
UNIVERSITY OF THESSALY THESIS

microRNA target prediction

by

Stathaki Areti-Anna

Supervisor

Prof. Hatzigeorgiou Artemis

Co-Supervisor

Prof. Potamianos Gerasimos



Department of Electrical and Computer Engineering

School of Engineering

June2014



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ & ΚΕΝΤΡΟ ΠΛΗΡΟΦΟΡΗΣΗΣ
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»**

Αριθ. Εισ.: 12672/1
Ημερ. Εισ.: 30-09-2014
Δωρεά: Συγγραφέα
Ταξιθετικός Κωδικός: ΠΤ - ΗΜΜΥ
2014
ΣΤΑ

ABSTRACT

This thesis investigates the performance of microRNA target prediction. More specifically, microRNAs are a class of post-transcriptional regulators. They are short ~22 nucleotide RNA sequences that bind to complementary sequences in the 3' UTR of multiple target mRNAs, usually resulting in their silencing.

MicroRNAs were first discovered in 1993 by Victor Ambros, Rosalind Lee and Rhonda Feinbaum during a study into development in the nematode *Caenorhabditis elegans* (*C. elegans*) regarding the gene *lin-14*.

Since then over 4000 miRNAs have been discovered in all studied eukaryotes including mammals, fungi and plants. More than 700 miRNAs have so far been identified in humans and over 800 more are predicted to exist.

The scope of this thesis is mainly to focus on the comparison of three different computational methods of miRNA targeting.

DEDICATIONS AND ACKNOWLEDGEMENTS

Πρώτα από όλα θα ήθελα να ευχαριστήσω θερμά την κυρία Χατζηγεωργίου Άρτεμις για την άμεση ανταπόκριση της στην επιθυμία μου να ασχοληθώ με το τομέα της Βιοπληροφορικής. Οι γνώσεις που αποκόμισα πάνω στους υπολογιστές όλα αυτά τα χρόνια, σε συνδυασμό με την ενασχόληση μου με ένα τομέα περισσότερο ανθρωποκεντρικό, όπως αυτός της βιολογίας, έκλεισαν ευχάριστα ένα από τα πιο αγαπημένα κομμάτια της ζωής μου, αυτό της φοιτητικής. Ευχαριστώ επίσης, θερμά τον κύριο Γεωργακίλα Γιώργο για την πάντα άμεση και πολύτιμη βοήθειά του.

Τέλος, δεν θα μπορούσα να ξεχάσω όλους αυτούς τους πολύ δικούς μου ανθρώπους που στάθηκαν δίπλα μου όλον αυτό τον καιρό με τόση υπομονή και αφοσίωση, ο καθένας με τον δικό του τρόπο.

TABLES OF CONTENTS

| <i>List of Tables</i> | <i>Pages</i> |
|---|--------------|
| Περίληψη..... | 1 |
| 1 Γενική Εισαγωγή και Βασικές έννοιες | |
| 1.1 Η επιστήμη της Βιοπληροφορικής..... | 2-3 |
| 1.2 Το DNA είναι το γενετικό υλικό..... | 4 |
| 1.3 Δομή του DNA..... | 5 |
| 1.4 Αντιγραφή, έκφραση και ρύθμιση της γενετικής πληροφορίας | |
| 1.4.1 Η αντιγραφή του DNA και ο ρόλος των ενζύμων..... | 6-7 |
| 1.4.2 Μεταγραφή του DNA..... | 8 |
| 2 Γενικές πληροφορίες για το RNA | |
| 2.1 Εισαγωγή..... | 9 |
| 2.2 Τα κυριότερα στάδια της ωρίμανσης του mRNA και ο φυσικός τους ρόλος..... | 10-12 |
| 2.3 Δομή του mRNA μετά την ωρίμανση..... | 13 |
| 3 Εισαγωγή στο MicroRNA | |
| 3.1 Εισαγωγή..... | 14 |
| 3.2 Βιογενετική..... | 15 |
| 3.3 Seed Regions..... | 16 |
| 3.4 Μεταγραφή του microRNA..... | 17 |
| 3.5 miRNA και γονιδιακή έκφραση..... | 18 |
| 3.6 Η εξέλιξη των microRNAs..... | 19 |
| 3.7 Ασθένειες και κλινικές εφαρμογές που συνδέονται με το miRNA..... | 19-20 |
| 4 MicroRNA target prediction | |
| 4.1 Online databases and tools..... | 21 |
| 4.2 Conservation..... | 21 |
| 4.3 Thermodynamics..... | 21 |

| | | |
|-----|--|-----------|
| 4.4 | Sensitivity and Specificity..... | 22 |
| 4.5 | Βασικές έννοιες για την ανάπτυξη αλγορίθμων στο τομέα της Βιοπληροφορικής..... | 23-24 |
| 4.6 | Παρουσίαση των πιο γνωστών αλγορίθμων που υπάρχουν σήμερα..... | 25-27 |
| 5 | Αλγόριθμοι και Μέθοδοι ανάπτυξης..... | 28 |
| 6 | Οι Βάσεις Δεδομένων που χρησιμοποιήθηκαν | |
| 6.1 | Η βάση δεδομένων miRBase..... | 29-30 |
| 6.2 | Η βάση δεδομένων Ensembl 69..... | 31 |
| 6.3 | MREs και interaction miRNA-gene..... | 32 |
| 6.4 | Ο αλγόριθμος miRanda..... | 33-34 |
| 6.5 | Ο αλγόριθμος targetscan..... | 35-36 |
| 6.6 | Ο αλγόριθμος DIANA-microT..... | 37-39 |
| 7 | Τελικό στάδιο εργασίας | |
| 7.1 | Μορφοποίηση εισόδων..... | 40-42 |
| 7.2 | Threshold..... | 43-45 |
| 8 | Validated Results | |
| 8.1 | TarBase v6.0..... | 46 |
| 8.2 | Final Results..... | 47 |
| 9 | Επίλογος..... | 48 |

Περίληψη

Τα διάφορα υπολογιστικά εργαλεία παίζουν καθοριστικό ρόλο γύρω απ την έρευνα των miRNAs. Το microRNA είναι μία μικρή μη κωδικοποιημένη μορφή του RNA, μικρού μεγέθους της τάξης των 22 νουκλεοτιδίων περίπου. Σύγχρονες έρευνες έχουν αποδείξει, ότι τα microRNAs παίζουν καθοριστικό ρυθμιστικό ρόλο σε διάφορες λειτουργίες του κυττάρου, εμφανίζοντας άμεση εμπλοκή με διάφορες ασθένειες. Ένα δεδομένο miRNA μπορεί να έχει πολλαπλούς διαφορετικούς στόχους mRNA και αντίστροφα, ένας συγκεκριμένος στόχος θα μπορούσε να απευθύνεται σε πολλαπλά miRNAs.

Τα πρώτα δείγματα miRNAs εντοπίστηκαν στις αρχές του 1990. Ωστόσο, τα miRNAs αναγνωρίστηκαν ως ξεχωριστή κατηγορία βιολογικών ρυθμιστών μόλις στις αρχές της δεκαετίας του 2000. Από τότε, η έρευνα γύρω από τα miRNAs αποκάλυψε πολλαπλούς ρόλους στη πιθανή συμμετοχή τους στην αρνητική και θετική ρύθμιση του κυττάρου (μεταγραφική καταστολή ή ενεργοποίηση).

Η παρακάτω διπλωματική εργασία παρουσιάζει αρχικά, μία εισαγωγή σε βασικές βιολογικές έννοιες, κάνοντας μία ανασκόπηση στην Ιστορία της Βιολογίας απ' τις αρχές της δεκαετίας του '80 μέχρι σήμερα.

Ιδιαίτερη έμφαση δίνεται σε μία μικρή κατηγορία RNA, που ονομάζεται όπως αναφέραμε microRNA. Στη συνέχεια, περιγράφονται τρεις διαφορετικοί αλγόριθμοι οι οποίοι καταγράφουν πιθανούς στόχους σύνδεσης του microRNA πάνω στο γονίδιο. Οι διάφοροι αλγόριθμοι παρουσιάζουν διαφορετικά αποτελέσματα ανάλογα με τις μεθόδους υπολογισμού που χρησιμοποιούν. Στο τέλος της παρουσίασης δίνονται αναλυτικοί πίνακες με τα αποτελέσματα της μελέτης.

1 Κεφάλαιο 1ο: Γενική εισαγωγή και Βασικές έννοιες

1.1 Η επιστήμη της Βιοπληροφορικής

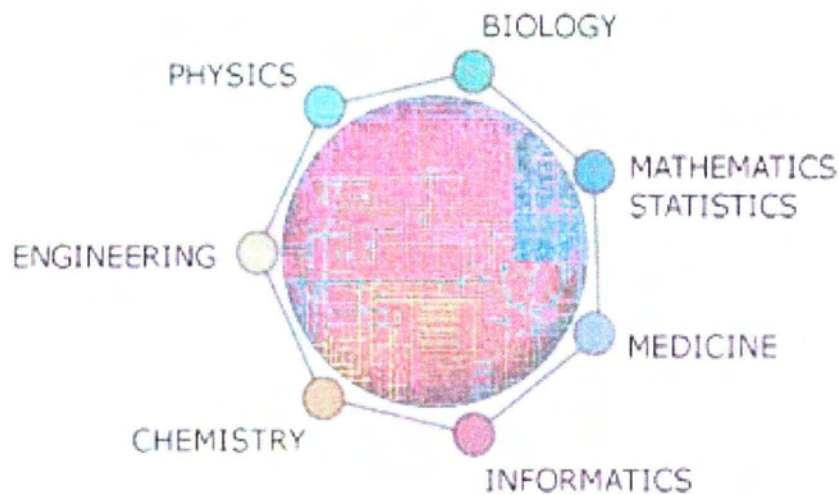
Βιοπληροφορική είναι ένα διεπιστημονικό πεδίο που αναπτύσσει και βελτιώνει τις μεθόδους για την αποθήκευση, την ανάκτηση, την οργάνωση και την ανάλυση βιολογικών δεδομένων. Μια σημαντική δραστηριότητα στη βιοπληροφορική είναι η ανάπτυξη εργαλείων λογισμικού για τη δημιουργία χρήσιμων βιολογικών γνώσεων.

Η βιοπληροφορική χρησιμοποιεί πολλούς τομείς της επιστήμης των υπολογιστών, τα μαθηματικά και την τεχνολογία για την επεξεργασία βιολογικών δεδομένων. Οι βάσεις δεδομένων και τα συστήματα πληροφοριών χρησιμοποιούνται για την αποθήκευση και οργάνωση των βιολογικών δεδομένων. Η ανάλυση βιολογικών δεδομένων συνδέεται με αλγόριθμους τεχνητής νοημοσύνης, με την εξόρυξη δεδομένων, την επεξεργασία εικόνας καθώς και τη προσομοίωση.

Οι αλγόριθμοι με τη σειρά τους εξαρτώνται από θεωρητικές μαθηματικές βάσεις, όπως διακριτά μαθηματικά, θεωρία υπολογισμού, θεωρία πληροφοριών και στατιστικά στοιχεία. Συνήθως χρησιμοποιούνται τεχνολογίες και εργαλεία λογισμικού όπως Java, C, C++, Perl, Python, MATLAB. Για τη παρούσα εργασία, χρησιμοποιήθηκε η γλώσσα Perl. Αποτελεί μία απ τις πολύ δημοφιλείς αντικειμενοστρεφής γλώσσες προγραμματισμού. Σχεδιάστηκε από το Larry Wall και η πρώτη έκδοση της γλώσσας ανακοινώθηκε στις 18 Δεκεμβρίου του 1987. Η Perl είναι προεγκατεστημένη στη συντριπτική πλειοψηφία των λειτουργικών συστημάτων τύπου Unix, ενώ υπάρχουν έτοιμες διανομές για τα περισσότερα λειτουργικά συστήματα.

Βασικές γνώσεις πάνω στην επιστήμη της Βιοπληροφορικής σχετίζονται με:

- εύρεση ομοιοτήτων σε ακολουθίες.
- ομολογία και ομοιότητα ακολουθιών και η σημασία τους.
- αλγόριθμοι δυναμικού προγραμματισμού.
- ευριστικές μέθοδοι για αναζήτηση ομοιοτήτων σε βάσεις δεδομένων όπως οι BLAST, FASTA κλπ.
- Πολυδιάστατοι αλγόριθμοι δυναμικού προγραμματισμού.
- φυλογενετικά δένδρα και πολλαπλές στοιχίσεις.
- Αλγόριθμοι πρόγνωσης στηριζόμενοι στην ακολουθία.
- εύρεση πιθανών γονιδίων σε ακολουθίες DNA, πολλαπλές στοιχίσεις ακολουθιών με χρήση Hidden Markov Models, κατάταξη ακολουθιών σε οικογένειες.



Σχήμα 1: Bioinformatics.

1.2 Το DNA είναι το γενετικό υλικό

Το δε(σ)οξυριβο(ζο)νουκλεϊ(νι)κό οξύ (Deoxyribonucleic acid – DNA) είναι ένα νουκλεϊκό οξύ που περιέχει τις γενετικές πληροφορίες που καθορίζουν τη βιολογική ανάπτυξη όλων των κυτταρικών μορφών ζωής και των περισσοτέρων ιών. Παρ' όλο που το DNA εντοπίστηκε στον πυρήνα των κυττάρων το 1869, έως και το 1944 δεν ήταν γνωστό ότι αποτελεί το γενετικό υλικό των οργανισμών. Οι επιστήμονες πίστευαν ότι τα μόρια που μεταφέρουν πληροφορία είναι οι πρωτεΐνες, που παρουσιάζουν μεγαλύτερη ποικιλομορφία, επειδή είναι αποτέλεσμα συνδυασμού είκοσι διαφορετικών αμινοξέων, ενώ το DNA είναι συνδυασμός τεσσάρων μόνο νουκλεοτιδίων.

Η ποσότητα του DNA σε κάθε οργανισμό είναι σταθερή και δε μεταβάλλεται από αλλαγές στο περιβάλλον. Η ποσότητα του DNA είναι επίσης ίδια σε όλα τα είδη κυττάρων ενός οργανισμού όπως στην περίπτωση του ανθρώπου σε αυτά του σπλήνα, της καρδιάς, του ήπατος κτλ.

Οι γαμέτες των ανώτερων οργανισμών, που είναι απλοειδείς, περιέχουν τη μισή ποσότητα DNA από τα σωματικά κύτταρα, που είναι διπλοειδή. Η ποσότητα του DNA είναι, κατά κανόνα, ανάλογη με την πολυπλοκότητα του οργανισμού. Συνήθως, όσο εξελικτικά ανώτερος είναι ο οργανισμός τόσο περισσότερο DNA περιέχει σε κάθε κύτταρό του.

Το 1953 οι Τζέιμς Γουάτσον και Φράνσις Κρικ δύο ερευνητές που εργάζονταν στο Πανεπιστήμιο του Καίμπριτζ παρουσίασαν ένα «μοντέλο» της δομής του DNA, που ονομάστηκε «μοντέλο της διπλής έλικας». Σύμφωνα με το μοντέλο αυτό, το μόριο του DNA παρουσιάζεται με τα ακόλουθα τέσσερα βασικά χαρακτηριστικά:

1. Αποτελείται από δύο πολυνουκλεοτιδικές αλυσίδες σε μορφή δύο αντιτακτών κλώνων που σχηματίζουν δεξιόστροφη διπλή έλικα.

2. Οι αζωτούχες βάσεις (ή πρωτεϊνικές) κάθε κλώνου είναι κάθετες ως προς τον άξονα του μορίου και προεξέχουν προς το εσωτερικό της συστροφής.

3. Οι δύο δημιουργούμενοι κλώνοι συγκρατούνται μεταξύ τους με δεσμούς υδρογόνου. Τα δε ζευγάρια των αζωτούχων βάσεων όπου αναπτύσσονται μεταξύ τους δεσμοί υδρογόνου είναι καθορισμένα: η αδενίνη με τη θυμίνη και η γουανίνη με την κυτοσίνη.

4. Μεταξύ της αδενίνης και της θυμίνης σχηματίζονται δύο δεσμοί υδρογόνου, ενώ μεταξύ της γουανίνης και της κυτοσίνης τρεις δεσμοί υδρογόνου.

1.3 Δομή του DNA

Η διαμόρφωση των μεγάλων μορίων του DNA στο χώρο έχει τη μορφή δύο επιμήκων αλυσίδων, οι οποίες συστρέφονται ελικοειδώς μεταξύ τους. Οι αζωτούχες βάσεις στο DNA είναι τέσσερις:

- κυτοσίνη C
- γουανίνη G
- θυμίνη T
- αδενίνη A

Τα συστατικά στοιχεία του DNA είναι τα εξής:

1. Αζωτούχα βάση
2. Σάκχαρο
3. Φωσφορική ομάδα

Οι αζωτούχες βάσεις, ανάλογα με την σειρά αλληλουχίας τους σε τριάδες, κωδικοποιούν το μήνυμα για τη σύνθεση των αμινοξέων του κυττάρου στα ριβοσώματα. Εκεί τα αμινοξέα συνδυάζονται, με τη σειρά κατά την οποία μεταφέρθηκαν στο ριβόσωμα και συντίθενται έτσι οι διαφορετικές πρωτεΐνες.

Το DNA έχει διπλή ελικοειδή δομή. Όμως, δεν είναι συμμετρικό. Έχει κατεύθυνση «προς τα εμπρός» και «προς τα πίσω». Τα άκρα ονομάζονται 5' και 3' με βάση τα άτομα του άνθρακα στο συστατικό στοιχείο του σακχάρου.

1869 Απομονώνεται DNA από τον κυτταρικό πυρήνα.

1902 Ο Emil Hermann Fischer κερδίζει το βραβείο Νόμπελ: δείχνει ότι τα αμινοξέα ενώνονται και σχηματίζουν πρωτεΐνες.

1903 Αποδεικνύεται ότι τα χρωμοσώματα είναι φορείς του κληρονομικού (γενετικού) υλικού.

1911 Ο Thomas Hunt Morgan ανακαλύπτει ότι τα γονίδια στα χρωμοσώματα είναι οι διακριτές μονάδες της κληρονομικότητας.

1911 Ο Pheobus Aaron Theodore Lerene ανακαλύπτει το RNA.

1944 Αποδεικνύεται ότι το DNA είναι το γενετικό υλικό.

1950 Ο Edwin Chargaff ανακαλύπτει ότι η κυτοσίνη είναι συμπληρωματική της γουανίνης, και η αδενίνη είναι συμπληρωματική της θυμίνης.

1953 Ανακαλύπτεται η δομή της διπλής έλικας του DNA. και συνεχίζεται μέχρι σήμερα...

2003 Ολοκληρώνεται το Πρόγραμμα Έρευνας του Ανθρώπινου Γονιδιώματος. Προσδιορίζεται η αλληλουχία του γονιδιώματος του ποντικού.

1.4 Αντιγραφή, έκφραση και ρύθμιση της γενετικής πληροφορίας

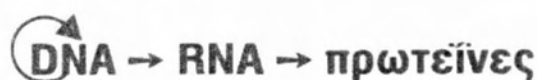
1.4.1 Η αντιγραφή του DNA και ο ρόλος των ενζύμων

Η διαδικασία της αντιγραφής φαίνεται απλή, όμως στη πραγματικότητα ύστερα από πολύχρονη ερευνητική μελέτη, διαπιστώθηκε ότι είναι ιδιαίτερα πολύπλοκη. Ο μηχανισμός της αντιγραφής έχει μελετηθεί πολύ περισσότερο στα προκαρυωτικά κύτταρα, γιατί το DNA τους είναι πολύ μικρότερο και απλούστερα οργανωμένο από το DNA των ευκαρυωτικών κυττάρων.

Η αντιγραφή του DNA αρχίζει από καθορισμένα σημεία, που ονομάζονται θέσεις έναρξης της αντιγραφής. Για να αρχίσει η αντιγραφή του DNA, είναι απαραίτητο να ξετυλιχθούν στις θέσεις έναρξης της αντιγραφής οι δύο αλυσίδες. Αυτό επιτυγχάνεται με τη βοήθεια ειδικών ενζύμων. Τα ένζυμα αυτά ονομάζονται DNA ελικάσες. Όταν ανοίξει η διπλή έλικα, δημιουργείται μια «θηλιά», η οποία αυξάνεται και προς τις δύο κατευθύνσεις. Τα κύρια ένζυμα που συμμετέχουν στην αντιγραφή του DNA ονομάζονται DNA πολυμεράσες.

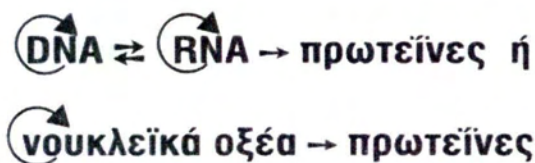
Οι DNA πολυμεράσες λειτουργούν μόνο προς καθορισμένη κατεύθυνση και τοποθετούν τα νουκλεοτίδια στο ελεύθερο 3' άκρο της δεοξυριβόζης του τελευταίου νουκλεοτιδίου κάθε αναπτυσσόμενης αλυσίδας. Έτσι, λέμε ότι αντιγραφή γίνεται με προσανατολισμό 5' προς 3'. Κάθε νεοσυντιθέμενη αλυσίδα θα έχει προσανατολισμό 5'→3'. Έτσι, σε κάθε διπλή έλικα που παράγεται οι δύο αλυσίδες θα είναι αντιπαράλληλες.

Το πρώτο βήμα για την έκφραση της πληροφορίας που υπάρχει στο DNA είναι η μεταφορά της στο RNA με τη διαδικασία της μεταγραφής. Το RNA μεταφέρει με τη σειρά του, μέσω της διαδικασίας της μετάφρασης, την πληροφορία στις πρωτεΐνες που είναι υπεύθυνες για τη δομή και λειτουργία των κυττάρων και κατ' επέκταση και των οργανισμών. Η σχέση αυτή συνοψίζεται στο ακόλουθο σχήμα, όπου τα βέλη δείχνουν την κατεύθυνση της μεταφοράς της γενετικής πληροφορίας:



Σχήμα 2: Ροή της γενετικής πληροφορίας (1)

Για αρκετό καιρό οι ερευνητές πίστευαν ότι όλη η ροή της γενετικής πληροφορίας γινόταν προς τη μία μόνο κατεύθυνση, δηλαδή ότι το DNA μεταγραφόταν σε RNA. Σήμερα είναι γνωστό ότι μερικοί ιοί έχουν RNA ως γενετικό υλικό. Ένα ένζυμο που υπάρχει στους ίδιους τους ιούς, η αντίστροφη μεταγραφάση, χρησιμοποιεί ως καλούπι το RNA, για να συνθέσει DNA. Επιπλέον, σε ορισμένους ιούς το RNA έχει την ικανότητα να αυτοδιπλασιάζεται. Έτσι σήμερα το κεντρικό δόγμα περιγράφεται ως εξής:



Σχήμα 3: Ροή της γενετικής πληροφορίας (2)

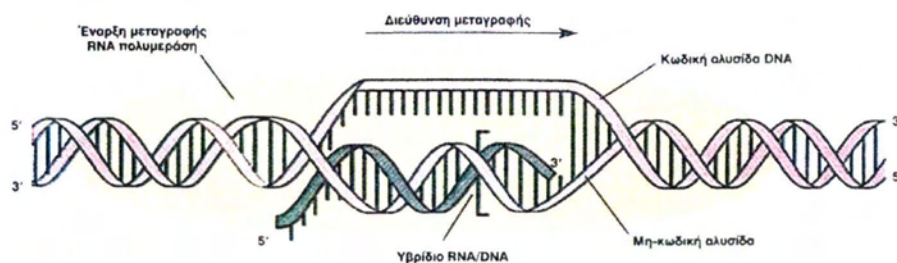
Αγγελιαφόρο RNA (mRNA). Αποτελεί λιγότερο από το 5% του ολικού RNA του κυττάρου. Τα περισσότερα mRNA έχουν μήκος 500-2.000 βάσεις.

Ριβοσωμικό RNA (rRNA). Είναι το πιο άφθονο από τα είδη RNA και αποτελεί το 80% περίπου του συνολικού RNA του κυττάρου. Έχει μέγεθος 100-3.000 βάσεις.

Μεταφορικό RNA (tRNA). Τα μόρια αυτά αποτελούν το 15% περίπου του συνολικού RNA στο κύτταρο. Αποτελείται από 70-85 βάσεις.

1.4.2 Μεταγραφή του DNA

Η μεταγραφή καταλύεται από ένα ένζυμο, την RNA πολυμεράση, η οποία προσδένεται σε ειδικές περιοχές του DNA, που ονομάζονται υποκινητές, με τη βοήθεια πρωτεϊνών που ονομάζονται μεταγραφικοί παράγοντες. Οι υποκινητές και οι μεταγραφικοί παράγοντες αποτελούν τα ρυθμιστικά στοιχεία της μεταγραφής του DNA και επιτρέπουν στην RNA πολυμεράση να αρχίσει σωστά τη μεταγραφή. Οι υποκινητές βρίσκονται πάντοτε πριν από την αρχή κάθε γονιδίου. Κατά την έναρξη της μεταγραφής ενός γονιδίου η RNA πολυμεράση προσδένεται στον υποκινητή και προκαλεί τοπικό ξετύλιγμα της διπλής έλικας του DNA. Στη συνέχεια, τοποθετεί τα ριβονουκλεοτίδια απέναντι από τα δεοξυριβονουκλεοτίδια μίας αλυσίδας του DNA σύμφωνα με τον κανόνα της συμπληρωματικότητας των βάσεων, όπως και στην αντιγραφή, με τη διαφορά ότι εδώ απέναντι από την αδενίνη τοποθετείται το ριβονουκλεοτίδιο που περιέχει ουρακίλη. Η RNA πολυμεράση συνδέει τα ριβονουκλεοτίδια που προστίθενται το ένα μετά το άλλο, με 3'-5'φωσφοδιεστερικό δεσμό. Η μεταγραφή έχει προσανατολισμό 5'→3' όπως και η αντιγραφή.



Σχήμα 4: Μηχανισμός της μεταγραφής του RNA.

Η σύνθεση του RNA σταματά στο τέλος του γονιδίου, όπου ειδικές αλληλουχίες οι οποίες ονομάζονται αλληλουχίες λήξης της μεταγραφής, επιτρέπουν την απελευθέρωσή του. Ο μηχανισμός της μεταγραφής είναι ο ίδιος τόσο στους προκαρυωτικούς όσο και στους ευκαρυωτικούς οργανισμούς.

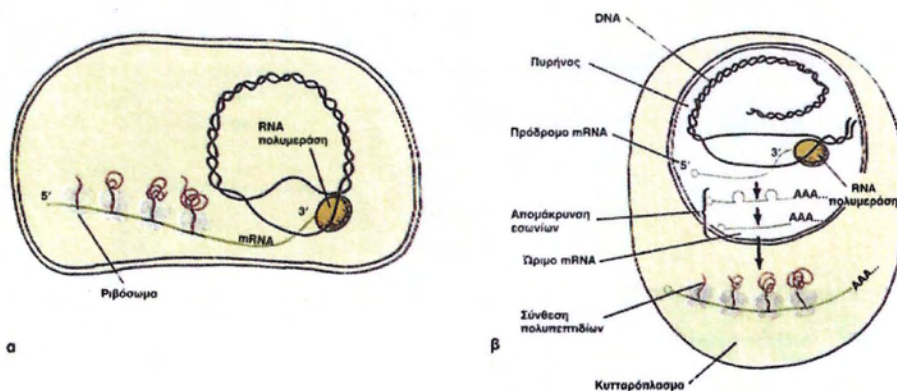
2 Κεφάλαιο 2ο: Γενικές πληροφορίες για το RNA

2.1 Εισαγωγή

Από χημικής άποψης το RNA είναι όμοιο με το DNA. Το RNA είναι το κινητό αντίγραφο της πληροφορίας ενός γονιδίου. Και οι δύο αυτές κατηγορίες νουκλεϊκών οξέων είναι μακρομοριακές ενώσεις μεγάλου μοριακού βάρους. Το μακρομόριο του RNA αποτελείται από επαναλαμβανόμενες δομικές μονάδες τα νουκλεοτίδια. Το μόριο του RNA περιλαμβάνει τέσσερις τύπους νουκλεοτιδίων που συνδέονται μεταξύ τους με 3'-5' φωσφοδιεστερικούς δεσμούς. Ωστόσο, κύρια διαφορά του RNA από το DNA είναι ότι το μόριό του είναι μονόκλωνο έναντι του δίκλωνου DNA, αποτελείται δηλαδή από μια μόνο αλυσίδα.

Βασική επίσης διαφορά, είναι ότι το σάκχαρο στα νουκλεοτίδια του είναι η ριβόζη, εξ ου και η ονομασία τους ριβονουκλεοτίδια, αντί της δεοξυριβόζης στο DNA, και ότι περιέχει την ουρακίλη αντί της θυμίνης. Το RNA χαρακτηρίζεται ως ο «αγγελιοφόρος» μεταξύ του DNA και των πρωτεϊνικών συμπλεγμάτων που είναι γνωστά σαν ριβοσώματα στο κυτταρόπλασμα του κυττάρου (αγγελιαφόρο RNA mRNA). Έτσι, το RNA μαζί με το DNA αποτελούν το γενετικό υλικό των οργανισμών.

Στους προκαρυωτικούς οργανισμούς το mRNA αρχίζει να μεταφράζεται σε πρωτεΐνη πριν ακόμη ολοκληρωθεί η μεταγραφή του. Αυτό είναι δυνατό, επειδή δεν υπάρχει πυρηνική μεμβράνη. Αντίθετα, στους ευκαρυωτικούς οργανισμούς, το mRNA που παράγεται κατά τη μεταγραφή ενός γονιδίου συνήθως δεν είναι έτοιμο να μεταφραστεί, αλλά υφίσταται μια πολύπλοκη διαδικασία ωρίμανσης.



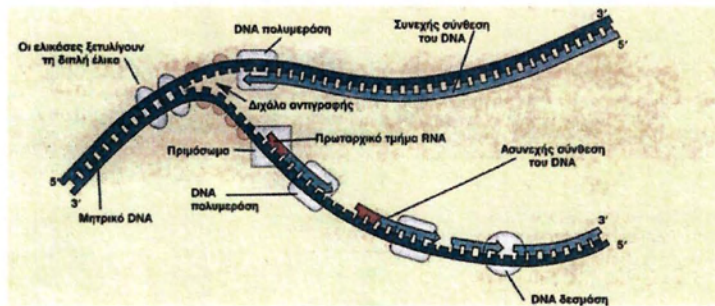
Σχήμα 5: Μεταγραφή - μετάφραση στους προκαρυωτικούς και ευκαρυωτικούς οργανισμούς.

2.2 Τα κυριότερα στάδια της ωρίμανσης του mRNA και ο φυσικός τους ρόλος.

Όπως έχει ήδη αναφερθεί, ο πρωταρχικός ρόλος του γονιδιώματος είναι να προδιαγράψει RNA μόρια. Συγκεκριμένες αλληλουχίες του DNA μεταγράφονται σε RNA αλληλουχίες οι οποίες κωδικοποιούνται σε πρωτεΐνες ή σε “δομικά” RNA όπως το tRNA ή το rRNA .

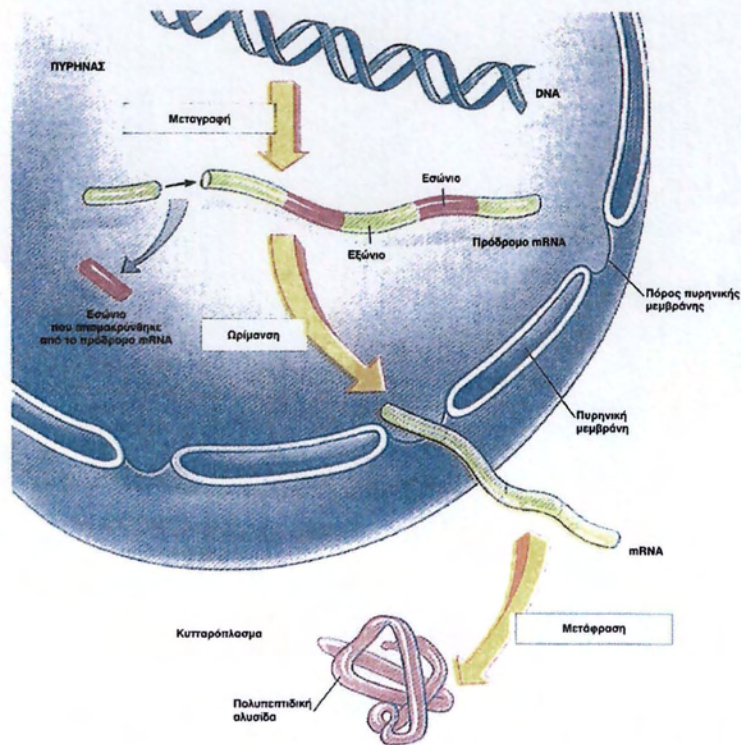
Στους προκαρυωτικούς οργανισμούς το μεγαλύτερο μέρος του DNA κωδικοποιεί αλληλουχίες RNA . Αντίθετα, στους πολυκύτταρους οργανισμούς το μεγαλύτερο μέρος του DNA δεν δείχνει να έχει κωδικοποιητική λειτουργία. Στην *Drosophila* (small fly) για παράδειγμα, μόνο το 5% του ολικού DNA κωδικοποιεί αλληλουχίες που μεταγράφονται σε RNA μόρια και στον άνθρωπο το ποσοστό αυτό είναι μικρότερο.

Στους ανώτερους ευκαρυωτικούς οργανισμούς το μέγεθος των γονιδίων ξεπερνά τα 100.000 bp (Base pairs), μάλιστα μερικά γονίδια περιέχουν παραπάνω από 2.000.000 bp ωστόσο περίπου 1.000 bp είναι απαραίτητα για την κωδικοποίηση μίας πρωτεΐνης μεσαίου μεγέθους (300-400 αμινοξέα). Οι περισσότερες από αυτές τις επιπλέον αλληλουχίες σχηματίζουν μεγάλου μήκους περιοχές οι οποίες δεν κωδικοποιούνται και παρεμβάλλονται μεταξύ των μικρών περιοχών που κωδικοποιούνται. Οι αλληλουχίες που κωδικοποιούνται ονομάζονται εξώνια (exons) ενώ οι περιοχές που δεν κωδικοποιούνται ονομάζονται εσώνια (introns).



Σχήμα 6: Έναρξη της αντιγραφής.

Όταν ένα γονίδιο που περιέχει εσώνια μεταγράφεται, δημιουργείται το πρόδρομο mRNA που περιέχει και εξώνια και εσώνια. Το πρόδρομο mRNA μετατρέπεται σε mRNA με τη διαδικασία της ωρίμανσης, κατά την οποία τα εσώνια κόβονται από μικρά ριβονουκλεοπρωτεϊνικά «σωματίδια» και απομακρύνονται. Τα ριβονουκλεοπρωτεϊνικά σωματίδια αποτελούνται από snRNA και από πρωτεΐνες και λειτουργούν ως ένζυμα: κόβουν τα εσώνια και συρράπτουν τα εξώνια μεταξύ τους. Έτσι σχηματίζεται το «ώριμο» mRNA.



Σχήμα 7: Ωρίμανση του πρόδρομου μορίου του mRNA στον πυρήνα και μεταφορά του στο κυτταρόπλασμα, στους ευκαρυωτικούς οργανισμούς.

Η μετάφραση της γενετικής πληροφορίας γίνεται στα ριβοσώματα. Στους προκαρυωτικούς οργανισμούς, όπου δεν υπάρχει πυρηνική μεμβράνη, η μετάφραση αρχίζει πριν τη λήξη της μεταγραφής. Στους ευκαρυωτικούς, όμως οργανισμούς το αρχικό μεταγράφημα (pre mRNA) υπόκειται ωρίμανση, πριν περάσει στο κυτταρόπλασμα και μεταφραστεί. Η ωρίμανση περιλαμβάνει την απομάκρυνση των εσωνίων, την οργάνωση της καλύπτρας (cap) του 5' άκρου και την προσθήκη της πολυ-A αλληλουχίας στο 3' άκρο (Poly-A tail). Η καλύπτρα είναι απαραίτητη για τη σωστή τοποθέτηση του mRNA στο ριβόσωμα, ενώ η πολυ-A αλληλουχία για τη σταθερότητα του mRNA.

Λόγω κυρίως των εσωνίων που υπάρχουν στα ευκαρυωτικά μεταγράφημα, το μέγεθος των πρωτογενών μεταγραφημάτων είναι κατά κανόνα πολύ πιο μεγάλο στα ευκαρυωτικά παρά στα προκαρυωτικά κύτταρα. Το μεγαλύτερο γνωστό ευκαρυωτικό γονίδιο περιέχει 2,5 εκατομμύρια ζεύγη βάσεων.

Η μετάφραση είναι μια περίπλοκη πορεία, στην οποία συμμετέχουν πολλές πρωτεΐνες, που εμπλέκονται στην έναρξη της μετάφρασης, στην πρόοδο της μετάφρασης και στον τερματισμό της μετάφρασης. Απαραίτητα στη μετάφραση είναι και δύο άλλα είδη RNA, το rRNA και το tRNA.

Σε ορισμένα γονίδια αρχίζει το μάτισμα του μεταγραφήματος, πριν ολοκληρωθεί η σύνθεση του. Ο τερματισμός της μεταγραφής γίνεται σε απόσταση αρκετών χιλιάδων ζευγών βάσεων από το 3' άκρο του ώριμου RNA, το mRNA. Το επιπλέον κομμάτι απομακρύνεται υδρολυτικά και ακολουθείται από πολυαδενυλίωση, δηλαδή από προσθήκη 200 περίπου αδενυλικών νουκλεοτιδίων, με αποτέλεσμα όλα σχεδόν τα μόρια mRNA να καταλήγουν στο 3' άκρο σε μια πολυ-A ουρά.

2.3 Δομή του mRNA μετά την ωρίμανση

- 5' Cap

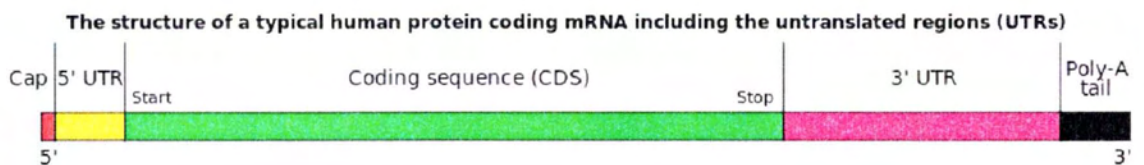
Η πρώτη τροποποίηση στο mRNA στους ευκαρυώτες είναι η προσθήκη της 5' cap. Το Capping εμφανίζεται μόνο στο mRNA.

- Coding DNA Sequence

Η περιοχή κωδικοποίησης ενός γονιδίου, γνωστή ως CDS, είναι το κομμάτι του DNA ή του RNA ενός γονιδίου, που αποτελείται από τα εξόνια, τα οποία κωδικοποιούν πρωτεΐνες. Η περιοχή βρίσκεται ανάμεσα στο 5' και 3' άκρα, ξεκινώντας και τελειώνοντας με τις περιοχές start codon και end codon αντίστοιχα. Η περιοχή κωδικοποίησης mRNA βρίσκεται ανάμεσα στην 5'UTR (untranslated region) και στην 3' UTR (untranslated region).

- Untranslated regions

Στη μοριακή γενετική, η 3' UTR περιοχή είναι το τμήμα του mRNA που ακολουθεί τη CDS περιοχή. Ένα mRNA μόριο μεταγράφεται από την ακολουθία DNA και αργότερα μεταφράζεται σε πρωτεΐνη. Διάφορες περιοχές του mRNA μορίου δεν μεταφράζονται σε πρωτεΐνη συμπεριλαμβανομένης της 5' cap, της 5' utr περιοχής, 3' utr και της poly(A) tail. Η 3' UTR περιέχει συχνά τις ρυθμιστικές περιοχές που επηρεάζουν την μετα-μεταγραφική έκφραση γονιδίων.



Σχήμα 8: Structure of mRNA

3 Κεφάλαιο 3ο: Εισαγωγή στο MicroRNA

3.1 Εισαγωγή

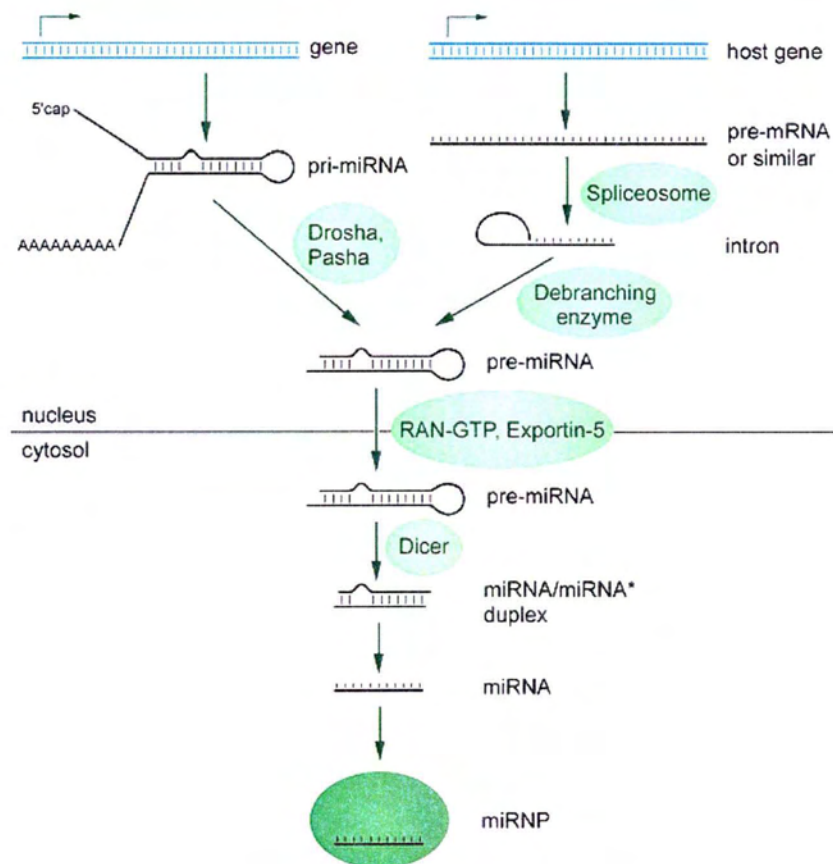
Επιστημονικές μελέτες των τελευταίων χρόνων υποδεικνύουν ότι τα κύτταρα περιέχουν ένα πλήθος από μη-κωδικοποιημένα κομμάτια RNA. Χαρακτηριστικό παράδειγμα αποτελεί το microRNA . Πάρα τη μελέτη που έχει γίνει πάνω στο microRNA και τις λειτουργίες του, πολλά ερωτήματα παραμένουν ακόμα αναπάντητα.

Το microRNA είναι μία μικρή μη κωδικοποιημένη μορφή του RNA μικρού μεγέθους της τάξης των 22 νουκλεοτιδίων περίπου. Μόλις το 1993, ανακαλύφθηκε το πρώτο μέλος της κατηγορίας lin-4 του microRNA μετά από μελέτη για το *C.elegans*. Επτά χρόνια μετά, έρχεται η ανακάλυψη του δεύτερου microRNA , αυτή τη φορά στη κατηγορία let-7, ξανά στο *C.elegans* . Πολύ σύντομα, ανακαλύφθηκαν νέες ομάδες microRNA σε φυτά, ζώα και ιούς. Με την ανακάλυψή τους, γρήγορα γεννήθηκε το ερώτημα του προσδιορισμού των θέσεων που στοχεύουν τα microRNAs .

Στην ουσία τίθεται το εξής ερώτημα, δεδομένου ενός microRNA και ενός γονιδιώματος, ποιά είναι τα γονίδια-στόχοι του συγκεκριμένου microRNA . Οι αρχικές ανακαλύψεις απέδειξαν ότι η lin-4 και let-7 “αντέδρασαν” με την 3' UTR περιοχή του αντίστοιχου γονιδίου. Οι πρώιμες αυτές παρατηρήσεις, οδήγησαν σχεδόν όλες τις ερευνητικές προσπάθειες στο να οριοθετήσουν την περιοχή του mRNA , όπου οι microRNA στόχοι αναζητήθηκαν. Το υπόλοιπο αυτής της παρουσίασης θα επικεντρωθεί στις διάφορες μεθοδολογίες που προτίθενται να αντιμετωπίσουν το πρόβλημα της πρόβλεψης-στόχου του microRNA. Με την εξέλιξη της τεχνολογίας σε συνδυασμό με την ανάπτυξη νέων αλγοριθμικών μεθόδων, ο προσδιορισμός αυτών των στόχων είναι πλέον παρελθόν.

3.2 Βιογενετική

Η πλειοψηφία των microRNAs μεταγράφονται ως ξεχωριστές οντότητες. Ωστόσο, σε ορισμένες περιπτώσεις, ένα microRNA γονίδιο μεταγράφεται μαζί με το γονίδιο-ξενιστή του, παρέχοντας μία συζευγμένη ρύθμιση του miRNA με το γονίδιο που κωδικοποιεί πρωτεΐνη. Περίπου, το 40% των miRNAs γονιδίων βρίσκονται στα ιντρόνια των γονιδίων που κωδικοποιούν πρωτεϊνικές και μη πρωτεϊνικές αλληλουχίες γονιδίων. Το 42-48% περίπου του συνόλου των miRNAs προέρχονται από polycistronic μονάδες που περιέχουν πολλαπλούς διακριτούς βρόχους (loops) από τα οποία τελικά, ωριμάζουν τα microRNAs. Παρόλα αυτά, miRNAs που ανήκουν στην ίδια οικογένεια δεν παρουσιάζουν απαραίτητα ομοιομορφία ως προς τη δομή και τη λειτουργία τους.



Σχήμα 9: Overview of microRNA processing

3.3 Seed Regions.

Σχεδόν μια δεκαετία μετά την ανακάλυψη των πρώτων δειγμάτων miRNA, άρχισε ο προσδιορισμός των στόχων του. Έτσι, τα νουκλεοτίδια 2-7 του 5' άκρου του microRNA, η λεγόμενη seed περιοχή, είναι ιδιαίτερα σημαντική για τον προσδιορισμό των mRNA στόχων. Προσδιορίστηκαν τρεις διαφορετικές κατηγορίες seed, με βάση τη χωρική διάταξη των interactions.

Οι διαφορές στα αποτελέσματα που παρέχονται από τους διαθέσιμους αλγόριθμους πρόβλεψης εξαρτώνται από το πόσο "αυστηρά" είναι τα κριτήρια αντιστοίχισης μεταξύ miPNA και το στοχευόμενο μPNA. Οι υλοποιήσεις επίσης μπορεί να ποικίλουν ανάλογα με το χειρισμό της θέσης διατήρησης του στόχου στις πειοχές αυτές. Η αποδοτική πρόβλεψη μειώνει κατά ένα μεγάλο βαθμό το ποσοστό των ψευδώς θετικών αποτελεσμάτων.

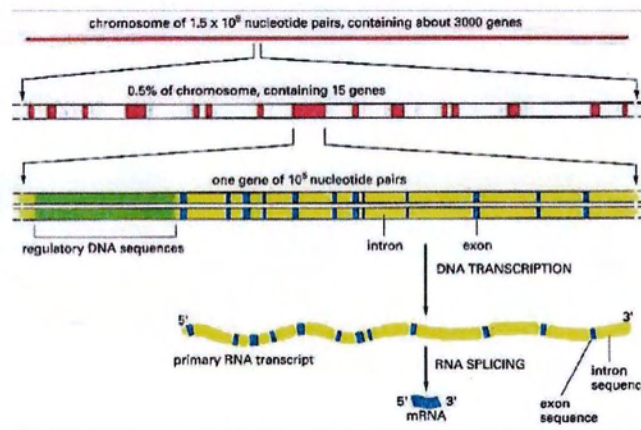
Τα περισσότερα miRNA "δένουν" με 7 νουκλεοτίδια με τον γονίδιο-στόχο. Υπάρχουν 4 διαφορετικά είδη ταιριάσματος: 8mer, 7mer-m8, 7mer-A1 και 6mer. Το 8mer ταιρίασμα αυξάνει την εξειδίκευση (specificity), ενώ σε αντίθεση το 6mer ταιρίασμα αποδίδει μεγαλύτερη ευαισθησία (sensitivity).

3.4 Μεταγραφή του microRNA

Τα γονίδια microRNA μεταγράφονται από την RNA πολυμεράση II ως μεγάλα πρωτογενή μετάγραφα (pre-microRNA), που υποβάλλονται σε επεξεργασία από ένα σύμπλοκο πρωτεϊνών που περιέχουν το ένζυμο RNase III Droscha. Μετά τη παραπάνω διαδικασία σχηματίζεται το πρόδρομο microRNA, της τάξης των 70 περίπου νουκλεοτιδίων. Το πρόδρομο αυτό miRNA μεταφέρεται στη συνέχεια στο κυτταρόπλασμα, όπου υποβάλλεται σε επεξεργασία από ένα δεύτερο ένζυμο το RNase III, DICER, για να σχηματιστεί τελικά το ώριμο microRNA περίπου 22 νουκλεοτιδίων.

Με την επιστήμη της βιοπληροφορικής, τα πρώτα δείγματα miRNAs προβλέφθηκαν με βάση τα διαρθρωτικά χαρακτηριστικά του miRNA και του αντίστοιχου γονιδιώματος. Ουσιαστικά, οι αλγόριθμοι “ταιριάζουν” τα υποψήφια miRNAs με μη κωδικοποιημένες περιοχές του γονιδιώματος. Στη συνέχεια, τα υποψήφια miRNAs φιλτράρονται ελέγχοντας τα επίπεδα διατήρησής τους, στα διάφορα είδη οργανισμών.

Τα ήδη γνωστά miRNAs, ρυθμίζουν και “εκπαιδεύουν” τους διάφορους αλγόριθμους, με σκοπό την διάκριση ανάμεσα στις πραγματικές προβλέψεις και τα ψευδώς θετικά αποτελέσματα. Μετά, τη πρόβλεψη, πολλές πειραματικές τεχνικές όπως μοριακή κλωνοποίηση, αλληλουχία ή υβριδισμός χρησιμοποιούνται με τελικό σκοπό την επικύρωση της πρόβλεψης. Οι βιολογικές προσεγγίσεις σε συνδυασμό με τους διάφορους αλγόριθμους προσέγγισης, οδήγησαν τα τελευταία χρόνια σε μία “έκρηξη” νέων miRNAs.



Σχήμα 11: mRNA mature process

3.5 miRNA και Γονιδιακή έκφραση

Τα miRNAs ελέγχουν ενεργά τη γονιδιακή έκφραση του mRNA, την αποσταθεροποίηση ή την καταστολή της μετάφρασης. Αυτή η αλληλεπίδραση αποτρέπει την παραγωγή πρωτεϊνών με την καταστολή της σύνθεσης τους ή ταυτόχρονα αποικοδομώντας το mRNA. Επιπλέον, κάθε mRNA μπορεί να περιέχει πολλαπλές θέσεις δέσμησης για διαφορετικά miRNAs, με αποτέλεσμα ένα πολύπλοκο δίκτυο αλληλεπιδράσεων. Τυπικά, στα ζώα το microRNA συνδέεται με το 3' untranslated region (UTR) του mRNA στόχου, με στόχο τη καταστολή ή όχι της γονιδιακής έκφρασης.

Αντίθετα, στα φυτά το miRNA συνδέεται με την coding region με σχεδόν τέλειο "ταίριασμα" με τον RNA στόχο. Η ανακάλυψη του miRNA και των πιθανών στόχων του οδήγησε σε σημαντική εξέλιξη την μοριακή διάγνωση και κατά συνέπεια την καλύτερη κατανόηση των βασικών μηχανισμών μπορεί να οδηγήσουν σε πιο εξελιγμένες θεραπευτικές μεθόδους.

Τα πρώτα δείγματα miRNAs εντοπίστηκαν στις αρχές της δεκαετίας του 1990. Ωστόσο, miRNAs αναγνωρίστηκαν ως διακριτή κατηγορία βιολογικών ρυθμιστών, μόλις αρχές του 2000. Από τότε, οι επιστημονικές έρευνες έδειξαν πολλαπλούς ρόλους στην αρνητική και θετική έκφραση του miRNA, με ενδεχόμενη συμμετοχή στη μεταφραστική ενεργοποίηση και καταστολή. Κατά τα τελευταία χρόνια, γύρω στο 500 miRNAs γονίδια έχουν ανακαλυφθεί σε άνθρωπο. Ένα κρίσιμο κομμάτι της λειτουργικότητας του miRNA είναι ο προσδιορισμός των miRNA στόχων με τους οποίους αλληλεπιδρούν άμεσα.

3.6 Η εξέλιξη των microRNAs

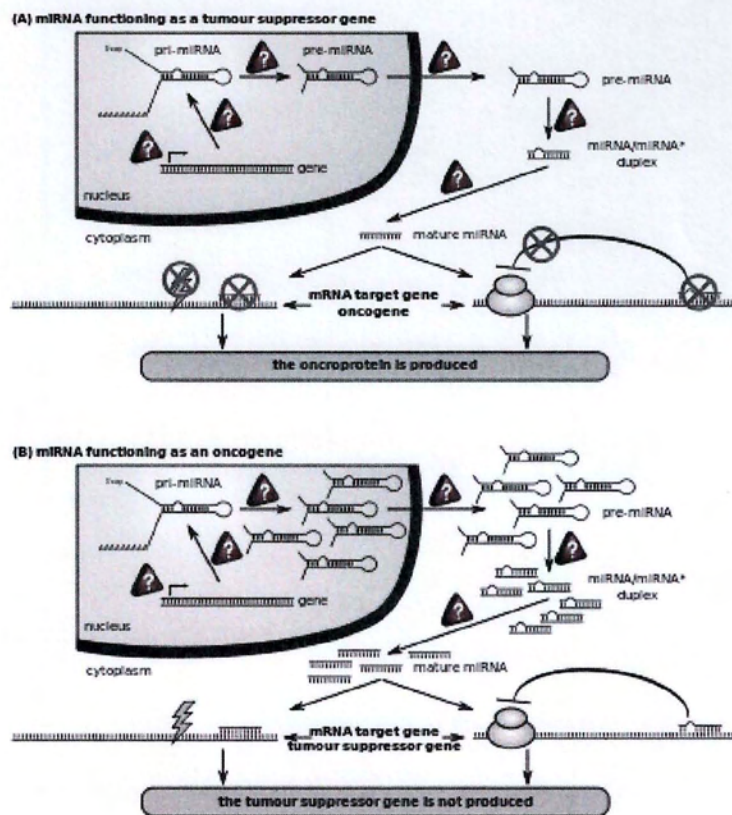
Τα miRNAs αποτελούν σημαντικούς φυλογενετικούς δείκτες λόγω των χαμηλών ποσοστών στην εξέλιξή τους. Τα miRNAs αρχικά, χρησιμοποιήθηκαν ως μέσο άμυνας ενάντια σε εξωγενείς παράγοντες, όπως οι ιοί. Με την ανακάλυψή τους, επιτράπηκε η ανάπτυξη νέων μορφολογικών καινοτομιών. Αντίθετα, τα νέα miRNAs δημιουργούνται με πολλαπλούς και διαφορετικούς τρόπους. Έτσι, τα miRNAs μπορεί να προέρχονται από τον τυχαίο σχηματισμό hairpins σε "μη-κωδικοποιημένα" τμήματα του DNA, αλλά και από την επικάλυψη και την τροποποίηση των ήδη υφιστάμενων microRNAs. Μπορούν επίσης να σχηματιστούν από ανεστραμμένες επαναλήψεις αλληλουχιών κωδικοποιημένης πρωτεΐνης, η οποία επιτρέπει τη δημιουργία μιας αναδιπλωμένης φουρκέτας (foldback hairpin structure).

3.7 Ασθένειες και κλινικές εφαρμογές που συνδέονται με το miRNA

Βασική λειτουργία του miRNA λοιπόν, είναι η ρύθμιση της μεταγραφικής έκφρασης των γονιδίων. Κάθε miRNA μπορεί να ελέγξει μεταγραφικά την έκφραση ενός σημαντικού αριθμού γονιδίων, τα οποία στην περίπτωση των θηλαστικών μπορεί να φτάσει έως και αρκετές εκατοντάδες. Σε είδη όπως ο Homo sapiens, περίπου το 50% των μεταγραφών, υπολογίζεται ότι είναι υπό τη ρύθμιση miRNA. Την τελευταία δεκαετία, η ρυθμιστική λειτουργία των miRNAs έχει αποκαλυφθεί σε πολλούς μηχανισμούς, όπως στην ανάπτυξη και το πολλαπλασιασμό των κυττάρων, το μεταβολισμό και την απόπτωση.

Ωστόσο, το miRNA εμφανίζει άμεση εμπλοκή με ανθρώπινες ασθένειες. Η σύγχρονη έρευνα έχει αποκαλύψει ότι τα miRNAs εμφανίζουν βασικές ρυθμιστικές λειτουργίες σε πολλές ανθρώπινες ασθένειες, όπως καρκίνος, καρδιαγγειακές ασθένειες, μεταβολικές διαταραχές νευρολογικές παθολογίες, ψυχιατρικές διαταραχές, νεφρική νόσο, αυτοάνοσες ασθένειες, γαστρεντερικές συνθήκες, παχυσαρκία, διαταραχές της αναπαραγωγής, μυοσκελετικές διαταραχές και γενικά ιογενείς ασθένειες.

Όσον αφορά, τις κλινικές μελέτες, το miRNA φαίνεται να λειτουργεί ως βιοδείκτης για την εκτίμηση της εξέλιξης μιας νόσου. Ταυτόχρονα, άλλες μελέτες έχουν επικεντρωθεί στη διερεύνηση της χρήσης των miRNAs ως δείκτες ανταπόκρισης φαρμάκων ωστόσο, οι πιο ελκυστικές προοπτικές περιλαμβάνουν την άμεση χρήση των miRNAs για την ανάπτυξη θεραπευτικών στρατηγικών. Μέρα με τη μέρα, καθώς η έρευνα ανακαλύπτει νέους μηχανισμούς του RNA, προσδιορίζεται η ακριβής επίπτωση των miRNAs σε πολλές παθοφυσιολογικές λειτουργίες, ξεπερνώντας τα εμπόδια για την αποτελεσματική διαχείρισή τους.



Σχήμα 12: Role of miRNA in a cancer cell

4 Κεφάλαιο 4ο: MicroRNA Target Prediction

4.1 Online databases and tools

Οι online βάσεις δεδομένων περιλαμβάνουν ευρύτερες εφαρμογές της έκφρασης των miRNAs. Αυτοί οι πόροι αποτελούν θεμελιώδη στοιχεία για κάθε ερευνητική διαδικασία, όπως για παράδειγμα, στην αναζήτηση γνωστών miRNAs για κάθε οργανισμό ή ακόμα στην αναζήτηση γονιδιωματικών περιοχών σε ιστούς και κυτταρικές περιοχές. Η πιο ευρέως χρησιμοποιούμενη βάση δεδομένων σε αυτή την κατηγορία είναι miRBase (<http://www.mirbase.org>). Περιέχει πάνω 18.226 miRNA εγγραφές, 21.643 mature miRNA ακολουθίες από 168 είδη, ενώ οι καταγραφές για το πρόδρομο και ώριμο miRNA στον άνθρωπο αγγίζει τις 1.527 και 1.929 αντίστοιχα.

4.2 Conservation

Η έννοια της διατήρησης του microRNA στα διάφορα είδη οργανισμών αποτελεί μία βασική απαίτηση. Αφού, εφόσον οι θέσεις που δένουν τα microRNA διατηρούνται μεταξύ των διαφόρων ειδών, είναι πιθανό να είναι και βιολογικά λειτουργικές. Η πρόβλεψη βάσει των συντηρημένων ακολουθιών μειώνει σημαντικά, το ποσοστό ψευδώς θετικών αποτελεσμάτων. Ωστόσο, τα διάφορα προγράμματα πρόβλεψης χρησιμοποιούν μερικές φορές ελαφρώς διαφορετικούς ορισμούς διατήρησης.

Οι θέσεις που συνήθως θεωρούνται ως διατηρούμενες, εμφανίζονται ακριβώς στην ίδια θέση στην 3'UTR περιοχή κάθε οργανισμού. Παρόλα αυτά, περιοχές που θεωρούνται διατηρημένες μπορούν επίσης να βρεθούν και σε ακολουθίες χωρίς πλήρη ευθυγράμμιση. Επίσης, είναι πιθανόν, κάποιες πιθανές θέσεις να αλλάζουν στους διάφορους οργανισμούς. Αυτές οι περιοχές θεωρούνται ως poorly conserved.

4.3 Thermodynamics

Μια άλλη προσέγγιση για την αναγνώριση των miRNA στόχων είναι η εξέταση της θερμοδυναμικής σταθερότητας. Μια συχνά χρησιμοποιούμενη παράμετρος είναι η ενέργεια που απελευθερώνεται ανάμεσα στο miRNA και το γονίδιο-στόχο. Όσο χαμηλότερη είναι η ελεύθερη ενέργεια που απελευθερώνεται μεταξύ δύο ζευγών RNA, τόσο περισσότερη ενέργεια απαιτείται για να διασπαστεί αυτός ο διπλός δεσμός.

Έτσι, ένα δίκλωνο RNA είναι σε θερμοδυναμικά περισσότερη σταθερότητα, το οποίο σημαίνει ότι η σύνδεση του miRNA με το mRNA είναι ισχυρότερη, όταν η ελεύθερη ενέργεια είναι χαμηλή. Κατά συνέπεια, ένα miRNA έχει μεγαλύτερες πιθανότητες να δέσει με το mRNA, όταν απελευθερώνεται χαμηλή ενέργεια.

4.4 Sensitivity and Specificity

Τα διάφορα προγράμματα πρόβλεψης στόχου προσεγγίζουν με διαφορετικό τρόπο τα δεδομένα στοιχεία. Οι διαφορές στα αποτελέσματα ανάμεσα στους διαθέσιμους αλγόριθμους πρόβλεψης εξαρτάται από το πόσο “αυστηρά” είναι τα κριτήρια ταιριάσματος μεταξύ miPNA και μPNA στόχου.

Επομένως, είναι δύσκολο να αποφασιστεί ποιες προβλεπόμενες αλληλεπιδράσεις miRNA-στόχων είναι πιο πιθανό να είναι ακριβής και ποια προγράμματα παρέχουν την καλύτερη δυνατή απόδοση.

Δύο βασικά στοιχεία που ελέγχονται είναι η ακρίβεια και η ευαισθησία που παρουσιάζουν οι διάφορες αλληλεπιδράσεις. Πιο συγκεκριμένα έχουμε:

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{false negatives}}$$

όπου το True positives είναι ο αριθμός των προβλεπόμενων αλληλεπιδράσεων που πραγματικά υπάρχουν, σε αντίθεση με το False negatives που είναι ο αριθμός των αλληλεπιδράσεων που υπάρχουν αλλά δεν είχαν προβλεφθεί.

Αντίθετα, το specificity ορίζεται ως εξής:

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{false positives}}$$

όπου το True negatives είναι ο αριθμός από τις μη-υπάρχουσες αλληλεπιδράσεις miRNA-στόχου που σωστά δεν ήταν μέρος των προβλέψεων, ενώ το False positives είναι ο αριθμός των αλληλεπιδράσεων που είχαν εσφαλμένα προβλεφθεί, αλλά στην πραγματικότητα δεν υπάρχουν.

Συνοπτικά, για την επίτευξη ενός καλού εργαλείου πρόβλεψης στόχου, τόσο η ακρίβεια όσο και η ευαισθησία πρέπει να μεγιστοποιηθούν. Πράγμα που σημαίνει, ότι η πρόβλεψη δεν πρέπει να περιέχει false positives και όσο το δυνατόν λιγότερα false negatives.

4.5 Βασικές έννοιες για την ανάπτυξη αλγορίθμων στο τομέα της Βιοπληροφορικής

Για την ανάλυση της λειτουργίας του miRNA, ένας μεγάλος αριθμός από μελέτες έχουν δημοσιευθεί με σκοπό την επικύρωση των miRNAs. Παρά τη σημαντική συμβολή των τεχνικών υψηλής απόδοσης, η πλειονότητα των στόχων miRNA, ακόμη και για οργανισμούς που έχουν μελετηθεί αναλυτικά, όπως ο *Homo sapiens* ή *Mus musculus*, παραμένουν ακόμα άγνωστοι. Η *in silico* μελέτη (μέσω προσομοίωσης σε υπολογιστή) αποτελεί ένα οικονομικό και αποδοτικό μέσο για την αναγνώριση πιθανών στόχων. Οι πρώτοι αλγόριθμοι βασίστηκαν σε απλούς κανόνες ταιριάσματος βάσεων (base-pairing) με αρνητικό όμως αποτέλεσμα τη μειωμένη ακρίβεια και τον υψηλό αριθμό των ψευδώς θετικών αποτελεσμάτων. Οι διάφοροι αλγόριθμοι πρόβλεψης εξελίχτηκαν μαζί με την ανακάλυψη όλο και νέων miRNAs. Τα βιολογικά πειράματα, σε συνδυασμό με την υπολογιστική ανάλυση αποτελούν τη βάση για όλους τους αλγόριθμους πρόβλεψης.

Οι διαφορές στα αποτελέσματα ανάμεσα στους διαθέσιμους αλγόριθμους πρόβλεψης εξαρτάται από το πόσο "αυστηρά" είναι τα κριτήρια ταιριάσματος μεταξύ miRNA και mRNA στόχου. Επίσης, οι διάφοροι αλγόριθμοι μπορεί να διαφέρουν ως προς το χειρισμό των miRNA στόχων. Παρόλα αυτά, ακόμα και οι καλύτεροι αλγόριθμοι αδυνατούν να αναγνωρίσουν έναν αρκετά μεγάλο αριθμό miRNAs αλληλεπιδράσεων. Επιπλέον, τα αποτελέσματα δεν μπορούν να χρησιμοποιηθούν άμεσα, αλλά πρέπει να επικυρωθούν πρώτα πειραματικά. Δύο βασικά επίσης χαρακτηριστικά που πρέπει να ερευνηθούν είναι τα επίπεδα ευαισθησίας και ακρίβειας πρόβλεψης (sensitivity, specificity).

4.6 Παρουσίαση των πιο γνωστών αλγορίθμων που υπάρχουν σήμερα

Τα τελευταία χρόνια, πάνω από είκοσι προγράμματα πρόβλεψης στόχων miRNA έχουν δημοσιευθεί. Τα περισσότερα από αυτά, βασίζονται κυρίως σε ευθυγράμμιση ακολουθίας της seed περιοχής του miRNA (2-7 νουκλεοτίδια του 5' άκρου του miRNA) προς το 3' UTR του υποψήφιου στόχου. Επιπλέον, το specificity των περιοχών αυτών, συνήθως αυξάνεται με την παρατήρηση της διατήρησης (conservation) στις περιοχές πρόσδεσης (binding sites).

Μερικοί από τους πιο γνωστούς αλγορίθμους πρόβλεψης, που μελετήθηκε η λειτουργία τους στην αρχή αυτής της εργασίας αποτελούν οι παρακάτω:

- **miRanda**

Ο δικτυακός τόπος microRNA.org είναι μια ολοκληρωμένη πηγή microRNA στόχων πρόβλεψης και προφίλ έκφρασης. Οι προβλέψεις βασίζονται στην εξέλιξη του αλγορίθμου Miranda που ενσωματώνει την τρέχουσα βιολογική γνώση με κανόνες και στόχους σχετικά με τη χρήση των microRNAs. Οι θέσεις στόχου που προβλέπεται από τον Miranda, προκύπτουν από το σκόρ με τη χρησιμοποίηση του μοντέλου mirSVR. Έχει εκπαιδευτεί για να ανιχνεύει συμφοραζόμενα χαρακτηριστικά της προβλεπόμενης miRNA και mRNA-στόχου.

Ο αλγόριθμος υπολογίζει ξεχωριστά κάθε σκορ μεταξύ πολλαπλών interactions για το ίδιο miRNA και γονίδιο (με διαφορετικές περιοχές δεσίματος στον mRNA-στόχο), και τελικά υπολογίζει και ένα συνολικό σκορ επιλέγοντας τελικά την μεγαλύτερη από τις παραπάνω τιμές.

- **TargetScan and TargetScanS**

Ο αλγόριθμος TargetScan αποτελεί πρώιμη έκδοση του TargetScanS. Ο TargetScan εξετάζει και το ταίριασμα με τις seed περιοχές και την ενέργεια που απελευθερώνεται κατά την αλληλεπίδραση. Ο TargetScan αναζητά σε πρώτη φάση ταίριασμα με την 3' UTR για να αποκλύσει περιοχές που δεν ανταποκρίνεται στο ταίριασμα. Αν υπάρχει ταίριασμα, έστω δηλαδή ότι υπάρχουν j seed matches, ο αλγόριθμος αυξάνει κάθε φορά το j έως ότου να βρεθεί μία αναντιστοιχία.

Το συνολικό σκορ υπολογίζεται με βάση τα ακόλουθα στοιχεία:

- site-type contribution
- 3' pairing contribution
- local AU contribution
- position contribution
- TA (target site abundance) contribution
- SPS (seed-pairing stability) contribution

Τελικά, το σκορ κάθε αλληλεπίδρασης miRNA- στόχου υπολογίζεται βάση ενός συνολικού σκορ. Κάθε οικογένεια miRNA ταξινομείται σε ένα κοινό «πλαίσιο» με κοινό σκορ την πιο ευνοϊκή τιμή. Αν και ένα μόνο miRNA επιλέγεται ως αντιπροσωπευτικό miRNA, τα υπόλοιπα της ίδιας οικογένειας miRNA προβλέπονται επίσης να στοχεύουν το ίδιο γονίδιο-στόχο.

- **DIANA-microT**

Ο αλγόριθμος DIANA-microT βασίζεται σε παραμέτρους που υπολογίζονται ξεχωριστά για κάθε miRNA, και για κάθε ξεχωριστό στοιχείο miRNA (MRE), ανάλογα με τα επίπεδα διατήρησης και διατήρησης. Το συνολικό προβλεπόμενο score ενός miRNA προκύπτει από την αλληλεπίδραση με το γονίδιο-στόχο και είναι το συντηρημένο και μη διατηρούμενο άθροισμα MREs (microRNA Elements) ενός γονιδίου. Αυτή η πληροφορία μπορεί εύκολα να ανακτηθεί στο φιλικό προς το χρήστη DIANA-microT web server, όπου τα αποτελέσματα πρόβλεψης οργανώνονται σε ομαδοποιημένες καρτέλες διαθέσιμων πληροφοριών.

Παρέχεται επίσης, ο λόγος σήματος ως προς θόρυβο (SNR) και το σκορ ακριβείας για κάθε αλληλεπίδραση miRNA και γονιδίου-στόχου που μπορεί να χρησιμοποιηθεί ως ένα χρήσιμο εργαλείο εκτίμησης της «ορθότητας» κάθε προβλεπόμενης miTG.

- **PITA**

Η βασική καινοτομία του PITA είναι το μοντέλο που χρησιμοποιεί για την αλληλεπίδραση miRNA-στόχου. Η αλληλεπίδραση αυτή, βασίζεται στην πειραματική παρατήρηση ότι σχηματίζεται μια ισχυρή δευτερογενή δομή από το ίδιο 3' UTR εμποδίζοντας τη σύνδεση του miRNA. Με βάση αυτή την παρατήρηση, ορίζεται ένα νέο θερμοδυναμικό μοντέλο για την αλληλεπίδραση miRNA-στόχος.

Η ενέργεια που απελευθερώνεται από την αλληλεπίδραση ορίζεται ως εξής:

$$\Delta\Delta G = \Delta G_{duplex} - \Delta G_{open}$$

όπου το ΔG_{duplex} είναι η ενέργεια του binding miRNA-στόχου, ενώ το ΔG_{open} είναι η απαιτούμενη ενέργεια για να καταστεί η περιοχή στόχος προσβάσιμος για miRNA πρόσδεση.

Το συνολικό σκορ ορίζεται ως: $T = \log(\sum_{i=1}^n e^{S_i})$

όπου το n συμβολίζει το σύνολο των υποψήφιων στόχων και το S_i το $\Delta\Delta G$ για κάθε i στόχο.

- **MultiMitar**

Αυτός ο αλγόριθμος βασίζεται στο μηχανισμό μηχανικής μάθησης πρόβλεψης miRNA - στόχου. Συχνά αποτυγχάνεται να επιτευχθεί ακρίβεια πρόβλεψης όσον αφορά τόσο την ευαισθησία και ειδικότητα. Επιπλέον, όλες οι αλγόριθμοι που βασίζονται στο μηχανισμό μηχανικής μάθησης αδυνατούν πολλές φορές να αναγνωρίσουν τις πραγματικά θετικές προβλέψεις. Ως εκ τούτου αυτού του είδους οι αλγόριθμοι πιθανόν να καταστούν αναξιόπιστοι σε βιολογικές μελέτες.

- **PicTar** Ο PicTar βασίζεται στον υπολογισμό ενός συνόλου από συγκεκριμένες miRNAs και 3' UTRs. Για τη δημιουργία του training database, ο PicTar καταγράφει πρώτα τις θέσεις που ικανοποιούν το seed match σε όλες τις 3' UTRs περιοχές. Δεύτερον, ελέγχει κατά πόσον αυτές οι περιοχές είναι τέλεια συντηρημένες ή όχι.

Ο αλγόριθμος βασίζεται στο Hidden Markov Model. Σύμφωνα με αυτό, χτίζεται ένα μοντέλο όπου διάφορα miRNAs μπορούν να δράσουν από κοινού για να καταστείλουν το ίδιο γονίδιο. Δεδομένου M μελών που αντιπροσωπεύουν το σύνολο των διαφορετικών miRNAs, ένας στόχος 3' UTR μπορεί να παραχθεί με τον ακόλουθο τρόπο :

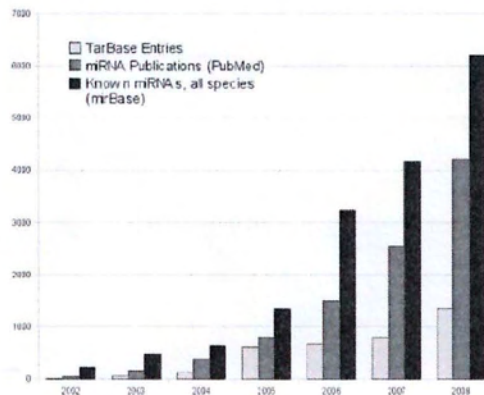
σε κάθε βήμα επιλέγεται μια πιθανότητα μετάβασης p_i για $i = 0$ έως M και μια συγκεκριμένη ακολουθία-στόχος. Με την επιλογή ενός miRNA προκύπτει ένα binding site, με απόλυτο ταιριασμα με τη seed περιοχή με πιθανότητα p ή ατελή με πιθανότητα $1-p$. Τελικά, το συνολικό σκορ προκύπτει από το μέσο όρο των πιθανοτήτων αλληλεπίδρασης των 3' UTRs στόχων με τα miRNAs.

5 Κεφάλαιο 5ο: Αλγόριθμοι και Μέθοδοι ανάπτυξης

Υπάρχει ένα πλήθος απο βάσεις δεδομένων που χρησιμοποιούνται ευρέως σήμερα. Εντούτις, στη παρούσα εργασία θα χρησιμοποιηθούν οι βάσεις δεδομένων Ensembl 69 και miRBase 19 για training και η Tarbase για testing.

Μια αλληλεπίδραση γονιδίου-στόχου miRNA προβλέπεται μόνο αν υπάρχει τουλάχιστον μία θέση δέσμευσης όπου αυτό το miRNA αλληλεπιδρά με τουλάχιστον μία από τις μεταγραφές του γονιδίου αυτού. Σε αυτή τη μελέτη, αναλύουμε την απόδοση τριών αλγορίθμων πρόβλεψης στόχου. Τα βιολογικά πειράματα και η υπολογιστική ανάλυση έχουν δώσει τα βασικά δομικά στοιχεία για όλους τους αλγόριθμους πρόβλεψης. Οι διαφορές στα αποτελέσματα που παρέχονται από τους διαθέσιμους αλγόριθμους πρόβλεψης εξαρτώνται από το πόσο "αυστηρά" είναι τα κριτήρια αντιστοίχισης μεταξύ miRNA και mRNA-στόχου. Οι υλοποιήσεις επίσης μπορεί να ποικίλλουν ανάλογα με το χειρισμό των περιοχών αυτών. (5' UTR , CDS , 3' UTR)

Οι αλγόριθμοι που εξετάστηκαν στη παρούσα διπλωματική είναι οι Miranda, TargetScan, DIANA-microT. Το κίνητρο για την επιλογή αυτών των τριών ως αντικείμενα για σύγκριση είναι ότι είναι οι πιο συχνά χρησιμοποιούμενοι αλγόριθμοι πρόβλεψης στόχου.



Σχήμα 13: Η ανάπτυξη των miRNA στην miRBase βάση δεδομένων(μαύρη μπάρα), η αύξηση των miRNA δημοσιεύσεων στο PubMed και η ανάπτυξη των πειραματικών miRNA στόχων στον άνθρωπο στην TarBase.

6 Κεφάλαιο 6ο: Οι Βάσεις Δεδομένων που χρησιμοποιήθηκαν

6.1 Η βάση δεδομένων miRBase

Στη βιοπληροφορική, η miRBase είναι μια βιολογική βάση δεδομένων που λειτουργεί ως ένα αρχείο των microRNA ακολουθιών. Από το Σεπτέμβριο του 2010, περιείχε πληροφορίες σχετικά 15.172 microRNAs. Το μητρώο miRBase παρέχει ένα κεντρικό σύστημα για την ανάθεση νέων ονομάτων στα microRNA γονίδια. Η miRBase έχει συσταθεί από τον Sam Griffiths-Jones το 2003. Η miRBase έχει τους εξής στόχους:

- Να παρέχει ένα συνεκτικό σύστημα ονοματοδοσίας για τα microRNAs.
- Να παρέχει μια κεντρική θέση συλλογή όλων των γνωστών αλληλουχιών microRNAs.
- Συγκέντρωση και σύνδεση των microRNAs στόχων των πληροφοριών τους.



Για τη παρούσα διπλωματική, χρησιμοποιήθηκε η έκδοση miRBase 19 και συγκεκριμένα αρχείο που περιέχει συνολικά 1.921 microRNAs για τον άνθρωπο. Παρακάτω, παρατίθεται κομμάτι του αρχείου.

```
>hsa-let-7a-2-3p MIMAT0010195 Homo sapiens let-7a-2-3p
CUGUACAGCCUCCUAGCUUUCC
>hsa-let-7b-5p MIMAT0000063 Homo sapiens let-7b-5p
UGAGGUAGUAGGUUGUGUGGUU
>hsa-let-7b-3p MIMAT0004482 Homo sapiens let-7b-3p
CUAUACAACCUACUGCCUUCCC
>hsa-let-7c MIMAT0000064 Homo sapiens let-7c
UGAGGUAGUAGGUUGUAUGGUU
>hsa-let-7d-5p MIMAT0000065 Homo sapiens let-7d-5p
AGAGGUAGUAGGUUGCAUAGUU
>hsa-let-7d-3p MIMAT0004484 Homo sapiens let-7d-3p
CUAUACGACCUGCUGCCUUUCU
>hsa-let-7e-5p MIMAT0000066 Homo sapiens let-7e-5p
UGAGGUAGGAGGUUGUAUAGUU
>hsa-let-7e-3p MIMAT0004485 Homo sapiens let-7e-3p
CUAUACGGCCUCCUAGCUUUCC
>hsa-let-7f-5p MIMAT0000067 Homo sapiens let-7f-5p
UGAGGUAGUAGAUUGUAUAGUU
>hsa-let-7f-1-3p MIMAT0004486 Homo sapiens let-7f-1-3p
CUAUACAUCUAUUGCCUUCCC
>hsa-let-7f-2-3p MIMAT0004487 Homo sapiens let-7f-2-3p
CUAUACAGUCUACUGUCUUUCC
```

Σχήμα 14: mirBase sample

όπου καταγράφεται σε κάθε γραμμή το microRNA ID, το accession number, το είδος του οργανισμού και το stem loop.

Τέλος, καταγράφεται και η ακολουθία.

6.2 Η βάση δεδομένων Ensembl 69

Το έργο Ensembl ξεκίνησε το 1999 , λίγο μετά το σχέδιο του ανθρώπινου γονιδιώματος. Ακόμη και σε αυτό το πρώιμο στάδιο , ήταν σαφές ότι η χειρονακτική επεξεργασία σε 3 δισεκατομμύρια ζεύγη βάσεων μιας αλληλουχίας δεν θα μπορούσε να προσφέρει στους ερευνητές έγκαιρη πρόσβαση στα τελευταία στοιχεία. Ο αρχικός στόχος της Ensembl λοιπόν ήταν να ενσωματωθούν σχόλια με άλλα διαθέσιμα βιολογικά δεδομένα και να είναι όλα αυτά διαθέσιμα στο κοινό μέσω του διαδικτύου. Μέτα την έναρξη λειτουργίας της ιστοσελίδας, τον Ιούλιο του 2000, έχουν έχουν προστεθεί πολλά γονιδιώματα στο Ensembl και το εύρος των διαθέσιμων δεδομένων έχει επεκταθεί συμπεριλαμβάνοντας συγκριτική γονιδιωματική, η διαφοροποίηση και ρύθμιση δεδομένων. Το έργο Ensembl διευθύνεται από τους Paul Flicek και Steve Searle.

Σήμερα, υπάρχουν καταγεγραμμένα γονίδια σε πάνω από 60 διαφορετικούς οργανισμούς. Για τη παρούσα διπλωματική χρησιμοποιήθηκε η Ensembl 69 που αφορά γονίδια για τον άνθρωπο. Περιέχει 19.414 γονίδια με διαφορετικά σημεία έναρξης και λήξης πρόσδεσης του microRNA.

Παρακάτω, παρουσιάζεται ένα κομμάτι του αρχείου:

```
>12|ENSG00000127720|METTL25|1|82832620|82832866|ENST00000550298
>4|ENSG00000109819|PPARGC1A|-1|23793644|23797444|ENST00000264867
>7|ENSG00000161057|PSMC2|1|103008502|103009842|ENST00000435765
>3|ENSG00000237787|C3orf79|1|153220272|153220486|ENST00000446603
>5|ENSG00000051596|THOC3|-1|175387222|175387961|ENST00000513482
>5|ENSG00000172244|C5orf34|-1|43486803|43487016|ENST00000306862
>6|ENSG00000189241|TSPYL1|-1|116597741|116599679|ENST00000368608
```

Σχήμα 15: Ensembl 69 sample

Το πρώτο στοιχείο κάθε γραμμής, αναφέρεται στο εκάστοτε χρωμόσωμα, το δεύτερο στην ονομασία που έχει δωθεί απο την Ensembl για το εκάστοτε ξεχωριστό γονίδιο με μοναδικό ξεχωριστό ENSG ID. Συγκεκριμένα, το αναγνωριστικό για τον άνθρωπο αρχίζει με «ENSG». Ένας κωδικός τριών γραμμάτων εισάγεται για άλλα είδη. (Για παράδειγμα, ENSMUSG είναι ένα γονίδιο ποντικού). Ο 11-ψήφιος αριθμός μετά από κάθε αναγνωριστικό είναι μοναδικός για κάθε γονίδιο.

Το τρίτο στοιχείο είναι η ονομασία του γονιδίου. Τα τρίτα επόμενα στοιχεία αναφέρονται στο strand, το utr start, utr stop. Το τελευταίο στοιχείο αναφέρεται σε όλες τις μεταγραφές για ένα γονίδιο (ENST, για τον άνθρωπο).

6.3 MREs και interaction miRNA-gene

- microRNA Recognition Elements

Αναφέρονται στα διαφορετικά σημεία πρόσδεσης του miRNA με το mRNA-στόχο. Υπάρχει ένα τουλάχιστον MREs για κάθε σχέση miRNA-gene.

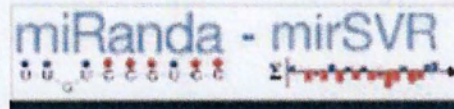
- interaction miRNA-gene.

Προκύπτει απο ένα ή περισσότερα MREs με διάφορες πιθανές περιοχές πρόσδεσης του miRNA πάνω στο γονίδιο.

Πράγματι, όπως φαίνεται και στο παρακάτω παράδειγμα, υπάρχουν 8 MREs και 3 miRNA-gene interactions.

| | | | | | |
|-----------------------|-----------|-----------|-----------------------|---|-------------------|
| chr5/ENSG00000164406 | 132210678 | 132210686 | hsa-miR-140-5p/LEAP2 | 1 | 0.393188182151969 |
| chr1/ENSG00000158864 | 161183762 | 161183770 | hsa-miR-140-5p/NDUFS2 | 1 | 0.472198960837444 |
| chr1/ENSG00000158864 | 161183868 | 161183876 | hsa-miR-140-5p/NDUFS2 | 1 | 0.472198960837444 |
| chr11/ENSG00000110719 | 67816755 | 67816763 | hsa-miR-140-5p/TCIRG1 | 1 | 0.412429082325869 |
| chr11/ENSG00000110719 | 67814872 | 67814880 | hsa-miR-140-5p/TCIRG1 | 1 | 0.412429082325869 |
| chr11/ENSG00000110719 | 67815758 | 67815766 | hsa-miR-140-5p/TCIRG1 | 1 | 0.412429082325869 |
| chr11/ENSG00000110719 | 67815309 | 67815317 | hsa-miR-140-5p/TCIRG1 | 1 | 0.412429082325869 |
| chr11/ENSG00000110719 | 67815523 | 67815531 | hsa-miR-140-5p/TCIRG1 | 1 | 0.412429082325869 |

6.4 Ο αλγόριθμος MiRanda



Αποτελεί ένα από τους παλαιότερους αλγόριθμους πρόβλεψης μεγάλης κλίμακας για σπονδυλωτά. Το [MicroRNA.org](http://www.mskcc.org/) είναι μία ιντερνετική πλατφόρμα που έχει δημιουργηθεί και διατηρείται από το cBio-the Computational Biology Center at Memorial Sloan-Kettering Cancer Center (<http://www.mskcc.org/>). Η ιστοσελίδα [microRNA.org](http://www.microRNA.org) είναι μια ολοκληρωμένη πηγή microRNA στόχων προβλέψης και προφίλ έκφρασης.

Παρέχονται οι εξής δυνατότητες:

- Εισαγωγή και αναζήτηση ενός συγκεκριμένου miRNA ανάμεσα σε διάφορους οργανισμούς.
- Με την εισαγωγή ενός mRNA και την προσθήκη ενός από τους 5 οργανισμούς (Homo sapiens, Mus musculus, Rattus norvegicus, Drosophila melanogaster, Caenorhabditis elegans) που παρέχει η εφαρμογή, εντοπίζονται πιθανοί στόχοι.
- Αναζήτηση ενός συγκεκριμένου ιστού σε έναν οργανισμό, επιτρέποντας την εξερεύνηση του επιπέδου έκφρασης όλων των miRNAs για ένα δεδομένο σύνολο των ιστών.

Η πρόβλεψη των στόχων βασίζεται στον αλγόριθμο miRanda. Ο αλγόριθμος πέρνει είσοδους από δύο αρχεία. Το 1ο περιέχει τις ακολουθίες των miRNAs, ενώ το 2ο του πιθανούς στόχους γονιδίων DNA/RNA.

Ο αλγόριθμος ακολουθεί τα εξής βήματα:

Σάρωση ενός ή περισσότερων αλληλουχιών miRNA έναντι όλων των αλληλουχιών που είναι πιθανές θέσεις στόχων. Οι πιθανές θέσεις-στόχοι προσδιορίζονται χρησιμοποιώντας δύο στάδια. Πρώτα μέσω δυναμικού προγραμματισμού γίνεται τοπική ευθυγράμμιση που πραγματοποιείται μεταξύ της αλληλουχίας miRNA και της αλληλουχίας αναφοράς. Η διαδικασία ευθυγράμμισης καταγράφει σκορ με βάση τη συμπληρωματικότητα της αλληλουχίας.

Για παράδειγμα:

```
Query: 3' gtCGAAAGTTTTACTAGAGTg 5' (eg. miRNA)
      |:||||| |||||:
Ref: 5' taGTTTTTCACAATGATCTCGg 3' (eg. 3'UTR)
```

Σε δεύτερη φάση ο αλγόριθμος κρατά μόνο τα σκορ με τιμή μεγαλύτερη από ένα κάτω όριο κατωφλίου. Αυτή η μέθοδος χρησιμοποιεί ρουτίνες από τη βιβλιοθήκη RNAlib, η οποία αποτελεί μέρος του πακέτου ViennaRNA γραμμένο από τον Ivo Hofacker.

Παρακάτω παρατίθεται κομμάτι της εξόδου του αλγόριθμου miranda:

Scores for this hit:

```
>gi|MIMAT0000062|emb|hsa-let-7a-5p|Homo |gi
|ENSG00000256574|gb|ENST00000536058|OR13A1 140.00 -17.93
2 21 466 487 19 73.68
```

Score for this Scan: Seq1.Seq2.Tot Score.Tot Energy.Max Score.Max Energy.Strand.Len1.Len2.Positions

```
>gi|MIMAT0000062|emb|hsa-let-7a-5p|Homo |gi
|ENSG00000256574|gb|ENST00000536058|OR13A1 140.00 -17.93
140.00 -17.93122 782 466 Complete
```

όπου φαίνεται το interaction μεταξύ ενός miRNA, συγκεκριμένα του hsa-let-7a-5p και του γονιδίου με id OR13A1. Το σκορ για το συγκεκριμένο interaction είναι 140.00

Στα πλαίσια της παρούσας διπλωματικής, έγινε εγκατάσταση της τελευταίας έκδοσης του αλγορίθμου miranda που παρέχεται στο site MicroRNA.org από τον Αύγουστο του 2010.

Ο αλγόριθμος miranda εντοπίζει συνολικά 9.747.981 MREs και 7.081.227 interactions gene-miRNA.

6.5 Ο αλγόριθμος Targetscan



Ο Targetscan είναι ένας απ' τους πρώτους αλγορίθμους που εμφάνισαν υψηλό sensitivity και specificity. Ο Targetscan προβλέπει τους miRNA στόχους βάσει του προσδιορισμού των "seed" περιοχών και της διατήρησης αυτών σε διάφορα είδη. Κάθε στόχος υπολογίζεται ανεξάρτητα των άλλων με τον υπολογισμό του τελικού context score. Το συνολικό score εξαρτάται από τα επίπεδα διατήρησης του miRNA στις περιοχές αυτές.

Ο Targetscan όπως και η πλειοψηφία των αλγορίθμων έχει εκπαιδευτεί για να εντοπίζει στόχους στην 3'UTR, εντούτις μελέτες έχουν δείξει ότι στόχοι βρίσκονται και σε CDS regions. Αποτελεί ένα online λογισμικό που παρέχεται από το πανεπιστήμιο του MIT για τη πρόβλεψη των miRNA στόχων. Η τελευταία έκδοση (Targetscan v.6.2), δημοσιευμένη τον Ιούνιο του 2012, αποτελεί μία βελτιωμένη και ενημερωμένη πηγή πληροφορίας με πάνω από 6.000 καταγεγραμμένες microRNA ακολουθίες, καθώς και πάνω από 700.000 ακολουθίες από UTR.

Το λογισμικό διατίθεται στη παρακάτω ιστοσελίδα <http://www.targetscan.org/>

Κατέβηξε η τελευταία έκδοση του αλγορίθμου targetscan6.2 που παρέχεται στο site <http://www.targetscan.org/> από τον Ιούνιο του 2012.

Ο αλγόριθμος targetscan εντοπίζει συνολικά 8.460.910 MREs και 6.340.066 interactions gene-miRNA..

Παρακάτω παρατίθεται κομμάτι της εξόδου του αλγορίθμου targetscan:

| Gene ID | Species ID | Mirbase ID | Site Type | UTR start | UTR end |
|-----------------|------------|--------------|-----------|-----------|---------|
| ENSG00000173876 | 9606 | hsa-miR-1179 | 7mer-m8 | 401 | 407 |
| ENSG00000173876 | 9606 | hsa-miR-1182 | 8mer-1a | 301 | 308 |
| ENSG00000173876 | 9606 | hsa-miR-1193 | 7mer-1a | 781 | 787 |
| ENSG00000173876 | 9606 | hsa-miR-1205 | 7mer-m8 | 50 | 56 |
| ENSG00000173876 | 9606 | hsa-miR-1206 | 7mer-m8 | 905 | 911 |

| 3' pairing contribution | local AU contribution | position contribution | context+ score |
|-------------------------|-----------------------|-----------------------|----------------|
| 0.003 | 0.073 | -0.005 | -0.049 |
| 0.003 | -0.005 | -0.038 | -0.287 |
| 0.004 | 0.026 | -0.027 | -0.071 |
| 0.003 | 0.070 | -0.055 | -0.102 |
| 0.012 | 0.010 | -0.049 | -0.147 |

Gene ID:ID γονιδίου

Species ID: για τον άνθρωπο 9606

Mirbase ID: από τη βάση δεδομένων που χρησιμοποιεί ο αλγόριθμος

UTR start: Σημείο που ξεκινά το binding του miRNA στη UTR περιοχή.

UTR start: Σημείο που σταματά το binding του miRNA στη UTR περιοχή.

Site Type: Τρεις διαφορετικές τιμές καταφλίου ανάλογα με το "δέσιμο" κάθε miRNA. 8mer ≥ 0.8 , 7mer-m8 ≥ 1.3 , 7mer-1A ≥ 1.6

3' pairing contribution: βρίσκει τη συμπληρωματικότητα μεταξύ miRNA-στόχου εκτός seed περιοχής.

position contribution: αντικατοπτρίζει την απόσταση από το πλησιέστερο UTR άκρο του στόχου.

local AU contribution: δείχνει τη μεταγραφή των 30 upstream και downstream νουκλεοτιδίων των προβλεπόμενων περιοχών.

Target-site abundance (TA) contribution

seed-pairing stability (SPS) contribution

Το context+score υπολογίζεται σύμφωνα με τα έξι τελευταία στοιχεία. Όσο πιο μικρή τιμή έχει το σκορ, τόσο πιο ευνοϊκή θέση έχει. Σε ένα γονίδιο με πολλαπλές θέσεις-στόχους της ίδιας οικογένειας miRNA, ένα συνολικό σκορ υπολογίζεται ως το άθροισμα του πιο αρνητικού-ευνοϊκού miRNA σε αυτή την οικογένεια.

6.6 Ο αλγόριθμος DIANA-microT

Το επίκεντρο του εργαστηρίου DIANA είναι η ανάπτυξη αλγορίθμων, βάσεων δεδομένων και εργαλείων για την ερμηνεία και την αρχειοθέτηση γονιδιωματικών δεδομένων. Ιδιαίτερη, έμφαση δίνεται στην ανάλυση των microRNA (miRNA) και στα γονίδια που κωδικοποιούν πρωτεΐνες. Ως εκ τούτου, οι δραστηριότητες του εργαστηρίου DIANA επικεντρώνεται στην ανάλυση της ρύθμισης της γονιδιακής έκφρασης καθώς και στην ερμηνεία του ρόλου των miRNAs σε διάφορες ασθένειες.

Τα διάφορα εργαλεία του DIANA σχετίζονται με αλγόριθμους πρόβλεψης στόχου (microT v4, microT-CDS), βάσεις δεδομένων επαληθεύοντας πειραματικούς miRNA στόχους για την κωδικοποίηση ή μη RNAs (TarBase, LncBase) καθώς και λογισμικό για την αναγνώριση ενδεχόμενων μονοπατιών έκφρασης ενός ή περισσότερων miRNAs (mirPath). Επιπλέον, ο πρόσφατα ανεπτυγμένος Web Server (v5.0) υποστηρίζει μια σειρά από εξελιγμένη ροή πληροφορίας που επιτρέπει σε χρήστες χωρίς απαραίτητη υποδομή βιοπληροφορικής να κατανοήσουν τις διάφορες λειτουργίες των miRNAs.



Σχήμα 16: Web applications.

Ο αλγόριθμος DIANA-microT-CDS (<http://www.microrna.gr/microT-CDS>) αποτελεί τη πέμπτη έκδοση του microT αλγορίθμου. Έχει εκπαιδευτεί για να ανυχνεύει στόχους τόσο στην 3'UTR περιοχή, όσο και στη CDS. Ο αλγόριθμος παρέχει σημαντικές βελτιώσεις ως προς το sensitivity σε σύγκριση με παλιότερες εκδόσεις (65% έναντι 52%).

Ο αλγόριθμος έχει ενημερωθεί σύμφωνα με την τελευταία έκδοση της miRBase (βάση δεδομένων που περιέχει το σύνολο των miRNA ακολουθιών). Παρέχει, επίσης, υπερσυνδέσμους σε on-line servers, όπως ο IHOP (Information Hyperlinked over Proteins) καθώς και επιπλέον πληροφορίες σε επιλεγμένα microRNAs σε ιστούς και κυτταρικές σειρές. Ο χρήστης μπορεί να εισάγει ένα microRNA ή ένα γονίδιο και να δει τις πιθανές αλληλεπιδράσεις. Παρακάτω, ακολουθεί μία αλληλεπίδραση του hsa-let-7a-5p miRNA με 987 πιθανούς στόχους και το threshold να έχει οριστεί 0.7.

| | Ensembl Gene Id | miRNA name | miTG score | Also Predicted |
|----|---------------------------|---------------|------------|--|
| 1 | ENSG00000156273 (BACH1) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 2 | ENSG00000187772 (LIN28B) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 3 | ENSG00000174498 (IGDCC3) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 4 | ENSG00000008323 (PLEKHG6) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 5 | ENSG00000135316 (SYNCRIP) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 6 | ENSG00000136603 (SKIL) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 7 | ENSG00000152894 (PTPRK) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 8 | ENSG00000075624 (ACTB) | hsa-let-7a-5p | 1.000 | <input checked="" type="checkbox"/> <input type="checkbox"/> ▼ |
| 9 | ENSG00000156466 (GDF6) | hsa-let-7a-5p | 1.000 | <input type="checkbox"/> <input checked="" type="checkbox"/> ▼ |
| 10 | ENSG00000198799 (LRIG2) | hsa-let-7a-5p | 1.000 | <input type="checkbox"/> <input checked="" type="checkbox"/> ▼ |
| 11 | ENSG00000180667 (YOD1) | hsa-let-7a-5p | 1.000 | <input type="checkbox"/> <input checked="" type="checkbox"/> ▼ |
| 12 | ENSG00000182263 (FIGN) | hsa-let-7a-5p | 1.000 | <input type="checkbox"/> <input checked="" type="checkbox"/> ▼ |
| 13 | ENSG00000159217 (IGF2BP1) | hsa-let-7a-5p | 1.000 | <input type="checkbox"/> <input checked="" type="checkbox"/> ▼ |

Σχήμα 17: Τα 13 πρώτα δείγματα. Το ροζ κουτί συμβολίζει ότι ο στόχος έχει επίσης βρεθεί και από τον αλγόριθμο miRanda, ενώ το μπλε από το targetscan.

7 Τελικό στάδιο εργασίας

7.1 Μορφοποίηση εισόδων

Τελικά, για την εξαγωγή αποτελεσμάτων χρησιμοποιήθηκαν οι τρεις αλγόριθμοι miRanda, targetscan, DIANA microT.

- **miranda**

Η πρόβλεψη γίνεται με τη χρήση του αλγορίθμου Miranda.

Score cut-off $S \geq 140$

Το σκορ χρησιμοποιεί το άθροισμα του «ταιριάσματος» των βάσεων μέσω δυναμικού προγραμματισμού.

Συγκεκριμένα:

A:U = 5

G:C = 5

G:U = 1 Άλλοι τύποι «ταιριάσματος» = -3

- **Targetscan**

Ο Targetscan είναι ένας από τους πρώτους αλγορίθμους που εμφάνισαν υψηλό sensitivity και specificity.

Αποτελεί ένα online λογισμικό που παρέχεται από το πανεπιστήμιο του MIT για τη πρόβλεψη των στόχων. Η τελευταία έκδοση (Targetscan v.6.2), δημοσιευμένη τον Ιούνιο του 2012, αποτελεί μία βελτιωμένη και ενημερωμένη πηγή πληροφορίας με πάνω από 6.000 καταγεγραμμένες microRNA ακολουθίες.

Το σκορ υπολογίζεται βάση του παρακάτω τύπου:

$$Z = \sum_{i=1}^i e^{-\frac{G_i}{T}}$$

όπου i : ο αριθμός των seed matches.

G_i : η ενέργεια που απελευθερώνεται μεταξύ του miRNA και του στόχου.

- **DIANA-microT**

Ο αλγόριθμος βασίζεται σε παραμέτρους που υπολογίζονται ξεχωριστά για κάθε miRNA, και για κάθε miRNA recognition element, ανάλογα με τα επίπεδα διατήρησης και διατήρησης. Το συνολικό σκορ ενός interaction miRNA: γονιδίου στόχου είναι το σταθμισμένο άθροισμα των συντηρημένων και μη ενός γονιδίου. Η πληροφορία αυτή, μπορεί να ανευρεθεί στο φιλικό προς το χρήστη DIANA microT web server, όπου τα αποτελέσματα πρόβλεψης οργανώνονται σε ομαδοποιημένες καρτέλες, μειώνοντας την πολυπλοκότητα αναζήτησης.

Τα βήματα που εκτελέστηκαν για την εξαγωγή των αποτελεσμάτων είναι τα ακόλουθα:

1. Τρέξιμο των αλγορίθμων με κοινές εισόδους τις βάσεις δεδομένων του training set. **Ensembl 69** και **mirBase release 19**.

2. Τα αποτελέσματα και των τριών αλγορίθμων επεξεργάστηκαν προκειμένου να εξαχθούν αυτά που χρειάζονταν για την συγκεκριμένη εργασία.

Στη προκειμένη περίπτωση η πληροφορία που χρειάστηκε αφορούσε τα εξής:

- **Chromosome**
- **Position on Chromosome**
- **MirbaseID**
- **GeneID**
- **strand**
- **score**

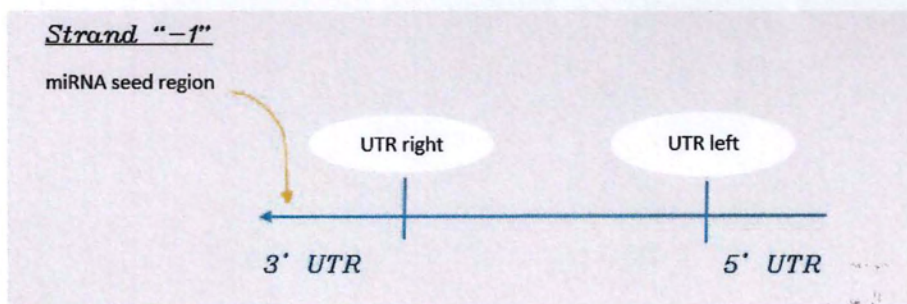
3. Και οι τρεις αλγόριθμοι, κατέγραφαν τη θέση του miRNA πάνω στη UTR ή CDs περιοχή(στη περίπτωση του microT αλγορίθμου).

4. Το τελικό αποτέλεσμα που μας ενδιαφέρει στη προκειμένη περίπτωση, είναι η θέση του miRNA πάνω στο χρωμόσωμα. Για το σκοπό αυτό, ιδιαίτερα σημαντική πληροφορία είναι το strand.

Η έννοια του strand

Στη μοριακή βιολογία αυτό που αποκαλούμε strand του DNA sense (positive +) σχετίζεται με τη δυνατότητα μεταγραφής της ίδιας ακολουθίας RNA σε πρωτεΐνη. Το συμπληρωματικό του ονομάζεται antisense (negative -). Το πρόσημο του strand, στη πραγματικότητα εξαρτάται από τη κατεύθυνση που γράφεται η ακολουθία που περιέχει τη πληροφορία για τη σύνθεση πρωτεϊνών. Έτσι, η βασική πληροφορία που είναι χρήσιμη στη προκειμένη περίπτωση, είναι η θέση των 5' και 3', διότι αυτά τα άκρα καθορίζουν την κατεύθυνση της μεταγραφής και μετάφρασης. Για παράδειγμα, η αλληλουχία 5' CGCTAT 3' είναι ισοδύναμη με την αλληλουχία 3' TATCGC 5'. Εάν τα άκρα δεν σημειώνονται, υποθέσουμε ότι η ακολουθία είναι γραμμένη στην 5' προς 3' κατεύθυνση.

Η κωδικοποίηση σε πρωτεΐνη μπορεί να μεταγράφει ομοίως και από τα δύο strands. Οι παρακάτω εικόνες δείχνουν γραφικά το binding του miRNA με το mRNA.



Σχήμα 19: Strand -1

Έτσι, σε κάθε αλγόριθμο, ο υπολογισμός της θέσης που “δένει” το miRNA με το χρωμόσωμα, υπολογίζεται ως εξής:

a) Strand +1:

Left position=UTR left + miRNA start

Right position=UTR left +miRNA end

b) Strand -1:

Left position=UTR right - miRNA start

Right position=UTR right - miRNA end

Έτσι, ο τελικός υπολογισμός της θέσης που δένει το miRNA πάνω στο χρωμόσωμα καθορίστηκε από το πρόσημο του strand και υπολογίστηκε σύμφωνα με τα παραπάνω.

7.2 Threshold

Το πρώτο βήμα για την εξαγωγή των αποτελεσμάτων, ήταν η καταγραφή των διάφορων αποτελεσμάτων με βάση το εκάστοτε threshold, το οποίο αυξάνονταν το ίδιο και στους τρεις αλγορίθμους.

Βρίσκοντας το χαμηλότερο και το υψηλότερο σκορ του συνόλου των αλληλεπιδράσεων, καταγράφονταν κάθε φορά τα εκάστοτε interactions των οποίων το σκορ έχει μεγαλύτερη τιμή από αυτήν της τιμής του κατωφλίου.

Παρακάτω, καταγράφονται οι πίνακες με τα συγκεντρωτικά αποτελέσματα. Η πρώτη στήλη καταγράφει το εκάστοτε threshold. Η δεύτερη στήλη απεικονίζει τον αριθμό των interactions για το σύνολο των MREs που βρίσκονται πάνω από το συγκεκριμένο όριο, ενώ η τρίτη στήλη, καταγράφει πόσα από αυτά τα Micro Recognition Elements υπάρχουν στην iarbase (βάση δεδομένων για την καταγραφή των testing αποτελεσμάτων). Η τέταρτη στήλη καταγράφει πόσα από αυτά τα δεδομένα είναι experimental validated. Αναλογικά οι επόμενες στήλες καταγράφουν τις αντίστοιχες μετρήσεις για gene-miRNA interaction.

A) microT results

| Threshold_microT | <u>MREs</u> | | <u>gene-miRNA interaction</u> | | |
|-------------------|-------------------|-------------------|--------------------------------|-------------------|-------------------|
| | Total_interaction | Validated_results | Experimental_validated_results | Total_interaction | Validated_results |
| 0.403462978042185 | 27.266.981 | 9.198 | 5.208 | 5.208 | 9.201 |
| 0.503710333616627 | 16.941.880 | 6.653 | 3.978 | 3.978 | 6.657 |
| 0.603957689191068 | 10.470.838 | 4.917 | 3.128 | 3.128 | 4.921 |
| 0.70420504476551 | 6.329.420 | 3.594 | 2.438 | 2.438 | 3.597 |
| 0.804452400339952 | 3.562.844 | 2.505 | 1.796 | 1.796 | 2.506 |
| 0.904699755914394 | 1.606.078 | 1.571 | 1.223 | 1.223 | 1.572 |

B) Miranda results

| Threshold_miranda | <u>MREs</u> | | <u>gene-miRNA interaction</u> | | |
|-------------------|-------------------|-------------------|--------------------------------|-------------------|-------------------|
| | Total_interaction | Validated_results | Experimental_validated_results | Total_interaction | Validated_results |
| 140.00 | 9.747.980 | 6.710 | 3.388 | 7.081.227 | 6.710 |
| 152 | 5.539.620 | 3.485 | 747 | 1.888.278 | 2.406 |
| 164 | 4.570.104 | 2.414 | 80 | 235.493 | 322 |
| 176 | 4.460.091 | 2.274 | 6 | 17.036 | 13 |
| 188 | 4.453.663 | 2.265 | 0 | 2.289 | 0 |
| 200 | 4.452.744 | 2.265 | 0 | 80 | 0 |

C) Targetscan results

| Threshold_targetscan | MREs | | | gene-miRNA interaction | |
|----------------------|-------------------|-------------------|--------------------------------|------------------------|-------------------|
| | Total_interaction | Validated_results | Experimental_validated_results | Total_interaction | Validated_results |
| -0.1855 | 6.344.809 | 5.833 | 2.601 | 5.419.486 | 5.107 |
| -0.0876 | 4.005.322 | 3.139 | 1.320 | 3.467.389 | 2.703 |
| 0.01029 | 1.458.535 | 1.146 | 312 | 1.018.625 | 705 |
| 0.1082 | 376.805 | 381 | 24 | 69.786 | 45 |
| 0.2061 | 126.256 | 128 | 1 | 443 | 0 |
| 0.304 | 54.316 | 74 | 0 | 5 | 0 |

8 TarBase v6.0

Η βάση δεδομένων DIANA-TarBase (<http://www.microrna.gr/tarbase>) αποτελεί μια ολοκληρωμένη βάση δεδομένων από miRNA στόχους σε είδη ζώων, φυτών και ιούς. Η τελευταία έκδοση (TarBase v6.0) είναι η μεγαλύτερη που διατίθεται σήμερα με περισσότερα από 65.000 miRNA αλληλεπιδράσεων γονιδίων. Η βάση δεδομένων περιλαμβάνει στόχους για 21 είδη που προέρχονται από πειράματα υψηλής απόδοσης. Ιδιαίτερη έμφαση έχει δοθεί στην ένταξη στόχων που προέρχονται από πειράματα όπως το HITS-CLIP και το PAR-CLIP. Είναι άψογα διασυνδεδεμένη με άλλα εργαλεία του DIANA-lab, όπως ο αλγόριθμος DIANA-microT. Είναι επίσης εξοπλισμένη με ισχυρά εργαλεία αναζήτησης και φιλτραρίσματος.

Η βάση δεδομένων παρέχει πειραματικά αλληλεπιδράσεις γονιδίων miRNA-στόχων. Ωστόσο, οι διάφοροι υπολογιστικοί αλγόριθμοι προβλέπουν υποθετικές θέσεις πρόσδεσης (binding). Αυτές οι αλληλεπιδράσεις (predicted or validated) πρέπει στη συνέχεια να συνδυαστούν με τους αλγόριθμους προκειμένου να συνδεθούν τα πειραματικά και υπολογιστικά αποτελέσματα από τα πειραματικά δεδομένα.

Ο web server χρησιμοποιεί για τη πρόβλεψη miRNA στόχων (σε CDs ή 3' UTR περιοχές) που παρέχονται από τον αλγόριθμο DIANA-microT πειραματικές ή

ακόμα και επικυρωμένες miRNA αλληλεπιδράσεις. Αυτές οι αλληλεπιδράσεις (προβλεπόμενες ή / και πειραματικά επικυρωμένες), μπορεί τη συνέχεια να συνδυαστούν με σκοπό τη μετά ανάλυση των αλγορίθμων.

Η TarBase v6.0 προσφέρει ένα ισχυρό σύστημα φιλτραρίσματος, το οποίο μπορεί να μειώσει τα αποτελέσματα με βάση: το είδος, τη πειραματική μέθοδο που χρησιμοποιείται, την μέθοδο επικύρωσης, τη πηγή δεδομένων, το έτος δημοσίευσης, το προβλεπόμενο σκορ αλληλεπίδρασης.

hsa-miR-1

Adv. options

Advanced options:

Select species:

- Caenorhabditis elegans
- Drosophila melanogaster
- Homo sapiens
- Mus musculus
- Arabidopsis thaliana
- Bombyx mori
- Bos taurus
- Danio rerio
- Epstein Barr virus
- Gallus gallus
- Glycine max
- HIV 1
- KSHV
- Medicago truncatula
- Oryza sativa
- Ovis aries
- Physcomitrella patens
- Rattus norvegicus
- Vitis vinifera
- Xenopus laevis
- Xenopus tropicalis

Apply Filter!

Select ver. method:

- Reporter Gene Assay
- Western Blot
- PCR
- Proteomics
- microarray
- Sequencing
- Other
- Degradome
- Northern Blot

Apply Filter!

Select regul. type:

- UP
- DOWN
- UNKNOWN

Apply Filter!

Select validation type:

- DIRECT
- INDIRECT
- UNKNOWN

Apply Filter!

Select those valid. as:

- POSITIVE
- NEGATIVE
- UNKNOWN

Apply Filter!

Select source:

- TarBase 6.0
- TarBase 5.0
- MiRecords
- miR2Disease
- miRTarBase

Apply Filter!

Select publ. year:

1900

Apply Filter!

Only publications published after the selected year will be presented.

Select pred. score thr.:

0

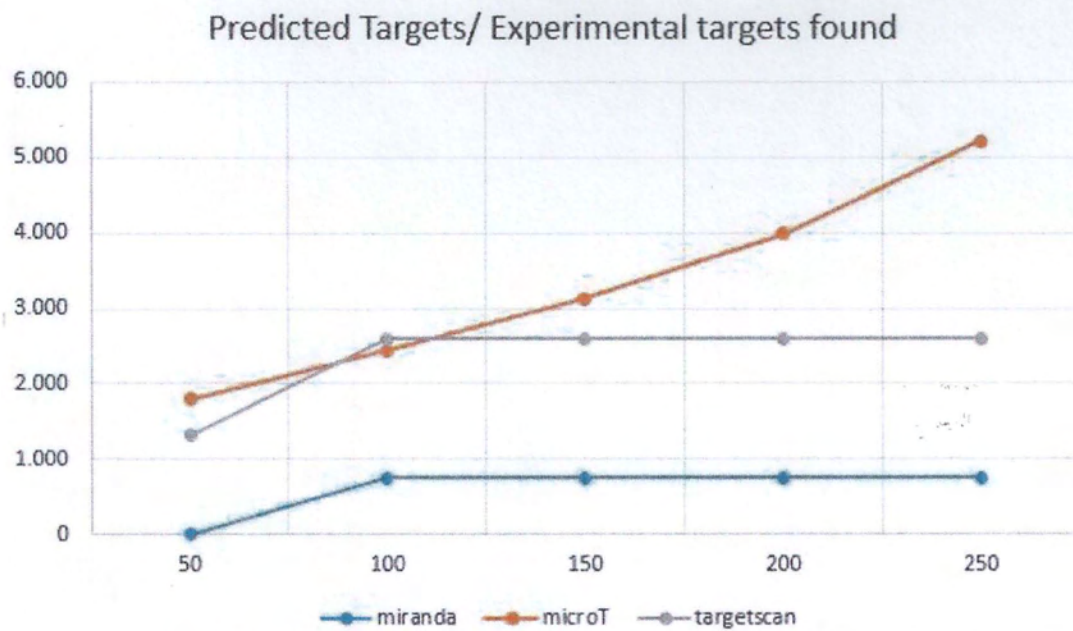
Apply Filter!

Only those having prediction score greater than the above will be presented.

Σχήμα 20

8.2 Final result

Τα πειραματικά δεδομένα που προέκυψαν, για τους τρεις αλγορίθμους απεικονίζονται στο παρακάτω διάγραμμα. Ο x άξονας απεικονίζει τις προβλεπόμενες αλληλεπιδράσεις κάθε αλγορίθμου (Predicted Targets) , ενώ ο y άξονας απεικονίζει τα experimental predicted targets που εντόπισε ο κάθε αλγόριθμος.



9 Επίλογος

Τα τελευταία χρόνια, η έρευνα γύρω από το miRNA έχει αποκτήσει νέα δυναμική. Νέα ερευνητικά δεδομένα, διερευνούν τη πιθανή εφαρμογή του miRNA σε κλινικές μελέτες, ως βιοδείκτης ή ως θεραπευτικό “εργαλείο”. Τα miRNAs έχουν αναγνωρισθεί ως βασικοί παράγοντες για την εξειδίκευση και την ευαισθησία της μεταγραφικής σίγασης των γονιδίων. Προσδιορίζοντας την αλληλεπίδραση των miRNAs με τα γονίδια-στόχους μπορούμε να εξάγουμε κρίσιμα συμπεράσματα για διάφορους ρυθμιστικούς μηχανισμούς, ζωτικής σημασίας για τη κατανόηση διάφορων βιολογικών διεργασιών με σκοπό την αποκρυπτογράφηση της παθοφυσιολογίας κάθε ασθένειας.

Με τον εντοπισμό των miRNAs και των στόχων τους, το ερευνητικό ενδιαφέρον για τον προσδιορισμό των λειτουργιών των miRNAs και ρυθμιστικών μηχανισμών τους, αυξάνεται συνεχώς. Τεχνολογίες υψηλής απόδοσης, έχουν συμβάλει σημαντικά στη σκιαγράφηση και στην αποκρυπτογράφηση καίριων βιολογικών λειτουργιών. Πρακτικά, όλες οι ερευνητικές εφαρμογές σήμερα, συνδέονται με τη χρήση υπολογιστικών εργαλείων και βάσεων δεδομένων.

Ένα σημαντικό μέρος αυτών των εφαρμογών έχει μεταφερθεί ή έχει σχεδιαστεί ειδικά για χρήση σε online εφαρμογές.

Τα διαθέσιμα εργαλεία παρέχουν εκτεταμένες δυνατότητες και λειτουργίες, ικανές για το χειρισμό διαφόρων σεναρίων ανάλυσης. Με την εξέλιξη των προγραμματιστικών μεθόδων, οι διεπαφές εφαρμογής έχουν αρχίσει να γίνονται όλο και πιο φιλικές και εύκολες στο χρήση. Επιπλέον, tutorials και blogs που σχετίζονται με τη βιοπληροφορική, καθώς και κοινωνικά δίκτυα για τους επιστήμονες, συμβάλλουν με τη σειρά τους προς αυτή τη κατεύθυνση.

1.evaluation_targetscan_mirbase.pl

```
#!/usr/bin/perl -w
use strict;

my $file=shift;
my $file1=shift;
my $file2=shift;

open (FINAL_FILE,'>miRNA_interaction_targetscan1.txt');
open (FINAL_FILE_1,'>validated_interaction_targetscan1.txt');
print FINAL_FILE "Threshold_Targetscan___Total_interaction___Validated_results\n";
#how to run: ./evaluation_targetscan_interaction_mirbase.pl final_targetscan.txt
human_mature-all.miRNAs.fa tarbase_human_interaction.txt

my %mres;
my $min=1000;
my $max=-1000;
my %same_gene_interaction;
warn "\t>Loading targetscan interactions\n";
open(IN,$file) or die "$!\n";
while(defined(my $line=<IN>)){
    chomp $line;

    if($line=~chr/){
        my @temp=split(/\\t/, $line);

        my $geneID=$temp[0];
        my $mirnaName=$temp[3];

        if(exists $mres{$geneID}{$mirnaName}){
```

```

        $same_gene_interaction{$geneID}{$mirnaName}++;
    }

    else{
        $same_gene_interaction{$geneID}{$mirnaName}=0;
    }

    #push (@{$mres{$geneID}{$mirnaName}}, "$temp[1]\t$temp[2]\t$temp[4]");
    $mres{$geneID}{$mirnaName}+=$temp[5];

    if($temp[5]<$min){
        $min=$temp[5];
    }
    if($temp[5]>$max){
        $max=$temp[5];
    }
}

#ENSG00000205916 27040599 27040605 hsa-miR-320a 1 -0.075
#ENSG00000205916 27021272 27021278 hsa-miR-320a 1 -0.075
#ENSG00000205916 27052480 27052486 hsa-miR-320a 1 -0.075
}

close IN;

#####
#loading human_miRNA hash with miRNA key and value.
warn "\t>Loading miRNA info\n";
my (%mirnasNAME,%mirnasMIMAT);
open(IN,$file1) or die "$!\n";
while(my $line=<IN>){ #human_mature-all.miRNAs.fa
    chomp $line;

```

```

if($line=~>){
    my @temp=split(/s/, $line);
    $temp[0]=~/>(./);

    $mirnasMIMAT{$temp[1]}=$1;
    $mirnasNAME{$1}=$temp[1];
}

# >hsa-let-7a-5p MIMAT0000062 Homo sapiens let-7a-5p
# UGAGGUAGUAGGUUGUAUAGUU
}
close IN;

#open tarbase_human_interaction.txt
#loading hash tarbase with 1
warn "\t>Loading tarbase interactions\n";
my %tarbase;
my $tarbaseInteraction=0;
open(IN,$file2) or die "$!\n";
while(my $line=<IN>){
    chomp $line;

    my @temp=split(/ |/, $line);
    $tarbase{$temp[1]}{$temp[0]}=1;
    $tarbaseInteraction++;

# MIMAT0000416 | ENSG00000101158 | Proteomics | Computational | 57569740_57569768
# MIMAT0000726 | ENSG00000146729 | microarray | Experimental | 3UTR
#
MIMAT0000646 | ENSG00000008282 | Proteomics | Computational | 105732177_105732205@10
5731507_105731535
#
MIMAT0000416 | ENSG00000137801 | Proteomics | Computational | 39887837_39887865@3988
8473_39888501@39891060_39891088@39874096_39874124@39882813;39883392_39882824;39883
408

```

```

#
MIMAT0000069 | ENSG00000170876 | Proteomics | Computational | 14183652_14183680@1418
3980_14184008@14184453_14184481@14184503_14184531

# MIMAT0000069 | ENSG00000100141 | Proteomics | Experimental | 3UTR

# MIMAT0000416 | ENSG00000196923 | Proteomics | Experimental | 3UTR

}

close IN;

my $step=(abs($min)+$max)/10;
my ($totalInteractions,$each_Interactions);
my $verifiedInteractions=0;
warn "\t>Crossreferencing targetscan with tarbase\n";
for(my $i=$min;$i<=$max;$i+=$step){
$totalInteractions=0;
    my $verifiedInteractions=0;
    foreach my $ensgid (keys %mres){
        foreach my $mirnaName (keys %{$mres{$ensgid}}){

$totalInteractions=$same_gene_interaction{$ensgid}{$mirnaName};

            if($mres{$ensgid}{$mirnaName}>=$i){
                $totalInteractions+=$each_Interactions+1;
                # print "$mres{$ensgid}{$mirnaName}\n";
                my $mimat=$mirnasNAME{$mirnaName};
                if(exists $tarbase{$ensgid}{$mimat}){

                    $verifiedInteractions++;
                    print FINAL_FILE_1 "$mimat\t$ensgid\n";
                }
            }
        }
    }
}
}

```

```
print FINAL_FILE "$i\t$totalInteractions\t$verifiedInteractions\n";  
}
```

2.miranda_seed_region.pl

```
#!/usr/bin/perl -w
```

```
use strict;
```

```
my $file=shift;
```

```
my $file1=shift;
```

```
my $file2=shift;
```

```
open (FINAL_FILE,'>seed_region_miranda_final_project.txt');
```

```
#how to run: ./target_region_codes.pl final_miranda_v1.txt human_mature-all.miRNAs.fa  
final_tarbase_human_interaction.txt
```

```
my %mres;
```

```
warn "\t>Loading targetscan interactions\n";
```

```
open(IN,$file) or die "$!\n";
```

```
while(defined(my $line=<IN>)){
```

```
    chomp $line;
```

```
    if($line=~chr/){
```

```
        my @temp=split(/\t/,$line);
```

```
        $temp[0]=~/(.+)\.(.+)/;
```

```
        my $geneID=$2;
```

```
        $temp[3]=~/(.+)\.(.+)/;
```

```

        my $mirnaName=$1;
        $mres{$geneID}{$mirnaName}="$Stemp[1]_ $Stemp[2]";

    }

}

close IN;

#####

#loading human_miRNA hash with miRNA key and value.
warn "\t>Loading miRNA info\n";
my (%mirnasNAME,%mirnasMIMAT);
open(IN,$file1) or die "$!\n";
while(my $line=<IN>){ #human_mature-all.miRNAs.fa
    chomp $line;

    if($line=~>){
        my @temp=split(/\\s/, $line);
        $temp[0]=~/>(./);

        $mirnasMIMAT{$temp[1]}=$1;
        $mirnasNAME{$1}=$temp[1];
    }

#    >hsa-let-7a-5p MIMAT0000062 Homo sapiens let-7a-5p
#    UGAGGUAGUAGGUUGUAUAGUU
}
close IN;

#open final_tarbase_human_interaction.txt
#loading hash tarbase with start_stop chromosome
warn "\t>Loading tarbase interactions\n";
my %tarbase;

```

```

open(IN,$file2) or die "$!\n";
while(my $line=<IN>){
    chomp $line;

    my @temp=split(/\t/,$line);
    $tarbase{$temp[1]}{$temp[0]}=$temp[2];

# MIMAT0000074      ENSG00000136986      124027310_124027338
#MIMAT0000252      ENSG00000153187      245016202_245016230
#MIMAT0000090      ENSG00000152503      114460514_114460542
#MIMAT0000252      ENSG00000166225      69969919_69969947
#MIMAT0000074      ENSG00000117335      207968193_207968221
#MIMAT0000090      ENSG00000158691      28348444_28348472

}
close IN;

my $verifiedInteractions=0;
warn "\t>Crossreferencing targetscan with tarbase\n";

    foreach my $ensgid (keys %mres){
        foreach my $mirnaName (keys %{$mres{$ensgid}}){
            my $mimat=$mirnasNAME{$mirnaName};

            if(exists $tarbase{$ensgid}{$mimat}){
                $verifiedInteractions++;
                print FINAL_FILE "$verifiedInteractions\t$mimat\t$ensgid\n";
            }
        }
    }
}

```

3.interaction_gene-miRNA_microT_algorithm.pl

```
#!/usr/bin/perl -w
```

```
use strict;
```

```
my $file=shift;
```

```
my $file1=shift;
```

```
my $file2=shift;
```

```
open (FINAL_FILE,'>validated_interactions_miRNA-gene_microT.txt');
```

```
#how to run: ./interaction_gene-miRNA_microT_algorithm.pl  
interaction_gene_miRNA_microT.txt human_mature-all.miRNAs.fa  
tarbase_human_Experimental.txt
```

```
my %mres;
```

```
warn "\t>Loading interactions\n";
```

```
open(IN,$file) or die "$!\n";
```

```
while(defined(my $line=<IN>)){
```

```
    chomp $line;
```

```
    my @temp=split(/\s/, $line);
```

```
    my $geneID=$temp[0];
```

```
    my $mirnaName=$temp[1];
```

```
    $mres{$geneID}=$mirnaName;
```

```
#ENSG00000127720 hsa-miR-20a-5p
```

```
#ENSG00000256574 hsa-let-7a-5p
```

```
#ENSG00000109819 hsa-let-7a-2-3p
```



```

}

close IN;

#####

#loading human_miRNA hash with miRNA key and value.
warn "\t>Loading miRNA info\n";
my (%mirnasNAME,%mirnasMIMAT);
open(IN,$file1) or die "$!\n";
while(my $line=<IN>){ #human_mature-all.miRNAs.fa
    chomp $line;

    if($line=~>/){
        my @temp=split(/\s/,$line);
        $temp[0]=~/>(.+)/;

        $mirnasMIMAT{$temp[1]}=$1;
        $mirnasNAME{$1}=$temp[1];
    }

#    >hsa-let-7a-5p MIMAT0000062 Homo sapiens let-7a-5p
#    UGAGGUAGUAGGUUGUAUAGUU
}

close IN;

#open tarbase_human_interaction.txt
#loading hash tarbase with 1
warn "\t>Loading tarbase interactions\n";
my %tarbase;
my $tarbaseInteraction=0;
open(IN,$file2) or die "$!\n";
while(my $line=<IN>){

```

```

chomp $line;

my @temp=split(/\/,$line);
$tarbase{$temp[1]}{$temp[0]}=1;
$tarbaseInteraction++;

#MIMAT0004801 | ENSG00000141551 | Sequencing | NA | UNKNOWN
#MIMAT0000091 | ENSG00000151233 | Sequencing | NA | UNKNOWN
#MIMAT0000255 | ENSG00000172728 | Sequencing | Experimental | 33229824_33229852
#MIMAT0000080 | ENSG00000123091 | Sequencing | Experimental | 51738603_51738631
#MIMAT0000093 | ENSG00000130962 | Sequencing | NA | UNKNOWN

}
close IN;

warn "\t>Finding interaction gene miRNAs in testing database tarbase \n";

foreach my $ensgid (keys %mres){
    #foreach my $mirnaName (keys %{$mres{$ensgid}}){
        my $microName=$mres{$ensgid};
        my $mimat=$mirnasNAME{$microName};
        if(exists $tarbase{$ensgid}{$mimat}){

                print FINAL_FILE "$mimat\t$ensgid\n";
        }
    }
}

```

Ο δεύτερος κώδικας γράφτηκε από τον κ. Γεωργακίλα.

Βιβλιογραφία κειμένων

- [1] <http://en.wikipedia.org/wiki/MicroRNA>
- [2] <http://en.wikipedia.org/wiki/Bioinformatics>
- [3] <http://ebooks.edu.gr/2013/ebooks.php?course=DSGL-C112>
- [4] <http://www.microna.org/microna/home.do>
- [5] <http://www.targetscan.org/>
- [6] <http://diana.cslab.ece.ntua.gr/microT/>
- [7] <http://el.wikipedia.org/wiki/Perl>
- [8] <http://www.perl.org/>
- [9] <http://en.wikipedia.org/wiki/Intron>
- [10] <http://en.wikipedia.org/wiki/DNA>
- [11] <http://en.wikipedia.org/wiki/MiRBase>
- [12] <http://www.mirbase.org/>
- [13] <http://www.ensembl.org/>
- [14] http://www.isical.ac.in/~bioinfo_miu/multimtar-download.htm
- [15] Tarbase 6.0: Capturing the Exponential Growth of miRNA Targets with Experimental Support. Nucl. Acids Res.

T. Vergoulis, T.I Vlachos , P. Alexiou, G. Georgakilas, M. Maragkakis , Martin Reczko, S.Gerangelos, N.Koziris, T. Dalamagas, and A. G. Hatzigeorgiou

2012

[14] Lost in translation: an assessment and perspective for computational microRNA target identification

Panagiotis Alexiou, Manolis Maragkakis, Giorgos L. Papadopoulos, Martin Reczko and Artemis Hatzigeorgiou

REVIEW September 29, 2009; p. 1-4

Associate Editor: Jonathan Wren

[15] Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role

Dimitrios Kleftogiannis, Aigli Korfiati, Kostantinos Theofilatos, Spiros Likothanassis, Athanasios Tsakalidis, Seferina Mavroudi

Journal of Biomedical Informatics (journal homepage: www.elsevier.com/locate/yjbin) 2013, p.1-3, 5

[16] Online resources for miRNA analysis

Ioannis S Vlachos, Artemis Hatzigeorgiou

Corresponding Author's Institution: Biomedical Sciences Research Center 'Alexander Fleming'

Elsevier Editorial System for Clinical
October-2012, p.4-11 & 21-24

[17] DIANA-microT web server: elucidating microRNA functions through target prediction

M. Maragkakis, M. Reczko, V. A. Simossis, P. Alexiou, G. L. Papadopoulos, T. Dalamagas, G. Giannopoulos, G. Goumas, E. Koukis, K. Kourtis, T. Vergoulis, N. Koziris, T. Sellis, P. Tsanakas and A. G. Hatzigeorgiou

Nucleic Acids Research, 2009, Vol. 37, Web Server issue, published online 30 April 2009,

p.1-2

Βιβλιογραφία εικόνων

[1]

https://www.google.gr/search?q=bioinformatics&source=lnms&sa=X&ei=DZKDU7HsIaSN0AXFolB4&ved=0CAcQ_AUoAA&biw=2051&bih=1001&dpr=0.67

[2,3,4,5,6,7] <http://ebooks.edu.gr/modules/ebook/show.php/DSGL-C112/479/3165,12729/>

[8] http://en.wikipedia.org/wiki/Messenger_RNA

[9] <http://en.wikipedia.org/wiki/MicroRNA>

[10]

https://www.google.gr/search?q=seed+dna+types&source=lnms&tbm=isch&sa=X&ei=dZODU6KDKqbl0AXbooCgBA&ved=0CAYQ_AUoAQ&biw=2051&bih=1001&dpr=0.67

[11]

https://www.google.gr/search?q=mrna+mature+process&source=lnms&tbm=isch&sa=X&ei=rZODU7PdB4OV0AW4IYH4BQ&ved=0CAYQ_AUoAQ&biw=2051&bih=1001&dpr=0.67

[12] <http://en.wikipedia.org/wiki/MicroRNA>

[13]

https://www.google.gr/search?q=tarbase+pubmed&source=lnms&tbm=isch&sa=X&ei=LpiDU9msA8fdPbbGgYAK&ved=0CAcQ_AUoAg&biw=2051&bih=1001&dpr=0.67

[14] <http://www.mirbase.org/>

[15] <http://www.ensembl.org/index.html>

[16] <http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=site/page&view=software>

[17] http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mircoT_CDS/results&keywords=hsa-let-7a-5p&genes=&mirnas=hsa-let-7a-5p&descr=&threshold=0.7

[18] <http://diana.cslab.ece.ntua.gr/>

[19] [Strand description](#)

[20] [16] Online resources for miRNA analysis

Ioannis S Vlachos, Artemis Hatzigeorgiou

Corresponding Author's Institution: Biomedical Sciences Research Center 'Alexander Fleming'

Elsevier Editorial System for Clinical
October-2012, p. 49

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ



004000121439

