

Γεφυρώνοντας μεγάλα ομικά, γενετικά και ιατρικά δεδομένα για την ευρεία εφαρμογή της Ιατρικής Ακριβείας στην Ελλάδα

ΠΑΡΑΔΟΤΕΟ Π2.3

«Μέθοδοι εντοπισμού του/των υπευθύνων γονιδίων σε ασθενείς με σπάνια νοσήματα -παρουσίαση σε συνέδριο»

| | |
|--------------------------|--|
| Φορέας | Πανεπιστήμιο Πατρών |
| Τύπος Παραδοτέου | Άλλο - Παρουσίαση σε συνέδριο |
| Τίτλος Παραδοτέου | Μέθοδοι εντοπισμού του/των υπευθύνων γονιδίων σε ασθενείς με σπάνια νοσήματα - παρουσίαση σε συνέδριο |
| Ενότητα Εργασίας | Ενότητα Εργασίας 2: Σπάνια νοσήματα: Ανάλυση δεδομένων αλληλούχισης εξονίων (WES) για την εύρεση του γενετικού υποβάθρου σπανίων νοσημάτων |

Περιεχόμενα

| | |
|---|-----------|
| <i>Εισαγωγή.....</i> | <i>3</i> |
| <i>Στόχος Παρούσας Μελέτης</i> | <i>3</i> |
| <i>Επέκταση κοόρτης ασθενών με σπάνια νοσήματα – συλλογή και ανάλυση δεδομένων Whole Exome Sequencing</i> | <i>3</i> |
| <i>Συνδυαστική ανάλυση πολλαπλών γενετικών τόπων και ομικών δεδομένων</i> | <i>6</i> |
| <i>Αλληλούχηση ολικού γονιδιώματος (Whole Genome Sequencing - WGS)</i> | <i>8</i> |
| <i>Επέκταση και επικαιροποίηση κοόρτης ασθενών με σπάνια νοσήματα.....</i> | <i>12</i> |
| <i>Παρουσίαση σε συνέδριο</i> | <i>14</i> |

Εισαγωγή

Η Αλληλούχιση Επόμενης Γενιάς (Next Generation Sequencing – NGS) αποτελεί μια επαναστατική τεχνολογία στη γονιδιωματική, η οποία έχει προσφέρει σημαντικά οφέλη στην κατανόηση και διάγνωση σπάνιων νοσημάτων. Μέσω της αλληλούχισης όλων των εξονίων του γονιδιώματος (Whole Exome Sequencing – WES) γίνεται δυνατός ο εντοπισμός παραλλαγών σε επίπεδο νουκλεοτιδίου, όπως αντικαταστάσεις, καθώς και μικρών ελλείψεων και προσθηκών (indels) εντός των κωδικοποιητικών περιοχών των γονιδίων. Αντίθετα, για την ανίχνευση σημειακών παραλλαγών εκτός κωδικοποιητικών περιοχών και δομικών παραλλαγών που δεν μπορούν να ταυτοποιηθούν με WES, γίνεται αλληλούχιση ολόκληρου του γονιδιώματος (Whole Genome Sequencing – WGS). Η αλληλούχιση Επόμενης Γενιάς είναι πολύτιμη για τη διάγνωση σπάνιων νοσημάτων, τα οποία στην πλειονότητά τους οφείλονται σε γενετικές παραλλαγές, αλλά εμφανίζουν μεγάλη γενετική και φαινοτυπική ετερογένεια με αποτέλεσμα να παραμένουν αδιάγνωστα επι μακρώ. Με την έγκαιρη διάγνωση μέσω NGS, γίνεται δυνατή στοχευμένη και πιο αποτελεσματική φροντίδα των ασθενών, επιλογή των κατάλληλων θεραπευτικών προσεγγίσεων, καθώς και γενετική συμβουλευτική για τους ίδιους και τις οικογένειές τους. Παρά τη μεγάλη χρησιμότητα των τεχνολογιών αυτών, η εφαρμογή τους στην κλινική πράξη παραμένει περιορισμένη, λόγω του σχετικά υψηλού κόστους τους αλλά κυρίως της δυσκολίας στην ανάλυση των δεδομένων καθώς και της ελλιπούς κατανόησης των ευρημάτων από τους λειτουργούς υγείας και τους ασθενείς.

Στόχος Παρούσας Μελέτης

Στόχος της παρούσας μελέτης είναι η συλλογή και ανάλυση δεδομένων αλληλούχισης νέας γενιάς με έμφαση στην αλληλούχιση όλων των εξονίων (Whole Exome Sequencing, WES), η ανάπτυξη μιας ολοκληρωμένης ροής ανάλυσης των δεδομένων αυτών και πρόβλεψης της παθογονικότητας των παραλλαγών που εντοπίζονται στο γονιδίωμα κάθε ασθενούς, με σκοπό την ταυτοποίηση γονιδίων και γενετικών παραλλαγών που σχετίζονται με σπάνια γενετικά νοσήματα. Μέσω της προσέγγισης αυτής, επιδιώκεται η ανάπτυξη αποδοτικών ροών ανάλυσης και η ανάδειξη της σημασίας των τεχνικών αλληλούχισης, όπως η αλληλούχιση όλων των εξονίων (Whole Exome Sequencing, WES) και η αλληλούχιση όλου του γονιδιώματος (Whole Genome Sequencing, WGS), στη διάγνωση και τη διαχείριση ασθενών με σπάνια γενετικά νοσήματα. Απώτερος στόχος του έργου είναι η διευκόλυνση της ευρείας εφαρμογής των τεχνολογιών αυτών στην καθημερινή κλινική πράξη.

Επέκταση κοόρτης ασθενών με σπάνια νοσήματα – συλλογή και ανάλυση δεδομένων Whole Exome Sequencing

Το Ινστιτούτο Ιατρικής Ακριβείας του Πανεπιστημίου Πατρών υλοποιεί αλληλούχιση όλων των εξονίων (WES) σε ασθενείς με σπάνια νοσήματα στην Ελλάδα, με σκοπό την εύρεση των υπεύθυνων γονιδίων, ένα βήμα καθοριστικό για την πρόγνωση, επιλογή θεραπείας, ένταξη σε κλινικές μελέτες νέων φαρμάκων και γενετική συμβουλευτική. Στα πλαίσια του έργου GoMedPrecision, η αρχική κοόρτη (βλέπε παραδοτέο 2.1) επεκτάθηκε με την πραγματοποίηση Whole exome sequencing σε 65 ασθενείς με σπάνια νοσήματα, οι περισσότεροι από τους οποίους προέρχονται από την περιοχή της Δυτικής Ελλάδας. Ο παρακάτω πίνακας περιλαμβάνει τους συνεργαζόμενους φορείς που παρείχαν δείγματα για την ανάλυση WES στα πλαίσια του έργου.

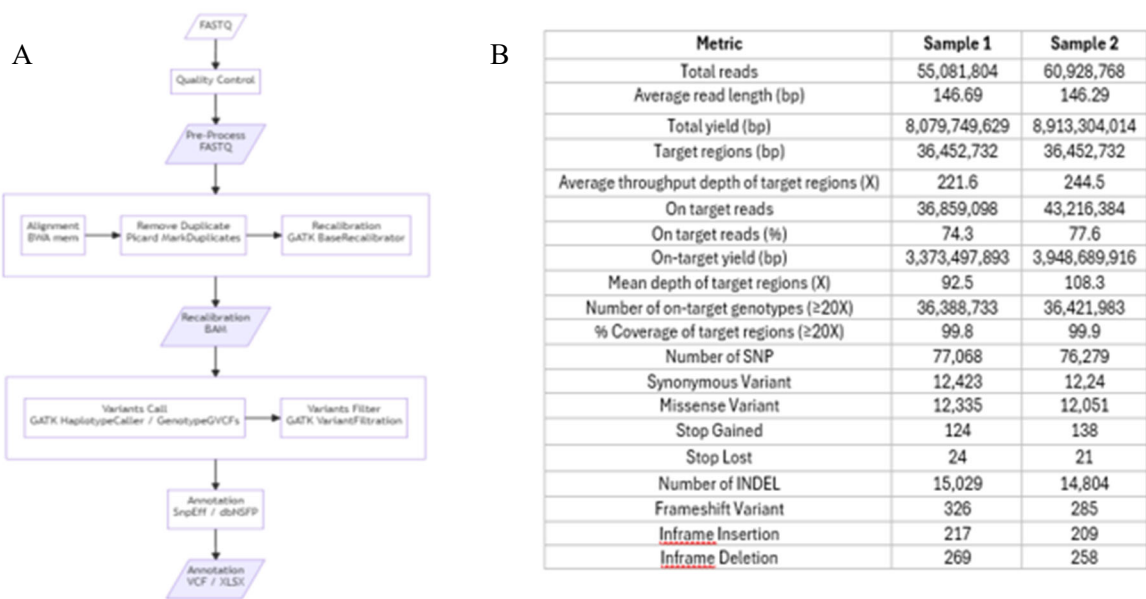
| Συμμετέχουσες Κλινικές Μονάδες |
|--|
| Νευρολογική Κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Πατρών |
| Καρδιολογική Κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Πατρών |
| Πνευμολογική Κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Πατρών |
| Καραμανδάνειο - Γενικό Νοσοκομείο Παίδων Πατρών |
| Νοσοκομείο Παίδων "Η Αγία Σοφία" |
| Ενδοκρινολογική Κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Πατρών |
| Αιματολογική Κλινική "Παπανικολάου" |

Πίνακας 1. Συνεργαζόμενες κλινικές και νοσοκομεία που συμμετείχαν παρέχοντας δείγματα για τις αναλύσεις της μελέτης.

Ιδιαίτερα σημαντική είναι η ενδεδειγμένη περιγραφή των κλινικών φαινοτύπων που εμφανίζει ο κάθε ασθενής. Για το σκοπό αυτό έγινε αναλυτική κλινική εκτίμηση από τους συνεργαζόμενους ιατρούς και για την περιγραφή των φαινοτύπων χρησιμοποιήθηκε η οντολογία Human Phenotypic Ontology (<https://hpo.jax.org/>).

Για την ανάλυση αλληλούχισης όλων των εξωνίων (WES) απομονώθηκε γονιδιωματικό DNA από ολικό περιφερικό αίμα. Η κατασκευή των γονιδιωματικών βιβλιοθηκών και η αλληλούχιση έγινε σε συνεργασία με τις εταιρίες 3billion και MacroGen, εταιρείες με μακρά εμπειρία στο χώρο των γονιδιωματικών αναλύσεων. Για τα περισσότερα περιστατικά, έγινε χρήση του πάνελ Twist Human Core Exome 2.0 (Twist Bioscience), το οποίο καλύπτει 36.5 Mb των κωδικοποιητικών περιοχών του ανθρώπινου γονιδιώματος. Η προετοιμασία της βιβλιοθήκης πραγματοποιήθηκε με τη χρήση εμπορικά διαθέσιμου κιτ της εταιρείας Twist Library Preparation EF Kit (96 samples, PN 101058). Η αλληλούχιση πραγματοποιήθηκε με τη χρήση του συστήματος Illumina NovaSeq X (PE150). Οι αναγνώσεις που δημιουργήθηκαν, συγκρίθηκαν με το Genome Reference Consortium Human Build 38 (GRCh38) και ευθυγραμμίστηκαν μέσω του αλγορίθμου BWA-MEM. Η κλήση των παραλλαγών πραγματοποιήθηκε μέσω του αλγορίθμου HaploType Caller από το Genome Analysis Toolkit.

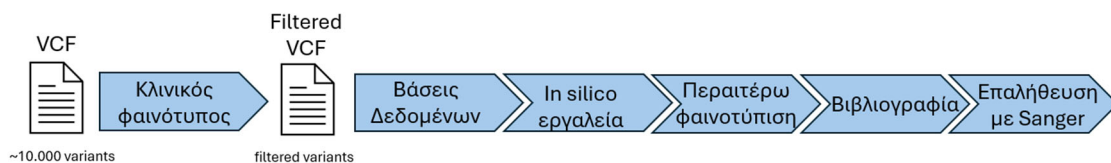
Η ανάλυση των αποτελεσμάτων καθώς και η κλινική επισημείωση πραγματοποιήθηκε με τη χρήση της διαδικτυακής πλατφόρμας Franklin (<https://franklin.genoox.com> - Franklin by Genoox Version 84.1) η οποία διαθέτει άμεση διασύνδεση με βάσεις δεδομένων όπως οι ClinVar, ExAC, gnomAD, OMIM και ClinGen. Επιπλέον, η συσχέτιση μεταξύ γονιδίων, νόσων και φαινοτύπων, πραγματοποιείται αντλώντας πληροφορίες από τις βάσεις δεδομένων: Orphanet, Monarch, HPO και DECIPHER. Η ερμηνεία των παραλλαγών πραγματοποιήθηκε με βάση τις κατευθυντήριες γραμμές των ACMG/AMP και τις βάσεις δεδομένων ClinVar και VarSome, στο πλαίσιο του φαινότυπου και του σχετικού οικογενειακού ιστορικού του ασθενούς. Στην Εικόνα 1 (Α) παρουσιάζεται συνοπτικά η ροή ανάλυσης που πραγματοποιείται κατά την αλληλούχιση και αλγόριθμοι που χρησιμοποιούνται καθώς επίσης (Β) οι τυπικές τιμές διάφορων μετρικών ποιότητας και κάλυψης που υποδεικνύουν την άριστη απόδοση και την υψηλή αξιοπιστία ανίχνευσης παραλλαγών.



Εικόνα 1. (Α) Ροή εργασιών αλληλούχισης, (Β) Μετρικές απόδοσης αλληλούχισης

Οι αναλύσεις εστίαστηκαν αρχικά σε αλληλούχιση του γονιδιώματος μόνο του ασθενούς (proband) και όχι και μελών της οικογένειάς του, μια προσέγγιση που θεωρείται η λιγότερο κοστοβόρα για την ευρύτερη δυνατή κάλυψη του μέγιστου αριθμού ασθενών.

Στην Εικόνα 2 περιγράφεται συνοπτικά η ροή εργασιών που πραγματοποιείται κατά την ανάλυση των δεδομένων. Η ανάλυση των δεδομένων από αλληλούχιση εξονίων (WES) ξεκινά συνήθως με ένα αρχείο VCF (Variant Calling File), το οποίο περιέχει χιλιάδες γενετικές παραλλαγές, πολλές εκ των οποίων δεν έχουν κλινική σημασία. Για τον εντοπισμό των παραλλαγών που μπορεί να σχετίζονται με την εκάστοτε νόσο ή φαινοτυπικό χαρακτηριστικό του εξεταζόμενου, εφαρμόζεται φιλτράρισμα βασισμένο στον φαινότυπο. Η διαδικασία αυτή περιλαμβάνει την αντιστοίχιση των φαινοτυπικών χαρακτηριστικών του ασθενούς σε βάσεις δεδομένων και γνώση σχετικά με γονίδια και παραλλαγές που έχουν αναφερθεί σε παρόμοιους κλινικούς φαινότυπους. Παραλλαγές που δεν συνδέονται με τα φαινοτυπικά χαρακτηριστικά αποκλείονται, μειώνοντας σημαντικά τον αριθμό των πιθανών παραλλαγών που είναι υπεύθυνες για το νόσημα. Το φιλτράρισμα με βάση τον φαινότυπο επιτρέπει, επομένως, την εστίαση σε παραλλαγές με υψηλότερη πιθανότητα κλινικής σημασίας, αυξάνοντας την ακρίβεια και την αποδοτικότητα της ερμηνείας των αποτελεσμάτων. Στη συνέχεια της ανάλυσης, κάθε παραλλαγή ελέγχεται για ποιοτικά χαρακτηριστικά, όπως το βάθος ανάγνωσης, η συνολική κάλυψη, η ποιότητα ευθυγράμμισης των reads και η ποιότητα βάσης, προκειμένου να διασφαλιστεί η αξιοπιστία της κλήσης. Ακολουθεί αναζήτηση της παραλλαγής σε βάσεις δεδομένων και βιβλιογραφική διερεύνηση, καθώς και αξιολόγηση με *in silico* εργαλεία πρόβλεψης, όπως AlphaMissense και SpliceAI, για την εκτίμηση της πιθανής λειτουργικής επίδρασης. Στη συνέχεια πραγματοποιείται περαιτέρω φαινοτύπωση του ασθενούς, ώστε να διαπιστωθεί η συμφωνία της παραλλαγής με τα κλινικά χαρακτηριστικά. Τέλος, τα ευρήματα επαληθεύονται με στοχευμένη αλληλούχιση Sanger, διασφαλίζοντας την ακρίβεια και την κλινική αξιοπιστία της ανάλυσης.



Εικόνα 2. Ροή εργασιών ανάλυσης δεδομένων

Παρακάτω παρουσιάζεται ένα ενδεικτικό παράδειγμα ανάλυσης δεδομένων από ασθενή ο οποίος εμφάνιζε μυϊκή αδυναμία των προσωπικών μυών, σημαντικά αυξημένα επίπεδα CPK, μειωμένη ικανότητα βάδισης στις μύτες των ποδιών, αναπνευστική ανεπάρκεια με επεισόδια άπνοιας, πτερυγοειδή ωμοπλάτη και αδυναμία των εγγύς μυϊκών ομάδων. Αρχικά πραγματοποιήθηκε φιλτράρισμα των παραλλαγών με βάση τον κλινικό φαινότυπο, από το οποίο προέκυψαν εννέα υποψήφιες παραλλαγές, όπως παρουσιάζονται στον Πίνακα 2. Ακολούθησε η αξιολόγηση της παθογονικότητάς τους μέσω αναζήτησης σε βάσεις δεδομένων, καθώς και η διερεύνησή τους με βιοπληροφορικά εργαλεία πρόβλεψης παθογονικότητας. Στη συνέχεια πραγματοποιήθηκε περαιτέρω φαινοτυπική συσχέτιση, σε συνδυασμό με εκτενή βιβλιογραφική ανασκόπηση σχετικά με τα εμπλεκόμενα γονίδια και τις συγκεκριμένες παραλλαγές. Μετά την ανάδειξη της πιθανώς αιτιολογικής παραλλαγής, εφαρμόστηκε στοχευμένη αλληλούχηση κατά Sanger για την επιβεβαίωση του ευρήματος, διασφαλίζοντας την αναλυτική εγκυρότητα του αποτελέσματος.

| Phenotype | Gene | Variant | Zygoty | Inheritance | Classification | Connection | Alphamissense | Disease |
|---|---------|--|--------|-------------|----------------------|--|---------------|--|
| Weakness of facial musculature, Highly elevated CPK, Impaired toe-walking ability, Respiratory insufficiency, Apnea, Scapular winging, Proximal muscle weakness | SELENON | NM_020451.3(SELENON):c.713dup (p.Asn238Lysfs*63) | Hom | AR | P | Impaired toewalking ability, Proximal muscle weakness and 5 more (Very High) | | Rigid spine muscular dystrophy 1 (Autosomal Recessive) |
| | AARS | NM_001605.3(AARS):c.1538_1540del (p.Lys513del) | Het | AD | VUS(PM2,PM4) | Proximal muscle weakness, Impaired toewalking ability and 2 more (High) | | Charcot-marie-tooth disease axonal type 2n |
| | TTN | NM_001267550.2(TTN):c.71452A>G (p.Ile23818Val) | Het | AD | LB | Impaired toewalking ability, Proximal muscle weakness and 6 more (Very High) | B:0.049 | Myopathy, myofibrillar, 9, with early respiratory failure (Autosomal Dominant) |
| | | NM_001267550.2(TTN):c.17338T>C (p.Phe5780Leu) | Het | AD | - | Impaired toewalking ability, Proximal muscle weakness and 6 more (Very High) | P:0.943 | Myopathy, myofibrillar, 9, with early respiratory failure (Autosomal Dominant) |
| | PRODH | NM_016335.6(PRODH):c.554G>A (p.Trp185*) | Het | AD | VUS(PVS1,BP6) | Muscle weakness (Low) | | Schizophrenia 4 (Autosomal Dominant) |
| | ASAH1 | NM_177924.5(ASAH1):c.592G>A (p.Val198Ile) | Het | AR | VUS(PM2,PM5,PP2,BP4) | Impaired toewalking ability, Proximal muscle weakness and 6 more (Very High) | LB:0.076 | Farber lipogranulomatosis |
| | C9 | NM_001737.5(C9):c.355T>G (p.Cys119Gly) | Het | AR/AD | LP(PP3,PP5,PM2,BP1) | not matching phenotype | LP:0.923 | |
| | SCN10A | NM_006514.4(SCN10A):c.955T>A (p.Cys319Ser) | Het | AD | VUS(PP3,PM2,BP1) | not matching phenotype | LP:0.995 | |
| | RP1L1 | NM_178857.6(RP1L1):c.1107G>A (p.Trp369*) | Het | AR | LP(PVS1,PM2) | not matching phenotype | | |

Πίνακας 2. Λίστα παραλλαγών έπειτα από φιλτράρισμα βάσει κλινικού φαινότυπου.

Συνδυαστική ανάλυση πολλαπλών γενετικών τόπων και ομικών δεδομένων

Αν και τα σπάνια νοσήματα θεωρούνται στις περισσότερες περιπτώσεις μονογονιδιακά, η ύπαρξη περισσότερων της μιας γενετικής παραλλαγής σε διαφορετικά γονίδια μπορεί αθροιστικά να οδηγούν στην εκδήλωση του προς μελέτη κλινικού φαινότυπου. Παράλληλα έχει αναγνωριστεί η σημασία παραλλαγών σε διαφορετικούς γενετικούς τόπους που επηρεάζουν ή και καθορίζουν την εκδήλωση του κύριου γενετικού τόπου ο οποίος συνδέεται με το νόσημα (τροποποιητικά γονίδια). Επίσης, συχνά απαντώμενες παραλλαγές (πολυμορφισμοί) που προδιαθέτουν για την εκδήλωση νοσημάτων μπορεί σε συνδυασμό με άλλους πολυμορφισμούς ή με σπάνιες παραλλαγές να εξηγούν τον κλινικό φαινότυπο του ασθενούς. Η ταυτόχρονη μελέτη πολλαπλών γενετικών τόπων μπορεί να οδηγήσει στον εντοπισμό παραλλαγών που η ταυτόχρονη εμφάνισή τους είναι υπεύθυνη για την πρόκληση κάποιας ασθένειας, ή δύναται να εξηγήσει τον εμφανιζόμενο κλινικό φαινότυπο του ασθενούς.

Στην κόρτη ασθενών της συγκεκριμένης μελέτης ανιχνεύθηκαν περιπτώσεις όπου η συμπτωματολογία των ασθενών ενδέχεται να εξηγείται από συνδυασμό πολλαπλών γενετικών τόπων καθώς και πολυμορφισμών που προδιαθέτουν για εμφάνιση νοσήματος. Για το σκοπό αυτό, εφαρμόσαμε συνδυαστική ανάλυση των παραλλαγών που ανευρίσκονται σε κάθε ασθενή.

Μελετήσαμε δύο περιστατικά με διατακτική μυοκαρδιοπάθεια, στα οποία ανιχνεύθηκαν παραλλαγές σε δύο γονίδια που έχουν συσχετιστεί με το νόσημα σε κάθε ασθενή. Οι περιπτώσεις αυτές παρουσιάζονται στον Πίνακα 3. Για να διερευνήσουμε εάν οι συγκεκριμένες παραλλαγές αθροιστικά μπορούν να οδηγήσουν στην εκδήλωση των αντίστοιχων φαινοτύπων χρησιμοποιήσαμε το εργαλείο VarCoPP2, το οποίο προβλέπει την πιθανή παθογένεια του συνδυασμού δύο ή περισσότερων παραλλαγών.

| A/A | 1 ^η Παραλλαγή | | | 2 ^η Παραλλαγή | | | VarCoPP2 |
|-----|--------------------------|--------------------------|-----|--------------------------|-----------------------|-----|----------|
| | Γονίδιο | Όνομα Παραλλαγής | | Γονίδιο | Όνομα Παραλλαγής | | |
| 1 | MYH6 | NM_002471.4:c.1408G>T | VUS | MYH7 | NM_000257.4:c.4348G>A | VUS | 0.98 |
| 2 | LAMA4 | NM_001105206.3:c.4492C>T | VUS | LMNA | NM_170707.4:c.1129C>T | P | 0.89 |

Πίνακας 3: Περιστατικά όπου εντοπίστηκαν δύο παραλλαγές.

Το 1^ο περιστατικό αφορά έναν ασθενή που πάσχει από Διατακτική Μυοκαρδιοπαθεία στον οποίο εντοπίστηκαν παραλλαγές στα γονίδια MYH6 και MYH7. Η βιβλιογραφική αναζήτηση ανέδειξε την δημοσίευση των Suzuki et al. όπου αναφέρεται ότι η ταυτόχρονη ύπαρξη παραλλαγών στα γονίδια MYH6 και MYH7 είναι υπεύθυνη για την εκδήλωση Υπερτροφικής Μυοκαρδιοπάθειας. Το σκορ που προέκυψε από το VarCoPP2 είναι 0.98, το οποίο συμφωνεί με την υπόθεση ότι ο συνδυασμός αυτών των παραλλαγών ενδέχεται να είναι υπεύθυνος για την εκδήλωση του εμφανιζόμενου κλινικού φαινοτύπου.

Το 2^ο περιστατικό αφορά ασθενή που πάσχει επίσης από Διατακτική Μυοκαρδιοπαθεία, στον οποίο εντοπίστηκαν παραλλαγές στα γονίδια LAMA4 και LMNA, δύο γονίδια που συνδέονται με το μελετούμενο φαινότυπο. Το σκορ που προέκυψε από το VarCoPP2 είναι 0.89, το οποίο συμφωνεί με την υπόθεσή ότι ο συνδυασμός αυτών των παραλλαγών ενδέχεται να είναι υπεύθυνος για την εκδήλωση του φαινοτύπου.

Παρακάτω παρουσιάζεται μια περίπτωση συνδυασμού πολυμορφισμών σε HLA γονίδιο που μπορεί να συνδέονται με τον κλινικό φαινότυπο του προς διάγνωση ασθενούς. Σε περιστατικό με παραπομπή για διερεύνηση νευροπάθειας και διαταραχής του αυτόνομου νευρικού συστήματος προέκυψαν 3 ομόζυγες παραλλαγές στο γονίδιο HLA-DQA1 που εδράζονται στο εξόνιο 2, στο εξωκυττάριο τμήμα της πρωτεΐνης. Αναλυτικότερα στον Πίνακα 4 παρουσιάζονται οι απαντώμενες παραλλαγές στο γονίδιο.

| Gene | Variant | Zygoty | Inheritance | Classification | Connection | Alphamissense | Disease |
|----------|------------------------------------|--------|-------------|----------------|------------------------|---------------|--|
| HLA-DQA1 | NM_002122.5:c.250G>T (p.Gly84Cys) | Hom | AR/Mu | VUS | Neuropathy (Very High) | - | Susceptibility to celiac disease 1 (Mu,AR) |
| | NM_002122.5:c.260G>C (p.Arg87Thr) | Hom | AR/Mu | VUS | Neuropathy (Very High) | - | |
| | NM_002122.5:c.308A>C (p.Tyr103Ser) | Hom | AR/Mu | VUS | Neuropathy (Very High) | - | |

Πίνακας 4. Παραλλαγές στο γονίδιο HLA-DQA1

Παραλλαγές στο γονίδιο HLA-DQA1 έχουν συσχετιστεί με γενετική προδιάθεση για δυσανεξία στη γλουτένη, ενώ η δυσανεξία στη γλουτένη έχει συσχετισθεί με εμφάνιση περιφερικής νευροπάθειας. Οι συγκεκριμένες παραλλαγές που ανιχνεύθηκαν δεν έχουν μελετηθεί ως προς την επίπτωσή τους στη δυσανεξία στη γλουτένη ή/και την εκδήλωση νευροπάθειας, αλλά εδράζονται σε περιοχή που γειτνιάζει με γνωστές παραλλαγές. Το εύρημα αυτό οδηγεί στην υπόθεση εργασίας ότι οι παραλλαγές στο HLA-DQA1 θα μπορούσαν να συνεισφέρουν από κοινού στη νευρολογική συμπτωματολογία του ασθενούς.

Συνδυαστική ανάλυση ομικών δεδομένων

Με το συνδυασμό διαφορετικών ομικών δεδομένων είναι δυνατή η πληρέστερη κατανόηση του τρόπου λειτουργίας των βιολογικών συστημάτων και μηχανισμών και πώς αυτοί επηρεάζονται από την ύπαρξη γενετικών παραλλαγών. Σε μια προσπάθεια να κατανοήσουμε τους παράγοντες που οδηγούν στην εμφάνιση διαφορετικών ασθενειών, όταν δεν έχει εντοπιστεί κάποια γενετική παραλλαγή από την ανάλυση WES, πραγματοποιήθηκε μεταβολομική ανάλυση με Nuclear Magnetic Resonance spectroscopy (NMR).

Σε δύο ασθενείς που πάσχουν από υπερχοληστερολαιμία, στους οποίους όμως δεν εντοπίστηκε κάποια παραλλαγή μέσω WES, πραγματοποιήθηκε μεταβολομική ανάλυση με NMR. Για αυτήν την ανάλυση, έγινε λήψη ολικού αίματος (πλάσμα και ορός) και ούρων τα οποία χρησιμοποιήθηκαν για τον εντοπισμό των μεταβολιτών και έγινε σύγκριση με τους μεταβολίτες που εντοπίζονται σε άτομα μάρτυρες. Περισσότερα για τη μεταβολομική ανάλυση παρουσιάζονται στο Παραδοτέο ΠΕ 2.2. Υπολογιστικές μέθοδοι συνδυασμού μεταβολιτών και πολυμορφισμών για κάθε ασθενή με σκοπό τη διάγνωση υπερχοληστερολαιμίας και διαβήτη παρουσιάζονται στην ενότητα εργασίας 6.

Αλληλούχιση ολικού γονιδιώματος (Whole Genome Sequencing - WGS)

Η ανάλυση WGS, παρά το υψηλότερο κόστος που απαιτεί για την υλοποίηση της, αποτελεί την τεχνική NGS με την καλύτερη διαγνωστική απόδοση. Μέσω της WGS μπορούν να εντοπιστούν παραλλαγές σε μη-κωδικοποιητικές περιοχές του γονιδιώματος, ενώ παράλληλα διευκολύνεται η ανίχνευση δομικών παραλλαγών (Structural Variations – SVs) και παραλλαγών αριθμού αντιγράφων (Copy Number Variations – CNVs).

Σε αυτό το κομμάτι της μελέτης επιλέχθηκαν δεκαπέντε ασθενείς στους οποίους πραγματοποιήθηκε WGS (Πίνακας 5). Οι ασθενείς αυτοί εμφανίζουν σαφές πρότυπο γενετικού νοσήματος και ισχυρό φαινότυπο στους οποίους όμως δεν εντοπίστηκε κάποια παραλλαγή που να συνδέεται με το φαινότυπό τους, μέσω WES ή panel sequencing. Σε κάποιες περιπτώσεις επιλέχθηκε να αναλυθεί μόνο το πάσχον άτομο (proband), ενώ σε τρεις περιπτώσεις στη μελέτη συμπεριλήφθηκαν συγγενείς 1^{ου} βαθμού, όπως φαίνονται και σκιαγραφημένοι στον Πίνακα 4. Συγκεκριμένα, οι περιπτώσεις 1 και 2 αφορούν πατέρα και κόρη πάσχοντες, οι 7,8 και 9 πάσχοντα αδέρφια και την υγιή μητέρα τους, ενώ οι περιπτώσεις 14 και 15 αφορούν πάσχοντα αδέρφια.

| A/A | Φύλο | Φαινότυπος |
|-----|---------|--|
| 1 | Άνδρας | Υπερχοληστερολαιμία <i>Αρνητικό WES</i> |
| 2 | Γυναίκα | Υπερχοληστερολαιμία <i>Αρνητικό WES</i> |
| 3 | Άνδρας | Υποτροπιάζουσες λοιμώξεις του αναπνευστικού συστήματος <i>Αρνητικό WES</i> |
| 4 | Γυναίκα | Διαταραχή βάδισης, Βάδισμα ευρείας βάσης, Υπερανακλαστικότητα των κάτω άκρων, Δυστονία των άκρων, Περικοιλιακές υποπυκνότητες λευκής ουσίας <i>Αρνητικό WES</i> |
| 5 | Γυναίκα | Αναπτυξιακή καθυστέρηση, Μυοκλονικοί σπασμοί, Μικροκεφαλία, Στοματοπροσωπική απραξία <i>Αρνητικό WES</i> |

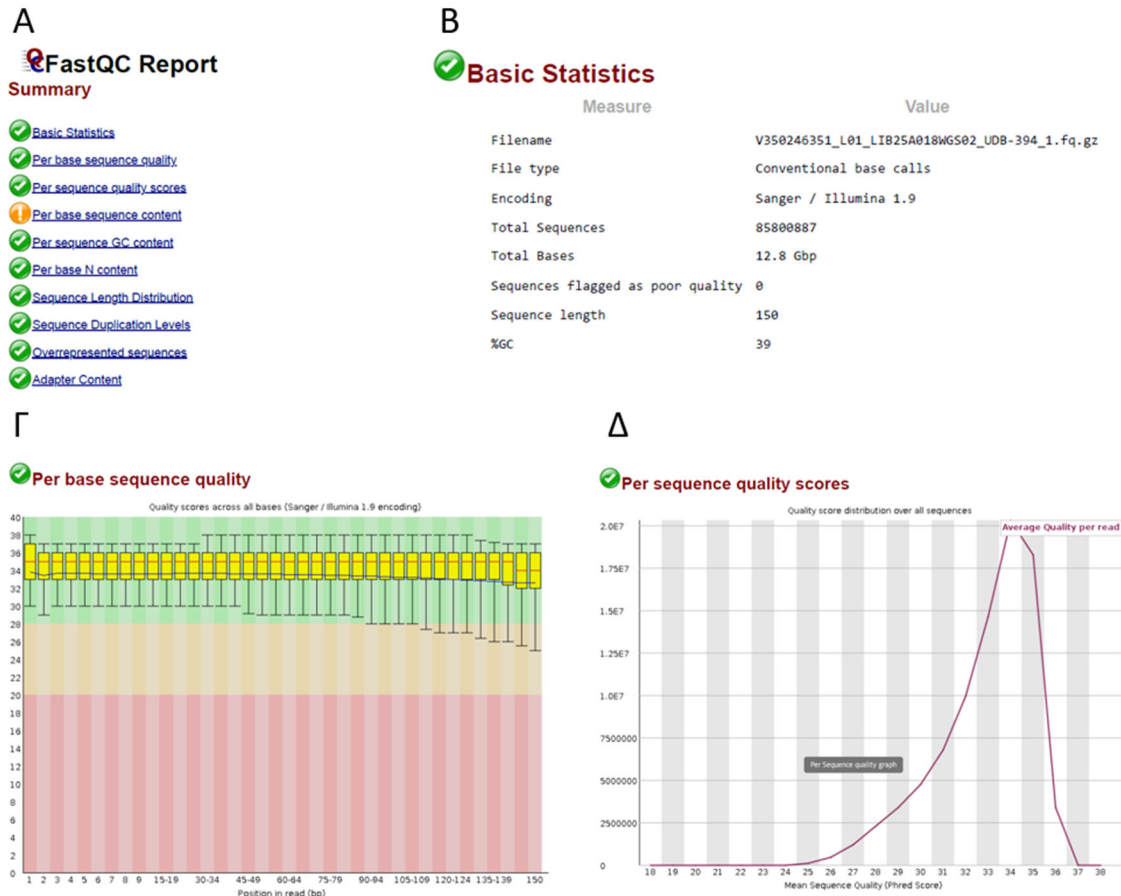
| | | |
|----|---------|--|
| 6 | Γυναίκα | Ηπατική ανεπάρκεια, Διαταραχή βιάδισης, Νοητική στέρση, Δυστονία, Εμφάνιση συμπτωμάτων κατά τη βρεφική ηλικία |
| 7 | Άνδρας | Σοβαρή νοητική αναπτυξιακή διαταραχή, Σοβαρή καθολική αναπτυξιακή καθυστέρηση, Γενικευμένες κρίσεις |
| 8 | Άνδρας | Σοβαρή νοητική αναπτυξιακή διαταραχή, Σοβαρή καθολική αναπτυξιακή καθυστέρηση, Γενικευμένες κρίσεις |
| 9 | Γυναίκα | Χωρίς φαινότυπο |
| 10 | Άνδρας | Δυστονία, Εγκεφαλική παράλυση, Νεφρική δυσπλασία, Μεταμόσχευση νεφρού, Ανωμαλία του μεσολοβίου, Υποθυρεοειδισμός <i>Αρνητικό WES</i> |
| 11 | Άνδρας | Υποτονία, Απουσία αντανακλαστικών, Άπνοια, Δυσφαγία <i>Αρνητικό WES</i> |
| 12 | Άνδρας | Σοβαρή νοητική αναπτυξιακή διαταραχή, Καθυστερημένα κινητικά ορόσημα, Καθυστερημένη ανάπτυξη λόγου/ομιλίας, Πυρετικοί σπασμοί, Γενικευμένη υποτονία, Υπερθερμία |
| 13 | Άνδρας | Σπασμοί, Επιληψία ανθεκτική σε φαρμακευτική αγωγή |
| 14 | Άνδρας | Σοβαρή νοητική αναπτυξιακή διαταραχή, Μαθησιακές δυσκολίες, Αυτισμός |
| 15 | Άνδρας | Σοβαρή νοητική αναπτυξιακή διαταραχή, Σοβαρή γενικευμένη αναπτυξιακή καθυστέρηση, Καθυστερημένα κινητικά ορόσημα, Καθυστερημένη ανάπτυξη λόγου/ομιλίας, Μαθησιακές δυσκολίες, Αυτισμός |

Πίνακας 5. Ασθενείς στους οποίους πραγματοποιήθηκε Whole Genome Sequencing.

Από ολικό περιφερικό αίμα απομονώθηκε γονιδιωματικό DNA. Η κατασκευή των γονιδιωματικών βιβλιοθηκών, η αλληλούχιση και η βιοπληροφορική ανάλυση των δεδομένων WGS έγινε σε συνεργασία με το Ινστιτούτο Φλέμινγκ (βλέπε και παραδοτέο 9.2). Μετά την ολοκλήρωση της αλληλούχισης γίνεται αξιολόγηση της ποιότητας των παραχθέντων δεδομένων.

Η αξιολόγηση της αλληλούχισης Whole Genome Sequencing με τεχνολογία αλληλούχισης Illumina έγινε εξετάζοντας διαφορετικές μετρικές ποιότητας πριν και μετά την ευθυγράμμιση με το ανθρώπινο γονιδίωμα αναφοράς GR37 (hg19). Αρχικά με την χρήση του εργαλείου FastQC (version 0.12.1) πραγματοποιείται ποιοτικός έλεγχος των ακατέργαστων δεδομένων αλληλούχισης (raw reads) και αξιολογούνται βασικά χαρακτηριστικά όπως η κατανομή της ποιότητας βάσεων, το GC content, η παρουσία adapters, οι διπλότυπες ακολουθίες και πιθανά τεχνικά artifacts. Με αυτόν τον τρόπο παράγονται αναλυτικά reports ανά δείγμα, τα οποία επιτρέπουν την εκτίμηση της καταλληλότητας των δεδομένων για περαιτέρω επεξεργασία (π.χ. trimming, filtering) πριν από τη χαρτογράφηση και την ανάλυση.

Παρακάτω, παρουσιάζονται τα αποτελέσματα για το πρώτο read (read 1) ενός paired-end πρωτοκόλλου (Εικόνα 3). Το αρχείο περιλαμβάνει 85.800.887 reads, συνολικού μήκους 12,8 Gb, με σταθερό μήκος ανά read 150 bp. Δεν έχουν επισημανθεί reads χαμηλής ποιότητας στο στάδιο του base calling, ενώ το ποσοστό GC περιεχομένου (39%) βρίσκεται εντός των αναμενόμενων τιμών για το ανθρώπινο γονιδίωμα, υποδηλώνοντας καλή ποιότητα βιβλιοθήκης. Συνολικά, τα δεδομένα χαρακτηρίζονται ως κατάλληλα για την βιοπληροφορική ανάλυση, συμπεριλαμβανομένης της ευθυγράμμισης, της κλήσης παραλλαγών και της λειτουργικής/κλινικής ερμηνείας.

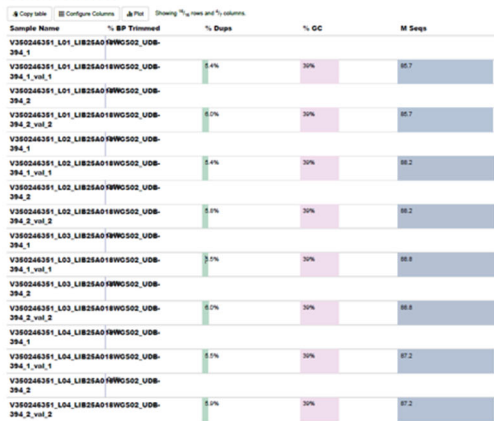


Εικόνα 3: Παράδειγμα ποιοτικού ελέγχου αλληλούχισης Whole Genome Sequencing. (Α) Όλα τα μετρικά στοιχεία που ελέγχονται κατά τον ποιοτικό έλεγχο των δεδομένων. (Β) Βασικά στατιστικά στοιχεία της αλληλούχισης. (Γ) Γράφημα στο οποίο φαίνεται η ποιότητα αλληλουχίας ανά βάση. (Δ) Γράφημα στο οποίο φαίνεται το σκορ ποιότητας ανά αλληλουχία.

Στην συνέχεια, χρησιμοποιείται το MultiQC Report ως μετα-εργαλείο που συγκεντρώνει και συνοψίζει τα αποτελέσματα του FastQC και άλλων εργαλείων σε ένα ενιαίο report, διευκολύνοντας τη σύγκριση πολλών δειγμάτων και τον συνολικό έλεγχο ποιότητας ενός pipeline. Το κάθε δείγμα έχει 4 lanes και το καθένα διαθέτει 2 reads (read 1, read 2). Ενδεικτικά παρουσιάζονται (Εικόνα 4) τα αποτελέσματα ενός δείγματος και παρουσιάζεται το ποσοστό των βάσεων που αφαιρέθηκαν με trimming (0.4–0.6%), το ποσοστό διπλοτύπων (5.4–6.0%), το GC content (39%) και τον αριθμό reads σε εκατομμύρια (M Seqs, 85–88 M ανά αρχείο). Οι τιμές αυτές δείχνουν ότι οι βιβλιοθήκες είναι υψηλής ποιότητας, με χαμηλό ποσοστό διπλοτύπων, ελάχιστο trimming και φυσιολογικό GC περιεχόμενο. Η εφαρμογή του Cutadapt (v3.5) χρησιμοποιείται για την αφαίρεση adapters, primers ή poly-A tails από το 3' άκρο έδειξε ότι η πλειονότητα των reads παρέμεινε σταθερή, με ελάχιστο αριθμό reads να αφαιρείται.

A

General Statistics



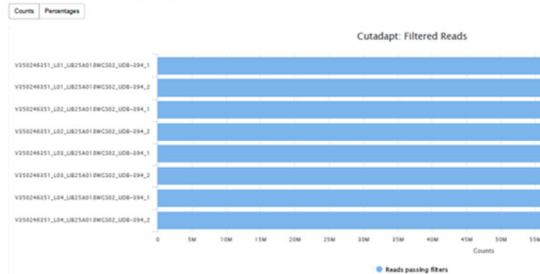
B

Cutadapt

Version: 3.5
Cutadapt is a tool to find and remove adapter sequences, primers, poly-A tails and other types of unwanted sequence from your high-throughput sequencing reads. DOI: 10.14806/q1.17.1.200

Filtered Reads

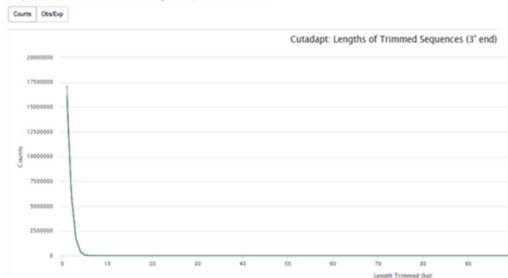
This plot shows the number of reads (SE) / pairs (PE) removed by Cutadapt.



Γ

Trimmed Sequence Lengths (3')

This plot shows the number of reads with certain lengths of adapter trimmed for the 3' end.



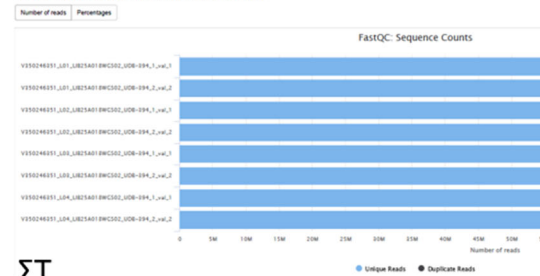
Δ

FastQC

Version: 0.12.1
FastQC is a quality control tool for high throughput sequence data, written by Simon Andrews at the Babraham Institute in Cambridge.

Sequence Counts

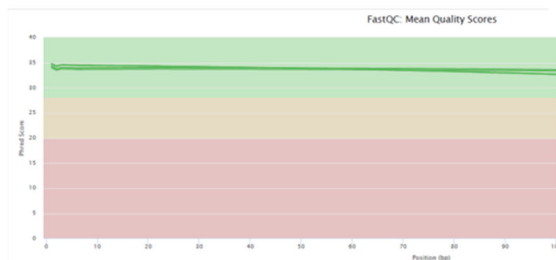
Sequence counts for each sample. Duplicate read counts are an estimate only.



E

Sequence Quality Histograms

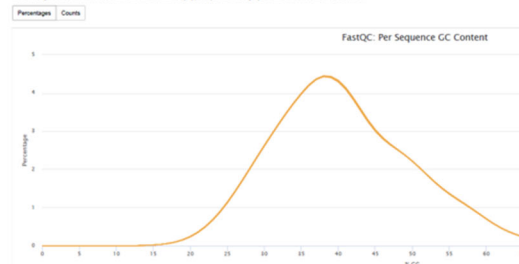
The mean quality value across each base position in the read.



ΣΤ

Per Sequence GC Content

The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content.

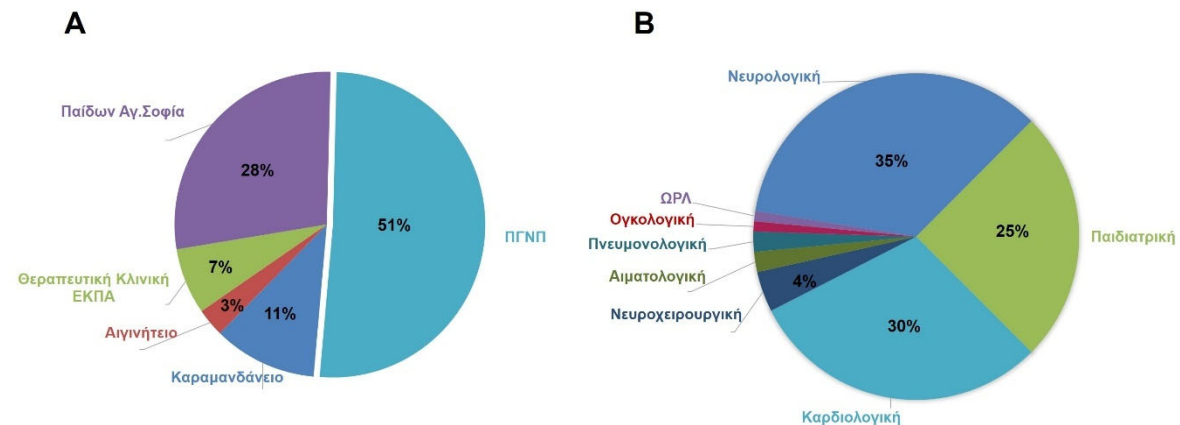


Εικόνα 4: Αποτελέσματα MultiQC Report. (Α) Γενικά στατιστικά στοιχεία πολλαπλών δειγμάτων. (Β) Γράφημα στο οποίο παρουσιάζεται ο αριθμός των βάσεων που αφαιρέθηκαν από το εργαλείο Cutadapt. (Γ) Γράφημα στο οποίο παρουσιάζεται το ποσοστό των βάσεων που αφαιρέθηκαν (trimming) στο 3' άκρο. (Δ) Γράφημα στο οποίο παρουσιάζεται το ποσό των διπλοτύπων. (Ε) Ιστόγραμμα στο οποίο εμφανίζεται η ποσότητα των βιβλιοθηκών. (ΣΤ) Γράφημα στο οποίο παρουσιάζεται το μέσο περιεχόμενο GC των αλληλουχιμένων βάσεων.

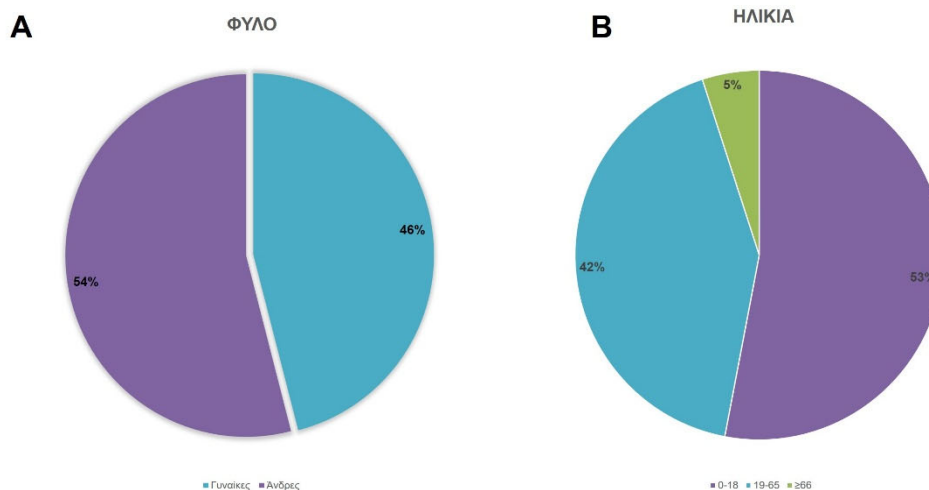
Επέκταση και επικαιροποίηση κόορτης ασθενών με σπάνια νοσήματα

Οι αναλύσεις που περιγράφονται ανωτέρω, καθώς και στα παραδοτέα 2.1 και 2.2, επέτρεψαν την επέκταση της κόορτης των ασθενών με σπάνια νοσήματα από τον ελληνικό πληθυσμό, που έχουν αναλυθεί με μεθόδους αλληλούχισης νέας γενιάς.

Στη συνέχεια παρουσιάζονται τα στοιχεία από την επικαιροποιημένη κόορτη ασθενών που έχει πλέον δημιουργηθεί και αφορά ένα σύνολο 465 ασθενών. Η κόορτη αποτελείται τόσο από παιδιατρικούς όσο και από ενήλικους ασθενείς, από διαφορετικά νοσηλευτικά ιδρύματα και κλινικές της χώρας (Εικόνα 5). Συμπεριλαμβάνει ποικίλους φαινοτύπους όπως νευροαναπτυξιακές και νευρολογικές/νευρομυϊκές διαταραχές, διαταραχές ανάπτυξης, μεταβολικές, καρδιαγγειακές και σκελετικές ανωμαλίες.

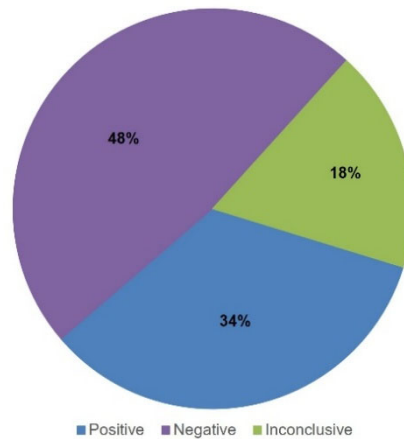


Εικόνα 5: Συμμετέχοντα νοσηλευτικά ιδρύματα και κλινικές. (Α) Ποσοστό εγγεγραμμένων ασθενών ανά νοσηλευτικό ίδρυμα. (Β) Ποσοστό εγγεγραμμένων ασθενών ανά κλινική του Πανεπιστημιακού Γενικού Νοσοκομείου Πατρών (ΠΓΝΠ). 35% των ασθενών προέρχεται από τη Νευρολογική Κλινική, 30% από την Καρδιολογική, 25% από την Παιδιατρική, 4% από τη Νευροχειρουργική, 2% από την Αιματολογική και Πνευμονολογική Κλινική αντίστοιχα και τέλος 1% από την Κλινική ΩΡΛ και την Ογκολογική αντίστοιχα.



Εικόνα 6: Δημογραφικά στοιχεία συμμετεχόντων ασθενών. (Α) Ποσοστό συμμετεχόντων ανδρών και γυναικών. (Β) Ηλικιακή κατανομή συμμετεχόντων.

Η ανάλυση των δεδομένων WES προσδιόρισε την υποκείμενη γενετική αιτία στο 34% των περιπτώσεων, παρέχοντας ένα θετικό αποτέλεσμα. Επιπλέον, στο 18% των περιπτώσεων, εντοπίστηκε παραλλαγή αγνώστου κλινικής σημασίας (Variant of Uncertain Significance – VUS), δίνοντας ένα μη καθοριστικό αποτέλεσμα, ενώ στο 48% των περιστατικών δεν εντοπίζονται σχετιζόμενες παραλλαγές υπογραμμίζοντας την ανάγκη περαιτέρω ανάλυσης (Εικόνα 7). Μέσω ανάλυσης σε οικογένειες, βαθύτερη φαινοτύπιση και βελτιωμένους αλγορίθμους ανάλυσης έγινε δυνατή η επανακατηγοριοποίηση 5 παραλλαγών, με αποτέλεσμα να δοθεί τελική διάγνωση σε 7 επιπλέον ασθενείς.



Εικόνα 7: Αποτελέσματα WES. Το 34% των περιστατικών ήταν θετικά, το 18% ήταν ασαφή (εντοπίστηκαν παραλλαγές αγνώστου κλινικής ή Παθογόνες (Pathogenic) /Πιθανώς Παθογόνες (Likely Pathogenic) παραλλαγές σε γονίδια αγνώστου σημασίας για το εκάστοτε νόσημα ή σε ετεροζυγωτία σε αυτοσωμικές υπολειπόμενες ασθένειες) και το 48% ήταν αρνητικά.

Παρουσίαση σε συνέδριο

Uncovering the Genetic Complexity of Rare Diseases through Whole-Exome Sequencing

Ioanna Papadionysiou¹, Valentini Tzimogianni¹, Eirini Veltsou¹, Zoi Lygerou¹

¹*Molecular Genetics Unit, General Biology Department, Medical School, University of Patras, Greece.*

Rare diseases (RDs) are conditions with low prevalence, affecting fewer than 1 in 2000 individuals. Globally, RDs impact nearly 400 million people, including an estimated 1 million in Greece. Their extensive phenotypic and genetic heterogeneity makes accurate diagnosis particularly challenging, leaving many patients without a definitive diagnosis. Since most RDs have a genetic origin, Whole-Exome Sequencing (WES) has become a powerful tool for identifying causative variants and achieving molecular diagnoses.

In this study, Whole-Exome Sequencing was conducted on a Greek cohort of more than 400 patients with rare diseases who remained undiagnosed at the time of analysis. The cohort included both pediatric and adult individuals, representing a wide spectrum of phenotypes, including neurodevelopmental, neurological, and neuromuscular disorders, as well as growth, metabolic, cardiovascular, and skeletal abnormalities.

Different data bases and in silico tools were used for the analysis of WES data, which identified the underlying genetic cause in 34% of cases (positive result) and additionally, in 18% of cases, a Variant of Uncertain Significance (VUS) was identified (inconclusive result). Furthermore, the pipeline analysis revealed six cases where more than one genetic locus was implicated in the disease etiology. Next step is to examine the presence of these genetic loci in population data and investigate potential correlation with the severity of the phenotype. Lastly, in negative cases with family history and a strong genetic indication Whole Genome Sequencing (WGS) was performed to identify rare and common variants that could be associated with the patient's phenotype.

Identifying the genetic cause of rare genetic diseases provides scientists and clinicians with valuable insights into the underlying mechanisms and the pathology of a disease, enabling improved prediction, prevention, and treatment strategies

Uncovering the Genetic Complexity of Rare Diseases through Whole-Exome Sequencing

Ioanna Papadionysiou¹, Valentini Tzimogianni¹, Eirini Vetslou¹, Zoi Lygerou¹

¹Molecular Genetics Unit, General Biology Department, Medical School, University of Patras, Greece

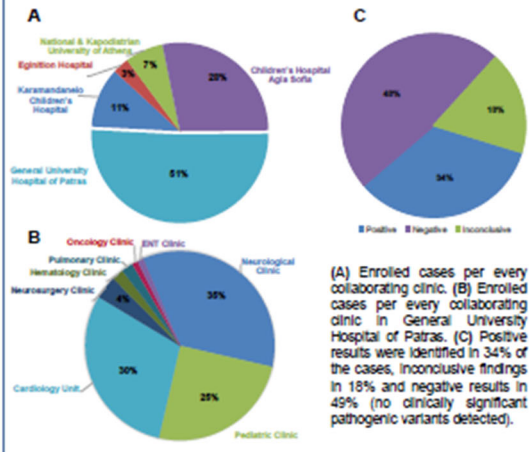


Introduction

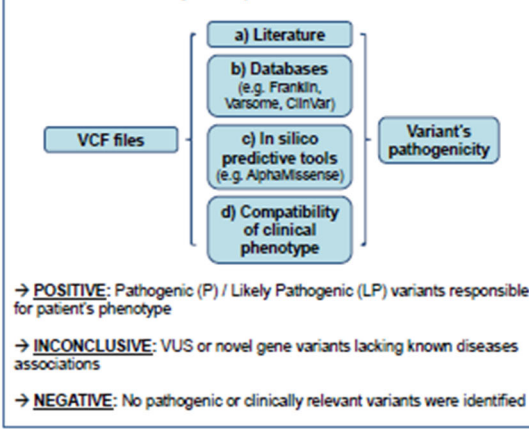
Rare diseases (RDs) impact millions of people around the globe, and most stem from underlying genetic causes. Although genetic research has advanced considerably, diagnosing these conditions remains difficult because of their complexity and infrequency. Whole Exome Sequencing (WES)—a genomic technique that targets the protein-coding regions of the genome—has become a crucial tool for uncovering the genetic roots of these disorders. By detecting pathogenic variants within coding sequences, WES has improved diagnostic yields and broadened our understanding of genetic diseases. Nevertheless, many individuals with rare conditions still experience diagnostic uncertainty, reflecting the limitations of current variant-interpretation methods and gaps in our knowledge of genetic and molecular mechanisms. Our analysis of the patients' WES data revealed cases in which multiple genetic loci contributed to disease etiology, while negative cases underwent further investigation through Whole Genome Sequencing (WGS), which also evaluates non-coding regions.

Results

1. Whole-Exome Sequencing (WES) in a Greek cohort of 465 patients



2. Analysis Pipeline for WES data



3. Whole-Genome Sequencing of 15 Negative WES Cases

- Advantages of Whole-Genome Sequencing:**
- Sequencing coding & non-coding regions
 - Identification of non-coding variants
 - Detection of Structural Variants (SVs)
 - Detection of Copy Number Variants (CNVs)
 - Improved variant calling in complex regions (repetitive sequences, GC-rich regions, homologous regions)
 - More uniform coverage and less sequencing bias
- BUT:** Cost & Time intensive, More challenging bioinformatic analysis

4. Multi-locus Contribution to the Disease Etiology

VarCoPP2: Identification of bilocus variant combinations with potential pathogenic significance

- 1 VUS in MYH6 & 1 VUS variant in MYH7**
Double heterozygous variant in MYH6 & MYH7 associated with the presence of Hypertrophic Cardiomyopathy (HCM) (Suzuki et al., 2021)
VarCoPP2 → Median Pathogenicity Score: 0.98
- 1 Pathogenic variant in LMNA & 1 VUS in LAMA4**
Both genes are associated with Dilated Cardiomyopathy (DCM).
VarCoPP2 → Median Pathogenicity Score: 0.89
- 1 Pathogenic variant & 1 VUS in SH3TC2**
Mutations in SH3TC2 is a frequent cause of the autosomal recessive demyelinating neuropathy (Adam et al., 2025, GeneReviews)
- 2 Pathogenic variants in DYSF**
Combined mutations in DYSF have been linked with dysferlinopathies.
- 1 Pathogenic variant & 1 VUS in GDAP1**
Severe cases of Charcot-Marie Tooth neuropathy have been linked with the presence of combined variants in GDAP1 gene. (Shumeri et al., 2025)

Summary

- In many cases, Whole-Exome sequencing provides a cost-effective solution in genetic diagnostics.
- VarCoPP2 can help us predict the pathogenicity of bilocus variant combinations.
- The next step involves comparing the prevalence of multilocus variants in a Greek population and examining their potential associations with disease severity and age of onset.

This project (TAEDR-0539180) is implemented within the framework of the Action "Actions in interdisciplinary scientific areas with special interests for the connection with the productive fabric", Greece 2.0 - National Recovery and Resilience Plan.