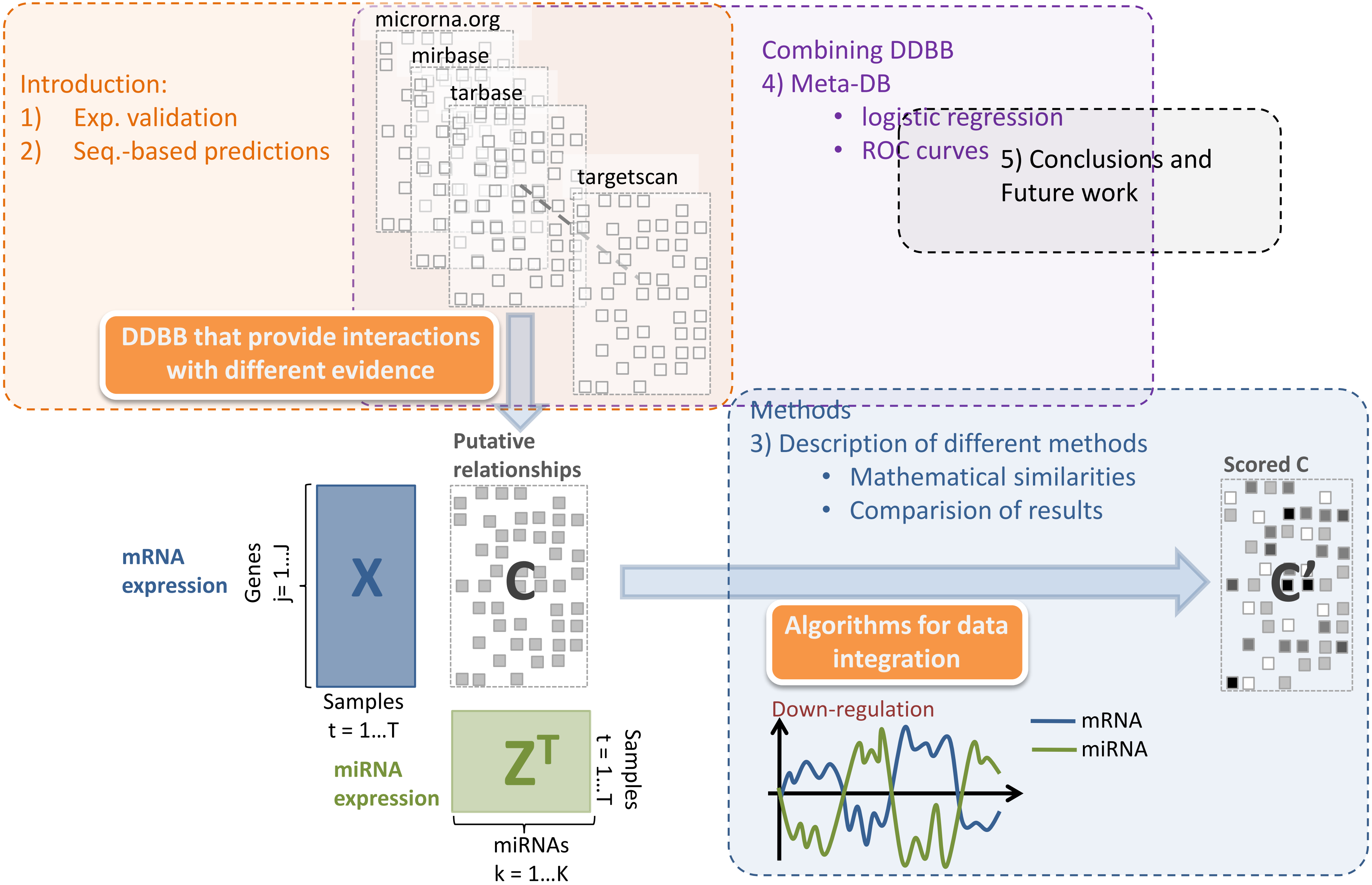
Joint analysis of miRNA and mRNA expression data

Angel Rubio, TECNUN-CEIT, Helsinki 9/01/2014

Summary



Introduction:

1) Exp. validation
2) Seq.-based predictions

Combining DDBB

4) Meta-DB
- logistic regression
- ROC curves

5) Conclusions and Future work

Methods

3) Description of different methods
- Mathematical similarities
- Comparision of results

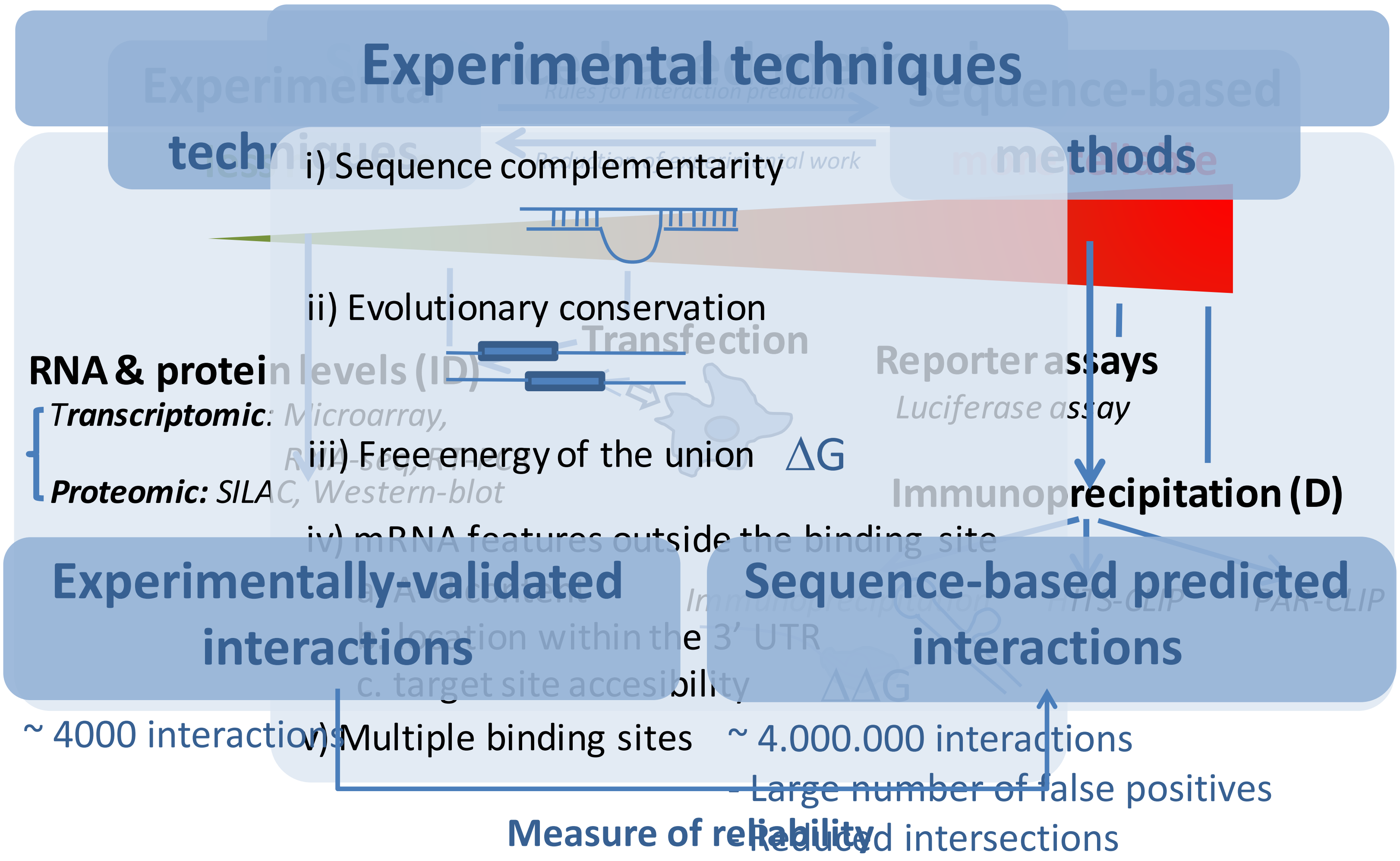
DDBB that provide interactions with different evidence

Algorithms for data integration

Methods

Introduction

Seeking miRNA-mRNA interactions



Experimental techniques

Sequence-based methods

Rules for interaction prediction

Reduction of experimental work

more reliable

i) Sequence complementarity

ii) Evolutionary conservation

iii) Free energy of the union ΔG

iv) mRNA features outside the binding site

a. A-U content

b. location within the 3’ UTR

c. target site accesibility

ΔΔG

v) Multiple binding sites

Transfection

RNA & protein levels (ID)

Transcriptomic: *Microarray, RNA-seq, RT-PCR*

Proteomic: *SILAC, Western-blot*

Reporter assays

Luciferase assay

Immunoprecipitation (D)

Immunoprecipitation *HITS-CLIP* *PAR-CLIP*

Experimentally-validated interactions

Sequence-based predicted interactions

~ 4000 interactions

~ 4.000.000 interactions

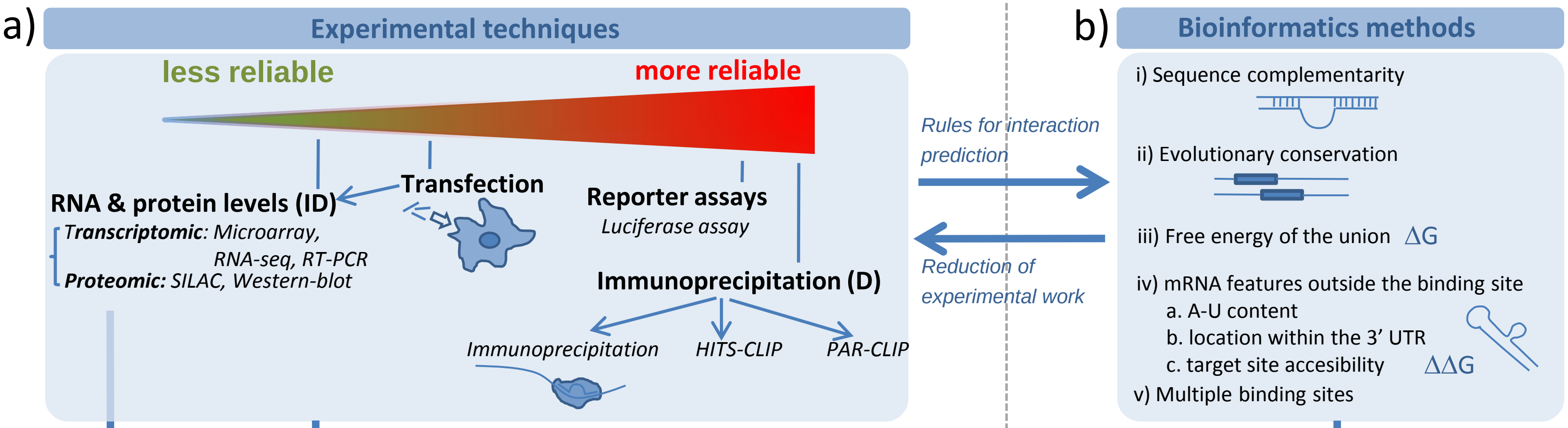
- Large number of false positives

- Reduced intersections

Measure of reliability

(Joint Analysis of miRNA-mRNA expression data; 2012; Briefings in Bioinformatics)

DDBB


a)
Experimental techniques
less reliable
more reliable
Transfection
RNA & protein levels (ID)
Transcriptomic: Microarray, RNA-seq, RT-PCR
Proteomic: SILAC, Western-blot
Reporter assays
Luciferase assay
Immunoprecipitation (D)
Immunoprecipitation
HITS-CLIP
PAR-CLIP
Rules for interaction prediction
Reduction of experimental work
b)
Bioinformatics methods
i) Sequence complementarity
ii) Evolutionary conservation
iii) Free energy of the union ΔG
iv) mRNA features outside the binding site
a. A-U content
b. location within the 3' UTR
c. target site accesibility ΔΔG
v) Multiple binding sites

c) **Experimentally validated interactions**

	Type	# int	TaRBase	miRecords	miRtarbase	miRWalk
TaRBase	m	877		⬛	🟥	🟩
miRecords	m	1273	🟩		🟥	⬛
miRtarbase	m	2781	🟩	🟩		🟩
miRWalk	t-m	5564	🟩	🟩	🟩	

Intersection: 🟥 70% ⬛ 35% 🟩 5%

m = manually curated t-m = text mining

Sequence-based predicted interactions

	based on										
	i)	ii)	iii)	iv)	v)	miRanda	mirBase	TargetScan	DIANA microT	PicTar	PITA
miRanda	x	x	x		x		🟩	🟩	🟥	🟩	🟩
mirBase	x	x	x		x	🟩		🟩	⬛	⬛	🟩
TargetScan	x	x	x	a),b)	x	🟥	🟩		🟥	🟥	🟥
DIANA microT	x	x	x			⬛	🟩	🟩		⬛	🟩
PicTar	x	x	x		x	🟩	⬛	🟥	🟩		🟩
PITA	x	x	x	c)	x	🟥	🟩	🟥	🟥	🟥	

Intersection: 🟥 70% ⬛ 35% 🟩 5%

d) **Filter putative targets**

HOCTAR	GenMiR++	MAGIA
TaLasso	Elastic-net	Partial Least Squares
Hctarget	Graphical Bayes	

Joint analysis of miRNA and mRNA expression data, Briefings in Bioinformatics, 2012, Muniategui et al.

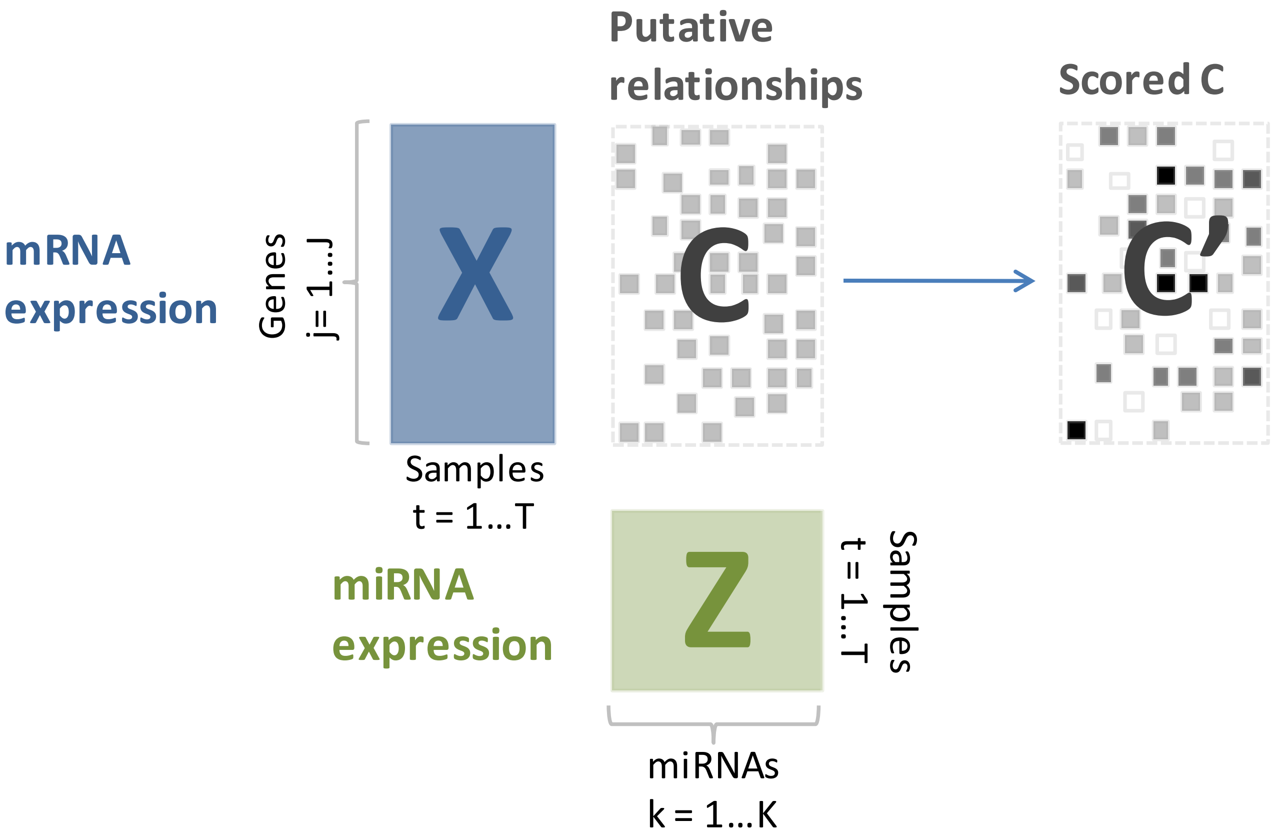
Questions regarding these databases

- They are VERY different when compared
  - Very different in sizes.
  - Different methodology
- Are they all equally reliable?
- How can we use the score that the DDBB provide?
- How do we combine them?
  - Union
  - Intersection
  - “At least in two of them”…
  - Any other arbitrary rule???
- We will return to answer these questions later.

Methods

Integration of expression data (miRNA and mRNA)

Methods to integrate expression



(Quantification of miRNA-mRNA interactions; 2012; PLoS ONE)

Methods

- **Pairwise analysis**
  - Correlation (Pearson and Spearman)
    - “Rank the annotated relationships according to the pairwise correlation: the more negative the correlation the higher the rank”
      - Absolute correlation? Are positive correlations significant?
  - Mutual information (MAGIA)
    - “Rank the annotated relationships according to the Mutual information”
      - Borrowed from information theory.
      - Ranking is similar to absolute correlation, i.e. the direction of the regulation is not taken into account.

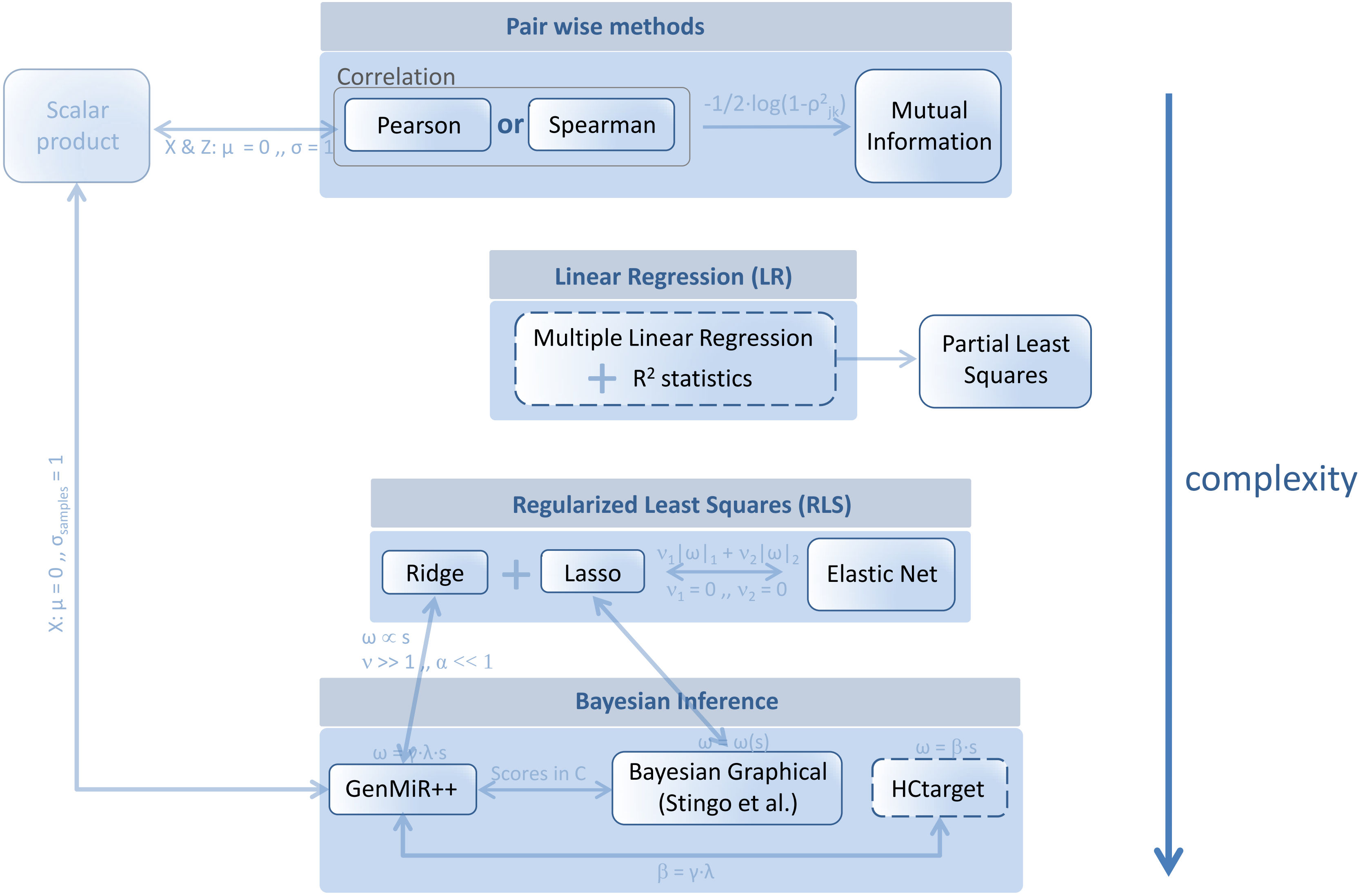
Methods

- **(Regularized) Linear models**
  - “Rank the annotated relationships according to its weight in a (regularized) linear model”
    - The p.value of the coefficient can also be used to rank the interactions.
    - mRNA expression as a linear combination of the miRNAs that putatively bind to it.
    - Sometimes the problem is stated as an inverse problem, i.e. the expression of the miRNA is a linear combination of their putative targets.
    - Usually more miRNAs than samples → Regularization, i.e., take only the most prominent interactions (Lasso, Ridge)
      - Some implementations cannot be applied simply because of this.

Methods

- **Bayesian Methods**
  - "Rank the annotated relationships according to their probabilities of being significant setting some sensible priors"
    - GenMir++, HCTarget, Graphical Bayesian

Methods to integrate expression in



Pair wise methods

Correlation

Pearson or Spearman

$-1/2\cdot\log(1-\rho^2_{jk})$

Mutual Information

Scalar product

X & Z: $\mu = 0$,, $\sigma = 1$

Linear Regression (LR)

Multiple Linear Regression + R^2 statistics

Partial Least Squares

Regularized Least Squares (RLS)

Ridge + Lasso

$\nu_1|\omega|_1 + \nu_2|\omega|_2$

$\nu_1 = 0$,, $\nu_2 = 0$

Elastic Net

X: $\mu = 0$,, $\sigma_{samples} = 1$

$\omega \propto s$

$\nu >> 1$,, $\alpha << 1$

Bayesian Inference

$\omega = \gamma\cdot\lambda\cdot s$

GenMiR++

Scores in C

$\omega = \omega(s)$

Bayesian Graphical (Stingo et al.)

$\omega = \beta\cdot s$

HCtarget

$\beta = \gamma\cdot\lambda$

complexity

(Joint analysis of miRNA and mRNA expression data; Briefings in bioinformatics; June 2012)

Talasso

Assumptions

1) Neglect any other regulator of gene expression but **miRNAs**

2) Only **down-regulatory** effects are considered

3) The aim is to **filter** putative interactions

4) **Linear relationship** between logarithms of expressions is assumed

Mathematical model

For each gene

$$\mathbf{x_j} = \sum_{k=1}^{K} \overset{\text{amount of down-regulation}}{\boxed{\beta_{jk}}} \cdot c_{jk} \cdot \mathbf{z_k} + \underset{\text{basal expression}}{\boxed{\mathbf{x_k^0}}} + \epsilon_{\mathbf{j}}$$

LASSO with non-negative constraints

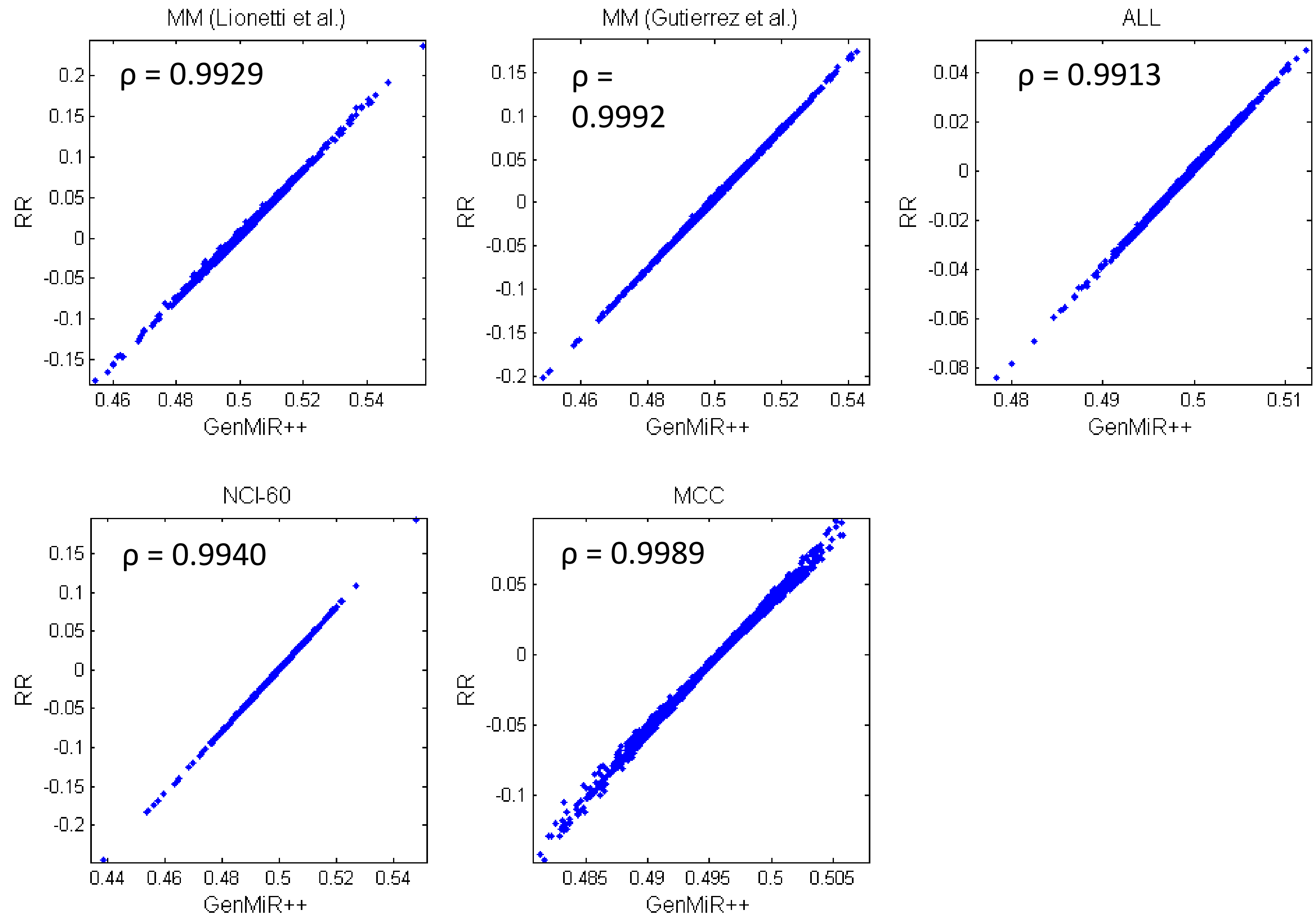
$$\underset{\beta_j, x_j^0}{\text{mín}} \left\{ \left\| \mathbf{x_j} - \sum_{k=1}^{K} \beta_{jk} \cdot c_j \cdot \mathbf{z_k} - \mathbf{x_j^0} \right\|_2 + \lambda_j \cdot \sum_{k=1}^{K} |\beta_{jk} \cdot c_{jk}| \right\}$$

$$\text{subject to } \beta_{jk} \leq 0, \text{ for k} = 1,2,\ldots,\text{K}.$$

Interesting relationships

- Genmir++ (with the values of the parameters of the authors) is equivalent to Ridge regression with a extremely large regularization parameter.
- In turn, both of them are equivalent to a scalar product.
  - Applying this shortcut, GenMir++ becomes 4 orders of magnitude faster.
- The only difference between correlation and GenMir++ is the normalization:
  - In correlation, both miRNA and mRNA are normalized.
  - In GenMir++, only mRNA are normalized
    - This subtle difference makes the results very different.

Other methods: GenMiR++ and a scalar product



Other methods: expression data used

Multi Class Cancer (MCC)

miRNA ,, bead-based flow cytometry (Lu J. et al. *Nature* 2005)
mRNA ,, Hu6800 and Hu35KsubA GeneChips (Affymetrix)
(Ramaswamy S. et al. *Proc.Natl.Acad.Sci.U.S.A)*

88 samples (paired):

normal and cancerous: bladder, breast, colon, kidney, lung, pancreas, prostate and uterus

cancerous without normal ref.: ovary cancer, melanoma and mesothelioma

NCI-60

miRNA ,, PCR (TaqMan) (Gaur A. et al *Cancer Res* 2007)

mRNA ,, HG-U95 A & HG-U133 (Affymetrix) (Shankavaram U.T. et al *Molecular Cancer Therapeutics* 2007) ,, http://discover.nci.nih.gov/cellmier/home.do

59 samples (paired):

(9 cancer types)
breast, glioblastoma, colon, lung, leukemia, melanoma, ovarian, prostate and renal

Acute Lymphoblastic Leukemia (LDS)

GEO (GSE14834) (Fulci V. et al *Genes Chromosomes Cancer 2009*)
miRNA ,, miRHuman 9.0 array (LC Sciences)
mRNA ,, Human Genome GeneChip U133 Plus 2.0 Array (Affymetrix)

19 samples (paired):

B-ALL: BCR/ABL ,, E2A/PBX1 ,, MLL/AF4 ,, no translocation

T-ALL: SIL/TAL ,, no translocation

Multiple Myeloma (MM) (Lionetti et al.)

miRNA ,, Agilent Human miRNA V2 (Lionetti et al. *Blood* 2009)
mRNA ,, Affymetrix GeneChip HG-U133A (Lionetti et al. *Blood* 2009)

40 samples (paired):

38 MM and 2 Plasma Cells divided into 5 groups attempting to translocations and gene expression values.

Multiple Myeloma (MM) (Gutierrez et al.)

miRNA ,, TaqMan low-density arrays (Gutierrez et al. *Leukemia* 2010)
mRNA ,, Affymetrix Human Gene 1.0 ST (Gutierrez et al. *Leukemia* 2010)

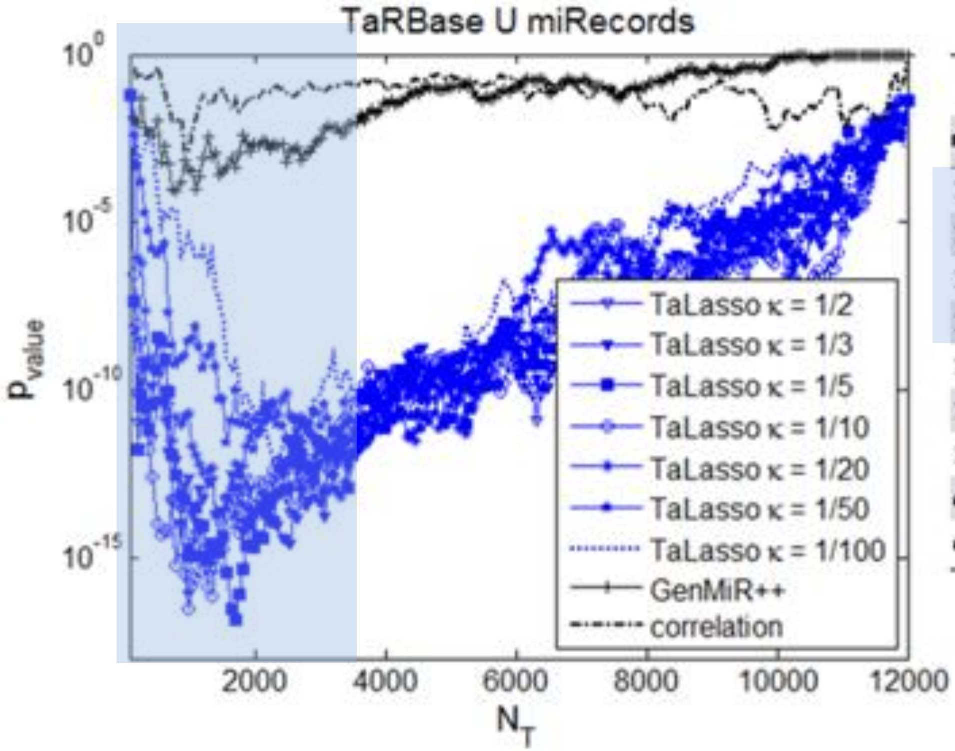
65 samples (paired):

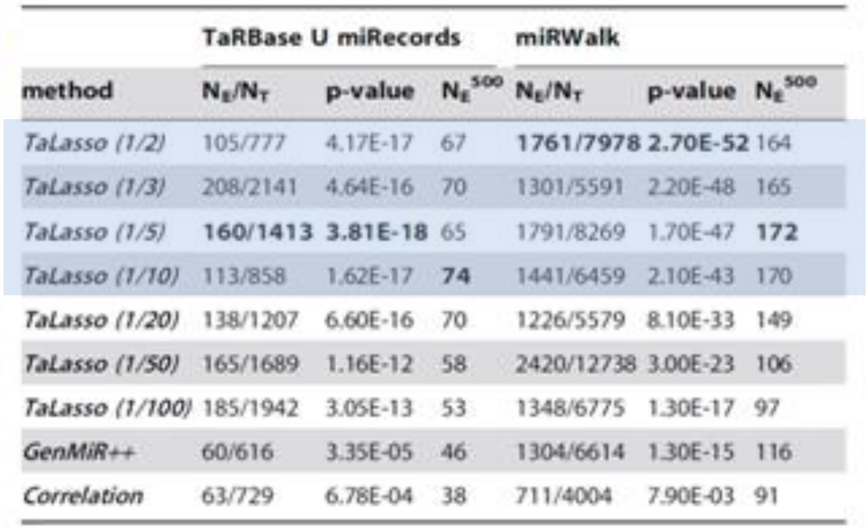
60 MM and 5 normal. MM samples divided into 4 groups: RB deletions, t(11;14), t(14;16) and t(4;14) translocations.

Results for MCC

Enrichment in EV interactions

“Good algorithm: top-ranked interactions more enriched in EV interactions”





	TaRBase U miRecords			miRWalk		
method	N_E/N_T	p-value	N_E^{500}	N_E/N_T	p-value	N_E^{500}
TaLasso (1/2)	105/777	4.17E-17	67	**1761/7978**	**2.70E-52**	164
TaLasso (1/3)	208/2141	4.64E-16	70	1301/5591	2.20E-48	165
TaLasso (1/5)	**160/1413**	**3.81E-18**	65	1791/8269	1.70E-47	**172**
TaLasso (1/10)	113/858	1.62E-17	**74**	1441/6459	2.10E-43	170
TaLasso (1/20)	138/1207	6.60E-16	70	1226/5579	8.10E-33	149
TaLasso (1/50)	165/1689	1.16E-12	58	2420/12738	3.00E-23	106
TaLasso (1/100)	185/1942	3.05E-13	53	1348/6775	1.30E-17	97
GenMiR++	60/616	3.35E-05	46	1304/6614	1.30E-15	116
Correlation	63/729	6.78E-04	38	711/4004	7.90E-03	91

N_E = # EV

N_T = # drawn

N_E^{500} = # EV in top-500

(Quantification of miRNA-mRNA interactions; 2012; PLoS ONE)

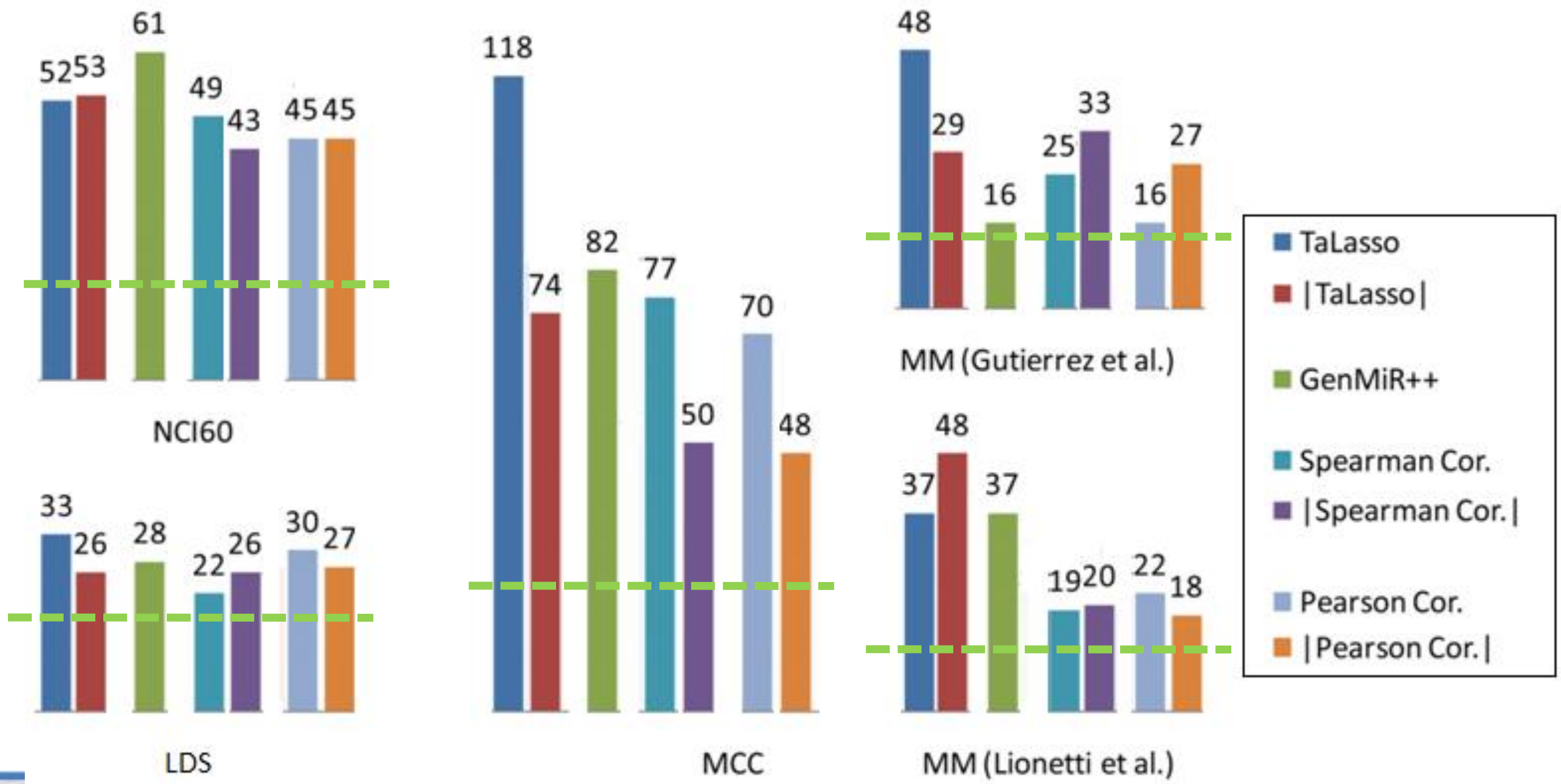
Results

The added value of using expression data

- Are results using expression data more enriched in EV interactions than the initial set of putative interactions?

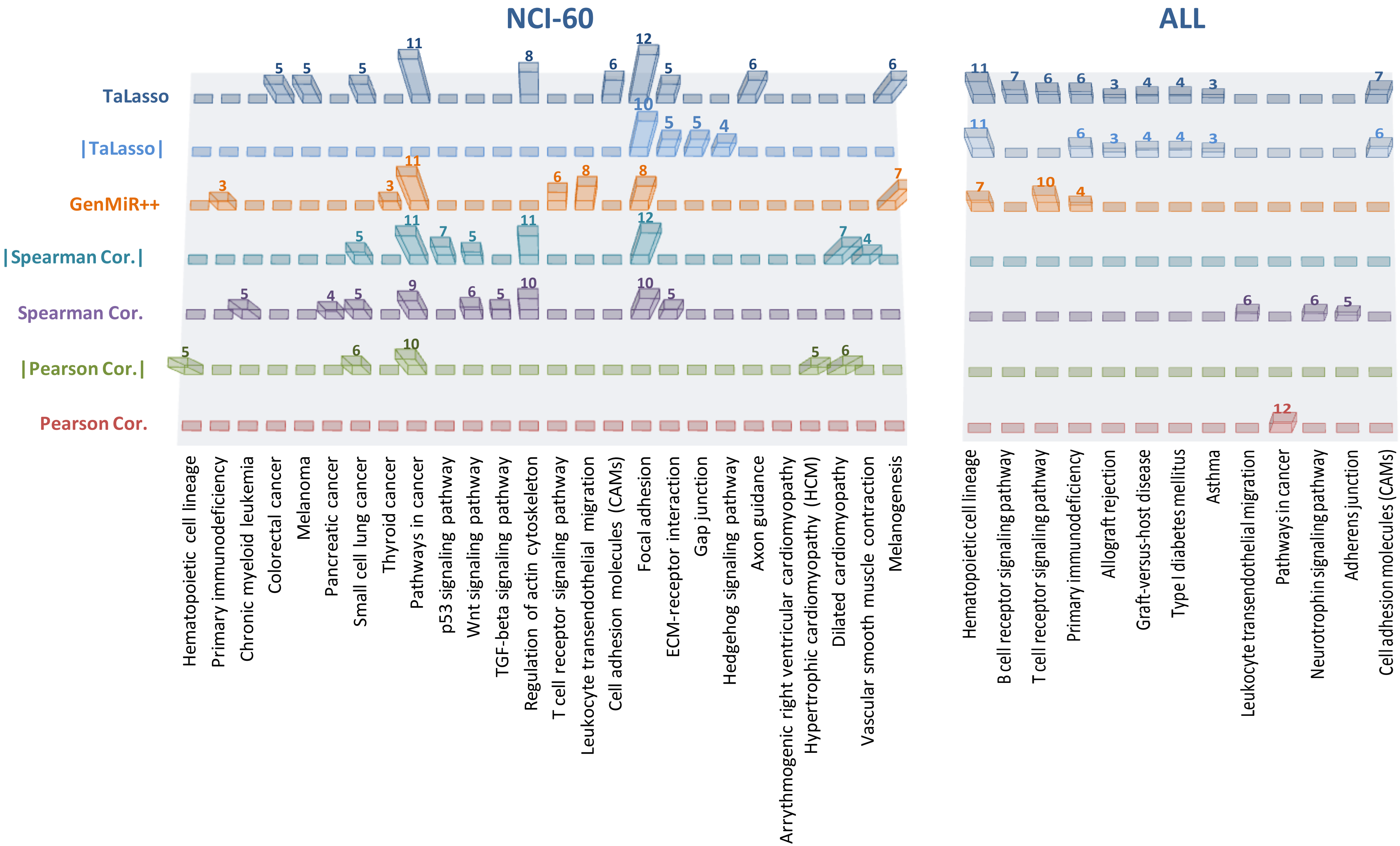
$$E = \frac{\overset{\#\text{ drawn}}{n} \cdot \overset{\#\text{ EV}}{m}}{\underset{\#\text{ union}}{N}}$$

Comparison of algorithms

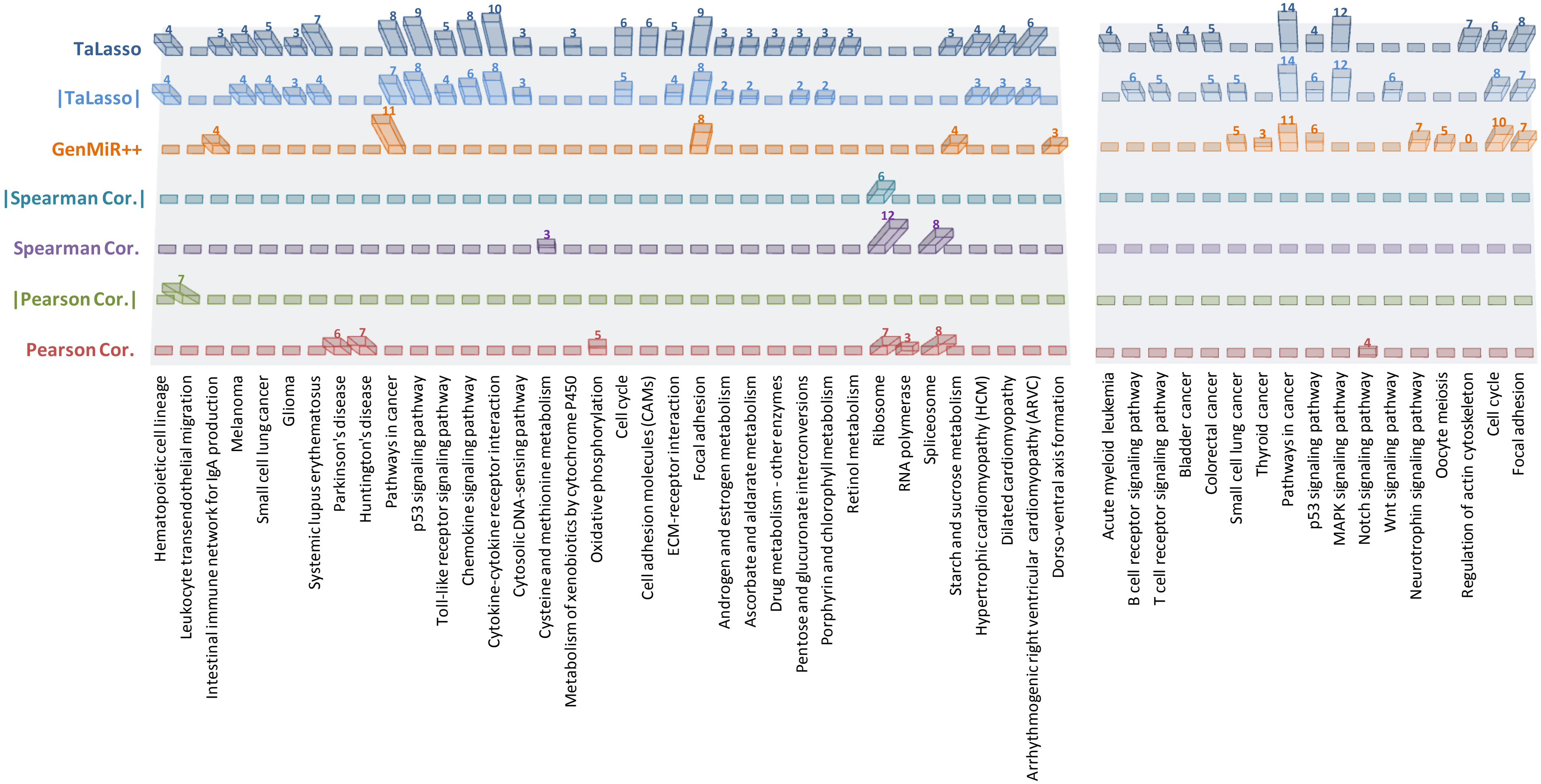


(Joint Analysis of miRNA-mRNA expression data; 2012; Briefings in Bioinformatics)

Comparisons: enrichment in KEGG pathways



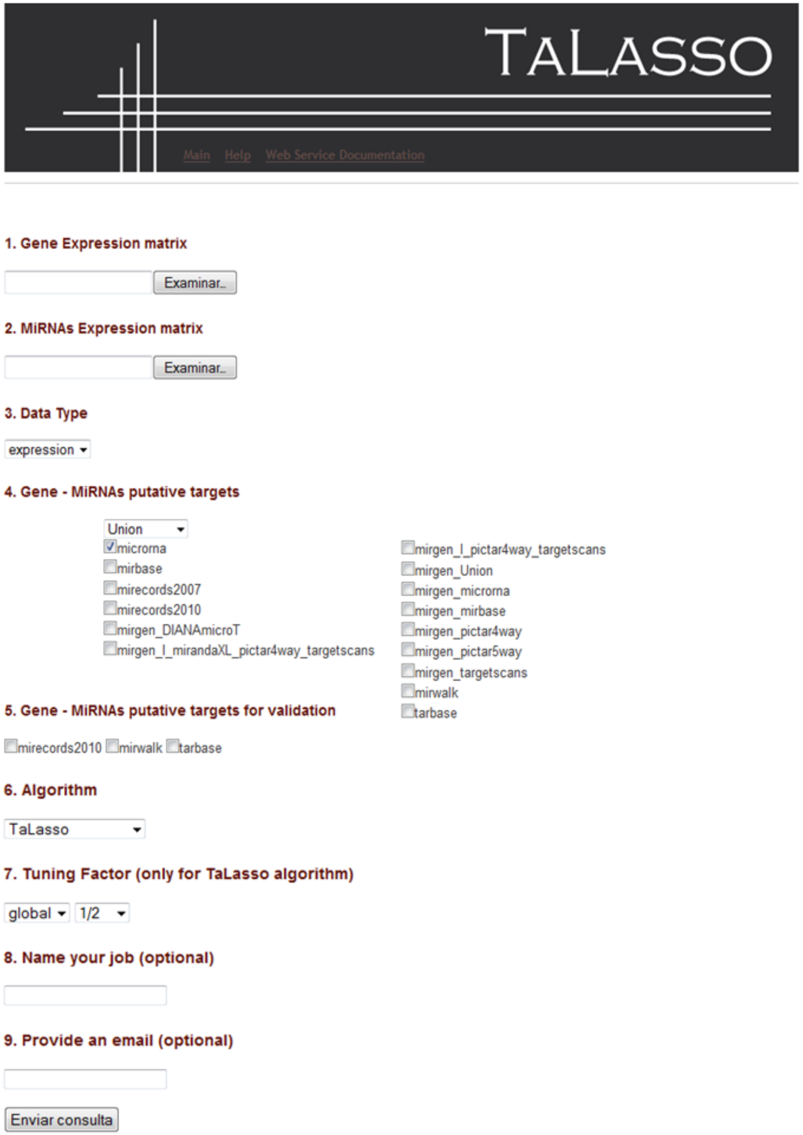
Comparisons: enrichment in KEGG pathways



MM (Gutierrez et al.)

MM (Lionetti et al.)

TaLasso: web application



1. Gene Expression matrix

Examinar...

2. MiRNAs Expression matrix

Examinar...

3. Data Type

expression

4. Gene - MiRNAs putative targets

Union

☑ microrna
□ mirbase
□ mirecords2007
□ mirecords2010
□ mirgen_DIANAmicroT
□ mirgen_I_mirandaXL_pictar4way_targetscans
□ mirgen_I_pictar4way_targetscans
□ mirgen_Union
□ mirgen_microrna
□ mirgen_mirbase
□ mirgen_pictar4way
□ mirgen_pictar5way
□ mirgen_targetscans
□ mirwalk
□ tarbase

5. Gene - MiRNAs putative targets for validation

□ mirecords2010 □ mirwalk □ tarbase

6. Algorithm

TaLasso

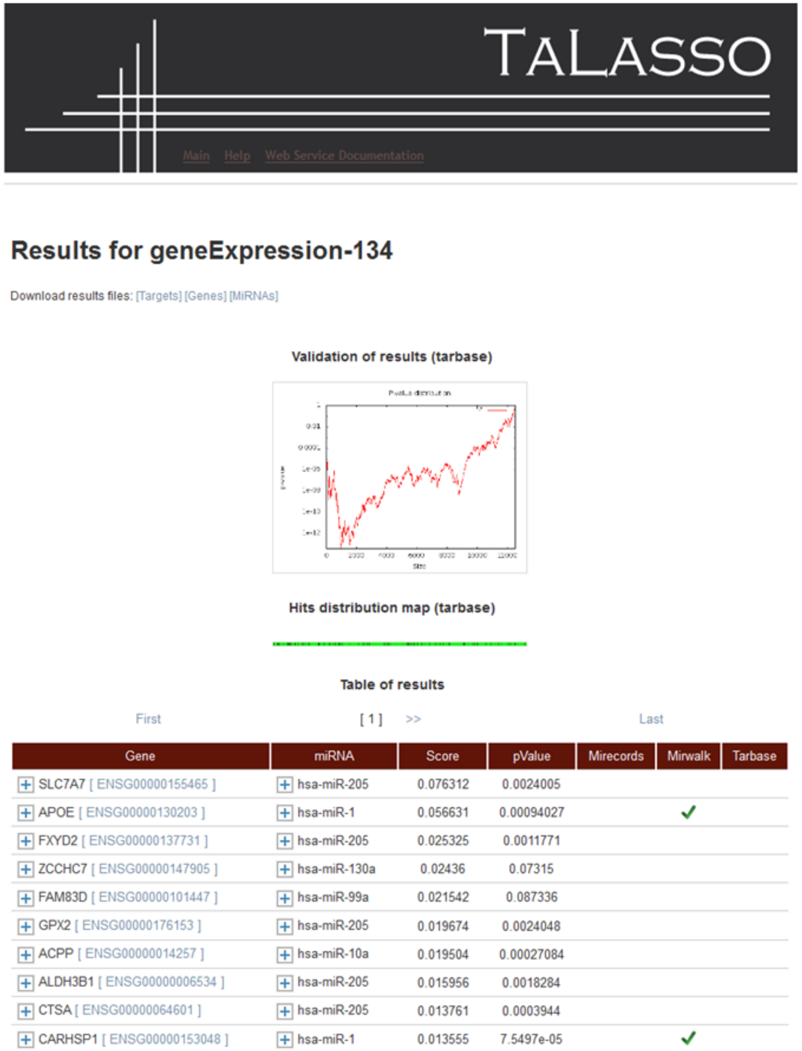
7. Tuning Factor (only for TaLasso algorithm)

global 1/2

8. Name your job (optional)

9. Provide an email (optional)

Enviar consulta



Results for geneExpression-134

Download results files: [Targets] [Genes] [MiRNAs]

Validation of results (tarbase)

Hits distribution map (tarbase)

Table of results

First [1] >> Last

Gene	miRNA	Score	pValue	Mirecords	Mirwalk	Tarbase
SLC7A7 [ENSG00000155465]	hsa-miR-205	0.076312	0.0024005			
APOE [ENSG00000130203]	hsa-miR-1	0.056631	0.00094027		✓	
FXYD2 [ENSG00000137731]	hsa-miR-205	0.025325	0.0011771			
ZCCHC7 [ENSG00000147905]	hsa-miR-130a	0.02436	0.07315			
FAM83D [ENSG00000101447]	hsa-miR-99a	0.021542	0.087336			
GPX2 [ENSG00000176153]	hsa-miR-205	0.019674	0.0024048			
ACPP [ENSG00000014257]	hsa-miR-10a	0.019504	0.00027084			
ALDH3B1 [ENSG00000006534]	hsa-miR-205	0.015956	0.0018284			
CTSA [ENSG00000064601]	hsa-miR-205	0.013761	0.0003944			
CARHSP1 [ENSG00000153048]	hsa-miR-1	0.013555	7.5497e-05		✓	

(Quantification of miRNA-mRNA interactions; Plos One; Feb. 2012)

http://talasso.cnb.csic.es/

Conclusions

- Imposing only negative relationship provides more biologically enriched pathways and validated interactions
  - Nevertheless, using positive correlations, the results are biologically interesting and complementary to the previous ones.
- Talasso seems to perform well the shown datasets.
- GenMir++ works better than plain correlation.
  - The only difference is the normalization method.
  - Is it more enriched in experimentally validated interactions because more expressed miRNA are easier to be validated?
    - Also biological significance.

Combination of DDBB

A meta-DB based on logistic regression

DDBB for interactions

- There are many databases of interactions of miRNA-mRNA
- Two main groups:
  - **Experimentally validated**
    - “Curated data”
    - High reliability…
    - **…but some experimental methods are more reliable than others**.
    - Very few interactions (1,000’s)
  - **Predicted** by sequence and other methods
    - Only computer predictions.
    - Low reliability…
    - **…but some of them are even less reliable**
    - Tons of interactions (100,000 to 1,000,000’s for each database)
    - Usually they provide a score for each interaction.

Questions to address

- Different DDBB provide different scores to rank the quality of the interactions. These scores cannot be compared among them.
  - Is it possible to have a unified score to compare the evidence of an interaction in different DDBB?
  - As a side effect, can this score also be used measure of the quality of the DDBB?

- In some cases (less than expected), a interaction is predicted by different DDBB (of course, with different scores).
  - Is it possible to provide a overall score that combines all the sources of evidence?

(Improving miRNA-mRNA Interaction Prediction; 2013; Bioinformatics; Submitted)

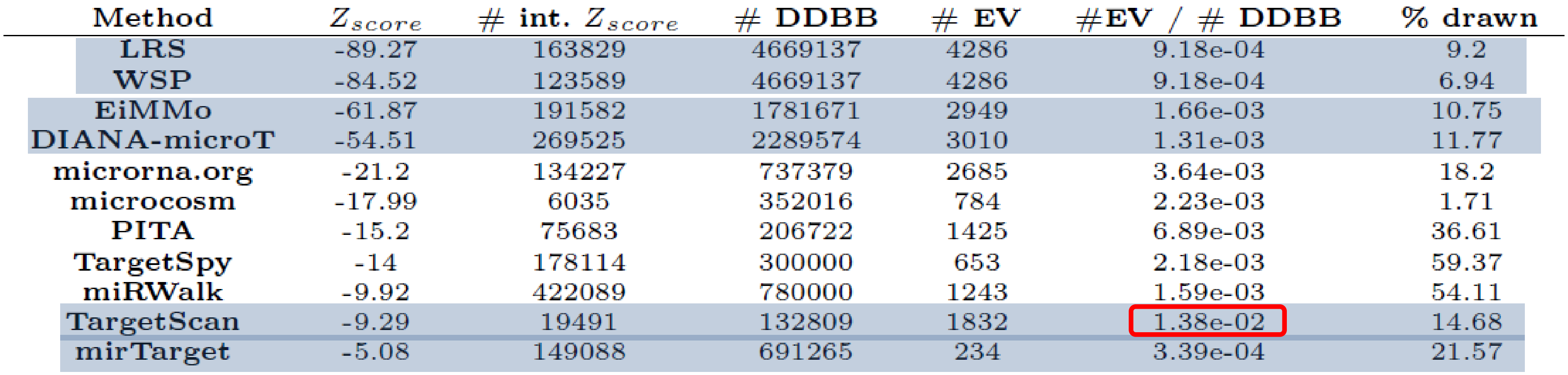
Reliability of the DDBB

- It is **difficult to compare the reliability** of DDBB due to:

    1) Differences in sizes
    2) Differences in qualities of the scores

- Compare DDBB using the **hypergeometric test**:

    1) Sort interactions by their scores
    2) Run hypergeometric test for each interaction
    3) Determine the position of the minimum p-value (# of interactions drawn)



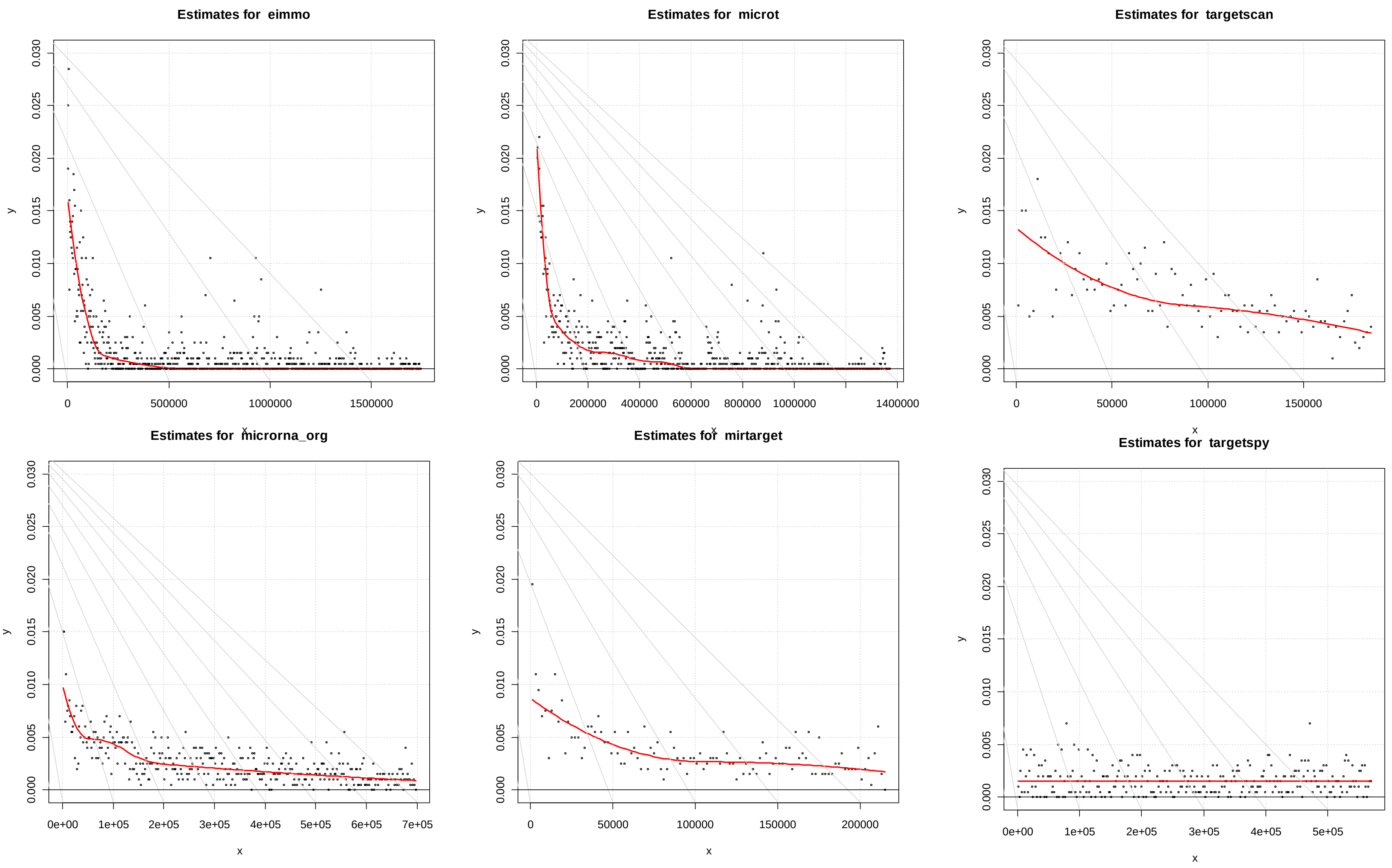
Method	Z_{score}	# int. Z_{score}	# DDBB	# EV	#EV / # DDBB	% drawn
LRS	-89.27	163829	4669137	4286	9.18e-04	9.2
WSP	-84.52	123589	4669137	4286	9.18e-04	6.94
EiMMo	-61.87	191582	1781671	2949	1.66e-03	10.75
DIANA-microT	-54.51	269525	2289574	3010	1.31e-03	11.77
microrna.org	-21.2	134227	737379	2685	3.64e-03	18.2
microcosm	-17.99	6035	352016	784	2.23e-03	1.71
PITA	-15.2	75683	206722	1425	6.89e-03	36.61
TargetSpy	-14	178114	300000	653	2.18e-03	59.37
miRWalk	-9.92	422089	780000	1243	1.59e-03	54.11
TargetScan	-9.29	19491	132809	1832	1.38e-02	14.68
mirTarget	-5.08	149088	691265	234	3.39e-04	21.57

(Improving miRNA-mRNA Interaction Prediction; 2013; Bioinformatics; Submitted)

New score of interactions for each database

- **Important assumption:**
  - The **quality score** used in this presentation is the **probability of being experimentally validated P(EV).**
  - Using the different scores, we can state P(EV).
  - P(EV) must be computed for every interaction in every database.
- **Recipe to get an estimate of P(EV)**
  - Rank the interactions according to their score (better are first).
  - Group them in bins of interactions and compute the proportion of experimentally validated interactions within each group.
  - Join the estimated probabilities by a smoothing spline that is constrained to be in [0,1] (since it is a probability), and non-increasing
    - The reason of this restriction, is that we assume that a better score provides a more reliable interaction.

Results of the ranking (identical y axis, different x axis for the DDBB)



Estimates for eimmo

Estimates for microt

Estimates for targetscan

Estimates for microrna_org

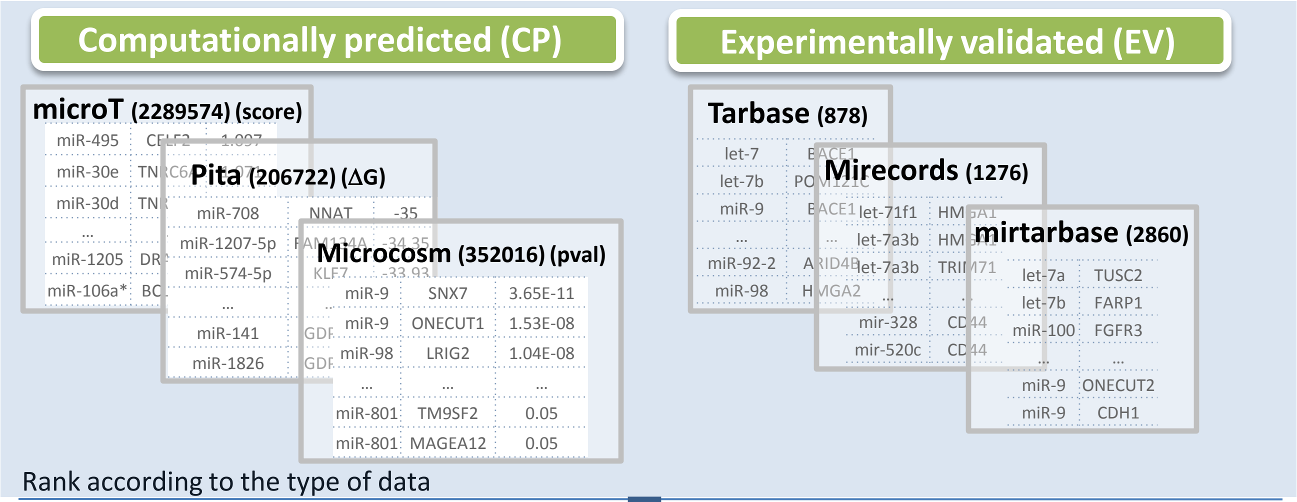
Estimates for mirtarget

Estimates for targetspy

Unified score

- Given that the score of each of the interactions are probabilities, it can be combined by applying a **logistic regression** to provide a unified probability.
  - We have run a logistic regression taking into account second order interactions
  - This approach helps to prevent the problem of the redundancy of the databases.
- After running the regression we have a **unified score** for each interaction that appear in all the databases
  - The number of interactions is the union of the interactions in all the databases.
  - The score is the probability of being experimentally validated.
- How good is this score?→ ROC curves comparing the unified database with each of the DDBBs.

Getting a global score


Computationally predicted (CP)
Experimentally validated (EV)
microT (2289574) (score)
miR-495 CELF2 1.097
miR-30e TNRC6
miR-30d TN
…
miR-1205 DR
miR-106a* BC
Pita (206722) (ΔG)
miR-708 NNAT -35
miR-1207-5p FAM134A -34.35
miR-574-5p KLF7 -33.93
…
miR-141 GDF
miR-1826 GDF
Microcosm (352016) (pval)
miR-9 SNX7 3.65E-11
miR-9 ONECUT1 1.53E-08
miR-98 LRIG2 1.04E-08
… … …
miR-801 TM9SF2 0.05
miR-801 MAGEA12 0.05
Tarbase (878)
let-7 BACE1
let-7b POMT21C
miR-9 BACE1
…
miR-92-2 ARID4B
miR-98 HMGA2
Mirecords (1276)
let-71f1 HMGA1
let-7a3b HMGA1
let-7a3b TRIM71
…
mir-328 CD44
mir-520c CD44
mirtarbase (2860)
let-7a TUSC2
let-7b FARP1
miR-100 FGFR3
… …
miR-9 ONECUT2
miR-9 CDH1

Rank according to the type of data

Re-score each interaction: probability of being experimentally-validated given its score in the database

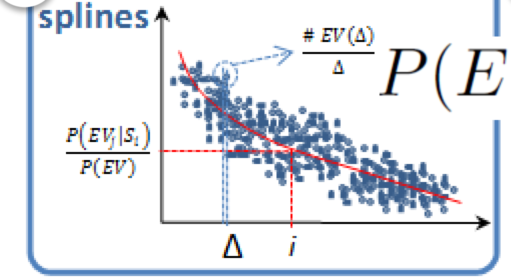
Combine computationally predicted interactions based on their scores

Assumptions:

Determine the probabil … ally-validated

Find the weights of each database for each score

1 splines



2 logistic regression

$$P(EV_j|S_{1j} \cap S_{2j} \cap \cdots \cap S_{nj})$$

$$\sum_{i=1} \qquad \sum_{ik}$$

$$x_{ij} = log\left(\frac{P(EV_j|S_{ij})}{P(EV_j)}\right)$$

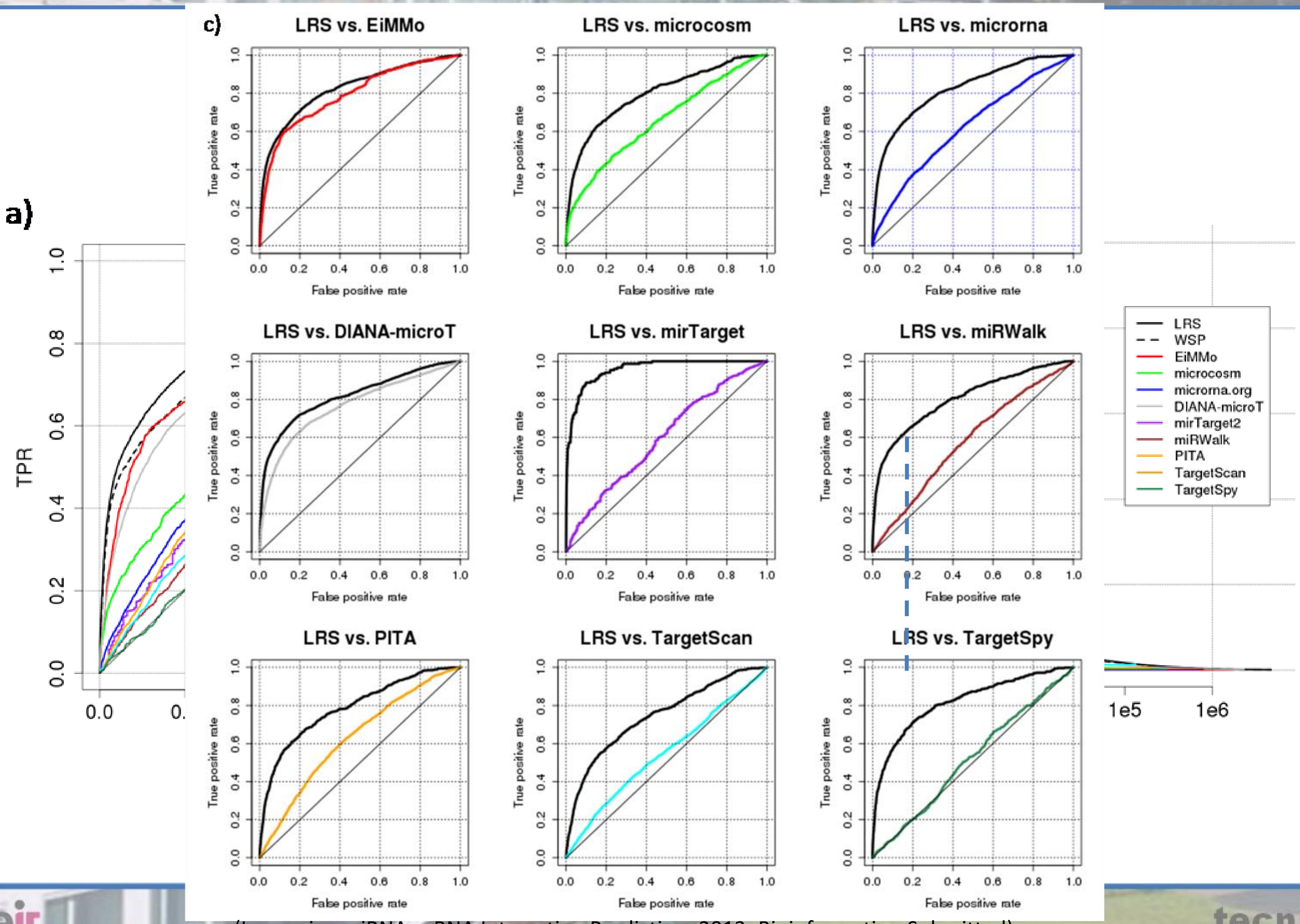
3 Combined score

$$\sum_{i=1}^{n} \widehat{\beta_i} \cdot x_{ij} + \sum_{ik}^{\binom{n}{2}} \widehat{\beta_{ik}} \cdot b_{ijk}$$

combined DB

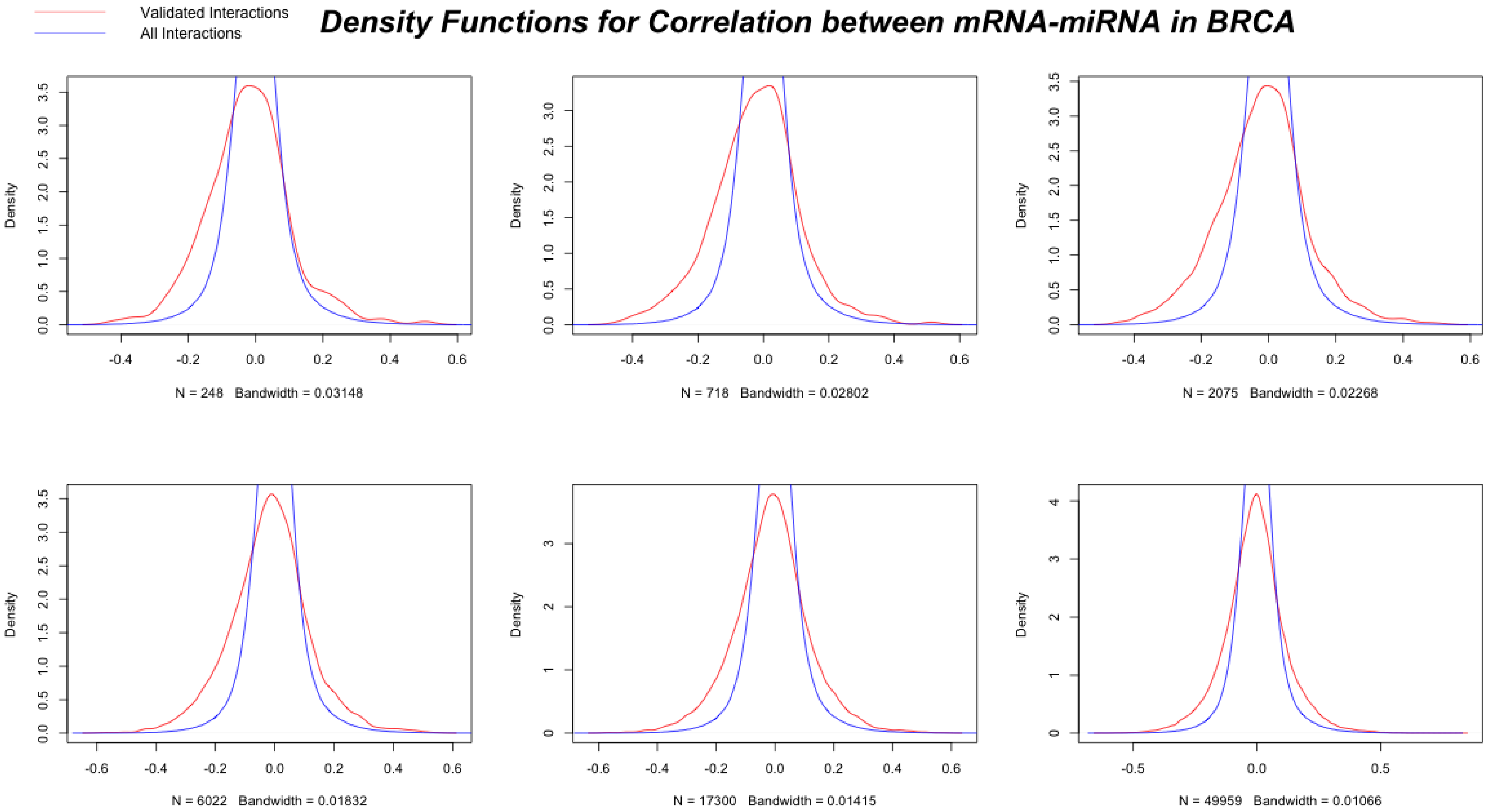
(Improving miRNA-mRNA Interaction Prediction; 2013; Bioinformatics; Submitted)

Results: ROC curves

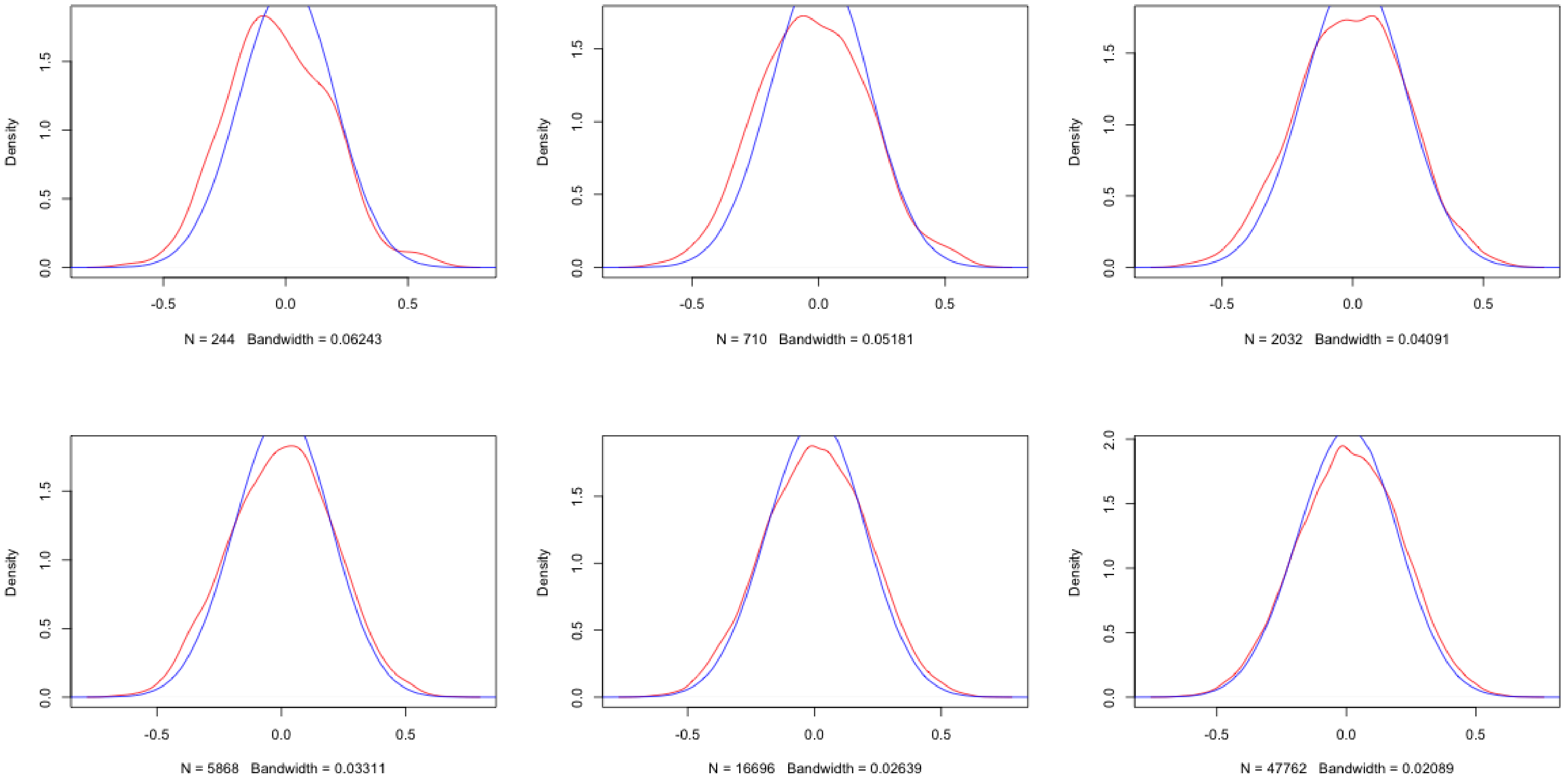


(Improving miRNA-mRNA Interaction Prediction; 2013; Bioinformatics; Submitted)

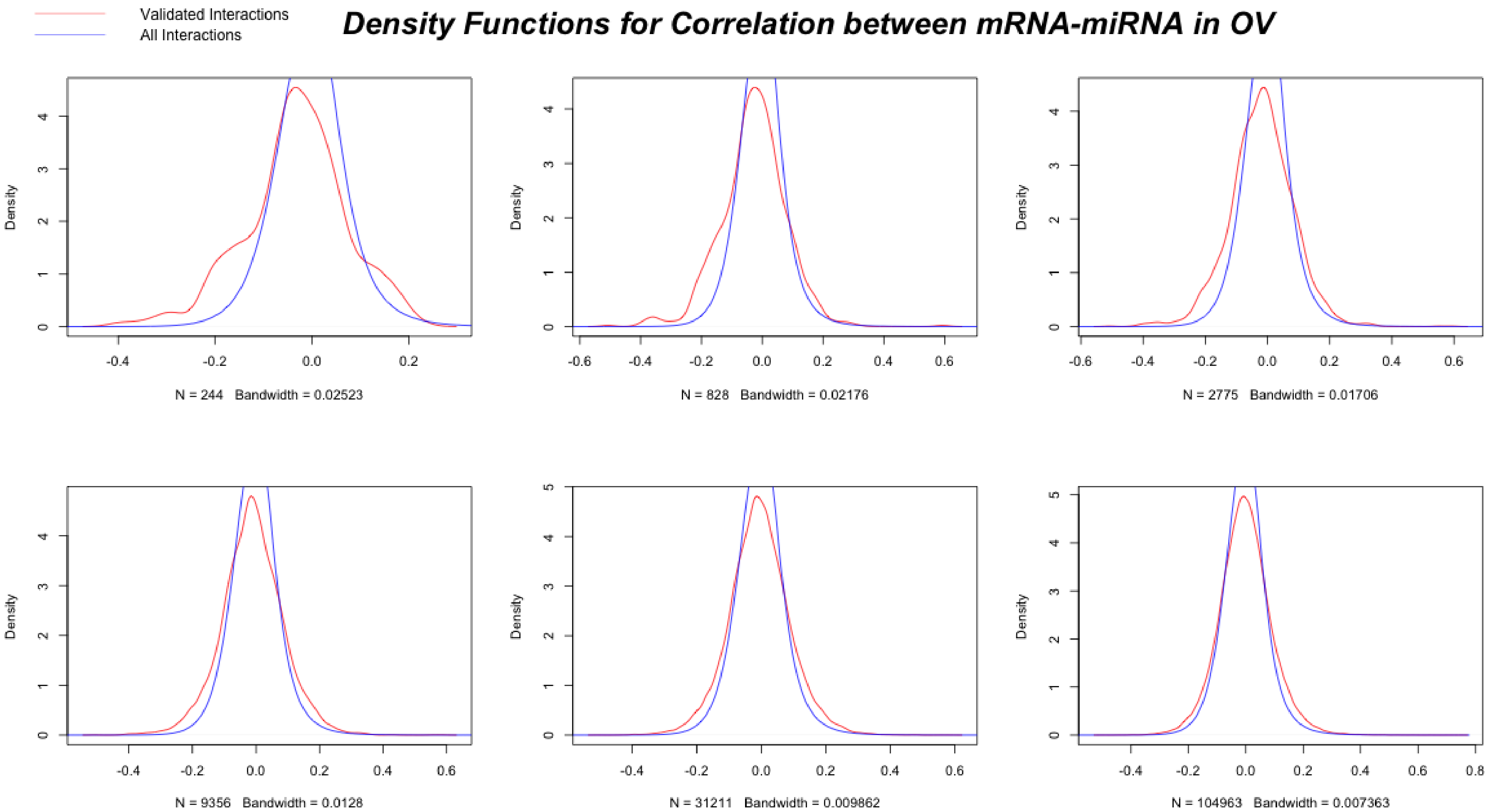
Correlation using these scores in TCGA



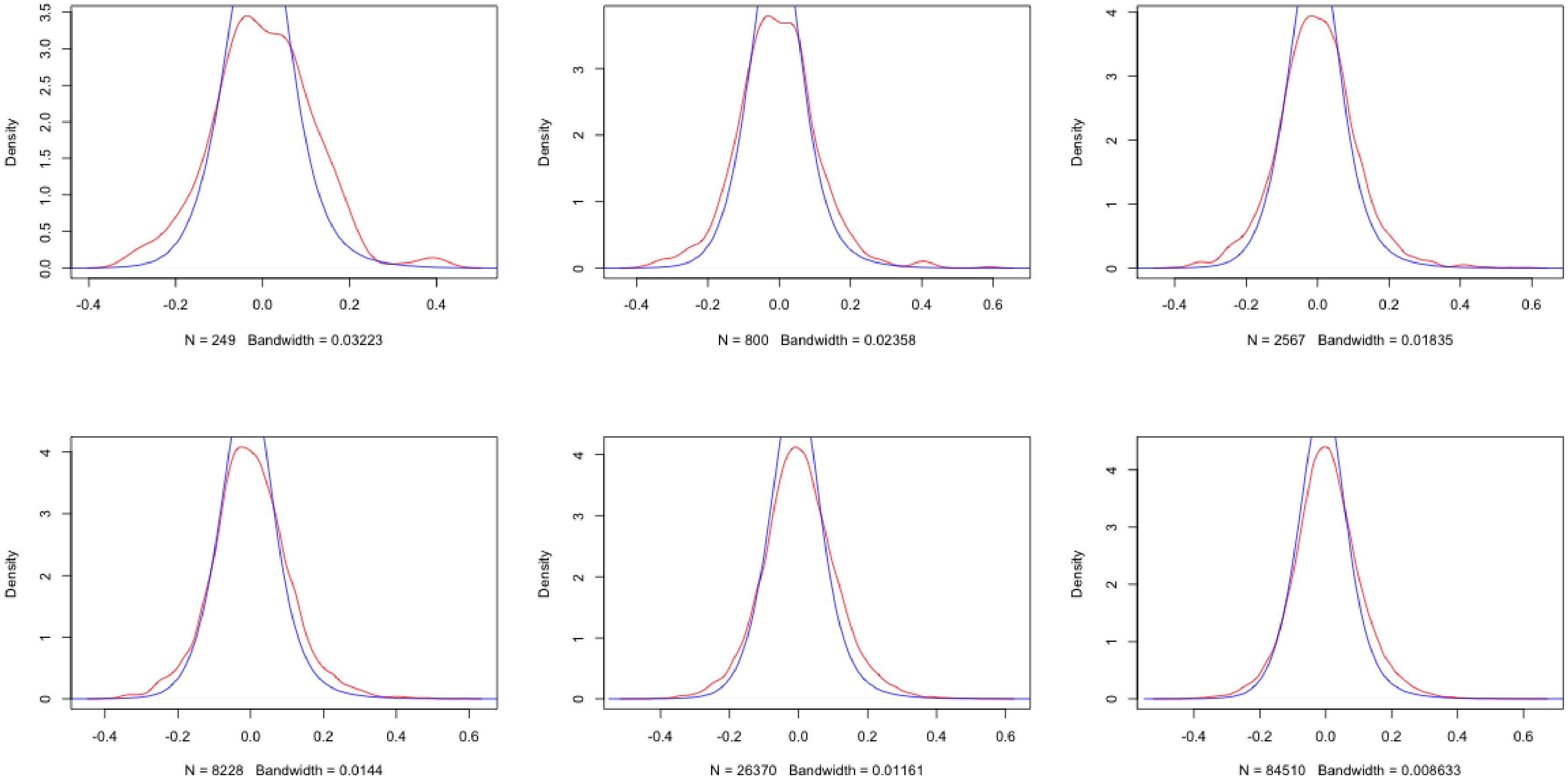
Density Functions for Correlation between mRNA-miRNA in LUAD



Density Functions for Correlation between mRNA-miRNA in OV

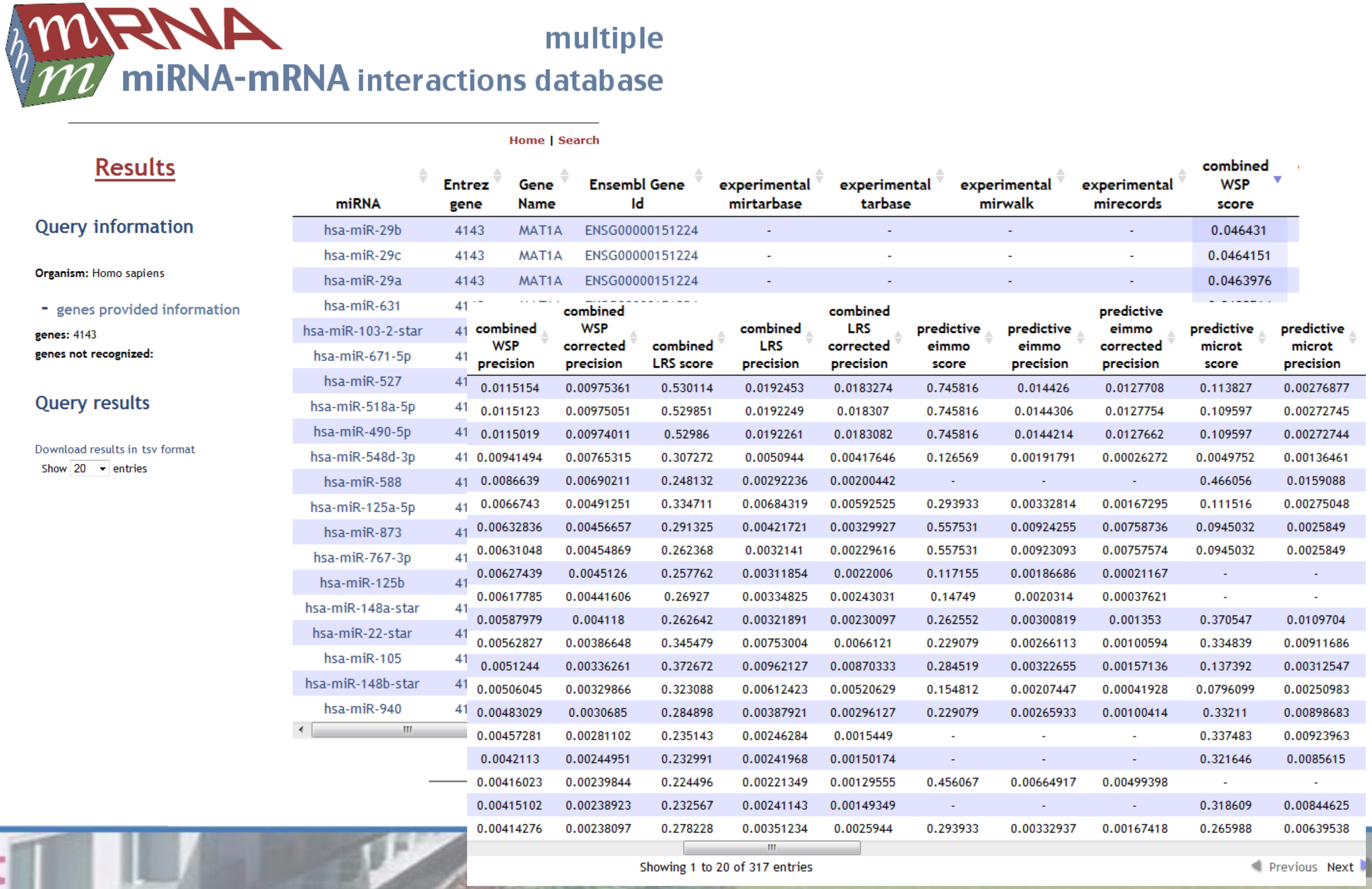


Density Functions for Correlation between mRNA-miRNA in GBM



Webpage for Join Database (m3RNA)

- http://m3rna.cnb.csic.es/



multiple miRNA-mRNA interactions database

Home | Search

Results

Query information

Organism: Homo sapiens

- genes provided information

genes: 4143

genes not recognized:

Query results

Download results in tsv format

Show 20 entries

miRNA	Entrez gene	Gene Name	Ensembl Gene Id	experimental mirtarbase	experimental tarbase	experimental mirwalk	experimental mirecords	combined WSP score
hsa-miR-29b	4143	MAT1A	ENSG00000151224	-	-	-	-	0.046431
hsa-miR-29c	4143	MAT1A	ENSG00000151224	-	-	-	-	0.0464151
hsa-miR-29a	4143	MAT1A	ENSG00000151224	-	-	-	-	0.0463976
hsa-miR-631	41							
hsa-miR-103-2-star	41							
hsa-miR-671-5p	41							
hsa-miR-527	41							
hsa-miR-518a-5p	41							
hsa-miR-490-5p	41							
hsa-miR-548d-3p	41							
hsa-miR-588	41							
hsa-miR-125a-5p	41							
hsa-miR-873	41							
hsa-miR-767-3p	41							
hsa-miR-125b	41							
hsa-miR-148a-star	41							
hsa-miR-22-star	41							
hsa-miR-105	41							
hsa-miR-148b-star	41							
hsa-miR-940	41							

combined WSP precision	combined WSP corrected precision	combined LRS score	combined LRS precision	combined LRS corrected precision	predictive eimmo score	predictive eimmo precision	predictive eimmo corrected precision	predictive microt score	predictive microt precision
0.0115154	0.00975361	0.530114	0.0192453	0.0183274	0.745816	0.014426	0.0127708	0.113827	0.00276877
0.0115123	0.00975051	0.529851	0.0192249	0.018307	0.745816	0.0144306	0.0127754	0.109597	0.00272745
0.0115019	0.00974011	0.52986	0.0192261	0.0183082	0.745816	0.0144214	0.0127662	0.109597	0.00272744
0.00941494	0.00765315	0.307272	0.0050944	0.00417646	0.126569	0.00191791	0.00026272	0.0049752	0.00136461
0.0086639	0.00690211	0.248132	0.00292236	0.00200442	-	-	-	0.466056	0.0159088
0.0066743	0.00491251	0.334711	0.00684319	0.00592525	0.293933	0.00332814	0.00167295	0.111516	0.00275048
0.00632836	0.00456657	0.291325	0.00421721	0.00329927	0.557531	0.00924255	0.00758736	0.0945032	0.0025849
0.00631048	0.00454869	0.262368	0.0032141	0.00229616	0.557531	0.00923093	0.00757574	0.0945032	0.0025849
0.00627439	0.0045126	0.257762	0.00311854	0.0022006	0.117155	0.00186686	0.00021167	-	-
0.00617785	0.00441606	0.26927	0.00334825	0.00243031	0.14749	0.0020314	0.00037621	-	-
0.00587979	0.004118	0.262642	0.00321891	0.00230097	0.262552	0.00300819	0.001353	0.370547	0.0109704
0.00562827	0.00386648	0.345479	0.00753004	0.0066121	0.229079	0.00266113	0.00100594	0.334839	0.00911686
0.0051244	0.00336261	0.372672	0.00962127	0.00870333	0.284519	0.00322655	0.00157136	0.137392	0.00312547
0.00506045	0.00329866	0.323088	0.00612423	0.00520629	0.154812	0.00207447	0.00041928	0.0796099	0.00250983
0.00483029	0.0030685	0.284898	0.00387921	0.00296127	0.229079	0.00265933	0.00100414	0.33211	0.00898683
0.00457281	0.00281102	0.235143	0.00246284	0.0015449	-	-	-	0.337483	0.00923963
0.0042113	0.00244951	0.232991	0.00241968	0.00150174	-	-	-	0.321646	0.0085615
0.00416023	0.00239844	0.224496	0.00221349	0.00129555	0.456067	0.00664917	0.00499398	-	-
0.00415102	0.00238923	0.232567	0.00241143	0.00149349	-	-	-	0.318609	0.00844625
0.00414276	0.00238097	0.278228	0.00351234	0.0025944	0.293933	0.00332937	0.00167418	0.265988	0.00639538

Showing 1 to 20 of 317 entries

Conclusions and Future work

Conclusions

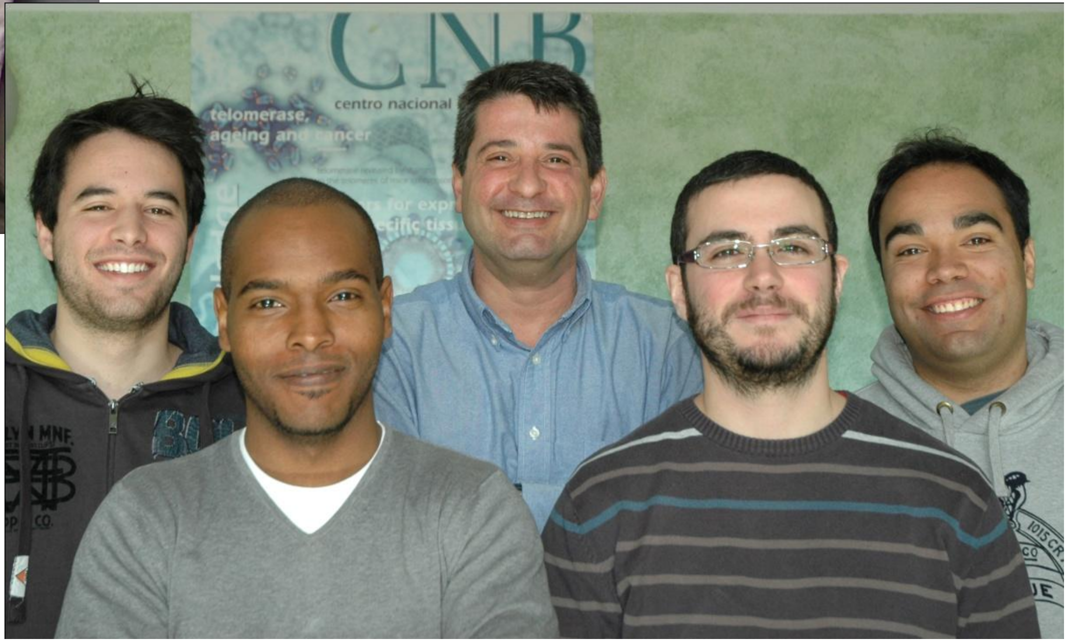
- **TaLasso** is a good alternative to find the outstanding miRNA – mRNA interactions using expression data and an initial set of putative interactions.
- **Normalization plays a major role**: the only difference between correlation and GenMir++ is whether the miRNAs are normalized or not.
- Focusing on **downregulation provides better results BUT**…
- … **positive regulation seem to exist** and also provides sound biological results.
- A **proper combination of the scores** of the databases provides a meta database with better features than any of its constituents

Future work

- Integration of the scores of the meta-base in the prediction methods
  - Include a weight in the Lasso regression that is inversely proportional to the probability of being validated.
- Migrate the implementation from `RCplex` to `glmnet`
  - `Rcplex` installation is cumbersome.
- The webpage will include several organisms
  - Now it only includes human.

Acknowledgements

CEIT (Centro de Estudios e Investigaciones Técnicas de Guipúzcoa)



CNB (Centro Nacional de Biotecnología)

Thanks for your attention. Questions?

21/03/2013