

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ ΤΟΜΕΑΣ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών "Κλινική Βιοχημεία - Μοριακή Διαγνωστική"

ΕΡΕΥΝΗΤΙΚΗ ΕΡΓΑΣΙΑ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Ταυτοποίηση, ανάλυση και κλινική αξιολόγηση εναλλακτικών μεταγράφων γονιδίων καθώς και μη κωδικών μορίων RNAs (ncRNAs), στο γενετικό τόπο 19q13.3-13.4, με χρήση δεδομένων αλληλούχησης νέας γενιάς, σε καρκινικά κύτταρα

> ΡΑΠΤΗΣ ΓΕΩΡΓΙΟΣ ΒΙΟΛΟΓΟΣ

> > Αθήνα 2017

ΠΡΟΛΟΓΟΣ

Η παρούσα ερευνητική εργασία διπλώματος ειδίκευσης εκπονήθηκε στον Τομέα Βιοχημείας και Μοριακής Βιολογίας του Τμήματος Βιολογίας της Σχολής Θετικών Επιστημών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών (Ε.Κ.Π.Α.), στο πλαίσιο του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών «Κλινική Βιοχημεία – Μοριακή Διαγνωστική», με επιστημονικό υπεύθυνο τον κύριο Ανδρέα Σκορίλα, Καθηγητή Κλινικής Βιοχημείας.

Η γνώση και η εμπειρία που αποκομίζει κανείς από την εκπόνηση μιας ερευνητικής εργασίας εξαρτώνται, όχι μόνο από το βαθμό ατομικής ενεργοποίησης, αλλά και από τη συμβολή μιας ομάδας ανθρώπων. Η παρούσα εργασία δε θα μπορούσε να ολοκληρωθεί χωρίς τη συμβολή ορισμένων ατόμων που θα ήθελα να ευχαριστήσω προσωπικά.

Έτσι, πρώτα από όλους, θέλω να ευχαριστήσω ολόψυχα τον κ. Ανδρέα Σκορίλα, Καθηγητή Κλινικής Βιοχημείας του Τμήματος Βιολογίας, (Ε.Κ.Π.Α.) αφενός μεν για την τιμή που μου έκανε να συμμετάσχω στην ερευνητική του ομάδα και να ασχοληθώ με ένα τόσο σύγχρονο και ενδιαφέρον αντικείμενο έρευνας, αφετέρου δε για το συνεχές του ενδιαφέρον, την επιστημονική του καθοδήγηση και υποστήριξη, και για τη συμβολή του στη βελτίωση του τρόπου σκέψης και κρίσης μου. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον κύριο Δημήτριο Γουργιώτη, Καθηγητή Κλινικής Βιοχημείας του Τμήματος Ιατρικής (Ε.Κ.Π.Α.), καθώς και τον κύριο Διαμάντη Σίδερη Αναπληρωτή Καθηγητή Βιοχημείας Ευκαρυωτικών Οργανισμών του Τμήματος Βιολογίας (Ε.Κ.Π.Α.), για τη συμμετοχή τους στην τριμελή επιτροπή και την αξιολόγηση της ερευνητικής μου εργασίας.

Ιδιαίτερες ευχαριστίες θέλω να εκφράσω στον Παναγιώτη Αδαμόπουλο, για τη βοήθεια και την πολύτιμη καθοδήγηση σε όλα τα στάδια εκπόνησης της ερευνητικής εργασίας, για το χρόνο που μου αφιέρωσε, για τις εποικοδομητικές παρατηρήσεις, για την συμβολή του με τις γνώσεις και την εμπειρία του στην επίλυση διαφόρων προβλημάτων που αντιμετώπισα, αλλά και για την υποστήριξή του σε ανθρώπινο επίπεδο. Η άριστη συνεργασία μεταξύ μας αποτέλεσε καθοριστικό παράγοντα για την επιτυχή και ομαλή διεξαγωγή της παρούσας ερευνητικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω και όλα τα μέλη του εργαστηρίου για την άψογη συνεργασία, τη βοήθειά τους και τη φιλία τους.

Τέλος, θερμές ευχαριστίες απευθύνω στην οικογένειά μου για τη συμπαράστασή της, στις προσπάθειες και τους αγώνες που έχω καταβάλει όλα αυτά τα χρόνια σε διάφορους τομείς.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΡΟΛΟΓΟΣ		
1. ΕΙΣΑΓΩΓΗ		1
1.1. АЛЛНЛ ВІОЛОГІКН ЕІ	ΟΥΧΗΣΗ DNA ΚΑΙ ΣΗΜΑΤΟΔΟΤΗΣΗ ΝΕΑΣ ΕΠΟΧΗΣ ΓΙΑ Τ ΡΕΥΝΑ	ГН 1
1.2. Н ЕЛЕҮ SEQUENCING,	ΣΗ ΤΗΣ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ (NEXT GENERA NGS)	TION
1.3. ΣΥΣΤΗΝ	ΜΑΤΑ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ	6
1.3.1. Roche	2/454 FLX	6
1.3.2. 5500 Fisher Scientif	Solid (Sequencing by Oligo Ligation and Detection) – The ic	ermo
1.3.3. Illumi	ina – Solexa	12
1.3.4. Ion Te	orrent – Thermo Fisher Scientific	14
1.3.5. Helico	os True Single Molecule Sequencing (tSMS)	16
1.3.6. Single	e molecule real time sequencing (SMRT) - Pacific Bioscien	nces18
1.3.7. Nanoj	pore DNA Sequencer - Oxford Nanopore Technologies	
1.4. ΒΙΟΠΛΗ ΑΛΛΗΛΟΥΧΗΣ	ΙΡΟΦΟΡΙΚΑ ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΉΣ ΝΕΑΣ ΓΕΝΙΑΣ	[2 1
1.5. ЕФАРМ	ΟΓΕΣ ΤΗΣ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ	22
1.5.1. Αλληλ	λούχηση καρκινικών γονιδιωμάτων	23
1.5.2. Αλληλ	λούχηση εξονίων (exomes sequencing)	23
1.5.3. Αλληλ	λούχηση RNA (RNA-Seq)	24
1.5.4. Ανακά	άλυψη πληθυσμών RNA με Αλληλούχηση νέας γενιάς	24
1.6. ENAЛЛ	AKTIKO MATIΣMA (ALTERNATIVE SPLICING)	25
1.6.1. Αλληλ	\ουχίες εντός του RNA καθοριστικές για το μάτισμα	26
1.6.2. Ο μηχ	ανισμός του ματίσματος	27
1.6.3. Basik	κοί τύποι εναλλακτικού ματίσματος	
1.6.4 Ο ρόλ	ος των πρωτεϊνών SR	31
1.7. ΓΟΝΙΔΙ	O SCAF1	
2 ΥΛΙΚΑ ΚΑΙ Ν	ΛΕΘΟΔΟΙ	4(
2.1. ΒΙΟΛΟΓ	ΊΚΟ ΥΛΙΚΟ	40
2.2. АПОМО	ΝΩΣΗ ΟΛΙΚΟΥ RNA ΑΠΟ ΑΝΘΡΩΠΙΝΕΣ ΚΥΤΤΑΡΙΚΕΣ ΣΕ	ΙΡΕΣ.41
2.2.1. Φασμ έλεγχος της κα	ατοφωτομετρικός προσδιορισμός της συγκέντρωσης κα χθαρότητας του απομονωμένου ολικού RNA	a 43

2]	2.3. ΓRAN	ΣΎΝΘΕΣΗ cDNA ME ANTIΣΤΡΟΦΗ ΜΕΤΑΓΡΑΦΗ ΤΟΥ mRNA (REVERS ISCRIPTION, RT)	SE .44
2	2.3.1.	Αρχή της μεθόδου	.44
2	2.3.2. από ο	Συνθήκες αντίδρασης αντίστροφης μεταγραφής για τη σύνθεση cD απομονωμένο ολικό RNA	NA .47
2 F	2.4. REAC	ΑΛΥΣΙΔΩΤΗ ΑΝΤΙΔΡΑΣΗ ΠΟΛΥΜΕΡΑΣΗΣ (POLYMERASE CHAIN TION, PCR)	.49
2	2.4.1.	Αρχή της μεθόδου	.49
2	2.4.2.	Αντιδραστήρια της PCR	.53
2	2.4.3.	Μοριακή κλωνοποίηση των μεταγράφων του γονιδίου SCAF1	.59
2	2.4.4.	Αλυσιδωτή Αντίδραση Ταχείας Ενίσχυσης 3'-Άκρου (3'-RACE PCR).	61
2	2.4.5.	Καθαρισμός προϊόντων PCR	.64
2	2.5.	ΚΑΤΑΣΚΕΥΉ ΒΙΒΛΙΟΘΗΚΗΣ	.65
2 F F	2.6. аллн and 1	ΠΡΟΕΤΟΙΜΑΣΙΑ ΕΚΜΑΓΕΙΟΥ ΑΛΛΗΛΟΥΧΗΣΗΣ, ΕΜΠΛΟΥΤΙΣΜΟΣ ΚΑΙ ΙΛΟΥΧΗΣΗ ΝΕΑΣ ΓΕΝΙΑΣ (TEMPLATE PREPARATION, ENRICHMENT, NEXT-GENERATION SEQUENCING)	67
2	2.7.	ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ	.73
2 4	2.8. ΑΛΥΣ	ΑΝΑΛΥΣΗ ΕΚΦΡΑΣΗΣ ΤΩΝ ΝΕΩΝ ΜΕΤΑΓΡΑΦΩΝ ΜΕ ΕΣΩΤΕΡΙΚΗ ΊΔΩΤΗ ΑΝΤΙΔΡΑΣΗ ΠΟΛΥΜΕΡΑΣΗΣ (nested RT-PCR)	.77
2 I	2.9. 10Л¥	ΑΝΙΧΝΕΥΣΗ ΤΩΝ ΠΡΟΪΟΝΤΩΝ ΤΗΣ ΑΛΥΣΙΔΩΤΗΣ ΑΝΤΙΔΡΑΣΗΣ (ΜΕΡΑΣΗΣ (PCR), ΜΕ ΗΛΕΚΤΡΟΦΟΡΗΣΗ ΣΕ ΠΗΚΤΩΜΑ ΑΓΑΡΟΖΗΣ	.83
2	2.10.	ΤΡΙΣΔΙΑΣΤΑΤΗ ΑΠΕΙΚΟΝΙΣΗ ΝΕΩΝ ΠΡΩΤΕΪΝΙΚΩΝ ΙΣΟΜΟΡΦΩΝ	.85
3.	АП	ΟΤΕΛΕΣΜΑΤΑ	.87
3	3.1.	ΝΕΕΣ ΘΕΣΕΙΣ ΣΥΡΡΑΦΗΣ ΕΞΟΝΙΩΝ.	.87
3	3.2.	ΤΑΥΤΟΠΟΙΗΣΗ 15 ΝΕΩΝ <i>SCAF1</i> ΜΕΤΑΓΡΑΦΩΝ	.89
3	3.3.	ΠΡΟΤΥΠΟ ΕΚΦΡΑΣΗΣ ΤΩΝ ΝΕΩΝ <i>SCAF1</i> ΜΕΤΑΓΡΑΦΩΝ	.93
3 N	3.4. MET <i>A</i>	ΔΙΕΡΕΥΝΗΣΗ ΤΩΝ 3' ΑΜΕΤΑΦΡΑΣΤΩΝ ΠΕΡΙΟΧΩΝ (UTRs) ΤΩΝ ΝΕΩΝ ΑΓΡΑΦΩΝ <i>SCAF1</i>	N .97
3	3.5. ΓΡΙΣΖ	ΝΕΕΣ ΠΙΘΑΝΕΣ ΠΡΩΤΕΪΝΙΚΕΣ ΙΣΟΜΟΡΦΕΣ ΚΑΙ ΟΙ ΕΠΙΚΡΑΤΕΣΤΕΡΕΣ ΔΙΑΣΤΑΤΕΣ ΔΟΜΕΣ ΤΟΥΣ	: .98
4.	ΣΥΖ	ΖΗΤΗΣΗ1	.00
5.	BIB	ЗЛІОГРАФІА1	.09
6.	ПA	PAPTHMA1	.28

1. ΕΙΣΑΓΩΓΗ

1.1. ΑΛΛΗΛΟΥΧΗΣΗ DNA ΚΑΙ ΣΗΜΑΤΟΔΟΤΗΣΗ ΝΕΑΣ ΕΠΟΧΗΣ ΓΙΑ ΤΗ ΒΙΟΛΟΓΙΚΗ ΕΡΕΥΝΑ

Τα τελευταία χρόνια είμαστε μάρτυρες μια συνεχούς προόδου στον τομέα της βιολογικής έρευνας, η οποία συνδυάζεται με την ανάπτυξη της τεχνολογίας και τη συνεχή επινόηση τεχνικών, στις οποίες συμπεριλαμβάνονται και οι τεχνικές που αφορούν την αλληλούχηση του DNA. Πρόκειται για αναμφισβήτητο γεγονός πως μία από τις πιο αξιοσημείωτες καινοτομίες των τελευταίων ετών αποτελεί η έλευση της αλληλούχησης νέας γενιάς (Next-Generation Sequencing, NGS).

Με την ολοκλήρωση της χαρτογράφησης του ανθρώπινου γονιδιώματος και της παρουσίασης των αποτελεσμάτων το 2001 [1, 2], υπήρξε μια ραγδαία ανάπτυξη τεχνολογιών που θα διευκόλυναν την υψηλής απόδοσης αλληλούχηση του DNA (high - throughput analysis). Στο σημείο αυτό αξίζει να παρουσιασθεί συνοπτικά η πορεία της αλληλούχησης στο πέρασμα των ετών, που θα αποτυπώσει τον τρόπο που καταφέραμε να φτάσουμε στην εποχή της αλληλούχησης νέας γενιάς.

Στην προσπάθεια κατανόησης των θεμελιωδών αρχών της ζωής, ο Ελβετός ιατρός και βιολόγος Friedrich Miescher απομόνωσε για πρώτη φορά DNA το 1869 [3]. Αργότερα άλλοι επιστήμονες θα ακολουθήσουν το παράδειγμα του που είχε ως αποτέλεσμα την περαιτέρω πρόοδο στο πρωτόκολλο απομόνωσης και καθαρισμού του DNA.

Το 1953 οι James Watson και Francis Crick θα παρουσιάσουν το μοντέλο της διπλής έλικας, σύμφωνα με το οποίο το μόριο του DNA αποτελείται από δύο συμπληρωματικές πολυνουκλεοτιδικές αλυσίδες που σχηματίζουν δεξιόστροφη διπλή έλικα και οι οποίες συγκρατούνται μεταξύ τους με δεσμούς υδρογόνου, συγκεκριμένα μεταξύ αδενίνης (Α) και θυμίνης (Τ) δημιουργείται διπλός δεσμός υδρογόνου, ενώ μεταξύ γουανίνης (G) και κυτοσίνης (C) τριπλός δεσμός [4]. Ουσιαστικά με την αλληλούχηση του DNA μπορούμε να προσδιορίσουμε τη σειρά των τεσσάρων αυτών βάσεων, σε μια αλυσίδα DNA.

Η πρώτη μέθοδος προσδιορισμού της ακολουθίας του DNA περιλάμβανε την επιμήκυνση ειδικών για συγκεκριμένες περιοχές του DNA εκκινητών, μια στρατηγική που αναπτύχθηκε από τον Ray Wu το 1970. Η καταλυτική ενεργότητα της DNA πολυμεράσης και η ειδική σήμανση νουκλεοτιδίων, οι οποίες χρησιμοποιούνται ευρέως και σήμερα για την αλληλούχηση, χρησιμοποιήθηκε για την αλληλούχηση DNA των άκρων του φάγου λ [5-7]. Μεταξύ του 1970 και 1973 οι Wu, Padmanabhan και οι ερευνητικές τους ομάδες, έδειξαν ότι η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για τον προσδιορισμό οποιασδήποτε ακολουθίας DNA χρησιμοποιώντας συνθετικούς ειδικούς για περιοχές του DNA εκκινητές [8-10].

To 1977, o Sanger και η ερευνητική του ομάδα, επιτυγχάνει τον προσδιορισμό της αλληλουχίας του γενετικού υλικού του βακτηριοφάγου φΧ174 της Ε. coli χρησιμοποιώντας μία απλή και γρήγορη μέθοδο αλληλούχησης μονόκλωνου DNA (κλώνοι «συν/πλην») [11]. Ο ίδιος δημοσιεύει μια νέα και αρκετά γρήγορη για τα δεδομένα της εποχής εκείνης μέθοδο προσδιορισμού της αλληλουχίας του DNA, κατά την οποία χρησιμοποιείται η μια αλυσίδα του DNA

ως καλούπι προκειμένου να συντεθεί μια συμπληρωματική αλυσίδα (Εικόνα 2), με χρήση όχι μόνο τριφωσφορικών δεοξυνουκλεοσιδίων αλλά και συγκεκριμένης ποσότητας από τα διδεόξυ ανάλογά τους τριφωσφορικά 2',3'διδεοξυνουκλεοσίδια, οποία τα σε αντίθεση φυσιολογικά με τα νουκλεοτίδια, δε φέρουν 3'-0H την (Εικόνα 1). Έτσι δρουν ως σημεία λήξης, αφού δε θα μπορεί να προστεθεί άλλο νουκλεοτίδιο καθώς στο άκρο της αλυσίδας δε θα υπάρχει ελεύθερο 3'-OH (http://archive.cnx.org/resources/a947) [12].

Η αλληλούχηση κατά Sanger βελτιστοποίησε σε μεγάλο βαθμό τις

P (P)-OCH2 Base (P)Dideoxynucleotide (ddNTP) OCH₂ Base P

Deoxynucleotide (dNTP) Εικόνα 1: Το διδεοξυνουκλεοσίδιο έχει παρόμοια δομή με το δεοξυνουκλεοσίδιο, αλλά λείπει 3'-OH του η 2231811262e673e8ac2ba87500049933 e09a/Figure_17_03_01.jpg).

τεχνικές αλληλούχησης που είχαν αναπτυχθεί νωρίτερα από τους Maxam και Gilbert [13]. Το εμφανές πλεονέκτημα της μείωσης της χρήσης τοξικών χημικών και ραδιοϊσοτόπων, κατέστησε άμεσα τη μέθοδο αλληλούχησης κατά Sanger, ως τη μόνη μέθοδο αλληλούχησης DNA που θα χρησιμοποιούταν για τα επόμενα τριάντα χρόνια [14].



Εικόνα 2: Ο προσδιορισμός της αλληλουχίας του DNA με τη μέθοδο Sanger με χρήση συσκευής αυτόματης αλληλούχησης όπου κάθε τμήμα ακτινοβολείται από το λέιζερ του ανιχνευτή. Οι διεγερμένες χρωστικές καθώς επανέρχονται σε ενεργειακό επίπεδο ηρεμίας εκπέμπουν φως συγκεκριμένου μήκους κύματος, χαρακτηριστικό για κάθε χρωστική. Η αυτοματοποίηση αυτή που πραγματοποιήθηκε μια δεκαετία αργότερα, έδωσε τη δυνατότητα στα εργαστήρια να εκτελούν αλληλούχηση του DNA σε επίπεδο ρουτίνας.

(https://www.google.gr/search?q=sanger+sequencing&espv=2&source=lnms&tbm=isc h&sa=X&ved=0ahUKEwio-Yylh7HTAhUMAsAKHcETBHQQ_AUIBigB&biw=1280&bih= 918#imgrc=eJq1W1vu2PRP2M).

Ωστόσο, παρότι πρόκειται για μια σχετικά οικονομική μέθοδο και χαρακτηρίζεται επίσης από μεγάλη ακρίβεια, η μέθοδος Sanger μπορεί να χρησιμοποιηθεί για την αλληλούχηση τμημάτων με μήκος μόνο μέχρι τα 1000 ζεύγη βάσεων. Για τον προσδιορισμό της αλληλουχίας του γονιδιώματος του φΧ174 με μήκος περίπου 5.375 νουκλεοτίδια [11], χρειάστηκαν πάνω από 24 μήνες. Αν θα θέλαμε να αλληλουχήσουμε ολόκληρο το γονιδίωμα του ανθρώπου θα χρειαζόταν να εκτελεστεί η παραπάνω μέθοδος 3.000.000 φορές [15].

Έχοντας ως απόλυτο στόχο την αποκρυπτογράφηση του ανθρώπινου γονιδιώματος, η απαίτηση για αλληλούχηση του DNA με μεγαλύτερη απόδοση αυξήθηκε σε μεγάλο βαθμό, οδηγώντας σε ανάπτυξη τεχνικών όπως η αυτοματοποιημένη ηλεκτροφόρηση τριχοειδούς (Εικόνα 2). Έτσι δεν άργησαν να δημιουργηθούν επιχειρήσεις που ονομάζονταν κέντρα αλληλούχησης, τα οποία χρησιμοποιούσαν εκατοντάδες μηχανήματα για αλληλούχηση του DNA. Παρ' όλα αυτά ακόμα και μετά την επιτυχή ολοκλήρωση του προγράμματος της χαρτογράφησης του ανθρώπινου γονιδιώματος, συνέχισε να αυξάνεται η επιθυμία για αλληλούχηση υψηλής απόδοσης και κυρίως για ανάπτυξη τεχνολογίας προκειμένου να μειωθεί το κόστος αλληλούχησης [14].

Οι νουκλεοτιδικές αλληλουχίες που προέκυπταν αυτά τα χρόνια, καταχωρούνταν σε δημόσιες βάσεις δεδομένων (GenBank, EBI, DBJL). Χαρακτηριστικά, η βάση δεδομένων GenBank περιείχε το 1983 2,2 × 10⁶ ζεύγη βάσεων, ενώ το 1994 ο αριθμός αυτός έφτασε στα 2,2 × 10⁸ ζεύγη βάσεων. Το έτος 2006, η GenBank περιείχε πάνω από 4,6 × 10¹⁰ ζεύγη βάσεων, για να φτάσουμε στο 2017 να ξεπεράσουν τα 2,3 × 10¹¹ ζεύγη βάσεων [16]. Το πρόγραμμα του Ανθρώπινου Γονιδιώματος έφερε ραγδαία ανάπτυξη των δυνατοτήτων αλληλούχησης, με αποτέλεσμα την αύξηση του αριθμού των νουκλεοτιδικών αλληλουχιών που κατατίθονταν στις βάσεις. Σημείο σταθμός στην προσπάθεια αυτή αποτελεί το έτος 2005, όπου εκατό δισεκατομμύρια ζεύγη βάσεων κατατέθηκαν στις βάσεις δεδομένων. Οι αλληλουχίες αυτές συνείσφεραν μεταξύ άλλων, στο σχεδιασμό πειραμάτων και στην ανάπτυξη μοντέλων που σχετίζονται με τη λειτουργία των γονιδίων, καθώς και τις μοριακές διαδικασίες του κυττάρου [17].

4

1.2. Η ΕΛΕΥΣΗ ΤΗΣ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ (NEXT GENERATION SEQUENCING, NGS)

Όπως έχουμε δει μέχρι τώρα, η μέθοδος αλληλούχησης κατά Sanger αποτέλεσε τη βασική μέθοδο αλληλούχησης για μια σειρά ετών, ωστόσο, δε παύει να εμφανίζει κάποια μειονεκτήματα τα οποία μας δημιουργούν ορισμένα εμπόδια κατά την αλληλούχηση. Για παράδειγμα, η ακρίβεια της μεθόδου επηρεάζεται αρνητικά, όταν παρατηρούνται αυξημένα επίπεδα γουανίνης – κυτοσίνης στην αλυσίδα του DNA που πρόκειται να αλληλουχηθεί. Όμοια, επαναλαμβανόμενες περιοχές στο DNA μπορούν επίσης να επηρεάσουν την ακρίβεια της αλληλούχησης [18]. Παρόλο που η "shotgun" αλληλούχηση ενίσχυσε το ρυθμό αλληλούχησης [19], και αργότερα αυτοματοποιήθηκε σε μεγαλύτερο βαθμό, εμφάνιζε ακόμα παρόμοια θέματα [20].

Για να ξεπεραστούν αυτά τα εμπόδια ήταν αναγκαία μια εντελώς διαφορετική προσέγγιση που συνεπαγόταν την αναζήτηση μια εντελώς νέας τεχνολογίας. Έτσι αναπτύσσονται σταδιακά οι τεχνολογίες της αλληλούχησης νέας γενιάς (Next Generation Sequencing, NGS), οι οποίες αποτελούν ένα μεγάλο βήμα κυρίως σε θέματα ταχύτητας της αλληλούχησης.

Η αλληλούχηση νέας γενιάς περιλαμβάνει το σπάσιμο του δείγματος DNA σε εκατομμύρια μικρότερα τμήματα. Έπειτα, ολιγονουκλεοτιδικά τμήματα καθορισμένης αλληλουχίας (adapters) συνδέονται με αυτά τα τμήματα DNA τα οποία θα δεσμευτούν σε μια μήτρα, όπου θα πραγματοποιηθούν αντιδράσεις ενίσχυσης και θα ακολουθήσει η αλληλούχηση. Η αλληλούχηση πραγματοποιείται ταυτόχρονα σε όλα τα προς αλληλούχηση τμήματα του DNA, διαδικασία που είναι γνωστή στη ξένη βιβλιογραφία ως «massively parallel sequencing» [21].

Κατά τη διάρκεια της αλληλούχησης ανιχνεύονται σήματα ο τύπος των οποίων εξαρτάται από την χρησιμοποιούμενη πλατφόρμα αλληλούχησης και την τεχνολογία που αυτή χρησιμοποιεί. Κάθε τμήμα του DNA που έχει αλληλουχηθεί χαρακτηρίζεται ως «read» (ανάγνωσμα) [22]. Η αλληλούχηση νέας γενιάς έχει την ικανότητα να παράγει εκατομμύρια "reads" τα οποία είναι γενικά μικρά, περίπου 35 έως 1000 βάσεις [23]. Ο τεράστιος αυτός αριθμός των reads απαιτεί σημαντική επεξεργαστική ισχύ, προκειμένου να στοιχηθούν τα reads που προκύπτουν στο γονιδίωμα αναφοράς. Ο αριθμός των reads που ευθυγραμμίζονται ή καλύπτουν τις βάσεις που απαρτίζουν το γονιδίωμα του οργανισμού που αλληλουχείται χαρακτηρίζεται ως «coverage» (κάλυψη). Για κάθε συγκεκριμένο βαθμό κάλυψης, μπορούμε να υπολογίσουμε το ποσοστό του γονιδιώματος που δε θα περιλαμβάνεται στις αλληλουχίες που έχουν αλληλουχηθεί, με χρήση μαθηματικής εξίσωσης που βασίζεται στην κατανομή Poisson [17, 24]. Ανάλογα με το λόγο για τον οποίο πραγματοποιείται η αλληλούχηση, οι απαιτήσεις του βαθμού κάλυψης της αλληλούχησης ποικίλουν. Όταν έχουμε επικάλυψη σε μεγάλο βαθμό, η κάθε βάση καλύπτεται από μεγαλύτερο αριθμό στοιχισμένων reads, έτσι ο προσδιορισμός της κάθε βάσης μπορεί να γίνει με μεγαλύτερο βαθμό εμπιστοσύνης [24, 25]. Για να παραχθούν πιο αξιόπιστα δεδομένα, η ίδια περιοχή μπορεί να χρειασθεί να αλληλουχηθεί [15].

1.3. ΣΥΣΤΗΜΑΤΑ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ

Μεταξύ των διαθέσιμων συστημάτων αλληλούχησης νέας γενιάς, οι Roche/454 FLX, Illumina/Solexa Genome Analyzer, 5500 SOLiD και Ion Torrent από την Thermo Fisher Scientific, είναι αυτές που προς το παρόν κυριαρχούν στην αγορά. Στα επόμενα χρόνια είναι πιθανή η προσθήκη νέων συστημάτων, τα οποία θα βελτιώσουν τις υπάρχουσες τεχνολογίες καθώς και θα δημιουργήσουν καινούριες, προκειμένου να συμβάλλουν στην αύξηση της ταχύτητας και της ακρίβειας της αλληλούχησης, αλλά και στη μείωση του κόστους αλληλούχησης. Στη συνέχεια θα αναφερθεί αναλυτικά, η μεθοδολογία που χρησιμοποιείται στα κυριότερα συστήματα αλληλούχησης.

1.3.1. Roche/454 FLX

Το 2005, ο Jonathan Rothberg και οι συνεργάτες του, ανέφεραν την ανάπτυξη του πρώτου εμπορικά διαθέσιμου συστήματος αλληλούχησης νέας γενιάς (454 Genome Sequencer) [26]. Το δεύτερο ολόκληρο γονιδίωμα ατόμου αλληλουχήθηκε με αυτό το σύστημα και ήταν του James D. Watson [27]. Το πρώτο βήμα της 454 τεχνικής είναι η παραγωγή βιβλιοθήκης DNA, που περιέχει αλληλουχίες αντάπτορες, οι οποίες χρησιμοποιούνται για να αγκυροβολούν τα θραύσματα DNA της βιβλιοθήκης σε σφαιρίδια. Έπειτα, καθένα από τα σύμπλοκα θραύσματος - σφαιριδίου που δημιουργήθηκαν, απομονώνεται μαζί με τα απαιτούμενα αντιδραστήρια PCR μέσα σε μίγμα νερού και ελαίου, δημιουργώντας κατά αυτό τον τρόπο κυστίδια. Έτσι εξασφαλίζεται ένα περιβάλλον που αποτελείται από πολλούς μικροαντιδραστήρες όπου τα θραύσματα ενισχύονται μέσω αλυσιδωτής αντίδρασης πολυμεράσης σε γαλάκτωμα (Emulsion PCR) [26, 28].

Προσθέτοντας τη σωστή ποσότητα από τη βιβλιοθήκη του DNA στο μίγμα της αντίδρασης, μπορεί κανείς να εξασφαλίσει την ενίσχυση κατά μέσο όρο ενός μόνο θραύσματος DNA σε κάθε σφαιρίδιο. Μετά από την ενίσχυση, το γαλάκτωμα διασπάται με τη προσθήκη διαλύτη και τα σφαιρίδια επωάζονται με μαγνητικά σφαιρίδια που καλύπτονται από στρεπταβιδίνη προς επιλεκτικό καθαρισμό των σφαιριδίων που περιέχουν το επισημασμένο με βιοτίνη ενισχυμένο προϊόν. Ένας εκκινητής ειδικός για την πραγματοποίηση της αλληλούχησης προσδένεται στο DNA που είναι δεσμευμένο στα σφαιρίδια και τα σφαιρίδια φορτώνονται σε μια ειδική πλάκα "picotiter" που περιέχει εκατομμύρια ξεχωριστά πηγαδάκια. Κάθε ένα τέτοιο πηγαδάκι έχει περίπου ίση διάμετρο με τη διάμετρο ενός σφαιριδίου, ώστε να εξασφαλιστεί ότι ένα μόνο σφαιρίδιο (που θα φέρει ενισχυμένο ένα συγκεκριμένο τμήμα DNA της βιβλιοθήκης) θα βρίσκεται σε κάθε πηγαδάκι [29].

Το 454 GS FLX χρησιμοποιεί έναν τρόπο αλληλούχησης που βασίζεται στο πυροφωσφορικό (pyrosequencing). Έχοντας για πρώτη φορά γίνει περιγραφή το 1996 [30, 31], η πυροαλληλούχηση αξιοποιεί το μόριο πυροφωσφορικού (PPi) που παράγεται κατά την ενσωμάτωση κάθε νουκλεοτιδίου από τη δράση της DNA πολυμεράσης κατά το στάδιο της επιμήκυνσης (sequencing by synthesis) (Εικόνα 3, ①). Το μόριο του πυροφωσφορικού μετατρέπεται σε ATP μέσω της δράσης της ATP σουλφυριλάσης παρουσία APS (adenosine 5´ phosphosulfate) και το ATP στη συνέχεια χρησιμοποιείται από τη λουσιφεράση για τη μετατροπή της λουσιφερίνης σε οξυλουσιφερίνη [32]. Αυτή η αντίδραση παράγει φως, το οποίο μπορεί να ανιχνευθεί και να ποσοτικοποιηθεί μέσω μιας υπερευαίσθητης κάμερας που υπάρχει μέσα στον αλληλουχητή (Εικόνα 3, ②,③) [28]. Τα νουκλεοτίδια που δεν έχουν ενσωματωθεί και το ATP αποικοδομούνται μέσω της απυράσης (Εικόνα 3, ④) και η αντίδραση μπορεί να ξανά ξεκινήσει με άλλο νουκλεοτίδιο. Για μικρές επαναλήψεις της ίδιας βάσης, η ένταση του φωτός που εκπέμπεται είναι ανάλογη του αριθμού των νουκλεοτιδίων που ενσωματώνονται. Ωστόσο, για μεγαλύτερες επαναλήψεις (>8 νουκλεοτίδια), το σήμα ξεκινά να παρουσιάζει απώλεια της γραμμικότητας και ταυτόχρονα αυξάνεται ο ρυθμός εμφάνισης λαθών, κατά των προσδιορισμό των βάσεων που αλληλουχούνται (base call error) [29].

Ένα πλεονέκτημα που εμφανίζει το σύστημα 454 όταν συγκρίνεται με άλλα συστήματα είναι το μεγαλύτερο μήκος αναγνώσματος (read length) που παρουσιάζει (μέχρι 1000 ζεύγη βάσεων) και ο μικρότερος χρόνος που απαιτείται για τη διαδικασία της αλληλούχησης [15].



Εικόνα 3: Σχηματική απεικόνιση της διαδικασίας και των βασικών αρχών της πυροαλληλούχησης (pyrosequencing) (http://ars.els-cdn.com/content/image/3-s2.0-B9780123745378000080-gr1.jpg?httpAccept=%2A%2F%2A).

8

1.3.2. 5500 Solid (Sequencing by Oligo Ligation and Detection) – Thermo Fisher Scientific

Η τεχνική SOLiD, η οποία αναπτύχθηκε για πρώτη φορά από το εργαστήριο του George Church το 2005 [33], διαφέρει από τα άλλα εμπορικά διαθέσιμα συστήματα αλληλούχησης υψηλής απόδοσης. Στο σύστημα αυτό, η αλληλουχία καθορίζεται μέσω μιας μεθόδου (sequencing by ligation) που εκμεταλλεύεται τη σύνδεση ανιχνευτών (probe ligation method). Παρόμοια με την προσέγγιση της πυροαλληλούχησης 454 που αναφέρθηκε προηγουμένως, το πρώτο βήμα είναι η ενίσχυση των θραυσμάτων της βιβλιοθήκης που είναι δεσμευμένα σε σφαιρίδια, μέσω αλυσιδωτής αντίδρασης πολυμεράσης σε γαλάκτωμα (Emulsion PCR). Το 3' άκρο του DNA εκμαγείου είναι τροποποιημένο με τέτοιο τρόπο ώστε να επιτρέπει τη σύνδεση των DNA – σφαιριδίων σε γυάλινη επιφάνεια [34]. Έπειτα, κατάλληλος εκκινητής για την αλληλούχηση συμπληρωματικός στην αλληλουχία του αντάπτορα προσδένεται στο εκμαγείο DNA, παρέχοντας 5' φωσφορικό υπόστρωμα για τη δράση της DNA λιγάσης. πραγματοποιηθεί η αντίδραση Προκειμένου να της αλληλούχησης, χρησιμοποιούνται φθορίζοντες ολιγονουκλεοτιδικοί (οκταμερείς) ανιχνευτές με ικανότητα να υβριδοποιούνται στα πρώτα δυο νουκλεοτίδια του DNA εκμαγείου, ακριβώς δίπλα από τον εκκινητή που αναφέρθηκε προηγουμένως [29].

Οι ανιχνευτές είναι κατασκευασμένοι ώστε οι δύο πρώτες θέσεις να αντιπροσωπεύουν το καθένα από τους δεκαέξι πιθανούς συνδυασμούς δινουκλεοτιδίων. Οι υπόλοιπες έξι θέσεις του ανιχνευτή είναι εκφυλισμένες και το 5' άκρο του είναι σημασμένο με μία από τις τέσσερις φθορίζοντες ετικέτες (Εικόνα 4 και Εικόνα 5). Μετά την υβριδοποίηση, η DNA λιγάση συνδέει ομοιοπολικά τον ανιχνευτή με τον εκκινητή και καταγράφεται φθορισμός. Ο ανιχνευτής έχει διασπαστεί μεταξύ των θέσεων 5 και 6 και το 5' φωσφορικό αναγεννάται για να καταστεί δυνατή η επακόλουθη αντίδραση σύνδεσης. Επαναλαμβάνονται αρκετοί κύκλοι αυτών των αντιδράσεων υβριδοποίησης και σύνδεσης [35]. Έπειτα η νέο-σχηματισθείσα αλυσίδα αποδιατάσσεται από το DNA εκμαγείο και ένας νέος εκκινητής υβριδοποιείται στο εκμαγείο. Το πολύ σημαντικό όμως είναι, ότι ο νέος αυτός εκκινητής υβριδοποιείται κατά ένα νουκλεοτίδιο πιο κοντά στο 5' άκρο του εκμαγείου σε σχέση με τον αρχικό εκκινητή (n-1) (Εικόνα 5). Αυτή η διαδικασία επαναλαμβάνεται τρεις φορές ακόμα, φτάνοντας τελικά στο σημείο να έχουν χρησιμοποιηθεί πέντε διαφορετικοί εκκινητές (n, n-1, n-2, n-3, n-4) [29, 35].

Probe Anatomy



Εικόνα 4: Σχηματική απεικόνιση ολιγονουκλεοτιδικού (οκταμερούς) ανιχνευτή που φέρει φθορίζουσα χρωστική και χρησιμοποιείται για τις ανάγκες της αλληλούχησης με τη μέθοδο SOLiD (Sequencing by Oligo Ligation and Detection) (https://image.slidesharecdn.com/nextgenerationsequencing-170219062723/95/next-

generation-sequencing-15-638.jpg?cb=1487485817).

Ένα από τα πλεονεκτήματα αυτής της προσέγγισης για τη διεκπεραίωση της αλληλούχησης είναι ότι κάθε νουκλεοτίδιο στην αλληλουχία διαβάζεται δύο φορές. Έτσι κάθε συγκεκριμένο νουκλεοτίδιο του προς αλληλούχηση εκμαγείου μπορεί να δημιουργεί δύο ξεχωριστά σήματα φθορισμού ανάλογα την ταυτότητα του διπλανού νουκλεοτιδίου. Ο ρυθμός εμφάνισης ψευδώς θετικών αποτελεσμάτων κατά την ανίχνευση μετάλλαξης είναι μειωμένος, καθώς ο πολυμορφισμός ενός νουκλεοτιδίου (SNP), θα δημιουργήσει δυο χρωματικές αλλαγές όταν συγκριθεί με την αλληλουχία αναφοράς [29]. Ο αλληλουχητής SOLiD είναι ικανός να παράγει 30 έως 40Gb πληροφορίας στο τέλος της κάθε ημέρας [36].

Ένα σχετικό μηχάνημα με τον αλληλουχητή SOLiD, αναπτύχθηκε από το εργαστήριο του Church (Polonator G.007), το οποίο χρησιμοποιεί παρόμοια προσέγγιση για να εκτελεί την αλληλούχηση (sequencing-by-ligation). Έχει την ικανότητα να παράγει 8 έως 10Gb πληροφορίας με την ολοκλήρωση της

αλληλούχησης [28]. Η βασική διαφορά μεταξύ των συστημάτων Polonator και SOLiD είναι το μειωμένο κόστος του αλληλουχητή και το περιβάλλον ανοιχτού κώδικα του λογισμικού του και των πακέτων ανάλυσης [37].



Εικόνα 5: Σχηματική απεικόνιση της βασικής αρχής της αλληλούχησης με τη μέθοδο SOLiD

(http://bebatutedu.github.io/metagenomique_cm/images/sequencage/metzker2010_f igure3a.png).

1.3.3. Illumina – Solexa

Ο αλληλουχητής αυτός διαφέρει από τα συστήματα 454 και SOLiD που είδαμε προηγουμένως, καθώς εδώ το στάδιο της ενίσχυσης των θραυσμάτων της βιβλιοθήκης γίνεται (in situ) σε στερεή επιφάνεια και όχι σε κυστίδια μέσω αλυσιδωτής αντίδρασης πολυμεράσης σε γαλάκτωμα. Όμοια με τα άλλα συστήματα, η DNA βιβλιοθήκη συνδέεται με ολιγονουκλεοτιδικούς αντάπτορες (Εικόνα 6, 1). Οι αντάπτορες περιέχουν αλληλουχία συμπληρωματική με αυτή των αγκυροβολημένων ολιγονουκλεοτιδίων, τα οποία είναι δεσμευμένα στη στερεή επιφάνεια (flow cell). Μετά από την υβριδοποίηση με τα αγκυροβολημένα ολιγονουκλεοτίδια, το εκμαγείο DNA ενισχύεται μέσω ενός τύπου αλυσιδωτής αντίδρασης πολυμεράσης που ονομάζεται "bridge PCR" [38, 39], στην οποία τα μόρια DNA είναι ελεύθερα να καμφθούν και να σχηματίσουν μια "γέφυρα" με ένα παρακείμενο αγκυροβολημένο ολιγονουκλεοτίδιο (Εικόνα 6, 3). Αυτή η διαδικασία έχει ως αποτέλεσμα την παραγωγή περισσότερων από 50.000.000 ξεχωριστών συστάδων που περιέχουν πάνω από χίλια αντίγραφα ενισχυμένων μορίων DNA πάνω στην στερεή επιφάνεια (Εικόνα 6, 6). Έπειτα, οι συστάδες αποδιατάσσονται για να παρέχουν μονόκλωνο εκμαγείο και ένας ειδικός για την αλληλούχηση εκκινητής υβριδοποιείται στην αλυσίδα. Κατά τη διάρκεια κάθε κύκλου αλληλούχησης, οι ενισχυμένες συστάδες εκτίθενται στη δράση της DNA πολυμεράσης και σε ένα μείγμα τεσσάρων νουκλεοτιδίων (A, T, G, C) όπου το καθένα φέρει μια μοναδική φθορίζουσα ετικέτα. Τα νουκλεοτίδια έχουν χημικά αδρανές το 3'-ΟΗ άκρο τους, με αποτέλεσμα να εμποδίζουν την ενσωμάτωση και άλλων νουκλεοτιδίων, οδηγώντας στον τερματισμό της σύνθεσης [40]. Στο τέλος κάθε κύκλου, το φθορίζον σήμα που εκπέμπεται, με χρήση ακτίνας λέιζερ, καταγράφεται (έγχρωμες κηλίδες, τεσσάρων διαφορετικών χρωμάτων) για κάθε συστάδα. Η φθορίζουσα ετικέτα και η 3' - ΟΗ τροποποίηση στα νουκλεοτίδια, που οδηγεί στο τερματισμό της σύνθεσης, διασπώνται και απομακρύνονται [41], αναγεννώντας την αναπτυσσόμενη αλυσίδα για τον επόμενο κύκλο ενσωμάτωσης νουκλεοτιδίου [29].

Με τη χρήση αυτής της χημείας αναστρέψιμου τερματισμού, το 2005 με τον Illumina Genome Analyzer ήταν εφικτή η παραγωγή αναγνωσμάτων "reads" μήκους 35 ζευγών βάσεων με ακρίβεια που ξεπερνά το 99%, καθώς και η παραγωγή δεδομένων περίπου ενός δισεκατομμυρίου βάσεων κατά την



The enzyme incorporates nucleotides to build double-stranded bridges on the solidphase substrate. Denaturation leaves single-stranded templates anchored to the substrate.

Several million dense dusters of doublestranded DNA are generated in each channel of the flow cell.

Εικόνα 6: Σχηματική απεικόνιση της τεχνικής bridge PCR που χρησιμοποιείται από το σύστημα αλληλούχησης της illumina, για την ενίσχυση του εκμαγείου DNA (http://seqanswers.com/forums/images/content/ilmn-step1-6.jpg).

ολοκλήρωση της αλληλούχησης. Το 2014 ο αριθμός αυτός αυξήθηκε φτάνοντας το 1.8×10¹² ζεύγη βάσεων. Αξιοσημείωτο γεγονός είναι πως το πρώτο ανθρώπινο γονιδίωμα, που συνδημοσιεύτηκε στο Science και το Nature το 2001, χρειάστηκε δεκαπέντε χρόνια για να αλληλουχηθεί και κόστισε σχεδόν τρία δισεκατομμύρια

δολάρια. Σε αντίθεση, ο HiSeqX[™] Ten, που κυκλοφόρησε το 2014, μπορεί να αλληλουχήσει πάνω από 45 ανθρώπινα γονιδιώματα σε μία μόνο ημέρα, με κόστος περίπου 1000 δολάρια [42].

Με αυτή την προσέγγιση αλληλούχησης, η μεγαλύτερη πηγή σφάλματος είναι η ενσωμάτωση λάθος νουκλεοτιδίων, επίσης, η ημιτελής αφαίρεση είτε της φθορίζουσας ετικέτας, είτε της τροποποίησης τερματισμού, έχει ως αποτέλεσμα τον αυξημένο θόρυβο στο σήμα φθορισμού που λαμβάνεται από κάθε συστάδα, οδηγώντας σε μικρότερης αξιοπιστίας προσδιορισμό της βάσης που αλληλουχήθηκε, και οδηγώντας επίσης σε μεγαλύτερου μήκους αναγνώσματα.

Ωστόσο, οι βελτιώσεις που είναι σε συνεχή εξέλιξη και αφορούν τη χημεία της αλληλούχησης, τα συστήματα απεικόνισης, καθώς και τα προγράμματα ανάλυσης, θα μειώσουν την ύπαρξη τέτοιων θεμάτων και θα επιτρέψουν την αξιόπιστη αύξηση του μήκους των αναγνωσμάτων [43].

1.3.4. Ion Torrent – Thermo Fisher Scientific

Πρόκειται για το σύστημα αλληλούχησης που χρησιμοποιεί το εργαστήριο μας, το οποίο κατασκευάζεται από την εταιρία Thermo Fisher Scientific. Με το εν λόγω σύστημα πραγματοποιήθηκαν τα πειράματα αλληλούχησης νέας γενιάς για τις ανάγκες της παρούσας εργασίας.

Το Ion Torrent καταφέρνει να μετατρέψει, όπως θα δούμε πιο κάτω, τη χημική πληροφορία (A, C, G, T) σε ψηφιακή πληροφορία (0, 1) και αποτελεί το πρώτο σύστημα αλληλούχησης που δεν απαιτεί φθορισμό και σάρωση με χρήση κάμερας, με αποτέλεσμα να επιτυγχάνεται ταχύτερη αλληλούχηση, με χαμηλό κόστος και χρήση εξοπλισμού μικρότερου μεγέθους [44, 45]. Επίσης δεν απαιτούνται τροποποιημένα νουκλεοτίδια όπως συμβαίνει σε κάποια άλλα συστήματα αλληλούχησης. Στη διεθνή βιβλιογραφία η συγκεκριμένη μεθοδολογία αλληλούχησης συναντάται και με τους όρους "pH-mediated sequencing" ή "silicon sequencing" [46].

Η ανάλυση με τη χρήση της πλατφόρμας Ion Torrent περιλαμβάνει την ενίσχυση των θραυσμάτων της βιβλιοθήκης στα σφαιρίδια με χρήση emulsion PCR και την αλληλούχηση μέσω σύνθεσης (sequencing by synthesis, SBS) μέσα σε ένα κύκλωμα (chip). Σημειώνεται πως τα ενισχυμένα θραύσματα DNA ποσοτικοποιούνται και αν κριθεί απαραίτητο εμπλουτίζονται πριν από την αλληλούχηση [47].

Το σύστημα αλληλούχησης Ion Torrent μπορεί να ανιχνεύσει μεταβολές στο pH, μέσα σε κάθε ξεχωριστό πηγαδάκι (well), που οφείλονται στην απελευθέρωση ιόντος υδρογόνου (H⁺) ως παραπροϊόν της ενσωμάτωσης κάθε νουκλεοτιδίου στη συντιθέμενη αλυσίδα, από την DNA πολυμεράση. Αυτή η μεταβολή στο pH ανιχνεύεται από έναν υπερευαίσθητο αισθητήρα. Το κύκλωμα του Ion Torrent κατακλύζεται από τέσσερα διαλύματα που φέρουν A, C, G, T αντίστοιχα και με συγκεκριμένη σειρά, καθώς δεν υπάρχει κάποια ετικέτα που να μαρτυρά την ταυτότητα της κάθε βάσης [48]. Μετά από κάθε χρησιμοποίηση των διαλυμάτων A, C, G, T, πραγματοποιείται έκπλυση προκειμένου να εξαλειφθούν πλήρως τα υπολείμματα βάσης και να χρησιμοποιηθεί το επόμενο διάλυμα [49].

Εάν το νουκλεοτίδιο δεν είναι το σωστό τότε δε θα υπάρχει μεταβολή pH. Εάν όμως το διάλυμα φέρει το νουκλεοτίδιο που απαιτείται, αν για παράδειγμα η επόμενη βάση της αλυσίδας εκμαγείου είναι η θυμίνη (T) και χρησιμοποιηθεί διάλυμα που φέρει αδενίνη (A), τότε ο αισθητήρας θα ανιχνεύσει τη μεταβολή pH



Εικόνα 7: Σχηματική απεικόνιση της έκλυσης ιόντος – ιόντων υδρογόνου (Η+) κατά την ενσωμάτωση βάσης – βάσεων (πάνω) και καταγραφή διπλάσιας τάσης έπειτα από ενσωμάτωση δυο νουκλεοτιδίων (κάτω)

(https://www.thermofisher.com/gr/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html).

και το δυναμικό που θα παρατηρηθεί, εξαιτίας της ενσωμάτωσης της αδενίνης (Α) και της απελευθέρωσης Η⁺. Εάν υπάρξουν δυο νουκλεοτίδια να ενσωματωθούν, θα υπάρξει ανίχνευση διπλάσιας τάσης (Εικόνα 7) [44].

Η μεταβολή στο pH που ανιχνεύεται από τον αισθητήρα δεν βρίσκεται σε τέλεια αναλογία με τον αριθμό των νουκλεοτιδίων που ενσωματώνονται, οπότε στις περιπτώσεις πολλαπλής ενσωμάτωσης της ίδιας βάσης (homopolymeric template sequence) θα απελευθερωθούν ανάλογα περισσότερα ιόντα υδρογόνου (H⁺) που θα οδηγήσουν σε λάθη, ιδιαίτερα σημαντικά ιδίως όταν πρόκειται για αλληλουχία που χαρακτηρίζεται από προσθήκες ή ελλείψεις [48]. Έτσι, ο προσδιορισμός των ομοπολυμερών περιοχών πραγματοποιείται με περιορισμένη ακρίβεια.

1.3.5. Helicos True Single Molecule Sequencing (tSMS)

Το σύστημα Helicos tSMS αναπτύχθηκε για πρώτη φορά από τον Stephen Quake και την ερευνητική του ομάδα το 2003 [50], και χαρακτηρίζεται από την ικανότητά του για αλληλούχηση από μη ενισχυμένο DNA εκμαγείο. Κατά την προετοιμασία του δείγματος, το γενωμικό DNA τεμαχίζεται τυχαία ώστε να δημιουργήσει μικρά θραύσματα περίπου 100 έως 200 ζεύγη βάσεων. Έπειτα, πολλά μόρια αδενοσινών προσδένονται στο 3' άκρο των μορίων του εκμαγείου για να επιτρέψουν σε αυτά τα μόρια DNA να υβριδοποιηθούν στα αγκυροβολημένα poly – Τ ολιγονουκλεοτίδια τα οποία είναι ομοιοπολικά προσδεμένα στην στερεή επιφάνεια (flow cell surface) (Εικόνα 8) [29].

Η τελευταία αδενοσίνη φέρει φθορίζουσα ετικέτα, έτσι ο αλληλουχητής μπορεί να αναγνωρίσει τη θέση κάθε μορίου του εκμαγείου στη συστοιχία πριν την αλληλούχηση. Η αρχική φθορίζουσα ετικέτα διασπάται και απομακρύνεται και οι κύκλοι της αλληλούχησης ξεκινούν με έκθεση του εκμαγείου στην DNA πολυμεράση και σε ένα από τα τέσσερα σημασμένα με φθορίζουσα ετικέτα νουκλεοτίδια. Παρόμοια με την 454 προσέγγιση, η αλληλούχηση είναι ασύγχρονη διότι δεν ενσωματώνουν όλα τα μόρια του εκμαγείου ένα νουκλεοτίδιο κατά τη διάρκεια ενός συγκεκριμένου κύκλου αλληλούχησης. Μετά από κάθε κύκλο, το σήμα φθορισμού μετρείται από κάθε μόριο του εκμαγείου από ένα ευαίσθητο σύστημα ανίχνευσης φθορισμού. Μετά από εκατοντάδες κύκλους αλληλούχησης, ο αλληλουχητής της Helicos μπορεί να επιτύχει παραγωγή πληροφορίας που ξεπερνά τα 20 Gb και αναγνώσματα μήκους κατά μέσο όρο 30 βάσεων [51, 52].

Η προσέγγιση της Helicos προσπερνά το πρόβλημα της εμφάνισης σφαλμάτων που οφείλονται σε artifacts της αλυσιδωτής αντίδρασης πολυμεράσης (PCR), καθώς δεν υπάρχει κανένα στάδιο ενίσχυσης κατά τη διάρκεια προετοιμασίας του δείγματος. Όπως και στο σύστημα αλληλούχησης της 454, σφάλματα μπορούν να προκύψουν από γεγονότα πολλαπλής ενσωμάτωσης νουκλεοτιδίων κατά τη διάρκεια αλληλούχησης της ομοπολυμερών περιοχών. Η Helicos έχει εισάγει τροποποιημένα "virtual terminator" νουκλεοτίδια [53], τα οποία αποτρέπουν την συνεχόμενη προσθήκη νουκλεοτιδίων κατά την αλληλούχηση ομοπολυμερών περιοχών. Η ακρίβεια αυτής της τεχνικής είναι αρκετά μεγάλη (>99,99%), επειδή κυρίως το εκμαγείο μπορεί να αλληλουχηθεί δύο φορές (two-pass sequencing) [54].





(https://www.researchgate.net/profile/Haseeb_Khan/publication/44684171/figure /fig3/AS:203231801155598@1425465645630/Basic-workflow-of-Helicos-single-molecule-sequencing-method-modified-from-30.png).

1.3.6. Single molecule real time sequencing (SMRT) - Pacific Biosciences

Η αλληλούχηση SMRT (Single molecule real time sequencing) έχει αναπτυχθεί από την Pacific Biosciences και βασίζεται στην παρακολούθηση της δραστηριότητας της πολυμεράσης καθώς ενσωματώνει νουκλεοτίδια στην αλυσίδα DNA, τα οποία είναι επισημασμένα με διαφορετική ετικέτα [55]. Κάθε νουκλεοτίδιο φέρει στη φωσφορική του ομάδα, μια φθορίζουσα ετικέτα χαρακτηριστική για κάθε βάση, η οποία απελευθερώνεται κατά την ενσωμάτωση του νουκλεοτιδίου από την DNA πολυμεράση. Τα νουκλεοτίδια που ενσωματώνονται ανιχνεύονται σε πραγματικό χρόνο (real time) κατά τη σύνθεση της αλυσίδας [56]. Όλη η διαδικασία πραγματοποιείται σε μικρά πηγαδάκια με διαφανή πυθμένα τα οποία περιβάλλονται από τοιχώματα αλουμινίου και ονομάζονται ZMW (Zero-mode waveguide) [57]. Τα πηγαδάκια αυτά είναι εξαιρετικά μικρά με διάμετρο μόλις 70nm και βάθος 100nm [56].

Ενώ άλλες τεχνολογίες αλληλούχησης δεσμεύουν το DNA και επιτρέπουν στην πολυμεράση να "ταξιδεύει" κατά μήκος του εκμαγείου, η PacBio σταθεροποιεί την πολυμεράση στον πυθμένα των ZMWs (Εικόνα 9) [58], καθώς εκεί μπορεί να παρακολουθηθεί η δραστηριότητά της. Η ενσωμάτωση των dNTPs οπτικοποιείται μέσω ενός συστήματος που φέρει λέιζερ και κάμερα, το οποίο καταγράφει το χρώμα και τη διάρκεια του φωτός που εκπέμπεται, καθώς τα σημασμένα νουκλεοτίδια σταματούν προς στιγμή, κατά τη διάρκεια της ενσωμάτωσης στο κάτω μέρος των ZMWs (Εικόνα 9). Η πολυμεράση διασπά κατά τη διάρκεια της ενσωμάτωσης το φθοροφόρο που είναι δεσμευμένο στο νουκλεοτίδιο, επιτρέποντας την απομάκρυνση του από την περιοχή του αισθητήρα, πριν από την επόμενη ενσωμάτωση σημασμένου νουκλεοτιδίου [58]. Τώρα, το επόμενο νουκλεοτίδιο μπορεί να ενσωματωθεί και να καταγραφεί το σήμα που εκπέμπεται λόγω της φθορίζουσας ετικέτας που φέρει [56].

Το φάσμα εκπομπής δε βοηθά μόνο στην ταυτοποίηση της αλληλουχίας του DNA εκμαγείου, αλλά επίσης αποκαλύπτει πιθανές επιγενετικές τροποποιήσεις εξαιτίας των διακριτών διαφορών στα φάσματα εκπομπής [59]. Η διαδικασία της αλληλούχησης είναι αρκετά γρήγορη και διαρκεί περίπου 4 ώρες. Επίσης, τα αναγνώσματα (reads) που παράγονται είναι αρκετά μεγαλύτερα από εκείνα που προκύπτουν από αλληλουχητές δεύτερης γενιάς (πχ illumina, Ion Torrent) και είναι κατά μέσο όρο μεγαλύτερα από 10.000 ζεύγη βάσεων, περιλαμβάνοντας και αναγνώσματα 54.000 ζευγών βάσεων [60].



Εικόνα 9: Σχηματική απεικόνιση της βασικής αρχής της αλληλούχησης με τη μέθοδο SMRT (Single molecule real time sequencing) της Pacific Biosciences. Με πράσινο χρώμα παρατηρούμε μια φ29 DNA πολυμεράση, ακινητοποιημένη στον πυθμένα των ZMWs που περιβάλλονται από τοιχώματα αλουμινίου (Al), να ενσωματώνει σημασμένα νουκλεοτίδια στο DNA εκμαγείο που φαίνεται με μαύρο χρώμα. Ο φθορισμός καταγράφεται μόνο κατά τη διάρκεια της ενσωμάτωσης του σωστού νουκλεοτιδίου καθώς το φως διεισδύει μόνο στα κατώτατα 20-30nm του ZMW. Πολλαπλά λέιζερ διεγείρουν τα φθοροφόρα μόνο όταν εισέρχονται στην περιοχή κάτω από την διακεκομμένη κόκκινη γραμμή (zero-mode waveguide layer)

(http://www.nature.com/nbt/journal/v27/n2/images/nbt0209-150-F1.gif).

1.3.7. Nanopore DNA Sequencer - Oxford Nanopore Technologies

Η ιδέα της αλληλούχησης nanopore ειπώθηκε για πρώτη φορά το 1989 από τον David Deamer του πανεπιστημίου της Καλιφόρνιας (Santa Cruz, CA, USA). Επρόκειτο για μια πολλά υποσχόμενη ιδέα και πολλοί επιστήμονες περίμεναν με ενθουσιασμό, τη χρησιμοποίηση αυτής της μεθόδου σε εμπορική κλίμακα [61]. Το 2014 έγινε διαθέσιμος για πρώτη φορά ο αλληλουχητής MinION από την Oxford Nanopore Technologies (ONT), που σε αντίθεση με άλλους αλληλουχητές, οι αλληλουχητές nanopore δε χρησιμοποιούν τη σήμανση των νουκλεοτιδίων όπως είδαμε σε άλλα συστήματα. Εκεί που άλλα συστήματα χρησιμοποιούν δευτερεύοντα σήματα, φως, χρώμα ή pH, οι αλληλουχητές nanopore μπορούν να ανιχνεύσουν απευθείας την αλληλουχία DNA ενός μονόκλωνου μορίου DNA.

Η τεχνολογία αλληλούχησης με τη μέθοδο nanopore περιλαμβάνει τη χρήση πολύ λεπτής μεμβράνης που περιέχει πολύ μικρούς πόρους διαμέτρου περίπου 1,5 – 2nm [61]. Για να πραγματοποιηθεί η αλληλούχηση, το DNA περνά μέσω ενός πρωτεϊνικού πόρου (nanopore) (Εικόνα 10), καθώς ιόντα που διέρχονται μέσω αυτού δημιουργούν ρεύμα [62]. Καθώς το DNA διέρχεται μέσω

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Εικόνα 10: Σχηματική απεικόνιση της βασικής αρχής της αλληλούχησης με τη μέθοδο Nanopore. Η αλυσίδα του DNA περνά μέσω του πόρου (nanopore). Το ρεύμα μεταβάλλεται καθώς οι βάσεις Α, Τ, G και C περνούν μέσω του πόρου (https://d267cvn3rvuq91.cloudfront.net/i/legacy/nanopore_x616[1].jpg).

αυτής της πρωτεΐνης – πόρου παρατηρείται αλλαγή του ηλεκτρικού σήματος [63] διότι εμποδίζεται η κίνηση των ιόντων μέσω του πόρου μεταβάλλοντας έτσι την τάση (V). Οι αλλαγές στην τάση είναι χαρακτηριστικές για κάθε συγκεκριμένη αλληλουχία DNA που περνά διαμέσου του πόρου [58, 61]. Η χρονική περίοδος για την οποία το ρεύμα μεταβάλλεται, καθώς και το επίπεδο μείωσής του, είναι χαρακτηριστικά για κάθε βάση, επιτρέποντας έτσι τον καθορισμό της αλληλουχίας του DNA [64] [65]. Το ρεύμα μειώνεται, κατά το πέρασμα μιας βάσης μέσω του πόρου, σε ένα από τα τέσσερα επίπεδα, καθένα από τα οποία φανερώνει ποιο από τα τέσσερα dNMPs (A, T, G και C) πέρασε κάθε φορά μέσω του πόρου [66].

Περαιτέρω βελτιώσεις και τροποποιήσεις στην τεχνική, για παράδειγμα, αύξηση των παραμέτρων που μετρούνται κατά τη διάρκεια της μεταφοράς του DNA επιτρέποντας τον καλύτερο προσδιορισμό της κάθε βάσης, μπορούν να οδηγήσουν σε αύξηση της ακρίβειας της τεχνικής αλληλούχησης.

1.4. ΒΙΟΠΛΗΡΟΦΟΡΙΚΑ ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΑΛΛΗΛΟΥΧΗΣΗΣ ΝΕΑΣ ΓΕΝΙΑΣ

Τα πλήρη οφέλη από την αλληλούχηση νέας γενιάς δε θα είχαν επιτευχθεί εάν δεν ήταν εφικτή η βιοπληροφορική ανάλυση, στοίχιση (alignment), σύνδεση (assembly) κλπ., των πολλών μικρών αναγνωσμάτων (short reads), που παράγονται κατά την αλληλούχηση νέας γενιάς (NGS) [67]. Συνήθως δεκάδες ή εκατοντάδες Gbp μικρών αναγνωσμάτων μπορούν να παραχθούν κατά τη διάρκεια της αλληλούχησης νέας γενιάς. Το αποτέλεσμα είναι η δημιουργία κατά μέσο όρο terabytes δεδομένων, που καθιστούν την ανάλυση καθώς και τη διαχείριση των δεδομένων δύσκολη. Επίσης, υπάρχουν διαφορές μεταξύ των συστημάτων αλληλούχησης που αφορούν τη μορφή των δεδομένων, το μήκος των αναγνωσμάτων κλπ. [28].

Η σωστή στοίχιση καθιστά τα δεδομένα από την αλληλούχηση νέας γενιάς να έχουν αξία. Ωστόσο, ένας περιορισμός της στοίχισης των NGS αναγνωσμάτων είναι ότι ένα μεγάλο μέρος από αυτά δε μπορεί να στοιχηθεί μοναδικά σε σχέση με την αλληλουχία αναφοράς, καθώς τα αναγνώσματα είναι πολύ μικρά και η αλληλουχία αναφοράς αρκετά σύνθετη [68]. Επίσης, η πιθανότητα μοναδικής στοίχισης μειώνεται όχι μόνο από την παρουσία επαναλαμβανόμενων αλληλουχιών σε ένα σύνθετο γονιδίωμα, αλλά επίσης από ομολογίες σε στενά συνδεδεμένες οικογένειες γονιδίων και ψευδογονιδίων [68].

Συμβατικές λύσεις στοίχισης όπως το βασικό υπολογιστικό εργαλείο αναζήτησης τοπικών στοιχίσεων ή αλγόριθμος BLAST (basic local alignment search tool) [69] και το BLAT (BLAST-like alignment tool) είναι επαρκή για στοίχιση μεγάλων αναγνωσμάτων, όπως αυτά που παράγονται από την αλληλούχηση με τη μέθοδο Sanger, αλλά είναι ανεπαρκή για να διαχειριστούν τα μικρά αναγνώσματα που προκύπτουν από την αλληλούχηση νέας γενιάς [28].

Μέχρι σήμερα, έχει αναπτυχθεί ποικιλία αλγορίθμων και πακέτων προγραμμάτων προκειμένου να διαχειριστούν τα εκατομμύρια μικρά αναγνώσματα NGS. Για παράδειγμα, τα MAQ [70] και Bowtie [71] χρησιμοποιούνται ευρέως ως εργαλεία στοίχισης. Το Tophat [72] μπορεί να χρησιμοποιηθεί για τη στοίχιση των RNA-Seq αναγνωσμάτων, να δημιουργήσει μια εικόνα των ενώσεων (junctions) ή να τα στοιχήσει σε ένα γνωστό σετ από ενώσεις (set of junction). Το Cufflinks [73] και το Scripture [74] είναι χρήσιμα για τη συναρμολόγηση του μεταγραφώματος και την ανίχνευση διαφορετικά εκφραζόμενων γονιδίων.

1.5. E ϕ APMOGES THS AAAHAOYXHSHS NEAS GENIAS.

Η αλληλούχηση νέας γενιάς χρησιμοποιείται όλο και περισσότερο σε πολλά πεδία έρευνας, που περιλαμβάνουν: τη de novo αλληλούχηση βακτηριακών και ιικών γονιδιωμάτων [51, 75], την αναζήτηση γενετικών πολυμορφισμών με επανεξέταση (resequencing) ολόκληρου του γονιδιώματος ή συγκεκριμένων περιοχών του γονιδιώματος [76], την κατανόηση των γενετικών μηχανισμών που σχετίζονται με την ποικιλία έκφρασης των ανθρώπινων γονιδίων [77], τον χαρακτηρισμό του μεταγραφώματος των κυττάρων, των ιστών και των οργανισμών μέσω RNA-Seq [78], τη δημιουργία προφίλ των πρωτεϊνών που δεσμεύονται στο DNA (DNA – binding proteins) και επιγενετικών δεικτών μέσω ChIP-Seq [79].

1.5.1. Αλληλούχηση καρκινικών γονιδιωμάτων

Η πλειοψηφία των καρκίνων στον άνθρωπο προέρχεται από σποραδικούς όγκους που εμφανίζονται μέσω μιας διαδικασίας πολλαπλών σταδίων, στην οποία σωματικές, γενετικές και επιγενετικές αλλοιώσεις οδηγούν σε αλλαγές στην αλληλουχία γονιδίων, στη δομή και στην έκφραση [80]. Επομένως, η αναγνώριση όλων των γενετικών γεγονότων που οδηγούν ή μπορούν ενδεχομένως να οδηγήσουν στην ανάπτυξη καρκίνου, θα αναθεωρήσουν την προσέγγιση μας στους διάφορους τύπους καρκίνου που εμφανίζονται στον άνθρωπο, και θα επιταχύνουν την ανακάλυψη νέων προσεγγίσεων για την κλινική διάγνωση, την πρόληψη και τον προσδιορισμό εξατομικευμένης θεραπείας, ανάλογα με το καρκινικό γονιδίωμα του κάθε ασθενούς [81, 82].

Η αλληλούχηση νέας γενιάς μπορεί να προσαρμοστεί για εφαρμογές μοριακής διαγνωστικής, ώστε να αποκτηθούν πληροφορίες που αφορούν αλληλουχίες γονιδίων σχετικών με τον καρκίνο, οι οποίες είναι το κλειδί για τη λήψη περαιτέρω αποφάσεων – προσεγγίσεων. Οι τελευταίες εξελίξεις που αφορούν την τεχνολογία της αλληλούχησης νέας γενιάς έχουν κάνει εφικτή την αλληλούχηση μεγάλου αριθμού ολόκληρων καρκινικών γονιδιωμάτων επιτρέποντας το χαρακτηρισμό τους σε γενωμικό, μεταγραφικό και επιγενετικό επίπεδο. Παρότι το κόστος αλληλούχησης είναι ακόμα υψηλό, υπάρχει ένας αυξανόμενος αριθμός προσπαθειών με χρήση της αλληλούχησης νέας γενιάς [83, 84].

1.5.2. Αλληλούχηση εξονίων (exomes sequencing)

Οι περιοχές όλων των γονιδίων του ανθρώπινου γονιδιώματος που κωδικοποιούν πρωτεΐνη (whole exome), αποτελούν κατά προσέγγιση μόνο το 1-2% του ολόκληρου γονιδιώματος του ανθρώπου, ωστόσο περιέχουν το 85% των μεταλλάξεων του DNA, οι οποίες έχουν μεγάλο αντίκτυπο στις ασθένειες στον άνθρωπο [85]. Μπορεί να υπάρχουν και άλλες μεταλλαγές που να βρίσκονται σε άλλες περιοχές, αλλά δεν έχουν μελετηθεί σε τέτοιο βαθμό όσο εκείνες στην κωδική περιοχή. Έτσι, η αλληλούχηση των περιοχών που κωδικοποιούν για πρωτεΐνη, όχι μόνο μείωσαν το κόστος αλληλούχησης, αφού η μελέτη είναι εστιασμένη στα εξόνια, αλλά αποτελεί και μια προσέγγιση για ανακάλυψη περισσότερων μεταλλάξεων, οι οποίες εμπλέκονται σε σπάνιες και μη, ανθρώπινες ασθένειες [80].

Ο αριθμός των πρόσφατα ταυτοποιημένων γονιδίων που σχετίζονται με ασθένειες, έχει αυξηθεί εκθετικά τα τελευταία χρόνια, μέσω της εφαρμογής της αλληλούχησης των εξονίων σε όλα τα επίπεδα της ιατρικής [86], περιλαμβάνοντας νευρολογικές ασθένειες [87] όπως η επιληψία [88], νοητικές ασθένειες [89], ασθένεια του Parkinson [90, 91] και άλλους τύπους νευροεκφυλισμού [92], δυστονία [93], δυσκινησία [94] ή νευροπάθειες [95].

1.5.3. Αλληλούχηση RNA (RNA-Seq)

Προφίλ έκφρασης mRNA με τη χρήση τεχνολογιών microarray, έχουν χρησιμοποιηθεί ως βιοδείκτες για τη διάγνωση και τη πρόγνωση ασθενειών, καθώς και για την επιλογή φαρμακευτικής θεραπείας [96, 97]. Η απευθείας αλληλούχηση RNA/cDNA προσφέρει μια εναλλακτική προσέγγιση για υψηλής ανάλυσης μελέτη του μεταγραφώματος [98]. Η υψηλή ανάλυση της RNA αλληλούχησης, βελτίωσε εκτός των άλλων και την ανακάλυψη νέων μεταγράφων [99].

Πρόσφατες μελέτες που χρησιμοποίησαν RNA – Seq για να χαρακτηρίσουν πληθυσμούς RNA, παρείχαν πιο σύνθετες απεικονίσεις της έκφρασης και της ρύθμισης του RNA, μέσω εναλλακτικού ματίσματος (alternative splicing), εναλλακτικής πολυαδενυλίωσης (alternative polyadenylation) και επεξεργασίας του RNA (RNA editing) [100, 101]. Αυτά τα ευρήματα έχουν διευρύνει την παραδοσιακή μας αντίληψη για την έκταση και την πολυπλοκότητα της γονιδιακής έκφρασης και βελτίωσε τη γνώση μας για τους μηχανισμούς ρύθμισης της έκφρασης του RNA τόσο στα ευκαρυωτικά [102] όσο και τα προκαρυωτικά [103] γονιδιώματα.

1.5.4. Ανακάλυψη πληθυσμών RNA με Αλληλούχηση νέας γενιάς

Τα τελευταία χρόνια έχουμε διευρύνει τη γνώση μας για όλες σχεδόν τις πτυχές του RNA, μεταξύ των οποίων περιλαμβάνονται η αναγνώριση ότι το εναλλακτικό μάτισμα είναι ο κανόνας και όχι η εξαίρεση και η ανακάλυψη αμέτρητων μη κωδικών RNAs που παίζουν ρόλο σχεδόν σε όλες τις πτυχές της γονιδιακής ρύθμισης και της κυτταρικής λειτουργίας [104].

Με τη χρήση της αλληλούχησης νέας γενιάς έχει γίνει δυνατή η ανακάλυψη πληθυσμών RNA που ποτέ δε φανταζόμασταν ότι υπάρχουν, ενδεικτικά αναφέρονται τα: miRNAs [105], lncRNAs [106], lincRNAs [107], piRNAs [108], circRNA [109].

Τα μη κωδικά RNAs μπορούν να ρυθμίσουν τη σταθερότητα του mRNA, την αποτελεσματικότητα της μετάφρασης και τη μεταφορά της πρωτεΐνης. Η κατηγορία των μη κωδικών μορίων RNA περιλαμβάνει τα: rRNA (ribosomal RNA), snRNA (small nuclear RNA), tRNA (transfer RNA), microRNA, siRNA (small interfering RNA) και snoRNA (small nucleolar RNA). Μεταξύ αυτής της αναπτυσσόμενης λίστας μη κωδικών RNA υπάρχει μια ομάδα που σχετίζεται με ανθρώπινες ασθένειες: DGCR5, εμπλέκεται στο σύνδρομο DiGeorge [110], KkvLQTA-AS, το οποίο εμπλέκεται στο σύνδρομο Beckwith-Wiedeman [111], SCA8, παρεγκεφαλιδονωτιαία αταξία τύπου 8 [112] και CMPD - σχετιζόμενο RNA, εμπλέκεται στη Campomelic δυσπλασία [113].

1.6. ΕΝΑΛΛΑΚΤΙΚΟ ΜΑΤΙΣΜΑ (ALTERNATIVE SPLICING)

Το εναλλακτικό μάτισμα ή εναλλακτική συρραφή (alternative splicing) είναι ένας θεμελιώδης μοριακός μηχανισμός, ο οποίος σχετίζεται με τη ρύθμιση της έκφρασης των ευκαρυωτικών γονιδίων [114]. Τα περισσότερα ευκαρυωτικά γονίδια εκφράζονται ως πρόδρομα mRNAs (pre-mRNAs), τα οποία μετατρέπονται σε mRNA μέσω μιας διαδικασίας, που πραγματοποιείται με μεγάλη ακρίβεια, αυτής του ματίσματος του RNA (RNA splicing) δηλαδή μέσω απομάκρυνσης των εσονίων από το πρόδρομο mRNA [115]. Οι αλυσίδες premRNA που απομονώνονται από τους πυρήνες είναι μεγαλύτερες από τα τελικά επεξεργασμένα mRNA [116].

Κατά το εναλλακτικό μάτισμα, οι κωδικοποιητικές (εξόνια) και μη κωδικοποιητικές (εσόνια) περιοχές ενός γονιδίου συνδυάζονται διαφορετικά, εκφράζοντας πολλές συγγενείς πρωτεΐνες με διαφορετικές βιολογικές λειτουργίες. Η πολυπλοκότητα του πρωτεώματος κωδικοποιείται από έναν περιορισμένο αριθμό γονιδίων, χωρίς να απαιτούνται έτσι νέα γονίδια για κάθε πρωτεΐνη [117]. Με αυτόν τον τρόπο, εξόνια ενός γονιδίου μπορούν να συμπεριληφθούν εντός ή να αποκλειστούν από το τελικό επεξεργασμένο mRNA.

Παρατηρείται συχνά, η παραγωγή δύο μορφών της ίδιας πρωτεΐνης μέσω εναλλακτικού ματίσματος. Οι μορφές αυτές που παράγονται μπορεί να απαιτούνται σε διαφορετικά αναπτυξιακά στάδια ή σε διαφορετικούς τύπους κυττάρων, όπως για παράδειγμα οι ανοσοσφαιρίνες της τάξης IgM, που συναντώνται στην επιφάνεια των κυττάρων ως μεμβρανο-συνδεδεμένες πρωτεΐνες ή εκκρίνονται στο πλάσμα (διαλυτές πρωτεΐνες)[17].

Πρωτεϊνικές ισομορφές και mRNA μπορούν να διαφέρουν σε πολλές ιδιότητες όπως η δομή, η λειτουργία και ο εντοπισμός. Αυτοί οι ασυνήθιστοι μηχανισμοί συντελούν στην ποικιλία των πρωτεϊνών και έχει βρεθεί να σχετίζονται με ποικίλες ασθένειες, από τις οποίες δεν εξαιρείται ο καρκίνος [118-120].

Η επίδραση του εναλλακτικού ματίσματος δεν περιορίζεται μόνο στις πρωτεϊνικές ισομορφές, αλλά εκτείνεται στα μη κωδικά RNAs (non-coding RNAs, ncRNAs), τα οποία δεσμεύονται σε παράγοντες μεταγραφής, όπου οι διαφορετικές ισομορφές ncRNA που παράγονται από το εναλλακτικό μάτισμα, μπορούν να ρυθμίσουν τη δραστηριότητα ενός παράγοντα μεταγραφής [121, 122].

Ο αριθμός των μορφών mRNA που εκφράζονται στον άνθρωπο είναι μεγαλύτερος από τον αριθμό των γονιδίων. Η τεχνολογία της αλληλούχησης υψηλής απόδοσης έχει αποκαλύψει ότι παραπάνω από το 90% των ανθρώπινων γονιδίων υποβάλλονται σε εναλλακτικό μάτισμα, αυτό το ποσοστό είναι πολύ μεγαλύτερο από το αναμενόμενο [123]. Εξελικτικές μελέτες για τη ρύθμιση του εναλλακτικού ματίσματος δείχνουν ότι η προέλευση ενός εξονίου μπορεί να επηρεάσει το πόσο συχνά αυτό ματίζεται σε ένα mRNA [124].

1.6.1. Αλληλουχίες εντός του RNA καθοριστικές για το μάτισμα

Το μάτισμα λαμβάνει χώρα σε σημεία που ορίζονται από ειδικές αλληλουχίες νουκλεοτιδίων που βρίσκονται εντός των πρόδρομων mRNA και χαρακτηρίζουν τα όρια μεταξύ των εξονίων και εσονίων. Στην Εικόνα 11, το όριο εξονίου – εσονίου στο 5' άκρο του εσονίου χαρακτηρίζεται από μία αλληλουχία που ονομάζεται 5' θέση ματίσματος (5' splice site). Όμοια, στο 3' άκρο του εσονίου συναντάται η 3' θέση ματίσματος (3' splice site) [17].

Μία ακόμα σημαντική αλληλουχία απαραίτητη για το μάτισμα, είναι η θέση σημείου διακλάδωσης (branch point site), η οποία βρίσκεται εντός των εσονίων και ακολουθείται όπως μπορούμε να δούμε στην ίδια εικόνα, από περιοχή πολυπυριμιδίνης (περιοχή Py, polypyrimidine tract) και είναι συνήθως μήκους 15-20 ζευγών βάσεων. Βρίσκεται περίπου 5 -40 ζεύγη βάσεων πριν το 3' άκρο του προς μάτισμα εσονίου [125].

Επιπρόσθετα, παρατηρούνται συντηρημένες αλληλουχίες (conserved sequences) στην 5' θέση ματίσματος που είναι η GU, στην 3' θέση ματίσματος που είναι η AG, και η A (αδενίνη), η οποία περιέχεται στην αλληλουχία που ονομάζεται θέση διακλάδωσης (branch site). Αυτές οι συντηρημένες αλληλουχίες που αναφέρθηκαν, βρίσκονται όλες μέσα στο ίδιο εσόνιο. Έτσι η αλληλουχία ενός τυπικού εσονίου ξεκινά με το δινουκλεοτίδιο GU και ολοκληρώνεται με το δινουκλεοτίδιο AG [17, 126], δηλαδή οι θέσεις ματίσματος ακολουθούν τον κανόνα GT-AG (GT-AG rule) [127]. Στο σημείο αυτό να αναφερθεί, ότι το εναλλακτικό μάτισμα πραγματοποιείται με τη διαφορική χρήση θέσεων ματίσματος [126].



Εικόνα 11: Σχηματική απεικόνιση των εξονίων και εσονίου σε ένα pre-mRNA (https://upload.wikimedia.org/wikipedia/commons/a/a8/Intron_miguelferig.jpg).

1.6.2. Ο μηχανισμός του ματίσματος

Μια ενναλακτική λειτουργία σε επίπεδο πρωτεΐνης, μπορεί να επιτευχθεί μετά την αναγνώριση ενός εξονίου από τη μηχανή ματίσματος και την ένταξή του στο mRNA [128]. Η διαδικασία αναγνώρισης εξονίου και απομάκρυνσης των εσονίων πραγματοποιείται από ένα μεγάλο μακρομοριακό σύμπλεγμα, μεγέθους 40S στον σακχαρομύκητα και 60S στα θηλαστικά [129], που ονομάζεται σωμάτιο ματίσματος (spliceosome), ενώ το pre-mRNA συντίθεται από την RNA πολυμεράση ΙΙ στον πυρήνα. Η διαδικασία του ματίσματος παρατηρείται και σε μόρια RNA που δε φέρουν καλύπτρα, καθώς και σε μη πολυαδενυλιωμένα μόρια [126]. Μόνο οι ευκαρυωτικοί οργανισμοί έχουν σωμάτια ματίσματος και μάλιστα

κάποιοι οργανισμοί έχουν και δεύτερο σωμάτιο ματίσματος, που ονομάζεται δευτερεύον σωμάτιο ματίσματος (minor spliceosome) [130, 131], το οποίο αναγνωρίζει διαφορετικές θέσεις ματίσματος από εκείνες που αναγνωρίζονται από το κύριο σωμάτιο ματίσματος [132].

Το σωμάτιο ματίσματος είναι ένα ριβονουκλεοπρωτεϊνικό σύμπλοκο, που αποτελείται από τουλάχιστον 170 πρωτεΐνες και πέντε μικρές πυρηνικές



Nature Reviews | Neuroscience

Εικόνα 12: Σχηματική απεικόνιση του μηχανισμού ματίσματος http://www.nature.com/nrn/journal/v8/n11/i

mages/nrn2237-i1.jpg.

ριβονουκλεοπρωτεΐνες (small nuclear ribonucleoproteins, snRNPs) τις U1, U2, U4, U5 και U6. Τα ονόματα αυτά προκύπτουν καθώς τα snRNAs είναι πλούσια σε κατάλοιπα ουριδίνης (U) [133]. Το συγκρότημα του σωματίου ματίσματος αναγνωρίζει ειδικά τις θέσεις ματίσματος (Εικόνα 12). Συγκεκριμένα, 5' θέση η ματίσματος, η 3' θέση ματίσματος και θέση διακλάδωσης n αναγνωρίζονται ειδικά από το σωμάτιο ματίσματος. Η U1 snRNP συνδέεται στην αλληλουχία του εσονίου δίπλα από τη 5' θέση ματίσματος (5' SS) που είναι η 5' GU, εκμεταλλευόμενη τη ζεύξη των βάσεων ανάμεσα στο snRNA της και του pre-mRNA [132].

Το σύμπλοκο που δημιουργείται για πρώτη φορά κατά τη διάρκεια του ματίσματος είναι το σύμπλοκο Ε (Early presplicing complex) και περιέχει

τη U1 snRNP, το βοηθητικό παράγοντα U2AF (U2 snRNP Auxiliary Factor, U2AF) και μέλη της οικογένειας πρωτεϊνών SR (SR proteins) [126]. Η U2 snRNP, με τη

βοήθεια του U2AF, δεσμεύεται στη θέση διακλάδωσης και το σύμπλοκο Ε μετατρέπεται σε σύμπλοκο ματίσματος Α. Ο βοηθητικός παράγοντας U2AF αναγνωρίζει την πολυπυριμιδινική (Py) περιοχή και το δινουκλεοτίδιο AG της 3' θέση ματίσματος [134, 135]. Οι U4, U5 και U6 snRNPs συνδυάζονται και σχηματίζουν το τριμερές snRNP (tri-snRNP) σωμάτιο [136]. Το πρόδρομο σύμπλοκο Α μετατρέπεται σε σύμπλοκο ματίσματος Β. Το U6 αντικαθιστά στην 5' θέση ματίσματος το U1, το οποίο εγκαταλείπει το σύμπλοκο [132]. Το U4 απομακρύνεται από το σύμπλοκο, κάνοντας εφικτή την αλληλεπίδραση του U6 με το U2 και έτσι έχουμε μια διάταξη, που ονομάζεται σύμπλοκο C, η οποία δημιουργεί το ενεργό κέντρο. Με το σχηματισμό του ενεργού κέντρου έρχονται κοντά η 5' θέση ματίσματος του pre-mRNA και η θέση διακλάδωσης και πραγματοποιείται η πρώτη αντίδραση μετεστεροποίησης (transesterification). Η δεύτερη αντίδραση πραγματοποιείται μεταξύ των 5' και 3' θέσεων ματίσματος. Στο σημείο αυτό είναι σημαντικός ο ρόλος της U5 snRNP, η οποία ευνοεί να έρθουν σε επαφή τα δύο εξόνια [132, 134].

1.6.3. Βασικοί τύποι εναλλακτικού ματίσματος

Πολλαπλά μετάγραφα μπορούν να παραχθούν από τον ίδιο γονιδιακό τόπο μέσω της χρήσης εναλλακτικών υποκινητών, εναλλακτικών θέσεων πολυαδενυλίωσης και εναλλακτικού ματίσματος [137]. Η συσχέτιση ρυθμιστικών παραγόντων ματίσματος μπορεί να επηρεαστεί από τη χρήση συγκεκριμένων υποκινητών ή θέσεων πολυαδενυλίωσης και αυτό έχει άμεσες συνέπειες στην αναλογία των εναλλακτικώς συραμμένων μεταγράφων. Επιπροσθέτως, ο συνδυασμός διαφορετικών 5'-, 3'-άκρων και εξονίων μπορεί να καθορίσει τη σύνδεση παραγόντων και να επηρεάσει την τύχη του RNA, που περιλαμβάνει την αποτελεσματικότητα της μετάφρασης, το ποσοστό αποικοδόμησης ή τον εντοπισμό του [137].

Όλα τα γεγονότα εναλλακτικού ματίσματος μπορούν να κατηγοριοποιηθούν σε έναν από τους ακόλουθους γενικά αποδεκτούς τύπους. Ο πρώτος τύπος είναι η παράλειψη εξονίου (exon skipping) ή εξονική κασέτα, στον οποίο ένα εξόνιο παραλείπεται από το μετάγραφο (Εικόνα 13, a). Παρότι είναι

ένας σπάνιος τύπος στους κατώτερους ευκαρυωτικούς οργανισμούς, είναι ο πιο συνήθης τύπος στους ανώτερους ευκαρυωτικούς [114, 138, 139].

Στο δεύτερο και τρίτο τύπο εναλλακτικού ματίσματος, συγκεκριμένα εξόνια κόβονται σε συγκεκριμένο αριθμό ζευγών βάσεων στη 3' θέση ματίσματος (alternative 3' SS selection) ή 5' θέση ματίσματος (alternative 5' SS selection) (Εικόνα 13, b & c). Αυτά τα εξόνια από τη μία πλευρά έχουν μια βασική θέση ματίσματος και στην άλλη πλευρά δύο ή περισσότερες ανταγωνιστικές εναλλακτικές θέσεις ματίσματος, καταλήγοντας σε μια εναλλακτική περιοχή που είτε περιλαμβάνεται στο μετάγραφο είτε όχι [140].

Ο τέταρτος τύπος είναι η παραμονή εσονίου (intron retention), στον οποίο ένα εσόνιο παραμένει στο ώριμο μετάγραφο mRNA (Εικόνα 13, d). Εναλλακτικά μετάγραφα μπορούν να προκύψουν και από λιγότερο συχνά και πολύπλοκα γεγονότα όπως: αμοιβαίως αποκλειόμενα εξόνια (mutually exclusive exons), χρήση εναλλακτικού υποκινητή (alternative promoter) και εναλλακτική πολυαδενυλίωση (alternative poly-A) [118, 141, 142]. Ένας άλλος σπάνιος τύπος εναλλακτικού ματίσματος περιλαμβάνει αντιδράσεις μεταξύ δυο πρωτογενών μεταγράφων in trans [122].

Στα αμοιβαίως αποκλειόμενα εξόνια (Εικόνα 13, e), ένα από τα δύο εξόνια παραμένει στο mRNA μετά το εναλλακτικό μάτισμα αλλά όχι και τα δύο. Στα γεγονότα χρησιμοποίησης εναλλακτικού υποκινητή (Εικόνα 13, f), η μεταγραφή ξεκινά από διαφορετικά σημεία παράγοντας μετάγραφα με διαφορετικό 5'. Στην εναλλακτική πολυαδενυλίωση (Εικόνα 13, g), μπορούν να παραχθούν μετάγραφα με διαφορετικό 3' άκρο. Οι δύο τελευταίοι μηχανισμοί συναντώνται σε συνδυασμό κατά το εναλλακτικό μάτισμα και παρέχουν επιπλέον ποικιλία στα mRNA που προέρχονται από ένα γονίδιο [118, 143].


Nature Reviews | Genetics

Εικόνα 13: Βασικοί τύποι γεγονότων εναλλακτικού ματίσματος. Με μπλε χρώμα απεικονίζονται τα εξόνια ενώ οι περιοχές εναλλακτικού ματίσματος απεικονίζονται με μωβ χρώμα. Τα εσόνια αντιπροσωπεύονται με συνεχείς γραμμές ενώ οι διακεκομμένες γραμμές παρουσιάζουν τις επιλογές ματίσματος

(http://www.nature.com/nrg/journal/v11/n5/images/nrg2776-i1.jpg).

1.6.4 Ο ρόλος των πρωτεϊνών SR

Η αναγνώριση των θέσεων ματίσματος υποβοηθείται από ρυθμιστικές πρωτεΐνες, οι οποίες προσδένονται σε ειδικές αλληλουχίες εξονίων ή εσονίων διευκολύνοντας ή αποτρέποντας με αυτό τον τρόπο την αναγνώριση του εξονίου. Όταν η αλληλουχία που ευνοεί την επιλογή μιας θέσης ματίσματος βρίσκεται μέσα σε ένα εξόνιο ονομάζεται εξονικός ενισχυτής ματίσματος (Exonic Splicing Enhancer, ESE) ενώ όταν βρίσκεται μέσα σε εσόνιο ονομάζεται εσονικός ενισχυτής ματίσματος (Intronic Splicing Enhancer, ISE) [144].

Μια αλληλουχία που βρίσκεται σε ένα εξόνιο και καταστέλλει την επιλογή μιας θέσης ματίσματος ονομάζεται εξονικός αποσιωπητής ματίσματος (Exonic Splicing Silencer, ESS), ενώ μια αλληλουχία που βρίσκεται σε εσόνιο και προκαλεί καταστολή στην επιλογή μιας θέσης ματίσματος ονομάζεται εσονικός αποσιωπητής ματίσματος (Intronic Splicing Silencer, ISS) [143, 144].

Για την αναγνώριση των εξονικών ενισχυτών ματίσματος (ESE) σημαντικός είναι ο ρόλος των πρωτεϊνών SR (splicing regulatory proteins, SR proteins) [145]. Οι πρωτεΐνες SR διαθέτουν στο Ν-τελικό άκρο τους ένα ή δύο μοτίβα αναγνώρισης RNA (RNA recognition motifs, RRM) και στο C-τελικό άκρο τους διαθέτουν επικράτεια πλούσια σε αργινίνη/σερίνη, η οποία ονομάζεται περιοχή RS [146]. Η περιοχή RS μεσολαβεί μεταξύ των αλληλεπιδράσεων της πρωτεΐνης SR και των πρωτεϊνών της μηχανής ματίσματος, οδηγώντας τη μηχανή σε μια θέση ματίσματος που βρίσκεται πλησίον. Συγκεκριμένα, οι πρωτεΐνες SR προσελκύουν τη U1 snRNP στην 5' θέση ματίσματος [147] και τον U2AF στην 3' θέση ματίσματος (Εικόνα 14). Η αναστρέψιμη φωσφορυλίωση των πρωτεϊνών SR, κυρίως στα κατάλοιπα σερίνης των περιοχών RS, μπορούν να επηρεάσουν τις αλληλεπιδράσεις πρωτεΐνης – RNA [148] και πρωτεΐνης – πρωτεΐνης [149].

Σε αντίθεση με τις πρωτεΐνες SR, μέλη της οικογένειας των ετερογενών πυρηνικών ριβονουκλεοπρωτεϊνών (heterogeneous nuclear ribonucleoprotein, hnRNP) μπορούν να ανταγωνιστούν τη θετική επίδραση των πρωτεϊνών SR, με τη δέσμευσή τους σε ESS ή ISS καταστέλλοντας τη χρήση συγκεκριμένων θέσεων συρραφής [150].

Η σημαντικότητα των πρωτεϊνών SR δεν άπτεται μόνο στην αποτελεσματικότητα του ματίσματος, αλλά και στη ρύθμιση του εναλλακτικού ματίσματος [132]. Έχει δειχθεί ότι, οι πρωτεΐνες SR εκφράζονται διαφορικά σε διάφορους ιστούς και εμπλέκονται στην παθοβιολογία του καρκίνου μέσω της συμμετοχής τους σε γεγονότα εναλλακτικού ματίσματος [151].



Εικόνα 14: Σχηματική απεικόνιση της ρύθμισης του εναλλακτικού ματίσματος. Με την προσέλκυση της U1 snRNP στην 5' θέση ματίσματος και του U2AF στην 3' θέση ματίσματος, ξεκινά η συγκρότηση της μηχανής ματίσματος στις σωστές θέσεις (http://h-invitational.jp/h-dbas/document/am_fig1b.png).

Εκτός από την κλασική κατηγορία των πρωτεϊνών SR, υπάρχει μια άλλη ομάδα πλούσιων σε αργινίνη/σερίνη πρωτεϊνών, οι SR σχετιζόμενες πρωτεΐνες (SR-related proteins) που ανήκουν στην ίδια υπεροικογένεια. Αυτή η ομάδα πρωτεϊνών χαρακτηρίζεται από επαναλαμβανόμενες αλληλουχίες καταλοίπων σερίνης/αργινίνης, όπως οι τυπικές πρωτεΐνες SR, αλλά μερικές από αυτές στερούνται το χαρακτηριστικό μοτίβο αναγνώρισης RNA (RRM). Ένας αριθμός λειτουργιών έχει αποδοθεί σε αυτούς τους παράγοντες. Μερικοί από αυτούς εμπλέκονται στη διαδικασία ματίσματος, όπως οι: U1-70K [152], SRm160 [153], και Sip1 [154], ενώ άλλοι εμπλέκονται στην 3' επεξεργασία και μεταφορά του mRNA όπως ο Np13p [155].

Μια ειδική ομάδα SR σχετιζόμενων πρωτεϊνών που ονομάζονται SCAFs (SR-related CTD-associated factors), χαρακτηρίζονται από μια περιοχή που αλληλεπιδρά με την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης II (RNAP II). Για το λόγο αυτό, έχει προταθεί ότι οι πρωτεΐνες αυτές σχετίζονται με το μεταγραφικό μηχανισμό, συνδέοντας τη μεταγραφή με το μάτισμα, διαδικασίες που εκτελούνται σε στενή εγγύτητα [156].

1.7. ΓΟΝΙΔΙΟ *SCAF1*

Το γονίδιο SCAF1 έχει ανακαλυφθεί και κλωνοποιηθεί από τον ερευνητή Δρ. Σκορίλα Α. και τους συνεργάτες του, το 2001 στο Τορόντο του Καναδά [157]. Το πρώτο όνομα του γονιδίου ήταν SR-A1, το οποίο άλλαξε όμως σε SCAF1, προς αποφυγή σύγχυσης με το SRA1 που αφορά άλλο γονίδιο (steroid receptor RNA activator 1) [158]. Το ανθρώπινο γονίδιο SCAF1 βρίσκεται στη χρωμοσωμική περιοχή 19q13.3 ανάμεσα στα γονίδια RRAS και IRF3 και κοντά στα γονίδια BCL2L12, TSKS και των γονιδίων της οικογένειας των Καλλικρεϊνών (Εικόνα 16) [157].

Μόνο ένα πειραματικά επικυρωμένο μετάγραφο του γονιδίου SCAF1 στον άνθρωπο και συγκεκριμένα το SCAF1 v.1 (GenBank[®] accession number: NM_021228.2) είναι επί του παρόντος γνωστό. Η κωδικοποιητική αλληλουχία του SCAF1 v.1 αποτελείται από έντεκα εξόνια και κωδικοποιεί τη SCAF1 πρωτεΐνη που αποτελείται από 1312 αμινοξέα [157]. Εκτός από το γνωστό SCAF1 v.1, έχουν προβλεφθεί τρία ακόμη μετάγραφα, συγκεκριμένα τα SCAF1 X1, X2 και X3 (GenBank[®] accession numbers: XM_011527194.2, XM_005259122.4 και XM_017027083.1, αντίστοιχα). Όλα αυτά τα προβλεπόμενα μετάγραφα φέρουν ανοιχτό πλαίσιο ανάγνωσης (ORF) και επομένως αναμένεται να κωδικοποιούν ισομορφές του SCAF1. Επιπλέον, κάθε προβλεπόμενο μετάγραφο έχει ξεχωριστή 5' UTR, αλλά και τα τρία περιέχουν την ίδια 3' UTR. Αυτή η προβλεφθείσα 3' UTR αποτελείται από 243 νουκλεοτίδια και είναι περικομμένη στο 3' άκρο της κατά 7 νουκλεοτίδια, σε σύγκριση με τη γνωστή, η οποία εκτείνεται σε περιοχή 250 νουκλεοτίδίων (Εικόνα 15).

Η έκφραση του γονιδίου *SCAF1* έχει εξεταστεί σε μια σειρά ανθρώπινων ιστών και έχει βρεθεί πως υπερεκφράζεται σε ένα υποσύνολο ωοθηκικών καρκίνων που είναι κλινικά πιο επιθετικοί [157]. Η εκφραζόμενη πρωτεΐνη χαρακτηρίζεται από την περιοχή αλληλεπίδρασης με την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης ΙΙ (RNAP II) και την SR περιοχή [156]. Η μεγάλη υπομονάδα της RNA πολυμεράσης ΙΙ περιέχει στην καρβοξυτελική περιοχή της επαναλαμβανόμενα μοτίβα του επταμερούς Tyr-Ser-Pro-Thr-Ser-Pro-Ser [159], ο αριθμός των οποίων ποικίλει μεταξύ των οργανισμών και κυμαίνεται από 26 με 27 στο ζυμομύκητα μέχρι 52 σε όλα τα θηλαστικά. Ο αριθμός αυτών των επαναλήψεων και συνεπώς το μήκος της CTD είναι σημαντικά για την αποτελεσματική επεξεργασία του pre-mRNA αντισταθμίζοντας την αρνητική επίδραση των ESS [160].

Το γονίδιο SCAF1 εκφράζεται ευρέως σε πολλούς φυσιολογικούς ιστούς, ωστόσο τα επίπεδα mRNA ποικίλουν αρκετά. Τα υψηλότερα επίπεδα των μεταγράφων του SCAF1 έχουν εντοπιστεί στον εμβρυικό εγκέφαλο και ήπαρ και τα χαμηλότερα στους σιελογόνους αδένες, το δέρμα, την καρδιά, τη μήτρα και τις ωοθήκες. Στο μαστικό και προστατικό αδένα, τα μετάγραφα mRNA του SCAF1 γονιδίου εμφανίζονται σε σχετικά υψηλά επίπεδα. Τα επίπεδα mRNA του γονιδίου SCAF1 φαίνονται να αυξάνουν σε καρκινικές σειρές κυττάρων που επιδρούν διάφορες στεροειδείς ορμόνες, συμπεριλαμβανομένων των οιστρογόνων, των ανδρογόνων και των γλυκοκορτικοειδών, και σε μικρότερο βαθμό με τις προγεστερόνες [157].



Εικόνα 15: Σχηματική απεικόνιση της δομής του γνωστού και όλων των προβλεπόμενων SCAF1 μεταγράφων. Τα εξόνια απεικονίζονται ως ορθογώνια τμήματα και τα εσόνια ως γραμμές. Τα γκρι και τα λευκά ορθογώνια τμήματα αντιπροσωπεύουν κωδικά και μη κωδικά εξόνια αντίστοιχα. Οι αριθμοί εντός αυτών των σχημάτων και πάνω από τις γραμμές δείχνουν το μήκος του κάθε εξονίου ή εσονίου σε νουκλεοτίδια. Τα βέλη δείχνουν τη θέση του κωδικονίου ATG και οι αστερίσκοι τη θέση του κωδικονίου λήξης. Η ανάλυση έκφρασης για το γονίδιο *SCAF1* έχει δείξει ότι μπορεί να θεωρηθεί ως ένας νέος δυσμενής προγνωστικός δείκτης, για τον ωοθηκικό καρκίνο [157] και τον καρκίνο του μαστού [161]. Η έκφραση του γονιδίου *SCAF1* σε καρκινικούς μαστικούς ιστούς, επηρεάζεται από το μέγεθος του όγκου και από την ύπαρξη μετάστασης στους λεμφαδένες. Επιπρόσθετα, η υψηλή έκφραση του *SCAF1* είναι ένας σημαντικός ανεξάρτητος προγνωστικός δείκτης επιβίωσης ελευθέρας νόσου (disease-free survival, DFS) ενώ η χαμηλή έκφραση mRNA του γονιδίου σχετίζεται με μακράς διάρκειας ελευθέρας νόσου επιβίωση (long DFS) και ολικής επιβίωσης (OS) [157].



Εικόνα 16: Το γονίδιο SCAF1 βρίσκεται στη χρωμοσωμική περιοχή 19q13.3 ανάμεσα στα γονίδια RRAS και IRF3. Αποτελείται από 11 εξόνια και 10 εσόνια. Τα εξόνια φαίνονται ως κουτιά ενώ τα εσόνια ως συνδετικές γραμμές

(http://atlasgeneticsoncology.org/Genes/png/SCAF1.png).

Όσον αφορά την έκφραση του γονιδίου SCAF1 στον καρκίνο των ωοθηκών, σχετίζεται θετικά με το στάδιο της νόσου και το μέγεθος του όγκου. Επιπρόσθετα, τα υψηλά επίπεδα έκφραση του SCAF1 είναι ένας σημαντικός ανεξάρτητος προγνωστικός δείκτης ολικής επιβίωσης, και τα χαμηλά επίπεδα έκφρασης του γονιδίου σχετίζονται με μακράς διάρκειας ελευθέρας νόσου επιβίωση (long DFS) και ολικής επιβίωσης (OS) [157]. Επίσης, τα επίπεδα έκφρασης του mRNA φαίνεται να σχετίζονται με την πρόοδο του καρκίνου του παχέος εντέρου, καθώς η έκφραση είναι υψηλή στα αρχικά στάδια της ογκογένεσης και μειώνεται καθώς ο όγκος εξελίσσεται [162].

Μεταβολές στα επίπεδα έκφρασης mRNA του γονιδίου SCAF1 έχουν παρατηρηθεί στην κυτταρική σειρά HL-60 με οξεία προμυελοκυτταρική λευχαιμία στον άνθρωπο, μετά από αγωγή με σισπλατίνη και βλεομυκίνη. Τα επίπεδα mRNA του SCAF1 διαμορφώνονται και στις δύο περιπτώσεις ως απόκριση στην απόπτωση που επάγεται από κάθε φάρμακο, και συγκεκριμένα με αυξημένη ρύθμιση στην απόπτωση που προκαλείται από τη βλεομυκίνη και μειωμένη ρύθμιση στην απόπτωση που προκαλείται από σισπλατίνη σε κύτταρα HL-60. Αυτή η διαφορική απόκριση των επιπέδων mRNA του SCAF1 στην απόπτωση που προκαλείται από κάθε φάρμακο, μπορεί να οφείλεται σε διακριτά αποπτωτικά μονοπάτια και συνεπώς σε διακριτές κυτταρικές ανάγκες για μετάγραφα συγκεκριμένων γονιδίων [163].

Η πρωτεΐνη SCAF1 αποτελείται από 1312 αμινοξέα και έχει μοριακό βάρος 139.1 kDa. Η ανωτέρω πρωτεΐνη, η οποία ανήκει στην υπεροικογένεια των πρωτεϊνών SR, περιέχει μια πλούσια σε αργινίνη/σερίνη περιοχή (SR) καθώς και περιοχή καρβοξυτελικής δέσμευσης (CTD-binding domain), παρούσα μόνο σε ένα υποσύνολο των πλούσιων σε Arg/Ser παραγόντων ματίσματος. Μέσω των αλληλεπιδράσεων με το pre-mRNA και την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης ΙΙ, έχει δειχθεί ότι οι πλούσιες σε Arg/Ser πρωτεΐνες ρυθμίζουν το εναλλακτικό μάτισμα.

Στην πρωτεΐνη SCAF1 έχουν εντοπιστεί δύο αρνητικά φορτισμένες περιοχές γλουταμικού οξέος και ένα μοτίβο πλούσιο σε Arg/Asp (Εικόνα 17). Το μοτίβο αυτό παρουσιάζεται επίσης σε άλλες RNA δεσμευτικές πρωτεΐνες όπως η U1-70K και η RD RNA δεσμευτική πρωτεΐνη. Έχει προβλεφθεί πως πρόκειται για πυρηνική πρωτεΐνη, χωρίς διαμεμβρανικές περιοχές.

Έχουν ταυτοποιηθεί διάφορες πιθανές θέσεις μετα-μεταφραστικής τροποποίησης, συμπεριλαμβανομένων πολυάριθμων πιθανών θέσεων είτε για Οείτε για Ν-γλυκοζυλίωση, και αρκετές πιθανές θέσεις φωσφορυλίωσης από cAMP εξαρτώμενη πρωτεϊνική κινάση Α (PKA), πρωτεϊνική κινάση C (PKC) και κινάση καζεΐνης 2 [157].

37

1	MEEEDESRGK	TEESGEDRGD	GPPDRDPTLS	PSAFILRAIQ	QAVGSSLQGD	LPNDKDGSRC
61	HGLRWRRCRS	PRSEPRSQES	GGTDTATVLD	MATDSFLAGL	VSVLDPPDTW	VPSRLDLRPG
121	ESEDMLELVA	EVRIGDRDPI	PLPVPSLLPR	LRAWRTGKTV	SPQSNSSRPT	CARHLTLGTG
181	DGGPAPPPAP	SSASSSPSPS	PSSSSPSPPP	PPPPPAPPAP	PAPRFDIYDP	FHPTDEAYSP
241	PPAPEQKYDP	FEPTGSNPSS	SAGTPSPEEE	EEEEEEEEE	EEDEEEEGL	SQSISRISET
301	LAGIYDDNSL	SQDFPGDESP	RPDAQPTQPT	PAPGTPPQVD	STRADGAMRR	RVFVVGTEAE
361	ACREGKVSVE	VVTAGGAALP	PPLLPPGDSE	IEEGEIVQPE	EEPRLALSLF	RPGGRAARPT
421	PAASATPTAQ	PLPQPPAPRA	PEGDDFLSLH	AESDGEGALQ	VDLGEPAPAP	PAADSRWGGL
481	DLRRKILTOR	RERYRORSPS	PAPAPAPAAA	AGPPTRKKSR	RERKRSGEAK	EAASSSSGTQ
541	PAPPAPASPW	DSKKHRSRDR	KPGSHASSSA	RRRSRSRSRS	RSTRRRSRST	DRRRGGSRRS
601	RSREKRRRR	RSASPPPATS	SSSSSRRERH	RGKHRDGGGS	KKKKKRSRSR	GEKRSGDGSE
661	KAPAPAPPPS	GSTSCGDRDS	RRRGAVPPSI	QDLTDHDLFA	IKRTITVGRL	DKSDPRGPSP
721	APASSPKREV	LYDSEGLSGE	ERGGKSSQKD	RRRSGAASSS	SSSREKGSRR	KALDGG <u>DRDR</u>
781	DRDRDRDRDR	SSKKARPPKE	SAPSSGPPPK	PPVSSGSGSS	SSSSSCSSRK	VKLQSKVAVL
841	IREGVSSTTP	AKDAASAGLG	SIGVKFSRDR	ESRSPFLKPD	ERAPTEMAKA	APGSTKPKKT
901	KVKAKAGAKK	TKGTKGKTKP	SKTRKKVRSG	GGSGGSGGQV	SLKKSKADSC	SQAAGTKGAE
961	ETSWSGEERA	AKVPSTPPPK	AAPPPPALTP	DSQTVDSSCK	TPEVSFLPEE	ATEEAGVRGG
1021	AEEEEEEEE	EEEEEEEQ	QPATTTATST	AAAAPSTAPS	AGSTAGDSGA	EDGPASRVSQ
1081	LPTLPPPMPW	NLPAGVDCTT	SGVLALTALL	FKMEEANLAS	RAKAQELIQA	TNQILSHRKP
1141	PSSLGMTPAP	VPTSLGLPPG	PSSYLLPGSL	PLGGCGSTPP	TPTGLAATSD	KREGSSSSEG
1201	rgdtdkyl <u>kk</u>	LHTQERAVEE	VKLAIKPYYO	KKDITKEEYK	DILRKAVHKI	CHSKSGEINP
1261	VKVSNLVRAY	VORYRYFRKH	GRKPGDPPGP	PRPPKEPGPP	DKGGPGLPLP	PL

Εικόνα 17: Σχηματική παρουσίαση της αμινοξικής ακολουθίας της πρωτεΐνης SCAF1 . Η πλούσια σε Arg/Ser περιοχή φαίνεται με έντονη γραφή και υπογράμμιση, η καρβοξυτελική περιοχή δέσμευσης (CTD-binding domain) είναι διπλά υπογραμμισμένη. Επιπροσθέτως, η πρωτεΐνη SCAF1 χαρακτηρίζεται από δύο αρνητικά φορτισμένες περιοχές γλουταμικού οξέος που φαίνονται με διακεκομμένη υπογράμμιση, και ένα πλούσιο σε Arg/Asp μοτίβο, το οποίο φέρει τυπική υπογράμμιση. Διάφορες πιθανές θέσεις μετα-μεταφραστικής τροποποίησης έχουν επίσης ταυτοποιηθεί, συμπεριλαμβανομένων πολυάριθμων πιθανών θέσεων είτε για Ο- είτε Ν-γλυκοζυλίωση και αρκετές πιθανές θέσεις φωσφορυλίωσης από cAMP εξαρτώμενη πρωτεϊνική κινάση (ΡΚΑ), πρωτεϊνική κινάση C (ΡΚC) και κινάση καζεΐνης 2

(http://atlasgeneticsoncology.org/Genes/SCAF1ID46074ch19q13.html).

ΣΚΟΠΟΣ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Σκοπός της παρούσας διπλωματικής εργασίας ήταν η εξαγωγή και ανάλυση αποτελεσμάτων της αλληλούχησης νέας γενιάς που πραγματοποιήθηκε, σε διαφορετικές καρκινικές κυτταρικές σειρές για το γονίδιο *SCAF1* στην πλατφόρμα Ion Semiconductor της εταιρείας Thermo Fisher Scientific του τομέα Βιοχημείας και Μοριακής Βιολογίας του τμήματος Βιολογίας του Πανεπιστημίου Αθηνών με επιστημονικό υπεύθυνο τον Καθηγητή κ. Ανδρέα Σκορίλα.

Το γονίδιο SCAF1 εκφράζεται ευρέως σε πολλούς φυσιολογικούς ιστούς, γεγονός που υποδεικνύει ότι η κωδικοποιούμενη πρωτεΐνη έχει σημαντική λειτουργία σε ποικίλους τύπους κυττάρων, ωστόσο τα επίπεδα mRNA ποικίλουν αρκετά. Η ανάλυση έκφρασης για το γονίδιο SCAF1 έχει δείξει ότι μπορεί να θεωρηθεί ως ένας νέος δυσμενής προγνωστικός δείκτης, για τον ωοθηκικό καρκίνο και τον καρκίνο του μαστού.

Η πρωτεΐνη SCAF1, η οποία ανήκει στην υπεροικογένεια των πρωτεϊνών SR, περιέχει μια πλούσια σε αργινίνη/σερίνη περιοχή (SR) καθώς και περιοχή καρβοξυτελικής δέσμευσης (CTD-binding domain). Μέσω των αλληλεπιδράσεων με το pre-mRNA και την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης ΙΙ, έχει δειχθεί ότι οι πλούσιες σε αργινίνη/σερίνη πρωτεΐνες ρυθμίζουν το εναλλακτικό μάτισμα.

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, θα αναλυθούν τα αποτελέσματα αλληλούχησης νέας γενιάς για το συγκεκριμένο γονίδιο (SCAF1) σε διαφορετικές καρκινικές κυτταρικές σειρές αναζητώντας νέα, άγνωστα εναλλακτικά μετάγραφα του συγκεκριμένου γονιδίου.

39

2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. ΒΙΟΛΟΓΙΚΟ ΥΛΙΚΟ

Για τη διεξαγωγή της συγκεκριμένης εργασίας χρησιμοποιήθηκαν 55 διαφορετικές ανθρώπινες κυτταρικές σειρές (53 καρκινικές και 2 φυσιολογικές) από συνολικά 19 ιστούς όπως φαίνεται στον παρακάτω πίνακα (Πίνακας 1).

Πίνακας 1 : Συνοπτική παρουσίαση του	βιολογικού υλικού που χρησιμοποιήθηκε.	
·····, ·····, ·····		

<u>ΠΡΟΕΛΕΥΣΗ (ΙΣΤΟΣ)</u>	ΚΥΤΤΑΡΙΚΕΣ ΣΕΙΡΕΣ
Μαστός	MCF-7, SK-BR-3, BT-20, BT-474, MDA-MB-231,
	MDA-MB-435S, MDA-MB-468, T-47D, ZR-75-1,
	MCF-12A
Ωοθήκη	OVCAR-3, SK-OV-3, ES-2, MDAH-2774
Μήτρα	Ishikawa, SK-UT-1B, HeLa
Τράχηλος της μήτρας	SiHa
Προστάτης	PC3, DU-145, LNCaP
Ουροδόχος κύστη	T24, RT4
Νεφρός	ACHN, 786-0, Caki-1
Παχύ έντερο	HCT-116, RKO, HT-29, DLD-1, Caco-2, SW-620,
	Colo-205
Στόμαχος	AGS
Ήπαρ	Hep G2, HuH-7
Εγκέφαλος	U-87 MG, U-251 MG, D54, H4, SH-SY5Y
Πνεύμονας	A549
Μελάνωμα	FM3
Λέμφωμα	Raji, Daudi, U937, SUDHL, K562, HL-60, Rec-1,
	Granta
Λευχαιμία Τ-κυττάρων	Jurkat
Εμβρυονικός νεφρός	HEK 293
Πάγκρεας	1.2B4
Κεφαλή και τράχηλος	BB-49, CAL-33

2.2. ANOMON $\Omega\Sigma$ H OAIKOY RNA ANO ANOP Ω NINE SKYTTAPIKE SEIPES

Τα απαιτούμενα υλικά και διαλύματα που χρησιμοποιήθηκαν για τη διεξαγωγή της συγκεκριμένης μεθόδου είναι τα εξής:

- Χλωροφόρμιο (Applichem)
- Ισοπροπανόλη (Applichem)
- Αιθανόλη (ACS) 75% (Applichem)
- RNA Storage Solution (Ambion)

Το πρώτο βήμα για την εφαρμογή της συγκεκριμένης μεθόδου είναι να μεταφέρουμε τα δείγματα (ομογενοποιήματα) από την θερμοκρασία των -80°C στον πάγο και έπειτα να προσθέσουμε 200 μl χλωροφορμίου. Έπειτα θα ακολουθήσει έντονη ανακίνηση των δειγμάτων με το χέρι για χρονικό διάστημα 15 δευτερολέπτων και στην συνέχεια θα επωαστούν σε θερμοκρασία δωματίου για 10 λεπτά.

Εν συνεχεία, φυγοκεντρούμε τα δείγματα στις 13,000 rpm για 15 λεπτά στους 4°C όπου το ομογενοποίημα διαχωρίζεται στις ακόλουθες φάσεις: μια υδατική φάση που περιέχει το RNA (πάνω φάση), μια οργανική φάση που περιλαμβάνει τις πρωτεΐνες (κάτω φάση) και τη μεσόφαση στην οποία βρίσκεται το DNA, η οποία παρεμβάλλεται μεταξύ της υδατικής και οργανικής φάσης (Εικόνα 18).

Μετά το στάδιο της φυγοκέντρησης, μεταφέρεται η άχρωμη υδατική φάση που περιέχει RNA, σε αποστειρωμένο eppendorf των 2 ml. Για τη μεταφορά χρησιμοποιείται αποστειρωμένη σύριγγα ή κατάλληλη πιπέτα με ιδιαίτερη προσοχή, ώστε να μη διαταραχθούν οι φάσεις και συλλεχθεί DNA ή πρωτεΐνες. Το αρχικό eppendorf με την μεσόφαση και την οργανική φάση επιστρέφεται σε θερμοκρασία -80°C για αποθήκευση.

Έπειτα, προσθέτουμε 500 μl ισοπροπανόλης στο eppendorf με την υδατική φάση, αναδεύουμε (vortex) για 5-10 δευτερόλεπτα και αφήνουμε το



Εικόνα 18: Απεικόνιση των τριών φάσεων που προκύπτουν έπειτα από την πρώτη φυγοκέντρηση (https://pbs.twimg.com/media/ChI1UkcU4AIpPGe.jpg).

δείγμα για επώαση σε θερμοκρασία δωματίου για 10 λεπτά. Στη συνέχεια, ακολουθεί φυγοκέντρηση στις 12,000 g για 8 λεπτά στους 4°C και προσεκτική απομάκρυνση του υπερκείμενου, με χρήση αποστειρωμένης σύριγγας, ώστε να μην ξεκολλήσει το ίζημα του RNA από τον πυθμένα του eppendorf.

Στο eppendorf που πλέον περιέχει μόνο το ίζημα, προσθέτουμε 1 ml αιθανόλης 75% και αφού αναδεύσουμε (vortex) για λίγα δευτερόλεπτα, φυγοκεντρούμε το δείγμα στις 13,000 rpm για 5 λεπτά στους 4°C. Μετά τη φυγοκέντρηση, απομακρύνεται το υπερκείμενο με την χρήση αποστειρωμένης σύριγγας των 5 ml. Το RNA ίζημα αφήνεται να στεγνώσει στον αέρα (χωρίς να ξεραθεί εντελώς), αναποδογυρίζοντας το eppendorf στο διηθητικό χαρτί.

Τέλος, προσθέτουμε συγκεκριμένο όγκο του αντιδραστηρίου RNA Storage Solution (RSS), ανάλογα με την ποσότητα του ιζήματος του RNA που σχηματίσθηκε. Έτσι ο όγκος του ανωτέρω αντιδραστηρίου (RSS), κυμαίνεται από 10μl για μικρή ποσότητα ιζήματος έως και 40 μl για μεγαλύτερη ποσότητα. Το RNA του ιζήματος επαναδιαλύεται στο RNA Storage Solution (RSS) και φυλάσσεται σε θερμοκρασία -80°C [164].

2.2.1. Φασματοφωτομετρικός προσδιορισμός της συγκέντρωσης και έλεγχος της καθαρότητας του απομονωμένου ολικού RNA

Το RNA που απομονώθηκε με την διαδικασία που αναλύθηκε στην προηγούμενη ενότητα (RNA extraction), μπορεί να προσδιοριστεί ποσοτικά και ποιοτικά με φωτομέτρηση. Το φωτόμετρο που χρησιμοποιήθηκε για αυτή την διαδικασία είναι το BioSpec-nano Micro-volume UV-Vis Spectrophotometer, της Shimadzu Scientific Instruments (Εικόνα 19).

Για να προσδιορίσουμε την ποιότητα του RNA υπολογίζουμε το λόγο A260/A280, ο οποίος είναι ενδεικτικός της καθαρότητας του RNA που απομονώθηκε.

Η απορρόφηση του δείγματος στα 260 nm οφείλεται κυρίως στην απορρόφηση των αζωτούχων βάσεων των νουκλεϊκών οξέων. Πιο συγκεκριμένα, πυριμιδίνες οι (κυτοσίνη και ουρακίλη στο RNA, κυτοσίνη και θυμίνη στο DNA) απορροφούν λίγο πιο κάτω από τα 260 nm, ενώ οι πουρίνες (αδενίνη και γουανίνη) λίγο πιο πάνω από τα 260 nm). Η απορρόφηση στα 280 nm οφείλεται κυρίως στην απορρόφηση τριών αμινοξέων των πρωτεϊνών της φαινυλαλανίνης, της τυροσίνης και τρυπτοφάνης [165].



Εικόνα 19: Φωτόμετρο BioSpec-nano Micro volume UV-Vis της Shimadzu Scientific Instruments

(http://www.informm.usm.my/images/infor mm/research_photo/Biospec_nano.jpg).

Το φασματοφωτόμετρο

BioSpec-nano Micro-volume UV-Vis Spectrophotometer, της Shimadzu Scientific Instruments που χρησιμοποιήθηκε (Εικόνα 19), επιτρέπει τον ακριβή και γρήγορο προσδιορισμό της συγκέντρωσης και της καθαρότητας κάθε δείγματος υπολογίζοντας αυτόματα τη συγκέντρωση του RNA σε μg/μL. Σε φασματοφωτόμετρα που υπολογίζουν μόνο οπτική απορρόφηση στα 260 nm, η συγκέντρωση του RNA σε μg/μL προσδιορίζεται πολλαπλασιάζοντας την τιμή της απορρόφησης αυτής με τον παράγοντα της αραίωσης και με την τιμή 0.04 μg/μL, η οποία είναι η συγκέντρωση του RNA που αντιστοιχεί σε μια μονάδα απορρόφησης.

Για να θεωρείται καθαρό το RNA που απομονώθηκε, δηλαδή απαλλαγμένο από προσμίξεις DNA και πρωτεϊνών, θα πρέπει σύμφωνα με το πρωτόκολλο της Ambion για το TRI Reagent®, ο λόγος A260/A280 να κυμαίνεται μεταξύ 1,8 και 2,2. Στην περίπτωση που ο λόγος A260/A280, λαμβάνει τιμές μικρότερες του 1.6, το εκχύλισμα RNA έχει υψηλή συγκέντρωση πρωτεϊνών, ενώ τιμές μεγαλύτερες του 2.2 φανερώνουν την παρουσία ποσότητας DNA στο υπό μελέτη δείγμα [166].

Τέλος σημειώνεται, πως για κάθε δείγμα η μέτρηση επαναλαμβάνεται τρεις φορές και υπολογίζεται ο μέσος όρος των μετρήσεων.

2.3. $\Sigma YN\Theta E\Sigma H cDNA ME ANTISTPO \Phi H METAFPA \Phi H TOY mRNA (REVERSE TRANSCRIPTION, RT)$

2.3.1. Αρχή της μεθόδου

Η αντίδραση της αντίστροφης μεταγραφής (reverse transcription, RT), βασίζεται στη δράση ενός ενζύμου που καλείται, αντίστροφη μεταγραφάση (Reverse Transcriptase, RTase), η οποία ανακαλύφθηκε το 1970 από τους Howard Temin και David Baltimore, οι οποίοι μοιράστηκαν το 1975 το Νόμπελ Ιατρικής [167]. Αυτό το ένζυμο μεταγράφει RNA σε συμπληρωματικό DNA (complementary DNA, cDNA). Η αντίστροφη μεταγραφάση είναι μια RNAεξαρτώμενη DNA πολυμεράση, η οποία δημιουργεί συμπληρωματικά μόρια DNA χρησιμοποιώντας ως εκμαγείο το RNA (Εικόνα 20). Επίσης, εκφράζεται μόνο από συγκεκριμένους τύπους ιών, οι οποίοι χρειάζονται αυτό το ένζυμο για να αναπαραχθούν [132].

Στο αρχικό στάδιο, στα μόρια RNA υβριδοποιούνται μόρια κατάλληλου εκκινητή, στα οποία η RTase προσθέτει τα συμπληρωματικά με το RNA δεοξυριβονουκλεοτίδια για τη δημιουργία του συμπληρωματικού DNA (cDNA). Συνεπώς, ο εκκινητής αυτός καθορίζει ποια μόρια RNA θα μεταγραφούν αντίστροφα σε cDNA, καθορίζει δηλαδή την ειδικότητα της αντίδρασης της αντίστροφης μεταγραφής.



Εικόνα 20: Σχηματική παρουσίαση της σύνθεσης cDNA με αντίστροφη μεταγραφή mRNA (reverse transcription, του RT) (https://www.thermofisher.com/content/da m/LifeTech/global/life-

sciences/Cloning/Images/0816/reverse_trans cription_process.png).

Όλα τα ευκαρυωτικά μόρια mRNA κατά την ωρίμανσή τους υφίστανται πολυαδενυλίωση, γεγονός στο οποίο βασίζεται η χρήση των ολιγονουκλεοτιδίων δεοξυθυμιδίνης [176]. Συγκεκριμένα, αυτοί οι εκκινητές υβριδοποιούνται στην poly(A) ουρά των ευκαρυωτικών μορίων mRNA, δημιουργώντας μόρια cDNA μόνο από τα μετάγραφα mRNA [177, 178]. Ο πληθυσμός των μορίων mRNA αποτελεί το ένα πολύ μικρό ποσοστό (1 έως 5 %) του ολικού RNA των κυττάρων, έτσι η χρήση oligo-dT ως εκκινητή έχει ως αποτέλεσμα τη μείωση της

Οι κατηγορίες των εκκινητών χρησιμοποιούνται που στην αντίδραση αντίστροφης μεταγραφής (RT), είναι τρεις και περιλαμβάνουν: τα τυχαία εξαμερή [168-172], τα ολιγομερή δεοξυθυμίδινης (oligo-dT) [168, 173], και ειδικό εκκινητή για συγκεκριμένο γονίδιο [174, 175] (Εικόνα 21).

Τα τυχαία εξαμερή, όπως δηλώνει και το όνομα τους, υβριδοποιούνται καμία χωρίς ειδικότητα (τυχαία) στο συνολικό πληθυσμό των μορίων RNA που έχουν προηγουμένως εκχυλιστεί, επιτρέποντας έτσι σε όλα τα μόρια RNA να δράσουν ως εκμαγεία για τον σχηματισμό μορίων cDNA. Στην περίπτωση αυτή, η ειδικότητα της αλυσιδωτής αντίδρασης πολυμεράσης αντίστροφης μεταγραφής (RT-PCR) καθορίζεται αποκλειστικά από τους εκκινητές της PCR που ακολουθεί.

πολυπλοκότητας του παραγόμενου πληθυσμού cDNA και τη βελτίωση της απόδοσης της αντίδρασης. Το γεγονός αυτό είναι πολύ σημαντικό για τον περαιτέρω ακριβή προσδιορισμό του αριθμού των μεταγράφων mRNA που εκφράζονται ελάχιστα.

Oligo(dT) primers	Random primers	Gene-specific primers
Standard		
S AAAA S	я N ₄ в ²	AAAA F
Anchored		
mRNA cDNA NY	VATGC Primer	

Εικόνα 21: Εκκινητές που χρησιμοποιούνται συνήθως στην αντίστροφη μεταγραφή (RT) (https://www.thermofisher.com/gr/en/home/life-science/cloning/cloninglearning-center/invitrogen-school-of-molecular-biology/rt-education/reversetranscription-setup.html).

Τέλος, μέγιστη ειδικότητα κατά την αντίδραση της αντίστροφης μεταγραφής επιτυγχάνεται με τη χρήση ειδικού εκκινητή για κάποιο γονίδιο, ο οποίος σχεδιάζεται έτσι ώστε να υβριδοποιείται μόνο στα μετάγραφα mRNA που προκύπτουν από τη μεταγραφή του συγκεκριμένου γονιδίου.

Πολύ σημαντική είναι η μη αντιστρεπτή αποδιάταξη της αντίστροφης μεταγραφάσης μετά την ολοκλήρωση της σύνθεσης του cDNA, καθώς σε αντίθετη περίπτωση τα ενζυμικά αυτά μόρια παραμένουν προσδεδεμένα με το νεοσύστατο cDNA, παρεμποδίζοντας έτσι τη μετέπειτα ενίσχυση αλληλουχιών με PCR.

2.3.2. Συνθήκες αντίδρασης αντίστροφης μεταγραφής για τη σύνθεση cDNA από απομονωμένο ολικό RNA

Τα απαιτούμενα υλικά και διαλύματα που χρησιμοποιήθηκαν για τη διεξαγωγή της συγκεκριμένης μεθόδου ήταν τα εξής:

- H₂O ελεύθερο από (δεοξυ)ριβονουκλεάσες (RNase/DNase-free H₂O)
- Μείγμα τριφωσφορικών δεοξυριβονουκλεοτιδίων (dNTPs)
- Ρυθμιστικό διάλυμα αντίδρασης 5X (RT-buffer)
- Διάλυμα 100 mM DTT
- Αναστολέας RNασών (Human Placental RNase Inhibitor) (Invitrogen)
- Αντίστροφη Μεταγραφάση, SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA)
- Θερμικός κυκλοποιητής (Applied Biosystems Veriti™ Thermal Cycler)

5μg RNA των διαλυμάτων που φωτομετρήσαμε προηγουμένως (53 καρκινικές κυτταρικές σειρές και 2 φυσιολογικές), χρησιμοποιούνται ως υπόστρωμα στη μέθοδο αυτή, με σκοπό τη σύνθεση του συμπληρωματικού DNA (complementary DNA, cDNA) με τη δράση της αντίστροφης μεταγραφάσης. Η αντίδραση της αντίστροφης μεταγραφής πραγματοποιείται σε δύο στάδια τα οποία πραγματοποιούνται στο θερμικό κυκλοποιητή.

Αρχικά σε ένα eppendorf των 0,2 ml απαλλαγμένο από DNάσες, προσθέτουμε:

- ✓ 1 μl ολιγονουκλεοτιδίων δεοξυθυμίνης με αντάπτορα (oligo-dT adaptor)
- ✓ 1 µl dNTPs
- ✓ 'Ογκο διαλύματος RNA
- Συμπληρώνουμε με RNase/DNase-free H₂O μέχρι ο συνολικός
 όγκος του διαλύματος να είναι 13 μl

Ο όγκος του διαλύματος RNA που προστίθεται στο μείγμα της αντίδρασης υπολογίζεται με βάση τη συγκέντρωση του RNA του συγκεκριμένου δείγματος. Η συγκέντρωση του RNA υπολογίζεται με βάση την απορρόφηση του δείγματος στα 260 nm (A260). Η σχέση που μας δίνει τον όγκο του διαλύματος RNA, ο οποίος αντιστοιχεί σε 1 μg είναι:

$$V_{RNA} (\mu l) = 1 / (8 \times A260)$$

Όταν προστεθούν όλα τα παραπάνω αντιδραστήρια στο eppendorf, τότε τοποθετείται στον θερμικό κυκλοποιητή στους 70°C για 5 λεπτά. Στο στάδιο αυτό, αποδιατάσσονται οι ανώτερες δομές που μπορεί να περιέχει το RNA λόγω συμπληρωματικότητας και υβριδοποιούνται τα oligo-dT στην poly(A) ουρά των mRNA μορίων του δείγματος.

Μετά το πέρας των 5 λεπτών, εξάγουμε το δείγμα από το θερμικό κυκλοποιητή και προσθέτουμε τα παρακάτω αντιδραστήρια:

- 4 μl RT buffer (ρυθμιστικό διάλυμα) που περιέχει:
 - 20 mM Tris-HCl (pH 7.5)
 - ➢ 100 mM NaCl
 - > 0.1 mM EDTA
 - ➤ 1 mM DTT
 - ➢ 0.01% (v/v) NP-40
 - 50% (v/v) γλυκερόλη
- ✓ 2μl διαλύματος DTT

- ✓ 0.5 μl διαλύματος αναστολέα RNασών (Human Placental RNase Inhibitor)
- 0.5 μl διαλύματος αντίστροφης μεταγραφάσης, SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA)

Ο συνολικός όγκος κάθε δείγματος, μετά την προσθήκη και των παραπάνω αντιδραστηρίων, είναι 20 μl. To eppendorf με τα 20 μl του μείγματος αντίδρασης τοποθετείται στο θερμικό κυκλοποιητή. Το πρωτόκολλο που ακολουθεί ο θερμικός κυκλοποιητής είναι το κάτωθι:

- 52 λεπτά στη θερμοκρασία των 37°C
- 15 λεπτά στη θερμοκρασία των 70°C

Τα cDNAs που θα παραχθούν αραιώνονται και αποθηκεύονται σε 11 ομάδες, ανάλογα με τον ιστό προέλευσης τους, στους -20°C, είτε χρησιμοποιούνται άμεσα σε άλλες διαδικασίες (π.χ. συμβατική PCR).

2.4. ΑΛΥΣΙΔΩΤΗ ΑΝΤΙΔΡΑΣΗ ΠΟΛΥΜΕΡΑΣΗΣ (POLYMERASE CHAIN REACTION, PCR)

2.4.1. Αρχή της μεθόδου

Η αλυσιδωτή αντίδραση πολυμεράσης (polymerase chain reaction, PCR) είναι μια επαναλαμβανόμενη ενζυμική διαδικασία, όπου μια συγκεκριμένη αλληλουχία DNA - στόχος αντιγράφεται in vitro, προκειμένου να παραχθεί μεγάλος αριθμός αντιγράφων της εν λόγω αλληλουχίας. Κάθε κύκλος αυτής της επαναλαμβανόμενης διαδικασίας, η οποία περιλαμβάνει συνήθως 25 – 40 κύκλους, αποτελείται από τα στάδια της αποδιάταξης του δίκλωνου DNA υποστρώματος, του υβριδισμού των εκκινητών και της σύνθεσης του DNA (Εικόνα 23), τα οποία θα αναλυθούν στη συνέχεια.

Σε κάθε κύκλο, κάθε μόριο που περιέχει την επιθυμητή προς ενίσχυση αλληλουχία DNA - στόχο, παράγει ένα αντίγραφο αυτής της αλληλουχίας. Στο τέλος μιας αντίδρασης PCR, μετά από n κύκλους θα παραχθούν, θεωρητικά, 2ⁿ δίκλωνα μόρια DNA, που αποτελούν αντίγραφα της αλληλουχίας - στόχου του DNA. Ωστόσο, δεν είναι απαραίτητο να απομονωθεί μόνο η προς ενίσχυση αλληλουχία, διότι αυτή μπορεί να καθοριστεί – οριοθετηθεί μέσω των εκκινητών που θα χρησιμοποιηθούν στην αλυσιδωτή αντίδραση πολυμεράσης. Η δράση της DNA πολυμεράσης κατευθύνεται μέσω κατάλληλων εκκινητών και πραγματοποιείται η σύνθεση συγκεκριμένης περιοχής του DNA με εκθετικό τρόπο.

Στην αλυσιδωτή αντίδραση πολυμεράσης, λόγω των υψηλών θερμοκρασιών που επιτυγχάνονται για την αποδιάταξη του DNA υποστρώματος, οι κοινές DNA πολυμεράσες αποδιατάσσονται, με αποτέλεσμα να απαιτείται σε κάθε κύκλο προσθήκη ενζύμου. Όμως, με την απομόνωση των θερμοανθεκτικών DNA πολυμερασών (π.χ. Taq πολυμεράση) από μικροοργανισμούς που ζουν σε θερμές πηγές [179] (Εικόνα 22), δεν απαιτείται πλέον προσθήκη ενζύμου σε κάθε κύκλο. Η αντίδραση PCR σχεδιάστηκε και παρουσιάστηκε από τον Dr Kary Mullis το 1983 [180].

Πλέον, για την πραγματοποίηση της αντίδρασης PCR χρησιμοποιούνται ειδικά όργανα, οι αυτόματοι θερμικοί κυκλοποιητές (thermal cyclers), που σε κάθε κύκλο μεταβαίνουν αυτόματα στις θερμοκρασίες που έχουμε επιλέξει, παρέχοντας τη δυνατότητα παραγωγής εκατομμυρίων αντιγράφων συγκεκριμένης αλληλουχίας DNA σε ελάχιστο χρονικό διάστημα.



Εικόνα 22: Thermus aquaticus https://upload.wikimedia.org/wikipedia/com mons/thumb/4/48/Thermus_aquaticus.JPG/30 0px-Thermus_aquaticus.JPG

Τα πρωτόκολλα που χρησιμοποιούνται για την αλυσιδωτή αντίδραση πολυμεράσης διαφέρουν ανάλογα με τις απαιτήσεις του ενζύμου που χρησιμοποιείται, τη δομή και το μέγεθος του DNA υποστρώματος, καθώς και τις βέλτιστες συνθήκες θερμοκρασίας πρόσδεσης των εκκινητών.

Ένα τυπικό πρωτόκολλο PCR περιλαμβάνει πολλούς κύκλους (25-40), ο καθένας εκ των οποίων περιλαμβάνει επώαση των δειγμάτων σε 3 διαφορετικές

θερμοκρασίες (αποδιάταξη του δίκλωνου DNA υποστρώματος, υβριδισμός εκκινητών, σύνθεση DNA), βάσει των παρακάτω βημάτων (Εικόνα 23):

- Αποδιάταξη του δίκλωνου DNA υποστρώματος (denaturation). Το μίγμα θερμαίνεται αρχικά στους 95°C για περίπου 15 λεπτά. Λόγω της υψηλής θερμοκρασίας που αναπτύσσεται στο διάλυμα, οι δεσμοί υδρογόνου που συγκρατούν τις δύο συμπληρωματικές αλυσίδες του DNA διασπώνται. Ως εκ τούτου, καταστρέφονται οι δευτεροταγείς και τριτοταγείς δομές, τα δίκλωνα μόρια DNA αποχωρίζονται τελείως και έτσι παράγονται οι μονόκλωνες αλυσίδες που θα χρησιμεύσουν ως εκμαγεία για τους εκκινητές και την DNA πολυμεράση.
- 2. Υβριδοποίηση των εκκινητών (annealing). Στη συνέχεια το διάλυμα ψύχεται απότομα και η θερμοκρασία μειώνεται μεταξύ 40-72°C (ανάλογα με το σχεδιασμό των εκκινητών, τη συγκέντρωση των αλάτων και τη συγκέντρωση των συστατικών του διαλύματος), ώστε οι εκκινητές να υβριδοποιηθούν με τις συμπληρωματικές τους αλληλουχίες στις μονόκλωνες, πλέον, αλυσίδες του DNA. Ο ένας εκκινητής προσδένεται στο 3' άκρο της μιας αλυσίδας, ενώ ο άλλος στο 3' άκρο της συμπληρωματικής αλυσίδας του DNA. Η θερμοκρασία υβριδισμού αποτελεί καθοριστικό παράγοντα για την εξειδίκευση της αντίδρασης PCR.
- 3. Επέκταση (extension) ή επιμήκυνση (elongation) του DNA. Στο επόμενο βήμα, η θερμοκρασία αυξάνεται στους 72°C, ώστε να εκτελεστεί η σύνθεση της συμπληρωματικής αλυσίδας DNA. Αυτή είναι η βέλτιστη θερμοκρασία για τη δράση της θερμοανθεκτικής DNA πολυμεράσης (συνήθως Taq πολυμεράση). Η DNA πολυμεράση επιμηκύνει και τους δύο εκκινητές προς την 3' κατεύθυνση και έτσι επιτυγχάνεται η σύνθεση της συμπληρωματικής αλυσίδας του DNA, καταναλώνοντας τα τριφωσφορικά δεοξυριβονουκλεοτίδια (dNTPs) του διαλύματος. Ο χρόνος επώασης στους 72°C ποικίλει από 1 έως 3 λεπτά, ανάλογα με το μήκος του στόχου που θέλουμε να ενισχύσουμε.
- Αποδιάταξη των προϊόντων. Η θερμοκρασία αυξάνεται και πάλι στους 95°C, ώστε τα δίκλωνα μόρια που μόλις συντέθηκαν, να αποχωριστούν

στις επιμέρους αλυσίδες. Αυτές οι μονόκλωνες αλυσίδες αποτελούν υποστρώματα για τον επόμενο κύκλο σύνθεσης DNA.

- 5. Επανάληψη των σταδίων 2 έως 4 για 25-40 φορές (25-40 κύκλοι).
- 6. Μετά το πέρας του τελευταίου κύκλου, ακολουθεί ένα στάδιο επιμήκυνσης διάρκειας περίπου 10 λεπτών, χρόνος απαραίτητος για την επέκταση όλων των προϊόντων από την θερμοανθεκτική DNA πολυμεράση.

Κατά τη διάρκεια του πρώτου κύκλου, το προϊόν που παράγεται είναι μικρότερο σε μέγεθος από το αρχικό δείγμα DNA (υπόστρωμα), αλλά μεγαλύτερο από την επιθυμητή προς ενίσχυση αλληλουχία. Στον δεύτερο κύκλο, έχουμε επίσης προϊόντα ενδιάμεσου μεγέθους και έπειτα εμφανίζονται αλυσίδες που αντιστοιχούν στην αλληλουχία - στόχο και περιλαμβάνουν στα άκρα τους εκκινητές. Πλέον, η ενίσχυση των προϊόντων επιθυμητού μεγέθους (που περιέχουν την αλληλουχία-στόχο) αυξάνεται εκθετικά, σε σύγκριση με την γραμμική συσσώρευση των προϊόντων ενδιάμεσου μεγέθους, που συνεπάγεται την ανίχνευση κυρίως των επιθυμητών προϊόντων στο τέλος της αντίδρασης [181, 182].

Με τη συνεχή αύξηση των εφαρμογών της PCR, η συγκεκριμένη μέθοδος αποτελεί πλέον μια απαραίτητη διαδικασία για ένα μεγάλο αριθμό επιστημονικών πεδίων, όπως για παράδειγμα στην: μοριακή βιολογία, μικροβιολογία, γενετική, κλινική βιοχημεία, μοριακή διαγνωστική, εγκληματολογία, επιστήμες του περιβάλλοντος, κληρονομικές μελέτες, τεστ πατρότητας.



Principle of the Polymerase Chain Reaction (PCR) Method

Εικόνα 23: Σχηματική παρουσίαση των βημάτων της αλυσιδωτής αντίδρασης πολυμεράσης (polymerase chain reaction, PCR) (http://classroom.sdmesa.edu/eschmid/images/Lab8-B1.jpg).

2.4.2. Αντιδραστήρια της PCR

Τα απαιτούμενα αντιδραστήρια και διαλύματα που χρησιμοποιήθηκαν για τη διεξαγωγή της αλυσιδωτής αντίδρασης πολυμεράσης ήταν τα εξής:

- H2O ελεύθερο από (δεοξυ)ριβονουκλεάσες (RNase/DNase-free H2O)
- 10X KAPA Taq buffer A (Kapa Biosystems Inc.), που περιέχει MgCl₂ τελικής συγκέντρωσης 1,5mM
- Μείγμα τριφωσφορικών δεοξυριβονουκλεοτιδίων (dNTPs)
- DNA Πολυμεράση-Τaq Πολυμεράση (Kapa Biosystems)
- Εκκινητές (ολιγονουκλεοτίδια)

 Θερμικός κυκλοποιητής (Veriti 96-Well Fast Thermal Cycler Applied Biosystems[™])

Ρυθμιστικό διάλυμα

Τα συστατικά που περιέχονται κατά κύριο λόγο είναι 100 mM Tris-HCl (pH 8,3 στους 25°C), 500 mM KCl, 1,5 mM MgCl₂ και 0,1% Tween-20. Η χρήση του ρυθμιστικού διαλύματος γίνεται για να βελτιστοποιηθεί η δράση της DNA πολυμεράσης, καθώς και για την βέλτιστη υβριδοποίηση των εκκινητών στο DNA εκμαγείο. Το Tris-HCl φέρει σε θερμοκρασία δωματίου pH 8,3, ενώ στους 72°C το pH του πέφτει στο 7,4 που αποτελεί το βέλτιστο pH για τις περισσότερες DNA πολυμεράσες. Τα μονοσθενή κατιόντα, όπως το K+, μειώνουν τις ηλεκτροαπωθητικές δυνάμεις που αναπτύσσονται μεταξύ των εκκινητών και του DNA υποστρώματος, εξαιτίας των αρνητικά φορτισμένων φωσφορικών ομάδων του σκελετού του DNA [181, 182].

Μείγμα τριφωσφορικών νουκλεοτιδίων (dNTPs)

Τα dNTPs (dATP, dCTP, dGTP, dTTP) ανήκουν στα απαραίτητα συστατικά για τη σύνθεση της νέας αλυσίδας του DNA στην αντίδραση PCR. Οι συγκεντρώσεις των dNTPs κυμαίνονται μεταξύ 50-200mM. Το μείγμα των τριφωσφορικών νουκλεοτιδίων πρέπει να περιέχει ίσες συγκεντρώσεις των dATP, dTTP, dCTP και dGTP, εκτός και αν η αλληλουχία DNA που πρόκειται να ενισχυθεί περιέχει κυρίως A/T ή G/C, οπότε και η αναλογία των dNTPs αλλάζει.

Σε υψηλότερες συγκεντρώσεις, η πιστότητα της αντιγραφής από την Taq πολυμεράση μειώνεται και ο ρυθμός εισαγωγής λάθος νουκλεοτιδίων αυξάνει προκαλώντας έτσι, την παραγωγή παραπροϊόντων από την πολυμεράση. Αντίθετα, αν η συγκέντρωση είναι χαμηλότερη μπορεί να επηρεάσει την αποδοτικότητα της αντίδρασης.

Η βέλτιστη συγκέντρωση των dNTPs εξαρτάται κάθε φορά από διάφορους παράγοντες, όπως τη συγκέντρωση ιόντων μαγνησίου, τη συγκέντρωση των εκκινητών, το μήκος των αντιγράφων του DNA και τον αριθμό των κύκλων της αντίδρασης.

Επειδή, τα dNTPs δεσμεύουν τα ιόντα Mg²⁺, σημαντική αλλαγή στη συγκέντρωση των dNTPs θα προκαλέσει αλλαγή στη συγκέντρωση του ελεύθερου ποσού του διαθέσιμου Mg²⁺ [181, 182].

Χλωριούχο μαγνήσιο (MgCl₂)

Ένα από τα σημαντικότερα συστατικά της PCR, αποτελούν τα ιόντα μαγνησίου (Mg²⁺), καθώς η συγκέντρωση αυτών επηρεάζει τόσο την ειδικότητα όσο και την αποδοτικότητα της αντίδρασης. Τα ιόντα μαγνησίου (Mg²⁺) προστίθενται στο μείγμα της αντίδρασης με την μορφή MgCl₂ και αποτελούν συμπαράγοντα της DNA πολυμεράσης.

Σε συγκέντρωση ελεύθερου Mg^{2+} 1,2-1,3 mM, η DNA πολυμεράση παρουσιάζει τη μέγιστη ενεργότητά της. Η συγκέντρωση του ελεύθερου μαγνησίου επηρεάζεται από την συγκέντρωση των dNTPs όπως έχει αναφερθεί και παραπάνω, καθώς τα ιόντα Mg^{2+} προσδένονται στοιχειομετρικά με τα dNTPs και σχηματίζουν διαλυτά σύμπλοκα που διευκολύνουν την ενσωμάτωση τους στην αλυσίδα του DNA. Η συγκέντρωση των ιόντων μαγνησίου κυμαίνεται από 0,5 ως 5 mM. Συγκεντρώσεις περίπου 1,5mM είναι συνηθέστερες, αλλά σε μερικές περιπτώσεις διαφορετικές συγκεντρώσεις Mg^{2+} μπορεί να αποδειχτούν απαραίτητες [182].

Χαμηλές συγκεντρώσεις των ιόντων μαγνησίου οδηγούν σε χαμηλή απόδοση της αντίδρασης. Αντίθετα, πολύ αυξημένες συγκεντρώσεις τους οδηγούν σε αναστολή της ενίσχυσης λόγω της σταθεροποίησης της διπλής έλικας, καθώς και σε παραγωγή μη ειδικών προϊόντων εξαιτίας της σταθεροποίησης και πρόσδεσης των εκκινητών σε μη συμπληρωματικές αλληλουχίες. Επίσης, σε υψηλές συγκεντρώσεις ιόντων Mg²⁺ αυξάνεται η συχνότητα των λαθών από την DNA πολυμεράση κατά την αντιγραφή, με αποτέλεσμα να μειώνεται η πιστότητα της αντίδρασης [181, 182].

DNA Πολυμεράση - Ταq Πολυμεράση

Η DNA πολυμεράση κατά την σύνθεση του DNA επιλέγει το σωστό νουκλεοτίδιο για την επιμήκυνση της ολιγονουκλεοτιδικής αλυσίδας σε συμφωνία με τους κανόνες συμπληρωματικότητας των βάσεων και ανάλογα με τη βάση που βρίσκεται στην αλυσίδα εκμαγείο. Συναντούμε δύο κατηγορίες DNA πολυμερασών: τις DNA εξαρτώμενες DNA πολυμεράσες και τις RNA εξαρτώμενες DNA πολυμεράσες ή αντίστροφες μεταγραφάσες.

Η DNA πολυμεράση καταλύει τη σύνθεση του DNA προς την 5' \rightarrow 3' κατεύθυνση. Ορισμένες πολυμεράσες διαθέτουν και μια 3' \rightarrow 5' εξωνουκλεολυτική ενεργότητα (επιδιορθωτική), η οποία ελέγχει την τοποθέτηση του σωστού νουκλεοτιδίου στην αναπτυσσόμενη αλυσίδα DNA. Αν η τοποθέτηση δεν έχει πραγματοποιηθεί σωστά, τότε μέσω της 3' \rightarrow 5' εξωνουκλεολυτικής ενεργότητας απομακρύνεται το νουκλεοτίδιο και εισάγεται το σωστό νουκλεοτίδιο λόγω της 5' \rightarrow 3' πολυμεριστικής ενεργότητας. Έτσι αυξάνεται η ακρίβεια, δηλαδή η πιστότητα, με την οποία η πολυμεράση αντιγράφει την αλυσίδα εκμαγείο [183].

Μετά την ανακάλυψη των θερμοανθεκτικών DNA πολυμερασών που διαθέτουν βακτήρια που ζουν σε υψηλές θερμοκρασίες σημειώθηκε μεγάλη πρόοδος στην αντίδραση PCR. Το θερμόφιλο βακτήριο Thermus aquaticus, απομονώθηκε και περιγράφτηκε πριν από σχεδόν 50 χρόνια [184]. Το συγκεκριμένο βακτήριο ζει στο νερό σε θερμοκρασία 75°C. Πλέον, η Taq πολυμεράση, με μοριακό βάρος 94 kDa, αποτελεί το πιο κοινό ένζυμο για χρήση σε αντιδράσεις PCR και διαθέτει δύο ενζυμικές ενεργότητες:

- μία 5΄→3΄ πολυμεριστική ενεργότητα με ρυθμό επιμήκυνσης 50-60
 νουκλεοτίδια/δευτερόλεπτο,
- μία 5΄→3΄ ενεργότητα εξωνουκλεάσης.

Το ένζυμο αυτό με χρόνο ημιζωής περίπου 130 λεπτά στους 92,5°C (40 λεπτά στους 95°C), παρουσιάζει γύρω στους 72-75°C τη βέλτιστη θερμοκρασία για την σύνθεση DNA. Στη θερμοκρασία αυτή αποτρέπονται διάφορες δευτεροταγείς ή τριτοταγείς δομές του υποστρώματος και εμποδίζεται η πρόσδεση των εκκινητών σε μη συμπληρωματικές θέσεις.

Η διαδικασία αντιγραφής του DNA, δεν είναι ιδανική και κάποιες φορές μπορεί να προστεθεί στην επιμηκυνόμενη αλυσίδα ένα λάθος νουκλεοτίδιο από την DNA πολυμεράση. Επειδή όμως όπως αναφέρθηκε, η Taq πολυμεράση δε διαθέτει την 3' \rightarrow 5' επιδιορθωτική ενεργότητα, τα λανθασμένα νουκλεοτίδια που τοποθετούνται στην αλυσίδα δεν απομακρύνονται. Η Taq πολυμεράση χαρακτηρίζεται από την προσθήκη μιας A στο 3' άκρο της νεοσυντιθέμενης αλυσίδας DNA [185]. Θερμοανθεκτικές DNA πολυμεράσες από άλλα βακτήρια, όπως οι Vent, Pfu και UlTma, διαθέτουν αφενός επιδιορθωτική ενεργότητα και αφετέρου είναι πιο θερμοσταθερές [179, 181, 182].

Με την αντοχή της Taq πολυμεράσης σε υψηλές θερμοκρασίες είναι δυνατή η αυτοματοποίηση της PCR με τη χρήση των θερμικών κυκλοποιητών (thermal cyclers), καθώς είναι δυνατή η προσθήκη της πολυμεράσης μια φορά στην αρχή της αντίδρασης αφού έχει την ικανότητα να παραμένει ενεργή σε όλη την διάρκεια των κύκλων.

Εκτός από τη δυνατότητα αυτοματοποίησης, με τη χρήση της Taq πολυμεράσης βελτιώθηκε η ευαισθησία της αντίδρασης PCR. Με τη Taq πολυμεράση χρησιμοποιούνται υψηλές θερμοκρασίες και όχι χαμηλές που απαιτούνται για την δράση της Ε. Coli DNA πολυμεράσης. Στις χαμηλές θερμοκρασίες οι εκκινητές θα μπορούσαν να υβριδοποιηθούν και σε διαφορετικές περιοχές από την περιοχή που βρίσκεται η αλληλουχία στόχος ενισχύοντας έτσι μη επιθυμητές περιοχές. Χρησιμοποιώντας λοιπόν υψηλές θερμοκρασίες εξαιτίας της θερμοανθεκτικότητας της Taq πολυμεράσης αποφεύγουμε την εμφάνιση των φαινομένων αυτών [181].

Σχεδιασμός και επιλογή του εκκινητή

Σημαντικό παράγοντα για την επιτυχή έκβαση μιας αντίδρασης PCR αποτελεί ο ακριβής σχεδιασμός και η επιλογή του κατάλληλου εκκινητή. Στις τυπικές αντιδράσεις PCR απαιτούνται δύο εκκινητές με διαφορετική αλληλουχία και τέλεια συμπληρωματικότητα με τις αλληλουχίες που βρίσκονται στα άκρα του τμήματος του DNA στόχου που επιθυμούμε να ενισχυθεί. Για να πολλαπλασιαστεί όμως μόνο ο επιθυμητός στόχος DNA, οι εκκινητές θα πρέπει να έχουν σχεδιαστεί σωστά ώστε να παρέχουν υψηλή εξειδίκευση. Όταν η αλληλουχία του DNA που χρησιμοποιείται ως υπόστρωμα είναι γνωστή, τότε είναι εύκολη η σχεδίαση κατάλληλων εκκινητών, με τη χρήση λογισμικού, προκειμένου να ενισχυθεί το τμήμα στόχος. Κατά τον σχεδιασμό των εκκινητών πρέπει να εξετάζονται κάποια κριτήρια, τα πιο σημαντικά από τα οποία ακολουθούν στη συνέχεια.

Το μήκος των εκκινητών πρέπει να κυμαίνεται από 18 έως 30 νουκλεοτίδια (συνήθως 20 νουκλεοτίδια), παρέχοντας αρκετά ικανή ειδικότητα για την ενίσχυση της αλληλουχίας-στόχου και μόνο. Μικρού μήκους εκκινητές ενδέχεται να συνδεθούν και με άλλα τμήματα του DNA δίνοντας μη ειδικά προϊόντα, ενώ μεγαλύτερου μήκους εκκινητές μειώνουν την απόδοση της αντίδρασης.

Κατά το σχεδιασμό των εκκινητών πρέπει να αποφεύγονται όσο είναι εφικτό επαναληπτικές ακολουθίες του ίδιου νουκλεοτιδίου, καθώς αυτό μπορεί να οδηγήσει σε «γλίστρημα» του εκκινητή στην αλληλουχία - στόχο. Επαναλαμβανόμενες δομές με υψηλό ποσοστό G+C είναι δυνατόν να οδηγήσουν σε αναδιπλώσεις φουρκέτας δημιουργώντας προβλήματα στην πρόσδεση γι' αυτό και το ποσοστό G+C πρέπει να κυμαίνεται μεταξύ 45% και 55% (συνήθως 50%). Το 3' άκρο του ενός εκκινητή δεν πρέπει να είναι συμπληρωματικό με το 3' άκρο του άλλου εκκινητή του ζεύγους, γιατί προκαλεί το σχηματισμό διμερών των εκκινητών.

Η αλληλουχία του εκκινητή μπορεί να μην είναι πάντα απόλυτα συμπληρωματική με την αλληλουχία DNA του υποστρώματος. Θα πρέπει όμως η αλληλουχία του εκκινητή που βρίσκεται στο 3' άκρο του, να διαθέτει τέλεια συμπληρωματικότητα (και ειδικότερα τα τρία πρώτα νουκλεοτίδια του), διότι η DNA πολυμεράση επιμηκύνει τον εκκινητή από το ελεύθερο 3'-OH του. Η ύπαρξη τριών δεσμών υδρογόνου στα ζευγάρια GC, τα καθιστά πιο σταθερά από τα AT, για το λόγο αυτό είναι προτιμότερο να μην υπάρχουν G's ή C's στο 3' άκρο του εκκινητή. Επίσης, επειδή η Τ εμφανίζει μεγαλύτερη ανοχή σε μη ειδικές δεσμεύσεις, πρέπει να αποφεύγεται ως τελευταία βάση, καθώς επίσης δε θα πρέπει να εμφανίζονται σε αυτή την περιοχή δευτεροταγείς δομές.

Σοβαρά υπόψιν πρέπει να ληφθεί η θερμοκρασία τήξεως των εκκινητών (melting temperature, Tm), δηλαδή η θερμοκρασία στην οποία το 50% των εκκινητών έχει υβριδοποιηθεί στην αλληλουχία-στόχο. Οι θερμοκρασίες τήξεως των εκκινητών που θα χρησιμοποιηθούν θα πρέπει να είναι παραπλήσιες (διαφορά της τάξης 1-2°C), ώστε και οι δύο εκκινητές να δεσμεύονται ειδικά στην ίδια θερμοκρασία αναδιάταξης και να κυμαίνονται από 55-80°C. Ένας απλός τρόπος για να υπολογιστεί η Tm ενός εκκινητή μήκους έως 20 νουκλεοτιδίων, είναι μέσω του τύπου που ακολουθεί:

Γενικά, οι δύο εκκινητές πρέπει να χρησιμοποιούνται στην ίδια συγκέντρωση (συνήθως 0,05 - 0,5 μΜ). Υψηλότερες συγκεντρώσεις θα αυξήσουν την πιθανότητα σχηματισμού διμερών - εκκινητών και μη ειδικών προϊόντων [181, 182].

2.4.3. Μοριακή κλωνοποίηση των μεταγράφων του γονιδίου SCAF1

Για τη μοριακή κλωνοποίηση των μεταγράφων του γονιδίου SCAF1 χρησιμοποιήθηκε η τεχνική της αλυσιδωτής αντίδρασης πολυμεράσης (PCR) στα cDNAs τμήματα που προέκυψαν από την διαδικασία της RT-PCR, και στη συνέχεια, η διεξαγωγή μιας εσωτερικής PCR αντίδρασης (nested PCR) στα προϊόντα που προέκυψαν αφού αυτά είχαν πρώτα αραιωθεί.

Έτσι για την αρχική αντίδραση PCR χρησιμοποιήθηκε ένας πρόσθιος εκκινητής (forward primer) με αλληλουχία 5'- GGTGACCATGGAGGAAGAAGATG-3' (Πίνακας 3), που στοχεύει το πρώτο ATG του δεύτερου εξονίου του SCAF1 και ένας ανάστροφος εκκινητής (reverse primer) με αλληλουχία 5'-GTCCAGAGTTTCAAAGAGGTGAGG-3' (Πίνακας 3), που στοχεύει το τελευταίο εξόνιο του SCAF1.

Έπειτα, στα προϊόντα της παραπάνω αντίδρασης πραγματοποιήθηκε αραίωση 1:30 με νερό ελεύθερο από νουκλεάσες (nuclease-free water) και στη συνέχεια τα αραιωμένα δείγματα χρησιμοποιήθηκαν ως εκμαγείο για την διεξαγωγή μιας εσωτερικής PCR αντίδρασης (nested PCR) προκειμένου να αυξηθεί η ειδικότητα της αντίδρασης για τα *SCAF1* mRNAs, δηλαδή να ελαχιστοποιηθεί η ύπαρξη μη ειδικών προϊόντων [186].

Έτσι, σχεδιάστηκε ένας δεύτερος ειδικός για το *SCAF1* πρόσθιος εκκινητής με αλληλουχία 5'-GAGGGAAGACAGAGGAGTCGG-3' (Πίνακας 3), που στόχευε έπειτα από έξι βάσεις από τον πρώτο πρόσθιο εκκινητή. Παράλληλα σχεδιάστηκε και ένας δεύτερος, λίγο πιο εσωτερικός ανάστροφος εκκινητής με αλληλουχία 5'-CAAAGAGGTGAGGGGGGAAG-3' (Πίνακας 3). **Πίνακας 2**: Πρωτόκολλο και συνθήκες αντιδράσεων PCR για τη μοριακή κλωνοποίηση των *SCAF1* μεταγράφων.

Πρωτόκολλο			Συνθήκες αντίδρασης			
Αντιδραστήρια	Αρχική συγκέντρωση	Όγκος (μl)	Στάδιο	Θερμ/σία (ºC)	Χρόνος	Κύκλοι
PCR-GRADE H ₂ 0	-	Μέχρι τα 25μl	Αρχική αποδιάταξη	95	3 min	1
10X KAPA Taq Buffer	10X	2,5				
dNTPs	10 mM	0,5	Αποδιάταξη	95	30 sec —	
MgCl ₂	-	Περιείχε το Buffer	Υβριδ/ση	60	30 sec	x35
Πρόσθιος εκκινητής	10µM	1	Επέκταση	72	1,1 min	J
Ανάστροφος εκκινητής	10µM	1				
ΚΑΡΑ Ταq πολυμεράση	5U/μL	0,2	Τελική επέκταση	72	1,1min	1
Σύνολο		25µL				

Πίνακας 3: Καταγραφή των εκκινητών που χρησιμοποιήθηκαν για τη μοριακή κλωνοποίηση των *SCAF1* μεταγράφων. Οι θερμοκρασίες τήξεως των εκκινητών (Tm) υπολογίστηκαν μέσω Primer-BLAST.

		Εκκινητές			
		Κατεύθυνση	Αλληλουχία (5΄→3΄)	Μήκος (nt)	T _m (°C)
ακή ποίηση	PCR	Πρόσθιος	GGTGACCATGGAGGAAGAAGATG	23	60.68
		Ανάστροφος	GTCCAGAGTTTCAAAGAGGTGAGG	24	61.04
Mop	Nested	Πρόσθιος	GAGGGAAGACAGAGGAGTCGG	21	61.29
кλ	PCR	Ανάστροφος	CAAAGAGGTGAGGGGGCGAAG	20	60.67

2.4.4. Αλυσιδωτή Αντίδραση Ταχείας Ενίσχυσης 3'-Άκρου (3'-RACE PCR)

Τα απαιτούμενα αντιδραστήρια και διαλύματα που χρησιμοποιήθηκαν για τη διεξαγωγή της αλυσιδωτής αντίδρασης πολυμεράσης ήταν τα εξής:

Η20 ελεύθερο από (δεοξυ)ριβονουκλεάσες (RNase/DNase-free H2O)

• KAPA Taq buffer A (Kapa Biosystems Inc.), το οποίο περιείχε MgCl2 σε τελική συγκέντρωση 1,5 mM

- Μείγμα τριφωσφορικών δεοξυριβονουκλεοτιδίων (dNTPs) 0,2mM
- Εκκινητές (ολιγονουκλεοτίδια) 0,4μM ο καθένας
- KAPA Taq DNA Polymerase (Kapa Biosystems) 1 unit
- Θερμικός κυκλοποιητής (Applied Biosystems Veriti™ Thermal Cycler)

Η 3'-RACE PCR (3' Rapid Amplification of cDNA ends PCR), είναι μια τεχνική που βασίζεται στην αλυσιδωτή αντίδραση πολυμεράσης (PCR) [187] και παρέχει ένα σχετικά φθηνό αλλά ταυτόχρονα δυνατό εργαλείο, για τη γρήγορη απόκτηση του πλήρους μήκους του cDNA, όταν η αλληλουχία είναι μόνο μερικώς γνωστή [188].

Η τεχνική της 3'-RACE PCR χρησιμοποιήθηκε για την ενίσχυση των 3'άκρων των cDNAs τμημάτων που προέκυψαν από την διαδικασία της RT-PCR. Με την ανωτέρω τεχνική, έχουμε τη δυνατότητα να χρησιμοποιήσουμε ως θέση εκκίνησης τον oligo-dT αντάπτορα, ο οποίος είναι υβριδοποιημένος στην poly-(A) ουρά των 3΄ άκρων των mRNAs (Εικόνα 24).

Η αλληλουχία του γενικευμένου ανάστροφου εκκινητή (reverse primer) που χρησιμοποιήθηκε για να εντοπίσει την αλληλουχία που υπάρχει στον oligodT αντάπτορα, είναι 5'-GCGAGCACAGAATTAATACGACT-3' (Πίνακας 4). Ωστόσο, για να εξασφαλίσουμε ειδικότητα μόνο για τα *SCAF1* mRNAs, σχεδιάστηκε ειδικός πρόσθιος εκκινητής (forward primer) με αλληλουχία 5'-AAGGTGAGCAACCTGGTGC-3' (Πίνακας 4), η οποία στόχευε στο τελευταίο εξόνιο του *SCAF1*. Έτσι η ενίσχυση του επιθυμητού τμήματος, από το τέλος της κωδικής αλληλουχίας, μέχρι το 3' άκρο του mRNA για το *SCAF1* κατέστη εφικτή.

Στη συνέχεια, στα προϊόντα της παραπάνω αντίδρασης πραγματοποιήθηκε αραίωση 1:50 σε νερό ελεύθερο νουκλεασών και έπειτα τα αραιωμένα δείγματα χρησιμοποιήθηκαν ως εκμαγείο για την διεξαγωγή μιας εσωτερικής PCR αντίδρασης (nested 3'-RACE). Ο λόγος της επιλογής της δεύτερης εσωτερικής PCR (nested 3'-RACE), ήταν για να αυξηθεί η ειδικότητα της αντίδρασης για τα *SCAF1* mRNAs, δηλαδή να ελαχιστοποιηθεί η ύπαρξη μη ειδικών προϊόντων [186].

Απαραίτητος ήταν ο σχεδιασμός και ενός δεύτερου ειδικού για το SCAF1 πρόσθιου εκκινητή με αλληλουχία 5'-GCCTACGTCCAGCGCTAC-3' (Πίνακας 4), που στόχευε έπειτα από δυο βάσεις από τον πρώτο πρόσθιο εκκινητή. Παράλληλα σχεδιάστηκε και ένας δεύτερος, λίγο πιο εσωτερικός, γενικευμένος ανάστροφος εκκινητής με αλληλουχία 5'-AGCACAGAATTAATACGACTCACTATAGG -3' (Πίνακας 4).

3'-RACE	Εκκινητές						
		Κατεύθυνση	Αλληλουχία (5'-3')	Μήκος (nt)	Т _т (°С)		
	3'-RACE	Πρόσθιος	AAGGTGAGCAACCTGGTGC	19	60.53		
		Ανάστροφος	GCGAGCACAGAATTAATACGACT	23	59.20		
	Nested 3'-RACE	Πρόσθιος	GCCTACGTCCAGCGCTAC	18	60.28		
		Ανάστροφος	AGCACAGAATTAATACGACTCACTATAGG	29	60.73		

Πίνακας 4: Καταγραφή των εκκινητών που χρησιμοποιήθηκαν για την 3' – RACE PCR. Οι θερμοκρασίες τήξεως των εκκινητών (Tm) υπολογίστηκαν μέσω Primer-BLAST. Οι συνθήκες της αντίδρασης περιλάμβαναν αρχική αποδιάταξη στους 95°C για 3 λεπτά και ακολουθούν 30 κύκλοι ενίσχυσης όπου έχουμε διαδοχικά: αποδιάταξη στους 95°C για 30 δευτερόλεπτα, υβριδοποίηση στους 60°C για 30 δευτερόλεπτα, επέκταση στους 72°C για 5 λεπτά και τελική επέκταση στους 72°C για 5 λεπτά

Στην εσωτερική PCR χρησιμοποιήθηκαν οι ανωτέρω συνθήκες αντίδρασης, με τη διαφορά ότι ακολούθησαν 40 κύκλοι ενίσχυσης. Έπειτα ίσοι όγκοι των nested PCR προϊόντων ηλεκτροφορήθηκαν σε πήκτωμα αγαρόζης (Cambrex Bio Science Rockland Inc., Rockland, ME, USA) και παρατηρήθηκαν με τη βοήθεια της UV και της χρώσης του βρωμιούχου αιθιδίου. Οι ζώνες αφαιρέθηκαν κατάλληλα και καθαρίστηκαν με χρήση του PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG) και αλληλουχήθηκαν κατά Sanger για την εξακρίβωση της αλληλουχίας των αμπλικονίων. Σε όλα τα αμλικόνια, οι αλληλουχήσεις κατά Sanger πραγματοποιήθηκαν χρησιμοποιώντας τον πρόσθιο ειδικό για το γονίδιο εκκινητή με αλληλουχία 5'-GCCTACGTCCAGCGCTAC-3' (Πίνακας 4).



Εικόνα 24: Σχηματική παρουσίαση της τεχνικής 3'-RACE PCR (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/3prime_race_man.pdf).

2.4.5. Καθαρισμός προϊόντων PCR

Για τον καθαρισμό όλων των προϊόντων της εσωτερικής (nested) PCR, τα προϊόντα αυτά συγκεντρώθηκαν σε ένα τελικό δείγμα για το οποίο χρησιμοποιήθηκε το σύστημα καθαρισμού NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, Duren, Germany). Ο σκοπός του καθαρισμού είναι να απομακρυνθούν εκκινητές, νουκλεοτίδια που δεν ενσωματώθηκαν κατά τη διαδικασία ενίσχυσης, ένζυμα και άλατα, από τα τελικά προϊόντα [186].

Αρχικά, πραγματοποιείται η πρόσδεση του DNA στη μεμβράνη πυριτίου που υπάρχει στην κολώνα καθαρισμού. Η πρόσδεση αυτή είναι εφικτή λόγω της παρουσίας παραγόντων, που επεμβαίνουν μεταξύ των μη ομοιοπολικών αλληλεπιδράσεων που δημιουργούνται μεταξύ DNA και νερού, καθώς και ενός ειδικού ρυθμιστικού διαλύματος NT1 Binding Buffer.

Η απομάκρυνση άχρηστων προϊόντων πραγματοποιείται στο επόμενο βήμα, όπου γίνεται έκπλυση με ρυθμιστικό διάλυμα που περιέχει αιθανόλη, NT3 Wash Buffer.

Στο τέλος, έχουμε την έκλουση υπερκάθαρου DNA από τη μεμβράνη πυριτίου, με χρήση ενός ελαφρώς αλκαλικού ρυθμιστικού διαλύματος (Elution Buffer NE, 5 mM Tris/HCl, pH 8.5) σε συνθήκες χαμηλής αλατότητας [189].

Μετά τη διαδικασία του καθαρισμού, καθορίζεται φασματοφωτομετρικά στα 260 και 280 nm, η συγκέντρωση και η καθαρότητα του τελικού προϊόντος και αποθηκεύεται σε θερμοκρασία -20°C.

Εικόνα 25: Βασικά στάδια καθαρισμού PCR προϊόντος (http://www.genemarkbio.com/1.hopegenbio/images/stories/products/1-2-2-PCR-Clean-up-kit/1-2-2-GERAL-PROCEDURE-PCR-Clean-up-kit..png).



2.5. ΚΑΤΑΣΚΕΥΗ ΒΙΒΛΙΟΘΗΚΗΣ

Μέχρι το στάδιο αυτό έχουν απομονωθεί τα επιθυμητά προς αλληλούχηση προϊόντα PCR και έχουν επίσης καθαριστεί αφού αυτά συγκεντρώθηκαν σε ένα τελικό δείγμα στο οποίο χρησιμοποιήθηκε το σύστημα καθαρισμού NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, Duren, Germany). Mia αρχική ποσότητα των 100ng από το καθαρισμένο προϊόν PCR χρησιμοποιήθηκε για την κατασκευή βιβλιοθήκης, η οποία χρησιμοποιήθηκε μετέπειτα στην αλληλούχηση νέας γενιάς. Η προετοιμασία βιβλιοθήκης NGS της πραγματοποιήθηκε με χρήση του συστήματος Ion Xpress™ Plus Fragment Library Kit (Applied Biosystems). Το εν λόγω Kit περιέχει τα παρακάτω αντιδραστήρια (Ion Shear[™] Plus Reagents):

- Ion Shear[™] Plus 10X Reaction Buffer
- Ion Shear[™] Plus Enzyme Mix II
- Ion Shear[™] Plus Stop Buffer
- Low TE

Ο στόχος μας σε αυτό το σημείο είναι να δημιουργήσουμε σε πρώτη φάση δίκλωνα θραύσματα με επιθυμητό μέγεθος περίπου 400 ζεύγη βάσεων. Τα θραύσματα αυτά προκύπτουν ύστερα από ενζυμική επίδραση στα PCR προϊόντα που επιτυγχάνεται με τη χρήση των αντιδραστηρίων Ion Shear[™] Plus Reagents. Αφού ακολουθήσει ο καθαρισμός του θρυμματισμένου (fragmented) DNA, έπειτα, στα άκρα των δίκλωνων θραυσμάτων συνδέουμε δύο ειδών ολιγονουκλεοτιδικά τμήματα καθορισμένης αλληλουχίας (adapters), τα A και P1 (Εικόνα 26) [190], σύμφωνα με το πρωτόκολλο του κατασκευαστή. Η σύνδεση αυτή επιτυγχάνεται εκμεταλλευόμενοι τη δράση της λιγάσης. Η δράση αυτού του ενζύμου έγκειται στη δημιουργία φωσφοδιεστερικού δεσμού μεταξύ του 3' υδροξυλικού άκρου ενός μορίου DNA με το 5' φωσφορικό άκρο ενός δεύτερου μορίου [191]. Στη συνέχεια θα γίνει η επιδιόρθωση των τμημάτων με χρήση της Nick Repair πολυμεράσης (nickrepair).

Στο σημείο αυτό είναι απαραίτητο να γίνει η επιλογή μόνο εκείνων των τμημάτων που φέρουν το επιθυμητό μέγεθος (400-base-read library) διαδικασία που πραγματοποιείται ηλεκτροφορητικά με τη χρήση του E-Gel® SizeSelectTM 2% Agarose Gel (Applied Biosystems). Η επιλογή αυτή είναι απαραίτητη να πραγματοποιηθεί, καθώς κατά την ενζυμική πέψη, παρόλο που έγινε για συγκεκριμένο χρόνο προκειμένου να παραχθούν θραύσματα επιθυμητού μήκους, μπορούν να παραχθούν και πολλά άλλα, είτε μεγαλύτερου, είτε μικρότερου μήκους θραύσματα, τα οποία δε θέλουμε να τα συμπεριλάβουμε στην αλληλούχηση που θα ακολουθήσει [190].

Τέλος, η ποσοτικοποίηση της βιβλιοθήκης η οποία περιέχει τα τμήματα επιθυμητού μεγέθους πραγματοποιήθηκε με χρήση του Ion Library TaqMan[™] Quantitation Kit (Applied Biosystems) στο σύστημα ABI 7500 Fast Real-Time PCR (Applied Biosystems)



Εικόνα 26: Σχηματική απεικόνιση διαδικασίας κατασκευής βιβλιοθήκης (https://tools.thermofisher.com/content/sfs/manuals/MAN0009847_IonXpressPl us_gDNA_FragLibraryPrep_UG.pdf).
2.6. ΠΡΟΕΤΟΙΜΑΣΙΑ ΕΚΜΑΓΕΙΟΥ ΑΛΛΗΛΟΥΧΗΣΗΣ, ΕΜΠΛΟΥΤΙΣΜΟΣ ΚΑΙ ΑΛΛΗΛΟΥΧΗΣΗ ΝΕΑΣ ΓΕΝΙΑΣ (TEMPLATE PREPARATION, ENRICHMENT, AND NEXT-GENERATION SEQUENCING)

Μετά την κατασκευή βιβλιοθήκης ακολούθησε η κατασκευή του εκμαγείου αλληλούχησης (Template preparation), όπου χρησιμοποιήθηκαν τα παρακάτω αντιδραστήρια και μηχανήματα:

- ✓ Ion PGM[™] Template OT2 Solutions 400 Kit:
 - Ion PGM[™] Template OT2 400 PCR Reagent B
 - Ion OneTouch[™] Wash Solution
 - Ion OneTouch[™] Reaction Oil
 - Ion OneTouch[™] Recovery Solution
 - Nuclease-Free Water
- ✓ Ion PGM[™] Template OT2 Reagents 400 Kit:
 - Ion PGM[™] Template OT2 400 Reagent Mix
 - Ion PGM[™] Template OT2 400 Enzyme Mix
 - Ion PGM[™] Template OT2 400 Reagent X
 - Ion PGM[™] Template OT2 400 Ion Sphere[™] Particles
 - Ion PGM[™] OneTouch Plus Reaction Filter Assembly
- ✓ Ion OneTouch[™] 2 instrument

To Ion OneTouch[™] 2 Instrument (Εικόνα 27) πραγματοποιεί την ενίσχυση της βιβλιοθήκης μέσω της μεθόδου αλυσιδωτής της αντίδρασης πολυμεράσης σε γαλάκτωμα (emulsion PCR), η οποία παρουσιάζει διαφοροποιήσεις σε σχέση με την κλασσική PCR. Η αντίδραση αλυσιδωτή πολυμεράσης σε γαλάκτωμα είναι μια



τίδραση
 Εικόνα 27: Ion OneTouch™ 2 Instrument
 άκτωμα
 (http://tools.thermofisher.com/content/sfs/prodIm
 ευρέως ages/high/Ion%20OneTouch%202.jpg)

χρησιμοποιούμενη μέθοδος για ενίσχυση του εκμαγείου αλληλούχησης σε πολλές πλατφόρμες αλληλούχησης νέας γενιάς συμπεριλαμβανομένων των Ion PGM/Proton, Roche/454, Life/APG, SOLiD, και Polonator. Η Illumina/Solexa χρησιμοποιεί μια εναλλακτική τεχνολογία που ονομάζεται ενίσχυση στερεής φάσης «bridge amplification» [192, 193].

Η EmPCR είναι βασισμένη στη διαμερισματοποίηση των θραυσμάτων DNA σε πολύ μικρά κυστίδια νερού, μέσα σε ένα γαλάκτωμα νερού και ελαίου (Ion OneTouch™ Reaction Oil και Nuclease-Free Water) [194]. Έτσι είναι δυνατή η δημιουργία ενός μεγάλου αριθμού μικροαντιδραστήρων, όπου μπορεί να πολλαπλασιάζεται εκλεκτικά ένας μόνο κλώνος της βιβλιοθήκης.

Η PCR αντίδραση μπορεί να γίνει στο εσωτερικό των προαναφερόμενων κυστιδίων, το οποίο λειτουργεί ως ένας απομονωμένος χώρος (μικροαντιδραστήρας). Θεωρητικά, μέσα σε κάθε μικροαντιδραστήρα θα υπάρχει κάθε φορά: α) ένα σφαιρίδιο που φέρει αγκυροβολημένους εκκινητές (ISP, Ion Sphere[™] Particles), που υβριδοποιούνται με τα ολιγονουκλεοτιδικά τμήματα καθορισμένης αλληλουχίας (adapters) που φέρουν τα μόρια της βιβλιοθήκης, β) τα απαιτούμενα ένζυμα και αντιδραστήρια για την αντίδραση και γ) ένας κλώνος της βιβλιοθήκης, ώστε με το πέρας της ενίσχυσης να υπάρχουν στο σφαιρίδιο πολλαπλά αντίγραφα ενός μόνο κλώνου της βιβλιοθήκης (Εικόνα 28).

Ωστόσο, υπάρχει και η περίπτωση δημιουργίας πολυκλωνικού σφαιριδίου όταν υπάρξουν περισσότεροι του ενός κλώνοι της βιβλιοθήκης μέσα στον μικροαντιδραστήρα. Τα πολυκλωνικά σφαιρίδια έχουν αρνητική επίδραση στην αλληλούχηση, καθώς δίνουν πολλαπλό σήμα, το οποίο πρακτικά δεν μπορεί να ερμηνευθεί [195].

Μετά την προετοιμασία του εκμαγείου αλληλούχησης, μέσω της διαδικασίας emPCR, ελέγχουμε φθορισμομετρικά το προϊόν που έχουμε παραλάβει, ώστε να διαπιστωθεί αν είναι κατάλληλο για να προχωρήσει σε αλληλούχηση.

Για τον παραπάνω έλεγχο χρησιμοποιούμε:

- Ion Sphere[™] Quality Control Kit
- Qubit® 2.0 Fluorometer



Εικόνα 28: Σχηματική απεικόνιση αρχής μεθόδου emulsion PCR

(https://users.ugent.be/~avierstr/nextgen/emulsionpcr_powerpoint.jpg).

Η ποιότητα του συγκεκριμένου εκμαγείου αλληλούχησης εκτιμάται φθορισμομετρικά στο Qubit[®] 2.0 Fluorometer (Invitrogen[™]) χρησιμοποιώντας δύο χρωστικές τις: Alexa Fluor[®] 647 και Alexa Fluor[®] 488. Το Qubit[®] 2.0 Fluorometer είναι μια μικρή συσκευή (Εικόνα 29), η οποία γενικά χρησιμοποιεί φθορίζουσες χρωστικές για τον ακριβή προσδιορισμό των DNA, RNA και πρωτεϊνών [196-199].



Εικόνα 29: Qubit® 2.0 Fluorometer

(https://en.wikipedia.org/wiki/Q ubit_fluorometer#/media/File:S0 03771_NoSkin.jpg) Η οικογένεια χρωστικών Alexa Fluor παράγονται από την Thermo Fisher Scientific και τα φάσματα εκπομπής και διέγερσης των χρωστικών αυτών, καλύπτουν το φάσμα του ορατού φωτός και εκτείνονται στο υπέρυθρο [190].

Η χρωστική Alexa Fluor® 488 σημαίνει την αλληλουχία που βρίσκεται αγκυροβολημένη πάνω στο ISP (εκκινητής ISP), καθώς και τα τμήματα που ενισχύονται πάνω στους αγκυροβολημένους εκκινητές των ISPs, που θα αποτελέσουν το εκμαγείο αλληλούχησης.

Η χρωστική Alexa Fluor® 647 υπάρχει μόνο στις περιοχές που δημιουργούνται πάνω στα ISPs. Επομένως μόνο τα σφαιρίδια στα οποία ενισχύονται τα τμήματα βιβλιοθήκης, φέρουν αυτήν τη χρωστική.

Το ποσοστό των σφαιριδίων ISPs που φέρουν πάνω τους εκμαγείο, καθορίζεται από το λόγο αυτών των δύο χρωστικών (Alexa Fluor® 647 και Alexa Fluor® 488). Ένα ποσοστό που κυμαίνεται μεταξύ 10-30% των σφαιριδίων που φέρουν εκμαγείο, είναι ικανοποιητικό (ανάλογα και με τις απαιτήσεις του κάθε πειράματος) για να προσχωρήσουμε αργότερα σε αλληλούχηση. Ποσοστά μικρότερα του 15% δείχνουν πολλά κενά ISPs, ενώ ποσοστά μεγαλύτερα του 30% δείχνουν μεγάλο αριθμό πολυκλωνικών ISPs και είναι πολύ πιθανό να μη δώσουν τα επιθυμητά αποτελέσματα. Στην περίπτωση που διαπιστώσουμε ότι το εκμαγείο αλληλούχησης (template) είναι κατάλληλο, μπορούμε να προχωρήσουμε στο επόμενο βήμα, το οποίο είναι ο εμπλουτισμός του εκμαγείου αλληλούχησης. Για τον εμπλουτισμό του εκμαγείου χρησιμοποιούμε:

- Ion OneTouch[™] Wash Solution
- MyOne[™] Beads Wash Solution
- Nuclease Free Water
- Dynabeads® MyOne [™] Streptavidin C1 Beads
- Ion OneTouch[™] ES (Εικόνα 30)



Εικόνα 30: Ion OneTouch[™] ES (αριστερά) (http://tools.thermofisher.com/content/sfs/prodImages/high/Ion %20OneTouch%202%20System.jpg).

Στο Ion OneTouch™ ES επιτελείται με αυτοματοποιημένο τρόπο, η διαδικασία εμπλουτισμού του εκμαγείου αλληλούχησης όπου απομακρύνονται τα κενά ISPs, εκείνα δηλαδή στα οποία δεν έχει γίνει ενίσχυση της βιβλιοθήκης μας. Αναφέρεται πως τα ISPs τα οποία φέρουν εκμαγείο, φέρουν τα (πολλαπλασιασμένα) θραύσματα της βιβλιοθήκης μας, βιοτινυλιωμένα από τη μία πλευρά. Έτσι, χρησιμοποιούνται μαγνητικά σφαιρίδια που είναι καλυμμένα με στρεπταβιδίνη (Dynabeads[®] MyOne[™] Streptavidin C1 Beads) (Εικόνα 31). Με την προσέγγιση αυτή, είναι εφικτή η θετική επιλογή των φορτωμένων ISPs με εκμαγείο, λόγω του συστήματος βιοτίνης - στρεπταβιδίνης. Το αποτέλεσμα είναι ο διαχωρισμός με τη χρήση μαγνήτη των καλυμμένων με στρεπταβιδίνη μαγνητικών σφαιριδίων, τα οποία έχουν δημιουργήσει σύμπλοκο με τα ISPs που φέρουν βιοτινυλιωμένο από τη μια πλευρά εκμαγείο (Εικόνα 31). Έπειτα ακολουθεί έκπλυση και απομάκρυνση των άδειων ISPs.





Τη διαδικασία του εμπλουτισμού του εκμαγείου αλληλούχησης διαδέχεται η αλληλούχηση νέας γενιάς σε ημιαγωγό, στο σύστημα Ion Torrent Personal Genome Machine[™] (PGM) (Εικόνα 32) με χρήση των αντιδραστηρίων που περιλαμβάνονται στο Ion PGM[™] Sequencing 400 Kit.

Αρχικά, απαιτείται ο εφοδιασμός του Ion Torrent Personal Genome Machine[™] με τα κατάλληλα αντιδραστήρια (ρυθμιστικό διάλυμα κατάλληλου pH και dNTPs). Έπειτα, προτού εισαχθεί το προς αλληλούχηση εκμαγείο (template), πραγματοποιείται ο αυτοματοποιημένος έλεγχος και ρύθμιση των συστημάτων του ανωτέρω μηχανήματος. Το εμπλουτισμένο εκμαγείο εισάγεται σε chip (Ion 316[™] Chip v2 για τους σκοπούς της παρούσας εργασίας) αποτελούμενο από εκατομμύρια πηγαδάκια – μικροαισθητήρες, που ανιχνεύουν τη μεταβολή του pH, εξαιτίας της έκλυσης ενός ή περισσότερων ιόντων υδρογόνου (Η+), έπειτα από την ενσωμάτωση μιας συγκεκριμένης βάσης. Με βάση λοιπόν αυτήν την ανιχνεύσιμη μεταβολή του pH, καθίσταται εφικτή η καταγραφή της ακολουθίας των βάσεων.



Εικόνα 32: Διάγραμμα απεικόνισης πυκνότητας – κάλυψης του chip, ύστερα από φόρτωμα δείγματος (αριστερά), σύστημα Ion Torrent Personal Genome Machine[™] (PGM) (δεξιά)

(https://media.licdn.com/mpr/Mpr/AAEAAQAAAAAAAAAAAAAAAJGI1YjBhYzEwLTVjNzA tNDYxNS1hNDBkLTljMTQ2MzY5NzEyNw.png).

2.7. ANAAYSH TQN $\Delta E \Delta O M E N Q N AAAHAOYXHSHS NEAS FENIAS$

Με την ολοκλήρωση της αλληλούχησης νέας γενιάς ακολουθεί η συλλογή και η ανάλυση των δεδομένων που παρήχθησαν. Ο αλληλουχητής, εκτός των όσων έχουν αναφερθεί, συνοδεύεται και από κατάλληλο λογισμικό Torrent Suite[™] software (Ion Torrent[™]), το οποίο επιτρέπει αφενός την οργάνωση και παρακολούθηση της πορείας αλληλούχησης μέσω ενός ηλεκτρονικού υπολογιστή (Torrent server), αφετέρου τη στοίχιση των δεδομένων (raw data) στο ανθρώπινο γονιδίωμα (GRCh38) που χρησιμοποιήθηκε ως γονιδίωμα αναφοράς και τέλος τη διαχείριση των δεδομένων με το πέρας της αλληλούχησης. Το λογισμικό αυτό παρέχεται από την εταιρεία Applied Biosystems.

Στα δεδομένα που παρήχθησαν έχουμε πρόσβαση μέσω μιας ειδικής πλατφόρμας περιήγησης (Torrent browser) και μπορούμε να τα μεταφέρουμε στον ηλεκτρονικό μας υπολογιστή ως αρχείο τύπου FASTQ (.fastq). Πρόκειται για ένα αρχείο κειμένου που αποθηκεύονται τα δεδομένα αλληλουχιών με τα quality scores τους και χρησιμοποιείται από πολλές πλατφόρμες αλληλούχησης [200, 201].

Τα αρχεία FASTQ περιλαμβάνουν ένα επαναλαμβανόμενο μοτίβο τεσσάρων (4) γραμμών (Εικόνα 33). Κάθε μια από αυτές τις τετράδες, αντιστοιχεί σε ένα ανάγνωσμα - read. Η πρώτη εκ των τεσσάρων γραμμών έχει ως πρώτο χαρακτήρα το σύμβολο «@», το οποίο ακολουθείται από τον αναγνωριστικό κωδικό της ακολουθίας (Εικόνα 33 κίτρινο χρώμα), η δεύτερη γραμμή περιλαμβάνει την αλληλουχία του read (Εικόνα 33 πράσινο χρώμα), η τρίτη γραμμή αποτελείται από το σύμβολο «+» (Εικόνα 33 πράσινο χρώμα), η τρίτη γραμμή περιλαμβάνει τα quality scores κάθε βάσης (Εικόνα 33 κόκκινο χρώμα) [200]. Η ποικιλία των συμβόλων που μπορούμε να συναντήσουμε στην τέταρτη γραμμή είναι τα: !"#\$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNO PQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~ (ASCII). Τα σύμβολα παρατέθηκαν με σειρά αυξανόμενου quality score, έτσι ο χαρακτήρας «-» το μεγαλύτερο quality score.

Το αρχείο FASTQ που καλούμαστε να διαχειριστούμε δεν περιέχει ακριβώς τα ακατέργαστα δεδομένα (raw data) όπως παρήχθησαν, αλλά τα δεδομένα που προκύπτουν έπειτα από αφαίρεση των αλληλουχιών των ολιγονουκλεοτιδικών τμημάτων καθορισμένης αλληλουχίας (adapters), τα οποία χρησιμοποιήθηκαν όπως έχει αναφερθεί κατά την κατασκευή της βιβλιοθήκης. Η επεξεργασία αυτή πραγματοποιείται μέσω ηλεκτρονικού υπολογιστή (Torrent server).

Ένα αρχείο FASTQ λόγω του εύρους και της πολυπλοκότητας της πληροφορίας που παρουσιάζει, είναι δύσκολο έως αδύνατο να αναλυθεί χωρίς τη χρήση υπολογιστικών συστημάτων. Για τον λόγο αυτό, έχουν αναπτυχθεί ειδικές πλατφόρμες επεξεργασίας και ανάλυσης δεδομένων που επιτρέπουν μεταξύ άλλων οπτικοποίηση στατιστικών που σχετίζονται με τα quality scores, φιλτράρισμα των reads με βάση το quality score ή άλλων παραμέτρων, στοίχιση των δεδομένων στο γονιδίωμα αναφοράς, μελέτη γονιδιακής έκφρασης, μελέτη μεταγραφώματος και πολλές ακόμη λειτουργίες [202].

Μια τέτοια πλατφόρμα αποτελεί και η ανοιχτού κώδικα, διαδικτυακή πλατφόρμα Galaxy [203], η οποία χρησιμοποιήθηκε στην επεξεργασία και ανάλυση των δεδομένων της παρούσας εργασίας και δεν έχει περιορισμούς στο μέγεθος των δεδομένων που μπορεί να διαχειριστεί και παρέχει βιοπληροφορικά εργαλεία καθώς και ειδικούς αλγορίθμους για την ανάλυση των δεδομένων αλληλούχησης νέας γενιάς [202].

@N40RV:00004:00038

TTCTGCCTTTATCCTGCGAGCCATCCAGCAGGCTGTGGGAAGCTCCCTGCAGGGGGGACCTGCCCAACGATAAA
GATGGCTCTCGGTGTCATGGCCTCGATGGCGGCGGCTGCCGGAGTCCACGGTCAGAGCCCCGTTCCAGGAAT
CAGGGGGCACTGACAGCGGCTACTGTGTTGGACATGGCCACGGACGAGCTTCCTCC
•
56:::505C:=::0:6568::0::5:5555555555686@65556=66666666666666666666
<=E=:1A+1+/826606606665>6>65660666066666666666666666
5::2:2:555505<@550533336<600*
@N40RV:00004:00039
CCGCCCGCTCCTGCGTGTGCAGCTTCTTCAGATACCGTTTTGCCCGTCCTCCAGGCCCTGAGACGGGGCAGCA
GGCTGGGCACAGGCAGAGGGATGGGATCTCTGTCCCCGATTCGGACCTCAGCCACCAGCTCCAGCATGTCCT
CACTTTCGCCAGGCCGCAGGTCCAGGCGGCTGGGAACCCAGGTATCCGGGGGGATCCAGGACACTCACCAGCT
GCGAGGAAGCTGTCCGTGGCAT
•
D@=<<4<:AA@DBBBDEEDCCEFDAAC>BD===BB@DEDD;DDD>DEEADEADD@DD>DEEDDBBCB;CDDDE
EADDDE?DDCCEAEEEEDE?DDDD?DBBBDDDDDDD;DDD>BB@<<9=DDDEADEBDDDEE?DDDDDDDDD@
DEDDDC:AAC@DD@D;==DC@ED?DD@EE=BCDD=D=BB<8D?DDBB29999-
9CC?DD@== ? ::3:::::BFEF@EAEG>>?K9AAC+01>
@N40RV:00005:00032
GGGTGGGCAGGGCCCCCGTCCCCGTGCCCAAGGTGAGGTGACGGGCACAGGTGGGCCTCGAATCGAATGT
AACAAGCTAACGTACGTCTGAGCAATCGAATCGATGTACA
-
://55,55:5;;;055.@@B0:::::1:4:4:???4::66:6,666::4::::06666606666666@@A=B?6?;::05=:555???<8-
.;/:55,55::,5;;;055:@@B0:::::1:4:4:???4::66:6,666::4:::,:06666666666666@@A=B?6?:::05=:555???<8— ?<=@=;};
::/:55,55::.5;;;055:@@B0:::::1:4:4:???4::66:6,666::4:::::066666666666@@A=B?6?:::05=:555???<8 ?<=@=;);: @N40RV:00005:00039
::/:55.55::.5:::055:@@B0:::::1:4:4:???4::66:6,666::4::::0666666666666@@A=B?6?:::05=:555???<8- ?<=@=;): @N40RV:00005:00039 CCTGGCTTGCGACCGTGCTTGCGGAAGTAGCGGTAGCGCTGGACGTAGGCCCGCACCAGGTTGTTCGCCTTC
ar/155.55::.5:::055:@@B0:::::1:4:4:???4::66:6,666::4:::::066666666666@@A=B?6?:::05=:555???<8 ?<=@=;}; @N40RV:00005:00039 CCTGGCTTGCGACCGTGCTTGCGGAAGTAGCGGTAGCGCTGGACGTAGGCCCGCACCAGGTTGTTCGCCTTC ACTGGGTTGATTTCCCCACTTTTGCTGTGGCAGATCTTGTGGACGGCCTTCCTCAGGATGTCCTTGTACTCCTCC
TIGGTGATG
*/:55,55::,5::,055:@@B0:::::1:4:4:???4::66:6,666::4::::,06656606666666@@A=B?6?:::05=:555???<8 ?<=@=:}: @N40RV:00005:00039 CCTGGCTTGCGACCGTGCTTGCGGAAGTAGCGGTAGCGCTGGACGTAGGCCCGCACCAGGTTGTTCGCCTTC ACTGGGTTGATTTCCCCACTTTTGCTGTGGCAGATCTTGTGGACGGCCTTCCTCAGGATGTCCTTGTACTCCTCC TTGGTGATG +
<pre>virting content is a set of the set of</pre>
<pre></pre>
<pre></pre>
<pre></pre>
<pre></pre>
<pre>//:55,55:5:::055:@@B0:::::1.4:4:???4::66:6,666::4::::0655606656606@@A=B?6?:::05=:555???<8- ?<=@=:)</pre>
<pre>//:55,55:5:::055:@@B0:::::1:4:4:???4::66:6,666::4::::06656606666666@@A=B?6?:::05=:555???<8- /<=@=:)</pre>
<pre> //:55,55::.5:;;;055:@@B0::::1:4:4:???4::66:6,666::4::::0666660666660@@A=B?6?:::05=:555???<8- ?<=@=:):; @N40RV:00005:00039 CCTGGCTTGCGACCGTGCTTGCGGAAGTAGCGGTAGCGCTGGACGTAGGCCCGCACCAGGTTGTTCGCCTTC ACTGGGTTGATTTCCCCACTTTTGCTGTGGGCAGATCTTGTGGACGGCCTTCCTCAGGATGTCCTTGTACTCCTCC TTGGTGATG + B>BD@CC>BBCCC>CBDDD@CDD@D@CCAACCADDCD==D>BC>>E@EF?EEBB>BEAE=BBAECD?C@E BBBEE?E>CCBB8@@@4AA?BB;CDDFEG?DCCC=>=BDEE>:::3:4:<b=ccee@dddeebibfdcaaa3::4>F< EKECC @N40RV:00006:00005 CTTGATCTCGCGAGCATCAGCGGCTGTGGGAAGCTCCGTCGCCAGGGGACCTGGCCCAATGGATAAAGATGG CTCTCGGTGTCATGGCCTTGATGGCGGCGCTGGCCGGAGTCGACGGTCGAGCCCCGTCCAGATCAGGGGCCCA F+ </b=ccee@dddeebibfdcaaa3::4></pre>
<pre> i/.t55,55::.,5;;;055;@@B0:::::1:4:4:???4::66:6,666::4::::066660666606@@A=B?6?;::05=:555???<8- ?<=@=;};; @N40RV:00005:00039 CCTGGCTTGCGACCGTGCTGCGGAAGTAGCGGTAGCGCTGGACGTAGGCCCGCACCAGGTTGTTCGCCTTC ACTGGGTTGATTTCCCCACTTTTGCTGTGGCAGATCTTGTGGACGGCCTTCCTCAGGATGTCCTTGTACTCCTCC TTGGTGATG + B>BD@CC>BBCCC>CBDDD@CDD@D@CCAACCADDCDD>BC>>>E@EF?EEBB>BEAE=BBAECD?C@E BBBEE?F>CCBB8@@@4AA?BB;CDDFEG?DCCC=>=BDEF>:::3:4<b=ccee@dddeebibfdcaaa3::4:>F< EKECC @N40RV:00006:00005 CTTGATCTCGCGAGCATCAGCGGCTGTGGGAAGCTCCGTCGCCAGGGGACCTGGCCCAATGGATAAAGATGG CTCTCGGTGTCATGGCCTTGATGGCGGCGCGGGGCCGGAGGTCGACGGTCGAGCCCCGTCCAGATCAGGGGCCCA TGACACGG + B)88/5555505:55555555555555555555555555</b=ccee@dddeebibfdcaaa3::4:></pre>

Εικόνα 33: Μορφή αρχείου FASTQ. Απεικονίζονται 5 από τα εκατομμύρια reads που περιείχε το FASTQ αρχείο που αναλύθηκε.

Η χρήση της πλατφόρμας GALAXY ξεκίνησε εισάγοντας το αρχείο FASTQ που περιείχε τα αποτελέσματα από την αλληλούχηση νέας γενιάς. Στη συνέχεια, χρησιμοποιήθηκε ο αλγόριθμος Tophat2 [204], καθώς μας ενδιαφέρει η μελέτη των εναλλακτικών μεταγράφων του γονιδίου *SCAF1*. Ο εν λόγω αλγόριθμος επιτυγχάνει τη στοίχιση των αλληλουχηθέντων αναγνωσμάτων και τον προσδιορισμό νέων θέσεων συρραφής, με την απευθείας χαρτογράφηση σε γνωστά μετάγραφα [72, 204]. Από τη χρήση του Tophat2 προκύπτουν πολλά αρχεία στα οποία περιλαμβάνονται και τα αρχεία BAM και BED. Το αρχείο της μορφής BAM (.bam) (Binary Alignment/Map) περιέχει την ίδια, αλλά με διαφορετική διαμόρφωση, πληροφορία με το αρχείο SAM (.sam) (Sequence Alignment/Map). Το αρχείο BAM αποτελεί συμπιεσμένη μορφή του αρχείου SAM [205]. Η πληροφορία που περιέχεται αφορά αλληλουχίες στοιχισμένες στο γονιδίωμα αναφοράς [206]. Το αρχείο BED (.bed) (Browser Extensible Data) [207], περιέχει όλα τα πιθανά καθώς και τα γνωστά σημεία συρραφής.

Για την οπτικοποίηση των αρχείων BAM και BED χρησιμοποιήθηκε το πρόγραμμα Integrative Genomics Viewer (IGV) (Εικόνα 34), το οποίο



Εικόνα 34: Οπτικοποίηση των αρχείων BAM και BED, με χρήση του προγράμματος IGV. Τα αρχεία BAM και BED που έχουν δημιουργηθεί από τον αλγόριθμο Tophat2 έχουν φορτωθεί στον IGV. Έτσι, μπορούμε να δούμε τη στοίχιση των αναγνωσμάτων, που περιέχονται στο BAM αρχείο, με το γονιδίωμα αναφοράς, καθώς και τα γεγονότα εναλλακτικού ματίσματος, που περιέχονται στο BED αρχείο. αναπτύχθηκε το 2007 για τις ανάγκες του "The Cancer Genome Atlas" (TCGA) [208]. Πρόκειται για ένα φιλικό προς το χρήστη πρόγραμμα, το οποίο υποστηρίζει αρκετές μορφές αρχείων δεδομένων και επιτρέπει στον χρήστη να περιηγηθεί (σαν στο Google Maps) κατά μήκος του γονιδιώματος σε οποιοδήποτε επίπεδο λεπτομέρειας, ακόμα και μέχρι το επίπεδο νουκλεοτιδικής βάσης [209]. Με το παραπάνω πρόγραμμα έχουμε επιπλέον τη δυνατότητα να εντοπίσουμε τα σημεία που έχουμε πρότυπα εναλλακτικής συρραφής.

2.8. ΑΝΑΛΥΣΗ ΕΚΦΡΑΣΗΣ ΤΩΝ ΝΕΩΝ ΜΕΤΑΓΡΑΦΩΝ ΜΕ ΕΣΩΤΕΡΙΚΗ ΑΛΥΣΙΔΩΤΗ ΑΝΤΙΔΡΑΣΗ ΠΟΛΥΜΕΡΑΣΗΣ (nested RT-PCR)

Έχοντας ολοκληρώσει την ανάλυση των δεδομένων που προέκυψαν από την αλληλούχηση νέας γενιάς, το επόμενο βήμα είναι η διερεύνηση του προφίλ έκφρασης κάθε καινούριου μεταγράφου η οποία πραγματοποιήθηκε με χρήση της μεθόδου της εσωτερικής αλυσιδωτής αντίδρασης της πολυμεράσης (nested RT-PCR) και της ηλεκτροφόρησης σε πήκτωμα αγαρόζης.

Για να πετύχουμε το σκοπό μας, όλα τα cDNAs που προέρχονται από τις κυτταρικές σειρές οι οποίες προκύπτουν από τον ίδιο ιστό αναμίχθηκαν. Έτσι προέκυψαν δεκαεπτά ¨δεξαμενές¨ cDNA (cDNA pools) από: αδενοκαρκίνωμα του μαστού, μετάσταση καρκίνου μαστού στον πνεύμονα, καρκίνο των ωοθηκών, αδενοκαρκίνωμα του ενδομητρίου, καρκίνο του τραχήλου της μήτρας, καρκίνο του προστάτη, καρκίνο της ουροδόχου κύστης, καρκίνο του νεφρού, καρκίνο του παχέος εντέρου, γαστρικό αδενοκαρκίνωμα, ηπατοκυτταρικό καρκίνωμα, καρκίνο του εγκεφάλου, αδενοκαρκίνωμα του πνεύμονα, μελάνωμα, λέμφωμα, λευχαιμία και καρκίνωμα εκ πλακωδών κυττάρων κεφαλής και τραχήλου, ενώ δύο φυσιολογικοί ιστοί (φυσιολογικός εμβρυϊκός νεφρός και φυσιολογικό πάγκρεας) αντιπροσωπεύονται από cDNA μίας κυτταρικής σειράς. Επίσης σχεδιάστηκαν κατάλληλοι εκκινητές, χρησιμοποιώντας το διαδικτυακό πρόγραμμα Primer – Blast [210], για κάθε εύρημα που προέκυψε από την ανάλυση των δεδομένων του γονιδίου *SCAF1*.

Έπειτα, πραγματοποιήθηκε αντίδραση RT-PCR με όγκους αντίδρασης 25μL που περιείχαν: KAPA Taq Buffer A (Kapa Biosystems Inc.), το οποίο περιελάμβανε MgCl₂ σε τελική συγκέντρωση 1.5 mM, 0.2 mM dNTPs, 0.4 μM από τους εκκινητές που χρησιμοποιήθηκαν κατά τη πρώτη PCR στη διαδικασία της μοριακής κλωνοποίησης (Πίνακας 3), και 1 unit από την KAPA Taq DNA πολυμεράση (Kapa Biosystems Inc.), στο θερμικό κυκλοποιητή Veriti 96-Well Fast Thermal Cycler (Applied Biosystems[™]), υπό τις ακόλουθες συνθήκες: αποδιάταξη στους 95°C για 3 λεπτά, ακολουθούμενο από 30 κύκλους στους 95°C για 30 δευτερόλεπτα, 72°C για 5 λεπτά, και στάδιο τελικής επιμήκυνσης στους 72°C για 5 λεπτά.

Έπειτα, αφού τα προϊόντα που προέκυψαν αραιώθηκαν κατά 1:30 σε νερό ελεύθερο νουκλεασών, χρησιμοποιήθηκαν ως εκμαγεία για τις εσωτερικές (nested) RT-PCR που ακολούθησαν, χρησιμοποιώντας τους ειδικούς εκκινητές (Πίνακας 7) που σχεδιάστηκαν προκειμένου να στοχεύουν κάθε νέο μετάγραφο *SCAF1*, και να εξακριβωθεί η τελική σύνθεση των νέων μεταγράφων.

Συγκεκριμένα, ο πρόσθιος εκκινητής 2/3F που στόχευε στο σημείο συρραφής των εξονίων 2 και 3, συνδυάστηκε ξεχωριστά με 5 ανάστροφους εκκινητές οι οποίοι στόχευαν το σημείο συρραφής των εξονίων: 10 και 11 (10/11R), 6 και 8 (6/8R), 6 και 10 (6/10R), 4 και 10 (4/10R) και τέλος 4 και 8 (4/8R). Όμοια, ο πρόσθιος εκκινητής 2/4F που στόχευε στο σημείο συρραφής των εξονίων 2 και 4, συνδυάστηκε σε ξεχωριστές αντιδράσεις με τους προαναφερόμενους ανάστροφους εκκινητές δηλαδή τους 10/11R, 6/8R, 6/10R, 4/10R και 4/8R. Οι παραπάνω αντιδράσεις PCR πραγματοποιήθηκαν σύμφωνα με το πρωτόκολλο (Πίνακας 6) για 35 κύκλους.

Ο προσδιορισμός τριών ακόμα εναλλακτικών μεταγράφων έγινε με βάση το πρωτόκολλο (Πίνακας 5), και κατά την αντίδραση PCR χρησιμοποιήθηκαν τα εξής ζεύγη εκκινητών: ο πρόσθιος εκκινητής 2/8F που στόχευε στο σημείο συρραφής των εξονίων 2 και 8 σε συνδυασμό με τον ανάστροφο εκκινητή 10/11R που στόχευε το σημείο συρραφής των εξονίων 10 και 11, ο πρόσθιος εκκινητής 2/11F που στόχευε στο σημείο συρραφής των εξονίων 2 και 11 σε συνδυασμό με τον ανάστροφο εκκινητή που χρησιμοποιήθηκε στην εσωτερική PCR κατά τη μοριακή κλωνοποίηση, 11R, όπου στόχευε μια συγκεκριμένη περιοχή στο εξόνιο 11 και τέλος, χρησιμοποιήθηκε ο πρόσθιος εκκινητής 3/8F που στόχευε στο σημείο συρραφής των εξονίων 3 και 8 σε συνδυασμό με τον ανάστροφο εκκινητή 10/11R που στόχευε το σημείο συρραφής των εξονίων 10 και 11. Στα προϊόντα που προέκυψαν με τη μέθοδο της αλυσιδωτής αντίδρασης πολυμεράσης και της χρήσης των ζευγών εκκινητών: 2/3F – 10/11R, 2/3F – 6/8R, 2/3F – 6/10R, 2/4F – 10/11R, 2/4F – 6/8R και 2/4F – 6/10R, έγινε αραίωση 1:50. Στα αραιωμένα προϊόντα πραγματοποιήθηκαν ξεχωριστές εσωτερικές αντιδράσεις PCR (nested PCR), με βάση το πρωτόκολλο (Πίνακας 6), και χρησιμοποιώντας τα κάτωθι ζευγάρια εκκινητών ως εξής:

- στα προϊόντα που προέκυψαν από τη χρήση των εκκινητών 2/3F 10/11R και αντίστοιχα των εκκινητών 2/4F 10/11R χρησιμοποιήθηκαν αντίστοιχα: ένας πρόσθιος εκκινητής 4/6F που στόχευε στο σημείο συρραφής των εξονίων 4 και 6 σε συνδυασμό με τον ανάστροφο εκκινητή 10R που στόχευε μια συγκεκριμένη περιοχή στο εξόνιο 10, ένας πρόσθιος εκκινητής 4/8F που στόχευε στο σημείο συρραφής των εξονίων 4 και 8 σε συνδυασμό με τον προαναφερόμενο ανάστροφο εκκινητή 10R και χρησιμοποιήθηκε ακόμα ένας πρόσθιος εκκινητής 4/10F που στόχευε στο σημείο συρραφής των εξονίων 4 και τον προαναφερόμενο ανάστροφο εκκινητή 10R και χρησιμοποιήθηκε ακόμα ένας πρόσθιος εκκινητής 4/10F που στόχευε στο σημείο συρραφής των εξονίων 4 και 10 σε συνδυασμό με τον προαναφερόμενο ανάστροφο εκκινητή 10R.
- στα προϊόντα που προέκυψαν από τη χρήση των εκκινητών 2/3F 6/8R
 και αντίστοιχα των εκκινητών 2/4F 6/8R χρησιμοποιήθηκαν
 αντίστοιχα: ο πρόσθιος εκκινητής 4/6F σε συνδυασμό με τον ανάστροφο
 εκκινητή 6R που στόχευε μια συγκεκριμένη περιοχή στο εξόνιο 6.
- στα προϊόντα που προέκυψαν από τη χρήση των εκκινητών 2/3F 6/10R
 και αντίστοιχα των εκκινητών 2/4F 6/10R χρησιμοποιήθηκαν
 αντίστοιχα: ο πρόσθιος εκκινητής 4/6F σε συνδυασμό με τον ανάστροφο
 εκκινητή 6R.

Πίνακας 5: Πρωτόκολλο και συνθήκες των ξεχωριστών αντιδράσεων PCR με εκκινητές τα ζεύγη 2/8F – 10/11R, 2/11F – 11R και 3/8F – 10/11R. Ως εκμαγείο DNA για τις αντιδράσεις χρησιμοποιήθηκε το αρχικό cDNA, με το οποίο είχε γίνει και η κατασκευή της βιβλιοθήκης.

Πρωτόκολλο		Συνθήκες αντίδρασης				
Αντιδραστήρια	Αρχική συγκέντρωση	Όγκος (μl)	Στάδιο	Θερμ/σία (ºC)	Χρόνος	Κύκλοι
PCR-GRADE H ₂ 0	-	Μέχρι τα 25μl	Αρχική αποδιάταξη	95	3 min	1
10X KAPA Taq Buffer	10X	2,5				
dNTPs	10 mM	0,5	Αποδιάταξη	95	30 sec —	
MgCl ₂	-	Περιείχε το Buffer	Υβριδ/ση	60	30 sec	x40
Πρόσθιος εκκινητής	10µM	1	Επέκταση	72	30 sec —	J
Ανάστροφος εκκινητής	10µM	1				
ΚΑΡΑ Τaq πολυμεράση	5U/μL	0,2	Τελική επέκταση	72	30 sec	1
Σύνολο		25µL				

Πίνακας 6: Πρωτόκολλο και συνθήκες εσωτερικής αντίδρασης Nested PCR στα υπό αραίωση 1:50 προϊόντα των ξεχωριστών αντιδράσεων PCR με εκκινητές 2/3F - 10/11R και 2/4F -10/11R αντίστοιχα.

Πρωτόκολλο		Συνθήκες αντίδρασης				
Αντιδραστήρια	Αρχική συγκέντρωση	Όγκος (μl)	Στάδιο	Θερμ/σία (ºC)	Χρόνος	Κύκλοι
PCR-GRADE H ₂ 0	-	Μέχρι τα 25μl	Αρχική αποδιάταξη	95	3 min	1
10X KAPA Taq Buffer	10X	2,5				
dNTPs	10 mM	0,5	Αποδιάταξη	95	30 sec —	
MgCl ₂	-	Περιείχε το Buffer	Υβριδ/ση	60	30 sec	x30
Πρόσθιος εκκινητής	10µM	1	Επέκταση	72	1 min —	J
Ανάστροφος εκκινητής	10µM	1				
ΚΑΡΑ Ταq πολυμεράση	5U/μL	0,2	Τελική επέκταση	72	1 min	1
Σύνολο		25µL				

Πίνακας 7: Καταγραφή των εκκινητών που χρησιμοποιήθηκαν για τον προσδιορισμό των νέων μεταγράφων του γονιδίου *SCAF1*. Κάθε αναφερόμενος εκκινητής που σχεδιάστηκε με σκοπό να στοχεύσει μια συγκεκριμένη θέση συρραφής μεταξύ δύο γνωστών εξονίων διαχωρίζεται με «/». Οι εκκινητές 6R, 10R και 11R, στοχεύουν μία συγκεκριμένη περιοχή στα εξόνια 6, 10 και 11 αντιστοίχως. Οι θερμοκρασίες τήξεως των εκκινητών (Tm) υπολογίστηκαν μέσω Primer-BLAST.

Εκκινητές					
Πρόσθιος εκκινητής	Αλληλουχία (5'→3')	Μήκος (nt)	T _m (°C)		
2/3F	GCCTTTATCCTGCGAGCCAT	20	60.54		
2/4F	CTGCCTTTATCCTGATGGCTCT	22	59.89		
2/8F	TGCCTTTATCCTGTGACTGCA	21	59.65		
2/11F	TGCCTTTATCCTGATCTGCCAC	22	60.42		
3/8F	TGCCCAATGATAAAGTGACTGCA	23	60.81		
4/6F	CTGACACGGCTACTAAGTGAGG	22	60.16		
4/8F	TGACACGGCTACTTGACTGC	20	60.32		
4/10F	CACTGACACGGCTACTTATCTGA	23	59.87		
Ανάστροφος	Αλληλουνία (5'→3')	Μήκος	Tm		
εκκινητής		(nt)	(°C)		
11R	CAAAGAGGTGAGGGGGGAAG	20	60.67		
10/11R	GCTGTGGCAGATCTTGTGGA	20	60.32		
10R	CCTCAGGATGTCCTTGTACTCCT	23	60.88		
6/8R	GAAGTGCAGTCACCGTTTTGC	21	60.86		
6R	GTTTTGCCCGTCCTCCAGG	19	60.97		
6/10R	AGCTTCTTCAGATACCGTTTTGC	23	59.56		
4/10R	AGCTTCTTCAGATAAGTAGCCGT	23	59.30		
4/8R	AGAGAAGTGCAGTCAAGTAGCC	22	60.03		

2.9. ΑΝΙΧΝΕΥΣΗ ΤΩΝ ΠΡΟΪΟΝΤΩΝ ΤΗΣ ΑΛΥΣΙΔΩΤΗΣ ΑΝΤΙΔΡΑΣΗΣ ΠΟΛΥΜΕΡΑΣΗΣ (PCR), ΜΕ ΗΛΕΚΤΡΟΦΟΡΗΣΗ ΣΕ ΠΗΚΤΩΜΑ ΑΓΑΡΟΖΗΣ

Τα απαιτούμενα υλικά και διαλύματα που χρησιμοποιήθηκαν για τη διεξαγωγή της συγκεκριμένης μεθόδου ήταν τα ακόλουθα:

- Αγαρόζη (Invitrogen)
- Ρυθμιστικό διάλυμα TBE (1X) που περιέχει:
 - ✓ 0,089M Tris (Applichem),
 - ✓ 0,089Μ βορικό οξύ (Merck),
 - ✓ 0,002M EDTA (Applichem).
- Βρωμιούχο αιθίδιο ethidium bromide (EtBr)
- DNA loading dye solution 2X (διάλυμα χρώσης του DNA)
- Συσκευή ηλεκτροφόρησης

Με την ηλεκτροφόρηση σε πήκτωμα αγαρόζης έχουμε τη δυνατότητα διαχωρισμού τμημάτων DNA που το μέγεθός τους κυμαίνεται περίπου από 75 ζεύγη βάσεων έως 30.000 ζεύγη βάσεων [211, 212]. Το DNA είναι ένα μόριο φορτισμένο αρνητικά, ως εκ τούτου όταν βρεθεί σε ηλεκτρικό πεδίο κινείται προς τη θετικά φορτισμένη άνοδο (Εικόνα 35) [213]. Η απόσταση που διανύουν τα μόρια DNA μέσα στο πήκτωμα αγαρόζης είναι αντιστρόφως ανάλογη του λογαρίθμου του μοριακού τους βάρους [214]. Ο ρυθμός της ηλεκτροφορητικής κινητικότητας των μορίων του DNA μέσω του πηκτώματος αγαρόζης καθορίζεται από: α) το μέγεθος του DNA, β) τη συγκέντρωση της αγαρόζης, γ) τη στερεοδιάταξη του DNA, δ) την ένταση του ρεύματος, ε) την παρουσία βρωμιούχου αιθιδίου, στ) το τύπο της αγαρόζης και ζ) το ρυθμιστικό διάλυμα της ηλεκτροφόρησης [213, 215].

Για να πραγματοποιηθεί η ηλεκτροφόρηση των προϊόντων DNA σε πήκτωμα αγαρόζης, ακολουθείται η παρακάτω πειραματική διαδικασία:

Σε κωνική φιάλη προσθέτουμε τα κατάλληλα g σκόνης αγαρόζης και τα ανάλογα ml ρυθμιστικού διαλύματος TBE 1X, ώστε να παραχθεί πήκτωμα αγαρόζης 1,5% - 3% w/v. (Για παράδειγμα, για την παραγωγή 2% w/v πηκτώματος αγαρόζης χρησιμοποιήθηκαν 1,4gr σκόνης αγαρόζης και 70ml ρυθμιστικού διαλύματος TBE 1X). Το επόμενο βήμα είναι να θερμάνουμε με συνεχή ανάδευση έως ότου διαλυθεί η αγαρόζη. Ακολουθεί η προσεκτική προσθήκη βρωμιούχου αιθιδίου (ισχυρό μεταλλαξογόνο) με τελική συγκέντρωση 0.5 μg/ml. Γίνεται η συναρμολόγηση της συσκευής, στην οποία θα πήξει η αγαρόζη και τοποθετείται "χτένα", η οποία δημιουργεί θέσεις φόρτωσης του δείγματος όταν ολοκληρωθεί η πήξη. Έπειτα, χύνουμε το διάλυμα της αγαρόζης στην συσκευή που φέρει τη "χτένα" και όταν πήξει η αγαρόζη, αφαιρούμε με προσοχή τη "χτένα" δημιουργώντας τις θέσεις φόρτωσης του δείγματος. Μετά, εισάγουμε το πήκτωμα αγαρόζης στη συσκευή ηλεκτροφόρησης που περιέχει ρυθμιστικό διάλυμα TBE 0,5Χ.

Τέλος, προσθέτουμε κατάλληλο όγκο διαλύματος χρώσης (DNA loading dye solution 6X) στα δείγματα που θα φορτώσουμε στο πήκτωμα. Εκτός από τη φόρτωση των δειγμάτων στο πήκτωμα, απαραίτητη είναι και η φόρτωση του κατάλληλου ladder. Η ηλεκτροφόρηση ξεκινά έχοντας ρυθμίσει στην συσκευή ηλεκτροφόρησης, τα κατάλληλα Volt (συνήθως 80 - 120V) και το απαιτούμενο χρονικό διάστημα (συνήθως 35 - 120 λεπτά) που θα χρειαστούν για το διαχωρισμό. Κάθε πήκτωμα φωτογραφήθηκε μετά το τέλος της ηλεκτροφόρησης με χρήση υπεριώδους ακτινοβολίας, εκμεταλλευόμενοι το φαινόμενο του φθορισμού, εξαιτίας της προσθήκης βρωμιούχου αιθιδίου κατά την κατασκευή του πηκτώματος αγαρόζης.



Εικόνα 35: Σχηματική απεικόνιση λειτουργίας ηλεκτροφόρησης σε πήκτωμα αγαρόζης (http://insights.globalspec.com/images/assets/161/1161/Image_1-Fullsize.jpg).

2.10. ΤΡΙΣΔΙΑΣΤΑΤΗ ΑΠΕΙΚΟΝΙΣΗ ΝΕΩΝ ΠΡΩΤΕΪΝΙΚΩΝ ΙΣΟΜΟΡΦΩΝ

Μετά την ανακάλυψη της πλήρους πρωτοταγούς δομής των νέων εναλλακτικών μεταγράφων για το γονίδιο *SCAF1*, πραγματοποιήθηκε πρόβλεψη της δομής των νέων πρωτεϊνικών ισομορφών στο χώρο. Για την ανωτέρω πρόβλεψη χρησιμοποιήθηκε ο I-TASSER (Iterative Threading ASSEmbly Refinement) [216], που αποτελεί την καλύτερη και πιο ολοκληρωμένη λύση στην πρόγνωση τριτοταγούς δομής [217], όπως πιστοποιείται από τις πρώτες θέσεις που καταλαμβάνει στις παγκόσμιες αξιολογήσεις του CASP (Critical Assessment of protein Structure Prediction) [218].

Ο Ι-TASSER πλεονεκτεί καθώς χαρακτηρίζεται από μεγάλη ακρίβεια στην πρόβλεψη, αλλά και από μεγάλη ταχύτητα στους υπολογισμούς. Για την παραγωγή μοντέλων μεγαλύτερης πιστότητας χρησιμοποιείται ο αλγόριθμος LOMETS (local meta-threading-server) [219]. Το εν λόγω εργαλείο εκμεταλλεύεται την πρωτοταγή αμινοξική αλληλουχία μια πρωτεΐνης και αποδίδει την τρισδιάστατη μορφή της καθώς και πληροφορίες σχετικές με τη λειτουργία της (Εικόνα 36). Ο Ι-TASSER, έχει τη δυνατότητα να εντοπίζει τμήματα τα οποία έχουν μεγάλη ομοιότητα με αντίστοιχα τμήματα πρωτεϊνών με γνωστή δομή από τη βάση δεδομένων PDB με χρήση του LOMETS, προβλέποντας τον τρόπο αναδίπλωσης αυτών των περιοχών [220, 221].

Έπειτα, ο I-TASSER συναρμολογεί αυτές τις περιοχές κατασκευάζοντας ένα μοντέλο με μια τεχνική που ονομάζεται replica exchange Monte Carlo simulations και όπου τα στοιχεία δεν είναι επαρκή, θα κατασκευάσει μοντέλο με μέθοδο ab initio για ολόκληρη την πρωτεΐνη [216]. Το επόμενο βήμα αφορά την αξιολόγηση μιας δομής ως επιθυμητή και σε αυτό το σημείο ξεχωρίζονται και προκρίνονται ως πιθανές δομές εκείνες που υπακούν στο κανόνα ελάχιστης ελεύθερης ενέργειας. Για τον προσδιορισμό των καταστάσεων χαμηλής ελεύθερης ενέργειας χρησιμοποιείται ο αλγόριθμος SPICKER [222]. Τα τελικά πλήρη πρωτεϊνικά μοντέλα λαμβάνονται από τον αλγόριθμο REMO, ο οποίος κατασκευάζει τις λεπτομέρειες σε επίπεδο ατόμου, μέσω της βελτιστοποίησης των δικτύων δεσμών υδρογόνου [223].



Εικόνα 36: Σχηματική απεικόνιση των βασικών σταδίων που μεσολαβούν για τη πρόγνωση τριτοταγούς δομής μέσω Ι – TASSER (<u>https://en.wikipedia.org/wiki/I-TASSER#/media/File:I-TASSER-pipeline.jpg</u>).

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1. ΝΕΕΣ ΘΕΣΕΙΣ ΣΥΡΡΑΦΗΣ ΕΞΟΝΙΩΝ

Έως τώρα έχει περιγραφεί ένα μετάγραφο του γονιδίου SCAF1 στον άνθρωπο (NM_021228.2), το οποίο κωδικοποιεί πρωτεΐνη (NP_067051.2) και αποτελείται από 11 εξόνια και 10 εσόνια. Εκτός από το ανωτέρω μετάγραφο του γονιδίου SCAF1, έχουν αναφερθεί τρία επιπλέον προβλεπόμενα μετάγραφα του γονιδίου τα XM_011527194.2, XM_005259122.4 και XM_017027083.1, τα οποία έχουν προκύψει από αυτοματοποιημένη υπολογιστική ανάλυση και προβλέπεται να κωδικοποιούν αντίστοιχα τις πρωτεΐνες XP_011525496.1, XP_005259179.1 και XP_016882572.1. Τα δεδομένα από την αλληλούχηση περιείχαν αναγνώσματα που επιβεβαίωναν όλα τα γνωστά εξόνια και τις γνωστές θέσεις συρραφής.

Εκτός από τις γνωστές θέσεις συρραφής, με τη βιοπληροφορική ανάλυση των δεδομένων που πραγματοποιήθηκε, βρέθηκαν εννέα νέες θέσεις συρραφής μεταξύ γνωστών εξονίων του γονιδίου SCAF1 (Πίνακας 8), υποστηρίζοντας την ύπαρξη εναλλακτικών μεταγράφων με συχνότητα μικρότερη του γνωστού μεταγράφου.

Πίνακας 8: Στον πίνακα καταγράφονται όλες οι γνωστές και νέες θέσεις συρραφής του γονιδίου *SCAF1* που ανιχνεύθηκαν στα δεδομένα αλληλούχησης νέας γενιάς. Επίσης, καταγράφεται ο αριθμός των αναγνώσεων που καλύπτουν τις θέσεις συρραφής.

Γνωστές θέσεις συρραφής		Νέες θέσεις συρραφής		
Θέση συρραφής	Αριθμός Αναγνώσεων	Θέση συρραφής	Αριθμός Αναγνώσεων	
2-3	54011	2-4	336	
3-4	84411	2-8	181	
4-5	108942	2-11	9	
5-6	80199	3-8	455	
6-7	22650	4-6	70	
7-8	8586	4-8	563	
8-9	26386	4-10	218	
9-10	28992	6-8	10738	
10-11	39175	6-10	9859	

Από τις εννέα νέες θέσεις συρραφής, η συρραφή μεταξύ των γνωστών εξονίων 2 και 4 ήταν το πιο συχνό εύρημα. Ωστόσο, το εξόνιο 2 φάνηκε επίσης να συρράπτεται εναλλακτικά με τα γνωστά εξόνια 8 και 11 αντίστοιχα. Επιπλέον, παρατηρήθηκε εναλλακτική συρραφή μεταξύ του εξονίου 3 και 8, του εξονίου 4 με τα εξόνια 6, 8 και 10 αντίστοιχα και τέλος του εξονίου 6 με τα εξόνια 8 και 10 αντίστοιχα.

Exon 2 - Exon 4 @N40RV:00907:01228 GGGCGATGGTCCGCCAGACAGAGAGACCCCACGCGTTCTCCTTCTGCCTTTATCCTGATGGCTCTCGGTGTCATGGCCTTCGATGGCGGCGCCTGCCGGAGTCCACGGTCAGAGC Exon 2 - Exon 8 @N40RV:00098:00097 gggcgatggtccgccagacagagaccccacgctttctcccttctgcctttatcctgtgactgcacttcttcttcaagatggaagaccaacccgagccgaaggccaaggccc Exon 2 - Exon 11 @N40RV:00568:00129 GGGCGATGGTCGGCCAGACAGAGACCCCACGCTTTCTCCTCTCTCCCCTTTATCCTGCCACAGCAAAAGTGGGGAAATCAACCCAGTGAAGGTGAGCAACCTGGTGCGG Exon 3 - Exon 8 @N40RV:00010:00385 gccatccagcaggctgtgggaagctcccccaggggggcccgaggggcccatgcactagatgaaggactgcacttcctcttcaagatggaagaagccaagccgaggcgaaggccaaggcaaggcca Exon 4 - Exon 8 @N40RV:00104:00656 GAGTCCACGGTCAGAGCCCCCGTTCCCAGGAATCAGGGGGCACTGACACGGCTACTTGACTGCACTTCTTCAAGATGGAAGAGCCAACCTGGCGAGCCGAGCGAAGGCCC Exon 4 - Exon 10@N40RV:00203:00102

Exon 4 - Exon 6 @N40RV:00062:00038 GGAGTCCACGGTCAGAGGCCCCGTTCCCAGGAATCAGGGGGGCACTGACACGGCTACTGTGTGGACATGGCCACGGACAGCTTCCTCGCAGGGCTGGTGAGTGTCCTGGATCC

Εικόνα 37: Σχηματική απεικόνιση αντιπροσωπευτικών αλληλουχιών από αναγνώσματα (reads) για κάθε νέα θέση συρραφής που ανιχνεύθηκε στο γονίδιο SCAF1. Για κάθε ανάγνωσμα αλληλούχησης, τα νουκλεοτίδια κάθε γνωστού εξονίου του *SCAF1* αναπαρίστανται με διαφορετικό χρώμα. Συνολικά, ανιχνεύθηκαν 9 νέες θέσεις συρραφής.

3.2. ΤΑΥΤΟΠΟΙΗΣΗ 15 ΝΕΩΝ SCAF1 ΜΕΤΑΓΡΑΦΩΝ

Όπως μέχρι τώρα έχουμε δει, από την αλληλούχηση νέας γενιάς προέκυψαν αναγνώσματα τα οποία, παρά τη συνεισφορά τους στην ανακάλυψη νέων ευρημάτων, δε μπορούσαν να επιβεβαιώσουν νέα μετάγραφα του γονιδίου SCAF1, καθώς οι αλληλουχίες mRNA των μεταγράφων του SCAF1 είναι μεγαλύτερες από 400 ζεύγη βάσεων, δηλαδή μεγαλύτερες από το μέγιστο μήκος θραυσμάτων DNA που μπορούν να αλληλουχηθούν με τη χρήση της αλληλούχησης νέας γενιάς. Κατ' επέκταση στο αρχείο FASTQ υπήρχαν αναγνώσματα που κάλυπταν συγκεκριμένης έκτασης (περίπου 400 bp) αλληλουχία mRNA του γονιδίου SCAF1. Για να ξεπεραστεί αυτός ο περιορισμός, χρησιμοποιήθηκαν ειδικοί εκκινητές για PCR, σχεδιασμένοι ώστε να στοχεύουν τις νέες θέσεις συρραφής, προκειμένου να γίνει η ταυτοποίηση των νέων μεταγράφων του SCAF1 (Πίνακας 7).

Έτσι, πραγματοποιήθηκαν αντιδράσεις PCR με τη χρήση ζευγών εκκινητών που επέτρεψαν την ειδική ενίσχυση ενός συγκεκριμένου νέου εναλλακτικού μεταγράφου του SCAF1. Μετά το τέλος της αντίδρασης PCR και της ηλεκτροφόρησης των PCR προϊόντων σε πήκτωμα αγαρόζης, πραγματοποιήθηκε η ανακάλυψη 7 νέων εναλλακτικών μεταγράφων του SCAF1 (SCAF1 v.4, v.6, v.10, v.12, v.14, v.15 και v.16). Ωστόσο, με βάση τα αποτελέσματα που αποκτήθηκαν από την αλληλούχηση νέας γενιάς, προέκυψαν και προβλεπόμενα μετάγραφα τα οποία δεν μπόρεσαν να επιβεβαιωθούν με μια αντίδραση PCR. Για τα μετάγραφα αυτά SCAF1 (SCAF1 v.2, v.3, v.5, v.7, v.8, v.9, v.11 και v.13), απαιτήθηκε να πραγματοποιηθούν εσωτερικές αντιδράσεις PCR (nested PCR) με τη χρήση ειδικών εσωτερικών εκκινητών και το αποτέλεσμα ήταν να παραχθούν αμπλικόνια με μήκος που κυμαίνεται μεταξύ 128 έως 458 ζεύγη βάσεων, όπως φαίνεται στον παρακάτω πίνακα (Πίνακας 9). **Πίνακας 9**: Απεικονίζονται τα ζεύγη εκκινητών που χρησιμοποιήθηκαν για την ανάλυση έκφρασης κάθε νέου μεταγράφου *SCAF1* (v.2 - v.16) με χρήση των μεθόδων RT-PCR και nested RT-PCR. Τα πεδία του πίνακα με γκρι χρώμα δηλώνουν πως για την ταυτοποίηση και την ανάλυση έκφρασης για το συγκεκριμένο μετάγραφο, δεν πραγματοποιήθηκε nested RT-PCR. Για κάθε νέο μετάγραφο που ταυτοποιήθηκε με τις μεθόδους που αναφέρθηκαν, φαίνεται το μέγεθος των προϊόντων που προέκυψαν με nested RT-PCR.

Νέα μετάγραφα SCAF1	RT-PCR		Nester	Μέγεθος	
	Πρόσθιος εκκινητής	Ανάστροφος εκκινητής	Πρόσθιος εκκινητής	Ανάστροφος εκκινητής	αμπλικονίων
v.2	2/3F	10/11R	4/8F	10R	429 bp
v.3	2/3F	10/11R	4/10F	10R	130 bp
v.4	2/3F	6/8R			394 bp
v.5	2/3F	6/8R	4/6F	6R	128 bp
v.6	2/3F	6/10R			396 bp
v.7	2/3F	6/10R	4/6F	6R	128 bp
v.8	2/4F	10/11R	4/8F	10R	429 bp
v.9	2/4F	10/11R	4/10F	10R	130 bp
v.10	2/4F	6/8R			338 bp
v.11	2/4F	6/8R	4/6F	6R	128 bp
v.12	2/4F	6/10R			340 bp
v.13	2/4F	6/10R	4/6F	6R	128 bp
v.14	2/8F	10/11R			456 bp
v.15	3/8F	10/11R			458 bp
v.16	2/11F	11R			240 bp

Σύμφωνα λοιπόν με τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν, τα οποία έχουν αναφερθεί ανωτέρω, προέκυψαν 15 νέα εναλλακτικά μετάγραφα του γονιδίου *SCAF1* (*SCAF1* v.2 – v.16) (Εικόνα 38), οι αλληλουχίες των οποίων είναι κατατεθειμένες στην βάση δεδομένων GenBank[®] (GenBank[®] accession numbers: KY849380 – KY849394 αντίστοιχα).



Εικόνα 38: Σχηματική απεικόνιση των 15 νέων εναλλακτικών μεταγράφων του γονιδίου *SCAF1* (*SCAF1* v.2 – v.16). Τα εξόνια παρουσιάζονται ως τετράγωνα, ενώ τα εσόνια ως γραμμές. Τα γκρι πεδία των τετραγώνων αντιπροσωπεύουν τις κωδικές περιοχές, ενώ τα λευκά τις μη κωδικές. Ο αριθμός κάθε μεταγράφου εμφανίζεται δεξιά από κάθε μετάγραφο.

Τα περισσότερα από τα νέα εναλλακτικά μετάγραφα που ανακαλύφθηκαν, μπορούν να ομαδοποιηθούν σε ζεύγη βάσει των κοινών δομικών χαρακτηριστικών που εμφανίζουν, καθώς και των ορίων των εξονίων και εσονίων τους, αλλά και βάσει της παρουσίας η απουσίας εναλλακτικού ματίσματος μεταξύ των εξονίων 2 και 4.

Συγκεκριμένα, στα νέα εναλλακτικά μετάγραφα *SCAF1* v.2 και v.8 παρατηρείται η εναλλακτική συρραφή μεταξύ των εξονίων 4 και 8. Ωστόσο, το μετάγραφο *SCAF1* v.8 περιλαμβάνει τη νέα θέση συρραφής μεταξύ των εξονίων 2 και 4, η οποία αποτελεί και τη μόνη διαφορά μεταξύ των δύο αυτών αλληλουχιών mRNA (Εικόνα 38). Όσον αφορά την ικανότητα τους να κωδικοποιούν για πρωτεΐνες, και τα δύο αυτά μετάγραφα περιέχουν ένα πρώιμο κωδικόνιο λήξης (premature termination codon, PTC) και ως εκ τούτου είναι υποψήφια να υπόκεινται σε αποικοδόμηση (non-sense-mediated mRNA decay, NMD).

Κατά τον ίδιο τρόπο, στα νέα μετάγραφα SCAF1 v.3 και v.9 παρατηρείται εναλλακτική συρραφή μεταξύ των εξονίων 4 και 10. Το SCAF1 v.9 χαρακτηρίζεται επίσης από την ύπαρξη εναλλακτικής συρραφής μεταξύ των εξονίων 2 και 4 (Εικόνα 38). Πέρα από αυτές τις δομικές διαφορές, το SCAF1 v.3 περιέχει ανοιχτό πλαίσιο ανάγνωσης (open reading frame, ORF) με αποτέλεσμα να προβλέπεται να κωδικοποιεί μια νέα πρωτεϊνική ισομορφή του γονιδίου SCAF1, αποτελούμενη από 293 αμινοξικά κατάλοιπα, ενώ το SCAF1 v.9 στερείται ανοιχτό πλαίσιο ανάγνωσης, καθώς περιέχει πρώιμο κωδικόνιο λήξης και ως εκ τούτου είναι υποψήφιο για NMD.

Δύο ακόμα νέα μετάγραφα, τα SCAF1 v.4 και v.10, φέρουν εναλλακτική συρραφή μεταξύ των εξονίων 6 και 8. Το SCAF1 v.10 περιέχει επίσης τη νέα θέση συρραφής μεταξύ των εξονίων 2 και 4 και ως εκ τούτου στερείται πλήρως το εξόνιο 3. Αυτή η διαφορά φαίνεται να έχει σημαντική επίδραση στη κωδικοποιητική ικανότητα των δύο αυτών νέων μεταγράφων, καθώς το SCAF1 v.4 περιέχει ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιεί για μια νέα πρωτεϊνική ισομορφή, αποτελούμενη από 366 αμινοξικά κατάλοιπα, ενώ το SCAF1 v.10 είναι υποψήφιο για NMD καθώς περιέχει πρώιμο κωδικόνιο λήξης.

Τα δύο επόμενα νέα μετάγραφα, τα *SCAF1* v.6 και v.12, περιέχουν μια νέα θέση συρραφής μεταξύ των εξονίων 6 και 10. Το *SCAF1* v.12 περιέχει επίσης μια ακόμα νέα θέση συρραφής μεταξύ των εξονίων 2 και 4 (Εικόνα 38). Το *SCAF1* v.6 έχει ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιεί για μια νέα πρωτεϊνική ισομορφή, αποτελούμενη από 344 αμινοξικά κατάλοιπα, ενώ το *SCAF1* v.12 περιέχει πρώιμο κωδικόνιο λήξης, έτσι θεωρείται ως μη κωδικό μετάγραφο mRNA (non – coding mRNA).

Τα νέα μετάγραφα SCAF1 v.5 και v.7 που ανακαλύφθηκαν παρουσιάζουν μία μόνο δομική διαφορά με τα προαναφερθέντα μετάγραφα SCAF1 v.4 και v.6, η οποία είναι η έλλειψη ολόκληρης της αλληλουχίας του εξονίου 5 από τις mRNA αλληλουχίες τους. Αυτό έχει ως αποτέλεσμα, τα μετάγραφα SCAF1 v.5 και v.7 να περιέχουν τη νέα θέση συρραφής μεταξύ των εξονίων 4 και 6, η οποία απουσιάζει από τα μετάγραφα SCAF1 v.4 και v.6. Συμπληρωματικά, τόσο το μετάγραφο SCAF1 v.5 όσο και το v.7 περιέχει πρώιμο κωδικόνιο λήξης (PTC), ως εκ τούτου είναι υποψήφια για NMD. Αντίστοιχα, τα δύο νέα μετάγραφα SCAF1 v.11 και v.13 διαφέρουν από τα νέα μετάγραφα SCAF1 v.10 και v.12 αντίσοιχα, μόνο ως προς την έλλειψη του εξονίου 5 στις αλληλουχίες του mRNA τους (Εικόνα 38). Τόσο το μετάγραφο SCAF1 v.11 όσο και το v.13 έχουν ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιούν για δύο νέες πρωτεϊνικές ισομορφές που αποτελούνται από 313 και 291 αμινοξικά κατάλοιπα, αντίστοιχα.

Τέλος, τρία ακόμα νέα εναλλακτικά μετάγραφα του γονιδίου SCAF1 ανακαλύφθηκαν, τα SCAF1 v.14, v.15 και v.16 (Εικόνα 38). Το νέο μετάγραφο SCAF1 v.14 περιλαμβάνει τη νέα θέση συρραφής μεταξύ των εξονίων 2 και 8, το SCAF1 v.15 περιλαμβάνει τη νέα θέση συρραφής μεταξύ των εξονίων 3 και 8, ενώ το SCAF1 v.16, το οποίο αποτελεί και το μικρότερο μετάγραφο που προσδιορίστηκε, περιλαμβάνει τη νέα θέση συρραφής μεταξύ των εξονίων 2 και 11. Παρά το περιορισμένο μήκος τους σε νουκλεοτίδια σε σχέση με το κυρίως μετάγραφο του γονιδίου SCAF1, τα μετάγραφα SCAF1 v.15 και v.16 εμφανίζουν ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιούν για δύο νέες πρωτεϊνικές ισομορφές που αποτελούνται από 262 και 99 αμινοξικά κατάλοιπα, αντίστοιχα. Ωστόσο, το μετάγραφο SCAF1 v.14 περιέχει πρώιμο κωδικόνιο λήξης (PTC) και ως εκ τούτου είναι υποψήφιο για NMD.

3.3. ΠΡΟΤΥΠΟ ΕΚΦΡΑΣΗΣ ΤΩΝ ΝΕΩΝ SCAF1 ΜΕΤΑΓΡΑΦΩΝ

Μετά από τον προσδιορισμό των δεκαπέντε νέων μεταγράφων (SCAF1 v.2 – v.16) εξετάστηκε το πρότυπο έκφρασής τους σε ένα σύνολο από cDNAs που

αντιστοιχούσαν σε 17 συγκεκριμένους ανθρώπινους καρκινικούς ιστούς, καθώς και σε 2 υγιείς. Στη συνέχεια, ακολουθούν τα αποτελέσματα ηλεκτροφόρησης σε πήκτωμα αγαρόζης των αμπλικονίων που δημιουργήθηκαν με τη χρήση ειδικών εκκινητών (Πίνακας 9), που επαληθεύουν την ύπαρξη και την έκφραση των νέων μεταγράφων (v.2 - v.16) του γονιδίου *SCAF1*, στο σύνολο των cDNAs που χρησιμοποιήθηκαν (Εικόνα 39).







Εικόνα 39: Απεικόνιση των αποτελεσμάτων της ηλεκτροφόρησης που επαληθεύουν την ύπαρξη των νέων εναλλακτικών μεταγράφων του γονιδίου *SCAF1*, v.2 έως v.16, καθώς και της έκφρασης των μεταγράφων αυτών στο σύνολο των cDNAs που χρησιμοποιήθηκαν και αντιστοιχούν σε καρκινικούς και φυσιολογικούς ανθρώπινους ιστούς. Εμφανίζονται κατά σειρά: 100bp DNA ladder, αδενοκαρκίνωμα του μαστού, πορογενές καρκίνωμα μαστού, καρκίνος των ωοθηκών, αδενοκαρκίνωμα του ενδομητρίου, καρκίνος του τραχήλου της μήτρας, καρκίνος του προστάτη, καρκίνος της ουροδόχου κύστης, καρκίνος του νεφρού, καρκίνος του παχέος εντέρου, γαστρικό αδενοκαρκίνωμα, ηπατοκυτταρικό καρκίνωμα, φυσιολογικός εμβρυϊκός νεφρός, φυσιολογικό πάγκρεας και καρκίνωμα εκ πλακωδών κυττάρων κεφαλής και τραχήλου.

3.4. ΔΙΕΡΕΥΝΗΣΗ ΤΩΝ 3' ΑΜΕΤΑΦΡΑΣΤΩΝ ΠΕΡΙΟΧΩΝ (UTRs) ΤΩΝ ΝΕΩΝ ΜΕΤΑΓΡΑΦΩΝ SCAF1

Έπειτα από τον προσδιορισμό των προτύπων έκφρασης των μεταγράφων SCAF1 v.2 – v.16, έγινε η διερεύνηση των 3' αμετάφραστων περιοχών (UTRs) (Εικόνα 40, Α) με χρήση nested 3' RACE, όπου κάθε αμπλικόνιο που προέκυψε (αντίστοιχο για κάθε καρκινικό και φυσιολογικό ιστό που χρησιμοποιήθηκε) καθαρίστηκε και αλληλουχήθηκε με Sanger αλληλούχηση. Σύμφωνα με τις κορυφές φθορισμού των χρωματογραφημάτων που προέκυψαν, τόσο η γνωστή όσο και η προβλεφθείσα 3' αμετάφραστη περιοχή ανιχνεύθηκαν σε κάθε αμπλικόνιο (Εικόνα 40, Β). Ωστόσο, σε κάθε δείγμα που αλληλουχήθηκε, το ύψος των κορυφών που αντιστοιχούσαν στην προβλεφθείσα 3' UTR, ήταν σημαντικά υψηλότερες συγκρινόμενες με εκείνες που αντιπροσώπευαν τη γνωστή 3' UTR (Εικόνα 40, Γ). Αυτό το εύρημα υποδηλώνει ότι η προβλεφθείσα 3' UTR παρουσιάζει υψηλότερα επίπεδα έκφρασης και ως εκ τούτου μπορεί να θεωρηθεί ότι τα νέα εναλλακτικά μετάγραφα που παρουσιάστηκαν στη συγκεκριμένη μελέτη, είναι πιο πιθανό να περιλαμβάνουν την προβλεφθείσα 3' UTR.



Εικόνα 40: Α) Σχηματική απεικόνιση της νουκλεοτιδικής αλληλουχίας του τελευταίου εξονίου του γονιδίου *SCAF1*, Β) Αποτελέσματα ηλεκτροφόρησης όλων των αμπλικονίων που προέκυψαν από εσωτερική 3' RACE, Γ) Χρωματογράφημα κορυφών φθορισμού που προέκυψε από την Sanger αλληλούχηση των αμπλικονίων.

3.5. ΝΕΕΣ ΠΙΘΑΝΕΣ ΠΡΩΤΕΪΝΙΚΕΣ ΙΣΟΜΟΡΦΕΣ ΚΑΙ ΟΙ ΕΠΙΚΡΑΤΕΣΤΕΡΕΣ ΤΡΙΣΔΙΑΣΤΑΤΕΣ ΔΟΜΕΣ ΤΟΥΣ

Η αλληλουχία καθενός από τα νέα μετάγραφα που ταυτοποιήθηκαν ελέγχθηκε όπως έχει ήδη αναφερθεί, για την ύπαρξη ανοιχτού πλαισίου ανάγνωσης (open reading frame, ORF). Κατόπιν του ελέγχου που πραγματοποιήθηκε, προέκυψε ότι από τα συνολικά 15 εναλλακτικά μετάγραφα του γονιδίου *SCAF1* που ταυτοποιήθηκαν και ελέγχθηκαν, τα 7 από αυτά (*SCAF1* v.3, v.4, v.6, v.11, v.13, v.15 και v.16) προβλέπεται να κωδικοποιούν νέες πρωτεϊνικές ισομορφές, σε αντίθεση με τα υπόλοιπα 8, που φαίνεται να μη κωδικοποιούν για κάποια πρωτεϊνική ισομορφή. η τρισδιάστατη απεικόνιση των επικρατέστερων δομών των 7 πιθανών νέων πρωτεϊνικών ισομορφών (Εικόνα 41), πραγματοποιήθηκε με τη βοήθεια του εργαλείου I-TASSER.

Πιθανές δομές ισομορφής 3









Εικόνα 41: Απεικόνιση των τρισδιάστατων δομών των 7 νέων πιθανών πρωτεϊνικών ισομορφών του γονιδίου *SCAF1*, όπως αυτές προέκυψαν από τη χρησιμοποίηση του εργαλείου Ι - TASSER. Για την κάθε νέα ισομορφή του γονιδίου *SCAF1* παρουσιάζονται τέσσερεις από τις επικρατέστερες πιθανές τρισδιάστατες δομές της ισομορφής.

4. ΣΥΖΗΤΗΣΗ

Τα τελευταία χρόνια με τη συμβολή διαφόρων επιτευγμάτων στα οποία βεβαίως συγκαταλέγεται και η έλευση της αλληλούχησης νέας γενιάς, έχουμε διευρύνει τις γνώσεις μας σε θέματα που αφορούν όλες σχεδόν τις πτυχές του RNA, περιλαμβανομένης της αναγνώρισης ότι το εναλλακτικό μάτισμα είναι ο κανόνας και όχι η εξαίρεση, και της ανακάλυψης ότι στα κύτταρά μας υπάρχουν αμέτρητα μη κωδικά μόρια RNA (non – coding RNAs), τα οποία διαδραματίζουν σημαντικό ρόλο σε κάθε σχεδόν πτυχή της γονιδιακής ρύθμισης και της κυτταρικής λειτουργίας. Οι μέθοδοι αλληλούχησης νέας γενιάς έχουν επιτρέψει τον προσδιορισμό πληθυσμών RNA που πριν δε φανταζόμασταν ότι υπάρχουν, μερικοί από τους οποίους είναι τα miRNAs, lncRNAs, lincRNAs, piRNAs και circRNA. Επίσης, έχουν αποκαλύψει ότι υπάρχει πολύ μεγαλύτερη ποικιλία και ποικίλη ρύθμιση στις διαδικασίες επεξεργασίας του RNA, από αυτή που πιστεύαμε κατά το παρελθόν.

Την επόμενη δεκαετία, η αλματώδης και καινοτόμος τεχνολογία της αλληλούχησης νέας γενιάς, θα επιταχύνει αναμφίβολα τις ανακαλύψεις σε γονιδιακό και μεταγραφικό επίπεδο και θα βοηθήσει τους ερευνητές να τις μετατρέψουν σε κλινικά πολύτιμες πληροφορίες, συμβάλλοντας κατά αυτόν τον τρόπο στη διάγνωση και την πρόγνωση πολλών ασθενειών, συμπεριλαμβανομένου του καρκίνου [224, 225].

Στην πραγματικότητα, η αλληλούχηση νέας γενιάς έχει ήδη προσφέρει νέες γνώσεις σχετικά με τη γονιδιωματική έρευνα επιτρέποντας τη διερεύνηση της πολυπλοκότητας του εναλλακτικού ματίσματος στο ανθρώπινο μεταγράφωμα και την ανακάλυψη νέων μεταγράφων, ακόμη και εκείνων που εμφανίζουν πολύ χαμηλή έκφραση [226-228]. Το εναλλακτικό μάτισμα συμβαίνει στην πλειοψηφία των κυρίων μεταγράφων των γονιδίων του ανθρώπου, ρυθμίζοντας τον εντοπισμό των πρωτεϊνών, τις ενζυμικές τους ιδιότητες, καθώς επίσης και την αλληλεπίδρασή τους με μόρια προσδέτες, επιδρώντας ισχυρά στον κυτταρικό πολλαπλασιασμό και στην κυτταρική επιβίωση. Σημαντικό είναι ακόμα το γεγονός ότι το εναλλακτικό μάτισμα που λαμβάνει χώρα στα κύρια mRNAs των γονιδίων που σχετίζονται με τον καρκίνο, μπορεί επίσης να επηρεάσει τον έλεγχο του κυτταρικού κύκλου, τον πολλαπλασιασμό, την απόπτωση, την αγγειογένεση, καθώς και την διήθηση και τη μετάσταση [229, 230]. Παραδείγματος χάριν, μια παραλλαγή της 3' UTR του ογκογονιδίου *KRAS* (*KRAS* – variant), φάνηκε να σχετίζεται τόσο με τον κίνδυνο εμφάνισης καρκίνου όσο και με τη μεταβολή της βιολογίας του όγκου, αποτελώντας βιοδείκτη κακής έκβασης σε ασθενείς με επιθηλιακό καρκίνο των ωοθηκών [231].

Το γονίδιο SCAF1 στον άνθρωπο έχει επίσης χαρακτηριστεί ως γονίδιο που σχετίζεται με τον καρκίνο, δεδομένου ότι οι κλινικές μελέτες έχουν δείξει σημαντική σύνδεση μεταξύ των επιπέδων έκφρασης του mRNA και του καρκίνου. Πιο συγκεκριμένα, μελέτες ανάλυσης έκφρασης του γονιδίου SCAF1 σε ασθενείς με καρκίνο μαστού και ωοθηκών, έχουν δείξει ότι η υπερέκφραση του SCAF1 mRNA είναι ένας ανεξάρτητος δυσμενής προγνωστικός δείκτης επιβίωσης ελευθέρας νόσου (disease-free survival, DFS), ενώ τα χαμηλά επίπεδα έκφρασης του mRNA συνδέονται με μακράς διάρκειας ελευθέρας νόσου επιβίωση (long DFS) και ολικής επιβίωσης (OS) [157, 161]. Επιπλέον, η έκφραση του SCAF1 mRNA βρέθηκε να σχετίζεται με την εξέλιξη του καρκίνου του παχέος εντέρου, καθώς η έκφρασή του είναι υψηλότερη στα αρχικά στάδια της ογκογένεσης και μειώνεται καθώς ο καρκίνος εξελίσσεται [162]. Τέλος, ασυνήθιστα επίπεδα έκφρασης SCAF1 mRNA έχουν επίσης προσδιοριστεί στη κυτταρική σειρά HL-60 οξείας προμυελοκυτταρικής λευχαιμίας ανθρώπου μετά από αγωγή με σισπλατίνη και βλεομυκίνη. Τα επίπεδα mRNA του SCAF1 βρέθηκαν σημαντικά υψηλότερα κατά την αγωγή με βλεομυκίνη, ενώ χαμηλότερα με σισπλατίνη. Επομένως, η έκφραση του mRNA του SCAF1 μπορεί να χρησιμεύσει ως ένας νέος πιθανός μοριακός δείκτης, διαδραματίζοντας σημαντικό ρόλο στο αποτέλεσμα της χημειοθεραπείας στη λευχαιμία στον άνθρωπο [163].

Στην παρούσα μελέτη, ανακαλύφθηκαν συνολικά 15 νέα μετάγραφα του γονιδίου SCAF1 (SCAF1 v.2 - v16) και οι αλληλουχίες τους κατατέθηκαν στη GenBank[®]. Αυτές τα νέα μετάγραφα μπορούν επίσης να έχουν ιδιότητες βιοδείκτη, αντιπροσωπεύοντας εν δυνάμει βιοδείκτες για πολλές κακοήθειες στον άνθρωπο. Για το λόγο αυτό, η ποσοτικοποίησή τους και η ανάλυση έκφρασης σε καρκινικά και μη καρκινικά ζεύγη δειγμάτων χρήζει περαιτέρω διερεύνησης.

Παρόλο που οι αλληλουχίες των νέων μεταγράφων που ανακαλύφθηκαν ξεκινούν από την περιοχή που στοχεύει ο εμπρόσθιος ειδικός για το γονίδιο εκκινητής που χρησιμοποιήθηκε κατά την μοριακή κλωνοποίηση με RT-PCR και συνεπώς η 5' UTR παραμένει απροσδιόριστη, επτά από τα νέα αυτά μετάγραφα έχουν ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιούν νέες πρωτεϊνικές ισομορφές. Αντίθετα, τα υπόλοιπα οκτώ μετάγραφα μπορεί να είναι υποψήφια για το μονοπάτι της μη νοηματικής διαμεσολαβούμενης αποικοδόμησης (Nonsense-mediated decay, NMD), καθώς εμφανίζουν πρόωρη αλληλουχία τερματισμού [232-234].

Παρά το γεγονός ότι, τα επτά νέα SCAF1 μετάγραφα που ανακαλύφθηκαν προβλέπεται να κωδικοποιούν για νέες πρωτεϊνικές ισομορφές, παρουσιάζουν κρίσιμες διαφορές στην αμινοξική τους αλληλουχία και τη δευτεροταγή δομή τους, σε σύγκριση με τη γνωστή - κύρια πρωτεΐνη SCAF1. Όπως αναφέρθηκε προηγουμένως, το γονίδιο SCAF1, κωδικοποιεί για την πρωτεΐνη SCAF1, μια πρωτεΐνη που αποτελείται από 1312 αμινοξέα και έχει τις SCAFs χαρακτηριστικές περιοχές και συγκεκριμένα την περιοχή αλληλεπίδρασης με την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης ΙΙ (RNAP II) και την πλούσια σε Arg / Ser περιοχή (SR). Έτσι φαίνεται να συνδέεται η μεταγραφή με το μάτισμα. Η σημασία αυτών των SR σχετιζόμενων πρωτεϊνών στη διαδικασία του ματίσματος επιβεβαιώνεται συνεχώς από διάφορες έρευνες. Επίσης, μελέτες έχουν υποστηρίξει τη συμμετοχή των SR πρωτεϊνών στη ρύθμιση του εναλλακτικού ματίσματος και κατ' επέκταση τη συσχέτισή τους με την ογκογένεση.

Επιπλέον, η πρωτεΐνη SCAF1 περιέχει δύο αρνητικά φορτισμένες περιοχές πολυγλουταμικού οξέος (Ε) και ένα πλούσιο σε Arg/Asp μοτίβο. Επίσης έχουν ταυτοποιηθεί διάφορες πιθανές θέσεις μετα-μεταφραστικής τροποποίησης, συμπεριλαμβανομένων πολυάριθμων πιθανών θέσεων είτε για Ο- είτε Νγλυκοζυλίωση και αρκετές πιθανές θέσεις φωσφορυλίωσης από cAMP εξαρτώμενη πρωτεϊνική κινάση (PKA), πρωτεϊνική κινάση C (PKC) και κινάση καζεΐνης 2 [157].

Με βάση τα πειραματικά αποτελέσματα της παρούσας μελέτης, τα επτά νέα SCAF1 μετάγραφα (SCAF1 v.3, v.4, v.6, v.11, v.13, v.15 και v.16) έχουν ανοιχτό πλαίσιο ανάγνωσης και προβλέπεται να κωδικοποιούν για νέες πρωτεϊνικές ισομορφές SCAF1 και χαρακτηρίζονται επίσης από σημαντικά μικρότερου μήκους αλληλουχίες mRNA, σε σχέση με το κύριο μετάγραφο SCAF1. Πρέπει να σημειωθεί
ότι όλα αυτά τα νέα μετάγραφα στερούνται το εξόνιο 7, το οποίο αποτελείται από 2838 νουκλεοτίδια και είναι με διαφορά το μεγαλύτερο εξόνιο του γονιδίου, το οποίο περιέχει σημαντικές περιοχές (domains) για τη λειτουργικότητα της πρωτεΐνης SCAF1. Επιπροσθέτως, όλες οι προβλεφθείσες νέες ισομορφές έχουν σημαντικά περικομμένες αμινοξικές ακολουθίες και τελείως διαφορετικές δευτεροταγείς δομές, σε σύγκριση με την κύρια SCAF1 πρωτεΐνη. Ακόμη, όλες αυτές οι ισομορφές αν και διαθέτουν περιοχή αλληλεπίδρασης με την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης ΙΙ (αμινοξική αλληλουχία KKLHTQERAVEEVKLAIKPYYQKKDITKEEYKDILRK AVHKICHSKSGEINPVKVSNLVRAYVQRYR), δε διαθέτουν την περιοχή πλούσια σε Arg / Ser (SR) και επομένως δε θεωρείται πιθανό να διαδραματίζουν κάποιο ρόλο στη ρύθμιση του εναλλακτικού ματίσματος. Παρ' όλα αυτά, το αν αυτά τα νέα SCAF1 μετάγραφα, τα οποία προβλέπεται να κωδικοποιούν νέες ισομορφές του SCAF1, μεταφράζονται πράγματι σε λειτουργικές πρωτεΐνες, μένει να διερευνηθεί περεταίρω πειραματικά.

Σε αντίθεση με τα επτά νέα μετάγραφα που συζητήθηκαν προηγουμένως (SCAF1 v.3, v.4, v.6, v.11, v.13, v.15 και v.16), τα υπόλοιπα οκτώ νέα μετάγραφα SCAF1 περιέχουν πρώιμο κωδικόνιο λήξης και ως εκ τούτου είναι υποψήφια για NMD. Δεδομένου ότι, δεν προβλέπεται να κωδικοποιούν πρωτεϊνικές ισομορφές του SCAF1, αντιπροσωπεύουν long non-coding RNAs (lncRNAs). Ωστόσο, η ύπαρξή τους μπορεί να έχει μεγάλη σημασία για την κυτταρική βιολογία και την ομοιόσταση. Σε αντίθεση με τα miRNAs, για τη βιογένεση των lncRNAs υπάρχει περιορισμένη βιβλιογραφία και ως εκ τούτου μένει να απαντηθούν πολλά ερωτήματα. Μέχρι πρόσφατα, πολύ λίγα lncRNAs επισημάνθηκαν στο ανθρώπινο γονιδίωμα. Παρά το γεγονός ότι η πλειονότητα των lncRNAs θεωρήθηκε αρχικά ως "μεταγραφικός θόρυβος", ο οποίος στερείται παντελώς οποιαδήποτε βιολογική λειτουργία, επιβεβαιώνεται πλέον από έναν αξιοσημείωτο αριθμό μελετών, ότι διαθέτουν την ικανότητα να αλληλεπιδρούν με το DNA, συμμετέχοντας έτσι σε πολλούς κυτταρικούς μηχανισμούς [235, 236].

Αν και οι εμπλεκόμενοι μηχανισμοί δεν είναι ακόμη απολύτως σαφείς, τα lncRNAs έχουν την ικανότητα να ρυθμίζουν την γονιδιακή έκφραση και να συμμετέχουν σε ένα ευρύ φάσμα βιολογικών διεργασιών, οι οποίες περιλαμβάνουν τη διατήρηση της ακεραιότητας του γονιδιώματος, το γενετικό εντύπωμα [237], την κυτταρική διαφοροποίηση [238], τη μεθυλίωση του DNA [239], την τροποποίηση των ιστονών [240] καθώς και την αναδόμηση της χρωματίνης [241]. Επίσης, τα νέα μη κωδικά μετάγραφα *SCAF1*, θα μπορούσαν να δράσουν ως συν-ενεργοποιητές, δεσμευόμενα σε μεταγραφικούς παράγοντες, ενισχύοντας έτσι την μεταγραφική τους ικανότητα [121, 242], ή να έχουν ανασταλτική επίδραση σε πολλά κρίσιμα στάδια της μεταγραφής ενός γονιδίου όπως την έναρξη, την επιμήκυνση ή τη διαδικασία τερματισμού, καταστέλλοντας έτσι τη γονιδιακή έκφραση [243, 244].

Εκτός από τη ρυθμιστική τους λειτουργία στη φυσιολογία των κυττάρων, έχουν βρεθεί στοιχεία που συνδέουν τα lncRNAs με την ανάπτυξη και την εξέλιξη του καρκίνου σε πολλές ανθρώπινες κακοήθειες [245], συμπεριλαμβανομένου του καρκίνου του πνεύμονα [246], του καρκίνου του παγκρέατος [247], του νευροβλαστώματος [248], του καρκίνου του παχέος εντέρου [249], καθώς και του καρκίνου του στομάχου [250]. Πολυάριθμες αναφορές έχουν δείξει ότι τα lncRNAs εκφράζονται ασυνήθιστα σε καρκινικούς ιστούς, εμπλεκόμενα έτσι σε ογκογονικά και ογκοκατασταλτικά μονοπάτια, όπως τα p53, MYC και NF-kB [251, 252].

Το γεγονός ότι το γονίδιο SCAF1 εμπλέκεται στο μάτισμα του pre-mRNA, καθιστά τον ρόλο του σημαντικό για την κυτταρική ομοιόσταση και όπως αναμενόταν, το κύριο μετάγραφό του εκφράζεται ευρέως σε πολλούς φυσιολογικούς ιστούς. Αντίστοιχα και τα δεκαπέντε νέα μετάγραφα SCAF1 (SCAF1 v.2 - v16) που ανακαλύφθηκαν κατά την παρούσα μελέτη, εμφάνισαν με βάση τα αποτελέσματα της ηλεκτροφόρησης, ευρύ προφίλ έκφρασης στην ομάδα των εξεταζόμενων cDNAs. Ωστόσο, δεδομένου ότι το γονίδιο SCAF1 εμπλέκεται σε πολλούς καρκίνους στον άνθρωπο, καθιστώντας το ως πιθανό βιοδείκτη, η ποσοτικοποίηση των παρουσιαζόμενων νέων μεταγράφων σε ανθρώπινα καρκινικά και μη καρκινικά δείγματα καθίσταται αναγκαία, καθώς μπορεί να έχει κλινικές εφαρμογές σε πολλές ανθρώπινες κακοήθειες.

104

Ταυτοποίηση, ανάλυση και κλινική αξιολόγηση εναλλακτικών μεταγράφων γονιδίων καθώς και μη κωδικών μορίων RNA (ncRNAs), στο γενετικό τόπο 19q13.3-13.4, με χρήση δεδομένων αλληλούχησης νέας γενιάς, σε καρκινικά κύτταρα.

Ράπτης Γεώργιος

Ερευνητική Εργασία Διπλώματος Ειδίκευσης Τμήμα Βιολογίας, Ε.Κ.Π.Α.

ΠΕΡΙΛΗΨΗ

Το γονίδιο *SCAF1* (SR-related CTD associated factor 1) στον άνθρωπο, αποτελεί ένα νέο μέλος της, πλούσιας σε Ser/arg, υπεροικογένειας SR παραγόντων ματίσματος του πρόδρομου mRNA, το οποίο έχει ανακαλυφθεί και κλωνοποιηθεί από μέλη της ερευνητικής μας ομάδας. Η πρωτεΐνη SCAF1 αλληλεπιδρά με την καρβοξυτελική περιοχή (CTD) της μεγάλης υπομονάδας της RNA πολυμεράσης II (RNAP II) και εμπλέκεται στο μάτισμα του πρόδρομου mRNA. Παρόλο που βρέθηκε να εκφράζεται ευρέως σε πολλούς ανθρώπινους ιστούς, τα επίπεδα του mRNA ποικίλουν πολύ.

Η σημαντική σχέση του *SCAF1* με τον καρκίνο έχει επιβεβαιωθεί από πολλές μελέτες, καθώς το μετάγραφο mRNA του *SCAF1* βρέθηκε να υπερεκφράζεται σε όγκους μαστού και ωοθηκών, επιβεβαιώνοντας με αυτόν τον τρόπο τη σημαντική προγνωστική του αξία, ως καρκινικό βιοδείκτη και στις δύο αυτές κακοήθειες στον άνθρωπο.

Στη μελέτη αυτή, περιγράφεται η ανακάλυψη και η μοριακή κλωνοποίηση δεκαπέντε νέων μεταγράφων του γονιδίου SCAF1 στον άνθρωπο (SCAF1 v.2 v.16), με χρήση εσωτερικής PCR (nested PCR) και της τεχνολογίας αλληλούχησης νέας γενιάς (next generation sequencing, NGS). Συγκεκριμένα, εκτενής βιοπληροφορική ανάλυση αποκάλυψε ότι αυτά τα νέα εναλλακτικά μετάγραφα SCAF1 περιλαμβάνουν ένα σύνολο εννέα νέων εναλλακτικών γεγονότων συρραφής μεταξύ των γνωστών εξονίων του γονιδίου, παράγοντας έτσι επτά νέα μετάγραφα SCAF1 με ανοιχτό πλαίσιο ανάγνωσης, τα οποία προβλέπεται να κωδικοποιούν για νέες ισομορφές SCAF1, και οκτώ νέα μετάγραφα SCAF1 με πρόωρα κωδικόνια λήξης που είναι πιθανώς μη κωδικά μόρια RNAs (long non – coding RNAs, lncRNAs). Επιπροσθέτως, ανακαλύφθηκε μια νέα 3' UTR χρησιμοποιώντας 3' RACE και η επικύρωσή της πραγματοποιήθηκε με αλληλούχηση με τη μέθοδο Sanger. Προκειμένου να επικυρωθούν τα ευρήματα της αλληλούχησης νέας γενιάς καθώς και να διερευνηθεί το πρότυπο έκφρασης κάθε νέου μεταγράφου, διεξήχθησαν πειράματα RT-PCR χρησιμοποιώντας ειδικούς εκκινητές για το κάθε μετάγραφο.

Τέλος, λαμβάνοντας υπόψιν ότι το *SCAF1* εμπλέκεται σε πολλές κακοήθειες στον άνθρωπο, γεγονός που το καθιστά ως δυνητικό βιοδείκτη, η ποσοτικοποίηση των νέων μεταγράφων σε ανθρώπινα δείγματα, μπορεί να έχει κλινικές εφαρμογές σε διάφορους τύπους καρκίνου.

Discovery and expression analysis of novel transcripts and non-coding RNAs (ncRNAs) on chromosome 19q13.3-13.4 in human cancer cells using Next-Generation Sequencing

Raptis Georgios

Master thesis

Department of Biochemistry and Molecular Biology, University of Athens

ABSTRACT

The human SR-related CTD associated factor 1 (*SCAF1*) gene is a new member of the human SR (Ser/Arg-rich) superfamily of pre-mRNA splicing factors, which has been discovered and cloned by members of our group. SCAF1 interacts with the CTD domain of the RNA polymerase II and is firmly involved in pre-mRNA splicing. Although it was found to be expressed widely in multiple human tissues, its mRNA levels vary a lot.

The significant relation of *SCAF1* with cancer has been confirmed by many studies, since *SCAF1* mRNA transcript was found to be overexpressed in breast and ovarian tumors, confirming its significant prognostic value as a cancer biomarker in both these human malignancies.

In this study, we describe the discovery and molecular cloning of fifteen novel transcripts of the human *SCAF1* gene (*SCAF1* v.2 - v.16), using nested PCR and NGS technology. In detail, extensive bioinformatic analysis revealed that these novel *SCAF1* splice variants comprise a total of nine novel alternative splicing events between the annotated exons of the gene, thus producing seven novel *SCAF1* transcripts with open-reading frames, which are predicted to encode novel SCAF1 isoforms and eight novel *SCAF1* transcripts with premature termination codons that are likely long non-coding RNAs. Additionally, a novel 3' UTR was discovered using 3' RACE and was validated with Sanger sequencing. In order to validate the NGS findings as well as to investigate the expression profile of each

novel transcript, RT-PCR experiments were carried out with the use of variantspecific primers.

In conclusion, since *SCAF1* is implicated in many human malignancies, qualifying as a potential biomarker, the quantification of the presented novel transcripts in human samples may have clinical applications in different types of cancer.

5. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Venter, J.C., et al., *The sequence of the human genome.* Science, 2001.
 291(5507): p. 1304-51.
- Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. 409(6822): p. 860-921.
- Dahm, R., Friedrich Miescher and the discovery of DNA. Dev Biol, 2005.
 278(2): p. 274-88.
- 4. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
- Padmanabhan, R., E. Jay, and R. Wu, Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4. Proc Natl Acad Sci U S A, 1974. 71(6): p. 2510-4.
- Onaga, L.A., *Ray Wu as Fifth Business: Deconstructing collective memory in the history of DNA sequencing.* Stud Hist Philos Biol Biomed Sci, 2014. 46: p. 1-14.
- Wu, R. and E. Taylor, Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA. J Mol Biol, 1971. 57(3): p. 491-511.
- Padmanabhan, R. and R. Wu, Nucleotide sequence analysis of DNA. IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis.
 Biochem Biophys Res Commun, 1972. 48(5): p. 1295-302.
- 9. Wu, R., C.D. Tu, and R. Padmanabhan, *Nucleotide sequence analysis of DNA. XII. The chemical synthesis and sequence analysis of a dodecadeoxynucleotide which binds to the endolysin gene of bacteriophage lambda.* Biochem Biophys Res Commun, 1973. **55**(4): p. 1092-9.
- Jay, E., et al., DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping. Nucleic Acids Res, 1974. 1(3): p. 331-53.
- Sanger, F., et al., Nucleotide sequence of bacteriophage phi X174 DNA. Nature, 1977. 265(5596): p. 687-95.

- 12. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chainterminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
- Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA*. Proc Natl Acad Sci U S A, 1977. **74**(2): p. 560-4.
- Schuster, S.C., *Next-generation sequencing transforms today's biology*. Nat Methods, 2008. 5(1): p. 16-8.
- 15. Sethi, N., et al., *Past and future impact of next-generation sequencing in head and neck cancer.* Head Neck, 2016. **38 Suppl 1**: p. E2395-402.
- https://<u>www.ncbi.nlm.nih.gov/genbank/statistics/</u> (τελευταία επίσκεψη Αύγουστος 2017).
- Watson, J.D., *Recombinant DNA: Genes and Genomes: A Short Course*. 2007:W. H. Freeman.
- 18. Bankier, A.T., et al., *The DNA sequence of the human cytomegalovirus genome.* DNA Seq, 1991. **2**(1): p. 1-12.
- Staden, R., A strategy of DNA sequencing employing computer programs. Nucleic Acids Res, 1979. 6(7): p. 2601-10.
- 20. Franca, L.T., E. Carrilho, and T.B. Kist, *A review of DNA sequencing techniques.* Q Rev Biophys, 2002. **35**(2): p. 169-200.
- 21. 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.
- 22. Muzzey, D., E.A. Evans, and C. Lieber, *Understanding the Basics of NGS: From Mechanism to Variant Calling.* Curr Genet Med Rep, 2015. **3**(4): p. 158-165.
- 23. Li, W. and J. Freudenberg, *Mappability and read length*. Front Genet, 2014.
 5: p. 381.
- 24. https://www.illumina.com/science/education/sequencing-coverage.html (τελευταία επίσκεψη Αύγουστος 2017).
- 25. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nat Rev Genet, 2014. **15**(2): p. 121-32.
- 26. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-80.
- 27. Wheeler, D.A., et al., *The complete genome of an individual by massively parallel DNA sequencing.* Nature, 2008. **452**(7189): p. 872-6.

- Zhang, J., et al., *The impact of next-generation sequencing on genomics*. J Genet Genomics, 2011. 38(3): p. 95-109.
- 29. Anderson, M.W. and I. Schrijver, *Next generation DNA sequencing and the future of genomic medicine.* Genes (Basel), 2010. **1**(1): p. 38-69.
- 30. Ronaghi, M., et al., *Real-time DNA sequencing using detection of pyrophosphate release.* Anal Biochem, 1996. **242**(1): p. 84-9.
- 31. Ronaghi, M., M. Uhlen, and P. Nyren, *A sequencing method based on real-time pyrophosphate.* Science, 1998. **281**(5375): p. 363, 365.
- 32. Siqueira, J.F., Jr., A.F. Fouad, and I.N. Rocas, *Pyrosequencing as a tool for better understanding of human microbiomes.* J Oral Microbiol, 2012. **4**.
- 33. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome.* Science, 2005. **309**(5741): p. 1728-32.
- 34. Valouev, A., et al., A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res, 2008.
 18(7): p. 1051-63.
- Ansorge, W.J., *Next-generation DNA sequencing techniques*. N Biotechnol, 2009. 25(4): p. 195-203.
- https://<u>www.thermofisher.com/order/catalog/product/4460730</u> (τελευταία επίσκεψη Αύγουστος 2017).
- <u>http://www.openwetware.org/wiki/Church Lab:PoloProt</u> (τελευταία επίσκεψη Αύγουστος 2017).
- 38. Adessi, C., et al., Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res, 2000. 28(20):
 p. E87.
- 39. Fedurco, M., et al., *BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.* Nucleic Acids Res, 2006. 34(3): p. e22.
- 40. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
- Chen, F., et al., *The history and advances of reversible terminators used in new generations of sequencing technology.* Genomics Proteomics Bioinformatics, 2013. **11**(1): p. 34-40.

- 42. https://www.illumina.com/documents/products/datasheets/datasheethiseq-x-ten.pdf (τελευταία επίσκεψη Αύγουστος 2017).
- 43. Quail, M.A., et al., *A large genome center's improvements to the Illumina sequencing system.* Nat Methods, 2008. **5**(12): p. 1005-10.
- 44. Liu, L., et al., *Comparison of next-generation sequencing systems*. J Biomed Biotechnol, 2012. 2012: p. 251364.
- 45. Rothberg, J.M., et al., *An integrated semiconductor device enabling nonoptical genome sequencing.* Nature, 2011. **475**(7356): p. 348-352.
- 46. Diaz-Sanchez, S., et al., Next-generation sequencing: the future of molecular genetics in poultry production and food safety. Poult Sci, 2013. 92(2): p. 562-72.
- 47. Matochko, W.L. and R. Derda, *Next-generation sequencing of phagedisplayed peptide libraries.* Methods Mol Biol, 2015. **1248**: p. 249-66.
- 48. Vigliar, E., et al., *Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective.* Cytopathology, 2015. **26**(5): p. 271-83.
- 49. Merriman, B., et al., *Progress in ion torrent semiconductor chip based sequencing.* Electrophoresis, 2012. **33**(23): p. 3397-417.
- 50. Braslavsky, I., et al., *Sequence information can be obtained from single DNA molecules.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3960-4.
- 51. Harris, T.D., et al., *Single-molecule DNA sequencing of a viral genome.* Science, 2008. **320**(5872): p. 106-9.
- 52. Pushkarev, D., N.F. Neff, and S.R. Quake, *Single-molecule sequencing of an individual human genome.* Nat Biotechnol, 2009. **27**(9): p. 847-50.
- 53. Bowers, J., et al., *Virtual terminator nucleotides for next-generation DNA sequencing.* Nat Methods, 2009. **6**(8): p. 593-5.
- 54. Metzker, M.L., *Sequencing technologies the next generation*. Nat Rev Genet, 2010. 11(1): p. 31-46.
- 55. Eid, J., et al., *Real-time DNA sequencing from single polymerase molecules.*Science, 2009. **323**(5910): p. 133-8.
- 56. Bleidorn, C., *Third generation sequencing: technology and its potential impact on evolutionary biodiversity research.* Systematics and Biodiversity, 2016. 14(1): p. 1-8.

- 57. Levene, M.J., et al., *Zero-mode waveguides for single-molecule analysis at high concentrations.* Science, 2003. **299**(5607): p. 682-6.
- Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nat Rev Genet, 2016. **17**(6): p. 333-51.
- 59. Flusberg, B.A., et al., *Direct detection of DNA methylation during singlemolecule, real-time sequencing.* Nat Methods, 2010. **7**(6): p. 461-5.
- 60. Lee, H., et al., *Error correction and assembly complexity of single molecule sequencing reads.* bioRxiv, 2014.
- 61. Gupta, P.K., *Single-molecule DNA sequencing technologies for future genomics research.* Trends Biotechnol, 2008. **26**(11): p. 602-11.
- 62. Clarke, J., et al., *Continuous base identification for single-molecule nanopore DNA sequencing.* Nat Nanotechnol, 2009. **4**(4): p. 265-70.
- 63. Pareek, C.S., R. Smoczynski, and A. Tretyn, *Sequencing technologies and genome sequencing*. J Appl Genet, 2011. **52**(4): p. 413-35.
- 64. Astier, Y., O. Braha, and H. Bayley, *Toward single molecule DNA sequencing:* direct identification of ribonucleoside and deoxyribonucleoside 5'monophosphates by using an engineered protein nanopore equipped with a molecular adapter. J Am Chem Soc, 2006. **128**(5): p. 1705-10.
- 65. Rusk, N., *Cheap third-generation sequencing.* Nat Meth, 2009. **6**(4): p. 244-244.
- 66. Branton, D., et al., *The potential and challenges of nanopore sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1146-53.
- 67. Pop, M. and S.L. Salzberg, *Bioinformatics challenges of new sequencing technology.* Trends Genet, 2008. **24**(3): p. 142-9.
- Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-generation* sequencing: from basic research to diagnostics. Clin Chem, 2009. 55(4): p. 641-58.
- 69. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990.
 215(3): p. 403-10.
- Li, H., J. Ruan, and R. Durbin, *Mapping short DNA sequencing reads and calling variants using mapping quality scores.* Genome Res, 2008. 18(11): p. 1851-8.

- 71. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
- 72. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
- 73. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation*. Nat Biotechnol, 2010. **28**(5): p. 511-5.
- 74. Guttman, M., et al., *Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs.* Nat Biotechnol, 2010. **28**(5): p. 503-10.
- 75. Chaisson, M.J. and P.A. Pevzner, *Short read fragment assembly of bacterial genomes.* Genome Res, 2008. **18**(2): p. 324-30.
- 76. Yi, X., et al., *Sequencing of 50 human exomes reveals adaptation to high altitude.* Science, 2010. **329**(5987): p. 75-8.
- 77. Pickrell, J.K., et al., Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature, 2010. **464**(7289): p. 768-72.
- 78. Hiller, D., et al., *Identifiability of isoform deconvolution from junction arrays and RNA-Seq.* Bioinformatics, 2009. **25**(23): p. 3056-9.
- 79. Kharchenko, P.V., M.Y. Tolstorukov, and P.J. Park, *Design and analysis of ChIP-seq experiments for DNA-binding proteins*. Nat Biotechnol, 2008.
 26(12): p. 1351-9.
- 80. Su, Z., et al., *Next-generation sequencing and its applications in molecular diagnostics*. Expert Rev Mol Diagn, 2011. **11**(3): p. 333-43.
- 81. Gambacorti-Passerini, C., *Part I: Milestones in personalised medicine-imatinib.* Lancet Oncol, 2008. **9**(6): p. 600.
- Gelmon, K., Part II: Milestones in personalised medicine--trastuzumab. Lancet Oncol, 2008. 9(7): p. 698.
- 83. Lee, W., et al., *The mutation spectrum revealed by paired genome sequences from a lung cancer patient.* Nature, 2010. **465**(7297): p. 473-7.
- 84. Pleasance, E.D., et al., *A comprehensive catalogue of somatic mutations from a human cancer genome.* Nature, 2010. **463**(7278): p. 191-6.

- 85. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.
- 86. Gilissen, C., et al., Unlocking Mendelian disease using exome sequencing.Genome Biol, 2011. 12(9): p. 228.
- 87. Kumar, K.R., K. Lohmann, and C. Klein, *Genetics of Parkinson disease and other movement disorders.* Curr Opin Neurol, 2012. **25**(4): p. 466-74.
- 88. Lemke, J.R., et al., *Targeted next generation sequencing as a diagnostic tool in epileptic disorders.* Epilepsia, 2012. **53**(8): p. 1387-98.
- 89. de Ligt, J., et al., *Diagnostic exome sequencing in persons with severe intellectual disability.* N Engl J Med, 2012. **367**(20): p. 1921-9.
- 90. Zimprich, A., et al., A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet, 2011. 89(1): p. 168-75.
- 91. Vilarino-Guell, C., et al., *VPS35 mutations in Parkinson disease.* Am J Hum Genet, 2011. **89**(1): p. 162-7.
- 92. Haack, T.B., et al., *Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA.* Am J Hum Genet, 2012. **91**(6): p. 1144-9.
- 93. Fuchs, T., et al., *Mutations in GNAL cause primary torsion dystonia*. Nat Genet, 2013. **45**(1): p. 88-92.
- 94. Chen, W.J., et al., *Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia.* Nat Genet, 2011. **43**(12): p. 1252-5.
- 95. Soong, B.W., et al., *Exome sequencing identifies GNB4 mutations as a cause of dominant intermediate Charcot-Marie-Tooth disease.* Am J Hum Genet, 2013. **92**(3): p. 422-30.
- 96. Consortium, M., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-61.
- 97. Shi, L., et al., *The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.* Nat Biotechnol, 2010. **28**(8): p. 827-38.

- 98. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
- 99. Roberts, A., et al., *Identification of novel transcripts in annotated genomes using RNA-Seq.* Bioinformatics, 2011. **27**(17): p. 2325-9.
- 100. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature, 2009. **458**(7235): p. 223-7.
- 101. Li, J.B., et al., Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science, 2009. 324(5931): p. 1210-3.
- 102. Jacquier, A., *The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs.* Nat Rev Genet, 2009. **10**(12): p. 833-44.
- 103. Sorek, R. and P. Cossart, *Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.* Nat Rev Genet, 2010. 11(1): p. 9-16.
- 104. Lynch, K.W., *Thoughts on NGS, alternative splicing and what we still need to know.* RNA, 2015. **21**(4): p. 683-4.
- 105. Cannell, I.G., Y.W. Kong, and M. Bushell, *How do microRNAs regulate gene expression?* Biochem Soc Trans, 2008. **36**(Pt 6): p. 1224-31.
- 106. Perkel, J.M., *Visiting "noncodarnia".* Biotechniques, 2013. **54**(6): p. 301, 303-4.
- Cabili, M.N., et al., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 2011. 25(18): p. 1915-27.
- 108. Meister, G., Argonaute proteins: functional insights and emerging roles. Nat Rev Genet, 2013. 14(7): p. 447-59.
- 109. Chen, L.L., *The biogenesis and emerging roles of circular RNAs.* Nat Rev Mol Cell Biol, 2016. **17**(4): p. 205-11.
- 110. Sutherland, H.F., et al., *Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome.* Am J Hum Genet, 1996. **59**(1): p. 23-31.
- 111. Lee, M.P., et al., Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann

syndrome and is independent of insulin-like growth factor II imprinting. Proc Natl Acad Sci U S A, 1999. **96**(9): p. 5203-8.

- 112. Nemes, J.P., et al., *The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1).* Hum Mol Genet, 2000. **9**(10): p. 1543-51.
- 113. Ninomiya, S., et al., Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. Hum Mol Genet, 1996. **5**(1): p. 69-72.
- 114. Sammeth, M., S. Foissac, and R. Guigo, *A general definition and nomenclature for alternative splicing events.* PLoS Comput Biol, 2008. **4**(8): p. e1000147.
- 115. Will, C.L. and R. Luhrmann, *Spliceosome structure and function*. Cold Spring Harb Perspect Biol, 2011. **3**(7).
- Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry, Fifth Edition*. 2002: W.H. Freeman.
- 117. Dutertre, M., S. Vagner, and D. Auboeuf, *Alternative splicing and breast cancer.* RNA Biol, 2010. **7**(4): p. 403-11.
- 118. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing*. Annu Rev Biochem, 2003. **72**: p. 291-336.
- 119. Pajares, M.J., et al., *Alternative splicing: an emerging topic in molecular and clinical oncology.* Lancet Oncol, 2007. **8**(4): p. 349-57.
- 120. Rajan, P., et al., *Alternative splicing and biological heterogeneity in prostate cancer.* Nat Rev Urol, 2009. **6**(8): p. 454-60.
- 121. 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.
- 122. 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.
- 123. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.

- Keren, H., G. Lev-Maor, and G. Ast, *Alternative splicing and evolution: diversification, exon definition and function.* Nat Rev Genet, 2010. **11**(5): p. 345-55.
- 125. Lodish, H., Molecular Cell Biology. 2004: W. H. Freeman.
- 126. Lewin, B., *Genes 8*. 2004: Pearson Prentice Hall.
- 127. King, R.C., W.D. Stansfield, and P.K. Mulligan, *A Dictionary of Genetics*. 2006: Oxford University Press.
- 128. Kelemen, O., et al., *Function of alternative splicing*. Gene, 2013. **514**(1): p. 1-30.
- 129. Klug, W.S., et al., *Concepts of Genetics*. 2014: Pearson Education.
- 130. Turunen, J.J., et al., *The significant other: splicing by the minor spliceosome.*Wiley Interdiscip Rev RNA, 2013. 4(1): p. 61-76.
- 131. Patel, A.A. and J.A. Steitz, *Splicing double: insights from the second spliceosome.* Nat Rev Mol Cell Biol, 2003. **4**(12): p. 960-70.
- 132. Watson, J.D., et al., *Molecular Biology of the Gene*. 2014: Pearson.
- 133. Behzadnia, N., et al., Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J, 2007.
 26(6): p. 1737-48.
- 134. McManus, C.J. and B.R. Graveley, *RNA structure and the mechanisms of alternative splicing.* Curr Opin Genet Dev, 2011. **21**(4): p. 373-9.
- 135. Gautam, A., et al., *Cwc21p promotes the second step conformation of the spliceosome and modulates 3' splice site selection*. Nucleic Acids Res, 2015.
 43(6): p. 3309-17.
- 136. Roscigno, R.F. and M.A. Garcia-Blanco, *SR proteins escort the U4/U6.U5 trisnRNP to the spliceosome.* RNA, 1995. **1**(7): p. 692-706.
- 137. Mendes Soares, L.M. and J. Valcarcel, *The expanding transcriptome: the genome as the 'Book of Sand'.* EMBO J, 2006. **25**(5): p. 923-31.
- 138. Sugnet, C.W., et al., *Transcriptome and genome conservation of alternative splicing events in humans and mice.* Pac Symp Biocomput, 2004: p. 66-77.
- Alekseyenko, A.V., N. Kim, and C.J. Lee, *Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes.* RNA, 2007. 13(5): p. 661-70.

- 140. Koren, E., G. Lev-Maor, and G. Ast, *The emergence of alternative 3' and 5' splice site exons from constitutive exons.* PLoS Comput Biol, 2007. 3(5): p. e95.
- 141. Ast, G., *How did alternative splicing evolve?* Nat Rev Genet, 2004. 5(10): p. 773-82.
- 142. Kim, E., A. Goren, and G. Ast, *Alternative splicing: current perspectives.* Bioessays, 2008. **30**(1): p. 38-47.
- 143. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.
- 144. Tropp, B.E., *Principles of Molecular Biology*. 2012: Jones & Bartlett Learning.
- 145. Huang, C.S., et al., *Increased expression of SRp40 affecting CD44 splicing is associated with the clinical outcome of lymph node metastasis in human breast cancer.* Clin Chim Acta, 2007. **384**(1-2): p. 69-74.
- 146. Birney, E., S. Kumar, and A.R. Krainer, *Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors.* Nucleic Acids Res, 1993. **21**(25): p. 5803-16.
- 147. Jamison, S.F., et al., U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. Nucleic Acids Res, 1995.
 23(16): p. 3260-7.
- 148. Xiao, S.H. and J.L. Manley, *Phosphorylation-dephosphorylation differentially* affects activities of splicing factor ASF/SF2. EMBO J, 1998. **17**(21): p. 6359-67.
- 149. Misteli, T., *RNA splicing: What has phosphorylation got to do with it?* Current Biology. **9**(6): p. R198-R200.
- 150. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.
- 151. Zhu, J. and A.R. Krainer, *Pre-mRNA splicing in the absence of an SR protein RS domain.* Genes Dev, 2000. **14**(24): p. 3166-78.
- 152. Theissen, H., et al., *Cloning of the human cDNA for the U1 RNA-associated 70K protein.* EMBO J, 1986. **5**(12): p. 3209-17.
- 153. Blencowe, B.J., et al., *A coactivator of pre-mRNA splicing.* Genes Dev, 1998.
 12(7): p. 996-1009.

- 154. Zhang, W.J. and J.Y. Wu, *Sip1, a novel RS domain-containing protein essential for pre-mRNA splicing.* Mol Cell Biol, 1998. **18**(2): p. 676-84.
- 155. Santos-Pereira, J.M., et al., *Npl3, a new link between RNA-binding proteins and the maintenance of genome integrity.* Cell Cycle, 2014. 13(10): p. 1524-9.
- 156. Yuryev, A., et al., *The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins.* Proc Natl Acad Sci U S A, 1996. **93**(14): p. 6975-80.
- 157. Scorilas, A., et al., *Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of pre-mRNA splicing factors: overexpression in aggressive ovarian cancer.* Br J Cancer, 2001. **85**(2): p. 190-8.
- <u>http://atlasgeneticsoncology.org/Genes/SCAF1ID46074ch19q13.html</u> (τελευταία επίσκεψη Αύγουστος 2017).
- 159. Katsarou, M.E., et al., *Expression of the C-terminal domain of novel human SR-A1 protein: interaction with the CTD domain of RNA polymerase II.* Biochem Biophys Res Commun, 2005. **334**(1): p. 61-8.
- 160. Rosonina, E. and B.J. Blencowe, Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. RNA, 2004. **10**(4): p. 581-9.
- 161. Leoutsakou, T., M. Talieri, and A. Scorilas, *Prognostic significance of the expression of SR-A1, encoding a novel SR-related CTD-associated factor, in breast cancer.* Biol Chem, 2006. **387**(12): p. 1613-8.
- 162. Mathioudaki, K., et al., SR-A1, a member of the human pre-mRNA splicing factor family, and its expression in colon cancer progression. Biol Chem, 2004. 385(9): p. 785-90.
- 163. Katsarou, M.E., et al., *Effect of bleomycin and cisplatin on the expression* profile of SRA1, a novel member of pre-mRNA splicing factors, in HL-60 human promyelocytic leukemia cells. Biol Chem, 2007. **388**(8): p. 773-8.
- 164. https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf
 (τελευταία επίσκεψη Αύγουστος 2017).
- 165. Hilario, E.C., et al., An Improved Method of Predicting Extinction Coefficients for the Determination of Protein Concentration. PDA J Pharm Sci Technol, 2016.

- 166. Wilfinger, W.W., K. Mackey, and P. Chomczynski, *Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity.* Biotechniques, 1997. 22(3): p. 474-6, 478-81.
- 167. Temin, H.M. and S. Mizutani, *RNA-dependent DNA polymerase in virions of Rous sarcoma virus.* Nature, 1970. **226**(5252): p. 1211-3.
- 168. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-70.
- 169. Luo, L., et al., *Gene expression profiles of laser-captured adjacent neuronal subtypes.* Nat Med, 1999. **5**(1): p. 117-22.
- 170. Sterrenburg, E., et al., *A common reference for cDNA microarray hybridizations*. Nucleic Acids Res, 2002. **30**(21): p. e116.
- 171. Xiang, C.C., et al., *Amine-modified random primers to label probes for DNA microarrays*. Nat Biotechnol, 2002. **20**(7): p. 738-42.
- 172. Yang, I.V., et al., *Within the fold: assessing differential expression measures and reproducibility in microarray assays.* Genome Biol, 2002. **3**(11): p. research0062.
- 173. Decraene, C., et al., *Reverse transcription in the presence of dideoxynucleotides to increase the sensitivity of expression monitoring with cDNA arrays.* Biotechniques, 1999. **27**(5): p. 962-6.
- 174. Brink, A.A., et al., *Multiprimed cDNA synthesis followed by PCR is the most suitable method for Epstein-Barr virus transcript analysis in small lymphoma biopsies.* Mol Cell Probes, 1997. **11**(1): p. 39-47.
- 175. Iturriza-Gomara, M., et al., *Comparison of specific and random priming in the reverse transcriptase polymerase chain reaction for genotyping group A rotaviruses.* J Virol Methods, 1999. **78**(1-2): p. 93-103.
- 176. Jia, X., et al., *The role of alternative polyadenylation in the antiviral innate immune response.* Nat Commun, 2017. **8**: p. 14605.
- 177. Glasel, J.A., *Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios.* Biotechniques, 1995. **18**(1): p. 62-3.
- 178. Freeman, W.M., S.J. Walker, and K.E. Vrana, *Quantitative RT-PCR: pitfalls and potential.* Biotechniques, 1999. **26**(1): p. 112-22, 124-5.

- 179. Chien, A., D.B. Edgar, and J.M. Trela, *Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus.* J Bacteriol, 1976. **127**(3): p. 1550-7.
- 180. Shampo, M.A. and R.A. Kyle, *Kary B. Mullis--Nobel Laureate for procedure to replicate DNA.* Mayo Clin Proc, 2002. **77**(7): p. 606.
- 181. McPherson, M.J. and S.G. Møller, *PCR*. 2001, Oxford: BIOS Scientific Publishers.
- 182. Σκορίλας, Α., Αρχές Κλινικής Βιοχημείας και Μοριακής Διαγνωστικής. 2005:
 ΣΥΜΜΕΤΡΙΑ.
- 183. Λεκανίδου Ρένα, Τ.Σ., Ροδάκης Γεώργιος, Εισαγωγή στη Μοριακή Βιολογία.
 2007.
- 184. Brock, T.D. and H. Freeze, *Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile.* J Bacteriol, 1969. **98**(1): p. 289-97.
- 185. Hu, G., DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. DNA Cell Biol, 1993. 12(8): p. 763-70.
- 186. van Pelt-Verkuil, E., A. van Belkum, and J.P. Hays, *Principles and Technical Aspects of PCR Amplification*. 2008: Springer Netherlands.
- 187. Friedrich, M., A. Grahnert, and S. Hauschildt, *Analysis of the 3' UTR of the ART3 and ART4 gene by 3' inverse RACE-PCR.* DNA Seq, 2005. 16(1): p. 53-7.
- 188. Yeku, O. and M.A. Frohman, *Rapid amplification of cDNA ends (RACE)*. Methods Mol Biol, 2011. **703**: p. 107-22.
- 189. <u>http://www.mn-</u> net.com/Products/DNAandRNApurification/Cleanup/NucleoSpinGelandPC <u>RCleanup/tabid/1452/language/en-US/Default.aspx</u> (τελευταία επίσκεψη Αύγουστος 2017).
- 190.

https://tools.thermofisher.com/content/sfs/manuals/MAN0007218_Ion_PG M_Template_OT2_400_Kit_UG.pdf (τελευταία επίσκεψη Αύγουστος 2017).

191. Lehman, I.R., *DNA ligase: structure, mechanism, and function.* Science, 1974.
186(4166): p. 790-7.

- 192. Kanagal-Shamanna, R., *Emulsion PCR: Techniques and Applications.* Methods Mol Biol, 2016. 1392: p. 33-42.
- 193. Buermans, H.P. and J.T. den Dunnen, Next generation sequencing technology: Advances and applications. Biochim Biophys Acta, 2014.
 1842(10): p. 1932-1941.
- 194. Luthra, R., R.R. Singh, and K.P. Patel. *Clinical applications of PCR*. 2016; Available from: <u>http://dx.doi.org/10.1007/978-1-4939-3360-0</u>.
- 195. Williams, R., et al., *Amplification of complex gene libraries by emulsion PCR.* Nat Methods, 2006. 3(7): p. 545-50.
- 196. Bakos, J., et al., Enriched environment influences hormonal status and hippocampal brain derived neurotrophic factor in a sex dependent manner. Neuroscience, 2009. **164**(2): p. 788-97.
- 197. Halaihel, N., et al., *A new real time PCR-based assay for diagnosing Renibacterium salmoninarum in rainbow trout (Oncorhynchus mykiss) and comparison with other techniques.* J Microbiol Methods, 2009. **76**(1): p. 75-80.
- 198. Hamza, I.A., et al., *Detection and quantification of human bocavirus in river water.* J Gen Virol, 2009. **90**(Pt 11): p. 2634-7.
- 199. Acar, E., et al., Optimization and validation studies of the Mentype Argus X-8 kit for paternity cases. Forensic Science International: Genetics Supplement Series. 2(1): p. 47-48.
- 200. Cock, P.J., et al., The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res, 2010.
 38(6): p. 1767-71.
- 201. Deorowicz, S. and S. Grabowski, *Compression of DNA sequence reads in FASTQ format.* Bioinformatics, 2011. **27**(6): p. 860-2.
- 202. Blankenberg, D., et al., *Manipulation of FASTQ data with Galaxy.* Bioinformatics, 2010. **26**(14): p. 1783-5.
- 203. https://usegalaxy.org/ (τελευταία επίσκεψη Αύγουστος 2017).
- 204. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.
- 205. Endrullat, C., et al., *Standardization and quality management in nextgeneration sequencing*. Appl Transl Genom, 2016. **10**: p. 2-9.

- 206. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
- 207. Zhang, H., Overview of Sequence Data Formats. Methods Mol Biol, 2016.1418: p. 3-17.
- 208. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
- 209. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011.
 29(1): p. 24-6.
- 210. Ye, J., et al., *Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.* BMC Bioinformatics, 2012. **13**: p. 134.
- 211. Sambrook, J. and D.W. Russell, *Molecular Cloning: A Laboratory Manual*.2001: Cold Spring Harbor Laboratory Press.
- 212. Brody, J.R. and S.E. Kern, *History and principles of conductive media for standard DNA electrophoresis.* Anal Biochem, 2004. **333**(1): p. 1-13.
- 213. Lee, P.Y., et al., Agarose gel electrophoresis for the separation of DNA fragments. J Vis Exp, 2012(62).
- 214. Helling, R.B., H.M. Goodman, and H.W. Boyer, *Analysis of endonuclease R-EcoRI fragments of DNA from lambdoid bacteriophages and other viruses by agarose-gel electrophoresis.* J Virol, 1974. **14**(5): p. 1235-44.
- 215. Sigmon, J. and L.L. Larcom, *The effect of ethidium bromide on mobility of DNA fragments in agarose gel electrophoresis.* Electrophoresis, 1996. **17**(10): p. 1524-7.
- 216. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction.* Nat Methods, 2015. **12**(1): p. 7-8.
- 217. Rigden, D.J., *From Protein Structure to Function with Bioinformatics*. 2017: Springer Netherlands.
- 218.

http://www.predictioncenter.org/casp12/zscores_final.cgi?model_type=firs t&gr_type=server_only (τελευταία επίσκεψη Αύγουστος 2017).

219. Wu, S. and Y. Zhang, *LOMETS: a local meta-threading-server for protein structure prediction.* Nucleic Acids Res, 2007. **35**(10): p. 3375-82.

- 220. Roy, A., J. Yang, and Y. Zhang, *COFACTOR: an accurate comparative algorithm for structure-based protein function annotation.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W471-7.
- 221. Yang, J. and Y. Zhang, *Protein Structure and Function Prediction Using I-TASSER.* Curr Protoc Bioinformatics, 2015. **52**: p. 5 8 1-15.
- 222. Zhang, Y. and J. Skolnick, *SPICKER: a clustering approach to identify nearnative protein folds.* J Comput Chem, 2004. **25**(6): p. 865-71.
- Li, Y. and Y. Zhang, *REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks.* Proteins, 2009. **76**(3): p. 665-76.
- 224. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
- 225. Hawkins, R.D., G.C. Hon, and B. Ren, *Next-generation genomics: an integrative approach.* Nat Rev Genet, 2010. **11**(7): p. 476-86.
- 226. Adamopoulos, P.G., C.K. Kontos, and A. Scorilas, *Identification and molecular cloning of novel transcripts of the human kallikrein-related peptidase 10 (KLK10) gene using next-generation sequencing.* Biochem Biophys Res Commun, 2017. **487**(4): p. 776-781.
- 227. Adamopoulos, P.G., et al., *Identification of novel alternative splice variants of the BCL2L12 gene in human cancer cells using next-generation sequencing methodology.* Cancer Lett, 2016. **373**(1): p. 119-129.
- 228. Feng, H., Z. Qin, and X. Zhang, Opportunities and methods for studying alternative splicing in cancer with RNA-Seq. Cancer Lett, 2013. 340(2): p. 179-91.
- 229. Venables, J.P., *Unbalanced alternative splicing and its significance in cancer*.Bioessays, 2006. **28**(4): p. 378-86.
- 230. Pal, S., R. Gupta, and R.V. Davuluri, *Alternative transcription and alternative splicing in cancer.* Pharmacol Ther, 2012. **136**(3): p. 283-94.
- 231. Ratner, E.S., et al., *A KRAS variant is a biomarker of poor outcome, platinum chemotherapy resistance and a potential target for therapy in ovarian cancer.* Oncogene, 2012. **31**(42): p. 4559-66.
- 232. Thermann, R., et al., *Binary specification of nonsense codons by splicing and cytoplasmic translation.* EMBO J, 1998. **17**(12): p. 3484-94.

- 233. 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.
- 234. 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.
- 235. Chen, L.L. and G.G. Carmichael, *Decoding the function of nuclear long noncoding RNAs.* Curr Opin Cell Biol, 2010. **22**(3): p. 357-64.
- 236. Yang, Y., L. Wen, and H. Zhu, *Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein.* Cell Biosci, 2015. **5**: p. 59.
- 237. Kanduri, C., Long noncoding RNAs: Lessons from genomic imprinting. Biochim Biophys Acta, 2016. 1859(1): p. 102-11.
- 238. Labott, A.T. and V. Lopez-Pajares, *Epidermal differentiation gene regulatory networks controlled by MAF and MAFB.* Cell Cycle, 2016. **15**(11): p. 1405-9.
- 239. Kumegawa, K., et al., *A genomic screen for long noncoding RNA genes epigenetically silenced by aberrant DNA methylation in colorectal cancer.* Sci Rep, 2016. **6**: p. 26699.
- 240. Sun, T.T., et al., *LncRNA GClnc1 Promotes Gastric Carcinogenesis and May Act as a Modular Scaffold of WDR5 and KAT2A Complexes to Specify the Histone Modification Pattern.* Cancer Discov, 2016. **6**(7): p. 784-801.
- 241. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.
- 242. Caretti, G., et al., *The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation.* Dev Cell, 2006.
 11(4): p. 547-60.
- 243. Mazo, A., et al., Transcriptional interference: an unexpected layer of complexity in gene regulation. Journal of Cell Science, 2007. 120(16): p. 2755-2761.
- 244. Willingham, A.T., et al., *A strategy for probing the function of noncoding RNAs finds a repressor of NFAT.* Science, 2005. **309**(5740): p. 1570-3.
- 245. Previdi, M.C., et al., *Noncoding RNAs as novel biomarkers in pancreatic cancer: what do we know?* Future Oncol, 2017. **13**(5): p. 443-453.

- 246. Chen, Y., et al., *The Emerging Role and Promise of Long Noncoding RNAs in Lung Cancer Treatment.* Cell Physiol Biochem, 2016. **38**(6): p. 2194-206.
- 247. Peng, J.F., et al., *Noncoding RNAs and pancreatic cancer*. World J Gastroenterol, 2016. **22**(2): p. 801-14.
- 248. Pandey, G.K. and C. Kanduri, *Long noncoding RNAs and neuroblastoma*. Oncotarget, 2015. **6**(21): p. 18265-75.
- 249. Saus, E., et al., Long Non-Coding RNAs As Potential Novel Prognostic Biomarkers in Colorectal Cancer. Front Genet, 2016. **7**: p. 54.
- 250. Wang, J., et al., *Long noncoding RNAs in gastric cancer: functions and clinical applications.* Onco Targets Ther, 2016. **9**: p. 681-97.
- 251. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.
- 252. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.

6. ПАРАРТНМА

Μέρος της παρούσας έρευνας δημοσιεύθηκε στο συνέδριο της Διεθνούς Εταιρείας Ενζυμολογίας (International Society for Enzymology, ISE) ADVANCES IN LABORATORY MEDICINE AND PATHOBIOLOGY που έλαβε χώρα στις 16-19 Ιουνίου 2017 στη Σαντορίνη.