



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικόν και Καποδιστριακόν  
Πανεπιστήμιον Αθηνών

— ΙΔΡΥΘΕΝ ΤΟ 1837 —

Σχολή Θετικών Επιστημών - Τμήμα Βιολογίας

Τομέας Βιοχημείας & Μοριακής Βιολογίας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Προσδιορισμός εναλλακτικών 3'-αμετάφραστον περιοχών (3'-UTRs) των  
ανθρώπινων γονιδίων της οικογένειας των καλλικρεϊνών, κατόπιν  
ανάλυσης δεδομένων αλληλούχησης νέας γενιάς (NGS)**

ΜΑΡΙΑ ΜΟΡΡΟΥ

Επιβλέπων καθηγητής: Χρήστος Κοντός

Αθήνα 2019



## Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στον Τομέα Βιοχημείας και Μοριακής Βιολογίας του Τμήματος Βιολογίας της Σχολής Θετικών Επιστημών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών (Ε.Κ.Π.Α.), κατά το ακαδημαϊκό έτος 2018-2019 , με επιβλέπων τον Επίκουρο Καθηγητή Μοριακής Βιολογίας Χρήστο Κοντό.

Πρωτίστως, θα ήθελα να ευχαριστήσω ολόψυχα τον Επίκ. Καθηγητή Χρήστο Κοντό και επιβλέποντα της παρούσας διπλωματικής εργασίας, για την ευκαιρία που μου προσέφερε, την εμπιστοσύνη που έδειξε στο πρόσωπό μου καθώς και την επιστημονική καθοδήγηση και υποστήριξή του, καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας.

Επιπλέον, ευχαριστώ θερμά τον Καθηγητή Ανδρέα Σκορίλα και διευθυντή του Τομέα Βιοχημείας και Μοριακής Βιολογίας, για την εμπιστοσύνη που μου έδειξε και την ευκαιρία που μου δόθηκε, να εργαστώ στον Τομέα και να υλοποιήσω την παρούσα διπλωματική εργασία.

Ακόμη, ένα μεγάλο ευχαριστώ στον διδάκτορα Παναγιώτη Αδαμόπουλο, για τη συνεισφορά του, τις πολύτιμες συμβουλές του και την επιστημονική υποστήριξή του κατά την εκπόνηση της παρούσας διπλωματικής εργασίας.

Επιπρόσθετα, εκφράζω τις ειλικρινείς ευχαριστίες μου σε όλους τους συναδέλφους που γνώρισα κατά τη διάρκεια της διπλωματικής μου εργασίας, για την άψογη συνεργασία, την υποστήριξή τους και πέραν του επαγγελματικού, προπάντων του φιλικού ενδιαφέροντος, το οποίο υπήρξε.

Τέλος, ένα μεγάλο ευχαριστώ στους φίλους μου, για όλη την αγάπη και υποστήριξή τους κατά τη διάρκεια της ακαδημαϊκής αυτής χρονιάς. Πάνω από όλα, θα ήθελα να εκφράσω την ευγνωμοσύνη και βαθιά εκτίμηση που έχω στους γονείς μου, Χριστούλλα και Ιωάννη, αλλά και στα δύο μου αδέρφια, Νατάσσα και Σάββα, για την απεριόριστη αγάπη και στήριξη την οποία μου έχουν προσφέρει.

## Πίνακας περιεχομένων:

1.	Εισαγωγή.....	1
1.1.	Εναλλακτική συρραφή.....	1
1.1.1.	Μηχανισμός εναλλακτικής συρραφής.....	2
1.1.2.	Βασικοί τύποι εναλλακτικής συρραφής.....	2
1.1.3.	Η σημασία της εναλλακτικής συρραφής.....	4
1.1.4.	Εναλλακτική συρραφή και ασθένειες.....	5
1.1.5.	Εναλλακτική συρραφή και καρκίνος.....	6
1.2.	Ρύθμιση έκφρασης από 3' αμετάφραστες περιοχές.....	7
1.2.1.	Ρύθμιση από AU-πλούσια στοιχεία.....	7
1.2.2.	Ρύθμιση από GU-πλούσια στοιχεία, CU-πλούσια, στοιχεία ελέγχου διαφοροποίησης και CA-πλούσια στοιχεία.....	8
1.2.3.	Ρύθμιση από στοιχεία απόκρισης σε σίδηρο και στοιχεία SECIS.....	9
1.2.4.	Ρύθμιση από RNA-binding πρωτεΐνες.....	9
1.2.5.	Ρύθμιση από microRNA .....	10
1.3.	Καλλικρεΐνες.....	11
1.3.1.	Δομή γονιδίων και πρωτεϊνών των καλλικρεΐνών.....	11
1.3.2.	Ρύθμιση έκφρασης των καλλικρεΐνών.....	13
1.3.3.	Εναλλακτικά μετάγραφα των καλλικρεΐνών.....	14
1.3.4.	Καλλικρεΐνες και ανθρώπινη παθογένεια.....	14
1.3.5.	Καλλικρεΐνες και καρκινικοί δείκτες.....	15
1.4.	Αλληλούχηση DNA.....	16
1.4.1.	Αλληλούχηση πρώτης γενιάς.....	17
1.4.2.	Αλληλούχηση νέας γενιάς.....	19
1.4.3.	Εργαλεία της αλληλούχησης νέας γενιάς.....	20
1.4.4.	Πλατφόρμες αλληλούχησης νέας γενιάς.....	21
2.	Υλικά και Μέθοδοι.....	23
2.1.	Βιολογικό Υλικό.....	23
2.2.	Φασματοφωτομετρικός προσδιορισμός του απομονωμένου ολικού RNA.....	23
2.3.	Αντίστροφη Μεταγραφή (Reverse Transcription, RT).....	24
2.4.	Βιοπληροφορική ανάλυση.....	26
2.5.	Έλεγχος έκφρασης γονιδίων μέσω ποσοτικής PCR .....	28

2.6	Μεθοδολογία SYBR-Green I για Real Time PCR.....	30
2.7	Αλυσιδωτή αντίδραση πολυμεράσης (PCR).....	32
2.8	Σχεδιασμός εκκινητών για PCR.....	33
2.9	3' Rapid amplification of cDNA ends (3' RACE).....	34
2.10	Nested PCR.....	35
2.11	Ενίσχυση εναλλακτικών 3' UTRs των ανθρώπινων γονιδίων <i>KLKs</i> με χρήση nested 3'RACE.....	35
2.12	Συνθήκες αντιδράσεων nested 3' RACE για τα ανθρώπινα γονίδια των <i>KLKs</i> .....	37
2.13	Ηλεκτροφόρηση των προϊόντων nested 3' RACE σε πήκτωμα αγαρόζης.....	40
2.14	Απομόνωση και καθαρισμός των nested 3' RACE προϊόντων από το πήκτωμα αγαρόζης και αλληλούχιση κατα Sanger.....	40
3.	Αποτελέσματα.....	42
3.1	Αποτελέσματα από τη βιοπληροφορική ανάλυση.....	42
3.2	Αποτελέσματα από αντιδράσεις με χρήση μιγμάτων cDNA.....	42
3.3	Αποτελέσματα από αντιδράσεις με χρήση cDNA από ανεξάρτητες κυτταρικές σειρές.....	44
3.4	Αποτελέσματα από εύρεση miRNA μορίων με χρήση αλγορίθμου miRDB.....	54
4.	Συζήτηση αποτελεσμάτων και συμπεράσματα.....	57
	Βιβλιογραφία.....	60
	Περίληψη.....	63

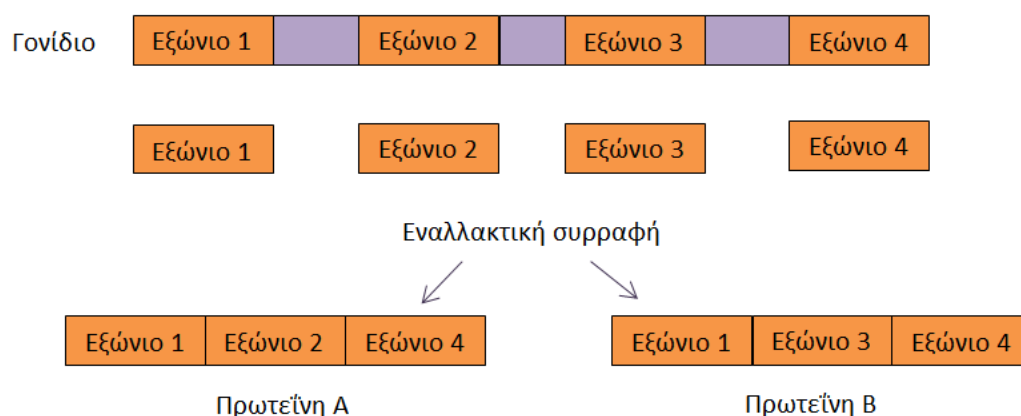


## 1.Εισαγωγή

### 1.1. Εναλλακτική συρραφή

Η εναλλακτική συρραφή ή μάτισμα (alternative splicing) θεωρείται θεμελιώδης μοριακός μηχανισμός για την ρύθμιση της έκφρασης των ευκαρυωτικών γονιδίων [1]. Η πλειοψηφία των γονιδίων στους ευκαρυωτικούς οργανισμούς εκφράζεται ως πρόδρομα mRNA μετάγραφα (pre-mRNAs) τα οποία και μετατρέπονται σε ώριμα mRNAs μέσω της διαδικασίας εναλλακτικού ματίσματος, η οποία πραγματοποιείται με μεγάλη ακρίβεια. Κατα τη διαδικασία αυτή γίνεται απομάκρυνση των μη κωδικών – κυρίως μεγάλου μήκους – αλληλουχιών που περιλαμβάνονται στα pre-mRNAs, τα εσώνια ( ή ιντρόνια ) και ακολουθεί συρραφή των κωδικών αλληλουχιών, τα λεγόμενα εξώνια, σχηματίζοντας το ώριμο mRNA [2].

Με την εφαρμογή της εναλλακτικής συρραφής, δηλαδή με τον σχηματισμό mRNAs που διαφέρουν είτε στις αμετάφραστες περιοχές (UTRs) είτε στις κωδικές περιοχές τους, επιτυγχάνεται η έκφραση πολλών συγγενών πρωτεϊνών (πρωτεϊνικές ισομορφές) που έχουν διαφορετικές βιολογικές λειτουργίες μεταξύ τους [3]. Οι πρωτεϊνικές αυτές ισομορφές μπορεί να διαφέρουν σε δομή, λειτουργία και θέση έκφρασης ή να απαιτούνται σε διαφορετικά αναπτυξιακά στάδια ή διαφορετικούς τύπους κυττάρων (**Εικόνα 1.1**). Ο μηχανισμός αυτός δίνει ένα εξαιρετικά σημαντικό πλεονέκτημα, τη δυνατότητα από ένα περιορισμένο αριθμό γονιδίων να μπορεί να κωδικοποιηθεί ένας σημαντικά αυξημένος αριθμός πρωτεϊνών, έτσι δεν απαιτούνται νέα γονίδια για κάθε πρωτεΐνη [4].



**Εικόνα 1.1** Ο σχηματισμός πρωτεϊνικών ισομορφών ως αποτέλεσμα της εναλλακτικής συρραφής

Στον άνθρωπο, ο αριθμός mRNA μεταγράφων που εκφράζονται είναι σημαντικά μεγαλύτερος από τον αριθμό γονιδίων, λόγω εναλλακτικού ματίσματος. Η τεχνολογία της αλληλούχησης υψηλής απόδοσης έχει αποκαλύψει ότι παραπάνω από το 90% των ανθρώπινων γονιδίων υποβάλλονται σε εναλλακτικό μάτισμα. Η επίδραση εναλλακτικού ματίσματος εκτείνεται και σε μη κωδικά mRNAs (non-coding RNAs, ncRNAs) τα οποία δέσμευονται σε παράγοντες μεταγραφής, έτσι οι διαφορετικές ισομορφές τους μπορούν να ρυθμίσουν τη δραστηριότητα ενός μεταγραφικού παράγοντα [5] [6].

Η πολύπλοκη αυτή διαδικασία της εναλλακτικής συρραφής πραγματοποιείται από τα σωματίδια συρραφής (spliceosomes), συγκροτήματα δηλαδή πρωτεϊνών και μικρών μορίων RNA. Τα σωματίδια αυτά μπορούν να αναγνωρίζουν ειδικά σήματα στο νεοσυντιθέμενο RNA τα οποία καθορίζουν τις θέσεις συρραφής. Γνωρίζουμε ότι τα εσώνια αρχίζουν σχεδόν πάντα με ζεύγος νουκλεοτιδίων GU και τελειώνουν με ζεύγος νουκλεοτιδίων AG, πριν από το οποίο υπάρχει αλληλουχία πλούσια σε πυριμιδίνες. Αυτό το μοτίβο αποτελεί ένα μέρος του σήματος για τα σωματίδια συρραφής

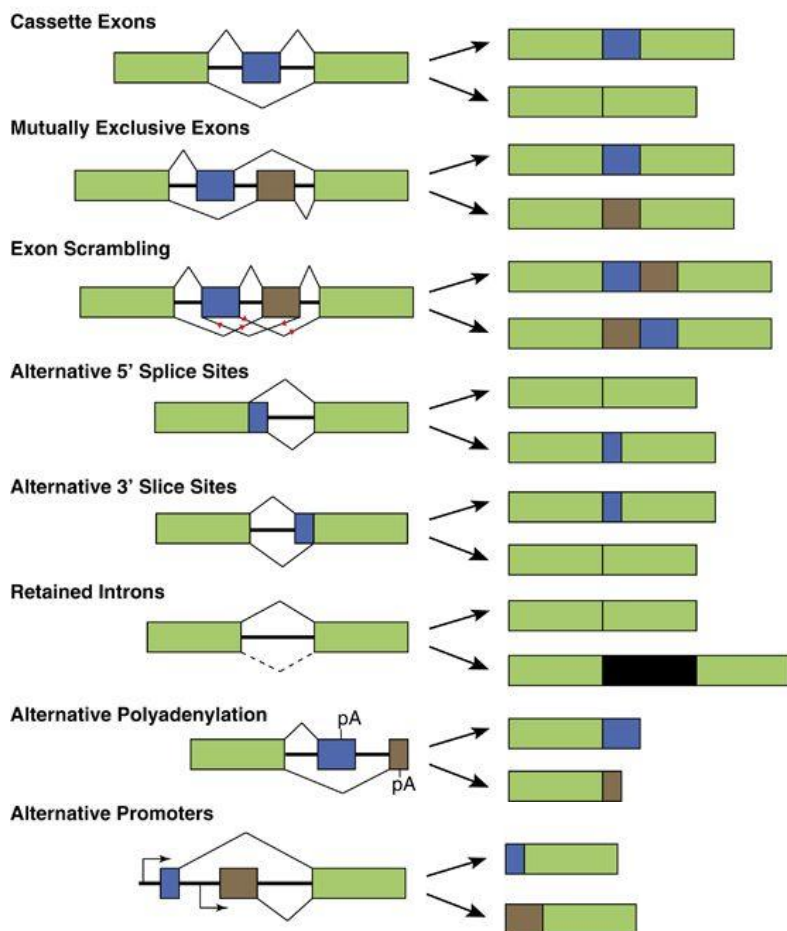
#### 1.1.1. Μηχανισμός εναλλακτικής συρραφής

Καθοριστικός παράγοντας για την οριοθέτηση των θέσεων συρραφής σε ένα εσώνιο ενός ευκαριωτικού γονιδίου είναι η θέση και η αλληλουχία cis- ρυθμιστικών στοιχείων, τα οποία αποτελούν νουκλεοτιδικές αλληλουχίες του πρόδρομου RNA μορίου. Ο μηχανισμός της συρραφής που εμπερικλείει την απομάκρυνση εσώνιων, καταλύεται από ένα υπερμοριακό σύμπλεγμα 200 τουλάχιστον πρωτεϊνών και πέντε μικρών ριβονουκλεοπρωτεϊνικών σωματιδίων (τα snRNPs) τα U1, U2, U4, U5 και U6, το οποίο ονομάζουμε σωματίδιο συρραφής (spliceosome) [7]. Η συναρμολόγηση αυτού του υπερμοριακού συμπλέγματος από τα ριβονουκλεοπρωτεϊνικά σωματίδια πραγματοποιείται κατά τα διάφορα στάδια του pre-mRNA μορίου.

#### 1.1.2 Βασικοί τύποι εναλλακτικής συρραφής

Ο μηχανισμός εναλλακτικής συρραφής δίνει τη δυνατότητα σύνθεσης διαφόρων ισομορφών μορίων mRNA τα οποία προκύπτουν από μόνο ένα γονίδιο. Από τη διαδικασία αυτή διακρίνονται αρκετοί τύποι εναλλακτικού ματίσματος, από τους οποίους οι βασικοί τύποι απεικονίζονται στην **Εικόνα 1.2**.





**Εικόνα 1.2** Οι βασικοί τύποι εναλλακτικού ματίσματος [7].

Ο πιο συχνά εμφανιζόμενος τύπος συρραφής στους ανώτερους ευκαρυωτικούς οργανισμούς είναι τα εξώνια-κασέτες (cassette exons) τα οποία είναι διακριτά εξώνια που μπορούν ανεξάρτητα να εμπεριέχονται ή όχι στο ώριμο μόριο mRNA, κατά τα αντίστοιχα γεγονότα εναλλακτικού ματίσματος. Ένας επίσης συχνά εμφανιζόμενος τύπος είναι το εναλλακτικό μάτισμα αμοιβαίου αποκλεισμού (mutually exclusive splicing), κατά τον οποίο μόνο ένα από τα δύο εναλλακτικά εξώνια εμπεριέχεται κάθε φορά στο ώριμο mRNA. Ακολούθως, δύο επίσης βασικοί τύποι εναλλακτικής συρραφής είναι η χρήση εναλλακτικού 5' ή 3' σημείου συρραφής εντός ενός εξωνίου (alternative 5' or 3' splice sites). Τα εναλλακτικά 5' και 3' σημεία συρραφής εντός των εξωνίων συχνά δημιουργούν mRNA μετάγραφο με επέκταση των εξωνίων (exon extensions) ή/και συντμήσεις τους. Ένας λιγότερο συχνός τύπος συρραφής είναι ο λεγόμενος exon scrambling, όπου τα εξώνια είναι συναρμολογημένα με μια διαφορετική σειρά από την αναμενόμενη στο mRNA μετάγραφο. Επιπλέον, ένας ακόμα τύπος εναλλακτικής συρραφής είναι η διατήρηση εσωνίων (intron retention), όπου η νουκλεοτιδική αλληλουχία ενός εσωνίου παραμένει στο ώριμο μετάγραφο

mRNA. Τέλος, εναλλακτικά μετάγραφα προκύπτουν και από χρήση εναλλακτικού υποκινητή (alternative promoter) ή εναλλακτικής πολυαδενυλίωσης (alternative poly-A) όπου στη πρώτη περίπτωση η μεταγραφή ξεκινά από διαφορετικά σημεία σχηματίζοντας μετάγραφα με διαφορετικό 5' άκρο, ενώ στη δεύτερη παράγονται μετάγραφα με διαφορετικό 3' άκρο. Οι δύο αυτοί μηχανισμοί συχνά συναντώνται συνδυαστικά με κάποιο από τους πιο πάνω μηχανισμούς εναλλακτικού ματίσματος εννοώντας την ποικιλομορφία στα mRNA που προέρχονται από ένα μόνο γονίδιο.

Αυτές οι αλλαγές μπορεί να επηρεάσουν την σταθερότητα του mRNA μεταγράφου, τον εντοπισμό του ή την μετάφραση [3]. Επιπλέον η εναλλακτική συρραφή μπορεί προκαλέσει αλλαγή στο πλαίσιο ανάγνωσης ή όχι. Τροποποιήσεις στο πλαίσιο ανάγνωσης (πλαισιοτροποποιητικές) μπορεί να οδηγήσουν σε μεγάλες αλλαγές στην πρωτεϊνική αλληλουχία άρα και στην λειτουργία της πρωτεΐνης ή να προκαλέσουν αποικοδόμηση του mRNA με σχηματισμό πρόωρου κωδικονίου λήξης [8-10]. Μη πλαισιοτροποποιητικές αλλαγές με διατήρηση ανοικτού πλαισίου ανάγνωσης οδηγούν συνήθως σε προσθήκες ή αφαιρέσεις λίγων μόνο αμινοξέων του πρωτεϊνικού προϊόντος, χωρίς ιδιαίτερες επιπτώσεις στη λειτουργικότητα της πρωτεΐνης.

### 1.1.3. Η σημασία της εναλλακτικής συρραφής

Η εναλλακτική συρραφή, δηλαδή η διαδικασία όπου εξώνια πρώιμων μεταγράφων (pre-mRNAs) γονιδίων, αποκόπτονται και συνδυάζονται με διαφορετικούς τρόπους σχηματίζοντας δομικά και λειτουργικά διακριτά mRNA μετάγραφα και αντίστοιχα πρωτεϊνικές ισομορφές, είναι ένας από τους μηχανισμούς που χρησιμοποιούνται εκτεταμένα κατά τη γονιδιακή έκφραση, προάγοντας τη μεγαλύτερη μακρομοριακή και κυτταρική πολυπλοκότητα στους ανώτερους ευκαρυωτικούς οργανισμούς [11]. Η διαδικασία αυτή απαιτεί ρύθμιση, η οποία γίνεται από τη συγχρονισμένη δράση παραγόντων που συμμετέχουν στην εναλλακτική συρραφή. Με την έκφραση συγκεκριμένων ρυθμιστικών παραγόντων για κάποιο κυτταρικό τύπο ή αναπτυξιακή περίοδο, εφαρμόζεται αυστηρός έλεγχος και διαφοροποίηση στα γεγονότα του εναλλακτικού ματίσματος. Η ανάπτυξη πολλών ιστών και οργάνων αλλά και η φυσιολογική λειτουργία τους εξαρτώνται κατά ένα μεγάλο βαθμό από τη σωστή ρύθμιση και λειτουργία της εναλλακτικής συρραφής έτσι μέσω δικτύων ματίσματος καθοδηγείται η ανάπτυξη νευρικού, μυϊκού και ανοσοποιητικού συστήματος όπως και η γαμετογένεση.

#### 1.1.4. Εναλλακτική συρραφή και ασθένειες

Έίναι πλέον γνωστό ότι η εναλλακτική συρραφή συνδέεται άμεσα με την ομοιόσταση των ανθρωπίνων κυττάρων και συγκεκριμένα με κυτταρικές διεργασίες όπως η διαφοροποίηση, η ανάπτυξη, η διακυτταρική επικοινωνία, η μετανάστευση, η απόπτωση και ο κυτταρικός θάνατος. Έτσι η αυστηρή ρύθμιση της λειτουργίας της είναι απαραίτητη. Δεν γνωρίζουμε πάντα το πώς και το γιατί, αλλά η εναλλακτική συρραφή μπορεί να παρεκκλίνει από την φυσιολογική πορεία της επηρεάζοντας και τις κυτταρικές διαδικασίες με τις οποίες συσχετίζεται. Πολλά γεγονότα εναλλακτικού ματίσματος ρυθμίζονται από διάφορους ρυθμιστικούς παράγοντες. Τα δίκτυα ματίσματος, αποτελούνται από μετάγραφα που κωδικοποιούν πρωτεΐνες οι οποίες συμμετέχουν σε συγχρονισμένες και συσχετιζόμενες διαδικασίες, έχοντας αντίκτυπο στην ανάπτυξη και τον φαινότυπο των διάφορων κυτταρικών τύπων. Αυτό συνεπάγεται ότι οποιαδήποτε αλλαγή στην συνηθισμένη πορεία ρύθμισης της εναλλακτικής συρραφής, μπορεί να οδηγήσει σε ασθένειες και καρκινικά φαινόμενα [12].

Συγκεκριμένα, πολλές μεταβολικές διαταραχές σχετίζονται άμεσα με λανθασμένο μάτισμα, κυρίως στην περίπτωση της παχυσαρκίας. Επιπλέον, λανθασμένα μοτίβα εναλλακτικής συρραφής φαίνεται να σχετίζονται και με την απόπτωση β κυττάρων του παγκρέατος, την παραγωγή και έκκριση ινσουλίνης ή με υποδοχείς ινσουλίνης, προκαλώντας την πιο κοινή μεταβολική ασθένεια, τον διαβήτη.

Το εναλλακτικό μάτισμα συνδέεται άμεσα και με αρκετές νευρολογικές ασθένειες στον άνθρωπο καθώς τα φαινόμενα εναλλακτικού ματίσματος συμβαίνουν πιο έντονα στον νευρικό ιστό σε σχέση με τους υπόλοιπους. Πιο συγκεκριμένα, το εναλλακτικό μάτισμα, σε νευρολογικές ασθένειες, σχετίζεται κυρίως με μετάγραφα που κωδικοποιούν νευροπεπτίδια, υποδοχείς νευροπεπτιδίων ή ένζυμα που ρυθμίζουν τη νευρική λειτουργία [13]. Αλλαγές στην αναμενόμενη εναλλακτική συρραφή των μεταγράφων αυτών, συνδέονται με νευρολογικές, νευρομυικές και νευροεκφυλιστικές ασθένειες όπως η νόσος του Alzheimer [14-18], η νόσος του Parkinson [19, 20] και η χορεία του Huntington [21-24], αλλά και ψυχικές ασθένειες όπως η σχιζοφρένεια και διαταραχές που εμπίπτουν στο φάσμα του Αυτισμού (Autism Spectrum Disorder, ASD) [25].

Επιπρόσθετα, σε μετάγραφα σαρκομερικών πρωτεϊνών, καναλιών ιόντων, πρωτεϊνών μεταγωγής σήματος ή πρωτεϊνών που σχετίζονται γενικά με την ανάπτυξη καρδιακού ιστού, οποιαδήποτε αλλαγή στη φυσιολογική πορεία εναλλακτικού ματίσματος, μπορεί να οδηγήσει σε καρδιακές παθήσεις. Επιπλέον, αρκετοί παράγοντες ελέγχουν τη σωστή λειτουργία του

εναλλακτικού ματίσματος, άρα οποιαδήποτε τροποποίησή τους, μπορεί να έχει συνέπειες στη σωστή ρύθμιση της γονιδιακής έκφρασης που συνδέεται με την καρδιακή ανάπτυξη κατά την εμβρυική περίοδο και την αύξηση παροχής αίματος στο σώμα κατά την ενήλικη περίοδο, οδηγώντας σε καρδιακή ανεπάρκεια.

#### 1.1.5. Εναλλακτική συρραφή και καρκίνος

Από πρόσφατες μελέτες φαίνεται μια σύνδεση μεταξύ γεγονότων εναλλακτικού ματίσματος και γεγονότων-σημάτων (hallmarks) καρκίνου. Ως hallmark ορίζεται ένα σύνολο βιολογικών ιδιοτήτων που ένα κύτταρο μπορεί να αποκτήσει κατά την ογκογένεση και την ανάπτυξη του καρκίνου, όπως η πολλαπλασιαστική ιδιότητα, η διαφυγή από την απόπτωση και από καταστολείς της ανάπτυξης, η απορρύθμιση του κυτταρικού μεταβολισμού και ιδιότητες που επιδρούν θετικά στην ανάπτυξη του καρκίνου όπως η αγγειογένεση, η διεισδυτικότητα και η μετάσταση. Έτσι έχει γίνει πρόταση ώστε η μη φυσιολογική συρραφή να θεωρηθεί ως ένα νέο hallmark. Γνωρίζουμε πλέον ότι η εναλλακτική συρραφή μπορεί να ρυθμιστεί και από καρκινικά κύτταρα. Από μελέτες έχει ανακαλυφθεί η ύπαρξη ειδικών καρκινικών μοτίβων εναλλακτικής συρραφής. Η δυνατότητα αυτή θα μπορούσε να είναι πολύ ωφέλιμη για τα καρκινικά κύτταρα εάν στα αρχικά αναπτυξιακά στάδια, οι ισομορφές που είναι απαραίτητες για τον πολλαπλασιασμό των κυττάρων, εκφράζονταν και εκτοπικά οδηγώντας σε ανεξέλεκτη αύξηση. Αυτή η αλλαγή στο μοτίβο εναλλακτικής συρραφής μπορεί να είναι κρίσιμη, καθώς πολλά γονίδια διαθέτουν ισομορφές που είναι ανταγωνιστικές μεταξύ τους [7]. Γνωρίζουμε πλέον ότι η μη φυσιολογική λειτουργία του εναλλακτικού ματίσματος δεν σχετίζεται απλά με την καρκινογένεση αλλά επηρεάζει αρνητικά και την εξέλιξη του καρκίνου καθώς και την αποτελεσματικότητα της θεραπείας [26, 27].

Ως συμπέρασμα, καταλήγουμε ότι ορισμένα εναλλακτικά μετάγραφα τα οποία εμφανίζονται σε συγκεκριμένους καρκινικούς ιστούς και σχετίζονται με την καρκινογένεση, μπορούν να χρησιμοποιηθούν ως καρκινικοί βιοδείκτες ή ακόμη και ως θεραπευτικοί στόχοι. Σήμερα, το πρόγραμμα Cancer Genome Atlas (TCGA) προσφέρει το μεγαλύτερο σύνολο εναλλακτικών προϊόντων συρραφής τα οποία έχουν αναγνωρισθεί ότι σχετίζονται με διαφορετικούς καρκινικούς τύπους. Όλη αυτή η πληροφορία μπορεί να χρησιμοποιηθεί για την αναγνώριση υποψήφιων καρκινικών βιοδεικτών και γονιδίων στα οποία τα εναλλακτικά μάτισμα μπορεί να προκαλέσει την εμφάνιση καρκίνου [28].

## 1.2. Ρύθμιση έκφρασης από 3'- αμετάφραστες περιοχές

Στα ευκαρυωτικά κύτταρα, κατα την μεταγραφή, πραγματοποιείται μια σειρά διαδικασιών για τον σχηματισμό του ώριμου mRNA. Αυτό περιλαμβάνει την απομάκρυνση εσωνίων και την συρραφή εξωνίων, την προσθήκη μιας “καλύπτρας” στο 5'- άκρο για την δημιουργία της 5'-UTR και την προσθήκη μιας σειράς νουκλεοτιδίων αδενυλικού οξέος (πολυαδενυλικό οξύ ή poly(A)) στο 3'- άκρο για την δημιουργία της 3'- UTR. Η προσθήκη της poly-A ουράς γίνεται μετά την ενδονουκλεοτιδική πέψη ενός τμήματος του 3'- άκρου του pre-mRNA το οποίο δεν αντιστοιχεί σε κωδική περιοχή. Κατά συνέπεια, μια 3'- αμετάφραστη περιοχή (ή 3'-UTR) θεωρείται το τμήμα μεταξύ του κωδικονίου λήξης και της αρχής της poly-A ουράς. Η δημιουργία 5' και 3'-ακρών στο ώριμο mRNA, έχει ως αποτέλεσμα την προστασία του mRNA από τυχόν αποικοδόμηση.

Σήμερα γνωρίζουμε ότι αυτές οι 3'-UTR περιοχές διαθέτουν πολύ σημαντικές αλληλουχίες, οι οποίες μπορούν να έχουν καθοριστικό ρόλο στην ρύθμιση της έκφρασης των γονιδίων, επηρεάζοντας την σταθερότητα του mRNA, την έξοδο του στο κυτόπλασμα, τον υποκυτταρικό εντοπισμό του και την αποτελεσματικότητα της μετάφρασής του [29]. Μπορούν επιπλέον να λειτουργήσουν ως μεγάλα μη κωδικά ή μικρά RNAs τόσο για ολόκληρες 3'-UTR περιοχές όσο και για θραύσματά τους. Ακόμη, οι 3'-UTR περιοχές μπορούν να καθορίσουν τη πορεία των πρωτεϊνών μέσω ρύθμισης των αλληλεπιδράσεων πρωτίνης-πρωτεΐνης, διευκολύνοντας τον σχηματισμό πρωτεϊνικών συμπλόκων τα οποία περιέχουν ποικιλία υπομονάδων [30].

Μια ακόμη ενδιαφέρουσα παρατήρηση είναι η σχέση μεταξύ μεταγραφής και θέσεων πολυαδενυλίωσης καθώς, γονίδια τα οποία εμφανίζουν υψηλά επίπεδα έκφρασης τείνουν να έχουν μικρότερες 3'-UTR περιοχές από ότι μεγάλες 3'-UTR ισομορφές σε γονίδια τα οποία έχουν χαμηλότερα επίπεδα έκφρασης, καθώς οι μικρότερες 3'-UTR ισομορφές έχουν μεγαλύτερη σταθερότητα [29]. Το γεγονός αυτό δικαιολογεί την εμφάνιση μικρότερων 3'-UTR περιοχών σε ογκογονίδια, καθώς η αυξημένη γονιδιακή έκφραση συνδέεται με τον κυτταρικό πολλαπλασιασμό.

### 1.2.1. Ρύθμιση από AU-πλούσια στοιχεία

Τα στοιχεία, τα οποία είναι πλούσια σε αδενίνη-κυτοσίνη (AU) ή αλλιώς AREs, θεωρούνται τα πιο κοινά ρυθμιστικά στοιχεία στα 3'-UTRs, καθώς επηρεάζουν την σταθερότητα του mRNA, τη διαδικασία εναλλακτικής συρραφής και μετάφρασης. Τα AREs

είναι κυρίως υπεύθυνα για την καταστολή της μετάφρασης και την αποσταθεροποίηση του mRNA. Ένα από τα βασικά μοτίβα είναι το αλληλουχικό τμήμα “AUUUA”. Γενικά, σχετίζονται με την διευκόλυνση της διαδικασίας απο-αδενυλίωσης, επιτυγχάνοντας την δημιουργία μικρότερης poly-A ουράς, η οποία είναι απαραίτητη για την σταθερότητα του mRNA μορίου. Μέχρι σήμερα, γνωρίζουμε περισσότερα από 3,700 ανθρώπινα ARE-γονίδια και περισσότερα από 6,000 ανθρώπινα AREmRNAs [29].

Τα ARE στοιχεία, αναγνωρίζονται από μια ομάδα πρωτεϊνών, η οποία ονομάζεται ARE-binding πρωτεΐνες (ARE-BPs). Η πρωτεΐνη, η οποία θα συνδεθεί με το στοιχείο ARE, καθορίζει εν τέλει την πορεία του μεταγράφου, οδηγώντας είτε σε σταθεροποίηση και ενίσχυση της μεταφραστικής διαδικασίας, είτε σε αποσταθεροποίηση και καταστολή της μετάφρασης. Οι πιο γνωστές ARE-BPs θεωρούνται οι πρωτεΐνες TTR-RBPs, οι οποίες έχουν πολυλειτουργικές δυνατότητες προσδίδοντας τους ρυθμιστική ικανότητα σε μεγάλο εύρος γονιδίων, συμμετέχοντας σε διάφορες κυτταρικές διαδικασίες. Οι πιο καλά μελετημένες είναι, η TTP, ο παράγοντας AUF1, η πρωτεΐνη KSRP, ο παράγοντας BRF1 και η οικογένεια των Hu πρωτεϊνών [29].

#### 1.2.2. Ρύθμιση από GU-πλούσια στοιχεία, CU-πλούσια, στοιχεία ελέγχου διαφοροποίησης και CA-πλούσια στοιχεία

Τα GU-πλούσια στοιχεία, ή αλλιώς GREs, εμφανίζονται σε τουλάχιστον 5% των ανθρωπινών mRNAs. Έχουν ομοιότητα με τα στοιχεία ARE καθώς η βασική πενταμελής αλληλουχία τους, είναι η “GUUUG”. Επιπλέον, εμφανίζουν λειτουργική ομοιότητα καθώς παρατηρούνται σε μικρής διάρκειας ζωής, mRNAs σε T-λεμφοκύτταρα, προκαλώντας την αποσύνθεσή του mRNA. Τοποθετούνται σε μια επικαλυπτόμενη διάταξη από δύο έως πέντε πενταμερών και σε αντίθεση με τα AREs, ο αριθμός των επαναλήψεων δεν καθορίζει το επίπεδο επίδρασής τους. Στα GREs, παρατηρείται πρόσδεση συγκεκριμένων ομάδων πρωτεϊνών, οι οποίες συμβάλλουν στη ρύθμιση μετα-μεταγραφικών γεγονότων όπως είναι η απο-αδενυλίωση, η αποσύνθεση του mRNA και η συρραφή του pre-mRNA [29].

Τα CU-πλούσια στοιχεία, ή αλλιώς CUREs και τα στοιχεία ελέγχου διαφοροποίησης (DICE), αρχικά θεωρήθηκαν ως τα ίδια ρυθμιστικά στοιχεία, ενώ σήμερα τα διακρίνουμε σε δύο ξεχωριστές κατηγορίες στοιχείων. Τα στοιχεία DICE εμφανίζουν συγκεκριμένες αλληλουχίες, στις οποίες γίνεται πρόσδεση από ειδικές αλληλεπιδρώσες πρωτεΐνες, γνωστές ως hnRNP K και hnRNP E1, οι οποίες αναστέλλουν την έναρξη της μετάφρασης μέσω ειδικής αναστολής της συναρμολόγησης του ριβοσώματος 80S. Μια πολύ καλά

χαρακτηρισμένη πρωτεΐνη η οποία συνδέεται στα στοιχεία CURE, είναι η PTB, η οποία έχει τη δυνατότητα να συνδέει μεταξύ τους, μακρινές θέσεις στο RNA μόριο, δημιουργώντας βρόγχους, ενισχύοντας τη διαδικασία αλληλεπίδρασης των παραγόντων που απαιτούνται για τη συναρμολόγηση του ριβοσώματος και την επιτυχία της μετάφρασης.

Τα CA-πλούσια στοιχεία, ή αλλιώς CAREs, είναι οι πιο κοινές δινουκλεοτιδικές επαναλήψεις που εντοπίζονται στο ανθρώπινο γονιδίωμα, τόσο σε κωδικές, όσο και σε μη κωδικές αλληλουχίες και λειτουργούν, προσδίδοντας σταθερότητα στα μόρια mRNA [29].

### 1.2.3. Ρύθμιση από στοιχεία απόκρισης σε σίδηρο και στοιχεία SECIS

Τα στοιχεία τα οποία ανταποκρίνονται στο σίδηρο ή αλλιώς iron responsive elements (IREs), αφορούν εξελικτικά διατηρημένες ρυθμιστικές θέσεις, οι οποίες συμμετέχουν στον έλεγχο του μεταβολισμού του σιδήρου. Συγκεκριμένα, διατηρούν την ομοιότητα του σιδήρου, ρυθμίζοντας την έξοδο, την ενδοκυτταρική μεταφορά και την είσοδο του σιδήρου στο κύτταρο, μέσω επιρροής της σταθερότητας του mRNA και της μετάφρασης συγκεκριμένων γονιδίων. Τα IREs είναι στην ουσία τμήματα αλληλουχίας σε μορφή βρόγχου τα οποία εμφανίζονται σε 5' και 3'-UTRs. Σε περιπτώσεις μη λειτουργίας των ρυθμιστικών μηχανισμών των στοιχείων αυτών, προκαλούνται διάφορες παθολογικές καταστάσεις όπως η νόσος Αλτσχάιμερ [29].

Τα στοιχεία αλληλουχίας εισαγωγής σεληνοκυστεΐνης, ή αλλιώς selenocysteine insertion sequence element (SECIS), είναι ουσιαστικά RNA δομές φουρκέτας, που εντοπίζονται στο 3'-UTR τμήμα των mRNA ευκαρυωτικών πρωτεϊνών, οι οποίες ονομάζονται selenoproteins. Οι πρωτεΐνες αυτές λειτουργούν αναγνωρίζοντας το κωδικόνιο λήξης "UGA" ως σεληνοκυστεΐνη, δηλαδή ως το πρόσφατα αναγνωρισμένο 21<sup>ο</sup> απαραίτητο αμινοξύ. Οι αλληλουχίες αυτές δεν εμφανίζουν συντηρημένη μορφή στα ευκαρυωτικά κύτταρα, αλλά έχουν δύο μορφές δευτεροταγούς δομής, με διατηρημένα μοτίβα, υπεύθυνα για την αναγνώριση του "UGA" ως σεληνοκυστεΐνη. Σε αυτά γίνεται η πρόσδεση πρωτεϊνών, καθοριστικών για την μετάφραση των selenoproteins [29].

### 1.2.4. Ρύθμιση από RNA-binding πρωτεΐνες

Στις 3'UTR περιοχές, εμφανίζεται πρόσδεση ειδικών πρωτεϊνών οι οποίες προσδένονται σε μονόκλωνο RNA ( RNA-binding proteins, RBPs). Οι πρωτεΐνες αυτές αλληλεπιδρούν με ομάδες πρωτεϊνών, οι οποίες επιδρούν με αποτέλεσμα τον εντοπισμό του mRNA σε διαφορετικά υποκυτταρικά διαμερίσματα. Επιπλέον, οι 3'-UTRs καθορίζουν τη

τύχη των νεοσυντιθέμενων πρωτεϊνών καθώς κατά τη διάρκεια σύνθεσης της πεπτιδικής αλυσίδας, οι RBPs στρατολογούν πρωτεΐνες οι οποίες επηρεάζουν τον εντοπισμό και τη λειτουργία της συντιθέμενης πρωτεΐνης [30]. Οι RBPs είναι πρακτικά οι πρωτεΐνες οι οποίες αναγνωρίζουν και προσδένονται στα ρυθμιστικά στοιχεία τα οποία περιλαμβάνονται στις 5' και 3'-UTR περιοχές των mRNAs και καθορίζουν τον μηχανισμό ρύθμισης τους.

#### 1.2.5. Ρύθμιση από microRNA

Τα microRNAs (miRNAs) είναι μόρια τα οποία ρυθμίζουν την έκφραση στο περίπου 60% των ανθρώπινων γονιδίων που κωδικοποιούν πρωτεΐνες, σε μετα-μεταγραφικό επίπεδο. Η ρύθμιση αυτή γίνεται με αποικοδόμηση του mRNA ή/και αναστολή μετάφρασης. Βάσει πρόσφατων μελετών, φαίνεται πως ο κύριος τους ρόλος είναι ο περιορισμός της πρωτεϊνοσύνθεσης, μέσω αποσταθεροποίησης των mRNA στόχων. Σε κάποιες περιπτώσεις βέβαια, τα μόρια miRNA εμφανίζονται ως ενεργοποιητές της μετάφρασης υπο συγκεκριμένες συνθήκες.

Όσον αφορά το ανθρώπινο γονιδίωμα, έχουν ταυτοποιηθεί miRNA γονίδια τα οποία αντιπροσωπεύουν το 5% περίπου του γονιδιώματος. Έμφανίζονται ως μονοκλωνικά μόρια RNA, μήκους 20-25 βάσεων. Υπάρχουν ειδικές θέσεις δέσμευσής τους, οι λεγόμενες “seed regions”, οι οποίες εντοπίζονται ως επι το πλήστων εντός 3'-UTR περιοχών και περιέχουν 6 διαδοχικά νουκλεοτίδια. Τα κατάλληλα μόρια miRNA προσδένονται σε αυτές μέσω συμπληρωματικότητας των βάσεων, στις θέσεις 2-7 από το 5' άκρο τους. Σπάνια μπορεί να ανιχνευθούν τέτοιες θέσεις δέσμευσης, σε 5'-UTR περιοχή. Επιπλέον, η ισχύς της ρύθμισης φαίνεται να σχετίζεται με τον αριθμό των θέσεων δέσμευσης, εντός της 3'-UTR.

Αρκετές μελέτες έχουν αποκαλύψει ισχυρά εξειδικευμένο προφίλ miRNA, κατά τη διάρκεια συγκεκριμένων σταδίων ανάπτυξης ή κατά την ογκογένεση. Σύμφωνα με αυτό, ένας αριθμός miRNAs, εμπλέκεται στη ρύθμιση αρκετών οδών μεταγωγής σήματος, θεμελιωδών κυτταρικών διεργασιών όπως η διαφοροποίηση, ο πολλαπλασιασμός, η απόπτωση και ο μεταβολισμός.

Τα τελευταία χρόνια έχει ανακαλυφθεί η εκτενής ρύθμιση από συγκεκριμένα miRNAs και των ανθρώπινων γονιδίων των *KLKs*. Η συμμετοχή των μορίων αυτών στη ρύθμιση των γονιδίων των *KLKs*, μπορεί να δώσει μια εξήγηση για αρκετά από τα φαινόμενα τα οποία χαρακτηρίζουν την έκφραση των γονιδίων αυτών, όπως οι ασυμφωνίες μεταξύ των επιπέδων mRNA και πρωτεΐνης, η κοινή απορρύθμιση στα επίπεδα έκφρασής τους όταν



βρίσκονται σε παθολογικές καταστάσεις καθώς και η παράλληλη έκφραση αρκετών από τα μέλη των *KLKs*, σε ίδιο τύπο ιστού.[31-33].

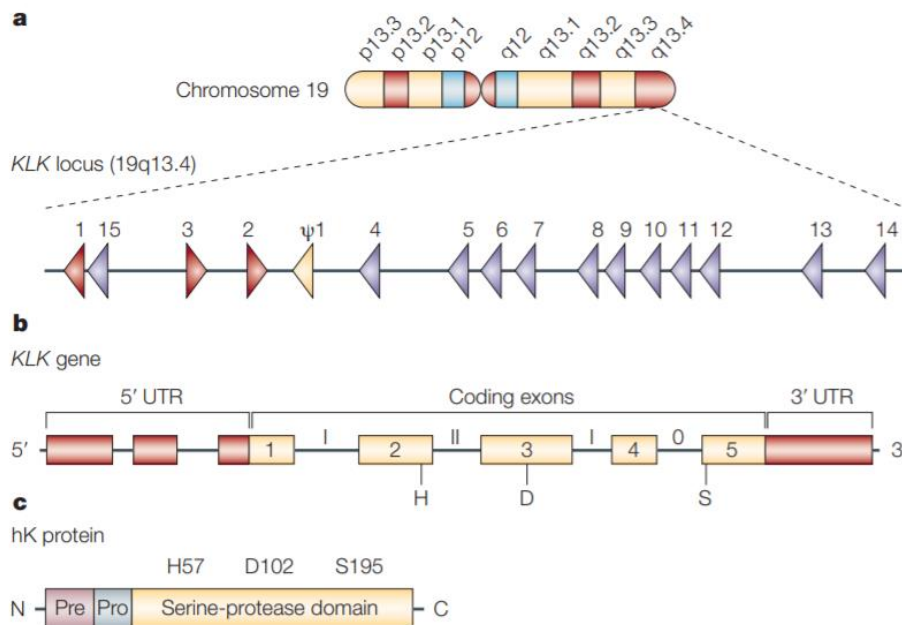
### 1.3. Καλλικρεΐνες

Οι καλλικρεΐνες (*KLKs*) ανθρώπινων ιστών αποτελούν μια υποομάδα πρωτεασών σερίνης, οι οποίες κωδικοποιούνται από δεκαπέντε δομικά όμοια, ορμονικά ρυθμιζόμενα γονίδια. Οι πρωτεάσες αυτές έχουν εντοπιστεί σε μεγάλο αριθμό ιστών και βιολογικών υγρών στον ανθρώπινο οργανισμό. Η ονομασία τους προέρχεται από την λέξη καλλίκρεας ή αλλιώς πάγκρεας, η οποία δόθηκε το 1930, μετά την ανακάλυψη μιας ουσίας που μπορούσε να ρυθμίζει την πίεση του αίματος και είχε υψηλή συγκέντρωση στο πάγκρεας, η οποία μετέπειτα ονομάστηκε καλλικρεΐνη 1 ή ιστική καλλικρεΐνη [34]. Χρόνια αργότερα έγινε και η ταυτοποίηση του γονιδίου που κωδικοποιεί το γονίδιο της καλλικρεΐνης 1, το *KLK1*, μαζί με ακόμη δύο γονίδια, τα *KLK2* και *KLK3*. Το *KLK3* ή αλλιώς *PSA*, θεωρείται το πιο καλά χαρακτηρισμένο γονίδιο καλλικρεΐνης μέχρι σήμερα, λόγω της κλινικής αξίας του ως βιοδείκτης για την διάγνωση του καρκίνου του προστάτη. Η ανακάλυψη νέων γονιδίων που φέρουν κοινά δομικά στοιχεία με τα τρία *KLKs*, συνέχισε μέχρι τα τέλη της δεκαετίας του '90 με συνεισφορά πολλών επιστημόνων. Η κλωνοποίηση, χαρτογράφηση και τελική ταυτοποίηση όλων των μελών της οικογένειας γονιδίων των καλλικρεΐνών ολοκληρώθηκε από την ομάδα των Yousef, Scorilas και Diamandis, παρουσιάζοντας την γνωστή πλέον δεκαπενταμελή γονιδιακή οικογένεια ανθρώπινων ιστικών καλλικρεΐνών (*KLK1-KLK15*) [35-39]. Μια επιπλέον ανακάλυψη, αφορά την καλλικρεΐνη του πλάσματος, η οποία κωδικοποιείται από το γονίδιο *KLKB1* αλλά δεν συμπεριλαμβάνεται στην ομάδα ιστικών καλλικρεΐνών καθώς εντοπίζεται σε διαφορετική χρωμοσωμική περιοχή και φέρει διαφορετικά δομικά και λειτουργικά στοιχεία [40]. Οι ιστικές καλλικρεΐνες ή *KLKs*, εκφράζονται σε πληθώρα ιστών στον άνθρωπο και συμμετέχουν σε διάφορες φυσιολογικές λειτουργίες όπως η αναδόμηση του δέρματος και η ρύθμιση της πίεσης του αίματος, αλλά και η επεξεργασία ορμονών και η ρευστοποίηση του σπέρματος [41, 42].

#### 1.3.1. Δομή γονιδίων και πρωτεϊνών των καλλικρεΐνών

Όπως αναφέρεται και πιο πάνω οι *KLKs* αποτελούν μια δεκαπενταμελή ομάδα πρωτεασών σερίνης και τα γονίδια που τις κωδικοποιούν, εδράζονται στη χρωμοσωμική περιοχή 19q13.3-13.4, σχηματίζοντας το μεγαλύτερο συνεχές σύμπλεγμα πρωτεασών

σερίνης στο ανθρώπινο γονιδίωμα [43]. Ο γενετικός τόπος των *KLKs* αφορά μια περιοχή περίπου 300 kb, σε ένα τμήμα του μεγάλου βραχίονα του χρωμοσώματος 19 και περιλαμβάνει 15 λειτουργικά γονίδια και ένα ψευδογονίδιο όπως παρουσιάζεται στην **Εικόνα 1.3**. Το μέγεθος των γονιδίων των *KLKs* κυμαίνεται μεταξύ 4 και 10 kb, λόγω των διαφορετικού μεγέθους εσωνίων που εμπεριέχονται σε αυτά. Τα γονίδια *KLK1*, *KLK2* και *KLK3* εντοπίζονται μαζί με το γονίδιο *KLK15* σε μια περιοχή μεγέθους 60 kb, ενώ όλα τα υπόλοιπα γονίδια (*KLK4-KLK14*) μαζί με τον ψευδογονίδιο *ψKLK1* εδράζονται τελομερικά. Το ψευδογονίδιο βρίσκεται στη περιοχή ανάμεσα στα *KLK2* και *KLK4*. Επιπλέον, η κατεύθυνση μεταγραφής όλων των γονιδίων πέραν των *KLK2* και *KLK3* είναι με φορά από το τελομερές προς το κεντρομερές.



**Εικόνα 1.3** Ο γονιδιακός τόπος της οικογένειας των ανθρώπινων γονιδίων *KLKs* στη χρωμοσωμική περιοχή 19q13.4 [34]

Οι *KLKs* φέρουν αρκετά κοινά δομικά χαρακτηριστικά, όπως στην οργάνωση εξωνίων/εσωνίων, στα διαστήματα μεταξύ ιντρονίων και στα μήκη εξωνίων. Επιπλέον, και τα 15 γονίδια, κωδικοποιούν προένζυμα μίας μόνο πολυπεπτιδικής αλυσίδας και μήκους μεταξύ 244 και 296 αμινοξικών καταλοίπων. Οι αμινοξικές τους αλληλουχίες, εμφανίζουν ομοιότητα σε ποσοστά 40-80%. Επιπρόσθετα, οι *KLKs* συχνά συνεκφράζονται στους φυσιολογικούς ιστούς και συνήθως απορυθμίζονται συντονισμένα σε παθολογικές

καταστάσεις, κάτι που υποδεικνύει και την ύπαρξη κοινών μηχανισμών ρύθμιση της έκφρασης τους [41].

### 1.3.2. Ρύθμιση έκφρασης των καλλικρεϊνών

Όπως σε κάθε γονίδιο, έτσι και στα *KLKs*, η ρύθμιση της έκφρασής τους πραγματοποιείται σε τρία επίπεδα, το μεταγραφικό, το μετα-μεταγραφικό και το μεταφραστικό επίπεδο.

Στο μεταγραφικό επίπεδο, τα *KLKs* ρυθμίζονται από ορμονικούς και επιγενετικούς παράγοντες (μεθυλιώσεις, τροποποίηση ιστονών). Η πλειονότητα των γονιδίων αυτών, ρυθμίζεται από στεροειδείς ορμόνες, οι οποίες μέσω υποδοχέων επιτυγχάνουν την πρόσδεσή τους σε *cis*-ρυθμιστικές αλληλουχίες του υποκινητή του γονιδίου ή του ενισχυτή του γονιδίου, γνωστές και ως “Hormonal response elements (HREs)”. Ως εκ τούτου, η απορύθμιση στο επίπεδο της μεταγραφής, συνδέεται άμεσα με αρκετές παθολογικές καταστάσεις όπως διάφορα καρκινώματα. Αρκετές επιγενετικές ανωμαλίες όπως υπερμεθυλιώσεις του DNA και ανωμαλίες στη χρωμοσωμική περιοχή των γονιδίων *KLKs* έχουν αναφερθεί σε διάφορους καρκίνους του μαστού και της λεμφοβλαστικής λευχαιμίας [41].

Στο μετα-μεταγραφικό και μεταφραστικό επίπεδο, η έκφραση των *KLKs* ρυθμίζεται από καθορισμένα μόρια miRNAs. Συγκεκριμένα, ένα γονίδιο μπορεί να ρυθμίζεται από περισσότερα του ενός miRNAs, ενώ παράλληλα ένα συγκεκριμένο miRNA μπορεί να ασκεί ρύθμιση σε περισσότερα του ενός *KLKs*. Πέραν της ρύθμισης που ασκείται από τα miRNAs, νέες μελέτες αποδεικνύουν και τον έλεγχο της έκφρασης των επιπέδων των miRNAs από τα *KLKs*, μέσω αλληλεπίδρασής τους με συγκεκριμένους υποκινητές. Επιπλέον, η διαθεσιμότητα των miRNAs ελέγχεται έμμεσα από τα *KLKs*, καθώς αυτά καταναλώνονται κατά την πρόσδεσή τους σε mRNA μόρια-στόχους. Απορρύθμιση της έκφρασης των *KLKs* σε επίπεδο mRNA επηρεάζει και την διαθεσιμότητα των miRNAs που στοχεύουν τα μέλη της οικογένειας των γονιδίων αυτών, άρα μπορεί να επηρεάσει και τα επίπεδα έκφρασής τους [31]. Η πολύπλοκη αυτή σχέση μεταξύ των *KLKs* και των miRNA μορίων, μπορεί να δώσει εξήγηση σε αρκετά από τα χαρακτηριστικά των *KLKs*, όπως είναι η συνέκφρασή τους σε αρκετούς ιστούς, οι ταυτόσημες διαφορές στα επίπεδα mRNA και πρωτεΐνης ανα ιστό και οι ανωμαλίες στην έκφρασή τους σε πολλές παθολογικές καταστάσεις.

Γνωρίζουμε ότι όλες οι KLKs συντίθενται ως προένζυμα, τα οποία εμφανίζουν ένα προ-πεπτιδίο 4-9 αμινοξικών καταλοίπων στο αμινοτελικό τους άκρο. Το προ-πεπτιδίο αυτό έχει την ιδιότητα να παρεμποδίζει την πρόσδεση υποστρώματος στο ενεργό κέντρο του ενζύμου, έτσι απαιτείται απομάκρυνσή του για την ενεργοποίηση του κάθε ανενεργού προενζύμου-ζυμογόνου (pro-KLKs). Η απομάκρυνση αυτή, γίνεται είτε από πρωτεάσες σερίνης είτε από άλλες KLKs, μέσω αυτοενεργοποίησης. Ωστόσο, για την απενεργοποίηση των KLKs απαιτείται τεμαχισμός, μέσω αυτόλυση ή μέσω άλλων πρωτεασών, ή ενεργότητα ενδογενών αναστολέων [42, 44].

### 1.3.3. Εναλλακτικά μετάγραφα των καλλικρεϊνών

Όπως έχουμε ήδη αναφέρει, η εναλλακτική συρραφή είναι ένας από τους σημαντικότερους μηχανισμούς στον ανθρώπινο οργανισμό καθώς επιτυγχάνει την αύξηση της ποικιλομορφίας σε πρωτεϊνικό επίπεδο, έχοντας έναν συγκριτικά μικρό αριθμό γονιδίων. Το πλεονέκτημα αυτό αξιοποιείται από τα περισσότερα ανθρώπινα γονίδια, δημιουργώντας πολλά εναλλακτικά μετάγραφα. Ο μηχανισμός εναλλακτικής συρραφής υπόκειται σε ρύθμιση, μέσω εναλλακτικής χρήσης υποκινητών, θέσεων ματίσματος και έναρξης/λήξης της μετάφρασης αλλά και εναλλακτικών θέσεων έναρξης της μεταγραφής και θέσεων πολυαδενυλίωσης [45].

Η οικογένεια των ανθρώπινων *KLKs*, εμφανίζει κατα μεγάλο βαθμό φαινόμενα εναλλακτικής συρραφής, δημιουργώντας πληθώρα mRNA μεταγράφων. Η δημιουργία τους προκύπτει ως επί το πλείστον από παράλειψη εξωνίου (κατα 35%), επέκταση εξωνίου (κατα 24%) ή αποκοπή μέρους του (κατα 18%), καθώς και από διατήρηση εσωνίου (κατα 16%). Επιπλέον έχουν ταυτοποιηθεί μετάγραφα που προέκυψαν από εναλλακτικές θέσεις έναρξης της μεταγραφής καθώς και από εναλλακτικές θέσεις πολυαδενυλίωσης. Μέχρι σήμερα, υπάρχουν 40 χαρακτηρισμένα εναλλακτικά μετάγραφα, τα οποία κωδικοποιούν μοναδικές πρωτεϊνικές ισομορφές των *KLKs*, οι οποίες κατα μεγάλο ποσοστό αναμένεται να παρουσιάζουν διαφορετική λειτουργία σε σχέση με τις κλασικές ισομορφές ή να είναι δυσλειτουργικές.

### 1.3.4. Καλλικρεΐνες και ανθρώπινη παθογένεια

Οι ανθρώπινες *KLK* πρωτεάσες είναι γνωστό πλέον ότι συμμετέχουν σε διαφορετικά μονοπάτια, με σκοπό τη ρύθμιση της πίεσης του αίματος και της λειτουργίας των νεφρών, την απολέπιση δέρματος, τη ρευστοποίηση του σπέρματος, τη συναπτική νευρική

πλαστικότητα και την ωρίμανση του σμάλτου των δοντιών [46]. Το γεγονός αυτό δικαιολογεί την έκφρασή τους στη πλειονότητα των ανθρώπινων ιστών.

Αρκετά μέλη της οικογένειας ανθρώπινων καλλικρεϊνών, σχετίζονται με μη νεοπλασματικές νόσους. Ένα γνωστό παράδειγμα αποτελεί η KLK6, η οποία συνδέεται με φλεγμονώδεις αντιδράσεις στο κεντρικό νευρικό σύστημα, οι οποίες προκαλούν απομυελίνωση, όπως για παράδειγμα οι πολλαπλή σκλήρυνση. Συνδέεται ακόμη, με νευροεκφυλιστικές νόσους, όπως η νόσος του Alzheimer και η νόσος του Parkinson. Επιπλέον, κύριο αίτιο σε διάφορες ασθένειες του δέρματος, όπως η χρόνια δερματίτιδα, η ψωρίαση, η ατοπική δερματίτιδα και το σύνδρομο Netherton, θεωρείται η μη φυσιολογική έκφραση των μελών KLKs, προκαλώντας μεταβολές στην πρωτεολυτική τους ικανότητα.

Πέραν των μη νεοπλασματικών ασθενειών που παρατηρείται να σχετίζονται με τα μέλη της οικογένειας των καλλικρεϊνών, έχει πλέον αποδειχθεί και η απορρύθμιση την οποία υφίστανται, σε περιπτώσεις ανθρώπινων νεοπλασιών, σε μοριακό και πρωτεϊνικό επίπεδο. Ακόμη, η ρυθμιστική δράση των KLKs ως προς την αύξηση των καρκινικών κυττάρων, την αγγειογένεση, τη διήθηση και την μετάσταση, καθιστά ξεκάθαρη την ενεργό δράση των πρωτεασών αυτών ως προς την καρκινογένεση [47]. Μια λειτουργία ορισμένων μελών των KLKs, είναι η ρύθμιση της βιοδιαθεσιμότητας των ινσουλινομορφων αυξητικών παραγόντων (IGF1 και 2) , προκαλώντας την πρωτεόλυση των δεσμευτικών πρωτεϊνών των IGFs (IGFBPs). Κατ' επέκταση, η απορρύθμιση της πρωτεόλυσης των IGFBPs από τις KLKs, μπορεί να επάγει τον κυτταρικό πολλαπλασιασμό και να αναστείλει την απόπτωση των καρκινικών κυττάρων, συμβάλλοντας ενεργά στην ανάπτυξη καρκινικών όγκων [48].

#### 1.3.5. Καλλικρεΐνες και καρκινικοί δείκτες

Όπως έχει ήδη αναφερθεί, οι KLKs συχετίζονται με αρκετούς μηχανισμούς καρκινογένεσης, γεγονός που τις καθιστά αρκετά υποσχόμενες ως υποψήφιους καρκινικούς βιοδείκτες. Αναμφισβήτητη απόδειξη θεωρείται ο πιο γνωστός διαγνωστικός βιοδείκτης της οικογένειας των KLKs, που αφορά τον καρκίνο του προστάτη, το PSA. Ακολούθως, υπάρχει μια σειρά από μέλη της ομάδας των KLKs, στα οποία παρουσιάζονται αλλαγές σε γονιδιακό ή/και πρωτεϊνικό επίπεδο, όταν συνδυάζονται με περιπτώσεις νεοπλασιών, εδραιώνοντας την προγνωστική ή/και διαγνωστική τους αξία. Ως καρκινικοί βιοδείκτες, οι KLKs μπορούν να θεωρηθούν ιδανική επιλογή και αυτό οφείλεται αποκλειστικά στη φύση τους. Η ανίχνευσή

τους, ως εκκρινόμενες πρωτεάσες, στον ορό του αίματος ή σε άλλα βιολογικά υγρά, είναι μια εύκολη και μη επεμβατική διαδικασία. Είναι σημαντικό να αναφερθεί ότι μέχρι σήμερα, όλα τα μέλη της οικογένειας των KLKs έχουν εισηγηθεί ως πιθανοί μοριακοί καρκινικοί δείκτες για μια τουλάχιστον μορφή καρκίνου [49].

#### 1.4. Αλληλούχηση DNA

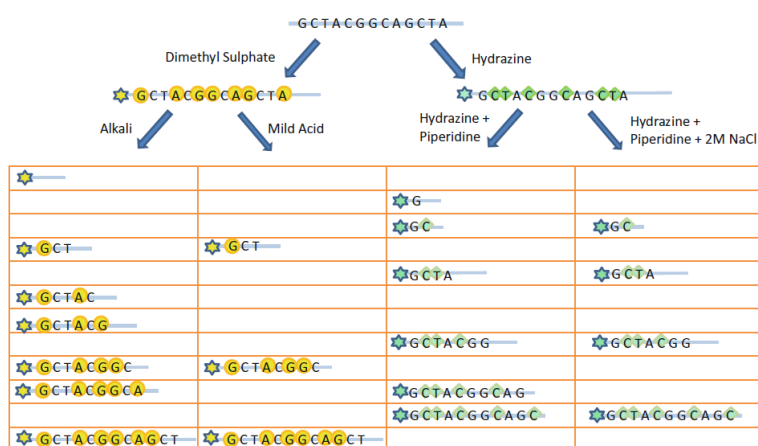
Αναφερόμενοι στην αλληλούχηση DNA περιγράφουμε την ανάλυση ενός τμήματος του DNA για τον προσδιορισμό της ακριβούς νουκλεοτιδικής του αλληλουχίας, δηλαδή τον καθορισμό της σειράς των τεσσάρων αζωτούχων βάσεων της αλυσίδας του DNA, αδενίνη (A), θυμίνη (T), γουανίνη (G), κυτοσίνη (C). Η ανάπτυξη τέτοιου είδους τεχνολογιών, έχει μια πλούσια ιστορία που περιλαμβάνει αρκετές επιστημονικές μελέτες και ανακαλύψεις που όχι μόνο επέτρεψαν την ταυτοποίηση της αλληλουχίας DNA, αλλά πλέον θεωρούνται σημεία αναφοράς για τον τομέα της Βιολογίας.

Η ανάπτυξη τεχνικών αλληλούχησης ήταν το στήριγμα για την ανάπτυξη και της “γονιδιωματικής” (genomics), δίνοντας απαντήσεις σε πολλά ερωτήματα σχετικά με την οργάνωση, τη δομή, τη λειτουργία και την εξέλιξη των γονιδιωμάτων. Το μεγαλύτερο επίτευγμα είναι ο προσδιορισμός του ανθρώπινου γονιδιώματος, ο οποίος ολοκληρώθηκε από το πρόγραμμα ανάλυσης του ανθρώπινου γονιδιώματος (Human Genome Project, HGP) το 2001 [50]. Σήμερα, η αλληλούχηση είναι το βασικότερο εργαλείο στη μοριακή βιολογία και απαιτείται για την μελέτη μεραγραφωμάτων και γονιδιωμάτων. Η γνώση που παρέχεται μπορεί να χρησιμοποιηθεί για τον εντοπισμό μεταλλαγών ή αλλαγών μήκους γονιδίων και μεταγράφων και τον συσχετισμό τους με ασθένειες και φαινότυπους. Επιπλέον, η μοριακή διαγνωστική στοχεύει στην ανίχνευση προδιάθεσης για γενετικές ασθένειες μέσω μελέτης συγκεκριμένων γενετικών περιοχών, χρωμοσωμάτων ή και ολόκληρων γονιδιωμάτων. Πέραν αυτού, η εξελικτική βιολογία έχει πλέον υιοθετήσει τη χρήση αλληλούχησης, για τη σύγκριση του DNA μεταξύ διαφορετικών οργανισμών και τη μελέτη του βαθμού συγγενείας τους, καθώς και τον τρόπο που έχουν εξελιχθεί γονιδιακά.

### 1.4.1 Αλληλούχηση πρώτης γενιάς

Η πρώτη επιτυχής αλληλούχηση DNA καταγράφεται το 1968, αν και οι προσπάθειες είχαν ξεκινήσει από τις αρχές του 1950 με την αλληλούχηση των πρώτων πρωτεΐνων και το 1960 με αλληλούχηση τμημάτων RNA. Το μεγαλύτερο βήμα όμως για την πορεία της αλληλούχησης, έγινε την δεκαετία του 1970 όπου δύο διαφορετικές μέθοδοι ήρθαν στο προσκήνιο, δίνοντας τη δυνατότητα αλληλούχησης εκατοντάδων βάσεων: μέθοδος των Sanger και Coulson και χημική αλληλούχηση των Maxam και Gilbert [50].

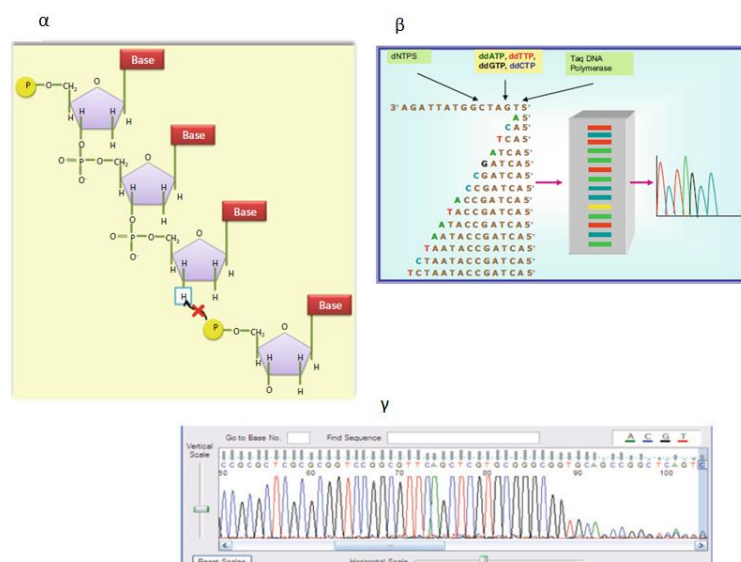
Η χημική αλληλούχηση, όπως ονομάστηκε από τους Allan-Maxam και Walter-Gilbert, αφορούσε την χημική τροποποίηση του DNA και διάσπασή του σε συγκεκριμένες βάσεις. Συγκεκριμένα, το δίκλωνο μόριο DNA σημαίνεται με ραδιοσημασμένο φώσφορο στο 5' ή το 3' άκρο. Έπειτα γίνεται χρήση χημικών ουσιών που διασπών το DNA σε συγκεκριμένες βάσεις (επιλεκτικά σε πουρίνες ή πυριμιδίνες), δημιουργώντας κολλώδη θραύσματα από μονόκλωνο DNA τα οποία είναι ραδιοσημασμένα στο ένα άκρο τους. Τα θραύσματα από τις τέσσερις αντιδράσεις, ηλεκτροφορούνται σε ειδικό πήκτωμα πολυακρυλαμίδης και διαχωρίζονται με βάση το μέγεθός τους και ακολοθεί αυτοραδιογραφία για την ανίχνευσή τους και τον προσδιορισμό της ακολουθίας, όπως παρουσιάζεται στην **Εικόνα 1.4**. Η τεχνική αυτή αν και επιτυχής, απαιτούσε την χρήση ραδιενέργειας για την σήμανση των μορίων DNA, έτσι αντικαταστάθηκε από μια λιγότερο πολύπλοκη τεχνική, αυτή της αλληλούχησης κατα Sanger [51].



**Εικόνα 1.4** Αλληλούχηση ενός ολιγονουκλεοτιδίου με την μέθοδο των Maxam-Gilbert [51].

Η αλληλούχηση κατα Sanger, η οποία και επικράτησε μέσα στα επόμενα χρόνια, λόγω της μειωμένης χρήσης χημικών και ραδιενέργειας, αναπτύχθηκε από τον Frederick Sanger το 1975 και είχε καθοριστικό ρόλο για την κατανόηση του ανθρώπινου γονιδιώματος.

Η τεχνική αυτή βασίζεται στην κύρια χρήση των τριφωσφορικών διδεοξυριβονουκλεοσιδίων, τα οποία στερούνται 3' υδροξυλομάδας (3'-OH) και έχουν την ιδιότητα να σταματούν την επιμήκυνση ενός κλώνου DNA. Συγκεκριμένα, παρασκευάζεται ένα μείγμα από μονόκλιωνα μόρια DNA, μια DNA πολυμεράση, ένας μικρός εκκινητής που βοηθάει την διαδικασία της αντιγραφής και τα τέσσερα τριφωσφορικά δεοξυριβονουκλεοσίδια (dATP, dCTP, dGTP και dTTP). Επιπλέον προστίθεται μικρή ποσότητα μείγματος από δεοξυριβονουκλεοσίδια, τα οποία με την ενσωμάτωσή τους στην αυξανόμενη αλυσίδα, παρεμποδίζουν την προσθήκη άλλου νουκλεοτιδίου, καθώς η αλυσίδα δεν διαθέτει πλέον ελεύθερη 3'-OH, με αποτέλεσμα την παύση της επιμήκυνσης στο σημείο αυτό. Το δίκλωνο DNA αρχικά διαχωρίζεται σε επιμέρους κλώνους, εκ των οποίων ο ένας χρησιμοποιείται ως εκμαγείο για τις περισσότερες αντιδράσεις. Τα τέσσερα τριφωσφορικά διδεοξυριβονουκλεοσίδια (ddATP, ddCTP, ddGTP και ddTTP) χρησιμοποιούνται για την σύνθεση αντιγράφων του ίδιου μονόκλωνου DNA εκμαγείου, έτσι η κάθε αντίδραση σταματάει σε διαφορετικά σημεία της αλληλουχίας παράγοντας ομάδα μορίων DNA διαφορετικού μήκους. Τα προϊόντα διαχωρίζονται με ηλεκτροφόρηση σε τέσσερις παράλληλες στήλες σε πήκτωμα πολυακρυλαμίδης. Ακολουθεί ανίχνευση των κλασμάτων DNA με χρήση ενός δείκτη (ραδιενεργού ή φθορίζοντα) και καθορισμός της αλληλουχίας του νεοσυντιθέμενου κλώνου με ανάγνωση των ζωνών στο πήκτωμα, όπως παρουσιάζεται στην **Εικόνα 1.5**.



**Εικόνα 1.5** (α) Σύνθεση νουκλεοτιδικής αλυσίδας με προσθήκη dNTPs. Με προσθήκη ενός ddNTP, η σύνθεση σταματά (β) Σχηματισμός μιας σειράς αλυσίδων με ενσωμάτωση ddNTPs



που μπορούν να διαχωριστούν με ηλεκτροφόρηση (γ) Ηλεκτροχρωματογράφημα από αλληλούχηση DNA [51]

Παρόλο που η μέθοδος κατα Sanger καθιερώθηκε ως η κύρια μέθοδος αλληλούχησης μέσα στα επόμενα χρόνια, αποκαλύπτοντας το γονιδίωμα πολλών οργανισμών, υπήρχαν μειονεκτήματα στη μεθοδολογία, τα οποία μέχρι σήμερα κατάφεραν να περιοριστούν αλλά όχι να ξεπεραστούν, όπως η μειωμένη ακρίβεια, όταν η αλυσίδα του DNA διαθέτει υψηλό ποσοστό γουανίνης – κυτοσίνης ή επαναλαμβανόμενες περιοχές. Αυτό οδήγησε στην ανάγκη ανάπτυξης νέων μεθόδων, για τη βελτίωση της ταχύτητας και ακρίβειας της αλληλούχησης.

#### 1.4.2 Αλληλούχηση νέας γενιάς

Το 2005 άρχισε επίσημα η χρήση μιας νέας μεθόδου, οι τεχνολογίες αλληλούχησης νέας γενιάς (Next-Generation Sequencing, NGS) οι οποίες κατάφεραν να διευρύνουν τους ορίζοντες για το μέλλον της αλληλούχησης, έχοντας καλύτερα αποτελέσματα σε ταχύτητα και ακρίβεια. Η λογική σε αυτή μεθοδολογία βασίζεται στο σπάσιμο του DNA σε τυχαία σημεία, μέσω ειδικών ενζύμων, δημιουργώντας εκατομμύρια θραύσματα με τυχαίο μέγεθος. Ακολούθως, γίνεται πρόσδεση μικρών ολιγονουκλεοτιδίων στα θραύσματα με καθορισμένη αλληλουχία, οι λεγόμενοι adapters και ακολουθεί η επιλογή και απομόνωση των θραυσμάτων με το κατάλληλο για την μέθοδο, μέγεθος (size-selection). Στο επόμενο στάδιο, τα επιλεγμένα αυτά θραύσματα δεσμεύονται το καθένα ξεχωριστά σε ειδικά σφαιρίδια, στα οποία θα ακολουθήσει μια σειρά αντιδράσεων που ενισχύουν το κάθε θραύσμα. Έπειτα ολοκληρώνεται η αλληλούχηση του καθενός, η οποία γίνεται ταυτόχρονα σε όλα τα προσδεδμεμένα θραύσματα, γνωστή και ως “massively parallel sequencing” [52]. Από την αλληλούχηση προκύπτει ένας τεράστιος αριθμός αναγνώσεων αλληλουχιών ή αλλιώς reads.

Η εξαγωγή τόσο μεγάλου όγκου πληροφορίας, απαιτεί επεξεργασία για την στοίχιση κάθε read, στο γονιδίωμα αναφοράς. Καθώς δεν είναι ίσος ο αριθμός των reads που στοιχίζονται για κάθε τμήμα του γονιδιώματος, καθορίζεται ένας βαθμός κάλυψης ή αλλιώς “coverage” της αλληλούχησης, που διαφέρει ανάλογα με τον σκοπό της αλληλούχησης. Όσο μεγαλύτερος ο βαθμός κάλυψης τόσο μεγαλύτερος και ο βαθμός εμπιστοσύνης για κάθε αλληλουχημένη βάση [53]. Ένα επιπλέον επίτευγμα της τεχνολογίας NGS είναι η απόδοση της αλληλούχησης σε λιγότερο χρόνο και η ακρίβεια, λόγω των επαναλήψεων ανάγνωσης της κάθε βάσης ξεχωριστά. Επιπρόσθετα, δίνεται η δυνατότητα μελέτης κάθε θέσης σε ένα γονιδίωμα, από μόνο μια αντίδραση, το οποίο είναι ανέφικτο με την μέθοδο κατα Sanger [52]. Ως επακόλουθο, η δυνατότητα ταυτόχρονης αλληλούχησης εκατομμυρίων τμημάτων

DNA, η οποία δίνεται μόνο με την αλληλούχηση νέας γενιάς, έκανε την τεχνική αυτή να ξεχωρίσει ανάμεσα στις τεχνικές αλληλούχησης.

Όπως κάθε μεθοδολογία, έτσι και η αλληλούχηση νέας γενιάς εμφανίζει μειονεκτήματα κατά την εφαρμογή της. Όπως έχουμε αναφέρει, η τεχνική βασίζεται στο σπάσιμο των μορίων DNA σε τυχαία σημεία, με επακόλουθο να έχουμε μεγάλο αριθμό μικρών αλληλουχιών προς αλληλούχηση. Επιπλέον, το ποσοστό λαθών είναι σχετικά αυξημένο σε σχέση με την αλληλούχηση κατά Sanger και το κόστος μιας αντίδρασης ξεφεύγει κατά πολύ σε σύγκριση με όλες τις υπόλοιπες μεθόδους αλληλούχησης.

### 1.4.3 Εργαλεία της αλληλούχησης νέας γενιάς

Ένα μεγάλο πλεονέκτημα των τεχνικών αλληλούχησης νέας γενιάς είναι η δυνατότητα χρήσης των διαφορετικών εφαρμογών που προσφέρει, ανάλογα με τον σκοπό της μελέτης. Πέραν του πιο βασικού εργαλείου της, την αλληλούχηση DNA (ή αλλιώς DNA-seq), προσφέρεται επιπλέον η αλληλούχηση RNA (ή αλλιώς RNA-seq) και η αλληλούχηση miRNA (ή αλλιώς miRNA-seq), η αλληλούχηση αμπλικονίων (ή αλλιώς amplicon sequencing, AmpliSeq) αλλά και η αλληλούχηση με ανοσοκατακρήμνιση χρωματίνης (ή αλλιώς Chromatin Immunoprecipitation Sequencing, ChIP-seq).

Η μεθοδολογία DNA-seq, ως βασική εφαρμογή του NGS δίνει τη δυνατότητα αλληλούχησης ολόκληρων γονιδιωμάτων οργανισμών, είτε είναι αυτό γνωστό είτε είναι άγνωστο. Η αλληλούχηση ολόκληρου γονιδιώματος ή Whole-Genome Sequencing (WGS) επιτρέπει την μελέτη και κατανόηση της δομής του, την συσχέτιση του με διάφορες ασθένειες και κυρίως του καρκίνου και δίνει τη δυνατότητα αντιμετώπισής τους με μεγαλύτερη αποτελεσματικότητα. Ιδιαίτερα για τον ανθρώπινο οργανισμό η κλινική μελέτη ασθενειών μέσω του WGS είναι μια επιτακτική ανάγκη για την πρόληψη, διάγνωση και την επιτυχή αντιμετώπισή τους. Πέραν την WGS, η οποία λόγω του μεγάλου όγκου πληροφορίας που εξάγεται, απαιτεί περισσότερο χρόνο και κόστος, εφαρμόζεται και μια άλλη προσέγγιση με την αλληλούχηση του συνόλου των εξωνίων ή Whole-Exome Sequencing (WES). Σε αυτή την μεθοδολογία, στοχεύονται μόνο όλες οι κωδικές περιοχές του γονιδιώματος, οι οποίες αντιστοιχούν στο 2% του γονιδιώματος αλλά συσχετίζονται άμεσα με τις περισσότερες ασθένειες που μελετώνται [54]. Το μειωμένο κόστος της και η στοχευμένη εφαρμογή της, κάνει τη μεθοδολογία αυτή ιδανική για την μελέτη μεταλλάξεων σε γονίδια που σχετίζονται με διάφορες ασθένειες και τη μελέτη νέων διαγνωστικών εργαλείων για αυτές.

Ένα ευρέως χρησιμοποιούμενο εργαλείο του NGS, είναι το RNA-seq. Το εργαλείο αυτό δίνει την δυνατότητα μελέτης των επιπέδων έκφρασης γονιδίων που μας ενδιαφέρουν καθώς και τα επίπεδα έκφρασης αλληλομόρφων τους και τυχόν εναλλακτικών μεταγράφων, γνωστών ή αγνώστων, κυρίως για τα long noncoding RNAs [55]. Το πλεονέκτημα σε αυτή τη τεχνική είναι η αποκάλυψη εναλλακτικών μεταγράφων τα οποία βρίσκονται σε πολύ χαμηλή έκφραση, τόσο που δεν μπορούν να ανιχνευτούν με κάποια άλλη τεχνική αλληλούχησης. Επιπλέον, ο ερευνητής έχει τη δυνατότητα ακριβούς ταυτοποίησης των γονιδίων (gene annotation) και των μεταγραφικών τους ορίων καθώς και των εκφραζόμενων μονονουκλεοτιδικών πολυμορφισμών που εμφανίζουν, τα λεγόμενα SNPs. Η μέθοδος miRNA-seq, μπορεί να θεωρηθεί ως μια υποκατηγορία της τεχνικής RNA-seq, η οποία όπως αποκαλύπτει και η ονομασία της στοχεύει στην ανίχνευση μικρών μη κωδικών μορίων RNA και των γονιδίων τους. Ο ρυθμιστικός τους ρόλος στην γονιδιακή έκφραση, καθιστά την μελέτη τους απαραίτητη για την προγνωστική ή/και διαγνωστική αξία που μπορεί να έχουν σε διάφορες ασθένειες, ιδιαίτερα στον καρκίνο.

Ένα ακόμη πολύ χρήσιμο εργαλείο είναι το AmpliSeq, με το οποίο αλληλουχούνται συγκεκριμένες περιοχές ενός panel γονιδίων-στόχων, οι οποίες αρχικά έχουν ενισχυθεί με τη μέθοδο PCR. Με αυτή τη προσέγγιση, επιτυγχάνεται η αλληλούχηση μόνο στις επιλεγμένες περιοχές του panel και όχι οποιαδήποτε άλλη περιοχή του γονιδιώματος. Μια διαφορετικού τύπου προσέγγιση είναι η τεχνική ChIP-seq, η οποία είναι ιδανική για την μελέτη των αλληλεπιδράσεων μεταξύ πρωτεϊνών και DNA in vivo, δίνοντας τη δυνατότητα για χαρτογράφηση ρυθμιστών χρωματίνης και μεταγραφικών παραγόντων τα οποία προσδένονται στο DNA.

#### 1.4.4 Πλατφόρμες αλληλούχησης νέας γενιάς

Στην αγορά προσφέρονται σήμερα αρκετές πλατφόρμες NGS, οι οποίες διαφέρουν κυρίως στη χημεία της αλληλούχησης, οδηγώντας σε διαφορές στο μήκος ανάγνωσης της αλληλουχίας, το ποσοστό σφάλματος και την απόδοση αλληλούχησης, την κάλυψη του γονιδιώματος αλλά και τον χρόνο εκτέλεσης και το κόστος [56]. Καθώς όλες οι πλατφόρμες εμφανίζουν πλεονεκτήματα και μειονεκτήματα, η επιλογή της κατάλληλης πλατφορμας NGS γίνεται με βάση τις ανάγκες της πειραματικής διαδικασίας.

Η πρώτη εταιρία που κυκλοφόρησε πλατφόρμα NGS στην αγορά ήταν η 454 Life Sciences/Roche, το 2005. Η τεχνολογία στην οποία βασίστηκε, αφορούσε την λεγόμενη πυροαλληλούχηση, δηλαδή την ανίχνευση πυροφωσφορικού το οποίο απελευθερώνεται κατα

την ενσωμάτωση κάθε νουκλεοτιδίου μέσω DNA πολυμεράσης και προκαλεί μια σειρά αντιδράσεων που οδηγούν σε παραγωγή φωτός μέσω του ενζύμου λουσιφεράσης, το οποίο απαιτείται για τον προσδιορισμό της αλληλουχίας του DNA. Σε αυτή τη τεχνική το δικλωνα DNA σπάει σε μικρότερα θραύσματα στα άκρα των οποίων επιτρέπεται η προσδεση από adapters, συμπληρωματικούς ως προς τις αλληλουχίες των εκκινητών. Οι adapters δρουν ως θέσεις πρόσδεσης εκκινητών οι οποίοι βρίσκονται στην επιφάνεια ειδικών σφαιριδίων, έτσι γίνεται η υβριδοποίησή τους. Έπειτα κάθε σύμπλοκο σφαιριδίου – θραύσματος, απομονώνεται σε μικκύλια που σχηματίζονται από νερό και έλαιο, όπου γίνεται ενίσχυση κάθε αλληλουχίας DNA του θραύσματος, μέσω PCR και ακολουθεί μαζική αλληλούχηση [51]. Μετά την τελευταία ανανέωση του συστήματος σε GS FLX, πραγματοποιείται παραγωγή μεγάλου μήκους αναγνώσεων (700bp) με μεγάλη ταχύτητα αντίδρασης (χρόνος αντίδρασης- 24 ώρες) που αποδίδουν μία υψηλής απόδοσης αλληλούχηση [51]. Ωστόσο η μεθοδολογία μειονεκτεί στο μεγάλο κόστος των αντιδραστηρίων, γεγονός που την καθιστά μη ιδανική για κάθε χρήση.

Μια ακόμη πλατφόρμα NGS είναι η Genome Analyzer (Solexa), της εταιρίας Illumina η οποία κυκλοφόρησε το 2006 και βασίζεται σε μια μέθοδο που ονομάζεται “Sequencing-by-synthesis”. Το DNA κατακερματίζεται σε μικρά θραύσματα στα οποία προσδέονται adapters και στα δύο τους άκρα. Τα θραύσματα αυτά φορτώνονται σε στερεή επιφάνεια γνωστή ως “flow cell” όπου τα άκρα του κάθε θραύσματος υβριδίζονται με τα συμπληρωματικά ολιγονουκλεοτίδια που βρίσκονται δεσμευμένα στην επιφάνεια αυτή, δημιουργώντας μια γέφυρα. Έπειτα γίνεται προσθήκη DNA πολυμεράσης και dNTPs και ακολουθεί PCR για την ενίσχυση του κάθε εκμαγείου, η λεγόμενη “bridge PCR” και την δημιουργία εκατομμυρίων πυκνών συστάδων δίκλωνου DNA, οι οποίες περιέχουν πολλά αντίγραφα ενισχυμένων μορίων DNA πάνω στην στερεή επιφάνεια [51]. Οι συστάδες αυτές αποδιατάσσονται και σε κάθε μονόκλωνο DNA υβριδοποιείται ένας ειδικός για την αλληλούχηση εκκινητής. Με δράση DNA πολυμεράσης γίνεται προσθήκη ενός νουκλεοτιδίου ανα κύκλο αλληλούχησης, το οποίο αναγνωρίζεται από μια μοναδική φθορίζουσα ετικέτα που φέρει. Η προσθήκη δεύτερου νουκλεοτιδίου εμποδίζεται καθώς τα νουκλεοτίδια φέρουν 3'-OH άκρο. Στο τέλος κάθε κύκλου καταγράφεται το σήμα που εκπέμπεται μέσω ακτίνας λέιζερ, για κάθε συστάδα ξεχωριστά και η 3'-OH τροποποίηση μαζί με την φθορίζουσα ετικέτα απομακρύνονται για να συνεχίσει η αλληλούχηση με την ενσωμάτωση του επόμενου νουκλεοτιδίου [57]. Η τεχνική αυτή καταφέρνει να αλληλουχεί τμήματα 36bp και να παράγει 50-60 εκατομμύρια reads σε κάθε αντίδραση.

## 2. Υλικά και Μέθοδοι

### 2.1 Βιολογικό Υλικό

Κατα την περάτωση της διπλωματικής εργασίας, έγινε χρήση συνολικά 32 κυτταρικών σειρών, οι οποίες αναγράφονται στον **Πίνακα 2.1** και προέρχονται από ανθρώπινους ιστούς.

**Πίνακας 2.1.** Καταγραφή 32 κυτταρικών ιστών από ανθρώπινους ιστούς, οι οποίοι ομαδοποιούνται με βάση τον τύπο καρκίνου και τον ιστό προέλευσης, σε 7 διαφορετικές ομάδες.

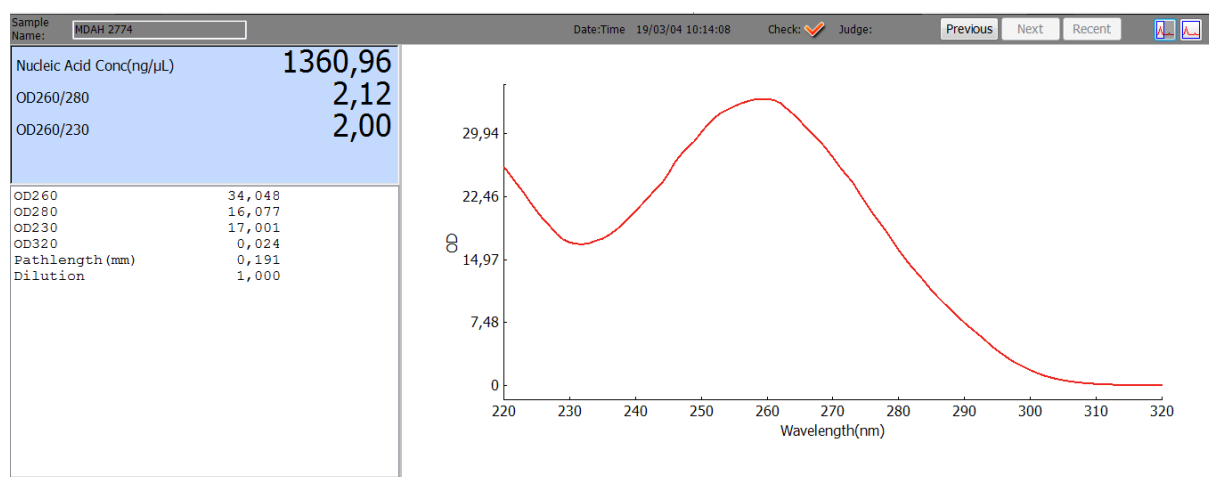
	Τύπος καρκίνου/Ιστός προέλευσης	Κυτταρικές Σειρές
1	Αδενοκαρκίνωμα μαστού	MCF-7, SK-BR-3, BT-20, MDA-MB-231, MDA-MB-468
2	Καρκίνος των ωοθηκών	OVCAR-3, SK-OV-3, ES-2, MDAH-2774
3	Καρκίνος του προστάτη	PC-3, DU 145, LNCaP
4	Καρκίνος του παχέος εντέρου	Caco-2, DLD-1, HT-29, HCT 116, SW 620, COLO 205, RKO
5	Καρκίνος του εγκεφάλου	U-87MG, U-251 MG, D54, H4, SH-SY5Y
6	Λευχαιμία	K-562, HL-60, Jurkat, REC-1, SU-DHL-1, GRANTA-519
7	Αδενοκαρκίνωμα του ενδομητρίου	Ishikawa, SK-UT-1B

Σε πρώτο στάδιο, χρησιμοποιήθηκαν έτοιμα pools από cDNA των πρώτων 6 ομάδων. Σε μετέπειτα στάδιο, λόγω των απαιτήσεων της πειραματικής διαδικασίας, έγινε χρήση συγκεκριμένων κυτταρικών σειρών ενδεικτικά από κάθε καρκινικό τύπο. Η απομόνωση ολικού RNA από ανθρώπινες αυτές κυτταρικές σειρές είχε προηγηθεί στο εργαστήριο, οπότε έγινε κατευθείαν χρήση του ολικού RNA.

### 2.2 Φασματοφωτομετρικός προσδιορισμός του απομονωμένου ολικού RNA

Πραγματοποιήθηκε φασματοφωτομετρικά, ο ποσοτικός προσδιορισμός της συγκέντρωσης του ολικού RNA το οποίο υπήρχε στο εργαστήριο, για να ελεγχθεί κατά πόσο είναι κατάλληλο για χρήση. Η μέτρηση γίνεται στο φωτόμετρο BioSpec-nano Micro-volume UV-Vis Spectrophotometer στο οποίο υπολογίζεται ο λόγος των απορροφήσεων σε μήκη κύματος 260 nm και 280 nm ( $\lambda=A_{260}/A_{280}$ ) για τον προσδιορισμό της συγκέντρωσης του

RNA σε  $\mu\text{g}/\mu\text{L}$ . Τα νουκλεϊκά οξέα έχουν μέγιστο απορρόφησης στα 260 nm εξαιτίας των αρωματικών δακτυλίων των πυριμιδινών και του ιμιδαζολίου του σκελέτου των αζωτούχων βάσεων τους, ενώ οι πρωτεΐνες έχουν μέγιστο απορρόφησης στα 280 nm λόγω των δακτυλίων των αρωματικών αμινοξικών καταλοίπων τους [58]. Για τα δείγματα που επισημαίνονται πιο πάνω, έγινε χρήση 1 $\mu\text{L}$  από κάθε δείγμα ολικού RNA και υπολογίστηκε αυτόματα τη συγκέντρωση του RNA σε  $\mu\text{g}/\mu\text{L}$  όπως παρουσιάζεται στην **Εικόνα 2.1**. Οι λόγοι απορροφήσεων για τα δείγματα RNA που μετρήθηκαν ήταν ικανοποιητικοί και κυμαίνονταν μεταξύ 1,8 και 2,2.



**Εικόνα 2.1.** Αντιπροσωπευτικό αποτέλεσμα φωτομέτρησης ενός δείγματος RNA για την κυτταρική σειρά MDAH-2774, με τη χρήση του BioSpec-nano Micro-volume UV-Vis Spectrophotometer

### 2.3 Αντίστροφη Μεταγραφή ( Reverse Transcription, RT )

Τα δείγματα ολικού RNA στα οποία έγινε ο ποσοτικός προσδιορισμός, χρησιμοποιήθηκαν για την αντίδραση RT, με σκοπό την αντίστροφη μεταγραφή τους σε cDNA και την μελέτη της έκφρασης των γονιδίων των KLKs σε αυτά.

Η αντίδραση της RT, βασίζεται στην δράση ενός ιικού ενζύμου, της αντίστροφης μεταγραφάσης (Reverse Transcriptase, RTase) η οποία είναι στην ουσία μια RNA-εξαρτώμενη DNA πολυμεράση που συνθέτει μόρια cDNA έχοντας ως εκμαγείο τα μόρια RNA. Για την δράση αυτού του ενζύμου, απαιτείται μια ολιγονουκλεοτιδική αλληλουχία που θα χρησιμοποιηθεί ως εκκινητής από την RTase. Οι εκκινητές σχεδιάζονται κατάλληλα για την αντίδραση και υβριδοποιούνται στα μόρια RNA. Γενικά για την αντίδραση αυτή υπάρχουν τρεις κατηγορίες εκκινητών που μπορούν να σχεδιαστούν, τα ολιγομερή δεοξυθυμιδίνης (oligo-dT) τα οποία υβριδοποιούνται στην poly-A ουρά των ευκαρυωτικών

mRNAs, τα τυχαία εξαμερή και οι ειδικοί εκκινητές για συγκεκριμένες νουκλεοτιδικές αλυσίδες RNA. Η πρώτη κατηγορία είναι η πιο συχνά επιλεγόμενη καθώς έχει τη μέγιστη απόδοση για τον σχηματισμό μόνο των mRNA μεταγράφων, τα οποία αντιπροσωπεύουν ένα πολύ μικρό ποσοστό του ολικού RNA των ανθρώπινων κυττάρων (1-5%). Σε αντίθεση με αυτό, τα τυχαία εξαμερή υβριδοποιούνται σε όλους τους τύπους RNA, καθιστώντας τη κατηγορία μη ειδική. Η μέγιστη ειδικότητα προσφέρεται με τη χρήση ειδικού εκκινητή για κάποιο συγκεκριμένο κλώνο RNA που είναι στόχος, με αποτέλεσμα την αποφυγή σχηματισμού μη επιθυμητών μορίων cDNA.

Η πειραματική διαδικασία των αντιδράσεων RT, αποτελείται από δύο διακριτά στάδια, τα οποία πραγματοποιούνται σε θερμικό κυκλοποιητή. Για την συγκεκριμένη αντίδραση RT, έγινε χρήση ενός τροποποιημένου oligo-dT εκκινητή (oligo-dT-adapter), με νουκλεοτιδική αλληλουχία 5'-GCGAGCACAGAATTAATACGACTCACTATAGGTTTTTTTTTTTNN-3', όπου V= G, A, C και N = G, A, T, C. Το 5' άκρο με το ολιγονουκλεοτιδικό τμήμα μήκους 32nt ονομάζεται αλληλουχία-προσαρμογέας (adapter sequence) και τα υπόλοιπα 12 κατάλοιπα θυμιδίνης του εκκινητή είναι το τμήμα που πρόκειται να υβριδοποιηθεί με τις αντίστοιχες αδενosίνες του 3' άκρου των mRNA μεταγράφων. Στο 3' άκρο του εκκινητή υπάρχουν τα δύο εκφυλισμένα nt όπου το V μπορεί να υβριδοποιηθεί με όλες τις αζωτούχες βάσεις πέραν της αδενίνης, ενώ το N υβριδοποιείται με όλες τις βάσεις. Αυτή η διαμόρφωση εξασφαλίζει την υβριδοποίηση του V με το τελευταίο νουκλεοτίδιο του 3' άκρου πριν ξεκινήσει η poly(A) ουρά του mRNA. Για τη συγκεκριμένη διπλωματική εργασία, η χρήση του oligo-dT-adapter είναι απαραίτητη για τη δημιουργία των κατάλληλων μορίων cDNAs, τα οποία θα χρησιμοποιηθούν αργότερα για την εκτέλεση μιας μεθοδολογίας αλυσιδωτής αντίδρασης ταχείας ενίσχυσης του 3' άκρου (3' Rapid Amplification of cDNA Ends, 3' RACE), με στόχο την ενίσχυση των μεταγράφων γονιδίων-στόχων και την μελέτη των 3' UTRs τους.

Στο πρώτο στάδιο των αντιδράσεων της RT, πραγματοποιείται ανάμιξη όγκου ολικού RNA που αντιστοιχεί σε 2 µg, 1 µL διαλύματος δεοξυνουκλεοτιδίων dNTPs (0.5 µM το καθένα), 1 µL oligo-dT-adapter (100 µM) και συμπληρώνεται με ultrapure-H<sub>2</sub>O μέχρι τελικό όγκο 12 µL. Ακολουθεί ήπια ανάδευση, επώαση στους 65°C για 5 min και γρήγορη μεταφορά των διαλυμάτων στον πάγο, επιτυγχάνοντας την πρόσδεση των μορίων του oligo-dT-adapter στην poly(A) ουρά των mRNAs. Στο δεύτερο στάδιο, κάθε διάλυμα συμπληρώνεται με 4 µL ρυθμιστικού διαλύματος 5X First-Strand Buffer, 2 µL DDT (0.1 M), 0.5 µL αναστολέα Rnaseών (RNase OUT, 40 units/µL), 0.5 µL αντίστροφης μεταγραφάσης

M-MLV (200 units/ $\mu$ L) και ultrapure- H<sub>2</sub>O έως τελικό όγκο 20  $\mu$ L. Ακολουθεί επώαση για 52 min στους 37°C, που αποτελεί τη βέλτιστη θερμοκρασία δράσης της αντίστροφης μεταγραφάσης M-MLV και 15 min στους 70°C, για τον τερματισμό της αντίδρασης. Ο θερμικός κυκλοποιητής που χρησιμοποιήθηκε για τις αντιδράσεις RT και τη σύνθεση των μορίων cDNAs, ήταν ο GeneAmp PCR System 9700 (Applied Biosystems™, Thermo Fisher Scientific Inc.). Μετά το τέλος του δεύτερου σταδίου, σχηματίστηκαν τα δείγματα cDNAs που προέκυψαν από την μεταγραφή των mRNAs των ανθρώπινων κυτταρικών σειρών που χρησιμοποιήθηκαν, με σκοπό την μελέτη έκφρασης γονιδίων-στόχων. Τα δείγματα αυτά αποθηκεύτηκαν στους -20°C.

#### 2.4 Βιοπληροφορική ανάλυση

Για την περάτωση της συγκεκριμένης διπλωματικής εργασίας, η μελέτη και ο προσδιορισμός εναλλακτικών 3'-αμετάφραστων περιοχών (3'-UTRs), στα γονίδια των KLKs, απαιτεί αρχικώς την ανάλυση δεδομένων από τα αποτελέσματα αλληλούχησης νέας γενιάς (NGS). Η αντιδράσεις NGS, πραγματοποιήθηκαν στα πλαίσια μιας διδακτορικής διατριβής και αποσκοπούν στην ταυτοποίηση και μελέτη της έκφρασης νέων mRNA ισομορφών για τα γονίδια των KLKs. Η βιοπληροφορική ανάλυση, η οποία ακολουθεί τις αντιδράσεις NGS, είναι ένα αρκετά χρονοβόρο στάδιο καθώς απαιτεί την επεξεργασία ενός τεράστιου όγκου δεδομένων το οποίο έχει παραχθεί.

Από κάθε αντίδραση αλληλούχησης νέας γενιάς, προκύπτουν αποτελέσματα τα οποία αποθηκεύονται σε αρχεία τύπου FASTQ, μέσω του λογισμικού της κάθε πλατφόρμας αλληλούχησης. Ένα αρχείο FASTQ εμφανίζει συγκεκριμένη μορφοποίηση στην οποία κάθε καταγραφή από ένα read αποτελείται από τέσσερις συνεχόμενες γραμμές. Η πρώτη γραμμή, αρχίζει με το σύμβολο «@» και περιέχει τον αναγνωριστικό κωδικό (sequence identifier) της νουκλεοτιδικής αλληλουχίας, ενώ η δεύτερη γραμμή περιλαμβάνει την νουκλεοτιδική αλληλουχία. Πέραν των τεσσάρων dNTPs, μπορεί να εμφανίζεται και το γράμμα N, το οποίο αντιπροσωπεύει ένα dNTP το οποίο δεν έχει διαβαστεί ικανοποιητικά. Η τρίτη γραμμή περιέχει μόνο το σύμβολο «+», ενώ η τέταρτη και τελευταία περιέχει χαρακτήρες που αντιστοιχούν στην ποιότητα αλληλούχησης (quality score) κάθε dNTP από την νουκλεοτιδική αλληλουχία (**Εικόνα 2.2**)



```
1 @1BBE2:00932:00575
2 CTAGTCTCAGGCTTCAACATCGAATACGCCGACGGCCCTTCGCCCTATTCTTCATAGCCGAATACACAAACATTATTATAATAAACACCCCTCACCCTACAATCTTCCTAGG
3 +
4 C<<====EEF>CC=B?:?@:>::=4:6E<<8=BBCAEEF8:5:::1010*0<6<@BBD?DD@ECBB:::0:::2::2:399333)3388)333=-3333336833-3-333-
```

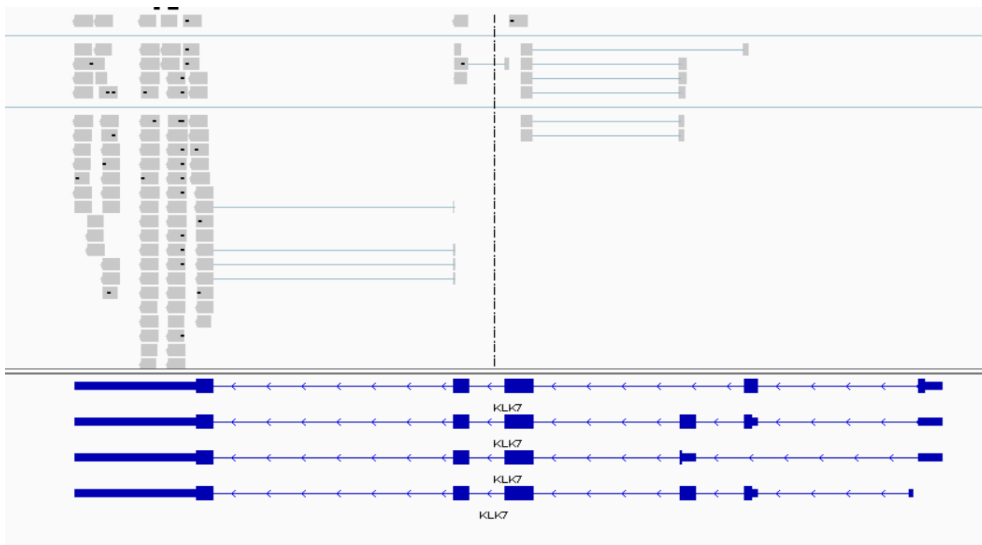
## Εικόνα 2.2. Δομή και μορφοποίηση μιας καταγραφής σε ένα αρχείο FASTQ.

Η βιοπληροφορική ανάλυση των αποτελεσμάτων του NGS, για την συγκεκριμένη διπλωματική εργασία, αποσκοπούσε στην ανίχνευση νέων πιθανών θέσεων 3'-αμετάφραστων περιοχών (3'-UTRs). Καθώς οι καταγραφές από την αλληλούχηση αντιστοιχούσαν σε θέσεις από όλα τα τμήματα των mRNAs των KLKs, εφαρμόστηκε μια μέθοδος διαχωρισμού και ομαδοποίησης των δεδομένων που αφορούσαν μόνο το 3'-άκρο των μεταγράφων. Αυτό επιτεύχθηκε χρησιμοποιώντας ένα λογάριθμο ο οποίος «ψαρεύει» όλες τις καταγραφές που διαθέτουν ένα συγκεκριμένο keyword «AAACCTATAGTGAGTCG» ως τμήμα της αλληλουχίας τους και το οποίο αντιστοιχεί στο συμπληρωματικό τμήμα από την αλληλουχία του oligo-dT-adapter. Έτσι δημιουργήθηκε αρχείο με όλες τις καταγραφές από αλληλουχίες που αντιστοιχούν στο τμήμα από το 3'-άκρο των mRNA των γνωστών μεταγράφων των γονιδίων-στόχων και πιθανόν και νέων εναλλακτικών 3'-UTRs για τα γονίδια αυτά.

Το επόμενο βήμα μετά την ομαδοποίηση των πειραματικών αλληλουχιών που μας ενδιαφέρουν, είναι η στοίχιση με το ανθρώπινο γονιδίωμα αναφοράς (GRCh38). Για την στοίχιση αλληλουχιών υπάρχουν διαθέσιμοι διάφοροι αλγόριθμοι ως εργαλεία σε πολλές πλατφόρμες. Για την στοίχιση των συγκεκριμένων πειραματικών αλληλουχιών έγινε χρήση του αλγόριθμου HISAT 2, μέσω της διαδικτυακής πλατφορμας Galaxy. Συγκεκριμένα, δίνεται ως input ένα αρχείο FASTQ και γίνεται μια επιλογή ανάμεσα στις διάφορες παραμέτρους που προσφέρει ο αλγόριθμος με στόχο την καλύτερη στοίχιση των αλληλουχιών. Με την ολοκλήρωση της στοίχισης, παράγεται ως output ένα αρχείο BAM (Binary Alignment/Map). Το αρχείο αυτό περιέχει τις επιτυχώς στοιχημένες αλληλουχίες σε οπτικοποιημένη μορφή. Με βάση το αρχείο αυτό μπορεί να δημιουργηθεί και ένα δεύτερο αρχείο BED (Browser Extensible Data), το οποίο περιέχει μόνο τις στοιχημένες αλληλουχίες που αφορούν εναλλακτικά σημεία συρραφής και είναι απαραίτητο για τη μελέτη του εναλλακτικού ματίσματος. Τελικώς, από την στοίχιση με τον αλγόριθμο HISAT 2, προέκυψαν δύο αρχεία, BAM και BED.

Όπως έχουμε αναφέρει τα δύο αυτά αρχεία περιέχουν τις οπτικοποιημένες πληροφορίες από την στοίχιση των πειραματικών αλληλουχιών με το ανθρώπινο γονιδίωμα. Η ανάλυσή τους μπορεί να επιτευχθεί μέσω προγράμματος που παρουσιάζει αυτές τις

πληροφορίες, όπως είναι το πρόγραμμα περιήγησης γονιδιωμάτων Integrative Genomics Viewer (IGV), με τη χρήση του οποίου γίνεται εύκολα και γρήγορα η οπτικοποίηση των αρχείων BAM και BED που έχουν προκύψει από τις τεχνικές NGS [59, 60]. Κατ' επέκταση γίνεται εύκολα εντοπίσιμη η θέση στοίχισης κάθε πειραματικής αλληλουχίας και δίνονται πληροφορίες για το μήκος και την νουκλεοτιδική της αλληλουχία (**Εικόνα 2.3**). Με αυτό τον τρόπο μπορούν να εντοπιστούν τόσο τα γνωστά όσο και τα πιθανά νέα σημεία εναλλακτικής συρραφής και να χρησιμοποιηθούν ως πληροφορία για τον σχεδιασμό πειραμάτων που να αποδεικνύουν την ύπαρξή τους.



**Εικόνα 2.3.** Οπτικοποίηση του αρχείου BAM μέσω προγράμματος IGV. Τα μπλε κουτιά αντιπροσωπεύουν τα εξώνια ενώ οι μπλε γραμμές τα εσώνια του γονιδιώματος αναφοράς. Επιλέον, τα γκρι κουτιά αντιπροσωπεύουν τις πειραματικές αλληλουχίες του αρχείου BAM, οι οποίες έχουν στοιχιστεί επιτυχώς στο γονιδίωμα αναφοράς μέσω του αλγορίθμου HISAT2.

Μετά το τέλος της βιοπληροφορικής ανάλυσης, επιλέχθηκαν συγκεκριμένα ανθρώπινα γονίδια *KLKs*, τα οποία εμφανίζουν πιθανά εναλλακτικά 3'-UTRs και στα οποία θα ακολουθήσει η ενίσχυση των 3'-UTRs για την περαιτέρω μελέτη και ταυτοποίησή τους.

## 2.5 Έλεγχος έκφρασης γονιδίων μέσω ποσοτικής PCR

Μετά την ολοκλήρωση των αντιδράσεων RT και τη σύνθεση των δειγμάτων cDNA, απαιτείται η μελέτη έκφρασης ενός γονιδίου σταθερής έκφρασης (housekeeping gene), το οποίο θα θεωρείται το γονίδιο αναφοράς (reference gene). Γενικότερα, τα γονίδια αναφοράς είναι γονίδια απαραίτητα για τη διατήρηση της βασικής κυτταρικής λειτουργίας έτσι διατηρούνται σε σχετικά σταθερά επίπεδα έκφρασης σε οποιοδήποτε ιστό του οργανισμού. Η

σύγκριση των επιπέδων έκφρασης των γονιδίων αυτών με τα επίπεδα έκφραση των γονιδίων-στόχων, για κάθε κυτταρική σειρά, μπορεί να δώσει μια πιο σαφή εικόνα για την ποιότητα των δειγμάτων cDNAs και να επιτρέψει την σχετική ποσοτικοποίηση της έκφρασης του κάθε γονιδίου-στόχου. Για την παρούσα διπλωματική εργασία, επιλέχθηκε το ανθρώπινο γονίδιο της αφυδρογονάσης της 3-φωσφορικής γλυκεραλδεΐδης (glyceraldehyde-3-phosphate dehydrogenase, *GAPDH*). Ο έλεγχος έκφρασης του ανθρώπινου γονιδίου αναφοράς *GAPDH* στα δείγματα cDNAs που μελετήθηκαν, πραγματοποιήθηκε μέσω αντιδράσεων Real-Time PCR για την ποσοτικοποίηση των προϊόντων που παράγονται.

Η αλυσιδωτή αντίδραση πολυμεράσης (Polymerase Chain Reaction, PCR) είναι μια τεχνική ενζυμικής ενίσχυσης μικρών τμημάτων DNA χρησιμοποιώντας ένα μεγαλύτερο δίκλωνο μόριο DNA. Συγκεκριμένα, γίνεται χρήση δύο εκκινητών (primers) οι οποίοι είναι ο καθένας συμπληρωματικός ως προς τη μια εκ των δύο αλυσίδων DNA αντίστοιχα και δράση του ενζύμου DNA πολυμεράσης, για την επιμήκυνση των αλληλουχιών και παραγωγή πολλών αντιγράφων τους. Η αντίδραση της PCR, περιλαμβάνει τρία επαναλαμβανόμενα στάδια, τους λεγόμενους κύκλους. Ανα κύκλο ενίσχυσης, γίνεται αποδιάταξη του DNA, πρόσδεση εκκινητών και σύνθεση της αλληλουχίας, με χρήση των υπάρχοντων αλληλουχιών ως εκμαγεία. Ως αποτέλεσμα, η ποσότητα συντιθέμενου DNA διπλασιάζεται ανα κύκλο αντίδρασης.

Η ποσοτική PCR (quantitative PCR, qPCR) θεωρείται μια γρήγορη και ευαίσθητη μέθοδος ποσοτικοποίησης μορίων DNA και cDNA, η οποία πραγματοποιείται σε εξειδικευμένο θερμικό κυκλοποιητή για την ανάλυση του φθορισμού που εκπέμπεται από διάφορες φθορίζουσες χρωστικές καθώς αυτές ενσωματώνονται στα προϊόντα της PCR. Η qPCR διαφέρει με την συμβατική PCR, στο στάδιο αντίδρασης όπου θα γίνει η συλλογή δεδομένων και η εξαγωγή αποτελεσμάτων. Σε αντίθεση με την συμβατική PCR όπου απαιτείται η ολοκλήρωση του θερμικού σταδίου της αντίδρασης μέχρι την φάση κορεσμού και έπειτα τα προϊόντα ηλεκτροφορούνται σε πήκτωμα αγαρόζης για την αξιολόγησή τους, στην qPCR τα δεδομένα της αντίδρασης μπορούν να συλλεχθούν και να αναλυθούν κατα την εκθετική φάση, με τη βοήθεια κατάλληλου προγράμματος στον ηλεκτρονικό υπολογιστή.

Υπάρχουν δύο τύποι qPCR, η qPCR τελικού σημείου (end-point qPCR) και η qPCR πραγματικού χρόνου (Real-Time PCR) οι οποίες εμφανίζουν διαφορές στη μεθοδολογία. Στην πρώτη περίπτωση, η ποσοτικοποίηση του προϊόντος γίνεται στο τέλος της αντίδρασης όπου η απόδοσή της έχει υποστεί μείωση ως συνέπεια της κατανάλωσης των αντιδρώντων και της συσσώρευσης των αναστολέων, έτσι τα αποτελέσματα δεν είναι απόλυτα αξιόπιστα.

Αντιθέτως, η Real-Time PCR προσφέρει πιο ακριβή ανάλυση των αποτελεσμάτων καθώς η μέτρηση των προϊόντων γίνεται κατά τη διάρκεια της αντίδρασης, όπου μετράται ο φθορισμός που εκπέμπεται ανα κύκλο και προκύπτει μια καμπύλη ενίσχυσης (amplification plot) μέσω προγράμματος. Η καμπύλη αυτή αποτελείται από την εκθετική φάση (exponential phase) όπου ανα κύκλο το προϊόν διπλασιάζεται σε ποσότητα, την γραμμική φάση όπου τα αντιδραστήρια μειώνονται σε ποσότητα και αρχίζει η συσσώρευση αναστολέων άρα η ταχύτητα αντίδρασης επιβραδύνεται και καταλήγει στη φάση κορεσμού όπου η αντίδραση σταματάει. Κατά την ποσοτικοποίηση η οποία γίνεται με τις τιμές από την εκθετική φάση της αντίδρασης, προκύπτει μια τιμή  $C_T$  (threshold cycle), η οποία αντιπροσωπεύει τον αριθμό των κύκλων που απαιτούνται ώστε η τιμή του φθορισμού που παρατηρείται να προσεγγίζει ένα συγκεκριμένο όριο (threshold) το οποίο ορίζεται πάνω από την αντίστοιχη του μη-ειδικού σήματος (background). Επιπλέον όσο μικρότερη είναι η τιμή  $C_T$ , τόσο υψηλότερη είναι η συγκέντρωση του αρχικού υποστρώματος. Για την σχετική ποσοτικοποίηση των συγκεντρώσεων των PCR προϊόντων, οι τιμές  $C_T$  των mRNA-στόχων, κανονικοποιούνται ως προς τις αντίστοιχες τιμές του ενδογενούς γονιδίου-ελέγχου. Η κανονικοποίηση ως προς το ενδογενές γονίδιο είναι απαραίτητη για τη διόρθωση των διαφορών μεταξύ των δειγμάτων, που μπορεί να προκύψουν από τη διαφορετική συγκέντρωση του αρχικού υποστρώματος (cDNA) ή λόγω διαφορών στην αποδοτικότητα της αντίδρασης ενίσχυσης.

## 2.6 Μεθοδολογία SYBR-Green I για Real Time PCR

Για τις απαιτήσεις της συγκεκριμένης διπλωματικής εργασίας, χρησιμοποιήσαμε την μεθοδολογία της SYBR-Green I, η οποία βασίζεται στην πρόσδεση της φθορίζουσας χρωστικής SYBR-Green στο δίκλωνο προϊόν PCR. Η χρωστική αυτή διεγείρεται σε ακτινοβολία μήκους κύματος 497 nm και εκπέμπει σε μήκος κύματος 520 nm, ενώ η ένταση φθορισμού είναι ανάλογη της συγκέντρωσης του προϊόντος. Έπιπλέον ο φθορισμός γίνεται μόνο κατα την ενσωμάτωσή της στο DNA και όχι όσο βρίσκεται ελεύθερη σε διάλυμα. Κατα την εκθετική φάση της PCR, ο φθορισμός διπλασιάζεται ανα κύκλο και μετά την ολοκλήρωση των 35 κύκλων, φτάνει στη φάση κορεσμού. Δεδομένου ότι η τιμή  $C_T$  είναι ανάλογη του λογαριθμού της αρχικής ποσότητας στόχου στο κάθε δείγμα, η σχετική συγκέντρωση ενός στόχου σε σχέση με μια άλλη, αντικατοπτρίζεται στην διαφορά στον αριθμό κύκλων ( $\Delta C_T$ ) η οποία είναι απαραίτητη για να φτάσουν τα δύο δείγματα στο ίδιο επίπεδο φθορισμού. Με αυτόν τον τρόπο μπορεί να γίνει μια σύγκριση των επιπέδων έκφρασης δύο γονιδίων-στόχων [61]. Κατα την εκτέλεση της Real-Time PCR με SYBR-

Green I, έγινε χρήση τριών σέτ εκκινητών για τα γονίδια *KLK9*, *KLK14* και του γονιδίου ελέγχου *GAPDH*, με συγκέντρωση 0.5μM. Η έκφραση των τριών γονιδίων μελετήθηκε σε 11 διαφορετικές κυτταρικές σειρές (**Πίνακας 2.2**), οι οποίες είναι αντιπροσωπευτικές από κάθε ομάδα όπως έχουν αναφερθεί πιο πάνω. Για τις αντιδράσεις Real-Time PCR, γίνεται ανάμειξη των δύο εκκινητών (1 μL ο κάθε ένας), του SYBR master mix (5 μL) και συμπλήρωση με ultrapure H<sub>2</sub>O μέχρι όγκο 8 μL. Έπειτα προστίθεται cDNA σε ποσότητα 2μL, στο οποίο έχει προηγηθεί αραίωση 1:10. Οι αντιδράσεις εκτελέστηκαν σε ειδικό θερμικό κυκλοποιητή, τον ABI 7500 Real-Time PCR System (Applied Biosystems™, Thermo Fisher Scientific Inc.). Το θερμικό πρωτόκολλο που εφαρμόστηκε, περιελάμβανε μια αρχική επώαση των δειγμάτων στους 95°C για 3 min, προκειμένου να ενεργοποιηθεί η πολυμεράση και στη συνέχεια 40 κύκλους, οι οποίοι αποτελούνταν από δύο θερμικά στάδια: ένα στους 95°C για 15 sec και ένα στους 60°C για 1 min. Μετά την ολοκλήρωση των κύκλων αντίδρασης, ακολουθεί η φάση αποδιάταξης για την εξαγωγή δεδομένων με σκοπό την μελέτη των καμπυλών αποδιάταξης (melting curves). Αυτό δίνει τη δυνατότητα να διαχωριστεί ο φθορισμός που προέκυψε από την ενίσχυση της αλληλουχίας-στόχου από τυχόν φθορισμούς που οφείλονται σε μη ειδικά προϊόντα ή σε διμερή εκκινητών.

**Πίνακας 2.2.** Καταγραφή κυτταρικών σειρών στις οποίες μελετήθηκε η έκφραση των γονιδίων *KLK9*, *KLK14* και *GAPDH*, με χρήση Real Time PCR.

	Τύπος καρκίνου/Ιστός προέλευσης	Κυτταρικές Σειρές
1	Αδενοκαρκίνωμα μαστού	MCF-7, SK-BR-3
2	Καρκίνος των ωοθηκών	SK-OV-3, ES-2, MDAH-2774
3	Καρκίνος του προστάτη	LNCaP
4	Καρκίνος του παχέος εντέρου	DLD-1, HT-29, HCT 116
5	Αδενοκαρκίνωμα του ενδομητρίου	Ishikawa, SK-UT-1B

Με τη βοήθεια ηλεκτρονικού υπολογιστή, ο οποίος ήταν συνδεδεμένος με τον θερμικό κυκλοποιητή, τα αποτελέσματα της αντίδρασης SYBR Green Real Time PCR, συλλέχθηκαν στο λογισμικό 7500 Software v2.3. Από τις τιμές C<sub>T</sub> των δειγμάτων για τα δύο γονίδια *KLK9* και *KLK14* και την κανονικοποίησή τους ως προς τις αντίστοιχες τιμές για το γονίδιο *GAPDH*, πραγματοποιήθηκε η σχετική ποσοτικοποίηση των συγκεντρώσεων των

δειγμάτων και ακολούθησε η επιλογή των κυτταρικών σειρών που εμφανίζουν πιο αυξημένη έκφραση των δύο πρώτων γονιδίων, με σκοπό την περαιτέρω ανάλυσή τους.

## 2.7 Αλυσιδωτή αντίδραση πολυμεράσης (PCR)

Όπως έχει αναφερθεί και νωρίτερα, η συμβατική PCR αφορά μια τεχνική ενίσχυσης των αλληλουχιών DNA *in vitro*, η οποία πραγματοποιείται με τη βοήθεια μιας θερμοανθεκτικής DNA πολυμεράσης, η οποία δρα σε κατάλληλο ρυθμιστικό διάλυμα το οποίο περιέχει συγκεκριμένη ποσότητα κατιόντων  $Mg^{2+}$ , dNTPs και ένα ζεύγος ολιγονουκλεοτιδίων, οι λεγόμενοι εκκινητές, που συνδέονται ειδικά στην αλληλουχία-στόχο. Ο σχεδιασμός κατάλληλων εκκινητών είναι κρίσιμος, καθώς αυτοί πρέπει να είναι συμπληρωματικοί ως προς τα άκρα των δύο αλυσίδων του τμήματος της αλληλουχίας-στόχου. Η αντίδραση της PCR, περιλαμβάνει τρία επαναλαμβανόμενα στάδια, τους λεγόμενους κύκλους κατά τους οποίους πραγματοποιείται η αποδιάταξη του δίκλωνου DNA, ο υβριδισμός των εκκινητών στην αλληλουχία-στόχο και η επιμήκυνση των εκκινητών μέσω πολυμερισμού, από την DNA πολυμεράση.

Αρχικά η αντίδραση PCR περιλαμβάνει το στάδιο αποδιάταξης (denaturation), το οποίο αφορά την θέρμανση του διαλύματος σε θερμοκρασίες 94-95°C, με σκοπό τον διαχωρισμό των αλυσίδων της διπλής έλικας του DNA, μέσω ρήξης των δεσμών υδρογόνου μεταξύ των συμπληρωματικών βάσεων στις δύο αλυσίδες του DNA. Ακολούθως εισέρχεται στο στάδιο των επαναλαμβανόμενων θερμικών κύκλων, συνήθως 30-40, με σκοπό τον υβριδισμό (annealing) των εκκινητών στην αλληλουχία DNA. Κάθε κύκλος αποτελείται από ένα σύντομο στάδιο αποδιάταξης DNA (94-95°C), το στάδιο υβριδισμού (annealing) των εκκινητών στην αλληλουχία DNA, το οποίο πραγματοποιείται συνήθως σε θερμοκρασίες 55-65°C και αφορά την θερμοκρασία τήξης των εκκινητών (Melting Temperature,  $T_m$ ) και τέλος το στάδιο προέκτασης (extension) των αλληλουχιών των εκκινητών που έχουν προσδεθεί στο DNA, από το ένζυμο DNA πολυμεράση, με προσθήκη δεοξυριβονουκλεοτιδίων, συμπληρωματικά με αυτά της αλυσίδας-εκμαγείου. Το στάδιο αυτό γίνεται σε θερμοκρασία 72°C, η οποία θεωρείται η βέλτιστη θερμοκρασία δράσης της Taq DNA πολυμεράσης, ενώ ο χρόνος του σταδίου καθορίζεται από το επιθυμητό μέγεθος της αλληλουχίας στόχου που θα επεκταθεί. Προαιρετικά μπορεί να πραγματοποιηθεί και το στάδιο τελικής προέκτασης (final extension) για τον πολυμερισμό πιθανών ανολοκλήρωτων προϊόντων της PCR [62].

Η επιτυχία μιας αντίδρασης PCR εξαρτάται από αρκετούς παράγοντες. Ο κυριότερος από αυτούς είναι η ποιότητα και η συγκέντρωση του DNA υποστρώματος. Ο τύπος υποστρώματος που επιλέγεται μπορεί να είναι γενωμικό DNA (genomic DNA), cDNA, πλασμίδια, φάγοι, BAC (Bacterial Artificial Chromosomes) ή YAC (Yeast Artificial Chromosomes) και η συγκέντρωσή του μπορεί να είναι λιγότερη από 1 ng έως και 1 μg ολικού γενωμικού DNA. Ένας ακόμα καθοριστικός παράγοντας αποτελεσματικότητας μιας αντίδρασης PCR, είναι η ειδικότητα των εκκινητών να υβριδοποιούνται αποκλειστικά στις συμπληρωματικές αλληλουχίες του υποστρώματος. Η θερμοκρασία υβριδοποίησης των εκκινητών, εξαρτάται τόσο από το μήκος της νουκλεοτιδικής αλυσίδας κάθε εκκινητή, όσο και από τη σύστασή του σε βάσεις GC. Έτσι επιλέγεται ως θερμοκρασία υβριδοποίησης, η θερμοκρασία  $T_m$  των εκκινητών, δηλαδή αυτή στην οποία η μισή ποσότητα εκκινητών είναι υβριδοποιημένη με την αλληλουχία στόχο. Γενικά, αυξάνοντας την θερμοκρασία υβριδοποίησης, αυξάνεται και η ειδικότητα προς το υπόστρωμα. Ένα αρκετά σημαντικό συστατικό κάθε αντίδρασης PCR είναι τα ιόντα μαγνησίου ( $Mg^{2+}$ ) τα οποία λειτουργούν ως μεταλλικός συμπαράγοντας της Taq DNA πολυμεράσης και η συγκέντρωση των οποίων επηρεάζει την ειδικότητα και την απόδοση της αντίδρασης. Επιλέγεται συγκέντρωση 0,5 έως 5 mM, με συνηθέστερη την συγκέντρωση 2,5 mM. Επιπλέον, οι υψηλές συγκεντρώσεις  $Mg^{2+}$  προκαλούν αυξημένη εμφάνιση λαθών κατά την δράση της Taq DNA πολυμεράσης, ενώ η χαμηλές συγκεντρώσεις προκαλούν μειωμένη αποδοτικότητα στην αντίδραση. Ένας ακόμη σημαντικός παράγοντας που επηρεάζει την εξέλιξη της αντίδρασης PCR, είναι το μίγμα των dNTPs, το οποίο συνήθως βρίσκεται σε ίσες συγκεντρώσεις των τεσσάρων δεοξυριβονουκλεοτιδίων, που κυμαίνονται από 50 έως 200 μM. Αύξηση της συγκέντρωσης των dNTPs μειώνει την πιστότητα της αντίδρασης ενώ με χαμηλές συγκεντρώσεις επηρεάζεται η αποδοτικότητα της αντίδρασης. Επιπλέον, τα dNTPs δεσμεύουν ισομοριακές ποσότητες  $Mg^{2+}$ , σχηματίζοντας σύμπλοκα dNTP- $Mg^{2+}$  μειώνοντας και την συγκέντρωση των ελεύθερων  $Mg^{2+}$ .

## 2.8 Σχεδιασμός εκκινητών για PCR

Ο κατάλληλος σχεδιασμός και σύνθεση των εκκινητών είναι το κλειδί για μια επιτυχημένη αντίδραση PCR. Η διαδικασία σχεδιασμού εκκινητών, για μια γνωστή αλληλουχία-στόχο δεν είναι πάντοτε απλή, έτσι έχουν αναπτυχθεί διάφορα λογισμικά προγράμματα και αλγόριθμοι για την διευκόλυνση της διαδικασίας αυτής. Κατά την περάτωση της συγκεκριμένης διπλωματικής εργασίας, έγινε χρήση του διαδουκτικού

αλγορίθμου Primer-BLAST (National Center for Biotechnology Information, NCBI). Ο σωστός και αποτελεσματικός σχεδιασμός εκκινητών προϋποθέτει την εφαρμογή των πιο κάτω κανόνων:

1. Ζεύγος εκκινητών, ειδικό για την επιθυμητή αλληλουχία-στόχο.
2. Ιδανικό μήκος νουκλεοτιδικής αλληλουχίας των εκκινητών κυμαίνεται από 16 έως 28 nt, παρόλο που σε ειδικές περιπτώσεις μπορεί να ξεπεράσουν τα 35 nt.
3. Κατάλληλη ποσοστιαία αναλογία GC ενός εκκινητή συνήθως στην περιοχή 40-60%.
4. Θερμοκρασία  $T_m$  εκκινητών κυμαίνεται από 55-65°C. Η ελάχιστη δυνατή διαφορά θερμοκρασιών  $T_m$  για τον πρόσθιο και ανάστροφο εκκινητή χωρίς να ξεπερνάει τους 3°C.
5. Θερμοκρασία υβριδισμού ίση ή με την ελάχιστη δυνατή διαφορά από τις θερμοκρασίες  $T_m$  των εκκινητών.
6. Αποφυγή χρήσης ζευγών εκκινητών με συμπληρωματικότητα στο 3' άκρο, καθώς προκαλεί τον σχηματισμό “διμερών εκκινητών” (primer dimers), όπου γίνεται προέκταση του εκκινητή με μήτρα τον εαυτό του ή τον άλλον εκκινητή [63].
7. Αποφυγή χρήσης νουκλεοτιδικών αλληλουχιών με επαναλαμβανόμενες περιοχές που διαθέτουν πολυπυρίνες ή πολυπυριμιδίνες, καθώς μπορούν να υβριδοποιηθούν και σε άλλες περιοχές του DNA υποστρώματος, δημιουργώντας μη ειδικά προϊόντα.
8. Αποφυγή εκκινητών που προβλέπεται να έχουν σημαντική δευτεροταγή δομή, ιδιαίτερα στο 3' άκρο.

Επιπρόσθετα, πολύ σημαντικό κριτήριο είναι η συγκέντρωση των εκκινητών σε μια αντίδραση PCR. Η συγκέντρωση των δύο εκκινητών πρέπει να είναι ίση και συνήθως κυμαίνεται από 0,1 έως 1  $\mu\text{M}$ .

## 2.9 3' Rapid amplification of cDNA ends (3' RACE)

Για τις απαιτήσεις της συγκεκριμένης διπλωματικής εργασίας, εφαρμόστηκε μια παραλλαγή της συμβατικής μεθόδου PCR, η μέθοδος της 3' RACE η οποία επιτρέπει την ενίσχυση των 3' άκρων σε τμήματα cDNA. Αυτό επιτυγχάνεται με τη χρήση ενός ανάστροφου μη ειδικού εκκινητή, ο οποίος συνδέεται με τις poly(A) ουρές των mRNA και ενός πρόσθιου εκκινητή που στοχεύει ειδικά την επιθυμητή αλληλουχία. Κατ' επέκταση, καταφέρνουμε να ενισχύσουμε τα 3' UTRs των γονιδίων-στόχων για την μελέτη τους και την ταυτοποίηση πιθανών εναλλακτικών των 3'-UTR μεταγράφων.



## 2.10 Nested PCR

Σε συνδυασμό με την μεθοδο της 3' RACE, εφαρμόστηκε και η μέθοδος “Nested PCR”, με σκοπό την αύξηση της ευαισθησίας και της ειδικότητας της αντίδρασης PCR. Η μεθοδολογία περιλαμβάνει δύο διαδοχικές αντιδράσεις PCR με δύο ζεύγη εκκινητών. Το πρώτο ζεύγος εκκινητών είναι σχεδιασμένο με σκοπό την ενίσχυση ενός μεγάλου τμήματος DNA που περιέχει την αλληλουχία-στόχο. Το ενισχυμένο προϊόν που προκύπτει από την πρώτη PCR υφίσταται αραίωση και χρησιμοποιείται ως υπόστρωμα για την δεύτερη αντίδραση. Σε αυτήν, το δεύτερο ζεύγος εκκινητών είναι σχεδιασμένο έτσι ώστε να ενισχύει ένα πιο εσωτερικό τμήμα της αλληλουχίας-στόχου. Αυτή η τεχνική επιτυγχάνει να αυξήσει αρκετά την ειδικότητα της αντίδρασης καθώς ενισχύει ένα ειδικό προϊόν ενώ παράλληλα εμποδίζει την ενίσχυση μη ειδικών παραπροϊόντων που μπορεί να έχουν σχηματιστεί στη πρώτη αντίδραση.

## 2.11 Ενίσχυση εναλλακτικών 3' UTRs των ανθρώπινων γονιδίων *KLKs* με χρήση nested 3' RACE

Από την βιοπληροφορική ανάλυση έγινε η επιλογή συγκεκριμένων ανθρώπινων γονιδίων *KLKs* τα οποία εμφανίζουν πιθανά εναλλακτικά 3'-UTRs, τα οποία μπορεί να είναι είτε μικρότερου μεγέθους νουκλεοτιδικής αλληλουχίας είτε μεγαλύτερου, σε σύγκριση με τις γνωστές 3'-UTRs. Ακολούθως, έγινε ο σχεδιασμός κατάλληλων ζευγών εκκινητών για την πραγματοποίηση αντιδράσεων nested 3' RACE (Πίνακας 2.3).

Αρχικά έγινε χρήση των 6 πρώτων ομάδων από cDNA pools (Πίνακας 2.1) τα οποία χρησιμοποιήθηκαν ως υποστρώματα για τις αντιδράσεις 3' RACE, με χρήση συγκεκριμένων ζευγών εκκινητών από τον Πίνακα 2.3. Τα προϊόντα PCR που προέκυψαν, αραιώθηκαν κατάλληλα για να χρησιμοποιηθούν ως υπόστρωμα στις αντιδράσεις nested 3'RACE που ακολούθησαν. Με τον ίδιο τρόπο, σε μετέπειτα στάδιο κατα την εκπόνηση της διπλωματικής αυτής εργασίας, έγινε επιλογή συγκεκριμένων κυτταρικών σειρών, οι οποίες είναι αντιπροσωπευτικές για κάθε τύπο καρκίνου/ιστό προέλευσης, οι οποίες χρησιμοποιήθηκαν ως υποστρώματα για τις αντιδράσεις 3' RACE και nested 3' RACE.

Όπως παρουσιάζεται και στον Πίνακα 2.3, για τις αντιδράσεις ενίσχυσης εναλλακτικών 3' UTRs, χρησιμοποιείται ένας γενικευμένος ανάστροφος εκκινητής (universal reverse primer) ο οποίος είναι σχεδιασμένος έτσι ώστε να υβριδοποιείται σε ένα τμήμα της αλληλουχίας του oligo-dT-adapter, ο οποίος χρησιμοποιήθηκε στις αντιδράσεις αντίστροφης μεταγραφής των mRNA. Οι πρόσθιοι εκκινητές σχεδιάστηκαν με στόχο να

υβριδοποιούνται στην περιοχή πριν το κωδικόνιο λήξης ή σε περιοχή πριν το τέλος εξωνίου όπου πιθανολογείται να υπάρχει εναλλακτικό μάτισμα με νέα θέση λήξης του μεταγράφου και νέο 3'-UTR. Έτσι, σε κάθε αντίδραση 3' RACE γίνεται χρήση ενός ειδικού πρόσθιου εκκινητή κάθε γονιδίου που έχει επιλεγθεί, σε συνδυασμό με τον universal reverse primer (3'RACE Outer Primer), ενισχύοντας το 3'- UTR κάθε γονιδίου.

	Όνομα εκκινητή	Αλληλουχία εκκινητή (5'→ 3')	Μήκος (nt)	Tm (°C)
3' RACE	KLK7 Ex7F	ACCGTTGGTGTGCAGAGGTA	20	61.40
Nested 3' RACE	KLK7 Ex7nF	GAGTCTACTCAAGTGTGCAAGT	24	60.74
3' RACE	KLK7 Ex4F	TCAGTGGCAATCAGCTCCAC	20	60.32
Nested 3' RACE	KLK7 Ex3/4F	CCACTGCAAGATGAATGAGTACAC	24	59.91
3' RACE	KLK8 Ex4F	TGTGTGATGGTGCCTCCAG	20	60.25
Nested 3' RACE	KLK8 Ex4nF	GCACTCCAGGGCATCACAT	19	60.08
3' RACE	KLK9 Ex5F	AGTCTACACCAGCGTATGCCA	21	61.23
Nested 3' RACE	KLK9 Ex5nF	CTGGATCCAAGAAATCATGGAGAAC	25	59.70
3' RACE	KLK14 Ex7F	ACCTGTGCAAGTACAGAAGCT	21	59.58
Nested 3' RACE	KLK14 Ex7nF	CAAGTACAGAAGCTGGATTGAGG	23	59.06
3' RACE	KLK15 Ex5F	ATTGTGTCTGGGGTGACGT	20	61.42
Nested 3' RACE	KLK15 Ex5nF	TGCCACTACTGGAGTGGATCA	22	61.09
3' RACE	3'RACE Outer Primer	GCGAGCACAGAATTAATACGACT	23	59.20
Nested 3' RACE	3'RACE Inner Primer	AGCACAGAATTAATACGACTCACTATAGG	29	60.73

**Πίνακας 2.3** Εκκινητές που χρησιμοποιήθηκαν για τις αντιδράσεις nested 3' RACE και ιδιότητές τους

Καθώς ο universal reverse primer δεν είναι ειδικός ως προς την κάθε αλληλουχία-στόχο, υβριδοποιείται με όλα τα μόρια cDNA που φέρουν την αλληλουχία του oligo-dT-adapter στο 3' άκρο, δημιουργώντας αρκετά παραπροϊόντα στην αντίδραση. Για το λόγο αυτό, τα προϊόντα από την 3' RACE PCR αραιώθηκαν σε ultrapure-H<sub>2</sub>O, κατά 1:100 και χρησιμοποιήθηκαν ως υπόστρωμα για τις αντιδράσεις nested 3' RACE. Σε αυτές τις αντιδράσεις έγινε χρήση ζεύγους εκκινητών με ένα πρόσθιο ειδικό εκκινητή για κάθε γονίδιο-στόχο, σχεδιασμένο λίγα nt έπειτα του προηγούμενου πρόσθιου εκκινητή, προς το 3' άκρο και ένας δευτέρος γενικευμένος ανάστροφος εκκινητής ο οποίος υβριδοποιείται μερικά

nt πιο εσωτερικά σε σχέση με τον πρώτο ανάστροφο εκκινητή. Με τον τρόπο αυτό επιτυγχάνεται η αύξηση της ειδικότητας της αντίδρασης ως προς τα προϊόντα-στόχους.

## 2.12 Συνθήκες αντιδράσεων nested 3' RACE για τα ανθρώπινα γονίδια των *KLKs*

Οι πρώτες αντιδράσεις των μεθοδολογιών 3' RACE και nested 3' RACE, στις οποίες έγινε χρήση των 6 pools cDNA, πραγματοποιήθηκαν σε μίγματα αντιδράσεων, τα οποία περιείχαν: ρυθμιστικό διάλυμα KAPA Taq Buffer A (Kapa Biosystems Inc), 0.2 mM από κάθε dNTP, 0.4 μM από κάθε εκκινητή, 0.2 units KAPA Taq DNA Polymerase (Kapa Biosystems Inc.), κατάλληλη ποσότητα DNA υποστρώματος και ultrapure-H<sub>2</sub>O σε τελικό όγκο 25μL. Επιπλέον, το θερμικό πρωτόκολο που χρησιμοποιήθηκε, ήταν ως εξής: Αρχική αποδιάταξη στους 95°C για 3 min, ακολουθεί το στάδιο 35 κύκλων που περιλάμβαναν ένα στάδιο αποδιάταξης στους 95°C για 30 sec, ένα στάδιο υβριδοποίησης για 30 sec και σε θερμοκρασία 60°C καθώς και ένα στάδιο επέκτασης σε 72°C και χρόνο βάσει του ρυθμού ενσωμάτωσης των nt από το ένζυμο (1 min/kb). Το στάδιο τελικής επέκτασης για όλα τα γονίδια-στόχους είχε θερμοκρασία 72°C για 1 min. Για τις αντιδράσεις αυτές έγινε χρήση του θερμικού κυκλοποιητή GeneAmp PCR System 9700 (Applied Biosystems™, Thermo Fisher Scientific Inc.). Οι αντιδράσεις αυτές πραγματοποιήθηκαν για τα έξι ζεύγη εκκινητών (**Πίνακα 2.4(α)**) και προέκυψαν συνολικά 36 διαφορετικά προϊόντα nested 3' RACE που αντιστοιχούν σε πέντε γονίδια *KLKs*.

Οι αντιδράσεις που ακολούθησαν, με τις μεθοδολογίες 3' RACE και nested 3' RACE, είχαν ως βασική διαφορά την χρήση κυτταρικών σειρών που επιλέχθηκαν (Πίνακας 2.2) και εξετάστηκαν σε ανεξάρτητες αντιδράσεις για δύο γονίδια των *KLKs*. Τα μίγματα αντιδράσεων ταυτίζονται με αυτά των πρώτων αντιδράσεων ενώ το θερμικό πρωτόκολο διαφοροποιείται. Για τις συγκεκριμένες αντιδράσεις, επιλέχθηκε ένα θερμικό πρωτόκολο το οποίο εξετάζει την ίδια αντίδραση σε έξι διαφορετικές θερμοκρασίες κατά το στάδιο υβριδοποίησης, τόσο για τις αντιδράσεις 3' RACE, όσο και για τις αντιδράσεις nested 3' RACE, οι οποίες είναι εύρους 58-63 °C. Οι υπόλοιπες συνθήκες του θερμικού πρωτοκόλου παραμένουν οι ίδιες. Από τις αντιδράσεις αυτές, προέκυψαν συνολικά 324 διαφορετικά προϊόντα nested 3' RACE από το γονίδιο *KLK9* που αντιστοιχούν σε 9 κυτταρικές σειρές και 72 διαφορετικά προϊόντα nested 3' RACE από το γονίδιο *KLK14* που αντιστοιχούν σε 2 κυτταρικές σειρές (**Πίνακας 2.4 (β)**).

**Πίνακας 2.4 (α)** Οι πρώτες αντιδράσεις μεθοδολογιών 3' RACE και nested 3'RACE, με χρήση 6 cDNA pools.

Pools cDNA	Τύπος καρκίνου/Ιστός προέλευσης	Κυτταρικές Σειρές	Πρόσθιοι εκκνητές για 3' RACE/nested 3'RACE
1	Αδενοκαρκίνωμα μαστού	MCF-7, SK-BR-3, BT-20, MDA-MB-231, MDA-MB-468	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF
2	Καρκίνος των ωοθηκών	OVCAR-3, SK-OV-3, ES-2, MDAH-2774	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF
3	Καρκίνος του προστάτη	PC-3, DU 145, LNCaP	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF
4	Καρκίνος του παχέος εντέρου	Caco-2, DLD-1, HT-29, HCT 116, SW 620, COLO 205, RKO	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF
5	Καρκίνος του εγκεφάλου	U-87MG, U-251 MG, D54, H4, SH-SY5Y	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF
6	Λευχαιμία	K-562, HL-60, Jurkat, REC-1, SU-DHL-1, GRANTA-519	KLK7 Ex7F/KLK7 Ex7nF, KLK7 Ex4F/KLK7 Ex3/4F KLK8 Ex4F/KLK8 Ex4nF, KLK9 Ex5F/KLK9 Ex5nF KLK14 Ex7F/KLK14 Ex7nF, KLK15 Ex5F/KLK15 Ex5nF

**Πίνακας 2.4 (β)** Οι αντιδράσεις μεθοδολογιών 3' RACE και nested 3'RACE, με διαβάθμιση θερμοκρασίας υβριδοποίησης (annealing gradient) σε 11 κυτταρικές σειρές

	Τύπος καρκίνου/Ιστός προέλευσης	Κυτταρικές Σειρές	Πρόσθιοι εκκινητές για 3' RACE/nested 3'RACE
1	Αδενοκαρκίνωμα μαστού	MCF-7	KLK14 Ex7F/ KLK14 Ex7nF
2	Αδενοκαρκίνωμα μαστού	SK-BR-3	KLK9 Ex5F/ KLK9 Ex5nF
3	Καρκίνος των ωοθηκών	SK-OV-3, ES-2	KLK9 Ex5F/ KLK9 Ex5nF
4	Καρκίνος των ωοθηκών	ES-2	KLK9 Ex5F/ KLK9 Ex5nF
5	Καρκίνος των ωοθηκών	MDAH-2774	KLK9 Ex5F/ KLK9 Ex5nF
6	Καρκίνος του προστάτη	LNCaP	KLK14 Ex7F/ KLK14 Ex7nF
7	Καρκίνος του παχέος εντέρου	DLD-1	KLK9 Ex5F/ KLK9 Ex5nF
8	Καρκίνος του παχέος εντέρου	HT-29	KLK9 Ex5F/ KLK9 Ex5nF
9	Καρκίνος του παχέος εντέρου	HCT 116	KLK9 Ex5F/ KLK9 Ex5nF
10	Αδενοκαρκίνωμα του ενδομητρίου	Ishikawa	KLK9 Ex5F/ KLK9 Ex5nF
11	Αδενοκαρκίνωμα του ενδομητρίου	SK-UT-1B	KLK9 Ex5F/ KLK9 Ex5nF

Η διαβάθμιση αυτή στη θερμοκρασία του σταδίου υβριδοποίησης (gradient), επιλέχθηκε με σκοπό την αύξηση της ειδικότητας της μεθόδου. Όπως έχει ήδη αναφερθεί, ένας από τους καθοριστικούς παράγοντες αποτελεσματικότητας μιας αντίδρασης PCR, είναι η ειδικότητα των εκκινητών να υβριδοποιούνται αποκλειστικά στις συμπληρωματικές αλληλουχίες του υποστρώματος, με αποτέλεσμα την ενίσχυση μόνο των αλληλουχιών-στόχων. Η ιδιότητα αυτή, επηρεάζεται από τη θερμοκρασία υβριδοποίησης των εκκινητών ή οποία θα πρέπει να ταυτίζεται όσο γίνεται στην θερμοκρασία  $T_m$  για κάθε ζεύγος εκκινητών. Η δοκιμή διαφορετικών θερμοκρασιών υβριδοποίησης για την ίδια αντίδραση, δίνει την δυνατότητα εύρεσης των ιδανικών συνθηκών για μια επιτυχή αντίδραση PCR.

### 2.13 Ηλεκτροφόρηση των προϊόντων nested 3' RACE σε πήκτωμα αγαρόζης

Μετά το τέλος των αντιδράσεων nested 3' RACE, με στόχο την ενίσχυση εναλλακτικών 3'-UTRs σε ανθρώπινα γονίδια *KLKs*, ακολούθησε η ηλεκτροφόρηση των προϊόντων των αντιδράσεων αυτών, σε πήκτωμα αγαρόζης έτσι ώστε να εκτιμηθεί η έκφρασή τους στα δείγματα cDNA που χρησιμοποιήθηκαν. Το πήκτωμα αγαρόζης έχει περιεκτικότητα 2% w/, σε ρυθμιστικό διάλυμα TBE 1X και περιέχει χρωστική Midori Green Advance σε ποσότητα 6 μL ανα 100 mL πηκτώματος. Μετά την παρασκευή του, γίνεται μεταφορά του πηκτώματος στη συσκευή ηλεκτροφόρησης (Mupid®-One electrophoresis system) και καλύπτεται πλήρως από ρυθμιστικό διάλυμα TBE 0.5X και ακολουθεί ηλεκτροφόρηση σε συνθήκες σταθερής τάσης 130V για 60 min. Τα δείγματα που ηλεκτροφορούνται, πριν το φόρτωμα υπόκεινται σε προετοιμασία, όπου προστίθεται ρυθμιστικό διάλυμα φόρτωσης 6X DNA Loading Dye (Thermo Fisher Scientific Inc.) ποσότητας 2 μL σε κάθε 10 μL προϊόντος PCR και ακολουθεί ήπια ανάδευση. Επιπλέον γίνεται ηλεκτροφόρηση ενός ακόμη δείγματος, το οποίο περιέχει μίγα τμημάτων DNA με γνωστό μοριακό βάρος, ο 100 bp DNA Ladder και χρησιμοποιείται ως μάρτυρας των ζωνών DNA. Μετά το τέλος της ηλεκτροφόρησης, το πήκτωμα μεταφέρεται σε τράπεζα υπεριώδους φωτός για τη συλλογή των αποτελεσμάτων της ηλεκτροφόρησης και την αξιολόγηση των αντιδράσεων nested 3' RACE.

### 2.14 Απομόνωση και καθαρισμός των nested 3' RACE προϊόντων από το πήκτωμα αγαρόζης και αλληλούχηση κατα Sanger

Μετά την ηλεκτροφόρηση των προϊόντων σε πήκτωμα αγαρόζης, από τις αντιδράσεις nested 3' RACE, ακολούθησε η ταυτοποίηση των πιθανών εναλλακτικών 3'-UTRs βάσει βιοπληροφορικής ανάλυσης των δεδομένων από NGS, μέσω αξιολόγησης των αποτελεσμάτων των ηλεκτροφορήσεων. Ακολούθως, έγινε η απομόνωση και ο καθαρισμός των προϊόντων τα οποία είναι υποσχόμενα και ταυτίζονται με τις ενδείξεις για εναλλακτικά 3'-UTRs στα γονίδια-στόχους, με σκοπό την αλληλούχηση τους κατα Sanger και την επιβεβαίωση της αλληλουχίας τους. Η επιλογή των ζωνών προς απομόνωση γίνεται αρχικώς βάσει του μήκους της αλληλουχίας της ζώνης σε σύγκριση με το ζητούμενο μήκος αλληλουχίας. Επιπλέον επιλέγονται έντονες ζώνες που υποδεικνύουν υψηλά επίπεδα έκφρασης του προϊόντος, συνεπώς και μεγαλύτερη ποσότητα προϊόντος και απορρίπτονται δείγματα με ελάχιστο θόρυβο (smear) τα οποία δεν αποδίδουν στην αλληλούχηση κατα Sanger. Η αποκοπή γίνεται με χρήση νυστεριού, σε τράπεζα υπεριώδους φωτός.

Έπειτα, κάθε κομμάτι ζώνης που κόβεται, ζυγίζεται σε αναλυτικό ζυγό και τοποθετείται σε φιαλίδιο Eppendorf 2 mL. Ακολουθεί καθαρισμός των δειγμάτων με χρήση του NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Inc). Συγκεκριμένα, σε κάθε φιαλίδιο γίνεται προσθήκη του αντιδραστηρίου NTI (200  $\mu$ L ανα 100 mg πηκτώματος) και επώαση στους 50°C, μέχρι να διαλυτοποιηθεί η αгарόζη. Το διάλυμα μεταφέρεται σε ειδικές στήλες καθαρισμού NucleoSpin Extract II οι οποίες τοποθετούνται σε νέα φιαλίδια τύπου Eppendorf 2,0 mL. Ακολουθεί φυγοκέντρηση σε 11.000  $\times$  g για 1 min, στην ψυχόμενη φυγόκεντρο Prism R (Labnet International Inc.), και απορρίψη του όγκου που διέτρεξε τη στήλη. Στη συνέχεια γίνεται προσθήκη 700  $\mu$ L του αντιδραστηρίου NT3 και φυγοκέντρηση σε 11.000  $\times$  g για 1 min, με απόρριψη του όγκου που διέτρεξε τη στήλη. Ακολουθεί επανάληψη της φυγοκέντρησης σε 11.000  $\times$  g για 1 min. Έπειτα οι στήλες θερμαίνονται για 2-5 min στους 70°C, προκειμένου να εξατμιστεί η αιθανόλη που περιέχει το αντιδραστήριο NT3 και τοποθετούνται σε ένα νέο φιαλίδιο τύπου Eppendorf 1,5 mL. Σε κάθε στήλη, προστίθεται το αντιδραστήριο Elution Buffer NE (5 mM Tris-HCl, pH = 8,5) ποσότητας 30  $\mu$ L και ακολουθεί επώαση σε θερμοκρασία δωματίου για 5 min. Μετά το τέλος της επώασης, επαναλαμβάνεται φυγοκέντρηση σε 11.000  $\times$  g για 1 min. Τέλος, τα διαλύματα που διατρέχουν τις στήλες, συλλέγονται με τη χρήση πιπέτας σε Eppendorfs των 0.2 mL και αποθηκεύονται στους -20°C.

Τα διαλύματα αυτά, αντιστοιχούν σε κάθε καθαρισμένο προϊόν από την nested 3' RACE που απομονώθηκε από το πήκτωμα αгарόζης. Ακολούθως, υποβάλλονται σε μια αντιδράση αλληλούχησης κατα Sanger, με χρήση του πρόσθιου εκκινητή που χρησιμοποιήθηκε για τον σχηματισμό του καθενός στην αντίδραση nested 3' RACE.

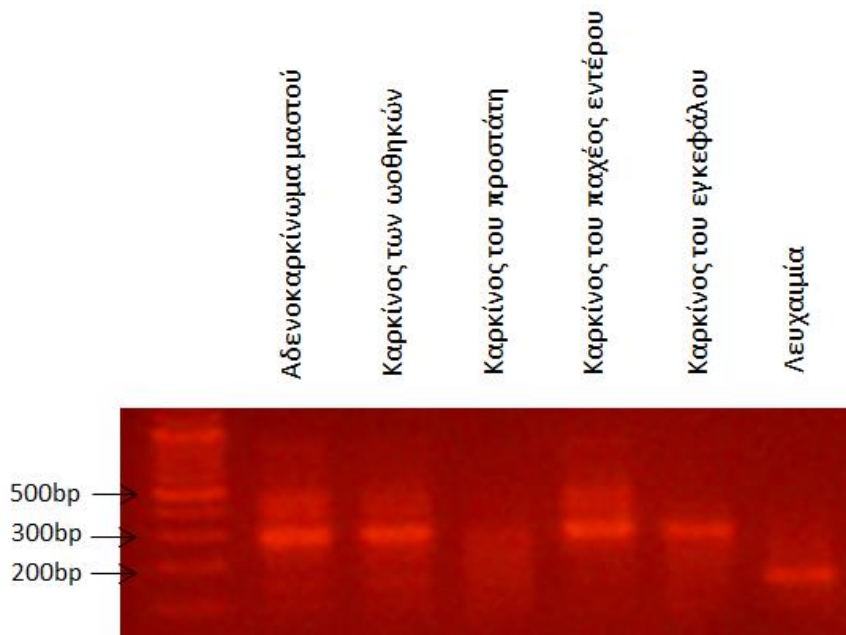
### 3. Αποτελέσματα

#### 3.1 Αποτελέσματα από τη βιοπληροφορική ανάλυση

Η βιοπληροφορική ανάλυση των δεδομένων από NGS, που πραγματοποιήθηκε για τους σκοπούς της περάτωσης της παρούσας διπλωματικής εργασίας, είχε στόχο τον εντοπισμό πιθανών θέσεων εναλλακτικών 3'-UTRs, στα ανθρώπινα γονίδια των *KLKs*. Με την ολοκλήρωσή της, επιλέχθηκαν συνολικά πέντε γονίδια *KLKs*, τα οποία έχουν μεγάλη πιθανότητα ύπαρξης εναλλακτικών 3'-UTRs, τα γονίδια *KLK7*, *KLK8*, *KLK9*, *KLK14* και *KLK15*. Για όλα τα γονίδια έγινε σχεδιασμός εκκινήτων με στόχο τον εντοπισμό εναλλακτικών 3'-UTRs, τα όποια εμφανίζουν μικρότερο ή μεγαλύτερο μήκος αλληλουχίας σε σύγκριση με το αναμενόμενο μήκος του τελευταίου εξωνίου στο κάθε μετάγραφο, ενώ για το γονίδιο *KLK7* έγινε σχεδιασμός ενός ακόμη εκκινήτη που στόχευε τη θέση ματίσματος δύο εξωνίων με στόχο τον εντοπισμό εναλλακτικού 3'-UTR σε νέα θέση στο μετάγραφο.

#### 3.2 Αποτελέσματα από αντιδράσεις με χρήση μιγμάτων cDNA

Οι πρώτες αντιδράσεις των μεθοδολογιών 3' RACE και nested 3' RACE έγιναν με την χρήση 6 pools cDNA και τα αποτελέσματα αξιολογήθηκαν από τις ηλεκτροφορήσεις των προϊόντων των αντιδράσεων nested 3' RACE για το προφίλ της έκφρασης εναλλακτικών 3'-UTR μεταγράφων στο γονίδιο *KLK7* όπως παρουσιάζεται στην **Εικόνα 3.1**.



**Εικόνα 3.1** Προφίλ έκφρασης των 3 νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK7* σε 6 pools cDNAs, που αντιστοιχούν σε 6 διαφορετικούς καρκινικούς ιστούς



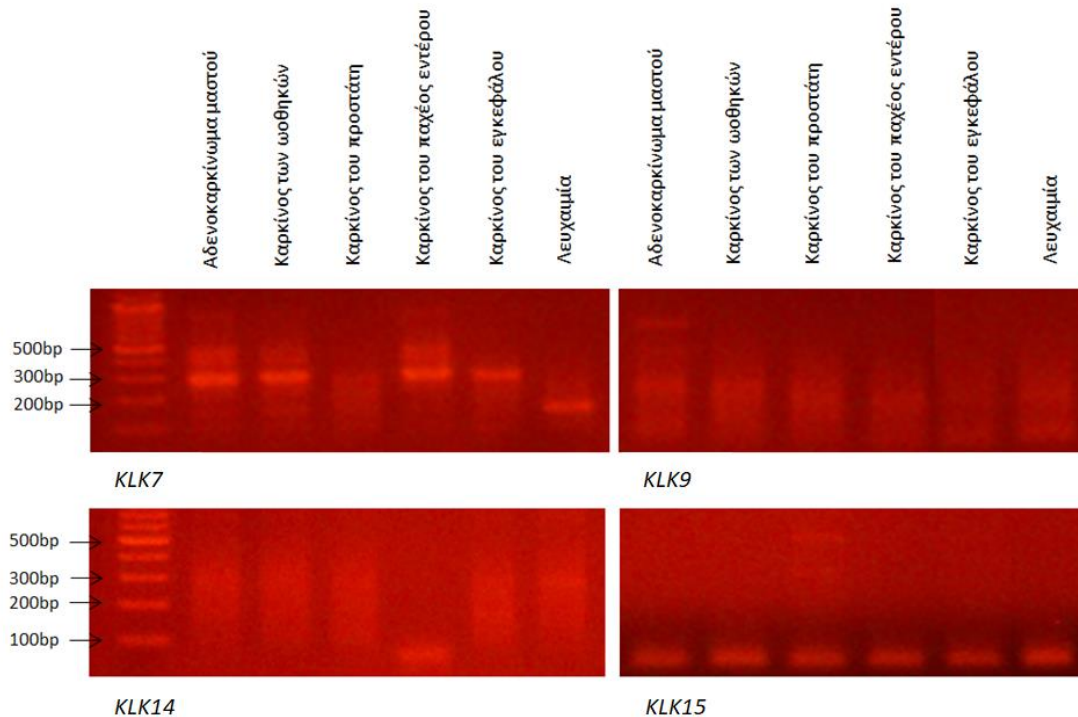
Μέχρι σήμερα, για το ανθρώπινο γονίδιο *KLK7* έχει ταυτοποιηθεί η ύπαρξη μιας 3'-UTR, μήκους 1057 nt, η οποία ακολουθεί το κωδικόνιο λήξης του τελευταίου εξωνίου στα μετάγραφα του γονιδίου *KLK7*. Με τη χρήση του ζεύγους εκκινητών *KLK7 Ex7nF / 3'RACE Inner Primer*, το αναμενόμενο μήκος αλληλουχίας που αντιστοιχεί στο 3'-UTR αναφοράς, είναι 1122 nt, άρα οποιαδήποτε ζώνη διαφορετικού μήκους, θεωρείται πιθανό προϊόν εναλλακτικού 3'-UTR μεταγράφου. Είναι σημαντικό να αναφερθεί ότι κάθε προϊόν από τις αντιδράσεις που αντιστοιχεί σε 3'-UTR, αναμένεται να έχει επιπλέον περίπου 50 nt καθώς συμπεριλαμβάνει την αλληλουχία της poly-A ουράς και την αλληλουχία του ανάστροφου εκκινητή.

Τα αποτελέσματα από την ηλεκτροφόρηση ήταν αρκετά υποσχόμενα, έχοντας τρεις έντονες ζώνες που αντιστοιχούν σε διαφορετικά μήκη αλληλουχίας και δεν εμφανίζουν θόρυβο. Ακολούθησε απομόνωση και καθαρισμός των ζωνών αυτών με σκοπό την αλληλούχησή τους κατα Sanger. Τα αποτελέσματα της αλληλούχησης κατα Sanger ταυτοποιούν την ύπαρξη 2 εναλλακτικών 3'-UTR μεταγράφων που αντιστοιχούν σε μήκη 204 nt και 35 nt, αρχίζοντας μετά το κωδικόνιο λήξης και μέχρι την έναρξη της poly-A ουράς. Η ζώνη που αφορά το μεγαλύτερο μήκος αλληλουχίας και αντιστοιχεί σε πιθανή εναλλακτική 3'-UTR μήκους 349 nt, πρόκειται για ψευδό αποτέλεσμα. Από τα αποτελέσματα της αλληλούχησης κατα Sanger, φαίνεται ξεκάθαρα ότι ο σχηματισμός του δεν οφείλεται στην ύπαρξη νέου σημείου πολυαδενυλίωσης. Ο ανάστροφος εκκινητής που χρησιμοποιήθηκε, στοχεύει τα τμήματα του oligo-dT adapter, άρα υβριδοποιείται με τμήμα της poly-A ουράς των mRNA μεταγράφων. Στην προκειμένη περίπτωση όμως, υβριδοποιήθηκε με τμήμα της αλληλουχίας του 3'-UTR το οποίο περιλαμβάνει μια σειρά επαναλαμβανόμενων αδενινών και το οποίο κατα τις αντιδράσεις αντίστροφης μεταγραφής (RT) θεωρήθηκε ως το 3' άκρο των mRNAs, με αποτέλεσμα την υβριδοποίηση του oligo-dT adapter, ο οποίος έπειτα κατα τις αντιδράσεις 3' RACE και nested 3' RACE, αναγνωρίστηκε και υβριδοποιήθηκε από τον ανάστροφο εκκινητή.

Με αντίστοιχες αντιδράσεις και τη χρήση 6 pools cDNA από 6 διαφορετικούς καρκινικούς ιστούς, εξετάστηκε το προφίλ της έκφρασης εναλλακτικών 3'-UTR μεταγράφων στα γονίδια *KLK9*, *KLK14* και *KLK15*, όπως παρουσιάζεται στην **Εικόνα 3.2**.

Σε σύγκριση με τα αποτελέσματα από την ηλεκτροφόρηση των προϊόντων για το γονίδιο *KLK7*, όπως φαίνεται και στην Εικόνα 3.2, οι ζώνες που σχηματίζονται για τα προϊόντα των τριών υπόλοιπων γονιδίων, δεν είναι ξεκάθαρες. Για τα γονίδια *KLK9* και

*KLK14* οι περισσότερες ζώνες εμφανίζουν αρκετό θόρυβο, το οποίο δεν επιτρέπει την αναγνώριση πιθανών 3'-UTRs και για το γονίδιο *KLK15* είναι εμφανής ο σχηματισμός διμερών εκκινητών, το οποίο υποδηλώνει χαμηλή απόδοση στις αντιδράσεις.



**Εικόνα 3.2** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs των γονιδίων *KLK7*, *KLK9*, *KLK14* και *KLK15* σε 6 pools cDNAs, που αντιστοιχούν σε 6 διαφορετικούς καρκινικούς ιστούς

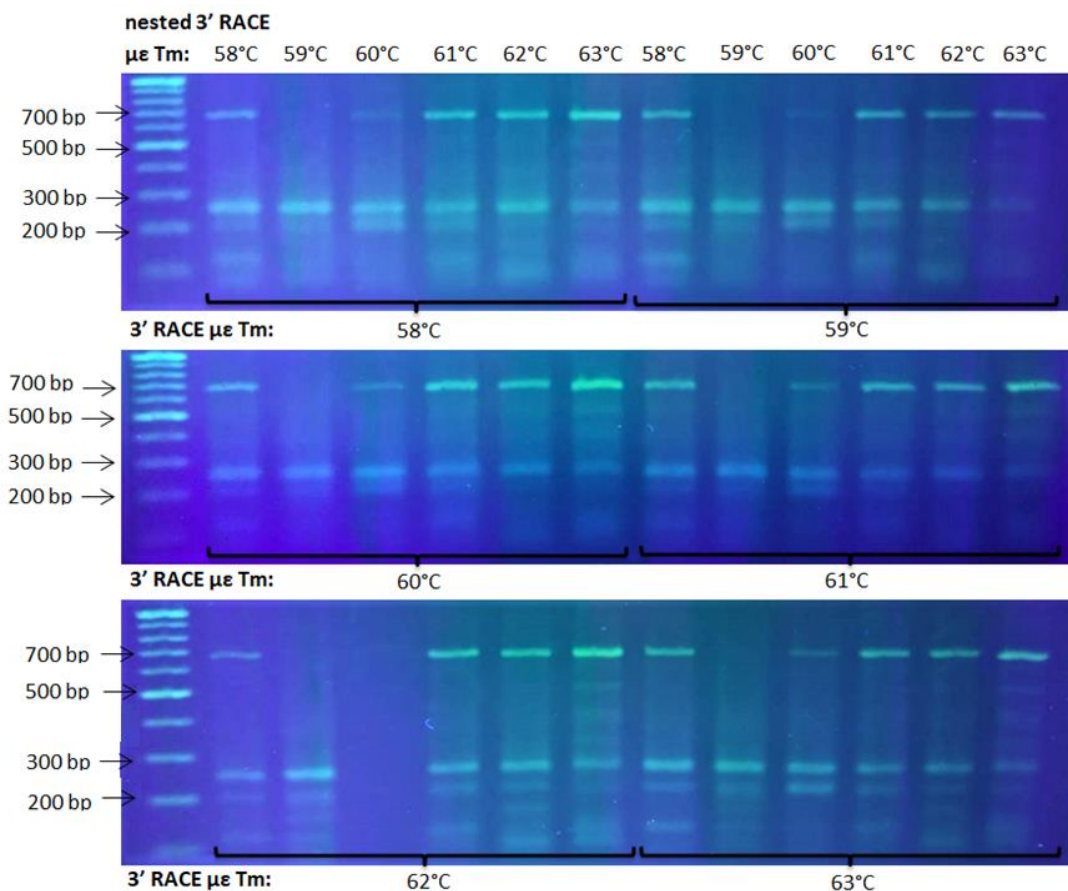
### 3.3 Αποτελέσματα από αντιδράσεις με χρήση cDNA από ανεξάρτητες κυτταρικές σειρές

Η χαμηλή απόδοση στις πρώτες αντιδράσεις, οδήγησε στη πραγματοποίηση αντιδράσεων κατά τις οποίες εφαρμόστηκε ένα διαφοροποιημένο πρωτόκολλο, με στόχο τη βελτιστοποίηση των συνθηκών στις μεθοδολογίες 3' RACE και nested 3' RACE .

Οι αντιδράσεις που ακολούθησαν, με τις μεθοδολογίες 3' RACE και nested 3' RACE, είχαν ως βασική διαφορά την χρήση 11 διαφορετικών κυτταρικών σειρών, ενδεικτικές από κάθε καρκινικό τύπο/ιστό προέλευσης και την εφαρμογή ενός μοντέλου διαβάθμισης της θερμοκρασίας υβριδοποίησης (annealing gradient). Οι αντιδράσεις αφορούσαν δύο γονίδια, τα *KLK9* και *KLK14*, ενώ τα αποτελέσματα αξιολογήθηκαν από τις ηλεκτροφορήσεις των προϊόντων των αντιδράσεων nested 3' RACE για το προφίλ της έκφρασης εναλλακτικών 3'-UTR μεταγράφων των γονιδίων αυτών.

Μέχρι σήμερα, για το ανθρώπινο γονίδιο *KLK9* έχει ταυτοποιηθεί η ύπαρξη ενός μόνο μεταγράφου το οποίο διαθέτει μια 3'-UTR, μήκους 598 nt, σε συνέχεια μετά το κωδικόνιο λήξης του εξωνίου 5. Με τη χρήση του ζεύγους εκκινήτων *KLK9* Ex5nF / 3'RACE Inner Primer, το αναμενόμενο μήκος αλληλουχίας που αντιστοιχεί στο 3'-UTR αναφοράς, είναι 626 nt, άρα οποιαδήποτε ζώνη διαφορετικού μήκους, θεωρείται πιθανό προϊόν εναλλακτικού 3'-UTR μεταγράφου. Αντίστοιχα, για το ανθρώπινο γονίδιο *KLK14*, έχουν ταυτοποιηθεί 3 μετάγραφα τα οποία διαθέτουν 3 διαφορετικά 3'-UTRs. Για τις αντιδράσεις που πραγματοποιήθηκαν έγινε χρήση του ζεύγους εκκινήτων *KLK14* Ex7nF / 3'RACE Inner Primer, για το οποίο το αναμενόμενο μήκος αλληλουχίας που αντιστοιχεί στο 3'-UTR αναφοράς, είναι 84 nt, άρα οποιαδήποτε ζώνη διαφορετικού μήκους, θεωρείται πιθανό προϊόν εναλλακτικού 3'-UTR μεταγράφου.

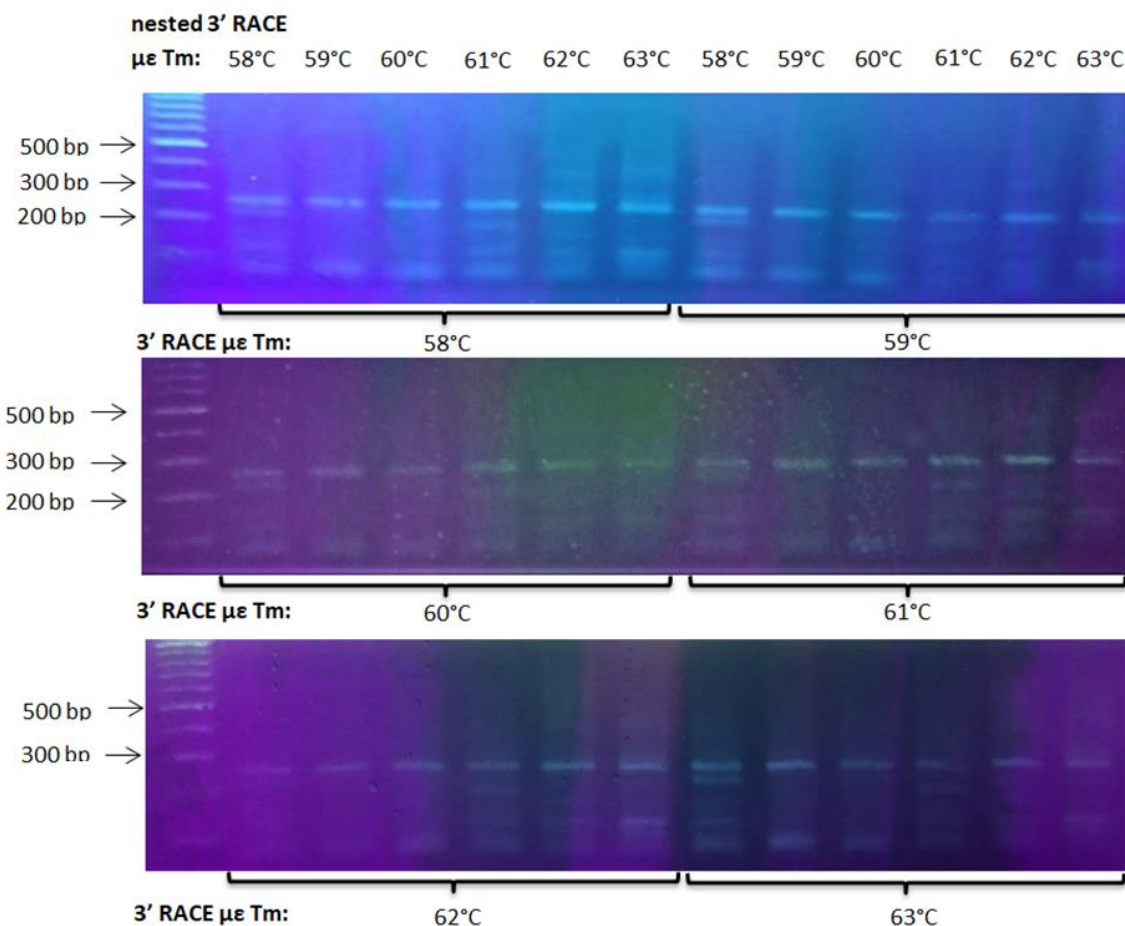
Για την κυτταρική σειρά SK-BR-3, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.3**.



**Εικόνα 3.3** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά SK-BR-3.

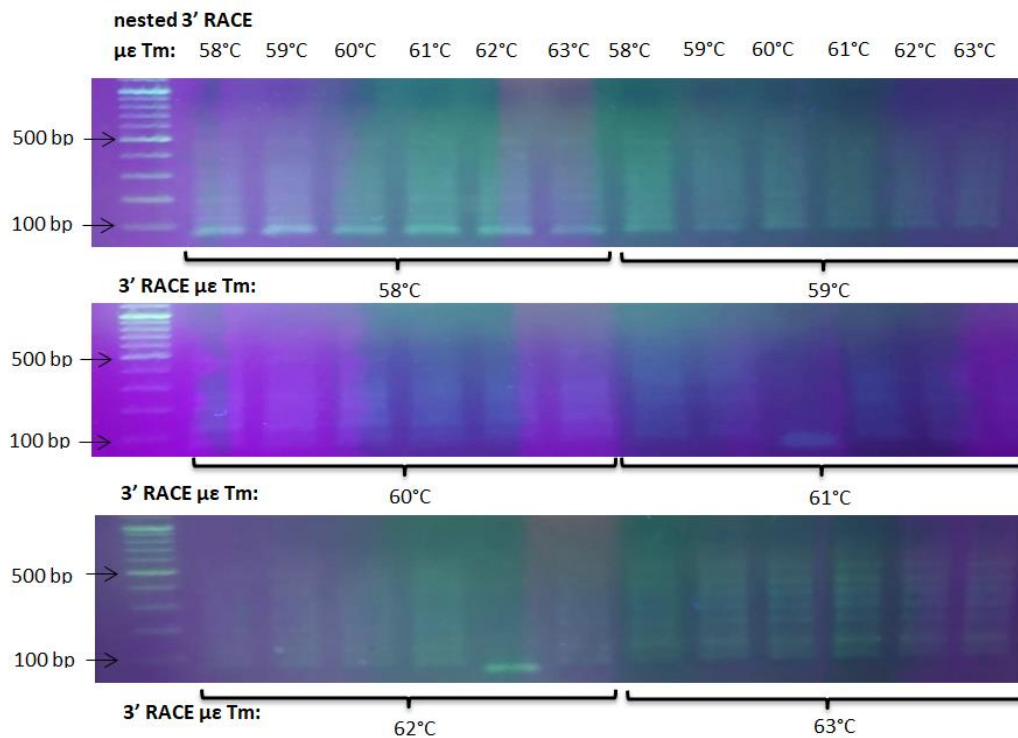
Για την κυτταρική σειρά SK-BR-3, οι ηλεκτροφορήσεις παρουσιάζουν εμφανώς μια ζώνη μήκους σχεδόν 700 bp που αντιστοιχεί στην 3'-UTR αναφοράς και μια ακόμη ζώνη περίπου 250 bp, η οποία εμφανίζεται σε όλες τις θερμοκρασίες υβριδοποίησης αλλά έχει μεγαλύτερη ευκρίνεια σε θερμοκρασία 63 °C για την αντίδραση 3' RACE. Αυτή η ζώνη αντιπροσωπεύει ένα πιθανόν εναλλακτικό 3'- UTR για το γονίδιο *KLK9*.

Για την κυτταρική σειρά SK-OV-3, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.4**. Για την κυτταρική σειρά SK-OV-3, οι ηλεκτροφορήσεις παρουσιάζουν λιγότερο ευκρινή ζώνη μήκους περίπου 250 bp, η οποία ταυτίζεται με τη ζώνη στην κυτταρική σειρά SK-BR-3. Αυτή η ζώνη αντιπροσωπεύει ένα πιθανόν εναλλακτικό 3'- UTR για το γονίδιο *KLK9*. . Γενικά η εμφάνιση ζωνών σε αντίστοιχα μήκη από διαφορετικές κυτταρικές σειρές είναι μια επιπλέον υπόδειξη της ύπαρξης των εναλλακτικών 3'-UTRs που αντιπροσωπεύουν.



**Εικόνα 3.4** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά SK-OV-3.

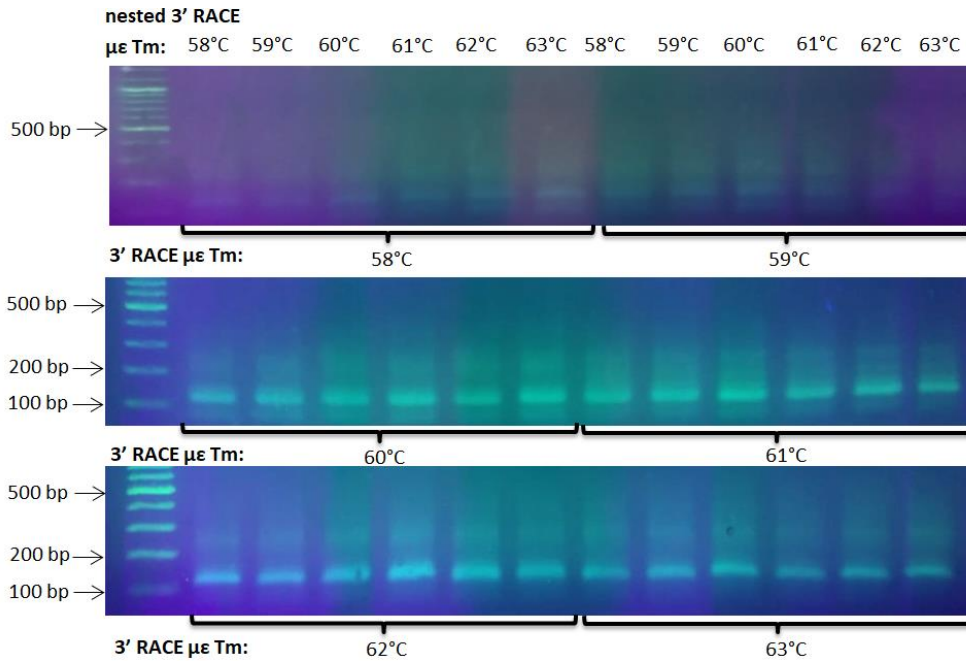
Για την κυτταρική σειρά ES-2, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλετροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.5**. Όπως φαίνεται και στην εικόνα, δεν υπάρχουν ζώνες που να αντιστοιχούν σε προϊόντα των αντιδράσεων αλλά θόρυβος. Αυτό μπορεί να οφείλεται στην μικρή έκφραση ή στην απουσία έκφρασης του γονιδίου που εξετάζεται στην συγκεκριμένη κυτταρική σειρά.



**Εικόνα 3.5** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά ES-2.

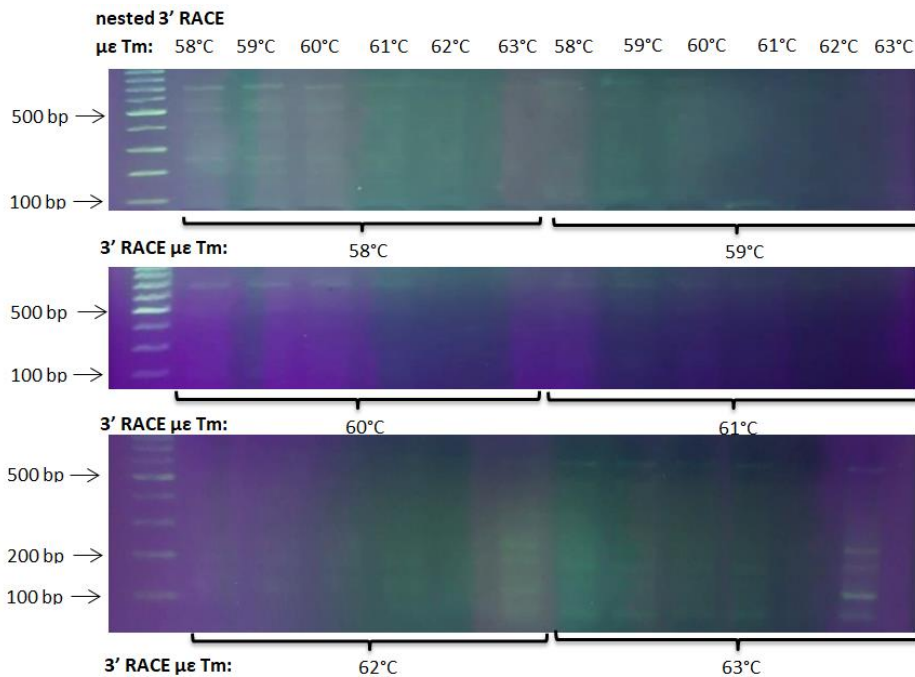
Για την κυτταρική σειρά MDAH-2774, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλετροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.6**.

Στη συγκεκριμένη κυτταρική σειρά, τα αποτελέσματα της ηλετροφόρησης για τις θερμοκρασίες υβριδοποίησης 58 °C και 59 °C ως προς τις αντιδράσεις 3' RACE, παρουσίασαν μόνο θόρυβο, ενώ για τις υπόλοιπες θερμοκρασίες, εμφανίζεται μια νέα ζώνη που αντιστοιχεί σε περίπου 120 bp και αντιπροσωπεύει ένα υποψήφιο πιθανό εναλλακτικό 3'-UTR.



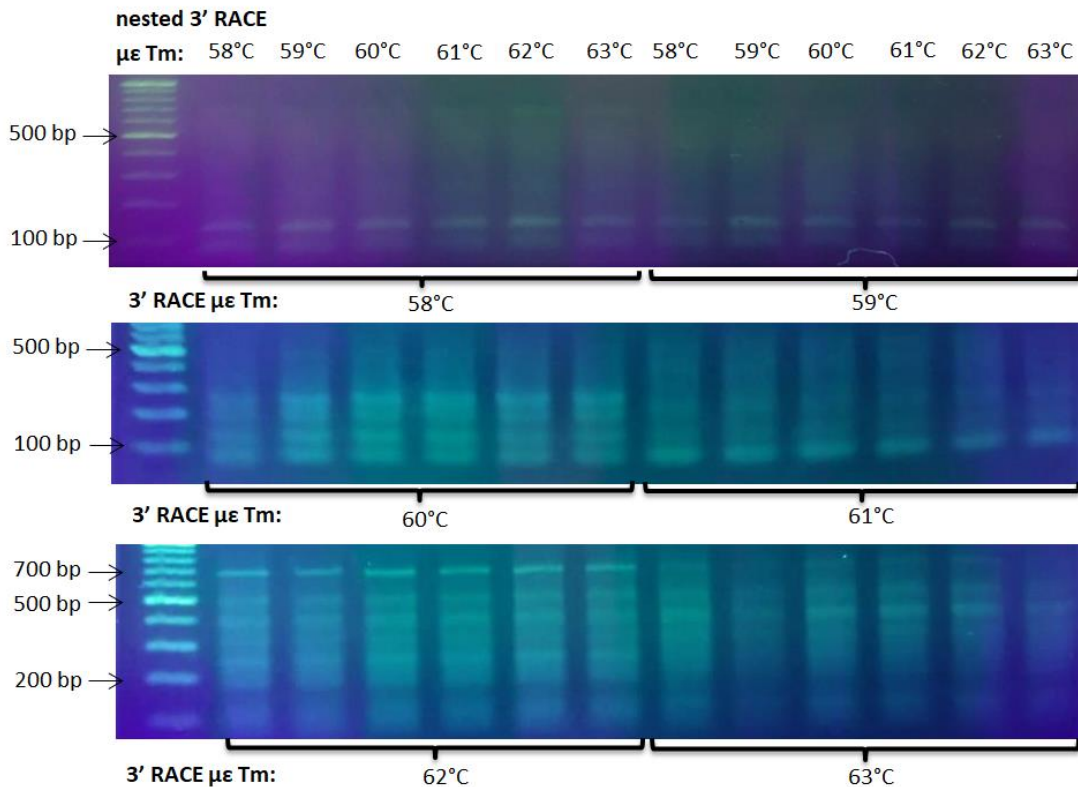
**Εικόνα 3.6** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά MDAH-2774.

Για την κυτταρική σειρά Ishikawa, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.7**. Τα αποτελέσματα των ηλεκτροφορήσεων για την κυτταρική σειρά Ishikawa, είναι αντίστοιχα με αυτά της κυτταρικής σειράς ES-2, παρουσιάζοντας αρκετό θόρυβο και πολύ θολές ζώνες οι οποίες είναι ακατάλληλες για καθαρισμό και περαιτέρω ανάλυσή τους



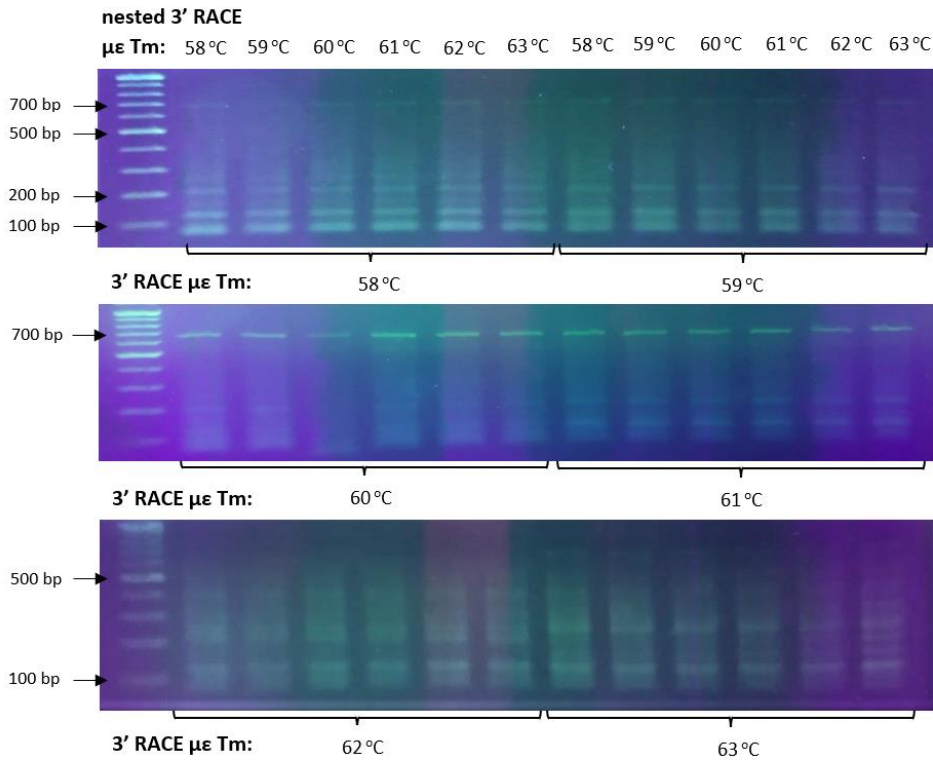
**Εικόνα 3.7** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά Ishikawa.

Για την κυτταρική σειρά SK-UT-1B, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.8**. Σε αυτή την κυτταρική σειρά, είναι εμφανής η ζώνη που αντιστοιχεί στην 3'-UTR αναφοράς, για συγκεκριμένες θερμοκρασίες υβριδοποίησης. Παρολ'αυτά εμφανίζεται αρκετά έντονος θόρυβος ιδιαίτερα σε θερμοκρασίες 60 – 63 °C ως προς τις αντιδράσεις 3'RACE, καθιστώντας μη ικανή την αξιολόγηση των αποτελεσμάτων.



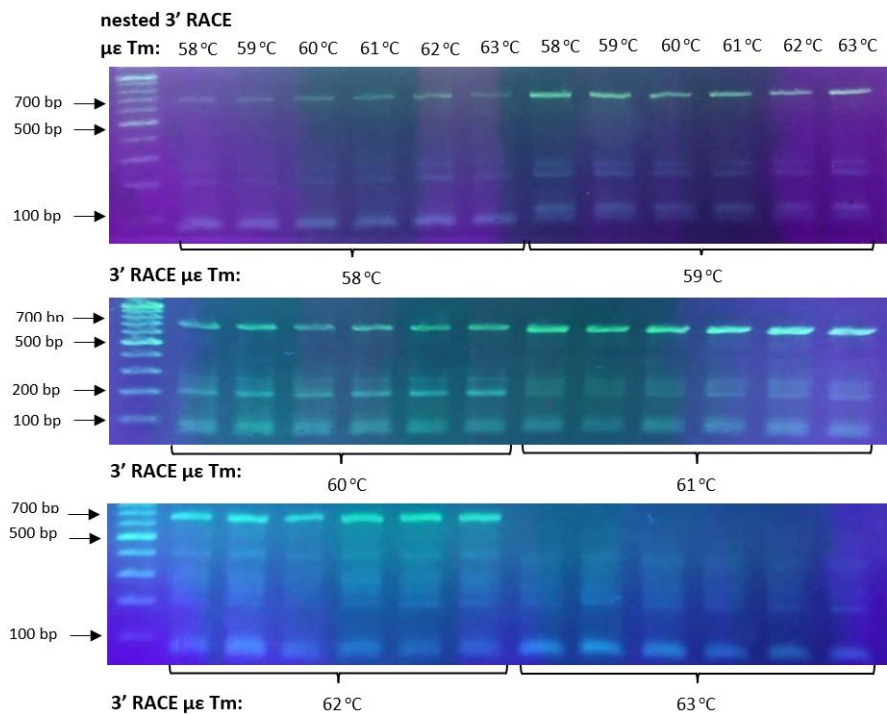
**Εικόνα 3.8** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά SK-UT-1B.

Για την κυτταρική σειρά HCT-116, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.9**. Απο τις ηλεκτροφορήσεις εμφανίζεται η ζώνη της 3'-UTR αναφοράς, αλλά και δύο ακόμη ζώνες σε θερμοκρασία υβριδοποίησης 58 °C ως προς τις αντιδράσεις 3'RACE, οι οποίες αντιστοιχούν σε 90 bp και 120 bp. Η δεύτερη ζώνη των 120 bp ταυτίζεται με την ζώνη που υπάρχει στις ηλεκτροφορήσεις της κυτταρικής σειράς MDAH-2774.



**Εικόνα 3.9** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά HCT-116.

Για την κυτταρική σειρά HT-29, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.10**.

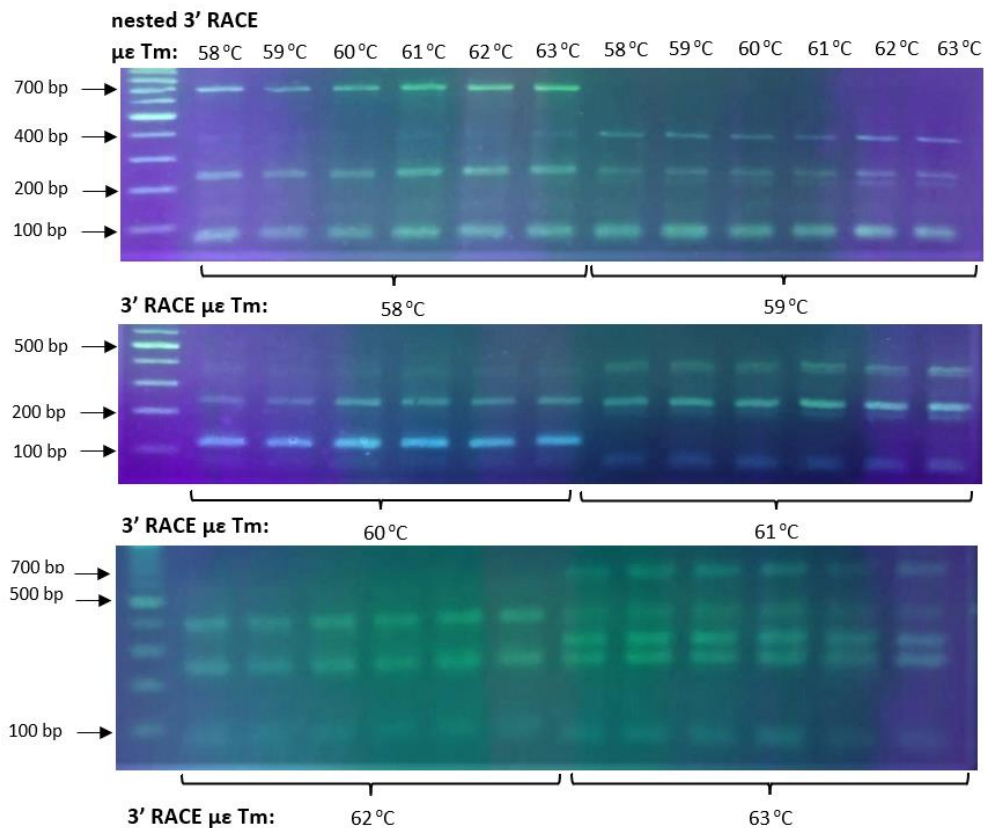


**Εικόνα 3.10** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά HCT-29.



Όπως παρουσιάζει και η εικόνα, στην κυτταρική σειρά HT-29, είναι ξεκάθαρη η ύπαρξη της 3'-UTR αναφοράς, η οποία αντιπροσωπεύεται από έντονη ζώνη. Ταυτόχρονα εμφανίζονται δύο ακόμη ζώνες οι οποίες αντιστοιχούν σε περίπου 200 bp και 90 bp. Η δεύτερη ζώνη ταυτίζεται με τα αποτελέσματα από την κυτταρική σειρά HCT-116.

Για την κυτταρική σειρά DLD-1, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK9* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.11**.

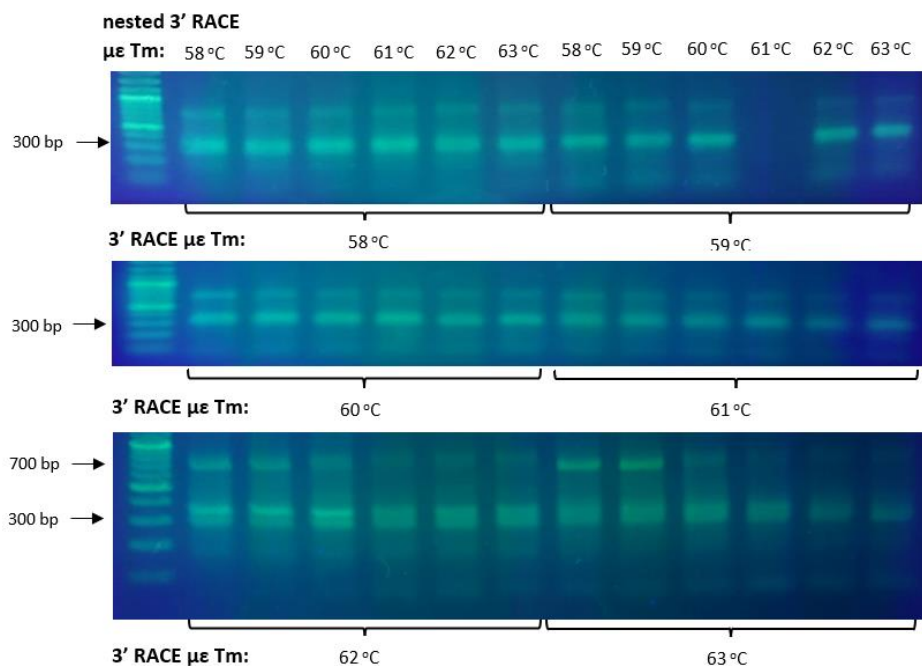


**Εικόνα 3.11** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK9*, για την κυτταρική σειρά DLD-1.

Στη συγκεκριμένη κυτταρική σειρά, εμφανίζεται έντονη καθαρή ζώνη που αντιστοιχεί στη 3'-UTR αναφοράς. Επιπλέον εμφανίζονται τέσσερις ακόμη ζώνες με μήκη περίπου 400 bp, 250 bp, 120 bp και 90 bp. Οι ζώνες αυτές εμφανίζουν πιο έντονη και καθαρή εικόνα μόνο σε συγκεκριμένους συνδυασμούς θερμοκρασιών, οι οποίοι επιλέχθηκαν για την απομόνωση και τον καθαρισμό των συγκεκριμένων ζωνών. Η ζώνη μήκους 250bp ταυτίζεται με τις ζώνες στις κυτταρικές σειρές SK-BR-3 και SK-OV-3, η ζώνη μήκους 120 bp ταυτίζεται με τις ζώνες στις κυτταρικές σειρές MDAH-2774 και HCT-116, ενώ η ζώνη

μήκους 90 bp ταυτίζεται με τις ζώνες στις κυτταρικές σειρές HCT-116 και HCT-29. Οι ζώνες αυτές αντιπροσωπεύουν πιθανά εναλλακτικά 3'-UTRs του γονιδίου *KLK9*.

Για την κυτταρική σειρά LNCaP, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK14* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.12**. Οι ηλεκτροφορήσεις για την κυτταρική σειρά LNCaP, παρουσιάζουν την ύπαρξη τριών ζωνών που αντιστοιχούν σε μήκη περίπου 300bp, 320 bp και 700 bp. Οι τρεις αυτές ζώνες αντιπροσωπεύουν πιθανά εναλλακτικά 3'-UTRs του γονιδίου *KLK14*.

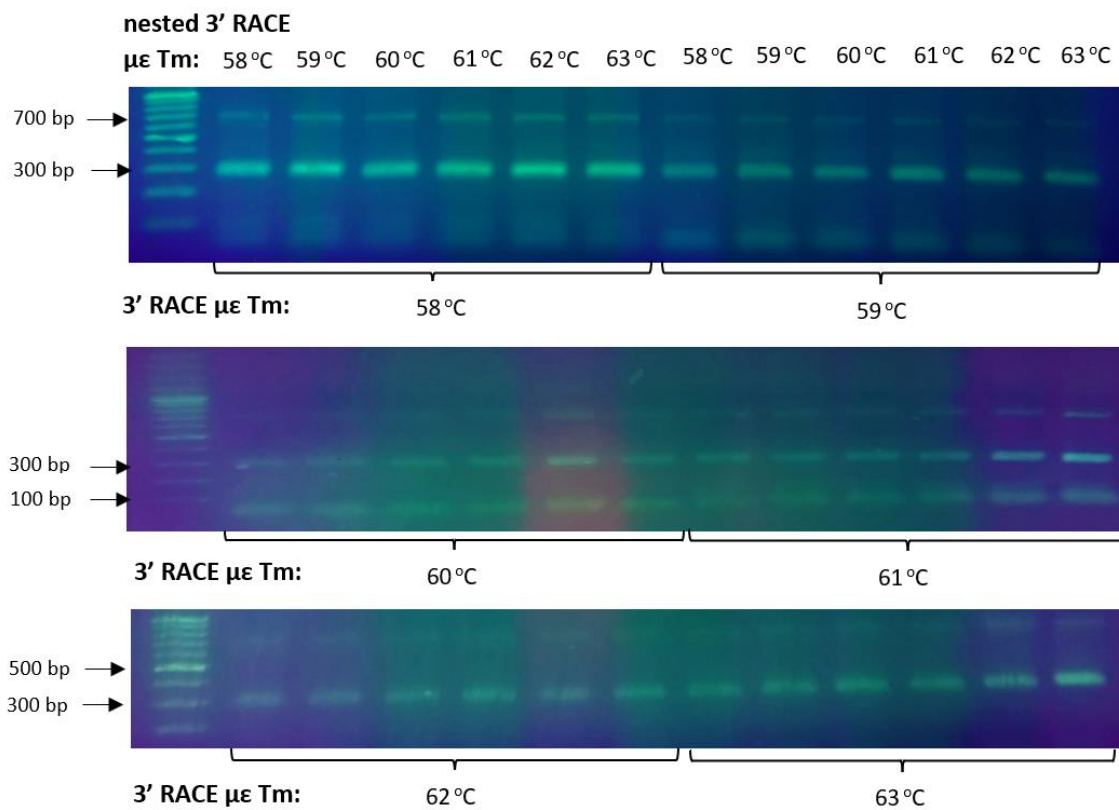


**Εικόνα 3.12** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK14*, για την κυτταρική σειρά LNCaP.

Για την κυτταρική σειρά MCF-7, πραγματοποιήθηκαν αντιδράσεις που αφορούσαν το γονίδιο *KLK14* και τα προϊόντα των αντιδράσεων nested 3' RACE, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης, όπως παρουσιάζεται στην **Εικόνα 3.13**. Στη συγκεκριμένη κυτταρική σειρά, εμφανίζονται δύο ζώνες που αντιστοιχούν σε μήκη περίπου 300 bp και 700 bp. Η δεύτερη ζώνη αντιστοιχεί με τη ζώνη των 700 bp που εμφανίζεται και στην κυτταρική σειρά LNCaP.

Με βάση τα αποτελέσματα των ηλεκτροφορήσεων, όπως παρουσιάζονται πιο πάνω, έγινε η επιλογή από πέντε κυτταρικές σειρές, MDAH-2774, HT-29, DLD-1, LNCaP και MCF-7 για κόψιμο και απομόνωση των ζωνών που είχαν την μεγαλύτερη ευκρίνεια και

καθαρότητα. Ακολούθησε η αλληλούχηση κατα Sanger για την ταυτοποίηση των αλληλουχιών που απομονώθηκαν. Απο τα αποτελέσματα της αλληλούχησης, ταυτοποιήθηκε ένα τμήμα ενός εναλλακτικού 3'-UTR, στο ανθρώπινο γονίδιο *KLK14*, το οποίο εκτείνεται έως 602 nt μετά το κωδικόνιο λήξης του τελευταίου εξωνίου και σε αυτό παρεμβάλλεται ένα νέο ιντρόνιο, μήκους 73 nt.



**Εικόνα 3.13** Προφίλ έκφρασης των νέων εναλλακτικών 3'-UTRs του γονιδίου *KLK14*, για την κυτταρική σειρά MCF-7.

Παρόλο που δεν κατάφερε να γίνει ταυτοποίηση των υπολοίπων πιθανών εναλλακτικών 3'-UTRs, τα αποτελέσματα από τις ηλεκτροφορήσεις υποδεικνύουν μεγάλη πιθανότητα ύπαρξής τους, κυρίως για τις ταυτόσημες ζώνες που παρουσιάζονται σε διαφορετικές κυτταρικές σειρές και με διαφορετικές θερμοκρασίες υβριδοποίησης.

### 3.4 Αποτελέσματα από εύρεση miRNA μορίων με χρήση αλγορίθμου miRDB

Μετά την επιτυχή ταυτοποίηση 2 νέων εναλλακτικών 3'-UTRs σε μετάγραφα του γονιδίου *KLK7* και ενός τμήματος νέου εναλλακτικού 3'-UTR σε μετάγραφα του γονιδίου *KLK14*, ακολούθησε η εύρεση miRNA μορίων, τα οποία πιθανώς στοχεύουν τη γνωστή 3'-UTR για τα δύο πιο πάνω γονίδια και τις νέες εναλλακτικές του γονιδίου *KLK7*. Η διαδικασία αυτή πραγματοποιήθηκε με τη χρήση του **αλγορίθμου miRDB**. Ως εύρος αποτελεσμάτων ορίστηκε η επιλογή των μορίων τα οποία εμφανίζουν βαθμό στόχευσης (target score) έως 70.

Συγκεκριμένα για το γονίδιο *KLK7*, όπως παρουσιάζεται στην **Εικόνα 3.14**, δίνονται ως αποτέλεσμα 19 μόρια miRNA τα οποία πιθανά στοχεύουν την γνωστή 3'-UTR του γονιδίου, με βάση την αλληλουχία της. Για την νέα εναλλακτική 3'-UTR του γονιδίου *KLK7*, μήκους 204 nt, εμφανίζονται ως αποτέλεσμα 7 μόρια miRNA τα οποία περιλαμβάνονται στα 19 μόρια τα οποία αντιστοιχούν στην 3'-UTR αναφοράς, όπως παρουσιάζει η **Εικόνα 3.15**, ενώ η νέα εναλλακτική 3'-UTR του γονιδίου, μήκους 35 nt, δεν προβλέπεται να στοχεύεται από κανένα miRNA μόριο.

Target Detail	Target Rank	Target Score	miRNA Name	Gene Symbol
Details	1	97	<a href="#">hsa-miR-153-5p</a>	submission
Details	2	90	<a href="#">hsa-miR-4679</a>	submission
Details	3	86	<a href="#">hsa-miR-3667-3p</a>	submission
Details	4	86	<a href="#">hsa-miR-561-5p</a>	submission
Details	5	85	<a href="#">hsa-miR-889-3p</a>	submission
Details	6	83	<a href="#">hsa-miR-3915</a>	submission
Details	7	80	<a href="#">hsa-miR-2681-3p</a>	submission
Details	8	80	<a href="#">hsa-miR-6888-3p</a>	submission
Details	9	78	<a href="#">hsa-miR-509-3p</a>	submission
Details	10	77	<a href="#">hsa-miR-335-3p</a>	submission
Details	11	75	<a href="#">hsa-miR-628-3p</a>	submission
Details	12	75	<a href="#">hsa-miR-6830-5p</a>	submission
Details	13	74	<a href="#">hsa-miR-1290</a>	submission
Details	14	73	<a href="#">hsa-miR-1914-5p</a>	submission
Details	15	72	<a href="#">hsa-miR-377-5p</a>	submission
Details	16	72	<a href="#">hsa-miR-4748</a>	submission
Details	17	72	<a href="#">hsa-miR-4464</a>	submission
Details	18	71	<a href="#">hsa-miR-6086</a>	submission
Details	19	71	<a href="#">hsa-miR-4743-3p</a>	submission

#### MicroRNA and Target Gene Description:

miRNA Name	<a href="#">hsa-miR-153-5p</a>	miRNA Sequence	UCAUUUUUGUGAUGUUGCAGCU
Target Score	97	Seed Location	800, 911
Target Length	1057		

#### Custom Target Sequence

```

1 cgccacactg agttaattaa ctgtgtcctt ccaacagaaa atgcacagga gtgaggacgc
61 cgatgacctg tgaagtcaaa ttgacttta cctttcctca aagatatatt taaaccaacc
121 tcatgccctg ttgataaacc aatcaaatg gtaaagacct aaacacaaaa caaataaaga
181 aacacaaaaa cctcagtctt ggagaagagt cagtggagac agcactctca aacactggaa
241 ctggacgttc gtacagcttt tacggaagac acttggtcaa cgtacaccga gacccttatt
301 caccaccttt gaccagtaa ctctaactt aggaagaacc tactgaaaca aaaaaaatcc
361 aaaatgtaga acaagacttg aatttaccat gatattatct atcacagaaa tgaagtgaaa
421 ccatcaaaac tgttccaaaa gtaccagatg gcttaataa tagtctggct tggcacaacg
481 atgtttttt tctttgagac agagtctctg ttgcttgggc tgcaatgcag tgaatgcaatc
541 ttggctcact gcaacctccg cctcctgggt tcaagtgatt ctctgtcttc agcctcccaa
601 gtacctggga ctacagggtg gcaccaccac accaggctaa tttttgtgt atttttacta
661 gagacagggt ttaccatgtt tggccagcat ggtcttgaac gcttgacctc agatgatcca
721 cccaccttgg cctcccaaaag tgcctgggatt acaggcatga gccaccacgg ccagcccaca
781 atgatattac aaacctatta aaaatgtac ttagacagaa ttgtcagtat tattcaagaa
841 catttaggct ataggatgtt aaatgacaaa aggaaggaca aaaatatata tgmtatgtgac
901 cctaccocata aaaatgaaa tattcacaga atcagatctg aaaaacatg tcccagactg
961 catactgggg tcgtcatgag gtgtctcctt ccttctgtgt acttttctt gaatgtgac
1021 ttttataaca tgaaaaataa aggtggggaa aaaagtc

```

**Εικόνα 3.14** Αποτελέσματα αλγορίθμου miRDB για τους πιθανούς στόχους από miRNAs στη 3'-UTR αναφοράς του γονιδίου *KLK7*.

Η μεγάλη διαφορά του αριθμού των μορίων miRNA τα οποία μπορούν πιθανά να στοχεύσουν τις νέες 3'-UTRs του γονιδίου *KLK7*, οφείλεται στη μείωση του μήκους της αλληλουχίας της νέας 3'-UTR, η οποία έχει ως αποτέλεσμα την ύπαρξη λιγότερων θέσεων πιθανής στόχευσης της 3'-UTR από miRNA μόρια.

**MicroRNA and Target Gene Description:**

Target Detail	Target Rank	Target Score	miRNA Name	Gene Symbol
Details	1	87	<a href="#">hsa-miR-3915</a>	submission
Details	2	83	<a href="#">hsa-miR-509-3p</a>	submission
Details	3	78	<a href="#">hsa-miR-1914-5p</a>	submission
Details	4	75	<a href="#">hsa-miR-6730-5p</a>	submission
Details	5	75	<a href="#">hsa-miR-6079</a>	submission
Details	6	74	<a href="#">hsa-miR-6086</a>	submission
Details	7	74	<a href="#">hsa-miR-377-5p</a>	submission

miRNA Name [hsa-miR-3915](#) miRNA Sequence UUGAGGAAAAGGUCUUAUU  
 Target Score 87 Seed Location 94  
 Target Length 204

**Custom Target Sequence**

```

1 cgccacactg agttaattaa ctgtgtgctt ccaacagaaa atgcacagga gtgaggacgc
61 cgatgaccta tgaagtcaaa tttgacttta cctttcctca aagatatatt taaaccaacc
121 tcatgccctg ttgataaacc aatcaaattg gtaaagacct aaaacaaaa caaataaaga
181 aacacaaaac cctcagtct ggag
  
```

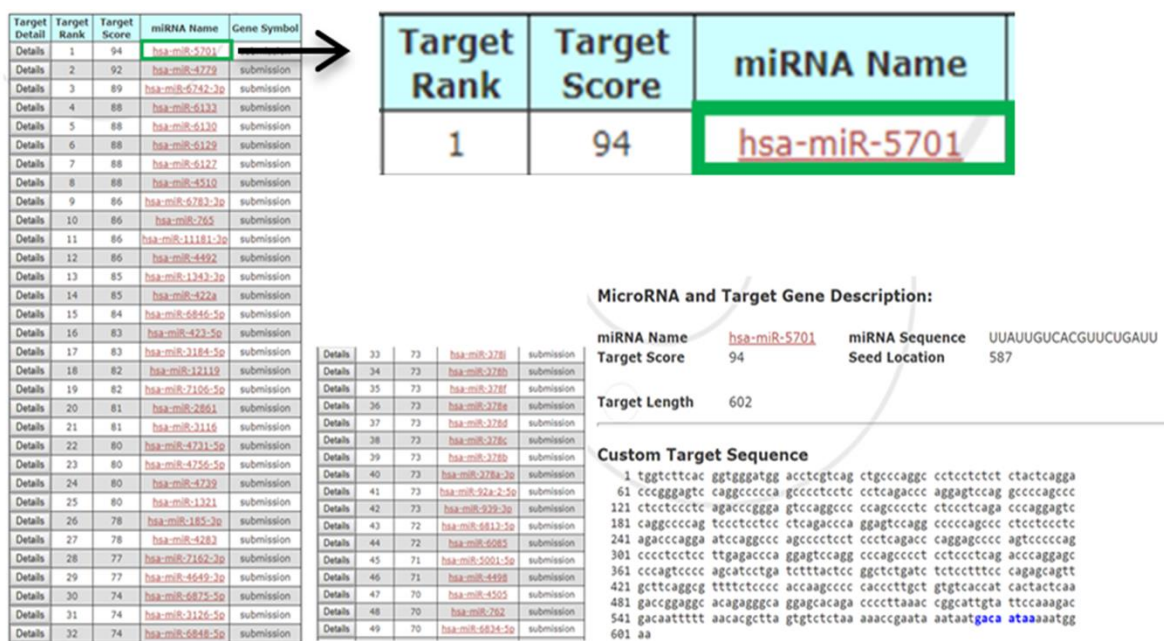
**Εικόνα 3.15** Αποτελέσματα αλγορίθμου miRDB για τους πιθανούς στόχους από miRNAs στη νέα 3'-UTR αναφοράς, μήκους 204 nt, του γονιδίου *KLK7*.

Target Detail	Target Rank	Target Score	miRNA Name	Gene Symbol
Details	1	97	<a href="#">hsa-miR-6742-3p</a>	submission
Details	2	95	<a href="#">hsa-miR-4779</a>	submission
Details	3	94	<a href="#">hsa-miR-6133</a>	submission
Details	4	94	<a href="#">hsa-miR-6130</a>	submission
Details	5	94	<a href="#">hsa-miR-6129</a>	submission
Details	6	94	<a href="#">hsa-miR-6127</a>	submission
Details	7	94	<a href="#">hsa-miR-4510</a>	submission
Details	8	94	<a href="#">hsa-miR-6783-3p</a>	submission
Details	9	93	<a href="#">hsa-miR-1343-3p</a>	submission
Details	10	93	<a href="#">hsa-miR-422a</a>	submission
Details	11	93	<a href="#">hsa-miR-765</a>	submission
Details	12	93	<a href="#">hsa-miR-11181-3p</a>	submission
Details	13	92	<a href="#">hsa-miR-4492</a>	submission
Details	14	91	<a href="#">hsa-miR-423-5p</a>	submission
Details	15	91	<a href="#">hsa-miR-3184-5p</a>	submission
Details	16	90	<a href="#">hsa-miR-12119</a>	submission
Details	17	90	<a href="#">hsa-miR-7106-5p</a>	submission
Details	18	89	<a href="#">hsa-miR-3116</a>	submission
Details	19	88	<a href="#">hsa-miR-4756-5p</a>	submission
Details	20	88	<a href="#">hsa-miR-4739</a>	submission
Details	21	88	<a href="#">hsa-miR-1321</a>	submission
Details	22	88	<a href="#">hsa-miR-2861</a>	submission
Details	23	88	<a href="#">hsa-miR-4731-5p</a>	submission
Details	24	87	<a href="#">hsa-miR-185-3p</a>	submission
Details	25	86	<a href="#">hsa-miR-7162-3p</a>	submission
Details	26	86	<a href="#">hsa-miR-4649-3p</a>	submission
Details	27	85	<a href="#">hsa-miR-4283</a>	submission
Details	28	85	<a href="#">hsa-miR-939-3p</a>	submission
Details	29	84	<a href="#">hsa-miR-378j</a>	submission
Details	30	84	<a href="#">hsa-miR-378h</a>	submission
Details	31	84	<a href="#">hsa-miR-378f</a>	submission
Details	32	84	<a href="#">hsa-miR-378e</a>	submission
Details	33	84	<a href="#">hsa-miR-378d</a>	submission
Details	34	84	<a href="#">hsa-miR-378c</a>	submission
Details	35	84	<a href="#">hsa-miR-378b</a>	submission
Details	36	84	<a href="#">hsa-miR-378a-3p</a>	submission
Details	37	81	<a href="#">hsa-miR-5001-5p</a>	submission
Details	38	81	<a href="#">hsa-miR-4498</a>	submission
Details	39	81	<a href="#">hsa-miR-6813-5p</a>	submission
Details	40	81	<a href="#">hsa-miR-6085</a>	submission
Details	41	80	<a href="#">hsa-miR-6875-5p</a>	submission
Details	42	80	<a href="#">hsa-miR-3126-5p</a>	submission
Details	43	80	<a href="#">hsa-miR-6852-5p</a>	submission
Details	44	80	<a href="#">hsa-miR-6834-5p</a>	submission
Details	45	80	<a href="#">hsa-miR-762</a>	submission
Details	46	77	<a href="#">hsa-miR-4505</a>	submission
Details	47	76	<a href="#">hsa-miR-5287</a>	submission
Details	48	75	<a href="#">hsa-miR-3690</a>	submission

**Εικόνα 3.16** Αποτελέσματα αλγορίθμου miRDB για τους πιθανούς στόχους από miRNAs στη 3'-UTR αναφοράς του γονιδίου *KLK14*.

Για το γονίδιο *KLK14*, όπως παρουσιάζεται στην **Εικόνα 3.16**, δίνονται ως αποτέλεσμα 148 μόρια miRNA τα οποία πιθανά στοχεύουν την γνωστή 3'-UTR του γονιδίου, με βαθμό στόχευσης έως 70. Για την νέα εναλλακτική 3'-UTR του γονιδίου *KLK7*, μήκους 602 nt, εμφανίζονται ως αποτέλεσμα 49 μόρια miRNA, όπως παρουσιάζει η **Εικόνα 3.17**.

Αξιοσημείωτο είναι το γεγονός ότι το επιπλέον μόριο το οποίο παρουσιάζεται να στοχεύει πιθανά την νέα 3'-UTR του *KLK14*, χωρίς να μπορεί να στοχεύσει τη γνωστή 3'-UTR, εμφανίζει το μεγαλύτερο βαθμό στόχευσης, από όλα τα πιθανά μόρια miRNA τα οποία στοχεύουν την νέα 3'-UTR. Το γεγονός αυτό οφείλεται στο σημείο στόχευσης του συγκεκριμένου μορίου miRNA, το οποίο περιλαμβάνει ένα τμήμα της αλληλουχίας της νέας 3'-UTR η οποία δεν περιλαμβάνεται στην 3'-UTR αναφοράς του γονιδίου.



**Εικόνα 3.17** Αποτελέσματα αλγορίθμου miRDB για τους πιθανούς στόχους από miRNAs στη νέα 3'-UTR αναφοράς, μήκους 602 nt, του γονιδίου *KLK14*.

#### 4. Συζήτηση αποτελεσμάτων και συμπεράσματα

Η παρούσα διπλωματική εργασία είχε ως βασικό στόχο την ανίχνευση και τον προσδιορισμό εναλλακτικών 3'-UTRs, στα ανθρώπινα γονίδια της οικογένειας των *KLKs*, μέσω ανάλυσης δεδομένων από αλληλούχηση νέας γενιάς (NGS). Κατά τη βιοπληροφορική ανάλυση των δεδομένων από NGS, έγινε εφαρμογή συγκεκριμένων αλγορίθμων για την επιλεκτική ομαδοποίηση των δεδομένων, με σκοπό την διευκόλυνση της μελέτης τους. Έπειτα, βάσει της ανάλυσης που έγινε ακολούθησε η επιλογή των πιθανών σημείων ύπαρξης εναλλακτικών 3'-UTRs σε συγκεκριμένα γονίδια των *KLKs* και σχεδιάστηκαν ειδικοί εκκινητές για την εφαρμογή των μεθοδολογιών 3'RACE και nested 3'RACE, με σκοπό την ειδική στόχευση και ενίσχυση των σημείων αυτών στα πιθανά εναλλακτικά μετάγραφα των γονιδίων *KLKs*. Στις μεθοδολογίες αυτές, έγινε εφαρμογή διαφοροποιημένων πρωτοκόλλων με σκοπό την αύξηση της ευαισθησίας τους. Τα προϊόντα από τις αντιδράσεις PCR, ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης και πραγματοποιήθηκε απομόνωση και καθαρισμός των επιθυμητών ζωνών που επιβεβαίωναν τα αναμενόμενα μήκη για τα εναλλακτικά 3'-UTRs. Ακολούθησε η αλληλούχηση τους κατα Sanger, για την ταυτοποίηση της αλληλουχίας τους και την επιβεβαίωση των αποτελεσμάτων της ύπαρξης των εναλλακτικών 3'-UTRs στα αντίστοιχα γονίδια.

Αθροιστικά, κατα την διπλωματικής αυτή εργασία επιτεύχθηκε η ταυτοποίηση 2 νέων εναλλακτικών 3'-UTRs σε μετάγραφα του γονιδίου *KLK7* και ενός τμήματος νέου εναλλακτικού 3'-UTR σε μετάγραφα του γονιδίου *KLK14*. Επιπλέον, έγινε εντοπισμός της θέσης και του μήκους αλληλουχίας, 7 επιπλέον πιθανών εναλλακτικών 3'-UTRs, σε μετάγραφα της οικογένειας ανθρώπινων γονιδίων των *KLKs*, όπως υποστηρίζουν τα αποτελέσματα από τις ηλεκτροφορήσεις των προϊόντων των αντιδράσεων που εφαρμόστηκαν, για τα οποία απαιτείται περαιτέρω μελέτη για την ταυτοποίησή τους. Οι αλληλουχίες των 2 νέων εναλλακτικών 3'-UTRs που ανακαλύφθηκαν, είναι γνωστές από ένα επιλεγμένο σημείο της αλληλουχίας του τελευταίου εξωνίου του μεταγράφου έως το τέλος του 3' άκρου. Για το τμήμα του νέου εναλλακτικού 3'-UTR στο γονίδιο *KLK14*, ενώ δεν έχει ταυτοποιηθεί η αλληλουχία του μέχρι το τελευταίο νουκλεοτίδιο του 3'-άκρου, πιθανολογείται ότι αυτό βρίσκεται λίγες μόνο βάσεις μετά το σημείο μέχρι το οποίο έχει γίνει αλληλούχηση, βάσει του μήκους αλληλουχίας του προϊόντος, όπως υποδεικνύεται από την ηλεκτροφόρηση στο πήκτωμα αγαρόζης και της ύπαρξης σημείου πολυαδενυλίωσης στο τέλος του τμήματος του εναλλακτικού 3'-UTR.

Από τη μελέτη των νέων εναλλακτικών 3'-UTRs τα οποία ταυτοποιήθηκαν, παρατηρείται η μεγάλη διαφορά τους, είτε στα μήκη αλληλουχιών είτε στην ίδια την αλληλουχία που περιλαμβάνουν. Οι διαφορές αυτές αποτελούν εξήγηση για τη διαφορά στα αποτελέσματα εύρεσης των μορίων miRNA, τα οποία πιθανολογείται ότι στοχεύουν τις αλληλουχίες αυτές στα νέα 3'-UTRs, σε σύγκριση με τα μόρια miRNA τα οποία πιθανώς στοχεύουν τις 3'-UTRs των γονιδίων *KLK7* και *KLK14*. Κατ'επέκταση, η στόχευση από διαφορετικά μόρια miRNAs στις 3'-UTRs των γονιδίων αυτών, υποδεικνύει και την διαφορετική ρύθμιση που υπόκεινται τα συγκεκριμένα γονίδια σε μετα-μεταγραφικό επίπεδο. Σημαντικό είναι το γεγονός ότι, η διαφορετική ρύθμιση ή απορρυθμηση που υπόκεινται τα συγκεκριμένα γονίδια και επηρεάζει την γονιδιακή και πρωτεϊνική τους έκφραση, αφορά κυρίως καρκινικές κυτταρικές σειρές.

Τα φαινόμενα εναλλακτικής συρραφής τα οποία παρουσιάζονται κατά τη πορεία ανάπτυξης καρκινικών όγκων έχουν άμεση σύνδεση με τις παθολογικές αυτές καταστάσεις καθώς παράγεται πληθώρα μεταγράφων ή πρωτεϊνικών ισομορφών, ενισχύοντας την διηθητική ικανότητα των καρκινικών κυττάρων σε παρακείμενους ιστούς και επιτρέπουν την απορρύθμιση σε διαδικασίες απόπτωσης. Η σχέση μεταξύ συγκεκριμένων θέσεων εναλλακτικής συρραφής και καρκινικών καταστάσεων δεν έχει διευκρινιστεί πλήρως ακόμη. Παρόλ'αυτά η μελέτη νέων εναλλακτικών μεταγράφων δίνει την ευκαιρία για ανακάλυψη νέων πιθανών μοριακών βιοδεικτών, με διαγνωστική ή / και προγνωστική αξία, νέων δυναμικών βιοδεικτών, ως προς την απόκριση ή την αντίσταση στη θεραπεία, είτε νέους θεραπευτικούς στόχους. Ιδιαίτερα στην οικογένεια των ανθρώπινων καλλικρεϊνών, υπάρχουν τα περιθώρια για εκτενή μελέτη, καθώς παρατηρείται έντονα το φαινόμενο εναλλακτικού ματίσματος, σε όλα τα γονιδιακά μέλη.

Από την παρούσα διπλωματική εργασία μπορούμε να θέσουμε ως μελλοντικούς στόχους, την ταυτοποίηση των νέων εναλλακτικών 3'-UTRs τα οποία υποδεικνύονται από τα αποτελέσματα των ηλεκτροφορήσεων, καθώς και την ανίχνευση και ταυτοποίηση νέων εναλλακτικών 3'-UTRs στα υπόλοιπα μέλη της οικογένειας των *KLKs* και την εύρεση ολόκληρης της αλληλουχίας των μεταγράφων στα οποία περιλαμβάνονται. Αυτό εμπεριέχει την περαιτέρω εφαρμογή των πρωτοκόλλων τα οποία έχουν χρησιμοποιηθεί κατά την διάρκεια της συγκεκριμένης διπλωματικής εργασίας, καθώς και την βελτιστοποίησή τους στο μέγιστο βαθμό. Ακολούθως, μπορεί να γίνει μελέτη των επιπέδων έκφρασης των μεταγραφικών και πρωτεϊνικών ισομορφών, σε διαφορετικούς τύπους καρκίνων. Θα ήταν ακόμη ιδιαίτερα ενδιαφέρον, να πραγματοποιηθεί η πειραματική επαλήθευση της στόχευσης



των νέων εναλλακτικών 3'-UTR μεταγράφων τα οποία έχουν ταυτοποιηθεί κατά τη διάρκεια της συγκεκριμένης διπλωματικής εργασίας, από τα miRNA μόρια τα οποία προβλέπεται ότι μπορούν να προσδεθούν σε συγκεκριμένες αλληλουχίες πρόσθεσης στη 3'-UTR περιοχή τους. Όπως γνωρίζουμε η στόχευση και πρόσδεση από συγκεκριμένα miRNA μόρια, έχει ως αποτέλεσμα τη ρύθμιση των επιπέδων έκφρασης των μεταγράφων-στόχων, η οποία κατ'επέκταση επηρεάζει και την πρωτεϊνική τους έκφραση.

## Βιβλιογραφία:

1. Sammeth, M., S. Foissac, and R. Guigo, *A general definition and nomenclature for alternative splicing events*. PLoS Comput Biol, 2008. **4**(8): p. e1000147.
2. Will, C.L. and R. Luhrmann, *Spliceosome structure and function*. Cold Spring Harb Perspect Biol, 2011. **3**(7).
3. Baralle, F.E. and J. Giudice, *Alternative splicing as a regulator of development and tissue identity*. Nat Rev Mol Cell Biol, 2017. **18**(7): p. 437-451.
4. Dutertre, M., S. Vagner, and D. Auboeuf, *Alternative splicing and breast cancer*. RNA Biol, 2010. **7**(4): p. 403-11.
5. Feng, J., et al., *The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator*. Genes Dev, 2006. **20**(11): p. 1470-84.
6. Hashimoto, K., et al., *A liver X receptor (LXR)-beta alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR-beta*. Biochem Biophys Res Commun, 2009. **390**(4): p. 1260-5.
7. Chen, J. and W.A. Weiss, *Alternative splicing in cancer: implications for biology and therapy*. Oncogene, 2015. **34**(1): p. 1-14.
8. Thermann, R., et al., *Binary specification of nonsense codons by splicing and cytoplasmic translation*. EMBO J, 1998. **17**(12): p. 3484-94.
9. Zhang, J., et al., *At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation*. Mol Cell Biol, 1998. **18**(9): p. 5272-83.
10. Zhang, J., et al., *Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm*. RNA, 1998. **4**(7): p. 801-15.
11. Blencowe, B.J., *Alternative splicing: new insights from global analyses*. Cell, 2006. **126**(1): p. 37-47.
12. Cieply, B. and R.P. Carstens, *Functional roles of alternative splicing factors in human disease*. Wiley Interdiscip Rev RNA, 2015. **6**(3): p. 311-26.
13. Beyer, K. and A. Ariza, *alpha-Synuclein posttranslational modification and alternative splicing as a trigger for neurodegeneration*. Mol Neurobiol, 2013. **47**(2): p. 509-24.
14. Vasquez, J.B., et al., *Alzheimer's Disease Genetics and ABCA7 Splicing*. J Alzheimers Dis, 2017. **59**(2): p. 633-641.
15. Kunkle, B.W., et al., *Targeted sequencing of ABCA7 identifies splicing, stop-gain and intronic risk variants for Alzheimer disease*. Neurosci Lett, 2017. **649**: p. 124-129.
16. Husain, M.A., et al., *NSAIDs Induced Regulation of Alternatively Spliced Transcript Isoforms: Possible Role in Cancer and Alzheimer Disease*. Curr Cancer Drug Targets, 2017. **17**(5): p. 467-478.
17. Ma, L., et al., *Expression and processing analyses of wild type and p.R47H TREM2 variant in Alzheimer's disease brains*. Mol Neurodegener, 2016. **11**(1): p. 72.
18. Lee, C., et al., *An isoform-specific role of FynT tyrosine kinase in Alzheimer's disease*. J Neurochem, 2016. **136**(3): p. 637-50.
19. Fu, R.H., et al., *Aberrant alternative splicing events in Parkinson's disease*. Cell Transplant, 2013. **22**(4): p. 653-61.
20. Gaweda-Walerych, K., et al., *Parkinson's disease-related gene variants influence pre-mRNA splicing processes*. Neurobiol Aging, 2016. **47**: p. 127-138.
21. Rindt, H., et al., *Optimization of trans-Splicing for Huntington's Disease RNA Therapy*. Front Neurosci, 2017. **11**: p. 544.
22. Chen, G.L., et al., *Modulation of nuclear REST by alternative splicing: a potential therapeutic target for Huntington's disease*. J Cell Mol Med, 2017. **21**(11): p. 2974-2984.

23. Fernandez-Nogales, M., et al., *Faulty splicing and cytoskeleton abnormalities in Huntington's disease*. Brain Pathol, 2016. **26**(6): p. 772-778.
24. Cabrera, J.R. and J.J. Lucas, *MAP2 Splicing is Altered in Huntington's Disease*. Brain Pathol, 2017. **27**(2): p. 181-189.
25. Reble, E., A. Dineen, and C.L. Barr, *The contribution of alternative splicing to genetic risk for psychiatric disorders*. Genes Brain Behav, 2018. **17**(3): p. e12430.
26. Pal, S., R. Gupta, and R.V. Davuluri, *Alternative transcription and alternative splicing in cancer*. Pharmacol Ther, 2012. **136**(3): p. 283-94.
27. Thorsen, K., et al., *Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis*. Mol Cell Proteomics, 2008. **7**(7): p. 1214-24.
28. Chang, J.T., Y.M. Lee, and R.S. Huang, *The impact of the Cancer Genome Atlas on lung cancer*. Transl Res, 2015. **166**(6): p. 568-85.
29. Matoulkova, E., et al., *The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells*. RNA Biol, 2012. **9**(5): p. 563-76.
30. Mayr, C., *Regulation by 3'-Untranslated Regions*. Annu Rev Genet, 2017. **51**: p. 171-194.
31. Chow, T.F., et al., *Kallikreins as microRNA targets: an in silico and experimental-based analysis*. Biol Chem, 2008. **389**(6): p. 731-8.
32. Pasic, M.D., et al., *Epigenetic regulation of kallikrein-related peptidases: there is a whole new world out there*. Biol Chem, 2012. **393**(5): p. 319-30.
33. Yousef, G.M., *microRNAs: a new frontier in kallikrein research*. Biol Chem, 2008. **389**(6): p. 689-94.
34. Borgono, C.A. and E.P. Diamandis, *The emerging roles of human tissue kallikreins in cancer*. Nat Rev Cancer, 2004. **4**(11): p. 876-90.
35. Yousef, G.M., et al., *Genomic organization of the human kallikrein gene family on chromosome 19q13.3-q13.4*. Biochem Biophys Res Commun, 2000. **276**(1): p. 125-33.
36. Diamandis, E.P., et al., *The new human kallikrein gene family: implications in carcinogenesis*. Trends Endocrinol Metab, 2000. **11**(2): p. 54-60.
37. Yousef, G.M., L.Y. Luo, and E.P. Diamandis, *Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4*. Anticancer Res, 1999. **19**(4B): p. 2843-52.
38. Clements, J., et al., *The expanded human kallikrein (KLK) gene family: genomic organisation, tissue-specific expression and potential functions*. Biol Chem, 2001. **382**(1): p. 5-14.
39. Paliouras, M. and E.P. Diamandis, *The kallikrein world: an update on the human tissue kallikreins*. Biol Chem, 2006. **387**(6): p. 643-52.
40. Yu, H., et al., *Genomic structure of the human plasma prekallikrein gene, identification of allelic variants, and analysis in end-stage renal disease*. Genomics, 2000. **69**(2): p. 225-34.
41. Sotiropoulou, G., G. Pampalakis, and E.P. Diamandis, *Functional roles of human kallikrein-related peptidases*. J Biol Chem, 2009. **284**(48): p. 32989-94.
42. Pampalakis, G. and G. Sotiropoulou, *Tissue kallikrein proteolytic cascade pathways in normal physiology and cancer*. Biochim Biophys Acta, 2007. **1776**(1): p. 22-31.
43. Emami, N. and E.P. Diamandis, *Human tissue kallikreins: a road under construction*. Clin Chim Acta, 2007. **381**(1): p. 78-84.
44. Yousef, G.M. and E.P. Diamandis, *Human tissue kallikreins: a new enzymatic cascade pathway?* Biol Chem, 2002. **383**(7-8): p. 1045-57.
45. Modrek, B. and C. Lee, *A genomic view of alternative splicing*. Nat Genet, 2002. **30**(1): p. 13-9.
46. Avgeris, M. and A. Scorilas, *Kallikrein-related peptidases (KLKs) as emerging therapeutic targets: focus on prostate cancer and skin pathologies*. Expert Opin Ther Targets, 2016. **20**(7): p. 801-18.
47. Kryza, T., et al., *The kallikrein-related peptidase family: Dysregulation and functions during cancer progression*. Biochimie, 2016. **122**: p. 283-99.

48. Avgeris, M., K. Mavridis, and A. Scorilas, *Kallikrein-related peptidases in prostate, breast, and ovarian cancers: from pathobiology to clinical relevance*. Biol Chem, 2012. **393**(5): p. 301-17.
49. Mavridis, K. and A. Scorilas, *Prognostic value and biological role of the kallikrein-related peptidases in human malignancies*. Future Oncol, 2010. **6**(2): p. 269-85.
50. Shendure, J., et al., *DNA sequencing at 40: past, present and future*. Nature, 2017. **550**(7676): p. 345-353.
51. Verma, M., S. Kulshrestha, and A. Puri, *Genome Sequencing*. Methods Mol Biol, 2017. **1525**: p. 3-33.
52. Tucker, T., M. Marra, and J.M. Friedman, *Massively parallel sequencing: the next big thing in genetic medicine*. Am J Hum Genet, 2009. **85**(2): p. 142-54.
53. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses*. Nat Rev Genet, 2014. **15**(2): p. 121-32.
54. Fernandez-Marmiesse, A., S. Gouveia, and M.L. Couce, *NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment*. Curr Med Chem, 2018. **25**(3): p. 404-432.
55. Chen, G., T. Shi, and L. Shi, *Characterizing and annotating the genome using RNA-seq data*. Sci China Life Sci, 2017. **60**(2): p. 116-125.
56. Ambardar, S., et al., *High Throughput Sequencing: An Overview of Sequencing Chemistry*. Indian J Microbiol, 2016. **56**(4): p. 394-404.
57. Anderson, M.W. and I. Schrijver, *Next generation DNA sequencing and the future of genomic medicine*. Genes (Basel), 2010. **1**(1): p. 38-69.
58. Sukumaran, S., *Concentration determination of nucleic acids and proteins using the micro-volume BioSpec-nano-spectrophotometer*. J Vis Exp, 2011(48).
59. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. Brief Bioinform, 2013. **14**(2): p. 178-92.
60. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011. **29**(1): p. 24-6.
61. Ponchel, F., et al., *Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions*. BMC Biotechnol, 2003. **3**: p. 18.
62. Ishmael, F.T. and C. Stellato, *Principles and applications of polymerase chain reaction: basic science for the practicing physician*. Ann Allergy Asthma Immunol, 2008. **101**(4): p. 437-43.
63. Chuang, L.Y., Y.H. Cheng, and C.H. Yang, *Specific primer design for the polymerase chain reaction*. Biotechnol Lett, 2013. **35**(10): p. 1541-9.

## ΠΕΡΙΛΗΨΗ

Σε όλους τους ευκαρυωτικούς οργανισμούς πραγματοποιείται η διαδικασία την συρραφής, στα πρόδρομα μετάγραφα των γονιδίων που κωδικοποιούν πρωτεΐνες, με στόχο τον σχηματισμό ώριμων mRNAs, μέσα από την αποβολή εσωνίων και συρραφή εξωνίων. Το φαινόμενο εναλλακτικής συρραφής δίνει την δυνατότητα δημιουργίας ποικιλίας mRNA μεταγράφων τα οποία προέρχονται από ένα μόνο γονίδιο. Η πολυπλοκότητα της συγκεκριμένης διαδικασίας, δικαιολογεί την αυστηρή ρύθμιση στην οποία υπόκειται. Ωστόσο, παρατηρούνται αρκετά φαινόμενα ρυθμιστικής δυσλειτουργίας, τα οποία μπορεί να σχετίζονται άμεσα με δυσπλασίες. Έτσι, πολλά εναλλακτικά μετάγραφα τα οποία παράγονται από μη αναμενόμενα γεγονότα συρραφής, μπορεί να αντιπροσωπεύουν πιθανούς καρκινικούς βιοδείκτες ή θεραπευτικούς στόχους.

Η οικογένεια ανθρώπινων γονιδίων των καλλικρεϊνών (*KLKs*), αποτελείται από 15 μέλη και κωδικοποιεί τη μεγαλύτερη ομάδα πρωτεασών σερίνης. Τα *KLKs* έχουν μεγάλη ομοιότητα και αποτελούνται από πέντε κωδικά εξώνια παρόμοιου μεγέθους καθώς και συντηρημένες αλληλουχίες εσωνίων. Επιπλέον, ως οικογένεια γονιδίων, παρουσιάζει αρκετά φαινόμενα εναλλακτικής συρραφής, με κύρια την παράλειψη εξωνίων, τη διατήρηση εσωνίων, την εναλλακτική 5' και 3' θέση συρραφής καθώς και τη χρήση εναλλακτικών θέσεων πολυαδενυλίωσης. Αρκετά από τα εναλλακτικά μετάγραφα, μπορεί να εμφανίζουν μη φυσιολογικά επίπεδα έκφρασης σε διάφορες κακοήθειες, έχοντας προγνωστική ή/και διαγνωστική αξία. Είναι πλέον γνωστό, ότι τα mRNAs των καλλικρεϊνών, στοχεύονται από συγκεκριμένα miRNAs μόρια, στις 3'-UTR περιοχές τους, ρυθμίζοντας την έκφρασή τους.

Η μελέτη της εναλλακτικής συρραφής έχει επεκταθεί αρκετά, κυρίως με τη χρήση της τεχνολογίας της αλληλούχησης νέας γενιάς (NGS). Συγκεκριμένα, η αλληλούχηση RNA (RNA seq) δίνει τη δυνατότητα ανάλυσης του mRNA εναλλακτικού ματίσματος και την ανίχνευση νέων εναλλακτικών μεταγράφων, με υψηλή ευαισθησία και ακρίβεια. Συνδυαστικά με την τεχνική ταχείας ενίσχυσης των 3'-άκρων cDNA (3'-RACE), μπορεί να προσφέρει την δυνατότητα μελέτης των εναλλακτικών θέσεων πολυαδενυλίωσης και την ανίχνευση εναλλακτικών 3'-αμετάφραστων περιοχών (3'-UTRs) στα *KLKs*.

Στόχος της παρούσας διπλωματικής εργασίας ήταν η ανίχνευση και ταυτοποίηση νέων εναλλακτικών 3'-αμετάφραστων περιοχών (3'-UTRs) στα ανθρώπινα γονίδια της οικογένειας των καλλικρεϊνών, με τη χρήση και ανάλυση δεδομένων αλληλούχησης νέας

γενιάς (NGS). Για την επίτευξη του στόχου αυτού, έγινε βιοπληροφορική ανάλυση των δεδομένων από NGS για την επιλογή των *KLK* γονιδίων και των θέσεών τους, τα οποία διαθέτουν πιθανά νέα εναλλακτικά 3'-UTRs. Έπειτα, πραγματοποιήθηκε nested 3' RACE με τη χρήση ειδικών ζευγών εκκινήτων, επιτρέποντας την ενίσχυση των ανθρώπινων μεταγράφων των γονιδίων *KLKs*, σε συνολικά 32 κυτταρικές σειρές από ανθρώπινους καρκινικούς ιστούς. Ακολούθως, η ταυτοποίηση των νέων εναλλακτικών 3'-UTRs τα οποία ανακαλύφθηκαν, έγινε μέσω αλληλούχησης κατά Sanger.

Συνολικά, ταυτοποιήθηκαν 2 νέα εναλλακτικά 3'-UTRs σε μετάγραφο του γονιδίου *KLK7* και ένα τμήμα νέου εναλλακτικού 3'-UTR σε μετάγραφο του γονιδίου *KLK14*. Επιπλέον, έγινε εντοπισμός της θέσης και του μήκους αλληλουχίας, 7 επιπλέον πιθανών εναλλακτικών 3'-UTRs, σε μετάγραφο των *KLKs*, όπως υποστηρίζουν τα αποτελέσματα από τις ηλεκτροφορήσεις των προϊόντων των αντιδράσεων 3'-RACE, για τα οποία απαιτείται περαιτέρω μελέτη για την ταυτοποίησή τους. Από την μελέτη των μεταγράφων των γονιδίων των καλλικρεϊνών, παρατηρήθηκε μεγάλη διαφορά στα μήκη και την αλληλουχία των 3'-UTRs που περιλαμβάνουν. Κατ' επέκταση, παρατηρούνται και διαφορετικά μόρια miRNA τα οποία δυνητικά μπορούν να στοχεύσουν τα εναλλακτικά 3'-UTRs, όπως αποκαλύπτουν διάφοροι αλγόριθμοι εύρεσης miRNA μορίων στόχευσης γονιδίων.

Τα νέα εναλλακτικά 3'-UTRs τα οποία ταυτοποιήθηκαν είναι μόνο λίγα από τα πληθώρα εναλλακτικών 3'-UTRs που πιθανά διαθέτουν τα διάφορα μετάγραφα των συγκεκριμένων γονιδίων. Η στόχευση των περιοχών αυτών από miRNA μόρια και η κατα συνέπεια ρύθμιση που υπόκεινται τα γονίδια αυτά, όσον αφορά τα επίπεδα έκφρασής τους, ιδιαίτερα σε περιπτώσεις κακοηθειών, οδηγεί στην αναγκαιότητα περαιτέρω μελέτης των 3'-εναλλακτικών μεταγράφων των γονιδίων των καλλικρεϊνών, καθώς μπορούν να αποτελέσουν σημαντικούς διαγνωστικούς ή/και προγνωστικούς καρκινικούς βιοδείκτες.

## SUMMARY

In all eukaryotic organisms, the process of splicing is carried out in the precursor translations of protein-coding genes, with the aim of generating mRNAs through the cleavage of insertions and the exon splicing. The alternative splicing phenomenon enables the generation of a variety of mRNA transcripts derived from a single gene. The complexity of the procedure justifies the strict regulation to which it is subject. However, there are several regulatory dysfunctions that may be directly related to malformations. Thus, many alternative translations produced by unexpected splice events may represent potential cancer biomarkers or therapeutic targets.

The human kallikrein gene family (*KLKs*) consists of 15 members and encodes the largest group of serine proteases. The *KLKs* are highly uniform and consist of five codons of similar size and conserved sequences. In addition, as a gene family, several alternative synergy phenomena occur, with major exons skipping, alternative 5' and 3' splicing sites as well as the use of alternative polyadenylation sites. Alternative translations may have abnormal expression values in various malignancies, with predictive and / or diagnostic value. It is now known that kallikrein mRNAs are targets of specific miRNAs in their 3'-UTR domains, regulating their expression.

The study of alternative stapling has been greatly expanded, notably with the use of next generation sequencing technology (NGS). In particular, the RNA sequence (RNA seq) enables the analysis of the alternative mRNA complex and the detection of new alternative transcripts, with high sensitivity and accuracy. Combined with the 3'-terminal cDNA rapid amplification technique (3'-RACE), it may offer the opportunity to study alternative polyadenylation sites and detect alternative 3-untranslated regions (3'-UTRs) in *KLKs*.

The aim of this thesis was to identify and identify new alternative 3'-translational regions (3'-UTRs) in the human genes of the kallikrein family, using next generation sequencing (NGS) data. To achieve this goal, bioinformatic analysis of NGS data on the selection of *KLK* genes and their locations, which may potentially have new alternative 3'-UTRs, was performed. Next, a nested 3'-RACE was performed using specific pairs of primers, allowing the amplification of human transcripts of *KLKs* in a total of 32 human cancer tissue lines. Subsequently, the identification of the new alternative 3'-UTRs discovered was carried out by Sanger sequencing.

In total, 2 new alternative 3'-UTRs were identified in translations of the *KLK7* gene and one portion of a new alternative 3'-UTR in transcripts of the *KLK14* gene. In addition, 7 potential

alternative 3'-UTRs, in translations of KLKs, support the results from the 3'-RACE product of electrophoresis, for which further study is needed to identify them. From the study of transcripts of the kallikrein genes, there was a large difference in the lengths and sequences of the 3'-UTRs they contain. By extension, different miRNA molecules are also observed which can potentially target alternative 3'-UTRs, as revealed by various algorithms for finding miRNA gene targeting molecules.

The new 3'-UTR variants identified are only a few of the numerous 3'-UTR alternatives that may possess the different transcripts of those genes. The detection of these regions by miRNA molecules and consequently the regulation of these genes with respect to their expression levels, especially in cases of malignancies, necessitates further study of the 3 alternative transcripts of the kallikerlin genes, as important diagnostic and / or and prognostic cancer biomarkers.