

UNIVERSITY OF THESSALY



Diploma Thesis

Web application for the visualization of microRNA-target interactions

*Διαδικτυακή εφαρμογή γραφικής
απεικόνισης για τις αλληλεπιδράσεις των
microRNA-στόχων*

Author:

Georgios TSIKRTEAS

Supervisor:

Artemis HATZIGEORGIOU

Gerasimos POTAMIANOS

*A thesis submitted in fulfillment of the requirements
for the degree of Diploma Thesis*

in the

Department of Electrical and Computer Engineering

October 12, 2018

Declaration of Authorship

I, Georgios TSIRTEAS, declare that this thesis titled, “Web application for the visualization of microRNA-target interactions” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“Our world is built on biology and once we begin to understand it, it then becomes a technology.”

Ryan Bethencourt

UNIVERSITY OF THESSALY

Abstract

Department of Electrical and Computer Engineering

Diploma Thesis

Web application for the visualization of microRNA-target interactions

by Georgios TSIKRTEAS

Being in 2018 we can safely say that we are in the age of information and data. The fast human progress in science and technology brings us, in a daily basis, face to face with a huge amount of information. This amount of information is also observable in the field of biology, which by nature encapsulates enormous sizes and complexity, which we do not have yet entirely clarified. Along with the help of bioinformatics, analyzing the data we come across from research laboratories, we came out that non-coding transcripts, those that do not code for any protein, are not junk data as we once thought, instead they are very important, as they are involved in a variety of biological processes. More specifically, microRNAs are small, single stranded, non-coding RNA molecules approximately 22 nucleotides long and they are considered central post-transcriptional gene regulators. They are involved in both physiological processes and in many pathological conditions. As a result, microRNAs are studied for their potential as biomarkers and therapeutic targets. The purpose of this diploma thesis is to create a Genome Browser, which will be specially implemented to represent graphically the microRNAs in relation to their predicted targets, based on the DIANA-microT-CDS prediction results. The present application at first, receives data from various sources and biological databases, and then processes them, combines them and finally presents a graphic depiction, from which the user can draw some useful conclusions. Briefly, in the final depiction of the application, an mRNA transcript is presented along with all the predicted microRNAs that target its MREs. Finally, various relevant information and statistics can be observed, interacting with the user interface of the application.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Περίληψη

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Διπλωματική Εργασία

**Διαδικτυακή Εφαρμογή Γραφικής Απεικόνισης για τις Αλληλεπιδράσεις των
microRNA-στόχων**

Γεώργιος Τσικριτέας

Εν έτει 2018 μπορούμε με σιγουριά να δηλώσουμε ότι βρισκόμαστε στην εποχή της πληροφορίας και των δεδομένων. Η ταχύτερη ανθρώπινη πρόοδος στις επιστήμες και την τεχνολογία μας φέρνει καθημερινά αντιμέτωπους με έναν τεράστιο όγκο πληροφοριών. Ο όγκος αυτός των πληροφοριών είναι ορατός και στο πεδίο της βιολογίας, η οποία εμπεριέχει τεράστια μεγέθη και πολυπλοκότητα, την οποία δεν έχουμε ακόμα αποσαφηνίσει στο σύνολό της. Με τη βοήθεια της βιοπληροφορικής λοιπόν και αναλύοντας τα δεδομένα τα οποία προκύπτουν από τα ερευνητικά εργαστήρια αποφανθήκαμε ότι τα μη κωδικά μετάγραφα (non-coding transcripts), δηλαδή αυτά τα οποία δεν κωδικοποιούν κάποια πρωτεΐνη, δε θεωρούνται άχρηστη πληροφορία, όπως πιστεύαμε στο παρελθόν, αντιθέτως είναι πολύ σημαντικά, καθώς συμμετέχουν σε μία πληθώρα βιολογικών διεργασιών. Πιο συγκεκριμένα, τα microRNAs είναι μη κωδικά, μονόκλιωνα μετάγραφα, τα οποία έχουν μήκος περίπου 22 νουκλεοτιδίων και θεωρούνται κεντρικοί μετά – μεταγραφικοί παράγοντες για τη ρύθμιση ορισμένων γονιδίων. Εμπλέκονται στις φυσιολογικές βιολογικές διεργασίες των οργανισμών, καθώς επίσης και σε πολλές παθολογικές καταστάσεις. Τα microRNAs κατά συνέπεια μελετούνται εκτενώς για τη δυναμική τους ως βιοδείκτες και θεραπευτικοί στόχοι. Σκοπός της παρούσας διπλωματικής εργασίας είναι η δημιουργία ενός Genome Browser, ο οποίος θα είναι ειδικά υλοποιημένος, έτσι ώστε να αναπαριστά τα microRNAs σε σχέση με τους προβλεπόμενους στόχους τους, βασιζόμενος στις προβλέψεις του προγράμματος DIANA-microT-CDS. Η παρούσα εφαρμογή λαμβάνει δεδομένα από διάφορες πηγές ή βιολογικές βάσεις δεδομένων, τα επεξεργάζεται, τα συνδυάζει, και τα αναπαριστά με όμορφο γραφικό τρόπο στον χρήστη, ερευνητή, με σκοπό ο ίδιος να καταλήξει σε χρήσιμα συμπεράσματα. Συνοπτικά, στην τελική απεικόνιση της εφαρμογής, αναπαριστώνται ένα μετάγραφο mRNA, μαζί με όλα τα πιθανά microRNAs, τα οποία στοχεύουν στα MREs αυτού του μεταγράφου. Τέλος, απεικονίζονται σχετικές πληροφορίες και στατιστικά, τα οποία ο χρήστης μπορεί να παρατηρήσει, αλληλεπιδρώντας με το γραφικό περιβάλλον της εφαρμογής και να αποφανθεί για τα αποτελέσματα.

Acknowledgements

Αρχικά, θα ήθελα να ευχαριστήσω θερμά την καθηγήτριά μου, κυρία Άρτεμις Χατζηγεωργίου, η οποία επέβλεψε τη διπλωματική μου εργασία και συνέβαλλε στο πέρας της. Θέλω να την ευχαριστήσω που με δέχτηκε με καλοσύνη στην ομάδα του ερευνητικού εργαστηρίου, στο οποίο είναι επικεφαλής, καθώς επίσης και για τις ευκαιρίες που μου έδωσε να συναναστραφώ με εξάισια άτομα και επιστήμονες. Στη συνέχεια, θα ήθελα να ευχαριστήσω την Δήμητρα Καραγκούνη, μέλος του ερευνητικού εργαστηρίου DIANA TOOLS, για τη συνεχή υποστήριξη που μου παρείχε και για τον πολύτιμο χρόνο που μου αφιέρωσε. Η συμβολή της ήταν καθοριστική για την ανάπτυξη και την αποπεράτωση της διπλωματικής μου εργασίας, καθώς με καθοδηγούσε και με συμβούλευε σε κάθε μου βήμα. Θέλω πραγματικά να της ευχηθώ ολόψυχα, τα καλύτερα για το μέλλον. Θα ήθελα ακόμα να ευχαριστήσω όλους τους φίλους μου, οι οποίοι με υποστήριζαν όλα αυτά τα χρόνια και οι οποίοι συνέβαλλαν και αυτοί στην ολοκλήρωση μίας μεγάλης προσπάθειας. Τέλος, θα ήθελα να πω ένα μεγάλο ευχαριστώ μέσα από την καρδιά μου, στην οικογένειά μου, τον πατέρα μου Βασίλη, τη μητέρα μου Χρυσούλα, την αδερφή μου Όλγα και τον μικρό μου αδερφό Στέλιο, για την συνεχή και αναντικατάστατη υποστήριξη, την οποία μου παρείχαν, καθώς και για την αλληπάλληλη στήριξη τους, σε κάθε μου επιλογή. Τους ευχαριστώ που μου έδωσαν την ευκαιρία να σπουδάσω και να μορφωθώ, να ταξιδέψω και να γνωρίσω νέα άτομα, να ζήσω ανεξάρτητος και να διαμορφώσω χαρακτήρα και προσωπικότητά.

Contents

Declaration of Authorship	ii
Abstract	iv
Περίληψη	v
Acknowledgements	vi
1 Εισαγωγή	1
1.1 Γιατί η βιοπληροφορική είναι σημαντική	1
1.2 Θεωρητικό βιολογικό υπόβαθρο	1
1.2.1 Από το DNA στην πρωτεΐνη	1
1.2.2 Δομή mRNA	3
1.3 microRNAs	4
1.3.1 Τί είναι τα microRNAs	4
1.3.2 Ονοματολογία των microRNAs	5
1.3.3 Η Βιογένεση των microRNAs	5
1.3.4 Ο Μηχανισμός Δράσης των microRNAs	8
1.3.5 Η Λειτουργία και ο Ρόλος των microRNAs	10
1.3.6 Seed Περιοχές (Seed Regions)	12
2 Αναγνώριση αλληλεπιδράσεων μεταξύ microRNA και mRNA	15
2.1 Μέθοδοι αναγνώρισης	15
2.1.1 Πειραματικές μέθοδοι για την αναγνώριση αλληλεπιδράσεων μεταξύ miRNA και mRNA (Experimental)	15
2.1.2 Υπολογιστικές μέθοδοι για την αναγνώριση των miRNA στόχων (Computational)	17
2.2 Αλγόριθμοι Πρόβλεψης στόχων microRNA	19
2.2.1 DIANA microT-CDS	19
2.2.2 Targetscan	22
3 Δεδομένα	23
3.1 Πηγές Δεδομένων	23
3.1.1 DIANA-microT-CDS	23
3.1.2 Targetscan	23
3.1.3 miRBase	24

3.1.4	NCBI	25
3.1.5	Ensembl	26
3.2	Προεπεξεργασία Δεδομένων	30
3.3	Η Βάση Δεδομένων της Εφαρμογής	38
4	Διαδικτυακή Εφαρμογή - Αποτελέσματα	40
4.1	Σκοπός της εφαρμογής	40
4.2	Παρουσίαση της Web εφαρμογής	40
4.2.1	Μενού Πλοήγησης	42
4.2.2	Φίλτρα Αναζήτησης	45
4.2.3	Ροή εφαρμογής	48
4.2.4	Ανάλυση τελικών αποτελεσμάτων	53
5	Διαδικτυακές Εφαρμογές - Django Framework	62
5.1	Διαδικτυακές Εφαρμογές	62
5.1.1	Πώς λειτουργούν οι διαδικτυακές εφαρμογές	63
5.1.2	Τα πλεονεκτήματα των διαδικτυακών εφαρμογών	64
5.2	Web Frameworks	64
5.2.1	Django Web Framework	65
6	Επίλογος	72
	Bibliography	73

List of Figures

1.1	Central Dogma of Biology	2
1.2	mRNA Structure	3
1.3	microRNA molecule	5
1.4	microRNA Biogenesis	7
1.5	microRNA Mechanisms	9
1.6	Rate of microRNA Mechanisms	9
1.7	microRNAs are involved in ...	11
1.8	microRNA-clinical candidates	11
1.9	microRNA-mRNA	12
1.10	Basic (microRNA-mRNA) bind types	13
1.11	microRNA-mRNA Binding Categories	14
2.1	Experimental target identification techniques	17
2.2	microRNA seed sequence	18
2.3	DIANA-microT-CDS flow algorithm	20
2.4	DIANA-microT-CDS parameters	21
3.1	microRNAs entries growth	25
3.2	Ensembl data flow	26
3.3	Ensembl vertebrate species	27
3.4	Ensembl biomart	29
3.5	Intron between microRNA bind site	31
3.6	Gene and Transcript annotations	32
3.7	Chromosome	37
3.8	Database Schema	38
4.1	App Home Page	41
4.2	App navigation bar	42
4.3	App Statistics window	43
4.4	App Help page part1	44
4.5	App Help page part2	45
4.6	App Filter Button	45
4.7	App Filter Window	46
4.8	App: Search by Gene	48
4.9	App: Search by Transcript	48
4.10	App: Search by microRNA	49

4.11 App: Choose transcript after gene input	49
4.12 App: Content of Result page	50
4.13 App: Choose transcript after microRNA input	51
4.14 App: Content of microRNA Result page	52
4.15 App: Content of microRNA Result page, zoomed	52
4.16 App: Flow diagram	53
4.17 App: Chromosome graphic	54
4.18 App: Gene graphic	54
4.19 App: Transcript graphic	55
4.20 App: Transcript region colours	55
4.21 App: Zoomed transcript regions graphic	55
4.22 App: Information about transcript region onMouse	56
4.23 App: Content of result page	56
4.24 App: microRNA colours	57
4.25 App: microRNA family name	58
4.26 App: microRNA family name	59
4.27 App: Information about microRNA onMouse	60
4.28 App: Scroll bar	61
5.1 Web application flow	63
5.2 MVC Diagram	66
5.3 MVT Diagram	67
5.4 Interactions in MVT	67
5.5 Django ORM	68
5.6 Python class to SQL table	69
5.7 Django Template	70
5.8 Django URL mapping to Views	71

List of Tables

3.1	microRNA-target interaction data	31
3.2	Gene and transcript annotations	33
3.3	Biomart data	34
3.4	Initial microRNA data	35
3.5	Processed microRNA data	35
3.6	Initial microRNA family data	35
3.7	Processed microRNA family data	36
3.8	Merged microRNA and microRNA family data	36
3.9	Chromosome data	37
3.10	Processed chromosome data	38

List of Abbreviations

3'UTR	5' Untranslated Region
5'UTR	3' Untranslated Region
AGO	Argonaute
BLAST	Basic Local Alignment Search Tool
CDS	Coding Sequence
DGCR8	DiGeorge syndrome Critical Region 8
ELISA	Enzyme Linked Immunosorbent Assay
FTP	File Transfer Protocol
HPC	High Performance Computing
HITS-CLIP	High Throughput Sequencing of RNA isolated by Crosslinking Immunoprecipitation
KEGG	microRNA Kyoto Encyclopedia of Genes and Genomes
miRISC	microRNA Induced Silencing Complex
miRNA	microRNA
MRE	MicroRNA Recognition Element
mRNA	messenger RNA
NGS	Next Generation Sequencing
PAR-CLIP	Photoactivatable-ribonucleoside-enhanced Crosslinking and Immunoprecipitation
poly(A)	polyadenylation
pre-microRNA	precursor microRNA
pSILAC	pulsed Stable Isotope by Aminoacids in Cell culture
qPCR	quantitative real-time Polymerase Chain Reaction
RNA-Seq	RNA Sequencing
RNA	Ribonucleic Acid
RISC	RNA Induced Silencing Complex
tRNA	transfer RNA

Dedicated to my family...

Chapter 1

Εισαγωγή

1.1 Γιατί η βιοπληροφορική είναι σημαντική

Ο όγκος των πληροφοριών και δεδομένων τα οποία προκύπτουν σήμερα, από τα πειράματα τα οποία γίνονται σε βιολογικό επίπεδο είναι τεράστιος. Έτσι έχουμε την ανάγκη να μπορούμε να αποθηκεύουμε αποδοτικά αυτά τα δεδομένα, να τα προσπελάζουμε γρήγορα και να τα επεξεργαζόμαστε. Επίσης η βιολογία είναι μία επιστήμη στην οποία εκ φύσεως υπάρχουν μεγάλα μεγέθη, αν αναλογιστεί κανείς για παράδειγμα ότι το ανθρώπινο DNA αποτελείται από 3 δισεκατομμύρια βάσεις. Έτσι οι συγκρίσεις μεταξύ αλληλουχιών DNA, οι αναλύσεις των γονιδιωματικών ή εξελικτικών δεδομένων, η μοντελοποίηση των δεδομένων μας, η συναρμογή τμημάτων αλληλουχιών DNA, απαιτούν μεγάλο χρόνο και ενέργεια για να εκτελεστούν. Εδώ έρχεται η βιοπληροφορική να πληρώσει αυτό το κενό, δημιουργώντας εργαλεία software και αξιοποιώντας τους διαθέσιμους υπολογιστικούς πόρους που ολοένα και αυξάνονται, εκμεταλλευόμενη ταυτοχρόνως νέες τεχνολογίες, όπως την τεχνητή νοημοσύνη, τα νευρωνικά δίκτυα, το HPC και την εξόρυξη δεδομένων.

1.2 Θεωρητικό βιολογικό υπόβαθρο

Είναι σκόπιμο να αναφερθούμε σε ορισμένες βιολογικές διεργασίες και να αναλύσουμε κάποιους βιολογικούς όρους σε θεωρητικό υπόβαθρο, τους οποίους θα χρησιμοποιήσουμε στη συνέχεια της εργασίας μας, καθώς επίσης θα αποτελέσουν και βασικό κομμάτι για την κατανόηση της εφαρμογής την οποία θα παρουσιάσουμε. Στην αρχή, θα μιλήσουμε για το κεντρικό δόγμα της βιολογίας καθώς και για το mRNA, ενώ στη συνέχεια θα αναφερθούμε στα microRNAs.

1.2.1 Από το DNA στην πρωτεΐνη

Πώς δημιουργείται μία πρωτεΐνη; Ποιά είναι η διαδικασία η οποία ακολουθείται ξεκινώντας από το DNA; Παρακάτω αναλύουμε αυτή τη διεργασία, η οποία είναι ζωτικής σημασίας για τον οργανισμό μας.

Στα ανθρώπινα κύτταρα υπάρχει μία δομή, η οποία ονομάζεται πυρήνας. Ο πυρήνας περιέχει το γονιδίωμα. Στους ανθρώπους το γονιδίωμα χωρίζεται σε

24 χρωμοσώματα. Κάθε χρωμόσωμα εμπεριέχει μια μακριά αλληλουχία DNA, η οποία είναι αναδιπλωμένη γύρω από κάποιες πρωτεΐνες επονομαζόμενες ιστόνες (histones). Μέσα στο DNA υπάρχουν τμήματα τα οποία καλούνται γονίδια και τα οποία εμπεριέχουν τις πληροφορίες για τη σύνθεση των πρωτεϊνών. Όταν ένα γονίδιο “ενεργοποιηθεί”, το ένζυμο RNA πολυμεράση προσδένεται στην αρχή του γονιδίου και κινείται κατά μήκος του DNA δημιουργώντας μία αλυσίδα (strand) messenger RNA (mRNA), με βάσεις συμπληρωματικές ως προς αυτές του DNA. Αυτή η διαδικασία ονομάζεται μεταγραφή (transcription). Πριν όμως το mRNA χρησιμοποιηθεί ως πρότυπο για την παραγωγή πρωτεϊνών πρέπει να υποστεί κάποια επεξεργασία, μέρος της οποίας είναι, η αφαίρεση κάποιων τμημάτων του (ιντρόνια) και η σύνδεση κάποιων άλλων μεταξύ τους (εξώνια). Αυτή η επεξεργασία ονομάζεται μάτισμα (splicing). Έπειτα το mRNA βγαίνει από τον πυρήνα και κατευθύνεται στο κυτταρόπλασμα, όπου και προσδένεται στα ριβοσώματα. Τα ριβοσώματα “διαβάζουν” καθε φορά μία τριάδα νουκλεοτιδίων του mRNA (κωδικόνιο) και το κατάλληλο αμινοξύ προστίθεται στην αλυσίδα του πολυπεπτιδίου, έτσι ώστε στο τέλος να παραχθεί η τελική αλυσίδα απο αμινοξέα (πρωτεΐνη), η οποία αποκτά μια τρισδιάστατη δομή και καθορίζει τη λειτουργία της. Το μόριο transfer RNA (tRNA) είναι αυτό, το οποίο μεταφέρει τα αμινοξέα στα ριβοσώματα. Αυτή η διαδικασία ονομάζεται μετάφραση (translation). Η παραπάνω διαδικασία στο σύνολό της αποκαλείται ως το “κεντρικό δόγμα της μοριακής βιολογίας” (Σχήμα 1.1) [44].

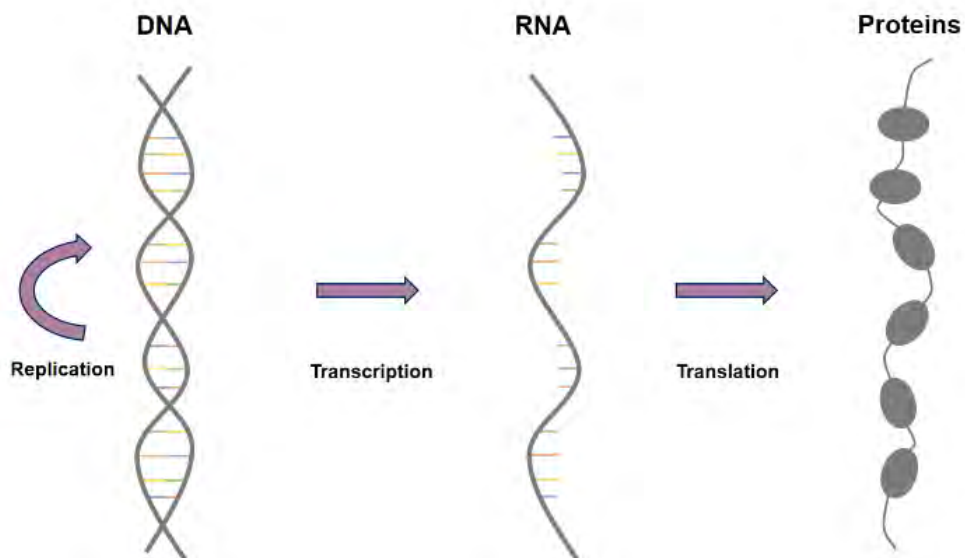


FIGURE 1.1: Αναπαράσταση του Κεντρικού Δόγματος της Βιολογίας.
[15]

1.2.2 Δομή mRNA

Όπως προαναφέραμε, το mRNA διαδραματίζει σημαντικότατο ρόλο στη διαδικασία δημιουργίας των πρωτεϊνών, παράλληλα όμως είναι και αυτό, όπως θα δούμε στη συνέχεια, το οποίο στοχεύουν τα microRNAs, με σκοπό την καταστολή της δημιουργίας μίας πρωτεΐνης. Αξίζει λοιπόν σε αυτό το σημείο να αναφερθούμε στη δομή του και να παρατηρήσουμε τις επιμέρους περιοχές του.

Αρχικά, είναι σημαντικό να πούμε, πως γενικά το μόριο του RNA είναι μία μονόκλωνη αλυσίδα, η οποία μπορεί να αποτελείται από τέσσερις πιθανές βάσεις (νουκλεοτίδια). Αυτές είναι η αδενίνη (A), η γουανίνη (G), η κυτοσίνη (C) και η ουρακίλη (U). Ας δούμε τις περιοχές οι οποίες συντελούν το mRNA, όπως αυτές αναπαρίστανται στο παρακάτω σχήμα (Σχήμα 1.2).

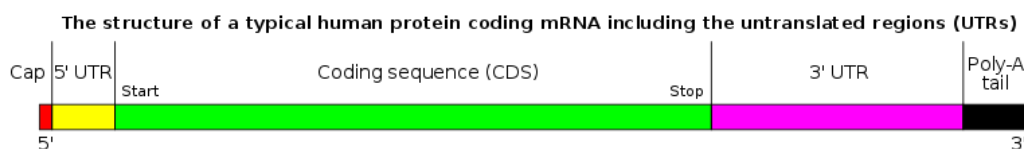


FIGURE 1.2: Η δομή ενός ώριμου ευκαριωτικού mRNA. Ένα πλήρως επεξεργασμένο mRNA αποτελείται από: 5'cap, 5'UTR, coding region, 3'UTR, and poly(A) tail.

Στο αριστερό μέρος του mRNA παρατηρούμε το “5'cap”. Το 5'cap είναι ένα τροποποιημένο νουκλεοτίδιο γουανίνης (G), το οποίο έχει προστεθεί στο εμπρόσθιο ή 5' άκρο, ενός ευκαριωτικού αγγελιαφόρου RNA (mRNA) λίγο μετά την έναρξη της μεταγραφής. Η παρουσία του είναι κρίσιμη για την αναγνώριση του mRNA από το ριβόσωμα, αλλά και για την προστασία του από τις ριβονουκλεάσες¹. Από την άλλη μεριά του mRNA παρατηρούμε την “3'poly(A)” ουρά, μία μακριά αλληλουχία νουκλεοτιδίων αδενίνης (συχνά αρκετές εκατοντάδες), η οποία προστίθεται στο 3' άκρο του pre-mRNA (διαδικασία πολυαδενυλίωσης). Η 3'poly(A) ουρά παίζει σημαντικό ρόλο στην εξαγωγή του mRNA από τον πυρήνα, στη μετάφραση και στον τερματισμό της μεταγραφής, καθώς επίσης προστατεύει το mRNA από την αποικοδόμηση (degradation) από εξωνουκλεάσες. Στη συνέχεια, παρατηρούμε την “περιοχή κωδικοποίησης” (coding region - CDS), η οποία αποτελείται από κωδικόνια². Οι περιοχές αυτές αποκωδικοποιούνται και μεταφράζονται σε πρωτεΐνες με τη βοήθεια των ριβοσωμάτων. Γενικά αρχίζουν με το κωδικόνιο έναρξης AUG (μεθειονίνη) και τελειώνουν με ένα κωδικόνιο τερματισμού, το οποίο μπορεί να είναι το UAA, UAG ή UGA. Τέλος, εκατέρωθεν της κωδικής περιοχής, βρίσκονται οι επονομαζόμενες “five prime untranslated region” (5' UTR) και “three prime untranslated region” (3' UTR), αμετάφραστες περιοχές, οι οποίες συνήθως δεν κωδικοποιούν κάποια πρωτεΐνη. Παρ' όλ' αυτά, τους έχουν αποδοθεί

¹Οι ριβονουκλεάσες (RNases) είναι μια μεγάλη ομάδα υδρολυτικών ενζύμων, οι οποίες αποικοδομούν μόρια RNA.

²Κωδικόνιο : μία τριάδα νουκλεοτιδίων, η οποία κωδικοποιεί ένα αμινοξύ στην πρωτεΐνη

αρκετοί ρόλοι στην έκφραση γονιδίου, όπως για παράδειγμα, η σταθερότητα του mRNA, ο εντοπισμός του και η αποτελεσματικότητα της μετάφρασης. Όσον αφορά στην αποτελεσματικότητα της μετάφρασης, οι πρωτεΐνες οι οποίες προσδένονται είτε στη 3'UTR είτε στη 5'UTR περιοχή, μπορεί να επηρεάσουν τη μετάφραση, επηρεάζοντας την ικανότητα του ριβοσώματος να δεσμεύεται στο mRNA. Τα microRNAs τα οποία συνδέονται με τη 3'UTR περιοχή, μπορούν επίσης να επηρεάζουν τη μεταφραστική αποτελεσματικότητα ή τη σταθερότητα του mRNA [41].

1.3 microRNAs

Σε αυτή την ενότητα θα αναφερθούμε στα microRNAs, τα οποία είναι το βασικό κομμάτι στην εργασίας μας. Θα πούμε τί είναι τα microRNAs, πως δημιουργούνται, ποιος είναι ο ρόλος και ο μηχανισμός δράσης τους, καθώς επίσης και άλλες πληροφορίες, οι οποίες σχετίζονται με αυτά.

1.3.1 Τί είναι τα microRNAs

Τα microRNAs είναι μικρά μονόκλιωνα μόρια RNA, μήκους 19-24 νουκλεοτιδίων, τα οποία δεν κωδικοποιούνται, δε συντελούν δηλαδή στην παραγωγή κάποιας πρωτεΐνης. Αντίθετα, εμπλέκονται στην “αρνητική” μετά-μεταγραφική ρύθμιση της έκφρασης των γονιδίων και ως εκ τούτου, παίζουν καθοριστικό ρόλο σε ένα πλήθος βιολογικών διεργασιών. Δρουν είτε μέσω της άμεσης αποδόμησης του RNA-στόχου (transcript degradation, cleavage), είτε αλληλεπιδρώντας με το mRNA και προκαλώντας την “καταστολή της μετάφρασης” (translation repression) [22]. Τα πρώτα microRNAs ανακαλύφθηκαν το 1993 στον οργανισμό “*Caenorhabditis elegans*” από τους Ambros, Lee και Feinbaum. Παρ’ όλ’ αυτά, ήταν το 2001 όπου έγινε φανερό ότι τα microRNAs είναι ευρύτερα διαδεδομένα και σε αφθονία στα κύτταρα (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001) [29]. Βρίσκονται στους περισσότερους ευκαριωτικούς οργανισμούς, συμπεριλαμβανομένου και του ανθρώπινου. Εικάζεται ότι αποτελούν το 1-5% του ανθρώπινου γονιδιώματος και ρυθμίζουν τουλάχιστον το 30% των γονιδίων που παράγουν κάποια πρωτεΐνη. Μέχρι σήμερα έχουν ανακαλυφθεί παραπάνω από 2500 microRNAs στον ανθρώπινο οργανισμό, σύμφωνα με τις καταχωρίσεις στη βάση δεδομένων miRBase. Τέλος, παρ’ όλο που έχουμε μία γενική κατανόηση για τη λειτουργία των microRNAs, οι λεπτομέρειες οι οποίες αφορούν στη βιογένεση τους και στη διαδικασία σίγασης των γονιδίων (gene silencing) είναι ακόμα εν μέρει ασαφείς [20].

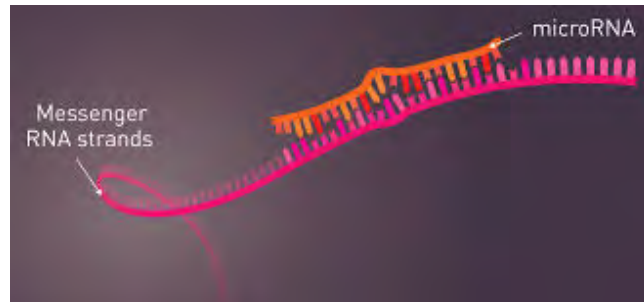


FIGURE 1.3: Πως μοιάζει ένα μόριο microRNA.

1.3.2 Ονοματολογία των *microRNAs*

Η ονοματολογία και κατ' επέκταση η αρίθμηση των *microRNAs* είναι ακολουθιακή. Αν θεωρήσουμε για παράδειγμα, ότι το τελευταίο δημοσιευμένο *microRNA* είναι το *mir-352*, τότε το επόμενο, το οποίο θα δημοσιευτεί θα λάβει τον αριθμό 353, δηλαδή θα είναι το *mir-353*. Τα ονόματα εν γένει είναι της μορφής “*hsa-mir-121*”. Τα πρώτα τρία γράμματα υποδηλώνουν τον οργανισμό. Εδώ το “*hsa*” υποδηλώνει τον ανθρώπινο οργανισμό (*homo sapiens*). Ύστερα, υπάρχει μία παύλα και έπειτα είτε το πρόθεμα “*mir*” ή “*miR*”. Το πρόθεμα “*miR*” υποδηλώνει ένα ώριμο (*mature*) *miRNA*, σε αντίθεση με το πρόθεμα “*mir*”, το οποίο αναφέρεται στο *miRNA gene* και *stem-loop* (δομή φουρκέτας) του πρώιμου μεταγράφου. Υπάρχει η περίπτωση δύο ή παραπάνω *miRNAs* να είναι σχεδόν ίδια, δηλαδή να διαφέρουν σε 1-2 νουκλεοτίδια, τότε χρησιμοποιούμε τον ίδιο αριθμό στο όνομα του *microRNA*, προσθέτωντας ένα ακόμα γράμμα στο τέλος του ονόματος. Για παράδειγμα *miR-120a* και *miR-120b*. Ακόμα, όταν δύο πρόδρομα *miRNAs* βρίσκονται σε διαφορετικές περιοχές, αλλά παράγουν το ίδιο ώριμο *miRNA*, τότε τα συμβολίζουμε με τον παρακάτω τρόπο. Λόγου χάρη, *hsa-mir-121-1* και *hsa-mir-121-2*, τα οποία καταλήγουν στο *hsa-miR-121*. Τέλος, με τις καταλήξεις “-3p” και “-5p” στο όνομα, υποδηλώνουμε δύο ώριμα *miRNAs*, τα οποία προέρχονται από αντίθετα “*arms*”, του ίδιου όμως *pre-miRNA*. Να σημειώσουμε ότι τα *microRNAs* με τα ονόματα “*let-7*” και “*lin4*” αποτελούν εξαίρεση στον κανόνα, καθώς ήταν αυτά που ανακαλύφθηκαν πρώτα και διατηρούμε το όνομά τους για ιστορικούς λόγους [40] [26].

1.3.3 Η Βιογένεση των *microRNAs*

Η βιογένεση των *microRNAs* είναι μία διαδικασία, για την οποία ξέρουμε τα βασικά βήματα τα οποία ακολουθούνται, παρ' όλ' αυτά κάποια σημεία παραμένουν ακόμα άγνωστα ή ασαφή. Τα μετάγραφα των *microRNAs* προέρχονται, είτε από τα εσώνια των γονιδίων τα οποία κωδικοποιούν κάποια πρωτεΐνη, είτε από περιοχές οι οποίες βρίσκονται εκτός των γονιδίων που κωδικοποιούν κάποια πρωτεΐνη και σπανιότερα από τα εξώνια των μη-κωδικών μεταγράφων είτε από περιοχές κωδικών εξωνίων, ή αμετάφραστες περιοχές (UTRs) [29].

Η βιογένεση των microRNAs συμβαίνει σε πολλαπλά βήματα ξεκινώντας από τον πυρήνα και καταλήγοντας στο κυτταρόπλασμα, όπου και λαμβάνεται η τελική του μορφή. Η μεταγραφή των microRNAs ξεκινάει με τη βοήθεια της RNA πολυμεράσης II. Τα πρώιμα παραγόμενα μετάγραφα miRNAs (pri-miRNAs) έχουν μήκος εκατοντάδων έως χιλιάδων βάσεων. Αυτά τα μετάγραφα αναδιπλώνονται, δημιουργώντας μία “δομή φουρκέτας” (hairpin/stem loop structure), η οποία γενικά έχει κάποια μη ζευγαρωμένα νουκλεοτιδία και μονόκλωνες προεκτάσεις τόσο στο 5’ άκρο, όσο και στο 3’ άκρο (5’ capped, spliced and polyadenylated at the 3’ end). Αυτή η “stem loop” δομή είναι η βάση για την περαιτέρω επεξεργασία που θα ακολουθήσει από ένα RNA III ένζυμο, το οποίο βρίσκεται στον πυρήνα του κυττάρου εν ονόματι “Drosha”. Η Drosha αλληλεπιδρά στον ανθρώπινο οργανισμό με την πρωτεΐνη “DGCR8”, μία εξειδικευμένη πρωτεΐνη με περιοχές δέσμευσης του δίκλωνου RNA και σχηματίζει ένα σύμπλεγμα, το οποίο ονομάζεται “microprocessor complex”. Αυτό το σύμπλεγμα εκτελεί μία αντίδραση διάσπασης (cleavage reaction), αφαιρώντας τις προεκτάσεις στο 3’ άκρο και 5’ άκρο, σχηματίζοντας έτσι ένα μετάγραφο μήκους 60-70 νουκλεοτιδίων, το οποίο ονομάζεται “precursor-miRNA”. Ο ενδοπυρηνικός παράγοντας με όνομα “Exportin 5”, αναγνωρίζει το pre-microRNA που δημιουργήθηκε και το μεταφέρει στο κυτταρόπλασμα για περαιτέρω επεξεργασία. Εκεί μία δεύτερη ενδονουκλεολυτική διάσπαση (endonucleolytic cleavage reaction) λαμβάνει χώρα, η οποία αναφέρεται συνήθως ως “dicing”, με τη βοήθεια ενός RNA III ενζύμου, το οποίο ονομάζεται “Dicer”. Το ένζυμο Dicer βρίσκεται σε ένα σύμπλεγμα με μία άλλη δίκλωνη πρωτεΐνη πρόσδεσης RNA, η οποία ονομάζεται “TRBR2” (tar RNA binding protein 2). Το προϊόν το οποίο προκύπτει ύστερα από την δεύτερη αυτή διάσπαση ονομάζεται “miR:miR*duplex”, έχει μήκος περίπου 22 νουκλεοτιδίων και δεν είναι πλήρως συμπληρωματικό. Οι δύο αλυσίδες (strands) του ονομάζονται “Guide strand” και “Passenger Strand”. Μόλις δημιουργηθεί το “miR:miR*duplex” προσδένεται στην πρωτεΐνη που ονομάζεται “Argonaut” και σχηματίζεται ένα σύμπλεγμα, το “miRISC complex”. Αν το “miR:miR*duplex” προσδέσει με τον κατάλληλο προσανατολισμό στην πρωτεΐνη Argonaut, τότε η μία strand, εν ονόματι “miR guide strand” διατηρείται, ενώ η άλλη (miR passenger strand) αφαιρείται. Αυτή η διαδικασία της διατήρησης της guide strand πάνω στην πρωτεΐνη AGO και η αφαίρεση της passenger strand ονομάζεται “Sorting”. Το “miRISC complex” είναι τώρα έτοιμο να προσδέσει σε ένα στόχο-mRNA και η miR guide strand είναι αυτή, η οποία τελικά θα καθορίσει τον στόχο RNA που θα σιγάσει (silenced). Στην αντίθετη περίπτωση, όπου έχουμε ένα “ατελή” miR:miR*duplex, δεν έχουμε ακόμα προσδιορίσει και κατανοήσει πλήρως, πώς η passenger strand αναγνωρίζεται και αφαιρείται από την πρωτεΐνη Argonaute με επιτυχία [27]. Παρακάτω φαίνεται σε σχηματική αναπαράσταση η βιογένεση ενός microRNA (Σχήμα 1.4).

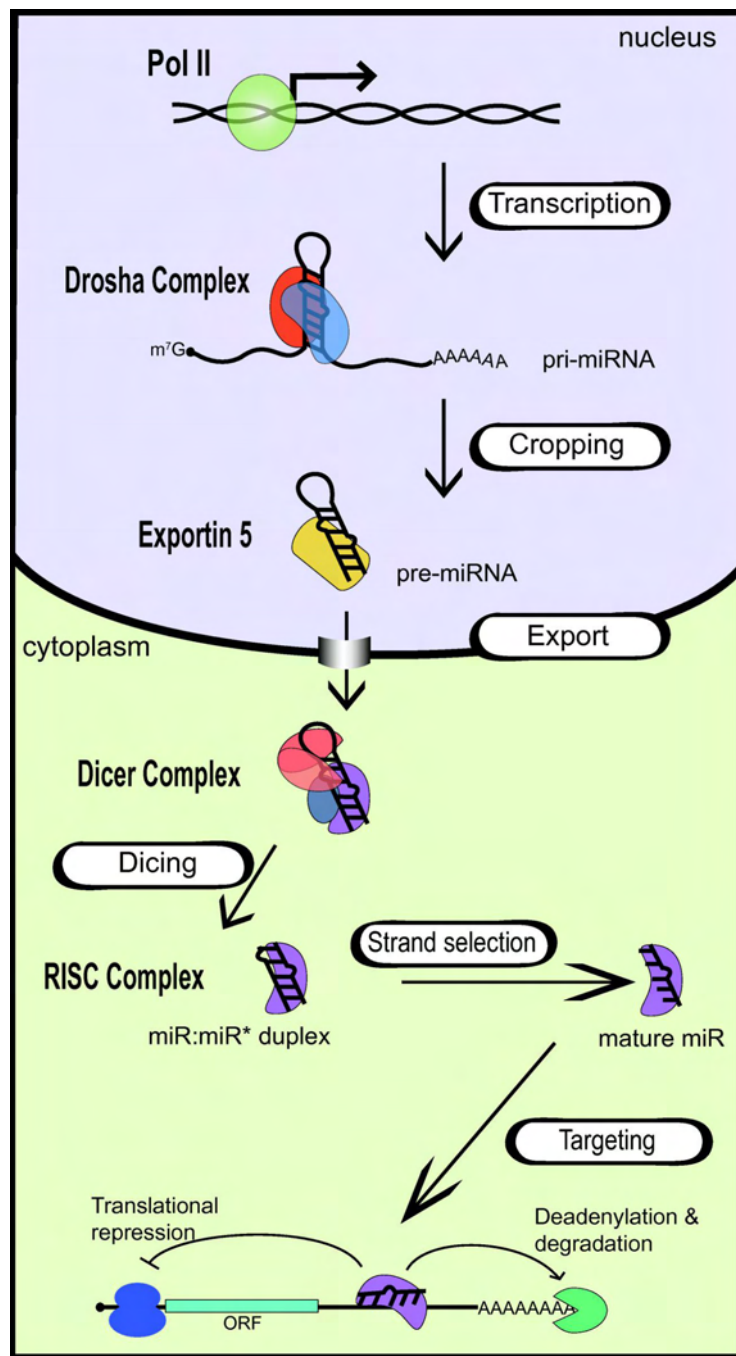


FIGURE 1.4: Σχηματική αναπαράσταση της βιογένεσης ενός mi-croRNA [5].

1.3.4 Ο Μηχανισμός Δράσης των microRNAs

Το “miRISC σύμπλεγμα”, το οποίο δημιουργήθηκε, όπως είδαμε παραπάνω, αναγνωρίζει το mRNA-στόχο και ρυθμίζει αρνητικά την έκφραση γονιδίου με έναν από τους παρακάτω μετά-μεταγραφικούς μηχανισμούς.

- Καταστολή της μετάφρασης (translation repression)
- Διάσπαση του mRNA (mRNA cleavage)

Μελέτες έχουν δείξει ότι οι περισσότερες θέσεις πρόσδεσης στα ζωικά mRNA βρίσκονται στη 3'UTR περιοχή, ως πολλαπλά αντίγραφα. Εν γένει, τα miRNAs των ζωικών οργανισμών προσδένονται στο mRNA με αναντιστοιχίες (mismatches) και εξογκώματα (bulges). Αντίθετα, οι θέσεις πρόσδεσης στα mRNA των φυτικών οργανισμών βρίσκονται στο κέντρο των “συμπληρωματικών περιοχών” (complementary regions), και τα περισσότερα miRNAs έχουν υψηλό βαθμό συμπληρωματικότητας των βάσεων, με τον mRNA στόχο τους [14].

Ο βαθμός συμπληρωματικότητας των βάσεων microRNA-mRNA είναι ο κύριος καθοριστικός παράγοντας, όσον αφορά στο ρυθμιστικό μηχανισμό που χρησιμοποιείται. Στην περίπτωση όπου ο βαθμός συμπληρωματικότητας είναι υψηλός, καθίσταται δυνατή η αποικοδόμηση των αλληλουχιών του mRNA στόχου, μέσω του μηχανισμού διάσπασης του mRNA. Σε αυτή την περίπτωση απαιτούνται οι πρωτεΐνες Ago, GW182, και ο κυτταρικός μηχανισμός απομάκρυνσης και αποδέσμευσης. Αντιθέτως, στην περίπτωση όπου υπάρχει μια κεντρική αναντιστοιχία, δρα ο μηχανισμός καταστολής της μετάφρασης [14] (Σχήμα 1.5).

Εκτός βέβαια από τη συμπληρωματικότητα μεταξύ miRNA και mRNA, διάφοροι άλλοι παράγοντες μπορούν να επηρεάσουν τη δράση του miRNA, όπως η εξασθενημένη επεξεργασία, η μεθυλίωση, οι γονιδιακοί πολυμορφισμοί, η γονιδιακή ενίσχυση, η έλλειψη της πρωτεΐνης Dicer και άλλα [36]. Μια πρόσφατη μελέτη έδειξε ότι η αποικοδόμηση του mRNA αντιπροσωπεύει τον κύριο μηχανισμό ρύθμισης του miRNA. Οι συγγραφείς έδειξαν, ότι περίπου το 84% όλων των στόχων mRNA που κωδικοποιούν κάποια πρωτεΐνη υποβάλλονται σε αποικοδόμηση όταν αναγνωρίζονται από τα σχετικά miRNAs [36]. Τέλος, αξίζει να σημειώσουμε ότι όσον αφορά στους μηχανισμούς δράσης των microRNAs υπάρχουν αρκετά “γκρίζα σημεία”. Για παράδειγμα, ο ακριβής μηχανισμός για την καταστολή της μετάφρασης από το miRISC σύμπλεγμα είναι ακόμα άγνωστος, ενώ πρέπει ακόμα να αποσαφηνιστεί και το κατά πόσο η καταστολή λαμβάνει χώρα στο μεταφραστικό ή στο μετα-μεταφραστικό επίπεδο. Επιπρόσθετα, η ακριβής διαδικασία της επιλογής των στόχων mRNA δεν έχει ακόμη καθοριστεί, ωστόσο έχει αποδειχθεί ότι ο αριθμός, ο τύπος και οι θέσεις των αναντιστοιχιών μεταξύ microRNA και mRNA, διαδραματίζουν κρίσιμο ρόλο στην επιλογή των μηχανισμών αποδόμησης ή καταστολής της μετάφρασης [14].

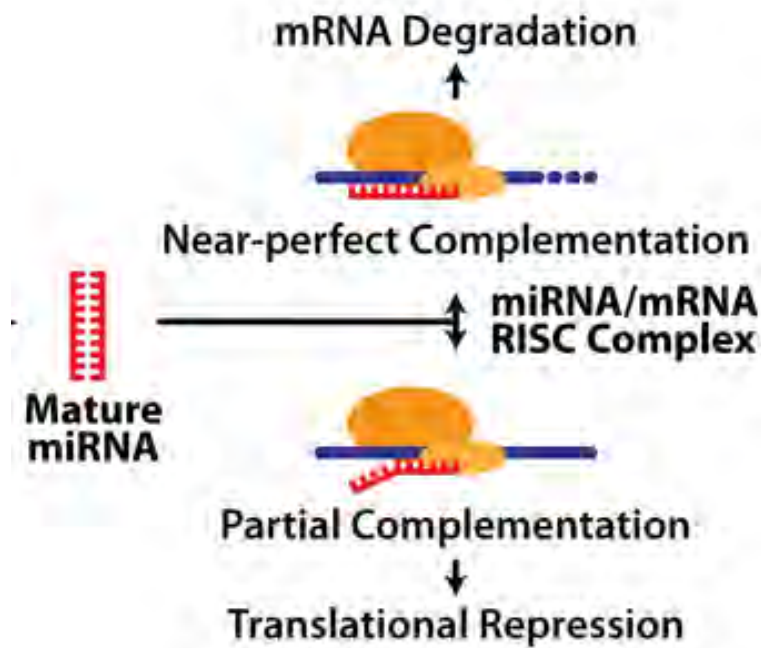


FIGURE 1.5: Μηχανισμός δράσης miRNA κατά του mRNA στόχου. Ανάλογα με τη συμπληρωματικότητα μεταξύ του miRNA και της 3'UTR περιοχής του mRNA στόχου, υπάρχουν δύο πιθανοί μηχανισμοί δράσης miRNA. [30]

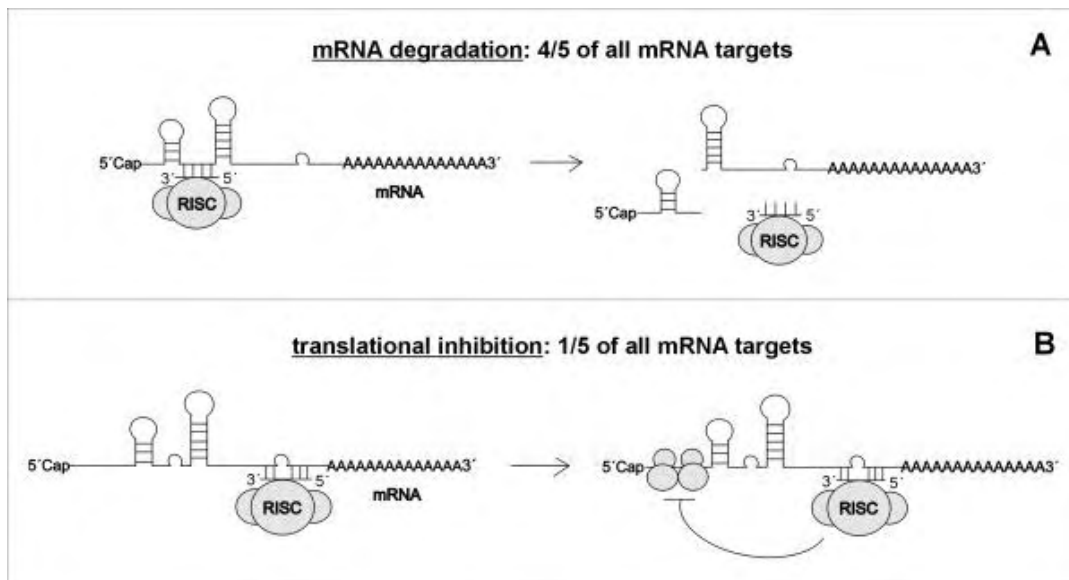


FIGURE 1.6: Μηχανισμός δράσης miRNA κατά του mRNA στόχου. Το 84% όλων των στόχων mRNA που κωδικοποιούν κάποια πρωτεΐνη υποβάλλονται σε αποικοδόμηση, σύμφωνα με μία πρόσφατη μελέτη [36].

1.3.5 Η Λειτουργία και ο Ρόλος των microRNAs

Η βασική λειτουργία των microRNAs είναι η ρύθμιση της μεταγραφικής έκφρασης των γονιδίων και καθένα από αυτά μπορεί να επηρεάζει όχι μόνο ένα, αλλά ένα σημαντικό αριθμό γονιδίων. Τα microRNAs διαδραματίζουν βασικό ρόλο σε πολλές βιολογικές διεργασίες, όπως στον πολλαπλασιασμό των βλαστοκύτταρων, στη διαίρεση και διαφοροποίηση, στην ανοσία, στην κυτταρική σηματοδότηση, απόπτωση και μεταβολισμό, στην ανάπτυξη του καρδιακού και σκελετικού μύος, στη νευρογένεση, έκκριση ινσουλίνης, στο μεταβολισμό της χοληστερόλης, στη γήρανση και σε πολλά ακόμη σημεία [22].

Εκτός όμως από τον φυσιολογικό τους ρόλο, ένας μεγάλος αριθμός μελετών συμπαιρνεί ότι τα miRNAs σχετίζονται με μία τεράστια ποικιλία ασθενειών, όπως ο καρκίνος, οι ιογενείς λοιμώξεις, οι καρδιαγγειακές παθήσεις, οι ασθένειες, οι μεταβολικές διαταραχές, οι αυτοάνοσες παθήσεις, οι μυοσκελετικές διαταραχές, καθώς και οι νευροψυχιατρικές παθολογικές καταστάσεις. Τα miRNAs μπορούν να επηρεάσουν τη γονιδιακή έκφραση σε διάφορους ιστούς. Έτσι, πιθανές μεταβολές στη συγκέντρωση των miRNAs που προκαλούνται από επιγενετική αποσιώπηση ή από απορυθμισμένους παράγοντες μεταγραφής, γενετικές διαταραχές / ανωμαλίες, μπορούν να οδηγήσουν στην απορρύθμιση των αντίστοιχων γονιδίων στόχων τους [22]. Ως αποτέλεσμα των παραπάνω, τα miRNAs αποτελούν αντικείμενο έντονης μελέτης τα τελευταία χρόνια, για τη δυναμική τους ως υποψήφιοι διαγνωστικοί ή και προγνωστικοί βιοδείκτες, αλλά και ως δείκτες για τα αποτελέσματα ή παρενέργειες των φαρμάκων [13].

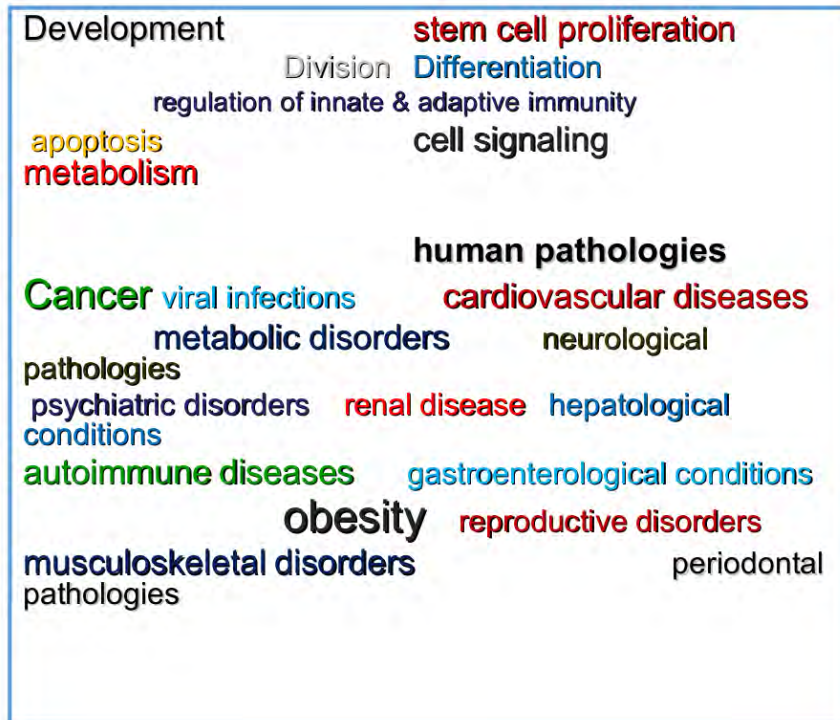
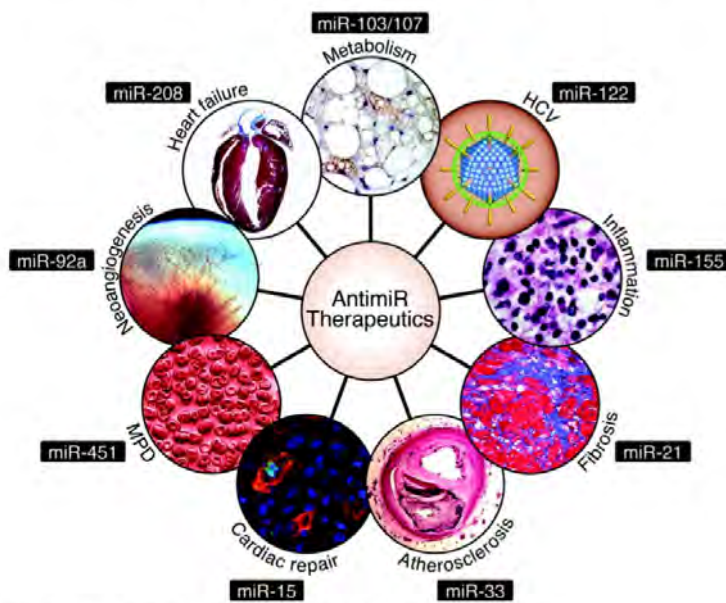


FIGURE 1.7: Τα *microRNAs* εμπλέκονται σε πολλαπλές βιολογικές διεργασίες, αλλά και ασθένειες. [19]

Specific *miRNAs* that are currently being pursued as clinical candidates.



Eva van Rooij et al. *Circ Res.* 2012;110:496-507

FIGURE 1.8: Τα *microRNAs* ως θεραπευτικοί στόχοι. [19]

1.3.6 Seed Περιοχές (Seed Regions)

Ο προσδιορισμός των θέσεων στόχων-mRNA είναι μία από τις πολλές προκλήσεις που υπάρχουν στο πεδίο της βιοπληροφορικής, όσον αφορά στα microRNAs και είναι απολύτως σημαντική για την κατανόηση και την πρακτική εφαρμογή τους. Τα νουκλεοτίδια 2-8, τα οποία αριθμούνται από το 5' άκρο προς το 3' άκρο του miRNA, αποτελούν τη λεγόμενη “seed περιοχή”, έχουν σχεδόν τέλεια συμπληρωματικότητα Watson-Crick με την 3'UTR περιοχή του mRNA και έχουν ιδιαίτερη σημασία για την εύρεση των στόχων-mRNA [4]. Η θέση του mRNA στην οποία προσδένεται το miRNA ονομάζεται MRE (microRNA recognition element) και βρίσκεται στο 3' άκρο του mRNA. Στο παρακάτω σχήμα (Σχήμα 1.9) βλέπουμε τόσο την seed περιοχή του microRNA, όσο και το MRE του mRNA.

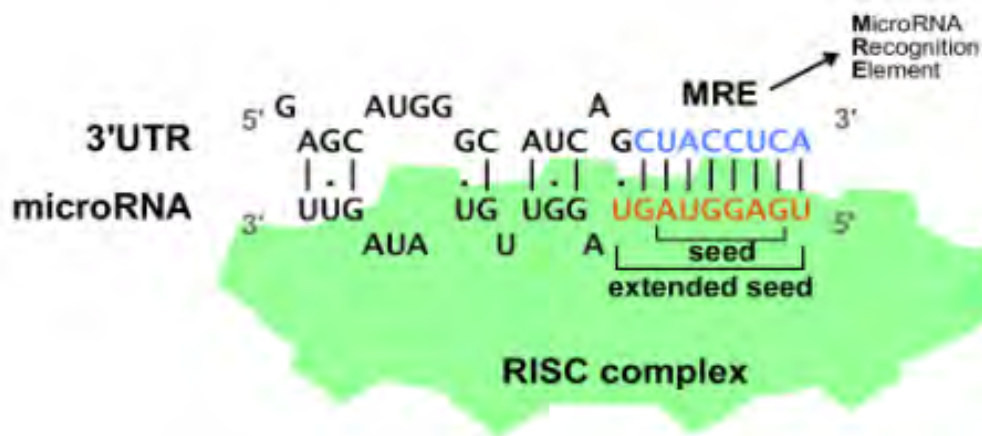


FIGURE 1.9: Seed περιοχή στο microRNA και MRE στο mRNA. [18]

Η περιοχή seed του microRNA δεν προσδένεται πάντα με τον ίδιο τρόπο πάνω στο mRNA. Υπάρχουν ποικίλες περιπτώσεις. Οι κυριότερες είναι οι: 6-mer (bases 2-7) , 7-mer (“7-mer-A1” being bases 1-7, and “7-mer-m8” being bases 2-8) και 8-mer (bases 1-8) [21] , οι οποίες φαίνονται στο παρακάτω σχήμα (Σχήμα 1.10) .

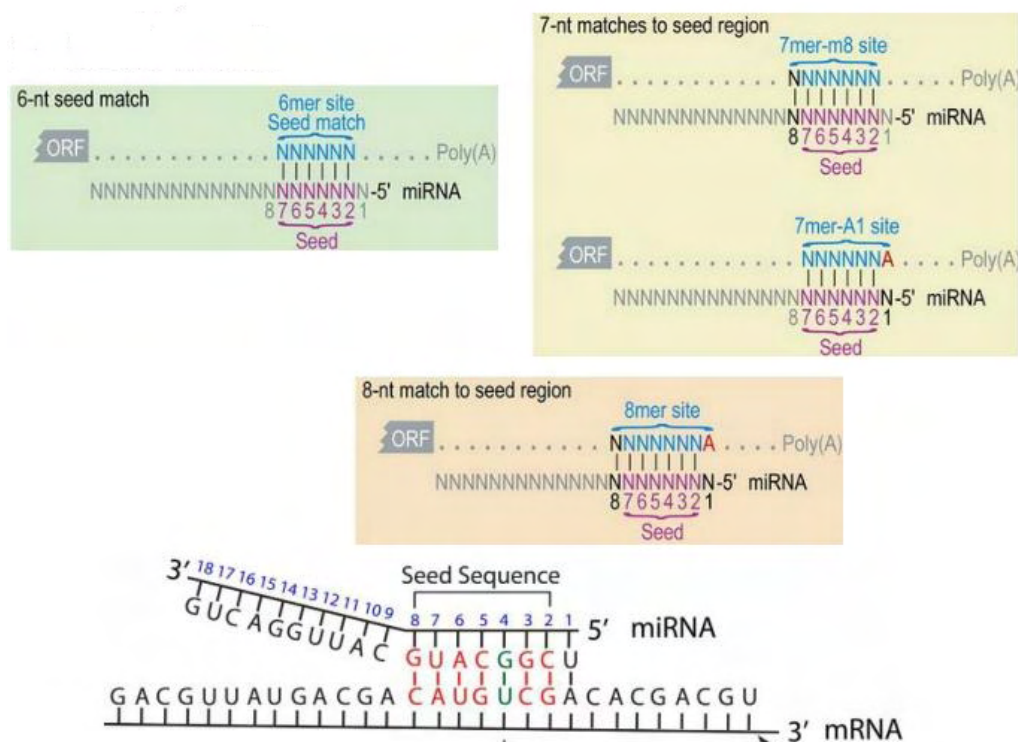


FIGURE 1.10: Βασικοί τύποι πρόσδεσης του microRNA seed στο mRNA (6mer , 7mer , 8mer) [7].

Εν γένει οι seed περιοχές του microRNA είναι τμήματα τα οποία έχουν διατηρηθεί σε υψηλό βαθμό κατά την εξέλιξη ανάμεσα σε διαφορετικά είδη και με βάση αυτές τις περιοχές είναι δυνατή η ταξινόμηση των miRNAs σε οικογένειες και είδη [4]. Πιο συγκεκριμένα έχει διαπιστωθεί ότι οι μεγαλύτερες σε έκταση περιοχές seed, για παράδειγμα περιοχές με 7 ή 8 νουκλεοτίδια, έχουν διατηρηθεί σε μεγαλύτερο βαθμό κατά την εξέλιξη, από ότι οι περιοχές seed μικρότερου μήκους (Ellwanger et al.) [21].

Ορισμένοι βασικοί κανόνες όσον αφορά στην πρόσδεση είναι οι εξής:

- Η πρόσδεση στο 5' άκρο (5'UTR) του microRNA είναι ισχυρή.
- Η πρόσδεση στο 3' άκρο (3'UTR) του microRNA μπορεί να είναι πιο χαλαρή.
- Το μήκος της κεντρικής διόγκωσης (central bulge) μπορεί να αλλάζει.

Στον παρακάτω πίνακα (Σχήμα 1.11) φαίνονται όλες οι περιπτώσεις των δεσμών μεταξύ microRNA – mRNA. Παρατηρούμε ότι, όταν η seed περιοχή είναι μικρού μήκους (4mer , 5mer) και ως εκ τούτου ο δεσμός στο 5' άκρο δεν είναι τόσο ισχυρός, τότε υπάρχει και επιπλέον αντισταθμιστικός δεσμός στο 3' άκρο του microRNA [18].

Binding category	Description	Graphic
9mer	9 consecutive matches	<pre> U GU AGUAA GCU GUGCGA ACUACCUCA - - - - UGA UAUGUU UGAUGGAGU </pre>
8mer	8 consecutive matches	<pre> U U AUUU CU UAUAACC CUACCUCA GA AUGUUGG GAUGGAGU UU U AU </pre>
7mer	7 consecutive matches	<pre> U AUU G AG GAUUGUG U UAC UACCUCA - - - - UUGAUAU G AUG AUGGAGU GUU G </pre>
6mer	6 consecutive matches	<pre> U C UUG U CUAUAC ACCU UACCUC GAUAUG UGGA AUGGAG UU U UG U </pre>
9mer with wobble (G-U)	8 matches + wobble + 3' binding	<pre> C ACA ACAGCC ACUGCCUCA UGUUGG UGAUGGAGU UUGAUA A </pre>
8mer with wobble (G-U)	7 matches + wobble + 3' binding	<pre> C C CC G GACU CAGCCU ACUGCCUC - UUGA GUUGGA UGAUGGAG UAU U </pre>
7mer with wobble (G-U)	6 matches + wobble + 3' binding	<pre> AA UC AUACGACCU UAUCUCA UAUGUUGGA AUGGAGU UUGA UG </pre>
8mer with miRNA bulge	8 matches + bulge + 3' binding	<pre> C ACA ACAGCC ACU CCUCA UGUUGG UGA GGAGU UUGAUA A U </pre>
8mer with mismatch	8 matches + mismatch + 3' binding	<pre> C ACA G ACAGCC ACU CCUCA UGUUGG UGA GGAGU UUGAUA A G </pre>
8mer with target bulge	8 matches + bulge + 3' binding	<pre> C ACA G ACAGCC ACU CCUCA UGUUGG UGA GGAGU UUGAUA A </pre>
5mer	5 matches + 3' binding	<pre> C AA ACAGCC ACGGCCUCA UGUUGG UGUUGGAGU UUGAUA A </pre>
4mer	4 matches + 3' binding	<pre> C CA GC ACAGCC ACU CUCA UGUUGG UGA GAGU UUGAUA GA </pre>

FIGURE 1.11: Περιπτώσεις δεσμών μεταξύ microRNA και mRNA.

[18]

Chapter 2

Αναγνώριση αλληλεπιδράσεων μεταξύ microRNA και mRNA

2.1 Μέθοδοι αναγνώρισης

Η αναγνώριση των στόχων mRNA αποτελεί ένα κρίσιμο μέρος της διαδικασίας, για την κατανόηση των φυσιολογικών βιολογικών διεργασιών των miRNAs και των ρόλων τους στην ανάπτυξη μίας ασθένειας. Οι μέθοδοι οι οποίες ακολουθούνται για την αναγνώριση των στόχων είναι δύο. Οι πειραματικές (experimental) και οι υπολογιστικές (computational-in silico). Στη συνέχεια θα αναλύσουμε κάθε μία από αυτές.

2.1.1 Πειραματικές μέθοδοι για την αναγνώριση αλληλεπιδράσεων μεταξύ miRNA και mRNA (Experimental)

Οι πειραματικές μέθοδοι είναι απαραίτητες για την εξακρίβωση των αλληλεπιδράσεων των microRNAs με τα γονίδια-στόχους, παρόλαυτα αυτές οι μέθοδοι υπόκεινται σε ορισμένους περιορισμούς, όπως θα δούμε στη συνέχεια [19]. Οι πειραματικές μέθοδοι-τεχνικές συνήθως χωρίζονται σε δύο κατηγορίες. Αυτές της χαμηλής απόδοσης (low-yield) και αυτές της υψηλής (high-throughput). Ο διαχωρισμός αυτός γίνεται με βάση το πεδίο της εφαρμογής τους και τον αριθμό των αποτελεσμάτων που λαμβάνονται ανά πείραμα [16]. Οι πιο συνηθισμένες τεχνικές χαμηλής απόδοσης που χρησιμοποιούνται είναι οι “reporter genes”, “qPCR” (quantitative polymerase chain reaction) και η “western blotting”. Η μέθοδος “reporter genes” χρησιμοποιείται για την επικύρωση των περιοχών πρόσδεσης (binding site validation), ενώ οι “qPCR”, “western blotting” ή “ELISA” (enzyme-linked immunosorbent assay) συνήθως συνδυάζονται, με σκοπό να αναγνωρίσουν αλληλεπιδράσεις οι οποίες οδηγούν στην καταστροφή του microRNA ή στην καταστολή της μετάφρασης [16].

Οι πρώτες τεχνικές υψηλής απόδοσης που ανακαλύφθηκαν θα μπορούσαν να θεωρηθούν ως τεχνικές αυξημένης απόδοσης μεν, μικρότερης ακρίβειας δε [37]. Για παράδειγμα, οι μικροσυστοιχίες (microarrays) μπορούν να χρησιμοποιηθούν για την αναγνώριση αλληλεπιδράσεων μεταξύ miRNA και γονιδίων, θεωρώντας τις ως μία έκδοση της “qPCR” και “northern blotting” υψηλότερης απόδοσης.

Από την άλλη, οι ποσοτικές πρωτεϊνωματικές τεχνικές (quantitative proteomic techniques) (π.χ. pSILAC [25]) μπορούν να θεωρηθούν ως μία γενίκευση υψηλής απόδοσης των “ELISA” και “Western blots”. Οι τεχνικές “επόμενης γενιάς αλληλούχισης” (NGS – next generation sequencing) προσφέρουν έναν συνδυασμό υψηλής απόδοσης και ταυτόχρονα αυξημένης ακρίβειας στα αποτελέσματά τους, με συνέπεια να έχει επέλθει επανάσταση στον τρόπο με τον οποίο αναγνωρίζουμε τις αλληλεπιδράσεις miRNA:γονιδίων. Αυτές οι τεχνικές βασίζονται στην NGS αλληλούχιση των mRNA περιοχών στις οποίες προσδένεται η πρωτεΐνη Argonaute (AGO) και συνήθως συνοδεύονται από την αλληλούχιση small-RNAs, καθώς επίσης και από συμπληρωματικά πειράματα, όπως “RNA-Seq” και “ribosome profiling” [34],[23].

Η τεχνική “HITS-CLIP” ήταν η πρώτη η οποία προσέφερε μία πλήρη εικόνα (transcriptome-wide map) των θέσεων πρόσδεσης των miRNAs [34]. Ωστόσο, οι περιοχές οι οποίες προσδιορίζονται είναι συνήθως ευρείες και έτσι περιπλέκουν την αναγνώριση της ακριβούς θέσης πρόσδεσης των miRNAs, η οποία πραγματοποιείται αλγοριθμικά. Η τεχνική “PAR-CLIP” είναι μια τροποποιημένη μεθοδολογία της “CLIP-Seq”, η οποία ενσωματώνει την 4-θειουριδίνη στα λανθάνοντα RNA, τα οποία στη συνέχεια ανιχνεύονται ως μεταβατικές θέσεις T σε C στις περιοχές διασταυρούμενης αλληλεπίδρασης (cross-linked regions) των AGO: miRNA: RNA [23]. Σε σύγκριση με τα αποτελέσματα τα οποία προκύπτουν από την HITS-CLIP, τα όρια-σύνορα των αναγνωρισμένων περιοχών πρόσδεσης είναι πιο σαφείς και στενότερα, ενώ οι μεταλλάξεις T-σε-C, οι οποίες βρίσκονται κοντά στην περιοχή, η οποία καλύπτεται από το RISC σύμπλεγμα (RNA-induced silencing complex) συμβάλλουν στην ακριβή αναγνώριση των MREs (miRNA recognition elements) [31]. Παρά τον ακριβή εντοπισμό των διασταυρωμένων περιοχών (crosslinked regions), αυτές οι μέθοδοι δεν μπορούν να αποκαλύψουν άμεσα το ακριβές miRNA το οποίο συμμετέχει στην αλληλεπίδραση, και πρέπει να αναγνωριστεί μέσω της βιοπληροφορικής, χρησιμοποιώντας αλγορίθμους ευθυγράμμισης [18]. Μία πιο πρόσφατη παραλλαγή της PAR-CLIP μεθοδολογίας [33] και του CLASH πρωτοκόλλου [1], ενσωματώνει ένα επιπλέον στάδιο σύνδεσης, συγκολλώντας το miRNA στην περιοχή πρόσδεσης του mRNA. Τα παραγόμενα χημικά θραύσματα miRNA: mRNA ακολουθώντας αλληλουχίζονται και διαχωρίζονται με τη βοήθεια της βιοπληροφορικής για την ταυτόχρονη ταυτοποίηση: των στοχευόμενων mRNAs, των θέσεων πρόσδεσης και των αλληλεπιδρώντων miRNAs. Τα CLIP-Seq / CLASH πειράματα μπορούν να αποκαλύψουν χιλιάδες αλληλεπιδράσεις miRNA:γονιδίου σε κάθε ανάλυση και πλέον έχουν τροποποιηθεί σημαντικά το εύρος και την κλίμακα των σχετικών ερευνητικών έργων. [16]

Συνοπτικά στον παρακάτω πίνακα (Σχήμα 2.1) παρουσιάζονται οι πειραματικές τεχνικές, συνοδευόμενες με σχόλια.

Table 1. Index of experimental techniques utilized for the identification of miRNA:gene interactions

Method	Throughput	Intended use
Reporter Genes (7)	Low	Validation of miRNA:UTR (or binding region) interaction
Northern Blotting (7)	Low	Relative effect of miRNA on mRNA levels
qPCR (7)	Low	Quantification of miRNA effect on mRNA levels
Western Blot (7)	Low	Relative assessment of miRNA effect on protein concentration
ELISA (7)	Low	Quantification of miRNA effect on protein concentration
5' RLM-RACE (7)	Low	Identification of cleaved mRNA targets
Microarrays (7)	High	High-throughput assessment of miRNA effect on mRNA expression
RNA-Seq (7)	High	High-throughput assessment of miRNA effect on mRNA expression
Quantitative Proteomics (e.g. pSILAC (9) AGO-IP)	High	High-throughput assessment of miRNA effects on protein concentration
HITS-CLIP (10)	High	Identification of enriched transcripts (miRNAs and mRNAs) in AGO immunoprecipitates
PAR-CLIP (11)	High	Sequencing of AGO binding regions on targeted transcripts
CLASH (14) / PAR-CLIP + Ligation (13)	High	Sequencing of AGO binding regions on targeted transcripts. Production of chimeric miRNA:mRNA reads for the identification of interacting pairs.
Biotin miRNA tagging (7)	High/Low	Pull-down of biotin-tagged miRNAs and estimation of bound transcript content using qPCR (Low yield), microarrays (High-throughput) and RNA-Seq (High-throughput)
IMPACT-Seq (25)	High	Pull-down of biotin-tagged miRNAs, identification of interacting pairs and binding regions.
PARE / Degradome-Seq (22)	High	High-throughput identification of cleaved mRNA targets
3Life (24)	High	High-throughput reporter gene assay
miTRAP (23)	High	miRNA trapping by RNA baiting

FIGURE 2.1: Πειραματικές τεχνικές σχετικά με την αναγνώριση των αλληλεπιδράσεων και των περιοχών στόχευσης. [16]

2.1.2 Υπολογιστικές μέθοδοι για την αναγνώριση των miRNA στόχων (Computational)

Παρά τη σημαντική αύξηση των πειραματικά επικυρωμένων miTGs (microRNA target genes), η πλειοψηφία των γονιδίων στόχων miRNA παραμένει άγνωστη και τα υπολογιστικά προγράμματα πρόβλεψης στόχων παραμένουν η μόνη πηγή για την ταχεία αναγνώριση ενός υποψήφιου στόχου miRNA. Επομένως, η ανάπτυξη των υπολογιστικών προγραμμάτων πρόβλεψης στόχων, προοδεύει σχεδόν παράλληλα με την κατανόηση της λειτουργίας των miRNAs [29].

Τα πρώτα προγράμματα αναπτύχθηκαν το 2003, αφού έγινε φανερό ότι τα miRNA είναι άφθονα στα κύτταρα. Αρκετές ομάδες (Doench and Sharp, 2004, Kiriakidou et al., 2004) έχουν δείξει πειραματικά, ότι τα νουκλεοτίδια κοντά στο 5'UTR άκρο του miRNA είναι τα πιο κρίσιμα για την αναγνώριση και πρόσδεση σε μια αλληλουχία στόχο. Επιπλέον, μια στατιστική ανάλυση από τους Lewis et al. (2005) αποκάλυψε ότι τα μοτίβα στο 3'UTR άκρο, των γονιδίων που κωδικοποιούν πρωτεΐνες, τα οποία αντιστοιχούν στα νουκλεοτίδια 2-7 του miRNA διατηρούνται επί το πλείστον σε διάφορα είδη. Αυτά τα έξι νουκλεοτίδια έχουν υποδειχθεί ως η αλληλουχία «σπόρου» (seed) του miRNA (Σχήμα 2.2). Ωστόσο, αργότερα οι Krek et al. (2005) χρησιμοποίησαν επτά νουκλεοτίδια ξεκινώντας από τη θέση 1 ή 2 ενός miRNA για να εντοπίσουν πιθανούς στόχους στο 3'UTR άκρο. [29]

Τα τελευταία χρόνια έχουν δημοσιευθεί αρκετά προγράμματα πρόβλεψης στόχων miRNA. Το κύριο χαρακτηριστικό πρόβλεψης που χρησιμοποιείται στα περισσότερα από αυτά τα προγράμματα είναι η ευθυγράμμιση της seed περιοχής του miRNA με το 3'UTR άκρο των υποψήφιων γονιδίων-στόχων. Η εξειδίκευση

αυτών των προγραμμάτων συνήθως αυξάνεται λαμβάνοντας υπόψη και αξιοποιώντας την “εξελικτική διατήρηση των θέσεων πρόσδεσης” ή τη χρήση επιπρόσθετων χαρακτηριστικών όπως : τη “δομική προσβασιμότητα” (structural accessibility) (Kertesz et al., 2007; Long et al., 2007), και τη “σύνθεση νουκλεοτιδίων” (nucleotide composition) (Grimson et al., 2007) ή τη θέση των περιοχών πρόσδεσης εντός της 3’UTR περιοχής (Baek et al., 2008 · Gaidatzis et al., 2007 · Grimson et al., 2007). [29]

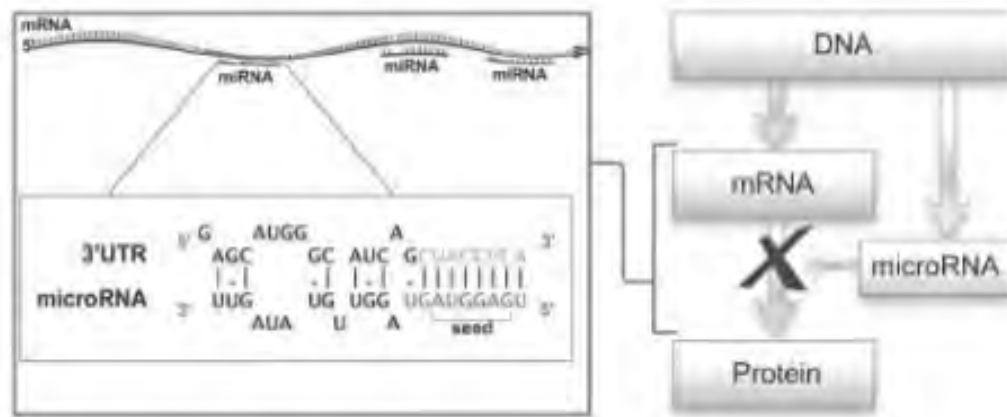


FIGURE 2.2: Η αλληλουχία seed του microRNA, η οποία κατά το πλείστον διατηρείται και είναι αυτή η οποία συμμετέχει στην αλληλεπίδραση με τον στόχο. [29]

2.2 Αλγόριθμοι Πρόβλεψης στόχων microRNA

Στη συνέχεια παραθέτουμε δύο υπολογιστικά προγράμματα - αλγόριθμους, με σκοπό να κατανοήσουμε καλύτερα τα χαρακτηριστικά τους σημεία.

2.2.1 DIANA microT-CDS

Το υπολογιστικό πρόγραμμα DIANA-microT-CDS <http://www.microrna.gr/microT-CDS> έχει αναπτυχθεί στο εργαστήριο DIANA TOOLS και βρίσκεται στην 5η έκδοσή του. Ο αλγόριθμος του microT-CDS είναι ειδικά εκπαιδευμένος στο να ανιχνεύει στόχους σε ένα θετικό και αρνητικό σύνολο MREs, τα οποία βρίσκονται τόσο στη 3'UTR, όσο και στη CDS περιοχή. Τα πειραματικά σύνολα δεδομένων τα οποία χρησιμοποιούνται για την εκπαίδευση του αλγόριθμου προέρχονται από το πείραμα “ενισχυμένης με φωτοβόλο-ριβονουκλεοζίτη διασταυρούμενης σύνδεσης και ανοσοκατακρήμνισης” (PAR-CLIP), το οποίο έγινε από τους Hafner et al., χρησιμοποιώντας κύτταρα ανθρώπινου εμβρυϊκού νεφρού 293 (HEK293) [23]. Η ταυτοποίηση των στόχων σε αυτές τις περιοχές (CDS, 3'UTR) εκτελείται ανεξάρτητα, χρησιμοποιώντας ξεχωριστά μοντέλα πρόβλεψης. [17]

Συνοπτικά, το πρόγραμμα DIANA-microT-CDS λαμβάνει υπόψην του χαρακτηριστικά όπως: το είδος της πρόσδεσης, το ποσοστό διατήρησης των MREs στις CDS και 3'UTR περιοχές, σε 30 και 16 είδη αντίστοιχα, την προσβασιμότητα των θέσεων στόχων, η οποία υπολογίζεται με βάση τον αλγόριθμο Sfold [6], τη σταθερότητα της υβριδοποίησης, και άλλα χαρακτηριστικά γνωρίσματα. Τα αποτελέσματα για τις 3'UTR και CDS περιοχές που παρέχει ο microT-CDS συνδυάζονται, χρησιμοποιώντας ένα γενικό γραμμικό μοντέλο, ειδικά εκπαιδευμένο σε σύνολα δεδομένων μικροσυστοιχιών (microarray datasets) για την αναγνώριση της συνδυαστικής δράσης των πολλαπλών θέσεων πρόσδεσης των miRNAs. Το πρόγραμμα DIANA-microT-CDS έχει σημειώσει μία σημαντική αύξηση στο μέτρο της ευαισθησίας (sensitivity) σε σύγκριση με την προηγούμενη του έκδοση (65% έναντι 52%), η οποία εκτιμάται με βάση πειραματικά δεδομένα πρωτεϊνωμάτωσης (proteomics) και έχει αναδείξει την υψηλότερη ευαισθησία σε οποιαδήποτε επίπεδο ακρίβειας, σε σύγκριση με άλλες state of the art υλοποιήσεις [24]. Η απόδοση του αλγόριθμου έχει ελεγχθεί περαιτέρω, με βάση την αλληλούχιση υψηλής απόδοσης του RNA (high throughput RNA sequencing) που απομονώθηκε με “ανοσοκαταβύθιση διασταυρωτής σύνδεσης” (HITS-CLIP) dataset of Chi et al. [34]. [17]

Στην παρούσα εργασία, η εφαρμογή η οποία έχει υλοποιηθεί και την οποία θα παρουσιάσουμε σύντομα, στηρίζεται στο υπολογιστικό πρόγραμμα DIANA-microT-CDS, αφού έχει λάβει τα αποτελέσματα από τις εκτελέσεις του, με σκοπό να τα χρησιμοποιήσει ως δεδομένα για την παραγωγή σχετικών αποτελεσμάτων. Στο παρακάτω σχήμα (Σχήμα 2.3) παρατηρούμε τη ροή την οποία ακολουθεί ο αλγόριθμος microT-CDS μαζί με τη σχετική περιγραφή.

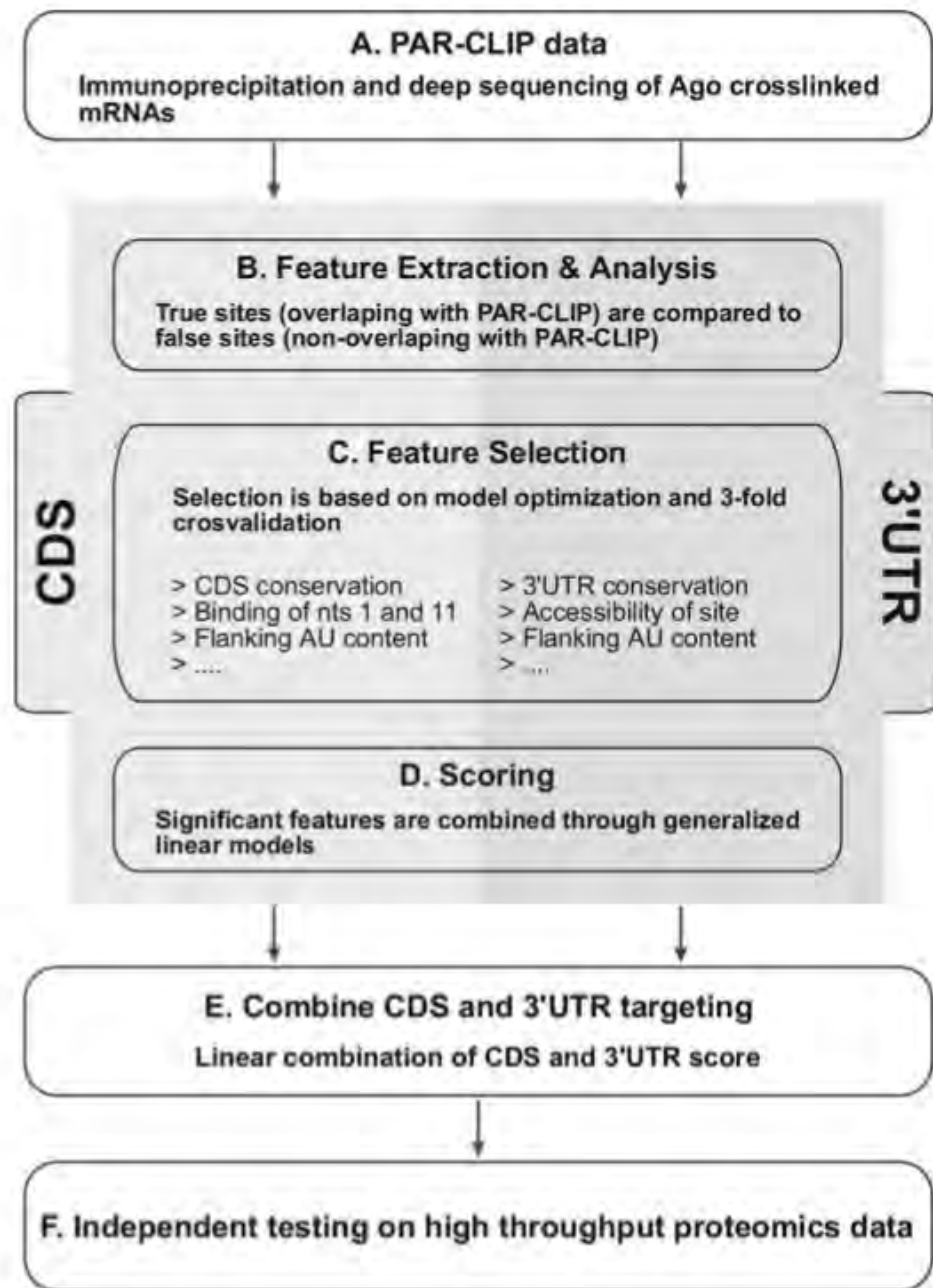


FIGURE 2.3: Ροή την οποία ακολουθεί το υπολογιστικό πρόγραμμα DIANA-microT-CDS. [24]

Περιγραφή Ροής Αλγορίθμου [24]:

Τα MREs τα οποία καθορίζονται από τα δεδομένα PAR-CLIP, χωρίζονται σε δύο κατηγορίες ανάλογα με την περιοχή γονιδιώματος στην οποία βρίσκονται (A). Για αυτά τα δύο σύνολα δεδομένων, εξάγονται διάφορα χαρακτηριστικά και επιλέγονται τα πιο χρήσιμα από αυτά, συγκρίνοντας τα “true MREs” με τα “false MREs” (B). Η επιλογή πραγματοποιείται μέσω “τριπλής διασταυρούμενης

επικύρωσης” (three-fold cross-validation) (C). Για κάθε αναγνωρισμένο miRNA MRE, τα επιλεγμένα χαρακτηριστικά (ανάλογα με την περιοχή γονιδίου που βρίσκεται) συνδυάζονται σε ένα “MRE score” μέσω γενικευμένων γραμμικών μοντέλων (generalized linear models) (D). Για κάθε γονίδιο, η βαθμολογία CDS και η βαθμολογία 3’UTR ορίζεται αθροίζοντας τις βαθμολογίες των MREs που βρίσκονται στις περιοχές CDS και 3’UTR, αντίστοιχα. Αυτές οι δύο βαθμολογίες συνδυάζονται γραμμικά σε μια τελική βαθμολογία (E). Για να ελεγχθεί η συνολική επίδοση αυτής της βαθμολογικής προσέγγισης, εκτελείται μια ανεξάρτητη δοκιμή στα υψηλής απόδοσης δεδομένα πρωτεϊνωματικής των Selbach et al. [25] (F)

Στο ακόλουθο σχήμα (Σχήμα 2.4) παρατηρούμε τις παραμέτρους, οι οποίες παίζουν σημαντικό ρόλο στη λειτουργία του αλγορίθμου microT-CDS .

Feature	Range	Description
miTG score	$0 < \text{miTG score} < 1$	Combined score that measures the potency of each miRNA-gene interaction. The greater the score and close to 1, the greater the confidence
Binding type	6mer, 7mer, 8mer, 9mer and miRNA bugle	The matching sites between the miRNA and the mRNA
Score	$0 < \text{Score} < 1$	The site contribution score in the miTG score
Conservation	≥ 0	Number of species where the predicted interaction is conserved
Signal-to-noise ratio	> 0	This score calculates the capacity of identification of the miTG score of each interaction without background noise
Precision	from 0 to 1	The score indicates the false-positive rate in a miTG interaction

FIGURE 2.4: Παράμετροι τις οποίες λαμβάνει υπόψη του ο αλγόριθμος DIANA microT-CDS.

2.2.2 Targetscan

Το πρόγραμμα TargetScan <http://www.targetscan.org> είναι μία από τις πρώτες διαθέσιμες εφαρμογές, οι οποίες παρουσιάζουν υψηλή ευαισθησία και ακρίβεια. Ο αλγόριθμος TargetScan χρησιμοποιεί ένα “context score¹”, το οποίο βασίζεται σε χαρακτηριστικά όπως, ο τύπος της θέσης-στόχου (target site type), η τοπική συνιστώσα AU (local AU content), η θέση της περιοχής του στόχου (target site position) και το συμπληρωματικό ζευγάρισμα στο 3' άκρο. Η έκτη έκδοση του αλγορίθμου TargetScan σχεδιάστηκε και υλοποιήθηκε ειδικά, προκειμένου να αντιμετωπιστεί το φαινόμενο των miRNAs μειωμένης αποδοτικότητας (low proficiency). Αυτά τα miRNAs εμφανίζουν μειωμένη ικανότητα καταστολής, ακόμη και σε στόχους με 7 ή 8 νουκλεοτίδια στις περιοχές όπου προσδένεται η seed αλληλουχία. Αυτό το συγκεκριμένο υποσύνολο των miRNAs, είναι ασυνήθιστα πλούσιο σε AU συνιστώσες στην seed περιοχή, με συνέπεια να μειώνεται η σταθερότητα των αλληλεπιδράσεων, καθώς και να παρουσιάζονται πολλαπλές θέσεις ως στόχοι.

Το πρόγραμμα TargetScan 6 χρησιμοποιεί τα παραπάνω χαρακτηριστικά, προκειμένου να αυξήσει περαιτέρω την ακρίβεια του αλγορίθμου πρόβλεψής του. Η εκπαίδευση των μοντέλων πραγματοποιήθηκε χρησιμοποιώντας πολλαπλή γραμμική παλινδρόμηση (multiple linear regression) έναντι δεδομένων microarray. Κάθε θέση-στόχος στη 3'UTR περιοχή θεωρείται ότι ενεργεί ανεξάρτητα των υπολοίπων και αυτή η υπόθεση χρησιμοποιείται για τον υπολογισμό του συνολικού “context score”. Στα θηλαστικά, οι τελικές προβλέψεις ταξινομούνται με βάση τα υπολογισμένα “context scores”, ενώ παρέχεται ακόμα και η δυνατότητα της ταξινόμησης των αποτελεσμάτων με βάση τη διατήρηση της θέσης-στόχου. Ο αλγόριθμος έχει δοκιμαστεί έναντι δεδομένων από microarray, τα οποία δεν χρησιμοποιήθηκαν κατά τη δημιουργία του μοντέλου. Το TargetScan, όπως και οι περισσότεροι από τους διαθέσιμους αλγόριθμους πρόβλεψης στόχου, έχουν εκπαιδευτεί για την ανίχνευση στόχων στην 3'UTR περιοχή των μεταγραφών, τα οποία κωδικοποιούν κάποια πρωτεΐνη, ωστόσο μελέτες έχουν δείξει ότι στόχοι βρίσκονται και στις CDS περιοχές. Τέλος, το πρόγραμμα TargetScan6 βασίζεται στην έκδοση miR-Base17 για την ονοματολογία των miRNAs και των σχετικών miRNA οικογενειών [17].

¹Ως context νοείται η θέση του στόχου στη 3'UTR περιοχή, καθώς και τα νουκλεοτίδια εκατέρωθεν αυτής.

Chapter 3

Δεδομένα

3.1 Πηγές Δεδομένων

Σε αυτήν την ενότητα θα πούμε λίγα λόγια σχετικά με τις πηγές από τις οποίες λάβαμε τα δεδομένα τα οποία χρησιμοποιούμε στην εφαρμογή μας. Οι δύο πρώτες πηγές μας είναι, τα υπολογιστικά προγράμματα “DIANA-microT-CDS” και “TargetScan”, στα οποία έχουμε ήδη αναφερθεί στην προηγούμενη ενότητα και ως εκ τούτου δε θα αφιερώσουμε πολύ χρόνο σε αυτό το σημείο.

3.1.1 DIANA-microT-CDS

Αποτελεί έναν web server για προβλέψεις “in silico” αλληλεπιδράσεων μεταξύ microRNA και γονιδίων-στόχων. Εμπεριέχει έναν από τους πρώτους εφαρμοσμένους αλγορίθμους για την εύρεση των αλληλεπιδράσεων των miRNA-γονιδίων, με αυξημένη απόδοση όσον αφορά στην ευαισθησία (sensitivity) και στην ακρίβεια (precision). Ανιχνεύει με ακρίβεια τους “κανονικούς” (canonical) στόχους των microRNAs τόσο στη CDS, όσο και στη 3’UTR περιοχή. Σημαντικά χαρακτηριστικά του microT-CDS είναι η συμπληρωματικότητα στη θέση στόχο, η προσβασιμότητα, η σταθερότητα υβριδοποίησης και η διατηρησιμότητα των CDS και 3’UTR περιοχών στόχων [19]. Από την ιστοσελίδα του [DIANA-microT-CDS](#), ο χρήστης έχει τη δυνατότητα να διεξάγει αναζητήσεις με βάση ένα microRNA, ένα γονίδιο, μία KEGG περιγραφή, ή συνδυασμό των προηγούμενων και να αποθηκεύσει τα αποτελέσματα της αναζήτησής του σε ένα “comma separated” αρχείο κειμένου. Τέλος, η νέα έκδοση του Web Server υποστηρίζει αυτόματα pipelines για την ενσωμάτωση δεδομένων υψηλής απόδοσης (high-throughput data) στις προηγμένες microRNA αναλύσεις.

3.1.2 Targetscan

Είναι ένας web server, ο οποίος προβλέπει τους στόχους των miRNAs με βάση την παρουσία διατηρημένων θέσεων 8mer, 7mer και 6mer, οι οποίες ταιριάζουν με την seed περιοχή του κάθε microRNA. Ως επιλογή, παρέχονται επίσης προβλέψεις μόνο για περιοχές οι οποίες είναι ανεπαρκώς διατηρημένες. Επίσης αναγνωρίζονται οι περιοχές με αναντιστοιχίες στην seed περιοχή, οι οποίες αντισταθμίζονται

με τη διατηρημένη υβριδοποίηση στο 3' άκρο και στις κεντρικές περιοχές. Από την ιστοσελίδα του **TargetScan** έχουμε τη δυνατότητα να κατεβάσουμε διάφορα σύνολα δεδομένων, τα οποία αφορούν στα γονίδια και τα microRNAs, στις προκαθορισμένες προβλέψεις (διατηρημένες περιοχές των διατηρημένων miRNA οικογενειών), καθώς και όλες τις προβλέψεις για τα αντιπροσωπευτικά μετάγραφα. Ακόμα, παρέχονται scripts, τα οποία αφορούν για παράδειγμα, στον προσδιορισμό των διατηρημένων microRNAs στόχων και των μη διατηρημένων περιοχών ή στον υπολογισμό των context++ scores για ένα σύνολο προβλεπόμενων microRNA περιοχών, χρησιμοποιώντας ένα “custom” σύνολο δεδομένων. [35]

3.1.3 miRBase

Η βάση δεδομένων της miRBase δημιουργήθηκε το 2002. Είναι το κυριότερο από τα διαδικτυακά αποθετήρια (online repositories) που υπάρχουν, το οποίο εμπεριέχει αλληλουχίες microRNAs και σχετικές πληροφορίες για αυτά (annotations). Παρέχει τις δημοσιευμένες microRNA ακολουθίες τόσο για αναζήτηση μέσω λέξεων κλειδιών, με τη χρήση ενός web interface, όσο και για αποθήκευση στον προσωπικό υπολογιστή μέσω FTP. Ακόμα, η miRBase χρησιμοποιεί τα deep sequence data sets, τα οποία έχει συλλέξει, για να αναθέσει “επίπεδα εμπιστοσύνης” (levels of confidence) ή scores στα microRNAs, που υπάρχουν στη βάση δεδομένων της, ανάλογα με το κατά πόσο αυτά τα microRNAs θεωρούνται σωστά. Αυτό συμβαίνει λόγω του ότι το κόστος του sequencing έχει μειωθεί, ενώ το “βάθος” του (depth of sequencing) έχει αυξηθεί και ως εκ τούτου οι ερευνητές ασχολούνται με microRNAs, τα οποία εκφράζονται όλο και σε πιο χαμηλό επίπεδο ή σε πιο συγκεκριμένους ιστούς ή κυτταρικούς τύπους. Κατά συνέπεια, καθίσταται όλο και πιο δύσκολη η διάκριση των πραγματικών microRNAs από θραύσματα άλλων μεταγράφων [2].

Οι κύριοι στόχοι της miRBase είναι [3] :

- Να προσδιορίσει ένα συνεπές σύστημα ονοματολογίας, με το οποίο θα ονομάζονται τα νεοανακαλυπτόμενα microRNAs.
- Να λειτουργήσει ως κεντρικό αποθετήριο για όλες τις δημοσιευμένες ακολουθίες microRNA και να διευκολύνει την ηλεκτρονική αναζήτηση και τη μαζική λήψη όλων των δεδομένων microRNA.
- Να παρέχει πληροφορίες, σχολιασμό (annotations) για τις αλληλουχίες των microRNAs (π.χ. αναφορές, λειτουργικά δεδομένα, χαρτογραφήσεις γονιδιωμάτων), οι οποίες θα είναι κατάλληλες προς ανάγνωση τόσο από τον άνθρωπο, όσο και για αναλύσεις από τον υπολογιστή.
- Να διασυνδέσει και να ομαδοποιήσει τις προβλέψεις και τις επικυρώσεις των microRNA στόχων.

Στο ακόλουθο σχήμα (Σχήμα 3.1) παρατηρούμε την αύξηση των microRNAs, τα οποία έχουν καταχωρηθεί στην miRBase ανά τα χρόνια.

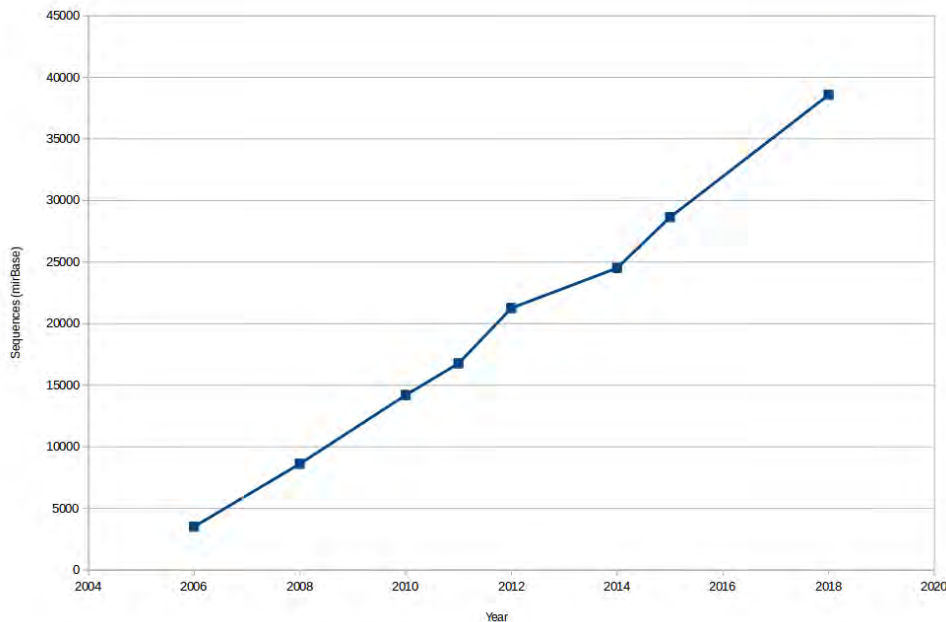


FIGURE 3.1: Σχεδιάστηκε για τους σκοπούς της παρούσας διπλωματικής. Αύξηση των microRNAs ακολουθιών στο αρχείο της miRBase ανά τα χρόνια. [3]

3.1.4 NCBI

Το ακρωνύμιο NCBI σημαίνει National Center for Biotechnology Information (Εθνικό Κέντρο Βιοτεχνολογικών Πληροφοριών). Το Εθνικό Κέντρο Βιοτεχνολογικών Πληροφοριών δημιουργήθηκε το 1988 και είναι μέρος του United States National Library of Medicine (NLM), το οποίο με τη σειρά του είναι μέρος του National Institutes of Health (NIH). Η NCBI εν γένει προωθεί την επιστήμη και την υγεία παρέχοντας πρόσβαση σε πληροφορίες βιοϊατρικής και γονιδιωματικής. Πιο συγκεκριμένα, φιλοξενεί μια σειρά βάσεων δεδομένων σχετικών με τη βιοτεχνολογία και τη βιοϊατρική και αποτελεί σημαντική πηγή εργαλείων και υπηρεσιών βιοπληροφορικής. Σημαντικές βάσεις δεδομένων είναι η “GenBank” για ακολουθίες DNA και η “PubMed”, μια βιβλιογραφική βάση δεδομένων για τη βιοϊατρική βιβλιογραφία. Ακόμα υποστηρίζει τις βάσεις δεδομένων OMIM, MMDB, UniGene, CGAP. Όλες οι βάσεις δεδομένων είναι διαθέσιμες στο διαδίκτυο μέσω της μηχανής αναζήτησης Entrez.

Η Entrez αποτελεί το σύστημα αναζήτησης και ανάκτησης της NCBI και παρέχει στους χρήστες πρόσβαση σε αλληλουχίες, mappings και δομικά δεδομένα (structural data). Επίσης, παρέχει γραφικές απεικονίσεις των αλληλουχιών και των χρωμοσωμάτων. Ένα αξιοσημείωτο χαρακτηριστικό της Entrez είναι ότι μπορεί να

ανακτά σχετικές αλληλουχίες, δομές και αναφορές. Η βιβλιογραφία είναι διαθέσιμη μέσω της PubMed όπως προαναφέραμε, η οποία παρέχει πρόσβαση σε πάνω απο έντεκα εκατομμύρια αναφορές στη MEDLINE, ενώ ακόμα περιέχει και συνδέσμους σε πλήρεις άρθρα στους ιστότοπους των εκδοτών. Ένα σημαντικό εργαλείο που παρέχει η NCBI είναι το BLAST, το οποίο χρησιμοποιείται για τον εντοπισμό παρόμοιων ακολουθιών, και έχει καθοριστική σημασία για τον εύρεση γονιδίων και γενετικών χαρακτηριστικών. Άλλα εργαλεία είναι τα : Open Reading Frame Finder (ORF Finder), Electronic PCR , Sequin και BankIt [28] [42].

3.1.5 Ensembl

Η Ensembl αναπτύχθηκε το 1999, πριν από τη δημοσίευση της πρώτης έκδοσης του “human genome project”. Είναι ένας “genome browser” ο οποίος εν γένει περιέχει πληροφορίες για τα γονιδιώματα των σπονδυλωτών οργανισμών (Σχήμα 3.3). Υποστηρίζει την έρευνα σε πολλά πεδία, όπως αυτά της συγκριτικής γονιδιωματικής, της εξέλιξης, της μεταβολής των αλληλουχιών και της μεταγραφικής ρύθμισης. Για παράδειγμα, υπολογίζει πολλαπλές ευθυγραμμίσεις, κάνει προβλέψεις για τη ρυθμιστική λειτουργία και συλλέγει δεδομένα τα οποία αφορούν σε ασθένειες. Ο genome browser της Ensembl παρέχει πρόσβαση σε οργανωμένες πληροφορίες, οι οποίες έχουν προκύψει απο την ανάλυση βιολογικών δεδομένων (Σχήμα 3.2) και συνήθως ανανεώνεται κάθε δύο με τρεις μήνες, μαζί με τις σχετικές βάσεις δεδομένων [11] [8].

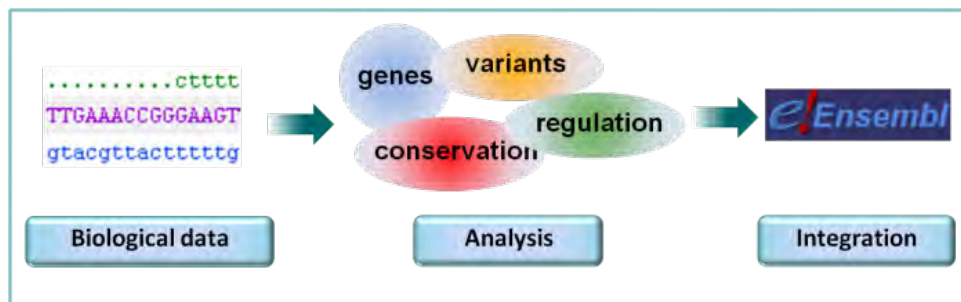


FIGURE 3.2: Η Ensembl παρέχει πρόσβαση σε οργανωμένες πληροφορίες από την ανάλυση βιολογικών δεδομένων.



FIGURE 3.3: Η Ensembl περιέχει κυρίως πληροφορίες σχετικά με τους σπονδυλωτούς οργανισμούς, συμπεριλαμβανομένων των ανθρώπων, ποντικών, αρουραίων, zebrafish και του πάντα. [12]

Για κάθε είδος οργανισμού, η Ensembl επισυνάπτει ένα μεγάλο πλήθος πληροφοριών και δεδομένων (annotations), στο γονιδίωμα του. Ένας από αυτούς τους τύπους δεδομένων είναι τα “μοντέλα γονιδίων” (gene models). Για την πρόβλεψη των γονιδιακών μοντέλων (a.k.a. “gene build”), οι αλληλουχίες από τις βάσεις δεδομένων “INSDC” (International Nucleotide Sequence Database Collaboration), “NCBI RefSeq” και “Uniprot” ευθυγραμμίζονται με το γονιδίωμα. Αφότου ολοκληρωθούν οι ευθυγραμμίσεις, δημιουργούνται κλάσεις μεταγράφων (transcripts), με βάση την επικάλυψη των κωδικών ακολουθιών (coding sequences), έτσι ώστε να προκύψουν τα Ensembl γονίδια. Αυτή η διαδικασία καλείται “The automated gene annotation pipeline”. Για ορισμένα είδη όπως τον άνθρωπο, το ποντίκι, τον αρουραίο, το γουρούνι και το zebrafish, η Ensembl ενσωματώνει επίσης χειροκίνητα μετάγραφα από το HAVANA project. Το συγχωνευμένο αυτό σύνολο γονιδίων, το οποίο δημιουργείται από την αυτόματη και χειροκίνητη πρόσθεση δεδομένων αποκαλείται ως “Gencode geneset” [12].

Στον browser της Ensembl μπορεί να γίνει αναζήτηση με βάση τις γονιδιωματικές συντεταγμένες, τα ονόματα γονιδίων, τους φαινότυπους, τα αναγνωριστικά (IDs) ρυθμιστικών χαρακτηριστικών ή των variant.

Τέλος, η Ensembl προσφέρει ένα σημαντικό σύνολο εργαλείων. Ορισμένα από αυτά είναι τα BLAST, BLAT, BioMart, Assembly Converter, File Chameleon και VEP (Variant Effect Predictor). Το εργαλείο BioMart επιτρέπει την ταχεία δημιουργία πινάκων δεδομένων και σου δίνει τη δυνατότητα να τα εξάγεις σε διάφορα formats, όπως για παράδειγμα σε html, csv, tsv και xls (Excel file) και FASTA format (Σχήμα 3.4). Ακόμα, σου παρέχει τη δυνατότητα να “μεταφράζεις” ένα αναγνωριστικό ID σε ένα άλλο, για παράδειγμα ένα Ensembl gene ID σε ένα NCBI RefSeq ID [10]. Με το εργαλείο VEP έχεις τη δυνατότητα να εισάγεις συντεταγμένες μεταγράφων ή γονιδιωματικές, έτσι ώστε να προσδιορίσεις την επίδραση στα μετάγραφα ή στις πρωτεΐνες, από τις αλλαγές στην αλληλουχία. [11]



FIGURE 3.4: Database: Μπορείς να επιλέξεις τη βάση δεδομένων, η οποία σε ενδιαφέρει. Λόγου χάρη “Ensembl genes database, version 89”. Dataset: Μπορείς να επιλέξεις το σύνολο δεδομένων που θέεις να προσπελάσεις. Λόγου χάρη το “Human data set”. Options: Μπορείς να φιλτράρεις το σύνολο δεδομένων σου, με σκοπό να επιλέξεις το σύνολο γονιδίων που σε ενδιαφέρει. Τα “Attributes” σου επιτρέπουν να επιλέξεις τί θα εμφανιστεί την έξοδο. Ακόμα, επιπρόσθετα σύνολα δεδομένων μπορούν να συνδυαστούν στην αναζήτηση. [10]

3.2 Προεπεξεργασία Δεδομένων

Η εφαρμογή η οποία υλοποιήθηκε στα πλαίσια της διπλωματικής εργασίας απαιτούσε ένα αρκετά μεγάλο όγκο δεδομένων, τα οποία και λάβαμε από τις πηγές τις οποίες περιγράψαμε παραπάνω. Χρειάστηκε να γίνει μια προεπεξεργασία πριν από την χρήση τους, έτσι ώστε να αποφύγουμε κυρίως τον πλεονασμό, αλλά και να αφαιρέσουμε αυτές τις εγγραφές, οι οποίες δε μας ήταν χρήσιμες. Σε αυτήν την ενότητα θα περιγράψουμε με παραδείγματα την επεξεργασία η οποία πραγματοποιήθηκε και θα δείξουμε τη μορφή των δεδομένων μας πριν και μετά. Η κατανόηση και η προεπεξεργασία των δεδομένων αποτέλεσε ένα πολύ σημαντικό κομμάτι για τη συνέχεια και την υλοποίηση της εφαρμογής.

1. Δεδομένα αλληλεπιδράσεων microRNA:γονιδίων

Ας αρχίσουμε λοιπόν, παρατηρώντας τα δεδομένα τα οποία αφορούν στις αλληλεπιδράσεις των microRNAs με τα μετάγραφα. Αυτά τα δεδομένα τα λάβαμε από τη βάση δεδομένων του DIANA-microT-CDS και ήταν πολύ μεγάλα σε μέγεθος. Επιπλέον, η μορφή τους δεν ήταν αρκετά βολική για τις απαιτήσεις της εφαρμογής μας, πράγμα που καθιστούσε αναγκαία την επεξεργασία τους. Χαρακτηριστικά, ένα στιγμιότυπο των δεδομένων αυτών αποτελούνταν από τις παρακάτω τιμές. Λόγω του μεγάλου μήκους των εγγραφών, αδυνατούμε να τα παρουσιάσουμε σε πίνακα και ως εκ τούτου στην παρακάτω λίστα θα δούμε τις χαρακτηριστικές τιμές από μία τυχαία εγγραφή.

1. chr1
2. 161184098
3. 161184103
4. hsa-let-7a-2-3p|ENSG00000158864|ENST00000392179|UTR3
5. 0.462072163848721
6. 1
7. 366|371|0.00822144254437157|6|panTro2, rn4, oryCun2, dasNov2, loxAfr3, echTel1|6mer|UGUGUGUGUGUGUGUAUGUUCA C|UGUACA|ACAUGU|G C|366|372

Τα δεδομένα τα οποία μας ενδιέφεραν και θέλαμε να εξάγουμε από εδώ ήταν: το χρωμόσωμα από το οποίο προέρχεται το γονίδιο-στόχος **1**, οι συντεταγμένες έναρξης **2** και λήξης **3** της αλληλεπίδρασης πάνω στο χρωμόσωμα, το όνομα του microRNA, του γονιδίου καθώς και του μεταγράφου, τα οποία λαμβάνουν μέρος στην αλληλεπίδραση, όπως επίσης και η περιογή αλληλεπίδρασης πάνω στο μετάγραφο **4**. Ακόμα μας ενδιέφερε το miTG score, το οποίο έχει αναθέσει ο αλγόριθμος του DIANA-microT-CDS στην παρούσα αλληλεπίδραση **5**, η strand **6**,

τα είδη οργανισμών, τα οποία παρουσιάζουν το παρόν MRE διατηρημένο, καθώς και ο αριθμός τους (“6”), και τέλος το είδος της πρόσδεσης που λαμβάνει χώρα (“6mer”) 7. Να σημειώσουμε ότι έχουμε λάβει μόνο εκείνες τις εγγραφές, στις οποίες ο αλγόριθμος microT-CDS έχει αναθέσει miTG score από “0.5” και πάνω.

Επιπρόσθετα, σε ορισμένες εγγραφές των δεδομένων μας υπήρχε μία ειδική περίπτωση, όπου οι συντεταγμένες έναρξης και λήξης είχαν την παρακάτω μορφή.

- Begin: 53468561;53468879
- End: 53468568;53468892

Αυτή η μορφή δείχνει ότι, ενδιάμεσα της περιοχής του mRNA, στην οποία γίνεται η πρόσδεση του microRNA, υπάρχει κάποιο ιντρόνιο. Η πρόσδεση αρχίζει στη θέση 53468561 και σταματάει προσωρινά στη θέση 53468568. Στη συνέχεια παρεμβάλλεται ένα ιντρόνιο και έπειτα η πρόσδεση ξαναξεκινάει από τη θέση 53468879 έως τη 53468892, όπου και τερματίζει. Σχηματικά η παραπάνω περιγραφή (Σχήμα 3.5).

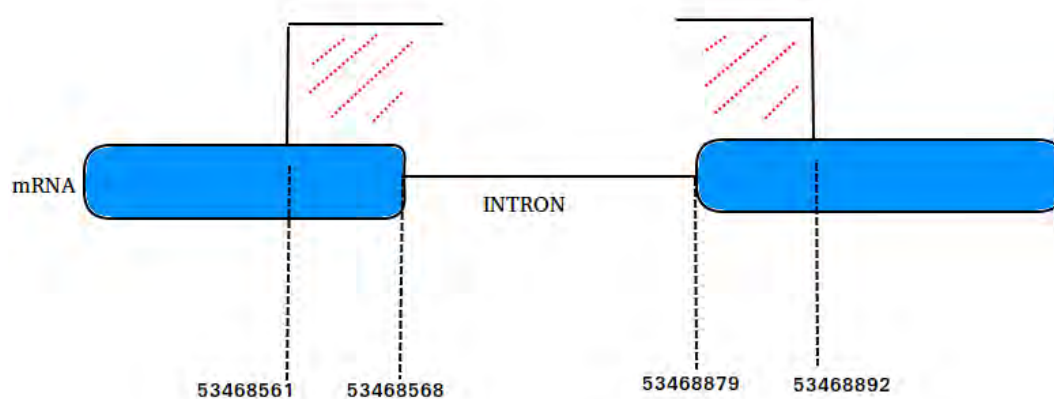


FIGURE 3.5: Αυτή η εικόνα έχει σχεδιαστεί για τους σκοπούς της παρούσας διπλωματικής.

Έτσι, τελικά δημιουργήσαμε ένα “tab-delimited” αρχείο, το οποίο εμπεριέχει τα δεδομένα μας σε κατάλληλη μορφή. Η παραπάνω επεξεργασία έγινε με τη βοήθεια του script “script_human_interactions.py”. Το τελικό αρχείο με όνομα “human_interactions.txt” το οποίο δημιουργήθηκε έχει την παρακάτω μορφή (Πίνακας 3.1).

TABLE 3.1: Επεξεργασμένα δεδομένα αλληλεπιδράσεων microRNA-γονιδίων στόχων.

chr	Bind Start	Bind Stop	miRNA name	ENS Gene ID	ENS Transcript ID	Bind Site	Score	Strand	# of Species Conserved	Species Conserved	Bind Type
8	81549128	81549154	hsa-let-7a-2-3p	ENSG00000164684	ENST00000327835	UTR3	0.892243274091082	-1	2	panTro2,rheMac2	6mer
12	50035679	50035702	hsa-let-7a-2-3p	ENSG00000161791	ENST00000335154	UTR3	0.774335484601486	-1	2	panTro2,loxAfr3	6mer
12	122627126	122627144	hsa-let-7f-5p	ENSG00000175727	ENST00000319080	UTR3	0.68208938900994	1	3	panTro2,rheMac2,loxAfr3	6mer

TABLE 3.2: Επεξεργασμένα δεδομένα, τα οποία αφορούν στα γονίδια και στα μετάγραφα.

chr	Type	Start	Stop	Strand	Gene ID	Gene Name	Transcript ID	Transcript Name
1	gene	69091	70008	+	ENSG00000186092	OR4F5		
1	transcript	69091	70008	+	ENSG00000186092	OR4F5	ENST00000335137	OR4F5-001
2	gene	860260	871173	+	ENSG00000187634	SAMD11		
2	transcript	860260	871173	+	ENSG00000187634	SAMD11	ENST00000420190	SAMD11-011
2	transcript	860260	871173	+	ENSG00000187634	SAMD11	ENST00000437963	SAMD11-003

3. Annotations των υποπεριοχών των μεταγράφων

Τα δεδομένα τα οποία αφορούν στα annotations των υποπεριοχών των μεταγράφων (3'UTR, 5'UTR, CDS) τα έχουμε λάβει από την [Ensembl-biomart \(release 75 – february 2014\)](#). Εκεί, επιλέξαμε την βάση δεδομένων: “Ensembl Genes 75” και το Dataset: “Homo Sapiens Genes (GRCh37. P13)”. Στη συνέχεια επιλέξαμε “attributes” και “structures”. Από εκεί, επιλέξαμε τα πεδία που θέλαμε να χρησιμοποιήσουμε, τα οποία ήταν τα παρακάτω. Ορισμένα από αυτά, τα κατεβάσαμε για λόγους testing και δεν τα χρησιμοποιούμε αργότερα στην εφαρμογή μας.

- Ensembl Gene ID
- Ensembl Transcript ID
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Transcript Start (bp)
- Transcript End (bp)
- Strand
- Associated Gene Name
- 5' UTR Start
- 5' UTR End
- 3' UTR Start
- 3' UTR End
- Exon Chr Start (bp)
- Exon Chr End (bp)
- Genomic coding start

- Genomic coding end
- gene biotype

Επεξεργαστήκαμε κατάλληλα τα δεδομένα μας, έτσι ώστε να ικανοποιούν τις ανάγκες της παρούσας εφαρμογής. Πιο συγκεκριμένα, αφαιρέσαμε όλες τις εγγραφές, οι οποίες είχαν ως χρωμόσωμα κάποιο “patch” (π.χ. HG506_HG507_HG1000_PATCH). Έπειτα, κρατήσαμε μόνο τις εγγραφές οι οποίες είχαν την τιμή “protein coding” στη στήλη “gene biotype” και στη συνέχεια διαγράψαμε αυτή τη στήλη, αφού δε μας ήταν πλέον χρήσιμη. Τέλος, αφαιρέσαμε όσες εγγραφές δεν περιείχαν πληροφορία, σε μία τουλάχιστον από τις παρακάτω στήλες:

- 5’ UTR Start
- 5’ UTR End
- 3’ UTR Start
- 3’ UTR End
- Genomic coding start
- Genomic coding end

Το script το οποίο χρησιμοποιήσαμε για να εκτελέσουμε τις παραπάνω ενέργειες ήταν το “script_transcr_region_annot.py”. Το τελικό αρχείο το οποίο προέκυψε το ονομάσαμε “biomart_transcr_region_annot.csv” και η μορφή του φαίνεται παρακάτω (Πίνακας 3.3):

TABLE 3.3: Τελική μορφή δεδομένων, τα οποία αφορούν στα annotations των περιοχών των μεταγράφων.

Gene ID	Transcript ID	Strand	5’UTR Start	5’UTR End	3’UTR Start	3’UTR End	Genomic coding start	Genomic coding end
ENSG00000215405	ENST00000427390	-1	20747024	20747114			20746826	20747023
ENSG00000215405	ENST00000427390	-1			20737094	20738840		
ENSG00000230031	ENST00000454856	-1	21071611	21071643			21071201	21071610
ENSG00000166351	ENST00000299443	1			15013888	15013906	15013666	15013887

4. Δεδομένα microRNA

Τα δεδομένα τα οποία αφορούσαν στα microRNAs τα έχουμε λάβει από την βάση δεδομένων της **miRBase version 21** και πιο συγκεκριμένα έχουμε χρησιμοποιήσει το αρχείο “mature.fa.zip”. Η αρχική μορφή των δεδομένων μας ήταν η παρακάτω (Πίνακας 3.4). Εμπεριείχε πληροφορίες για πολλά είδη οργανισμών εκτός του ανθρώπινου, όπως και πληροφορία που δε χρειαζόμασταν, για την παρούσα εφαρμογή.

TABLE 3.4: Τα δεδομένα των microRNAs, πριν την επεξεργασία.

miRNA Name	Accession Number	Species	Stem loop	miRNA Sequence
>cel-let-7-5p	MIMAT0000001	Caenorhabditis elegans	let-7-5p	UGAGGUAGUAGGUUGUAUAGUU
>hsa-let-7d-5p	MIMAT0000065	Homo sapiens	let-7d-5p	AGAGGUAGUAGGUUGCAUAGUU
>hsa-miR-26a-5p	MIMAT0000082	Homo sapiens	miR-26a-5p	UUCAAGUAAUCCAGGAUAGGCU
>dme-miR-9a-3p	MIMAT0020792	Drosophila melanogaster	miR-9a-3p	UAAAGCUAGCUUACCGAAGUUA

Κατά την επεξεργασία των δεδομένων μας, αφαιρέσαμε τη στήλη “Stem loop” και κρατήσαμε μόνο τις εγγραφές οι οποίες αφορούσαν στον ανθρώπινο οργανισμό. Με το πέρας της επεξεργασίας, δημιουργήθηκε το αρχείο “21_final_processed_mature.fa.txt”, του οποίου οι στήλες φαίνονται παρακάτω (Πίνακας 3.5). Για την εν λόγω επεξεργασία χρησιμοποιήθηκε το script “script_mature-fa.py”.

TABLE 3.5: Τα δεδομένα των microRNAs, μετά την επεξεργασία.

miRNA Name	Accession Number	miRNA Sequence
hsa-let-7a-5p	MIMAT0000062	UGAGGUAGUAGGUUGUAUAGUU
hsa-miR-21-5p	MIMAT0000076	UAGCUUAUCAGACUGAUGUUGA
hsa-miR-26a-1-3p	MIMAT0004499	CCUAUUCUUGGUUACUUGCACG
hsa-miR-101-3p	MIMAT0000099	UACAGUACUGUGAUAACUGAA

5. Δεδομένα microRNA families

Τα δεδομένα τα οποία αφορούσαν στα microRNA families τα έχουμε λάβει από την ιστοσελίδα του [Targetscan](#) και πιο συγκεκριμένα έχουμε χρησιμοποιήσει το αρχείο “miR Family - (0.16 MB)”. Η αρχική τους μορφή είναι η παρακάτω (Πίνακας 3.6).

TABLE 3.6: Τα δεδομένα των microRNA families, πριν την επεξεργασία.

miR family	Seed+m8	Species ID	miRNA name	mature Sequence	Family Conservation	Accession Number
let-7/98	GAGGUAG	8364	xtr-let-7f	UGAGGUAGUAGAUUGUAUAGUU	2	MIMAT0003645
let-7-5p/98-5p	GAGGUAG	9606	hsa-let-7b-5p	UGAGGUAGUAGGUUGUGUGUU	2	MIMAT0000063
miR-101-3p.1	ACAGUAC	9606	hsa-miR-101-3p.1	UACAGUACUGUGAUAACUGAA	2	MIMAT0000099
miR-101-3p.2	UACAGUA	9606	hsa-miR-101-3p.2	GUACAGUACUGUGAUAACUGA	2	
miR-101-3p	ACAGUAC	9544	mml-miR-101-3p	UACAGUACUGUGAUAACUGAAG	2	MIMAT0002431

Χρησιμοποιώντας το python script “script_miR_family.py” φιλτράραμε τις εγγραφές μας και κρατήσαμε μόνο αυτές, οι οποίες αφορούν στον ανθρώπινο οργανισμό, δηλαδή αυτές που έχουν Species ID = 9606, όπως φαίνεται παραπάνω, καθώς επίσης διαγράψαμε και τη στήλη με το όνομα “Family Conservation”, εφόσον

δε μας ήταν χρήσιμη. Το νέο αρχείο το οποίο δημιουργήθηκε ονομάζεται “processed_miR_Family_Info.txt.csv”. (Πίνακας 3.7).

TABLE 3.7: Τα δεδομένα των microRNA families, μετά την επεξεργασία.

miR family	Seed+m8	Species ID	miRNA name	mature Sequence	Accession Number
let-7-5p/98-5p	GAGGUAG	9606	hsa-let-7b-5p	UGAGGUAGUAGGUUGUGUGGUU	MIMAT0000063
miR-101-3p.1	ACAGUAC	9606	hsa-miR-101-3p.1	UACAGUACUGUGUAACUGAA	MIMAT0000099
miR-101-3p.2	UACAGUA	9606	hsa-miR-101-3p.2	GUACAGUACUGUGUAACUGA	

Στην πρώτη στήλη του πίνακα παρατηρούμε την “οικογένεια” (family), στην οποία ανήκει το εκάστοτε microRNA, ενώ στη δεύτερη, την αλληλουχία της seed περιοχής του. Στην τρίτη στήλη παρατηρούμε το αναγνωριστικό ID του οργανισμού και στην τέταρτη το όνομά του εκάστοτε microRNA. Τέλος, η πέμπτη στήλη εμπεριέχει την αλληλουχία του microRNA, ενώ η έκτη μας παρέχει το “accession number” του. Παρατηρούμε ότι σε κάποιες εγγραφές το “accession number” είναι κενό. Αυτό συμβαίνει διότι μέσα στα δεδομένα μας υπάρχουν microRNAs της μορφής “hsa-miR-101-3p.1” και “hsa-miR-101-3p.2”, τα οποία καταλήγουν στο ίδιο mature microRNA “hsa-miR-101-3p”. Έτσι, μόνο μία από τις προηγούμενες εγγραφές θα παρουσιάζει το “accession number”.

Αφού τα δεδομένα μας που αφορούν στα microRNAs και στα microRNA families είναι έτοιμα, έχει έρθει η ώρα να τα συγχωνεύσουμε σε ένα αρχείο δεδομένων, το οποίο θα εμπεριέχει όλα τα microRNAs από το αρχείο “21_final_processed_mature.fa.txt” συνδυασμένα με τις οικογένειές τους. Την εν λόγω συγχώνευση την κάνουμε με βάση τα κοινά τους πεδία: “accession number” και “mature Sequence”, χρησιμοποιώντας το script “script_merge_mirna_families”. Έτσι, προκύπτει το νέο αρχείο “merged_mirna_families.txt.csv”, με τα δεδομένα όπως αυτά παρουσιάζονται στον πίνακα (Πίνακας 3.8).

TABLE 3.8: Τα δεδομένα μετά τη συγχώνευση των microRNA και mir families δεδομένων.

miRNA Name	Accession Number	miRNA Sequence	miR family	seed+m8
hsa-miR-20a-3p	MIMAT0004493	ACUGCAUUAUGAGCACUUAAG	miR-20a-3p	CUGCAUU
hsa-let-7a-3p	MIMAT0004481	CUAUACAAUCUACUGUCUUUC	let-7a-3p/let-7b-3p/let-7f-1-3p/98-3p	UAUACAA
hsa-miR-28-5p	MIMAT0000085	AAGGAGCUCACAGUCUAUUGAG	miR-28-5p/708-5p	AGGAGCU
hsa-miR-101-3p	MIMAT0000099	UACAGUACUGUGUAACUGAA	miR-101-3p.1	ACAGUAC

6. Δεδομένα χρωμοσωμάτων

Τέλος, τα δεδομένα τα οποία χρησιμοποιήσαμε για να δημιουργήσουμε τα χρωμοσώματα στην εφαρμογή μας, τα έχουμε λάβει από την **NCBI** και πιο συγκεκριμένα χρησιμοποιήσαμε το αρχείο “ideogram_9606_GCF_000001305.13_850_V1”.

Η αρχική μορφή των δεδομένων μας είναι αυτή, η οποία φαίνεται στον παρακάτω πίνακα (Πίνακας 3.9)

TABLE 3.9: Αρχικά δεδομένα χρωμοσωμάτων.

chromosome	arm	band	iscn start	iscn stop	bp start	bp stop	stain	density
1	p	36.23	344	459	7200000	9200000	gpos	25
2	q	14.2	7036	7334	118800000	122400000	gneg	

Η πρώτη στήλη υποδηλώνει τον αριθμό του χρωμοσώματος, με τιμές από 1 έως 22, X, Y. Η δεύτερη στήλη υποδηλώνει τον βραχίονα του χρωμοσώματος, και μπορεί να πάρει τις τιμές ‘p’ ή ‘q’, όπου με ‘p’ συμβολίζεται ο αριστερός βραχίονας και με ‘q’ ο δεξιός (Σχήμα 3.7).

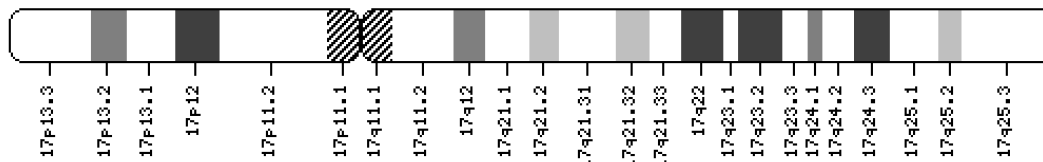


FIGURE 3.7: Στο χρωμόσωμα φαίνονται οι βραχίονες ‘p’, ‘q’, η αρίθμηση και τα χρώματα της κάθε περιοχής.

Η τρίτη στήλη μας δείχνει τον αριθμό της περιοχής του χρωμοσώματος, ο οποίος είναι φθίνων για τον αριστερό βραχίονα και αύξων για τον δεξιό, κοιτάζοντας από αριστερά προς τα δεξιά. Η τέταρτη και η πέμπτη στήλη μας δείχνουν τις συντεταγμένες αρχής και τέλους της κάθε περιοχής του χρωμοσώματος, όπως και οι στήλες έξι και επτά, μόνο που οι τελευταίες ξεκινούν από τη θέση νούμερο 1 (are 1-based). Η όγδοη και η ένατη στήλη θα μας καθοδηγήσουν στο να δώσουμε τους κατάλληλους χρωματισμούς στην κάθε περιοχή (band) του χρωμοσώματος, όπως φαίνεται στην εικόνα (Σχήμα 3.7). Η όγδοη στήλη μπορεί να πάρει τις παρακάτω τιμές : gneg, gpos, acen, gvar, stalk, ενώ η ένατη στήλη παίρνει τις τιμές 25, 50, 75, 100 και μόνο όταν η όγδοη στήλη έχει την τιμή “gpos”. Σε κάθε άλλη περίπτωση, η ένατη στήλη παραμένει κενή.

Μετά απο επεξεργασία των δεδομένων, έχουμε αφαιρέσει τις στήλες “iscn_start” και “iscn_stop”, καθώς θα χρησιμοποιήσουμε τις “bp_start” και “bp_stop” για τις συντεταγμένες της κάθε περιοχής του χρωμοσώματος (Πίνακας 3.10). Χρησιμοποιήσαμε το script “script_chromosome.py” και το νέο αρχείο το οποίο δημιουργήθηκε ονομάστηκε “proc_ideogram_9606_GCF_000001305.13_850_V1.csv”.

TABLE 3.10: Επεξεργασμένα δεδομένα χρωμοσωμάτων.

chromosome	arm	band	bp start	bp stop	stain	density
1	p	36.23	7200000	9200000	gpos	25
2	q	14.2	118800000	122400000	gneg	

Όλα τα scripts που χρησιμοποιήθηκαν βρίσκονται στο github στο παρακάτω [link](#).

3.3 Η Βάση Δεδομένων της Εφαρμογής

Σύμφωνα λοιπόν με τα παραπάνω δεδομένα, έχουμε υλοποιήσει τη βάση δεδομένων της παρούσας εφαρμογής, από την οποία θα αντλούμε από εδώ και πέρα, τις πληροφορίες που πρέπει να παρουσιάσουμε στον χρήστη, σύμφωνα με τα κριτήρια της αναζήτησής του. Το σχήμα της βάσης δεδομένων μας είναι αρκετά απλό, όπως παρουσιάζεται στο παρακάτω σχήμα (Σχήμα 3.8).

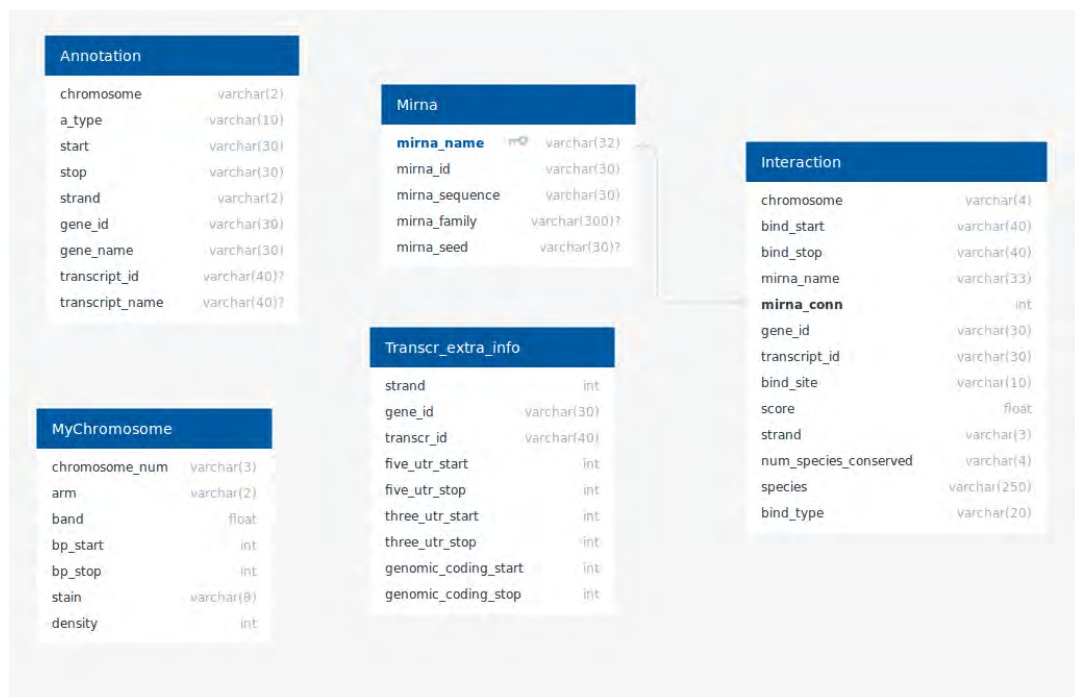


FIGURE 3.8: Το σχήμα της βάσης δεδομένων της εφαρμογής.

Παρατηρούμε ότι αποτελείται από πέντε πίνακες, από τους οποίους οι δύο είναι μεταξύ τους συνδεδεμένοι με ξένο κλειδί (foreign key). Ονομαστικά οι πίνακες είναι οι παρακάτω:

1. MyChromosome
2. Mirna
3. Interaction
4. Annotation
5. Transcr_extra_info

Ο πίνακας 1 εμπεριέχει τις πληροφορίες για τα χρωμοσώματα. Ο πίνακας 2 περιέχει τις πληροφορίες για κάθε microRNA και συνδέεται με τον πίνακα 3, μέσω του πεδίου “mirna_name”, το οποίο αποτελεί το κυρίως κλειδί του 2. Ο πίνακας 3 εμπεριέχει τις πληροφορίες που αφορούν στις συσχετίσεις των microRNAs με τα γονίδια στόχους, ενώ ο πίνακας 4 και 5, κρατούν τις πληροφορίες σχετικά με τα annotations των γονιδίων και των μεταγράφων. Ο μεγαλύτερος πίνακας της εφαρμογής μας είναι ο πίνακας “Interaction” 3, ο οποίος αποτελείται από 24.646.090 εγγραφές.

Chapter 4

Διαδικτυακή Εφαρμογή - Αποτελέσματα

4.1 Σκοπός της εφαρμογής

Σκοπός της παρούσας εφαρμογής είναι να παρουσιάσουμε γραφικά όλα εκείνα τα microRNAs, τα οποία έχουν προβλεφθεί με βάση τον αλγόριθμο DIANA-microT-CDS, ότι στοχεύουν ορισμένα MREs σε ένα mRNA μετάγραφο. Δίνοντας λοιπόν, ο χρήστης ένα μετάγραφο ως είσοδο, θα παρατηρεί στην έξοδο όλα τα microRNAs, τα οποία αλληλεπιδρούν με αυτό, καθώς επίσης και όλες τις πιθανές θέσεις πρόσδεσης τους (MREs) πάνω σε αυτό. Πρόσθετες σχετικές πληροφορίες παρουσιάζονται επιπλέον όπως θα δούμε στην επόμενη ενότητα, η οποία αφορά στην ανάλυση των αποτελεσμάτων.

Αυτή η απεικόνιση λαμβάνει χώρα, με σκοπό ο χρήστης ή ερευνητής να μπορεί να παρατηρήσει άμεσα, ποια και πόσα microRNAs αλληλεπιδρούν με ένα συγκεκριμένο mRNA, καθώς επίσης και την κατανομή αυτών των microRNAs κατά μήκος των υποπεριοχών (3'UTR, CDS, 5'UTR) του mRNA. Μπορεί ακόμα να παρατηρήσει με βάση το miR family, ποια microRNAs από αυτά τα οποία προβάλλονται, έχουν ίδιο seed sequence και πόσο κοντά ή μακριά το ένα από το άλλο, αυτά προσδέονται πάνω στο mRNA. Έπειτα, μπορεί να διεξάγει μία νέα αναζήτηση με βάση ένα microRNA από τα προηγούμενα αποτελέσματα και να δει με ποια άλλα mRNAs μετάγραφα αλληλεπιδρά. Με αυτόν τον τρόπο, μπορεί να βγάλει χρήσιμα συμπεράσματα ή σχέσεις, σχετικά με τα microRNAs και τα seed sequences αυτών, τις περιοχές όπου μπορεί να προσδένει το ίδιο microRNA σε διαφορετικά μετάγραφα, καθώς επίσης και τον τύπο της πρόσδεσης κάθε φορά.

4.2 Παρουσίαση της Web εφαρμογής

Σε αυτή την ενότητα θα παρουσιάσουμε αναλυτικά την διαδικτυακή εφαρμογή, η οποία υλοποιήθηκε στα πλαίσια της διπλωματικής εργασίας, με παραδείγματα και εικόνες. Θα δούμε το γραφικό περιβάλλον της, τον τρόπο με τον οποίο μπορούμε να αλληλεπιδρούμε με αυτή και θα αναλύσουμε τα τελικά αποτελέσματα.

τα οποία παρουσιάζονται στον χρήστη.

Η αρχική σελίδα (home page) της εφαρμογής μας είναι η παρακάτω (Σχήμα 4.1):



FIGURE 4.1: Η αρχική σελίδα της εφαρμογής.

Στο μενού πλοήγησης (navigation bar), το οποίο φαίνεται στην κορυφή έχουμε τρεις επιλογές (Σχήμα 4.2), οι οποίες είναι πάντα εκεί, ανεξάρτητα σε ποια σελίδα της εφαρμογής βρισκόμαστε :

1. Home
2. Statistics
3. Help

ενώ στα δεξιά υπάρχει ένα κρυφό – αναδυόμενο παράθυρο, το οποίο θα εξετάσουμε στη συνέχεια.



FIGURE 4.2: Το μενού πλοήγησης φαίνεται στο κέντρο της κορυφής της σελίδας, ενώ το αναδυόμενο παράθυρο εμφανίζεται πατώντας το κουμπί, που φαίνεται πάνω, δεξιά στη σελίδα.

4.2.1 Μενού Πλοήγησης

Ας δούμε λοιπόν κάθε μία από τις τρεις επιλογές του navigation bar ξεχωριστά.

1. Επιλέγοντας τον πρώτο σύνδεσμο “Home” από το μενού πλοήγησης, ανακατευθυνόμαστε πάντα στην αρχική σελίδα της εφαρμογής (Σχήμα 4.1). Αυτή αποτελεί την κεντρική σελίδα της εφαρμογής μας, από την οποία μπορούμε να ξεκινήσουμε την πλοήγηση. Εδώ έχουμε τη δυνατότητα να δώσουμε τρεις πιθανές εισόδους με σκοπό να εκτελέσουμε μία αναζήτησή:
 - Η πρώτη περίπτωση είναι να εισάγουμε είτε ένα όνομα γονιδίου (ZNF704) ή ένα “ENSEMBL ID”, του τύπου “ENSG00000164684”.
 - Η δεύτερη περίπτωση είναι να εισάγουμε ένα μετάγραφο. Αυτό το εισάγουμε με τη μορφή “ENSEMBL TRANSCRIPT ID”, όπως για παράδειγμα “ENST00000327835”.
 - Τρίτη και τελευταία περίπτωση είναι να εισάγουμε είτε ένα όνομα microRNA (π.χ. “has-let-7a-5p”), είτε ένα MIMAT ID - accession number (π.χ. “MIMAT0000062”).
2. Ο δεύτερος σύνδεσμος (“Statistics”) είναι αρχικά απενεργοποιημένος και ενεργοποιείται όταν στον χρήστη έχουν εμφανιστεί τα τελικά αποτελέσματα της αναζήτησής του. Σε αυτό το στάδιο έχει τη δυνατότητα να επιλέξει την καρτέλα “Statistics”, η οποία θα του εμφανίσει ένα παράθυρο με στατιστικά αποτελέσματα, τα οποία είναι σχετικά με την εκάστοτε αναζήτηση, την οποία πραγματοποιεί (Σχήμα 4.3).

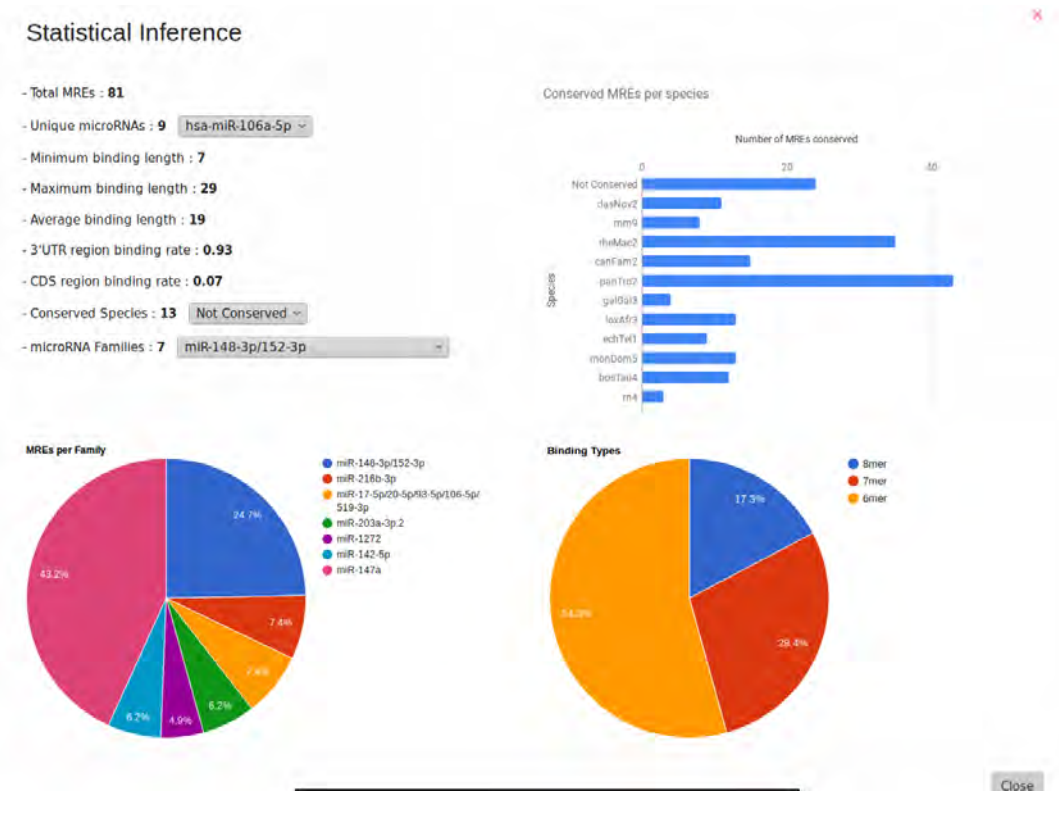


FIGURE 4.3: Στατιστικά στοιχεία που εμφανίζονται στον χρήστη και τα οποία αφορούν την εκάστοτε αναζήτηση, την οποία διεξάγει.

Αναλυτικότερα, σε αυτό το παράθυρο βλέπουμε στατιστικά στοιχεία όπως :

- Τον συνολικό αριθμό των προβλεπόμενων MREs (predicted MREs), στα οποία προσδένονται τα microRNAs.
- Τον αριθμό των microRNAs, τα οποία αλληλεπιδρούν, καθώς και τα ονόματά τους.
- Το ελάχιστο και μέγιστο μήκος πρόσδεσης, όπως επίσης και το μέσο.
- Το ποσοστό των MREs ανά mRNA περιοχή (3'UTR , CDS).
- Τον συνολικό αριθμό των οργανισμών (species), των οποίων τα νουκλεοτίδια πρόσδεσης (MRE sequence nucleotides) της εκάστοτε θέσης στόχου έχουν διατηρηθεί κατά την εξέλιξη, καθώς και τα ονόματά τους.
- Τον αριθμό των microRNA οικογενειών (microRNA families) οι οποίες συμμετέχουν και τα ονόματά τους.

Ακόμα τα στατιστικά αποτελέσματα μας παρέχουν 2 γραφήματα πίτας και ένα ιστόγραμμα. Το πρώτο γράφημα πίτας (κάτω αριστερά) μας δείχνει ποσά predicted MREs ανήκουν σε κάθε microRNA οικογένεια (microRNA family). Το δεύτερο γράφημα (κάτω δεξιά) μας δείχνει το ποσοστό που

καταλαμβάνει ο κάθε τύπος πρόσδεσης, όσον αφορά στα predicted MREs, τα οποία εμφανίζονται. Για παράδειγμα το 13.7% των MREs είναι 8mer, το 28.4% είναι 7mer και το υπόλοιπο 54.3% είναι 6mer. Τέλος, το ιστόγραμμα μας δείχνει για κάθε είδος οργανισμού, σε πόσα predicted MREs είναι διατηρημένο. Με βάση αυτά τα στατιστικά, ο χρήστης μπορεί να βγάλει συμπεράσματα και να κάνει συσχετίσεις μεταξύ των αναζητήσεων του.

3. Η τρίτη και τελευταία επιλογή (“Help”) μας οδηγεί σε μία σελίδα, η οποία μας παρέχει επεξηγηματικές πληροφορίες για τη χρήση της εφαρμογής. Αναλυτικότερα, παρέχει πληροφορίες σχετικά με τα φίλτρα αναζήτησης και τις προκαθορισμένες τιμές τους, τα οποία θα δούμε στην επόμενη υποενότητα, καθώς επίσης και για τα χρώματα, τα οποία χρησιμοποιούμε για την αναπαράσταση των αποτελεσμάτων μας. Περιέχει έναν σύντομο οδηγό για τη ροή και τη χρήση της εφαρμογής και πληροφορίες οι οποίες επεξηγούν τα τελικά αποτελέσματα. Η μορφή αυτής της σελίδας φαίνεται στις παρακάτω εικόνες (Σχήμα 4.4 , 4.5).

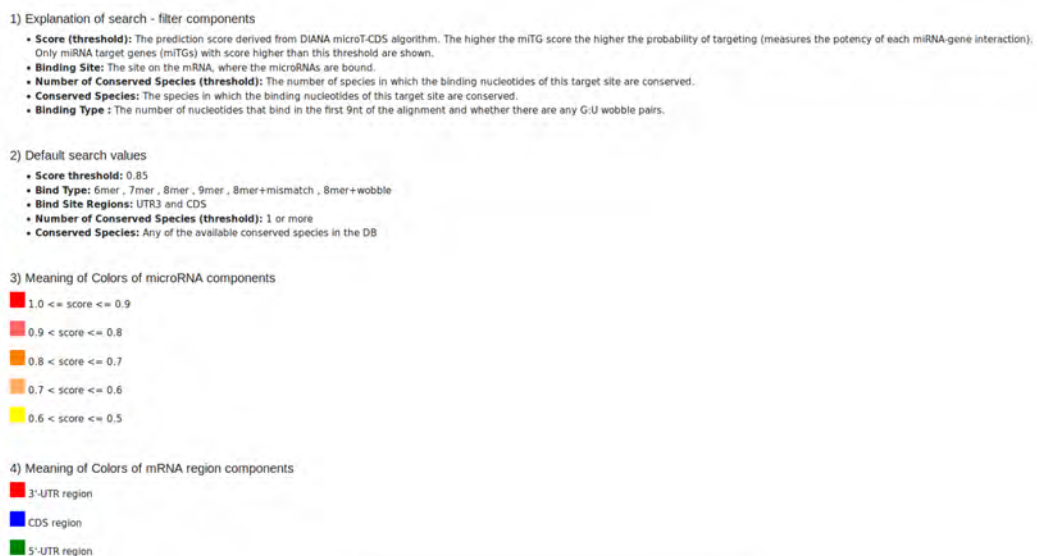


FIGURE 4.4: Στη σελίδα που παρέχει βοηθητικές πληροφορίες μπορούμε να ενημερωθούμε για τη σημασία του κάθε φίλτρου και για τις προκαθορισμένες τιμές του. Ακόμα, βλέπουμε τη σημασία των χρωμάτων, τα οποία αναθέτουμε στα “microRNA components”, καθώς και στις υποπεριοχές του μεταγράφου.

5) Web Application Flow

At the home page we have three possible inputs , and we have to fill one to continue ...

- Gene (Gene Ensembl ID or Gene Name)
- Transcript (Ensembl Transcript ID)
- microRNA (microRNA Name or microRNA ID)

Case 1: Input a Gene and observe the available transcripts that show up. Choose one of them and check the final results of your search.

Case 2: Input a Transcript and direct check the final results of your search.

Case 3: Input a microRNA and observe the available transcripts that show up ,along with extra information. Choose one from the list and check the final results of your search.

6) Results Page - What we observe

- On the chromosome we observe the location of the current gene , with a red vertical line.
- Next, we observe the gene, along with its coordinates and arrows indicating the strand (left arrows - minus strand , right arrows - plus strand).
- Below, we observe the whole transcript along with its coordinates and its regions (3'-UTR , CDS , 5'-UTR).
Leaving the mouse onto a region component, the coordinates of this region are shown up.
- Next, we observe only the part of the transcript, where the microRNAs are bound.
Similarly , leaving the mouse onto a region component, the coordinates of that region are shown up.
The coordinates shown here, right and left of the partial-transcript are the coordinates of the transcript regions that are depicted.
- Lastly, the microRNAs are displayed as small rectangular components. The color of the rectangular component indicates the score as it is described above.
Next to each microRNA component is either its name or its family name with bold letters.
In case a family name is shown instead of microRNA name, that means in that position, more than one microRNAs belonging to the same family, start their interaction with the current miTG.
- Leaving the mouse onto a microRNA rectangular component, useful information are shown up.
Clicking on a microRNA name , all the microRNAs belonging to the same miR family, as the one that was clicked, are noted with a rectangular.

7) Statistics

In this part, we observe some statistical inferences, that are calculated on the fly, each time the user makes a new search.

FIGURE 4.5: Στη βοηθητική σελίδα μπορούμε ακόμα να ενημερωθούμε για τη ροή την οποία ακολουθεί η εφαρμογή, καθώς και για τα τελικά αποτελέσματα, τα οποία εμφανίζονται.

4.2.2 Φίλτρα Αναζήτησης

Ας δούμε τώρα το κρυφό – αναδυόμενο παράθυρο (Σχήμα 4.6), το οποίο αναφέραμε στην αρχή του παρόντος κεφαλαίου και το οποίο εμπεριέχει τα φίλτρα με τα οποία ο χρήστης μπορεί να ρυθμίσει ή να περιορίσει τα αποτελέσματα της αναζήτησής του. Πρέπει να σημειώσουμε ότι αυτά τα φίλτρα έχουν κάποιες αρχικά προκαθορισμένες τιμές, οι οποίες υποθέσαμε ότι αντιπροσωπεύουν μία μέση αναζήτηση ενός χρήστη.



FIGURE 4.6: Το κουμπί με το οποίο εμφανίζεται το παράθυρο, το οποίο περιέχει τα φίλτρα αναζήτησης.

Close x

Search-Filters

Search

Score threshold:

Binding site :

Conserved Species threshold :

Conserved Species :

-
- panTro2
- rheMac2
- oryCun2
- canFam2
- dasNov2
- loxAfr3
- echTel1
- monDom5
- mm9

Binding type :

-
- 6mer
- 7mer
- 8mer
- 9mer
- 8mer+mismatch
- 8mer+wobble

FIGURE 4.7: Το παράθυρο το οποίο περιέχει τα φίλτρα αναζήτησης, καθώς και μία φόρμα εισόδου για τη διεξαγωγή νέας αναζήτησης.

Στην παραπάνω εικόνα (Σχήμα 4.7) φαίνεται το παράθυρο, το οποίο εμπεριέχει τα φίλτρα αναζήτησης. Στη συνέχεια θα παρουσιάσουμε ποια είναι αυτά, καθώς επίσης και τις προκαθορισμένες τιμές τους.

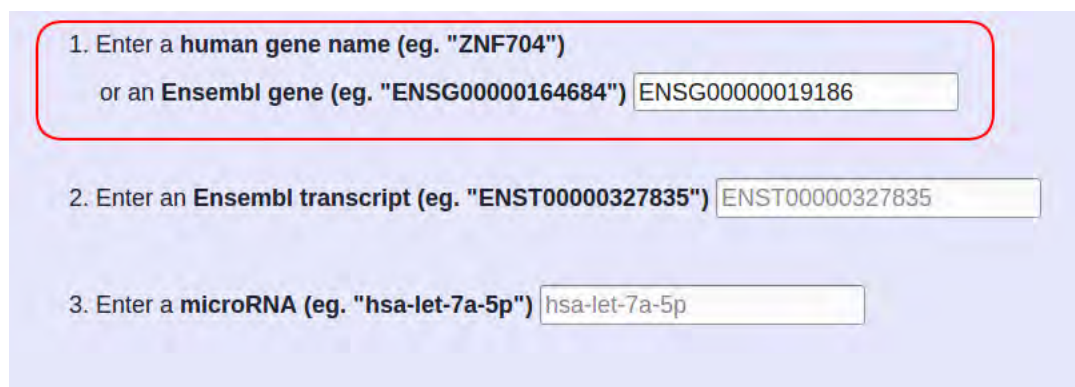
Αρχικά, υπάρχει η δυνατότητα, στις αναζητήσεις μας να εισάγουμε ένα κατώφλι, το οποίο αφορά στο miTG score. Αυτό το κατώφλι μπορούμε να το εισάγουμε στο πεδίο με όνομα “Score threshold” και με αυτόν τον τρόπο να εμφανίζονται μόνο οι αλληλεπιδράσεις microRNA:στόχων, οι οποίες έχουν βαθμολογία υψηλότερη από αυτό το όριο. Το miTG score αντιπροσωπεύει ένα score πρόβλεψης και προέρχεται από τον αλγόριθμο DIANA-microT-CDS. Όσο υψηλότερο είναι το miTG score, τόσο μεγαλύτερη είναι η πιθανότητα στόχευσης (targeting). Με άλλα λόγια αυτή η παράμετρος μετρά την ισχύ κάθε αλληλεπίδρασης miRNA-γονιδίου. Η προκαθορισμένη τιμή του είναι 0.85 και το εύρος των τιμών του είναι από 0.5 έως 1.0. Στη συνέχεια, μπορούμε να καθορίσουμε την περιοχή του mRNA, η οποία μας ενδιαφέρει (3’UTR ή CDS ή και τις δύο ταυτόχρονα) και να δούμε τις αλληλεπιδράσεις που συμβαίνουν μόνο εκεί. Οι δυνατές τιμές επιλογής είναι είτε 3’UTR περιοχή, είτε CDS περιοχή, είτε και οι δύο στην περίπτωση που αφήσουμε αυτή την επιλογή κενή στο πεδίο με όνομα “Binding Site”. Ως προκαθορισμένη τιμή έχουμε επιλέξει να εμφανίζονται αλληλεπιδράσεις σε οποιαδήποτε από αυτές τις δύο περιοχές. Ακολούθως, διακρίνουμε το φίλτρο με όνομα “Conserved Species threshold”. Αυτό το φίλτρο παίρνει αριθμητικές τιμές από 1 έως 18, όσα δηλαδή και τα είδη των οργανισμών που υπάρχουν στη βάση δεδομένων μας. Εισάγωντας έναν αριθμό στο πεδίο αυτό, τα αποτελέσματα φιλτράρονται, έτσι ώστε να εμφανιστούν μόνο οι αλληλεπιδράσεις με τα MREs, τα οποία έχουν διατηρημένα τα νουκλεοτίδια πρόσδεσης στη θέση στόχο, σε τόσα ή και παραπάνω είδη οργανισμών, όσο και ο εισαχθέν αριθμός. Ως προεπιλεγμένη τιμή εδώ έχουμε επιλέξει τον αριθμό 1, που σημαίνει ότι θα εμφανιστούν οι αλληλεπιδράσεις που έχουν σε τουλάχιστον έναν οργανισμό διατηρημένα τα νουκλεοτίδια πρόσδεσης στη θέση στόχο. Με το επόμενο φίλτρο εν ονόματι “Conserved Species”, μπορούμε να διεξάγουμε μία αναζήτηση, έτσι ώστε τα αποτελέσματα τα οποία θα εμφανίζονται να έχουν διατηρήσει τα νουκλεοτίδια πρόσδεσης στη θέση στόχο, τουλάχιστο για τα επιλεγμένα είδη οργανισμών. Ως προκαθορισμένη τιμή έχουμε επιλέξει να εμφανίζονται αποτελέσματα με οποιοδήποτε είδος οργανισμού. Τέλος, μπορούμε να επιλέξουμε τον τύπο πρόσδεσης, τον οποίο μας ενδιαφέρει να μελετήσουμε από το πεδίο “Binding Type”. Οι δυνατές τιμές επιλογής εδώ είναι : 6mer , 7mer , 8mer , 9mer , 8mer+mismatch , 8mer+wobble. Η προκαθορισμένη ρύθμιση είναι η επιλογή όλων. Δηλαδή θα εμφανιστούν όλα τα πιθανά MREs. Ένα τελευταίο σημείο που πρέπει να σημειώσουμε εδώ, είναι ότι στην κορυφή αυτού του παραθύρου (Σχήμα 4.7) υπάρχει ένα “πεδίο εισαγωγής κειμένου” με όνομα “Search”, από το οποίο μπορούμε να εκτελέσουμε μία νέα αναζήτηση, χωρίς να χρειάζεται να ανακατευθυνθούμε στην αρχική σελίδα της εφαρμογής.

4.2.3 Ροή εφαρμογής

Όπως είδαμε και παραπάνω έχουμε τρεις δυνατές επιλογές με τις οποίες μπορούμε να ξεκινήσουμε μία αναζήτηση.

1. Εισαγωγή γονιδίου
2. Εισαγωγή μεταγράφου
3. Εισαγωγή microRNA

Ας εξετάσουμε λοιπόν, κάθε μία από αυτές ξεχωριστά με σκοπό να δούμε τη ροή της εφαρμογής μας.

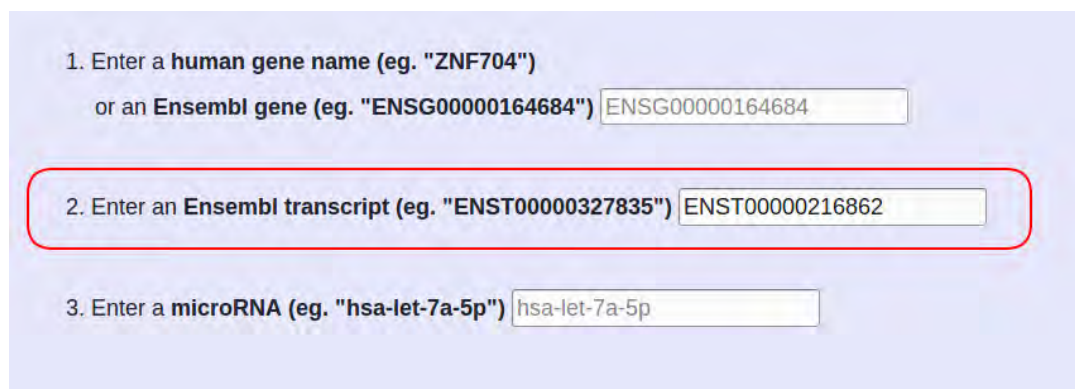


1. Enter a **human gene name** (eg. "ZNF704")
or an **Ensembl gene** (eg. "ENSG00000164684")

2. Enter an **Ensembl transcript** (eg. "ENST00000327835")

3. Enter a **microRNA** (eg. "hsa-let-7a-5p")

FIGURE 4.8: Αναζήτηση με βάση ένα γονίδιο.

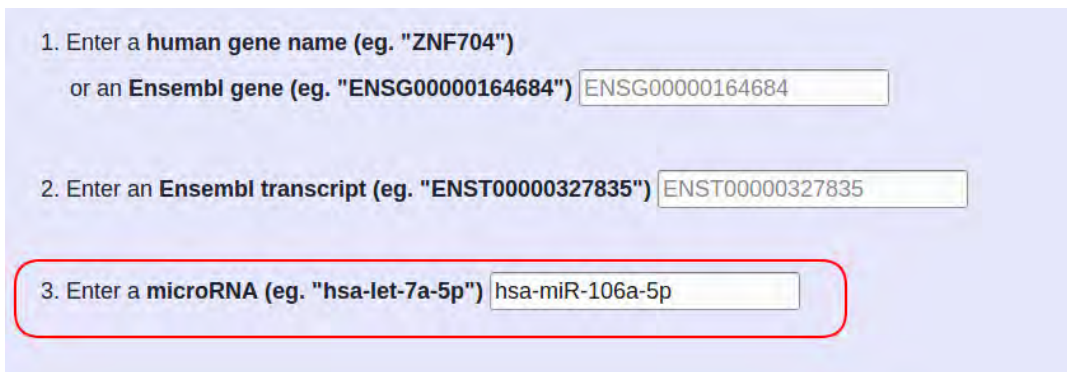


1. Enter a **human gene name** (eg. "ZNF704")
or an **Ensembl gene** (eg. "ENSG00000164684")

2. Enter an **Ensembl transcript** (eg. "ENST00000327835")

3. Enter a **microRNA** (eg. "hsa-let-7a-5p")

FIGURE 4.9: Αναζήτηση με βάση ένα μετάγραφο.



1. Enter a **human gene name** (eg. "ZNF704")
or an **Ensembl gene** (eg. "ENSG00000164684")

2. Enter an **Ensembl transcript** (eg. "ENST00000327835")

3. Enter a **microRNA** (eg. "hsa-let-7a-5p")

FIGURE 4.10: Αναζήτηση με βάση ένα microRNA.

1. Εισαγωγή γονιδίου

Ο χρήστης εισάγει ένα όνομα ή ένα αναγνωριστικό (ID) ενός γονιδίου (Σχήμα 4.8). Η σελίδα η οποία εμφανίζεται είναι η παρακάτω (Σχήμα 4.11).



DIANA TOOLS
microRNA - miTG Visualization

Home Statistics Help

Choose a relative transcript

Transcript ID	Transcript Name
ENST00000216862	CYP24A1-003
ENST00000395955	CYP24A1-004
ENST00000460643	CYP24A1-005
ENST00000395954	CYP24A1-006
ENST00000487593	CYP24A1-002
ENST00000472970	CYP24A1-001

FIGURE 4.11: Μετά την εισαγωγή ενός γονιδίου ως είσοδο, εμφανίζονται όλα τα μετάγραφα του, τα οποία αλληλεπιδρούν με κάποιο microRNA.

Σε αυτή τη σελίδα εμφανίζονται όλα τα πιθανά μετάγραφα, τα οποία ανήκουν στο εισαχθέν γονίδιο και για τα οποία έχουμε πληροφορίες στη βάση δεδομένων μας ότι αλληλεπιδρούν με κάποιο microRNA. Ο χρήστης σε αυτό το στάδιο πρέπει να επιλέξει ένα από τα διαθέσιμα μετάγραφα για να συνεχίσει την αναζήτησή του.



DIANA TOOLS
microRNA - miTG Visualization

Home Statistics Help

microRNA hsa-miR-106a-5p interacts with the following transcripts

Search filter...

Transcript ID	Gene ID	Score	Chromosome
ENST00000369615	ENSG00000081026	0.99821831	1
ENST00000366518	ENSG00000162849	0.99783077	1
ENST00000369317	ENSG00000117289	0.99693264	1
ENST00000261441	ENSG00000081019	0.99654502	1
ENST00000367701	ENSG00000185278	0.99304543	1
ENST00000482352	ENSG00000242485	0.99195130	1
ENST00000370521	ENSG00000065243	0.99121141	1
ENST00000303635	ENSG00000171735	0.98523805	1
ENST00000349213	ENSG00000054267	0.98415463	1
ENST00000368006	ENSG00000158796	0.98381024	1
ENST00000360851	ENSG00000143344	0.98094393	1
ENST00000368985	ENSG00000163131	0.97916016	1

FIGURE 4.13: Μετά την εισαγωγή ενός microRNA ως είσοδο, εμφανίζονται όλα τα μετάγραφα με τα οποία αλληλεπιδρά και ο χρήστης σε αυτό το σημείο πρέπει να κάνει μία επιλογή.

Σε αυτή τη σελίδα εμφανίζονται όλα τα μετάγραφα, τα οποία αλληλεπιδρούν με το εισαχθέν microRNA, σύμφωνα με τα στοιχεία που έχουμε στη βάση δεδομένων μας και τα οποία τα έχουμε λάβει από τις προβλέψεις του προγράμματος DIANA-microT-CDS. Δίπλα από κάθε μετάγραφο εμφανίζεται το γονίδιο στο οποίο ανήκει, το χρωμόσωμα, καθώς επίσης και το miTG score της αλληλεπίδρασης, όπως αυτό περιγράφηκε παραπάνω. Επειδή τα αποτελέσματα τα οποία εμφανίζονται σε αυτή την περίπτωση είναι συνήθως πολλά, έχουμε εισάγει ένα πεδίο αναζήτησης, με σκοπό να διευκολύνουμε τον χρήστη, στην περίπτωση την οποία ψάχνει να δει αν υπάρχει αλληλεπίδραση με ένα συγκεκριμένο γονίδιο ή μετάγραφο. Σε αυτό το πεδίο αναζήτησης έχεις επίσης τη δυνατότητα να ψάξεις με βάση το miTG score. Στη συνέχεια, και αφού ο χρήστης επιλέξει ένα μετάγραφο, εμφανίζονται τα τελικά αποτελέσματα της αναζήτησής του (Σχήμα 4.14), με παρόμοια μορφή όπως τα παραπάνω (Σχήμα 4.12).

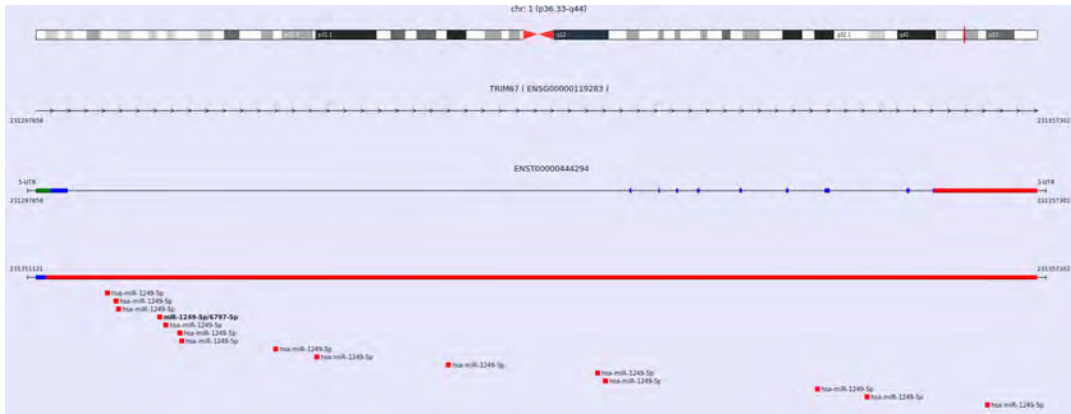


FIGURE 4.14: Το περιεχόμενο της σελίδας των αποτελεσμάτων, όταν έχουμε αρχικά δώσει ως είσοδο ένα microRNA.

Η μόνη διαφορά είναι ότι σε αυτή την περίπτωση, στα τελικά αποτελέσματα εμφανίζεται μόνο το microRNA, το οποίο είχε επιλέξει στην αρχή και τα MREs στα οποία στοχεύει. Αντιθέτως, δεν εμφανίζονται όλα τα microRNAs, τα οποία αλληλεπιδρούν με το μετάγραφο, το οποίο διάλεξε στην πορεία της αναζήτησής του. Ενδεικτικά αποτελέσματα φαίνονται στην παρακάτω εικόνα (Σχήμα 4.15).

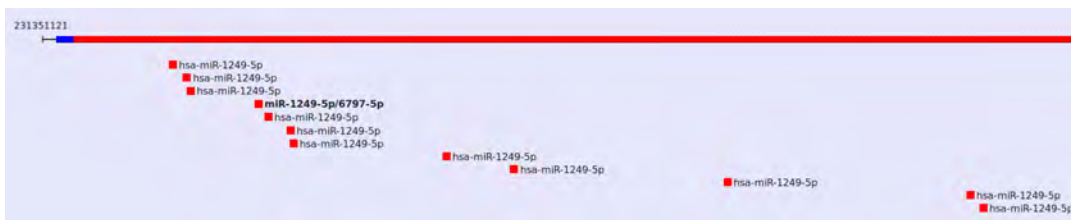


FIGURE 4.15: Το περιεχόμενο των αποτελεσμάτων σε αυτή την περίπτωση σχετίζεται μόνο με το microRNA, το οποίο είχε δώσει ο χρήστης ως αρχική είσοδο.

Στο επόμενο σχήμα (Σχήμα 4.16) φαίνεται ένα διάγραμμα της ροής της web εφαρμογής, όπως την περιγράψαμε παραπάνω:

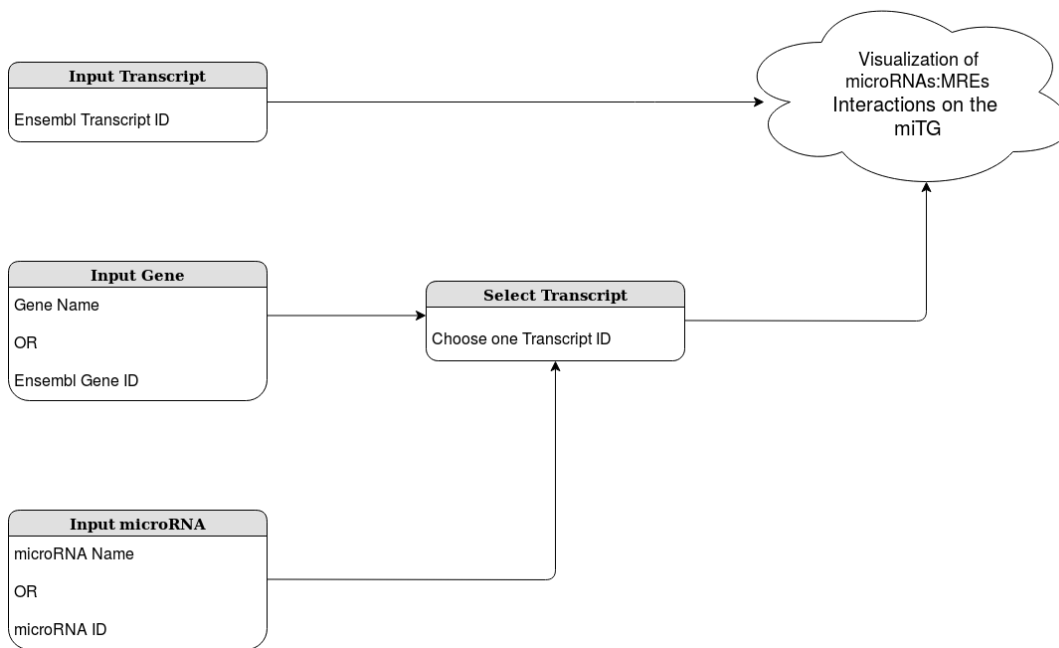


FIGURE 4.16: Το διάγραμμα ροής, το οποίο ακολουθεί η εφαρμογή.

4.2.4 Ανάλυση τελικών αποτελεσμάτων

Σε αυτό το σημείο ήρθε η ώρα να παρατηρήσουμε και να αναλύσουμε τα τελικά αποτελέσματα της εφαρμογής, όπως αυτά εμφανίζονται στις εικόνες 4.12 ή 4.15 που παρουσιάσαμε παραπάνω. Ας δούμε λοιπόν πιο αναλυτικά τι απεικονίζεται σε αυτή τη σελίδα. Αρχικά και στην κορυφή της σελίδας εμφανίζεται ένα χρωμόσωμα όπως αυτό φαίνεται στην επόμενη εικόνα (Σχήμα 4.17). Το παρόν χρωμόσωμα είναι αυτό, πάνω στο οποίο βρίσκεται το γονίδιο, το οποίο συμμετέχει στην παρούσα αλληλεπίδραση. Αυτό το γονίδιο το απεικονίζουμε με μία κάθετη κόκκινη γραμμή πάνω στο χρωμόσωμα, στην ακριβή θέση στην οποία βρίσκεται. Πάνω ακριβώς από το χρωμόσωμα υπάρχει το όνομά του και το εύρος των συντεταγμένων του. Ακόμα φαίνονται οι κυτταρογενετικές ζώνες του (cytogenetics bands¹), καθώς και η αρίθμησή τους, η οποία είναι φθίνουσα προς το κέντρο του χρωμοσώματος.

¹Όποιαδήποτε από τις υποπεριοχές ενός χρωμοσώματος που είναι ορατές κάτω από ένα μικροσκόπιο μετά την χρώση

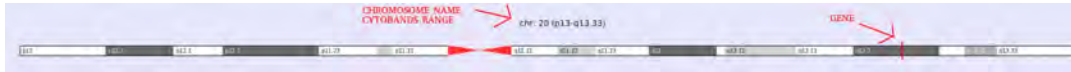


FIGURE 4.17: Το χρωμόσωμα στο οποίο ανήκει το γονίδιο της έκαστης αλληλεπίδρασης. Πάνω στο χρωμόσωμα φαίνεται με μία κόκκινη κάθετη γραμμή η θέση του γονιδίου.

Στη συνέχεια και ακριβώς από κάτω απεικονίζουμε το γονίδιο, το οποίο συμμετέχει στην αλληλεπίδραση. Τα βέλη τα οποία υπάρχουν πάνω στο γονίδιο υποδεικνύουν τη “strand” από την οποία έχει προέλθει. Όταν τα βέλη έχουν δεξιά φορά σημαίνει ότι το γονίδιο έχει προέλθει από τη θετική ή positive strand (+1), ενώ όταν έχουν φορά προς τα αριστερά το γονίδιο έχει προέλθει από την αρνητική ή negative ή strand (-1). Στο προκείμενο παράδειγμα και με βάση την παρακάτω εικόνα (Σχήμα 4.18) βλέπουμε ότι βρισκόμαστε στην αρνητική strand. Ακόμα, εκατέρωθεν του γονιδίου μπορούμε να διακρίνουμε τις συντεταγμένες του. Αυτές οι συντεταγμένες μας υποδεικνύουν σε ποιες θέσεις πάνω στο χρωμόσωμα ξεκινάει και τελειώνει το παρόν γονίδιο. Τέλος, πάνω ακριβώς από το γονίδιο φαίνεται το όνομά του, καθώς επίσης και το αναγνωριστικό του με βάση την ensembl (EN-SEMBL ID).

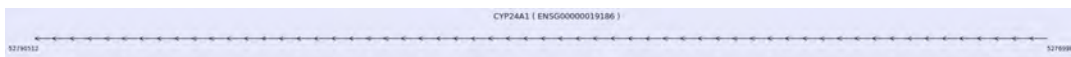


FIGURE 4.18: Το γονίδιο όπως παρουσιάζεται, με το όνομά του, τις συντεταγμένες του και την κατεύθυνσή του, η οποία υποδηλώνει τη strand.

Το επόμενο στοιχείο το οποίο διακρίνουμε στη σελίδα των αποτελεσμάτων είναι το μετάγραφο mRNA (Σχήμα 4.19). Πάνω από το μετάγραφο διακρίνουμε το αναγνωριστικό (ID) της ensembl που το προσδιορίζει, ενώ εκατέρωθεν αυτού φαίνονται οι συντεταγμένες του (συντεταγμένες που προσδιορίζουν τη θέση του πάνω στο γονίδιο), όπως επίσης και τα άκρα 3' prime και 5' prime στην δεξιά και αριστερή μεριά, ανάλογα με τη strand στην οποία βρισκόμαστε. Πάνω στο μετάγραφο φαίνονται με διαφορετικό χρώμα οι περιοχές, οι οποίες το απαρτίζουν (Σχήμα 4.20). Πιο συγκεκριμένα με κόκκινο χρώμα σημειώνουμε τις 3'UTR περιοχές, με πράσινο τις 5'UTR, και με μπλε τις CDS (coding sequence) περιοχές. Δύναται κάποιες από αυτές τις περιοχές οι οποίες έχουν σχετικά μικρό μήκος σε σχέση με τις υπόλοιπες, να μη φαίνονται πολύ αναλυτικά, λόγω περιορισμού

στην ανάλυση της οθόνης μας και ταυτόχρονα, λόγω του πολύ μεγάλου φυσικού μεγέθους αυτών των περιοχών, σε βάσεις (bases).



FIGURE 4.19: Το μετάγραφο όπως παρουσιάζεται, με το αναγνωριστικό ID του, τις συντεταγμένες του και τις περιοχές, οι οποίες το απαρτίζουν (3'UTR, 5'UTR, CDS).

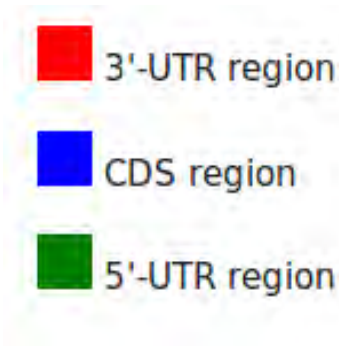


FIGURE 4.20: Τα χρώματα που έχουμε αναθέσει σε κάθε υποπεριοχή του μεταγράφου (3'UTR, 5'UTR, CDS), έτσι ώστε να τις διαχωρίζουμε.

Παρ'ολ'αυτά όμως, αν κάποια αλληλεπίδραση συμβαίνει σε μία από αυτές τις περιοχές θα είμαστε σε θέση να τη δούμε αναλυτικότερα στο επόμενο γραφικό στοιχείο που παρουσιάζεται (Σχήμα 4.21).



FIGURE 4.21: Οι περιοχές του μεταγράφου, οι οποίες συμμετέχουν στην αλληλεπίδραση, εμφανίζονται σε αυτό το γραφικό σχήμα σε μεγέθυνση.

Όπως παρατηρούμε το κάθε “microRNA στοιχείο” που απεικονίζεται, συνδέεται με ένα χρώμα. Αυτό το χρώμα μας υποδεικνύει το miTG score της συγκεκριμένης αλληλεπίδρασης, μεταξύ του εκάστοτε microRNA και του στόχου. Το miTG score όπως έχουμε αναφέρει και παραπάνω, είναι ένα υπολογισθέν score (prediction score), το οποίο εξάγεται από την εκτέλεση του αλγορίθμου DIANA-microT-CDS και το οποίο μας δείχνει την ισχύ της καθε microRNA:miTG αλληλεπίδρασης. Με άλλα λόγια μας δείχνει την πιθανότητα να υφίστανται αυτή η αλληλεπίδραση. Αναλυτικότερα η σημασία των χρωμάτων φαίνεται παρακάτω (Σχήμα 4.24) :

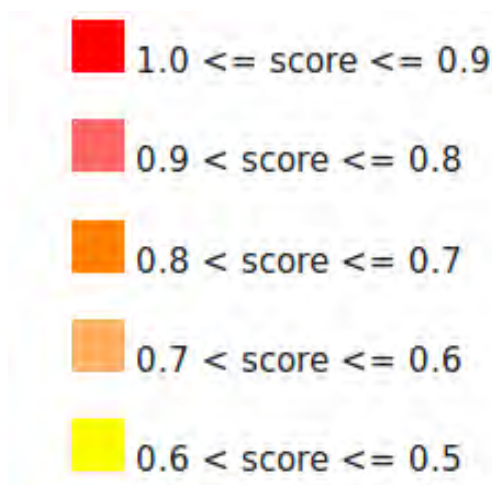


FIGURE 4.24: Τα διαφορετικά χρώματα τα οποία αναθέτουμε σε κάθε microRNA, με σκοπό να τα κατατάξουμε σε κατηγορίες, ανάλογα με το miTG Score.

Όπως ήδη αναφέραμε προ λίγου, στα αποτελέσματα φαίνονται τα ονόματα των microRNAs και των οικογενειών (microRNA families). Σε αυτό το σημείο όμως πρέπει να αναφέρουμε μία σύμβαση την οποία έχουμε ακολουθήσει για την απεικόνισή τους. Κατά βάση δίπλα από κάθε “microRNA component” (τετραγωνάκι) παρουσιάζουμε το όνομα του microRNA. Στην περίπτωση όμως όπου υπάρχουν πολλά microRNAs, τα οποία ανήκουν στην ίδια οικογένεια και ταυτόχρονα αλληλεπιδρούν με το μετάγραφο ξεκινώντας από την ίδια θέση πρόσδεσης, αντί να εμφανίζουμε όλα αυτά τα “microRNAs components” με το όνομά τους, εμφανίζουμε ένα component με το όνομα της οικογένειας (miRNA family). Το όνομα της οικογένειας φαίνεται πάντα με έντονα (bold) γράμματα (Σχήμα 4.25).

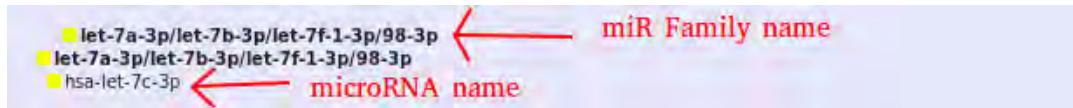


FIGURE 4.25: Τα ονόματα των microRNA οικογενειών εμφανίζονται πάντοτε με έντονα γράμματα, ενώ τα ονόματα των microRNAs με κανονική γραμματοσειρά.

Στη συνέχεια, μπορούμε να εξάγουμε διάφορα χρήσιμα συμπεράσματα αλληλεπιδρώντας με τα γραφικά. Πιο συγκεκριμένα πατώντας (click) πάνω σε ένα microRNA ή miR family όνομα, όλα τα microRNAs, τα οποία ανήκουν στην ίδια οικογένεια με αυτό το οποίο “κλικάραμε”, σημειώνονται με ένα ορθογώνιο σχήμα γύρω από το όνομά τους (Σχήμα 4.26). Ακόμα, αφήνοντας το ποντίκι πάνω σε ένα “microRNA component” (τετραγωνάκι), εμφανίζονται σχετικές πληροφορίες, οι οποίες αφορούν στο συγκεκριμένο microRNA (Σχήμα 4.27). Πιο συγκεκριμένα μπορούμε να παρατηρήσουμε το MIMAT ID του microRNA ή των microRNAs, όταν πρόκειται για miR family, την seed αλληλουχία του miRNA, το predicted miTG score της αλληλεπίδρασής του (το μεγαλύτερο από τα scores, όταν πρόκειται για miR family που απαρτίζεται από πολλά microRNAs), σε ποιά θέση ξεκινάει και τελειώνει η πρόσδεσή του πάνω στο στόχο, καθώς επίσης και σε ποια περιοχή (3’UTR, CDS) πάνω στο στόχο προσδένεται και με ποιόν τύπο πρόσδεσης (6mer, 7mer ...).

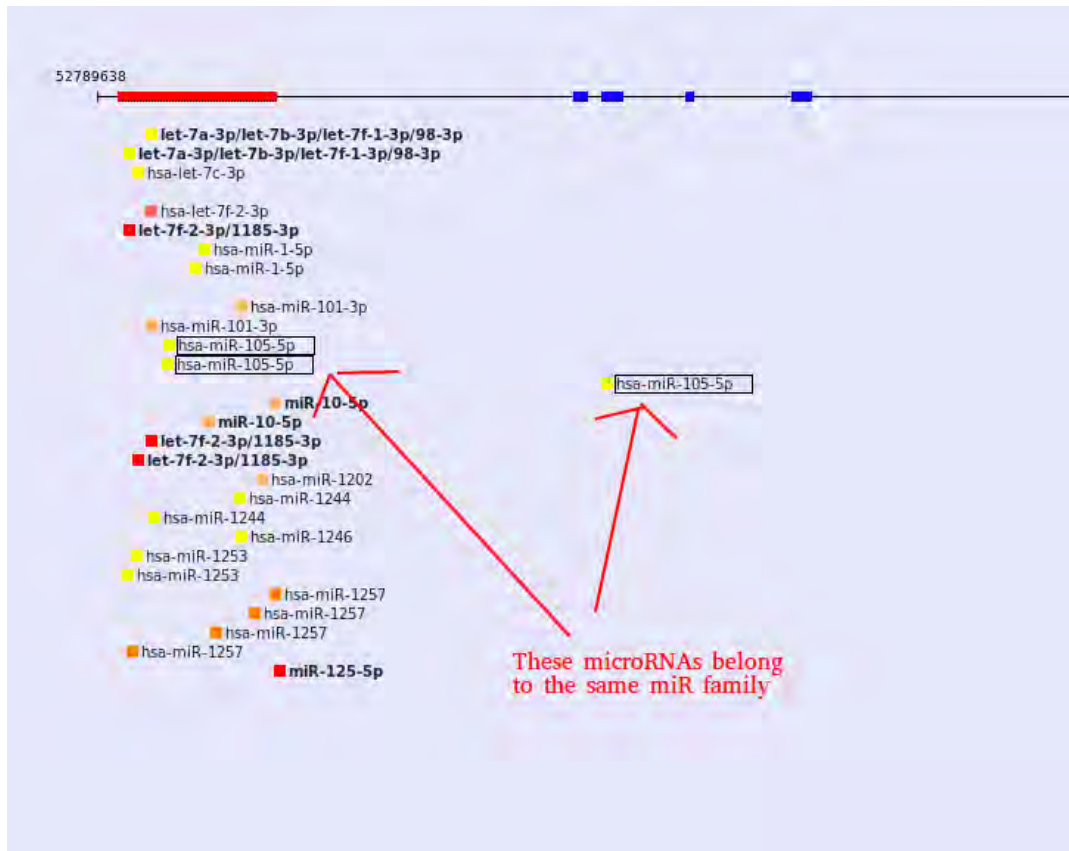


FIGURE 4.26: “Κιλικάροντας” πάνω σε ένα microRNA, σημειώνονται με ένα ορθογώνιο περίγραμμα, όλα τα microRNAs, τα οποία ανήκουν στην ίδια οικογένεια με αυτό.

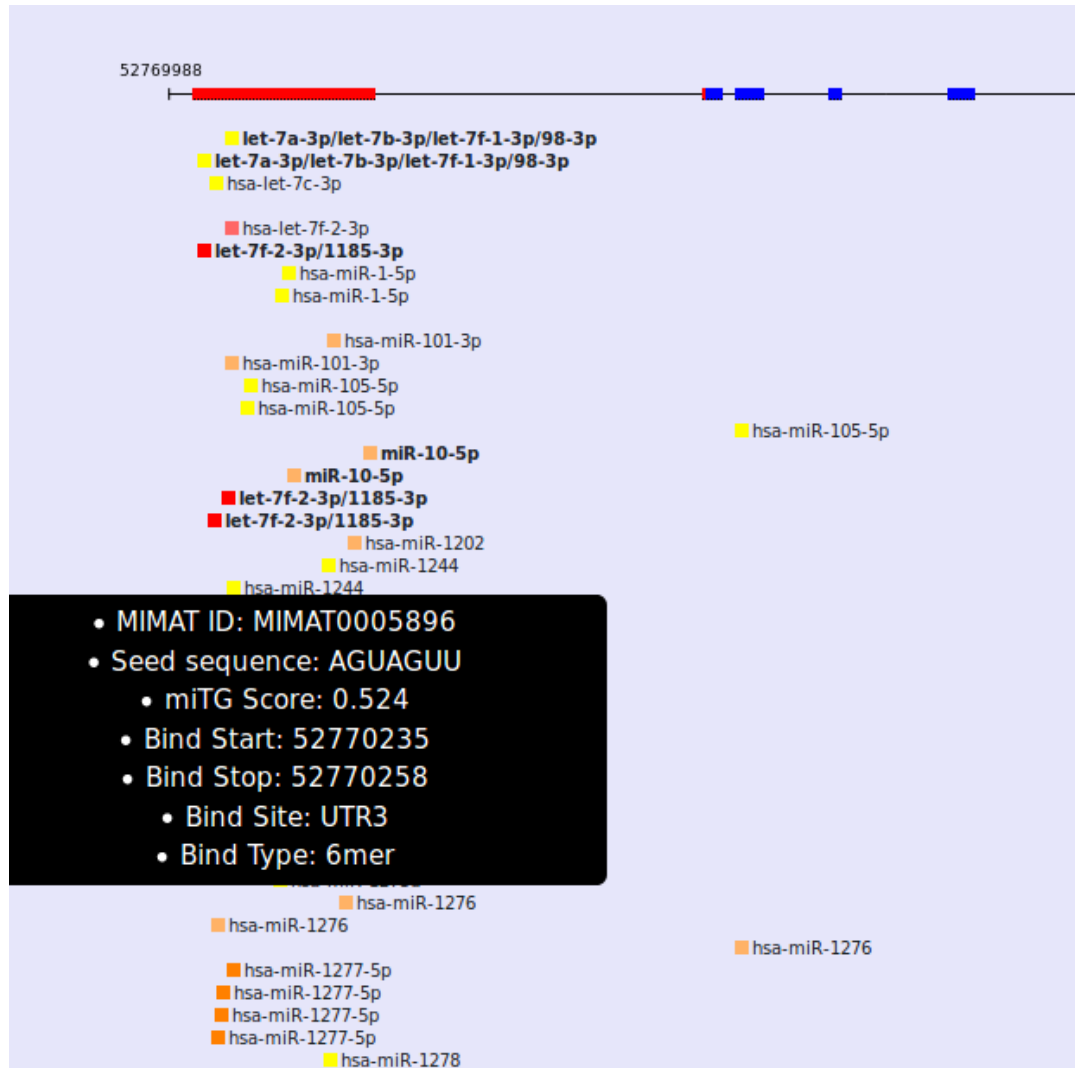


FIGURE 4.27: Οι σχετικές πληροφορίες οι οποίες εμφανίζονται όταν αφήνουμε το ποντίκι πάνω σε ένα microRNA component είναι: αναγνωριστικό ID του microRNA, η seed αλληλουχία του, το predicted miTG score της αλληλεπίδρασής του με το μετάγραφο, η αρχική και η τελική θέση πρόσδεσής του πάνω στον στόχο, η υποπεριοχή του στόχου στην οποία προσδένεται (3'UTR, CDS) και τέλος ο τύπος πρόσδεσης (6mer, 9mer ..).

Τέλος, σημειώνουμε ότι έχουμε εγκαταστήσει μια μπάρα κύλισης στη δεξιά πλευρά του παραθύρου, έτσι ώστε όταν τα αποτελέσματα των microRNAs που εμφανίζονται είναι πολλά, ο χρήστης να έχει τη δυνατότητα να κάνει “scroll” και να βλέπει πάντα τα microRNAs σε σχέση με το μετάγραφο – στόχο και τις υποπεριοχές του (Σχήμα 4.28).

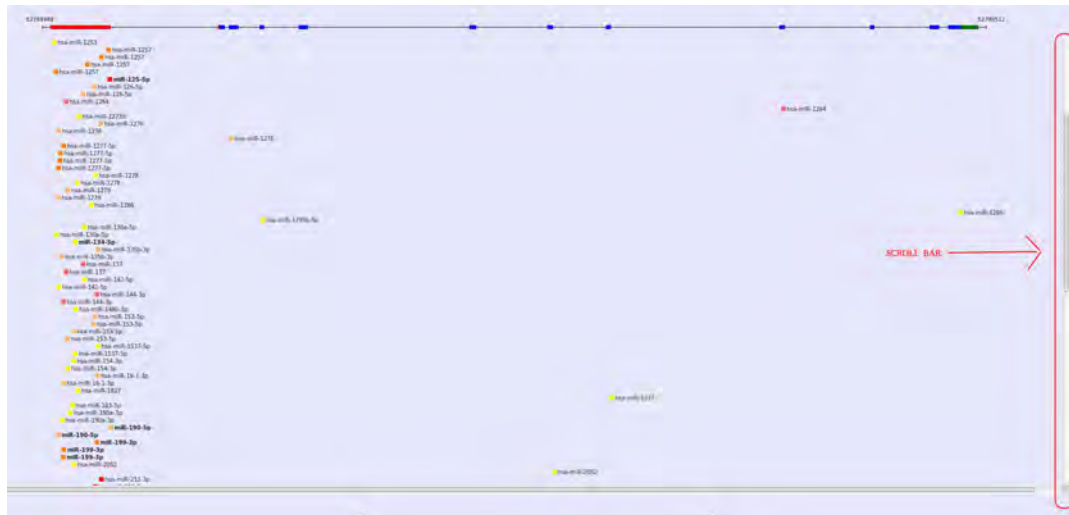


FIGURE 4.28: Η μπάρα κύλισης βοηθάει στην περίπτωση όπου στα αποτελέσματα υπάρχουν πολλά microRNA components, καθώς μπορούμε να κάνουμε scroll προς τα κάτω με σκοπό να δούμε όλα τα microRNA components, πάντα όμως σε σχέση με τις περιοχές του μεταγράφου - στόχου.

Chapter 5

Διαδικτυακές Εφαρμογές - Django Framework

5.1 Διαδικτυακές Εφαρμογές

Στην παρούσα διπλωματική εργασία υλοποιήσαμε μία διαδικτυακή εφαρμογή (Web Application), η οποία εκμεταλλεύεται τα δεδομένα, στα οποία αναφερθήκαμε στις προηγούμενες ενότητες και παρουσιάζει τα σχετικά αποτελέσματα στον χρήστη. Είναι σκόπιμο λοιπόν, σε αυτό το σημείο να αναφερθούμε γενικότερα στις διαδικτυακές εφαρμογές και να δούμε ποια είναι η αρχιτεκτονική τους, καθώς επίσης και ποιές τεχνολογίες χρησιμοποιούνται εν γένει σε κάθε επίπεδο.

Αν θέλαμε να δώσουμε έναν ορισμό στο τί είναι μία διαδικτυακή εφαρμογή, θα μπορούσαμε να πούμε το εξής: “Μία διαδικτυακή εφαρμογή είναι ένα πρόγραμμα σε έναν υπολογιστή, το οποίο χρησιμοποιεί τους web browsers (προγράμματα περιηγητή ιστού) και την διαδικτυακή τεχνολογία (web technology), με σκοπό να εκτελέσει ορισμένες εργασίες μέσω του διαδικτύου”. Οι περισσότερες από τις επιχειρήσεις στον σημερινό κόσμο χρησιμοποιούν το Internet, ως ένα οικονομικά αποδοτικό κανάλι για επικοινωνία. Για παράδειγμα, ανταλλάσσουν άμεσα πληροφορίες με τους πελάτες και την αγορά γενικότερα, καθώς επίσης πραγματοποιούν και ασφαλείς συναλλαγές. Παρ’ όλ’ αυτά, η εν λόγω αποτελεσματικότητα καθίσταται υπαρκτή και δυνατή, μόνο αν η εκάστοτε επιχείρηση είναι σε θέση να καταγράφει και να αποθηκεύει όλα τα απαραίτητα δεδομένα και να έχει ένα μέσο για την επεξεργασία αυτών των πληροφοριών και την παρουσίασή τους στον χρήστη. Σε αυτό το σημείο εμφανίζονται οι διαδικτυακές εφαρμογές, οι οποίες χρησιμοποιούν ένα συνδυασμό, με τον οποίο είναι σε θέση, τόσο να διαχειρίζονται την αποθήκευση και την ανάκτηση των πληροφοριών, χρησιμοποιώντας server-side τεχνολογίες (PHP,ASP...), όσο και να τις παρουσιάζουν στον χρήστη, μέσω client-side τεχνολογιών (Javascript και HTML). Αυτός ο συνδυασμός, επιτρέπει στους χρήστες να αλληλεπιδρούν με την εταιρία χρησιμοποιώντας ηλεκτρονικές φόρμες για παράδειγμα, ή διαχειρίζοντας το καλάθι των αγορών τους. Επιπρόσθετα οι εφαρμογές δίνουν τη δυνατότητα στους υπαλλήλους λόγω χάρη, να δημιουργούν διάφορα έγγραφα, να τα μοιράζονται μεταξύ τους ή να συνεργάζονται σε projects και

να δουλεύουν πάνω σε κοινά έγγραφα, ανεξάρτητα από το πού βρίσκονται ή τί συσκευή χρησιμοποιούν. [32]

5.1.1 Πώς λειτουργούν οι διαδικτυακές εφαρμογές

Οι web εφαρμογές συνήθως γράφονται σε γλώσσα, την οποία καταλαβαίνουν και υποστηρίζουν τα προγράμματα περιήγησης ιστού (web browsers), όπως οι Javascript και HTML, καθώς αυτοί είναι, οι οποίοι θα εκτελέσουν και θα καταστήσουν λειτουργικά τα προγράμματα. Παλαιότερα, τα περισσότερα προγράμματα ήταν στατικά, δηλαδή παρουσίαζαν την ίδια πληροφορία κάθε φορά στον χρήστη. Σήμερα όμως, το μεγαλύτερο πλήθος των προγραμμάτων είναι δυναμικά και απαιτούν κάποιου είδους επεξεργασία στη μεριά του server, με σκοπό να ικανοποιήσουν τις επιλογές και τις απαιτήσεις των χρηστών.

Μία διαδικτυακή εφαρμογή εν γένει, απαιτεί έναν “web server”, ο οποίος διαχειρίζεται τις αιτήσεις (requests) του χρήστη, έναν “application server”, ο οποίος είναι υπεύθυνος για την εκτέλεση των λειτουργιών, οι οποίες είναι αναγκαίες για να ικανοποιηθεί το αίτημα του χρήστη, και σε ορισμένες περιπτώσεις απαιτείται και μία βάση δεδομένων για να αποθηκεύονται πληροφορίες και δεδομένα. Οι τεχνολογίες οι οποίες αφορούν στον application server, μπορεί να είναι ASP.NET, ASP, ColdFusion, PHP, JSP, Python και άλλες πολλές.

Παρακάτω θα δούμε τη ροή, την οποία ακολουθεί μία τυπική εφαρμογή [32].

1. Ο χρήστης στέλνει ένα αίτημα στον web server, χρησιμοποιώντας το Internet, είτε μέσω του web browser, είτε μέσω της διεπαφής χρήστη της εφαρμογής.
2. Ο web server λαμβάνει το αίτημα και το προωθεί στον κατάλληλο web application server.
3. Ο web application server εκτελεί την απαιτούμενη λειτουργία, η οποία ζητήθηκε από τον χρήστη, με το να παίρνει δεδομένα απο τη βάση δεδομένων και να τα επεξεργάζεται για παράδειγμα, έτσι ώστε να εξάγει τα αποτελέσματα.
4. Ο web application server στη συνέχεια, στέλνει τα αποτελέσματα που εξήγαγε πίσω στον web server.
5. Ο web server απαντάει στον “client” με τις απαιτούμενες πληροφορίες και τότε εμφανίζονται τα αποτελέσματα στην οθόνη του χρήστη.

Η παραπάνω περιγραφή φαίνεται σχηματικά στην εικόνα που ακολουθεί (Σχήμα 5.1).



FIGURE 5.1: Ένα παράδειγμα μίας τυπικής ροής, η οποία ακολουθεί μία διαδικτυακή εφαρμογή. [32]

5.1.2 Τα πλεονεκτήματα των διαδικτυακών εφαρμογών

Τα πλεονεκτήματα των διαδικτυακών εφαρμογών είναι πολλά και μας διευκολύνουν σε πολλούς τομείς. Στη συνέχεια θα απαριθμήσουμε ορισμένα από αυτά, με σκοπό να δείξουμε τη σημαντικότητά τους.

- Οι web εφαρμογές μπορούν να εκτελεστούν σε οποιαδήποτε πλατφόρμα, ανεξάρτητα το λειτουργικό σύστημα, αρκεί να είναι συμβατό το πρόγραμμα περιήγησης ιστού.
- Όλοι οι χρήστες έχουν άμεση πρόσβαση στην ίδια έκδοση της εφαρμογής.
- Είναι προσβάσιμες από παντού, κάθε ώρα της ημέρας και ανεξαρτήτου συσκευής (π.χ. κινητό ή laptop).
- Είναι ασφαλείς σε μεγάλο επίπεδο, καθώς αναπτύσσονται σε servers, τους οποίους παρακολουθούν και συντηρούν συνήθως έμπειροι διαχειριστές.
- Δε χρειάζεται να τις αποθηκεύσεις στις προσωπικές σου συσκευές και συνήθως δεν απαιτούν μεγάλη υπολογιστική δύναμη.
- Είναι εύκολο να τις αναπτύξεις, αλλά και να τις επεκτείνεις στο μέλλον.
- Παρέχουν μειωμένο κόστος για τις επιχειρήσεις, καθώς ξοδεύεται λόγου χάρη, λιγότερος χρόνος στην τηλεφωνική εξυπηρέτηση πελατών, στη διαφήμιση και συνεχή ενημέρωση, ή στο έντυπο υλικό.

5.2 Web Frameworks

Σε αυτή την υποενότητα θα μιλήσουμε για τα Web frameworks και πιο συγκεκριμένα για το Django Web Framework, στο οποίο στηρίχθηκε η ανάπτυξη και υλοποίηση της παρούσας εφαρμογής. Ένα web framework είναι ένα εργαλείο λογισμικού, το οποίο είναι σχεδιασμένο, έτσι ώστε να υποστηρίζει την ανάπτυξη και τη λειτουργία διαδικτυακών εφαρμογών. Τα web frameworks παρέχουν έναν τυπικό (standard) τρόπο ανάπτυξης και υποστήριξης της εφαρμογής. Στοχεύουν στο να αυτοματοποιήσουν το “overhead”, το οποίο προκύπτει από κοινές διαδικασίες, οι οποίες λαμβάνουν μέρος κατά την ανάπτυξη της εφαρμογής. Αρκετά web frameworks παρέχουν βιβλιοθήκες για την πρόσβαση στη βάση δεδομένων με απλοϊκό τρόπο, υποστηρίζουν το “session management”⁴ και προωθούν την επαναχρησιμοποίηση κώδικα [43].

⁴Ένα web session είναι η διαδικασία της διασφάλισης πολλαπλών αιτήσεων σε μια υπηρεσία από τον ίδιο χρήστη ή την ίδια οντότητα.

5.2.1 Django Web Framework

Το Django είναι ένα υψηλού επιπέδου web framework, το οποίο στηρίζεται στη γλώσσα Python και ενθαρρύνει την ταχεία ανάπτυξη της εφαρμογής με σαφή σχεδιασμό. Παρέχει ένα σύνολο εργαλείων και λειτουργιών, τα οποία δίνουν λύση σε πολλές πτυχές, όπως σε θέματα ασφαλείας, στην προσπέλαση της βάσης δεδομένων, στη δρομολόγηση διευθύνσεων URL (URL routing), στην επεξεργασία των templates (template processing). Το web framework Django υποστηρίζει το MVC πρότυπο (Model-View-Controller pattern), με μια μικρή διαφορά, την οποία θα εξετάσουμε παρακάτω. Το MVC είναι ένα πρότυπο σχεδίασης λογισμικού για ανάπτυξη web εφαρμογών και αποτελείται από τα παρακάτω μέρη [38]:

- **Model:** Αποτελεί το κατώτερο επίπεδο του προτύπου (pattern), το οποίο είναι υπεύθυνο για τη διατήρηση και διαχείριση των δεδομένων της εφαρμογής. Ανταποκρίνεται στα αιτήματα, τα οποία έρχονται από το μέρος του “View”, όπως επίσης ανταποκρίνεται και στις εντολές τις οποίες στέλνει ο “Controller”.
- **View:** Αυτό το μέρος είναι υπεύθυνο για την απεικόνιση όλων ή ενός μέρους από τα δεδομένα της εφαρμογής. Πυροδοτείται από τις εντολές του “Controller”.
- **Controller:** Ο “Controller” είναι υπεύθυνος για να λαμβάνει τα αιτήματα του χρήστη, να ελέγχει αν είναι έγκυρα και ακολούθως να απαντάει σε αυτά. Σε αυτό το μέρος βρίσκεται ο κώδικας του λογισμικού, ο οποίος ελέγχει τις αλληλεπιδράσεις μεταξύ των “Model” και “View”.

Στο ακόλουθο σχήμα (Σχήμα 5.2) βλέπουμε ένα τυπικό σχεδιάγραμμα του προτύπου MVC.

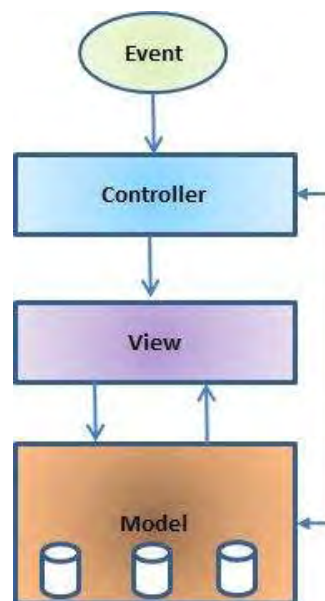


FIGURE 5.2: Ένα τυπικό σχεδιάγραμμα του μοντέλου MVC. [38]

Όπως προαναφέρθηκε το Django framework έχει μια ιδιαιτερότητα, στο MVC πρότυπο το οποίο υποστηρίζει, και αυτή ονομάζεται MVT (Model-View-Tempalte). Το πρότυπο MVT είναι ελαφρώς διαφορετικό από το MVC. Η διαφορά έγκειται στο ότι το ίδιο το Django framework φροντίζει για το τμήμα του “Controller”, αφήνοντας σε εμάς τον έλεγχο του “Template”. Το “Template” δεν είναι τίποτα άλλο, παρά ένα αρχείο HTML, το οποίο εμπεριέχει κώδικα με μορφή Django Template Language (DTL) και το οποίο είναι υπεύθυνο για τον τρόπο με τον οποίο θα παρουσιαστούν τα περιεχόμενα στην οθόνη του χρήστη [39]. Το παρακάτω διάγραμμα (Σχήμα 5.3) απεικονίζει τον τρόπο αλληλεπίδρασης καθενός από τα συστατικά μέρη του προτύπου MVT μεταξύ τους για την εξυπηρέτηση ενός αιτήματος χρήστη, ενώ το ακόλουθο σχήμα (Σχήμα 5.4), δείχνει τις αλληλεπιδράσεις μεταξύ των στοιχείων model,view,template.

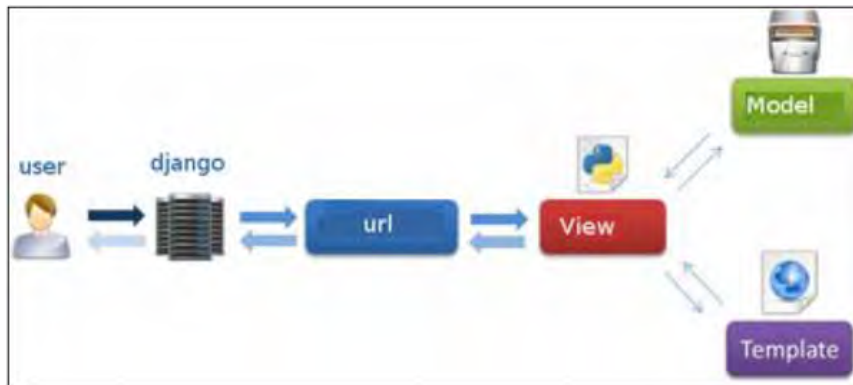


FIGURE 5.3: Ένα τυπικό σχεδιάγραμμα του μοντέλου MVT. [39]

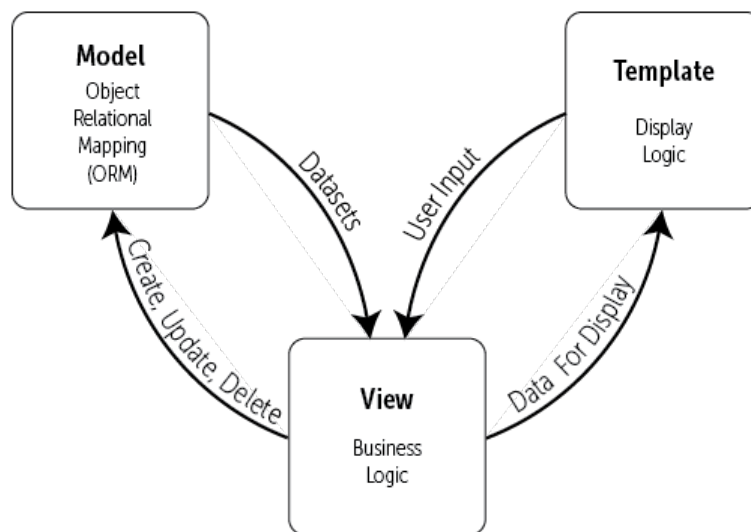


FIGURE 5.4: Οι αλληλεπιδράσεις μεταξύ του model, view και template στο MVT πρότυπο. Model: Το Object-Relational Mapping παρέχει τη διεπαφή για τη βάση δεδομένων της εφαρμογής. Template: παρέχει τη “display logic” και είναι η διεπαφή ανάμεσα στον χρήστη και της Django εφαρμογής. View: διαχειρίζεται την επεξεργασία των δεδομένων της εφαρμογής, τη λογική της εφαρμογής (application logic) και των μηνυμάτων [9]

Στη συνέχεια, θα αναλύσουμε συνοπτικά ορισμένα χαρακτηριστικά του Django Framework.

Django Models

Το “Django Framework” χρησιμοποιεί το ORM (Object-Relational Mapping). Το ORM είναι μία ισχυρή τεχνική προγραμματισμού, η οποία καθιστά πολύ πιο εύκολη την εργασία με τα δεδομένα και τις σχεσιακές βάσεις δεδομένων. Οι περισσότερες κοινές βάσεις δεδομένων προγραμματίζονται με κάποια μορφή δομημένης γλώσσας ερωτημάτων (SQL), ωστόσο κάθε βάση δεδομένων υλοποιεί την SQL με τον δικό της τρόπο. Η SQL κάποιες φορές δύναται να γίνει δύσκολη ή και περίπλοκη και εδώ είναι που κερδίζει έδαφος το ORM, αφού ο προγραμματιστής δε χρειάζεται να γνωρίζει τη δομή της βάσης ή να γράφει πολύπλοκη SQL για να χειρίζεται και να ανακτά τα δεδομένα από αυτή [9].

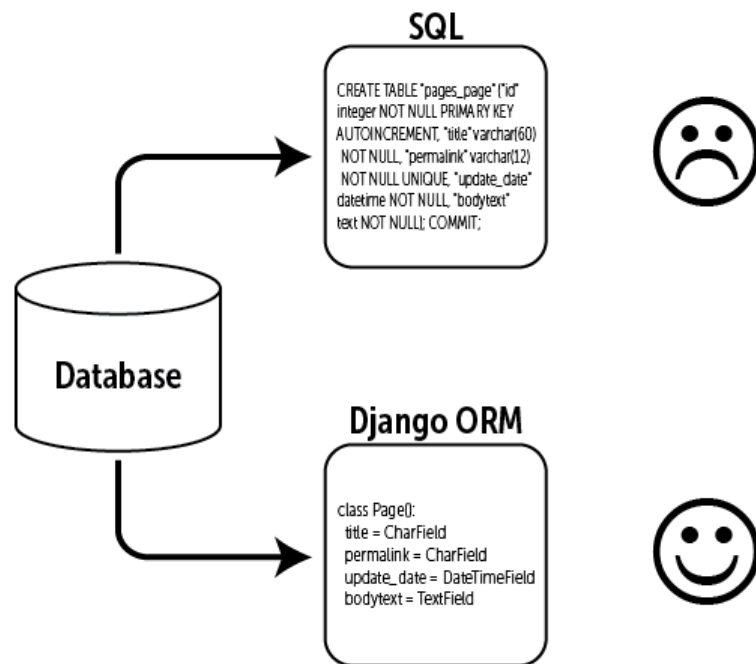


FIGURE 5.5: Η πολύπλοκη πρόσβαση στη βάση δεδομένων και η διαχείρισή της μέσω SQL γίνεται αρκετά πιο εύκολη με το ORM του Django framework. [9]

Το ORM παρέχει μία αντιστοίχιση μεταξύ ενός αντικειμένου (object) και της βάσης δεδομένων (the ‘O’ in ORM). Από προγραμματιστικής άποψης τα μοντέλα τα ορίζουμε μέσα σε ένα python αρχείο με όνομα “models.py”. Όταν δημιουργούμε ένα μοντέλο (model), ορίζοντας μία κλάση python, δημιουργείται αυτόματα ένας πίνακας στη βάση δεδομένων (Σχήμα 5.6). Ακόμα, πολλά μοντέλα (πίνακες στη ΒΔ) μπορούν να δημιουργηθούν και να συνδεθούν μεταξύ τους (the ‘R’ in ORM) μέσω ενός “ξένου κλειδιού” (foreign key). Με αυτόν τον τρόπο μπορούμε να δημιουργήσουμε συσχετίσεις μεταξύ των δεδομένων μας και να αποφύγουμε τον

πιθανό πλεονασμό στη βάση. Βλέπουμε λοιπόν, ότι όσον αφορά στο κομμάτι της βάσης δεδομένων τα πράγματα απλοποιούνται πολύ χρησιμοποιώντας το Django framework, διατηρώντας παράλληλα μία σωστή και σαφής δομή [9].

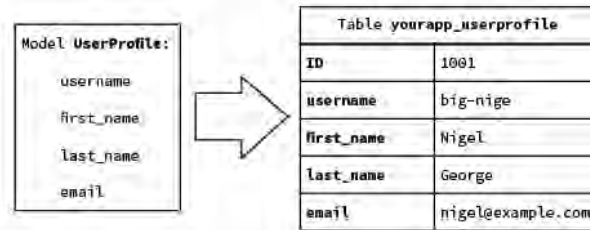


FIGURE 5.6: Η κλάση η οποία ορίζουμε στην python, μετατρέπεται αυτόματα σε έναν πίνακα στη βάση δεδομένων. [9]

Django Templates

Τα “Django Templates” είναι αρχεία κειμένου σχεδιασμένα για να διαχωρίζουν τα δεδομένα της εφαρμογής από τον τρόπο παρουσίασής τους. Όπως είχε προαναφερθεί είναι αρχεία HTML για την παρουσίαση των δεδομένων της εφαρμογής σε ένα πρόγραμμα περιήγησης ιστού, ωστόσο δεν περιορίζονται μόνο στην HTML, αλλά μπορούν να χρησιμοποιηθούν για την απόδοση πολλών διαφορετικών μορφών κειμένου (text formats).

Η σχεδίαση των “Django Templates” βασίζεται σε τρεις βασικές αρχές [9].

1. Το “template” πρέπει να διαχωρίζει τη “λογική” του προγράμματος από τον “σχεδιασμό”. Στις περισσότερες περιπτώσεις το “web design” και το “web programming” γίνεται από διαφορετικά άτομα, με συνέπεια αυτός ο διαχωρισμός να καθίσταται αναγκαίος. Χρησιμοποιείται λοιπόν, μία μορφή κώδικα, την οποία ονομάζουμε “plain-text scripting language” και η οποία μέσω των “tags” παρέχει τη λογική με την οποία θα παρουσιαστούν τα δεδομένα (presentation logic). Έτσι ο σχεδιαστής (designer) δε χρειάζεται να γνωρίζει τίποτα όσον αφορά στο “Django Framework”, ώστε να δημιουργήσει ένα “template”.
2. Όχι στην επανάληψη και τον πλεονασμό – DRY (Don’t Repeat Yourself). Εδώ χρησιμοποιούμε αυτό που ονομάζουμε “template inheritance”. Άς δούμε τί σημαίνει. Άς υποθέσουμε ότι έχουμε μία εφαρμογή, η οποία αποτελείται από πολλές ιστοσελίδες, οι οποίες όμως έχουν όλες ένα κοινό layout. Για παράδειγμα, έχουν κοινό header, footer και top navigation και στην καθεμία αλλάζει μόνο το περιεχόμενο, το οποίο προβάλλει. Σε αυτή την περίπτωση, αν δε χρησιμοποιήσουμε το “template inheritance” θα χρειαστεί να γράψουμε τον ίδιο κώδικα επανειλημμένα, με συνέπεια το έργο της συντηρησιμότητας

στο μέλλον να γίνεται χρονοβόρο και δύσκολο. Για να λύσουμε αυτό το πρόβλημα λοιπόν, μπορούμε να δημιουργήσουμε ένα “parent template”, το οποίο θα περιέχει όλα τα κοινά στοιχεία, τα οποία θέλουμε να εμφανίζονται οι ιστοσελίδες της εφαρμογής (header, footer, top navigation). Εν συνεχεία, τα “child templates” τα οποία θα δημιουργήσουμε μπορούν να κληρονομήσουν τα κοινά χαρακτηριστικά από το “parent”, έχοντας ταυτόχρονα το δικό τους ξεχωριστό περιεχόμενο. Στην ακόλουθη εικόνα (Σχήμα 5.7) φαίνεται η διαδικασία που περιγράφουμε.

- Τέλος, το σύστημα των “templates” πρέπει να είναι ασφαλές. Δεν πρέπει να εκτελείται κώδικας σε αυτό το σημείο. Τα “Django Template tags” παρέχουν μόνο “display logic”, δηλαδή την λογική για το τί θα εμφανιστεί στον χρήστη. Πιο συγκεκριμένα, μέσα στα templates μπορείς να εμφανίσεις μεταβλητές (όπως π.χ. κάποιο όνομα χρήστη), να επιλέξεις ποιά περιεχόμενα θα απεικονίσεις βασιζόμενος σε λογικές πράξεις (if-else if statements), να διατρέξεις μία λίστα δεδομένων. Δεν είναι δυνατό να εκτελέσεις κώδικα python, να ορίσεις τιμή σε μία μεταβλητή ή να εκτελέσεις κάποια πιο προηγμένη λογική (advanced logic). Τέλος, τα “Django Template tags” παρέχουν και άλλα χαρακτηριστικά, τα οποία καθιστούν τα templates ασφαλή, όπως το Cross-Site Scripting και Cross-Site Request Forgery protection.

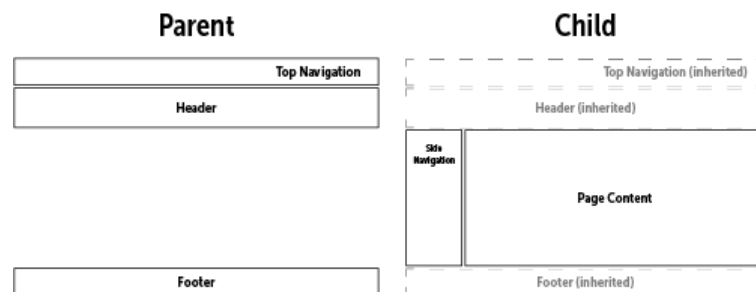


FIGURE 5.7: Το “child template” προσθέτει μόνο τη δομή και τα περιεχόμενα, όλα τα κοινά στοιχεία κληρονομούνται από το “parent template”. [9]

Django Views

Τα Django Views βρίσκονται συνήθως μέσα σε ένα αρχείο, το οποίο ονομάζεται “views.py”. Κάθε “view” εκτελεί μία συγκεκριμένη λειτουργία και σχετίζεται με ένα μοναδικό “template”. Από προγραμματιστικής άποψης ένα “view” μπορεί να υλοποιηθεί, είτε μέσω μίας συνάρτησης, είτε μέσω μίας κλάσης. Αντλεί δεδομένα από τη βάση δεδομένων της εφαρμογής (ή από εξωτερική πηγή δεδομένων), τα επεξεργάζεται αν χρειαστεί και στη συνέχεια τα παραδίδει σε ένα “template”. Στο μέρος του “view” μπορούν να ληφθούν αποφάσεις σχετικά με το ποιά δεδομένα θα σταλούν στο “template”, ανάλογα είτε με την είσοδο την οποία έχει δώσει

ο χρήστης, είτε με κάποια άλλη λογική (business logic), η οποία ορίζει κάποιους κανόνες για τη δημιουργία, αποθήκευση, επεξεργασία και τροποποίηση των δεδομένων [9].

URL configuration

Έχοντας λοιπόν ήδη τα “views” και “templates” έτοιμα στη διάθεσή μας, θα πρέπει να έχουμε και έναν τρόπο για να μπορούμε να περιηγηθούμε στις ιστοσελίδες της εφαρμογής. Η περιήγηση γίνεται με τα URL, όπως σε κάθε ιστοσελίδα. Δίνοντας ο χρήστης ένα URL, ένα αίτημα αποστέλλεται στο Django Framework για την συγκεκριμένη διεύθυνση. Μόλις το Django λάβει αυτό το αίτημα, πρέπει να είναι σε θέση να αποφασίσει ποιό από τα διαθέσιμα “views” θα χρησιμοποιήσει για να ικανοποιήσει το αίτημα του χρήστη. Είναι υποχρέωση του προγραμματιστή να δημιουργήσει μία αντιστοίχιση ανάμεσα στα URL και τα “views”, έτσι ώστε κάθε φορά να απαντάμε στο αίτημα του χρήστη κατάλληλα. Από προγραμματιστικής άποψης η διαδικασία του “url configuration - mapping” γίνεται σε ένα python αρχείο εν ονόματι “urls.py”, μέσα στο οποίο δηλώνουμε το url και το όνομα του view, το οποίο θέλουμε να αντιστοιχίσουμε (Σχήμα 5.8) [9].

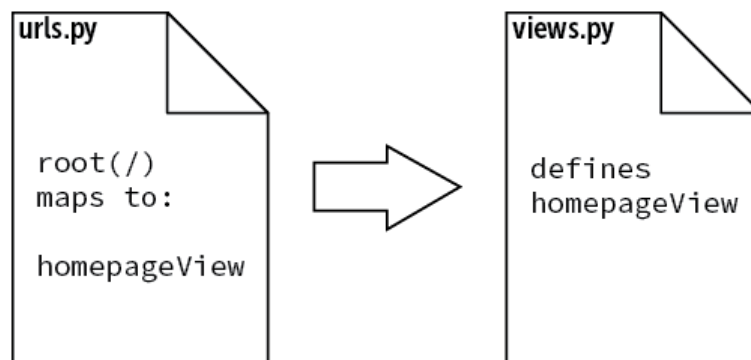


FIGURE 5.8: Η αντιστοίχιση των URLs σε συγκεκριμένα views λαμβάνει χώρα μέσα στο αρχείο urls.py [9]

Ο κώδικας της web εφαρμογής βρίσκεται στο github, στο παρακάτω [link](#)

Chapter 6

Επίλογος

Από το 2001 και μετά, αφού έγινε ευρέως γνωστό ότι τα microRNAs υπάρχουν σε αφθονία στον οργανισμό μας και παίζουν ένα πολύ σημαντικό ρόλο, καθώς ρυθμίζουν πολλές ζωτικές λειτουργίες, έχει ξεκινήσει μία τεράστια προσπάθεια, με στόχο την καλύτερη κατανόηση της βιογένεσής τους, της λειτουργίας τους, των μηχανισμών δράσης τους. Έρευνες που διεξάγονται, λαμβάνουν υπόψη τους τα microRNAs, ως θεραπευτικούς στόχους, αλλά και ως βιοδείκτες για την ανίχνευση, την πρόγνωση και τη διάγνωση διάφορων ασθενειών. Σε όλη αυτή την προσπάθεια, η οποία λαμβάνει χώρα τα τελευταία περίπου 15 χρόνια, σημαντικό ρόλο διαδραματίζουν τα υπολογιστικά εργαλεία, τα οποία αναπτύσσονται στον τομέα της βιοπληροφορικής και τα οποία αξιοποιώντας τεχνολογίες αιχμής παρέχουν σημαντική βοήθεια στο έργο των ερευνητών βιολόγων. Παράλληλα όμως και online εργαλεία, όπως οι Genome Browsers βοηθούν στην απεικόνιση της βιολογικής πληροφορίας και κατ'επέκταση στην καλύτερη κατανόησή της και στην εξαγωγή χρήσιμων συμπερασμάτων. Η παρούσα διπλωματική εργασία αποτέλεσε την αφορμή για την υλοποίηση ενός genome browser, ο οποίος είναι σχεδιασμένος ειδικά για την αναπαράσταση των microRNAs και των προβλεπόμενων στόχων τους, με βάση τις προβλέψεις του προγράμματος DIANA-microT-CDS. Στο μέλλον, σχεδιάζουμε να επεκτείνουμε τις δυνατότητες του genome browser μας, εντάσσοντας πειραματικά αποδεδειγμένα δεδομένα από άλλες βιολογικές βάσεις, όπως για παράδειγμα την DIANA-TarBase. Ακόμα, ο χρήστης θα έχει τη δυνατότητα να διοχετεύει τον genome browser με δικά του βιολογικά δεδομένα, όπως δεδομένα σχετικά με SNPs και να παρατηρεί γραφικά τη συνδυαστική δράση των microRNAs σε συγκεκριμένες περιοχές των γονιδίων. Τέλος, θα είναι δυνατή η συνδυαστική αναζήτηση με σκοπό την εμφάνιση συσχετίσεων και η αναζήτηση με βάση ασθενειών. Εν κατακλείδι, η μελέτη και η κατανόηση των microRNAs είναι πολύ σημαντική, έτσι ώστε να μπορέσουμε να αποσαφηνίσουμε εξ'ολοκλήρου το ρόλο και τη λειτουργία τους στους οργανισμούς των ειδών, τους μηχανισμούς ρύθμισης τους οποίους χρησιμοποιούν, καθώς επίσης και τα γονίδια τα οποία στοχεύουν, με στόχο να μπορούμε να τα χρησιμοποιήσουμε πρακτικά για την ρύθμιση των βιολογικών μας διεργασιών, αλλά και για την αντιμετώπιση ασθενειών.

Bibliography

- [1] Helwak Aleksandra and Tollervey David. “Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH)”. In: *Nature* 9 (2014). [View online](#), pp. 711–728.
- [2] Kozomara Ana and Griffiths-Jones Sam. “miRBase: annotating high confidence microRNAs using deep sequencing data”. In: *Nucleic Acids Research* 42 (2013). [View online](#), pp. D68–D73.
- [3] Kozomara Ana and Griffiths-Jones Sam. “miRBase: integrating microRNA annotation and deep-sequencing data”. In: *Nucleic Acids Research* 39 (2010). [View online](#), pp. D152–D157.
- [4] Riffo-Campos Ángela L., Riquelme Ismael, and Brebi-Mieville Priscilla. “Tools for Sequence-Based miRNA Target Prediction: What to Choose?” In: *Molecular Sciences* 17.12 (2016). [View online](#).
- [5] Davis Brandi N and Hata Akiko. “Regulation of MicroRNA Biogenesis: A miRiad of mechanisms”. In: *Cell Communication and Signaling* (2009). [View online](#).
- [6] Yu Chan Chi, E. Lawrence Charles, and Ding Ye. “Structure clustering features on the Sfold Web server”. In: *Oxford Journals* 21 (2005). [View online](#), pp. 3926–3928.
- [7] “Common features of microRNA target prediction tools”. Oct. 2018. URL: <https://www.slideserve.com/riordan-nyhan/common-features-of-microRNA-target-prediction-tools>.
- [8] Zerbino Daniel et al. “Ensembl 2018”. In: *Oxford Academic* 46 (2018). [View online](#), pp. D754–D751.
- [9] DjangoBook. “Django overview”. Oct. 2018. URL: https://www.tutorialspoint.com/django/django_overview.htm.
- [10] Embl-ebi. “Download data with BioMart”. Oct. 2018. URL: <https://www.ebi.ac.uk/training/online/course/ensembl-browsing-chordate-genomes/download-data-biomart>.
- [11] Ensembl. “Browsing genomes”. Oct. 2018. URL: <https://www.ebi.ac.uk/training/online/course/ensembl-browsing-chordate-genomes>.
- [12] Ensembl. “What is Ensembl”. Oct. 2018. URL: <https://www.ebi.ac.uk/training/online/course/ensembl-browsing-chordate-genomes/what-ensembl>.

- [13] Exiqon. "What are microRNAs". Oct. 2018. URL: <http://www.exiqon.com/what-are-microRNAs>.
- [14] Wahida Fazli et al. "MicroRNAs: Synthesis, mechanism, function, and recent clinical trials". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803 (2010). View online, pp. 1231–1243.
- [15] Georgakilas Georgios. "Pattern recognition for non-coding RNA promoters". In: *University of Thessaly. Department of Electrical and Computer Engineering*. (2015). View online.
- [16] Vlachos Ioannis S. et al. "DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions". In: *Oxford Journals* 43 (2014). View online, pp. D153–D159.
- [17] S. Vlachos Ioannis and Artemis G. Hatzigeorgiou. "Online resources for miRNA analysis". In: *Clinical Biochemistry* 46 (2013). View online, pp. 879–900.
- [18] DIANA lab. "Noncoding RNAs and their targets". Oct. 2018.
- [19] DIANA lab. "Tools and databases for analysis of non-coding RNAs". Oct. 2018.
- [20] MacFarlane Leigh-Ann and Murphy Paul R. "MicroRNA: Biogenesis, Function and Role in Cancer". In: *Current Genomics* 11 (2010). View online, pp. 537–561.
- [21] Mullany Lila E. et al. "MicroRNA Seed Region Length Impact on Target Messenger RNA Expression and Survival in Colorectal Cancer". In: *PLOS Journal* 11.4 (2016). View online.
- [22] Paraskevopoulou Maria. "Αλγόριθμοι για την υπολογιστική ανάλυση της λειτουργίας των μη κωδικών μεταγραφών". In: *University of Thessaly. Department of Electrical and Computer Engineering*. (2016). View online, pp. 6, 30.
- [23] Hafner Markus et al. "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP". In: *Cell* 141 (2010). View online, pp. 129–141.
- [24] Reczko Martin et al. "Functional microRNA targets in protein coding sequences". In: *Oxford Journals* 28 (2012). View online, pp. 771–776.
- [25] Selbach Matthias et al. "Widespread changes in protein synthesis induced by microRNAs". In: *Nature* 455 (2008). View online, pp. 58–63.
- [26] miRBase. "What do the miRNA names/identifiers mean". Oct. 2018. URL: <http://www.mirbase.org/help/nomenclature.shtml>.
- [27] Craig Nancy et al. "Molecular Biology Principles of Genome Function". In: *Oxford University Press* (2014). View online.
- [28] NCBI. "Programs and Activities". Oct. 2018. URL: <https://www.ncbi.nlm.nih.gov/home/about/programs>.

- [29] Alexiou Panagiotis et al. “Lost in translation: an assessment and perspective for computational microRNA target identification”. In: *Oxford Journals* 25 (2009). [View online](#), pp. 3049–3055.
- [30] ReasearchGate. “*microRNA Mechanisms*”. Oct. 2018. URL: https://www.researchgate.net/figure/MicroRNA-biogenesis-and-mechanism-of-action-See-text-for-details_fig2_273384811.
- [31] Kishore Shivendra et al. “A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins”. In: *Nature* 8 (2011). [View online](#), pp. 559–564.
- [32] Stackpath. “*What is a Web Application*”. Oct. 2018. URL: <https://www.maxcdn.com/one/visual-glossary/web-application/>.
- [33] Grosswendt Stefanie et al. “Unambiguous Identification of miRNA:Target Site Interactions by Different Types of Ligation Reactions”. In: *Cell* 54 (2014). [View online](#), pp. 1042–1054.
- [34] Wook Chi Sung et al. “Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps”. In: *Nature* 460 (2009). [View online](#), pp. 479–486.
- [35] “*TargetsScan*”. Oct. 2018. URL: http://www.targetsScan.org/vert_72/.
- [36] Tomankov Tereza, Petrek Martin, and Kriegova Eva. “Involvement of microRNAs in physiological and pathological processes in the lung”. In: *Respiratory Research* 11 (2010). [View online](#).
- [37] Vergoulis Thanasis et al. “TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support”. In: *Oxford Journals* 40 (2012). [View online](#), pp. D222–D229.
- [38] Tutorialspoint. “*Basic MVC Architecture*”. Oct. 2018. URL: http://www.tutorialspoint.com/struts_2/basic_mvc_architecture.htm.
- [39] Tutorialspoint. “*Django overview*”. Oct. 2018. URL: https://www.tutorialspoint.com/django/django_overview.htm.
- [40] Ambros Victor et al. “A uniform system for microRNA annotation”. In: *RNA Journal* 9 (2003). [View online](#), pp. 277–279.
- [41] Wikipedia. “*Messenger RNA*”. Oct. 2018. URL: https://en.wikipedia.org/wiki/Messenger_RNA.
- [42] Wikipedia. “*National Center for Biotechnology Information*”. Oct. 2018. URL: https://en.wikipedia.org/wiki/National_Center_for_Biotechnology_Information.
- [43] Wikipedia. “*Web Framework*”. Oct. 2018. URL: https://en.wikipedia.org/wiki/Web_framework.
- [44] Yourgenome. “*From DNA to protein*”. Oct. 2018. URL: <https://www.yourgenome.org/search/results/from%20dna%20to%20protein>.