



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCE  
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS**

**POSTGRADUATE PROGRAM  
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"**

**MASTER THESIS**

**A computational approach to identify long non-coding RNAs  
acting as microRNA sponges**

**Anna A. Karavangeli**

***Supervisor:*** **Prof. Artemis Hatzigeorgiou**, Professor of Bioinformatics,  
Department of Electrical & Computer Engineering,  
Telecommunications and Networks, University of Thessaly

**ATHENS**

**JUNE 2019**



**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**

**Υπολογιστική μέθοδος για την αναγνώριση μακρών μη  
κωδικών RNA που δρουν σαν «σφουγγάρια» των microRNA**

**Άννα Α. Καραβαγγέλη**

**Επιβλέπουσα :** **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής,  
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του  
Πανεπιστημίου Θεσσαλίας

**ΑΘΗΝΑ**

**ΙΟΥΝΙΟΣ 2019**

## MASTER THESIS

**A computational approach to identify long non-coding RNAs acting as microRNA sponges**

**Anna A. Karavangeli**

**SRN.: ΠΙΒ0174**

**Supervisor :** **Prof. Artemis Hatzigeorgiou**, Professor of Bioinformatics,  
Department of Electrical & Computer Engineering,  
Telecommunications and Networks, University of Thessaly

**EXAMINATION  
COMMITTEE:**

**Prof. Artemis Hatzigeorgiou**, Professor of Bioinformatics,  
Department of Electrical & Computer Engineering,  
Telecommunications and Networks, University of Thessaly  
**Martin Reczko**, Head of the bioinformatics group of the  
genomics facility, Alexander Fleming  
**Dr. Ema Anastasiadou**, Researcher-Lecturer Level,  
Biomedical Research Foundation of the Academy of  
Athens (BRFAA)

June 2019

## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υπολογιστική μέθοδος για την αναγνώριση μακρών μη κωδικών RNA που δρουν σαν «σφουγγάρια» των microRNA

**Άννα Α. Καραβαγγέλη**

**A.M.: ΠΙΒ0174**

**Επιβλέπουσα :** **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής,  
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του  
Πανεπιστημίου Θεσσαλίας

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:** **Καθ. Άρτεμις Χατζηγεωργίου**, Καθηγήτρια  
Βιοπληροφορικής,  
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του  
Πανεπιστημίου Θεσσαλίας  
**Martin Reczko**, Επικεφαλής Τμήματος Βιοπληροφορικής  
Ερευνητικού Κέντρου Βιοϊατρικών Επιστημών, Αλέξανδρος  
Φλέμινγκ  
**Δρ. Έμα Αναστασιάδου**, Ερευνήτρια Δ',  
Ιδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών  
(ΙΙΒΕΑΑ)

Ιούνιος 2019

## ABSTRACT

Long non-coding RNAs (lncRNAs) are transcribed non-coding RNAs (ncRNAs) that are more than 200 nucleotides long. Among their reported functions, their ability to act as molecular “sponges” for microRNAs (miRNAs) in several physiological processes and pathological conditions has gained attention over the past few years. According to the ceRNA hypothesis, by sequestering miRNAs, lncRNAs can reduce the number of the available miRNAs that target mRNAs and indirectly prevent target gene repression.

In order to investigate in what extent lncRNAs reduce the amount of miRNAs available to other targets and whether individual lncRNAs can be characterized as “sponges”, a mathematical quantitative model of binding site competition was employed. Argonaute Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (AGO-PAR-CLIP), RNA-Seq and small RNA-Seq (sRNA-Seq) experiments were used to identify targets and quantify target and miRNA abundances for to 2 different tissues. Site occupancies (fraction of sites bound by the miRNA) for protein coding targets were predicted in the presence and absence of lncRNAs. Increased occupancies of miRNA targeted mRNAs, observed in the lack of lncRNAs, lead to potential lncRNA “sponges”.

This analysis resulted in the identification of 8 lncRNAs acting as potential sponges for 38 miRNAs. The abundance of most individual targets was insufficient to alter miRNA levels and the reported sponges were highly expressed. Two well-studied nuclear lncRNAs, XIST and MALAT1, were among them suggesting additional functionalities of lncRNAs in certain settings, but also highlighting the need for separate analysis of nuclear and cytoplasmic RNAs.

**SUBJECT AREA:** Computational Biology, Bioinformatics

**KEYWORDS:** microRNA, lncRNA, molecular sponges, competition, mathematical modeling

## ΠΕΡΙΛΗΨΗ

Τα μακρά μη κωδικά RNAs (lncRNAs) είναι μεταγραφόμενα μη κωδικά RNAs (ncRNAs) με μήκος πάνω από 200 νουκλεοτίδια. Μεταξύ των αναγνωρισμένων λειτουργιών τους, η ικανότητά τους να δρουν σαν μοριακά "σφουγγάρια" για τα microRNAs (miRNAs) σε διάφορες φυσιολογικές διεργασίες και παθολογικές καταστάσεις έχει κερδίσει προσοχή τα τελευταία χρόνια. Σύμφωνα με την υπόθεση του ceRNA, με την πρόσδεση των miRNAs, τα lncRNAs μπορούν να μειώσουν τον αριθμό των διαθέσιμων miRNAs που στοχεύουν mRNAs και να αποτρέψουν έμμεσα τη καταστολή του γονιδίου στόχου

Για να διερευνηθεί σε ποιο βαθμό τα lncRNAs μειώνουν την ποσότητα των miRNAs που είναι διαθέσιμα σε άλλους στόχους και αν μεμονωμένα lncRNAs μπορούν να χαρακτηριστούν ως "σφουγγάρια", χρησιμοποιήθηκε ένα μαθηματικό ποσοτικό μοντέλο για τον ανταγωνισμό των θέσεων πρόσδεσης. Argonaute Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (AGO-PAR-CLIP), RNA-Seq και small RNA-Seq (sRNA-Seq) πειράματα χρησιμοποιήθηκαν για τον προσδιορισμό των στόχων και την ποσοτικοποίηση της αφθονίας τόσο των ίδιων όσο και των miRNA για 2 διαφορετικούς ιστούς. Το ποσοστό των θέσεων που δεσμεύονται από τα miRNA για πρωτεϊνικούς στόχους προβλέφθηκε παρουσία και απουσία των lncRNA. Αυξημένα ποσοστά πρόσδεσης των mRNA στόχων, που παρατηρούνται απουσία των lncRNA, οδηγούν σε πιθανά lncRNA «σφουγγάρια»

Η ανάλυση αυτή είχε σαν αποτέλεσμα την ταυτοποίηση 8 lncRNAs με πιθανή λειτουργία σφουγγαριού για 38 miRNAs. Η αφθονία των περισσότερων επιμέρους στόχων ήταν ανεπαρκής για να μεταβάλει τα επίπεδα του miRNA και τα αναφερθέντα σφουγγάρια εκφράζονταν σε μεγάλο βαθμό. Δύο καλά μελετημένα πυρηνικά lncRNAs, το XIST και το MALAT1, αναγνωρίστηκαν μεταξύ των άλλων, υποδεικνύοντας επιπλέον λειτουργίες των lncRNA σε συγκεκριμένα περιβάλλοντα, αλλά και υπογραμμίζοντας την ανάγκη ξεχωριστής ανάλυσης πυρηνικών και κυτταροπλασματικών RNAs.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Υπολογιστική Βιολογία, Βιοπληροφορική

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** microRNA, lncRNA, μοριακά σφουγγάρια, ανταγωνισμός, μαθηματική μοντελοποίηση