



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

Πρόβλεψη στόχων μικρών RNAs σε μακρά μη κωδικά μετάγραφα

Μαρία Δ Παρασκευοπούλου

Επιβλέποντες: Άρτεμις Χατζηγεωργίου, Καθηγήτρια

ΑΘΗΝΑ

ΝΟΕΜΒΡΙΟΣ 2012

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Πρόβλεψη στόχων μικρών RNAs σε μακρά μη κωδικά μετάγραφα

Μαρία Δ Παρασκευοπούλου

A.M.: 061

ΕΠΙΒΛΕΠΟΝΤΕΣ: Άρτεμις Χατζηγεωργίου, Καθηγήτρια

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Άρτεμις Χατζηγεωργίου, Καθηγήτρια
Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας
(βαθμίδα Α΄)
Χρυσίνα Ευαγγελία, Επιστημονικό Προσωπικό του ΕΙΕ

Νοέμβριος 2012

ΠΕΡΙΛΗΨΗ

Τα microRNAs είναι μικρά μονόκλινα μόρια RNA μήκους περίπου 22 νουκλεοτιδίων και ανήκουν στην ευρύτερη κατηγορία των μη κωδικών μεταγράφων. Διαδραματίζουν καθοριστικό ρόλο σε πληθώρα από βιολογικές διαδικασίες, με κύρια τη ρύθμιση έκφρασης των περισσότερων γονιδίων. Τα lncRNAs, από την άλλη μεριά, είναι μακρά μη κωδικά μετάγραφα, μήκους άνω των 200 νουκλεοτιδίων, τα οποία εντοπίζονται σε γενωμικές περιοχές με ή χωρίς επικάλυψη με κωδικά γονίδια. Η αλληλεπίδραση των lncRNAs με τα microRNAs μπορεί να αποτελέσει ένα ενδιάμεσο επίπεδο ρύθμισης της γονιδιακής έκφρασης.

Σκοπός της παρούσας μελέτης είναι να διερευνήσει τη λειτουργική αλληλεπίδραση μεταξύ των microRNAs και των lncRNAs βασισμένη τόσο σε πειραματικά δεδομένα αλλά και σε υπολογιστικές μεθόδους αιχμής. Ακόμη, παρέχει για πρώτη φορά ένα ολοκληρωμένο κατάλογο από πιθανούς στόχους των microRNA (miRNA) σε περιοχές των lncRNAs, στους οργανισμούς του ανθρώπου και του ποντικού. Η αναγνώριση miRNA στόχων, μέσω της ανάλυσης πειραματικών δεδομένων, για μεγάλο αριθμό από lncRNA μετάγραφα, δεν έχει μελετηθεί στο παρελθόν και υπάρχουν λίγες σχετικές αναφορές στη βιβλιογραφία.

Το πρώτο σκέλος της ανάλυσης περιελάμβανε τη συλλογή όλων των μορίων lncRNAs από τις πιο πρόσφατα ενημερωμένες βιολογικές βάσεις δεδομένων καθώς και από σχετικές δημοσιεύσεις. Στη συνέχεια, προσδιορίστηκαν οι στόχοι των miRNAs σε lncRNA μέσω της ανάλυσης δεδομένων από γονοτύπηση RNA μεγάλης-κλίμακας, και συγκεκριμένα των πειραμάτων HITS-CLIP και PAR-CLIP. Παράλληλα, χρησιμοποιήθηκε ένας από τους καλύτερους αλγόριθμους, σε ευαισθησία και ειδικότητα, ο DIANA-microT-CDS, για την ανακάλυψη υπολογιστικά προβλεπόμενων στόχων των microRNAs σε lncRNAs. Δεδομένα που προήλθαν από την ανάλυση των miRNAs/lncRNAs, καθώς και οι πιθανές αλληλεπιδράσεις τους είτε πειραματικά υποστηριζόμενες (πάνω από 5,000) ή υπολογιστικά (*in silico*) προβλεπόμενες (περισσότερες από 10^6), ενοποιήθηκαν και εισήχθησαν σε μία καινούρια βάση δεδομένων, την DIANA-LncBase. Πρόκειται για μια εκτενή βάση, η οποία δύναται να παρέχει υψηλής ποιότητας δεδομένα για πολλαπλές αναλύσεις και μελλοντικά ερευνητικά έργα μέσα στην επιστημονική κοινότητα.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Βιολογία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: microRNA, lncRNA, HITS-CLIP, PAR-CLIP, πρόβλεψη στόχων, πειραματικά επιβεβαιωμένοι στόχοι.

ABSTRACT

microRNAs (miRNAs) are small endogenous RNA molecules (~22nt) that belong in the wide ncRNA family. They play a major role in a plethora of biological processes and especially act as post-transcriptional gene regulators. lncRNAs, on the other hand, are long non coding transcripts, larger than 200 nucleotides. They are identified either on intergenic regions, or in genomic locations that exhibit overlap with protein-coding genes. Possible interactions of lncRNAs-microRNAs can indirectly regulate gene expression.

Aim of this study, is to explore microRNA – lncRNA putative functional interactions, based on high-throughput experimental data, as well as on state of the art *in silico* target predictions. Moreover, it provides for the first time, a comprehensive annotation of microRNA (miRNA) targets on human and mouse lncRNAs. miRNA targets of large collections of lncRNA transcripts have not yet been extensively studied through experimental data analysis, and there are only few lncRNA-miRNA interactions reported in the available literature.

The first part of the study includes the collection of the available lncRNA resources, derived from the most recently updated biological databases and the relevant literature. Subsequently, miRNA targets on lncRNAs have been identified with the analysis of high-throughput data, and more precisely HITS/PAR-CLIP experimental data. Moreover, one of the best target prediction algorithms in terms of sensitivity and specificity, DIANA-microT-CDs, has been utilized for computational detection of miRNA targets on lncRNAs. Finally, miRNA/lncRNA-related information, as well as experimentally verified and *in silico* predicted miRNA-lncRNA interactions have populated a novel database named DIANA-LncBase. The experimentally supported entries available in DIANA-LncBase correspond to more than 5,000 interactions, while the computationally predicted interactions exceed 10 million. DIANA-LncBase is an extensive repository of miRNA-lncRNA interactions that can provide high quality data for numerous analyses and future research projects throughout the community.

SUBJECT AREA: Computational Biology

KEYWORDS: microRNA, lncRNA, HITS-CLIP, PAR-CLIP, target prediction, experimentally verified targets.