

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ (ΠΜΣ)
«ΜΕΘΟΔΟΛΟΓΙΑ ΒΙΟΪΑΤΡΙΚΗΣ ΈΡΕΥΝΑΣ, ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΚΛΙΝΙΚΗ
ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Διπλωματική Εργασία

Ανάλυση υπολογιστικών μεθόδων για
ανεύρεση στόχων μικρών RNA
(Analysis of computational methods for
finding miRNA targets)

Καναβού Άννα

**Επιβλέπουσα καθηγήτρια:
Χατζηγεωργίου Άρτεμις**

Περιεχόμενα

Εισαγωγή.....	2
DIANA-microT-CDS	4
mirMark	7
MBSTAR.....	10
PACCMIT/PACCMIT-CDS.....	14
Συσχέτιση των υπολογιστικών μεθόδων	17
Βιβλιογραφία	18

Εισαγωγή

Τα micro-RNA (miRNA) είναι μικρά ενδογενή μόρια που παίζουν καθοριστικό ρόλο στην ανάπτυξη και τις ασθένειες μέσω της μετα-μεταγραφικής ρύθμισης της γονιδιακής έκφρασης. Αρχικά ανακαλύφθηκαν το 1993 στο εργαστήριο του Victor Ambros ενώ μελετούσαν το γονίδιο lin-14. Την ίδια χρονική περίοδο ο Gary Ruvkun προσδιόρισε το πρώτο miRNA target. Από τότε, ένας αυξανόμενος αριθμός των miRNA έχουν εντοπιστεί και γνωρίζουμε πλήρως την αλληλουχία σε πάνω από 700 miRNA στους ανθρώπους, ενώ υπολογίζεται ότι ο εκτιμώμενος αριθμός των miRNA σε ένα ανθρώπινο γονιδίωμα είναι κάτι παραπάνω από χίλια.

Τα miRNA γονίδια δεν κωδικοποιούν κάποια πρωτεΐνη αλλά αντί αυτού παράγουν μικρά μόρια μεγέθους περίπου 22 νουκλεοτιδίων. Αυτά τα μόρια αλληλεπιδρούν με το mRNA και συγκεκριμένα βρίσκουν στόχους στις 3' UTR (untranslated region) και CDS (coding sequences) περιοχές άλλων γονιδίων και αλληλεπιδρούν με βάση την αρχή της συμπληρωματικότητας Watson-Crick. Ως αποτέλεσμα αυτής της πρόσδεσης έχουμε την καταστολή της πρωτεϊνικής έκφρασης του γονιδίου στόχου μέσω αναστολής της μετάφρασης ή μέσω της υποβάθμισης του mRNA.

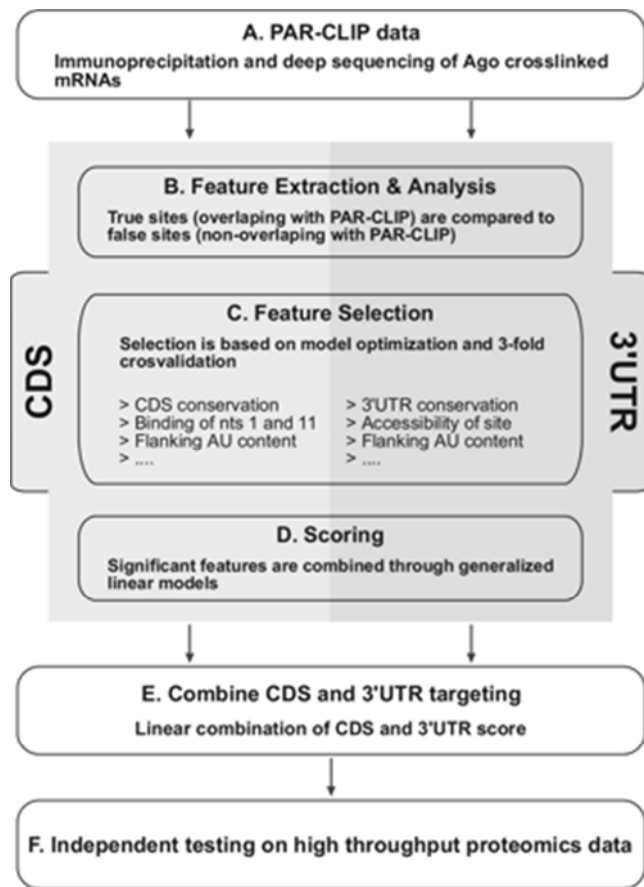
Έχει αποδειχθεί ότι τα miRNA εμπλέκονται σε ένα ευρύ φάσμα βιολογικών διεργασιών όπως έλεγχος του κυτταρικού κύκλου, απόπτωση και αρκετές αναπτυξιακές και φυσιολογικές διαδικασίες όπως η διαφοροποίηση των βλαστικών κυττάρων, η αιματοποίηση, η υποξία, καρδιακή και μυϊκή σκελετική ανάπτυξη, έκκριση ινσουλίνης, μεταβολισμό της χοληστερόλης, γήρανση και αντιγραφή του ιού. Εκτός από τους σημαντικούς ρόλους στα υγιή άτομα, τα miRNA έχουν επίσης ενοχοποιηθεί για έναν αριθμό ασθενειών που περιλαμβάνει ένα ευρύ φάσμα καρκίνων, καρδιαγγειακές παθήσεις, νευρολογικές ασθένειες, διαταραχές της νεφρικής λειτουργίας, πολυκαϊτιμία, διαβήτη, χρόνια ηπατίτιδα.

Όλα τα προηγούμενα μας οδηγούν στο πόσο αναγκαίο είναι να υπάρξει ανάπτυξη των εργαλείων βιοπληροφορικής πρόβλεψης των

αλληλεπιδράσεων του miRNA και των γονιδίων. Τα τελευταία χρόνια υπάρχει αλματώδη ανάπτυξη στον τομέα της μελέτης του miRNA και ο αριθμός των σχετικών επιστημονικών δημοσιεύσεων αυξάνεται με ταχύ ρυθμό. Στην παρούσα εργασία θα μελετήσουμε 4 υπολογιστικές μεθόδους για την ανεύρεση miRNA targets, DIANA microT-CDS, mirMark, MBSTAR και PACCMIT/PACCMIT-CDS.

DIANA-microT-CDS

Έχει αποδεχθεί από έρευνες πως το miRNA ελέγχει την έκφραση των γονιδίων τόσο στην περιοχή 3' UTR (untranslated region) όσο και στην περιοχή CDS (coding sequences). Παρ' όλα αυτά τα περισσότερα προγράμματα πρόβλεψης miRNA targets περιορίζουν την αναζήτησή τους για MREs(miRNA recognition elements) μόνο στο 3' UTR. Εδώ περιγράφουμε έναν αλγόριθμο για την πρόβλεψη των miRNA targets και στις δύο περιοχές (3' UTRs και CDS).



Η ανάλυση που πραγματοποιείται, γίνεται ανεξάρτητα για τα MREs στη περιοχή CDS και στην 3' UTR, και επιτρέπει την ταυτοποίηση των miRNA:mRNA χαρακτηριστικών δεσμών ειδικά για CDS ή UTRs. Για κάθε μία από αυτές τις περιοχές, ένα ξεχωριστό μοντέλο πρόβλεψης δημιουργείται και έπειτα τα μοντέλα συνδυάζονται για τον τελικό απολογισμό βαθμού αλληλεπίδρασης miRNA:γονιδίου.

Μέθοδοι που χρησιμοποιούνται:

Όσον αφορά τις συλλογές δεδομένων, χρησιμοποιούνται: PAR-CLIP data, Microarray data, Proteomics data. HITS-CLIP data, miRNA αλληλουχίες, πολλαπλές συστοιχίσεις και προβλέψεις miRNA targets από άλλα προγράμματα.

Όσον αφορά την εξαγωγή χαρακτηριστικών: alignment for putative MRE identification, primary analysis of PAR-CLIP data and training set construction, detection of binding categories with significant PAR-CLIP reads enrichment, conservation measure of the MRE sequence in CDS, conservation measure of the MRE sequence in 3'-UTR, detection of significantly accessible locations within MREs and other MRE features.

Έπειτα έχουμε την επιλογή των χαρακτηριστικών και στην συνέχεια την κατάρτιση και βαθμολόγηση όπου έχουμε την εξής λειτουργία: combining CDS and 3'-UTR targeting.

Το DIANA-microT-CDS εμφανίζει την υψηλότερη ευαισθησία σε οποιοδήποτε επίπεδο εξειδίκευσης σε σχέση με άλλα προγράμματα. Ένα μεγάλο μέρος νέων στόχων προβλέπονται μόνο από το συγκεκριμένο, και ειδικότερα στα χαμηλότερα επίπεδα ακρίβειας ο αριθμός των σωστών προβλέψεων έχει σχεδόν διπλασιαστεί χρησιμοποιώντας το DIANA-microT-CDS.

Τα υψηλής απόδοσης πρωτεομικά πειράματα που μετρούν τις αλλαγές χιλιάδων γονιδίων και στο επίπεδο miRNA και στο επίπεδο των πρωτεϊνών, αποκαλύπτουν ότι τα μισά από τα γονίδια που η έκφρασή τους αυξήθηκε/μειώθηκε μετά την επιμόλυνση/νοκ-άουτ του miRNA, δεν φέρουν ούτε ένα ανταποκρινόμενο miRNA seed match στην 3'-UTR περιοχή τους. Το πρόγραμμα αυτό παρουσιάζει το 12% αυτών των γονιδίων ως πρόσθετους στόχους των miRNAs, που έχουν τους στόχους τους σε περιοχή κωδικοποίησης. Η λίστα που περιέχει όλα τα γονίδια που προβλέπονται να έχουν στόχους μόνο στα CDS, έχει προβλεφθεί με ακρίβεια 50% και περιέχει κατά μέσο όρο 64 τέτοια γονίδια ανά miRNA.

Μία ανάλυση χαρακτηριστικών MREs σε 3'-UTRs αποκαλύπτει έναν αριθμό από νέα σημαντικά χαρακτηριστικά, όπως η απαίτηση για

αυξημένη προσβασιμότητα στη δευτερογενή δομή του miRNA κατά την έναρξη ενός MRE. Επίσης, η ανάλυση αποκαλύπτει ότι τα λειτουργικά MREs στα CDS απαιτούν έναν ισχυρότερο δεσμό από ότι τα MREs στα 3'-UTR. Συγκεκριμένα απαιτούν ένα τέλειο δέσιμο κατά μήκος του miRNA seed region και αναντιστοιχίες διαταράσσουν την λειτουργικότητά τους (Fang και Rajewsky (2001).

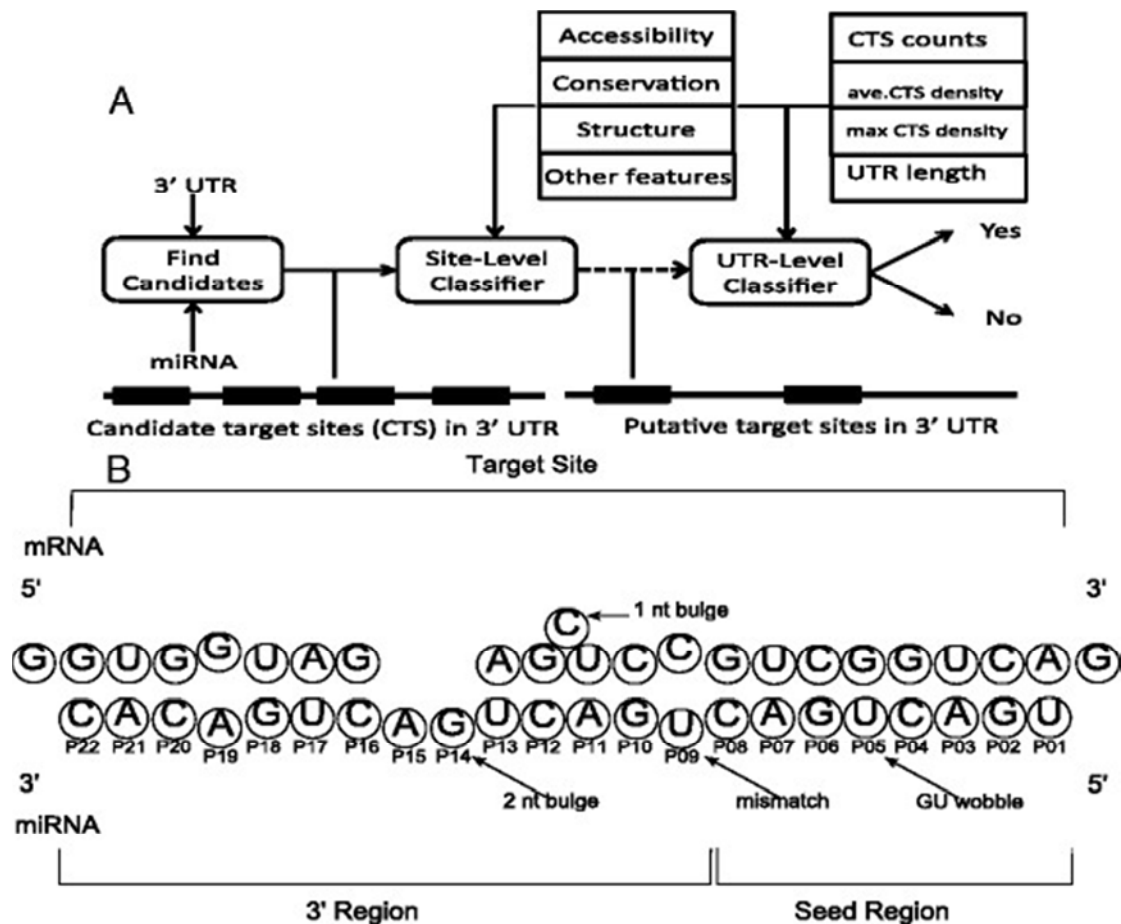
Για να έχουμε πρόσβαση στα αποτελέσματα του DIANA-microT-CDS, που είναι ο μοναδικός πόρος παροχής πρόβλεψης στόχων υπολογίζοντας επίσης και target sites στα CDS, και ακόμη περισσότερο προσδιορίζοντας τις προβλεπόμενες δεσμευτικές θέσεις όλων των περιοχών, μπορούμε να επισκεφθούμε το DIANA web server (Maragkakis et al., 2009) στο url www.microrna.gr/microT-CDS.

mirMark

Το mirMark είναι μια νέα μηχανή μάθησης με βάση τη μέθοδο πρόβλεψης miRNA target στην περιοχή και τα επίπεδα UTR. Αυτή η μέθοδος χρησιμοποιεί τα επαληθευμένα πειραματικώς miRNA targets από το miRecords και το mirTarBase (databases for miRNA targets).

Συνδυάζοντας την συσχέτιση με βάση την δυνατότητα της επιλογής με μια ποικιλία στατιστικών ή μηχανικών μεθόδων μάθησης για τους ταξινομητές των site και UTR επιπέδων, το mirMark βελτιώνει σημαντικά την συνολική προγνωστική απόδοση σε σύγκριση με τις υπάρχοντες μεθόδους.

Λαμβάνονται υπόψη μόνο τα 3' UTR targets sites λόγω της αφθονίας δεδομένων εκπαίδευσης στο 3' UTR. Η ταξινόμηση στο UTR επίπεδο χρησιμοποιείται αντί του gene-level. Ωστόσο, η μέθοδος είναι προσαρμόσιμη στην πρόβλεψη των target sites εντός CDSs, όταν όμως είναι διαθέσιμο ικανοποιητικό ποσό από δεδομένα εκπαίδευσης. Έχει ως στόχο να βελτιώσει την ικανότητα πρόβλεψης miRNA targets, τόσο στο site όσο και στο UTR level λαμβάνοντας υπόψη μια εκτενή λίστα με πάνω από 700 προγνωστικά χαρακτηριστικά και χρησιμοποιώντας την πιο πρόσφατη συλλογή των πειραματικά αποδεδειγμένων miRNA targets.



Στην δομή του το mirMark αποτελείται από δύο ταξινομητές, έναν για το site-level και έναν για το UTR-level, ανάλογα από τον επιθυμητό τύπο πρόβλεψης. Πρώτα εντοπίζονται τα CTSs (candidate target sites) του miRNA σε 3' UTR. Στη συνέχεια, η ευθυγράμμιση των CTSs και διάφορα άλλα χαρακτηριστικά που αφορούν την προσβασιμότητα, την διατήρηση και άλλες δομικές πληροφορίες, χρησιμοποιούνται για να βρεθεί το ισχυρότερο CTS. Από την άλλη πλευρά, ο ταξινομητής UTR-level ενσωματώνει το CTS για να προσδιοριστεί αν το γονίδιο είναι στόχος του δοθέντος miRNA. Στην παραπάνω εικόνα βλέπουμε μια απεικόνιση της σύνδεσης μεταξύ των miRNAs και τις περιφέρειες στόχου του 3' UTR σε επίπεδο site-level.

Η μέθοδος για την επιλογή των χαρακτηριστικών που χρησιμοποιείται, βρίσκει το πιο σχετικό, ακόμη λιγότερο περιττό, σύνολο χαρακτηριστικών γνωρισμάτων για την πρόβλεψη των site και UTR level. Διάφορες μέθοδοι στατιστικής ή μηχανικής μάθησης χρησιμοποιούνται για την ενσωμάτωση των επιλεγμένων χαρακτηριστικών και συγκρίνονται οι λειτουργίες τους.

Το mirMark είναι χτισμένο πάνω στην προσέγγιση μηχανικής μάθησης, έτσι τα αποτελέσματα του μοντέλου εξαρτώνται από τα δεδομένα εισόδου, όπως και όλα τα στατιστικά μοντέλα. Τα θετικά στοιχεία για το mirMark λαμβάνονται μέσω των συνδυασμένων αποτελεσμάτων από το miRecords και μια αυστηρή επιλογή από το miRTarBase (όπως αναφέρθηκε προηγουμένως). Αναπόφευκτα, τα miRNA και οι αλληλεπιδράσεις στόχων από τις βάσεις αυτές μπορεί απλώς να αντανακλούν τα miRNAs και γονίδια που παρουσιάζουν ενδιαφέρον για τους πειραματιστές που πραγματοποίησαν την επικύρωση, και μπορεί να μην είναι αντιπροσωπευτικά του τοπίου των αλληλεπιδράσεων στόχων γενικά. Επιπλέον, η επιλογή του αρνητικού συνόλου δεδομένων ενδεχομένως εισάγει μεροληψία στο μοντέλο. Το mirMark χρησιμοποιεί την mock προσέγγιση miRNA για τα αρνητικής γενιάς σύνολο δεδομένων. Mock miRNAs είναι αποτέλεσμα προσομοίωσης υπολογιστή και δεν βρίσκονται στη φύση, και σύμφωνα με τις τρέχουσες γνώσεις των miRNAs, είναι πιθανό ότι έχουν διαφορετικές αλληλουχίες και ιδιότητες σε σύγκριση με «πραγματικά αρνητικά miRNAs», τα οποία δεν γνωρίζουμε την πλήρη σειρά ακόμα. Επιπλέον, χρησιμοποιείται η miRanda για να βρεθούν υποψήφιοι target sites με αντιστοίχιση σπόρων σε τόσο θετικά όσο και αρνητικά σύνολα δεδομένων των mirMark, ως εκ τούτου, τα επιλεγμένα χαρακτηριστικά μπορεί να είναι προκατειλημμένα εναντίον της αντιστοίχισης σπόρων, αλλά ευνοούν και άλλα χαρακτηριστικά που δεν χρησιμοποιούνται από το φίλτράρισμα miRanda.

Αν και υπάρχουν τα πιθανά προβλήματα λόγω των δεδομένων εισόδου, οι κατασκευαστές του mirMark είναι αισιόδοξοι ότι η προσέγγιση μηχανικής μάθησης είναι η state-of-art μεθοδολογία για την πιο ακριβή πρόβλεψη miRNA targets. Η ποιότητα και η ποσότητα των δεδομένων εκπαίδευσης βελτιώνονται συνεχώς, καθώς όλο και περισσότερο τα δεδομένα αλληλεπίδρασης για τους miRNA-targets καταγράφονται από τις βάσεις δεδομένων, όπως miRecords και miRTarBase. Επίσης σχεδιάζουν να διατηρήσουν και να ενημερώνουν τακτικά το mirMark μόλις καταστούν διαθέσιμα νέα δεδομένα εκπαίδευσης. Επιπλέον, θα επεκτείνουν το mirMark από την πρόβλεψη του ανθρώπινου σε άλλα είδη, όπως ποντικού, στο εγγύς μέλλον.

MBSTAR

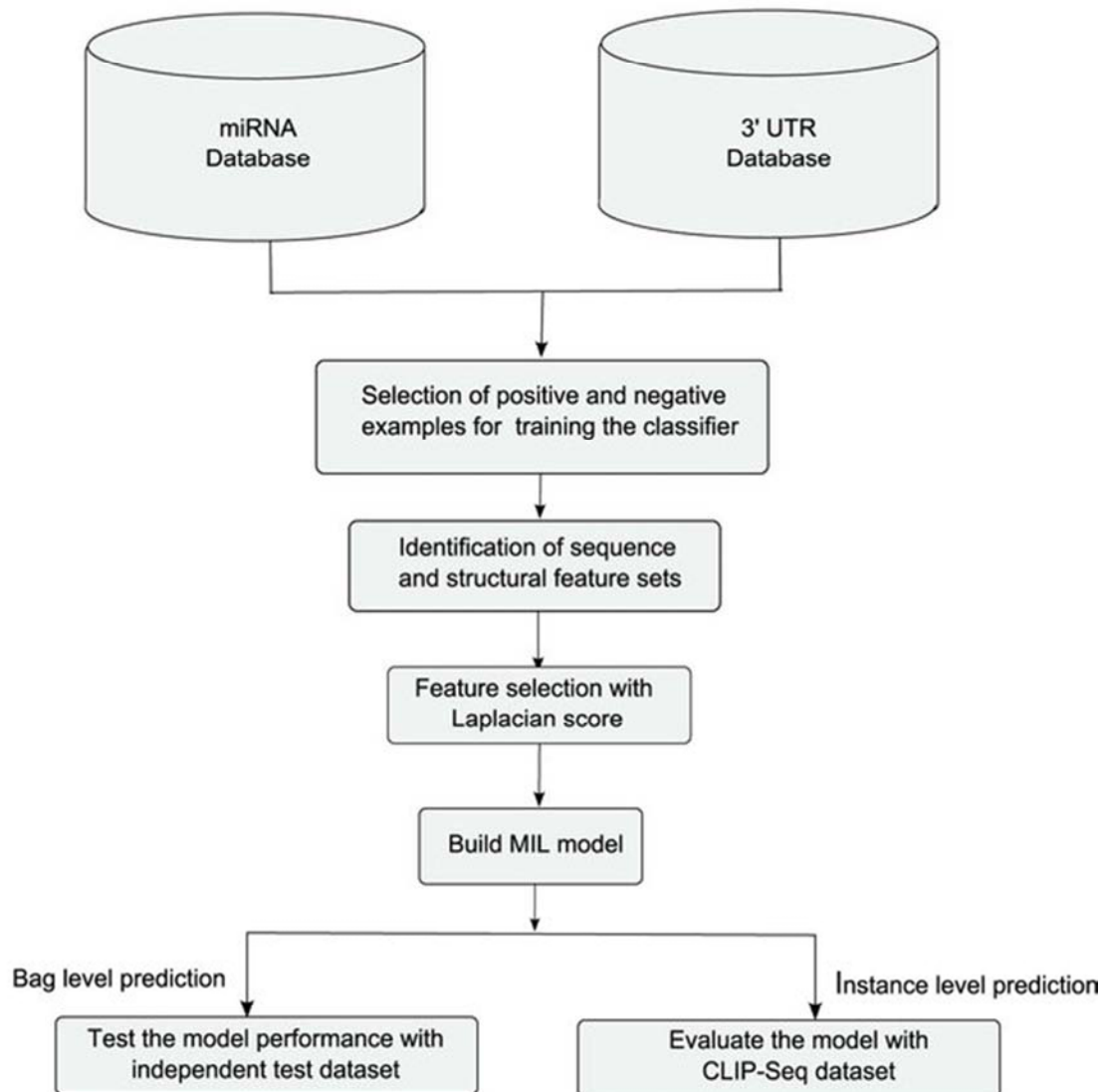
Το MBSTAR (Multiple instance learning of Binding Sites of miRNA TARgets) χρησιμοποιείται για την ακριβή πρόβλεψη των πραγματικών ή λειτουργικών miRNA θέσεων πρόσδεσης (binding sites). Το πολλαπλό πλαίσιο εκμάθησης υιοθετήθηκε για να χειριστεί την έλλειψη πληροφοριών για τις πραγματικές θέσεις σύνδεσης στα mRNAs targets. Έχουν βιολογικά επικυρωθεί 9531 αλληλεπιδρώντα και 973 μη-αλληλεπιδρώντα ζεύγη miRNA-mRNA προσδιοριζόμενα από την Tarbase 6.0 και επιβεβαιωμένα από το PAR-CLIP dataset. Αρκετοί μηχανισμοί μάθηση με στόχο την πρόβλεψη των miRNA targets έχουν αναπτυχθεί την τελευταία δεκαετία.

Το γενικό διάγραμμα που ακολουθείται έχει ως εξής:

- i. Για κάθε miRNA, εντόπισε τα PBSs (potential binding sites) από τον επικυρωμένο στόχο (θετικά) και μη-στόχο (αρνητικά) mRNAs που βασίζονται στη συμπληρωματικότητα seed site.
- ii. Απέσπασε χαρακτηριστικά από αυτά τα PBSs (ανεξάρτητα αν είναι λειτουργικά ή όχι).
- iii. Εκπαίδευσε έναν ταξινομητή να διακρίνει τους στόχους από τους μη-στόχους.
- iv. Για ένα άγνωστο ζευγάρι miRNA-mRNA, χρησιμοποίησε τον ταξινομητή για να το ονομάσει ως θετικό (στόχος) ή ως αρνητικό (μη-στόχος).

Όσον αφορά τα 2 πρώτα βήματα, ορισμένα από τα PBSs ενός πραγματικού στόχου θα μπορούσαν να είναι απρόσιτα για miRNA binding επειδή φράσσονται από δευτερογενή δομή των miRNA ή πρωτεϊνών που δεσμεύονται με RNA.

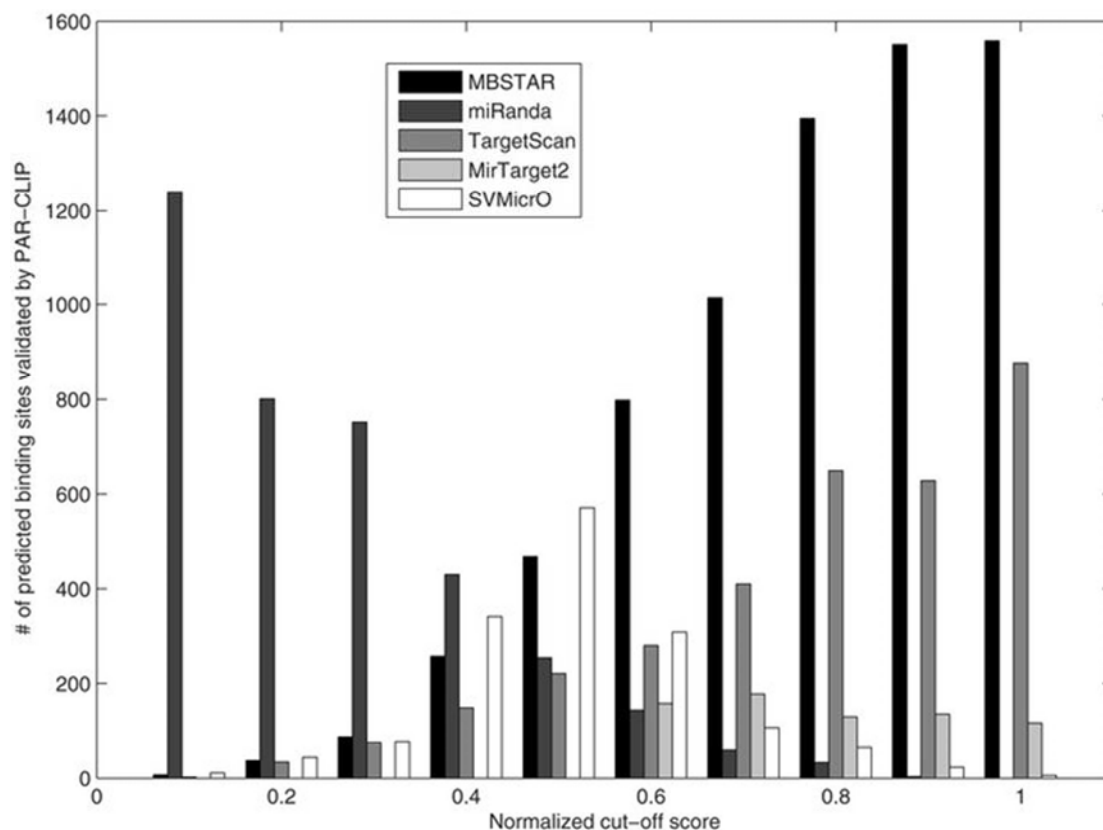
Process workflow of proposed method



Το MBSTAR έρχεται να δείξει μια εναλλακτική multiple instance learning framework για την πρόβλεψη των ειδικών λειτουργικών θέσεων πρόσδεσης των miRNAs. Η προτεινόμενη μέθοδος αντιμετωπίζει κάθε PBS για την αλληλεπίδραση miRNA-mRNA ως ξεχωριστή περίπτωση και χρησιμοποιεί multiple instance learning προσέγγιση για να ανακαλύψει τις λειτουργικές περιοχές. Και η ειδική αλληλουχία και τα διαρθρωτικά χαρακτηριστικά εξάγονται από το 3'-UTR για την εκπαίδευση του MIL μοντέλου. Η μέθοδος Laplacian (αξιολογεί τα χαρακτηριστικά σύμφωνα με την δύναμη διατήρησης θέσης) χρησιμοποιείται για τη λήψη των 40 υψηλότερων σχετικών χαρακτηριστικών. Η προκύπτουσα συλλογή δεδομένων

χρησιμοποιείται για την εκπαίδευση 6 αλγορίθμων MIL και συγκρίνονται οι επιδόσεις τους. Η απόδοση του MBSTAR μελετάται σε ένα ανεξάρτητο σύνολο δεδομένων που αποτελείται, όπως προείπαμε, από 9531 θετικές αλληλεπιδράσεις καθώς και από 973 αρνητικές. Το F-score που ισούται με 0,337 μας δείχνει ότι η μέθοδος αυτή είναι ικανή να πετύχει πρόβλεψη μεγάλης ακρίβειας ενώ προσδιορίζει 7156 FBSs (Functional Binding Sites) που συμπίπτουν με τις συσπειρώσεις PAR-CLIP. Επίσης, στον προσδιορισμό των targets και non-targets σε επίπεδο mRNA ξεπερνά άλλους αλγορίθμους όπως: TargetScan, miRanda, MirTarget και SVMicrO. Το Wilcoxon rank sum test δείχνει ότι το MBSTAR έχει μεγαλύτερο median score rank σε σχέση με άλλους αλγορίθμους. Επίσης, με normalized score cut-off το 0,5, το MBSTAR προβλέπει τα FBSs με ακρίβεια 38% ενώ το TargetScan επιτυγχάνει μόνο 18%.

Number of predicted sites verified by PAR-CLIP for MBSTAR, miRanda, MirTarget2, SVMicrO and TargetScan according to normalized cut-off score intervals:



Το MBSTAR βασιζόμενο στην λειτουργία του θα έχει πολύτιμη προσφορά στα μελλοντικά εργαστηριακά πειράματα για την ανακάλυψη των λειτουργικών miRNA binding sites. Ο αλγόριθμος μελετάει μόνο τα binding sites σε 3'-UTR και canonical binding sites. Πρόσφατες έρευνες έδειξαν ότι τα binding sites σε περιοχές κωδικοποίησης έχουν επίσης σημαντικές επιδράσεις ρύθμισης, αυτό θα διερευνηθεί στο μέλλον. Επίσης μαθαίνοντας από το PAR-CLIP dataset μέσα στο πλαίσιο MIL μπορεί να αυξηθεί η ακρίβεια των αποτελεσμάτων πρόβλεψης.

PACCMIT/PACCMIT-CDS

Το PACCMIT/PACCMIT-CDS είναι ένα web server που είναι διαθέσιμο στο κοινό στο url <http://pacccmit.epfl.ch>. Μια φιλική προς τον χρήστη διεπαφή σε 2 αλγορίθμους για την πρόβλεψη των mRNA που ρυθμίζονται από miRNA. Τον PACCMIT (Prediction of Accessible and/or Conserved MicroRNA Targets) που στοχεύει στις 3' UTR και είναι σχεδιασμένος να βρίσκει τα mRNA που απευθύνονται στις κωδικές αλληλουχίες (CDSs). Η κύρια συμβολή του PACCMIT είναι διπλή: παρέχει ένα ακριβές εργαλείο για την πρόβλεψη στόχων και των ασθενών συντηρημένων ή μη-συντηρημένων miRNAs, καθώς αντιμετωπίζει και την έλλειψη παρόμοιων portals που εξειδικεύονται για ανεύρεση στόχων στα CDSs.

Το PACCMIT είναι σχεδιασμένο για την αναγνώριση στόχων κυρίως σε περιοχές 3'UTR, ενώ το PACCMIT-CDS είναι προσαρμοσμένο ειδικά για κωδικοποίηση αλληλουχιών. Αν και υπάρχουν αρκετά web sites που παρέχουν πρόσβαση σε αλγορίθμους πρόβλεψης miRNA targets σε 3' UTR περιοχή, ο συγκεκριμένος web server για το PACCMIT συμπληρώνει την ανάγκη για ένα ακριβές εργαλείο για την πρόβλεψη targets των ασθενών συντηρημένων miRNAs, αντιστοίχως και ο web server για το PACCMIT-CDS βρίσκει τη θέση του ανάμεσα σε άλλα ακριβή εργαλεία που έχουν σχεδιαστεί ειδικά για την πρόβλεψη miRNA targets στην περιοχή κωδικοποίησης.

PACCMIT αλγόριθμος

Ο αρχικός αλγόριθμος κατατάσσει τις προβλέψεις σύμφωνα με την υπερ-εκπροσώπηση των συμπληρωματικών χώρων των miRNA seeds σε σχέση με ένα τυχαίο υπόβαθρο που βασίζεται στο μοντέλο Markov. Η ακρίβεια αυξάνεται περαιτέρω με την εξέταση μόνο διατηρημένων, εν μέρει προσιτών, ή συντηρημένων και μερικώς προσβάσιμων sites. Με αυτόν τον τρόπο βελτιώνει την ακρίβεια των προβλεπόμενων στόχων των ασθενών συντηρημένων miRNAs αλλά και βελτιώνει τις προβλέψεις των στόχων των εξαιρετικά συντηρημένων miRNAs.

PACCMIT-CDS αλγόριθμος

Αυτός ο αλγόριθμος βρίσκει πιθανούς στόχους miRNA στα CDS με την έρευνα για συντηρημένα μοτίβα συμπληρωματικά στη miRNA seed περιοχή και τους κατατάσσει σύμφωνα με την υπερ-εκπροσώπηση σε σχέση με ένα τυχαίο φόντο διατηρώντας τόσο την χρήση των κωδικονίων όσο και την αλληλουχία των αμινοξέων.

Η ακρίβεια και η ευαισθησία των 2 αλγορίθμων αξιολογήθηκαν με την κατασκευή επικυρωμένων data sets που είναι βασισμένα στις θέσεις πρόσδεσης που αναφέρονται στα πειράματα PAR-CLIP και στις αλλαγές στην πρωτεϊνική έκφραση από πρωτεομικά πειράματα.

Figure showing the main user interface of the PACCMIT/PACCMIT-CDS web server, divided into four panels (a, b, c, d) illustrating the workflow from algorithm selection to output specification.

- (a) Selection of the algorithm and database. The interface shows options for Algorithm (PACCMIT (3'UTR)), Use conservation (checked), Use accessibility ("loose" filter (higher sensitivity)), Genome (Human), Assembly (NCBI/hg18), and Track (Ensembl Genes).
- (b) Input of miRNAs of interest. The interface shows a field for "Accession numbers" with a "Browse" button and a text area containing miRNA accession numbers (e.g., X08710.1, X08710.2, X08710.3).
- (c) Input of mRNAs of interest. The interface shows a field for "Transcript IDs" with a "Browse" button and a text area containing transcript IDs (e.g., NM_001001, NM_001002, NM_001003, NM_001004).
- (d) Specification of the output parameters. The interface shows options for "Output specification", "Maximum number of pairs" (200), "FDR threshold" (0.05), and "Output file" (Excel 97-2003).

Figure:

Overview of the main user interface of the PACCMIT/PACCMIT-CDS web server. (a) Selection of the algorithm and database. (b) Input of miRNAs of interest. (c) Input of mRNAs of interest. (d) Specification of the output parameters.

Όσον αφορά το περιβάλλον του web server, είναι αρκετό φιλικό προς τον χρήστη. Ο τελευταίος ζητείται να επιλέξει έναν από τους 2 διαθέσιμους αλγορίθμους, είτε PACCMIT για 3' UTRs, είτε PACCMIT-CDS για μεταγραφές targets σε CDS. Επίσης είναι δυνατόν να αυξηθεί η

ακρίβεια εφαρμόζοντας φίλτρα για την διατήρηση (διαθέσιμο και στους 2 αλγορίθμους) ή/και προσβασιμότητα (διαθέσιμο μόνο για PACCMIT). Έπειτα ο χρήστης καθορίζει τα miRNAs και mRNAs που πρόκειται να αναλυθούν. Από αυτή την επιλογή ο web server εντοπίζει τα υποψήφια ζεύγη miRNA-target, τα κατατάσσει σύμφωνα με το P-value, αξιολογεί τη στατιστική σημασία τους σύμφωνα με το false discovery rate που βασίζεται στη διαδικασία Benjamini – Hochberg και διαδραστικά επιστρέφει μία λίστα ζεύγων miRNA/mRNA σε καθορισμένη μορφή. Είναι διαθέσιμη και η «Βοήθεια» του web server όπου υπάρχουν αναφορές στη σχετική βιβλιογραφία καθώς και μία σύντομη επεξήγηση των βασικών λειτουργιών των αλγορίθμων. Τέλος, στο τμήμα «Tutorial» προσφέρεται ένα μινιμαλιστικό παράδειγμα που δείχνει μια τυπική χρήση του web server.

Συσχέτιση των υπολογιστικών μεθόδων

Ένα μεγάλο ποσοστό ερευνών προτείνει ότι τα miRNA bindings προτιμούν να τοποθετούνται στη 3' UTR περιοχή των target genes και για αυτόν ακριβώς τον λόγο πολλά εργαλεία έχουν ειδικά σχεδιαστεί για να εντοπίζουν τέτοιου είδους θέσεις. Επίσης έχει αποδειχθεί ότι λειτουργικές θέσεις εντοπίζονται πλησίον των 2 άκρων του 3' UTR αλλά όχι πάρα πολύ κοντά στο κωδικόνιο λήξης. Τα εργαλεία mirMark και MBSTAR βασίζονται σε αυτά τα χαρακτηριστικά. Ειδικότερα, οι δημιουργοί του mirMark ήταν σε θέση να διακρίνουν αρκετά διακρίνοντας site-level και UTR – level συμπεριλαμβανόμενων των αντιστοιχιών, αναντιστοιχιών και εξογκώματα σε ορισμένες θέσεις της ακολουθίας miRNA καθώς και την απόσταση της θέσης δέσμευσης από την αρχή του UTR. Το MBSTAR, αντιθέτως, θεωρεί τον αριθμό και το μέγεθος των εξογκωμάτων ως συγκεκριμένα μοτίβα νουκλεοτιδίων που πλευρίζουν τις θέσεις δέσμευσης σαν σχετικά χαρακτηριστικά των λειτουργικών αλληλεπιδράσεων. Και οι δύο προηγούμενες μέθοδοι εστιάζουν κυρίως στον εντοπισμό στόχων στην περιοχή 3' UTR.

Όμως έχει αποδειχθεί ότι το miRNA ελέγχει την έκφραση των γονιδίων και στην περιοχή CDS (coding sequence). Οι υπολογιστικές μέθοδοι όπως η DIANA-microT-CDS και το PACCMIT-CDS έχουν ειδικά σχεδιαστεί για την πρόβλεψη binding sites σε αυτή την περιοχή. Ειδικά το PACCMIT-CDS βρίσκει υποψήφια binding sites στο CDS ψάχνοντας για συντηρημένα μοτίβα συμπληρωματικά στην αλληλουχία του miRNA, και αυτό αποτελεί μία πρόκληση λόγω της ισχυρής επιλεκτικής πίεσης που αποσκοπεί στη διατήρηση της αλληλουχίας αμινοξέων και χρήση του κωδικονίου. Η μέθοδος DIANA-microT-CDS, από την άλλη μεριά, βασίζεται σε προσεγγιστική μηχανή μάθησης που στηρίζεται στα δεδομένα CLIP για τον εντοπισμό των πιο σχετικών χαρακτηριστικών που σχετίζονται με τα miRNAs binding στην περιοχή CDS, στην περιοχή 3' UTR, ή και στις δύο.

Βιβλιογραφία

- Martin Reczko, Manolis Maragkakis, Panagiotis Alexiou, Ivo Grosse and Artemis G. Hatzigeorgiou (2012) Functional microRNA targets in protein coding sequences
- Mark Menor, Travers Ching, Xun Zhu, David Garmire and Lana X Garmire (2014) mirMark: a site level and UTR-level classifier for miRNA target prediction
- Miroslav Sulc, Ray M. Marin, Harlan S. Robins and Jiri Vanicek (2015), PACCMIT/PACCMIT-CDS: identifying microRNA targets in 3' UTRs and coding sequences
- Sanghamitra Bandyopadhyay, Dip Ghosh, Ramkrishna Mitra & Zhongming Zhao (2015) MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-233.
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 116, 281–297 (2004)
- Gaetano Santulli (2015) microRNA: Basic Science: From Molecular Biology to Clinical Practice
- Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P. & Stoffel, M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell*. DOI:10.1016/j.molcel.2014.03.045 (2014).
- Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 141, 129–141 (2010).
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 39, D152–D157 (2011).
- Sengupta, D. & Bandyopadhyay, S. Topological patterns in microRNA-gene regulatory network: studies in colorectal and breast cancer. *Mol. BioSyst*. 9, 1360–1371 (2013)

- Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, 120:15–20.
- Friedman RC, Farh KK, Burge CB, Bartel DP: Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009, 19:92–105.
- Hüntgen S, Siemens H, Kaller M, Hermeking H: The p53/microRNA Network in Cancer: Experimental and Bioinformatics Approaches. In *MicroRNA Cancer Regulation*. Dordrecht: Springer; 2013:77–101
- Wang KC, Garmire LX, Young A, Nguyen P, Trinh A, Subramaniam S, Wang N, Shyy JY, Li YS, Chien S: Role of microRNA-23b in flow-regulation of Rb phosphorylation and endothelial cell growth. *Proc Natl Acad Sci U S A* 2010, 107:3234–3239.