



Ο ρόλος των microRNA στον καρκίνο.

Μια υπολογιστική προσέγγιση

Παναγιώτης Βαρελάς
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Στην Επιστήμη

Περιεχόμενα

| | |
|---|----|
| 1.Εισαγωγή..... | 5 |
| 1.1 Περίληψη..... | 5 |
| 1.2 miRNAs | 5 |
| 1.3 Ιστορικά Στοιχεία | 6 |
| 1.4 Ονοματολογία Συμβολισμοί | 6 |
| 1.5 Φυσιολογική Βιογένεση miRNA..... | 7 |
| 1.6 Απορρύθμιση και ανωμαλίες στην Βιογένεση των miRNA..... | 8 |
| 1.7 SNPs..... | 9 |
| 1.8 miRNAs και SNPs..... | 9 |
| 2. WebServers..... | 11 |
| 2.1 MirSNP..... | 11 |
| 2.2 PolymiRTS Database 3.0..... | 15 |
| 2.3 miRdSNP..... | 19 |
| 2.4 dbSMR..... | 24 |
| 2.5 MiRvar..... | 25 |
| 2.6 Micro Sniper..... | 26 |
| 2.7 mrSNP..... | 29 |
| 2.8 dPore..... | 31 |
| 2.9 miRNASNP..... | 35 |
| 3.Αποτελέσματα..... | 38 |
| 3.1 Ποιοτικά Χαρακτηριστικά..... | 38 |
| 3.1.1 Έκδοση του miRBase..... | 39 |
| 3.1.2 Πηγή δεδομένων SNPs..... | 40 |
| 3.1.3 Genome Browser..... | 41 |
| 3.1.4 Πειραματική Υποστήριξη..... | 41 |
| 3.1.5 Αλγόριθμοι Πρόβλεψης..... | 41 |
| 3.1.6 Human Genome Diseases..... | 42 |
| 3.1.7 Αριθμός Ειδών..... | 42 |
| 3.1.8 Αλληλεπιδράσεις με ασθένειες..... | 43 |
| 3.1.9 Διαφορετικοί τρόποι αναζήτησης..... | 43 |
| 3.1.10 Ανακάλυψη Αλληλεπιδράσεων ανά WebServer..... | 44 |
| 3.2 Αριθμητικά Χαρακτηριστικά..... | 46 |
| 3.2.1 Αριθμός miRNAs..... | 46 |
| 3.2.2 Αριθμός miRNA target sites..... | 47 |
| 3.2.3 Αριθμός Γονιδίων..... | 47 |
| 3.2.4 miRNA-related SNPs..... | 47 |
| 3.2.5 Αριθμός SNPs στην 3'UTR..... | 47 |
| 3.2.6 Ενδεικτική αναζήτηση για το hsa-miR-55 σύγκριση..... | 48 |
| 4.Συμπεράσματα..... | 49 |
| 5.Ευχαριστίες..... | 50 |
| 6.Βιβλιογραφία..... | 51 |

Περιεχόμενα Πινάκων-Εικόνων-Γραφημάτων

Πίνακες:

Κεφαλαίο 2

- Πίνακας 2.1: Αποθηκευμένα δεδομένα στον MirSNP 13
Πίνακας 2.2: Επιβεβαιωμένες Αλληλεπιδράσεις στον MirSNP 15
Πίνακας 2.3: Αποθηκευμένα δεδομένα στον PolymiRTS 16
Πίνακας 2.4: Αποθηκευμένα δεδομένα στον miRdSNP 20
Πίνακας 2.5: Συγκεντρωτικά αποτελέσματα της αναζήτησης του hsa-miR-155 στον miRdSNP 23
Πίνακας 2.6: Ασθένειες που σχετίζονται με τον μεγαλύτερο αριθμό miRNAs στον miRdSNP 25
Πίνακας 2.7: Εξωτερικές πηγές πληροφοριών που χρησιμοποιεί ο miRvar 27
Πίνακας 2.8: Δεδομένα που είναι αποθηκευμένα στον MicroSniper 28
Πίνακας 2.9: Πολυμορφισμοί του hsa-miR-155 στον MicroSniper 29
Πίνακας 2.10: Ασθένειες που σχετίζονται με το hsa-miR-155 στον dPore 35

Κεφαλαίο 3

- Πίνακας 3.1: Πληροφορίες για την συλλογή των δεδομένων ανά WebServer 40
Πίνακας 3.2: Εκδόσεις του miRBase με χρονολογική σειρά 42
Πίνακας 3.3: Δυνατότητα επιλογής μελέτης και άλλων οργανισμών πέραν του ανθρώπου 44
Πίνακας 3.4: Μελέτη ασθενειών 44
Πίνακας 3.5: Τρόποι αναζήτησης ανά WebServer 45
Πίνακας 3.6: Ανακάλυψη αλληλεπιδράσεων ανά WebServer 45
Πίνακας 3.7: Αριθμός γονιδίων στόχων ανά WebServer 49
Πίνακας 3.8: Ποσοστό ταύτισης προβλέψεων ανά ζεύγος WebServers 49

Εικόνες:

Κεφάλαιο 1

- Εικόνα 1.1: Οπτική Απεικόνιση του miRNA 7
Εικόνα 1.2: Φυσιολογική Βιογένεση του miRNA 7
Εικόνα 1.3 Πολυμορφισμός του ίδιου του miRNA στα διάφορα στάδια της Βιογένεσης του 9
Εικόνα 1.4: Πολυμορφισμός στην seed περιοχή και στην περιοχή στόχο 10

Κεφάλαιο 2

- Εικόνα 2.1: Interface του Webserver MirSNP 12
Εικόνα 2.2: Γραφική απεικόνιση ταιριασμάτων μεταξύ miRNA-mRNA 12
Εικόνα 2.3: Τρόποι αναζήτησης στον MirSNP , μέσω αρχείου 12
Εικόνα 2.4: Αναζήτηση για ασθένειες στον MirSNP 13
Εικόνα 2.5: Αποτέλεσμα αναζήτησης για το miRNA hsa-miR-155-3p στο MirSNP 14
Εικόνα 2.6: Interface του WebServer PolymiRTS 17
Εικόνα 2.7: Αναζήτηση για το miRNA hsa-miR-155-3p στον PolymiRTS 17
Εικόνα 2.8: Καταγεγραμμένες αλληλεπιδράσεις γονιδίων-ασθενειών στον PolymiRTS 18
Εικόνα 2.9: Συγκεντρωτικά αποτελέσματα ανά miRNA-ασθένεια στον PolymiRTS 19
Εικόνα 2.10: Επιλογές αναζήτησης στον miRdSNP 20
Εικόνα 2.11: Γραφική απεικόνιση των SNPs, των target sites και των dSNPs ανά χρωμόσωμα στον άνθρωπο 21
Εικόνα 2.11: Αποτέλεσμα αναζήτησης στον miRvar 25
Εικόνα 2.12: Interface του MicroSniper 27
Εικόνα 2.13: Γραφική απεικόνιση των miRNAs ανά χρωμόσωμα 28

Εικόνα 2.14: Interface του mrSNP 30

Εικόνα 2.15: Αποτέλεσμα αναζήτησης για το SNP 69018917 στον mrSNP 30

Εικόνα 2.16: Αποτέλεσμα για την αναζήτηση του hsa-miR-155 στον dPore 32

Εικόνα 2.17: SNPs στον dPore για το hsa-miR-155 33

Εικόνα 2.18: SNPs ανά περιοχή για το hsa-miR-155 στον miRNASNP 35

Εικόνα 2.19: SNPs και γονίδια που σχετίζονται με το hsa-miR-155-3p στον miRNASNP 36

Εικόνα 2.20: Αποτελέσματα σχετικά με τον πολυμορφισμό του ίδιου του miRNA στον miRNASNP 36

Εικόνα 2.21: Συσχέτιση με ασθένειες του hsa-miR-155 στον miRNASNP 37

Κεφάλαιο 3

Εικόνα 3.1: Interface του miRBase 39

Γραφήματα

Κεφάλαιο 2

Γράφημα 2.1: Είδη των SNPs στον dPore 32

Κεφάλαιο 3

Γράφημα 3.1: πλήθος των miRNAs ανά WebServer 45

Γράφημα 3.2: miRNA target sites ανά WebServer 46

Γράφημα 3.3: αριθμός γονιδίων ανά WebServer 46

Γράφημα 3.4: αριθμός SNPs που σχετίζονται με miRNAs ανά WebServer 47

Γράφημα 3.5: Αριθμός πολυμορφισμών στην 3'UTR 47

1.Εισαγωγή

Τα τελευταία χρόνια λόγω της ραγδαίας εξέλιξης των υπολογιστικών συστημάτων και των υπολογιστικών αλγορίθμων, η βιολογία επικεντρώθηκε στην ανάπτυξη μεθόδων για την συλλογή γονιδιωματικών δεδομένων υψηλής απόδοσης. Ο συνδυασμός αυτών των δύο επιστημών οδήγησε στην δημιουργία της Βιοπληροφορικής, τα βιολογικά δεδομένα, δηλαδή οι αλληλουχίες (DNA, RNA, miRNA, SNPs κλπ.), αποθηκεύονται στην μορφή ψηφιακής πληροφορίας και μπορούμε να εφαρμόζουμε σε αυτά τα δεδομένα, διάφορους αλγορίθμους για την επεξεργασία τους αλλά και για την παραγωγή συμπερασμάτων Ένα τμήμα της έρευνας έχει επικεντρωθεί γύρω από τα miRNAs και τον πολύ σημαντικό ρόλο που παίζουν στην δημιουργία η μη ασθενειών. Για αυτό τον λόγο οι επιστήμονες της πληροφορικής ανέπτυξαν μεθόδους, αλγορίθμους και εργαλεία γύρω από αυτό το ζήτημα ώστε η έρευνα να είναι πιο γρήγορη και πιο αποτελεσματική Ανέπτυξαν βάσεις αναφοράς όπου είναι καταγεγραμμένο σε ψηφιακή μορφή το ανθρώπινο γονιδίωμα, ανέπτυξαν αλγόριθμους πρόβλεψης περιοχών στόχων όπου τα miRNAs έχουν προτίμηση, βάσεων δεδομένων όπου είναι καταγεγραμμένα όλα τα γνωστά miRNAs και τα γνωστά SNPs, βάσεων δεδομένων που καταγράφουν όλες τις πειραματικές επιβεβαιωμένες συσχετίσεις μεταξύ miRNAs και ασθενειών. Τα τελευταία 5-6 χρόνια διάφορα πανεπιστήμια και ερευνητικές ομάδες επικεντρώνονται στην δημιουργία WebServers που συνδυάζουν όλα τα παραπάνω.

1.1 Περίληψη

Στην παρούσα διπλωματική εργασία θα περιγράψουμε και θα συγκρίνουμε 9 διαφορετικούς WebServers. Η λειτουργία τους εστιάζει στον εντοπισμό περιοχών στόχων των miRNAs που εμφανίζουν πολυμορφισμό καθώς και σε ορισμένους από αυτούς στον πολυμορφισμό των ίδιων των miRNAs. Οι περιοχές αυτές ονομάζονται SNPs και σε πολλές περιπτώσεις σχετίζονται με ασθένειες. Αφού παρουσιάσουμε τον κάθε WebServer ξεχωριστά, στην συνέχεια θα τους συγκρίνουμε, με βάση μια σειρά παραμέτρων αριθμητικές όπως είναι ο συνολικός αριθμός των SNPs, ο αριθμός των miRNAs, οι αλγόριθμοι πρόβλεψης που χρησιμοποιούνται και ποιοτικές όπως είναι οι εκδόσεις των βάσεων που χρησιμοποιούνται. Επίσης θα συγκρίνουμε τα αποτελέσματα της αναζήτησης για το miR-155.

1.2 miRNAs

Τα miRNAs είναι μη κωδικοποιημένα μόρια RNA, δηλαδή ακολουθίες οι οποίες δεν κωδικοποιούν πρωτεΐνες, μήκους κατά μέσο όρο 22 βάσεων. Μέσω πειραμάτων που έχουν γίνει, έχει ανακαλυφθεί ότι είναι ρυθμιστές πάνω από το 60% όλων των γονιδίων, αυτό σημαίνει ότι επηρεάζουν σε πολύ σημαντικό βαθμό την λειτουργία του κυττάρου και κατ' επέκταση πολλές βιολογικές διαδικασίες όπως είναι ο πολλαπλασιασμός του κυττάρου, η απόπτωση, ο μεταβολισμός και πολλές ασθένειες όπως είναι ο καρκίνος.

Τα miRNAs εμφανίζονται στο στάδιο της μετα-μεταγραφής. Τα miRNAs συνήθως δεσμεύουν την 3' αμετάφραστη περιοχή ενός mRNA το οποίο έχει σαν αποτέλεσμα την μειωμένη έκφραση του συγκεκριμένου γονιδίου, η υποέκφραση του συγκεκριμένου λοιπόν γονιδίου μπορεί να οδηγήσει στην εμφάνιση κάποιας ασθένειας, η οποία δεν θα είχε εμφανισθεί χωρίς την παρέμβαση του συγκεκριμένου miRNA, αντίθετα σε ορισμένες περιπτώσεις η παρουσία ενός miRNA μπορεί να λειτουργήσει ακριβώς αντίθετα, δηλαδή να εμποδίσει την εμφάνιση μια ασθένειας. Μερικές φορές για παράδειγμα τα miRNAs λειτουργούν σαν ογκογονίδια ενώ άλλες φορές σαν ογκοκατασταλτικά γονίδια. Όπως είπαμε και στην περίληψη στην συγκεκριμένη διπλωματική εργασία θα παρουσιάσουμε και θα συγκρίνουμε WebServers που εντοπίζουν στόχους των miRNAs (binding sites), τα οποία διαφέρουν μεταξύ τους κατά μια βάση (SNPs), δηλαδή πολυμορφικές περιοχές των mRNAs στις οποίες τα miRNAs δείχνουν να έχουν προτίμηση

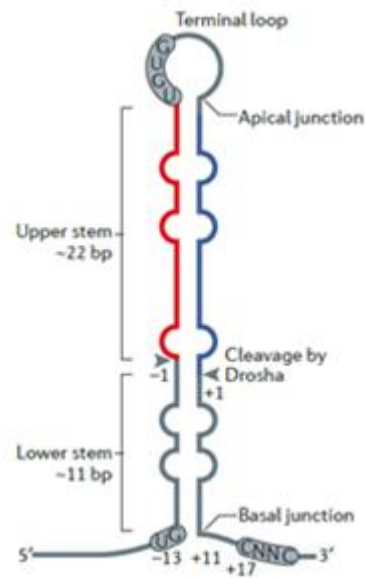
1.3 Ιστορικά Στοιχεία

Το πρώτο miRNA ανακαλύφθηκε το 1993 από τον Victor Ambros και την ερευνητική ομάδα, όπως συμβαίνει πολύ συχνά στην επιστήμη με εντελώς τυχαίο τρόπο. Συγκεκριμένα μελετούσαν το γονίδιο lin-4, το οποίο όπως ήταν γνωστό ότι ελέγχει την ανάπτυξη του σκουληκιού *Caenorhabditis elegans* μέσω της κατασταλτικής του δράσης πάνω στο γονίδιο lin-14. Κατά την διάρκεια του πειράματος, όταν απομόνωσαν το lin-4 συνειδητοποίησαν ότι αντί να παραχθεί ένα κανονικό μόριο mRNA, το οποίο φυσιολογικά κωδικοποιεί μια πρωτεΐνη, παρήχθησαν μικρά μη-κωδικοποιημένα RNAs, εκ των οποίων το ένα είχε μήκος 22 νουκλεοτίδια και περιείχε ακολουθίες σχεδόν συμπληρωματικές με διάφορα τμήματα του 3' άκρου του lin-14. Θεώρησαν λανθασμένα ότι το συγκεκριμένο περιστατικό αποτελεί, ιδιομορφία του συγκεκριμένου σκουληκιού. Πέρασαν αρκετά χρόνια και εν τέλει το 2000, εντοπίστηκε ένα δεύτερο μικρό RNA, το οποίο ονομάστηκε let-7, το οποίο αποτελεί σήμερα μια ολόκληρη ομάδα miRNAs και εμφανίζεται σε πολλούς οργανισμούς συμπεριλαμβανομένου και του ανθρώπου. Το let-7 παρατηρήθηκε ότι καταστέλλει το γονίδιο lin-41.

1.4 Ονοματολογία – Συμβολισμοί

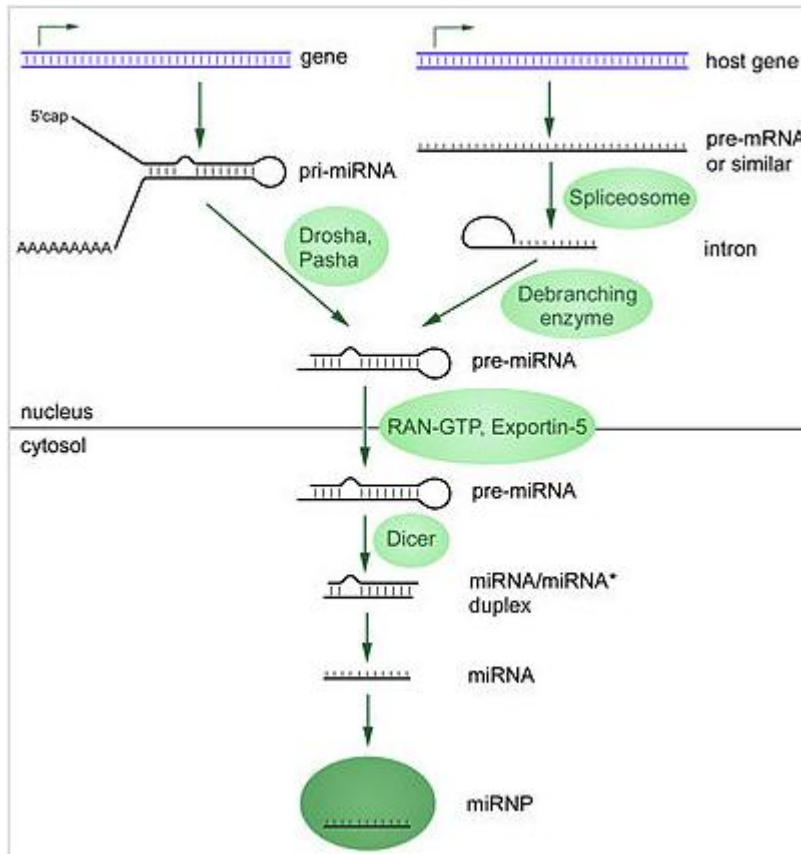
Για να δοθεί όνομα σε ένα miRNA πρέπει πρώτα αυτό να έχει αποδειχθεί πειραματικά. Το πρόθεμα "miR" ακολουθούμενο από παύλα και ένα αριθμό, ο αριθμός συνήθως υποδηλώνει την σειρά ανακάλυψης πχ το miR-145 ανακαλύφθηκε πριν το miR-189. Όταν χρησιμοποιούμε R, αυτό υποδηλώνει ότι αναφερόμαστε στο ώριμο miRNA και όταν χρησιμοποιούμε r αναφερόμαστε στο πρόδρομο miRNA. Όταν miRNAs είναι σχεδόν ίδια, δηλαδή διαφέρουν σε 1-2 νουκλεοτίδια τότε χρησιμοποιούμε τον ίδιο αριθμό και να τα ξεχωρίσουμε προσθέτουμε ένα επιπλέον γράμμα στην κατάληξη πχ miR-135a και miR-135b. Όταν δύο πρόδρομα miRNAs βρίσκονται σε διαφορετικές περιοχές, αλλά καταλήγουν στην παραγωγή του ίδιου ώριμου miRNA, τότε τα συμβολίζουμε για παράδειγμα, hsa-miR-194-1 και hsa-miR-194-2 που καταλήγουν στο hsa-miR-194, το

πρόθεμα hsa αναφέρεται στον οργανισμό στον οποίο ανήκει το miRNA στην συγκεκριμένη περίπτωση, hsa=homosariens(άνθρωπος).



(Εικόνα 1.1: Δομή του πρόδρομου miRNA, Πηγή: Minju Ha& V. Narry Kim: Regulation of microRNA biogenesis Nature Reviews Molecular Cell Biology 15, 509–524 (2014))

1.5 Φυσιολογική Βιογένεση miRNA



(Εικόνα 1.2: Βιογένεση του miRNA και τρόπος δράσης, Πηγή:Wikipedia)
<https://en.wikipedia.org/wiki/MicroRNA#/media/File:MiRNA-biogenesis.jpg>

Τα περισσότερα γονίδια των miRNAs εμφανίζονται σε εσόνια γονιδίων που κωδικοποιούν πρωτεΐνες ή σε μεταγραφικές μονάδες που εντοπίζονται μεταξύ γονιδίων και δεν κωδικοποιούν πρωτεΐνες. Τα μόρια miRNA συνήθως μεταγράφονται από το ένζυμο RNA πολυμεράση II, ένα ένζυμο που εμφανίζεται στα κύτταρα των ευκαρυωτικών οργανισμών. Η λειτουργία της είναι να καταλύει την μεταγραφή του DNA και να παράγει πρόδρομα mRNA και microRNA. Η RNA πολυμεράση, δεσμεύει μια πολύ συγκεκριμένη περιοχή του DNA και συγκεκριμένα στο σημείο εκείνο, το οποίο μετέπειτα θα γίνει το hair-pin loop του pre-miRNA. Σε πρώτη φάση παράγεται το pri-miRNA(primary), το οποίο περνάει από πολλά στάδια ωρίμανσης, έχει μήκος πάνω από 1000 βάσεις και περιέχει ένα loop, στο οποίο είναι ενσωματωμένη η ακολουθία του ώριμου miRNA. Στην συνέχεια ένα ένζυμο η Drosha κόβει το pri-miRNA και παράγει το pre-miRNA μήκους περίπου 65 νουκλεοτιδίων. Το σημείο στο οποίο η Drosha (σε συνδυασμό με άλλα ένζυμα) κόβει το pri-miRNA δεν είναι καθόλου τυχαίο, πρέπει να απέχει 11 βάσεις από την περιοχή που ξεκινάει η 3' περιοχή και 22 βάσεις από την περιοχή του loop, όπως φαίνεται και στο Σχήμα 1. Στην συνέχεια για να μεταφερθεί το pre-miRNA από τον πυρήνα στο κυτταρόπλασμα παρεμβαίνει μια πρωτεΐνη που ονομάζεται exportin-5. Στο επόμενο στάδιο ένα άλλο σύμπλεγμα πρωτεϊνών που περιλαμβάνει τα ένζυμα DICER και TRBP, κόβουν περαιτέρω το pre-miRNA και παράγονται 2 αντίγραφα του miRNA, μέσω της απομάκρυνσης του ενός από τα δύο, έχουμε την παραγωγή του ώριμου miRNA, το οποίο μέσω της σύνδεσης του με άλλες

πρωτεΐνες όπως είναι η AGO2 και η GW182 δημιουργεί το μπλόκο RISC , το οποίο μέσω της πρόσδεσης του απέναντι από ένα μονόκλωνο RNA προκαλεί σίγαση σε πολλά γονίδια που βρίσκονται κατά κύριο λόγο στην 3' αμετάφραστη περιοχή. Το ώριμο miRNA που παράχθηκε μπορεί να προκαλέσει μετα-μεταγραφικές ανωμαλίες από την σίγαση διάφορων γονιδίων μέσω της πρόσδεσης του μπλόκου RISC σε σχεδόν συμπληρωματικές περιοχές διάφορων mRNA και κυρίως στην 3' αμετάφραστη περιοχή

1.6 Απορρύθμιση και Ανωμαλίες στην Βιογένεση των miRNA

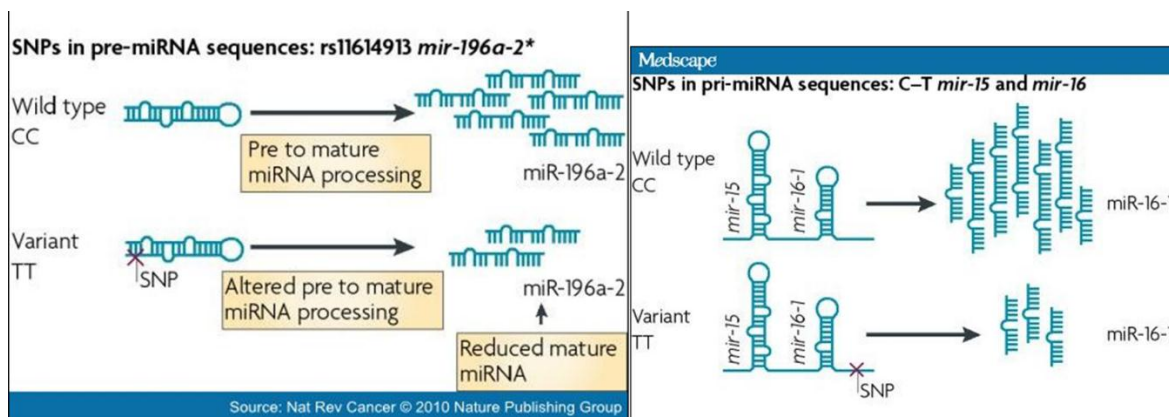
Εκτός από την φυσιολογική βιογένεση ενός miRNA όπως περιγράψαμε αναλυτικά παραπάνω , υπάρχουν και διαφορετικά μονοπάτια μη φυσιολογικά που μπορούν να παράγουν μόρια miRNA. Ένα τέτοιο μονοπάτι προκύπτει όταν υπάρχει για παράδειγμα απουσία της Droscha και συγκεκριμένα το σημείο εκείνο στο οποίο η Droscha κόβει το pri-miRNA παραλείπεται , τον ρόλο αυτό τον αναλαμβάνει ένα άλλο μοριακό εργαλείο που ονομάζεται «κοπτόσωμα»(spliceosome). Άλλη περίπτωση διαφορετικού μονοπατιού είναι το pre-miRNA να μην εξάγεται στο κυτταρόπλασμα από την πρωτεΐνη EXP-5 αλλά από την EXP-1 , τέλος υπάρχουν περιπτώσεις όπου το τελικό ώριμο miRNA προέρχεται από άλλου τύπου μικρά RNAs όπως είναι τα tRNAs και τα snoRNAs

1.7 SNPs(Single Nucleotide Polymorphism)

Οι μονονουκλεοτιδικοί πολυμορφισμοί , είναι παραλλαγές νουκλεοτιδίων στην ίδια γενετική θέση της αλληλουχίας DNA, είτε είναι σε γονίδιο είτε σε άλλες θέσεις του γονιδιώματος Για να θεωρηθεί μια μονονουκλεοτιδική αλλαγή πολυμορφική θα πρέπει να παρατηρείται σε συχνότητα >1% του γενικού πληθυσμού. Τα τελευταία χρόνια φάνηκε ότι κάποια SNPs ενδεχομένως να μην είναι τυχαία απλές παραλλαγές από τυχαίες μεταλλάξεις αλλά να συνδέονται με επιβάρυνση ή προστασία από μια παθολογική κατάσταση επηρεάζοντας αλλά όχι αλλάζοντας τη δράση κάποιας σημαντικής πρωτεΐνης. Μέχρι σήμερα έχουν καταγραφεί πάνω από 5 εκατομμύρια SNPs ενώ έχει ξεκινήσει μια προσπάθεια καταγραφής των διαφορετικών SNPs (HarMap) από 4 διαφορετικούς πληθυσμούς. Επίσης επειδή τα SNPs δεν είναι επικίνδυνα για την ζωή , κληρονομούνται με σταθερό τρόπο από γενιά σε γενιά από την στιγμή που δημιουργούνται , με αποτέλεσμα να χρησιμοποιούνται και σε πληθυσμιακές μελέτες για την καταγωγή των ανθρώπων ιστορικά. Η μελέτη τους είναι πολύ σημαντική γιατί μπορούμε μέσω αυτής να βγάλουμε πολλά συμπεράσματα για την ανθρώπινη υγεία. Ένας ή περισσότεροι πολυμορφισμοί μπορεί να καθορίζουν μια σειρά από χαρακτηριστικά , όπως τον βαθμό απόκρισης μας σε μια φαρμακευτική αγωγή, τις πιθανότητες εκδήλωσης συγκεκριμένων ασθενειών και πολλά ακόμη. Ιδιαίτερα στις ασθένειες που εξαρτώνται από πολλούς παράγοντας όπως είναι ο διαβήτης και ο καρκίνος , η μελέτη των SNPs μπορεί να μας οδηγήσει σε πολύ χρήσιμα συμπεράσματα

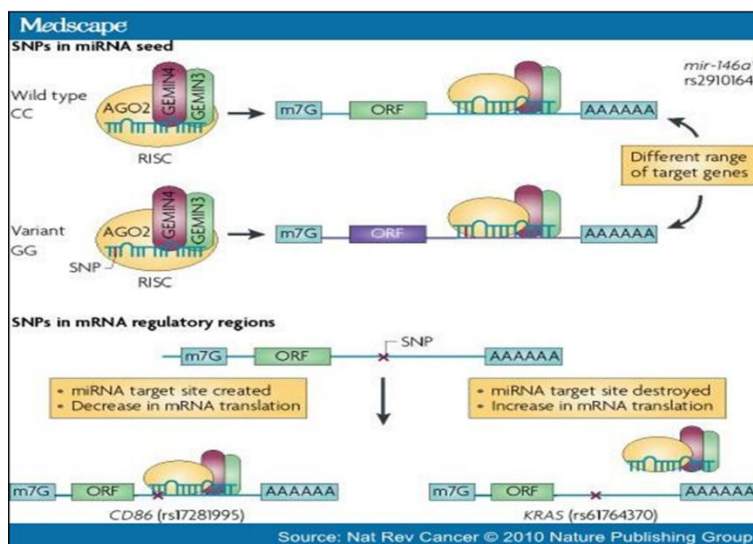
1.8 miRNAs και SNPs

Σε αυτήν την υπο-ενότητα θα προσπαθήσουμε να εξηγήσουμε τον τρόπο με τον οποίο συνδέονται τα miRNAs με τα SNPs και την ιδιαίτερη αξία που έχει η συνδυαστική τους μελέτη. Οι πολυμορφισμοί μπορεί λοιπόν να εμφανισθούν και στα ίδια τα miRNAs κατά τα διάφορα στάδια της Βιογένεσης τους όπως φαίνεται στην Εικόνα 1.3 αλλά και στις περιοχές στόχους τους όπως φαίνεται στην Εικόνα 1.4



(Εικόνα 1.3 Τρόπος δράσης πολυμορφισμών σε miRNA στα διάφορα στάδια της Βιογένεσης του, Πηγή: Bríd M. Ryan; Ana I. Robles; Curtis C. Harris, Genetic Variation in microRNA Networks: the Implications for Cancer Research)

Όπως παρατηρούμε στην πρώτη περίπτωση, η αλλαγή αυτή δηλαδή η αλλαγή μια μοναδικής βάσης, είτε στο pri-miRNA είτε στο pre-miRNA, μπορεί να οδηγήσει εν τέλει στην παραγωγή μικρότερης ποσότητας ώριμου (mature) miRNA, κατ'επέκταση λοιπόν θα επηρεαστούν λειτουργίες του κυττάρου που ρυθμίζονται από το συγκεκριμένο miRNA.



(Εικόνα 1.4: Πολυμορφισμός στην seed περιοχή και στην περιοχή στόχο, Πηγή: Bríd M. Ryan; Ana I. Robles; Curtis C. Harris, Genetic Variation in microRNA Networks: the Implications for Cancer Research)

Επίσης ο πολυμορφισμός μπορεί να εμφανισθεί στην seed περιοχή του mature miRNA ή σε κάποια περιοχή στόχο, δηλαδή πάνω στο mRNA και στις δύο περιπτώσεις έχουμε πρακτικά αλλαγή των περιοχών στόχων των miRNAs δηλαδή μια περιοχή που πριν ήταν, τώρα δεν αποτελεί ενδιαφέρον και το αντίστροφο.

Σύμφωνα με τα παραπάνω λοιπόν γίνεται εύκολα αντιληπτό ότι η μελέτη τους πρέπει να γίνεται παράλληλα, διότι η κατανόηση τους αλλά και η ανακάλυψη νέων αλληλεπιδράσεων μεταξύ τους, μπορεί να οδηγήσει στην θεραπεία ασθενειών.

2. Webservers

Στην συγκεκριμένη διπλωματική εργασία στόχος μας ήταν να συγκεντρώσουμε όσο το δυνατόν περισσότερες πληροφορίες για μια σειρά από WebServers με παρόμοια λειτουργία. Οι WebServers είναι OnLine εργαλεία που μας βοηθούν να αναζητήσουμε συνδυαστικές πληροφορίες για miRNAs, SNPs, ασθένειες, γονίδια, πειράματα, διαφορετικά είδη κλπ. Οι WebServers τους οποίους μελετήσαμε είναι οι PolymiRTS, dPore, miRdSNP, miRNASNP, mrSNP, miRvar, dbSMR, miRNA Sniper, MirSNP. Βασικός στόχος τους είναι να παρουσιάσουν αλλά να ανακαλύψουν και νέες συσχετίσεις μεταξύ των miRNAs και περιοχών στόχων των miRNAs που εμφανίζουν πολυμορφισμό, περιοχές γνωστές και ως SNPs, το πως το πετυχαίνει ο καθένας ξεχωριστά θα το

αναλύσουμε στην συνέχεια. Μερικοί από αυτούς ανακαλύπτουν και πιθανές συσχετίσεις μεταξύ αυτών των περιοχών και την εμφάνιση ασθενειών. Σε γενικές γραμμές συλλέγουν πληροφορίες από άλλους WebServers και από την βιβλιογραφία και καταλήγουν σε ένα dataset. Έχουν αρκετές ομοιότητες μεταξύ τους αλλά και σημαντικές διαφορές για αυτό πέρα από την παρουσίαση που θα κάνουμε για τον καθένα ξεχωριστά, θα επιχειρήσουμε και μια σύγκριση μεταξύ τους.

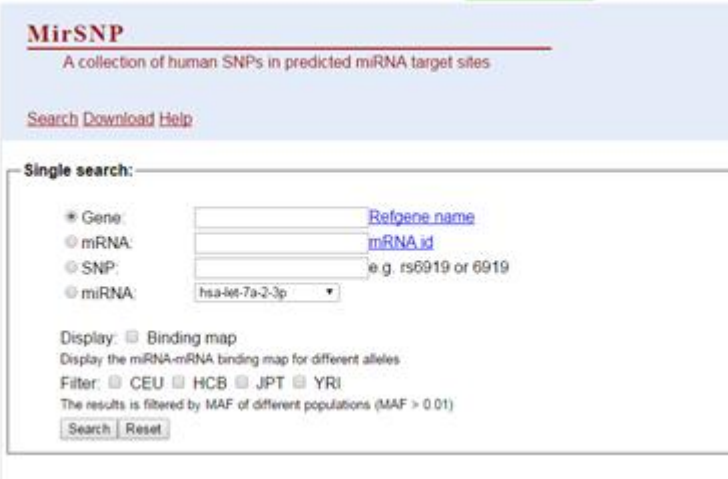
2.1 MirSNP

Μία βάση δεδομένων που εντοπίζει SNPs τα οποία επηρεάζουν τις περιοχές στόχους(binding sites) των miRNAs και γενικότερα miRNAs που εξαρτώνται από SNPs. Όπως έχουμε ήδη αναφέρει τα miRNAs αποτελούν ρυθμιστικούς παράγοντες στην έκφραση των γονιδίων. Με την σειρά τους τα SNPs επηρεάζουν την seed περιοχή των miRNA στόχων. Οι επιρροές αυτές έχουμε παρατηρήσει πως σχετίζονται με διάφορες ασθένειες, όπως με καρκίνους αλλά και ψυχικές διαταραχές, καρδιακά νοσήματα αλλά και με το άσθμα.

Ο κάθε χρήστης μπορεί να κατασκευάσει το δικό του GWAS(Genome wide association study) ή το δικό του eQTL και να κατευθύνει την μελέτη του, όπως αυτός θέλει.

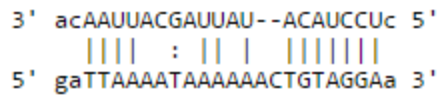
Το MirSNP περιλαμβάνει πάνω από 414,510 SNPs που σχετίζονται με miRNAs, υπάρχει και η δυνατότητα εντοπισμού νέων διασυνδέσεων από τον χρήστη, κάνοντας αναζήτηση με τα δικά του δεδομένα.

Όπως μπορεί να παρατηρήσει κάποιος, μας δίνεται η δυνατότητα να κάνουμε αναζήτηση με 4 διαφορετικούς τρόπους(γονίδιο, mRNA, SNP, miRNA), μέσω του εκάστοτε ονόματος.



(Εικόνα 2.1: Interface του Webservice MirSNP)

Επίσης υπάρχει η δυνατότητα της γραφικής απεικόνισης, αρκεί να επιλέξουμε την επιλογή binding map, τότε το αποτέλεσμα της αναζήτησης θα μας δώσει και την ακολουθία και τα ταιριάσματα μεταξύ miRNA και mRNA.



(Εικόνα 2.2: Γραφική απεικόνιση ταιριασμάτων μεταξύ miRNA-mRNA)

Επίσης υπάρχει η δυνατότητα για σύνθετη αναζήτηση μέσω αρχείου. Στο αρχείο μπορούμε να δώσουμε σε μορφή λίστας ταυτόχρονα πολλά διαφορετικά γονίδια , mRNAs ή SNPs

(Εικόνα 2.3: Τρόποι αναζήτησης στον MirSNP, μέσω αρχείου)

Τέλος μπορούμε να κάνουμε αναζήτηση ξανά μέσω αρχείου , ώστε να δούμε για συγκεκριμένα SNPs , με ποιες ασθένειες και ποια χαρακτηριστικά σχετίζονται ή έχουν άμεση επιρροή.

(Εικόνα 2.4: Αναζήτηση για ασθένειες στον MirSNP)

Στην συνέχεια παρουσιάζουμε και μερικά στατιστικά στοιχεία

| | |
|------------------|--------|
| Αριθμός Γονιδίων | 17,569 |
| Αριθμός mRNAs | 29,273 |

| | |
|--------------------|-----------|
| Αριθμός miRNAs | 1,921 |
| MirSNPs: | 414,510 |
| Records: | 2,709,054 |
| Records(MAF>0.01): | 227,062 |

(Πίνακας 2.1:Αποθηκευμένα δεδομένα στον MirSNP)

Το MirSNP δημιουργήθηκε με το MySQL software , τα miRNAs και οι πληροφορίες για αυτά , εισήχθησαν από το miRBase 18.0 , για την ονοματολογία και τις απαραίτητες πληροφορίες των SNPs έγινε χρήση 4 πινάκων από την βάση dbSNP135. Για την καλύτερη αξιοπιστία των δεδομένων , επιπλέον αρχεία με ακολουθίες γνωστών mRNA εισήχθησαν από τους NCBI και UCSC GenomeBrowsers. Συνολικά 42,733 mRNA και 513,249 SNPs είναι στην διάθεση του χρήστη για να κάνει την ανάλυσή του. Η πρόβλεψη των πιθανών στόχων των miRNAs έγινε μέσω της μεθόδου mirSVR η οποία στην ουσία αποτελεί ένα machine learning . Το machine learning όπως υποδηλώνει και η ονομασία του , είναι μια επιστημονικής μέθοδος που βελτιώνει τα αποτελέσματά της με την πάροδο του χρόνου, δηλαδή όσο περισσότερο επεξεργάζεται δεδομένα , τόσο καλύτερα αποτελέσματα μας δίνει. Για να βελτιωθεί το αποτέλεσμα και να μην υπάρχουν επικαλύψεις των SNPs , το MirSNP χρησιμοποιεί κάποιου είδους φίλτρου. Περισσότερες πληροφορίες για το κάθε SNP ξεχωριστά εισήχθησαν από το HarMap Project στο οποίο κάναμε αναφορά και νωρίτερα , και συγκεκριμένα από το Phase 2 HarMap fileset. Άλλα εργαλεία και software που χρησιμοποιήθηκαν στην κατασκευή του site ήταν το PLINK software για μαθηματικούς υπολογισμούς , καθώς και το Django.

Μία ενδεικτική αναζήτηση για το has-miR-155-3p , ο πίνακας μας δίνει πληροφορίες για το γονίδιο , το miRNA , το όνομα του SNP , το score του αλγορίθμου πρόβλεψης και την επίδραση την οποία έχει το SNP.

| Gene | miRNA | Snp | mirSVR | Effect | Allele | Score | Energy | Conservation |
|-----------|----------------|-------------|--------|-----------------|--------|--------|--------|--------------|
| ABCG5 | hsa-miR-155-3p | rs4148195 | | create | C | | | |
| | | | | | T | 151.00 | -8.15 | 0.001 |
| ACAD8 | hsa-miR-155-3p | rs139716145 | | create | C | | | |
| ACOX1 | hsa-miR-155-3p | rs71382188 | | -0.005 break | G | 141.00 | -11.75 | 0.000 |
| | | | | | C | | | |
| ACOX1 | hsa-miR-155-3p | rs187624130 | | -0.005 break | T | 141.00 | -11.75 | 0.000 |
| | | | | | C | | | |
| ACSS2 | hsa-miR-155-3p | rs192584896 | | -0.608 enhance | C | 140.00 | -11.29 | 0.767 |
| | | | | | T | 144.00 | -11.22 | 0.767 |
| ADAMTS5 | hsa-miR-155-3p | rs78716320 | | -0.391 break | T | 156.00 | -11.64 | 0.464 |
| | | | | | A | | | |
| ADARB1 | hsa-miR-155-3p | rs28404209 | | -0.086 break | A | 143.00 | -9.72 | 0.041 |
| | | | | | G | | | |
| ADARB1 | hsa-miR-155-3p | rs28663337 | | -0.086 create | G | | | |
| | | | | | T | 143.00 | -9.72 | 0.041 |
| ADAT2 | hsa-miR-155-3p | rs139120374 | | -0.068 decrease | T | 160.00 | -11.32 | 0.022 |
| | | | | | C | 152.00 | -9.67 | 0.022 |
| ADM2 | hsa-miR-155-3p | rs8136908 | | -0.001 break | A | 140.00 | -13.80 | 0.000 |
| | | | | | G | | | |
| ADPRHL2 | hsa-miR-155-3p | rs111335009 | | create | C | | | |
| | | | | | T | 160.00 | -12.62 | 0.004 |
| AEN | hsa-miR-155-3p | rs189668489 | | -0.184 create | A | | | |
| | | | | | G | 143.00 | -11.87 | 0.000 |
| AFF4 | hsa-miR-155-3p | rs192634524 | | break | G | 162.00 | -16.86 | 0.999 |
| | | | | | A | | | |
| ALG10B | hsa-miR-155-3p | rs1046069 | | -0.079 break | G | 163.00 | -12.16 | 0.011 |
| | | | | | T | | | |
| ALKBH1 | hsa-miR-155-3p | rs11848383 | | create | C | | | |
| | | | | | A | 156.00 | -11.97 | 0.000 |
| ANKH | hsa-miR-155-3p | rs149623281 | | break | T | 155.00 | -14.75 | 0.000 |
| | | | | | C | | | |
| ANKRD20A3 | hsa-miR-155-3p | rs4582814 | | -0.392 break | T | 142.00 | -14.59 | 0.003 |
| | | | | | A | | | |

(Εικόνα 2.5: Αποτέλεσμα αναζήτησης για το miRNA hsa-miR-155-3p στο MirSNP)

Επίσης μας δίνει και στο δεξί πίνακα πληροφορίες αποκλειστικά για το SNP , όπως σε ποια βάση διαφέρει, το score , την ενέργεια και το επίπεδο διατήρησης του.(ο πίνακας αποτελεί μόνο ένα κομμάτι των αποτελεσμάτων)

Ο WebServer δεν μας δίνει την δυνατότητα αναζήτησης διασυνδέσεων μεταξύ miRNAs και ασθενειών , αλλά μόνο μεταξύ SNPs και ασθενειών για αυτό και παραθέτουμε έναν ενδεικτικό πίνακα με πειραματικά αποδεδειγμένα SNPs τα οποία σχετίζονται με την εμφάνιση ασθενειών που εντοπίστηκαν από τον MirSNP WebServer

| miRNA | Γονίδιο | SNP | Ασθένεια |
|-------------|---------|-----------|---------------------------------------|
| hsa-mir-629 | NBS1 | rs2735383 | Καρκίνος του Πνεύμονα |
| hsa-mir-184 | TNFAIP2 | rs8126 | Καρκίνωμα του κεφαλιού και του λαιμού |

| | | | |
|-----------------|---------|-------------|-----------------------------|
| hsa-mir-1827 | MYCL1 | rs3134615 | Τύπος καρκίνου του πνεύμονα |
| hsa-mir-148a | HLA-C | rs67384697 | HIV |
| hsa-mir-125b | BMPR1B | rs1434536 | Καρκίνος του μαστού |
| hsa-mir-148/152 | HLA-G | Rs1063320 | Παιδικό άσθμα |
| Has-mir24 | SLITRK1 | rs193302862 | Σύνδρομο Tourette |

(Πίνακας 2.2:Επιβεβαιωμένες Αλληλεπιδράσεις στον MirSNP)

2.2 PolymiRTS Database 3.0

Ο συγκεκριμένος WebServer με την σειρά του έχει ως στόχο να εντοπίσει και να αναλύσει την επίδραση που έχουν οι πολυμορφισμοί(μεταξύ αυτών και τα SNPs) στην seed περιοχή των miRNAs. Για να το πετύχει αυτό αποτελεσματικά χρησιμοποιεί κάποιες τεχνικές και επιστημονικές μεθόδους. Μεταξύ αυτών τα δεδομένα CLASH. Το CLASH είναι μία τεχνική που χαρτογραφεί της αλληλεπιδράσεις μεταξύ RNAs και συγκεκριμένα στην περίπτωση την δική μας αλληλεπιδράσεις τύπου miRNA-mRNA κατευθείαν σε ζευγάρια. Συγκεκριμένα το dataset περιλαμβάνει 18,514 περιοχές στόχους για 399 διαφορετικά miRNA(homo).Σε αυτό το σημείο παρόλο που δεν θα ασχοληθούμε αξίζει να αναφερθούμε σε ένα άλλο είδος πολυμορφισμών που ονομάζονται INDELS και ο συγκεκριμένος WebServer τους εντοπίζει, λέγονται έτσι γιατί οι ακολουθίες διαφέρουν μεταξύ τους κατά μία διαγραφή ή προσθήκη βάσης.

Ο αλγόριθμος πρόβλεψης που χρησιμοποιείται είναι του TargetScan, ο οποίος υπολογίζει έναν αριθμό context+score προκειμένου να εκτιμήσει και να εντοπίσει πιθανές miRNA-mRNA αλληλεπιδράσεις. Επίσης κάνοντας χρήση του KEGG database, χρησιμοποιεί όλα τα γνωστά βιολογικά μονοπάτια για να αντλεί περισσότερες πληροφορίες. Για παράδειγμα γνωρίζουμε ότι το SNP rs5186, το οποίο εντοπίζεται στο miR-155, αλλάζει την έκφραση ενός γονιδίου, συγκεκριμένα του AGTR1, το οποίο σχετίζεται με την πίεση του αίματος.

Γνωρίζοντας τα βιολογικά μονοπάτια όλων των γονιδίων και άλλων λειτουργιών, μας είναι πιο εύκολο να αντιληφθούμε και γραφικά πως γίνεται η παραπάνω διαδικασία Το PolymiRTS Database 3.0 χρησιμοποιεί την 20^η έκδοση του miRBase, για τις απαραίτητες πληροφορίες και την ονοματολογία των SNPs την βάση δεδομένων dbSNP 126, πληροφορίες σχετικά με τα γονίδια αντλεί από τον Ensembl και από το Refgene table of UCSC GenomeBrowser hg18, για την πειραματική υποστήριξη το TarBase και το

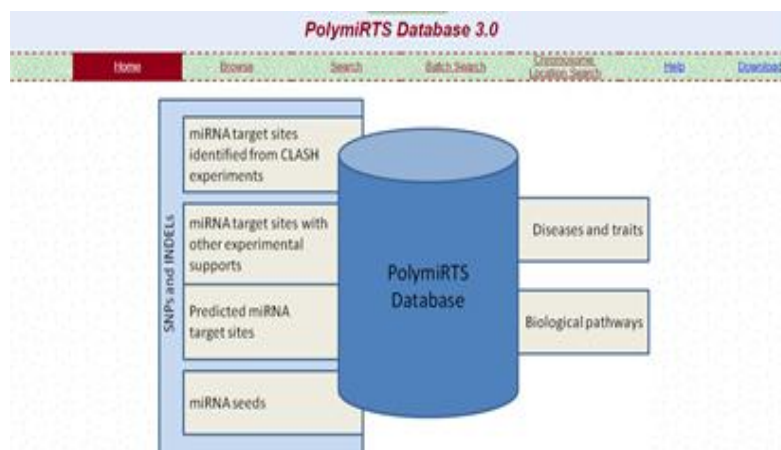
mirTarBase καθώς και της βάσεις δεδομένων του ανθρώπινου γονιδιώματος NHGRI GWAS Catalog, dbGap, GTEx Eqtl. Επιπλέον διατίθενται πληροφορίες και για το ποντίκι πέρα από τον άνθρωπο.

Στον παρακάτω πίνακα παρουσιάζονται κάποια επιπλέον στοιχεία

| | |
|--|---|
| SNPs στις περιοχές στόχους των miRNAs | 22979(CLASH) 358874(TargetScan) 1900(από άλλες μεθόδους) |
| SNPs στις περιοχές seed των miRNAs | 271 |
| Αριθμός miRNAs | 399 |
| Αριθμός Γονιδίων | 13119 |

(Πίνακας 2.3:Αποθηκευμένα δεδομένα στον PolymiRTS)

Στο αρχικό μενού του site υπάρχουν όλες οι διαθέσιμες πληροφορίες αρκεί να επιλέξουμε συγκεκριμένα αυτό που μας ενδιαφέρει, η κάθε επιλογή μας δίνει μια συγκεντρωτική λίστα, με όση πληροφορία δηλαδή όσες αλληλεπιδράσεις είναι αποθηκευμένες στον server. Όπως φαίνεται και από την παρακάτω εικόνα, μπορούμε να επιλέξουμε να δούμε πληροφορίες για miRNA target sites τα οποία έχουν εντοπιστεί με την μέθοδο CLASH, miRNA target sites που έχουν εντοπιστεί από άλλες πειραματικές μεθόδους , miRNA target sites που έχουν προβλεφθεί από κάποιον αλγόριθμο, περιοχές seed των miRNA, ασθένειες και χαρακτηριστικά καθώς και Βιολογικά μονοπάτια.



(Εικόνα 2.6:Interface του WebServer PolymiRTS)

Φυσικά στον χρήστη δίνονται και άλλες δυνατότητες ώστε να αναλύσει τα δικά του δεδομένα και να βγάλει τα δικά του συμπεράσματα Αναζήτηση είτε σύνθετη , είτε μεμονωμένη Οι επιλογές αναζήτησης είναι για SNPs, miRNAs, Γονίδια (και με τον κωδικό του και περιγραφικά), χαρακτηριστικά. Επίσης δίνεται η δυνατότητα στον χρήστη για απευθείας αναζήτηση θέσης σε κάποιο χρωμόσωμα.

Η παρακάτω εικόνα μας δείχνει ένα τμήμα της αναζήτησης για το miRNA , hsa-miR-155-3p. Όπως βλέπουμε η έξοδος περιλαμβάνει μια λίστα από γονίδια που αποτελούν περιοχές στόχους του συγκεκριμένου miRNA.

| Transcript ID | Gene Symbol | Description |
|---------------|-------------|---|
| NM_000080 | CHRNE | Homo sapiens cholinergic receptor, nicotinic, epsilon (CHRNE), mRNA |
| NM_000108 | DLD | Homo sapiens dihydrolipoamide dehydrogenase (DLD), mRNA |
| NM_000112 | SLC26A2 | Homo sapiens solute carrier family 26 (sulfate transporter), member 2 (SLC26A2), mRNA |
| NM_000337 | SGCD | Homo sapiens sarcoglycan, delta (35kDa dystrophin-associated glycoprotein) (SGCD), transcript variant 1, mRNA |
| NM_000351 | STS | Homo sapiens steroid sulfatase (microsomal), isozyme S (STS), mRNA |
| NM_000388 | CASR | Homo sapiens calcium-sensing receptor (CASR), transcript variant 2, mRNA |
| NM_000428 | LTBP2 | Homo sapiens latent transforming growth factor beta binding protein 2 (LTBP2), mRNA |
| NM_000448 | RAG1 | Homo sapiens recombination activating gene 1 (RAG1), mRNA |
| NM_000449 | RFX5 | Homo sapiens regulatory factor X, 5 (influences HLA class II expression) (RFX5), transcript variant 1, mRNA |
| NM_000543 | SMPD1 | Homo sapiens sphingomyelin phosphodiesterase 1, acid lysosomal (SMPD1), transcript variant 1, mRNA |
| NM_000618 | IGF1 | Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1), transcript variant 4, mRNA |
| NM_000724 | CACNB2 | Homo sapiens calcium channel, voltage-dependent, beta 2 subunit (CACNB2), transcript variant 1, mRNA |
| NM_000743 | CHRNA3 | Homo sapiens cholinergic receptor, nicotinic, alpha 3 (CHRNA3), transcript variant 1, mRNA |
| NM_000744 | CHRNA4 | Homo sapiens cholinergic receptor, nicotinic, alpha 4 (CHRNA4), mRNA |
| NM_000790 | DDC | Homo sapiens dopa decarboxylase (aromatic L-amino acid decarboxylase) (DDC), transcript variant 2, mRNA |
| NM_000798 | DRD5 | Homo sapiens dopamine receptor D5 (DRD5), mRNA |
| NM_000813 | GABRB2 | Homo sapiens gamma-aminobutyric acid (GABA) A receptor, beta 2 (GABRB2), transcript variant 2, mRNA |
| NM_000914 | OPRM1 | Homo sapiens opioid receptor, mu 1 (OPRM1), transcript variant MOR-1, mRNA |
| NM_001001412 | CALHM1 | Homo sapiens calcium homeostasis modulator 1 (CALHM1), mRNA |
| NM_001001662 | ZNF782 | Homo sapiens zinc finger protein 782 (ZNF782), mRNA |
| NM_001002269 | EXOSC3 | Homo sapiens exosome component 3 (EXOSC3), transcript variant 2, mRNA |
| NM_001002843 | ZNF280D | Homo sapiens zinc finger protein 280D (ZNF280D), transcript variant 2, mRNA |
| NM_001005163 | OR52D1 | Homo sapiens olfactory receptor, family 52, subfamily D, member 1 (OR52D1), mRNA |
| NM_001007097 | NTRK2 | Homo sapiens neurotrophic tyrosine kinase, receptor, type 2 (NTRK2), transcript variant b, mRNA |
| NM_001007156 | NTRK3 | Homo sapiens neurotrophic tyrosine kinase, receptor, type 3 (NTRK3), transcript variant 3, mRNA |
| NM_001008393 | C4orf46 | Homo sapiens chromosome 4 open reading frame 46 (C4orf46), mRNA |
| NM_001010915 | PTPLAD2 | Homo sapiens protein tyrosine phosphatase-like A domain containing 2 (PTPLAD2), mRNA |
| NM_001012446 | C9orf128 | Homo sapiens chromosome 9 open reading frame 128 (C9orf128), mRNA |
| NM_001012729 | DUXA | Homo sapiens double homeobox A (DUXA), mRNA |
| NM_001012959 | DISC1 | Homo sapiens disrupted in schizophrenia 1 (DISC1), transcript variant S, mRNA |
| NM_001018020 | TPM1 | Homo sapiens tropomyosin 1 (alpha) (TPM1), transcript variant 7, mRNA |
| NM_001029875 | RGS7BP | Homo sapiens regulator of G-protein signaling 7 binding protein (RGS7BP), mRNA |
| NM_001031835 | PHKB | Homo sapiens phosphorylase kinase, beta (PHKB), transcript variant 2, mRNA |
| NM_001037160 | CYS1 | Homo sapiens cystin 1 (CYS1), mRNA |

(Εικόνα 2.7:Αναζήτηση για το miRNA hsa-miR-155-3p στον PolymiRTS)

Για να δούμε τώρα συσχετίσεις μεταξύ ασθενειών επιλέγουμε την ασθένεια που θέλουμε π.χ. τον καρκίνο και παρατηρούμε με ποια γονίδια σχετίζεται , αν το γονίδιο αυτό αποτελεί και πιθανό στόχο του miRNA και εμφανίζει και πολυμορφισμό , τότε έχουμε εντοπίσει αυτό που θέλουμε.

| Transcript ID | Gene Symbol | Disease/Trait |
|---------------|-------------|--|
| NM_005845 | ABCC4 | Breast cancer |
| NM_006720 | ABLIM1 | Prostate cancer (gene x gene interaction) |
| NM_000970 | AC004086-1 | Esophageal cancer |
| NM_000690 | ALDH2 | Esophageal cancer |
| NM_001148 | ANK2 | Prostate cancer (gene x gene interaction) |
| NM_015245 | ANKS1A | Non-small cell lung cancer |
| NM_006305 | ANP32A | Esophageal cancer |
| NM_000044 | AR | Prostate cancer |
| NM_020732 | ARID1B | Pancreatic cancer |
| NM_020801 | ARRDC3 | Breast cancer (prognosis) |
| NM_005171 | ATF1 | Colorectal cancer |
| NM_018179 | ATF7IP | Testicular germ cell cancer |
| NM_001186 | BACH1 | Pancreatic cancer |
| NM_003658 | BARX2 | Breast cancer |
| NM_001714 | BICD1 | Pancreatic cancer |
| NM_182641 | BPTF | Lung cancer |
| NM_014292 | CBX6 | Bladder cancer |
| NM_001105563 | CCHCR1 | Prostate cancer |
| NM_053056 | CCND1 | Breast cancer |
| NM_001759 | CCND2 | Colorectal cancer |
| NM_001238 | CCNE1 | Bladder cancer |
| NM_000389 | CDKN1A | Colorectal cancer |
| NM_000077 | CDKN2A | Breast cancer |
| NM_000745 | CHRNA5 | Lung cancer |
| NM_030782 | CLPTM1L | Testicular germ cell cancer |
| NM_017623 | CNNM3 | Erectile dysfunction and prostate cancer treatment |
| NM_017738 | CNTLN | Ovarian cancer |
| NM_001122633 | CPS1 | Non-small cell lung cancer |
| NM_001025105 | CSNK1A1 | Esophageal cancer |
| NM_020943 | CWC22 | Prostate cancer (gene x gene interaction) |
| NM_173674 | DCBLD1 | Lung cancer |
| NM_001933 | DLST | Non-small cell lung cancer |
| NM_001080463 | DYNC2H1 | Small-cell lung cancer |
| NM_001002030 | ECHDC1 | Breast cancer |

(Εικόνα 2.8: Καταγεγραμμένες αλληλεπιδράσεις γονιδίων-ασθενειών στον PolymiRTS)
Στον παρακάτω πίνακα έχουμε μερικά συγκεντρωτικά αποτελέσματα

| Pathway ID | Pathway Class | Description | #PolymiRTS |
|------------|----------------------------------|--|------------|
| hsa04930 | Endocrine and metabolic diseases | Type II diabetes mellitus | 26 |
| hsa04940 | Endocrine and metabolic diseases | Type I diabetes mellitus | 9 |
| hsa04950 | Endocrine and metabolic diseases | Maturity onset diabetes of the young | 5 |
| hsa05010 | Neurodegenerative diseases | Alzheimer's disease | 93 |
| hsa05012 | Neurodegenerative diseases | Parkinson's disease | 68 |
| hsa05014 | Neurodegenerative diseases | Amnortrophic lateral sclerosis (ALS) | 25 |
| hsa05016 | Neurodegenerative diseases | Huntington's disease | 101 |
| hsa05020 | Neurodegenerative diseases | Prion diseases | 18 |
| hsa05100 | Infectious diseases: Bacterial | Bacterial invasion of epithelial cells | 49 |
| hsa05110 | Infectious diseases: Bacterial | Vibrio cholerae infection | 25 |
| hsa05120 | Infectious diseases: Bacterial | Epithelial cell signaling in Helicobacter pylori infection | 31 |
| hsa05130 | Infectious diseases: Bacterial | Pathogenic Escherichia coli infection | 37 |
| hsa05131 | Infectious diseases: Bacterial | Shigellosis | 50 |
| hsa05140 | Infectious diseases: Parasitic | Leishmaniasis | 31 |
| hsa05200 | Cancers: Overview | Pathways in cancer | 203 |
| hsa05210 | Cancers: Specific types | Colorectal cancer | 54 |
| hsa05211 | Cancers: Specific types | Renal cell carcinoma | 57 |
| hsa05212 | Cancers: Specific types | Pancreatic cancer | 60 |
| hsa05213 | Cancers: Specific types | Endometrial cancer | 42 |
| hsa05214 | Cancers: Specific types | Glioma | 47 |
| hsa05215 | Cancers: Specific types | Prostate cancer | 71 |
| hsa05216 | Cancers: Specific types | Thyroid cancer | 25 |
| hsa05217 | Cancers: Specific types | Basal cell carcinoma | 21 |
| hsa05218 | Cancers: Specific types | Melanoma | 45 |
| hsa05219 | Cancers: Specific types | Bladder cancer | 36 |
| hsa05220 | Cancers: Specific types | Chronic myeloid leukemia | 60 |
| hsa05221 | Cancers: Specific types | Acute myeloid leukemia | 45 |
| hsa05222 | Cancers: Specific types | Small cell lung cancer | 56 |
| hsa05223 | Cancers: Specific types | Non-small cell lung cancer | 45 |
| hsa05310 | Immune diseases | Asthma | 1 |
| hsa05320 | Immune diseases | Autoimmune thyroid disease | 5 |
| hsa05322 | Immune diseases | Systemic lupus erythematosus | 50 |
| hsa05330 | Immune diseases | Allograft rejection | 5 |
| hsa05332 | Immune diseases | Graft-versus-host disease | 7 |
| hsa05340 | Immune diseases | Primary immunodeficiency | 5 |

(Εικόνα 2.9: Συγκεντρωτικά αποτελέσματα ανά miRNA-ασθένεια στον PolymiRTS)

2.3 miRdSNP

Μία βάση δεδομένων για SNPs που σχετίζονται με ασθένειες και για περιοχές στόχους των miRNAs που εντοπίζονται στην 3' αμετάφραστη περιοχή των ανθρώπινων γονιδίων πληροφορίες για τα dSNPs έχουν εισαχθεί από την PubMed, ποιο συγκεκριμένα ο WebServer μας περιλαμβάνει 786 σχέσεις μεταξύ dSNPs και ασθενειών, εκ των οποίων σχέσεων προκύπτουν 630 μοναδικά dSNPs και 204 διαφορετικοί τύποι ασθενειών. Στόχος του miRdSNP είναι να προβάλλει όλες τις διαθέσιμες πηγές δεδομένων που υπάρχουν για τα dSNPs και πως επηρεάζουν της μετα-μεταγραφικές ρυθμίσεις των γονιδίων που σχετίζονται με ασθένειες και σε τι βαθμό απέχουν αυτά τα dSNPs από τις περιοχές στόχους των miRNAs. Οι αλγόριθμοι πρόβλεψης που χρησιμοποιούνται για να εντοπίσουν τις πιθανές περιοχές στόχους των miRNAs είναι οι αλγόριθμοι TargetScan(5.1) και PicTar. Για την καλύτερη αναζήτηση ο WebServer χρησιμοποιεί όλα τα γνωστά SNPs και ας μην σχετίζονται άμεσα με ασθένειες καθώς και γονίδια τα οποία έχουν αποδειχθεί πειραματικά ότι αποτελούν στόχους των miRNAs. Επιπλέον πληροφορίες για τα miRNAs εισάγονται από το miRBase 18, για τα SNPs από το NCBI dbSNP database 130, για τα γονίδια από το Refgene table of UCSC GenomeBrowser hg18, πειραματικά δεδομένα εισάγονται από τα TarBase, mirTarBase, miRecords, miR2disease καθώς και πληροφορίες για τις ασθένειες από το HarMap Project(rel 27). Ο συγκεκριμένος WebServer χρησιμοποιεί 19,834 διαφορετικά γονίδια μαζί με τις συντεταγμένες για το χρωμόσωμα στο οποίο βρίσκονται. Αν ένα γονίδιο διαθέτει παραπάνω από μία μεταγραφικές περιοχές επιλέγεται εκείνη με το μεγαλύτερο μήκος. Από ένα σύνολο 18,833,531 SNPs, καταλήγουμε σε ένα υποσύνολο 175,351 όλα στην 3' αμετάφραστη περιοχή σε σύνολο 16,810 γονιδίων. Επιπλέον πληροφορίες εισάγονται από 2,785 δημοσιεύσεις που αναφέρονται σε 16,447 μοναδικά SNPs, στην συνέχεια καταλήξαμε μέσω αυτών των δημοσιεύσεων στην καταγραφή 630 dSNPs για 204 διαφορετικές ανθρώπινες ασθένειες.

| | |
|---------------------------|-----------|
| Αριθμός Γονιδίων | 19.834 |
| Αριθμός SNPs 3'UTR | 174,351 |
| Disease-SNP | 786 |
| dSNPs | 630 |
| Ασθένειες | 204 |
| Ζευγάρια SNP-miRNA | 3.361.770 |
| miRNAs | 249 |
| miRNA Targets | 117,929 |

(Πίνακας 2.4:Αποθηκευμένα δεδομένα στον miRdSNP)

Όπως φαίνεται στην παρακάτω εικόνα υπάρχει η δυνατότητα αναζήτησης με μια πληθώρα επιλογών τα πεδία αναζήτησης περιλαμβάνουν , γονίδιο , miRNA , Μέθοδο πρόβλεψης , SNP , ασθένεια

The image shows a search interface with five input fields: 'Gene', 'miRNA', 'Prediction Method', 'SNP', and 'Disease'. Each field is represented by a rectangular box with a small downward arrow on the right side, indicating a dropdown menu. The 'miRNA' field is currently selected and highlighted with a blue border.

(Εικόνα 2.10: Επιλογές αναζήτησης στον miRdSNP)

Ο παρακάτω πίνακας αποτελεί το συγκεντρωτικό αποτέλεσμα της αναζήτησης για το miR-155 , το αποτέλεσμα είναι ταυτόχρονα για γονίδια, SNPs και ασθένειες. Επίσης υπάρχει και η ένδειξη για το αν οι συσχετίσεις αυτές είναι και πειραματικά επιβεβαιωμένες.

| Gene | refseq_id | miR | SNP | Diseases | Distance | experimentally_confirmed |
|--------|--------------|-------------|-----------|------------------------------|----------|--------------------------|
| APC | NM_000038 | hsa-miR-155 | rs41116 | Breast cancer | 127 | Yes |
| CHAF1A | NM_005483 | hsa-miR-155 | rs2992 | Prostate cancer | -184 | Yes |
| ETS1 | NM_001143820 | hsa-miR-155 | rs1128334 | Systemic lupus erythematosus | 298 | No |
| ETS1 | NM_001143820 | hsa-miR-155 | rs1128334 | Systemic lupus erythematosus | 298 | No |

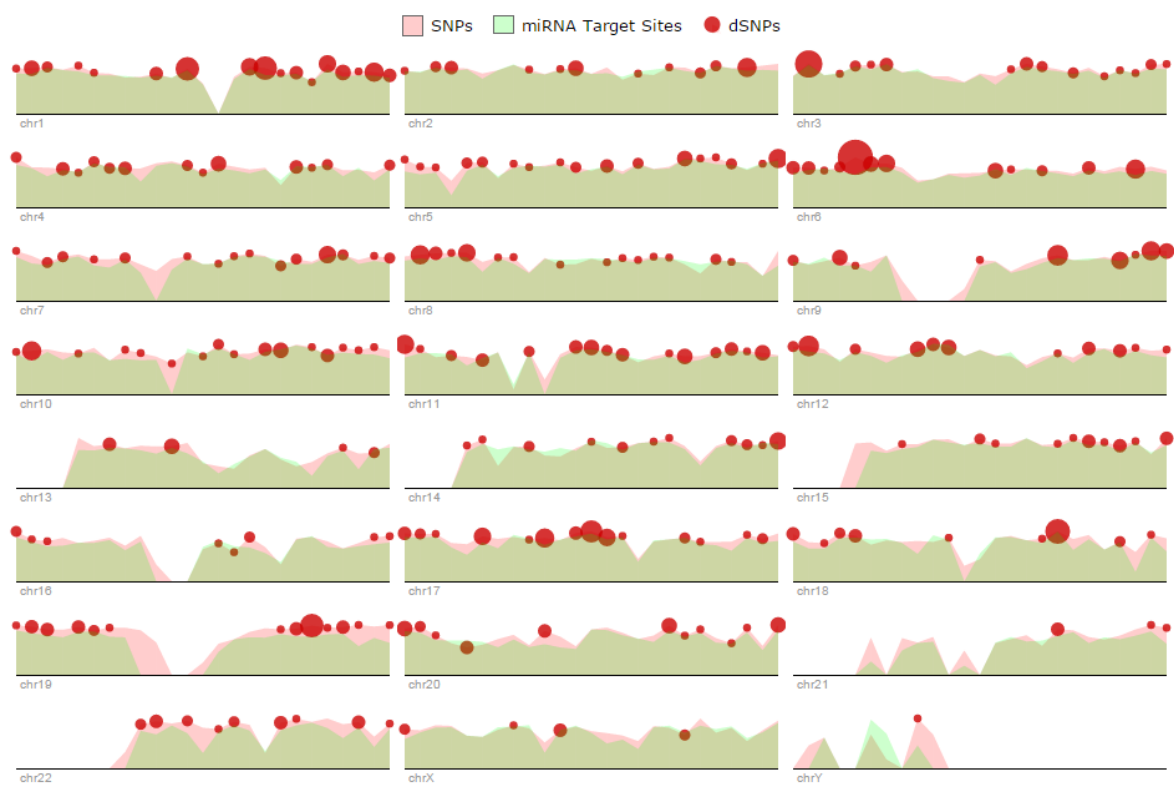
| | | | | | | |
|-------|--------------|-------------|------------|------------------------------|-------|-----|
| CCND1 | NM_053056 | hsa-miR-155 | rs678653 | Breast cancer | 324 | Yes |
| BDNF | NM_170731 | hsa-miR-155 | rs7124442 | Depression, Eating Disease | -470 | No |
| APC | NM_000038 | hsa-miR-155 | rs1804197 | Autism | -879 | Yes |
| BDNF | NM_170731 | hsa-miR-155 | rs11030100 | Asthma | -1015 | No |
| GNAS | NM_016592 | hsa-miR-155 | rs7121 | Prostate cancer | -1112 | No |
| MEF2A | NM_001130928 | hsa-miR-155 | rs1059759 | Coronary artery disease | 1351 | No |
| GPR85 | NM_001146266 | hsa-miR-155 | rs56080411 | Schizophrenia | 1404 | No |
| BAG5 | NM_004873 | hsa-miR-155 | rs7693 | Non-Hodgkin lymphoma | 2141 | No |
| BDNF | NM_170731 | hsa-miR-155 | rs41282918 | Major depressive disorder | -2209 | No |
| ETS1 | NM_001143820 | hsa-miR-155 | rs1128334 | Systemic lupus erythematosus | 2370 | No |
| CCND1 | NM_053056 | hsa-miR-155 | rs7178 | Ovarian cancer | 2617 | Yes |

| | | | | | | |
|----------|------------------|---------------------|--------|-----------------|------|----|
| C10orf26 | NM_00108391 3 | hsa- miR- 155 | rs6892 | Prostate cancer | 2710 | No |
|----------|------------------|---------------------|--------|-----------------|------|----|

Πίνακας 2.5:(Συγκεντρωτικά αποτελέσματα της αναζήτησης του hsa-miR-155 στον miRdSNP)

Παρατηρούμε λοιπόν ότι το αποτέλεσμα δίνεται σε μορφή πίνακα. Στην πρώτη στήλη αναγράφεται το γονίδιο στο οποίο βρισκόμαστε , ο κωδικό του γονιδίου , το όνομα του miRNA για το οποίο εκτελούμε την αναζήτηση , το SNP με το οποίο σχετίζεται , η απόσταση του SNP από την περιοχή στόχο του miRNA. Επιπλέον όσα εγγραφές του πίνακα έχουν την πράσινη ένδειξη σημαίνει ότι είναι πειραματικά αποδεδειγμένα. Επίσης μας δίνεται η δυνατότητα ώστε να δούμε την ακολουθία του miRNA για το οποίο κάνουμε την αναζήτηση

Ο Webserver επιτρέπει ακόμα την αναζήτηση ανά χρωμόσωμα , απευθείας σε όποια θέση θέλει ο χρήστης . Παρατηρούμε στην παρακάτω φωτογραφία την γραφική απεικόνιση όλων των χρωμοσωμάτων του ανθρώπου. Όπως βλέπουμε σε κάθε χρωμόσωμα υπάρχουν τρεις διαφορετικές χρωματικές ενδείξεις για τα SNPs , τους στόχους των miRNAs και τα dSNPs , δηλαδή τα SNPs τα οποία σχετίζονται με ασθένειες.



(Εικόνα 2.11: Γραφική απεικόνιση των SNPs, των target sites και των dSNPs ανά χρωμόσωμα στον άνθρωπο)

Οι ασθένειες στις οποίες γίνονται οι περισσότερες αναφορές

| | |
|------------------------------|----|
| Καρκίνος του Μαστού | 75 |
| Σχιζοφρένεια | 43 |
| Διαβήτης τύπου 2 | 28 |
| Πάρκινσον | 20 |
| Systemic lupus erythematosus | 18 |
| Alzheimer | 18 |
| Παχυσαρκία | 16 |
| Καρκίνος του Προστάτη | 15 |
| HIV-1 | 14 |
| Άσθμα | 13 |
| Καρκίνος του Πνεύμονα | 13 |
| Μυοκαρδίου | 11 |
| Καρκίνος του παχέος εντέρου | 11 |
| Ρευματική Αρθρίτιδα | 11 |
| Αυτισμός | 10 |

(Πίνακας 2.6: Ασθένειες που σχετίζονται με τον μεγαλύτερο αριθμό miRNAs στον miRdSNP)

2.4 dbSMR

Μία OnLine πηγή πληροφοριών , με μεγάλο αριθμό SNPs τα οποία επηρεάζουν την επίδραση των miRNAs.

Οι πληροφορίες για την ονομασία και τους πιθανούς στόχους κάθε ανθρώπινου miRNA έχουν εισαχθεί από το miRBase 9 , βρίσκονται όλα στην 3' αμετάφραστη περιοχή και επιπλέον πληροφορίες για τα γονίδια έχουν εισαχθεί από την βάση δεδομένων Ensembl. Εξαιτίας του γεγονότος ότι πολλά εργαλεία-αλγόριθμοι πρόβλεψης πιθανών στόχων των miRNAs σχετίζονται με λανθασμένες προβλέψεις αρκετών ζευγαριών miRNAs-στόχων, ο WebServer χρησιμοποιεί το συνδυαστικό αποτέλεσμα 3 διαφορετικών αλγορίθμων για να εξαντλήσει την πιθανότητα λάθους πρόβλεψης , οι αλγόριθμοι που χρησιμοποιεί είναι οι miRanda, RNAHybrid και TargetScan. Χρησιμοποιήθηκαν μόνο όσα ζευγάρια ήταν κοινά στο αποτέλεσμα και των 3 εργαλείων πρόβλεψης.

Στην συνέχεια και έχοντας καταλήξει σε έναν αριθμό 470 ανθρώπινων miRNAs και 22,234 περιοχών στόχων , εντοπίζονται όλα τα SNPs τα οποία βρίσκονται ±200 βάσεις απόσταση από κάθε περιοχή στόχο, καταλήγοντας σε έναν αριθμό 30,432 SNPs. Στην συνέχεια χωρίζονται σε δύο υποσύνολα , εκείνο με την τυπική-φυσιολογική του μορφή(wild type) και εκείνο με την πολυμορφική-αλλαγμένη μορφή(polymorphic type). Υπολογίζονται στην συνέχεια με την βοήθεια των μαθηματικών ο βαθμός αλλαγής μια περιοχής κάνοντας χρήση μόνο του υποσυνόλου με τα wild types , καθώς με τα φυσιολογικά εκ των πραγμάτων δεν υπάρχουν αλλαγές στην ακολουθία.

$$\text{Degree of Change} = \frac{\text{Number of bases changing the conformation}}{\text{Total number of bases involved in miR binding}}$$

Αντιλαμβανόμαστε ότι ο παρονομαστής παραμένει σταθερός , συνεπώς όσες περισσότερες βάσεις αλλάζουν, τόσο μεγαλύτερος γίνεται ο βαθμός αλλαγής και συνεπώς τόσο πιθανότερο είναι , ένα miRNA να «χάσει» μια περιοχή στόχο εξαιτίας του πολυμορφισμού. Για να πάρουμε πληροφορίες για τα ζεύγη miRNAs-στόχων που έχουν επιβεβαιωθεί πειραματικά χρησιμοποιούμε την βάση δεδομένων TarBase Επιπλέον λογισμικό που χρησιμοποιείται είναι το RNAfold program. Στο στάδιο αφότου έχουμε χωρίσει τα SNPs που εντοπίσαμε σε 2 υποσύνολα , το συγκεκριμένο πρόγραμμα εντοπίζει περισσότερα ζεύγη-βάσεων.

2.5 miRvar

Η συλλογή των δεδομένων για τον συγκεκριμένο WebServer έγινε εξίσου από OnLine πηγές αλλά και από την πρόσφατη βιβλιογραφία , δηλαδή της δημοσιεύσεις. Για την ονοματολογία και για τις πληροφορίες των γενετικών παραλλαγών δηλαδή στην προκειμένη περίπτωση των SNPs , χρησιμοποιήθηκε το NCBI dbSNP database 130. Οι απαραίτητες αλληλουχίες βάσεων «κατέβηκαν» από το Refgene table of UCSC GenomeBrowser hg18 , η αποσφαλμάτωση των παραπάνω δειγμάτων έγινε μέσω προγραμμάτων(scripts) τα οποία είναι γραμμένα σε γλώσσα Perl και μπορεί κανείς να τα δει αναλυτικά στην διεύθυνση (<http://genome.igib.res.om/mirvar/pipeline.com>). Για τις πληροφορίες σχετικά με την ονοματολογία των miRNAs αλλά και των pre-miRNAs χρησιμοποιήθηκε όπως πλέον είναι φανερό η βάση miRBase 13 , για βελτίωση των ονομάτων και των δεδομένων χρησιμοποιήθηκαν επιπλέον πληροφορίες όπως ορίζονται από τον HUGO , έχοντας πάντα υπόψιν τους κανόνες που ορίζουν διεθνείς οργανισμοί όπως είναι ο Human Genome Variation Society. Η βάση δεδομένων στηρίζεται από το σύστημα LOVD(Leiden Open Variation Database) , το οποίο είναι διαδραστικό με τον χρήστη και παρέχει ξεχωριστές σελίδες και παραπομπές για κάθε γονίδιο ξεχωριστά και για όλες τις πιθανές παραλλαγές του. Επιπλέον πληροφορίες πάρθηκαν από τις βάσεις δεδομένων miR2Diseases, miRecords , 1000 Genome Project. Η χαρτογράφηση των SNPs από το dbSNP και το UCSC Genome Browser κατέληξε στον

εντοπισμό 106 SNPs που σχετίζονται με 86 διαφορετικά miRNAs, τα SNPs αυτά βρίσκονται σε διαφορετικές περιοχές των miRNAs και όχι πάντα στην ίδια. Στην συνέχεια χρησιμοποιήθηκαν μια σειρά από επιπλέον εργαλεία/λογισμικά και μεθόδους για να εντοπίσουμε miRNAs με ποιο συγκεκριμένα χαρακτηριστικά. Για παράδειγμα μέσω του εργαλείου-αλγορίθμου PHDclean , εντοπίσαμε 2 miRNAs, τα οποία στο σημείο εκείνο της αλληλουχίας που τα «κόβει» η πρωτεΐνη Dicer ήταν διαφοροποιημένα. Παρατηρούμε πως δίνεται η δυνατότητα στον χρήστη να επιλέξει από την λίστα των miRNAs τα οποία διαθέτει ο WebServer, για το hsa-miR-155 δεν υπάρχει καταγραφή στον WebServer , συνεπώς επιλέγουμε την αναζήτηση για το hsa-miR-151

| General information | |
|--|---|
| Gene name | hsa-mir-151 |
| Gene symbol | MIR151 |
| Chromosome Location | 8 |
| Database location | http://fishmap2.igib.res.in/mirlovd |
| Curator | Deeksha Bhartiya |
| Date of creation | May 08, 2010 |
| Last update | December 15, 2010 |
| Version | MIR151 101215 |
| Add sequence variant | Submit a sequence variant |
| First time submitters | Register here |
| Reference sequence | Genomic reference sequence for describing sequence variants |
| Total number of unique DNA variants reported | 1 |
| Total number of variants reported | 3 |
| Subscribe to updates of this gene |  |

(Εικόνα 2.11:Αποτέλεσμα αναζήτησης στον miRvar)

Στην συνέχεια και αφού έχουμε επιλέξει γονίδιο , δηλαδή κάποιο γνωστό miRNA , μας εμφανίζει πληροφορίες σχετικές με αυτό , όπως είναι , το όνομα του, ο συμβολισμός του , το χρωμόσωμα στο οποίο εντοπίζεται , ένα εσωτερικό link για το που είναι αποθηκευμένο στον Server, ημερομηνία που δημιουργήθηκε η εγγραφή κ.α. Επίσης για περισσότερες πληροφορίες δίνονται όλα τα link των εξωτερικών πηγών.

| | |
|-----------------------------|---|
| miRBase | http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=hsa-let-7a-1 |
| UCSC Genome Browser | http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg18&position=chr9%3A95978060-95978139&hgside=469623433_1IthFpnDpoENVCXfX11NkQfo89a1 |
| Mir2Disease | http://watson.compbio.iupui.edu:8080/miR2Disease/searchMiRNA.jsp?checkbox=hsa-let-7a-1 |
| 1000 Genomes Browser | http://browser.1000genomes.org/Homo_sapiens/Location/View?db=core;r=9:95978060-95978139 |
| miRecords | |
| HapMap | http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap27_B36/?name=Chr9:95978060-95978139 |

(Πίνακας 2.7: Εξωτερικές πηγές πληροφοριών που χρησιμοποιεί ο miRvar)

2.6 MicroSniper

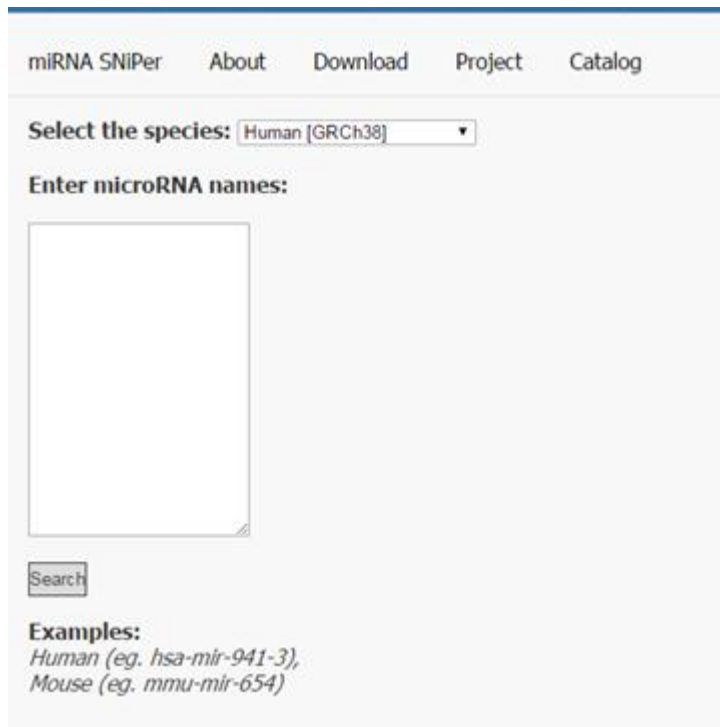
Ένα εργαλείο για την πρόβλεψη των επιδράσεων των SNPs σε περιοχές στόχους των miRNAs. Δίνει την δυνατότητα δηλαδή στον χρήστη να εκτιμήσει για κάθε SNP ξεχωριστά, την επίδραση σου σε μια περιοχή στόχο(3' UTR) και να παρατηρήσει κατά πόσο μια τέτοια περιοχή δημιουργείται η καταστρέφεται λόγω της παρουσίας των πολυμορφισμών. Η συγκεκριμένη πλατφόρμα δίνει αρκετές επιλογές εισόδου για την επιλογή περιοχών στόχων δηλαδή 3' UTRs, ο χρήστης μπορεί να επιλέξει μεταξύ του University of California Santa Cruz(UCSC), του HarMap Project η μπορεί να κάνει μια χειροκίνητη αναζήτηση. Να αναφέρουμε επίσης ότι επιπλέον πληροφορίες για τα miRNAs και το γονιδίωμα , εισάγονται από το miRBase 21 και το Ensembl 78 , αντίστοιχα. Ο αλγόριθμος πρόβλεψης που χρησιμοποιείται για τον εντοπισμό των 3'UTRs είναι ο TargetScan 6.2

Αρχικά ο χρήστης καλείται να επιλέξει , αν θα κάνει χειροκίνητη ή αυτόματη αναζήτηση για κάποιο miRNA. Στην πρώτη περίπτωση μπορεί να πληκτρολογήσει αυτός μια ακολουθία και στην άλλη να φορτώσει μια από τις ήδη υπάρχουσες. Η δυνατότητα αναζήτησης είναι εφικτή μέσω του ονόματος του miRNA , του Refseq_id και του ονόματος του SNP. Για την αποθήκευση των 3' UTRs και των SNPs δημιουργήθηκε μια βάση δεδομένων SQL , μέσω του λογισμικού MySQL DBMS. Από το σύνολο των 3'UTRs και το SNPs που έγιναν download από το dbSNP (version 29) και από το HarMap Project μέσω της χρήσης του UCSC, μόνο εκείνα τα SNPs τα οποία βρίσκονται μέσα στις περιοχές στόχους εν τέλει αποθηκεύτηκαν στην βάση δεδομένων. Υπάρχουν συνολικά 18,026 γονίδια και 26.357 3'UTRs αποθηκευμένα στην βάση μας.

| | |
|-------------------------------------|--------|
| Αριθμός miRNAs | 1881 |
| SNPs in pre-miRNAs | 1532 |
| SNPs in mature region | 991 |
| SNPs in seed region | 437 |
| miRNA target sites (3' UTRs) | 26.357 |
| Number of Genes | 18.026 |

(Πίνακας 2.8: Δεδομένα που είναι αποθηκευμένα στον MicroSniper)

Στην παρακάτω φωτογραφία βλέπουμε την κεντρική σελίδα του miRNA Sniper , μπορούμε να επιλέξουμε , μεταξύ 15 διαφορετικών ειδών και στην συνέχεια να δώσουμε το όνομα του miRNA για το οποίο θέλουμε να αντλήσουμε πληροφορίες



(Εικόνα 2.12: Interface του MicroSniper)

Έστω για παράδειγμα ότι θέλουμε να βρούμε πληροφορίες , όπως και στα προηγούμενα παραδείγματα για το miRNA hsa-mir-155

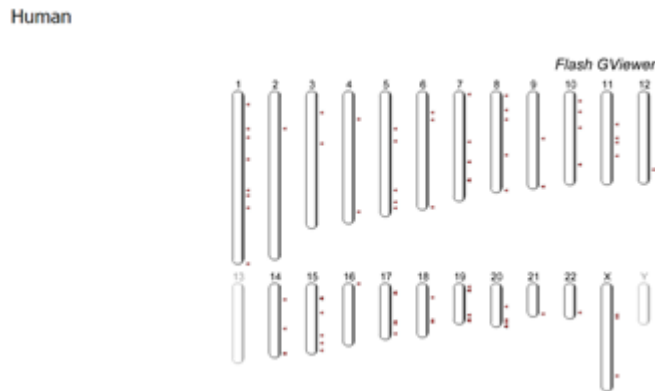
| | | | | |
|-------------|----------------|-------------|-------|------------|
| hsa-mir-155 | hsa-miR-155-5p | rs140021681 | T > C | pre-mature |
| hsa-mir-155 | hsa-miR-155-5p | rs200351615 | A>G | pre-mature |
| hsa-mir-155 | hsa-miR-155-5p | rs377265631 | A>G | pre-mature |
| hsa-mir-155 | hsa-miR-155-3p | rs140021681 | T>C | pre-mature |
| hsa-mir-155 | hsa-miR-155-3p | rs200351615 | A>G | seed |
| hsa-mir-155 | hsa-miR-155-3p | rs377265631 | A>G | mature |

(Πίνακας 2.9: Πολυμορφισμοί του hsa-miR-155 στον MicroSniper)

Το αποτέλεσμα της αναζήτησης είναι ένας πίνακας που μας δίνει πληροφορίες για τα SNPs που εμφανίζονται στο συγκεκριμένο miRNA. Υπάρχουν λοιπόν πληροφορίες για το όνομα του miRNA για το οποίο κάνουμε την αναζήτηση , για το ώριμο miRNA από το

οποίο προέρχεται , για το όνομα του SNP , για την βάση η οποία αλλάζει , καθώς και για τον τύπο.

Επίσης ο WebServer διαθέτει τα χρωμοσώματα του κάθε οργανισμού και έτσι ο χρήστης μπορεί να πάρει επιπλέον πληροφορίες για την θέση κάθε miRNA , αλλά και για τον αριθμό των miRNAs που βρίσκονται σε κάθε χρωμόσωμα.



(Εικόνα 2.13: Γραφική απεικόνιση των miRNAs ανά χρωμόσωμα)

2.7 mrSNP

Ένα λογισμικό εντοπισμού των επιπτώσεων των SNPs σε περιοχές στόχους των miRNAs. Οι αλληλουχίες 3'UTRs και οι πληροφορίες σχετικά με αυτές κατέβηκαν από την βάση δεδομένων UCSC , μέσω της χρήσης του Genome Table Browser. Κάθε χρωμόσωμα αποθηκεύεται σε ξεχωριστό αρχείο , όπου κάθε αρχείο έχει πληροφορίες σχετικές με το όνομα του κάθε γονιδίου και τις συντεταγμένες της κάθε 3' UTR περιοχής μέσα στο χρωμόσωμα. Οι πληροφορίες για την ονοματολογία των miRNAs κατέβηκαν από το miRBase και ταξινομήθηκαν ανάλογα με την διατήρησή τους, δηλαδή την εμφάνιση τους ανάμεσα στους διάφορους οργανισμούς με βάση τις προβλέψεις του εργαλείου-αλγόριθμου TargetScan.

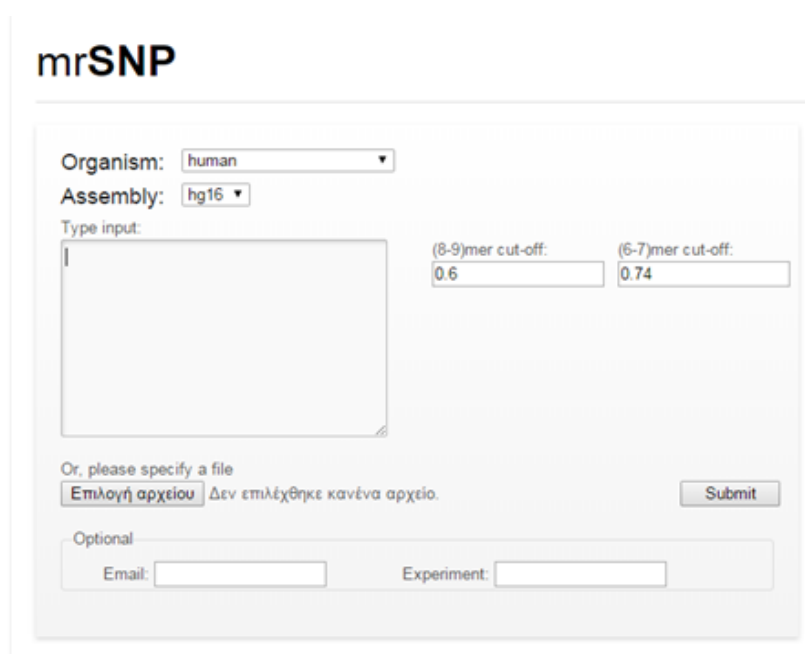
Η είσοδος μπορεί να δοθεί με αρκετούς τρόπους , το όνομα του SNP , την μεθοδολογία με την οποία έγινε η χαρτογράφηση του DNA, το χρωμόσωμα στο οποίο εντοπίζεται το SNP, η θέση του SNP στο χρωμόσωμα και τις διάφορες παραλλαγές του SNP.

Στην συνέχεια το λογισμικό προσπαθεί να εντοπίσει την αλληλουχία SNP , για την οποία κάνουμε την αναζήτηση. Σε περίπτωση που το SNP δεν ανήκει σε κάποιο καταγεγραμμένο 3'UTR , η διαδικασία σταματάει εκεί και το λογισμικό βγάζει σχετικό μήνυμα. Σε αντίθετη περίπτωση επιστρέφεται μια ακολουθία 79 βάσεων , στο κέντρο της οποίας βρίσκεται το SNP , για το οποίο ψάχνουμε. Η λογική με την οποία επιλέχθηκε το συγκεκριμένο μήκος έχει ως εξής. Το μέσο miRNA έχει μήκος 25βάσεων και ένα

στέλεχος βρόχο(loop) μήκους 15 βάσεων , συνεπώς η μέγιστη απόσταση από την αρχή ή το τέλος του miRNA θα είναι 39 βάσεις , οπότε $25+15+39=79$ για αυτόν τον λόγο επιλέγεται αυτό το μήκος.

Στο τελευταίο μας βήμα για κάθε miRNA που έχουμε αποθηκεύσει , ελέγχουμε για την ύπαρξη μια αλληλουχίας 6 συνεχόμενων ζευγών βάσεων, το κάνουμε και για 7,8,9. Αν μια περιοχή δεν έχει κατ' ελάχιστο ταύτιση 6 συνεχόμενων βάσεων τότε δεν αποτελεί περιοχή στόχο του miRNA. Αντίθετα αν βρεθούν 7 συνεχόμενα ζεύγη βάσεων στην seed περιοχή του miRNA , τότε κατατάσσεται αυτομάτως σαν περιοχή ενδιαφέροντος.

Τέλος χρησιμοποιείτε το λογισμικό RNAHybrid για τον υπολογισμό του εκάστοτε binding Energy , όταν το εκάστοτε binding Energy είναι μεγαλύτερο από το 74% του συνολικού , τότε η συγκεκριμένη περιοχή(SNP) , αποτελεί περιοχή στόχο του miRNA.



(Εικόνα 2.14: Interface του mrSNP)

Στην παραπάνω φωτογραφία , βλέπουμε το interface του συγκεκριμένου WebServer .Όπως αναφέραμε και νωρίτερα ο χρήστης αρχικά επιλέγει το είδος του οργανισμού στον οποίο θέλει να κάνει την αναζήτηση και την μεθοδολογία με την οποία έχει γίνει η χαρτογράφηση. Επίσης μπορεί να θέσει τιμές στο binding energy , όπως βλέπουμε το default energy για τα 6μερή και τα 7μερή είναι στο 74% . όπως αναφέραμε, ενώ για τα 8μερη και τα 9μερη είναι το 60%. Επίσης παρατηρούμε ότι η αναζήτηση μπορεί να γίνει είτε χειροκίνητα , είτε μέσω αρχείου. Δίνεται ακόμα η δυνατότητα επιλογής φίλτρου στο αποτέλεσμα της αναζήτησης.

Αποτέλεσμα αναζήτησης για input: “chr15 69018917 C T”, δηλαδή κάνουμε αναζήτηση στο chr15 , για τον εντοπισμό του SNP rs69018917.

mrSNP Results:

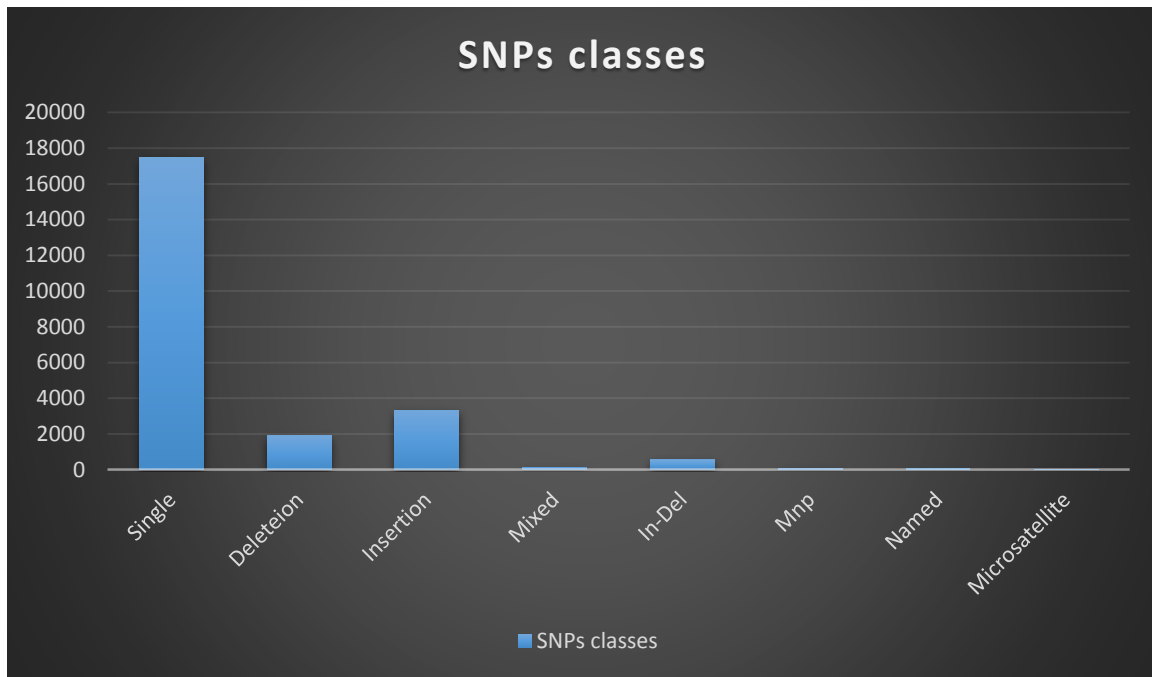
| Chromosome | SNP Position | Target Gene | microRNA | Strand | Binding Energy Difference | SNP1 | SNP1 Binding Energy | Cut Off 1 | Alignment 1 | is Binding 1 Valid |
|------------|--------------|-------------|-----------------|--------|---------------------------|------|---------------------|-----------|-------------|--------------------|
| chr15 | 69018917 | CORO2B | hsa-miR-2661 | * | 0.0 | C | -41.5 | 0 | N/A | True |
| chr15 | 69018917 | CORO2B | hsa-miR-4271 | + | 0.0 | C | -43.3 | 0 | N/A | True |
| chr15 | 69018917 | CORO2B | hsa-miR-3936 | + | 0.0 | C | -47.2 | 0 | N/A | True |
| chr15 | 69018917 | CORO2B | hsa-miR-4747-5p | + | 0.0 | C | -48.0 | 0 | N/A | True |
| chr15 | 69018917 | CORO2B | hsa-miR-432 | + | 31.5 | C | -49.2 | 0 | N/A | True |
| chr15 | 69018917 | CORO2B | hsa-miR-2110 | + | 0.0 | C | -50.6 | 0 | N/A | True |

(Εικόνα 2.15: Αποτέλεσμα αναζήτησης για το SNP 69018917 στον mrSNP)

Στο αποτέλεσμα της αναζήτησης παίρνουμε αρκετές πληροφορίες σχετικά με το SNP. Συγκεκριμένα το γονίδιο που βρίσκεται στην συγκεκριμένη περιοχή του χρωμοσώματος που στην περίπτωση την δική μας είναι το CORO2B , τα miRNAs που η συγκεκριμένη περιοχή αποτελεί στόχο και κάποιες επιπλέον πληροφορίες σχετικές με το SNP.

2.8 dPore

Σε αυτόν τον WebServer η ανάλυση και η συλλογή των αντικειμένων προέκυψε με τον εξής τρόπο. Αρχικά ενδιαφερόμαστε μόνο για SNPs τα οποία έχει προβλέψει και έχει επιβεβαιώσει πειραματικά το TarBase και εάν επικαλύπτονται με περιοχές στόχους των miRNAs. Τα SNPs και κατ' επέκταση τα miRNAs που σχετίζονται με ασθένειες εντοπίστηκαν αποκλειστικά από την βιβλιογραφία μέσω παλαιότερων πειραμάτων που έχουν γίνει για παράδειγμα σε δείγμα ασθενών με myasthenia gravis, νόσο του Κρον και διάφορους τύπους καρκίνων. Οι περιοχές του γονιδιώματος (ολόκληρες περιοχές), χρησιμοποιήθηκαν από το Human Genome build HG18. Το αρχικό δείγμα των SNPs που πάρθηκε από το dbSNP ήταν 18,833,531 και κατέληξαν μόλις σε 22.325 που βρίσκονται στις περιοχές στόχους των miRNAs. Αναλυτικά τα είδη των SNPs παρουσιάζονται στο παρακάτω γράφημα.



(Γράφημα 2.1: Είδη των SNPs στον dPore)

Το αποτέλεσμα της αναζήτησης για το has-miR-155

Details of miRNA *hsa-mir-155*

Associated information for miRNA:

Associated diseases: 35

Associated targets: 96

Associated pathways: 64

5 promoters found

| View SNP/TFBS associations | Chromosome | Strand | Start | Stop | Reference | UCSC Genome Browser |
|--|------------|--------|----------|----------|------------------------|---|
| | chr21 | + | 25867713 | 25868062 | Zhou et al. | view annotated promoter |
| | chr21 | + | 25856186 | 25856386 | Marson et al. | view annotated promoter |
| | chr21 | + | 25867620 | 25868098 | Marson et al. | view annotated promoter |
| | chr21 | + | 25863163 | 25868227 | UCSC | view annotated promoter |
| | chr21 | + | 25851291 | 25856290 | Ozsolak et al. (MALME) | view annotated promoter |

(Εικόνα 2.16: Αποτέλεσμα για την αναζήτηση του hsa-miR-155 στον dPore)

Παρατηρούμε λοιπόν το εξής , στην έξοδο έχουμε όλες τις αναφορές που με το συγκεκριμένο miRNA και τις ασθένειες τις οποίες σχετίζεται

| Diseases |
|--|
| Pancreatic cancer |
| Glioblastoma multiforme, somatic |
| Hepatocellular carcinoma |
| Down syndrome, risk of |
| Colorectal cancer |
| Breast cancer |
| Gastrointestinal |
| Thyroid carcinoma, papillary |
| Thyroid carcinoma, follicular |
| Cancer |
| Lung cancer |
| Leukemia, acute myeloid |
| Leukemia, chronic myeloid |
| Multiple myeloma |
| Leukemia/lymphoma, chronic B-cell |
| Ovarian cancer |
| Burkitt lymphoma |
| Rapp-Hodgkin syndrome |
| Squamous cell carcinoma, head and neck |
| Rheumatoid arthritis, susceptibility to |
| Non-Hodgkin lymphoma, somatic |
| Leukemia, chronic lymphatic, susceptibility to |

| |
|---|
| Cervical cancer, somatic |
| Duchenne muscular dystrophy |
| Muscular |
| Leukemia, acute myelogenous |
| Muscular dystrophy, limb-girdle, type 2A |
| Muscular dystrophy, limb-girdle, type 2B |
| Miyoshi myopathy |
| Myopathy, nemaline, 3 |
| Myelofibrosis, idiopathic |
| Melanoma and neural system tumor syndrome |
| Prostate cancer |
| Hematological |

(Πίνακας 2.10: Ασθένειες που σχετίζονται με το hsa-miR-155 στον dPore)

Περιοχές στόχοι δηλαδή γονίδια και τα αντίστοιχα pathways , (ενδεικτικά από σύνολο 96):

TACD2_HUMAN
CEBPB_HUMAN
RAB6A_HUMAN
VATC1_HUMAN
5NTD_HUMAN
DNJB1_HUMAN
AGTR1_HUMAN
CTNB1_HUMAN
KPCI_HUMAN
MATR3_HUMAN
RAB5C_HUMAN
PPP5_HUMAN
ARFP2_HUMAN
ARFP1_HUMAN
DCAF7_HUMAN
PTPRJ_HUMAN
MYO1E_HUMAN
TRI32_HUMAN

CAF1A_HUMAN
 PICAL_HUMAN
 CUL4B_HUMAN

Τέλος τις περιοχές που εμφανίζουν πολυμορφισμό και πληροφορίες για κάθε SNP

| SNPs overlapping with existing TFBS: | | | | | | |
|--------------------------------------|------------|-----------|----------------|-------------|---------------------------------|--|
| View details | SNP | SNP class | UCSC reference | observation | absolute position on chromosome | effect(s) on TFBS(s) [?] |
| | rs5843159 | insertion | - | G | 25856231 | loss |
| | rs66843086 | insertion | - | G | 25856232 | unmodified |
| | rs72558339 | single | G | C | 25856234 | unmodified, loss, modified |

(Εικόνα 2.17: SNPs στον dPore για το hsa-miR-155)

2.9 miRNASNP

Ο συγκεκριμένος WebServers εντοπίζει πολυμορφισμούς στα ίδια τα miRNA και στις περιοχές στόχους των miRNAs. Τις πληροφορίες για τα miRNAs τις φόρτωσε από το miRBase και αντίστοιχα για τα SNPs από την βάση dbSNP137.

Αποτελείται από δύο κομμάτια την βάση δεδομένων που εμπεριέχει το σύνολο των SNPs. Εντοπίζει 2257 SNPs τα οποία προέρχονται από ένα σύνολο 1596 διαφορετικών miRNAs. Επίσης κατηγοριοποιεί τα SNPs ανάλογα με την περιοχή στην οποία εντοπίστηκαν και συγκεκριμένα 706 βρίσκονται στην “ώριμη” περιοχή ενώ 227 στην seed περιοχή των miRNAs, ταυτόχρονα επιτρέπει την αναζήτηση σε άλλα 8 είδη πέρα του ανθρώπου. Το δεύτερο κομμάτι είναι που συμμετέχει και ο χρήστης και μπορεί να κάνει την αναζήτηση του. Στην παρακάτω εικόνα παρατηρούμε το αποτέλεσμα της αναζήτησης για το hsa-mir-155, όπως αναφέραμε και νωρίτερα κατηγοριοποιεί τα SNPs ανάλογα με την περιοχή στην οποία βρίσκονται.

SNPs in human pre-miRNA regions

| Pre-miRNA | SNP ID | Chromosome | SNP Positon | Allele | Location | Energy change(kcal/mol) | Predicted product |
|-------------|-------------|------------|-------------|--------|-----------|-------------------------|-------------------|
| hsa-mir-155 | rs140021681 | chr21 | 26946325 | T/C | pre-miRNA | 0 | mild |
| hsa-mir-155 | rs200351615 | chr21 | 26946341 | A/G | in_seed | 1.4 | mild |

SNPs in the pre-miRNA flanking regions (-1kb ~ +1kb)

| miRNA | ACC | Chromosome | Start | End | miRNA strand | Numbers of SNPs up/down |
|-------------|-----------|------------|----------|----------|--------------|-------------------------|
| hsa-mir-155 | MI0000681 | chr21 | 26946292 | 26946356 | + | 17/10 |

(Εικόνα 2.18: SNPs ανά περιοχή για το hsa-miR-155 στον miRNASNP)

| miRNA | miRNA exp. \ominus | SNP in gene 3'UTR | Gene exp. \ominus | $\Delta\Delta G$ \ominus | miRNA/SNP-target duplexes | LD SNP | Effect |
|----------------|----------------------|--|---------------------|-----------------------------|---|--------|--------|
| hsa-miR-155-3p | 0.86 | CLCN6 NM_001286 rs138217633 chr1:11902225 | 511.17 | Wild: 0.00 SNP: -21.90 | rs138217633: C --> U miRNA:3'acaaUUACGAUUUU--ACAUCCUc5' : : UTR: 5'tggcAGAGCTAACGCCCTGUAGGAg3' | ✓ | gain |
| hsa-miR-155-3p | 0.86 | CLCN6 NM_001256959 rs138217633 chr1:11902225 | 511.17 | Wild: 0.00 SNP: -21.90 | rs138217633: C --> U miRNA:3'acaaUUACGAUUUU--ACAUCCUc5' : : UTR: 5'tggcAGAGCTAACGCCCTGUAGGAg3' | ✓ | gain |
| hsa-miR-155-3p | 0.86 | VPS13D NM_015378 rs149546990 chr1:12569407 | 1867.22 | Wild: 0.00 SNP: -24.50 | rs149546990: A --> G miRNA:3'acaauuACGAUUUAUCAUCCUc5' : : UTR: 5'tgtgttTGTAATATGTAGGAa3' | ✗ | gain |
| hsa-miR-155-3p | 0.86 | VPS13D NM_018156 rs149546990 chr1:12569407 | 1867.22 | Wild: 0.00 SNP: -24.50 | rs149546990: A --> G miRNA:3'acaauuACGAUUUAUCAUCCUc5' : : UTR: 5'tgtgttTGTAATATGTAGGAa3' | ✗ | gain |
| hsa-miR-155-3p | 0.86 | ESPNP NR_026567 rs557690 chr1:17022952 | 4.21 | Wild: 0.00 SNP: -25.10 | rs557690: C --> A miRNA:3'acaauuACGA-UUAUCAUCCUc5' : : UTR: 5'agatgcaGCTGAAGATGAGGAg3' | ✓ | gain |
| hsa-miR-155-3p | 0.86 | ADPRHL2 NM_017825 rs111335009 chr1:36559208 | 892.65 | Wild: -18.30 SNP: -20.40 | rs111335009: C --> U miRNA:3'acAAUUACGAUUUAUCAUCCUc5' : : UTR: 5'gtTTGCTGAGAGCUGTAGGAc3' | ✓ | gain |
| hsa-miR-155-3p | 0.86 | TTC39A NM_001144832 rs185953946 chr1:51753170 | 905.70 | Wild: 0.00 SNP: -18.40 | rs185953946: U --> C miRNA:3'acaAUUACGAUUUAUCAUCCUc5' : : UTR: 5'atcTGA-GCTCTTCTGTAGCAa3' | ✗ | gain |
| hsa-miR-155-3p | 0.86 | TTC39A NM_001080494 rs185953946 chr1:51753170 | 905.70 | Wild: 0.00 SNP: -18.40 | rs185953946: U --> C miRNA:3'acaAUUACGAUUUAUCAUCCUc5' : : UTR: 5'atcTGA-GCTCTTCTGTAGCAa3' | ✗ | gain |

(Εικόνα 2.19: SNPs και γονίδια που σχετίζονται με το hsa-miR-155-3p στον miRNASNP)

| miRNA | miRNA exp. \ominus | Target site on gene 3'UTR | Gene exp. \ominus | $\Delta\Delta G$ \ominus |
|-------------------------------|----------------------|-------------------------------------|---------------------|-----------------------------|
| hsa-miR-155-3p rs200351615 | 0.86 | C1orf21 NM_030806 588-610 | 1369.18 | Wild: -18.30 SNP: -22.30 |
| hsa-miR-155-3p rs200351615 | 0.86 | IL28RA NM_173064 1284-1305 | 206.56 | Wild: -17.10 SNP: -19.70 |
| hsa-miR-155-3p rs200351615 | 0.86 | IL28RA NM_173065 1981-2002 | 206.56 | Wild: -17.10 SNP: -19.70 |
| hsa-miR-155-3p rs200351615 | 0.86 | IL28RA NM_170743 1284-1305 | 206.56 | Wild: -17.10 SNP: -19.70 |
| hsa-miR-155-3p rs200351615 | 0.86 | FBNP1L NM_017737 1972-1991 | 1642.36 | Wild: -20.20 SNP: -22.90 |
| hsa-miR-155-3p rs200351615 | 0.86 | FBNP1L NM_001164473 1972-1991 | 1642.36 | Wild: -20.20 SNP: -22.90 |
| hsa-miR-155-3p rs200351615 | 0.86 | FBNP1L NM_001024948 3263-3282 | 1642.36 | Wild: -20.20 SNP: -22.90 |
| hsa-miR-155-3p rs200351615 | 0.86 | GPR89C NM_001097616 327-348 | 85.50 | Wild: -17.70 SNP: -22.00 |

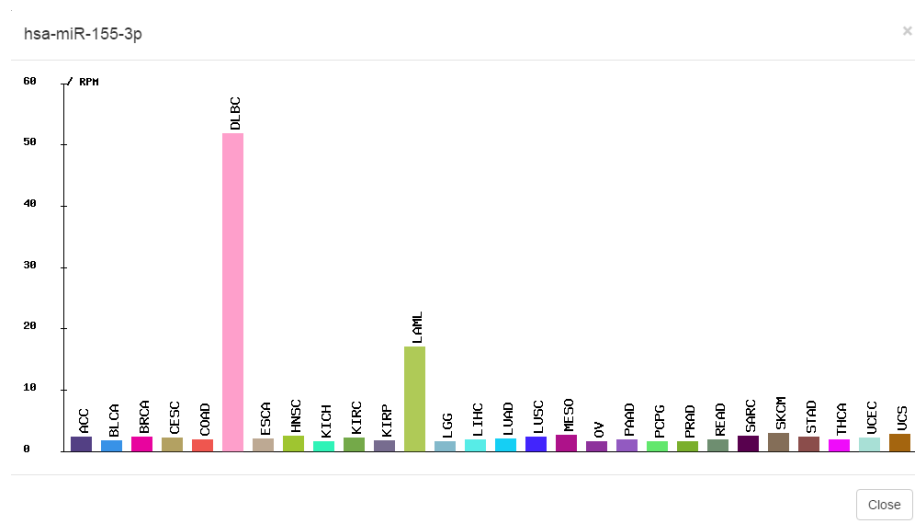
(Εικόνα 2.20: Αποτελέσματα σχετικά με τον πολυμορφισμό του ίδιου του miRNA στον miRNASNP)

Στην συνέχεια παραθέτουμε και ένα γράφημα που δείχνει της αρρώστιες που η εμφάνιση τους σχετίζεται με το συγκεκριμένο miRNA.

DLBC=Lymphoid_Neoplasm_Diffuse_Large_B-Cell_Lymphoma

LAML=Acute_Myeloid_Leukemia

Με αυτές τις δύο ασθένειες σχετίζεται περισσότερο το hsa-miR-155-3p



(Εικόνα 2.21: Συσχέτιση με ασθένειες του hsa-miR-155 στον miRNASNP)

3.Αποτελέσματα

Στο συγκεκριμένο κεφάλαιο θα παρουσιάσουμε τα αποτελέσματα της μελέτης μας, τα χωρίσαμε σε δύο κατηγορίες στα «ποιοτικά» και στα «αριθμητικά». Στην συνέχεια εξηγούμε πως έγινε αυτός ο διαχωρισμός.

3.1 Ποιοτικά Χαρακτηριστικά

Σε αυτόν τον πίνακα , καταγράφουμε τα χαρακτηριστικά εκείνα που αφορούν στις πηγές των Webservers, δηλαδή από πού συγκέντρωσαν τα δεδομένα τους. Αυτά τα δεδομένα αφορούν στην ονοματολογία των miRNAs και των SNPs , στο GenomeBrowser που χρησιμοποιήθηκε για την ονοματολογία αλλά και τις θέσεις των γονιδίων , όπως επίσης και τα εργαλεία πρόβλεψης και η πειραματική επιβεβαίωση μεταξύ miRNAs και στόχων. Στην συνέχεια ακολουθεί αναλυτική περιγραφή κάθε πεδίου ξεχωριστά. Ο πίνακας αυτός θα χρησιμοποιηθεί στην συνέχεια για εξαγωγή συμπερασμάτων σε σχέση με την εγκυρότητα , την ενημερότητα και την αξιοπιστία του κάθε Webserver.

| Name/Info | miRBase Edition | NCBI dbSNP database Edition | GenomeBrowser | Experimentally Support | Prediction Method |
|------------------|------------------------|------------------------------------|-----------------------------------|-------------------------------|---------------------------|
| PolymiRTS | 20 | 126 | Ensembl, UCSC Genome Browser hg18 | TarBase mirTarBase | TargetScan(6.2) |
| dPore | 20 | 130 | UCSC Genome Browser hg18 | TarBase | TargetScan(5.1) PicTar |

| | | | | | |
|-----------------|----|--------------|--------------------------|---|--------------------------------------|
| miRdSNP | 18 | 130 | UCSC Genome Browser hg18 | TarBase mirTarBase miRecords miR2disease | TargetScan(6.2),miRanda |
| miRNASNP | 19 | 137 | UCSC Genome Browser hg19 | TarBase Starbase miRecords mirTarBase miR2disease | TargetScan |
| mrSNP | 20 | 138 | UCSC Genome Browser hg19 | Βιβλιογραφία, RNA-seq | TargetScan |
| miRvar | 13 | 130 | UCSC Genome Browser hg18 | miR2disease miRecords 1000 Genome Project | PHDcleav,RISCbinder |
| dbSMR | 9 | βιβλιογραφία | Ensembl | TarBase | TargetScan ,RNAHybrid ,miRanda |
| Sniper | 21 | βιβλιογραφία | Ensembl 78 | βιβλιογραφία | TargetScan(6.2) |
| MirSNP | 18 | 135 | UCSC Genome Browser hg18 | GWAS, eQTLs | miRanda |

(Πίνακας 3.1: Πληροφορίες για την συλλογή των δεδομένων ανά Webserver)

3.1.1 Έκδοση του miRBase



(Εικόνα 3.1: Interface του miRBase πηγή: <http://www.mirbase.org/>)

Οι πληροφορίες για τα miRNAs όπως ονομασία, το έτος ανακάλυψης, η ερευνητική ομάδα που το ανακάλυψε, η πειραματική τους μη επιβεβαίωση, η οικογένεια, το είδος, το πρόδρομο miRNA, το ώριμο miRNA, το hairpin sequence για κάθενα ξεχωριστά

εισάγονται συνήθως από κάποια έκδοση του miRBase (<http://www.mirbase.org/>) από αρκετά παλαιότερες όπως είναι η 9^η μέχρι τις πιο πρόσφατες όπως είναι η 20^η έκδοση και η τελευταία η οποία είναι η 21^η. Μερικοί από τους WebServers εκτός από το miRBase εισάγουν επιπλέον πληροφορίες από την βιβλιογραφία

Δηλαδή στο αποτέλεσμα κάθε αναζήτησης οι webservers έχουν έναν εξωτερικό σύνδεσμο με επιπλέον πληροφορίες για το miRNA , το οποίο κάναμε την αναζήτηση. Επειδή αναφερόμαστε στις εκδόσεις του miRNA ώστε να έχουμε μια εικόνα για την ενημερότητα και την ανανέωση του Webserver , παραθέτουμε έναν πίνακα με τις εκδόσεις του miRBase , από την πρώτη μέχρι την πιο πρόσφατη. Να σημειώσουμε ότι λόγω της ραγδαίας εξέλιξης της Βιοπληροφορικής η ανακάλυψη νέων miRNAs , γίνεται με ταχύτατους ρυθμούς , συνεπώς η ενημέρωση το servers πρέπει να είναι αρκετά συχνή για υπάρξει αξιοπιστία

| Version | Date | Entries |
|---------|-------|---------|
| 1.0 | 12/02 | 218 |
| 1.1 | 01/03 | 262 |
| 1.2 | 04/03 | 295 |
| 1.3 | 05/03 | 332 |
| 1.4 | 07/03 | 345 |
| 2.0 | 07/03 | 506 |
| 2.1 | 09/03 | 558 |
| 2.2 | 11/03 | 593 |
| 3.0 | 01/04 | 719 |
| 3.1 | 04/04 | 899 |
| 4.0 | 07/04 | 1185 |
| 5.0 | 09/04 | 1345 |
| 5.1 | 12/04 | 1420 |
| 6.0 | 04/05 | 1650 |
| 7.0 | 06/05 | 2909 |
| 7.1 | 10/05 | 3424 |
| 8.0 | 02/06 | 3518 |
| 8.1 | 05/06 | 3963 |
| 8.2 | 07/06 | 4039 |
| 9.0 | 10/06 | 4361 |
| 9.1 | 02/07 | 4449 |
| 9.2 | 05/07 | 4584 |
| 10.0 | 08/07 | 5071 |
| 10.1 | 12/07 | 5395 |
| 11.0 | 04/08 | 6396 |
| 12.0 | 09/08 | 8619 |
| 13.0 | 03/09 | 9539 |
| 14 | 09/09 | 10883 |
| 15 | 04/10 | 14197 |
| 16 | 08/10 | 15172 |
| 17 | 04/11 | 16772 |

| | | |
|----|-------|-------|
| 18 | 11/11 | 18226 |
| 19 | 08/12 | 21264 |
| 20 | 06/13 | 24521 |
| 21 | 06/14 | 28645 |

(Πίνακας 3.2: Εκδόσεις του miRBase με χρονολογική σειρά)

3.1.2 Πηγή δεδομένων SNPs

Υπάρχει μια αντίστοιχη βάση δεδομένων με την miRBase, η οποία έχει συγκεντρωμένες πληροφορίες για όλα τα γνωστά SNPs που υπάρχουν. Ονομάζεται NCBI dbSNP database και όπως και πριν υπάρχουν παλιότερες και νεότερες εκδόσεις. Αντίστοιχα με τα miRNAs επειδή και η ανακάλυψη νέων SNPs γίνεται με ταχύτερους αριθμούς, η ενημέρωση του εκάστοτε server με τις νέες εκδόσεις του dbSNP database, αποτελεί πολύ βασικό κριτήριο αξιοπιστίας.

3.1.3 Genome Browser

Γραφική αναπαράσταση του γονιδιώματος για διαφορετικά είδη, διαφορετικά χρωμοσώματα, διαφορετικά γονίδια, περιοχές μεταγραφής, ασθένειες. Άλλες λειτουργίες τους είναι η πρόβλεψη γονιδίων, πρωτεϊνών. Αυτοί που είναι ευρέως γνωστοί και χρησιμοποιούνται και από τους περισσότερους WebServers είναι ο Ensembl και ο UCSC Genome Browser.

3.1.4 Πειραματική Υποστήριξη

Άλλοι WebServers που έχουν ως στόχο την πειραματική επιβεβαίωση των miRNAs και την συσχέτιση τους με ασθένειες και στόχους. Όπως TarBase, mirTarBase, miRecords, Starbase, miR2disease, 1000 Genome Project. Μερικές από τις παραμέτρους που χρησιμοποιούνται είναι μεταξύ άλλων: το όνομα του γονιδίου στο οποίο βρίσκεται το miRNA, οι πειραματικές μέθοδοι που χρησιμοποιήθηκαν και ποιο ήταν το prediction score που είχε αρχικά το συγκεκριμένο miRNA, καθώς και την πηγή του πειράματος.

3.1.5 Αλγόριθμοι Πρόβλεψης

Ένα από τα σημαντικότερα κομμάτια των WebServers πρέπει να είναι ο αλγόριθμος πρόβλεψης, που χρησιμοποιούν ούτε ώστε να προβλέπουν πιθανούς στόχους των miRNA, αυτό όπως γίνεται εύκολα αντιληπτό, γλιτώνει πάρα πολύ χρόνο από το να ψάχνουμε εντελώς τυχαίες περιοχές του γενετικού υλικού δεδομένου του μεγέθους του. Συνήθως οι αλγόριθμοι αυτοί κάνοντας χρήση κάποιων παραμέτρων που δεν χρειάζεται να επεκταθούμε, υπολογίζουν έναν τελικό αριθμό (score), το οποίο κυμαίνεται συνήθως μεταξύ του 0 και του 1 ή σε άλλες περιπτώσεις παίρνει μια αρνητική τιμή, όσο πιο κοντά στην μονάδα είναι μια ακολουθία ή αντίστοιχα όσο πιο αρνητική τιμή έχει, τόσο πιο

πιθανό στόχο του εκάστοτε miRNA αποτελεί. Φυσικά υπάρχει και η εργαστηριακή - πειραματική επιβεβαίωση που αποτελεί ένα κριτήριο αξιοπιστίας του εκάστοτε αλγορίθμου. Μεταξύ αυτών των αλγορίθμων είναι οι TargetScan , PicTar, miRanda. Συνεπώς οι WebServers μας δεν ψάχνουν στην τύχη φορτώνουν όλες τις περιοχές που έχουν προβλέψει οι αλγόριθμοι που αναφέρθηκαν και στην συνέχεια ελέγχουν εάν αυτές οι περιοχές εμφανίζουν πολυμορφισμό και εάν ο πολυμορφισμός αυτός σχετίζεται με την δημιουργία ασθενειών.

3.1.6 Human Genome Diseases

Της πληροφορίες για τις διάφορες ασθένειες οι WebServers τις αντλούν από διάφορες εκδόσεις ενός διεθνούς Project , το οποίο ονομάζεται HarMap , στην ουσία είναι ένας κατάλογος από κοινές γενετικές παραλλαγές (variants) που συναντάμε συχνά στον άνθρωπο. Μας δίνει πληροφορίες για τις περιοχές του DNA στις οποίες εντοπίζονται αυτές οι παραλλαγές και πως κατανέμονται στους πληθυσμούς της γης.

Κρατήσαμε επίσης μια σειρά αριθμητικών δεδομένων όπως τον αριθμό των miRNAs, τον αριθμό των γονιδίων , τον αριθμός των SNPs κλπ, τα δεδομένα αυτά θα τα παρουσιάσουμε στο τελευταίο κόμμα της της εργασίας αυτής που αφορά στην σύγκριση των WebServers.

3.1.7 Αριθμός Ειδών

Οι WebServers που μελετάμε όπως μπορεί να παρατηρήσει κανείς , έχουν πολλά κοινά χαρακτηριστικά αλλά και αρκετές διαφορές. Μια πολύ σημαντική διαφορά αποτελεί ο διαφορετικός αριθμός των ειδών για τα οποία μπορούν να δώσουν πληροφορίες. Για αυτό τον λόγο τους χωρίσαμε σε δύο κατηγορίες σε αυτούς που μελετάνε μόνο τον άνθρωπο(homo sapiens) και σε εκείνους που διαθέτουν πληροφορίες και για περισσότερους οργανισμούς. Σε πρώτη φάση η ταξινόμηση γίνεται με το εάν μελετάνε μόνο τον άνθρωπο ή και περισσότερα είδη.

| Όνομα WebServer | Μόνο Homosapiens | Περισσότεροι Οργανισμοί |
|---------------------|------------------|-------------------------|
| PolymiRTS | | + |
| dPore | + | |
| miRdSNP | + | |
| miRNASNP | | + |
| mrSNP | | + |
| miRvar | + | |
| dbSMR | + | |
| Micro Sniper | | + |

| | | |
|---------------|---|--|
| MirSNP | + | |
|---------------|---|--|

(Πίνακας 3.3: Δυνατότητα επιλογής μελέτης και άλλων οργανισμών πέραν του ανθρώπου)

Το γεγονός ένας WebServer να μελετάει περισσότερους οργανισμούς από τον homo sapiens δεν τον καθιστά και αυτόματα, αποδοτικότερο ή «καλύτερο» από κάποιον άλλον που μελετάει μόνο τον homosapiens. Σε κάθε περίπτωση όμως βοηθάει περισσότερο ερευνητικές ομάδες που επεκτείνονται στην ανακάλυψη miRNAs και σε άλλα ζώα. Έχει επίσης ενδιαφέρον να μελετήσει κανείς την τελείως διαφορετική επίδραση που μπορεί να έχει το ίδιο miRNA σε διαφορετικούς οργανισμούς.

3.1.8 Αλληλεπιδράσεις με ασθένειες

Όλοι οι Webserver οι οποίοι μελετήσαμε, όπως έχουμε ήδη αναφέρει αρκετές φορές ανακαλύπτουν αλληλεπιδράσεις μεταξύ miRNAs και πολυμορφικών περιοχών. Αυτό έχει σαν απώτερο στόχο την ανακάλυψη σχέσεων μεταξύ miRNAs-SNPs και ασθενειών. Δεν καταγράφουν όμως όλοι οι WebServers συσχετίσεις με ασθένειες, αλλά το αφήνουν στον χρήστη. Για αυτό τους χωρίζουμε ξανά σε δύο κατηγορίες σε αυτούς που καταγράφουν τις συσχετίσεις με ασθένειες και σε αυτούς που δεν το κάνουν. Τις πληροφορίες για τις ασθένειες τις συλλέγουν από την βιβλιογραφία και από τα πειράματα που έχουν γίνει στο παρελθόν, καθώς και από άλλες βάσεις δεδομένων που έχουν σαν αποκλειστική δουλειά να καταγράφουν ότι επιβεβαιώνεται πειραματικά και αναφέρθηκαν νωρίτερα.

| Όνομα WebServer | Μελέτη Ασθενειών |
|---------------------|------------------|
| PolymiRTS | + |
| dPore | + |
| miRdSNP | + |
| miRNASNP | + |
| mrSNP | - |
| miRvar | + |
| dbSMR | + |
| Micro Sniper | + |
| MirSNP | + |

(Πίνακας 3.4: Μελέτη ασθενειών)

3.1.9 Διαφορετικοί τρόποι αναζήτησης

Στον παρακάτω πίνακα συνοψίζονται οι διαφορετικοί τρόποι αναζήτησης που είναι εφικτοί σε κάθε Webserver ξεχωριστά.

| Όνομα WebServer | Αναζήτηση μέσω miRNA | Αναζήτηση μέσω SNPs | Αναζήτηση μέσω Gene | Αναζήτηση μέσω Ασθένειας | Αναζήτηση μέσω Μεθόδου Πρόβλεψης | Σύνθετη Αναζήτηση |
|-----------------|----------------------|---------------------|---------------------|--------------------------|----------------------------------|-------------------|
| MirSNP | + | + | + | - | - | + |
| PolymiRTS | + | + | + | + | - | + |
| miRdSNP | + | + | + | + | + | - |
| miRvar | + | - | - | - | - | - |
| Micro Sniper | + | - | - | - | - | + |
| dPore | + | + | + | + | - | - |
| mrSNP | - | + | - | - | - | - |
| miRNASNP | + | + | + | - | - | - |

(Πίνακας 3.5: Τρόποι αναζήτησης ανά Webserver)

3.1.10 Ανακάλυψη Αλληλεπιδράσεων ανά WebServer

| Όνομα WebServer | miRNA / Γονίδιο | SNP/Γονίδιο | miRNA/Ασθένεια | Ασθένεια / Γονίδιο | miRNA / SNP | SNP/miRNA | Γονίδιο/ miRNA | Γονίδιο/ Ασθένεια | Γονίδιο/ SNP | SNP/ Ασθένεια |
|-----------------|-----------------|-------------|----------------|--------------------|-------------|-----------|----------------|-------------------|--------------|---------------|
| MirSNP | + | + | - | + | + | + | + | - | + | + |
| PolymiRTS | + | + | - | + | + | + | + | - | + | - |
| miRdSNP | + | + | + | + | + | + | + | + | + | + |
| miRvar | + | - | + | - | - | - | - | - | - | - |
| MicroSniper | - | - | - | - | + | - | - | - | - | - |
| dPore | + | - | + | - | + | + | + | - | - | + |
| mrSNP | - | + | - | - | - | + | - | - | - | - |
| miRNASNP | + | + | + | - | + | + | + | + | + | - |

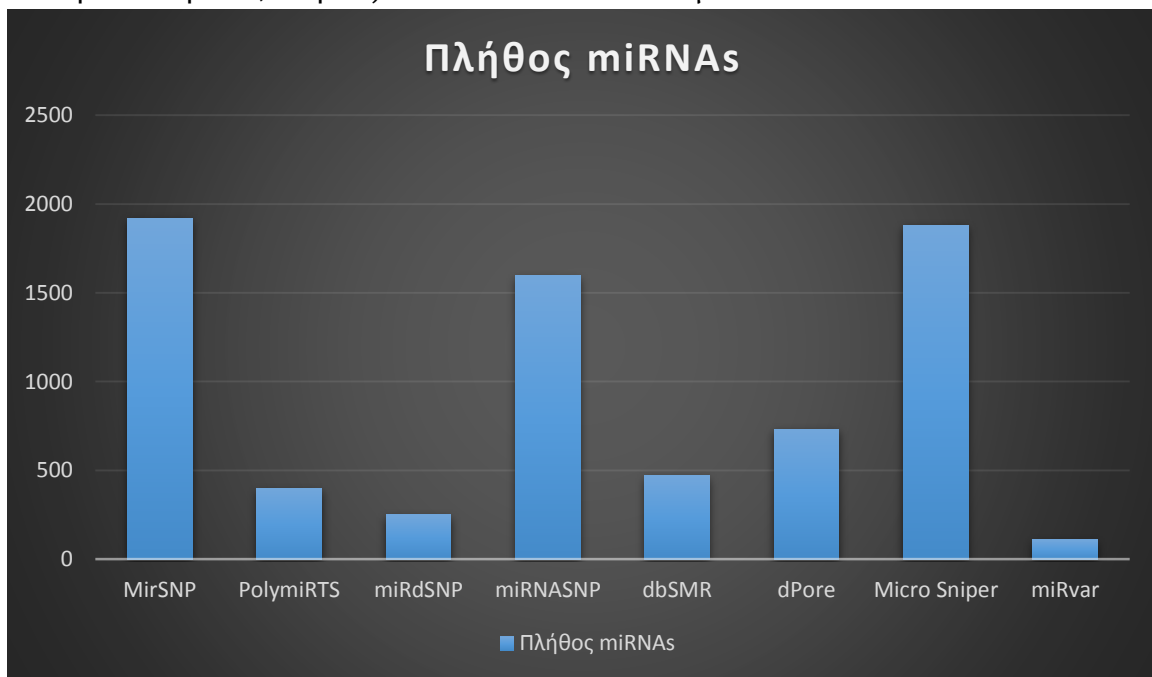
(Πίνακας 3.6:Ανακάλυψη αλληλεπιδράσεων ανά WebServer)

3.2 Αριθμητικά Αποτελέσματα

Σε αυτό το σημείο θα παρουσιάσουμε κάποια στατιστικά στοιχεία των WebServers των οποίων μελετήσαμε. Πέρα από τις πηγές των δεδομένων, έχει αξία να μελετήσουμε και μερικά αριθμητικά δεδομένα των WebServers

3.2.1 Αριθμός miRNAs

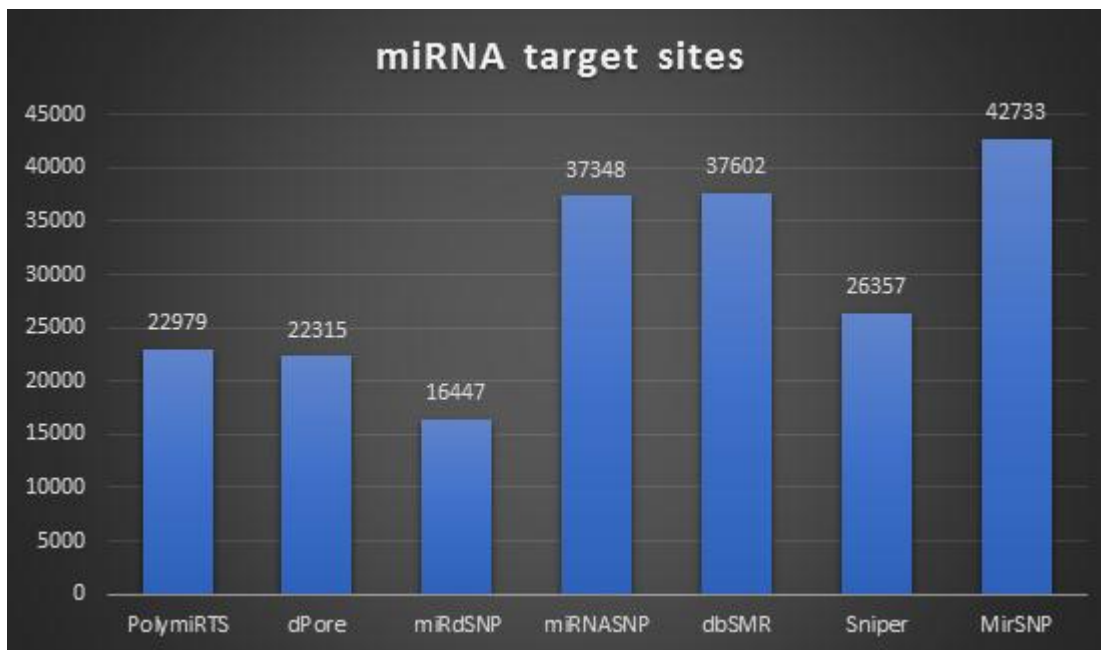
Ένα από τα σημαντικότερα στατιστικά για την μελέτη μας. Από την στιγμή που μελετάμε, αλληλεπιδράσεις μεταξύ miRNAs και SNPs, το πλήθος τους παίζει καθοριστικό ρόλο, στην αξιοπιστία των αποτελεσμάτων του WebServer.



(Γράφημα 3.1: πλήθος των miRNAs ανά WebServer)

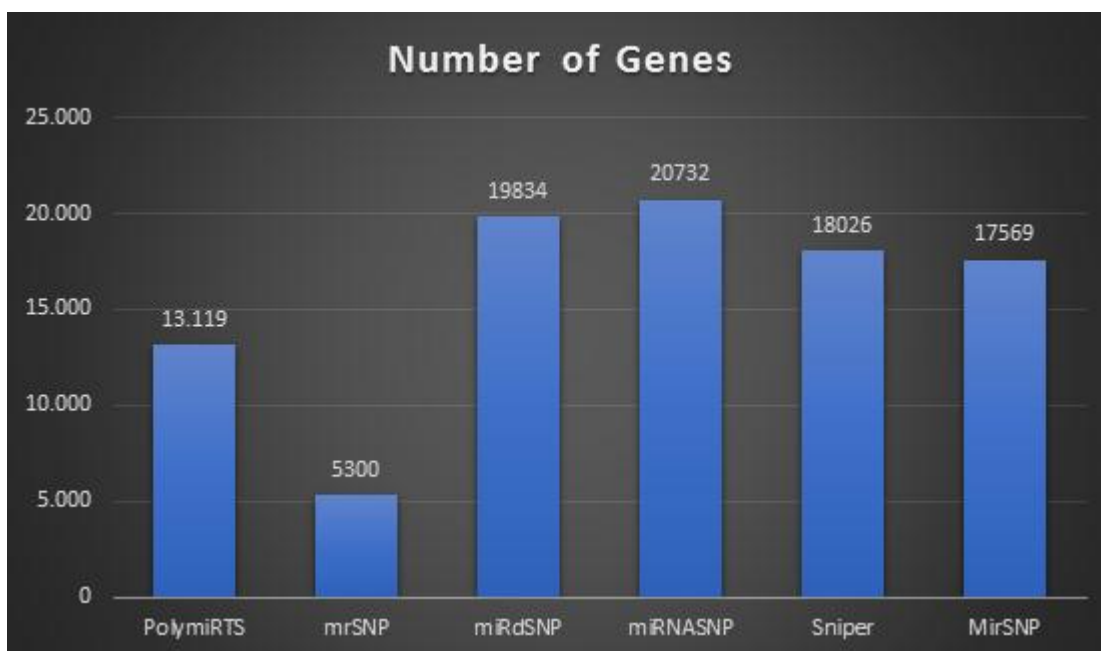
3.2.2 Αριθμός miRNA target sites

Στο παρακάτω γράφημα παραθέτουμε τον συνολικό αριθμό των περιοχών στόχων επί του συνόλου των miRNAs που διαθέτει ο κάθε webserver.



(Γράφημα 3.2: miRNA target sites ανά WebServer)

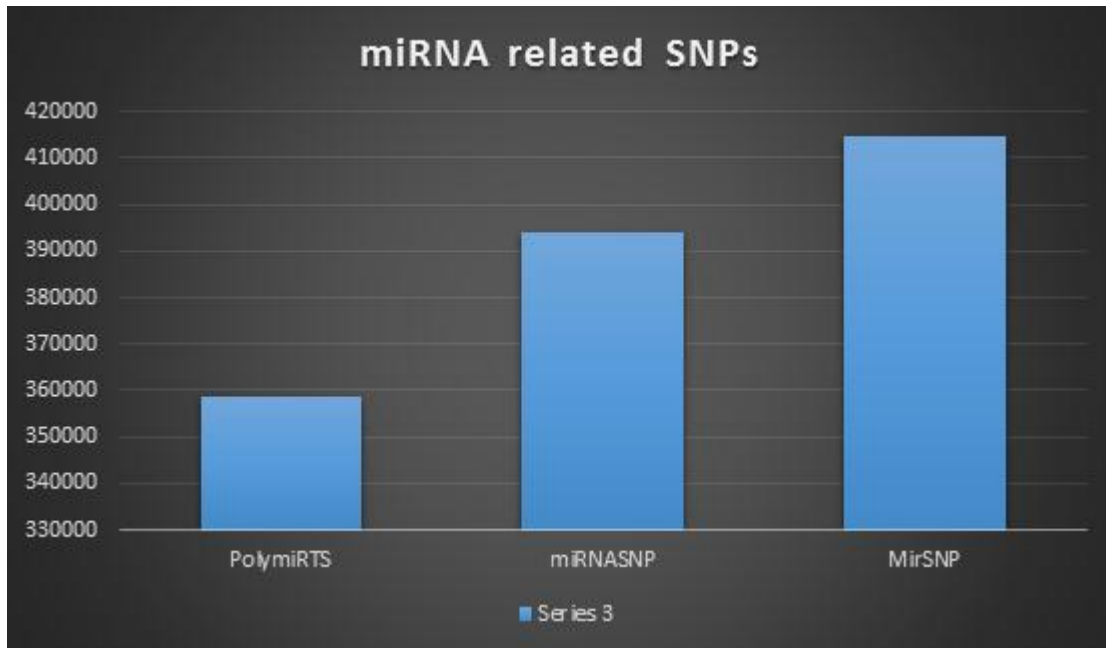
3.2.3 Αριθμός Γονιδίων



(Γράφημα 3.3: αριθμός γονιδίων ανά Webserver)

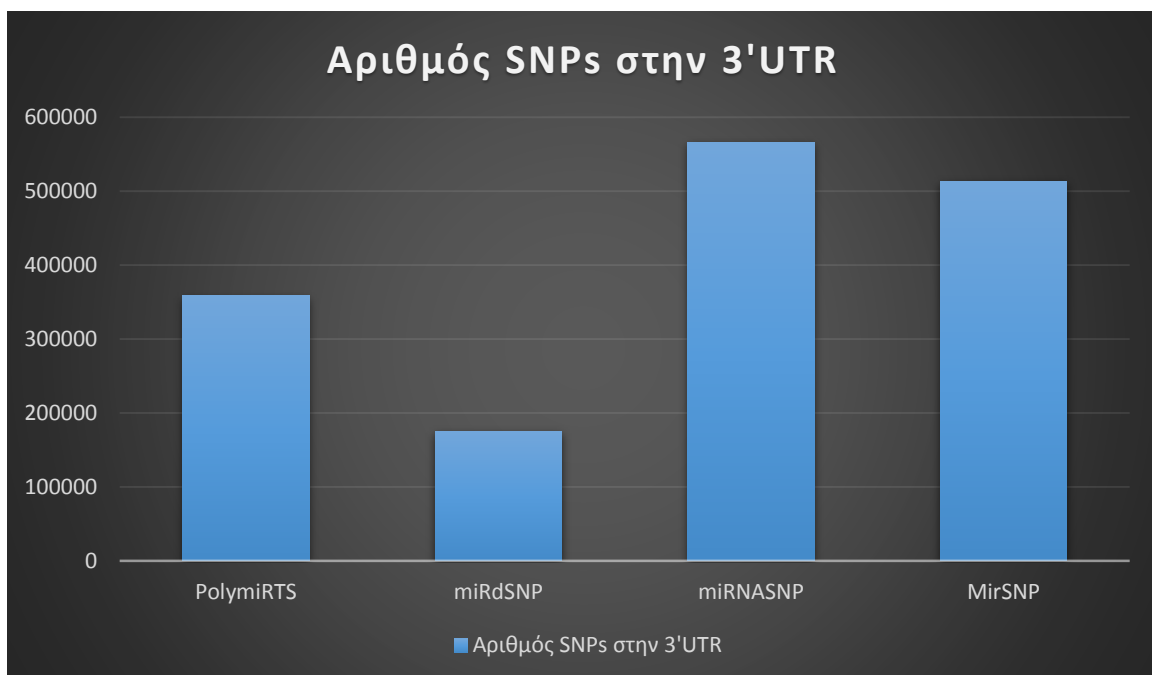
3.2.4 Validated Targets/miRNA - related SNPs

Αριθμός που μας δείχνει πόσοι είναι επιβεβαιωμένοι στόχοι , δηλαδή πόσα miRNAs σχετίζονται άμεσα από τα SNPs στον εκάστοτε WebServer.



(Γράφημα 3.4: αριθμός SNPs που σχετίζονται με miRNAs ανά WebServer)

3.2.5 Αριθμός SNPs στην 3' αμετάφραστη περιοχή



(Γράφημα 3.5:Αριθμός πολυμορφισμών στην 3'UTR)

3.2.6 Ενδεικτική αναζήτηση για το hsa-miR-55 και στατιστική σύγκριση

| Όνομα WebServer | hsa-mir-155-3p | hsa-mir-155-5p | hsa-mir-155(unique) |
|------------------|----------------|----------------|---------------------|
| PolymiRTS | 375 | 296 | 673 |
| MirSNP | 492 | 469 | 933 |
| dPore | - | - | 96 |
| miRdSNP | - | - | 399 |
| miRNASNP | 569 | 475 | 1044 |

(Πίνακας 3.7: Αριθμός γονιδίων στόχων ανά WebServer)

Στο παραπάνω πίνακα έχουμε καταγράψει τα αποτελέσματα για κάθε webserver για την αναζήτηση του hsa-mir-155 και αλληλεπιδράσεις μεταξύ γονιδίων.

| Κοινές Προβλέψεις | Αριθμός | Ποσοστό 1 ^{ου} | Ποσοστό 2 ^{ου} |
|--------------------------|---------|-------------------------|-------------------------|
| miRdSNP/dPore | 16 | 4% | 16,6% |
| miRdSNP/PolymiRTS | 57 | 14,2% | 8,4% |
| miRdSNP/MirSNP | 67 | 16,7% | 7,1% |
| dPore/MirSNP | 8 | 8,3% | 0,85% |
| dPore/PolymiRTS | 9 | 9,3% | 1,3% |
| MirSNP/PolymiRTS | 409 | 43,8% | 60,7% |

(Πίνακας 3.8: Ποσοστό ταύτισης προβλέψεων ανά ζεύγος WebServers)

4.Συμπεράσματα

1.Μπορούμε να συμπεράνουμε ότι η μελέτη και εξέλιξη τέτοιων Webservers με την παράλληλη εξέλιξη των αλγορίθμων πρόβλεψης, αποτελεί ένα από τα σημαντικότερα πεδία έρευνας στον τομέα της Βιοπληροφορικής. Η δημιουργία νέων αλλά και η βελτίωση τέτοιων εργαλείων , θα βοηθήσει τα επόμενα χρόνια στην περαιτέρω κατανόηση ασθενειών και στην πιθανή θεραπεία τους. Η κατευθυνόμενη έρευνα σε συγκεκριμένα σημεία του γονιδιώματος γλιτώνει πάρα πολύ χρόνο στο εργαστήριο του Βιολόγου.

2.Οι Webservers που μελετήσαμε ενώ έχουν πολλά κοινά χαρακτηριστικά ο καθένας έχει και την δική του ιδιαιτερότητα και ανάλογα με την έρευνα που θέλουμε να κάνουμε , οφείλουμε να επιλέξουμε κάποιον ή κάποιους από αυτούς. Σίγουρα ένα στοιχείο πολύ βασικό είναι το πόσο ενημερωμένος είναι ο server σε σχέση με τα νέα πράγματα που ανακαλύπτονται σχεδόν σε καθημερινή βάση. Συνεπώς ένας server ο οποίος χρησιμοποιεί παλαιότερες εκδόσεις , τόσο για την εισαγωγή πληροφοριών(miRBase, dbSNP) αλλά και για την πρόβλεψη νέων στόχων , δεν είναι πολύ αξιόπιστος. Με την έννοια ότι όχι μόνο δεν θα είναι επαρκής αλλά και πολλά πράγματα μπορεί να έχουν αλλάξει συνεπώς τα αποτελέσματά του να είναι λανθασμένα. Οι servers λοιπόν που χρησιμοποιούν τις πιο πρόσφατες εκδόσεις όπως είναι ο miRNA Sniper ο οποίος χρησιμοποιεί την πιο πρόσφατη έκδοση(21^η) είναι σίγουρα καλύτερα ενημερωμένος , από τον dbSMR που χρησιμοποιεί την 9^η έκδοση. Αντίστοιχα και για τις υπόλοιπες πληροφορίες.

3.Οι αλγόριθμοι πρόβλεψης σίγουρα αποτελούν ένα από τα σημαντικότερα χαρακτηριστικά κάθε WebServer γιατί αποτελούν την αφετηρία. Είναι αυτοί που εντοπίζουν τα binding sites των miRNAs αλλά και αυτοί που καταγράφουν όσα από αυτά επιβεβαιώνονται πειραματικά. Η σύγκριση τέτοιων αλγορίθμων δεν αποτελεί αντικείμενο της συγκεκριμένης διπλωματικής εργασίας, θα μπορούσε να αποτελεί ένα εντελώς ξεχωριστό θέμα. Αυτό που μπορούμε να πούμε είναι ότι οι Webservers που χρησιμοποιούν συνδυασμό αποτελεσμάτων , τόσο για την πρόβλεψη αλλά και για την επιβεβαίωση μειώνουν την πιθανότητα λάθους.

Ένα δεύτερο στοιχείο είναι τα αριθμητικά δεδομένα που χρησιμοποιεί ο κάθε webserver. Προφανώς όση περισσότερη πληροφορία διαθέτει ένας WebServer και όσα περισσότερα miRNAs , SNPs , ασθένειες κλπ. έχει καταγεγραμμένα αυτομάτως , αποτελεί καλύτερη επιλογή δεδομένου , ότι και ο πιο απαιτητικός χρήστης θα μπορέσει να πάρει τις πληροφορίες και για το πιο σπάνιο miRNA , η μια ασθένεια που δεν εμφανίζεται συχνά. Για παράδειγμα όπως είδαμε και στα γραφήματα οι miRNASNP και miRdSNP διαθέτουν πληροφορίες για τον διπλάσιο αριθμό γονιδίων απ' ότι οι miRdSNP και PolyMiRTS , συνεπώς αν κάποιος επικεντρώνεται στην αναζήτηση αλληλεπιδράσεων για τα γονίδια

θα απευθυνθεί στους δύο πρώτους. Αντίστοιχο είναι το σκεπτικό και για τα υπόλοιπα αριθμητικά γνωρίσματα. Συνεπώς ανάλογα με το που θέλει να κατευθύνει την έρευνα του ο χρήστης , αντίστοιχα θα επιλέξει και τον κατάλληλο Webserver

Βασικό επίσης χαρακτηριστικό είναι το Interface του κάθε WebServer , ένα εύχρηστο περιβάλλον και ως προς την αναζήτηση και ως προς τα αποτελέσματα , σίγουρα διευκολύνει τον χρήστη και την έρευνά του. Επίσης ο τρόπος που γίνεται η αναζήτηση , δηλαδή αν διατίθενται πολλές επιλογές όπως ανά miRNA/SNP/γονίδιο/ασθένεια, αν μπορεί να γίνει μεμονωμένη ή σύνθετη αναζήτηση όπως είδαμε στο κεφάλαιο 2 κλπ. Αντίστοιχα το αποτέλεσμα της αναζήτησης , αν είναι εύχρηστο , αν ο χρήστης μπορεί να το κατεβάσει σε αρχείο στον υπολογιστή του , αν βγάζει συγκεντρωτικά ή ταυτόχρονα αποτελέσματα κλπ.

Δεδομένου λοιπόν ότι το κοινό στο οποίο απευθύνονται οι συγκεκριμένοι WebServers ,έχει γνώση του αντικειμένου και κρίση στο τι θα επιλέξει για την έρευνα του , ανάλογα με τα ιδιαίτερα χαρακτηριστικά που επιθυμεί οφείλει να επιλέξει και κατάλληλα.

5.Ευχαριστίες

Θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια κυρία Άρτεμις Χατζηγεωργίου και τον επιβλέποντα καθηγητή κύριο Γεράσιμο Ποταμιάνο, καθώς επίσης και την διδακτορική φοιτήτρια Δήμητρα Καραγκούνη χωρίς την βοήθεια της οποίας δεν θα είχε ολοκληρωθεί αυτή η διπλωματική εργασία. Ακόμη θα ήθελα να ευχαριστήσω το Πανεπιστήμιο Θεσσαλίας και τον πρόεδρο του τμήματος κύριο Ελευθέριο Τσουκαλά, όπως επίσης και τους εργαζόμενους του τμήματος στην Γραμματεία, την Φοιτητική μέριμνα , την καθαριότητα και την φύλαξη. Τέλος του φίλους μου και την οικογένειά μου για την στήριξη που μου παρείχαν όλα τα χρόνια των σπουδών μου.

6. Βιβλιογραφία

- 1 Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 2014; 42(D1):D86-D91.
- 2 Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.* 2012; 40(D1):D216-221. PMID: 22080514.
- 3 Saunders, M.A., Liang, H. and Li, W.-H. (2007) Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl Acad. Sci. USA*, 104, 3300–3305.
- 4 Sethupathy, P. and Collins, F.S. (2008) MicroRNA target site polymorphisms and human disease. *Trends Genet.*, 24, 489–497.
- 5 Ryan, B.M., Robles, A.I. and Harris, C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer*, 10, 389–402.
- 6 Mendell, J.T. and Olson, E.N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, 148, 1172–1187.
- 7 Li, J. and Zhang, Z. (2013) miRNA regulatory variation in human evolution. *Trends Genet.*, 29, 116–124.
- 8 Esquela-Kerscher, A. and Slack, F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, 6, 259–269.
- 9 Nicoloso, M.S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P., Shimizu, M., Wojcik, S.E., Ferdin, J., Kunej, T., Xiao, L. et al. (2010) Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.*, 70, 2789–2798.
- 10 Zhao, X., Ye, Q., Xu, K., Cheng, J., Gao, Y., Li, Q., Du, J., Shi, H. and Zhou, L. (2013) Single-nucleotide polymorphisms inside microRNA target sites influence the susceptibility to type 2 diabetes. *J. Hum. Genet.*, 58, 135–141.
- 11 Delay, C., Mandemakers, W. and Hebert, S.S. (2012) MicroRNAs in Alzheimer's disease. *Neurobiol. Dis.*, 46, 285–290.
- 12 Gao, Y., He, Y., Ding, J., Wu, K., Hu, B., Liu, Y., Wu, Y., Guo, B., Shen, Y., Landi, D. et al. (2009) An insertion/deletion polymorphism at miRNA-122-binding site in the interleukin-

- 1alpha 30 untranslated region confers risk for hepatocellular carcinoma. *Carcinogenesis*, 30, 2064–2069.
- 13 Bao,L., Zhou,M., Wu,L., Lu,L., Goldowitz,D., Williams,R.W. and Cui,Y. (2007) PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res.*, 35, D51–D54.
 - 14 Ziebarth,J.D., Bhattacharya,A., Chen,A. and Cui,Y. (2012) PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.*, 40, D216–D221.
 - 15 Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460, 479–486.
 - 16 Hafner,M., Landthaler,M., Burger,L., Khorshid,M., Hausser,J., Berninger,P., Rothballer,A., Ascano,M. Jr, Jungkamp,A.C., Munschauer,M. et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141, 129–141.
 - 17 Kudla,G., Granneman,S., Hahn,D., Beggs,J.D. and Tollervey,D. (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl Acad. Sci. USA*, 108, 10010–10015.
 - 18 Helwak,A., Kudla,G., Dudnakova,T. and Tollervey,D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153, 654–665.
 - 19 Mullaney,J.M., Mills,R.E., Pittard,W.S. and Devine,S.E. (2010) Small insertions and deletions (INDELS) in human genomes. *Hum. Mol. Genet.*, 19, R131–R136.
 - 20 Dawson,E., Chen,Y., Hunt,S., Smink,L.J., Hunt,A., Rice,K., Livingston,S., Bumpstead,S., Bruskiwich,R., Sham,P. et al. (2001) A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. *Genome Res.*, 11, 170–178
 - 21 Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, 40, D109–D114.
 - 22 Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, 32, D493–D496.

- 23 Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, 37, D105–D110.
- 24 Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, 39, D163–D169.
- 25 Sherry,S.T., Ward,M.-H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, 29, 308–311
- 26 Georges M, Coppieters W, Charlier C (2007) Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Current Opinion in Genetics & Development* 17: 166–176.
- 27 Iwai N, Naraba H (2005) Polymorphisms in human pre-miRNAs. *Biochemical and Biophysical Research Communications* 331: 1439–1444.
- 28 Hiard S, Charlier C, Coppieters W, Georges M, Baurain D (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Research* 38: D640–D651.
- 29 Bhartiya D, Laddha SV, Mukhopadhyay A, Scaria V (2011) miRvar: A comprehensive database for genomic variations in microRNAs. *Hum Mutat* 32: E2226–2245.
- 30 Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 39: D152–D157.
- 31 McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26: 2069–2070.
- 32 Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Research* 37: D155–D158.
- 33 Twigger SN, Shimoyama M, Bromberg S, Kwitek AE, Jacob HJ, et al. (2007) The Rat Genome Database, update 2007—Easing the path from disease to data and back again. *Nucleic Acids Research* 35: D658–D662.

- 34 Jiang Q, Wang Y, Hao Y, Juan L, Teng M, et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research* 37: D98–D104.
- 35 Min H, Yoon S: Got target? Computational methods for microRNA target prediction and their extension. *Exp Mol Med* 2010, 42:233–244. 10.3858/emm.2010.42.4.032
- 36 Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R: Fast and effective prediction of microRNA/target duplexes. *RNA* 2004,10:1507–1517. 10.1261/rna.5248604
- 37 Rusinov V, Baev V, Minkov IN, Tabler M: MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res* 2005, 33:696–700. 10.1093/nar/gki364
- 38 Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: Combinatorial microRNA target predictions. *Nat Genet* 2005,37(5):495–500. 10.1038/ng1536
- 39 Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG: DIANA-microT web server: elucidating microRNA, functions through target prediction. *Nucleic Acids Res* 2009, 37:273–276. 10.1093/nar/gkp292
- 40 Maragkakis M, Alexiou P, Papadopoulos GL, Reczko M, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Simossis VA, Sethupathy P, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG: Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics* 2009.,295(1):
- 41 Sethupathy P, Collins FS: MicroRNA target site polymorphisms and human disease. *Trends Genet* 2008, 24:489–497. 10.1016/j.tig.2008.07.004
- 42 Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W: Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. *Carcinogenesis* 2008,29(7):1306–1311. 10.1093/carcin/bgn116
- 43 Barenboim M, Zoltick BJ, Guo Y, Weinberger DR: MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Hum Mutat* 2010,31(11):1223–1232. 10.1002/humu.21349
- 44 Bruno AE, Li L, Kalabus JL, Pan Y, Yu A, Hu Z: mirdsnp: a database of disease-associated snps and microRNA target sites on 3' utrs of human genes. *BMC Genomics* 2012,13(1):44. 10.1186/1471-2164-13-44

- 45 Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, Zhang D: Mirsnap, a database of polymorphisms altering mirna target sites, identifies mirna-related snps in gwas snps and eqtls. *BMC Genomics* 2012,13(1):661. 10.1186/1471-2164-13-661
- 46 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002, 12(6):996-1006.
- 47 Consortium IH, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JTF, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007, 449(7164):851-861.

- 48 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008, 24(24):2938-2939.
- 49 Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG: Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* 2009, 25(23):3049-3055.
- 50 Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006, 34(Database):D140-D144.
- 51 Sethupathy P, Corda B, Hatzigeorgiou AG: TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 2006, 12(2):192-197
- 52 Zorc M., Skok D.J., Godnic I., et al. (2012) Catalog of microRNA seed polymorphisms in vertebrates. *PLoS One*, 7, e30737
- 53 Deveci M., Catalyurek U.V., Toland A.E. (2014) mrSNP: software to detect SNP effects on microRNA binding. *BMC Bioinform.*
- 54 Vergoulis T., Vlachos I.S., Alexiou P., et al. (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.*, 40, D222–D229.
- 55 Dweep H., Sticht C., Pandey P., et al. (2011) miRWalk–database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.*, 44, 839–847
- 56 Hsu S.D., Tseng Y.T., Shrestha S., et al. . (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.*, 42, D78–D85
- 57 Barrett J.C., Fry B., Maller J., et al. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263–265
- 58 Jing Gong, Chunjie Liu,^{1,†} Wei Liu,¹ Yuliang Wu,¹ Zhaowu Ma,¹ Hu Chen,¹ and An-Yuan Guo (2015) An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools
- 59 Ahmed F, Ansari HR, Raghava GP. 2009. Prediction of guide strand of microRNAs from its sequence and secondary structure. *BMC Bioinformatics* 10:105

- 60 Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S, Birney E. 2008. The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res* 36:D445-D448.
- 61 Griffiths-Jones S, Saini HK, van DS, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36:D154-D158
- 62 Lai EC: Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* 2002,30(4):363–364.
- 63 Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin DY, Duan J, Ophoff RA, Andreassen OA, et al.: Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011,43(10):969–976.
- 64 Betel D, Wilson M, Gabow A, Marks DS, Sander C: The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008,36(Database issue):D149-D153.
- 65 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al.: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007,81(3):559–575.
- 66 Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology* 13: 1097–1101
- 67 Canella D, Praz V, Reina JH, Cousin P, Hernandez N (2010) Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Research* 20: 710–721
- 68 Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a RanGTPdependent dsRNA-binding protein that mediates nuclear export of premiRNAs. *RNA* 10: 185–191.
- 69 Lee Y, Ahn C, Han J, Choi H, Kim J, et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425: 415–419.
- 70 Lau N, Lim L, Weinstein E, Bartel D (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858–862
- 71 Hariharan M, Scaria V, Brahmachari SK (2009) dbSMR: A novel resource of genome-wide SNPs affecting microRNA mediated regulation. *BMC Bioinformatics* 10: 108.

- 72 Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, et al. (2006) PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Research* 34: W621–625.
- 73 Alexiou P, Vergoulis T, Gleditsch M, Prekas G, Dalamagas T, et al. (2010) miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Research* 38: D137–D141.
- 74 Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, et al. (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biology* 11: R6.
- 75 Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, et al. (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Research* 38: D613–619.
- 76 Schmeier S, Schaefer U, MacPherson CR, Bajic VB (2011) dPORE-miRNA: Polymorphic Regulation of MicroRNA Genes. *PLoS ONE* 6(2): e16657. doi:10.1371/journal.pone.0016657
- 77 Lai EC: Predicting and validating microRNA targets. *Genome Biol* 2004, 5:115
- 78 Thadani R, Tammi MT: MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics* 2006,7(Suppl 5):S20.
- 79 Flicek P, et al.: Ensembl 2008. *Nucleic Acids Res* 2008, 36:D707–14.