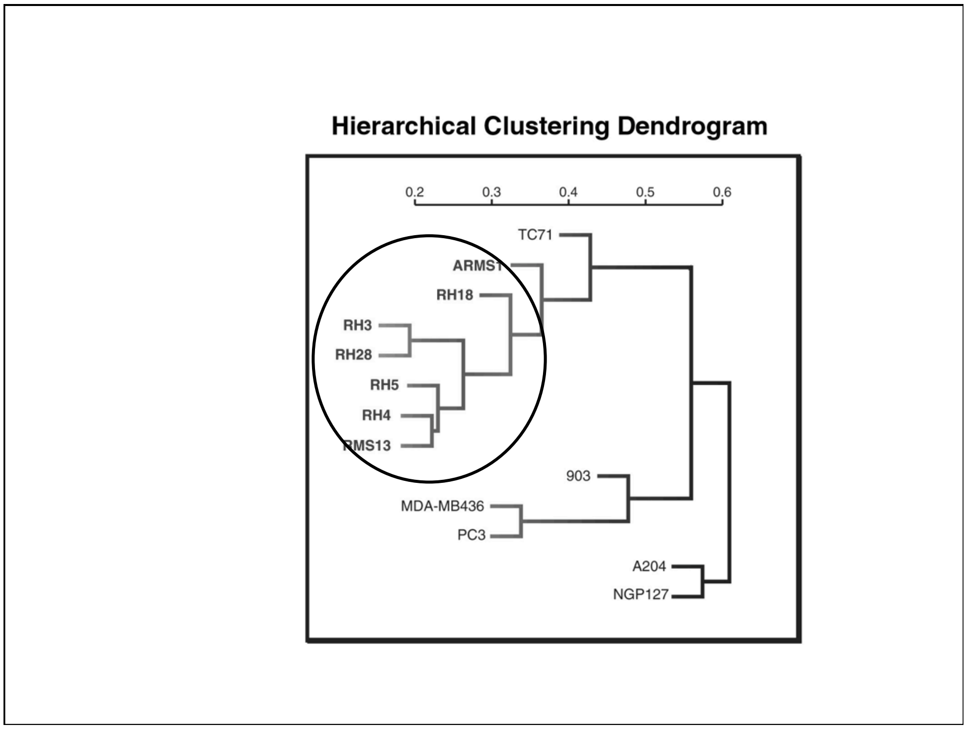
UNSUPERVISED CLUSTERING IS BASED ON A GLOBAL SIMILARITY METRIC

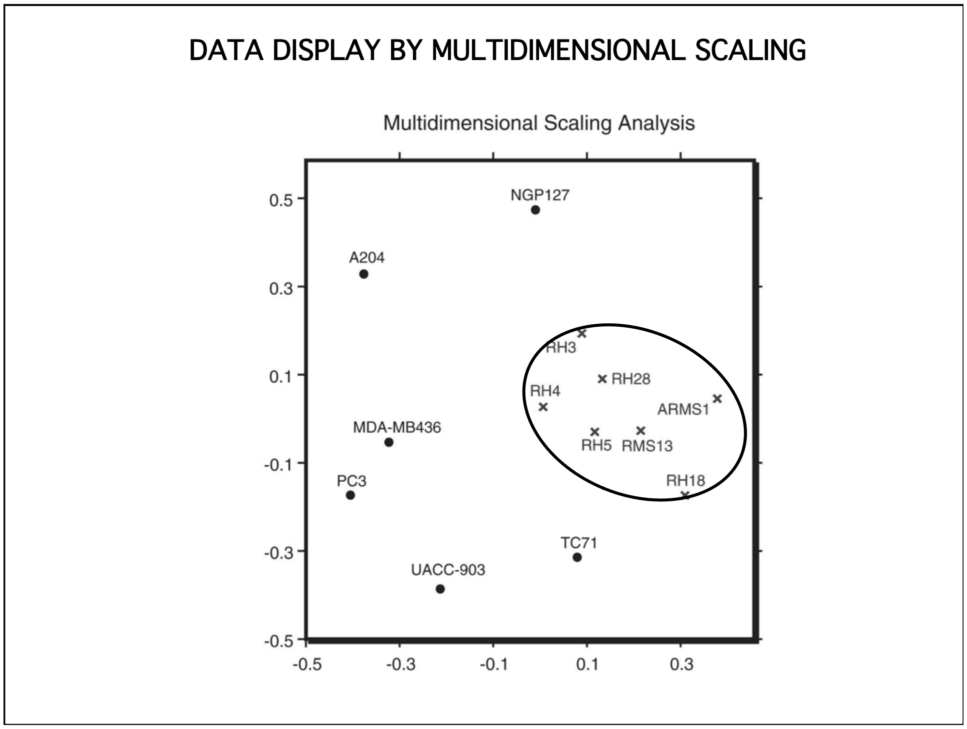


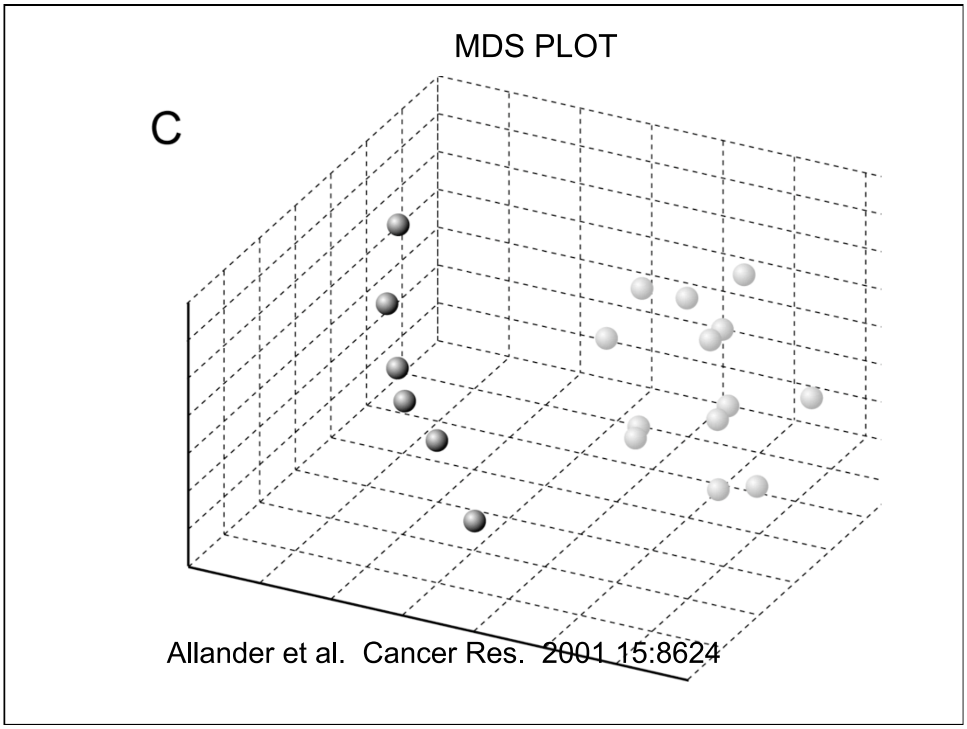
Khan et al. Can Res 58:5009

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

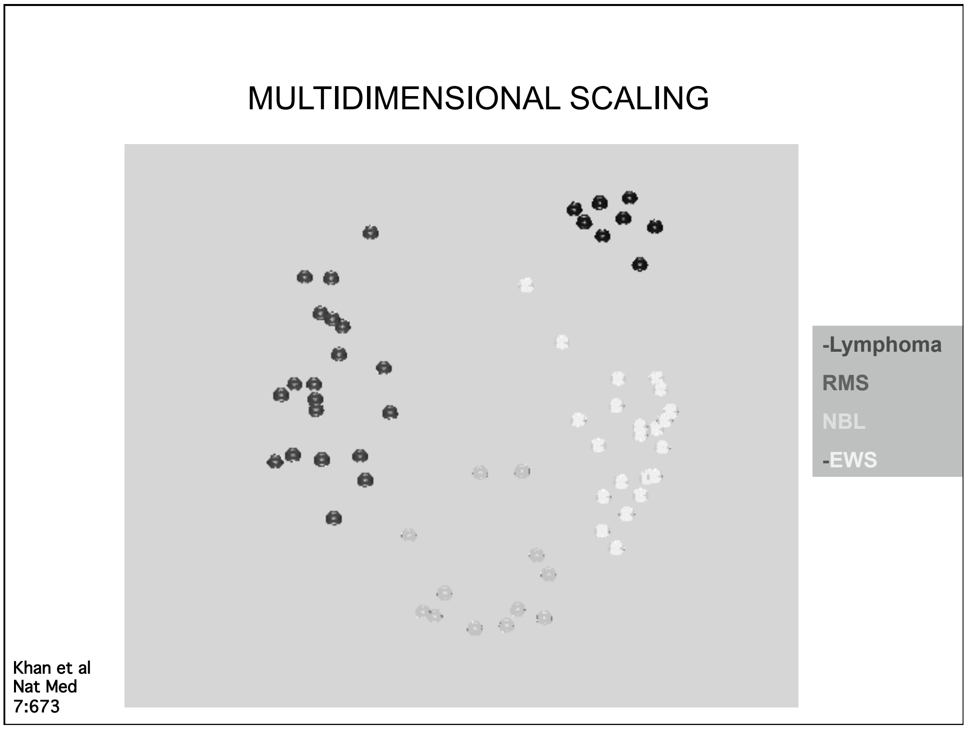



DATA DISPLAY BY MULTIDIMENSIONAL SCALING



MDS PLOT

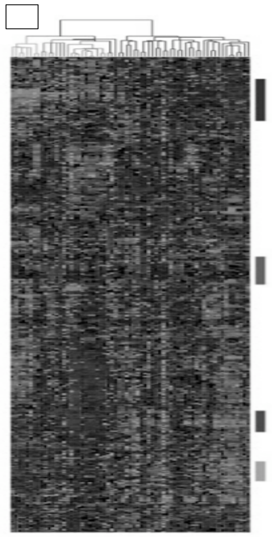
Allander et al. Cancer Res. 2001 15:8624



MULTIDIMENSIONAL SCALING

Khan et al
Nat Med
7:673

CLUSTERING GENES AND SAMPLES



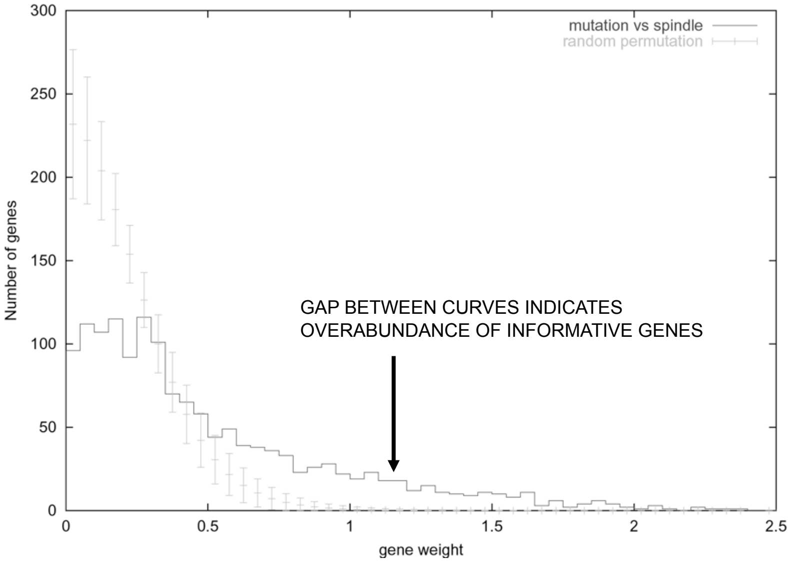
Perou et al. Nature 2000 406:747

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.**

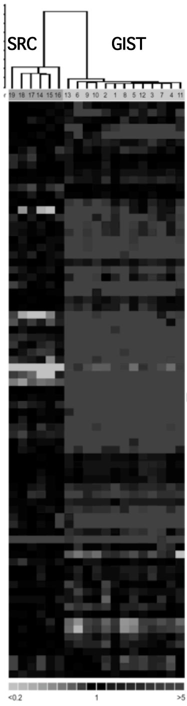
OVERABUNDANCE OF INFORMATIVE GENES DEMONSTRATED BY RANDOM PERMUTATION TEST



GAP BETWEEN CURVES INDICATES
OVERABUNDANCE OF INFORMATIVE GENES

Allander et al. Cancer Res. 2001 15:8624

HIERARCHICAL CLUSTERING OF SAMPLES/GENES DISCRIMINATING GASTROINTESTINAL STROMAL TUMOR USING THE GENES SELECTED BY SUPERVISED ANALYSIS

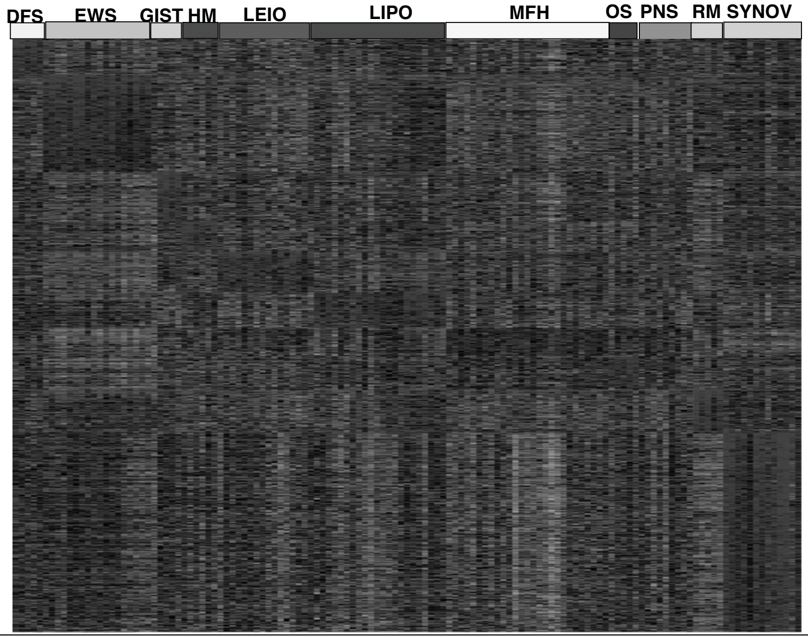


cloneID	Gene Symbol	Weight	α-value
378565	GHR	1	0
795309	SOD3	1.4	0
486641	MOCS2	1.5	0.00003
629896	MAP1B	1.8	0.00008
67769	KCNK3	1.7	0.00003
40304	SDHA	1.6	0.00006
2800976	CDX1	1.6	0
2556956	GOT2	1.1	0.0001
841370	GOT2	1.2	0.00003
42059	DLD	1.9	0.00001
41940	CXORF1	1.5	0.00006
878815	ARF3	1.4	0.00004
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

KIT, THE MUTATIONAL AND THERAPEUTIC TARGET IS HIGHLY SIGNIFICANT.

Allander et al. Cancer Res. 2001 15:8624

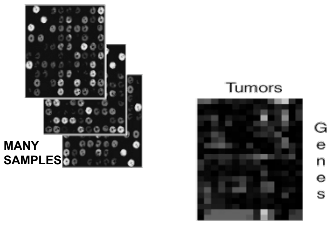
CHARACTERISTIC PATTERNS OF GENE EXPRESSION IN DIFFERENT SARCOMAS



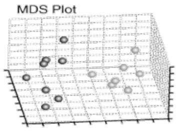
BAIRD ET AL.
CANCER RES
2005

GENOMICS FROM BENCH TO BEDSIDE

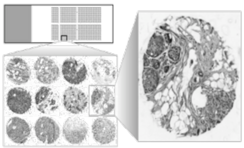
WHOLE GENOME

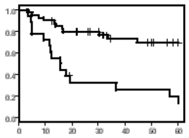


GENE SELECTION



GENE VALIDATION



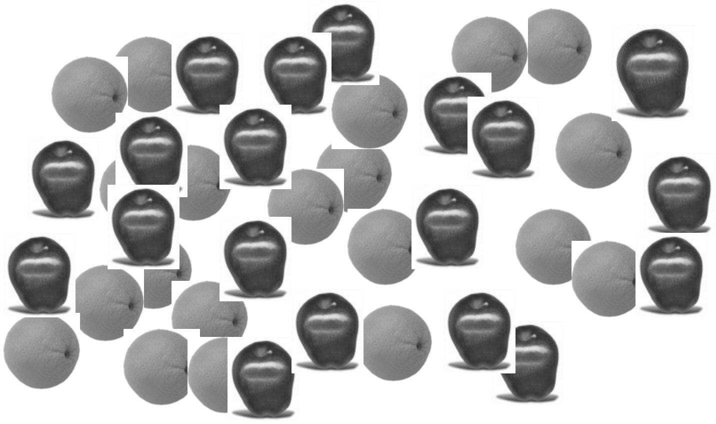


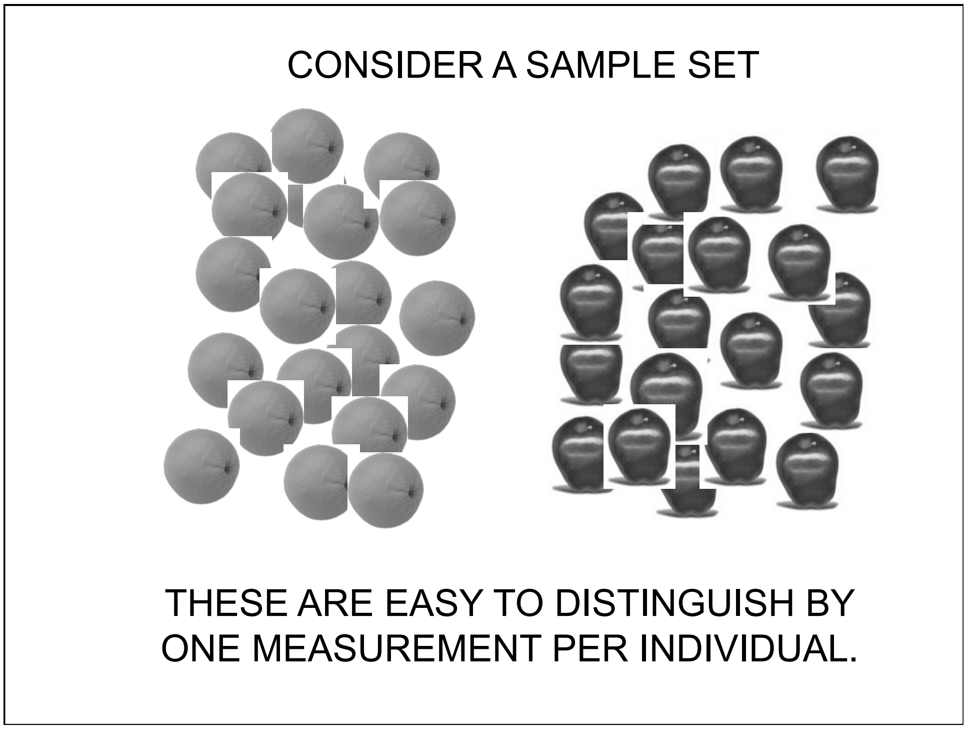
ASSAY DEVELOPMENT

SIGNAL STRENGTH VARIES IN TISSUE PROFILING EXPERIMENTS

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

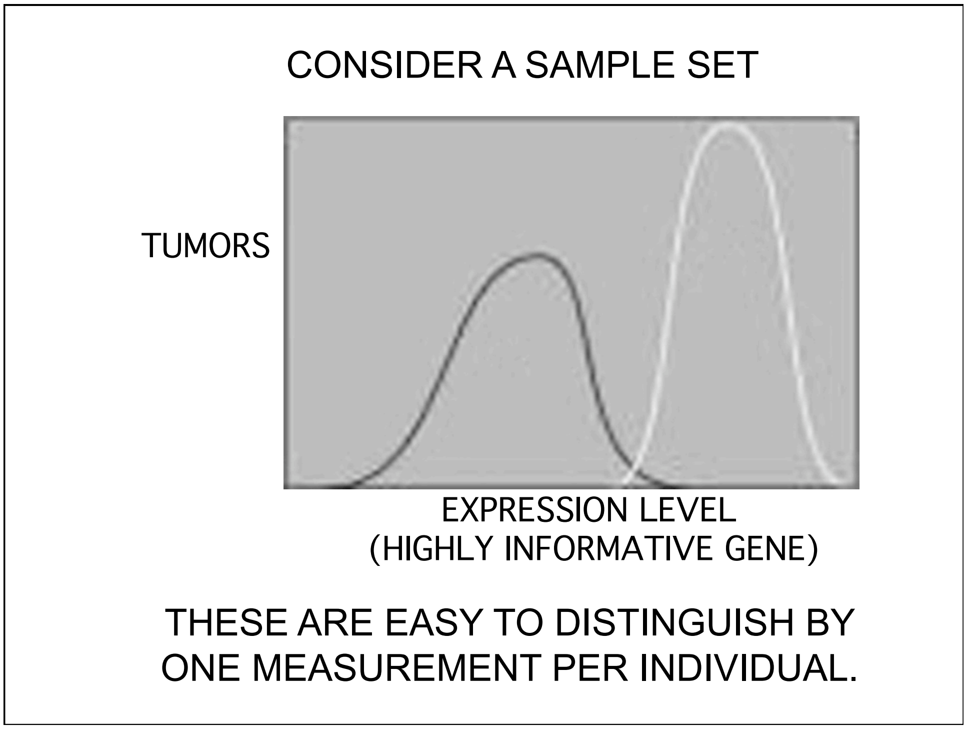
CONSIDER A SAMPLE SET





CONSIDER A SAMPLE SET

THESE ARE EASY TO DISTINGUISH BY ONE MEASUREMENT PER INDIVIDUAL.

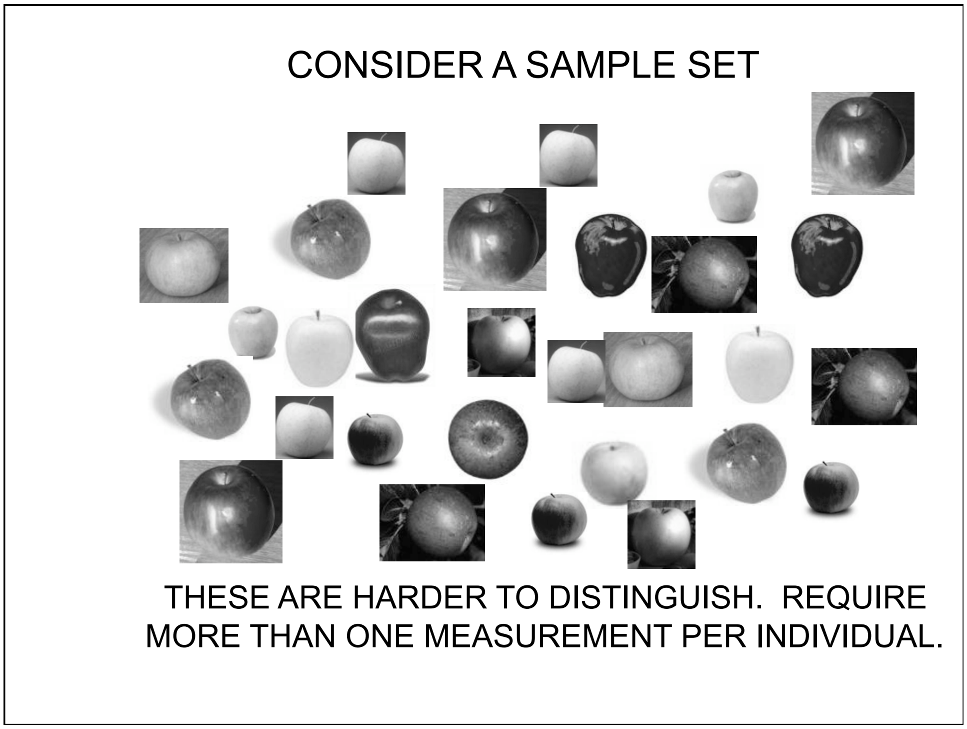


CONSIDER A SAMPLE SET

TUMORS

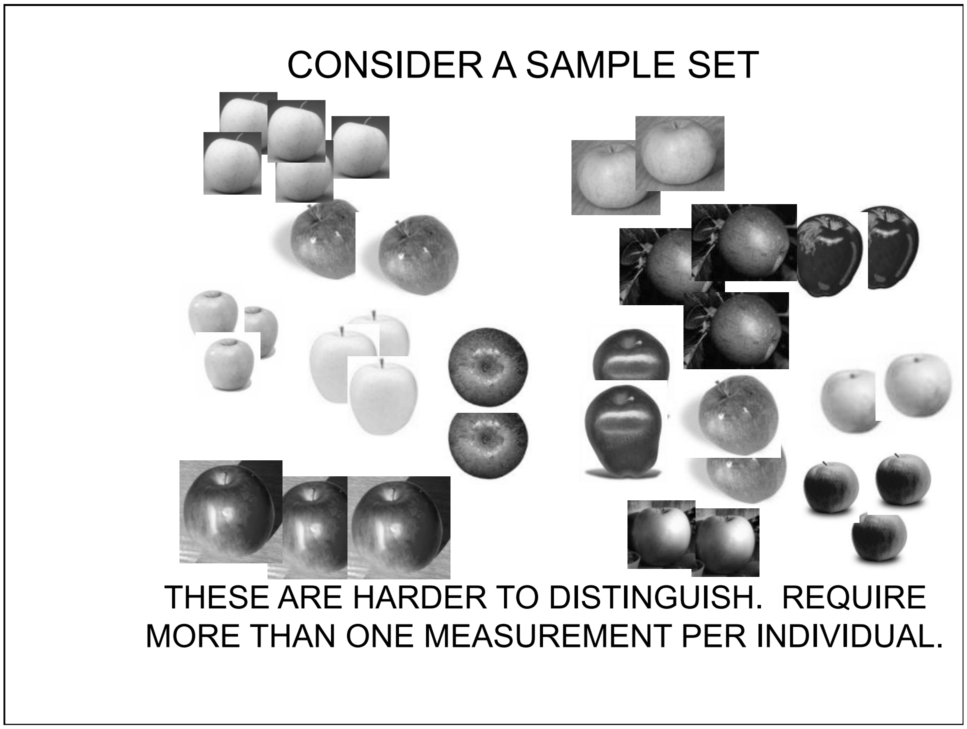
EXPRESSION LEVEL
(HIGHLY INFORMATIVE GENE)

THESE ARE EASY TO DISTINGUISH BY ONE MEASUREMENT PER INDIVIDUAL.



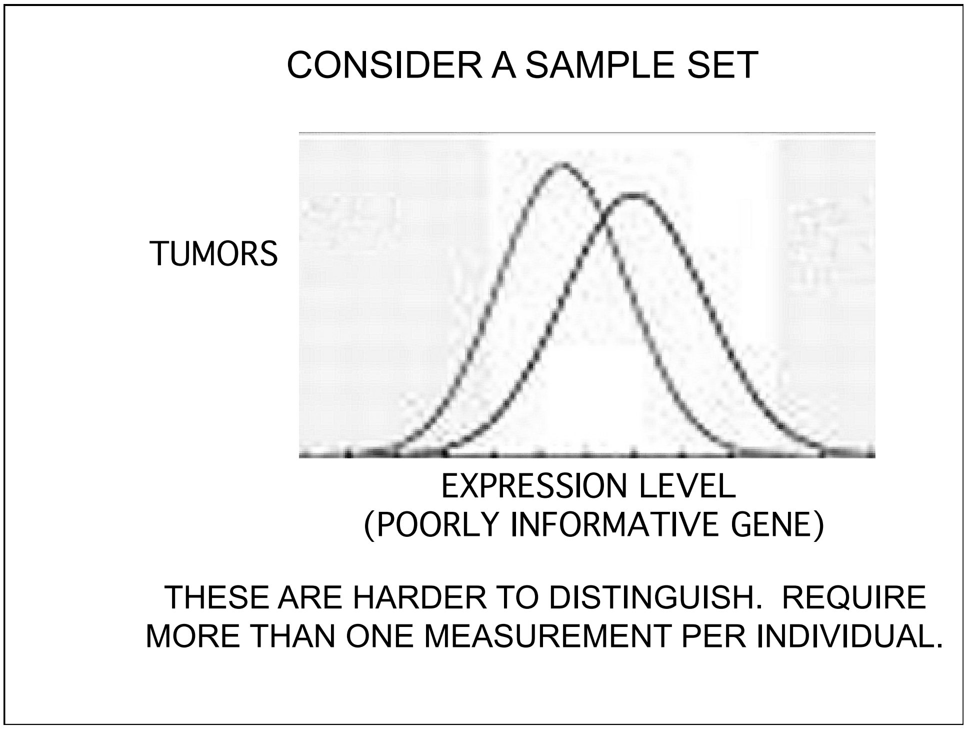
CONSIDER A SAMPLE SET

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

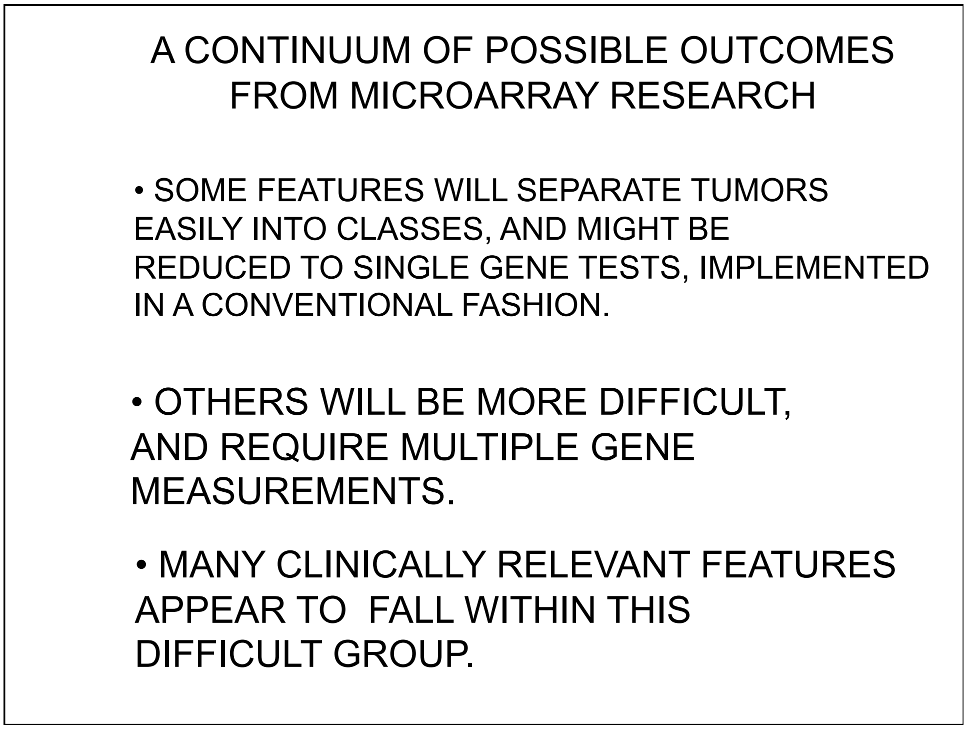
TUMORS

EXPRESSION LEVEL
(POORLY INFORMATIVE GENE)

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?

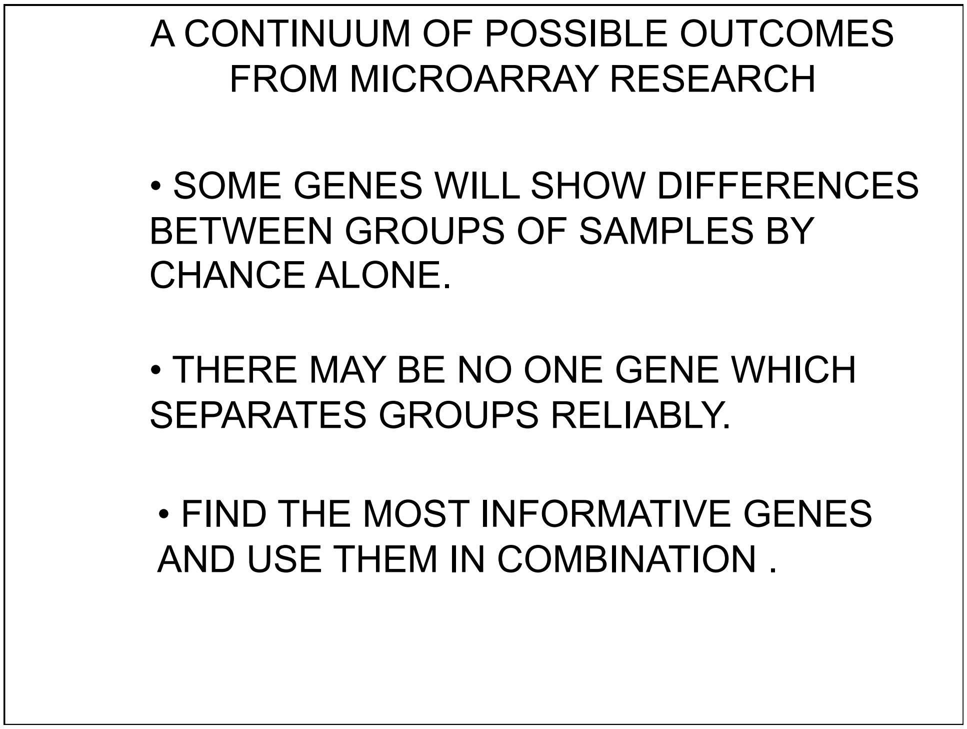


A CONTINUUM OF POSSIBLE OUTCOMES FROM MICROARRAY RESEARCH

- 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.

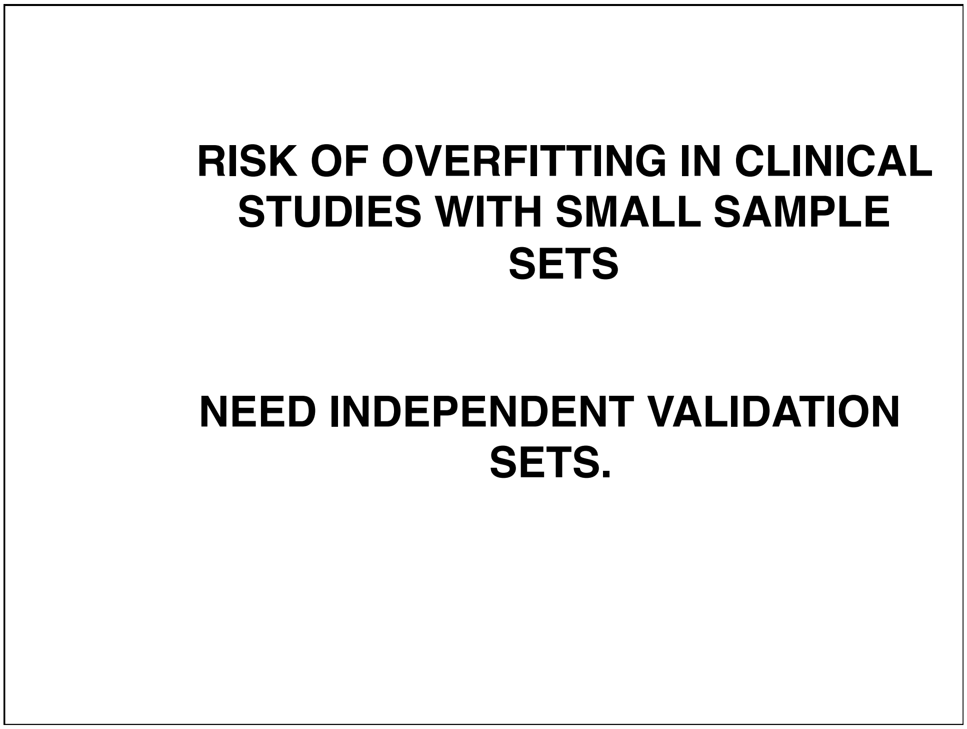


A CONTINUUM OF POSSIBLE OUTCOMES FROM MICROARRAY RESEARCH

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

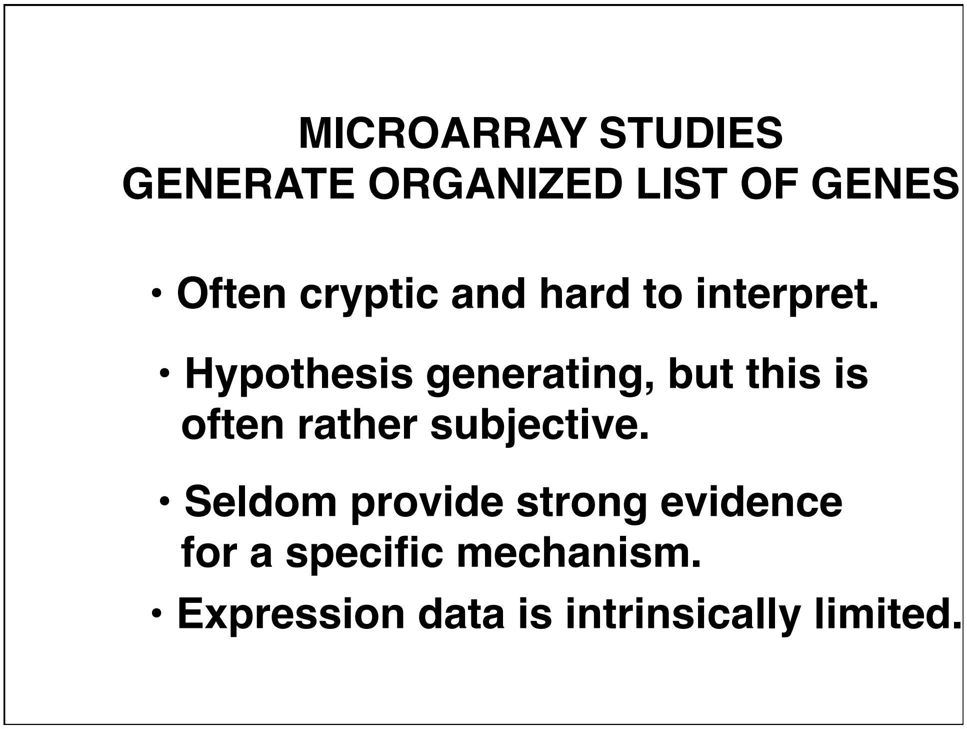
- THERE MAY BE NO ONE GENE WHICH SEPARATES GROUPS RELIABLY.

- 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

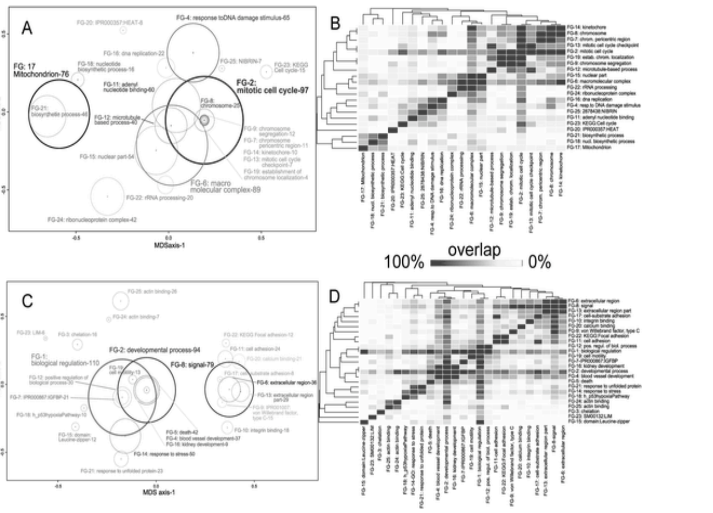
- Often cryptic and hard to interpret.
- Hypothesis generating, but this is often rather subjective.
- Seldom provide strong evidence for a specific mechanism.
- Expression data is intrinsically limited.

GETTING BEYOND GENE LISTS

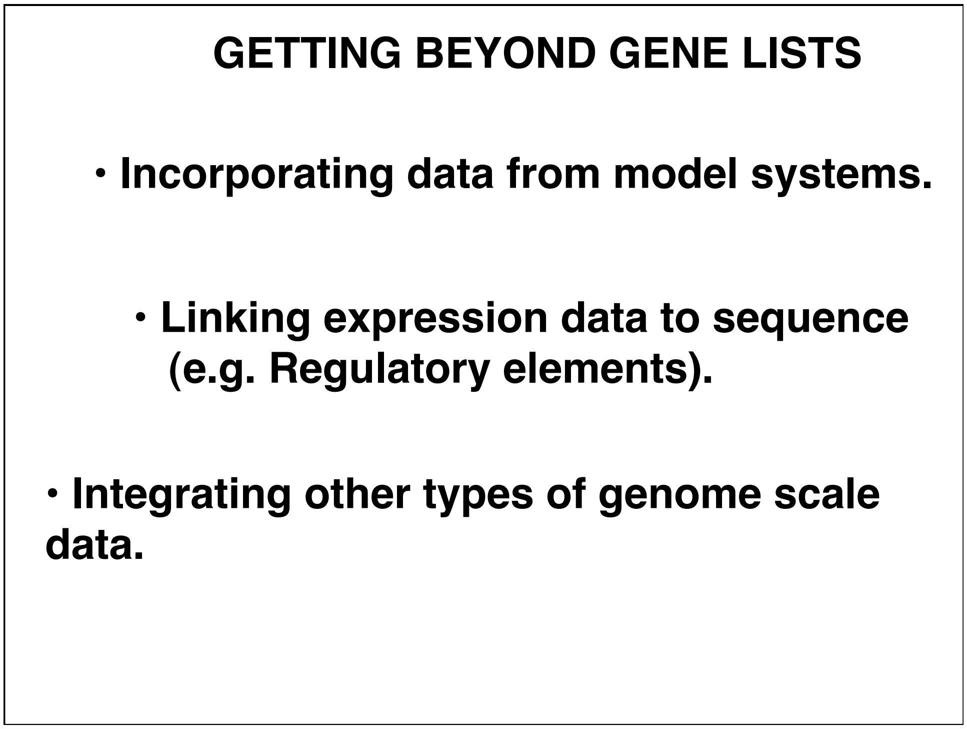
- **Optimal use of gene annotations.**
  - **Gene Ontology** (http://david.abcc.ncifcrf.gov/)
- **Optimizing use of public data.**
  - **GEO, ARRAY EXPRESS, ACADEMIC DATA**
  - **GENE SIGNATURE BASED METHODS (Gene Set Enrichment Analysis).**

GENE ONTOLOGY AND PROMOTER DATABASES CAN HELP FIND BIOLOGY

GENE ONTOLOGY CATEFORIES AFFECTED BY ONCOGENE KNOCKDOWN IN EWING'S SARCOMA

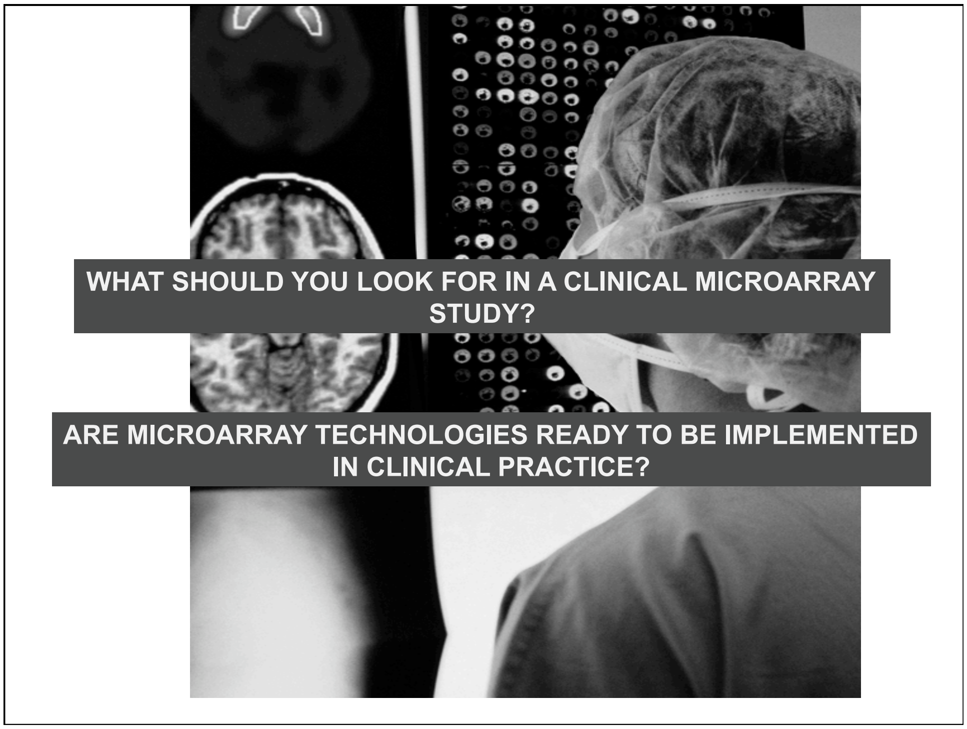


KAUER ET AL. PLOS ONE 4:e5415 2009



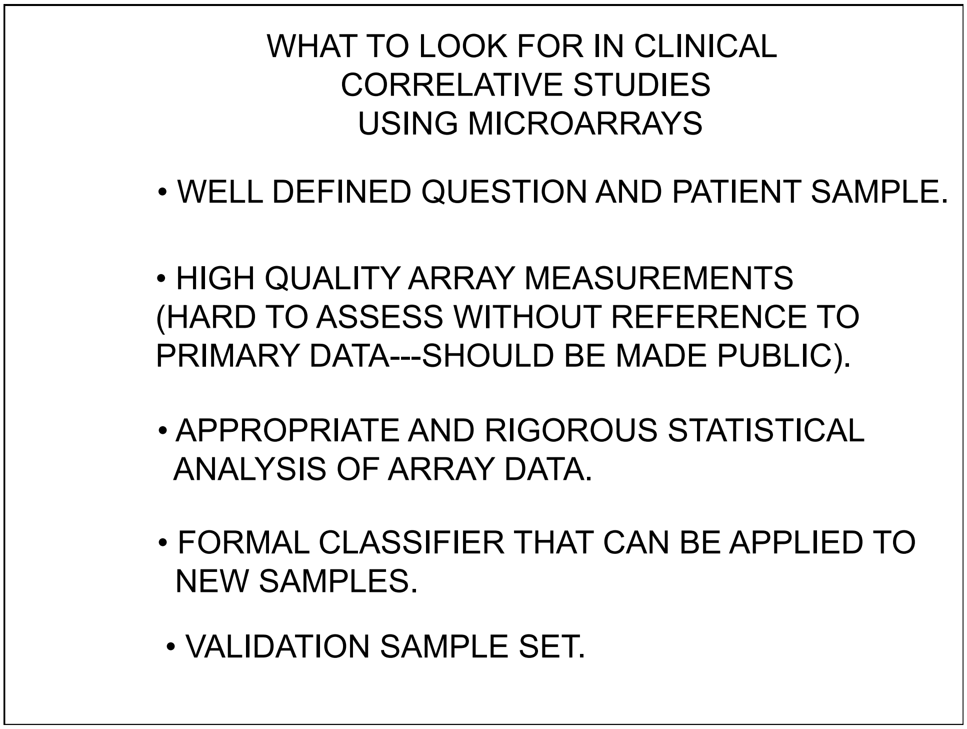
GETTING BEYOND GENE LISTS

- **Incorporating data from model systems.**
- **Linking expression data to sequence (e.g. Regulatory elements).**
- **Integrating other types of genome scale data.**



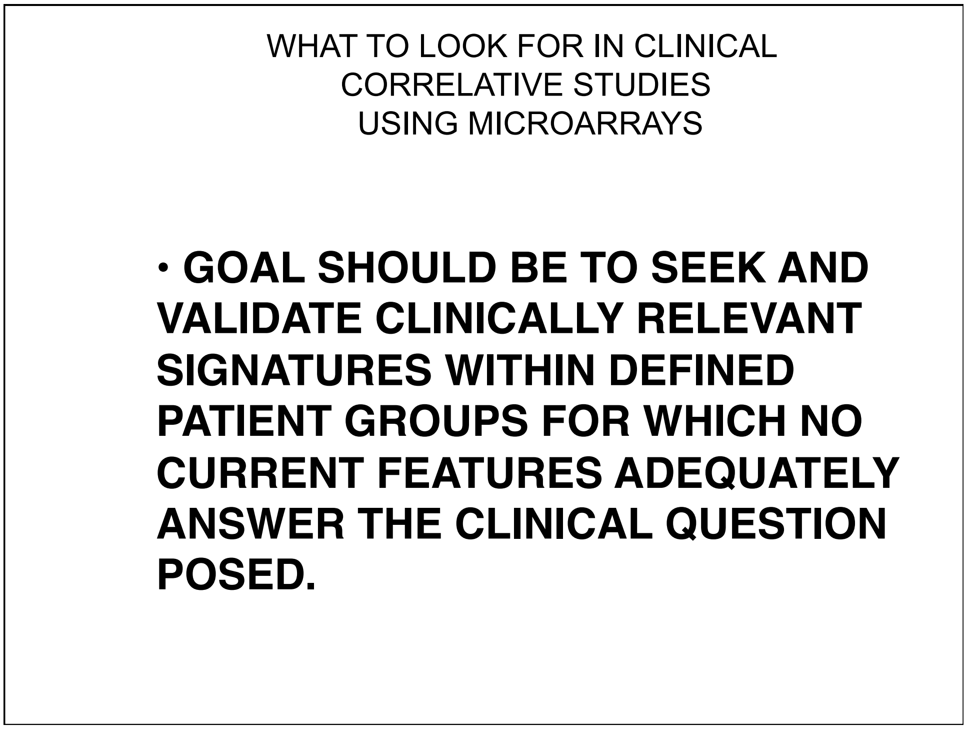
WHAT SHOULD YOU LOOK FOR IN A CLINICAL MICROARRAY STUDY?

ARE MICROARRAY TECHNOLOGIES READY TO BE IMPLEMENTED IN CLINICAL PRACTICE?



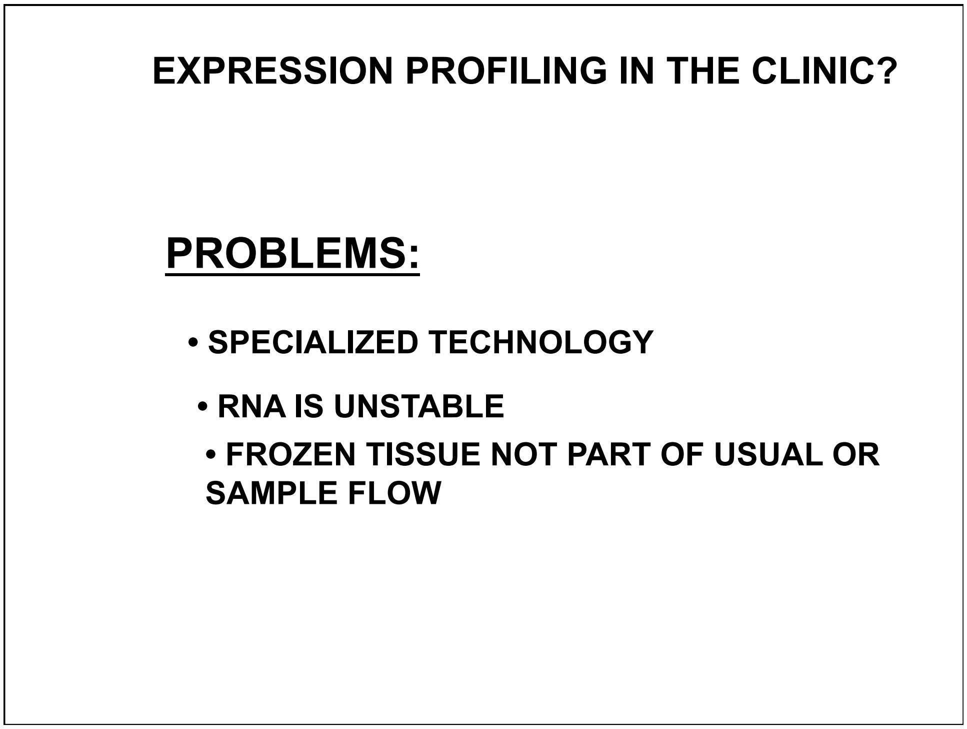
WHAT TO LOOK FOR IN CLINICAL
CORRELATIVE STUDIES
USING MICROARRAYS

- 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

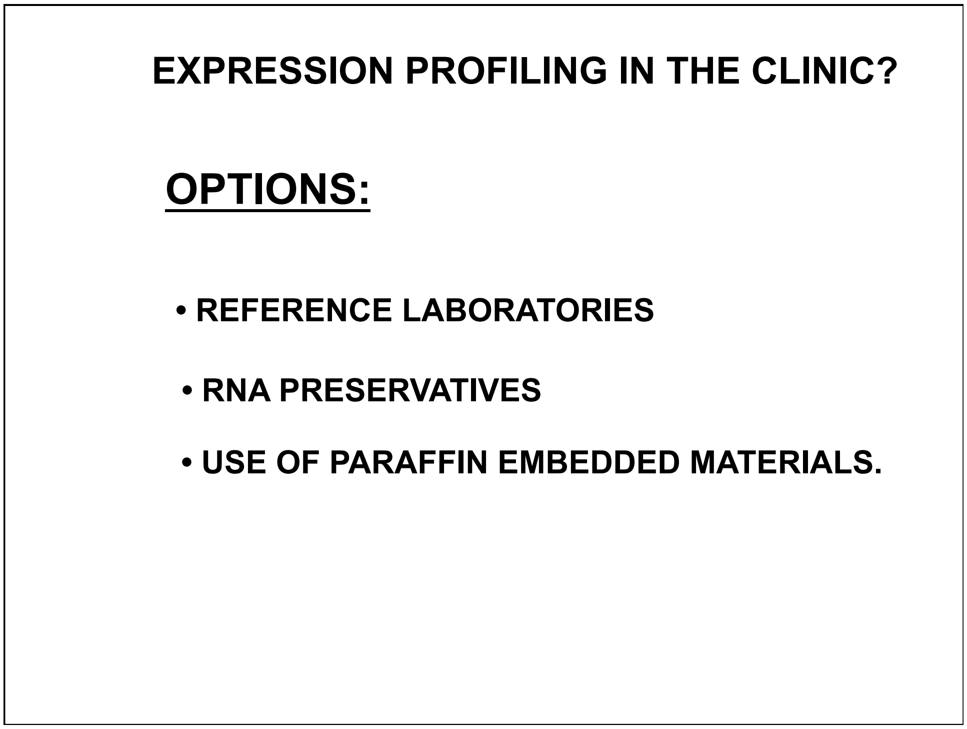
- **GOAL SHOULD BE TO SEEK AND VALIDATE CLINICALLY RELEVANT SIGNATURES WITHIN DEFINED PATIENT GROUPS FOR WHICH NO CURRENT FEATURES ADEQUATELY ANSWER THE CLINICAL QUESTION POSED.**



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:

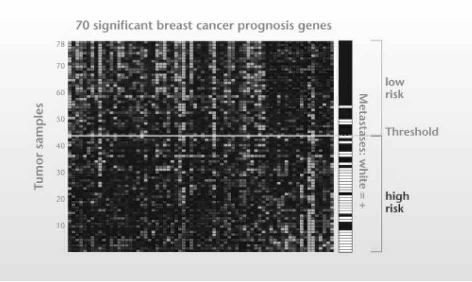
- REFERENCE LABORATORIES
- RNA PRESERVATIVES
- USE OF PARAFFIN EMBEDDED MATERIALS.

EXPRESSION PROFILING IN THE CLINIC?

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

FDA APPROVED TESTS FOR BREAST CANCER BASED ON EXPRESSION STUDIES

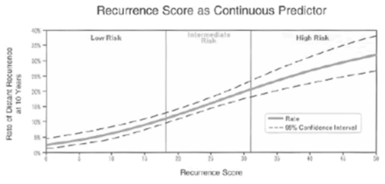
70 GENE MICROARRAY SIGNATURE



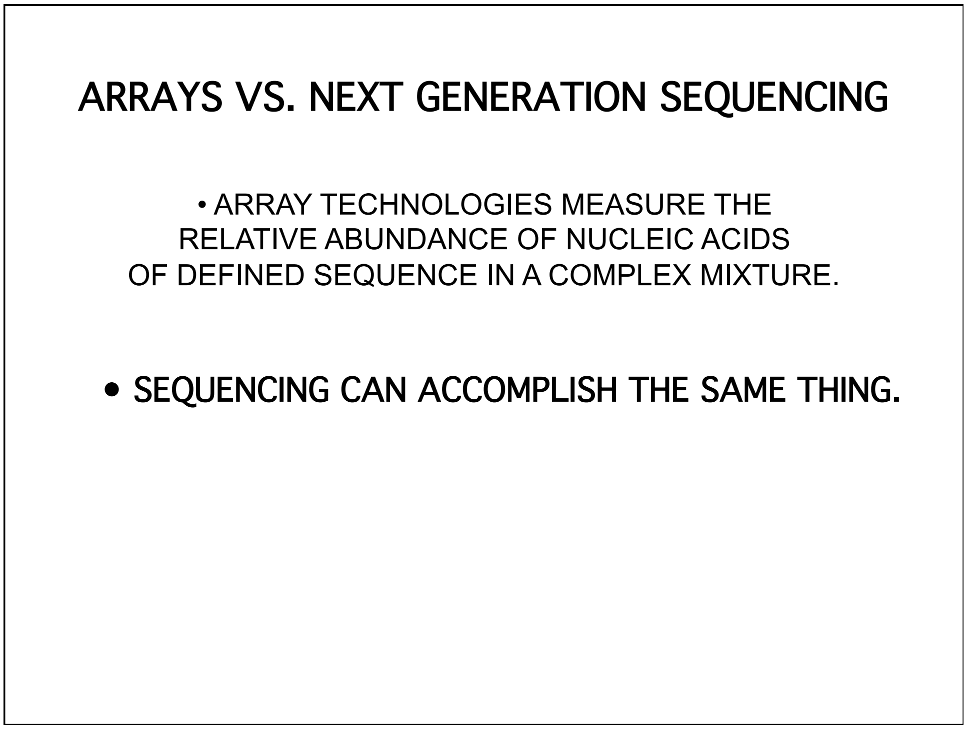
Van de Vijver et al

NEJM 347:1999 .

Muiltgene RT-PCR SIgnature



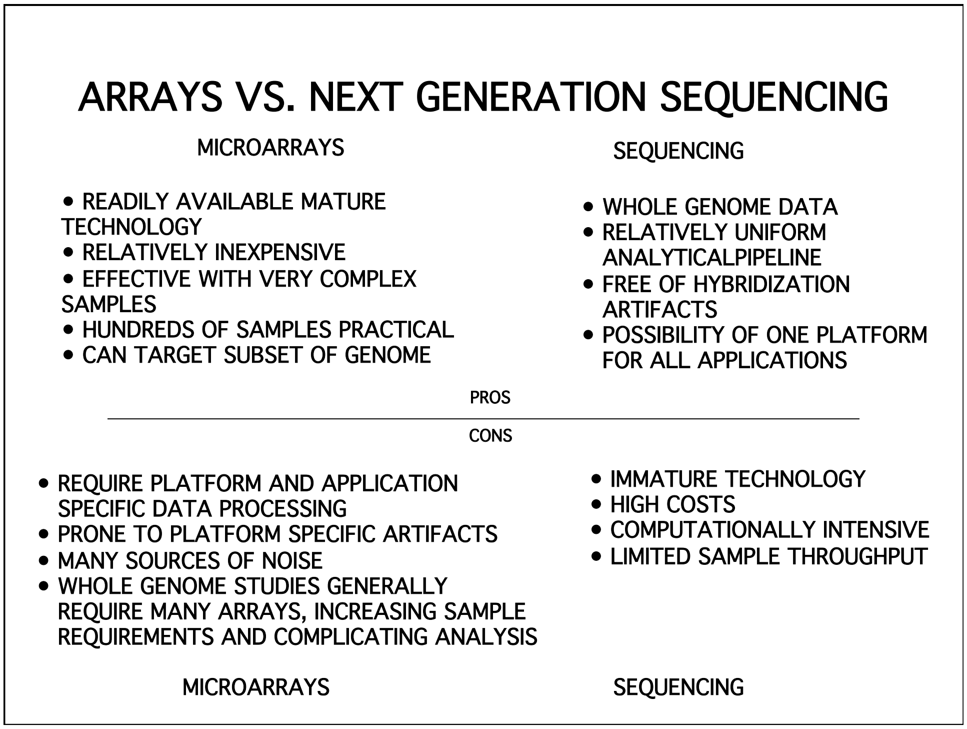
Paik et al NEJM 351:2817



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
- RELATIVELY UNIFORM ANALYTICALPIPELINE
- FREE OF HYBRIDIZATION ARTIFACTS
- POSSIBILITY OF ONE PLATFORM FOR ALL APPLICATIONS

PROS

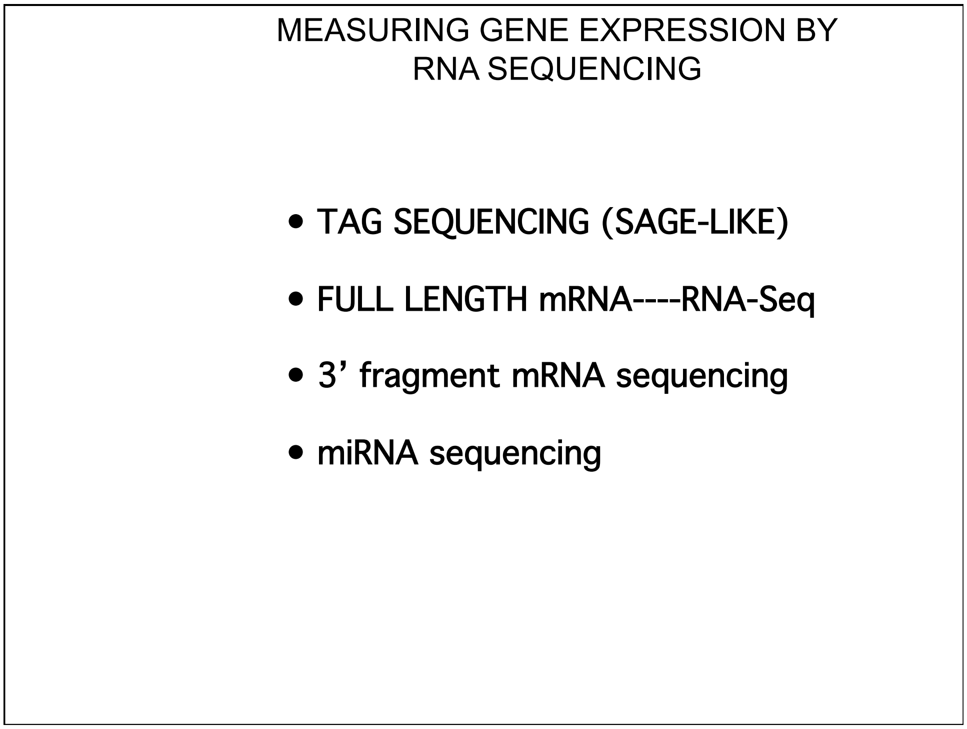
CONS

- 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

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

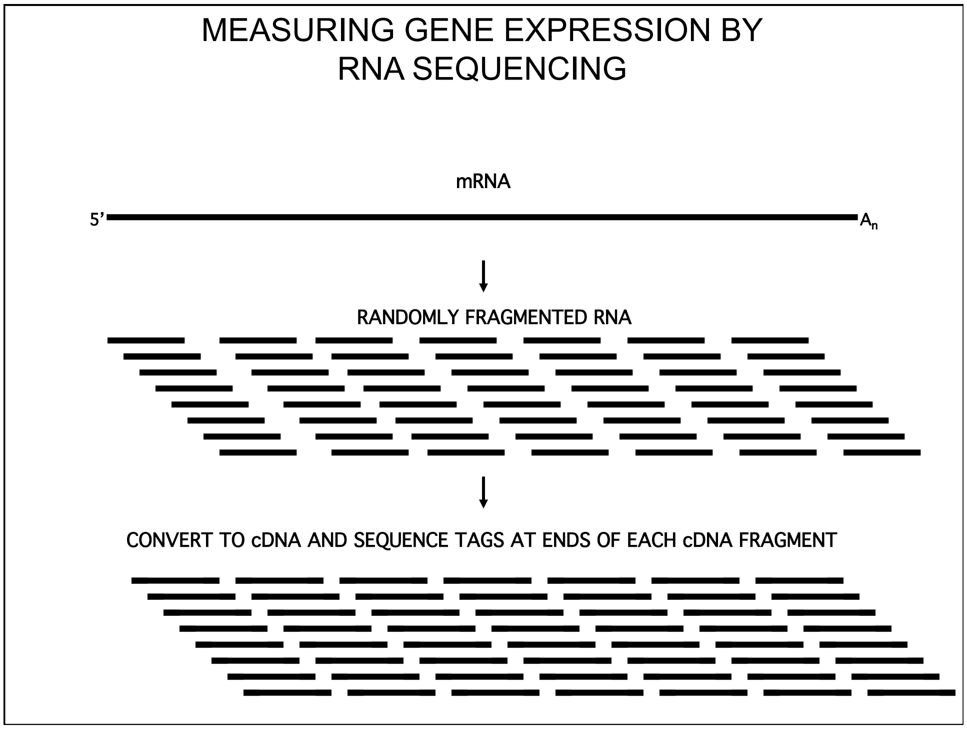
MICROARRAYS

SEQUENCING



MEASURING GENE EXPRESSION BY RNA SEQUENCING

- TAG SEQUENCING (SAGE-LIKE)
- FULL LENGTH mRNA----RNA-Seq
- 3’ fragment mRNA sequencing
- miRNA sequencing

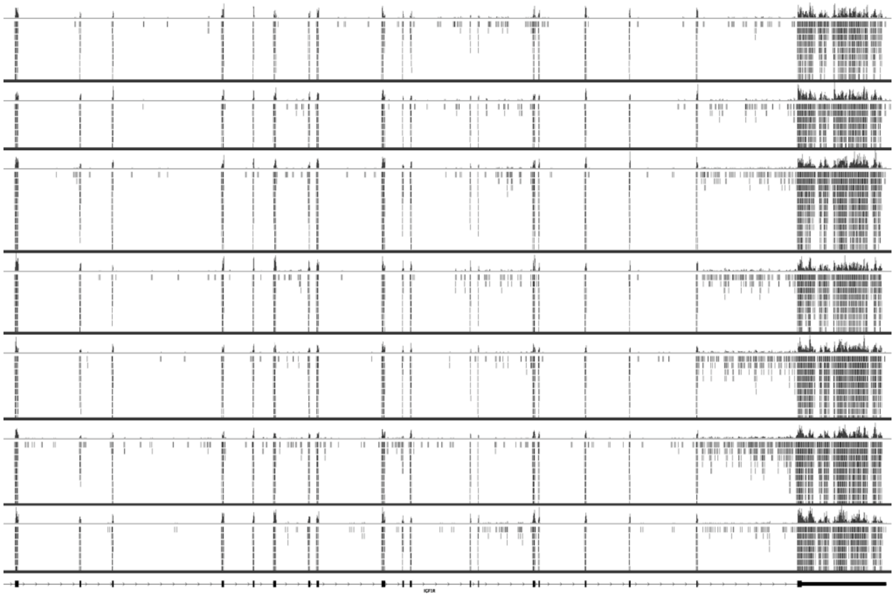


MEASURING GENE EXPRESSION BY RNA SEQUENCING

RANDOMLY FRAGMENTED RNA

CONVERT TO cDNA AND SEQUENCE TAGS AT ENDS OF EACH cDNA FRAGMENT

RNA-seq IGF1R



MEASURING GENE EXPRESSION BY RNA SEQUENCING: PROS AND CONS

ADVANTAGES

- RNA SEQUENCE VARIATIONS DETECTED AT SINGLE NUCLEOTIDE RESOLUTION

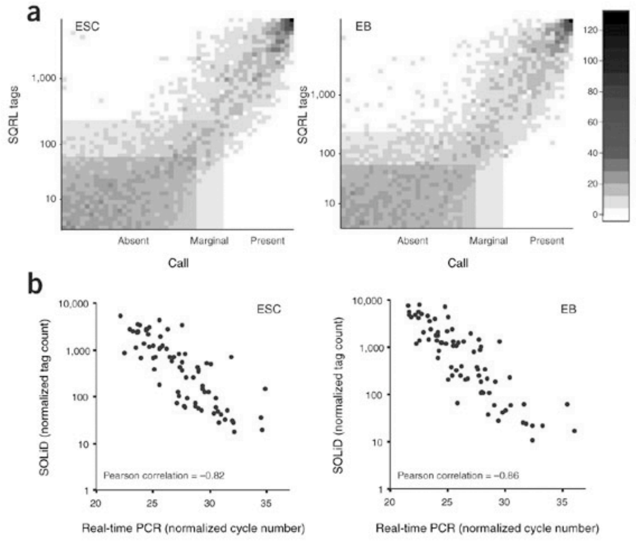
  -ALLELE SPECIFIC EXPRESSION
  -MUTATIONS

- RNA STRUCTURE: SPLICING, START SITES, TERMINATION SITES; REARRANGEMENTS

- DETECTED SIGNALS ARE RELATIVELY UNAMBIGUOUS; POTENTIAL TO OUTPERFORM MICROARRAY

- DE NOVO ASSEMBLY IS POSSIBLE

MEASURING GENE EXPRESSION BY RNA SEQUENCING



Cloonan et al.

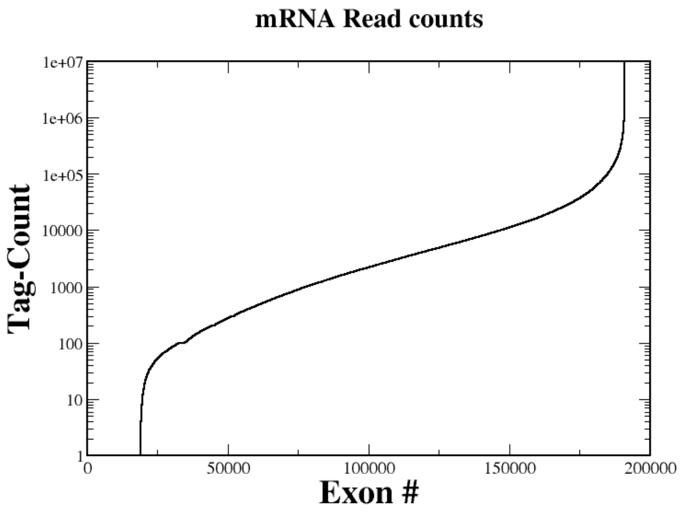
Nature Methods - 5, 613 - 619 (2008)

MEASURING GENE EXPRESSION BY RNA SEQUENCING: PROS AND CONS

LIMITATIONS

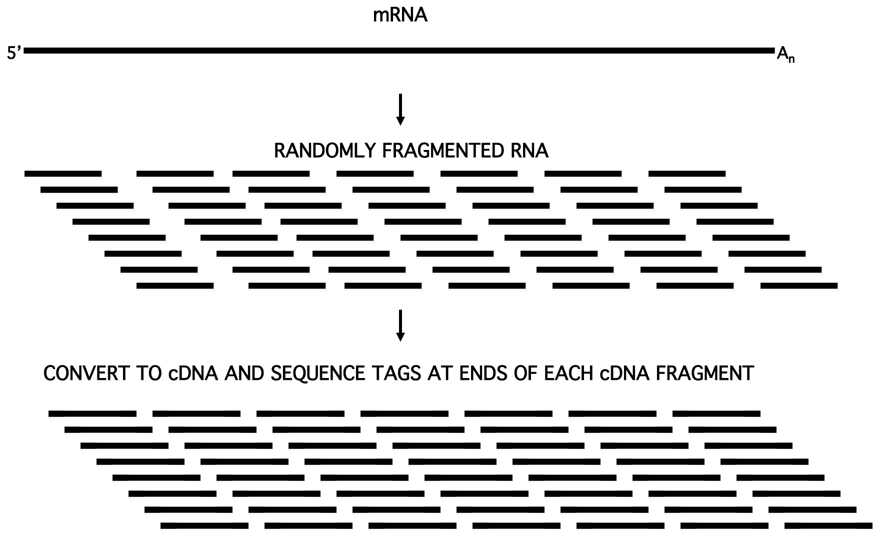
- LOWER LIMIT OF DETECTION IS CONSTRAINED BY THE mRNA ABUNDANCE DISTRIBUTION AND THE NUMBER OF ALIGNED READS PER SAMPLE.
- LARGE SAMPLE NUMBERS DIFFICULT TO ACHIEVE, EXCEPT IN TAG MODE.
- SOFTWARE IS STILL DEVELOPMENTAL: REQUIRES SOPHISTICATED BIOINFORMATICS COLLABORATION. [For review see Pepke et al. Nat Methods 6:S22 (2009)]

mRNA ABUNDANCE VARIES OVER A LARGE DYNAMIC RANGE



Sven Bilke

MEASURING GENE EXPRESSION BY RNA SEQUENCING



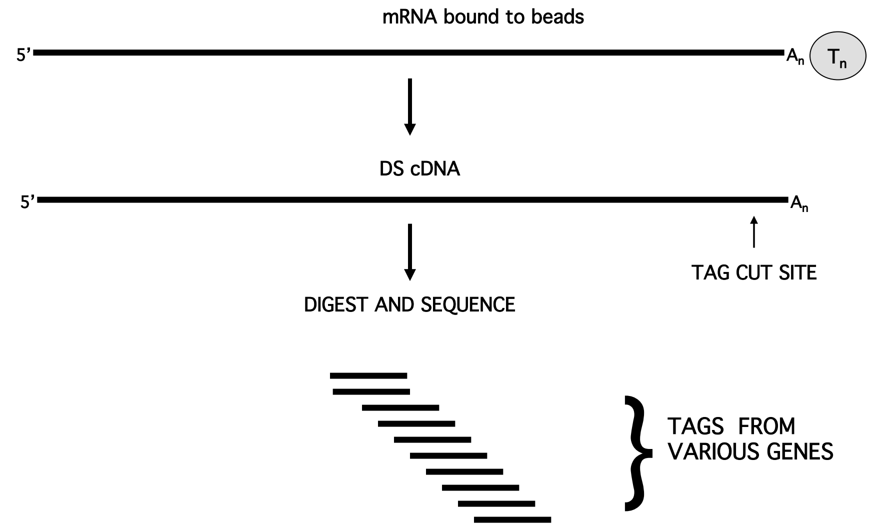
mRNA

5’ A_n

RANDOMLY FRAGMENTED RNA

CONVERT TO cDNA AND SEQUENCE TAGS AT ENDS OF EACH cDNA FRAGMENT

3’ TAG SEQUENCING



mRNA bound to beads

5’ A_n T_n

DS cDNA

5’ A_n

TAG CUT SITE

DIGEST AND SEQUENCE

TAGS FROM VARIOUS GENES

3’ TAG SEQUENCING

- SEQUENCES ALIGNED AND COUNTED
- LIBRARIES OF TAGS FROM MANY SAMPLES CAN BE IDENTIFIED BY ADDING A “BARCODE” AND POOLED BEFORE SEQUENCING
- POTENTIAL TO ANALYZE LARGE NUMBERS OF SAMPLES IN PARALLEL

THE FUTURE?

AS SEQUENCE THROUGHPUT INCREASES AND COSTS PER READ DECLINE, SEQUENCING IS LIKELY TO BECOME AN ATTRACTIVE ALTERNATIVE TO MICROARRAYS IN MORE AND MORE APPLICATIONS.

USEFUL WEB SITES

MGEGD The Microarray Gene Expression Data Society:

http//www.mged.org/

NCBI Gene Expression Omnibus:

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

NCBI Sequence Read Archive (SRA):

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

EBI Microarray informatics:

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

Stanford Microarray Database:

http://smd.stanford.edu/

UCSF DeRisi lab:

http://derisilab.ucsf.edu/data/microarray/index.html

Broad Institute:

Gene Set Enrichment Analysis (GSEA)
http://www.broadinstitute.org/gsea/

Connectivity Map:

http://www.broadinstitute.org/cmap/