

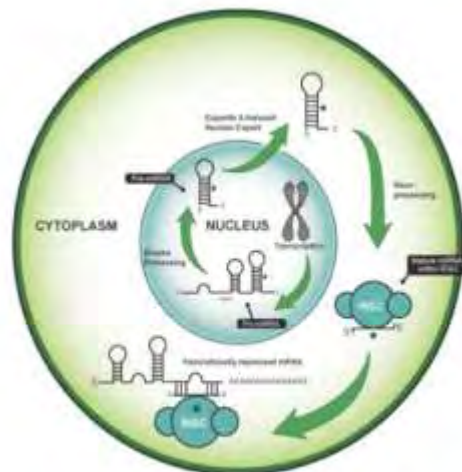
**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών  
Τηλεπικοινωνιών & Δικτύων



Διπλωματική εργασία

**ΠΡΟΒΛΕΨΗ ΓΟΝΙΔΙΩΝ ΣΤΟΧΩΝ ΓΙΑ ΙΪΚΑ  
ΜΙRNA ΣΕ ΙΟΥΣ ΚΑΙ ΘΗΛΑΣΤΙΚΑ**



**Τζεβελεκίδης Κωνσταντίνος**

Επιβλέποντες καθηγητές: Χούστης Ηλίας  
Καθηγητής Πανεπιστημίου Θεσσαλίας  
Χατζηγεωργίου Άρτεμις  
Συμβασιούχος 407

Βόλος, 2011

Στην οικογένεια και στους φίλους μου,

## ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία αποτελεί την ολοκλήρωση των σπουδών μου στο τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου Θεσσαλίας. Η συγκεκριμένη εργασία αποτελεί προϊόν εκτενούς αναζήτησης και έρευνας και πραγματοποιήθηκε με την υποστήριξη και καθοδήγηση συγκεκριμένων ανθρώπων, τους οποίους θα ήθελα να ευχαριστήσω θερμά.

Αρχικά, θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου, κυρία Άρτεμις Χατζηγεωργίου και κύριο Ηλία Χούστη για την εμπιστοσύνη που μου έδειξαν, αναθέτοντάς μου τη συγκεκριμένη εργασία. Θα ήθελα να ευχαριστήσω, ακόμη, το Μανώλη Μαραγκάκη για την καθοδήγηση και την ουσιαστική συμβολή του στην περάτωση της εργασίας, καθώς και τον Παναγιώτη Αλεξίου για τις παρεμβάσεις του.

Ο Δρ. Στ. Τσόχας με βοήθησε να κατανοήσω την περιπλοκότητα των βιολογικών εννοιών. Επίσης, θα ήθελα να ευχαριστήσω τον πολύ καλό φίλο και συνάδελφο Σπέντζα Θεοχάρη, για την υποστήριξή του σε κάθε βήμα και τις επισημάνσεις του στην εξέλιξη της διπλωματικής μου εργασίας.

Η μορφοποίηση της παρούσας διπλωματικής έγινε με τη βοήθεια της αδερφής μου, Δρ Β. Τζεβελεκίδη.

Τέλος, οφείλω να ευχαριστήσω την οικογένεια και τους φίλους μου για την υποστήριξη και τη βοήθεια τους, καθώς βρίσκονταν πάντα δίπλα μου σε όλη τη διάρκεια των σπουδών μου και με βοήθησαν να ανταπεξέλθω και να φέρω σε πέρας τις ακαδημαϊκές μου σπουδές.

Τζεβελεκίδης Κωνσταντίνος

Βόλος, 2011

## ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η υλοποίηση του πρώτου τμήματος του προγράμματος αναγνώρισης γονιδιακών στόχων ιικών microRNAs σε ιούς και θηλαστικά. Τα microRNAs είναι μικρά μόρια RNA μέσου μήκους 22 νουκλεοτιδίων, τα οποία παρουσιάζουν ιδιαίτερο ενδιαφέρον τα τελευταία χρόνια και έχει αποδειχθεί ότι διαδραματίζουν σημαντικό ρόλο σε πολλές βιολογικές λειτουργίες του κυττάρου. Το ενδιαφέρον εστιάζεται κυρίως στον εντοπισμό των στόχων τους, δηλαδή τα μόρια που κωδικοποιούν συγκεκριμένες πρωτεΐνες (mRNA). Η δράση των microRNAs ενδέχεται να μεταβάλλει δραματικά τα επίπεδα έκφρασης των παραπάνω πρωτεϊνών, γεγονός που έχει επίδραση σε διάφορες κυτταρικές λειτουργίες, ενώ μπορεί ακόμα και να αποτελέσει τη βάση εκδήλωσης παθήσεων, όπως ο καρκίνος. Για το σκοπό αυτό, λοιπόν, δημιουργήθηκαν τρία προγράμματα, τα οποία εντοπίζουν τους ιούς, βρίσκουν τα ιικά microRNAs και υπολογίζουν τα γονίδια των ιών.

Λέξεις-κλειδιά: μόρια microRNA, βιολογία, γονίδια, ιικά microRNA, προγραμματισμό

## **ABSTRACT**

This thesis attempted to develop the first part of a program used for the recognition of gene targets of viral microRNAs in viruses and mammals. MicroRNAs are short molecules of RNA with an average length of 22 nucleotides and have been proven to play an important role in many biological functions of cells. Scientific interest is mainly focused on identifying microRNA targets, that is the molecules that encode specific proteins (mRNA). MiRNAs activity can alter the expression of such proteins significantly, affecting, thus, various cellular functions, while it can also trigger the development of various diseases, such as cancer. Therefore, three programs have been created that detect viruses and viral microRNAs and locate the genes of viruses.

Keywords: microRNA, biology, genes, targets microRNA, viral microRNA, programming

## ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΠΛΩΜΑΤΙΚΗΣ

Η παρούσα διπλωματική εργασία και το λογισμικό που αναπτύχθηκε μέσα στο πλαίσιο της εκπόνησής της, έχει ως στόχο να συνεισφέρει στον τομέα της βιοϊατρικής επιστήμης γύρω από το αντικείμενο των microRNAs. Τα προγράμματα που δημιουργήθηκαν αποτελούν ένα ευρύτερο κομμάτι της πρόβλεψης γονιδιακών στόχων ιικών microRNAs σε ιούς και θηλαστικά, η οποία δεν έχει επιχειρηθεί και επιτευχθεί μέχρι τώρα. Μετά την ολοκλήρωσή του, το συγκεκριμένο εγχείρημα θα αποτελέσει ένα επιπλέον εργαλείο στα χέρια των ερευνητών, προσφέροντάς τους τη δυνατότητα να εξάγουν χρήσιμα συμπεράσματα για την επιρροή των microRNAs σε σοβαρές ασθένειες, όπως ο καρκίνος.

Η εκπόνηση της διπλωματικής εργασίας πραγματοποιήθηκε με τη βοήθεια του Ερευνητικού Κέντρου Βιοϊατρικών Επιστημών «Αλέξανδρος Φλέμινγκ», το οποίο έχει αναπτύξει αλγόριθμο για την πρόβλεψη γονιδιακών στόχων microRNAs.

## ΟΡΓΑΝΩΣΗ ΚΕΙΜΕΝΟΥ

Η παρούσα διπλωματική εργασία έχει οργανωθεί σε τέσσερα κεφάλαια, τα οποία περιγράφουν βασικές βιολογικές έννοιες, που βοηθούν στην κατανόηση του στόχου και της υλοποίησης της εργασίας και των πιθανών μελλοντικών επεκτάσεων.

Στο **1ο κεφάλαιο** περιλαμβάνονται ορισμοί βασικών βιολογικών εννοιών (νουκλεϊκά οξέα (DNA, RNA), κεντρικό δόγμα της μοριακής βιολογίας), των οποίων η γνώση είναι απαραίτητη, προκειμένου να γίνει αντιληπτή η αξία των προγραμμάτων που δημιουργήθηκαν.

Στο **2ο κεφάλαιο** εισάγεται η έννοια των microRNAs, η πρόβλεψη των γονιδίων στόχων των microRNAs (συγκεκριμένα των ιικών) και η αξία της πρόβλεψης στόχων για τον άνθρωπο.

Στο **3ο κεφάλαιο** ακολουθεί η ανάλυση της πρόβλεψης γονιδιακών στόχων ιικών microRNAs, με βήματα που αποτελούν το καθένα ξεχωριστά προγράμματα.

Στο **4ο κεφάλαιο** συνοψίζονται οι στόχοι που επιτεύχθηκαν και παρατίθενται οι δυνατότητες για μελλοντικές επεκτάσεις.

## ΛΙΣΤΑ ΕΙΚΟΝΩΝ

### **ΚΕΦΑΛΑΙΟ 1:**

Εικ. 1.1.: Το μοντέλο της διπλής έλικας.....	11
Εικ. 1.2: Απεικόνιση του RNA .....	12
Εικ. 1.3: Τρισδιάστατη απεικόνιση μίας πρωτεΐνης .....	13
Εικ. 1.4: Το κεντρικό δόγμα της μοριακής βιολογίας.....	14
Εικ. 1.5: Η διαδικασία ωρίμανσης ενός μορίου mRNA .....	16
Εικ. 1.6: Ο γενετικός κώδικας.....	17
Εικ. 1.7: Δημιουργία πρωτεΐνης από DNA .....	18

### **ΚΕΦΑΛΑΙΟ 2:**

Εικ. 2.1: Λειτουργία των μορίων microRNAs.....	19
Εικ. 2.2: MicroRNA ανάλυση σε καρκινικούς όγκους .....	22
Εικ. 2.3: MicroRNAs λειτουργώντας ως ογκογονίδια και ογκοκατασταλτικά.....	23
Εικ. 2.4: PCR (Polymerase Chain Reaction).....	25
Εικ. 2.5: Microarrays .....	27
Εικ. 2.6: SILAC (Stable Isotope Labeling with Amino acids in Cell culture) .....	28
Εικ. 2.7: Φόρμα αναζήτησης του προγράμματος Diana MicroT 3.0 .....	30
Εικ. 2.8: Αποτελέσματα του Diana microT 3.0.....	31
Εικ. 2.9: EIMMo .....	32
Εικ. 2.10: Miranda .....	33
Εικ. 2.11: MiRBase.....	35
Εικ. 2.12: PicTar.....	36
Εικ. 2.13: PITA.....	37
Εικ. 2.14: RNA22.....	38
Εικ. 2.15: Target Scan.....	39

### **ΚΕΦΑΛΑΙΟ 3:**

Εικ. 3.1: Χαρακτηριστικά υπολογιστικού συστήματος που χρησιμοποιήθηκε .....	41
Εικ. 3.2: Αρχείο 'organisms.txt'.....	42
Εικ. 3.3: Αποτελέσματα πρώτου προγράμματος - αναγνωριστικά ιών.....	43
Εικ. 3.4: Αρχείο σε μορφή fasta.....	44
Εικ. 3.5: Αποτελέσματα δεύτερου προγράμματος.....	46
Εικ. 3.6: Αρχείο σε μορφή genbank.....	49
Εικ. 3.7: Αποτελέσματα τρίτου προγράμματος.....	50



## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>10</b>
1.1 ΝΟΥΚΛΕΪΚΑ ΟΞΕΑ ΚΑΙ ΠΡΩΤΕΪΝΕΣ.....	10
1.2 ΤΟ ΚΕΝΤΡΙΚΟ ΔΟΓΜΑ ΤΗΣ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ.....	14
<b>2. MICRORNA .....</b>	<b>19</b>
2.1 ΕΙΣΑΓΩΓΗ ΣΤΑ MICRORNAS (MIRNAS).....	19
2.2 ΠΡΟΒΛΕΨΗ ΣΤΟΧΩΝ MICRORNAS (TARGET PREDICTION) .....	23
2.2.1 <i>DIANA-microT</i> .....	30
2.2.2 <i>EIMMo</i> .....	32
2.2.3 <i>miRanda</i> .....	32
2.2.4 <i>miRBase</i> .....	34
2.2.5 <i>Pictar</i> .....	36
2.2.6 <i>PITA</i> .....	37
2.2.7 <i>RNA 22</i> .....	38
2.2.8 <i>TargetScan 5.2</i> .....	39
<b>3. ΑΝΑΛΥΣΗ ΚΑΙ ΤΕΧΝΙΚΕΣ ΑΠΑΙΤΗΣΕΙΣ.....</b>	<b>40</b>
3.1 ΠΕΡΙΓΡΑΦΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΣΥΣΤΗΜΑΤΟΣ.....	40
3.2 ΑΝΑΛΥΣΗ ΠΡΟΓΡΑΜΜΑΤΩΝ.....	41
<b>4. ΕΠΙΛΟΓΟΣ .....</b>	<b>51</b>
4.1 ΣΥΝΟΨΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ.....	51
4.2 ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ.....	52

## 1. ΕΙΣΑΓΩΓΗ

### 1.1 Νουκλεϊκά οξέα και πρωτεΐνες

Αρχικά, τα νουκλεϊκά οξέα ανακαλύφθηκαν το 1869 από τον Φρίντριχ Μίσερ, ο οποίος ανακάλυψε σε πυρήνες κυττάρων την ύπαρξη μίας ουσίας με συγκεκριμένη όξινη αντίδραση, την οποία ονόμασε νουκλεΐνη (από το λατινικό *nucleus*, που σημαίνει πυρήνας), ενώ το 1889 ο μαθητής του Ρίτσαρντ Άλτμαν την μετονόμασε σε νουκλεϊκό οξύ. Τα πιο κοινά νουκλεϊκά οξέα είναι το δεσοξυριβονουκλεϊκό οξύ (DNA) και το ριβονουκλεϊκό οξύ (RNA) (Alberts *et al.* 2000).

Μετά το πέρας μιας σειράς επιστημονικών ερευνών που διήρκησαν χρόνια ανακαλύφθηκε ότι το DNA (Εικ. 1.1) είναι ο φορέας της γενετικής πληροφορίας. Ενώ είχε πιστοποιηθεί από το 1869 η ύπαρξή του στον πυρήνα των κυττάρων, μόλις στα μέσα του 20ου αιώνα ξεκίνησαν οι ερευνητές να αντιλαμβάνονται ότι μπορεί να αποθηκεύει γενετικές πληροφορίες. Το 1953 οι Τζέιμς Γουάτσον (J.Watson) και Φράνσις Κρικ (F.Crick), δύο ερευνητές που εργάζονταν στο πανεπιστήμιο του Καίμπριτζ, παρουσίασαν ένα μοντέλο της δομής του DNA, που ονομάστηκε “*μοντέλο της διπλής έλικας*”. Σύμφωνα με το μοντέλο αυτό το μόριο του DNA έχει τα εξής χαρακτηριστικά:

- Αποτελείται από δύο πολυνουκλεοτιδικές αλυσίδες σε μορφή δύο αντιτακτών κλώνων, που σχηματίζουν δεξιόστροφη διπλή έλικα.
- Οι αζωτούχες βάσεις (πρωτεϊνικές) κάθε κλώνου είναι κάθετες ως προς τον άξονα του μορίου και προεξέχουν προς το εσωτερικό της συστροφής.
- Οι δύο κλώνοι συγκρατούνται μεταξύ τους με δεσμούς υδρογόνου.

Έχει τέσσερις αζωτούχες βάσεις: κυτοσίνη (C), γουανίνη (G), θυμίνη (T), αδενίνη (A), οι οποίες αναπτύσσονται μεταξύ τους καθορισμένους δεσμούς υδρογόνου: η αδενίνη με την θυμίνη και η γουανίνη με την κυτοσίνη. Πιο συγκεκριμένα, μεταξύ της γουανίνης και της κυτοσίνης σχηματίζονται τρεις δεσμοί υδρογόνου, ενώ μεταξύ της αδενίνης και της θυμίνης δύο (Alberts *et al.* 2000).

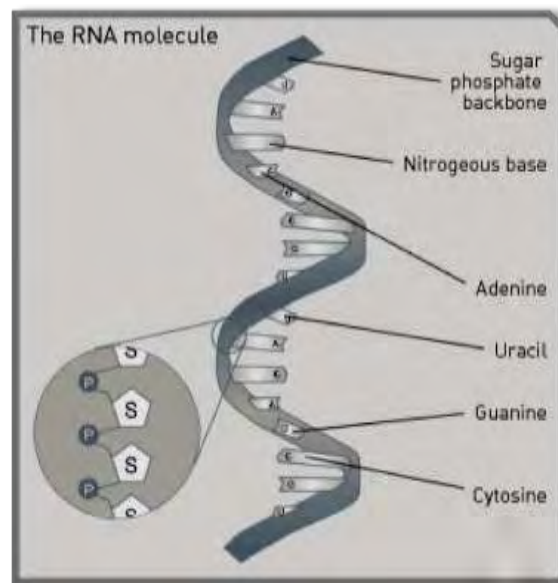


**Εικ. 1.1.:** Το μοντέλο της διπλής έλικας  
(<http://www.achievement.org/autodoc/page/wat0bio-1>).

Σε ό,τι αφορά στο RNA (**Εικ. 1.2**), οι επιστήμονες υποπτεύονταν τον ειδικό του ρόλο στην πρωτεϊνοσύνθεση από το 1939, χάρη σε μελέτες των Κάσπερσον (Torbjörn Oskar Caspersson), Μπρασέ (Jean Brachet) και Σουλτζ (Jack Schultz). Ο Χιούμπερτ Σαντρέν ήταν αυτός που απέδειξε τελικά τους μηχανισμούς δράσης και τον ρόλο του RNA στην πρωτεϊνοσύνθεση μέσα στα ριβοσώματα (Alberts *et al.* 2000).

Το RNA από χημική άποψη είναι παρόμοιο με το DNA. Το RNA περιλαμβάνει τέσσερις τύπους νουκλεοτιδίων, που συνδέονται μεταξύ τους με 3'-5' φωσφοδιεστερικούς δεσμούς: κυτοσίνη (C), γουανίνη (G), αδενίνη (A) και, αντί της θυμίνης (T), την ουρακίλη (U). Οι κυριότερες διαφορές του από το DNA είναι ότι 1) το μόριο του είναι μονόκλωνο και 2) το σάκχαρο στα νουκλεοτίδιά του είναι η ριβόζη αντί της δεσοξυριβόζης. Επίσης, όπως θα αναφερθεί παρακάτω, τα μονομερή νουκλεοτίδια του RNA παίζουν σημαντικό ρόλο στη διαδικασία της μεταγραφής και της μετάφρασης του γενετικού κώδικα από το DNA σε πρωτεΐνη. Τέλος, μέσα στο κύταρο υπάρχουν τέσσερα είδη RNA : α) το **αγγελιαφόρο** RNA (messenger RNA, mRNA), που μεταφέρει πληροφορίες του DNA για την παραγωγή μίας πολυπεπτιδικής αλυσίδας, β) το **ριβosomal** RNA (ribosomal RNA, rRNA), το οποίο συνδέεται με τις πρωτεΐνες και σχηματίζει το

ριβόσωμα, που είναι απαραίτητο για την πρωτεϊνοσύνθεση, γ) το **μεταφορικό RNA** (transfer RNA, tRNA), που συνδέεται με ένα συγκεκριμένο αμινοξύ και το μεταφέρει στη θέση της πρωτεϊνοσύνθεσης και δ) το **μικρό πυρηνικό RNA** (small nuclear RNA, snRNA), το οποίο συνδέεται με μικρά μόρια RNA, τα οποία συνδέονται με πρωτεΐνες και σχηματίζουν μικρά ριβονουκλεοπρωτεϊνικά σωματίδια. Τα σωματίδια αυτά, καταλύουν την ωρίμανση του mRNA, η οποία θα αναφερθεί παρακάτω.(Alberts *et al.* 2000).



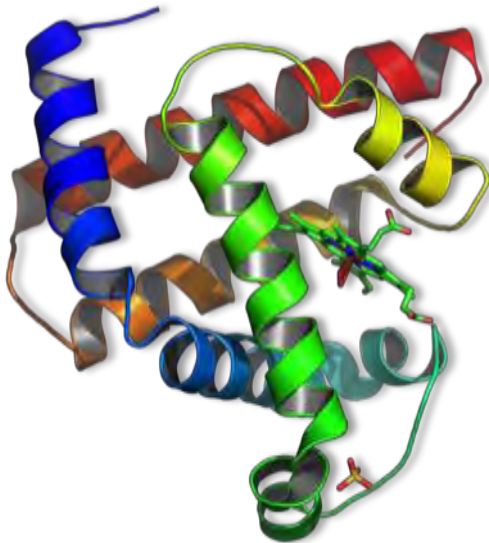
**Εικ. 1.2:** Απεικόνιση του RNA

([http://www.makingthemodernworld.org.uk/learning\\_modules/biology/01.TU.03/?section=6](http://www.makingthemodernworld.org.uk/learning_modules/biology/01.TU.03/?section=6))

Μπορούμε, λοιπόν, να πούμε ότι το DNA ενός κυττάρου είναι ο μοριακός σκληρός δίσκος που περιέχει αποθηκευμένες οδηγίες, οι οποίες καθορίζουν τη δομή και τη λειτουργία του κυττάρου και κατ' επέκταση ολόκληρου του οργανισμού. Με συγκεκριμένες διαδικασίες, τις οποίες θα αναλύσουμε στη συνέχεια (κεντρικό δόγμα της μοριακής βιολογίας), το DNA μεταγράφεται σε RNA, το οποίο έπειτα μεταφράζεται σε πρωτεΐνες.

Οι πρωτεΐνες (Εικ. 1.3) αποτελούν τα πιο διαδεδομένα και πολυδιάστατα, τόσο στη μορφή όσο και στη λειτουργία τους, μακρομόρια (μόρια υψηλού μοριακού βάρους). Είναι μεγάλα σύνθετα βιομόρια, με μοριακό βάρος πάνω από 10000 μέχρι και περισσότερο από 1000000, αποτελούμενα από αμινοξέα, τα οποία ενώνονται μεταξύ τους με πολυπεπτιδικούς δεσμούς, σχηματίζοντας μία γραμμική αλυσίδα (αλυσίδα πολυπεπτιδίων). Στους πιο πολλούς οργανισμούς

το γονιδίωμα ορίζει 20 αμινοξέα, με τα οποία κατασκευάζονται όλες οι απαραίτητες πρωτεΐνες. Οι πρωτεΐνες κατηγοριοποιούνται α) ανάλογα με τη μορφή τους σε *ινώδεις* και *σφαιρικές*, β) ανάλογα με τη σύνθεσή τους σε *απλές* (όταν αποτελούνται μόνο από αμινοξέα) και *σύνθετες* (όταν στο μόριό τους περιλαμβάνονται και μη πρωτεϊνικά τμήματα όπως μέταλλα, σάκχαρα και λίπη) και γ) με βάση τη λειτουργία τους σε *δομικές* (όταν αποτελούν τα δομικά υλικά του κυττάρου) και *λειτουργικές* (όταν συμβάλλουν σε κάποιες λειτουργίες του). Οι πρωτεΐνες είναι απαραίτητες για όλους τους ζωντανούς οργανισμούς και συμμετέχουν σε κάθε διαδικασία μέσα στα κύτταρα. Παραδειγματικά, πολλές πρωτεΐνες δρουν ως ένζυμα που καταλύουν τις βιοχημικές αντιδράσεις και είναι σημαντικές για το μεταβολισμό, άλλες οι δομικές πρωτεΐνες, συμβάλλουν στη διατήρηση της μορφής των κυττάρων και το σχηματισμό του κυτταρικού σκελετού, ενώ ακόμα υπάρχουν πρωτεΐνες που συμμετέχουν στη διακυτταρική επικοινωνία, στη δράση του ανοσοποιητικού συστήματος, στο σχηματισμό κυτταρικών ιστών και στον κυτταρικό κύκλο (Alberts *et al.* 2000).



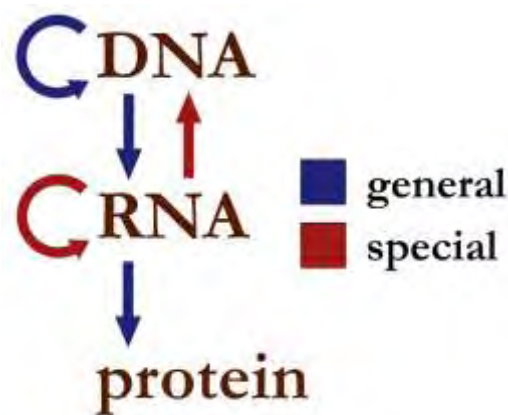
**Εικ. 1.3:** Τρισδιάστατη απεικόνιση  
μίας πρωτεΐνης  
(<http://en.wikipedia.org/wiki/Protein>)

## 1.2 Το κεντρικό δόγμα της μοριακής βιολογίας

Το κεντρικό δόγμα της μοριακής βιολογίας (Εικ. 1.4), το οποίο διατυπώθηκε για πρώτη φορά από τον Φράνσις Κρικ το 1958, αποτελεί μία αναπαράσταση της ροής της γενετικής πληροφορίας. Το δόγμα στην αρχική του μορφή περιλάμβανε τα εξής:

- **Αντιγραφή** του DNA.
- **Μεταγραφή**, μεταφορά δηλαδή της γενετικής πληροφορίας από μορφή DNA σε μορφή αγγελιαφόρου RNA (mRNA).
- **Μετάφραση**, έκφραση δηλαδή της πληροφορίας σε αμινοξέα (πρωτεΐνες) με βάση τον γενετικό κώδικα.

Πλέον, οι ερευνητές έχουν εμπλουτίσει την αρχική μορφή του δόγματος, με την **αντίστροφη μεταγραφή** και τον **αυτοδιπλασιασμό**. Συγκεκριμένα, έχουν εντοπιστεί ορισμένοι ιοί, που έχουν ως γενετικό υλικό RNA και χρησιμοποιούν ένα ειδικό ένζυμο, την αντίστροφη μεταγραφάση, για να χρησιμοποιήσουν αυτό το RNA και να συνθέσουν DNA. Επίσης, έχει διαπιστωθεί ότι σε κάποιους ιούς το RNA μπορεί να αυτοδιπλασιάζεται (Alberts *et al.* 2000).



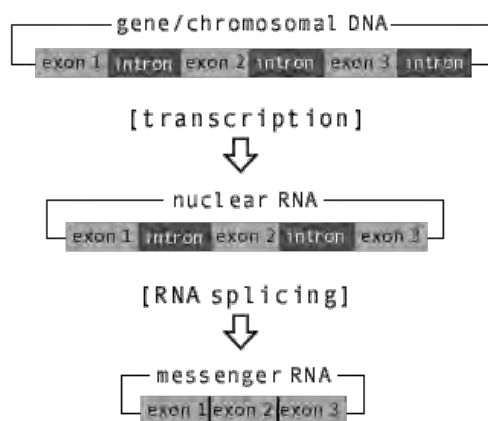
**Εικ. 1.4:** Το κεντρικό δόγμα της μοριακής βιολογίας.  
(<http://www.molecularstation.com/molecular-biology-images/506-molecular-biology-pictures/93-central-dogma-molecular-biology.html>)

Αναλυτικότερα, το πρώτο βήμα στην έκφραση ενός τμήματος των γενετικών οδηγιών ενός κυττάρου είναι η μετατροπή με βάση την αρχή της συμπληρωματικότητας (A-T, C-G), της κατάλληλης αλληλουχίας των νουκλεοτιδίων του DNA, δηλαδή του κατάλληλου γονιδίου, σε μία αλληλουχία νουκλεοτιδίων του RNA. Η διαδικασία αυτή λέγεται **μεταγραφή** (transcription)

επειδή, παρόλο που οι πληροφορίες μετατρέπονται σε μία διαφορετική χημική μορφή, εξακολουθούν να γράφονται στην ίδια γλώσσα: αυτή των νουκλεοτιδίων. Όλο το RNA ενός κυττάρου παράγεται με μεταγραφή, η οποία εμφανίζει αρκετές ομοιότητες με την αντιγραφή του DNA. Συγκεκριμένα, η μεταγραφή αρχίζει με τη διάνοιξη και το ξεδίπλωμα ενός μικρού τμήματος της διπλής έλικας του DNA, έτσι ώστε να αποκαλυφθούν οι βάσεις του κάθε κλώνου. Έπειτα, ένας από τους δύο κλώνους δρα ως εκμαγείο για τη σύνθεση του RNA, ακολουθώντας τους κανόνες συμπληρωματικότητας των βάσεων. Τα ένζυμα που επιτελούν την μεταγραφή ονομάζονται RNA πολυμεράσες (RNA polymerases), οι οποίες καταλύουν το σχηματισμό των φωσφοδιεστερικών δεσμών που συνδέουν τα νουκλεοτίδια μεταξύ τους και δημιουργούν τον σακχαροφωσφορικό σκελετό της αλυσίδας του RNA. Η διαδικασία της μεταγραφής ξεκινά από συγκεκριμένα σημεία πάνω στο DNA, που ονομάζονται υποκινητές και εντοπίζονται στην αρχή κάθε γονιδίου. Επιπλέον, προκειμένου η RNA πολυμεράση να αρχίσει σωστά τη μεταγραφή, απαιτείται η δράση συγκεκριμένων πρωτεϊνών, που ονομάζονται μεταγραφικοί παράγοντες. Κατά την έναρξη της μεταγραφής ενός γονιδίου η RNA πολυμεράση προσδένεται στον υποκινητή και προκαλεί τοπικό ξετύλιγμα της διπλής έλικας του DNA. Στη συνέχεια, τοποθετεί τα ριβονουκλεοτίδια απέναντι από τα δεοξυριβονουκλεοτίδια μίας αλυσίδας του DNA, σύμφωνα με τον κανόνα της συμπληρωματικότητας των βάσεων. Η RNA πολυμεράση συνδέει τα ριβονουκλεοτίδια, που προστίθενται το ένα μετά το άλλο, με 3'→5' φωσφοδιεστερικό δεσμό. Η μεταγραφή έχει προσανατολισμό 5'→3'. Η σύνθεση του mRNA σταματά στο τέλος του γονιδίου, όπου ειδικές αλληλουχίες (αλληλουχίες λήξης της μεταγραφής) επιτρέπουν την απελευθέρωσή του. Η σχεδόν άμεση απελευθέρωση του κλώνου του mRNA από το DNA, ενώ εξελίσσεται η σύνθεσή του, σημαίνει ότι, μέσα σε σχετικά σύντομο χρονικό διάστημα, είναι δυνατόν να παραχθούν πολλά μόρια mRNA από το ίδιο γονίδιο, ενώ η σύνθεση του επόμενου μορίου mRNA συνήθως αρχίζει προτού ολοκληρωθεί το προηγούμενο (Alberts *et al.* 2000).

Στους προκαρυωτικούς οργανισμούς το mRNA αρχίζει να μεταφράζεται σε πρωτεΐνη πριν ακόμη ολοκληρωθεί η μεταγραφή του εξαιτίας της έλλειψης

πυρηνικής μεμβράνης. Αντίθετα, στους ευκαρυωτικούς οργανισμούς, το RNA, που παράγεται κατά τη μεταγραφή ενός γονιδίου, συνήθως δεν είναι έτοιμο να μεταφραστεί, αλλά υφίσταται μία πολύπλοκη διαδικασία ωρίμανσης. Η διαδικασία αυτή αποτελεί ένα από τα πιο ενδιαφέροντα ευρήματα της μοριακής βιολογίας, γιατί οδήγησε στο συμπέρασμα ότι τα περισσότερα γονίδια των ευκαρυωτικών οργανισμών είναι ασυνεχή ή διακεκομμένα. Ειδικότερα, η αλληλουχία που μεταφράζεται σε αμινοξέα (εξώνια) διακόπτεται από ενδιάμεσες αλληλουχίες, οι οποίες δεν μεταφράζονται σε αμινοξέα (εσώνια). Όταν ένα γονίδιο που περιέχει εσώνια μεταγράφεται, δημιουργείται το πρόδρομο mRNA που περιέχει τόσο εξωνικές όσο και εσωνικές αλληλουχίες. Η ωρίμανση του πρόδρομου mRNA περιλαμβάνει τη δράση μικρών ριβονουκλεοπρωτεϊνικών σωματιδίων (snRNA και πρωτεΐνες), τα οποία κόβουν τα εσώνια, τα απομακρύνουν και συρράπτουν τα εξώνια μεταξύ τους. Έτσι, δημιουργείται το ώριμο mRNA, το οποίο περιλαμβάνει δύο περιοχές που δε μεταφράζονται σε αμινοξέα. Η μία βρίσκεται στο 5' άκρο και η άλλη στο 3' άκρο και ονομάζονται 5'- και 3'- αμετάφραστες περιοχές, αντίστοιχα. Το ώριμο mRNA μεταφέρεται από τον πυρήνα στο κυτταρόπλασμα και ειδικότερα στα ριβοσώματα, όπου πραγματοποιείται και η πρωτεϊνοσύνθεση (Εικ. 1.5) (Alberts *et al.* 2000).



Εικ. 1.5: Η διαδικασία ωρίμανσης ενός μορίου mRNA

Μετά την ολοκλήρωση της μεταγραφής, το δεύτερο στάδιο είναι η **μετάφραση**. Η αλληλουχία των βάσεων του mRNA καθορίζει την αλληλουχία των αμινοξέων στις πρωτεΐνες, με βάση έναν κώδικα αντιστοίχισης, το γενετικό κώδικα: μία τριπλέτα νουκλεοτιδίων mRNA (κωδικόνιο) καθορίζει ένα αμινοξύ. Επειδή, ο αριθμός των διαφορετικών συνδυασμών των τεσσάρων νουκλεοτιδίων ανά



τρία ( $4^3 = 64$ ) είναι πολύ μεγαλύτερος από τον αριθμό των αμινοξέων που συγκροτούν τις πρωτεΐνες (20), γίνεται κατανοητό ότι κάποια αμινοξέα θα πρέπει να κωδικοποιούνται από περισσότερες από μία τριπλέτες. Για το λόγο αυτό, ο γενετικός κώδικας (Εικ. 1.6) χαρακτηρίζεται ως εκφυλισμένος. Η αντιστοίχιση των κωδικονίων σε αμινοξέα και η διαδοχική σύνδεση των αμινοξέων σε πολυπεπτιδική αλυσίδα (μετάφραση του mRNA), πραγματοποιείται στα ριβοσώματα με τη βοήθεια των μορίων tRNA, τη συμμετοχή αρκετών πρωτεϊνών και την κατανάλωση ενέργειας (Alberts *et al.* 2000).

		Second letter						
		U	C	A	G			
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC		UCC		UAC		UGC	
	UUA	Leu	UCA		UAA	Stop	UGA	Stop
	UUG		UCG		UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC		CCC		CAC		CGC	
	CUA		CCA		CAA	Gin	CGA	
	CUG		CCG		CAG		CGG	
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC		ACC		AAC		AGC	
	AUA		ACA		AAA	Lys	AGA	Arg
	AUG	Met	ACG		AAG		AGG	
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC		GCC		GAC		GGC	
	GUA		GCA		GAA	Glu	GGA	
	GUG		GCG		GAG		GGG	

Εικ. 1.6: Ο γενετικός κώδικας.

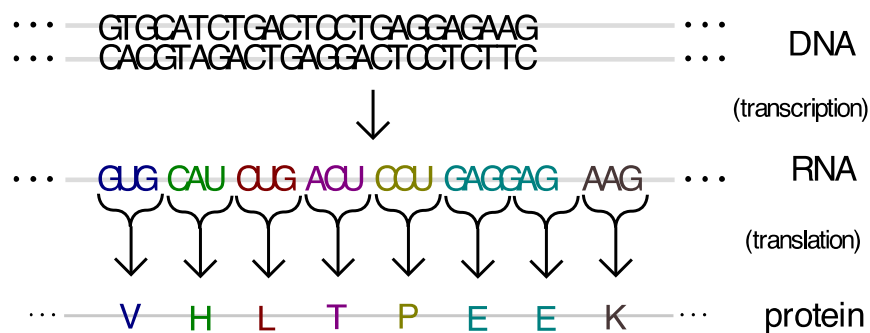
(<http://www.mun.ca/biology/scarr/MGA2-03-28.jpg>)

Η πρωτεϊνόςύνθεση διακρίνεται σε τρία στάδια (Alberts *et al.* 2000):

1. **Έναρξη:** Κατά την έναρξη της μετάφρασης του RNA σε πρωτεΐνες, το mRNA συνδέεται μέσω μίας αλληλουχίας που υπάρχει στην 5'-αμετάφραστη περιοχή του, με το ριβοσωμικό RNA της μικρής υπομονάδας του ριβοσώματος, ακολουθώντας τους κανόνες της συμπληρωματικότητας. Το AUG αποτελεί το πρώτο κωδικόνιο του mRNA για όλα τα γονίδια και σε αυτό προσδένεται το tRNA που φέρει το αμινοξύ μεθειονίνη. Όμως, όλες οι πρωτεΐνες του οργανισμού δεν έχουν πάντα ως πρώτο αμινοξύ την μεθειονίνη, κάτι που οφείλεται στην απομάκρυνση ορισμένων αμινοξέων από το αρχικό αμινικό άκρο τους. Το σύμπλοκο που

δημιουργείται αφού γίνει η πρόσδεση του mRNA και του tRNA ονομάζεται σύμπλοκο έναρξης.

2. **Επιμήκυνση:** Στο στάδιο αυτό, η μεγάλη υπομονάδα συνδέεται με τη μικρή και ένα δεύτερο μόριο tRNA, το οποίο έχει αντικωδικόνιο συμπληρωματικό του δεύτερου κωδικονίου του mRNA τοποθετείται στην κατάλληλη εισδοχή του ριβοσώματος μεταφέροντας το δεύτερο αμινοξύ. Μεταξύ της μεθειονίνης και του δεύτερου αμινοξέος σχηματίζεται πεπτιδικός δεσμός και αμέσως μετά το πρώτο tRNA αποσυνδέεται από το ριβόσωμα και απελευθερώνεται στο κυτταρόπλασμα, όπου συνδέεται με τη μεθειονίνη. Το ριβόσωμα και το mRNA έχουν ένα tRNA, πάνω στο οποίο είναι προσδεμένα δύο αμινοξέα. Έτσι, ξεκινάει η επιμήκυνση της πολυπεπτιδικής αλυσίδας. Στη συνέχεια το ριβόσωμα κινείται κατά μήκος του mRNA κατά ένα κωδικόνιο. Ένα τρίτο tRNA έρχεται να προσδεθεί μεταφέροντας το αμινοξύ του. Ανάμεσα στο δεύτερο και στο τρίτο αμινοξύ σχηματίζεται ένας πεπτιδικός δεσμός. Η πολυπεπτιδική αλυσίδα συνεχίζει να αναπτύσσεται, καθώς νέα tRNA φέρουν αμινοξέα τα οποία προσδένονται μεταξύ τους.
3. **Λήξη:** Η διαδικασία της επιμήκυνσης σταματά σε ένα κωδικόνιο λήξης (UGA, UAG ή UAA), εξαιτίας του ότι δεν υπάρχουν tRNA που να αντιστοιχούν σε αυτά. Το τελευταίο tRNA απομακρύνεται από το ριβόσωμα και η πολυπεπτιδική αλυσίδα απελευθερώνεται από αυτό (Alberts *et al.* 2000).

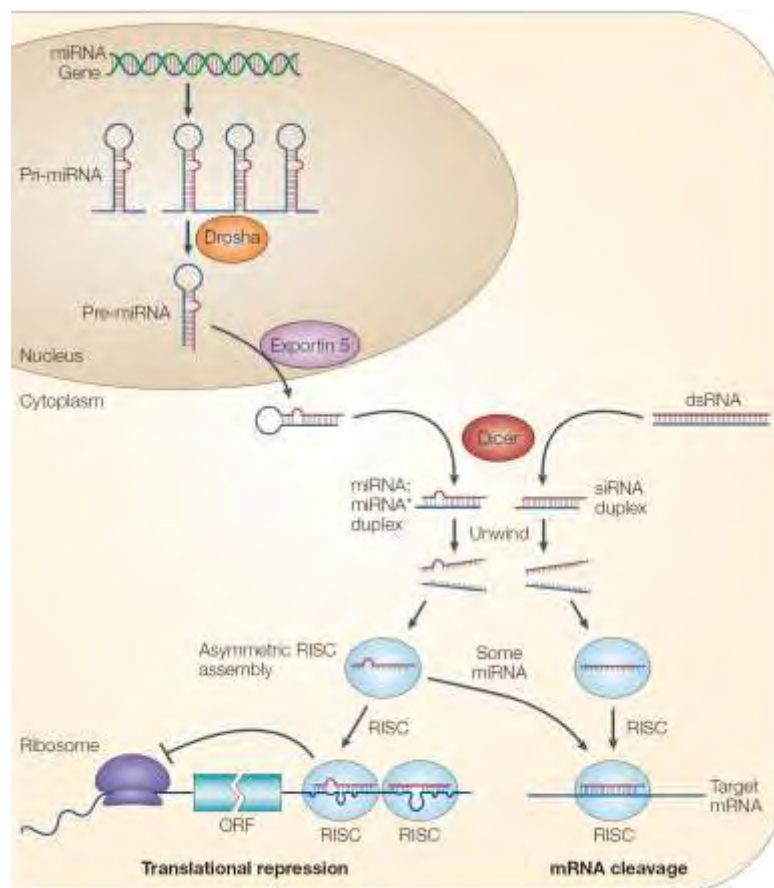


**Εικ. 1.7:** Δημιουργία πρωτεΐνης από DNA  
 (<http://www.websters-online-dictionary.org/definitions/Gene%20Expression>)

## 2. MICRORNA

### 2.1 Εισαγωγή στα MicroRNAs (miRNAs)

Η διαδικασία της μεταγραφής του DNA σε RNA αφορά, στη συντηρητική πλειονότητά της, τη δημιουργία μορίων mRNA και τελικά την πρωτεϊνοσύνθεση. Παρόλα αυτά, είναι γνωστό πως ένα μέρος της μεταγραφής αφορά τη δημιουργία μη κωδικοποιητών μορίων RNA, των ncRNA (non-coding RNA). Σε αυτή την κατηγορία ανήκουν και τα microRNAs (Εικ. 2.1).



**Εικ. 2.1:** Λειτουργία των μορίων microRNAs (He-Hannon 2004)

Τα microRNAs (miRNAs) είναι μικρά ενδογενή μη κωδικοποιητικά μόρια ριβονουκλεϊκού οξέος (RNA), που κατά μέσο όρο αποτελούνται από 22 νουκλεοτίδια και τα οποία συναντώνται σε όλα τα ευκαρυωτικά κύτταρα. Ανακαλύφθηκαν το 1993 από τους Victor Ambros, Rosalind Lee και Rhonda Feinbaum κατά τη διάρκεια μελέτης του γονιδίου *lin-14* στον γεωσκόληκα *Caenorhabditis Elegans* (Maragkakis *et al.* 2009).

Έχει αποδειχθεί μέσω πειραμάτων ότι το πιο σημαντικό κομμάτι του microRNA για την αναγνώριση και τον εντοπισμό της θέσης πρόσδεσης στο γονίδιο-στόχο είναι τα νουκλεοτίδια κοντά στο 5' άκρο του. Έχει ακόμα αποδειχθεί ότι τα μοτίβα στην 3'- αμετάφραστη περιοχή της πρωτεΐνης που αντιστοιχούν στα νουκλεοτίδια 2-7 των microRNAs, είναι εξελικτικά διατηρημένα σε αρκετά είδη και ονομάζονται κεντρική περιοχή (seed region) (Alexiou *et al.* 2009). Τα γονίδια των microRNAs μεταγράφονται από την RNA πολυμεράση II στον πυρήνα του κυττάρου, δημιουργώντας το pri-microRNA μετάγραφο. Στη συνέχεια, από το pri-microRNA δημιουργείται το pre-microRNA (μήκους 70 νουκλεοτιδίων) με τη χρήση ενός πρωτεϊνικού συμπλέγματος, που περιλαμβάνει τα ένζυμα Drosha και Pasha. Τα pre-microRNAs μετατρέπονται σε ώριμα microRNAs στο κυτταρόπλασμα με τη βοήθεια ενός ενζύμου (Dicer) (Alexiou *et al.* 2009). Με τη βοήθεια του συμπλόκου RISC (RNA-induced Silencing Complex), το microRNA μπορεί να συνδεθεί με συμπληρωματικό τρόπο στην αλληλουχία του mRNA στόχου και να ρυθμίσει αρνητικά την γονιδιακή έκφραση. Υπάρχουν δύο τρόποι για να το πετύχει αυτό:

α) Το microRNA συνδέεται με ατελή συμπληρωματικότητα στο mRNA στόχο στην 3'- αμετάφραστη περιοχή, καταστέλλοντας τη μετάφραση και, κατά συνέπεια, την παραγωγή πρωτεΐνης, και

β) το microRNA συνδέεται με πλήρη συμπληρωματικότητα στο ανοιχτό πλαίσιο ανάγνωσης του mRNA, με αποτέλεσμα να επάγει την αποικοδόμηση του mRNA (Alexiou *et al.* 2009, Σουρβίνου κ.ά. 2010 ).

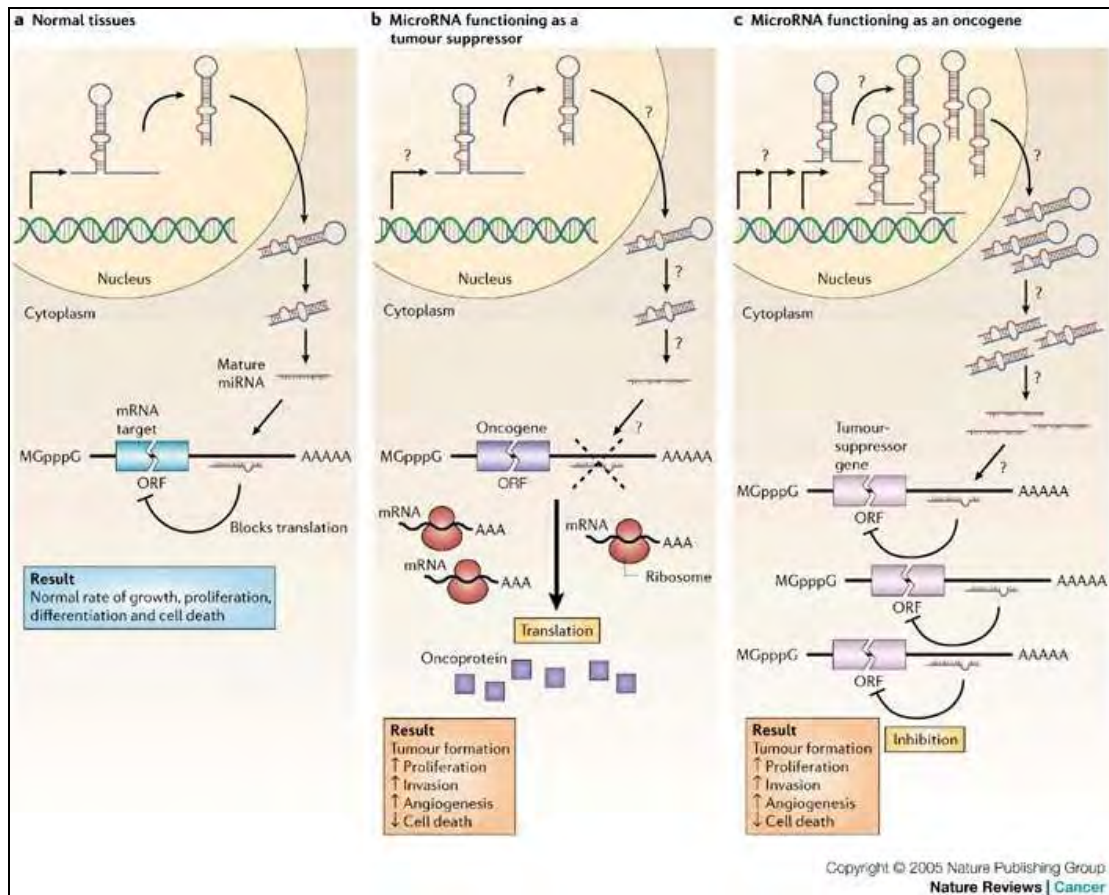
Τα microRNAs ρυθμίζουν σημαντικές βιολογικές διεργασίες, όπως η κυτταρική ανάπτυξη, η διαφοροποίηση και ο πολλαπλασιασμός, ενώ επιπλέον διαδραματίζουν σημαντικό ρόλο στην εξέλιξη των ασθενειών (καρδιαγγειακές, νευροεκφυλιστικές, κακοήθειες). Συγκεκριμένα, οι επιστήμονες πιστεύουν ότι τα microRNAs μπορούν να δράσουν εξίσου ως ογκοκατασταλτικά γονίδια και ως ογκογονίδια. Στην πρώτη περίπτωση μεταγράφονται, ωριμάζουν και προσδένονται στο mRNA ενός ογκογονιδίου, με συνέπεια την καταστολή της έκφρασής του -ογκοκατασταλτική δράση- διατηρώντας τους φυσιολογικούς ρυθμούς ανάπτυξης του κυττάρου (αύξηση, πολλαπλασιασμός, διαφοροποίηση και θάνατος). Αντίθετα, η καταστολή της έκφρασης ενός ογκοκατασταλτικού

γονιδίου που αποτελεί στόχο κάποιου μορίου microRNA, έχει σαν αποτέλεσμα το σχηματισμό όγκου –δράση ογκογονιδίου- (Εικ. 2.2). Πράγματι, έχει αποδειχθεί πως αυξημένα επίπεδα microRNA σε λανθασμένες χρονικές στιγμές, σε λανθασμένους ιστούς, μπορούν να ελαττώσουν την έκφραση ενός ογκοκατασταλτικού γονιδίου-στόχου (Λογοθέτη κ.ά. 2006). Επιπρόσθετα, τα microRNAs αποτελούν δείκτες για την ταξινόμηση των διάφορων νεοπλασμάτων και την πρόγνωση της κλινικής τους εξέλιξης, ενώ παράλληλα μπορούν να μας δώσουν πολλές πληροφορίες για την προέλευση και το στάδιο διαφοροποίησης του όγκου. Μερικά παραδείγματα microRNA, που λειτουργούν ως ογκοκατασταλτικά, αποτελούν τα microRNAs lin-4 και lin-7, που ρυθμίζουν τον κυτταρικό πολλαπλασιασμό και τη διαφοροποίηση στο γεωσκόληκα *Caenorhabditis Elegans*, το miR-4 που καταστέλλει την απόπτωση στη μύγα του ξυδιού *Drosophila Melanogaster* και το microRNA-21 που υπερεκφράζεται σε τουλάχιστον 11 διαφορετικούς τύπους συμπαγών όγκων, στο μαστό, το στομάχι, το παχύ έντερο, τον πνεύμονα κ.α.. Αξίζει να σημειωθεί, τέλος, ότι το γονιδίωμα του ανθρώπου μπορεί να κωδικοποιήσει πάνω από 1000 microRNAs.

Ιδιαίτερη σημασία έχει η ανακάλυψη στην οποία προέβει μία ομάδα ερευνητών από το αντικαρκινικό κέντρο Memorial Sloan-Kettering της Νέας Υόρκης με επικεφαλής τον Δρ Σοχάι Ταβαζόιε, η οποία απέδειξε ότι τα microRNAs εμποδίζουν τον καρκίνο του μαστού να κάνει μετάσταση και σε άλλα όργανα του ανθρώπινου οργανισμού όπως είναι οι πνεύμονες αλλά και τα οστά που αποτελούν τα βασικά σημεία μετάστασης. Η συσχέτισή τους με τη δημιουργία καρκίνου ενισχύθηκε από το ότι κατανέμονται σε εύθραυστες ή σχετιζόμενες με τον καρκίνο γενωμικές περιοχές και ότι σε πολλούς καρκίνους του ανθρώπου είναι κατεσταλμένα (Τσίντου 2011). Είναι λοιπόν κατανοητό ότι τα microRNA κατέχουν σημαντικό ρόλο ως ρυθμιστές πολλών κυτταρικών λειτουργιών μέσω της μετα-μεταγραφικής καταστολής διάφορων γονιδίων (Εικ. 2.3).

Tumor type	Profiling method	miRNAs deregulated in tumor		note	ref
		up-regulated	down-regulated		
breast cancer	bead			classification of breast tumors into five sub-types	(93)
breast tumor	qRT-PCR array	miR-21			(166)
breast cancer	array			classification according to ErbB2/ER status	(158)
breast cancer	array	miR-21	miR-125b, -145	clinicopathologic features	(185)
prostate cancer	array		let-7c, miR-125b, miR-145		(186)
prostate cancer	membrane			classification according to androgen dependency	(66)
prostate cancer	array				(158)
pancreatic cancer	array	miR-21, -221, -222, -181a, -181b, -181d, -155		vs. chronic pancreatitis and normal tissues	(187)
pancreatic cancer	array	miR-196a	miR-217	pancreatic ductal adenocarcinoma vs. chronic pancreatitis and normal tissues	(180)
pancreatic cancer	qRT-PCR of pre-miRNA	let-7f, miR-100, -155, -221, -301, -21, -181a, -125b, -212, -378a	miR-375	pancreatic adenocarcinoma, paired benign tissue, normal pancreas, chronic pancreatitis	(188)
pancreatic tumor	array	miR-103, -107, -23a, -28b, -342, -192, -204, -211, -21	miR-155	endocrine tumors vs. acinar cell carcinoma	(189)
lung cancer	array	miR-21, -205	miR-126*	clinicopathological features	(190)
ovarian cancer	array	miR-200a, -141	miR-199a, -140, -145, -125b1	clinicopathological features	(191)
cervical cancer	array			vs. Drosha overexpression	(95)
uterine leiomyoma	array	let-7, miR-21, -23b		clinicopathological features (tumor size, race, etc)	(192)
hepatocellular cancer	array	miR-21			(134)
hepatocellular cancer	array	miR-221	let-7a, miR-122a	vs. liver cirrhosis	(103)
hepatocellular cancer	array	miR-224, -18	miR-199a, -199a*, -200a		(194)
thyroid cancer	array		miR-30d, -125b, -28a, and -30a-5p	anaplastic thyroid carcinomas vs. normal tissues	(67)
thyroid cancer	array	miR-197, -346 in FTC relative to FA		follicular thyroid carcinoma (FTC) vs. follicular adenoma (FA)	(195)
thyroid cancer	array	miR-221, -222, -181b		papillary thyroid carcinoma vs. normal tissues	(159)
thyroid cancer	array	miR-221, -222, -148b		papillary thyroid carcinoma vs. normal tissues	(100)
colorectal cancer	array	miR-25, -92 in MSS relative to MSI-H		microsatellite stability (MSS) vs. high microsatellite instability (MSI-H)	(196)
colorectal cancer	qRT-PCR	miR-31, -96, -135b, -183	miR-133b, -145	miR-31 according to cancer stage	(197)
colorectal cancer	cloning miSAGE				(17)
pituitary adenomas	array	miR-26a, -149	miR-21, -141, -144	clinicopathological features	(168)
neuroblastoma	qRT-PCR			neuroblastoma subtypes	(89)
glioblastoma	array	miR-221, -10b	miR-128, -181b, -181a		(169)
leukemia	bead	miR-128a, -128b in ALL compared with AML	let-7b, miR-223 in ALL compared with AML	acute lymphoblastic leukemia (ALL) vs. acute myeloid leukemia (AML)	(200)
leukemia	cloning	miR-21, -150, -155	miR-92, -222	various chronic lymphocytic leukemias (CLL)	(201)
leukemia	array		miR-29b, -181b in aggressive CLL with 11q deletion	three subtypes of CLL	(202)
leukemia	array		miR-16-1, -15a	CLL according to ZAP-70 and IgVH	(203)
leukemia	array			B-cell CLL	(65)
various tumors	array cloning			various sarcomas	(204)
various tumors	array			solid tumors (lung, breast, stomach, prostate, colon, and pancreatic tumors)	(122)
various tumors	bead			comprehensive collection of various cancers and normal tissues	(54)

Εικ. 2.2: MicroRNA ανάλυση σε καρκινικούς όγκους (Sun Lee and Dutta 2009)



Εικ. 2.3: MicroRNAs λειτουργώντας ως ογκογονίδια και ογκοκατασταλτικά. (Esquela- Kerscher and Slack 2006)

## 2.2 Πρόβλεψη στόχων microRNAs (target prediction)

Όπως αναφέρθηκε παραπάνω, τα microRNAs προσδένονται πάνω σε μόρια του αγγελιοφόρου RNA, τα οποία ονομάζονται στόχοι (targets). Η αναγνώριση των στόχων μπορεί να γίνει είτε με πειραματικές μεθόδους είτε με τη χρήση εξειδικευμένων αλγορίθμων. Οι πειραματικές μέθοδοι χωρίζονται σε δύο κατηγορίες, τις άμεσες και τις έμμεσες. Οι άμεσες μέθοδοι επιτρέπουν την εκτίμηση συγκεκριμένων αλληλεπιδράσεων microRNA με mRNA, ενώ οι έμμεσες μέθοδοι βασίζονται σε ευρείας κλίμακας πειράματα με μικροσυστοιχίες ή με πειράματα ποσοτικοποίησης των πρωτεϊνών, εξετάζοντας αν η έκφραση ενός γονιδίου αυξάνεται ή μειώνεται, χωρίς να εντοπίζεται άμεσα η αλληλεπίδραση (Alexiou *et al.* 2009).

Αναλυτικότερα, στην άμεση μέθοδο χρησιμοποιούνται πρωτεϊνικά σκευάσματα, όπως είναι η λουσιφεράση (Luciferase) και η GFP (Green Fluorescent Protein), η

γενετική πληροφορία για τα οποία ενσωματώνεται στην 3'-αμετάφραστη περιοχή (3'-UTR) του εν δυνάμει γονιδιακού στόχου. Έτσι, όταν θα συντεθεί το mRNA του γονιδίου-στόχος και θα μεταφραστεί σε πρωτεΐνη, θα δημιουργηθεί παράλληλα και το πρωτεϊνικό σκεύασμα που επιλέχθηκε. Στην περίπτωση της λουσιφεράσης, η οποία εκπέμπει φως, μπορεί να μετρηθεί η ποσότητα αυτού του φωτός, ενώ στην περίπτωση της GFP, μετράται φθορισμός πράσινου χρώματος. Η ποσότητα του φωτός ή φθορισμού αντανακλά άμεσα τα επίπεδα έκφρασης του mRNA του γονιδίου-στόχος και άρα την καταστολή (ή μη) που προκάλεσε η δράση του microRNA που μελετάται (Alexiou *et al.* 2009).

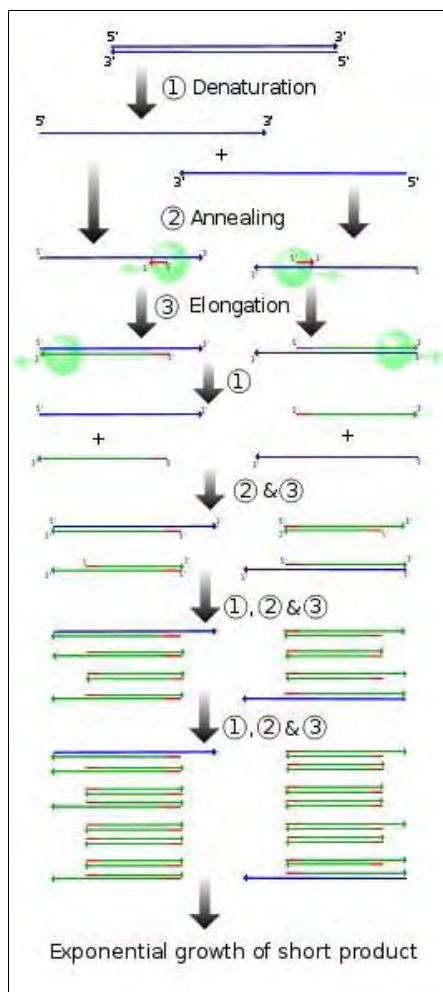
Άλλη άμεση μέθοδος για να μετρηθεί η γονιδιακή έκφραση είναι η RT-PCR (Reverse Transcriptase Polymerase Chain Reaction), η οποία αποτελεί παραλλαγή της απλής μεθόδου PCR, η οποία περιλαμβάνει τρία στάδια (Εικ. 2.4):

- Αποδιάταξη (Denaturation)
- Επαναδιάταξη (Annealing)
- Επιμήκυνση (Extension)

Αυτά τα τρία στάδια επαναλαμβάνονται 20 έως 40 φορές και ουσιαστικά εκθέτουν το δείγμα του DNA σε διαφορετικές θερμοκρασίες. Για να επιτευχθούν οι αλλαγές αυτές χρησιμοποιείται η μέθοδος "Peltier effect", η οποία πετυχαίνει τη θέρμανση και ψύξη των δειγμάτων, αντιστρέφοντας το ηλεκτρικό ρεύμα. Συγκεκριμένα, στην αποδιάταξη το δείγμα εκτίθεται σε θερμοκρασία 94-98°C για 20 έως 30 δευτερόλεπτα. Οι δεσμοί υδρογόνου του DNA διασπώνται με αποτέλεσμα να έχουμε δύο κλώνους DNA. Έπειτα, στο στάδιο της επαναδιατάξης, η θερμοκρασία πέφτει στους 55 έως 72°C για 20 έως 40 δευτερόλεπτα για να δώσει τη δυνατότητα στους εκκινήτες (primers) να προσδεθούν στους ελεύθερους κλώνους οι οποίοι αποτελούνται από 18 έως 22 νουκλεοτίδια και χρησιμοποιούνται ως σημεία εκκίνησης για τη δράση της πολυμεράσης. Τέλος, στο στάδιο της επιμήκυνσης το δείγμα εκτίθεται σε μία θερμοκρασία που εξαρτάται από τις ιδιότητες της DNA πολυμεράσης που χρησιμοποιείται (συνήθως χρησιμοποιείται η Taq polymerase που ενεργοποιείται στους 72°C). Η πολυμεράση προσθέτει νουκλεοτίδια με βάση την



αρχή της συμπληρωματικότητας, ξεκινώντας από την αρχή κάθε κλώνου. Με αυτό τον τρόπο ενισχύεται η ποσότητα του δείγματος και μπορεί να εντοπιστεί ευκολότερα το DNA. Επειδή όμως, το RNA από τη φύση του είναι ευαίσθητο στις υψηλές θερμοκρασίες, προκειμένου να μετρηθεί η ποσότητα μορίων mRNA ενός γονιδίου, θα πρέπει πρώτα το mRNA να μετατραπεί σε cDNA (copy DNA) μέσω της διαδικασίας της ανάστροφης μεταγραφάσης και έπειτα να εφαρμοστεί σε αυτό η PCR (Alberts *et al.* 2000). Η μέθοδος αυτή λέγεται RT-PCR και μπορεί να προσφέρει μια ποσοτικοποίηση των μορίων cDNA που αντιστοιχούν σε κάθε γονίδιο. Κατόπιν, μπορούν να συγκριθούν τα επίπεδα έκφρασης διαφόρων γονιδίων σε κύτταρα που φέρουν το microRNA που εξετάζεται και σε άλλα που δεν το φέρουν. Με αυτόν τον τρόπο, μπορούν να ταυτοποιηθούν ορισμένα από τα γονίδια -στόχοι, καθώς αυτά είναι που θα παρουσιάσουν τη μεγαλύτερη μεταβολή στα επίπεδα έκφρασης.



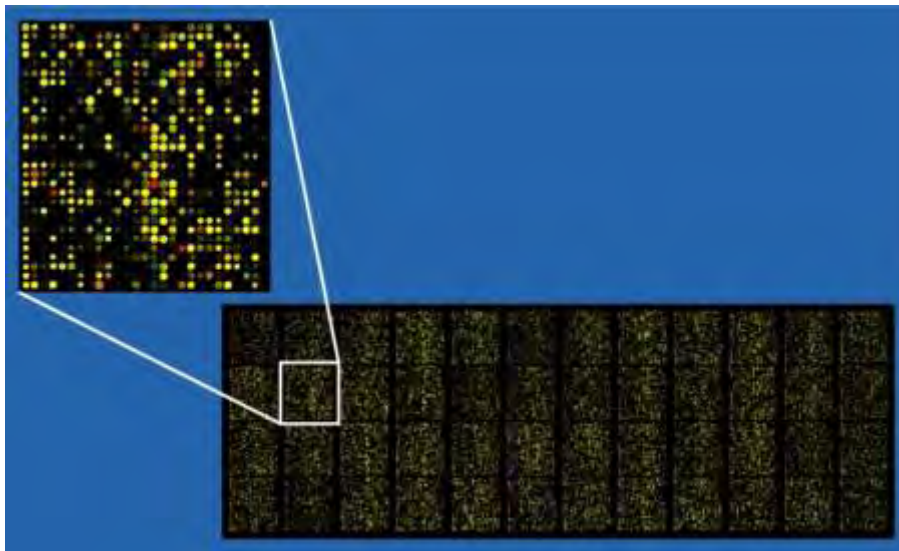
**Εικ. 2.4:** PCR (Polymerase Chain Reaction).  
([http://en.wikipedia.org/wiki/Polymerase\\_chain\\_reaction](http://en.wikipedia.org/wiki/Polymerase_chain_reaction))

Αυτές οι μέθοδοι ποσοτικοποίησης δεν δίνουν πληροφορίες για τα σημεία του mRNA στα οποία προσδένονται τα microRNAs (MREs, MicroRNA Recognition Elements). Για να γίνουν γνωστά τα σημεία αυτά, χρειάζεται η κατευθυνόμενη μεταλλαξιγέννηση, μια διακασία εξαιρετικά επίπονη. Πιο συγκεκριμένα, εισάγονται τεχνητά σημειακές μεταλλάξεις στην αλληλουχία ενός mRNA που αποτελεί στόχο για το υπό μελέτη miRNA, και ελέγχεται η δυνατότητα πρόσδεσης του τελευταίου στις αλληλουχίες αυτές. Επειδή η πρόσδεση είναι απόλυτα ειδική, αλλαγή ακόμη και ενός από τα νουκλεοτίδια που απαιτούνται για την παραπάνω πρόσδεση θα υποβαθμίσει την ισχύ της πρόσδεσης ή ακόμα θα την αποτρέψει εντελώς. Με τον τρόπο αυτό μπορούμε να εντοπίσουμε την ακριβή αλληλουχία του mRNA που αναγνωρίζεται από το miRNA και απαιτείται για πλήρη και ισχυρή πρόσδεσή του.

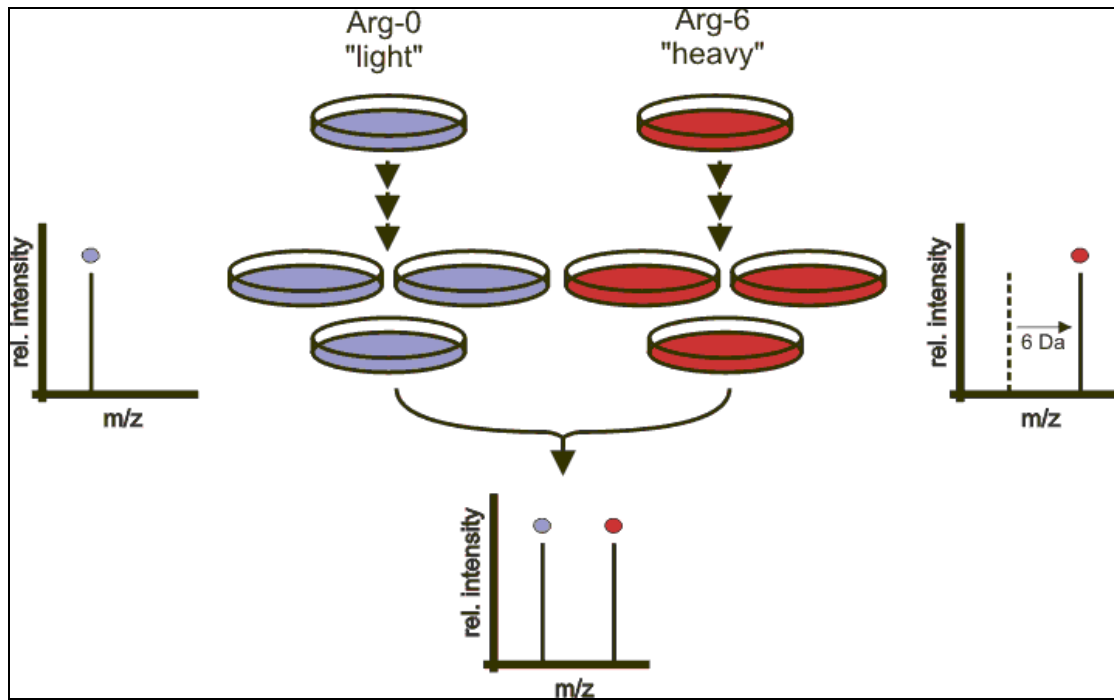
Εκτός, όμως, από τις άμεσες μεθόδους υπάρχουν και οι έμμεσες, οι οποίες δίνουν πληροφορίες για την επίδραση που θα έχει στα γονίδια ολόκληρου του κυττάρου το microRNA, αν δηλαδή η γονιδιακή έκφραση αυξήθηκε, μειώθηκε ή παρέμεινε αμετάβλητη. Η υπερέκφραση ενός microRNA μπορεί να επιτευχθεί με τη χρήση ώριμου microRNA (mature microRNA) ή με precursor microRNA (hairpin) ή με pre-microRNA in vivo (για παράδειγμα εισάγοντας ενέσιμα το microRNA σε ένα ζωντανό οργανισμό) και in vitro (μέσα σε δοκιμαστικό σωλήνα σε κύτταρα εκτός του οργανισμού).

Η πρώτη έμμεση μέθοδος είναι αυτή της αποσιώπησης (silencing). Η μέθοδος αυτή γίνεται με δύο τρόπους : α) Με μικρές αλληλουχίες απόλυτα συμπληρωματικές με το microRNA οι οποίες τοποθετούνται στο κύτταρο και εξουδετερώνουν το microRNA και β) αποσιωπώντας το γονίδιο και μην επιτρέποντάς του να εκφραστεί (knocking down). Έχοντας, λοιπόν, δύο δείγματα, ένα με υπερέκφρασμένο το microRNA και ένα δεύτερο με αποσιωπημένο το microRNA, μπορεί κανείς να συγκρίνει τα επίπεδα έκφρασης ενός πολύ μεγάλου αριθμού γονιδίων, ανιχνεύοντας με τον τρόπο αυτό την επίδραση του miRNA σε ολόκληρο το κύτταρο. Η μελέτη των επιπέδων μπορεί να γίνει με μικροσυστοιχίες γονιδίων (microarrays, γονιδιακά τσιπ, **Εικ. 2.5**), μία μεθοδολογία με σχετικά υψηλό κόστος και με συχνά ανακριβή αποτελέσματα λόγω τεχνικών προβλημάτων. Οι μικροσυστοιχίες είναι μία διάταξη

μικροσκοπικών σημείων που αντιπροσωπεύουν μοναδικά γονίδια και ακινητοποιούνται με ομοιοπολικούς δεσμούς σε μία στερεή επιφάνεια (συνήθως γυάλινη). Η μέθοδος αυτή δίνει σημαντικές πληροφορίες για τους στόχους των microRNA, αφού γίνεται αντιληπτή η ελάττωση της έκφρασης των mRNA, χωρίς όμως να μπορούν να διακριθούν οι στόχοι όπου η καταστολή της έκφρασης οφείλεται στην καταστολή της μετάφρασης από αυτούς που οφείλεται σε καταστολή της πρωτεϊνσύνθεσης. Για να εντοπιστούν αυτοί οι στόχοι αναπτύχθηκε μία καινούργια μέθοδος, γνωστή ως SILAC (Stable Isotope Labeling with Amino acids in Cell culture, **Εικ. 2.6**). Με αυτή τη μέθοδο αναπτύσσονται δύο κυτταρικοί πληθυσμοί σε δύο ίδια περιβάλλοντα, με τη διαφορά ότι το ένα περιλαμβάνει ένα “ελαφρύ” αμινοξύ ενώ το άλλο ένα “βαρύ” (για παράδειγμα  $^{12}\text{C}$  και  $^{13}\text{C}$  L-lysine). Όταν το αμινοξύ εισάγεται στο κύτταρο αντί του φυσικού αμινοξέος, ενσωματώνεται σε όλες τις πρωτεΐνες που έχουν δημιουργηθεί. Μετά από κάποιες διαιρέσεις του κυττάρου, κάθε στοιχείο από το συγκεκριμένο αμινοξύ αντικαθίσταται από το ανάλογο ισότοπο (για παράδειγμα deuterium,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ). Αυτό έχει σαν αποτέλεσμα τη μεταβολή της μάζας των συγκεκριμένων πεπτιδίων, η οποία εντοπίζεται μέσω ειδικών ανιχνευτών (<http://www.silac.org>).



**Εικ. 2.5:** Microarrays  
([http://www.liv.ac.uk/lmf/about\\_microarrays.htm](http://www.liv.ac.uk/lmf/about_microarrays.htm))



**Εικ. 2.6:** SILAC (Stable Isotope Labeling with Amino acids in Cell culture)  
(<http://en.wikipedia.org/wiki/File:Silac.gif>)

Η δεύτερη έμμεση μέθοδος είναι η ανοσοκατακρήμνιση, η οποία χρησιμοποιείται προκειμένου να απομονωθεί άμεσα ολόκληρο το σύμπλοκο miRNA-mRNA. Για να καταστεί δυνατό κάτι τέτοιο, χρησιμοποιείται ένα αντίσωμα που αναγνωρίζει το ένα από τα δύο μέρη. Αν για παράδειγμα, το αντίσωμα αναγνωρίζει το microRNA τότε το κατακρημνίζει και επειδή αποτελεί σύμπλοκο με το mRNA θα το έχει μαζί του, οπότε μπορεί να μελετηθεί (Alexiou *et al.* 2009).

Τέλος, η τρίτη μέθοδος είναι αυτή της ανάλυσης αλληλουχίας (sequencing), η οποία απομονώνει τα σύμπλοκα miRNA-mRNA, π.χ. με ανοσοκατακρήμνιση, και εν συνεχεία οι αλληλουχίες τους αναλύονται, προκειμένου να αναγνωριστούν και να ταυτοποιηθούν οι θέσεις πρόσδεσης του miRNA, με βάση την αρχή της συμπληρωματικότητας (Alexiou *et al.* 2009).

Οι έμμεσες μέθοδοι μπορούν να μας δώσουν ένα σύνολο από στόχους των microRNA, οι οποίοι είναι δύσκολο να ανιχνευτούν με τις άμεσες μεθόδους, αλλά, παρόλα αυτά, είναι λιγότερο εξειδικευμένες επειδή υπεισέρχονται σφάλματα (Alexiou *et al.* 2009).

Όλες οι παραπάνω πειραματικές μέθοδοι είναι ιδιαίτερα δαπανηρές και χρονοβόρες. Επίσης, η πλειοψηφία των γονιδίων-στόχων των microRNA παραμένει άγνωστη, με αποτέλεσμα η χρήση εξειδικευμένων αλγορίθμων να αποτελεί τη μόνη πηγή για γρήγορη και οικονομική αναγνώριση microRNA στόχων. Τα πρώτα προγράμματα αναγνώρισης στόχων microRNA δημιουργήθηκαν το 2003, όταν επιβεβαιώθηκε ότι τα microRNAs υπάρχουν στα κύτταρα. Τα τελευταία χρόνια έχουν δημοσιευτεί αρκετά προγράμματα πρόβλεψης γονιδιακών στόχων, χρησιμοποιώντας διάφορα κριτήρια. Μερικά από αυτά είναι τα ακόλουθα: 1) η δομική συμβατότητα, 2) η σύνθεση των νουκλεοτιδίων, 3) η θέση των σημείων πρόσδεσης στην 3'-αμετάφραστη περιοχή, ενώ το κυριότερο είναι 4) η εξέταση της συμπληρωματικότητας της ακολουθίας του microRNA με την 3' αμετάφραστη περιοχή των υποψήφιων γονιδιακών στόχων. Στην ενότητα που ακολουθεί, αναφέρονται οχτώ από τους πιο γνωστούς αλγορίθμους.

### 2.2.1 DIANA-microT (Εικ. 2.7, Εικ. 2.8)

Ο αλγόριθμος αυτός έχει πετύχει τη μέγιστη ακρίβεια στην αναγνώριση στόχων microRNA σε σύγκριση με τους υπόλοιπους αλγορίθμους. Βασίζεται σε παραμέτρους που υπολογίζονται ξεχωριστά για κάθε microRNA και για κάθε MRE (microRNA Recognition Element) με βάση χαρακτηριστικά ή ιδιότητες της μοριακής σύνδεσης του δεσμάτος και της εξελικτικής διατήρησης. Το σκορ της πρόβλεψης ενός πιθανού στόχου είναι ο σταθμισμένος μέσος των σκορ των διατηρούμενων και μη MREs ενός γονιδίου. Για κάθε αλληλεπίδραση υπολογίζεται ένας λόγος (SNR-Signal to Noise Ratio) και ένα σκορ ακρίβειας μέσω των οποίων βρίσκεται το ποσοστό των ψευδώς θετικών (false positive rate) αποτελεσμάτων για κάθε microRNA στόχο (Alexiou *et al.* 2009, Maragkakis *et al.*, 2009, <http://microrna.gr/microT>).

DIANA LAB DNA Intelligent Analysis

HOME SOFTWARE DATABASES MEMBERS PUBLICATIONS HELP

DIANA - microT v3.0

How to cite:

1/ M. Maragkakis, P. Alexiou, G. L. Papadopoulos, M. Recchi, T. Dalamagas, G. Giannopoulos, G. Georvas, E. Koukik, K. Kourtis, V. A. Sitsos, F. Selbach, T. Vergidis, N. Koutra, T. Selli, P. Tsanakas, A. G. Hatzigeorgiou. Accurate microRNA target prediction correlates with protein expression levels. *BMC Bioinformatics* 2009, 10:199 [View](#)

2/ M. Maragkakis, M. Recchi, V. A. Sitsos, P. Alexiou, G. L. Papadopoulos, T. Dalamagas, G. Giannopoulos, G. Georvas, E. Koukik, K. Kourtis, T. Vergidis, N. Koutra, T. Selli, P. Tsanakas, A. G. Hatzigeorgiou. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Research* 2009 Jul 1; 37(Web Server issue): W229-8. [View](#)

For issues regarding this application contact [maragkakis@iif.med.uoi.gr](mailto:maragkakis@iif.med.uoi.gr)

Download all microT v3.0 results

Search for targets of a specific miRNA

Insert a miRNA identifier (eg. hsa-let-7a)

OR predict for your miRNA sequence

Search for miRNA targets on a specific gene


Enter gene name (eg. ENSG00000159217, IGF2BP1, NM\_006546)  OR: Enter a gene description

Search for targets of a specific miRNA on a specific gene

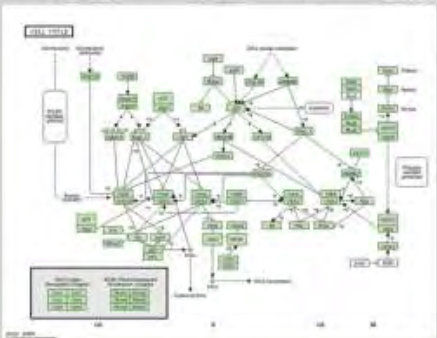
Insert a miRNA identifier (eg. hsa-let-7a)


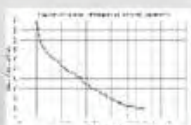
Enter gene name (eg. ENSG00000159217, IGF2BP1, NM\_006546)  OR: Enter a gene description

Εικ. 2.7: Φόρμα αναζήτησης του προγράμματος Diana MicroT 3.0 (<http://diana.cslab.ece.ntua.gr/microT/>)

**Ensembl Gene ID:** ENSG00000110092  
**Gene name:** CCND1  ①  
**Refseq ID:** NM\_053056  
**Description:** G1/S-specific cyclin-D1 (PRAD1 oncogene) (BCL-1 oncogene). [Source:UniProt/SWISSPROT;Acc:P24385]  
**KEGG pathways:** Cell cycle → p53 signaling pathway, Wnt signaling pathway, Focal adhesion, Jak-STAT signaling pathway, Colorectal cancer, Pancreatic cancer, Endometrial cancer, Glioma, Prostate cancer, Thyroid cancer, Melanoma, Bladder cancer, Chronic myeloid leukemia, Acute myeloid leukemia, Small cell lung cancer, Non-small cell lung cancer

**Species:** human  
**Chromosome:** 11



**miRNA name:** hsa-let-7b  ③ **Figure:**  ④

**Alternative Description:** MIMAT0000063  
**miRNA Sequence:** UGAGGUAGUAGGUUGUGUGUU

Rank	Gene name	miRNA name	⑤ miTG score	⑥ Precision	⑦ SNR
136	CCND1	hsa-let-7b	24.86	0.92	15

**PubMed Links:** miRNA | gene | both ⑧

⑨	Binding Type	UTR position	Score	Conservation	Also predicted by
<input type="checkbox"/>	7mer(pos 1)	2191 - 2219	8	9	VERIFIED pictar targetscan ⑩

⑪ **Position on Chromosome** 11: 69177423 - 69177451  
**Conservation Info** rn4, mm8, oryCun1, bosTau2, canFam2, loxAfr1, echTel1, monDom4, galGal2  
**3'UTR:** 5' U GUUU 3'  
 GCU CACAA UACCUCA  
 -1- ||||| |||||  
 UGG GUGUU AUGGAGU  
**miRNA:** 3' U U GUAUG 5'

<input type="checkbox"/>	7mer(pos 1)	2850 - 2878	5	5	VERIFIED pictar targetscan
--------------------------	-------------	-------------	---	---	----------------------------------

Εικ. 2.8: Αποτελέσματα του Diana microT 3.0 (Maragkakis et al. 2009)

### 2.2.2 EIMMo (Εικ. 2.9)

Ο αλγόριθμος αυτός χρησιμοποιεί μία γενική Bayesian μέθοδο, η οποία μετράει τη διατήρηση των θέσεων πρόσβασης με βάση ένα εξελικτικό μοντέλο, το οποίο αξιοποιεί την πιθανή φυλογενετική σχέση μεταξύ των διαφόρων ειδών (Alexiou *et al.* 2009, Gaidatzis *et al.*, 2007, <http://www.mirz.unibas.ch>).

**Εικ. 2.9:** EIMMo  
(<http://www.mirz.unibas.ch/EIMMo3/>)

### 2.2.3 miRanda (Εικ. 2.10)

Ο αλγόριθμος αυτός περιλαμβάνει δύο βήματα: α) Όλο το μήκος του microRNA στοιχίζεται με βάση την 3'-αμετάφραστη περιοχή. Οι στοιχίσεις που περιλαμβάνουν G:U ζευγάρια (wobble pairs) επισημαίνονται με μικρότερο βάρος. β) Υπολογίζεται η θερμοδυναμική σταθερότητα του συμπλόκου για τις στοιχίσεις με τα υψηλότερα σκορ (Alexiou *et al.* 2009, John *et al.* 2004, <http://www.microrna.org/>).





#### 2.2.4 miRBase (Εικ. 2.11)

Για να αναγνωρίσει πιθανές θέσεις πρόσδεσης για κάθε microRNA, ο miRBase χρησιμοποιεί τον miRanda. Χρησιμοποιεί δυναμικό προγραμματισμό ευθυγράμμισης για να εντοπίσει τις συμπληρωματικές θέσεις, απαιτείται αυστηρή συμπληρωματικότητα στην 5'-περιοχή και υπολογίζει την θερμοδυναμική σταθερότητα κάθε θέσης (Alexiou *et al.* 2009, Griffiths-Jones *et al.*, 2008, <http://microrna.sanger.ac.uk>).



### Search miRBase

#### By miRNA identifier or keyword

Enter a miRNA accession, name or keyword:

Submit Query Reset Example

#### By genomic location

Select organism, chromosome and start and end coordinates. Leave the start/end boxes blank to retrieve all miRNAs on the selected chromosome.

Choose species: Chr: Start: End: Get sequences

#### For clusters

Select organism and the desired inter-miRNA distance.

Choose species: Inter-miRNA distance: 10000 Get clusters

#### By tissue expression

Select organism and tissue.

Choose species: Select tissue: Get experiments

#### By sequence

##### Single sequence searches:

Paste a sequence here to search against miRNA sequences. You can choose to search against the intact precursor sequences or just the mature miRNAs. This search may take a few minutes. (Max size 1000bps)

Search sequences:

Mature miRNAs

Search method:

BLASTN

Choose BLASTN to search for a miRNA in a longer sequence. SSEARCH is useful for finding a short sequence within the library of miRNAs (for instance, find a short motif in a miRNA or precursor stem-loop, or find mature sequences that are related to your query).

E-value cutoff:

10

Maximum no. of hits:

100

Show results only from specific organisms:

human  mouse  worm  fly  Arabidopsis

or choose a taxonomic classification:

No species filter

Or: Select the sequence file you wish to use

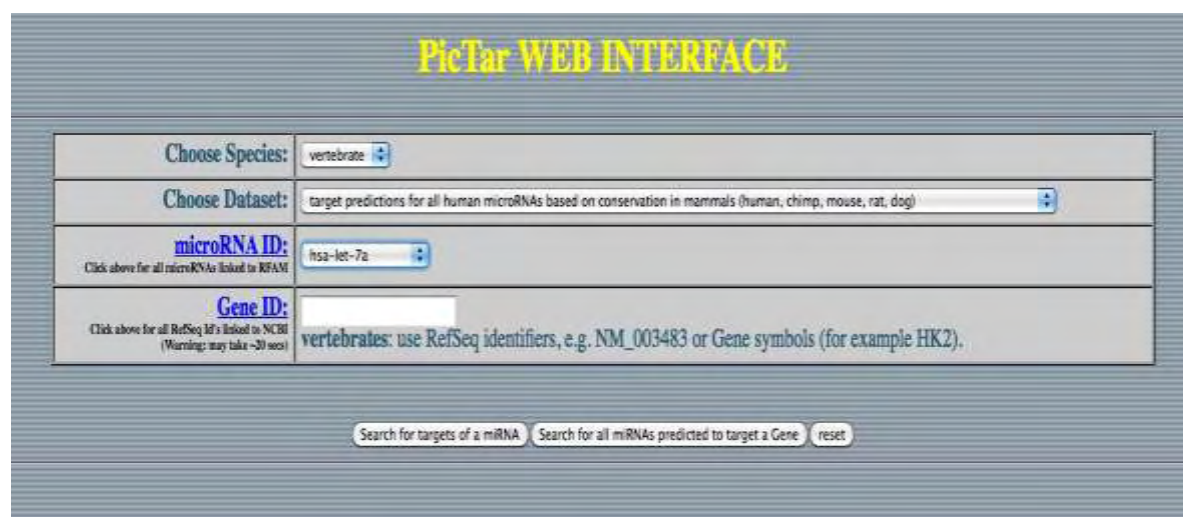
Browse...

Search miRNAs Reset Example

Fig. 2.11: MiRBase  
(<http://www.mirbase.org/search.shtml>)

### 2.2.5 Pictar (Εικ. 2.12)

Ο αλγόριθμος αυτός αναγνωρίζει δύο τύπους microRNA: α) αυτά που έχουν απόλυτη συμπληρωματικότητα μεταξύ της κεντρικής περιοχής (seed region) του microRNA και της 3'-αμετάφραστης περιοχής και β) αυτά των οποίων η απόλυτη συμπληρωματικότητα διακόπτεται από ζευγάρι G:U ή αν δεν υπάρχει συμπληρωματικότητα σε κάποια νουκλεοτιδική θέση (mismatch). Και στις δύο περιπτώσεις η θερμοδυναμική ενέργεια της ένωσης θα πρέπει να ξεπερνάει κάποιο κατώτατο όριο. Για κάθε αλληλεπίδραση microRNA με mRNA που εντοπίζεται, ο αλγόριθμος σημειώνει ως «άγκυρες» (anchors) τις υψηλά διατηρούμενες θέσεις στόχους και αφήνει εκτός εκείνες τις 3'-αμετάφραστες περιοχές που δεν έχουν ένα συγκεκριμένο αριθμό από άγκυρες. Χρησιμοποιεί τέλος, ένα μοντέλο Markov με το οποίο υπολογίζει την πιθανότητα μία 3'-αμετάφραστη περιοχή να αποτελεί στόχο με ένα συνδυαστικό τρόπο για ένα αριθμό διαφορετικών ειδών. Τα αποτελέσματα του Markov συνδυάζονται για να υπολογιστεί το τελικό σκορ (Alexiou *et al.* 2009, Lall *et al.*, 2006, <http://pictar.mdc-berlin.de>).



The image shows the PicTar WEB INTERFACE. At the top, the title "PicTar WEB INTERFACE" is displayed in yellow. Below the title is a search form with four rows of input fields:

Choose Species:	vertebrate
Choose Dataset:	target predictions for all human microRNAs based on conservation in mammals (human, chimp, mouse, rat, dog)
microRNA ID: <small>Click above for all microRNAs linked to Rfam</small>	hsa-let-7a
Gene ID: <small>Click above for all RefSeq IDs linked to NCBI (Warning: may take ~20 secs)</small>	vertebrates: use RefSeq identifiers, e.g. NM_003483 or Gene symbols (for example HK2).

At the bottom of the form, there are three buttons: "Search for targets of a miRNA", "Search for all miRNAs predicted to target a Gene", and "reset".

**Εικ. 2.12:** PicTar  
([http://pictar.mdc-berlin.de/cgi-bin/new\\_PicTar\\_vertibrate.cgi](http://pictar.mdc-berlin.de/cgi-bin/new_PicTar_vertibrate.cgi))

## 2.2.6 PITA (Εικ. 2.13)

Ο αλγόριθμος PITA λαμβάνει υπόψη την επίδραση που έχει ο βαθμός προσβασιμότητας της θέσης-στόχου από το microRNA, σε σχέση με την καταστολή της έκφρασης που προκαλεί το τελευταίο. Για κάθε θέση του στόχου, χρησιμοποιείται ένα ενεργειακό μέτρο που υπολογίζει τη διαφορά μεταξύ της ενέργειας που παράγεται από την ένωση του microRNA με το mRNA και της ενέργειας που χάνεται λόγω του μη ταιριάσματος των νουκλεοτιδίων (Alexiou *et al.* 2009, Kertesz *et al.*, 2007, <http://genie.weizmann.ac.il/pubs/mir07>).

The screenshot shows the web interface for the PITA online microRNA prediction tool. The header includes the Segal Lab of Computational Biology logo and navigation links: HOME, RESEARCH, PEOPLE, PUBLICATIONS, GENOMICS, SOFTWARE, INTERNAL. A sidebar on the left lists menu items: Main, Supplementary Data, Download Predictions, Search Predictions, Browse Predictions, Predict your UTR (highlighted), Download executable, and FAQ / Notes. The main content area is titled "Online microRNA prediction tool" and contains the following elements:

- A descriptive paragraph explaining the tool's function and its limitation to 100 targets per run.
- A dropdown menu for "Select an organism:" with "Human" selected.
- Input fields for "Paste UTR sequence (fasta format)" and "Or upload:" with a "Browse..." button and a "Paste your example" link.
- Radio buttons for "Use all known human microRNAs (470)" (selected) and "Upload microRNAs to use".
- A "Submit" button.
- An "Advanced Parameters" section with radio buttons for "Use flanking upstream sequence" (selected) and "Specify upstream sequence (optional, improves results accuracy)".
- Input fields for "Minimum seed size" (set to 8), "Allow single G-U" (checked), and "Allow single mismatch" (checked).
- A dropdown menu for "Flank settings" set to "No Flank".
- A note: "For more flexible seed-matching criteria (e.g. allow two G-U wobbles or loops) and flank settings please download and run the PITA executable."
- A footer note: "For problems and questions, please e-mail Eran Segal".

A vertical "SOFTWARE" label is visible on the right side of the interface.

Εικ. 2.13: PITA

([http://genie.weizmann.ac.il/pubs/mir07/mir07\\_prediction.html](http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html))

### 2.2.7 RNA 22 (Εικ. 2.14)

Ο συγκεκριμένος αλγόριθμος ενσωματώνει την αναγνώριση περιττών μοτίβων σε ώριμες ακολουθίες microRNA. Μία δεύτερης τάξης αλυσίδα Markov χρησιμοποιείται για να αναλύσει τα αποτελέσματα της αναγνώρισης. Έπειτα, γίνεται αναγνώριση του συμπληρώματος όλων των μοτίβων μέσα στην 3'-αμετάφραστη περιοχή. Η περιοχή της 3'-UTR, στην οποία υπάρχουν αρκετά συμπληρώματα μοτίβων ονομάζεται νησίδα-στόχος (Target Island). Τα microRNAs αντιστοιχίζονται με τέτοιες περιοχές και η δύναμη της αντιστοίχισης μετριέται βάσει της ενέργειας και του αριθμού των νουκλεοτιδίων που εμπλέκονται (Alexiou *et al.* 2009, Miranda *et al.*, 2006, <http://cbcsrv.watson.ibm.com/rna22.html>).

**IBM Research**

## RNA22 microRNA target detection

Input *single* microRNA sequence (fasta format):

Input *single* target sequence (fasta format):

Maximum number of allowed UN-paired bases  in seed/nucleus of  nucleotides

Minimum number of paired-up bases in heteroduplex:

Maximum folding energy for heteroduplex (Kcal/mol):

[License Terms](#)

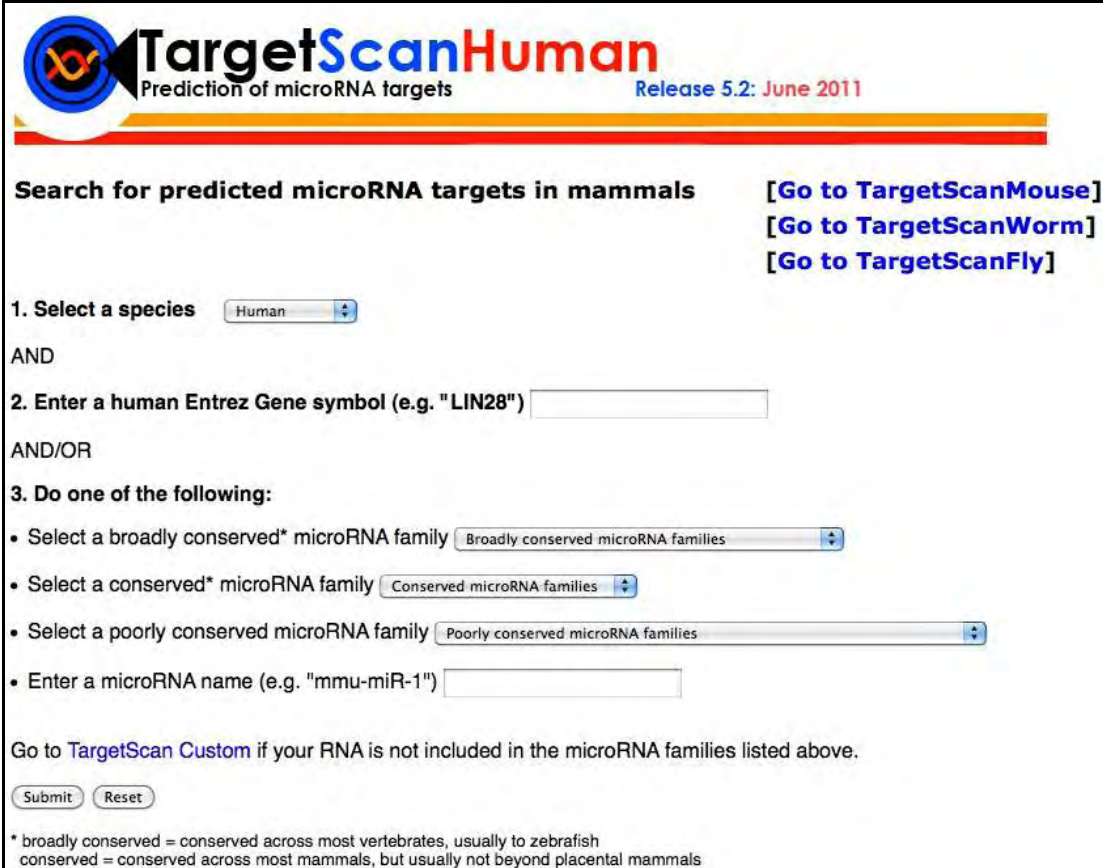
We gratefully acknowledge use of the Vienna RNA Package originally described in: Fast Folding and comparison of RNA secondary structures, Hofacker et al. *Monatshefte fuer Chemie*, Vol 125, 167-188, 1996.

Εικ. 2.14: RNA22

(<http://cbcsrv.watson.ibm.com/rna22.html>)

## 2.2.8 TargetScan 5.2 (Εικ. 2.15)

Ο TargetScan προβλέπει τους στόχους των microRNA βασιζόμενος στην αναγνώριση των στοιχισμένων ακολουθιών και την εξελικτική διατήρησή τους σε διάφορα είδη. Το σκορ ενός microRNA στόχου εξαρτάται από: α) το επίπεδο της διατήρησης, είτε αν δένεται στη θέση οχτώ του microRNA είτε αν έχει την αδενίνη στη θέση ένα, β) την απόσταση του στόχου από το τέλος της 3'-UTR και από την AU σύνθεση της περιοχής δεξιά και αριστερά του γονιδίου (flanking area) (Alexiou *et al.* 2009, Friedman *et al.*, 2009, <http://www.targetscan.org>).



The image shows the TargetScanHuman web interface. At the top left is the logo, a blue circle with a stylized 'X' and 'Y' inside. To its right is the text 'TargetScanHuman' in blue and red, with 'Prediction of microRNA targets' below it. Further right is 'Release 5.2: June 2011'. Below the header is a search section titled 'Search for predicted microRNA targets in mammals'. On the right of this section are three blue links: '[Go to TargetScanMouse]', '[Go to TargetScanWorm]', and '[Go to TargetScanFly]'. The main search area has three steps: 1. 'Select a species' with a dropdown menu showing 'Human'. 2. 'Enter a human Entrez Gene symbol (e.g. "LIN28")' with a text input field. 3. 'Do one of the following:' with four options: 'Select a broadly conserved\* microRNA family' (dropdown: 'Broadly conserved microRNA families'), 'Select a conserved\* microRNA family' (dropdown: 'Conserved microRNA families'), 'Select a poorly conserved microRNA family' (dropdown: 'Poorly conserved microRNA families'), and 'Enter a microRNA name (e.g. "mmu-miR-1")' (text input field). Below these is a link 'Go to TargetScan Custom if your RNA is not included in the microRNA families listed above.' At the bottom left are 'Submit' and 'Reset' buttons. At the bottom is a footnote: '\* broadly conserved = conserved across most vertebrates, usually to zebrafish; conserved = conserved across most mammals, but usually not beyond placental mammals'.

**Εικ. 2.15:** Target Scan  
(<http://www.targetscan.org/>)

### 3. ΑΝΑΛΥΣΗ ΚΑΙ ΤΕΧΝΙΚΕΣ ΑΠΑΙΤΗΣΕΙΣ

#### 3.1 Περιγραφή αρχιτεκτονικής συστήματος

Στο παρόν κεφάλαιο αναλύεται η δομή του προγραμματιστικού τμήματος της παρούσας εργασίας. Τα βήματα που πραγματοποιήθηκαν στην παρούσα διπλωματική εργασία σχετικά με την πρόβλεψη γονιδίων στόχων για ιικά microRNA σε ιούς και θηλαστικά είναι τα εξής:

- 1.Εξαγωγή των οργανισμών που χαρακτηρίζονται ως ιοί από ένα αρχείο που περιλαμβάνει διάφορους οργανισμούς.
- 2.Εξαγωγή ιικών microRNAs από ένα αρχείο, που περιλαμβάνει τα microRNAs διάφορων οργανισμών, χρησιμοποιώντας τα αποτελέσματα του προηγούμενου βήματος.
- 3.Εξαγωγή από ένα αρχείο των γονιδίων των ιών, καθώς επίσης και πληροφοριών για την κωδική περιοχή του κάθε γονιδίου και υπολογισμός της 3'-αμετάφραστης περιοχής.

Για την περάτωση της συγκεκριμένης διπλωματικής εργασίας χρησιμοποιήθηκε η γλώσσα προγραμματισμού PERL (<http://www.perl.org>), η οποία είναι μία υψηλού επιπέδου αντικειμενοστραφής γλώσσα προγραμματισμού. Αναπτύχθηκε το 1987 από τον Larry Wall, ενώ η έκδοση που χρησιμοποιήθηκε είναι η 5.10.0. Αυτό που διακρίνει την PERL είναι το γεγονός ότι είναι διαθέσιμη σχεδόν για όλα τα λειτουργικά συστήματα σε αντίθεση με άλλες γλώσσες προγραμματισμού.

Επίσης, χρησιμοποιήθηκε η διαθέσιμη βιβλιοθήκη της BioPerl ([http://www.bioperl.org/wiki/Main\\_Page](http://www.bioperl.org/wiki/Main_Page)), από την οποία αντλήθηκαν διαθέσιμες μέθοδοι, προκειμένου να εξαχθούν οι απαραίτητες πληροφορίες από τα διαθέσιμα αρχεία. Το υπολογιστικό σύστημα που χρησιμοποιήθηκε ήταν ένας apple iMac με τα χαρακτηριστικά που φαίνονται στην εικόνα (Εικ. 3.1).

Στο τέλος κάθε προγράμματος παρατίθεται και ο χρόνος εκτέλεσης του αντίστοιχου προγράμματος που έχει υπολογιστεί ως ο μέσος όρος δέκα μετρήσεων.





**Εικ. 3.1:** Χαρακτηριστικά υπολογιστικού συστήματος που χρησιμοποιήθηκε.

### 3.2 Ανάλυση προγραμμάτων

Στην ενότητα αυτή θα γίνει αναλυτική περιγραφή των προγραμμάτων που δημιουργήθηκαν για την διεξαγωγή κάθε βήματος για την προβλεψη στόχων ιικών microRNA.

1. Για το **πρώτο βήμα**, χρησιμοποιήθηκε το αρχείο "organisms.txt" (<http://www.mirbase.org/>), το οποίο περιλαμβάνει ονόματα από διάφορους οργανισμούς, μεταξύ των οποίων και ιούς (Εικ. 3.2). Το αρχείο χωρίζεται σε τέσσερις στήλες: η πρώτη και δεύτερη περιλαμβάνουν το αναγνωριστικό του κάθε οργανισμού (συμβολισμός του είδους), η τρίτη δίνει το επίσημο όνομα του οργανισμού και η τέταρτη την ταξονομική κατηγοριοποίηση (taxonomy) του οργανισμού (π.χ. ιός). Συγκεκριμένα, ένα παράδειγμα εγγραφής του αρχείου είναι το εξής : **ebv VRL Epstein Barr virus Viruses**, το οποίο αφορά σε έναν ιό που ανήκει στην οικογένεια των ερπητοϊών. Στόχος του πρώτου βήματος είναι η εξαγωγή και αποθήκευση σε ένα νέο αρχείο όλων των ιών.

aqu	AQU	<i>Amphimedon queenslandica</i>	Metazoa;Porifera;
nve	NVE	<i>Nematostella vectensis</i>	Metazoa;Cnidaria;
hna	HMA	<i>Hydra magnipapillata</i>	Metazoa;Cnidaria;
sko	SKO	<i>Saccoglossus kowalevskii</i>	Metazoa;Bilateria;Deuterostoma;Hemichordata;
spu	SPU	<i>Strongylocentrotus purpuratus</i>	Metazoa;Bilateria;Deuterostoma;Echinodermata;
cin	CIN	<i>Ciona intestinalis</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Urochordata;
csa	CSA	<i>Ciona savignyi</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Urochordata;
odi	ODI	<i>Oikopleura dioica</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Urochordata;
bfl	BFL	<i>Branchiostoma floridae</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Cephalochordata;
xla	XLA	<i>Xenopus laevis</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Amphibia;
xtr	XTR	<i>Xenopus tropicalis</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Amphibia;
gga	GGA	<i>Gallus gallus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Aves;
cfa	CFA	<i>Canis familiaris</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Carnivora;
mdo	MDO	<i>Monodelphis domestica</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Metatheria;
age	AGE	<i>Ateles geoffroyi</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Atelidae;
lla	LLA	<i>Lagothrix lagotricha</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Atelidae;
sla	SLA	<i>Saguinus labiatus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Cebidae;
ml	MML	<i>Macaca mulatta</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Cercopithecoidea;
mme	MME	<i>Macaca nemestrina</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Cercopithecoidea;
pb	PBI	<i>Pygathrix hietzi</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Cercopithecoidea;
gga	GGO	<i>Gorilla gorilla</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
hsa	HSA	<i>Homo sapiens</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
ppa	PPA	<i>Pan paniscus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
ppy	PPY	<i>Pongo pygmaeus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
ptr	PTR	<i>Pan troglodytes</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
ssy	SSY	<i>Symphalangus syndactylus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Hominoidea;
lca	LCA	<i>Iemur catta</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Primates;Lemnidae;
gan	GAN	<i>Ornithorhynchus anatinus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Prototheria;
cgr	CGR	<i>Cricetulus griseus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Rodentia;
mmu	MMU	<i>Mus musculus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Rodentia;
rno	RND	<i>Rattus norvegicus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Rodentia;
bta	BTA	<i>Bos taurus</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Ruminantia;
oar	OAR	<i>Ovis aries</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Ruminantia;
ssc	SSC	<i>Sus scrofa</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Mammalia;Suina;
dre	DRE	<i>Danio rerio</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Pisces;
fru	FRU	<i>Fugu rubripes</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Pisces;
tnt	TNI	<i>Tetraodon nigroviridis</i>	Metazoa;Bilateria;Deuterostoma;Chordata;Vertebrata;Pisces;
aga	AGA	<i>Anopheles gambiae</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
ame	AME	<i>Apis mellifera</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
bmo	BMD	<i>Boerbyx mori</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
dan	DAN	<i>Drosophila ananassae</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
der	DER	<i>Drosophila erecta</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
agr	DGR	<i>Drosophila grimshawi</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
dme	DME	<i>Drosophila melanogaster</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
dmo	DMD	<i>Drosophila mojavensis</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;
dpe	DPE	<i>Drosophila persimilis</i>	Metazoa;Bilateria;Ecdysozoa;Arthropoda;Hexapoda;

Εικ. 3.2: Αρχείο 'organisms.txt'.

Για την επεξεργασία του επιθυμητού αποτελέσματος ακολουθήθηκε η εξής διαδικασία:

- Άνοιγμα του αρχείου που περιλαμβάνει όλους τους οργανισμούς ("organisms.txt", <ftp://mirbase.org/pub/mirbase/CURRENT/>).
- Έλεγχος για το αν ο συγκεκριμένος οργανισμός ανήκει στην κατηγορία των ιών, αν δηλαδή περιλαμβάνεται στην τελευταία στήλη η λέξη "Viruses". Εφόσον πρόκειται για ιό, το αναγνωριστικό του ιού, με άλλα λόγια η τιμή που υπάρχει στην πρώτη στήλη, αποθηκεύεται σε μία δομή δεδομένων (πίνακας).
- Δημιουργία, τέλος, ενός καινούργιου αρχείου, που περιέχει όλα τα ζητούμενα αναγνωριστικά, προκειμένου να χρησιμοποιηθούν στο επόμενο βήμα (Εικ. 3.3).

Ο χρόνος εκτέλεσης του συγκεκριμένου προγράμματος είναι 0.01 δευτερόλεπτα.

```

bkn
ebv
hcmv
hiv1
hsv1
jcv
kshv
mcmv
mcy
mdv1
mdv2
mghv
rlcy
rrv
sv40
hsv2
hbv
iltv
hvt
bhv1

```

**Εικ. 3.3:** Αποτελέσματα πρώτου προγράμματος – αναγνωριστικά ιών.

2. Το **δεύτερο βήμα** περιλαμβάνει την εξαγωγή των ιικών microRNAs. Χρησιμοποιείται ένα αρχείο σε μορφή fasta (Εικ. 3.4) (<ftp://mirbase.org/pub/mirbase/CURRENT/>), στην οποία κάθε ακολουθία ξεκινάει με μια γραμμή, που περιέχει πληροφορίες για το microRNA και έπειτα ακολουθεί η αλληλουχία των νουκλεοτιδίων. Οι περιγραφικές γραμμές ξεκινούν πάντα με το σύμβολο ">" και ακολουθεί, στη συνέχεια, χωρίς την ύπαρξη κενού, το αναγνωριστικό της ακολουθίας. Είναι υποχρεωτικό όλες οι γραμμές του αρχείου να μην ξεπερνούν τους 80 χαρακτήρες. Για παράδειγμα:

```

>cel-let-7 MIMAT0000001 Caenorhabditis elegans let-7
UGAGGUAGUAGGUUGUAUAGUU.

```

Η πρώτη γραμμή είναι αυτή με τα σχόλια, εφόσον ξεκινάει με το χαρακτήρα ">". Στη συγκεκριμένη περίπτωση, περιγράφεται το microRNA cel-let-7 (το αναγνωριστικό του) το οποίο εμφανίζεται στο χρωμόσωμα X στον γεωσκόληκα *Caenorhabditis elegans* και εντάσσεται στην κατηγορία MIMAT0000001. Στη δεύτερη γραμμή ακολουθεί η αλληλουχία του microRNA UGAGGUAGUAGGUUGUAUAGUU.

```

>cel-let-7 MIMAT0000001 Caenorhabditis elegans let-7
UGAGGUAGUAGGUUGUUAUAGUU
>cel-let-7* MIMAT0015091 Caenorhabditis elegans let-7*
UGAACUAUGCAAUUUUUCUACCUUAC
>cel-lin-4 MIMAT0000002 Caenorhabditis elegans lin-4
UCCUUGAGACCUCAAGUGUGA
>cel-lin-4* MIMAT0015092 Caenorhabditis elegans lin-4*
ACACCUGGGCUCUCCGGGUAC
>cel-miR-1 MIMAT0000003 Caenorhabditis elegans miR-1
UGGAAUGUAAAAGAAGUAUGUA
>cel-miR-2 MIMAT0000004 Caenorhabditis elegans miR-2
UAUCACAGCCAGCUUUUGAUGUGC
>cel-miR-34 MIMAT0000005 Caenorhabditis elegans miR-34
AGGCAGUGUGGUUAGCUGGUU
>cel-miR-34* MIMAT0015093 Caenorhabditis elegans miR-34*
ACGGCUACCUUCACUGCCACCC
>cel-miR-35 MIMAT0000006 Caenorhabditis elegans miR-35
UCACCGGGUGGAAACUAGCAGU
>cel-miR-36 MIMAT0000007 Caenorhabditis elegans miR-36
UCACCGGGUGAAAAUUCGCAUG
>cel-miR-37* MIMAT0015094 Caenorhabditis elegans miR-37*
UGUGGGUGUCCGUUGCGGUGC
>cel-miR-37 MIMAT0000008 Caenorhabditis elegans miR-37
UCACCGGGUGAACACUUGCAGU
>cel-miR-38 MIMAT0000009 Caenorhabditis elegans miR-38
UCACCGGGAGAAAAACUGGAGU
>cel-miR-39 MIMAT0000010 Caenorhabditis elegans miR-39
UCACCGGGUGUAAAUCAGCUUG
>cel-miR-40 MIMAT0000011 Caenorhabditis elegans miR-40
UCACCGGGUGUACAUAGCUAA
>cel-miR-41 MIMAT0000012 Caenorhabditis elegans miR-41
UCACCGGGUGAAAAUACCUA
>cel-miR-42* MIMAT0015095 Caenorhabditis elegans miR-42*
GUGGGUGUUUGCUUUUUCGGUGAAG
>cel-miR-42 MIMAT0000013 Caenorhabditis elegans miR-42
UCACCGGGUUAACAUCUACAGA
>cel-miR-43 MIMAT0000014 Caenorhabditis elegans miR-43
UAUCACAGUUUACUUGCUGUCGC
>cel-miR-44 MIMAT0000015 Caenorhabditis elegans miR-44
UGACUAGAGACACAUUCAGCU
>cel-miR-45 MIMAT0000016 Caenorhabditis elegans miR-45
UGACUAGAGACACAUUCAGCU

```

Εικ. 3.4: Αρχείο σε μορφή fasta.

Για την εξαγωγή, λοιπόν, όλων των microRNAs ακολουθούνται τα εξής βήματα:

- Αρχικά, χρησιμοποιείται το αρχείο που περιέχει όλα τα microRNAs (“mature.fa”).
- Έπειτα, χρησιμοποιείται το αρχείο που δημιουργήθηκε στο προηγούμενο βήμα, το οποίο περιέχει όλους τους ιούς.
- Δημιουργείται ένα hash, το οποίο έχει ως keys τα ονόματα των ιών και αρχικοποιούνται οι τιμές του με ένα.
- Για κάθε γραμμή του αρχείου “mature.fa” ελέγχεται αν ξεκινάει με το σύμβολο “>”. Εάν ξεκινάει, τότε τίθεται σε μία μεταβλητή όλη η γραμμή χωρίς όμως το σύμβολο “>”.
- Από κάθε γραμμή του αρχείου και αφού έχει αφαιρεθεί το σύμβολο “>”, επιλέγεται η πρώτη στήλη (συγκεκριμένα μέχρι την

πρώτη "-" ) και αποθηκεύεται σε μία άλλη μεταβλητή. Έπειτα, ελέγχεται αν υπάρχει στο hash. Εάν υπάρχει, δημιουργείται ένα νέο hash, το οποίο στο τέλος, και αφού θα διατρέξει όλο το αρχείο, θα περιλαμβάνει όλα τα ονόματα των ιών (τα οποία υπάρχουν από το πρώτο hash) μαζί με τις αντίστοιχες ακολουθίες των microRNAs. Δημιουργείται ακόμα για κάθε ιό ένα ξεχωριστό αρχείο με όνομα το όνομα του ιού, το οποίο περιλαμβάνει τις ακολουθίες των microRNA του.

- Στο τέλος δημιουργείται ένα συγκεντρωτικό αρχείο για όλους τους ιούς (Εικ. 3.5).

Ο χρόνος εκτέλεσης του συγκεκριμένου προγράμματος είναι 0.17 δευτερόλεπτα.

```

>rlcv-miR-rL1-20-5p MIMAT0014065 Rhesus lymphocryptovirus miR-rL1-20-5p
UAAAGGGUAGUGUGGUUACACAG
>hsv2-miR-H23 MIMAT0014705 Herpes Simplex miR-H23
AGGCCGUGGAGCUUGCCAGC
>rlcv-miR-rL1-34-5p MIMAT0016960 Rhesus lymphocryptovirus miR-rL1-34-5p
UAGAUAGCGUAACAGGCAACUU
>mdv2-miR-M29-3p MIMAT0012545 Mareks disease miR-M29-3p
CAUAGUGAGGUACGUGUAGG
>mdv2-miR-M28 MIMAT0004474 Mareks disease miR-M28
UUUUCUCGACGCCUACCCUCGG
>kshv-miR-K12-5 MIMAT0002190 Kaposi sarcoma-associated miR-K12-5
UAGGAUGCCUGGAACUUGCCGGU
>hsv2-miR-H9-3p MIMAT0014352 Herpes Simplex miR-H9-3p
UUCCACCUCGGUUCUCCUC
>rlcv-miR-rL1-23-3p MIMAT0016946 Rhesus lymphocryptovirus miR-rL1-23-3p
UUAGUUGUCUGCACUGGAGAGU
>mdv1-miR-M11-3p MIMAT0005983 Mareks disease miR-M11-3p
UGAGUUAACUGGUCAGGGAAU
>mdv2-miR-M20 MIMAT0004462 Mareks disease miR-M20
UCAAGUACUGCGCAAGACCG
>mdv2-miR-M16 MIMAT0004455 Mareks disease miR-M16
AUCCAGUCUGUUUGGCAUCUGA
>mdv1-miR-M2-5p MIMAT0003921 Mareks disease miR-M2-5p
GUUGUAAUUCGCCCGUAGUCCG
>kshv-miR-K12-5* MIMAT0015218 Kaposi sarcoma-associated miR-K12-5*
AGGUAGUCCUGGUGCCUAAGG
>hsv2-miR-H11* MIMAT0014697 Herpes Simplex miR-H11*
AAGCGUUCGCACUUUGUCCUA
>hsv1-miR-H6-5p MIMAT0015281 Herpes Simplex miR-H6-5p
GGUGGAAGGCAGGGGGUGUA
>rlcv-miR-rL1-5-3p MIMAT0003433 Rhesus lymphocryptovirus miR-rL1-5-3p
CGCACCAUUUACUAGGUGU
>rrv-miR-rR1-5 MIMAT0004412 Rhesus monkey miR-rR1-5
CCGGAACCAAAGACACGUGCCG
>kshv-miR-K12-8* MIMAT0015216 Kaposi sarcoma-associated miR-K12-8*
ACUCCUCACUAACGCCCGCU
>hvt-miR-H14 MIMAT0012866 Herpesvirus of miR-H14
UCAUUCAGCGGCAAGUAGACUGU
>rrv-miR-rR1-4* MIMAT0004411 Rhesus monkey miR-rR1-4*
CUCGUUAACGCCUCCGAGA
>ebv-miR-BART8* MIMAT0003418 Epstein Barr miR-BART8*
GUCACAAUCUAUGGGGUCGUA
>ebv-miR-BART11-5p MIMAT0003421 Epstein Barr miR-BART11-5p
UCAGACAGUUUGGUGCGUAGUUG
>mcmv-miR-m01-3* MIMAT0005537 Mouse cytomegalovirus miR-m01-3*
CGAGGAACGCUCGCUACCGC
>mdv2-miR-M26* MIMAT0012544 Mareks disease miR-M26*
UCGGGCACCGCACCGAAGGAUG
>mdv1-miR-M12 MIMAT0005984 Mareks disease miR-M12
UGCAUAAUACGGAGGUUCU
>mdv1-miR-M2-3p MIMAT0003922 Mareks disease miR-M2-3p
CGGACUGCCGAGAAUAGCUU
>hsv1-miR-H13 MIMAT0014690 Herpes Simplex miR-H13
UUAGGGCGAAGUGCGAGCACUGG
>mcmv-miR-m88-1* MIMAT0005551 Mouse cytomegalovirus miR-m88-1*
AUGACCGACCCUAGACUCCG

```

Εικ. 3.5: Αποτελέσματα δεύτερου προγράμματος.

3. Στο **τρίτο βήμα** της διαδικασίας υπάρχει για κάθε ιό, ένα αρχείο genbank (Εικ. 3.6) από τη βάση NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov>). Το genbank αρχείο έχει την παρακάτω μορφή: αρχικά, έχει πληροφορίες σχετικά με την κατηγορία στην οποία ανήκει (π.χ. NC\_009334) ο ιός, πόσα ζευγάρια βάσεων έχει το γονιδίωμα (π.χ. 172764), αν είναι κυκλικός (π.χ. circular VRL), το όνομα του (π.χ. Human herpesvirus 4) και γενικές πληροφορίες για τους

συγγραφείς του αρχείου. Στο τέλος του αρχείου υπάρχει το γονιδίωμα του ιού, το οποίο βρίσκεται στην ετικέτα origin. Κάτω από τα αρχικά σχόλια η πρώτη ετικέτα (tag) που έχει είναι η source, η οποία δίνει το μήκος του γονιδιώματος. Μετά το source υπάρχουν όλα τα γονίδια του συγκεκριμένου ιού. Για να διαχωρίσει ότι είναι γονίδιο έχει μια ετικέτα gene. Σε αυτό το πεδίο δίνει πληροφορίες σχετικά με το που ξεκινάει και που τελειώνει το γονίδιο στο γονιδίωμα του ιού που μπορεί να είναι ένα απλό διάστημα, ένωση ή και συμπλήρωμα διαστημάτων (π.χ. 59..1680 ή join (167587..172764,1..1680) ή complement (43807..44763)), το όνομα του γονιδίου και κάποιες πληροφορίες για τη βάση δεδομένων. Μετά την ετικέτα gene, έχει ετικέτα cds για την κωδική περιοχή του γονιδίου. Σε αυτή την ετικέτα δίνει πληροφορίες για την αρχή και το τέλος της κωδικής περιοχής, που ακολουθούν τους ίδιους κανόνες με τα αντίστοιχα του γονιδίου, το όνομα του γονιδίου στο οποίο ανήκει η περιοχή, μερικές πληροφορίες για τη βάση δεδομένων, το αναγνωριστικό της πρωτεΐνης και τέλος την αλληλουχία της πρωτεΐνης. Στόχος είναι να εξαχθούν από αυτά τα αρχεία όλα τα γονίδια του ιού, η κωδική περιοχή κάθε γονιδίου, η αρχή και το τέλος τόσο του γονιδίου όσο και της κωδικής περιοχής, η αλληλουχία των νουκλεοτιδίων για την κωδική περιοχή και για το γονίδιο και ο υπολογισμός της 3' αμετάφραστη περιοχή. Στο τέλος, θα δημιουργηθεί ένα fasta αρχείο, το οποίο θα περιλαμβάνει το όνομα του κάθε γονιδίου, το strand, την κωδική περιοχή, την αρχή και το τέλος της κωδικής περιοχής, την 3' αμετάφραστη περιοχή, από που αρχίζει και που τελειώνει η 3' αμετάφραστη περιοχή και τέλος τις αναπαραστάσεις τους σε αλληλουχίες νουκλεοτιδίων. Ας αναλύσουμε τα βήματα που ακολουθήσαμε για να δημιουργήσουμε το συγκεκριμένο αρχείο.

- Αρχικά, χρησιμοποιείται το αρχείο genbank, από το οποίο θα εξαχθούν τα δεδομένα που έχουν αναφερθεί παραπάνω.
- Για κάθε γονίδιο του ιού δημιουργείται ένα αντικείμενο το οποίο έχει τα εξής χαρακτηριστικά : α) Το όνομα του γονιδίου, β) Το όνομα της κωδικής περιοχής, γ) την 3' αμετάφραστη περιοχή, δ) την αλληλουχία της κωδικής περιοχής, ε) την

αλληλουχία του γονιδίου, ζ) τα διαστήματα όπου βρίσκεται το γονίδιο στο γονιδίωμα, και η) τα διαστήματα όπου βρίσκεται η κωδική περιοχή στο γονιδίωμα. Αντικείμενα δημιουργούνται μόνο για γονίδια που έχουν και κωδική περιοχή, ειδάλλως αγνοούνται.

- Έχει κατασκευαστεί μία συνάρτηση parse, μέσω της οποίας διαβάζεται όλο το γονιδίωμα, το οποίο χρησιμοποιείται για την δημιουργία της αλληλουχίας του γονιδίου, της κωδικής περιοχής και της 3' αμετάφραστης περιοχής.
- Ακόμη, έχει υλοποιηθεί μία συνάρτηση reversecomplement, η οποία καλείται όταν το strand του γονιδίου είναι -1 και η οποία αντιστρέφει την αλληλουχία του γονιδίου.
- Έχει δημιουργηθεί και η συνάρτηση translateprotein, η οποία χρησιμοποιείται για να μεταφραστεί η αλληλουχία της κωδικής περιοχής σε πρωτεΐνη και να συγκριθεί με την υπάρχουσα μετάφραση που είναι διαθέσιμη στο αρχείο. Ο έλεγχος αυτός γίνεται ως επαλήθευση ότι έχουν παρθεί σωστά από το γονιδίωμα, τα νουκλεοτίδια της κωδικής περιοχής.
- Τέλος, δημιουργήθηκε ένα fasta αρχείο με την μορφή που φαίνεται παρακάτω:

```
">GENENAME_X|STRAND "
```

```
"CDS START1;START2; STOP1;STOP2;
```

```
"UTR3 START1;STOP1"
```

Ο χρόνος εκτέλεσης του συγκεκριμένου προγράμματος είναι 7.9 δευτερόλεπτα.



LOCUS	NC_009334	172764 bp	DNA	circular	VRL 26-MAR-2010
DEFINITION	Human herpesvirus 4, complete genome.				
ACCESSION	NC_009334				
VERSION	NC_009334.1 GI:139424470				
DBLINK	Project: 20959				
KEYWORDS	.				
SOURCE	Human herpesvirus 4 type 2 (Epstein-Barr virus type 2)				
ORGANISM	Human herpesvirus 4 type 2				
	Viruses; dsDNA viruses, no RNA stage; Herpesvirales; Herpesviridae; Gammaherpesvirinae; <u>Lymphocryptovirus</u> .				
REFERENCE	1 (bases 1 to 172764)				
AUTHORS	Dolan, A., Addison, C., Gatherer, D., Davison, A.J. and McGeach, D.J.				
TITLE	The genome of Epstein-Barr virus type 2 strain AG876				
JOURNAL	Virology 350 (1), 164-170 (2006)				
PUBMED	16490228				
REFERENCE	2 (bases 1 to 172764)				
AUTHORS	Dolan, A., Addison, C., Davison, A.J., Rickinson, A.B. and McGeach, D.J.				
TITLE	DNA sequence of Epstein-Barr virus type 2				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 172764)				
CONSTRM	NCBI Genome Project				
TITLE	Direct Submission				
JOURNAL	Submitted (03-APR-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA				
REFERENCE	4 (bases 1 to 172764)				
AUTHORS	McGeach, D.J. and Davison, A.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-MAY-2006) MRC Virology Unit, Church Street, Glasgow G11 5JR, UK				
REFERENCE	5 (bases 1 to 172764)				
AUTHORS	McGeach, D.J. and Davison, A.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-NOV-2005) MRC Virology Unit, Church Street, Glasgow G11 5JR, UK				
COMMENT	PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from DQ279927.				
	COMPLETENESS: full length.				
FEATURES	Location/Qualifiers				
source	1..172764				
	/organism="Human herpesvirus 4 type 2"				
	/mol_type="genomic DNA"				
	/strain="AG876"				
	/host="Homo sapiens"				
	/db_xref="taxon:12589"				
	/acronym="EBV"				
	/acronym="HHV-4"				
gene	join(167587..172764,1..1680)				
	/locus_tag="HHV4tp2_gp01"				
	/db_xref="GeneID:5176182"				
CDS	join(167587..167942,58..272,368..458,540..788,871..951,1026..1196,1280..1495,1574..1680)				
	/locus_tag="HHV4tp2_gp01"				
	/function="latency and B cell survival"				
	/note="multiple transmembrane protein; LMP-2A"				
	/codon_start=1				
	/product="K15"				
	/protein_id="YP_001129436.1"				
	/db_xref="GI:139424472"				
	/db_xref="GeneID:5176182"				
	/translation="MGSLEMVFMGAGPPSPGGDPDGGDGGNNSQYPSASGSSGNTPTP PNDEERESNEEPPPPYEDPYWGNDRHSQYQPLGTQQSLYLGLQHDGNDGLPPPPYS PRDOSSQHTYEEAGRSMNPVCLPVIVAPYLFMLAAIAASCFTASVTVVTATGLALS LLLLAAVASSYAAQRKLLTPVTYLVAVVTFFAICLTWRIEDPPFNLSLLFALLAAAGG LQGIYVLMVLLILAYRRRWRRLTVCGGIMFLACVLVIVDAVQLSPLLGAIVTVS MTLLAFVLMVLSPPGGLTGAALLTAAALALLASLILGTLNLTTFLLMLLWTLV VLLICSSCSCLSKILLARLFLYALALLSAL IAGGSILQTNFKSLSTEFIPNL FCMLLIVAGILFILAILTEWGSNRTYGPVFMCLGGLTMVAGAVMLVMTNTLLSA WLTAGFLIFLIGFALFGVIRCCRYCCYCLTLESEERPPTPYRNTV"				
gene	59..1680				
	/locus_tag="HHV4tp2_gp02"				
	/db_xref="GeneID:5176195"				
CDS	join(59..272,368..458,540..788,871..951,1026..1196,1280..1495,1574..1680)				
	/locus_tag="HHV4tp2_gp02"				
	/function="latency and B cell survival"				
	/note="N-terminally truncated form of LMP-2A; multiple transmembrane protein"				
	/codon_start=1				
	/product="LMP-2B"				
	/protein_id="YP_001129437.1"				
	/db_xref="GI:139424474"				
	/db_xref="GeneID:5176195"				
	/translation="MNPVCLPVIVAPYLFMLAAIAASCFTASVTVVTATGLALSLLL AAVASSYAAQRKLLTPVTYLVAVVTFFAICLTWRIEDPPFNLSLLFALLAAAGGLQ IYVLMVLLILAYRRRWRRLTVCGGIMFLACVLVIVDAVQLSPLLGAIVTVSMTL LLAFVLMVLSPPGGLTGAALLTAAALALLASLILGTLNLTTFLLMLLWTLVLL ICSSCSCLSKILLARLFLYALALLSAL IAGGSILQTNFKSLSTEFIPNLFCM LLLIVAGILFILAILTEWGSNRTYGPVFMCLGGLTMVAGAVMLVMTNTLLSAIL TAGFLIFLIGFALFGVIRCCRYCCYCLTLESEERPPTPYRNTV"				

Εικ. 3.6: Αρχείο σε μορφή genbank.

```

>GENENAME_X1STRAND
CDS      START1;START2  STOP1;STOP2  SEQUENCE
UTR3     START1  STOP1  SEQUENCE
>HHV4tp2_gp0111
CDS      167587;58;360;540;871;1026;1280;1574  167942;272;458;788;951;1196;1495;1680
atgggggtccctagaaatgggtgccaatgggagcgggtccccctagccccggcggggatccggatggggagcagatggcggaacaactcccaatatccatctgcttctggctcttctgggaacacccccaccacccaacgatgaggaacgtgaatctaatgaaga
gccccaccgccttatgaggaccatattggggcaatggcgaccgtcactcggactatcaaccactaggaacccaagatcaaatgctgtacttgggattgcaacacgacgggaatgacgggctccctccccctccctactctccacgggatgactcatctcaac
acatatacgaagaagcgggagaggaatgaatccagatgctgctgtaattgttgcgcctacctgttttggctggcggtattgcccctcgtgttccagggcctcagttagtaccgttgtgaccgccaccggcttggccctctcacttctactcttgg
cagcagtgggccagctcatatgccgtgcacaaggaactgctgacacgggtgacagttcttactgcggtgtcacttcttgcatttgcctaacatggaggattgaggaccacacctttaaattctcttctgtttgattgctggccgagctggcggaactac
aaggcatttacttctggtgatgcttgtgctcctgatactagcgtacagaaggagatggcgccgttggactgtttgtggcggcatcatgtttttggcatgtgtactgtccttatcgtcgacgctgttttgacgtgagtcctcctctggagctgaactgtgg
tttccatgacgctgctgctactggcttctgctcctggctctcttcgcccagggggcctaggtaactcttgggtgcagccctttaaactgtggcagcactctggcactgtagcgtcactgattttgggcacacttaactgactacaatgttctctcatgctcc
tatggacactttggttctctgatttgcctctcgtgctcttcatgtccactgagcaagatccttctggcacgactgttctatatgctctcgactcttgttgcagcctccgcgctaatcgtggtggcagatatttgcaaacaaacttcaagagtttaagca
gcactgaattataaccatttggcttctgcatgttattactgattgctgctggcactcttcttcttggcactcctgacccaatggggcagtggaatagaacatacggccagttttatgtgctcgggtggcctgctcaccatggtagcggcgctgtgtggc
tgacggtgatgactaacacgcttttctgctgctggattcttacagcaggattcctgatttctcattgctttggcctcttgggggtcattagatgctgccgctactgctgctactactgcttacactggaagtgaggagcggccaccgaccccatatcgca
acactgtata
UTR3     1680    1680    a

>HHV4tp2_gp0211
CDS      59;360;540;871;1026;1280;1574  272;458;788;951;1196;1495;1680
atgggggtccctagaaatgggtgccaatgggagcgggtccccctagccccggcggggatccggatggggagcagatggcggaacaactcccaatatccatctgcttctggctcttctgggaacacccccaccacccaacgatgaggaacgtgaatctaatgaaga
gccccaccgccttatgaggaccatattggggcaatggcgaccgtcactcggactatcaaccactaggaacccaagatcaaatgctgtacttgggattgcaacacgacgggaatgacgggctccctccccctccctactctccacgggatgactcatctcaac
acatatacgaagaagcgggagaggaatgaatccagatgctgctgtaattgttgcgcctacctgttttggctggcggtattgcccctcgtgttccagggcctcagttagtaccgttgtgaccgccaccggcttggccctctcacttctactcttgg
cagcagtgggccagctcatatgccgtgcacaaggaactgctgacacgggtgacagttcttactgcggtgtcacttcttgcatttgcctaacatggaggattgaggaccacacctttaaattctcttctgtttgattgctggccgagctggcggaactac
aaggcatttacttctggtgatgcttgtgctcctgatactagcgtacagaaggagatggcgccgttggactgtttgtggcggcatcatgtttttggcatgtgtactgtccttatcgtcgacgctgttttgacgtgagtcctcctctggagctgaactgtgg
tttccatgacgctgctgctactggcttctgctcctggctctcttcgcccagggggcctaggtaactcttgggtgcagccctttaaactgtggcagcactctggcactgtagcgtcactgattttgggcacacttaactgactacaatgttctctcatgctcc
tatggacactttggttctctgatttgcctctcgtgctcttcatgtccactgagcaagatccttctggcacgactgttctatatgctctcgactcttgttgcagcctccgcgctaatcgtggtggcagatatttgcaaacaaacttcaagagtttaagca
gcactgaattataaccatttggcttctgcatgttattactgattgctgctggcactcttcttcttggcactcctgacccaatggggcagtggaatagaacatacggccagttttatgtgctcgggtggcctgctcaccatggtagcggcgctgtgtggc
tgacggtgatgactaacacgcttttctgctgctggattcttacagcaggattcctgatttctcattgctttggcctcttgggggtcattagatgctgccgctactgctgctactactgcttacactggaagtgaggagcggccaccgaccccatatcgca
acactgtataatgaatccagatgctgctgtaattgttgcgcctacctgttttggctggcggtattgcccctcgtgttccagggcctcagttagtaccgttgtgaccgccaccggcttggccctctcacttctactcttggcagcagtgggcagctcat
atgcccgtgcacaagaactgctgacacgggtgacagttcttactgcggtgtcacttcttgcatttgcctaacatggaggattgaggaccacacctttaaattctcttctgtttgattgctggccgagctggcggaactacaaggcatttacttctgggt
gatgcttgtgctcctgatactagcgtacagaaggagatggcgccgttggactgtttgtggcggcatcatgtttttggcatgtgtactgtccttatcgtcgacgctgttttgacgtgagtcctcctctggagctgaactgtggtttccatgacgctgctgc
tactggcttctgctccttggctctcttcgcccagggggcctaggtaactcttgggtgcagccctttaaactgtggcagcactctggcactgtagcgtcactgattttgggcacacttaactgactacaatgttctctcatgctcctatggacactttggttct
cctgatttgccttctgctcttcatgtccactgagcaagatccttctggcacgactgttctatatgctctcgactcttgttgcagcctccgcgctaatcgtggtggcagatatttgcaaacaaacttcaagagtttaagcagcactgaattataacca
tttgttctgcatgttattactgattgctgctggcactcttcttcttggcactcctgacccaatggggcagtggaatagaacatacggccagttttatgtgctcgggtggcctgctcaccatggtagcggcgctgtgtggctgacgggtgatgactaaca
cgcttttctgctgctggattcttacagcaggattcctgatttctcattgctttggcctcttgggggtcattagatgctgccgctactgctgctactactgcttacactggaagtgaggagcggccaccgaccccatatcgcaacactgtata
UTR3     1680    1680    a

```

Εικ. 3.7: Αποτελέσματα τρίτου προγράμματος.

## 4. ΕΠΙΛΟΓΟΣ

### 4.1 Σύνοψη και συμπεράσματα

Καθώς η διεξαγωγή πειραμάτων είναι μια χρονοβόρα, ακριβή και επίπονη διαδικασία της βιολογικής έρευνας, η συμβολή της πληροφορικής σε αυτόν τον τομέα είναι ιδιαίτερα σημαντική. Έχουν, λοιπόν, αναπτυχθεί συστήματα λογισμικού για την επεξεργασία των βιολογικών δεδομένων, τα οποία αποτελούν απαραίτητα εργαλεία στα χέρια των ερευνητών, μέσω των οποίων μπορούν να πραγματοποιηθούν με εύκολο και γρήγορο τρόπο μια σειρά από μελέτες ή πειράματα, που θα ήταν δύσκολο ή/και αδύνατο να ολοκληρωθούν χωρίς τα συγκεκριμένα λογισμικά.

Μέσα στο πλαίσιο αυτό, στην παρούσα διπλωματική εργασία αναπτύχθηκε ένα τμήμα της πρόβλεψης των γονιδιακών στόχων των ιϊκών microRNA σε ιούς και θηλαστικά. Το κοινό στο οποίο απευθύνεται αποτελείται από ερευνητές που ασχολούνται με τη μοριακή βιολογία, καθώς επίσης και ερευνητές του κλάδου της πληροφορικής, που επιθυμούν να συνδράμουν στον τομέα της μοριακής βιολογίας.

Οι στόχοι που επιτεύχθηκαν κατά την ανάπτυξη των συγκεκριμένων προγραμμάτων παρουσιάζονται επιγραμματικά παρακάτω:

- Υλοποίηση ενός προγράμματος εξαγωγής των ιών από ένα αρχείο με διάφορους οργανισμούς.
- Δημιουργία ενός προγράμματος, μέσω του οποίου εντοπίζονται τα ιϊκά microRNAs, χρησιμοποιώντας τα αποτελέσματα του προηγούμενου βήματος.
- Τέλος, η συγγραφή κώδικα, ο οποίος εξάγει από αρχεία ιών (μορφής genbank) τα γονίδια του κάθε ιού, καθώς και πληροφορίες για την κωδική και την αμετάφραστη περιοχή του.

Για την υλοποίηση των παραπάνω βημάτων μελετήθηκαν μία σειρά από σύγχρονες μεθοδολογίες ανάπτυξης λογισμικού, οι οποίες και εφαρμόστηκαν

για την επίτευξη ενός καλύτερου αποτελέσματος. Τέλος, ο σωστός σχεδιασμός των παραπάνω βημάτων δίνει τη δυνατότητα, με την υλοποίηση μερικών ακόμα προγραμμάτων, να δημιουργηθεί ένας αλγόριθμος πρόβλεψης γονιδιακών στόχων ιϊκών microRNA που θα αποτελέσει ένα σημαντικό εργαλείο για την αντιμετώπιση σοβαρών ασθενειών, όπως ο καρκίνος.

#### **4.2 Μελλοντικές επεκτάσεις**

Τα προγράμματα που δημιουργήθηκαν στα πλαίσια της παρούσας διπλωματικής εργασίας αποτελούν ένα σημαντικό κομμάτι της πρόβλεψης γονιδιακών στόχων των ιϊκών microRNAs. Βασική επιδίωξη είναι να δημιουργηθεί και ένα πρόγραμμα, το οποίο να εντοπίζει τους στόχους χρησιμοποιώντας τα αποτελέσματα των προγραμμάτων αυτής της διπλωματικής. Με την ολοκλήρωση του συγκεκριμένου αλγορίθμου θα υπάρχει η δυνατότητα πρόβλεψης στόχων, παρακάμπτοντας την ανάγκη πραγματοποίησης χρονοβόρων και δαπανηρών πειραμάτων, προκειμένου να διεξαχθούν συμπεράσματα, που μπορούν να συνδράμουν στην αντιμετώπιση σοβαρών ασθενειών

Αφού ολοκληρωθεί ο αλγόριθμος, μία πιθανή προγραμματιστική επέκταση περιλαμβάνει την υλοποίηση στα πλαίσια της αντικειμενοστρέφειας του προγράμματος που παρουσιάζει όλα τα γονίδια ενός ιού. Συγκεκριμένα, για κάθε γονίδιο υπάρχει ένα αντικείμενο, το οποίο θα μπορούσε να υλοποιηθεί σε ένα διαφορετικό πρόγραμμα μαζί με τις μεθόδους που χρειάζεται (για παράδειγμα μέθοδος για να επιστρέφεται το όνομα του γονιδίου, η αλληλουχία της κωδικής περιοχής), δίνοντας τη δυνατότητα διαχωρισμού του κώδικα του αντικειμένου από τον κώδικα ανάγνωσης και επεξεργασίας του αρχείου.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Alberts, B., D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, 2000.** *Βασικές Αρχές Κυτταρικής Βιολογίας: Εισαγωγή στη Μοριακή Βιολογία του Κυττάρου.* Αθήνα: Ιατρικές εκδόσεις Π.Χ. Πασχαλίδης.
- Alexiou, P. M. Maragkakis, G. L. Papadopoulos, M. Reczko and A. Hatzigeorgiou, 2009.** Lost in translation: an assessment and perspective for computational microRNA target identification. In: *Bioinformatics*, 25(23).
- Esquela-Kerscher, A. & F. J. Slack 2006.** Oncomirs — microRNAs with a role in cancer. In: *Nature Reviews Cancer* 6: 259-269.
- He, L. and Hannon, G.J. 2004.** MicroRNAs: Small RNAs with a big role in gene regulation. In: *Nat. Rev. Genet.* 5(7): 522--531.
- Λογοθέτη, Σ., Α. Γιαρίκα, Δ. Μήκας, Σ. Βλαχόπουλος και Β. Ζουμπουρλής, 2006.** Ογκογόνα Μικρο - RNAs (oncomirs) : microRNAs με Ρόλο στον Καρκίνο. Στο: *Ογκολογική Ενημέρωση* 8:4, 220-226.
- Maragkakis, M., P. Alexiou, G. L. Papadopoulos, M. Reczko, T. Dalamagas, G. Giannopoulos, G. Goumas, E. Koukis, K. Kourtis, V. A. Simossis, P. Sethupathy, T. Vergoulis, N. Koziris, T. Sellis, P. Tsanakas and A. Hatzigeorgiou, 2009.** Accurate microRNA target prediction correlates with protein repression levels. In: *BMC Bioinformatics* 10:295.
- Σουρβίνου, Ι. Σ., Α. Μάρκου, Λ. Κακλαμάνης, Α. Γ. Τσαρούχα και Ε. Σ. Λιανίδου, 2010.** Επίδραση του microRNA-21 στα επίπεδα έκφρασης της πρωτεΐνης PDCD4 στο μη μικροκυτταρικό καρκίνο του πνεύμονα. 9<sup>ο</sup> Πανελληνικό Συνέδριο Κλινικής Χημείας, Οκτώβριος 2010, Αθήνα.
- Sun Lee, Y. And A. Dutta 2009.** MicroRNAs in cancer. In: *Annu Rev Pathol.* 4: 199–227.
- Tisdall, J. 2001.** *Beginning Perl for Bioinformatics.* O'Reilly publisher.
- Τσίντου, Μ. 2011.** Καρκινικά βλαστικά κύτταρα και microRNAs σε εγκεφαλικούς όγκους. Στο: *Helmedica* 8 (Άρθρο 5).