



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Υπολογιστικά Εργαλεία για την Αναγνώριση του
Λειτουργικού Ρόλου Πειραματικά Επιβεβαιωμένων
Γονιδίων – Στόχων των microRNAs**

Ιωάννης Σ Βλάχος

Επιβλέποντες: Άρτεμις Χατζηγεωργίου, Καθηγήτρια

ΑΘΗΝΑ

ΜΑΙΟΣ 2012

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υπολογιστικά Εργαλεία για την Αναγνώριση του Λειτουργικού Ρόλου Πειραματικά
Επιβεβαιωμένων
Γονιδίων – Στόχων των microRNAs

Ιωάννης Σ. Βλάχος

A.M.: 056

ΕΠΙΒΛΕΠΟΝΤΕΣ: Άρτεμις Χατζηγεωργίου, Καθηγήτρια

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Άρτεμις Γ Χατζηγεωργίου, Καθηγήτρια
Ηλίας Μανωλάκος, Αναπληρωτής Καθηγητής
Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας
(βαθμίδα Α')

Μάιος 2012

ΠΕΡΙΛΗΨΗ

Τα microRNAs είναι μικρά μη κωδικά μόρια ριβονουκλεϊκού οξέος (~23bp), τα οποία ρυθμίζουν ενεργά τη γονιδιακή έκφραση. Τα microRNAs έχουν θεωρηθεί ως κεντρικοί ρυθμιστές για πληθώρα βιολογικών διεργασιών, ενώ φαίνεται να διαδραματίζουν καίριο ρόλο και σε ένα πολύ μεγάλο αριθμό ανθρωπίνων παθολογιών.

Σκοπός της παρούσης μελέτης είναι να δημιουργήσει για πρώτη φορά τη σύνδεση μεταξύ πειραματικά επιβεβαιωμένων στόχων των microRNAs με μοριακά μονοπάτια, μέσω της δόμησης αλγορίθμων και εφαρμογών απόλυτα στοχευμένων στις ανάγκες των ερευνητών του συγκεκριμένου γνωστικού πεδίου. Για την επίτευξη αυτού του στόχου ήταν απαραίτητος ο σχεδιασμός και η υλοποίηση 2 σημαντικά εξελιγμένων εφαρμογών: μίας βάσης συγκέντρωσης πειραματικά επιβεβαιωμένων στόχων των microRNAs, με σημαντικά περισσότερους στόχους και δυνατότητες από οποιαδήποτε ήδη υπάρχουσα καθώς και μίας web εφαρμογής εξειδικευμένης στον εντοπισμό της επίδρασης των microRNAs στα μοριακά μονοπάτια. Η μεταξύ τους διασύνδεση ολοκληρώνει τη λειτουργικότητα των εφαρμογών και επιτρέπει την επίτευξη του τελικού στόχου της μελέτης. Με αυτόν τον τρόπο επεκτείνονται σημαντικά οι παρεχόμενες υπηρεσίες, με λειτουργικότητες πρώτη φορά διαθέσιμες στην ερευνητική κοινότητα.

Στην παρούσα μελέτη υλοποιήθηκε το TarBase v6.0, η μεγαλύτερη διαθέσιμη βάση πειραματικά επιβεβαιωμένων στόχων αυτή τη στιγμή στον κόσμο. Για τον εντοπισμό και την υποβοήθηση της εξαγωγής πληροφοριών από τη βιβλιογραφία δομήθηκε διαδικασία αυτόματης ανάλυσης των εγγράφων, που υποστήριξε την επισκόπηση της βιβλιογραφίας (text mining assisted manual curation). Κάνοντας χρήση αυτής της διαδικασίας συλλέχθηκαν 65.814 πειραματικά επιβεβαιωμένες αλληλεπιδράσεις.

Κατόπιν της δημιουργίας της βάσεως δεδομένων που ήταν ικανή για να υποστηρίξει μία τέτοια προσπάθεια, σχεδιάστηκε και υλοποιήθηκε μια σειρά αλγορίθμων και εργαλείων που ενσωματώθηκαν υπό τη μορφή της διαδικτυακής εφαρμογής "DIANA-miRPath v2.0". Το DIANA-miRPath v2.0 είναι μια εφαρμογή σχεδιασμένη για να παρέχει λειτουργίες και εργαλεία απόλυτα ειδικά στην ανάλυση και τον εντοπισμό μοριακών μονοπατιών ρυθμιζόμενων από microRNAs. Η δεύτερη έκδοση του miRPath υλοποιήθηκε από την αρχή και είναι η πρώτη που ενσωματώνει προβλεπόμενους και πειραματικά επιβεβαιωμένους στόχους, μέσω της διασύνδεσής της με το TarBase, ενώ παράλληλα προσφέρει πληθώρα καινοτόμες λειτουργικότητες, όπως η διενέργεια ταξινομήσεων των μονοπατιών ή/και των microRNAs με βάση τις μεταξύ τους αλληλεπιδράσεις και τη σχεδίαση heat maps των microRNAs έναντι των μονοπατιών.

Η ενοποίηση των δύο υλοποιήσεων μπορεί να εξυπηρετήσει πολλαπλά διαφορετικά σενάρια ανάλυσης μέσω στατιστικών μεθόδων υψηλής ακριβείας, ενώ προσφέρει ουσιαστικά εργαλεία που θα επιτρέψουν στους ερευνητές να εξετάσουν και να κατανοήσουν ευκολότερα και σε μεγαλύτερο βάθος δεδομένα σχετικά με τη φύση και τη λειτουργία των microRNAs.

Οι δύο υλοποιήσεις είναι προσβάσιμες από τις διευθύνσεις:

<http://www.microrna.gr/tarbase/> και <http://www.microrna.gr/miRPathv2/>

Και έχουν ανακοινωθεί μέσω 2 δημοσιεύσεων: T. Vergoulis, I. S. Vlachos *et al*, *Nucleic acids research*, vol. 40, pp. D222-9, Jan 2012. & Ioannis S Vlachos, Nikos Kostoulas *et al*, *Nucleic acids research*, Accepted for Publication.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Βιολογία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: microRNA, μοριακά μονοπάτια, γονίδια – στόχοι, text mining, heat maps.

ABSTRACT

microRNAs are small non coding RNA species (~23bp), which actively regulate gene expression. They are esteemed as central regulators for a large number of biological processes and human pathologies.

Aim of the present study is to actively link for the first time experimentally validated microRNA targets with molecular pathways. This connection will be established via the formation of highly specific algorithms and applications to microRNA centered research. An imperative prerequisite for achieving this goal is the design and implementation of 2 distinct applications: a data base collecting significantly more experimentally validated targets than any other available today; as well as a web application allowing the accurate determination of the functional role of microRNAs in molecular pathways. The interconnection between these two on-line applications will significantly extend their functionality and usefulness to the research community.

Initially, the largest available microRNA gene target dataset was semi-manually extracted from the available literature. A text mining assisted curation pipeline was implemented, in order to optimize the laborious manual process. The collected targets represent a 15.4 to 175 – fold increase, compared to all available manually curated databases available today. This wealth of information was inserted in TarBase v6.0, an online version of the database, in order to be freely accessible to the public.

Following the creation of a database capable of supporting the project's aim, a series of highly specific tools and algorithms were implemented, enabling the identification of microRNA regulated molecular pathways. All these modules were incorporated to DIANA-miRPath v2.0, the second version of the miRPath web server, which was redesigned from the ground up to meet the project's specifications. This application is the first available to include predicted, as well as experimentally validated microRNA targets. It was especially designed to offer tools and functionalities highly specific to the detection and analysis of microRNA – pathway interactions. DIANA-miRPath v2.0 offers a plethora of novel tools, such as clustering of microRNAs and pathways based on their interactions, design of microRNA vs pathways heat maps and more.

The integration of these two implementations can support a large number of different use case scenarios and research pipelines by sophisticated statistics and relevant algorithms, while offering for the first time essential tools to the research community. These tools can significantly assist researchers examine and understand data relevant to microRNA nature and function.

The applications are accessible from the following URLs:

<http://www.microrna.gr/tarbase/> and <http://www.microrna.gr/miRPathv2/>

The results have been published in two journal articles: T. Vergoulis, I. S. Vlachos *et al*, *Nucleic acids research*, vol. 40, pp. D222-9, Jan 2012. &

Ioannis S Vlachos, Nikos Kostoulas *et al*, *Nucleic acids research*, Accepted for Publication.

SUBJECT AREA: Computational Biology

KEYWORDS: microRNA, molecular pathway, gene targets, text mining, heat maps