



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
ΤΟΜΕΑΣ ΧΕΙΡΟΥΡΓΙΚΟΣ
ΕΡΓΑΣΤΗΡΙΟ ΙΑΤΡΙΚΗΣ ΓΕΝΕΤΙΚΗΣ ΣΤΗΝ ΚΛΙΝΙΚΗ ΠΡΑΞΗ

ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΚΑΙ ΣΥΣΧΕΤΙΣΗ ΔΟΜΩΝ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΜΕ
ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΕΣ ΑΛΛΗΛΟΥΧΙΕΣ ΤΟΥ DNA ΣΕ ΚΑΤΑΣΤΑΣΕΙΣ
ΓΕΝΕΤΙΚΗΣ ΑΣΤΑΘΕΙΑΣ ΜΕ ΕΡΓΑΛΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΑΡΗΣ Π. ΛΑΔΙΑΣ
ΒΙΟΛΟΓΟΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ
ΙΩΑΝΝΙΝΑ 2019



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
ΤΟΜΕΑΣ ΧΕΙΡΟΥΡΓΙΚΟΣ
ΕΡΓΑΣΤΗΡΙΟ ΙΑΤΡΙΚΗΣ ΓΕΝΕΤΙΚΗΣ ΣΤΗΝ ΚΛΙΝΙΚΗ ΠΡΑΞΗ

ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΚΑΙ ΣΥΣΧΕΤΙΣΗ ΔΟΜΩΝ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΜΕ
ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΕΣ ΑΛΛΗΛΟΥΧΙΕΣ ΤΟΥ DNA ΣΕ ΚΑΤΑΣΤΑΣΕΙΣ
ΓΕΝΕΤΙΚΗΣ ΑΣΤΑΘΕΙΑΣ ΜΕ ΕΡΓΑΛΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΑΡΗΣ Π. ΛΑΔΙΑΣ
ΒΙΟΛΟΓΟΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ
ΙΩΑΝΝΙΝΑ 2019

«Η έγκριση της διδακτορικής διατριβής από το Τμήμα Ιατρικής του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα Ν. 5343/32, άρθρο 202, παράγραφος 2 (νομική κατοχύρωση του Ιατρικού Τμήματος) ».

Ημερομηνία αίτησης του κ. Λαδιά Πάρη: 28-5-2013

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 753^α/18-6-2013

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων

Γεωργίου Ιωάννης, Καθηγητής Ιατρικής Γενετικής και Υποβοηθούμενης Αναπαραγωγής της Μαιευτικής-Γυναικολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.

Μέλη

Τζαβάρας Θεόδωρος, Επίκουρος Καθηγητής Γενικής Βιολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.

Παπαλουκάς Κωνσταντίνος, Επίκουρος Καθηγητής Βιοπληροφορικής του Τμήματος Βιολογικών Εφαρμογών και Τεχνολογιών του Πανεπιστημίου Ιωαννίνων.

Ημερομηνία ορισμού θέματος: 29-11-2013

«Χαρακτηρισμός και Συσχέτιση δομών ανασυνδυασμού με επαναλαμβανόμενες αλληλουχίες του DNA σε καταστάσεις γενετικής αστάθειας με εργαλεία Βιοπληροφορικής»

ΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ 906^α/25-11-2019

Γεωργίου Ιωάννης	Καθηγητής Ιατρικής Γενετικής και Υποβοηθούμενης Αναπαραγωγής της Μαιευτικής-Γυναικολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων
Τζαβάρας Θεόδωρος	τ. Καθηγητής Γενικής Βιολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων
Μιχαηλίδης Θεολόγος	Αναπληρωτής Καθηγητής Μοριακής Γενετικής του Τμήματος Βιολογικών Εφαρμογών και Τεχνολογιών του Πανεπιστημίου Ιωαννίνων
Παπαλουκάς Κωνσταντίνος	Αναπληρωτής Καθηγητής Βιοπληροφορικής του Τμήματος Βιολογικών Εφαρμογών και Τεχνολογιών του Πανεπιστημίου Ιωαννίνων
Τσίπουρας Μάρκος,	Αναπληρωτής Καθηγητής Ιατρικής Πληροφορικής του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτικής Μακεδονίας
Τζάλλας Αλέξανδρος	Επίκουρος Καθηγητής Βιοϊατρικής Τεχνολογίας με έμφαση στην Επεξεργασία και Ανάλυση Βιοϊατρικών Δεδομένων του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων
Χατζημιχαήλ Ελευθερία	Επίκουρη Καθηγήτρια Αιματολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων

Έγκριση Διδακτορικής Διατριβής με βαθμό «ΑΡΙΣΤΑ» στις 9-12-2019

Ιωάννινα 11-12-2019

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ

Άννα Μπατιστάτου

Καθηγήτρια Παθολογικής Ανατομίας



Η Γραμματέας του Τμήματος

ΜΑΡΙΑ ΚΑΠΙΤΟΠΟΥΛΟΥ

....στους γονείς μου Παναγιώτη και Ευαγγελία

...στον αδερφό μου Αλέξανδρο

...στη Σοφία

ΠΡΟΛΟΓΟΣ

Το παρόν πόνημα ξεκίνησε κατοπιν εισήγησης του επιβλέποντα καθηγητή μου Γεωργίου Ι. ως συνέχεια της πτυχιακής μου εργασίας όταν ήμουν τελειόφοιτος της σχολής Β.Ε.Τ. Παράλληλα για μένα ήταν ένας φιλόδοξος στόχος για διεύρυνση του γνωστικού μου πεδίου στα πλαίσια των επιστημών της βιοπληροφορικής και της γενετικής. Κατά την πορεία της εκπόνησης της διατριβής εδραιώνεται ολοένα και περισσότερο η πεποίθησή μου ότι και οι 2 επιστήμες είναι πολύ σπουδαία ζητήματα που σημάδεψαν τη ζωή μου και ότι στο εξής είτε σε επίπεδο επιστημονικής διερεύνησης είτε σε πρακτικό επίπεδο θα αποτελούσαν μια δια βίου ενασχόληση για να οικειωθώ όσο γίνεται περισσότερους από τους πνευματικούς καρπούς που θα μου πρόσφεραν οι ποικίλες εφαρμογές της. Οφείλω να ομολογήσω πως στο ταξίδι μου αυτό στον κόσμο των βιολογικών επιστημών υπήρξαν δυσκολίες που καθεμιά τους ξεχωριστά, καθώς ξεπερνιόταν χάρη στη συνεχή καθοδήγηση του μέντορά και επιβλέποντα μου κ.Γεωργίου και στην καλόπιστη κριτική της συντονιστικής επιτροπής, ξεκλείδωνε και μια διαφορετική πόρτα που άφηνε ελεύθερο το βλέμμα μου για να ατενίσει ενδιαφέρουσες πτυχές αυτού του κόσμου.

Ήταν όμως αυτό το ταξίδι για μένα και μια μεγάλη πρόκληση στην οποία συνειδητοποίησα την ανάγκη να απαντήσω με μία υπόσχεση: ότι θα έθετα τις γνώσεις και τα συμπεράσματα από τη διηλεκτή έρευνά μου στο χώρο της βιολογίας ως επιστήμης του ανθρώπου στην υπηρεσία του ίδιου, που τόσο ταλανίζεται στην εποχή μας σε θέματα υγείας και επιδράσεων του περιβάλλοντος στη ζωή του.

Πέρα από την ευγνωμοσύνη μου στον επιβλέποντά μου και στις 2 συντονιστικές επιτροπές θα ήθελα να ευχαριστήσω και μερικούς ανθρώπους που χωρίς τη συνδρομή τους τίποτα από όλα αυτά δεν θα είχε πραγματοποιηθεί. Θα ήθελα να ευχαριστήσω θερμά τους γονείς και τον αδερφό μου για την αδιάκοπη στήριξη τους σε όλα μου τα όνειρα και σε όλες μου τις προσπάθειες, τους συναδέλφους και φίλους μου εντός και εκτός εργαστηρίου για την συμπαράσταση καθώς και τη Σοφία που με την ενθάρρυνση της διαδραμάτισε καθοριστικό ρόλο για την τελική διεκπεραίωση της.

“Η ανθρωπότητα εξελίχτηκε όχι γιατί υπήρξε σοβαρή, υπεύθυνη και προσεκτική, αλλά γιατί υπήρξε κεφάτη, επαναστατική και ανώριμη.”

Tom Robbins,

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1	6
ΑΝΑΣΥΝΔΥΑΣΜΟΣ	6
1.1 ΟΡΙΣΜΟΣ ΚΑΙ ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ.....	6
1.1.1 Η ΕΥΡΕΣΗ ΤΟΥ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΚΑΙ ΟΙ ΣΗΜΕΡΙΝΕΣ ΓΝΩΣΕΙΣ ΠΕΡΙ ΑΥΤΟΥ ...	8
1.1.2 ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΠΟΙΚΙΛΟΜΟΡΦΙΑ.....	13
1.1.3 ΓΕΝΕΤΙΚΗ ΣΥΝΔΕΣΗ ΚΑΙ ΕΚΤΙΜΗΣΗ.....	15
1.1.4 ΜΕΤΡΗΣΗ ΓΕΝΕΤΙΚΗΣ ΑΠΟΣΤΑΣΤΑΣΗΣ.....	16
1.1.5 Ο ΜΕΙΩΤΙΚΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ	18
1.2 ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΔΟΜΕΣ HOLLIDAY	22
1.2.1 ΤΡΟΠΟΙ ΔΗΜΙΟΥΡΓΙΑΣ HOLLIDAY JUNCTIONS	23
1.2.1.1 ΜΟΝΤΕΛΛΟ ΚΛΑΣΣΙΚΟΥ ΟΜΟΛΟΓΟΥ ΑΝΑΣΥΝΔΥΑΣΜΟΥ	23
1.2.1.2 ΤΟ ΜΟΝΤΕΛΛΟ ΔΙΚΛΩΝΗΣ ΘΡΑΥΣΗΣ ΕΝΟΣ ΜΟΡΙΟΥ (DSBS) ΚΑΙ	
ΕΠΙΔΙΟΡΘΩΣΗΣ ΤΟΥ.....	26
1.2.1.3 ΤΟ ΜΟΝΤΕΛΛΟ ΔΙΕΙΣΔΥΣΗΣ ΜΟΝΗΣ ΑΛΥΣΙΔΑΣ	29
1.2.2 ΔΥΝΑΜΙΚΗ ΚΑΙ ΔΟΜΗ ΤΩΝ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ (IR) ΠΟΥ	
ΣΧΗΜΑΤΙΖΟΥΝ ΗJ	31
1.2.2.1 ΑΡΧΙΚΑ ΜΟΝΤΕΛΛΑ ΤΩΝ ΔΟΜΩΝ HOLLIDAY	31
1.2.2.2 ΠΡΩΤΕΣ ΜΟΝΟ-ΚΡΥΣΤΑΛΛΙΚΕΣ ΔΟΜΕΣ ΤΩΝ HOLLIDAY JUNCTION	32
1.2.2.3 ΑΛΛΗΛΟΥΧΙΑΚΕΣ ΚΑΙ ΙΟΝΤΙΚΕΣ ΑΛΛΗΛΕΠΙΔΡΑΣΕΙΣ ΣΤΟ ΜΟΤΙΒΟ ΤΩΝ ΥΠΟ	
ΜΕΛΕΤΗ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ ΣΤΟΝ ΣΧΗΜΑΤΙΣΜΟ ΚΑΙ ΤΗΝ ΔΙΑΜΟΡΦΩΣΗ	
ΤΩΝ ΗJ.....	33
1.3 HOTSPOTS ΚΑΙ ΑΛΛΑ ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΩΣ ΣΗΜΕΙΑ ΑΝΑΣΥΝΔΥΑΣΜΟΥ	
ΚΑΤΑ ΜΗΚΟΣ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ.....	38
ΚΕΦΑΛΑΙΟ 2	43
ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ	43
2.1 ΕΙΣΑΓΩΓΗ ΣΤΑ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ (TES, TRANSPOSABLE ELEMENTS) ΤΟΥ	
ΓΟΝΙΔΙΩΜΑΤΟΣ	43
2.1.1 DNA ΤΡΑΝΣΠΟΖΟΝΙΑ	44
2.1.1.2.1 ΜΗΧΑΝΙΣΜΟΙ ΜΕΤΑΘΕΣΗΣ ΤΩΝ DNA ΤΡΑΝΣΠΟΖΟΝΙΩΝ.....	45
2.1.1.2.2 ΜΗΧΑΝΙΣΜΟΣ “COPY-AND-PASTE”	45
2.1.2 ΒΙΟΛΟΓΙΑ ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ	47

2.1.2.1 LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΑ.....	48
2.1.2.1.1 ΜΗΧΑΝΙΣΜΟΣ ΜΕΤΑΘΕΣΗΣ LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ ΜΕΣΑ ΣΤΑ VIRUS-LIKE ΣΩΜΑΤΙΔΙΑ ΣΤΟ ΚΥΤΤΑΡΟΠΛΑΣΜΑ	48
2.1.2.1.2 ΡΕΤΡΟΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΜΕ ΜΑΚΡΙΕΣ ΤΕΡΜΑΤΙΚΕΣ ΕΠΑΝΑΛΗΨΕΙΣ (LTRS) ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ.....	50
2.1.2.2 NON-LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΑ.....	51
2.1.2.2.1 LINES (LONG INTERSPERSED ELEMENTS).....	51
2.1.2.2.2 SINES (SHORT INTERSPERSED ELEMENTS).....	54
2.1.2.2.3 ΔΟΜΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ALU ΚΑΙ SVA ΤΗΣ ΥΠΕΡΟΙΚΟΓΕΝΕΙΑΣ ΤΩΝ SINES ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ	55
2.1.3 ΔΙΑΣΠΟΡΑ ΜΕΤΑΘΕΤΩΝ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΙΚΑΝΟΤΗΤΑ ΜΕΤΑΚΙΝΗΣΗΣ ΣΤΟ ΓΟΝΙΔΙΩΜΑ ΤΩΝ ΘΗΛΑΣΤΙΚΩΝ	56
2.1.3.1 ΣΤΟΙΧΕΙΑ ΣΥΣΧΕΤΙΣΗΣ HOLLIDAY JUNCTIONS ΜΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ	58
ΚΕΦΑΛΑΙΟ 3	60
ΓΕΝΕΤΙΚΗ ΑΣΤΑΘΕΙΑ	60
3.1 ΧΡΩΜΟΣΩΜΙΚΕΣ ΑΝΑΔΙΑΤΑΞΕΙΣ ΜΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΥΝΔΡΟΜΩΝ	60
3.1.1 CNVS (COPY NUMBER VARIATIONS) ΚΑΙ LCR (LOW COPY REPEATS).	62
3.1.2 CNVS ΚΑΙ ΡΕΤΡΟΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ.....	63
3.1.3 ΠΑΘΟΓΕΝΕΙΑ ΤΩΝ CNVS.....	65
3.1.4 ΑΝΑΔΥΟΜΕΝΟΙ ΑΡΙΘΜΟΙ ΜΜΣ	66
3.2 ΔΙΚΛΩΝΕΣ ΘΡΑΥΣΕΙΣ	67
3.2.1 ΑΙΤΙΑ ΠΡΟΚΛΗΣΗΣ ΔΙΚΛΩΝΩΝ ΘΡΑΥΣΕΩΝ (DSBS).....	69
3.2.1.1 ΕΝΔΟΓΕΝΕΙΣ ΠΑΡΑΓΟΝΤΕΣ	70
3.2.1.1 ΕΞΩΓΕΝΕΙΣ ΠΑΡΑΓΟΝΤΕΣ	74
3.2.2 ΓΕΝΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΤΩΝ ΤΡΙΩΝ ΕΙΔΩΝ ΚΑΡΚΙΝΟΥ ΠΟΥ ΜΕΛΕΤΗΘΗΚΑΝ	75
3.3 ΜΗΧΑΝΙΣΜΟΙ ΕΠΙΔΙΟΡΘΩΣΗΣ ΓΕΝΕΤΙΚΗΣ ΑΣΤΑΘΕΙΑΣ	77
3.3.1 ΜΗ ΟΜΟΛΟΓΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ / ΜΗ ΟΜΟΛΟΓΗ ΣΥΝΔΕΣΗ ΑΚΡΩΝ ΝΗΕJ (NON-HOMOLOGOUS END JOINING).....	77
3.3.2 ΟΜΟΛΟΓΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ HR (HOMOLOGOUS RECOMBINATION)	82
ΣΚΟΠΟΣ.....	86

ΚΕΦΑΛΑΙΟ 4	89
ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	89
4.1 ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ DECODE GENETICS ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΤΗ ΜΕ ΔΕΙΚΤΕΣ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΣΤΑ 2 ΦΥΛΑ	89
4.2 ΕΥΡΕΣΗ ΤΗΣ ΣΥΝΤΗΡΗΜΕΝΗΣ ΜΟΡΦΗΣ ΤΟΥ ΑΛΛΗΛΟΥΧΙΑΚΟΥ ΜΟΤΙΒΟΥ IRs ΚΑΙ ΤΩΝ ΜΟΡΦΩΝ ΠΟΥ ΑΥΤΟ ΕΜΦΑΝΙΖΕΤΑΙ ΩΣ HOLLIDAY JUNCTION	102
4.2.1 ΕΥΡΕΣΗ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΜΟΤΙΒΩΝ IRs-HJs ΣΕ ΟΛΟ ΤΟ ΜΗΚΟΣ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R	103
4.2.2 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ ΤΟΥ HARMAP PROJECT	106
4.2.3 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ REPEATMASKER ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.....	108
4.2.4 ΦΥΛΟΓΕΝΕΤΙΚΗ ΑΝΑΛΥΣΗ ΣΥΜΦΩΝΑ ΜΕ ΤΟ ΜΟΝΤΕΛΟ TAMURA-NEI	109
4.2.5 ΚΑΤΑΝΟΜΗ ΤΩΝ IRs ΠΟΥ ΣΧΗΜΑΤΙΖΟΥΝ HOLLIDAY JUNCTIONs ΣΕ TSS ΚΑΙ ΓΟΝΙΔΙΑ ΜΕ ΠΑΡΑΜΕΤΡΟ ΣΤΑΤΙΣΤΙΚΗΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ	109
4.2.6 ΒΙΟΠΛΗΡΟΦΟΡΙΚΑ ΕΡΓΑΛΕΙΑ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΕΞΟΥΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΑΛΛΗΛΟΥΧΙΩΝ IRs-HJs ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΕΝΤΟΣ ΤΩΝ ΟΡΙΩΝ ΣΥΝΔΡΟΜΩΝ ΜΙΚΡΟΕΛΛΕΙΨΩΝ ΚΑΙ ΜΙΚΡΟΔΙΠΛΑΣΙΑΣΜΩΝ	110
4.2.7 ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΕΓΓΥΤΗΤΑΣ ΣΕ ΓΟΝΙΔΙΑ ΚΑΙ TSS.....	114
4.3 ΕΥΡΕΣΗ ΟΛΩΝ ΤΩΝ ΠΑΡΑΛΛΑΓΩΝ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΜΟΤΙΒΩΝ ΤΟΥ PRDM9	114
4.3.1 ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΑΚΟΛΟΥΘΙΩΝ ΜΕΣΩ ΤΗΣ ΚΟΝΣΟΛΑΣ R ΚΑΙ ΑΝΑΦΟΡΑ ΤΩΝ ΑΠΑΡΑΙΤΗΤΩΝ ΠΑΚΕΤΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ.....	115
4.3.2 ΣΥΛΛΟΓΗ ΑΡΧΕΙΩΝ ΚΑΡΚΙΝΟΥ ΚΑΙ ΧΡΗΣΗ ΤΟΥΣ ΣΤΗΝ R.....	118
ΚΕΦΑΛΑΙΟ 5	123
ΑΠΟΤΕΛΕΣΜΑΤΑ.....	123
5.1 ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΤΗ ΜΕ ΔΕΙΚΤΕΣ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΣΤΑ 2 ΦΥΛΑ	123
5.2 ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΞΥ ΦΥΣΙΚΗΣ ΚΑΙ ΓΕΝΕΤΙΚΗΣ ΑΠΟΣΤΑΣΗΣ ΤΩΝ ΓΕΝΕΤΙΚΩΝ ΔΕΙΚΤΩΝ ΣΤΑ 2 ΦΥΛΑ	125
5.3 ΠΛΗΡΗΣ ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ IRs-HJs ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ	136
5.3.1 ΣΥΣΧΕΤΙΣΗ IRs-HJs ΜΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ	137
5.3.2 ΦΥΛΟΓΕΝΕΤΙΚΟ ΚΑΙ ΕΞΕΛΙΚΤΙΚΟ ΥΠΟΒΑΘΡΟ IRs-HJs ΠΟΥ ΕΝΤΟΠΙΖΟΝΤΑΙ ΣΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ	139

5.3.3 ΚΑΤΑΝΟΜΗ ΤΩΝ IRs-HJs ΚΑΙ ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΗΣ ΜΕΣΗΣ ΑΠΟΣΤΑΣΗΣ ΜΕΤΑΞΥ ΤΟΥΣ ΣΤΑ ΑΝΘΡΩΠΙΝΑ ΧΡΩΜΟΣΩΜΑΤΑ	141
5.3.4 ΑΝΑΛΥΣΗ ΕΓΓΥΤΗΤΑΣ ΤΩΝ IRs-HJs ΣΕ ΘΕΣΕΙΣ ΕΝΑΡΞΗΣ ΤΗΣ ΜΕΤΑΓΡΑΦΗΣ ΓΟΝΙΔΙΩΝ	144
5.4 ΣΥΣΧΕΤΙΣΗ ΤΩΝ IRs-HJs ΑΛΛΗΛΟΥΧΙΩΝ ΜΕ ΣΥΝΔΡΟΜΑ ΜΙΚΡΟΕΛΛΕΙΨΕΩΝ ΚΑΙ ΜΙΚΡΟΔΙΠΛΑΣΙΑΣΜΩΝ	145
5.5 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΠΑΡΑΛΛΑΓΩΝ ΤΟΥ PRDM9 ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΟΥΣ ΚΑΡΚΙΝΟΥΣ ΤΟΥ ΜΑΣΤΟΥ, ΤΗΣ ΚΕΦΑΛΗΣ ΚΑΙ ΤΡΑΧΗΛΟΥ ΚΑΙ ΤΟΥ ΑΔΕΝΟΚΑΡΚΙΝΩΜΑΤΟΣ ΤΟΥ ΠΝΕΥΜΟΝΑ	154
ΚΕΦΑΛΑΙΟ 6	158
ΣΥΖΗΤΗΣΗ.....	158
6.1 ΜΕΓΑΛΥΤΕΡΗ ΣΥΣΣΩΡΕΥΣΗ HOTSPOTS ΑΝΑΥΝΔΥΑΣΜΟΥ ΟΔΗΓΕΙ ΣΕ ΛΙΓΟΤΕΡΑ ΣΥΜΒΑΝΤΑ ΑΝΑΣΥΝΔΥΑΣΜΟΥ. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΠΑΡΑΔΕΙΓΜΑ ΤΩΝ 2 ΦΥΛΩΝ.....	158
6.2 ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ DNA ΩΣ ΦΟΡΕΙΣ ΑΛΛΗΛΟΥΧΙΩΝ ΜΕ ΑΝΑΣΥΝΔΥΑΣΤΙΚΗ ΔΡΑΣΗ.....	160
6.3 ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ ΠΟΥ ΠΕΡΙΓΡΑΦΗΚΑΝ ΠΡΟΗΓΟΥΜΕΝΩΣ ΟΤΙ ΣΧΗΜΑΤΙΖΟΥΝ ΔΟΜΕΣ HOLLIDAY (IRs-HJs) ΩΣ ΠΙΘΑΝΑ HOTSPOTS ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΣΥΝΔΡΟΜΩΝ ΜΕ ΑΝΑΠΤΥΞΙΑΚΕΣ ΔΙΑΤΑΡΑΧΕΣ ΚΑΙ ΝΟΗΤΙΚΗ ΣΤΕΡΗΣΗ.	165
6.4 ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΠΟΥ ΑΝΑΓΝΩΡΙΖΟΝΤΑΙ ΑΠΟ ΤΗΝ PRDM9 ΩΣ ΠΙΘΑΝΑ HOTSPOTS ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΔΙΚΛΩΝΩΝ ΘΡΑΥΣΕΩΝ (ΚΑΡΚΙΝΟΓΕΝΕΣΗ).....	169
ΣΥΜΠΕΡΑΣΜΑΤΑ	171
ΠΕΡΙΛΗΨΗ.....	173
ABSTRACT.....	176
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	179
ΒΙΒΛΙΟΓΡΑΦΙΑ	181

ΚΕΦΑΛΑΙΟ 1

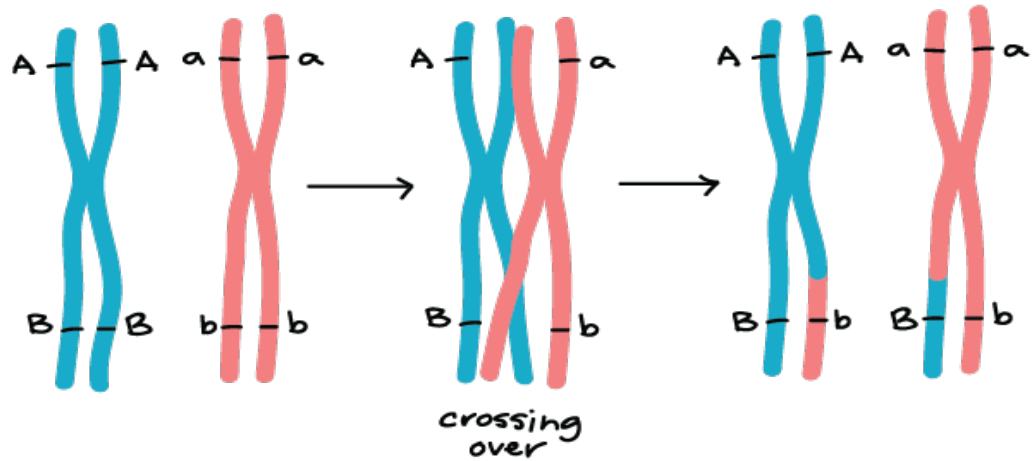
ΑΝΑΣΥΝΔΥΑΣΜΟΣ

1.1 ΟΡΙΣΜΟΣ ΚΑΙ ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

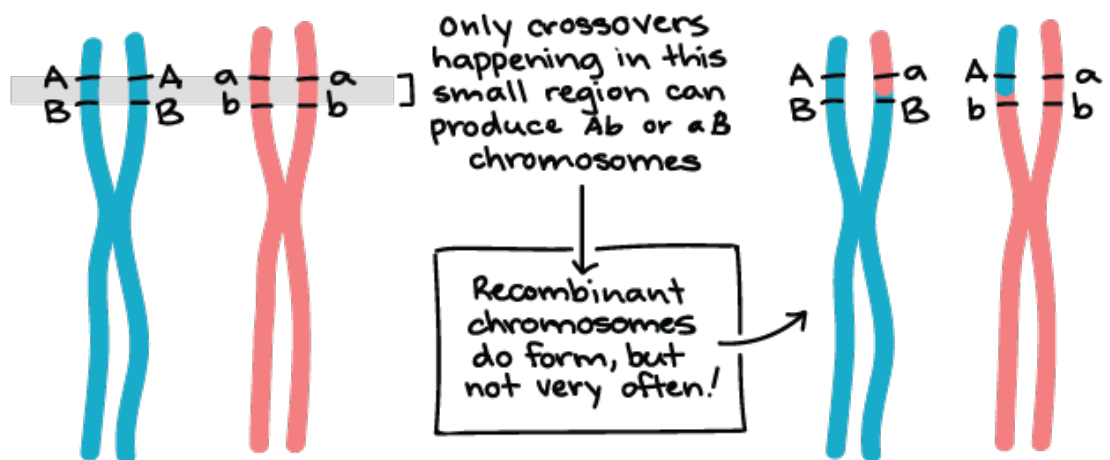
Ο ανασυνδυασμός γονιδίων, ή γενετικός ανασυνδυασμός, είναι η ανταλλαγή γενετικού υλικού (γονιδίων) μεταξύ των ομόλογων χρωμοσωμάτων. Κατά τον ανασυνδυασμό γονιδίων, τμήμα ενός πατρικού χρωμοσώματος αλλάζει θέση με το αντίστοιχο τμήμα του μητρικού χρωμοσώματος. Ο ανασυνδυασμός γονιδίων πραγματοποιείται μέσω μιας βιολογικής πορείας που ονομάζεται διασκελισμός και συμβαίνει στο στάδιο της ζυγοταινίας στην μειωτική διαίρεση. Συμβάλλει στην αύξηση της ποικιλομορφίας στην φύση και καθιστά αδύνατη την περίπτωση δύο άτομα του ίδιου είδους να είναι απόλυτα όμοια. Πιο συγκεκριμένα, κατά τη διάρκεια αυτής της διεργασίας, ανταλλάσσονται τμήματα ομόλογων χρωμοσωμάτων, καθώς η διπλή έλικα του DNA κόβεται και μετά επανασυνδέεται τόσο σε μία μητρική όσο και σε μία ομόλογη πατρική χρωματίδα. Με αυτόν τον τρόπο γίνεται αμοιβαία ανταλλαγή κλασμάτων DNA μεταξύ των 2 μη αδελφών χρωματίδων(1).

Ένας από τους πρώτους που επιχείρησαν να εξγήσουν τις παρεκκλίσεις από τις κλασσικές αναλογίες της Μεντελικής κληρονομικότητας ήταν ο Morgan σε άρθρο του το 1911(2). Στο συγκεκριμένο άρθρο πρότεινε πως οι παράγοντες που προσδίδουν συγκεκριμένους φαινοτύπους (γνωστοί σήμερα ως γονίδια) βρίσκονται πάνω στα χρωμοσώματα σε σειρά και κατά την διαδικασία της μείωσης και τον ανασυνδυασμό μεταξύ των ομόλογων χρωμοσωμάτων, τα γονίδια που είναι πολύ κοντά μεταξύ τους είναι πιο πιθανό να παραμείνουν στην ίδια πλευρά τηςσχάσης που προκαλείται από τον ανασυνδυασμό και να κληρονομηθούν μαζί, ενώ γονίδια που απέχουν πολύ μεταξύ τους έχουν αρκετές πιθανότητες να βρεθούν στην αντίθετη μεριά τηςσχάσης, δηλαδή το ένα

γονίδιο να παραμείνει στο αρχικό χρωμόσωμα και το άλλο να μεταπηδήσει στο ομόλογο χρωμόσωμα. Αυτό εξηγεί γιατί παρατηρείται σύνδεση μεταξύ μερικών χαρακτήρων ενώ δεν μπορεί να αποδειχθεί η σύνδεση μεταξύ άλλων.



Εικόνα 1.1: Ανασυνδυασμός μεταξύ δύο απομακρυσμένων γονιδίων. Το γονίδιο B αλλάζει θέση με το γονίδιο b δημιουργώντας δύο νέους γονότυπους (Ab, aB) (3).

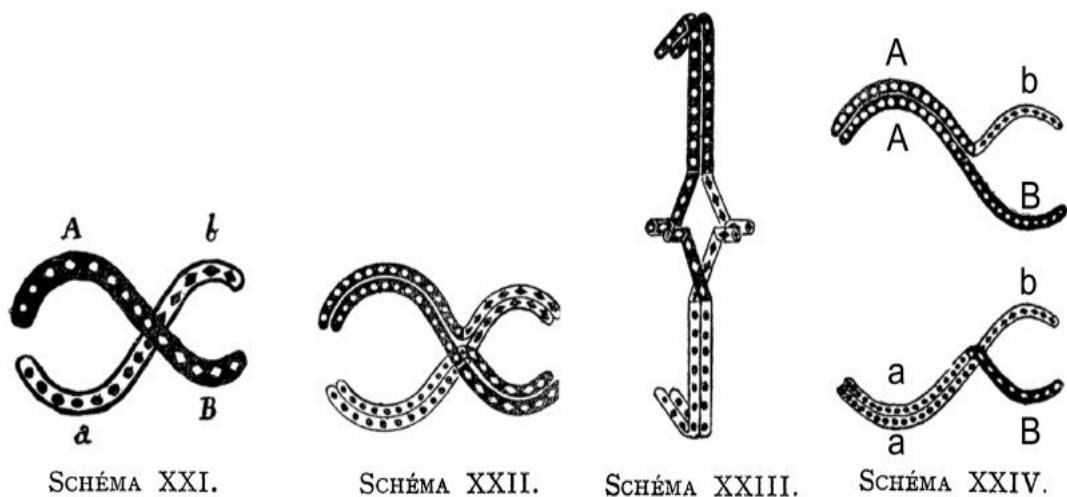


Εικόνα 1.2: Ανασυνδυασμός μεταξύ δύο κοντινών γονιδίων. Για να δημιουργηθούν οι γονότυποι Ab, aB θα πρέπει ο ανασυνδυασμός να εντοπίζεται στην μικρή γκρι περιοχή. Αρκετά σπανιότερος από τον προηγούμενο (3).

1.1.1 Η ΕΥΡΕΣΗ ΤΟΥ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΚΑΙ ΟΙ ΣΗΜΕΡΙΝΕΣ ΓΝΩΣΕΙΣ

ΠΕΡΙ ΑΥΤΟΥ

Για την παραπάνω υπόθεσή του, ο Morgan βασίστηκε στο έργο του Frans Alfons Janssens, ο οποίος με άρθρο του το 1909 περιέγραψε την δομή των σχηματιζόμενων χιασμάτων κατά τη μείωση(4). Συνοπτικά, μελέτησε διαφορετικές διατάξεις χρωμοσωμάτων από κύτταρα του είδους *B. attenuatus* κατά τη μείωση I και II. Ειδικότερα, στάθηκε στο στάδιο της ανάφασης I όπου και παρατήρησε τον σχηματισμό των χιασμάτων, δηλαδή τις συνδέσεις μεταξύ ομόλογων περιοχών που τείνουν να παραμείνουν στο κέντρο της μιτωτικής ατράκτου, σε αντίθεση με τις κεντρομερικές περιοχές που έχουν διαχωριστεί λόγω της μετακίνησής τους προς τους πόλους.



Εικόνα 1.3: Schema XXI: Χίασμα μεταξύ δύο ομόλογων χρωμοσωμάτων AB, ab. Schema XXII: Τέσσερις χρωματίδες (δύο ζεύγη αδελφών χρωματίδων) και το σημείο του χιάσματος που πραγματοποιείται αμοιβαία ανταλλαγή τμημάτων των ομόλογων χρωμοσωμάτων. Schema XXIII: Ανάφαση I. Οι κεντρομερικές περιοχές μετακινούνται στους πόλους ενώ η περιοχή του χιάσματος παραμένει στο κέντρο. Schema XXIV: Η ανταλλαγή οδηγεί στην δημιουργία δύο πατρικών χρωματίδων AB, ab και δύο ανασυνδυασμένων aB, Ab(4).

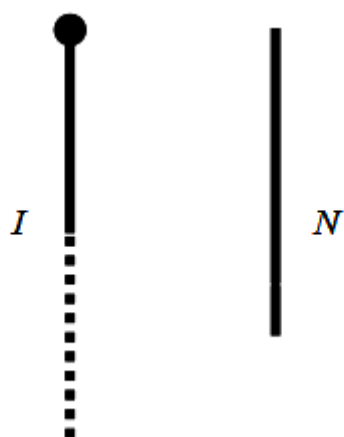
Στην παραπάνω δημοσίευση ο Janssens καταλήγει σε μερικά συμπεράσματα. Αναφέρει ότι στις περιοχές των χιασμάτων, τα χρωμοσώματα διεισδύουν το ένα στο άλλο και προκαλούνται θραύσεις και συντήξεις των χρωματίδων. Οι αλληλεπιδράσεις

αυτές μπορούν να αλλάξουν την οργάνωση των χρωμοσωμάτων δημιουργώντας νέες αλληλουχίες, διαφορετικές από τις πατρικές. Έτσι, έθεσε τα θεμέλια για την έρευνα του γενετικού ανασυνδυασμού.

Την καθοριστική απόδειξη της ύπαρξης ανασυνδυασμού μεταξύ των χρωμοσωμάτων, έδωσαν το 1931 οι Harriet Creighton και Barbara McClintock όπου σχεδιάζοντας ένα ευφάνταστο πείραμα κατάφεραν να συσχετίσουν δομικές αντάλλαγες που συνέβαιναν στα χρωμοσώματα με γενετικές ανταλλαγές. Για να το αποδείξουν αυτό, έπρεπε να πραγματοποιήσουν τα ακόλουθα βήματα:

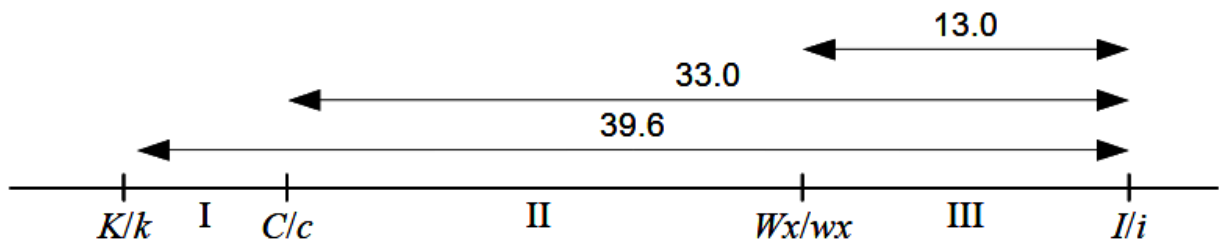
1. Να ανακαλύψουν ζεύγη ομόλογων χρωμοσωμάτων τα οποία θα εμφάνιζαν κάποιου είδους ετερομορφία, δηλαδή θα υπήρχε κάποια δομική διαφορά μεταξύ τους που θα μπορούσε να λειτουργήσει ως δείκτης για τον εντοπισμό τυχόν ανταλλαγής γενετικού υλικού μεταξύ των ομόλογων χρωμοσωμάτων.
2. Να εντοπίσουν επιχιασμούς στα χρωμοσώματα αυτά ώστε να αποδείξουν ότι πράγματι συμβαίνει ανταλλαγή των χρωματίδων ή μέρους αυτών.
3. Με την ύπαρξη των παραπάνω δομικών ανταλλαγών, θα πρέπει να πραγματοποιήσουν επιχιασμούς σε χρωμοσώματα που θα έχουν και δομικούς και γενετικούς δείκτες και να προσδιορίσουν τυχόν σχέση συσχέτισης μεταξύ των δύο ειδών ανταλλαγών.

Για τα πειράματά τους χρησιμοποίησαν το καλαμπόκι (*Zea mays*) και έπειτα από αρκετές διασταυρώσεις κατάφεραν να απομονώσουν ένα στέλεχος καλαμποκιού στο οποίο το ένα από τα δύο χρωμοσώματα του ζεύγους 9 (συνολικά το καλαμπόκι έχει 10 ζεύγη χρωμοσωμάτων) είχε μια διακριτή δυσμορφία στο ένα άκρο του ενώ το άλλο άκρο ήταν μεγαλύτερο σε μήκος λόγω προηγούμενης μετατόπισης τμήματος του χρωμοσώματος 8 σε αυτό(5).



Εικόνα 1.4: Σχηματική απεικόνιση του ζεύγους 9. Το αφύσικο χρωμόσωμα (I) με την στρογγυλή κεφαλή στο ένα άκρο και την εκτεταμένη ουρά στο άλλο και το φυσιολογικό χρωμόσωμα (N) (6).

Για να αποκτήσουν άτομα με ετερόμορφα ζεύγη του χρωμοσώματος 9, διασταύρωσαν άτομα που είχαν μόνο αφύσικα χρωμοσώματα στο ζεύγος αυτό με άτομα που είχαν φυσιολογικά χρωμοσώματα. Ακολούθως, αφού είχαν ήδη αποδείξει το πρώτο από τα τρία βήματα που αναφέρθηκαν παραπάνω, διασταύρωσαν τα ετερόμορφα άτομα με σκοπό να εντοπίσουν κάποιον δομικό ανασυνδυασμό. Τα αποτελέσματά τους επιβεβαίωσαν και την δεύτερη υπόθεσή τους καθώς εμφανίστηκε ανασυνδυασμός στο 39.6% των απογόνων. Τέλος, έπρεπε να διαπιστώσουν εάν υπάρχει σχέση μεταξύ της δομικής ανταλλαγής που ανακάλυψαν με τον γενετικό ανασυνδυασμό όπως ήταν γνωστός μέχρι τότε από την ερευνητική ομάδα του Morgan. Για να το κάνουν αυτό, χαρτογράφησαν δύο γονίδια στο χρωμόσωμα 9, ένα που δίνει χρωματιστή αλευρώνη (ιστός αποτελούμενος από την εξωτερική στοιβάδα κυττάρων του ενδοσπερμίου) και συμβολίζεται με C ή άχρωμη αλευρώνη που συμβολίζεται με c και ένα που δίνει κέρινο (wx) ή αμυλούχο (Wx) ενδοσπέρμιο.



Εικόνα 1.5: Ο χάρτης των δύο γονιδίων. K/k προσδιορίζει την ύπαρξη ή μη της αφύσικης κεφαλής ενώ I/i την ύπαρξη ή μη της μετατόπισης (7).

Οι δύο πιθανότητες που υπάρχουν είναι:

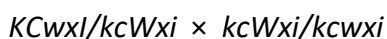
1. Ο δομικός και ο γενετικός ανασυνδυασμός να πραγματοποιούνται με την ίδια διαδικασία.
2. Ο δομικός και ο γενετικός ανασυνδυασμός να είναι ανεξάρτητοι μεταξύ τους και να ακολουθούν διαφορετικά στάδια κατά την πραγματοποίησή τους.

Συνεπώς:

1. Ένας δομικός ανασυνδυασμός στην περιοχή I ή III όπως αυτές αναγράφονται στην εικόνα 6, θα έδινε ανασυνδυασμένους γαμέτες όσον αφορά τις δυσμορφίες του χρωμοσώματος αλλά θα είχαν τα ίδια γονίδια με τους προγόνους. Ανασυνδυασμός στην περιοχή II ωστόσο, θα έδινε γαμέτες ανασυνδυασμένους τόσο για τα γονίδια όσο και για τις δυσμορφίες. Εάν ισχύει αυτή η υπόθεση και λαμβάνοντας υπόψιν την πιθανότητα για διπλό ανασυνδυασμό: (i) Απόγονοι ανασυνδυασμένοι και για τα δύο γονίδια, θα πρέπει να είναι στην πλειοψηφία τους ανασυνδυασμένοι και για τις δυσμορφίες. (ii) Οι μισοί από τους απογόνους που θα είναι ανασυνδυασμένοι για τις δυσμορφίες θα πρέπει να είναι ανασυνδυασμένοι και για τα γονίδια. (iii) Ελάχιστοι έως και καθόλου από τους γαμέτες ανασυνδυασμένους για τα γονίδια θα εμφανίζουν φυσιολογική δομή του χρωμοσώματος.
2. Στην περίπτωση που οι δύο διαδικασίες δεν συσχετίζονται, οι συχνότητες ανασυνδυασμού μπορούν να υπολογιστούν με χρήση του γενετικού χάρτη και

των πιθανοτήτων. (i) 39% απογόνων ανασυνδυασμένοι για τα γονίδια και για τις δυσμορφίες. (ii) 20% απογόνων ανασυνδυασμένοι για τις δυσμορφίες θα είναι και ανασυνδυασμένοι για τα γονίδια. (iii) 61% απογόνων θα είναι ανασυνδυασμένοι για τα γονίδια αλλά όχι για τις δυσμορφίες, δηλαδή θα έχουν φυσιολογική δομή χρωμοσωμάτων(7).

Η καθοριστική διασταύρωση που πραγματοποιήθηκε ήταν η εξής:



Πρόεκυψαν 27 απόγονοι, οι φαινότυποι των οποίων παρουσιάζονται στον παρακάτω πίνακα:

		heteromorphologies			
		par	rec	?	
genetic markers	par	8	2	0	10
	rec	0	3	1	4
	?	6	6	1	13
		14	11	2	27

Εικόνα 1.6: Οι φαινότυποι των 27 απογόνων. Heteromorphologies είναι οι χρωμοσωμικές δυσμορφίες ενώ genetic markers τα υπό εξέταση γονίδια. Par είναι οι πατρικοί ενώ rec οι ανασυνδυασμένοι. Με ? είναι αυτοί που δεν ανήκουν ξεκάθαρα σε κάποια κατηγορία (7).

Τα αποτελέσματα μοιάζουν εκπληκτικά με τις προβλέψεις της πρώτης υπόθεσης καθώς:

(i) Όλοι (3/3) οι ανασυνδυασμένοι για τα γονίδια απόγονοι είναι ανασυνδυασμένοι και για τις δυσμορφίες.

(ii) Οι μισοί (3/5) ανασυνδυασμένοι για τις δυσμορφίες είναι ανασυνδυασμένοι και για τα γονίδια.

(iii) Κανένας ανασυνδυασμένος για τα γονίδια απόγονος δεν είχε φυσιολογική δομή χρωμοσωμάτων.

Έτσι, η Harriet Creighton και η Barbara McClintock συμπέραναν ότι δύο ομόλογα χρωμοσώματα που παρουσιάζουν ετερομορφία σε μια περιοχή, ανταλλάσσουν τμήματά τους καθώς και γονίδια που εντοπίζονται στα τμήματα αυτά(8).

Σήμερα, είναι γνωστό πως ο ομόλογος ανασυνδυασμός δεν σχετίζεται μόνο με την μείωση, αλλά παρατηρείται και μιτωτικός ομόλογος ανασυνδυασμός, κυρίως σε διεργασίες όπως η επιδιόρθωση δίκλωνων θραύσεων του DNA, η διάσωση της διχάλας αντιγραφής του DNA όταν παρατηρείται μονόκλωνη τομή (nick) στην αλυσίδα εκμαγείο και η συντήρηση των τελομερών. Ο μειωτικός ανασυνδυασμός συμβαίνει από 100 έως και 1000 φορές συχνότερα από ότι ο μιτωτικός και η κύρια διαφορά μεταξύ των δύο είναι ότι στον πρώτο παρατηρούνται επιχιασμοί οι οποίοι είναι απαραίτητοι για τον διαχωρισμό των χρωμοσωμάτων, ενώ στον δεύτερο οι επιχιασμοί καταστέλλονται από DNA ελικάσες στην πλειοψηφία των περιπτώσεων (4, 5(9)).

1.1.2 ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΠΟΙΚΙΛΟΜΟΡΦΙΑ

Η γονιδιωματική ποικιλομορφία εξαρτάται άμεσα από τη διαδικασία του γονιδιακού ανασυνδυασμού η οποία συμβάλλει στην παραγωγή απλότυπων. Οι απλότυποι αποτελούν συνδυασμό από αλληλικές οντότητες πολυμορφισμών ενός μορίου DNA. Ειδικότερα είναι νέοι συνδυασμοί γονιδίων που συνδέονται μεταξύ τους και προκύπτουν από τον επιχιασμό γενετικών τόπων.

Το αν δύο αυτοσωμικές θέσεις θα ανασυνδυαστούν είναι αποτέλεσμα και της σχετικής τους απόστασης. Ο ανασυνδυασμός συνήθως οδηγεί στον διαχωρισμό

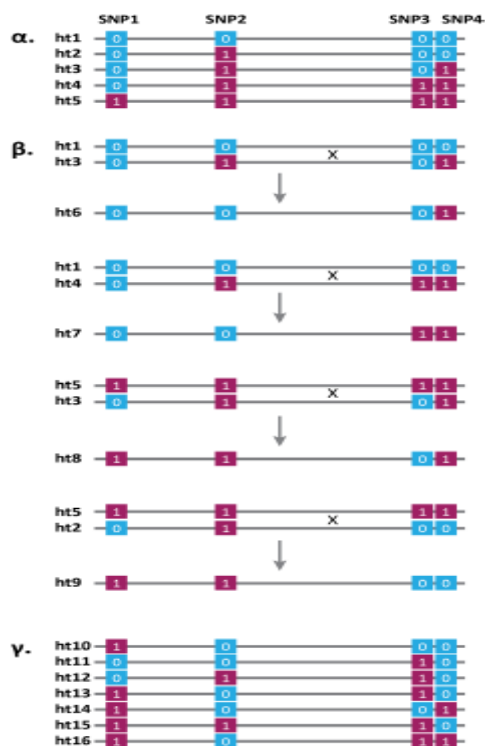
γενετικών τόπων με μεγάλη απόσταση μεταξύ τους. Η φυσική απόσταση των γενετικών τόπων δεν είναι, παρ' όλα αυτά, γραμμικά συνδεδεμένη με τη γενετική τους απόσταση, δηλαδή με την πιθανότητα να ανασυνδυαστούν. Ο μειωμένος ανασυνδυασμός οδηγεί στο φαινόμενο της συχνότερης ταυτόχρονης κληρονόμησης γενετικών τόπων με μικρή απόσταση.

Όταν τα αλληλόμορφα έχουν μικρή απόσταση σε ένα χρωμόσωμα και άρα μεγάλη πιθανότητα ταυτόχρονης κληρονόμησης κατά τη διαίρεση υπάρχει γενετική σύνδεση (linkage). Αυτή μας δείχνει εάν είναι εφικτό να προβλέψουμε την ύπαρξη συγκεκριμένων αλληλομόρφων σε έναν απλότυπο. Εάν γνωρίζω την ύπαρξη ενός αλληλομόρφου σε έναν απλότυπο και μπορώ να προβλέψω την μη τυχαία ύπαρξη ενός δεύτερου στον ίδιο απλότυπο, τότε τα δύο αλληλόμορφα βρίσκονται σε *ανισορροπία σύνδεσης*. Αντιθέτως, εάν δεν μπορώ να προβλέψω την ύπαρξη ενός δεύτερου αλληλομόρφου, τότε αυτά βρίσκονται σε *ισορροπία σύνδεσης*. Εφόσον μελετώνται SNPs, τότε θα πρέπει να ληφθεί υπόψιν ότι τα περισσότερα βρίσκονται σε ανισορροπία σύνδεσης διότι οι αλγόριθμοι ανάλυσης της γενετικής σύνδεσης έχουν ως δεδομένο ότι οι δείκτες που χρησιμοποιούνται βρίσκονται σε ισορροπία. Ένας τρόπος είναι να αποκλείσουμε τα SNPs που βρίσκονται σε μεγάλη ανισορροπία ώστε το σετ δεδομένων υπό εξέταση να αποτελείται μόνο από SNPs που βρίσκονται κοντά στην ισορροπία. Εάν η μέτρηση δύο κοντινών SNPs δείξει ότι βρίσκονται σε μεγάλη ανισορροπία, τότε ένας από τους δύο δείκτες απορρίπτεται. (10, 11).

Διαδεδομένη είναι η άποψη περί ύπαρξης περιοχών με μεγάλη πιθανότητα ανασυνδυασμού διεσπαρμένων στο γονιδίωμα του ανθρώπου μεταξύ απλοτύπων, δηλαδή περιοχών όπου κυριαρχεί ανισορροπία σύνδεσης (LD / Haplotypeblocks) (7). Ωστόσο, η διαφορά μεταξύ των πιθανών προτύπων απλοτυπικών συμπλόκων ενδέχεται να μην οφείλεται μόνο στην επίδραση του ανασυνδυασμού αλλά και στην γενετική απόκλιση ή τη διαφοροποιημένη διαστρωμάτωση του πληθυσμού.

Αξιοσημείωτες διαφορές ως προς τη διαφορά των συμπλόκων αυτών παρουσιάζονται κατά την συνεξέταση ξεχωριστών πληθυσμών. Έτσι, οι πληθυσμοί της Ευρώπης και της Ασίας φαίνονται να έχουν μεγαλύτερα απλοτυπικά σύμπλοκα σε

σύγκριση με αυτούς της Αφρικής καθώς οι πρώτοι είναι νεότεροι ως προς τη γενετική τους εξέλιξη(12, 13).



Εικόνα 1.7. Γενετικός ανασυνδυασμός και απλότυπα. α.) Δεν πραγματοποιείται ανασυνδυασμός, καθώς 4 σημειακοί νουκλεοτιδικοί πολυμορφισμοί (SNPs) δίδουν 5 απλότυπους. β.) Ένα σημείο ανασυνδυασμού δίνει 4 επιπλέον ανασυνδυασμούς. γ.) Ελεύθερος ανασυνδυασμός. Δίνονται 16 απλότυποι (14).

1.1.3 ΓΕΝΕΤΙΚΗ ΣΥΝΔΕΣΗ ΚΑΙ ΕΚΤΙΜΗΣΗ

Έως σήμερα έχουν προταθεί πολλοί διαφορετικοί υπολογισμοί για την εκτίμηση της ισχύος της LD, οι περισσότεροι από τους οποίους αφορούν την ισχύ του βαθμού συσχέτισης μεταξύ των ζευγών διαλληλικών τόπων. Δύο τέτοιοι σημαντικοί όροι είναι οι LD και r^2 (μερικές φορές συμβολίζονται και ως $|D'|$ και Δ^2 , αντίστοιχα), οι τιμές των οποίων κυμαίνονται από 0 (δεν υφίσταται ανισορροπία σύνδεσης) έως 1 (απόλυτη ανισορροπία σύνδεσης)(15). Η ερμηνεία τους, ωστόσο, είναι ελαφρώς διαφορετική. Παίρνουν τιμές μεταξύ παντελούς απουσίας ανισορροπίας σύνδεσης, δηλαδή 0, και

πλήρους ανισορροπίας σύνδεσης, δηλαδή 1. Αν 2 ή 3 πιθανοί απλότυποι είναι παρόντες, το LD θα έχει τιμή 1 ενώ αν και οι 4 πιθανοί απλότυποι υπάρχουν, θα λάβει τιμή μικρότερη της μονάδας, υποδεικνύοντας ύπαρξη ανασυνδυασμού ανάμεσα σε 2 γενετικούς τόπους. Η σχέση, ωστόσο μεταξύ 2 τιμών του LD που είναι μικρότερες της μονάδας δεν ερμηνεύεται αυτόματα. Το r^2 εκφράζει τη στατιστική συσχέτιση ανάμεσα σε 2 τόπους και γίνεται ίσο με την μονάδα στην περίπτωση που παρουσιάζονται ακριβώς 2 απλότυποι. Το μέγεθος του δείγματος που χρειάζεται για την ανακάλυψη συσχέτισης μεταξύ γενετικών τόπων και μονονουκλεοτιδικών πολυμορφισμών είναι αντίστροφο του r^2 . Αυτός είναι άλλωστε και ο λόγος που το r^2 προτιμάται κατά την χαρτογράφηση(16).

Έτσι, αν μία πάθηση συνδέεται στην πραγματικότητα με έναν SNP αλλά έχει αποδοθεί λανθασμένα στον γειτονικό του SNP, το μόνο που χρειάζεται για να διαπιστωθεί η συσχέτιση για τον τελευταίο, είναι να μεγαλώσουμε το δείγμα μας κατά $1/r^2$. Για την εκτίμηση όμως, της LD σε μία περιοχή με διαφορετικές θέσεις πολυμορφισμού απαιτείται ιδιαίτερη πολυπλοκότητα που αυξάνεται σε περίπτωση που η LD παρουσιάζει μεγάλη διαφοροποίηση ανάμεσα σε διαφορετικούς τόπους ή ακόμα και πληθυσμούς ή γενικότερα απέχει πολύ από την αρχική μας πρόβλεψη.

Μία ακόμα βοηθητική εκτίμηση μπορεί να γίνει και στο ρ , το οποίο χρησιμοποιείται όταν παρατηρούμε ανισορροπία σύνδεσης στα δεδομένα ενός μοντέλου ενός πληθυσμού και εκφράζει το πόσο εκτεταμένος πρέπει να είναι ο ανασυνδυασμός για να δικαιολογηθεί αυτή(17).

1.1.4 ΜΕΤΡΗΣΗ ΓΕΝΕΤΙΚΗΣ ΑΠΟΣΤΑΣΗΣ

Υπάρχουν διάφορα μέτρα που ποσοτικοποιούν την απόσταση μεταξύ δύο συνδεδεμένων τόπων. Το πρώτο και το πιο απλό είναι απλά ο αριθμός των νουκλεοτιδίων που χωρίζουν τους γενετικούς τόπους. Συνήθως, οι μονάδες μέτρησης

εδώ εκφράζονται σε ζεύγη βάσεων (ή bp), χιλιάδες ζευγών βάσης (kilobases ή kb) ή εκατομμύρια ζευγών βάσης (megabases ή Mb). Για να μετρηθούν τυχόν προσθήκες και διαγραφές γονοδιωματικών περιοχών, ο αριθμός των νουκλεοτιδίων που διαχωρίζουν τους τόπους βασίζεται στην αλληλουχία ανθρώπινου γονιδιώματος αναφοράς. Η δεύτερη μονάδα μέτρησης, είναι το κλάσμα ανασυνδυασμού θ και είναι μία δεσμευμένη πιθανότητα δύο γονιδίων που ισούται με την πιθανότητα ένας γαμέτης να περιέχει ένα αλληλόμορφο από το αντίθετο χρωμόσωμα, δεδομένου ότι περιέχει ένα αλληλόμορφο από το χρωμόσωμα που μας ενδιαφέρει. Οι τιμές που παίρνει είναι από 0, που σημαίνει ότι τα δύο γονίδια είναι απολύτως συνδεδεμένα και δεν παρατηρείται ξεχωριστή κληρονομία αυτών, έως 0,5 που είναι και το ανώτερο όριο και υποδεικνύει ότι τα δύο γονίδια είτε βρίσκονται πολύ μακριά στο ίδιο χρωμόσωμα είτε σε διαφορετικά χρωμοσώματα. Γονίδια με αυτήν την τιμή θ ακολουθούν τους νόμους κληρονομικότητας του Mendel. Το τρίτο μέτρο είναι το centimorgan ή cM. Ένα centimorgan ορίζεται ως η φυσική απόσταση που αντιστοιχεί σε μια τιμή θ που ισούται με το 0,01. Στη γενετική, μια μονάδα centimorgan (cM) ή μονάδα χάρτη (m.u.) είναι μια μονάδα μέτρησης της γενετικής σύνδεσης. Ορίζεται ως η απόσταση μεταξύ χρωμοσωμικών θέσεων (που ονομάζονται επίσης τόποι ή δείκτες) για τους οποίους ο αναμενόμενος μέσος όρος παρεμβαλλόμενων χρωμοσωμικών διασταυρώσεων σε μία μόνο γενιά είναι 0,01. Συχνά χρησιμοποιείται για να υπολογίσει την απόσταση κατά μήκος ενός χρωμοσώματος. Ωστόσο, δεν είναι μια πραγματική φυσική απόσταση(18). Ο αριθμός των ζευγών βάσεων στους οποίους αντιστοιχεί, ποικίλλει ευρέως σε όλο το γονιδίωμα (οι διαφορετικές περιοχές ενός χρωμοσώματος έχουν διαφορετική δυναμική ανασυνδυασμού) και επίσης εξαρτάται από το κατά πόσον ο ανασυνδυασμός είναι μέρος της ωογένεσης (σχηματισμός θηλυκοί γαμέτες) ή σπερματογένεσης (σχηματισμός αρσενικών γαμετών). Ένα centimorgan αντιστοιχεί σε περίπου 1 εκατομμύριο ζεύγη βάσεων στους ανθρώπους κατά μέσο όρο. Η σχέση είναι απλή, καθώς η φυσική χρωμοσωμική απόσταση που αντιστοιχεί σε ένα centimorgan ποικίλλει από τόπο σε τόπο στο γονιδίωμα και επίσης ποικίλλει μεταξύ ανδρών και γυναικών, δεδομένου ότι ο ανασυνδυασμός κατά το σχηματισμό γαμετών σε θηλυκά είναι σημαντικά συχνότερος

από τους άνδρες. Ο Morton υπολόγισε ότι το θηλυκό γονιδίωμα είναι 4782 cM σε μήκος, ενώ το αρσενικό γονιδίωμα έχει μήκος μόνο 2809 cM (19).

Επειδή ο γενετικός ανασυνδυασμός μεταξύ δύο δεικτών ανιχνεύεται μόνο εάν υπάρχει ένας περιττός αριθμός χρωμοσωματικών διασταυρώσεων μεταξύ των δύο δεικτών, η απόσταση σε centimorgans δεν αντιστοιχεί ακριβώς στην πιθανότητα γενετικού ανασυνδυασμού. Υποθέτοντας τη λειτουργία χάρτη του J. B. S. Haldane, στην οποία ο αριθμός των χρωμοσωματικών διασταυρώσεων κατανέμεται σύμφωνα με την κατανομή Poisson (4), μια γενετική απόσταση d σε centimorgans θα οδηγήσει σε έναν περιττό αριθμό χρωμοσωματικών διασταυρώσεων και συνεπώς σε έναν ανιχνεύσιμο γενετικό ανασυνδυασμό, με πιθανότητα:

$$\begin{aligned} \Pr[\text{recombination}|\text{linkage of } d \text{ cM}] &= \sum_{k=0}^{\infty} \Pr[2k + 1 \text{ crossovers}|\text{linkage of } d \text{ cM}] \\ &= \sum_{k=0}^{\infty} e^{-d/100} \frac{(d/100)^{2k+1}}{(2k+1)!} = e^{-d/100} \sinh(d/100) = \frac{1 - e^{-2d/100}}{2}, \end{aligned}$$

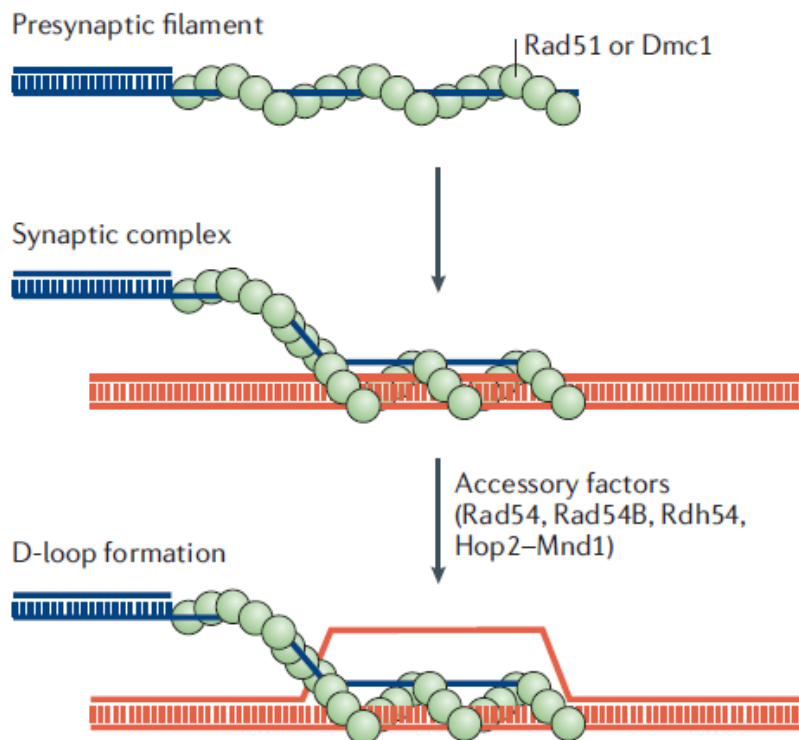
όπου \sinh είναι η συνάρτηση υπερβολής του συνημιτονίου. Η πιθανότητα ανασυνδυασμού είναι περίπου $d / 100$ για μικρές τιμές d και πλησιάζει το 50% καθώς το d πηγαίνει στο άπειρο. Ο τύπος μπορεί να αναστραφεί, δίνοντας την απόσταση σε centimorgans ως συνάρτηση της πιθανότητας ανασυνδυασμού:

$$d = 50 \ln \left(\frac{1}{1 - 2 \Pr[\text{recombination}]} \right). \quad (20)$$

1.1.5 Ο ΜΕΙΩΤΙΚΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ

Όπως αναφέρθηκε παραπάνω, για να ξεκινήσει η διαδικασία του ομόλογου ανασυνδυασμού, είναι απαραίτητη η ύπαρξη μονόκλωνου DNA το οποίο δημιουργείται από την επεξεργασία της δίκλωνης θραύσης. Ο ρόλος του είναι να προσελκύσει τις ρεκομπινάσες Rad51 και Dmc1. Η δομή αυτών είναι ένας δακτύλιος αποτελούμενος από

έξι και οκτώ μονομερή αντίστοιχα που τους επιτρέπει να δεσμεύουν και να υδρολύουν ATP. Αυτές πολυμερίζονται πάνω στο μονόκλωνο DNA δημιουργώντας ένα δεξιόστροφο πρωτεϊνικό νημάτιο, το λεγόμενο προσυναπτικό νημάτιο. Ο σχηματισμός αυτού απαιτεί μόνο την πρόσδεση του ATP στις ρεκομπινάσες και όχι την υδρόλυσή του. Το προσυναπτικό νημάτιο παρέχει το καταλυτικό κέντρο για τον σχηματισμό συνδέσεων των δύο μορίων DNA που πρόκειται να ανασυνδυαστούν και περιέχει δύο περιοχές δέσμευσης DNA: Μία στο αρχικό μονόκλωνο τμήμα και μία επικουρική που βοηθάει στην σύλληψη του δεύτερου μορίου. Το νεοσχηματιζόμενο σύμπλοκο που απαρτίζεται από το προσυναπτικό νημάτιο και το δίκλωνο μόριο DNA ονομάζεται συναπτικό σύμπλοκο και στα όρια αυτού αρχίζει η αναζήτηση για την ομόλογη περιοχή(21).

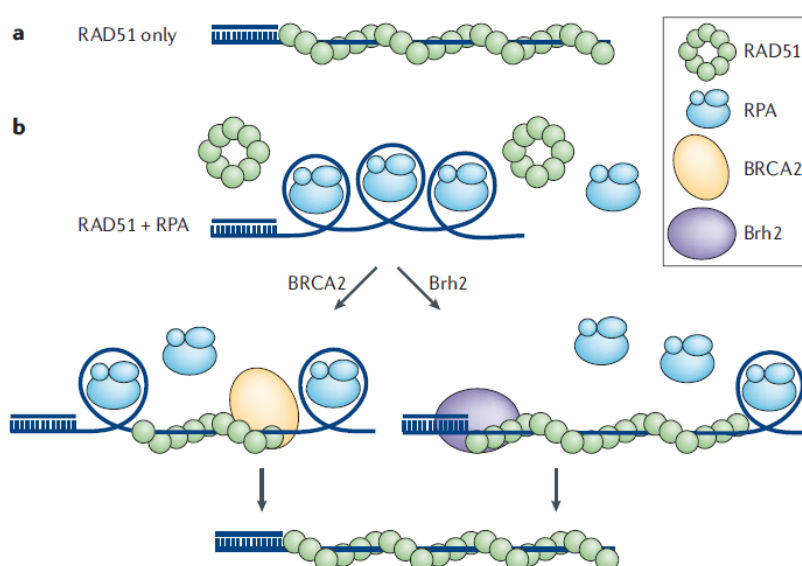


Εικόνα 1.8: Ο σχηματισμός του προσυναπτικού νηματίου και η είσοδός του στο μόριο δότη δημιουργώντας μία θηλιά D (22).

Επιπρόσθετα, έγινε αναφορά και σε μία πρωτεΐνη ικανή να δεσμεύει μονόκλωνο DNA η οποία βρίσκεται σε πολύ μεγαλύτερες ποσότητες από ότι οι ρεκομπινάσες, την RPA. Για να μπορέσει να συνδεθεί η Rad51 στο μονόκλωνο τμήμα το οποίο αναπόφευκτα θα είναι καλυμμένο με την RPA χρειάζεται έναν βοηθητικό παράγοντα,

την Rad52. Αυτή αλληλεπιδρά με την Rad51 και της επιτρέπει να συνδεθεί στο μονόκλωνο DNA. Ελάχιστη ποσότητα αυτής είναι απαραίτητη, καθώς χρησιμοποιείται μόνο στα αρχικά στάδια της σύνδεσης της Rad51 στο μονόκλωνο DNA ενώ ο πολυμερισμός της Rad51 συνεχίζεται από μόρια που δεν έχουν δεσμευμένη Rad52(23). Το σύμπλοκο Rad55-Rad57 έχει διακριτό ρόλο από την Rad52 καθώς σταθεροποιεί το ήδη σχηματισμένο προσυναπτικό νημάτιο παρεμποδίζοντας την πρόσδεση της RPA(24).

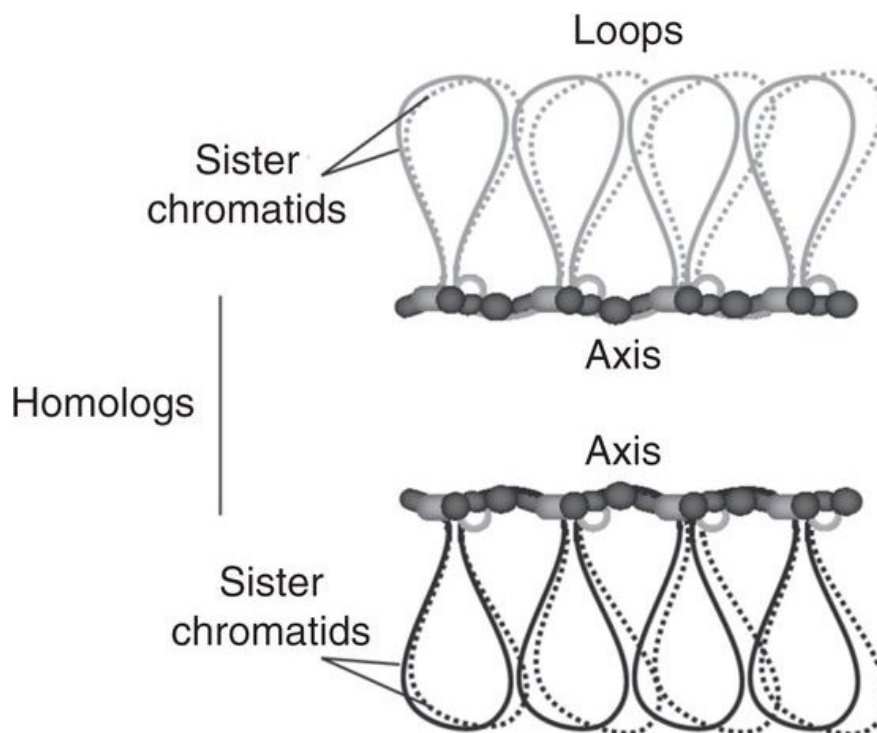
Στον άνθρωπο, η λειτουργία της RAD52 αντικαθίσταται από την BRCA2, με την RAD52 να κατέχει βοηθητικό ρόλο. Η δράση της παρουσιάζεται σχηματικά παρακάτω



Εικόνα 1.9: (a) In vitro σχηματισμός προσυναπτικού νηματίου χωρίς καμία παρεμπόδιση. (b) Η ύπαρξη της RPA εμποδίζει την πρόσδεση της RAD51. Η BRCA2 και η Brh2 (ανάλογη της BRCA2 που απομονώθηκε από την μύκητα *Ustilago maydis*) επιτρέπουν την σύνδεση της RAD51 στο μονόκλωνο DNA με δύο τρόπους (25).

Η παραπάνω διαδικασία ακολουθείται τόσο στον μιτωτικό όσο και στον μειωτικό ανασυνδυασμό. Στην μείωση όμως, ο επιθυμητός ανασυνδυασμός είναι μεταξύ ομόλογων χρωμοσωμάτων και όχι αδερφών χρωματίδων. Για να γίνει αυτό, τα χρωμοσώματα πρέπει να είναι οργανωμένα με τέτοιο τρόπο ώστε να προτιμηθεί η ανταλλαγή μεταξύ ομόλογων χρωμοσωμάτων. Έτσι, στην αρχή της μείωσης και στο στάδιο της λεπτοταινίας της πρόφασης I, οι αδερφές χρωματίδες δημιουργούν μια σειρά από θηλιές πάνω σε έναν πρωτεϊνικό άξονα, με αποτέλεσμα τα ομόλογα χρωμοσώματα

να αρχίσουν να έρχονται κοντά και να βρίσκονται σε απόσταση 400nm. Στον πρωτεϊνικό αυτό άξονα, το σύμπλοκο Red1/Hop1/Mek1 καθώς και οι ρεκομπινάσες Dmc1 και Rad51 παίζουν καθοριστικό ρόλο στο να μετατρέψουν τον ανασυνδυασμό από την μιτωτική του μορφή στην μειωτική του(26).



Εικόνα 1.10: Οι αδερφές χρωματίδες ενός χρωμοσώματος σχηματίζουν θηλιές κάτω από τις οποίες σχηματίζεται ένας άξονας δομικών πρωτεϊνών (27).

Κατά το στάδιο αυτό, τα τελομερή συνδέονται στην εσωτερική πλευρά του πυρηνικού φακέλου με την βοήθεια των μικροσωληνίσκων και συσσωματώνονται σε ένα σημείο στο οποίο μελλοντικά θα σχηματιστεί η μιτωτική άτρακτος. Στο επόμενο στάδιο, το στάδιο της ζυγοταινίας, τα χρωμοσώματα συμπυκνώνονται όλο και περισσότερο και τα ομόλογα χρωμοσώματα ευθυγραμμίζονται πλήρως με την διαδικασία της σύναψης και την δημιουργία του συναπτονηματικού συμπλόκου, στο οποίο οι δύο πρωτεϊνικοί άξονες, ένας από κάθε χρωμόσωμα, συνδέονται. Στην παχυταινία, η σύναψη ολοκληρώνεται και οι περιοχές στις οποίες επρόκειτο να γίνει ανασυνδυασμός, δημιουργούν δομές Holliday και κατά το τέλος της παχυταινίας με την διάσπαση των συναπτονηματικών συμπλόκων δημιουργούνται τα χιάσματα(28). Αυτό

είναι και το στάδιο στο οποίο παρατηρείται ανταλλαγή γενετικού υλικού μεταξύ των ομόλογων χρωμοσωμάτων.

1.2 ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΔΟΜΕΣ HOLLIDAY

Οι δομή Holliday είναι στην ουσία μία κινητή διασταύρωση μεταξύ 4 αλυσίδων DNA. Η δομή ονομάστηκε έτσι από τον Robin Holliday ο οποίος το πρότεινε το 1964 καθώς παρατηρούσε έναν συγκεκριμένο τύπο ανταλλαγής γενετικών πληροφοριών στους μύκητες, γνωστό και ως ομόλογο ανασυνδυασμό (29). Τα Holliday Junctions (όπως έχει επικρατήσει η ονομασία στην ευρύτερη επιστημονική κοινότητα) είναι ειδικά διατηρημένες δομές από τους προκαρυωτικούς οργανισμούς έως τα θηλαστικά. Πλέον έχουν αναγνωριστεί ως σημαντικά ενδιάμεσα που συμμετέχουν σε αρκετές κυτταρικές διαδικασίες, όπως ο ανασυνδυασμός (29), η ιϊκή ενσωμάτωση γενετικού υλικού (30) και η επιδιόρθωση του DNA (31). Ολοένα και αυξανόμενα αποδεικτικά στοιχεία δείχνουν ότι, πολλές από τις πρωτεΐνες που σχετίζονται με νεοπλασματικά νοσήματα, όπως τα προϊόντα των γονιδίων BLM και BRCA 2 (32, 33) που σχετίζονται με το σύνδρομο Bloom και καρκίνο του μαστού αντίστοιχα, φέρουν μια επιπλέον λειτουργία όπως η ενεργοποίηση ή καταστολή του ανασυνδυασμού, γεγονός που δείχνει ότι οι αλληλουχίες που είναι πιο ευαίσθητες στη δημιουργία των Holliday Junctions διαδραματίζουν σημαντικό ρόλο σε κάποια νοσήματα. Ένας μεγάλος αριθμός μεμονωμένων κρυσταλλικών δομών Holliday έχει βρεθεί τα τελευταία χρόνια, αποδεικνύοντας έτσι την ιδιότητα της μεταβλητότητας στην διαμόρφωσή τους (34). Από τα στοιχεία που έχουν συλλεχθεί μελετώντας τις δομές αυτές έχουν βγει ασφαλή συμπεράσματα που δηλώνουν κατά πόσο η εσωτερική διαμόρφωση της διασταύρωσης ποικίλλει ανάλογα με την αλληλουχία των βάσεων και την επίδρασή της ίδιας με φάρμακα και ιόντα και πώς οι αυτές μπορούν να επηρεάσουν ίσως την προσβασιμότητα της δομής σε πρωτεΐνες. Ο σχηματισμός της διασταύρωσης μελετήθηκε στη παρούσα

εργασία σε συγκεκριμένες δομές που έχουν αποδειχθεί ότι σχηματίζουν Holliday Junctions, τις λεγόμενες ανεστραμμένες επαναλήψεις (Inverted Repeats) και φαίνεται να εξαρτώνται από ένα κοινό τρινουκλεοτιδικό μοτίβο (ACC-triplet) στις θέσεις 6, 7 και 8 της δεκανουκλεοτιδικής αλληλουχίας d(CCn₁nnN6N7C8GG), σχηματίζοντας ισχυρούς υδρογονικούς δεσμούς, που σταθεροποιούν την εν λόγω διαστάυρωση (35).

1.2.1 ΤΡΟΠΟΙ ΔΗΜΙΟΥΡΓΙΑΣ HOLLIDAY JUNCTIONS

1.2.1.1 ΜΟΝΤΕΛΟ ΚΛΑΣΣΙΚΟΥ ΟΜΟΛΟΓΟΥ ΑΝΑΣΥΝΔΥΑΣΜΟΥ

Ο ομόλογος ανασυνδυασμός ίσως είναι ο πιο θεμελιώδης μηχανισμός ανασυνδυασμού. Τα βασικά χαρακτηριστικά του φαίνεται ότι είναι τα ίδια σε όλους οργανισμούς πάνω στη Γη. Παρόλο που ο μηχανισμός του ομόλογου ανασυνδυασμού δεν έχει διαλευκανθεί πλήρως, οι ακόλουθες ιδιότητες πιθανώς είναι κοινές σε όλα τα κύτταρα :

1. Δύο δίκλινα μόρια DNA που έχουν περιοχές με πολύ όμοια (ομόλογη) αλληλουχία νουκλεοτιδίων παρατάσσονται έτσι ώστε οι ομόλογες αλληλουχίες τους να βρίσκονται αντικριστά. Στη συνέχεια, τα μόρια αυτά μπορεί να επιχιαστούν: συγκεκριμένα, σε μια πολύπλοκη αντίδραση, οι δύο κλώνοι κάθε διπλής έλικας σπάνε και τα σπασμένα άκρα ενώνονται με τα άκρα του απέναντι μορίου του DNA και ξανασχηματίζουν δύο ακέραιες διπλές έλικες που αποτελούνται από τμήματα των δύο, διαφορετικών, μορίων του DNA.
2. Η θέση της ανταλλαγής μπορεί να βρίσκεται οπουδήποτε μέσα στις ομόλογες αλληλουχίες των νουκλεοτιδίων των δύο ανασυνδυασμένων μορίων DNA.

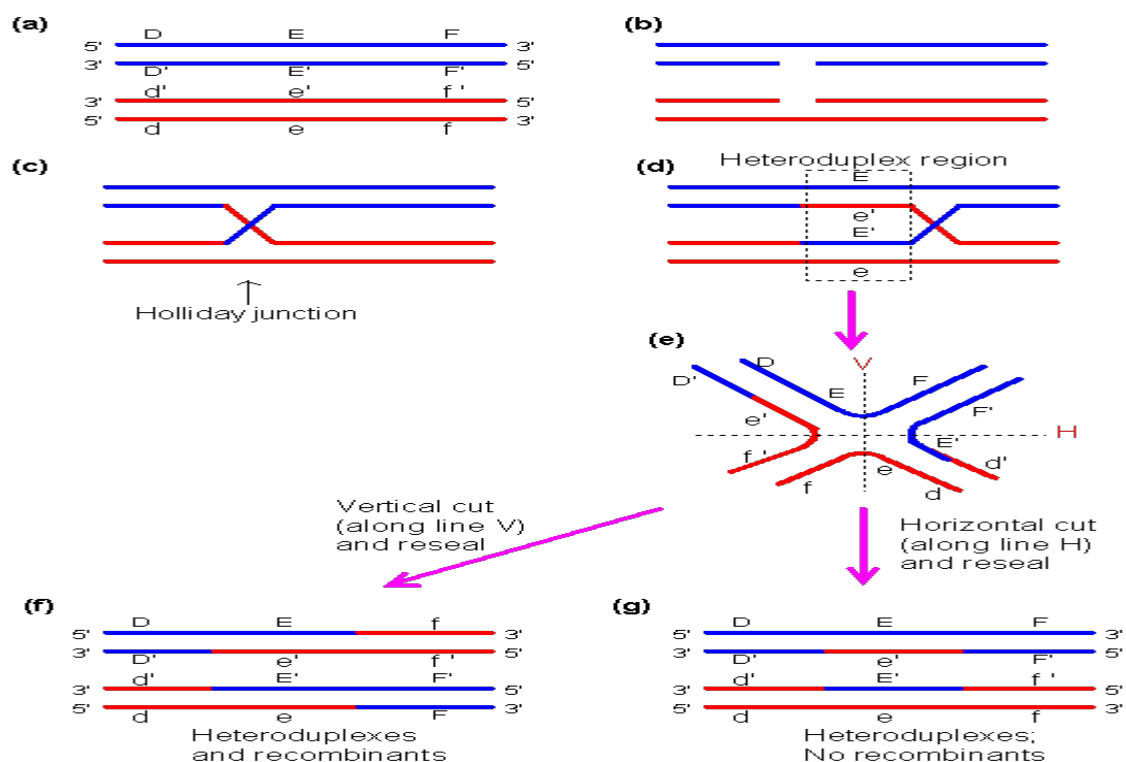
3. Στην θέση της ανταλλαγής, οι αλληλουχίες των νουκλεοτιδίων δεν μεταβάλλονται. Το σπάσιμο και η επανασύνδεση γίνονται με τέτοια ακρίβεια ώστε να μην χάνεται ούτε και να κερδίζεται έστω και ένα νουκλεοτίδιο.

Ο ομόλογος ανασυνδυασμός αρχίζει με ένα αποφασιστικό χτύπημα: ένα ειδικό ένζυμο (DNAse) κόβει τον ένα κλώνο από κάθε έλικα σε αντίστοιχες θέσεις και δημιουργεί ένα πλήρες ρήγμα στο μόριο του DNA. Στη συνέχεια τα 5' άκρα του ρήγματος αποδομούνται από ένα ένζυμο που πέπτει το DNA, αφήνοντας προεξέχοντα μονόκλωνα 3' άκρα. Καθένα από αυτά τα μονόκλωνα τμήματα αναζητά μια ομόλογη συμπληρωματική έλικα DNA για να ζευγαρώσει. Αν συμβεί αυτό, τότε μεταξύ των δύο χρωμοσωμάτων θα δημιουργηθεί ένα κοινό μόριο (joint molecule). Οι εγκοπές στους κλώνους του DNA σφραγίζονται ώστε τα δύο μόρια να συγκρατούνται τώρα μέσω επιχιασμού μεταξύ των δύο κλώνων, έναν από κάθε μόριο. Αυτό το κρίσιμο ενδιάμεσο στον ομόλογο ανασυνδυασμό αναφέρεται ως δομή ανταλλαγής των διασταυρούμενων κλώνων (cross-strand exchange) ή σύνδεση του Holliday (Holliday junction)(Εικ. 1.11).

Για να σχηματιστούν ξανά δύο ανεξάρτητα μόρια DNA, οι δύο επιχιαζόμενοι κλώνοι πρέπει να κοπούν. Αν η τομή γίνει οριζόντια της ετεριδιμερούς περιοχής, τα δύο αρχικά μόρια θα διαχωρίζονται το ένα από το άλλο σχεδόν αμετάβλητα. Ωστόσο, η ενδιάμεση δομή ανταλλαγής μπορεί να πραγματοποιήσει διάφορες περιστροφικές κινήσεις (ισομερείωση), ώστε οι δύο αρχικά μη διασταυρούμενοι κλώνοι να επιχιαστούν ή και το αντίστροφο. Αν οι διασταυρούμενοι κλώνοι κοπούν μετά από μία τέτοια περιστροφή, ένα τμήμα της κάθε αρχικής έλικας DNA θα συνδέεται με ένα τμήμα της άλλης έλικας. Με άλλα λόγια, τα δύο μόρια του DNA θα έχουν επιχιαστεί και έτσι θα έχουν παραχθεί δύο μόρια DNA με νέα αλληλουχία.

Τα κύτταρα χρησιμοποιούν εξειδικευμένες πρωτεΐνες για να διευκολύνουν τον ομόλογο ανασυνδυασμό. Οι πρωτεΐνες αυτές προκαλούν θραύσεις στο DNA, καταλύουν την ανταλλαγή κλώνων και κόβουν τις δομές Holliday. Επειδή τα βασικά χαρακτηριστικά του ομόλογου ανασυνδυασμού είναι πολύ διατηρημένα, οι πρωτεΐνες που διεκπεραιώνουν αυτή την διεργασία σε διαφορετικούς οργανισμούς συχνά έχουν πολύ ομόλογη αλληλουχία αμινοξέων.

Ο ομόλογος ανασυνδυασμός προσφέρει πολλά πλεονεκτήματα στα κύτταρα και τους οργανισμούς καθώς επιτρέπει στον οργανισμό να επιδιορθώσει το DNA που έχει υποστεί βλάβες και στους δύο κλώνους της διπλής έλικας και μπορεί να αντικαταστήσει άλλες γενετικές βλάβες που συμβαίνουν σχεδόν σε κάθε γύρο αντιγραφής του DNA. Επίσης είναι αναγκαίος για τον ακριβή διαχωρισμό των χρωμοσωμάτων που συμβαίνει κατά τη μείωση. Ο χρωμοσωμικός επιχιασμός προκαλεί την ανταλλαγή γενετικών πληροφοριών μεταξύ ομόλογων χρωμοσωμάτων, δημιουργώντας έτσι νέους συνδυασμούς αλληλουχιών DNA σε κάθε χρωμόσωμα. Στην παρακάτω εικόνα βλέπουμε τον σχηματισμό Holliday Junction κατά τη διαδικασία ομόλογου ανασυνδυασμού (36):



Εικόνα 1.11 (36)

- i. 2 ομόλογα μόρια DNA παρατάσσονται κατά τη μείωση
- ii. Αντίστοιχες τομές στα 2 δίκλινα μόρια
- iii. Δημιουργία σταυροειδούς τομής (δομή Holliday)
- iv. Διαδικασία μετανάστευσης (branch migration)
- v. Λύση της δομής Holliday .
- vi. Η κάθετη τομή θα δημιουργήσει ανασυνδυασμένα μόρια
- vii. Η οριζόντια τομή θα δημιουργήσει 2 ετεροδιμερή μόρια

1.2.1.2 ΤΟ ΜΟΝΤΕΛΟ ΔΙΚΛΩΝΗΣ ΘΡΑΥΣΗΣ ΕΝΟΣ ΜΟΡΙΟΥ (DSBS) ΚΑΙ ΕΠΙΔΙΟΡΘΩΣΗΣ ΤΟΥ.

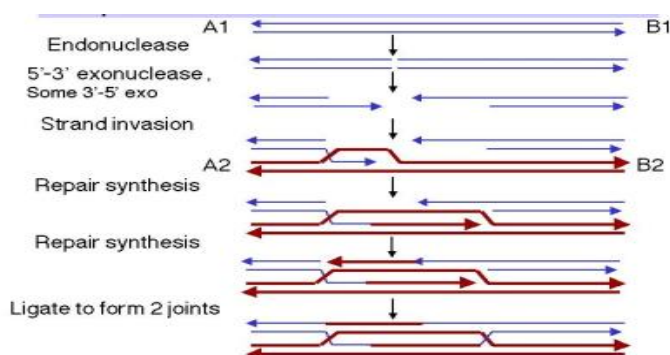
Το DNA στα κύτταρα εκτίθεται συνεχώς σε βλαπτικούς παράγοντες για το ίδιο. Ιονίζουσα ακτινοβολία όπως ακτίνες X και ακτίνες γ, καθώς και φάρμακα όπως η βλεομυκίνη, δημιουργούν δραστικές μορφές οξυγόνου, όπως ελεύθερες ρίζες οξυγόνου που προκαλούν δίκλωνη θραύση του μορίου DNA (DSBs) . Τέτοιου είδους βλαπτικές συνέπειες μπορούν να προκληθούν επίσης όταν η DNA πολυμεράση φτάσει σε μη διορθωμένη τομή αλυσίδας καθώς και από αναστολείς τοποϊσομερασών, όταν αυτοί συμβάλλουν στην εμπόδιση της τοποϊσομεράσης να επανενώσει τα άκρα μιας τομής στην αλυσίδα του DNA. Επίσης τα ίδια τα κύτταρα προάγουν τις δίκλωνες θραύσεις στο δικό τους DNA ως αποτέλεσμα φυσικών λειτουργιών των ίδιων , όπως πχ. κατά το συνδυασμό γονιδίων το οποίο συμβαίνει κατά τη διάρκεια ωρίμανσης των λεμφοκυττάρων και συμβάλλει στη δημιουργία ποικιλίας σε αντισώματα, υποδοχείς T-λεμφοκυττάρων καθώς και σε πρωτεΐνες του ανοσοποιητικού. Αυτές οι δίκλωνες θραύσεις (DSBs) μπορούν να δημιουργήσουν πολλά προβλήματα στα κύτταρα. Μία μόνο θραύση σε ένα σημαντικό γονίδιο μπορεί να σκοτώσει ένα κύτταρο, ή να το οδηγήσει στην αυτοκαταστροφή του μέσω ενεργοποίησης της διαδικασίας της απόπτωσης (37).

Για αυτό το λόγο τα κύτταρα έχουν αναπτύξει ταχύτατους και δραστικούς μηχανισμούς στην προσπάθεια αντιμετώπισης των συγκεκριμένων βλαβών. Συγκεκριμένα χρησιμοποιούν 2 μεθόδους επιδιόρθωσης των δίκλωνων θραύσεων, τον ομόλογο ανασυνδυασμό και τον μη-ομόλογο ανασυνδυασμό τα οποία θα αναλυθούν εκτενώς παρακάτω (37, 38). Πιο συγκεκριμένα, κατά τον ομόλογο ανασυνδυασμό το κατεστραμμένο χρωμόσωμα έρχεται σε άμεση επαφή με ένα ομόλογο μη κατεστραμμένο τμήμα DNA συνήθως της αδερφής χρωματίδας, το οποίο χρησιμεύει ως μήτρα για την επιδιόρθωση. Ο ομόλογος ανασυνδυασμός ξεκινά με την

νουκλεολυτική εκτομή των 2 σπασμένων αλυσίδων στην κατεύθυνση 5'-3'. Η κάθε τομή πραγματοποιείται σε διαφορετική αλυσίδα και είναι αποτέλεσμα της δράσης μιας ενδονουκλεάσης. Οι τομές μεγενθύνονται λόγω της υδρολυτικής δράσης μιας εξωνουκλεάσης με αποτέλεσμα να προκύψουν 2 κενά και προς τις 2 κατευθύνσεις του δίκλωνου μορίου, τα οποία εμφανίζουν 3' ελεύθερα άκρα. Το ένα από τα 3' ελεύθερα άκρα διεισδύει στο εσωτερικό του δεύτερου, άθικτου μορίου. Όταν η αλυσίδα που διεισδύει συναντήσει μια περιοχή με αλληλουχιακή ομολογία, σχηματίζεται μια δομή που ονομάζεται "D loop". Σε αυτό το σημείο δεσμεύεται η DNA πολυμεράση και χρησιμοποιεί:

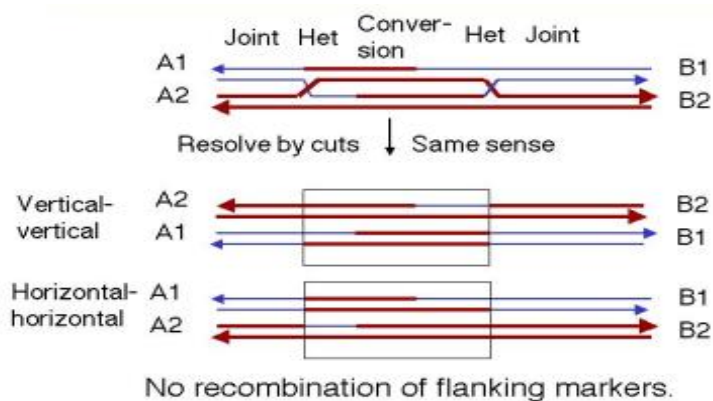
- ως εκκινητή το διεισδύσαν 3' άκρο
- τη συμπληρωματική αλυσίδα ως εκμαγείο
- ένα μηχανισμό επιδιορθωτικής σύνθεσης

Έτσι συνθέτει ένα μόριο μονόκλωνου DNA, εκτοπίζοντας παράλληλα την αρχική αλυσίδα του δίκλωνου μορίου. Η σύνθεση του νέου τμήματος και η εκτόπιση της πατρικής αλυσίδας συνεχίζονται έως ότου συναντήσουν το άκρο της δεύτερης τομής στην άλλη αλυσίδα του μορίου δέκτη, που δεν έχει διεισδύσει στο μόριο του δότη. Η εκτοπισμένη αλυσίδα υβριδοποιείται με τις συμπληρωματικές τις αλληλουχίες στη μη-διεισδύουσα αλυσίδα και σχηματίζονται 2 ετεροδιμερή μόρια. Το κενό που υφίσταται στη διεισδύουσα αλυσίδα του δέκτη, συμπληρώνεται με ένα δεύτερο συμβάν επιδιορθωτικής σύνθεσης DNA μέσω της διαδικασίας της μετανάστευσης (branch migration). Σε αυτό το σημείο (Εικόνα 2) η ενδιάμεση ανασυνδυαστική δομή έχει 2 συνδέσεις Holliday (Holliday Junctions).



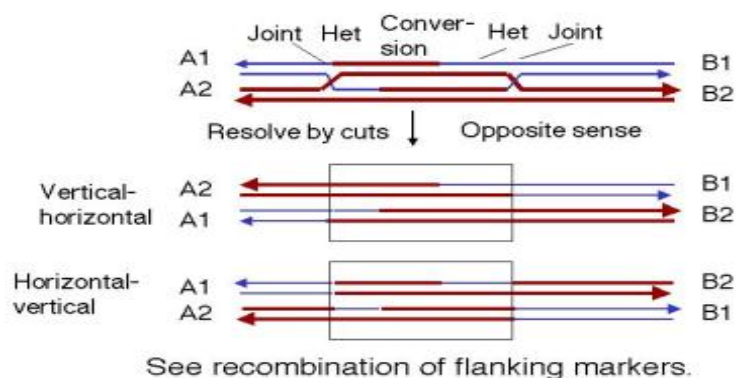
Εικόνα 1.12. Βήματα στο μοντέλο δίκλωνης θραύσης και επιδιόρθωσης μέσω ανασυνδυασμού, από την έναρξη στον σχηματισμό των δομών Holliday (39)

Η παρουσία των 2 δομών σίγουρα αυξάνει τον βαθμό πολυπλοκότητας επίλυσης των δομών, αλλά η διαδικασία ουσιαστικά είναι ίδια όπως συζητήθηκε για το απλό μοντέλο Holliday. Κάθε δομή Holliday θα υποστεί τομή κάθετη ή οριζόντια. Το κλειδί της υπόθεσης είναι όμως κατά πόσο οι 2 δομές θα λυθούν με τον ίδιο τρόπο (και οι 2 κάθετα ή και οι 2 οριζόντια) ή με αντίθετους τρόπους (η μία κάθετα και η άλλη οριζόντια). Αν και οι 2 δομές Holliday λυθούν με τον ίδιο τρόπο, τα αρχικά δίκλιωνα μόρια DNA θα απελευθερωθούν, το καθένα με μία περιοχή διαφορετικής γενετικής σύστασης, αποτέλεσμα της αλληλεπίδρασης των 2 μορίων. Πιο ειδικά, αν οι 2 δομές υποστούν τομή οριζόντια, τότε δεν έχουμε ανασυνδυασμό σε κανένα από τις 2 δομές. Αν υποστούν τομή κάθετη, τότε έχουμε ανασυνδυασμό και στις 2 δομές. Επειδή όμως ο ανασυνδυασμός γίνεται σε μικρό διάστημα βάσεων, το τελικό αποτέλεσμα είναι η παραγωγή μη ανασυνδυασμένων αρχικών μορίων.



Εικόνα 1.13. Παρουσίαση των δομικών ενδιάμεσων στο μοντέλο δίκλωνης θραύσης μετά από ίδιας φύσης τομή των δομών Holliday (και οι 2 κάθετα ή και οι 2 οριζόντια). Δεν υπάρχει ανασυνδυασμός μεταξύ των 2 μορίων (39)

Αντίθετα, αν η μία δομή Holliday υποστεί κάθετη τομή και η άλλη οριζόντια, τότε θα προκύψουν ανασυνδυασμένα αρχικά μόρια

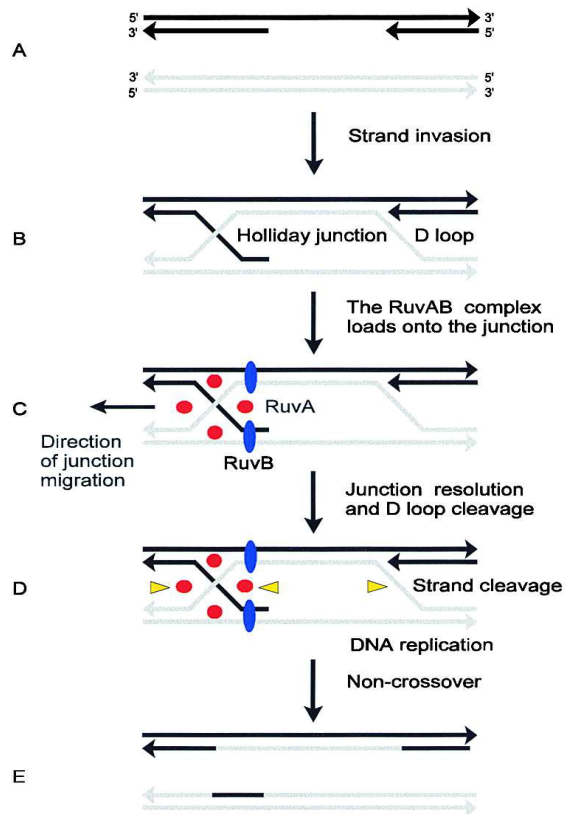


Εικόνα 1.14. Παρουσίαση των δομικών ενδιάμεσων στο μοντέλο δίκλωνης θραύσης μετά από αντίθετης φύσης τομή των δομών Holliday (η μία κάθετα και η άλλη οριζόντια). Υπάρχει ανασυνδυασμός μεταξύ των 2 μορίων (39)

Ωστόσο, παρατηρείται ακόμη διαφορετική αντιμετώπιση στην επιδιόρθωση σχηματιζόμενων δίκλωνων θραύσεων όσον αφορά διαφορετικά στάδια του ίδιου μονοπατιού όπως αυτό της νευρικής ανάπτυξης ποντικού. Πράγματι, κατά τα πρώιμα στάδια της νευρωνικής ανάπτυξης επάγεται ο ομόλογος ανασυνδυασμός, όπως αποδείχθηκε από πειράματα, όπου σημειώθηκε απόπτωση σε πολλαπλασιαζόμενα πρόδρομα νευρικά κύτταρα, τα οποία εμφάνισαν ελαττωματικό ομόλογο ανασυνδυασμό (40). Αντίστοιχα ήταν τα αποτελέσματα ως προς την απόπτωση, αλλά σε διαφοροποιημένα νευρικά κύτταρα μεταγενέστερων σταδίων ανάπτυξης, αναφορικά με ελαττωματικό μη ομόλογο ανασυνδυασμό. Επομένως, είναι προφανές πως ο ομόλογος ανασυνδυασμός συμμετέχει κατά τα πρώιμα στάδια νευρικής ανάπτυξης σε κύτταρα ποντικού, ενώ κατά τα όψιμα στάδια αυτής λαμβάνει χώρα ο μη ομόλογος ανασυνδυασμός.

1.2.1.3 ΤΟ ΜΟΝΤΕΛΟ ΔΙΕΙΣΔΥΣΗΣ ΜΟΝΗΣ ΑΛΥΣΙΔΑΣ

Η αδυναμία κληρονομικής φύσης στην επιδιόρθωση βλαβών του DNA είναι αποτέλεσμα διαφόρων ασθενειών, πολλές από τις οποίες χαρακτηρίζονται από νευρολογικές δυσλειτουργίες και/ή γενετική αστάθεια και καρκίνους. Από τους διάφορους τύπους βλαβών που προκύπτουν σε κύτταρα, η θραύση μονής αλυσίδας (SSB) είναι η πιο κοινή, και φθάνει σε μία συχνότητα των δεκάδων χιλιάδων ανά κύτταρο αποτέλεσμα ενδοκυτταρικών μεταβολών και αυθόρμητων θραύσεων του DNA. Ο μοριακός μηχανισμός επιδιόρθωσης περιγράφεται παρακάτω (41). Αρχικά η μία αλυσίδα ενός από τα 2 δίκλινα μόρια υφίσταται μια τομή και το ελεύθερο άκρο που προκύπτει διεισδύει στο εσωτερικό της δεύτερης διπλής έλικας, η οποία είναι άθικτη. Η διεισδύουσα αλυσίδα κινείται κατά μήκος του δεύτερου μορίου στο εσωτερικό του, έως ότου συναντήσει τη συμπληρωματική της περιοχή. Όταν συμβεί αυτό, η διεισδύουσα αλυσίδα εκτοπίζει την ομόλογή της και καταλαμβάνει τη θέση της. Το εκτοπισμένο τμήμα της ομόλογης αλυσίδας υδρολύεται και προκύπτει ελεύθερο άκρο. Το κενό που αφήνει η διεισδύουσα αλυσίδα στην έλικα από την οποία προέρχεται, συμπληρώνεται από την DNA πολυμεράση που χρησιμοποιεί ως εκμαγείο την υφιστάμενη αλυσίδα. Το ελεύθερο άκρο που έχει προκύψει από την υδρόλυση του εκτοπισμένου τμήματος, ενώνεται με το άκρο του τμήματος που συνέθεσε η DNA πολυμεράση. Με τον τρόπο αυτόν δημιουργείται η δομή Holliday (42). Ο τρόπος λύσης της δομής είναι ακριβώς ίδιος με τον τρόπο που προαναφέρθηκε.



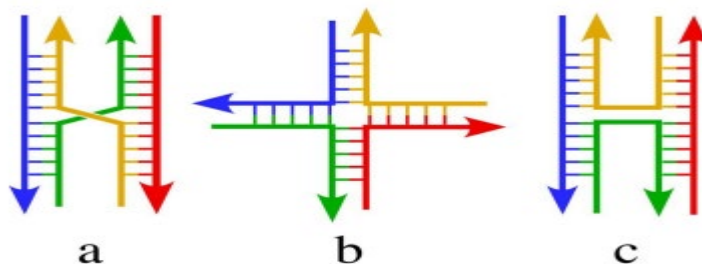
Εικόνα 1.15. Δημιουργία HJ ως συνέπεια μονόκλωνης θραύσης DNA (43)

1.2.2 ΔΥΝΑΜΙΚΗ ΚΑΙ ΔΟΜΗ ΤΩΝ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ

(IR) ΠΟΥ ΣΧΗΜΑΤΙΖΟΥΝ HJ

1.2.2.1 ΑΡΧΙΚΑ ΜΟΝΤΕΛΑ ΤΩΝ ΔΟΜΩΝ HOLLIDAY

Η δομή και η διαμόρφωση των HJ ήταν στο επίκεντρο του ενδιαφέροντος από την στιγμή που προτάθηκαν για πρώτη φορά από τον Robin Holliday το 1964. Τα βασικά δομικά χαρακτηριστικά των DNA junctions έγιναν για πρώτη φορά κατανοητά, μετά από μελέτες σε διαλύματα την δεκαετία του 1980 (44). Πιο συγκεκριμένα, διάφορες επιστημονικές ομάδες χρησιμοποίησαν ασύμμετρα κατασκευάσματα ακολουθιών DNA, τα οποία αποτρέπουν την μετανάστευση (branch immigration) και την επιδιορθωση των junctions με ανοιχτά άκρα. Η γενική δομή βρέθηκε να εξαρτάται τόσο από τον τύπο, όσο και από την συγκέντρωση των κατιόντων που υπάρχουν στο διάλυμα, με την Holliday junction να υιοθετεί είτε μια ανοικτή διαμόρφωση σε συνθήκες χαμηλής συγκέντρωσης ιόντων (open-X), είτε μια κλειστή διαμόρφωση σε συνθήκες υψηλής συγκέντρωσης ιόντων (stacked-X) (45, 46). Η open-X διαμόρφωση κυριαρχεί σε συνθήκες χαμηλής συγκέντρωσης, με τους 4 B-DNA κλώνους διατεταγμένους ουσιαστικά σε μία τετραγωνική επίπεδη διαμόρφωση έτσι ώστε να ελαχιστοποιηθούν οι ηλεκτρικές αλληλεπιδράσεις ανάμεσα στις φωσφορικές ομάδες. Υψηλές συγκεντρώσεις μονοσθενών και δισθενών κατιόντων, βοηθούν στην προστασία των φωσφορικών ομάδων, επιτρέποντας στον κόμβο Holliday να λάβει την πιο συμπαγή stacked-X διαμόρφωση στην οποία οι βραχίονες στοιβάζονται ομοαξονικά για να δημιουργήσουν 2 ημισυνεχή B-DNA διμερή. Στην stacked-X διαμόρφωση, οι αλυσίδες των B-DNA κλώνων, θεωρητικά, είναι είτε παράλληλες είτε αντιπαράλληλες. Μέχρι σήμερα έχει διαπιστωθεί μόνο σχέση αντιπαράλληλότητας μεταξύ των αλυσίδων βάσει καταγεγραμμένων πειραματικών μελετών. Ο αντιπαράλληλος κόμβος μπορεί να πάρει είτε μία ειδική ισομερής μορφή είτε να ταλαντεύεται μεταξύ των 3 καταστάσεων με την open-X διαμόρφωση να λειτουργεί ως ενδιάμεσο σε διάφορες διαδικασίες. Ειδικότερα, αυτός ο αντιπαράλληλος σχηματισμός του κόμβου, τοπολογικά είναι ανίκανος να μεταναστεύσει κατά μήκος των ομόλογων αλυσίδων DNA, ενώ τόσο η παράλληλη μορφή (parallel stacked-X) όσο και η ανοιχτή (open-X) μπορούν να επιτελέσουν αυτήν την λειτουργία (branch migration) (47). Ποιοι είναι όμως οι παράγοντες που καθορίζουν την διαμόρφωση της διαστάυρωσης? Θα δούμε παρακάτω ότι τόσο η αλληλουχία των HJ όσο και η συγκέντρωση κατιόντων είναι οι βασικοί καθοριστές.



Εικόνα 1.16. Δομικές μορφές των DNA Holliday Junctions. a) Η παράλληλη κλειστή μορφή (stacked-X) η οποία αρχικά προτάθηκε από τον Holliday ως ενδιάμεση δομή του ανασυνδυασμού b) Η ανοιχτή μορφή (open-X) c) Η αντιπαράλληλη κλειστή μορφή που δεν επιτρέπει την μετακίνηση των δομών Holliday κατά μήκος των αλυσίδων που γίνεται η ανταλλαγή (branch migration) (48)

1.2.2.2. ΠΡΩΤΕΣ ΜΟΝΟ-ΚΡΥΣΤΑΛΛΙΚΕΣ ΔΟΜΕΣ ΤΩΝ HOLLIDAY JUNCTION

Η πλήρης κατανόηση της μονο-κρυσταλλικής δομής των Holliday junction θεωρήθηκε τεράστιο επίτευγμα στην επιστημονική κοινότητα. Η ακριβής μοριακή δομή των HJ διευκρινίστηκε στα μέσα της δεκαετίας του 1990 ως συμπλέγματα DNA με διάφορες πρωτεΐνες με ανασυνδυαστικές και επιδιορθωτικές ιδιότητες, συμπεριλαμβανομένου της RunA επιδιορθωτικής πρωτεΐνης, RunC, Cre ρεκομπινάση και Flp ρεκομπινάση (47). Σε όλα αυτά τα σύμπλοκα τα HJ υφίστανται σε ανοιχτή μορφή διαμόρφωσης (open-X), πιθανώς για να είναι δυνατή η μετακίνηση κατά μήκος των αλληλουχιών (branch migration). Η κρυσταλλική δομή των κόμβων απουσία πρωτεϊνών ανιχνεύθηκε πρώτα σε ένα σύμπλοκο RNA/DNA, όπου δομές των HJ στην φυσική τους μορφή καθορίστηκαν ταυτόχρονα από 2 διαφορετικά εργαστήρια. Και στις 2 περιπτώσεις η εύρεση των δομών Holliday αποτέλεσαν τυχαία ευρήματα καθώς στην πρώτη περίπτωση η εν λόγω διασταύρωση ήταν κομμάτι μιας ευρύτερης αλληλουχίας

που δημιουργήθηκε για να μελετήσει την δομή παρακείμενων G-A αταίριαστων βάσεων, ενώ στην δεύτερη περίπτωση ήταν αποτέλεσμα μελέτης της δομής που δημιουργείται απο ενδοκλωνική ένωση θυμίνης-θυμίνης υπο την επίδραση του φαρμάκου ψωραλενίου. Πιο συγκεκριμένα η πρώτη δομή αποτελούσε μια ανεστραμμένη επαναλαμβανόμενη ακολουθία (IR) με αναντιστοιχίες G-A (5'-CCGGGACCGG-3') , ενώ η δεύτερη αποτελούσε κανονική ανεστραμμένη επαναλαμβανόμενη αλληλουχία (5'-CCGGTACCGG-3') .

1.2.2.3 ΑΛΛΗΛΟΥΧΙΑΚΕΣ ΚΑΙ ΙΟΝΤΙΚΕΣ ΑΛΛΗΛΕΠΙΔΡΑΣΕΙΣ ΣΤΟ ΜΟΤΙΒΟ ΤΩΝ ΥΠΟ ΜΕΛΕΤΗ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ ΣΤΟΝ ΣΧΗΜΑΤΙΣΜΟ ΚΑΙ ΤΗΝ ΔΙΑΜΟΡΦΩΣΗ ΤΩΝ ΗJ

Από την σύγκριση των δύο προαναφερθέντων αλληλουχιών του προηγούμενου αποσπάσματος με παρόμοιες ακολουθίες , οι οποίες είχαν χαρακτηριστικά διαμόρφωσης B-DNA προκύπτει ότι το τρινουκλεοτίδιο ACC στις θέσεις N6N7N8 του κοινού δεκανουκλεοτιδιακού μοτίβου διαμόρφωσης d(CCnnnN6N7N8GG) αποτελεί τον πυρήνα σταθεροποίησης των ΗJ. Η υπόθεση αυτή υποστηρίχθηκε από την παρατήρηση ότι αυτός ο κεντρικός πυρήνας (ACC-triplet) βρέθηκε ακριβώς στο σημείο της διασταύρωσης με την κυτοσίνη στην θέση N8 να σχηματίζει απευθείας δεσμούς υδρογόνου με τις φωσφορικές ομάδες των διειδύουσων αλυσίδων απο το σημείο U-turn(σημείο της διασταύρωσης). Τα ευρήματα αυτά εγείρουν ερωτήσεις του πώς η αλληλουχία , οι ομάδες υποκαταστατών και τα κατιόντα επηρεάζουν την διαμόρφωση και τον σχηματισμό των Holliday κόμβων.

Οι κρυσταλλικές δομές των Holliday Junctions πλέον προσδιορίζονται από ανεστραμμένες επαναλήψεις (IR sequences) στο δεκανουκλεοτίδιο d(CCnnnN6N7C8GG) οι οποίες περιέχουν τον κοινό ACC πυρήνα (ACC-triplet) και μπορούν να σχηματιστούν (i) με αντικατάσταση των τερματικών βάσεων από G-C σε T-A, (ii) με αντικατάσταση της

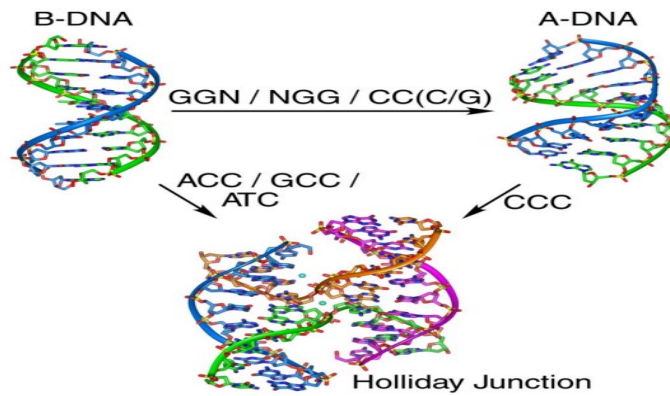
κυτοσίνης με μεθυλιωμένη κυτοσίνη στην θέση N8, (iii) με αντικατάσταση βάσεων από ινোসίνη 2-αμινοπουρίνη και βρωμιούχα ανάλογα στις θέσεις N6N7N8 και (iv) υπό την παρουσία δισθενών κατιόντων συμπεριλαμβανομένου του μαγνησίου, ασβεστίου και στροντίου. Οι δομές επιβεβαιώνουν ότι (i) ο ACC πυρήνας είναι απαραίτητος (αλλά όχι και απόλυτα αναγκαίος) για την δημιουργία των HJ, (ii) οι αλληλεπιδράσεις της αμινομάδας της κυτοσίνης C8 με τις φωσφορικές ομάδες στα σημεία της διασταύρωσης είναι σημαντικές όχι μόνο για την δημιουργία των διασταυρώσεων (HJ) αλλά και για τον καθορισμό της γεωμετρίας των ίδιων, και (iii) και ύπαρξη κατιόντων συμβάλλει στην τελική γεωμετρία των κόμβων.

Είδαμε ότι η ύπαρξη του ACC πυρήνα στο προαναφερθέν δεκανουκλεοτίδιο αποτελεί βασική προϋπόθεση για την δημιουργία Holliday Junctions σίγουρα όμως όχι αναγκαία. Μετά από την διεξαγωγή πολλών ερευνών διαπιστώθηκε ότι και η ύπαρξη διαφορετικών τριπλετων στις συγκεκριμένες θέσεις οδηγούν στον σχηματισμό Holliday κόμβων. Πειραματικές μελέτες στο κοινό δεκανουκλεοτιδιακό μοτίβο d(CCnnnN6N7C8GG) (όπου NNC είναι οποιοδήποτε από τα 4 νουκλεοτίδια, και nnn είναι οι ανεστραμμένες επαναλήψεις των NNC) έδειξαν την ύπαρξη 3 διαφορετικών τρινουκλεοτιδιακών πυρήνων που σχηματίζουν HJ υπό συγκεκριμένες συνθήκες. Τονίζουμε το υπο συγκεκριμένες συνθήκες γιατί σε αντίθεση με το ACC τρινουκλεοτίδιο που σχηματίζει μόνο HJ, η ύπαρξη των 3 άλλων πυρήνων δεν οδηγεί μόνο σε σχηματισμό HJ αλλά και σε δομές με χαρακτηριστικά A-DNA και B-DNA, πρόκειται δηλαδή για αμφιμορφικές αλληλουχίες. Πιο συγκεκριμένα τα τρινουκλεοτίδια $N_6 N_7 N_8 = ACC, GCC, ATC, CCC$ και οι αντίστοιχες ευρύτερες αλληλουχίες μπορούν υπό κατάλληλες συνθήκες να σχηματίσουν HJ. Η ακολουθία με τον πυρήνα GCC σχηματίζει HJ υπό την παρουσία Ca^{2+} (κατιόντων ασβεστίου) ενώ έμφανίζει χαρακτηριστικά B-DNA υπό την παρουσία Mg^{2+} (κατιόντων μαγνησίου). Επιπλέον η αλληλουχία με πυρήνα ATC σχηματίζει μόνο υπό την παρουσία Ca^{2+} και δομή Holliday και δομή B-DNA, όπου ευνοείται η κλειστή (stacked-X) διαμόρφωση σε υψηλές συγκεντρώσεις ιόντων. Αντίθετα σε μικρότερες συγκεντρώσεις ιόντων ευνοείται η ανοιχτή (open-X) διαμόρφωση, η οποία στη συνέχεια επιτρέπει την μετακίνηση (branch migration) και την επακόλουθη μετατροπή της "διασταύρωσης" σε δομή με χαρακτηριστικά B-DNA.

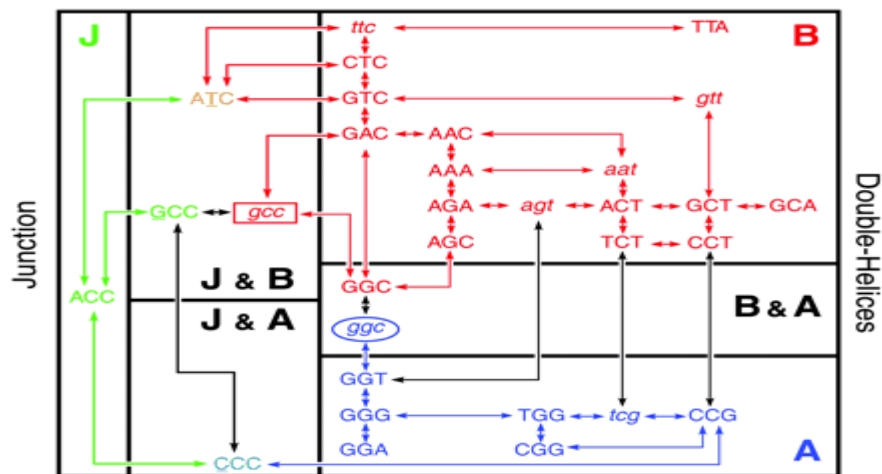
Τέλος, η αλληλουχία με πυρήνα CCC σχηματίζει HJ μόνο υπό την παρουσία Ca^{2+} σε υψηλή συγκέντρωση, ενώ δομή A-DNA σε χαμηλή συγκέντρωση Ca^{2+} αυτή τη φορά. Μετά από συγκρίσεις των αλληλουχιών με τους 4 διαφορετικούς πυρήνες σε πειραματικό επίπεδο, προέκυψαν αποτελέσματα τα οποία αποτελούν γενικά χαρακτηριστικά των αλληλουχιών που σχηματίζουν Holliday Junctions. Τα συμπεράσματα που προέκυψαν είναι τα εξής :

Αρχικά, όλες οι αλληλουχίες περιέχουν στη θέση N7 πυριμιδίνη. Από τη σύγκριση ACC με το αμφιμορφικό ATC, προέκυψε ότι η κυτοσίνη σταθεροποιεί την ένωση σε καλύτερο βαθμό από ότι η θυμίνη. Η σταθεροποιητική δράση της πυριμιδινικής βάσης είναι κυρίως ηλεκτροστατική με την αμινομάδα της κυτοσίνης να υπερτερεί της μεθυλομάδας της θυμίνης, όσον αφορά την αντίδρασή τους με το οξυγόνο της φωσφορικής ομάδας. Έτσι προκύπτει ο γενικός αλληλουχιακός κανόνας $NNC > NNT$ στον σχηματισμό δομών Holliday κύριως όπως αναφέρθηκε πριν για ηλεκτροστατικούς λόγους.

Με μία μόνο εξαίρεση, η θέση N6 του τρινουκλεοτιδιακού πυρήνα καλύπτεται πάντα από πουρίνη, με $A > G$ στην σταθεροποίηση της διασταύρωσης. Η σειρά αυτή προκύπτει από την παρατήρηση ότι ο ATC πυρήνας είναι αμφιμορφικός και ικανός να σχηματίσει HJ, αλλά αντίστοιχα ο GTC σχηματίζει μόνο δομές με χαρακτηριστικά B-DNA. Η μόνη εξαίρεση που αναφέρθηκε πριν είναι ότι την θέση N6 μπορεί να την καταλάβει και η βάση C. Και πάλι, τα αποτελέσματα δείχνουν ότι μια κυτοσίνη μπορεί να σταθεροποιήσει σε μικρότερο βαθμό μια διασταύρωση από ότι η αδερίνη ή γουανίνη καθώς ο πυρήνας CCC είναι αμφιμορφικός, ενώ ο CTC δημιουργεί μόνο δομές με χαρακτηριστικά B-DNA. Έτσι η σειρά προτεραιότητας για την θέση N6 είναι $A > G > C$ (35, 49).



Εικόνα 1.17. Αλληλουχιακή επίδραση σε δομές B-DNA, A-DNA και Holliday Junctions. Η σταθερότητα των δομών εξαρτάται από την αλληλουχία του δεκανουκλεοτιδίου $d(\text{CCn}n\text{nN}_6\text{N}_7\text{N}_8\text{GG})$ και ειδικότερα από την αλληλουχία της πυρηνικής τριπλέτας (N₆N₇N₈). Στην εικόνα παρατηρούμετα αμφιμορφικά τρινουκλεοτίδια τα οποία οδηγούν στο σχηματισμό HJ από B-DNA, A-DNA από B-DNA, και HJ από A-DNA (35, 49)



Εικόνα 1.18. Χάρτης των διαφορετικών δομών του δεκανουκλεοτιδίου $d(\text{CCn}n\text{nN}_6\text{N}_7\text{N}_8\text{GG})$. Ο χάρτης χωρίζεται σε 3 διαφορετικές δομικές κατηγορίες (σημειωμένο με B για δομές B-DNA, με A για δομές A-DNA και J για Holliday Junctions) καθώς και τις ενδιάμεσες καταστάσεις μεταξύ των δομών (αναφέρεται στις αμφιμορφικές ακολουθίες). Το ορθογώνιο γύρω από τον πυρήνα GCC υποδεικνύει ότι η δομή σχηματίζεται από μια αλλαγή στα δισθενή ιόντα (δηλ. Από Ca^{2+} σε Mg^{2+}). Εμείς επικεντρωνόμαστε στο αριστερό σκέλος της εικόνας (Junction) όπου και μας δείχνει τους τρινουκλεοτιδιακούς πυρήνες που σχηματίζουν HJ (49).

Αξίζει στο τέλος του συγκεκριμένου κεφαλαίου να παρουσιάσουμε τις αλληλουχίες των ανεστραμμένων επαναλήψεων που σχηματίζουν Holliday Junctions. Παρέχονται επίσης επιπλέον πληροφορίες, όπως η παρουσία συγκεκριμένων ιόντων

που σταθεροποιούν τη διασταύρωση καθώς και το είδος των δεσμών που δημιουργούνται.

Structure ¹	Description of Structure				Geometry ²			Core Interactions ³			
	Sequence	asu	Cation	Ref ⁴	IDA	J_{roll}	J_{slide}	C ₈ -C ₇	G ₃ -A ₆	C ₇ -A ₆	Na-Cavity
ACC-4Na	CCGGTACCGG	4	Na	(46)	41.4°	159.6°	0 Å	Direct	1H ₂ O	Direct	Na ⁺
ACC-2Na	CCGGTACCGG	2	Na	(78)	42.7°	163.5°	2.0 Å	Direct	1H ₂ O	Direct	Absent
ACC-2Ca1	CCGGTACCGG	2	Ca	UD0024	39.6°	158.5°	0 Å	Direct	1H ₂ O	Absent	H ₂ O
ACC-2Ca2	CCGGTACCGG	2	Ca	UD0025	39.4°	158.8°	0 Å	Direct	1H ₂ O	1H ₂ O	H ₂ O
tACC-4Ca	TCGGTACCGA	4	Ca	(74)	39.6°	159.9°	0 Å	Direct	1H ₂ O	Direct	H ₂ O
tACC-2Ca	TCGGTACCGA	2	Ca	UD0023	39.6°	168°	3.6 Å	Direct	Absent	Absent	H ₂ O
tACC-2Sr1	TCGGTACCGA	2	Sr	(74)	43.0°	169.0°	2.4 Å	Direct	2H ₂ O	1H ₂ O	Absent
tACC-2Sr2	TCGGTACCGA	2	Sr	UD0026	43.3°	169.5°	2.4 Å	Direct	2H ₂ O	1H ₂ O	H ₂ O
gACC-2Na	CCGGGΔCCGG	2	Na	(45)	40.6°	174.5°	0 Å	Direct	1H ₂ O	Absent	H ₂ O
ACmC-2Ca	CCGGTACm ⁵ CGG	2	Ca	(47)	41.3°	170.4°	3.4 Å	1H ₂ O	2H ₂ O	1H ₂ O	Absent
GCC-2Na	CCGGCGCCGG	2	Na	(75)	40.0°	159.8°	0 Å	Direct	1H ₂ O	Direct	Absent
ACbU-2Ca	CCAGTACbr ⁵ UGG	2	Ca	(75)	38.8°	164.6°	2.4 Å	Absent	Absent	1H ₂ O	Na ⁺

Εικόνα 1.19. Στην στήλη **Structure**, βλέπουμε τρινουκλεοτιδικούς πυρήνες καθώς και τα κύρια ιόντα τα οποία βοηθούνε στο σχηματισμό των HJ, στην στήλη **Description of structure**, βλέπουμε την ολοκληρωμένη μορφή της αλληλουχίας που σχηματίζει HJ καθώς επίσης και πληροφορίες όπως τον αριθμό εύρεσης των συγκεκριμένων δομών στη Nucleic Acid Database, στην στήλη **Geometry** διάφορες πληροφορίες για την γεωμετρία των δομών και τέλος στην στήλη **Core Interactions**, τις διάφορες αλληλεπιδράσεις των τρινουκλεοτιδικών πυρήνων και το είδος των δεσμών που σχηματίζονται (49)

1.3 HOTSPOTS ΚΑΙ ΑΛΛΑ ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΩΣ ΣΗΜΕΙΑ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΚΑΤΑ ΜΗΚΟΣ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

Η συχνότητα εμφάνισης ανασυνδυασμού δεν είναι ίδια σε όλο το γονιδίωμα. Υπάρχουν περιοχές που εμφανίζουν ανασυνδυασμούς σε πολύ μεγαλύτερες συχνότητες (recombination hotspots) από ότι άλλες (recombination coldspots) και η ύπαρξη ενός θερμού σημείου είναι στενά συνδεδεμένη με την ύπαρξη σημείων που πραγματοποιούνται δίκλωνες θραύσεις, καθώς σε ένα hotspot υπάρχουν αρκετά τέτοια

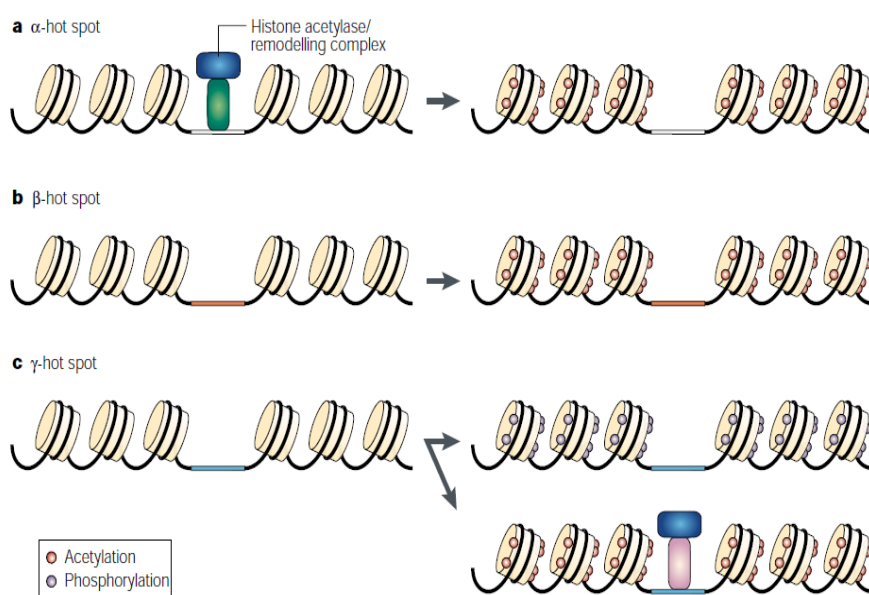
σημεία κατά μήκος μιας περιοχής 100 έως 500 ζευγών βάσεων(50). Τα hotspots έχουν χωριστεί σε τρεις κατηγορίες οι οποίες παρουσιάζονται παρακάτω:

Ως *α-hotspots* ορίζονται τα σημεία τα οποία απαιτούν την σύνδεση κάποιου μεταγραφικού παράγοντα για να παρουσιάσουν αυξημένα επίπεδα ανασυνδυασμού. Ένα παράδειγμα είναι το θερμό σημείο *HIS4* του *S.cerevisiae* το οποίο απαιτεί την σύνδεση των μεταγραφικών παραγόντων Bas1, Bas2, Gcn4 και Rap1 ανοδικά του. Αξιοσημείωτο είναι πως απαραίτητη είναι μόνο η σύνδεση των μεταγραφικών παραγόντων και όχι τόσο η μεταγραφή καθώς πειράματα στα οποία η απαραίτητη για την μεταγραφή ακολουθία TATAA έχει εξαλειφθεί, δεν παρουσιάζουν μείωση στην ενεργότητα του σημείου *HIS4* ως hotspot. Τα σημεία αυτά χρησιμοποιούν τους μεταγραφικούς παράγοντες για να προσελκύσουν σύμπλοκα HAT που τροποποιούν τις ιστόνες (50, 51).

Ως *β-hotspots* ορίζονται τα σημεία χρωματίνης με υπερευαισθησία σε νουκλεάσες τα οποία δεν απαιτούν πρόσδεση κάποιου μεταγραφικού παράγοντα. Η ακολουθία (CCGNN) επαναλαμβανόμενη 12 φορές, είναι ικανή να επάγει ανασυνδυασμό και μεταγραφή χωρίς την ύπαρξη μεταγραφικού παράγοντα καθώς κανένας αναγνωρισμένος μεταγραφικός παράγοντας δεν αναγνωρίζει την ακολουθία αυτή. Επιπρόσθετα, η παραπάνω ακολουθία δεν ευνοεί τον σχηματισμό νουκλεοσωμάτων με αποτέλεσμα η χρωματίνη στο σημείο αυτό να είναι υπερευαίσθητη στην δράση της DNase I και επακόλουθα να είναι ευπαθής σε δίκλωνες θραύσεις. Το γεγονός ότι δεν υπάρχουν νουκλεοσώματα στην περιοχή που πρόκειται να συμβεί ανασυνδυασμός διευκολύνει τα σύμπλοκα HAT να τροποποιήσουν τις ιστόνες. Επίσης, το σημείο αυτό είναι πιο προσβάσιμο για την Sro11 λόγω έλλειψης μεταγραφικών παραγόντων (50, 52).

Τέλος, ως *γ-hotspots* ορίζονται τα σημεία που συσχετίζονται με περιοχές του χρωμοσώματος υψηλές σε G και C. Στα σημεία αυτά η χρωματίνη δεν είναι εύκολα προσβάσιμη. Μάλιστα τα σημεία του DNA με υψηλό ποσοστό GC πιθανώς να δημιουργούν καθυστερήσεις στην διχάλα της αντιγραφής.

Πειράματα δείχνουν ότι οι ιστόνες σε αυτές τις διχάλες αντιγραφής μπορούν να τροποποιηθούν και ειδικότερα να φωσφορυλιωθούν από την κινάση Mec1 με σκοπό την χαλάρωση της χρωματίνης, ξεμπλοκάροντας έτσι την διαδικασία της αντιγραφής. Η φωσφορυλίωση αυτή αναγνωρίζεται και από τους μηχανισμούς του ανασυνδυασμού. Άλλη πιθανότητα είναι η ύπαρξη μιας πρωτεΐνης που αναγνωρίζει περιοχές DNA πλούσιες σε G και C η οποία είτε αλληλεπιδρά απευθείας με το σύμπλοκο ρεκομπινασών που πραγματοποιούν τον ανασυνδυασμό είτε οδηγεί σε τροποποιήσεις των ιστονών οι οποίες με τη σειρά τους αλληλεπιδρούν με τις ρεκομπινάσες (50, 53, 54).



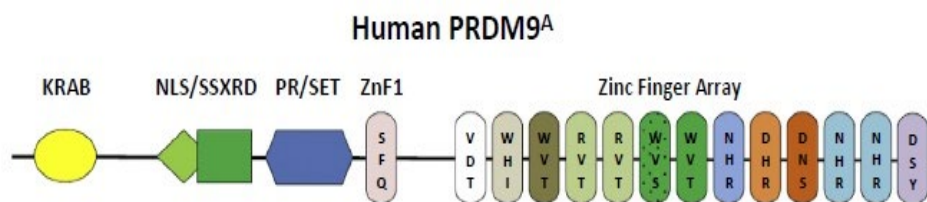
Εικόνα 1.20: Τα τρία είδη θερμών σημείων. Οι κίτρινοι κύλινδροι απεικονίζουν τα νουκλεοσώματα ενώ οι σφαιρικές δομές πάνω σε αυτά δείχνουν μετα-μεταφραστικές τροποποιήσεις, δηλαδή ακετυλίωση και φωσφορυλίωση των ιστονών. Οι περιοχές μη κωδικού DNA εντοπίζονται με λευκό (a), πορτοκαλί (b) και μπλέ (c) χρώμα (55).

Στον άνθρωπο έχουν εντοπιστεί θερμά σημεία των οποίων οι ακολουθίες DNA παρουσιάζουν υψηλά ποσοστά σε G και C και θα μπορούσαν έτσι να θεωρηθούν ως γ-hotspots. Έρευνες δείχνουν θετικές συσχετίσεις μεταξύ των ποσοστών GC και της συχνότητας του ανασυνδυασμού, όμως η συσχέτιση αυτή είναι αρκετά ασθενής.

Παρόλο που ελάχιστα θερμά σημεία έχουν μελετηθεί με ακρίβεια, υπολογίζεται πως στο ανθρώπινο γονιδίωμα υπάρχουν περισσότερα από 25000 τέτοια σημεία. Για

τον καθορισμό ενός σημείου ως hotspot ευθύνεται η πρωτεΐνη PRDM9. Η δομή της πρωτεΐνης αυτής έχει ως εξής:

- i. Μία περιοχή KRAB στο αμινο-τελικό άκρο της, η οποία σχετίζεται με αλληλεπιδράσεις μεταξύ πρωτεϊνών. Ειδικότερα, στην περιοχή αυτή συνδέονται οι πρωτεΐνες EWSR1, CDYL, EHMT και CXX1. Οι CDYL και EHMT δεσμεύονται στην PRDM9 στα αρχικά στάδια της μείωσης και ο ρόλος τους εικάζεται πως είναι να ορίζουν τα όρια κάθε hotspot, περιορίζοντας το πόσα νουκλεοσώματα θα μεθυλιωθούν πλευρικά από την επιθυμητή ακολουθία. Σε επόμενη φάση, η EWSR1 δημιουργεί σύμπλοκο με την PRDM9 το οποίο αλληλεπιδρά με την κοχεζίνη REC8 με αποτέλεσμα η αλληλουχία του hotspot να έρθει κοντά στον άξονα των δομικών πρωτεϊνών ώστε να ξεκινήσει η διαδικασία του ανασυνδυασμού.
- ii. Ένα μοτίβο SSXRD που επιτρέπει στην πρωτεΐνη να μεταφέρεται στον πυρήνα.
- iii. Μία περιοχή SET που της προσδίδει ενεργότητα μεθυλοτρανσφεράσης.
- iv. Μία συντηρημένη περιοχή zinc finger.
- v. Μία σειρά από οκτώ έως είκοσι υψηλά πολυμορφικών zinc fingers, τα οποία καθορίζουν την εξειδίκευση της πρωτεΐνης για διαφορετικές νουκλεοτιδικές αλληλουχίες. Λόγω της μεγάλης πολυμορφίας που παρατηρείται, σε έναν πληθυσμό διαφορετικά άτομα θα έχουν διαφορετικές νουκλεοτιδικές αλληλουχίες που θα δρουν ως hotspot ανάλογα με τα αλληλόμορφα που έχουν κληρονομήσει ως προς τα zinc fingers αυτά. Η αναγνώριση των αλληλουχιών από την PRDM9 γίνεται στις θηλιές της χρωματίνης, όμως η διαδικασία του ανασυνδυασμού ξεκινάει στον πρωτεϊνικό άξονα όπου η SPO11 δημιουργεί δίκλωνες θραύσεις με την βοήθεια των TOPOVIB, MEI4, IHO1 καθώς και των πρωτεϊνών που αναφέρθηκαν στο i).



Εικόνα 1.21: Η δομή της ανθρώπινης PRDM9. Στα zinc fingers αναφέρονται τα αμινοξικά κατάλοιπα των θέσεων 1,3 και 6 της κάθε α -έλικας που σχετίζονται με την δέσμευση του DNA (56).

Χρονικά, η PRDM9 εμφανίζεται στην αρχή της μείωσης και ειδικότερα στο στάδιο της λεπτοταινίας, ενώ μέχρι το τέλος της ζυγοταινίας έχει εξαλειφθεί. Η παρουσία της στον πυρήνα της επιτρέπει να δεσμεύεται στο DNA σε συγκεκριμένες ακολουθίες, ανάλογα με τα zinc fingers που κατέχουν. Ο ρόλος των zinc fingers είναι να δεσμεύουν το DNA και να σταθεροποιούν την δέσμευση αυτή, αλλά για την εξειδίκευση της ακολουθίας στην οποία δεσμεύονται ευθύνονται μερικά και όχι όλα. Επιπρόσθετα, όπως αναφέρθηκε και στην δομή της, η PRDM9 έχει την ικανότητα να μεθυλιώνει τις λυσίνες 4 και 36 της ιστόνης H3, αναδιοργανώνοντας έτσι την μορφή της χρωματίνης στα τροποποιημένα νουκλεοσώματα. Ως αποτέλεσμα, τα νουκλεοσώματα αυτά απομακρύνονται το ένα από το άλλο αφήνοντας ένα τμήμα της χρωματίνης προσβάσιμο στην SPO11 η οποία θα δημιουργήσει την δίκλωνη θραύση και θα οδηγήσει στην ανταλλαγή γενετικού υλικού μεταξύ των ομόλογων χρωμοσωμάτων. Τέλος, μια γνωστή λειτουργία της αλλά με άγνωστη έως σήμερα σημασία, είναι η ικανότητά της να αυτομεθυλιώνει τις ιστόνες της. Εικάζεται πως η αυτομεθυλίωση σχετίζεται με την σταθερότητα της και την απότομη εξαφάνισή της κατά την ζυγοταινία ή με την ικανότητά της να μεθυλιώνει τις ιστόνες και να δεσμεύει το DNA, η οποία πιθανώς να ελαττώνεται.

Δύο νουκλεοτιδικές ακολουθίες που σχετίζονται με την ύπαρξη ενός θερμού σημείου είναι τα ολιγονουκλεοτίδια CCTCCCT και CCCACCCC. Η ανακάλυψή τους προέκυψε από μελέτες στα πολυμορφικά hotspots DNA2 και NID1 όπου SNPs που εκφυλίζουν τις παραπάνω ακολουθίες οδηγούν σε σημαντική μείωση της ικανότητας ανασυνδυασμού. Η πρώτη ακολουθία εμφανίζεται περισσότερες φορές σε ρετροτρανσποζόνια THE1A/B μέσα στα hotspots ενώ συνολικά και οι δύο εντοπίζονται στο ~10% των αναγνωρισμένων hotspot. Το μικρό αυτό ποσοστό και η ανικανότητά τους

να προβλέψουν την ύπαρξη των περισσότερων ανθρώπινων hotspots, ώθησε τους επιστήμονες να ψάξουν για άλλες ακολουθίες έως ότου κατέληξαν στην ακολουθία CCNCCNTNNCCNC, απαρτιζόμενη από δεκατρία νουκλεοτίδια και συγγενική με την πρώτη προηγουμένως ανακαλυφθείσα ακολουθία.

Για την ανακάλυψη της συναινετικής αυτής ακολουθίας, αρχικά αναγνωρίστηκαν διάφορες επαναλαμβανόμενες ακολουθίες DNA που εντοπίζονται συχνά μέσα σε θερμά σημεία ανασυνδυασμού και έπειτα αναζητήθηκαν ακολουθίες που σχετίζονται με την ύπαρξη ή μη θερμών σημείων σε διάφορες κατηγορίες επαναλαμβανόμενου DNA, όπως τα μεταθετά στοιχεία και το δορυφορικό DNA. Αναλύοντας μη επαναλαμβανόμενο DNA και χρησιμοποιώντας ως βάση το επταμερές CCTCCCT και θέτοντας πλευρικά όρια μήκους 50 ζευγών βάσεων, βρέθηκε ότι το τετραμερές CCAC δύο ζεύγη βάσεων καθοδικά από το επταμερές είναι το πιο ισχυρό επιπρόσθετο αποδεικτικό στοιχείο για την ύπαρξη θερμού σημείου. Έτσι, η ακολουθία CCTCCCTNNCCAC σχετίζεται τόσο με hotspots σε επαναλαμβανόμενο DNA όπως τα ρετροτρανσποζόνια THE1A/B όσο και με μη επαναλαμβανόμενες αλληλουχίες.

Η συναινετική ακολουθία CCNCCNTNNCCNC προέκυψε ύστερα από έρευνες εκτός επαναλαμβανόμενων ακολουθιών με την μέθοδο του mismatching στο οποίο αντικαθιστώντας μία μόνο βάση του παραπάνω μοτίβου στις θέσεις 3, 6 και 12 εξακολουθούσε να παρατηρείται ενεργότητα του σημείου ως hotspot. Με την μέθοδο της μέγιστης πιθανοφάνειας, οι ερευνητές κατέληξαν στο συμπέρασμα ότι το 41% των hotspots που αναγνώρισαν (τα 9.306 από τα 22.699 σε όλο το γονιδίωμα) απαιτεί την ύπαρξη ενός ή περισσότερων αντιγράφων της συναινετικής ακολουθίας για την διεκπεραίωση επιχιασμού στο hotspot αυτό. Τέλος, συμπέραναν ότι η ύπαρξη της ακολουθίας αυτής μπορεί να προβλέψει την ύπαρξη θερμού σημείου πολύ καλύτερα σε επαναλαμβανόμενες αλληλουχίες (73% στο THE1A) παρά σε κανονικές (10% σε τυχαίες αλληλουχίες). Αυτό οφείλεται στην φύση των επαναλαμβανόμενων στοιχείων, καθώς έχουν νουκλεοτίδια με μεγαλύτερη ικανότητα ανασυνδυασμού, κάτι που είναι λογικό εφόσον αναλογιστούμε την ικανότητά τους να πολλαπλασιάζονται και να αλλάζουν θέση μέσα στο γονιδίωμα(56-58).

ΚΕΦΑΛΑΙΟ 2

ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΑΝΤΙΓΡΑΦΙΚΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ

2.1 ΕΙΣΑΓΩΓΗ ΣΤΑ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ (TES, TRANSPOSABLE ELEMENTS) ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

Ως γενικός όρος τα μεταθετά στοιχεία αποτελούν διάσπαρτες επαναλαμβανόμενες DNA αλληλουχίες, οι οποίες έχουν τη δυνατότητα να μεταπηδούν σε διαφορετικές θέσεις μέσα στο γονιδίωμα ενίοτε προκαλώντας γενετικές αναδιατάξεις. Ανάλογα με το μηχανισμό που χρησιμοποιούν για τη μετακίνησή τους τα μεταθετά στοιχεία μπορούν να κατηγοριοποιηθούν σε ρετροτρανσποζόνια (Class I TEs) και σε DNA τρανσποζόνια (Class II TEs). Σε ένα γενικότερο πλαίσιο τα DNA τρανσποζόνια μετακινούνται μέσω ενός “copy-and-paste” μηχανισμού, ενώ τα ρετροτρανσποζόνια παρόλο που χρησιμοποιούν τον ίδιο μηχανισμό, για τη μετάθεσή τους απαιτείται και η μεσολάβηση ενός RNA ενδιάμεσου. Τόσο τα DNA τρανσποζόνια όσο και τα ρετροτρανσποζόνια διαιρούνται σε επιμέρους υπερ-οικογένειες, υπο-οικογένειες κλπ, με τα τελευταία να αποτελούνται από τις πιο σημαντικές οικογένειες ρετρομεταθετών που θα μας απασχολήσουν περισσότερο παρακάτω, όπως τα LTR (Long Terminal Repeats) ρετροτρανσποζόνια, τα LINEs (Long Interspersed Nuclear Elements) και τα SINEs (Short Interspersed Nuclear Elements) στοιχεία(59, 60).

2.1.1 DNA ΤΡΑΝΣΠΟΖΟΝΙΑ

2.1.1.1 ΔΡΑΣΤΗΡΙΟΤΗΤΑ ΤΩΝ DNA ΤΡΑΝΣΠΟΖΟΝΙΩΝ ΣΤΟ ΓΟΝΙΔΙΩΜΑ ΤΩΝ ΘΗΛΑΣΤΙΚΩΝ ΚΑΤΑ ΤΗ ΔΙΑΡΚΕΙΑ ΤΗΣ ΕΞΕΛΙΞΗΣ

Αν και κατά τα πρώιμα στάδια της εξέλιξης, τα DNA τρανσποζόνια διατηρούσαν μεγαλύτερη ευελιξία ως προς τη μετακίνησή τους, τα τελευταία 37 εκατομμύρια χρόνια η δραστηριότητά τους είναι μηδενική στη γενεαλογία των πρωτευόντων οργανισμών, στους οποίους μεταξύ άλλων συγκαταλέγεται και ο άνθρωπος. Παρόλο που κατά τη διαφοροποίηση των πρωτευόντων περίπου 100.000 αντίγραφα από τέσσερις διαφορετικές υπεροικογένειες DNA τρανσποζονίων ενσωματώθηκαν στο DNA αυτών των οργανισμών, κανένα από αυτά δεν κατάφερε φαινομενικά να επιβιώσει έπειτα από τη μαζική εξάλειψή τους από το γονιδίωμα ενός προγονικού ανθρωποειδούς, που έλαβε χώρα πριν από 40 εκατομμύρια χρόνια. Αναλύσεις και σε γονιδιώματα ποντικού, αρουραίου και σκύλου δείχνουν πως τα DNA τρανσποζόνια έχουν απολέσει τη δυνατότητα μεταπήδησης στο γονιδίωμα τα τελευταία 40 έως 50 εκ. χρόνια. Τα αποτελέσματα αυτά συγκλίνουν στην άποψη πως τα DNA τρανσποζόνια εξαφανίστηκαν ανεξάρτητα σε διαφορετικές γενεαλογίες θηλαστικών κατά την ίδια χρονική περίοδο (Ηώκαινος εποχή 35-55 εκ. χρόνια), ενώ παράλληλα δεν έχει παρατηρηθεί οποιοδήποτε περιστατικό μετέπειτα ενσωμάτωσης σε κάποια από αυτές τις γενεαλογίες από εκείνη την εποχή και ύστερα. Ωστόσο, αυτό δεν ισχύει για όλα τα είδη των θηλαστικών μιας και έχει παρουσιαστεί έντονη δραστηριότητα DNA τρανσποζονίων, και δη ορισμένων που έχουν εξαφανιστεί από τα υπόλοιπα θηλαστικά, στο απλοειδές γονιδίωμα της νυχτερίδας *Myotis lucifigus* (36, 61, 62).

2.1.1.2 ΔΡΑΣΤΗΡΙΟΤΗΤΑ ΤΩΝ DNA ΤΡΑΝΣΠΟΖΟΝΙΩΝ ΣΤΟ ΓΟΝΙΔΙΩΜΑ ΤΩΝ ΘΗΛΑΣΤΙΚΩΝ ΚΑΤΑ ΤΗ ΔΙΑΡΚΕΙΑ ΤΗΣ ΕΞΕΛΙΞΗΣ

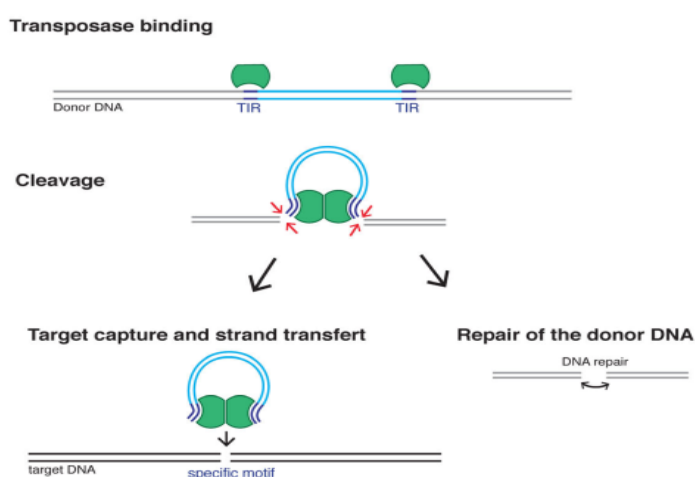
2.1.1.2.1 ΜΗΧΑΝΙΣΜΟΙ ΜΕΤΑΘΕΣΗΣ ΤΩΝ DNA ΤΡΑΝΣΠΟΖΟΝΙΩΝ

Τα DNA τρανσποζόνια που εδράζονται στους ευκαρυώτες, διαχωρίζονται σε τρεις υποκατηγορίες ανάλογα με τον τρόπο που χρησιμοποιούν για τη μετακίνησή τους: 1) αυτά τα οποία αποκόπτονται ως δίκλινα τμήματα DNA και ενσωματώνονται ξανά σε άλλο σημείο του γονιδιώματος, 2) εκείνα που χρησιμοποιούν ένα μηχανισμό κυκλικού αναδιπλασιασμού (“rolling cycle” mechanism) και ονομάζονται *Helitrons*, και 3) τα λεγόμενα *Mavericks*, τα οποία αντιγράφονται με τη βοήθεια μιας DNA πολυμεράσης την οποία κωδικοποιούν τα ίδια. Στην παρούσα εργασία θα μας απασχολήσει κυρίως η πρώτη κατηγορία DNA τρανσποζονίων, τα οποία αντιγράφονται με το “copy-and-paste” μηχανισμό μετακίνησης(63, 64).

2.1.1.2.2 ΜΗΧΑΝΙΣΜΟΣ “COPY-AND-PASTE”

Το σύνηθες μέγεθος των DNA τρανσποζονίων κυμαίνεται μεταξύ 1,5 έως 5 kb. Η δυνατότητα μετακίνησης των DNA τρανσποζονίων εξασφαλίζεται από το γονίδιο του ενζύμου της τρανσποζάσης που κωδικοποιούν, η οποία αναγνωρίζει τις δύο τυχαίες ανάστροφες επαναλήψεις (TIRs, Tandem Inverted Repeats) που πλαισιώνουν το γονίδιο

της και προσδένεται σε αυτές. Μετά την πρόσδεση της στις επαναλήψεις, η τρανσποζάση κόβει το 5' άκρο και των δύο, σχηματίζοντας ένα διμερές DNA-τρανσποζάσης, το οποίο ακολούθως αφού αποκοπεί, αναγνωρίζει ειδικές αλληλουχίες-στόχους σε κάποιο άλλο σημείο του γονιδιώματος, διαφορετικό ανάλογα με την εκάστοτε οικογένεια που μετακινείται. Τέλος, το διμερές ενσωματώνεται σε νέα θέση στο γονιδίωμα και συνθέτει μικρές θέσεις-στόχους διπλασιασμού (TSDs, Tandem Site Duplications) στα άκρα της ένθεσης μήκους 4-8 pb.



Εικόνα 2.1 Μηχανισμός “copy-and-paste” μετάθεσης (65)

Γενικότερα, τα DNA τρανσποζόνια που χρησιμοποιούν τον παραπάνω μηχανισμό, μετακινούνται χωρίς να καθίσταται απαραίτητη η αντιγραφή τους σε σχέση με τις άλλες δύο κατηγορίες. Παραδείγματος χάρη, σε περίπτωση που η μετάθεσή τους πραγματοποιηθεί κατά το στάδιο της αντιγραφής, όπου ένα DNA τρανσποζόνιο μετατεθεί από μία προσφάτως συντεθειμένη χρωματίδα σε μία άλλη που δεν έχει προλάβει ακόμη να συντεθεί, τότε στο θυγατρικό κύτταρο που θα προκύψει, θα περιέχεται ένα επιπλέον αντίγραφο του εν λόγω τρανσποζονίου. Άλλος ένας τρόπος με τον οποίο τα DNA τρανσποζόνια καταφέρνουν να συσσωρεύουν πολλαπλά αντίγραφα στο γονιδίωμα των ξενιστών τους, είναι μέσω της επιδιόρθωσης των δίκλωνων θραύσεων που προκαλούν καθώς αποκόπτονται από το γονιδίωμα καθώς μεταναστεύουν σε άλλη θέση. Πιο συγκεκριμένα, οι δίκλωνες θραύσεις αυτές επιδιορθώνονται μέσω του ομόλογου ανασυνδυασμού και της χρήσης του ομόλογου

χρωμοσώματος (ή της αδελφής χρωματίδας, αναλόγως το στάδιο του κυτταρικού κύκλου που λαμβάνει χώρα η μετάθεση) ως μήτρα(66). Με τον τρόπο αυτό το μεταθετό αναγεννάται και αυξάνονται τα αντίγραφα του. Διαφορετικά, η δίκλωνη θραύση που προκύπτει μπορεί να επιδιορθωθεί και μέσω του μη ομόλογου ανασυνδυασμού αφήνοντας “αποτυπώματα” τρανσποζονίων από τις TSDs περιοχές που παραμένουν από το αποκομμένο τρανσποζόνιο(63, 64, 67).

2.1.2 ΒΙΟΛΟΓΙΑ ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ

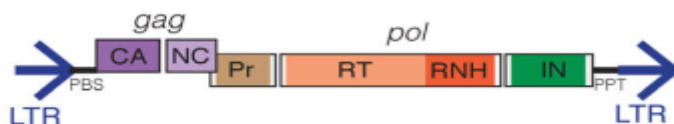
2.1.2.1 LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΑ

Το χαρακτηριστικό αυτών των ρετροτρανσποζονίων αποτελεί η ύπαρξη, εκατέρωθεν της περιοχής που κωδικοποιεί τα γονίδια που επάγουν τη μετακίνησή τους, των μακριών τερματικών επαναλήψεων LTR (Long Terminal Repeats). Με μια πρώτη ματιά γίνεται αντιληπτό πως τα συγκεκριμένα ρετροτρανσποζόνια χρησιμοποιούν για τη μετάθεσή τους ένα μηχανισμό που μοιράζεται πολλά κοινά στοιχεία με τον τρόπο με τον οποίο μετακινούνται οι ρετροϊοί. Ωστόσο, αντίθετα με τους ρετροϊούς, τα LTR ρετροτρανσποζόνια δε σχηματίζουν σωματίδια ικανά να μεταναστεύσουν και εκτός του κυττάρου και να μολύνουν γειτονικά κύτταρα.

Όλα τα λειτουργικά LTR ρετροτρανσποζόνια κωδικοποιούν δύο γονίδια, τα οποία είναι απαραίτητα για την αντιγραφή τους, τα gag και pol γονίδια. Οι πολυπρωτεΐνες, με καψίδιο και με μια περιοχή πυρηνικού καψιδίου (nucleocapsid domain), που κωδικοποιεί το γονίδιο gag, σχηματίζουν σωματίδια παρόμοια με τα ιικά σωματίδια (virus-like particles) μέσα στα οποία λαμβάνει χώρα η αντίστροφη μεταγραφή. Το γονίδιο pol κωδικοποιεί τρεις διαφορετικές πρωτεΐνες: μία πρωτεάση (PR, protease), μία

αντίστροφη μεταγραφή RT (Reverse Transcriptase) με RNase H περιοχές και τέλος μια ιντεγκράση (IN, Integrase).

LTR-retrotransposon (5-20 kb)



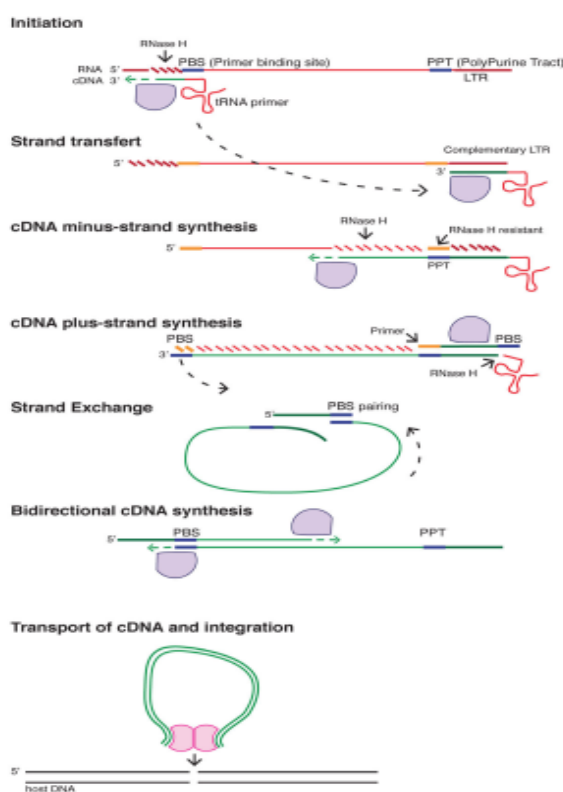
LTR: Long terminal repeat. **gag** and **pol:** polyprotein-coding genes **CA** and **NC:** capsid and nucleocapsid protein. **Pr** and **IN:** Protease and Integrase protein **RT** and **RNH:** Reverse transcriptase and RNase H domain. **PBS:** Primer binding site **PPT:** Poly Purine track

Εικόνα 2.2 Δομή LTR ρετροτρανσποζονίων (68)

2.1.2.1.1 ΜΗΧΑΝΙΣΜΟΣ ΜΕΤΑΘΕΣΗΣ LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ ΜΕΣΑ ΣΤΑ VIRUS-LIKE ΣΩΜΑΤΙΔΙΑ ΣΤΟ ΚΥΤΤΑΡΟΠΛΑΣΜΑ

Αρχικά, το mRNA των LTR ρετροτρανσποζονίων συντίθεται μέσω της RNA πολυμεράσης II του ξενιστή, η οποία αναγνωρίζει μια περιοχή εκκινήτη στο 5' LTR του ρετροτρανσποζονίου. Ακολούθως, η αντίστροφη μεταγραφή για τη σύνθεση του cDNA τμήματος που θα ενσωματωθεί στο τέλος σε άλλο σημείο του γονιδιώματος του ξενιστή, ξεκινάει με την αναγνώριση από ένα αγγελιοφόρο RNA, t-RNA, του ξενιστή, μιας μικρής περιοχής καθοδικά του 5' LTR. Το t-RNA, που προσδέεται στην περιοχή πρόσδεσης του υποκινήτη (PBS, Promoter Binding Site), χρησιμοποιείται ως εκκινήτης για τη σύνθεση της μίας cDNA αλυσίδας (minus-strand synthesis). Κατόπιν η νεοσυντιθέμενη cDNA αλυσίδα που αντιγράφηκε από το 5' LTR μεταφέρεται στο συμπληρωματικό του 3' LTR

άκρο και χρησιμοποιείται ως εκκινητής για την αντίστροφη μεταγραφή και της υπόλοιπης cDNA minus αλυσίδας (69). Μετά τη σύνθεση της cDNA minus αλυσίδας ακολουθεί η σύνθεση της cDNA plus αλυσίδας, μέσω της βοήθειας του πολύ-πουρινικού τμήματος της RNase H-resistant περιοχής, το οποίο λειτουργεί ως υποκινητής για τη σύνθεση του 3' LTR και του συμπληρωματικού PBS τμήματος. Η RNase H παράλληλα με την αντίστροφη μεταγραφή του cDNA αποσυνθέτει το mRNA ενδιάμεσο που δε χρησιμοποιείται πλέον ως εκμαγείο και αποδεσμεύει το αγγελιοφόρο tRNA από τη cDNA minus αλυσίδα. Η cDNA plus αλυσίδα, τέλος, αλληλεπιδρά με τη cDNA minus αλυσίδα μέσω των PBS περιοχών και ολοκληρώνεται η σύνθεση του δίκλωνου cDNA μορίου. Στο σημείο αυτό το cDNA μόριο είναι έτοιμο να μεταφερθεί στον πυρήνα και να ενσωματωθεί σε μία νέα θέση στο γονιδίωμα του ξενιστή, με τις δύο αυτές διαδικασίες να καταλύονται από τις ιντεγκράσες (69).



Εικόνα 2.3 Μηχανισμός μετάθεσης ενός LTR ρετροτρανσποζονίου (68)

Τα LTR ρετροτρανσποζόνια μπορούν να διαχωριστούν σε αυτόνομα και μη αυτόνομα ως προς τη δυνατότητα μετακίνησής τους. Απαραίτητη προϋπόθεση για τη

μετάθεση των μη αυτόνομων LTR ρετροτρανσποζονίων από άλλα στοιχεία είναι η ύπαρξη των πλευρικών LTR αλληλουχιών καθώς και PBS περιοχές έτσι ώστε να καθίσταται εφικτή η εκκίνηση της αντίστροφης μεταγραφάσης.

Αξίζει να σημειωθεί ότι στα σπονδυλωτά και συγκεκριμένα στα θηλαστικά, η πλειονότητα των LTR ρετροτρανσποζονίων που εδράζονται στο γονιδίωμα των οργανισμών αυτών, είναι αποτέλεσμα της εγκόλπωσης ρετροϊών και της μετέπειτα απενεργοποίησης ή διαγραφής των περιοχών που εξασφάλιζαν την εξωκυττάρια μετακίνηση αυτών. Στα θηλαστικά συναντώνται τρεις κλάσεις (I, II, III) των ενδογενών ρετροϊών (ERVs, Endogenous Retroviruses) με τους οποίους συνεξελίχθηκαν τα LTR ρετροτρανσποζόνια, η κατηγοριοποίηση των οποίων έγινε βάσει της ομοιότητας τους με τους εξωγενείς ρετροϊούς (70-73).

2.1.2.1.2 ΡΕΤΡΟΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΜΕ ΜΑΚΡΙΕΣ ΤΕΡΜΑΤΙΚΕΣ ΕΠΑΝΑΛΗΨΕΙΣ (LTRS) ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ.

Η ύπαρξη των ενδογενών ρετροϊών είναι αποτέλεσμα ρετροϊικών λοιμώξεων που προσβάλλουν κύτταρα της βλαστικής σειράς και ακολουθεί η επακόλουθη ενσωμάτωσή τους στο γονιδίωμα του ξενιστή. Το ανθρώπινο γονιδίωμα εκτιμάται να περιέχει δεκάδες χιλιάδες αντίγραφα ανθρώπινων ενδογενών ρετροϊών (HERVs) που αντιπροσωπεύουν περίπου το 9% του ανθρώπινου γονιδιώματος. Τα περισσότερα HERVs στοιχεία φαίνεται να έχουν εισέλθει στο γονιδίωμα των προγόνων μας, 35-40 εκατομμύρια χρόνια πριν. Λόγω της σχετικά μακράς παραμονής τους στο ανθρώπινο γονιδίωμα, η πλειονότητα των HERVs είναι γεμάτη με επιβλαβείς μεταλλάξεις, μεγάλες αφαιρέσεις γονιδιώματος και με ενθέσεις διαφόρων επαναλλαμβανόμενων στοιχείων. Η δομή των HERVs κυρίως έχει την μορφή 5' LTR-gag-pro-pol-env-3'-LTR, όπου τα 4 γονίδια (gag:group-specific antigen, pro:protease, pol:polymerase και env:envelope)

κωδικοποιούν δομικές/λειτουργικές πρωτεΐνες απαραίτητες για την αντιγραφή του ρετροϊού, και οι μακριές επαναλαμβανόμενες επαναλήψεις (LTRs) διαφοροποιούν τους ρετροϊούς από άλλα επαναλαμβανόμενα στοιχεία όπως τα LINEs. Τα περισσότερα υπολείμματα ρετροϊών υπάρχουν στην μορφή απομωνομένων LTRs αντιγράφων, με την ενδιάμεση ακολουθία να έχει χαθεί κατά την ενσωμάτωση των ίδιων στο γονιδίωμα ή κατά τη διαδικασία ομόλογου ανασυνδυασμού (74).

2.1.2.2 NON-LTR ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΑ

Υπενθυμίζεται ότι στην ομάδα των Non-LTR ρετροτρανσποζονίων κατατάσσονται τα ρετρομεταθετά, τα οποία δε διαθέτουν στα άκρα τους τις LTR αλληλουχίες και τα οποία αποτελούνται από τα LINEs, τα SINEs και SVAs στοιχεία. Οι τρεις αυτές υποομάδες κατηγοριοποιούνται εκ νέου ανάλογα με τη δυνατότητα εγγενούς μετακίνησης μέσα στο γονιδίωμα, με τα LINEs στοιχεία να είναι αυτόνομα, ενώ τα SINEs και SVAs στοιχεία να εξαρτώνται από τα ένζυμα των πρώτων για τη μετάθεσή τους(75-77).

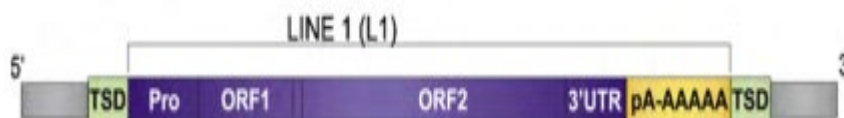
2.1.2.2.1 LINES (LONG INTERSPERSED ELEMENTS)

Τα LINEs στοιχεία συναντώνται σε όλους τους ευκαρυώτες και διαχωρίζονται σε πέντε υπερικογένειες, R2, RTE, Jockey, L1 και I, ανάλογα με τη φυλογένεια της περιοχής αντίστροφης μεταγραφάσης, RT (Reverse Transcriptase) που διαθέτουν. Τα LINEs στο σύνολο τους κωδικοποιούν τουλάχιστον την πρωτεΐνη ORF2, η οποία περιέχει μία περιοχή αντίστροφης μεταγραφάσης RT και μια περιοχή ενδονουκλεάσης EN (Endonuclease), ενώ αν εξαιρεθούν οι εξελικτικά αρχαίες υπερικογένειες των R2 και RTE, τα υπόλοιπα LINEs κωδικοποιούν και μία επιπλέον πρωτεΐνη, την ORF1. Η L1 υπερικογένεια έχει κατακλύσει το γονιδίωμα των σπονδυλωτών και κυριότερα των

θηλαστικών, καταλαμβάνοντας τουλάχιστον το 17% του συνολικού γονιδιώματος, ενώ η δραστηριότητά τους σε αυτό χρονολογείται τουλάχιστον 160 εκατομμύρια χρόνια πριν (78, 79).

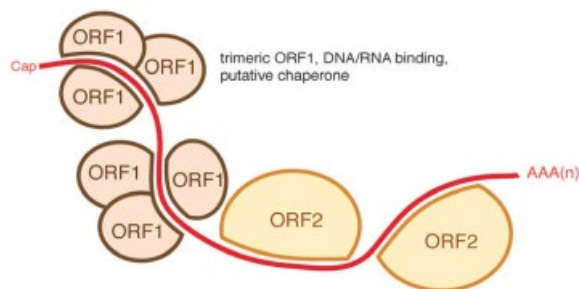
Ένα πλήρες σε μέγεθος L1 ρετροτρανσποζόνιο αποτελείται από 6 kb και περιέχει εσωτερικά έναν υποκινητή RNA πολυμεράσης II στην 5' αμετάφραστη περιοχή UTR (Untranslated Region), δύο ανοιχτά πλαίσια ανάγνωσης (ORFs, Open Reading Frames) και τέλος μια 3' αμετάφραστη περιοχή UTR, η οποία απαρτίζεται από ένα πολύ-αδενυλικό σήμα ακολουθούμενο από μια πολύ-αδενυλική ουρά (polyA-tail). Τέλος, τα άκρα των L1 περιστοιχίζονται από TSDs (Tandem Site Duplications) περιοχές. Η ORF1 πρωτεΐνη που κωδικοποιούν τα L1 ρετροτρανσποζόνια, έχει μέγεθος 40 kDa, μία συντηρημένη καρβοξυτελική περιοχή (C-terminus) και μία υπερ-ελικωμένη αμινοτελική περιοχή (N-terminus), η οποία ευθύνεται για το σχηματισμό των τριμερών ORF1 συμπλόκων. Η λειτουργία των τριμερών ORF1 συμπλόκων σχετίζεται με τη ρετρομετάθεση μέσω της περιοχής πρόσδεσης RNA που διαθέτουν, καθώς και από την ενεργότητα μεταφοράς του DNA στον πυρήνα της πρωτεΐνης. Όπως αναφέρθηκε προηγουμένως, η ORF2 πρωτεΐνη διαθέτει ενεργότητα αντίστροφης μεταγραφάσης και ενδονουκλεάσης (80-82).

Η cis-αλληλεπίδραση ενός αγνώστου αριθμού τριμερών ORF1 και δύο ORF2 πρωτεϊνών μαζί με τα mRNA τους σχηματίζει το ριβονουκλεοπρωτεϊνικό RNP (ribonucleoprotein) σύμπλοκο, το οποίο κατόπιν μεταφέρεται στον πυρήνα.



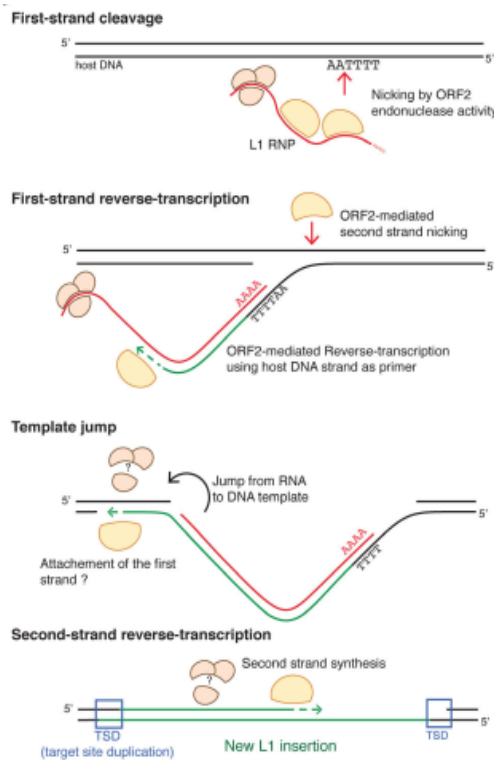
Εικόνα 2.4. Δομή ενός LINE (L1) ρετροτρανσποζονίου (83)

L1 ribonucleoprotein complex



Εικόνα 2.5. Μορφή του L1 ριβονουκλεοπρωτεϊνικού συμπλόκου (83)

Σύμφωνα με τον μηχανισμό στοχευμένης αντίστροφης μεταγραφής “TPRT” (target-primed reverse transcription) των L1 ρετροτρανσποζονίων, το L1 RNP σύμπλοκο αναγνωρίζει το συμπληρωματικό ως προς αυτό εξανουκλεοτίδιο, με κατεύθυνση 5’ ΤΠΤΤΑΑ 3’, του DNA του ξενιστή και η ORF2 πρωτεΐνη με την ενεργότητα ενδονουκλεάσης κόβει το εξανουκλεοτίδιο ανάμεσα στη T (θυμίνη) και την A (αδενίνη) με φορά 5’ προς 3’. Η poly-A ουρά του mRNA του L1 προσδένεται στο 3’ προεξέχον άκρο της minus αλυσίδας του DNA του ξενιστή, το οποίο χρησιμοποιείται ως εκκινήτης για την αντίστροφη μεταγραφή μέσω της ενεργότητας αντίστροφης μεταγραφάσης της ORF2 πρωτεΐνης. Πιθανόν, η ORF2 ενδονουκλεάση να συμμετέχει και στη θραύση της δεύτερης (plus) αλυσίδας DNA του ξενιστή και να κόβει πάλι με την ίδια κατεύθυνση το εξανουκλεοτίδιο, χρησιμοποιώντας ξανά την αλυσίδα του ξενιστή ως εκκινήτη για την αντίστροφη μεταγραφή του mRNA του ρετροτρανσποζονίου. Ωστόσο, τα στάδια του μηχανισμού, που διαμεσολαβούν τη δεύτερη θραύση της αλυσίδας του ξενιστή, την αντίστροφη μεταγραφή και τη σύνδεση των άκρων του ρετροτρανσποζονίου με το DNA του δότη, δεν έχουν διευκρινιστεί πλήρως. Η ένθεση του L1 ολοκληρώνεται με τη σύνθεση των TSDs αλληλουχιών στα άκρα των 5’ και 3’ UTR του. Οι περισσότερες από τις ενθέσεις των L1 βρίσκονται ανεστραμμένες με κομμένες τις 5’ UTR περιοχές τους με αποτέλεσμα να καθίστανται μη λειτουργικές (80, 84).



Εικόνα 2.6 Στάδια μηχανισμού ένθεσης “TPRT” (target-primed reverse transcription) των L1 ρετροτρανσποζονίων στο DNA του δότη (85)

2.1.2.2 SINES (SHORT INTERSPERSED ELEMENTS)

Τα SINEs ρετροτρανσποζόνια αποτελούν τα μη αυτόνομα μεταθετά στοιχεία της οικογένειας των non-LTR ρετροτρανσποζονίων με μικρό μήκος που κυμαίνεται μεταξύ των 80-500pb για τα ποντίκια, ενώ το στοιχείο Alu στους ανθρώπους φτάνει έως και τα 300bp σε μέγεθος. Το στοιχείο Alu, χαρακτηριστικά, παρουσιάζει μια προτίμηση για μετάθεση κατά κόρον στο γονιδίωμα των τάξεων των πρωτευόντων, ενώ φαίνεται πως χρονολογεί 65 εκατομμύρια χρόνια ύπαρξης. Λόγω της εγγενούς αδυναμίας τους για την αντιγραφή του γενετικού τους υλικού φαίνεται πως εξελίχθηκαν ανεξάρτητα από τα LINEs στοιχεία, τους μηχανισμούς των οποίων όμως εκμεταλλεύονται για το σκοπό αυτό. Παρόλα αυτά, ο αριθμός των αντιγράφων τους καλύπτει περίπου έως και το 15% του γονιδιώματος των θηλαστικών. Όσον αφορά τη μεταγραφή, τα SINEs στηρίζονται στην ύπαρξη ενός υποκινητή για την πρόσδεση της RNA πολυμεράσης III ανοδικά του 5' άκρου τους. Η τυχαία ρετρομετάθεση ορισμένων μεταγράφων της RNA πολυμεράσης

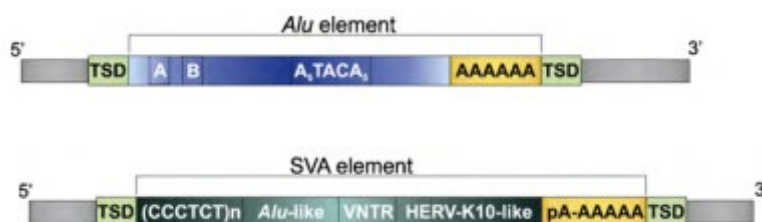
III κατά τακτά χρονικά διαστήματα στην πορεία της εξέλιξης ευθύνεται για την προέλευση των SINEs στοιχείων. Πιο συγκεκριμένα, οι υπεροικογένειες των SINEs μεταξύ τους διαφοροποιούνται ανάλογα με τον τύπο προέλευσης (από tRNA, ριβοσωμικό 5S RNA ή από το σωματίδιο αναγνώρισης σήματος 7SL RNA) του υποκινητή της πολυμεράσης III τον οποίο διαθέτουν. Για παράδειγμα, το στοιχείο Alu, το οποίο αριθμεί 1.1 εκατομμύρια αντίγραφα στο ανθρώπινο γονιδίωμα διαθέτει υποκινητή 7SL τύπου. Αναφορικά με τις 3' περιοχές των SINEs στοιχείων παρατηρείται μεγάλη ποικιλομορφία, καθώς ορισμένες αλληλουχίες του 3' άκρου αποτελούν θέσεις αναγνώρισης για τις πρωτεΐνες των LINEs στοιχείων, ενώ άλλες αποτελούνται από περιοχές πλούσιες σε αδενίνες (A) ή σε επαναλήψεις των βάσεων αδενίνης-θυμίνης (AT). Ακόμα έχουν καταγραφεί και επαναλήψεις θυμίνης (T) (poly-T tail) στο 3' άκρο, οι οποίες αποτελούν το σήμα τερματισμού για την πολυμεράση III.

2.1.2.2.3 ΔΟΜΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ALU ΚΑΙ SVA ΤΗΣ ΥΠΕΡΟΙΚΟΓΕΝΕΙΑΣ ΤΩΝ SINES ΡΕΤΡΟΤΡΑΝΣΠΟΖΟΝΙΩΝ

Ο κοινός παρονομαστής όσον αφορά τη δομή των Alu όσο και των SVA (SINE-R, VNTR και Alu) στοιχείων αποτελεί το γεγονός ότι περιστοιχίζονται από τις TSDs αλληλουχίες στα 5' και 3' άκρα τους. Το Alu στοιχείο από το 5' προς το 3' άκρο αποτελείται από τα A και B κουτιά του υποκινητή της RNA πολυμεράσης III (RNA pol III promoter boxes), ακολούθως τη θέση καταλαμβάνει η A₅TACA₅ αλληλουχία, ενώ προς τη μεριά του 3' άκρου, το Alu καταλήγει σε μια πολυαδενυλική ουρά (poly(A) tail).

Αντίστοιχα για το στοιχείο SVA η δομή του εκκινώντας από το 5' προς το 3' άκρο είναι ως εξής: Αρχικά το SVA στοιχείο απαρτίζεται από πολλαπλές διαδοχικές επαναλήψεις του CCCTCT εξαμερούς, ενώ έπειτα ακολουθεί μία ομόλογη ως προς το Alu στοιχείο περιοχή που αποτελείται από δύο αντισημαινόντα θραύσματα, τα οποία ενδεχομένως να περιλαμβάνουν και αλληλουχίες αγνώστου προελεύσεως. Στη συνέχεια, συναντάται μια περιοχή ποικίλου αριθμού τυχαίων επαναλήψεων VNTR

(variable number of tandem repeats) και η δομή του SVA ολοκληρώνεται με την ύπαρξη μιας SINE περιοχής, η οποία αποτελείται από τμήματα του ενδογενούς ρετροϊού, HERV-K10, που παρουσιάζεται στα κύτταρα του ανθρώπου. Το μέγεθος των στοιχείων SVA είναι μικρότερο των 3kb με τα αντίγραφα του να αριθμούνται γύρω στα 2500 με 3000 στο ανθρώπινο γονιδίωμα, όπου εμφανίζουν μεγάλη ετερογένεια και μετατίθενται ακόμη και σήμερα χρησιμοποιώντας κυρίως το μηχανισμό των L1 ρετροτρανσποζονίων (86-91).



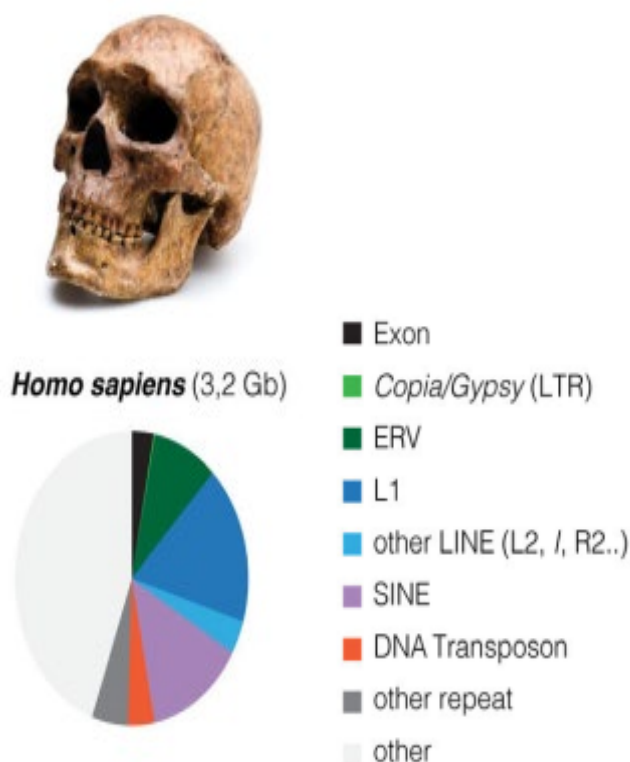
Εικόνα 2.7. Δομή των στοιχείων Alu και SVA (πάνω και κάτω αντίστοιχα) (92)

2.1.3 ΔΙΑΣΠΟΡΑ ΜΕΤΑΘΕΤΩΝ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΙΚΑΝΟΤΗΤΑ ΜΕΤΑΚΙΝΗΣΗΣ ΣΤΟ ΓΟΝΙΔΙΩΜΑ ΤΩΝ ΘΗΛΑΣΤΙΚΩΝ

Αν και υπολογίζεται ότι το ανθρώπινο γονιδίωμα αποτελείται κατά το ήμισυ από μεταθετά στοιχεία, στην πραγματικότητα πιθανόν να κατακλύζεται από πολλά περισσότερα μιας και η αναγνώριση και ταυτοποίηση αρχαίων μεταθετών στοιχείων φαίνεται να αποτελεί μία ιδιαίτερα περίπλοκη διαδικασία. Παρόλη την εξαιρετικά μεγάλη εξάπλωση και τα πολυάριθμα αντίγραφα που εμφανίζουν μέσα στο γονιδίωμα, τα περισσότερα μεταθετά στοιχεία βρίσκονται κατακερματισμένα, ενώ οι μεταλλάξεις που έχουν συσσωρεύσει στην πορεία της εξέλιξης, τα έχουν καταστήσει ανενεργά με ταυτόχρονη αδυναμία μετακίνησης. Τέτοιου είδους παράδειγμα αποτελεί η οικογένεια των LINEs, τα οποία στο γονιδίωμα θηλαστικών εμφανίζονται σε μεγάλο ποσοστό

κομμένα στο 5' άκρο τους. Ενδεικτικά, η υποοικογένεια LINE1 αν και διαθέτει περίπου 500.000 αντίγραφα στο ανθρώπινο γονιδίωμα, μόνο τα 5-7.000 διατηρούν ολόκληρο το μήκος τους, εκ των οποίων τα 80-100 είναι ικανά για μετακίνηση, ενώ ένας πυρήνας 6 μόνο LINE1 ("hot L1s") ευθύνεται για την πλειονότητα των μεταθέσεων στον ανθρώπινο πληθυσμό.

Εκτός από τα "hot L1s" στοιχεία τα οποία βρίσκονται ενεργά μέσα στο ανθρώπινο γονιδίωμα, ιδιαίτερη εξάπλωση με δυνατότητα μετακίνησης διαθέτει η υποοικογένεια Alu (SINEs), ενώ τα LTR ρετροτρανσποζόνια, αν εξαιρεθεί ο ρετροϊός HERVK-HML2, ο οποίος εκφράζεται σε διάφορες ασθένειες ή ορισμένους τύπους καρκίνου, κινδυνεύουν να εξαφανιστούν. Χαρακτηριστικά, τα μεταθετά στοιχεία ευθύνονται μόλις για το 0,3% των μεταλλάξεων στο ανθρώπινο γονιδίωμα, οι περισσότερες από τις οποίες αποτελούν ενθέσεις των L1, Alu και SVA, και σχετίζονται με ασθένειες.



Εικόνα 2.8. Σύνθεση ανθρώπινου γονιδιώματος από μεταθετά στοιχεία ή άλλες επαναλαμβανόμενες αλληλουχίες (τα αποτελέσματα αντλήθηκαν από τη βάση δεδομένων του προγράμματος Repeat Masker) (93)

2.1.3.1 ΣΤΟΙΧΕΙΑ ΣΥΣΧΕΤΙΣΗΣ HOLLIDAY JUNCTIONS ΜΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

Ένας από τους γρίφους της βιολογίας των ρετροτρανσποζονίων είναι η διευκρίνιση της σχέσης των ίδιων με τα Holliday Junctions. Εκτός από το να επηρεάζουν την έκφραση και την λειτουργικότητα των γονιδίων μέσω παρεμβολών, ο μεγάλος αριθμός στοιχείων Alu και σε μικρότερο βαθμό τα στοιχεία L1, φαίνεται να προωθούν συμβάντα ομόλογου ανασυνδυασμού. Υπάρχουν πολλές αποδεδειγμένες περιπτώσεις ομόλογου ανασυνδυασμού μεταξύ των στοιχείων Alu, και μόνο 3 καταγεγραμμένες περιπτώσεις ομόλογου ανασυνδυασμού μεταξύ των στοιχείων L1. Η ύπαρξη υψηλότερης συχνότητας των Alu ανασυνδυασμών μπορεί να προκύπτει από την ύπαρξη περισσότερων αντιγράφων των ίδιων στο γονιδίωμα, την τάση των Alus να αποικούν σε περιοχές πλούσιες σε γουανίνες και κυτοσίνες καθώς επίσης και για την περιεκτικότητα σε αλληλουχίες που προωθούν τον ανασυνδυασμό (74, 84).

Ένα χαρακτηριστικό παράδειγμα του πώς τα TEs μπορούν να συμμετέχουν σε γεγονότα ανασυνδυασμού αποδεικνύεται στον γονιδιακό τόπο που ευθύνεται για την ασθένεια Charcot-Marie-Tooth. Η ασθένεια CMT αποτελεί μια κληρονομική νευροπάθεια, και είναι αποτέλεσμα ενός διπλασιασμού και μίας διαγραφής, αντίστοιχα, επαναλαμβανόμενων αλληλουχιών που πλευρίζουν το γονίδιο CMT. Το σημείο της ανταλλαγής εντοπίστηκε στο 75% των ασθενών σε μια περιοχή μήκους 1,7 kb και αποτελεί σημείο που επάγει τον ανασυνδυασμό (hot spot). Η διπλασιασμένη περιοχή φαίνεται να περιέχει εκτός από αλληλουχίες που επάγουν τον ανασυνδυασμό και ένα αδρανοποιημένο ρετρομεταθετό στοιχείο, με αποτέλεσμα ο Reiner και οι συνεργάτες του να συμπεραίνουν εν τέλει ότι το αδρανοποιημένο μεταθετό στοιχείο είναι αυτό που προκαλεί τον ανασυνδυασμό στο σημείο αυτό (74, 84).

Ένα άλλο σημαντικό στοιχείο είναι ότι, πρόσφατα αποδείχθηκε ότι η εισαγωγή L1 στοιχείων σε καλλιέργειες κυττάρων θηλαστικών έχει ως αποτέλεσμα την δημιουργία

δίκλωνων θραύσεων (DSBs) και αυτό οφείλεται στην ενεργότητα της ενδονουκλεάσης που κωδικοποιείται στο ORF2 των L1. Ο αριθμός των DSBs που προκαλούνται σε πειραματικές συνθήκες (in vitro) είναι μεγαλύτερος από τον αριθμό των DSBs που προκαλούνται από ενθέσεις L1 σε πραγματικές συνθήκες (in vivo). Σίγουρα αυτή η διαφορά προσδίδει αβεβαιότητα στο ερώτημα κατά πόσο οι DSBs επαγόμενες από L1 συμβάλλουν στην πλαστικότητα του ανθρώπινου γονιδιώματος. Ωστόσο, η επιδιόρθωση των δίκλωνων θραύσεων που επάγονται από τα στοιχεία L1 , υποδεικνύουν τη συμμετοχή των προαναφερθέντων στοιχείων σε αυτή τη διαδικασία , έτσι ένα σημαντικό κλάσμα των μεταβολών που υφίσταται το ανθρώπινο γονιδίωμα που σχετίζονται με την δημιουργία DSBs , οι οποίες εμφανίζουν υψηλή μεταλλαξιγόνο και ανασυνδυαστική ιδιότητα, οφείλονται στην δράση των στοιχείων L1 (74, 84).

Εν κατακλείδι, τα μεταθετά στοιχεία και συγκεκριμένα τα ρετροτρανσποζόνια μπορούν να αποδειχθούν χρήσιμα εργαλεία παρεμβατικής μεταλλαξιγένεσης. Επίσης επειδή η ρετρομετάθεση είναι αντιγραφική , μπορούν να συμβάλλουν στην καταπολέμηση διάφορων μορφών καρκίνου, στις οποίες απαιτείται η συσσώρευση διαδοχικών μεταλλάξεων σε καρκινικά κύτταρα. Έτσι αν και έχουν παρασιτικό χαρακτήρα στη φύση , οι αλληλουχίες τους συμμετεχουν με αρκετούς τρόπους στην ρύθμιση της κυτταρικής λειτουργίας. Στο μέλλον, η συμβολή τους στην βελτίωση της ανθρώπινης ζωής μπορεί να επεκταθεί στην χρήση τους ως φορείς κρίσιμων αλληλουχιών σε γονιδιώματα ξενιστών (74, 84).

ΚΕΦΑΛΑΙΟ 3

ΓΕΝΕΤΙΚΗ ΑΣΤΑΘΕΙΑ

3.1 ΧΡΩΜΟΣΩΜΙΚΕΣ ΑΝΑΔΙΑΤΑΞΕΙΣ ΜΕ ΧΑΡΑΚΤΗΡΗΣΤΙΚΑ ΣΥΝΔΡΟΜΩΝ.

Το τελικό κομμάτι του θεωρητικού μέρους της παρούσας διατριβής αποσκοπεί στη μελέτη και κατανόηση των χρωμοσωματικών αναδιατάξεων με χαρακτηριστικά συνδρόμων, που συνδέονται κυρίως είτε με υπερηχογραφικά ευρήματα στα έμβρυα είτε με συγγενείς ανωμαλίες στα νεογνά είτε νοητική υστέρηση και καθυστέρηση της ανάπτυξης στα βρέφη και παιδιά, γνωστά και ως σύνδρομα μικροελλειμμάτων και μικροδιπλασιασμών (MMS). Αυτές οι αναδιατάξεις είναι στην ουσία οι υπο-μικροσκοπικές παραλλαγές στον αριθμό των αντιγράφων ενός ή πολλών τμημάτων DNA (Copy Number Variations, CNVs), τα οποία έχουν ως αποτέλεσμα τη δημιουργία ελλειμμάτων ή διπλασιασμών στα χρωμοσώματα σε περιοχές με ένα ή περισσότερα γονίδια. Είναι στην ουσία δομικοί γενωμικοί δείκτες, οι οποίοι ποικίλλουν σε μέγεθος από 1 kb έως πολλά Mbs και χαρακτηρίζονται είτε ως φυσιολογικοί πολυμορφισμοί πολλών αντιγράφων (Copy Number Polymorphisms CNPs) είτε ως σπάνιοι πολυμορφισμοί υπεύθυνοι για πολλά γενετικά σύνδρομα (94-99). Στην μεγάλη πλειοψηφία των περιπτώσεων, τα CNVs προκύπτουν ως αποτέλεσμα πολύπλοκων μηχανισμών με ανασυνδυαστική δράση. Οι δύο μεγαλύτερες κατηγορίες των CNVs είναι :

α) τα επαναλαμβανόμενα (recurrent) CNVs, που είναι ανακατατάξεις στις οποίες τα σημεία των θραύσεων (breakpoints) είναι κοινά μεταξύ πασχόντων ατόμων που δεν συνδέονται με κάποια συγγενική σχέση και

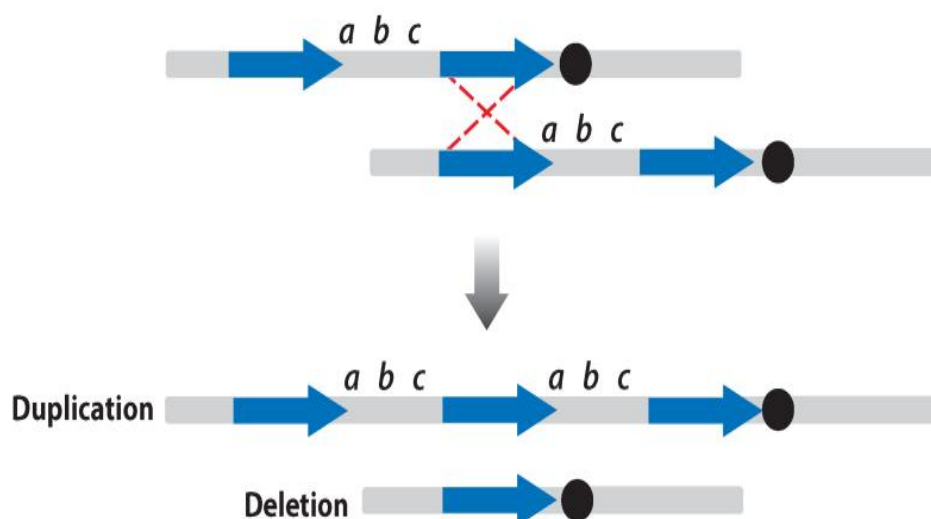
β) τα μη-επαναλαμβανόμενα CNVs (non-recurrent) , δηλαδή ανακατατάξεις όπου τα σημεία των θραύσεων δεν είναι κοινά μεταξύ των ατόμων που έχουν το γενετικό νόσημα (100, 101). Έχουν προταθεί τρεις κύριοι μηχανισμοί υπεύθυνοι για τη δημιουργία των CNVs στο ανθρώπινο γένωμα:

- Ομόλογος Ανασυνδυασμός (Homologous Recombination Mechanism, HRM) και κατά κύριο λόγο Μη- Αλληλόμορφος Ομόλογος Ανασυνδυασμός (Non-Allelic Homologous Recombination, NAHR) ,
- Μη Ομόλογος Ανασυνδυασμός (Non-homologous recombination mechanism (NHRM)) και κυρίως ο μη ομόλογος ανασυνδυασμός κατά τη διαδικασία της επιδιόρθωσης των θραύσεων του DNA (Non- homologous End joining, NHEJ) (101-103) και
- Ανασυνδυασμός διαμεσολαβούμενος από αντιγραφικούς μηχανισμούς (replication based mechanisms (RBM)) με κύριο τον μηχανισμό που χαρακτηρίζεται ως Fork Stalling and Template Switching/mediated break induced replication FoSTeS/MMBIR.

Η ίδια η φύση των CNVs, τα καθιστά επιρρεπή στους ανασυνδυασμούς. Η υψηλή περιεκτικότητα επαναλαμβανόμενων αλληλουχιών μέσα στα CNVs είναι η αιτία που:

α) τα γεγονότα NAHR είναι υπεύθυνα για την δημιουργία επαναλαμβανόμενων (recurrent) ανακατατάξεων (με ίδιο μέγεθος και ίδια breakpoints) που συμβαίνουν κατά κύριο λόγο κατά τη μείωση και σπανιότερα σε σωματικά κύτταρα όπου δημιουργείται το φαινόμενο του μωσαϊκισμού

β) τα γεγονότα NHEJ που είναι υπεύθυνα για τη δημιουργία μη-επαναλαμβανόμενων (non-recurrent) ανακατατάξεων (με διαφορετικό μέγεθος και διαφορετικά breakpoints) και προκύπτουν κυρίως από λάθη στην επιδιόρθωση των διπλών θραύσεων του DNA από μηχανισμούς μη ομόλογης επιδιόρθωσης (101-103).



Εικόνα 3.1. Μη αλληλικός ομόλογος ανασυνδυασμός (NAHR), ο κύριος μηχανισμός που αποτελεί τη βάση για την παραγωγή παραλλαγών του αριθμού αντιγράφων (CNVs). Οι τμηματικές επαναλήψεις (segmental duplications), που ονομάζονται επίσης επαναλήψεις χαμηλών αντιγράφων, αντιπροσωπεύουν περιοχές με εκτεταμένη ομολογία αλληλουχίας που μπορούν να παρέχουν ένα υπόστρωμα για χρωμοσωμικές αναδιατάξεις με τη μεσολάβηση του NAHR. Σε αυτό το σχήμα, δύο μεγάλες τμηματικές επαναλήψεις (μπλε βέλη) με υψηλή ομοιότητα αλληλουχίας πλαισιώνουν μια περιοχή που περιέχει τα γονίδια a, b και c. Μετά την εσφαλμένη ευθυγράμμιση των ομόλογων, αυτές οι επαναλήψεις τμημάτων DNA διευκολύνουν μέσω του NAHR τη δημιουργία 2 προϊόντων: ένα χρωμόσωμα με ένα επιπλέον αντίγραφο των γονιδίων a, b και c και ένα δεύτερο χρωμόσωμα που φέρει μια διαγραφή της ίδιας περιοχής. Τέτοιες αναδιατάξεις είναι κοινές αιτίες πολλών επαναλαμβανόμενων γενωμικών διαταραχών που χαρακτηρίζονται από αμοιβαίες αναδιατάξεις συγκεκριμένων χρωμοσωμικών περιοχών (104)

3.1.1 CNVS (COPY NUMBER VARIATIONS) ΚΑΙ LCR (LOW COPY REPEATS).

Τα LCRs είναι ομόλογες αλληλουχίες, μεγαλύτερες της μιας κιλοβάσης, οι οποίες βρίσκονται σε πολλά αντίγραφα στο γονιδίωμα και είναι το αποτέλεσμα γεγονότων διπλασιασμού. Ορισμένες ιδιότητες των LCRs όπως το μέγεθος και το είδος των ομόλογων αλληλουχιών, επηρεάζουν την συχνότητα των γεγονότων NAHR. Όπως αποδείχθηκε από τους Carvalho και Lupski, η συχνότητα των γεγονότων NAHR

εξαρτάται από το μήκος της αλληλουχίας LCR, την απόσταση μεταξύ κοντινών LCRs καθώς και την ίδια την ακολουθία της αλληλουχίας νουκλεοτιδίων που τα περιέχει. Προκύπτει ότι η συχνότητα των NAHR είναι ανάλογη του μήκους της LCR ακολουθίας και αντιστρόφως ανάλογη της απόστασης μεταξύ κοντινών LCRs. Υπάρχει ένας συγκριτικά μεγάλος αριθμός από CNVs τα οποία δημιουργούνται από LCRs και σχετίζονται άμεσα με σύνδρομα. Για παράδειγμα, ο διπλασιασμός ενός σύνθετου συμπλέγματος αλληλουχιών LCR που φθάνει σε συνολικό μέγεθος 4 Mb και επεκτείνεται γύρω από το MECP2 γονίδιο, έχει συσχετισθεί με αναπτυξιακές και άλλες διαταραχές στους άρρενες (MECP2 duplication syndrome). Άλλα χαρακτηριστικά παραδείγματα είναι δυο από τα πιο κοινά γενετικά νοσήματα, τα σύνδρομα Angelman (AS) και Prader – Willi (PWS), τα οποία προκύπτουν από ελλείμματα τμημάτων μεγέθους 4 Mb στον γενετικό τόπο 15q11-q13. Παρατηρήθηκε ότι σε αυτά τα σύνδρομα, η ελλειμματική περιοχή πλαισιώνεται από ακολουθίες LCR, οι οποίες ευθύνονται για το 70% των περιπτώσεων του AS και το 75 % του PWS, ανάλογα με το γονέα από τον οποίο προέρχονται (102, 105-107).

3.1.2 CNVS ΚΑΙ ΡΕΤΡΟΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ.

Οι Braun και συνεργάτες παρατήρησαν ότι οι περιοχές που γίνονται οι θραύσεις των CNVs που προκαλούν την εκδήλωση του συνδρόμου Reutz-Jeghers και συγκεκριμένα η περιοχή του γονιδίου SKT11, είναι εμπλουτισμένη με ρετρομεταθετά στοιχεία της οικογένειας ALUs (που καταλαμβάνουν το 11% του γενώματος, ενώ όλες οι οικογένειες ρετρομεταθετών στοιχείων καταλαμβάνουν μαζί σχεδόν το 45% ολόκληρου του ανθρώπινου γενώματος) (108).

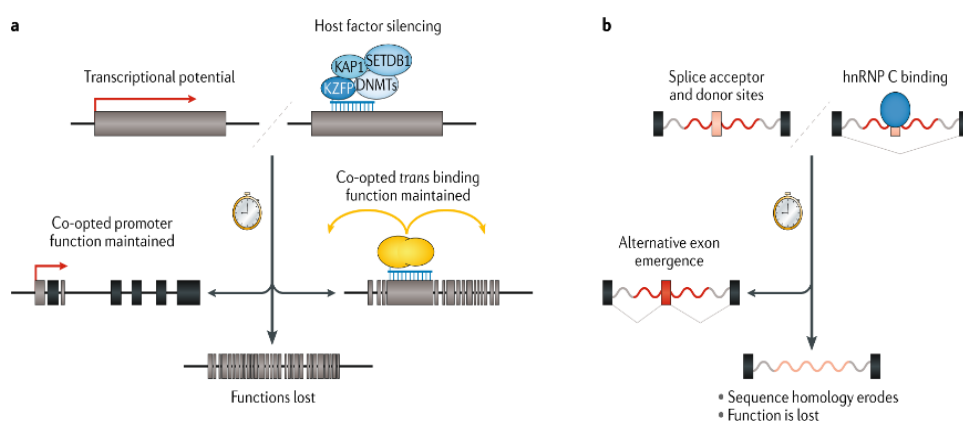
Επίσης στο γενετικό τόπο 17p13.3 παρατηρήθηκε πληθώρα γεγονότων CNVs τα οποία οδηγούν σε σημαντικά γενετικά σύνδρομα, όπως είναι το Mieler- Diecker. Τα σημεία θραύσης είναι πλούσια σε ALUs , γεγονός που τα κάνει ευάλωτα σε μεταξύ τους ανασυνδυασμό, λόγω υψηλής ομολογίας (109). Επίσης , περίπου το 70% των CNVs που βρέθηκαν στο γονίδιο SPAST της σπαστικής παραπληγίας έχουν συσχετιστεί με

ανασυνδυασμούς μεταξύ των στοιχείων ALU . Η πλούσια σε ALUs αρχιτεκτονική στην προαναφερθείσα περιοχή, προδιαθέτει στη δημιουργία παθογόνων αναδιατάξεων οι οποίες σχετίζονται άμεσα με την σπαστική παραπληγία (110).

Χαρακτηριστικό επίσης είναι και το παράδειγμα όπου ανασυνδυασμός NAHR μεταξύ συγκεκριμένων ρετρομεταθετών στοιχείων της οικογένειας HERV (που αποτελούν το 8% του γενώματος), στον γενετικό τόπο AZFa του χρωμοσώματος Y, οδηγεί στη δημιουργία παθογόνων ελλειμμάτων που σχετίζονται με την ανδρική υπογονιμότητα (111, 112). Τα ρετρομεταθετά στοιχεία συμμετέχουν από κοινού με τα LCR στη δημιουργία των CNVs.

Σε αναλύσεις στα σημεία θραύσης πολλών CNVs σε γενετικά σύνδρομα, εντοπίζονται και LCRs και HCRs τα οποία επηρεάζουν τα υπεύθυνα γονίδια για τα MMS. Το σύνδρομο Williams – Beuren χαρακτηρίζεται από έλλειμμα ενός τμήματος DNA μήκους μεταξύ 1,6 και 1,8 Mb . Η διαγραφείσα περιοχή περιέχει ένα σύνολο 28 δοσοεξαρτώμενων γονιδίων με απλοανεπάρκεια και μεγάλες αλληλουχίες LCR οι οποίες προέκυψαν από ανασυνδυασμούς μεταξύ δύο στοιχείων ALU (113).

Στο πιο συχνό σύνδρομο μικροελλείμματος το DiGeorge , περίπου το 96% των περιπτώσεων φέρει έλλειμμα ενός τμήματος μήκους μεταξύ 1.5 και 1.7 Mb στον γενετικό τόπο 22q11.2. Στη διαγραφείσα περιοχή εντοπίζονται 24-30 γονίδια, ενώ στα σημεία θραύσης αναγνωρίστηκαν αφ'ενός η συγκεκριμένη αλληλουχία LCR22 και το μέλος της οικογένειας ALU το ALUY (114).



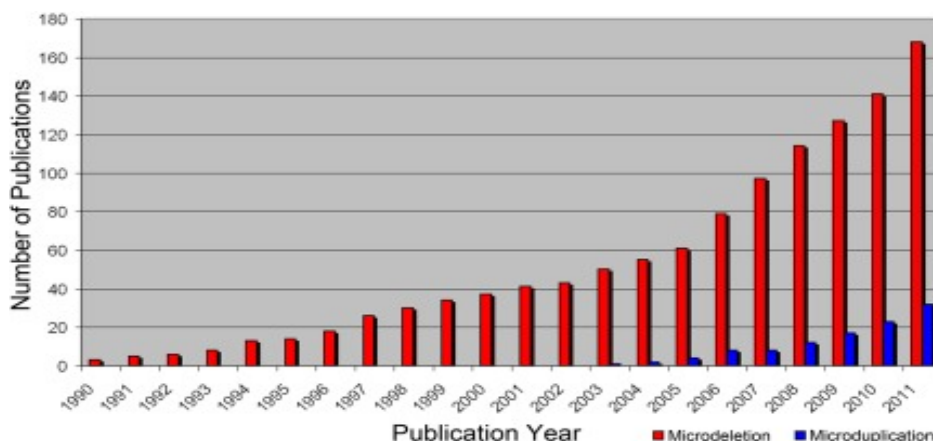
Εικόνα 3.2 Απεικόνιση της σχέσης ρετρομεταθετών στοιχείων και γενετικής αστάθειας (115)

3.1.3 ΠΑΘΟΓΕΝΕΙΑ ΤΩΝ CNVs

Διάφορα κριτήρια μπορούν να χρησιμοποιηθούν στη διαδικασία χαρακτηρισμού ενός CNV ως παθογόνου, σε ένα σύνδρομο. Τα πιο βασικά κριτήρια είναι η κληρονομικότητα, το μέγεθος και ο τύπος (έλλειμμα ή διπλασιασμός) της αναδιάταξης και το γονιδιακό της περιεχόμενο (116). Μια εκ νέου δημιουργία ενός CNV έχει περισσότερες πιθανότητες να είναι παθογόνος από ένα κληρονομούμενο CNV. Επειδή όμως στην κλινική πράξη, είναι πολλές φορές δύσκολο να εξετάσεις και τους δύο γονείς, για οικονομικούς αλλά και για κοινωνικούς λόγους, πρέπει να εισάγουμε άλλα κριτήρια για τον καθορισμό της παθογένειας τους, όπως είναι το μέγεθος και ο τύπος του CNV. Μεγάλα CNVs είναι πιο πιθανό να προκαλέσουν γενετικά νοσήματα σε σχέση με μικρότερα CNVs. Αυτό συμβαίνει γιατί, τα μεγάλα CNVs περιέχουν περισσότερα γονίδια (δοσο-εξαρτώμενο αποτέλεσμα στη μεταγραφή και μετάφραση). Οι ελλείψεις μπορεί επίσης να οδηγήσουν σε απλοανεπάρκεια, οι συνέπειες της οποίας είναι γνωστές μόνο για συγκεκριμένα γονίδια. Τους διπλασιασμούς είναι πιο δύσκολο να τους χαρακτηρίσουμε και να τους ερμηνεύσουμε καθώς έχουμε περιορισμό της εκδήλωσης των φαινοτυπικών τους χαρακτηριστικών. Τέλος το γονιδιακό περιεχόμενο πρέπει να ληφθεί υπόψη, διότι CNVs τα οποία περιέχουν πολλά γονίδια ή γονίδια τα οποία ευθύνονται για γνωστά νοσήματα, είναι πολύ περισσότερο παθογόνα από CNVs που περιέχουν λιγότερα γονίδια. CNVs που εντοπίζονται σε μη κωδικοποιούσες περιοχές είναι εξίσου δύσκολο να ερμηνευτούν και θα γίνει πιο κατανοητή η λειτουργία τους όταν αποκωδικοποιηθούν οι ρυθμιστικές περιοχές εντός μη-κωδικοποιητικού DNA το οποίο ενδεχομένως εκφράζει ρυθμιστικά μη-κωδικοποιητικά νουκλεϊκά οξέα (ncRNAs) (117-119)

3.1.4 «ΑΝΑΔΥΟΜΕΝΑ» MMS

Αν και πριν από το 2000 ήταν γνωστές μόνο μερικές δεκάδες MMS, όλο και περισσότερα εντοπίστηκαν με τη διαθεσιμότητα νέων τεχνολογιών για αναλύσεις υψηλής ανάλυσης ολόκληρων γονιδιωμάτων. Κατά τα τελευταία 10 χρόνια, υπήρξαν ραγδαίες εξελίξεις στις τεχνολογίες γονιδιώματος υψηλής απόδοσης. Η εκτεταμένη χρήση της ανάλυσης ολόκληρου του γονιδιώματος βασισμένη στο συγκριτικό γονιδιωματικό υβριδισμό (aCGH) και στη μαζική πολλαπλή αλληλούχιση (Massive Parallel Sequencing, NGS) σε διάγνωση και έρευνα έχει οδηγήσει σε έναν συνεχώς αυξανόμενο αριθμό συνδρόμων μικροελλείψεων και μικροδιπλασιασμών (MMΣ) που συνδέονται με συγκεκριμένους φαινοτύπους. Από το 1990 έως το 2011, δημοσιεύθηκαν 200 άρθρα σχετικά με νέα σύνδρομα (εικ. 3.3). Έχοντας υπ όψιν ότι μια κοινή αιτία για τη δημιουργία των συνδρόμων είναι η δομή του ανθρώπινου γονιδιώματος, η οποία έχει αμέτρητες επαναληπτικές αλληλουχίες του DNA που μπορούν να οδηγήσουν σε μη αλληλικό ομόλογο ανασυνδυασμό (NAHR), ο αναμενόμενος συνολικός αριθμός των MMΣ θα πρέπει να είναι πολύ μεγαλύτερος (101). Θεωρητικά, για κάθε σύνδρομο μικροελλείψης θα πρέπει να υπάρχει ένα αμοιβαίο σύνδρομο μικροδιπλασιασμού. Παρ' όλα αυτά, έχουν διαγνωστεί μέχρι σήμερα 212 σύνδρομα μικροελλείψεων έναντι μόνο 84 σύνδρομα μικροδιπλασιασμών που αναφέρθηκαν (Πίνακας 1,2 Εικόνα 3). Αυτή είναι μια αναλογία 2,5: 1 για ένα σύνολο 267 διαφορετικών γονιδιωματικών τόπων που εμφανίζουν σύνδρομα. Μόνο για 56 από αυτούς, οι τόποι αναφέρονται ως αμοιβαία MMΣ, δηλαδή 21%. Αυτό οφείλεται σε διάφορους λόγους. για παράδειγμα, τα μειωτικά σφάλματα που οδηγούν σε αλληλεπικάλυψη και διαγραφή πρέπει να εμφανίζονται με ίσες συχνότητες, αλλά πρόσφατες μελέτες δείχνουν ότι η πρόωμη επιλογή κατά τη διάρκεια της γαμετογένεσης ευνοεί είτε την μία είτε την άλλη (120). Στην πραγματικότητα, είναι μια γενική παρατήρηση ότι οι διπλασιασμοί φαίνεται να οδηγούν σε έναν ηπιότερο ή όχι κλινικό φαινότυπο σε σύγκριση με την αμοιβαία διαγραφή τμημάτων DNA.



Εικόνα 3.3 Αυξανόμενος αριθμός δημοσιεύσεων ετησίως που σχετίζονται με την αναζήτηση νέων σύνδρομων στο Pub Med (121)

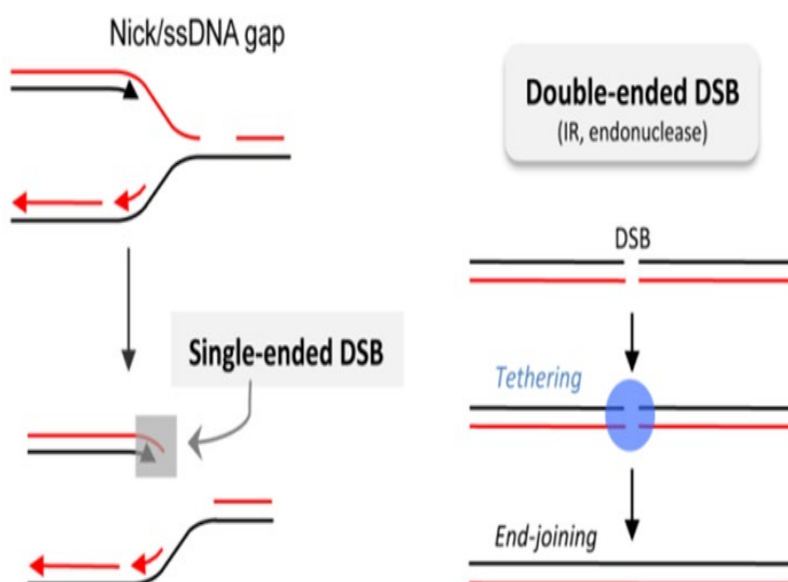
3.2 ΔΙΚΛΩΝΕΣ ΘΡΑΥΣΕΙΣ

Ως δίκλωνη θραύση (Double Strand Breaks, DSBs) ορίζεται η διάσπαση δύο φωσφοδιεστερικών δεσμών στις δύο αντικρινές αλυσίδες της δίκλωνης έλικας του DNA. Σε γενικότερη ανάλυση, οι δίκλωνες θραύσεις επιδιορθώνονται δυσκολότερα από τις θραύσεις στη μονόκλωνη αλυσίδα, λόγω του ότι στην τελευταία κατηγορία, η μία αλυσίδα παραμένει ανέπαφη και μπορεί να λειτουργήσει ως εκμαγείο χωρίς να χαθεί γενετική πληροφορία, σε αντίθεση με τα κύτταρα που έχουν υποστεί δίκλωνες θραύσεις από τα οποία αν έχει χαθεί πληροφορία, η έλλειψη θα είναι διαρκής (122). Άλλη μια παράμετρος που δυσχεραίνει την επιδιόρθωση των δίκλωνων θραύσεων αποτελεί η απομάκρυνση των άκρων των αλυσίδων στο σημείο της θραύσης με αποτέλεσμα να ελλοχεύει κίνδυνος λανθασμένου ανασυνδυασμού με άλλες περιοχές του γονιδιώματος. Τα συνήθως άναρχα κομμένα άκρα των δίκλωνων θραύσεων είναι ακόμη ένας λόγος που εν τέλει παρεμποδίζεται η επανένωση τους λόγω των κατεστραμμένων βάσεων που διαθέτουν, και οι οποίες θα πρέπει πρώτα να επεξεργαστούν καταλλήλως από DNA πολυμεράσες ή νουκλεάσες πριν επανασυνδεθούν (123).

Οι δίκλωνες θραύσεις μπορούν να διαχωριστούν σε δύο κατηγορίες ανάλογα με τον τρόπο που εμφανίζονται και την απόσταση / θέση που μεσολαβεί ανάμεσα τους στο γονιδίωμα:

- Double-ended DSBs → παρατηρούνται δύο άκρα δίκλωνων θραύσεων σε κοντινή απόσταση μεταξύ τους που προέρχονται από τη δράση ενδονουκλεασών ή μετά από έκθεση σε ιονίζουσα ακτινοβολία
- Single-ended DSBs → αποτελούν δίκλωνες θραύσεις με ένα και μοναδικό δίκλωνο άκρο που προκύπτουν ή από αντιγραφικό στρες ή από μη επικαλυπτόμενα άκρα τελομερών (124)

Εικόνα 3.4 Σχηματισμός Single-ended και Double-ended DSBs (125)



3.2.1 ΑΙΤΙΑ ΠΡΟΚΛΗΣΗΣ ΔΙΚΛΩΝΩΝ ΘΡΑΥΣΕΩΝ (DSBS)

Οι δίκλωνες θραύσεις παρουσιάζουν μεγάλη ποικιλομορφία ως προς τους παράγοντες από τους οποίους σχηματίζονται, είτε αυτοί αφορούν εξωγενείς είτε ενδογενείς παράγοντες.

3.2.1.1 ΕΝΔΟΓΕΝΕΙΣ ΠΑΡΑΓΟΝΤΕΣ

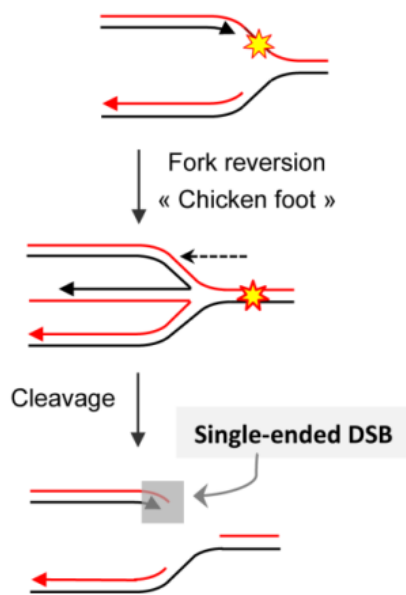
Αξίζει να σημειωθεί πως δίκλωνες θραύσεις επάγονται ενδογενώς και σε φυσιολογικές διεργασίες του κυττάρου όπως η αντιγραφή. Επομένως, μη επιδιορθωμένες θραύσεις μονόκλωνης αλυσίδας σε κοντινή απόσταση μεταξύ των αλυσίδων του δίκλωνου μορίου DNA κατά την αντιγραφή ή σε περιπτώσεις κατάρρευσης της διακλάδωσης αναδιπλασιασμού, μπορούν να οδηγήσουν στο σχηματισμό δίκλωνων θραύσεων (126, 127).

- Αντιγραφικό Στρες

Κατά τη διάρκεια της αντιγραφής, η αντιγραφική διχάλα καλείται να αντιμετωπίσει διαφόρου τύπου εμπόδια, τα οποία τείνουν να καθυστερήσουν την εξέλιξη της. Παραδείγματα αυτού του τύπου αποτελούν ειδικές περιοχές με δυσκολία αντιγραφής, πρωτεΐνες προσδεμένες σφιχτά στο DNA, καρκινογόνες ουσίες που σχηματίζουν σύμπλοκα ή πρόσθετα με το DNA (DNA adducts), καθώς και κακώσεις του γονιδιώματος. Για το λόγο αυτό, η καθυστέρηση της αντιγραφής με σκοπό την αντιμετώπιση τους μέσω ανασυνδυασμού, επάγει το σχηματισμό δίκλωνων θραύσεων.

Σε περίπτωση ήδη σχηματισμένων εγκοπών στη μία αλυσίδα, η δράση της αντιγραφικής μηχανής έχει ως αποτέλεσμα τη δημιουργία δίκλωνων θραύσεων. Πιο

αναλυτικά, οι δίκλωνες θραύσεις αυτές σχηματίζονται έπειτα από την ενεργοποίηση νουκλεασών όπως οι MUS81, GEN-1 και SLX4-SLX1. Αυτή η ενέργεια πραγματοποιείται έτσι ώστε να καταλυθεί η δομή, που έχει σχηματιστεί από την παλινδρόμηση της αντιγραφικής διχάλας λόγω του υβριδισμού με τις νεοσυντιθέμενες συμπληρωματικές αλυσίδες και το σχηματισμό μιας τετρασκελούς δομής (chicken feet).



Εικόνα 3.5 Σχηματισμός τετρασκελούς δομής (chicken feet) λόγω παλινδρόμησης της αντιγραφικής διχάλας (128)

Παράλληλα όλο και περισσότερες μελέτες υποστηρίζουν πως η παρέμβαση της μεταγραφής στη διαδικασία της αντιγραφής λόγω του ότι και οι δύο χρησιμοποιούν την ίδια μήτρα DNA έχει ως αποτέλεσμα την εμφάνιση γενωμικής αστάθειας. Ειδικότερα, σε γενετικούς τόπους όπου εδράζονται επαναλαμβανόμενα στοιχεία, CpG islands ή ομάδες γονίδιων, τα οποία αντιγράφονται σε πρώιμο στάδιο, παρατηρείται μια ιδιαίτερη αλληλεπίδραση της μεταγραφής και της αντιγραφής. Αυτή η παρεμβολή της μεταγραφής διαταράσσει τη γενωμική ακεραιότητα και καθιστά τις περιοχές αυτές hot spot για θραύσεις χρωμοσωμάτων σε περιοχές γονιδίων που σχετίζονται με τον καρκίνο και προκαλούνται κάτω από συνθήκες αντιγραφικού στρες.

Επιπλέον, η ύπαρξη εύθραυστων γενετικά περιοχών ή αλλιώς CFS (Common Fragile Sites) έχει προταθεί ως μια σοβαρή αιτία εμφάνισης δίκλωνων θραύσεων σε περιπτώσεις εκδήλωσης αντιγραφικού στρες. Χαρακτηριστικά, πάνω από 120 τέτοιου είδους περιοχές έχουν καταγραφεί στο ανθρώπινο γονιδίωμα και συχνά παίζουν ρόλο στην εμφάνιση όγκων.

Τέλος, αναφορικά με το οξειδωτικό, κατά το οποίο παράγονται δραστικές μορφές οξυγόνου μέσω του μεταβολισμού, και το αντιγραφικό στρες, υπάρχουν ενδείξεις, οι οποίες παρουσιάζουν μια στενή σχέση ανάμεσα τους, με το οξειδωτικό στρες να αποτελεί μια από τις αιτίες για την πρόκληση του αντιγραφικού στρες. Παράλληλα και τα δύο μπορούν να επαχθούν από ενεργοποίηση ογκογονιδίων, ενώ έχουν προταθεί να λαμβάνουν μέρος στην εμφάνιση καρκίνου και στην εκκίνηση της διαδικασίας του γηρασμού του κυττάρου.

- Ειδικές Δομές

Δομές οι οποίες θυμίζουν Holliday Junctions και προέρχονται είτε από G-τετραπλέγματα (G-quadruplexes, γουανίνες που διατάσσονται σε σειρά σχηματίζοντας μια συνεπτυγμένη δομή που συνδέεται μέσω θηλειών) είτε από διάφορες παλίνδρομες αλληλουχίες με δομές σε σχήμα σταυρού, μετασχηματίζονται σε δίκλωνες θραύσεις έπειτα από τη δράση ενζύμων ρεσολβασών.

Παράλληλα, η Z-DNA δομή (αριστερόστροφη έλικα DNA), που σχηματίζεται μέσω των αναδιατάξεων επαγόμενων από τη μεταγραφή σε αλληλουχίες πλούσιες σε πουρίνες και πυριμιδίνες, έχει βρεθεί σε περιοχές, οι οποίες έχουν υποστεί θραύσεις.

- Ρετροτρανσποζόνια

Τα L1 (Long Interspersed Elements-1) αποτελούν αυτόνομα ρετροτρανσποζόνια, τα οποία έχουν τη δυνατότητα να εξασφαλίσουν μαζί με τη δική τους μετακίνηση και τη μετακίνηση άλλων μη αυτόνομων ρετροτρανσποζονίων όπως τα ALUs ή τα SINES, μέσα στο γονιδίωμα. Για να πραγματοποιηθεί η αντίστροφη μεταγραφή και η ενσωμάτωση του μεταθετού αυτού στοιχείου, η L1 ενδονουκλεάση προκαλεί μια δίκλωνη θραύση στο DNA. Επιπροσθέτως, τα L1 ρετροτρανσποζόνια μπορούν να ενσωματωθούν στο γονιδίωμα και σε ήδη υπάρχοντα σημεία δίκλωνων θραύσεων. Αυτού του είδους η ρετρομετάθεση έχει παρατηρηθεί σε περιπτώσεις καρκίνου του πνεύμονα και σε άλλα είδη καρκίνου (124).

- Τελομερή

Τα τελομερή αποτελούν περιοχές στα άκρα των χρωμοσωμάτων και ο λόγος ύπαρξης τους εξυπηρετεί τη γενωμική ακεραιότητα του γονιδιώματος. Έπειτα από κάθε κυτταρική διαίρεση όπου έχει προηγηθεί διπλασιασμός του γενετικού υλικού, παρατηρείται μείωση των τελομερών λόγω της αδυναμίας της αντιγραφικής μηχανής να αντιγράψει το τελευταίο σε έκταση DNA των τμημάτων Okazaki (129). Η αποφυγή της εξάλειψης των τελομερών πραγματοποιείται μέσω της δράσης του ενζύμου της τελομεράσης, μίας αντίστροφης μεταγραφάσης, η οποία φέρει ένα RNA εκμαγείο συμπληρωματικό ως προς τα τελομερή και με τη βοήθεια του οποίου τα επιμηκύνει προσθέτοντάς τους νουκλεοτίδια στο 3'-OH άκρο.

Αναφορικά με τη δομή τους, τα τελομερή αποτελούνται από επαναλαμβανόμενες αλληλουχίες πλούσιες σε γουανίνη (G-rich) καθώς και μία 3' προεξέχουσα μονόκλωνη αλυσίδα. Η αλυσίδα αυτή στο 3' άκρο εξασφαλίζει την προστασία των τελομερών σε θηλαστικά, φυτά και άλλους οργανισμούς μέσω της εισβολής της στην περιοχή της δίκλωνης αλυσίδας του τελομερούς, σχηματίζοντας μία προστατευτική δομή βρόχου, την t-loop.

Παρόλα αυτά, υπάρχουν περιπτώσεις κατά τις οποίες αυτή η δομή χάνεται ή είναι μη λειτουργική και τα τελομερή μένουν μη επικαλυμμένα (uncapped telomeres) με αποτέλεσμα να αναγνωρίζονται από το μηχανισμό DDR ως δίκλωνες θραύσεις με τη μορφή Double-ended

DSBs και να ενεργοποιείται η επιδιόρθωση τους μέσω του μη ομόλογου ανασυνδυασμού (124).

Στους επιμέρους ενδογενείς παράγοντες πρόκλησης δίκλωνων θραύσεων συγκαταλέγονται το οξειδωτικό στρες από δραστικές μορφές οξυγόνου και η εφαρμογή μηχανικής πίεσης κατά το διαχωρισμό των χρωμοσωμάτων στο στάδιο της ανάφασης της μίτωσης. Ωστόσο, δίκλωνες θραύσεις προκαλούνται φυσιολογικά και σε άλλες διεργασίες του κυττάρου πέραν της αντιγραφής. Ανάλογα παραδείγματα αποτελούν ο ανασυνδυασμός στο στάδιο της μείωσης ή κατά την ισοτυπική μεταστροφή των βαριών αλυσίδων των ανοσοσφαιρινών των Β λεμφοκυττάρων, καθώς και ο V(D)J ανασυνδυασμός των ανοσοσφαιρινών (Ig) των Β λεμφοκυττάρων και του υποδοχέα TCR (T Cell Recceptor) των Τ λεμφοκυττάρων (130).

Ενδεικτικά,

- V(D)J ανασυνδυασμός των ανοσοσφαιρινών (Ig) των Β λεμφοκυττάρων

Τα γονίδια που κωδικοποιούν τις ανοσοσφαιρίνες (Ig) και τους υποδοχείς των Τ λεμφοκυττάρων βρίσκονται κατακερματισμένα σε διάφορους γενετικούς τόπους και συναρμολογούνται σε λειτουργικά γονίδια μέσα στα Β και Τ λεμφοκύτταρα, αντίστοιχα (126). Στόχος του ανοσοποιητικού συστήματος είναι ο εξοπλισμός του από ένα ευρύ ρεπερτόριο αντιγονικών υποδοχέων μέσω της διαδικασίας του V(D)J ανασυνδυασμού.

Απαραίτητο βήμα για την εκκίνηση του ανασυνδυασμού αποτελεί η πρόκληση δίκλωνων θραύσεων από ένζυμα που συναντώνται αποκλειστικά στα Β/Τ κύτταρα, τις ρεκομπινάσες RAG1 (Recombination Active Gene 1) και RAG-2. Οι ρεκομπινάσες αυτές αναγνωρίζουν ειδικές αλληλουχίες ανασυνδυασμού (RSS, Recombination Signal Sequences) και κόβουν μεταξύ αυτών και του θραύσματος του επιθυμητού γονιδίου ανοσοσφαιρίνης αφήνοντας τυφλά άκρα, ενώ παράλληλα σχηματίζεται μία δομή φουρκέτας στην πλευρά της

κωδικοποιούσας αλληλουχίας των εμπλεκόμενων γονιδίων. Οι δίκλωνες θραύσεις επιδιορθώνονται και πραγματοποιείται συρραφή των σπασμένων άκρων μέσω του μη ομόλογου ανασυνδυασμού και τη συμμετοχή πρωτεϊνών αυτού όπως οι KU, DNA PKcs, Ligase IV και XRCC4 (X-Ray Repair Cross Complementing 4), ενώ οι ενδιάμεσες αλληλουχίες απαλείφονται.

3.2.1.2 ΕΞΩΓΕΝΕΙΣ ΠΑΡΑΓΟΝΤΕΣ

Στους εξωγενείς παράγοντες συμπεριλαμβάνονται ξενοβιοτικά, τα οποία μπορούν να μετατραπούν σε βιοδραστικές ουσίες στον οργανισμό, ιονίζουσα ακτινοβολία (ακτίνες X και γ), η οποία παράγει δραστικές μορφές οξυγόνου, ROS (Reactive Oxygen Species), υπεριώδης ακτινοβολία (UV, ultraviolet radiation), χημικές ουσίες, αντικαρκινικά φάρμακα (αναστολή της δράσης της τοποϊσομεράσης II), διατάραξη της ομοιόστασης του συστήματος καθώς και εξωγενώς ενεργοποιημένος κυτταρικός θάνατος (130). Παρακάτω θα αναλυθεί εκτενέστερα η επίδραση της ιονίζουσας και της υπεριώδους ακτινοβολίας, καθώς και ορισμένων χημικών ουσιών στην εμφάνιση βλάβης στο DNA. Μία από τις πιο σοβαρές αιτίες πρόκλησης πολλαπλών κακώσεων στο DNA αποτελεί η ιονίζουσα ακτινοβολία (IR, Ionizing Radiation). Οι κακώσεις αυτές που πλήττουν ταυτόχρονα και τις δύο αλυσίδες και εδράζονται σε έκταση μιας στροφής του DNA, ονομάζονται θέσεις πολλαπλής κάκωσης (MDS, Multiple Damage Sites) ή μπορούν διαφορετικά να αναφερθούν ως ομαδοποιημένες περιοχές βλάβης του DNA (clustered DNA damage). Ο σχηματισμός τέτοιου είδους κάκωσης είναι χαρακτηριστικό της ιονίζουσας ακτινοβολίας μιας και είναι σχεδόν αδύνατο να επαχθούν λόγω τυχαίων χημικών τροποποιήσεων όπως οξειδωση, απαμίνωση ή υδρόλυση βάσης, δηλαδή λόγω διαδικασιών κυτταρικής αύξησης ή διατήρησης που πραγματοποιούνται φυσιολογικά στον οργανισμό (131).

Η ιονίζουσα ακτινοβολία εναποθέτει ενέργεια και στις δύο αλυσίδες του DNA με αποτέλεσμα την απευθείας καταστροφή του μορίου. Παράλληλα, η ιονίζουσα ακτινοβολία

είναι υπεύθυνη για το σχηματισμό δραστικών μορφών οξυγόνου (ROS) μέσω της υδρόλυσης του νερού, καθώς και για το σχηματισμό ομαδοποιημένων εγκοπών, οξειδωμένων βάσεων και θέσεων όπου απουσιάζουν είτε οι πυριδίνες είτε οι πυριμιδίνες (abasic sites). Τέλος, το οξειδωτικό στρες που επάγεται από την IR έχει την ικανότητα πρόκλησης αντιγραφικού στρες αντίστοιχα. Όλες οι παραπάνω επιπτώσεις της έκθεσης σε ιονίζουσα ακτινοβολία συντελούν στην εμφάνιση δίκλωνων θραύσεων στο DNA. Αναφορικά με την υπεριώδη ακτινοβολία και τις χημικές ουσίες, έχει αποδειχθεί πως δημιουργούν πρόσθετα με το DNA και η επιδιόρθωση αυτών μπορεί να σχηματίσει ενδιάμεσες δομές που διαθέτουν τη δυνατότητα μετατροπής σε δίκλωνες θραύσεις, αλλά και να προκαλέσουν δίκλωνες θραύσεις μέσω επαγωγής αντιγραφικού στρες, όπως ήδη αναφέρθηκε. Πιο συγκεκριμένα, ορισμένα πρόσθετα DNA που προκαλούνται από χημικά μπορούν να παρεμποδίσουν την εξέλιξη της αντιγραφικής διχάλας ή της διαδικασίας της μεταγραφής, όπως συμβαίνει με τις μιτομυκίνη C και τραμπεκτιδίνη, αντίστοιχα (132).

3.2.2 ΓΕΝΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΤΩΝ ΤΡΙΩΝ ΕΙΔΩΝ ΚΑΡΚΙΝΟΥ ΠΟΥ ΜΕΛΕΤΗΘΗΚΑΝ

Οι μη μεταδοτικές ασθένειες είναι πλέον υπεύθυνες για την πλειοψηφία των θανάτων παγκοσμίως και ο καρκίνος αναμένεται να καταταχθεί ως η κύρια αιτία θανάτου και το σημαντικότερο εμπόδιο για την αύξηση του προσδόκιμου ζωής σε κάθε χώρα του κόσμου τον 21ο αιώνα. Σύμφωνα με εκτιμήσεις του Παγκόσμιου Οργανισμού Υγείας (ΠΟΥ) το 2015, ο καρκίνος είναι η πρώτη ή η δεύτερη αιτία θανάτου πριν από την ηλικία των 70 ετών σε 91 από τις 172 χώρες και κατατάσσεται τρίτη ή τέταρτη σε άλλες 22 χώρες. Την τελευταία δεκαετία έχουν γίνει πολλά βήματα στην κατανόηση της γενετικής βάσης της ανθρώπινης νόσου. Αναμφισβήτητα, μεγαλύτερο αντίκτυπο υπήρξε στον τομέα της γενετικής του καρκίνου, όπου ο τεράστιος όγκος των δεδομένων γονιδιωματικών αλληλουχιών και μοριακών προφίλ, έδειξε την πολυπλοκότητα των ανθρώπινων κακοηθειών. Σε ένα κύτταρο όγκου, δεκάδες διαφορετικά γονίδια μπορεί να είναι ανώμαλα σε δομή ή αριθμό

αντιγράφων και εκατοντάδες ή χιλιάδες γονίδια μπορεί να εκφράζονται διαφορεικά. Έχουν ταυτοποιηθεί πολλές γνωστές αλληλουχίες με μεταλλάξεις υψηλής διείσδυσης, αλλά η συμβολή των γενετικών παραλλαγών ή πολυμορφισμών χαμηλής διείσδυσης στον κίνδυνο ανάπτυξης σποραδικού καρκίνου παραμένει ασαφής. Μελέτες των πολύπλοκων σωματικών γενετικών συμβάντων που λαμβάνουν χώρα στο αναπτυσσόμενο καρκινικό κύτταρο μπορεί να βοηθήσουν στην αναζήτηση των πιο δυσδιάκριτων παραλλαγών της γενετικής γραμμής που προσδίδουν αυξημένη ευαισθησία στην εμφάνιση καρκίνου. Οι ενδείξεις για τη μοριακή παθογένεση του καρκίνου έχουν δώσει νέες στρατηγικές για τη θεραπεία, αλλά μια βαθύτερη κατανόηση αυτής της ασθένειας θα απαιτήσει νέες στατιστικές και υπολογιστικές προσεγγίσεις για την ανάλυση των γενετικών και σηματοδοτικών δικτύων που ενορχηστρώνουν την επιδεκτικότητα του καρκίνου και τη συμπεριφορά του όγκου. Συγκεκριμένα έγινε μελέτη για τον καρκίνο του μαστού, τον καρκίνο της κεφαλής και τραχήλου (head and neck carcinoma) καθώς και για το αδενοκαρκίνωμα του πνεύμονος. Αναφορικά με το διηθητικό καρκίνο του μαστού BIC (Breast Invasive Carcinoma) ,αποτελεί έναν από τους πιο κοινούς καρκίνους που πλήττουν σχεδόν κατ' αποκλειστικότητα το γυναικείο πληθυσμό, ο οποίος εκτός από εξωγενείς παράγοντες, μπορεί να οφείλεται και σε κληρονομικότητα, όπως η ύπαρξη μεταλλάξεων ή παρεμπόδισης της έκφρασης στα BRCA1 και BRCA2 γονίδια. Όσον αφορά το αδενοκαρκίνωμα του πνεύμονα LUAD (Lung Adenocarcinoma) αποτελεί έναν από τους πιο θανατηφόρους καρκίνους τόσο στους άνδρες όσο και στις γυναίκες. Τέλος, ο τρίτος καρκίνος που μελετήθηκε, πλήττει τις υγρές βλεννώδεις μεμβράνες εσωτερικά του στόματος, της μύτης και του λαιμού και οφείλεται στο καρκίνωμα που εμφανίζεται στα πλακώδη κύτταρα της κεφαλής και του τραχήλου (HN, Head and Neck squamous cell carcinoma), τα οποία απαρτίζουν τις προαναφερθείσες μεμβράνες (133, 134)

3.3 ΜΗΧΑΝΙΣΜΟΙ ΕΠΙΔΙΟΡΘΩΣΗΣ ΓΕΝΕΤΙΚΗΣ ΑΣΤΑΘΕΙΑΣ

3.3.1 ΜΗ ΟΜΟΛΟΓΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ / ΜΗ ΟΜΟΛΟΓΗ ΣΥΝΔΕΣΗ ΑΚΡΩΝ ΝΗΕJ (NON-HOMOLOGOUS END JOINING)

Οι οργανισμοί έχουν εξελίξει ένα ιδιαίτερα ποικιλόμορφο σύστημα επιδιόρθωσης, το οποίο απαρτίζεται κυρίως από τον μη ομόλογο και τον ομόλογο ανασυνδυασμό, καθώς και διάφορα άλλα παρακλάδια τους. Για παράδειγμα, σύμφωνα με μελέτες σε κύτταρα πιθήκων από τα οποία απουσίαζαν παράγοντες του μη ομόλογου ανασυνδυασμού που εμπλέκονται στην ένωση δύο κατατετμημένων άκρων, οι δίκλωνες θραύσεις επιδιορθώθηκαν επιτυχώς χωρίς να προκύψει κάποιο πρόβλημα στην ένωση των άκρων. Παρόμοια σχεδόν αποτελέσματα σημειώθηκαν και σε άλλους ανώτερους ευκαρυωτικούς οργανισμούς (όρνιθα, ποντικός), από όπου είτε απουσίαζαν παράγοντες του μη ομόλογου ανασυνδυασμού είτε εμφάνιζαν κάποια δυσλειτουργία. Ωστόσο, παρατηρήθηκαν και περιπτώσεις όπου έπειτα από επαγωγή ιονίζουσας ακτινοβολίας αυξήθηκαν τα ποσοστά γενετικών αναδιατάξεων. Συνοψίζοντας όλα τα παραπάνω στοιχεία γεννήθηκε η υποψία πως ένας άλλος μηχανισμός επιδιόρθωσης εφαρμόζονταν στις προκύπτουσες δίκλωνες θραύσεις με ορισμένα από τα χαρακτηριστικά του μη ομόλογου, αλλά και του ομόλογου ανασυνδυασμού.

Πράγματι, σε περιπτώσεις όπου ο κλασσικός μη ομόλογος ανασυνδυασμός αδυνατούσε να λάβει δράση για τον οποιοδήποτε λόγο, είτε λόγω ανεπάρκειας / απουσίας του ίδιου, είτε λόγω της φάσης του κυτταρικού κύκλου κατά την οποία φυσιολογικά είναι απενεργοποιημένος, τη θέση του καταλαμβάνει ένας άλλος μηχανισμός. Ο μηχανισμός αυτός είναι ο εναλλακτικός μη ομόλογος ανασυνδυασμός (A-EJ, Alternative End Joining), για τον οποίο έχουν χρησιμοποιηθεί διάφορες ονομασίες, όπως microhomology-mediated end-joining(MMEJ), KU-independent end-joining, LigIV-independent NHEJ, και backup NHEJ (B-NHEJ). Από το σημείο αυτό κι έπειτα θα αναφέρεται ως εφεδρικός μη ομόλογος ανασυνδυασμός, διότι κυρίως επιλέγεται σε καταστάσεις όπου ο κλασσικός μη ομόλογος ανασυνδυασμός καθίσταται δυσλειτουργικός. Τα χαρακτηριστικά του προαναφερθέντος μηχανισμού θα αναλυθούν εκτενέστερα παρακάτω (131).

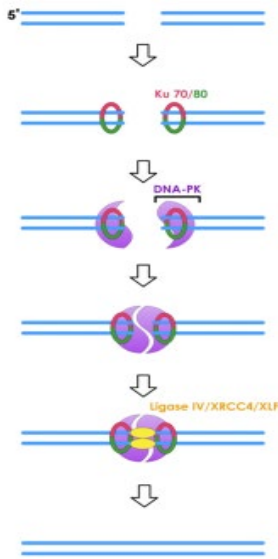
Αν και στον κλασσικό μη ομόλογο ανασυνδυασμό έχουν χρεωθεί οι περισσότερες χρωμοσωμικές ανωμαλίες με τη μορφή διαγραφής ή προσθήκης βάσεων, μεταθέσεων

τμημάτων DNA, αλλά και διάφορες άλλες γενετικές αναδιατάξεις, χαρακτηρίζοντάς τον ως επιρρεπή σε λάθη “error-prone”, η ιδιότητά του αυτή έχει μια ιδιαίτερη σκοπιμότητα. Πράγματι, την έλλειψη ολοκληρωτικής ακεραιότητας που χαρακτηρίζει τον κλασικό μη ομόλογο ανασυνδυασμό εκμεταλλεύεται ο V (D) J ανασυνδυασμός, έτσι ώστε να ενισχυθεί η ποικιλομορφία σε αντιγονικούς υποδοχείς και να θωρακιστεί όσο το δυνατόν καλύτερα το ανοσοποιητικό σύστημα στα ανώτερα θηλαστικά. Καθίσταται επομένως προφανής η λειτουργική, αλλά και η εξελικτική σημασία της μειωμένης ακεραιότητας του μη ομόλογου ανασυνδυασμού.

Η βασική έννοια γύρω από το μηχανισμό του κλασικού μη ομόλογου ανασυνδυασμού σχετίζεται με την επανένωση δύο κομμένων άκρων, χωρίς να αποτελεί απαραίτητη προϋπόθεση η ύπαρξη ομόλογης αλληλουχίας ως μήτρα για να πραγματοποιηθεί η σύνδεση μεταξύ τους. Κυρίως χρησιμοποιείται για την επιδιόρθωση απλών θραύσεων ή θραύσεων που αφήνουν τυφλά άκρα, αλλά με τον τρόπο αυτό αυξάνεται ο κίνδυνος για αλλαγές στη νουκλεοτιδική αλληλουχία στα σημεία των θραύσεων, όταν τα άκρα δεν είναι συμβατά. Παρόλα αυτά, ο όρος “μη ομόλογος” για το συγκεκριμένο τύπο ανασυνδυασμού κρύβει παρερμηνείες μιας και η ύπαρξη μιας μικρής σε έκταση περιοχής ομολογίας (microhomology) μόλις 1-6 bp ταυτοποιείται κοντά στα κατεστραμμένα άκρα, εξασφαλίζοντας τη σύνδεση μεταξύ τους. Η μικρή έκταση αναφορικά με την περιοχή ομολογίας αποτελεί ειδοποιό διαφορά ανάμεσα στον μη ομόλογο και τον ομόλογο ανασυνδυασμό, διότι στον ομόλογο ανασυνδυασμό η συγκεκριμένη περιοχή καταλαμβάνει μέχρι 100 bp ή και περισσότερο (135).

Η διαδικασία του μη ομόλογου ανασυνδυασμού ξεκινάει με την πρόσδεση στα κατεστραμμένα άκρα, του ετεροδιμερούς Ku70/80, η οποία συνοδεύεται με την πρόσδεση της DNA-PK πρωτεΐνης για να σχηματιστεί μία δομή δακτυλίου που προσδένεται πιο εξειδικευμένα στα DNA άκρα. Η σύζευξη του ετεροδιμερούς με την DNA-PK δρομολογεί τη στρατολόγηση της καταλυτικής υπομονάδας της DNA-PK, μίας κινάσης σερίνης/ θρεονίνης, την DNA-PKcs, με σκοπό το σχηματισμό του ενεργού πλέον DNA-PK ολοενζύμου. Η ενεργότητα κινάσης του ολοενζύμου προωθεί τη φωσφορυλίωση μιας σειράς πρωτεϊνών όπως οι RPA2, WRN, XLF/Cernunnos, DNA Ligase IV (LigIV) και XRCC4, η δράση των οποίων θα αναλυθεί παρακάτω. Το ολοένζυμο διαθέτει και τη δυνατότητα αυτοφωσφορυλίωσης, η

οποία προσφέρει την απαραίτητη ενέργεια για την αλλαγή στη δομή του ολοενζύμου από ανοιχτή σε κλειστή διαμόρφωση σε περιπτώσεις όπου τα άκρα των θραύσεων χρειάζεται πρωτίστως να επεξεργαστούν για να μπορέσει να λάβει χώρα η σύνδεση μεταξύ τους. Ακολούθως, μετά την αυτοφωσφορυλίωση του ολοενζύμου, στα άκρα προσδένεται το σύμπλοκο της λιγάσης Ligase IV, το οποίο περιλαμβάνει τους XRCC4 και XLF συμπαραγόντες, πιθανόν για την αποτελεσματικότερη σύνδεση της λιγάσης IV στα άκρα των θραύσεων.



Εικόνα 3.6 Το μοντέλο του κλασικού μη ομόλογου ανασυνδυασμού (NHEJ) (136)

Ωστόσο, στις περισσότερες περιπτώσεις τα άκρα στα όρια των θραύσεων δεν είναι σε κατάσταση που να επιτρέπει την ένωση από τη λιγάση, αλλά χρειάζονται περαιτέρω επεξεργασία, διότι είτε έχει καταστραφεί μέρος του σκελετού του σακχάρου της αλυσίδας, είτε οι βάσεις τους έχουν υποστεί βλάβη. Ακόμη, μπορεί να διαθέτουν 3' φωσφορικά ή 5' υδροξυλικά άκρα, τα οποία δεν βρίσκονται στην κατάλληλη δομή για να πραγματοποιηθεί η διαδικασία της σύνδεσης των άκρων. Για να γίνουν συμβατά προς επανένωση τα παραπάνω άκρα, μπορούν να επεξεργαστούν από μια πολυνουκλεοτιδική κινάση, η οποία θα αλληλεπιδράσει έπειτα με τον XRCC4 παράγοντα. Άλλα δυσλειτουργικά άκρα με ασύμβατες DNA δομές μπορεί να τα επιμεληθεί μια εξειδικευμένη ως προς τη δομή νουκλεάση, η Artemis, η οποία έχει τη δυνατότητα να κόβει 3' προεξέχοντα μονόκλωνα άκρα. Μια επιμέρους πρωτεΐνη με ικανότητα επεξεργασίας άκρων αποτελεί η WRN (Werner syndrome proteins), η οποία διαθέτει ενεργότητα εξωνουκλεάσης. Επιπλέον, διάφορες άλλες DNA

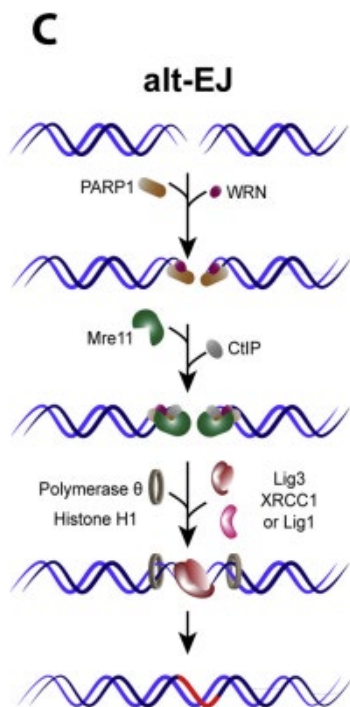
πολυμεράσες, όπως οι μ και λ πολυμεράσες, έχουν χαρακτηριστεί για το ρόλο τους στην επιμήκυνση 5' προεξέχοντων μονόκλωνων άκρων. Αξιοσημείωτο, τέλος, αποτελεί το γεγονός ότι σε ορισμένες πολύπλοκες κακώσεις των άκρων των θραύσεων έχει παρατηρηθεί να συμμετέχουν πρωτεΐνες όπως οι ATM, 53BP1, το MRN σύμπλοκο (MRE11/RAD50/NBS1) και η φωσφορυλιωμένη H2AX ιστόνη. Οι δράσεις τους όμως θα πρέπει να διερευνηθούν περαιτέρω (135).

Αντίθετα με όσα αναφέρθηκαν σχετικά με το μη ομόλογο ανασυνδυασμό, κατά τον εφεδρικό μη ομόλογο ανασυνδυασμό (Back-up NHEJ, B-NHEJ) δεν παρατηρείται πρόσδεση του διμερούς Ku70/80, αλλά φαίνεται πως υπάρχει ανταγωνισμός για την πρόσδεση στα άκρα τη δίκλωνης θραύσης με την PARP1 (Poly [ADP-ribose] polymerase 1) πρωτεΐνη. Αν και οι Ku πρωτεΐνες παρουσιάζουν μεγαλύτερη συγγένεια ως προς τα άκρα συγκριτικά με την PARP1, φαίνεται πως το εμπόδιο αυτό υπερκεράται, σε περιπτώσεις όπου τόσο ο μη ομόλογος όσο και ο ομόλογος ανασυνδυασμός έχουν αποτύχει να επιδιορθώσουν τις θραύσεις. Πράγματι, η PARP1 πρωτεΐνη αναγνωρίζει τις θέσεις των δίκλωνων θραύσεων, προσδέεται σε αυτές και τις φέρνει κοντά, ενώ παράλληλα επάγει τη στρατολόγηση των XRCC1 και DNA Ligase III, για να λάβει χώρα η σύνδεση μεταξύ τους (υπάρχει περίπτωση η ένωση των άκρων να διαμεσολαβεί και από τη DNA Ligase I). Ακόμα, μόνο σε περιπτώσεις όπου αυξάνονται τα επίπεδα της WRN (Werner syndrome RecQ like helicase) πρωτεΐνης, η ίδια με την ενεργότητα ελικάσης που διαθέτει, αρχικά ξετυλίγει το DNA στις θέσεις των θραύσεων. Ενδεχομένως, η WRN μπορεί να δράσει και ως εξωνουκλεάση, αφαιρώντας τις ανεπιθύμητες δομές DNA που δημιουργήθηκαν κατά τη θραύση.

Γενικότερα, ο B-NHEJ ανασυνδυασμός είναι ιδιαίτερα μεταλλαξογόνος μηχανισμός επιδιόρθωσης και συχνά χρησιμοποιεί μικρο-ομολογίες (2 έως 8 νουκλεοτίδια) που εμφανίζουν συμπληρωματικότητα και βρίσκονται απομακρυσμένες από τη δίκλωνη θραύση. Για να μπορέσουν να αποκαλυφθούν οι θέσεις των μικρο-ομολογίων, να ευθυγραμμιστούν μεταξύ τους και να υβριδοποιηθούν θα πρέπει πρώτα να δημιουργηθούν 3' προεξέχοντα μονόκλινα άκρα μέσω εκτομής. Με τον τρόπο αυτό, σχηματίζεται μια ενδιάμεση δομή όπως αυτή που παρατηρείται στο SSA μηχανισμό επιδιόρθωσης. Η αντίδραση αυτή καταλύεται από τη φωσφορυλιωμένη / ενεργή CtIP πρωτεΐνη, η οποία ενεργοποιεί την ενεργότητα ενδονουκλεάσης του MRN συμπλόκου. Μετά την επεξεργασία των άκρων από τις

ενδονουκλεάσες, οι μικρο-ομολογίες ευθυγραμμίζονται μεταξύ τους μέσω της H1 ιστόνης και σταθεροποιούνται από τη θ πολυμεράση. Η πολυμεράση θ είναι μια παρακαμπτήρια DNA πολυμεράση, με μεγαλύτερο ποσοστό λάθους (τοποθετεί έως και περισσότερα από 10 νουκλεοτίδια) από τις άλλες ομοειδείς μ και λ . Έχει τη δυνατότητα να προσδεθεί στο 3' προεξέχον άκρο και να το χρησιμοποιήσει ως εκκινήτη εισάγοντας ακόμη και μικρο-ομολογίες που δεν υφίσταντο προηγουμένως. Με τον τρόπο αυτό, εισάγονται προσθήκες βάσεων που αλλοιώνουν τη νουκλεοτιδική αλληλουχία.

Διαφορετικά, εκτιμάται ότι τα μη συμβατά άκρα που απομένουν μετά την υβριδοποίηση των μικρο-ομολογίων απομακρύνονται πιθανόν μέσω της ERCC1 (DNA excision repair protein ERCC-1) με αποτέλεσμα τη διαγραφή περιοχών DNA στο σημείο της σύνδεσης των μικρο-ομολογίων και την απώλεια γενετικής πληροφορίας. Τέλος, για τη σύνδεση των άκρων των αλυσίδων χρησιμοποιείται το σύμπλοκο της DNA Ligase III (ενίοτε και DNA Ligase I) με την XRCC1, όπου η τελευταία δρα ως πρωτεΐνη σκαλωσιάς και διαθέτει θέσεις τόσο για τη DNA Ligase III όσο και για την PARP1 πρωτεΐνη (137, 138).



Εικόνα 9. Μονοπάτι εφεδρικού μη ομόλογου ανασυνδυασμού, B-NHEJ [29](136)

3.3.2 ΟΜΟΛΟΓΟΣ ΑΝΑΣΥΝΔΥΑΣΜΟΣ HR (HOMOLOGOUS RECOMBINATION)

Αντίθετα με όσα συζητήθηκαν στο μη ομόλογο ανασυνδυασμό, κατά την επιδιόρθωση από τον ομόλογο ανασυνδυασμό η χρήση μήτρας DNA με ομόλογη ακολουθία αποτελεί απαραίτητη προϋπόθεση για να συντεθούν εκ νέου τμήματα κατεστραμμένης ή χαμένης νουκλεοτιδικής αλληλουχίας. Η εύρεση αυτής της ομόλογης αλληλουχίας αναζητείται κατά προτεραιότητα στα πλαίσια αδελφών χρωματίδων ή ακολούθως σε οποιαδήποτε ομόλογη ακολουθία του γονιδιώματος. Το κύριο μονοπάτι επιδιόρθωσης δίκλωνων θραύσεων του ομόλογου ανασυνδυασμού (DSBR, Double-Ended Breaks Repair) περιλαμβάνει τη διείσδυση και δεύτερης αλυσίδας και το σχηματισμό δομών Holliday.

Ωστόσο έχουν καταγραφεί και επιπλέον μονοπάτια, τα οποία στηρίζονται στη χρήση ομόλογης αλληλουχίας για να επιδιορθώσουν δίκλωνες θραύσεις, τα οποία έχουν ως εξής:

1. Υβριδοποίηση μονής αλυσίδας (SSA, Single-Strand Annealing)
2. Μονοπάτι επιδιόρθωσης αλυσίδας εξαρτώμενο από τη σύνθεση (SDSA, Synthesis-Dependent Strand Annealing pathway)
3. Αντιγραφή επαγόμενη από θραύση (BIR, Break-Induced Replication)

Ο πρώτος μηχανισμός χρησιμοποιείται αποκλειστικά για την επιδιόρθωση δίκλωνων θραύσεων με δύο άκρα (double-ended DSBs), ο δεύτερος συμμετέχει και στην αντιμετώπιση δίκλωνων θραύσεων με μονά άκρα (single-ended DSBs), ενώ ο τελευταίος μηχανισμός εξειδικεύεται κυρίως στην επούλωση δίκλωνων θραύσεων με μονά άκρα.

Ο ομόλογος ανασυνδυασμός, αρχικά, μελετήθηκε στο ζυμομύκητα και παρόλο που οι παράγοντες που τον απαρτίζουν στο ζυμομύκητα είναι λιγότεροι από αυτούς που συμμετέχουν στους ανώτερους ευκαρυωτικούς οργανισμούς, οι αρχές που τον διέπουν είναι σχεδόν οι ίδιες. Σε ένα γενικό πλαίσιο, ο ομόλογος ανασυνδυασμός διαθέτει κάποια συγκεκριμένα βήματα:

1. Νουκλεοτιδική εκτομή των άκρων των δίκλωνων θραύσεων
2. Σχηματισμός ενός νουκλεοτιδικού νηματίου
3. Ανίχνευση ομόλογης αλληλουχίας και προσβολή αλυσίδας
4. Κατάλυση του ανασυνδυασμού και κατάληξη σε διαφορετικούς υποτύπους ανασυνδυασμού

Πιο αναλυτικά, ο μηχανισμός του ομόλογου ανασυνδυασμού κατά τα αρχικά στάδια περιλαμβάνει την επεξεργασία των άκρων έτσι ώστε να σχηματιστούν 3' προεξέχοντα μονόκλινα άκρα με τη βοήθεια του MRN συμπλόκου, της EXO1 (Exonuclease 1), DNA2 (DNA Replication helicase/nuclease 2) και της CtIP πρωτεΐνης, όπως είχε αναφερθεί και προηγουμένως. Η πρώτη εγκοπή καταλύεται από τη δράση ενδονουκλεάσης της MRE11 σε κατεύθυνση 3' προς 5', έπειτα από ενεργοποίηση από τη CtIP πρωτεΐνη. Ακολούθως, η συνεργιστική δράση της EXO1 εξωνουκλεάσης σε κατεύθυνση 5' προς 3', αλλά και η ενεργότητα εξωνουκλεάσης που επιπλέον διαθέτει η MRE11, εξυπηρετούν την περαιτέρω εκτομή των άκρων των θραύσεων προς δύο κατευθύνσεις. Η μονόκλινη αλυσίδα που δημιουργείται μετά την εκτομή των νουκλεοτιδίων καλύπτεται από την RPA (Replication Protein A) πρωτεΐνη. Το στάδιο αυτό αποτελεί το στάδιο της προ-σύναψης (pre-synapsis).

Η πρόσδεση της RPA στα μονόκλινα άκρα ευνοείται λόγω του ότι έχει μεγαλύτερη χημική συγγένεια προς αυτά συγκριτικά με τη RAD51 πρωτεΐνη. Η πρόσδεση της RAD51 στα μονόκλινα άκρα πραγματοποιείται συνεργιστικά και σχηματίζεται ένα τριμερές σύμπλοκο στο οποίο συμμετέχει και το ATP. Σύμφωνα με πειράματα *in vitro*, η RPA παρεμποδίζει την

πρόσδεση της RAD51, αλλά σε *in vivo* καταστάσεις το πρόβλημα αυτό μπορεί να υπερκεραστεί μετά από την επέμβαση συμπαραγόντων της RAD51, όπως οι XRCC2 (X-ray Repair Cross Complementing Protein), XRCC3, RAD51B, RAD51C, RAD51D, RAD52, RAD54 και η BRCA2 πρωτεΐνη. Οι παραπάνω πρωτεΐνες χαρακτηρίζονται και ως διαμεσολαβητές και συγκροτούν δύο διαφορετικά σύμπλοκα. Το πρώτο απαρτίζεται από τις RAD51B, RAD51C, RAD51D και XRCC2 και το δεύτερο από τις RAD51C και XRCC3 πρωτεΐνες. Η δεύτερη σειρά πρωτεϊνών-διαμεσολαβητών εκπροσωπείται από την RAD52, η οποία έχει φυσική αλληλεπίδραση με την RAD51 και την RPA προωθώντας το σχηματισμό ενός νουκλεοπρωτεϊνικού νηματίου. Η τρίτη τάξη διαμεσολαβητών αποτελείται από την BRCA2 πρωτεΐνη, η οποία αλληλεπιδρά άμεσα με την RAD51 και τη μεταφέρει στα άκρα, τα οποία επικαλύπτονται από την RPA πρωτεΐνη.

Κατά το δεύτερο στάδιο – στάδιο σύναψης– πραγματοποιείται η αναζήτηση συμπληρωματικής αλληλουχίας, που θα χρησιμοποιηθεί ως μήτρα για την επιδιόρθωση των δίκλωνων ρηγμάτων, καθώς και η διείσδυση στο εσωτερικού ενός άθικτου μορίου DNA που περιέχει την εν λόγω αλληλουχία, με τη βοήθεια του προαναφερθέντος προσυναπτικού νηματίου. Η ακόλουθη αυτή διείσδυση ολοκληρώνεται με το σχηματισμό μιας ενδιάμεσης θηλιάς – D (Displacement loop, D-loop) όπου το αρχικό 3' προεξέχον άκρο της πληγείσας αλυσίδας εκτοπίζει την ομοειδή της αλυσίδα και καταλαμβάνει τη θέση της. Έπειτα αρχίζει τη σύνθεση DNA έχοντας ως εκμαγείο τη συμπληρωματική αλυσίδα του δίκλωνου μορίου στο οποίο διείσδυσε. Η σταθεροποίηση του RAD51- νουκλεοπρωτεϊνικού νηματίου και ο σχηματισμός της D θηλιάς καταλύεται από τη RAD54B κινητική πρωτεΐνη. Ακολούθως, η RAD54B αποδεσμεύει το νημάτιο για να ξεκινήσει η σύνθεση DNA κυρίως από τη DNA πολυμεράση (δ) ή (η), οι οποίες χρησιμοποιούν το 3' προεξέχον άκρο ως εκκινητή, έχουν όμως ταυτοποιηθεί να συμμετέχουν και παρακαμπτήριες DNA πολυμεράσες (error-prone) για την επέκταση του κλώνου.

ΣΚΟΠΟΣ

Η παρούσα διατριβή εκπονήθηκε για να δώσει απαντήσεις σχετικά με το πώς διάφορες αλληλουχίες μπορούν εν δυνάμει να αποτελέσουν συστατικά στοιχεία της εξέλιξης και της παθογένεσης. Ειδικότερα μέσω ουσιαστικής και τεκμηριωμένης μεθοδολογίας χαρακτηρίστηκαν αλληλουχιακά μοτίβα μέσω βιοπληροφορικών αναλύσεων και συσχετίστηκαν με σύνδρομα αναπτυξιακών διαταραχών και νοητικής στέρησης καθώς και με 3 τύπους καρκίνου. Πιο αναλυτικά έγιναν με σειρά τα παρακάτω βήματα:

1. Δημιουργία φυσικού και γενετικού χάρτη ανασυνδυασμού στα 2 φύλα
2. Ταυτοποίηση αλληλουχιών ανασυνδυασμού και χαρακτηρισμός των λειτουργικών τους χαρακτηριστικών
3. Συσχετισμός των αλληλουχιών ανασυνδυασμού με σύνδρομα αναπτυξιακών διαταραχών και νοητικής στέρησης καθώς και με λειτουργικά γονίδια της ανάπτυξης
4. Ταυτοποίηση αλληλουχιών με ανασυνδυαστική δράση σε τρεις μορφές καρκίνου.

Σύμφωνα με τα παραπάνω απομονώθηκαν τα αλληλουχιακά μοτίβα που βάσει στατιστικής σημαντικότητας βρέθηκε να έχουν καθοριστικό ρόλο στη παθογένεση των αναπτυξιακών συνδρόμων και του καρκίνου καθώς και στη δημιουργία ισοζυγισμένου ανασυνδυασμού.

ΕΙΔΙΚΟ ΜΕΡΟΣ

ΚΕΦΑΛΑΙΟ 4

ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

4.1 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ DECODE GENETICS ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΤΗ ΜΕ ΔΕΙΚΤΕΣ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΣΤΑ 2 ΦΥΛΑ

deCODE genetics

Στην ιστοσελίδα της *deCODE genetics* (139) υπάρχουν αναρτημένοι έξι γενετικοί χάρτες, τρεις για κάθε φύλο και εννέα τυποποιημένοι για τον ρυθμό ανασυνδυασμού χάρτες, τρεις για κάθε φύλο και τρεις για τις μέσες τιμές των δύο φύλων. Υπάρχουν τρεις χάρτες για κάθε φύλο διότι πέρα από τον έλεγχο του φύλου, ερευνήθηκε και η ύπαρξη ή μη του αλληλομόρφου PRDM9 με 14 ή 15 επαναλήψεις των zinc fingers. Συνεπώς υπάρχει ένα χάρτης για χρωμοσώματα μητρικής σειράς ή χρωμοσώματα πατρικής χειράς, ένας για θήλεους ή άρρενες φορείς του αλληλομόρφου και ένας για θηλυκά ή αρσενικά άτομα τα οποία δεν φέρουν το συγκεκριμένο αλληλόμορφο.

Οι γενετικοί χάρτες είναι της παρακάτω μορφής:

1. chr	2. snp	3. pos	4. cM
5. chr1	6. rs4648527	7. 3497721	8. 0,121450
9. ...	10. ...	11. ...	12. ...

Πίνακας 4.1: Η μορφή ενός γενετικού χάρτη.

Όπου:

- *chr* = Το χρωμόσωμα που στεγάζεται ο γενετικός δείκτης.
- *snp* = Το αναγνωριστικό κάθε σημειακού πολυμορφισμού όπως αναγνωρίζεται από την βάση δεδομένων *dbSNP*.
- *pos* = Η θέση του SNP σε συντεταγμένες βάσης όπως εμφανίζονται στην έκδοση 36 (hg18) του ανθρώπινου γονιδιώματος.
- *cM* = Η απόσταση σε *centiMorgan* του συγκεκριμένου SNP από το προηγούμενό του.

Αξίζει να αναφερθεί πως τα SNPs του χρωμοσώματος X, των άκρων μήκους 2,5 Mb κάθε χρωμοσώματος και περιοχές οι οποίες παρέμεναν χωρίς σαφή αλληλούχιση στην έκδοση 36 του ανθρώπινου γονιδιώματος (10.254 τμήματα των 10 Kb, 8.891 εκ των οποίων κεντρομερικά) δεν έχουν ληφθεί υπόψιν στους υπολογισμούς (τιμή σε cM = NA).

Οι τυποποιημένοι χάρτες ανασυνδυασμού είναι της εξής μορφής:

13. chr	14. pos	15. seqbin	16. stdrate
17. chr1	18. 6000669	19. 1	20. 0,588856
21. ...	22. ...	23. ...	24. ...

Πίνακας 4.2: Η μορφή ενός τυποποιημένου χάρτη ανασυνδυασμού.

Όπου:

- *chr* = Το χρωμόσωμα που στεγάζεται το κάθε τμήμα μήκους 10 Kb.
- *pos* = Η θέση του κάθε τμήματος σε συντεταγμένες βάσης όπως εμφανίζονται στην έκδοση 36 (hg18) του ανθρώπινου γονιδιώματος. Η θέση αυτή αντικατοπτρίζει το κέντρο κάθε τμήματος.

- *seqbin* = Υποδεικνύει την ύπαρξη ή όχι μη αλληλουχημένων περιοχών σε κάθε τμήμα. Ίσο με 1 εάν δεν υπάρχουν αχαρτογράφητες περιοχές ή ίσο με 0 εάν υπάρχουν.
- *stdrate* = Ο τυποποιημένος ρυθμός ανασυνδυασμού κάθε τμήματος. Η τυποποίηση έγινε αναφορικά με τις πλήρως αλληλουχημένες περιοχές.

Microsoft Excel

Το πρόγραμμα ανάλυσης δεδομένων Excel χρησιμοποιήθηκε για τον καθαρισμό των παραπάνω δεδομένων. Έτσι, για τους γενετικούς χάρτες αρσενικών και θηλυκών ατόμων και μετά την αφαίρεση των μη απαραίτητων δεδομένων, ο αριθμός των SNPs μειώθηκε από τα 289.658 και 298.070 αντίστοιχα στα 276.280 και για τα δύο φύλα. Αναλυτικότερα, τα SNPs για κάθε χρωμόσωμα παρουσιάζονται στον κάτωθι πίνακα:

25. Chr1	26. 21.588
27. Chr2	28. 23.322
29. Chr3	30. 19.596
31. Chr4	32. 17.391
33. Chr5	34. 17.491
35. Chr6	36. 18.644
37. Chr7	38. 15.241
39. Chr8	40. 16.496
41. Chr9	42. 14.176
43. Chr10	44. 13.973
45. Chr11	46. 13.047
47. Chr12	48. 13.497
49. Chr13	50. 10.302
51. Chr14	52. 8.731
53. Chr15	54. 8.048
55. Chr16	56. 7.950

57. Chr17	58. 7.290
59. Chr18	60. 9.178
61. Chr19	62. 4.882
63. Chr20	64. 6.663
65. Chr21	66. 4.859
67. Chr22	68. 3.915

Πίνακας 4.3: Ο αριθμός των διαφορετικών SNP για κάθε χρωμόσωμα

Στους γενετικούς χάρτες (πίνακας 3.1) δημιουργήθηκε επιπρόσθετα μία νέα στήλη, η οποία είχε τις αθροιστικές τιμές των γενετικών αποστάσεων μεταξύ των SNPs. Με άλλα λόγια, το άθροισμα των αποστάσεων δύο γειτονικών SNPs και επακόλουθα το άθροισμα που προέκυπτε με το επόμενο στη σειρά SNP κ.ό.κ. Η στήλη ήταν της παρακάτω μορφής:

69. cM	70. addcM
71. 0,083175	72. 0,134111
73. 0,050936	74. 0,134115
75. 0,000004	76. 0,150207
77. 0,016092	78. 0,151958
79. 0,001751	80. 0,154545

Πίνακας 4.4: Η στήλη addcM (additive cM) στο πρώτο κελί της οποίας βρίσκεται το άθροισμα των πρώτων δύο κελιών της στήλης cM, ενώ στο δεύτερο κελί της το άθροισμα του πρώτου κελιού της στήλης addcM με το τρίτο κελί της στήλης cM κ.ό.κ.

Έτσι για κάθε χρωμόσωμα, το συνολικό μήκος που καλύπτουν οι γενετικοί δείκτες σε centiMorgan είναι:

81.	82. ♂	83. ♀
84. Chr1	85. 190,03	86. 333,97
87. Chr2	88. 175,19	89. 312,87
90. Chr3	91. 151,20	92. 268,749
93. Chr4	94. 134,96	95. 259,99
96. Chr5	97. 130,19	98. 248,97

99. Chr6	100.	116,17	101.	233,32
102. Chr7	103.	123,75	104.	220,96
105. Chr8	106.	108,86	107.	204,19
108. Chr9	109.	104,69	110.	187,88
111. Chr10	112.	114,66	113.	206,15
114. Chr11	115.	99,17	116.	188,58
117. Chr12	118.	115,40	119.	197,40
120. Chr13	121.	91,91	122.	146,77
123. Chr14	124.	85,67	125.	123,99
126. Chr15	127.	82,53	128.	139,33
129. Chr16	130.	87,97	131.	149,40
132. Chr17	133.	86,34	134.	152,89
135. Chr18	136.	73,71	137.	137,85
138. Chr19	139.	70,83	140.	115,23
141. Chr20	142.	56,29	143.	110,94
144. Chr21	145.	47,67	146.	67,03
147. Chr22	148.	39,86	149.	66,19

Πίνακας 4.5: Το συνολικό άθροισμα των γενετικών αποστάσεων κάθε γενετικού δείκτη, για κάθε χρωμόσωμα των δύο φύλων.

Οι χάρτες ανασυνδυασμού αποτελούνται από τμήματα μήκους 10 Kb και ο καθένας από αυτούς έχει έναν τυποποιημένο ρυθμό ανασυνδυασμού. Συνολικά υπάρχουν 254.700 τέτοια τμήματα και για να οριστεί ένα τμήμα ως θερμό σημείο ανασυνδυασμού (hotspot) η τιμή του ρυθμού ανασυνδυασμού του θα πρέπει να είναι μεγαλύτερη ή ίση με 10. Με την παράμετρο αυτή ως φίλτρο, τα τμήματα που πληρούν τις προδιαγραφές ανέρχονται σε 4.763 στα αρσενικά και 4.131 στα θηλυκά άτομα. Αναλυτικότερα για το κάθε χρωμόσωμα, τα hotspots παρουσιάζονται στον παρακάτω πίνακα:

150.	151.	♂	152.	♀
153. Chr1	154.	400	155.	319
156. Chr2	157.	357	158.	285

159.	Chr3	160.	314	161.	277
162.	Chr4	163.	286	164.	248
165.	Chr5	166.	298	167.	271
168.	Chr6	169.	224	170.	229
171.	Chr7	172.	239	173.	216
174.	Chr8	175.	210	176.	212
177.	Chr9	178.	232	179.	198
180.	Chr10	181.	227	182.	209
183.	Chr11	184.	213	185.	191
186.	Chr12	187.	208	188.	168
189.	Chr13	190.	192	191.	139
192.	Chr14	193.	201	194.	126
195.	Chr15	196.	199	197.	166
198.	Chr16	199.	204	200.	168
201.	Chr17	202.	171	203.	173
204.	Chr18	205.	149	206.	156
207.	Chr19	208.	89	209.	97
210.	Chr20	211.	147	212.	150
213.	Chr21	214.	100	215.	68
216.	Chr22	217.	102	218.	65

Πίνακας 4.6: Ο αριθμός των τμημάτων που χαρακτηρίζονται ως hotspots για τα χρωμοσώματα κάθε φύλου.

Επιπρόσθετα, για την περαιτέρω ανάλυση των δεδομένων δημιουργήθηκαν τα κατώτερα και τα ανώτερα όρια των τμημάτων από το σημείο που δίνεται από τα δεδομένα και αναλογεί στο κέντρο κάθε τμήματος αφαιρώντας 5000 ζεύγη βάσεων για το κατώτερο όριο και προσθέτοντας 5000 ζεύγη βάσεων για το ανώτερο όριο. Δηλαδή, δημιουργήθηκε ο παρακάτω πίνακας:

219.	chr	221.	lpos	222.	pos	223.	upos
224.	chr1	226.	6045669	227.	6050669	228.	6055669

229.	...	231.	...	232.	...	233.	...
------	-----	------	-----	------	-----	------	-----

Πίνακας 4.7: Το εύρος των 10 Kb που καλύπτει κάθε θερμό σημείο ανασυνδυασμού. Ως προς καθορίζεται το κατώτερο όριο ενώ ως προς το ανώτερο όριο.

RStudio

Η γλώσσα προγραμματισμού R αποτελεί μέρος του εγχειρήματος GNU και μοιάζει με τη διάλεκτο/περιβάλλον S. Είναι ένα διεπιστημονικό πρόγραμμα που δημιουργήθηκε στα Bell Laboratories από τους John Chambers κ.ά.. Στην παρούσα μελέτη χρησιμοποιήθηκε ως εργαλείο στατιστικής και βιοπληροφορικής ανάλυσης. Αποτελεί μία ολοκληρωμένη κονσόλα λογισμικού κατάλληλη για τη διαχείριση δεδομένων, την εκτέλεση υπολογισμών και τη γραφική απεικόνιση αποτελεσμάτων. Πιο αναλυτικά, περιλαμβάνει μονάδα διαχείρισης και αποθήκευσης δεδομένων, κονσόλα τελεστών υπολογισμών σε συστοιχίες συγκεκριμένων πινάκων, πληθώρα μεσάζοντων εργαλείων, που επιτρέπουν την ολοκληρωμένη ανάλυση δεδομένων, καθώς και μονάδα γραφικών που επιτρέπει τη γραφική απεικόνιση των δεδομένων τόσο στην οθόνη, όσο και σε έντυπη μορφή, μία εύχρηστη, απλή και αποτελεσματική γλώσσα προγραμματισμού που επιτρέπει τη χρήση μεταβλητών, βρόχων, επαναλαμβανόμενων λειτουργιών όπως ορίζονται από τους χρήστες, και μονάδες εισόδου και εξόδου. Αν και θα μπορούσε να θεωρηθεί πως πρόκειται για ένα σύστημα στατιστικής ανάλυσης, είναι ένα περιβάλλον στο οποίο εφαρμόζονται τεχνικές στατιστικής. Σημαντικό πλεονέκτημα της γλώσσας R που έχει συμβάλει στην ευρεία χρήση της είναι η πολυπληθής κοινότητα χρηστών στο διαδίκτυο. Το στοιχείο αυτό βοηθά τόσο στην εκμάθηση της γλώσσας, όσο και στην αντιμετώπιση των εμποδίων που προκύπτουν κατά τη χρήση της. Τέλος, τόσο η R, όσο και το Rstudio διατίθενται δωρεάν στο κοινό, επιτρέποντας τη χρήση από οποιοδήποτε άτομο το επιθυμεί. Αξίζει να σημειωθεί πως οι τεχνικές αυτές μπορούν εύκολα να επεκταθούν με την εγκατάσταση επιπλέον πακέτων, τα αποτελούν συλλογές εύχρηστων εντολών που έχουν δημιουργηθεί στο παρελθόν και αντικαθιστούν συγκεκριμένους κώδικες με ευρεία εφαρμογή (140). Η χρήση της R ήταν απαραίτητη για την συσχέτιση των SNPs με τα hotspots. Με άλλα λόγια, από τα 276.280 συνολικά SNPs κατά

μήκος του γονιδιώματος επιλέχθηκαν μόνο εκείνα που εντοπίζονται στα 4.763 και 4.131 θερμά σημεία αντίστοιχα. Αυτό έγινε με τις παρακάτω εντολές:

1. `hotspots <- read.delim("hotspotlimits.txt")`
2. `snps <- read.delim("snps.txt")`
3. `hotsnps <- merge(hotspots, snps)`
4. `result <- subset(hotsnps, pos >= lpos & pos <= upos)`
5. `write.table(result, "hotsnps.txt", sep = "\t", row.names = FALSE)`

Οι παραπάνω εντολές εκτελέστηκαν δύο φορές, μία για κάθε φύλο και τα αποτελέσματα μετονομάστηκαν σε "hotsnpsmale" και "hotsnpsfemale" αντίστοιχα. Αναλυτικότερα για το κάθε βήμα:

1. Δημιουργία ενός αντικειμένου "hotspots" το οποίο είναι ένα data frame και περιέχει τα δεδομένα του αρχείου "hotspotlimits.txt" τα οποία είναι της μορφής του πίνακα 3.7 (χωρίς την στήλη pos).
2. Ίδιο με το βήμα 1. Το data frame "snps" είναι της μορφής του πίνακα 3.1 (χωρίς την στήλη pos).
3. Το αντικείμενο "hotsnps" είναι ένα νέο data frame το οποίο είναι το αποτέλεσμα της συνένωσης των data frames "hotspots" και "snps".
4. Το αντικείμενο "result" αποτελείται μόνο από δεδομένα του "hotsnps" στα οποία η τιμή pos ενός SNP είναι μεγαλύτερη από το κατώτερο όριο ενός hotspot και μικρότερη από το ανώτερο όριο του ίδιου hotspot.
5. Η εντολή αυτή μας δίνει το data frame που προέκυψε σε αρχείο text που ονομάζεται "hotsnps(male ή female αντίστοιχα).txt"

Ο πίνακας είναι της παρακάτω μορφής:

234.	chr	235.	lpos	236.	upos	237.	snp	238.	pos
239.	chr1	240.	6245669	241.	6255669	242.	rs3789512	243.	6252602
244.	chr1	245.	7165669	246.	7175669	247.	rs4908446	248.	7168256
249.	...	250.	...	251.	...	252.	...	253.	...

Πίνακας 4.8: Ο παραπάνω πίνακας μας δείχνει ποιό SNP βρίσκεται στα όρια ενός hotspot, καθώς και την ακριβή θέση αυτού.

Με τον τρόπο αυτόν εντοπίζονται 9.413 SNPs σε 4.763 hotspot στα άρρενα άτομα και 8.776 SNPs σε 4.131 hotspot στα θήλεα άτομα.

Στην συνέχεια, στόχος ήταν η εύρεση της απόστασης μεταξύ κάθε διαδοχικού hotspot τόσο σε ζεύγη βάσεων όσο και σε centiMorgan. Οι εντολές που παρατίθενται αφορούν τα θηλυκά άτομα, αλλά είναι οι ίδιες που χρησιμοποιήθηκαν και για τα αρσενικά άτομα, με μόνη διαφορά την ονοματολογία των αντικειμένων.

1. `hotsnpsfemale <- read.delim("hotsnpsfemale.txt")`
2. `gmapfemale <- read.delim("femgmap.txt")`
3. `hotgmapfemale <- merge(gmapfemale, hotsnpsfemale, by = "snp")`
4. `hotgmapfemale$cM = NULL`
5. `hotgmapfemale$chr.y = NULL`
6. `hotgmapfemale$pos.y = NULL`
7. `hotgmapfemale$upos = NULL`

8. `hotgmapfemale$lpos = NULL`
9. `colnames(hotgmapfemale) <- c("snp", "chr", "pos", "addcM")`
10. `hotgmapfem <- hotgmapfemale[order(hotgmapfemale$chr, hotgmapfemale$pos, decreasing = FALSE),]`
11. `rownames(hotgmapfem) <- 1:nrow(hotgmapfem)`
12. `result <- as.data.frame(lapply(hotgmapfem[3:4], diff, lag=1))`
13. `colnames(result) <- c("difference(bp)", "difference(cM)")`
14. `write.table(result, "differencesfemale.txt", sep = "\t", row.names = FALSE)`

Το αποτέλεσμα των παραπάνω παρουσιάζεται στον παρακάτω πίνακα:

254.	difference(bp)	255.	difference(cM)
256.	915654	257.	1,510025
258.	26581	259.	0,026511
260.	215499	261.	0,580570
262.	...	263.	...

Πίνακας 4.9: Τα κελιά της πρώτης γραμμής δείχνουν την απόσταση του δεύτερου SNP από το πρώτο σε ζεύγη βάσεων και σε centiMorgan. Τα κελιά της δεύτερης γραμμής την απόσταση του τρίτου SNP από το δεύτερο κ.ό.κ.

Η σημασία των εντολών αναλυτικότερα:

1 & 2: Όπως και οι προηγούμενες, φόρτωση των δεδομένων ως data frames. Το data frame “gmarfemal” είναι της μορφής του πίνακα 3.1. Επιπρόσθετα, περιέχει και την στήλη “addcM” του πίνακα 3.4.

3: Συνένωση των δύο data frames ως προς την στήλη “snr”. Έτσι, από τα 276.280 SNPs του “gmarfemal” κρατάμε μόνο τα 8.776 που βρίσκονται σε hotspots καθώς και την στήλη “addcM” που μας ενδιαφέρει.

4 - 8: Με τις εντολές αυτές απαλείφονται οι ολόιδιες στήλες “chr” και “pos” καθώς και στήλες οι οποίες δεν μας ενδιαφέρουν όπως η “cM”, η “uros” και η “lros” από το data frame “hotgmarfemal”.

9: Αλλαγή των ονομάτων των στηλών σε “snr”, “chr”, “pos” και “addcM”.

10: Δημιουργία νέου data frame “hotgmarfem”, το οποίο αποτελείται από τα δεδομένα του “hotgmarfemal” σε αύξουσα σειρά για τις στήλες “chr” και “pos”.

11: Σωστή αρίθμηση των γραμμών για περαιτέρω ανάλυση καθώς η εντολή 10 βάζει τα δεδομένα σε αύξουσα σειρά, χωρίς όμως να αλλάξει την πραγματική τους θέση μέσα στο data frame. Δηλαδή, παρόλο που δείχνει μία τιμή πρώτη σύμφωνα με τα κριτήρια που τέθηκαν, η θέση που καταλαμβάνει στο data frame με την οποία και αναγνωρίζεται από επακόλουθες εντολές μπορεί να είναι οποιαδήποτε (π.χ 342).

12: Με την εντολή αυτή, δημιουργείται ένα καινούριο data frame, στο οποίο περιέχονται οι διαφορές δύο διαδοχικών τιμών (εντολή diff) των στηλών “pos” και “addcM”.

13: Αλλαγή των ονομάτων των στηλών σε difference(bp) και difference(cM).

14: Εξαγωγή του αποτελέσματος στο αρχείο “differencesfemale.txt”.

Σημαντικό είναι να αναφερθεί ότι επειδή τα δεδομένα σε κάθε data frame είναι γραμμικά, δηλαδή όταν τελειώνουν τα δεδομένα για το ένα χρωμόσωμα ξεκινάνε αμέσως τα δεδομένα για το άλλο, στα σημεία αυτά οι διαφορές σε ζεύγη βάσεων και σε centiMorgan θα βγαίνουν αρνητικές καθώς θα αφαιρείται η αρχική θέση ενός χρωμοσώματος από την τελική θέση του προηγούμενου. Για παράδειγμα στην μετάβαση από το χρωμόσωμα 18 στο 19,

μετρείται η απόσταση του 1^{ου} SNP σε hotspot του χρωμοσώματος 19 από το τελευταίο SNP σε hotspot του χρωμοσώματος 18, δίνοντας μια απόσταση -65656113 ζευγών βάσεων και -124.699 centiMorgan. Επειδή οι τιμές αυτές δεν έχουν νόημα αφαιρούνται και αντικαθίστανται με NA καθώς δεν υπάρχει SNP μετά από αυτά ώστε να γίνει μέτρηση.

Σε επόμενο στάδιο, τα παραπάνω αποτελέσματα πρέπει να τροποποιηθούν καταλλήλως, καθώς υπάρχουν περισσότερα από ένα SNP σε κάθε hotspot, με αποτέλεσμα ο υπολογισμός της μέσης απόστασης τόσο σε ζεύγη βάσεων όσο και σε centiMorgan να βγεί μικρότερος από τον αναμενόμενο καθώς θα λαμβάνονται υπόψιν και τα SNP που έχουν απόσταση ελάχιστων βάσεων μεταξύ τους και η απόσταση σε centiMorgan μεταξύ τους είναι 0. Ακολουθήθηκαν τα παρακάτω βήματα:

1. `dif <- read.delim("differencesfemale.txt")`
2. `dif1 <- dif[order(difchr, difpos, decreasing = TRUE)]`
3. `dif2 <- dif1[!duplicated(dif1$addcM),]`
4. `hotspots <- read.delim("hotspotlimitsfem.txt")`
5. `hotdif2 <- merge(dif2, hotspots)`
6. `hotdif3 <- subset(hotdif2, pos >= lpos & pos <= upos)`
7. `hotdif4 <- hotdif3[!duplicated(hotdif3$upos),]`
8. `with(hotdif4, mean(hotdif4$difference.bp. [hotdif4$chr == "chr1"], na.rm = TRUE)`
9. `with(hotdif4, mean(hotdif4$difference.cM. [hotdif4$chr == "chr1"], na.rm = TRUE)`

Αναλυτικότερα:

1. Δημιουργία ενός data frame με τα αποτελέσματα των μέσων αποστάσεων για 8776 SNP.
2. Ταξινόμηση του data frame "dif" σε ένα νέο με όνομα "dif1", στο οποίο η στήλη "chr" και "pos" βρίσκονται σε φθίνουσα σειρά.

Η ταξινόμηση αυτή είναι απαραίτητη για την εκτέλεση της επόμενης εντολής.

3. Απαλοιφή των γραμμών στις οποίες η τιμή της στήλης addcM είναι ίδια με αυτή κάποιας άλλης γραμμής. Με άλλα λόγια δύο γειτονικά SNP που απέχουν 0 cM μεταξύ τους θα έχουν ίδια τιμή addcM, όπως τα SNP rs618962 και rs1864709 στις θέσεις 132385430 και 132389259 του χρωμοσώματος 9 που έχουν την ίδια τιμή addcM και είναι 171,0835.

Αυτό έχει ως αποτέλεσμα η διαφορά τους σε cM να είναι ίση με 0. Με την εντολή αυτή παραμένει η πρώτη ενώ απαλείφεται η δεύτερη ίδια τιμή που εντοπίζεται. Έτσι μένει μόνο το SNP το οποίο στη θέση 171,0835 cM απέχει από το επόμενο γειτονικό SNP απόσταση μεγαλύτερη του 0. Στην συγκεκριμένη περίπτωση, το γειτονικό του rs1864709 (αφού απαλείφθηκε το rs618962) είναι το rs10901125 από το οποίο απέχει 0,125720 cM. Εάν δεν είχε πραγματοποιηθεί ταξινόμηση του data frame σε φθίνουσα σειρά, τότε θα απαλειφόταν το άλλο SNP της θέσης 171,0835 cM και θα παρέμενε η τιμή 0 στην στήλη "difference.cM". Με την εντολή αυτή, ο αριθμός των SNP μείωθηκε από τα 8776 στα 6471

4. Δημιουργία data frame με τα όρια των hotspot.
5. Δημιουργία data frame με το όνομα "hotdif2" που αποτελείται από την συνένωση των δύο παραπάνω data frames, "dif2" και "hotspots" .
6. Δημιουργία ενός υποσυνόλου του "hotdif2" με όνομα "hotdif3" στο οποίο υπάρχουν μόνο τα SNPs που εντοπίζονται στα όρια ενός hotspot.
7. Τέλος, δημιουργείται ένα υποσύνολο του "hotdif3" ονόματι "hotdif4", στο οποίο έχουν απαλειφθεί οι γραμμές στις οποίες εντοπίζονται ίδιες τιμές στην μία από τις δύο στήλες των ορίων των hotspot. Με τον τρόπο αυτόν, δεν θα υπάρχουν δύο SNPs σε ένα hotspot, με αποτέλεσμα η απόσταση από δύο γειτονικά SNP να είναι η ίδια με την απόσταση δύο γειτονικών hotspot. Συνολικά μετρήθηκαν 3.717 στα θηλυκά και 4.075 στα αρσενικά αποστάσεις διαδοχικών SNP και επομένως hotspot.
8. Εύρεση της μέσης τιμής της απόστασης σε ζεύγη βάσεων δύο γειτονικών SNP για το χρωμόσωμα 1 χωρίς να λαμβάνονται υπόψιν οι τιμές NA.

9. Έυρεση της μέσης τιμής της απόστασης σε centiMorgan δύο γειτονικών SNP για το χρωμόσωμα 1 χωρίς να λαμβάνονται υπόψιν οι τιμές NA.

Τα βήματα 8 και 9 επαναλήφθηκαν για τα υπόλοιπα χρωμοσώματα, ενώ η όλη διαδικασία επαναλήφθηκε και για τα δεδομένα των αρσενικών ατόμων. Οι μέσες αποστάσεις δύο γειτονικών hotspots που προέκυψαν παρατίθενται στην επόμενη ενότητα των αποτελεσμάτων. Η συσχέτιση μεταξύ της φυσικής και της γενετικής απόστασης των γενετικών δεικτών που χρησιμοποιήθηκαν για την εύρεση των μέσων αποστάσεων σε κάθε χρωμόσωμα, έγινε με την βοήθεια του πρόσθετου πακέτου "MareyMap". Με τις δύο παρακάτω εντολές:

1. library(MareyMap)
2. startMareyMapGUI()

Ανοίγει ένα γραφικό περιβάλλον χρήστη στο οποίο εισάγονται δεδομένα της μορφής:

264.	"HomoSapiens"	265.	"chr1"	266.	"rs12117802"	267.	"239931842"	268.
269.	"HomoSapiens"	270.	"chr1"	271.	"rs663814"	272.	"239876209"	273.
274.	"HomoSapiens"	275.	"chr1"	276.	"rs602676"	277.	"239250516"	278.
279.	"HomoSapiens"	280.	"chr1"	281.	"rs605792"	282.	"239242378"	283.
284.	...	285.	...	286.	...	287.	...	288.

Πίνακας 4.10: Η απαραίτητη μορφή που πρέπει να έχουν τα δεδομένα ώστε να τα επεξεργαστεί ο αλγόριθμος του προγράμματος Marey Map.

Έπειτα, με την μέθοδο *Loess* (Lowess for LOcally WEighted Scatterplot Smoothing) που παρέχεται από την R, μπορούν να υπολογιστούν τοπικά οι ρυθμοί ανασυνδυασμού, προσαρμόζοντας μία πολυωνυμική καμπύλη 1^{ου} ή 2^{ου} βαθμού. Τα γραφήματα για κάθε φύλο και για κάθε χρωμόσωμα παρουσιάζονται στην ενότητα των αποτελεσμάτων.

4.2 ΕΥΡΕΣΗ ΤΗΣ ΣΥΝΤΗΡΗΜΕΝΗΣ ΜΟΡΦΗΣ ΤΟΥ ΑΛΛΗΛΟΥΧΙΑΚΟΥ ΜΟΤΙΒΟΥ IRs ΚΑΙ ΤΩΝ ΜΟΡΦΩΝ ΠΟΥ ΑΥΤΟ ΕΜΦΑΝΙΖΕΤΑΙ ΩΣ HOLLIDAY JUNCTION

Μετά από εκτενής αναζήτηση σε βάσεις δεδομένων όπως η NCBI και άλλες μικρότερης έκτασης βρέθηκε ότι οι δομές Holliday αποτελούνται συνήθως από ανάστροφες επαναλήψεις (IR) ενός συντηρημένου δεκανουκλεοτιδίου που έχει την μορφή d[CCnnnN(6)N[7]C[8]GG]. Όπως αναφέρθηκε και στο θεωρητικό κομμάτι της μελέτης η τριπλέτα στις θέσεις N(6)N(7)C(8) είναι υπεύθυνη για την σταθεροποίηση των εν λόγω δομών και συνήθως εντοπίζεται ο πυρήνας ACC. Άλλες μορφές που μπορεί να πάρει ο πυρήνας είναι οι N(6)N(7)C(8)= GCC, ATC, CCC αλλά η σταθεροποίηση των αλληλουχιών σε δομές Holliday προϋποθέτει συγκεκριμένες συνθήκες, όπως η παρουσία δισθενών ιόντων όπως πχ. Ca και Mg. Επίσης, αποδείχθηκε ότι τροποποιήσεις στις μερικές βάσεις του δεκανουκλεοτιδίου όπως πχ. η αντικατάσταση των τερματικών βάσεων G-C σε T-A ή αναντιστοιχία μεταξύ A-G στις ανάστροφες ακολουθίες πάλι οδηγεί στον σχηματισμό των επιθυμητών δομών. Συνεπώς οι αλληλουχίες που εντοπίστηκαν ότι σχηματίζουν HJ μετά την διαξαγωγή της βιβλιογραφικής έρευνας, είναι οι εξής :

1. d(CCGGG**ACCGG**)
2. d(CCGGT**ACCGG**)
3. d(CCGGG**CCCGG**)
4. d(CCGGC**GCCGG**)
5. d(TCGGT**ACCGA**)
6. d(TCCGT**ACCGA**)

4.2.1 ΕΥΡΕΣΗ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΜΟΤΙΒΩΝ IRs-HJs ΣΕ ΟΛΟ ΤΟ ΜΗΚΟΣ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R

Τα πακέτα που εγκαταστάθηκαν και χρησιμοποιήθηκαν για την διεξαγωγή της μελέτης είναι τα εξής :

1. AnnotationDbi_1.16.19

2. Biobase_2.14.0
3. biomaRt_2.10.0
4. Biostrings_2.22.0
5. bitops_1.0-4.1
6. BSgenome.Hsapiens.UCSC.hg18_1.3.17
7. BSgenome_1.22.0
8. GenomeGraphs_1.14.0
9. GenomicRanges_1.6.7
10. IRanges_1.12.6
11. RCurl_1.91-1.1
12. XML_3.9-4.1
13. BiocGenerics

Στη συνέχεια έγινε αναζήτηση των επιθυμητών αλληλουχιών με την εισαγωγή διάφορων εντολών οι οποίες θα παρουσιαστούν στη συνέχεια του κειμένου. Η διαδικασία αναζήτησης διεξήχθη 2 φορές για κάθε χρωμόσωμα του ανθρώπινου γονιδιώματος. Οι εντολές που δώθηκαν κατά τη διαδικασία της πρώτης αναζήτησης είναι οι εξής:

1. library(BSgenome.Hsapiens.UCSC.hg18)
2. x ← Hsapiens\$chr**
3. active(masks(x)) ← TRUE
4. x
5. m1 ← matchPattern("*****", x)
6. as.matrix(m1)

1^η εντολή → Δόθηκε για την φόρτωση ολόκληρου του ανθρώπινου γονιδιώματος στο R project

2^η εντολή → Θέτω ως μεταβλητή X το επιθυμητό χρωμόσωμα. ** = 1-22 ή X ή Y

3^η εντολή → Ενεργοποίηση 4 ειδών επικαλύψεων μέσα στις οποίες εντάσσονται και επικαλύψεις που δίνονται από το Repeat Masker (Το Repeat Masker αναλύεται και πιο κάτω και αποτελεί ένα πρόγραμμα το οποίο συμβάλλει στην αλληλούχιση διάσπαρτων και χαμηλής πολυπλοκότητας επαναλήψεων του DNA)

4^η εντολή → Παρουσίαση του αποτελέσματος συνδυασμού των 3^{ων} προηγούμενων εντολών

5^η εντολή → Εντολή για την διαδικασία εύρεσης των συντεταγμένων της επιθυμητής αλληλουχίας στο δωθέν χρωμόσωμα με *****= οποιαδήποτε από τις 7 ακολουθίες που σχηματίζουν HJ

6^η εντολή → Παρουσίαση τελικών αποτελεσμάτων

Ακολουθεί παράδειγμα για την κατανόηση της λειτουργίας του προγράμματος.

```
> library(BSgenome.Hsapiens.UCSC.hg18)
> x <- Hsapiens$chr1
> active(masks(x)) <- TRUE
> x
247249719-letter "MaskedDNAString" instance (# for masking)
seq: #####...#####
masks:
  maskedwidth maskedratio active names desc
1 22250000 0.089989991 TRUE AGAPS assembly gaps
2 0 0.000000000 TRUE AMB intra-contig ambiguities (empty)
3 109628227 0.443390704 TRUE RM RepeatMasker
4 1513562 0.006121592 TRUE TRF Tandem Repeats Finder [period<=12]
all masks together:
  maskedwidth maskedratio
131963053 0.5337238
> m1 <- matchPattern("ccggtaccgg", x)
> as.matrix(m1)
      [,1] [,2]
[1,] 43409825 10
[2,] 45249263 10
[3,] 93960333 10
[4,] 154894092 10
[5,] 177318269 10
[6,] 199566988 10
[7,] 242280213 10
```

Εικόνα 4.1 Η παραπάνω εικόνα αποτελεί παράδειγμα αναζήτησης της αλληλουχίας d(CCGGTACCGG) που αποτελεί μία από τις 7 αλληλουχίες που σχηματίζουν HJ, στο χρωμόσωμα 1 του ανθρώπινου γονιδιώματος. Παρατηρούμε ότι η συγκεκριμένη αλληλουχία βρίσκεται 7 φορές στο χρωμόσωμα 1, όπου δίνονται και οι συντεταγμένες.

Έτσι λοιπόν επαναλήφθηκε αυτή η διαδικασία και για τις 6 αλληλουχίες για κάθε χρωμόσωμα του ανθρώπινου γονιδιώματος, και έγινε συλλογή των τελικών δεδομένων. Στην δεύτερη διαδικασία αναζήτησης, δόθηκαν ακριβώς οι ίδιες εντολές με την μόνη διαφορά, ότι ανάμεσα στην 3^η και την 4^η εντολή προστέθηκε μία νέα εντολή η οποία είναι η εξής:

1. library(BSgenome.Hsapiens.UCSC.hg18)
2. x <- Hsapiens\$chr**
3. active(masks(x)) <- TRUE
4. active(masks(x)) [3] <- FALSE
5. x
6. m1 <- matchPattern("*****", x)
7. as.matrix(m1)

Σε αυτήν την συγκεκριμένη περίπτωση η νέα εντολή που δώσαμε απενεργοποιεί τις επικαλύψεις που δίνονται από το πρόγραμμα RepeatMasker. Οπότε τώρα η αλληλουχία μας μπορεί να εντοπιστεί και σε περιοχές επαναλαμβανόμενων ακολουθιών. Η εικόνα που ακολουθεί βοηθάει στην κατανόηση της διαδικασίας. Η δεύτερη αναζήτηση έγινε πάλι και για τις 6 αλληλουχίες σε όλα τα χρωμοσώματα του ανθρώπινου γονιδιώματος. Με συλλογή και σύγκριση των νέων δεδομένων με τα αρχικά, παρατηρούμε ότι προκύπτουν παραπάνω παρατηρήσεις από τις αρχικές, πράγμα το οποίο δηλώνει ότι οι επιπλέον παρατηρήσεις είναι στην ουσία οι συντεταγμένες των χρωμοσωμάτων στις οποίες ΗJ βρίσκονται σε επαναλαμβανόμενες αλληλουχίες, συνεπώς και σε ρετρομεταθετά στοιχεία που αναζητάμε.

```
> library(BSgenome.Hsapiens.UCSC.hg18)
> x <- Hsapiens$chr1
> active(masks(x)) <- TRUE
> active(masks(x))[3] <- FALSE
> x
> 247249719-letter "MaskedDNAString" instance (# for masking)
seq: #####...#####
mask:
  maskedwidth maskedratio active names desc
1 22250000 0.089989991 TRUE AGAPS assembly gaps
2 0 0.000000000 TRUE AMB intra-contig ambiguities (empty)
3 109628227 0.443390704 FALSE RM RepeatMasker
4 1513562 0.006121592 TRUE TRF Tandem Repeats Finder [period<=12]
all masks together:
  maskedwidth maskedratio
131963053 0.5337238
all active masks together:
  maskedwidth maskedratio
23763562 0.09611158
> m1 <- matchPattern("ccggtaccgg", x)
> as.matrix(m1)
      [,1] [,2]
[1,] 42299510 10
[2,] 43409825 10
[3,] 45249263 10
[4,] 61043799 10
[5,] 93960333 10
[6,] 119678276 10
[7,] 154894092 10
[8,] 177318269 10
[9,] 199566988 10
[10,] 205339389 10
[11,] 242280213 10
> |
```

Εικόνα 4.2. Η παραπάνω εικόνα αποτελεί παράδειγμα της προηγούμενης ακολουθίας στο ίδιο χρωμόσωμα κατά την διαδικασία της δεύτερης αναζήτησης. Πράγματι ενώ πριν προέκυψαν 7 παρατηρήσεις, τώρα βλέπουμε εικόνα με τις ίδιες συντεταγμένες αλλά με τέσσερις επιπλέον. Οι τέσσερις επιπλέον αποτελούν συντεταγμένες όπου ΗJ σε επαναλαμβανόμενες αλληλουχίες

4.2.2 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ ΤΟΥ HARMAP PROJECT.

Το διεθνές HarMap project αποτελεί μια συνεργασία επιστημόνων και φορέων χρηματοδότησης από Καναδά, Κίνα, Ιαπωνία, Νιγηρία, Ηνωμένο Βασίλειο και τις Ηνωμένες Πολιτείες, στην προσπάθεια δημιουργίας μίας κοινής διαδικτυακής βάσης δεδομένων, η οποία αρχικά θα βοηθούσε τους ερευνητές να βρουν γονίδια που σχετίζονται με ασθένειες, αλλά σίγουρα παίρνει και άλλες προεκτάσεις. Ο στόχος της προαναφερθείσας βάσης δεδομένων είναι η ανάπτυξη ενός χάρτη του απλοειδούς ανθρώπινου γονιδιώματος, ο οποίος θα περιγράφει τα κοινά pattern των παραλλαγών της ανθρώπινης ακολουθίας DNA. Το HarMap αναμένεται να αποτελέσει τον βασικό παράγοντα έρευνας στην προσπάθεια εύρεσης γενετικών τόπων που σχετίζονται με την υγεία, ασθένειες καθώς και τρόπους απόκρισης σε φάρμακα και περιβαλλοντικούς παράγοντες (141). Στην παρούσα εργασία χρησιμοποιήθηκε για τον εντοπισμό και την ταυτοποίηση των IRs-HJs και των ευρύτερων αλληλουχιών τους (θα αναλυθεί πιο κάτω) σε κάθε χρωμόσωμα του ανθρώπινου γονιδιώματος, σε μορφή κατάλληλη (μορφή FASTA) η οποία συνέβαλλε στην συνέχιση της έρευνας.

Μετά την σύγκριση των δεδομένων που αντλήθηκαν από την R, όπως αναφέρθηκε και στο προηγούμενο απόσπασμα συλλέχθηκαν σε μορφή text οι παρατηρήσεις που δηλώνουν την ύπαρξη HJ σε επαναλαμβανόμενες αλληλουχίες. Έπειτα, ακολούθησε η διαδικασία ταυτοποίησης των συλλεχθέντων στοιχείων στο διαδικτυακό πρόγραμμα HarMap σύμφωνα με την εξής διαδικασία:

1. Η διαδικασία αναζήτησης διεξήχθη στη ενότητα (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28_B36/).
2. Έγινε μία πρώτη ταυτοποίηση και έλεγχος και των 6 αλληλουχιών που σχηματίζουν HJ, σε όλα τα χρωμοσώματα σε όλα τα σημεία που προέκυψαν από τις προηγούμενες συγκρίσεις. Η κάθε αναζήτηση δεν αφορούσε μόνο τα 10 νουκλεοτίδια που σχηματίζουν τους Holliday κόμβους αλλά και τα 40 νουκλεοτίδια που ακολουθούσαν, καθαρά για την καλύτερη αναζήτηση στην μετέπειτα επεξεργασία τους στο RepeatMasker και για την διερεύνηση της φύσης των γειτονικών περιοχών των Holliday Junctions. Πχ για την αλληλουχία d(CCGATATCGG) η αναζήτηση είχε την μορφή 5'- d(CCGATATCGG)-40nt- 3'

Search
 Help links: [- LD -](#) [- TopSNPs -](#) [- Phased Haplotype -](#) [- Genotype data -](#) [- Frequency data -](#) [- Symbols and colours used -](#)

Landmark or Region:

Data Source

Reports & Analysis:

Scroll/Zoom: Flip

Εικόνα 4.3. Παράδειγμα παρουσίασης της φόρμας αναζήτησης

3. Απόκτηση σε μορφή FASTA όλων των αλληλουχιών που σχηματίζουν HJ , με σημείο αναφοράς των συντεταγμένων τους σε κάθε χρωμόσωμα και συλλογή των δεδομένων σε φάκελο .txt.

```

>chr6:1555582,1555632
ccgggccccgggactcgggactcgggcccggccggcgcgggcgcgggcccccggcccccggagc
>chr6:3417408,3417458
ccgggccccgggtgggctcacacacctgtaatcccagcatttttgggaggccaaggt
>chr6:7487961,7488011
ccgggccccgggtctcgaagttcgggccgggacgtgcctgttacacctcggtgoc
>chr6:10916003,10916053
ccgggccccgggtgggctcacgcctgtaatcccagcactatgcataggaggcca
>chr6:11167899,11167949
ccgggccccgggtgggctcatgtctgttaatcccggcacttttgggaggccgagac
>chr6:12117428,12117478
ccgggccccgggccgggcccggccgcgctggggaaccggagccggtggacgcagcggc
>chr6:16834795,16834845
ccgggccccggtagctcacacacctatagttcccagcacttttgggaggccgaggc
>chr6:18138868,18138918
ccgggccccggccagggttaggaaaacatttctaagacaagtcacaaaaggca
>chr6:19573074,19573124
ccgggccccggccctagcagtaacttttatagcaaatgggacccttggatgata
>chr6:20408340,20408390
ccgggccccgggtgggtggggcgtttagtagtcccagctactcaggagggtgaggc
>chr6:21304008,21304058
ccgggccccggccactctggagatttttaaggcccttgccaatgagactggg
>chr6:21495373,21495423
ccgggccccgggtgggctcacgcccataatcccagcaattggtgggttacgccc
>chr6:33487638,33487688
ccgggccccgggtgggctcacgcctgtaatcccagcacttttgggaagctgaggt
>chr6:33658667,33658717
ccgggccccgggtgggctctcgcctgtaatcccagcacttttgggaggccgaggc
>chr6:34772416,34772466
ccgggccccggggaccctggaggggggccgcgctgctagcttagcttccgacccc
  
```

Εικόνα 4.4. Παράδειγμα το οποίο παρουσιάζει την FASTA μορφή της αλληλουχίας d(CGGGCCCGG)-40nt στο χρωμόσωμα 6.

Έτσι λοιπόν με τη βοήθεια του HapMap project και της διμορφικού χαρακτήρα αναζήτησης, έγινε συγκέντρωση και των 6 αλληλουχιών που σχηματίζουν Holliday Junctions στη FASTA μορφή τους και με τις συντεταγμένες σε κάθε χρωμόσωμα.

4.2.3 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ REPEATMASKER ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Το RepeatMasker αποτελεί ένα διαδικτυακό πρόγραμμα το οποίο είναι ικανό να αλληλουχίσει διάσπαρτες και χαμηλής πολυπλοκότητας επαναλήψεις στο ανθρώπινο γονιδίωμα. Τα αποτελέσματα που προκύπτουν από την επεξεργασία αλληλουχιών στο RepeatMasker αποτελούν μία λεπτομερή ποσοστιαία παρουσίαση των επαναλήψεων που υπάρχουν στην αλληλουχία που υποβλήθηκε προς επεξεργασία (query-sequence) ή μία τροποποιημένη παρουσίαση σε FASTA μορφή, στην οποία οι επαναλήψεις επικαλύπτονται και αντικαθίστανται από γράμματα N. Οι αλληλουχιακές συγκρίσεις στο RepeatMasker εκτελούνται από το πρόγραμμα `cross_match`, μια αποτελεσματική εφαρμογή του αλγορίθμου Smith-Waterman, ή από τον αλγόριθμο WU-BLAST. Στην παρούσα εργασία έγινε αναζήτηση μέσω του αλγορίθμου `abblast` (WU-BLAST) (93).

Τα βασικά κριτήρια της αναζήτησης ήταν τα εξής:

Search Engine: `abblast`

Speed Sensitivity: `default`

DNA source: `Human`

Return Format: `html file`

Τα επίπλεον κριτήρια (`advanced`) της αναζήτησης είναι τα εξής :

Masking Options: `Repetitive sequences replaces by strings of N`

Repeat Options: `Mask interspersed and simple repeats`

Artifact Check: `Report E.coli insertion elements artifacts`

Matrix: `RepeatMasker choice`

4.2.4 ΦΥΛΟΓΕΝΕΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΜΟΝΤΕΛΟ ΑΝΤΙΚΑΤΑΣΤΑΣΗΣ ΝΟΥΚΛΕΟΤΙΔΙΩΝ

Η πολλαπλή ευθυγράμμιση επαναλαμβανόμενων στοιχείων (TEs) που περιέχουν αλληλουχίες ευάλωτες στη δημιουργία HJs πραγματοποιήθηκε με τον αλγόριθμο Clustal W2 (142). Η φυλογενετική ανάλυση έγινε σύμφωνα με τη μέθοδο μέγιστης πιθανότητας με βάση το μοντέλο Tamura-Nei αντικατάστασης νουκλεοτιδίων (143). Ο έλεγχος bootstrap (500 επαναλήψεις) εφαρμόστηκε για να επιβεβαιωθεί η φυλογένεση (144) και όλες οι εξελικτικές αναλύσεις διεξήχθησαν με το λογισμικό ανάλυσης μοριακής εξελικτικής γενετικής (MEGA5) (145).

4.2.5 ΚΑΤΑΝΟΜΗ ΤΩΝ IRs ΠΟΥ ΣΧΗΜΑΤΙΖΟΥΝ HOLLIDAY JUNCTIONs ΣΕ TSS ΚΑΙ ΓΟΝΙΔΙΑ ΜΕ ΠΑΡΑΜΕΤΡΟ ΣΤΑΤΙΣΤΙΚΗΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Η σχετική κατανομή των HJs σε TSS των ανθρώπινων γονιδίων αναλύθηκε από το GREAT (146). Τα δεδομένα συσχέτισης επιβεβαιώθηκαν και εξετάστηκαν περαιτέρω από το UCSC Genome Browser.

Η στατιστική ανάλυση της κατανομής των IRs-HJs σε ανθρώπινα χρωμοσώματα διεξήχθη στο στατιστικό λογισμικό EasyFit 5.5 καθώς και με στατιστικό αλγόριθμο στη γλώσσα προγραμματισμού R. Η δοκιμή χ^2 χρησιμοποιήθηκε για να ελέγξει τη σημασία του αριθμού των IRs-HJs που σχετίζονται με TEs και του αριθμού των IRs-HJs κοντά στα γονίδια, σε αντίθεση με τους αναμενόμενους αριθμούς.

4.2.6 ΒΙΟΠΛΗΡΟΦΟΡΙΚΑ ΕΡΓΑΛΕΙΑ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΑΛΛΗΛΟΥΧΙΩΝ IRS-HJS ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΕΝΤΟΣ ΤΩΝ ΟΡΙΩΝ ΣΥΝΔΡΟΜΩΝ ΜΙΚΡΟΕΛΛΕΙΨΕΩΝ ΚΑΙ ΜΙΚΡΟΔΙΠΛΑΣΙΑΣΜΩΝ

Για να προσδιορίσουμε με ακρίβεια τις συντεταγμένες των MMS στο ανθρώπινο γονιδίωμα, έγινε «εξόρυξη» δεδομένων για όλα τα σύνδρομα μικροελλείψεων και μικροδιπλασιασμών (107, 121). Αυτό οδήγησε στη δημιουργία 2 αρχείων σε bed format για όλα τα MMS που έχουν καταχωρηθεί σε δημόσιες βάσεις δεδομένων και επιστημονικά άρθρα. Χρησιμοποιώντας το εργαλείο liftover στο πρόγραμμα περιήγησης UCSC, μετατράπηκε το σύνολο των συντεταγμένων των συνδρόμων από την έκδοση hg18 στην έκδοση hg19 του γονιδιώματος.

Οι permutation και overlap αναλύσεις μεταξύ των IRs-HJs και των συνδρόμων (MMS), έγιναν στην R 3.5.1. Τα πακέτα που χρησιμοποιήθηκαν στην R για τη διεξαγωγή της έρευνας είναι τα εξής:

- BiocManager
- regioneR
- biomaRt
- ggplot2
- magrittr
- dplyr
- digest
- rlang
- AnnotationDbi
- Biobase
- Biostrings
- bitops
- BSgenome.Hsapiens.UCSC.hg19/38
- BSgenome
- GenomeGraphs
- GenomicRanges
- IRanges_1.12.6
- RCurl

- XML
- BiocGenerics

Χρησιμοποιήθηκε κυρίως το πακέτο `Regioner` που δημιουργήθηκε για την αντιμετώπιση αυτού του προβλήματος και παρέχει λειτουργίες για την στατιστική αξιολόγηση των συσχετίσεων μεταξύ των συνόλων μας με την εφαρμογή δοκιμών τυχαιοποίησης (147). Το πλαίσιο δοκιμών τυχαιοποίησης (`permutation test`) έχει σχεδιαστεί ειδικά για να λειτουργεί με γονιδιωματικές περιοχές και περιλαμβάνει μια σειρά από προκαθορισμένες λειτουργίες τυχαιοποίησης και αξιολόγησης που καλύπτουν τις συχνότερες περιπτώσεις, αλλά ο χρήστης μπορεί επίσης να παρέχει προσαρμοσμένες λειτουργίες για να επεκτείνει τη λειτουργικότητά του.

Ο ακριβής αλγόριθμος για την εύρεση των IRs-HJ σε σύνδρομα μεκροελλείψεων και μικροδιπλασιασμών παρατίθεται στη συνέχεια:

```
library(BSgenome.Hsapiens.ucsc hg19)

genome <- BSgenome.Hsapiens.ucsc hg19

seqnames <- seqnames(genome)

hjtable <- read.table("C:\\hj.txt", header = TRUE)

for (i in 1:nrow(hjtable))
{
  for (seqname in seqnames)
  {
    x <- genome[[seqname]]

    #print(paste("Run HJ no. ", seqname, i, hjtable$HJ[i]))

    Lpattern <- toString(hjtable$HJ[i])

    M1 <- matchPattern(Lpattern, x, max.mismatch=0, min.mismatch=0,
```

```
with.indels=FALSE, fixed=TRUE, algorithm="auto")
#print(paste("M1 = ", M1))
myfile <- paste("C:\\hj", hjtable$HJ[i],
".bed", sep="")
sink(myfile, append=TRUE)
if (length(M1) != 0)
{
  chrvalue <- seqname
  for (k in 1:length(M1))
    cat(chrvalue)
    cat("\t")
    cat(start(M1[k]))
    cat("\t")
    cat(end(M1[k]))
    cat("\t")
    cat(toString(hjtable$HJ[i]))
    cat("\n")
  }
}
sink()
}
library("regioner")
test <- import(myfile)
test
Microdel.hg19_path <- system.file("tests", package = "rtracklayer")
Microdel.hg19_bed <- file.path(Microdel.hg19_path, "Microdel.hg19.bed")
```

```

Microdel.hg19 <- import(Microdel.hg19_bed)

Microdel.hg19

pt <- overlapPermTest(A=Microdel.hg19, B=test, ntimes=1000)

pt

lz <- localZScore(pt=pt, A=Microdel.hg19, B=test)

plot(lz)

plot(pt)

output <- overlapRegions(test, Microdel.hg19)

output <- paste("C:\\op", hjtable$HJ[i],
".bed", sep="")
}

```

Επιπλέον, τα βιοπληροφορικά εργαλεία, που είναι διαθέσιμα στον UCSC (<http://genome.ucsc.edu/>) (148), χρησιμοποιήθηκαν για τις γονιδιακές αναλύσεις στο ανθρώπινο γονιδιόμα (έκδοση hg19) για αλληλουχίες HJ που βρίσκονται εντός των ορίων των συνδρόμων και τις περιβάλλουσες αλληλουχίες τους σε κάθε ένα από τα 24 ανθρώπινα χρωμοσώματα, σε συνδυασμό με αναλύσεις εγγύτητας ρυθμιστικών παραγόντων του ENCODE (149). Η Repeatmasker Analysis (93) χρησιμοποιήθηκε για να προσδιορίσει την ύπαρξη επαναλαμβανόμενων στοιχείων στη γύρω περιοχή του HJ καθώς και τον τύπο και τη θέση ενός επαναλαμβανόμενου στοιχείου σε σχέση με τις υπο μελέτη ακολουθίες.

4.2.7 ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΕΓΓΥΤΗΤΑΣ ΣΕ ΓΟΝΙΔΙΑ ΚΑΙ TSS

Η γονιδιωματική ανάλυση στατιστικού εμπλουτισμού σε συγκεκριμένες γονιδιωματικές περιοχές πραγματοποιήθηκε χρησιμοποιώντας το εργαλείο ELDorado, στο διακομιστή Genomatix (www.genomatix.de) (150). Η γονιδιακή ειδική ανάλυση διεξήχθη με

το βιοπληροφορικό εργαλείο GREAT. Το GREAT χρησιμοποιήθηκε για να αναλύσει τη σχετική κατανομή μεταξύ των περιοχών IRs-HJs-MMS σε περιοχές έναρξης της μεταγραφής (TSS) των ανθρώπινων γονιδίων (153), χρησιμοποιώντας τις προεπιλεγμένες ρυθμίσεις για τη στατιστική ανάλυση στην έκδοση του ανθρώπινου γονιδιώματος hg19. Τα γονίδια που σχετίζονται με τις IRs-HJs-MMS λήφθηκαν από το διακομιστή GREAT.

4.3 ΕΥΡΕΣΗ ΟΛΩΝ ΤΩΝ ΠΑΡΑΛΛΑΓΩΝ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΜΟΤΙΒΩΝ ΤΟΥ PRDM9

Η πρωτεϊνική δομή του PRDM9 είναι εξαιρετικά διατηρημένη σε όλα τα είδη εκτός από την τερματική διάταξη δακτύλων ψευδαργύρου, η οποία είναι ιδιαίτερα πολυμορφική, τόσο στον αριθμό των δακτύλων που παρουσιάζονται όσο και στην ταυτότητα των τριών αμινοξέων σε κάθε δάκτυλο που έρχονται σε επαφή με το DNA και προσδιορίζουν την ειδικότητα πρόσδεσης DNA του PRDM9. Συνεπώς το γονίδιο παρουσιάζει πολλές παραλλαγές στην αλληλουχία του. Μετά από εκτενή αναζήτηση σε βάσεις δεδομένων όπως η NCBI, Science, Nature και άλλες μικρότερης έκτασης βρέθηκε ότι οι δομές του PRDM9 έχουν την μορφή $d[CCNCCNTNNCCNC]$. Για τις πιθανές παραλλαγές του γονιδίου, βρέθηκαν όλες οι δυνατές αλληλουχίες για $N=A$, $N=C$, $N=G$, $N=T$ αντίστοιχα. Αυτό κατέστη δυνατό με την δημιουργία script στην γλώσσα προγραμματισμού C, το οποίο «έτρεξε» σε [OnLine compiler](#). Από το Script βρέθηκαν 1024 παραλλαγές του νουκλεοτιδίου οι οποίες αποθηκεύτηκαν σε αρχείο .txt οι οποίες θα αξιοποιηθούν αργότερα στην R.

The screenshot shows a C program on the left and its output on the right. The program defines a character array 'num' with values 'A', 'G', 'T', 'C'. It uses four nested for loops to iterate through all combinations of indices i, j, k, and m from 0 to 3. A printf statement prints each combination as a 4-character string. A counter variable 'count' tracks the total number of permutations, and a final printf statement prints the total count.

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <string.h>
4
5 int main(int argc, char *argv[])
6 {
7
8     char num[10];
9     num[0]= 'A';
10    num[1]= 'G';
11    num[2]= 'T';
12    num[3]= 'C';
13    int i,j,k,l,m;
14    int count=0;
15
16    for (i = 0; i <= 3; i++) {
17        for (j = 0; j <= 3; j++) {
18            for (k = 0; k <= 3; k++) {
19                for (l = 0; l <= 3; l++) {
20                    for (m = 0; m <= 3; m++) {
21
22                        printf("CC%CC%t%c%CC%CC \n", num[i], num[j], num[k], num[l], num[m]
23                    );
24                        count++;
25                    }
26                }
27            }
28        }
29    }
30    printf("total combinations %d ",count);
31    return 0;
32 }
33

```

The output on the right shows 24 lines of 4-character strings, each representing a permutation of 'AGTC', followed by the line 'total combinations 1024'.

Εικόνα 2. Script στην γλώσσα προγραμματισμού C με το οποίο βρέθηκαν. Όλες οι πιθανές παραλλαγές του δεκατριονουκλεοτιδίου PRDM9 .

4.3.1 ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΑΚΟΛΟΥΘΙΩΝ ΜΕΣΩ ΤΗΣ ΚΟΝΣΟΛΑΣ R 3.5.2 ΚΑΙ ΑΝΑΦΟΡΑ ΤΩΝ ΑΠΑΡΑΙΤΗΤΩΝ ΠΑΚΕΤΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ.

Για την μελέτη της συσχέτισης των αλληλουχιών του PRDM9 με τα καρκινικά γονίδια στο ανθρώπινο γονιδίωμα χρησιμοποιήθηκε η γλώσσα προγραμματισμού R 3.5.2. με τα εξής πακέτα:

1. BiocManager_ 1.30.4

2. Biocgenerics_0.28.0
3. BSgenome_1.50.0
4. BSgenome.Hsapiens.UCSC.hg19_ 1.4.0
5. Rtracklayer_ 1.42.2
6. regioneR_ 1.14.0

Η διαδικασία αναζήτησης διεξήχθη μία φορά για όλο το ανθρώπινο DNA ,με την χρήση ειδικού αλγορίθμου. Με την χρήση αυτού, εντοπίστηκαν όλες οι παραλλαγές του PRDM9 ,που βρέθηκαν από το script, για κάθε χρωμόσωμα του ανθρώπινου γονιδιώματος. Οι εντολές που δόθηκαν κατά τη διαδικασία της αναζήτησης είναι οι εξής:

```

1) library(BSgenome.Hsapiens.UCSC.hg19)

genome <- BSgenome.Hsapiens.UCSC.hg19

2) seqnames <- seqnames(genome)

for (seqname in seqnames)
{
x <- genome[[seqname]]

3) prdm9table <- read.table("C:\\prdm9.txt", header = TRUE)

for (i in 1:1024)
{

#print(paste("Run PRDM9 no. ", seqname, i, prdm9table$PRDM9[i]))

Lpattern <- toString(prdm9table$PRDM9[i])

4) M1 <- matchPattern(Lpattern, x, max.mismatch=0, min.mismatch=0,with.indels=FALSE,
fixed=TRUE, algorithm="auto")

```

```
5) #print(paste("M1 = ", M1))

myfile <- paste("C:\\prdm9", prdm9table$PRDM9[i], ".bed", sep="")

sink(myfile, append=TRUE)

if (length(M1) != 0)

{

chrvalue <- seqname

for (k in 1:length(M1))

{

cat(chrvalue)

cat("\t")

cat(start(M1[k]))

cat("\t")

cat(end(M1[k]))

cat("\t")

cat(toString(prdm9table$PRDM9[i]))

cat("\n")

}

}

sink()

}
```

- 1^η Εντολή → Δόθηκε για την φόρτωση ολόκληρου του ανθρώπινου γονιδιώματος στο R Project
- 2^η Εντολή → Τίθεται ως μεταβλητή X όλο το ανθρώπινο γονιδίωμα ορίζοντας του να ξεκινάει να διαβάζει από το χρωμόσωμα 1
- 3^η Εντολή → Ζητείται να διαβαστεί το txtαρχείο με όλες τις 1024 αλληλουχίες του PRDM9 ξεκινώντας με την σειρά από την πρώτη αλληλουχία να διαβαστεί ως τμήμα DNA(DNA string)
- 4^η Εντολή → Ζητείται να συσχετιστούν(match) οι αλληλουχίες του PRDM9 με αυτές του κάθε χρωμοσώματος
- 5^η Εντολή → Εμφάνιση των αποτελεσμάτων ως .bed αρχεία σε συγκεκριμένο pathway

4.3.2 ΣΥΛΛΟΓΗ ΑΡΧΕΙΩΝ ΚΑΡΚΙΝΟΥ ΚΑΙ ΧΡΗΣΗ ΤΟΥΣ ΣΤΗΝ R

Για την υλοποίηση της συλλογής αρχείων καρκίνου χρησιμοποιήθηκε η διαδικτυακή βάση δεδομένων BreCAN από την οποία συλλέχθηκε το σύνολο των δεδομένων (151). Στη συγκεκριμένη βάση δεδομένων βρίσκονται αποθηκευμένα προφίλ ασθενών που πάσχουν από καρκίνο, το γονιδίωμα των οποίων έχει υποστεί σημειακές δίκλωνες θραύσεις μονής βάσης. Πιο αναλυτικά, τα προφίλ που κατατέθηκαν στον ιστότοπο δεν είναι τίποτα άλλο από τα χαρτογραφημένα γονιδιώματα σωματικών κυττάρων με δίκλωνες θραύσεις, καθενός από τους καρκινοπαθείς. Οι 99 ασθενείς που μελετήθηκαν συνολικά πάσχουν από 3 διαφορετικά είδη καρκίνου, δηλαδή από διηθητικό καρκίνο του μαστού (BRC, Breast Invasive Carcinoma, από αδενοκαρκίνωμα του πνεύμονα (LUAD, Lung Adenocarcinoma), και από καρκίνωμα στα πλακώδη κύτταρα της κεφαλής και του τραχήλου (HN, Head and Neck squamous cell carcinoma).

Από την βάση δεδομένων χρησιμοποιήθηκαν 192,621 δίκλωνες θραύσεις DNA και αποθηκεύτηκαν σε μορφή .bed τα εξής αρχεία:

- Γονιδιωματικό προφίλ για τον καρκίνο του στήθους(breast cancer breakpoint profiles)

- Γονιδιωματικό προφίλ για το καρκίνο της κεφαλής και τραχήλου (head and neck cancer breakpoint profiles)
- Γονιδιωματικό προφίλ για το αδενοκαρκίνωμα πνεύμονος (luad breakpoint profiles)

Έπειτα έγινε εκ νέου χρήση ειδικού αλγορίθμου στη κονσόλα R 3.5.2 για την ανάλυση και ταυτοποίηση των αρχείων καρκίνου με τις παραλλαγές του γονιδίου PRDM9

- BRC100 (breast cancer)
- BRC500
- HN100 (head and neck cancer)
- HN500
- LUAD100 (lung adenocarcinoma)
- LUAD500

Οι αριθμοί δίπλα από τα αρχεία δηλώνουν την επέκταση του εύρους της γονιδιωματικής περιοχής κατά 100 νουκλεοτίδια και 500 νουκλεοτίδια αντίστοιχα. Οι εντολές που δόθηκαν κατά τη διαδικασία της αναζήτησης είναι οι εξής:

```
1) library(BSgenome.Hsapiens.UCSC.hg19)

genome <- BSgenome.Hsapiens.UCSC.hg19

seqnames <- seqnames(genome)

2) prdm9table <- read.table("C:\\Work\\GeneScreening\\prdm9.txt", header =TRUE)

for (i in 1:1024){for (seqname in seqnames)

x <- genome[[seqname]]

#print(paste("Run PRDM9 no. ", seqname, i, prdm9table$PRDM9[i]))

Lpattern <- toString(prdm9table$PRDM9[i])

M1 <- matchPattern(Lpattern, x, max.mismatch=0, min.mismatch=0,
```

```

with.indels=FALSE, fixed=TRUE, algorithm="auto")

#print(paste("M1 = ", M1))

3) myfile <- paste("C:\\Work\\GeneScreening\\prdm9\\",
prdm9table$PRDM9[i], ".bed", sep="")

sink(myfile, append=TRUE)

if (length(M1) != 0)

{chrvalue <- seqname

for (k in 1:length(M1))

{cat(chrvalue)

cat("\t")

cat(start(M1[k]))

cat("\t")

cat(end(M1[k]))

cat("\t")

cat(toString(prdm9table$PRDM9[i]))

cat("\n")

}

}

sink(

myfile2 <- paste("C:\\Work\\GeneScreening\\prdm9\\lz",
prdm9table$PRDM9[i], ".jpg", sep="")

myfile3 <- paste("C:\\Work\\GeneScreening\\prdm9\\pt",
prdm9table$PRDM9[i], ".jpg", sep="")

myfile4 <- paste("C:\\Work\\GeneScreening\\prdm9\\output",

```

```

prdm9table$PRDM9[i], ".txt", sep="")

library("regioner")

test <- import(myfile)

test

4) cancer_path <- system.file("tests", package = "rtracklayer")

***cancer_bed <- file.path(**cancer_path, "**cancer.bed")

***cancer <- import(**cancer_bed)

***cancer

5) pt <- overlapPermTest(A=***cancer, B=PRDM9, ntimes=1000)

pt

6) lz <- localZScore(pt=pt, A=***cancer, B=PRDM9)

jpeg(myfile2)

plot(lz)

dev.off()

jpeg(myfile3)

plot(pt)

dev.off()

7) output <- overlapRegions(test, ****cancer)

output

write.table(output, myfile4, append = FALSE, quote = TRUE, sep = " ",

eol = "\n", na = "NA", dec = ".", row.names = TRUE,

col.names = TRUE, qmethod = c("escape", "double"),

fileEncoding = "")

```

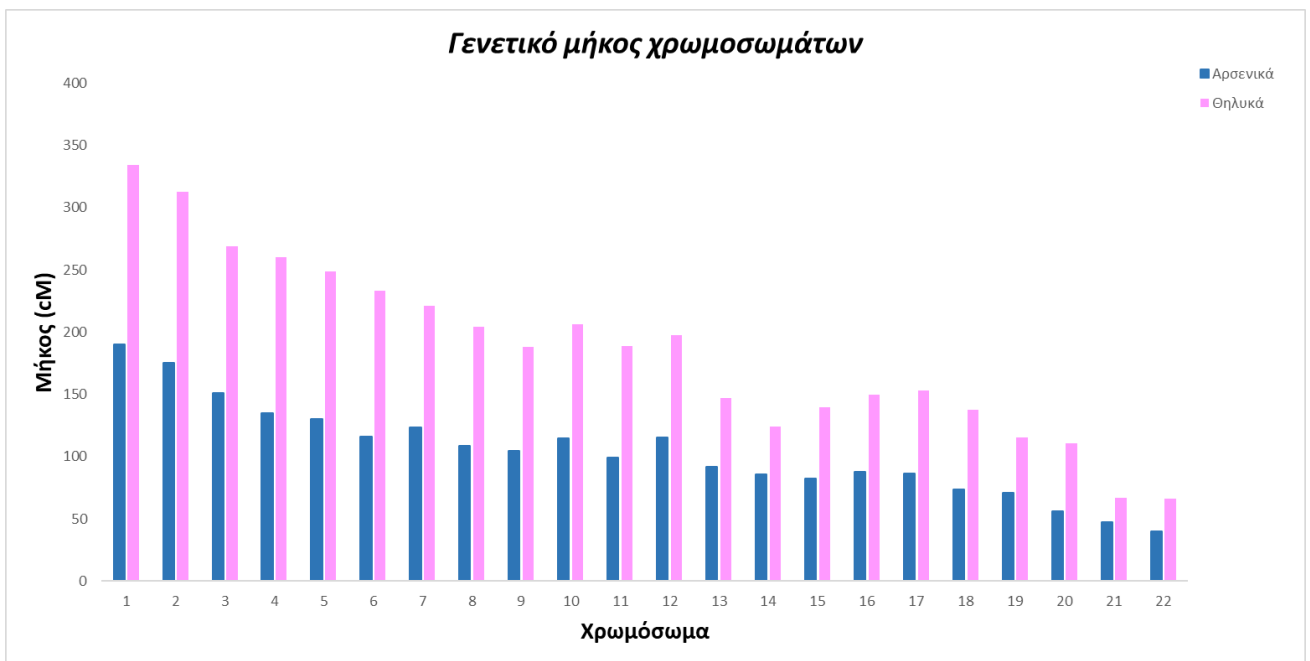
- 1^η Εντολή → Φόρτωση όλου του ανθρώπινου γονιδιώματος στην R
- 2^η Εντολή → Φόρτωση των αλληλουχιών του γονιδίου PRDM9 που εντοπίζονται στο ανθρώπινο γονιδίωμα
- 3^η Εντολή → Φόρτωση αυτών των αλληλουχιών σε συγκεκριμένο pathway
- 4^η Εντολή → Φόρτωση των αρχείων καρκίνου. Όπου ***cancer ένα διαφορετικό αρχείο καρκίνου για κάθε ανάλυση
- 5^η Εντολή → Σύγκριση του PRDM9 και του αρχείου καρκίνου και εμφάνιση των αποτελεσμάτων σε διαγράμματα εύρους z-score σε jpg μορφή
- 6^η Εντολή → Σύγκριση του PRDM9 και του αρχείου καρκίνου και χρήση overlap permutation test των αποτελεσμάτων σε .jpg μορφή
- 7^η Εντολή → Σύγκριση του PRDM9 και του αρχείου καρκίνου και εμφάνιση των αποτελεσμάτων σε .outrput μορφή

ΚΕΦΑΛΑΙΟ 5

ΑΠΟΤΕΛΕΣΜΑΤΑ

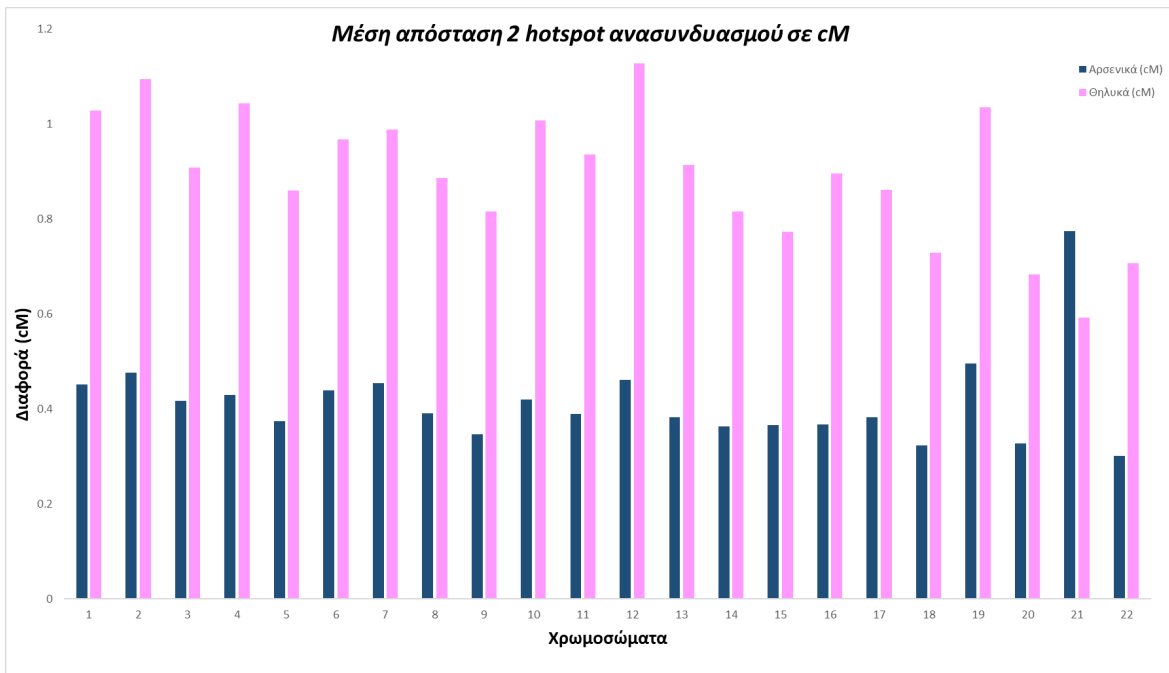
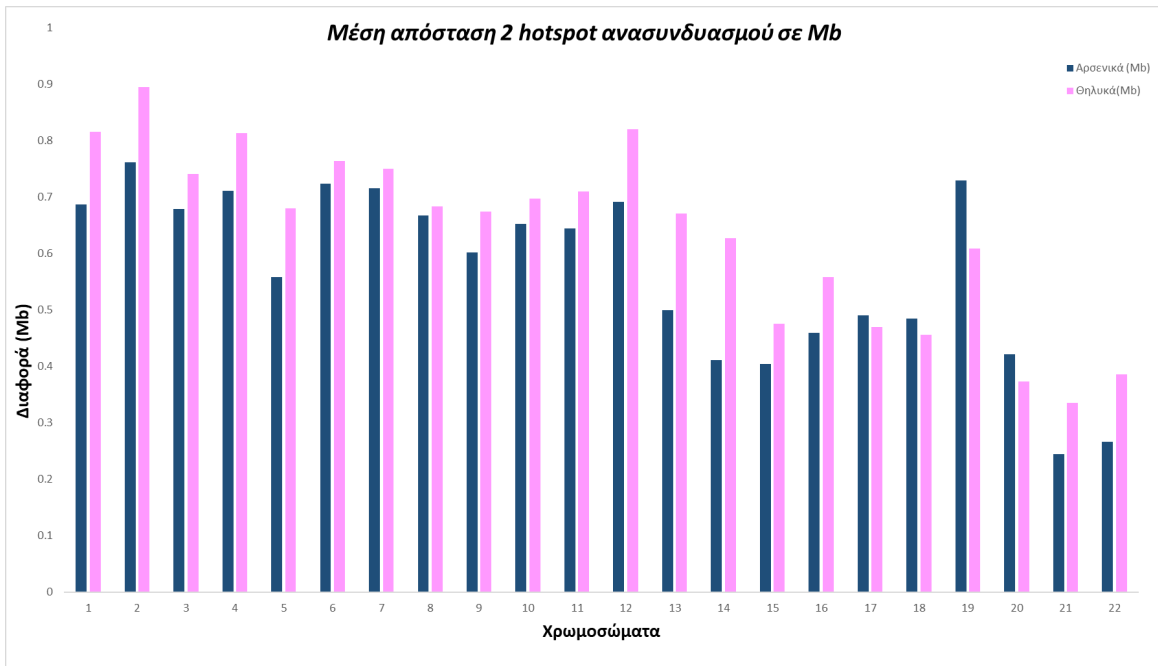
5.1 ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΤΗ ΜΕ ΔΕΙΚΤΕΣ ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΣΤΑ 2 ΦΥΛΑ

Το συνολικό μήκος που καλύπτουν οι γενετικοί δείκτες που χρησιμοποιήθηκαν, όπως αυτό υπολογίστηκε, παρατίθεται παρακάτω:



Εικόνα 5.1: Σύγκριση των γενετικών μηκών των χρωμοσωμάτων μεταξύ των δύο φύλων.

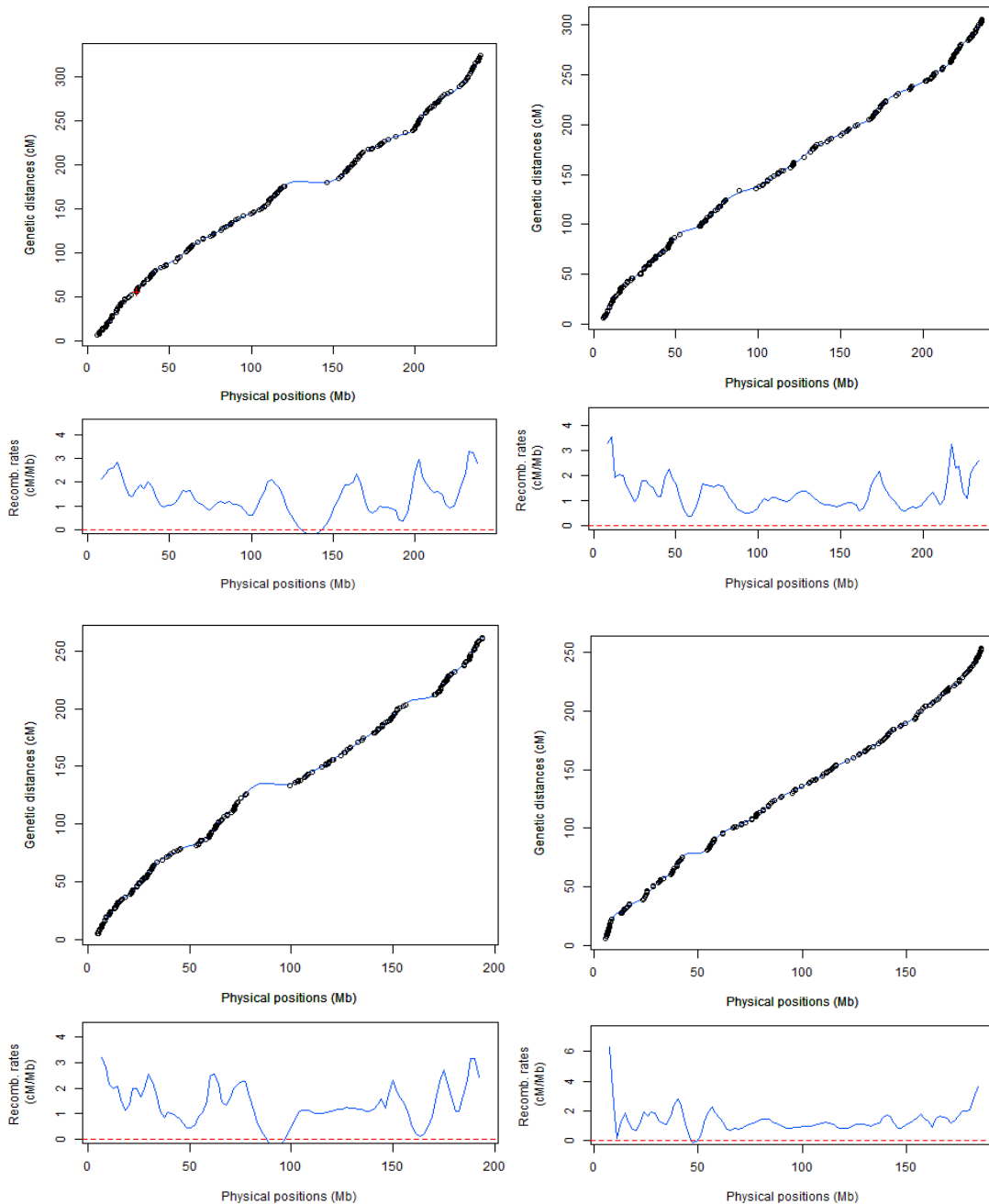
Τα αποτελέσματα σε γράφημα στηλών:



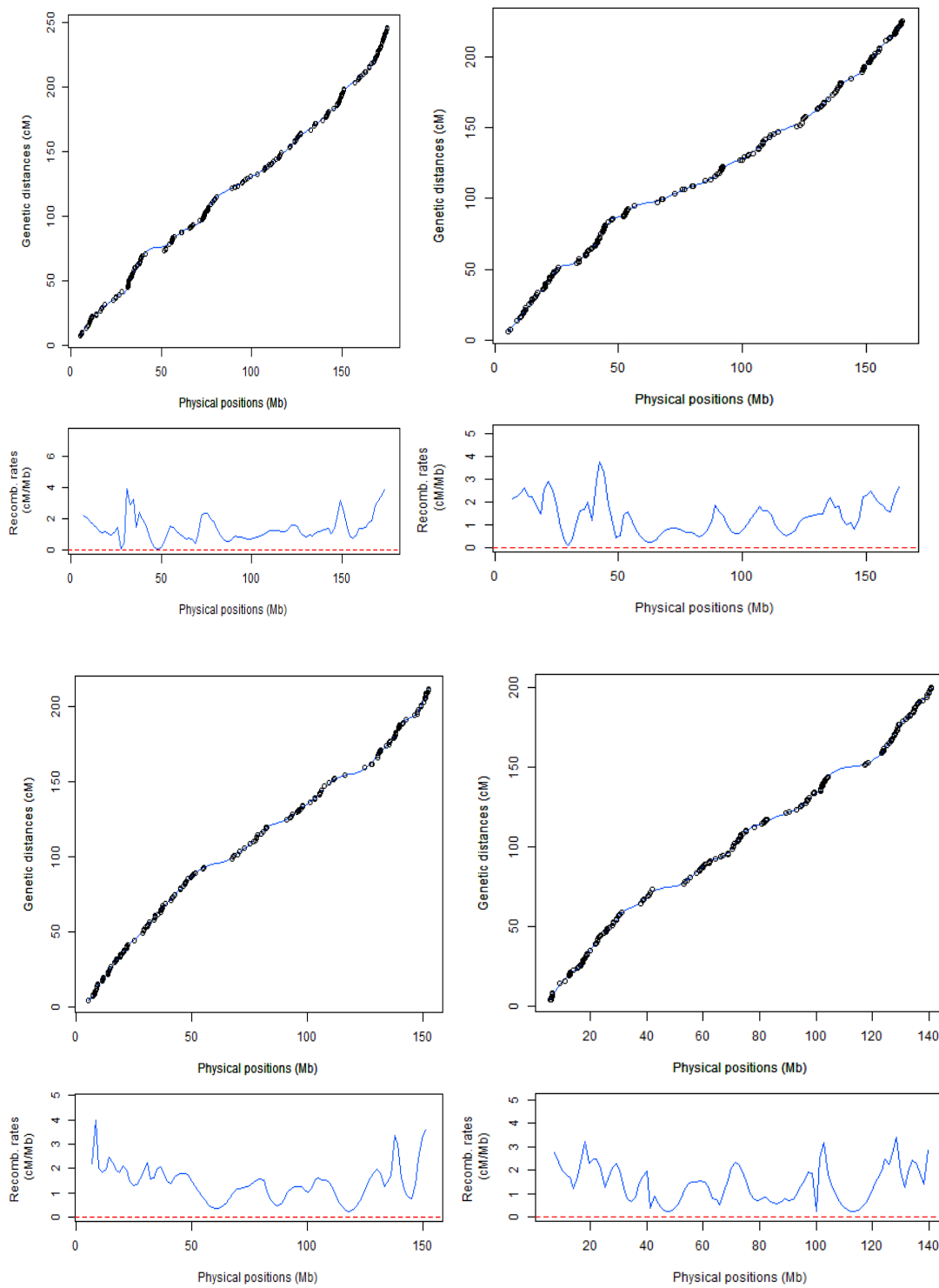
Εικόνα 5.2: Η μέση απόσταση δύο διαδοχικών hotspot σε κάθε χρωμόσωμα για κάθε φύλο σε Mb και cM. Με μπλε παρουσιάζονται τα αρσενικά άτομα ενώ με ροζ τα θηλυκά.

5.2 ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΞΥ ΦΥΣΙΚΗΣ ΚΑΙ ΓΕΝΕΤΙΚΗΣ ΑΠΟΣΤΑΣΗΣ ΤΩΝ ΓΕΝΕΤΙΚΩΝ ΔΕΙΚΤΩΝ ΣΤΑ 2 ΦΥΛΛΑ

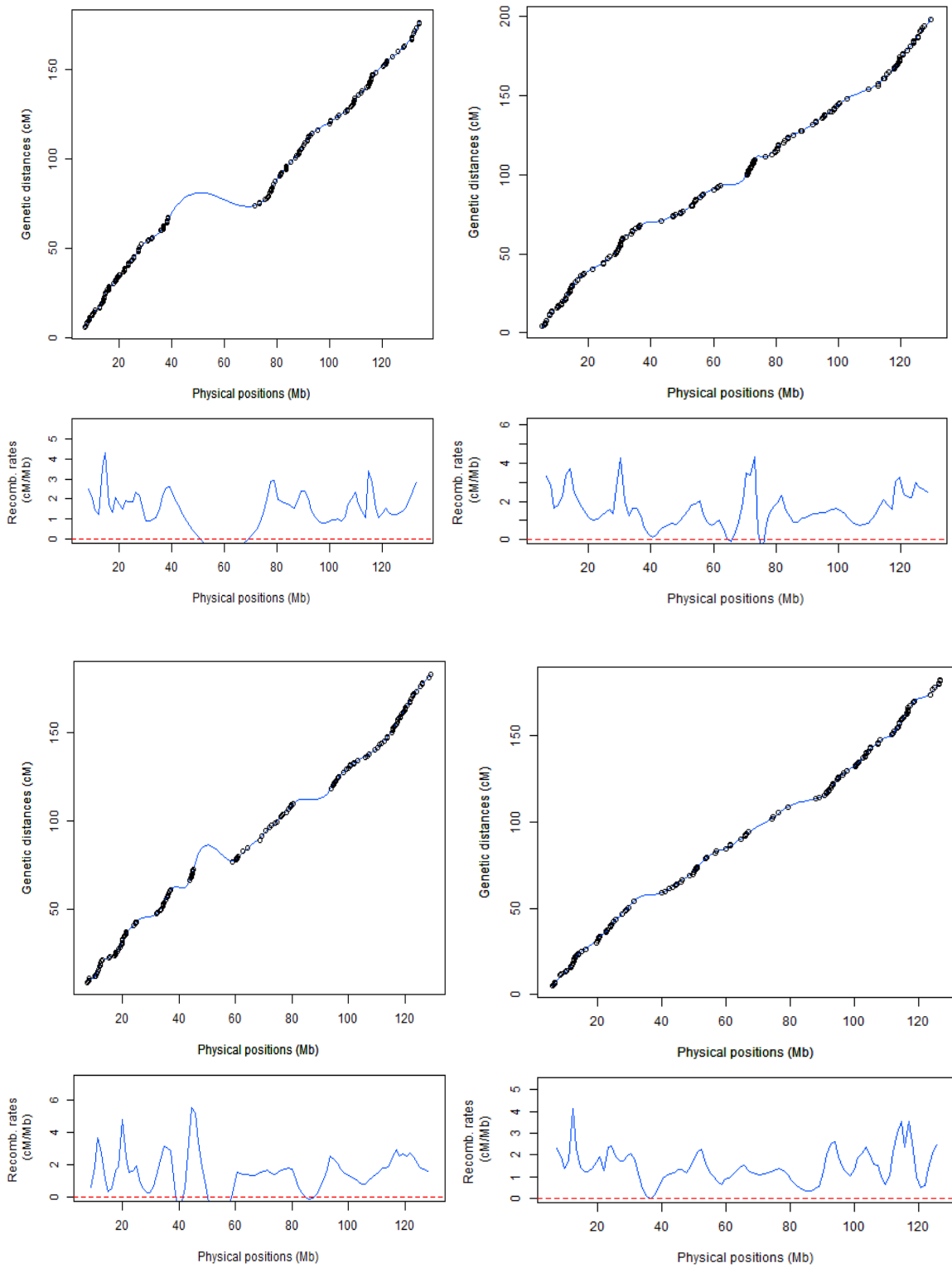
Η συσχέτιση μεταξύ φυσικής και γενετικής απόστασης των γενετικών δεικτών, όπως αναλύθηκε στην ενότητα 4:



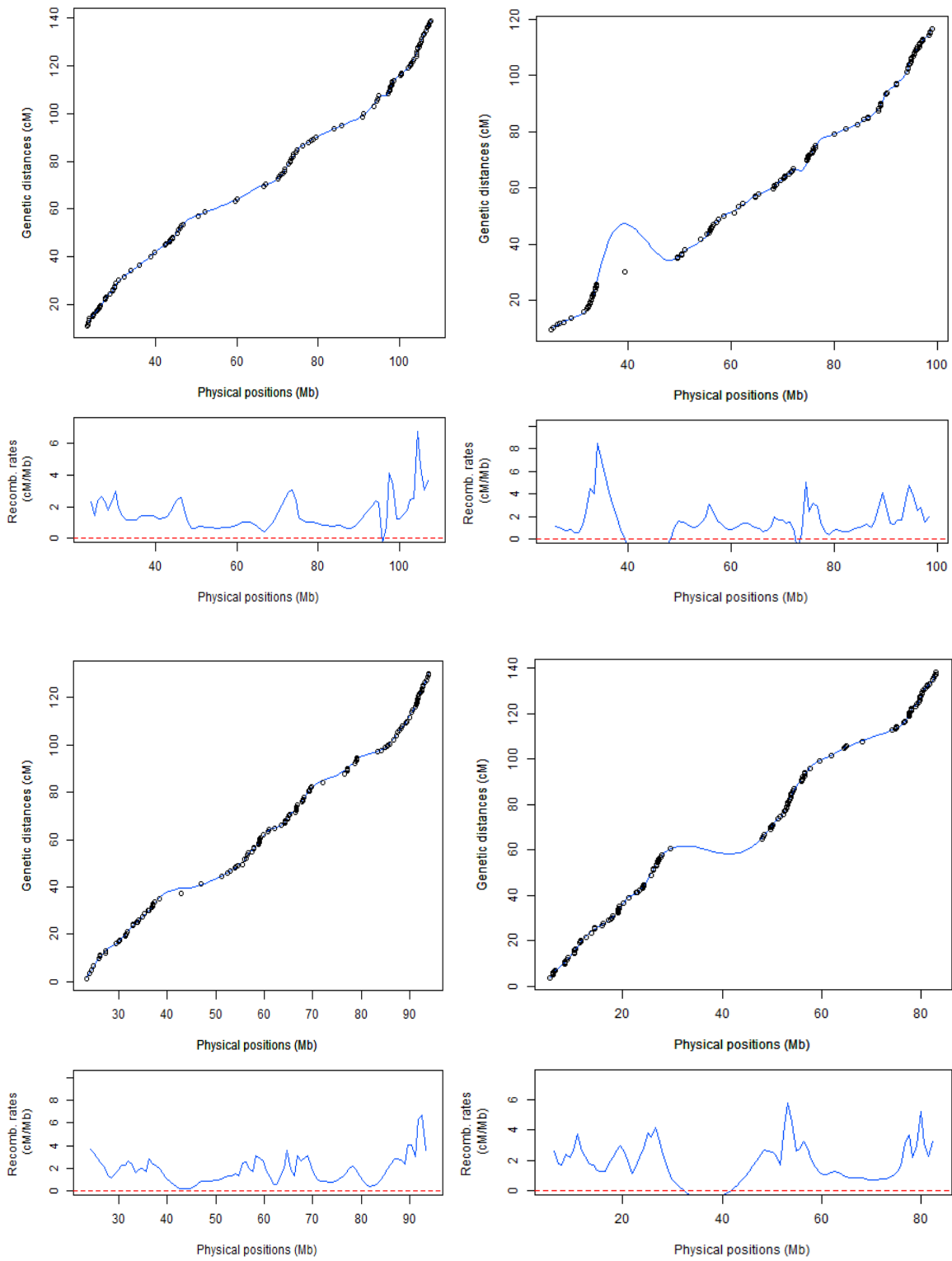
Εικόνα 5.3: Θηλυκά χρωμοσώματα 1 (επάνω αριστερά) έως 4 (κάτω δεξιά)



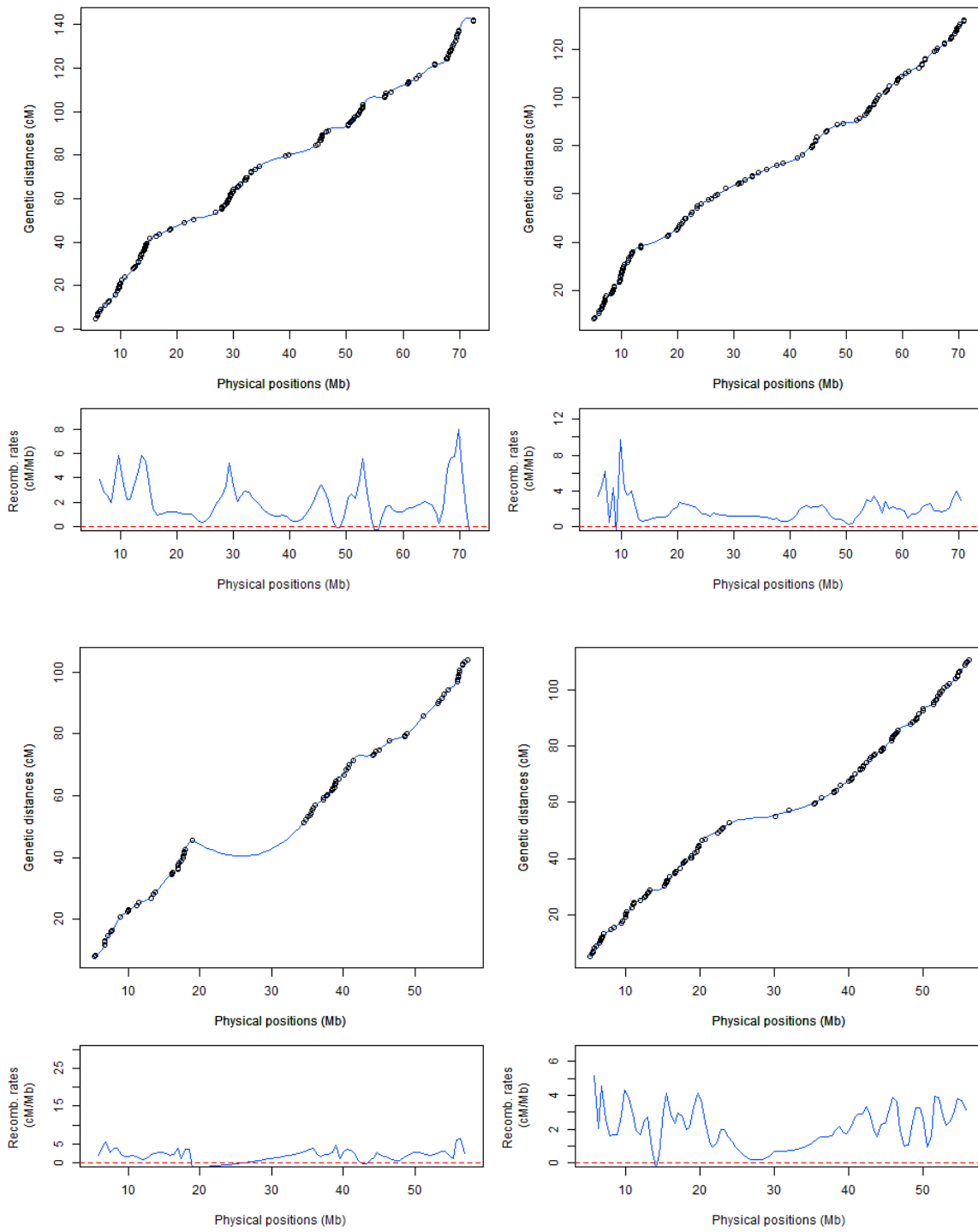
Εικόνα 5.4: Θηλυκά χρωμοσώματα 5 (επάνω αριστερά) έως 8 (κάτω δεξιά).



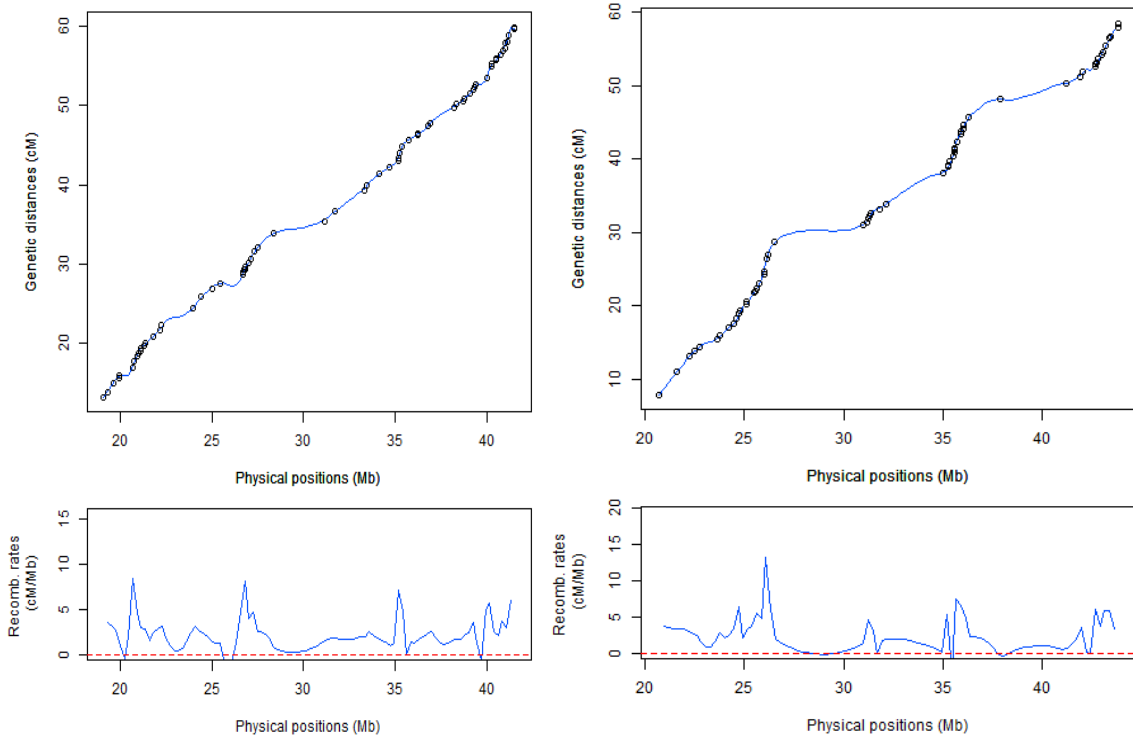
Εικόνα 5.5: Θηλυκά χρωμοσώματα 9 (επάνω αριστερά) έως 12 (κάτω δεξιά).



Εικόνα 5.6 Θηλυκά χρωμοσώματα 13 (επάνω αριστερά) έως 16 (κάτω δεξιά).

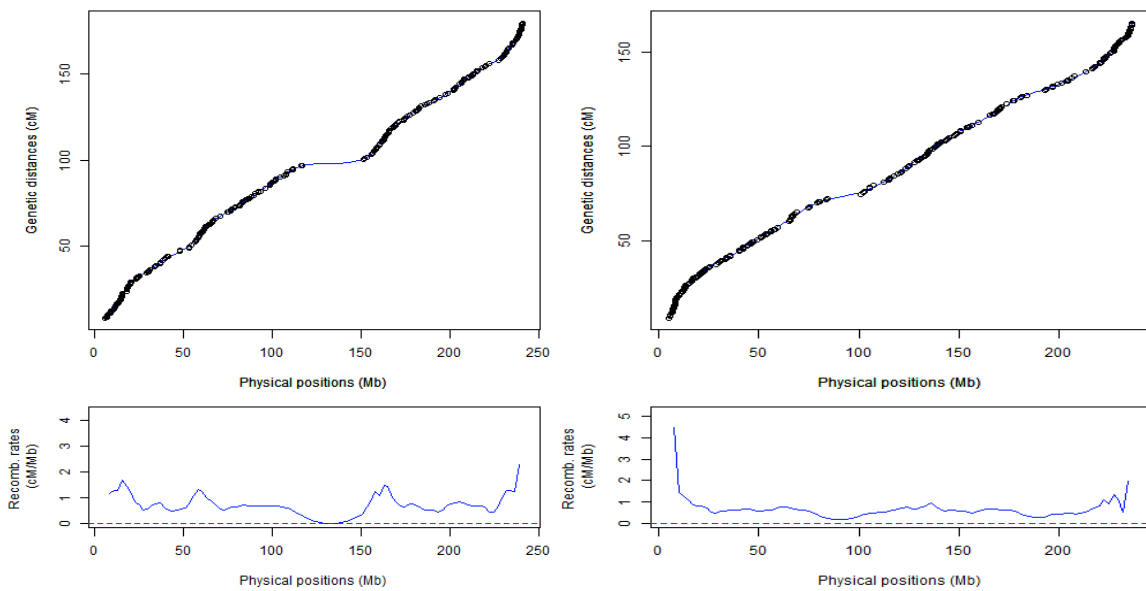


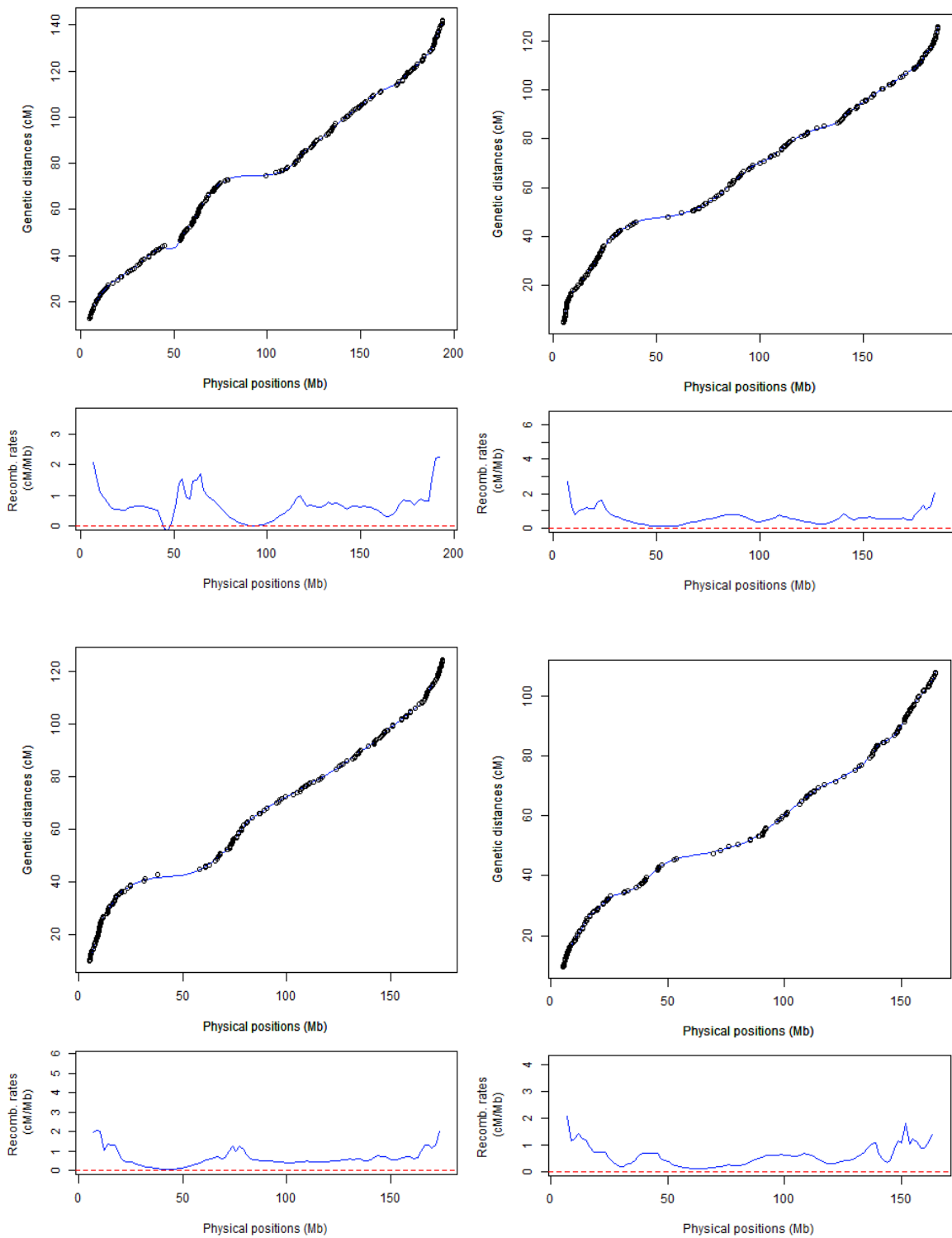
Εικόνα 5.7 Θηλυκά χρωμοσώματα 17 (επάνω αριστερά) έως 20 (κάτω δεξιά).



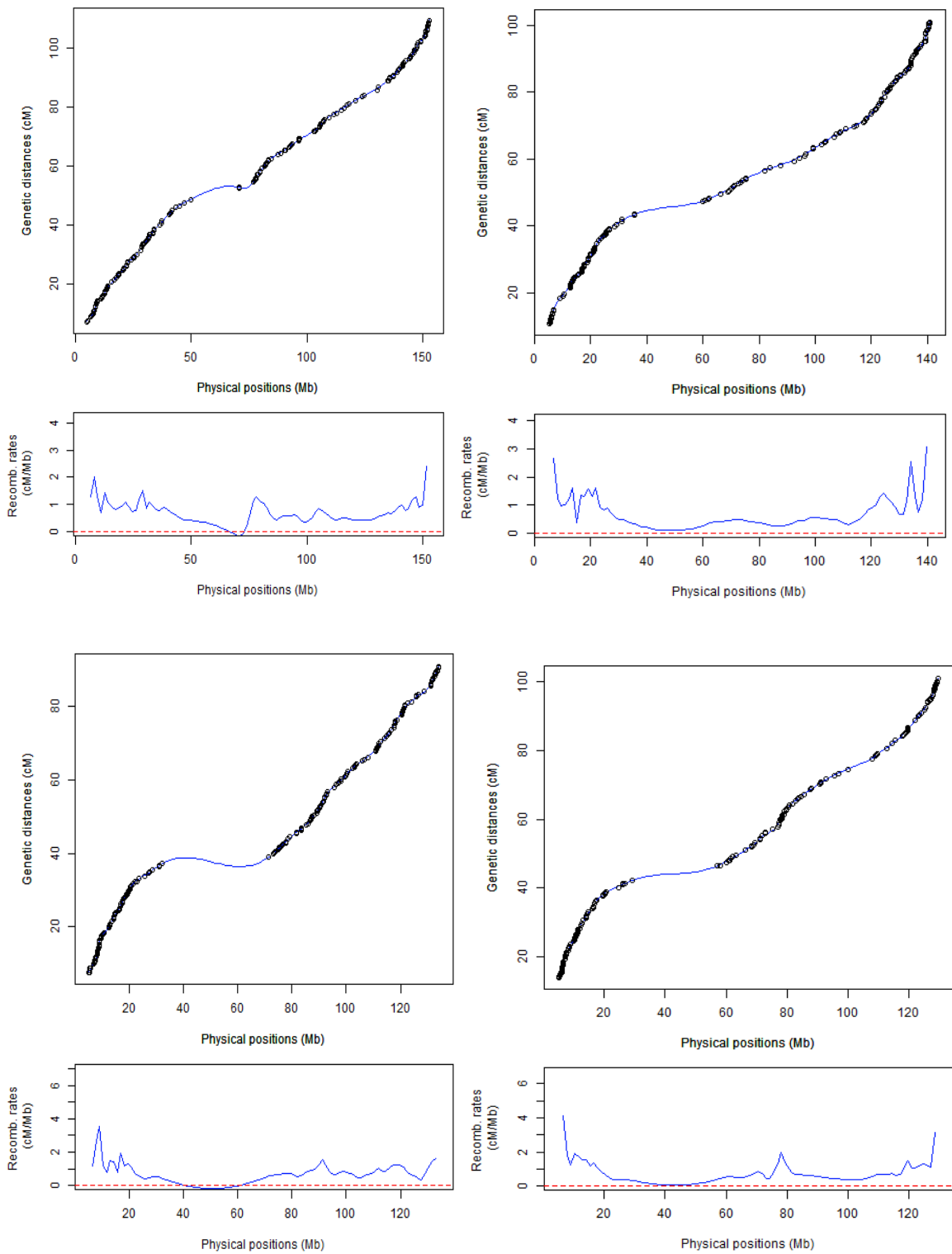
Εικόνα 5.8 Θηλυκά χρωμοσώματα 21 (αριστερά) και 22 (δεξιά).

Οι χάρτες Marey για τα αρσενικά άτομα:

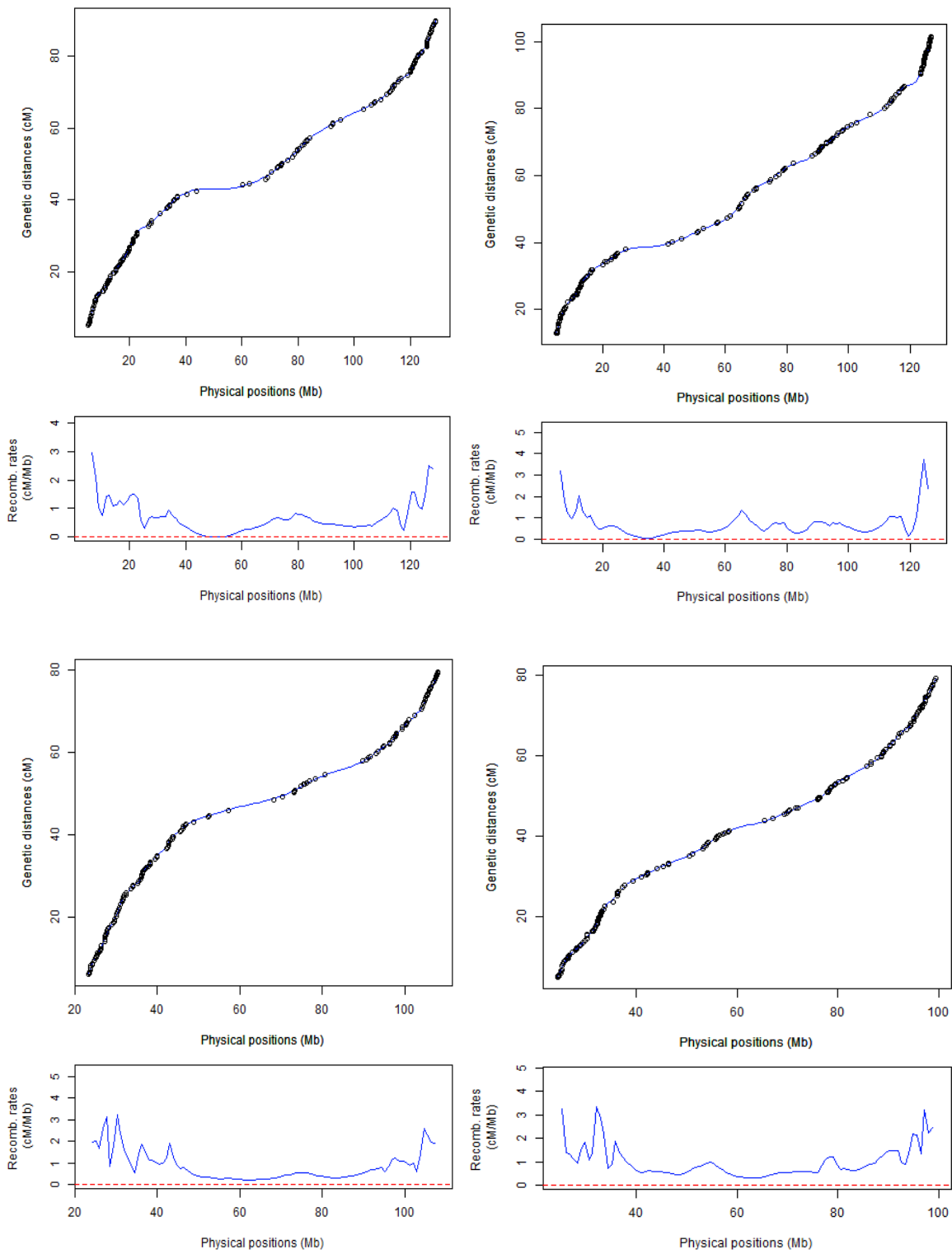




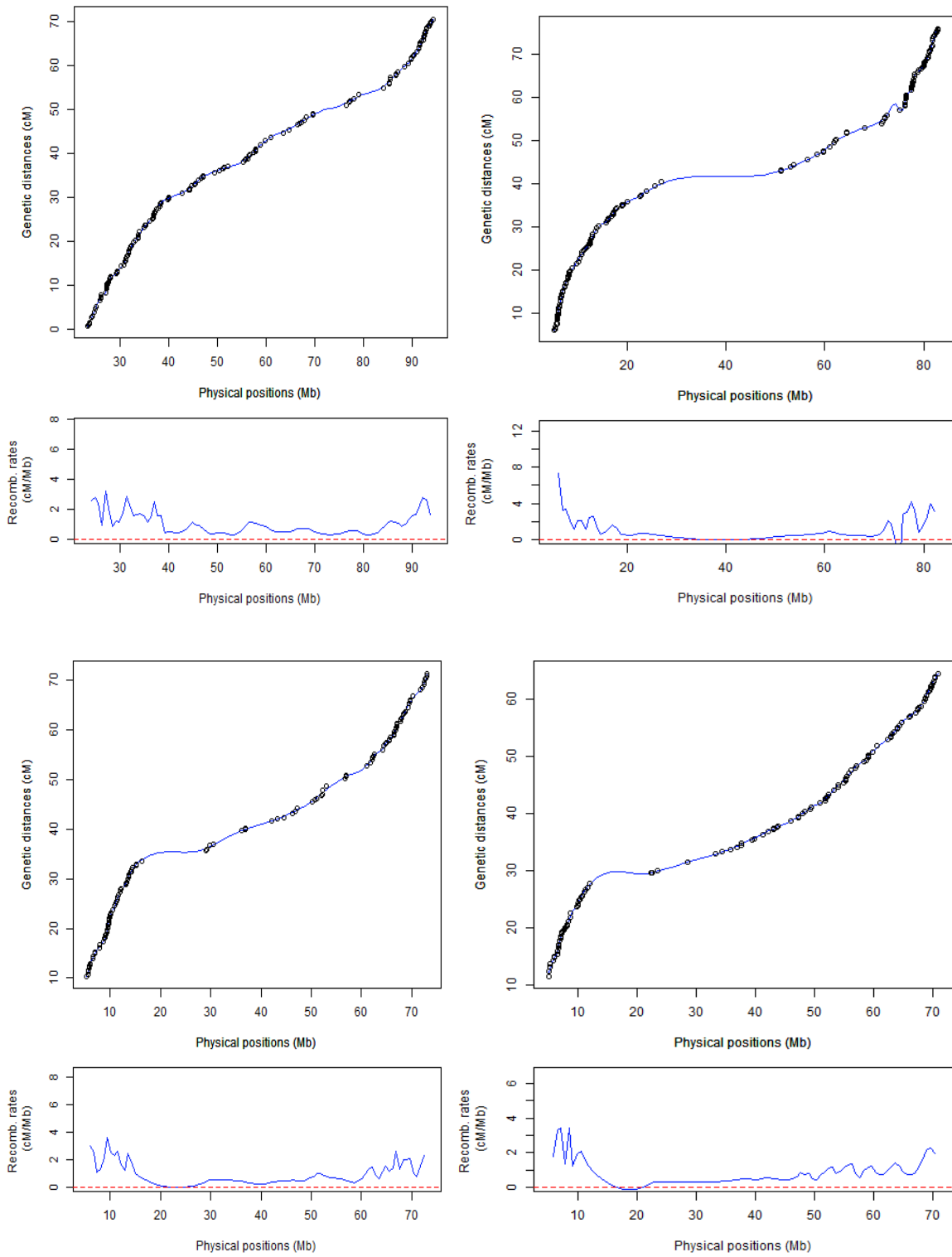
Εικόνα 5.9: Αρσενικά χρωμοσώματα 1 (επάνω αριστερά) έως 6 (κάτω δεξιά).



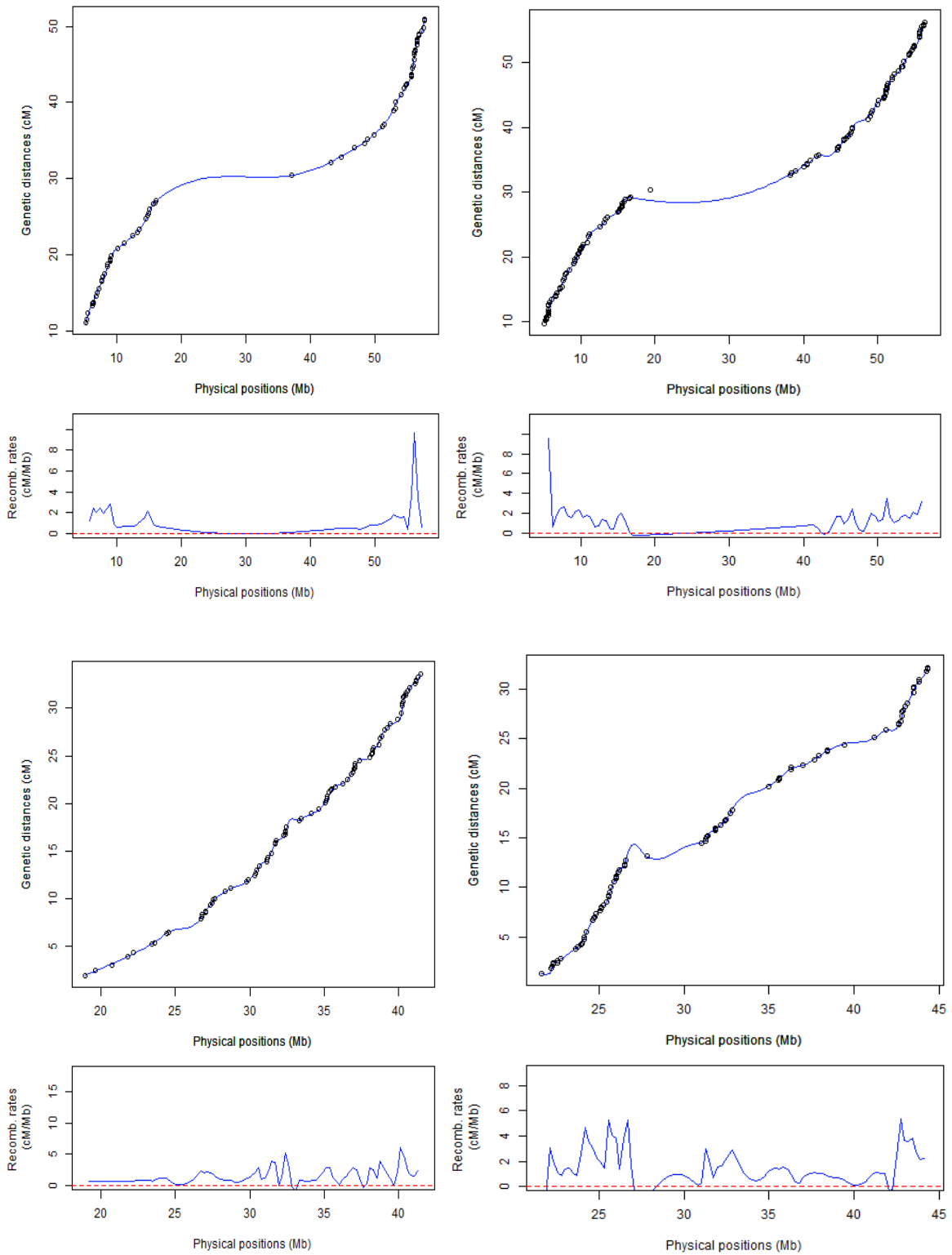
Εικόνα 5.10 Αρσενικά χρωμοσώματα 7 (επάνω αριστερά) έως 10 (κάτω δεξιά).



Εικόνα 5.11 Αρσενικά χρωμοσώματα 11 (επάνω αριστερά) έως 14 (κάτω δεξιά).



Εικόνα 5.12 Αρσενικά χρωμοσώματα 15 (επάνω αριστερά) έως 18 (κάτω δεξιά).



Εικόνα 5.13 Αρσενικά χρωμοσώματα 19 (επάνω αριστερά) έως 22 (κάτω δεξιά).

5.3 ΠΛΗΡΗΣ ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ IRs-HJs ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ

5.3.1 ΣΥΣΧΕΤΙΣΗ IRs-HJs ΜΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ

Οι δομές Holliday που μελετήθηκαν συνίστανται από ανεστραμμένες επαναλήψεις (IRs) μιας διατηρημένης μορφής ενός δεκαμερούς αλληλουχιακού μοτίβου που έχει τη μορφή (CCnnN6N7C8GG). Προηγούμενες πειραματικές μελέτες στο ελεύθερο DNA, το προαναφερθεν δεκαμερές (CCnnN6N7C8GG), έδειξε την ύπαρξη 4 τρινουκλεοτιδίων στις θέσεις του N6N7N8 του πυρήνα, δηλαδή ACC, GCC, ATC, CCC και οι αντίστοιχες εκτεταμένες αλληλουχίες τους, μπορεί επίσης υπό ευνοϊκές συνθήκες να σχηματίσουν HJs. Αντίστοιχα, το συντηρημένο μοτίβο IRs-HJs έχει τη μορφή (CCDRBVYCGG) σύμφωνα με τον κώδικα International Union of Pure and Applied Chemistry. Με βάση τα παραπάνω δεδομένα, σχεδιάστηκε πίνακας που περιέχει όλα τα πιθανά μοτίβα HJ, τις κατανομές καθώς και τα ποσοστά αλληλουχιών IRs-HJ που βρίσκονται σε μεταθετά στοιχεία

IRs-HJs	Total hits	TE class (% of total hits)					
		ALUs	LINE 1	ERV MaLRs	ERV class 1	DNA TEs	Other repeats
d(CCGGGCCCGG)(1)	818	81.1	1.93	—	3.34	0.28	13.4
d(CCGGCGCCCGG)(2)	360	1.7	—	—	84.5	—	13.8
d(CCGGGACCGG)(3)	72	32.19	—	5.01	19.71	7.62	32.56
d(CCGGTACCGG)(4)	39	—	42.94	3.22	11.88	41.94	
d(CCGATATCGG)(5)	17	72.72	18.18	—	—	—	9
d(TCGGTACCGG)(6)	13	—	36.61	25	—	17.4	48.5

Πίνακας 5.1. Συνολικός αριθμός παρατηρήσεων στο γονιδίωμα και ποσοστό συσχέτισης κάθε ακολουθίας IRs-HJ με συγκεκριμένες οικογένειες TEs

Παρατηρήθηκε άμεση σχέση μεταξύ των IRs-HJs και μεταθετών στοιχείων (TEs) του ανθρώπινου γονιδιώματος και διαπιστώθηκε πως ορισμένα μοτίβα του δεκανουκλεοτιδίου

των αντίστοιχων ανεστραμμένων επαναλήψεων που σχηματίζουν Holliday Junctions εδράζονται μέσα σε ανενεργά, αλλά και ενεργά ρετρομεταθετά στοιχεία. Αυτό το αποτέλεσμα οδηγεί στο συμπέρασμα ότι τα ρετροτρανσποζόνια λειτουργούν ως φορείς που διασπείρουν τις δομές ανασυνδυασμού μέσα στο γονιδίωμα ενισχύοντας την πλαστικότητα του. Παρόλα αυτά, η αλλαγή της ρύθμισης τους από τυχόν μεθυλιώσεις μπορεί να επάγει την άναρχη μετακίνηση τους μέσα στο γονιδίωμα. Τα αποτελέσματα της παραπάνω ενέργειας μπορεί να αποβούν μοιραία για τη σταθερότητα του γονιδιώματος και κατ' επέκταση για την επιβίωση του κυττάρου, καθώς έχουν καταγραφεί περιστατικά εμφάνισης καρκινικών όγκων, τα οποία σχετίζονται με την άναρχη μετακίνηση των ρετροτρανσποζονίων. Ενδεικτικά παραδείγματα όπου μπορεί να πληγεί η ακεραιότητα του γονιδιώματος λόγω της απορρύθμισης της μετακίνησης των ρετροτρανσποζονίων αποτελούν οι ενθέσεις των προαναφερθέντων ρετροτρανσποζονίων σε ογκογονίδια ή σε ογκοκατασταλτικά γονίδια, καθώς και η πρόκληση δίκλωνων θραύσεων κατά τη μετακίνηση τους όπως αναφέρθηκε προηγουμένως.

Μέχρι σήμερα, είναι άγνωστο αν οι ανεστραμμένες αλληλουχίες (IRs) που σχηματίζουν HJs συνδέονται με μεταθετά στοιχεία (TEs) του DNA. Η παρουσία των IRs-HJs εντός των TEs μπορεί να δώσει έναν πιθανό μηχανισμό διασποράς των υπό μελέτη αλληλουχιών, μέσω συμβάντων ρετρομετάθεσης. Για το λόγο αυτό, μέσω της γλώσσας προγραμματισμού R εντοπίστηκαν οι συγκεκριμένες θέσεις στο γονιδίωμα κάθε ακολουθίας IRs-HJs. Πιο συγκεκριμένα, αναπτύχθηκε αλγόριθμός για τον προσδιορισμό των γονιδιωματικών τόπων σε κάθε χρωμόσωμα όπως επίσης και για τον ακριβή εντοπισμό των IRs-HJs που βρίσκονται σε επαναλαμβανόμενες και μη επαναλαμβανόμενες ακολουθίες του DNA, όπως ορίζεται από το λογισμικό RepeatMasker. Για να εντοπιστούν όλες οι ακολουθίες IRs-HJs που σχετίζονται ή δεν σχετίζονται με μεταθετά στοιχεία (TEs), ο αλγόριθμος εφαρμόστηκε δύο φορές για κάθε ανθρώπινο χρωμόσωμα. Η πρώτη εφαρμογή πραγματοποιήθηκε για την προβολή όλων των IRs-HJs, με τη «μάσκα» RepeatMasker απενεργοποιημένη, ενώ η δεύτερη εφαρμογή πραγματοποιήθηκε για την προβολή των IRs-HJs με ενεργοποιημένη τη RepeatMasker «μάσκα». Οι πλεονάζουσες συντεταγμένες που δεν εντοπίστηκαν και στις δύο αναζητήσεις οδήγησαν στον ακριβή εντοπισμό του αριθμού των συντεταγμένων των HJs που σχετίζονται με μεταθετά στοιχεία. Ως αποτέλεσμα,

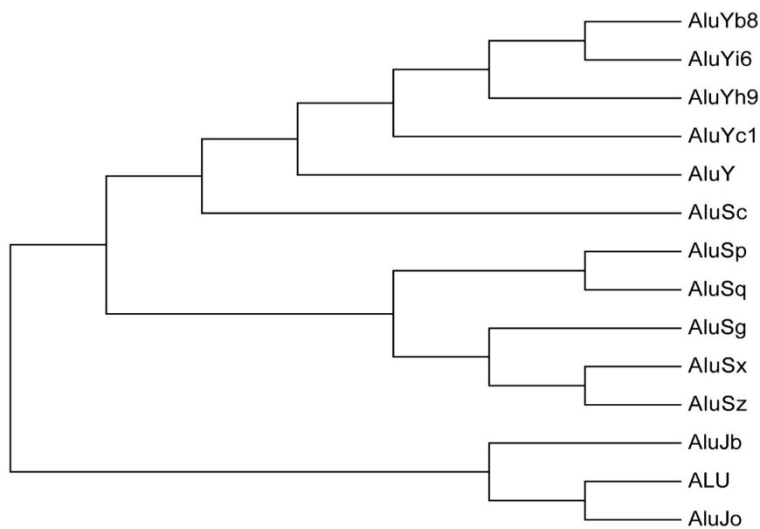
διαπιστώθηκε ότι ο συνολικός αριθμός των έξι προαναφερθέντων αλληλουχιών IRs-HJs στο ανθρώπινο γονιδίωμα είναι 2982 και εκείνες που συνδέονται με TEs είναι 1319. Στη συνέχεια βρέθηκε ότι ο αριθμός IRs-HJs που σχετίζονται με TEs είναι στατιστικά σημαντικός σε σύγκριση με τον συνολικό αριθμό των IRs-HJs στο ανθρώπινο γονιδίωμα. Για το σκοπό αυτό, υπολογίστηκε το συνολικό ποσοστό των μεταθετών στοιχείων TEs σε γειτνιάζουσες περιοχές εύρους μέχρι 1000 bp και στην 3' και στην 5' κατεύθυνση όλων των 2982 HJs. Το 30,9% αυτών των περιοχών είναι άμεσα συνδεδεμένο με TEs, έτσι ο αριθμός των 1319 HJ που σχετίζονται με TE είναι στατιστικά σημαντικός για τη συγκεκριμένη περιοχή ($p < 0.0001$).

Ο περαιτέρω χαρακτηρισμός των οικογενειών των μεταθετών στοιχείων που «φιλοξενούν» τις 1319 αλληλουχίες IRs-HJs έγινε με το λογισμικό RepeatMasker, αφού επεξεργάστηκαν οι συγκεκριμένες ακολουθίες μετά από επέκταση εύρους 50-bp 3' και 5' για καλύτερη ανάλυση. Το ποσοστό κατανομής της κάθε ακολουθίας IRs-HJs στα μεταθετά στοιχεία παρατίθεται στον Πίνακα 5.1. Βρέθηκε ότι τρεις συγκεκριμένες αλληλουχίες IRs-HJs (ακολουθίες # 1, # 2 και # 5) συνδέονται με έναν συγκεκριμένο τύπο μεταθετού στοιχείου. Αυτά τα HJs έχουν μεγάλη εξειδίκευση για τα ALU (# 1 και # 5) και ERV class 1 (# 2) και σχετίζονται πολύ λιγότερο με άλλα μεταθετά στοιχεία. Τα ρετρομεταθετά στοιχεία της οικογένειας ALU σχετίζονται με τρεις διαφορετικές αλληλουχίες IRs-HJs (# 1, # 3 και # 5) και τα LINE σχετίζονται με τρία HJs (# 4, # 5 και # 6). Τέλος, τα ERVs class 1 σχετίζονται με την HJ # 2, που επίσης την βρίσκομαι σε στοιχεία της οικογένειας ALU.

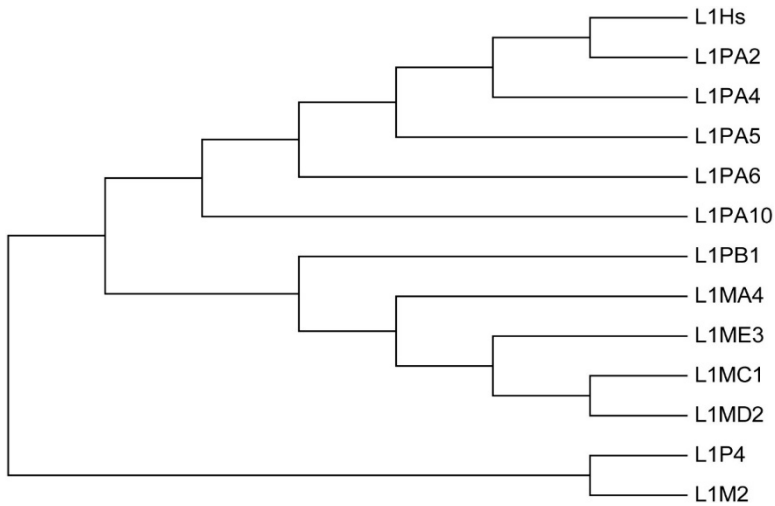
Επιπλέον, καθώς δεν υπάρχουν δεδομένα που να συσχετίζουν το περιεχόμενο % GC του γονιδιώματος με τις αλληλουχίες που είναι πιο ευαίσθητες σε ανασυνδυασμό, έγινε αλληλουχιακή ανάλυση των ευρύτερων περιοχών εκατέρωθεν των ακολουθιών IRs-HJs για την περιεκτικότητά τους σε GC (Πίνακας 1). Σύμφωνα με τον Πίνακα 5.1, κάθε υπο εξέταση αλληλουχία IRs-HJs δείχνει προτίμηση για μια διαφορετική οικογένεια μεταθετών στοιχείων. Σε ορισμένες περιπτώσεις, η εξειδίκευση φθάνει σχεδόν στο 100%. Για παράδειγμα, η αλληλουχία # 2 με πυρήνα το τριονουκλεοτίδιο GCC έχει σχεδόν απόλυτη εξειδίκευση με τον ενδογενή ρετροϊό HERV-H και σπάνια εντοπίζεται σε άλλες οικογένειες μεταθετών στοιχείων.

5.3.2 ΦΥΛΟΓΕΝΕΤΙΚΟ ΚΑΙ ΕΞΕΛΙΚΤΙΚΟ ΥΠΟΒΑΘΡΟ IRs-HJs ΠΟΥ ΕΝΤΟΠΙΖΟΝΤΑΙ ΣΕ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΣΤΟ ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ

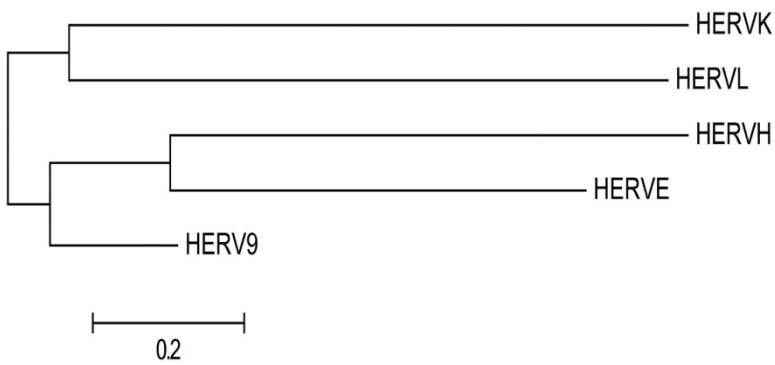
Για να αναλυθεί περαιτέρω το φυλογενετικό και εξελικτικό υπόβαθρο των IRs-HJs που σχετίζονται με τα μεταθετά στοιχεία, απομονώθηκαν όλες οι αλληλουχίες των LTR και των μη-LTR στοιχείων που στοιχειοθετούνται στη βάση δεδομένων Repbase που περιέχουν IRs-HJs όπως παρατίθενται στην προηγούμενη ανάλυση RepeatMasker και αναλύθηκαν με το MEGA5 λογισμικό. Διαπιστώθηκε ότι τα IRs-HJs υπάρχουν τόσο σε ενεργά όσο και σε αδρανή, εξελικτικά παλιά και νέα, κινητά στοιχεία, τα οποία παρουσιάζονται σε φυλογενετικά/εξελικτικά δέντρα στις παρακάτω εικόνες (Εικ.5.14, 5.15 και 5.16).



Εικόνα 5.14. Φυλογενετικό δέντρο 14 υποοικογενειών των ρετροεμαθετών στοιχείων ALUs που περιέχουν IR-HJs.



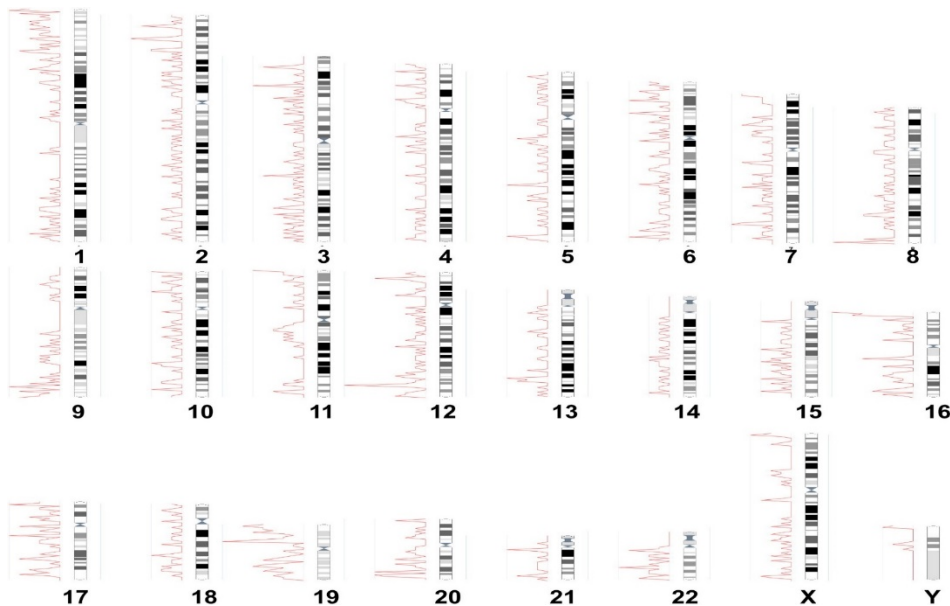
Εικόνα 5.15. Φυλογενετικό δέντρο 13 υποοικογενειών των ρετροεμαθετών στοιχείων LINEs που περιέχουν IR-HJs.



Εικόνα 5.16. Φυλογενετικό δέντρο 5 υποοικογενειών των ενδογενών ρετροϊών HERVs που περιέχουν IR-HJs.

5.3.3 ΚΑΤΑΝΟΜΗ ΤΩΝ IRs-HJs ΚΑΙ ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΗΣ ΜΕΣΗΣ ΑΠΟΣΤΑΣΗΣ ΜΕΤΑΞΥ ΤΟΥΣ ΣΤΑ ΑΝΘΡΩΠΙΝΑ ΧΡΩΜΟΣΩΜΑΤΑ

Δεδομένου ότι η γονιδιωματική κατανομή των IRs-HJs που σχετίζονται με TEs είναι άγνωστη, έγινε απεικόνιση των ίδιων στον ανθρώπινο καρυότυπο, αναλύοντας τις συντεταγμένες HJ στο Ensembl Genome Browser. Βρέθηκε μια σχεδόν ομοιόμορφη κατανομή των IRs-HJs που βρίσκονται σε μεταθετά στοιχεία σε όλα τα χρωμοσώματα με αξιοσημείωτη εξαίρεση αυτή του χρωμοσώματος Y, που περιέχει τον χαμηλότερο αριθμό των υπο μελέτη αλληλουχιών κατά μήκος του χρωμοσώματος (Πίνακας 5.2). Από την άλλη, τα χρωμοσώματα 1 και 19 είχαν τις υψηλότερες συχνότητες εμφάνισης. Αξίζει να σημειωθεί ότι τα ακροκεντρικά χρωμοσώματα δεν εμφάνισαν καμία αλληλουχία στον μικρό βραχίονα εκτός από το χρωμόσωμα 21 που εμφανίζει μόνο μια πρατήρηση στον βραχίονα p (Εικ. 5.17). Επιπλέον, οι περιοχές γύρω από το κεντρομερίδιο του χρωμοσώματος 1 και ιδιαίτερα του χρωμοσώματος 9, τόσο p και q, δεν διαθέτουν IRs-HJs .



Εικόνα 5.17 Η κατανομή των IRs-HJs στα ανθρώπινα χρωμοσώματα. Οι ακολουθίες έχουν επισημανθεί στις αντίστοιχες θέσεις γονιδιώματός και προβάλλονται στο παραπάνω ιδιογράφημα χρωμοσωμάτων, χρησιμοποιώντας τον Ensembl Genome Browser. Οι κόκκινες γραμμές υποδηλώνουν τη συχνότητα των IRs-HJs αλληλουχιών

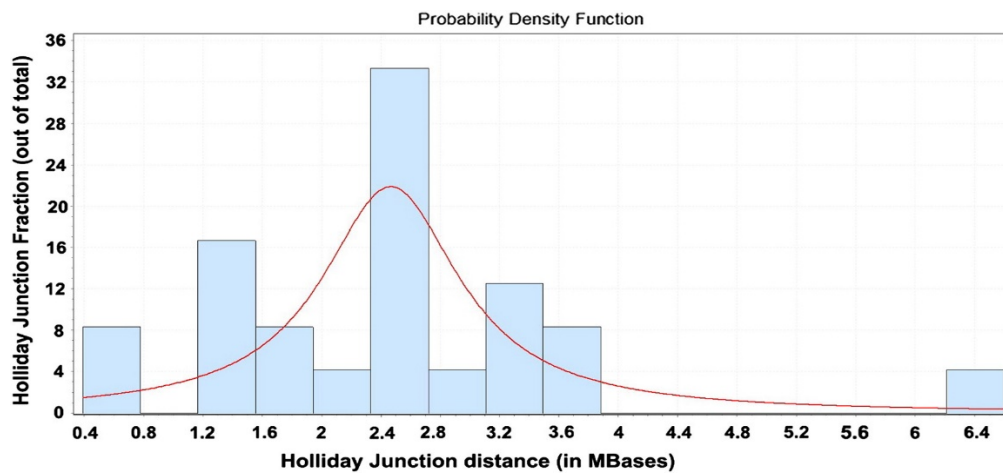
Η συχνότητα ανασυνδυασμού δεν παραμένει ίδια σε όλες της περιοχές κατά μήκος των χρωμοσωμάτων. Προκειμένου να προσδιοριστεί η μέση απόσταση μεταξύ των IRs-HJs μεταξύ των χρωμοσωμάτων, πραγματοποιήθηκε κατανομή των IRs-HJs σύμφωνα με ένα περιβάλλον αριθμητικής υπολογιστικής και μια προγραμματιστική γλώσσα τέταρτης γενιάς

το MATLAB και EasyFit , χρησιμοποιώντας τις χρωμοσωμικές συντεταγμένες που εντοπίστηκαν(Πίνακας 5.2). Η Εικόνα 5.18 δείχνει μια σχεδόν ομοιόμορφη κατανομή μεταξύ των υπο μελέτη αλληλουχιών και υποδεικνύει ότι οι περισσότερες αλληλουχίες εμφανίζουν μια μέση απόσταση της τάξης των 2,4 Mb. Οι ράβδοι της κατανομής παρουσιάζουν επίσης σημαντικές υψηλότερες κορυφές στις αποστάσεις των HJs της τάξης των 1,2 Mb και 3,2 Mb.

Chromosomes	Number of Hits	HJ/Mb
1	637	0.39
2	176	1.38
3	80	2.475
4	56	3.414
5	52	3.478
6	82	2.086
7	65	2.447
8	58	2.524
9	55	2.567
10	51	2.656
11	55	2.454
12	75	1.78
13	30	3.84
14	28	3.832
15	36	2.847
16	65	1.39
17	64	1.268

18	31	2.519
19	91	0.649
20	39	1.619
21	18	2.672
22	34	1.508
X	48	3.235
Y	9	6.6

Πίνακας 5.2 Συνολικός αριθμός των IRs-HJs στο ανθρώπινο γονιδίωμα και απεικόνιση της μέσης απόστασης μεταξύ τους στα ανθρώπινα χρωμοσώματα

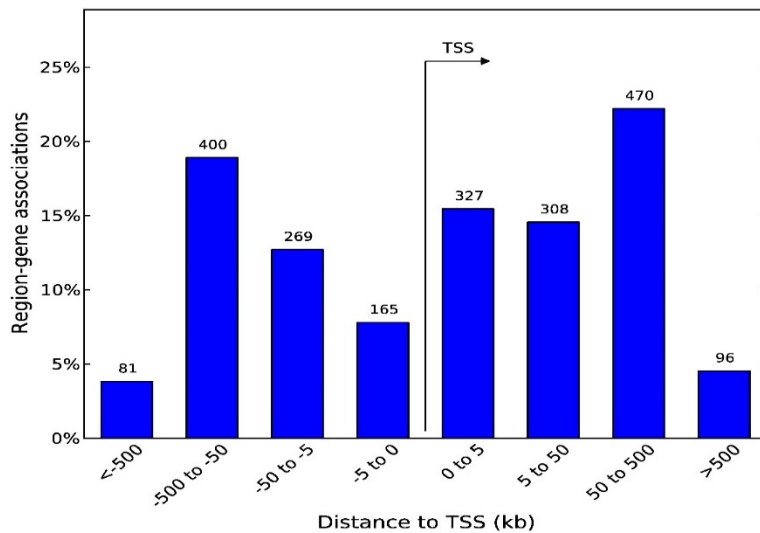


Εικόνα 5.18. Απεικόνιση της μέσης απόστασης μεταξύ των αλληλουχιών IRs-HJs στο γονιδίωμα. Παρατηρείται μια κατανομή σχεδόν στα όρια της γκαουσιανής κατανομής με συγκέντρωση των περισσότερων παρατηρήσεων στην τιμή που ισοούται με μέση απόσταση ίση με 2.4 Mb

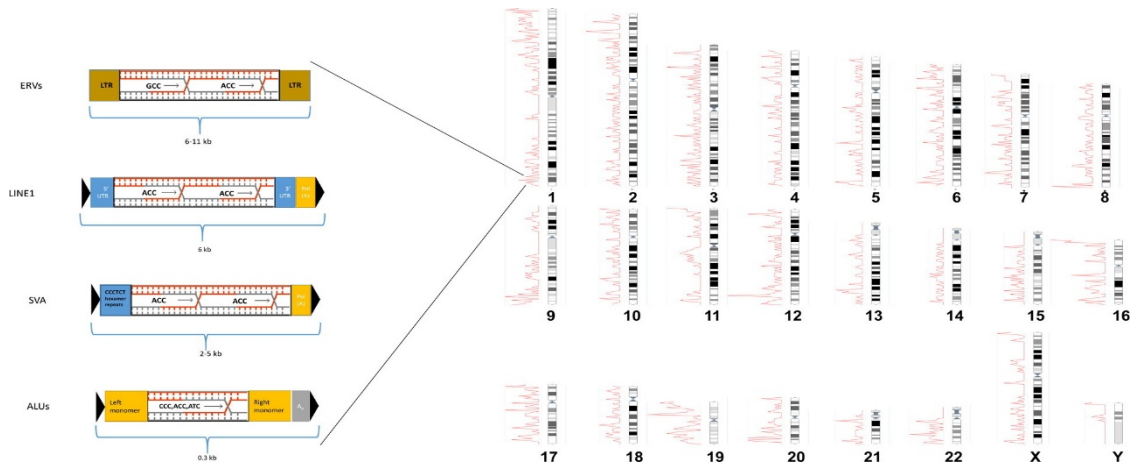
5.3.4 ΑΝΑΛΥΣΗ ΕΓΓΥΤΗΤΑΣ ΤΩΝ IRs-HJs ΣΕ ΘΕΣΕΙΣ ΕΝΑΡΞΗΣ ΤΗΣ ΜΕΤΑΓΡΑΦΗΣ ΓΟΝΙΔΙΩΝ

Τα συμβάντα ανασυνδυασμού που εμφανίζονται κοντά ή μέσα στα γονίδια συνήθως οδηγούν σε απώλεια γονιδίου, σύντηξη γονιδίων ή σε οποιαδήποτε αλλαγή στην γονιδιακή έκφραση. Για να αναλυθεί η σχετική κατανομή των HJs κοντά σε TSS γονιδίων (θέση έναρξης

της μεταγραφής), χρησιμοποιήθηκε το εργαλείο βιοπληροφορικής GREAT. Η παρουσία των IRs-HJs που σχετίζονται με TEs στις περιοχές γονιδίων και ειδικότερα σε TSS παρουσιάζεται στην Εικόνα 5.19. Βρέθηκε ότι ένας σημαντικός αριθμός HJs ($p < 0,005$) βρίσκεται κοντά σε 1854 ανθρώπινα γονίδια σε απόσταση μικρότερη από 1 Mb του γονιδίου TSS και μόνο 23 παρατηρήσεις από τις 1319 αλληλουχίες δεν σχετίζονται με κανένα γονίδιο. Από τις υπόλοιπες 1298 ακολουθίες HJ, 492 κατανέμονται σε αποστάσεις 0 έως 5 kb κοντά σε TSS γονιδίων, ακολούθως 577 κατανέμονται σε αποστάσεις 5 έως 50 kb, 870 κατανέμονται σε αποστάσεις 50 έως 500 kb και 177 κατανέμονται σε αποστάσεις μεγαλύτερες από 500 kb μακριά από το TSS.



Εικόνα 5.19. Κατανομή των IRs-HJs αλληλουχιών με κριτήριο την απόστασή τους από τις θέσεις έναρξης της μεταγραφής σε ανθρώπινα γονίδια. Οι ακολουθίες αναλύθηκαν με το βιοπληροφορικό εργαλείο GREAT στις προεπιλεγμένες ρυθμίσεις. Το διάγραμμα δείχνει τη σχετική απόσταση (0-5, 5-50, 50-500 και 500-1000 kb) των υπο μελέτη αλληλουχιών από TSS γονιδίων.

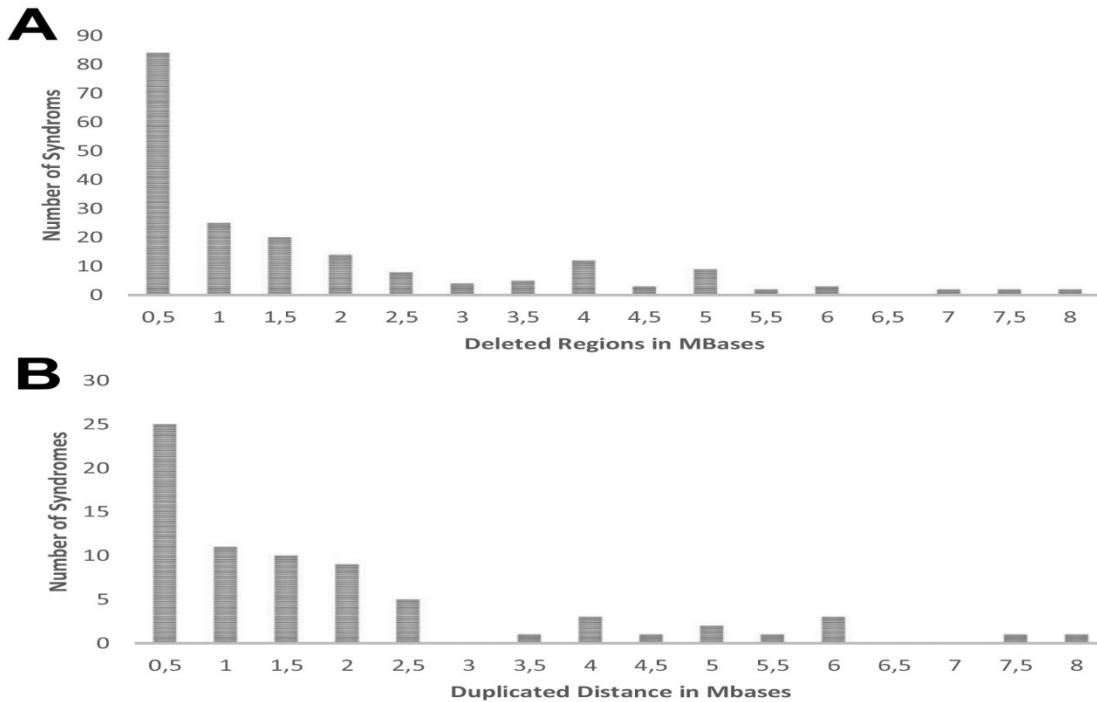


Εικόνα 5.20 Γραφική απεικόνιση των IRs-HJs σε ρετρομεταθετα στοιχεία που λειτουργούν ως φορείς για τη διασπορά στο γονιδίωμα

5.4 ΣΥΣΧΕΤΙΣΗ ΤΩΝ IRs-HJs ΑΛΛΗΛΟΥΧΙΩΝ ΜΕ ΣΥΝΔΡΟΜΑ ΜΙΚΡΟΕΛΛΕΙΨΕΩΝ ΚΑΙ ΜΙΚΡΟΔΙΠΛΑΣΙΑΣΜΩΝ

Οι γενετικές διαταραχές είναι ασθένειες που προκύπτουν από την απώλεια ή πρόσθεση γενετικού υλικού στο ανθρώπινο γονιδίωμα. Οι πιο κοινές γενετικές διαταραχές χωρίζονται σε δύο βασικές κατηγορίες: τα σύνδρομα μικροελλείψεων και μικροδιπλασιασμών που αντιπροσωπεύουν περισσότερο από το 10% όλων των γενετικών ανωμαλιών. Πραγματοποιήθηκε «εξόρυξη δεδομένων» για όλα τα σύνδρομα MMS και διαπιστώθηκε ότι τα περισσότερα σύνδρομα (40% των συνολικών συνδρόμων) είναι μικρότερα από 1 Mb γράφημα (Εικόνα 5.21). Για να προσδιοριστεί περαιτέρω η σχέση μεταξύ των συνδρόμων και των υπο μελέτη αλληλουχιών, δημιουργήθηκαν 2 βάσεις δεδομένων σε bed format, η μια να περιέχει όλα τα σύνδρομα μικροελλείψεων σε όλο το ανθρώπινο γονιδίωμα με 213 καταχωρήσεις και η άλλη περιέχει όλα τα σύνδρομα μικροδιπλασιασμών με 82 καταχωρήσεις, καθεμία από τις οποίες αποτελείται από το όνομα του συνδρόμου και τις συντεταγμένες των σημείων θραύσεων. Αντίστοιχα, δημιουργήθηκαν 7 βάσεις δεδομένων σε bed format, 1 για τις 6 IRs-Hjs σε όλο το ανθρώπινο γονιδίωμα από τα προηγούμενα δημοσιευμένα δεδομένα και 6 επιμέρους βάσεις για τις 6 μεμονωμένες αλληλουχίες IR-Hjs με τις συντεταγμένες τους σε κάθε χρωμόσωμα. Όλες οι βάσεις δεδομένων δημιουργήθηκαν

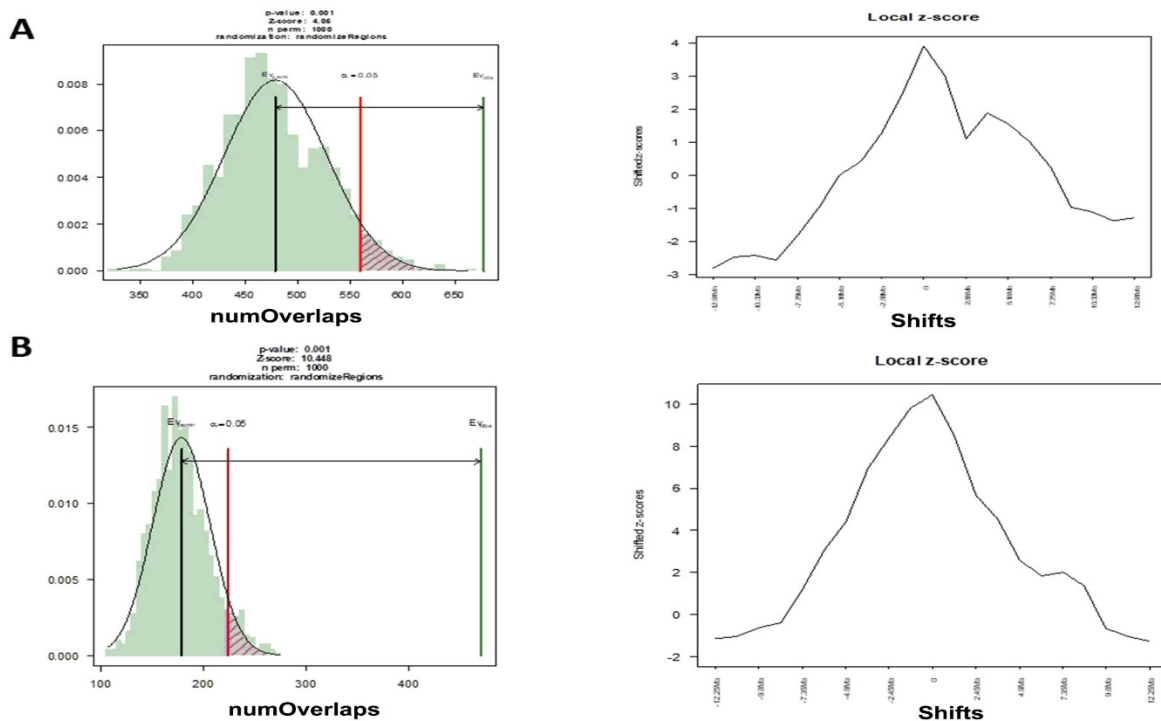
σύμφωνα με δύο εκδόσεις του ανθρώπινου γονιδιώματος, δηλαδή hg19 και hg38. Περισσότερα σύνδρομα ταυτοποιήθηκαν με τον αλγόριθμό χρησιμοποιώντας την hg19 έκδοση, επομένως εκμεταλλευτήκα κυρίως τη hg19 έκδοση για μεγαλύτερη δυνατότητα κάλυψης συνδρόμων.



Εικόνα 5.21. Διαγράμματα κατανομής συχνότητας που περιέχουν όλα τα σύνδρομα μικροελλείψεων και μικροδιπλασιασμών. Τα περισσότερα σύνδρομα συνδέονται με CNVs μεγέθους από 1 Mb και κάτω. Στο A διάγραμμα απεικονίζονται τα σύνδρομα μικροελλείψεων και στο B διάγραμμα τα σύνδρομα μικροδιπλασιασμών

Μέχρι σήμερα δεν είναι γνωστό εάν τα IRs-HJs συνδέονται με την εμφάνιση συνδρόμων. Η παρουσία IRs που σχηματίζουν HJs εντός των σημείων θραύσης ενός συνδρόμου (MMS) μπορεί να αποτελέσει ένδειξη ενός πιθανού μηχανισμού δημιουργίας των εν λόγω συνδρόμων. Για το λόγο αυτό, μέσω της γλώσσας προγραμματισμού R καθορίσαμε την αλληλεπικάλυψη και στατιστική συσχέτιση μεταξύ των δεδομένων στις προσωπικές βάσεις με τις IRs-HJs και τα MMS. Ειδικότερα μέσω του πακέτου Regioner αποκαλύφθηκε ποιες ακολουθίες IRs-HJs συνδέονται με σύνδρομα βάσει στατιστικής σημαντικότητας και ποιες βάσει τυχαιότητας. Το πακέτο regioneR απαιτείται για την αξιολόγηση της συσχέτισης μεταξύ ομάδων γενετικών περιοχών και γενετικών χαρακτηριστικών. Η αξιολόγηση

πραγματοποιείται με στατιστική ανάλυση που βασίζεται κυρίως σε μεθόδους τυχαιοποίησης των δεδομένων. Ειδικότερα, βρέθηκαν 677 από τα συνολικά 2982 IRs-HJs σε 116 σύνδρομα μικροελλείψεων, που εμφανίζονται πιο συχνά στα ακόλουθα σύνδρομα (55 IRs-HJs στο GABRD, 28 IRs-HJs σε 10q25q26, 26 IRs-HJs σε JBS, 23 IRs-HJs στο Di George κ.λπ.), καθώς και 420 IRs-HJs σε 54 σύνδρομα μικροδιπλασιασμών που εμφανίζονται πιο συχνά στα ακόλουθα σύνδρομα (65 IRs-HJs σε 16p13.3, 55 IRs-HJs σε GABRD, 42 IRs-HJs σε 22q13, 40 IRs-HJs σε 17p13.3, 23 IRs-HJs σε 22q11.2, κ.λπ.) με τιμές p-values / z-scores ίσες με 0.00019 / 4.05 and 0.000000 / 10,44 αντίστοιχα. Στο σχήμα 3 απεικονίζονται τα διαγράμματα τιμών p / z-scores που αντιστοιχούν στα IRs-HJs που βρίσκονται στα σύνδρομα ΜΜΣ.



Εικόνα 5.22. Διαγράμματα p-values και z-score που αποδεικνύουν τη στατιστική σημασία της θέσης των IRs-HJs που εντοπίζονται σε σύνδρομα. Α) Διάγραμμα p-value και z-score για όλα τα IRs-HJs σε σύνδρομα μικροελλείψεων Β) Διάγραμμα p-value και z-score για όλα τα IRs-HJs σε σύνδρομα μικροδιπλασιασμών

Το διάγραμμα με την τιμή του Z-Score δηλώνει ότι η μετακίνηση των ακολουθιών IRs-HJs έξω από τα όρια των συνδρόμων παράγει μια πτώση στο Z-Score και έτσι, δείχνει ότι η σχέση μεταξύ συνδρόμων και υπο μελέτη περιοχών εξαρτάται από την ακριβή θέση των IRs-HJs μέσα στα όρια των συνδρόμων και δεν είναι τυχαία. Αναφορικά με τα διαγράμματα των p-values, με το σιέλ χρώμα εντός της γκαουσιανής κατανομής αναπαριστώνται ο αριθμός των

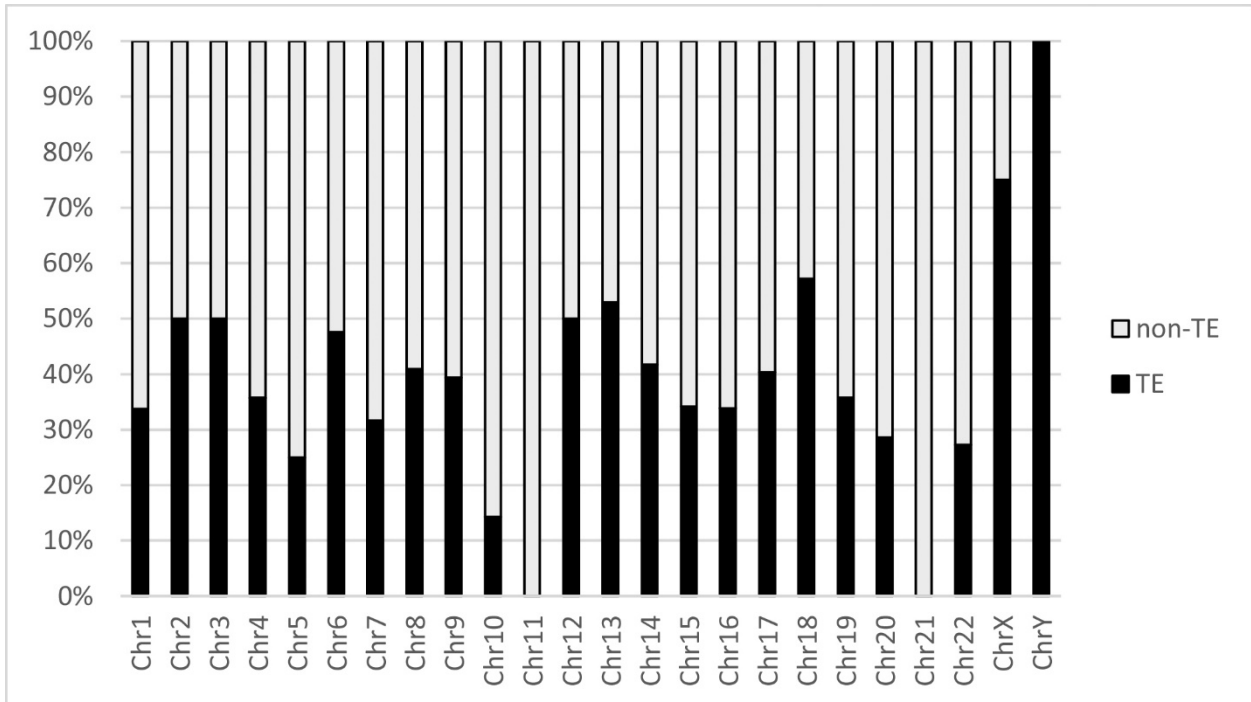
επικαλύψεων τυχαίων δεκανουκλεοτιδίων με τη μορφή ανεστραμμένων επαναλήψεων (που δημιουργούνται αυτομάτως από τους αλγορίθμους ανάλυσης) με τα υπο μελέτη σύνδρομα οι οποίες είναι συγκεντρωμένες γύρω από τη κόκκινη μπάρα που αντιπροσωπεύει το μέσο όρο, ενώ η πράσινη μπάρα δηλώνει τον αριθμό των επικαλύψεων με το εξεταζόμενο πακέτο δεδομένων, δηλαδή με τις περιοχές των IRs-HJs, ο οποίος φαίνεται να είναι πολύ μεγαλύτερος από το αναμενόμενο. Η κόκκινη μπάρα αποτελεί το επίπεδο σημαντικότητας (significant limit) α , το οποίο έχει οριστεί σε 0,05. Τα αποτελέσματα αποδεικνύουν ότι τα IRs-HJs αλληλοεπικαλύπτονται με τα MMS, λόγω των στατιστικά σημαντικών αποτελεσμάτων που προκύπτουν από τις τιμές p-value και Z-score, οι οποίες πληρούν τις προϋποθέσεις για τη στατιστικά σημαντική συσχέτιση δύο πακέτων δεδομένων ($p\text{-value} < \alpha = 0.05$, $Z\text{-score} > 2.5$). Τα σύνδρομα μικροελλείψεων (Α γράφημα) δείχνουν μια στατιστικά σημαντική συσχέτιση με τις IRs-HJs αλληλουχίες. Όπως τα σύνδρομα μικροελλείψεων έτσι και τα σύνδρομα μικροδιπλασιασμών (γράφημα Β) δείχνουν μια πολύ σημαντική συσχέτιση με ακόμα μεγαλύτερο Z-σκορ και πολύ χαμηλότερη p-value με τις υπο μελέτη αλληλουχίες (IRs-HJs). Από τα συνολικά 2982 IRs-HJs στο ανθρώπινο γονιδίωμα, τα 831 βρίσκονται μέσα στις περιοχές των εν λόγω συνδρόμων MMS. Το κάθε αλληλουχιακό IR-HJs δεκαμερές δείχνει διαφορετική συχνότητα εμφάνισης στα MMS. Ειδικότερα, οι αλληλουχίες # 1, # 2, # 3 έχουν βρεθεί ότι έχουν άμεση σχέση με 165 MMS σύνδρομα και πράγματι με πολύ χαμηλό p-value. Συγκεκριμένα, η αλληλουχία # 1 εμφανίζεται 302 φορές στα σύνδρομα Μικροελλείψεων (τιμή $p < 0,001$, $z\text{-score} = 3,147$) και 258 φορές στα σύνδρομα Μικροδιπλασιασμών (τιμή $p < 0,001$, $z\text{-score}: 13,107$). Η αλληλουχία # 2 εμφανίζεται 162 φορές στα σύνδρομα Μικροελλείψεων (τιμή $P < 0,001$, Z-βαθμολογία: 2,529) και 101 φορές στα σύνδρομα Μικροδιπλασιασμών (P-τιμή $< 0,001$, Z-βαθμολογία: 5,599). Η αλληλουχία # 3 εμφανίζεται 89 φορές σε σύνδρομα Μικροελλείψεων (τιμή $p < 0,001$, z-βαθμολογία: 3,543) και 51 φορές στα σύνδρομα Μικροδιπλασιασμών ($p\text{-value} < 0,001$, ζ-βαθμολογία: 5,228). Ο πίνακας 5.3 δείχνει ποιες από τις IRs-HJs αλληλουχίες σχετίζονται με τα σύνδρομα (MMS) βάσει στατιστικής σημαντικότητας και ποιες βάσει τύχης επιδεικνύοντας τα p-values, z-scores και τον αριθμό παρατηρήσεων στο σύνολο των συνδρόμων μικροελλείψεων και μικροδιπλασιασμών.

IRs-HJs Sequences	Microdeletion Syndromes p-values/ Z-scores/hits in Syndromes	Microduplication Syndromes p-values/ Z-scores/hits in syndromes
CCGGGCCCGG #1	<0.001/ =3.147/ 302	<0.001/ =13.107/ 258
CCGGCGCCGG #2	<0.001/ =2.529/ 162	<0.001/ =5.599/ 101
CCGGGACCGG #3	<0.001/ =3.543/ 89	<0.001/ =5.228/ 51
CCGGTACCGG #4	= 0.2/ =0.827/ 20	=0.332/ =0.551/ 8
CCGATACCGG #5	=0.255/ =0.430/ 9	=0.231/ =-0.946/ 4
TCGGTACCGG #6	=0.445/ =0.319/ 11	=0.377/ =0.430/ 4

Πίνακας 5.3. Ο πίνακας δείχνει ποιες από τις IRs-HJs αλληλουχίες σχετίζονται με τα σύνδρομα (MMS) βάσει στατιστικής σημαντικότητας και ποιες βάσει τύχης επιδεικνύοντας τα p-values, z-scores και τον αριθμό παρατηρήσεων στο σύνολο των συνδρόμων μικροελλείψεων και μικροδιπλασιασμών.

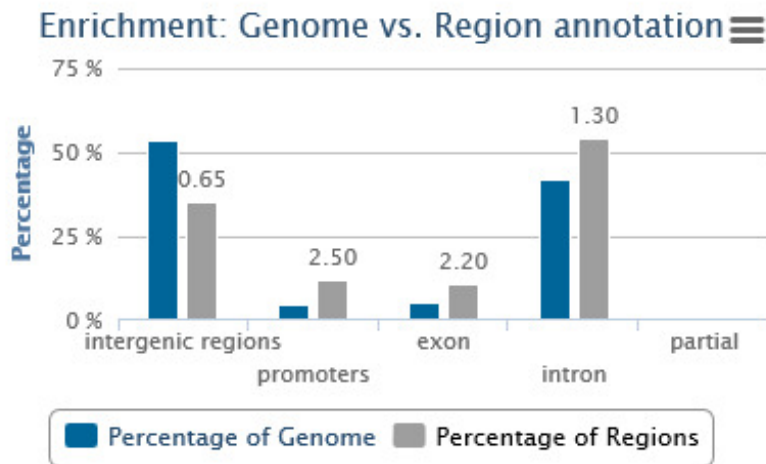
5.4.1 REPEATMASKER ANALYSIS ΚΑΙ ENRICHMENT ANALYSIS ΤΩΝ IRs-HJs ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΣΕ ΣΥΝΔΡΟΜΑ (MMS)

Τα μεταθετά στοιχεία του DNA (TEs) χρησιμεύουν ως υποστρώματα για τον μη ομόλογο ανασυνδυασμό εντός ή μεταξύ των χρωμοσωμάτων κατά τη διάρκεια της μείωσης και της μίτωσης. Στην εικόνα 5.23 παρουσιάζεται το ποσοστό των IRs-HJs που συνδέονται με TEs και αυτών που δεν συνδέονται με TEs και βρίσκονται μέσα σε MMS σε όλα τα ανθρώπινα χρωμοσώματα. Η κατανομή των TE και των μη-TE IRs-HJs σε MMS είναι χρωμοσωμοεξαρτώμενη.



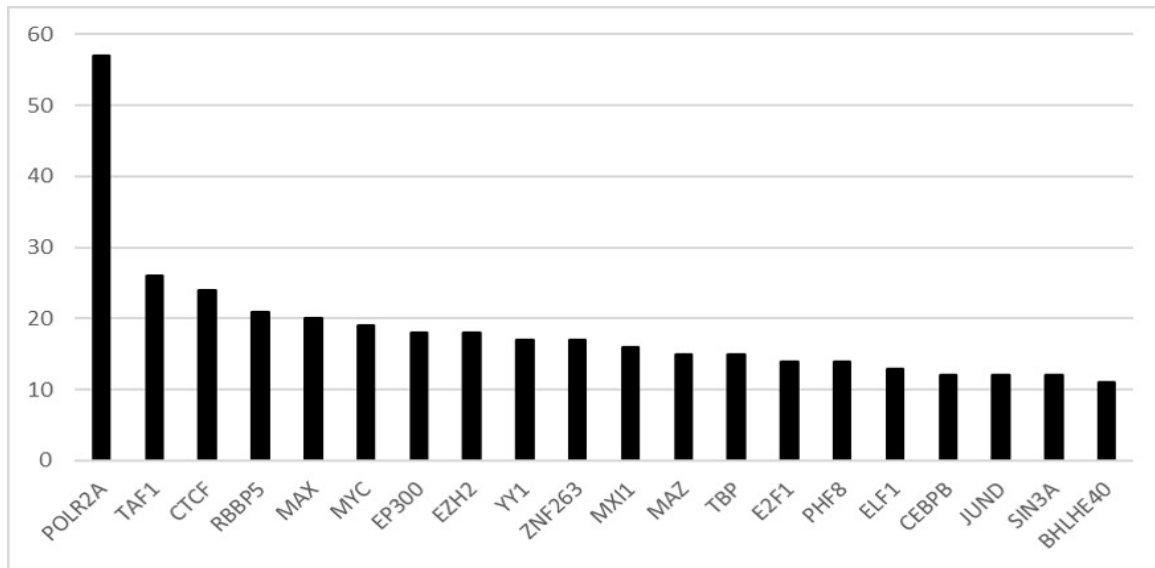
Εικόνα 5.23. Η χρωμοσωμοεξαρτώμενη κατανομή των TE και των μη-TE (IRs-HJs) σε MMS..

Η Repeatmasker ανάλυση έδειξε ότι οι κυρίαρχες 3 οικογένειες των ρετρομεταθετών στοιχείων που συνυπάρχουν με τα IRs-HJs μέσα στα υπο μελέτη σύνδρομα εντοπίζονται στις οικογένειες των ALUs, LINEs, HERVs. Για να αποκτήσουμε περαιτέρω γνώση της γονιδιακής κατανομής των περιοχών IRs-HJs-MMS, πραγματοποιήθηκε γονιδιωματική ανάλυση εμπλουτισμού (enrichment analysis) χρησιμοποιώντας το εργαλείο EIDorado, από το διακομιστή Genomatix. Αυτό το εργαλείο ταξινομεί κάθε περιοχή ανάλογα με το αν βρίσκεται σε διαγονιδιακές περιοχές, υποκινητές, εξόνια γονιδίων ή εσόνια ή εν μέρει σε περισσότερες από μία κατηγορίες. Διαπιστώσαμε ότι 54,1% των γονιδιωματικών περιοχών βρίσκονται σε ιντρόνια (1,3χ εμπλουτισμένο), 35,0% σε διαγονιδιακές περιοχές (εμπλουτισμένο κατά 0,7χλ.), 12,0% σε υποκινητές (2,5χ εμπλουτισμένο) και 10,8% σε εξόνια (2,2χ εμπλουτισμένο). Τα συνοπτικά αποτελέσματα φαίνονται στην εικόνα 5.24. Γενικά, οι υπό μελέτη αλληλουχίες είναι πιο εμπλουτισμένες και πιο πιθανό να εμφανιστούν σε περιοχές που σχετίζονται με γονίδια.



Εικόνα 5.24 Αναμενόμενος αριθμός παρατηρήσεων έναντι παρατηρούμενος αριθμός παρατηρήσεων όσον αφορά την παρουσία των IRs-HJs στο ανθρώπινο γονιδίωμα (54.1 % των περιοχών σε ιντρόνια, 35% σε διαγονιδιακές περιοχές, 12% σε υποκινητές και 10.8% σε εξόνια)

Η εικόνα 5.24 δηλώνει ότι οι περισσότερες υπο μελέτη αλληλουχίες μπορεί να διαδραματίσουν και ρυθμιστικό ρόλο. Τα δεδομένα από το ENCODE υποστηρίζουν ότι τουλάχιστον το 80% των αλληλουχιών συμμετέχει σε τουλάχιστον ένα βιοχημικό μονοπάτι. Για να αποκτήσουμε περαιτέρω γνώση σχετικά με την πιθανή ρυθμιστική δράση των υπο μελέτη περιοχών, προσδιορίστηκε αν τα IR-HJs-MMS εμφανίζονται σε περιοχές πρόσδεσης των μεταγραφικών παραγόντων. Τα αποτελέσματά συνοψίζονται στην εικόνα 5.25 με κύριο εύρημα τον εντοπισμό των 20 κορυφαίων θέσεων πρόσδεσης με χαρακτηριστικές τις POLR2A (57 θέσεις), TAF1 / TFIID (26 θέσεις) και TBP Πρωτεΐνη σύνδεσης TATA) (15 θέσεις). Ένα πρόσθετο ενδιαφέρον εύρημα είναι ότι τα IRs-HJs βρέθηκαν σε περιοχές που συνδέονται με τον μεταγραφικό παράγοντα CTCF (24 θέσεις), ένας παράγοντας που είναι γνωστό ότι εμπλέκεται σε πολλές κυτταρικές διεργασίες, συμπεριλαμβανομένης της μεταγραφικής ρύθμισης, του V (D) J ανασυνδυασμού και της ρύθμισης της αρχιτεκτονικής της χρωματίνης.



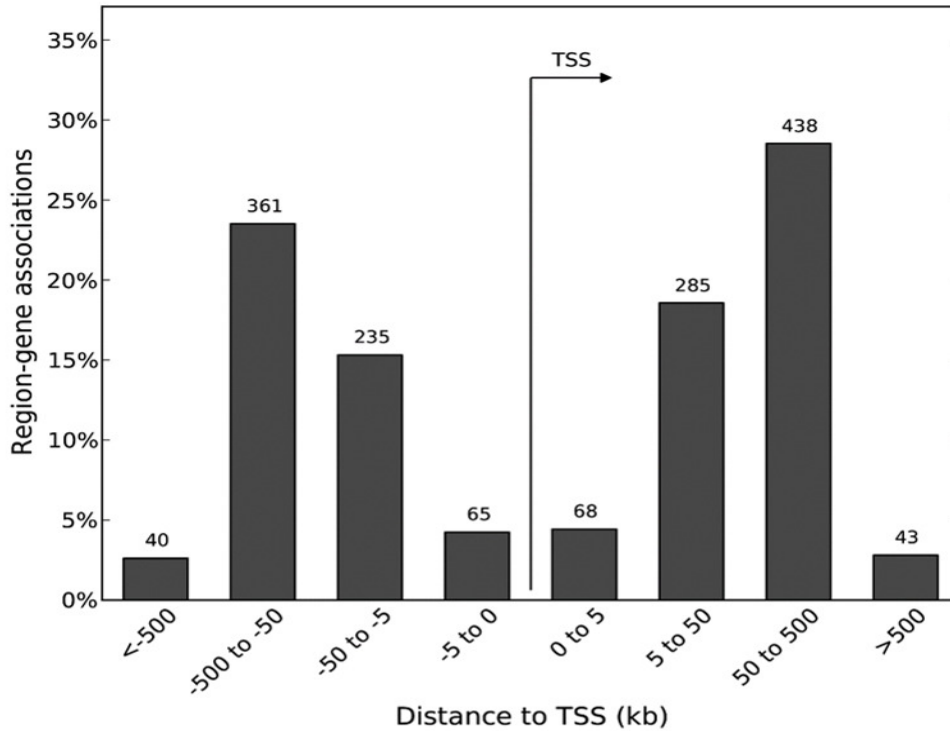
Εικόνα 5.25. Προσδιορισμός του ρυθμιστικού δυναμικού των γονιδιωματικών περιοχών που περιέχουν τις υπο μελέτη IRs-HJs. Οι IRs-HJs εμφανίζονται σε συγκεκριμένες περιοχές πρόσδεσης παραγόντων μεταγραφής.

Στη συνέχεια, πραγματοποιήθηκε ανάλυση για να διαπιστωθεί εάν οι αλληλουχίες IRs-HJs βρίσκονται μέσα ή κοντά σε ανθρώπινα γονίδια. Η ανάλυσή παρουσιάζεται στην εικόνα 5.26. Βρέθηκε ότι οι γονιδιωματικές θέσεις των 831 IRs-HJs συνδέονται με 877 γονίδια συνολικά. Η επακόλουθη ανάλυση εμπλουτισμού έδειξε ότι ένας σημαντικά μεγάλος αριθμός αυτών των γονιδίων συμμετέχουν σε μονοπάτια που ελέγχουν τα αρχικά στάδια της εμβρυικής ανάπτυξης και της μορφογένεσης οργάνων. Μεταξύ άλλων, βρέθηκε ότι συμμετέχουν στην:

- μορφογένεση εμβρυικών οργάνων (30 γονίδια)
- κυτταρική διαφοροποίηση (12 γονίδια)
- στην ανάπτυξη νεφρονικού επιθηλίου (9 γονίδια)
- διατήρηση σωματικών στελεχιαίων κυττάρων βασικών στοιβάδων (9 γονίδια)
- θετική ρύθμιση του κυτταρικού πολλαπλασιασμού στο μυελό των οστών (3 γονίδια)
- πολυκυτταρική οργανική απόκριση στο στρες (11 γονίδια)
- ανάπτυξη εμβρυϊκού πεπτικού συστήματος (8 γονίδια)
- ανάπτυξη μυϊκών κυττάρων (3 γονίδια).

Συνολικά, τα IRs-HJs-MMS σχετίζονται με τις περιοχές 877 γονιδίων, πολλά από τα οποία ελέγχουν την φυσιολογική ανθρώπινη εμβρυϊκή ανάπτυξη.

A



B

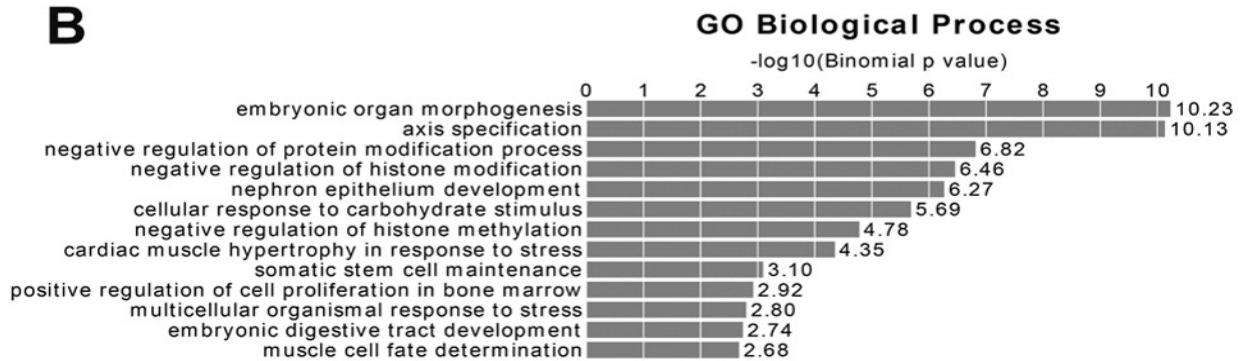


Figure 5.26. A) Κατανομή των αλληλουχιών IR-HJ σε σχέση με τις TSS ανθρώπινων γονιδίων. Το διάγραμμα δείχνει τη σχετική απόσταση (0-5, 5-50, 50-500 και 500-1000 kb) αλληλουχιών IR-HJ από τις TSS. B) Η ανάλυση εμπλουτισμού GO αποκάλυψε ότι ένας σημαντικά μεγάλος αριθμός αυτών των γονιδίων συμμετέχουν σε μονοπάτια που ελέγχουν τα αρχικά στάδια ανάπτυξης και μορφογένεσης οργάνων.

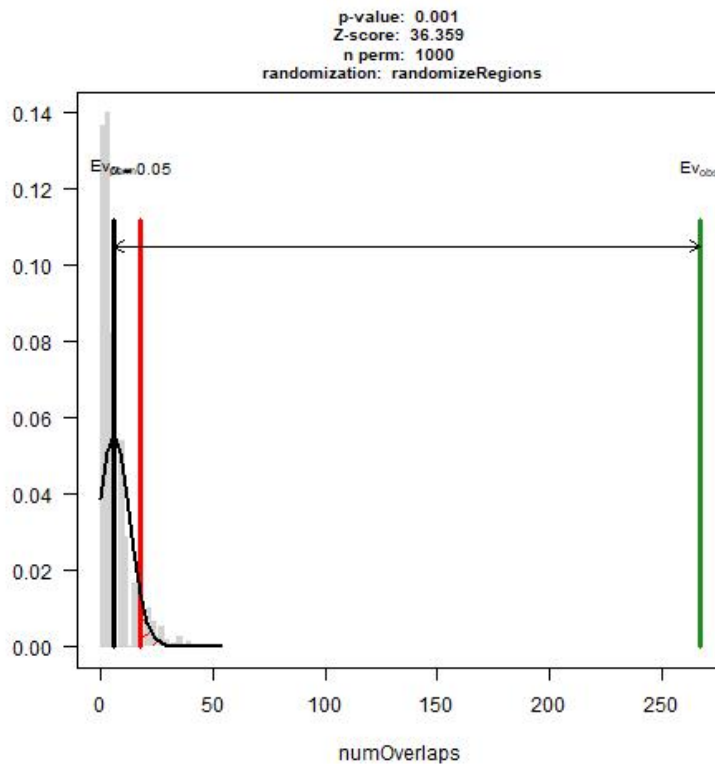
5.5 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΑΛΛΗΛΟΥΧΙΑΚΩΝ ΠΑΡΑΛΛΑΓΩΝ ΤΟΥ PRDM9 ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΟΥΣ ΚΑΡΚΙΝΟΥΣ ΤΟΥ ΜΑΣΤΟΥ, ΤΗΣ ΚΕΦΑΛΗΣ ΚΑΙ ΤΡΑΧΗΛΟΥ ΚΑΙ ΤΟΥ ΑΔΕΝΟΚΑΡΚΙΝΩΜΑΤΟΣ ΤΟΥ ΠΝΕΥΜΟΝΑ

Η πρωτεϊνική δομή του PRDM9 είναι εξαιρετικά διατηρημένη σε όλα τα είδη εκτός από την τερματική διάταξη δακτύλων ψευδαργύρου, η οποία είναι ιδιαίτερα πολυμορφική, τόσο στον αριθμό των δακτύλων που παρουσιάζονται όσο και στην ταυτότητα των τριών αμινοξέων σε κάθε δάκτυλο που έρχονται σε επαφή με το DNA και προσδιορίζουν την ειδικότητα πρόσδεσης DNA του PRDM9. Συνεπώς το γονίδιο παρουσιάζει πολλές παραλλαγές στην αλληλουχία του. Μετά από εκτενή αναζήτηση σε βάσεις δεδομένων όπως η NCBI, βρέθηκε ότι η μορφή του αλληλουχιακού μοτίβου πρόσδεσης του PRDM9 έχει τη μορφή d[CCNCCNTNNCCNC] . Για τις πιθανές παραλλαγές του γονιδίου, βρέθηκαν όλες οι δυνατές αλληλουχίες για N=A , N=C ,N=G, N=T αντίστοιχα . Αυτό κατέστη δυνατό με την δημιουργία script στην γλώσσα προγραμματισμού C. Από το Script βρέθηκαν 1024 παραλλαγές του νουκλεοτιδίου οι οποίες εντοπίστηκαν σε όλο το ανθρώπινο γονιδίωμα.

Από τις αναλύσεις για κάθε αρχείο καρκίνου βρέθηκε ότι για τον καρκίνο του μαστού, 13 παραλλαγές του PRDM9 είναι στατιστικά σημαντικές, μέσα στις ευρύτερες περιοχές δίκλωνων θραύσεων:

CCACCATCACCAC	CCACCTTAACCTC
CCACCATGGCCGC	CCACCTTGGCCTC
CCACCCTACCCAC	CCACCTTTCCCCC
CCACCCTGCCCCC	CCACCTTTCCCGC
CCACCGTACCCAC	CCACCTTTCCCC
CCACCGTCCCGC	CCACCTTTTCCTC
CCACCGTTGCCAC	

Βρέθηκε ότι η αλληλουχία d(CCACCATCACCAC) που σχετίζεται με τον καρκίνο του μαστού είναι η πιο σημαντική στατιστικά καθώς έχει το μικρότερο p-value και μεγαλύτερο z-score από τις υπόλοιπες.



Εικόνα 5.27. Γραφική απεικόνιση για την πιο στατιστικά σημαντική (p -value= 0.001 και z -score=36.359) αλληλουχία d(CCACCATCACCAC) που εμφανίζεται στον καρκίνο του μαστού.

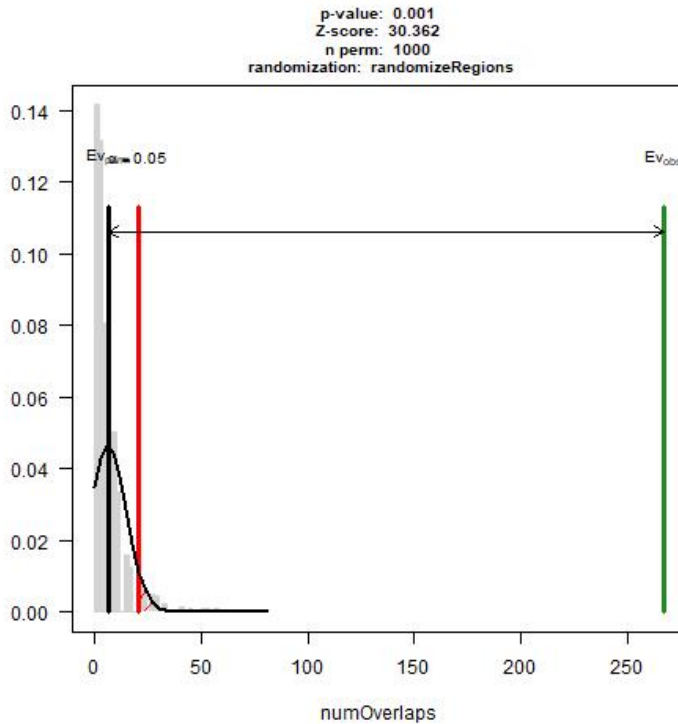
Για τον καρκίνο της κεφαλής και τραχήλου (HN) βρέθηκαν 5 παραλλαγές του PRDM9 που είναι στατιστικά σημαντικές στις ευρύτερες περιοχές των δίκλωνων θραύσεων οι οποίες παρατίθενται πιο κάτω:

CCACCATACCTC CCACCCTAACCGC

CCACCATCACCAC CCACCTTCACCAC

CCACCATTGCCCC

Βρέθηκε ότι η αλληλουχία d(CCACCATCACCAC) που σχετίζεται με τον καρκίνο της κεφαλής και τραχήλου (HN) είναι η πιο σημαντική στατιστικά καθώς έχει το μικρότερο p-value και μεγαλύτερο z-score από τις υπόλοιπες.

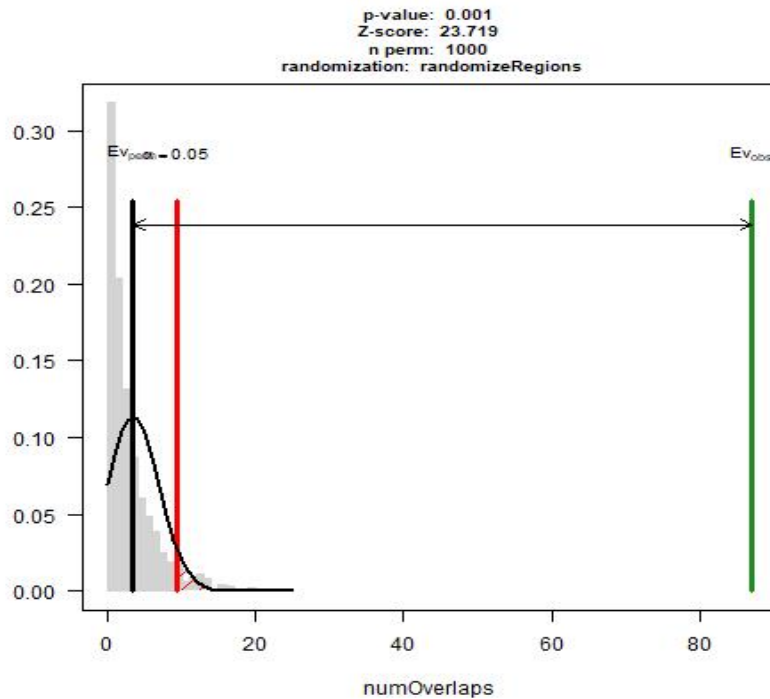


Εικόνα 5.28. Διάγραμμα p-value και z-score για την πιο στατιστικά σημαντική (p-value= 0.001 και z-score=30.362) αλληλουχία d(CCACCATCACCAC) που εμφανίζεται στο αδενοκαρκίνωμα του πνεύμονα (LUAD)

Τέλος για την περίπτωση του αδενοκαρκινώματος στο πνεύμονα βρέθηκαν 8 παραλλαγές του PRDM9 που είναι στατιστικά σημαντικές στις ευρύτερες περιοχές των δίκλωνων θραύσεων οι οποίες παρατίθενται πιο κάτω:

CCACCATCACCAC	CCACCATCTCCAC
CCACCATCACCTC	CCACCATGACCAC
CCACCATCCCCAC	CCACCCTACCCAC
CCACCATCGCCGC	CCACCCTCGCCGC

με την αλληλουχία d(CCACCATCACCAC) να είναι η πιο σημαντική στατιστικά καθώς έχει το μικρότερο p-value και μεγαλύτερο z-score από τις υπόλοιπες.



Εικόνα 5.29. Διάγραμμα p-value και z-score για την πιο στατιστικά σημαντική (p-value= 0.001 και z-score=23.719) αλληλουχία d(CCACCATCACCAC) που εμφανίζεται στον καρκίνο της κεφαλής και του τραχήλου (HN).

Μελετώντας και συγκρίνοντας όλες τις αλληλουχίες του PRDM9 που εμφανίζονται στις ευρύτερες περιοχές των δίκλωνων θραύσεων των καρκίνων που μελετήθηκαν, βρέθηκε μία μοναδική κοινή αλληλουχία από τις 1024 η οποία εμφανίζεται και στις 3 διαφορετικές περιπτώσεις καρκίνου και είναι της μορφής : **d(CCACCATCACCAC)**

ΚΕΦΑΛΑΙΟ 6

ΣΥΖΗΤΗΣΗ

6.1 ΜΕΓΑΛΥΤΕΡΗ ΣΥΣΣΩΡΕΥΣΗ HOTSPOTS ΑΝΑΣΥΝΔΥΑΣΜΟΥ ΟΔΗΓΕΙ ΣΕ ΛΙΓΟΤΕΡΑ ΣΥΜΒΑΝΤΑ ΑΝΑΣΥΝΔΥΑΣΜΟΥ. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΠΑΡΑΔΕΙΓΜΑ ΤΩΝ 2 ΦΥΛΩΝ.

Από το σύνολο των αποτελεσμάτων, το πρώτο σημαντικό εύρημα, είναι η διαφορά που παρατηρείται μεταξύ των γενετικών μηκών των χρωμοσωμάτων των δύο φύλων όπως παρουσιάζεται στην εικόνα 5.1. Το αποτέλεσμα αυτό συνάδει με συμπεράσματα προηγούμενων μελετών (152) και η διαφορά αυτή οφείλεται στους διαφορετικούς ρυθμούς ανασυνδυασμού μεταξύ της σπερματογένεσης και της ωογένεσης. Τα αρρεν άτομα έχουν μικρότερους ρυθμούς ανασυνδυασμού στους γαμέτες από ότι τα θήλεα δηλαδή οι γενετικοί δείκτες που μελετούνται είναι πιο κοντά μεταξύ τους, με αποτέλεσμα να μην εμφανίζονται το ίδιο συχνά συμβάντα ανασυνδυασμού. Λόγω της συγκριτικά μικρότερης απόστασης των θερμών σημείων ανασυνδυασμού, το συνολικό άθροισμα των γενετικών αποστάσεων από δείκτη σε δείκτη θα είναι μικρότερο από το αντίστοιχο των θήλεων γαμετικών χρωμοσωμάτων.

Παρά το μικρότερο γενετικό μήκος των αρρεν χρωμοσωμάτων, εμφανίζουν περισσότερα θερμά σημεία ανασυνδυασμού καθώς υπολογίστηκε ότι υπάρχουν 4.763 τέτοια σημεία σε αντίθεση με 4.131 στα θήλεα άτομα. Υπενθυμίζεται ότι ως θερμό σημείο ανασυνδυασμού ορίστηκε αυτό που έχει ρυθμό ανασυνδυασμού μεγαλύτερο του 10. Συνδυαστικά, τα παραπάνω αποτελέσματα προδιαθέτουν στην «συσσώρευση» των θερμών σημείων ανασυνδυασμού πάνω στο γονιδίωμα των αρρένων ενώ η πιο ομοιόμορφη κατανομή αυτών στα θηλυκούς γαμέτες, εξηγεί τον διαφορετικό ρυθμό ανασυνδυασμού μεταξύ των φύλων.

Η εικόνα 5.2 παρουσιάζει την μέση απόσταση μεταξύ οποιωνδήποτε δύο διαδοχικών hotspot ανασυνδυασμού για κάθε χρωμόσωμα των δύο φύλων. Αξιοσημείωτο είναι πως οι φυσικές αποστάσεις μεταξύ των θερμών σημείων είναι περίπου ίσες και στα δύο φύλα, με αυτές των αρσενικών να είναι ελαφρώς μικρότερες. Οι παραπάνω μέσες αποστάσεις έχουν πτωτική πορεία από χρωμόσωμα σε χρωμόσωμα, παρόμοια με αυτή του συνολικού μήκους των χρωμοσωμάτων. Με άλλα λόγια, τα μεγαλύτερα χρωμοσώματα έχουν τα θερμά σημεία ανασυνδυασμού τους πιο διάσπαρτα από ότι τα μικρότερα χρωμοσώματα.

Στις γενετικές αποστάσεις μεταξύ δύο διαδοχικών hotspot ανασυνδυασμού τα αποτελέσματα είναι πολύ διαφορετικά. Οι αρρενες γαμέτες εμφανίζουν σχεδόν την μίση γενετική απόσταση από hotspot σε hotspot από ότι οι θηλυκοί γαμέτες. Αυτό υποδεικνύει ότι τα θερμά σημεία ανασυνδυασμού είναι πολύ κοντά μεταξύ τους και οι πιθανότητες για ύπαρξη συμβάντος ανασυνδυασμού μεταξύ οποιωνδήποτε δύο διαδοχικών τέτοιων σημείων είναι μικρότερη στα χρωμοσώματα της σπερματογένεσης από αυτά της ωογένεσης.

Στα επόμενα γραφήματα των εικόνων 5.3 έως 5.13 παρουσιάζεται η συσχέτιση μεταξύ γενετικών και των φυσικών αποστάσεων που καλύπτουν οι γενετικοί δείκτες που χρησιμοποιήθηκαν καθώς και μια εκτίμηση για τους ρυθμούς ανασυνδυασμού κατά μήκος κάθε χρωμοσώματος. Εδώ παρατηρείται αυξομείωση του ρυθμού ανασυνδυασμού ομοιόμορφα σε όλο το μήκος των χρωμοσωμάτων των ωαρίων, δηλαδή εντοπίζονται μέγιστα σημεία ($cM/Mb > 2$) τόσο στις πρώτες 10-20 Mb κάθε χρωματίδας όσο και στις κεντρικές περιοχές των 40-50 Mb και στις τελικές περιοχές που ξεπερνούν τις 150 Mb στα μεγαλύτερα χρωμοσώματα.

Αντιθέτως, στα χρωμοσώματα των σπερματοζωαρίων, τοπικά μέγιστα ($cM/Mb > 2$) παρατηρούνται μόνο στα αρχικά και στα τελικά τμήματα αυτών, με τα κεντρικά να μην έχουν μεγάλους ρυθμούς ανασυνδυασμού. Τα παραπάνω αποτελέσματα ενισχύουν περαιτέρω την υπόθεση ότι τα αρσενικά hotspot εντοπίζονται πολύ κοντά το ένα με το άλλο σε συστάδες, ενώ τα θηλυκά hotspots είναι κατανεμημένα ομοιόμορφα στο γονιδίωμα ώστε όλα να έχουν τις ίδιες πιθανότητες για την δημιουργία ανασυνδυασμού μεταξύ τους. Η ομοιόμορφη κατανομή των θήλεων hotspot ίσως να είναι αναγκαία για την διατήρηση της ποικιλομορφίας στους απογόνους καθώς περισσότεροι ομόλογοι ανασυνδυασμοί οδηγούν σε περισσότερους

και διαφορετικούς απλοτύπους με μεγάλο εύρος συνδυασμών γονιδίων, κάτι που είναι απαραίτητο για την διαδικασία της εξέλιξης και την εμφάνιση νέων χαρακτήρων ή τον αφανισμό άλλων.

6.2 ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ DNA ΩΣ ΦΟΡΕΙΣ ΑΛΛΗΛΟΥΧΙΩΝ ΜΕ ΑΝΑΣΥΝΔΥΑΣΤΙΚΗ ΔΡΑΣΗ.

Πιο συγκεκριμένα, οι δομές Holliday είναι οι ενδιάμεσες δομές που δημιουργούνται κατά τη διάρκεια γεγονότων ομόλογου ανασυνδυασμού. Παρόλο που έχουν διερευνηθεί εκτενώς οι συγκεκριμένες δομές, μέχρι στιγμής δεν είναι γνωστή η ακριβής τους κατανομή στο γονιδίωμα και η σχέση τους, με μεταθετά και ρετρομεταθετά στοιχεία του DNA (TEs) ή άλλες αλληλουχίες με ανασυνδυαστική δράση, όπως η θέση πρόσδεσης της PRDM9.

Στην τρέχουσα μελέτη, έγινε ο καθορισμός των συντεταγμένων των συγκεκριμένων αλληλουχιών IRs που δημιουργούν υπό συνθήκες HJs στο ανθρώπινο γονιδίωμα και αναλύσαμε αν σχετίζονται με TEs. Ένα από τα κύρια ευρήματα είναι η προνομιακή συνύπαρξή τους μέσα σε αλληλουχίες ρετρομεταθετών στοιχείων, που πιθανώς αποδεικνύει τη συνεργιστική δράση και προσαρμοστική φύση των αλληλουχιών που σχετίζονται με γονιδιωματικές αναδιατάξεις, που απαντώνται σχεδόν σε όλους τους τύπους μειωτικών και μιτωτικών ανασυνδυασμών. Το αποτέλεσμα μας προτείνει ότι τα μεταθετά στοιχεία μπορούν να δράσουν ως φορείς που διασπείρουν τις ευαίσθητες σε ανασυνδυασμό αλληλουχίες σε όλο το γονιδίωμα.

Σε ένα πρώτο βήμα, με βάση την αλληλουχία του πυρήνα των ανεστραμμένων επαναλήψεων (IRs) που σχηματίζουν HJ, εντοπίστηκαν όλα τα IRs-HJs στο ανθρώπινο γονιδίωμα βάσει του αλληλουχιακού μοτίβου (CCDRBVYCGG) και βρέθηκαν αυτά που σχετίζονται με μεταθετά και ρετρομεταθετά στοιχεία του DNA (πίνακας 5.1). Τα αποτελέσματα αποδεικνύουν ότι κάθε μοναδική αλληλουχία IRs-HJs σχετίζεται με διαφορετικές οικογένειες μεταθετών στοιχείων (TEs). Αυτό το εύρημα ενδέχεται να σημαίνει ότι η διασπορά των μεταθετών στοιχείων μέσω της ρετρομετάθεσης, οδηγεί σε ταυτόχρονη

διασπορά των IRs-HJs στο ανθρώπινο γονιδίωμα. Όσον αφορά την κινητικότητα και τη διασπορά των HJs με τη συνδρομή των TEs ως αποτελεσματικών φορέων τους, βρήκαμε επίσης μια σχεδόν 100% ειδικότητα της κάθε ακολουθίας IRs-HJ με συγκεκριμένα ενεργά και ανενεργά ρετρομεταθετά στοιχεία. Και τα δύο παραπάνω, δείχνουν ότι κατά τη διάρκεια της εξέλιξης, διάφορες αλληλουχίες IRs-HJs έχουν «φιλοξενηθεί» από ορισμένα μεταθετά στοιχεία για να λειτουργήσουν μαζί ως "μονάδες ανασυνδυασμού" εντός των χρωμοσωμάτων και ενδεχομένως μεταφέρονται ενεργητικά κατά τη διαδικασία της ρετρομετάθεσης.

Ενδιαφέρον είναι επίσης ότι υπάρχει μια θετική συσχέτιση των συχνοτήτων εμφάνισης κάθε αλληλουχίας IRs-HJs με τη συχνότητα εμφάνισης των συγκεκριμένων ρετρομεταθετών στοιχείων στο ανθρώπινο γονιδίωμα. Τα παραπάνω επαληθεύονται στην εκδοση HG.19 του ανθρώπινου γονιδιώματος, όπου οι συχνότητες εμφάνισης κάθε ακολουθίας IRs-HJs σε μεταθετά στοιχεία εξαρτώνται από το ποσοστό των δινουκλεοτιδίων CG στο αντίστοιχο αλληλουχιακό μοτίβο των IRs-HJs. Αυτό που διαπιστώθηκε είναι ότι όταν οι Αδενίνες ή οι Θυμίνες αυξάνονται από 1 σε 4 στο συντηρημένο αλληλουχιακό μοτίβο των IRs αυτό αυτομάτως επιφέρει κατακόρυφη μείωση στη συχνότητα εμφάνισης των HJs-IRs στο ανθρώπινο γονιδίωμα. Τα μεταθετά στοιχεία της οικογένειας ALU φαίνεται ότι έχουν εξελιχθεί ταυτόχρονα με 3 διαφορετικές αλληλουχίες IRs-HJs, ενώ τα LINEs έχουν εξελιχθεί ταυτόχρονα με 2 από τις αλληλουχίες IRs-HJs με χαρακτηριστικά χαμηλότερο περιεχόμενο των δινουκλεοτιδίων GC και επακόλουθη μείωση του αριθμού αντιγράφων τους στο γονιδίωμα. Επιπλέον, τα LINEs συσχετίζονται με εντελώς διαφορετικές ακολουθίες IRs-HJs από ότι τα ALUs όσον αφορά την αλληλουχία του πυρήνα των IRs-HJs.

Και οι τρεις πιο συνηθισμένοι τύποι ανασυνδυασμού, ο ομόλογος, ο μη-ομόλογος και ο αντιγραφικός ανασυνδυασμός, πυροδοτούνται από τις IRs-HJs ακολουθίες και εφαρμόζονται σε μια ποικιλία κυττάρων και ιστών. Τα ευρήματά της συγκεκριμένης μελέτης δείχνουν έναν μηχανισμό διασποράς που ανακατανέμει όλες τις σημαντικές αλληλουχίες IRs-HJs στο γονιδίωμα, προκειμένου να προκαλέσουν νέα γεγονότα ανασυνδυασμού με υπόστρωμα τα TEs που περιέχουν τα IRs-HJs. Ο ομόλογος ανασυνδυασμός, ένας από τους πιο θεμελιώδεις μηχανισμούς στη φύση, είναι υπεύθυνος για τον μειωτικό επιχιασμό, την ενσωμάτωση του DNA σε γονιδιώματα ξενιστή και την επιδιόρθωση του DNA ως απόκριση

στις δίκλωνες θραυσεις του DNA (DSBs). Παρόλο που ο λεπτομερής μηχανισμός παραμένει ακόμα εν μέρει ασαφής, είναι πιθανώς κοινός σε όλα τα κύτταρα. Ο ομόλογος ανασυνδυασμός είναι ένας τύπος γονιδιωματικής αναδιάταξης που συμβαίνει μέσω της θραύσης και επανασύνδεσης αλυσίδων DNA, σε περιοχές με ταυτόσημες ή πολύ παρόμοιες αλληλουχίες. Αυτός είναι ένας κοινός μηχανισμός που επιτρέπει την αυτο-επιδιόρθωση του DNA που έχει υποστεί βλάβη και στις δύο έλικες της διπλής έλικας και μπορεί να καθορίσει βλάβες που συμβαίνουν σχεδόν σε κάθε γύρο αντιγραφής του DNA. Είναι επίσης απαραίτητος για τον σωστό και ακριβή διαχωρισμό των χρωμοσωμάτων κατά τη διάρκεια της μείωσης.

Ο μη ομόλογος ανασυνδυασμός (NHR) παίζει επίσης κεντρικό ρόλο στη βιολογία των γονιδιωμάτων των θηλαστικών. Για παράδειγμα, η ανάπτυξη μεταστατικών όγκων φαίνεται να περιλαμβάνει τη συσσώρευση μιας ευρείας ποικιλίας τροποποιημένων χρωμοσωμικών μοτίβων σχετιζόμενων με το NHR. Το NHR λαμβάνει χώρα σε αλληλουχίες του DNA όπου δεν υπάρχει ομοιότητα μεγάλης κλίμακας, π.χ. μετατοπίσεις μεταξύ διαφορετικών χρωμοσωμάτων ή διαγραφών που αφαιρούν αρκετά γονίδια κατά μήκος ενός χρωμοσώματος. Ωστόσο, η ομοιότητα μικρής έκτασης των αλληλουχιών βρίσκεται στα σημεία θραύσης όπου λαμβάνει χώρα η διαδικασία του μη ομόλογου ανασυνδυασμού και εξαρτάται επίσης από την παρουσία των IRs-HJs. Επιπλέον, ο NHR έχει βρεθεί ότι είναι υπεύθυνος για τον μειωτικό ανασυνδυασμό που έχει ως αποτέλεσμα χρωμοσωμικές διαταραχές με συνδρομικά χαρακτηριστικά. Μια εξίσωση που περιέχει IRs-HJs, CNVs, ανεστραμμένους διπλασιασμούς και ενεργά ρετρομεταθετά στοιχεία, όπως το LINE1P (153) θα μπορούσε να αποδώσει τη συχνότητα και την πλαστικότητα του ανασυνδυασμού.

Ο αντιγραφικός ανασυνδυασμός είναι ένας τύπος ανασυνδυασμού ο οποίος δημιουργεί ένα νέο αντίγραφο ενός τμήματος DNA. Πολλά TEs χρησιμοποιούν μια διαδικασία αντιγραφικού ανασυνδυασμού για τη δημιουργία ενός νέου αντιγράφου τους σε μια νέα θέση. Αυτά τα στοιχεία συνήθως διασκορπίζονται στο γονιδίωμα με ένα επαναλαμβανόμενο πρότυπο αντιγραφής και επικόλλησης δημιουργώντας νέες θέσεις ενσωμάτωσης, νέα εξώνια γονιδίων και εκτεταμένες γονιδιωματικές αλλοιώσεις που περιλαμβάνουν επίσης IRs-HJs. Τα εν λόγω TEs εκφράζονται στους γαμέτες και ειδικότερα εκφράζονται στα μη γονιμοποιημένα ανθρώπινα ωάρια εν αναμονή της γονιμοποίησης από

το απλοειδές άρρεν γονιδίωμα κατά την είσοδο του στο κυτταρόπλασμα του ωαρίου πιθανώς ως εκμαγεία θέσεων ανασυνδυασμού και τροποποίησης του γονιδιώματος των βλαστομεριδίων που θα προκύψουν (154).

Η εξέλιξη σε όλα τα γονιδιώματα είναι αποτέλεσμα της διαδικασίας του ανασυνδυασμού και επιτρέπει τη συσσώρευση πολλών μειωτικών και μιτωτικών αλλαγών με το χρόνο. Η σταδιακή συσσώρευση γενετικών μεταλλάξεων για μεγάλο χρονικό διάστημα σε αναπαραγωγικά ή σωματικά κύτταρα θα είχε ως αποτέλεσμα την εκτεταμένη αναδιάρθρωση της ακολουθίας νουκλεοτιδίων του γονιδιώματος με συνέπεια τον μωσαϊκισμό των κυττάρων που θα προκύψουν. Η πλαστικότητα του γονιδιώματος και η ταυτόχρονη ικανότητα να δημιουργεί νέα χαρακτηριστικά για το ίδιο μέσω του ανασυνδυασμού δεν θα υπήρχε και το εξελικτικό του δυναμικό θα ήταν περιορισμένο αν δεν υπήρχαν ευρέως διασκορπισμένα IRs-HJs που φαίνεται να μεταφέρονται παράλληλα με τη διαδικασία της ρετρομετάθεσης των TEs.

Επιπλέον, δημιουργήθηκε ένας γονιδιωματικός χάρτης κατανομής των HJs και βρέθηκε ότι αυτά εμφανίζονται σε μια μέση απόσταση 2,4 Mb. Και τα δύο ευρήματά δείχνουν ότι τα HJs δεν κατανέμονται ομοιόμορφα στα ανθρώπινα χρωμοσώματα, η συχνότητα ανασυνδυασμού δεν είναι η ίδια σε όλο το μήκος του γονιδιώματος και αυτό υποστηρίζεται με την ύπαρξη διαφόρων θερμών σημείων ανασυνδυασμού όπως προαναφέρθηκε (155). Στη συγκεκριμένη διατριβή προτείνεται ότι η ρετρομετάθεση των TEs μπορεί να διαμορφώσει τις διαφορετικές συχνότητες ανασυνδυασμού στο γονιδίωμα, μέσω της μετεγκατάστασης και διασποράς των IRs-HJs στα ανθρώπινα χρωμοσώματα. Αξίζει να σημειωθεί ότι η κατανομή των IRs-HJs στα TEs είναι σχεδόν ίδια σε όλα τα χρωμοσώματα με εξαίρεση το χρωμόσωμα Y, το οποίο είναι γνωστό ότι έχει εξελιχθεί αργότερα από τα άλλα αυτοσωμικά χρωμοσώματα και το χρωμόσωμα X (156). Παρόλο που τα χρωμοσώματα 1 και 19 είχαν τις υψηλότερες συχνότητες HJ ανά μεγαβάση, οι περιοχές γύρω από το κεντρομερίδιο του χρωμοσώματος 1 και ειδικότερα στους p και q βραχίονες του χρωμοσώματος 9 δεν έχουν IRs-HJs, γεγονός που παρατηρείται και στα ακροκεντρικά χρωμοσώματα. Με δεδομένο ότι η κατανομή των IRs-HJs που βρίσκονται σε TEs έχει μια μέση απόσταση μεταξύ των 2,4 Mbs άλλες μικρές κορυφές της γκαουσιανής κατανομής των αποστάσεων βρέθηκαν επίσης έως 1,2 Mbs και 3,2 Mbs, ενώ στο χρωμόσωμα Y οι

αποστάσεις μεταξύ των κατανεμημένων HJ βρίσκονται έως 6,4 Mbs. Αυτά τα ευρήματα σημαίνουν ότι, τα σχεδόν σταθερά διαστήματα μεταξύ των ευαίσθητων περιοχών σε ανασυνδυασμό που καλύπτουν το γονιδίωμα δημιουργούν συμβάντα ανασυνδυασμού κατά μήκος της μακράς έκτασης του ανθρώπινου γονιδιώματος με σχεδόν σταθερό ρυθμό ή καθορισμένες αποστάσεις.

Η παρουσία των IRs-HJs σε TEs κοντά σε γονίδια ήταν επίσης εξαιρετικής σημασίας και ειδικότερα γύρω από τη θέση έναρξης της μεταγραφής (TSS) καθώς και μεταξύ των γονιδίων. Η συντριπτική πλειονότητα των IRs-HJs σε TEs βρίσκεται σε απόσταση μικρότερη των 500 Kb από περιοχές γονιδίων, επιδεικνύοντας τη σημασία του αντιγραφικού ανασυνδυασμού στη διαμόρφωση του γονιδιακού περιεχομένου και των γονιδιακών συστοιχιών στη διάρκεια της εξέλιξης των πρωτεύοντων και του ανθρώπου.

Η σταθερότητα του γονιδιώματος από την άλλη, χαρακτηρίζεται από την εγγενή ικανότητα του να επιδιορθώνει τις βλάβες, να αποφεύγει την απώλεια της ετεροζυγωτίας και να επιτρέπει την αναδιάταξη χρωμοσωμάτων, χωρίς επιβάρυνσεις και δημιουργία επιβλαβών φαινοτύπων. Αυτές οι ισχυρές αλλά αμοιβαία εξαρτώμενες ιδιότητες του γονιδιώματος, είναι προφανείς σε όλες σχεδόν τις κυτταρικές διεργασίες από τη μείωση μέχρι τη μίτωση, σε γαμετικά και σωματικά κύτταρα. Ο μιτωτικός ανασυνδυασμός είναι ένα λιγότερο συχνό φαινόμενο ανά κυτταρική διαίρεση σε υγιή κύτταρα, αλλά είναι πιο συχνό φαινόμενο σε γενετικά αποσταθεροποιημένα καρκινικά κύτταρα. Αντίθετα, ο μειωτικός ανασυνδυασμός είναι πιο συχνός στα αναπαραγωγικά γεννητικά κύτταρα και είναι η βάση για τη δημιουργία και εμφάνιση νέων αναδιατάξεων σε επόμενες γενιές ανθρώπων. Και οι δύο διαδικασίες έχουν την ικανότητα να διαμορφώνουν το γονιδιακό περιεχόμενο ειδικά όταν εμπλέκονται TEs που λειτουργούν ως φορείς για τα IRs-HJs σε μια διαδικασία παρεμβατικής μεταλλαξιγένεσης.

Αυτή η δυναμική του γονιδιώματος να αναδιαμορφώνεται συνεχώς είναι ταυτόχρονα αποτέλεσμα της συνύπαρξης και της αλληλεπίδρασης των TEs με τα IRs-HJs. Οι ανεπιθύμητες συνέπειες των ανωμαλιών στη στενή συνεργασία μεταξύ TEs και HJs ή η αδυναμία των κυττάρων να ελέγχουν τις ανασυνδυαστικές δράσεις των ίδιων μπορεί να οδηγήσει σε καρκίνο, μη ισορροπημένες μετατοπίσεις, μωσαϊκισμό και άλλες ανεπιθύμητες καταστάσεις.

Παρόλο που συνεργάζονται στενά στην εξέλιξη των θηλαστικών και των ανθρώπων για να δημιουργήσουν σταθερότητα, ποικιλομορφία και πλαστικότητα, μπορεί μερικές φορές να αποσταθεροποιήσουν το γονιδίωμα, να εμφανίσουν φαινοτύπους με κλινικά σημαντικά χαρακτηριστικά και διαταραχές δυνητικά επηρεασμένες από μεθυλίωση ή διαταραχές που απορρυθμίζουν την μετακίνηση των μεταθετών στοιχείων.

6.3 ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΑΝΕΣΤΡΑΜΜΕΝΩΝ ΕΠΑΝΑΛΗΨΕΩΝ ΠΟΥ ΠΕΡΙΓΡΑΦΗΚΑΝ ΠΡΟΗΓΟΥΜΕΝΩΣ ΟΤΙ ΣΧΗΜΑΤΙΖΟΥΝ ΔΟΜΕΣ HOLLIDAY (IRS-HJs) ΩΣ ΠΙΘΑΝΑ HOTSPOTS ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΣΥΝΔΡΟΜΩΝ ΜΕ ΑΝΑΠΤΥΞΙΑΚΕΣ ΔΙΑΤΑΡΑΧΕΣ ΚΑΙ ΝΟΗΤΙΚΗ ΣΤΕΡΗΣΗ

Στην 3^η φάση της τρέχουσας μελέτης, καταγράφηκαν ποιες απο τις προαναφερθείσες ανεστραμμένες αλληλουχίες (IRS) που σχηματίζουν HJs σχετίζονται με σύνδρομα μικροδιπλασιασμών και μικροελλείψεων (MMS), και αναλύθηκε κατά πόσον η ακριβής θέση των IRS-HJs συνδέεται βάσει στατιστικής σημαντικότητας ή βάσει τυχαιότητας με τα εν λόγω σύνδρομα. Το κύριο εύρημα είναι η ύπαρξη ενός σημαντικά υψηλού αριθμού IRS-HJs σε περιοχές συγκεκριμένων χρωμοσωμάτων και γονιδιωματικών τόπων που εμφανίζονται τα MMS. Ειδικότερα οι αλληλουχίες # 1, # 2, # 3 IRS-HJs σχετίζονται με πολύ χαμηλή p-value και υψηλό z-score συγκριτικά με το σύνολο των αλληλουχιών IR-HJ που αναφέρονται παραπάνω. Τα συγκεκριμένα αποτελέσματά σε συνδυασμό με τα προηγούμενα ότι, τα TEs μπορούν να δρουν ως φορείς που διασπείρουν ευαίσθητες σε ανασυνδυασμό αλληλουχίες σε όλο το γονιδίωμα (157) δηλώνουν ότι η σχέση μεταξύ IRS-HJs-MMS και TEs δεν είναι τυχαία. Η ανάλυση στα σημεία θραύσης των συνδρόμων έδειξε την παρουσία των 3 κύριων οικογενειών TEs (ALUs, LINEs, HERVs) στους ανασυνδυασμένους γονιδιωματικούς τόπους, υποδεικνύοντας ότι ο ανασυνδυασμός μεταξύ αυτών των πλούσιων σε IRS-HJs μεταθετών στοιχείων (TEs) μπορεί να οδηγήσει στη δημιουργία συγκεκριμένων συνδρόμων.

Προηγούμενες μελέτες έχουν δείξει ότι συγκεκριμένα μεταθετά στοιχεία συμμετέχουν σε χρωμοσωμικές αναδιατάξεις που εντοπίζονται στην κυτταρογενετική σε μεγαλύτερου μεγέθους ανωμαλίες (>5 Mb) (158) επομένως τα TEs αναμένεται να διαδραματίζουν κάποιο ρόλο σε μικρές αλλά κλινικά σημαντικές αναδιατάξεις.

Για να κατανοήσουμε περαιτέρω τη σημασία των γονιδιωματικών τόπων που περιέχουν IRs-HJs, εκτελέστηκαν τρεις τύποι ανάλυσης.

Πρώτον, πραγματοποιήθηκε ανάλυση εμπλουτισμού με το ELDorado, ένα βιοπληροφορικό εργαλείο που παρέχει στατιστικά στοιχεία για τον εμπλουτισμό και τη συγκέντρωση των υπο μελέτη περιοχών σε ρυθμιστικές περιοχές του γονιδιώματος. Διαπιστώθηκε ότι οι αλληλουχίες IRs-HJs σε MMS είναι εξαιρετικά εμπλουτισμένες σε περιοχές που σχετίζονται με λειτουργικά γονίδια της εμβρυογένεσης και της μορφογένεσης. Εξετάζοντας τα δεδομένα εμπλουτισμού, μπορούμε επίσης να αποκτήσουμε γνώση για πιθανούς μηχανισμούς εμφάνισης των MMS. Ο εμπλουτισμός κατά 2,7x σε υποκινητές γονιδίων μπορεί ενδεχομένως να σημαίνει ότι οι διαγραφές σε τέτοιες περιοχές μπορούν να διαταράξουν τους ρυθμιστικούς μηχανισμούς γονιδίων, επιπρόσθετα ο εμπλουτισμός τους κατά 2,2x και 1,3x σε εξόνια και ιντρόνια αντίστοιχα, μπορεί ενδεχομένως να σημαίνει ότι στα σύνδρομα MMS η διαταραχή της γονιδιακής αλληλουχίας μέσω αντιγραφικού ανασυνδυασμού επηρεάζει την γονιδιακή λειτουργία των εμβρυολογικά σημαντικών γονιδίων που περιέχονται στα MMS.

Δεύτερον, πραγματοποιήθηκε ανάλυση για την κατανόηση της σχέσης των υπό εξέταση περιοχών με ρυθμιστικές περιοχές του DNA χρησιμοποιώντας εργαλεία βιοπληροφορικής στον περιηγητή γονιδιώματος UCSC. Βρέθηκε ότι συνδέονται κυρίως με τον παράγοντα POLR2A που είναι η θέση δέσμευσης πολυμεράσης της μεγαλύτερης υπομονάδας της RNA πολυμεράσης II, που είναι υπεύθυνη για τη σύνθεση του αγγελιαφόρου RNA σε ευκαρυωτικά κύτταρα. Η έναρξη της μεταγραφής με RNA πολυμεράση II απαιτεί τις δράσεις περισσότερων από 70 πολυπεπτιδίων. Η πρωτεΐνη που συντονίζει αυτές τις δραστηριότητες είναι ο βασικός μεταγραφικός παράγοντας TFIID, ο οποίος αποτελείται από την πρωτεΐνη TBP και από μια ομάδα εξελικτικά συντηρημένων πρωτεϊνών TAFs, που σχετίζονται επίσης με περιοχές IRs-HJs-MMS όπως φαίνεται από το σχήμα 5.25.

Αξιοσημείωτη είναι επίσης η σχέση των προαναφερθέντων περιοχών με τον παράγοντα μεταγραφής CTCF, υπεύθυνο για τη ρύθμιση της τρισδιάστατης δομής της χρωματίνης, με αποτέλεσμα να επηρεάζεται στη συνέχεια η ρύθμιση των γονιδίων. Αυτά τα ευρήματα ενισχύουν την παραδοχή ότι η σχέση μεταξύ των IRs-HJs-MMS-TEs και των μεταγραφικών παραγόντων (>20) επηρεάζει τη συνολική ανασυνδυαστική ικανότητα του ανθρώπινου γονιδιώματος.

Τρίτον, πραγματοποιήθηκε ειδική ανάλυση για την ύπαρξη και κατηγοριοποίηση γονιδίων στις κοινές περιοχές που εμφανίζονται τα σύνδρομα και οι IRs-HJs χρησιμοποιώντας το εργαλείο GREAT. Ένα σημαντικό εύρημα της ανάλυσης γονιδιακής εγγύτητας αποκάλυψε ότι όλες οι υπο μελέτη περιοχές IRs-HJs-MMS, είναι κοντά σε γονίδια ή εντός γονιδίων που ελέγχουν αναπτυξιακές διεργασίες (5.26). Αυτό έρχεται σε συμφωνία με το γεγονός ότι τα MMS είναι αλληλένδετα με φαινότυπους που εκδηλώνουν αναπτυξιακές διαταραχές στη διάρκεια της κύησης και αργότερα. Έχουμε λόγους να πιστεύουμε ότι τα MMS-IR-HJs σε περιοχές γονιδίων που ελέγχουν αναπτυξιακές διεργασίες προσφέρουν ένα υπόστρωμα για ανασυνδυασμό και μπορεί να οδηγήσουν σε άνισους διασκελισμούς και CNVs εντός των σημείων θραύσεων των υπό μελέτη συνδρόμων. Οι προκύπτουσες μικροελλείψεις και μικροδιπλασιασμοί μπορεί να επηρεάσουν τη γονιδιακή λειτουργικότητα, είτε διακόπτοντας απευθείας τις αλληλουχίες γονιδίων ή τις γειτονικές ρυθμιστικές περιοχές γονιδίων οδηγώντας έτσι σε αναπτυξιακές διαταραχές. Ένα νέο εύρημα που εξάγεται από αυτή τη μελέτη είναι ότι οι σχετιζόμενες με τα MMS ειδικές αλληλουχίες IRs-HJs κοντά ή μέσα στα γονίδια που ελέγχουν αναπτυξιακές διεργασίες μπορεί να συμμετέχουν σε έναν πιθανό μηχανισμό που να εξηγεί τον υποκείμενο φαινότυπο που παρατηρείται σε κάθε ένα από αυτά τα σύνδρομα. Πιο σημαντικά, μπορούν να βοηθήσουν στην ταυτοποίηση των ορίων που περιέχουν γονίδια που προκαλούν νοσήματα των οποίων η αλληλουχία ή παρακείμενες ρυθμιστικές περιοχές έχουν αλλοιωθεί. Σε συνδυασμό με τις αναλύσεις εμπλουτισμού τόσο γονιδιωματικής όσο και γονιδιακής ανάλυσης, αποκαλύφθηκε ότι η δημιουργία των MMS μπορεί να συμβεί μέσω της αλλοίωσης ρυθμιστικών ή κωδικοποιητικών αλληλουχιών. Τα προσβεβλημένα γονίδια σε κάθε περίπτωση μπορεί να προκαλούν νοσήματα και να εξηγήσουν, άμεσα ή έμμεσα, τον υποκείμενο φαινότυπο.

Η παρούσα μελέτη υποκινήθηκε από την κλινική σημασία των MMS που πρόσφατα ήρθε στο φως με βάση την ανάλυση NGS και τον μοριακό καρυότυπο. Η σημασία αυτή επισημαίνεται επίσης από το γεγονός ότι η συχνότητα των μικροελλείψεων και μικροδιαπλασιασμών έχει υπολογιστεί πρόσφατα σε 1/992 και 1/850 γεννήσεις αντίστοιχα. Η βιοπληροφορική ανάλυση του NGS επιτρέπει την εξαγωγή πολλαπλών αποτελεσμάτων σε όλες τις ηλικίες εγκύων είτε στο επίπεδο της αλλαγής της νουκλεοτιδικής αλληλουχίας είτε στο επίπεδο των δομικών αλλαγών στα χρωμοσώματα (121). Κατά την έλευση της τεχνολογίας υψηλής απόδοσης στη γενετική διάγνωση και της εφαρμογής της σε προγενετικούς ελέγχους, γίνεται ευρέως γνωστό στην κλινική πράξη ότι τα συγκεκριμένα MMS δεν είναι σπάνια και μπορεί να έχουν αθροιστικά συχνότητες υψηλότερες από τις ανευπλοειδίες (148). Χαρακτηριστική είναι η περίπτωση του συνδρόμου μικροέλλειψης Di-George, το οποίο τώρα θεωρείται δεύτερο σε συχνότητα εμφάνισης μετά το συνδρόμο Down και πιο συχνό από τις ανευπλοειδίες 13 και 18 (159). Το φάσμα των MMS είναι πολύ μεγάλο και ενδεχομένως ακόμη και τα σπάνια MMS αθροιστικά έχουν μεγάλη επίδραση στην προγεννητική έρευνα και διάγνωση. Η κλινική σημασία των MMS αποδεικνύεται από το ευρύ φαινοτυπικό τους φάσμα είτε σε προγεννητικές μελέτες υπερηχογραφήματος σε νεογνά είτε σε άλλες μελέτες, υποδεικνύοντας την ανάγκη για πιο ισχυρή προσέγγιση στη μοριακή διάγνωση. Πιο συγκεκριμένα, το υπερηχογράφημα που εκτελείται ακόμη και από έμπειρους γυναικολόγους ενδέχεται να μην διαγιγνώσκει ποσοστό 30% ή υψηλότερο από τα MMS (158).

Η κατανόηση της εξέλιξης και της βιολογίας των MMS θα βοηθούσε στην ανάπτυξη και εφαρμογή διαγνωστικών προσεγγίσεων βασισμένων στη βιοπληροφορική ανάλυση των δεδομένων NGS και στην αντιμετώπιση των CNVs που εμπλέκονται στην παραγωγή MMS χρησιμοποιώντας αυστηρά κριτήρια. Οι πλατφόρμες NGS που χρησιμοποιούνται για να στοχεύσουν ταυτόχρονα τις χρωμοσωμικές ανευπλοειδίες και τα MMS, καθώς και σημειακές μεταλλάξεις περιέχουν έως και 50 γονίδια (158). Συνεπώς, οι εφαρμογές του NGS που στοχεύουν στην εντοπισμό των μεταλλάξεων, τον αριθμό χρωμοσωματικών αντιγράφων και των MMS μπορούν να εμπλουτιστούν με περισσότερα MMS είτε σε μη επεμβατικούς είτε σε επεμβατικούς ελέγχους. Από την άλλη πλευρά, τέτοιες εφαρμογές μπορεί επίσης να ωφελήσουν την έρευνα για τον καρκίνο καθώς η ίδια τεχνολογία και τα δυνητικά παρόμοια

μοριακά γεγονότα αποτελούν τη βάση πολλαπλών γενετικών εκτροπών και μωσαϊκισμού σε έμβρυα και σωματικά κύτταρα.

6.4 ΑΛΛΗΛΟΥΧΙΑΚΑ ΜΟΤΙΒΑ ΠΟΥ ΑΝΑΓΝΩΡΙΖΟΝΤΑΙ ΑΠΟ ΤΗΝ PRDM9 ΩΣ ΠΙΘΑΝΑ HOTSPOTS ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΔΙΚΛΩΝΩΝ ΘΡΑΥΣΕΩΝ (ΚΑΡΚΙΝΟΓΕΝΕΣΗ)

Στο τελευταίο κομμάτι της διατριβής μελετήθηκε το φαινόμενο της παρουσίας διαφορετικών αλληλουχιών του μοτίβου πρόσδεσης της PRDM9 σε σημεία δίκλωνων θραύσεων αλυσίδων του DNA στο ανθρώπινο γονιδίωμα. Πραγματοποιήθηκε μελέτη για την εύρεση των θέσεων πρόσδεσης της PRDM9 μέσα σε ευρύτερες αλληλουχίες γύρω από δίκλωνες θραύσεις του DNA σε προφίλ 99 ασθενών (151) που πάσχουν από 3 είδη καρκίνου καθώς και για την στατιστική σημασία των διαφορών παραλλαγών της θέσης πρόσδεσης του PRDM9 στην καρκινογένεση.

Η αρχική μελέτη για την συντηρημένη μορφή του PRDM9 ως δεκατριανουκλειοτίδιο αποτέλεσε την βάση του τελικού σταδίου της έρευνας, καθώς με συγκεκριμένο script βρέθηκαν όλες οι πιθανές παραλλαγές του. Στη συνέχεια η υπολογιστική μέθοδος, υπέδειξε τις ακριβείς συντεταγμένες των παραλλαγών με σημείο αναφοράς τις δίκλωνες θραύσεις που σχετίζονται με 3 είδη καρκίνου (μαστού, πνεύμονα, κεφαλής και τραχήλου). Τα τελικά αποτελέσματα που προέκυψαν σε μορφή γραφημάτων, δείχνουν μια σημαντική συσχέτιση συγκεκριμένων αλληλουχιών του PRDM9 με τους καρκίνους του πνεύμονα, κεφαλής και τραχήλου και του πνεύμονα.

Η γενετική του καρκίνου έχει επί πολλά έτη επικεντρωθεί σε μεταλλακτικά συμβάντα που έχουν την πρωταρχική τους επίδραση στο καρκινικό κύτταρο. Πρόσφατα αυτή η εστίαση έχει διευρυνθεί, με την απόδειξη της σημασίας των επιγενετικών συμβάντων και των κυτταρικών αλληλεπιδράσεων στην ανάπτυξη του καρκίνου. Ο ρόλος της κοινής γενετικής ποικιλίας στον προσδιορισμό του φάσματος της ατομικής ευαισθησίας εντός του πληθυσμού αναγνωρίζεται όλο και περισσότερο και θα αντιμετωπιστεί χρησιμοποιώντας πληροφορίες από το Πρόγραμμα Ανθρώπινου Γονιδιώματος (Human Genome Project). Αυτές οι νέες κατευθύνσεις έρευνας θα αναδείξουν τους καθοριστικούς παράγοντες της γενετικής

αστάθειας στη καρκινογένεση που θα μπορούν να εντοπιστούν πέρα από τη φυσιολογία του καρκινικού κυττάρου, θα προτείνουν νέους στόχους παρέμβασης και στρατηγικής πρόληψης σε ομάδες με αυξημένο κίνδυνο.

Στην παρούσα έρευνα αποδείχθηκε η συσχέτιση των διάφορων παραλλαγών του PRDM9 στην εμφάνιση ορισμένων ειδών καρκίνου, και υποδεικνύει κατα πόσο αλληλουχιακά μοτίβα μπορούν να χρησιμεύσουν σαν βιοδείκτες για την μερική ή πλήρη πρόγνωση συγκεκριμένων περιπτώσεων καρκίνου.

ΣΥΜΠΕΡΑΣΜΑΤΑ

- Τα χρωμοσώματα της πατρικής σειράς έχουν μικρότερους ρυθμούς ανασυνδυασμού από ότι τα αντίστοιχα της μητρικής σειράς επειδή οι γενετικοί δείκτες ανασυνδυασμού είναι πιο κοντά μεταξύ τους από ότι στα θηλυκά χρωμοσώματα, με αποτέλεσμα να μην εμφανίζονται συχνά συμβάντα ανασυνδυασμού.
- Τα hotspots στα χρωμοσώματα της πατρικής σειράς εντοπίζονται πολύ κοντά το ένα με το άλλο σε συστάδες, ενώ τα hotspots στα χρωμοσώματα της μητρικής σειράς είναι κατανεμημένα ομοιόμορφα στο γονιδίωμα ώστε όλα να έχουν σχεδόν τις ίδιες πιθανότητες για την ύπαρξη ανασυνδυασμού μεταξύ τους.
- Τα ρετροτρανσποζόνια λειτουργούν ως φορείς που διασπείρουν τις αλληλουχίες (IRs-HJs) με ανασυνδυαστική δράση μέσα στο γονιδίωμα ενισχύοντας την πλαστικότητα του. Κάθε αλληλουχιακή μορφή των IRs-HJs εμφανίζει ειδικότητα και συγγένεια με συγκεκριμένες οικογένειες των TEs, παρέχοντας νέα δεδομένα για την ταυτόχρονη εξέλιξη των IRs-HJs και TEs
- Ο μηχανισμός του αντιγραφικού ανασυνδυασμού των μεταθετών στοιχείων πιθανώς να οδηγεί στην αύξηση του αριθμού των αλληλουχιών που σχηματίζουν HJs, συμβάλλοντας στον ανασυνδυασμό, στην πλαστικότητα του γενόματος και στην επιδιόρθωση του DNA μέσω ομόλογου ανασυνδυασμού.
- Η δημιουργία ενός ολοκληρωμένου γονιδιωματικού χάρτη των IRs-HJs που σχετίζονται με TEs, συμβάλλει στην κατανόηση της σημασίας της διαφορικής κατανομής των IRs-ανθρώπινα χρωμοσώματα.
- Η συντριπτική πλειονότητα των IRs-HJs που σχετίζονται με συγκεκριμένα TEs βρίσκεται σε απόσταση μικρότερη των 500 Kb από θέσεις έναρξης μεταγραφής γονιδίων (TSS), αναδεικνύοντας τη σημασία του αντιγραφικού ανασυνδυασμού στη διαμόρφωση του γονιδιακού περιεχομένου.
- Κυρίως 3 αλληλουχιακά μοτίβα ανεστραμμένων επαναλήψεων που περιγραφηκαν προηγουμενωσ οτι σχηματιζουν δομες Holliday (IRs-HJs) χαρακτηριζονται ως hotspots

με αυξημένη πιθανότητα για τη δημιουργία συνδρόμων με αναπτυξιακές διαταραχές και νοητική στέρηση

- Οι αλληλουχίες IRs-HJs εντοπίζονται σε συστάδες γονιδίων ελέγχουν αναπτυξιακές διαδικασίες της εμβρυογένεσης και μορφογένεσης. Από τα δεδομένα γονιδιακού εμπλουτισμού ενδέχεται να αποκρυπτογραφηθούν και να προβλεφθούν οι πιθανοί μηχανισμοί εμφάνισης συνδρόμων μικροελλείψεων και μικροδιπλασιασμών.
- Η συσχέτιση των IRs-HJ με μεταγραφικούς παράγοντες και κυρίως με τον POLR2A, την πρωτεΐνη δέσμευσης TAB (TBP), τον TAFs, και τον CTCF, ενισχύει την υπόθεση ότι η σχέση μεταξύ των TFs (μεταγραφικών παραγόντων)-IRs-HJs-MMS-TEs επηρεάζει τη συνολική ανασυνδυαστική ικανότητα του ανθρώπινου γονιδιώματος.
- Το αλληλουχιακό μοτίβο d(CCNCCNTNNCCNC) που προσδένεται η πρωτεΐνη PRDM9 σε περιοχές δίκλωνων θραυσεων που οδηγούν στη δημιουργία 3 τύπων καρκίνου, ενισχύει την υπόθεση ότι μερικές μορφές του PRDM9 σχετίζονται με τον ισοζυγισμένο ανασυνδυασμό και συγκεκριμένες με τον παθογόνο και μη ισοζυγισμένο ανασυνδυασμό.
- Η ύπαρξη κοινής αλληλουχίας πρόσδεσης της PRDM9 με τη μεγαλύτερη στατιστική σημαντικότητα [d(CCACCATCACAC)] και στους 3 τύπους καρκίνου μελετήθηκαν δηλώνει την εμπλοκή της ίδιας στη διαδικασία της καρκινογένεσης.

ΠΕΡΙΛΗΨΗ

Η συγκεκριμένη ερευνητική εργασία διεκπεραιώθηκε σε 4 στάδια. Στο πρώτο στάδιο χρησιμοποιήθηκαν βιοπληροφορικά εργαλεία για την ανάλυση δεδομένων ανασυνδυασμού από γενετικούς χάρτες, προκειμένου να εκτιμηθεί η μέση απόσταση μεταξύ ενός hotspot ανασυνδυασμού και των επόμενων, κατά μήκος του γονιδιώματος. Ο πρωταρχικός σκοπός αυτής της έρευνας ήταν η επικύρωση της θεωρίας του ανασυνδυασμού του T. H Morgan. Συνδυάζοντας σύνολα δεδομένων από γεγονότα ανασυνδυασμού που έχουν συγκεντρωθεί από μειώσεις κατασκευάστηκαν γενετικοί χάρτες ειδικοί για το κάθε φύλο. Αποδείχθηκε ότι υπάρχει διαφορετική εκδήλωση της συχνότητας του ανασυνδυασμού στα 2 φύλα. Πιο συγκεκριμένα, η μεγάλη πλειοψηφία των hotspots στους άρρενες συγκεντρώνεται σε μικρές γενετικές αποστάσεις, αντιθέτως με τα hotspots των θήλεων τα οποία κατανέμονται ομοιόμορφα κατά μήκος των χρωμοσωμάτων.

Στο δεύτερο στάδιο της έρευνας χαρακτηρίστηκαν ενδιάμεσες δομές ανασυνδυασμού, οι δομές Holliday οι οποίες προέρχονται από αλληλουχιακά μοτίβα ανεστραμμένων επαναλήψεων. Οι ανεστραμμένες επαναλήψεις είναι παρούσες σε αφθονία τόσο σε προκαρυωτικά όσο και σε ευκαρυωτικά γονιδιώματα και μπορούν να σχηματίσουν δευτερογενείς δομές του DNA που εμπλέκονται σε πολλές σημαντικές βιολογικές διεργασίες όπως η αντιγραφή και η μεθυλίωση του DNA. Σε αυτό το στάδιο, διερευνήθηκε κατά πόσο οι αλληλουχίες των IRs-HJs συνδέονται με μεταθετά στοιχεία (TEs) και κυρίως με ALUs, LINEs, SVAs και HERVs. Έξι διαφορετικές μορφές του αλληλουχιακού μοτίβου των IRs-HJs ταυτοποιήθηκαν και εντοπίστηκαν οι γονιδιωματικές συντεταγμένες αλληλουχιών που περιέχουν τόσο IRs-HJs όσο και TEs. Από τα 2982 συνολικά HJs, βρέθηκε ένας σημαντικός αριθμός 1319 HJs που σχετίζονται με TEs, με μέση κατανομή 1 HJ ανά 2,4 Mb. Ένα υψηλό ποσοστό IRs-HJs συσχετίστηκε με όλες τις κύριες οικογένειες TEs, με ειδικότητα για συγκεκριμένα ενεργά ή ανενεργά μεταθετά στοιχεία: τα DNA μεταθετά και τα μεταγενέστερα στοιχεία των οικογενειών των ALUs, LINEs και HERVs έως 41,94%, 72,72%, 42,94% και 84,5% αντίστοιχα. Η φυλογενετική ανάλυση αποκάλυψε ότι τα HJs εμφανίζονται τόσο σε ενεργά όσο και σε αδρανή TEs. Επιπλέον, οι σχετιζόμενες με TEs IRs-HJs αλληλουχίες εντοπίστηκαν σχεδόν αποκλειστικά σε απόσταση μικρότερη από 1 Mb από τα ανθρώπινα γονίδια, ενώ μόνο

23 δεν συσχετίστηκαν με κανένα γονίδιο. Αυτή είναι η πρώτη αναφορά που συνδέει τα ανθρώπινα HJs με μεταθετά στοιχεία. Από τα παραπάνω δεδομένα προκύπτει ότι συγκεκριμένες αλληλουχιακές μορφές των HJs δείχνουν προτίμηση για συγκεκριμένα ενεργά στοιχεία των οικογενειών των ALUs και LINEs, αποδεικνύοντας ότι τα ενσωματωμένα σε ρετρομεταθετά IRs-HJs μπορούν να μεταφερθούν στο γονιδίωμα μέσω της ρετρομετάθεσης, συμβάλλοντας στον ανασυνδυασμό, στην πλαστικότητα του γονιδιώματος και στην επιδιόρθωση του DNA.

Στο 3^ο στάδιο της έρευνας χαρακτηρίστηκαν αλληλουχίες με ανασυνδυαστική δράση που ενοχοποιούνται για την εμφάνιση γονιδιωματικής αστάθειας ειδικότερα μέσω των συνδρόμων μικροελλείψεων και μικροδιπλασιασμών (MMS). Τα MMS προέρχονται από σπάνια γεγονότα ανασυνδυασμού, αλλά αθροιστικά αντιπροσωπεύουν μια σημαντική ομάδα γενετικών ανωμαλιών σχεδόν ισοδύναμων με ανευπλοειδίες. Από τις προσωπικές βάσεις δεδομένων των MMS διαπιστώθηκε ότι σχεδόν όλα τα MMS έχουν μήκος μικρότερο από 5 Mbs. Επιπλέον, ένα σημαντικό ποσοστό αυτών έχει μήκος μικρότερο από 0.5Mbs και όλα μαζί εμφανίζουν μεγάλη πιθανότητα να φιλοξενούν TEs και HJs-IR. Πιο σημαντικά, η βιοπληροφορική ανάλυση εγγύτητας γονιδίων αποκάλυψε ότι ένας σημαντικά υψηλός αριθμός IRs-HJs που σχετίζονται με MMS εντοπίζονται σε περιοχές γονιδίων που ελέγχουν βασικές αναπτυξιακές διεργασίες της εμβρυογένεσης και της μορφογένεσης οργάνων. Συνολικά, προτείνεται ότι τα IRs-HJs και άλλες μονάδες ανασυνδυασμού, όπως τα TEs που εντοπίζονται σε περιοχές των MMS, ίσως δηλώνουν έναν πιθανό μηχανισμό για την εμφάνισή των συγκεκριμένων συνδρόμων κατά τη διάρκεια της γαμετογένεσης.

Στο τελικό στάδιο της έρευνας εντοπίστηκαν παραλλαγές του μοτίβου αλληλουχίας PRDM9 (CCNCCNTNNCCNC) οι οποίες ενοχοποιούνται για τη εμφάνιση καρκίνου του μαστού (BRC), κεφαλής και τραχήλου (HN) καθώς και για το αδενοκαρκίνωμα του πνεύμονα LUAD). Τα τελικά αποτελέσματα έδειξαν την ύπαρξη συγκεκριμένων αλληλουχιών PRDM9 που βρίσκονται εντός των ορίων των δίκλωνων θραύσεων που είναι υπεύθυνες για τη δημιουργία ειδικών τύπων καρκίνου αφού αποδείχθηκε με βιοπληροφορική ανάλυση η στατιστική σημασία της παρουσίας κάθε αλληλουχίας PRDM9 εντός των BRC, HN και LUAD . Για πρώτη φορά, αποδεικνύεται άμεση συσχέτιση μεταξύ του αλληλουχιακού μοτίβου

d(CCACCATCACCAC) και της εμφάνισης του καρκίνου του μαστού (BRC), του αδενοκαρκινώματος του πνεύμονα (LUAD) και του καρκίνου κεφαλής και τραχήλου(HN).

PARIS P. LADIAS

PhD, THESIS

A bioinformatics approach for the characterization of recombination structures and how they are affected by transposable element sequences in Genetic Instability

ABSTRACT

In the beginning of the 20th century, the experiments conducted by Thomas Hunt Morgan set the foundation for the creation of a new field in biology, namely the field of genetics as we know it today. Correlating the process of transferring a chromosome with the process of transferring characters from parental generations to their progeny, led to the discovery of many important phenomena that occur during the stages of meiosis, the most important of which, is possibly the existence of recombination events between homologous pairs of chromosomes. In this paper, bioinformatic tools and recombination datasets from genetic maps were utilized, in order to estimate the average distance between a recombination hotspot and its subsequent, along the genome. The first purpose of this research, was the validation of T. H Morgan's theory of recombination, as it was developed in 1911, with contemporary methods that were unavailable at the time. By combining data sets, recombination events have collected from meioses and sex-specific genetic maps have constructed. It was shown that, a substantial fraction of the genome shows some degree of sexually-specific varied recombination frequency. More specifically the vast majority of male hotspots are clustered in small genetic distances as long as the female hotspots are distributed uniformly along the chromosomes.

Recombination is occurred in Holliday junctions (HJs) that constitute important intermediate structures for many cell functions such as DNA recombination and DNA repair. The second stage of the research was conducted in order to characterize Holliday junctions that derive from a 10-nt degenerate inverted repeat sequence, with a 3-nt core motif. Inverted repeats are present in abundance in both prokaryotic and eukaryotic genomes and can form DNA secondary structures – hairpins and cruciforms that are involved in many important biological processes such as DNA replication, DNA transition and DNA methylation. In this study, the human genome was explored to find whether the IRs-HJ degenerate sequence associates with transposable elements (TEs) and mainly with those of the active and inactive ALU, LINE, SVA

and HERV families. Six different forms of the IRs-HJs sequence motif were identified, and the genomic coordinates of sequences containing both IRs-HJs and TEs were located. From 2982 total HJs, a significant number of 1319 TE-associated HJs were found, with a median distribution of 1 per 2.4 Mb. The HJs with higher GC content were observed more frequently at the genome. A high percentage of HJs were associated with all main TE families, with specificity for particular active or inactive elements: DNA elements and the retroelements ALUs, LINEs and HERVs up to 41.94%, 72.72%, 42.94% and 84.5%, respectively. Phylogenetic analysis revealed that HJs occur in both active and inactive TEs. Furthermore, the TE-associated HJs were almost exclusively found within a distance less than 1 Mb from human genes, while only 23 were not associated with any genes. This is the first report associating human HJs, with mobile elements. These data pinpoint that particular HJ forms show preference for specific active retrotransposon families of ALUs and LINEs, suggesting that retrotransposon-incorporated HJs may relocate or replicate in the genome through retrotransposition, contributing to recombination, genome plasticity and DNA repair.

The third stage of this research was conducted in order to characterize sequences that promote genomic instability through particular Microdeletion and Microduplication Syndromes (MMS). MMS are rare recombinations, but cumulatively they represent a significant group of genetic abnormalities almost equal to aneuploidies. Some of them and in particular Di George Microdeletion can be found in frequencies second only to trisomy 21. Based on our database of MMS we found that almost all MMS span in length less than 5Mbs which is the cut-off point of resolution in standard cytogenetics. Furthermore, a significant percentage of them are in length less than 0.5Mbs and all together, short and long MMS, have a high probability to harbor TE-associated and non-TE associated HJs-IRs. More importantly, gene proximity bioinformatics analysis revealed that a significantly high number of MMS associated IRs are proximal to regions of genes that control key developmental processes and transcription factors in humans. Collectively, it is proposed that IRs and other recombination primers such as TEs that exist in regions spanning MMS hotspots underlie a possible mechanism for their occurrence during gametogenesis.

The final stage of the research was conducted in order to discover variations of PRDM9 sequence motif (CCNCCNTNNCCNC) which are responsible for the generation of breast invasive carcinoma (BRC), head and neck squamous cell carcinoma (HN) as well as for lung

adenocarcinoma (LUAD). The final results showed a preference for specific PRDM9 sequences that reside within the boundaries of Double Strand Breaks and are responsible for the generation of specific types of cancer after proving by bioinformatics analysis the statistical significance of each PRDM9 sequence variation inside the BRC, HN and LUAD. For the first time, an immediate correlation between the PRDM9 sequence motif d(CCACCATCACCAC) and the emergence of breast invasive carcinoma (BRC), lung adenocarcinoma (LUAD) and head and neck squamous cell carcinoma (HN) is proved.

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

HR, Homologous Recombination

NHRM, Non Homologous Recombination

NAHR, Non Allelic Homologous Recombination,

RBM, Replicative Recombination

cM, CentiMorgan

LD, Linkage Disequilibrium

HJs, Holliday Junctions

IRs, Inverted Repeats

IRs-HJs, Inverted Repeats that form Holliday Junctions

SSB, Single Strand Break

DSBs, Double Strand Breaks

CFS, Common Fragile Sites

MDS, Multiple Damage Sites

TEs, Transposable Elements

LINEs, Long Interspersed Nuclear Elements

LTRs, Long Terminal Repeats

NHEJ, non homologous end joining

non LTRs, Non Long Terminal Repeats

ORF, open reading frame

RT, reverse Transcriptase

SINEs, Short Interspersed Nuclear Elements

TSD, Target Site Duplications

VNTR, Variable Number Tandem Repeats

CNVs, Copy Number Variations

LCR, Low Copy Repeats

HCR, High Copy Repeats

CNPs, Copy Number Polymorphisms

MMS, Microdeletion and Microduplication Syndromes

TSS, Transcription Start Sites

TFs, Transcription Factors

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Grelon M. Meiotic recombination mechanisms. *Comptes Rendus Biologies*. 2016;339(7):247-51.
2. Bellen HJ, Yamamoto S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell*. 2015;163(1):12-4.
3. <https://www.khanacademy.org/science/biology/classical-genetics/chromosomal-basis-of-genetics/a/linkage-mapping>.
4. Koszul R, Meselson M, Van Doninck K, Vandenhautte J, Zickler D. The centenary of Janssens's chiasmatype theory. *Genetics*. 2012;191(2):309-17.
5. Ravindran S, Barbara McClintock and the discovery of jumping genes. *Proceedings of the National Academy of Sciences*. 2012;109(50):20198.
6. Hampshire DJ, Ayub M, Springell K, Roberts E, Jafri H, Rashid Y, et al. MORM syndrome (mental retardation, truncal obesity, retinal dystrophy and micropenis), a new autosomal recessive disorder, links to 9q34. *European Journal of Human Genetics*. 2006;14(5):543-8.
7. Creighton HB, McClintock B. A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. *Proceedings of the National Academy of Sciences of the United States of America*. 1931;17(8):492-7.
8. Coe E, Kass LB. Proof of physical exchange of genes on the chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(19):6641.
9. Chowdhury R, Bois PR, Feingold E, Sherman SL, Cheung VG. Genetic analysis of variation in human meiotic recombination. *PLoS genetics*. 2009;5(9):e1000648.
10. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature genetics*. 2005;37(5):549-54.
11. Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, et al. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nature genetics*. 2010;42(10):827-9.
12. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature genetics*. 1999;22(2):139-44.
13. Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*. 1964;49(1):49-67.
14. Murray SS, Oliphant A, Shen R, McBride C, Steeke RJ, Shannon SG, et al. A highly informative SNP linkage panel for human genetic studies. *Nature methods*. 2004;1(2):113-7.
15. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik*. 1968;38(6):226-31.
16. Jorgenson E, Witte JS. A gene-centric approach to genome-wide association studies. *Nature reviews Genetics*. 2006;7(11):885-91.

17. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nature genetics*. 2005;37(11):1217-23.
18. Li W, Freudenberg J. Two-parameter characterization of chromosome-scale recombination rate. *Genome Res*. 2009;19(12):2300-7.
19. Morton NE. Parameters of the human genome. *Proceedings of the National Academy of Sciences of the United States of America*. 1991;88(17):7474-6.
20. Casares P. A corrected Haldane's map function to calculate genetic distances from recombination data. *Genetica*. 2007;129(3):333-8.
21. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. Genetic recombination is directed away from functional genomic elements in mice. *Nature*. 2012;485(7400):642-5.
22. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Reviews Molecular Cell Biology*. 2006;7(10):739-50.
23. Liu J, Renault L, Veaute X, Fabre F, Stahlberg H, Heyer W-D. Rad51 paralogues Rad55-Rad57 balance the antirecombinase Srs2 in Rad51 filament formation. *Nature*. 2011;479(7372):245-8.
24. Brick K, Pratto F, Sun C-Y, Camerini-Otero RD, Petukhova G. Chapter Sixteen - Analysis of Meiotic Double-Strand Break Initiation in Mammals. In: Spies M, Malkova A, editors. *Methods in Enzymology*. 601: Academic Press; 2018. p. 391-418.
25. Godin SK, Sullivan MR, Bernstein KA. Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. 2016;94(5):407-18.
26. Hong S, Sung Y, Yu M, Lee M, Kleckner N, Kim KP. The logic and mechanism of homologous recombination partner choice. *Mol Cell*. 2013;51(4):440-53.
27. Dluzewska J, Szymanska M, Ziolkowski PA. Where to Cross Over? Defining Crossover Sites in Plants. 2018;9(609).
28. Cromie GA, Hyppa RW, Taylor AF, Zakharyevich K, Hunter N, Smith GR. Single Holliday junctions are intermediates of meiotic recombination. *Cell*. 2006;127(6):1167-78.
29. Meselson MS, Radding CM. A general model for genetic recombination. *Proceedings of the National Academy of Sciences of the United States of America*. 1975;72(1):358-61.
30. Subramaniam S, Tewari AK, Nunes-Duby SE, Foster MP. Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. *Journal of molecular biology*. 2003;329(3):423-39.
31. Déclais A-C, Fogg JM, Freeman ADJ, Coste F, Hadden JM, Phillips SEV, et al. The complex between a four-way DNA junction and T7 endonuclease I. *EMBO J*. 2003;22(6):1398-409.
32. Karow JK, Constantinou A, Li JL, West SC, Hickson ID. The Bloom's syndrome gene product promotes branch migration of holliday junctions. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(12):6504-8.
33. Abaji C, Cousineau I, Belmaaza A. BRCA2 Regulates Homologous Recombination in Response to DNA Damage: Implications for Genome Stability and Carcinogenesis. *Cancer Research*. 2005;65(10):4117.

34. Eichman BF, Ortiz-Lombardia M, Aymami J, Coll M, Ho PS. The inherent properties of DNA four-way junctions: comparing the crystal structures of holliday junctions. *Journal of molecular biology*. 2002;320(5):1037-51.
35. Eichman BF, Vargason JM, Mooers BHM, Ho PS. The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions. *Proceedings of the National Academy of Sciences*. 2000;97(8):3971.
36. Chaffey N, Alberts B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. *Molecular biology of the cell*. 4th edn. Ann Bot. 2003;91(3):401-.
37. Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N. Double Holliday junctions are intermediates of DNA break repair. *Nature*. 2010;464(7290):937-41.
38. Kassem HS, Girolami F, Sanoudou D. Molecular genetics made simple. *Glob Cardiol Sci Pract*. 2012;2012(1):6-.
39. Marsolier-Kergoat MC, Khan MM, Schott J, Zhu X, Llorente B. Mechanistic View and Genetic Control of DNA Recombination during Meiosis. *Mol Cell*. 2018;70(1):9-20.e6.
40. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell research*. 2008;18(1):134-47.
41. Fleck O, Nielsen O. DNA repair. *Journal of Cell Science*. 2004;117(4):515.
42. Shah Punatar R, Martin MJ, Wyatt HDM, Chan YW, West SC. Resolution of single and double Holliday junction recombination intermediates by GEN1. *Proceedings of the National Academy of Sciences*. 2017;114(3):443.
43. Cromie GA, Leach DR. Control of crossing over. *Mol Cell*. 2000;6(4):815-26.
44. Seeman NC, Kallenbach NR. Design of immobile nucleic acid junctions. *Biophysical journal*. 1983;44(2):201-9.
45. Duckett DR, Murchie AI, Diekmann S, von Kitzing E, Kemper B, Lilley DM. The structure of the Holliday junction, and its resolution. *Cell*. 1988;55(1):79-89.
46. Ho P, Eichman B. The crystal structures of DNA Holliday junctions. *Current opinion in structural biology*. 2001;11:302-8.
47. Khuu P, Voth A, Hays F, Ho P. The stacked-X DNA Holliday junction and protein recognition. *Journal of molecular recognition : JMR*. 2006;19:234-42.
48. Hyeon C, Lee J, Yoon J, Hohng S, Thirumalai D. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nature Chemistry*. 2012;4:907.
49. Hays F, Teegarden A, Jones Z, Harms M, Raup D, Watson J, et al. How Does Sequence Define Structure? a Crystallographic map of DNA structure and conformation. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:7157-62.
50. Petes TD. Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*. 2001;2(5):360-9.
51. Mieczkowski PA, Dominska M, Buck MJ, Gerton JL, Lieb JD, Petes TD. Global analysis of the relationship between the binding of the Bas1p transcription factor and meiosis-specific double-strand DNA breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2006;26(3):1014-27.
52. Li J, Zhang MQ, Zhang X. A new method for detecting human recombination hotspots and its applications to the HapMap ENCODE data. *Am J Hum Genet*. 2006;79(4):628-39.

53. De Iorio M, de Silva E, Stumpf MPH. Recombination hotspots as a point process. *Philos Trans R Soc Lond B Biol Sci.* 2005;360(1460):1597-603.
54. McVean G. What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc Lond B Biol Sci.* 2010;365(1544):1213-8.
55. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology.* 2017;18:407.
56. Paigen K, Petkov PM. PRDM9 and Its Role in Genetic Recombination. *Trends in genetics : TIG.* 2018;34(4):291-300.
57. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science (New York, NY).* 2010;327(5967):836-40.
58. Grey C, Baudat F, de Massy B. PRDM9, a driver of the genetic map. *PLoS genetics.* 2018;14(8):e1007479.
59. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature.* 1980;284(5757):601-3.
60. Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics.* 1982;101(3-4):519-31.
61. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics.* 2007;41(1):331-68.
62. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
63. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell.* 1993;72(4):595-605.
64. Whitcomb JM, Hughes SH. Retroviral reverse transcription and integration: progress and problems. *Annual review of cell biology.* 1992;8:275-306.
65. https://commons.wikimedia.org/wiki/File:DNA_Transposon.png.
66. Pace JK, 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 2007;17(4):422-32.
67. Wessler S. Plant Transposable Elements. A Hard Act to Follow. *Plant physiology.* 2001;125:149-51.
68. https://commons.wikimedia.org/wiki/File:LTR_retrotransposon.png.
69. Doh JH, Lutz S, Curcio MJ. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. *PLoS genetics.* 2014;10(3):e1004219-e.
70. de Parseval N, Heidmann T. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenetic and genome research.* 2005;110(1-4):318-32.
71. Deininger PL, Moran JV, Batzer MA, Kazazian HH, Jr. Mobile elements and mammalian genome evolution. *Current opinion in genetics & development.* 2003;13(6):651-8.
72. Pelisson A, Mejlumian L, Robert V, Terzian C, Bucheton A. Drosophila germline invasion by the endogenous retrovirus gypsy: Involvement of the viral env gene. *Insect biochemistry and molecular biology.* 2002;32:1249-56.

73. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 1990;9(10):3353-62.
74. Luning Prak ET, Kazazian Jr HH. Mobile elements and the human genome. *Nature Reviews Genetics.* 2000;1:134.
75. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH, Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet.* 2003;73(6):1444-51.
76. Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet.* 2001;35:501-38.
77. Rowold DJ, Herrera RJ. Alu elements and the human genome. *Genetica.* 2000;108(1):57-72.
78. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet.* 2011;12:187-215.
79. Lutz SM, Vincent BJ, Kazazian HH, Jr., Batzer MA, Moran JV. Allelic heterogeneity in LINE-1 retrotransposition activity. *Am J Hum Genet.* 2003;73(6):1431-7.
80. Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* 1996;87(5):905-16.
81. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol.* 2001;21(2):467-75.
82. Peddigari S, Li PW-L, Rabe JL, Martin SL. hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* 2013;41(1):575-85.
83. <https://en.wikipedia.org/wiki/LINE1>.
84. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nature reviews Genetics.* 2009;10(10):691-703.
85. Viollet S, Monot C, Cristofari G. L1 retrotransposition: The snap-velcro model and its consequences. *Mob Genet Elements.* 2014;4(1):e28907-e.
86. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nature reviews Genetics.* 2002;3(5):370-9.
87. Weiner AM, Deininger PL, Efstratiadis A. Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annual review of biochemistry.* 1986;55:631-61.
88. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays : news and reviews in molecular, cellular and developmental biology.* 2000;22(2):148-60.
89. Weiner AM. SINEs and LINEs: the art of biting the hand that feeds you. *Current opinion in cell biology.* 2002;14(3):343-50.
90. Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, et al. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* 2003;13(9):1998-2004.
91. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, et al. SVA elements: a hominid-specific retroposon family. *Journal of molecular biology.* 2005;354(4):994-1007.

92. Savage AL, Schumann GG, Breen G, Bubb VJ, Al-Chalabi A, Quinn JP. Retrotransposons in the development and progression of amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*. 2019;90(3):284.
93. <http://www.repeatmasker.org/>.
94. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-54.
95. Nowakowska B. Clinical interpretation of copy number variants in the human genome. *Journal of applied genetics*. 2017;58(4):449-57.
96. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. *Journal of human genetics*. 2012;57(1):6-13.
97. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451-81.
98. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Human molecular genetics*. 2009;18(R1):R1-8.
99. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet*. 2009;84(2):148-61.
100. Lindsay SJ, Khajavi M, Lupski JR, Hurles ME. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am J Hum Genet*. 2006;79(5):890-902.
101. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *PathoGenetics*. 2008;1(1):4.
102. Liu P, Lacaria M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am J Hum Genet*. 2011;89(4):580-8.
103. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nature reviews Genetics*. 2016;17(4):224-38.
104. Watson CT, Marques-Bonet T, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: an update. *Annu Rev Genomics Hum Genet*. 2014;15:215-44.
105. Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its roles in human copy number variation. *Genomics Inform*. 2014;12(4):136-44.
106. Lee JA, Lupski JR. Genomic Rearrangements and Gene Copy-Number Alterations as a Cause of Nervous System Disorders. *Neuron*. 2006;52(1):103-21.
107. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524-33.
108. Borun P, De Rosa M, Nedoszytko B, Walkowiak J, Plawski A. Specific Alu elements involved in a significant percentage of copy number variations of the STK11 gene in patients with Peutz-Jeghers syndrome. *Familial cancer*. 2015;14(3):455-61.
109. Gu S, Yuan B, Campbell IM, Beck CR, Carvalho CM, Nagamani SC, et al. Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Human molecular genetics*. 2015;24(14):4061-77.

110. Boone PM, Yuan B, Campbell IM, Scull JC, Withers MA, Baggett BC, et al. The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am J Hum Genet.* 2014;95(2):143-61.
111. Campbell IM, Gambin T, Dittwald P, Beck CR, Shuvarikov A, Hixson P, et al. Human endogenous retroviral elements promote genome instability via non-allelic homologous recombination. *BMC biology.* 2014;12:74.
112. Bosch E, Jobling MA. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Human molecular genetics.* 2003;12(3):341-7.
113. Antonell A, de Luis O, Domingo-Roura X, Perez-Jurado LA. Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. *Genome Res.* 2005;15(9):1179-88.
114. Uddin RK, Zhang Y, Siu VM, Fan Y-S, O'Reilly RL, Rao J, et al. Breakpoint Associated with a novel 2.3 Mb deletion in the VCFS region of 22q11 and the role of Alu (SINE) in recurring microdeletions. *BMC Med Genet.* 2006;7:18-.
115. Payer LM, Burns KH. Transposable elements in human genetic disease. *Nature Reviews Genetics.* 2019;20(12):760-72.
116. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet.* 2010;86(5):749-64.
117. de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet.* 2005;77(4):606-16.
118. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature.* 2008;455(7210):237-41.
119. Dunham I, Kundaje A, Aldred S, Collins P, Davis C, Doyle F, et al. The ENCODE Project Consortium: An integrated encyclopedia of DNA elements in the human genome. 2012. *Nature* 489: 57–74. *Nature.* 2012;489:57-74.
120. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, et al. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature genetics.* 2008;40(1):90-5.
121. Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, Jr., Hardekopf D, et al. Microdeletion and microduplication syndromes. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society.* 2012;60(5):346-58.
122. Lieber MR. Warner-Lambert/Parke-Davis Award Lecture. Pathological and physiological double-strand breaks: roles in cancer, aging, and the immune system. *Am J Pathol.* 1998;153(5):1323-32.
123. Bohgaki T, Bohgaki M, Hakem R. DNA double-strand break signaling and human disorders. *Genome Integr.* 2010;1(1):15-.
124. So A, Le Guen T, Lopez BS, Guirouilh-Barbat J. Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. *The FEBS Journal.* 2017;284(15):2324-44.
125. Weiner A, Zauberman N, Minsky A. Recombinational DNA repair in a cellular context: a search for the homology search. *Nature Reviews Microbiology.* 2009;7:748.

126. van Gent DC, Hoeijmakers JH, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nature reviews Genetics*. 2001;2(3):196-206.
127. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature genetics*. 2001;27(3):247-54.
128. Gelot C, Magdalou I, Lopez B. Replication Stress in Mammalian Cells and Its Consequences for Mitosis. *Genes*. 2015;6:267-98.
129. Holstein E-M, Clark KRM, Lydall D. Interplay between nonsense-mediated mRNA decay and DNA damage response pathways reveals that Stn1 and Ten1 are the key CST telomere-cap components. *Cell Rep*. 2014;7(4):1259-69.
130. Vamvakas S, Vock EH, Lutz WK. On the Role of DNA Double-Strand Breaks in Toxicity and Carcinogenesis. *Critical Reviews in Toxicology*. 1997;27(2):155-74.
131. Mladenov E, Iliakis G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutation research*. 2011;711(1-2):61-72.
132. So A, Le Guen T, Lopez BS, Guirouilh-Barbat J. Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. *Febs j*. 2017;284(15):2324-44.
133. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543-50.
134. O'Donovan PJ, Livingston DM. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*. 2010;31(6):961-7.
135. Helleday T, Lo J, van Gent DC, Engelward BP. DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA repair*. 2007;6(7):923-35.
136. Gonzalez-Marin C, Gosálvez J, Roy R. Types, Causes, Detection and Repair of DNA Fragmentation in Animal and Human Sperm Cells. *International journal of molecular sciences*. 2012;13:14026-52.
137. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*. 2017;18:495.
138. Grabarz A, Barascu A, Guirouilh-Barbat J, Lopez BS. Initiation of DNA double strand break repair: signaling and single-stranded resection dictate the choice between homologous recombination, non-homologous end-joining and alternative end-joining. *American journal of cancer research*. 2012;2(3):249-68.
139. Hakonarson H, Gulcher JR, Stefansson K. deCODE genetics, Inc. *Pharmacogenomics*. 2003;4(2):209-15.
140. <https://cran.r-project.org/>.
141. <https://www.genome.gov/10001688/international-hapmap-project>.
142. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*. 2007;23(21):2947-8.
143. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*. 1993;10(3):512-26.

144. Efron B, Halloran E, Holmes S. Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(14):7085-90.
145. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer applications in the biosciences* : CABIOS. 1994;10(2):189-91.
146. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology*. 2010;28(5):495-501.
147. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics (Oxford, England)*. 2016;32(2):289-91.
148. <https://genome.ucsc.edu/>.
149. <https://genome.ucsc.edu/ENCODE/>.
150. <https://www.genomatix.de/>.
151. Narang P, Dhapola P, Chowdhury S. BreCAN-DB: a repository cum browser of personalized DNA breakpoint profiles of cancer genomes. *Nucleic Acids Res*. 2016;44(D1):D952-D8.
152. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. *Nature genetics*. 2002;31(3):241-7.
153. Klein SJ, O'Neill RJ. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res*. 2018;26(1-2):5-23.
154. Finley J. Transposable elements, placental development, and oocyte activation: Cellular stress and AMPK links jumping genes with the creation of human life. *Medical hypotheses*. 2018;118:44-54.
155. Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci*. 2017;372(1736):20160455.
156. Pandey RS, Wilson Sayres MA, Azad RK. Detecting evolutionary strata on the human x chromosome in the absence of gametologous y-linked sequences. *Genome Biol Evol*. 2013;5(10):1863-71.
157. Ladias P, Markopoulos G, Lazaros L, Markoula S, Tzavaras T, Georgiou I. Holliday Junctions Are Associated with Transposable Element Sequences in the Human Genome. *Journal of molecular biology*. 2016;428(3):658-67.
158. Neofytou MC, Tsangaras K, Kypri E, Loizides C, Ioannides M, Achilleos A, et al. Targeted capture enrichment assay for non-invasive prenatal testing of large and small size sub-chromosomal deletions and duplications. *PloS one*. 2017;12(2):e0171319.
159. Kraus C, Vanicek T, Weidenauer A, Khanaqa T, Stamenkovic M, Lanzenberger R, et al. DiGeorge syndrome : Relevance of psychiatric symptoms in undiagnosed adult patients. *Wien Klin Wochenschr*. 2018;130(7-8):283-7.