THE UNIVERSITY of EDINBURGH
informatics

Computational approaches to improving miRNA – mRNA interaction predictions

Maciej Pajak[1,*], *Clive R. Bramham*[2], *T. Ian Simpson*[1,3]

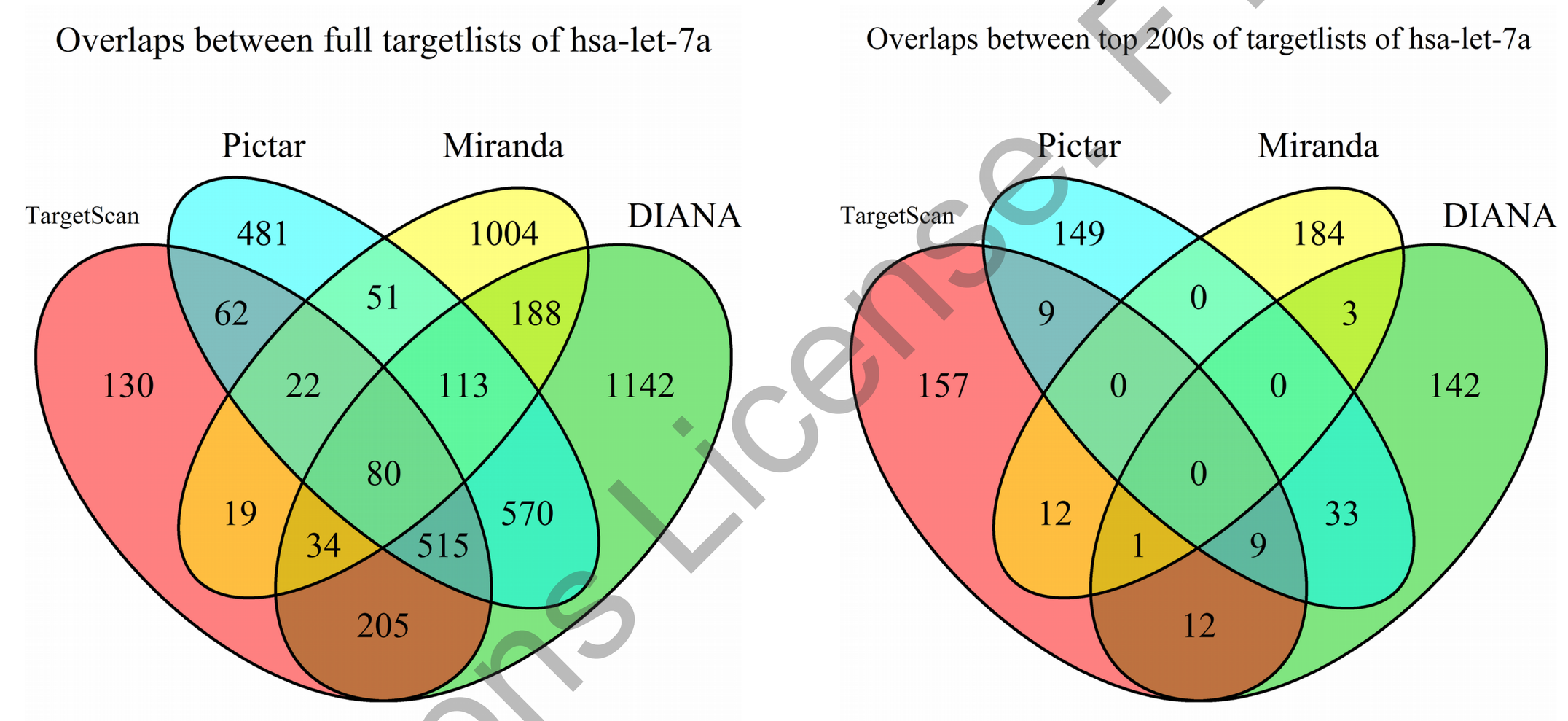
[1]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, [2]Department of Biomedicine and K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, [3]Biomathematics and Statistics Scotland (BioSS), *Correspondence: maciej.pajak@ed.ac.uk

Background

MicroRNAs (miRNAs) are 20-22nt long transcripts that form protein complexes which bind to mRNA and decrease the translation rate of the target mRNA. An ongoing problem in miRNA research is target prediction, as identifying targets is necessary for biological interpretation of the findings.

Target list aggregation

Four main target prediction sources at the moment include: *Pictar, Miranda, TargetScan, DIANA*. However, there is poor convergence between them even though they are all based on canonical binding – 2-7 nt 5' end of miRNA to 3' UTR of mRNA, as well as site conservation.



In order to improve predictions and reduce noise at the top of the list we propose a method of aggregating existing predictions, where the aggregated rank is calculated according to the equation:

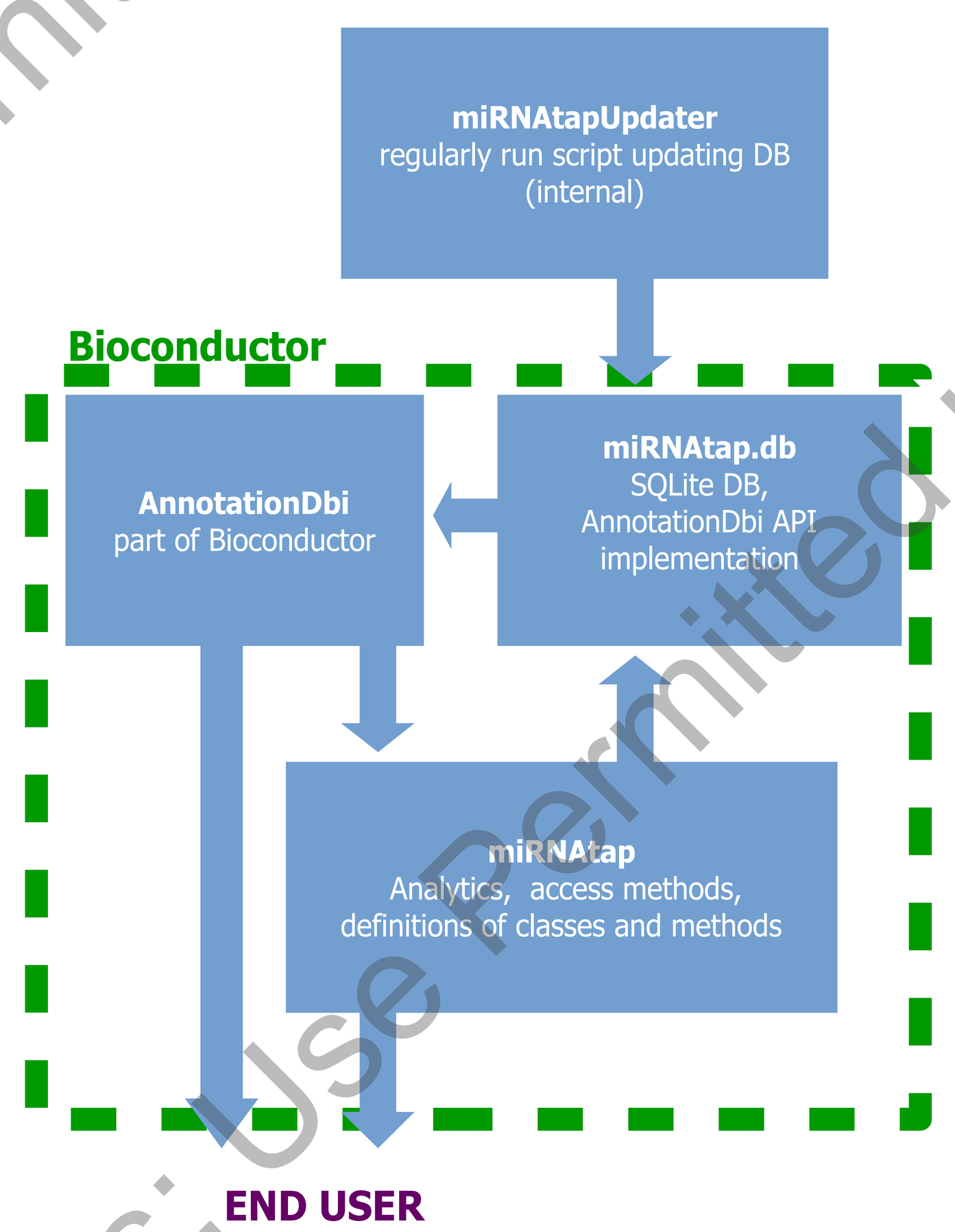
$$Aggregate=\frac{1}{M}\cdot geom(X)=\frac{1}{M}\sqrt[n]{\prod_{i\le n} X_i}$$

where M is the number of sources which returned the gene and X_i is its rank in i-th source. To appear in the aggregated list a target needs to appear in at least 2 or 3 source lists.

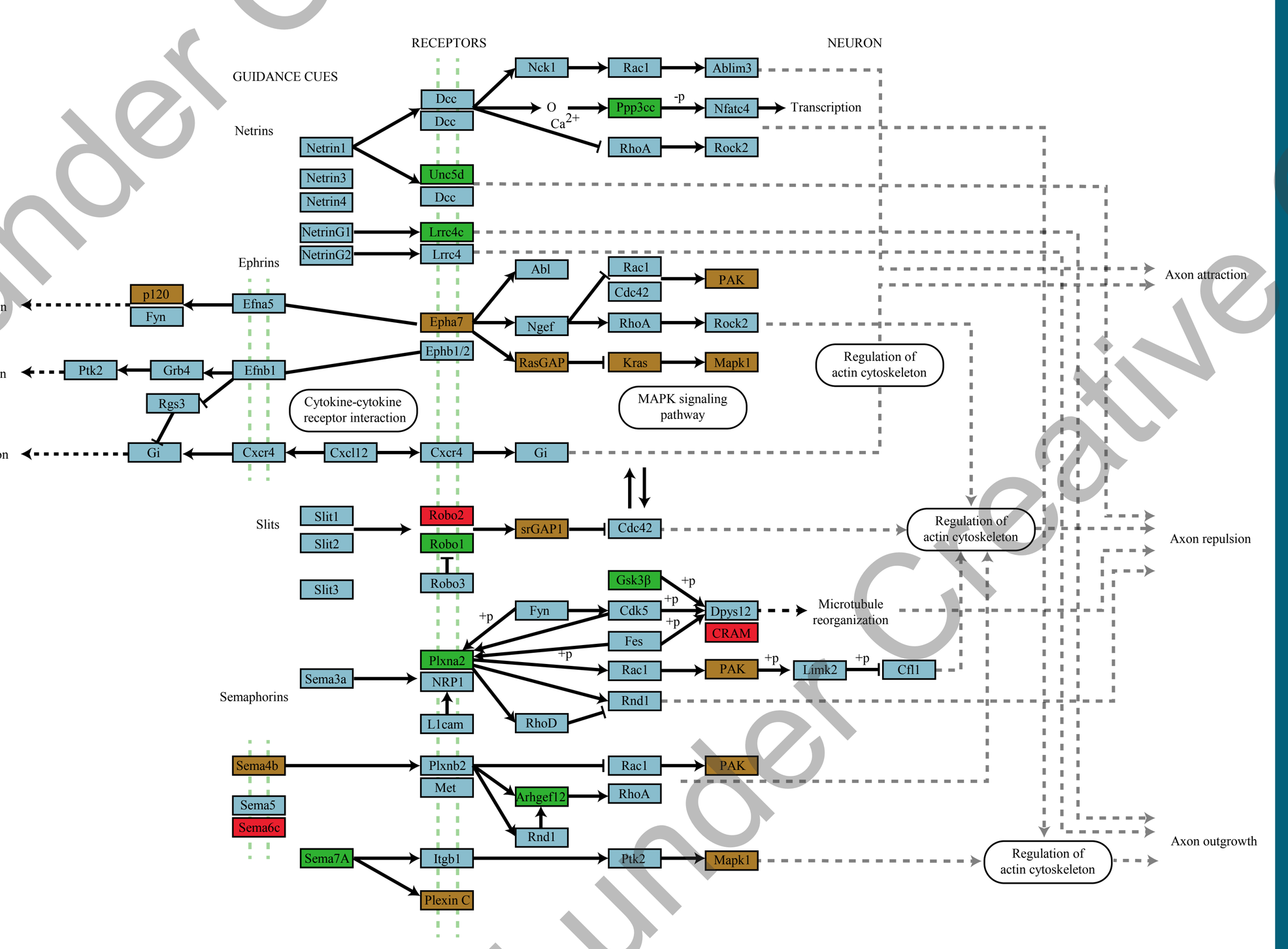
Bioconductor R package `miRNAtap`

miRNAtap provides the implementation of ranked target list aggregation method. It allows for programmatic access to aggregated miRNA target predictions. Direct predictions are available for *Homo sapiens* and *Mus musculus*, predictions for *Rattus norvegicus* (and other species in the future) are derived through homology transfer.

It is particularly useful for longer workflows where ranked target data can be used in further analysis, e.g. interaction networks, GO or pathways. The package is available on Bioconductor, the data is stored in an annotation package `miRNAtap.db` (also on Bioconductor).



High level view of the architecture

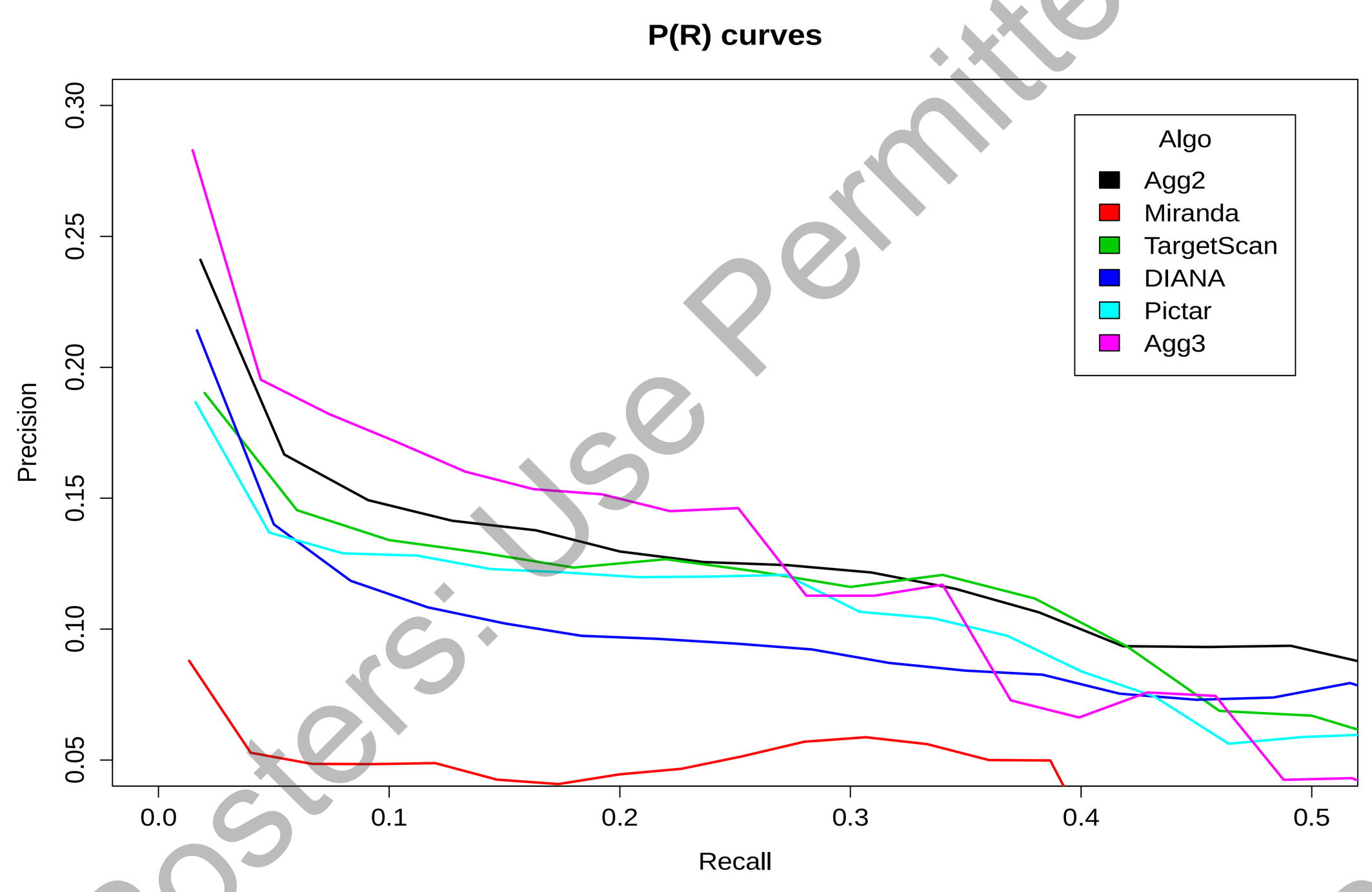


Example use in pathway analysis from Pai et al. (Front. Cell. Neurosci. 2014)

Evaluation

We evaluated our system based on Precision vs. Recall and we compared it to the four algorithms used as sources of predictions (*Pictar, Miranda, TargetScan, DIANA*). We used a sample of 160 human miRNAs which had at least 10 confirmed targets. Ground truth data come from *miRTarBase (2014)* and include experimentally confirmed interactions observed in various different data collection protocols, we did not differentiate between data collection protocols.

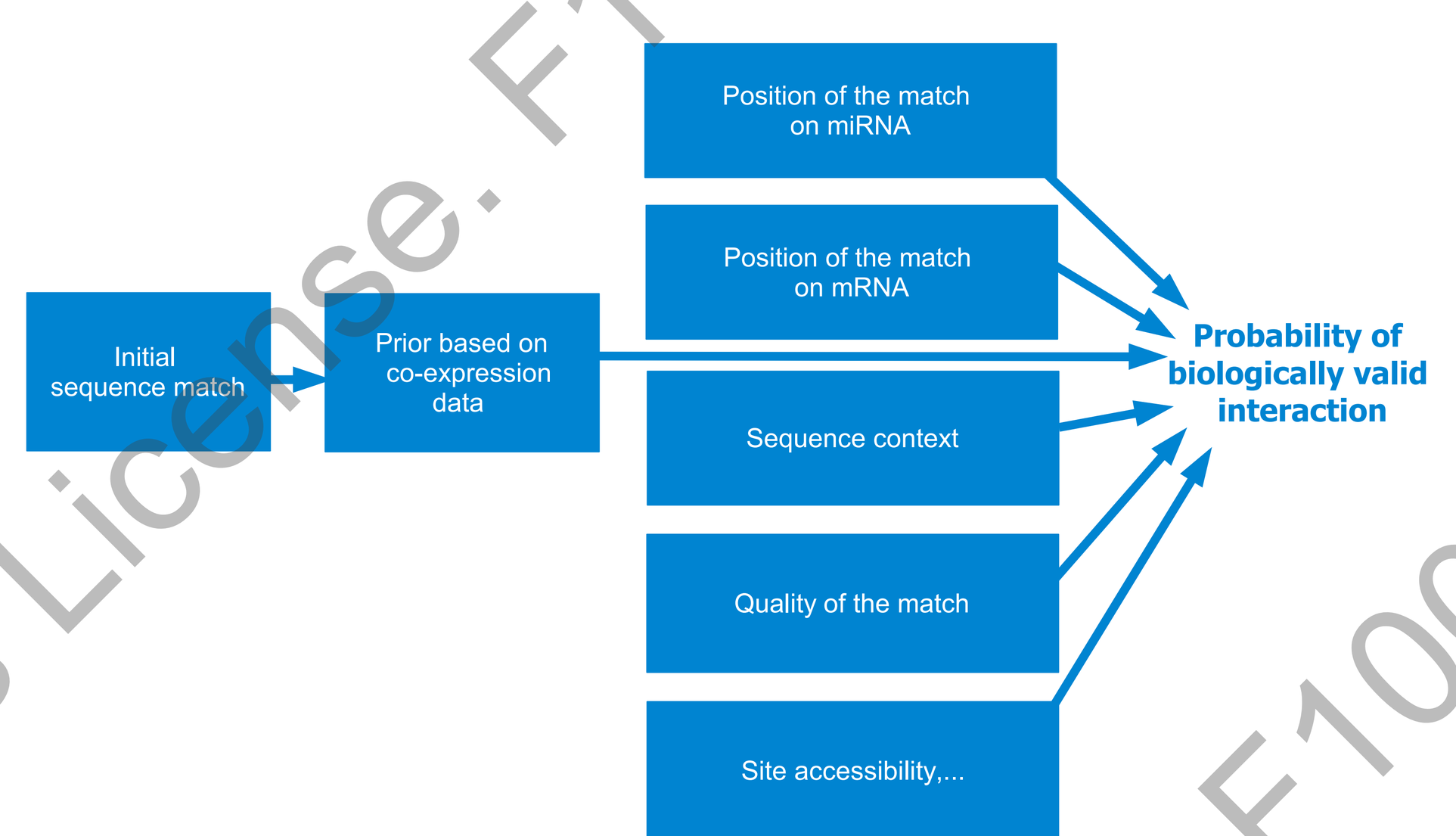
Overall, we observed improvement over single source predictions, *miRNAtap* aggregation with minimum 3 sources (*Agg3*) achieves higher precision at the top of the list than aggregation with minimum 2 sources (*Agg2*) but there is a trade off with recall. We recommend using different parameters depending on the use case.



Learning features from data

In our future work we propose an alternative approach to miRNA target prediction by training a classifier on experimental data.

First, prior on initial sequence matches is determined by co-expression data. Then features such as sequence signatures, quality of the match and binding site context (as used in *TargetScan*) are used to calculate the probability of the initial match being a meaningful interaction. Informative features are learned from the training set.



This method benefits from allowing more mismatches and non-canonical bindings than conventional algorithms, as suggested in a recent CLASH study by Helwak and colleagues (2013), it is also open to discovery of new sequence patterns.

Gold standard data which form the training set for our classifier are rapidly accumulating in multiple species and conditions thanks to high throughput techniques such as HITS-CLIP and CLASH, these data are likely to greatly improve the accuracy and reliability of miRNA target prediction in the near future.

Bioconductor package: Research group page:

References:

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. S. (2003). MicroRNA targets in Drosophila. Genome Biology, 5(1), R1.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Research, 19(1), 92–105.

Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell, 153(3), 654–65.

Hsu, S.-D., Tseng, Y.-T., Shrestha, S., Lin, Y.-L., Khaleel, A., Chou, C.-H., … Huang, H.-D. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Research, 42(Database issue), D78–85.

Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C. N., … Rajewsky, N. (2006). A genome-wide map of conserved microRNA targets in C. elegans. Current Biology : CB, 16(5), 460–71.

Maragkakis, M., Vergoulis, T., Alexiou, P., Reczko, M., Plomaritou, K., Gousis, M., … Hatzigeorgiou, A. G. (2011). DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. Nucleic Acids Research, 39(Web Server issue), W145–8.

Pai, B., Siripornmongcolchai, T., Berentsen, B., Pakzad, A., Vieuille, C., Pallesen, S., … Bramham, C. R. (2014). NMDA receptor-dependent regulation of miRNA expression and association with Argonaute during LTP in vivo. Frontiers in Cellular Neuroscience, 7(January), 1–15.

Acknowledgements:
BBSRC
EPSRC Engineering and Physical Sciences Research Council
MRC Medical Research Council