

Υπολογιστικές προσεγγίσεις για την ανακάλυψη και παραγωγή γνώσης από βάσεις Βιολογικών και Μοριακών Δεδομένων

Λευτέρης Κουμάκης

Υποψήφιος Διδάκτωρ ΜΠΔ

Επιβλέπων Καθηγητής: Βασίλειος Μουστάκης

Μέλος Τριμελούς: Μιχαήλ Ζερβάκης (Καθηγητής)

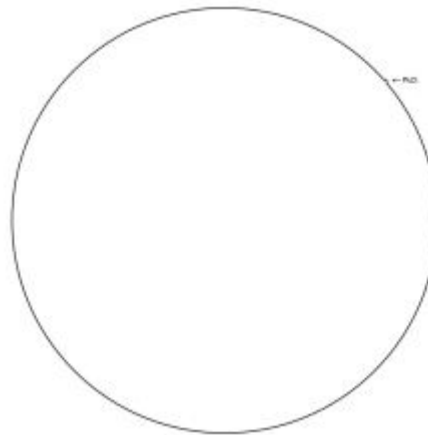
Μέλος Τριμελούς: Γεώργιος Ποταμιάς (Ερευνητής Β)



Το Διδακτορικό σε εικόνες (Matt Might)

Φαντάσου έναν κύκλο, που περιέχει όλη την ανθρώπινη γνώση.

Δημοτικό
Γυμνάσιο και το Λύκειο
Πανεπιστήμιο
Μεταπτυχιακό
Έρευνα
Στο άκρο
Πιέζεις το όριο
Βρίσκεις κάτι νέο !
Διδακτορικό
Ο κόσμος φαίνεται διαφορετικός



Αλλά μην ξεχνάς την μεγάλη εικόνα!

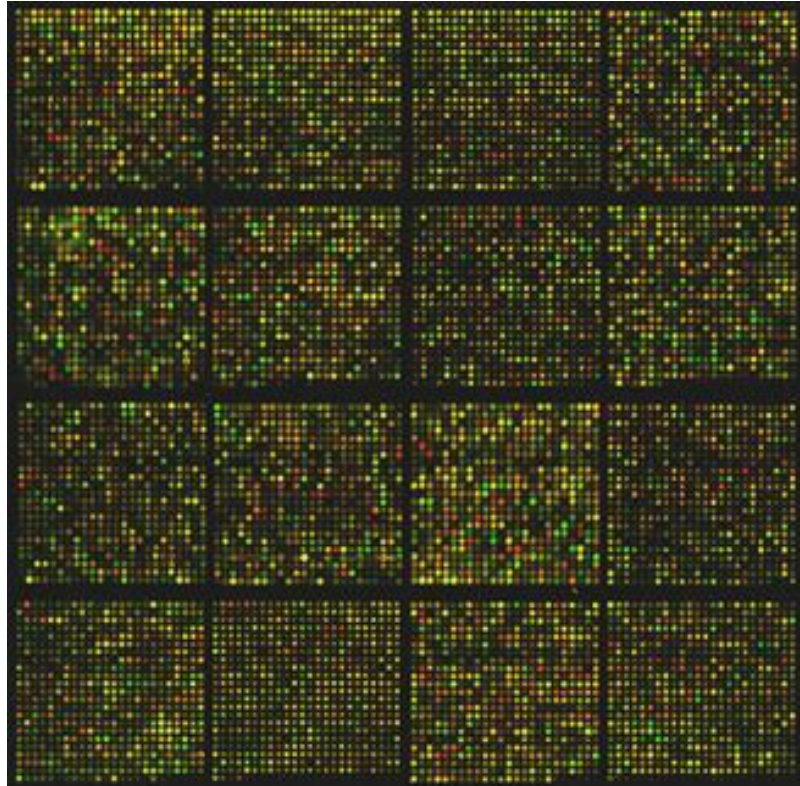


Υπολογιστικές προσεγγίσεις για την
ανακάλυψη και παραγωγή γνώσης από
βάσεις Βιολογικών και Μοριακών
Δεδομένων

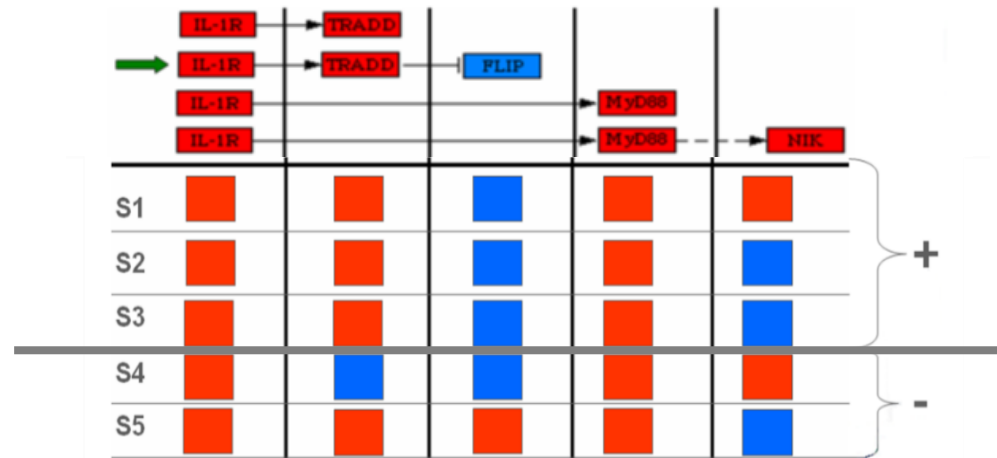
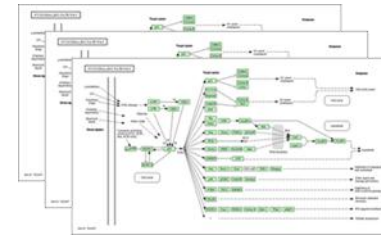
Εξατομικευμένη Ιατρική

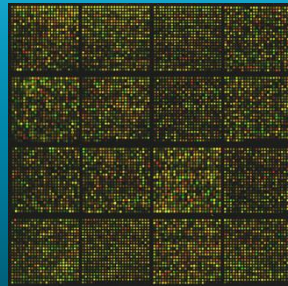


Μικροσυστοιχίες γονιδίων

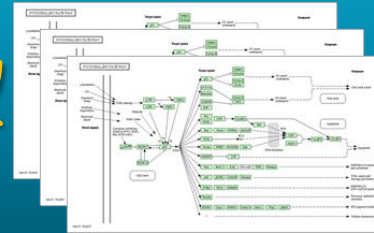


Γενετικά κανονιστικά μονοπάτια





**ΓΕΝΕΤΙΚΑ
ΚΑΝΟΝΙΣΤΙΚΑ
ΜΟΝΟΤΥΠΙΑ**



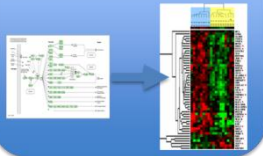
Advanced Gene Set Analysis (GSA)

Gene Regulatory Network reconstruction

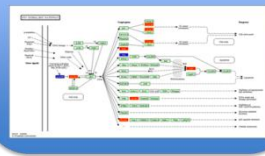
**Discriminant
pathways
subpathways**

GSA & GRNs

As **list** of genes



GSA & GRNs
With **Topology** info



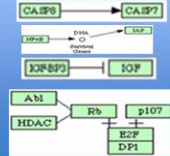
Pathway selection



Sub-pathway selection Using **Topology**



Sub-pathway selection Using Regulatory mechanisms



Σύντομη περιγραφή

Δύο από τις πιο σημαντικές μορφές μοριακών δεδομένων προέρχονται από τις μικροσυστοιχίες γονιδίων και τα γενετικά κανονιστικά μονοπάτια (GRNs).

Σε αυτή τη μελέτη παρουσιάζουμε μια καινοτόμα μεθοδολογία επεξεργασίας των πόρων αυτών.

1. εντοπίζουμε όλα τα λειτουργικά μονοπάτια που είναι κωδικοποιημένα σε GRNs.
2. Εκτιμάμε ποια από τα μονοπάτια είναι συμβατά με τις εκφάσεις των γονιδίων στις μικροσυστοιχίες των δειγμάτων σε διαφορετικές κλινικές κατηγορίες (για παράδειγμα ασθενής – υγιείς).
3. Υπολογίζουμε την διαφορική ισχύς των επιλεγμένων μονοπατιών και αξιολογούμε την βιολογική σημασία τους.



Διεργασίες

Συλλογή Δεδομένων

Προ-επεξεργασία

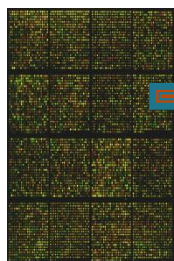
Ονοματολογία

Συνένωση

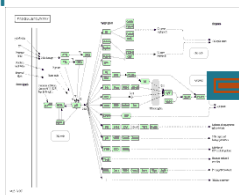
Ανάλυση

Οπτικοποίηση

Μικροσυστοιχίες



Γενετικά
κανονιστικά
μονοπάτια



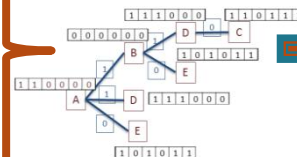
Διακριτοποίηση

	Normal			Disease		
	S1	S2	S3	S4	S5	S6
g1						
g2						
g3						
g9						
gN						

Αποσύνθεση

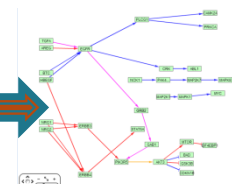


g1=hsa1
g2=hsa2
g3=hsa3
...
g7=hsa1



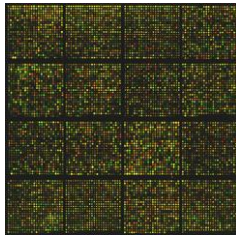
	Normal			Disease		
	S1	S2	S3	S4	S5	S6
p1	1	1	0	0	0	0
p2	0	0	0	0	0	0
p3	1	1	0	1	1	1
p4	1	1	1	0	0	0
pN	1	0	1	0	1	1

Αναζήτηση





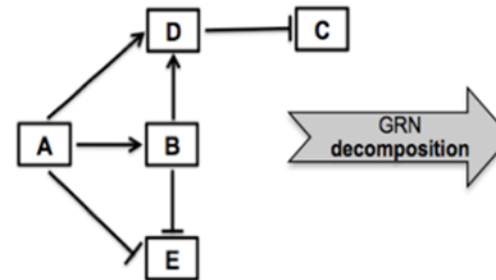
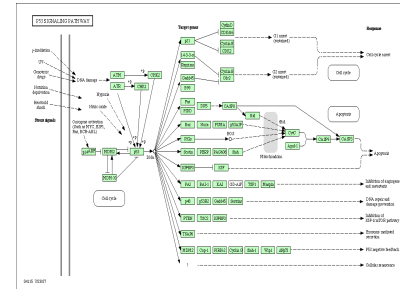
Προ-επεξεργασία



	Normal			Disease		
	S1	S2	S3	S4	S5	S6
A	98	78	23	43	1	9
B	34	23	3	22	11	12
C	79	66	12	80	82	67
D	89	91	77	12	43	33
E	80	20	78	12	89	99

Binary representation

	Normal			Disease		
	S1	S2	S3	S4	S5	S6
A	1	1	0	0	0	0
B	0	0	0	0	0	0
C	1	1	0	1	1	1
D	1	1	1	0	0	0
E	1	0	1	0	1	1



GRN decomposition

- 1 $A \rightarrow B \rightarrow D \rightarrow C$
- 2 $A \rightarrow B \rightarrow D$
- 3 $A \rightarrow B$
- 4 $A \rightarrow D \rightarrow C$
- 5 $A \rightarrow D$
- 6 $A \rightarrow B \rightarrow E$
- 7 $A \rightarrow E$
- 8 $B \rightarrow D \rightarrow C$
- 9 $B \rightarrow D$
- 10 $B \rightarrow E$
- 11 $D \rightarrow C$

Ονοματολογία

Probe-U133A	KEGG id
212607_at	hsa:10000
212609_s_at	hsa:10000
219393_s_at	hsa:10000
209706_at	hsa:4824
211497_x_at	hsa:4824
211498_s_at	hsa:4824
207163_s_at	hsa:207
203808_at	hsa:208
203809_s_at	hsa:208
211453_s_at	hsa:208
202723_s_at	hsa:2308
202724_s_at	hsa:2308

hsa:4824<-->hsa:10000 hsa:207 hsa:208-|hsa:2308

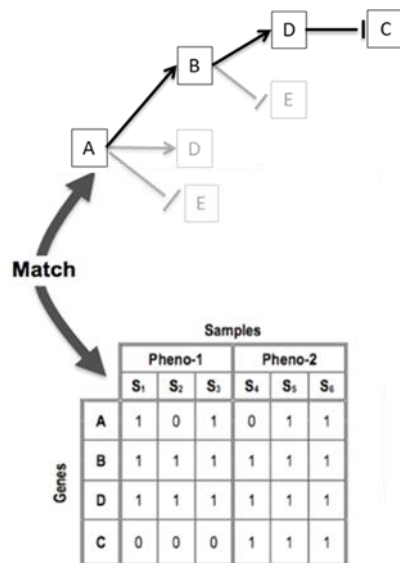
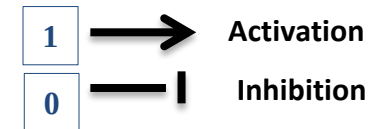
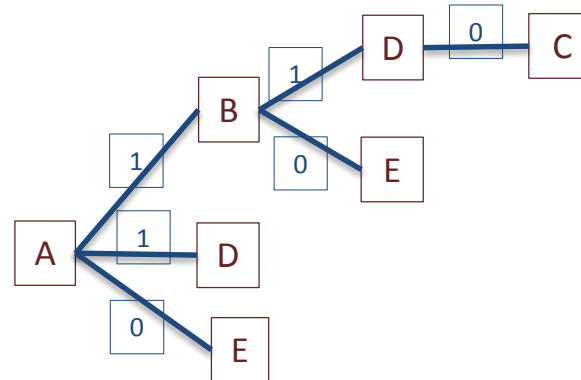
209706_at	212607_at	207163_s_at	203808_at	202723_s_at
211497_x_at	212609_s_at		203809_s_at	202724_s_at
211498_s_at	219393_s_at		211453_s_at	

ID	Περιγραφή	Γονίδια U133A	Υπο- μονοπάτια	Κλώνοι
hsa04010	MAPK signaling	481	1291	21109
hsa04012	ErbB signaling	164	486	4277
hsa04020	Calcium	335	157	189
hsa04110	Cell cycle	231	161	437
hsa04115	p53 signaling	123	277	1939
hsa04150	mTOR signaling	91	65	365
hsa04210	Apoptosis	157	145	1505
hsa04310	Wnt signaling	256	277	371
hsa04350	TGF-beta signaling	140	57	79
hsa04370	VEGF signaling	129	61	187
hsa04510	Focal adhesion	404	420	1275
hsa04520	Adherens junction	179	442	10873
hsa04912	GnRH signaling	205	145	1488
hsa05200	Pathways in cancer	634	988	16014

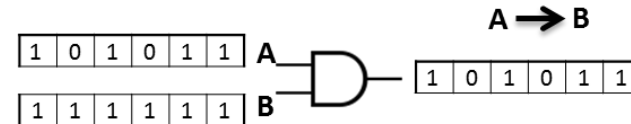


Συνένωση πηγών

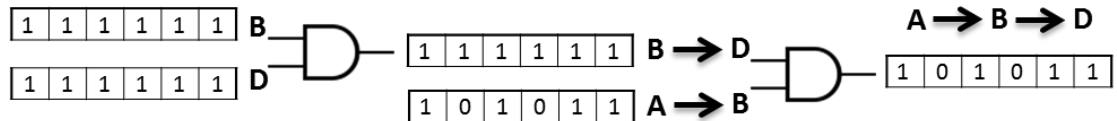
	Normal			Disease		
	S1	S2	S3	S4	S5	S6
A	1	1	0	0	0	0
B	0	0	0	0	0	0
C	1	1	0	1	1	1
D	1	1	1	0	0	0
E	1	0	1	0	1	1



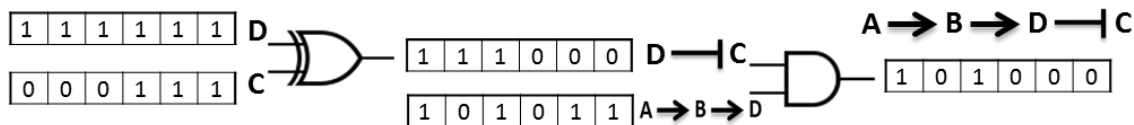
1



2



3

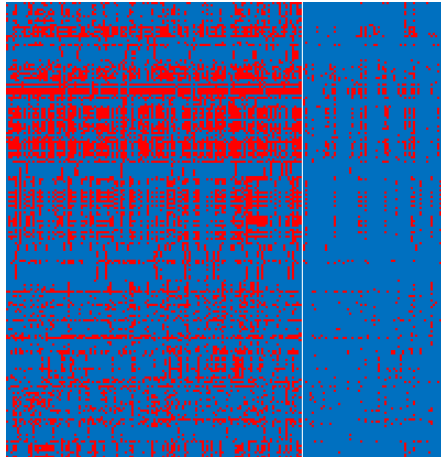


	Pheno-1			Pheno-2		
	S1	S2	S3	S4	S5	S6
A → B → D → C	1	0	1	0	0	0

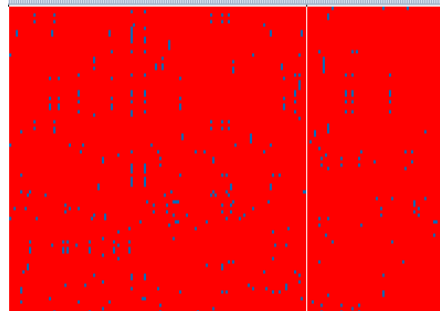
Ανάλυση

- ▶ Αλγόριθμοι μηχανικής μάθησης (για επαλήθευση)
 - Decision tree (C4.5), Support Vector Machines, naïve Bays
- ▶ Υπολογισμός μονοπατιών που χαρακτηρίζουν το φαινότυπο

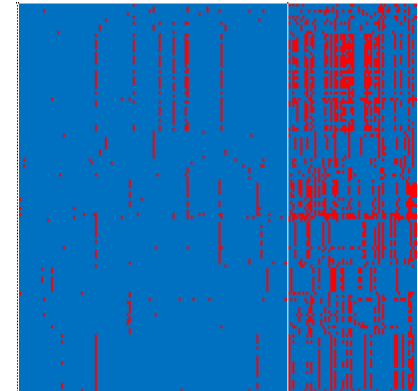
Φαινότυπος 1



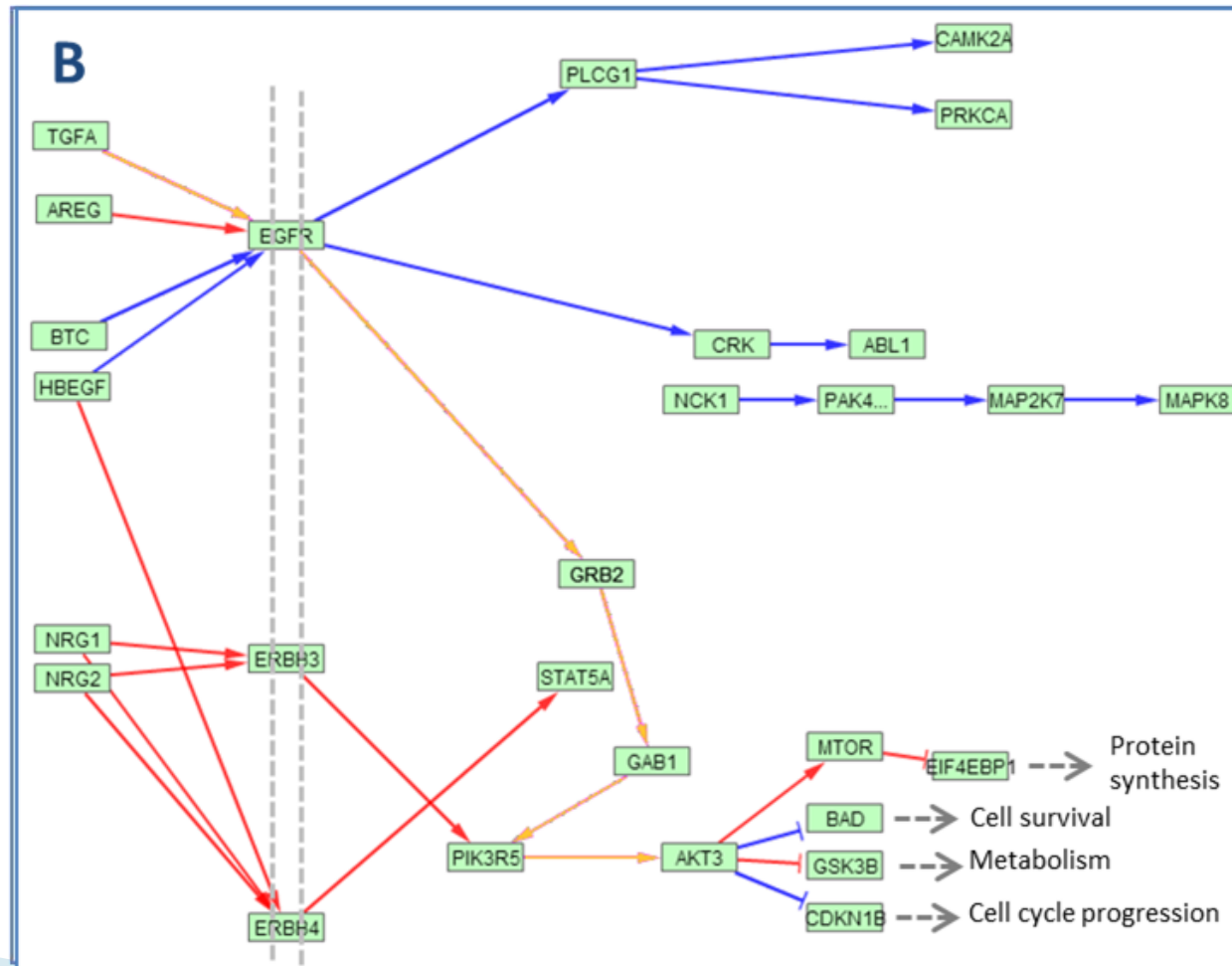
Κοινά μονοπάτια



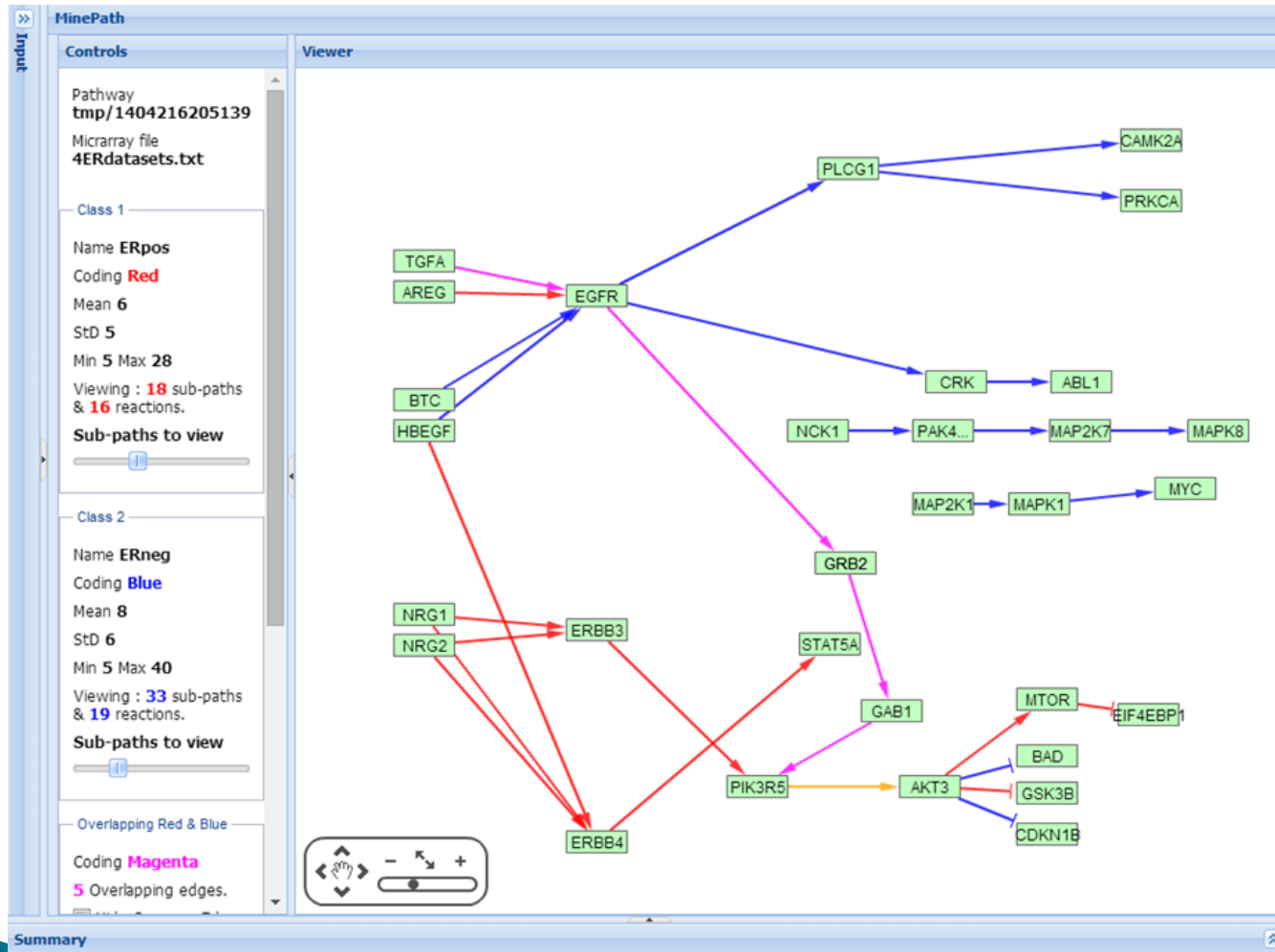
Φαινότυπος 2



Κοινά μονοπάτια



MinePath (web based tool)



ErbB pathway for
'4ERdatasets' dataset

<http://minepath.org>

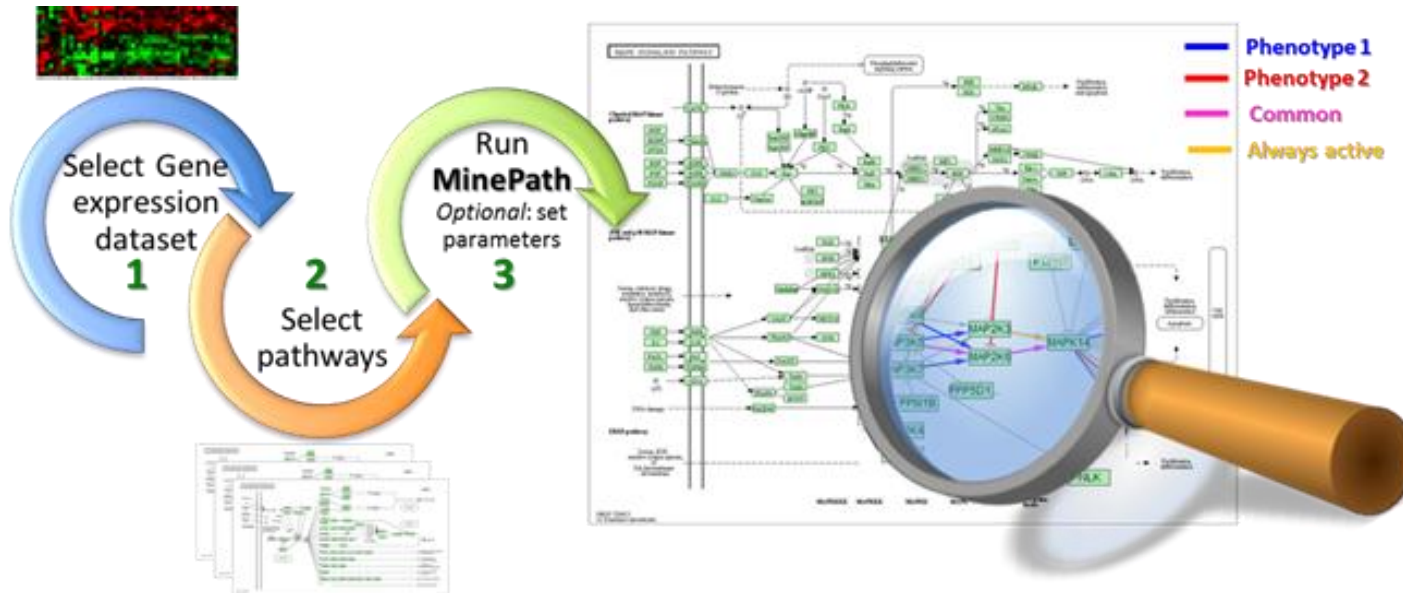
Λ. Κουμάκης
Χανιά 3 Ιουλίου 2014

Δημοσιεύσεις

- ▶ Koumakis L., Potamias G., Tsiknakis M., Zervakis M. and Moustakis V. Integrating Microarray Data and GRNs. Methods in Molecular Biology (under review)
- ▶ Koumakis, L., Moustakis, V.A., Zervakis, M.E., Kafetzopoulos, D., & Potamias, G.A. Coupling Regulatory Networks and Microarrays: revealing Molecular Regulations of Breast Cancer Treatment Responses. Artificial Intelligence: Theories and Applications. Lecture Notes in Computer Science, 7297, 239–246 (2012).
- ▶ Koumakis, L., Potamias, G.A., Zervakis, M.E., & Moustakis, V.A. (2011). Integrating microarray data and gene regulatory networks: Survey and critical considerations. 10th International Workshop on Biomedical Engineering. Kos, Greece 5–7 October 2011.
- ▶ K. Kalantzaki, L. Koumakis, E. S. Bei, M. Zervakis, G. Potamias and D. Kafetzopoulos. Experimental Model Construction and Validation of the ErbB Signaling Pathway. 13th IEEE International Conference on Bioinformatics and Bioengineering. Chania, Greece, November 10–13, 2013



Ευχαριστώ



<http://minepath.org>



Η παρούσα έρευνα έχει συγχρηματοδοτηθεί από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο - ΕΚΤ) και από εθνικούς πόρους μέσω του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» του Εθνικού Στρατηγικού Πλαισίου Αναφοράς (ΕΣΠΑ) – Ερευνητικό Χρηματοδοτούμενο Έργο: Ηράκλειτος ΙΙ . Επένδυση στην κοινωνία της γνώσης μέσω του Ευρωπαϊκού Κοινωνικού Ταμείου.

Λ. Κουμάκης
Χανιά 3 Ιουλίου 2014