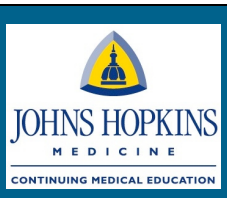
LARGE SCALE ANALYSIS OF GENE EXPRESSION

Evolution and Revolution



Current Topics in Genome Analysis 2012

Paul Meltzer

No Relevant Financial Relationships with Commercial Interests

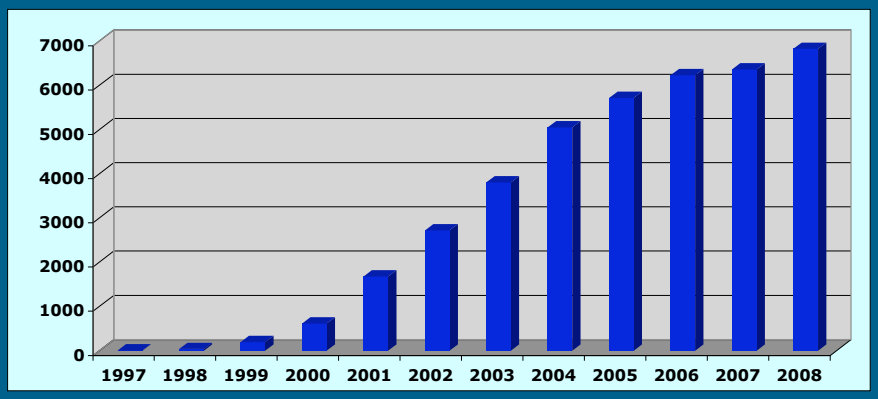
AFTER THE SEQUENCE:
WHOLE GENOME APPROACHES TO BIOLOGICAL QUESTIONS

GENE EXPRESSION
GENE VARIATION
GENE FUNCTION

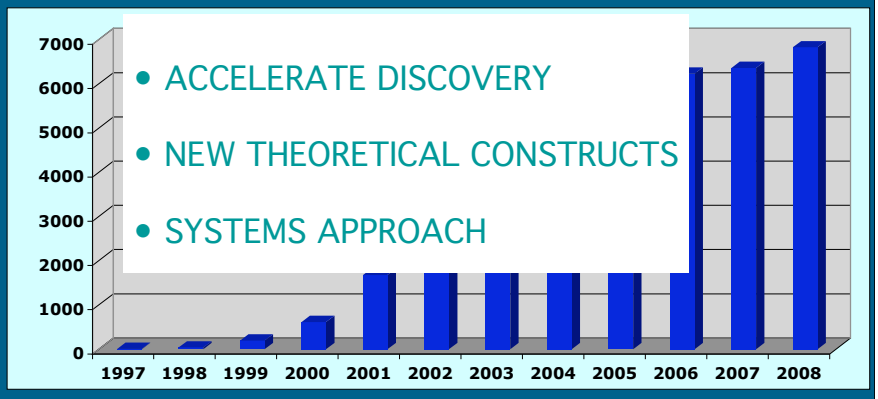
MICROARRAYS PROVIDE A TOOL FOR WHOLE GENOME ANALYSIS

PRIMARY IMPACT:
ACCELERATED DISCOVERY AND HYPOTHESIS GENERATION

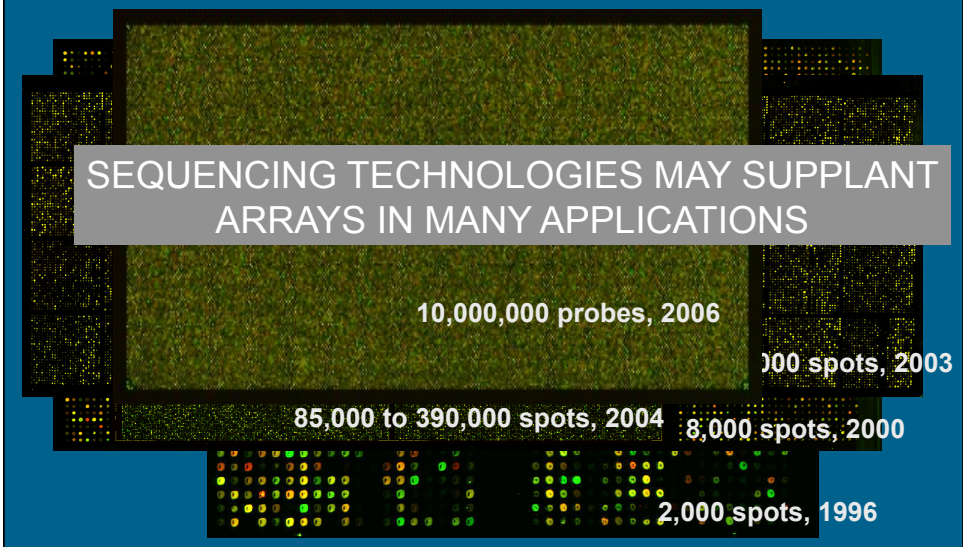
PUBMED CITATIONS FOR DNA MICROARRAYS



PUBMED CITATIONS FOR DNA MICROARRAYS



INCREASE IN FEATURE DENSITY



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

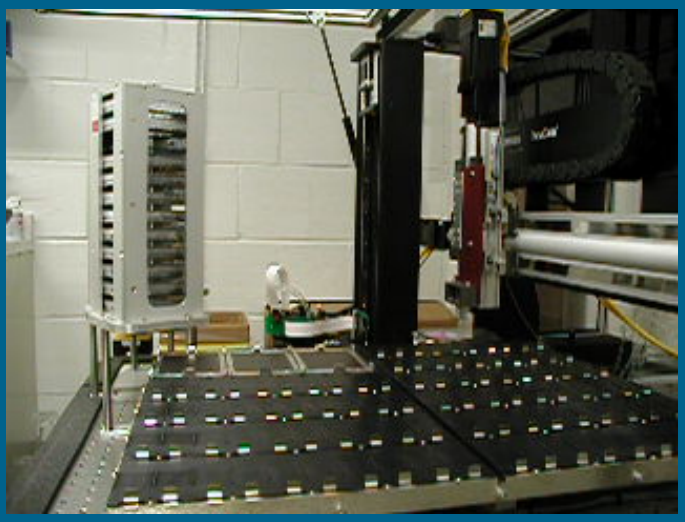
- Synthetic Oligonucleotides
- PCR products from
  - Cloned DNAs
  - Genomic DNA
- Cloned DNA

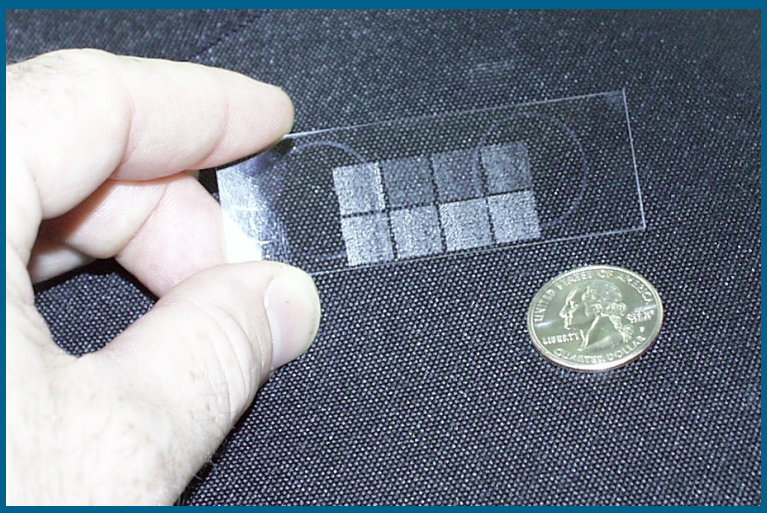
OLIGONUCLEOTIDE ARRAY DESIGN

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

Microarray Manufacture

- **Printing**

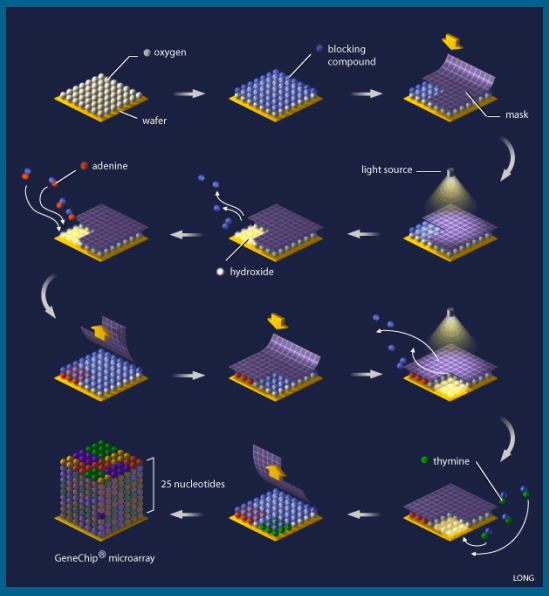




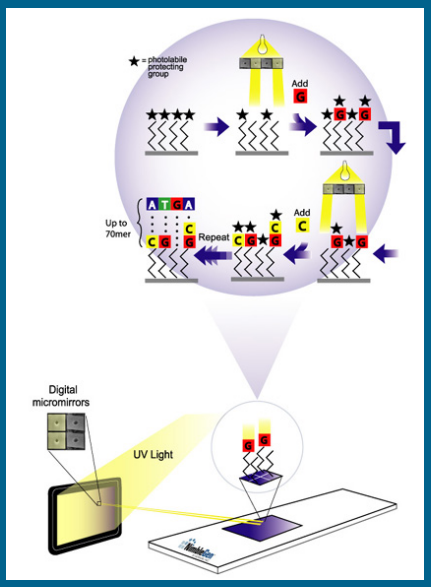
Microarray Manufacture

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

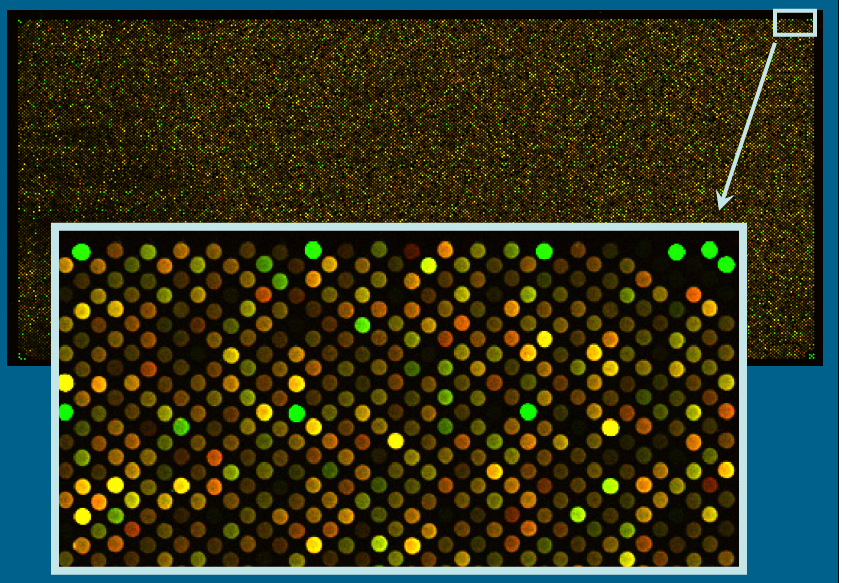
LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

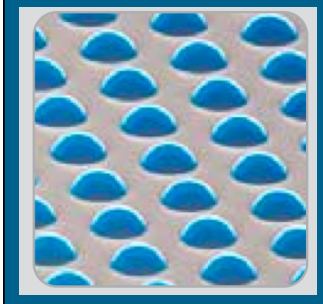


LIGHT DIRECTED OLIGONUCLEOTIDE SYNTHESIS

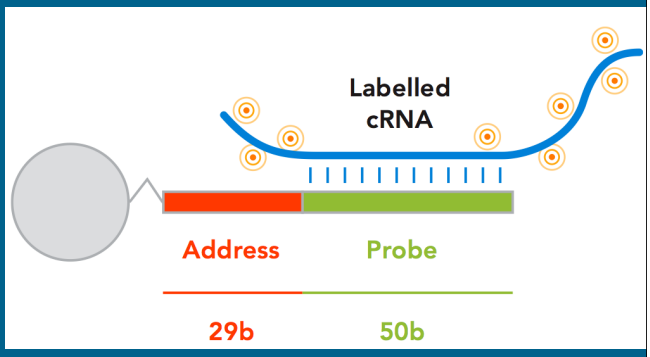


INK JET DIRECTED SYNTHESIS





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/bead arrays: millions
- Array design is linked to purpose.

Laboratory Essentials

- Arrays
- Hybridization and Wash Equipment
- 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

Reports on Microarray Data Quality

Nature Biotechnology

September 2006

Accessing Expression Data

•Individual Lab and Journal Sites; public databases

GEO

Currently contains expression data on 592,204 sample sets.

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

Accessing Expression Data

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

Publishing Expression Data

•MIAME standard

Minimum Information about a Microarray Experiment

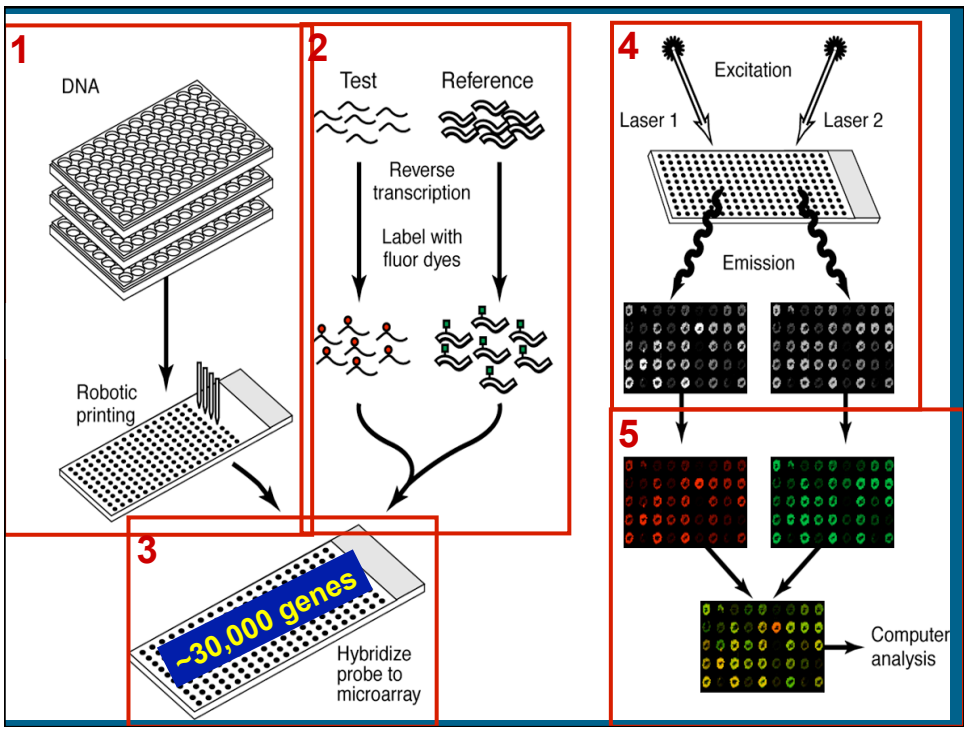
• Format required by many journals

• 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**


1
DNA
Robotic printing
2
Test
Reference
Reverse transcription
Label with fluor dyes
3
~30,000 genes
Hybridize probe to microarray
4
Excitation
Laser 1
Laser 2
Emission
5
Computer analysis


mRNA Extracted from Cell
Reverse Transcription
cDNA
Fragmented, Biotin-labeled cRNA
Biotin-labeled cRNA
Fragmentation
Transcription
GeneChip Expression Array
Hybridize
Wash and Stain
Scan and Quantitate
LONG
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

?????

Downstream Genes

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

EXPRESSION DATA ANALYSIS

•Large amount of data

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

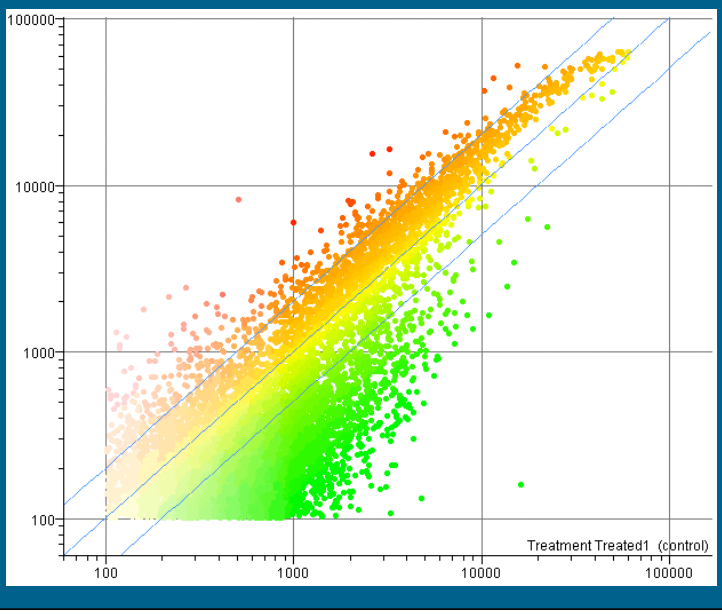
•Requires analysis and visualization tools

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**

MICROARRAY SCATTER PLOT



Unsupervised Clustering

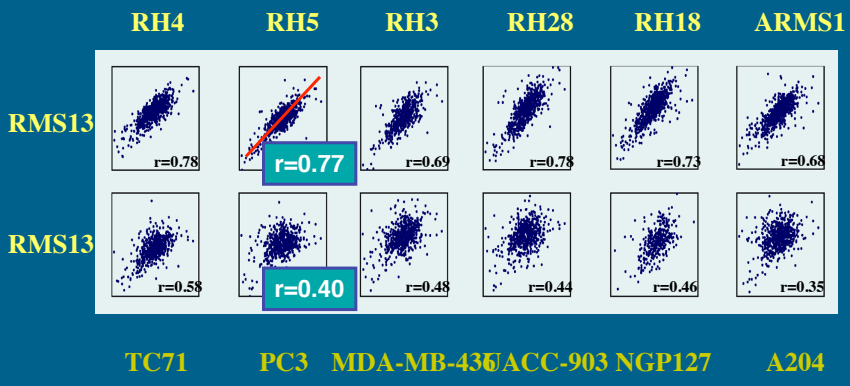
How do genes and samples organize into groups?

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

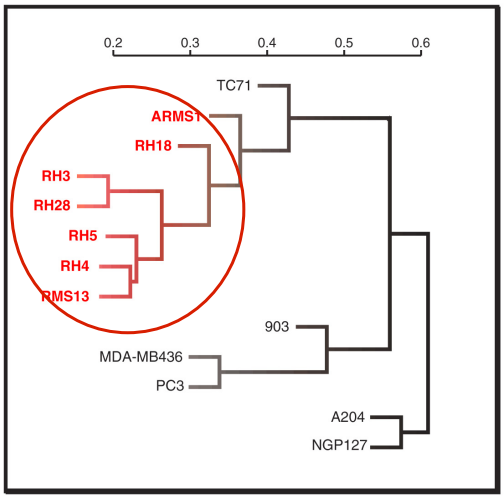


Khan et al. Can Res 58:5009

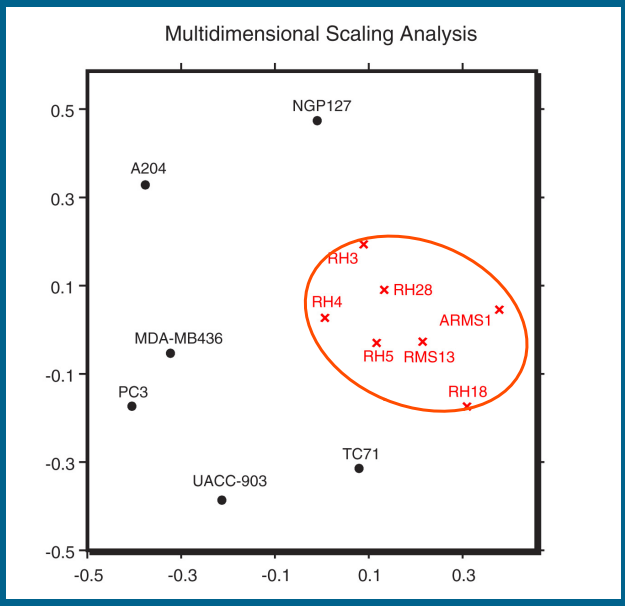
Matrix of Pearson Correlation Coefficients Distance Map

	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												

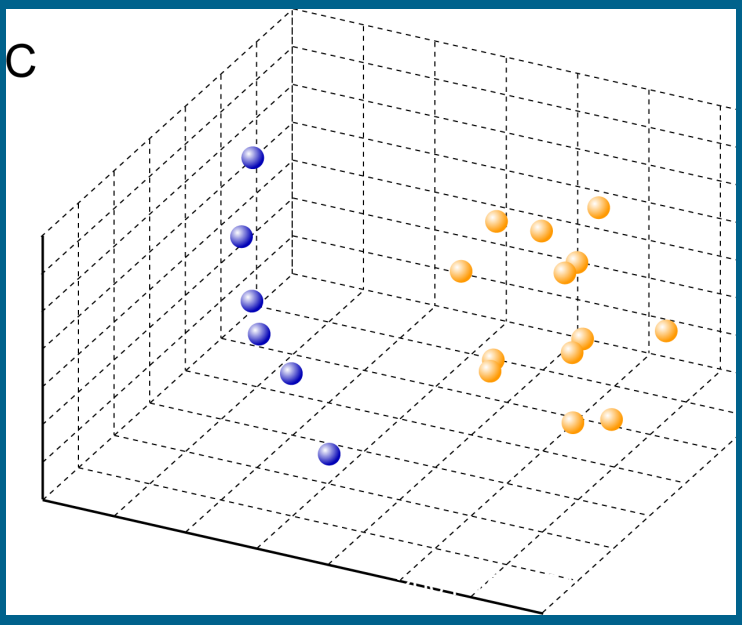
Hierarchical Clustering Dendrogram



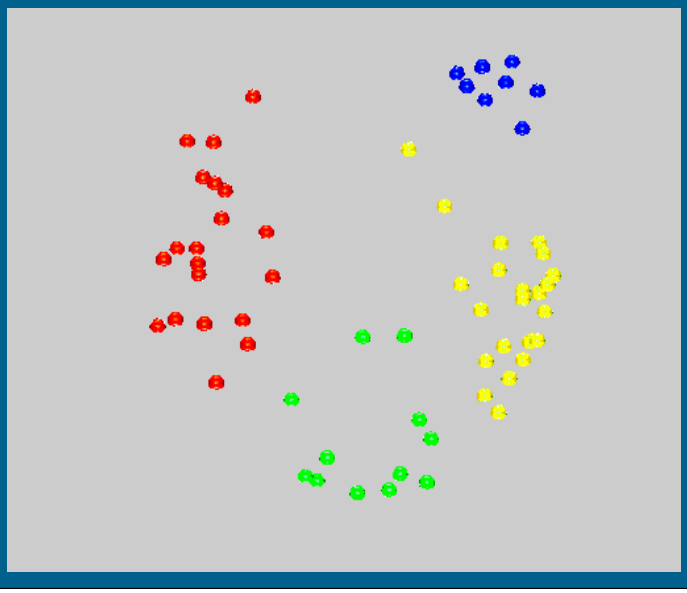
DATA DISPLAY BY MULTIDIMENSIONAL SCALING



MDS PLOT

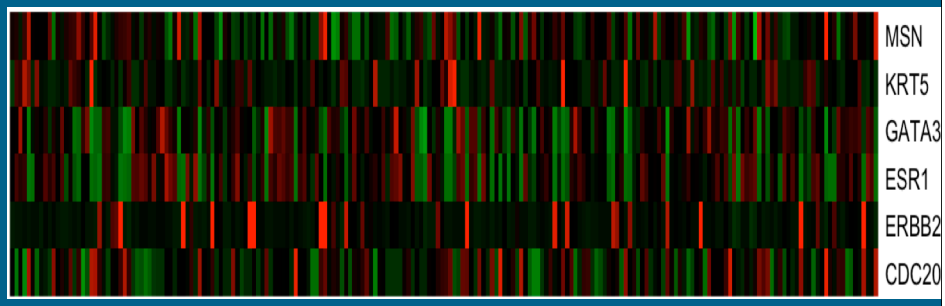


MULTIDIMENSIONAL SCALING OR PRINCIPLE COMPONENT ANALYSIS CAPTURE VARIATION IN SAMPLES AND ARE EXCELLENT VISUALIZATION TOOLS

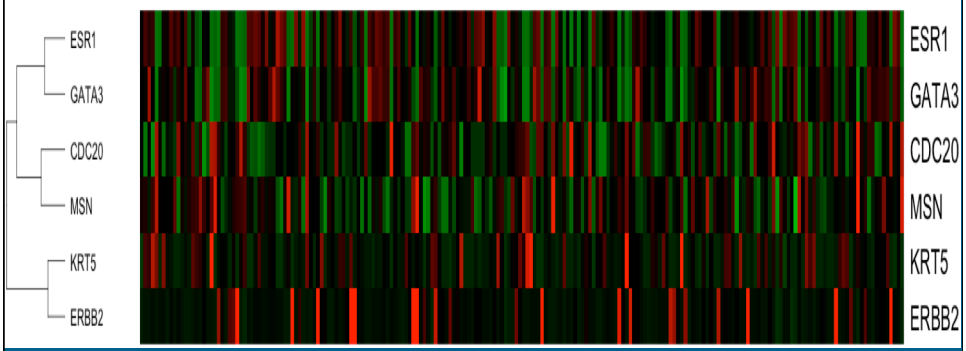


Khan et al
Nat Med
7:673

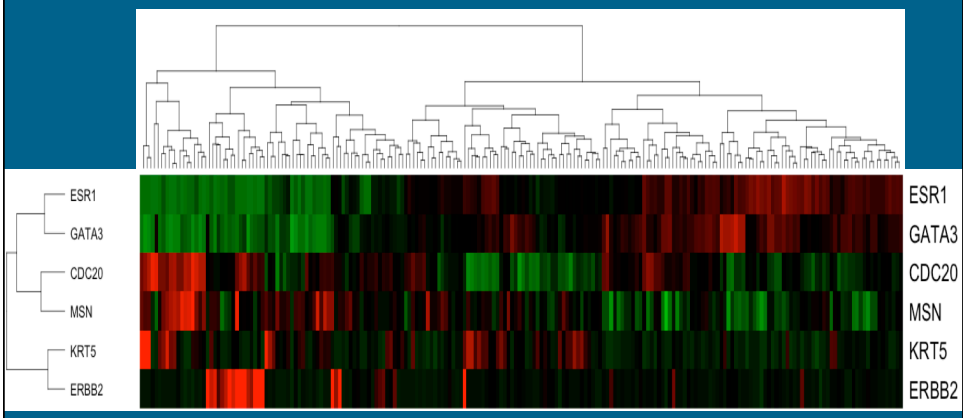
UNCLUSTERED DATA (6 GENES IN 200 SAMPLES)



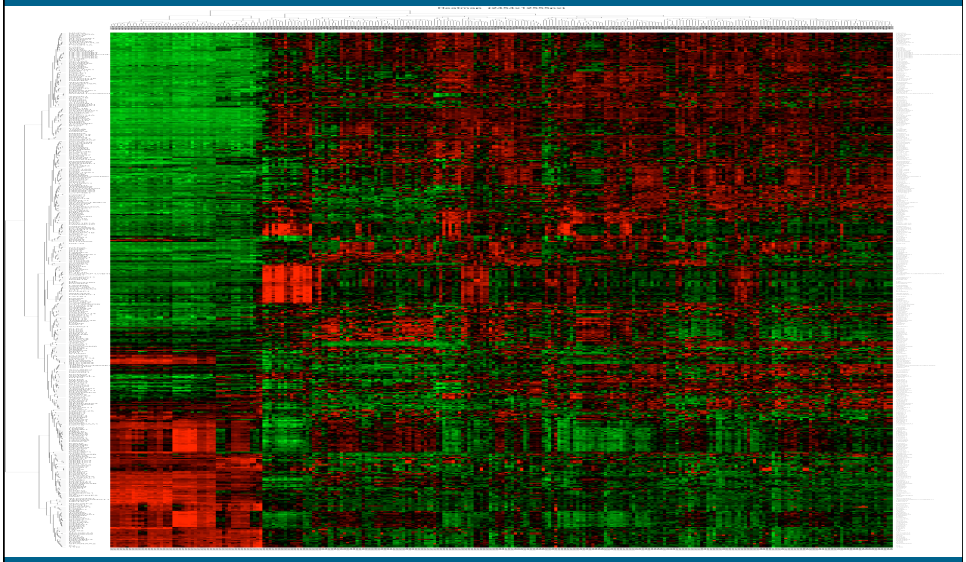
CLUSTERING GENES ONLY



CLUSTERING GENES AND SAMPLES



CLUSTERING GENES AND SAMPLES



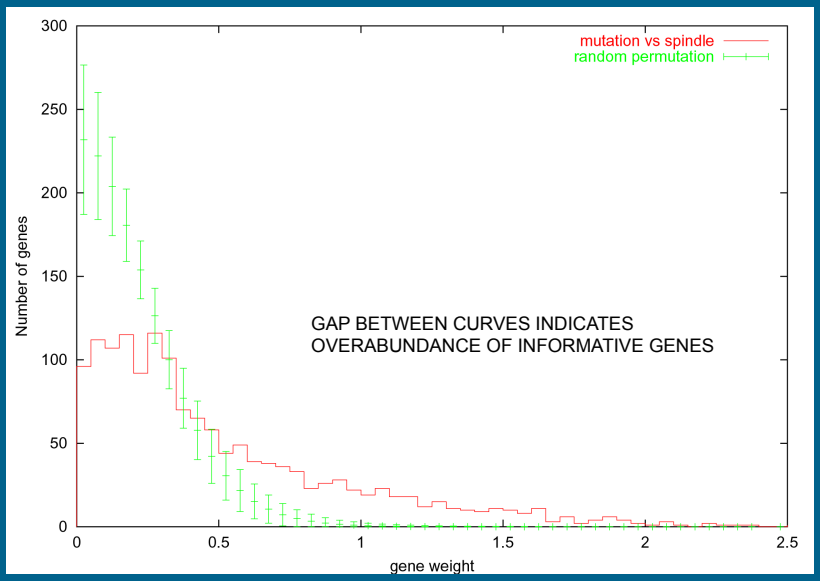
DATA FROM GEO

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

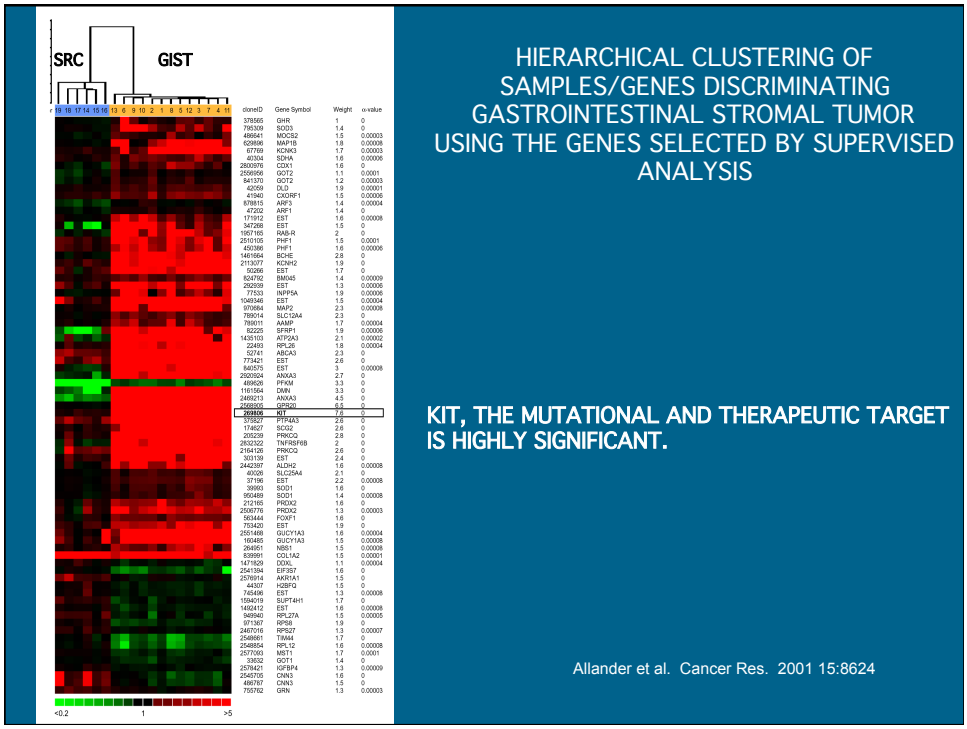


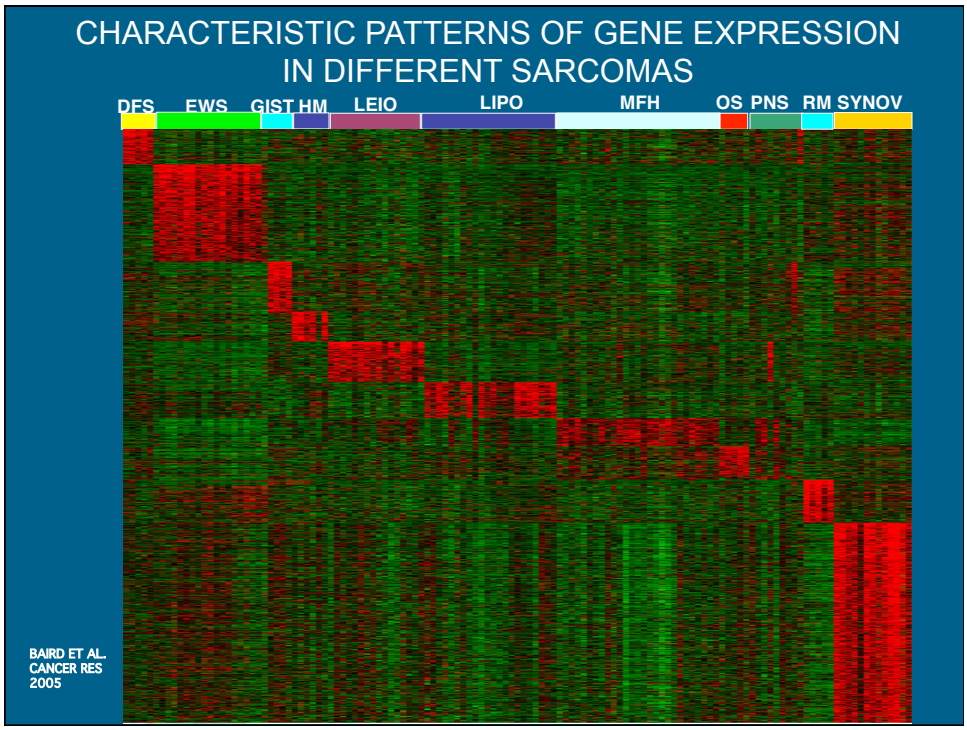
Allander et al. Cancer Res. 2001 15:8624

SUPERVISED METHODS GENERATE RANKED GENE LISTS

TOP DISCRIMINATORS FOR GIST

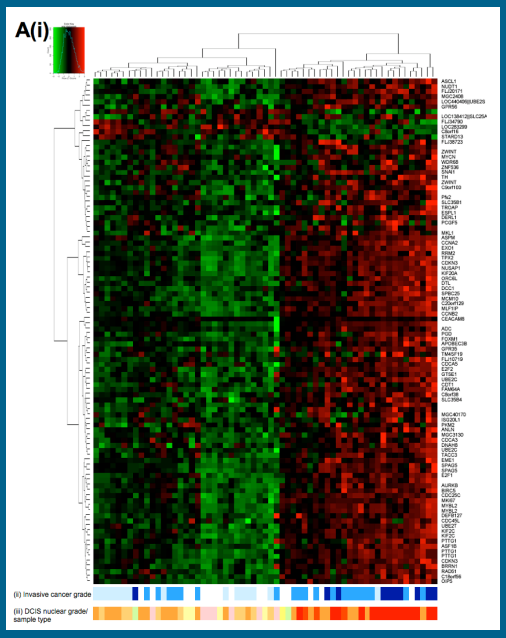
Rank	Weight	Gene Description
1	7.55575	v-kit sarcoma oncogene
2	6.48306	G coupled receptor 20
3	4.60057	G coupled receptor 20
4	4.51681	annexin A3
5	3.33057	KIAA0353 protein
6	3.31734	phosphofructokinase
7	2.95095	DKFZP434N161 n
8	2.83435	protein kinase C, theta
9	2.79721	butyrylcholinesterase
10	2.72752	annexin A3


SRC
GIST
HIERARCHICAL CLUSTERING OF SAMPLES/GENES DISCRIMINATING GASTROINTESTINAL STROMAL TUMOR USING THE GENES SELECTED BY SUPERVISED ANALYSIS
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
DFS
EWS
GIST
HM
LEIO
LIPO
MFH
OS
PNS
RM
SYNOV
BAIRD ET AL. CANCER RES 2005

CLUSTERING GENES AND SAMPLES

Separation Of High and Low Grade Ductal Carcinoma In Situ

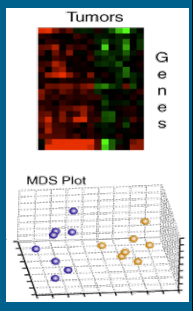


Balleine et al. 14:8244 Clin Can Res 2008

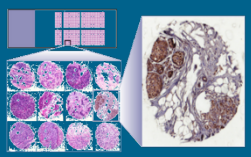
GENOMICS FROM BENCH TO BEDSIDE

WHOLE GENOME

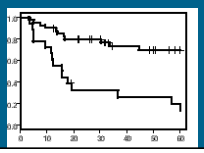
GENE SELECTION



GENE VALIDATION



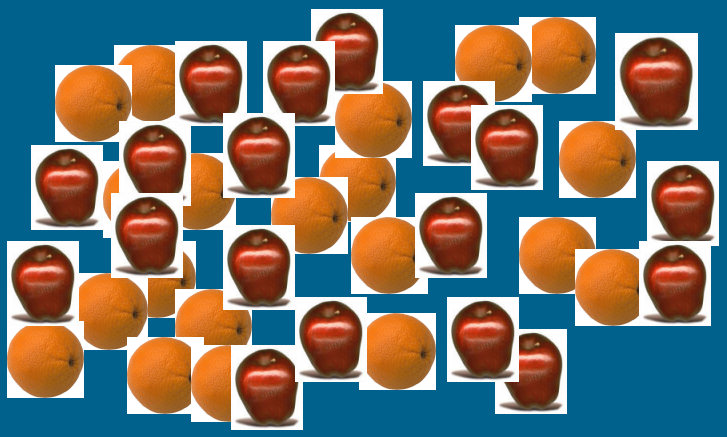
ASSAY DEVELOPMENT



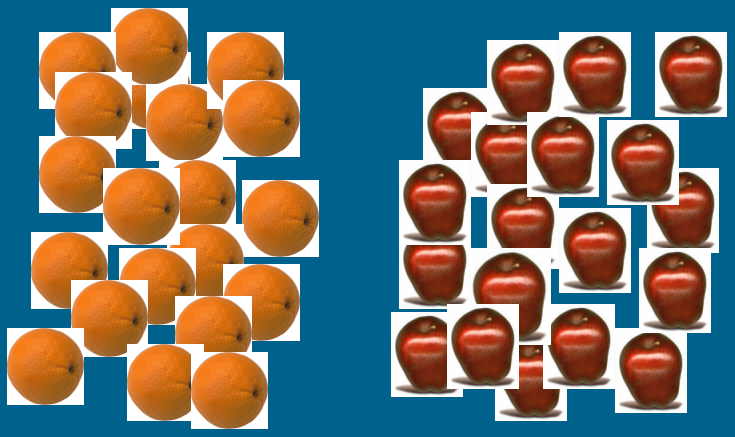
SIGNAL STRENGTH VARIES IN TISSUE PROFILING EXPERIMENTS

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

CONSIDER A SAMPLE SET



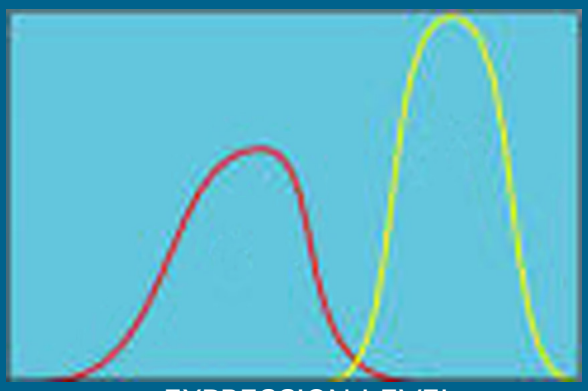
CONSIDER A SAMPLE SET



THESE ARE EASY TO DISTINGUISH BY ONE MEASUREMENT PER INDIVIDUAL.

CONSIDER A SAMPLE SET

TUMORS



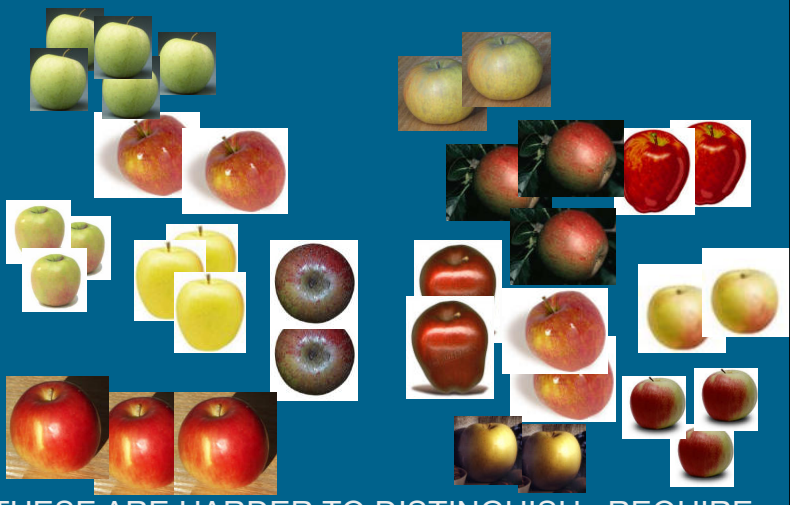
EXPRESSION LEVEL
(HIGHLY INFORMATIVE GENE)

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.

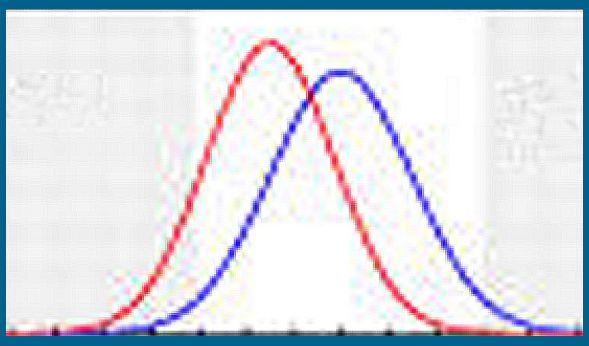
CONSIDER A SAMPLE SET



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.

J Natl Cancer Inst. 2007 Jan 17;99(2):147-57.
Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting.
Dupuy A, Simon RM.

BACKGROUND: Both the validity and the reproducibility of microarray-based clinical research have been challenged. There is a need for critical review of the statistical analysis and reporting in published microarray studies that focus on cancer-related clinical outcomes. METHODS: Studies published through 2004 in which microarray-based gene expression profiles were analyzed for their relation to a clinical cancer outcome were identified through a Medline search followed by hand screening of abstracts and full text articles. Studies that were eligible for our analysis addressed one or more outcomes that were either an event occurring during follow-up, such as death or relapse, or a therapeutic response. We recorded descriptive characteristics for all the selected studies. A critical review of outcome-related statistical analyses was undertaken for the articles published in 2004. RESULTS: Ninety studies were identified, and their descriptive characteristics are presented. Sixty-eight (76%) were published in journals of impact factor greater than 6. A detailed account of the 42 studies (47%) published in 2004 is reported. Twenty-one (50%) of them contained at least one of the following three basic flaws: 1) in outcome-related gene finding, an unstated, unclear, or inadequate control for multiple testing; 2) in class discovery, a spurious claim of correlation between clusters and clinical outcome, made after clustering samples using a selection of outcome-related differentially expressed genes; or 3) in supervised prediction, a biased estimation of the prediction accuracy through an incorrect cross-validation procedure. CONCLUSIONS: The most common and serious mistakes and misunderstandings recorded in published studies are described and illustrated. Based on this analysis, a proposal of guidelines for statistical analysis and reporting for clinical microarray studies, presented as a checklist of "Do's and Don'ts," is provided.

MICROARRAY STUDIES GENERATE ORGANIZED LIST OF GENES

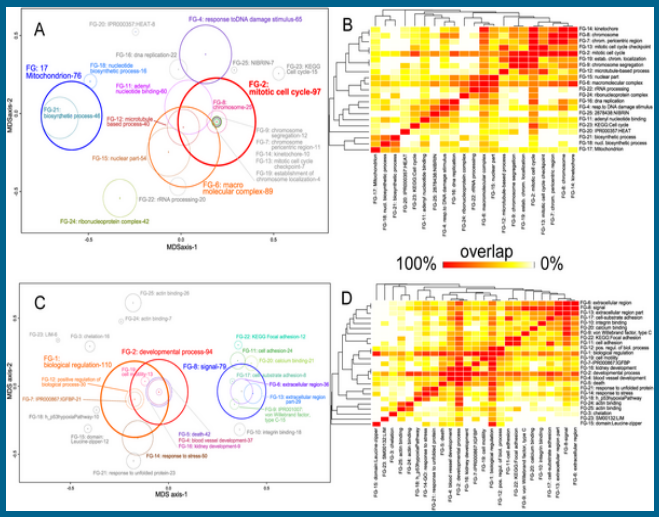
- Often cryptic and hard to interpret.
- Hypothesis generating, but this is often rather subjective.
- Seldom provide strong evidence for a specific mechanism.
- Expression data is intrinsically limited.

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).

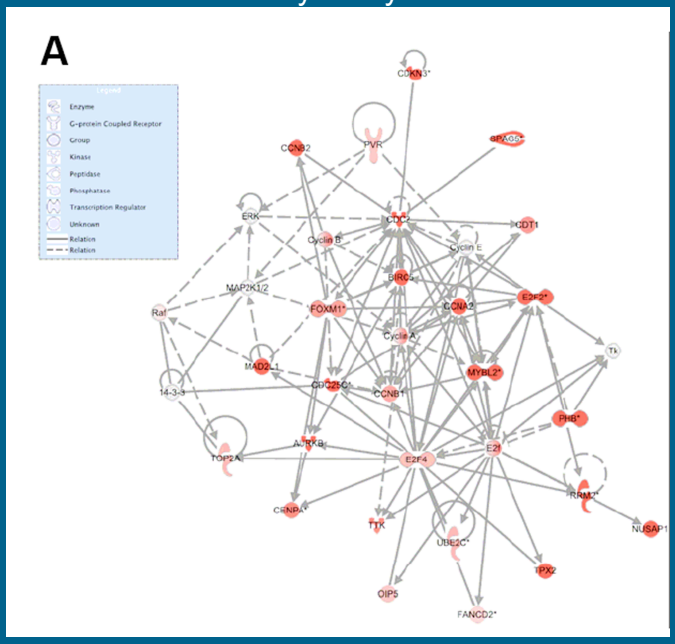
GENE ONTOLOGY AND PROMOTER DATABASES CAN HELP FIND BIOLOGY

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



KAUER ET AL. PLOS ONE 4:e5415 2009

Pathway Analysis



Balleine et al. 14:8244 Clin Can Res 2008

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?

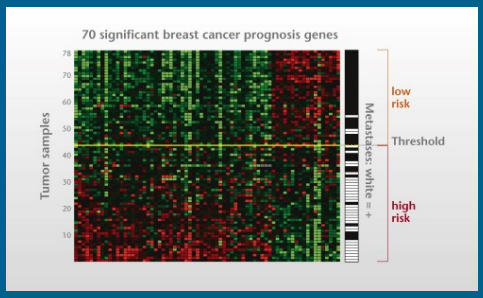
OPTIONS:

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

•USE ARRAYS FOR DISCOVERY TO EXTRACT SIGNATURES WHICH CAN BE ASSAYED WITH ALTERNATIVE TECHNOLOGIES.

FDA APPROVED TESTS FOR BREAST CANCER BASED ON EXPRESSION STUDIES

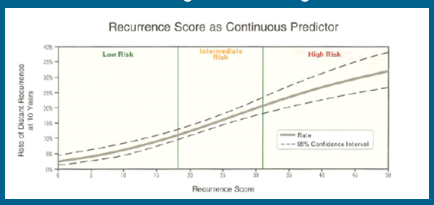
70 GENE MICROARRAY SIGNATURE



Van de Vijver et al

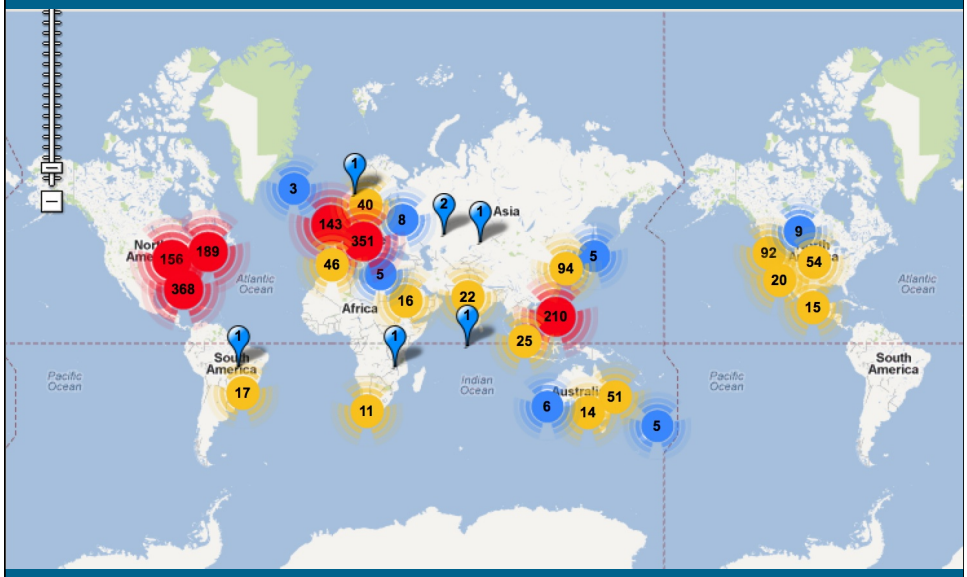
NEJM 347:1999 .

Muiltgene RT-PCR Signature

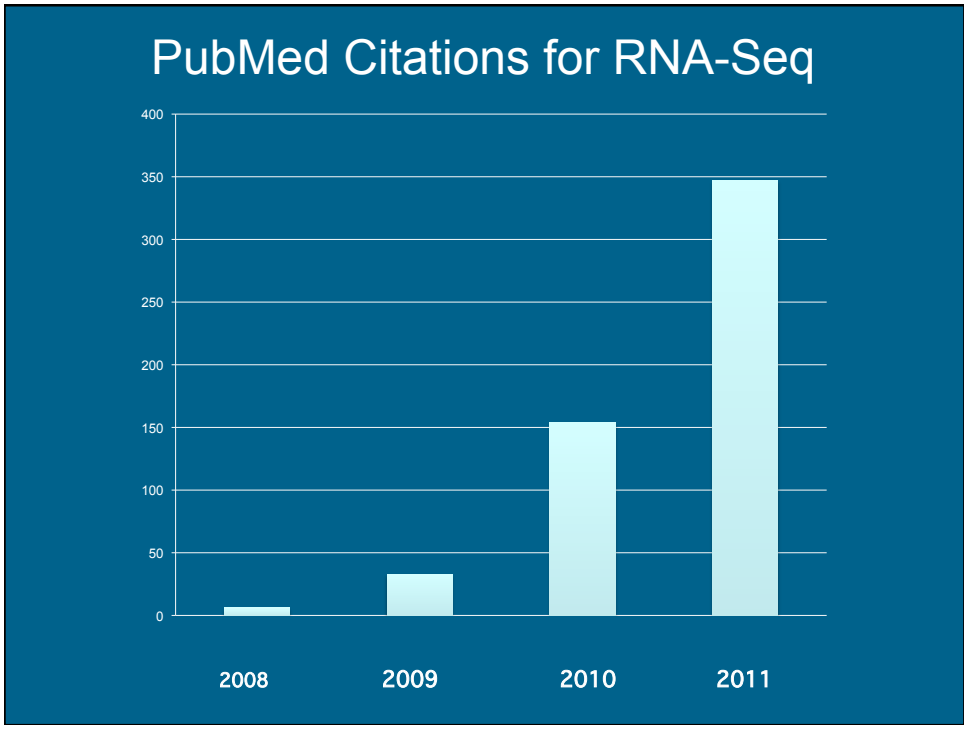


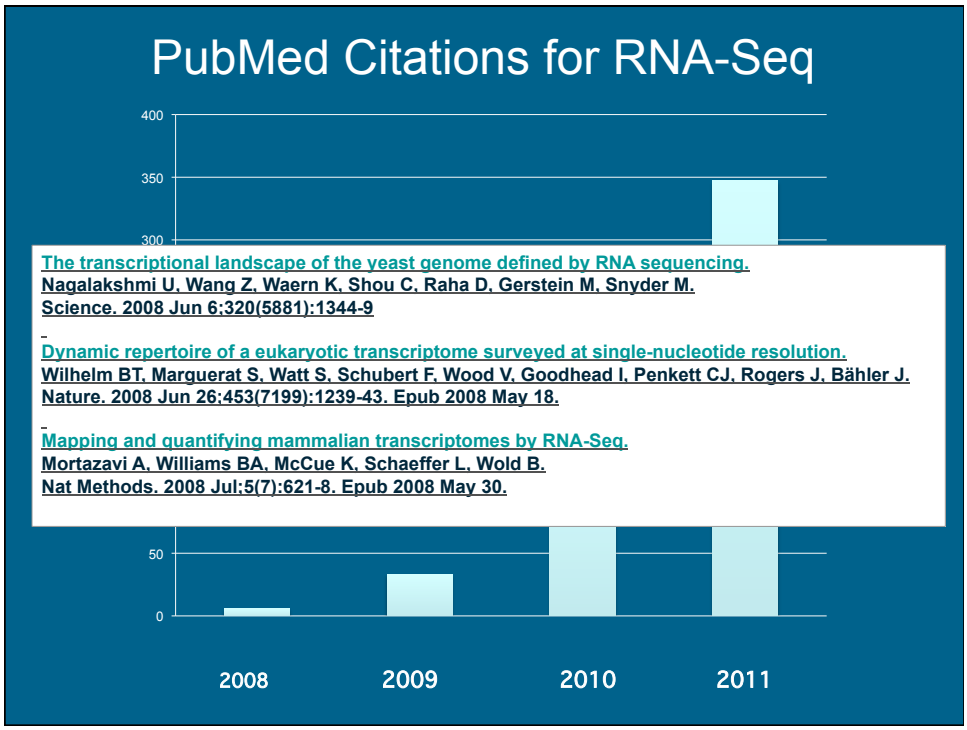
Paik et al NEJM 351:2817

THEY' RE EVERYWHERE!



http://pathogenomics.bham.ac.uk/hts/


PubMed Citations for RNA-Seq
400
350
300
250
200
150
100
50
0
2008
2009
2010
2011


PubMed Citations for RNA-Seq
400
350
300
The transcriptional landscape of the yeast genome defined by RNA sequencing.
Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M.
Science. 2008 Jun 6;320(5881):1344-9
Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution.
Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J.
Nature. 2008 Jun 26;453(7199):1239-43. Epub 2008 May 18.
Mapping and quantifying mammalian transcriptomes by RNA-Seq.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.
Nat Methods. 2008 Jul;5(7):621-8. Epub 2008 May 30.
50
0
2008
2009
2010
2011

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	SEQUENCING
• READILY AVAILABLE MATURE TECHNOLOGY • RELATIVELY INEXPENSIVE • EFFECTIVE WITH VERY COMPLEX SAMPLES • HUNDREDS OF SAMPLES PRACTICAL • CAN TARGET SUBSET OF GENOME	• WHOLE GENOME DATA • RELATIVELY 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	• IMMATURE TECHNOLOGY • TECHNOLOGY SPECIFIC ARTIFACTS • RESOURCE INTENSIVE • COMPUTATIONALLY INTENSIVE • NO STANDARD ANALYSIS YET • LOWER SAMPLE THROUGHPUT
MICROARRAYS	SEQUENCING

MEASURING GENE EXPRESSION BY RNA SEQUENCING

ADVANTAGES

- RNA SEQUENCE VARIATIONS DETECTED AT SINGLE NUCLEOTIDE RESOLUTION
  - -ALLELE SPECIFIC EXPRESSION
  - -MUTATIONS
  - -RNA EDITING
- 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

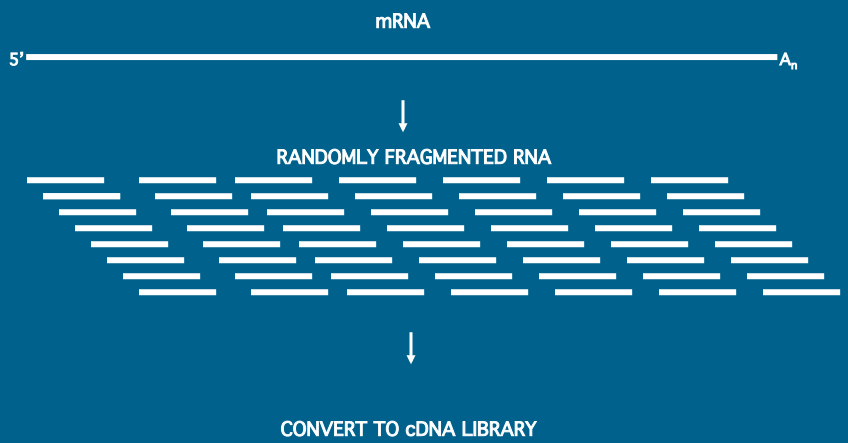
- FULL LENGTH mRNA----RNA-Seq
- TAG SEQUENCING (SAGE-LIKE)
- PolyA vs. Total (ribosomal depleted)
- Strand specific vs. non-strand specific
- miRNA sequencing
- lincRNA sequencing

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)]
- COMPUTATIONAL HARDWARE REQUIREMENTS ARE SUBSTANTIAL
- LIBRARY PREP METHODS EVOLVING
- DATA MAY NOT MERGE WELL IF NOT GENERATED WITH THE SAME METHOD

MEASURING GENE EXPRESSION BY RNA SEQUENCING



RNASeq COMPUTATIONAL WORKFLOW

RAW READS

↓

ALIGN

RNASeq COMPUTATIONAL WORKFLOW

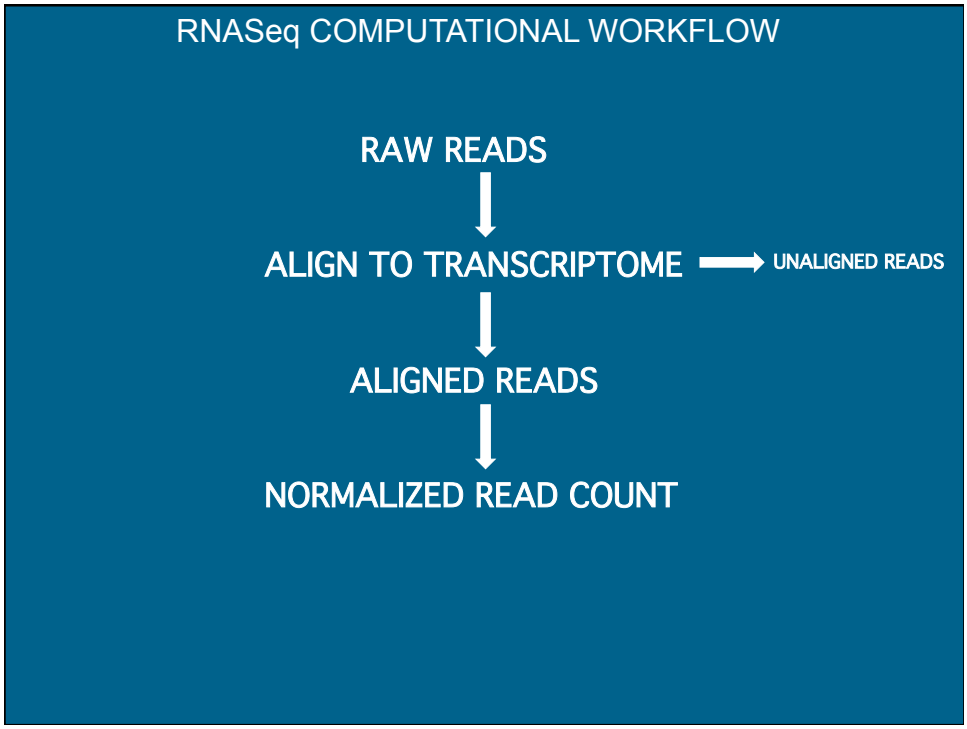
RAW READS

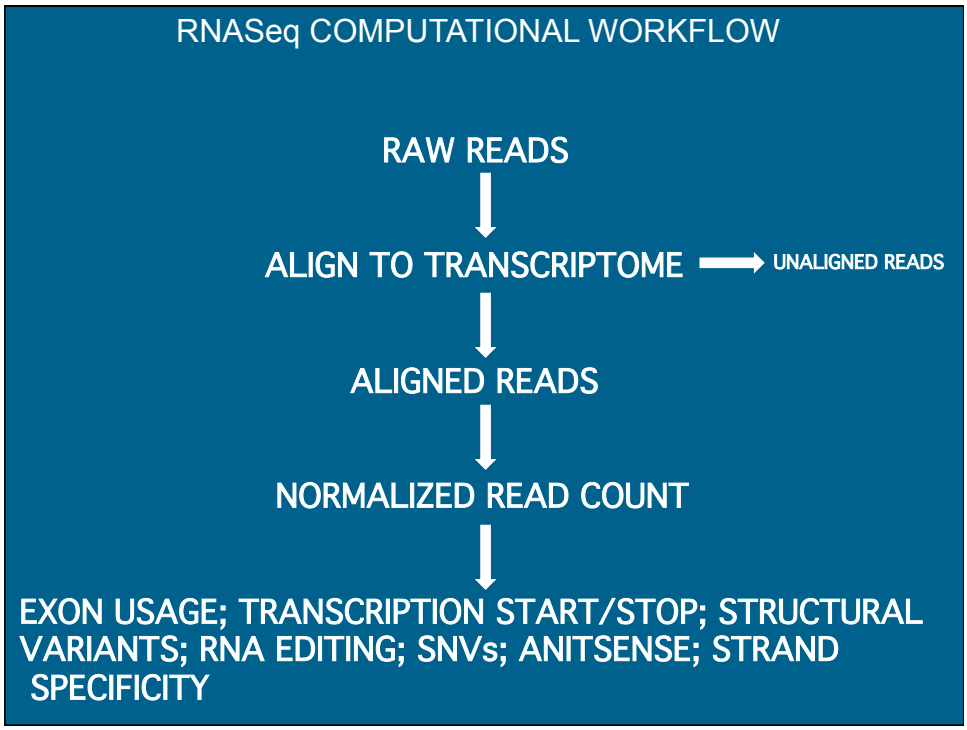
↓

ALIGN

TO WHAT???

- GENOME?
- TRANSCRIPTOME?
- WHICH TRANSCRIPTOME? REFSEQ? EMBL? CUSTOM?


RNASeq COMPUTATIONAL WORKFLOW
RAW READS
ALIGN TO TRANSCRIPTOME
UNALIGNED READS
ALIGNED READS
NORMALIZED READ COUNT


RNASeq COMPUTATIONAL WORKFLOW
RAW READS
ALIGN TO TRANSCRIPTOME
UNALIGNED READS
ALIGNED READS
NORMALIZED READ COUNT
EXON USAGE; TRANSCRIPTION START/STOP; STRUCTURAL VARIANTS; RNA EDITING; SNVs; ANITSENSE; STRAND SPECIFICITY

RNASeq COMPUTATIONAL WORKFLOW

RAW READS

ALIGN TO TRANSCRIPTOME → UNALIGNED READS

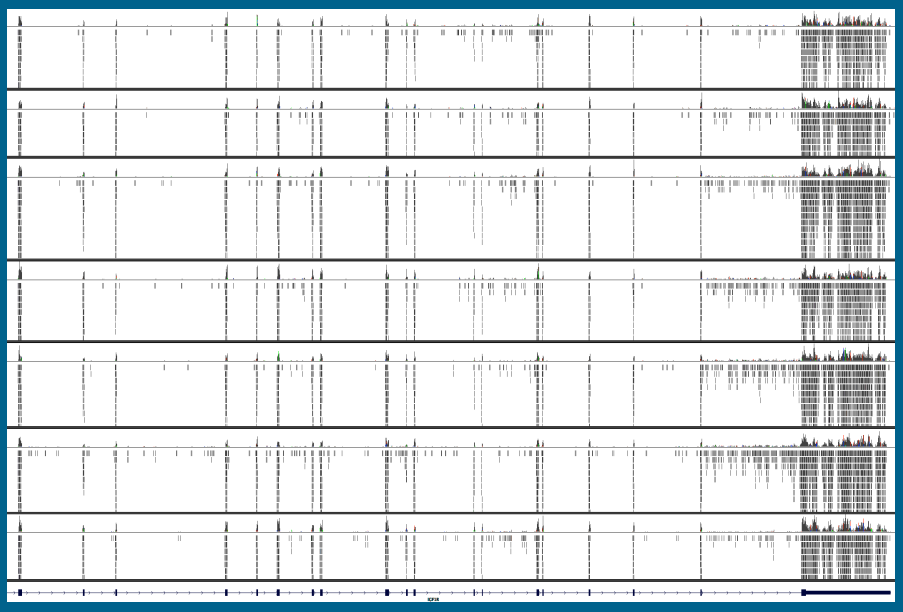
ALIGNED READS

DE NOVO ALIGNMENT

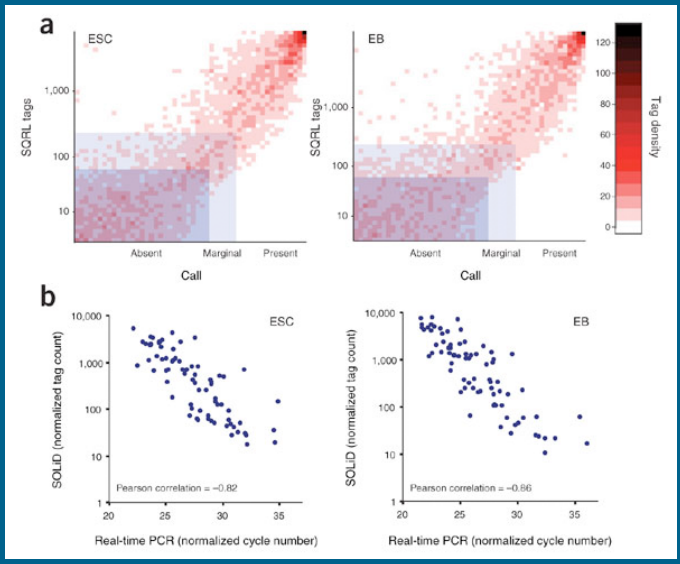
NORMALIZED READ COUNT

EXON USAGE; TRANSCRIPTION START/STOP; STRUCTURAL VARIANTS; RNA EDITING; SNVs; ANITSENSE; STRAND SPECIFICITY

RNA-seq IGF1R



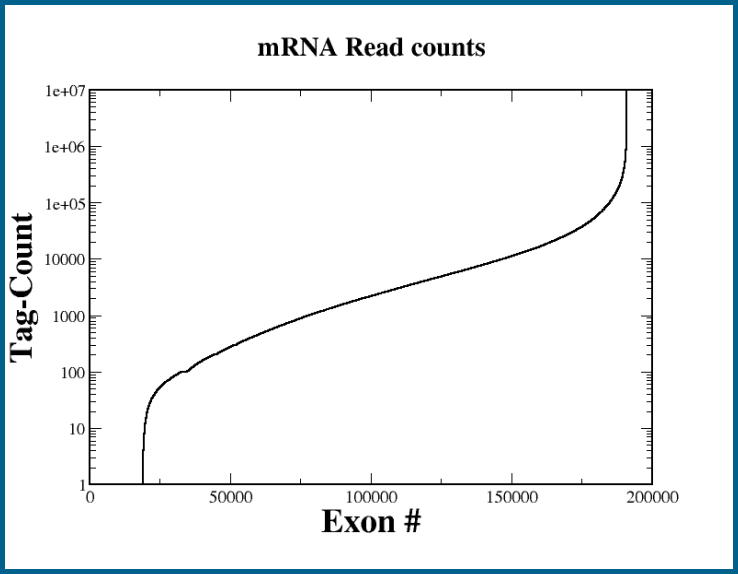
MEASURING GENE EXPRESSION BY RNA SEQUENCING



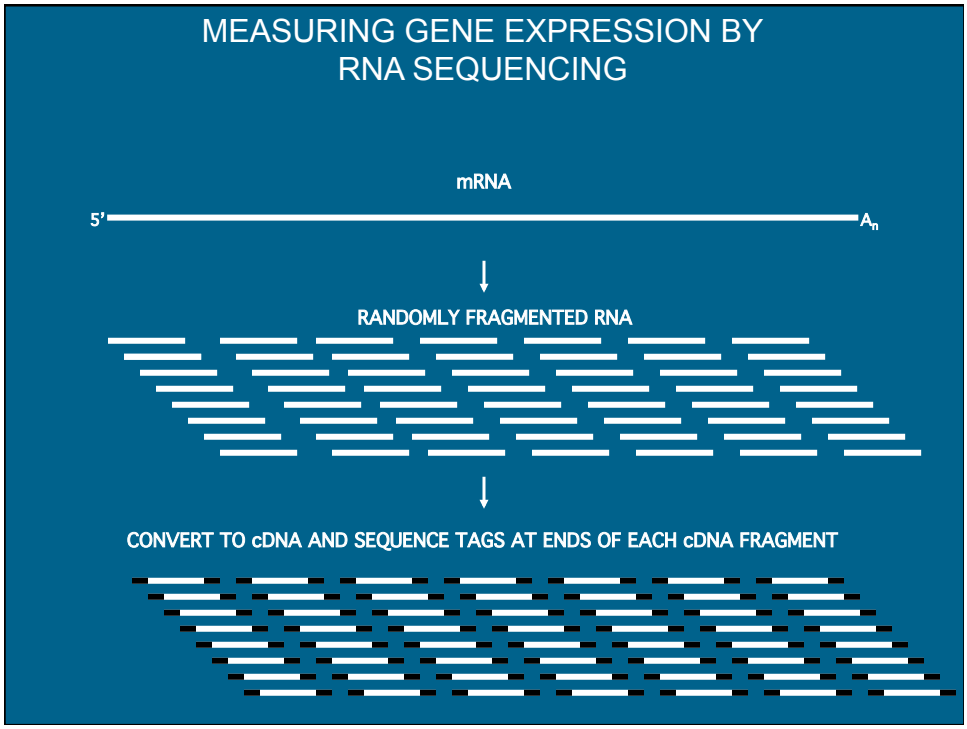
Cloonan et al.

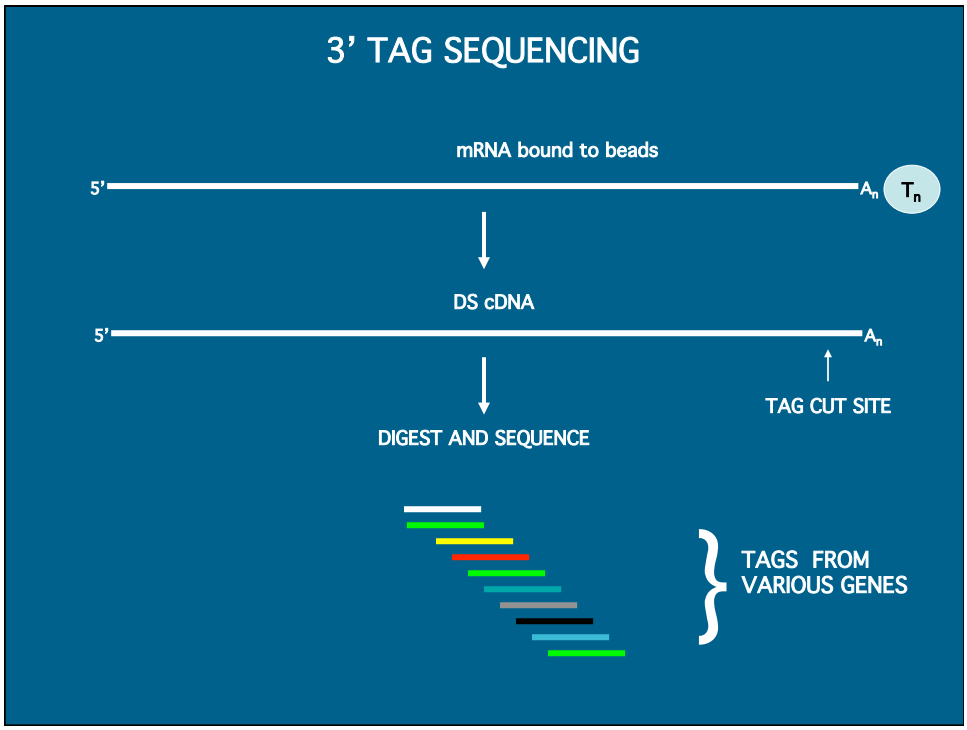
Nature Methods - 5, 613 - 619 (2008)

mRNA ABUNDANCE VARIES OVER A LARGE DYNAMIC RANGE



Sven Bilke


MEASURING GENE EXPRESSION BY
RNA SEQUENCING
mRNA
5'
An
RANDOMLY FRAGMENTED RNA
CONVERT TO cDNA AND SEQUENCE TAGS AT ENDS OF EACH cDNA FRAGMENT


3' TAG SEQUENCING
mRNA bound to beads
5'
An
Tn
DS cDNA
5'
An
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/