



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
—ΙΔΡΥΘΕΝ ΤΟ 1837—

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Στοχευμένος προσδιορισμός αλληλουχίας, με
τεχνολογία δεύτερης γενιάς (Next Generation
Sequencing), σε ασθενείς με μυελικές νεοπλασίες»

ΛΟΥΠΗΣ Ν. ΘΕΟΔΩΡΟΣ

Πτυχιούχος Βιολογίας, ΕΚΠΑ

ΑΘΗΝΑ 2019



HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
— EST. 1837 —

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY

MASTER IN «BIOINFORMATICS»

Biomedical Research Foundation of the Academy of Athens (BRFAA),
Clinical, Experimental Surgery and Translational Research Center,
Haematology Research Lab

Master Diploma Thesis

«Study of molecular abnormalities in patients with
myeloid malignancies utilizing NGS targeted re-
sequencing»

THEODOROS N. LOUPIS
Biologist, University of Athens

ATHENS 2019



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
—ΙΔΡΥΘΕΝ ΤΟ 1837—

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)
Κέντρο Κλινικής Έρευνας, Πειραματικής Χειρουργικής και Μεταφραστικής Έρευνας,
Εργαστήριο Αιματολογίας

Διευθυντής: Καθηγητής Δημήτρης Λουκόπουλος
Επιστημονική Υπεύθυνη: Ειδική Λειτουργική Επιστήμονας Β' Κατερίνα Ζώη

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Στοχευμένος προσδιορισμός αλληλουχίας, με τεχνολογία δεύτερης γενιάς (Next Generation Sequencing), σε ασθενείς με μυελικές νεοπλασίες»

ΘΕΟΔΩΡΟΣ Ν. ΛΟΥΠΗΣ
Α.Μ:71705

Τριμελής εξεταστική επιτροπή

Καθηγητής Ιωάννης Π. Τρουγκάκος (Επιβλέπων)
Τομέας Βιολογίας Κυττάρου & Βιοφυσικής,
Τμήμα Βιολογίας, ΕΚΠΑ

Επίκουρη Καθηγήτρια Βασιλική Α. Οικονομίδου
Τομέας Βιολογίας Κυττάρου & Βιοφυσικής,
Τμήμα Βιολογίας, ΕΚΠΑ

Επίκουρος Καθηγητής Χρήστος Κοντός
Τομέας Βιοχημείας & Μοριακής Βιολογίας,
Τμήμα Βιολογίας, ΕΚΠΑ

ΑΘΗΝΑ 2019

Πρόλογος

Η εργασία αυτή πραγματοποιήθηκε στο Εργαστήριο Αιματολογίας του Κέντρου Κλινικής Έρευνας, Πειραματικής Χειρουργικής και Μεταφραστικής Έρευνας του Ιδρύματος Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών κατά την διάρκεια του ακαδημαϊκού έτους 2018-2019.

Από τη θέση αυτή θα ήθελα να ευχαριστήσω τον Διευθυντή του Εργαστηρίου Αιματολογίας Ομότιμο Καθηγητή κ. Δημήτρη Λουκόπουλο για την δυνατότητα που μου έδωσε να πραγματοποιήσω την Διπλωματική μου εργασία στο εργαστήριο του.

Επίσης, ένα μεγάλο ευχαριστώ από καρδιάς την επιστημονική μου υπεύθυνη κ. Κατερίνα Ζώη για την εμπιστοσύνη και πολύτιμη καθοδήγησή που μου έχει δείξει όλα αυτά τα χρόνια. Ακόμα θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς μου επιτροπής για την άψογη συνεργασία και βοήθεια τους. Τον κ. Ιωάννη Τρουγκάκο που δέχτηκε να είναι ο επιβλέπων της διπλωματικής μου εργασίας καθώς και την κα. Βασιλική Οικονομίδου και τον κ. Χρήστο Κοντό.

Τελειώνοντας, ένα τεράστιο ευχαριστώ στον Ανδρέα Γιαννόπουλο, μέτα-διδακτορικό φοιτητή του εργαστηρίου, που από την πρώτη στιγμή μου στο εργαστήριο με έκανε να νιώθω μέρος του και με υπομονή με δίδαξε ό,τι χρειαζόταν να ξέρω, καθώς και για την πολύτιμη βοήθεια και συνεισφορά του σε όλη την πορεία εκπόνησης της διπλωματικής μου εργασίας και παρουσίας μου στο εργαστήριο τα τελευταία χρόνια.

Τέλος, να ευχαριστήσω την Γεωργία, την Ελπίδα, την Μαρία και την Μαριλένα που με την βοήθεια τους και την παρουσία τους δημιούργησαν ένα υπέροχο κλίμα στο εργαστήριο.

Πίνακας Περιεχομένων

ΠΕΡΙΛΗΨΗ	10
1 ΕΙΣΑΓΩΓΗ	14
1.1 Μυελικές Νεοπλασίες	14
1.1.1 Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ)	14
1.1.2 Μυελοδυσπλαστικά Σύνδρομα (ΜΔΣ).....	16
1.1.3 Μυελοδυσπλαστικά Σύνδρομα / Μυελοϋπερπλαστικά Νεοπλάσματα (ΜΔΣ/ΜΥΝ)	18
1.2 Μοριακές βλάβες στις Μυελικές Νεοπλασίες	19
1.2.1 Μεταλλάξεις μικρής κλίμακας - Σημειακές αντικαταστάσεις βάσεων και μεταλλάξεις προσθήκης/διαγραφής νουκλεοτιδίων	21
1.3 Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με τεχνολογία επόμενης (δεύτερης) γενιάς (Next Generation Sequencing, NGS)	40
1.3.1 Εισαγωγικά στοιχεία	40
1.3.2 Στοχευμένος προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων (Targeted-Sequencing) με τεχνολογία NGS	41
1.3.3 Εφαρμογές στην κλινική πράξη	42
2 ΣΚΟΠΟΣ	44
3 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	45
3.1 Αντιδραστήρια	45
3.1.1 Χημικά αντιδραστήρια, βιολογικά αντιδραστήρια και ολοκληρωμένα συστήματα αντιδραστηρίων	45
3.1.2 Εργαστηριακός αναλυτικός εξοπλισμός	45
3.2 Κλινικό Υλικό	46
3.2.1 Προέλευση Δειγμάτων Ασθενών	46
3.3 Βασικές Αρχές Τεχνικών	46
3.3.1 Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με τεχνολογία επόμενης (δεύτερης) γενιάς (Next Generation Sequencing, NGS)	46
3.3.2 Το Docker	53
3.3.3 Η γλώσσα CWL (Common Workflow Language).....	55
3.4 Μεθοδολογία	56
3.4.1 Απομόνωση γενομικού DNA	56
3.4.2 Στοχευμένος Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με Τεχνολογία Δεύτερης Γενιάς (targeted Next Generation Sequencing, targeted NGS)	57
4 ΑΠΟΤΕΛΕΣΜΑΤΑ	68
4.1 Ενδεικτικά αποτελέσματα της ανάλυσης	68
4.1.1 Προσδιορισμός της ποιότητας των διαβασμάτων (quality control, QC).....	68
4.1.2 Επεξεργασία των διαβασμάτων	69
4.1.3 Αντιπαραβολή των επεξεργασμένων διαβασμάτων επί του πρότυπου γονιδιώματος (alignment to reference)	70
4.1.4 Προετοιμασία του αρχείου αντιπαραβεβλημένων διαβασμάτων (alignment file) για ανίχνευση μεταλλάξεων	71
4.1.5 Ανάδειξη μεταλλάξεων	72
4.1.6 Φιλτράρισμα σύμφωνα με ποιοτικά και ποσοτικά δεδομένα	73
4.1.7 Ενημέρωση του αρχείου μεταλλάξεων με βιολογικά και κλινικά δεδομένα.....	73
4.1.8 Φιλτράρισμα μεταλλάξεων σύμφωνα με βιολογικά και κλινικά δεδομένα.	74
4.2 Συγκεντρικά αποτελέσματα	75
4.2.1 Ανίχνευση μεταλλάξεων σε ασθενείς με ΜΔΣ.....	75
4.2.2 Ανίχνευση μεταλλάξεων σε ασθενείς με ΜΥΝ.	77

4.2.3	Ανίχνευση μεταλλάξεων σε ασθενή με ΜΔΣ/ΜΥΝ.	79
4.2.4	Ανίχνευση μεταλλάξεων σε ασθενείς με ΟΜΛ.....	80
5	ΣΥΖΗΤΗΣΗ.....	82
5.1	Σύγκριση αποτελεσμάτων της ανάλυσης μεταξύ των δύο διαφορετικών ορίων ποσοστού εμφάνισης της μετάλλαξης στο δείγμα.	82
5.2	Ανάλυση των μοριακών ευρημάτων 83	83
5.2.1	Σχολιασμός αποτελεσμάτων σε ομάδες ασθενών 84	84
5.3	Γενικά συμπεράσματα και προτάσεις..... 87	87
5.3.1	Μεθοδολογία κατασκευής των βιβλιοθηκών..... 87	87
5.3.2	Επιλογή κατάλληλου μηχανήματος αλληλούχισης και αναλώσιμων 88	88
5.3.3	Βιοπληροφορική ανάλυση των δεδομένων 88	88
5.4	Μελλοντικές προοπτικές..... 91	91
6	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	93
	ΠΑΡΑΡΤΗΜΑ Ι.....	102

ΠΕΡΙΛΗΨΗ

Σκοπός: Οι μυελικές νεοπλασίες είναι κλωνικές διαταραχές του αρχέγονου αιμοποιητικού κυττάρου (AAK) και χαρακτηρίζονται από άσκοπο και άμετρο πολλαπλασιασμό ή/και διαταραχή της ωρίμανσης. Οι διαταραχές αυτές χωρίζονται σε δύο μεγάλες ομάδες νοσημάτων, στην Οξεία Μυελογενή Λευχαιμία (ΟΜΛ) και στα Χρόνια Μυελικά Νεοπλάσματα. Η ταξινόμηση αυτή βασίζεται κυρίως στον αριθμό των βλαστών στο περιφερικό αίμα ή τον μυελό των οστών. Με τη σειρά τους, τα Χρόνια Μυελικά Νεοπλάσματα ταξινομούνται σε τρεις κύριες κατηγορίες, τις Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ), τα Μυελοδυσπλαστικά Σύνδρομα (ΜΔΣ) και τα νοσήματα εκείνα που παρουσιάζουν χαρακτηριστικά και των δύο προηγούμενων κατηγοριών (ΜΔΣ/ΜΥΝ) σύμφωνα με κλινικά, μορφολογικά, γενετικά και μοριακά δεδομένα, όπως καταγράφονται στην πρόσφατη αναθεώρηση των διαγνωστικών κριτηρίων που εκδόθηκε πρόσφατα, από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ., 2016).

Μέχρι πρόσφατα, η διάγνωση και ταξινόμηση αυτών των νοσημάτων, βασιζόταν σχεδόν αποκλειστικά σε κυτταρολογικά χαρακτηριστικά και καρυοτυπικές αναλύσεις, με εξαίρεση πολύ συγκεκριμένες μοριακές βλάβες, όπως εκείνες στα γονίδια *JAK2*, *CALR* και *MPL*, που ανιχνεύονται στην πλειοψηφία των ασθενών με ΜΥΝ, σπανιότερα δε στα ΜΔΣ/ΜΥΝ. Η πρόοδος της τεχνολογίας οδήγησε σε καλύτερο χαρακτηρισμό της παθογένειας των μυελικών νεοπλασιών, με την ανάδειξη πληθώρας μεταλλάξεων σε γονίδια με ετερογενείς κυτταρικές λειτουργίες και μάλιστα με διαφορετική συχνότητα σε κάθε περίπτωση. Χαρακτηριστικό παράδειγμα αποτελούν οι συχνές μεταλλάξεις στα γονίδια *JAK2*, *MPL* και *CALR*, που ανιχνεύονται στην πλειοψηφία των ασθενών με Ιδιοπαθή Θρομβοκυτταραιμία (ΙΘ) ή Μυελοϊνωση (ΜΙ), και οι οποίες θεωρούνται φαινοτυπικές (phenotypic drivers) μεταλλάξεις. Επιπλέον, μεταλλάξεις έχουν αναφερθεί σε γονίδια που εμπλέκονται στη ρύθμιση του επιγενετικού μηχανισμού ρύθμισης του κυττάρου, στον κυτταρικό κύκλο, στην ωρίμανση του mRNA και σε άλλες σημαντικές διεργασίες. Οι μεταλλάξεις αυτές θεωρείται πως διαμορφώνουν τον κλινικό φαινότυπο και συμβάλλουν στην πρόοδο και εξέλιξη της νόσου. Σκοπό της παρούσας Διπλωματικής Εργασίας, αποτέλεσε η μελέτη του Μοριακού τοπίου των ασθενών με μυελικές νεοπλασίες, μέσω στοχευμένου προσδιορισμού της νουκλεοτιδικής ακολουθίας, συχνά μεταλλαγμένων γονιδίων, με χρήση τεχνολογίας δεύτερης γενιάς (targeted panel Next Generation Sequencing) και η ανάπτυξη μίας συμβατής βιοπληροφορικής ροής, για την ανάλυση των δεδομένων.

Η διαδικασία αυτή συνεισφέρει στην εύκολη, γρήγορη και αξιόπιστη Μοριακή τυποποίηση των ασθενών με μυελικές νεοπλασίες, συμβάλλοντας έτσι στην ορθή διάγνωση και ταξινόμηση, αλλά και σε πρόγνωση και διαστρωμάτωση του κινδύνου εκτροπής αυτών προς ΟΜΛ, ενώ ιδανικά, δύναται να παρέχει πληροφορίες σχετικά με τις διαθέσιμες θεραπευτικές επιλογές.

Μέθοδοι: Στην παρούσα εργασία αναπτύχθηκε μία βιοπληροφορική ροή για την ανάλυση δεδομένων NGS, με στόχο την ανάδειξη μοριακών εξαλλαγών. Για το σκοπό αυτό χρησιμοποιήθηκε η προγραμματιστική γλώσσα Common Workflow Language (CWL) σε συνδυασμό με το λογισμικό Docker, που διευκολύνει τη χρήση, μεταφορά και επαναληψιμότητα της αναλυτικής διαδικασίας.

Συνολικά, μελετήθηκαν 51 ασθενείς με μυελικές νεοπλασίες σε Χρόνια ή Οξεία φάση. Αρχικά, απομονώθηκε DNA από περιφερικό αίμα ή μυελό των οστών. Ακολούθησε κατασκευή βιβλιοθηκών DNA, ενώ τέλος πραγματοποιήθηκε στοχευμένος προσδιορισμός της νουκλεοτιδικής ακολουθίας 28 γονιδίων ή μεμονωμένων περιοχών αυτών. Οι περιοχές των

γονιδίων που μελετήθηκαν αναφέρονται συχνά μεταλλαγμένα στους ασθενείς με μυελικές νεοπλασίες, όπως υποστηρίζεται από τη σύγχρονη βιβλιογραφία.

Αποτελέσματα: Συνολικά αναδείχθηκαν 223 μοριακές εξαλλαγές, εκ των οποίων 29 είχαν ήδη ανιχνευθεί προηγουμένως στο εργαστήριο μας, με ήδη εγκαθιδρυμένες τεχνικές Μοριακής Βιολογίας (post Real-time PCR HRMA και Sanger Sequencing). Από αυτές, 21 μεταλλάξεις εντοπίστηκαν σε 9 ασθενείς με ΜΔΣ, 3 στον μοναδικό ασθενή με ΜΔΣ/ΜΥΝ, 173 σε 33 ασθενείς με ΜΥΝ και 26 στους 8 ασθενείς με ΟΜΛ.

Συμπεράσματα: Με την παρούσα εργασία αναδείξαμε τη μεγάλη ετερογένεια του μοριακού τοπίου των μυελικών νεοπλασιών, ακόμα και σε ίδιες ή παρόμοιες κοορτές ασθενών. Παράλληλα εγκαθιδρύσαμε την τεχνική NGS, για την ταυτόχρονη μελέτη πολλών γενομικών περιοχών, ολόκληρων εξονίων ή και γονιδίων, για περισσότερους από έναν ασθενείς, σε μία μόνο αντίδραση και μάλιστα με υψηλή ευαισθησία.

Ταυτόχρονα, προτείνουμε βελτιώσεις, τόσο στο τεχνικό όσο και στο αναλυτικό υπόβαθρο της διαδικασίας, αποσκοπώντας σε βελτίωση της ευαισθησίας και ακρίβειας της μεθοδολογίας, κατά την ανάδειξη μοριακών εξαλλαγών, όπως επίσης σε εμπειριστατωμένη αξιολόγηση της επίπτωσης αυτών στο φαινότυπο της νόσου, με απώτερο στόχο την ενσωμάτωση της τεχνολογίας NGS στην κλινική πράξη.

SUMMARY

Objective: Myeloid neoplasms are clonal disorders of the hematopoietic stem cell (HSC) and are characterized by excess proliferation and/or disruption of maturation of one or more myeloid cell lineages. These disorders are divided into two major categories, acute myeloid leukemia (AML) and chronic myeloid neoplasms. This classification is mainly based on the number of blasts in peripheral blood or bone marrow. Chronic myeloid neoplasms are classified into three main categories according to clinical, morphological, genetic and molecular data, as reported in the recent revision of the World Health Organization (WHO, 2016), Myeloproliferative Neoplasms (MPNs), Myelodysplastic Syndromes (MDS) and a category with overlapping features of both MDS and MPNs, referred to as Myelodysplastic/Myeloproliferative neoplasms (MDS/MPN).

Until recently, the diagnosis and classification of these disorders was based almost exclusively on morphological and cytogenetic features, with the exception of very specific molecular lesions, such as those in the *JAK2*, *CALR* and *MPL* genes, detected in the majority of MPN patients, and more rarely in MDS/MPN. The recent advances in technology have led to a better characterization of the pathogenesis of the myeloid neoplasms, with the identification of an increasing number of genetic defects in genes with heterogeneous cellular functions. A typical example is the frequent mutations in the *JAK2*, *MPL* and *CALR* genes detected in the majority of patients with essential thrombocytosis (ET) and myelofibrosis (MF), which are considered as driver mutations.

Additionally, several other mutated genes have been identified and most of these mutations affect a range of essential, interrelated cellular mechanisms such as epigenetic regulation, RNA splicing, transcription and DNA damage response. The various combinations of mutations suggest a multistep pathogenesis and may account for the clinical heterogeneity. The delineation of the complex clonal architectures of these disorders could serve as the cornerstone for better risk stratification of these patients and for the identification of novel therapeutic targets within the context of Precision Medicine.

The purpose of this Diploma Thesis was to study the molecular landscape of patients with myeloid neoplasms through targeted re-sequencing, in commonly mutated genes, using Next Generation Sequencing (NGS) technology and developing a portable and user-friendly bioinformatics pipeline for the data analysis.

This process enables an affordable, rapid and reliable molecular screening of patients with myeloid neoplasms, contributing thus in a more accurate diagnosis and classification, prognosis and patients' risk stratification, and could also provide information regarding the available therapeutic options.

Methods: In this study a bioinformatics pipeline was developed for the analysis of NGS data aiming at variant calling. For this purpose, the Common Workflow Language (CWL) programming language was used in conjunction with the Docker software, which facilitates the easy use, transfer and repeatability of the analytical process.

Overall, 51 patients with myeloid neoplasms in chronic or blast phase were studied. DNA was extracted from peripheral blood or bone marrow. DNA libraries were constructed, and targeted re-sequencing was performed on 28 genomic targets (full genes or specific exons). The selection of the genomic regions under investigation represent mutational hotspots according to recent data and WHO and European LeukemiaNET guidelines.

Results: In total, 223 variants were identified, of which 29 were previously detected in our laboratory, with already established methods (Post Real-time PCR HRMA and Sanger

Sequencing). Of these, 21 variants were detected in 9 MDS patients, 3 in the only MDS/MPN patient, 173 in 33 MPN patients and 26 in 8 AML patients.

Conclusions: In this study, we highlighted the molecular heterogeneity of the genomic landscape of myeloid neoplasms, either within the same or similar cohorts of patients. We also established a custom NGS methodology for the simultaneous analysis of many genomic regions, entire exons and/or genes, for more than one patient, in a single experiment.

Furthermore, we propose amendments in both technical and analytical workflow of the process, aiming thus to improve the sensitivity and accuracy of the method regarding variant calling, as well as for the assessment of the impact of these variants on disease phenotype, in order the NGS technology to be reliably incorporated in clinical practice.

1 ΕΙΣΑΓΩΓΗ

1.1 Μυελικές Νεοπλασίες

Οι μυελικές νεοπλασίες είναι κλωνικές διαταραχές του αρχέγονου αιμοποιητικού κυττάρου (AAK) και χαρακτηρίζονται από άσκοπο και άμετρο πολλαπλασιασμό ενός ή περρισσοτέρων κυττάρων της μυελικής σειράς με συνοδή ή μη διαταραχή της ωρίμανσής του. Οι διαταραχές αυτές χωρίζονται σε δύο μεγάλες ομάδες νοσημάτων, την Οξεία Μυελογενή Λευχαιμία (ΟΜΛ) και στα Χρόνια Μυελικά Νεοπλάσματα. Η ταξινόμηση αυτή βασίζεται κυρίως στον αριθμό των βλαστών στο περιφερικό αίμα ή τον μυελό των οστών. Τα Χρόνια Μυελικά Νεοπλάσματα ταξινομούνται σε τρεις κύριες κατηγορίες σύμφωνα με κλινικά, μορφολογικά, γενετικά και μοριακά δεδομένα, όπως καταγράφονται στην πρόσφατη αναθεώρηση των διαγνωστικών κριτηρίων που εκδόθηκε πρόσφατα, από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ, 2016).

1.1.1 Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ)

Οι Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ) αποτελούν κλωνικές διαταραχές του AAK, που χαρακτηρίζονται από άμετρο και άσκοπο κλωνικό πολλαπλασιασμό ενός ή περισσότερων προγονικών κυττάρων της μυελικής σειράς, τα οποία, ωστόσο, διατηρούν σχετικά φυσιολογική ωρίμανση. Ο πολλαπλασιασμός αυτός οδηγεί σε αυξημένο αριθμό κοκκιοκυττάρων, ερυθροκυττάρων ή αιμοπεταλίων στο περιφερικό αίμα.

1.1.1.1 Ταξινόμηση των Μυελοϋπερπλαστικών Νεοπλασιών

Ταξινόμηση κατά Dameshek

Ο όρος Μυελοϋπερπλαστικά Σύνδρομα (ΜΥΣ) εντάσσει την Ιδιοπαθή Θρομβοκυτταραιμία (ΙΘ), την Αληθή Πολυκυτταραιμία (ΑΠ), την Πρωτοπαθή Μυελοϊνωση (ΠΜΙ) και τη Χρόνια Μυελογενή Λευχαιμία (ΧΜΛ), σε μία κοινή ομάδα νοσημάτων, με βάση το σχετικά όμοιο κλινικό τους φαινότυπο και την πεποίθηση ότι όλες υπόκεινται σε ένα κοινό αδιευκρίνιστο ερέθισμα, που οδηγεί στον ανεξέλεγκτο πολλαπλασιασμό των κυττάρων του μυελού των οστών (Dameshek 1951).

Ταξινόμηση από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.) το 2001 και η αναθεώρηση του 2008

Το 2001, ο Παγκόσμιος Οργανισμός Υγείας (Π.Ο.Υ.) (World Health Organization, WHO) κατέταξε τα ΜΥΣ σε μία ευρύτερη κατηγορία νοσημάτων, γνωστή ως Χρόνια Μυελοϋπερπλαστικά Σύνδρομα (ΧΜΥΣ) (Chronic Myeloproliferative Diseases, CMPDs). Στην ίδια ομάδα προστέθηκαν η Χρόνια Ουδετεροφιλική Λευχαιμία (ΧΟΛ), η Χρόνια Ηωσινοφιλική Λευχαιμία/Υπερηωσινοφιλικό Σύνδρομο (ΧΗΛ/ΥΗΣ), καθώς και τα αταξινόμητα Χρόνια Μυελοϋπερπλαστικά Σύνδρομα (αΧΜΥΣ).

Κοινό χαρακτηριστικό αυτής της ετερογενούς ομάδας νοσημάτων αποτελεί η κλωνική υπερπλασία της μυελικής σειράς, η οποία στο περιφερικό αίμα ανιχνεύεται ως αύξηση των κοκκιοκυττάρων, των αιμοπεταλίων ή/και των ερυθροκυττάρων, χωρίς όμως να συνοδεύεται από δυσερυθροποίηση, δυσπλασία της κοκκιδώδους σειράς ή μονοκυττάρωση.

Τέλος, οι διάφορες μορφές Μαστοκυττάρωσης αποτελούν μία ετερογενή ομάδα νοσημάτων που χαρακτηρίζεται από αυξημένο αριθμό μαστοκυττάρων και ταξινομείται ανεξάρτητα (Vardiman et al. 2001).

Η αναθεώρηση της κατάταξης αυτής από τον Π.Ο.Υ. το 2008 στηρίχθηκε στις παραπάνω βασικές πληροφορίες, αλλά περιέλαβε και νεότερα κλινικά, γενετικά και μοριακά δεδομένα. Συγκεκριμένα, ενσωματώθηκαν στα διαγνωστικά κριτήρια ορισμένοι μοριακοί δείκτες, περισσότερο ειδικοί των MYN. Σημειακές μεταλλάξεις στα εξώνια 12 και 14 του γονιδίου *JAK2* και στο εξόνιο 10 του γονιδίου *MPL* παρατηρούνται σε υψηλό ποσοστό ασθενών με ΙΘ, ΑΠ και ΠΜΙ και φαίνεται πως διαδραματίζουν καθοριστικό ρόλο στην εμφάνιση και εξέλιξη αυτών.

Στο ίδιο πλαίσιο εντάσσεται και η ανίχνευση του χρωμοσώματος Φιλαδέλφεια (Ph^1 , μη-φυσιολογικό χρωμόσωμα 22), που χαρακτηριστικά οδηγεί στη διάγνωση της ΧΜΛ. Το χρωμόσωμα αυτό σχηματίζεται από την αμοιβαία μετάθεση τμημάτων των χρωμοσωμάτων 9 και 22, η οποία επιφέρει σύντηξη του 5' τμήματος του γονιδίου *BCR* με το 3' τμήμα του γονιδίου *ABL*, προς σχηματισμό του υβριδικού γονιδίου *BCR-ABL* που θεωρείται το γενεσιουργό αίτιο της νόσου.

Παρόμοιες γενετικές και, κατ' επέκταση, μοριακές αναδιατάξεις συσχετίζονται με την εμφάνιση μυελικών νεοπλασιών με συνοδό ηωσηνοφιλία όπως η ΧΗΛ, ενώ η σημειακή μετάλλαξη *KITD816V*, αποτελεί μοριακό γνώρισμα της συστηματικής μαστοκυττάρωσης.

Τέλος, ο όρος "σύνδρομα" αντικαταστάθηκε από τον όρο "νεοπλάσματα", οπότε οι κύριες κατηγορίες ΧΜΥΣ μετονομάστηκαν σε Μυελοϋπερπλαστικά Νεοπλάσματα (MYN) (Tefferi and Vardiman 2008).

Ταξινόμηση από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.). Αναθεώρηση του 2016

Η ταξινόμηση του 2008 συμπληρώθηκε με πρόσφατα δεδομένα που προέκυψαν από τον προσδιορισμό της αλληλουχίας του DNA πολλών ασθενών με MYN, με τεχνολογίες αλληλούχισης δεύτερης γενιάς (Next Generation Sequencing, NGS).

Κύρια σημεία της αναθεώρησης είναι:

α) Η εύρεση μεταλλάξεων στο εξόνιο 9 του γονιδίου *CALR*, σε μεγάλο ποσοστό ασθενών με MYN, κυρίως ΙΘ και ΠΜΙ, οι οποίοι ήταν αρνητικοί για μεταλλάξεις *JAK2* και *MPL*, μέχρι πρότινος.

β) Η διάκριση της ΠΜΙ σε δύο περαιτέρω υποκατηγορίες, ανάλογα με το βαθμό της ίνωσης που παρατηρείται στο μυελό (προϊνωτικό στάδιο και ινωτικό στάδιο). Σημαντικό γεγονός αποτελεί πως και στις δύο περιπτώσεις, κύριο κριτήριο για την κατηγοριοποίηση των ασθενών αποτελούν οι μεταλλάξεις στα γονίδια *JAK2*, *MPL* και *CALR*. Σε περιπτώσεις τριπλά αρνητικών ασθενών για αυτές τις μεταλλάξεις, μείζονος σημασίας θεωρούνται, τροποποιητικές μεταλλάξεις σε γονίδια όπως τα *ASXL1*, *EZH2*, *TET2*, *IDH1/IDH2*, *SRSF2*, *SF3B1* κ.α, προκειμένου να αναδειχθεί η κλωνική φύση της νόσου.

γ) Η ένταξη της συστηματικής μαστοκυττάρωσης σε χωριστή κατηγορία, λόγω της μοναδικής κλινικής και παθολογοανατομικής εικόνας που παρουσιάζει (Arber et al. 2016).

Π.Ο.Υ. 2016: Αναθεώρηση ταξινόμησης Μυελικών Νεοπλασιών
Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ)
Χρόνια Μυελογενής Λευχαιμία (ΧΜΛ), BCR-ABL ⁺
Χρόνια Ουδετεροφιλική λευχαιμία (ΧΟΛ)
Αληθής Πολυκυτταραιμία (ΑΠ)
Πρωτοπαθής Μυελοϊνωση (ΠΜΙ) 4.1 ΠΜΙ σε προ-ινωτικό στάδιο 4.2 ΠΜΙ σε ίνωση
Ιδιοπαθής Θρομβοκυτταραιμία (ΙΘ)
Χρόνια Ηωσινοφιλική Λευχαιμία, μη αλλιώς ταξινομημένη (ΧΗΛ)
Αταξινόμητα ΜΥΝ (αΜΥΝ)
Μαστοκυττάρωση
Μυελικά Νεοπλάσματα με Ηωσινοφιλία και αναδιατάξεις των γονιδίων <i>PDGFRA</i>, <i>PDGFRB</i>, <i>FGFR1</i> ή <i>PCM1-JAK2</i>
Μυελικά Νεοπλάσματα με αναδιάταξη των γονιδίων <i>PDGFRA</i> , <i>PDGFRB</i> , <i>FGFR1</i> Προσωρινή οντότητα: Μυελικές ή Λεμφικές Νεοπλασίες με αναδιάταξη των γονιδίων <i>PCM1-JAK2</i>

Πίνακας 1. Η ταξινόμηση των ΜΥΝ σύμφωνα με τον Π.Ο.Υ (2016).

1.1.2 Μυελοδυσπλαστικά Σύνδρομα (ΜΔΣ)

Τα Μυελοδυσπλαστικά Σύνδρομα (ΜΔΣ) περιλαμβάνουν μια ανομοιογενή ομάδα αιματολογικών νοσημάτων. Αποτελούν διαταραχές του αρχέγονου αιμοποιητικού κυττάρου (ΑΑΚ), που οδηγούν σε αναποτελεσματική αιμοποίηση, ενώ παράλληλα τα κύτταρα μίας ή περισσότερων μυελικών σειρών χαρακτηρίζονται από δυσπλαστικά μορφολογικά χαρακτηριστικά και τάση μετασχηματισμού προς Οξεία Μυελογενή Λευχαιμία (Kaushansky K. et al. 2016)

Τα κλινικά συμπτώματα ασθενών με ΜΔΣ ποικίλουν από αναιμία, μέχρι λευκοπενία ή θρομβοπενία, μεμονωμένα ή συνολικά, ενώ ο μυελός είναι κυτταροβριθής. Τα ΜΔΣ εμφανίζονται κυρίως σε μεγάλη ηλικία, με μέση ηλικία εμφάνισης τα 70 έτη, ενώ κατά προσέγγιση 1 στους 500 νοσεί σε ηλικία άνω των 60 ετών, χωρίς ωστόσο αυτό να σημαίνει ότι δεν εμφανίζονται και σε νεότερες ηλικίες. Η αύξηση εμφάνισης των ΜΔΣ με την αύξηση της ηλικίας υποδηλώνει την συσσώρευση γενετικών και επιγενετικών διαταραχών, ενώ περίπου 50% των ασθενών φέρουν μια καρυοτυπική ανωμαλία και σε ποσοστό 72%-90% εντοπίζεται τουλάχιστον μια μετάλλαξη (Χατζημιχαηλ Ε et al. 2015).

Τα ΜΔΣ δεν απαντούν στους συνήθεις αιμοποιητικούς παράγοντες και στις παλαιότερες ταξινομήσεις ο όρος «ανθηκτική αναιμία», χρησιμοποιούταν μέχρι πρόσφατα για τον χαρακτηρισμό των διαφόρων υποομάδων αυτών (Kaushansky K. et al. 2016, Λουκόπουλος Δ. et al. 2015).

1.1.2.1 Ταξινόμηση των Μυελοδυσπλαστικών Συνδρόμων

Ταξινόμηση από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.) το 2001 και η αναθεώρηση του 2008

Η ταξινόμηση των ΜΔΣ σύμφωνα με τον ΠΟΥ έγινε το 2001 και αναθεωρήθηκε το 2008. Αυτή, βασίζεται τόσο στην μορφολογία των δυσπλαστικών κυττάρων, όσο και στον αριθμό των βλαστικών κυττάρων στο περιφερικό αίμα και στον μυελό των οστών.

Επιπροσθέτως συνυπολογίζει τον αριθμό και τον τύπο των δυσπλαστικών κυτταρικών σειρών. Επιπλέον, ένας μόνο κλινικός υπότυπος χαρακτηρίζεται από απώλεια μεγάλου τμήματος του μεγάλου βραχίονα του χρωμοσώματος 5 (del(5q)) (Kaushansky K. et al. 2016, Λουκόπουλος Δ. et al. 2015).

Τύπος ΜΔΣ	Περιφερικό Αίμα	Μυελός	Συχνότητα
<i>Κυτταροπενίες που αφορούν μία μόνο σειρά</i>			
Ανθεκτική Αναιμία ¹ Ανθεκτική ουδετεροπενία Ανθεκτική θρομβοπενία	Αναιμία Ουδετεροπενία Θρομβοπενία Βλάστες <1%	Δυσπλασία της αντίστοιχης σειράς (ερυθράς, μυελικής και μεγακαρυοκυτταρικής) σε >10% των κυττάρων). Βλάστες ≤ 5%	10-12%
Ανθεκτική Αναιμία με δακτυλιωτούς σιδηροβλάστες (RARS) ¹	Αναιμία Βλάστες <1%	Δυσπλασία μόνον της ερυθράς σειράς Δακτυλιωτοί σιδηροβλάστες >15% Βλάστες <5%	3-10%
<i>Κυτταροπενίες που αφορούν σε περισσότερες της μιάς σειρές</i>			
Ανθεκτική Αναιμία με δυσπλασία περισσότερων της μίας σειρών	Κυτταροπενία(ες) της αντίστοιχης σειράς <1% βλάστες	Δυσπλασία περισσότερων της μιάς σειρών ± δακτυλιωτοί σιδηροβλάστες <5% βλάστες χωρίς ραβδία Auer	30%
Ανθεκτική Αναιμία με περίσσεια βλαστών (RAEB-1) ²	Κυτταροπενία(ες) 2-9% βλάστες χωρίς ραβδία Auer	Δυσπλασία περισσότερων της μιάς σειρών ± δακτυλιωτοί σιδηροβλάστες 5-9% βλάστες χωρίς ραβδία Auer	40%
Ανθεκτική Αναιμία με περίσσεια βλαστών (RAEB-2) ²	Κυτταροπενία(ες) 10-19% βλάστες +/- ραβδία Auer	Δυσπλασία περισσότερων της μιάς σειρών ± δακτυλιωτοί σιδηροβλάστες. 10-19% βλάστες, ± ραβδία Auer	
Αταξινόμητο Μυελοδυσπλαστικό σύνδρομο	Κυτταροπενίες Βλάστες <1%	Δυσπλασία στο <10% των κυττάρων, αλλά τυπική κυτταρογενετική ανωμαλία ή RCUF/ RCMD με 1% βλάστες στο αίμα ή RCUD με παγκυταροπενία	<5%
Μυελοδυσπλαστικό Σύνδρομο με μεμονωμένη ανωμαλία [del(5q)]	Αναιμία Φυσιολογικά ή αυξημένα σιμοπετάλια Όχι ή σπάνιοι βλάστες	Φυσιολογικά ή αυξημένα μεγακαρυοκύτταρα με υπολοβωμένους πυρήνες. Βλάστες <5%, χωρίς ραβδία Auer Μεμονωμένη κυτταρογενετική ανωμαλία [del(5q)].	<1%
Μυελοδυσπλαστικά/ Μυελοϋπερπλαστικά Νεοπλάσματα (ΜΔΣ/MYN)	Χρόνια Μυελομονοκυτταρική Λευχαιμία Άτυπη χρόνια μυελοκυτταρική Λευχαιμία Νεανική Μυελομονοκυτταρική Λευχαιμία Αταξινόμητο ΜΔΣ/MYN Ανθεκτική αναιμία με δακτυλιωτούς σιδηροβλάστες και θρομβοκυττάρωση		
1. Refractive anemia with ringed sideroblasts			
2. Refractive anemia with excess of blasts			

Πίνακας 2. Η ταξινόμηση των ΜΔΣ σύμφωνα με τον ΠΟΥ 2008. Λουκόπουλος Δ., Πολίτου Μ., Κεφ. 6 Μυελοδυσπλαστικά Σύνδρομα, Μαθήματα Αιματολογίας 2015

Ταξινόμηση από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.). Αναθεώρηση του 2016

Το 2016 ο Π.Ο.Υ. προχώρησε σε μια νέα αναθεώρηση σύμφωνα με την οποία ο όρος RA (ανθεκτική αναιμία) αντικαταστάθηκε με τον όρο ΜΔΣ. Αυτή η νέα ταξινόμηση παρουσιάζεται στον πίνακα 3.

Π.Ο.Υ. 2016: Αναθεώρηση ταξινόμησης Μυελικών Νεοπλασιών
Μυελοδυσπλαστικά Σύνδρομα (ΜΔΣ)
ΜΔΣ με δυσπλασία μίας σειράς
ΜΔΣ με δυσπλασία περισσότερων της μίας σειράς
ΜΔΣ με δακτυλιωτούς σιδηροβλάστες (ΜΔΣ-ΔΣ)
ΜΔΣ-ΔΣ με δυσπλασία μίας σειράς
ΜΔΣ-ΔΣ με δυσπλασία περισσότερων της μίας σειράς
ΜΔΣ με διαγραφή 5q
ΜΔΣ με περίσσεια βλαστών (ΜΔΣ-ΠΒ)
ΜΔΣ-ΠΒ 1
ΜΔΣ-ΠΒ 2
Αταξινόμητα ΜΔΣ
Με 1% βλάστες στο αίμα
Με δυσπλασία μίας σειράς και πανκυτταροπενία
Βασισμένα σε χαρακτηριστική κυτταρογενετική ανωμαλία
Παιδική ανθεκτική κυτταροπενία

Πίνακας 3. Η ταξινόμηση των ΜΔΣ σύμφωνα με τον Π.Ο.Υ (2016).

1.1.3 Μυελοδυσπλαστικά Σύνδρομα / Μυελοϋπερπλαστικά Νεοπλάσματα (ΜΔΣ/ΜΥΝ)

Όταν η μυελική νεοπλασία συνοδεύεται από δυσερυθροποίηση, δυσπλασία ή μονοκυττάρωση, τότε τα αντίστοιχα σύνδρομα περιγράφονται από μια τρίτη κατηγορία μυελικών νεοπλασιών, γνωστά ως Μυελοδυσπλαστικά Σύνδρομα/Μυελοϋπερπλαστικές διαταραχές (ΜΔΣ/ΜΥΔ) (Myelodysplastic /Myeloproliferative diseases, MDS/MPD).

Στην κατηγορία αυτή εντάσσονται η άτυπη Χρόνια Μυελοειδής Λευχαιμία (αΧΜΛ), η Χρόνια Μυελομονοκυτταρική Λευχαιμία (ΧΜΜΛ), η Νεανική Μυελομονοκυτταρική Λευχαιμία (ΝΜΜΛ), καθώς και αταξινόμητες διαταραχές (αΜΔΣ/ΜΥΔ), όπως η προσωρινά ταξινομημένη οντότητα, Ανθεκτική Αναιμία με Δακτυλιωτούς Σιδηροβλάστες και Θρομβοκυττάρωση (ΑΑΔΣ-Θ).

Μυελοδυσπλαστικά Σύνδρομα/Μυελοϋπερπλαστικά Νεοπλάσματα (ΜΔΣ/ΜΥΝ)
Χρόνια Μυελομονοκυτταρική Λευχαιμία (ΧΜΜΛ)
Άτυπη Χρόνια Μυελογενής Λευχαιμία (αΧΜΛ), <i>BCR-ABL1</i> αρνητική
Νεανική Μυελομονοκυτταρική Λευχαιμία (ΝΜΜΛ)
ΜΔΣ/ΜΥΝ με δακτυλιωτούς σιδηροβλάστες και θρομβοκυττάρωση (ΜΔΣ/ΜΥΝ-ΔΣ-Θ)
Αταξινόμητα ΜΔΣ/ΜΥΝ (αΜΔΣ/ΜΥΝ)

Πίνακας 4. Η ταξινόμηση των ΜΔΣ/ΜΥΝ σύμφωνα με τον Π.Ο.Υ (2016).

1.2 Μοριακές βλάβες στις Μυελικές Νεοπλασίες

Σε μία γενική συνοπτική θεώρηση, η παθογένεια του καρκίνου οφείλεται στη δημιουργία και επικράτηση κλωνικών κυτταρικών πληθυσμών που ξεφεύγουν από τους μηχανισμούς ελέγχου του φυσιολογικού τους πολλαπλασιασμού, μέσω απόκτησης μεταλλάξεων που οδηγούν σε διαφορετικά γονιδιακά μετάγραφα, τα οποία μετατρέπουν τη λειτουργία αυτή σε αυτόνομη και ανεξέλεγκτη.

Τα εναλλακτικά αυτά μετάγραφα, που υπό περιπτώσεις είναι η αιτία για την παθογένεια των επιμέρους νοσημάτων, συντίθενται κυρίως μέσω δύο ευρέων κατηγοριών βιολογικών φαινομένων:

α) Η πρώτη κατηγορία αφορά σε μεταλλάξεις μεγάλης κλίμακας, όπου παρατηρούνται είτε διαγραφές/προσθήκες ολόκληρων χρωμοσωμάτων ή εκτεταμένων τμημάτων αυτών, είτε σε αναδιατάξεις χρωμοσωμικού υλικού, κυρίως προς σχηματισμό υβριδικών γονιδίων. Στη δεύτερη περίπτωση, συχνά, παρατηρείται η γεινίαση και αλληλεπίδραση ισχυρών υποκινητών, με γονίδια που προάγουν τον κυτταρικό πολλαπλασιασμό, όπως στην περίπτωση της ΧΜΛ που εκδηλώνεται λόγω δόμησης του χιμαιρικού γονιδίου *BCR-ABL1*, εξαιτίας της χρωμοσωμικής αναδιάταξης $t(9;22)(q34;q11)$.

β) Η δεύτερη κατηγορία αφορά σε μεταλλάξεις μικρής κλίμακας, όπου περιλαμβάνονται είτε σημειακές αντικαταστάσεις μοναδικών βάσεων, όπως εκείνες στα γονίδια *JAK2* και *MPL*, είτε διαγραφές/προσθήκες λίγων νουκλεοτιδίων, όπως οι μεταλλάξεις *CALR*.

Ωστόσο, σε μία προσπάθεια βαθύτερης κατανόησης της δράσης των μεταλλάξεων που ανιχνεύονται στους ασθενείς που πάσχουν από ΜΥΝ, αυτές μπορούν να ταξινομηθούν περαιτέρω, με βάση της συνεισφορά τους στη βαρύτητα της νόσου, σε δύο ανεξάρτητες των προηγούμενων κατηγορίες:

α) Στην πρώτη κατηγορία ανήκουν μεταλλάξεις που έχουν συσχετιστεί και είναι πλέον γνωστό πως ευθύνονται σε μεγάλο βαθμό για την εμφάνιση του παθολογικού φαινότυπου. Οι μεταλλάξεις αυτές χαρακτηρίζονται ως κύριες ή φαινοτυπικές (*driver mutations*), και εντοπίζονται στα γονίδια *JAK2*, *MPL* και *CALR*, ενώ ανάμεσά τους συμπεριλαμβάνονται και χρωμοσωμικές αναδιατάξεις, όπως το χρωμόσωμα Φιλαδέλφεια στη ΧΜΛ.

β) Στη δεύτερη κατηγορία ανήκουν μεταλλάξεις σε επιπρόσθετα γονίδια, που διαμορφώνουν και τροποποιούν τον φαινότυπο των μυελικών νεοπλασιών. Οι

μεταλλάξεις αυτές χαρακτηρίζονται ως δευτερεύουσες ή συνοδές μεταλλάξεις (passenger mutations) και ενδεχομένως οδηγούν σε εξέλιξη της νόσου και σε εκτροπή προς Οξεία Μυελογενή Λευχαιμία (ΟΜΛ).

Πιο συγκεκριμένα, τα τελευταία χρόνια, η δυνατότητα λεπτομερέστερης ανάλυσης του γονιδιώματος, έδειξε ότι σε ορισμένες παθολογικές καταστάσεις, μερικά γονίδια εμφανίζουν συχνότερα μεταλλάξεις, σε σχέση με φυσιολογικούς μάρτυρες.

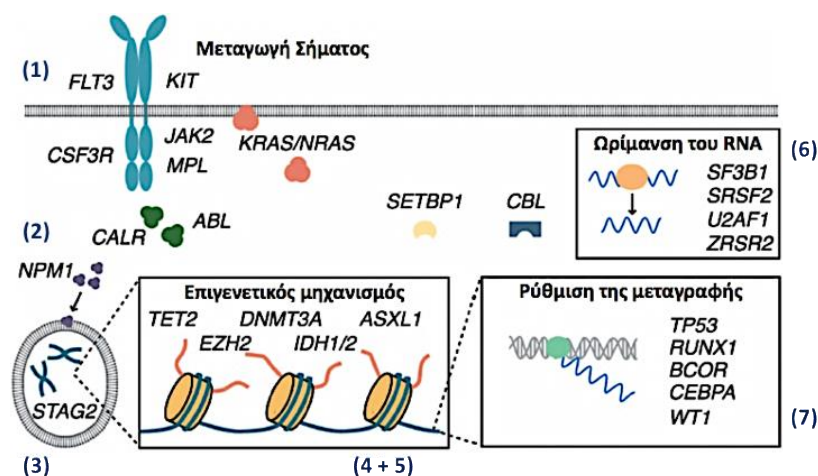
Οι μεταλλάξεις αυτές δεν έχουν άμεση σχέση με την παθογένεια των νοσημάτων στα οποία διαπιστώνονται, προφανώς όμως συμβάλλουν στην εμφάνιση και εξέλιξη αυτών, αφού κατά κανόνα είναι συχνότερες στα νοσήματα με βαρύτερη πρόγνωση. Άλλωστε, η παρουσία ορισμένων εξ αυτών, διαπιστώνεται ακόμη και σε άτομα χωρίς ενδείξεις νόσου, μάλιστα, με συχνότητα που αυξάνεται με την ηλικία.

Τα δεδομένα που υπάρχουν μέχρι σήμερα, για το είδος των μεταλλάξεων και κατ' επέκταση τη σημασία των μεταβολών που αυτές επιφέρουν σε πρωτεϊνικό επίπεδο, είναι περιορισμένα.

Ωστόσο, η μελέτη τους αποτελεί επίκεντρο ερευνητικού ενδιαφέροντος και βρίσκει ήδη εφαρμογή στην κλινική πράξη ως προγνωστική παράμετρος διαφόρων νοσημάτων του αίματος.

Τα γονίδια στα οποία παρατηρούνται συνοδές μεταλλάξεις είναι πολλά και άσχετα μεταξύ τους, ενώ για λόγους απλούστευσης, μπορούν να ομαδοποιηθούν ανάλογα με την κυτταρική λειτουργία στην οποία μετέχουν, όπως φαίνεται και στη γραφική αναπαράσταση που ακολουθεί. Έτσι διακρίνουμε τις εξής κατηγορίες:

- 1) Ενδοκυττάρια μεταγωγή σήματος (*JAK2*, *MPL*, *CSF3R*, *KIT*, *RAS*, *FLT3* κ.α.)
- 2) Γονίδια με περισσότερες από μία λειτουργίες (*CALR*, *SETBP1*, *NPM1* κ.α.)
- 3) Κυτταρικός Πολλαπλασιασμός (*STAG2* κ.α.)
- 4) Μεθυλίωση του DNA (*DNMT3A*, *TET2*, *IDH1/2* κ.α.)
- 5) Τροποποίηση της χρωματίνης (*ASXL1*, *EZH2* κ.α.)
- 6) Ωρίμανση του πρόδρομου RNA (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2* κ.α.)
- 7) Μεταγραφή (*TP53*, *RUNX1*, *BCOR*, *CEBPA*, *WT1* κ.α.)
- 8) Αποικοδόμηση των πρωτεϊνών
- 9) Απόπτωση του κυττάρου (Harada and Harada 2015).



Εικόνα 1. Οι κυτταρικές λειτουργίες σημαντικών γονιδίων της αιμοποίησης. Πληθώρα σημαντικών κυτταρικών λειτουργιών επηρεάζονται από συχνά εμφανιζόμενες, αλλά και πιο σπάνιες μεταλλάξεις σε διάφορα γονίδια (Tremblay et al. 2018).

1.2.1 Μεταλλάξεις μικρής κλίμακας - Σημειακές αντικαταστάσεις βάσεων και μεταλλάξεις προσθήκης/διαγραφής νουκλεοτιδίων

1.2.1.1 Μεταλλάξεις σε γονίδια που συμβάλλουν στη μεταγωγή σήματος

1.2.1.1.1 Το γονίδιο JAK2

Το γονίδιο *JAK2* εντοπίζεται στη χρωμοσωμική περιοχή 9p24.1 και απαρτίζεται από 26 εξώνια.

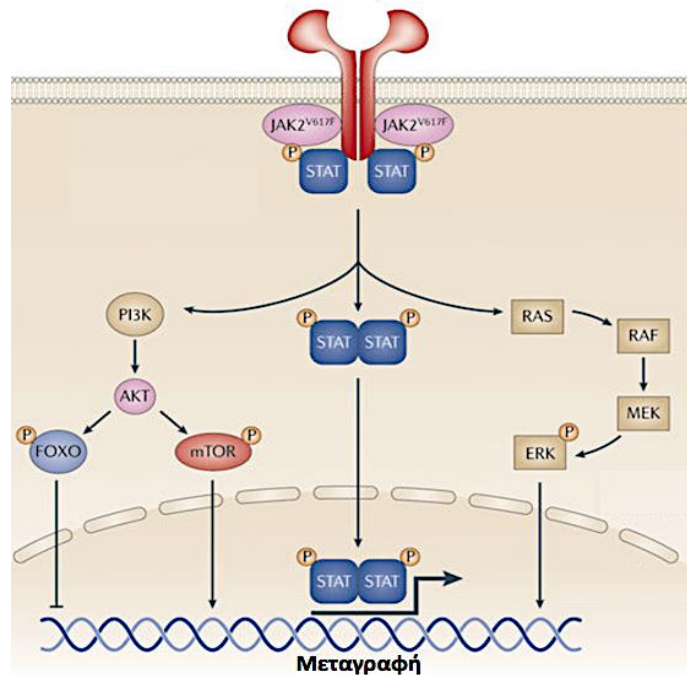
Το γονίδιο *JAK2* (Janus Kinase 2) κωδικοποιεί για την ομώνυμη πρωτεΐνη, με δράση κινάσης τυροσίνης, η οποία διαδραματίζει σημαντικό ρόλο στην επιβίωση, στον πολλαπλασιασμό και στη διαφοροποίηση των αιμοποιητικών κυττάρων.

Η σύνδεση της κατάλληλης κυτταροκίνης στο εξωκυτταρικό τμήμα του υποδοχέα, οδηγεί σε διμερισμό και ενεργοποίησή του μέσω φωσφορυλίωσης. Με τον τρόπο αυτό ενεργοποιείται η κινάση *JAK2*, γεγονός που έχει ως αποτέλεσμα την ενεργοποίηση πληθώρας σηματοδοτικών μονοπατιών, εντός του κυττάρου, μέσω διαδοχικών φωσφορυλίσεων.

Η επίκτητη σωματική μετάλλαξη V617F στο εξόνιο 14 του γονιδίου *JAK2*, χαρακτηρίστηκε το 2005 (Vardiman et al. 2001, Baxter et al. 2005, Kralovics et al. 2005, Jones et al. 2005, Levine et al. 2005).

Η μετάλλαξη έχει ως αποτέλεσμα την αντικατάσταση ενός αμινοξέος βαλίνης, από φαινυλαλανίνη στη θέση 617 (V617F) της αμινοξικής αλληλουχίας. Η θέση αυτή εντοπίζεται σε μια υψηλά συντηρημένη περιοχή της πρωτεΐνης, η οποία δρα ως ψευδοκινάση (JH2).

Η αμινοξική αλλαγή V617F έχει ως αποτέλεσμα την απώλεια της ανασταλτικής λειτουργίας της περιοχής JH2. Ως συνέπεια, ενεργοποιούνται αδιάκοπα, μέσω αυτοφωσφορυλίωσης και διαδοχικών ετεροφωσφορυλίσεων, υποκείμενα σηματοδοτικά μονοπάτια, όπου εμπλέκονται μεταγραφικοί παράγοντες και άλλες κινάσες (STATs, RAS/ERK(MAPK) και PI3K/AKT), οδηγώντας εν τέλει σε ανεξέλεγκτο πολλαπλασιασμό των αιμοποιητικών κυττάρων.



Εικόνα 2. Η σηματοδότηση JAK2V617F. Διαρκώς ενεργοποιημένα σηματοδοτικά μονοπάτια λόγω απρόσκοπτης δράσης κινάσης που επιδεικνύει το μεταλλαγμένο πρωτεϊνικό προϊόν JAK2V617F, απουσία εξωκυτταρικού ερεθίσματος (Τροποποίηση από Quintás-Cardama et al. 2011).

Η μετάλλαξη *JAK2V617F* ανευρίσκεται στο 95% τουλάχιστον των περιπτώσεων ΑΠ και στο 50-60% των περιπτώσεων ΙΘ και ΠΜΙ (Jones et al. 2005, Zhao et al. 2005) και σε ασθενείς με ΜΔΣ/ΜΥΝ, όπως αΧΜΛ (4-8%), ΧΜΜΛ (10%) και ΜΔΣ/ΜΥΝ-ΔΣ-Θ (60%) (Jones et al. 2005, Zoi and Cross 2015).

Επιπρόσθετες μεταλλάξεις παρατηρούνται αποκλειστικά σε ασθενείς με ΑΠ, στο εξόνιο 12 του γονιδίου *JAK2*, σε σαφώς χαμηλότερα ποσοστά (1-2%). Τόσο η V617F, όσο και οι μεταλλάξεις στο εξόνιο 12 του γονιδίου *JAK2*, είναι μεταλλάξεις απόκτησης λειτουργίας (gain-of-function mutations, GOF) (Scott et al. 2007).

1.2.1.1.2 Το γονίδιο *MPL*

Το γονίδιο *MPL* εντοπίζεται στη χρωμοσωμική περιοχή 1p34.2 και απαρτίζεται από 11 εξώνια. Το ανθρώπινο γονίδιο *c-MPL* αναγνωρίστηκε το 1992 και φέρει ομολογία με το ογκογονίδιο *v-mpl*, που ανιχνεύθηκε το 1990, σε στελέχη ιού που προκαλεί λευχαιμογένεση στα τρωκτικά (murine myeloproliferative leukemia virus, MMLV).

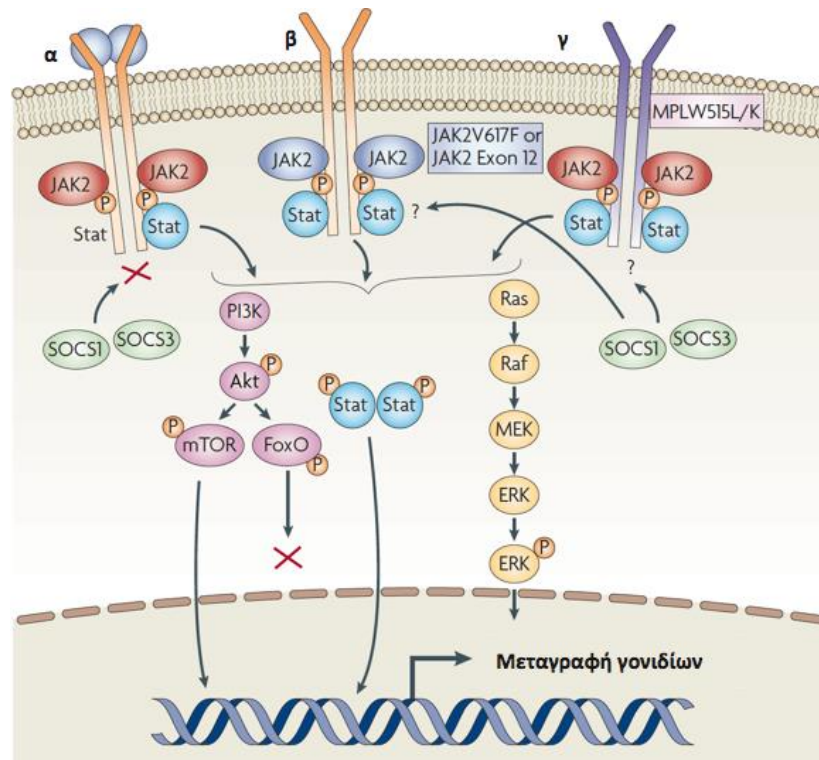
Δεδομένα προσδιορισμού αλληλουχίας αποκάλυψαν πως το γονίδιο *c-MPL* κωδικοποιεί για μία πρωτεΐνη, η οποία παρουσιάζει υψηλή ομολογία με μέλη της υπερικογένειας των υποδοχών των αιμοποιητικών κυττάρων.

Το πρώτο-ογκογονίδιο *MPL* (*c-MPL*) (myeloproliferative leukemia virus), κωδικοποιεί για τον κυτταρικό υποδοχέα της θρομβοποιητίνης (TPOR), ο οποίος διαδραματίζει σημαντικό ρόλο στην αιμοποίηση.

Η πρόσδεση θρομβοποιητίνης (TPO) στον υποδοχέα της, επιφέρει το διμερισμό αυτού. Ακολουθεί ενεργοποίηση της τυροσινικής κινάσης *JAK2*, η οποία εντοπίζεται στο ενδοκυτταρικό τμήμα του υποδοχέα. Έτσι, μέσω διαδοχικών φωσφορυλιώσεων,

δραστηριοποιούνται σημαντικά σηματοδοτικά μονοπάτια, όπως περιεγράφηκε στο προηγούμενο κεφάλαιο (He et al. 2013).

Η αντικατάσταση του αμινοξέος τρυπτοφάνη (W), από ένα μόριο λευκίνης (L) ή λυσίνης (K), στη θέση 515 (W515L/K) της διαμεμβρανική περιοχής του υποδοχέα TPOR, οδηγεί σε απρόσκοπτη ενεργοποίηση των σηματοδοτικών μονοπατιών JAK/STAT, PI3K/AKT/mTOR, RAS/MAPK, δίχως να είναι απαραίτητη η πρόσδεση TPO στον υποδοχέα, με συνέπεια την ανεξέλεγκτη, μη φυσιολογική, αύξηση των μεγακαρυοκυττάρων.



Εικόνα 3. Φυσιολογική και μη φυσιολογική ενδοκυτταρική σηματοδότηση. α. Φυσιολογική ρύθμιση της δράσης των υποδοχέων κυτταροκινών. β. Η μεταλλαγμένη JAK2 ενεργοποιείται αυτόνομα, χωρίς να είναι απαραίτητη η πρόσδεση εξωτερικών σηματοδοτικών παραγόντων, στους υποδοχείς όπου εδράζει η κινάση γ. Μεταλλάξεις στο γονίδιο MPL, επίσης, οδηγούν σε έκφραση τροποποιημένων υποδοχέων θρομβοποιητίνης, ικανών να ενεργοποιήσουν απρόσκοπτα, φυσιολογικά σε αυτή την περίπτωση, μόρια JAK2, απουσία θρομβοποιητίνης (Τροποποίηση από Levine et al. 2007).

Οι μεταλλάξεις απόκτησης λειτουργίας (GOF) στο γονίδιο *MPL*, ανιχνεύονται συχνά σε ασθενείς με MYN οι οποίοι ελέγχονται αρνητικοί για άλλες κύριες μεταλλάξεις.

Πιο συγκεκριμένα, ανιχνεύονται σε ποσοστό 5% και 8% των ασθενών με ΙΘ και ΠΜΙ, αντίστοιχα, καθώς και σε ποσοστό περίπου 5% ασθενών με ΜΔΣ/MYN-ΔΣ-Θ (He et al. 2013, Zoi and Cross 2015).

1.2.1.1.3 Το γονίδιο *CSF3R*

Το γονίδιο *CSF3R* εδράζει στη χρωμοσωμική περιοχή 1p34.3 και αποτελείται από 19 εξώνια (Maxson et al. 2013) που κωδικοποιούν για τον υποδοχέα της αυξητικής ορμόνης των κοκκιοκυττάρων (G-CSF3R, granulocyte colony-stimulating factor 3 receptor), ο οποίος

διαδραματίζει καθοριστικό ρόλο στη διαφοροποίηση και πολλαπλασιασμό των ουδετερόφιλων. Οι μεταλλάξεις απόκτησης λειτουργίας (GOF) που ανιχνεύονται στο γονίδιο *CSF3R*, ανήκουν σε δύο βασικές κατηγορίες σε αυτές δίχως νόημα (nonsense mutations) (Y752X, W791X) και σε αυτές που μετατοπίζουν το ανοιχτό πλαίσιο ανάγνωσης (frameshift mutations) (D771fs, S783fs), οι οποίες όλες οδηγούν σε πρόωρο τερματισμό της μεταγραφής και απώλεια τμήματος του κυτταροπλασματικού άκρου της πρωτεΐνης. Στη δεύτερη κατηγορία ανήκουν σημειακές μεταλλάξεις με λάθος νόημα (missense mutations), με συχνότερη την T618I, οι οποίες εντοπίζονται στο κυτταροπλασματικό άκρο της πρωτεΐνης, πλησίον της κυτταροπλασματικής μεμβράνης. Οι μεταλλάξεις αυτές ενεργοποιούν το μονοπάτι JAK2-STAT.

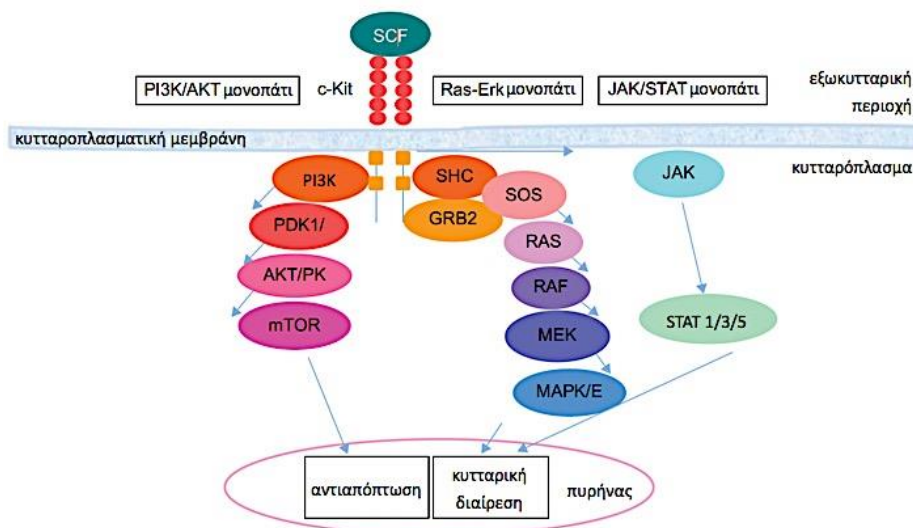
Οι μεταλλάξεις *CSF3R*, φαίνεται να εμφανίζονται αποκλειστικά σε ασθενείς με ΧΟΛ (80-90%). Επομένως, η ανίχνευση αυτών αποτελεί ένα πολύτιμο κριτήριο διαφοροδιάγνωσης, αφού πολλές φορές η διάκριση μεταξύ ΧΟΛ, αΧΜΛ και ΧΜΜΛ είναι ιδιαίτερα δύσκολη (Maxson et al. 2013, Wang et al. 2014, Dao et al. 2017).

1.2.1.1.4 Το γονίδιο *KIT*

Το ανθρώπινο κυτταρικό πρωτο-ογκογονίδιο *KIT* (*c-KIT*), φέρει ομολογία και προφανώς αποτελεί εξελικτικό κατάλοιπο του ομώνυμου ογκογονιδίου (*v-KIT*) μίας συγκεκριμένης κατηγορίας ρετροϊών, που έχει βρεθεί πως επιμολύνουν οριζόντια τα αιλουροειδή, επάγοντας τη λευχαιμογένεση σε αυτά (feline leukemia viruses, FeLV) (Besmer et al. 1986, d'Auriol et al. 1988). Εντοπίζεται στο χρωμόσωμα 4q11 και απαρτίζεται από 21 εξόνια. Το πρώτο από αυτά συμπεριλαμβάνει το κωδικόνιο έναρξης της μετάφρασης. Τα εξόνια 2-9 κωδικοποιούν το εξωκυτταρικό τμήμα του υποδοχέα, το εξόνιο 10 τη διαμεμβρανική περιοχή, ενώ τα 11-21 το ενδοκυτταρικό κομμάτι της πρωτεΐνης.

Μέχρι στιγμής, στον άνθρωπο έχουν παρατηρηθεί τέσσερις διαφορετικές ισομορφές μεταγράφων του γονιδίου που κωδικοποιεί σε έναν τύπου III υποδοχέα με εγγενή δράση κινάσης τυροσίνης. Ο υποδοχέας *KIT* ανιχνεύεται τόσο σε διάφορους τύπους μη αιμοποιητικών κυττάρων, όσο και στα αρχέγονα αιμοποιητικά κύτταρα, όπου με τη σταδιακή ωρίμανση αυτών, η έκφρασή του μειορρυθμίζεται. Από τα ώριμα αιμοποιητικά κύτταρα, μόνο τα μαστοκύτταρα, συνεχίζουν να εκφράζουν διαρκώς τον υποδοχέα σε υψηλά επίπεδα (Sun et al. 2009, Pardanani 2016).

Υπό φυσιολογικές συνθήκες, η πρόσδεση του παράγοντα των βλαστικών κυττάρων (SCF) στο εξωκυττάριο τμήμα του υποδοχέα, οδηγεί σε διμερισμό και ενεργοποίηση μέσω αυτοφωσφορυλίωσης του τελευταίου. Με τον τρόπο αυτό δραστηριοποιεί καταρροϊκά σηματοδοτικά μονοπάτια, όπως τα PI3K/AKT/mTOR, JAK/STAT και RAS/MAPK. Αυτά οδηγούν σε μεταγραφή γονιδίων, τα οποία συμμετέχουν σε διάφορες κυτταρικές λειτουργίες, όπως ο πολλαπλασιασμός, η επιβίωση, η διαφοροποίηση και η μετακίνηση (Figueira et al. 2014).



Εικόνα 4. Η δράση του ενεργοποιημένου φυσιολογικού υποδοχέα KIT (τροποποίηση από Babaei et al. 2016).

Στην πλειονότητα των ασθενών με συστηματική μαστοκυττάρωση (82-89%), ανιχνεύονται σωματικές μεταλλάξεις απόκτησης λειτουργίας (GOF), με συχνότερη την D816V, καθώς και άλλες πιο σπάνιες, όπως οι D816(H/Y), στο εξόνιο 17 του γονιδίου *KIT*. Οι μεταλλάξεις αυτές συμβάλλουν στη διαρκή ενεργότητα του υποδοχέα, αφού εντοπίζονται στην περιοχή με δράση κινάσης τυροσίνης (Soucie et al. 2012, Jawhar et al. 2016).

Έτσι, έχουν ως αποτέλεσμα την ενεργοποίηση ενδοκυτταρικών σηματοδοτικών μονοπατιών που όλα συμβάλλουν στον ίδιο φαινότυπο, όπου περιλαμβάνονται η αναστολή της απόπτωσης και η προαγωγή της κυτταρικής αύξησης, του πολλαπλασιασμού και της μετακίνησης των κυττάρων.

Σήμερα, υπάρχουν διάφοροι θεραπευτικοί παράγοντες, οι οποίοι είναι περισσότερο ή λιγότερο κατάλληλοι για την αναστολή του εκάστοτε μονοπατιού σηματοδότησης, όπως φαίνεται στην εικόνα που ακολουθεί (Babaei et al. 2016).

1.2.1.1.5 Τα γονίδια *RAS*

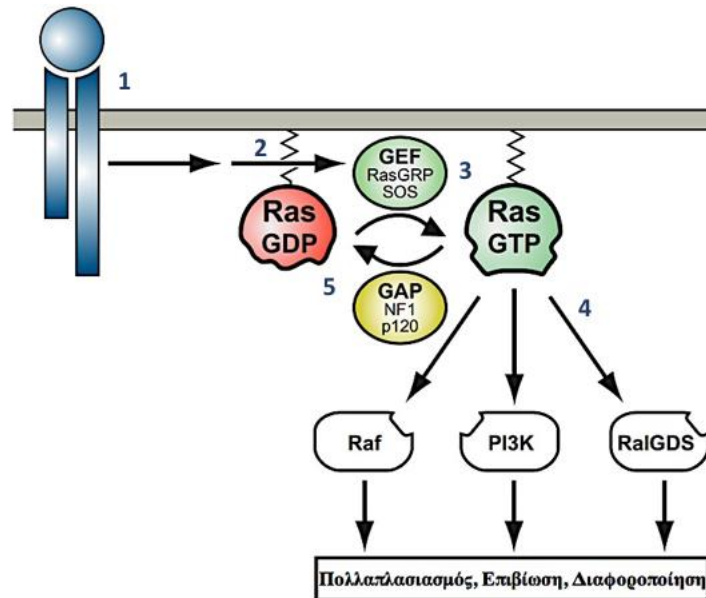
Η οικογένεια *RAS* (*rat sarcoma*) απαρτίζεται από υψηλά συντηρημένα γονίδια που ανευρίσκονται και εκφράζονται σε όλους τους ευκαρυωτικούς οργανισμούς. Τα γονίδια αυτά λόγω της εμπλοκής τους στην καρκινογένεση, χαρακτηρίζονται ως πρωτο-ογκογονίδια. Η οικογένεια *RAS* χωρίζεται σε τρεις υποοικογένειες, την *H-RAS*, *K-RAS* και *N-RAS* και εδράζουν στα χρωμοσώματα 11p15.5, 12p12.1 και 1p13.2 αντίστοιχα. Όλες οι τρεις προαναφερθείσες υποοικογένειες γονιδίων εκφράζονται στον άνθρωπο (Barbacid 1987, Scott et al. 2016).

Οι πρωτεΐνες *RAS* αποτελούν καθοριστικούς ρυθμιστές, σημαντικών κυτταρικών διεργασιών, όπως ο πολλαπλασιασμός, η διαφοροποίηση και η επιβίωση, αφού μεσολαβούν της μετάδοσης εξωγενών ερεθισμάτων από διάφορους διαμεμβρανικούς υποδοχείς προς τον πυρήνα του κυττάρου, μέσω ενεργοποίησης πολυάριθμων πολύπλοκων σηματοδοτικών μονοπατιών.

Ουσιαστικά αποτελούν μοριακούς διακόπτες, όπου, μέσω μεταβολής της στερεοδομής των συμπλόκων που απαρτίζουν, εναλλάσσονται μεταξύ μίας ενεργούς δομής, η οποία διατηρεί

δεσμευμένη μία τριφωσφορική γουανοσίνη (guanosine triphosphate, GTP) (Ras-GTP) και μίας ανενεργούς δομής, όπου η GTP έχει απωλέσει ένα μόριο φωσφορικού οξέος (guanosine diphosphate, GDP) (Ras-GDP).

Υπό φυσιολογικές συνθήκες, η ενεργός σηματοδότηση σταματά, με την υδρόλυση του GTP προς GDP, που μεσολαβείται από την ενδογενή δράση GTPάσης της RAS, η οποία ενισχύεται από επιπρόσθετους παράγοντες (GTPase activating proteins, GAPs), κυρίως τους NF1 και p120 GAP (Εικόνα 5, στάδιο 5) (Ward et al. 2012).



Εικόνα 5. Μοντέλο δράσης των πρωτεϊνών RAS. Συντμήσεις: GAP: GTPase activating proteins, GDP: guanosine diphosphate, GEF: guanine nucleotide exchange factors, GTP: guanosine triphosphate (τροποποίηση από Ward et al. 2012).

Οι σημειακές μεταλλάξεις που ανιχνεύονται συχνότερα στο γονίδιο RAS, ευνοούν τη διατήρηση της ενεργής διαμόρφωσης GTP-RAS, μέσω μειωμένης υδρόλυσης του GTP, κυρίως λόγω ανθεκτικότητας στη δράση των GAPs.

Σε αυτή την περίπτωση, η θεραπευτική στόχευση είναι δυσχερής, αφού η μοριακή βλάβη ουσιαστικά αφορά στον ενδογενή μηχανισμό αρνητικής ρύθμισης της πρωτεϊνικής δραστηριότητας και όχι σε ενισχυμένη ενεργότητα κινάσης.

Για το λόγο αυτό, η ερευνητική κοινότητα έχει εστιάσει κυρίως στην αναστολή στοιχείων, που εντοπίζονται καταρροϊκά των σηματοδοτικών μονοπατιών Ras, όπως τα PI3K, AKT, mTOR και MEK (Ward et al. 2012, Al-Kali et al. 2013, Akutagawa et al. 2016).

Στις μυελικές διαταραχές παρατηρούνται συχνότερα μεταλλάξεις NRAS με αυτές να εντοπίζονται κυρίως σε περιπτώσεις NMMΛ και ΧΜΜΛ, σε ποσοστό 25-30% και >20%, αντίστοιχα.

Επιπλέον, τα γονίδια NRAS και KRAS, ανευρίσκονται αντίστοιχα μεταλλαγμένα σε ποσοστό ασθενών 10-11% και 5%, στην ΟΜΛ.

Λιγότερο συχνά ανιχνεύονται μεταλλάξεις σε ασθενείς με ΜΔΣ (5%), οπότε και σχετίζονται με αυξημένο κίνδυνο εκτροπής προς ΟΜΛ (Ward et al. 2012, Al-Kali et al. 2013, Oshima et al. 2016, Meggendorfer et al. 2018).

1.2.1.2 Μεταλλάξεις σε γονίδια που εμπλέκονται σε πληθώρα κυτταρικών διεργασιών

1.2.1.2.1 Το γονίδιο *CALR*

Το γονίδιο *CALR* εντοπίζεται στη χρωμοσωμική περιοχή 19p13.13 και απαρτίζεται από 9 εξώνια.

Το γονίδιο *CALR*, κωδικοποιεί στην ομώνυμη πρωτεΐνη, η οποία ανιχνεύεται στο ενδοπλασματικό δίκτυο, στην πλασματική μεμβράνη και στο κυτταρόπλασμα των κυττάρων. Στη φυσιολογική της μορφή δρα ως μοριακός συνοδός (chaperon), συμμετέχοντας στη σωστή αναδίπλωση πρωτεϊνών, στην ομοιόσταση της συγκέντρωσης του Ca^{2+} , ενώ παράλληλα αλληλοεπιδρά με ορμονικούς υποδοχείς του πυρήνα (Heuser et al. 2014).

Το καρβοξυτελικό άκρο της πρωτεΐνης περιλαμβάνει μια αρνητικά φορτισμένη, ανεξάρτητη λειτουργική περιοχή για τη δέσμευση ιόντων ασβεστίου, καθώς και ένα υψηλά συντηρημένο μοτίβο, για τη σταθερή πρόσδεση αυτής στο ενδοπλασματικό δίκτυο (KDEL).

Πρόσφατα, από δυο ανεξάρτητες ομάδες, δημοσιεύθηκε η παρουσία συχνών μεταλλάξεων στο εξόνιο 9 του γονιδίου της καλρετικούλίνης (calreticulin, *CALR*), κυρίως σε ασθενείς με MYN (Nangalia et al. 2013, Klampfl et al. 2013).

Μέχρι στιγμής έχουν ανιχνευτεί πάνω από 50 διαφορετικές μεταλλάξεις στο εξόνιο 9 του γονιδίου *CALR*, οι οποίες είναι μεταλλάξεις έλλειψης ή προσθήκης βάσεων, οι οποίες ανεξαιρέτως οδηγούν σε μετατόπιση του ανοικτού πλαισίου ανάγνωσης (Open Reading Frame, ORF).

Όλες αυτές οι μεταλλάξεις έχουν ως αποτέλεσμα τη μετατροπή των αρνητικά φορτισμένων αμινοξέων του καρβοξυτελικού άκρου της *CALR*, σε αμινοξέα με θετικό και ουδέτερο φορτίο, με συνέπεια την απώλεια του μοτίβου πρόσδεσης στο ενδοπλασματικό δίκτυο.

Αυτό οδηγεί την πρωτεΐνη, σε απώλεια της ικανότητας πρόσδεσης ιόντων ασβεστίου, αλλά και σε διαταραχή της τοπογραφίας αυτής στο κύτταρο μέσω της απώλειας της ρυθμιστικής ακολουθίας KDEL, κάτι που φαίνεται να διαδραματίζει σημαντικό ρόλο στο μη-φυσιολογικό πολλαπλασιασμό των κυττάρων που φέρουν τέτοιες μεταλλάξεις (Cazzola and Kralovics 2014).

Πρόσφατα δεδομένα δείχνουν πως η θετικά φορτισμένη μεταλλαγμένη *CALR*, που εντοπίζεται εντός του ενδοπλασματικού δικτύου, αλληλοεπιδρά με τον υποδοχέα της θρομβοποιητίνης και τον συνοδεύει στην πλασματική μεμβράνη, όπου εκεί διμερίζεται απουσία προσδέτη.

Αυτό συντελεί σε διαρκή ενεργοποίηση, τόσο του ίδιου του υποδοχέα, όσο και των υποκείμενων σηματοδοτικών μονοπατιών που περιεγράφηκαν στα προηγούμενα κεφάλαια και τα οποία οδηγούν σε μη φυσιολογικό πολλαπλασιασμό των κυττάρων (Cazzola 2016, Araki et al. 2016).

Οι μεταλλάξεις στο εξόνιο 9 του γονιδίου της *CALR* ανιχνεύονται σχεδόν αποκλειστικά σε ασθενείς με ΙΘ (15-25%), ΠΜΙ (20-25%) και ΜΔΣ/MYN-ΔΣ-Θ (5%) (Patnaik and Tefferi 2015). Αυτό αντιστοιχεί σε ποσοστό 70-85% των ασθενών με ΙΘ και ΠΜΙ, που ελέγχονται αρνητικοί για μεταλλάξεις στα γονίδια *JAK2* και *MPL*.

Οι ίδιες μεταλλάξεις εντοπίζονται σε μικρό αριθμό ασθενών με ΜΔΣ (8%), αλλά χαρακτηριστικά όχι στην ΑΠ (Nangalia et al. 2013, Klampfl et al. 2013).

1.2.1.2.2 Το γονίδιο *SETBP1*

Το γονίδιο *SETBP1* εντοπίζεται στη χρωμοσωμική περιοχή 18q12.3 και απαρτίζεται από 15 εξώνια. Κατόπιν εναλλακτικής ωρίμανσης προκύπτουν διαφορετικοί συνδυασμοί εξωνίων, οι οποίοι αποδίδουν διάφορες ισομορφές ώριμων προϊόντων mRNA.

Το γονίδιο *SETBP1* κωδικοποιεί για μία πρωτεΐνη η οποία εντοπίζεται κυρίως εντός του πυρήνα του κυττάρου, αλλά και στο κυτταρόπλασμα, επιτελώντας πληθώρα δράσεων. Το γονίδιο εκφράζεται τόσο σε βλαστικά και προγονικά αιμοποιητικά κύτταρα, όσο και σε κύτταρα που έχουν ήδη δεσμευτεί προς μία συγκεκριμένη αιμοποιητική σειρά.

Η πρωτεΐνη *SETBP1* σχηματίζει ετεροδιμερές με το προϊόν του ογκογονιδίου *SET*, το οποίο ρυθμίζει αρνητικά την ογκοκατασταλτική πρωτεΐνη φωσφατάση 2A (*PP2A*) που αναστέλλει τον κυτταρικό πολλαπλασιασμό.

Η συμπλοκοποίηση των πρωτεϊνών *SETBP1* και *SET* προστατεύει την τελευταία από αποικοδόμηση, αναστέλλοντας περαιτέρω τη δράση της *PP2A*, ενισχύοντας παράλληλα τον κυτταρικό πολλαπλασιασμό.

Οι μεταλλάξεις *SETBP1* διαταράσσουν ένα χαρακτηριστικό μοτίβο της πρωτεΐνης, όπου υπό φυσιολογικές συνθήκες, προσδένεται η λιγάση ουβικουιτίνης E3. Η ουβικουιτινυλίωση των πρωτεϊνών από τις λιγάσες E3, αποτελεί μείζονος σημασίας τροποποίηση, αφού ρυθμίζει πληθώρα κυτταρικών διεργασιών, όπως η μετακίνηση, η επιδιόρθωση του DNA, η κυτταρική σηματοδότηση και η ρύθμιση του κυτταρικού κύκλου.

Η απώλεια ουβικουιτινυλίωσης του μεταλλαγμένου προϊόντος, οδηγεί σε μειωμένη αποικοδόμηση της πρωτεΐνης *SETBP1* και κατά συνέπεια της *SET*. Συγχρόνως, η *PP2A* παρουσιάζει ιδιαίτερα μειωμένη δράση, με αποτέλεσμα τα κύτταρα αυτά να παρουσιάζουν έντονα αυξημένο ρυθμό αυτοανανέωσης.

Οι συχνότερες επίκτητες μεταλλάξεις *SETBP1* ανιχνεύονται μεταξύ των αμινοξέων Glu858 to Ile871, στην υψηλά συντηρημένη SKI-ομόλογη περιοχή, που παρουσιάζει μερική ομολογία με την ογκοπρωτεΐνη SKI.

Υπό φυσιολογικές συνθήκες, το τμήμα αυτό είναι χρήσιμο για την πρόσδεση της E3 λιγάσης ουβικουιτίνης, που προάγει την αποικοδόμηση της *SETBP1*.

Οι μεταλλάξεις αυτές δεν επιτρέπουν την πρόσδεση της λιγάσης E3, οδηγώντας έτσι σε υπερέκφραση της πρωτεΐνης *SETBP1*, με αποτέλεσμα τη διαρκή συμπλοκοποίηση με τον συμπάραγοντα *SET* και την ογκοκατασταλτική πρωτεΐνη *PP2A*, με συνέπεια την απώλεια της ογκοκατασταλτικής δράσης της τελευταίας (Maxson JE and Tyner 2017).

Οι μεταλλάξεις *SETBP1* ανιχνεύονται σε υψηλό ποσοστό ασθενών με αΧΜΛ (24-32%) ή με ΧΟΛ (25%), στην τελευταία κατάσταση σε συνδυασμό με μεταλλάξεις *CSF3R*, σπανιότερα δε παρατηρούνται σε περιπτώσεις ΧΜΜΛ (4-7%), ΜΔΣ (2%), ΠΜΙ και ΟΜΛ.

Αποτελούν παράγοντα ιδιαίτερα δυσμενούς πρόγνωσης, με ασθενείς με μεταλλάξεις *SETBP1* να παρουσιάζουν σημαντικά χαμηλότερη μέση επιβίωση 22 μηνών, συγκριτικά με τους 77 μήνες που επιβιώνουν κατά μέσο όρο οι ασθενείς χωρίς τέτοιες μεταλλάξεις (Makishima et al. 2013, Piazza et al. 2013, Zoi and Cross 2015, Clara et al. 2016, Elliott and Tefferi 2016, Shou et al. 2017).

1.2.1.3 Μεταλλάξεις σε γονίδια που εμπλέκονται στην επιγενετική τροποποίηση του DNA

Η επιγενετική αναφέρεται στη μελέτη αναστρέψιμων κληρονομήσιμων αλλαγών στη λειτουργία των γονιδίων χωρίς αυτά να παρουσιάζουν διαφορές στην νουκλεοτιδική αλληλουχία (Jaenisch and Bird 2003, Kaushansky et al. 2016).

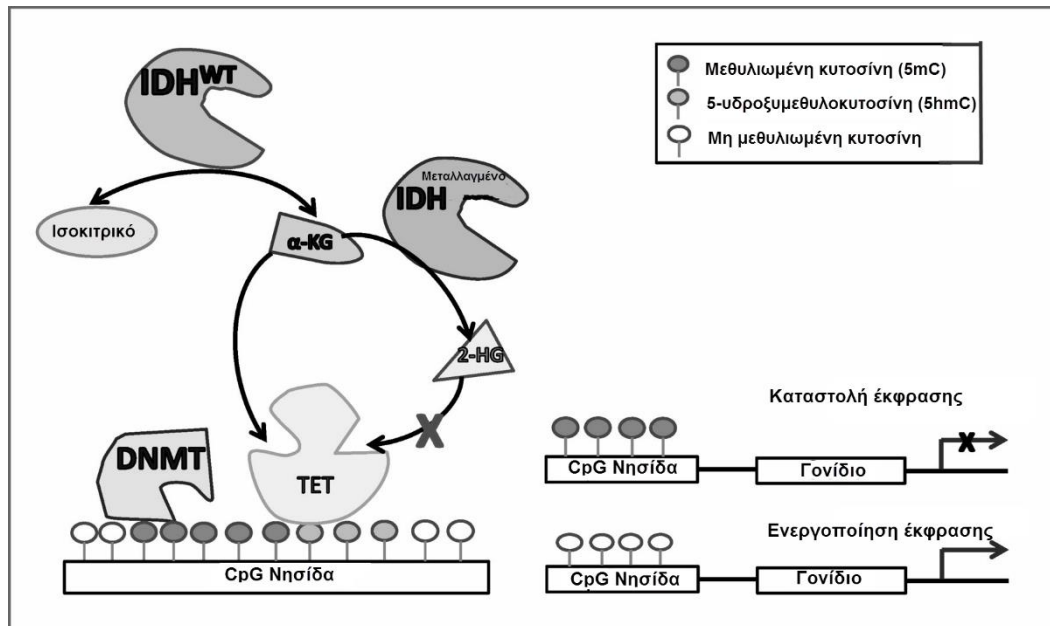
Αυτό επιτυγχάνεται με διεργασίες όπως η μεθυλίωση και η απομεθυλίωση του DNA και οι τροποποιήσεις των ιστονών πρωτεϊνών, που οδηγούν σε αναδιάταξη της χρωματίνης, η οποία πλέον αλληλοεπιδρά διαφορετικά με πρωτεΐνες που ρυθμίζουν άμεσα τη γονιδιακή έκφραση, όπως οι μεταγραφικοί παράγοντες (Jaenisch and Bird 2003).

Μεταλλάξεις σε γονίδια που σχετίζονται με τη φυσιολογική λειτουργία των παραπάνω μηχανισμών, έχουν αναφερθεί σε διάφορες μεταβολικές, αυτοάνοσες και νευρολογικές ασθένειες, όπως επίσης στον καρκίνο.

Με τον όρο μεθυλίωση του DNA νοείται η προσθήκη μιας μεθυλομάδας (CH₃) σε μια βάση κυτοσίνης (C) του DNA προς σχηματισμό 5-μεθυλο-κυτοσίνης (5mC), όπως φαίνεται και στην εικόνα 4. Στα σπονδυλόζωα η προσθήκη της μεθυλομάδας γίνεται σε βάσεις κυτοσίνης που ακολουθούνται από μία βάση γουανίνης, δηλαδή σε δι-νουκλεοτίδια κυτοσίνης-γουανίνης (CpG dinucleotides). Αυτά τα δινουκλεοτίδια κατά κύριο λόγο απαντώνται σε περιοχές του γονιδιώματος γνωστές και ως νησίδες CpG (CpG islands). Ένα μεγάλο ποσοστό αυτών των νησίδων φαίνεται να βρίσκεται σε ρυθμιστικές θέσεις και υποκινητές γονιδίων. Τα ένζυμα DNA μεθυλοτρανσφεράσες (DNA Methyltransferases; DNMTs) είναι υπεύθυνα για την μεθυλίωση του DNA (Santini V. et al. 2015).

Αποτέλεσμα της μεθυλίωσης του DNA είναι η αποσιώπηση γονιδίων. Αυτό συμβαίνει είτε μέσω παρεμπόδισης της πρόσδεσης μεταγραφικών παραγόντων σε ρυθμιστικές περιοχές γονιδίων είτε μέσω προσέλκυσης ειδικών καταστολέων που είναι σε θέση να αναγνωρίζουν και να προσδένονται σε μεθυλιωμένες αλληλουχίες δι-νουκλεοτιδίων CpG (Sansom OJ et al 2007).

Η αντίστροφη διαδικασία ονομάζεται απομεθυλίωση του DNA και καταλύεται από το ένζυμο TET2, του οποίου η λειτουργία θα αναλυθεί παρακάτω



Εικόνα 6. Ο ρόλος των DNMT, TET και IDH πρωτεϊνών στη μεθυλίωση και απομεθυλίωση του DNA (Τροποποίηση από Connie A. et al., 2013).

1.3.2.3.1 Το γονίδιο *DNMT3A*

Το γονίδιο *DNMT3A* εδράζει στο χρωμόσωμα 2p23.3 ανήκει στην οικογένεια των γονιδίων *DNMT*, τα οποία κωδικοποιούν μεθυλομεταφοράσες του DNA, που είναι υπεύθυνες για την μεθυλίωση του DNA, μέσω της μετατροπής των βάσεων κυτοσίνης (Cyt) προς 5-μεθυλοκυτοσίνη (5-mC) (Walter et al. 2011, Cross 2011).

Οι πρωτεΐνες *DNMT3A/B* οδηγούν σε de novo μεθυλίωση του DNA με την *DNMT3A*, εκφράζεται σταθερά σε διαφοροποιημένους ιστούς, ενώ φαίνεται πως είναι χρήσιμη για τη φυσιολογική διαφοροποίηση των κυττάρων (Moore et al. 2013).

Η πλειονότητα των μεταλλάξεων που ανιχνεύονται στο γονίδιο *DNMT3A* είναι παρανοηματικές μεταλλάξεις (missense mutations) κυρίως στο εξόνιο 23. Σχεδόν σε κάθε περίπτωση εντοπίζονται σε ετεροζυγωτία. Πιο συχνά εμφανίζονται οι μεταβολές R882(H/C/S) (Im et al. 2014), πλησίον του καρβοξυτελικού άκρου της πρωτεΐνης, όπου εντοπίζεται η λειτουργική περιοχή με ενεργότητα μεθυλομεταφοράσης (Russler-Germain et al. 2014).

Διάφορες μελέτες αποδεικνύουν πως η απώλεια δράσης της *DNMT3A*, σε πολυδύναμα αιμοποιητικά (Hematopoietic Stem Cells, HSCs) και εμβρυϊκά βλαστικά κύτταρα (Embryonic Stem, ES) ποντικών, συντελεί σε μειωμένη ικανότητα διαφοροποίησης, καθώς και σε ενισχυμένη ικανότητα αυτοανανέωσης (Chen et al. 2003, Challen et al. 2011).

Οι μεταλλάξεις στο γονίδιο *DNMT3A*, ανευρίσκονται κυρίως σε ασθενείς με ΟΜΛ (20%) (Yamashita et al. 2010, Im et al. 2014) και ΜΔΣ (8-10%) (Walter et al. 2011), σπανιότερα δε σε ασθενείς με ΠΜΙ, είτε αυτή είναι πρωτοπαθής, είτε αποτελεί εξέλιξη από ΑΠ (5% και 8% αντίστοιχα) (Brecqueville et al. 2014).

Συνιστούν μεταλλάξεις με κακή πρόγνωση, ιδίως για τους ασθενείς με ΟΜΛ, ενώ στις περιπτώσεις αυτές, συνανιχνεύονται συχνά με μεταλλάξεις στα γονίδια *NPM1*, *FLT3* και *IDH1* (Koya et al. 2016).

1.2.1.3.1 Το γονίδιο *TET2*

Το γονίδιο *TET2* (Ten-Eleven Translocation 2) εδράζεται στο γενετικό τόπο 4q24 και ανήκει στην οικογένεια των *TET* γονιδίων. Οι *TET* πρωτεΐνες είναι ένζυμα με δράση υδροξυλάσης και καταλύουν την μετατροπή της 5-μεθυλοκυτοσίνης (5mC) προς 5-υδροξυμεθυλοκυτοσίνη (5hmC). Με τον τρόπο αυτό επάγουν την απομεθυλίωση του DNA (Tahiliani M. et al. 2014).

Το ένζυμο *TET2* απαιτεί για την λειτουργία του και την κατάλυση 5-μεθυλοκυτοσίνης προς 5-υδροξυμεθυλοκυτοσίνη, α -Κετογλουταρικό (α -KG) και δισθενή κατιόντα σιδήρου (Fe^{2+}).

Μεταλλάξεις στο γονίδιο *TET2* είναι αρκετά συχνές σε ασθενείς με ΧΟΛ (30%) ΧΜΜΛ (40-60%), αΧΜΛ (30-40%) και αταξινόμητα ΜΔΣ/ΜΥΝ (30%) (Zoi and Cross 2015, Maxson and Tyner 2017), σπανιότερα δε ανιχνεύονται σε ασθενείς με ΠΜΙ (14%), ή με μυελοϊνωση ως εξέλιξη ΑΠ (15%) ή ΙΘ (21%) (Brecqueville et al. 2014).

Αυτές αφορούν συνήθως μεταλλάξεις με λάθος νόημα (missense mutations), κοντά στο καταλυτικό εξελικτικά συντηρημένο καρβοξυτελικό άκρο (-COOH) της πρωτεΐνης, οι οποίες έχουν ως αποτέλεσμα την απώλεια ενζυμικής λειτουργίας αυτής (LOF mutations), ενώ παρατηρούνται και στα δύο αντίγραφα του γονιδίου στον παθολογικό πληθυσμό κυττάρων.

Οι μεταλλάξεις στα γονίδια *TET2* και *DNMT3A* είναι αμοιβαία αποκλειόμενες (Tefferi et al. 2009, Soucie et al. 2012, Larsson et al. 2013, Solary et al. 2014).

1.2.1.3.2 Τα γονίδια *IDH1* και *IDH2*

Το γονίδιο *IDH1* εδράζεται στο γενετικό τόπο 2q33, ενώ το γονίδιο *IDH2* στο 15q26. Και τα δυο αυτά γονίδια κωδικοποιούν ένζυμα με λειτουργία NADP+-εξαρτώμενων ισοκιτρικών αφυδρογονασών (Isocitrate Dehydrogenase; IDH) τα οποία καταλύουν την αντίδραση:



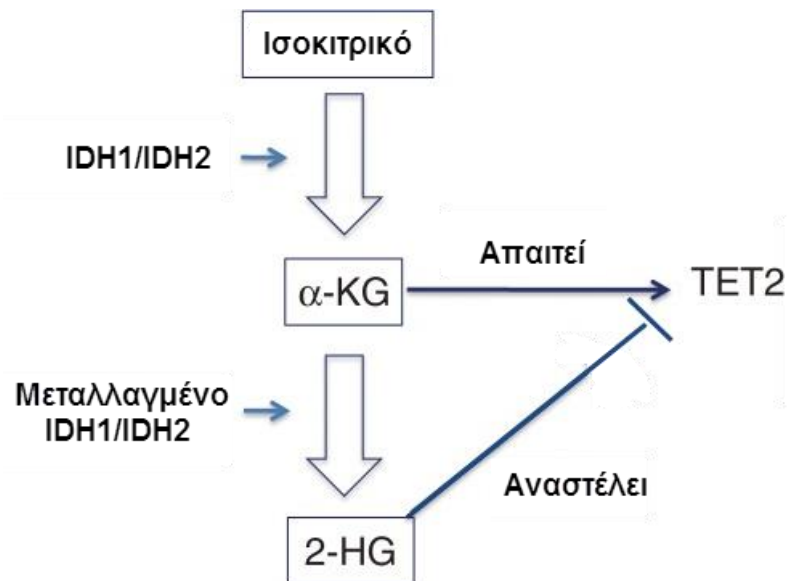
Η δράση του ενζύμου που κωδικοποιείται από το γονίδιο *IDH1* εντοπίζεται στο κυτταρόσωμο, ενώ η δράση του ενζύμου που κωδικοποιείται από το γονίδιο *IDH2* εντοπίζεται στα μιτοχόνδρια και τα υπεροξυσώματα.

Η αντίδραση που καταλύουν έχει σημαντικό ρόλο τόσο για τον κύκλο του κιτρικού οξέος (κύκλος Krebs) όσο και για την μεθυλίωση του DNA. Το α-Κετογλουταρικό (αKG) ή αλλιώς 2-οξογλουταρικό μαζί με οξυγόνο (O₂) και κατιόντα δισθενούς σιδήρου (Fe²⁺) είναι απαραίτητα για την λειτουργία του ενζύμου TET2.

Στο γονίδιο *IDH1*, συχνά παρατηρείται η αντικατάσταση μίας αργινίνης (R) στη θέση 132 της αμινοξικής ακολουθίας, από ιστιδίνη (R132H), ή από κυστεΐνη (R132C). Αντίστοιχα στο γονίδιο *IDH2*, συχνότερα ανιχνεύεται η μετάλλαξη R140Q (80% των περιπτώσεων), ενώ σπανιότερα η R172K. Όλες οι προαναφερθείσες μεταλλάξεις είναι μεταλλάξεις απόκτησης λειτουργίας (GOF).

Συνέπεια αυτών είναι η μετατροπή του α-KG προς 2-υδροξυ-γλουταρικό (2-HG), από τα ένζυμα *IDH1/2*. Οι δύο αυτές ενώσεις είναι δομικά ανάλογες, ενώ η μοναδική τους διαφορά, έγκειται στην αντικατάσταση της α-κετοομάδας του α-KG, από μια υδροξυλομάδα (-OH), οπότε και προκύπτει το 2-HG.

Η συσσώρευση του 2HG με την αντίστοιχη μείωση των επιπέδων του αKG επηρεάζει την λειτουργία του ενζύμου TET2, του οποίου η δράση εξαρτάται από το αKG και συγκεκριμένα εμποδίζει την σύνδεση του με τον δισθενή σίδηρο. Ως συνέπεια τα επίπεδα 5hmC μειώνονται, λόγω της μετατροπής 5mC προς 5hmC. Αυτό οδηγεί σε υπέρμετρη μεθυλίωση του DNA, κάτι που φαίνεται να επηρεάζει πληθώρα σημαντικών κυτταρικών λειτουργιών, όπως η διαφοροποίηση και ο πολλαπλασιασμός των κυττάρων. Οι λειτουργίες αυτές στη συγκεκριμένη περίπτωση επηρεάζονται πιθανότατα λόγω αποσιώπησης ογκοκατασταλτικών γονιδίων και γονιδίων που προάγουν την διαφοροποίηση των κυττάρων (Medeiros et al. 2017, Kaushansky et al. 2016).



Εικόνα 7. Σχηματική απεικόνιση της δράσης των φυσιολογικών και μεταλλαγμένων IDH1/2. (Τροποποίηση από Toshio Kitamura et al., 2014).

Μεταλλάξεις στα γονίδια *IDH1/2* ανιχνεύονται τόσο σε ασθενείς με ΟΜΛ, όσο και στις υπόλοιπες κατηγορίες μυελικών νεοπλασιών όπως ΜΔΣ, ΜΥΝ και αλληλεπικαλυπτόμενα σύνδρομα, σε διαφορετικές συχνότητες.

Οι μεταλλάξεις στα *IDH1/2* ανιχνεύονται σε ετερόζυγη κατάσταση, ενώ παράλληλα είναι αμοιβαία αποκλειόμενες με τις μεταλλάξεις *TET2* (Ward et al. 2010, Gaidzik et al. 2012).

1.2.1.4 Μεταλλάξεις σε γονίδια που συμμετέχουν στην τροποποίηση των ιστονών πρωτεϊνών

Η διαμόρφωση της χρωματίνης στο χώρο έχει σημαντικό ρόλο στην έκφραση των γονιδίων. Ο βαθμός συμπύκνωσης της καθορίζει εάν οι μεταγραφικοί παράγοντες και τα ένζυμα της μεταγραφής θα προσδεθούν στις κατάλληλες λειτουργικές και ρυθμιστικές περιοχές του γονιδιώματος και κατ' επέκταση εάν θα εκφραστούν συγκεκριμένα γονίδια (Kaushansky et al. 2016).

Δομικοί λίθοι της χρωματίνης είναι τα νουκλεοσώματα. Αυτά είναι οκταμερή μόρια ιστονών που αποτελούνται από 4 διπλές υπομονάδες πρωτεϊνών H2A, H2B, H3 και H4. Γύρω από αυτές τις δομές περιελίσσεται το ευκαρυωτικό DNA και συμπυκνώνεται στον πυρήνα.

Οι ιστόνες μπορούν να τροποποιηθούν με την προσθήκη πλευρικών ομάδων κυρίως στο καρβοξυτελικό τους άκρο. Συνδυασμός διαφορετικών τροποποιήσεων στις ιστόνες επιφέρει αλλαγές στην διάταξη της χρωματίνης και κατά συνέπεια αλλαγές στην έκφραση γονιδίων.

Τα πιο βασικά είδη τροποποιήσεων των ιστονών είναι:

- (απο)ακετυλίωση
- (απο)μεθυλίωση
- (απο)φωσφορυλίωση
- (απο)ουβικουΐτινυλίωση

Στις τροποποιήσεις αυτές συμμετέχουν αντίστοιχα διάφορα ένζυμα όπως ακετυλοτρανσφεράσες των ιστονών (histone acetyltransferases; HATs), αποακετυλάσες των ιστονών (histone deacetylases; HDACs), μεθυλοτρανσφεράσες των ιστονών (histone methyltransferases; HMTs), απομεθυλάσες των ιστονών (histone demethylases; HDMs),

κινάσες, φωσφατάσες, λιγάσες ουβικουτινυλίωσης και από-ουβικουτινιάσες των ιστονών (Verma M. et al. 2015).

Παραδείγματα τέτοιων τροποποιήσεων είναι η τριμεθυλίωση της λυσίνης 4 και 6 της H3 ιστόνης (H3K4me3, H3K6me3) που αποτελεί ένδειξη ενεργής γονιδιακής έκφρασης ενώ αντίθετα η δι- και τρι-μεθυλίωση της λυσίνης 9 της ιστόνης H3 (H3K9me2/3) όπως και η τριπλή μεθυλίωση της λυσίνης στη θέση 27 της ιστόνης 3 (H3K27me3) οδηγεί σε καταστολή της έκφρασης γονιδίων.

1.2.1.5 Τα γονίδια *ASXL1* και *EZH2*

Το γονίδιο *ASXL1* (Additional sex combs-like 1) εδράζεται στο γενετικό τόπο 20q11 και είναι ένα από τα 3 μέλη της οικογένειας γονιδίων *ASXL* (*ASXL1*, *ASXL2*, *ASXL3*) που ανιχνεύονται στα θηλαστικά. Τα γονίδια αυτά είναι ομόλογα των *Asx* (additional sex combs) γονιδίων της *Drosophila*. Οι πρωτεΐνες αυτές δρουν γενικά ως κριώματα για την προσέλκυση και συγκρότηση επιγενετικών ρυθμιστών και μεταγραφικών παραγόντων σε περιοχές του γονιδιώματος όπου παρατηρούνται συγκεκριμένα πρότυπα τροποποίησης των ιστονών (Kato M. et al. 2013). Η πρωτεΐνη *ASXL1* φαίνεται να παίζει σημαντικό ρόλο στην ρύθμιση της έκφρασης των ομοιοτικών γονιδίων, όπως τα γονίδια *Hox*, συμβάλλοντας στην ενεργοποίηση ή καταστολή τους ανάλογα με το στάδιο ανάπτυξης και τον ιστό που αυτή εκφράζεται (Inoue D. et al. 2013). Αυτό επιτυγχάνεται με τη συμμετοχή της πρωτεΐνης *ASXL1* στην συγκρότηση των πρωτεϊνικών συμπλόκων PRC (Polycomb-group repressor complex) και TRXG (trithorax-group, activator complex) και την αλληλεπίδραση αυτής με υποδοχείς ορμονών και άλλων μορίων, όπως επιγενετικούς ρυθμιστές του DNA και των ιστονών. Ταυτόχρονα διαθέτει την ικανότητα να προσδέεται άμεσα στην χρωματίνη μέσω της PHD (plant homeodomain) λειτουργικής/δομικής περιοχής του (domain) (Kato M. et al. 2013). Τα PRC και TRXG είναι συμπλέγματα πρωτεϊνών, υπεύθυνα για την διατήρηση της έκφρασης γονιδίων κατά την εμβρυογένεση και την ανάπτυξη αλλά και για την διατήρηση των επιγενετικών τροποποιήσεων. Τα PRC δρουν διατηρώντας την καταστολή στην έκφραση ενός γονιδίου, ενώ τα TRXG διατηρώντας την επαγωγή της μεταγραφής τροποποιώντας τις ιστόνες ανάλογα.

Χαρακτηριστικό παράδειγμα αποτελεί το σύμπλοκο PRC2 για την συγκρότηση του οποίου σημαντικό ρόλο κατέχει η πρωτεΐνη *ASXL1*. Το σύμπλοκο PRC2 αποτελείται από τις υπομονάδες *E(z)* (Enhancer of zeste), *Su(z)12* (Suppressor of zeste 12), *Esc* (Extra sex combs), *Jarid2* (jumonji, AT rich interactive domain 2) και *p55* οι οποίες κωδικοποιούνται αντίστοιχα από τα γονίδια *EZH1/2*, *SUZ12*, *EED*, *JARID2* και *RBAP46/48*. Το σύμπλοκο αυτό καθοδηγεί την τριπλή μεθυλίωση της λυσίνης στη θέση 27 της ιστόνης 3 (H3K27me3) μέσω της δράσης της *E(z)* υπομονάδας, ενώ η υπομονάδα *Su(z)12* ενισχύει την δράση της *E(z)*. Η τροποποίηση αυτή δρα κατασταλτικά επί της γονιδιακής μεταγραφής με την υπομονάδα *Esc* του PRC2 συμπλόκου να προσδέεται στην τριπλά μεθυλιωμένη ιστόνη και να ενισχύει την κατασταλτική δράση αυτής της τροποποίησης (Abdel-Wahab et al. 2012, Inoue et al. 2013, Blackledge et al. 2014, Greenblatt and Nimer 2014, Geisler and Paro 2015).

Σε διάφορους τύπους συμπαγών όγκων (καρκίνος προστάτη, στήθους, ουροδόχου κύστης), οι μεταλλάξεις *EZH2* έχει αποδειχθεί πως οδηγούν σε απόκτηση νέων λειτουργιών. Αντιθέτως, στα αιματολογικά νοσήματα, φαίνεται πως συντελούν σε απώλεια (LOF mutations) της ογκοκατασταλτικής δράσης της πρωτεΐνης.

Μάλιστα, αφορούν κυρίως σε μεταλλάξεις πρόωρου τερματισμού της μετάφρασης ή άλλες μεταλλάξεις που επηρεάζουν άμεσα τη λειτουργική περιοχή με δράση μεθυλομεταφοράσης (Ernst et al. 2010, You and Jones 2012, Rinke et al. 2017).



Εικόνα 8. Το σύμπλοκο PRC2. Τροποποίηση από Greenblatt S. M. and Nimer S. D., 2014.

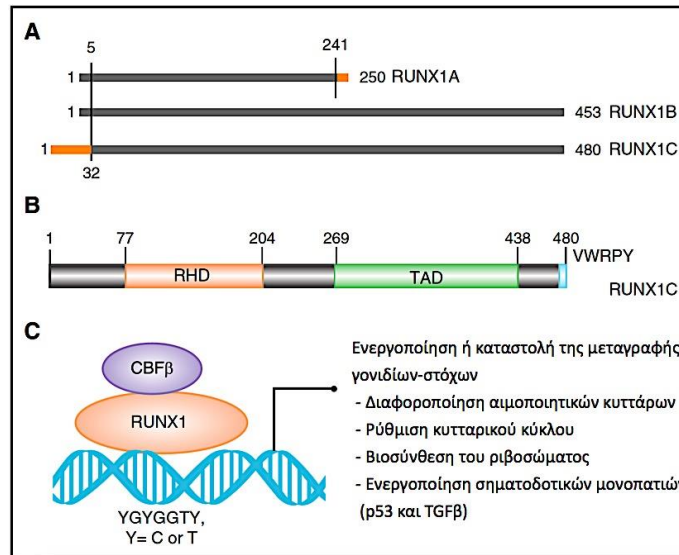
Μεταλλάξεις του γονιδίου *ASXL1* έχουν ανιχνευθεί σε ένα ευρύ φάσμα νεοπλασιών του μυελού των οστών όπως Μυελοϋπερπλαστικές Νεοπλασίες (ΜΥΝ), Χρόνια Μυελομονοκυτταρική Λευχαιμία (ΧΜΜΛ), Οξεία Μυελογενής Λευχαιμία (ΟΜΛ) και στα ΜΔΣ (Gelsi-Boyer V. et al. 2009, Boulwood J. et al. 2010). Οι μεταλλάξεις αυτές είναι κυρίως μεταλλάξεις μη νοηματικές-ανερμηνεύσιμες (nonsense) που οδηγούν σε πρόωρη λήξη και μετατόπισης του πλαισίου ανάγνωσης (frame shift) ενώ εντοπίζονται κυρίως στο καρβοξυτελικό άκρο του *ASXL1* και συγκεκριμένα στο τελευταίο εξώνιο (εξώνιο 12) κοντά στην περιοχή PHD (PHD domain). Ανιχνεύονται δε σχεδόν πάντα σε ετεροζυγωτία (Inoue et al. 2013).

1.2.1.6 Μεταλλάξεις σε γονίδια μεταγραφικών παραγόντων

1.2.1.6.1 Το γονίδιο *RUNX1*

Το *RUNX1*, αποτελεί το πρώτο μέλος μίας οικογένειας τριών γονιδίων, εδράζει στο χρωμόσωμα 21q22.12 και κωδικοποιεί για τρεις διαφορετικές ισομορφές πρωτεϊνών, οι οποίες προκύπτουν είτε εξαιτίας εναλλακτικής ωρίμανσης, είτε ως προϊόντα δύο ανεξάρτητων υποκινητών που απέχουν διαφορετική απόσταση από το γονίδιο.

Όλες οι ισομορφές ανεξαιρέτως περιλαμβάνουν μία συντηρημένη περιοχή 128 αμινοξέων (runt-homology domain, RHD), που χρησιμεύει για την πρόσδεση της πρωτεΐνης στο DNA, ενώ οι 1B και 1C, περιέχουν μία επιπλέον περιοχή (transactivating domain, TAD) (Εικόνα 9B). Η τελευταία είναι χρήσιμη για τη δέσμευση και αλληλεπίδραση με άλλες πρωτεΐνες, που οδηγούν σε ενεργοποίηση ή καταστολή της μεταγραφής γονιδίων-στόχων του μεταγραφικού παράγοντα *RUNX1* (Sood et al. 2017).



Εικόνα 9. Δράση των ισομορφών RUNX1 (τροποποίηση από Sood et al. 2017).

Οι μεταλλάξεις απώλειας λειτουργίας (LOF mutations), στο γονίδιο *RUNX1*, έχουν συσχετιστεί κυρίως με τα ΜΔΣ και με τη ΧΜΜΛ (37%), καθώς και με την εκτροπή αυτών προς ΟΜΛ.

Στις περιπτώσεις ασθενών με ΧΜΜΛ ανιχνεύονται μεταλλάξεις παρανοηματικές (missense mutations) στις περιοχές RHD ή TAD, ή ακόμη και μεταλλάξεις χωρίς νόημα (nonsense mutations) ή μετατόπισης του πλαισίου ανάγνωσης (frameshift mutations), τόσο στο αμινοτελικό (-NH₂), όσο και στο καρβοξυτελικό άκρο (-COOH) της πρωτεΐνης.

Σε ένα επίσης σημαντικό ποσοστό ασθενών με ΧΜΜΛ και μεταλλάξεις *RUNX1* παρατηρείται συχνότερα εκτροπή νόσου (53%), συγκριτικά με αντίστοιχους ασθενείς χωρίς μεταλλάξεις στο συγκεκριμένο γονίδιο (33%). Το γεγονός αυτό, ωστόσο, φαίνεται πως είναι ανεξάρτητο της επιβίωσης, η οποία δε διαφέρει ανάμεσα στις δύο κοορτές (Kuo et al. 2009, Tsai et al. 2015, Cai et al. 2016, Reinig et al. 2016).

1.2.1.7 Μεταλλάξεις σε γονίδια που συμμετέχουν στην ωρίμανση (μάτισμα) του RNA

Μάτισμα (splicing), ή αλλιώς ωρίμανση του πρόδρομου mRNA (pre-mRNA) στους ευκαρυωτικούς οργανισμούς, ονομάζουμε την διαδικασία αποκοπής των εσωνίων ενός πρόδρομου mRNA και την συρραφή των εξωνίων αυτού προς το σχηματισμό ώριμου μορίου mRNA.

Σε πολλές περιπτώσεις ανώτερων ευκαρυωτικών οργανισμών το μάτισμα ενός pre-mRNA μπορεί να δώσει πολλά διαφορετικά mRNAs μέσω του εναλλακτικού ματίσματος. Σε αυτή τη διαδικασία έχουμε την διαφορετική συρραφή των εξωνίων ενός pre-mRNA και άρα την παραγωγή διαφορετικών mRNA. Αποτέλεσμα του εναλλακτικού ματίσματος είναι η παραγωγή πολλών διαφορετικών ισομορφών πρωτεϊνών από ένα μόνο γονίδιο (Will CL. et al. 2011).

Υπεύθυνο για το μάτισμα του pre-mRNA είναι το σωματίο ματίσματος (spliceosome). Το σωματίο ματίσματος είναι ένα μεγαλομοριακό ριβονουκλεοπρωτεϊνικό (RNP) σύμπλοκο το οποίο απαρτίζεται από επιμέρους συμπλέγματα μικρών πυρηνικών RNA (small nuclear RNA, snRNA) και πρωτεϊνών, γνωστά ως μικρά ριβονουκλεοπρωτεϊνικά σωματίδια (snRNP), αλλά και τουλάχιστον άλλων 50 πρωτεϊνών. Όλα μαζί συγκεντρώνονται και συναρμολογούνται στο

σημείο συνένωσης εξωνίου-εσωνίου ενός pre-mRNA για να σχηματίσουν το σωματίο ματίσματος (Abdel-Wahab O. et al. 2011).

Στα περισσότερα ευκαρυωτικά κύτταρα έχουμε δύο διαφορετικά είδη σωματίων ματίσματος ανάλογα με την ειδικότητα που έχουν:

Το U2-εξαρτώμενο σωματίο, το οποίο καταλύει την αφαίρεση των εσωνίων τύπου U2 και αποτελείται από τα snRNPs U1, U2, U5, και U4/U6 αλλά και από πολυάριθμες non-snRNP πρωτεΐνες.

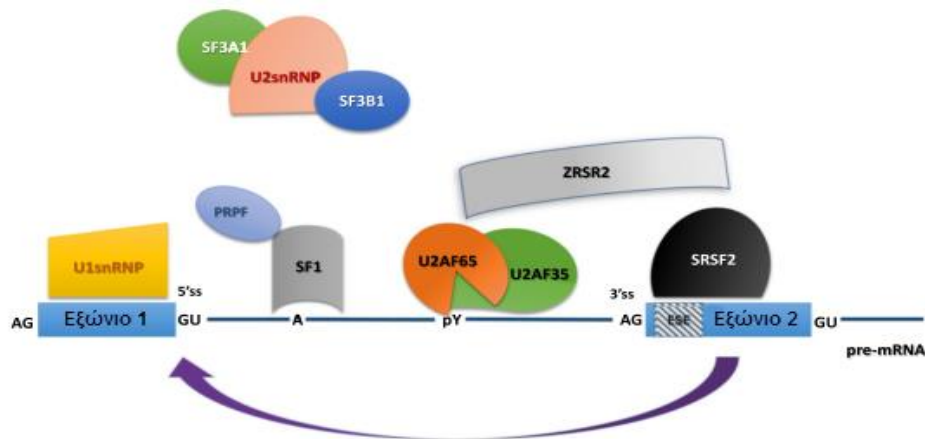
Το U12-εξαρτώμενο σωματίο, καταλύει την αφαίρεση των όχι τόσο συχνών εσωνίων τύπου U12 και αποτελείται από τα snRNPs U11, U12, U5, και U4atac/U6atac (Abdel-Wahab O. et al. 2011, Patel AA et al. 2003).

Το U2-εξαρτώμενο σωματίο καταλύει το μάτισμα σχεδόν του 99% των εσωνίων με το U12-εξαρτώμενο σωματίο να καταλύει το μάτισμα στο υπόλοιπο 1% των σπάνιων αυτών εσωνίων.

Σημαντικό ρόλο στην όλη διαδικασία έχουν πέρα από τα trans στοιχεία που συμμετέχουν στο μάτισμα και τα cis στοιχεία του pre-mRNA όπως, η 3' θέση ματίσματος του εσωνίου (3' splicing site; 3' ss), της οποίας χαρακτηριστικό είναι η νουκλεοτιδική ακολουθία AG στο σημείο ένωσης με το επόμενο εξώνιο, η 5' θέση ματίσματος (5' splicing site, 5' ss), της οποίας χαρακτηριστικό είναι η ακολουθία GU στο σημείο ένωσης με το προηγούμενο εξώνιο, η θέση διακλάδωσης (Branch site; BS) του εσωνίου, η οποία εντοπίζεται 18-40 νουκλεοτίδια πριν την 3' ss και περιέχει μια βάση αδενοσίνης υπεύθυνη για την πρώτη αντίδραση εστεροποίησης και την δημιουργία του ενδιάμεσου βρόγχου.

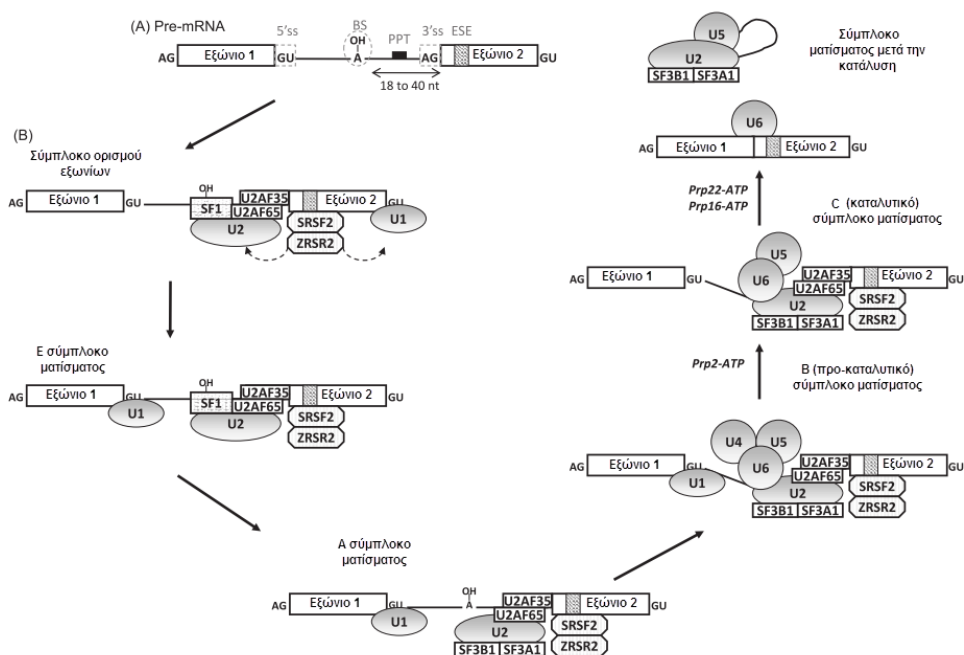
Ακολουθείται δε από μια πολυπυριμιδική αλληλουχία (polypyrimidine tract; PPT) περίπου 10 βάσεων και από αλληλουχίες εξωνικών και εσωνικών ενισχυτών, ή αποσιωπητών, του ματίσματος (exonic and intronic splicing enhancers; ESEs and ISEs, exonic and intronic splicing silencers; ESSs and ISSs) που είναι μικρές και ποικίλες αλληλουχίες που ρυθμίζουν το βασικό, ή εναλλακτικό μάτισμα, προσδένοντας ρυθμιστικές πρωτεΐνες (όπως αυτές της οικογένειας SR) (Will CL. et al. 2011, Smith CW et al. 2000).

Το σωματίο ματίσματος οργανώνεται στο πρώιμο-mRNA μέσω διαδοχικών προσδέσεων snRNPs και το μάτισμα του RNA ξεκινά με την αναγνώριση της 5' θέσης ματίσματος από το U1 snRNP και της 3' θέσης ματίσματος από τον βοηθητικό παράγοντα U2AF35 (U2-auxiliary factor 1) του ετεροδιμερούς βοηθητικού παράγοντα U2AF. Ο παράγοντας ματίσματος 1 (splicing factor 1, SF1) προσδένεται στο σημείο διακλάδωσης μέσω του SF3B1 και λειτουργεί ως γέφυρα μεταξύ της 5' και 3' θέσης ματίσματος προστατεύοντας την περιοχή του εσωνίου πριν την έναρξη του ματίσματος. Τέλος, το σύμπλοκο που σχηματίζεται από το U2AF, την πρωτεΐνη Zinc finger RNA binding motif and serine/arginine rich 2 (ZRSR2) και έναν από τους παράγοντες ματίσματος πλούσιους σε κατάλοιπα σερίνης/αργινίνης (serine/arginine rich splicing factors) SRSF1 ή SRSF2 προσδένεται στην πολυπυριμιδική αλληλουχία (βλ. Εικόνα 10) (Zoi and Cross 2015).



Εικόνα 10. Οργάνωση σωματίου ματίσματος (Τροποποίηση από Zoi and Cross 2015).

Τα εσώνια αφαιρούνται από το pre-mRNA μέσω δυο αντιδράσεων τρανσ-εστεροποίησης που καταλύονται από το σωματίο ματίσματος. Το σωματίο σχηματίζει σύμπλοκα πάνω στο pre-mRNA τα οποία με τη σειρά είναι τα E, A, B και C με το τελευταίο να είναι το σύμπλοκο το οποίο καταλύει τις δυο αντιδράσεις.



Εικόνα 11. Η διαδικασία της ωρίμανσης του RNA. Συντμήσεις: A, branch site, AG, splice acceptor site, ESE, exon splicing enhancer, GU, splice donor site, SF1, splicing factor 1, PPT, polypyrimidine tract, Prp2, Prp22 and Prp16 elicases, SF3A1, splicing factor 3a subunit 1, SF3B1, splicing factor 3b subunit 1, SRSF2, serine and arginine rich splicing factor 2, U1-U6, U1-U6 small nuclear ribonucleo proteins, U2AF35, U2 small nuclear RNA auxiliary factor 35, U2AF65, U2 small nuclear RNA auxiliary factor 65, YYY, polypyrimidine tract, ZRSR2, zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2 (Τροποποίηση από Chesnais et al. 2012).

1.2.1.7.1 Το γονίδιο *SF3B1*

Το γονίδιο *SF3B1* εδράζεται στο γενετικό τόπο 2q33 και κωδικοποιεί την υπομονάδα 1 του παράγοντα ματίσματος 3b (subunit 1 of the splicing factor 3b-SF3B1). Μαζί με τον παράγοντα

ματίσματος 3α (SF3a) και ένα 12S RNA, σχηματίζουν το snRNP U2 του σωματίου ματίσματος. Το σύμπλεγμα των παραγόντων SF3a και SF3b προσδένεται ανοδικά του σημείου διακλάδωσης (branch site) στο εσώνιο ενός pre-mRNA με τρόπο ανεξάρτητο της αλληλουχίας και έχει ως αποτέλεσμα την πρόσδεση του U2 snRNP στο pre-mRNA (Cazzola M. et al. 2013).

Μεταλλάξεις στο γονίδιο *SF3B1* είναι συνήθως παρανοηματικές σημειακές μεταλλάξεις (missense mutations) και εμφανίζονται κυρίως στο συντηρημένο καρβοξυτελικό άκρο της πρωτεΐνης, το οποίο αποτελείται από 22 Huntington, Elongation Factor 3, PR65/A, TOR (HEAT) επαναλήψεις από το εξώνιο 14 έως το 16. Η μετάλλαξη με τη μεγαλύτερη συχνότητα είναι η c.2098A>G στο εξώνιο 15 του γονιδίου, που οδηγεί σε πρωτεϊνικό επίπεδο στην αντικατάσταση της λυσίνης στην θέση 700 από γλουταμινικό οξύ (p.K700E), ενώ και άλλες μεταλλάξεις παρουσιάζονται με μικρότερη συχνότητα στο εξώνιο 14 όπως οι p.R625H/C/L, p.H662Q, p.H662Q/D, p.E622D, p.K666N/E/T/R/M/Q, p.N626Y, p.T663I, p.Y623C (Youzhong Wan 2013, Virginie C. et al. 2012). Η απουσία μεταλλάξεων μετατόπισης του πλαισίου ανάγνωσης (frameshift), πρόωρης λήξης (nonsense) καθώς και το ότι δεν εντοπίζονται σε σημεία άκρως σημαντικά για την λειτουργία της πρωτεΐνης αλλά και το γεγονός πως από in silico μελέτες φαίνεται να έχουν μικρότερη επίπτωση από άλλες τυχαίες μεταλλάξεις οδηγούν στο συμπέρασμα ότι η πρωτεΐνη διατηρεί την δομική της ακεραιότητα και μερική ή/και αλλαγμένη λειτουργικότητα.

Μεταλλάξεις στο γονίδιο *SF3B1* έχουν αναφερθεί σε ασθενείς με ΧΛΛ (5-18%), με ΜΔΣ (20%) (Paraemmanuil et al. 2011), καθώς και με ΠΜΙ (5%) ή δευτεροπαθή ΜΙ από ΙΘ (8%).

Ωστόσο, συχνότερα απαντώνται σε ασθενείς με ΜΔΣ-ΔΣ (60-80%) (Paraemmanuil et al. 2011, Chesnais et al. 2012, Wan and Wu 2013, Yoshimi and Abdel-Wahab 2017), με συνέπεια οι μεταλλάξεις του γονιδίου να συσχετίζονται σημαντικά με αυτόν τον παθολογικό φαινότυπο και πλέον τα *SF3B1* θετικά ΜΔΣ-ΔΣ δύνανται να θεωρηθούν ως ξεχωριστή κλινική οντότητα (Harada and Harada 2015).

Σε παρόμοια ποσοστά (>80%) ανιχνεύονται μεταλλάξεις *SF3B1* και στα ΜΔΣ/ΜΥΝ-ΔΣ-Θ νεοπλάσματα. Σε αυτά, η συνανίχνευση της JAK2V617F με *SF3B1* μεταλλάξεις, κυρίως την K700E και άλλες, αγγίζει περίπου το 64% των περιπτώσεων (Broséus et al. 2013, Jeromin et al. 2013).

1.2.1.7.2 Το γονίδιο *U2AF1*

Το γονίδιο *U2AF1* εντοπίζεται στο γενετικό τόπο 21q22 και κωδικοποιεί την μικρή υπομονάδα του U2AF (U2 auxiliary factor), μεγέθους 35kDalton γνωστή ως U2AF1 ή U2AF35. Το U2AF35 μαζί με την μεγάλη υπομονάδα U2AF65 σχηματίζουν το ετεροδιμερές U2AF. Το U2AF έχει σημαντικό ρόλο στην οργάνωση του σωματίου ματίσματος και στην διαδικασία του ματίσματος αναγνωρίζοντας σημαντικά cis στοιχεία για το μάτισμα του pre-mRNA. Η υπομονάδα U2AF65 αναγνωρίζει και προσδένεται στην πολυπυριμιδική αλληλουχία (PPT) του εσωνίου, ενώ το U2AF35 αναγνωρίζει την και προσδένεται στην 3' ss στα αρχικά στάδια οργάνωσης του σωματίου ματίσματος (Wu S. et al. 1999). Επίσης φαίνεται να παίζει σημαντικό ρόλο τόσο στο κύριο όσο και στο εναλλακτικό μάτισμα μεσολαβώντας σε αλληλεπιδράσεις πρωτεΐνης-πρωτεΐνης και πρωτεΐνης-RNA για την αναγνώριση της 3' ss περιοχής αλλά και λειτουργώντας ως διαμεσολαβητής μεταξύ του U2AF65 και παραγόντων του ματίσματος όπως των πρωτεϊνών SR (serine/arginine rich). Αυτό επιτυγχάνεται με δυο λειτουργικές δομικές περιοχές αλληλεπίδρασης πρωτεΐνη με πρωτεΐνη του U2AF35, η μία στο πλούσιο σε κατάλοιπα σερίνης αργινίνης (SR rich) αμινοτελικό του άκρο, το οποίο

χρησιμεύει για την αλληλεπίδραση του με πρωτεΐνες SR και η άλλη στο καρβοξυτελικό του άκρο που επιτρέπει την σχηματισμό του ετεροδιμερούς συμπλόκου U2AF35/U2AF65 (Zuo P. et al. 1996).

Μεταλλάξεις στο γονίδιο U2AF1 έχουν βρεθεί σε ασθενείς με Χρόνια μυελομονοκυτταρική λευχαιμία (ΧΜΜΛ), ΟΜΛ, ΜΥΝ και ΜΔΣ, ενώ στην τελευταία κατηγορία οι μεταλλάξεις συσχετίζονται με κακή πρόγνωση και πιθανότητα εκτροπής της νόσου σε ΟΜΛ (Yoshida K. et al 2011, Przychodzen B. et al. 2013, Graubert TA et al. 2012).

Αυτές οι μεταλλάξεις του U2AF35 έχουν ανιχνευθεί κυρίως σε δυο περιοχές με δομή δακτύλου ψευδαργύρου (Zinc fingers) σημαντικές για την πρόσδεση του στην 3' ss. Η πρώτη είναι παρανοηματική μετάλλαξη που αλλάζει το συντηρημένο κατάλοιπο σερίνης στην θέση 34 της πρωτεΐνης (S34) στο πρώτο δάκτυλο ψευδαργύρου, ενώ η δεύτερη εντοπίζεται στα κατάλοιπα R156, Q157 του δεύτερου δακτύλου (Graubert TA et al. 2012).

Αποτέλεσμα αυτών μεταλλάξεων είναι η αλλαγή του μηχανισμού ματίσματος και η παράλειψη εξωνίων ή η μειωμένη έκφραση γονιδίων λόγω μη ματίσματος του pre-mRNA. Πρόσφατα δεδομένα δείχνουν πως πολλά από αυτά τα γονίδια συσχετίζονται με τον καρκίνο (Przychodzen B. et al. 2013). Αξίζει να σημειωθεί πως οι μεταλλάξεις αυτές εμφανίζονται κυρίως σε ετεροζυγωτία και πως δεν συνδυάζονται με άλλες μεταλλάξεις γονιδίων του σωματίου ματίσματος καθώς φαίνεται πως κάτι τέτοιο οδηγεί στον κυτταρικό θάνατο (Makishima H. et al. 2012).

1.2.1.7.3 Το γονίδιο *SRSF2*

Το γονίδιο *SRSF2* εντοπίζεται στο γενετικό τόπο 17q25 και κωδικοποιεί μια πρωτεΐνη 25,5-kDa, η οποία ανήκει στην οικογένεια των πρωτεϊνών SR (serine/arginine rich). Οι πρωτεΐνες αυτές έχουν ως κύριο ρόλο την αναγνώριση των εξωνίων από το σωματίο ματίσματος μέσω πρόσδεσης τους σε ρυθμιστικές περιοχές ματίσματος του RNA.

Η πρωτεΐνη *SRSF2* προσδένεται στις αλληλουχίες ESE (exonic splicing enhancers) και δρα ουσιαστικά ως ενισχυτής για την αναγνώριση της 3' περιοχής ματίσματος από την υπομονάδα U2AF1 και βοηθά στην πρόσδεση της στο δινουκλεοτίδιο AG.

Μεταλλάξεις στο γονίδιο *SRSF2* στα ΜΔΣ είναι οι δεύτερες πιο συχνές μοριακές βλάβες στα γονίδια του σωματίου ματίσματος, εντοπίζονται δε σε αρκετές κατηγορίες ΜΔΣ όπως ΧΜΜΛ, ΟΜΛ και ΜΔΣ/ΜΥΝ-ΔΣ-Θ νεοπλάσματα.

Οι μεταλλάξεις αυτές απαντώνται σχεδόν πάντα σε ετερόζυγη κατάσταση. Συνήθως αφορούν σε παρανοηματικές μεταλλάξεις, στο αμινοξύ προλίνη, στην θέση 95 της πρωτεΐνης και οδηγούν σε συσσώρευση βλαβών στο DNA και αποδιοργάνωση του. Συνύπαρξη μεταλλάξεων στο γονίδιο *SRSF2* έχει παρατηρηθεί τόσο με το γονίδιο *IDH1*, όσο και με το γονίδιο *RUNX1* (Larsson et al. 2013).

1.2.1.7.4 Το γονίδιο *ZRSR2*

Το γονίδιο *ZRSR2* εντοπίζεται επί του χρωμοσώματος Χ στη θέση p22.2 και απαρτίζεται από 14 εξώνια.

Ο παράγοντας *ZRSR2* που κωδικοποιείται από το ομώνυμο γονίδιο, συμμετέχει στη διαδικασία ωρίμανσης του RNA, μέσω αναγνώρισης της 3'-ss. Αλληλεπιδρά με άλλα στοιχεία του πρόδρομου σωματίου ωρίμανσης, όπως το ετεροδιμερές U2AF2/U2AF1 και η πρωτεΐνη *SRSF2*, ενώ φαίνεται πως είναι απαραίτητος για το σωστό μάτισμα και των δύο τύπων εσωνίων (εσώνια τύπου U2 και U12) (Madan et al. 2015).

Σε όλη την έκταση της κωδικής περιοχής, καθώς και στις θέσεις ματίσματος του γονιδίου ZRSR2, ανιχνεύονται μεταλλάξεις παρανοηματικές, χωρίς νόημα και μετατόπισης του πλαισίου ανάγνωσης, οι οποίες οδηγούν σε απώλεια λειτουργίας (LOF mutations) (Hong et al. 2015, Saez et al. 2017).

Οι μεταλλάξεις αυτές επηρεάζουν την ωρίμανση των εσωνίων τύπου U12 (U12-type intron) κυρίως, οδηγώντας σε διατήρηση εσωνικών περιοχών στο ώριμο mRNA και κατά συνέπεια σε διαφορετικό πρωτεϊνικό προϊόν (Saez et al. 2017).

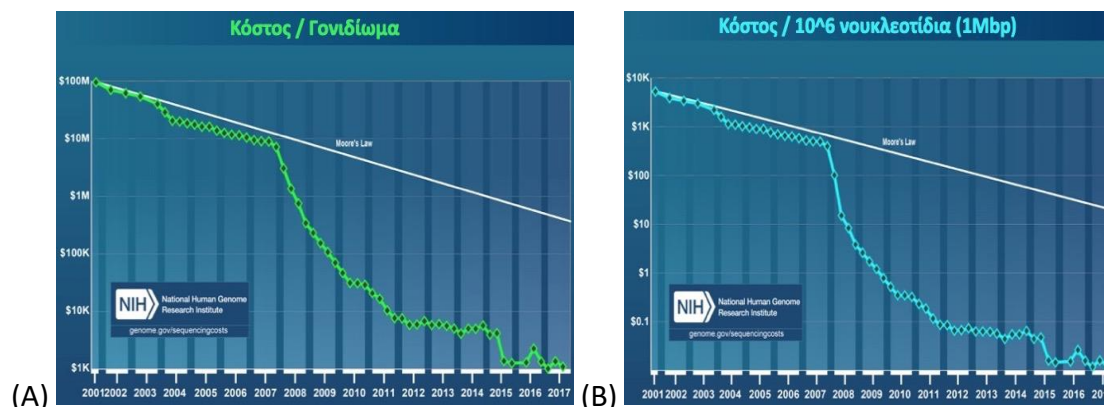
1.3 Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με τεχνολογία επόμενης (δεύτερης) γενιάς (Next Generation Sequencing, NGS)

1.3.1 Εισαγωγικά στοιχεία

Τριάντα περίπου χρόνια μετά την ανακάλυψη της μεθόδου προσδιορισμού της πρωτοδιάταξης των νουκλεϊκών οξέων με χρήση δι-δεοξυ-νουκλεοτιδίων ddNTPs, πραγματοποιήθηκε μια εκ νέου σημαντική πρόοδος στο συγκεκριμένο τομέα, με την εμφάνιση εξελιγμένων αναλυτικών συσκευών.

Ο προσδιορισμός της πρωτοδιάταξης νουκλεϊκών οξέων με τεχνολογία επόμενης γενιάς (Next Generation Sequencing, NGS), μείωσε δραματικά το κόστος και το χρόνο μελέτης του γονιδιώματος, αφού επέτρεψε την ταυτόχρονη διερεύνηση περισσότερων γενομικών περιοχών, για μεγαλύτερο αριθμό δειγμάτων, ταυτόχρονα σε μία αντίδραση.

Παράλληλα, η τεχνολογία NGS, παρέχει ποσοτικά στοιχεία σχετικά με το ποσοστό των μεταλλαγμένων αλληλίων σε ένα δείγμα (VAF, variant allele frequency ή AF, allele frequency), συγκρίνοντας το πλήθος των αλυσίδων, κατά συνέπεια τον αριθμό των κυττάρων, που φέρουν μετάλλαξη, σε σχέση με τις αντίστοιχες φυσιολογικές αλυσίδες (Luthra et al. 2015)



Εικόνα 12. Οικονομικοτεχνική ανάλυση της τεχνολογίας NGS. (A) Κόστος χαρτογράφησης ενός γονιδιώματος, ανά έτος και (B) κόστος προσδιορισμού της πρωτοδιάταξης νουκλεϊκών οξέων ανά μεγαβάση (1 Megabase (1 Mbp)), ανά έτος (Wetterstrand 2017).

Ωστόσο, σε συνδυασμό με τα θετικά στοιχεία της νέας τεχνολογίας, παράλληλα, επήλθαν καινούργιες τεχνικές προκλήσεις, ενώ άλλες προϋπάρχουσες μεγιστοποιήθηκαν.

Δύο μειονεκτήματα των νέων συσκευών, αφορούν στις σχετικά μικρότερου μήκους νουκλεοτιδικές αλληλουχίες που μπορούν να διαβάσουν (read length) (35-700 bp), όπως και

στο σχετικά υψηλότερο ποσοστό λάθους, κατά το διάβασμα των επί μέρους νουκλεοτιδίων (error rate) (0.1-2.0%), συγκριτικά με τον προσδιορισμό της πρωτοδιάταξης νουκλεϊκών οξέων κατά Sanger (έως 900 bp μήκος διαβάσματος και λιγότερο από 0.001% ποσοστό λάθους).

Εν τω μεταξύ, σημαντική βελτίωση, στην αξιοπιστία (confidence) των αποτελεσμάτων, έχει σημειωθεί με την αυτοματοποίηση των σταδίων προετοιμασίας των βιβλιοθηκών, μέσω χρήσης ρομποτικών συστημάτων, γεγονός που συνέβαλε σε μειωμένη διακύμανση μεταξύ των πειραματικών διαδικασιών, αλλά και με τη βελτίωση της χημείας που χρησιμοποιεί κάθε τεχνολογία, οδηγώντας έτσι σε αύξηση του συνολικού αριθμού διαβασμάτων και ταυτόχρονα σε αυξημένο βάθος διαβάσματος (read depth) κάθε νουκλεοτιδικής θέσης. Οι παραπάνω, καθώς και επιπρόσθετες λιγότερο σημαντικές καινοτομίες, συνετέλεσαν σε ενισχυμένη ακρίβεια (accuracy), επαναληψιμότητα (precision), ειδικότητα (specificity) και ευαισθησία (sensitivity) της μεθόδου, ενώ παράλληλα μείωσαν τα όρια ανίχνευσης μεταλλάξεων (limit of detection, LOD).

Παράλληλα, προς την ίδια κατεύθυνση συνεισέφερε και η ανάπτυξη περισσότερο εξελιγμένων αλγόριθμων βιοπληροφορικής ανάλυσης των δεδομένων. Η απουσία συγκεκριμένης αναλυτικής μεθοδολογίας, καθιστά μονόδρομο τη χρήση πακέτων βιοπληροφορικών αναλύσεων. Έτσι, η αναζήτηση για ακόμη πιο εξειδικευμένο προσωπικό σε συνδυασμό με την ανάγκη για αυξημένη υπολογιστική ισχύ, καθώς και μεγαλύτερους αποθηκευτικούς χώρους, λόγω του τεράστιου όγκου δεδομένων (big data), που απορρέουν από αυτές τις τεχνολογίες, αυξάνει εν μέρη το κόστος χρήσης της νέας τεχνολογίας (Luthra et al. 2015, Park and Kim 2016, Wartman and Mardis 2016).

1.3.2 Στοχευμένος προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων (Targeted-Sequencing) με τεχνολογία NGS

Αρχικά, οι πλατφόρμες NGS χρησιμοποιήθηκαν για την ανάλυση οργανισμών με σχετικά μικρό μέγεθος γονιδιώματος. Καθώς η τεχνολογία εξελίχθηκε, κατέστη εφικτή η ανάλυση διαρκώς μεγαλύτερων γονιδιωμάτων. Ωστόσο, το αυξημένο μέγεθος και η πολυπλοκότητα που χαρακτηρίζει τα πιο σύνθετα γονιδιώματα, όπως το ανθρώπινο, δυσχεραίνει την ερμηνεία των αποτελεσμάτων.

Οι περιορισμοί που προκύπτουν, από τη χρήση της υπάρχουσας τεχνολογίας, για τη μελέτη ολόκληρων ανθρώπινων γονιδιωμάτων (Whole Genome Sequencing, WGS), αφορούν κυρίως στο πεπερασμένο εύρος γενομικών περιοχών που αναλύονται επαρκώς, σε ικανοποιητικό βάθος διαβάσματος, όπως επίσης στο είδος των μεταλλάξεων που είναι εφικτό να ανιχνευτούν αξιόπιστα, με τις υπάρχουσες μεθόδους βιοπληροφορικής ανάλυσης.

Μεταλλάξεις μικρής κλίμακας, όπως SNVs (single nucleotide variants), MNVs (multi nucleotide variants) και περιορισμένες σε μέγεθος συνθετες προσθήκες ή/και διαγραφές (Indels), μπορούν εύκολα να αναδειχθούν, σε αντίθεση με μεγαλύτερες χρωμοσωμικές αλλοιώσεις και με CNVs, (copy number variations), συμβάντα που γίνονται πιο δύσκολα ανιχνεύσιμα.

Τα παραπάνω, σε συνδυασμό με το υψηλότερο κόστος, που προκύπτει εξαιτίας της ανάγκης για περισσότερα διαβάσματα, οδήγησαν στην ανάπτυξη προσεγγίσεων, που προσβλέπουν στο χαρακτηρισμό μόνον του συνόλου των εξωνικών περιοχών (Whole Exome Sequencing, WES), ή στην ανάλυση ακόμη πιο περιορισμένων, προκαθορισμένων, γενομικών περιοχών (Targeted DNA Resequencing).

Δύο κυρίως μεθοδολογίες έχουν αναπτυχθεί προς αυτή την κατεύθυνση.

1) Σε μία περίπτωση χρησιμοποιούνται τεχνητοί εκκινήτες (primers), που στοχεύουν στην ενζυμική επέκταση των περιοχών ενδιαφέροντος προτού αυτές αναλυθούν (amplicon sequencing).

2) Η άλλη τεχνική αφορά στην υβριδοποίηση λόγω συμπληρωματικότητας, τμήματος των περιοχών ενδιαφέροντος (Hybridization Capture), με συνθετικά ολιγονουκλεοτίδια, γνωστά ως ιχνηθέτες (probes). Οι ιχνηθέτες είναι τροποποιημένοι, ώστε φέρουν μία ομάδα βιοτίνης, μέσω της οποίας προσδένονται σε μαγνητικά σφαιρίδια με επικάλυψη στρεπταβιδίνης, επιτρέποντας την διατήρηση μόνο των περιοχών ενδιαφέροντος, καθώς το υπόλοιπο γονιδίωμα απορρίπτεται και αποκλείεται από την ανάλυση (Luthra et al. 2015, Wartman and Mardis 2016).

1.3.3 Εφαρμογές στην κλινική πράξη

Η τεχνολογία NGS προσφέρει τη δυνατότητα για την ανάπτυξη ιδιαίτερα εξειδικευμένων διαγνωστικών εξετάσεων, με την προοπτική της στοχευμένης θεραπευτικής παρέμβασης.

Η χρήση της, ωστόσο, οφείλει να ακολουθεί υψηλές προδιαγραφές, όπως εκείνες που τηρούν τα διαγνωστικά εργαστήρια σχετικά με τις συνθήκες εκτέλεσης των πειραματικών διαδικασιών, τις συνθήκες φύλαξης των κλινικών δειγμάτων, τις τεχνικές απομόνωσης νουκλεϊκών οξέων και την ανάγκη για συγκατάθεση του ασθενούς, για την πραγματοποίηση των εξετάσεων.

Εφόσον τα παραπάνω κριτήρια πληρούνται, τότε οι αναλύσεις NGS μπορούν να χρησιμεύσουν ιδιαίτερα, σε διάφορες πτυχές της κλινικής πράξης, εν προκειμένω σε ασθενείς με αιματολογικές κακοήθειες, όπως περιγράφεται στη συνέχεια:

1) Ταυτόχρονη διερεύνηση περισσότερων γενετικών τόπων, σε αντίθεση με παλαιότερες τεχνολογίες που παρείχαν δεδομένα για επιμέρους γενετικούς στόχους. Με τον τρόπο αυτό αυξάνεται η πιθανότητα ανάδειξης δεδομένων, που μπορεί να αποδειχθούν χρήσιμα για τον ασθενή, μάλιστα, σε μικρότερη χρονική διάρκεια από διάφορες συμβατικές τεχνικές.

2) Παραγωγή σημαντικών δεδομένων, σχετικά με την πολυπλοκότητα, σε επίπεδο μεταλλάξεων, που παρουσιάζουν οι επιμέρους παθολογικοί πληθυσμοί κλωνικών κυττάρων.

Έχει αποδειχθεί πως οι αιματολογικοί ασθενείς διακρίνονται από έναν κυτταρικό πληθυσμό, με υπαιτιότητα του οποίου εκδηλώνεται η διαταραχή, ωστόσο ενδέχεται να αναπτύξουν δευτερεύοντες κυτταρικούς υποπληθυσμούς, που έχουν εξελιχθεί από τον αρχικό, έπειτα από απόκτηση επιπρόσθετων μεταλλάξεων, οι οποίοι συμβάλλουν στην πρόοδο της νόσου και σε αλλαγές στα κλινικά χαρακτηριστικά. Το φαινόμενο αυτό χαρακτηρίζεται ως γενετική ετερογένεια (genomic heterogeneity) και μπορεί να γίνει αντιληπτό με διαδοχικούς μοριακούς ελέγχους μέσω NGS, σε δείγματα του ίδιου ασθενούς που προέρχονται από διαφορετικά χρονικά σημεία κατά την εξέλιξη της νόσου ή τη θεραπευτική αντιμετώπιση αυτής.

Η στοχευμένη θεραπευτική αντιμετώπιση ενός δευτερεύοντος κλωνικού κυτταρικού υποπληθυσμού, που φέρει μία συγκεκριμένη μετάλλαξη, μπορεί να οδηγήσει σε αριθμητική ελάττωση ή και εξάλειψη των κυττάρων αυτών, θα αφήσει όμως ανεπηρέαστο τον κλώνο κυττάρων που ευθύνεται, κατά βάση, για την εκδήλωση της εκάστοτε αιματολογικής διαταραχής και που διακρίνεται από επιπλέον μεταλλάξεις, οι οποίες εμπλέκουν γονίδια που συμμετέχουν σε διαφορετικές κυτταρικές λειτουργίες.

Αυτές οι πληροφορίες είναι ιδιαίτερα σημαντικές για την κλινική πράξη αφού καθορίζουν το θεραπευτικό σχήμα που πρέπει να χορηγηθεί σε κάθε περίπτωση, ώστε να επιτευχθεί ύφεση ή να αποτραπεί υποτροπή της νόσου.

3) Σαφέστατα πιο ευαίσθητη αναζήτηση της ελάχιστης υπολειπόμενης νόσου (minimal residual disease, MRD). Με τον όρο MRD περιγράφεται η παρουσία, ανθεκτικών στη θεραπεία, παθολογικών κυττάρων, που παραμένουν κατά την ύφεση, ενώ παράλληλα, είναι μη ανιχνεύσιμα με συμβατικές κλινικές μεθόδους.

4) Μελέτες φαρμακοτοξικότητας και μεταβολισμού των φαρμάκων με βάση το γενετικό προφίλ του ασθενούς, αναδεικνύουν μία ακόμη χρησιμότητα των νέων τεχνολογιών (Wartman and Mardis 2016).

2 ΣΚΟΠΟΣ

Σκοπό της παρούσας Διπλωματικής Εργασίας αποτέλεσε η μελέτη του Μοριακού τοπίου των ασθενών με μυελικές νεοπλασίες, μέσω στοχευμένου προσδιορισμού της νουκλεοτιδικής ακολουθίας συχνά μεταλλαγμένων γονιδίων, με χρήση τεχνολογίας δεύτερης γενιάς (targeted panel Next Generation Sequencing) και η ανάπτυξη συμβατής βιοπληροφορικής ροής για την ανάλυση των δεδομένων.

Η αναλυτική αυτή ροή θα μπορούσε να συνεισφέρει στην εύκολη, γρήγορη και αξιόπιστη Μοριακή τυποποίηση των ασθενών με μυελικές νεοπλασίες, συμβάλλοντας έτσι στην ορθή τους διάγνωση και ταξινόμηση, αλλά και σε πρόγνωση και διαστρωμάτωση του κινδύνου εκτροπής αυτών προς ΟΜΛ, ενώ ιδανικά, δύναται να παρέχει πληροφορίες σχετικά με τις διαθέσιμες θεραπευτικές επιλογές.

3 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

3.1 Αντιδραστήρια

3.1.1 Χημικά αντιδραστήρια, βιολογικά αντιδραστήρια και ολοκληρωμένα συστήματα αντιδραστηρίων

Εταιρεία	Κωδικός	Αντιδραστήριο
DEMO	4328	Water Vial
VWR Int. GmbH	20821	Ehtanol Absolute
Panreac-Applichem	CR131090	Propanol-2 Molecular Biology Grade/5Lt
Qiagen	51206	FlexiGene blood genomic DNA isolation kit
Macherey-Nagel	744970.5	NucleoMag NGS Clean-up and Size Select (150-800 bp)
Biotium	31006-T	AccuBlue High Sensitivity dsDNA Quantitation Kit (0,2-100 ng)
Agilent Tech.	5067-4626	Agilent High Sensitivity Kit
NEB	M0348L	dsDNA Fragmentase (250 Rxns)
Macrogen Inc.	-	Oligo synthesis (50 nmol), MOPC purified (up to 30 bases)
Bioline	BIO-21111	MyTaq HS DNA Polymerase
Biotium	31000	Evagreen Dye, 20X in water
Thermo-Fischer	4316357	3100 POP-6 TM Performance Optimized Polymer
Thermo-Fischer	4337450	ABIPrism BigDye Terminator v1.1
Thermo-Fischer	402824	Genetic Analyzer 10x Running Buffer with EDTA
NuGen	-	Ovation Target Enrichment custom panel
Illumina	MS-102-2003	MiSeq Reagent Kit v2 (500 cycles) (15M reads)

Πίνακας 5. Χημικά αντιδραστήρια, βιολογικά αντιδραστήρια και ολοκληρωμένα συστήματα αντιδραστηρίων.

3.1.2 Εργαστηριακός αναλυτικός εξοπλισμός

Εταιρεία	Συσκευή	Λειτουργία
Roche	LightCycler® 480	Real Time PCR, HRMA
Bio-Rad	T100 Thermal Cycler for PCR	PCR
Grant	Grant Sub waterbath	Υδατόλουτρο
Thermo-Fischer	Eppendorf centrifuge 5804	Φυγόκεντρος σωληναρίων τύπου Falcon
Thermo-Fischer	Biofuge pico Heraeus	Φυγόκεντρος σωληναρίων τύπου Eppendorf
Merck	Eppendorf concentrator 5301	Περιστροφική εξάτμιση διαλυμάτων υπό κενό
Illumina	MiSeq System	Σύστημα προσδιορισμού αλληλουχίας
Agilent Genomics	2100 Bioanalyzer	Ανάλυση μεγέθους νουκλεϊκών οξέων
Thermo-Fischer	Qubit Fluorometer	Ποσοτικός προσδιορισμός DNA
Thermo-Fischer	NanoDrop	Ποσοτικός προσδιορισμός ολικού RNA
Thermo-Fischer	Biomate3 Spectrometer	UV-Vis Ποσοτικός προσδιορισμός ολικού DNA

Πίνακας 6. Εργαστηριακός αναλυτικός εξοπλισμός

3.2 Κλινικό Υλικό

Ο στοχευμένος προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με αναλυτικές συσκευές δεύτερης γενιάς (targeted Next Generation Sequencing, targeted NGS), πραγματοποιήθηκε σε 51 ασθενείς εκ των οποίων:

- 8 ασθενείς με ΟΜΛ
- 9 ασθενείς με ΜΔΣ
- 33 ασθενείς με ΜΥΝ
 - 1 ασθενής με ΧΟΛ
 - 5 ασθενείς με ΙΘ
 - 11 ασθενείς με ΜΙ
 - 3 ασθενείς με ΣΜ
 - 5 ασθενείς με ΑΠ
 - 8 ασθενείς με αΜΥΝ
- 1 ασθενής με ΜΔΣ/ΜΥΝ (ΧΜΜΛ)

3.2.1 Προέλευση Δειγμάτων Ασθενών

Οι ασθενείς που μελετήθηκαν στην παρούσα εργασία προέρχονται από συνεργαζόμενες Αιματολογικές Κλινικές και Τμήματα Νοσοκομείων της Αττικής και της Περιφέρειας, συγκεκριμένα, από την Αιματολογική Κλινική της Ευρωκλινικής Αθηνών, την Αιματολογική Κλινική και το Αιματολογικό Τμήμα της Κλινικής Παθολογικής Φυσιολογίας του ΠΓΝ Αθηνών "Λαϊκό", το Αιματολογικό Εργαστήριο της Υπηρεσίας Αιμοδοσίας του ΠΝ "Αρεταίειο", την Αιματολογική Κλινική του ΠΓΝ Λάρισας, την Αιματολογική Κλινική του ΠΓΝ Αλεξανδρούπολης, την Αιματολογική Κλινική του ΠΓΝ Ιωαννίνων, την Αιματολογική Κλινική του ΠΓΝ "Αττικόν", την Αιματολογική Κλινική του ΓΝ Αθηνών "Γ. Γεννηματάς" και την Αιματολογική Κλινική του ΓΝ Αθηνών "Ο Ευαγγελισμός".

Η διαχείριση των δειγμάτων έγινε σύμφωνα με τη Διακήρυξη του Ελσίνκι. Κατόπιν ενημέρωσης από τον θεράποντα ιατρό, ζητήθηκε και ελήφθη, έγγραφη συγκατάθεση, από όλους τους ασθενείς που μελετήθηκαν.

3.3 Βασικές Αρχές Τεχνικών

3.3.1 Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με τεχνολογία επόμενης (δεύτερης) γενιάς (Next Generation Sequencing, NGS)

3.3.1.1 Αρχή λειτουργίας της μεθόδου

Η τεχνολογία που χρησιμοποιήθηκε στην παρούσα εργασία, αναπτύχθηκε το 2005 από την εταιρεία Solexa (ιδρύθηκε το 1998 από τους Balasubramanian και Klenerman), η οποία, την επόμενη χρονιά, εξαγοράστηκε από την Illumina, οπότε και έγινε ευρύτερα γνωστή και εμπορικά διαθέσιμη (Shendure et al. 2017).

Η εταιρεία Illumina παρέχει ένα ευρύ φάσμα συσκευών, με δυνατότητα κυμαινόμενου αριθμού διαβασμάτων (reads), από περίπου 25×10^6 έως 3×10^9 , τα οποία χρησιμεύουν στον ταυτόχρονο προσδιορισμό της νουκλεοτιδικής ακολουθίας πολλαπλών γενετικών στόχων

ανά αντίδραση, ενδεικτικά από 0.5×10^9 (gigabase, Gb), έως πάνω από 1.5×10^{12} (terabase, Tb) βάσεις.

Η κάθε αντίδραση διαρκεί από λίγες ώρες, έως μερικές ημέρες, γεγονός που εξαρτάται από το πλήθος των διαβασμάτων και κατά συνέπεια των δεδομένων που εξάγονται σε κάθε περίπτωση (data output) (Wartman and Mardis 2016).

Οι συσκευές Illumina βασίζονται στην τεχνολογία που είναι ευρύτερα γνωστή ως προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων, μέσω σύνθεσης νέων προϊόντων, διαμέσου προσθήκης μονο-νουκλεοτιδίων, με την τεχνολογία του αντιστρεπτού τερματισμού της αντιγραφής (SBS/CRT, Sequencing by Synthesis/Cyclic Reversible Termination).

Συνοπτικά, κατά τη διαδικασία αυτή, οι αλληλουχίες που πρόκειται να αναλυθούν, προσδένονται σε μία επιφάνεια πυριτίου και χρησιμεύουν ως καλούπι για τη σύνθεση συμπληρωματικών αλυσίδων, που επεκτείνονται μέσω ενζυμικής προσθήκης μονο-νουκλεοτιδίων, τα οποία φέρουν διαφορετικά φθοριοχρώματα.

Οι ευρύτερα διαδεδομένες συσκευές Illumina ακολουθούν δύο διαφορετικές μεθοδολογίες ανίχνευσης των νουκλεοτιδίων που προστίθενται στη νεοσυντιθέμενη αλυσίδα. Στη μία περίπτωση συμπεριλαμβάνονται τέσσερα διαφορετικά φθοριοχρώματα (4-channel sequencing), ένα για κάθε νουκλεοτίδιο (συσκευές MiSeq και HiSeq).

Στη δεύτερη περίπτωση, τα νουκλεοτίδια κυτοσίνης και θυμίνης, φέρουν δύο διαφορετικά φθοριοχρώματα (2-channel sequencing), κόκκινο και πράσινο, αντίστοιχα, παράλληλα τα νουκλεοτίδια αδενίνης φέρουν και τις δύο αυτές χρώσεις ταυτόχρονα, αποδίδοντας συνδυαστικά, κίτρινο χρώμα, ενώ η μη απόδοση ακτινοβολίας εκλαμβάνεται ως προσθήκη νουκλεοτιδίου γουανίνης, αφού αυτά δε φέρουν κανένα από τα παραπάνω φθοριοχρώματα.

3.3.1.2 Πειραματική και αναλυτική διαδικασία

3.3.1.2.1 Προετοιμασία της αντίδρασης

Αναλυτικά, η διαδικασία της προετοιμασίας βιβλιοθηκών νουκλεϊκών οξέων για προσδιορισμό της αλληλουχίας αυτών περιλαμβάνει:

1) Απομόνωση του κατάλληλου κλάσματος νουκλεϊκών οξέων (DNA ή RNA, το τελευταίο σε συνδυασμό με προετοιμασία cDNA), η οποία εξαρτάται από το στόχο της εκάστοτε πειραματικής διαδικασίας.

2) Κατακερματισμός των νουκλεϊκών οξέων, προς σχηματισμό αλυσίδων που κυμαίνονται σε μέγεθος, από 200 έως 500 βάσεις, ανάλογα με το πρωτόκολλο και το είδος της ανάλυσης.

3) Αντίδραση λιγάσης, για ενσωμάτωση, στα άκρα των κατακερματισμένων τμημάτων, διακριτών για κάθε δείγμα, συνθετικών ολιγονουκλεοτιδικών αλληλουχιών (adapters), οι οποίες έχουν ιδιαίτερη σημασία τόσο στο πειραματικό, όσο και στο αναλυτικό σκέλος της τεχνικής.

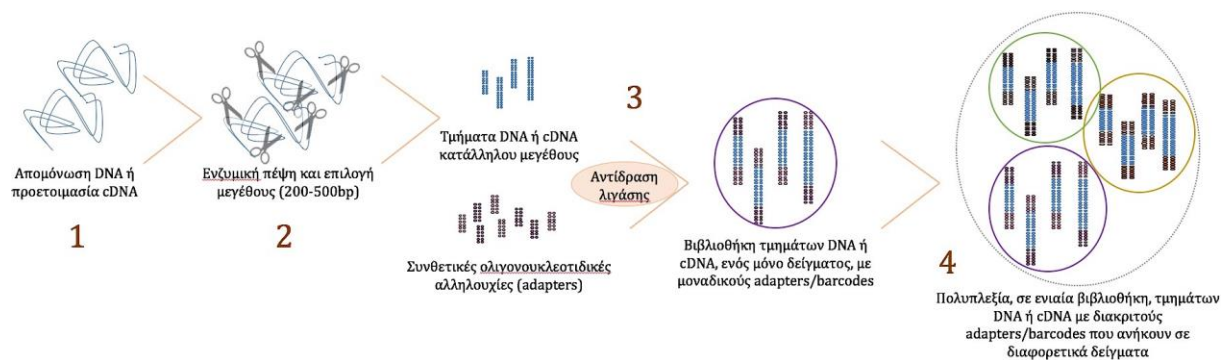
3α) Ένα μικρό τμήμα της συνθετικής αλληλουχίας (barcode), που περιλαμβάνει 6-8 τυχαία τοποθετημένα νουκλεοτίδια, προσδίδει την απαραίτητη ποικιλομορφία στα δείγματα μίας βιβλιοθήκης, εάν αυτή πρόκειται να συμπεριλάβει περισσότερα από ένα, τα οποία πρόκειται να αναμειχθούν και διαβαστούν ταυτόχρονα (πολυπλεξία). Η διαδικασία κατά την οποία συγκεκριμένα διαβάσματα αποδίδονται στο εκάστοτε δείγμα από το οποίο προέρχονται, είναι γνωστή ως αποπολυπλεξία και

πραγματοποιείται μέσω μεθόδων βιοπληροφορικής, κατά την ανάλυση των δεδομένων.

3β) Η υπόλοιπη συνθετική αλληλουχία, είναι όμοια για όλα τα δείγματα, ενώ περιλαμβάνει τμήματα χρήσιμα για ποσοτικές αναλύσεις, για διάφορα στάδια ενζυμικής ενίσχυσης των επιμέρους δειγμάτων, πριν και μετά την πολυπλεξία (multiplexing), όπως επίσης για την πρόσδεση της βιβλιοθήκης σε μία μοναδικής χρήσης επιφάνεια πυριτίου (flow cell), όπου θα πραγματοποιηθεί ο προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων.

4) Προαιρετικό στάδιο, προετοιμασίας μίας ενιαίας βιβλιοθήκης, η οποία απαρτίζεται από περισσότερα δείγματα, μέσω ανάμειξης ισομοριακής ποσότητας από το καθένα.

Διάφορα στάδια ποιοτικών και ποσοτικών αναλύσεων, παρεμβάλλονται σε διάφορα βήματα της διαδικασίας προετοιμασίας της αντίδρασης (Goodwin et al. 2016, Heather and Chain 2016).



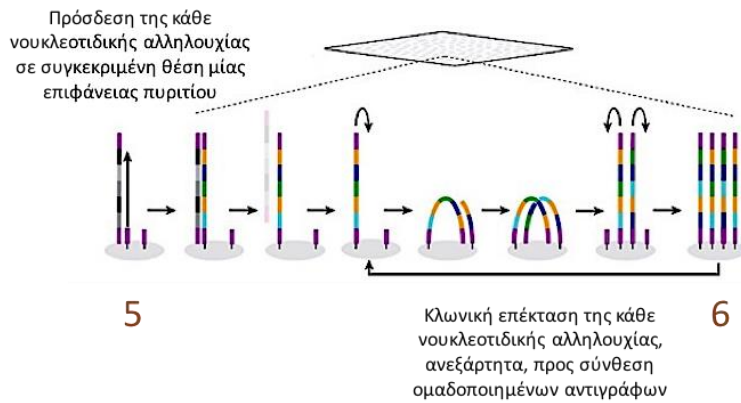
Εικόνα 13. Προετοιμασία βιβλιοθήκης DNA/cDNA, των δειγμάτων που προορίζονται για προσδιορισμό πρωτοδιάταξης νουκλεϊκών οξέων, με τεχνολογία NGS.

3.3.1.2.2 Κυρίως αντίδραση

Αφού προετοιμαστεί η βιβλιοθήκη, η οποία απαρτίζεται από ένα ή περισσότερα δείγματα, ακολουθεί το κυρίως μέρος της διαδικασίας:

5) Χημική αποδιάταξη των νουκλεϊκών οξέων, του δείγματος ή των δειγμάτων, που απαρτίζουν τη βιβλιοθήκη και πρόσδεση αυτών στο flow cell, μέσω συγκεκριμένων κοινών τμημάτων των adapters, που αναγνωρίζουν συμπληρωματικές τους αλληλουχίες, οι οποίες εντοπίζονται σταθερά προσδεμένες στο υπόστρωμα, εκ των προτέρων.

6) Κλωνική επέκταση (clonal amplification) της κάθε νουκλεοτιδικής αλληλουχίας μέσω μιας διαδικασίας που ονομάζεται bridge amplification, που οδηγεί σε σύνθεση ομαδοποιημένων αντιγράφων (clusters).



Εικόνα 14. Κλωνική επέκταση του συνόλου των νουκλεοτιδικών αλληλουχιών μέσω **bridge amplification**, που αφορούν στη βιβλιοθήκη DNA/cDNA, των δειγμάτων που προορίζονται για προσδιορισμό πρωτοδιάταξης νουκλεϊκών οξέων, με τεχνολογία NGS. (προσαρμογή από Shendure et al. 2017).

7) Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων και καταγραφή των δεδομένων. Ο προσδιορισμός της νουκλεοτιδικής ακολουθίας όλων των cluster που έχουν σχηματιστεί πραγματοποιείται ταυτόχρονα (*massive parallel sequencing*), για όλα τα δείγματα της βιβλιοθήκης, εφόσον αυτά είναι περισσότερα από ένα, ενώ περιλαμβάνει επιμέρους στάδια, όπως αυτά περιγράφονται στη συνέχεια:

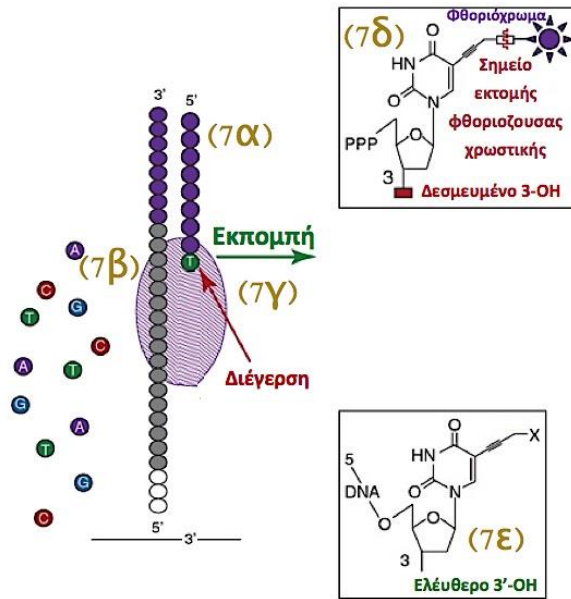
7α) Οι εκκινητές προσδένονται, μέσω συμπληρωματικότητας με κοινό τμήμα των adapters, σε καθεμία από τις προς ανάλυση ομαδοποιημένες νουκλεοτιδικές αλληλουχίες, οι οποίες βρίσκονται προσδεμένες στο υπόστρωμα.

7β) Χημικά τροποποιημένα φθορίζοντα dNTPs, ενσωματώνονται ενζυμικά, από πολυμεράσες, στο 3' άκρο των εκκινητών, προς σχηματισμού των νεοσυντιθέμενων αλληλουχιών. Κάθε νουκλεοτίδιο που προστίθεται, οδηγεί σε προσωρινή διακοπή της διαδικασίας επέκτασης, λόγω της ύπαρξης μίας αντιστρεπτής χημικής τροποποίησης, στο 3' άκρο κάθε dNTP.

7γ) Ένα σύστημα ακτινοβολίας διεγείρει τα φθοριοχρώματα των τεσσάρων διαφορετικών dNTPs, καθένα από τα οποία εκπέμπει φωτόνια σε διακριτό μήκος κύματος. Η ανίχνευση της εκπεμπόμενης ακτινοβολίας πραγματοποιείται μέσω μίας κάμερας υψηλής ευκρίνειας (*Charge-coupled device*, CCD).

7δ) Ακολουθεί απομάκρυνση των φθοριοχρωμάτων.

7ε) Η διαδικασία ολοκληρώνεται με την απομάκρυνση της χημικής τροποποίησης από το 3' άκρο των dNTPs, ώστε επιτρέπεται ο επόμενος κύκλος ενσωμάτωσης (Goodwin et al. 2016, Heather and Chain 2016).



Εικόνα 15. Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων, των δειγμάτων που απαρτίζουν τη βιβλιοθήκη DNA/cDNA (τροποποίηση από Mardis 2013).

στ) Προαιρετικά, μόλις ολοκληρωθεί το διάβασμα των αλληλουχιών για κάθε cluster, προς μία κατεύθυνση (single-end sequencing, SE-Seq), υπάρχει η δυνατότητα διαβάσματος της νουκλεοτιδικής ακολουθίας και προς την αντίθετη κατεύθυνση (paired-end sequencing, PE-Seq). Αυτή η διάταξη αυξάνει την αξιοπιστία της χαρτογράφησης των δεδομένων κατά την αντιπαραβολή με το πρότυπο γονιδίωμα και τις ποιότητας των διαβασμάτων, σε μεταγενέστερο στάδιο, κατά τη βιοπληροφορική ανάλυση, όπως περιγράφεται παρακάτω (Yohe and Thyagarajan 2017).

3.3.1.2.3 Ανάγνωση του αποτελέσματος

Τα δεδομένα που προκύπτουν από NGS αναλύσεις καταγράφονται σε ειδικά διαμορφωμένα υπολογιστικά αρχεία (fastq files), όπου για κάθε διάβασμα, η πρώτη σειρά περιλαμβάνει στοιχεία σχετικά με την πλατφόρμα που χρησιμοποιήθηκε, με την ημερομηνία και ώρα διεξαγωγής του πειράματος, καθώς και στοιχεία ειδικά της εκάστοτε αλληλουχίας, του κάθε δείγματος, όπως η αλληλουχία barcode του adapter, στη δεύτερη σειρά καταγράφεται η υπό διερεύνηση νουκλεοτιδική ακολουθία που διαβάστηκε, ενώ τα σύμβολα στην τέταρτη σειρά αφορούν σε ποιοτικά δεδομένα, σχετικά με το διάβασμα κάθε νουκλεοτιδίου.

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((( (**+))%%%++) (%%%) .1***-+*!))**55CCF>>>>>CCCCCC65
```

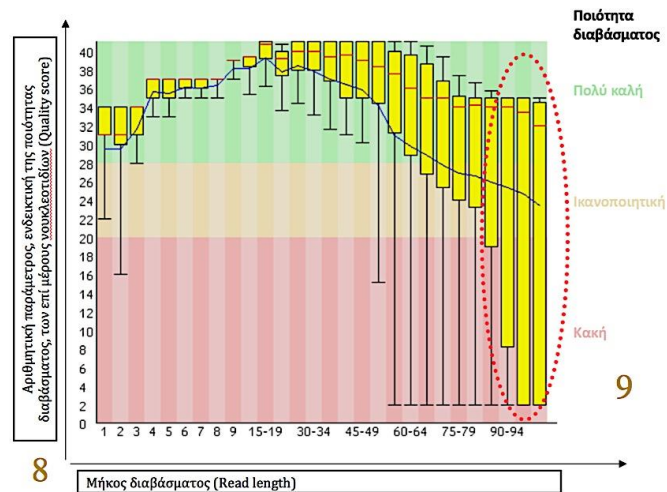
Εικόνα 16. Αρχεία διαβασμάτων (fastq files). Ενδεικτική μορφή, με την οποία αποτυπώνονται τα δεδομένα, που προκύπτουν από προσδιορισμό αλληλουχίας με NGS, σε ειδικά υπολογιστικά αρχεία τύπου fastq. Η εικόνα αφορά σε ένα μόνον διάβασμα.

Ο μεγάλος όγκος πολύπλοκων δεδομένων που πηγάζει από τις σύγχρονες συσκευές προσδιορισμού αλληλουχίας, απαιτεί την ανάπτυξη ολοένα και πιο εκλεπτυσμένων

αλγόριθμων, ωστόσο, σε γενικές γραμμές, η βιοπληροφορική αναλυτική διαδικασία ακολουθεί ορισμένα τυπικά στάδια, όπως αυτά περιγράφονται στη συνέχεια:

8) Ποιοτική ανάλυση των δεδομένων.

9) Απομάκρυνση των νουκλεοτιδίων κατώτερης ποιότητας, καθώς και των νουκλεοτιδίων που ανήκουν σε adapters ή σε άλλες τεχνητές αλληλουχίες, που χρησιμεύουν μόνον στο τεχνικό κομμάτι της μεθοδολογίας, ενώ δεν αποτελούν τμήμα των υπό διερεύνηση περιοχών.



Εικόνα 17. Ποιοτική ανάλυση δεδομένων διαβασμάτων. Διαγραμματική απεικόνιση, μίας ενδεικτικής στατιστικής ανάλυσης (boxplot analysis), της ποιότητας των διαβασμάτων NGS, προκειμένου να απομακρυνθούν τα νουκλεοτίδια που διαβάστηκαν με μειωμένη αξιοπιστία, εκ των υστέρων. Η τιμή της ποιότητας (Quality ή Q score) προκύπτει από τον τύπο $Q = -10 \log_{10} P$, όπου P είναι η πιθανότητα λανθασμένα διαβασμένων νουκλεοτιδίων, έτσι, ενδεικτικά, εάν απαντάται 1 λάθος ανά 100 διαβάσματα, τότε η τιμή Q ισούται με 20, που αντιπροσωπεύει 99% ακρίβεια διαβάσματος (accuracy). Όπως φαίνεται και στην εικόνα, η ποιότητα διαβάσματος μειώνεται όσο προχωρά το μήκος της νεοσυντιθέμενης αλληλουχίας, γεγονός που σχετίζεται με το τεχνικό κομμάτι της τεχνολογίας NGS, με συσκευές της Illumina (Leggett et al. 2013).

10) Αντιπαραβολή των διαβασμάτων (sequence alignment) επί πρότυπου γονιδιώματος (reference genome), για τη σύνθεση ενός χάρτη διαβασμάτων. Στο σημείο αυτό, αξίζει να τονίσουμε ξανά, τη σημασία του βάθους με το οποίο διαβάζονται τα νουκλεοτίδια (read depth), σε μία μεθοδολογία NGS, αφού αποτελεί μία ιδιαίτερα σημαντική παράμετρο, για την αξιόπιστη ανάδειξη νουκλεοτιδικών αποκλίσεων (calling confidence) σε σύγκριση με το πρότυπο γονιδίωμα.

ΑΛΛΗΛΟΥΧΙΑ ΠΡΟΤΥΠΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

TCACCCAAGAGGAAAGGGACAGG**G**TAGGCACCCCAGGTGA**G**TCTGACTCAAAAATGGTACTTCTTGTA AAC



- **ΒΑΘΟΣ ΔΙΑΒΑΣΜΑΤΟΣ (Read depth):**
- Πόσες φορές έχει διαβαστεί μία γενομική θέση
- Υψηλότερες τιμές = ΠΙΟ ΑΞΙΟΠΙΣΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ
- Η τιμή εξαρτάται από την πολυπλοκότητα της αλληλουχίας του DNA (π.χ. πολυεπαναλαμβανόμενα στοιχεία), από τη μέθοδο προετοιμασίας της βιβλιοθήκης (amplicon vs. probe), από το πλήθος διαβασμάτων και άλλους παράγοντες.

Εικόνα 18. Αντιπαραβολή των διαβασμάτων επί πρότυπου γονιδιώματος (Sequence alignment).

11) Ανάδειξη αποκλίσεων από το πρότυπο γονιδίωμα, μέσω διαφορετικών αλγόριθμων που αναγνωρίζουν διάφορους τύπους γενετικών αλλαγών, όπως εξαλλαγές ενός (SNVs, single nucleotide variants) ή περισσότερων νουκλεοτιδίων (MNVs, multiple nucleotide variants), καθώς και περιορισμένες ή ευρύτερες προσθήκες ή διαγραφές νουκλεοτιδίων (Indels, Insertion or deletion events), τόσο σε εξώνια, όσο και σε περιοχές σημαντικές για την ωρίμανση του pre-mRNA, αλλά και αριθμητικές αλλαγές που αφορούν σε εξώνια ή ολόκληρα γονίδια (Copy Number Variations, CNVs), καθώς και δομικές χρωμοσωμικές αλλαγές, όπως η δημιουργία χμαιρικών γονιδίων (fusion genes) λόγω μεταθέσεων γενομικού υλικού, ή άλλες αναδιατάξεις. Επισκόπηση και επιβεβαίωση των μεταλλάξεων σε προγράμματα απεικόνισης γονιδιωματικών χαρτών.



Εικόνα 19. Απεικόνιση διαβασμάτων και μεταλλάξεων σε γονιδιωματικούς χάρτες. Επιβεβαίωση ενός SNV που ανιχνεύθηκε στο εργαστήριό μας, εν προκειμένω στο γονίδιο TP53, σε ασθενή με

ΟΜΛ, όπως αυτή παρουσιάζεται, σε υπολογιστικό πρόγραμμα απεικόνισης γονιδιωματικών χαρτών, συγκεκριμένα στο IGV (Integrative genomic viewer) (Robinson et al. 2011, Thorvaldsdóttir et al. 2013).

12) Εκτίμηση της επίδρασης των μεταλλαγών που ανιχνεύθηκαν, στο φαινότυπο των ασθενών μέσω:

12α) διασταύρωσης στοιχείων με βάσεις δεδομένων (annotation databases).

12β) προβλεπτικών αλγόριθμων (prediction algorithms).

Οι περισσότερες μελέτες που αφορούν στην ανίχνευση επίκτητων μεταλλάξεων (σωματικές μεταλλάξεις) σε καρκινικά κύτταρα από συμπαγείς όγκους, συνεκτιμούν τυχόν προϋπάρχουσες κληρονομούμενες γενετικές αλλαγές (γαμετικές μεταλλάξεις), μέσω παράλληλης ανάλυσης ενός πληθυσμού φυσιολογικών κυττάρων του ασθενούς (T/N pair, Tumor/Normal pair).

Ωστόσο, στους ασθενείς με αιματολογικές κακοήθειες, τα καρκινικά κύτταρα διηθούν επιπρόσθετους ιστούς, όπως το δέρμα ή το επιθήλιο της στοματικής κοιλότητας, οπότε η απομόνωση φυσιολογικού δείγματος, για αντιπαραβολή με το παθολογικό, είναι μία ιδιαίτερα δυσχερής διαδικασία.

Οπότε, αναμένεται η χρονική στιγμή που ο ασθενής θα βρεθεί σε ύφεση νόσου, εφόσον αυτό συμβεί, ώστε ενδεχομένως υπάρχει η δυνατότητα συλλογής επαρκούς ποσότητας φυσιολογικού δείγματος, από τους προαναφερθέντες ιστούς.

Εναλλακτικά απομονώνεται καθαρός πληθυσμός φυσιολογικών αιμοποιητικών κυττάρων, όπως κύτταρα της λεμφικής σειράς, για τους ασθενείς με μυελικές νεοπλασίες, μέσω εξειδικευμένων κυτταρόμετρων ροής ή με άλλες τεχνικές (Luthra et al. 2015, Wartman and Mardis 2016).

3.3.2 Docker

Η ανάλυση των δεδομένων που προέρχονται από τις παραπάνω διαδικασίες (κεφάλαιο 3.3.1.2.3) με μεθόδους βιοπληροφορικής, στηρίζεται στην ορθή λειτουργία ενός σημαντικού αριθμού προγραμμάτων και λογισμικών με διαφορετικές διακριτές λειτουργίες, οι οποίες πολλές φορές είναι μη συμβατές μεταξύ τους.

Για να επιτευχθεί αυτό, απαιτείται η προετοιμασία κατάλληλου υπολογιστικού περιβάλλοντος (γλώσσες προγραμματισμού, αρχεία κ.α) ώστε κάθε πρόγραμμα ή/και λογισμικό να μπορεί να λειτουργήσει τόσο σε τοπικό επίπεδο, όσο και σε υπολογιστικό δίκτυο (computer cluster) ή στο υπολογιστικό νέφος (cloud computing) εφόσον αυτό κρίνεται ωφέλιμο.

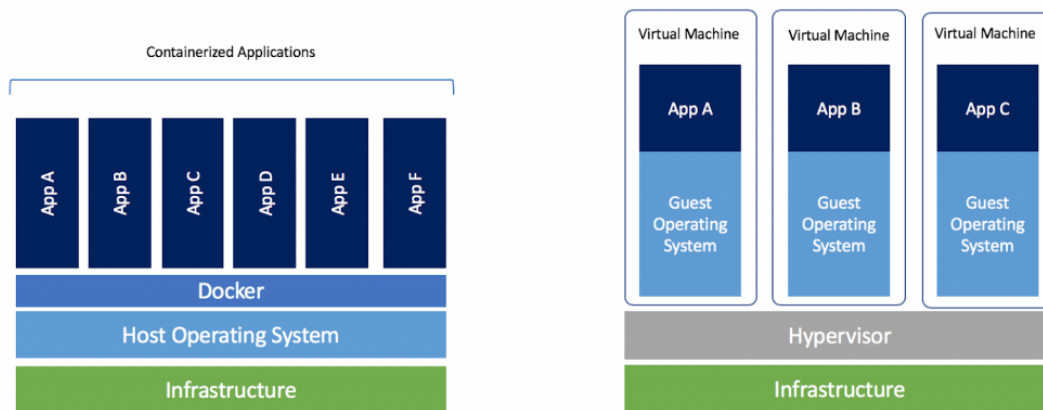
3.3.2.1 Σύγκριση Εικονικής Μηχανής και Docker

Μία λύση για την προετοιμασία ενός τέτοιου κατάλληλου περιβάλλοντος, αποτελεί η χρήση εικονικοποίησης και η δημιουργία ξεχωριστών εικονικών περιβαλλόντων, μέσα στα οποία θα εγκατασταθεί και θα λειτουργεί κάθε πρόγραμμα. Εξειδικευμένα προγράμματα, γνωστά ως εικονικές μηχανές (Virtual Machine, VM), επιτρέπουν την εγκατάσταση πολλαπλών εικονικών λειτουργικών υπολογιστών, σε μία μόνο συσκευή. Οι εικονικοί υπολογιστές κατά τη δημιουργία τους, προαπαιτούν ένα λειτουργικό σύστημα ως βάση, ώστε στη συνέχεια να ανατεθούν συγκεκριμένες δυνατότητες και λειτουργίες. Επίσης, παρόλο που ο χρήστης

επιθυμεί συγκεκριμένο λογισμικό και δυνατότητες από τον εικονικό υπολογιστή που δημιουργεί, είναι αναγκασμένος να εγκαταστήσει επιπλέον πακέτα τα οποία δεν τα χρησιμοποιεί. Αυτό καθιστά τον εικονικό υπολογιστή αργό στην εκκίνηση και μεγάλο σε μέγεθος, ενώ πολλές φορές υπάρχουν επιπλοκές κατά τη φορητότητα τους και δύσκολα επεκτείνονται.

Για τους παραπάνω λόγους, το λογισμικό Docker εμφανίζεται ως πιο ολοκληρωμένη λύση, αφού δεν απαιτεί ολόκληρο λειτουργικό σύστημα, αλλά μόνο τη βάση αυτού, καθώς και τις απαραίτητες βιβλιοθήκες για να λειτουργήσει, εξοικονομώντας έτσι, χρόνο κατά την εκκίνηση και χώρο στην μνήμη. Μεγάλο πλεονέκτημα αποτελεί το γεγονός πως το Docker δεσμεύει δυναμικά μνήμη από το σύστημα, δηλαδή όσο προστίθενται λειτουργίες στην εικόνα, ανάλογα το μέγεθος, δεσμεύεται ο αντίστοιχος χώρος από το σύστημα.

Αντίθετα, στις εικονικές μηχανές, η δέσμευση χώρου γίνεται από την αρχή, ανάλογα με τις ρυθμίσεις του χειριστή, οδηγώντας έτσι, είτε σε περίσσεια χώρου που παραμένει ανεκμετάλλευτος, είτε σε πλήρωση της διαθέσιμης χωρητικότητας, με συνέπεια την ανάγκη για δημιουργία ενός νέου εικονικού υπολογιστή από την αρχή.



Εικόνα 20. Σχηματική αναπαράσταση του τρόπου λειτουργίας των VM και του Docker (<https://blog.docker.com/2018/08/containers-replacing-virtual-machines>).

3.3.2.2 Τι είναι το Docker

Το Docker (<https://www.docker.com>) είναι ένα λογισμικό που παρέχει εικονικοποίηση σε επίπεδο λειτουργικού συστήματος, γνωστή και ως containers. Πρόκειται για ένα λογισμικό ανοιχτού κώδικα που αναπτύχθηκε από την εταιρεία Docker Inc. το 2013. Είναι γραμμένο σε γλώσσα προγραμματισμού Go και μπορεί να εκτελεστεί από τα λειτουργικά συστήματα Linux, MacOS και Windows x86-64.

Η τεχνολογία docker παρέχει τη διαχείριση υπολογιστικών πόρων (CPU, μνήμη, δίκτυο) και την απομόνωση των εφαρμογών στο λειτουργικό περιβάλλον που χρησιμοποιούνται. Αυτός είναι και ο στόχος των containers, δηλαδή να εκτελούνται πολλαπλές διεργασίες και εφαρμογές ξεχωριστά, ώστε να επιτυγχάνεται καλύτερη χρήση των υπολογιστικών πόρων του συστήματος και παράλληλα να διατηρείται η ασφάλεια. Επειδή τα containers, τα οποία είναι εικόνες σε εκτέλεση, είναι τόσο ελαφριά, ένας και μόνο φυσικό μηχάνημα ή VM να τρέξει ταυτόχρονα πολλά containers.

Τέλος, η χρήση του Docker για τη δημιουργία και τη διαχείριση container μπορεί να απλοποιήσει τη δημιουργία εξαιρετικά καταναεμημένων συστημάτων, επιτρέποντας σε

πολλαπλές εφαρμογές, και άλλες διαδικασίες να εκτελούνται αυτόνομα σε ένα μονό φυσικό μηχάνημα ή σε πολλαπλές εικονικές μηχανές.

3.3.2.3 Docker Εικόνες (Docker Images)

Η κάθε εικόνα που δημιουργείται μέσω του Docker Engine, μπορεί να λειτουργήσει σε οποιοδήποτε Linux σύστημα, είτε αυτό είναι ένας φυσικός υπολογιστής, είτε σε εικονικούς. Το γεγονός αυτό βελτιώνει τη φορητότητα και επεκτασιμότητα της εφαρμογής. Οι εικόνες είναι, εφαρμογές που δημιουργούνται για να εκτελούν μια συγκεκριμένη διεργασία και μετά το πέρας αυτής να απενεργοποιούνται μέχρι να ξανακαλεστούν. Κατά την εντολή να τρέξει μια εικόνα, αυτή δέχεται ένα μοναδικό κωδικό (ID), και αυτό δίνει τη δυνατότητα να τρέξει η ίδια εικόνα πολλές φορές, με διαφορετικό ID κάθε φορά. Έτσι μπορεί να τρέξει ταυτόχρονα μεγάλος αριθμός ιδίων εικόνων, με την κάθε μια να εκτελεί τη διαδικασία για την οποία καλέστηκε, ξεχωριστά.

Είναι σημαντικό να αναφερθεί ότι, οι εικόνες είναι στατικές, δηλαδή κατά την απενεργοποίησή τους, οποίες αλλαγές και να πραγματοποιηθήκαν μέσα σε αυτές αναιρούνται, και η εικόνα επιστρέφει στην αρχική της κατάσταση (<https://docs.docker.com/get-started/>).

3.3.2.4 Docker Container

Οι εικόνες που δημιουργούνται, όπως αναφέρθηκε παραπάνω, όταν τρέχουν ονομάζονται containers. Πακετάροντας με τέτοιο τρόπο, ώστε να μπορούν να εκτελεστούν και να είναι προσβάσιμες, είτε μέσω διαδικτύου (άμα πρόκειται για κάποια διαδικτυακή εφαρμογή), είτε μέσω τόμων (volumes), για την επικοινωνία μεταξύ των containers, και την πρόσβαση τους σε κοινά αρχεία. (<https://docs.docker.com/get-started/>) (https://en.wikipedia.org/wiki/Linux_containers).

Το Container είναι ένα απομονωμένο περιβάλλον εκτέλεσης εφαρμογών λογισμικού. Περιέχουν τα άκρως απαραίτητα εργαλεία και εξαρτήσεις που χρειάζονται για να εκτελεστεί μια εφαρμογή. Έτσι η εικόνα (image) που δημιουργείται, είναι αρκετά μικρότερη σε μέγεθος, σε σχέση με έναν εικονικό υπολογιστή, που θα έπρεπε να δημιουργηθεί για να εκτελέσει την ίδια εφαρμογή. Αυτό προσφέρει περισσότερο ελεύθερο χώρο σε σκληρό δίσκο, καθώς και λιγότερη κατανάλωση της RAM και επεξεργαστή (CPU).

3.3.3 Η γλώσσα CWL (Common Workflow Language)

Όπως αναφέρεται και από τους δημιουργούς της (<https://www.commonwl.org>) “Η Common Workflow Language είναι μια γλώσσα για την περιγραφή ροών εργασίας και των επιμέρους εργαλείων της ανάλυσης με τρόπο που τις καθιστά φορητές και επεκτάσιμες σε μια ποικιλία λογισμικών και υπολογιστικών συστημάτων, από σταθερούς Η/Υ, υπολογιστικά νέφη έως cluster υπολογιστών. Η CWL έχει σχεδιαστεί για να καλύπτει τις ανάγκες των επιστημών που επεξεργάζονται μεγάλο όγκο δεδομένων, όπως η βιοπληροφορική, η ιατρική απεικόνιση, η αστρονομία, η φυσική και η χημεία”.

Η CWL περιγράφει και προτυποει αναλυτικές ροές με τρόπο ανεξάρτητο του υπολογιστικού συστήματος ενώ είναι συμβατή με την τεχνολογία του Docker καθιστώντας έτσι δυνατή την επαναληψιμότητα και την σταθερή διατήρηση των αναλυτικών ροών με μόνη προϋπόθεση

την ύπαρξη του Docker και ενός λογισμικού εκτέλεσης CWL αρχείων (script) στο επιθυμητό υπολογιστικό σύστημα (Peter Amstutz et al. 2016).

3.4 Μεθοδολογία

3.4.1 Απομόνωση γενομικού DNA

Η απομόνωση γενομικού DNA μεγάλου μοριακού βάρους, από εμπύρνηνα κύτταρα, περιφερικού αίματος ή μυελού των οστών πραγματοποιήθηκε απευθείας από δείγματα ολικού περιφερικού αίματος ή μυελού των οστών με τη βοήθεια του FlexiGene blood genomic kit, από την εταιρεία Qiagen, σύμφωνα με το πρωτόκολλο που ακολουθεί.

Αντιδραστήρια

- ddH₂O για λύση των κυττάρων
- Διάλυμα θραύσης κυτταρικών μεμβρανών FG1
- Διάλυμα θραύσης πυρήνων FG2
- Πρωτεάση Qiagen (PrK)
- Ισοπροπανόλη 100%
- Διάλυμα αιθανόλης 70%
- Διάλυμα ενυδάτωσης DNA FG3 (10mM Tris.Cl , pH 8.5)

Τεχνική

1. Λύση των κυττάρων του περιφερικού αίματος ή του μυελού των οστών. Προσθήκη H₂O στο falcon tube που βρίσκεται το δείγμα του αίματος έως το σημείο των 40ml.
2. Φυγοκέντρηση 20 λεπτά, 4.000 rpm σε θερμοκρασία δωματίου (room temperature-RT).
3. Απόρριψη υπερκείμενου.
4. Μηχανική επαναιώρηση pellet.
5. Προσθήκη 5 ml FG1, για την θραύση πυρηνικών μεμβρανών.
6. Έντονη ανάδευση, Vortex.
7. Φυγοκέντρηση 10 λεπτά, 2.000 rpm, RT.
8. Απόρριψη υπερκείμενου.
9. Προσθήκη 1 ml διαλύματος FG2/PrK (100:1), για απομάκρυνση πρωτεϊνών.
10. Μηχανική επαναιώρηση pellet.
11. Επώαση σε υδατόλουτρο 65° C, 10 λεπτά.
12. Προσθήκη propanol-2, συμπύκνωση του DNA.
13. Ανάδευση, Πιστοποίηση οπτικοποίησης-καθίζησης DNA.
14. Μεταφορά σε erpendorf tube 2 ml.
15. Φυγοκέντρηση 10 min, 3.000 rpm, RT.
16. Απόρριψη υπερκείμενου.
17. DNA pellet, επαναδιαλύεται με προσθήκη δ/τος EtOH 70%.
18. Φυγοκέντρηση 10 min, 3.000 rpm, RT.
19. Απόρριψη υπερκείμενου.
20. Αφαίρεση περισσειας EtOH με πιπέτα.
21. Πλήρης εξάτμιση του διαλύτη σε speed vac. (~60 sec).
22. Το καθαρό DNA επαναδιαλύεται σε κυμαινόμενη ποσότητα δ/τος FG3 (50-500 μl).

23. Thermoblock στους 65° C overnight (12-24 h).

24. Φωτομέτρηση δείγματος σε δύο μήκη κύματος (260nm και 280nm). Ο λόγος O.D₂₆₀/O.D₂₈₀ πρέπει να είναι περίπου 1.9 για υψηλής καθαρότητας ακέραια δείγματα DNA.

25. Φύλαξη στους 4 °C.

Τροποποίηση πρωτοκόλλου FlexiGene DNA Handbook (QIAGEN).

3.4.2 Στοχευμένος Προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με Τεχνολογία Δεύτερης Γενιάς (targeted Next Generation Sequencing, targeted NGS)

Οι τεχνολογίες προσδιορισμού αλληλουχίας, με συσκευές νέας γενιάς, έχουν αναπτυχθεί σε βαθμό που μπορούν πλέον να καλύψουν ευρύτερες γενομικές περιοχές, αναζητώντας πληθώρα διαφορετικών τύπων μεταλλάξεων, με στόχο να καλύψουν το κενό που υπάρχει, στην κατανόηση των αιματολογικών κακοηθειών, όπως και άλλων νοσημάτων.

Στην παρούσα εργασία μελετήθηκαν γονίδια τα οποία, σύμφωνα με τη διεθνή βιβλιογραφία, εμφανίζουν μεταλλάξεις σε υψηλό ποσοστό ασθενών με μυελικές νεοπλασίες. Τα εξόνια των γονιδίων που διερευνήθηκαν με τεχνολογία NGS, καταγράφονται στον Πίνακα που ακολουθεί.

Πολλά από αυτά έχουν περιγραφεί ήδη, ενώ ορισμένα, αναλύθηκαν ταυτόχρονα, με τεχνικές που εφαρμόζονταν ήδη ή που αναπτύχθηκαν παράλληλα, στο εργαστήριο Αιματολογίας του ΙΒΕΑΑ, και χρησιμοποιήθηκαν ως θετικοί μάρτυρες για την ευαισθησία της μεθόδου και της βιοπληροφορικής ανάλυσης.

<i>ASXL1</i> (exon 12)	<i>BCOR</i> (exons 2-15)	<i>BCORL1</i> (exon 3 + 4)	<i>CALR</i> (exon 9)	<i>CBL</i> (exons 1-16)	<i>CSF3R</i> (exons 14-17)
<i>CUX1</i> (exons 1-23)	<i>DNMT3A</i> (exons 3-23)	<i>EZH2</i> (exons 2-20)	<i>GATA1</i> (exons 2+3)	<i>GATA2</i> (exons 3-6)	<i>IDH1</i> (exon 4)
<i>IDH2</i> (exon 4)	<i>JAK2</i> (exons 12-14)	<i>KRAS</i> (exons 2+3)	<i>MPL</i> (exon 10)	<i>NRAS</i> (exons 2+3)	<i>RUNX1</i> (exons 2-9)
<i>SF3B1</i> (exons 13-18)	<i>SH2B3</i> (exons 2-8)	<i>SRSF2</i> (exons 1+2)	<i>STAG2</i> (exons 3-35)	<i>TET2</i> (exons 3-11)	<i>TP53</i> (exons 2-11)
<i>U2AF1</i> (exons 2+6)	<i>ZRSR2</i> (exons 1-11)	<i>KIT</i> (exons 17)	<i>SETBP1</i> (exon 4)		

Πίνακας 7. Εξόνια γονιδίων που διερευνήθηκαν με τεχνολογία NGS.

Η διαδικασία που ακολουθήθηκε περιγράφεται στη συνέχεια. Σε κάθε πειραματική πορεία προετοιμάζονται δείγματα, σε πλήθος πολλαπλάσιο του 8, αφού έτσι επιβεβαιώνεται η ελάχιστη δυνατή απώλεια αναλωσίμων, κατά την προετοιμασία των αντιδράσεων, σύμφωνα με το πρωτόκολλο του κατασκευαστή (Ovation Custom Target Enrichment System, NuGen).

3.4.2.1 Παρασκευή διαλύματος 0.5M EDTA

Το διάλυμα προετοιμάζεται κατάλληλα όπως περιγράφεται στη συνέχεια.

- Υπολογισμοί:

Ισχύει ότι $0.5M = 0.5\text{mol}$ ουσίας σε 1000mL διαλύματος EDTA

Άρα σε 10mL διαλύματος EDTA, θα υπάρχουν $0.5\text{mol}/100 = 0.005\text{mol}$ (ή 5mmol) ουσίας

Το μοριακό βάρος του EDTA είναι $372,24\text{g/mol}$.

Άρα, 1mol EDTA περιλαμβάνεται σε $372,24\text{g}$ αυτού.

Έτσι 0.005mol EDTA θα περιλαμβάνονται σε $0.005\text{mol} \cdot 372,24\text{g}/1\text{mol} = 1.8612\text{g}$

- Προετοιμασία διαλύματος σε γυάλινο δοχείο Duran 50mL :

Συστατικό	Ποσότητα/Όγκος
EDTA	1.8612g
ddH ₂ O	10mL
Συνολικός όγκος	10mL

3.4.2.2 Ενζυμικός κατακερματισμός DNA

Ο ενζυμικός κατακερματισμός του DNA αποσκοπεί στο δημιουργία νουκλεοτιδικών ακολουθιών που κυμαίνονται μεταξύ 300 και $1,000\text{bp}$, με μέσο μέγεθος 500bp .

- Προετοιμασία ενός μείγματος αντιδραστηρίων σε σωληνάριο PCR 0.2mL , για κάθε δείγμα ξεχωριστά, ανάδευση σε συσκευή vortex και τοποθέτηση σε πάγο:

Αντιδραστήριο PCR	Όγκος
Υπόστρωμα DNA (μέχρι $3\mu\text{g}$)	$1.0\text{--}16.0\mu\text{L}$
Ρυθμιστικό διάλυμα (Fragmentase Rxn Buffer v2, NEB)	$2.0\mu\text{L}$
WFI	μέχρι συμπλήρωσης $18\mu\text{L}$
Συνολικός όγκος	$18.0\mu\text{L}$

- Ανάδευση πεπτικών ενζύμων, NEBNext dsDNA Fragmentase (NEB), σε συσκευή vortex και γρήγορη κατακρήμνιση (Quick spin), πριν από άμεση προσθήκη $2\mu\text{L}$ στο διάλυμα buffer-DNA, που παρακευάστηκε μόλις. Συνολικός όγκος αντίδρασης ίσος με $20\mu\text{L}$.
- Ανάδευση του κάθε μείγματος, σε συσκευή vortex, γρήγορη κατακρήμνιση και τοποθέτηση σε πάγο, μέχρι να προστεθούν πεπτικά ένζυμα και στα υπόλοιπα δείγματα.
- Ταυτόχρονη επώαση όλων των δειγμάτων, σε συσκευή PCR, σε θερμοκρασία 37°C , για χρονικό διάστημα $23'$.
- Άμεση απομάκρυνση των σωληνάρων από τη συσκευή PCR και τοποθέτηση σε πάγο.
- Προσθήκη $5\mu\text{L}$ EDTA $0.5M$ και ανάδευση με πιπέτα, ώστε να τερματίσει η αντίδραση πέψης. Συνολικός όγκος αντίδρασης ίσος με $25\mu\text{L}$.
- Άμεσος καθαρισμός των προϊόντων.

3.4.2.3 Απομάκρυνση προσμείξεων της αντίδρασης πέψης και διπλή επιλογή μεγέθους τμημάτων DNA

Έπειτα από την αντίδραση πέψης, ακολουθεί απομάκρυνση των προσμείξεων, με ταυτόχρονη διπλή επιλογή ($0.53x\text{--}0.77x$) των τμημάτων DNA κατάλληλου μήκους ($300\text{--}1,000\text{bp}$), μέσω μαγνητικών σφαιριδίων.

- Επώαση διαλύματος μαγνητικών σφαιριδίων σε θερμοκρασία περιβάλλοντος χώρου (ΘΠ), για 20' πριν από τη χρήση αυτών.
- Προσθήκη 75uL WFI, στα 25uL αντίδρασης. Συνολικός όγκος αντίδρασης ίσος με 100uL.
- Προσθήκη 53uL διαλύματος μαγνητικών σφαιριδίων. Συνολικός όγκος αντίδρασης ίσος με 153uL.
- Ανάδευση με πιπέτα.
- Επώαση μείγματος για 5' σε ΘΠ.
- Τοποθέτηση των σωληνάρων PCR 0.2mL σε κατάλληλες υποδοχές μίας αυτοσχέδιας βάσης, όπου ειδικοί μαγνήτες (rare-earth magnets) ασκούν ελκτική δύναμη, επί των μαγνητικών σφαιριδίων, επάνω στα οποία προσδέονται τμήματα DNA μεγαλύτερα από 1,000bp.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 5' σε ΘΠ.
- Ποσότητα ίση με 143uL διαλύματος, που περιέχει τμήματα DNA, μικρότερα από 1,000bp, μεταφέρονται σε νέο προσημασμένο σωληνάριο PCR 0.2mL. Τα σωληνάρια με τα μαγνητικά σφαιρίδια, από το προηγούμενο στάδιο της διαδικασίας, απορρίπτονται.
- Προσθήκη 24uL διαλύματος μαγνητικών σφαιριδίων. Συνολικός όγκος αντίδρασης ίσος με 167uL.
- Ανάδευση με πιπέτα.
- Επώαση μείγματος για 5' σε ΘΠ.
- Τοποθέτηση των σωληνάρων PCR 0.2mL επί της μαγνητικής βάσης. Τα μαγνητικά σφαιρίδια, αυτή τη φορά, προσδέουν τα τμήματα DNA με μέγεθος μεγαλύτερο από 300bp.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 5' σε ΘΠ.
- Ποσότητα ίση με 157uL διαλύματος, που περιέχει τμήματα DNA, μικρότερα από 300bp, απορρίπτονται. Πλέον στα μαγνητικά σφαιρίδια εντοπίζονται προσδεμένα, τμήματα DNA μεταξύ 300 και 1,000 νουκλεοτιδίων (μέσο μέγεθος 500bp).
- Ακολουθεί προσθήκη 180uL αιθανόλης (EtOH) 85%.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 30'' σε ΘΠ.
- Απομάκρυνση 180uL αιθανόλης (EtOH) 85%.
- Ακολουθεί δεύτερη προσθήκη 180uL αιθανόλης (EtOH) 85%.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 30'' σε ΘΠ.
- Απομάκρυνση 180uL αιθανόλης (EtOH) 85%.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 7' σε ΘΠ και απομάκρυνση περίσσιας αιθανόλης από τη βάση του σωληναρίου.
- Απομάκρυνση του σωληναρίου από τη μαγνητική βάση και ενυδάτωση των σφαιριδίων με το DNA, μέσω προσθήκης 22uL WFI.
- Ανάδευση με πιπέτα.
- Επώαση μείγματος για 3' σε ΘΠ.
- Τοποθέτηση του σωληναρίου στη μαγνητική βάση.
- Επώαση μείγματος επί της μαγνητικής βάσης, για 3' σε ΘΠ.
- Μεταφορά 20uL διαλύματος DNA σε νέο προσημασμένο σωληνάριο PCR των 0.2mL.
- Αποθήκευση σε θερμοκρασία +4°C για μετέπειτα επεξεργασία, ή άμεση έναρξη της διαδικασίας προετοιμασίας της βιβλιοθήκης δειγμάτων DNA, κατόπιν ποσοτικής και ποιοτικής ανάλυσης των δειγμάτων.

3.4.2.4 Ποσοτικός προσδιορισμός

Ο ποσοτικός προσδιορισμός των κατακερματισμένων τμημάτων DNA, πραγματοποιήθηκε μέσω κατάλληλων φθοριοχρωμάτων, σε συσκευή Qubit.

Η ανάλυση του DNA με χρήση φθοριοχρωμάτων, αποτελεί μία γρήγορη και αξιόπιστη τεχνική για τον ακριβή προσδιορισμό της ποσότητας των νουκλεϊκών οξέων.

Αντιδραστήρια

1. DNA: Καθαρισμένο προϊόν της αντίδρασης PCR
2. Ρυθμιστικό διάλυμα με χρωστική: AccuBlue dsDNA Quantitation Solution (Biotium, USA)
3. Ενισχυτής σήματος: 100x Enhancer (Biotium, USA)
4. Πρότυπο δείγμα συγκεκριμένης συγκέντρωσης για καθορισμό της μέγιστης τιμής σήματος: Pre-diluted dsDNA standard (10ng/uL) (Biotium, USA)
5. ddH₂O

Τεχνική

1. Αρχικά προετοιμάζεται 1:10 αραιωμένο διάλυμα καθαρισμένου προϊόντος της αντίδρασης PCR, με προσθήκη ddH₂O.
2. Έπειτα, για κάθε δείγμα που πρόκειται να αναλυθεί, αναμειγνύονται 198uL ρυθμιστικού διαλύματος με χρωστική, με 2uL ενισχυτή σήματος (100x Enhancer). Δύο επιπλέον αντιδράσεις, μία που αφορά στην ελάχιστη τιμή μέτρησης και μία στη μέγιστη, συνυπολογίζονται στο στάδιο αυτό. Οι μετρήσεις των δύο ακραίων τιμών σήματος είναι απαραίτητες για τη βαθμονόμηση (calibration) της συσκευής (Qubit 1.0, Thermo Fisher Scientific).
3. Ακολουθεί vortex και σύντομη φυγοκέντρηση του μείγματος.
4. Το μείγμα διαμοιράζεται σε προκαθορισμένα, ανάλογα με τον αριθμό των δειγμάτων διαφανή σωληνάρια τύπου Eppendorf, τα οποία είναι ειδικά κατασκευασμένα για μετρήσεις σε φθοριόμετρο. Η ποσότητα είναι συγκεκριμένη, και περιλαμβάνει 190uL μείγματος ανά σωληνάριο. Παράλληλα με τα σωληνάρια των δειγμάτων, προετοιμάζονται δύο επιπλέον σωληνάρια για τις δύο αντιδράσεις βαθμονόμησης της συσκευής.
5. Στο σωληνάριο που προορίζεται για βαθμονόμηση η οποία αφορά στην ελάχιστη τιμή σήματος προστίθενται 10uL ddH₂O.
6. Στο σωληνάριο που προορίζεται για βαθμονόμηση η οποία αφορά στη μέγιστη τιμή σήματος προστίθενται 10uL πρότυπου δείγματος DNA (10ng/uL).
7. Στα σωληνάρια που προορίζονται για μετρήσεις φθορισμού άγνωστων δειγμάτων προστίθενται 7uL ddH₂O και 3uL 1:10 αραιωμένου διαλύματος καθαρισμένου προϊόντος της αντίδρασης PCR.
8. Οι μετρήσεις (fluorescence quantity, Qf) αναφέρονται σε συγκεντρώσεις ng/mL και αφορούν αραιωμένες ποσότητες DNA, οπότε και ανάγονται σύμφωνα με τον παρακάτω τύπο.

$$c \text{ (PCR product) (ng/uL)} = ((Qf) \times (200/\gamma) \times (1/z))/1000$$

Όπου,

- Qf είναι η μέτρηση που έδωσε η συσκευή για ένα συγκεκριμένο δείγμα σε ng/mL

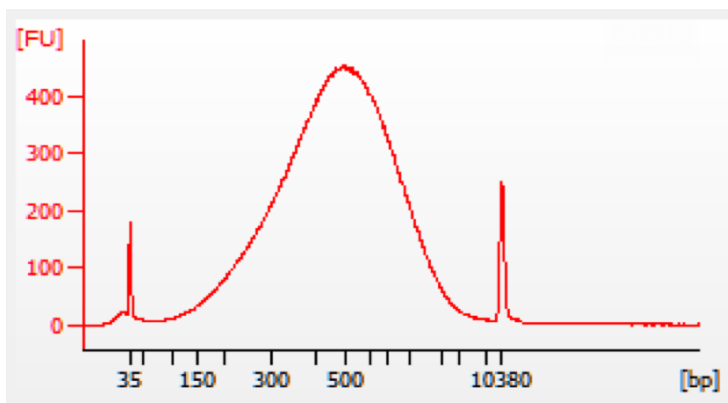
- γ είναι η ποσότητα 1:10 αραιωμένου διαλύματος καθαρισμένου προϊόντος της αντίδρασης PCR που προστέθηκε στην αντίδραση μέτρησης
- z είναι η αραιώση που πραγματοποιήθηκε στο καθαρισμένο προϊόν της αντίδρασης PCR

Οπότε η εξίσωση αυτομάτως μετατρέπεται στην ακόλουθη μορφή.

$$c \text{ (PCR product) (ng/uL) } = [(Qf) \times (200/3) \times (1/1/10)]/1000$$

3.4.2.5 Ποιοτική ανάλυση

Η ανάλυση της ποιότητας και του μεγέθους των κατακερματισμένων τμημάτων DNA, πραγματοποιήθηκε μέσω μικροηλεκτροφόρησης σε συσκευή Bioanalyzer, ακολουθώντας το πρωτόκολλο που περιγράφεται στο ολοκληρωμένο σύστημα High Sensitivity DNA kit, σύμφωνα με τις οδηγίες του κατασκευαστή (Agilent).



Εικόνα 21. Αποτέλεσμα μικροηλεκτροφόρησης σε συσκευή Bioanalyzer. Στο διάγραμμα δίνεται τιμή FU ανάλογη της ποσότητας του DNA και το αντίστοιχο μήκος του. Φαίνεται πως μετά την ενζυμική πέψη και τον καθαρισμό το μέσο μήκος των τμημάτων του DNA είναι 500 ζεύγη βάσεων.

3.4.2.6 Προετοιμασία βιβλιοθήκης δειγμάτων

Η προετοιμασία κάθε βιβλιοθήκης DNA, για προσδιορισμό της νουκλεοτιδικής ακολουθίας των υπό διερεύνηση γενομικών περιοχών, με αναλυτική συσκευή δεύτερης γενιάς, πραγματοποιήθηκε σύμφωνα με το πρωτόκολλο που περιγράφεται στο ολοκληρωμένο σύστημα Ovation Custom Target Enrichment System, ακολουθώντας τις οδηγίες του κατασκευαστή (NuGEN).

Τα δείγματα που αναλύονται ταυτόχρονα στη συσκευή, φέρουν διαφορετικές διακριτές τεχνητές ολιγονουκλεοτιδικές αλληλουχίες, για να ξεχωρίζουν κατά την ανάλυση. Αυτές οι αλληλουχίες προσαρτώνται στα άκρα των υπό μελέτη κατακερματισμένων νουκλεϊκών οξέων, σύμφωνα με το πρωτόκολλο παρασκευής των βιβλιοθηκών.

3.4.2.7 Προσδιορισμός αλληλουχίας

Ο προσδιορισμός της αλληλουχίας των κατακερματισμένων δειγμάτων DNA, που απαρτίζουν τη βιβλιοθήκη, πραγματοποιήθηκε σε συσκευή MiSeq (Illumina), ακολουθώντας το πρωτόκολλο, όπως αυτό περιγράφεται στο ολοκληρωμένο σύστημα MiSeq Reagent Kit v2 (2x500 cycles), σύμφωνα με τις οδηγίες του κατασκευαστή (Illumina). Για κάθε δείγμα παρήχθησαν δύο fastq αρχεία καθώς η αλληλούχιση ήταν paired-end, ένα <όνομα>_r1.fastq και ένα <όνομα>_r2.fastq.

3.4.2.8 Βιοπληροφορική ανάλυση των δεδομένων NGS Targeted-Seq δεδομένων

Η βιοπληροφορική ανάλυση των δεδομένων πραγματοποιήθηκε με την χρήση βιοπληροφορικών εργαλείων σε Docker containers. Τα images που χρειάστηκαν για την λειτουργία των αντίστοιχων containers βρέθηκαν από το ανοιχτό στο κοινό εναποθετήριο DockerHub, αναλυτικά τα διάφορα images δίνονται στον πίνακα 8. Η ενοποίηση των διαφόρων βιοπληροφορικών εργαλείων σε μια ροή εργασίας (pipeline) έγινε με τη χρήση CWL με τη βοήθεια του δωρεάν λογισμικού Rabix (<https://rabix.io>). Τέλος, για την εκτέλεση των ροών σε CWL χρησιμοποιήθηκε το rabix executor (Kaushik G. et al. 2016).

REPOSITORY	TAG	IMAGE ID
teoloup/myfiltervcf	latest	1bf9423974e8
ensemblorg/ensembl-vep	release_96.2	2672013fcddd
broadinstitute/gatk	4.1.2.0	9e737a9f562c
comics/trimmomatic	0.36	c32f9bdd077b
biocontainers/fastqc	0.11.5	35a358eb52f5
biocontainers/samtools	1.3.1	64c3e43e3483
biocontainers/bwa	0.7.15	5cf656e225f7

Πίνακας 8. Τα images των εργαλείων που χρησιμοποιήθηκαν καθώς και η συγκεκριμένη έκδοσή τους

Η ροή εργασίας όπως σχεδιάστηκε στο Rabix φαίνεται σε συγκεντρωτική μορφή στην εικόνα 22. Ο CWL κώδικας της ροής και των επιμέρους εργαλείων παρέχεται διαδικτυακά στο <https://github.com/teoloup/cwltools>.

Σημαντικό στοιχείο της ροής εργασίας αποτελεί η ευκολία χρήσης της χωρίς υψηλές απαιτήσεις γνώσεων ηλεκτρονικών υπολογιστών και βιοπληροφορικής καθώς και η φορητότητα της από ένα υπολογιστικό περιβάλλον σε ένα άλλο με ελάχιστη προσπάθεια χωρίς την ανάγκη εγκατάστασης όλων των εργαλείων και περιβαλλόντων εκ νέου αφού αυτό πραγματοποιείται αυτόματα από την ίδια τη ροή. Επιπλέον η δυνατότητα απόλυτης επαναληψιμότητας και η αναλυτική καταγραφή με σταθερό και εύκολα ιχνηλάσιμο τρόπο όλων των εργαλείων, εντολών, βάσεων δεδομένων και αποτελεσμάτων σε κάθε βήμα της ανάλυσης. Στοιχεία που αποτελούν την απαραίτητη βάση κάθε αναλυτικής ροής για την αξιοπιστία των αποτελεσμάτων.

ανιχνευθεί γνωστό και ως vcf αρχείο (variant call format file) είναι το τελικό προϊόν της αναλυτικής διαδικασίας και στην εικόνα φαίνεται ως ο γκρι κόμβος στο κάτω μέρος της εικόνας με όνομα output.

Πιο αναλυτικά:

Προσδιορισμός της ποιότητας των διαβασμάτων (quality control)

Ο έλεγχος της ποιότητας των διαβασμάτων που περιέχονται στα fastq αρχεία έγινε με το λογισμικό FastQC. Το λογισμικό αυτό αναλύει ένα μέρος του συνόλου των διαβασμάτων και παράγει μια αναφορά σε HTML η οποία δίνει αναλυτικά γραφήματα για ένα πλήθος χαρακτηριστικών όπως την ποιότητα του διαβάσματος για κάθε βάση ανά θέση, την αναλογία GC ανά θέση, το αν εμφανίζονται συχνά (και άρα όχι τυχαία) ολιγονουκλεοτιδικές αλληλουχίες οι οποίες συνήθως ανήκουν στις τεχνητές αλληλουχίες που χρησιμοποιήθηκαν κατά την κατασκευή των βιβλιοθηκών και θα πρέπει να αφαιρεθούν στα επόμενα βήματα ώστε να μην επηρεάσουν την ανάλυση.

Απομάκρυνση adapters (adapters clipping) και νουκλεοτιδίων χαμηλής ποιότητας (trimming of low quality nucleotides)

Η αφαίρεση τέτοιων τμημάτων έγινε με το πρόγραμμα Trimmomatic (Bolger A. et al. 2014). Το πρόγραμμα δέχεται ενδεικτικά ως δεδομένα εισόδου:

- τα fastq αρχεία και προαιρετικά
- ένα αρχείο με τα τεχνητά ολιγονουκλεοτίδια που χρησιμοποιήθηκαν για την κατασκευή της βιβλιοθήκης ώστε να αφαιρεθούν αν υπάρχουν
- επιλογές για την ελάχιστη ποιότητα των νουκλεοτιδίων στα άκρα κάθε διαβάσματος κάτω από την οποία τα νουκλεοτίδια θα αφαιρούνται.

Στη συνέχεια γίνεται ένας ακόμη έλεγχος των νέων fastq αρχείων ώστε να διαπιστωθεί η ποιότητά των διαβασμάτων μετά την επεξεργασία

Αντιπαραβολή των επεξεργασμένων διαβασμάτων επί του πρότυπου γονιδιώματος (alignment to reference)

Ο αλγόριθμος bwa (Burrows-Wheeler Aligner) έχει παρατηρηθεί πως αποδίδει τα βέλτιστα αποτελέσματα, όσον αφορά στην αντιπαραβολή των διαβασμάτων επί ενός πρότυπου γονιδιώματος. Η αντιπαραβολή έγινε επί της έκδοσης hg19 του ανθρώπινου πρότυπου γονιδιώματος.

Προετοιμασία του αρχείου αντιπαραβεβλημένων διαβασμάτων (alignment file) για ανίχνευση μεταλλάξεων

- Μετατροπή του αρχείου κειμένου, το οποίο προέκυψε κατά το προηγούμενο στάδιο της αναλυτικής διαδικασίας και περιέχει διαβάσματα αντιπαραβεβλημένα επί ενός πρότυπου γονιδιώματος (SAM alignment file), στη δυαδική μορφή αυτού (BAM alignment file).
- Ταξινόμηση κάθε αρχείου BAM, σύμφωνα με τις χρωμοσωμικές συντεταγμένες (genomic coordinates) των αντιπαραβεβλημένων διαβασμάτων, κατά αύξουσα σειρά.
- Δημιουργία ενός αρχείου καταλόγου (.bam.bai), που λειτουργεί ως ευρετήριο.

Και οι τρεις διεργασίες έγιναν με τη χρήση του προγράμματος samtools (Li H. et al. 2009).

Προσθήκη πληροφορίας στην κεφαλίδα του δείγματος

Η κεφαλίδα των διαβασμάτων ενός δείγματος περιλαμβάνει στοιχεία για την τεχνολογία και τη συσκευή NGS που χρησιμοποιήθηκε, για το μήκος και την κατεύθυνση διαβάσματος, για το εκάστοτε δείγμα που αναλύεται επί του παρόντος, καθώς και επιπρόσθετα στοιχεία, που είναι χρήσιμα τόσο για την αναλυτική διαδικασία, όσο και για τη μετέπειτα αρχειοθέτηση των δεδομένων.

Η προσθήκη των στοιχείων αυτών γίνεται με τη χρήση εργαλείων του προγράμματος Picardtools (<http://broadinstitute.github.io/picard/>) το οποίο έχει πλέον ενσωματωθεί στο Genome Analysis Toolkit (GATK) του Broad Institute (<https://software.broadinstitute.org/gatk>)

Αναβαθμόνωση των δεδομένων ποιότητας των νουκλεοτιδίων (Base Quality Score Recalibration, BQSR)

- Η διαδικασία αυτή χρησιμεύει στην αναβαθμόνωση της ποιότητας με την οποία διαβάστηκε κάθε νουκλεοτίδιο, κατά τη διάρκεια μίας αντίδρασης NGS, προτού αναδειχθούν οι νουκλεοτιδικές εξαλλαγές στα ακόλουθα στάδια της βιοπληροφορικής αναλυτικής πορείας.
- Με τον τρόπο αυτό πραγματοποιείται μία εξομάλυνση των δεδομένων στο σύνολό τους, αφού οι επι μέρους τιμές, που αφορούν στην ποιότητα του εκάστοτε διαβάσματος (quality scores), ενδεχομένως εμπεριέχουν συστηματικά λάθη, που μπορεί να οφείλονται στην ίδια τη συσκευή, όπως και σε άλλες παραμέτρους που επηρεάζουν το κάλεσμα των νουκλεοτιδίων σε κάθε γύρο διαβάσματος.
- Ακολουθεί, επισκόπηση της αναβαθμόνωσης, των τιμών της ποιότητας των διαβασμάτων και καταγραφή αυτών σε κατάλληλο αρχείο (τύπου pdf).
- Ο αλγόριθμος του BQSR καθώς και των εργαλείων που χρησιμοποιούνται παρακάτω για την ανάδειξη μοριακών εξαλλαγών, παρέχεται από το Ινστιτούτο BROAD, και έχουν αναπτυχθεί από το συγκεκριμένο φορέα, με στόχο την ανίχνευση μεταλλάξεων, σε δεδομένα αντιδράσεων NGS. Το σύνολό των εργαλείων που παρέχονται τελεί υπό τη γενική ονομασία εργαλειοθήκη ανάλυσης του γονιδιώματος (Genome Analysis Toolkit, GATK)(<https://software.broadinstitute.org/gatk>) (DePristo et al. 2011).

Ανάδειξη μεταλλάξεων με τον αλγόριθμο Mutect (GATK 4.1.2.0)

Η ανάδειξη των μεταλλάξεων έγινε με τη χρήση των εργαλείων Mutect2 (Cibulskis K. et al. 2013), LearnOrientationModel και FilterMutectCalls όπως περιγράφονται και προτείνονται στις βέλτιστες πρακτικές (Best Practises) για την ανίχνευση επίκτητων μεταλλάξεων μικρής κλίμακας στο GATK (<https://software.broadinstitute.org/gatk/best-practices/workflow?id=11146>).

Πιο συγκεκριμένα:

- Χρησιμοποιήθηκε το Mutect2 σε tumor only mode. Η επιλογή αυτή έγινε γιατί όπως αναφέρθηκε στο κεφάλαιο 3.3.1.3.3 η απόκτηση φυσιολογικού ιστού δεν ήταν

δυνατή τόσο τεχνικά όσο και οικονομικά αφού θα διπλασίαζε το κόστος της ανάλυσης.

- Φυσικό είναι πολλές από τις διαφορές των δειγμάτων που θα βρεθούν σε σχέση με το γονιδίωμα αναφοράς να μην είναι επίκτητες μεταλλάξεις (somatic mutations), αλλά πολυμορφισμοί (Single Nucleotide Polymorphisms, SNPs) ή μεταλλάξεις της γαμετικής σειράς (germline mutations). Για αυτό το λόγο το Mutect χρησιμοποιεί την βάση δεδομένων GnomAD (Konrad J. et al. BioRxiv 531210, <https://gnomad.broadinstitute.org>). Η βάση αυτή περιέχει δεδομένα από 125,748 αλληλουχίσεις του συνόλου των εξωνίων (Whole Exome Sequencing-WES) και 15,708 αλληλουχίσεις ολόκληρων γονιδιωμάτων (Whole Genome Sequencing-WGS) ανεξάρτητων ανθρώπων με σκοπό την ανίχνευση και καταγραφή το δυνατόν περισσότερων πολυμορφισμών στο γενικό πληθυσμό και σε υποπληθυσμούς. Σύμφωνα με τα δεδομένα αυτής το Mutect επισημαίνει όσες μοριακές εξαλλαγές θεωρεί πως δεν είναι somatic mutations.
- Επίσης για την αποφυγή όσο το δυνατόν περισσότερων τεχνητών μοριακών εξαλλαγών λόγω συστηματικών σφαλμάτων της τεχνικής δημιουργήθηκε, εκμεταλλευόμενοι το εργαλείο CreatePanelOfNormals, ένα σύνολο δεδομένων τεχνητών μοριακών εξαλλαγών (Panel of Normals) τα οποία ανευρίσκονται κοινά τουλάχιστον στο 50% των δειγμάτων (25 δείγματα). Το σύνολο αυτό χρησιμοποιήθηκε ώστε τέτοιες τεχνητές μοριακές εξαλλαγές να αφαιρεθούν από τα τελικά αποτελέσματα της ανάλυσης.
- Οι αναδεδειγμένες μεταλλάξεις καταγράφονται σε αρχείο τύπου vcf.
- Τέλος οι διαγραφές/προσθήκες νουκλεοτιδίων εφόσον υπάρχουν, κάθε επί μέρους γενομικής περιοχής ενδιαφέροντος, ευθυγραμμίζονται στο αριστερό τους άκρο (left alignment), μία πάγια τακτική που απαλείφει παρερμηνείες του εναρκτήριου σημείου, αυτού του είδους των μεταλλάξεων. Το πρόβλημα αυτό ανακύπτει, λόγω της αδυναμίας που παρουσιάζουν οι αλγόριθμοι ανάδειξης μεταλλάξεων, να αναγνωρίζουν επαναλαμβανόμενα νουκλεοτιδικά μοτίβα, στα άκρα των μεταλλάξεων, με αποτέλεσμα να μην αντιλαμβάνονται ότι πρόκειται για την ίδια μετάλλαξη, όπως επίσης ποιο είναι το εναρκτήριο σημείο αυτής. Η διαδικασία αυτή, αποτελεί μία σύμβαση, που ενισχύει την ομοιογένεια των δεδομένων που προέρχονται από διαφορετικές NGS πειραματικές διαδικασίες.

Φιλτράρισμα των αναδεδειγμένων μεταλλάξεων.

Το φιλτράρισμα των αναδεδειγμένων μεταλλάξεων έγινε με βάση τρία χαρακτηριστικά.

Το βάθος που έχει διαβαστεί η μετάλλαξη (Depth, DP), το ποσοστό αυτής (Allele frequency, AF) καθώς και με το αν η μετάλλαξη έχει μαρκαριστεί από το Mutect2 πως βρίσκεται στο σύνολο του Panel of normal.

Ως ελάχιστο βάθος διαβάσματος ορίστηκε το 100 (ελάχιστο βάθος διαβάσματος με το οποίο ανιχνεύθηκε μετάλλαξη, προηγουμένως επιβεβαιωμένη από το εργαστήριο μας μέσω HRMA και Sanger sequencing), ενώ ως ελάχιστο ποσοστό ανίχνευσης της μετάλλαξης στο δείγμα το 10% (ελάχιστο ποσοστό, 11%, με το οποίο ανιχνεύθηκε μετάλλαξη προηγουμένως επιβεβαιωμένη από το εργαστήριο μας μέσω HRMA και Sanger sequencing) (Γιαννόπουλος Α. 2019)). Ενώ πραγματοποιήθηκε και μια επιπλέον ανάλυση θέτοντας 5% ως ελάχιστο

ποσοστό ανίχνευσης μεταλλάξεων στο δείγμα, με στόχο τη λεπτομερέστερη μελέτη του μοριακού τοπίου των ασθενών.

Για την διεργασία αυτή καθώς δεν βρέθηκε κάποιο διαθέσιμο εργαλείο που να μπορεί να φιλτράρει ταυτόχρονα με τα τρία αυτά χαρακτηριστικά, δημιουργήθηκε ένα πρόγραμμα σε γλώσσα προγραμματισμού Perl που να εκτελεί αυτή την λειτουργία. Ενώ για την προσθήκη του στη ροή εργασίας δημιουργήθηκε Docker image αυτού το οποίο είναι διαθέσιμο στο <https://hub.docker.com/r/teoloup/myfiltervcf>. Τόσο ο πηγαίος κώδικας όσο και το Dockerfile για την δημιουργία του Docker image δίνονται στο ΠΑΡΑΡΤΗΜΑ Ι.

Ενημέρωση των αρχείων μεταλλάξεων με βιολογικά και κλινικά δεδομένα

- Στο σημείο αυτό πραγματοποιείται ενημέρωση των αρχείων μεταλλάξεων με βιολογικά και κλινικά στοιχεία, που υπάρχουν διαθέσιμα σε διεθνείς βάσεις δεδομένων, προκειμένου οι αναδεδειγμένες μεταλλάξεις να αξιολογηθούν κατάλληλα, σε επόμενο στάδιο της αναλυτικής διαδικασίας.
- Η ενημέρωση αυτή πραγματοποιείται μέσω του προγράμματος VEP (Variant Effect Predictor) της Ensembl (<https://www.ensembl.org/info/docs/tools/vep/index.html>) (McLaren W et al., 2016). Αρχικά προστίθενται στοιχεία που περιλαμβάνονται στην προεγκατεστημένη βάση δεδομένων του προγράμματος και περιέχει πληροφορίες για τον τύπο της μετάλλαξης, την αλλαγή στην κωδική αλληλουχία και την πρωτεϊνική εφόσον υπάρχει, καθώς και το αποτέλεσμα της πρόβλεψης δύο αλγορίθμων (SIFT, PolyPhen) για την σοβαρότητα της κάθε μετάλλαξης εφόσον αυτή είναι σημειακή.
- Ο SIFT (Sorting Intolerant from Tolerant) προβλεπτικός αλγόριθμος κάνει προβλέψεις για την επίπτωση σημειακών μεταλλάξεων στην λειτουργία μιας πρωτεΐνης βασιζόμενος στην ομολογία της αλληλουχίας και τις φυσικοχημικές ιδιότητες των αμινοξέων (<https://sift.bii.a-star.edu.sg/www/publications.html>) (Vaser R et al., 2016).
- Αντίστοιχα ο PolyPhen (Polymorphism Phenotyping) προβλεπτικός αλγόριθμος κάνει προβλέψεις για την επίπτωση σημειακών μεταλλάξεων στην λειτουργία μιας πρωτεΐνης βασιζόμενος στις πιθανές δομικές αλλαγές της πρωτεΐνης.
- Στη συνέχεια, προστίθενται στοιχεία, σχετικά με την επίπτωση των αναδεδειγμένων μεταλλάξεων, πληθυσμιακά δεδομένα, βιβλιογραφικές αναφορές και επιπλέον προβλεπτικοί αλγόριθμοι με την βοήθειά της ολοκληρωμένης βάσης δεδομένων dbNSFP3.5 (Liu X et al 2011, Liu X et al 2017) που λειτουργεί ως plugin του VEP.
- Μεταλλάξεις χωρίς επίπτωση στο γενετικό προφίλ του ασθενούς, όπως εκείνες που συντελούν σε συνώνυμα κωδικώνια (silent mutations), 5' ή 3' αμετάφραστες περιοχές ή σε περιοχές εσωνίων καθώς και μοριακές εξαλλαγές απομακρύνονται.

4 ΑΠΟΤΕΛΕΣΜΑΤΑ

Ο στοχευμένος προσδιορισμός της πρωτοδιάταξης των νουκλεϊκών οξέων με τεχνολογία δεύτερης γενιάς πραγματοποιήθηκε σε 51 ασθενείς εκ των οποίων:

- 8 ασθενείς με ΟΜΛ
- 9 ασθενείς με ΜΔΣ
- 33 ασθενείς με ΜΥΝ
 - 1 ασθενής με ΧΟΛ
 - 5 ασθενείς με ΙΘ
 - 11 ασθενείς με ΜΙ
 - 3 ασθενείς με ΣΜ
 - 5 ασθενείς με ΑΠ
 - 8 ασθενείς με αΜΥΝ
- 1 ασθενής με ΜΔΣ/ΜΥΝ (ΧΜΜΛ)

Σε 16 ασθενείς της μελέτης, είχαν ήδη αναγνωριστεί 29 μεταλλάξεις εκ των προτέρων, με μεθόδους που έχουν αναπτυχθεί και επιβεβαιωθεί στο εργαστήριο Αιματολογίας και περιλαμβάνουν post Real-time PCR HRMA και Sanger Sequencing (Γιαννόπουλος Α. 2019). Οι μεταλλάξεις αυτές χρησιμοποιήθηκαν ως θετικά δείγματα ελέγχου για την ευαισθησία της μεθόδου αλλά και για να προσδιορισθεί το ελάχιστο βάθος διαβάσματος (DP) και ποσοστό της μετάλλαξης στο δείγμα (AF) για την αναγνώριση μοριακών εξαλλαγών όπως αναφέρθηκε στο κεφάλαιο 3.4.2.8.

Σημαντικό γεγονός αποτελεί πως θέτοντας ως κατώτατο όριο βάθους διαβάσματος μίας μετάλλαξης το 100 και ελάχιστο ποσοστό εμφάνισης της μετάλλαξης στο δείγμα το 10%, η ανάλυση κατάφερε να αναδείξει και τις 29 επιβεβαιωμένες μεταλλάξεις στους 16 ασθενείς και επιπλέον 194 μεταλλάξεις.

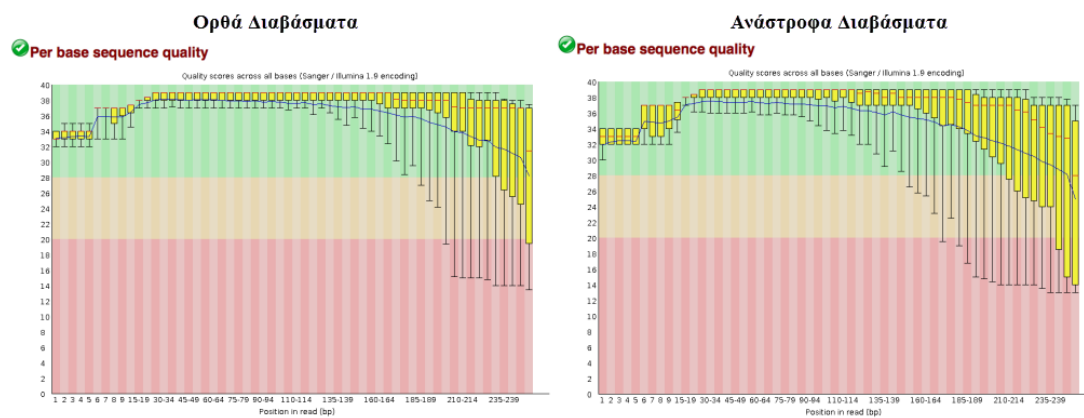
4.1 Ενδεικτικά αποτελέσματα της ανάλυσης

Στη συνέχεια θα περιγραφούν ενδεικτικά αποτελέσματα, από τη διαδικασία βιοπληροφορικής ανάλυσης δεδομένων, όπως αυτή περιεγράφηκε ήδη στο κεφάλαιο 3.4.2.8, με στόχο την ανάδειξη μεταλλάξεων, σε ένα από τα δείγματα με ΑΠ που διερευνήθηκαν. Στο συγκεκριμένο ασθενή είχε ήδη ανιχνευθεί η μετάλλαξη *JAK2V617F*. Η ανάλυση των δεδομένων NGS, ανέδειξε δύο επιπρόσθετες μεταλλάξεις, την c.1129A>C, p. p.T377P στο γονίδιο *TP53* μια μετάλλαξη άγνωστης ακόμη σημασίας σύμφωνα με βάσεις κλινικών δεδομένων όπως οι COSMIC (<https://cancer.sanger.ac.uk/cosmic/>) και CKB (<https://ckb.jax.org>), καθώς και την σιωπηλή μετάλλαξη c.627C>T στο γονίδιο *STAG*, η οποία και απομακρύνθηκε σε μεταγενέστερο αναλυτικό στάδιο, όπως αναφέρθηκε ήδη στο κεφάλαιο 3.4.2.8. Η ίδια αναλυτική πορεία ακολουθήθηκε ανεξαιρέτως και για τα υπόλοιπα δείγματα.

4.1.1 Προσδιορισμός της ποιότητας των διαβασμάτων (quality control, QC)

Ο προσδιορισμός της ποιότητας των διαβασμάτων του δείγματος πραγματοποιήθηκε όπως περιγράφηκε στο κεφάλαιο 3.4.2.8.

Δεδομένα από διάφορες στατιστικές αναλύσεις λαμβάνονται υπόψιν κατά τον έλεγχο ποιότητας των δεδομένων, με απώτερο στόχο τη βελτιστοποίηση της ποιότητας των διαβασμάτων και τη συνολική αξιολόγηση του πειράματος.



Εικόνα 23. Γραφική αναπαράσταση με ραβδογράμματα (boxplot analysis), όπου απεικονίζεται η μέση τιμή (mean value) και η τυπική απόκλιση (standard deviation, SD) της ποιότητας των νουκλεοτιδίων σε κάθε θέση, επί του συνόλου των ορθών και των ανάστροφων διαβασμάτων (paired-end sequencing), για ένα δείγμα, έπειτα από αποπολυπλεξία, αμέσως μετά τη λήψη των δεδομένων από τη συσκευή, χωρίς προηγούμενη επεξεργασία.

Η τιμή της ποιότητας, μέσω της παρακάτω λογαριθμικής εξίσωσης, μεταφράζεται σε πιθανότητα με την οποία κάθε νουκλεοτίδιο ενδεχομένως έχει διαβαστεί λανθασμένα.

$$Q = -10 \log_{10} P, \text{ όπου } P \text{ είναι η πιθανότητα λανθασμένα διαβασμένων νουκλεοτιδίων}$$

Για παραδείγματι, τιμή ποιότητας Q ίση με 20, ουσιαστικά σημαίνει, πως εάν ένα νουκλεοτίδιο σε μία συγκεκριμένη θέση, διαβάστηκε συνολικά εκατό φορές, τότε τις 99 διαβάστηκε σωστά, ενώ τη μία λάθος. Αυτό, εν δυνάμει, θα μπορούσε να εκληφθεί ως πραγματική νουκλεοτιδική αλλαγή στην αλληλουχία, ενώ δεν είναι τίποτα περισσότερο από ανακρίβεια της τεχνολογίας.

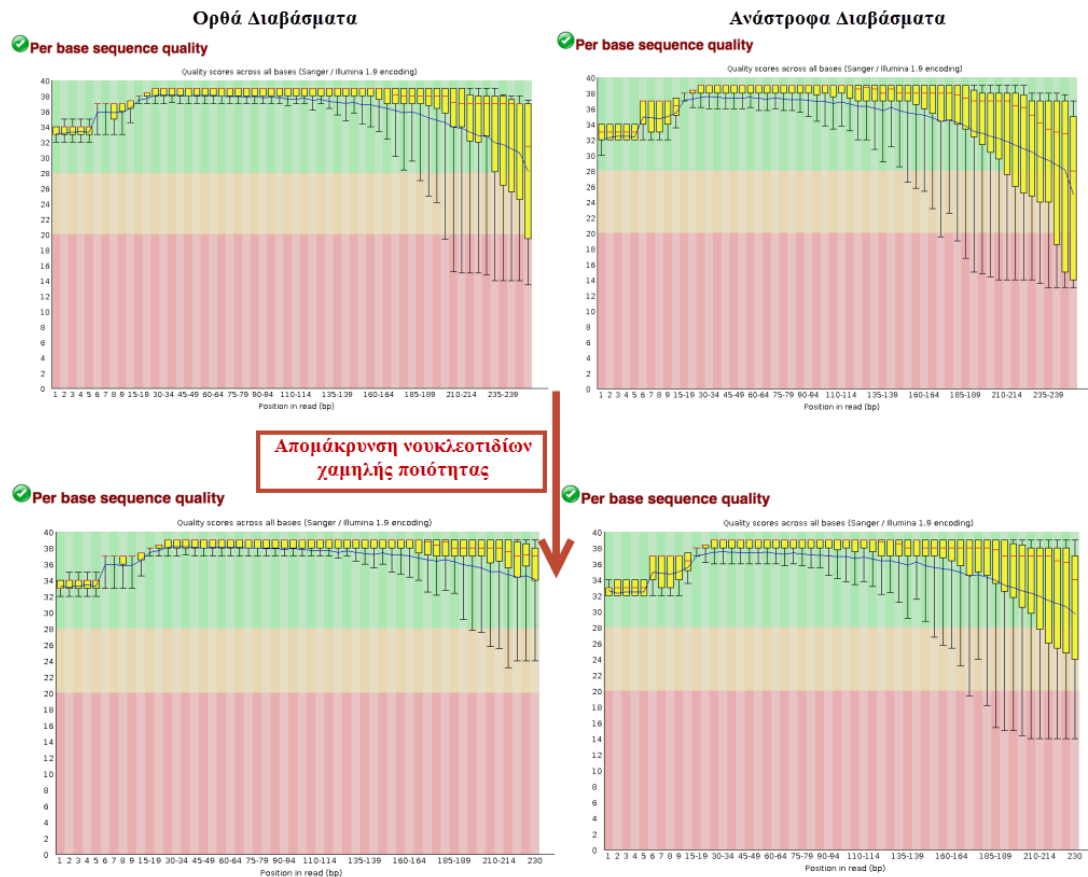
Για το λόγο αυτό, μεγαλύτερες τιμές ποιότητας, κοντά στο 30, μεταφράζονται σε πιο αξιόπιστα διαβάσματα. Αυτή είναι μία ιδιαίτερα σημαντική παράμετρος που λαμβάνεται υπόψιν κατά την ανάλυση δεδομένων NGS.

4.1.2 Επεξεργασία των διαβασμάτων

Ακολουθεί απομάκρυνση των adapters, όπου εμπεριέχεται η μεταβλητή αλληλουχία (barcode) για τη διάκριση των επιμέρους δειγμάτων, καθώς και επιπρόσθετες μη-μεταβλητές αλληλουχίες. Οι τελευταίες συμμετέχουν τόσο σε διάφορα στάδια PCR ενίσχυσης, κατά την προετοιμασία των βιβλιοθηκών δειγμάτων, όσο και στην πρόσδεση των νουκλεϊκών οξέων στο υπόστρωμα, κατά τον προσδιορισμό της πρωτοδιάταξης αυτών.

Επομένως, οι αλληλουχίες των adapters δεν είναι πλέον χρήσιμες, αφού έχουν ήδη εξυπηρετήσει το σκοπό τους και για το λόγο αυτό απορρίπτονται υπολογιστικά εφόσον υπάρχουν στα διαβάσματα.

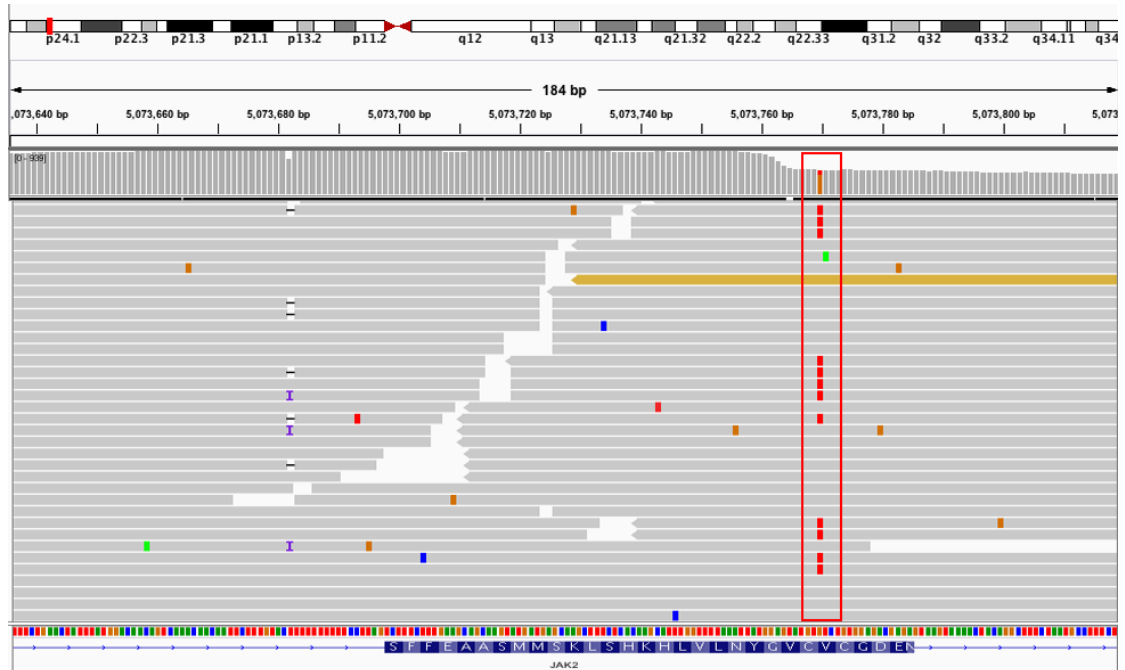
Επιπροσθέτως γίνεται απομάκρυνση (trimming) βάσεων στα άκρα των διαβασμάτων που έχουν Q-score μικρότερο του 20, έτσι ώστε να αυξηθεί η ποιότητα των διαβασμάτων και τελικά η αξιοπιστία των αποτελεσμάτων.



Εικόνα 24. Γραφική αναπαράσταση με ραβδογράμματα (boxplot analysis), όπου απεικονίζεται η μέση τιμή (mean value) και η τυπική απόκλιση (standard deviation, SD) της ποιότητας των νουκλεοτιδίων σε κάθε θέση, εκ του συνόλου των ορθών και ανάστροφων διαβασμάτων ενός δείγματος, έπειτα από απομάκρυνση των αλληλουχιών που ανήκουν σε adapters-barcodes.

4.1.3 Αντιπαραβολή των επεξεργασμένων διαβασμάτων επί του πρότυπου γονιδιώματος (alignment to reference)

Στη συνέχεια τα διαβάσματα αντιπαραβάλλονται επί ενός πρότυπου ανθρώπινου γονιδιώματος και πλέον μπορούν να οπτικοποιηθούν μέσω κατάλληλων υπολογιστικών προγραμμάτων, γνωστοί ως γενομικοί χάρτες, όπως το IGV (Integrative Genomics Viewer, Broad Institute). Με τον τρόπο αυτό γίνεται αντιπαραβολή της πρότυπης αλληλουχίας όλων των χρωμοσωμάτων, με τις υπό διερεύνηση νουκλεοτιδικές ακολουθίες (διαβάσματα).

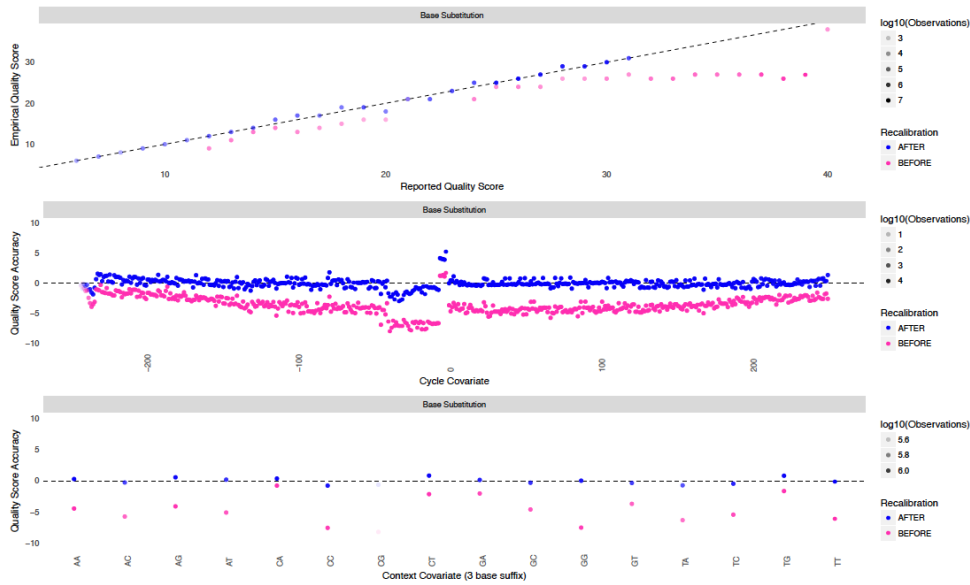


Εικόνα 25. Γενομικός χάρτης (*genetic map*) όπου αντιπαραβάλλονται τα διαβάσματα του προς ανάλυση δείγματος ως προς το πρότυπο γονιδίωμα (*reference genome*). Στο κάτω μέρος της εικόνας, με μπλε χρώμα, φαίνεται το υπό διερεύνηση γονίδιο, εν προκειμένω το *JAK2*. Τα διαβάσματα, με γκρι χρώμα στο κέντρο του γραφήματος, περικλείονται από εκκινητές και καλύπτουν εξώνια, αλλά και εσώνια, αφού οι εκκινητές προσδένονται ευρύτερα, ώστε να επιτύχουν πλήρη κάλυψη του γονιδίου. Με κόκκινο φαίνονται διαβάσματα που στην αλληλουχία τους έχουν νουκλεοτίδιο θυμίνη έναντι του φυσιολογικού της γουανίνης οδηγώντας στη ανίχνευση της γνωστής σημειακής μετάλλαξης *c.1849G>T* στο γονίδιο *JAK2* (*JAK2p.V617F*) στο συγκεκριμένο δείγμα.

4.1.4 Προετοιμασία του αρχείου αντιπαραβεβλημένων διαβασμάτων (*alignment file*) για ανίχνευση μεταλλάξεων

Κατά την προετοιμασία του αρχείου των αντιπαραβεβλημένων διαβασμάτων πραγματοποιούνται τα βήματα «Προετοιμασία του αρχείου αντιπαραβεβλημένων διαβασμάτων (*alignment file*) για ανίχνευση μεταλλάξεων», «Προσθήκη πληροφορίας στην κεφαλίδα του δείγματος» και «Αναβαθμονόμηση των δεδομένων ποιότητας των νουκλεοτιδίων (*Base Quality Score Recalibration, BQSR*)» όπως περιγράφονται στο κεφάλαιο 3.5.1.2.8.

Δεδομένα από την αναβαθμονόμηση της ποιότητας των διαβασμάτων του δείγματος δίνονται στην παρακάτω εικόνα.



Εικόνα 26. Κανονικοποίηση των δεδομένων ποιότητας των νουκλεοτιδίων, με βάση τη δυσκολία που παρουσιάζουν οι επιμέρους περιοχές του γονιδιώματος ως προς τον προσδιορισμό της αλληλουχίας τους.

4.1.5 Ανάδειξη μεταλλάξεων

Η ανάδειξη των μεταλλάξεων έγινε όπως περιγράφεται στο κεφάλαιο 3.5.1.2.8 σύμφωνα με τις βέλτιστες πρακτικές (Best Practices) για την ανίχνευση επίκτητων μεταλλάξεων μικρής κλίμακας στο GATK (<https://software.broadinstitute.org/gatk/best-practices/workflow?id=11146>) με τη χρήση του εργαλείου ανίχνευσης επίκτητων μεταλλάξεων Mutect2 (DePristo et al. 2011).

Με την ανάλυση αυτή προέκυψαν 23 πιθανές μεταλλάξεις.

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	PV_sample
chr2	25505275	.	C	A,T	.	base_qual;clustered_events;germline;	CONTQ=93;DP=3 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,2,1,1,1:0.33:
chr2	25505294	.	G	GA	.	clustered_events;germline;haplotype;	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:PGT:PID:PS:SB		0/1,15,2,0.158:1
chr2	25505295	.	G	GAGCGATCC	.	clustered_events;germline;haplotype;	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:PGT:PID:PS:SB		0/1,15,2,0.155:1
chr3	128204877	.	C	G	.	germline;orientation;strand_bias	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,4,6,0.584:10
chr4	106157912	.	C	CCCACCTGCT	.	germline;haplotype;weak_evidence	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:PGT:PID:PS:SB		0/1,1,6,2,0.161:1
chr4	106157914	.	C	CTTGTGCAA	.	germline;haplotype;weak_evidence	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:PGT:PID:PS:SB		0/1,1,6,2,0.161:1
chrX	39933339	.	A	G	.	base_qual;germline;panel_of_normals	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,0,125,0.992:
chrX	39933357	.	C	T	.	germline;orientation	CONTQ=93;DP=2 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,7,12,0.651:1
chrX	123179178	.	C	T	.	orientation;strand_bias;weak_evidence	CONTQ=93;DP=5 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,49,3,0.074:5
chr9	5073770	.	G	T	.	orientation	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,142,37,0.205
chr11	119170224	.	A	G	.	orientation;strand_bias;weak_evidence	CONTQ=93;DP=5 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,91,3,0.041:9
chr12	111856004	.	C	T	.	germline;orientation;strand_bias;weak_evidence	CONTQ=93;DP=4 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,2,2,0.499:4
chr12	111856181	.	G	A	.	germline;orientation;strand_bias;weak_evidence	CONTQ=93;DP=3 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,1,2,0.600:3
chr12	111856251	.	A	G	.	germline;orientation;strand_bias;weak_evidence	CONTQ=93;DP=2 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,0,2,0.750:2
chr12	111884608	.	T	C	.	germline;orientation;panel_of_normals	CONTQ=93;DP=5 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,33,18,0.353:
chr12	111884965	.	A	G	.	base_qual;germline;orientation;strand	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,9,2,0.227:11
chr12	111886069	.	A	G	.	orientation;strand_bias;weak_evidence	CONTQ=93;DP=5 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,88,3,0.043:9
chr17	7572980	.	T	G	.	base_qual;orientation;strand_bias	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,146,16,0.083
chr17	7578645	.	C	T	.	germline;panel_of_normals	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,0,132,0.993:
chr17	7578711	.	CTTTTTTT	C,CTT,CTTT,C	.	germline;multiallelic;panel_of_normals	CONTQ=93;DP=4 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,2/3/4,0.5,8,;
chr17	7579588	.	G	T	.	base_qual;germline;orientation;strand	CONTQ=93;DP=7 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,5,1,0.246:6
chr17	74732507	.	A	G	.	germline;orientation;panel_of_normals	CONTQ=93;DP=5 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,0,52,0.982:5
chr20	31023303	.	T	C	.	germline;orientation;strand_bias;weak_evidence	CONTQ=93;DP=1 GT:AD:AF:DP:F1R2:F2R1:SB		0/1,12,2,0.186:1

Εικόνα 27. Τμήμα ενός αρχείου μεταλλάξεων (vcf file). Από αριστερά προς τα δεξιά καταγράφονται το χρωμόσωμα, η θέση σε αυτό όπου ανιχνεύτηκε μία νουκλεοτιδική εξαλλαγή, ο αριθμός καταχώρησης αυτής στη βάση δεδομένων πολυμορφισμών dbSNP, εφόσον είναι καταχωρημένη, το κανονικό νουκλεοτίδιο, η μετάλλαξη, η ποιότητα με την οποία κατέστηκε η μετάλλαξη (τιμή που το εργαλείο Mutect2 δεν παράγει), τα φίλτρα του εργαλείου για τα οποία είναι θετική η μετάλλαξη και στη συνέχεια διάφορα ποιοτικά και ποσοτικά δεδομένα.

4.1.6 Φιλτράρισμα σύμφωνα με ποιοτικά και ποσοτικά δεδομένα

Το αρχικό φιλτράρισμα έγινε σύμφωνα με τα κριτήρια που αναφέρθηκαν στο κεφάλαιο 3.4.2.8 «Φιλτράρισμα των αναδεδειγμένων μεταλλάξεων» και συγκεκριμένα με AF μεγαλύτερο ή ίσο του 10%.

Από αυτό προέκυψαν τρεις (3) μεταλλάξεις που πληρούν τα κριτήρια καθώς από τις 23 μεταλλάξεις:

- Πέντε (5) συμπεριλαμβάνονταν στο τεχνητό σύνολο μοριακών εξαλλαγών (Panel of Normals) της μεθόδου
- Δεκατρία (13) είχαν βάθος διαβάσματος μικρότερο από 100
- Δύο (2) είχαν ποσοστό μετάλλαξης μικρότερο από 10%.

4.1.7 Ενημέρωση του αρχείου μεταλλάξεων με βιολογικά και κλινικά δεδομένα

Η ενημέρωση του αρχείου γίνεται με την βοήθεια του προγράμματος VEP (Variant Effect Predictor) της Ensembl (<https://www.ensembl.org/info/docs/tools/vep/index.html>) (McLaren W et al., 2016) και της ολοκληρωμένης βάσης δεδομένων dbNSFP3.5 (Liu X et al 2011, Liu X et al 2017) που λειτουργεί ως plugin του VEP, όπως αναφέρθηκε στο κεφάλαιο 3.4.2.8.

Από τις σχολιασμένες πλέον μεταλλάξεις του αρχείου, μπορούμε να εξάγουμε χρήσιμες πληροφορίες για την αλλαγή που προκαλούν τόσο σε νουκλεοτιδικό όσο και σε πρωτεϊνικό επίπεδο. Επίσης για την καλύτερη αξιολόγησή τους το αρχείο ενημερώνεται με βιβλιογραφικές αναφορές και κωδικούς καταχωρήσεων αυτών των μεταλλάξεων σε γνωστές βάσεις δεδομένων όπου έχουν καταχωρηθεί ώστε να γίνει χειροκίνητα η αξιολόγησή τους σε διεθνείς επιστημονικές βάσεις δεδομένων (COSMIC, UNIPROT, dbSNP κ.α). Επίσης παρέχονται πληθυσμιακά δεδομένα για τα ποσοστά ανίχνευσης αυτών των μεταλλάξεων στον γενικό πληθυσμό.

Επιπρόσθετα, αξιολογή προσθήκη για σημειακές μεταλλάξεις, που οδηγούν σε αλλαγή της αμινοξικής ακολουθίας, αποτελούν μια πληθώρα αλγορίθμων πρόβλεψης της παθογένειας τέτοιων μεταλλάξεων. Οι αλγόριθμοι αυτοί βασίζονται σε φυλογενετικά δεδομένα συντήρησης της νουκλεοτιδικής και αμινοξικής ακολουθίας (evolutionary approach), σε βιολογικά χαρακτηριστικά (biological principles), όπως η περιοχή της πρωτεΐνης που επηρεάζεται από τη μετάλλαξη και οι φυσικοχημικές ιδιότητες του αμινοξέως ή ακόμα και συνδυασμό αυτών των δύο προσεγγίσεων. Τέλος, υπάρχουν αλγόριθμοι, οι οποίοι λειτουργούν με πιο ετερογενείς προσεγγίσεις ή ακόμη και συνδυάζοντας περισσότερους από έναν από τους προηγούμενους αλγόριθμους σε μία προβλεπτική μηχανή.

Τα αποτελέσματα των προβλεπτικών αυτών αλγορίθμων για τις δύο παρανοηματικές μεταλλάξεις που βρέθηκαν στο προς ανάλυση δείγμα δίνονται στα παρακάτω αραχνοειδή διαγράμματα.



Εικόνα 28. Αραχνοειδή διαγράμματα (spider graphs) με τις αξιολογήσεις της παθογένειας των μεταλλάξεων. Σε κάθε περίπτωση, αριστερά παρουσιάζονται απλοί αλγόριθμοι που βασίζονται σε δεδομένα που σχετίζονται με τις επιμέρους νουκλεοτιδικές και αμινοξικές αλλαγές, καθώς και με δεδομένα φυλογένεσης-συντήρησης, ενώ δεξιά απεικονίζονται πιο σύνθετοι αλγόριθμοι που συνδυάζουν χαρακτηριστικά των προηγούμενων. Βαθμολογίες κοντά στο 0 ή στο 1, χαρακτηρίζουν μεταλλάξεις οι οποίες σύμφωνα με τον κάθε αλγόριθμο ενδεχομένως δεν συμβάλλουν ή συμβάλλουν, αντίστοιχα στην παθογένεια της νόσου. Από επάνω προς τα κάτω απεικονίζονται οι προβλέψεις για τις μεταλλάξεις JAK2p.V617F και TP53T377P. Παρατηρούμε πως η γνωστή κύρια μετάλλαξη JAK2V617F έχει τιμές πρόβλεψης της παθογένειας κοντά στο 1, ενώ η μετάλλαξη TP53p.T377P σύμφωνα με τους προβλεπτικούς αλγόριθμους φαίνεται να μην έχει υψηλό σκορ παθογένειας.

4.1.8 Φιλτράρισμα μεταλλάξεων σύμφωνα με βιολογικά και κλινικά δεδομένα.

Τελευταίο βήμα της ανάλυσης αποτελεί το φιλτράρισμα των μεταλλάξεων σύμφωνα με τα δεδομένα, με τα οποία εμπλουτίστηκαν, από την παραπάνω ανάλυση. Το εργαλείο VEP δίνει την δυνατότητα επιλογής μεταλλάξεων με συγκεκριμένα χαρακτηριστικά όπως πληθυσμιακά ποσοστά, σοβαρότητα επίπτωσης της μετάλλαξης, είδος της μετάλλαξης κ.α.

Στην παρούσα μελέτη επιλέχθηκαν μόνο μεταλλάξεις, οι οποίες έχουν αντίκτυπο στην πρωτεϊνική ακολουθία, καθώς προς το παρόν η μεγαλύτερη πλειοψηφία των μεταλλάξεων που μπορούν να αξιολογηθούν επαρκώς από την τρέχουσα βιβλιογραφία και τους προβλεπτικούς αλγόριθμους είναι τέτοιου τύπου (δηλαδή παρανοηματικές μεταλλάξεις και μικρές προσθήκες ή διαγραφές νουκλεοτιδίων).

Έτσι, στο υπό μελέτη δείγμα, καταλήγουμε σε δύο μεταλλάξεις, την JAK2V617F και την c.1129A>C (p.T377P) στο γονίδιο TP53.

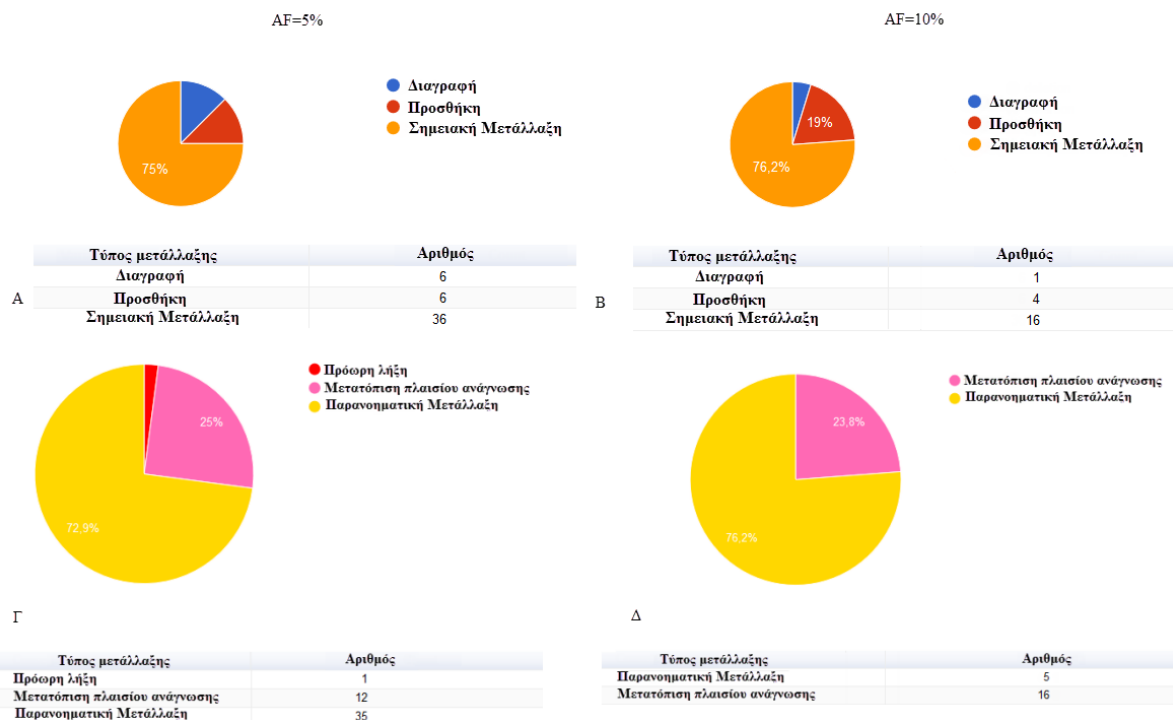
Η JAK2V617F αποτελεί μια κύρια μετάλλαξη με γνωστή παθογόνο δράση όπως αναφέρθηκε και στο κεφάλαιο 1.2 με τους προβλεπτικούς αλγόριθμους να συμφωνούν όπως φαίνεται και στην εικόνα 27. Αντίθετα η μετάλλαξη c.1129A>C (p.T377P) στο γονίδιο TP53 τόσο από την βιβλιογραφία (<https://www.ncbi.nlm.nih.gov/snp/rs774269719>) όσο και από του προβλεπτικούς αλγόριθμους, φαίνεται να μην διαδραματίζει σημαντικό ρόλο στην νόσο.

4.2 Συγκεντρωτικά αποτελέσματα

Παρακάτω δίνονται συγκεντρωτικά τα αποτελέσματα ανά ομάδα ασθενών σύμφωνα με δύο διαφορετικά ελάχιστα ποσοστά ανίχνευσης της μετάλλαξης στο δείγμα (AF) όπως αναφέρθηκε και στο κεφάλαιο 3.4.2.8 «Φιλτράρισμα των αναδεδειγμένων μεταλλάξεων», ένα με ελάχιστο ποσοστό το 5% και ένα με 10%.

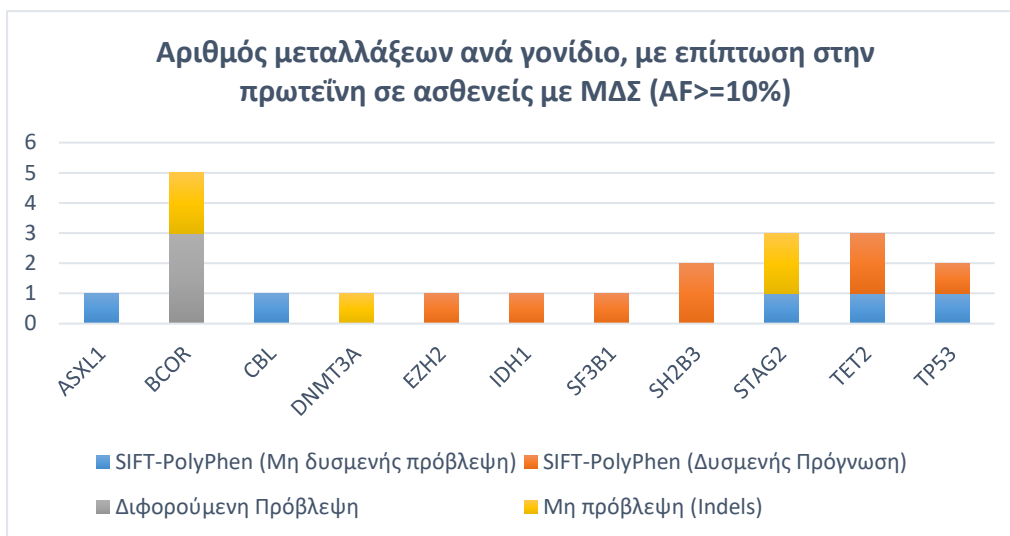
4.2.1 Ανίχνευση μεταλλάξεων σε ασθενείς με ΜΔΣ.

Συνολικά αναλύθηκαν εννέα (9) δείγματα ασθενών με ΜΔΣ. Στις παρακάτω εικόνες δίνονται αναλυτικά τα αποτελέσματα των μεταλλάξεων που ανιχνεύθηκαν κατά τις δύο αναλύσεις με διαφορετικά όρια AF.



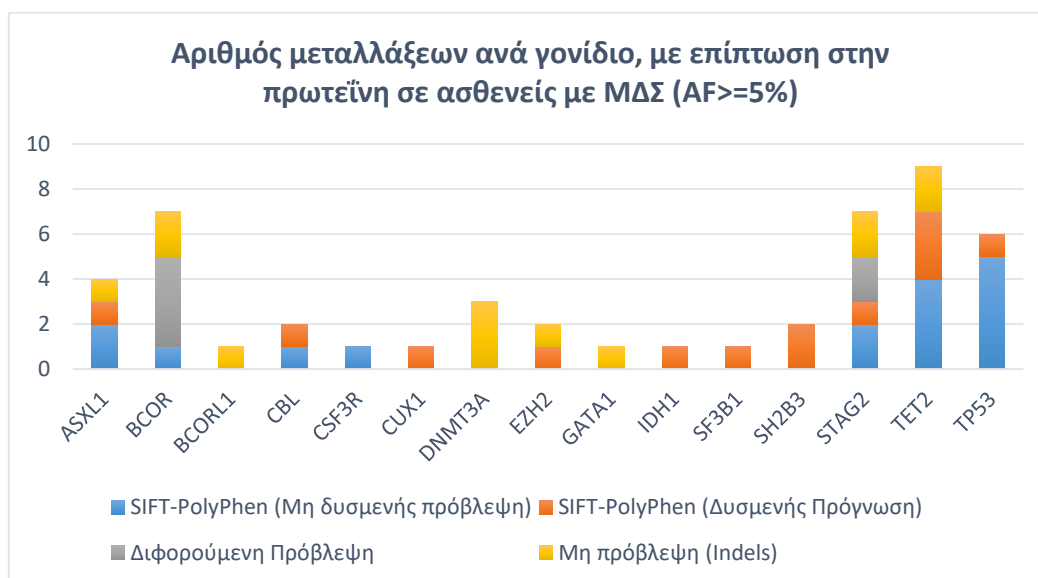
Εικόνα 29. Διαγράμματα απεικόνισης του τύπου και του πλήθους των μεταλλάξεων για τις δύο διαφορετικές αναλύσεις σε ασθενείς με ΜΔΣ. Αριστερά (A, Γ) αποδίδονται τα αποτελέσματα με

όριο AF 5% ενώ δεξιά (B, Δ) τα αποτελέσματα με AF 10%. Οι σημειακές παρανοηματικές μεταλλάξεις (missense mutations) ανέρχονται σε ποσοστό 75% με το υπόλοιπο 25% να είναι μεταλλάξεις διαγραφής ή προσθήκης νουκλεοτιδίων που οδηγούν σε μετατόπιση του πλαισίου ανάγνωσης (open reading frame, ORF).



Εικόνα 30. Ραβδογράμμο με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με ΜΔΣ ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Από την ανάλυση με κατώτατο όριο AF 10% αναδείχθηκαν 21 μεταλλάξεις στο σύνολο των 9 ασθενών με μέσο όρο 2,3 μεταλλάξεις/ασθενή.



Εικόνα 31. Ραβδογράμμο με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με ΜΔΣ ανά γονίδιο με όριο AF 5%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

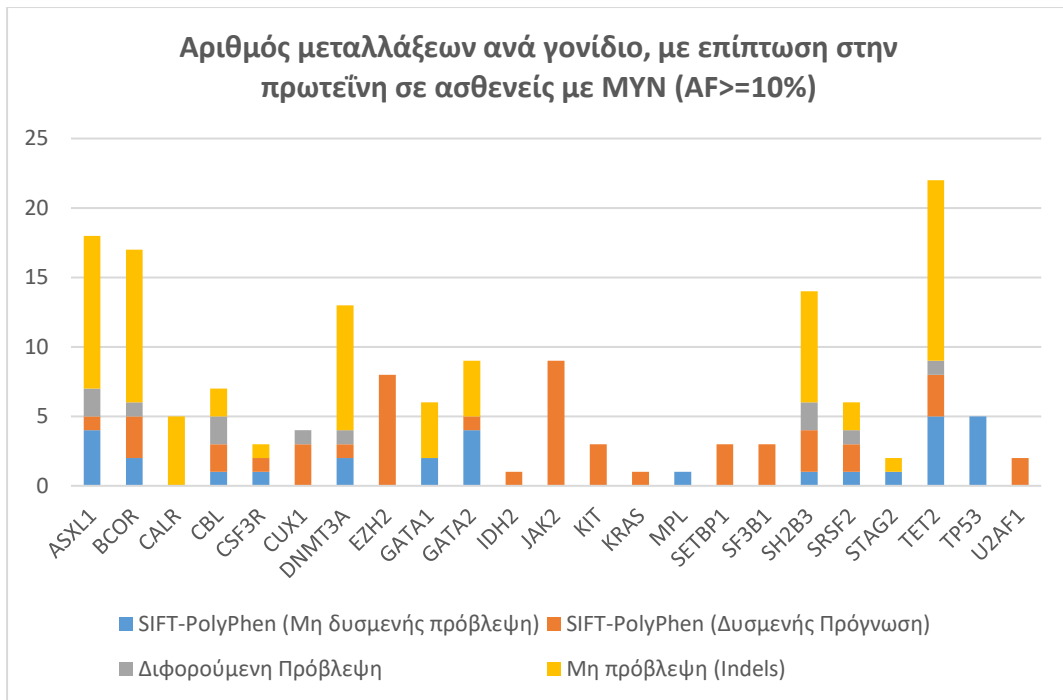
Από την ανάλυση με κατώτατο όριο AF 5% αναδείχθηκαν 48 μεταλλάξεις στο σύνολο των 9 ασθενών με μέσο όρο 5,3 μεταλλάξεις/ασθενή.

4.2.2 Ανίχνευση μεταλλάξεων σε ασθενείς με ΜΥΝ.

Συνολικά αναλύθηκαν τριάντα τρία (33) δείγματα ασθενών με ΜΥΝ. Στις παρακάτω εικόνες δίνονται αναλυτικά τα αποτελέσματα των μεταλλάξεων που ανιχνεύθηκαν κατά τις δύο αναλύσεις με διαφορετικά όρια AF.

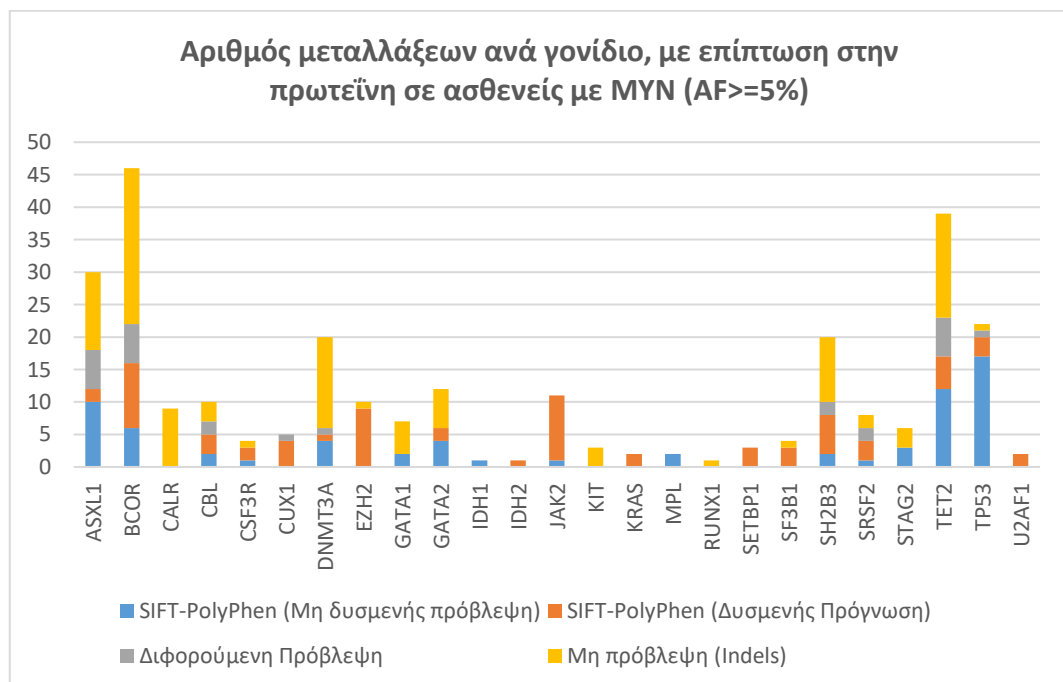


Εικόνα 32. Διαγράμματα απεικόνισης του τύπου και του πλήθους των μεταλλάξεων για τις δύο διαφορετικές αναλύσεις σε ασθενείς με ΜΥΝ. Αριστερά (A, Γ) αποδίδονται τα αποτελέσματα με όριο AF 5% ενώ δεξιά (B, Δ) τα αποτελέσματα με AF 10%. Οι σημειακές μεταλλάξεις (point mutations) ανέρχονται σε ποσοστό 65,4% και 60,6% αντίστοιχα, με ένα 3% αυτών στο σύνολο να οδηγούν στην δημιουργία πρόωρων κωδικονίων λήξης (nonsense mutations). Οι υπόλοιπες μεταλλάξεις είναι προσθήκης-διαγραφής νουκλεοτιδίων όπου το 10% δεν έχει επίπτωση στο ORF (inframe indels).



Εικόνα 33. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με MYN ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Από την ανάλυση με κατώτατο όριο AF 10% αναδείχθηκαν 162 μεταλλάξεις στο σύνολο των 33 ασθενών με μέσο όρο 4,9 μεταλλάξεις/ασθενή.

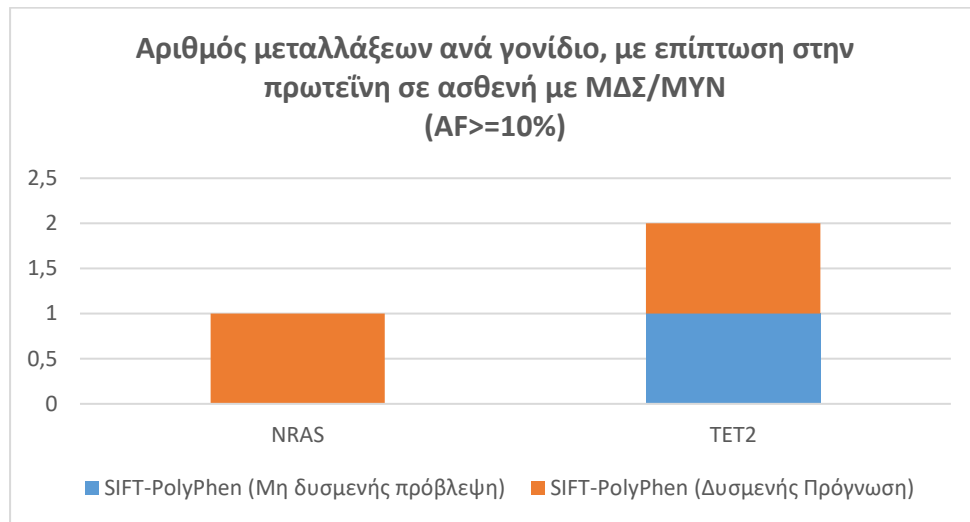


Εικόνα 34. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με MYN ανά γονίδιο με όριο AF 5%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Από την ανάλυση με κατώτατο όριο AF 5% αναδείχθηκαν 278 μεταλλάξεις στο σύνολο των 33 ασθενών με μέσο όρο 8,4 μεταλλάξεις/ασθενή.

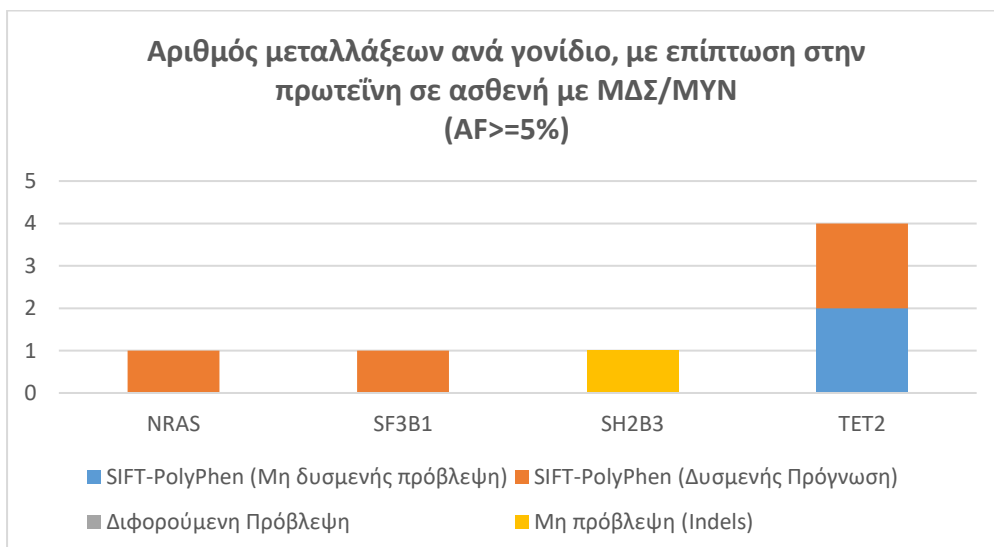
4.2.3 Ανίχνευση μεταλλάξεων σε ασθενή με ΜΔΣ/ΜΥΝ.

Στην ομάδα των ΜΔΣ/ΜΥΝ μελετήθηκε μόνο ένας ασθενής για τον οποία τα αποτελέσματα των αναλύσεων δίνονται στις παρακάτω εικόνες.



Εικόνα 35. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στον ασθενή με ΜΔΣ/ΜΥΝ ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Ανιχνεύθηκαν συνολικά τρεις (3) σημειακές παρανοηματικές μεταλλάξεις με κατώτατο όριο AF 10%.

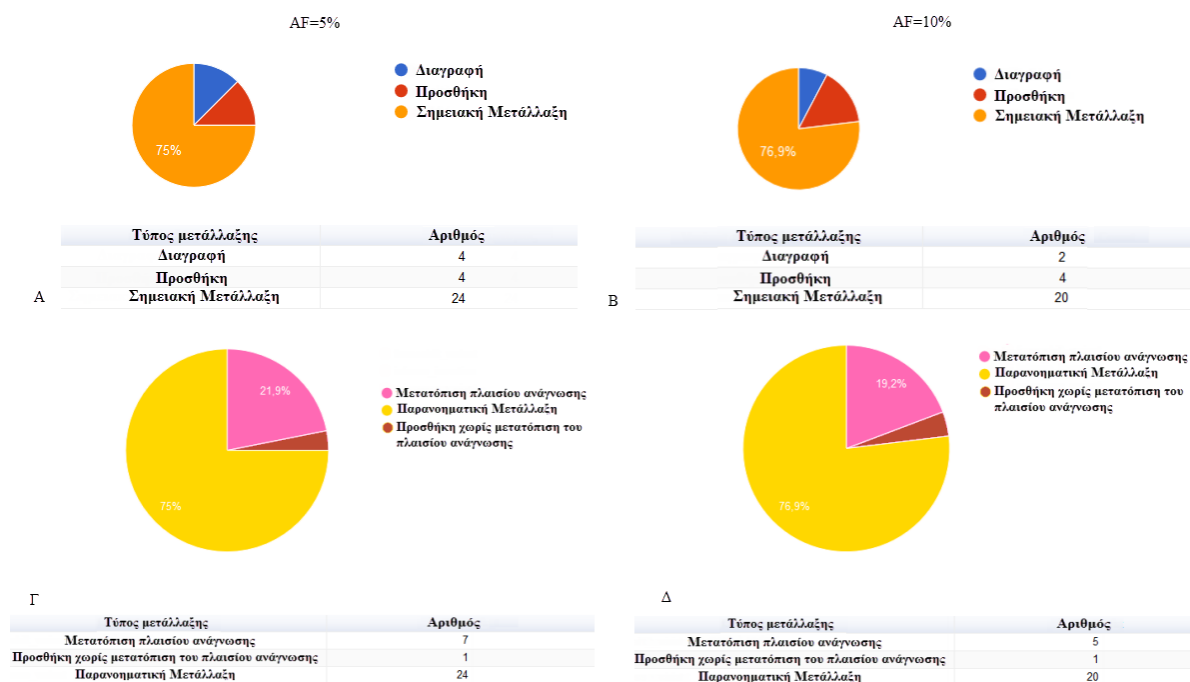


Εικόνα 36. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στον ασθενή με ΜΔΣ/ΜΥΝ ανά γονίδιο με όριο AF 5%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

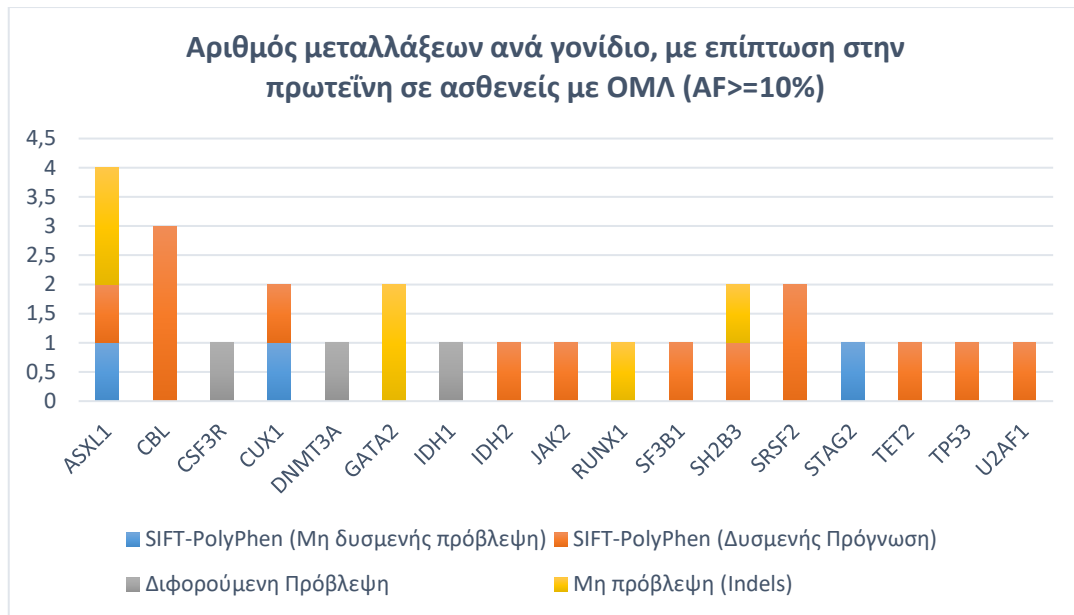
Ανιχνεύθηκαν συνολικά με κατώτατο όριο AF 5% τέσσερις (4) σημειακές παρανοηματικές μεταλλάξεις και μια (1) διαγραφή νουκλεοτιδίων που οδηγεί σε μετατόπιση του ORF.

4.2.4 Ανίχνευση μεταλλάξεων σε ασθενείς με ΟΜΛ.

Συνολικά αναλύθηκαν οκτώ (8) δείγματα ασθενών με ΟΜΛ. Στις παρακάτω εικόνες δίνονται αναλυτικά τα αποτελέσματα των μεταλλάξεων που ανιχνεύθηκαν κατά τις δύο αναλύσεις με διαφορετικά όρια AF.

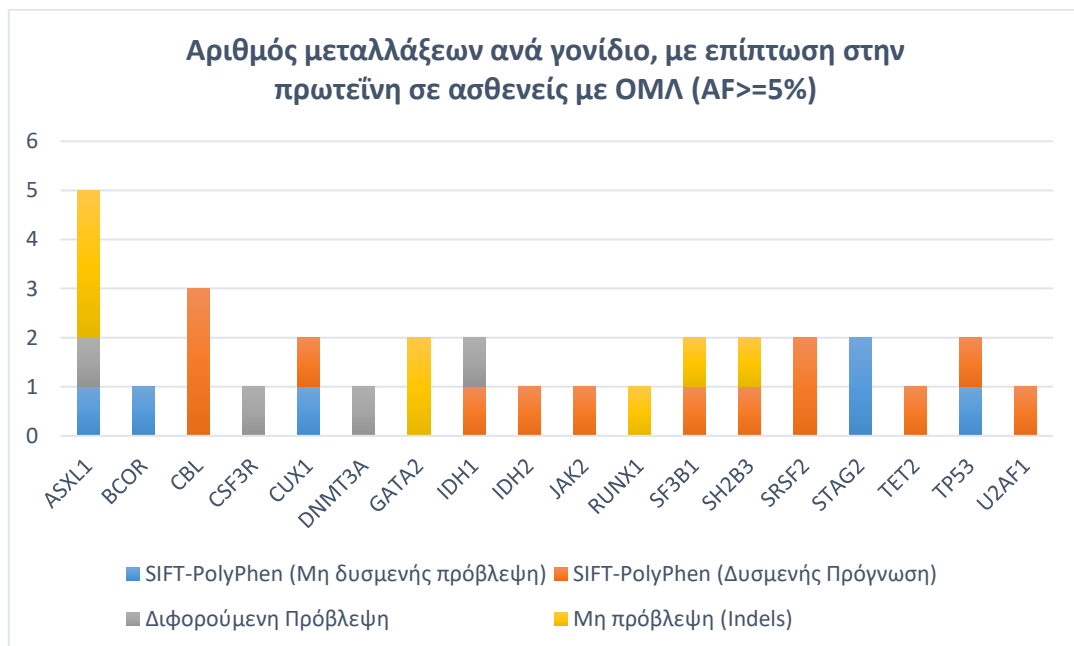


Εικόνα 37. Διαγράμματα απεικόνισης του τύπου και του πλήθους των μεταλλάξεων για τις δύο διαφορετικές αναλύσεις σε ασθενείς με ΟΜΛ. Αριστερά (Α, Γ) αποδίδονται τα αποτελέσματα με όριο AF 5% ενώ δεξιά (Β, Δ) τα αποτελέσματα με AF 10%. Οι σημειακές μεταλλάξεις (point mutations) ανέρχονται σε ποσοστό περίπου 75% και στις δύο περιπτώσεις με το υπόλοιπο 25% να είναι μεταλλάξεις διαγραφής ή προσθήκης νουκλεοτιδίων που ως επί το πλείστον οδηγούν σε μετατόπιση του ORF.



Εικόνα 38. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με ΟΜΛ ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Από την ανάλυση με κατώτατο όριο AF 10% αναδείχθηκαν 26 μεταλλάξεις στο σύνολο των 8 ασθενών με μέσο όρο 3,25 μεταλλάξεις/ασθενή.



Εικόνα 39. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με ΟΜΛ ανά γονίδιο με όριο AF 5%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, PolyPhen) του εργαλείου σχολιασμού VEP.

Από την ανάλυση με κατώτατο όριο AF 5% αναδείχθηκαν 32 μεταλλάξεις στο σύνολο των 8 ασθενών με μέσο όρο 4 μεταλλάξεις/ασθενή.

5 ΣΥΖΗΤΗΣΗ

5.1 Σύγκριση αποτελεσμάτων της ανάλυσης μεταξύ των δύο διαφορετικών ορίων ποσοστού εμφάνισης της μετάλλαξης στο δείγμα.

Όπως έχει ήδη περιγραφεί προηγουμένως (κεφάλαια 3, 4), η ανάλυση πραγματοποιήθηκε με δυο διαφορετικά όρια AF. Το πρώτο όριο θεσπίστηκε στο 10%, ως ελάχιστη τιμή AF με την οποία μία συγκεκριμένη μετάλλαξη αναδείχθηκε μέσω NGS, ενώ παράλληλα επιβεβαιώθηκε με άλλες αναλυτικές διαδικασίες (post Real-time PCR HRMA και Sanger Sequencing) (Γιαννόπουλος Α. 2019). Σε αυτό το αναλυτικό όριο πετύχαμε μέγιστη ευαισθησία και ικανοποιητική ακρίβεια. Η τελευταία τιμή μεταφράζεται ως το πλήθος των αληθώς θετικών γενετικών εξαλλαγών, σε συνδυασμό με τις ψευδώς θετικές καταχωρήσεις. Σε αναζήτηση μεταλλαγών με χαμηλότερο AF επιλέχθηκε ένα δεύτερο χαμηλότερο όριο 5%.

Κατώτατο όριο AF 10%

Κατηγορία αιμοποιητικής διαταραχής /Τύπος Μετάλλαξης	Σημειακή	Προσθήκης	Διαγραφής	Σύνολο
ΜΔΣ	16	4	1	21
ΜΔΣ/ΜΥΝ	3	0	0	3
ΜΥΝ	105	56	12	173
ΟΜΛ	20	4	2	26
Σύνολο	144	64	15	223

Πίνακας 9. Πλήθος μεταλλάξεων που αναδείχθηκαν στις διαφορετικές κατηγορίες αιμοποιητικών διαταραχών που μελετήθηκαν, ταξινομημένες ανά τύπο μεταλλαγής, αλλά και συνολικά, σε 10% κατώτατο AF.

Κατώτατο όριο AF 5%

Κατηγορία αιμοποιητικής διαταραχής /Τύπος Μετάλλαξης	Σημειακή	Προσθήκης	Διαγραφής	Σύνολο
ΜΔΣ	36	6	6	48
ΜΔΣ/ΜΥΝ	6	0	1	7
ΜΥΝ	182	65	31	278
ΟΜΛ	24	4	4	32
Σύνολο	248	75	42	365

Πίνακας 10. Πλήθος μεταλλάξεων που αναδείχθηκαν στις διαφορετικές κατηγορίες αιμοποιητικών διαταραχών που μελετήθηκαν, ταξινομημένες ανά τύπο μεταλλαγής, αλλά και συνολικά, σε 5% κατώτατο AF.

Από τους παραπάνω πίνακες παρατηρούμε μια σημαντική διαφορά στο πλήθος των μεταλλάξεων που ανιχνεύονται, με μια μόνο μείωση κατά 5% στο όριο του AF.

Κατηγορία αιμοποιητικής διαταραχής	Σημειακή	Προσθήκης	Διαγραφής	Σύνολο
ΜΔΣ	+20 (+55,6%)	+2 (+33,3%)	+5 (+88,3%)	+27 (+56,3%)
ΜΔΣ/ΜΥΝ	+3 (+50%)	+0 (0%)	+1 (+100%)	+4 (+57,1%)
ΜΥΝ	+77 (+42,3%)	+9 (+13,8%)	+19 (+61,3%)	+105 (+37,8%)
ΟΜΛ	+4 (+16,7%)	+0 (0%)	+2 (+50%)	+6 (+18,8%)
Σύνολο	+104 (+41,9%)	+11 (+14,7%)	+27 (+64,3%)	+142 (+38,9%)

Πίνακας 11. Διαφορές στο πλήθος των μεταλλάξεων που αναδείχθηκαν στις διαφορετικές κατηγορίες αιμοποιητικών διαταραχών που μελετήθηκαν σε απόλυτες τιμές και επί τις εκατό, ταξινομημένες ανά τύπο μεταλλαγής, αλλά και συνολικά, κατά τη μετάβαση από κατώτατο AF 10% σε 5%.

Μεταβιβάζοντας το ελάχιστο όριο AF από 10% σε 5% (πίνακας 11), καταγράφονται συνολικά 142 επιπρόσθετες γενετικές αλλαγές (αύξηση 38,9%) οι οποίες δεν είχαν προηγουμένως παρατηρηθεί. Από αυτές, οι 104 (+41,9%) αφορούν σε σημειακές μεταλλαγές και οι 27 (+64,3%) σε μεταλλάξεις διαγραφής νουκλεοτιδίων.

Υπάρχουν διάφορες ερμηνείες που μπορεί να οδηγήσουν σε μία τέτοια παρατήρηση:

1. Ανίχνευση αληθώς θετικών μεταλλάξεων σε χαμηλό ποσοστό λόγω κλωνικότητας. Για παράδειγμα, μία μετάλλαξη σε ετεροζυγωτία, σε ποσοστό 10% επί του συνόλου των κυττάρων, θεωρητικά θα ανιχνευόταν σε AF=5% (10% διπλοειδών κυττάρων *0,5 λόγω ετεροζυγωτίας= 5% AF).
2. Στην δυσκολία που αντιμετωπίζουν οι αλγόριθμοι στοίχισης (bwa), όπως επίσης οι αλγόριθμοι ανάδειξης μοριακών εξαλλαγών (Mutect2), κατά τη χαρτογράφηση και κατά την ανάδειξη μεταλλάξεων διαγραφής ή προσθήκης νουκλεοτιδίων, αντίστοιχα. Τα χαρακτηριστικά αυτά, αυξάνουν το πλήθος των ψευδώς θετικών αποτελεσμάτων (Krigerød AB et al. 2016).
3. Η ολοκληρωμένη λύση που χρησιμοποιήθηκε για την κατασκευή των βιβλιοθηκών είναι βασισμένη στην τεχνολογία νουκλεοτιδικής επέκτασης (amplicon based protocol). Βασίζεται στην επιλογή των επιθυμητών τμημάτων προς αλληλούχιση, μέσω ενίσχυσης αυτών με αλυσιδωτή αντίδραση πολυμεράσης (Polymerase Chain Reaction, PCR). Η τεχνική PCR ενέχει πάντα την πιθανότητα εισαγωγής λανθασμένων νουκλεοτιδίων, τα οποία ενισχύονται μέσω της διαδικασίας αυτής καθαυτής. Ως αποτέλεσμα, κατά τον προσδιορισμό νουκλεοτιδικής ακολουθίας, το λάθος αυτό αναδεικνύεται ως μετάλλαξη, μάλιστα με AF εκθετικά εξαρτώμενο από τον κύκλο της PCR αντίδρασης στον οποίο εισήχθη το λάθος.

5.2 Ανάλυση των μοριακών ευρημάτων

Το πρωτόκολλο NGS και η αναλυτική διαδικασία που αναπτύχθηκαν στην παρούσα εργασία, αποσκοπούν στην ανάδειξη μεταλλάξεων, σε γονίδια που συμμετέχουν σε διαφορετικές κυτταρικές διεργασίες, ενώ ταυτόχρονα εμπλέκονται στην εμφάνιση και εξέλιξη ετερογενών κατηγοριών μυελικών διαταραχών.

Με τη μέθοδο αυτή ταυτοποιήσαμε μεταλλάξεις, σε ασθενείς για τους οποίους προηγουμένως, στο εργαστήριο μας, δεν είχαμε καταφέρει να αποκομίσουμε μοριακά δεδομένα, με τη χρήση συμβατικών τεχνικών, οι οποίες μελετούν μεμονωμένες γενομικές περιοχές. Παράλληλα εντοπίσαμε τις ήδη αναδειγμένες από το εργαστήριο μας μεταλλάξεις, στο σύνολό τους.

Συγκεκριμένα ανιχνεύσαμε με την μέθοδο του στοχευμένου προσδιορισμού νουκλεοτιδικής ακολουθίας (targeted sequencing), συνολικά, 2519 μοριακές εξαλλαγές. Ακολούθησε φιλτράρισμα με βάση τα ακόλουθα κριτήρια:

- Το αν ανήκουν σε τεχνητά συστηματικά σφάλματα (PanelOfNormals)
- Το βάθος διαβάσματος (DP >=100)
- Το ποσοστό ανίχνευσης κάθε μετάλλαξης (AF>=10%)
- Την επίδραση στο πρωτεϊνικό προϊόν(coding regions)

Έτσι, καταλήξαμε σε 223 μεταλλάξεις στους 53 ασθενείς που μελετήθηκαν, καθώς οι υπόλοιπες 2296 (90% του συνόλου) δεν πληρούν τα κριτήρια ελέγχου, αφού πιθανότατα αποτελούν ψευδώς θετικά αποτελέσματα.

Από τις 223, οι 29, είχαν επιβεβαιωθεί ήδη από το εργαστήριο μας, ενώ οι 194 είναι πιθανές νέες γενετικές αλλαγές, χωρίς ωστόσο να αποκλείεται το ενδεχόμενο ορισμένες από αυτές να αποτελούν ψευδώς θετικά στοιχεία ακόμη και στην ανάλυση με όριο AF 10%, όπως περιγράφεται χαρακτηριστικά στο κεφάλαιο 5.1.

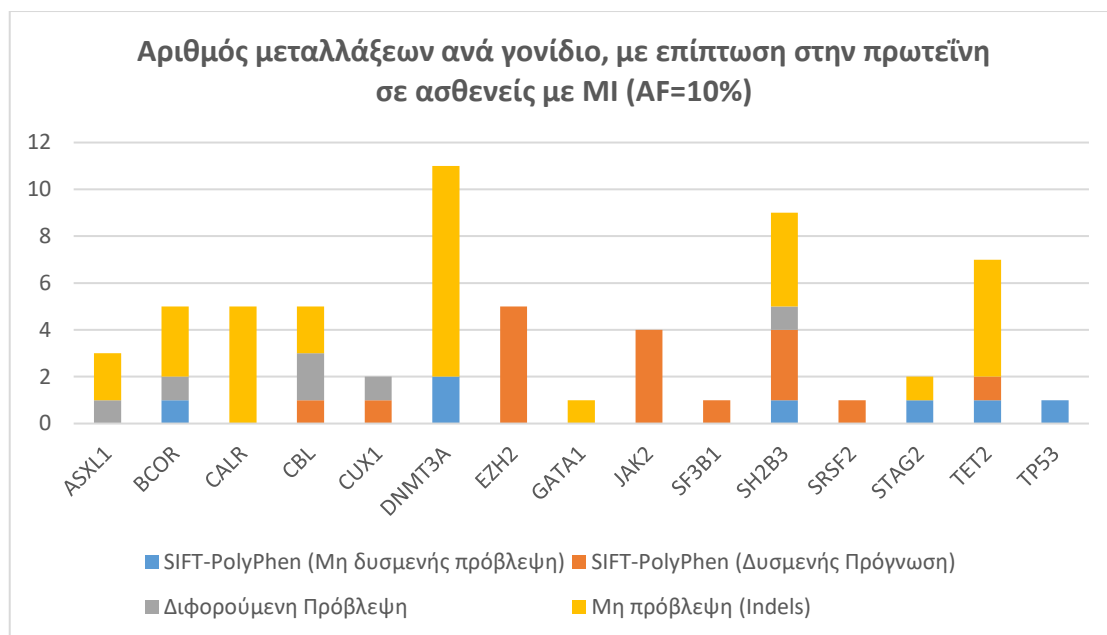
Στο σημείο αυτό θα πρέπει να τονιστεί ιδιαίτερα η σημασία της απουσίας αντίστοιχου φυσιολογικού ιστού, για κάθε δείγμα και πως αυτή συμβάλλει στην ανάδειξη επιπρόσθετων νουκλεοτιδικών εξαλλαγών (κεφάλαιο 3.4.2.8). Ως εκ τούτου δεν είναι εφικτό να αφαιρεθούν από την ανάλυση τυχόν κληρονομήσιμες μεταλλάξεις και πιθανοί πολυμορφισμοί με χαμηλή εμφάνιση στον γενικό πληθυσμό.

5.2.1 Σχολιασμός αποτελεσμάτων σε ομάδες ασθενών

Στον μοναδικό ασθενή με ΧΜΜΛ, ανιχνεύθηκε μια σημειακή παθογόνος μετάλλαξη στο γονίδιο *NRAS* (G12C) και δύο σημειακές μεταλλάξεις στο γονίδιο *TET2* (V218M και H1778R), με τους προβλεπτικούς αλγόριθμους να χαρακτηρίζουν τη δράση των δύο τελευταίων ως μη παθογόνο και πιθανώς παθογόνο, αντίστοιχα. Μεταλλάξεις σε αυτά τα γονίδια θα ήταν δύσκολο να ανιχνευθούν στο εργαστήριο μας με ήδη εγκαθιδρυμένες τεχνικές, εξαιτίας του μεγέθους των γονιδίων και της έλλειψης συχνά μεταλλαγμένων περιοχών σε εκείνα. Η συνεμφάνιση μεταλλάξεων στα γονίδια που προαναφέρθηκαν έχει ήδη περιγραφεί στη βιβλιογραφία σε ασθενείς με ΜΔΣ/ΜΥΝ (Meggendorfer et al. 2018).

Ιδιαίτερο ενδιαφέρον εμφανίζουν οι ασθενείς με Μυελοϊνωση (MI) και εκείνοι με Οξεία Μυελογενή Λευχαιμία (ΟΜΛ).

Στην πρώτη κατηγορία ανιχνεύθηκαν 63 μεταλλάξεις με επίπτωση στην πρωτεϊνική ακολουθία, στο σύνολο 11 ασθενών (μέσο όρο 5,6 μεταλλάξεις/ασθενή).



Εικόνα 40. Ραβδογράμμα με τις μεταλλάξεις που ανιχνεύθηκαν στο σύνολο των ασθενών με MI ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, Polyphen) του εργαλείου σχολιασμού VEP.

Η MI αναπτύσσεται άλλοτε *de novo* και άλλοτε ως εξέλιξη της ΑΠ (post-PV MF) ή συχνότερα, της ΙΘ (post-ET MF), με ίδια περίπου κλινική εικόνα. Η έναρξη της νόσου οφείλεται σε ανεξέλεγκτη υπερπλασία των μεγακαρυοκυττάρων, τα οποία εκκρίνουν κυτταροκίνες που προάγουν την ανάπτυξη των ινοβλαστών και κατά συνέπεια τη δημιουργία ινών κολλαγόνου, στο μυελό των οστών. Σε μοριακό επίπεδο ανιχνεύονται συχνά μεταλλάξεις κλασικές των MYN (*JAK2*, *MPL* και *CALR*), αλλά και επιπρόσθετες τροποποιητικές μεταλλάξεις, σε γονίδια που εμπλέκονται σε ποικίλες κυτταρικές διεργασίες, όπως τα *ASXL1*, *EZH2*, *TET2*, *IDH1/IDH2*, *SRSF2*, *SF3B1* και άλλα. Επιπρόσθετες μελέτες έχουν δείξει πως σε νόσους όπως η ΑΠ, απώλεια της λειτουργίας του ενζύμου *DNMT3A* σε συνδυασμό με την μετάλλαξη *JAK2V617F* επάγουν την εκτροπή της νόσου προς μυελοϊνώση (Jacquelin S. et al. 2018, Radek C. Skoda, 2018).

Στην ομάδα που μελετήσαμε εντοπίσαμε ένα πλήθος μεταλλάξεων σε 15 γονίδια όπως φαίνεται και στην εικόνα 39. Ανάμεσά τους, διακρίνονται κύριες μεταλλάξεις στα γονίδια *JAK2* (V617F) και *CALR* (μεταλλάξεις προσθήκης/διαγραφής νουκλεοτιδίων) καθώς και πληθώρα τροποποιητικών μεταλλάξεων. Σημαντικό είναι να τονιστεί πως σε τρεις περιπτώσεις η μετάλλαξη *JAK2V617F* συνοδεύονταν από μια τουλάχιστον μετάλλαξη στο γονίδιο *DNMT3A*, με πιθανή απώλεια λειτουργίας του ενζύμου που κωδικοποιεί, υποδηλώνοντας πως σε αυτές τις περιπτώσεις ενδεχομένως δεν πρόκειται για *de novo* εκδήλωση της νόσου, αλλά πιθανότητα ως εκτροπή από ΑΠ ή ΙΘ.

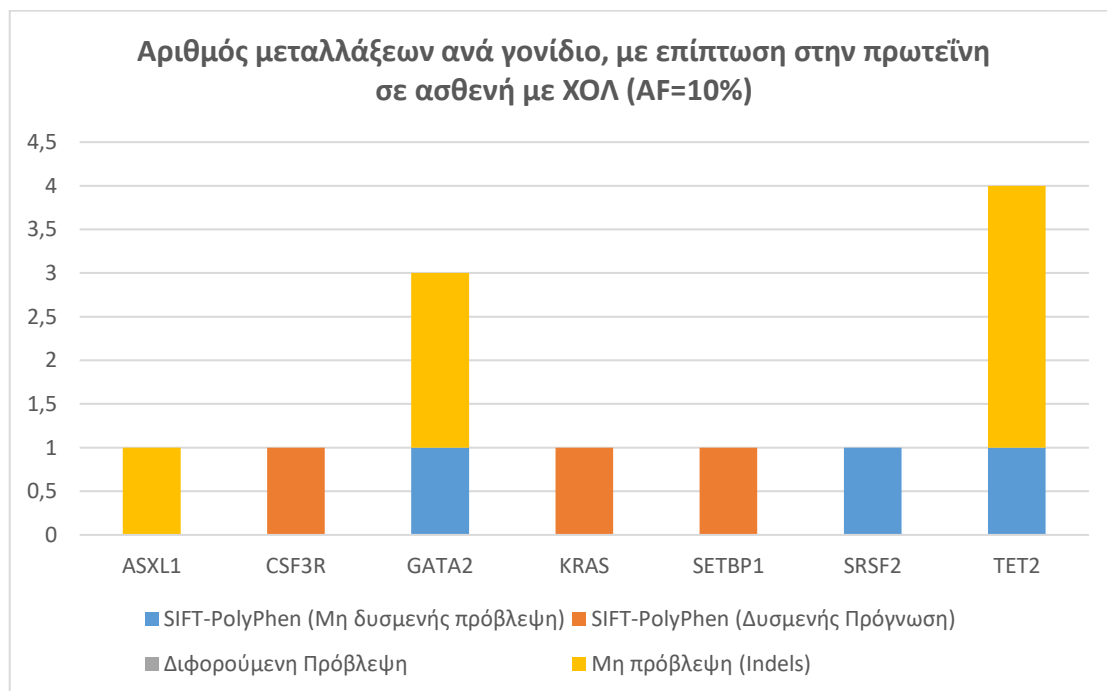
Στην ομάδα της ΟΜΛ ανιχνεύσαμε συνολικά 26 μεταλλάξεις με επίπτωση στην πρωτεϊνική ακολουθία, σε σύνολο 9 ασθενών που μελετήθηκαν (μέσος όρος 2,9 μεταλλάξεις/ασθενής). Η χαμηλότερη τιμή μέσου όρου μεταλλάξεων, συγκριτικά πάντα με άλλες νεοπλασίες, μπορεί να αποδοθεί στο γεγονός πως στην ΟΜΛ παρατηρούνται επιπρόσθετες ευρείες κυτταρογενετικές αναδιατάξεις (Blau O. 2015), όπως επίσης στο σχεδιασμό του πάνελ, όπου

τα γονίδια επελέγησαν με στόχο τη βαθύτερη κατανόηση των χρόνιων μυελικών νεοπλασιών. Γονίδια με μεγάλη συχνότητα εμφάνισης μεταλλάξεων, σχεδόν αποκλειστικά στην ΟΜΛ, όπως τα *NPM1* (16.7%), *FLT3* (11.3%), *ABL2* (14.29%), *ETV6* (17.38%), *CEBPA* (4.19%), δεν συμπεριλήφθηκαν στο πάνελ γονιδίων (AACR Project GENIE Consortium 2017).

Παρ' όλα αυτά, ως επί το πλείστον, ανιχνεύθηκαν σημειακές μεταλλάξεις, με πιθανή παθογόνο δράση, σύμφωνα με τους προβλεπτικούς αλγόριθμους και μάλιστα σε γονίδια τα οποία εμφανίζονται αρκετά συχνά στην βιβλιογραφία (*ASXL1*, *IDH1*, *IDH2*, *RUNX1*, *TP53* και γονίδια που εμπλέκονται στην ωρίμανση του mRNA), με δυσμενή μάλιστα πρόγνωση για την εξέλιξη της νόσου σύμφωνα με την βιβλιογραφία (DiNardo CD. 2016).

Αξιοσημείωτη είναι η περίπτωση ασθενούς με Χρόνια Ουδετεροφιλική Λευχαιμία (ΧΟΛ) που αναλύθηκε.

Η ΧΟΛ αποτελεί ένα ιδιαίτερος σπάνιο κλωνικό μυελικό νεόπλασμα, το οποίο χαρακτηρίζεται από αυξημένο αριθμό ώριμων ουδετερόφιλων στο περιφερικό αίμα, υπερπλασία του μυελού των οστών και ηπατοσπληνομεγαλία. Χωρίς θεραπεία η νόσος εξελίσσεται σε ΟΜΛ σε διάστημα μόλις 12-24 μηνών με μέση επιβίωση των ασθενών μεταξύ 21-24 μηνών (Elliott and Tefferi 2016). Πρόσφατες μελέτες ανέδειξαν μεταλλάξεις στο γονίδιο που κωδικοποιεί στον υποδοχέα των κοκκιοκυττάρων (*CSF3R*), στην εκδήλωση της νόσου, σε ποσοστό άνω του 80% των ασθενών με ΧΟΛ, συμβάλλοντας έτσι στην ορθή ταξινόμηση αυτής της σπάνιας νόσου (Dao et al. 2017, Maxson and Tyner 2017).



Εικόνα 41. Ραβδογράμμο με τις μεταλλάξεις που ανιχνεύθηκαν σε ασθενή με ΧΟΛ ανά γονίδιο με όριο AF 10%. Με διαφορετικό χρώμα απεικονίζεται το πλήθος των μεταλλάξεων σύμφωνα με την πρόγνωση της παθογένειας τους από τους δύο βασικούς αλγόριθμους (SIFT, Polyphen) του εργαλείου σχολιασμού VEP.

Στον συγκεκριμένο ασθενή ανιχνεύσαμε 12 μεταλλάξεις εκ των οποίων 4 είχαν ήδη αναδειχθεί με χρήση συμβατικών τεχνικών από το εργαστήριο μας (*ASXL1* p.G646fs*12,

CSF3R p.T618I, *SETBP1* p.G870S και *SRSF2* p.P95L). Με την παρούσα διαδικασία επιβεβαιώσαμε τις προαναφερθείσες και ακόμα 8 μεταλλάξεις στα γονίδια *GATA2*, *KRAS* και *TET2*, όπως παρουσιάζεται στην εικόνα 41. Το πλήθος των μεταλλάξεων αλλά και η βιολογική σημασία των γονιδίων όπου αυτές εμφανίζονται, αντικατοπτρίζονται από τη σοβαρότητα της νόσου και την ταχεία εκτροπή αυτής προς ΟΜΛ, όπως και συνέβει στον συγκεκριμένο ασθενή.

5.3 Γενικά συμπεράσματα και προτάσεις

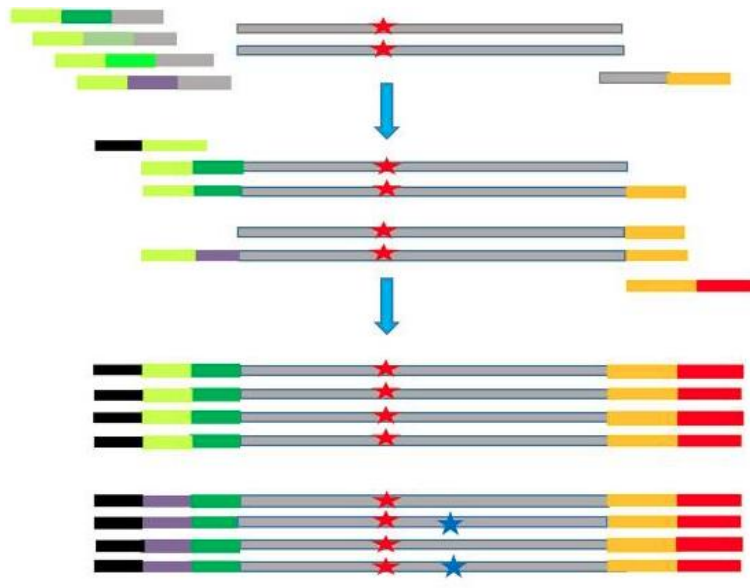
Στην συγκεκριμένη εργασία καταφέραμε με την πειραματική και αναλυτική μέθοδο που χρησιμοποιήσαμε να αναδείξουμε με υψηλή ευαισθησία μεταλλάξεις με $AF \geq 10\%$, ενώ επαληθεύσαμε στο σύνολο τους τις μεταλλάξεις που είχαμε αναδείξει με τη χρήση συμβατικών τεχνικών του εργαστηρίου μας.

Στην προσπάθεια μας να μελετήσουμε με μεγαλύτερη ευαισθησία το μοριακό τοπίο των επίκτητων μεταλλάξεων στις μυελικές νεοπλασίες, ανέκυψαν προβληματισμοί σχετικά με το πλήθος των ψευδώς θετικών δεδομένων, οι οποίοι θα αναλυθούν εκτενώς στη συνέχεια.

5.3.1 Μεθοδολογία κατασκευής των βιβλιοθηκών

Αρχικό κριτήριο αποτελεί ο τρόπος κατασκευής των βιβλιοθηκών για αλληλούχιση στοχευμένων περιοχών σε μεγάλα βάθη. Οι δύο βασικές μεθοδολογίες που υπάρχουν για την στοχευμένη αλληλούχιση βασίζονται στην απομόνωση των επιθυμητών τμημάτων είτε μέσω PCR ενίσχυσης αυτών, είτε με τον κατακερματισμό του γενομικού DNA και την επιλογή των επιθυμητών τμημάτων με τη χρήση ιχνηθετών, μεθόδους γνωστές ως Amplicon και Hybridization Capture αντίστοιχα. Για την αλληλούχιση μικρών σετ γονιδίων σε μεγαλύτερα βάθη από την βιβλιογραφία προτείνεται η χρήση amplicon based μεθόδων. Όπως όμως αναλύσαμε και στο κεφάλαιο 5.1, τέτοιες μέθοδοι ανάμεσά τους αυτή που χρησιμοποιήθηκε στην παρούσα εργασία ενέχουν τον κίνδυνο προσθήκης λανθασμένων γονιδίων κατά την PCR ενίσχυση. Τα λάθη αυτά, ενδέχεται να αναδειχθούν ως πραγματικές μεταλλάξεις, εάν δεν είναι συστηματικά, ώστε να εντοπιστούν και απομακρυνθούν με κατάλληλη επεξεργασία, όπως μέσω τροποποιημένης χρήσης του Panel Of Normals (Samorodnitsky E. 2015).

Μια μικρή βελτίωση τέτοιων amplicon based μεθόδων αποτελεί η χρήση μοναδικών ολιγονουκλεοτιδικών αλληλουχιών γνωστών ως Unique Molecular Identifiers ή εν συντομία UMIs. Με τον τρόπο αυτό είναι δυνατόν, τα σφάλματα κατά την PCR να ανιχνευθούν και να εξαιρεθούν από την ανάλυση (Kou R et al. 2016)



Εικόνα 42. Απεικόνιση ανίχνευσης σφαλμάτων κατά την ενίσχυση των τμημάτων προς αλληλούχιση. Σε περίπτωση που μια αληθινή μετάλλαξη υπάρχει στην περιοχή τότε θα βρίσκεται σε όλα τα διαβάσματα με την ίδια μοναδική αλληλουχία (συμβολίζεται στην εικόνα με κόκκινο αστέρι), σε διαφορετική περίπτωση αν πρόκειται για σφάλμα κατά την διάρκεια της PCR τότε μια τέτοια αλλαγή θα εμφανίζεται σε ένα μόνο μέρος των διαβασμάτων με ίδια μοναδική αλληλουχία (με μπλε αστέρι) (Τροποποίηση από Kou R et al. 2016)

5.3.2 Επιλογή κατάλληλου μηχανήματος αλληλούχισης και αναλώσιμων

Στην παρούσα εργασία η αλληλούχιση πραγματοποιήθηκε σε συσκευή MiSeq (Illumina), ακολουθώντας το πρωτόκολλο, όπως αυτό περιγράφεται στο ολοκληρωμένο σύστημα MiSeq Reagent Kit v2 (500 cycles), σύμφωνα με τις οδηγίες του κατασκευαστή (Illumina).

Μερικές επιλογές αν και σίγουρα **όχι οικονομικά, ούτε χρονικά ρεαλιστικές** είναι:

- η χρήση κατάλληλων αναλωσίμων για μεγαλύτερα βάθη διαβάσματος και βελτίωση της ποιότητας αυτών (π.χ με την χρήση του MiSeq Reagent Kit v3 (600 cycles)) που μπορούν να αποδώσουν σχεδόν τον διπλάσιο αριθμό διαβασμάτων συγκριτικά με την τεχνολογία που χρησιμοποιήθηκε
- Η μείωση του αριθμού των δειγμάτων ανά πειραματική διαδικασία ώστε να αυξηθεί βάθος αλληλούχισης

5.3.3 Βιοπληροφορική ανάλυση των δεδομένων

Η βιοπληροφορική ανάλυση των δεδομένων αποτελεί ένα από τα κυριότερα βήματα της μελέτης. Η ανάλυση στην παρούσα εργασία έγινε ακολουθώντας τις βέλτιστες πρακτικές και εργαλεία σύμφωνα με τις οδηγίες του Eli and Edythe L. Broad Institute of MIT and Harvard γνωστού και ως απλά Broad Institute. Τα βιοπληροφορικά αυτά εργαλεία καθώς και η μεθοδολογία που προτείνεται από το Broad Institute θεωρούνται σήμερα ως μια από τις καλύτερες δωρεάν διαθέσιμες επιλογές ανάλυσης μεγάλου όγκου δεδομένων από τεχνολογίες αλληλούχισης δεύτερης γενιάς στην επιστημονική κοινότητα. Με το πρόγραμμα Mutect2 να έχει αρκετά υψηλή ευαισθησία και ειδικότητα στην ανάδειξη επίκτητων μεταλλάξεων συγκριτικά με διαφορετικά προγράμματα παρόμοιας λειτουργίας (Krøigård AB et al. 2016).

5.3.3.1 Διαχωρισμός somatic και germline μεταλλάξεων

Ένα κρίσιμο σημείο κατά την διαδικασία ανάδειξης των μεταλλάξεων σε ένα δείγμα αποτελεί η διάκριση των επίκτητων μεταλλάξεων (somatic mutations) από τις κληρονομούμενες μεταλλάξεις και τους πολυμορφισμούς που υπάρχουν (germline mutations και SNPs αντίστοιχα). Η διάκριση αυτή είναι σημαντική για αρκετούς λόγους (Βλ. Κεφάλαιο 1.3.3.) όπως:

- Μόνο οι επίκτητες μεταλλάξεις μπορούν να αποδείξουν την επέκταση κλώνων κυττάρων κατά την αιμοποίηση στις μυελικές νεοπλασίες.
- Συχνά μεταλλάξεις χρησιμοποιούνται ως ποσοτικοί δείκτες για την παρακολούθηση της ελάχιστης υπολιμματικής νόσου (minimal residual disease, MRD) κατά την θεραπεία ασθενών. Κληρονομούμενες μοριακές εξαλλαγές δεν είναι δυνατόν να παρέχουν τέτοια πληροφορία (παρά μόνο σε περίπτωση αλλογενούς μεταμόσχευσης μυελού των οστών).
- Η μελέτη κληρονομούμενων μοριακών εξαλλαγών μπορεί να παρέχει χρήσιμες πληροφορίες για την προδιάθεση που αυτές μπορεί να προσδίδουν για την εμφάνιση νεοπλασιών.

Ένας από τους πιο αποδοτικούς τρόπους διάκρισης μεταξύ αυτών, αποτελεί η ταυτόχρονη αλληλούχιση μαζί με το προς μελέτη δείγματος και ενός αντίστοιχου πληθυσμού φυσιολογικών κυττάρων του ασθενούς (T/N pair, Tumor/Normal pair).

Με αυτό τον τρόπο μοριακές εξαλλαγές που εμφανίζονται στο φυσιολογικό δείγμα να αφαιρεθούν από την ανάλυση. Οδηγώντας έτσι τελικά στην ανίχνευση μόνο των επίκτητων μεταλλάξεων που συμβάλλουν στην νόσο.

Ωστόσο, στους ασθενείς με αιματολογικές κακοήθειες, τα καρκινικά κύτταρα διηθούν επιπρόσθετους ιστούς, όπως το δέρμα ή το επιθήλιο της στοματικής κοιλότητας, οπότε η απομόνωση φυσιολογικού δείγματος, για αντιπαραβολή με το παθολογικό, είναι μία ιδιαίτερα δυσχερής διαδικασία και εξ'ορισμού διπλασιάζει το κόστος της ανάλυσης. Αναπόφευκτα λοιπόν υπάρχει μια σημαντική διακριτική αδυναμία της ανάλυσης μεταξύ αυτών των τύπων μεταλλάξεων. Εναλλακτικοί τρόποι με τους οποίους θα μπορούσε να γίνει διαχωρισμός τους είναι:

- Με την χρήση στατιστικών μέτρων που δίνουν την πιθανότητα μιας μετάλλαξης να είναι germline που παρέχει η ανάλυση του Mutect, σε συνδυασμό με το AF που εντοπίζεται η μετάλλαξη. Λαμβάνοντας υπόψιν πως μεταλλάξεις με AF κοντά στο 50% ή 100% θα μπορούσαν να είναι germline μεταλλάξεις σε ετεροζυγωτία ή ομοζυγωτία, αντίστοιχα, χωρίς φυσικά αυτό να αποτελεί κανόνα, καθώς επίκτητες μεταλλάξεις σε μεγάλο φορτίο επίσης ανιχνεύονται σε ετεροζυγωτία ενώ έχουν αναφερθεί περιπτώσεις ακόμη και κύριων μεταλλάξεων, όπως η *JAK2V617F*, να εντοπίζονται ακόμα και σε ομοζυγωτία (Godfrey AL et al. 2012). Ενώ μη ομοιογενής ενίσχυση τμημάτων κατά την κατασκευή της βιβλιοθήκης μπορεί να οδηγήσει σε παρεκκλίσεις του AF από τα αναμενόμενα ποσοστά του 50% και 100% των germline μεταλλάξεων (Constance Baer et al. 2019).

- Με τον όσο το δυνατόν καλύτερο σχολιασμό των αναδεδειγμένων μεταλλάξεων με πλήθος πληροφοριών από βάσεις δεδομένων όπως πληθυσμιακών, κλινικών και φαρμακευτικών δεδομένων που θα οδηγήσουν σε καλύτερη αξιολόγηση.

5.3.3.2 Επιλογή κριτηρίων φιλτραρίσματος

Το φιλτράρισμα των μοριακών εξαλλαγών είναι το τελευταίο στάδιο της ανίχνευσης μεταλλάξεων και αυτό που τελικά θα οδηγήσει στα τελικά αποτελέσματα.

Τα δύο βασικά κριτήρια για την ανάδειξη πραγματικών μεταλλάξεων είναι το βάθος (DP) και το ποσοστό στο οποίο ανιχνεύεται η εκάστοτε μοριακή εξαλλαγή (AF). Για την επιλογή του καλύτερου ορίου αυτών των τιμών θα πρέπει να πραγματοποιηθούν ιδανικά έλεγχοι με γνωστά βιολογικά και *in silico* δείγματα γνωστών μεταλλάξεων και του ποσοστού αυτών, βάσει των οποίων θα μπορεί να γίνει η επιλογή των βέλτιστων ορίων για μέγιστη ευαισθησία και ειδικότητα της αναλυτικής διαδικασίας.

Διαφορετικά μπορεί να πραγματοποιηθεί αναδρομική επιβεβαίωση των αποτελεσμάτων με εγκαθιδρυμένες μεθόδους όπως PCR ή/και Sanger sequencing. Κάτι τέτοιο βέβαια πολλές φορές είναι οικονομικά και πρακτικά πολύ δύσκολο ή ακόμη και ακατόρθωτο όταν η έκταση των γενομικών περιοχών που αλληλουχούνται είναι μεγάλη και κατά συνέπεια και ο αριθμός των μοριακών εξαλλαγών που εντοπίζονται. Με αποτέλεσμα να μην είναι δυνατός ο υπολογισμός των αληθώς θετικών και ψευδώς θετικών μεταλλάξεων και κατ' επέκταση ο ακριβής υπολογισμός της ευαισθησίας και ειδικότητας των αναλύσεων.

Στη παρούσα εργασία, όπως έχει ήδη αναφερθεί, ένα σημαντικό κομμάτι κύριων και τροποποιητικών μεταλλάξεων με κλινική σημασία, είχε ανιχνευθεί εκ των προτέρων, στην προσπάθεια μας να καθορίσουμε όσο το δυνατόν πιο αξιόπιστα επίπεδα.

5.3.3.3 Ενημέρωση των αρχείων μεταλλάξεων με βιολογικά και κλινικά δεδομένα

Μετά την ανάδειξη και την επιλογή των αξιόπιστων μεταλλάξεων, εξαιρετικής σημασίας βήμα αποτελεί η ενημέρωση των αρχείων μεταλλάξεων με βιολογικά και κλινικά στοιχεία, που υπάρχουν διαθέσιμα σε διεθνείς βάσεις δεδομένων, προκειμένου οι αναδεδειγμένες μεταλλάξεις να αξιολογηθούν για την βιολογική και κλινική σημασία τους.

Στην παρούσα εργασία η ενημέρωση των αρχείων μεταλλάξεων με την χρήση του VEP και των επεκτάσεών του έδωσε ένα σημαντικό όγκο στοιχείων για την επίπτωση των μεταλλάξεων τόσο σε επίπεδο DNA όσο και πρωτεΐνης. Έγινε ενημέρωση με πληθυσμιακά στοιχεία για την συχνότητα εμφάνισης αυτών στον γενικό πληθυσμό (με σκοπό την διάκριση σπάνιων μεταλλάξεων και πολυμορφισμών) ενώ όπου ήταν δυνατό με βιβλιογραφικά στοιχεία για την επίπτωση συγκεκριμένων μεταλλάξεων και της κλινικής τους σημασίας.

Επιπλέον, για όλες τις σημειακές μεταλλάξεις με επίπτωση στην πρωτεϊνική ακολουθία έγινε πρόβλεψη της σοβαρότητας αυτών των αλλαγών στην λειτουργία της πρωτεΐνης και κατ' επέκταση στην παθογένεια τους.

Επιπλέον ενημέρωση των αρχείων με φαρμακολογικά δεδομένα, θεραπευτικές επιλογές και κλινικές δοκιμές για συγκεκριμένες μεταλλάξεις είναι δυνατόν να προστεθούν, ενισχύοντας έτσι την κλινική σημασία των αναδεδειγμένων μεταλλάξεων.

5.3.3.4 Πρότυπα και κατευθυντήριες γραμμές για την ερμηνεία και απόδοση επίκτητων μοριακών εξαλλαγών στον καρκίνο

Το 2017 η Αμερικανική Ένωση Μοριακής Παθολογοανατομίας (Association for Molecular Pathology, AMP) μετά από τη συλλογή απόψεων από ειδικούς της μοριακής παθολογοανατομίας εξέδωσε τα «Πρότυπα και κατευθυντήριες γραμμές για την ερμηνεία και απόδοση επίκτητων μοριακών εξαλλαγών στον καρκίνο», γνωστά και ως AMP οδηγίες (AMP guidelines) (Li MM et al. 2017). Σύμφωνα με αυτές τις οδηγίες οι επίκτητες μεταλλάξεις μπορούν να ταξινομηθούν σε 4 κατηγορίες ανάλογα με την κλινική τους σημασία με σκοπό την τυποποιημένη και όσο το δυνατόν καλύτερη αξιολόγηση και αναφορά σωματικών μεταλλάξεων στην κλινική πράξη.



Εικόνα 43. Η ταξινόμηση των επίκτητων σωματικών μεταλλάξεων σύμφωνα με τα AMP κριτήρια (Τροποποίηση από Li MM et al. 2017).

Όπως είναι προφανές, για την ακριβή ταξινόμηση σύμφωνα με τα AMP κριτήρια που φαίνονται στην εικόνα 43, απαιτείται επαρκής σχολιασμός των αναδεδειγμένων μεταλλάξεων, με ενημερωμένες βάσεις δεδομένων, όπως αναφέρεται στο προηγούμενο κεφάλαιο.

5.4 Μελλοντικές προοπτικές

Η τεχνική του στοχευμένου προσδιορισμού της νουκλεοτιδικής ακολουθίας, με τεχνολογία δεύτερης γενιάς (NGS), παρέχει τη δυνατότητα ταυτόχρονης διερεύνησης περισσότερων γενομικών περιοχών, ολόκληρων εξωνίων ή γονιδίων, για περισσότερους από έναν ασθενείς σε μία μόνο αντίδραση με μια αρκετά καλή ευαισθησία που κυμαίνεται από 1% έως 10% ανάλογα με την έκταση της αλληλούχισης (Whole Genome Sequencing-WGS, Whole Exome Sequencing-WES και Targeted re-sequencing σε συγκεκριμένες γενομικές περιοχές-γονίδια),

το αντίστοιχο βάθος που επιτυγχάνεται, την μέθοδο προετοιμασίας της βιβλιοθήκης και φυσικά της βιοπληροφορικής ανάλυσης των δεδομένων.

Η βελτιστοποίηση της αναλυτικής διαδικασίας μέσω αντιδραστηρίων καλύτερης χημείας, ακριβέστερων αλγόριθμων βιοπληροφορικής ανάλυσης των δεδομένων NGS, πληρέστερων βάσεων δεδομένων, αλλά και πιο εκλεπτυσμένων στατιστικών μοντέλων, με βάση τα οποία χαρακτηρίζεται η παθογένεια των ανιχνευόμενων μοριακών βλαβών όπως αναφέρθηκε στα παραπάνω κεφάλαια είναι καίριας σημασίας για την αξιοπιστία των αποτελεσμάτων.

Όλα τα παραπάνω, σε συνδυασμό με τη σταδιακή μείωση του κόστους λειτουργίας των συσκευών NGS καθώς και της θέσπισης διεθνών κριτηρίων ταξινόμησης και αξιολόγησης των επίκτητων μεταλλάξεων, έχουν ωθήσει τα εργαστήρια, να στραφούν προς αυτή την τεχνολογία για τη διάγνωση πληθώρας νοσημάτων και την εισαγωγή τους στην κλινική πράξη. (Goodwin et al. 2016, Wartman and Mardis 2016).

Στόχος του εργαστηρίου είναι η συνεχής βελτίωση της αναλυτικής διαδικασίας του στοχευμένου προσδιορισμού νουκλεοτιδικής ακολουθίας μέσω NGS, λαμβάνοντας υπόψιν τα προαναφερθέντα δεδομένα, με απώτερο σκοπό την πλήρη ενσωμάτωση αυτής, στη διαγνωστική ρουτίνα του εργαστηρίου. Η διαδικασία αυτή θα μπορούσε να συμβάλλει στην εύκολη, γρήγορη και αξιόπιστη ανίχνευση μοριακών εξαλλαγών, οι οποίες συμβάλλουν στη διάγνωση και ορθή ταξινόμηση, των αιματολογικών διαταραχών, στην πρόγνωση και διαστρωμάτωση του κινδύνου και ιδανικά στην επιλογή στοχευμένης θεραπείας, για τους ασθενείς με μυελικές νεοπλασίες.

6 ΒΙΒΛΙΟΓΡΑΦΙΑ

- AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* 2017;7(8):818–831. doi:10.1158/2159-8290.CD-17-0151
- Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, Pandey S, Patel JP, Chung YR, Koche R et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell.* 2012;22(2):180-93.
- Abdel-Wahab O, Levine R. The spliceosome as an indicted conspirator in myeloid malignancies. *Cancer cell.* 2011;20(4):420-422.
- Akutagawa J, Huang TQ, Epstein I, Chang T, Quirindongo-Crespo M, Cottonham CL, Dail M, Slusher BS, Friedman LS, Sampath D, et al. Targeting the PI3K/Akt pathway in murine MDS/MPN driven by hyperactive Ras. *Leukemia.* 2016;30(6):1335-43.
- Al-Kali A, Quintás-Cardama A, Luthra R, Bueso-Ramos C, Pierce S, Kadia T, Borthakur G, Estrov Z, Jabbour E, Faderl S et al. Prognostic impact of RAS mutations in patients with myelodysplastic syndrome. *Am J Hematol.* 2013;88(5):365-9.
- Araki M, Yang Y, Masubuchi N, Hironaka Y, Takei H, Morishita S, Mizukami Y, Kan S, Shirane S, Edahiro Y, et al. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. *Blood.* 2016;10;127(10):1307-16
- Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood.* 2016;127(20):2391-2405.
- Babaei MA, Kamalidehghan B, Saleem M, Huri HZ, Ahmadipour F. Receptor tyrosine kinase (c-Kit) inhibitors: a potential therapeutic target in cancer cells. *Drug Des Devel Ther.* 2016;10:2443-59.
- Barbacid M. Ras genes. *Annu Rev Biochem.* 1987;56:779-827.
- Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet.* 2005;365(9464):1054-1061.
- Besmer P, Murphy JE, George PC, Qiu FH, Bergold PJ, Lederman L, Snyder HW Jr, Brodeur D, Zuckerman EE, Hardy WD. A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family. *Nature.* 1986;320(6061):415-21.
- Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, Ito S, Cooper S, Kondo K, Koseki Y et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell.* 2014;157(6):1445-59.
- Blau, O. (2015). Gene Mutations in Acute Myeloid Leukemia — Incidence, Prognostic Influence, and Association with Other Molecular Markers in Leukemias - Updates and New Insights. InTech. <https://doi.org/10.5772/60928>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

- Boultonwood J, et al. Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia*. 2010;24(5):1062–1065.
- Brecqueville M, Rey J, Devillier R, Guille A, Gillet R, Adélaïde J, Gelsi-Boyer V, Arnoulet C, Chaffanet M, Mozziconacci MJ et al. Array comparative genomic hybridization and sequencing of 23 genes in 80 patients with myelofibrosis at chronic or acute phase. *Haematologica* 2014;99(1):37-45.
- Broséus J, Alpermann T, Wulfert M, Florensa Brichs L, Jeromin S, Lippert E, Rozman M, Lifermann F, Grossmann V, Haferlach T et al. Age, JAK2(V617F) and SF3B1 mutations are the main predicting factors for survival in refractory anaemia with ring sideroblasts and marked thrombocytosis. *Leukemia*. 2013;27(9):1826-31.
- Cai X, Hayashi Y, Wunderlich M, Speck NA, Mulloy JC, Huang G, Zheng Y. Loss of Function RUNX1 Mutations Restrict Protein Biosynthesis during Pre-Leukemia and MDS Transition but Not after Leukemic Transformation. *Blood* 2016 128:3860.
- Cazzola M, Kralovics R. From Janus kinase 2 to calreticulin: the clinically relevant genomic landscape of myeloproliferative neoplasms. *Blood*. 2014;123(24):3714-3719.
- Cazzola M, Rossi M, on behalf of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative L. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood*. 2013;121(2):260-269.
- Cazzola M. Mutant calreticulin: when a chaperone becomes intrusive. *Blood*. 2016;127(10):1219-1220.
- Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, Bock C, Vasanthakumar A, Gu H, Xi Y, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*. 2011;44(1):23-31.
- Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol*. 2003;23(16): 5594-5605.
- Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219. doi:10.1038/nbt.2514
- Clara JA, Sallman DA, Padron E. Clinical management of myelodysplastic syndrome/myeloproliferative neoplasm overlap syndromes. *Cancer Biol Med*. 2016;13(3):360-372.
- Cross NC. Genetic and epigenetic complexity in myeloproliferative neoplasms. *Hematology Am Soc Hematol Educ Program*. 2011;2011:208-214.
- Dameshek W. Some speculations on the myeloproliferative syndromes. *Blood*. 1951;6(4):372-375.
- Dao KT, Tyner JW, Gotlib J. Recent Progress in Chronic Neutrophilic Leukemia and Atypical Chronic Myeloid Leukemia. *Curr Hematol Malig Rep*. 2017;12(5):432-441.
- d'Auriol L, Mattei MG, Andre C, Galibert F. Localization of the human c-kit protooncogene on the q11-q12 region of chromosome 4. *Hum Genet*. 1988;78(4):374-6.
- DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M, (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *NATURE GENETICS* 43:491-498

- DiNardo CD, Cortes JE. Mutations in AML: prognostic and therapeutic implications. *Hematology Am Soc Hematol Educ Program*. 2016;2016(1):348–355. doi:10.1182/asheducation-2016.1.348
- Elliott MA, Tefferi A. Chronic neutrophilic leukemia 2016: Update on diagnosis, molecular genetics, prognosis, and management. *Am J Hematol*. 2016;91(3):341-9.
- Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, Waghorn K, Zoi K, Ross FM, Reiter A, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8):722-726.
- Figueira MI, Cardoso HJ, Correia S, Maia CJ, Socorro S. Hormonal regulation of c-KIT receptor and its ligand: implications for human infertility? *Prog Histochem Cytochem*. 2014;49(1-3):1-19.
- Gaidzik VI, Paschka P, Späth D, Habdank M, Köhne CH, Germing U, von Lilienfeld-Toal M, Held G, Horst HA, Haase D, et al. TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J Clin Oncol*. 2012;30(12):1350-1357.
- Geisler SJ, Paro R. Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development*. 2015;142(17):2876-2887.
- Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccia N, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol*. 2009;145(6):788-800.
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-51.
- Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature genetics*. 2012;44(1):53-7.
- Greenblatt SM, Nimer SD. Chromatin modifiers and the promise of epigenetic therapy in acute leukemia. *Leukemia*. 2014;28(7):1396-406.
- Harada H, Harada Y. Recent advances in myelodysplastic syndromes: Molecular pathogenesis and its implications for targeted therapies. *Cancer Sci*. 2015;106(4):329-336.
- He X, Chen Z, Jiang Y, Qiu X, Zhao X. Different mutations of the human c-mpl gene indicate distinct haematopoietic diseases. *J Hematol Oncol*. 2013;6:11.
- Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1-8.
- Heuser M, Panagiota V, Koenecke C, Fehse B, Alchalby H, Badbaran A, Shahswar R, Stadler M, Eder M, Göhring G, et al. Low frequency of calreticulin mutations in MDS patients. *Leukemia*. 2014;28(9):1933-1934.
- Hong JY, Seo JY, Kim SH, Jung HA, Park S, Kim K, Jung CW, Kim JS, Park JS, Kim HJ et al. Mutations in the Spliceosomal Machinery Genes SRSF2, U2AF1, and ZRSR2 and Response to Decitabine in Myelodysplastic Syndrome. *Anticancer Res*. 2015;35(5):3081-9.
- Im AP, Sehgal AR, Carroll MP, Smith BD, Tefferi A, Johnson DE, Boyiadzis M. DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies. *Leukemia* 2014; 28(9):1774-1783.

- Inoue D, Kitaura J, Togami K, Nishimura K, Enomoto Y, Uchida T, et al. Myelodysplastic syndromes are induced by histone methylation-altering ASXL1 mutations. *The Journal of clinical investigation*. 2013;123(11):4627-40.
- Jacquelin S, Straube J, Cooper L, et al. Jak2V617F and Dnmt3a loss cooperate to induce myelofibrosis through activated enhancer-driven inflammation. *Blood*. 2018;132(26):2707-2721
- Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33 Suppl:245-254.
- Jawhar M, Schwaab J, Schnittger S, Meggendorfer M, Pfirrmann M, Sotlar K, Horny HP, Metzgeroth G, Kluger S, Naumann N. Additional mutations in SRSF2, ASXL1 and/or RUNX1 identify a high-risk group of patients with KIT D816V(+) advanced systemic mastocytosis. *Leukemia*. 2016;30(1):136-43.
- Jeromin S, Haferlach T, Grossmann V, Alpermann T, Kowarsch A, Haferlach C, Kern W and Schnittger S. High frequencies of SF3B1 and JAK2 mutations in refractory anemia with ring sideroblasts associated with marked thrombocytosis strengthen the assignment to the category of myelodysplastic/myeloproliferative neoplasms. *Haematologica*. 2013;98(2): e15-e17.
- Jones AV, Kreil S, Zoi K, Waghorn K, Curtis C, Zhang L, Score J, Seear R, Chase AJ, Grand FH, et al. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood*. 2005;106(6):2162-2168.
- Katoh M. Functional and cancer genomics of ASXL family members. *British journal of cancer*. 2013;109(2):299-306.
- Kaushansky K, Josef T, Marcel M. L, Oliver W. P, Linda J. B, Caligiuri M. *Williams Hematology*, 9th ed. 9th ed: McGraw-Hill; 2016.
- Kaushansky K, Lichtman M, Prchal J, Levi MM, Press O, Burns L, Caligiuri M (eds.). *Williams Hematology*. New York (US): McGraw-Hill Education. 2016;Ed:9
- Kaushik G, Ivkovic S, Simonovic J, Tijanac N, Davis-Dusenbery B, Kural D. RABIX: AN OPEN-SOURCE WORKFLOW EXECUTOR SUPPORTING RECOMPUTABILITY AND INTEROPERABILITY OF WORKFLOW DESCRIPTIONS. *Pac Symp Biocomput*. 2016;22:154–165. doi:10.1142/9789813207813_0016
- Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NC, Berg T, Gisslinger B, Pietra D, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med*. 2013;369(25):2379-2390.
- Konrad J Karczewski, Laurent C Francioli, Grace Tiao, Beryl B Cummings, JessicaAlföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel PBirnbaum, Laura D Gauthier, Harrison Brand, Matthew Solomonson, Nicholas AWatts, Daniel Rhodes, Moriel Singer-Berk, Eleanor G Seaby, Jack A Kosmicki, Raymond K Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X Chong, Kaitlin E Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S Ware, Christopher Vittal, Irina MArmean, Louis Bergelson, Kristian Cibulskis, Kristen M Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, ThibaultJeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Mushi, Sam Novod, NikellePetrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, JoseSoto, Kathleen Tibbetts, Charlotte Tolone

n, Gordon Wade, Michael E Talkowski, The Genome Aggregation Database Consortium, Benjamin M Neale, Mark J Daly, Daniel GMacArthur. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 531210; doi: <https://doi.org/10.1101/531210>

- Kou R, Lam H, Duan H, et al. Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations. *PLoS One*. 2016;11(1):e0146638. Published 2016 Jan 11. doi:10.1371/journal.pone.0146638
- Koya J, Kataoka K, Sato T, Bando M, Kato Y, Tsuruta-Kishino T, Kobayashi H, Narukawa K, Miyoshi H, Shirahige K et al. DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. *Nat Commun*. 2016;7:10924.
- Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med*. 2005;352(17):1779-1790.
- Krøigård AB, Thomassen M, Lænkholm AV, Kruse TA, Larsen MJ. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLoS One*. 2016;11(3):e0151664. Published 2016 Mar 22. doi:10.1371/journal.pone.0151664
- Kuo MC, Liang DC, Huang CF, Shih YS, Wu JH, Lin TL, Shih LY. RUNX1 mutations are frequent in chronic myelomonocytic leukemia and mutations at the C-terminal region might predict acute myeloid leukemia transformation. *Leukemia*. 2009;23(8):1426-31.
- Larsson CA, Cote G, Quintás-Cardama A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol Cancer Res*. 2013;11(8):815-827.
- Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing Quality Assessment Tools to Enable Data-driven Informatics for High Throughput Genomics. *Front Genet*. 2013; 4: 288.
- Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, Boggon TJ, Wlodarska I, Clark JJ, Moore S, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell*. 2005;7(4):387-397.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-9. [PMID: 19505943]
- Li MM, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017;19(1):4–23. doi:10.1016/j.jmoldx.2016.10.002
- Liu X, Jian X, and Boerwinkle E. 2011. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Human Mutation*. 32:894-899.
- Liu X, Wu C, Li C and Boerwinkle E. 2016. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Human Mutation*. 37:235-241

- Luthra R, Chen H, Roy-Chowdhuri S, Singh RR. Next-Generation Sequencing in Clinical Molecular Diagnostics of Cancer: Advantages and Challenges. *Cancers* 2015;7:2023–2036.
- Madan V, Kanojia D, Li J, Okamoto R, Sato-Otsubo A, Kohlmann A, Sanada M, Grossmann V, Sundaresan J, Shiraishi Y et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun.* 2015;6:6042.
- Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood.* 2012;119(14):3203-10.
- Makishima H, Yoshida K, Nguyen N, Przychodzen B, Sanada M, Okuno Y, Ng KP, Gudmundsson KO, Vishwakarma BA, Jerez A, et al. Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet.* 2013;45(8):942-6.
- Maxson J, Gotlib J, Pollyea DA, Fleischman AG, Agarwal A, Eide CA, Bottomly D, Wilmot B, McWeeney SK, Tognon CE. Oncogenic CSF3R Mutations in Chronic Neutrophilic Leukemia and Atypical CML. *Engl J Med.* 2013;368(19):1781–1790.
- Maxson JE, Tyner JW. Genomics of chronic neutrophilic leukemia. *Blood.* 2017;129(6):715-722.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biology* Jun 6;17(1):122. (2016). doi:10.1186/s13059-016-0974-4
- Medeiros BC, Fathi AT, DiNardo CD, Pollyea DA, Chan SM, Swords R. Isocitrate dehydrogenase mutations in myeloid malignancies. *Leukemia.* 2017;31(2):272-281.
- Meggendorfer M, Jeromin S, Haferlach C, Kern W, Haferlach T. The mutational landscape of 18 investigated genes clearly separates four subtypes of myelodysplastic/myeloproliferative neoplasms. *Haematologica.* 2018;103(5):e192-e195.
- Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology.* 2013;38(1):23-38.
- Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, Avezov E, Li J, Kollmann K, Kent DG, Aziz A, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med.* 2013;369(25):2391-2405.
- Oshima K, Khiabani H, da Silva-Almeida AC, Tzoneva G, Abate F, Ambesi-Impiombato A, Sanchez-Martin M, Carpenter Z, Penson A, Perez-Garcia A, et al. Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A.* 2016;113(40):11306-11311.
- Papaemmanuil E, Cazzola M, Boultonwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, et al. Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts. *N Engl J Med.* 2011;365(15):1384-1395.
- Park ST, Kim J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *Int Neurolog J.* 2016;20(Suppl 2):S76-83.
- Patel AA, Steitz JA. Splicing double: Insights from the second spliceosome. *Nat Rev Mol Cell Biol* 2003;4: 960–970.
- Peter Amstutz, Michael R. Crusoe, Nebojša Tijanić (editors), Brad Chapman, John Chilton, Michael Heuer, Andrey Kartashov, Dan Leehr, Hervé Ménager, Maya Nedeljkovich, Matt Scales, Stian Soiland-Reyes, Luka Stojanovic (2016): *Common*

Workflow Language, v1.0. Specification, Common Workflow Language working group. <https://w3id.org/cwl/v1.0/doi:10.6084/m9.figshare.3115156.v2>

- Piazza R, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A, Antolini L, Mologni L, Donadoni C, Papaemmanuil E, et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat Genet.* 2013;45(1):18-24.
- Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood.* 2013;122(6):999-1006.
- Radek C, Skoda. Accelerating myelofibrosis through loss of Dnmt3a. *Blood Dec* 2018, 132 (26) 2703-2704; DOI: 10.1182/blood-2018-11-883470
- Reinig E, Yang F, Traer E, Arora R, Brown S, Rattray R, Braziel R, Fan G, Press R, Dunlap J. Targeted Next-Generation Sequencing in Myelodysplastic Syndrome and Chronic Myelomonocytic Leukemia Aids Diagnosis in Challenging Cases and Identifies Frequent Spliceosome Mutations in Transformed Acute Myeloid Leukemia. *Am J Clin Pathol.* 2016;145(4):497-506.
- Rinke J, Müller JP, Blaess MF, Chase A, Meggendorfer M, Schäfer V, Winkelmann N, Haferlach C, Cross NCP, Hochhaus A et al. Molecular characterization of EZH2 mutant patients with myelodysplastic/myeloproliferative neoplasms. *Leukemia.* 2017;31(9):1936-1943.
- Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK, et al. The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits WT DNMT3A by Blocking its Ability to Form Active Tetramers. *Cancer Cell.* 2014;25(4):442-454.
- Saez B, Walter MJ, Graubert TA. Splicing factor gene mutations in hematologic malignancies. *Blood.* 2017;129(10):1260-1269.
- Samorodnitsky E, Jewell BM, Hagopian R, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum Mutat.* 2015;36(9):903–914. doi:10.1002/humu.22825
- Sansom OJ, Maddison K, Clarke AR. Mechanisms of disease: methylbinding domain proteins as potential therapeutic targets in cancer. *Nature Clinical Practice Oncology* 2007;4:305–15.
- Santini V, Melnick A, Maciejewski JP, Duprez E, Nervi C, Cocco L, et al. Epigenetics in focus: Pathogenesis of myelodysplastic syndromes and the role of hypomethylating agents. *Critical Reviews in Oncology/Hematology.* 2013;88(2):231-45.
- Scott AJ, Lieu CH and Messersmith WA. Therapeutic Approaches to RAS Mutation. *Cancer J.* 2016; 22(3): 165–174.
- Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR, Futreal PA, Erber WN, McMullin MF, Harrison CN, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med.* 2007;356(5):459-468.
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. DNA sequencing at 40: past, present and future. *Nature.* 2017;550(7676):345-353.
- Shou LH, Cao D, Dong XH, Fang Q, Wu Y, Zhang Y, Fei JP, Xu BL. Prognostic significance of SETBP1 mutations in myelodysplastic syndromes, chronic myelomonocytic leukemia, and chronic neutrophilic leukemia: A meta-analysis. *PLoS One.* 2017;12(2): e0171608.
- Smith CW, Valca'rcel J. Alternative pre-mRNAsplicing: The logic of combinatorial control. *Trends Biochem Sci.* 2000;25: 381–388.

- Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*. 2014;28(3):485-96.
- Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. *Blood*. 2017;129(15):2070-2082.
- Soucie E, Hanssens K, Mercher T, Georgin-Lavialle S, Damaj G, Livideanu C, Chandesris MO, Acin Y, Létard S, de Sepulveda P, Hermine O, Bernard OA, Dubreuil P. In aggressive forms of mastocytosis, TET2 loss cooperates with c-KITD816V to transform mast cells. *Blood*. 2012;120(24):4846-9.
- Sun YG, Gracias NG, Drobish JK, Vasko MR, Gereau RW, Chen ZF. The c-kit signaling pathway is involved in the development of persistent pain. *Pain*. 2009;144(1-2):178-86.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009 May 15;324(5929):930-5.
- Tefferi A, Levine RL, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, Finke CM, Mullally A, Li CY, Pardanani A et al. Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRα correlates. *Leukemia*. 2009;23(5):900-4.
- Tefferi A, Vardiman JW. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia*. 2008;22(1):14-22.
- Tsai SC, Shih LY, Liang ST, Huang YJ, Kuo MC, Huang CF, Shih YS, Lin TH, Chiu MC, Liang DC. Biological Activities of RUNX1 Mutants Predict Secondary Acute Leukemia Transformation from Chronic Myelomonocytic Leukemia and Myelodysplastic Syndromes. *Clin Cancer Res*. 2015;21(15):3541-51.
- Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellström-Lindberg E, Tefferi A, Bloomfield CD. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114(5):937-51.
- Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016 Jan;11(1):1-9. doi: 10.1038/nprot.2015.123. Epub 2015 Dec 3.
- Verma M. Cancer Epigenetics; Risk Assessment, Diagnosis, Treatment, and Prognosis (1st ed.). *Methods in Molecular Biology*. 2015;1238:XVI, 804.
- Virginie Chesnais, Olivier Kosmider, Frederik Damm, Raphael Itzykson, Olivier A. Bernard, Eric Solary, Michaela Fontenay, Institut Cochin, et al. Spliceosome mutations in myelodysplastic syndromes and chronic myelomonocytic leukemia. *Oncotarget*. 2012.
- Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-veizer J, O'Laughlin M, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 2011;25(7):1153-1158.
- Wang M, He N, Tian T, Liu L, Yu S, Ma D. Mutation Analysis of JAK2V617F, FLT3-ITD, NPM1, and DNMT3A in Chinese Patients with Myeloproliferative Neoplasms. *Biomed Res Int*. 2014; 2014: 485645.
- Ward AF, Braun BS, Shannon KM. Targeting oncogenic Ras signaling in hematologic malignancies. *Blood*. 2012;120(17):3397-406.
- Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Collier HA, Cross JR, Fantin VR, Hedvat CV, Perl AE, et al. The common feature of leukemia-associated IDH1 and IDH2

mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*. 2010;17(3):225-234.

- Wartman LD, Mardis ER. Genomics. In: Kaushansky K, Lichtman M, Prchal J, Levi MM, Press O, Burns L, Caligiuri M (eds.). *Williams Hematology*. New York (US): McGraw-Hill Education. 2016;Ed:9;pp:155-164
- Will CL, Lührmann R. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*. 2011;3(7):a003707.
- Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*. 1999;402(6763):832-5.
- Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, Ueno T, Soda M, Hamada T, Haruta H, et al. Array-based genomic resequencing of human leukemia. *Oncogene*. 2010;29(25):3723-3731.
- Yohe S and Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med*. 2017;141(11):1544-1557.
- Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-9.
- You JS, Jones PA. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? *Cancer Cell*. 2012;22(1):9-20.
- Youzhong Wan. SF3B1 mutations in chronic lymphocytic leukemia. *BLOOD*. 2013;121.
- Zoi K, Cross NC. Molecular pathogenesis of atypical CML, CMML and MDS/MPN-unclassifiable. *Int J Hematol*. 2015;101(3):229-242.
- Zuo P, Maniatis T. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & development*. 1996;10(11):1356-68.
- Γιαννόπουλος Α. 2019. Μελέτη των μοριακών διαταραχών του επιγενετικού μηχανισμού ρύθμισης των κυττάρων στην παθογένεια των μυελοϋπερπλαστικών νεοπλασιών. URI: https://www.openarchives.gr/aggregator-openarchives/edm/pergamos/000005-_uoa_dl_object_uoadl%3A2851210
- Λουκόπουλος Δ., Πολίτου Μ. Κεφ. 6, Μυελοδυσπλαστικά Σύνδρομα. Μαθήματα Αιματολογίας, 2015.
- Χατζημιχαηλ Ε, Δοντσίδου Χ, Παπαθανασίου Κ. Διαταραχές Μεταγραφικών Παραγόντων, Σηματοδοτικών και Μεταβολικών Οδών στα ΜΔΣ, Θεραπευτική παρέμβαση. *ΗΑΕΜΑ*, 2015; 6(4):21-25

ΠΑΡΑΡΤΗΜΑ Ι – Πηγαίος κώδικας του προγράμματος φιλτραρίσματος και δημιουργία του Docker Image

I.1 Πηγαίος κώδικας του προγράμματος φιλτραρίσματος

```
#!/usr/bin/perl -w
use File::Basename;

$numberOfArgs=scalar(@ARGV);
if($ARGV[0] eq "-h"){

    print(
    "
NOTE: You have to split multiallelic sites before using!!

usage: perl filter_vcf.pl [-h] [-I input_vcf] [-O output_vcf]

required arguments:

-I the input vcf in 4.2 format
-O the output vcf after filtering

optional arguments:
-h show this help message and exit
-F the filter on FILTER column to use to exclude, you can give
only one option
-A the AF from which to filter from the AF in FORMAT
-D the min Depth of the DP in INFO\n"
);
die("\nTry again with some arguments\n");
}
for($i=0;$i<$numberOfArgs;$i++){
    if($ARGV[$i] eq "-I"){
        $invcf=$ARGV[$i+1];
    }
    if($ARGV[$i] eq "-O"){
        $outvcf=$ARGV[$i+1];
    }
    if($ARGV[$i] eq "-A"){
        $minAF=$ARGV[$i+1];
    }
    if($ARGV[$i] eq "-D"){
        $minDP=$ARGV[$i+1];
    }
    if($ARGV[$i] eq "-F"){
        $excludeFilter=$ARGV[$i+1];
    }
}

open (INPUT,"<$invcf") or die ("could not find vcf file");
open (OUT,">$outvcf") or die ("could not create vcf file");
$failToFilters=0;
$failToAF=0;
```



```

        print $_;
        print OUT "$. Something is not matching\n";
    }
}
}
$basename = basename($outvcf);
print "\n $basename";
$dirname = dirname($outvcf);
print "\n $dirname";
if($dirname eq '.'){ $dirname = "";}
$statstxt="$dirname"."$basename"."_stats.txt";
print "\n $statstxt";
open (STATS,"> $statstxt") or die ("could not create stats file");
$remaining=$total_variants-$failToAF-$failToDP-$failToFilters;

print "\nTotal Variants Processed : $total_variants
Excluded $failToFilters due to filters
Excluded $failToDP due to depth
Excluded $failToAF due to AF
Remaining varints: $remaining
";

print STATS "Total Variants Processed : $total_variants
Excluded $failToFilters due to filters
Excluded $failToDP due to depth
Excluded $failToAF due to AF
Remaining varints: $remaining
";

close STATS;
close INPUT;
close OUT;

```

I.2 Δημιουργία του Docker Image

I.2.1 Το Dockerfile

Είναι ένα αρχείο με ονομασία Dockerfile μέσα στο οποίο δίνονται οι οδηγίες για την κατασκευή του image. Παρακάτω δίνονται οι οδηγίες που χρησιμοποιήθηκαν για την δημιουργία του myfilter image.

```

FROM ubuntu:18.04
RUN apt-get update && \
    apt-get upgrade -y && \
    apt-get install -y libimage-magick-perl

# install modules that are needed

WORKDIR /opt
COPY filter_vcf.pl /opt/filter_vcf.pl

CMD ["perl","filter_vcf.pl","-h"]

```


1.2.2 Δημιουργία του Docker Image

Το μόνο που χρειάζεται για την δημιουργία ενός image, είναι σε έναν φάκελο να υπάρχει το Dockerfile και το αρχείο με τον πηγαίο κώδικα.

Στη συνέχεια εκτελούμε την παρακάτω εντολή όπου λέμε στο Docker να κατασκευάσει το docker image με όνομα myfilter σύμφωνα με τις οδηγίες που δίνονται στο dockerfile που βρίσκεται στον ίδιο φάκελο (.).

```
docker build -t myfilter .
```