A bioinformatics approach to predict the Influence of multiple conjoint mirnAs on caNcer diseAse: the DIANA project

Antinisca Di Marco
Department of Biotechnological and Applied Clinical Sciences
University of L'Aquila
antinisca.dimarco@univaq.it

Microsoft

DISIM

nam et ipsa scientia potestas est DISCAB

Outline

- The context: (Hepatocellular) carcinoma desease and MicroRNA
- MicroRNA-target predicton: State of the art
- Goal: the prediction of the influence of multiple conjoint mirnas on cancer disease
- DIANA Project
- Conclusions and Future Work

The context

- Cancer diseases are very diffuse and affect even in serious mode many people.
- In particular, Hepatocellular carcinoma (HCC) is a highly aggressive epithelial tumor originating both from mature hepatocytes and stem cells.
- It is characterized by poor prognosis and very high rate of recurrence. Epidemiological studies indicate that HCC is the fifth most common cancer and the third most common cause of cancer-related death worldwide.
- The use of animal models helped to better understand the different phases of the entire cancerous process.

The context

- Animals' treatment with diethylnitrosamine (DEN) is one of the most frequently used approaches.
- DEN is a well-known hepatic carcinogen. *At the cellular level, particularly inside the hepatocyte, it produces lesions and DNA mutations.*
- A study demonstrated that *DEN administration* for several weeks *induces a rapid cancer development* and promotes HCC formation in 100 % of male and 10–30 % of female mice.
- Literature reports indicate that *tumor* molecular profile *of mice exposed to DEN are comparable to those related to human HCC* cases characterized by a poor prognosis.

The context

- MicroRNAs (miRNAs) are a class of small, non-coding RNAs.
- They can negatively regulate the expression of their target genes in a post-transcriptional manner, inducing mRNA degradation or inhibiting mRNA translation.
- miRNAs have the ability to regulate almost every aspect of cellular functionality, such as differentiation, development, apoptosis and proliferation.
- MiRNA deregulated activity has been described in various pathologies including cancer. For this reason, miRNAs functions started to be investigated with the help of bioinformatics approaches that allows to predict interaction with potential target genes [mirbase.org, microrna.org, genemania.org].

State of the art

- These tools are able to analyse a particular sequence (located on the 5’ end) of miRNA, called seed region, in order to predict the most probable genes interacting with it.
- Apart from complementarity, these tools take into account other important characteristics such as site accessibility, sequence conservation, multiple binding sites.
- In literature, there are some reports describing the exploitation of these algorithms to make prediction about miRNAs-target gene interactions for different common cancers.

State of the art

- However:
  - the majority of these studies halted to the miRNA profiling.
  - The online tools predict miRNAs-target gene interactions starting from one miRNA
  - in many conducted experimental studies, in presence of cancer, several miRNAs are deregulated and their conjoint influence is still little investigated.

Research Goal

We want to obtain a list of potential genes, and relative functions, all together related to a (small) group of significantly altered miRNAs in (HCC) cancer.

We want to predict the influence of multiple conjoint mirnas in presence of cancer disease

We want to implement a user-friendly tool that allow biologist to make this analysis.

A first attempt

We conducted a study [1] to specify the analysis method.

In this study, we combined miRNA expression analysis, obtained by an in vivo HCC mouse model, with a bioinformatics-based workflow. New genes, pathways and protein interactions, putatively involved in HCC initiation and progression, were identified and explored.

[1] F. Del Vecchio, F. Gallo, A. Di Marco, V. Mastroiaco, P. Caianiello, F. Zazzeroni, E. Alesse, A. Tessitore: Bioinformatics approach to predict target genes for dysregulated microRNAs in hepatocellular carcinoma: study on a chemically-induced HCC mouse model. BMC Bioinformatics 16: 408 (2015)

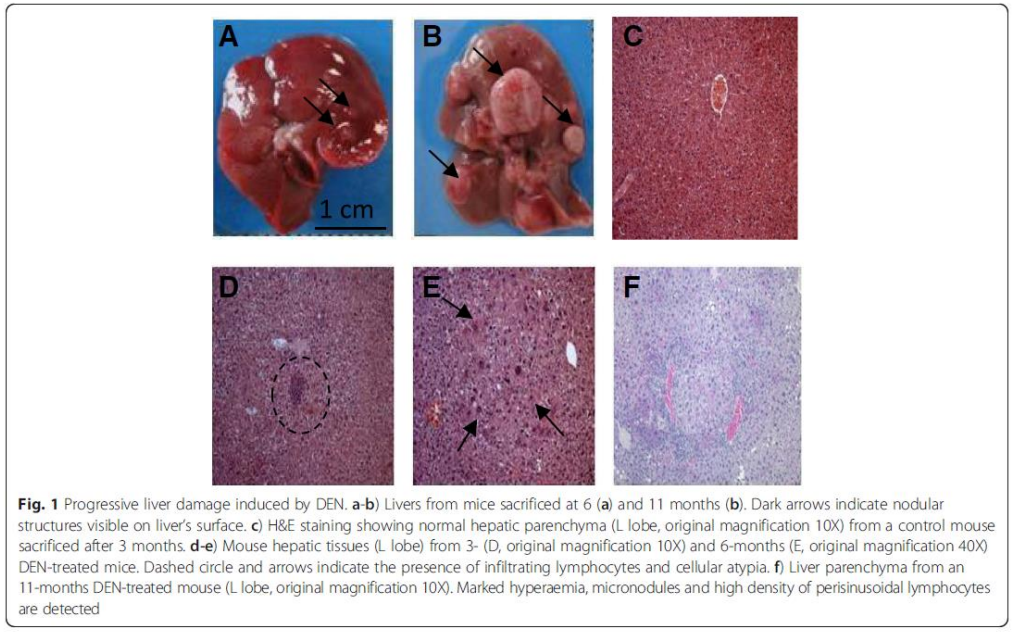


Fig. 1 Progressive liver damage induced by DEN. **a**-**b**) Livers from mice sacrificed at 6 (**a**) and 11 months (**b**). Dark arrows indicate nodular structures visible on liver's surface. **c**) H&E staining showing normal hepatic parenchyma (L lobe, original magnification 10X) from a control mouse sacrificed after 3 months. **d**-**e**) Mouse hepatic tissues (L lobe) from 3- (D, original magnification 10X) and 6-months (E, original magnification 40X) DEN-treated mice. Dashed circle and arrows indicate the presence of infiltrating lymphocytes and cellular atypia. **f**) Liver parenchyma from an 11-months DEN-treated mouse (L lobe, original magnification 10X). Marked hyperaemia, micronodules and high density of perisinusoidal lymphocytes are detected

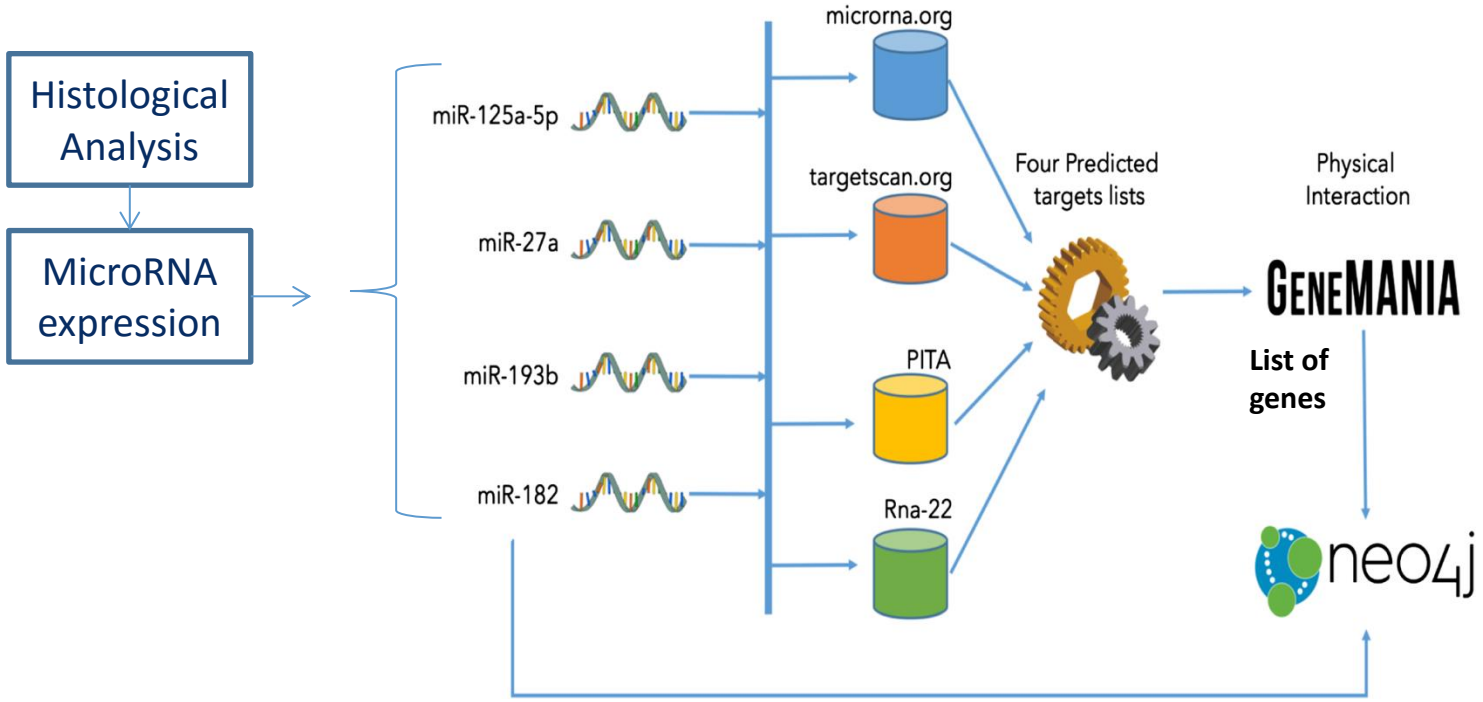
A first attempt

In particular, the analysis method is composed by the following phases:

- we predicted putative target genes using different bioinformatics mirna-gene prediction algorithms [pitar, rna22, targetscan, and miranda].
- We conducted enrichment annotation analysis to identify functional clusters which could be related to those target genes.
- We built up networks to visualize the possible circuits and pathways where the selected miRNAs could be involved, providing a resource for further functional studies on HCC pathogenesis.

[1] F. Del Vecchio, F. Gallo, A. Di Marco, V. Mastroiaco, P. Caianiello, F. Zazzeroni, E. Alesse, A. Tessitore: Bioinformatics approach to predict target genes for dysregulated microRNAs in hepatocellular carcinoma: study on a chemically-induced HCC mouse model. BMC Bioinformatics 16: 408 (2015)

Followed Workflow

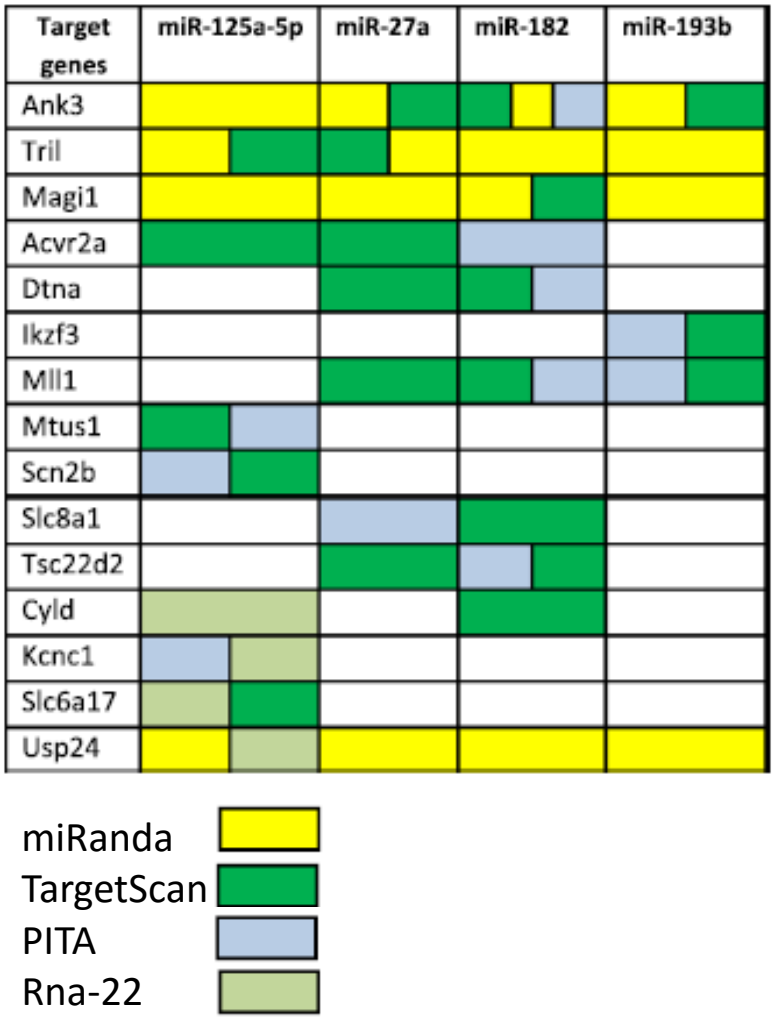


1. After MicroRNA expression, only 4 MicroRNA have been considered: MiR-125a-5p, miR-27a, miR-193b and miR-182.
2. We look for their common targets on four on-line DB (microrna, Targetscan, PITA and rna-22) in order to obtain four predicted targets lists, one for each DB.
3. From these lists, only 15 targets have been considered.
4. We gave the filtered list of targets as input to GENEMANIA in order to obtain a physical relation network.
5. Finally, we built up a network showing the relationships between miRNAs and targets, as well as those among targets by using Neo4j

Considered targets

The table illustrating the resulting 15 potential top targets for the selected microRNAs.
The list includes only genes predicted by at least 2 of 4 prediction tools.
Blank boxes represent too low (under the considered cut-off) or null association with microRNAs.

For the final analysis, we decided to consider only genes predicted by at least 2 of the above-mentioned 4 prediction programs. ANK3 mRNA was the unique target predicted by three different programs (MiRanda, TargetScan, PITA). In addition to ANK3, fourteen mRNAs were predicted by 2 different programs.



Validation

To validate the correctness of the described workflow, liver tissues from 11 months DENtreated and control mice as well as tumors from 11 months DEN-treated mice were analyzed for the expression of ANKG, which is the protein product of ANK3.

The results are in line with those evidenced in miRNA expression analysis, which, on the contrary, show a corresponding miRNAs' expression level increase in DEN tissues and tumors. The evidences obtained provide a validation of in silico data.

[1] F. Del Vecchio, F. Gallo, A. Di Marco, V. Mastroiaco, P. Caianiello, F. Zazzeroni, E. Alesse, A. Tessitore: Bioinformatics approach to predict target genes for dysregulated microRNAs in hepatocellular carcinoma: study on a chemically-induced HCC mouse model. BMC Bioinformatics 16: 408 (2015)

DIANA project

DIANA project deals with the development on the AZURE platform of the method described, and applied on mus musculus data, that permits to obtain a list of potential genes relating to a group of significantly altered miRNAs in cancer diseases for different species.

Moreover, the system will build up networks to visualize the possible circuits and pathways in which selected miRNAs could be involved, providing a potential resource for other researches focused on cancer disease.

DIANA project

Expected outcomes

i) a new graph DB, based on NEO4J technology collecting all the information related to MiRNAs, their target genes, and all the functional relations and annotations;

ii) a new functional prediction techniques to determine putative target genes influenced by a set of conjoint MicroRNA;

iii) a user-friendly and agile graphical interface easy to use for biotechnologies and biologists that helps them to query the DB in order to find new miRNAs-targets (multi-hop) relations and hence miRNA-pathways relations ;

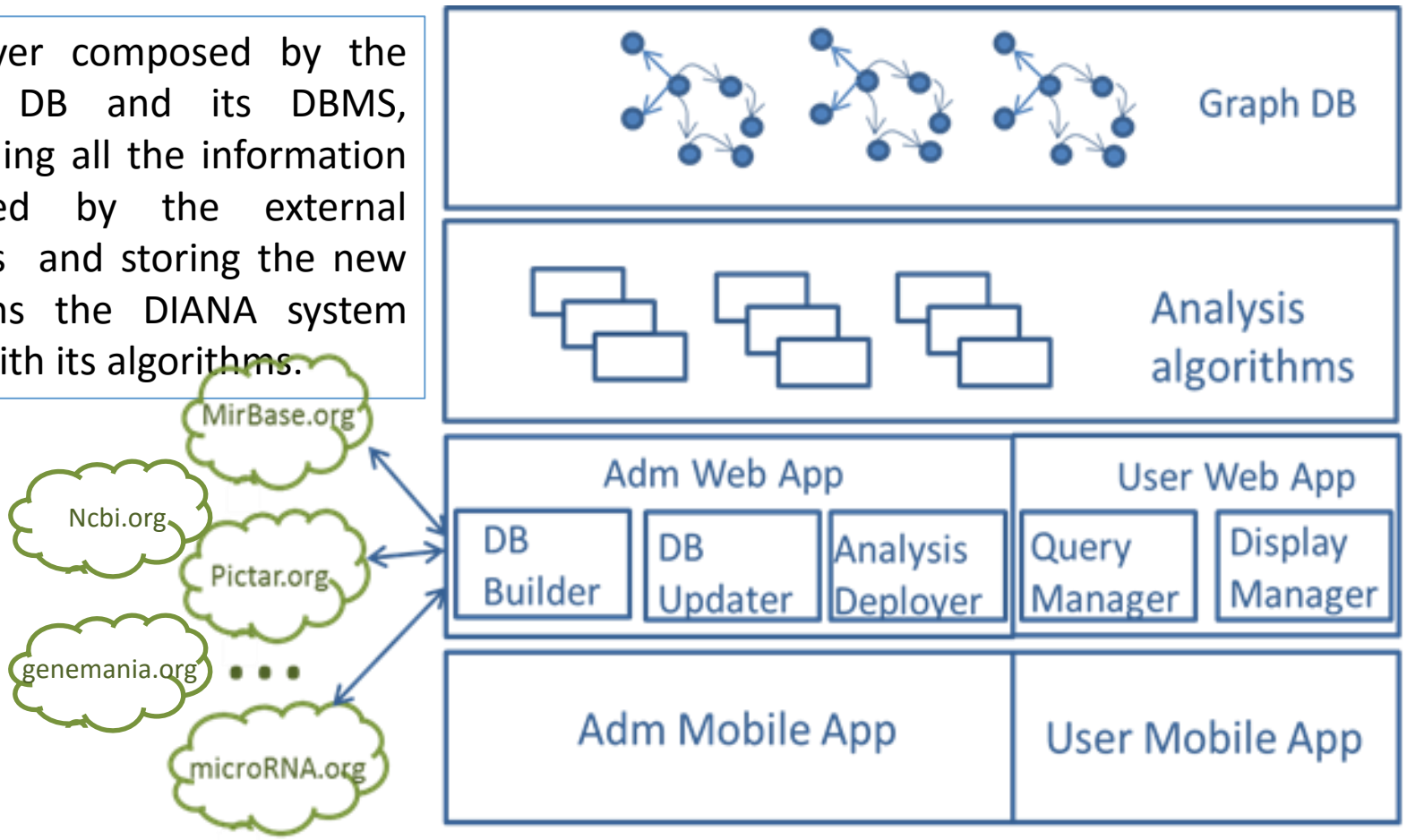
iv) possibly, identify new and unobserved miRNAs-targets (multi-hop) relations that can guide towards new directions in in-vivo experiments;

v) provide DIANA API to push towards cloud paradigm for future services built on top of DIANA.

DIANA project

The Diana System is composed by four software layers

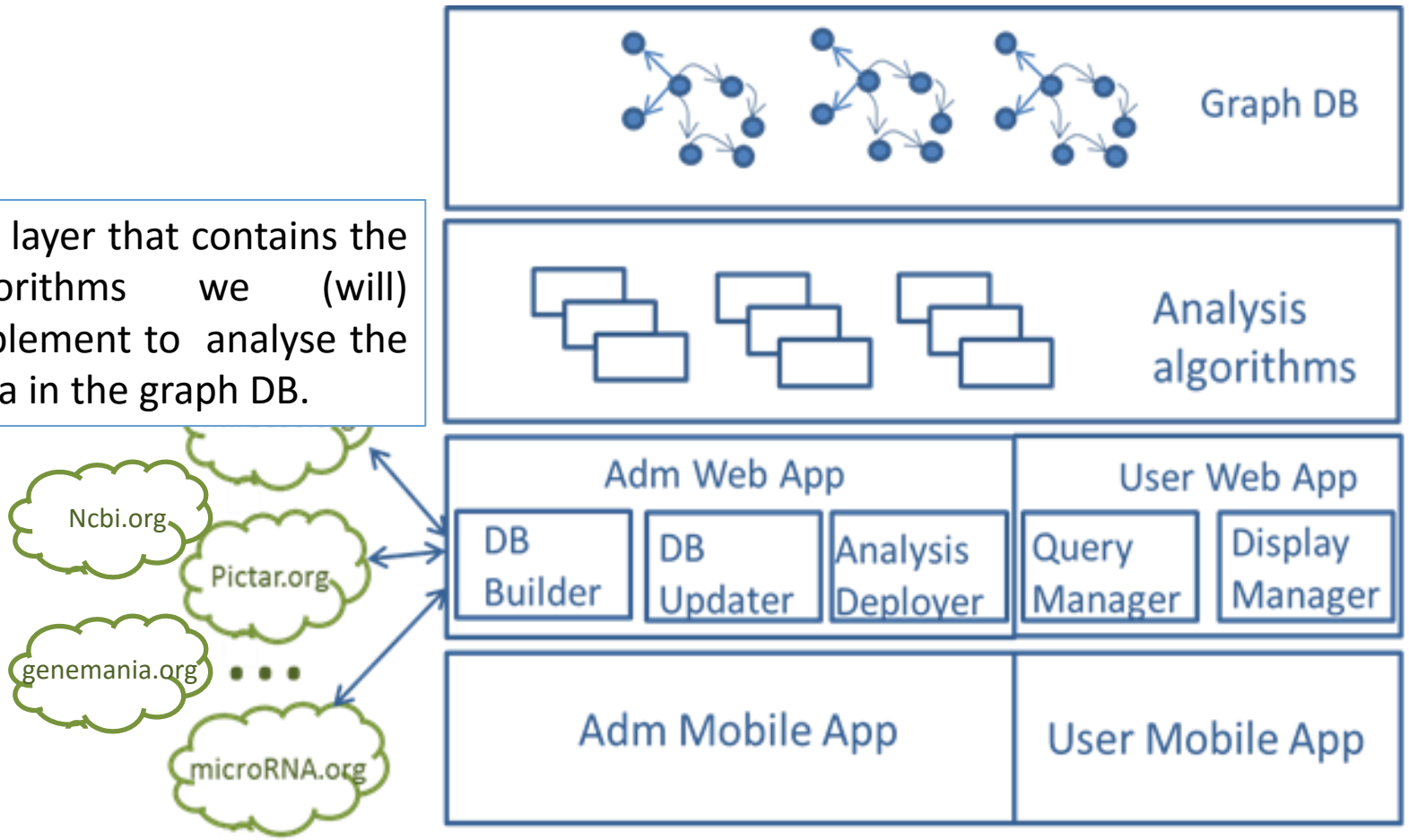
the layer composed by the Graph DB and its DBMS, containing all the information collected by the external sources and storing the new relations the DIANA system finds with its algorithms.



DIANA project

The Diana System is composed by four software layers

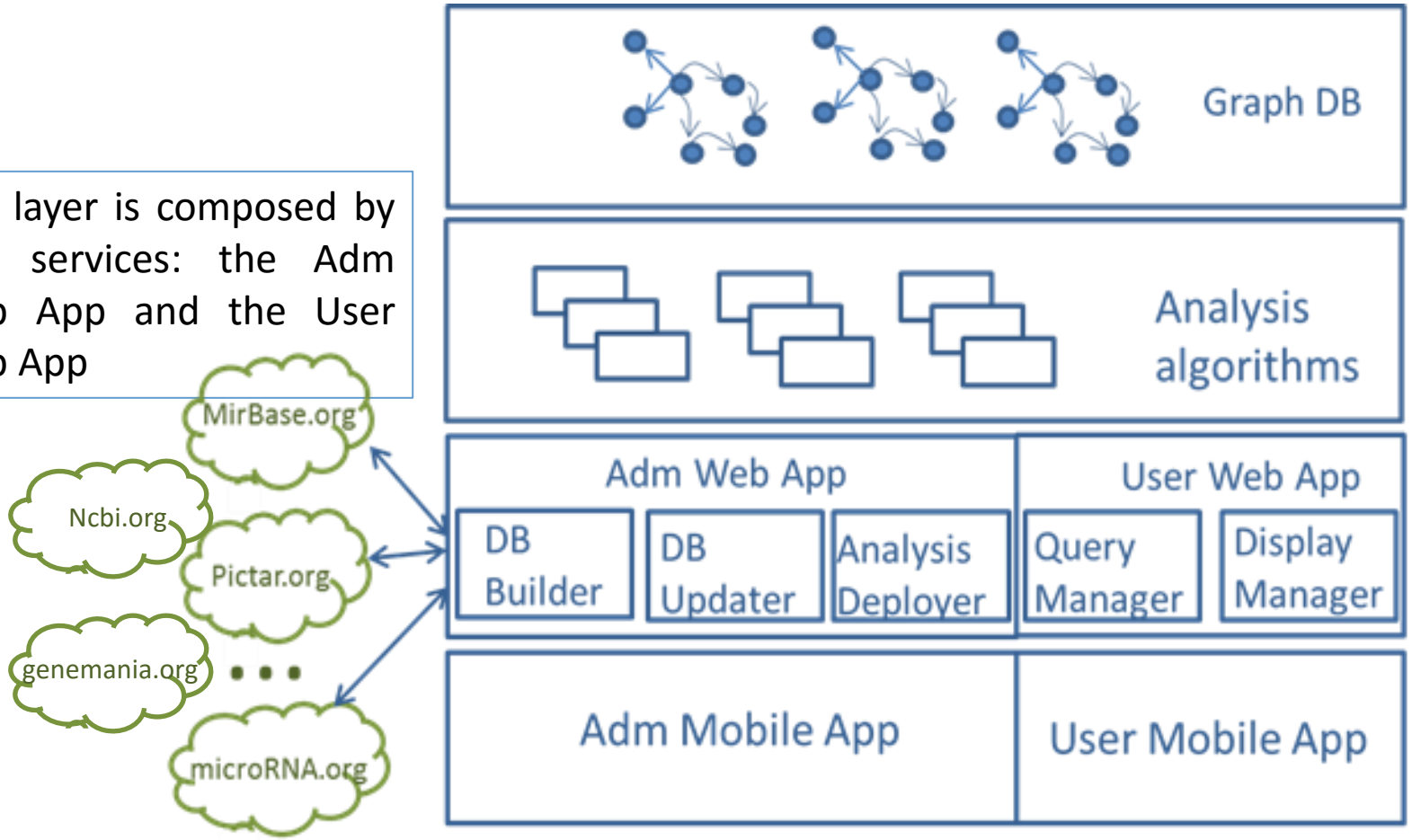
the layer that contains the algorithms we (will) implement to analyse the data in the graph DB.



DIANA project

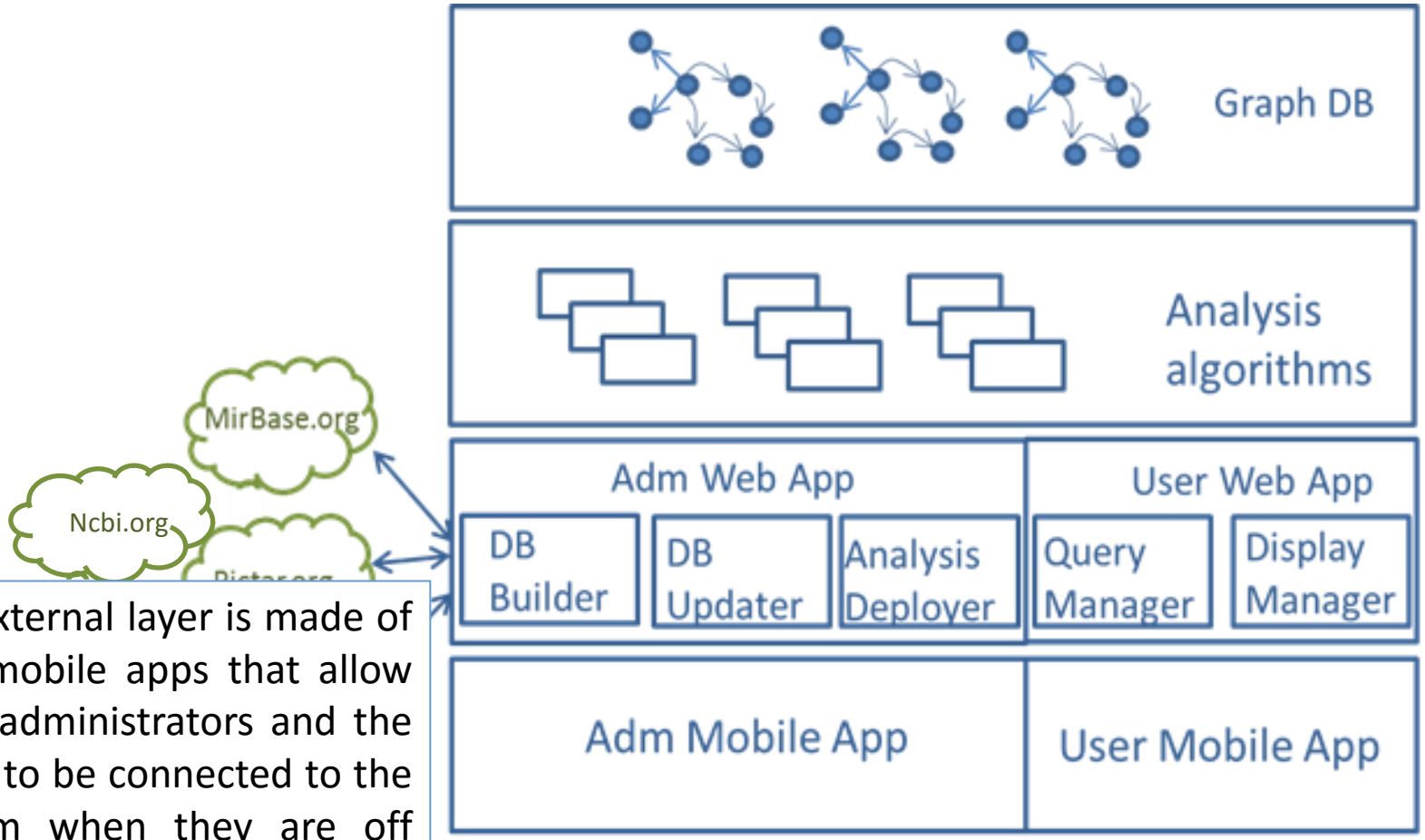
The Diana System is composed by four software layers

This layer is composed by two services: the Adm Web App and the User Web App



DIANA project

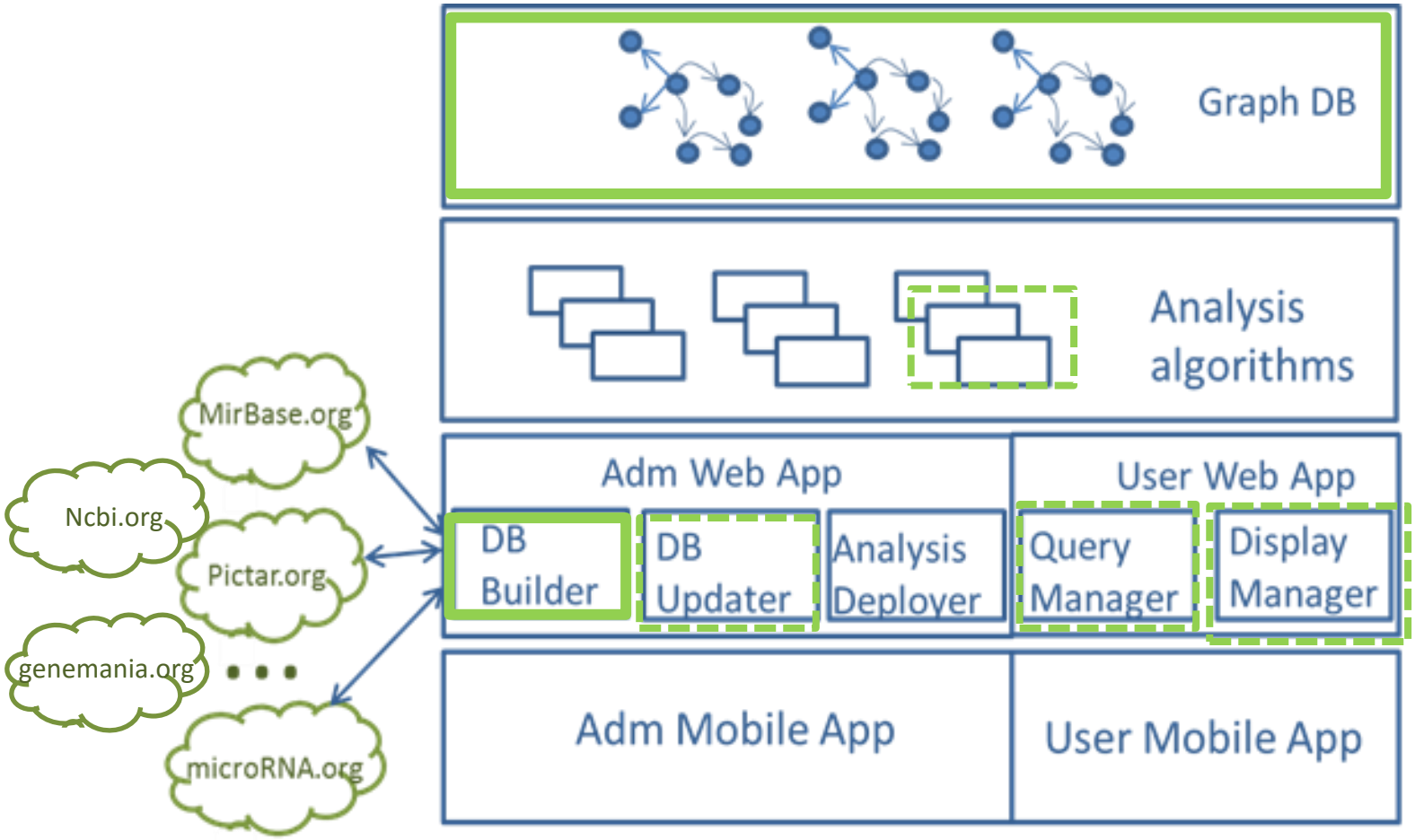
The Diana System is composed by four software layers



the external layer is made of two mobile apps that allow both administrators and the users to be connected to the system when they are off from the office.

DIANA project

Current Status



DIANA project

On-line DB integrated for

- Microrna information

Database	Descrizione	Download
miRBase	miRNA.dat (all published miRNA data in EMBL format)	ftp://mirbase.org/pub/mirbase/CURRENT/miRNA.dat.gz

- Target information

Database	Descrizione	Download
UniProt (Swiss-Prot)	UniProtKB - Reviewed (Swiss-Prot) - text format	ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz

DIANA project

On-line DB integrated for

- Microrna-Microrna relation

Database	Descrizione	Download
PicTar	picTarMiRNADog_mm7.bed e picTarMiRNAChicken_mm7.bed - BED format	https://genome.ucsc.edu/cgi-bin/hgTables clade: Mammal genome: Mouse assembly: Aug. 2005 (NCBI35/mm7) group: Expression and Regulation track: PicTar miRNA table: PicTar 7 Species (picTarMiRNADog) and PicTar 13 Species (picTarMiRNAChicken)

Database	Descrizione	Download
RNA22	MusMusculus,mRNA,ENSEMBL 65,miRbase18,RNA22v2	https://cm.jefferson.edu/data-tools-downloads/rna22-full-sets-of-predictions/

DIANA project

On-line DB integrated for

- Microrna-Microrna relation

Database	Descrizione	Download
TargetScan	Predicted (conserved) targets of conserved miRNA families. Includes positions on UTRs (without gaps) and UTR multiple sequence alignments (MSA; with gaps)	http://www.targetscan.org/mmu_71/mmu_71_data_download/Conserved_Family_Conserved_Targets_Info.txt.zip

Database	Descrizione	Download
miRTarBase	Mus musculus	http://mirtarbase.mbc.nctu.edu.tw/cache/download/6.1/mmu_MTI.xls

DIANA project

On-line DB integrated for

- Target-target relations

Database	Description	link
genemania	Mus Musculus	www.genemania.org

DIANA project

DB structure: two types of node

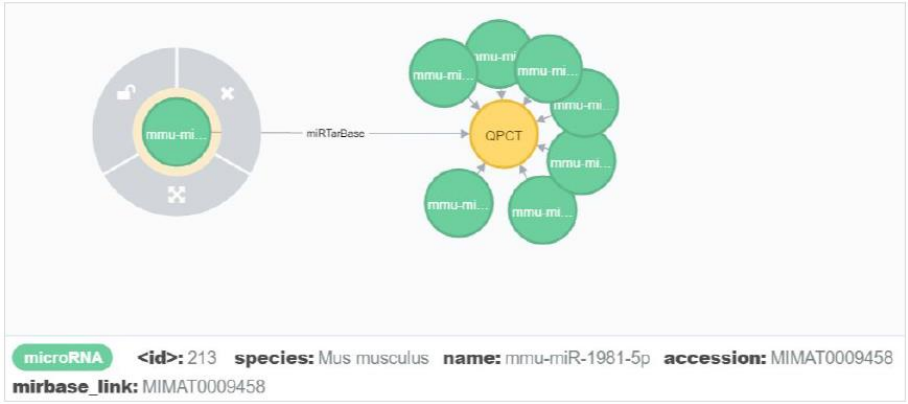
microRNA (miRBase)	**Target** (UniProt)
• Id • Name • Synonyms • Accession • Species • Mirbase_link	• Id • Name • Geneid • Ens_code • Species • Ncbi_link

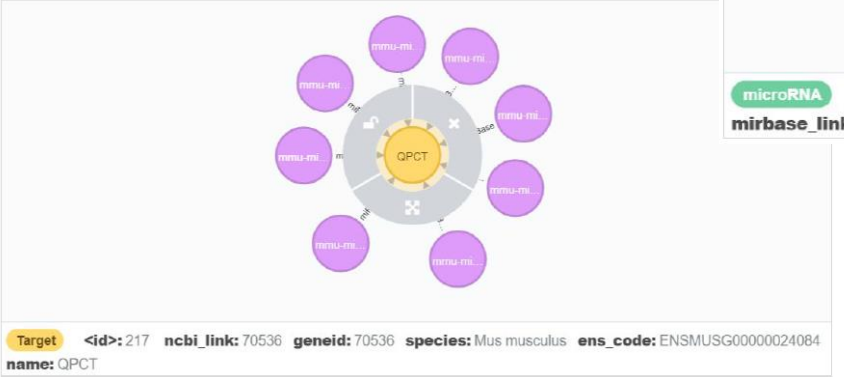
Several types of (labelled) node-to-node relations

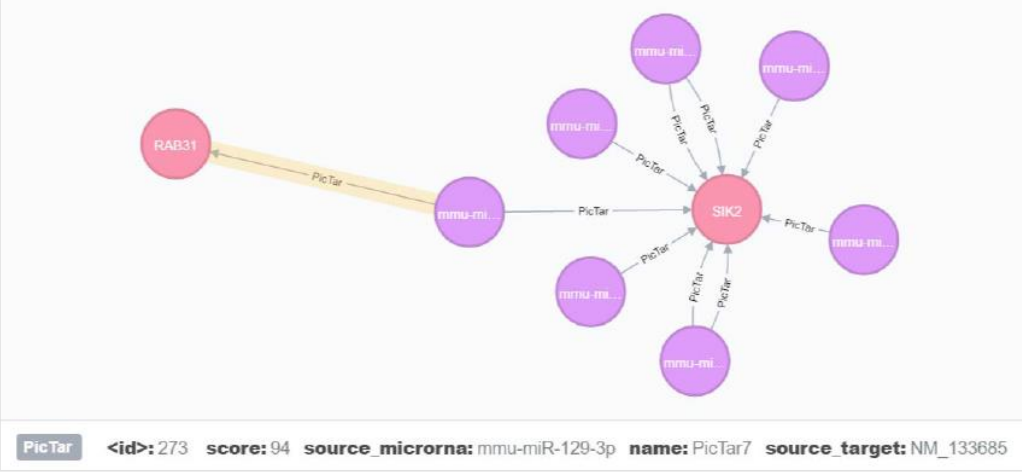
Relation
• ID • Name • Score • Source_MicroRNA • Source_target

DIANA project

NEO4J DB data







DIANA Team

- STAFF
  - Francesco Gallo, Post-Doc in Computer Science @ DISIM
  - Pasquale Caianiello, Assistant Professor in Computer Science @ DISIM
  - Alessandra Tessitore, Assistant Professor in Molecular Biology, DISCAB
  - Edoardo Alesse, Biologist, Full Professor in General Pathology @ DISCAB

- Students of the Bio-informatics Course:
  - Riccardo Rubei, Roberta Capuano, Michele Tucci, Federico Flaiano, Domenico Di Cesare, Andrea Bianchi, Alessandro Liberato, Gabriele Fargioni.

DIANA Project has been recently awarded by Microsoft that assigned its Azure Research Grant.

Conclusions

- There is evidence of the influence of MicroRNA and (HCC) cancer desease.
- In literature, there exist many algoritms that predict, in silico, gene targets that are negatively influenced by the deregulation of Mirna.
- DIANA is a project that aims to combine the existing predicition algorithms to determine the most probable gene targets.
- The resulting system is designed to be used from biologists in an easy way.
- We started from HCC cancer study but the resulting system is generic and can be used bradly in different domain where it is useful to predict gene target of deregulated microrna.