Bioinformatics — Lecture 6
Microarray analysis
(EG Ch. 13, MM. Ch. 10)

Krzysztof Bartoszek

Linköping University

krzysztof.bartoszek@liu.se

4 XII 2023 (U6)

Today

- Introduction
- Microarray analysis: preprocessing
  - Image analysis
  - Data
  - Normalization
- Microarray analysis: Statistical analysis
- Functional genomics
- Software
- RNA–seq

Additional reading

BH P. Baldi, G. W. Hatfield. DNA Microarray and Gene Expression, 2002, Cambridge University Press

L A.M. Lesk. Introduction to Bioinformatics, Oxford, 2014. Oxford University Press

R S. Raychaudhuri. Computational Text Analysis for Functional Genomics and Bioinformatics, 2006, Oxford University Press

S D. Stekel. Microarray Bioinformatics, 2003, Cambridge University Press

X J. Xiong. Essential Bioinformatics, 2006, Cambridge University Press

Motivating questions (EG Ch. 13.1.1)

Gene expression: process of changing gene into gene product

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein}$$

Sample: biological material under some condition

e.g. disease, environment, stimulation, stimulus, e.t.c.

1. What genes are expressed in a given sample?
2. What genes are differentially expressed between different samples?
3. Identify clusters of genes whose expression is correlated.
4. Identify gene–gene interactions in networks of activity over time.

Microarrays

Idea from 1995

Hybridization used to measure abundance of target molecule

Array will contain probes for thousands of different sequences

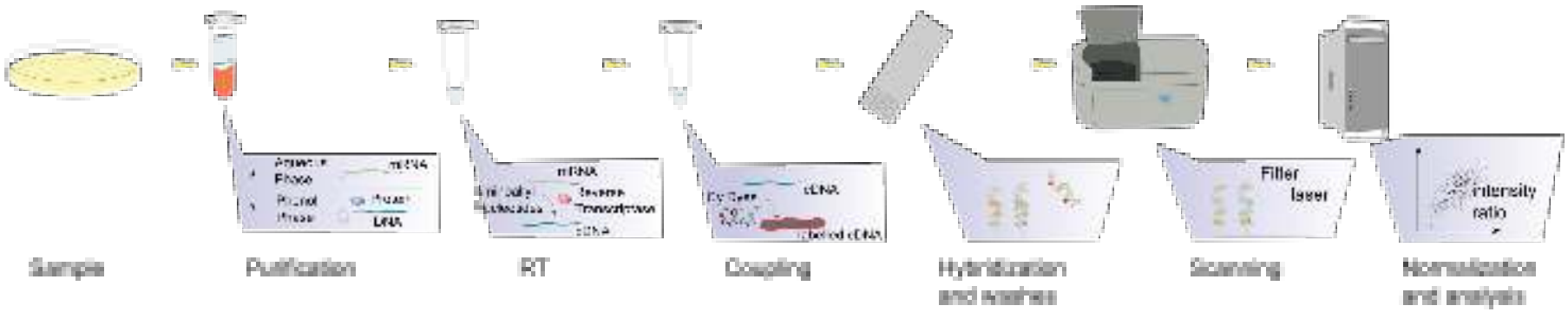
If sequence present, hybridization takes place and "glows"

probe complimentary sequence to target molecule's sequence

based on N. L. Barbosa–Morais' slides from Functional Genomics module of MPhil in Comp. Biol., Univ. Cambridge

NOT A MATRIX!

Stages of microarray analysis (see also MM Fig. 10–5)



https://commons.wikimedia.org/w/index.php?curid=39423104 by Squidonius, public domain

Binding

Access for free at https://openstax.org/books/microbiology/pages/1-introduction.
N. Parker, M. Schneegurt, A.-H. Thi Tu, P. Lister, B. M. Forster, OpenStax, Microbiology, Houston Texas, 2016
Fig. 12.13 in Section 12.2

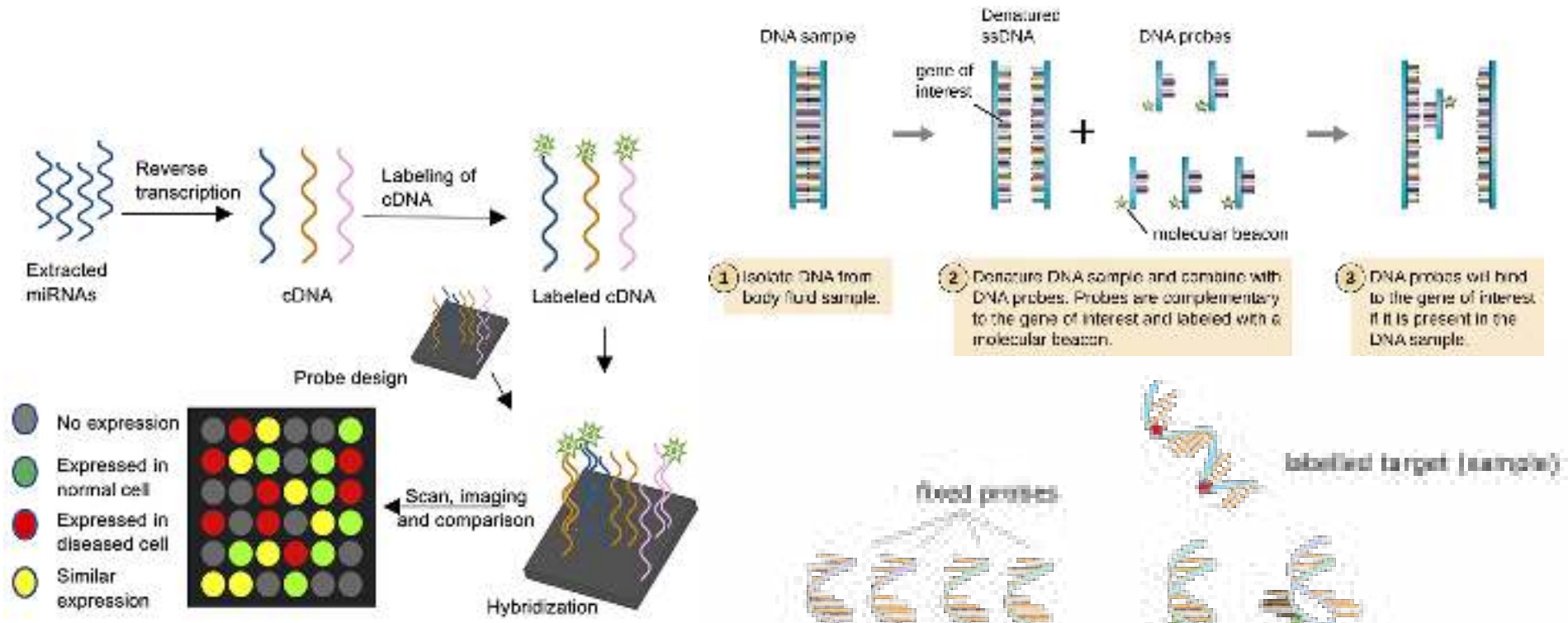
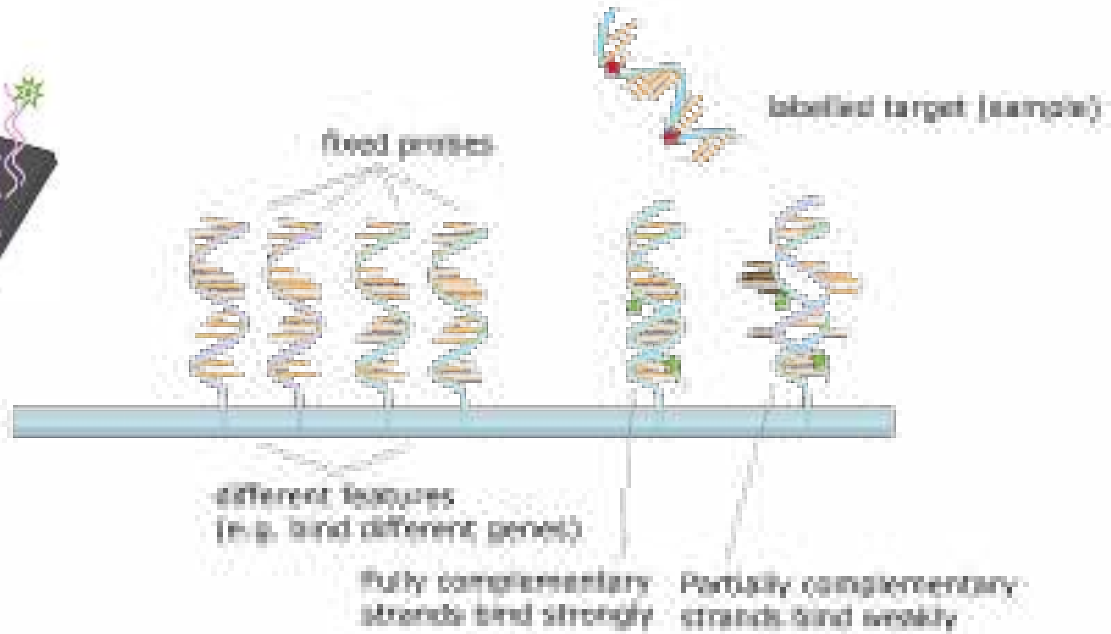


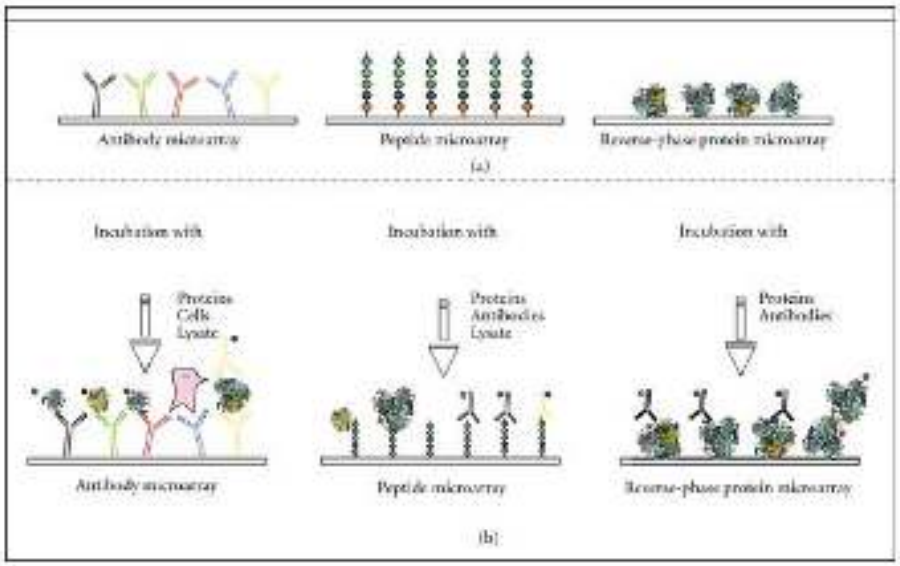
Fig. 1 of T. Siddika, I.U. Heinemann IU (2021)
Bringing MicroRNAs to Light: Methods for MicroRNA Quantification and Visualization in Live Cells.
Front. Bioeng. Biotechnol. 8:619583.
doi: 10.3389/fbioe.2020.619583, CC BY 4.0



https://en.wikipedia.org/wiki/DNA_microarray graphic by Squidonius, public domain

Binding (see also S Fig. 1.9)

Fig. 1 of R. Wellhausen, H. Seitz (2012) Facing Current Quantification Challenges in Protein Microarrays, BioMed Res. Int., 2012:831347, doi: 10.1155/2012/831347 CC BY 3.0



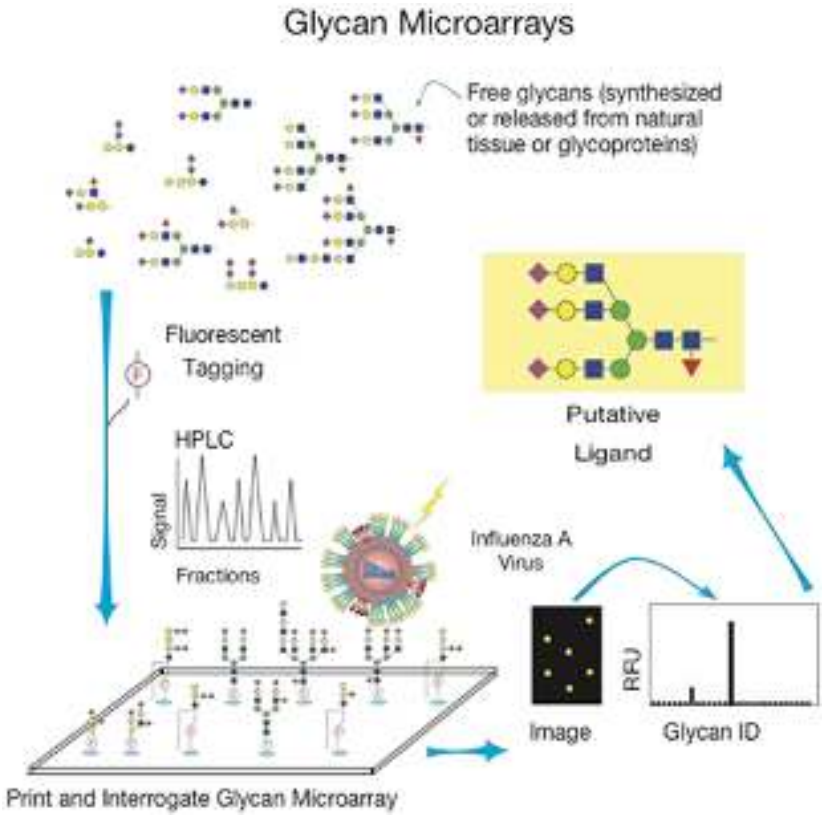
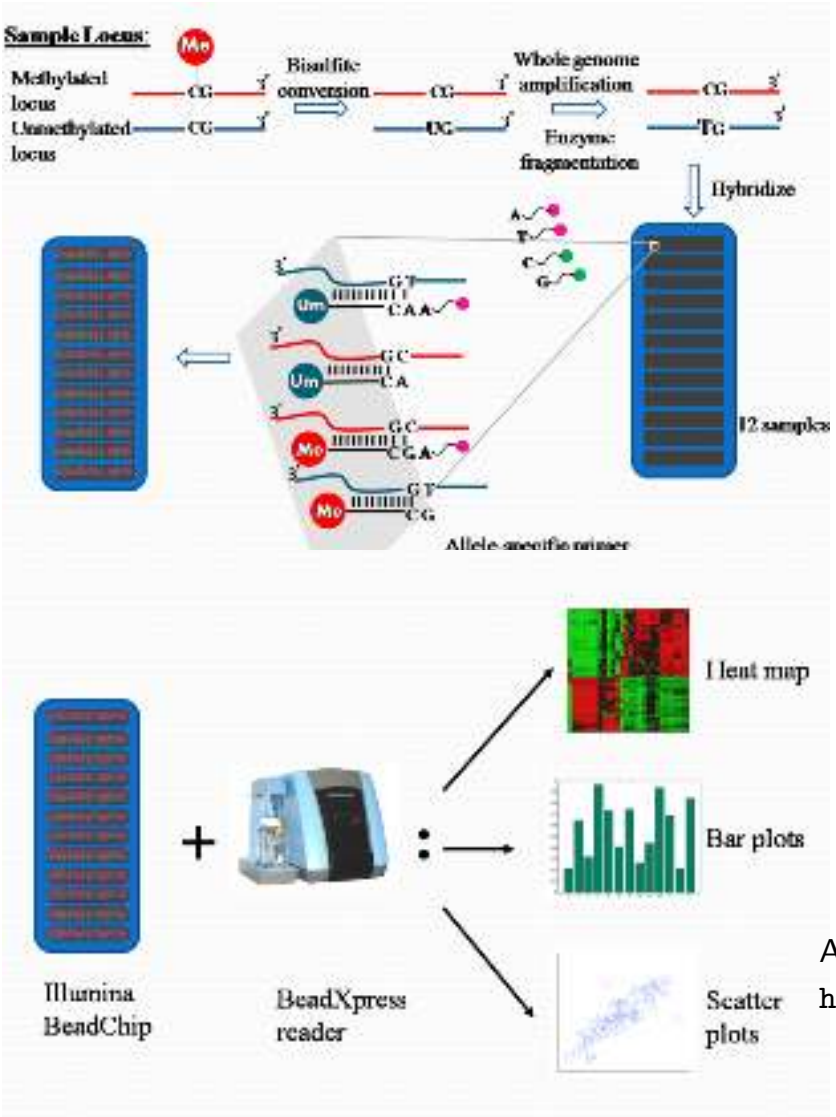
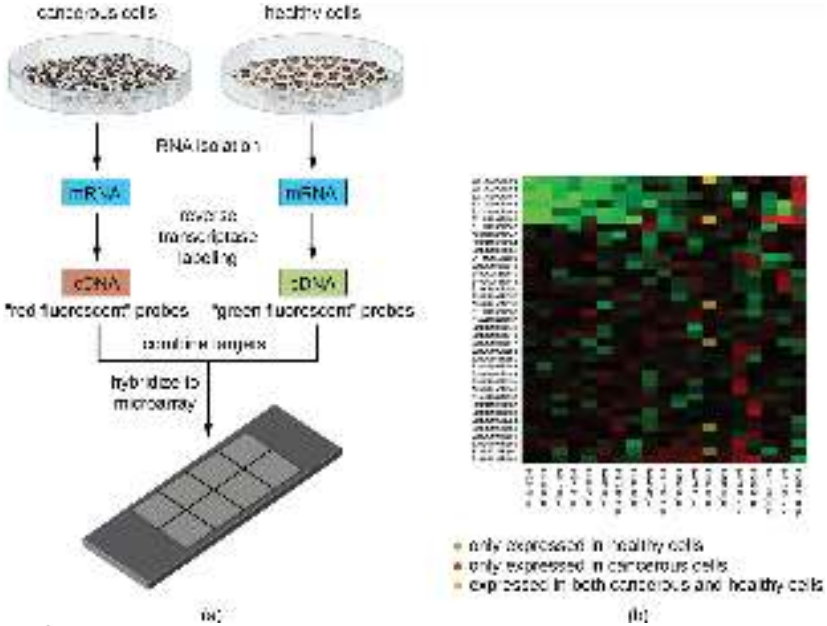


Fig. 2 of A.M. McQuillan, L. Byrd-Leotis, J. Heimburg-Molinaro, R.D. Cummings RD (2019) Natural and Synthetic Sialylated Glycan Microarrays and Their Applications. Front. Mol. Biosci. 6:88. doi: 10.3389/fmolb.2019.00088, CC BY 4.0

Microarray procedures (see also R Plate 2.6)



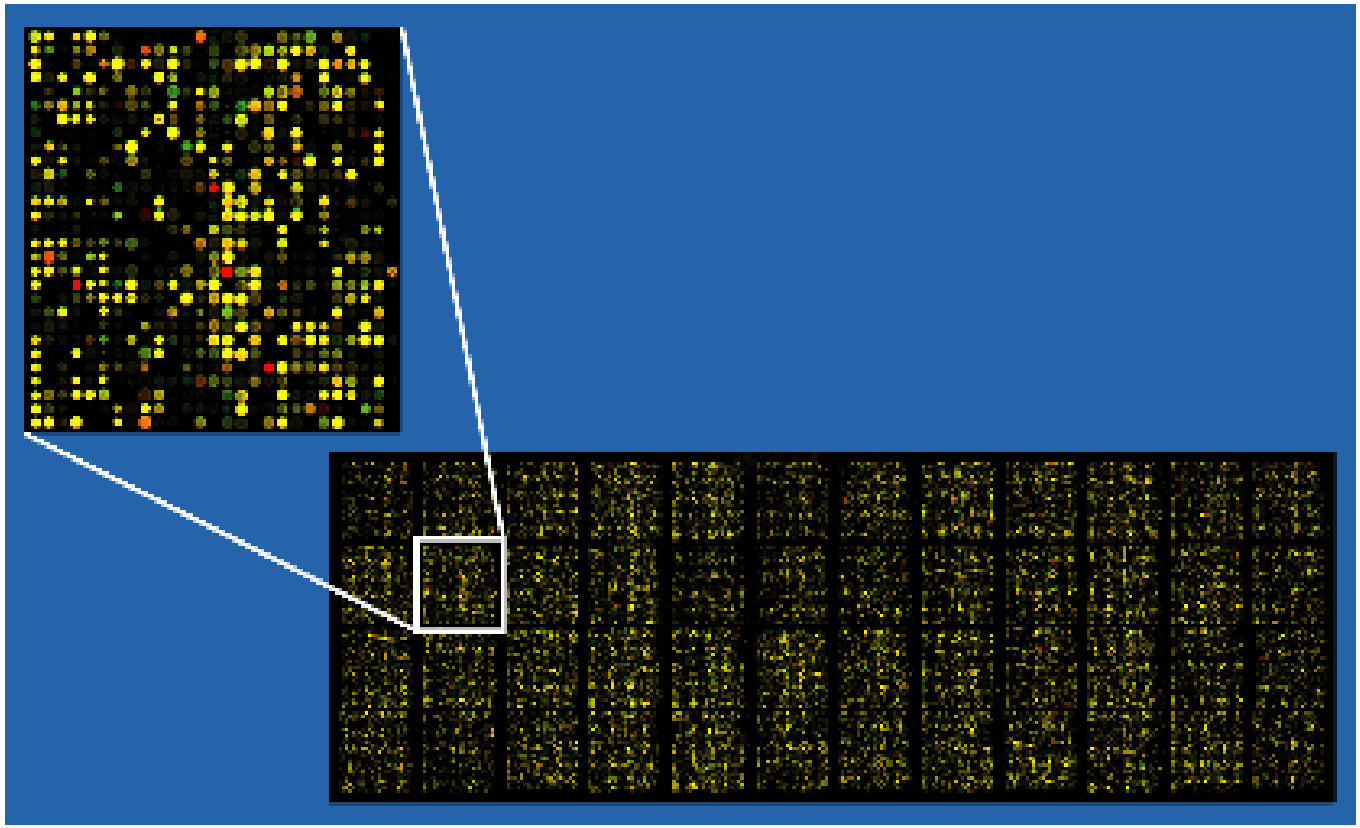
N. Parker, M. Schneegurt, A.-H. Thi Tu, P. Lister, B. M. Forster,
OpenStax, Microbiology, Houston Texas, 2016
Fig. 12.17 in Section 12.2



Access for free at
https://openstax.org/books/microbiology/pages/1-introduction.

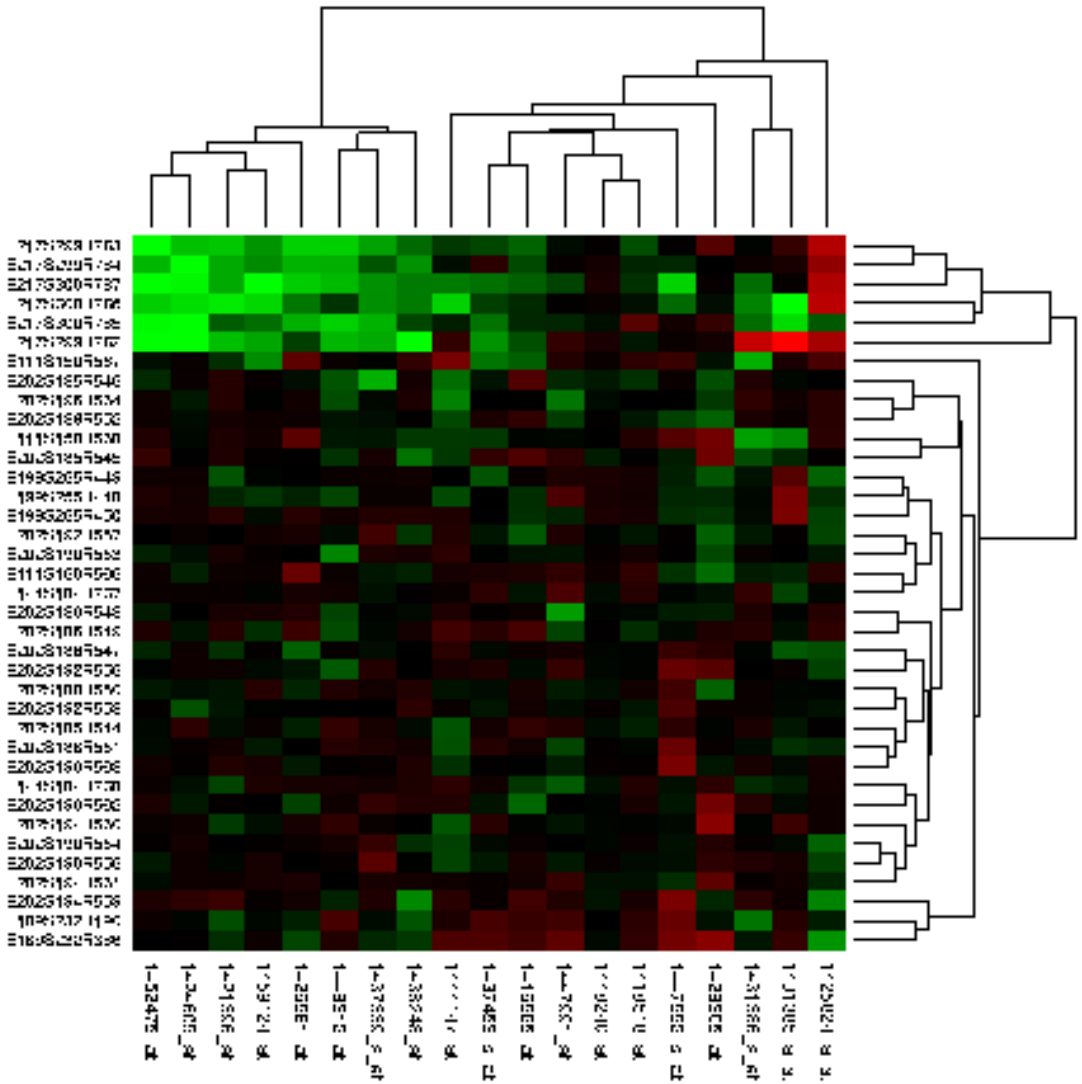
https://en.wikipedia.org/wiki/Illumina_Methylation_Assay by Phoebe Lu, public domain

Microarray results



https://commons.wikimedia.org/w/index.php?curid=1612185, by Paphrag, public domain

Microarray results



https://commons.wikimedia.org/w/index.php?curid=1612199, by Miguel Andrade, public domain

Microarray example technologies

Affymetrix

https://en.wikipedia.org/wiki/DNA_microarray, by Schutz, CC BY 2.5

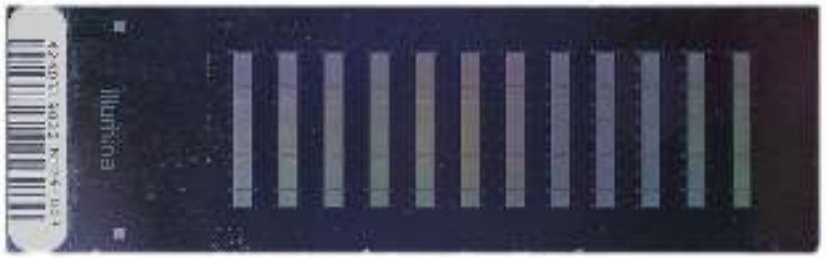


Fig. 1. in
W. Shi, A. Banerjee, M.E. Ritchie, S. Gerondakis, G.K. Smyth (2009), Illumina WG–6 BeadChip strips should be normalized separately.
BMC Bioinformatics 10, 372,
doi: 10.1186/1471-2105-10-372, CC BY 2.0

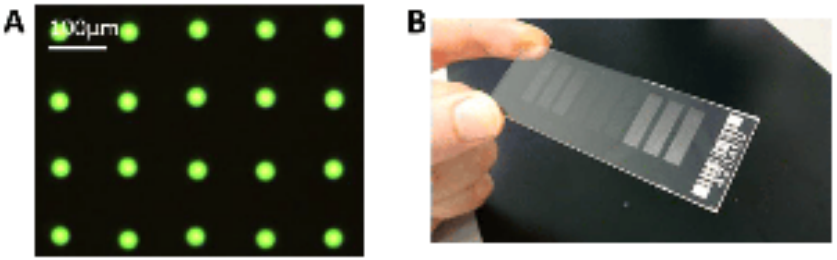
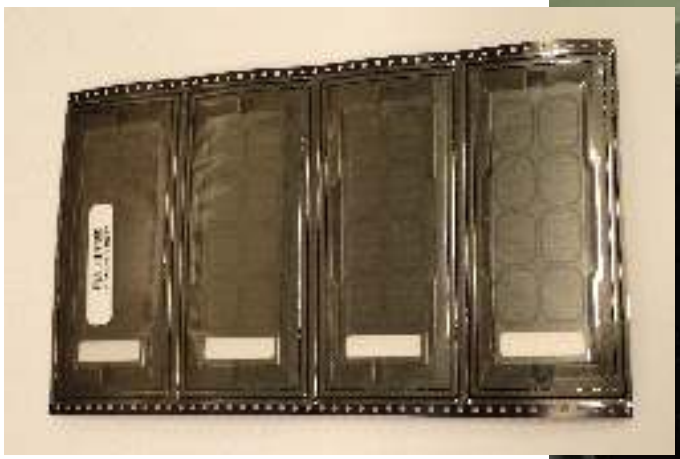


Fig. 6 in I. Hospach, Y. Joseph, M.K. Mai, N. Krasteva, G. Nelles (2014), Fabrication of Homogeneous High-Density Antibody Microarrays for Cytokine Detection.
Microarrays, 3:282-301, doi: 10.3390/microarrays3040282
CC BY 4.0

others: Agilent, Eppendorf, TeleChem, e.t.c.

Microarray example technologies



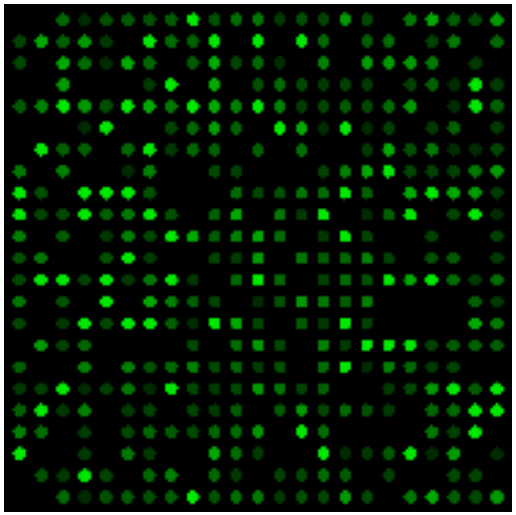
Roche

photographs by Dr. Piotr Madanecki, reproduced with kind permission

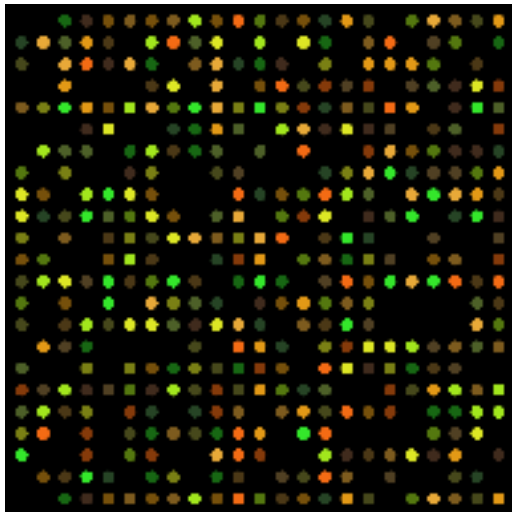
Double vs single channel microarray

Double channel: cDNA from two samples, each labelled with different fluorescent (Cy3 green, Cy5 red)

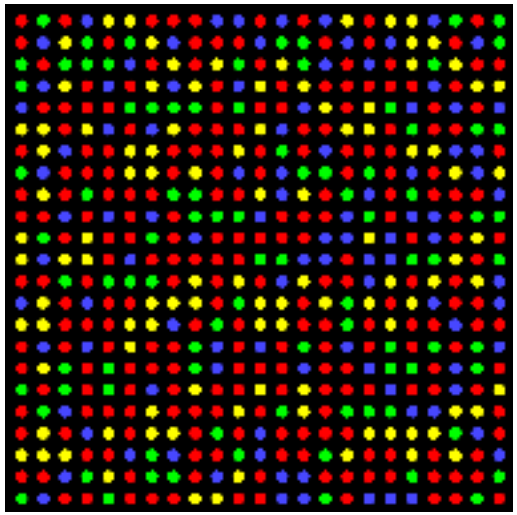
Single channel: intensity data for each probe (not absolute but relative to other samples from same experiment)



Wikimedia Commons,
by Thomas Shafee, CC BY 4.0



Wikimedia Commons,
by Guillaume Paumier (user:guillom), GFDL



Wikimedia Commons,
by Thomas Shafee, CC BY 4.0

See also: 1 H. van Bakel, F. C. P. Holstege. A Tutorial for DNA Microarray Expression Profiling, 2007, Cell Press

Double vs single channel microarray

Double channel

- 2 samples compared directly on same plate
- one sample can affect the raw data of the other
  - one sample high quality, other low quality?
- i samples: $i(i-1)/2$ comparisons

Single channel

- 1 sample per plate
- no interactions between samples
- easier to compare between experiments
- i samples: i runs, choose one as reference
- ensure same conditions for comparison between samples

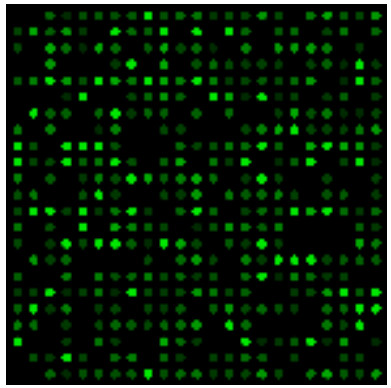
https://en.wikipedia.org/wiki/DNA_microarray

Microarray analysis workflow

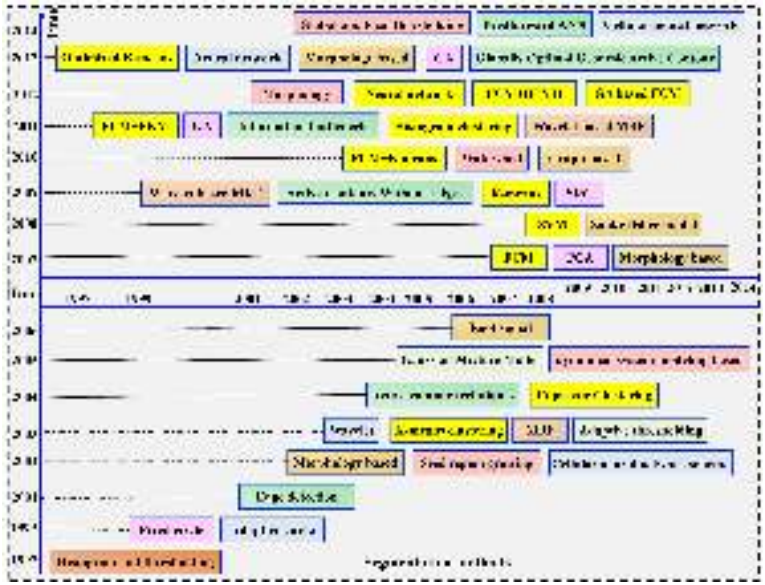
1. Biological question
2. Experimental design
3. Microarray procedure
4. Image analysis (acquiring numerical data)
5. **Normalization, data preprocessing**
6. **Data analysis**
7. Biological interpretation

Image analysis

MATLAB workflow



Wikimedia Commons,
by Thomas Shafee, CC BY 4.0

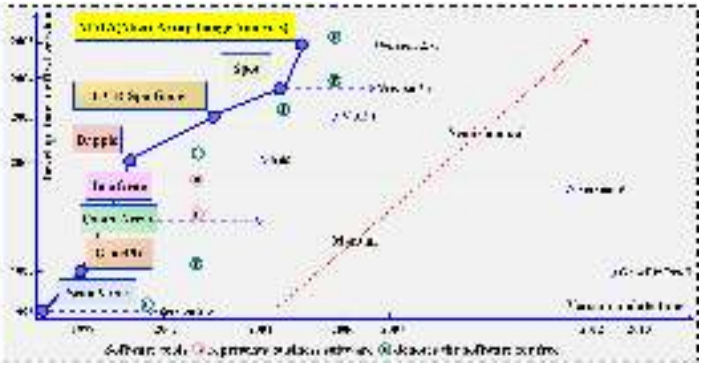


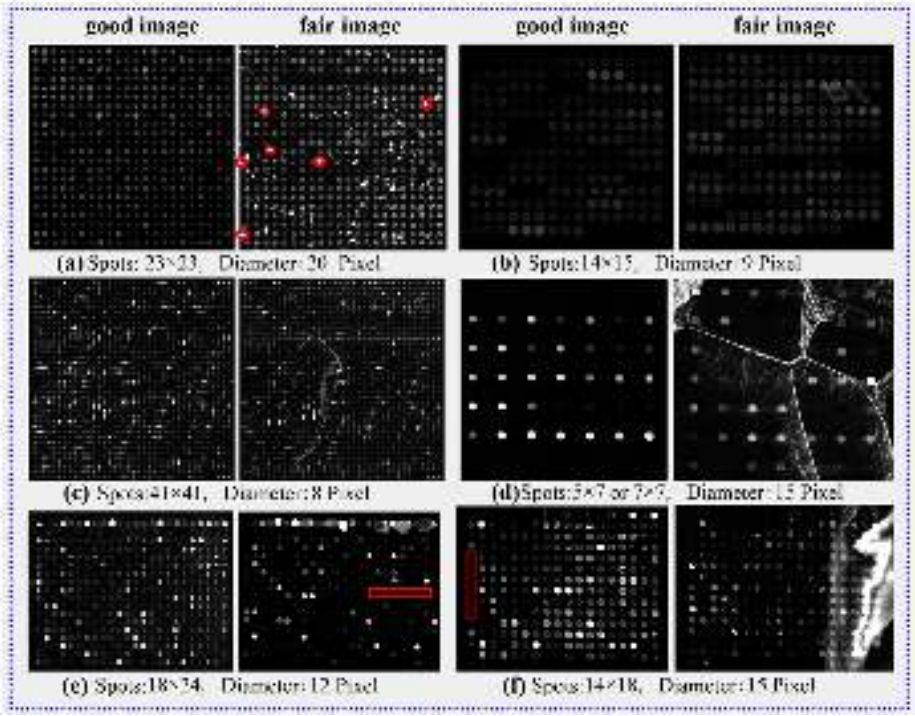
- Addressing
- Gridding and segmentation

e.g, G. Shao, T. Li, W. Zuo, S. Wu, T. Liu (2015) A Combinational Clustering Based Method for cDNA Microarray Image Segmentation. PLoS ONE 10(8): e0133025. doi:10.1371/journal.pone.0133025 : software, segmentation methods history (Fig. 3, above right, public domain); flowcharts; discussion of methods

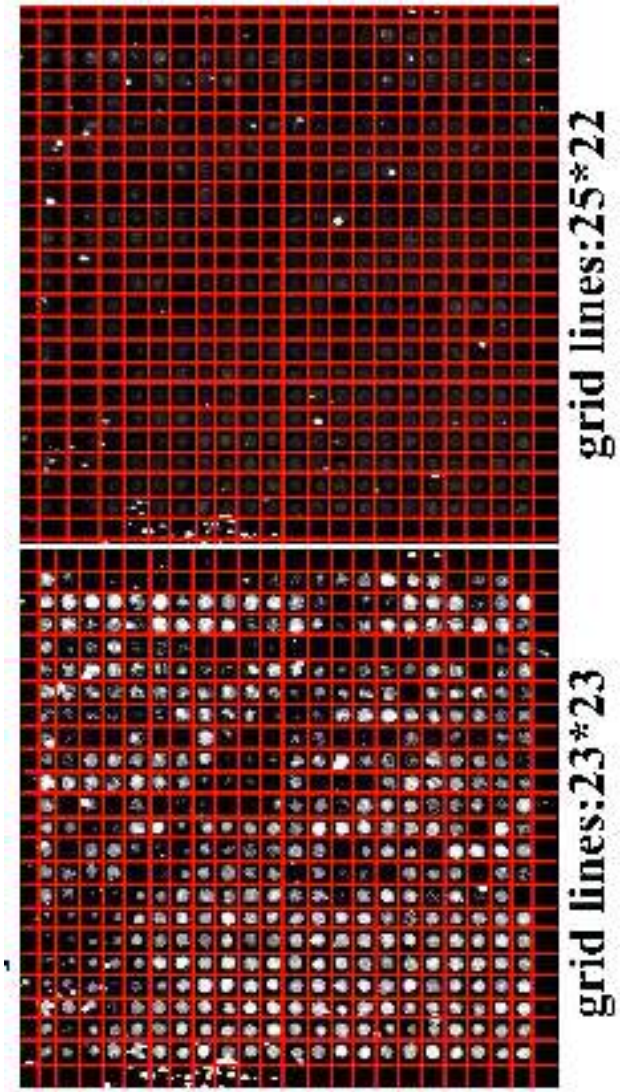
- Problems:
  - Uneven spot sizes, spacings, grid positions, curves in a grid
  - image analysis: fixed/variable size spots, failed signal (no hybridization)
  - image analysis: dust on images (brighter)
  - spatial errors (bias)

Gridding





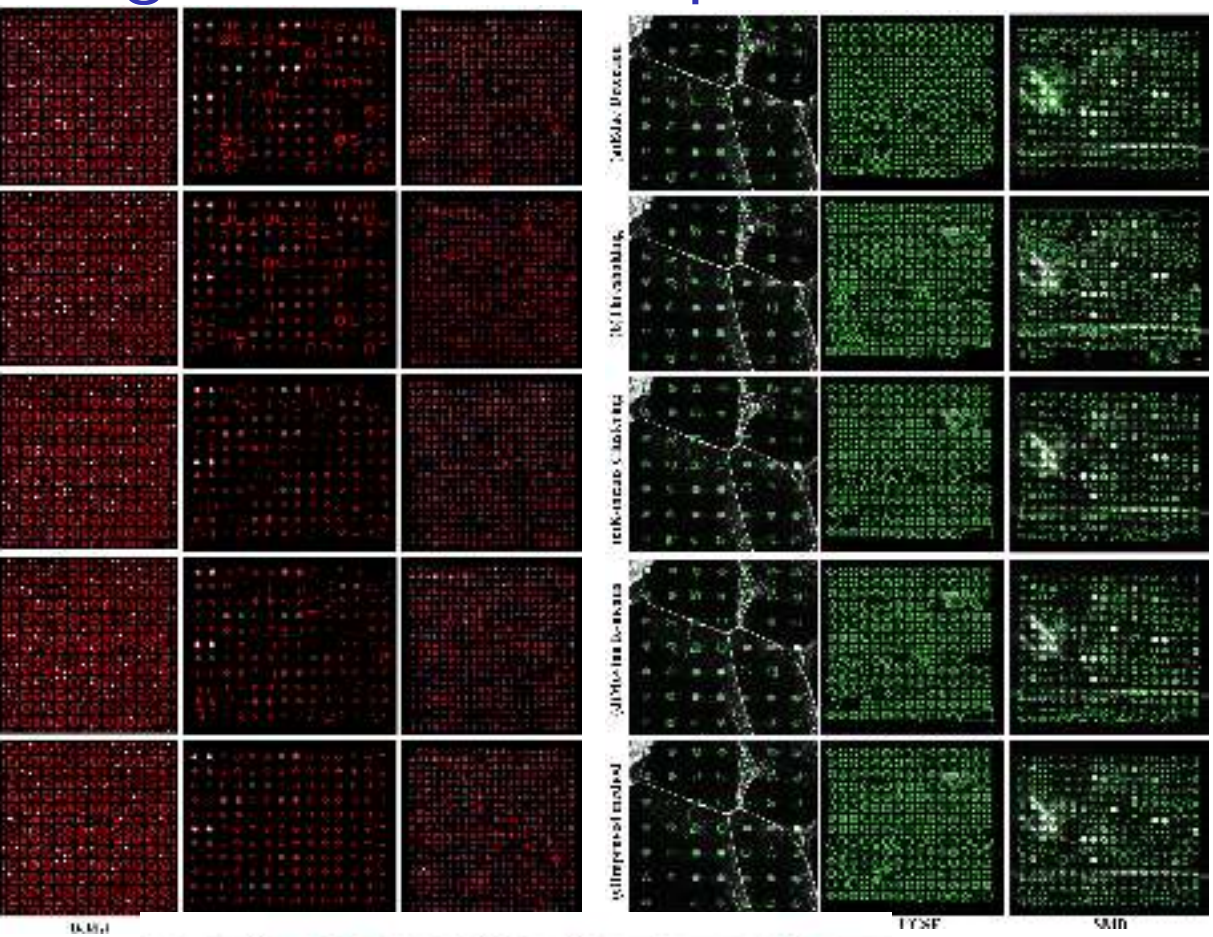
Grids with various quality.

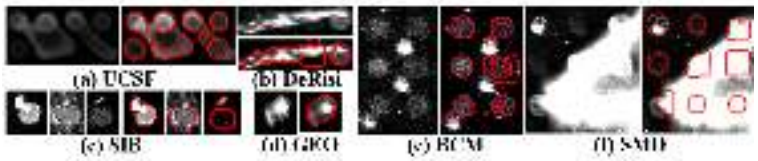


Segmentation without (top) and with (bottom) contrast enhancement.

Figs. 1,2 and 5 in G. Shao, T. Li, W. Zuo, S. Wu, T. Liu (2015) A Combinational Clustering Based Method for cDNA Microarray Image Segmentation. PLoS ONE 10(8): e0133025. doi:10.1371/journal.pone.0133025, public domain

Segmentation and spots





Figs. 7, 9, 10 and 11 in G. Shao, T. Li, W. Zuo, S. Wu, T. Liu (2015) A Combinational Clustering Based Method for cDNA Microarray Image Segmentation. PLoS ONE 10(8): e0133025. doi:10.1371/journal.pone.0133025, public domain

Data to analyze

For each spot, i, two values :

red intensity; $Cy5_i$

green intensity; $Cy3_i$

$$S_i = \log \frac{Cy5_i}{Cy3_i}$$

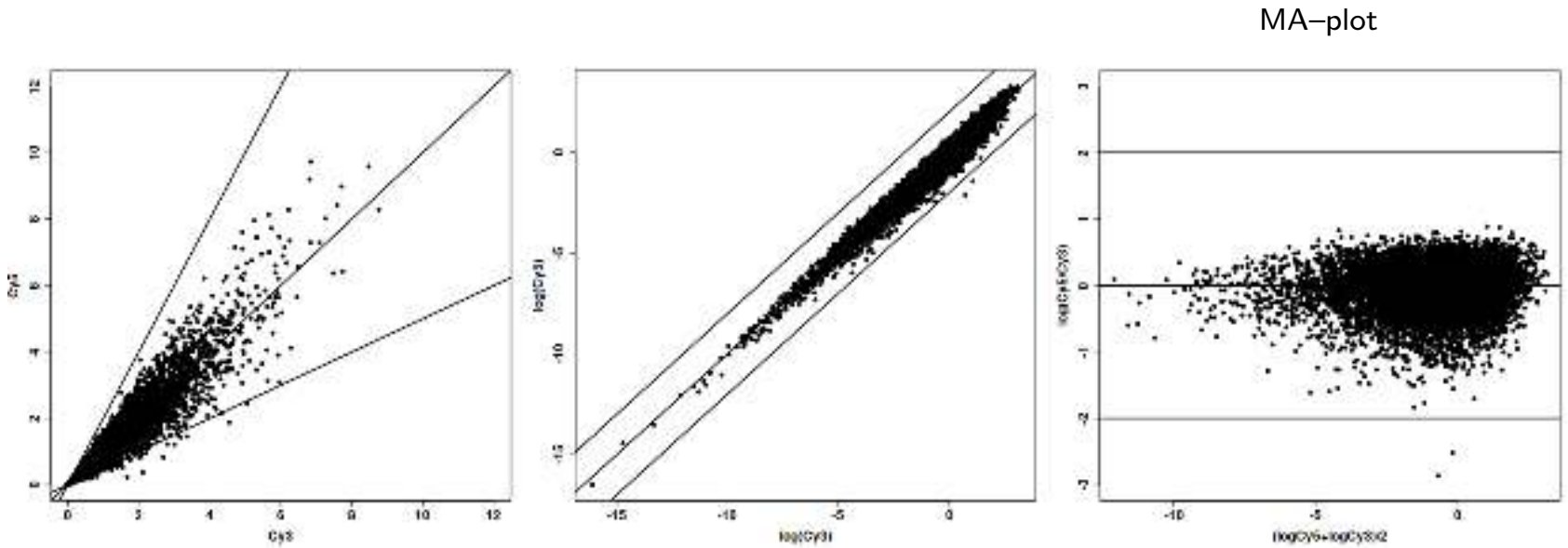
See also MATLAB workflow

https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/2573/versions/3/previews/R14_MicroarrayImage_CaseStudy/html/R14_MicroarrayImage_CaseStudy.html?access_key=

Normalization (EG 13.1.3)

1. Array–specific effects: no two arrays are identical
2. Gene–specific effects: hybridization conditions cannot be optimized at once for all elements
3. Dye–specific effects
4. Background noise
5. Preparation effects: operator, weather, time of day, e.t.c. (microarrays are sensitive)

Close to ideal (see also X: Fig. 18.5)



Cy3: green, Cy5: red

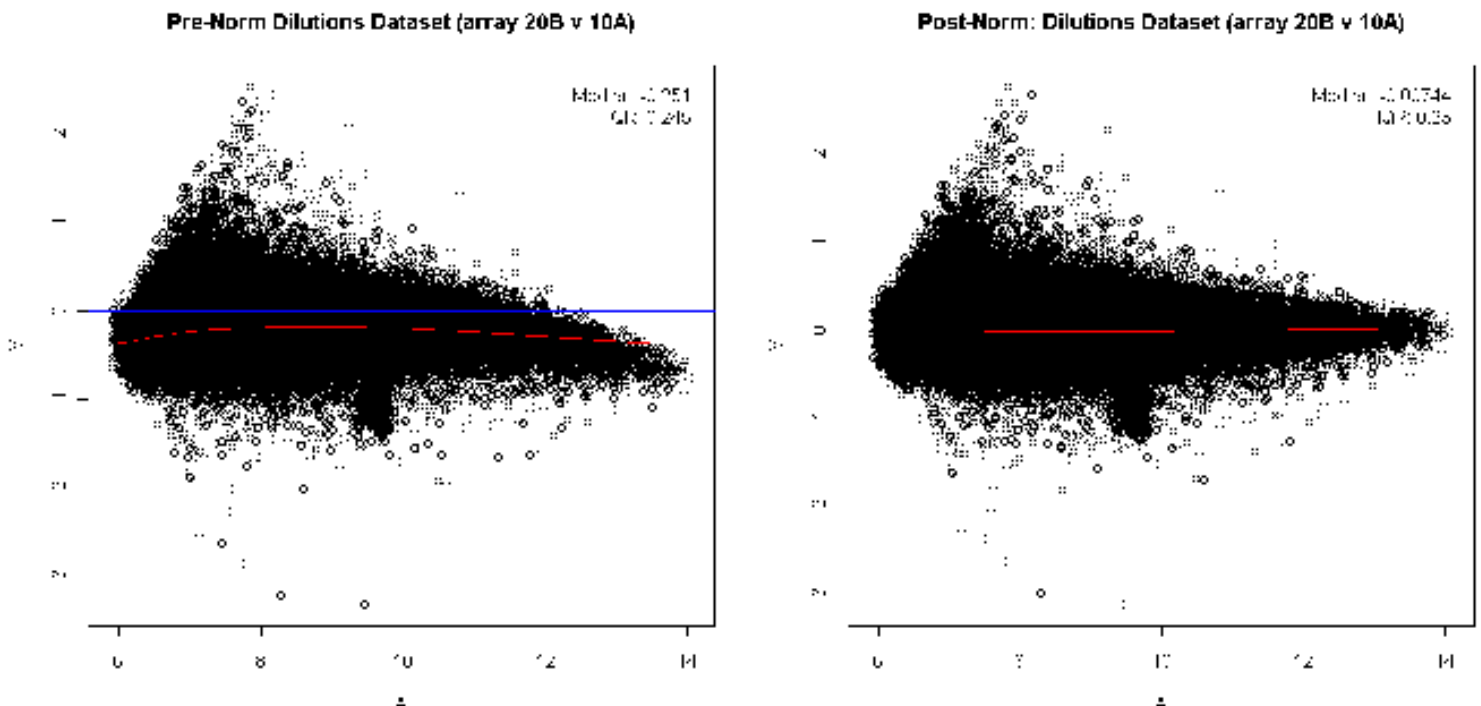
MA–plot: $M \equiv \log_2(Cy5/Cy3), \quad A \equiv (\log_2 Cy5 + \log_2 Cy3)/2$

code: `732A51_BioinformaticsHT2023_Lecture06codeSlide22CloseIdealMicroarray.R`

Normalization MA–plot, loess curve (see also EG Figs 13.4–6)

$$\text{y}-\text{axis } M = \log_2\left(\frac{\text{Green}(gene)}{\text{Red}(gene)}\right) = \log_2 \text{Green}(gene) - \log_2 \text{Red}(gene)$$

$$\text{x}-\text{axis } A = (\log_2 \text{Green}(gene) + \log_2 \text{Red}(gene))/2$$



Bioconductor R code: https://en.wikipedia.org/wiki/MA_plot, by Zoolium, public domain

Array level normalization

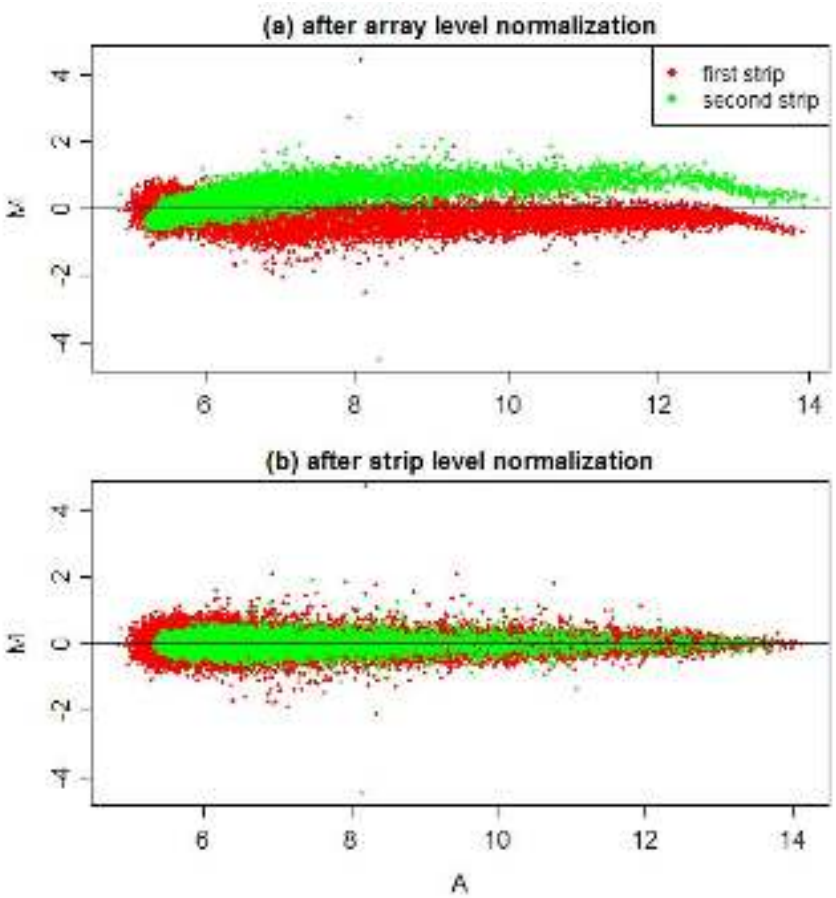


Fig. 3 in
W. Shi, A. Banerjee, M.E. Ritchie, S. Gerondakis, G.K. Smyth (2009), Illumina WG–6 BeadChip strips should be normalized separately.
BMC Bioinformatics 10:372,
doi: 10.1186/1471-2105-10-372, CC BY 2.0

Removing intensity dependent bias

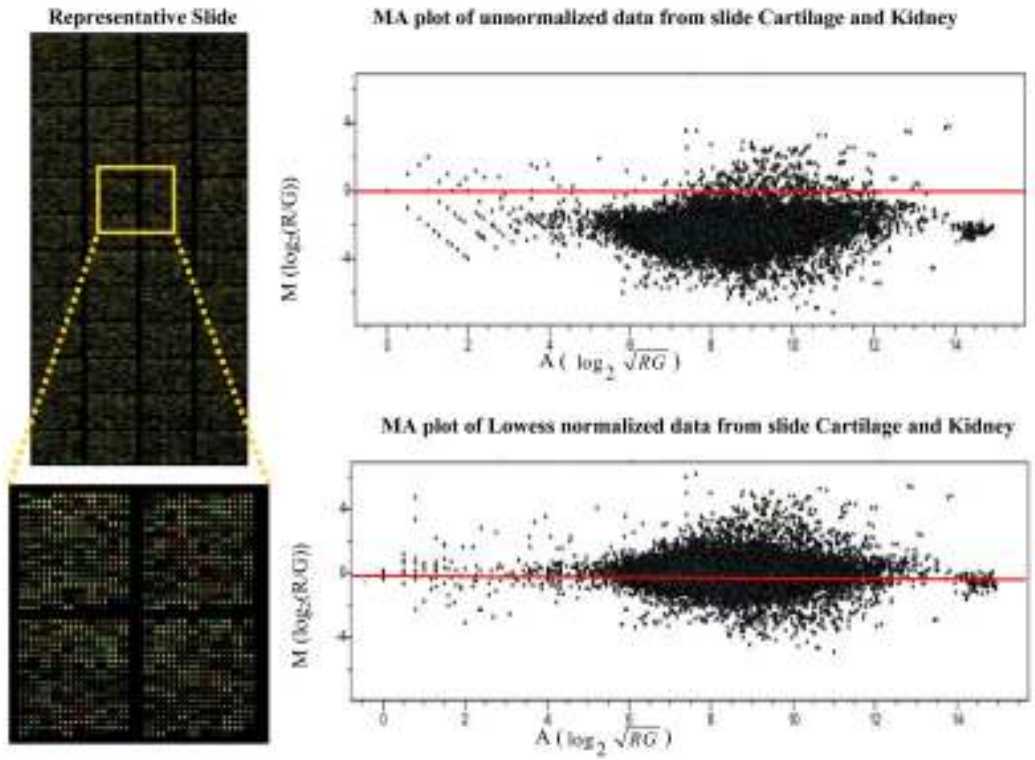


Fig. 1 in
L. Huang, W. Zhu, C.P. Saunders, J.N. MacLeod, M. Zhou, A.J. Stromberg, A.C. Bathke (2008), A novel application of quantile regression for identification of biomarkers exemplified by equine cartilage microarray data.
BMC Bioinformatics 9, 300,doi: 10.1186/1471-2105-9-300, CC BY 2.0

Quality control

1. Number of outliers (expect few)

2. Few missing values, no empty portions on array

3. Controls on microarray
    a. housekeeping genes: expect expressed (positive control)
    b. negative controls: no signal expected
    c. Cy3 labelled control: signal independent of sample control
    d. low stringency: should give low signal
    e. and others

based on Analysing data from Illumina BeadArrays, Matt Ritchie, MPhil in Computational Biology, University of Cambridge
See also `https://www.illumina.com/documents/products/technotes/technote_gene_expression_data_quality_control.pdf`

Is a gene expressed (Affymetrix)? (EG Ch. 13.2.2)

Gene–specific negative controls/mismatch probes

Multiple match and mismatch probes for each gene

Is the "match probes" expression significantly higher than the "mismatch probes" expression?

H_0: gene is not expressed

Wilcoxon signed–rank test

Interpret failure of rejecting H_0 as *Absent*, reject H_0 as *Present*

Differential expression (EG Ch. 13.2.3)

Single gene, two samples (case and control)

Is the case expression significantly different from the control one?

In principle: same as previous slide, Wilcoxon signed–rank test

- Normalize arrays w.r.t. each other
  - Two channel probes.
- What if only one probe per gene?
  - Multiple genes at once
  - (assume most are not differentially expressed)

Tests

1. t–test (difference in mean)

2. ANOVA approach

$$X_{gik} = \mu + \tau_k + A_i + D_{ik} + E_{gik}$$

gene g, array i, condition k
(array specific effects)

See `limma` package in `Bioconductor` for setting up linear models

Multiple genes (EG Ch. 13.3.1)

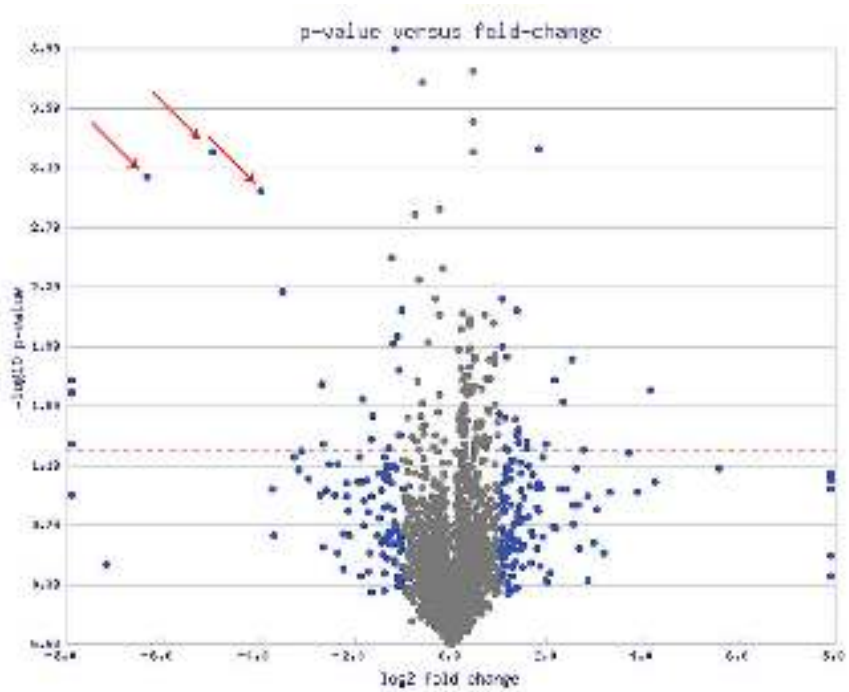
Rank genes

in decreasing order of t–test statistic

in increasing order of p–values
(if different null hypothesis distributions for genes)

Cut–off: 0.05 significance?

Volcano plot



" Example of a Volcano plot, here showing metabolomic data. The three red arrows indicate points that display both large–magnitude fold–changes (x-axis) as well as high statistical significance ($-\log_{10}$ of p–value, y–axis). The dashed red–line shows where $p = 0.05$ with points above the line having $p < 0.05$ and points below the line having $p > 0.05$. This plot is colored such that those points having a fold-change less than 2 ($log_2 2 = 1$) are shown in gray."
`https://en.wikipedia.org/wiki/Volcano_plot_(statistics)`, by Roadnottaken, public domain
See also `http://bioinformatics.knowledgeblog.org/2011/06/21/volcano-plots-of-microarray-data/`

Multiple testing (EG Ch. 3.11)

Assume type I error of $\alpha = 0.01$

Perform $n = 1000$ tests: 1000 p–values

Under null distribution: p–value$\sim \mathrm{Unif}[0, 1]$

Significant calls expected by chance $S \sim \mathrm{Binomial}(n, \alpha)$

$\mathrm{E}[\mathrm{S}] = n\alpha = 10$

FWER (EG Ch. 3.11)

Family–wise error rate (FWER): probability of at least one null hypothesis rejected when all are true

Aim: FWER at level α

Bonferroni correction
No assumption of independence between tests

Individual significance call at α/n

FWER$\leq \sum_{i=1}^{n}(\alpha/n) = \alpha$

Šidák procedure (EG Ch. 3.11)

Assume independent tests

Individual significance call at $K(n, \alpha) = 1 - \sqrt[n]{1-\alpha}$

$$\text{FWER}=1-\text{accept all} = 1-\prod_{i=1}^{n}(1-K(n,\alpha)) = 1-\prod_{i=1}^{n}\sqrt[n]{1-\alpha} = \alpha$$

FWER control (EG Ch. 13.3.5)

g: number of genes

1. Bonferroni: individual cut–off$= \alpha/g$
2. Sidák: individual cut–off$= 1 - \sqrt[g]{1-\alpha}$

But these approaches can be too conservative!

Few (ca 3, 4) replicates per gene

FWER framework: cannot allow assumption of *some* false positives

FDR (EG Ch. 13.3.6)

False discovery rate (FDR): controls % of false positives

10000 genes, 100 differentially expressed

FWER for $\alpha = 0.01$ will give 1 true positive

FDR of 50% will give 50 *candidate* genes

FOLLOW UP!!

Benjamini–Hochberg (EG Ch. 13.3.6)

g tests, α desired FDR

order p–values $(P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(g)})$
let $H_{(i)}$ be corresponding null hypothesis

define $q_i = \frac{i}{g}\alpha$

Procedure:

1. If for all i $p_{(i)} \leq q_i$ do **not** reject **any** H_0
2. Else find *largest* k s.t. $p_{(k)} \leq q_k$
3. Reject $H_{(1)}, \ldots, H_{(k)}$

NOTE: there may be j s.t. $p_{(j)} > q_j$

SAM (EG Ch. 13.3.6)

Significance analysis of microarrays for gene i

$$d(i) = \frac{\bar{x}_i - \bar{y}_i}{s(i) - s_0}$$

$$s(i) = \frac{1}{n_{ix} + n_{iy} - 2} \left(\sum_{j=1}^{n_{ix}} (x_{ji} - \bar{x}_i)^2 + \sum_{j=1}^{n_{iy}} (y_{ji} - \bar{y}_i)^2 \right)$$

SAM (EG Ch. 13.3.6)

1. Order test statistics, $d(i)$ according to magnitude

2. Consider all permutations of the data's columns (i.e. between the groups)
3. Calculate for each permutation the $d(i)$ and rank

4. Calculate for each row (gene) the average $d_E(i)$ and then take $d_{org}(i) - d_E(i)$
5. Choose threshold $\Delta < |d_{org}(i) - d_E(i)|$ for calling a gene significant
6. Find FDR (EG p. 465)

choice of s_0 "moderate" value so it has an effect but not too large
ideally to maximize power, no known formula

https://en.wikipedia.org/wiki/Significance_analysis_of_microarrays

ANOVA (EG Ch. 13.3.7)

$$X_{ijkg} = \mu + A_i + \delta_j + \tau_k + \gamma_g + B_{ig} + \psi_{kg} + E_{gik}$$

array i, dye j, condition k, gene g

X: logarithm of gene expression

Replicates

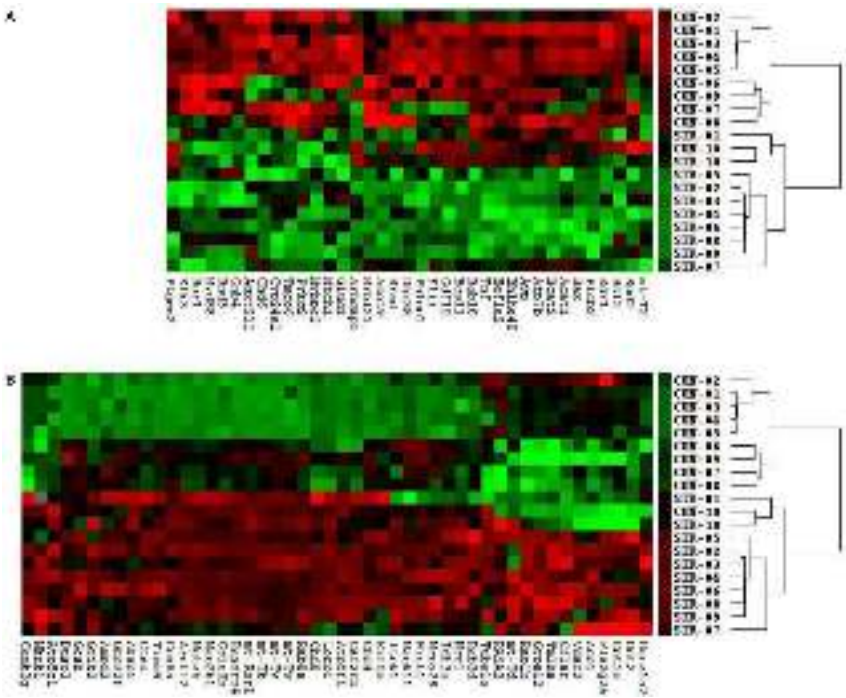
Biological replicates

versus

Technical replicates

Clustering (see also MM Fig 10–6)

cluster samples: e.g. identify tumour classes
cluster genes: e.g. groups of co–regulated genes



Genes distinguishing stressed (A) amygdala *(part of brain)* specimens from controls (B).

Fig. 2 in H. Li, X. Li, S.E. Smerin, L. Zhang, M. Jia, G. Xing, Y.A. Su, J. Wen, D. Benedek, R. Ursano (2014), Mitochondrial expression profiles and metabolic pathways in the amygdala associated with exaggerated fear in an animal model of PTSD. Front. Neurol. 5:164. doi: 10.3389/fneur.2014.00164, CC BY 4.0.

Clustering (EG Ch. 13.3.8)

Create a dissimilarity metric between genes/samples :

1−correlation, Euclidean distance

Hierarchical clustering

e.g. Tree construction algorithms (clades are clusters)

Partitioning algorithms

e.g. K–means

Cluster validation:

Statistical

Biological: look at functional categories of clustered genes

Preprocessing for clustering

1. normalize (remove background)

2. filter: remove genes with low variability and with many missing values

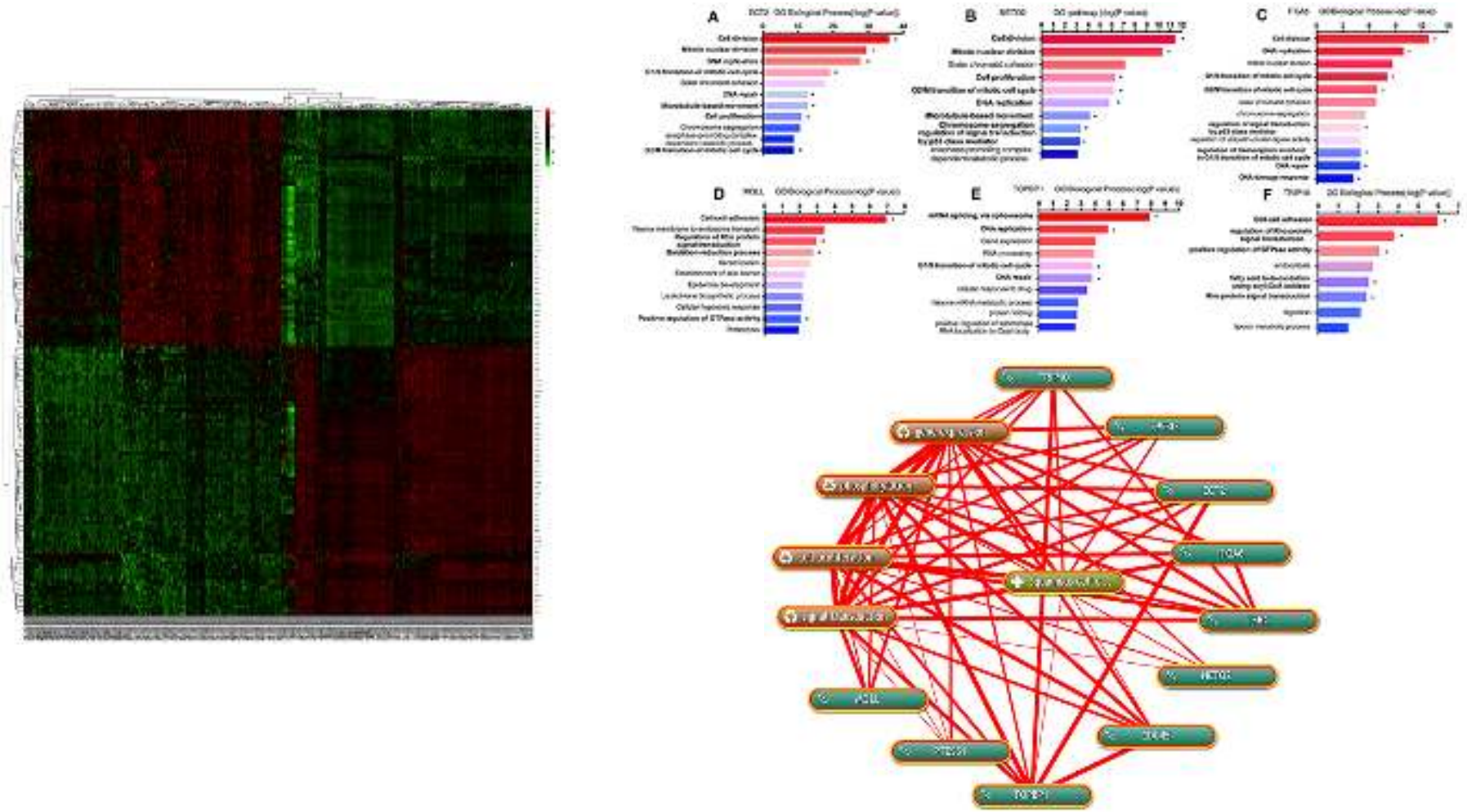
3. impute missing values

4. standardize (z–score)

$$\frac{x - \bar{x}}{\mathrm{sd}(x)}$$

based on Clustering microarray data, MPhil in Computational Biology, University of Cambridge

Text analysis of scientific literature (see also R Plate 1.2)



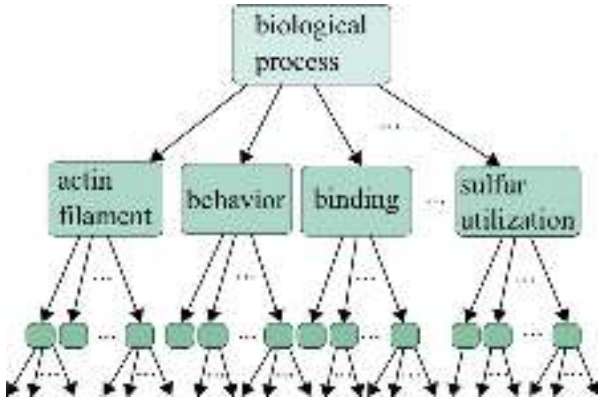
Figs. 2, 5, 8 in Qixing M., Gaochao D., Wenjie X., Anpeng W., Bing C., Weidong M., Lin X., Feng J. Microarray analyses reveal genes related to progression and prognosis of esophageal squamous cell carcinoma. Oncotarget. 2017; 8: 78838-78850. doi: 10.18632/oncotarget.20232, CC BY 3.0.

Gene Ontology (see also R Fig. 8.1)

Gene Ontology Consortium `http://www.geneontology.org/`

`http://geneontology.org/docs/ontology-documentation/`,
CC BY 4.0.

Applied Bioinformatics by D. A. Hendrix , CC BY 4.0.



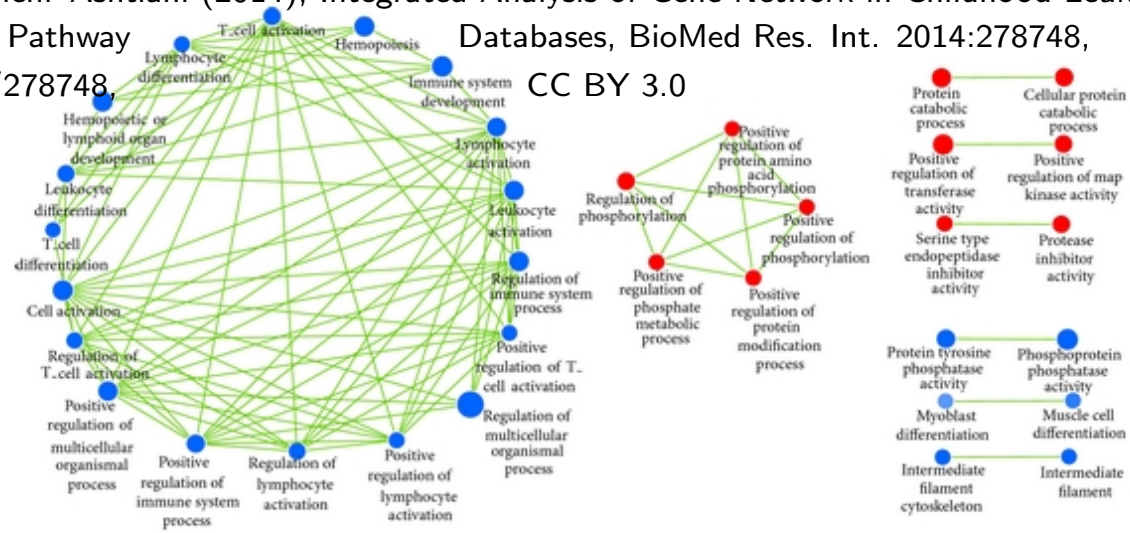
KEGG: Kyoto Encyclopedia of Genes and Genomes

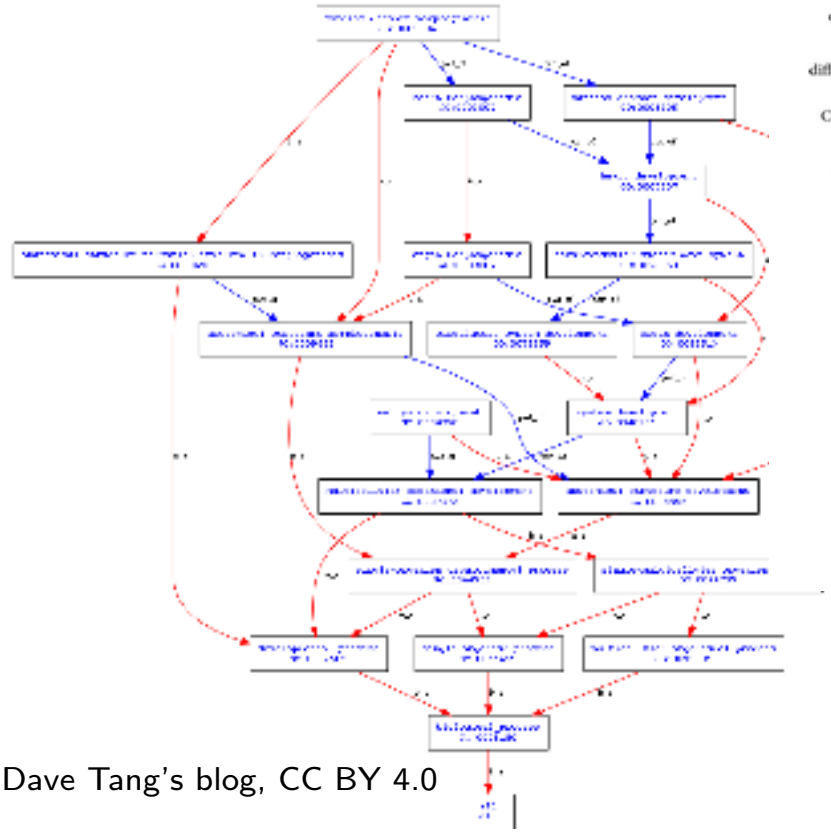
`https://www.genome.jp/kegg/`

"KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular–level information, especially large–scale molecular datasets generated by genome sequencing and other high–throughput experimental technologies."

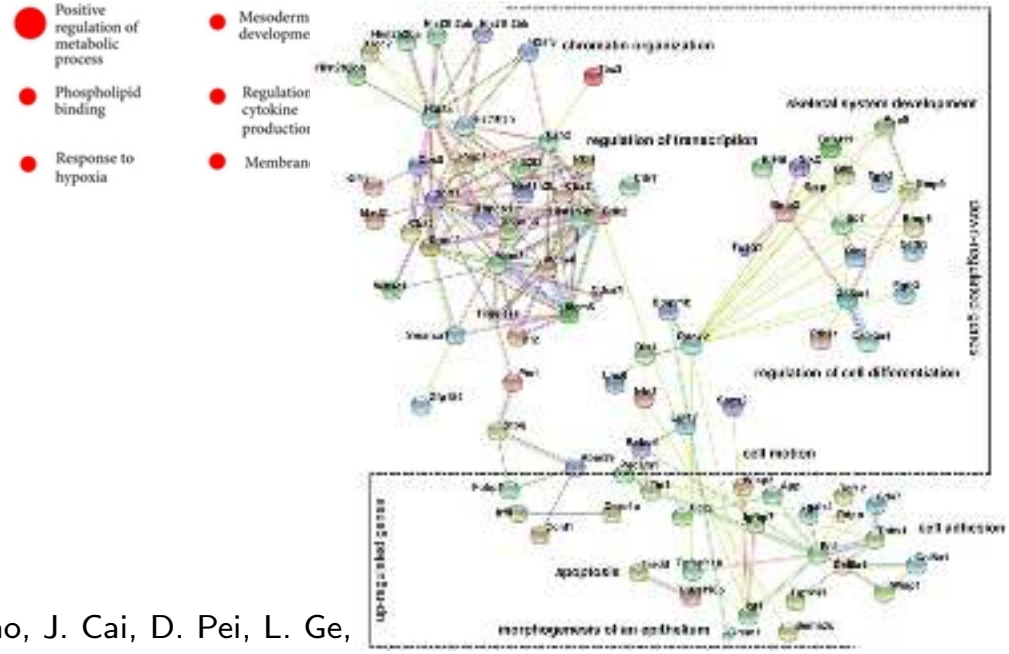
Gene Ontology

A. Chaiboonchoe, S. Samarasinghe, D. Kulasiri, K. Salehi–Ashtiani (2014), Integrated Analysis of Gene Network in Childhood Leukemia from Microarray and Pathway Databases, BioMed Res. Int. 2014:278748, doi: 10.1155/2014/278748, CC BY 3.0





Dave Tang's blog, CC BY 4.0



Y. Zheng, L. Jia, P. Liu, D. Yang, W. Hu, S. Chen, Y. Zhao, J. Cai, D. Pei, L. Ge, S. Wei (2016),Insight into the maintenance of odontogenic potential in mouse dental mesenchymal cells based on transcriptomic analysis. PeerJ 4:e1684, doi: 10.7717/peerj.1684, CC BY 3.0

Functional enrichment analysis

Gene set enrichment analysis (GESA)

Statistically identify GO–terms that are over/under–represented in a set (e.g up/down–regulated) of genes

Tools:
`en.wikipedia.org/wiki/Gene_set_enrichment_analysis#Tools_for_performing_GSEA`
`www.geneontology.org/page/go-enrichment-analysis`

Functional enrichment analysis: Basic method

Input: gene list ordered by expression levels,
gene set (e.g. sample, pathway, location, GO terms)

1. Calculate the *enrichment score* (ES) as the amount of overrepresentation of the genes from the set at the bottom or top of the list ordered by expression levels.
2. Find statistical significance of *ES* by a permutation test (assign random group to gene).
3. Correct for multiple testing (if analyzing multiple gene sets). Normalize ES values and calculate FDR.

`https://en.wikipedia.org/wiki/Gene_set_enrichment_analysis#Methods_of_GSEA`

A. Subramanian et al., 2005.Gene set enrichment analysis: A knowledge–based approach for interpreting genome–wide expression profiles. PNAS 102:43, 15545–15550., grouping called phenotype

Functional enrichment analysis

	Term2	Count	P-value	Genes	Fold Enrichment	FDR
ECT	GO:0004871~signal transducer activity	10	1.59E-15	STAT6, STAT4, STAT5A, STAT5B, SH2B3, SH2B2, SH2B1, STAT1, STAT3, STAT2	66.13636	1.20E-12
ECT	GO:0003700~transcription factor activity, sequence-specific DNA binding	7	4.19E-06	STAT6, STAT4, STAT5A, STAT5B, STAT1, STAT3, STAT2	13.34279	0.003235
ECT	GO:0035591~signaling adaptor activity	3	2.96E-05	SH2B3, SH2B2, SH2B1	327.375	0.022992
ECT	GO:0003677~DNA binding	6	3.80E-04	STAT6, STAT5A, STAT5B, STAT1, STAT3, STAT2	8.184375	0.293076
ECT	GO:0035870~SH3/SH2 adaptor activity	2	0.005335	SH2D1A, BLNK	349.2	4.046197
ECT	GO:0019221~cytokine-mediated signaling pathway	5	1.23E-05	STAT4, STAT5A, STAT5B, SH2B2, STAT1	31.91136	0.014889
ECT	GO:0007259~JAK-STAT cascade	3	1.24E-04	STAT5A, STAT5B, STAT1		
ECT	GO:0035556~intracellular signal transduction	5	3.81E-04	SH2B3, SH2B2, CLNK, SH		
ECT	GO:0045931~positive regulation of mitotic cell cycle	3	4.73E-04	SH2, STAT5A, STAT5B		
ECT	GO:0006351~transcription, DNA-templated	6	8.29E-04	STAT6, STAT5A, STAT5B, STAT2		
ECT	GO:0005737~cytoplasm	9	0.01126	SH2D4A, STAT6, SH2D1A, STAT5B, STAT1, STAT3, ST		
ECT	GO:0000790~nuclear chromatin	3	0.011556	STAT6, STAT1, STAT3		

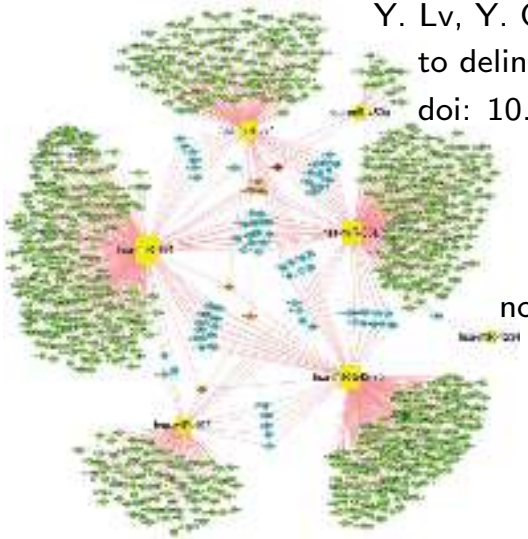
ions; BP, biological processes; CC, cellular component.

Category	Term2	Count	P-Value	Genes	Fold Enrichment	FDR
KEGG_PATHWAY	bta05161:Hepatitis B	8	8.53E-09	STAT6, STAT4, STAT5A, STAT5B, PIK3R3, STAT1, STAT3, STAT2	24.09836	1.00E-06
KEGG_PATHWAY	bta04630:Jak-STAT signaling pathway	8	1.10E-08	STAT6, STAT4, STAT5A, STAT5B, PIK3R3, STAT1, STAT3, STAT2	23.52941	1.15E-06
KEGG_PATHWAY	bta05162:Measles	7	2.51E-07	SH2D1A, STAT5A, STAT5B, PIK3R3, STAT1, STAT3, STAT2	22.20588	2.63E-04
KEGG_PATHWAY	bta04917:Prolactin signaling pathway	5	1.42E-05	STAT5A, STAT5B, PIK3R3, STAT1, STAT3	30.00795	0.014858
KEGG_PATHWAY	bta05221:Acute myeloid leukemia	4	2.02E-04	STAT5A, STAT5B, PIK3R3, STAT3	31.72269	0.211984
KEGG_PATHWAY	bta05321:Inflammatory bowel disease (IBD)	4	3.92E-04	STAT6, STAT4, STAT1, STAT3	25.37815	0.410699
KEGG_PATHWAY	bta04062:Chemokine signaling pathway	5	5.04E-04	STAT5B, PIK3R3, STAT1, STAT3, STAT2	12.00319	0.527802
KEGG_PATHWAY	bta04650:Natural killer cell mediated cytotoxicity	4	0.001753	SH2D1A, PIK3R3, SH2D1B, SH3BP2	15.18351	1.824584
KEGG_PATHWAY	bta04722:Neurotrophin signaling pathway	4	0.002119	SH2B3, SH2B2, SH2B1, PIK3R3	14.21176	2.201949
KEGG_PATHWAY	bta05160:Hepatitis C	4	0.03253	PIK3R3, STAT1, STAT3, STAT2	13.39802	2.823742

KEGG, Kyoto Encyclopedia of Genes and Genomes.

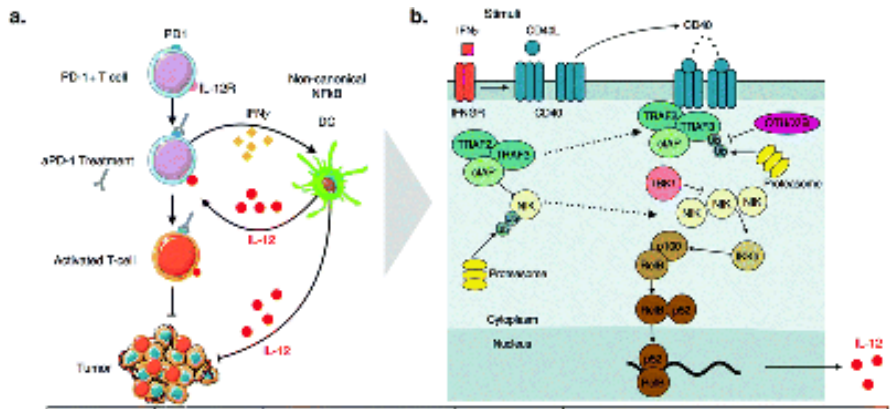
W. Ji, Y. Liu, B. Xu, J. Mei, C. Cheng, Y. Xiao, K. Yang, W. Huang, J. Jiao, H. Liu, J. Shao J (2021) Bioinformatics Analysis of Expression Profiles and Prognostic Values of the Signal Transducer and Activator of Transcription Family Genes in Glioma. Front. Genet. 12:625234, doi: 10.3389/fgene.2021.625234, CC BY 4.0

Network examples (see also L Plate XIV)



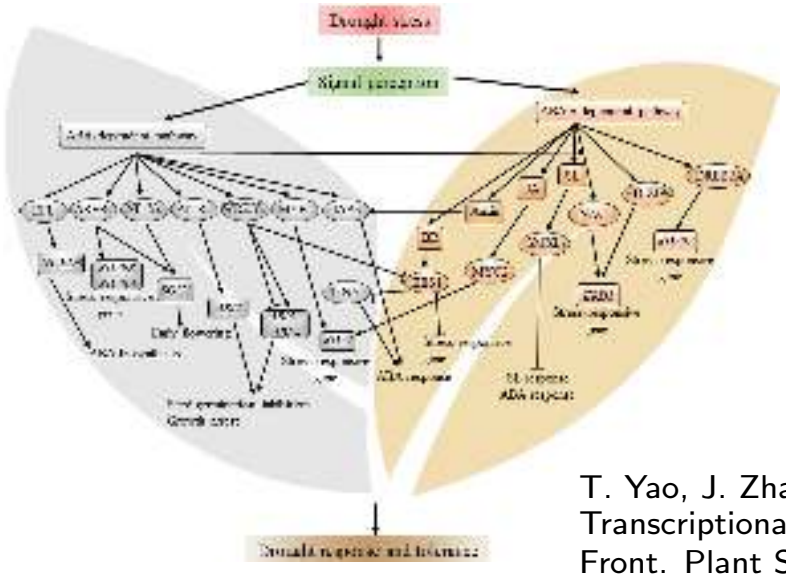
Y. Lv, Y. Que, Q. Su, Q. Li, X. Chen, H. Lu, (2016), Bioinformatics facilitating the use of microarrays to delineate potential miRNA biomarkers in aristolochic acid nephropathy. Oncotarget, 7:52270-52280, doi: 10.18632/oncotarget.10586 CC BY 3.0

P.D. Koch, M. J. Pittet, R. Weissleder, (2020), The chemical biology of IL-12 production via the non–canonical NFkB pathway, RSC Chem. Biol., 1:166-176, doi: 10.1039/D0CB00022A CC BY 3.0



c.

IL12 modulators	Ligand/Target/ Interactions	Functions	Cells	Drugs
CD40	CD40L (CD154); IFNg from T cells	TNF superfamily receptor	DC, Mo, B	CP-870,893, SGN-40, ChiLob 7/4, ABBV-927, SEA-CD40
cIAP	NIK; TRAF3; others	Cellular inhibitor of apoptosis protein-1	Multiple	LCL161, Birinapant, GDC-0152, AZD5582
OTUD7B	TRAF3	Multifunctional deubiquitinase	Multiple	None to date
Proteasome	Ub marked targets	Degrades ubiquitinated targets	Multiple	Bortezomib, carfilzomib, ixazomib
TBK1	NIK; others	TANK-binding kinase 1	Multiple	MRT67307, BX795
CD137 (4-1BB)	4-1BBL	TNF superfamily receptor	DC, NK, T, G	BMS663513, PF-05082566
GITR	GITRL	TNF superfamily receptor	DC, T, NK	MEDI1873, TRX518, GWN323
NIK	p100	NF-κB inducing kinase	DC, NK, T, B	NIK SMI1
RelB	DNA promoters	Transcription factor	Multiple	None: likely not directly druggable



T. Yao, J. Zhang, M. Xie, G. Yuan, T.J. Tschaplinski, W. Muchero, J.-G. Chen (2021), Transcriptional Regulation of Drought Response in Arabidopsis and Woody Plants. Front. Plant Sci. 11:572137. doi: 10.3389/fpls.2020.572137, CC BY 4.0

Annotation

Assigning function to particular regions of genome

Need to know *expressed* sequence of gene to design probe

RefSeq: `https://www.ncbi.nlm.nih.gov/refseq/`
(Reference Sequences)
dbEST: `https://www.ncbi.nlm.nih.gov/dbEST/`
(Expressed Sequence Tags database, short usually $<$ 1000bp)

Probe: short, subsequence specific to gene
- uniqueness (?)
- alternative splicing

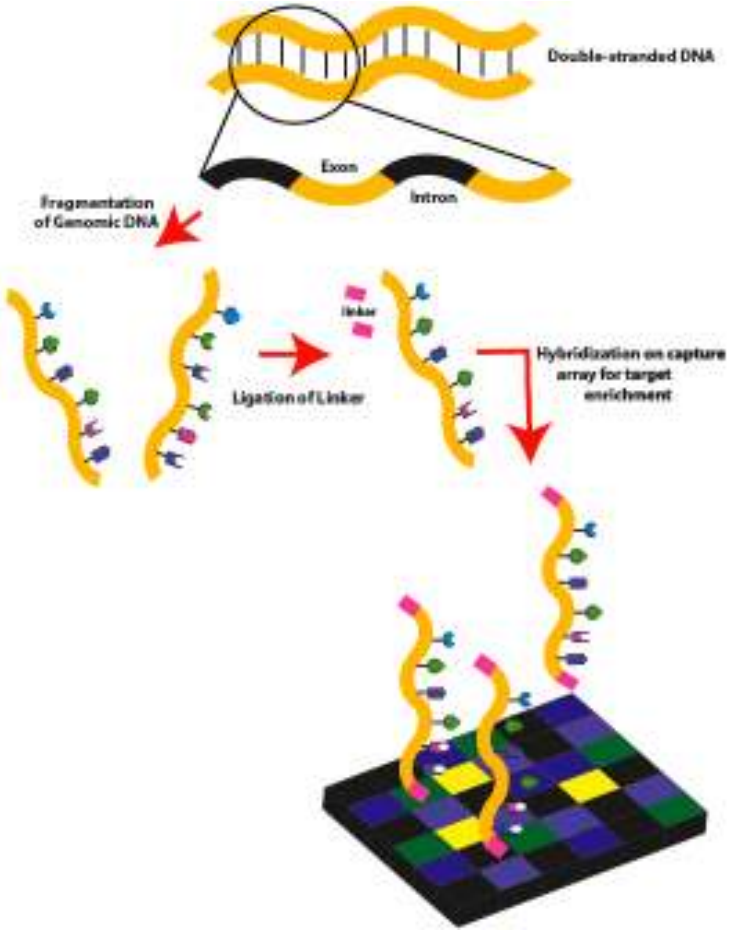
Software

Bioconductor (R methods for microarray analysis)
`www.bioconductor.org`
limma: Linear Models for Microarray and RNA-Seq Data

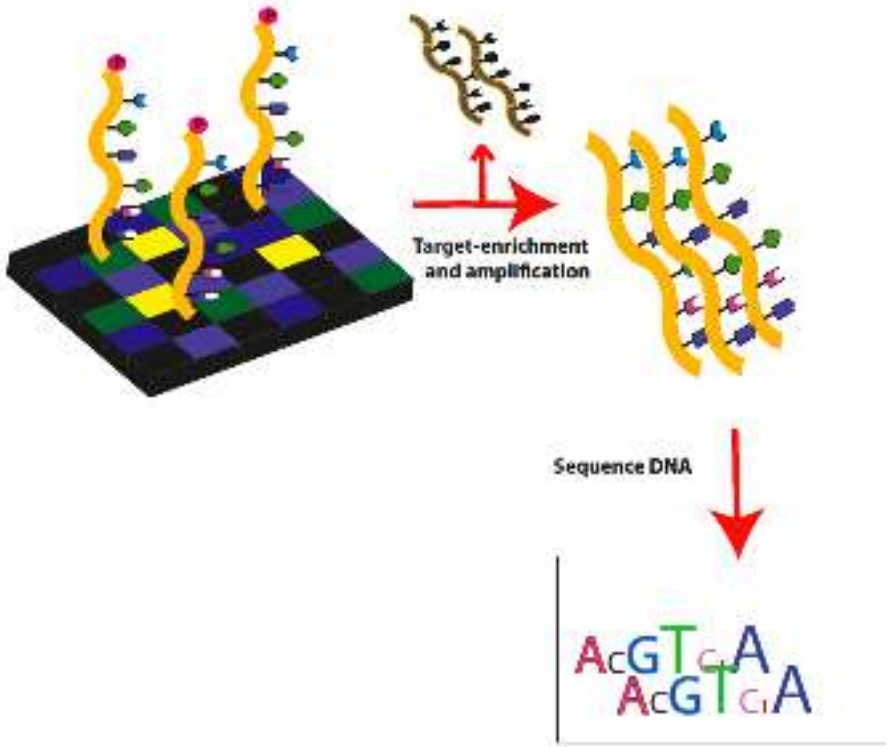
Microarray producers provide software

New technology (see also MM Fig. 10–7)

https://commons.wikimedia.org/w/index.php?curid=9642877,
by SarahKusala, CC BY 3.0



https://commons.wikimedia.org/w/index.php?curid=9642932,
by SarahKusala, CC BY 3.0



No reference genome needed, useful for new unsequenced organisms

Questions?