



ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ  
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ  
ΣΠΟΥΔΩΝ  
«ΚΛΙΝΙΚΗ ΒΙΟΧΗΜΕΙΑ - ΜΟΡΙΑΚΗ ΔΙΑΓΝΩΣΤΙΚΗ»

ΙΑΤΡΙΚΗ ΣΧΟΛΗ ΑΘΗΝΩΝ - ΕΡΓΑΣΤΗΡΙΟ ΙΑΤΡΙΚΗΣ ΓΕΝΕΤΙΚΗΣ  
ΧΩΡΕΜΕΙΟ ΕΡΕΥΝΗΤΙΚΟ ΕΡΓΑΣΤΗΡΙΟ  
ΝΟΣΟΚΟΜΕΙΟ ΠΑΙΔΩΝ «Η ΑΓΙΑ ΣΟΦΙΑ»

Ερευνητική Εργασία Διπλώματος Ειδίκευσης:

«Πιλοτική εφαρμογή της αλληλούχησης επόμενης γενιάς (NGS: Next Generation Sequencing) σε όλα τα εξόνια του γονιδιώματος (WES: whole exome sequencing) για την τυποποίηση και αξιολόγηση της μεθοδολογίας για διαγνωστική χρήση»

“Pilot application of next-generation sequencing (Next Generation Sequencing) in all exons of the genome (Whole Exome Sequencing) for standardization and evaluation methodology for diagnostic use”

Μαρινάκης Νικόλαος  
*Μοριακός Βιολόγος και Γενετιστής*

**Επιβλέπουσα Καθηγήτρια:** Ε. Λιανίδου, Καθηγήτρια Αναλυτικής Χημείας, Τμήμα Χημείας, ΕΚΠΑ

**Εργαστηριακή Καθοδήγηση:** Jan Traeger-Synodinos, Αναπλ. Καθηγήτρια Γενετικής, Εργαστήριο Ιατρικής Γενετικής, Χωρέμειο Ερευνητικό Εργαστήριο, Τμήμα Ιατρικής, ΕΚΠΑ

**Τριμελής Εξεταστική Επιτροπή:**

- Ε. Λιανίδου, Καθηγήτρια Αναλυτικής Χημείας, Τμήμα Χημείας, ΕΚΠΑ
- Α. Σκορίλας, Καθηγητής Κλινικής Βιοχημείας, Τμήμα Βιολογίας, ΕΚΠΑ
- Jan Traeger-Synodinos, Αναπλ. Καθηγήτρια Γενετικής, Εργαστήριο Ιατρικής Γενετικής, Χωρέμειο Ερευνητικό Εργαστήριο, Τμήμα Ιατρικής, ΕΚΠΑ

## Περιεχόμενα

Περίληψη.....	σελ. 3
<b>Κεφάλαιο 1: Εισαγωγή</b>	
1.1) Αλληλούχηση του DNA.....	σελ. 8
1.1.1) Χημική Μέθοδος ή Μέθοδος κατά Maxam-Gilbert .....	σελ. 8
1.1.2) Ενζυμική Μέθοδος ή Μέθοδος κατά Sanger.....	σελ. 10
1.2) Αλληλούχηση Επόμενης Γενιάς (Next Generation Sequencing).....	σελ. 12
1.2.1) Αλληλούχηση κατά Sanger και Αλληλούχησης Επόμενης Γενιάς.....	σελ. 13
1.2.2) Τεχνολογίες Αλληλούχησης Νέας Γενιάς .....	σελ. 14
1.3) Αλληλούχηση Επόμενης Γενιάς στην κλινική πράξη.....	σελ. 17
1.3.1) Προ-αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς.....	σελ. 17
1.3.2) Αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς.....	σελ. 19
1.3.3) Μετα-αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς.....	σελ. 22
1.4) Κύρια επίπεδα Αλληλούχησης Επόμενης Γενιάς.....	σελ. 26
1.4.1) Αλληλούχηση των εξονίων του γονιδιώματος (Whole Exome Sequencing-WES).....	σελ. 39
1.5) WES ως διαγνωστικό εργαλείο.....	σελ. 30
<b>Κεφάλαιο 2: Υλικά και Μέθοδοι</b>	
2.1) Εφαρμογή της Αλληλούχησης Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES).....	σελ. 34
2.1.1) Whole Exome Sequencing (WES) για δείγμα ασθενούς με Νευρομυϊκές Διαταραχές (NMDs).....	σελ. 34
2.1.2) Τεχνική Επισκόπηση του Whole Exome Sequencing (WES).....	σελ. 35
2.2) Απομόνωση γενωμικού DNA από λεμφοκύτταρα περιφερικού αίματος με χρήση του πρωτόκολλο MagAttract DNA Blood M48 Kit και ρομποτικού συστήματος BioRobot M48 της QIAGEN.....	σελ. 37

2.3) Ποσοτικός προσδιορισμός του DNA.....σελ.	38
2.4) Ποιοτικός έλεγχος του DNA με Bioanalyzer 2100 της Agilent.....σελ.	39
2.5) Προετοιμασία βιβλιοθήκης με χρήση του TruSeq Exome Library Prep Kit της Illumina.....σελ.	41
2.6) Αλληλούχηση με σύστημα NextSeq 500 της Illumina.....σελ.	53
2.7) Ανάλυση δεδομένων με χρήση Βιοπληροφορικών Συστημάτων.....σελ.	57
2.7.1) Πρωτοβάθμια ανάλυση (Primary analysis).....σελ.	58
2.7.2) Δευτεροβάθμια ανάλυση (Secondary analysis).....σελ.	60
2.7.3) Τριτοβάθμια ανάλυση (Tertiary analysis).....σελ.	68

### **Κεφάλαιο 3: Αποτελέσματα**

3.1) Αποτελέσματα της Αλληλούχησης Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES).....σελ.	72
3.2) Προετοιμασία βιβλιοθήκης με TruSeq Exome Library Prep Kit της Illumina.....σελ.	72
3.2.1) Ποσοτικός έλεγχος DNA.....σελ.	72
3.2.2) Ποιοτικός έλεγχος DNA.....σελ.	73
3.3) Βιοπληροφορική ανάλυση των δεδομένων Αλληλούχησης Επόμενης Γενιάς.....σελ.	74
3.3.1) Ποιοτικός έλεγχος της κλήσης βάσεων (Base calling).....σελ.	75
3.3.2) Ποιοτικός έλεγχος χαρτογράφησης/στοίχισης.....σελ.	78
3.3.3) Μελέτη παραλλαγών που προσδιορίστηκαν με τους τρεις Ροές Εργασίας.....σελ.	83
3.4) Συγκριτική μελέτη των παραλλαγών που προσδιορίστηκαν με τις τρεις Ροές Εργασίας.....σελ.	89

### **Κεφάλαιο 4: Συμπεράσματα - Συζήτηση**

4.1) Αξιολόγηση της ροής εργασίας της Αλληλούχησης της Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES) – Συμπεράσματα.....σελ.	98
4.2) Συζήτηση.....σελ.	100

### **Κεφάλαιο 5: Βιβλιογραφία**

5.1) Βιβλιογραφία.....σελ.	103
----------------------------	-----

## Περίληψη

Η ολοκλήρωση της ανάλυσης του ανθρώπινου γονιδιώματος, το 2001, έθεσε τη βάση για την αναζήτηση διαφορετικών τεχνικών για την αλληλούχηση, όχι πλέον σε επίπεδο γονιδίου ή συγκεκριμένων περιοχών αλλά σε επίπεδο γονιδιώματος. Από το 2005 μέχρι σήμερα, έχουν αναπτυχθεί διαφορετικές τεχνολογίες έτσι ώστε να μιλάμε πλέον για μαζική παράλληλη αλληλούχηση (massive parallel sequencing) πολλών περιοχών του γονιδιώματος ή και ακόμα ολόκληρου του γονιδιώματος. Αν και η τεχνολογία του κάθε συστήματος είναι διαφορετική, έχουν παρόμοια στάδια κατά τη διαδικασία τους, όπως την κοπή του DNA σε τυχαία τμήματα, την προσθήκη ειδικών αλληλουχιών-ανταπτόρων (προσαρμογέων) στα άκρα των τμημάτων του DNA, την μεταφορά των τμημάτων σε σφαιρίδια ή σε πλακίδιο, την ενίσχυση των τμημάτων με σκοπό την κατασκευή DNA βιβλιοθήκης, δηλαδή βιβλιοθήκης που περιλαμβάνει τα θραύσματα DNA που τελικώς θα αλληλουχηθούν.

Η ροή εργασίας για την εφαρμογή αλληλούχησης επόμενης γενιάς στην κλινική πράξη περιλαμβάνει τρεις κύριες φάσεις, την προ-αναλυτική φάση, την αναλυτική φάση και την μετά-αναλυτική φάση. Η προ-αναλυτική φάση περιλαμβάνει τη συλλογή πληροφοριών σχετικά με το οικογενειακό ιστορικό του ασθενούς και τη συστηματική αξιολόγηση του φαινοτύπου, το είδος της δοκιμασίας που θα εφαρμοστεί, τη λήψη δείγματος, την ενημέρωση του ασθενούς και τέλος τη συγκατάθεση του. Η αναλυτική φάση περιλαμβάνει έναν αριθμό εργαστηριακών βημάτων συμπεριλαμβανομένου την προετοιμασία του δείγματος, την κατασκευή «βιβλιοθήκης», τον εμπλουτισμό του στόχου και την αντίδραση αλληλούχησης. Επίσης, περιλαμβάνει τρία επίπεδα ανάλυσης δεδομένων, την πρωτοβάθμια (primary), τη δευτεροβάθμια (secondary) και την τριτοβάθμια ανάλυση και την κατηγοριοποίηση των παραλλαγών που εξάχθηκαν. Η μετά-αναλυτική φάση ξεκινάει με την αξιολόγηση και ερμηνεία των αποτελεσμάτων της αναλυτικής φάσης και τελειώνει με τη σύνταξη της αναφοράς και παράδοσής της στον ασθενή με γενετική συμβουλευτική.



Η αλληλούχηση επόμενης γενιάς μπορεί να χρησιμοποιηθεί σε διάφορα επίπεδα πολυπλοκότητας. Αυτά περιλαμβάνουν αλληλούχηση επιλεγμένων γονιδίων (πάνελ) που σχετίζονται με κάποια ασθένεια (Disease-targeted gene panel), αλληλούχηση των εξονίων όλων των γονιδίων του γονιδιώματος (Whole Exome Sequencing-WES) και αλληλούχηση όλου του γονιδιώματος (Whole Genome Sequencing-WGS). Συγκρίνοντας τα τρία επίπεδα, είναι ξεκάθαρο ότι το WGS είναι ανώτερο από τα άλλα δύο παράγοντας πιο πλήρη δεδομένα για το γονιδίωμα που εξετάζεται. Το WES είναι σημαντικά πιο οικονομικό από το WGS και αποτελεί μια αποτελεσματική στρατηγική για την ανίχνευση μεταλλαγών που ευθύνονται για ασθένειες. Η ανάλυση πάνελ γονιδίων, αντιπροσωπεύει μια εναλλακτική τακτική, η οποία έχει σημαντικά χαμηλότερο κόστος αλληλούχησης, αλλά θα επιτύχει μόνο εάν το γονίδιο που προκαλεί την ασθένεια περιλαμβάνεται στο πάνελ.

Οι τεχνολογίες αλληλούχησης επόμενης γενιάς έχουν αλλάξει δραματικά το πεδίο της μοριακής διαγνωστικής. Η αλληλούχηση των εξονίων του γονιδιώματος (WES) έχει εξελίξει τη βιοϊατρική έρευνα επεκτείνοντας τα αποτελέσματα και τις εφαρμογές στην κλινική διάγνωση. Μέσω του WES μπορεί κανείς να αποκαλύψει νέες παραλλαγές που δεν είναι γνωστό ότι σχετίζονται με κάποια νόσο ή προκαλούν εξαιρετικά σπάνιες ασθένειες και να ταυτοποιήσει παθογόνες παραλλαγές για ασθένειες οι οποίες δεν έχουν σχέση με το κύριο σκοπό της εξέτασης που ζητήθηκε (δευτερεύοντα ευρήματα). Επίσης, μπορεί να προσδιοριστεί η ύπαρξη φορέας για υπολειπόμενη ασθένεια, αλλά και φαρμακογενετικές παραλλαγές που μπορεί να επηρεάσουν την ανταπόκριση ενός ατόμου σε κάποιο φαρμακευτικό παράγοντα. Ωστόσο είναι σημαντικό τα υποβληθέντα άτομα σε τέτοια εξέταση να ενημερωθούν σχετικά με τα αναμενόμενα αποτελέσματα, την πιθανότητα και το είδος των τυχαίων ευρυμάτων, ποια αποτελέσματα μπορούν να ανακοινωθούν και ποια όχι, καθώς ζητήματα διαγνωστικής απόδοσης και βιοηθικής παραμένουν για την χρήση του WES στην κλινική πράξη.

Για τη χρήση Αλληλούχησης Νέας Γενιάς στην κλινική πράξη είναι αναγκαία η βελτιστοποίηση των εργαστηριακών πρωτοκόλλων, η κατάλληλη κατάρτιση του προσωπικού και η ανάπτυξη βιοπληροφορικών εργαλείων για τη

διαχείριση και ερμηνεία των πολύπλοκων δεδομένων που παράγονται. Επίσης, απαιτείται εντατική αλληλεπίδραση μεταξύ εργαστηριακών ειδικών, κλινικών γενετιστών, βιοπληροφορικών και άλλων ιατρικών ειδικοτήτων.

Στη συγκεκριμένη μελέτη εφαρμόστηκε πιλοτική δοκιμή της Αλληλούχησης Νέας Γενιάς στην κλινική πράξη προσδιορίζοντας τις απαιτήσεις, τις ανάγκες και τους περιορισμούς της τεχνολογίας αυτής μέσα από τη σύγκριση τριών ροών εργασίας ('Workflows'). Παρόλο λοιπόν που δεν υπάρχει αμφιβολία ότι οι τεχνολογίες προσδιορισμού της αλληλουχίας έχουν τη δυνατότητα να ωφελήσουν την υγεία, απαιτούνται σαφείς κατευθυντήριες οδηγίες για τις μεθοδολογίες NGS σε όλα τα στάδια.

## **Abstract**

The completion of human genome analysis in 2001 has set the stage for searching for different sequencing techniques, not limited to the specific regions or genes, but at the level of the entire genome. From 2005 until today, different technologies have been developed, leading to the procedure of massively parallel sequencing or next generation sequencing, applicable to many regions of the genome or even the entire genome. Although the technology of each commercially available system is different, overall all have similar stages in the process, including random DNA fragmentation, addition of adaptors at the ends of the DNA fragments, transfer of the fragments onto beads or microarray surfaces, fragment enrichment for the purpose of construction DNA library, the final fragments that will be sequenced.

The workflow for implementing massively parallel sequencing or next generation sequencing in clinical practice includes three main phases: the pre-analytical phase, the analytical phase and the post-analytical phase. The pre-analytical phase includes the collection of information about the phenotype, family history and the systematic evaluation of the patients, the type of test to be applied, acquiring of the sample of DNA, genetic counseling and signing the informed consent. The analytical phase includes a number of laboratory steps including DNA preparation, 'library' construction, target enrichment and sequencing reaction. It also includes three levels of data analysis, primary, secondary and tertiary analysis and categorization of exported variants. The post-analytic phase begins with the evaluation and interpretation of the variants from the analytical phase and ends with the writing of the report and delivery to the patient with genetic counseling.

The next generation sequencing can be used with various levels of complexity with respect to the extent of the genome to be analysed. These levels include Disease-targeted gene panels, Whole Exome Sequencing-WES and Whole Genome Sequencing-WGS. By comparing the three options, it is clear that WGS is superior to the other two, as it gives more comprehensive data on the entire genome under consideration. On the other hand, WES is significantly more cost-effective than WGS and it is an effective strategy for detecting

disease-causing mutations. The analysis of gene panels that have been specifically targeted, represents an alternative tactic that has significantly lower sequencing costs but will only succeed if the gene causing the disease is included in the panel.

The next-generation sequencing technologies have dramatically changed the field of molecular diagnostics. Whole Exome Sequencing (WES) has revolutionized biomedical research by extending the results of sequencing and applications to clinical diagnosis. Through WES one can reveal new variants that have not been previously associated with a disease or cause extremely rare diseases. Another diagnostic result from the use of WES is the identification of pathogenic variants for diseases that are unrelated to the main purpose of the test requested (secondary findings). Also, the existence of a carrier for residual disease, as well as pharmacogenomic variants that may affect an individual's response to a pharmaceutical agent can be determined. However, it is important for the patients undergoing such an examination to be informed of the expected results, the likelihood and type of secondary findings, which results can be announced and which are not. Many issues of diagnostic yield and bioethics remain to be addressed prior to the wider use of WES in clinical practice .

Before the use of Next Generation Sequencing in clinical practice it is necessary to optimize laboratory protocols, appropriate staff training and development of bioinformatic tools to manage and interpret the data that generated. Also, intensive interaction between laboratory specialists, clinical geneticists, bioinformatics and other medical specialists is required.

In this study, was carried out a pilot test to implement Next Generation Sequencing into clinical practice. It took place comparing of different laboratory and bioinformatics methodologies, so were identified the requirements and the needs. Although, there is no doubt that sequencing technologies have the potential to benefit health, but guidelines are needed for NGS methodologies at all stages.

# **Κεφάλαιο 1:**

## ***Εισαγωγή***

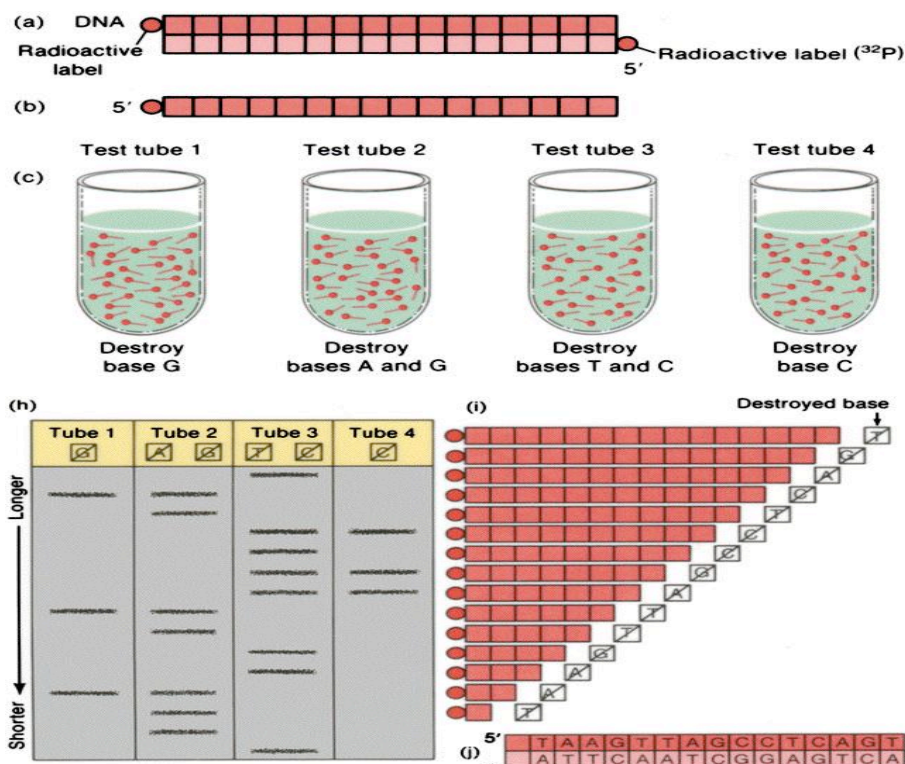
## **1.1) Αλληλούχηση του DNA**

Η ανάλυση της πρωτοταγούς δομής (ή αλληλούχηση) του DNA είναι ο ακριβής προσδιορισμός των βάσεων του DNA μιας περιοχής (γονίδιο, χρωμοσωμική περιοχή, γονιδίωμα). Παρόλο που μέχρι τα μέσα της δεκαετίας του 1970 ήταν δυνατή η απομόνωση γονιδίων και η παραγωγή τους σε πολλαπλά αντίγραφα προκειμένου να αναλυθούν, δεν υπήρχε τρόπος να διαβαστεί η νουκλεοτιδική αλληλουχία του DNA. Στην πραγματικότητα, οι πρώτες νουκλεοτιδικές αλληλουχίες είχαν προσδιοριστεί κατά τη δεκαετία του 1960, αλλά αφορούσαν τα σχετικά μικρά μόρια tRNA, το μήκος των οποίων ήταν μόνο 75-80 νουκλεοτίδια. Στα τέλη της δεκαετίας του 1960 ο Frederick Sanger έστρεψε την προσοχή του από την αλληλούχηση πρωτεϊνών στην ανάπτυξη γρήγορων και απλών μεθόδων για την αλληλούχηση μεγαλύτερων μορίων RNA. Δημιουργήθηκε έτσι ένας έμμεσος τρόπος προσδιορισμού της αλληλουχίας του DNA, χρησιμοποιώντας αρχικά RNA πολυμεράση για τη σύνθεση μιας συμπληρωματικής αλυσίδας RNA. Η επανάσταση στις διαδικασίες εύρεσης της αλληλουχίας νουκλεϊκών οξέων ήρθε λίγα χρόνια αργότερα, με την ανάπτυξη μεθόδων που επέτρεπαν την άμεση αλληλούχηση τμημάτων DNA μήκους μεταξύ 100 και 500 νουκλεοτιδίων. Η πρώτη από αυτές, η λεγόμενη μέθοδος συν-πλην, αναπτύχθηκε από τον Sanger το 1975 και σε συνδυασμό με τη χρήση ηλεκτροφόρησης σε πήκτωμα πολυακρυλαμίδης, επέτρεψε το γρήγορο προσδιορισμό μεγάλου μέρους της αλληλουχίας του γονιδιώματος (μεγέθους 5.386 bp) του μικρού φάγου φX174. Δύο χρόνια αργότερα, οι Allan Maxam και Walter Gilbert στο Harvard και ο Sanger με τους συνεργάτες του στο Cambridge της Αγγλίας συνέβαλαν σημαντικά στην ανάπτυξη νέων μεθόδων πολύ πιο εύχρηστων και αποτελεσματικών από τη μέθοδο συν-πλην.

### **1.1.1) Χημική Μέθοδος ή Μέθοδος κατά Maxam-Gilbert**

Η μέθοδος των Maxam-Gilbert βασίζεται στη χημική αποικοδόμηση των αλυσίδων DNA. Κατά τη μέθοδο αυτή, τα μόρια DNA σημαίνονται στο ένα τους άκρο και κατόπιν χωρίζονται σε τέσσερις επιμέρους αντιδράσεις. Σε κάθε αντίδραση πραγματοποιείται διαφορετική χημική επεξεργασία, με σκοπό την

κατάργηση των φωσφοδιεστερικών δεσμών μίας ή δύο βάσεων. Είναι μια διαδικασία 3 βημάτων, τροποποίηση της αζωτούχας βάσης, αφαίρεση της τροποποιημένης αζωτούχας βάσης από τη δεοξυριβόζη και κοπή του κλώνου στη συγκεκριμένη δεοξυριβόζη. Αν και η διαδικασία αποτελείται από τρία στάδια, η χημική τροποποίηση δεν είναι ίδια και για τα 4 νουκλεοτίδια, εξαρτάται από το είδος του νουκλεοτιδίου και από το αν η βάση ανήκει στις πουρίνες ή στις πυριμιδίνες. Συνολικά, πραγματοποιούνται 4 αντιδράσεις: α) στο νουκλεοτίδιο της γουανίνης (αντίδραση G) β) στα νουκλεοτίδια της γουανίνης και της αδενίνης –πουρίνες (αντίδραση G και A) γ) στα νουκλεοτίδια της κυτοσίνης και της θυμίνης –πυριμιδίνες (αντίδραση C και T) και δ) στο νουκλεοτίδιο της κυτοσίνης (αντίδραση C). Μετά την ολοκλήρωση της παραπάνω επεξεργασίας τα τμήματα DNA διαχωρίζονται μέσω ηλεκτροφόρησης σε πήκτωμα πολυακρυλαμίδης και ανιχνεύονται μέσω αυτοραδιογραφίας. Η σήμανση του ενός άκρου και η κοπή σε συγκεκριμένες θέσεις ανάλογα με την αντίδραση έχει ως αποτέλεσμα την δημιουργία ενός φάσματος τμημάτων DNA. Το μήκος του κάθε τμήματος καθορίζεται από την απόσταση που υπάρχει από το σημασμένο άκρο του DNA έως τη θέση κοπής (Εικόνα 1.1) (Watson, 2007).



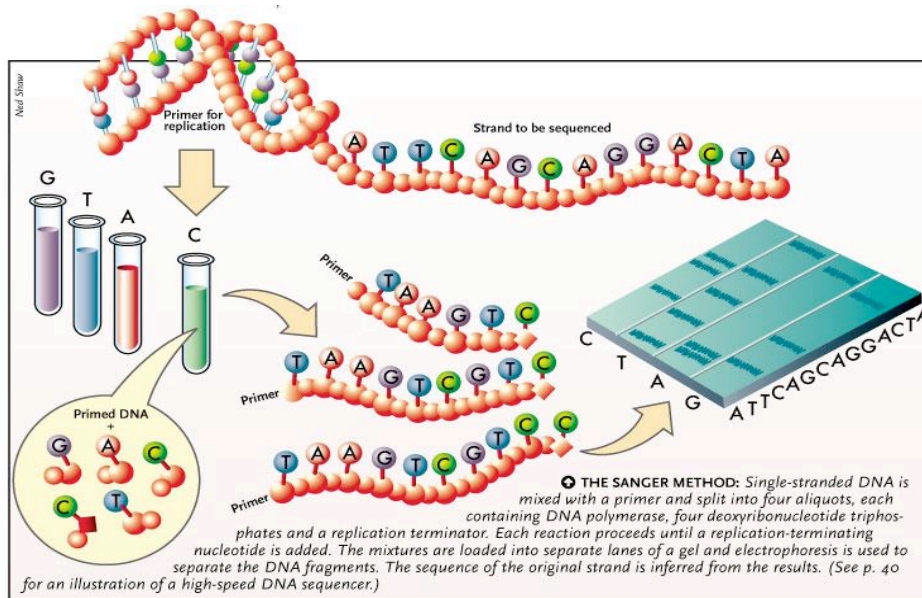
Πηγή: <http://blog.ub.ac.id/jatmikoeokotbp/files/2014/04/Picture1.png>

**Εικόνα 1.1:** Μέθοδος Maxam-Gilbert.

### 1.1.2) Ενζυμική Μέθοδος ή Μέθοδος κατά Sanger

Είναι η πιο διαδεδομένη τεχνική για την ανάλυση της πρωτοταγούς δομής του DNA. Έχει τη δυνατότητα προσδιορισμού της πρωτοταγούς δομής μιας αλληλουχίας η οποία μπορεί να φτάνει τις 1000 βάσεις. Η τεχνική βασίζεται στην ενζυμική ενσωμάτωση ενός νουκλεοτιδίου στην επιμηκυνόμενη αλυσίδα του DNA. Η αρχή της τεχνικής στηρίζεται στην ύπαρξη ενός συνθετικού σημασμένου εκκινητή (αρχικά ήταν ραδιενεργά σημασμένος), ο οποίος υβριδίζεται στο μονόκλωνο DNA. Επίσης για την πραγματοποίηση της αντίδρασης απαιτείται πολυμεράση, τα 4 φυσιολογικά τριφωσφορικά ολιγονουκλεοτίδια (dNTPs) καθώς και, σε μικρή αναλογία, 2',3' διδεόξυνουκλεοτίδια (dideoxynucleotides, ddNTPs). Το διδεόξυνουκλεοτίδιο έχει ένα άτομο υδρογόνου στον 3' άνθρακα της δεοξυριβόζης αντί για ένα υδροξύλιο. Με τον τρόπο αυτό, όταν γίνεται η ενσωμάτωση του διδεόξυνουκλεοτιδίου στο νεοσυντιθέμενο κλώνο του DNA δεν είναι δυνατή η δημιουργία άλλου φωσφωδιεστερικού δεσμού και επομένως η πρόσδεση του επόμενου νουκλεοτιδίου (λόγω της έλλειψης του υδροξυλίου) και η αντίδραση σταματάει. Για το κάθε νουκλεοτίδιο πραγματοποιείται μια αντίδραση (επομένως πραγματοποιούνται 4 αντιδράσεις, μια για κάθε νουκλεοτίδιο). Σε κάθε αντίδραση υπάρχει, σε μικρή αναλογία, ένα διδεοξυνουκλεοτίδιο (για την αντίδραση της γουανίνης- το διδεόξυνουκλεοτίδιο της γουανίνης-ddGTPs κτλ). Κατά τη σύνθεση των νέων αλυσίδων, δημιουργούνται διαφορετικά, σε μήκος, τμήματα DNA λόγω της τυχαίας προσθήκης στην νεοσυντιθέμενη αλυσίδα των διδεόξυνουκλεοτιδίων. Το μήκος του κάθε τμήματος καθορίζεται από την απόσταση που υπάρχει από το 5' άκρο του εκκινητή έως τη θέση προσθήκης του διδεόξυνουκλεοτιδίου. Αυτό έχει ως αποτέλεσμα να έχουμε συγκεκριμένα μήκη μορίων στα οποία γνωρίζουμε το νουκλεοτίδιο που υπάρχει κάθε φορά. Ο διαχωρισμός των μορίων των παραπάνω αντιδράσεων γίνεται με τη διαδικασία της ηλεκτροφόρησης (Εικόνα 1.2) (Watson, 2007).

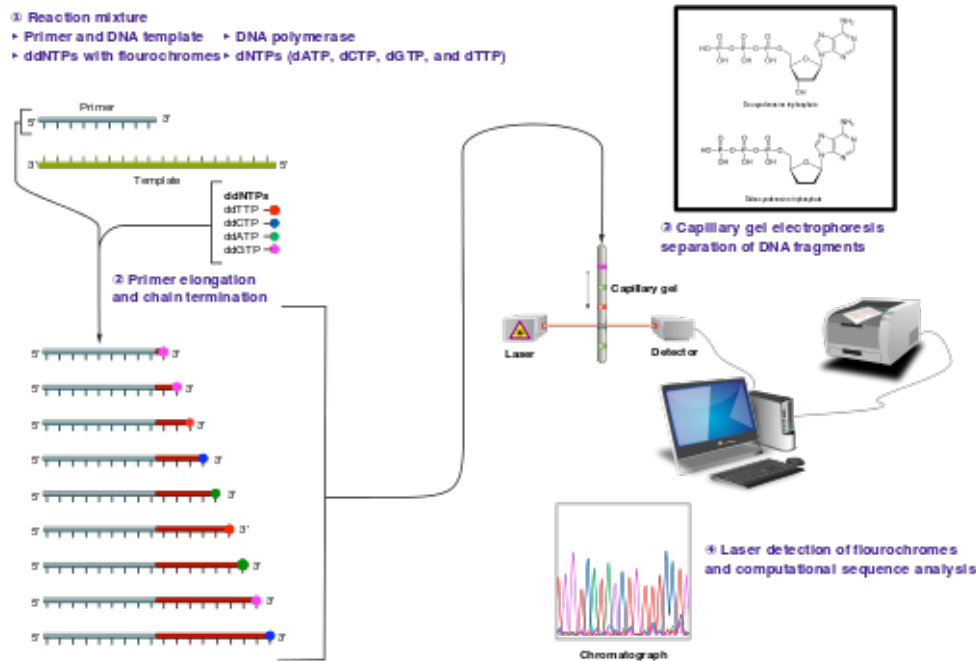




Πηγή: <http://images.the-scientist.com/content/figures/0890-3670-040927-44-1-1.jpg>

**Εικόνα 1.2:** Μέθοδος Sanger.

Από την πρώτη εφαρμογή της μεθόδου μέχρι σήμερα, έχουν πραγματοποιηθεί πολλές αλλαγές στην τεχνική οι οποίες είχαν ως σκοπό την αυτοματοποίηση της. Ο εκκινητής δε χρειάζεται πλέον να είναι σημασμένος με ραδιενεργό στοιχείο αλλά με ειδική χρωστική. Η χρωστική αποτελείται από μια χρωστική δότη (φλουορεσκεΐνη) συνδεδεμένη με μια χρωστική δέκτη (διχλωροροδαμίνη) μέσω ενός συνδέτη αμινοβενζοϊκού οξέος. Υπάρχουν 4 διαφορετικές χρωστικές δέκτη οι οποίες έχουν διαφορετικό φάσμα εκπομπής και η κάθε μια από αυτές ενσωματώνεται σε ένα συγκεκριμένο διδεοξυνουκλεοτίδιο. Η ανίχνευση του σήματος πραγματοποιείται από ειδικό laser και το σύστημα ηλεκτροφόρησης χρησιμοποιεί τριχοειδή σωληνάρια. Η δέση φωτός από το laser κατευθύνεται κατά μήκος των τριχοειδών. Ως αποτέλεσμα αυτού, τα σημασμένα ddNTPs απορροφούν μια μικρή ποσότητα από τη δέση φωτός και την αποδίδουν ως δέση φωτός με μεγαλύτερο μήκος κύματος προς όλες τις κατευθύνσεις. Με το κατάλληλο λογισμικό γίνεται μετατροπή των δεδομένων στις 4 χρωστικές που έχουν καθοριστεί για την ανάλυση της πρωτοταγούς δομής. Τα αποτελέσματα παρέχονται σε χρωματογράφημα το οποίο απαιτεί συγκεκριμένο λογισμικό για την ανάλυση του (Εικόνα 1.3) (Watson, 2007).



Πηγή: <https://upload.wikimedia.org/wikipedia/commons/thumb/b/b2/Sangersequencing.svg/673px-Sanger-sequencing.svg.png>

**Εικόνα 1.3:** Αυτοματοποιημένη μέθοδος αλληλούχησης Sanger.

## 1.2) Αλληλούχηση Επόμενης Γενιάς (Next Generation Sequencing)

Η ολοκλήρωση της ανάλυσης του ανθρώπινου γονιδιώματος το 2001, έθεσε την 'αφορμή' για την αναζήτηση διαφορετικών τεχνικών για την αλληλούχηση, όχι πλέον σε επίπεδο γονιδίου ή συγκεκριμένων περιοχών αλλά σε επίπεδο γονιδιώματος. Από το 2005 μέχρι σήμερα, έχουν αναπτυχθεί διαφορετικές τεχνολογίες αλληλούχησης για τον προσδιορισμό πολλών περιοχών του γονιδιώματος ή και ακόμα ολόκληρου του γονιδιώματος. Οι βελτιώσεις με την πάροδο των ετών στις παραδοσιακές τεχνικές επέτρεψε στους επιστήμονες να προσδιορίσουν αλληλουχία DNA έως περίπου 600 βάσεις σε μήκος. Ωστόσο, επειδή η αλληλούχηση τέτοιων κομματιών για τον προσδιορισμό ολόκληρων γονιδιωμάτων απαιτεί μεγάλο κόστος και χρόνο, υπάρχουν τεράστιοι περιορισμοί για την εφαρμογή τους. Η λύση αυτών των προβλημάτων επιτεύχθηκε με την Αλληλούχηση Επόμενης Γενιάς, η οποία επιτρέπει την ανάλυση εκατοντάδων κομματιών DNA στον ίδιο χρόνο και είναι γνωστή ως Μαζική Παράλληλη Αλληλούχηση. Οι επαναστατικές αυτές τεχνολογίες είναι πολλά υποσχόμενες για χρήση στην κλινική πράξη με στόχο τη βελτίωση της δημόσιας υγείας. Ο όρος Αλληλούχηση Επόμενης Γενιάς

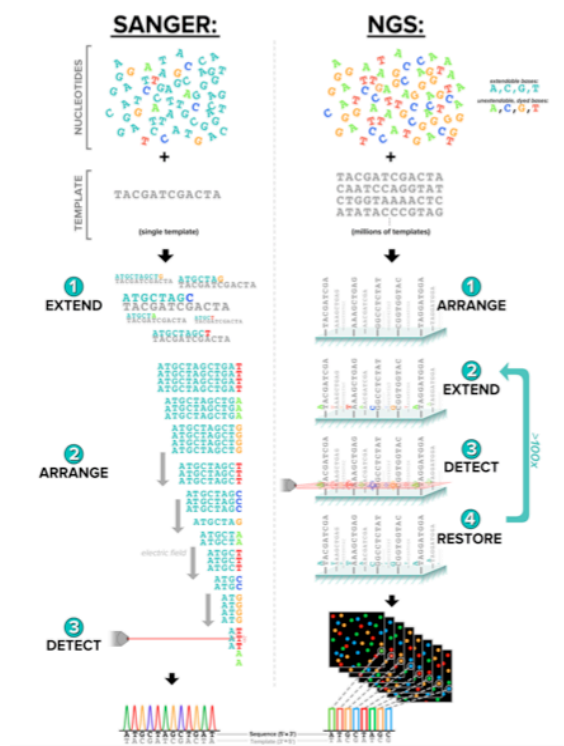
(NGS) δε σημαίνει μία ενιαία τεχνική, αλλά αναφέρεται σε συλλογή τεχνολογιών προσδιορισμού αλληλουχιών (Rabbani, Tekin et al. 2014).

### **1.2.1) Αλληλούχηση κατά Sanger και Αλληλούχησης Επόμενης Γενιάς**

Η αλληλούχηση κατά Sanger αποτελεί μέχρι πρότινος τη μέθοδο αναφοράς για το κλινικό εργαστήριο, όμως αντικαθίσταται ολοένα και περισσότερο από τις τεχνολογίες NGS. Οι δύο τεχνικές έχουν κοινή προέλευση καθώς χρησιμοποιούν την ικανότητα του DNA να αντιγράφεται. Η αντιγραφή απαιτεί κλώνο-εκμαγείο, ελεύθερα νουκλεοτίδια και ένζυμο πολυμεράση που τοποθετεί τα νουκλεοτίδια προς τη δημιουργία του συμπληρωματικού κλώνου. Η καινοτομία μετατροπής της αντιγραφής σε στρατηγική αλληλούχησης είναι η χρήση τροποποιημένων νουκλεοτιδικών βάσεων με φθορισμό. Στην ανάλυση κατά Sanger, μόνο ένα μικρό ποσοστό των βάσεων είναι τροποποιημένο, ενώ στο NGS όλες οι βάσεις έχουν τροποποιηθεί. Και στις δυο τεχνικές προσδιορισμού, όταν η πολυμεράση ενσωματώσει μια τροποποιημένη βάση στο νεοσυντιθέμενο κλώνο, η επέκταση του κλώνου σταματάει και ανιχνεύεται ο φθορισμός από το νουκλεοτίδιο που προστέθηκε (Εικόνα 4) (Muzzey, Evans et al. 2015).

Η βασική πρόκληση είναι η ερμηνεία του σήματος φθορισμού από τον αλληλουχητή. Στην ανάλυση Sanger τα τμήματα DNA που προκύπτουν έχουν διαφορετικό μέγεθος λόγω λήξης σε διαφορετικές θέσεις ανάλογα με το τροποποιημένο νουκλεοτίδιο και διατάσσονται σε ηλεκτρικό πεδίο για να διαχωριστούν. Καθώς τα μόρια DNA διαχωρίζονται, ένας ανιχνευτής καταγράφει την ένταση και το χρώμα του φθορισμού δίδοντας μια σειρά από κορυφές που αντιστοιχούν στην αλληλουχία του DNA (Εδάφιο 1.1.2). Η Αλληλούχηση Επόμενης Γενιάς χρησιμοποιεί διαχωρισμό βάσει θέσης, δηλαδή εκατομμύρια διαφορετικά τμήματα DNA προσδένονται σταθερά σε συγκεκριμένες θέσεις σε ειδική επιφάνεια. Σε κάθε τμήμα ο συμπληρωματικός κλώνος συντίθεται χρησιμοποιώντας τις τροποποιημένες βάσεις, οι οποίες ανιχνεύονται από ένα μικροσκόπιο-ανιχνευτή, το οποίο αναγνωρίζει τη θέση του κάθε κλώνου στην επιφάνεια και την ένταση σήματος που αυτό δίνει. Στη συνέχεια, κάτι που συμβαίνει μόνο στο NGS, τα τροποποιημένα νουκλεοτίδια

που έχουν προστεθεί μετατρέπονται σε απλά νουκλεοτίδια με ενζυμική επεξεργασία, επιτρέποντας την περαιτέρω επέκταση. Αυτό επιτρέπει να πραγματοποιηθούν πολλοί κύκλοι επέκτασης και απεικόνισης σε σύντομο χρόνο, με τελικό αποτέλεσμα μια ακολουθία νουκλεοτιδίων που ονομάζεται read (Εικόνα 1.4) (Muzzey, Evans et al. 2015).



**Εικόνα 1.4:** Το NGS αποτελεί μια τροποποιημένη, ψηφιακή και κατά πολύ μεγαλύτερη εφαρμογή της αλληλούχησης κατά Sanger. Επέκταση, διεύθετηση και ανίχνευση είναι κοινά βήματα και στα δύο πρωτόκολλα, αλλά με διαφορετική σειρά και με ένα επιπλέον βήμα, αυτό της αποκατάστασης στο NGS, που μετατρέπει τα νουκλεοτίδια επεκτάσιμα (Muzzey, Evans et al. 2015).

### 1.2.2) Τεχνολογίες Αλληλούχησης Επόμενης Γενιάς

Κάποιες από τις τεχνολογίες (πλατφόρμες) που έχουν αναπτυχθεί και χρησιμοποιούνται είναι η Roche 454 FLX, η Illumina Genome Analyzer (GA)/Solexa, η AB SOLiD και η PGM /Ion Torrent. Κάθε πλατφόρμα ενσωματώνει μια σύνθετη αλληλεπίδραση ενζυμολογίας, χημείας, υψηλής ανάλυσης ανίχνευση, υπολογιστικής και λογισμικής μηχανικής. Τα μέσα αυτά επιτρέπουν τη βελτιωμένη προετοιμασία του δείγματος πριν από την αλληλούχηση του DNA, την ίδια τη διαδικασία αλληλούχησης, αλλά και την επεξεργασία των αποτελεσμάτων, εξοικονομώντας χρόνο και απαιτώντας ελάχιστο εξοπλισμό πέρα του αυτοματοποιημένου συστήματος (Mardis 2008).

Το σύστημα Roche 454 FLX ήταν το πρώτο που χρησιμοποιήθηκε για εμπορική χρήση. Η αρχή της συγκεκριμένης τεχνολογίας βασίζεται στην

πυροανάλυση-pyrosequencing. Το 2005, όταν και κυκλοφόρησε το συγκεκριμένο μηχάνημα είχε τη δυνατότητα ανάγνωσης μέχρι και 150 βάσεις και τα δεδομένα από την κάθε αλληλούχιση έφταναν τα 20Mb. Το 2008, κυκλοφόρησε ένα νέο μηχάνημα το 454 GS FLX Titanium όπου το μέγεθος ανάγνωσης μπορούσε να φτάσει μέχρι και τις 700 βάσεις με ακρίβεια 99.9%, τα δεδομένα από την κάθε αντίδραση έφταναν τα 0.7Gb και ο χρόνος για την ολοκλήρωση της αντίδρασης ήταν μέσα σε 24 ώρες. Στα τέλη του 2009, η εταιρία Roche προχώρησε στον συνδυασμό 2 μηχανημάτων του GS Junior και του 454 συστήματος αλληλούχισης με σκοπό την απλούστευση του τρόπου παρασκευής της βιβλιοθήκης και της επεξεργασίας των δεδομένων. Τα δεδομένα από την κάθε αντίδραση φτάνουν τα 14Gb. Σημαντικό πλεονέκτημα σε σχέση με τις υπόλοιπες τεχνολογίες παραμένει ο χρόνος που απαιτείται από την έναρξη της αντίδρασης μέχρι την ολοκλήρωση της (10 ώρες) αλλά σημαντικό μειονέκτημα της παραμένει το υψηλό κόστος της (Liu, Li et al. 2012).

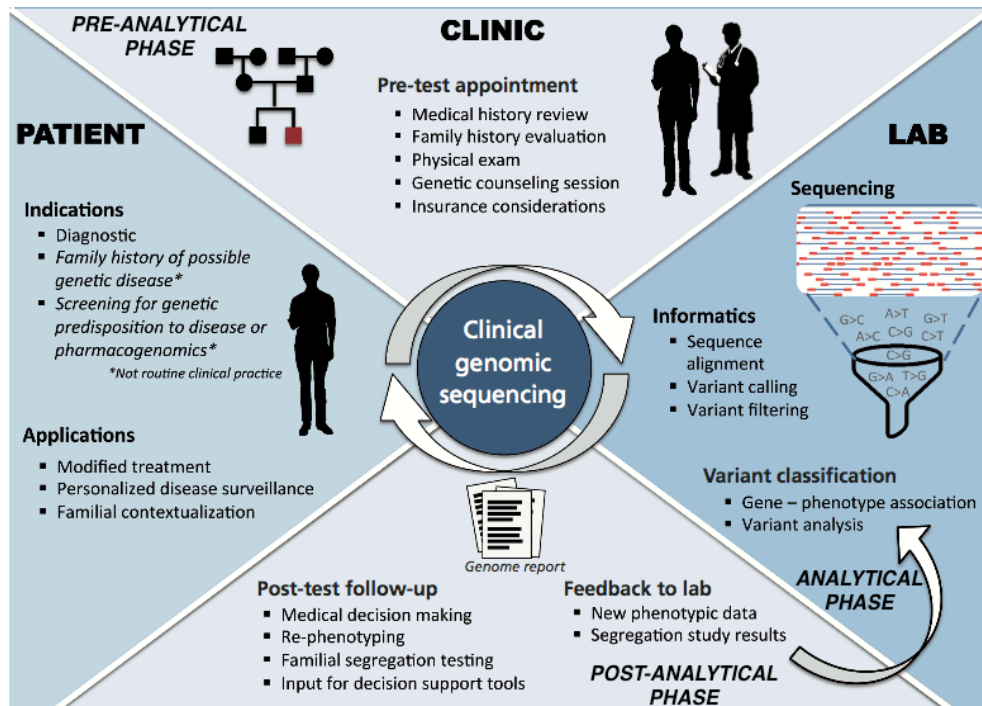
Το σύστημα Genome Analyzer (GA) εμφανίστηκε για πρώτη φορά το 2006 από την εταιρεία Solexa. Ένα χρόνο αργότερα, η εταιρία εξαγοράστηκε από την Illumina. Η αρχής της συγκεκριμένης πλατφόρμας για την αντίδραση της αλληλούχισης στηρίζεται στην αλληλούχιση μέσω σύνθεσης (sequencing by synthesis). Αρχικά, τα δεδομένα του μηχανήματος έφταναν το 1G/αντίδραση. Με διάφορες αλλαγές που πραγματοποιήθηκαν σε διάφορα στάδια της διαδικασίας τα δεδομένα έφταναν τα 20G/αντίδραση στην αρχή του 2009 ενώ μέχρι το τέλος του 2009 έφταναν τα 50G/αντίδραση. Στις αρχές του 2010, η Illumina παρουσίασε ένα νέο μηχάνημα το HiSeq 2000 (είχε την ίδια βασική αρχή με το GA) του οποίου τα δεδομένα έφταναν τα 200G/αντίδραση ενώ στη συνέχεια έφτανε έως τα 600G/αντίδραση (αντίδραση η οποία μπορούσε να έχει ολοκληρωθεί σε 8 ημέρες). Το 2011 κυκλοφόρησε στην αγορά το MiSeq, ένα μικρό σε μέγεθος μηχάνημα αλληλούχισης σχεδιασμένο για μικρά εργαστήρια και στοχεύει κυρίως σε κλινικές εφαρμογές, παρέχοντας δεδομένα μέσα σε 10 ώρες συμπεριλαμβανομένου και του χρόνου προετοιμασίας του δείγματος και της βιβλιοθήκης (Liu, Li et al. 2012).

Η πλατφόρμα του SOLiD (Sequencing by Oligonucleotide Ligation and Detection) κυκλοφόρησε εμπορικά το 2008. Σε αντίθεση με τις προηγούμενες πλατφόρμες η τεχνολογία της δε στηρίζεται στην αλληλούχιση με τη χρήση μιας πολυμεράσης αλλά στην αλληλούχιση με τη χρήση μιας λιγάσης. Στην αρχή η συγκεκριμένη πλατφόρμα είχε τη δυνατότητα αλληλούχισης έως 35 βάσεις και τα δεδομένα από την κάθε αντίδραση έφταναν τα 3Gb. Η ακρίβεια της μεθόδου έφτανε το 99.85%. Στο τέλος του 2010, κυκλοφόρησε μια νέα παραλλαγή του μηχανήματος το SOLiD 5500xl με βελτιωμένα χαρακτηριστικά τα οποία αφορούσαν το μέγεθος των παραγόμενων τμημάτων (85 βάσεις), την ακρίβεια (99.9%) και τα δεδομένα (30G/ αντίδραση).

Το 2010, η Ion Torrent κυκλοφόρησε στην αγορά το PGM (Personal Genome Machine). Όπως και το MiSeq, το Ion Torrent είναι ένα μικρό σε μέγεθος μηχάνημα το οποίο απευθύνεται σε μικρά εργαστήρια. Χρησιμοποιεί την τεχνολογία αλληλούχισης ημιαγωγού (semiconductor sequencing technology). Στηρίζεται στο γεγονός πως κατά την ενσωμάτωση ενός νουκλεοτιδίου από την πολυμεράση στην έλικα του DNA απελευθερώνεται ένα πρωτόνιο. Η ανίχνευση του πρωτονίου γίνεται μέσω αλλαγών που παρατηρούνται στο pH. Αποτελεί ένα από τα πρώτα μηχανήματα τα οποία δεν χρειάζονται κάμερες για την ανίχνευση των νουκλεοτιδίων, ούτε χρωστικές για την αλληλούχιση των τμημάτων, ούτε τροποποιήσεις των νουκλεοτιδίων με αποτέλεσμα να μειώνεται το κόστος και να γίνεται πιο απλή η διαδικασία της αλληλούχισης. Για την αλληλούχιση 200 βάσεων απαιτούνται 2 ώρες ενώ η διαδικασία για την παρασκευή της βιβλιοθήκης είναι λιγότερη από 6 ώρες για 8 δείγματα. Μέχρι σήμερα, η εταιρία έχει βελτιώσει τα «chips» καθώς το 318 μπορεί να δώσει δεδομένα που θα κυμαίνονται από 600Mb- 2Gb σε χρονικό διάστημα το οποίο κυμαίνεται από 4-7 ώρες (Liu, Li et al. 2012).

### 1.3) Αλληλούχηση Επόμενης Γενιάς στην κλινική πράξη

Η ροή εργασίας για την εφαρμογή αλληλούχησης επόμενης γενιάς στην κλινική πράξη περιλαμβάνει τρεις κύριες φάσεις, την προ-αναλυτική φάση, την αναλυτική φάση και την μετά-αναλυτική φάση (Εικόνα 1.5).



**Εικόνα 1.5:** Ροή εργασίας (Workflow) για αλληλούχηση επόμενης γενιάς στην κλινική πράξη (Krier, 2016).

#### 1.3.1) Προ-αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς

Ένα σημαντικό ερώτημα το οποίο πρέπει να απαντηθεί για την επιλογή των ασθενών που πιθανότατα θα ωφεληθούν από την εφαρμογή τέτοιου είδους εξέτασης. Η κλινική εφαρμογή αλληλούχησης επόμενης γενιάς, ενδείκνυται για την ανίχνευση παραλλαγών σε ασθενείς με φαινότυπο ύποπτο για Μενδελική (μονογονιδιακή) διαταραχή και όταν η πολυγονιδιακή προσέγγιση είναι δαπανηρή. Οι ασθενείς μπορεί να είναι οποιασδήποτε ηλικίας, αλλά κυρίως παιδιά, δεδομένου ότι πολλές γενετικές παθήσεις εκδηλώνονται στην παιδική ηλικία.

Η προ-αναλυτική φάση περιλαμβάνει εκ μέρους του κλινικού τη συλλογή πληροφοριών σχετικά με το οικογενειακό ιστορικό του ασθενούς και τη συστηματική αξιολόγηση του φαινοτύπου, τον προσδιορισμό των ενδείξεων για τη χρήση της συγκεκριμένης ανάλυσης, το είδος της δοκιμασίας που θα εφαρμοστεί, τη λήψη δείγματος από το οποίο θα απομονωθεί γενετικό υλικό, την ενημέρωση του ασθενούς και τέλος τη συγκατάθεση του. Επίσης, σημαντικό μέρος της φάσης αυτής αποτελεί η γενετική συμβουλευτική του ασθενούς πριν τη δοκιμασία για όλες τις παραμέτρους και περιορισμούς της εξέτασης. Η λήψη του οικογενειακού ιστορικού θα πρέπει να γίνει διαζοδικά για να εκτιμηθεί κατά πόσον υπάρχουν παρόμοιοι φαινότυποι σε άλλα μέλη της οικογένειας, καθώς και για την εκτίμηση-αξιολόγηση του μοτίβου κληρονομικότητας (Shashi, 2014).

Η επιλογή των μελών της οικογένειας που θα αλληλουχηθούν και θα αναλυθούν επηρεάζεται από το μοτίβο κληρονομικότητας που έχει βρεθεί για το συγκεκριμένο νόσημα (επικρατές ή υπολειπόμενο) και αν άλλα μέλη της οικογένειας με ή χωρίς παρόμοιο φαινότυπο είναι διαθέσιμα για ανάλυση. Σε μερικές περιπτώσεις μπορεί να είναι σημαντικό να εξεταστούν μη προσβεβλημένοι συγγενείς. Όταν υπάρχει υπόνοια για *de novo* παραλλαγή προτιμάται να εξετάζονται οι βιολογικοί γονείς του ασθενή και αν κανείς από τους δυο δεν την έχει τότε η παραλλαγή είναι σίγουρα *de novo*. Όταν υπάρχει υπόνοια για επικρατική ή υπολειπόμενη κληρονομικότητα, οι γονείς, τα αδέρφια ακόμα και μακρινοί συγγενείς μπορούν να εξεταστούν ανάλογα με το οικογενειακό ιστορικό, έτσι ώστε να μειωθούν οι πιθανές υποψήφιες παραλλαγές. Έτσι, η αλληλούχηση σε δύο προσβεβλημένα ξαδέρφια θα δώσει λιγότερες υποψήφιες παραλλαγές σε σχέση με την αλληλούχηση σε δύο προσβεβλημένα αδέρφια (Πίνακας 1.1). Σε άλλες περιπτώσεις, αλληλούχηση επόμενης γενιάς μπορεί να γίνει μόνο στο προσβεβλημένο άτομο και γονοτύπηση ορισμένων παραλλαγών σε ορισμένους συγγενείς (Biesecker, 2014).



### Πίνακας 1.1: Γενετικοί Παράγοντες για την επιλογή των συγγενών προς εξέταση στην Αλληλούχησης Επόμενης Γενιάς (Biesecker, 2014).

#### Μελέτη για Αυτοσωμική Επικρατής Κληρονομήση:

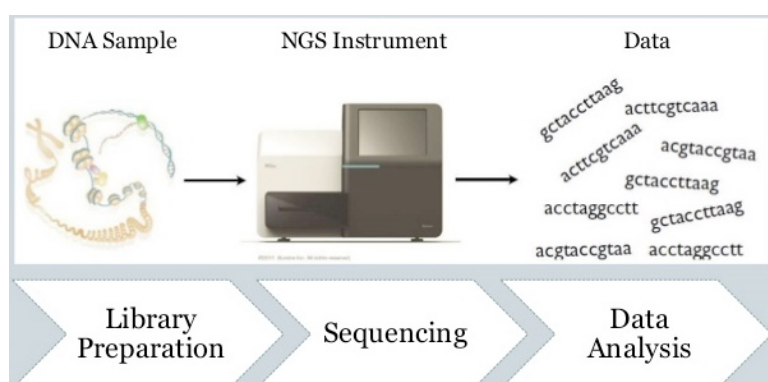
- Αν υπάρχει υποψία για *de novo* παραλλαγή του σχετιζόμενου γονιδίου με τη διαταραχή να προκαλείται από ετερόζυγη (μονοαλληλική) μεταλλαγή, η δοκιμή με μεγαλύτερη αξία είναι η ανάλυση σε τρίο που αποτελείται από ένα παιδί και τους δύο βιολογικούς γονείς.
- Αν αξιολογηθεί ένας φαινότυπος για αυτοσωμική επικρατή κληρονομήση σε μια οικογένεια, αλληλούχηση δύο προσβεβλημένων συγγενών έχει μεγαλύτερη αξία από την αλληλούχηση δύο κοντινών συγγενών. Η ανάλυση των μακρινών συγγενών συνίσταται γιατί θα προκύψουν λιγότερες κοινές παραλλαγές να μοιράζονται.

#### Μελέτη για Αυτοσωμική Υπολειπόμενη Κληρονομήση:

- Αν μελετάται η ύπαρξη αυτοσωμικού υπολειπόμενου γονιδίου τότε είναι χρήσιμη η δοκιμή σε τρίο, εξετάζοντας το παιδί και τους δύο βιολογικούς γονείς.
- Αν υπάρχει συγγένεια τότε η δοκιμή σε τρίο είναι λιγότερο χρήσιμη, επειδή η αλληλουχία του παιδιού θα έχει πολλές περιοχές ομοζυγωτίας.

### 1.3.2) Αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς

Η αναλυτική φάση της ροής εργασίας για την Αλληλούχηση Επόμενης Γενιάς στην κλινική πράξη περιλαμβάνει έναν αριθμό εργαστηριακών βημάτων συμπεριλαμβανομένου την προετοιμασία του δείγματος, την κατασκευή βιβλιοθήκης, τον εμπλουτισμό του στόχου και την αντίδραση αλληλούχησης. Επίσης, περιλαμβάνει τρία επίπεδα ανάλυσης δεδομένων, την πρωτοβάθμια (primary), τη δευτεροβάθμια (secondary) και την τριτοβάθμια ανάλυση και την κατηγοριοποίηση των παραλλαγών που εξάχθηκαν (Εικόνα 1.6).



**Εικόνα 1.6:** Ροή εργασίας αναλυτικής φάσης Αλληλούχησης Επόμενης Γενιάς.

Η εργαστηριακή επεξεργασία του δείγματος ξεκινάει με την απομόνωση του γενετικού υλικού, τον προσδιορισμό της κατάλληλης ποσότητας και της απαιτούμενης ποιότητας. Η κατασκευή της βιβλιοθήκης περιλαμβάνει την παραγωγή θραυσμάτων DNA συγκεκριμένου μεγέθους με ενζυμική κατεργασία ή χρήση υπερήχων και προσθήκη ανταπτόρων σε κάθε άκρο.

Οι αντάπτορες-προσαρμοστές είναι ειδικές αλληλουχίες συμπληρωματικές με εκκινητές PCR, αλληλούχησης. Επίσης, πραγματοποιείται επισήμανση (Barcoding) των θραυσμάτων κάθε δείγματος με μοναδικές αλληλουχίες 3-5 βάσεων (tags), επιτρέποντας την ανάμειξη-ομαδοποίηση (pooling) των δειγμάτων ελαττώνοντας το κόστος της αντίδρασης (Rehm, 2013).

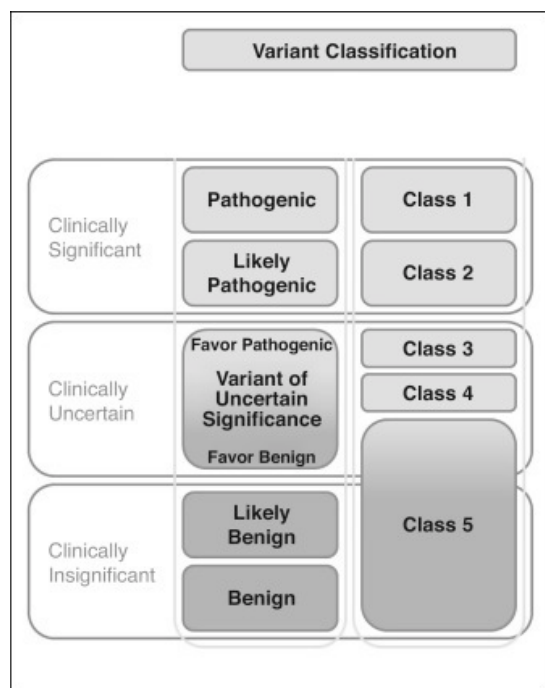
Στις περισσότερες εφαρμογές/πλατφόρμες αλληλούχησης είναι αναγκαίος ο εμπλουτισμός των δειγμάτων πριν την αντίδραση αλληλούχησης ενισχύοντας το στόχο (επιθυμητά γονίδια, σύνολο εξονίων). Οι μέθοδοι που χρησιμοποιούνται για τον εμπλουτισμό βασίζονται στην PCR ή στον υβριδισμό. Οι μέχρι τώρα τεχνολογίες αλληλούχησης (Εδάφιο 1.2.2) βασίζονται στην ικανότητα να πραγματοποιούν πολλές χημικές αντιδράσεις παράλληλα. Οι χημικές αυτές αντιδράσεις δίνουν τη δυνατότητα για αλληλούχηση μέσω σύνθεσης ή αλληλούχηση μέσω σύνδεσης με ανάστροφους τερματιστές, επιλογή σφαιριδίων και ανίχνευση ιόντων. Κάθε πλατφόρμα έχει συγκεκριμένες παραμέτρους σχετικές με τις απαιτήσεις κάθε εργαστηρίου, όπως το κόστος του μηχανήματος και των αναλωσίμων, το χρόνο αντίδρασης, το μέγεθος των αλληλουχιών που εξάγει και το κόστος ανά δείγμα (Rehm, 2013, Endrullat, 2016).

Ο προσδιορισμός της αλληλουχίας του γονιδιώματος είναι μόνο η αρχή της διαδικασίας για τη μετατροπή ενός δείγματος DNA σε ουσιαστική γενετική πληροφορία. Το επόμενο βήμα της συλλογής, ανάλυσης και αποθήκευσης των δεδομένων περιλαμβάνει την εκτεταμένη χρήση διάφορων υπολογιστικών μεθόδων για τη μετατροπή των πρώτων δεδομένων σε πληροφορίες αλληλουχίας και την εφαρμογή τεχνικών βιοπληροφορικής για την ερμηνεία αυτής της αλληλουχίας. Η τεράστια ποσότητα δεδομένων που παράγονται από το NGS έχει μεταφέρει το φόρτο εργασίας από τον πάγκο και την προετοιμασία του δείγματος, στον υπολογιστή και την ανάλυση των δεδομένων. Παράλληλα με την ανάπτυξη τεχνολογιών αλληλούχησης, απαιτείται η ταυτόχρονη ανάπτυξη κατάλληλων βιοπληροφορικών προγραμμάτων για να γίνει πράξη η κλινική ερμηνεία των γενωμικών πληροφοριών (Moorthie, 2013).

Το πρώτο επίπεδο ανάλυσης, η πρωτοβάθμια ανάλυση (Primary analysis), περιλαμβάνει την ταυτοποίηση παρουσίας ενός συγκεκριμένου νουκλεοτιδίου σε συγκεκριμένη θέση στην αλληλουχία που παράγεται (Base calling). Το συγκεκριμένο επίπεδο ανάλυσης πραγματοποιείται από την κάθε πλατφόρμα αλληλούχησης χωρίς δυνατότητα εξωτερικής επέμβασης. Στο δεύτερο επίπεδο ανάλυσης, δευτεροβάθμια ανάλυση (Secondary analysis), περιλαμβάνεται η στοίχιση των παραγόμενων αλληλουχιών (reads) με την αλληλουχία του γονιδιώματος αναφοράς GRCh37 ή με την πιο πρόσφατη έκδοσή του GRCh38 (Read alignment/mapping) και την ανίχνευση των παραλλαγών που προκύπτουν σε σύγκριση με το γονιδίωμα αναφοράς (Variant calling). Οι παραλλαγές αυτές καταγράφονται σε ένα ειδικό αρχείο VCF (Variant Call Format), το οποίο εν συνεχεία επεξεργάζεται στο επόμενο επίπεδο ανάλυσης. Στο τρίτο επίπεδο ανάλυσης, η τριτοβάθμια ανάλυση (Tertiary analysis), πραγματοποιείται η ταυτοποίηση των παραλλαγών από τη δευτεροβάθμια ανάλυση προσθέτοντας πληροφορίες για τη χρωμοσωμική θέση, το γονίδιο, το εξόνιο που εδράζονται, το είδος κάθε παραλλαγής, την επίπτωση στην πρωτεΐνη και άλλες βιβλιογραφικές πληροφορίες (Variant analysis/annotation) (Kassahn, 2014). Η τριτοβάθμια ανάλυση εμφανίζει τα σημαντικότερα εμπόδια για την ευρύτερη κλινική εφαρμογή του NGS. Η ερμηνεία των δεδομένων αλληλουχίας εξακολουθεί να είναι ένα αναπτυσσόμενο πεδίο όσον αφορά αλγορίθμους, κατάλληλα αναλυτικά εργαλεία και αποτελεσματικών βάσεων τεκμηρίωσης γονοτύπου-φαινοτύπου του ανθρώπου (Moorthie, 2013).

Το τελευταίο μέρος της αναλυτικής φάσης είναι η ταξινόμηση των παραλλαγών ανάλογα με την επίπτωσή τους στο φαινότυπο του ασθενούς (Variant interpretation). Σύμφωνα με τις τελευταίες οδηγίες (Richards, 2015), εφαρμόζεται ένα σύστημα ταξινόμησης πέντε βαθμίδων (ήπια, πιθανώς ήπια, παραλλαγή αβέβαιης σημασίας-VUS, πιθανώς παθολογική, παθολογική) (Εικόνα 1.7) (Hoskinson et al., 2017). Είναι σημαντικό να γίνει διάκριση μεταξύ των παραλλαγών και αξιολόγηση του κατά πόσο είναι επιβλαβής για την ερμηνεία του κλινικού αποτελέσματος. Οι καλοήθειες παραλλαγές πιθανότατα δε θα ερμηνευθούν περαιτέρω. Αντίθετα, αυτές που προσδιορίστηκαν ότι σχετίζονται με το φαινότυπο, θα αξιολογηθούν και θα

ερμηνευθούν ενδελεχώς στη μετα-αναλυτική φάση (Rehm, 2013). Η αναλυτική φάση θα αναλυθεί λεπτομερώς στο Κεφάλαιο 2.



**Εικόνα 1.7:** Ταξινόμηση των παραλλαγών σύμφωνα με το σύστημα των πέντε βαθμίδων (Hoskinson et al., 2017).

### 1.3.3) Μετα-αναλυτική φάση κλινικής εφαρμογής Αλληλούχησης Επόμενης Γενιάς

Η μετα-αναλυτική φάση ξεκινάει με την αξιολόγηση και ερμηνεία των αποτελεσμάτων της αναλυτικής φάσης και τελειώνει με τη σύνταξη της αναφοράς και παράδοσής της στον ασθενή με γενετική συμβουλευτική. Τα αποτελέσματα που προκύπτουν από την αναλυτική φάση ποικίλουν σημαντικά. Σε ορισμένες περιπτώσεις θεωρείται ότι μία μόνο παραλλαγή είναι η πιθανή αιτία της νόσου, ενώ σε άλλες περιπτώσεις περιπτώσεις εντοπίζονται περισσότερες υποψήφιες παραλλαγές οι οποίες πρέπει να αξιολογηθούν. Σε πολλές περιπτώσεις όμως, δεν εντοπίζονται υπεύθυνες μεταλλάξεις.

Οι δύο κύριες εκτιμήσεις για την αξιολόγηση των αποτελεσμάτων είναι η αναλυτική εγκυρότητα και η κλινική εγκυρότητα. Η αναλυτική εγκυρότητα είναι μέτρο πιθανότητας ότι ο ασθενής έχει στην πραγματικότητα το συγκεκριμένο γονότυπο που εμφανίζεται στα αποτελέσματα. Η κλινική εγκυρότητα είναι πιο περίπλοκη και αφορά τον προσδιορισμό ότι μια συγκεκριμένη ασθένεια προκαλείται πραγματικά από παραλλαγές σε ένα συγκεκριμένο γονίδιο και ότι

η συγκεκριμένη παραλλαγή που ανιχνεύθηκε είναι πράγματι παθογόνος (Biesecker, 2014). Τα θετικά ευρήματα από WES ή WGS είναι εξαιρετικά ακριβή (αληθώς θετικά), αλλά υπάρχουν αρκετά ψευδώς αρνητικά αποτελέσματα ανάλογα με την περιοχή του γονιδιώματος που εξετάζεται. Για το λόγο αυτό το WES ή WGS δεν μπορούν ακόμα να υποκαταστήσουν τη στοχευμένη ανάλυση. Αξίζει να σημειωθεί ότι τα περισσότερα εργαστήρια επιβεβαιώνουν τα θετικά αποτελέσματα με καθιερωμένες μεθόδους αναφοράς, όπως Sanger αλληλούχηση (Kassahn, 2014). Έτσι, οι παράμετροι ευαισθησίας και ειδικότητας αποτελούν κύριο παράγοντα για να εφαρμοστεί αυτή η τεχνολογία.

Όπως προαναφέρθηκε, ο προσδιορισμός της κλινικής εγκυρότητας είναι πιο δύσκολος από της αναλυτικής εγκυρότητας. Η γενική προσέγγιση είναι να συγκριθούν οι παραλλαγές που εξετάζονται με βάσεις δεδομένων γνωστών παραλλαγών, οι οποίες βασίζονται σε βιβλιογραφικά δεδομένα και σε συσχετίσεις γονοτύπου-φαινοτύπου (Πίνακας 1.2) (Brookes, 2015). Ωστόσο, στη βιβλιογραφία υπάρχουν πολλές ψευδείς καταχωρίσεις παραλλαγών συσχετιζόμενες με φαινοτύπους, γι' αυτό το λόγο θα πρέπει να γίνεται σωστή και έγκυρη αξιολόγηση των παραλλαγών που ανιχνεύθηκαν. Η ταυτοποίηση των παραλλαγών που ευθύνονται για την κληρονομική διαταραχή ή το προηγουμένως μη διαγνωσθέν σύνδρομο, έχει μεγάλη κλινική χρησιμότητα καθώς μπορεί να οδηγήσει σε στοχευμένη θεραπεία ή σε σωστή διαχείριση της νόσου και να αλλάξει το κλινικό αποτέλεσμα. Στις περιπτώσεις όπου το εύρημα δεν αλλάζει τη θεραπεία ή τη διαχείριση της νόσου, μπορεί να τερματίσει μια δαπανηρή και επεμβατική πορεία διάγνωσης ή μπορεί να δώσει εκτιμήσεις κινδύνου επανεμφάνισης, ακόμα και πληροφορίες πρόληψης, όπως γενετική ευαισθησία για καρκίνο (Delaney, Hultner et al. 2016).

Σε αυτό το στάδιο οι κλινικοί σε συνεργασία με το εργαστήριο θα πρέπει να συσχετίσουν τα κλινικά αποτελέσματα με τα ιατρικά στοιχεία του ασθενούς. Μερικές φορές τα ευρήματα από το WES ή WGS μπορεί να καθοδηγήσουν τον κλινικό για επιπλέον λήψη ιστορικού από τον ασθενή ή διεξαγωγή εξειδικευμένων εξετάσεων ή μπορεί να οδηγήσουν και σε μη συσχέτιση με το

φαινότυπο του ασθενούς. Όταν τα ευρήματα περιλαμβάνουν παραλλαγές που θεωρούνται παθογόνες και υποστηρίζουν τη διαγνωστική σχέση με την ασθένεια, η αναφορά που παραδίδεται στον ασθενή (ή στην οικογένεια), θα πρέπει να περιέχει πληροφορίες για την κληρονομικότητα, τη διεισδυτικότητα, την πρόγνωση, τη θεραπεία αν υπάρχει, τη διαχείριση της κατάστασης με παραπομπή σε ειδικό, βιβλιογραφικές αναφορές και τις παραμέτρους της ανάλυσης. Όταν τα ευρήματα περιλαμβάνουν παραλλαγές που πιθανώς να σχετίζονται ή μη με το φαινότυπο, μπορεί να χρειαστεί πρόσθετη βιβλιογραφία, φαινοτυπικές μελέτες, ακόμα και λειτουργικές μελέτες για να αποκλειστούν ή να επιβεβαιωθούν τα αποτελέσματα. Αντίθετα, στις περιπτώσεις που τα ευρήματα δεν καταλήγουν σε συγκεκριμένο συμπέρασμα, μελλοντικές μελέτες με ανανεωμένες βάσεις δεδομένων μπορεί να οδηγήσουν στην αναγνώριση ότι το αρνητικό αποτέλεσμα είναι εν τέλει παθογόνο. Επομένως, είναι σημαντικό να γίνεται τακτική επανεξέταση τέτοιων αποτελεσμάτων με στόχο τον εντοπισμό προηγουμένως άγνωστων παραλλαγών. Επίσης, αξίζει να σημειωθεί ότι η γνωστοποίηση στον ασθενή ενός μη θετικού αποτελέσματος με δοκιμή WES ή WGS δεν αποκλείει την παρουσία μιας αιτιολογικής μετάλλαξης (MacArthur 2014, Posey 2016).

Η αλληλούχηση επόμενης γενιάς σε επίπεδο WES ή WGS μπορεί να παράγει αποτελέσματα που δε σχετίζονται με τις ενδείξεις για τις οποίες ζητήθηκε η συγκεκριμένη ανάλυση (Δευτερεύοντα ευρήματα). Για το σκοπό αυτό έχει προταθεί από το American College of Medical Genetics and Genomics να αναφέρονται ευρήματα σε ένα σύνολο 56 γονίδιων μετά από συγκατάθεση του εξεταζόμενου. Τα γονίδια αυτά σχετίζονται με καλά χαρακτηρισμένες ασθένειες (καρδιοπάθειες, καρκίνο κ.α.). Η ενημέρωση και συγκατάθεση του ασθενούς για τα δευτερεύοντα ευρήματα είναι ζωτικής σημασίας, καθώς εμπίπτουν σε ευαίσθητα βιοηθικά ζητήματα. Εκτιμάται ότι το 1-3% των ασθενών που υποβάλλονται σε WES έχουν τέτοιο εύρημα, για το οποίο θα πρέπει να γίνει από τον κλινικό πρόσθετη λήψη οικογενειακού ιστορικού, αλλά και γενετική συμβουλευτική (Green, 2013).

**Πίνακας 1.2:** Βάσεις δεδομένων που επικεντρώνονται στη σχέση γονοτύπου-φαινοτύπου (Brookes, 2015).

Database	Scope and scale	Standards	Data entry	Data access policies	Refs
<b>Gene variation database (LSDB or MDB)</b>					
ClinVar	<ul style="list-style-type: none"> <li>Genetic variants and phenotypes</li> <li>125,520 variants</li> </ul>	HGVS, HPO, MeSH, OMIM, RefSeq, SO	Curators and users	P	63
Human Gene Mutation Database (HGMD)	<ul style="list-style-type: none"> <li>Genetic variants and phenotypes</li> <li>163,610 variants</li> </ul>	HGNC, HGVS	Curators	Com, P	96
Leiden Open Variation Databases (LOVD)	<ul style="list-style-type: none"> <li>Genetic variants and phenotypes</li> <li>3,334,104 variants (2,400,084 unique) in 248,807 individuals in 86 LOVD installations</li> </ul>	HGVS, Mutalyzer	Curators	P, Cs	12
Universal Mutation Database (UMD)	<ul style="list-style-type: none"> <li>Genetic variants and phenotypes</li> <li>90,383 variants in 40 databases</li> </ul>	HGVS	Curators	P, Cs	13
Amyotrophic Lateral Sclerosis Online genetics Database (ALSoD)	<ul style="list-style-type: none"> <li>LSDB</li> <li>116 genes associated with amyotrophic lateral sclerosis</li> <li>569 variations</li> </ul>	HGNC	Curators	P	97
CFTR2	<ul style="list-style-type: none"> <li>Cystic fibrosis LSDB</li> <li>88,000 patients</li> </ul>	HGNC, HGVS	Curators	P	23
Fanconi Anemia Mutation Database	<ul style="list-style-type: none"> <li>LSDB</li> <li>16 genes associated with Fanconi anaemia–BRCA pathway</li> <li>~3,000 variations</li> </ul>	HGNC, HGVS, LRG	Curators	P	98
Osteogenesis Imperfecta Variant Database	<ul style="list-style-type: none"> <li>LSDB</li> <li>16 genes associated with osteogenesis imperfecta</li> <li>~1,500 variations</li> </ul>	HGNC, HGVS, LRG	Curators	P	99
IDbases	<ul style="list-style-type: none"> <li>LSDB</li> <li>131 genes associated with immunodeficiency</li> <li>Data for 7,292 patients</li> </ul>	HGNC, HGVS, VariO	Curators	P	100
MITOMAP	<ul style="list-style-type: none"> <li>LSDB</li> <li>Mitochondrial DNA variation</li> <li>1,746 variants</li> </ul>		Curators	P	101
FINDbase	<ul style="list-style-type: none"> <li>Aggregated information on national and ethnic variation frequencies</li> <li>~100 NEMDBs</li> </ul>	HGNC, HGVS	Curators	P	102
<b>Array-CGH, WES, WGS (rare disease)</b>					
DECIPHER	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>Diagnostics and discovery.</li> <li>42,815 cases</li> </ul>	HGVS, HGNC, HPO	Users	P, Cs, RG, MM	38, 103
PhenomeCentral	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>Genomic "matchmaking"</li> <li>600 cases</li> </ul>	HPO, VCF	Users	MM	65
PhenoDB	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>Diagnostics and discovery.</li> <li>3,300 cases</li> </ul>	EoM, HPO, OMIM, PhenoDB	Users	MM	104
GeneMatcher	<ul style="list-style-type: none"> <li>Gene matching (discovery)</li> <li>668 genes</li> </ul>	HGNC, Ensembl, Entrez Gene, OMIM	Users	MM	105
<b>Mendelian and other rare disease knowledge bases</b>					
Online Mendelian Inheritance in Man (OMIM)	<ul style="list-style-type: none"> <li>Knowledge base</li> <li>22,644 entries (genes or diseases)</li> </ul>	HGNC, HPO, ICD, OMIM, PhenoDB, SNoMED, UMLS	Curators	Ac	106
Orphanet	<ul style="list-style-type: none"> <li>Knowledge base</li> <li>5,833 disease entries, copious data on other rare disease topics</li> </ul>	HGNC, ICD, MedDRA, MeSH, OMIM, UMLS, Uniprot	Curators	Ac	73
Monarch Initiative knowledge base	<ul style="list-style-type: none"> <li>Human and model organism genetics and phenotypes</li> <li>36K diseases, 33K phenotypes, 500K genotypes, 30K genes, 2M curated phenotype associations, &gt;100 species</li> </ul>	HPO, MPO	Curators	P	107

**Πίνακας 1.2 (συνέχεια):** Βάσεις δεδομένων που επικεντρώνονται στη σχέση γονοτύπου-φαινοτύπου (Brookes, 2015).

Database	Scope and scale	Standards	Data entry	Data access policies	Refs
<b>Cancer genomics and variations</b>					
Cancer Genomics Hub	<ul style="list-style-type: none"> <li>Genetic and phenotypic data repository</li> <li>82,140 files (1870 Tb)</li> </ul>	Sequence Read Archive Metadata XML	NCI projects, curators	P, CA	108
Catalogue Of Somatic Mutations In Cancer (COSMIC)	<ul style="list-style-type: none"> <li>Variation and genetic and phenotypic data</li> <li>2,139,424 unique variants</li> </ul>	HGNC, CCDS	Curators	P	109
DriverDB	<ul style="list-style-type: none"> <li>Variation and genetic/phenotypic data</li> <li>6,079 datasets</li> </ul>	HGNC	Curators	P	110
<b>Genotype-phenotype information for GWAS and other studies</b>					
Database of Genotypes and Phenotypes (dbGAP)	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>508 studies</li> </ul>	dbGAP, XML	Curators (focus on NIH-funded projects)	P, CA	43
European Variation Archive (EVA)	<ul style="list-style-type: none"> <li>All types of genetic variants from any species</li> <li>~40 studies, representing 35 species, describing ~400 million unique alleles from more than 150,000 samples</li> </ul>	VCF, dbSNP	Users, curators	P	
European Genome-Phenome Archive (EGA)	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>1,555 datasets</li> </ul>	VCF, FASTQ, BAM, EFO	Users	P, CA	42
GWAS Catalog	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>18,697 associations</li> </ul>	dbSNP, HGNC	Curators	P	111
GWAS Central	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>&gt;75 million associations</li> </ul>	dbSNP, HGNC, HPO, MeSH	Curators	P, CA	47
GWASdb	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>272,918 associations</li> </ul>	dbSNP, DO, HPO	Curators	P	112
Human Genome Variation Database	<ul style="list-style-type: none"> <li>Genetic and phenotypic data</li> <li>~100 datasets in 6 integrated databases</li> </ul>	HGNC, HGVS, dbSNP	Users	P, CA	113
<b>Pharmacogenomics</b>					
PharmacoGenomics Database (PharmGKB)	<ul style="list-style-type: none"> <li>Pharmacogenomics knowledge resource</li> <li>Extensive data on variants, pathways, dosing, clinical annotations, drug labels</li> </ul>	dbSNP, HGNC, MeSH, SNoMED, UMLS	Curators, NLP	Ac	114

Ac, open to academics, but commercial entities require license; CA, controlled access; CCDS, Consensus CDS; CGH, comparative genome hybridization; Com, commercial; Cs, restricted to consortium members; dbSNP, Database of Single Nucleotide Polymorphism; DO, Disease Ontology<sup>9</sup>; EFO, Experimental Factor Ontology; EoM, Elements of Morphology<sup>115</sup>; GWAS, genome-wide association study; HGNC, HUGO Gene Nomenclature Committee; HGVS, Human Genome Variation Society; HPO, Human Phenotype Ontology; ICD, International Classification of Diseases; LRG, Locus Reference Genomic; LSDB, Locus-specific database; MDB, variation (mutation) database; MedDRA, Medical Dictionary for Regulatory Activities; MeSH, Medical Subject Headings; MM, MatchMaking (finding similar patients); MPO, Mammalian Phenotype Ontology; NCI, National Cancer Institute; NEMDB, national and ethnic variation (mutation) databases; NIH, National Institutes of Health; OMIM, Online Mendelian Inheritance in Man; P, public; RG, access restricted to a specific research group; SNoMED, Systematized Nomenclature of Medical Terms; SO, Sequence Ontology; UMLS, Unified Medical Language System; VariO, Variation Ontology; VCF, Variant Call Format; WES, whole-exome sequencing; WGS, whole-genome sequencing; XML, Extensible Markup Language. \*Further details are available at the websites of the databases. Owing to space constraints, many important databases had to be omitted. Databases are divided into major categories, but it is recognized that many of the databases can be considered to belong to multiple categories. Scope refers to the major focus of the database. If applicable, information is given about the number of items currently contained in the database. Standards indicates the major terminologies or ontologies that the database uses to annotate and organize data. Data entry indicates whether the data in the database are primarily entered by curators or by users of the database, and if natural language processing (NLP) is employed to gather data. Note that for data access policies, databases with more than one access level tend to provide more information as data access becomes stricter.

#### 1.4) Κύρια επίπεδα Αλληλούχησης Επόμενης Γενιάς

Το ανθρώπινο γονιδίωμα περιλαμβάνει  $3 \times 10^9$  βάσεις έχοντας κωδικοποιούσες και μη κωδικοποιούσες αλληλουχίες. Περίπου  $3 \times 10^7$  ζεύγη βάσεων (1%) του γονιδιώματος είναι οι κωδικοποιούσες περιοχές. Επίσης το λιγότερο από το 10% της αλληλουχίας ολόκληρου του γονιδιώματος έχει χαρακτηριστεί κλινικά προσφέροντας λειτουργική γνώση. Επιπλέον, εκτιμάται ότι το 85% των μεταλλάξεων που προκαλούν ασθένειες βρίσκονται στις κωδικοποιούσες και λειτουργικές περιοχές του γονιδιώματος (Majewski,



Schwartzentruber et al. 2011). Η αλληλούχηση επόμενης γενιάς μπορεί να χρησιμοποιηθεί σε διάφορα επίπεδα πολυπλοκότητας, όσον αφορά το τμήμα του γονιδιώματος που θα ακολουθηθεί. Τα επίπεδα αυτά περιλαμβάνουν αλληλούχηση πάνελ γονιδίων που σχετίζονται με κάποια ασθένεια (Disease-targeted gene panel), αλληλούχηση των εξονίων του γονιδιώματος (Whole Exome Sequencing-WES) και αλληλούχηση όλου του γονιδιώματος (Whole Genome Sequencing-WGS) (Zhang, Chiodini et al. 2011).

Η αλληλούχηση με πάνελ γονιδίων περιλαμβάνει ανάλυση γονιδίων που έχουν συσχετιστεί με κάποια ασθένεια. Πρόκειται για την πιο ευρέως χρησιμοποιούμενη εφαρμογή NGS στα κλινικά εργαστήρια, διότι παρέχει μια οικονομική, αποδοτική και ολοκληρωμένη διαγνωστική προσέγγιση για τον προσδιορισμό της αιτιολογίας μιας ασθένειας. Η ανάλυση με πάνελ γονιδίων επιτρέπει μεγαλύτερο βάθος κάλυψης (Depth coverage) για αυξημένη αναλυτική ευαισθησία και ειδικότητα. Το βελτιωμένο βάθος αυξάνει την πεποίθηση για ετεροζυγωτία και την πιθανότητα προσδιορισμού μωσαϊκισμού ή ετερογένειας σε μιτοχονδριακές ή ογκολογικές εφαρμογές. Επιπλέον, η αλληλούχηση μόνο των γονιδίων που έχουν συσχετιστεί με μια ασθένεια δίνει μεγαλύτερη ικανότητα ερμηνείας των αποτελεσμάτων σε κλινικό πλαίσιο, ανάλυση περισσότερων δειγμάτων ανά κύκλο αντίδρασης και διαχείριση δεδομένα (Rehm, Bale et al. 2013).

Η ανάλυση των εξονίων του γονιδιώματος καλύπτει όλες τις κωδικοποιούσες περιοχές του γονιδιώματος, οι οποίες καλύπτουν το 1-2% αυτού. Η ανάλυση με WES απαιτεί την προετοιμασία του δείγματος πριν την αντίδραση αλληλούχησης με εμπλουτισμό των περιοχών κωδικοποίησης με υβριδισμό ή PCR. Χρησιμοποιείται για τον εντοπισμό παραλλαγών (variants) σε γονίδια που σχετίζονται με συγκεκριμένη ασθένεια, όπως και τα πάνελ γονιδίων, αλλά και σε ερευνητικό πεδίο για την εύρεση νέων υποψήφιων γονιδίων σχετιζόμενων με κάποια ασθένεια. Το βάθος κάλυψης και το κόστος της αλληλούχησης των εξονίων είναι μεταξύ του πάνελ γονιδίων και της αλληλούχησης ολόκληρου του γονιδιώματος (Rehm, Bale et al. 2013).

Η αλληλούχηση όλου του γονιδιώματος καλύπτει τις κωδικοποιούσες και τις μη κωδικοποιούσες περιοχές του γονιδιώματος. Το βασικό πλεονέκτημα του WGS είναι ότι δεν απαιτείται προετοιμασία του δείγματος με εμπλουτισμό περιοχών πριν την αντίδραση. Τα δεδομένα που προκύπτουν από την ανάλυση μπορούν να διερευνηθούν για παραλλαγές στις κωδικές και στις μη κωδικές περιοχές, για Copy-Number Variants (CNVs) ή δομικές παραλλαγές. (Rehm, Bale et al. 2013).

Συγκρίνοντας τα τρία επίπεδα, είναι ξεκάθαρο ότι το WGS είναι ανώτερο από τα άλλα δύο, παράγοντας πιο πλήρη δεδομένα για το γονιδίωμα που εξετάζεται. Ωστόσο, παρά τις πολλές διαφορετικές τεχνολογίες που έχουν αναπτυχθεί, το κόστος είναι πολύ υψηλό σε σχέση με το WES και ο χρόνος ανάλυσης του μεγάλου όγκου δεδομένων που εξάγονται, εξακολουθούν να περιορίζουν την εφαρμογή του στην κλινική πράξη. Σε αντίθεση, το WES είναι σημαντικά πιο οικονομικό από το WGS και αποτελεί μια αποτελεσματική στρατηγική για την ανίχνευση μεταλλαγών που ευθύνονται για ασθένειες. Η ανάλυση πάνελ γονιδίων που απομονώθηκαν στοχευμένα, αντιπροσωπεύει μια εναλλακτική τακτική, η οποία έχει σημαντικά χαμηλότερο κόστος αλληλούχησης, αλλά θα επιτύχει μόνο εάν το γονίδιο που προκαλεί την ασθένεια περιλαμβάνεται στο πάνελ. Ένα πλεονέκτημα είναι ότι η στόχευση περιορίζει την πιθανότητα εύρεσης δευτερευόντων ευρημάτων και επιτρέπει υψηλότερη κάλυψη (Sun, Ruivenkamp et al. 2015).

Η μέση οδός μεταξύ αυτών των προσεγγίσεων είναι η χρήση WES βασισμένη σε βιοπληροφορικά πάνελ γονιδίων. Σε περίπτωση που η βιοπληροφορική αυτή προσέγγιση αδυνατεί να προσδιορίσει μια παραλλαγή, η προσοχή μπορεί να επικεντρωθεί εκ νέου σε παραλλαγές γονιδίων έξω από τα αρχικά πάνελ. Ένα επιπλέον πλεονέκτημα είναι ότι νέα γονίδια που σχετίζονται με τη νόσο μπορούν να προστεθούν πολύ εύκολα και γρήγορα με τροποποίηση του βιοπληροφορικού αγωγού ανάλυσης (pipeline) (Sun, Ruivenkamp et al. 2015).

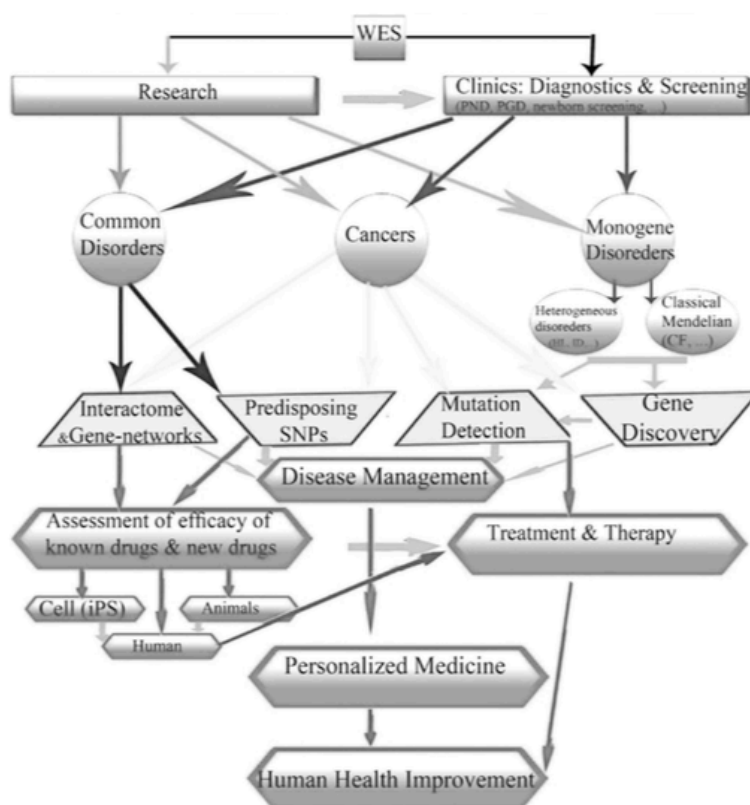
#### **1.4.1) Αλληλούχηση των εξονίων του γονιδιώματος (Whole Exome Sequencing-WES)**

Όπως έχει ήδη αναφερθεί, περίπου  $3 \times 10^7$  ζεύγη βάσεων (1%) του γονιδιώματος είναι οι κωδικοποιούσες περιοχές με το 85% των μεταλλάξεων που προκαλούν ασθένειες να βρίσκονται στις περιοχές αυτές και σε λειτουργικές περιοχές του γονιδιώματος (Majewski, Schwartzenuber et al. 2011). Για το λόγο αυτό ο προσδιορισμός της πλήρους αλληλουχίας των εξονίων του γονιδιώματος δίνει τη δυνατότητα να αποκαλυφθούν τα αίτια μεγάλου αριθμού σπάνιων και ως επί το πλείστον μονογονιδιακών γενετικών διαταραχών καθώς και προσδιορισμό γενετικής προδιάθεσης για καρκίνο (Bonasio, Tu et al. 2010).

Η αλληλούχηση των εξονίων του γονιδιώματος (WES) έχει εξελίξει τη βιοϊατρική έρευνα επεκτείνοντας τα αποτελέσματα και τις εφαρμογές στην κλινική διάγνωση. Ο χαρακτηρισμός γονιδίων που εμπλέκονται σε μια ασθένεια γίνεται με μεγαλύτερη αποτελεσματικότητα χρησιμοποιώντας τεχνολογίες αλληλούχησης επόμενης γενιάς. Όλες οι στρατηγικές αυτές μπορούν να χρησιμοποιηθούν στη διάγνωση ανάλογα με το είδος της ασθένειας και τη διαθέσιμη πληροφορία. Για παράδειγμα, στοχευμένη αλληλούχηση με πάνελ γονιδίων μπορεί να χρησιμοποιηθεί για διάγνωση φυλοσύνδετης νοητικής υστέρησης, αλλά αλληλούχηση ολόκληρου του γονιδιώματος είναι καταλληλότερη για διάγνωση νοητική υστέρηση και συγγενής δυσπλασία που οφείλονται σε δομικές μεταλλαγές (Rabbani, Tekin et al. 2014).

Το WES έχει διάφορους τομείς εφαρμογής στην έρευνα και στην κλινική γενετική, συγκεκριμένα στη διάγνωση περιστατικών στην προγεννητική διάγνωση, στην ανίχνευση φορέων, στην πρόγνωση για προκλινικά περιστατικά (εξατομικευμένη-στοχευμένη θεραπεία, διαχείριση θεραπείας) και στον έλεγχο τύπου screening σε νεογνά. Επίσης η ανακάλυψη γονιδίων για άγνωστες διαταραχές και εύρεση Single Nucleotide Variants (SNVs) προδιάθεσης για κοινά νοσήματα μπορεί να διευκρινίσει τη γενετική βάση και το μοριακό μηχανισμό διάφορων διαταραχών (βιολογικά δίκτυα). Επιπλέον, η

χρήση του WES θα μπορούσε να βελτιώσει γενικώς την υγειονομική περίθαλψη επηρεάζοντας τη διαχείριση της νόσου, την ανακάλυψη νέων φαρμάκων και την εφαρμογή εξατομικευμένης θεραπείας (Εικόνα 5). Τα παραπάνω συστήματα αναλύσεων μπορούν να υποστηρίξουν την εξαγωγή αποτελεσμάτων από Μενδελιανά νοσήματα και ετερογενείς διαταραχές μέχρι πολυγονιδιακές και πολυπαραγοντικές ασθένειες, όπως ο καρκίνος (Rabbani, Tekin et al. 2014).



**Εικόνα 1.5:** Χρήση του WES στην έρευνα και την κλινική διάγνωση. Εφαρμογή στη διάγνωση, πρόγνωση και διαχείριση νόσου. Εύρεση νέων φαρμάκων και εξατομικευμένη θεραπεία. Στόχος των παραπάνω εφαρμογών είναι η βελτίωση της ανθρώπινης υγείας (Rabbani, Tekin et al. 2014).

### 1.5) WES ως διαγνωστικό εργαλείο

Η γενετική διάγνωση ως διαδικασία στην ιατρική υπηρεσία πρέπει να πληρεί τις υψηλότερες προδιαγραφές, γι' αυτό το λόγο απαιτείται η χρήση εξελιγμένων τεχνολογιών. Η αλληλούχηση κατά Sanger είναι σήμερα η μέθοδος αναφοράς (gold standard) για δοκιμασίες βασισμένες στην αλληλούχηση. Ωστόσο, με ένα συνεχώς αυξανόμενο αριθμό γονιδίων που

πρόκειται να ελεγχθούν για κάθε διαταραχή, η μέθοδος Sanger γίνεται λιγότερο αποτελεσματική και πολύ δαπανηρή. Αντίθετα, το NGS επιτρέπει την ταυτόχρονη ανάλυση μεγάλου αριθμού γονιδίων (Panel) ή όλων των εξονίων του γονιδιώματος (WES), ακόμα και ολόκληρου του γονιδιώματος (WGS) σε μια δοκιμασία, σε σύντομο χρονικό διάστημα και με μειωμένο κόστος.

Οι τεχνολογίες αλληλούχησης επόμενης γενιάς έχουν αλλάξει δραματικά το πεδίο της μοριακής διαγνωστικής. Το είδος της ανάλυσης που χρησιμοποιείται ήδη ως διαγνωστικό τεστ είναι η αλληλούχηση συγκεκριμένων γονιδίων που σχετίζονται με μια ασθένεια ή φαινότυπο (πάνελ γονιδίων). Η ισχυρή αυτή τεχνολογία έχει επεκταθεί και στη χρήση του WES (και WGS) για διαγνωστικούς σκοπούς, προσφέροντας αξιόπιστα δεδομένα για το φαινότυπο που μελετάται. Οι διαφορετικές διαγνωστικές εξετάσεις που χρησιμοποιούν NGS έχουν ποικίλη κάλυψη και εκτίμηση των γονιδιακών περιοχών και απαιτείται προσεκτική εξέταση των παραμέτρων κατά την επιλογή της πλέον κατάλληλης δοκιμασίας (Lapin, Mighion et al. 2016), καθώς δεν πρόκειται απλά για μία νέα τεχνολογία, αλλά μια διαδικασία που επηρεάζει το χειρισμό των δειγμάτων, την εξαγωγή και παρακολούθηση των δεδομένων. Επιτρέπει όχι μόνο την αλληλούχηση πολλαπλών γονιδίων σε μια μόνο δοκιμασία, αλλά και το συνδυασμό δειγμάτων DNA διαφορετικών ασθενών έτσι ώστε να αναλυθούν σε ένα πείραμα.

Η χρήση του WES ως διαγνωστικό εργαλείο έχει αλλάξει τη στρατηγική αναλύσεων, από την αλληλούχηση μόνο των σχετιζόμενων με ένα φαινότυπο γονιδίων σε σάρωση όλων των σχετιζόμενων γονιδίων που δυνητικά μπορούν να εξηγήσουν μια συγκεκριμένη διαταραχή. Το σύνολο των εξονίων ενός μόνο ατόμου μπορεί να δίδει περίπου 30.000 διαφορετικές παραλλαγές σε σύγκριση με την αλληλουχία αναφοράς του ανθρώπου, τα οποία στη συνέχεια αναλύονται βιοπληροφοριακά, φιλτράρονται, αξιολογούνται και ερμηνεύονται για να δώσουν τη διάγνωση (Gilissen, Hoischen et al. 2012).

Μέσω του WES μπορεί κανείς να αποκαλύψει νέες παραλλαγές που δεν είναι γνωστό ότι σχετίζονται με κάποια νόσο ή προκαλούν εξαιρετικά σπάνιες ασθένειες. Γι' αυτό το λόγο, αποτελεί σήμερα σημαντική πρόκληση η ανάλυση

και ταξινόμηση τέτοιων παραλλαγών που δεν έχουν αναφερθεί προηγουμένως στη βιβλιογραφία. Η εκτίμηση της παθογένειας μιας παραλλαγής πρέπει να ακολουθεί αποδεκτές κατευθυντήριες οδηγίες για την ταξινόμηση του, όμως όταν τα υπάρχον στοιχεία για τη φύση της παθογένειας είναι ανεπαρκή, τότε παραμένει ως μια παραλλαγή αγνώστου σημασίας (Variant of Unknown Significance - VUS). Ο μεγάλος αριθμός των αγνώστου σημασίας παραλλαγών οδηγεί συνήθως σε αβέβαιο αποτέλεσμα της ζητούμενης διαγνωστικής δοκιμασίας, το οποίο απαιτεί περαιτέρω διερεύνηση (Richards, Aziz et al. 2015).

Ένα ακόμη διαγνωστικό όφελος από τη χρήση WES, είναι η ταυτοποίηση παθογόνων παραλλαγών για ασθένειες οι οποίες δεν έχουν σχέση με το κύριο σκοπό της εξέτασης που ζητήθηκε (δευτερεύοντα ευρήματα). Τα ευρήματα αυτά μπορεί να είναι για κάποια όψιμης έναρξης θεραπεύσιμης ή μη ασθένειας, προδιάθεση για καρκίνο ή για ασθένεια με χαμηλή διεισδυτικότητα (Green, 2013). Επίσης μπορεί να προσδιοριστεί η ύπαρξη φορείας για υπολειπόμενη ασθένεια, αλλά και φαρμακογενετικές παραλλαγές που μπορεί να επηρεάσουν την ανταπόκριση ενός ατόμου σε κάποιο φαρμακευτικό παράγοντα. Σημαντική παράμετρος για την εφαρμογή τέτοιου είδους αναλύσεων είναι οι συμμετέχοντες ενήλικες, οι γονείς ή κηδεμόνες ανήλικων παιδιών, αλλά και όποιος άλλος υποβληθεί σε τέτοια εξέταση να ενημερωθούν σχετικά με τα αναμενόμενα αποτελέσματα, την πιθανότητα και το είδος των τυχαίων ευρημάτων, ποια αποτελέσματα μπορούν να ανακοινωθούν και ποια όχι (Green, 2013).

Η απόκτηση υψηλής ποιότητας διάγνωσης δεν είναι πλέον στην αποκλειστική αρμοδιότητα των κλινικών, αλλά απαιτεί εντατική αλληλεπίδραση μεταξύ εργαστηριακών ειδικών, κλινικών γενετιστών, βιοπληροφορικών και άλλων ιατρικών ειδικοτήτων. Ειδικά ο τομέας της βιοπληροφορικής, είναι ζωτικής σημασίας με τη δημιουργία κατάλληλων αγωγών ανάλυσης (pipeline) και διασφάλισης ποιότητας μέσα από τα υπολογιστικά εργαλεία που χρησιμοποιούνται (Delaney, Hultner et al. 2016).

## **Κεφάλαιο 2:**

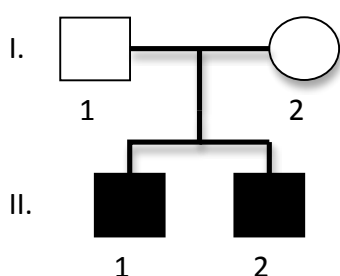
### ***Υλικά και Μέθοδοι***

## 2.1) Εφαρμογή της Αλληλούχησης Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES)

### 2.1.1) Whole Exome Sequencing (WES) σε δείγματα ασθενών με Νευρομυϊκό Νόσημα (NMDs)

Η εφαρμογή του Whole Exome Sequencing (WES) πραγματοποιήθηκε σε 6 δείγματα (3 μεμονωμένα δείγματα και 1 trio ανάλυση) με φαινότυπο νευρομυϊκού νοσήματος. Τα Νευρομυϊκά Νοσήματα (NMDs) είναι νοσήματα, τα οποία επηρεάζουν το περιφερικό νευρικό σύστημα και τους μύες. Μπορούν να οδηγήσουν σε απώλεια ελέγχου των μυών ή της λειτουργίας τους ή σε ατροφία μυών. Η ηλικία έναρξης των συμπτωμάτων κυμαίνεται από την προγεννητική έως αργότερα στην ενήλικη ζωή, με την πλειονότητα να αρχίζει κατά τη βρεφική ή τη παιδική ηλικία. Τα βιοχημικά δεδομένα που υποστηρίζουν τα νευρομυϊκά νοσήματα είναι οι μετρήσεις των επιπέδων CPK (κρεατινική φωσφοκινάση), η οποία καταλύει τη μετατροπή της κρεατίνης χρησιμοποιώντας το ενεργειακό νόμισμα του κυττάρου, το ATP, σε φωσφοκρεατίνη και ADP. Κυρίως εμφανίζεται σε ιστούς με μεγάλη κατανάλωση ATP, όπως οι σκελετικοί μύες και το μυοκάρδιο.

Στην παρούσα μελέτη εξετάζεται μόνο η trio ανάλυση με το πάσχον παιδί και τους δύο βιολογικούς γονείς του. Το πάσχον παιδί που εξετάζεται είναι αγόρι ηλικίας 7 ετών (II-1), ασυμπτωματικό, με τιμές CPK 3.500-9.000 U/L. Επίσης, έχει ένα αδερφό ηλικίας 10 ετών (II-2) με την ίδια κλινική εικόνα. Οι γονείς τους (I-1, I-2) είναι επίσης ασυμπτωματικοί και με φυσιολογικές τιμές CPK. Οι φυσιολογικές τιμές CPK είναι για τους άνδρες 80 U/L και για τις γυναίκες 70 U/L.



**Εικόνα 2.1:** Γενεαλογικό δέντρο trio ανάλυσης της οικογένειας που μελετήθηκε.



### 2.1.2) Τεχνική επισκόπηση του Whole Exome Sequencing (WES)

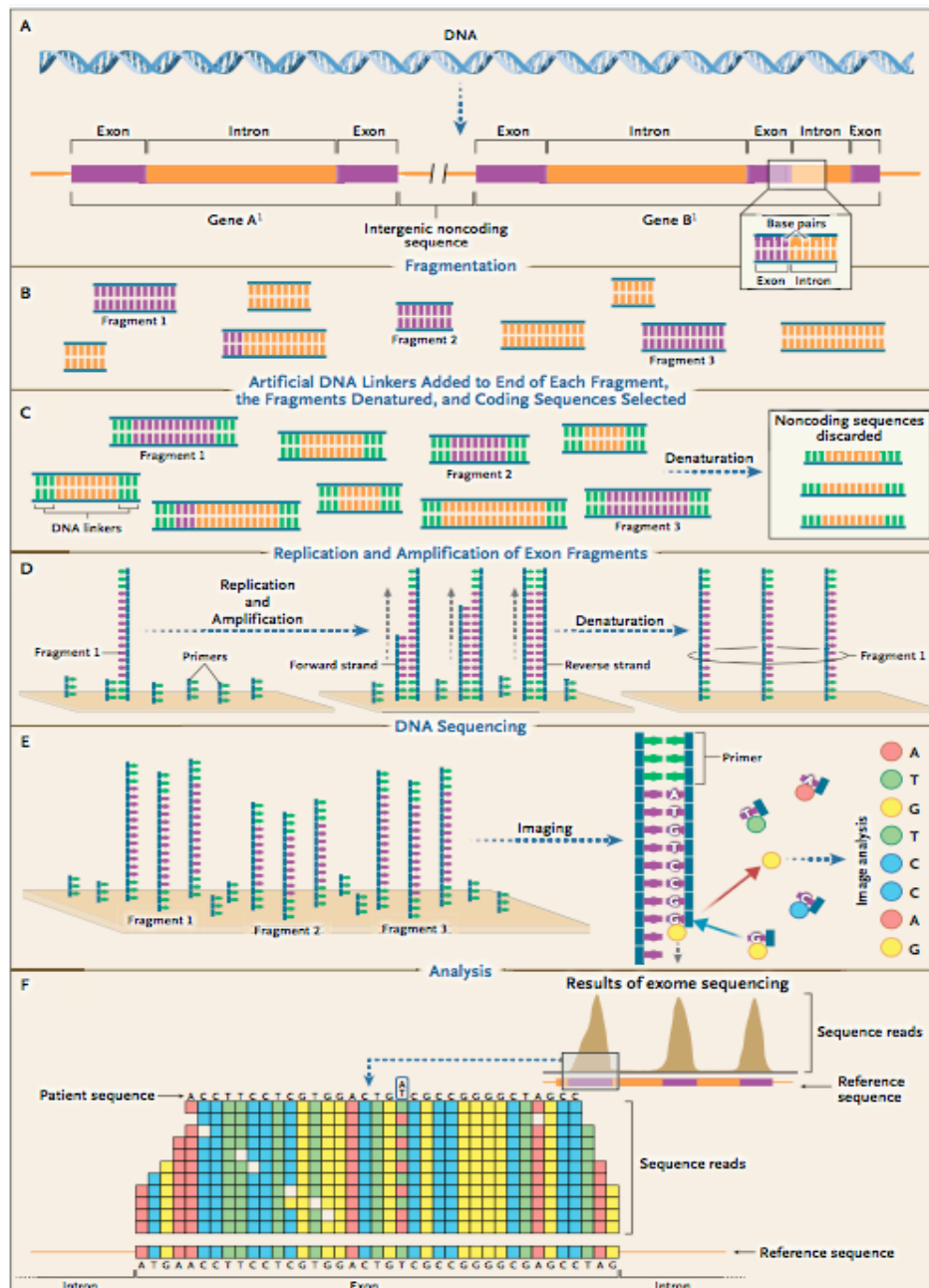
Η διαδικασία της ανάλυσης των εξονίων του γονιδιώματος με αλληλούχηση νέας γενιάς ξεκινάει με την εκχύλιση του DNA των προς εξέταση ατόμων από δείγμα αίματος, σάλιου, ιστού ή άλλου βιολογικού υλικού. Το DNA τεμαχίζεται σε μικρά θραύσματα και αλληλουχείται χρησιμοποιώντας κάποια από τις διαθέσιμες τεχνολογίες. Έτσι, παράγονται εκατομμύρια μικρές αλληλουχίες (reads), οι οποίες αντιπροσωπεύουν τη σειρά των νουκλεοτιδίων σε κάθε θραύσμα. Στη συνέχεια, αυτά τα reads στοιχίζονται με την ανθρώπινη αλληλουχία αναφοράς χρησιμοποιώντας υπολογιστικά συστήματα. Ομοιότητες και διαφορές μεταξύ των δύο αλληλουχιών κατατάσσονται σε πίνακες και προσδιορίζεται ο συγκεκριμένος γονότυπος για κάθε γενετική θέση. Επιπλέον εξάγονται δεδομένα σχετικά με τις ποιοτικές παραμέτρους για κάθε νουκλεοτιδική θέση, όπως η ακρίβεια και το βάθος κάλυψης. Το εξαχθέν αρχείο με τα κλινικά δεδομένα φιλτράρεται για τις διάφορες παραλλαγές, τα οποία προβλέπεται να ευθύνονται για την εμφάνιση ενός συγκεκριμένου φαινοτύπου ή είναι σπάνια και δεν έχουν αναφερθεί προηγουμένως στη βιβλιογραφία (Εικόνα 2.2) (Stein 2011, Duzkale, Shen et al. 2013).

Η μελέτη των προς εξέταση δειγμάτων πραγματοποιήθηκε με δύο διαφορετικά εργαστηριακά πρωτόκολλα (Illumina TruSeq και Sophia Genetics®) και δύο διαφορετικά βιοπληροφορικά συστήματα ανάλυσης (VarAFT και Sophia DDM®). Πιο αναλυτικά, τα παραπάνω συστήματα συνδυάστηκαν και προέκυψαν τρεις ροές εργασίας ('Workflows').

A) Η πρώτη ροή εργασίας περιλαμβάνει προετοιμασία βιβλιοθήκης με TruSeq Exome Library Prep Kit της Illumina (Εδάφιο 2.5), αλληλούχηση με σύστημα NextSeq 500 της Illumina (Εδάφιο 2.6) και βιοπληροφορική ανάλυση με υπολογιστικά συστήματα που διατίθενται στον ακαδημαϊκό χώρο (Bowtie2, Samtools, VarAFT), σύμφωνα με το διάγραμμα της Εικόνας 2.12 (Εδάφιο 2.7).

B) Η δεύτερη ροή εργασίας βιοπληροφορική ανάλυση των δεδομένων (FastQ αρχεία) της πρώτης ροή εργασίας με το αυτοματοποιημένο σύστημα Sophia DDM® της Sophia Genetics®.

Γ) Η τρίτη ροή εργασίας περιλαμβάνει προετοιμασία της βιβλιοθήκης, αλληλούχηση και βιοπληροφορική ανάλυση από την εταιρεία Sophia Genetics®.



**Εικόνα 2.2:** Τεχνική επισκόπηση του WES

Το DNA απομονώνεται από βιολογικά δείγματα (Πάνελ A) και τεμαχίζεται σε θραύσματα (Πάνελ B). Στα άκρα των θραυσμάτων συνδέονται τεχνητά κομμάτια DNA (DNA linkers) (Πάνελ C) και υβριδοποιούνται με ειδικές αλληλουχίες σε στερεή επιφάνεια με στόχο τον εμπλουτισμό των περιοχών-στόχων (Πάνελ D). Αλληλούχηση των θραυσμάτων με σύνθεση αλυσίδας τοποθετώντας ένα από τα τέσσερα κάθε φορά νουκλεοτίδια απέναντι από το συμπληρωματικό του. Το κάθε νουκλεοτίδιο που τοποθετείται είναι σημασμένο με διαφορετικό φθορίζον μόριο (Πάνελ E). Η ταυτοποίηση του έγχρωμου προϊόντος από το φθορίζον μόριο για κάθε σύμπλεγμα γίνεται με τη χρήση λέιζερ και κάμερας συνδεδεμένα με υπολογιστικό σύστημα. Αυτό επαναλαμβάνεται για κάθε θραύσμα και προκύπτει η νουκλεοτιδική αλληλουχία, η οποία στοιχίζεται με την αλληλουχία αναφοράς (Πάνελ F). Στη συγκεκριμένη εικόνα οι περισσότερες θέσεις είναι ομόλογες με την αλληλουχία αναφοράς, εκτός από μία βάση η οποία εμφανίζεται σε ετεροζυγωτία (A/T) (Biesecker and Green 2014).

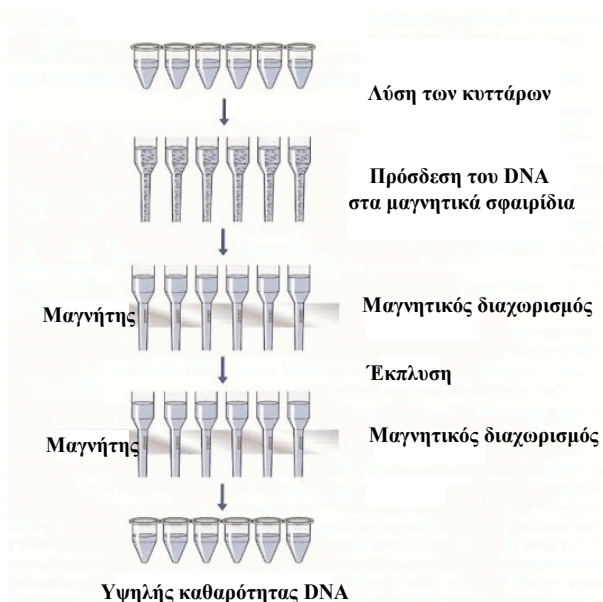
## 2.2) Απομόνωση γενωμικού DNA από λεμφοκύτταρα περιφερικού αίματος με χρήση του πρωτόκολλο MagAttract DNA Blood M48 Kit και ρομποτικού συστήματος BioRobot M48 της QIAGEN

Το DNA που χρησιμοποιήθηκε απομονώθηκε από λεμφοκύτταρα του περιφερικού αίματος των ασθενών. Το αίμα συλλέχθηκε σε φιαλίδια όγκου 3mL με χρήση EDTA ως αντιπηκτικού, ενός χηλικού παράγοντα που επηρεάζει τη λειτουργία του ινοδωγόνου και κατά συνέπεια την πήξη του αίματος. Επιλέχθηκε EDTA, διότι σε αντίθεση με άλλα αντιπηκτικά (π.χ. ηπαρίνη) δεν παρεμποδίζει την περαιτέρω χρήση του DNA που απομονώνεται σε μοριακές τεχνικές, όπως η PCR. Η απομόνωση του DNA έγινε με τη χρήση του BioRobot M48, σύμφωνα με το πρωτόκολλο MagAttract DNA Blood M48 Kit της εταιρείας QIAGEN.

Το MagAttract DNA Blood M48 Kit επιτρέπει την απόλυτα αυτοματοποιημένη απομόνωση ολικού DNA από το περιφερικό αίμα, το οποίο μπορεί να χρησιμοποιηθεί απευθείας σε περαιτέρω μοριακές εφαρμογές και ενζυμικές αντιδράσεις. Η τεχνολογία MagAttract συνδυάζει την ταχύτητα και την ακρίβεια της απομόνωσης που βασίζεται στη χρήση επιφανειών πυριτίου, με την ευκολία στον χειρισμό των μαγνητικών σωματιδίων. Το DNA προσδένεται στην πυριτική επιφάνεια των μαγνητικών σφαιριδίων, παρουσία ενός χαοτροπικού άλατος. Στη συνέχεια το προσδεμένο DNA εκπλένεται με δύο ρυθμιστικά διαλύματα και με απιονισμένο νερό, το οποίο βελτιώνει την καθαρότητα του DNA. Τελικά, DNA υψηλής ποιότητας εκπλένεται σε νερό

που παρέχεται από την εταιρία. Το DNA που αποδίδει η τεχνική εξαρτάται από τον τύπο του δείγματος, τις συνθήκες φύλαξής του και την περιεκτικότητα του αίματος σε λευκοκύτταρα. Στην Εικόνα 2.3 φαίνονται τα στάδια της ροής εργασίας MagAttract DNA Blood M48.

**Εικόνα 2.3:** Η διαδικασία απομόνωσης DNA στο MagAttract Blood M48.



### 2.3) Ποσοτικός προσδιορισμός του DNA

Ο ποσοτικός προσδιορισμός του DNA έγινε φθορισμομετρικά από τον αναλυτή Qubit Fluorimeter της εταιρείας Invitrogen (ΗΠΑ), χρησιμοποιώντας το Quant-iT™ dsDNA HS Assay Kit (Εικόνα 2.4). Το Kit αυτό μπορεί να μετρήσει ποσοτικά γενωμικό DNA, ιϊκό dsDNA καθώς και PCR προϊόντα. Η ποσότητα του δείγματος που μπορεί να χρησιμοποιηθεί είναι από 1 έως 20μl, ενώ τα όριο ανίχνευσης είναι 10 pg/μl. Η πειραματική διαδικασία μπορεί να γίνει σε θερμοκρασία δωματίου (22-28 °C) ενώ η φωτοσταθερότητα των αντιδραστηρίων είναι υψηλή. Στην παρούσα μελέτη χρησιμοποιήθηκαν 1μl δείγματος σε σωληνάρια PCR των 0.5 ml, με καινούρια βαθμονόμηση για κάθε χρήση του. Ο προσδιορισμός της συγκέντρωσης του DNA γίνεται με βάση την καμπύλη αλγορίθμου χρησιμοποιώντας πρότυπα διαλύματα συγκέντρωσης 0ng/ml και 100ng/ml. Τέλος το Qubit® fluorometer δίνει τις τιμές των συγκεντρώσεων είτε σε μg/ml είτε σε ng/ml. Για τον υπολογισμό της συγκέντρωσης χρησιμοποιείται η σχέση :

$$C = QF \text{ value} \times (200) / n$$

Όπου QF value= η τιμή που δίνεται στον αναλυτή, n = ο αριθμός των μl του δείγματος που χρησιμοποιήθηκαν.



**Εικόνα 2.4:** Qubit® Φθορισμόμετρο.

#### **Υλικά (αντιδραστήρια, εξοπλισμός, αναλώσιμα)**

- Πιπέτες (Gilson) των 10 μl και 200 μl
- Ακροφύσια για πιπέτες
- Σωληνάρια τύπου Eppendorf των 500 μl
- Αντιδραστήρια ds DNA BR βαθμονομητής 1 (0 ng/μl), ds DNA BR βαθμονομητής 2 (100 ng/ μl), ds ρυθμιστικό διάλυμα DNA BR και ds

DNA BR αντιδραστήριο (παρέχονται από το Quant-iT ds DNA BR Assay Kit, Q32850)

- Συσκευή Vortex

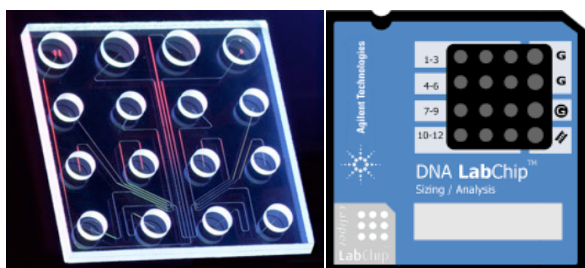
### **Πειραματική διαδικασία**

Αρχικά όλα τα διαλύματα έρχονται σε θερμοκρασία δωματίου.

Προστίθενται 190 x n ( n= αριθμός δειγμάτων) μl αντιδραστηρίου ds DNA BR και 10 μL αντιδραστηρίου ds DNA BR αντιδραστηρίου σε σωληνάριο erpendorf. Από το παραπάνω διάλυμα λαμβάνονται 199μl για κάθε δείγμα, τοποθετούνται σε νέο σωληνάριο erpendorfs και προστίθεται 1μl δείγματος DNA. Ακολουθεί ανάδευση του διαλύματος, τοποθέτηση στο φθορισμόμετρο και καταγραφή των ενδείξεων.

### **2.4) Ποιοτικός έλεγχος του DNA με Bioanalyzer 2100 της Agilent**

Το Bioanalyzer 2100 της Agilent είναι όργανο ανάλυσης υψηλής ακρίβειας της ποιότητας του DNA ή του RNA. Το μηχάνημα λειτουργεί με μικρές πλακέτες (τσιπ), ειδικά διαμορφωμένες για το υλικό που θα εξεταστεί και βασίζεται στην αρχή της τριχοειδούς ηλεκτροφόρησης. Οι πλακέτες έχουν την ικανότητα να μειώνουν το χρόνο διαχωρισμού και την κατανάλωση δείγματος και αντιδραστηρίων με δραματικό τρόπο. Η πλακέτα διαθέτει πηγαδάκια για το δείγμα, για το πήκτωμα και ένα πηγαδάκι για τον δείκτη (ladder). Μέσα στο γυαλί της πλακέτας υπάρχουν μικροκανάλια που δημιουργούν ένα εκτεταμένο, διασυνδεδεμένο δίκτυο μεταξύ των πηγαδιών (Εικόνα 2.5).



**Εικόνα 2.5:** Άποψη εσωτερικού μέρους πλακέτας ηλεκτροφόρησης σε μικροκανάλια (αριστερά) και μακροσκοπική άποψη της πλακέτας (δεξιά).

Κατά την ετοιμασία της πλακέτας, τα μικροκανάλια γεμίζουν με το πορώδες, πολυμερές πήκτωμα και μια φθορίζουσα χρωστική. Μόλις γεμίσουν τα

πηγάδια και τα μικροκανάλια, η πλακέτα γίνεται ένα κλειστό ηλεκτρικό κύκλωμα. Τα δεκαέξι ηλεκτρόδια του φυσιγγίου είναι τοποθετημένα ώστε να προσαρμόζονται μέσα στα δεκαέξι πηγάδια της πλακέτας. Κάθε ηλεκτρόδιο έχει ξεχωριστή παροχή ηλεκτρικής ισχύος. Τα διοχετευόμενα στα πηγάδια βιομόρια (DNA ή RNA) οδηγούνται ηλεκτροφορητικά από μια διαβάθμιση τάσης, όπως στην κοινή ηλεκτροφόρηση. Επέρχεται, λοιπόν, ο διαχωρισμός των μορίων από το πορώδες υλικό, με κριτήριο το μέγεθος. Τα μικρά μόρια μετακινούνται ταχύτερα, εντός του υλικού, από τα μεγαλύτερα. Τα μόρια της χρωστικής αλληλεπιδρούν με τα βιομόρια και τα σύμπλοκα που προκύπτουν, ανιχνεύονται με επαγόμενο από ακτίνα λέιζερ φθορισμό.

Τα δεδομένα μετατρέπονται σε κλασσικές εικόνες ηλεκτροφόρησης (μπάντες) και διαγράμματα. Με τη βοήθεια του δείκτη (ladder) που περιέχει στοιχεία γνωστού μεγέθους, προκύπτει καμπύλη αναφοράς του χρόνου μετακίνησης προς το μέγεθος των στοιχείων. Από το χρόνο μετακίνησης υπολογίζεται το μέγεθος κάθε στοιχείου. Η ποσοτικοποίηση γίνεται με τη βοήθεια της περιοχής που ορίζει ο δείκτης στο διάγραμμα (ladder). Η περιοχή αυτή συγκρίνεται με το άθροισμα των περιοχών που ορίζουν οι κορυφές του δείγματος. Η περιοχή που ορίζεται από πολύ χαμηλές κορυφές, δε λαμβάνεται υπ' όψιν στους υπολογισμούς.

#### ***Υλικά (αντιδραστήρια, εξοπλισμός, αναλώσιμα)***

- Πιπέτες (Gilson) των 2 μl, 10 μl και 200 μl
- Ακροφύσια για πιπέτες
- Σωληνάρια τύπου Eppendorf των 500 μl
- Agilent DNA 7500 assay kit
- Agilent Bioanalyzer 2100
- Συσκευή Vortex
- Φυγόκεντρος τύπου Eppendorf
- Ηλεκτρονικός Υπολογιστής

### **Πειραματική διαδικασία**

Αρχικά όλα τα διαλύματα έρχονται σε θερμοκρασία δωματίου.

- Προετοιμασία μίγματος πήκτωμα-χρωστική:

Ανάδευση με Vortex της DNA χρωστικής και προσθήκη 25μl στο DNA πήκτωμα. Φιλτράρισμα του μίγματος πήκτωμα-χρωστική σε ειδική στήλη με φυγοκέντρηση στα 1500g για 10min. (Το μίγμα είναι φωτοευαίσθητο, αποθηκεύεται στους 4°C)

- Φόρτωση μίγματος πήκτωμα-χρωστική:

Τοποθέτηση ενός DNA chip στη συσκευή εκφόρτωσης με το έμβολο και προσθήκη 9μl από το μίγμα πήκτωμα-χρωστική στο πηγάδι με την ένδειξη **G**. Τοποθέτηση του εμβόλου στην ένδειξη για 1ml και κατέβασμα του εμβόλου για 1min. Άνοιγμα του εμβόλου και προσθήκη 9μl από το μίγμα πήκτωμα-χρωστική στα πηγάδια με την ένδειξη G.

- Φόρτωση των μαρτύρων (markers):

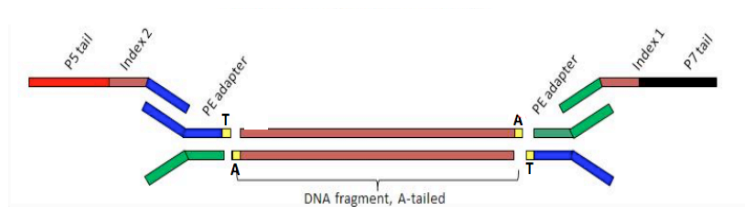
Προσθήκη 5μl από το διάλυμα μαρτύρων σε κάθε πηγάδι 1-12 και στο πηγάδι με την ένδειξη /=/.

- Φόρτωση του DNA ladder και των δειγμάτων:

Προσθήκη 1μl από το DNA ladder στο πηγάδι με την ένδειξη /=/. Προσθήκη 1μl από κάθε δείγμα στα πηγάδια 1-12 (ή προσθήκη dH<sub>2</sub>O στα κενά πηγάδια). Στη συνέχεια, ανάδευση του chip με Vortex για 1min, τοποθέτησή του στον Bioanalyzer 2100 για ανάλυση και καταγραφή των αποτελεσμάτων.

### **2.5) Προετοιμασία βιβλιοθήκης με χρήση του TruSeq Exome Library Prep Kit της Illumina**

Το TruSeq Exome Library Prep Kit προσφέρει μια απλουστευμένη οικονομικά αποδοτική μέθοδο για την αναγνώριση και κατανόηση των κωδικών παραλλαγών με εξαιρετική ακρίβεια. Η εξειδίκευση του kit αποτελεί μια ισχυρή μέθοδο για την ανακάλυψη δυνητικών παραλλαγών που προκαλούν γενετικές ασθένειες. Ως μέρος μιας ολοκληρωμένης ροής εργασίας, το kit περιλαμβάνει προετοιμασία βιβλιοθήκης, εμπλουτισμό εξονίων και ανάλυση αλληλουχίας [Εικόνα 2.2 (A-C)]. Η χρήση του TruSeq Exome Library Prep Kit είναι συμβατή με τα συστήματα αλληλούχησης MiSeq, NextSeq, HiSeq της Illumina. Αυτά τα συστήματα χρησιμοποιούν χημεία για αλληλούχηση μέσω σύνθεσης, την πιο ευρέως υιοθετούμενη τεχνολογία, η οποία αποδίδει το υψηλότερο ποσοστό



<https://www.fimm.fi/en/services/technology-centre/sequencing/next-generation-sequencing/dna-library-preparation>

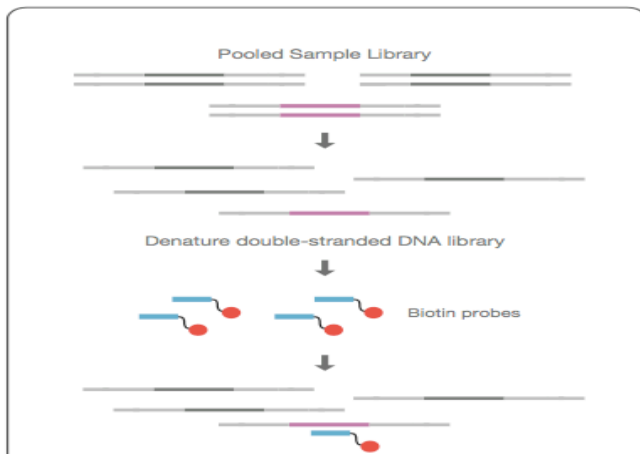
**Εικόνα 2.6:** Θραύσμα DNA με προσαρμοσμένους αντάπτορες και εκκινήτες αλληλούχησης και PCR.

βάσεων στην αλληλουχία με βαθμολογία Q30 (1 λάθος στις 1000 βάσεις, ακρίβεια 99,9%) ελαχιστοποιώντας τα ψευδώς θετικά και ψευδώς αρνητικά αποτελέσματα. Το kit έχει βελτιωθεί ώστε να παρέχει ομοιόμορφη κάλυψη για 45Mb εξονικού περιεχομένου και να εμπλουτίζει 214.405 εξόνια. Το kit υποστηρίζει την ανάμειξη έως 12 διαφορετικών δειγμάτων σε μια αντίδραση αλληλούχησης, επιτρέποντας τον εντοπισμό των παραλλαγών σε σύντομο χρονικό διάστημα και μεγαλύτερη οικονομία.

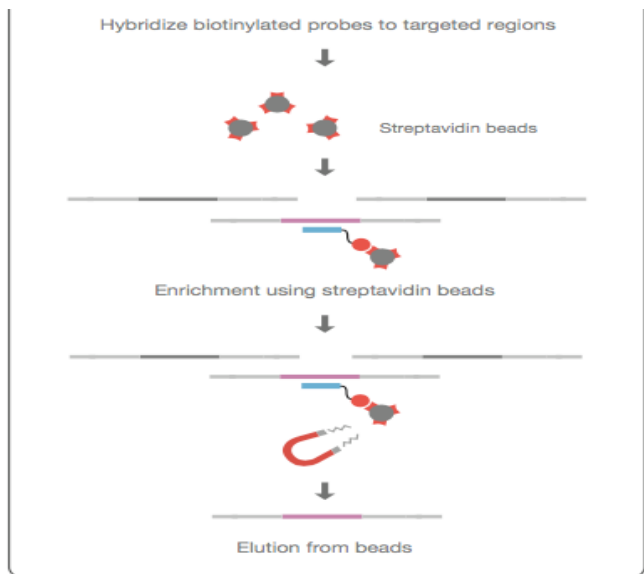
Το TruSeq Exome Library Prep Kit ξεκινάει με την προετοιμασία της βιβλιοθήκης αρχίζοντας από 100ng DNA, το οποίο έχει κατακερματιστεί δημιουργώντας ομοιόμορφα μεγέθη θραυσμάτων. Η μηχανική κατάτμηση γίνεται με χρήση υπερηχογραφικού μηχανήματος Covaris ή με άλλες παρόμοιες μεθόδους. Τα θραύσματα που παράγονται έχουν μικρές προεξοχές στα άκρα τους, τα οποία επιδιορθώνονται με χρήση πολυμεράσης ή εξωνουκλεάσης. Ακολουθεί επιλογή βάση μεγέθους με παρεχόμενα σφαιρίδια αντιστρεπτής ακινητοποίησης AMPure (SPRI) (BeckmanCoulter). Στη συνέχεια, πραγματοποιείται προσθήκη μιας A (Αδεΐνης) στα τυφλά 3' άκρα κάθε κλώνου για να μπορέσουν να υβριδοποιηθούν οι ειδικοί αντάπτορες, οι οποίοι διαθέτουν μονόκλωνο άκρο με T (Θυμίνη). Αυτοί οι αντάπτορες αποτελούνται από συμπληρωματικές αλληλουχίες που υβριδίζονται με τους εκκινήτες αλληλούχησης και εκκινήτες PCR (Εικόνα 2.6). Οι αντάπτορες-προσαρμοστές προσδένονται στα θραύσματα και τα προϊόντα που προκύπτουν ενισχύονται με PCR. Οι βιβλιοθήκες κάθε δείγματος αναμειγνύονται σε ένα σωληνάριο και βιοτινυλιωμένοι ανιχνευτές υβριδοποιούνται στις περιοχές στόχους. Χρησιμοποιώντας σφαιρίδια στρεπταβιδίνης γίνεται επιλογή των βιοτινυλιωμένων περιοχών και



εμπλουτισμός με PCR. Τα θραύσματα εκλούνται από τα σφαιρίδια και είναι έτοιμα για αλληλούχηση (Εικόνα 2.7). Αυτή η πορεία εργασίας παράγει περιοχές στόχους μεγέθους 150bp και πραγματοποιείται σε περίπου 2,5 ημέρες.



**Εικόνα 2.7:** Απεικόνιση της υβριδοποίησης των βιοτινυλιωμένων ανιχνευτών στις περιοχές στόχους και χρήση μαγνητικών σφαιριδίων στρεπταβιδίνης για επιλογή των βιοτινυλιωμένων περιοχών και έκλουση αυτών.



[www.illumina.com/technology/next-generation-sequencing.html](http://www.illumina.com/technology/next-generation-sequencing.html)

Η ροή εργασίας του TruSeq Exome Library Prep Kit περιλαμβάνει 12 ξεχωριστές διαδικασίες, οι οποίες χωρίζονται σε 3 ημέρες με ασφαλή σημεία παύσης και ποιοτικούς ελέγχους της βιβλιοθήκης, όπως φαίνεται στην Εικόνα 2.8.



**Εικόνα 2.8:** Ροή εργασίας για προετοιμασία βιβλιοθήκης με χρήση του TruSeq Exome Library Prep Kit.

## Υλικά (αντιδραστήρια, εξοπλισμός, αναλώσιμα)

**Πίνακας 2.1:** Αντιδραστήρια TruSeq Exome Library Prep Kit

Αντιδραστήριο	Περιγραφή	Συνθήκες Αποθήκευσης	Συνθήκες Χρήσης
<b>ATL2</b>	A Tailing Mix	-25°C έως -15°C	θ°C Δωματίου
<b>LIG2</b>	Ligation Mix 2	-25°C έως -15°C	Στον πάγο
<b>STL</b>	Stop Ligation Buffer	-25°C έως -15°C	θ°C Δωματίου
<b>PPC</b>	PCR Primer Cocktail	-25°C έως -15°C	θ°C Δωματίου
<b>EPM</b>	Enhanced PCR Mix	-25°C έως -15°C	Στον πάγο
<b>CT3</b>	Capture Target Buffer 3	-25°C έως -15°C	θ°C Δωματίου
<b>EE1</b>	Enrichment Elution Buffer	-25°C έως -15°C	θ°C Δωματίου
<b>HP3</b>	2N NaOH	-25°C έως -15°C	θ°C Δωματίου
<b>NEM</b>	Enrichment Amplification Mix	-25°C έως -15°C	Στον πάγο
<b>RSB</b>	Resuspension Buffer	2°C έως 8°C	θ°C Δωματίου
<b>SWS</b>	Streptavidin Wash Solution	-25°C έως -15°C	θ°C Δωματίου
<b>ERP3</b>	End Repair Mix	-25°C έως -15°C	θ°C Δωματίου
<b>EDTA</b>	EDTA 0.5M	-25°C έως -15°C	θ°C Δωματίου
<b>ET2</b>	Elute Target Buffer 2	2°C έως 8°C	θ°C Δωματίου
<b>SPB</b>	Sample Purification Beads	2°C έως 8°C	θ°C Δωματίου
<b>SMB</b>	Streptavidin Magnetic Beads	2°C έως 8°C	θ°C Δωματίου
<b>CEX</b>	Coding Exome Oligos	-25°C έως -15°C	θ°C Δωματίου

**Πίνακας 2.2:** Αντάπτορες για 8x9plex με τις αλληλουχίες τους

Αντάπτορας	Αλληλουχία
<b>AD002</b>	CGATGT(A)
<b>AD004</b>	TGACCA(A)
<b>AD005</b>	ACAGTG(A)
<b>AD006</b>	GCCAAT(A)
<b>AD007</b>	CAGATC(A)
<b>AD012</b>	CTTGTA(A)
<b>AD013</b>	AGTCAA(C)
<b>AD014</b>	AGTTCC(G)
<b>AD015</b>	ATGTCA(G)

Όλα τα αντιδραστήρια μετά τη χρήση επιστρέφονται στις συνθήκες αποθήκευσης. Το SPB χρειάζεται καλή ανάδευση (Vortex) πριν από κάθε χρήση και αργό πιπετάρισμα λόγω του ιξώδους του διαλύματος. Το LIG2 χρησιμοποιείται άμεσα και επιστρέφεται στους -20°C. Το CT3 χρειάζεται καλή ανάδευση (Vortex) για να επαναιωρηθεί το παγωμένο διάλυμα έως ότου γίνει διαυγές. Το SMB χρειάζεται ήπια ανάδευση πριν τη χρήση. Η αιθανόλη 80% πρέπει να είναι πρόσφατα παρασκευασμένη κάθε φορά. Το PPC και το EPM δεν πρέπει να αναδεύονται με Vortex. Οι DNA αντάπτορες-προσαρμοστές χρησιμοποιούνται σε θερμοκρασία δωματίου για 10min.

- Θερμοκυκλοποιητής με θερμαινόμενο καπάκι
- Μικροφυγόκεντρος
- DynaMag-2 Magnet
- Vortex
- Covaris σύστημα S220
- Σωληνάρια microTUBE AFA Fiber Snap-Cap για Covaris
- Heat Block
- Πιπέτες (Gilson) των 2μl, 10μl, 20μl, 200μl και 1000μl
- Ακροφύσια για πιπέτες
- Σωληνάρια τύπου Eppendorf των 500μl
- Σωληνάρια τύπου Eppendorf των 1,5ml
- Σωληνάρια τύπου Falcon των 15ml
- Αιθανόλη 80%
- dH<sub>2</sub>O

### ***Πειραματική διαδικασία***

#### **1. Κατάτμηση του gDNA**

Αρχικά γίνεται ποσοτικοποίηση του gDNA χρησιμοποιώντας φθορισμομετρική μέθοδο (Εδάφιο 2.2) και παρασκευή ρυθμιστικού διαλύματος σε σωληνάριο 15ml με 5ml RSB και 10ml EDTA. Σε 50μl τελικού όγκου ρυθμιστικού διαλύματος προστίθενται 100ng gDNA, αναδεύονται με πιπέτα (pipette up and down) και φυγοκεντρούνται σύντομα (spin down). Για την θραυσματοποίηση μεταφέρονται τα 50μl στα ειδικά σωληνάρια microTUBE AFA Fiber Snap-Cap. Η θραυσματοποίηση του DNA γίνεται σε σύστημα υπερήχων Covaris S220, το οποίο έχει Duty factor 10%, Peak Power (W) 175, Cycles 200, Duration 280sec, Θερμοκρασία 7 °C, Water level 12. Μετά από σύντομη φυγοκέντρηση, μεταφέρονται 50 μl από τα ειδικά σωληνάρια του Covaris σε σωληνάρια των 1,5ml για να γίνει καθαρισμός των θραυσμάτων DNA. Για τον καθαρισμό προστίθενται 100μl από SPB σε κάθε σωληνάριο, αναδεύονται καλά με πιπέτα και επωάζονται για 5min σε θερμοκρασία δωματίου. Ακολουθεί σύντομη φυγοκέντρηση, τοποθέτηση σε θήκη σωληναρίων με μαγνήτη για ~8min μέχρι το δείγμα να γίνει διαυγές και απόρριψη του υπερκείμενου. Στη συνέχεια, πραγματοποιείται έκπλυση του κάθε δείγματος 2

φορές με προσθήκη 200μl αιθανόλης 80%, επώαση στη μαγνητική θήκη για 30sec και απόρριψη του υπερκλειμένου. Γίνεται σύντομη φυγοκέντρηση και χρήση πιπέτας 20μl για απομάκρυνση της περίσσειας αιθανόλης και αναμονή στη μαγνητική θήκη για 5min με ανοικτό καπάκι για εξάτμιση της εναπομείναντας αιθανόλης. Τέλος, προστίθενται σε κάθε δείγμα 62,5μl RSB, αναδεύονται με πιπέτα, επωάζονται για 2min σε θερμοκρασία δωματίου, φυγοκεντρώνονται σύντομα και τοποθετούνται στη μαγνητική θήκη για 2-5min έως ότου να γίνει διαυγές. Από κάθε δείγμα μεταφέρονται 60μl υπερκλειμένου σε νέο σωληνάριο 0,5ml.

## 2. Επιδιόρθωση των 3' άκρων και επιλογή βάσει μεγέθους

Αρχικά γίνεται προσθήκη 40μl ERP3 σε κάθε σωληνάριο, ανάδευση με πιπέτα και σύντομη φυγοκέντρηση. Ακολουθεί επώαση σύμφωνα με το ακόλουθο πρόγραμμα ERP σε θερμοκυκλοποιητή:

- Προθέρμανση για το καπάκι στους 100 °C
- 30 °C για 30min
- Αναμονή στους 4 °C
- Κάθε σωληνάριο περιέχει 100μl

Μετά την επώαση πραγματοποιείται βελτιστοποίηση του μήκους των θραυσμάτων. Τα δείγματα φυγοκεντρώνονται σύντομα και προστίθενται σε αυτά 90μl SPB, αναδεύονται με πιπέτα επωάζονται για 5min και φυγοκεντρώνονται σύντομα. Ακολουθεί τοποθέτηση στη μαγνητική θήκη, αναμονή 2-5min μέχρι να γίνουν διαυγή και μεταφέρονται 185μl υπερκλειμένου σε νέο σωληνάριο 1,5ml. Στη συνέχεια, γίνεται προσθήκη 125μl SPB σε κάθε σωληνάριο, ανάδευση με πιπέτα, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min μέχρι να γίνουν διαυγή και απόρριψη του υπερκλειμένου. Ακολουθεί έκπλυση του κάθε δείγματος 2 φορές όπως αναφέρεται παραπάνω, προσθήκη 20μl RSB, ανάδευση με πιπέτα, σύντομη φυγοκέντρηση και τοποθέτηση σε μαγνήτη για 2-5min. Από κάθε δείγμα μεταφέρονται 17,5μl υπερκλειμένου σε νέο σωληνάριο 0,5ml.

### 3. Αδενυλίωση των 3' άκρων κάθε κλώνου

Αρχικά γίνεται προσθήκη 12,5μl ATL2 σε κάθε δείγμα, ανάδευση με πιπέτα και σύντομη φυγοκέντρηση. Ακολουθεί επώαση σύμφωνα με το ακόλουθο πρόγραμμα ATAIL70 σε θερμοκυκλοποιητή:

- Προθέρμανση για το καπάκι στους 100 °C
- 37 °C για 30min
- 70 °C για 5min
- 4 °C για 5min
- Αναμονή στους 4 °C
- Κάθε σωληνάριο περιέχει 30μl

Τα δείγματα φυγοκεντρώνται σύντομα και οδηγούνται για την πρόσδεση των ανταπτόρων.

### 4. Πρόσδεση ανταπτόρων-προσαρμοστών

Αρχικά γίνεται φυγοκέντρηση των DNA ανταπτόρων για 5sec και μεταφορά του LIG2 από τους -20 °C. Ακολουθεί προσθήκη σε κάθε δείγμα με τη σειρά 2,5μl RSB, 2,5μl LIG2 και 2,5μl DNA αντάπτορα-προσαρμοστή διαφορετικό σε κάθε δείγμα, ανάδευση με πιπέτα, σύντομη φυγοκέντρηση και επώαση σύμφωνα με το ακόλουθο πρόγραμμα LIG σε θερμοκυκλοποιητή:

- Προθέρμανση για το καπάκι στους 100 °C
- 30 °C για 10min
- Αναμονή στους 4 °C
- Κάθε σωληνάριο περιέχει 37,5μl

Μετά την επώαση γίνεται σύντομη φυγοκέντρηση και προσθήκη 3μl STL σε κάθε δείγμα, ανάδευση με πιπέτα και σύντομη φυγοκέντρηση. Στη συνέχεια, πραγματοποιείται καθαρισμός των τμημάτων σε 2 γύρους. Στον πρώτο γύρο προστίθενται αρχικά 42,5μl SPB, ενώ στο δεύτερο 50μl SPB. Ακολουθεί ανάδευση με πιπέτα, επώαση για 5min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min και απόρριψη του υπερκειμένου. Πραγματοποιείται έκπλυση του κάθε δείγματος 2 φορές με προσθήκη 200μl αιθανόλης 80%, επώαση στη μαγνητική θήκη για 30sec και απόρριψη του υπερκειμένου. Γίνεται σύντομη φυγοκέντρηση και χρήση πιπέτας 20μl για απομάκρυνση της περίσσειας αιθανόλης και αναμονή στη μαγνητική θήκη για 5min με ανοικτό καπάκι για εξάτμιση της εναπομείναντας

αιθανόλης. Έπειτα προστίθενται 52,5μl RSB για τον πρώτο γύρο, ενώ για τον δεύτερο 27,5μl RSB. Ακολουθεί ανάδευση με πιπέτα, επώαση 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και μεταφορά 50μl υπερκειμένου σε νέο σωληνάριο 1,5ml. Τα παραπάνω βήματα επαναλαμβάνονται για το δεύτερο γύρο και τελικά μεταφέρονται 25μl υπερκειμένου σε νέο σωληνάριο 0,5ml.

#### 5. Εμπλουτισμός των θραυσμάτων DNA

Τα σωληνάρια των δειγμάτων τοποθετούνται στον πάγο και προστίθενται 5μl PPC, 20μl EPM, αναδεύονται με πιπέτα και φυγοκεντρούνται σύντομα. Ακολουθεί πρόγραμμα PCR σε θερμοκυκλοποιητή PCRNano:

- Προθέρμανση για το καπάκι στους 100°C
- 95 °C για 3min
- 8 κύκλοι από:
  - 98 °C για 20sec
  - 60 °C για 15sec
  - 72 °C για 30sec
- 72 °C για 5min
- Αναμονή στους 4 °C
- Κάθε σωληνάριο περιέχει 50μl

Μετά την PCR ακολουθεί καθαρισμός των τμημάτων DNA με προσθήκη 35μl SPB, ανάδευση με πιπέτα, επώαση για 5min, σύντομη φυγοκέντρηση και τοποθέτηση στη μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές. Έπειτα μεταφέρονται 82μl υπερκειμένου σε νέο σωληνάριο 1,5ml και γίνεται προσθήκη 82μl SPB με ήπια ανάδευση και επώαση για 5min σε θερμοκρασία δωματίου. Στη συνέχεια, τα δείγματα τοποθετούνται σε μαγνητική θήκη για 2-5min μέχρι να γίνει το διάλυμα διαυγές απορρίπτεται το υπερκείμενο και ακολουθεί έκλυση 2 φορές όπως αναφέρεται παραπάνω. Τέλος, γίνεται προσθήκη 17,5μl RSB, ανάδευση με πιπέτα, επώαση για 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min και μεταφορά 15μl σε νέο σωληνάριο 1,5ml.

Σε αυτό το σημείο γίνεται ποσοτικός και ποιοτικός έλεγχος της βιβλιοθήκης με μεθόδους που αναφέρονται στο Εδάφιο 2.3 και 2.4.

Ασφαλής σημείο διακοπής με αποθήκευση του σωληναρίου στους  $-25^{\circ}\text{C}$  έως  $-15^{\circ}\text{C}$  μέχρι 7 ημέρες (Τέλος 1<sup>ης</sup> ημέρας).

#### 6. Πρώτη υβριδοποίηση ανιχνευτών (probes)

Αρχικά πραγματοποιείται ανάμειξη των δειγμάτων σε ένα σωληνάριο (pooling), σύμφωνα με τον παρακάτω πίνακα:

**Πίνακας 2.3:** Αριθμός δειγμάτων που αναμειγνύονται με τις κατάλληλες ποσότητες βιβλιοθήκης

Αριθμός Δειγμάτων	Ποσότητα κάθε βιβλιοθήκης	Συνολική ποσότητα Pool
3-plex	250ng	750ng
6-plex	200ng	1200ng
<b>9-plex</b>	<b>150ng</b>	<b>1350ng</b>
12-plex	100ng	1200ng

Αν ο τελικός όγκος είναι μεγαλύτερος από 40μl θα πρέπει να χρησιμοποιείται συμπυκνωτής κενού ή Amicon Ultra-0,5 centrifugal filter unit (0,5ml, 30kDa) μέχρι να συγκεντρωθεί το pool στα 40μl. Αν ο συνολικός όγκος είναι μικρότερος των 40μl, θα πρέπει να προστεθεί RSB έως τα 40μl.

Στη συνέχεια, γίνεται ανάμειξη σε νέο σωληνάριο 0,5ml, 40μl από DNA pool, 50μl CT3 και 10μl CEX, σύντομη φυγοκέντρηση και τοποθέτηση στον θερμοκυκλοποιητή με το ακόλουθο πρόγραμμα TE HYB:

- Προθέρμανση για το καπάκι στους  $100^{\circ}\text{C}$
- $95^{\circ}\text{C}$  για 10min
- 18 κύκλους του 1min ο καθένας, ξεκινώντας στους  $94^{\circ}\text{C}$  και μειώνοντας  $2^{\circ}\text{C}$  ανά κύκλο
- $58^{\circ}\text{C}$  για 90min
- Αναμονή στους  $58^{\circ}\text{C}$  μέχρι 24hrs
- Κάθε σωληνάριο περιέχει 100μl

#### 7. Πρόσδεση στους υβριδοποιημένους ανιχνευτές των σφαιριδίων στρεπταβιδίνης

Η πρώτη πρόσδεση πραγματοποιείται με προσθήκη 250μl SMB σε νέο σωληνάριο 1,5ml και μεταφορά των 100μl από την παραπάνω αντίδραση PCR. Ακολουθεί επώαση 25min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση σε μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Στη συνέχεια, γίνεται η πρώτη πλύση μετά από απομάκρυνση από τη μαγνητική θήκη, με προσθήκη 200μl SWS στο σωληνάριο και ανάδευση με πιπέτα, τοποθέτηση σε heat block στους 50°C για 30min, άμεση τοποθέτηση σε ζεστή μαγνητική θήκη μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Η παραπάνω πλύση επαναλαμβάνεται άλλη μια φορά και ακολουθεί η πρώτη έκλυση, με δημιουργία αρχικά του διαλύματος έκλυσης, το οποίο αποτελείται από 28,5μl EE1 και 1,5μl σε σωληνάριο 1,5ml. Από το συγκεκριμένο διάλυμα μεταφέρονται 23μl στο σωληνάριο του δείγματος με τα σφαιρίδια, γίνεται ανάδευση με πιπέτα, επώαση για 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για ~5min και μεταφορά 21μl υπερκειμένου σε νέο σωληνάριο 0,5ml. Στο νέο αυτό σωληνάριο προστίθενται 4μl ET2, αναδεύονται ήπια με πιπέτα και φυγοκεντρούνται σύντομα.

#### 8. Δεύτερη υβριδοποίηση ανιχνευτών (probes)

Για τη δεύτερη υβριδοποίηση προστίθενται στο σωληνάριο, το οποίο περιέχει 25μl από την DNA βιβλιοθήκη, 15μl RSB, 50μl CT3, 10μl CEX, φυγοκεντρούνται σύντομα και τοποθετούνται στο θερμοκυκλοποιητή για την υβριδοποίηση σύμφωνα με το πρόγραμμα TE HYB.

#### 9. Πρόσδεση στους υβριδοποιημένους ανιχνευτές των σφαιριδίων στρεπταβιδίνης

Η δεύτερη πρόσδεση πραγματοποιείται με προσθήκη 250μl SMB σε νέο σωληνάριο 1,5ml και μεταφορά των 100μl από την παραπάνω αντίδραση PCR. Ακολουθεί επώαση 25min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση σε μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Στη συνέχεια, γίνεται η



δεύτερη πλύση μετά από απομάκρυνση από τη μαγνητική θήκη, με προσθήκη 200μl SWS στο σωληνάριο και ανάδευση με πιπέτα, τοποθέτηση σε heat block στους 50°C για 30min, άμεση τοποθέτηση σε ζεστή μαγνητική θήκη μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Η παραπάνω πλύση επαναλαμβάνεται άλλη μια φορά και ακολουθεί η δεύτερη έκλουση, με δημιουργία αρχικά του διαλύματος έκλουσης, το οποίο αποτελείται από 28,5μl EE1 και 1,5μl σε σωληνάριο 1,5ml. Από το συγκεκριμένο διάλυμα μεταφέρονται 23μl στο σωληνάριο του δείγματος με τα σφαιρίδια, γίνεται ανάδευση με πιπέτα, επώαση για 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για ~5min και μεταφορά 21μl υπερκειμένου σε νέο σωληνάριο 0,5ml. Στο νέο αυτό σωληνάριο προστίθενται 4μl ET2, αναδεύονται ήπια με πιπέτα και φυγοκεντρώνονται σύντομα.

Ασφαλής σημείο διακοπής με αποθήκευση του σωληναρίου στους -25 °C έως -15 °C μέχρι 7 ημέρες (Τέλος 2<sup>ης</sup> ημέρας).

#### 10. Καθαρισμός της DNA βιβλιοθήκης με τα προσδεδεμένα σφαιρίδια στρεπταβιδίνης στους υβριδοποιημένους ανιχνευτές

Για τον καθαρισμό αρχικά προστίθενται 45μl SPB στο σωληνάριο, γίνεται επώαση για 5min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση σε μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Ακολουθεί έκπλυση του κάθε δείγματος 2 φορές με προσθήκη 200μl αιθανόλης 80%, επώαση στη μαγνητική θήκη για 30sec και απόρριψη του υπερκειμένου. Γίνεται σύντομη φυγοκέντρηση και χρήση πιπέτας 20μl για απομάκρυνση της περίσσειας αιθανόλης και αναμονή στη μαγνητική θήκη για 5min με ανοικτό καπάκι για εξάτμιση της εναπομείναντας αιθανόλης. Έπειτα προστίθενται 27,5μl RSB, γίνεται ανάδευση με πιπέτα, επώαση 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και μεταφορά 25μl υπερκειμένου σε νέο σωληνάριο 0,5ml.

### 11. Ενίσχυση των εμπλουτισμένων βιβλιοθηκών

Στο συγκεκριμένο βήμα γίνεται προσθήκη 5μl PPC και 20μl NEM στο σωληνάριο, σύντομη φυγοκέντρηση και τοποθέτηση στο θερμοκυκλοποιητή με το ακόλουθο πρόγραμμα AMP8:

- Προθέρμανση για το καπάκι στους 100 °C
- 98 °C για 30sec
- 8 κύκλοι των:
  - 98 °C για 10sec
  - 60 °C για 30sec
  - 72 °C για 30sec
- 72 °C για 5min
- Αναμονή στους 4 °C
- Το σωληνάριο περιέχει 50μl

### 12. Καθαρισμός των ενισχυμένων εμπλουτισμένων βιβλιοθηκών

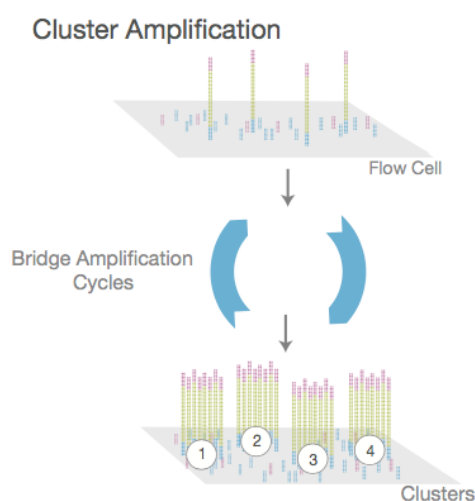
Για τον καθαρισμό αρχικά προστίθενται 45μl SPB στο σωληνάριο, γίνεται επώαση για 5min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση σε μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και απόρριψη του υπερκειμένου. Ακολουθεί έκπλυση του κάθε δείγματος 2 φορές με προσθήκη 200μl αιθανόλης 80%, επώαση στη μαγνητική θήκη για 30sec και απόρριψη του υπερκειμένου. Γίνεται σύντομη φυγοκέντρηση και χρήση πιπέτας 20μl για απομάκρυνση της περίσσειας αιθανόλης και αναμονή στη μαγνητική θήκη για 5min με ανοικτό καπάκι για εξάτμιση της εναπομείναντας αιθανόλης. Έπειτα προστίθενται 22μl RSB, γίνεται ανάδευση με πιπέτα, επώαση 2min σε θερμοκρασία δωματίου, σύντομη φυγοκέντρηση, τοποθέτηση στη μαγνητική θήκη για 2-5min μέχρι το διάλυμα να γίνει διαυγές και μεταφορά 20μl υπερκειμένου σε νέο σωληνάριο 0,5ml.

Σε αυτό το σημείο γίνεται ποσοτικός και ποιοτικός έλεγχος των εμπλουτισμένων βιβλιοθηκών με μεθόδους που αναφέρονται στο Εδάφιο 2.3 και 2.4.

Ασφαλής σημείο διακοπής με αποθήκευση του σωληναρίου στους -25 °C έως -15 °C μέχρι 7 ημέρες (Τέλος 3<sup>ης</sup> ημέρας).

## 2.6) Αλληλούχηση με σύστημα NextSeq 500 της Illumina

Στην αντίδραση αλληλούχησης νέας γενιάς, όπως έχει αναφερθεί στα Εδάφια 1.2.1, 1.2.2 και 2.1, η DNA πολυμεράση καταλύει την ενσωμάτωση επισημασμένων με φθορισμό τριφωσφορικών δεοξυριβονουκλεοτιδίων σύμφωνα με την αλυσίδα-μήτρα DNA κατά τη διάρκεια διαδοχικών κύκλων σύνθεσης DNA. Κατά τη διάρκεια κάθε κύκλου, στο σημείο ενσωμάτωσης, τα νουκλεοτίδια αναγνωρίζονται από φθορίζουσα διέγερση. Με αυτή την τεχνολογία δεν αλληλουχείται ένα μοναδικό θραύσμα DNA, αλλά η διαδικασία αυτή επεκτείνεται σε εκατομμύρια θραύσματα με μαζικό παράλληλο τρόπο (Εικόνα 2.2Ε). Η χημεία της αλληλούχησης της Illumina μέσω σύνθεσης είναι η ευρύτερα υιοθετούμενη μέθοδος στη βιομηχανία και παρέχει την υψηλότερη ακρίβεια και απόδοση χωρίς σφάλματα. Η αντίδραση αλληλούχησης περιλαμβάνει τη δημιουργία συμπλέγματος θραυσμάτων (Cluster) σε ειδική επιφάνεια (flow cell), στην οποία φορτώνεται η βιβλιοθήκη των δειγμάτων που προετοιμάστηκε στο Εδάφιο 2.5. Στην επιφάνεια του flow cell βρίσκονται ακινητοποιημένα τμήματα ολιγονουκλεοτιδίων, τα οποία υβριδοποιούνται συμπληρωματικά με τους αντάπτορες των θραυσμάτων. Κάθε θραύσμα, στη συνέχεια ενισχύεται ξεχωριστά μέσω ενίσχυσης γέφυρας (Bridge Amplification) με στόχο να δημιουργηθούν τα κλωνικά συμπλέγματα, τα οποία θα αλληλουχηθούν και θα δώσουν τα αντίστοιχα reads (Εικόνα 2.9).

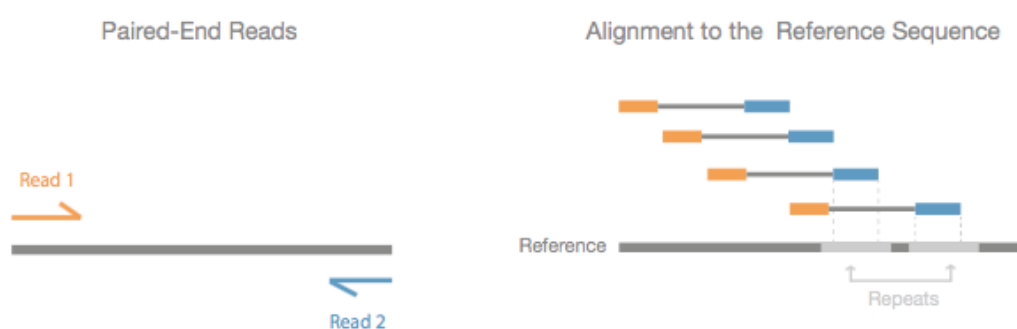


**Εικόνα 2.9:** Ενίσχυση κλωνικών συμπλεγμάτων. Η βιβλιοθήκη φορτώνεται στο flow cell, τα θραύσματα υβριδοποιούνται στην επιφάνεια και ενισχύονται για τη δημιουργία κλωνικών συμπλεγμάτων με ενίσχυση γέφυρας.

Η τεχνολογία αλληλούχησης της Illumina χρησιμοποιεί τη μέθοδο που βασίζεται σε αντιστρεπτό τερματισμό, στην οποία ανιχνεύονται μεμονωμένες βάσεις καθώς ενσωματώνονται στην αλυσίδα DNA. Δεδομένου ότι υπάρχουν

και οι 4 τροποποιημένες βάσεις που προκαλούν τερματισμό κατά τη διάρκεια κάθε κύκλου προσδιορισμού αλληλουχίας, ο φυσικός ανταγωνισμός ελαχιστοποιεί το θόρυβο ενσωμάτωσης και μειώνει σημαντικά τα επίπεδα σφαλμάτων σε σύγκριση με άλλες τεχνολογίες (Bentley 2008, Ross 2013).

Μια σημαντική πρόοδος στην τεχνολογία NGS αποτελεί η αλληλούχηση paired-end (Εικόνα 2.10). Η ανάλυση paired-end περιλαμβάνει αλληλούχηση και από τα δύο άκρα των θραυσμάτων DNA και ευθυγράμμιση των εμπρόσθιων και ανάστροφων αλληλουχιών με χρήση αλγορίθμων. Εκτός από την παραγωγή του διπλάσιου αριθμού ακολουθιών στον ίδιο χρόνο, οι ακολουθίες αυτές ευθυγραμμίζονται ως ζεύγη και καθιστούν δυνατή την ακριβέστερη ευθυγράμμιση με μεγαλύτερη ικανότητα προσδιορισμού INDELS, που δεν είναι δυνατόν να προσδιοριστούν με single-end αλληλούχηση. Ακόμα, η συγκεκριμένη προσέγγιση παράγει μεγαλύτερο αριθμό SNVs κατά την ευθυγράμμιση των ζευγών των ακολουθιών. Επιπλέον, επειδή είναι γνωστή η απόσταση μεταξύ των ζευγών ακολουθιών στα άκρα, οι αλγόριθμοι ευθυγράμμισης μπορούν να χρησιμοποιήσουν τις πληροφορίες αυτές για τη χαρτογράφησή τους σε επαναλαμβανόμενες περιοχές του γονιδιώματος με μεγαλύτερη ακρίβεια, που σε άλλες περιπτώσεις ο προσδιορισμός τους θα ήταν δύσκολος (Nakazato, 2013).



**Εικόνα 2.10:** Η ανάλυση paired-end περιλαμβάνει αλληλούχηση από τα δύο άκρα των θραυσμάτων DNA και ευθυγράμμιση των εμπρόσθιων-ανάστροφων ακολουθιών στην αλληλουχία αναφοράς. Επειδή είναι γνωστή η απόσταση μεταξύ των ζευγών ακολουθιών στα άκρα, οι αλγόριθμοι ευθυγράμμισης μπορούν να χρησιμοποιήσουν τις πληροφορίες αυτές για τη χαρτογράφησή τους σε επαναλαμβανόμενες περιοχές του γονιδιώματος με μεγαλύτερη ακρίβεια (Nakazato, 2013).

Το σύστημα NextSeq 500 παρέχει υψηλή ανάλυση αλληλουχιών με την απλότητα ενός επιτραπέζιου αναλυτή. Η χρήση του συστήματος επιτρέπει την εκτέλεση πολλών εφαρμογών σε υψηλότερο βάθος και ανάλυση. Για να εκτελεστεί μια σειρά ανάλυσης στο συγκεκριμένο σύστημα, ετοιμάζεται μια κασέτα αντιδραστηρίων, το flow cell και οι απαιτούμενες ρυθμίσεις του λογισμικού. Η δημιουργία του κλωνικού συμπλέγματος στην επιφάνεια του flow cell και η αντίδραση αλληλούχησης πραγματοποιούνται στο μηχάνημα. Τα συμπλέγματα απεικονίζονται με τη χρήση χημείας 2-καναλιών και συνδυασμό φίλτρων ειδικών για κάθε ένα από τους επισημασμένους με φθορισμό τερματιστές. Αφού ολοκληρωθεί η απεικόνιση μιας σειράς στο flow cell απεικονίζεται η επόμενη σειρά και η διαδικασία επαναλαμβάνεται για κάθε κύκλο αλληλούχησης. Μετά την ανάλυση εικόνας το λογισμικό εκτελεί την αντιστοίχιση βάσεων (base calling), το φιλτράρισμα και ο έλεγχος ποιότητας. Η διάρκεια εκτέλεσης της αντίδρασης εξαρτάται από τον αριθμό των κύκλων που πραγματοποιούνται. Το μέγιστο μήκος αντίδρασης για paired-end ανάλυση είναι 150 κύκλοι για κάθε άκρο. Σε μια σειρά αντίδρασης ο αριθμός των κύκλων που εκτελούνται για μία ακολουθία (read) είναι ένας επιπλέον κύκλος από αυτόν που αναλύονται, δηλαδή 151 κύκλοι. Ο επιπλέον κύκλος απαιτείται για υπολογισμούς φάσης και πρόφασης.

#### ***Υλικά (αντιδραστήρια, εξοπλισμός, αναλώσιμα)***

- Single-use NextSeq 500 Kit (flow cell, reagent cartridge, buffer cartridge)
- HT1 (Hybridization Buffer)
- RSB (Resuspension Buffer)
- NaOH 1N
- Tris-Hcl 200mM pH 7.0
- PhiX control 10nM
- Πιπέτες (Gilson) των 2μl, 10μl, 20μl, 200μl και 1000μl
- Ακροφύσια για πιπέτες
- Σωληνάρια τύπου Eppendorf των 500μl
- Σωληνάρια τύπου Eppendorf των 1,5ml
- Σωληνάρια τύπου Falcon των 15ml

- Σύστημα NextSeq 500
- Vortex
- Φυγόκεντρος
- Ηλεκτρονικός Υπολογιστής με ενσωματωμένη την πλατφόρμα BaseSpace της Illumina

### ***Πειραματική διαδικασία***

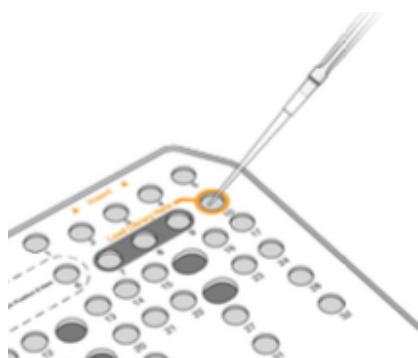
#### **1. Αραίωση και προετοιμασία της βιβλιοθήκης για αλληλούχηση**

Αρχικά γίνεται η προετοιμασία των αντιδραστηρίων και του PhiX control. Σε σωληνάριο 1,5ml προστίθενται 800μl dH<sub>2</sub>O, 200μl NaOH 1N για τελική συγκέντρωση 0,2N NaOH. Το HT1 και το RSB χρησιμοποιούνται σε θερμοκρασία δωματίου. Σε σωληνάριο 0,5ml προστίθενται 10μl PhiX 10nM και 15μl RSB, αναδεύονται και προκύπτει διάλυμα με τελική συγκέντρωση 4nM σε όγκο 25μl. Στη συνέχεια, συνδυάζονται σε σωληνάριο 5μl PhiX 4nM, 5μl NaOH, αναδεύονται με Vortex, φυγοκεντρώνονται σύντομα και επωάζονται για 5min σε θερμοκρασία δωματίου. Έπειτα, γίνεται προσθήκη 5μl Tris-HCl 200mM pH 7.0, ανάδευση και φυγοκέντρηση για 1min στα 280g. Ακολουθεί αραίωση του PhiX στα 20pM σε όγκο 1ml με προσθήκη 985μl HT1 και δεύτερη αραίωση στα 1.8pM με ανάμειξη 117μl PhiX 20pM και 1183μl HT1 και φυγοκέντρηση για 1min στα 280g.

Για την προετοιμασία της βιβλιοθήκης λαμβάνεται υπόψιν η συγκέντρωση την βιβλιοθήκης που προκύπτει από τον ποσοτικό έλεγχο (Εδάφιο 2.5). Για αρχική συγκέντρωση βιβλιοθήκης 4nM, αναδεύονται σε σωληνάριο 0,5ml 5μl 0.2 NaOH και 5μl από τη βιβλιοθήκη, αναδεύονται σύντομα και φυγοκεντρώνονται για 1min στα 280g. Ακολουθεί επώαση για 5min σε θερμοκρασία δωματίου και προσθήκη 5μl Tris-HCl 200mM pH 7.0 με σύντομη ανάδευση και φυγοκέντρηση για 1min στα 280g. Στη συνέχεια, γίνεται αραίωση της βιβλιοθήκης στα 20pM με προσθήκη 985μl HT1, σύντομη ανάδευση και φυγοκέντρηση για 1min στα 280g. Η βιβλιοθήκη τοποθετείται στους 4°C μέχρι την τελική αραίωση. Για την τελικά αραίωση η βιβλιοθήκη με συγκέντρωση 20pM αραιώνεται στα 1.8pM με ανάδευση 128μl βιβλιοθήκης 20pM, 1170μl HT1 και για 1,5% PhiX 2μl.

## 2. Αντίδραση αλληλούχησης

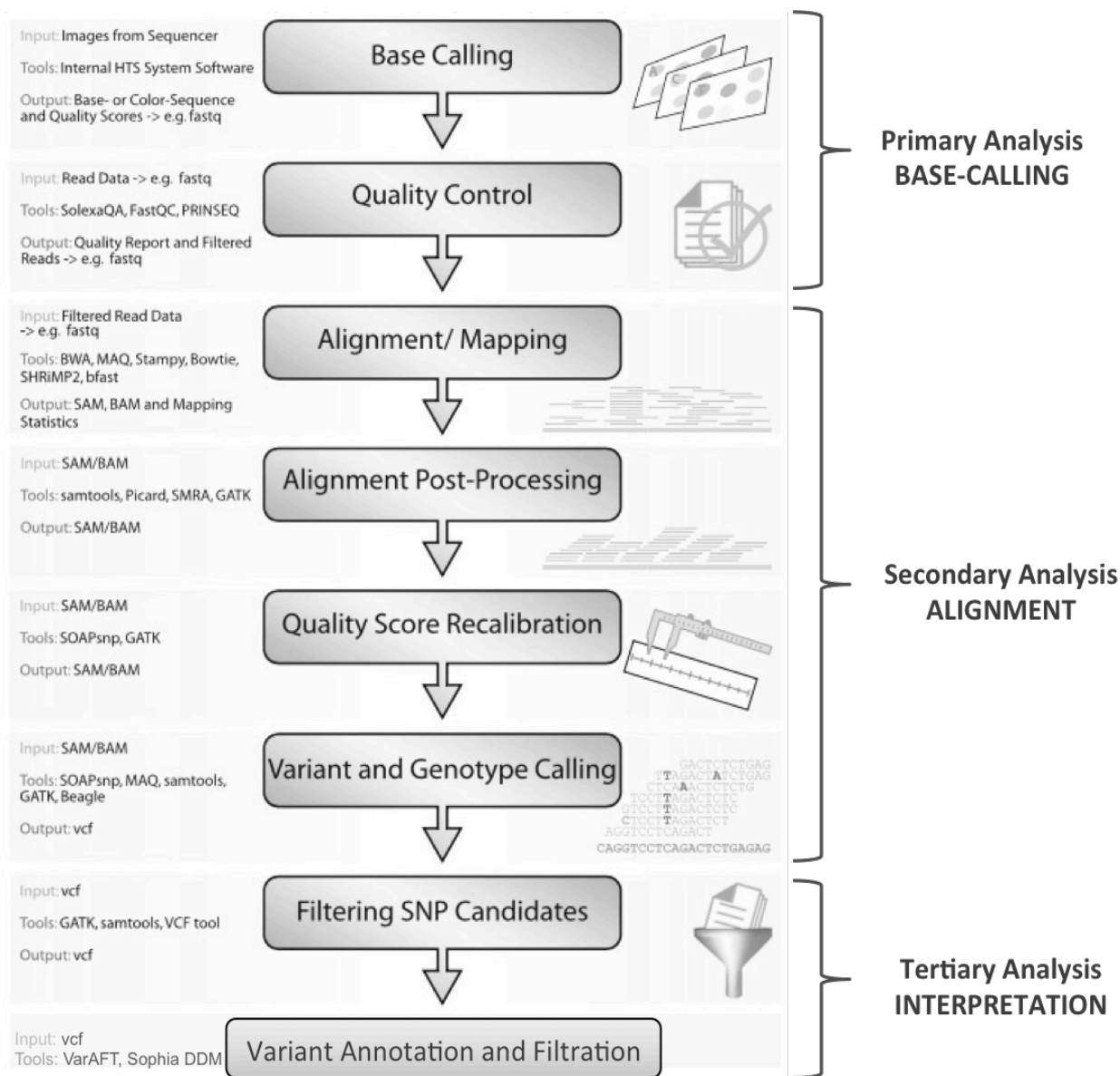
Για την αντίδραση αλληλούχησης ρυθμίζεται η εκτέλεση από την πλατφόρμα BaseSpace Prep Tab και προετοιμάζεται η κασέτα αντιδραστηρίων με επώαση σε θερμοκρασία δωματίου για ~60min. Στη συνέχεια γίνεται η φόρτωση της βιβλιοθήκης 1,3ml στην κασέτα (Εικόνα 2.11) και η τοποθέτηση του flow cell στο μηχάνημα. Έπειτα ελέγχονται τα αντιδραστήρια του μηχανήματος και ρυθμίζεται το μηχάνημα στις παραμέτρους Sequence για να εκτελεστεί η αντίδραση.



**Εικόνα 2.11:** Φόρτωση της βιβλιοθήκης στην κασέτα, στην εισδοχή #10.

### 2.7) Ανάλυση δεδομένων με χρήση Βιοπληροφορικών Συστημάτων

Η ανάλυση δεδομένων χωρίζεται σε τρία επίπεδα, όπως αναφέρεται στο Εδάφιο 1.4.2. Στην πρωτοβάθμια ανάλυση (Primary analysis), η οποία αφορά τη μετατροπή των raw data, βασιζόμενη στην ένταση του φθορισμού σε μικρές ακολουθίες νουκλεοτιδίων. Στη δευτεροβάθμια ανάλυση (Secondary analysis) στην οποία γίνεται η στοίχιση και η χαρτογράφηση των μικρών νουκλεοτιδικών ακολουθιών (reads) στο γονιδίωμα αναφοράς και ο προσδιορισμός των παραλλαγών (Variant calling). Στην τριτοβάθμια ανάλυση (Tertiary analysis), στην οποία πραγματοποιείται η ερμηνεία των παραλλαγών που εξάχθησαν στη δευτεροβάθμια ανάλυση με βάση τη λειτουργία, την προέλευση, τη συχνότητα εμφάνισης κ.α. (Εικόνα 2.12) (Endrullat, 2016).



**Εικόνα 2.12:** Διαγραμματική απεικόνιση της βιοπληροφορικής ροής εργασίας για την ανάλυση των δεδομένων που παράγονται από τα συστήματα αλληλούχησης νέας γενιάς (Altmann et al. 2012, Moorthie et al., 2013).

### 2.7.1) Πρωτοβάθμια ανάλυση (Primary analysis)

#### Βήμα 0: Base calling

Όπως αναφέρθηκε και παραπάνω, η τεχνολογία της Illumina χρησιμοποιεί αλληλούχηση βάσει σύνθεσης. Η χημεία αυτή χρησιμοποιεί τέσσερα σημασμένα με φθορισμό νουκλεοτίδια, τα οποία χρησιμοποιούνται για να αλληλουχηθούν τα δισεκατομμύρια συμπλέγματα κλώνων που έχουν



δημιουργηθεί στο flow-cell. Κατά τη διάρκεια κάθε κύκλου προσδιορισμού αλληλουχίας προστίθεται μια μόνο σημασμένη βάση στη νεοσυντιθέμενη αλυσίδα DNA. Τα τροποποιημένα νουκλεοτίδια περιέχουν μια αντιστρέψιμη ομάδα αποκλεισμού που χρησιμεύει ως τερματιστής για τον πολυμερισμό. Έτσι μετά από κάθε ενσωμάτωση, η ένταση της φθορίζουσας χρωστικής απεικονίζεται σε ειδική συσκευή για να ταυτοποιηθεί η βάση και εν συνεχεία η ομάδα αποκλεισμού διασπάται ενζυματικά για να επιτραπεί η ενσωμάτωση του επόμενου νουκλεοτιδίου. Ο προσδιορισμός της βάσης γίνεται απευθείας από τις μετρήσεις έντασης σήματος φθορισμού κατά τη διάρκεια κάθε κύκλου. Το base calling εκτελείται αυτομάτως από την πλατφόρμα αλληλούχησης χωρίς τη δυνατότητα εξωτερικής παρέμβασης (Altmann 2012, Gargis 2015).

**Πίνακας 2.4:** Software tools για base calling

Όνομα	Περιγραφή	Link
Ibis (Kircher et al. 2009)	Base calling: Illumina	<a href="http://bioinf.eva.mpg.de/Ibis/">http://bioinf.eva.mpg.de/Ibis/</a>
naiveBayesCall (Kao and Song 2011)	Base calling: Illumina	<a href="http://bayescall.sourceforge.net">http://bayescall.sourceforge.net</a>

### Βήμα 1: Ποιοτικός έλεγχος (Quality scoring)

Οι περισσότερες πλατφόρμες αλληλούχησης παρέχουν τα δεδομένα ακολουθιών απευθείας σε μορφή FASTQ (Εικόνα 2.13), τα οποία μπορούν να χρησιμοποιηθούν για μετρήσεις ποιότητας και να παρέχουν σημαντικές πληροφορίες σχετικά με την ακρίβεια του base calling. Η ακρίβεια αυτή εκτιμάται με τη βαθμολογία ποιότητας Phred (Qscore,  $Q = -10 \log_{10} P$ ), την πιο κοινή μέτρηση αξιολόγησης, υποδεικνύοντας την πιθανότητα μια βάση να μην έχει προσδιοριστεί σωστά από τον αναλυτή. Για παράδειγμα, αν το Phred δώσει τιμή Q30 σε μία βάση, η πιθανότητα να είναι λανθασμένη είναι 1 στα 1000. Αυτό σημαίνει ότι η ακρίβεια βάσης είναι 99,9%. Η χαμηλότερη αποδεκτή ακρίβεια βάσης είναι 99% (Q20) και σημαίνει ότι η πιθανότητα λανθασμένης βάσης είναι 1 στις 100, δηλαδή ότι ανά 100bp θα περιέχεται 1 σφάλμα. Όταν η βαθμολογία ποιότητας είναι μεταξύ Q30-Q40, σχεδόν όλες οι βάσεις έχουν τοποθετηθεί σωστά με μηδενικά σφάλματα (Altmann 2012).

Label: @FORJUSP02AJWD1
   
 Sequence: CCGTCAATTCATTTAAGTTTAAACCTTGCGGCCGTACTCCCAGGCGGT
   
 Base = T, Q = A = 25: +
   
 Q Scores (as ASCII charts): AAAAAAAAAAAAAA::99@: ::::??@: ::FFAAAAACCAA: ::::BB@@?A?

Πηγή: [https://www.abmgood.com/marketing/knowledge\\_base/next\\_generation\\_sequencing\\_data\\_analysis.php](https://www.abmgood.com/marketing/knowledge_base/next_generation_sequencing_data_analysis.php)

**Εικόνα 2.13:** FASTQ μορφή των raw data αλληλούχησης. Η πρώτη σειρά αποτελεί την ταυτότητα της ακολουθίας. Η δεύτερη σειρά είναι η ακολουθία που παρήγαγε ο αναλυτής. Η τρίτη σειρά αποτελεί μέρος της ταυτότητας. Η τέταρτη σειρά είναι η βαθμολογία ποιότητας (Qscore) σύμφωνα με τους χαρακτήρες ASCII (<http://www.asciitable.com/>).

### Πίνακας 2.5: Software tools για έλεγχο ποιότητας

Όνομα	Περιγραφή	Link
FastQC	Quality control	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
SolexaQA (Cox et al. 2010)	Quality control	<a href="http://solexaqa.sourceforge.net">http://solexaqa.sourceforge.net</a>

#### **FastQC:**

Η εφαρμογή FastQC παρέχει ποιοτικό έλεγχο για τα δεδομένα που παράγονται από τις πλατφόρμες αλληλούχησης και μπορεί να εντοπίσει προβλήματα που προέρχονται είτε από τον αναλυτή, είτε από το υλικό της βιβλιοθήκης. Τα αποτελέσματα του ελέγχου παρέχονται με απλό τρόπο μέσα από ένα σύνολο αναλύσεων δίνοντας γρήγορη εικόνα για την ποιότητα των δεδομένων. Οι κύριες λειτουργίες του FastQC είναι η εισαγωγή δεδομένων σε FASTQ μορφή (και σε άλλες μορφές, BAM, SAM), γρήγορη επισκόπηση των περιοχών που εμφανίζουν πρόβλημα (duplicates, adapters, dimers κ.α.), περίληψη γραφημάτων και πινάκων, στατιστικά στοιχεία για την ανάλυση, Qscore για κάθε ακολουθία κ.α..

### 2.7.2) Δευτεροβάθμια ανάλυση (Secondary analysis)

#### Βήμα 2: Alignment/mapping reads to reference genome

Το επόμενο βήμα για σχεδόν όλες τις εφαρμογές είναι η στοίχιση/χαρτογράφηση των ακολουθιών που παράγονται στην αλληλουχία αναφοράς. Για να πραγματοποιηθεί όμως αυτό το βήμα, θα πρέπει οι ακολουθίες να είναι όσο το δυνατόν πιο 'καθαρές' και αξιόπιστες, δηλαδή χωρίς να περιλαμβάνουν εσφαλμένα τμήματα στα άκρα ή τμήματα

ανταπτόρων-προσαρμοστών και εκκινήτων. Η ύπαρξη τέτοιων περιοχών μειώνει την ικανότητα στοίχισης και διαπιστώνεται από τον ποιοτικό έλεγχο στο βήμα 1. Για την απομάκρυνση τέτοιων περιοχών χρησιμοποιούνται εργαλεία, όπως το FASTX-toolkit (Trimmer, Clipper), το PRINSEQ κ.α.. Εφόσον οι ακολουθίες τηρούν τα ποιοτικά κριτήρια, οδηγούνται για στοίχιση χρησιμοποιώντας ειδικούς αλγορίθμους.

Για τη στοίχιση των ακολουθιών χρησιμοποιούνται δύο προσεγγίσεις: η πρώτη χρησιμοποιεί το μετασχηματιστή Burrows-Wheeler (BWT) (Burrows and Wheeler 1994) για την αποτελεσματική ανάλυση των δεδομένων και η δεύτερη βασίζεται στο hashing για να επιτευχθεί το βήμα στοίχισης, επιτρέποντας τη γρήγορη πρόσβαση στις πληροφορίες σχετικά με τη θέση των ακολουθιών στο γονιδίωμα αναφοράς. Το κύριο πλεονέκτημα των αλγορίθμων που βασίζονται σε BWT, σε σχέση με τους hash-αλγόριθμους είναι η ταχύτητα επεξεργασίας. Ωστόσο οι BWT αλγόριθμοι δεν είναι τόσο ευαίσθητοι, όσο οι hash-αλγόριθμοι και συνεπώς μπορεί να εισάγουν σφάλματα χαρτογράφησης, ειδικά σε περιοχές με υψηλή πολυμορφικότητα.

Γενικά, η επιλογή του εργαλείου στοίχισης επηρεάζει σημαντικά το αποτέλεσμα της ανάλυσης, κάτι που φαίνεται στον προσδιορισμό των παραλλαγών, καθώς λανθασμένη στοίχιση οδηγεί σε απόκλιση από την αλληλουχία αναφοράς και σε εσφαλμένη ταξινόμηση των παραλλαγών για την παρακάτω επεξεργασία. Επιπλέον, πέρα από την επιλογή του αλγορίθμου στοίχισης, είναι απαραίτητο να επιλεγούν και οι κατάλληλες ρυθμίσεις παραμέτρων. Αν επιτραπούν μόνο τέλειες στοιχίσεις μεταξύ ακολουθιών και αλληλουχίας αναφοράς, η ανάλυση δε θα εξάγει διαφορές, επομένως δε θα μπορούν να ανιχνευθούν SNVs. Αντιστρόφως, αν επιτραπούν πολλές αναντιστοιχίες θα οδηγήσουν σε λανθασμένες στοιχίσεις και κατά συνέπεια σε πολλά ψευδώς θετικά SNVs. Μόλις οι ακολουθίες στοιχιστούν με την αλληλουχία αναφοράς, τα δεδομένα αποθηκεύονται σε μορφή SAM (Sequence Alignment Map). Η μορφή αυτή αποθηκεύει πληροφορίες σχετικά τη στοίχιση/χαρτογράφηση και την ποιότητά της (Altmann 2012, Gargis 2015).

**Πίνακας 2.6:** Software tools για στοίχιση/χαρτογράφηση

Όνομα	Περιγραφή	Link
Bowtie2 (Langmead et al. 2012)	Mapping with BWT	<a href="http://bowtie-bio.sourceforge.net">http://bowtie-bio.sourceforge.net</a>
BWA (Li H. and Durbin 2009)	Mapping with BWT	<a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a>
MAQ (Li H. et al. 2008)	Mapping hash-based	<a href="http://maq.sourceforge.net">http://maq.sourceforge.net</a>
Stampy (Lunter and Goodson 2011)	Mapping hash-based with optional speedup using BWA	<a href="http://www.well.ox.ac.uk/project-stampy">http://www.well.ox.ac.uk/project-stampy</a>
SOAP (Li R. et al. 2009)	Mapping hash-based	<a href="http://soap.genomics.org.cn">http://soap.genomics.org.cn</a>

### **Bowtie2:**

Ο Bowtie2 αλγόριθμος είναι ένα εξαιρετικά γρήγορο και αποτελεσματικό εργαλείο για τη στοίχιση των ακολουθιών σε γονιδιώματα αναφοράς. Πολλοί αλγόριθμοι χρησιμοποιούν γονιδιωματικούς δείκτες για να περιορίσουν τη λίστα με τις υποψήφιες θέσεις χαρτογράφησης. Ο δείκτης Full-text Minute Index (FM-index) είναι ένας γρήγορος και αποτελεσματικός δείκτης που χρησιμοποιείται ευρέως και βασίζεται στο μετασχηματισμό BWT, επιτρέποντας τη διατήρηση του memory footprint της RAM μικρό. Η στοίχιση με τη βοήθεια δείκτη μπορεί να είναι αρκετά ανεπαρκής όταν επιτρέπεται η ύπαρξη κενών (gaps) στην αλληλουχία, τα οποία προκύπτουν είτε από σφάλματα της αλληλούχησης, είτε από πραγματικά Indels, επιβραδύνοντας την ανάλυση. Ωστόσο, ο Bowtie2 επεκτείνοντας τον Bowtie και χρησιμοποιώντας FM-index, επιτρέπει τη στοίχιση με κενά χρησιμοποιώντας 2 στάδια: ένα αρχικό ungapped seed-finding, το οποίο ωφελείται από την ταχύτητα και αποτελεσματικότητα της χρήσης του δείκτη και ένα δεύτερο στάδιο gapped-extension, το οποίο χρησιμοποιεί δυναμικό προγραμματισμό και ωφελείται από την αποτελεσματικότητα της παράλληλης επεξεργασίας πολλαπλών δεδομένων (SIMD). Ο συνδυασμός αυτών των 2 σταδίων επιταχύνει το συνδυασμό ταχύτητας, ευαισθησίας και ακρίβειας που απαιτείται για το NGS. Για κάθε ακολουθία ο Bowtie2 ακολουθεί ακολουθεί τέσσερα βήματα: στο πρώτο βήμα, ο αλγόριθμος εξάγει 'seeds', υποσυστοιχίες των ακολουθιών που θα ιεραρχηθούν για στοίχιση, στο δεύτερο βήμα, οι εξαγόμενες υποσυστοιχίες στοιχίζονται στην αλληλουχία αναφοράς με ungapped τρόπο χρησιμοποιώντας FM-index, στο τρίτο βήμα τα

στοιχιζόμενα seeds ταξινομούνται και υπολογίζεται η θέση τους στο γονιδίωμα αναφοράς και στο τέταρτο βήμα, τα seeds επεκτείνονται σε πλήρης αλληλουχίες στοίχισης χρησιμοποιώντας δυναμικό προγραμματισμό με επιτάχυνση SIMD (Langmead et al. 2012).

**Πίνακας 2.7:** Γραμμή εντολών του αλγορίθμου Bowtie2.

<b>Bowtie2 command line</b> (Langmead et al. 2012)	
<b>--sensitive</b> -D 15 -R 2 -N0 -L 22 -i S,1,1.15	Settings for alignment
<b>-l 0</b>	Minimum fragment length
<b>-X 500</b>	Maximum fragment length
<b>-S [file].sam</b>	Output file
<b>-x</b>	Reference genome
<b>-1</b>	Input file (pair 1)
<b>-2</b>	Input file (pair 2)

### Βήμα 3: Alignment post-processing

Πριν το πραγματικό variant calling, απαιτείται οι στοιχιζόμενες ακολουθίες να ταξινομηθούν αναφορικά με τη χρωμοσωμική θέση. Σε αυτό το βήμα τα SAM αρχεία μετατρέπονται σε BAM (Binary Alignment Map), τα οποία έχουν μικρότερο μέγεθος και είναι πιο εύχρηστα για τα βιοπληροφορικά εργαλεία. Ακολουθώς, επειδή η PCR που χρησιμοποιείται για την ενίσχυση της βιβλιοθήκης παράγει πολλά αντίγραφα των περιοχών (duplicates), γίνεται απομάκρυνσή τους με ειδικά προγράμματα για την αποφυγή ψευδώς θετικών αποτελεσμάτων. Επίσης, σε αυτό το βήμα υπάρχει η δυνατότητα επαναστοίχισης των ακολουθιών κυρίως γύρω από τα indels για μεγαλύτερη ειδικότητα.

**Πίνακας 2.8:** Software tools για alignment post-processing

Όνομα	Περιγραφή	Link
Samtools (Li H. et al. 2009)	Manipulation of SAM files, viewing, sorting, filtering, SNP calling, duplicate removal and more. Likely the fastest tool form SAM file manipulation	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
Picard	Manipulation of SAM files, viewing, sorting, filtering, duplicate removal, generates statistics and more.	<a href="http://picard.sourceforge.net">http://picard.sourceforge.net</a>

### Βήμα 4: Ποιοτικός έλεγχος στοίχισης/χαρτογράφησης

Μετά τα βήματα 2 και 3 ακολουθεί επαναληπτικός ποιοτικός έλεγχος χρησιμοποιώντας τα BAM αρχεία που παρήχθησαν στο παραπάνω βήμα, για

να διαπιστωθεί η ακρίβεια και η ειδικότητα της στοίχισης στην αλληλουχία αναφοράς. Αυτό γίνεται εύκολα δημιουργώντας στατιστικά στοιχεία με χρήση ειδικών εφαρμογών, υπολογίζοντας το ποσοστό των ακολουθιών που χαρτογραφήθηκαν επιτυχώς και αυτών που απέτυχαν. Ειδικότερα, σε paired-end αλληλούχηση δίνονται πληροφορίες για το ποσοστό των ακολουθιών που αναλύθηκαν ανά ζεύγη, καθώς και την κατανομή της κάλυψης κατά μήκος των χρωμοσωμάτων.

**Πίνακας 2.9:** Software tools για ποιοτικό έλεγχο στοίχισης/χαρτογράφησης

Όνομα	Περιγραφή	Link
Samtools (Li H. et al. 2009)	Manipulation of SAM files, viewing, sorting, filtering, SNP calling, duplicate removal, generates statistics. Likely the fastest tool form SAM file manipulation	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
QualiMap (Okonechnikov et al. 2015)	Quality control of alignment sequencing data.	<a href="http://qualimap.bioinfo.cipf.es/">http://qualimap.bioinfo.cipf.es/</a>

### **QualiMap:**

Το QualiMap είναι μια εφαρμογή η οποία παρέχει γραφικό περιβάλλον χρήστη (GUI) για τον ποιοτικό έλεγχο των δεδομένων στοίχισης των ακολουθιών. Εξετάζει τα δεδομένα αυτά από BAM αρχείο και παρέχει μια γενική εικόνα για τις περιοχές που είχαν προβληματική στοίχιση, ώστε να συνεχιστεί η μετέπειτα ανάλυση με τα σωστά κριτήρια.

### **Βήμα 5: Προσδιορισμός (Ανάκληση) παραλλαγών (Variant calling)**

Το τελευταίο βήμα της δευτεροβάθμιας ανάλυσης είναι ο προσδιορισμός των παραλλαγών, όπου εντοπίζονται οι διαφορές μεταξύ του γονιδιώματος αναφοράς και των ακολουθιών που έχουν στοιχηθεί σε αυτό. Αρκετά προγράμματα λογισμικού που χρησιμοποιούν ποικιλία αλγορίθμων είναι διαθέσιμα για να εκτιμηθεί η πιθανότητα εμφάνισης ή απουσίας μιας παραλλαγής. Αυτοί οι αλγόριθμοι χρησιμοποιούν ποικίλες μεθόδους για να εκτελέσουν τον προσδιορισμό, που μπορεί να περιλαμβάνουν τον υπολογισμό του αριθμού των ακολουθιών που σχετίζονται με κάθε αλληλόμορφο, αφού καθοριστούν κατάλληλα όρια και ποιότητες χαρτογράφησης, συχνότητες αλληλομόρφων, ιδιότητες πλατφόρμας

αλληλούχησης, δεδομένα ανισορροπίας σύνδεσης κ.α. Αυτές οι προσεγγίσεις εκμεταλλεύονται τη διπλοειδή φύση του ανθρώπου. Ωστόσο, άλλες προσεγγίσεις χρησιμοποιούν Bayesian μοντέλα για τον προσδιορισμό των παραλλαγών (Coonrod, 2013). Επίσης, συχνά χρησιμοποιούνται παράμετροι όπως η ελάχιστη συχνότητα παραλλαγών και η ελάχιστη κάλυψη (coverage). Η λίστα των αναγνωρισμένων παραλλαγών καταγράφεται σε ψηφιακό αρχείο VCF (Variant Call Format) και χρησιμοποιείται ευρύτατα περιέχοντας πληροφορίες που θα χρησιμοποιηθούν στο επόμενο επίπεδο ανάλυσης, οι οποίες μπορεί να είναι διαφορετικές από εργαστήριο σε εργαστήριο ή από χειριστή σε χειριστή μέσα στο ίδιο εργαστήριο (Gargis 2015).

**Πίνακας 2.10:** Software tools για προσδιορισμό παραλλαγών

Όνομα	Περιγραφή	Link
Samtools (Li H. et al. 2009)	Manipulation of SAM files, viewing, sorting, filtering, SNP calling, duplicate removal, generates statistics. Likely the fastest tool form SAM file manipulation	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
BCFtools	Manipulation of BCF/VCF files.	<a href="https://samtools.github.io/bcftools/">https://samtools.github.io/bcftools/</a>
VCFtools (Danecek et al. 2011)	Manipulate VCF files, generate statistics on SNVs	<a href="http://vcftools.sourceforge.net">http://vcftools.sourceforge.net</a>

### **Samtools:**

Πρόκειται για ένα σύνολο βοηθητικών προγραμμάτων που χειρίζονται SAM και BAM αρχεία. Πραγματοποιεί ταξινόμηση, συγχώνευση, διαλογή, κατηγοριοποίηση, αφαίρεση PCR duplicates, εξάγει στατιστικά και ποιοτικά αποτελέσματα. Επίσης, παραγεί BCF αρχεία από τα BAM (Binary Call Format), τα οποία περιέχουν τις πιθανές παραλλαγές που θα καταγραφούν στη συνέχεια στα VCF αρχεία.

**Πίνακας 2.11:** Γραμμή εντολών του αλγορίθμου Samtools για χειρισμό και παραγωγή αρχείων Bam.

<i>Samtools command line</i> (Li H. et al. 2009)	
<code>view -bS [file].sam &gt; [file].bam</code>	Convert SAM format to Bam format
<code>sort [file].bam &gt; [sortfile].bam</code>	Sort alignments
<code>rmdup -S [sortfile].bam &gt; [uniquefile].bam</code>	Remove duplicates from paired-end alignments
<code>index -b [uniquefile].bam</code>	Index a BAM file for fast random access
<code>flagstat [uniquefile].bam</code>	Calculate and print statistics for the input file
<code>mpileup -g -f /hg19.fa [uniquefile].bam &gt; [uniquefile].bcf</code>	Generate BCF file for BAM file

### **BCFtools:**

Είναι ένα σύνολο βοηθητικών προγραμμάτων που χειρίζονται τα αρχεία με τις παραλλαγές BCF ή VCF.

**Πίνακας 2.12:** Γραμμή εντολών των αλγορίθμων Bcftools για χειρισμό και παραγωγή αρχείων VCF.

<i><b>Bcftools command line</b></i> (Danecek et al. 2011)	
<code>view -bvcg [uniquefile].bcf &gt; [newfile].vcf</code>	Generate VCF file with the variants.

Η μορφή αρχείων VCF είναι η πρότυπη μορφή αποθήκευσης παραλλαγών. Το αρχείο αυτό περιλαμβάνει στήλες και γραμμές με πληροφορίες για το είδος της παραλλαγής, αλλά και τις ποιοτικές παραμέτρους που αντιστοιχούν σε αυτές (Πίνακας 2.12).

**Πίνακας 2.13:** Περιεχόμενο στηλών και γραμμών που περιλαμβάνονται σε αρχείο VCF.

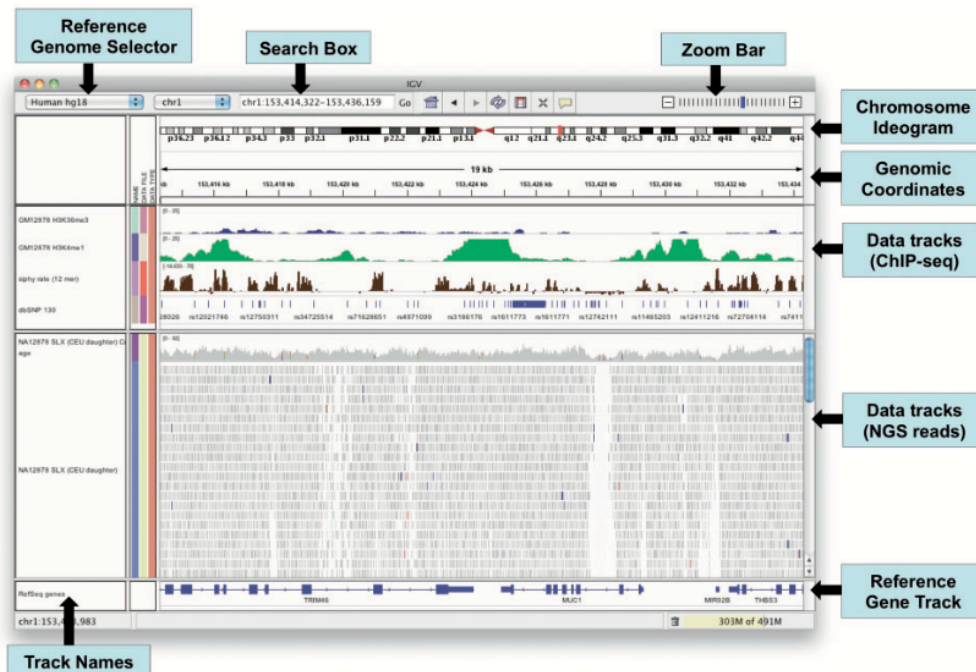
Col	Field	Description
1	CHROM	Chromosome name
2	POS	1-based position. For an indel, this is the position preceding the indel.
3	ID	Variant identifier. Usually the dbSNP rsID.
4	REF	Reference sequence at POS involved in the variant. For a SNP, it is a single base.
5	ALT	Comma delimited list of alternative sequence(s).
6	QUAL	Phred-scaled probability of all samples being homozygous reference.
7	FILTER	Semicolon delimited list of filters that the variant fails to pass.
8	INFO	Semicolon delimited list of variant information.
9	FORMAT	Colon delimited list of the format of individual genotypes in the following fields.
10+	Sample(s)	Individual genotype information defined by FORMAT.

### **IGV:** Οπτικοποίηση δεδομένων στοίχισης

Το Integrated Genome Viewer (IGV) είναι ένα πρόγραμμα οπτικοποίησης των γονιδιωματικών δεδομένων του BAM αρχείου σε μορφή στοίχισης στο γονιδίωμα αναφοράς. Παρέχει πληροφορίες για τις συντεταγμένες των χρωμοσωμάτων πάνω στα οποία στοιχίζονται τα reads της αλληλούχησης,



πληροφορίες για την ποιότητα κάλυψης, αλλά και για τις παραλλαγές που εντοπίζονται (Εικόνα 2.14) (Thorvaldsdottir et al., 2013).



**Εικόνα 2.14:** Παράθυρο εφαρμογής IGV (Thorvaldsdottir et al., 2013).

#### Βήμα 6: Φιλτράρισμα των υποψήφιων παραλλαγών (Filtering SNVs)

Το φιλτράρισμα των υποψήφιων παραλλαγών είναι ένα σημαντικό βήμα για τη μείωση των ψευδώς θετικών αποτελεσμάτων. Τα τυπικά εφαρμοζόμενα φίλτρα ελέγχουν για αποκλίσεις από την ισορροπία Hardy-Weinberg, ελάχιστο και μέγιστο βάθος ανάγνωσης, θόρυβος αλυσίδας. Αν και στο συγκεκριμένο βήμα μπορεί να αφαιρεθούν και πραγματικές παραλλαγές, είναι ένα σημαντικό εργαλείο για τη μείωση των παραλλαγών και την ευκολότερη εφαρμογή των επόμενων βημάτων.

#### **VCFTools:**

Είναι ένα πακέτο προγραμμάτων που έχει σχεδιαστεί για να δουλεύει με VCF αρχεία. Ο στόχος του είναι να παρέχει εύκολα προσβάσιμες μεθόδους για την εργασία με τα σύνθετα δεδομένα των γενετικών παραλλαγών. Εκτελεί λειτουργίες, όπως φιλτράρισμα συγκεκριμένων παραλλαγών, συγκρίνει αρχεία, συγκεντρώνει παραλλαγές κ.α..

**Πίνακας 2.14:** Γραμμή εντολών των αλγορίθμων Vcftools για χειρισμό και παραγωγή αρχείων VCF.

<b>Vcftools command line</b> (Danecek et al. 2011)	
<code>vcfutils.pl varFilter -Q30 -d [depthnormfactor] -D 1000 [newfile].vcf &gt; [file].vcf</code>	Filtering BCF file including the minimum and maximum number of reads, base calling quality.

### 2.7.3) Τριτοβάθμια ανάλυση (Tertiary analysis)

#### Βήμα 7: Σχολιασμός παραλλαγών (Variant annotation)

Η τριτοβάθμια ανάλυση χρησιμοποιεί τα αρχεία που παράγονται στη δευτεροβάθμια ανάλυση (VCF) για την ερμηνεία των παραλλαγών που έχουν καταγραφεί σε αυτά. Η ερμηνεία των γονιδιωματικών δεδομένων περιλαμβάνει την ανάλυση παραλλαγών για την εκτίμηση της προέλευσης τους, τη μοναδικότητα τους και τις πιθανές λειτουργικές επιπτώσεις. Η διαδικασία αυτή ξεκινάει με το variant annotation όπου γίνεται η συλλογή και η σύνδεση όλων των διαθέσιμων πληροφοριών με κάθε παραλλαγή. Αυτές οι πληροφορίες περιλαμβάνουν δεδομένα ποιότητας, τη θέση της παραλλαγής στο γονιδίωμα, δεδομένα για τα επίπεδα γονιδίου και μεταγράφου, το γονότυπο, τη συχνότητα στο γενικό πληθυσμό, την επίπτωση στο γονίδιο και στην πρωτεΐνη, την πρόβλεψη παθογένειας και τη συσχέτιση με ασθένειες. Σε γονιδιακό επίπεδο περιλαμβάνουν τη λειτουργία του γονιδίου όπου βρίσκεται η παραλλαγή, πρότυπο έκφρασης, εμπλοκή σε μοριακά μονοπάτια και συσχέτιση γονοτύπου-φαινοτύπου (Moorthie et al., 2013). Τα περισσότερα συστήματα για σχολιασμό παραλλαγών χρησιμοποιούν βάσεις δεδομένων πρόβλεψης, όπως, SIFT (Sim et al., 2012), Polyphen2 (Adzhubei et al., 2010), LRT (Chun and Fay, 2009), Mutation-Taster (Schwarz et al., 2014), Mutation Assessor (Reva et al., 2011), FATHMM (Shihab et al., 2013), MetaSVM and MetaLR (Dong et al., 2015), CADD (Kircher et al., 2014), PROVEAN (Choi et al., 2012), fathmm-MKL (Shihab et al., 2013), and DANN (Quang et al., 2015), UMD-Predictor (only for VarAFT) (Salgado et al., 2016).

Το πιο ευρέως χρησιμοποιούμενο σύστημα για variant annotation είναι το ANNOVAR. Πρόκειται για ένα αποτελεσματικό εργαλείο που χρησιμοποιεί ενημερωμένες πληροφορίες για να σχολιάσει λειτουργικά τις γενετικές παραλλαγές που ανιχνεύονται σε σύγκριση με το γονιδίωμα αναφοράς (hg19,

hg38). Λαμβάνοντας υπόψη τη λίστα των παραλλαγών με τη χρωμοσωμική θέση και τη νουκλεοτιδική αλλαγή, το ANNOVAR μπορεί να εκτελέσει:

- Σχολιασμό βάσει-γονιδίων: Προσδιορίζεται αν τα SNVs προκαλούν αλλαγή στην κωδικοποιούσα περιοχή και πως αυτή επηρεάζει τα αμινοξέα. Για αυτό το σχολιασμό αντλούνται πληροφορίες από βάσεις δεδομένων, όπως RefSeq genes, UCSC genes, Ensembl genes κ.α..
- Σχολιασμό βάσει-θέσης: Προσδιορίζονται παραλλαγές σε συγκεκριμένες περιοχές, όπως συντηρημένες περιοχές, ρυθμιστικές περιοχές, περιοχές δέσμευσης μεταγραφικών παραγόντων κ.α. Για αυτό το σχολιασμό αντλούνται πληροφορίες από βάσεις δεδομένων όπως, ENCODE.
- Σχολιασμό βάσει-φίλτρων: Προσδιορίζονται παραλλαγές που τεκμηριώνονται σε συγκεκριμένες βάσεις δεδομένων όπως dbSNP, 1000 Genome Project, ESP6500 και υπολογίζονται βαθμολογίες πρόβλεψης σύμφωνα με τις ανάλογες βάσεις δεδομένων.

**Πίνακας 2.15:** Annotation systems για WES με τις πληροφορίες που παρέχει το κάθε ένα (Salgado et al., 2016). Με πλαίσιο είναι το σύστημα που χρησιμοποιήθηκε στην παρούσα μελέτη.

	Annovar	SNPeff	Ensembl VEP	SeattleSeq	AnnTools	Oncotator	Vanno	Variant Annotation Tools
Availability	Command line	Command line	Command line Webservice Web	web	command line	Command line Web	Web	Command line
Variant quality	Yes	Yes	Yes	-	Yes	-	Yes	Yes
Variant localization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gene/transcript annotation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Genotype	Yes	Yes	Yes	Yes	Yes	-	Yes	Yes
Population frequency	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes
Impact at the RNA level	Yes	Yes	Yes	-	-	-	-	-
Impact at the protein level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conservation	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes
Reported impact	-	-	Yes	Yes	-	Yes	Yes	Yes
Predicted pathogenicity	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes
Gene ontology	-	-	-	-	-	Yes	Yes	-
Pathways	-	-	-	Yes	-	-	Yes	Yes
Tissue expression	-	-	-	-	-	-	-	-

### Βήμα 8: Ερμηνεία παραλλαγών (Variant interpretation)

Ακολουθεί ταξινόμηση, φιλτράρισμα και κατηγοριοποίηση των παραλλαγών αυτών χρησιμοποιώντας γενετικά και λειτουργικά φίλτρα. Τα γενετικά φίλτρα περιλαμβάνουν σύγκριση με γονιδιωματικές βάσεις δεδομένων (1000G, dbSNP, HarMap κ.α.), με άτομα που έχουν παρόμοιο φαινότυπο, αλλά και με άτομα της ίδιας οικογενείας. Τα λειτουργικά φίλτρα περιλαμβάνουν προσδιορισμό γενωμικής ή εξονικής θέσης των παραλλαγών, πρόβλεψη για

τη δομή, τη λειτουργία, την επίπτωση στην πρωτεΐνη, τη θέση ματίσματος και αξιολόγηση με βάση τη συχνότητα στον πληθυσμό. Για τις πληροφορίες αυτές αντλούνται δεδομένα από βάσεις δεδομένων που παρουσιάζονται στον πίνακα 1.2 (Moorthie et al., 2013).

Το σύστημα Sophia DDM<sup>®</sup> είναι ένα αυτοπατοποιημένο σύστημα ανάλυσης το οποίο παρέχεται από την εταιρεία Sophia Genetics<sup>®</sup> και προσφέρει πλήρες πακέτο βιοπληροφορικής ανάλυσης από τη στοίχιση/χαρτογράφηση μέχρι και το variant interpretation.

**Πίνακας 2.16:** Filtration tools για WES με αυτοματοποιημένο σύστημα (Salgado et al., 2016). Με πλαίσιο είναι το σύστημα που χρησιμοποιήθηκε στην παρούσα μελέτη.

Software name	Availability	Mode of inheritance	Custom analysis	Mutation localization	Mutation type	Mutation frequency	Pathogenicity predictions	Functional evidences	Clinical report	Prioritization score
ExomeWalker	Web App	Yes	-	-	-	Yes	No but provided	No but provided	Yes	Yes
Exomiser	Command line	Yes	-	-	-	Yes	Yes	-	Yes	Yes
eXtasy	Command line Web App	-	-	-	-	-	No but provided	No but provided	-	Yes
MirTRIOS	Web App	Yes	-	Yes	Yes	Yes	Yes	-	No - But provided	Yes
OMIM Explorer	Web App	Yes	-	-	-	-	-	-	Yes	Yes
Sophia DDM	Web App	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
wKGGSeq	Web App	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Το VarAFT είναι ένα ημιαυτόνομο σύστημα ανάλυσης, το οποίο παρέχει σχολιασμό των παραλλαγών (Variant annotation) και φιλτράρισμα. Επίσης, δίνει τη δυνατότητα για συνδυασμό μεμονωμένων δειγμάτων, trio analysis, ταξινόμηση παραλλαγών μέσω χρήσης πολλαπλών φίλτρων και παρουσίαση της κάλυψης των εξονίων (coverage analysis) από τα BAM αρχεία.

**Πίνακας 2.17:** Filtration tools για WES με ημι-αυτοματοποιημένο ή χειροκίνητο σύστημα (Salgado et al., 2016). Με πλαίσιο είναι το σύστημα που χρησιμοποιήθηκε στην παρούσα μελέτη.

Software name	Availability	Mode of inheritance	Custom analysis	Mutation localization	Mutation type	Mutation frequency	Pathogenicity predictions	Functional evidences	Clinical report
ANNOVAR	Command line	-	-	-	-	Yes	Yes	-	Yes
BIERapp	Web App	Yes	Yes	Yes	Yes	Yes	-	-	-
FILTUS	GUI	Yes	Yes	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided
FMFilter	GUI	Yes	-	Yes - if provided	Yes - if provided	Yes - if provided	-	-	-
Gemini	Command line Web App	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vanno	Web App	-	-	Yes	Yes	Yes	No but provided	Yes	Yes
VarAFT	GUI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No - But provided
VarSifter	Command line Web App	Yes	Yes	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided
VCF-MINER	Local Web App	Yes	Yes	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided	Yes - if provided

# **Κεφάλαιο 3:**

## ***Αποτελέσματα***

### **3.1) Αποτελέσματα της Αλληλούχησης Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES)**

Η ανάλυση των αποτελεσμάτων των προς εξέταση δειγμάτων πραγματοποιήθηκε με δύο διαφορετικά βιοπληροφορικά συστήματα ανάλυσης (VarAFT και Sophia DDM<sup>®</sup>). Σύμφωνα με το Εδάφιο 2.1.2, στην πρώτη ροή εργασίας, η οποία περιλαμβάνει προετοιμασία βιβλιοθήκης με TruSeq Exome Library Prep Kit της Illumina και αλληλούχηση με σύστημα NextSeq 500 της Illumina πραγματοποιήθηκε βιοπληροφορική ανάλυση με υπολογιστικά συστήματα που διατίθενται στον ακαδημαϊκό χώρο (Bowtie2, Samtools, VarAFT), σύμφωνα με το διάγραμμα της Εικόνας 2.12. Στη δεύτερη ροή εργασίας πραγματοποιήθηκε βιοπληροφορική ανάλυση των δεδομένων (FastQ αρχεία) της πρώτης ροής εργασίας με το αυτοματοποιημένο σύστημα Sophia DDM<sup>®</sup> της Sophia Genetics<sup>®</sup>. Στην τρίτη ροή εργασίας, η οποία περιλαμβάνει προετοιμασία της βιβλιοθήκης και αλληλούχηση από την εταιρεία Sophia Genetics<sup>®</sup>, πραγματοποιήθηκε βιοπληροφορική ανάλυση με το αυτοματοποιημένο σύστημα Sophia DDM<sup>®</sup>.

### **3.2) Προετοιμασία βιβλιοθήκης με TruSeq Exome Library Prep Kit της Illumina**

Για την προετοιμασία της βιβλιοθήκης απομονώθηκε DNA από τα λεμφοκύτταρα περιφερικού αίματος με χρήση του πρωτόκολλο MagAttract DNA Blood M48 Kit και ρομποτικού συστήματος BioRobot M48 της QIAGEN (Εδάφιο 2.2) και πραγματοποιήθηκε προσδιορισμός της συγκέντρωσης του DNA (Εδάφιο 2.3) με τις μετρήσεις να παρουσιάζονται στον Πίνακα 3.1. Ακολούθησε το πρωτόκολλο για την κατασκευή της βιβλιοθήκης όπως αναφέρεται στο Εδάφιο 2.5. Επίσης, πραγματοποιήθηκε ποσοτικός και ποιοτικός έλεγχος των θραυσμάτων της βιβλιοθήκης και της εμπλουτισμένης βιβλιοθήκης (pool library) (Εδάφιο 2.3, 2.4).

#### **3.2.1) Ποσοτικός έλεγχος DNA**

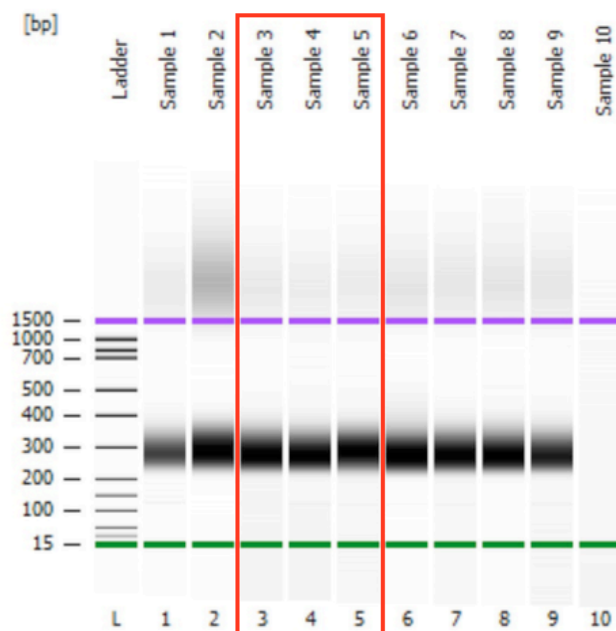
Ο ποσοτικός έλεγχος πραγματοποιήθηκε με αναλυτή Qubit Fluorimeter της Invitrogen. Για την κατασκευή της βιβλιοθήκης απαιτούνται 100ng αρχικής ποσότητας DNA από κάθε δείγμα, ενώ για το pool library απαιτούνται 150ng DNA από κάθε βιβλιοθήκη δείγματος.

**Πίνακας 3.1:** Συγκέντρωση DNA μετά την απομόνωση από λεμφοκύτταρα του περιφερικού αίματος (Στήλη 2), της βιβλιοθήκης κάθε δείγματος (Στήλη 3) και της εμπλουτισμένης βιβλιοθήκης (pool) (Στήλη 4).

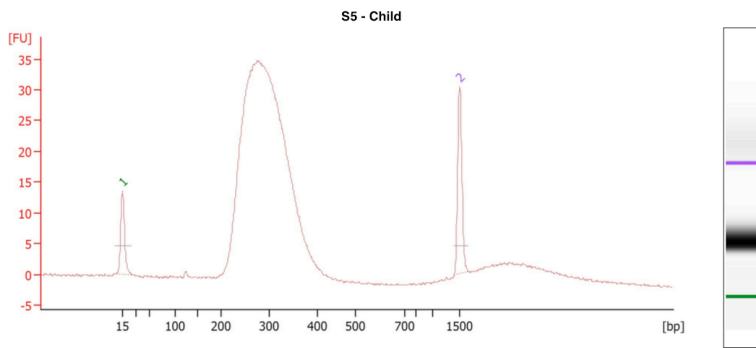
Δείγμα	Συγκέντρωση DNA (μετά την απομόνωση)	Συγκέντρωση DNA βιβλιοθηκών	Συγκέντρωση DNA εμπλουτισμένης βιβλιοθήκης (pool)
Δείγμα 3 - S5 (Child)	47,5 ng/μl	40,20 ng/μl	8,05 ng/μl
Δείγμα 4 - S6 (Father)	55 ng/μl	35,40 ng/μl	
Δείγμα 5 - S7 (Mother)	51 ng/μl	44,60 ng/μl	

### 3.2.2) Ποιοτικός έλεγχος DNA

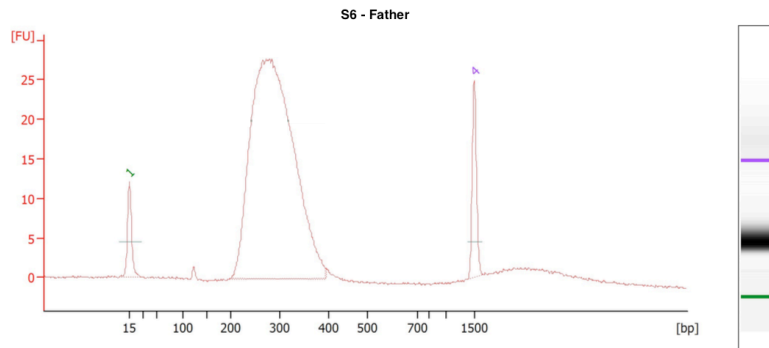
Ο ποιοτικός έλεγχος των βιβλιοθηκών των δειγμάτων πραγματοποιήθηκε με Bioanalyzer 2100 της Agilent (Εδάφιο 2.4).



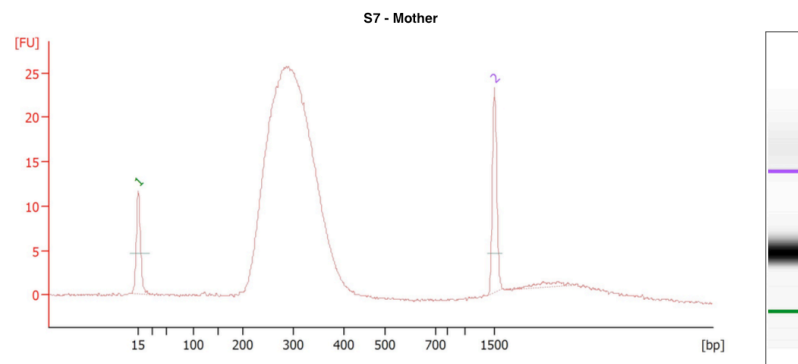
**Εικόνα 3.2:** Ηλεκτροφόρηση των βιβλιοθηκών. Η στήλη L παρουσιάζει τον μοριακό μάρτυρα (bp). Η στήλη 3 παρουσιάζει τη βιβλιοθήκη του δείγματος 3-S5 (Child). Η στήλη 4 παρουσιάζει τη βιβλιοθήκη του δείγματος 4-S6 (Father). Η στήλη 5 παρουσιάζει τη βιβλιοθήκη του δείγματος 5-S7 (Mother). Η πράσινη ζώνη παρουσιάζει τον κατώτερο μοριακό μάρτυρα και η μωβ τον ανώτερο μοριακό μάρτυρα.



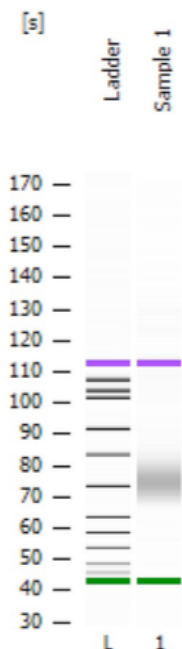
**Εικόνα 3.3:** Ηλεκτροφόρημα της βιβλιοθήκης του δείγματος 3-S5 (Child).



**Εικόνα 3.4:** Ηλεκτροφόρημα της βιβλιοθήκης του δείγματος 4-S6 (Father).

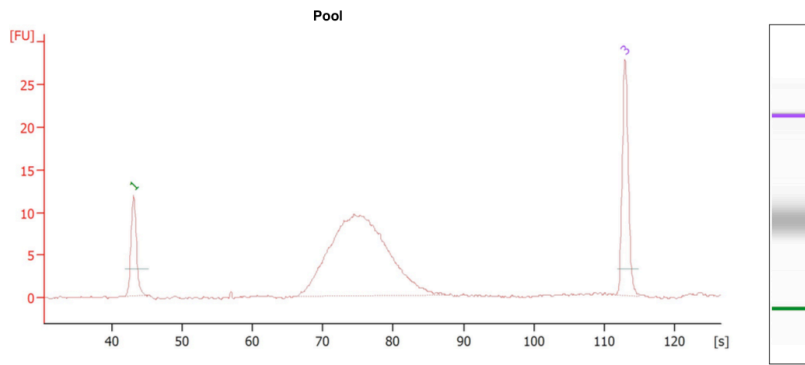


**Εικόνα 3.5:** Ηλεκτροφόρημα της βιβλιοθήκης του δείγματος 5-S7 (Mother).



**Εικόνα 3.6:** Ηλεκτροφόρηση της εμπλουτισμένης συνολικής βιβλιοθήκης (pool library). Η στήλη L παρουσιάζει τον μοριακό μάρτυρα (s-size). Η στήλη 1 παρουσιάζει το pool library.





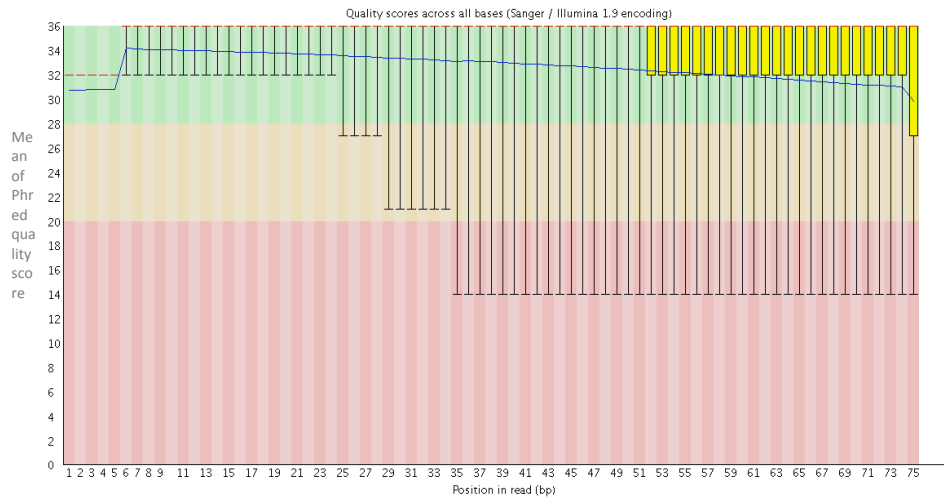
**Εικόνα 3.7:** Ηλεκτροφόρημα της εμπλουτισμένης συνολικής βιβλιοθήκης (pool library).

### 3.3) Βιοπληροφορική ανάλυση των δεδομένων αλληλούχησης επόμενης γενιάς

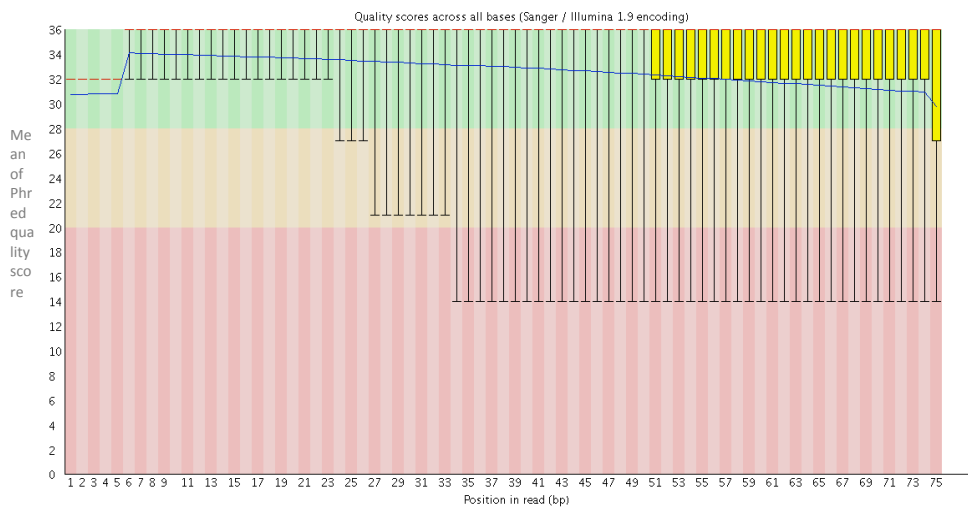
Η αντίδραση αλληλούχησης πραγματοποιήθηκε με σύστημα NextSeq 500 της Illumina για την πρώτη και δεύτερη ροή εργασίας και στο σύστημα της Sophia Genetics® για την τρίτη ροή εργασίας.

#### 3.3.1) Ποιοτικός έλεγχος της κλήσης βάσεων (Base calling)

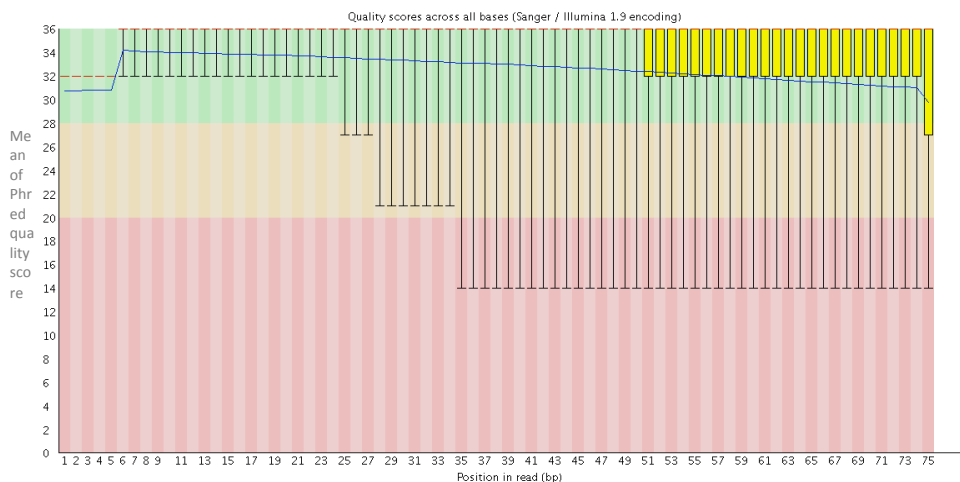
Ο έλεγχος ποιότητας του Base calling πραγματοποιείται με την εφαρμογή FastQC (Εδάφιο 2.7.1), χρησιμοποιώντας τα FastQ ή τα BAM αρχεία από κάθε δείγμα. Οι Εικόνες 3.8, 3.9, 3.10 παρουσιάζουν τα αποτελέσματα από την πρώτη ροή εργασίας και οι Εικόνες 3.11, 3.12, 3.13 παρουσιάζουν τα αποτελέσματα από την Τρίτη ροή εργασίας. Για κάθε θέση σχεδιάζεται ένα BoxWhisker type plot. Η κεντρική κόκκινη γραμμή είναι η μέση τιμή. Το κίτρινο κουτί παριστά το ενδοτεταρτομοριακό εύρος (25-75%). Το ανώτερο και κατώτερο μουστάκι παριστά το 10% και το 90% αντίστοιχα. Η μπλε γραμμή παριστά τη μέση ποιότητα. Όσο υψηλότερη είναι η βαθμολογία τόσο καλύτερο είναι το base calling. Το φόντο του γραφήματος διαιρεί τον άξονα y σε τρεις ζώνες, την πράσινη ζώνη για πολύ καλή ποιότητα, την πορτοκαλί ζώνη για λογική ποιότητα και την κόκκινη ζώνη για χαμηλή ποιότητα. Η ποιότητα των base calling στις περισσότερες πλατφόρμες μειώνεται καθώς το τρέξιμο εξελίσσεται, επομένως παρατηρείται πτώση στην πορτοκαλί περιοχή καθώς τελειώνει η αντίδραση.



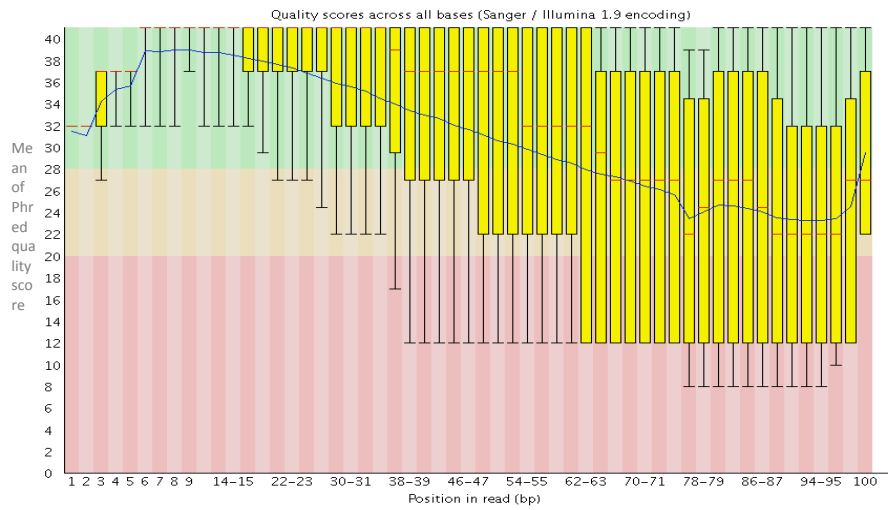
**Εικόνα 3.8:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S5 (Child) με την 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας.



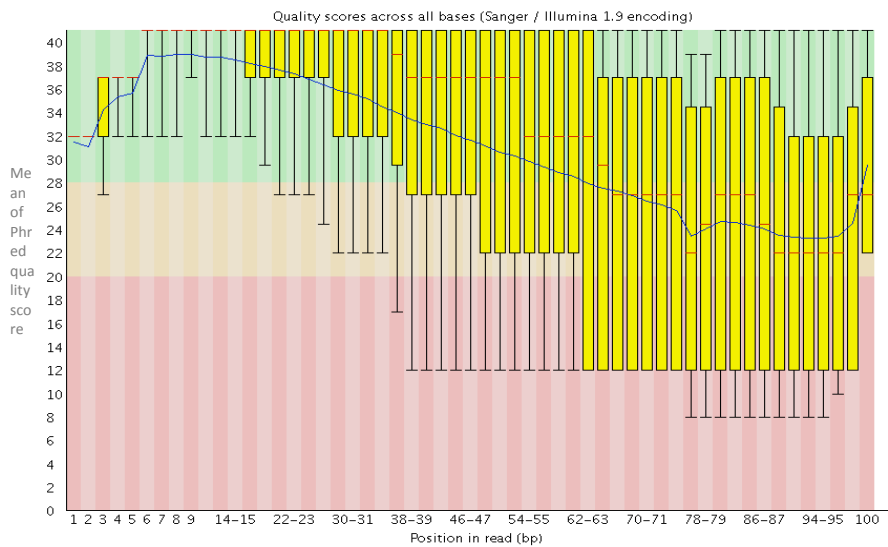
**Εικόνα 3.9:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S6 (Father) με την 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας.



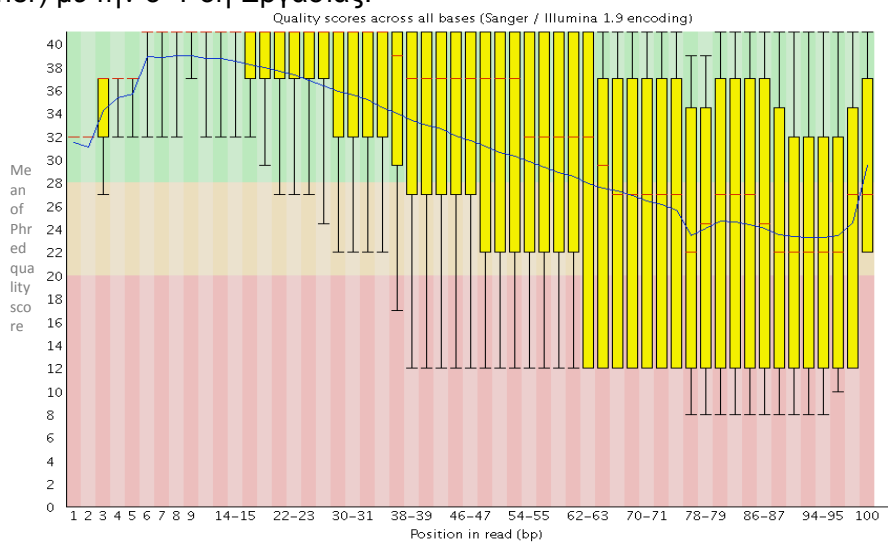
**Εικόνα 3.10:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S7 (Mother) με την 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας.



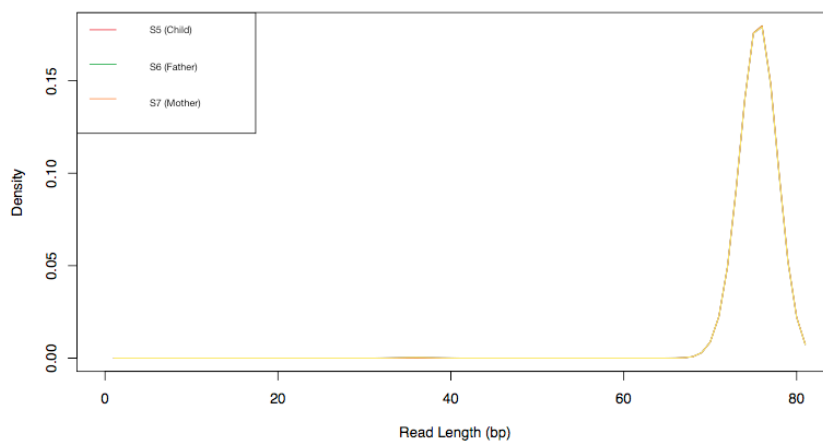
**Εικόνα 3.11:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S5 (Child) με την 3<sup>η</sup> Ροή Εργασίας.



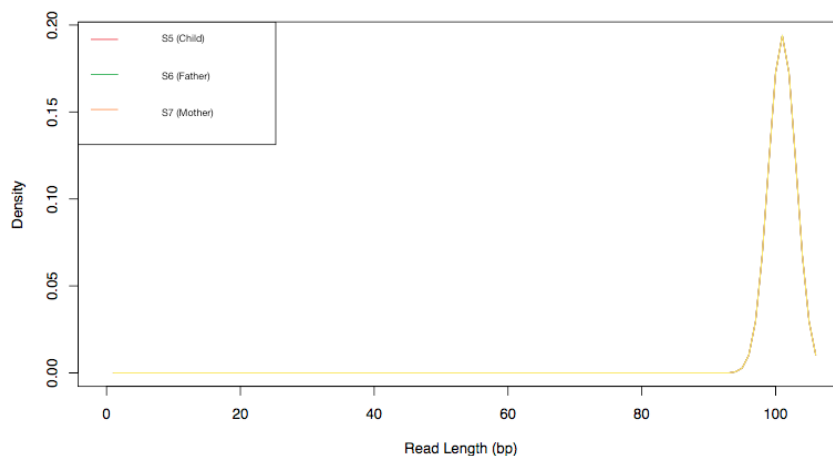
**Εικόνα 3.12:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S6 (Father) με την 3<sup>η</sup> Ροή Εργασίας.



**Εικόνα 3.13:** Μέση βαθμολογία ποιότητας base calling (phred score) χρησιμοποιώντας την εφαρμογή FastQC για τα FastQ αρχείο του δείγματος S7 (Mother) με την 3<sup>η</sup> Ροή Εργασίας.



**Εικόνα 3.14:** Κατανομή μήκους αναγνώσεων (reads) που εξάγονται από την πλατφόρμα αλληλούχησης. Παρουσιάζονται τα δείγματα S5, S6 και S7 με την 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας.



**Εικόνα 3.15:** Κατανομή μήκους αναγνώσεων (reads) που εξάγονται από την πλατφόρμα αλληλούχησης. Παρουσιάζονται τα δείγματα S5, S6 και S7 με την 3<sup>η</sup> Ροή Εργασίας.

### 3.3.2) Ποιοτικός έλεγχος χαρτογράφησης/στοίχισης

Η χαρτογράφηση των δειγμάτων με την 1<sup>η</sup> Ροή Εργασίας έγινε με τον αλγόριθμο Bowtie2, ενώ στη 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας έγινε με τον αλγόριθμο BWA. Ο ποιοτικός έλεγχος της χαρτογράφησης/στοίχισης πραγματοποιήθηκε με τα βιοπληροφορικά εργαλεία Samtools και QualiMap για την 1<sup>η</sup> Ροή Εργασίας, και με την πλατφόρμα Sophia DDM<sup>®</sup> για τη 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας.

**Πίνακας 3.2:** Συγκριτικός πίνακας του αριθμού των αναγνώσεων (reads) για κάθε δείγμα με τις τρεις διαφορετικές Ροές Εργασίας.

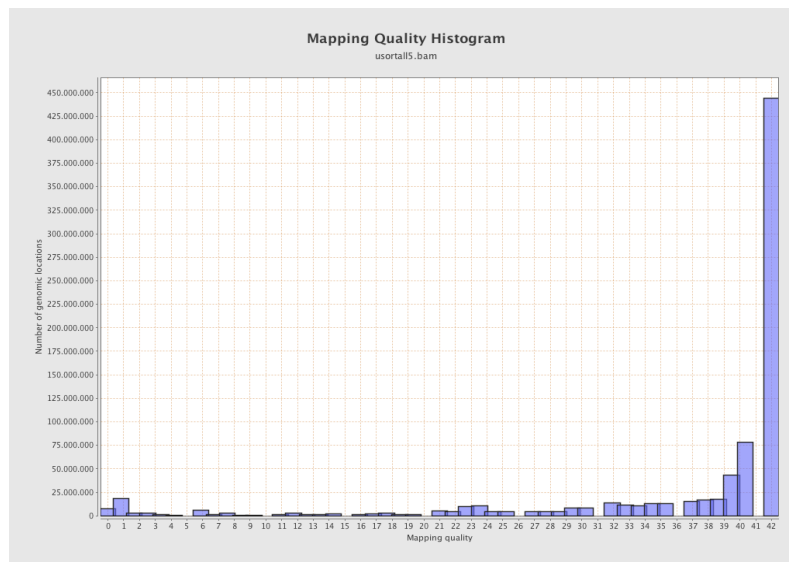
Αριθμός αναγνώσεων (reads)

Δείγμα	1η Ροή Εργασίας	2η Ροή Εργασίας	3η Ροή Εργασίας
<b>S5 (Child)</b>	81.606.989	121.978.642	214.060.508
<b>S6 (Father)</b>	77.058.581	113.610.292	220.171.050
<b>S7 (Mother)</b>	77.003.587	112.595.628	217.952.598

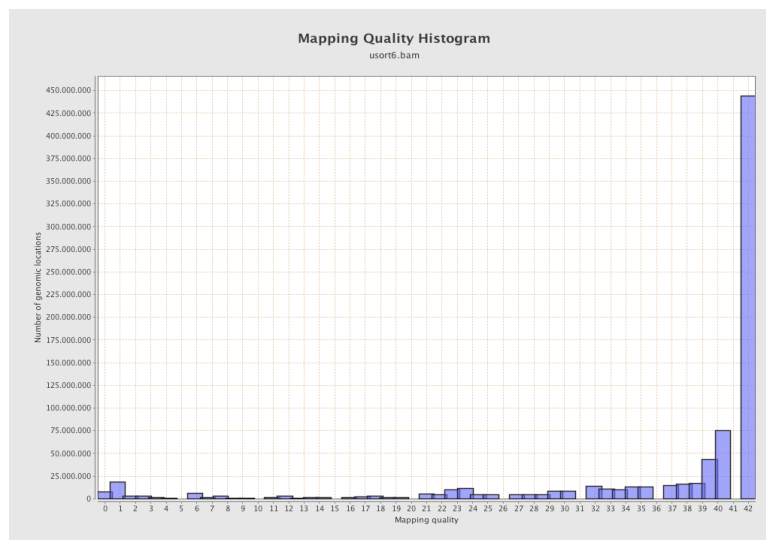
**Πίνακας 3.3:** Συγκριτικός πίνακας του αριθμού των στοιχισμένων αναγνώσεων (reads) στην αλληλουχία αναφοράς GRCh19 για κάθε δείγμα, με τις τρεις διαφορετικές Ροές Εργασίας.

*Αριθμός στοιχισμένων αναγνώσεων (Mapped reads)*

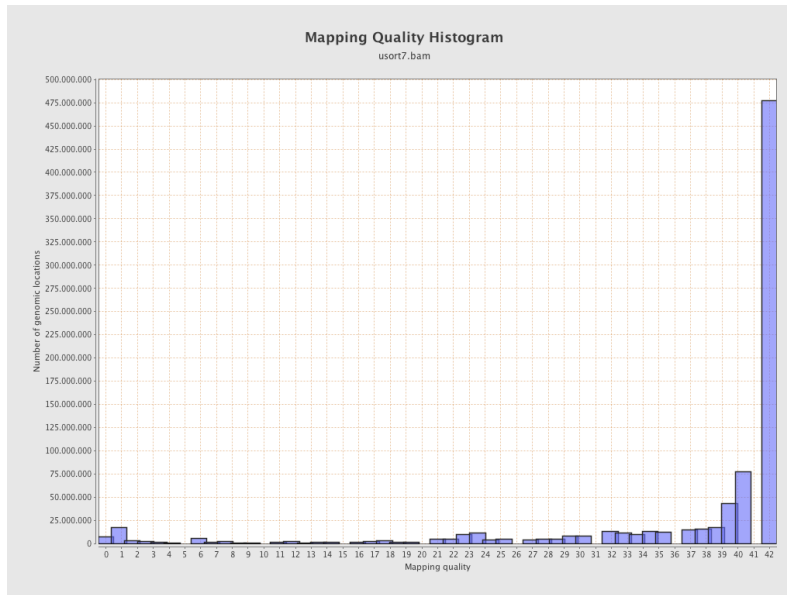
Δείγμα	1η Ροή Εργασίας		2η Ροή Εργασίας		3η Ροή Εργασίας	
<b>S5 (Child)</b>	76.279.976	93,47%	116.790.475	95,75%	208.675.458	97,48%
<b>S6 (Father)</b>	72.019.178	93,46%	108.736.090	95,71%	213.763.172	97,09%
<b>S7 (Mother)</b>	72.187.422	93,75%	108.026.756	95,94%	212.457.670	97,49%



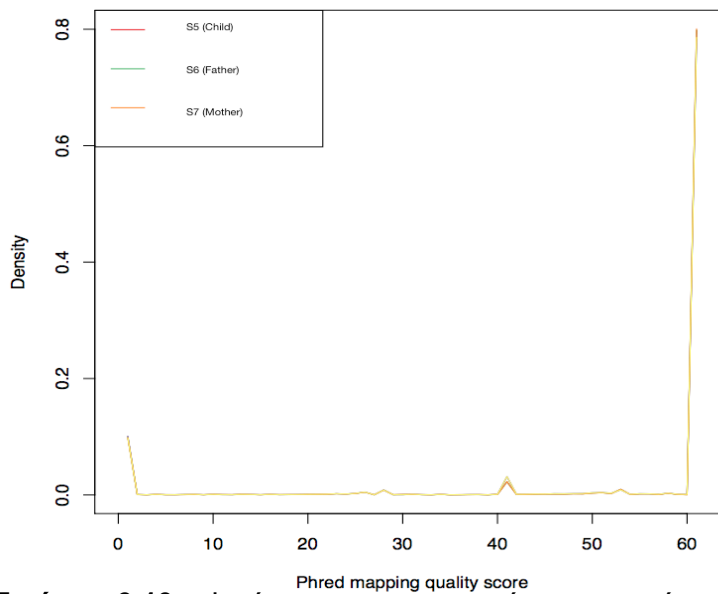
**Εικόνα 3.16:** Ιστόγραμμα της ποιότητας στοίχισης του δείγματος S5 με το πρόγραμμα QualiMap (1<sup>η</sup> Ροή Εργασίας). Οι περισσότερες αντιστοιχίσεις πρέπει να είναι υψηλής ποιότητας με αποτέλεσμα να εμφανίζεται μια κορυφή γύρω στα 40 Phred score. Η μέση ποιότητα στοίχισης είναι 34,68.



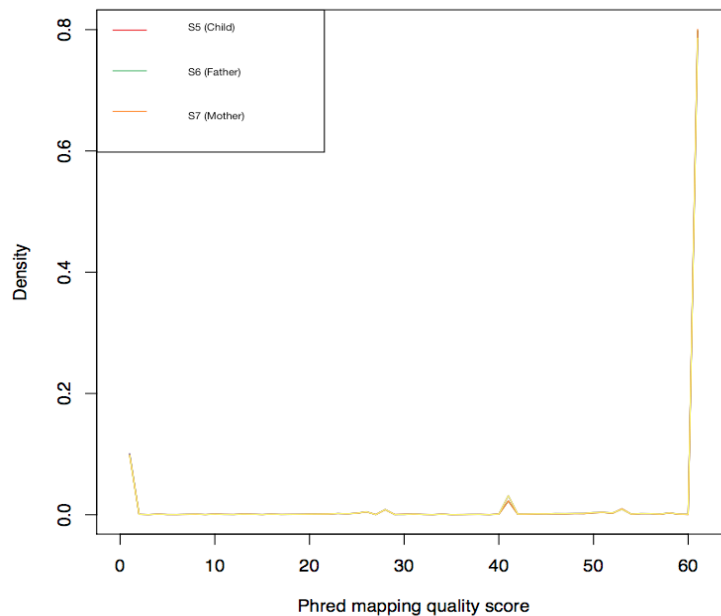
**Εικόνα 3.17:** Ιστόγραμμα της ποιότητας στοίχισης του δείγματος S6 με το πρόγραμμα QualiMap (1<sup>η</sup> Ροή Εργασίας). Οι περισσότερες αντιστοιχίσεις πρέπει να είναι υψηλής ποιότητας με αποτέλεσμα να εμφανίζεται μια κορυφή γύρω στα 40 Phred score. Η μέση ποιότητα στοίχισης είναι 34,61.



**Εικόνα 3.18:** Ιστόγραμμα της ποιότητας στοίχισης του δείγματος S7 με το πρόγραμμα QualiMap (1<sup>η</sup> Ροή Εργασίας). Οι περισσότερες αντιστοιχίσεις πρέπει να είναι υψηλής ποιότητας με αποτέλεσμα να εμφανίζεται μια κορυφή γύρω από τα 40 Phred score. Η μέση ποιότητα στοίχισης του δείγματος είναι 34,78.



**Εικόνα 3.19:** Ιστόγραμμα της ποιότητας στοίχισης των δειγμάτων S5, S6, S7 από την πλατφόρμα Sophia DDM® (2<sup>η</sup> Ροή Εργασίας). Οι περισσότερες αντιστοιχίσεις πρέπει να είναι υψηλής ποιότητας με αποτέλεσμα να εμφανίζεται μια κορυφή γύρω από τα 40 Phred score. Η μέση ποιότητα στοίχισης του δείγματος είναι 43,91.



**Εικόνα 3.20:** Ιστόγραμμα της ποιότητας στοίχισης των δειγμάτων S5, S6, S7 από την πλατφόρμα Sophia DDM<sup>®</sup> (3<sup>η</sup> Ροή Εργασίας). Οι περισσότερες αντιστοιχίσεις πρέπει να είναι υψηλής ποιότητας με αποτέλεσμα να εμφανίζεται μια κορυφή γύρω από τα 40 Phred score. Η μέση ποιότητα στοίχισης του δείγματος είναι 44,83.

**Πίνακας 3.4:** Ποσοστό των στοιχισμένων αναγνώσεων (reads) στις περιοχές-στόχους (Table Browser by UCSC) και ποσοστό των PCR duplicates για κάθε δείγμα με την 1<sup>η</sup> Ροή Εργασίας. Για τις παρακάτω μετρήσεις χρησιμοποιήθηκε η εφαρμογή QualiMap.

*Στατιστικά χαρτογράφησης (Εντός περιοχών-στόχου) – 1<sup>η</sup> Ροή Εργασίας*

Δείγμα	Mapped reads (on target)	Duplicated reads
<b>S5 (Child)</b>	38,38%	22,91%
<b>S6 (Father)</b>	38,42%	29,08%
<b>S7 (Mother)</b>	37,64%	21,58%

**Πίνακας 3.5:** Ποσοστό των στοιχισμένων αναγνώσεων (reads) στις περιοχές-στόχους και ποσοστό των PCR duplicates για κάθε δείγμα με τη 2<sup>η</sup> Ροή Εργασίας. Η πλατφόρμα Sophia DDM παρέχει ποσοστά και για τις πλευρικές περιοχές των target regions (flank target).

*Στατιστικά χαρτογράφησης (Εντός περιοχών-στόχου) – 2<sup>η</sup> Ροή Εργασίας*

Δείγμα	Mapped reads (on target)	Mapped reads (flank target)	Duplicated reads
<b>S5 (Child)</b>	36,26%	28,14%	8,25%
<b>S6 (Father)</b>	36,23%	27,81%	8,81%
<b>S7 (Mother)</b>	35,20%	28,45%	8,58%

**Πίνακας 3.6:** Ποσοστό των στοιχισμένων αναγνώσεων (reads) στις περιοχές- στόχους και ποσοστό των PCR duplicates για κάθε δείγμα με την 3<sup>η</sup> Ροή Εργασίας. Η πλατφόρμα Sophia DDM παρέχει ποσοστά και για τις πλευρικές περιοχές των target regions (flank target).

*Στατιστικά χαρτογράφησης (Εντός περιοχών-στόχου) – 3<sup>η</sup> Ροή Εργασίας*

Δείγμα	Mapped reads (on target)	Mapped reads (flank target)	Duplicated reads
S5 (Child)	62,65%	30,30%	19,27%
S6 (Father)	62,76%	30,32%	19,56%
S7 (Mother)	62,81%	29,79%	18,27%

Μια σημαντική προϋπόθεση για την αξιόπιστη ανίχνευση παραλλαγών από δεδομένα αλληλούχησης επόμενης γενιάς είναι η κάλυψη των στοχευμένων περιοχών. Η κάλυψη αυτή ορίζεται ως ο αριθμός των αναγνώσεων που καλύπτουν μια βάση. Η κάλυψη μιας περιοχής ορίζεται ως η μέση κάλυψη των βάσεων της.

**Πίνακας 3.7:** Ποσοστά περιοχών-στόχων για κάλυψη 25X, 50X και 100X των δειγμάτων S5, S6, S7 χρησιμοποιώντας την εφαρμογή QualiMap για την 1<sup>η</sup> Ροή Εργασίας. Η πέμπτη στήλη παρουσιάζει την κάλυψη για το 90% των περιοχών.

*Κάλυψη περιοχών-στόχων – 1<sup>η</sup> Ροή Εργασίας*

Δείγμα	25X	50X	100X	Coverage 10% quantile
S5 (Child)	85%	50%	25%	19X
S6 (Father)	82%	46%	22%	18X
S7 (Mother)	82%	44%	21%	17X

**Πίνακας 3.8:** Ποσοστά περιοχών-στόχων για κάλυψη 25X, 50X και 100X των δειγμάτων S5, S6, S7 μέσω της πλατφόρμας Sophia DDM για τη 2<sup>η</sup> Ροή Εργασίας. Η πέμπτη στήλη παρουσιάζει την κάλυψη για το 90% των περιοχών και η έκτη στήλη τις περιοχές με κάλυψη μικρότερη του 20X.

*Κάλυψη περιοχών-στόχων – 2<sup>η</sup> Ροή Εργασίας*

Δείγμα	25X	50X	100X	Coverage 10% quantile	Low coverage regions
S5 (Child)	88%	69%	32%	23X	89.625
S6 (Father)	87%	65%	28%	21X	101.625
S7 (Mother)	86%	64%	27%	20X	105.138

**Πίνακας 3.9:** Ποσοστά περιοχών-στόχων για κάλυψη 25X, 50X και 100X των δειγμάτων S5, S6, S7 μέσω της πλατφόρμας Sophia DDM για τη 3<sup>η</sup> Ροή Εργασίας. Η πέμπτη στήλη παρουσιάζει την κάλυψη για το 90% των περιοχών και η έκτη στήλη τις περιοχές με κάλυψη μικρότερη του 20X.

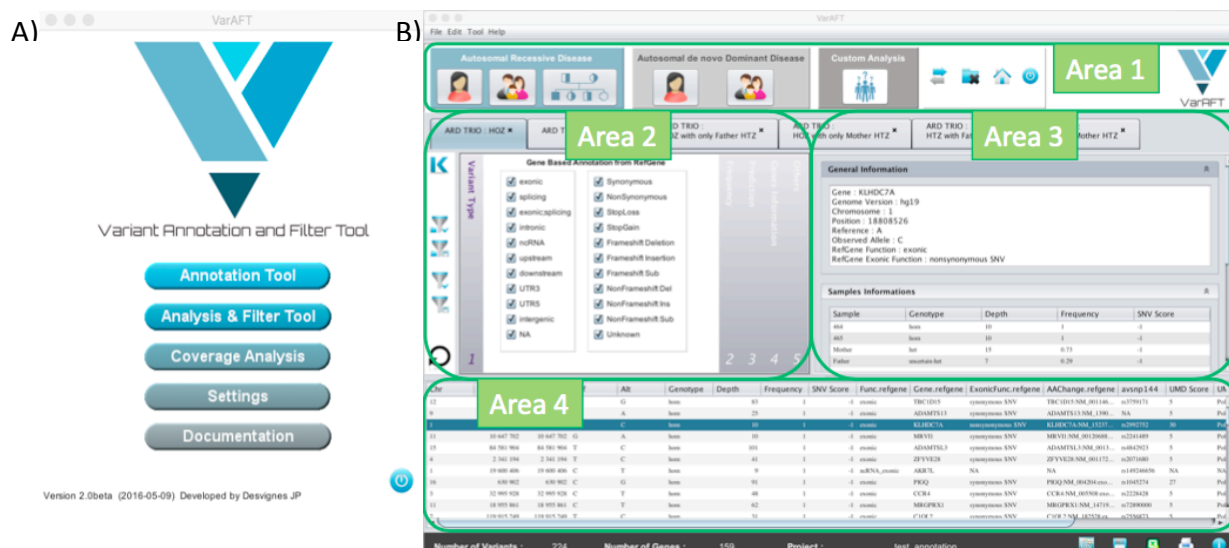
*Κάλυψη περιοχών-στόχων – 3<sup>η</sup> Ροή Εργασίας*

Δείγμα	25X	50X	100X	Coverage 10% quantile	Low coverage regions
S5 (Child)	99,7%	97%	83%	82X	1944
S6 (Father)	99,7%	98%	84%	84X	1550
S7 (Mother)	99,5%	98%	84%	85X	1726



### 3.3.3) Μελέτη παραλλαγών που προσδιορίστηκαν με τις τρεις Ροές Εργασίας

Ο σχολιασμός και ανάλυση των παραλλαγών με την 1<sup>η</sup> Ροή Εργασίας πραγματοποιήθηκε με την πλατφόρμα VarAFT (Εικόνα 3.21)



**Εικόνα 3.21:** Πλατφόρμα VarAFT. Α) Αρχική σελίδα με τα διαθέσιμα εργαλεία (Σχολιασμός παραλλαγών, Ανάλυση και Φιλτράρισμα παραλλαγών, Ανάλυση κάλυψης). Β) Σελίδα ανάλυσης και φίλτρων, Area 1: Περιοχή επιλογής συνδυασμών δειγμάτων (Αυτοσωμική υπολειπόμενη, Αυτοσωμική Επικρατής, Custom ανάλυση), Area 2: Περιοχή Φίλτρων (Τύπος παραλλαγών, Συχνότητα πληθυσμού, Πρόβλεψη, Πληροφορίες γονιδίων κ.α.), Area 3: Περιοχή πληροφοριών για κάθε παραλλαγή που έχει επιλεγεί, Area 4: Περιοχή αποτελεσμάτων σε μορφή πίνακα.

Τα αρχεία VCF (Variant Calling File) των δειγμάτων, σχολιάστηκαν (Variant Annotation) με βάση το γονιδίωμα αναφοράς GRCh37/hg19, την έκδοση 147dbSNP ([http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=606266489\\_VI4Mvke aPUlze0RyCNSSbQATTBf&g=snp147&hgTracksConfigPage=configure](http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=606266489_VI4Mvke aPUlze0RyCNSSbQATTBf&g=snp147&hgTracksConfigPage=configure)) και τις βάσεις δεδομένων όπως φαίνονται στην Εικόνα 3.22.

Η ανάλυση/φιλτράρισμα των δειγμάτων έγινε αρχικά επί του συνόλου των παραλλαγών που ανιχνεύθηκαν και στη συνέχεια με τη χρήση λίστας γονιδίων που έχουν συσχετιστεί με νευρομυϊκό φαινότυπο (gene list – bioinformatic panel). Η συγκεκριμένη λίστα αποτελείται από 504 γονίδια, τα οποία έχουν αναφερθεί στη βιβλιογραφία ή χρησιμοποιούνται ήδη σε αντίστοιχα πάνελ γονιδίων. Επίσης, η ανάλυση του δείγματος S5 (Child) έγινε ως μεμονωμένο

δείγμα και σε Trio ανάλυση συσχετιζόμενο με τους δύο γονείς. Η προσέγγιση με Trio ανάλυση είναι εξαιρετικά χρήσιμη διότι μελετάται η ύπαρξη αυτοσωμικού υπολειπόμενου γονιδίου με την ομόζυγη παραλλαγή του παιδιού να εμφανίζεται ως ετερόζυγη στους δύο γονείς (Πίνακας 1.1). Οι πίνακες που ακολουθούν προέκυψαν με την εφαρμογή φίλτρων όπως φαίνονται στην Εικόνα 3.22.

**Εικόνα 3.22:** Επιλεγμένες βάσεις δεδομένων που χρησιμοποιήθηκαν μέσω της πλατφόρμας VarAFT για το σχολιασμό των παραλλαγών με βάση τα γονίδια, πρόβλεψη και συχνότητα.

**Πίνακας 3.10:** Αριθμός παραλλαγών που αναλύθηκαν με την πλατφόρμα VarAFT για το δείγμα S5 (Child) με την 1<sup>η</sup> Ροή Εργασίας. Η ανάλυση έγινε ως μεμονωμένο δείγμα (Index) και ως Trio, με τη χρήση λίστας γονιδίων (panel) και χωρίς.

S5 (Child)	Index		Index - Panel		Trio		Trio - Panel	
	All	Pathog.	All	Pathog.	All	Pathog.	All	Pathog.
<b>Hom+Het</b>	34.102	322	1311	<b>17</b>	4.529	20	267	<b>16</b>
<b>Homozygous</b>	12.780	109	512	2	1302	9	57	2
<b>Heterozygous</b>	21.322	213	799	15	3.227	11	210	14

**Πίνακας 3.11:** Υποψήφιες παθολογικές παραλλαγές μετά από ανάλυση με την πλατφόρμα VarAFT για το δείγμα S5 (Child) με την 1<sup>η</sup> Ροή Εργασίας. Αφορούν 17 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως μεμονωμένο και εφαρμόζοντας το πάνελ γονιδίων.

Chr	Start	End	Ref	Alt	Func.refgene	Gene.refgene	ExonicFunc.refgene	Genotype	Depth	AAChange.refgene
1	236899904	236899904	C	-	exonic	ACTN2	frameshift deletion	het	43	NM_001278344:exon10:c.95delC;p.P32fs
1	100349983	100349983	C	T	exonic	AGL	nonsynonymous SNV	het	89	NM_000645:exon17:c.2471C>T;p.S824F
2	73675230	73675230	-	CTC	exonic	ALMS1	nonframeshift insertion	het	67	NM_015120:exon8:c.1573_1574insCTC;p.L525delinsPL
2	202598051	202598051	C	T	exonic	ALS2	nonsynonymous SNV	het	80	NM_020919:exon13:c.2528G>A;p.R843Q
2	241696871	241696873	tcc	-	exonic	KIF1A	nonframeshift deletion	het	25	NM_001244008:exon27:c.2721_2723del;p.907_908del
6	129722481	129722481	T	G	exonic	LAMA2	nonsynonymous SNV	het	101	NM_000426:exon38:c.5558T>G;p.I1853R
22	33673227	33673227	G	A	exonic	LARGE	nonsynonymous SNV	het	42	NM_133642:exon14:c.1892C>T;p.T631M
11	47371414	47371414	C	T	exonic	MYBPC3	nonsynonymous SNV	het	33	NM_000256:exon5:c.565G>A;p.V189I
2	152507189	152507189	C	G	exonic	NEB	nonsynonymous SNV	het	91	NM_001164507:exon53:c.7126G>C;p.V2376L
8	24811071	24811071	G	-	exonic	NEFL	frameshift deletion	hom	78	NM_006158:exon3:c.1408delC;p.P470fs
10	13323113	13323113	G	T	splicing	PHYH	NA	het	83	NM_001037537:exon7:c.529_3C.A
1	46656162	46656162	T	G	exonic	POMGNT1	nonsynonymous SNV	het	70	NM_001290129:exon18:c.1566A>C;p.E522D
19	39002998	39002998	C	T	exonic	RYR1	nonsynonymous SNV	het	86	NM_000540:exon63:c.9347C>T;p.S3116L
11	118898437	118898437	C	-	exonic;splicing	SLC37A4	frameshift deletion	hom	85	NM_001164280:exon3:c.527delG;p.C176fs
13	36900750	36900750	G	A	exonic	SPG20	nonsynonymous SNV	het	94	NM_001142294:exon5:c.1250C>T;p.P417L
6	152655372	152655372	C	T	exonic	SYNE1	nonsynonymous SNV	het	85	NM_033071:exon76:c.12352G>A;p.V4118M
14	68248167	68248167	C	G	exonic	ZFYVE26	nonsynonymous SNV	het	71	NM_015346:exon22:c.4452G>C;p.R1484S

**Πίνακας 3.12:** Υποψήφιες παθολογικές παραλλαγές μετά από ανάλυση με την πλατφόρμα VarAFT για το δείγμα S5 (Child) με την 1<sup>η</sup> Ροή Εργασίας. Αφορούν 16 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως Τριο (χρήση αρχείων S6 & S7) και εφαρμόζοντας το πάνελ γονιδίων.

Chr	Start	End	Ref	Alt	Func.refgene	Gene.refgene	ExonicFunc.refgene	Genotype	Depth	AChange.refgene	Inherited from
1	236899904	236899904	C	-	exonic	ACTN2	frameshift deletion	het	43	NM_001278344:exon10:c.95delC:p.P32fs	S7
1	100349983	100349983	C	T	exonic	AGL	nonsynonymous SNV	het	89	NM_000645:exon17:c.2471C>T:p.S824F	S7
2	73675230	73675230	-	CTC	exonic	ALMS1	nonframeshift insertion	het	67	NM_015120:exon8:c.1573_1574insCTC:p.L525delinsPL	S7
2	202598051	202598051	C	T	exonic	ALS2	nonsynonymous SNV	het	80	NM_020919:exon13:c.2528G>A:p.R843Q	S6
2	241696871	241696873	tcc	-	exonic	KIF1A	nonframeshift deletion	het	25	NM_001244008:exon27:c.2721_2723del:p.907_908del	S6
6	129722481	129722481	T	G	exonic	LAMA2	nonsynonymous SNV	het	101	NM_000426:exon38:c.5558T>G:p.I1853R	S6
22	33673227	33673227	G	A	exonic	LARGE	nonsynonymous SNV	het	42	NM_133642:exon14:c.1892C>T;p.T631M	S7
2	152507189	152507189	C	G	exonic	NEB	nonsynonymous SNV	het	91	NM_001164507:exon53:c.7126G>C:p.V2376L	S7
8	24811071	24811071	G	-	exonic	NEFL	frameshift deletion	hom	78	NM_006158:exon3:c.1408delC:p.P470fs	S6, S7
10	13323113	13323113	G	T	splicing	PHYH	NA	het	83	NM_001037537:exon7:c.529_3C.A	S6
1	46656162	46656162	T	G	exonic	POMGNT1	nonsynonymous SNV	het	70	NM_001290129:exon18:c.1566A>C:p.E522D	S6
19	39002998	39002998	C	T	exonic	RYR1	nonsynonymous SNV	het	86	NM_000540:exon63:c.9347C>T:p.S3116L	S7
11	118898437	118898437	C	-	exonic;splicing	SLC37A4	frameshift deletion	hom	85	NM_001164280:exon3:c.527delG:p.C176fs	S6, S7
13	36900750	36900750	G	A	exonic	SPG20	nonsynonymous SNV	het	94	NM_001142294:exon5:c.1250C>T:p.P417L	S7
6	152655372	152655372	C	T	exonic	SYNE1	nonsynonymous SNV	het	85	NM_033071:exon76:c.12352G>A:p.V4118M	S6
14	68248167	68248167	C	G	exonic	ZFYVE26	nonsynonymous SNV	het	71	NM_015346:exon22:c.4452G>C:p.R1484S	S6

The image displays five panels (A-E) from the VarAFT platform used for variant filtering:

- A) Gene Based Annotation from RefGene:** A panel with two columns of checkboxes. The left column includes 'exonic', 'splicing', 'exonic;splicing', 'intronic', 'ncRNA', 'upstream', 'downstream', 'UTR3', 'UTR5', 'intergenic', and 'NA'. The right column includes 'Synonymous', 'NonSynonymous', 'StopLoss', 'StopGain', 'Frameshift Deletion', 'Frameshift Insertion', 'Frameshift Sub', 'NonFrameshift Del', 'NonFrameshift Ins', 'NonFrameshift Sub', and 'Unknown'.
- B) Public databases:** A panel for selecting public databases (ESP6500, 1000G, ExAC, KaViar, HRC) with dropdown menus for frequency and checkboxes for inclusion/exclusion. It also includes a 'Local database' section with radio buttons for 'HOM', 'HET', and 'ALL'.
- Γ) Prediction:** A panel for various prediction tools. UMD Predictor includes 'Polymorphism', 'Probable Polymorphism', 'Probable Pathogenic', and 'Pathogenic'. SIFT includes 'Tolerate' and 'Damaging'. PolyPhen 2 HumDiv includes 'Benign', 'Possibly Damaging', and 'Probably Damaging'. PolyPhen 2 HumVar includes 'Benign', 'Possibly Damaging', and 'Probably Damaging'. Mutation Taster is also present.
- Δ) Genes Information:** A panel for gene-related filters including RVIS Score, GDI Score, LoFTool, and GHIS. It also features logos for OMIM, REACTOME, GO, and GTEx.
- E) Others:** A panel for additional filters like 'Position query' (Chr, Position), 'SNV Score', 'Clinvar', and 'COSMIC'. It includes buttons for 'Get Compound Heterozygous' and 'Get Genes with threshold'.

**Εικόνα 3.23:** Περιοχή φίλτρων στην πλατφόρμα VarAFT, τα οποία εφαρμόστηκαν κατά την ανάλυση και φιλτράρισμα των παραλλαγών του δείγματος S5 (Child). Α) Είδος παραλλαγής (exonic, splicing, frameshift etc.), Β) Συχνότητα (1000G, ExAC), Γ) Πρόβλεψη (UMD-Predictor, SIFT, PolyPhen, Provean, CADD etc.) Δ) Πληροφορίες γονιδίων (OMIM, GTEc etc.), Ε) Άλλα φίλτρα (ClinVar, SNV score etc.)

Ο σχολιασμός και ανάλυση των παραλλαγών της 2<sup>ης</sup> και 3<sup>ης</sup> Ροής Εργασίας πραγματοποιήθηκε με την πλατφόρμα Soria DDM. Οι συγκεκριμένες αναλύσεις έγιναν ως μεμονωμένα δείγματα και με τη χρήση της λίστας γονιδίων, όπως παραπάνω. Η πλατφόρμα αυτή χωρίζει τις παραλλαγές σε εναπομείναντες ισχυρά παθογενείς, δυνητικά παθογενείς, αβέβαιης σημασίας και ήπιες.

**Πίνακας 3.13:** Αριθμός παραλλαγών που αναλύθηκαν με την πλατφόρμα Soria DDM για το δείγμα S5 (Child) με την 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας. Η ανάλυση έγινε ως μεμονωμένο δείγμα (Index), με τη χρήση λίστας γονιδίων (panel) και χωρίς.

S5 (Child)	2 <sup>η</sup> Ροή Εργασίας		3 <sup>η</sup> Ροή Εργασίας	
	No Panel	Panel	No panel	Panel
<b>All</b>	100.334	3.757	149.445	5.462
<b>Retained</b>	32.228	1.589	30.370	985
<b>Highly Pathogenic</b>	298	<b>8</b>	389	<b>11</b>
<b>Potentially Pathog.</b>	1.197	<b>19</b>	1.199	<b>20</b>
<b>VUS</b>	24.923	1.410	22.465	793
<b>Likely Bening</b>	5.810	141	6.317	148

**Πίνακας 3.14:** Ισχυρά παθολογικές υποψήφιες παραλλαγές μετά από ανάλυση με την πλατφόρμα Soria DDM για το δείγμα S5 (Child) με την 2<sup>η</sup> Ροή Εργασίας. Αφορούν 8 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως μεμονωμένο και εφαρμόζοντας το πάνελ γονιδίων.

chr	genome_position	ref	alt	codingConsequence	gene	type	depth	var_percent	exon_rank	c.DNA	protein	refSeqId
11	111742145	CG	C	splice_acceptor_cds_indel	ALG9	INDEL	47	91,49/HOM	2	c.61_1delC	p.?	NM_001077690
X	31893307	T	G	splice_donor +1	DMD	SNP	54	100/HOM	48	c.7071-1A.C	p.?	NM_000109
8	24811064	AGGGGGG	G	AGGGGGG splice_acceptor_cds_indel	NEFL	INDEL	184	82,07/HOM	4	c.1414_1delC	p.?	NM_006158
2	69659126	A	T	no_start	NFU1	SNP	140	99,29/HOM	2	c.2T.A	p.Met1?	NM_015700
12	122064773	CCCGCCA	C	inframe_6	ORAI1	INDEL	25	92/HOM	1	c.138_143delACCGCC	p.Pro48_Pro49del	NM_032790
3	38645420	T	C	missense	SCN5A	SNP	38	39,47/HET	12	c.1673A.G	p.His558Arg	NM_000335
11	118898435	ACC	AC	splice_acceptor_cds_indel	SLC37A4	INDEL	181	92,27/HOM	6	c.528_1delG	p.?	NM_001164277
1	17985844	G	A	splice_acceptor -1	TOR1AIP1	SNP	70	57,14/HET	3	c.554_1G.A	p.?	NM_001267578

**Πίνακας 3.15:** Δυνητικά παθολογικές υποψήφιες παραλλαγές μετά από ανάλυση με την πλατφόρμα Soria DDM για το δείγμα S5 (Child) με την 2<sup>η</sup> Ροή Εργασίας. Αφορούν 19 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως μεμονωμένο και εφαρμόζοντας το πάνελ γονιδίων.

chr	genome_position	ref	alt	codingConsequence	gene	type	depth	var_percent	exon_rank	c.DNA	protein	refSeqId
1	100349983	C	T	missense	AGL	SNP	164	49,39/HET	19	c.2522C.T	p.Ser841Phe	NM_000642
1	986732	G	A	missense	AGRN	SNP	35	42,86/HET	31	c.5353G.A	p.Asp1785Asn	NM_198576
2	202598051	C	T	missense	ALS2	SNP	160	50,62/HET	13	c.2528G.A	p.Arg843Gln	NM_020919
2	241696840	ATCC	A	inframe_3	KIF1A	INDEL	72	19,44/HET	27	c.2751_2753delGGA	p.Glu917del	NM_001244008
6	129722481	T	G	missense	LAMA2	SNP	205	56,1/HET	38	c.5558T.G	p.Ile1853Arg	NM_000426
22	33673227	G	A	missense	LARGE	SNP	61	50,82/HET	15	c.1892C.T	p.Thr631Met	NM_004737
11	47371414	C	T	missense	MYBPC3	SNP	41	39,02/HET	5	c.565G.A	p.Val189Ile	NM_000256
19	50784953	C	T	missense	MYH14	SNP	25	56/HET	32	c.4294C.T	p.Arg1432Cys	NM_001077186
2	152507189	C	G	missense	NEB	SNP	169	50,3/HET	53	c.7126G.C	p.Val2376Leu	NM_001164507
1	46656162	T	G	missense	POMGNT1	SNP	126	46,03/HET	19	c.1632A.C	p.Glu544Asp	NM_001243766
19	39002998	C	T	missense	RYR1	SNP	132	51,52/HET	63	c.9347C.T	p.Ser3116Leu	NM_000540
5	236649	C	T	missense	SDHA	SNP	175	22,29/HET	10	c.1367C.T	p.Ser456Leu	NM_004168
11	118897348	C	T	missense	SLC37A4	SNP	40	47,5/HET	8	c.835G.A	p.Ala279Thr	NM_001164277
13	36900750	G	A	missense	SPG20	SNP	174	54,02/HET	5	c.1250C.T	p.Pro417Leu	NM_001142294
6	152655372	C	T	missense	SYNE1	SNP	153	60,78/HET	76	c.12352G.A	p.Val4118Met	NM_033071
4	184615166	T	C	missense	TRAPP11	SNP	72	45,83/HET	22	c.2483T>C	p.Val828Ala	NM_021942
2	179472223	A	G	missense	TTN	SNP	184	50/HET	227	c.48269T.C	p.Ile16090Thr	NM_001256850
22	19865926	G	A	missense	TXNRD2	SNP	134	58,21/HET	15	c.1310C.T	p.Thr437Met	NM_006440
14	68248167	C	G	missense	ZFYVE26	SNP	116	41,38/HET	22	c.4452G.C	p.Arg1484Ser	NM_015346

**Πίνακας 3.16:** Ισχυρά παθολογικές υποψήφιες παραλλαγές μετά από ανάλυση με την πλατφόρμα Soria DDM για το δείγμα S5 (Child) με την 3<sup>η</sup> Ροή Εργασίας. Αφορούν 11 που εξάχθηκαν με την ανάλυση του δείγματος ως μεμονωμένο και εφαρμόζοντας το πάνελ γονιδίων.

chr	genome_position	ref	alt	codingConsequence	gene	type	depth	var_percent	exon_rank	c.DNA	protein	refSeqId
11	111742145	CG	C	splice_acceptor_cds_indel	ALG9	INDEL	276	95,29/HOM	2	c.61_1delC	p.?	NM_001077690
X	31893307	T	G	splice_donor_+1	DMD	SNP	40	100/HOM	48	c.7071-1A.C	p.?	NM_000109
6	7542148	CA	CAA	splice_acceptor_cds_indel	DSP	INDEL	196	47,45/HET	1	c.1dupA	p.?	NM_001008844
7	128587351	CACTCTGCAGCCGCC	C	inframe_30	IRF5	INDEL	240	47/HET	6	c.572_601del	p.Arg191_Leu200del	NM_001098629
8	24811064	AGGGGGGG	AGGGGGGG	splice_acceptor_cds_indel	NEFL	INDEL	248	84,27/HOM	4	c.1414_1delC	p.?	NM_006158
2	69659126	A	T	no_start	NFU1	SNP	65	100/HOM	2	c.2T.A	p.Met1?	NM_015700
12	122064773	CCCGCCA	C	inframe_6	ORAI1	INDEL	195	77,44/HET	1	c.138_143delACCGCC	p.Pro48_Pro49del	NM_032790
3	38645420	T	C	missense	SCN5A	SNP	248	47,58/HET	12	c.1673A.G	p.His558Arg	NM_000335
11	118898435	ACC	AC	splice_acceptor_cds_indel	SLC37A4	INDEL	220	90,91/HOM	6	c.528_1delG	p.?	NM_001164277
1	179858444	G	A	splice_acceptor_-1	TOR1AIP1	SNP	63	44,44/HET	3	c.554_1G.A	p.?	NM_001267578
12	974308	TCCC	TCCC	frameshift	WNK1	INDEL	249	37,75/HET	9	c.2175dupC	p.Ile726Hisfs*45	NM_213655

**Πίνακας 3.17:** Δυνητικά παθολογικές υποψήφιες παραλλαγές μετά από ανάλυση με την πλατφόρμα Soria DDM για το δείγμα S5 (Child) με την 3<sup>η</sup> Ροή Εργασίας. Αφορούν 20 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως μεμονωμένο και εφαρμόζοντας το πάνελ γονιδίων.

chr	genome_position	ref	alt	codingConsequence	gene	type	depth	var_percent	exon_rank	c.DNA	protein	refSeqId
1	100349983	C	T	missense	AGL	SNP	116	49,14/HET	19	c.2522C.T	p.Ser841Phe	NM_000642
1	986732	G	A	missense	AGRN	SNP	174	47,13/HET	31	c.5353G.A	p.Asp1785Asn	NM_198576
2	202598051	C	T	missense	ALS2	SNP	91	51,65/HET	13	c.2528G.A	p.Arg843Gln	NM_020919
19	47258718	C	G	missense	FKRP	SNP	148	41,22/HET	4	c.11C.G	p.Thr4Ser	NM_001039885
2	241696840	ATCC	A	inframe_3	KIF1A	INDEL	199	19,1/HET	27	c.2751_2753delGGA	p.Glu917del	NM_001244008
6	129722481	T	G	missense	LAMA2	SNP	45	44,44/HET	38	c.5558T.G	p.Ile1853Arg	NM_000426
22	33673227	G	A	missense	LARGE	SNP	104	45,19/HET	15	c.1892C.T	p.Thr631Met	NM_004737
11	47371414	C	T	missense	MYBPC3	SNP	227	44,93/HET	5	c.565G.A	p.Val189Ile	NM_000256
19	50784953	C	T	missense	MYH14	SNP	243	44,44/HET	32	c.4294C.T	p.Arg1432Cys	NM_001077186
2	152507189	C	G	missense	NEB	SNP	126	48,41/HET	53	c.7126G.C	p.Val2376Leu	NM_001164507
1	46656162	T	G	missense	POMGNT1	SNP	49	42,86/HET	19	c.1632A.C	p.Glu544Asp	NM_001243766
19	39002998	C	T	missense	RYR1	SNP	226	48,23/HET	63	c.9347C.T	p.Ser3116Leu	NM_000540
11	118897348	C	T	missense	SLC37A4	SNP	91	47,25/HET	8	c.835G.A	p.Ala279Thr	NM_001164277
13	36900750	G	A	missense	SPG20	SNP	109	44,95/HET	5	c.1250C.T	p.Pro417Leu	NM_001142294
6	152655372	C	T	missense	SYNE1	SNP	145	44,83/HET	76	c.12352G.A	p.Val4118Met	NM_033071
4	184615166	T	C	missense	TRAPPC11	SNP	108	46,3/HET	22	c.2483T>C	p.Val828Ala	NM_021942
2	179472223	A	G	missense	TTN	SNP	246	48,78/HET	227	c.48269T.C	p.Ile16090Thr	NM_001256850
2	179441038	C	T	missense	TTN	SNP	216	44,44/HET	276	c.64898G.A	p.Gly21633Asp	NM_001256850
22	19865926	G	A	missense	TXNRD2	SNP	77	45,45/HET	15	c.1310C.T	p.Thr437Met	NM_006440
14	68248167	C	G	missense	ZFYVE26	SNP	153	43,14/HET	22	c.4452G.C	p.Arg1484Ser	NM_015346

Τα αρχεία VCF από την πλατφόρμα Soria DDM αναλύθηκαν και μελετήθηκαν επίσης, με τη χρήση της πλατφόρμας VarAFT. Αρχικά πραγματοποιήθηκε σχολιασμός των παραλλαγών (variant annotation) με βάση το γονιδίωμα αναφοράς (hg19), την έκδοση 147dbSNP και τις βάσεις δεδομένων όπως φαίνονται στην Εικόνα 3.22. Στη συνέχεια, οι παραλλαγές αναλύθηκαν/φιλτραρίστηκαν χρησιμοποιώντας το πάνελ γονιδίων και έγινε σύγκριση των αποτελεσμάτων.

**Πίνακας 3.18:** Υποψήφιες παθολογικές παραλλαγές μετά από ανάλυση του VCF αρχείου του δείγμα S5 (Child) της 2<sup>ης</sup> Ροής Εργασίας με την πλατφόρμα VarAFT. Αφορούν 17 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως Τριο (χρήση αρχείων S6 & S7) και εφαρμόζοντας το πάνελ γονιδίων.

Chr	Start	End	Ref	Alt	Func.refgene	Gene.refgene	ExonicFunc.refgene	Genotype	Depth	AACChange.refgene	Inherited from
1	236899904	236899904	C	-	exonic	ACTN2	frameshift deletion	het	63	NM_001278344:exon10:c.95delC:p.P32fs	S7
1	100349983	100349983	C	T	exonic	AGL	nonsynonymous SNV	het	164	NM_000645:exon17:c.2471C>T:p.S824F	S7
2	202598051	202598051	C	T	exonic	ALS2	nonsynonymous SNV	het	160	NM_020919:exon13:c.2528G>A:p.R843Q	S6
6	129722481	129722481	T	G	exonic	LAMA2	nonsynonymous SNV	het	205	NM_000426:exon38:c.5558T>G:p.I1853R	S6
22	33673227	33673227	G	A	exonic	LARGE	nonsynonymous SNV	het	61	NM_133642:exon14:c.1892C>T:p.T631M	S7
11	47371414	47371414	C	T	exonic	MYBPC3	nonsynonymous SNV	het	41	NM_000256:exon5:c.565G>A:p.V189I	S7
2	152507189	152507189	C	G	exonic	NEB	nonsynonymous SNV	het	169	NM_001164507:exon53:c.7126G>C:p.V2376L	S7
8	24811071	24811071	G	-	exonic	NEFL	frameshift deletion	hom	183	NM_006158:exon3:c.1408delC:p.P470fs	S6, S7
10	13323113	13323113	G	T	splicing	PHYH	NA	het	173	NM_001037537:exon7:c.529_3C.A	S6
1	46656162	46656162	T	G	exonic	POMGNT1	nonsynonymous SNV	het	126	NM_001290129:exon18:c.1566A>C:p.E522D	S6
19	39002998	39002998	C	T	exonic	RYR1	nonsynonymous SNV	het	132	NM_000540:exon63:c.9347C>T:p.S3116L	S7
11	118897348	118897348	C	T	exonic	SLC37A4	nonsynonymous SNV	het	40	NM_001164280:exon6:c.835G>A:p.A279T	S6
11	118898437	118898437	C	-	exonic;splicing	SLC37A4	frameshift deletion	hom	180	NM_001164280:exon3:c.527delG:p.C176fs	S6, S7
13	36900750	36900750	G	A	exonic	SPG20	nonsynonymous SNV	het	174	NM_001142294:exon5:c.1250C>T:p.P417L	S7
2	179441038	179441038	C	T	exonic	TTN	nonsynonymous SNV	het	14	NM_003319:exon154:c.4262G>A:p.G14209D	S7
12	974356	974356	-	C	exonic	WNK1	frameshift insertion	hom	46	NM_213655:exon9:c.2220dupC:p.L740fs	S6, S7
14	68248167	68248167	C	G	exonic	ZFYVE26	nonsynonymous SNV	het	116	NM_015346:exon22:c.4452G>C:p.R1484S	S6

**Πίνακας 3.19:** Υποψήφιες παθολογικές παραλλαγές μετά από ανάλυση του VCF αρχείου του δείγμα S5 (Child) της 3<sup>ης</sup> Ροής Εργασίας με την πλατφόρμα VarAFT. Αφορούν 21 παραλλαγές που εξάχθηκαν με την ανάλυση του δείγματος ως Τριο (χρήση αρχείων S6 & S7) και εφαρμόζοντας το πάνελ γονιδίων.

Chr	Start	End	Ref	Alt	Func.refgene	Gene.refgene	ExonicFunc.refgene	Genotype	Depth	AACChange.refgene	Inherited from
1	236899904	236899904	C	-	exonic	ACTN2	frameshift deletion	het	82	NM_001278344:exon10:c.95delC:p.P32fs	S7
1	100349983	100349983	C	T	exonic	AGL	nonsynonymous SNV	het	116	NM_000645:exon17:c.2471C>T:p.S824F	S7
2	202598051	202598051	C	T	exonic	ALS2	nonsynonymous SNV	het	90	NM_020919:exon13:c.2528G>A:p.R843Q	S6
6	7542149	7542149	-	A	exonic	DSP	frameshift insertion	het	194	NM_001008844:exon1:c.1dupG:p.M1fs	S6
6	129722481	129722481	T	G	exonic	LAMA2	nonsynonymous SNV	het	45	NM_000426:exon38:c.5558T>G:p.I1853R	S6
22	33673227	33673227	G	A	exonic	LARGE	nonsynonymous SNV	het	104	NM_133642:exon14:c.1892C>T:p.T631M	S7
11	47371414	47371414	C	T	exonic	MYBPC3	nonsynonymous SNV	het	227	NM_000256:exon5:c.565G>A:p.V189I	S7
2	152507189	152507189	C	G	exonic	NEB	nonsynonymous SNV	het	126	NM_001164507:exon53:c.7126G>C:p.V2376L	S7
8	24811071	24811071	G	-	exonic	NEFL	frameshift deletion	hom	243	NM_006158:exon3:c.1408delC:p.P470fs	S6, S7
10	13323113	13323113	G	T	splicing	PHYH	NA	het	135	NM_001037537:exon7:c.529_3C.A	S6
1	46656162	46656162	T	G	exonic	POMGNT1	nonsynonymous SNV	het	49	NM_001290129:exon18:c.1566A>C:p.E522D	S6
19	39002998	39002998	C	T	exonic	RYR1	nonsynonymous SNV	het	226	NM_000540:exon63:c.9347C>T:p.S3116L	S7
11	118898437	118898437	C	-	exonic;splicing	SLC37A4	frameshift deletion	hom	220	NM_001164280:exon3:c.527delG:p.C176fs	S6, S7
11	118897348	118897348	C	T	exonic	SLC37A4	nonsynonymous SNV	het	91	NM_001164280:exon6:c.835G>A:p.A279T	S6
13	36900750	36900750	G	A	exonic	SPG20	nonsynonymous SNV	het	109	NM_001142294:exon5:c.1250C>T:p.P417L	S7
14	64443278	64443278	C	T	splicing	SYNE2	NA	het	35	NM_015180:exon12:c.1129_3C.T	S6
2	179441038	179441038	C	T	exonic	TTN	nonsynonymous SNV	het	216	NM_003319:exon154:c.4262G>A:p.G14209D	S7
2	179616770	179616770	G	A	splicing	TTN	NA	het	36	NM_133379:exon46:c.10361_4C.T	S7
12	974356	974356	-	C	exonic	WNK1	frameshift insertion	hom	215	NM_213655:exon9:c.2220dupC:p.L740fs	S6, S7
12	974311	974311	-	C	exonic	WNK1	frameshift insertion	het	243	NM_213655:exon9:c.2175dupC:p.P725fs	S7
14	68248167	68248167	C	G	exonic	ZFYVE26	nonsynonymous SNV	het	153	NM_015346:exon22:c.4452G>C:p.R1484S	S6

**Πίνακας 3.20:** Κοινές υποψήφιες παθολογικές παραλλαγές μετά από ανάλυση με την πλατφόρμα VarAFT των αρχείων της 1<sup>ης</sup> Ροής Εργασίας, της 2<sup>ης</sup> Ροής Εργασίας και της 3<sup>ης</sup> Ροής Εργασίας του δείγματος S5 (Child). Αφορά 14 παραλλαγές που εξάχθηκαν με εφαρμογή του πάνελ γονιδίων.

Chr	Start	End	Ref	Alt	Func.refgene	Gene.refgene	ExonicFunc.refgene	Genotype	Depth	AACChange.refgene
1	236899904	236899904	C	-	exonic	ACTN2	frameshift deletion	het	82	NM_001278344:exon10:c.95delC:p.P32fs
1	100349983	100349983	C	T	exonic	AGL	nonsynonymous SNV	het	116	NM_000645:exon17:c.2471C>T:p.S824F
2	202598051	202598051	C	T	exonic	ALS2	nonsynonymous SNV	het	90	NM_020919:exon13:c.2528G>A:p.R843Q
6	129722481	129722481	T	G	exonic	LAMA2	nonsynonymous SNV	het	45	NM_000426:exon38:c.5558T>G:p.I1853R
22	33673227	33673227	G	A	exonic	LARGE	nonsynonymous SNV	het	104	NM_133642:exon14:c.1892C>T:p.T631M
11	47371414	47371414	C	T	exonic	MYBPC3	nonsynonymous SNV	het	227	NM_000256:exon5:c.565G>A:p.V189I
2	152507189	152507189	C	G	exonic	NEB	nonsynonymous SNV	het	126	NM_001164507:exon53:c.7126G>C:p.V2376L
8	24811071	24811071	G	-	exonic	NEFL	frameshift deletion	hom	243	NM_006158:exon3:c.1408delC:p.P470fs
10	13323113	13323113	G	T	splicing	PHYH	NA	het	135	NM_001037537:exon7:c.529_3C.A
1	46656162	46656162	T	G	exonic	POMGNT1	nonsynonymous SNV	het	49	NM_001290129:exon18:c.1566A>C:p.E522D
19	39002998	39002998	C	T	exonic	RYR1	nonsynonymous SNV	het	226	NM_000540:exon63:c.9347C>T:p.S3116L
11	118898437	118898437	C	-	exonic;splicing	SLC37A4	frameshift deletion	hom	220	NM_001164280:exon3:c.527delG:p.C176fs
13	36900750	36900750	G	A	exonic	SPG20	nonsynonymous SNV	het	109	NM_001142294:exon5:c.1250C>T:p.P417L
14	68248167	68248167	C	G	exonic	ZFYVE26	nonsynonymous SNV	het	153	NM_015346:exon22:c.4452G>C:p.R1484S

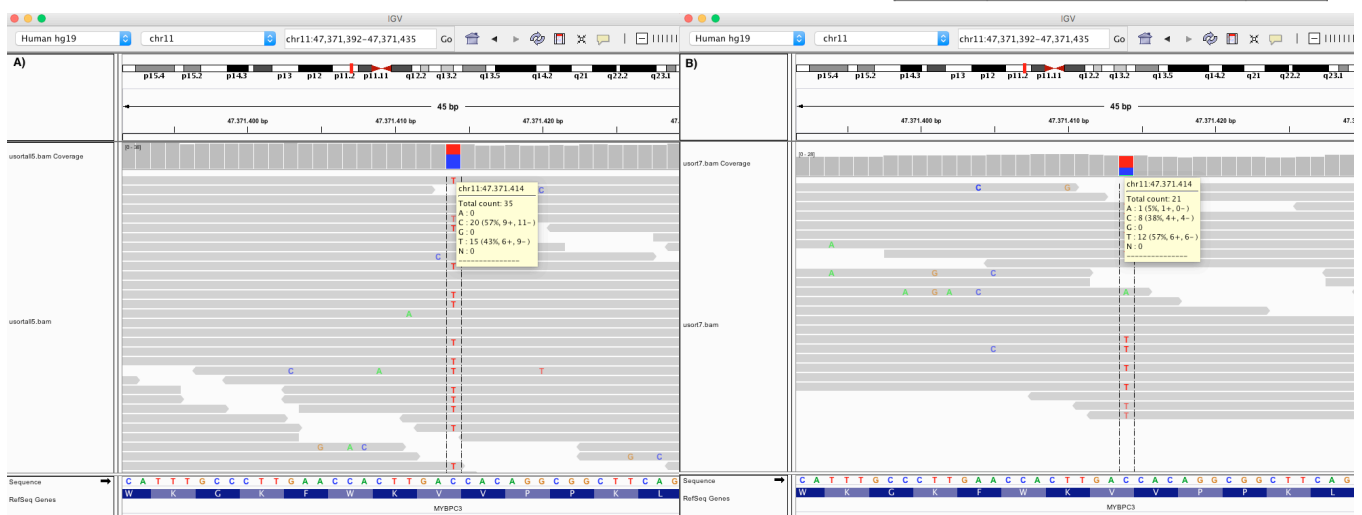


### 3.4) Συγκριτική μελέτη των παραλλαγών που προσδιορίστηκαν με τις τρεις Ροές Εργασίας

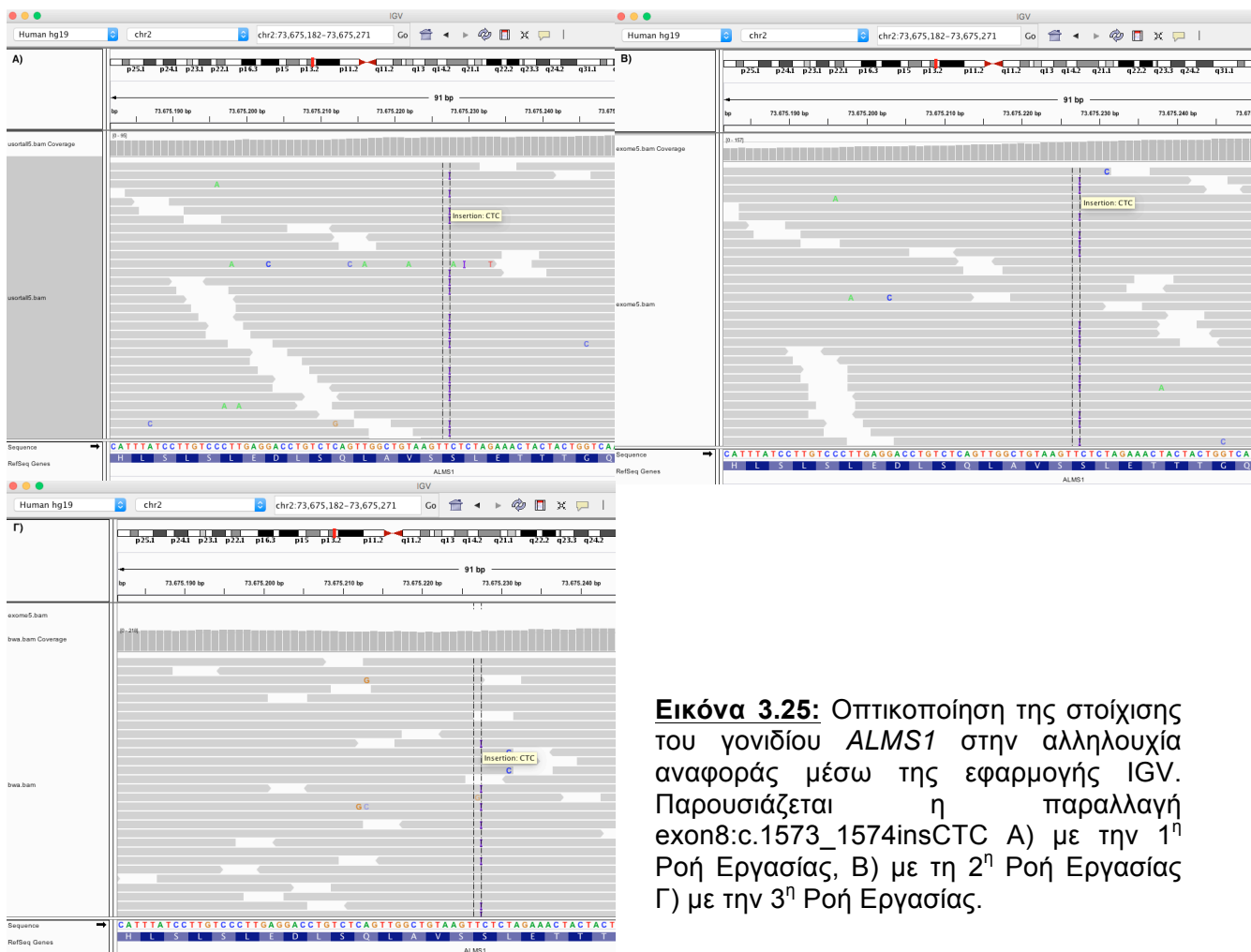
Τα αρχεία των παραλλαγών που εξάχθηκαν από κάθε Ροή Εργασίας αναλύθηκαν με την πλατφόρμα VarAFT και εφαρμόστηκε το πάνελ γονιδίων (Πίνακας 3.21). Οι διαφορετικές παραλλαγές που εντοπίστηκαν μεταξύ των τριών Ροών Εργασίας αναλύθηκαν με την απεικονιστική εφαρμογή IGV (Integrative Genomics Viewer) (Εικόνα 3.24-3.33). Οι διαφορές στις παραλλαγές βρίσκονται στα γονίδια: *ALMS1* (NM\_015120), *KIF1A* (NM\_001244008), *SYNE1* (NM\_033071), *SLC37A4* (NM\_001164280), *WNK1* (NM\_213655), *DSP* (NM\_001008844), *SYNE2* (NM\_015180), *TTN* (NM\_003319).

**Πίνακας 3.21:** Σύγκριση των υποψήφιων παθολογικών παραλλαγών των τριών Ροών Εργασίας του δείγματος S5 (Child), μετά από ανάλυση με την πλατφόρμα VarAFT και εφαρμόζοντας το πάνελ γονιδίων.

1 <sup>η</sup> Ροή Εργασίας			2 <sup>η</sup> Ροή Εργασίας			3 <sup>η</sup> Ροή Εργασίας		
Gene.refgene	AAChange.refgene	Genotype	Gene.refgene	AAChange.refgene	Genotype	Gene.refgene	AAChange.refgene	Genotype
ACTN2	exon10:c.95delC;p.P32fs	het	ACTN2	exon10:c.95delC;p.P32fs	het	ACTN2	exon10:c.95delC;p.P32fs	het
AGL	exon17:c.2471C>T;p.S824F	het	AGL	exon17:c.2471C>T;p.S824F	het	AGL	exon17:c.2471C>T;p.S824F	het
ALMS1	exon8:c.1573_1574insCTC;p.L525delinsPL	het						
ALS2	exon13:c.2528G>A;p.R843Q	het	ALS2	exon13:c.2528G>A;p.R843Q	het	ALS2	exon13:c.2528G>A;p.R843Q	het
KIF1A	exon27:c.2721_2723del;p.907_908del	het						
LAMA2	exon38:c.5558T.G;p.I1853R	het	LAMA2	exon38:c.5558T.G;p.I1853R	het	LAMA2	exon38:c.5558T.G;p.I1853R	het
LARGE	exon14:c.1892C>T;p.T631M	het	LARGE	exon14:c.1892C>T;p.T631M	het	LARGE	exon14:c.1892C>T;p.T631M	het
MYBPC3	exon5:c.565G>A;p.V189I	het	MYBPC3	exon5:c.565G>A;p.V189I	het	MYBPC3	exon5:c.565G>A;p.V189I	het
NEB	exon53:c.7126G>C;p.V2376L	het	NEB	exon53:c.7126G>C;p.V2376L	het	NEB	exon53:c.7126G>C;p.V2376L	het
NEFL	exon3:c.1408delC;p.P470fs	hom	NEFL	exon3:c.1408delC;p.P470fs	hom	NEFL	exon3:c.1408delC;p.P470fs	hom
PHYH	exon7:c.529_3C.A	het	PHYH	exon7:c.529_3C.A	het	PHYH	exon7:c.529_3C.A	het
POMGNT1	exon18:c.1566A>C;p.E522D	het	POMGNT1	exon18:c.1566A>C;p.E522D	het	POMGNT1	exon18:c.1566A>C;p.E522D	het
RYR1	exon63:c.9347C>T;p.S3116L	het	RYR1	exon63:c.9347C>T;p.S3116L	het	RYR1	exon63:c.9347C>T;p.S3116L	het
SLC37A4	exon3:c.527delG;p.C176fs	hom	SLC37A4	exon3:c.527delG;p.C176fs	hom	SLC37A4	exon3:c.527delG;p.C176fs	hom
			SLC37A4	exon6:c.835G>A;p.A279T	het	SLC37A4	exon6:c.835G>A;p.A279T	het
SPG20	exon5:c.1250C>T;p.P417L	het	SPG20	exon5:c.1250C>T;p.P417L	het	SPG20	exon5:c.1250C>T;p.P417L	het
			TTN	exon154:c.42626G>A;p.G14209D	het	TTN	exon154:c.42626G>A;p.G14209D	het
			WNK1	exon9:c.2220dupC;p.L740fs	hom	WNK1	exon9:c.2220dupC;p.L740fs	hom
SYNE1	exon76:c.12352G>A;p.V4118M	het						
ZFYVE26	exon22:c.4452G>C;p.R1484S	het	ZFYVE26	exon22:c.4452G>C;p.R1484S	het	ZFYVE26	exon22:c.4452G>C;p.R1484S	het
						TTN	exon46:c.10361_4C.T	het
						WNK1	exon9:c.2175dupC;p.P725fs	het
						DSP	exon1:c.1dupA;p.M1fs	het
						SYNE2	exon12:c.1129_3C.T	het



**Εικόνα 3.24:** Οπτικοποίηση της στοίχισης του γονιδίου *MYBPC3* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζεται η παραλλαγή exon5:c.565G>A με την 1<sup>η</sup> Ροή Εργασίας. Α) Δείγμα S5 (Child) Β) Δείγμα S7 (Mother).

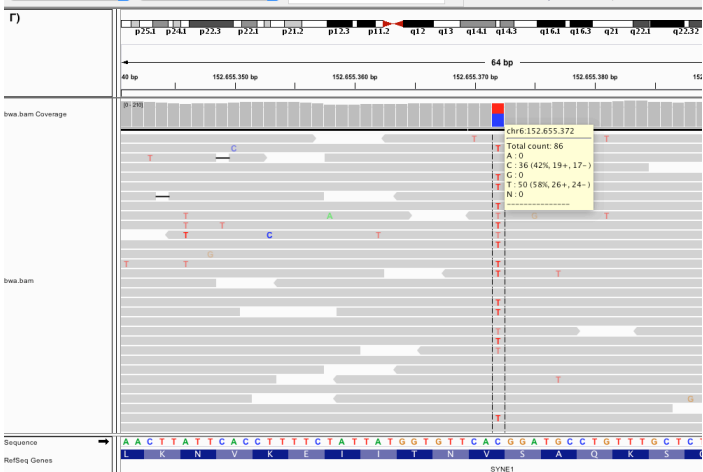
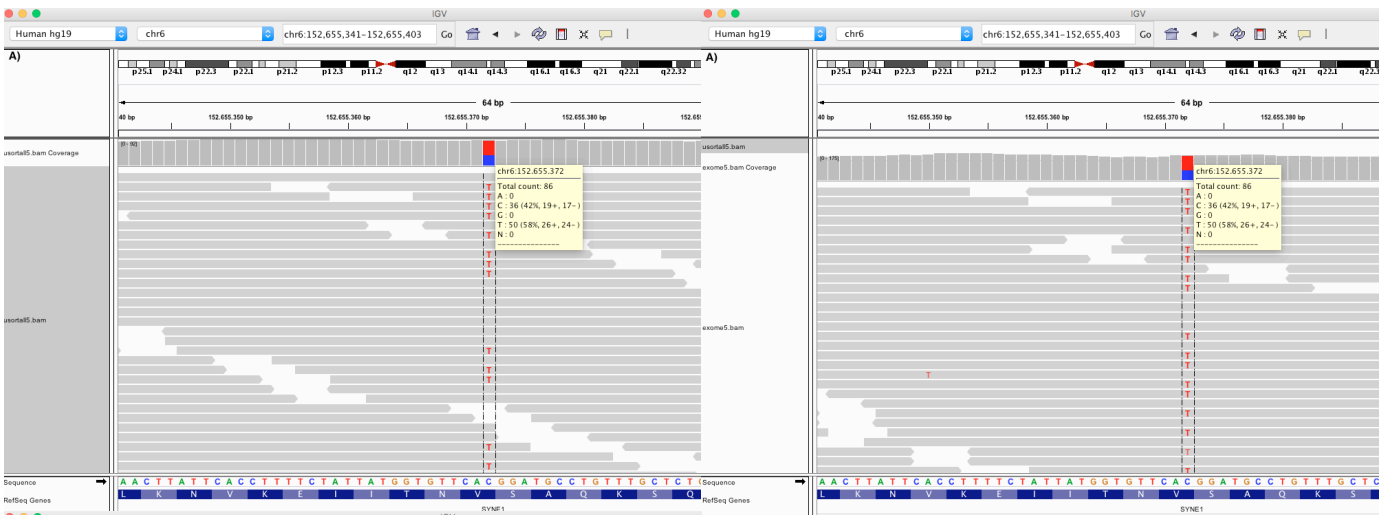


**Εικόνα 3.25:** Οπτικοποίηση της στοίχισης του γονιδίου *ALMS1* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζεται η παραλλαγή exon8:c.1573\_1574insCTC A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

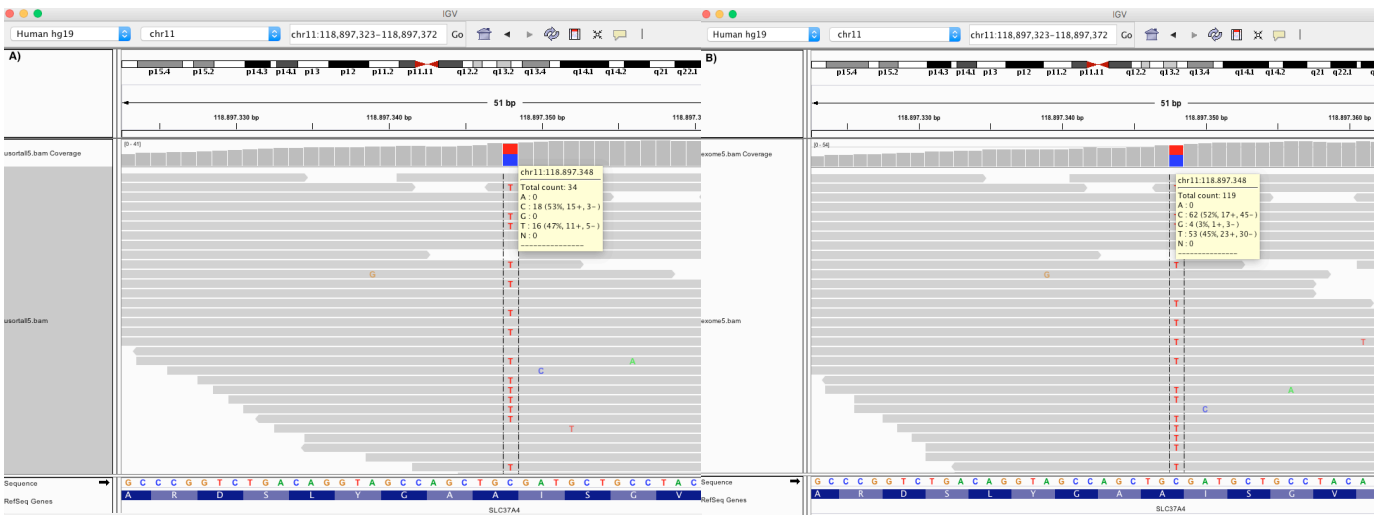


**Εικόνα 3.26:** Οπτικοποίηση της στοίχισης του γονιδίου *KIF1A* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζεται η παραλλαγή exon27:c.2721\_2723delTCC A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

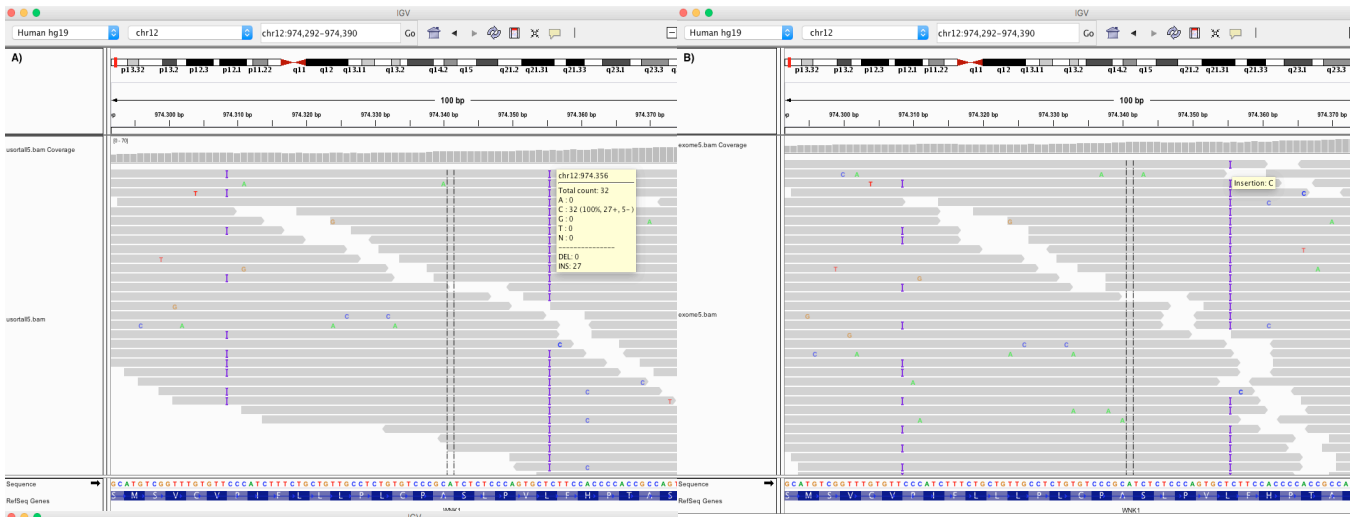




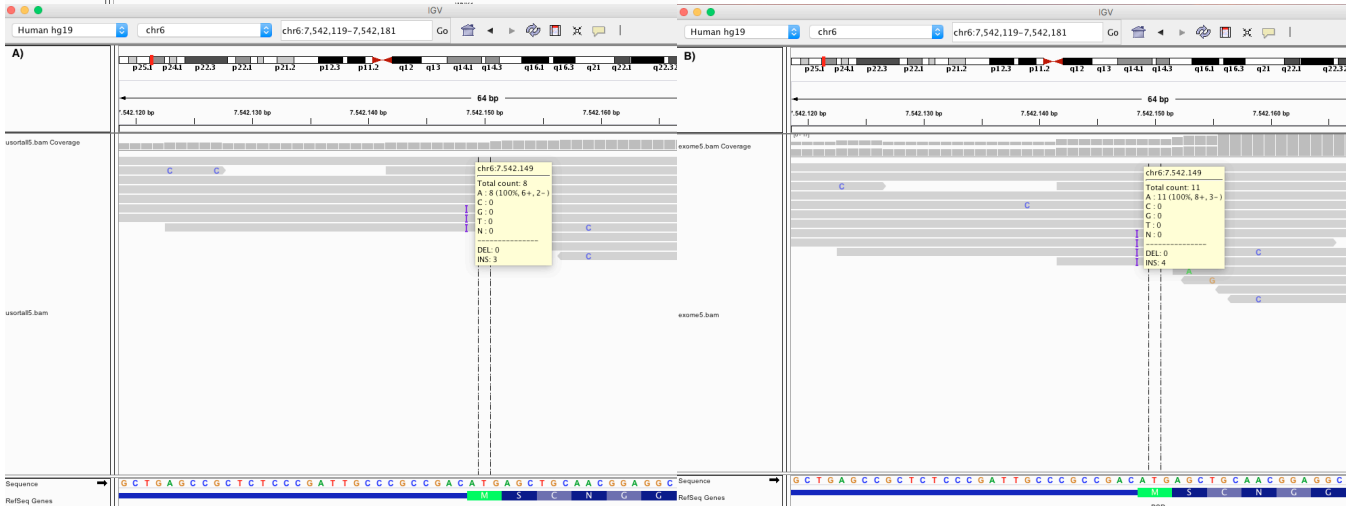
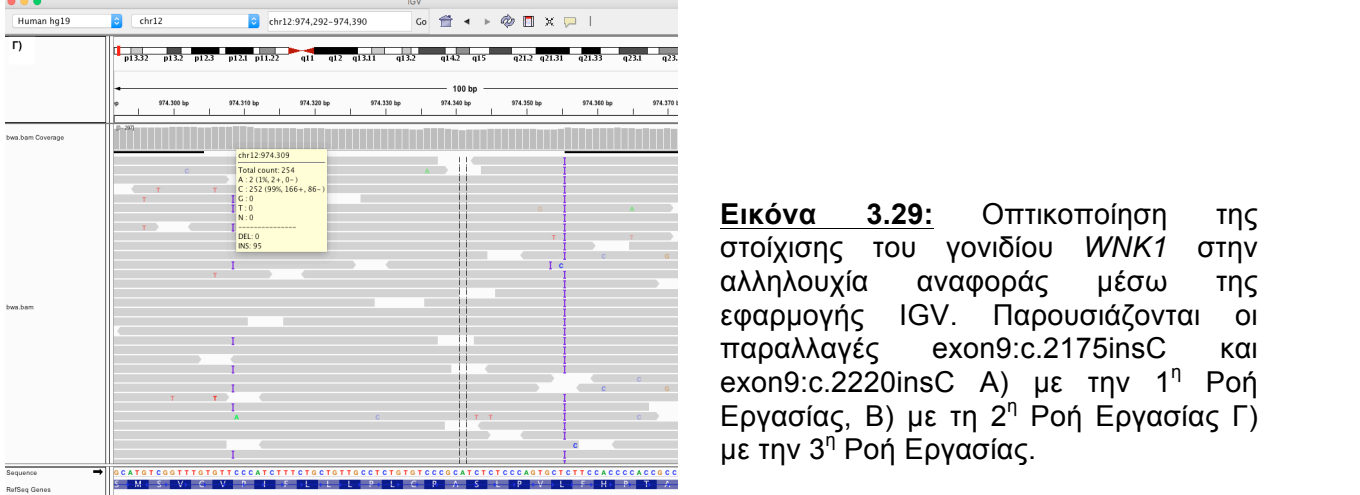
**Εικόνα 3.27:** Οπτικοποίηση της στοίχισης του γονιδίου *SYNE1* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζεται η παραλλαγή exon76:c.12352G>A A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.



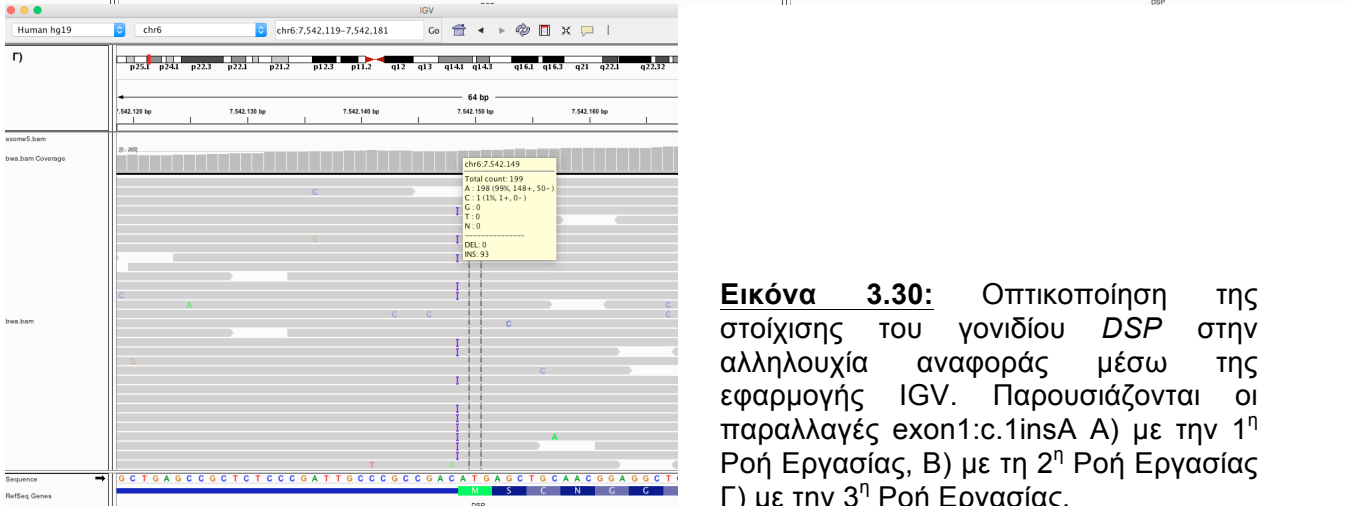
**Εικόνα 3.28:** Οπτικοποίηση της στοίχισης του γονιδίου *SLC37A4* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζεται η παραλλαγή exon6:c.835G>A A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

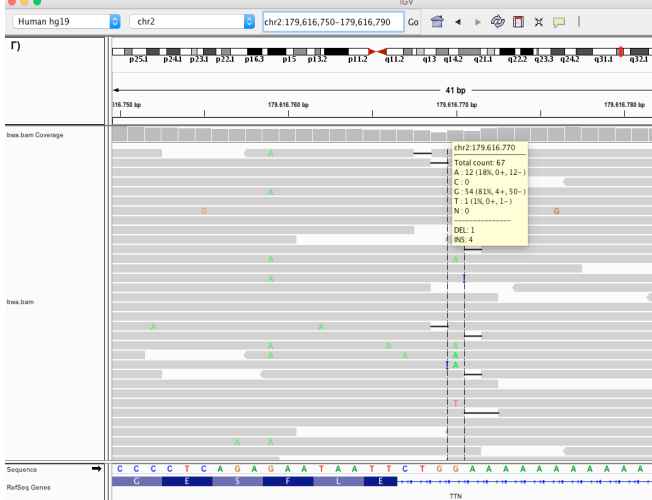
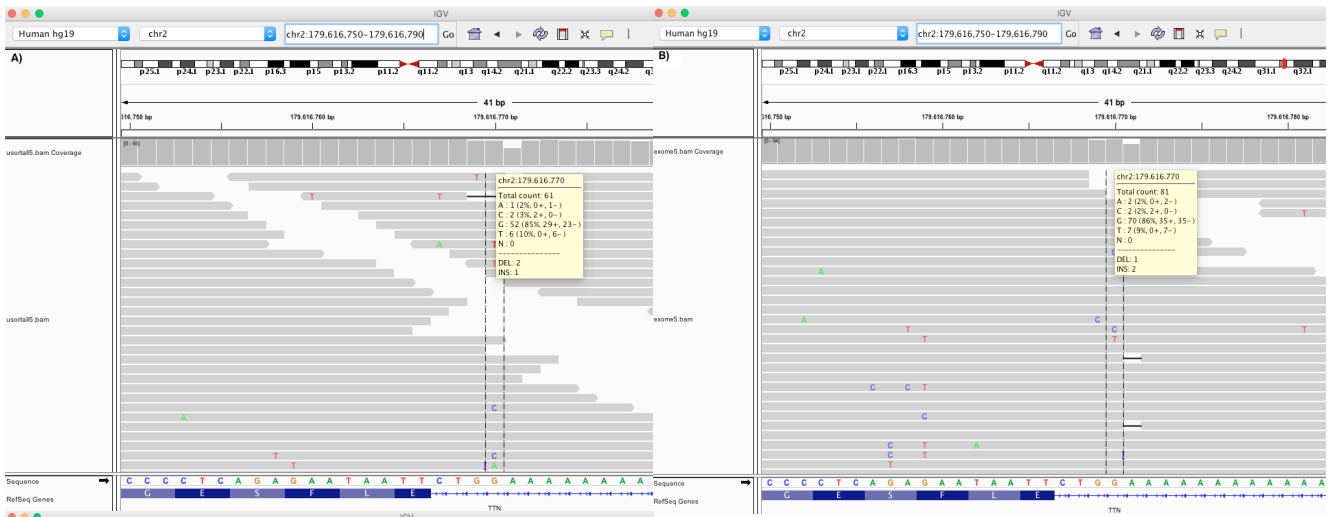


**Εικόνα 3.29:** Οπτικοποίηση της στοίχισης του γονιδίου *WNK1* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζονται οι παραλλαγές exon9:c.2175insC και exon9:c.2220insC A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

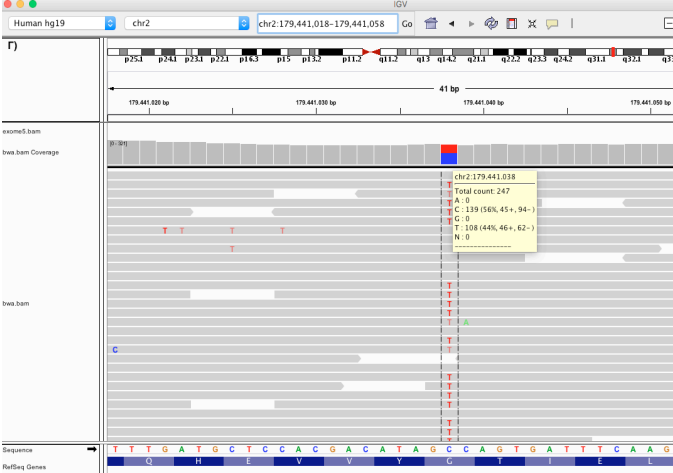
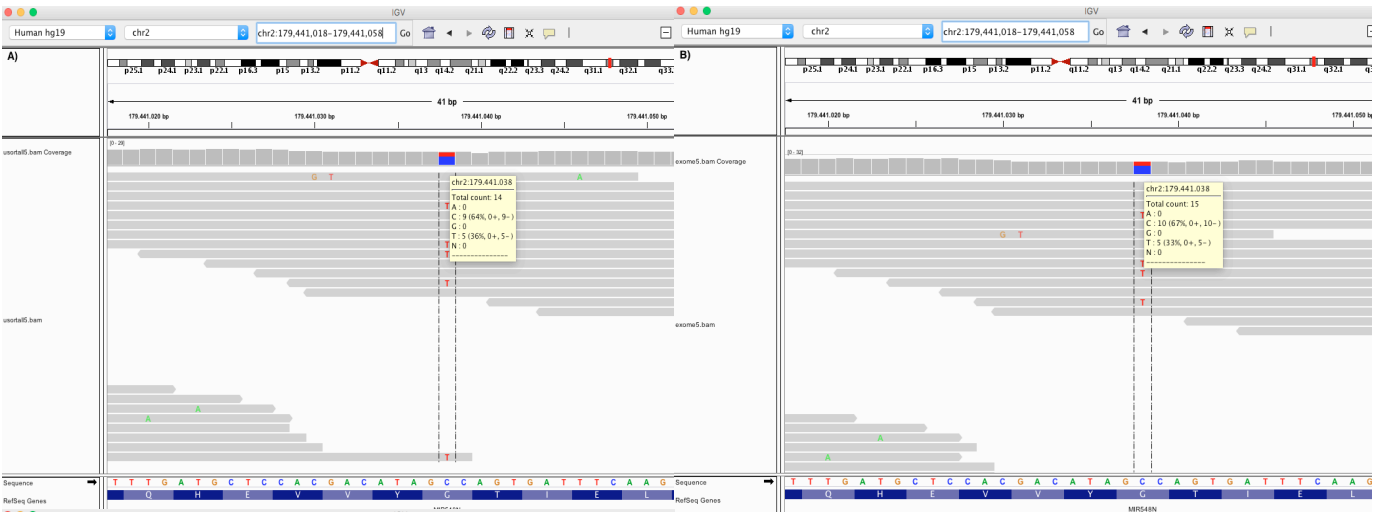


**Εικόνα 3.30:** Οπτικοποίηση της στοίχισης του γονιδίου *DSP* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζονται οι παραλλαγές exon1:c.1insA A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

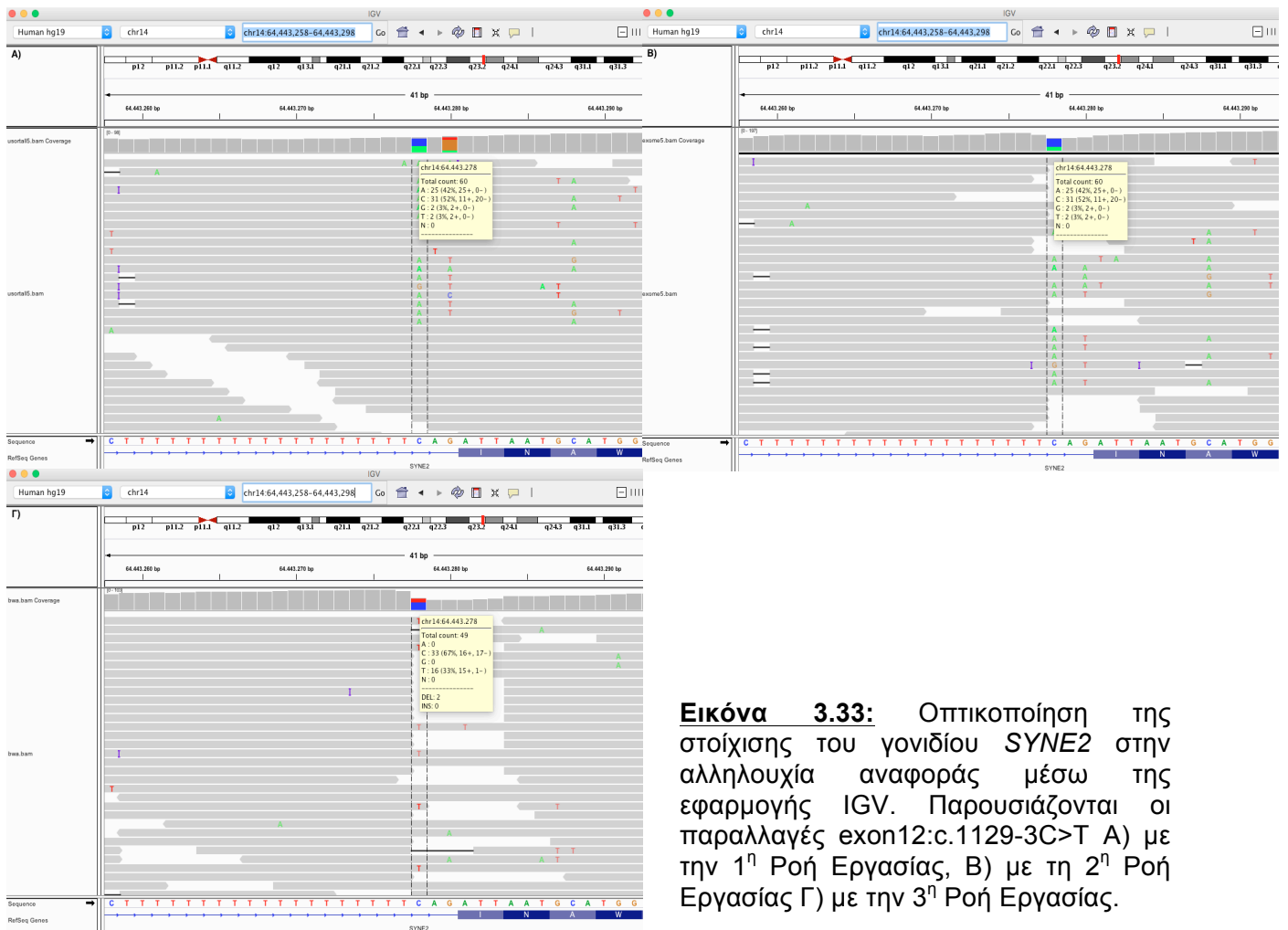




**Εικόνα 3.31:** Οπτικοποίηση της στοίχισης του γονιδίου *TTN* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζονται οι παραλλαγές exon46:c.10361-4C>T A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.



**Εικόνα 3.32:** Οπτικοποίηση της στοίχισης του γονιδίου *TTN* στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζονται οι παραλλαγές exon154:c.42626G>A A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.



**Εικόνα 3.33:** Οπτικοποίηση της στοίχισης του γονιδίου SYNE2 στην αλληλουχία αναφοράς μέσω της εφαρμογής IGV. Παρουσιάζονται οι παραλλαγές exon12:c.1129-3C>T A) με την 1<sup>η</sup> Ροή Εργασίας, B) με τη 2<sup>η</sup> Ροή Εργασίας Γ) με την 3<sup>η</sup> Ροή Εργασίας.

Μελετώντας τους Πίνακες 3.10, 3.13 παρατηρείται ότι οι παραλλαγές, ιδίως οι φιλτραρισμένες-υποψήφιες παραλλαγές δε διαφέρουν αρκετά, κάτι που αποδεικνύεται και από τους αναλυτικούς πίνακες. Πιο αναλυτικά, στους Πίνακες 3.11 και 3.12 υπάρχει διαφοροποίηση σε μια παραλλαγή, στο γονίδιο MYBPC3 ex5:c.565 G>A, η οποία δεν εμφανίζεται στην ανάλυση ως Τρίο. Όμως, σύμφωνα με τον Πίνακα 3.18 και 3.19 φαίνεται ότι κληρονομείται από τη μητέρα (S7-Mother). Αυτό οφείλεται στη χαμηλή κάλυψη που εμφανίζει η συγκεκριμένη περιοχή στο δείγμα S7 καθιστώντας αδύνατη την απόδοση της παραλλαγής (Εικόνα 3.24). Στους Πίνακες 3.14, 3.16 όπου εμφανίζονται οι ισχυρά παθολογικές υποψήφιες παραλλαγές εφαρμόζοντας φιλτράρισμα βάσει συχνότητας εμφάνισης στον πληθυσμό (1000G) απομακρύνονται όλες οι παραλλαγές του Πίνακα 3.14, ενώ στον Πίνακα 3.16 παραμένει μόνο η παραλλαγή WNK1 ex9:c.2175insC. Οι Πίνακες 3.15 και 3.17 που παρουσιάζουν τις δυνητικά παθολογικές υποψήφιες παραλλαγές έχουν εξαχθεί με την εφαρμογή του φίλτρου της συχνότητας και είναι τελικοί.

Συγκρίνοντας τις παραλλαγές που εξάγονται με τις τρεις Ροές Εργασίας (Πίνακας 3.21 και Εικόνα 3.25-3.33) προκύπτει για κάθε παραλλαγή που διαφοροποιείται ότι:

- *ALMS1* exon8:c.1573\_1574insCTC: Στην 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας η παραλλαγή εμφανίζεται στη χρωμοσωμική θέση chr2:73.675.227 και απομακρύνεται λόγω υψηλής συχνότητας στον πληθυσμό (1000G). Ενώ στην 1<sup>η</sup> Ροή Εργασίας εμφανίζεται στη χρωμοσωμική θέση chr2:73.675.230 η οποία είναι εσφαλμένη και γι' αυτό το λόγο δεν υπάρχουν στοιχεία στις βάσεις δεδομένων.
- *KIF1A* exon27:c.2721\_2723delTCC: Στην 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας η παραλλαγή εμφανίζεται στη χρωμοσωμική θέση chr2:241.696.841-843 και απομακρύνεται λόγω υψηλής συχνότητας στον πληθυσμό (1000G). Ενώ στην 1<sup>η</sup> Ροή Εργασίας εμφανίζεται στη χρωμοσωμική θέση chr2:241.696.871 η οποία είναι εσφαλμένη και γι' αυτό το λόγο δεν υπάρχουν στοιχεία στις βάσεις δεδομένων.
- *SYNE1* exon76:c.12352G>A: Στην 2<sup>η</sup> και 3<sup>η</sup> Ροή Εργασίας η παραλλαγή απομακρύνεται κατά τη διαδικασία του φιλτραρίσματος καθώς τα *in silico* tools το εμφανίζουν ως tolerate.
- *SLC37A4* exon6:c.835G>A: Στην 1<sup>η</sup> Ροή Εργασίας δεν εμφανίζεται η παραλλαγή καθώς έχει χαμηλή κάλυψη στην περιοχή.
- *WNK1* exon9:c.2175insC και exon9:c.2220insC: Στην 1<sup>η</sup> Ροή Εργασίας δεν εμφανίζονται οι παραλλαγές καθώς έχουν χαμηλή κάλυψη στην περιοχή.
- *DSP* exon1:c.1insA: Στην 1<sup>η</sup> Ροή Εργασίας δεν εμφανίζεται η παραλλαγή καθώς έχει χαμηλή κάλυψη στην περιοχή.
- *TTN* exon46:c.10361-4C>T και exon154:c.42626G>A: Στην 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας δεν εμφανίζονται οι παραλλαγές καθώς έχουν χαμηλή κάλυψη στην περιοχή αντίστοιχα.
- *SYNE2* exon12:c.1129-3C>T: Στην 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας έχει γίνει ετερόζυγη αντικατάσταση του C>A, η οποία είναι πολυμορφισμός (ExAC). Ενώ στην 3<sup>η</sup> Ροή Εργασίας από όπου καταγράφεται η συγκεκριμένη παραλλαγή παρατηρείται ετερόζυγη αντικατάσταση C>T, η οποία εμφανίζεται στα *in silico* εργαλεία ως παθολογική.

**Κεφάλαιο 4:**

***Συμπεράσματα –  
Συζήτηση***

#### **4.1) Αξιολόγηση της ροής εργασίας της Αλληλούχησης της Επόμενης Γενιάς σε όλα τα εξόνια του γονιδιώματος - Whole Exome Sequencing (WES) – Συμπεράσματα**

Σύμφωνα με τα δεδομένα που εξάχθηκαν από τον ποιοτικό έλεγχο του DNA πριν από την προετοιμασία της βιβλιοθήκης, αλλά και κατά τη διάρκεια της κατασκευής η ποσότητα σε κάθε δείγμα επαρκεί για την πραγματοποίηση του πειράματος σύμφωνα με το πρωτόκολλο, τόσο μετά την απομόνωση του από τα κύτταρα του αίματος, στην κάθε βιβλιοθήκη και στην εμπλουτισμένη βιβλιοθήκη (Πίνακας 3.1). Ο ποιοτικός έλεγχος των βιβλιοθηκών έδειξε ότι το μέγεθος των τμημάτων DNA μετά τη θραυσματοποίηση είναι κατάλληλα για τη συνέχεια του πειράματος. Επίσης, η ποιότητα της εμπλουτισμένης βιβλιοθήκης (pooled library) είναι η αναμενόμενη σύμφωνα με το πρωτόκολλο (Εδάφιο 3.2.2).

Ο ποιοτικός έλεγχος του Base calling, δηλαδή των δεδομένων που εξάγονται αμέσως μετά την αλληλούχηση (Εδάφιο 3.3.1), έδειξε ότι η μέση ποιότητα κλήσης βάσεων με την 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας για τα δείγματα S5, S6, S7 είναι άνω του 30 Phred score (πράσινη ζώνη), που σημαίνει πολύ καλή ποιότητα, ενώ με την 3<sup>η</sup> Ροή Εργασίας είναι άνω του 20 Phred score (πορτοκαλί ζώνη), που σημαίνει αποδεκτή ποιότητα.

Ο ποιοτικός έλεγχος της στοίχισης/χαρτογράφησης έδειξε ότι ο αριθμός των αναγνώσεων (reads) είναι επαρκής στις τρεις Ροές Εργασίας (Πίνακας 3.2) και ο αριθμός των στοιχιζόμενων αναγνώσεων στην αλληλουχία αναφοράς είναι ικανοποιητικός (Πίνακας 3.3, Εικόνα 3.16-3.20). Πιο αναλυτικά, μελετώντας τα δεδομένα στοίχισης των αναγνώσεων στις περιοχές-στόχους, δηλαδή στις περιοχές που αποτελούν εξόνια, παρατηρείται ότι με την 1<sup>η</sup> Ροή Εργασίας (Πίνακας 3.4) τα ποσοστά των στοιχίσεων είναι αρκετά χαμηλά που σημαίνει ότι πολλές περιοχές έχουν καλυφθεί με λίγα reads ή και καθόλου και δε θα αποδίδουν παραλλαγές (Ψευδώς Αρνητικά). Επίσης, παρατηρούνται αυξημένα ποσοστά από duplicated reads, τα οποία θα αποδίδουν πολλά Ψευδώς Θετικά αποτελέσματα. Με τη 2<sup>η</sup> Ροή Εργασίας (Πίνακας 3.5) τα ποσοστά των στοιχίσεων στις περιοχές-στόχους είναι εξίσου χαμηλά, όμως σε συνδυασμό με τα ποσοστά στοίχισης των πλευρικών περιοχών (flanking regions) το αποτέλεσμα είναι πιο ικανοποιητικό από αυτο της 1<sup>η</sup> Ροής

Εργασίας. Με την 3<sup>η</sup> Ροή Εργασίας (Πίνακας 3.6) τα ποσοστά των στοιχίσεων στις περιοχές στόχους σε συνδυασμό με αυτά των πλευρικών περιοχών δίνουν ένα αρκετά ικανοποιητικό αποτέλεσμα σε σύγκριση με τις προηγούμενες δύο αποδίδοντας αξιόπιστα αποτελέσματα παραλλαγών.

Μια σημαντική παράμετρος για την αξιόπιστη ανίχνευση των παραλλαγών αποτελεί η κάλυψη των περιοχών-στόχων με τα αντίστοιχα reads. Μελετώντας τα ποσοστά για κάλυψη 25X, 50X, 100X (Πίνακας 3.7-3.9) παρατηρείται ότι η 1<sup>η</sup> και 2<sup>η</sup> Ροή Εργασίας εμφανίζει χαμηλά ποσοστά σε σύγκριση με αυτά της 3<sup>ης</sup> Ροής Εργασίας. Το γεγονός αυτό αφείλεται πιθανότατα στο TruSeq Exome Library Prep Kit της Illumina, το οποίο εμφανίζει αδυναμίες κάλυψης συγκεκριμένων περιοχών (π.χ. GC-rich), αφήνοντας “κενά” σε αυτές και δίνοντας μη αξιόπιστα αποτελέσματα (Ψευδώς Θετικά και Ψευδώς Αρνητικά).

Αξιολογώντας τη συγκριτική μελέτη των παραλλαγών που προσδιορίστηκαν με τις τρεις Ροές Εργασίας συμπεραίνει κανείς ότι οι διαφοροποιήσεις στον Πίνακα 3.21 και στις Εικόνες 3.24-3.33 οφείλονται συνοπτικά, σε ανεπαρκή στοιχεία των βάσεων δεδομένων και χρήση μη ενημερωμένων πληροφοριών, διαφορετική προσέγγιση και κριτήρια φιλτραρίσματος, καθώς και χρήση διαφορετικών βιοπληροφορικών εργαλείων. Επιπλέον, η χαμηλή κάλυψη ορισμένων περιοχών καθιστά αδύνατη την απόδοση παραλλαγών.

Συγκρίνοντας επί του συνόλου τις δύο πλατφόρμες ανάλυσης, VarAFT και Sophia DDM συμπαιραίνει κανείς πως είναι εύκολες στη χρήση με γραφική απεικόνιση των επιλογών που παρέχουν και σαφείς κατηγοριοποιήσεις των φίλτρων και των δυνατοτήτων που παρέχουν. Οι πλατφόρμες αυτές δίνουν τη δυνατότητα ταξινόμησης των παραλλαγών με διάφορους τρόπους, όπως παθογένεια, συχνότητα εμφάνισης στον πληθυσμό, τύπο παραλλαγής κ.α., δίνουν πληροφορίες για την κάλυψη των συγκεκριμένων περιοχών και σύνδεση με βάσεις δεδομένων. Επίσης, υπάρχει δυνατότητα και στις δύο πλατφόρμες να χρησιμοποιηθεί λίστα γονιδίων ενδιαφέροντος (βιοπληροφορικό πάνελ) καθιστώντας πιο εύκολη την ανάλυση και διερεύνηση των υποψήφιων παραλλαγών μειώνοντας αρκετά τον αριθμό τους.



## 4.2) Συζήτηση

Σκοπός της παρούσας μελέτης είναι η εφαρμογή της Αλληλούχησης Επόμενης Γενιάς (NGS) στην κλινική πράξη προσδιορίζοντας τις απαιτήσεις, τις ανάγκες και τους περιορισμούς της τεχνολογίας αυτής μέσα από τη σύγκριση των τριών Ρωών Εργασίας που εφαρμόστηκαν σύμφωνα με το Κεφάλαιο 2, αλλά και από την ανάλυση των αποτελεσμάτων σύμφωνα με το Κεφάλαιο 3.

Οι τεχνολογίες NGS χρησιμοποιούνται ήδη για να βοηθήσουν στη διάγνωση πολλών κληρονομικών ασθενειών και ειδικότερα η εφαρμογή του WES έχει αποδειχθεί πολύτιμη στην κλινική εφαρμογή. Ωστόσο, η εφαρμογή του NGS στην κλινική πράξη εξακολουθεί να αντιμετωπίζει κάποια εμπόδια. Πρώτον, η ταχεία παραγωγή τεράστιων ποσοτήτων δεδομένων αποτελεί τεράστια πρόκληση για τη διαχείρισή και την ανάλυσή τους. Δεύτερον, δεν υπάρχει πρότυπο για τον ποιοτικό έλεγχο των δεδομένων αλληλουχίας. Το πρόβλημα των σφαλμάτων προσδιορισμού αλληλουχίας παραμένει σημαντικό, αφού τέτοια σφάλματα δεν διακρίνονται από τις γενετικές παραλλαγές και θα μπορούσαν να αποδοθούν λανθασμένα ως μεταλλάξεις που σχετίζονται με το φαινότυπο. Είναι γνωστό ότι όλες οι εμπορικά διαθέσιμες πλατφόρμες NGS έχουν διαφορετικούς τύπους σφαλμάτων και ποσοστά σφάλματος, τα οποία πρέπει να αξιολογηθούν και να διορθωθούν. Επιπλέον, η εφαρμογή διαφορετικών στρατηγικών βιοπληροφορικής επηρεάζει σημαντικά την ανάλυση δεδομένων NGS. Συνεπώς, είναι απαραίτητο να κατανοηθούν οι αρχές, τα πλεονεκτήματα και οι περιορισμοί των εργαλείων βιοπληροφορικής, ώστε να δημιουργηθούν οι κατάλληλοι αγωγοί ανάλυσης 'pipelines' που μπορούν να παράγουν αξιόπιστα αποτελέσματα. Αυτό σημαίνει ότι όλα τα αποτελέσματα που πρόκειται να αναφερθούν σε διαγνωστικό πλαίσιο πρέπει να αξιολογηθούν εφαρμόζοντας τα κατάλληλα κριτήρια και σε ορισμένες περιπτώσεις να γίνει περαιτέρω αξιολόγηση με αλληλούχης Sanger, λειτουργικές μελέτες ή μελέτη οικογενειών.

Η μαζική συσσώρευση γονιδιωματικών δεδομένων εγείρει και ηθικά ζητήματα. Το εάν και πώς μπορούν να επιστραφούν τα αποτελέσματα αλληλουχίας στους ασθενείς παραμένει αμφίβολο. Η προστασία της ιδιωτικής ζωής και της

εμπιστευτικότητας των ατομικών γονιδιωματικών πληροφοριών προκαλεί επίσης ανησυχίες. Μόνο όταν οι ασθενείς είναι πρόθυμοι να μοιραστούν τις ιατρικές τους πληροφορίες και τα δεδομένα αλληλουχίας τους μπορούν να αξιοποιηθούν και να αξιολογηθούν. Τέλος, είναι σημαντικό να σημειωθεί ότι οι αποδεδειγμένες τεχνικές επιδόσεις του NGS ή οποιονδήποτε άλλων τεχνολογιών γονιδιωματικής δεν μπορούν να μεταφερθούν αυτόματα σε διαγνωστικές δοκιμασίες κλινικών υπηρεσιών. Παρά τις προκλήσεις αυτές, η NGS έχει προσφέρει πρωτοφανείς ευκαιρίες για κλινική διάγνωση.

Αξίζει λοιπόν να σημειωθεί, ότι η υψηλής ποιότητας διάγνωση, δεν είναι πλέον αποκλειστική ευθύνη των κλινικών γενετιστών, αλλά χρειάζεται συνεργασία μεταξύ εργαστηριακών γενετιστών, βιολόγων, βιοπληροφορικών και κλινικών γενετιστών. Επίσης, καθίσταται αναγκαία η τυποποίηση και ενσωμάτωση αυτοματοποιημένων εργαλείων ανάλυσης, ενημερωμένων βάσεων δεδομένων και ταυτόχρονα κατάλληλων πληροφορικών υποδομών για τη διευκόλυνση προσδιορισμού παραλλαγών. Οι παραπάνω ανάγκες απαιτούν επενδύσεις στην τεχνολογία την βιο/πληροφορικής, σε υπολογιστικές υποδομές, σε εξειδίκευση προσωπικού και κατάρτισή του στο πλαίσιο υπηρεσιών υγείας, στην ανάπτυξη πολιτικής σχετικά με την ανταλλαγή ευαίσθητων γονιδιωματικών δεδομένων και την προστασία τους, καθώς και στην καθιέρωση ισχυρής βάσης τεκμηρίωσης για την κλινική ερμηνεία των παραλλαγών.

Παρόλο που δεν υπάρχει αμφιβολία ότι οι τεχνολογίες προσδιορισμού της αλληλουχίας έχουν τη δυνατότητα να ωφελήσουν την υγεία, απαιτούνται σαφείς κατευθυντήριες οδηγίες για τις μεθοδολογίες NGS, όσον αφορά την ποιότητα των δεδομένων που απαιτούνται για την καταγραφή των παραλλαγών, τους βιοπληροφορικούς αλγορίθμους και την αναφορά των αποτελεσμάτων.

# **Κεφάλαιο 5:**

## ***Βιβλιογραφία***

## 5.1) Βιβλιογραφία

1. Altmann, A., et al. (2012). "A beginners guide to SNP calling from high-throughput DNA-sequencing data." *Hum Genet* 131(10): 1541-1554.
2. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53-59.
3. Biesecker, L. G. and R. C. Green (2014). "Diagnostic clinical genome and exome sequencing." *N Engl J Med* 370(25): 2418-2425.
4. Bonasio, R., et al. (2010). "Molecular signals of epigenetic states." *Science* 330(6004): 612-616.
5. Brookes, A. J. and P. N. Robinson (2015). "Human genotype-phenotype databases: aims, challenges and opportunities." *Nat Rev Genet* 16(12): 702-715.
6. Coonrod, E. M., et al. (2013). "Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach." *Arch Pathol Lab Med* 137(3): 415-433.
7. Cox, M. P., et al. (2010). "SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data." *BMC Bioinformatics* 11: 485.
8. Danecek, P., et al. (2011). "The variant call format and VCFtools." *Bioinformatics* 27(15): 2156-2158.
9. Delaney, S. K., et al. (2016). "Toward clinical genomics in everyday medicine: perspectives and recommendations." *Expert Rev Mol Diagn* 16(5): 521-532.
10. Endrullat, C., et al. (2016). "Standardization and quality management in next-generation sequencing." *Appl Transl Genom* 10: 2-9.
11. Gargis, A. S., et al. (2015). "Good laboratory practice for clinical next-generation sequencing informatics pipelines." *Nat Biotechnol* 33(7): 689-693.
12. Gilissen, C., et al. (2012). "Disease gene identification strategies for exome sequencing." *Eur J Hum Genet* 20(5): 490-497.
13. Green, R. C., et al. (2013). "ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing." *Genet Med* 15(7): 565-574.
14. Hoskinson, D. C., et al. (2017). "The current state of clinical interpretation of sequence variants." *Curr Opin Genet Dev* 42: 33-39.

15. Kao, W. C. and Y. S. Song (2011). "naiveBayesCall: an efficient model-based base-calling algorithm for high-throughput sequencing." *J Comput Biol* 18(3): 365-377.
16. Kassahn, K. S., et al. (2014). "Integrating massively parallel sequencing into diagnostic workflows and managing the annotation and clinical interpretation challenge." *Hum Mutat* 35(4): 413-423.
17. Kircher, M., et al. (2009). "Improved base calling for the Illumina Genome Analyzer using machine learning strategies." *Genome Biol* 10(8): R83.
18. Krier, J. B., et al. (2016). "Genomic sequencing in clinical practice: applications, challenges, and opportunities." *Dialogues Clin Neurosci* 18(3): 299-312.
19. Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." *Nat Methods* 9(4): 357-359.
20. Lapin, V., et al. (2016). "Regulating whole exome sequencing as a diagnostic test." *Hum Genet* 135(6): 655-673.
21. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18:1851–1858
22. Li, H., et al. (2009). "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25(16): 2078-2079.
23. Li, R., et al. (2009). "SNP detection for massively parallel whole-genome resequencing." *Genome Res* 19(6): 1124-1132.
24. Liu, L., et al. (2012). "Comparison of next-generation sequencing systems." *J Biomed Biotechnol* 2012: 251364.
25. Lunter, G. and M. Goodson (2011). "Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads." *Genome Res* 21(6): 936-939.
26. MacArthur, D. G., et al. (2014). "Guidelines for investigating causality of sequence variants in human disease." *Nature* 508(7497): 469-476.
27. Majewski, J., et al. (2011). "What can exome sequencing do for you?" *J Med Genet* 48(9): 580-589.
28. Mardis, E. R. (2008). "Next-generation DNA sequencing methods." *Annu Rev Genomics Hum Genet* 9: 387-402.

29. Moorthie, S., et al. (2013). "Informatics and clinical genome sequencing: opening the black box." *Genet Med* 15(3): 165-171.
30. Muzzey, D., et al. (2015). "Understanding the Basics of NGS: From Mechanism to Variant Calling." *Curr Genet Med Rep* 3(4): 158-165.
31. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One*. 2013;22;8(10):e77910
32. Okonechnikov, K., et al. (2016). "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." *Bioinformatics* 32(2): 292-294.
33. Posey, J. E., et al. (2016). "Molecular diagnostic experience of whole-exome sequencing in adult patients." *Genet Med* 18(7): 678-685.
34. Rabbani, B., et al. (2014). "The promise of whole-exome sequencing in medical genetics." *J Hum Genet* 59(1): 5-15.
35. Rehm, H. L., et al. (2013). "ACMG clinical laboratory standards for next-generation sequencing." *Genet Med* 15(9): 733-747.
36. Richards, S., et al. (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of
37. Medical Genetics and Genomics and the Association for Molecular Pathology." *Genet Med* 17(5): 405-424.
38. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. *Gen Biol*. 2013;14:R51
39. Salgado, D., et al. (2016). "How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era." *Hum Mutat* 37(12): 1272-1282.
40. Shashi, V., et al. (2014). "The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders." *Genet Med* 16(2): 176-182.
41. Sun, Y., et al. (2015). "Next-generation diagnostics: gene panel, exome, or whole genome?" *Hum Mutat* 36(6): 648-655.
42. Thorvaldsdottir, H., et al. (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." *Brief Bioinform* 14(2): 178-192.
43. Zhang, J., et al. (2011). "The impact of next-generation sequencing on genomics." *J Genet Genomics* 38(3): 95-109.