Λειτουργική γονιδιωματική

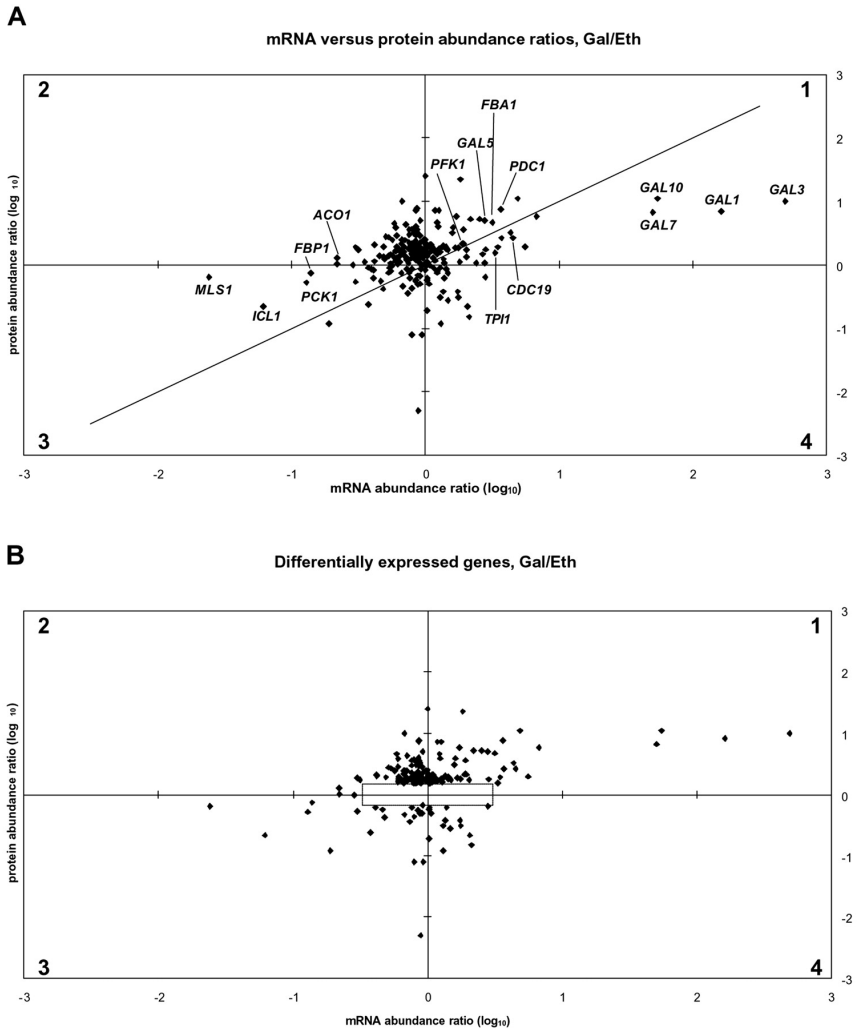
Λειτουργική γονιδιωματική: Τι είναι

- Προσπαθεί να κατανοήσει τις λειτουργίες των βιολογικών μορίων, σε επίπεδο ολόκληρου του γονιδιώματος.
- Γίνονται μετρήσεις για το σύνολο των γονιδίων, σε μια συγκεκριμένη στιγμή ή κατάσταση.
- Αρχικά, οι μετρήσεις γίνονταν για ένα βιομόριο. Σήμερα μελετάμε την συμπεριφορά ολόκληρου του συστήματος.
- Η μελέτη της μεταγραφής του συνόλου των γονιδίων ονομάζεται μεταγραφωματική ή transcriptomics.

Transcriptomics

- Expressed sequence tags (ESTs)
- Serial analysis of gene expression (SAGE)
- Μικροσυστοιχίες (microarrays)
- RNA-seq (whole transcriptome shotgun sequencing)

mRNA abundance ratios versus protein-abundance ratios.

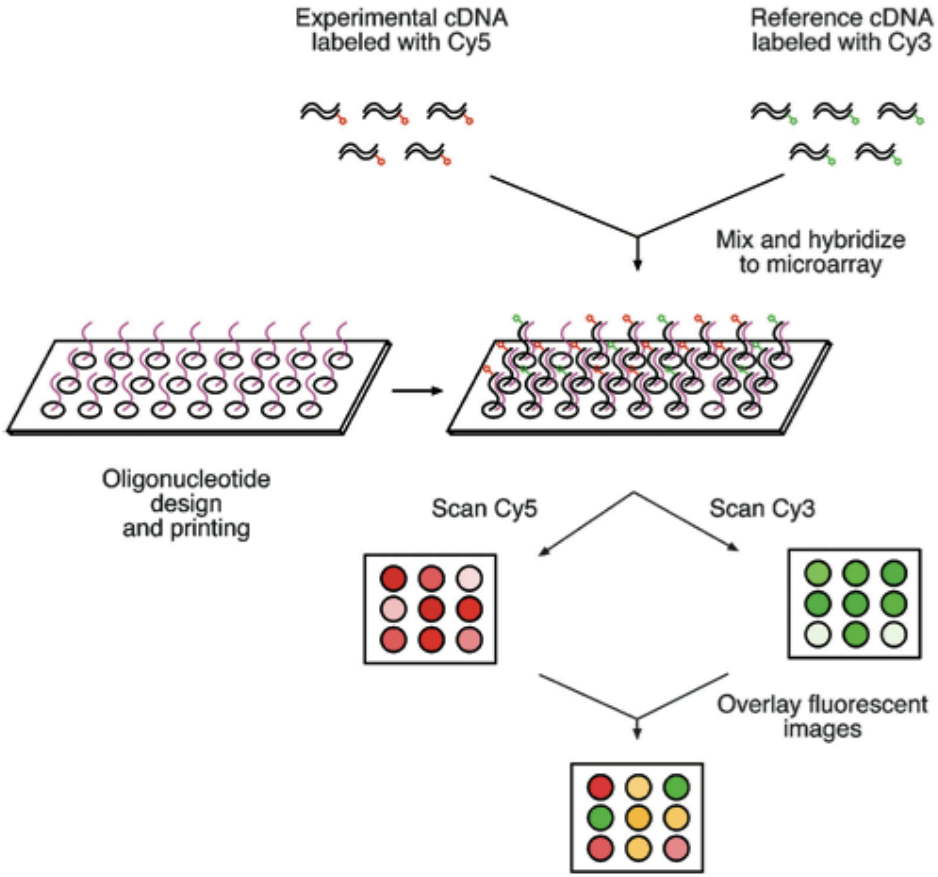


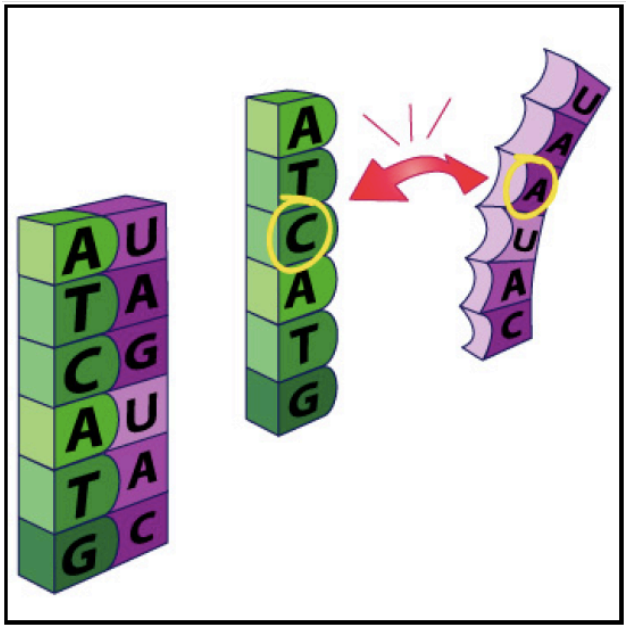
Griffin T J et al. Mol Cell Proteomics 2002;1:323-333

Διαφορική έκφραση γονιδίων

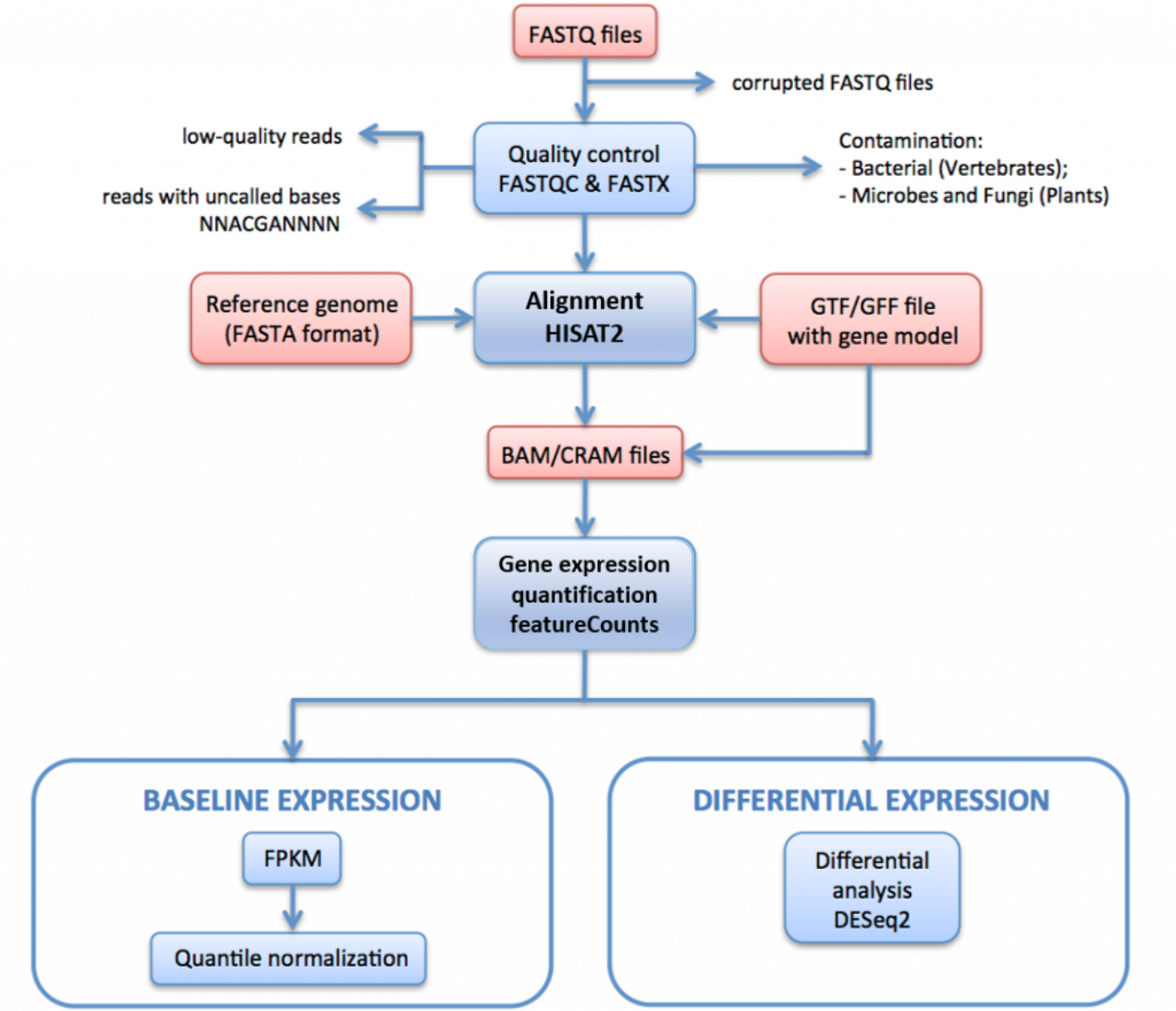
Microarrays & RNA-Sequencing

Μικροσυστοιχίες

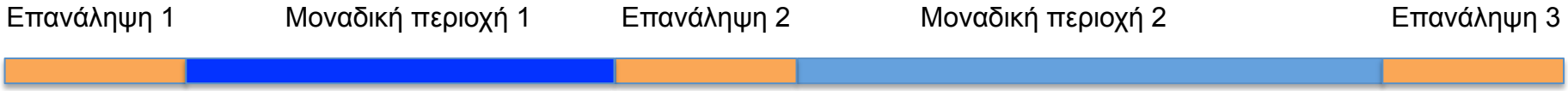




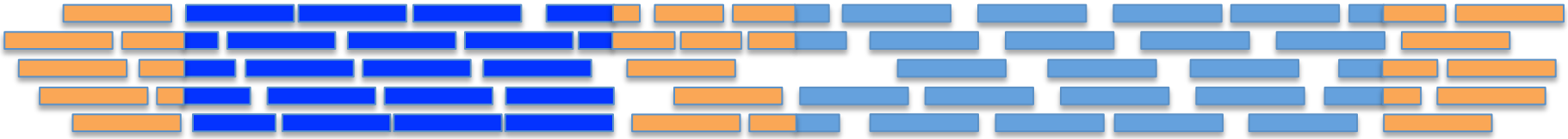
RNA-SEQ


FASTQ files
corrupted FASTQ files
low-quality reads
reads with uncalled bases
NNACGANNNN
Quality control
FASTQC & FASTX
Contamination:
- Bacterial (Vertebrates);
- Microbes and Fungi (Plants)
Reference genome
(FASTA format)
Alignment
HISAT2
GTF/GFF file
with gene model
BAM/CRAM files
Gene expression
quantification
featureCounts
BASELINE EXPRESSION
FPKM
Quantile normalization
DIFFERENTIAL EXPRESSION
Differential
analysis
DESeq2

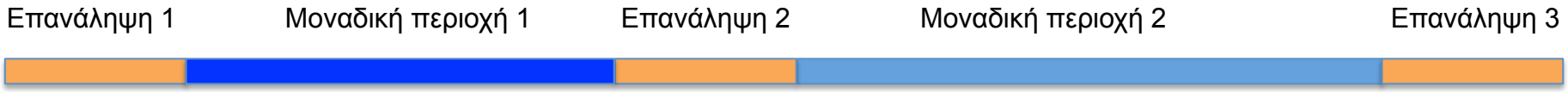
Reference assembly/alignment



Αλληλούχιση με Sequence Reads

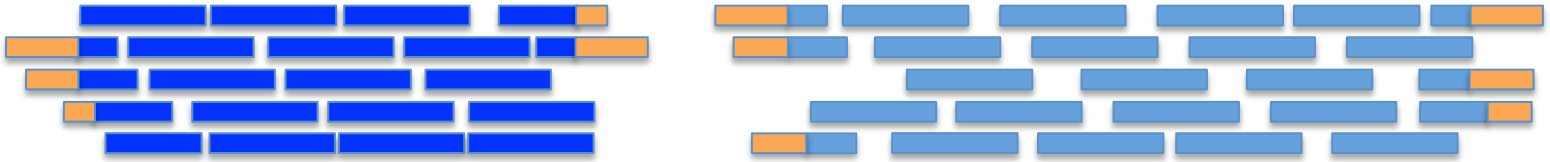


Συναρμολόγιση με βάση γονιδίωμα αναφοράς



Sequence Reads που μπορούν να στοιχιθούν
σε περισσότερες από μια θέσεις δεν στοιχίζονται

Μόνο στοίχιση των Sequence Reads που έχουν μια μοναδική θέση

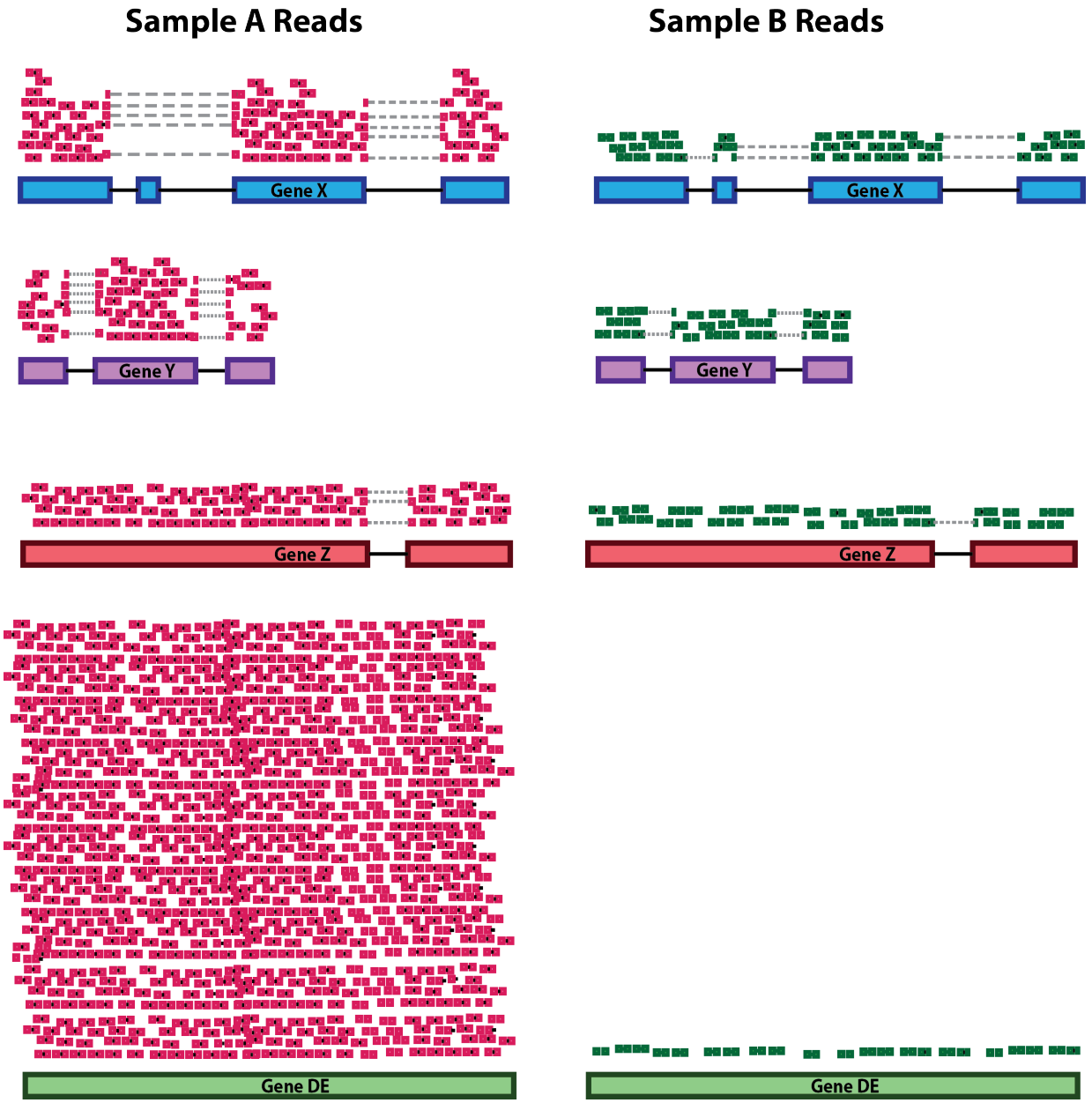


Reference assembly

Short read aligners

- Bowtie
- BWA
- STAR

- RPKM – Reads per kilobase million
- FPKM – fragments per kilobase million
- TPM - Transcripts per million (TPM)

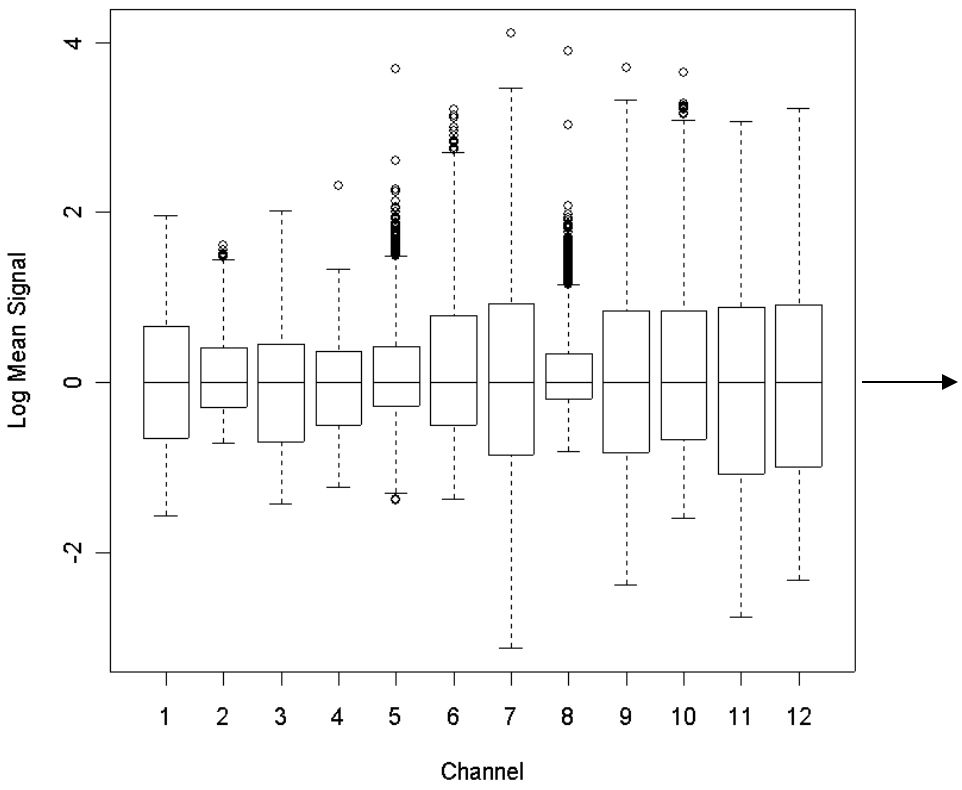


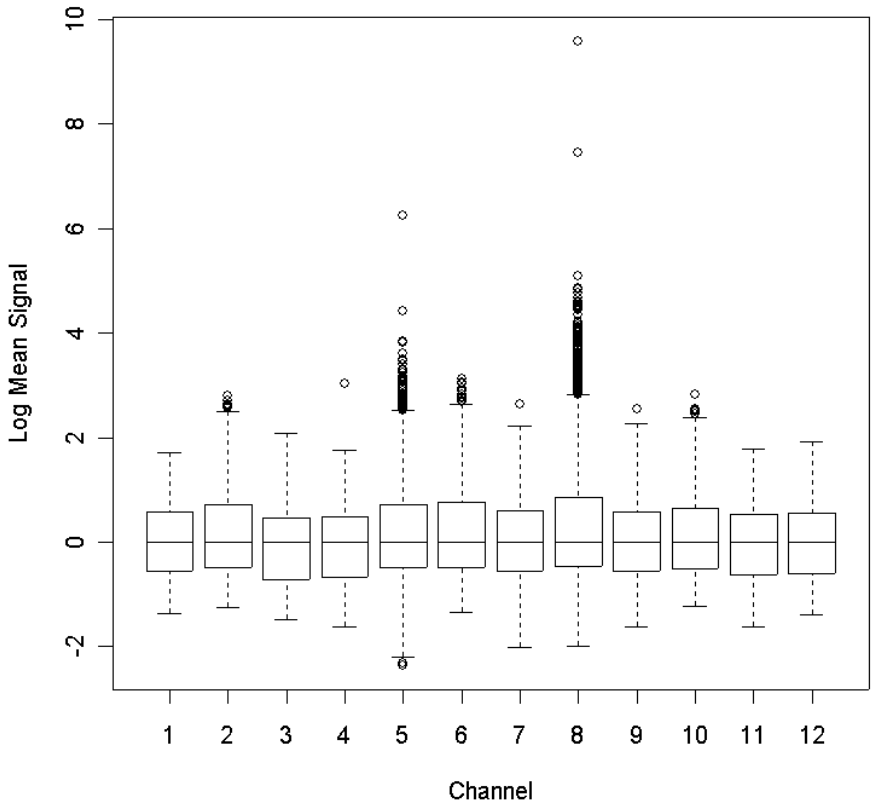
Log2

- Αν το γονίδιο εκφράζεται περισσότερο στην Α συνθήκη (κόκκινη χρωστική) από ότι στην control (πράσινη χρωστική), τότε ο λόγος συνθήκη_Α/control (κόκκινη/πράσινη) θα είναι $\lambda>1$, αλλιώς σε αντίθετη περίπτωση $0<\lambda<1$.
- Αν το γονίδιο εκφράζεται με διπλάσια ένταση στην συνθήκη Α, σε σχέση με την συνθήκη control, τότε ο λόγος θα είναι $\lambda=2$.
- Αν το γονίδιο εκφράζεται με τη μισή ένταση στην συνθήκη Α, σε σχέση με την συνθήκη control, τότε ο λόγος θα είναι $\lambda=0.5$.
- Μετατρέποντας τους λόγους σε $\log_2$, έχουμε:
  - $\lambda=2$ -> $\log_2\lambda=1$
  - $\lambda=0.5$ -> $\log_2\lambda=-1$
  - Με την κανονικοποίηση σε $\log_2$ τα δεδομένα γίνονται συμμετρικά.

Κανονικοποίηση κλίμακας
Scale normalization

Data after Median Centering and Scale Normalizing

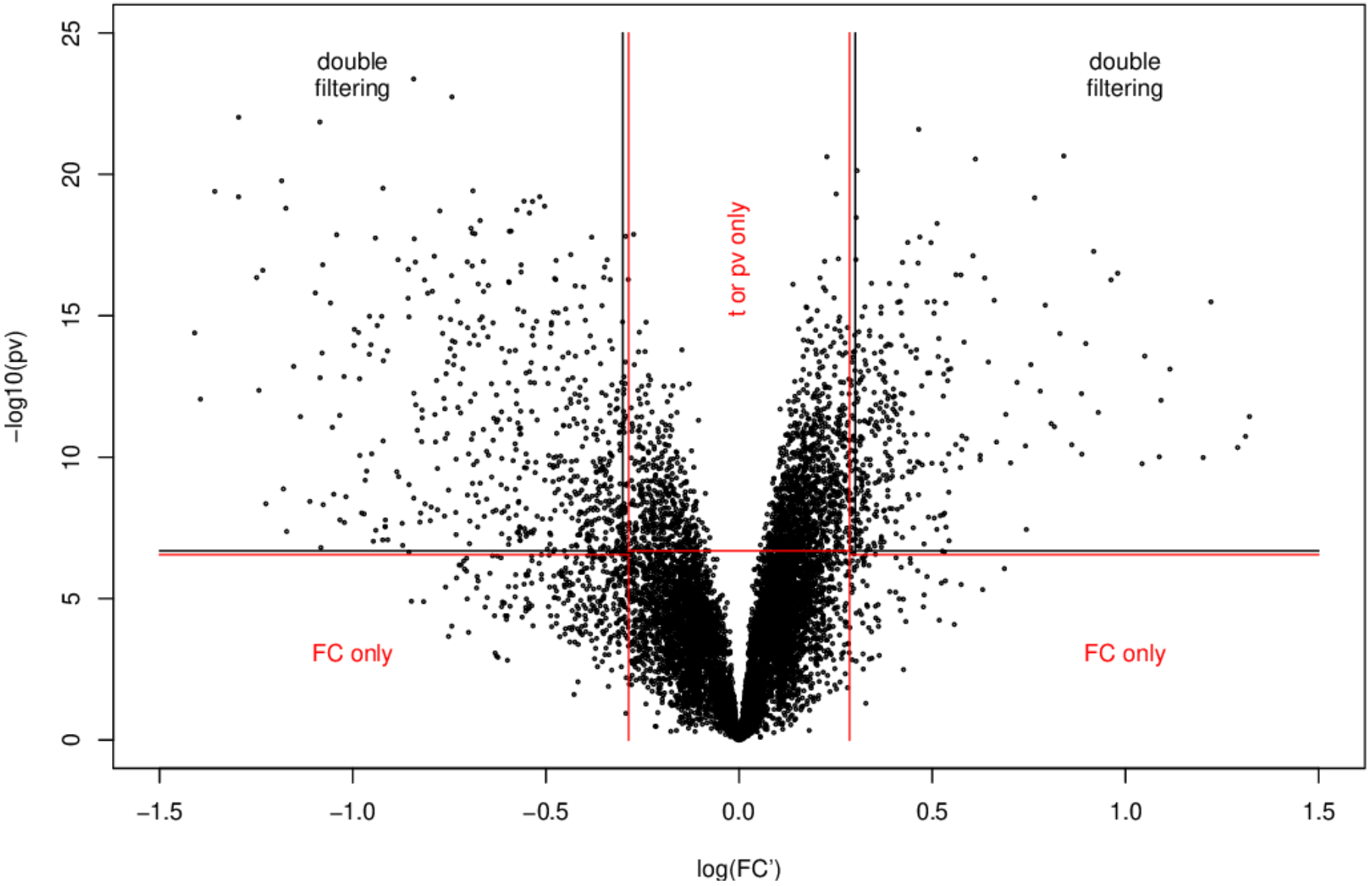




Υπερ/υπο-έκφραση

- Πότε θεωρούμε ότι ένα γονίδιο υπερ/υπό-εκφράζεται σε μια συγκεκριμμένη συνθήκη.
  - $Log_2\lambda > 1$ ή $Log_2\lambda < -1$ (διπλάσια/υποδιπλάσια έκφραση σε σχέση με τη συνθήκη control).
  - Με στατιστικές μεθόδους (t-test, ANOVA).

Volcano Plot



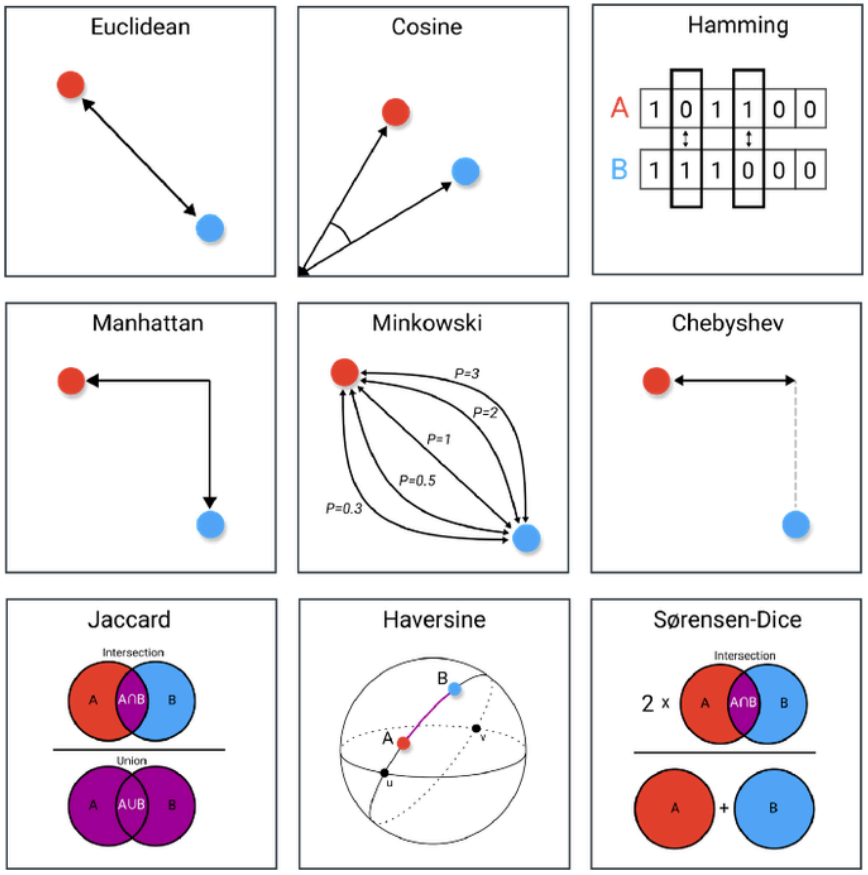
Ομαδοποίηση γονιδίων/συνθηκών με την ίδια συμπεριφορά.

- Χρειαζόμαστε αρκετά σημεία (διαφορετικές συνθήκες ή χρονικές στιγμές)
- Με μεθόδους αποστάσεων, όπου οι μετρήσεις ενός γονιδίου για διαφορετικές συνθήκες αποτελούν ένα διάνυσμα.
- Υπολογίζουμε αποστάσεις μεταξύ διαφορετικών διανυσμάτων (γονιδίων).
  - Ευκλείδια απόσταση
  - Συντελεστής συσχέτισης Pearson (Pearson correlation coefficient).
  - Δημιουργείται πίνακας αποστάσεων μεταξύ των γονιδίων.
  - Το αντίστοιχο μπορεί να γίνει και για να ομαδοποιήσουμε κοινές συνθήκες.

9 Distance Measures in Data Science

The advantages and pitfalls of common distance measures

Maarten Grootendorst Feb 1 · 10 min read ★



Distance Measures. Image by the author.

	Condition1	Condition2	Condition3	Condition4	Condition5
Gene1	1	−3	10	0	0
Gene2	−7	−2	−1	10	−8
Gene3	2	1	9	−9	5
Gene4	10	10	−4	0	−9
Gene5	−2	9	−7	0	−7
Gene6	−6	6	−5	−3	9
Gene7	2	1	8	−1	−2
Gene8	−3	−8	−1	−6	2
Gene9	−10	0	9	6	0
Gene10	−2	4	5	−7	−6
Gene11	−2	−2	0	−9	10
Gene12	−6	−10	−5	8	5
Gene13	2	−8	1	−1	2
Gene14	−7	−9	−7	1	1
Gene15	−6	4	−8	−1	−6
Gene16	−5	2	−5	8	−8
Gene17	8	−2	−7	0	2
Gene18	2	9	−9	9	3
Gene19	−3	−1	7	−1	6
Gene20	10	−4	3	−3	−1

	Condition1	Condition2
Gene1	1	−3
Gene2	−7	−2
Gene3	2	1
Gene4	10	10
Gene5	−2	9
Gene6	−6	6
Gene7	2	1
Gene8	−3	−8
Gene9	−10	0
Gene10	−2	4
Gene11	−2	−2
Gene12	−6	−10
Gene13	2	−8
Gene14	−7	−9
Gene15	−6	4
Gene16	−5	2
Gene17	8	−2
Gene18	2	9
Gene19	−3	−1
Gene20	10	−4

	Condition1	Condition2	Condition3	Condition4	Condition5
Gene1	1	−3	10	0	0
Gene2	−7	−2	−1	10	−8

	Condition1	Condition2	Condition3	Condition4	Condition5
Gene1	1	−3	10	0	0
Gene2	−7	−2	−1	10	−8

	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10	Gene11	Gene12	Gene13	Gene14	Gene15	Gene16	Gene17	Gene18	Gene19	Gene20
Gene1																				
Gene2																				
Gene3																				
Gene4																				
Gene5																				
Gene6																				
Gene7																				
Gene8																				
Gene9																				
Gene10																				
Gene11																				
Gene12																				
Gene13																				
Gene14																				
Gene15																				
Gene16																				
Gene17																				
Gene18																				
Gene19																				
Gene20																				

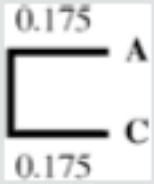
	Condition1	Condition2
Gene1	1	−3
Gene2	−7	−2
Gene3	2	1
Gene4	10	10
Gene5	−2	9
Gene6	−6	6
Gene7	2	1
Gene8	−3	−8
Gene9	−10	0
Gene10	−2	4
Gene11	−2	−2
Gene12	−6	−10
Gene13	2	−8
Gene14	−7	−9
Gene15	−6	4
Gene16	−5	2
Gene17	8	−2
Gene18	2	9
Gene19	−3	−1
Gene20	10	−4

	Condition1	Condition2	Condition3	Condition4	Condition5
Condition1					
Condition2					
Condition3					
Condition4					
Condition5					

UPGMA

	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

1. Using a distance matrix involving four taxa, A, B, C, and D, the UPGMA method first joins two closest taxa together which are A and C (0.35 in grey). Because all taxa are equidistant from the node, the branch length for A to the node is AC/2 = 0.35/2 = 0.175.



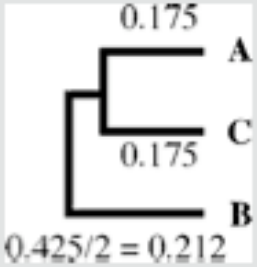
2. Because A and C are joined into a cluster, they are treated as one new composite taxon, which is used to create a reduced matrix. The distance of A-C cluster to every other taxa is one half of a taxon to A and C, respectively. That means that the distance of B to A-C is (AB + BC)/2; and that of D to A-C is (AD + CD)/2.

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

UPGMA

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

3. In the newly reduced-distance matrix, the smallest distance is between B and A-C (in grey), which allows the grouping of B and A-C to create a three-taxon cluster. The branch length for the B is one half of B to the A-C cluster.

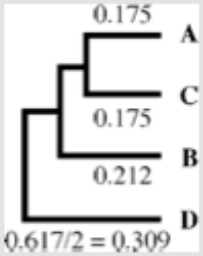


UPGMA

4. When B and A-C are grouped and treated as a single taxon, this allows the matrix to reduce further into only two taxa, D and B-A-C. The distance of D to the composite taxon is the average of D to every single component which is (BD + AD + CD)/3.

	B-A-C
D	$\frac{0.7 + 0.6 + 0.55}{3} = 0.617$

5. D is the last branch to add to the tree, whose branch length is one half of D to B-A-C.



6. Because distance trees allow branches to be additive, the resulting distances between taxa from the tree path can be used to create a distance matrix. Obviously, the estimated distances do not match the actual evolutionary distances shown, which illustrates the failure of UPGMA to precisely reflect the experimental observation.

	A	B	C
B	0.42		
C	0.35	0.42	
D	0.62	0.62	0.62

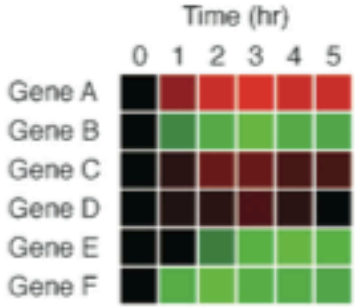
	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

Ομαδοποίηση

	0 hr	1 hr	2 hr	3 hr	4 hr	5 hr
Gene A	1	4	6	8	6	6
Gene B	1	0.6	0.3	0.1	0.3	0.4
Gene C	1	2	4	4	3	3
Gene D	1	1.5	2	3	2	1
Gene E	1	1	0.5	0.2	0.1	0.2
Gene F	1	0.3	0.1	0.2	0.3	0.4

convert to false colors

$\log_2$ conversion



	Gene B	Gene C	Gene D	Gene E	Gene F
Gene A	-0.82	0.96	0.65	-0.68	-0.79
Gene B		-0.85	-0.86	0.66	0.67
Gene C			0.70	-0.65	-0.87
Gene D				-0.41	-0.72
Gene E					0.26

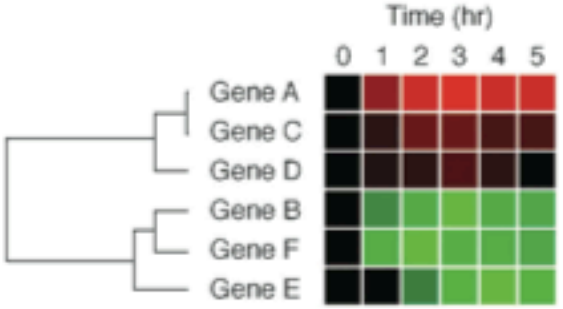
calculating Pearson correlation coefficients between genes

	0 hr	1 hr	2 hr	3 hr	4 hr	5 hr
Gene A	0	2	2.6	3	2.6	2.6
Gene B	0	-0.7	-1.7	-3.3	-1.7	-1.3
Gene C	0	1	2	2	1.6	1.6
Gene D	0	0.6	1	1.6	1	0
Gene E	0	0	-1	-2.3	-3.3	-2.3
Gene F	0	-1.7	-3.3	-2.3	-1.7	-1.3

conversion of coefficients to positive distance values

	Gene B	Gene C	Gene D	Gene E	Gene F
Gene A	1.82	0.04	0.35	1.68	1.79
Gene B		1.85	1.86	0.34	0.33
Gene C			0.30	1.65	1.87
Gene D				1.41	1.72
Gene E					0.74

hierarchical clustering



Οντολογίες

- www.geneontology.org

- Ελεγχόμενο λεξιλόγιο για την περιγραφή των ιδιοτήτων των γονιδίων και των πρωτεϊνών.

- Περιγράφουν:
  - Μοριακές λειτουργίες του βιομορίου (1 ή περισσότερες).
  - Βιολογικές διαδικασίες στις οποίες εμπλέκεται το βιομόριο (1 ή περισσότερες).
  - Κυτταρικό διαμέρισμα στο οποίο συναντάται το βιομόριο (1 ή περισσότερα).

Gene ontology

REVIEWS

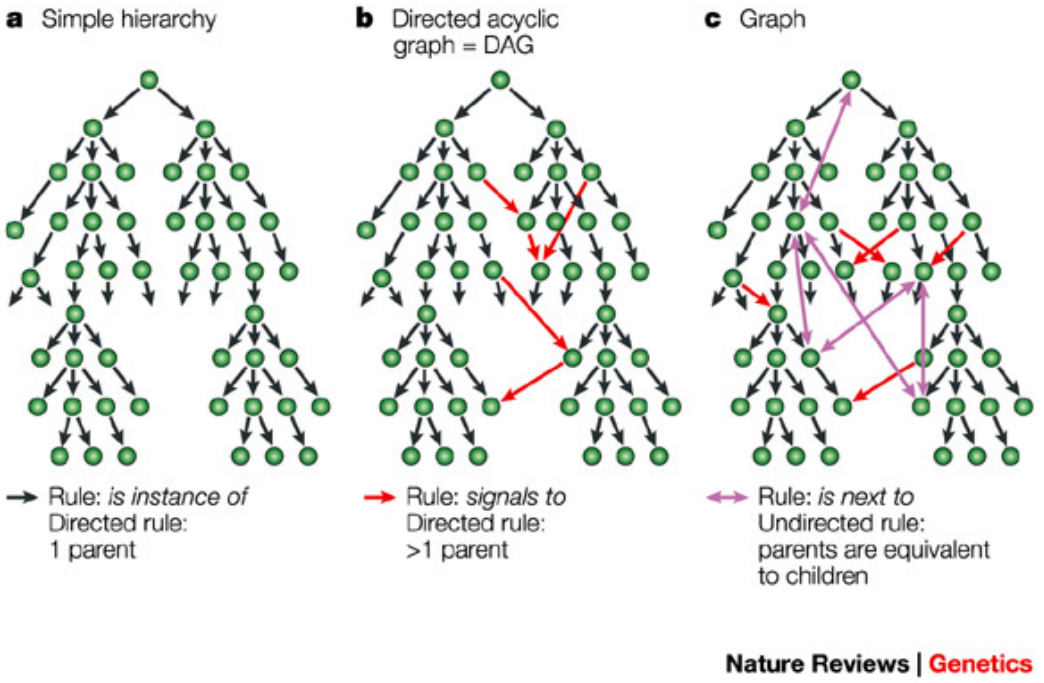
Use and misuse of the gene ontology annotations

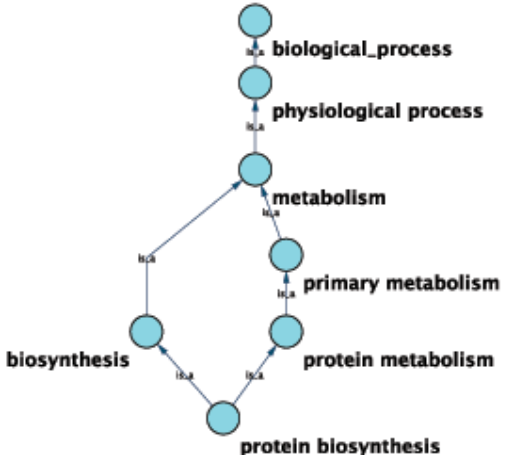
Seung Yon Rhee, Valerie Wood‡, Kara Dolinski§ and Sorin Draghici||*

Abstract | The Gene Ontology (GO) project is a collaboration among model organism databases to describe gene products from all organisms using a consistent and computable language. GO produces sets of explicitly defined, structured vocabularies that describe biological processes, molecular functions and cellular components of gene products in both a computer- and human-readable manner. Here we describe key aspects of GO, which, when overlooked, can cause erroneous results, and address how these pitfalls can be avoided.

Οντολογίες: Η δομή τους

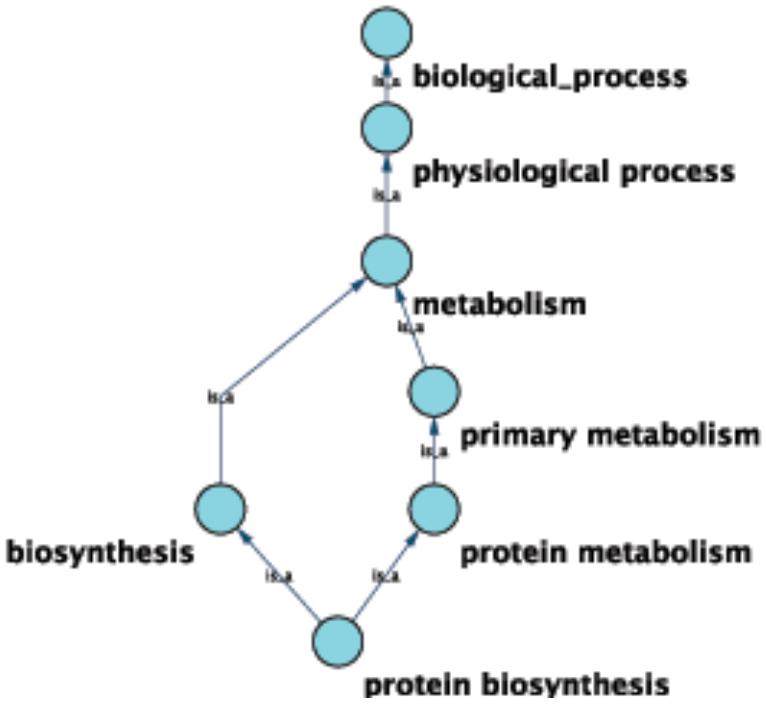
- Δείχνει τις σχέσεις μεταξύ των διαφορετικών όρων.
- Ένας όρος μπορεί να αποτελεί πιο εξειδικευμένη περιγραφή ενός άλλου όρου.
- Είναι κατευθυνόμενα ακυκλικά γραφήματα (DAG).
- Παρόμοια με ιεραρχίες.
- Η διαφορά είναι ότι ένας κόμβος-απόγονος μπορεί να έχει περισσότερους από έναν προγόνους.





Οντολογίες: Η δομή τους

- Θεωρούμε ότι αν σε ένα βιομόριο αντιστοιχεί ένα όρος-οντολογία, τότε σε αυτό το βιομόριο ανήκουν και όλοι οι πρόγονοι του όρου-οντολογίας.



Gene ontology

NATURE REVIEWS | GENETICS VOLUME 9 | JULY 2008 | 509

REVIEWS

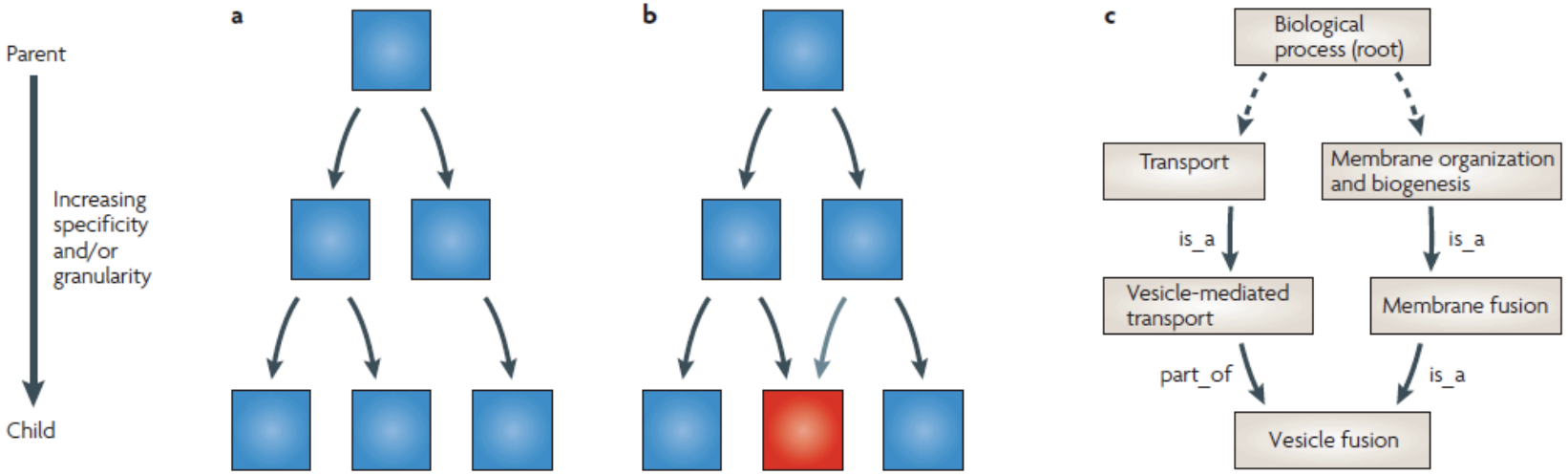


Figure 1 | **Simple trees versus directed acyclic graphs.** Boxes represent nodes and arrows represent edges. **a** | An example of a simple tree, in which each child has only one parent and the edges are directed, that is, there is a source (parent) and a destination (child) for each edge. **b** | A directed acyclic graph (DAG), in which each child can have one or more parents. The node with multiple parents is coloured red and the additional edge is coloured grey. **c** | An example of a node, vesicle fusion, in the biological process ontology with multiple parentage. The dashed edges indicate that there are other nodes not shown between the nodes and the root node (biological process). A root is a node with no incoming edges, and at least one leaf (also called a sink). A leaf node is a node with no outgoing edges, that is, a terminal node with no children (vesicle fusion). Similar to a simple tree, A DAG has directed edges and does not have cycles, that is, no path starts and ends at the same node, and will always have at least one root node. The depth of a node is the length of the longest path from the root to that node, whereas the height is the length of the longest path from that node to a leaf[41]. is_a and part_of are types of relationships that link the terms in the GO ontology. More information about the relationships between GO terms are found online (An Introduction to the Gene Ontology).

Gene ontology

Table 1 | **Evidence codes used by GO**

Evidence code	Evidence code description	Source of evidence	Manually checked	Current number of annotations*
IDA	Inferred from direct assay	Experimental	Yes	71,050
IEP	Inferred from expression pattern	Experimental	Yes	4,598
IGI	Inferred from genetic interaction	Experimental	Yes	8,311
IMP	Inferred from mutant phenotype	Experimental	Yes	61,549
IPI	Inferred from physical interaction	Experimental	Yes	17,043
ISS	Inferred from sequence or structural similarity	Computational	Yes	196,643
RCA	Inferred from reviewed computational analysis	Computational	Yes	103,792
IGC	Inferred from genomic context	Computational	Yes	4
IEA	Inferred from electronic annotation	Computational	No	15,687,382
IC	Inferred by curator	Indirectly derived from experimental or computational evidence made by a curator	Yes	5,167
TAS	Traceable author statement	Indirectly derived from experimental or computational evidence made by the author of the published article	Yes	44,564
NAS	Non-traceable author statement	No 'source of evidence' statement given	Yes	25,656
ND	No biological data available	No information available	Yes	132,192
NR	Not recorded	Unknown	Yes	1,185

*October 2007 release

Gene ontology

Table 2 | **Distribution of gene ontology (GO) annotations for species with more than 5,000 annotations**

Species (NCBI taxon ID)	Genes* with experimental annotations‡	Total annotated genes*	Percentage of genes* with at least one experimental annotation	Total genes*	Percentage annotated§	Percentage known in genome‖
Schizosaccharomyces pombe (4896)	4,482	4,930	90.9%	4,930	100%	90.9%
Saccharomyces cerevisiae (4932)	4,947	5,794	85.4%	5,794	100%	85.4%
Mouse (10090)	10,621	18,386	57.8%	27,289	67.4%	38.9%
Caenorhabditis elegans (6239)	4,614	14,154	32.6%	20,163	70.2%	22.9%
Human¶ (9606)	4,780	17,021	28.1%	20,887	81.5%	22.9%
Arabidopsis thaliana# (3702)	5,530	26,637	20.8%	27,029	98.5%	20.5%
Rat (10116)	3,566	17,243	20.7%	17,993	95.8%	19.8%
Fruitfly (7227)**	2,790	9,563	29.2%	14,141	67.6%	19.7%
Candida albicans (5476)	806	3,756	21.4%	6,166	60.9%	13.0%
Pseudomonas aeruginosa PAO1 (208964)	491	2,506	19.6%	5,568	45.0%	8.82%
Slime mold (44689)	797	6,892	11.6%	13,625	50.6%	5.9%
Trypanosoma brucei (5691)	449	3,914	11.5%	9,154	42.8%	4.92%
Zebrafish (7955)	1,235	13,574	5.8%	21,322	63.7%	3.7%
Plasmodium falciparum (5833)	188	3,243	5.8%	5,420	59.8%	3.47%
Rice (39947)	654	29,877	2.2%	41,908	71.3%	1.57%
Chicken¶ (9031)	75	6,063	1.2%	16,737	36.2%	0.4%
Cow¶ (9913)	96	8,536	1.1%	21,756	39.2%	0.4%

*Total genes in genomes include only those that encode proteins. These numbers were obtained from the databases that contribute annotations to GO and are listed on the GO annotations download page (http://www.geneontology.org/GO.current.annotations.shtml). ‡Experimental annotations include those only with the following evidence codes: IDA (inferred from direct assay), IEP (inferred from expression pattern), IGI (inferred from genetic interaction), IMP (inferred from mutant phenotype) and IPI (inferred from physical interaction). §Percentage annotated is determined by dividing the number of genes annotated by total genes. ‖Percentage known in genome is determined by multiplying the percentage of experimentally derived annotations by the percentage of the genome annotated. This is an approximation of the extent of knowledge about the portion of the genome that encodes proteins in an organism with a complete genome sequence that is captured by annotation. ¶Numbers are from the GO annotation project at the European Bioinformatics Institute, human data last updated 14 September 2007, cow data last updated 17 January 2007, chicken data last updated 10 July 2007. #Numbers are from The *Arabidopsis* Information Resource (TAIR), last updated 14 December 2007. **Numbers are based on release 5.4 of the *Drosophila melanogaster* genome and GO annotations from FlyBase release FB2007_03 (dated 11 January 2007). NCBI, National Center for Biotechnology Information.

Gene ontology

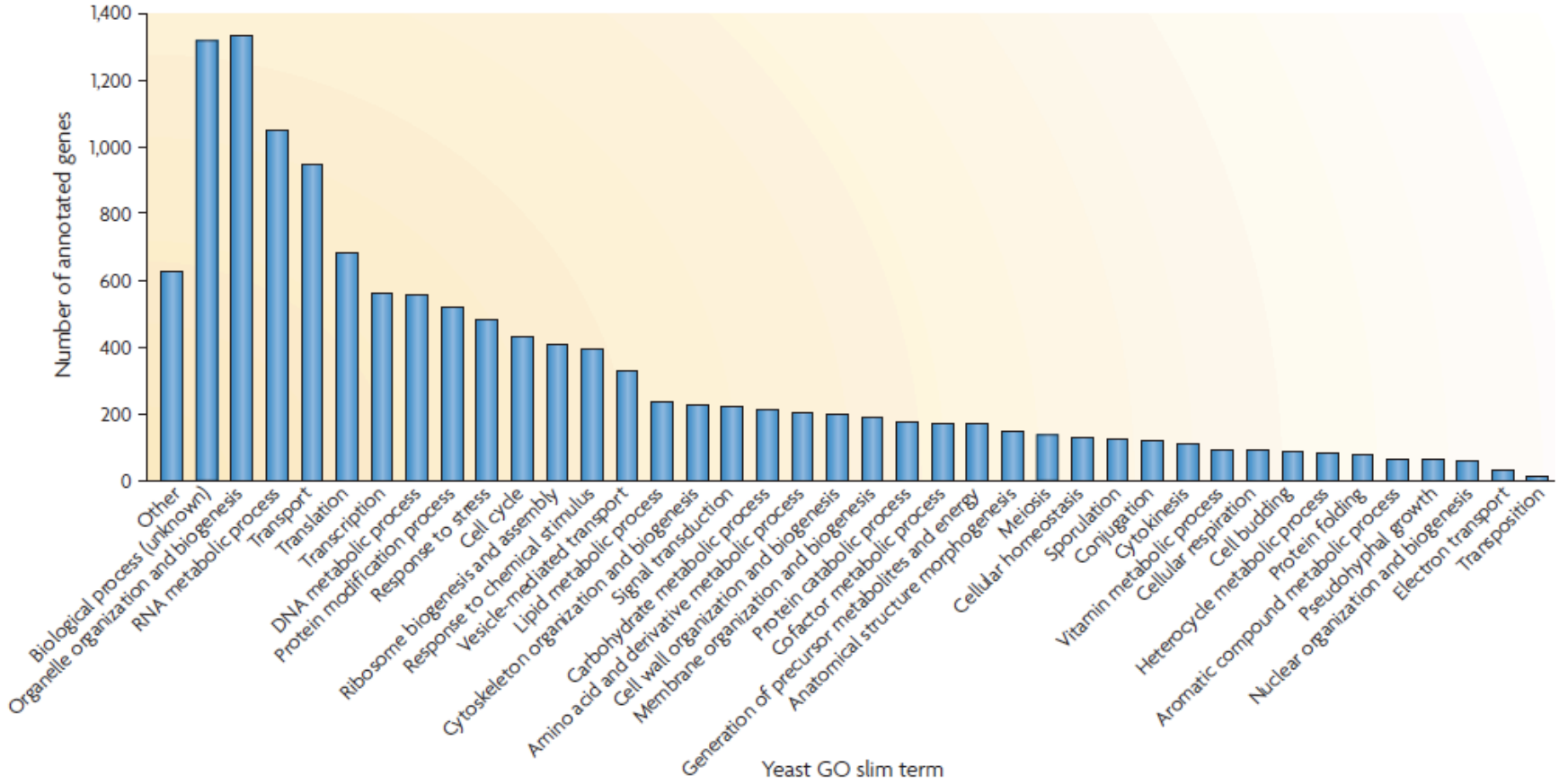


Figure 2 | **Using gene ontology (GO) to bin the yeast genome into broad biological process categories.** This example was generated by downloading the go_slim_mapping.tab file from the *Saccharomyces* genome database ftp site (dated 19 January 2008). This file maps every gene in the yeast genome to the yeast GO slim ontology available from the GO website. The number of genes (6,200 in total, including RNAs but excluding 'dubious' genes) annotated to a particular term in the yeast GO slim ontology is indicated on the graph. Dubious genes are those that were originally predicted to exist, on the basis of ORF length, but that are now thought to be unlikely to encode an expressed protein, on the basis of functional and comparative genomics data. The 'other' term is used when genes are annotated to terms other than those included in the GO slim ontology, and the 'biological process' term, the root node in the biological process ontology, indicates that genes annotated to it are not yet characterized. Note that because genes can be binned to more than one category, there are more annotations (13,074) than total genes (6,200) with annotations.

Οντολογίες: στατιστική ανάλυση

- Παράδειγμα:
  - 1 γονιδίωμα με 10.000 γονίδια.
  - 1.000 γονίδια εμπλέκονται στον κυτταρικό κύκλο (GO_term: cell-cycle). (10% του γονιδιώματος).
  - Αν επιλέξουμε τυχαία έναν αριθμό Χ γονιδίων, θα περιμέναμε (από τύχη) περίπου το 10% (με κάποιες διακυμάνσεις) να έχουν τον όρο “κυτταρικός κύκλος”.
  - Η τυχαία διακύμανση εξαρτάται από τον αριθμό των γονιδίων.
  - Έστω ότι με τα microarrays σε ένα πείραμα βρήκαμε ότι Χ αριθμός γονιδίων υπερεκφράζονται.
  - Σε αυτό τον Χ αριθμό, βρήκαμε ότι 20% των γονιδίων ανήκουν στον κυτταρικό κύκλο.
  - Αυτή η απόκλιση (20% παρατηρούμενο - 10% αναμενόμενο) είναι στα όρια των τυχαίων διακυμάνσεων, ή είναι στατιστικά σημαντική?
    - Στατιστικά σημαντική, σημαίνει ότι τα υπερεκφρασμένα γονίδια είναι εμπλουτισμένα για την κατηγορία “κυτταρικός κύκλος ”. Δηλαδή, ο κυτταρικός κύκλος εμπλέκεται στην διαδικασία που μελετάμε.

Οντολογίες: στατιστική ανάλυση

- Η στατιστική ανάλυση γίνεται με το υπεργεωμετρικό τεστ.
- Παίρνουμε ένα p-value.
- Αν p-value < 0.05, τότε είναι στατιστικά σημαντικό.

- Αν στις οντολογίες μας είχαμε 100 όρους, θα επαναλαμβάναμε τα παραπάνω τεστς για τον κάθε όρο.
- Όμως, όσο περισσότερα τεστ κάνουμε για το πείραμά μας, τόσο αυξάνει ή πιθανότητα να βρούμε κάτι στατιστικά σημαντικό (p-value < 0.05) καθαρά από λάθος.
- Άρα, πρέπει να λάβουμε υπόψην μας πόσα τεστ διενεργούμε και να διορθώσουμε τα p-values (multiple testing correction).
  - False discovery rate (Benjamini-Hochberger)
  - Bonferroni correction

In vitro διαγνωστικά τεστ που βασίζονται σε μικροσυστοιχίες

FDA: In Vitro Diagnostic Multivariate Index Assays (IVDMIAs)

- FDA's **In Vitro Diagnostic Product Database**
- http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfivd/index.cfm

- http://www.ivdtechnology.com/article/exploring-fda-approved-ivdmias

- Some IVDMIAs are laboratory-developed tests (LDTs). LDTs are tests that are developed by a single clinical laboratory for use only in that laboratory.

- http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079148.htm

- IVDMIAs raise significant issues of safety and effectiveness. These types of tests are developed based on observed correlations between multivariate data and clinical outcome, such that the clinical validity of the claims is not transparent to patients, laboratorians, and clinicians who order these tests. Additionally, IVDMIAs frequently have a high risk intended use. FDA is concerned that patients are relying upon IVDMIAs with high risk intended uses to make critical healthcare decisions when FDA has not ensured that the IVDMIA has been clinically validated and the healthcare practitioners are unable to clinically validate the test themselves. Therefore, there is a need for FDA to regulate these devices to ensure that the IVDMIA is safe and effective for its intended use.

Mammaprint - Tissue of origin

- http://www.ivdtechnology.com/article/exploring-fda-approved-ivdmias
- **MammaPrint.**
  The first IVDMIA, the MammaPrint system, made by Agendia Inc., is a qualitative IVD test service performed in a single lab outside the United States using a 70-gene expression profile of fresh frozen breast cancer tissue samples to assess a breast cancer patient¬s risk for distant metastasis. FDA approved MammaPrint in February 2007 under de novo classification procedures.

- **Tissue of Origin Test**
  In July 2008, the Tissue of Origin Test, made by Pathwork Diagnostics, was cleared. This microarray RNA profiling test is to be used on clinical, formalin-fixed, paraffin-embedded (FFPE) biopsy tissue to aid in the classification of the origin of the tumor tissue. In June 2010 a second clearance introduced a different specimen and specimen-preparation method, and the algorithm for analysis of the expression data to create a diagnostics report and interpretation. The test uses microarray technology by Affymetrix Inc. and advanced analytics to measure the gene-expression patterns of challenging tumors, including metastatic, poorly differentiated, and undifferentiated cancer. It is intended to measure the degree of similarity between the RNA expression patterns in a patient's tumor tissue with the RNA expression patterns in a database of fifteen known tumor types.

Mammaprint


70 significant breast cancer prognosis genes
Tumor samples
78
70
60
50
40
30
20
10
Metastases: white = +
low risk
Threshold
high risk

Καρκίνοι αγνώστου προελεύσεως

- Σε κάποιες περιπτώσεις εμφάνισης/επανεμφάνισης καρκίνου είναι άγνωστη η πρωταρχική πηγή (ιστός), ακόμα και μετά από μια σειρά διαγνωστικών τεστ/βιοψία.
- Αυτό δεν επιτρέπει να χρησιμοποιηθεί ένα κατάλληλο θεραπευτικό σχήμα.
- Οι μικροσυστοιχίες επιτρέπουν να δημιουργηθεί το προφίλ γονιδιακής έκφρασης του συγκεκριμένου καρκίνου και να συγκριθεί με το προφίλ καρκίνων γνωστής προέλευσης.

Καρκίνοι αγνώστου προελεύσεως

- Δημιουργείται μια βάση από δεδομένα μεταγραφωμικής (από άλλες βάσεις δεδομένων και βιβλιογραφία).
- Τα δεδομένα είναι από γνωστούς καρκίνους, κανονικούς ιστούς, και από άλλες ασθένειες.
- Τα δεδομένα φιλτράρονται, κανονικοποιούνται.
- Στη συνέχεια γίνεται σύγκριση.

Καρκίνοι αγνώστου προελεύσεως

- http://genomemedicine.com/content/3/9/63/abstract
- **Classification of unknown primary tumors with a data-driven method based on a large microarray reference database**
- **Kalle A Ojala**, **Sami K Kilpinen** and **Olli P Kallioniemi**

IVDMIA - FDA

- http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2007/ucm108836.htm
- The MammaPrint is the first cleared in vitro diagnostic multivariate index assay (IVDMIA) device.

- http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2008/ucm116931.htm
- **FDA Clears Test that Helps Identify Type of Cancer in Tumor Sample**
- The Pathwork Tissue of Origin test compares the genetic material of a patient's tumor with genetic information on malignant tumor types stored in a database.It uses a microarray technology to analyze thousands of pieces of genetic material at one time. The test considers 15 common malignant tumor types, including bladder, breast, and colorectal tumors.