



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ**  
**ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**Βιοπληροφορική Ανάλυση Φαρμακογονιδιωματικών Δεικτών σε  
Έλληνες Ασθενείς**

**Γέρου Αλεξάνδρα**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**  
**Υπεύθυνοι**  
**Μπάγκος Παντελής, Πατρινός Γιώργος**  
**Καθηγητές**

**Λαμία, 2021**





**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ  
ΒΙΟΙΑΤΡΙΚΗ**

**Βιοπληροφορική Ανάλυση Φαρμακογονιδιωματικών Δεικτών σε  
Έλληνες Ασθενείς**

**Γέρου Αλεξάνδρα**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Επιβλέποντες  
Μπάγκος Παντελής, Πατρινός Γιώργος  
Καθηγητές**

**Λαμία, 2021**

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις <sup>(1)</sup>, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: ...../...../20.....

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

**Βιοπληροφορική Ανάλυση Φαρμακογονιδιωματικών Δεικτών σε  
Έλληνες Ασθενείς  
Γέρου Αλεξάνδρα**

**Τριμελής Επιτροπή:**

Μπάγκος Παντελής, Καθηγητής (επιβλέπων)

Μπράλιου Γεωργία, Επίκουρος Καθηγήτρια

Χατζηγεωργίου Άρτεμις, Καθηγήτρια

## Πίνακας Περιεχομένων:

Περίληψη.....	7
Abstract.....	8
1 Κεφάλαιο. Θεωρητικό μέρος.....	9
1.1 Εισαγωγή.....	10
1.2 Whole exome studies (Μελέτη του συνόλου των εξονίων).....	11
1.3 Whole genome studies (Μελέτη ολόκληρου του γονιδιώματος).....	12
1.4 Γονιδιωματικοί και Φαρμακογονιδιωματικοί Δείκτες.....	13
1.5 Βάσεις δεδομένων και εργαλεία για τον χαρακτηρισμό μεταλλάξεων.....	16
1.5.1 dbNSFP.....	17
1.5.2 BLAST.....	17
1.5.3 BLAT (Seach Genome).....	18
1.5.4 BioMart.....	18
1.5.5 ENCODE.....	18
1.5.6 GENCODE.....	18
1.5.7 Consensus Coding Sequence.....	18
1.5.8 ANNOVAR.....	19
1.5.9 VAT (Variant Annotation Tool).....	19
1.5.10 VAAST (Variant Annotation, Analysis and Search Tool).....	19
2 Κεφάλαιο. Υλικά και Μέθοδοι.....	20
2.1 Βάσεις Δεδομένων για Εξόρυξη Πληροφοριών με σκοπό τον Χαρακτηρισμό Γονιδιωματικών και Φαρμακογονιδιωματικών δεικτών στο δείγμα.....	21
2.1.1 PharmGKB.....	21
2.1.2 Ensembl.....	21
2.1.3 Variant Effect Predictor.....	21
2.2 Διαδικασία επεξεργασίας δεδομένων πενήντα έξι (56) φακέλων ασθενών.....	22
2.2.1 Πρώτο Στάδιο Επεξεργασίας.....	22
2.2.2 Δεύτερο Στάδιο Επεξεργασίας.....	25
2.2.3 Τρίτο Στάδιο Επεξεργασίας.....	25
3 Κεφάλαιο. Αποτελέσματα.....	26
4 Κεφάλαιο. Συμπεράσματα.....	31
5 Κεφάλαιο. Βιβλιογραφία.....	35
6 Κεφάλαιο. Παράρτημα.....	39

## Περίληψη

Η φαρμακογονιδιωματική αποτελεί κλάδο της γενετικής ιατρικής, ο οποίος στοχεύει στην εξατομίκευση της προτεινόμενης θεραπείας και μελετά την γενετικά καθοριζόμενη απόκριση των ασθενών σε αυτή, παρέχοντας τη δυνατότητα προσαρμογής στις μοναδικές ανάγκες τους. Οι διαφορετικές μεταλλάξεις που εντοπίζονται στο γονιδίωμα, καθορίζουν σημαντικά τις φαινοτυπικές διαφορές των ατόμων. Για τον σκοπό αυτό, έχουν αναπτυχθεί μέθοδοι για τον γρήγορο, σωστό και οικονομικό προσδιορισμό της αλληλουχίας του ανθρώπινου γονιδιώματος. Ένα τέτοιο εργαλείο αποτελεί ο προσδιορισμός αλληλουχίας επόμενης γενιάς (next generation sequencing), ο οποίος μακροπρόθεσμα μπορεί ακόμα να προσφέρει μία ολοκληρωμένη πλατφόρμα φαρμακογονιδιωματικής ταυτοποίησης. Σκοπός της παρούσας εργασίας είναι η μελέτη και παρουσίαση τέτοιων εργαλείων βιοπληροφορικής, προκειμένου να διασφαλιστούν αυτοματοποιημένα μοντέλα φιλικά τόσο για τον ίδιο τον ασθενή όσο και για τον επιστήμονα υγείας στο εγγύς μέλλον. Συγκεκριμένα επιλέχθηκε για την πρακτική εφαρμογή η VEP της ensemble, με στόχο των εντοπισμό ενός συνόλου διαφορετικών φαρμακογονιδιωματικών πολυμορφισμών, σε ένα πλήθος 56 (πενήντα έξι) Ελλήνων ασθενών και στην συνέχεια έγινε στατιστική αποτύπωση των αποτελεσμάτων που προέκυψαν. Με σκοπό την επιλογή πολυμορφισμών με χαρακτηριστική κλινική σημασία, τα αποτελέσματα που προέκυψαν από την αναζήτηση της VEP συγκρίθηκαν με αντίστοιχες καταχωρήσεις που υπάρχουν στην PharmGKB. Κατόπιν συζητήθηκαν περεταίρω μερικοί πολυμορφισμοί οι οποίοι φάνηκαν να υπάρχουν στο δείγμα σε ιδιαίτερα υψηλά ποσοστά.

## **Abstract.**

Pharmacogenomics is a branch of genetic medicine that aims to personalize the proposed treatment and studies the genetically determined response of patients to it, providing the ability to adapt to their unique needs. The different mutations found in the genome significantly determine the phenotypic differences of individuals. For this purpose, methods have been developed for the rapid, accurate and economical sequencing of the human genome. One out of all these methods is next generation sequencing, which in the long run can still provide a complete pharmacogenomic identification platform. Functional interpretation of the results obtained from the next generation sequences, however, is an obstacle to clinical application. Bioinformatics comes to face this problem, as it offers the possibility of processing and interpreting a huge volume of data. It implements advanced computational tools for the analysis of biologically important data, in order to convert the signals, which it receives into data and then the data into comprehensible and useful information that can be practically applied in the clinical study. The purpose of this paper is to study and present such bioinformatics tools, in order to ensure automated models that are friendly to both the patient and the health scientist in the near future. Specifically, the VEP of ensembl was selected for practical application, with the aim of identifying a set of different pharmacogenomic polymorphisms in 56 (fifty-six) Greek patients and then a statistical record of the results was made. To select polymorphisms of characteristic clinical significance, the results obtained from the VEP search were compared with corresponding entries in PharmGKB. Some polymorphisms were then discussed which appeared to be present in the sample in very high percentages.



## **1 Κεφάλαιο. Θεωρητικό μέρος**

## 1.1 Εισαγωγή

Η απόκτηση μιας πλήρους εικόνας του συνόλου των μεταλλάξεων ενός ατόμου, είναι απαραίτητη ιδιαίτερα στην περίπτωση εμφάνισης νέων μοναδικών μεταλλάξεων, οι οποίες μπορούν είτε να καταστήσουν ανενεργό κάποιο ένζυμο ή/και τον μεταφορέα που μεταβολίζει κάποιο φάρμακο που χορηγείται ή πρόκειται να χορηγηθεί στο άτομο, είτε να απενεργοποιήσουν την έκφρασή του και συνεπώς τον φορέα του παραλλακτικού ενδιάμεσου, ή ακόμα να οδηγήσουν σε μερικό του μεταβολισμό.

Η φαρμακογονιδιοματική είναι ένας κλάδος που συνδυάζει τη μοριακή φαρμακολογία και την μοριακή βιολογία με βάση την φαρμακογενετική, για την μελέτη της διαφορικής απόκρισης των πληθυσμών στα φάρμακα [27].

Παρόλο που πολλές φορές η Φαρμακογονιδιοματική και η Φαρμακογενετική συγχέονται, δεν αποτελούν στην πραγματικότητα ταυτόσημους όρους. Η Φαρμακογενετική χρησιμοποιείται προκειμένου να περιγράψει τη μελέτη συγκεκριμένων φαινοτύπων, όσον αφορά την απάντηση του οργανισμού στα φάρμακα. Ενασχόληση της είναι η εξατομίκευση της ιατρικής στα πλαίσια της συνεχούς αναζήτησης για ασφαλέστερη, περισσότερο αποτελεσματική και ατομική φαρμακοθεραπεία στους ασθενείς. Αυτό πραγματοποιείται με την παροχή ιατρικών αποτελεσματικών επιστημονικών εργαλείων στους κλινικούς, που θα τους καθοδηγούν να επιλέξουν το σωστό φάρμακο και στην επαρκή δόση για κάθε ασθενή. Είναι ένα πεδίο, στο οποίο η κλινική φαρμακολογία εισάγει τη γενετική σε μια προσπάθεια παροχής καλύτερης φαρμακευτικής αγωγής στους ασθενείς. Αντίθετα, η Φαρμακογονιδιοματική (Pharmacogenomics) αναφέρεται σε ολόκληρο το φάσμα των γονιδίων που καθορίζουν την απάντηση του οργανισμού στα φάρμακα, συμπεριλαμβανομένης της εκτίμησης της ποικιλομορφίας της αλληλουχίας του ανθρώπινου γονιδιώματος και των κλινικών συνεπειών της ποικιλομορφίας αυτής. Η Φαρμακογενετική εστιάζει στο φαινότυπο ενός κληρονομούμενου γενετικού στοιχείου, ενώ η Φαρμακογονιδιοματική έχει ευρύτερο στόχο και καλύπτει τη μελέτη της ρύθμισης και της έκφρασης των γονιδίων που εμπλέκονται σε αυτόν το φαινότυπο [28].

Η Φαρμακογονιδιοματική στοχεύει στην ανάπτυξη μεθόδων για την ατομική θεραπεία, με σκοπό να αυξήσει την αποδοτικότητα των φαρμάκων και παράλληλα να ελαχιστοποιήσει την τοξικότητα τους, σύμφωνα με το πως οι γονιδιακές μεταλλάξεις επηρεάζουν την δράση αυτών [10]. Μια ολοκληρωμένη προσέγγιση στη φαρμακογονιδιοματική έρευνα θα πρέπει να περιλαμβάνει την εξέταση της επίδρασης των γονιδίων που εμπλέκονται στην απορρόφηση φαρμάκων, τη μεταφορά, το μεταβολισμό, την απέκκριση και ακόμα την εξέταση των πρωτεϊνών που σχετίζονται με τον μηχανισμό δράσης [15].

Οι φαρμακογονιδιοματικές έρευνες έχουν παρουσιάσει πρωτοφανή άνοδο κατά την πάροδο των χρόνων, γεγονός που οδηγεί στην διάθιση της ανακάλυψης και έρευνας ποικίλων γενετικών μεταλλάξεων. Παρά τα εμπόδια που συναντάει, η φαρμακογονιδιοματική φαίνεται να είναι ένας τομέας της γενετικής ιατρικής όπου η ακρίβεια της εφαρμογής ενός φαρμάκου θα μπορούσε να έχει άμεσο αντίκτυπο στο εγγύς μέλλον [23].

## 1.2 Whole exome studies (Μελέτη του συνόλου των εξωνίων)

Οι προσεγγίσεις των ακολουθιών επόμενης γενιάς (next generation sequences) αρχίζουν σταδιακά να υιοθετούνται από την φαρμακογονιδιοματική και περιλαμβάνουν αλληλούχιση:

- Ολόκληρου του γονιδιώματος
- Ολόκληρου του πλήθους των εξωνίων του οργανισμού

Για τον εντοπισμό μεταλλάξεων στο DNA, η ανάλυση του συνόλου των εξωνίων (Whole exome sequencing – WES) ενός ατόμου είναι μία πολύ σημαντική τεχνική. Ωστόσο αποτελεί ενδιάμεσο στάδιο της μελέτης, καθώς ακόμα και αν εφαρμοστεί πλήρης τεχνική ακρίβεια, το κομμάτι αυτό (των εξωνίων) αποτελεί μόλις το 1% του συνολικού γονιδιώματος του ανθρώπου [16].

Η ανάλυση του συνόλου των εξωνίων αρχίζει να χρησιμοποιείται, αν και σταδιακά, ολοένα και περισσότερο ως διαγνωστικό τεστ για τον εντοπισμό ετερογενών διαταραχών. Παρόλο που χρησιμοποιείται κυρίως για την ανίχνευση σημειακών μεταλλάξεων, καθώς και για την ανίχνευση μικρών διαφοροποιήσεων στο γονιδίωμα (π.χ. προσθήκες, διαγραφές κλπ), το WES μπορεί επίσης να χρησιμοποιηθεί και για τον χαρακτηρισμό CNVs (copy number variations), επιτρέποντας την ανάλυση μικρών και μεγάλων μεταλλάξεων [11].

Η αλληλούχιση του συνόλου των εξωνίων εντοπίζεται και στον τομέα της προγεννητικής διάγνωσης, η οποία παρέχει πληροφορίες για την εγκυμοσύνη και τη λήψη και διαχείριση των περιγεννητικών αποφάσεων. Σε αρκετές περιπτώσεις, οι προσεγγίσεις προγεννητικής αλληλουχίας ολόκληρου του πλήθους των εξωνίων, (WES) έχουν εντοπίσει γενετικές διαγνώσεις όταν οι συμβατικές εξετάσεις (καρυότυπος και μικροσυστοιχία) δεν επέφεραν την απαραίτητη διάγνωση. Πέρα από την ικανότητά του να βελτιώνει τα διαγνωστικά ποσοστά, το WES αποτελεί πιθανό εργαλείο ώστε να βελτιωθεί η κατανόηση των προγεννητικών εμφανίσεων γενετικών διαταραχών και θανατηφόρων εμβρυϊκών συνδρόμων στο εγγύς μέλλον της προγεννητικής γενετικής διάγνωσης [4]. Η προγεννητική αλληλούχιση όλων των εξωνίων, προσφέρει ακόμα την δυνατότητα αύξησης της παροχής μεγαλύτερου πλήθους διαγνωστικών ικανοτήτων σε έμβρυα με υπερηχογραφικές ανωμαλίες, οι οποίες μπορούν να συμβάλουν στην ικανότητα παροχής συμβουλών σε οικογένειες. Είναι επίσης συχνά το πρώτο βήμα για τη βελτίωση της πορείας προς την ενημερωμένη διάγνωση και θεραπεία, η οποία είναι ιδιαίτερα σημαντική στην εποχή της προόδου για την εμβρυϊκή θεραπεία [13].

Παρά το γεγονός ότι η ανάλυση του πλήθους των εξωνίων (whole exome sequencing) αποτελεί μία πολύ αποδοτική μέθοδο όσον αφορά τον εντοπισμό και την μελέτη των διάφορων μεταλλάξεων στα άτομα, συχνά συναντώνται αρκετοί περιορισμοί, ειδικά όσον αφορά την φαρμακογονιδιοματική. Ο λόγος είναι ότι υπάρχουν ακόμα τμήματα στην κωδικοποίηση των πρωτεϊνών που παραμένουν άγνωστα, με αποτέλεσμα με την μέθοδο αυτή να μπορούν να εντοπιστούν μόνο εξώνια τα οποία είναι ήδη γνωστά, ενώ ρυθμιστικές περιοχές ή περιοχές γονιδίων οι οποίες δεν έχουν μεταφραστεί ακόμα δεν αλληλουχίζονται.

Επιπρόσθετα υπάρχει το ενδεχόμενο σημαντικής απόκλισης κατά το στάδιο εμπλουτισμού του στόχου, καθώς η αποτελεσματικότητα των ανιχνευτών δέσμευσης ποικίλει σημαντικά, με αποτέλεσμα πολλές φορές να παραλείπονται ορισμένες ακολουθίες/ στόχοι. Το παραπάνω μπορεί να είναι παράγωγο μίας ενδεχομένως κακής κάλυψης σε ορισμένες περιοχές κωδικοποίησης, ή ακόμα μηδενικής κάλυψης σε πιθανά σημαντικές και κλινικά σχετικές γονιδιοματικές παραλλαγές που βρίσκονται

σε γονιδιωματικές περιοχές εκτός των εξωνίων. Τέλος, για τη συγκεκριμένη μελέτη, υπάρχουν πολλά σε πλήθος και σε διαφορές (αναφορικά με την κάλυψη, την ομοιομορφία στο αποτέλεσμα, την αξιοπιστία κοκ) «kit» εμπλουτισμού εμπορικού στόχου (commercial target enrichment kits), ένα κρίσιμο στοιχείο της στοχευμένης αλληλουχίας επόμενης γενιάς για την οικονομικά αποδοτική και ευαίσθητη ανίχνευση μεταλλάξεων, η οποία εκτελείται κυρίως από συλλογή δείγματος που βασίζεται στην υβριδική επιλογή ή με τεχνικές PCR, που αποτελούν διαδεδομένες τεχνικές. Ανάμεσα στις δύο τεχνικές πιο οικονομική χαρακτηρίζεται η μέθοδος PCR, διότι η χρήση βιοτινυλιωμένων ολιγονουκλεοτιδίων (biotinylated oligonucleotides) συμπληρωματικών ως προς περιοχές στόχους και μαγνητικών σφαιριδίων επικαλυμμένων με στρεπταβιδίνη για τον εντοπισμό των επιθυμητών περιοχών η οποία απαιτείται στον εντοπισμό μέσω υβριδικής επιλογής, είναι ακριβότερη, περιλαμβάνει περισσότερα στάδια και απαιτεί μεγαλύτερο δείγμα DNA [5].

Η ποικιλότητα αυτή στις τεχνικές και τα εργαλεία, έχει ως αποτέλεσμα τα δείγματα να μην αλληλουχίζονται όλα με την ίδια αποδοτικότητα. Κατά συνέπεια ένα σημαντικό ποσοστό των μεταλλάξεων μπορεί να μην ανιχνευθεί καθόλου. Το γεγονός αυτό μπορεί να είναι ιδιαίτερα προβληματικό για τα παράλογα γονίδια, δηλαδή ομόλογα γονίδια τα οποία προήλθαν από την δημιουργία ενός δεύτερου αντιγράφου της αρχικής αλληλουχίας, όπου η αναζήτηση τους θα πρέπει να πραγματοποιηθεί μέσα από συγκεκριμένες πλατφόρμες αλληλούχησης επόμενης γενιάς (next generation sequencing platforms). Χαρακτηριστικό παράδειγμα η γονιδιακή επανεξέταση του *CYP2D6*. [16].

### 1.3 Whole genome studies (Μελέτη ολόκληρου του γονιδιώματος)

Η ανάλυση ολόκληρου του γονιδιώματος (whole genome sequencing) αποτελεί ως επί το πλείστον μία πολλά υποσχόμενη τεχνική, που βρίσκει εφαρμογή σε διάφορα είδη κλινικών μελετών και χάρη στην οποία παρέχονται άκρως σχετικές πληροφορίες για την κλινική μικροβιολογία σε σχεδόν πραγματικό χρόνο, από τη δοκιμή φαινοτύπων έως την παρακολούθηση εξάρσεων στον οργανισμό [2].

Στην φαρμακογονιδιωματική υπάρχουν ελάχιστες μελέτες αλληλούχησης ολόκληρου του γονιδιώματος, παρά το γεγονός ότι οι αλληλουχίσεις αυτές ωφελούν σημαντικά τις μελέτες σχετικά με τις γενετικές ανωμαλίες. Με την μέθοδο αυτή αναλύεται ολόκληρο το γονιδίωμα των ατόμων, με στόχο την δημιουργία του φαρμακογονιδιωματικού προφίλ του καθενός. Σε αντίθεση με μεθόδους που προσεγγίζουν συμβατικά ή ακόμα και με υψηλή απόδοση την διαλογή γενετικών πληροφοριών, η ανάλυση ολόκληρου του γονιδιώματος επιτρέπει την δημιουργία μίας πλήρους εικόνας σχετικά με τα ποικίλα γενετικά χαρακτηριστικά των ατόμων [16].

Η αλληλούχιση ολόκληρου του γονιδιώματος χρησιμοποιήθηκε για την ταυτοποίηση νέων και πιθανώς αιτιολογικά υπεύθυνων γονιδιωματικών μεταλλάξεων, οι οποίες επηρεάζουν την δομή και την λειτουργία διακοσίων τριάντα ένα (231) φαρμακογονιδίων σε ένα μεγάλο δείγμα ατόμων, με διαφορετικά εθνικά υπόβαθρα [10]. Επιπλέον με την μέθοδο αυτή παρέχονται πληροφορίες θέσης και προσανατολισμού των μεταλλάξεων [11]. Η ανάλυση του WGS σε χρόνια ασθενή παιδιά, παραδείγματος χάριν, αποτελεί ένα διαγνωστικό εργαλείο πολύ σημαντικό για τον εντοπισμό υποκείμενων νοσημάτων που μπορεί να μην είχαν εντοπιστεί μέχρι τώρα. Σε μελέτες μάλιστα που πραγματοποιήθηκαν τόσο στο πυρηνικό όσο και στο μιτοχονδριακό γονιδίωμα τέτοιων ασθενών, βρέθηκε ότι σαράντα (40) περιπτώσεις εξ αυτών (ανάμεσα στις οποίες συμπεριλαμβάνονταν παθολόγες ή πιθανά παθολόγες

παραλλαγές σε ένα ευρύ φάσμα σπάνιων διαταραχών όπως εγκεφαλοπάθειες, μυοπάθειες, σκελετική δυσπλασία και διάφορα σύνδρομα) έλαβαν διάγνωση μέσω ανάλυσης WGS (διαγνωστικός ρυθμός 21%), η οποία είτε πλήρως (95%) είτε μερικώς (5%) εξήγησε τον φαινότυπο τους [9].

Αυτό είναι σημαντικό, διότι είναι πιθανό κάθε άτομο να εμφανίζει σπάνιες ή και νέες παραλλαγές σε γνωστά φαρμακογονίδια, γεγονός που συνδέεται άμεσα με το πως ανταποκρίνεται σε ορισμένα φάρμακα (π.χ. υπερμεταβολιστής, μη ανταποκρινόμενο κ.ο.κ.), κάτι που ενδέχεται να μην εντοπιστεί καθόλου με την χρήση άλλων γενετικών μεθόδων. Ωστόσο, από μία σειρά ερευνών, διαπιστώνεται ότι οποιοδήποτε αποτέλεσμα προκύπτει από μία τρέχουσα διαθέσιμη έκθεση βασισμένη σε φαρμακογονιδιωμάτη διαδικασία διαλογής ενός ασθενούς να ανταποκρίνεται σε ορισμένα φάρμακα, δεν είναι απαραίτητα ενδεικτική, συνεπώς πρέπει να ερμηνεύεται με σύνεση [10].

Πρόσφατες αποδείξεις, αν και περιορισμένες σε έκταση, υποστηρίζουν την θέση ότι η αλληλούχιση ολόκληρου του γονιδιώματος είναι ικανή να αποκαλύψει σπάνιους ή και μοναδικούς ακόμα φαρμακογονιδιωματικούς δείκτες, που υπό άλλες συνθήκες μπορεί να μην εντοπίζονταν αν εφαρμόζονταν τεχνικές συνηθισμένου γενετικού ελέγχου (π.χ. μέθοδοι PCR ή μικροσυστοιχίες). Οι γενετικές διαταραχές έχουν ήδη αρχίσει να επωφελούνται σε μεγάλο βαθμό από εφαρμογές αλληλουχίας ολόκληρου του γονιδιώματος και γονιδιωματικά δεδομένα από επιλεγμένους ασθενείς, με σαφώς καθορισμένους φαινοτύπους να συγκεντρώνονται με ταχύ ρυθμό. Ακόμα, μελέτες που κάνουν χρήση της τεχνολογίας της αλληλουχίας ολόκληρου του γονιδιώματος οδήγησαν τα τελευταία χρόνια σε εκθετική αύξηση της παραγωγής δεδομένων μεταλλάξεων του ανθρώπινου γονιδιώματος από ερευνητικά κέντρα και διαγνωστικά εργαστήρια, εμπλουτίζοντας έτσι ακόμα περισσότερο τις γνώσεις της επιστημονικής κοινότητας τόσο για τη γενετική ετερογένεια που διέπει τις μονογονικές διαταραχές, όσο και για τα πολύπλοκα κλινικά χαρακτηριστικά [20].

Η διαδικασία αλληλούχισης ολόκληρου του γονιδιώματος φαίνεται ωστόσο πως αντιμετωπίζει προκλήσεις στο να γίνει κλινικά αποδεκτή. Τις πιο σημαντικές από αυτές αποτελούν το κόστος της διαδικασίας, η δημιουργία ρυθμίσεων σχετικά με την χρήση της μεθόδου ως ιατρική συσκευή, η δημιουργία κατάλληλων εγκαταστάσεων και τέλος η εκπαίδευση προσωπικού ώστε να μπορεί να φέρει τις κλινικές μελέτες εις πέρας.

#### 1.4 Γονιδιωματικοί και Φαρμακογονιδιωματικοί Δείκτες.

Οι τομείς της παθολογίας και της κλινικής φαρμακολογίας βρίσκονται σε μια εποχή ταχείας εξέλιξης, αντικατοπτρίζοντας τις συνεχείς τεχνολογικές εξελίξεις στη μοριακή διάγνωση, που εστιάζουν στη βελτίωση της αποτελεσματικότητας των φαρμάκων (όπως η έγκριση του trastuzumab (Herceptin) για τη θεραπεία του HER-2 /neu που υπερεκφράζει τον καρκίνο του μαστού, του imatinib (Gleevec) για τη θεραπεία της χρόνιας μυελογενούς λευχαιμίας και των bcr/abl μεταστάσεων που επιφέρει, όπως και των γαστρεντερικών στρωματικών όγκων με επιλεκτικές μεταλλάξεις ενεργοποίησης ογκογόνου c-kit) και στην μείωση της τοξικότητας τους στους ασθενείς. Αυτό επιτυγχάνεται με προσαρμογή της δοσολογίας και του τρόπου χορήγησης υπαρχόντων και νέων θεραπευτικών παραγόντων, ενώ παράλληλα διευκολύνει και την περαιτέρω ανάπτυξη εξατομικευμένης ιατρικής περίθαλψης. Η ανακάλυψη και η βελτίωση της αλληλουχίας του ανθρώπινου γονιδιώματος, η περαιτέρω κατανόηση των επιγενετικών γεγονότων (π.χ. στον κλάδο της καρδιολογίας σε περιπτώσεις που παρατηρούνται φλεγμονώδεις καταστάσεις,

νευροεκφυλιστικές ασθένειες και ψυχιατρικές διαταραχές) και η επέκταση της πρωτεϊνικής έρευνας σε συνδυασμό με αναδυόμενες τεχνολογίες (όπως λειτουργική απεικόνιση, biomarkers και εξελιγμένη υπολογιστική βιολογία) φαίνεται να έχουν πρωτοφανές αντίκτυπο στη φαρμακευτική βιομηχανία [29].

Ανάλογα με το αντικείμενο του ενδιαφέροντος διακρίνονται οι φαρμακογενετικές και οι φαρμακογονιδιωματικές μελέτες. Η διαφορά είναι πως οι φαρμακογενετικές μελέτες αξιολογούν την επίδραση των μεταβολών της αλληλουχίας DNA στην απόκριση του φαρμάκου, ενώ οι φαρμακογονιδιωματικές μελέτες επικεντρώνονται στον εντοπισμό και την ανάλυση των παραλλαγών τόσο του DNA όσο και του RNA που επηρεάζουν την αποτελεσματικότητα και τοξικότητα της φαρμακευτικής θεραπείας.

Ο προσδιορισμός και ο έλεγχος των φαρμακογονιδιωματικών βιοδεικτών (PGx biomarkers), συνήθως περιλαμβάνει προσεγγίσεις ανίχνευσης DNA και RNA για τις δύο θέσεις ενός γονιδίου, μία στοχευμένη ομάδα γονιδίων (targeted gene panel) ή την σάρωση ολόκληρου του γονιδιώματος, ανάλογα με την έρευνα ή την κλινική εφαρμογή [17]. Οι βιοδείκτες (biomarkers) αποτελούν στην πραγματικότητα μια βιολογική παρατήρηση που αντικαθιστά και προβλέπει ιδανικά ένα κλινικά σχετιζόμενο endpoint ή ένα ενδιάμεσο αποτέλεσμα που είναι δύσκολο να παρατηρηθεί (intermediate outcome). Ένας επικυρωμένος βιοδείκτης έχει αποδοθεί σε ένα αναλυτικό σύστημα δοκιμών με καλά εδραιωμένα χαρακτηριστικά απόδοσης και για τα οποία υπάρχει ένα καθιερωμένο επιστημονικό πλαίσιο ή ένα σύνολο αποδεικτικών στοιχείων που διευκρινίζει την φυσιολογική, τοξικολογική, φαρμακολογική ή κλινική σημασία των αποτελεσμάτων των δοκιμών. Η χρήση κλινικών βιοδεικτών είναι ευκολότερη και λιγότερο δαπανηρή συγκριτικά με την άμεση μέτρηση του κλινικού endpoint (καταληκτικού σημείου). Μετριούνται συνήθως σε μικρότερο χρονικό διάστημα και μπορούν να χρησιμοποιηθούν στον έλεγχο, την διάγνωση, τον χαρακτηρισμό και την παρακολούθηση ασθενειών καθώς και ως προγνωστικοί δείκτες. Ακόμα βρίσκουν εφαρμογή στην ανάπτυξη εξατομικευμένων θεραπευτικών παρεμβάσεων, την πρόβλεψη και την αντιμετώπιση ανεπιθύμητων ενεργειών που μπορεί να προκληθούν από τα φάρμακα και τον προσδιορισμό των διαφορών τύπων κυττάρων, όπως ακόμα και για μελέτες φαρμακοδυναμικής και μελέτες απόκρισης δόσης. Σημαντική προϋπόθεση για την κατανόηση της αξίας ενός βιοδείκτη, είναι η γνώση της παθοφυσιολογικής σχέσης μεταξύ του βιοδείκτη και του σχετικού κλινικού endpoint. Με ποιόν τρόπο δηλαδή σχετίζονται οι διαταραχές στην φυσιολογία του βιοδείκτη με την εκδήλωση κάποιου κλινικού endpoint ή νόσου. Οι χρήσιμοι βιοδείκτες θα πρέπει να είναι μετρήσιμοι με μικρή ή καμία μεταβλητότητα, να έχουν υψηλή αναπαραγωγικότητα και να μεταβάλλονται άμεσα και αξιόπιστα ως απάντηση σε αλλαγές σε μια κατάσταση ή την θεραπεία αυτής [1].

Το 1998, στις ΗΠΑ, η Ομάδα Εργασίας του Εθνικού Ινστιτούτου για τις Ορολογίες των Βιοδεικτών, (National Institutes of Health Biomarkers Definitions Working Group) όρισε τον βιοδείκτη ως «Χαρακτηριστικό που μετράται αντικειμενικά και αξιολογείται ως ένας δείκτης φυσιολογικών βιολογικών διεργασιών, παθογόνων διεργασιών ή φαρμακολογικών αποκρίσεων σε θεραπευτική επέμβαση». Στην έκθεσή τους για την εγκυρότητα των βιοδεικτών στην εκτίμηση του περιβαλλοντικού κινδύνου, ο Παγκόσμιος Οργανισμός Υγείας ανακοίνωσε ότι ένας πραγματικός ορισμός των βιοδεικτών περιλαμβάνει «σχεδόν οποιαδήποτε μέτρηση αντικατοπτρίζει μια αλληλεπίδραση μεταξύ ενός βιολογικού συστήματος και ενός δυνητικού κινδύνου, ο οποίος μπορεί να είναι χημικός, φυσικός, ή βιολογικός. Η

μετρούμενη απόκριση μπορεί να είναι λειτουργική, φυσιολογική, βιοχημική στο κυτταρικό επίπεδο ή και μια μοριακή αλληλεπίδραση» [21].

Στην φαρμακογονιδιωματική η χρήση γενετικών βιοδεικτών για την εξατομίκευση και βελτιστοποίηση των απαιτήσεων δοσολογίας ορισμένων φαρμάκων στους ασθενείς, μέχρι στιγμής φαίνεται απογοητευτική. Για παράδειγμα, στο πολύπλοκο μοντέλο των επιδράσεων της βαρφαρίνης (warfarin) στην πήξη που περιλάμβανε δύο γενετικούς βιοδείκτες, παρουσιάστηκε απλώς μία μικρή βελτίωση στο θεραπευτικό αντιπηκτικό. Ωστόσο οι βιοδείκτες βρίσκουν άλλες εφαρμογές, όπως σε διάφορα στάδια για την ανακάλυψη και την ανάπτυξη φαρμάκων όπου χρησιμοποιούνται ως στόχοι (targets) ή για τον έλεγχο των ενώσεων κατά τη διάρκεια της ανακάλυψης φαρμάκων (π.χ. μέτρηση της δράσης της κυκλοοξυγενάσης για τον εντοπισμό πιθανών αντιφλεγμονωδών παραγόντων). Ακόμα εμφανίζονται ως τελικά σημεία (endpoints) για φαρμακοδυναμικές μελέτες (π.χ. στον ορό της χοληστερόλης ως δείκτης για τη δράση ενός φαρμάκου που προορίζεται για την πρόληψη των καρδιαγγειακών παθήσεων στις φαρμακοκινητικές/ φαρμακοδυναμικές μελέτες). Οι βιοδείκτες συναντώνται επίσης στην μελέτη της σχέσης μεταξύ της συγκέντρωσης ή της δόσης ενός φαρμάκου και του αποτελέσματός του, στην μέτρηση της αποτελεσματικότητας σε κλινικές δοκιμές αλλά και στην συνεισφορά στον καθορισμό των δυσμενών επιπτώσεων των υποψήφιων φαρμάκων [1].

Υπάρχουν πολλά γονίδια, γνωστά και ως «φαρμακογονίδια», τα οποία σχετίζονται με διάφορες λειτουργίες των φαρμάκων όπως παραδείγματος χάριν την απορρόφηση, διανομή και τον μεταβολισμό από τον οργανισμό (κ.ο.κ.). Τα γονίδια που σχετίζονται με τις παραπάνω λειτουργίες, χαρακτηρίζονται εν συντομία ως ADMET γονίδια (absorption, distribution, metabolism, excretion, toxicity). Τα πιο σημαντικά από αυτά τα ένζυμα/ πρωτεΐνες που κωδικοποιούνται από τα συγκεκριμένα γονίδια, χωρίζονται σε τέσσερις κύριες κατηγορίες:

- Τροποποιητές (modifiers).
- Ένζυμα μεταβολισμού φάσης 1 (phase I metabolism).
- Ένζυμα μεταβολισμού φάσης 2 (phase II metabolism).
- Μεταφορείς.

Η κλινική απόκριση σε ένα συγκεκριμένο φάρμακο προκαλείται από την αλληλεπίδραση μεταφορέων, ενζύμων του μεταβολισμού και των οδών αποτοξίνωσης (detoxification pathways). Ως εκ τούτου, οι πολυμορφισμοί που υπάρχουν σε ένα ή περισσότερα από τα εμπλεκόμενα γονίδια, θα μπορούσαν ενδεχομένως να επηρεάσουν την απόκριση ενός ατόμου στη θεραπεία [24]. Σκοπός της φαρμακογονιδιωματικής είναι ο εντοπισμός και η αξιοποίηση αυτών των διαφορετικών γονιδιωματικών πολυμορφισμών, ως εργαλείο στην εξατομικευμένη θεραπεία. Με γνώμονα τα παραπάνω και παρόλο που η φαρμακογονιδιωματική ανάλυση αλληλούχισης ολόκληρου του γονιδιώματος βρίσκεται σε πρώιμα στάδια, είναι πλέον αποδεκτή η θέση ότι η καθοριστική φαρμακογονιδιωματική δοκιμασία θα συμπεριλαμβάνει την επαναλληλούχιση των φαρμακογονιδίων που σχετίζονται με ADMET μεταλλάξεις, ιδιαίτερα εκείνων που έχουν αναγνωριστεί ως αξιόπιστοι φαρμακογονιδιωματικοί βιοδείκτες από ρυθμιστικούς οργανισμούς [10].

Οι βιοδείκτες (biomarkers) που διακρίνονται σε «καλούς βιοδείκτες» πρέπει να είναι μετρήσιμοι με μικρή ή καθόλου μεταβλητότητα, να έχουν μια σημαντική αναλογία σήματος προς θόρυβο και να μεταβάλλονται άμεσα και αξιόπιστα σύμφωνα με τις αλλαγές που παρατηρούνται κατά την διάρκεια της εκάστοτε θεραπείας [1]. Για το λόγο αυτόν έχουν δημιουργηθεί μέθοδοι (π.χ. μέθοδος PCR, μέθοδοι βασισμένες

σε μικροσυστοιχίες και άλλες) για τον εντοπισμό γνωστών φαρμακογονιδιωματικών δεικτών σε καλώς προσδιορισμένα φαρμακογονίδια [20].

Από μελέτες που έχουν πραγματοποιηθεί σχετικά με τους φαρμακογονιδιωματικούς δείκτες που αφορούν τα πιο διαδεδομένα φάρμακα, έχουν εντοπιστεί αρκετές μεταλλάξεις που οφείλονται στην αλληλεπίδραση φαρμάκου – γονιδίου. Παραδείγματος χάριν σε μελέτες σχετικά με τις αντικαρκινικές θεραπείες που αφορούν τις Ηνωμένες Πολιτείες, παρατηρήθηκε περιορισμένη αποτελεσματικότητα περίπου στο 70% των συμμετεχόντων ασθενών και ανεπιθύμητες αντιδράσεις στον οργανισμό από τα φάρμακα, είτε λόγω αδυναμίας πρόβλεψης της τοξικότητας του φαρμάκου στους διαφορετικούς οργανισμούς, είτε εξαιτίας των αντιδράσεων των φαρμάκων μεταξύ τους (drug to drug interactions) [29]. Ωστόσο οι μελέτες αυτές δεν συμπεριλαμβάνουν τους υπόλοιπους ή νέους τέτοιους δείκτες. Η ουσιαστική πρόκληση βρίσκεται στο να πραγματοποιηθεί όχι μόνο ο εντοπισμός της συσχέτισης αυτών των φαρμακογονιδίων με την μετάλλαξη, αλλά η αλληλούχιση ολόκληρων των φαρμακογονιδίων. Με την προσέγγιση αυτή ενδέχεται να εντοπιστούν νέες και πιθανώς επιβλαβείς μεταλλάξεις στα ήδη αναγνωρισμένα φαρμακογονίδια.

Ο συσχετισμός παραλλαγών των γονιδίων με συγκεκριμένες αντιδράσεις φαρμάκων σε μεμονωμένους ασθενείς είναι ύψιστης σημασίας. Με τον τρόπο αυτό συγκεντρώνονται πληροφορίες για την λήψη κλινικών αποφάσεων, προσαρμόζοντας την δοσολογία ή και αλλάζοντας εντελώς την φαρμακευτική αγωγή του ασθενούς [10]. Παρά το γεγονός αυτό, έχει παρατηρηθεί ότι σε πολλές Ευρωπαϊκές χώρες, ιδιαίτερα σε εκείνες με χαμηλό εισόδημα, οι πληροφορίες σχετικά με τους φαρμακογονιδιωματικούς δείκτες είναι ελλιπείς. Οι μελέτες που πραγματοποιούνται εστιάζουν κυρίως στις γενετικές μεταλλάξεις που εμφανίζονται στο γονιδίωμα, δίνοντας ελάχιστη έμφαση στην παγκόσμια επικράτηση και συνάφεια με τους φαρμακογονιδιωματικούς βιοδείκτες. Έτσι υπάρχουν πολύ λίγες πληροφορίες σχετικά με την συχνότητα των αλληλομόρφων των PGx βιοδεικτών που εμφανίζεται σε διαφορετικούς πληθυσμούς, έναντι των διαφορετικών φυλετικών ομάδων [20].

## 1.5 Βάσεις δεδομένων και εργαλεία για τον χαρακτηρισμό μεταλλάξεων.

Ο χαρακτηρισμός μεταλλάξεων είναι ένα κρίσιμο βήμα στην ανάλυση δεδομένων της γονιδιακής αλληλούχισης και πρακτικά αποτελεί διαδικασία εκχώρησης λειτουργικών πληροφοριών σε παραλλαγές DNA. Τα λειτουργικά συμπεράσματα του χαρακτηρισμού μεταλλάξεων μπορεί να έχουν καταλυτική επίδραση στα τελικά συμπεράσματα των μελετών διάφορων ασθενειών. Για τους πάσχοντες από κάποια ασθένεια, σπάνια ή διαδεδομένη, τα πιθανά οφέλη της ανάλυσης αυτής συμπεριλαμβάνουν την βελτίωση της φροντίδας των ασθενών, την παρακολούθηση και την αποτελεσματική τους θεραπεία. Χαρακτηριστικά αναφέρονται περιπτώσεις καρκίνου, όπου παρατηρούνται θετικά αποτελέσματα χρησιμοποιώντας δεδομένα από γενετικές εξετάσεις. Για παράδειγμα, οι ασθενείς που είναι θετικοί ως προς την κληρονομικότητα των μεταλλάξεων BRCA έχουν την επιλογή της επιλεκτικής προληπτικής χειρουργικής, ενώ σε ασθενείς με καρκίνο του πνεύμονα στους οποίους εντοπίζονται μεταλλάξεις γονιδίου *EGFR* ή σε τριπλά αρνητικούς (triple negative) ασθενείς για τον καρκίνο του μαστού παρέχεται η δυνατότητα προσαρμοσμένης φαρμακευτικής αγωγής με σκοπό την καλύτερη και εξατομικευμένη θεραπεία. Ακόμα σε περιπτώσεις σπάνιων ασθενειών, που ενδεχομένως να ήταν δύσκολη η αντιμετώπιση τους, η ανάλυση του γονιδιώματος ή



του συνόλου των εξωνίων, μπορεί να αποκαλύψει μία σειρά νέων γενετικών μεταλλάξεων και να οδηγήσει σε πιο ακριβή πρόγνωση. Ο εντοπισμός μιας τέτοιας μετάλλαξης είναι επωφελής τόσο για την έρευνα όσο και για την θεραπεία και πιθανόν την μελλοντική ανακάλυψη φαρμάκων.

Οι λανθασμένοι ή ελλιπείς σχολιασμοί μπορούν, σε αντίθεση, να οδηγήσουν τους ερευνητές στο να παραβλέψουν δυνητικές μεταλλάξεις DNA που σχετίζονται με την εκάστοτε ασθένεια και να αποδυναμώσουν ενδιαφέρουσες μεταλλάξεις σε μία ομάδα ψευδών θετικών αποτελεσμάτων. Για το λόγο αυτό είναι απαραίτητη η ανάπτυξη επαρκών εργαλείων (variant annotation tools), τα οποία θα επιτρέπουν στους χρήστες να εντοπίσουν και να χαρακτηρίσουν μεταλλάξεις που αφορούν το γονιδίωμα σε ένα περιβάλλον υπολογιστικής αποθήκης (cloud) [18].

Υπάρχουν διαθέσιμα ποικίλα μέσα για τον σχολιασμό μεταλλάξεων, ανάμεσα στα οποία ανήκουν σύνολα εργαλείων, πολύ – πλατφόρμες, βάσεις δεδομένων, γονιδιωματικές βιβλιοθήκες και λογισμικά προγράμματα. Μερικά από αυτά αναφέρονται στην συνέχεια, με σκοπό την ενημέρωση και την κατανόηση της λειτουργικότητας τους.

### 1.5.1 dbNSFP.

Ένα ευρέως διαδεδομένο εργαλείο είναι το dbNSFP. Συγκεκριμένα πρόκειται για μία βάση δεδομένων, η οποία αναπτύχθηκε για την λειτουργική πρόβλεψη και τον σχολιασμό όλων των πιθανών μη συνώνυμων μεταλλάξεων ενός νουκλεοτιδίου (nsSNVs) στο ανθρώπινο γονιδίωμα. Η πιο πρόσφατη έκδοση βασίζεται στην έκδοση Gencode 29 / Ensembl 94 και περιλαμβάνει συνολικά 84.013.490 nsSNVs και ssSNVs (splicing-site SNVs). Συλλέγει predicting scores (αναμενόμενα αποτελέσματα) από 37 (τριάντα επτά) αλγόριθμους πρόβλεψης (συγκεκριμένα τους SIFT, SIFT4G, Polyphen2 - HDIV, Polyphen2-HVAR, LRT, MutationTaster2, MutationAssessor, FATHMM, MetaSVM, MetaLR, CADD, CADD\_hg19, VEST4, PROVEAN, FATHMM-cod, FATHMM-MKL, fitCons x 4, LINSIGHT, DANN, GenoCanyon, Eigen, Eigen-PC, M-CAP, REVEL, MutPred, MVP, MPC, PrimateAI, GEOGEN2, BayesDel\_addAF, BayesDel\_noAF, ClinPred, LIST-S2, ALoFT), 9 (εννέα) conservation scores - βαθμολογίες διατήρησης (ex. PhyloP x 3, phastCons x 3, GERP ++, SiPhy και bStatistic) και διάφορες άλλες σχετικές πληροφορίες, συμπεριλαμβανομένων των συχνοτήτων αλληλόμορφων που παρατηρήθηκαν στα δεδομένα φάσης 3 του 1000 Genomes Project, UK10K cohorts δεδομένα, δεδομένα κοινοπραξίας ExAC, δεδομένα gnomAD και δεδομένα NHLBI Exome Sequencing Project ESP6500. Ακόμα διάφορα αναγνωριστικά γονιδίων (gene IDs) από διαφορετικές βάσεις δεδομένων, λειτουργικές περιγραφές γονιδίων, γονδιακή έκφραση, πληροφορίες αλληλεπίδρασης γονιδίων κ.α. Ορισμένα περιεχόμενα dbNSFP (ενδέχεται να μην είναι ενημερωμένα) είναι επίσης προσβάσιμα μέσω [variant tools](#), ANNOVAR, KGGSeq, [VarSome](#), UCSC Genome Browser [Variant Annotation](#) Integrator, [Ensembl Variant Effect Predictor](#), [SnpSift](#) και [HGMD](#). Ωστόσο ότι τμήματα του dbNSFP έχουν συγκεκριμένες απαιτήσεις ή χρειάζονται άδεια για μη ακαδημαϊκή χρήση και ενδέχεται να μην μπορούν να χρησιμοποιηθούν [30].

### 1.5.2 BLAST

Εντοπίζει περιοχές ομοιότητας μεταξύ βιολογικών αλληλουχιών. Το πρόγραμμα συγκρίνει αλληλουχίες νουκλεοτιδίων ή πρωτεϊνών με βάσεις δεδομένων

αλληλουχίας και υπολογίζει την στατιστική σημαντικότητα [31].

### 1.5.3 BLAT (Search Genome)

Αναφορικά με το DNA, το πρόγραμμα έχει σχεδιαστεί ώστε να εντοπίζει γρήγορα αλληλουχίες με ομοιότητα μεγαλύτερη ή ίση του 95%, σε μήκος 25 βάσεων και άνω. Λειτουργεί με αντίστοιχο τρόπο στο RNA όπως και στο DNA, με μικρές διαφορές στον δείκτη. Είναι σημαντικό να τονιστεί ότι δεν είναι το ίδιο πρόγραμμα με το BLAST [32].

### 1.5.4 BioMart

Προσφέρει διεπαφή με την αυξανόμενη συλλογή βάσεων δεδομένων, που χρησιμοποιούν το πακέτο λογισμικού BioMart. Το πακέτο επιτρέπει την ανάκτηση μεγάλης ποσότητας δεδομένων με ομοιομορφία [12].

### 1.5.5 ENCODE

Το ENCODE αποτελεί μία βάση δεδομένων η οποία παράγει δεδομένα υψηλής ποιότητας και παράλληλα αναλύει τα δεδομένα αυτά με ολοκληρωμένο τρόπο. Η εγκυκλοπαίδεια ENCODE οργανώνει τα κυριότερα προϊόντα ανάλυσης σε σχολιασμούς και παρέχει εργαλεία για αναζήτηση και οπτικοποίηση. Παρέχει δύο επίπεδα σχολιασμού:

1. Σχολιασμούς ολοκληρωμένου επιπέδου (Integrative-level annotations), που ενσωματώνουν πολλούς τύπους πειραματικών δεδομένων και σχολιασμούς επιπέδου εδάφους.
2. Σχολιασμούς σε επίπεδο εδάφους (Ground-level annotations), οι οποίοι προέρχονται απευθείας από τα πειραματικά δεδομένα, που παράγονται συνήθως από ομοιόμορφους αγωγούς επεξεργασίας [33].

### 1.5.6 GENCODE

Πρόκειται για βάση δεδομένων, στην οποία τα αποτελέσματα της ανάλυσης της βασίζονται στο γονιδίωμα. Στόχος της είναι η ταυτοποίηση και πιστοποίηση όλων των γονιδιακών χαρακτηριστικών στο ανθρώπινο γονιδίωμα και το γονιδίωμα του ποντικού, με μεγάλη σχετικότητα στηριζόμενοι σε βιολογικές αποδείξεις, με αποτέλεσμα τον σχολιασμό προς όφελος της βιοϊατρικής και γονιδιοματικής έρευνας [34].

### 1.5.7 Consensus Coding Sequence

Η Consensus Coding Sequence περιέχει ένα σύνολο δεδομένων υψηλής ποιότητας περιοχών, που κωδικοποιούν πρωτεΐνες. Οι πρωτεΐνες αυτές επισημαίνονται πανομοιότυπα τόσο στο γονιδίωμα του ανθρώπου όσο και στο γονιδίωμα του ποντικού, σε σχολιασμούς γονιδιώματος που διεξάγονται ανεξάρτητα από την ομάδα του Ensembl στο NCBI και την ομάδα του Ensembl στο EMBL-EBI. Αυτό το σύνολο δεδομένων είναι το προϊόν μίας διεθνούς συνεργασίας που

περιλαμβάνει τα NCBI, Ensembl, HUGO Gene Nomenclature Committee, Mouse Genome Informatics και το University of California, Santa Cruz [31].

### 1.5.8 ANNOVAR

Αποτελεσματικό και ευρεία διαδεδομένο εργαλείο λογισμικού, για τη χρήση ενημερωμένων πληροφοριών ως προς τον λειτουργικό σχολιασμό γενετικών μεταλλάξεων. Οι μεταλλάξεις αυτές ανιχνεύονται από διαφορετικά γονιδιώματα (συμπεριλαμβανομένων του ανθρώπινου γονιδιώματος *hg18*, *hg19*, *hg38*, του ποντικού, του σκώληκα, της μύγας, της μαγιάς (ζυμωμύκητας) και άλλων). Το ANNOVAR είναι εργαλείο για τον σχολιασμό μεμονωμένων νουκλεοτιδικών παραλλαγών (SNVs) καθώς και την μελέτη εισαγωγών ή διαγραφών (π.χ. εξέταση της λειτουργικής συνέπειας των παραλλαγών στα γονίδια, και παρεμβολές / διαγραφές, συμπερίληψη κυτταρογενετικών ζωνών, αναφορά αποτελεσμάτων λειτουργικής σημασίας, εύρεση παραλλαγών σε διατηρημένες περιοχές, προσδιορισμός παραλλαγών που αναφέρονται στο 1000 Genomes Project και dbSNP και άλλα) [26]. Λαμβάνοντας υπόψη τα παραπάνω, το λογισμικό αυτό μπορεί να πραγματοποιήσει:

1. Σχολιασμό βάση γονιδίων (Gene-based annotation).
2. Σχολιασμό βάση περιοχής (Region-based annotation).
3. Σχολιασμό βάση φίλτρου (Filter-based annotation).
4. Άλλες λειτουργίες.

### 1.5.9 VAT (Variant Annotation Tool)

Εργαλείο λογισμικού για τον λειτουργικό σχολιασμό μεταλλάξεων πολλαπλών γονιδιωμάτων που βρίσκονται στο επίπεδο της μεταγραφής, σε ένα υπολογιστικό περιβάλλον. Επιπλέον χρησιμοποιείται για την λήψη στατιστικών στοιχείων μεταξύ γονιδίων και ατόμων. Το VAT επιτρέπει την οπτικοποίηση των αποτελεσμάτων των διάφορων μεταλλάξεων και ενσωματώνει συχνότητες αλληλομόρφων και δεδομένα γονοτύπων διαφορετικών ατόμων, διευκολύνοντας κατά αυτόν τον τρόπο την συγκριτική ανάλυση μεταξύ διαφορετικών ομάδων ατόμων [22].

### 1.5.10 VAAST (Variant Annotation, Analysis and Search Tool)

Εργαλείο το οποίο χρησιμοποιεί ένα τεστ αθροιστικής συσχέτισης των μεταλλάξεων, το οποίο συνδυάζει την υποκατάσταση του αμινικού οξέος (AAS) και τις συχνότητες των αλληλομόρφων. Έχει σχεδιαστεί αποκλειστικά για την ταυτοποίηση αλληλομόρφων που σχετίζονται με ασθένειες από δεδομένα αλληλούχισης επόμενης γενιάς (next generation sequencing). Αντίστοιχα προγράμματα που βασίζονται σε αυτό είναι τα VAAST 2.0 και pVAAST.

Το **VAAST 2.0** είναι η πιο πρόσφατη έκδοση του VAAST. Ενσωματώνει ισχυρά μοντέλα αλληλουχίας διάφορων ειδών, τα οποία βελτιώνουν περαιτέρω την ακρίβεια στη διαφοροποίηση μεταξύ καλοηθών και κακοηθών παραγόντων που προκαλούν ασθένειες.

## **2 Κεφάλαιο. Υλικά και Μέθοδοι**

## 2.1 Βάσεις Δεδομένων για Εξόρυξη Πληροφοριών με σκοπό τον Χαρακτηρισμό Γονιδιωματικών και Φαρμακογονιδιωματικών δεικτών στο δείγμα.

Οι ακολουθίες ολόκληρου του γονιδιώματος ή οι ακολουθίες των εξωνίων μπορούν εύκολα να αναλυθούν χρησιμοποιώντας εμπορικά διαθέσιμες ή ιδιόκτητες πλατφόρμες, οι οποίες προσφέρουν μεγάλο βαθμό αξιοπιστίας και είναι αρκετά οικονομικές δεδομένου των όσων προσφέρουν [20]. Στην συγκεκριμένη εργασία, κατά το στάδιο της μελέτης των πολυμορφισμών που εμφανίζονται στους 56 (πενήντα έξι) ασθενείς του δείγματος, χρησιμοποιήθηκαν για τον σκοπό αυτό τα εξής εργαλεία σχολιασμού: η PharmGKB και η VEP της ensembl.

### 2.1.1 PharmGKB

Η PharmGKB χαρακτηρίστηκε το 2000 ως η πρώτη μετά – γονιδιωματική βάση δεδομένων για την περιγραφή, αποθήκευση και επιμέλεια γονοτυπικών και φαινοτυπικών δεδομένων που αφορούν φαρμακογονιδιωματικές έρευνες. Συνδυαστικά με την CPIC και την PGRN αποτελούν την πιο περιεκτική πηγή σχετικά με τα φαρμακογονίδια, τις μεταλλάξεις που σχετίζονται με αυτά καθώς και για τα φαρμακοκινητικά και φαρμακοδυναμικά μονοπάτια που μας ενδιαφέρουν. Ακόμα εξετάζει τις επιδράσεις των παραπάνω σε φαινότυπους σχετικούς με τα φάρμακα και προσφέρει στην φαρμακογονιδιωματική οδηγίες για κλινικούς σχολιασμούς και δοσολογίες φαρμάκων. Στην πράξη αποτελεί τράπεζα βιοδεικτών, οι οποίοι μπορεί να σχετίζονται με την εμφάνιση μεταλλάξεων [10].

### 2.1.2 Ensembl

Το Ensembl είναι ένα πρόγραμμα γονιδιωματικής περιήγησης, το οποίο αφορά γονιδιώματα σπονδυλωτών και υποστηρίζει την έρευνα σε συγκριτική γονιδιωματική, εξέλιξη, αλληλούχιση μεταλλάξεων και μεταγραφικών ρυθμίσεων. Το πρόγραμμα σχολιάζει (annotates) γονίδια, υπολογίζει πολλαπλές ευθυγραμμίσεις, προβλέπει τη ρυθμιστική ακολουθία και συλλέγει δεδομένα ασθενειών. Στο σύνολο του συμπεριλαμβάνονται τα εργαλεία BLAST, BLAT, BioMart και Variant Effect Predictor (VEP), για όλα τα υποστηριζόμενα είδη [25].

### 2.1.3 Variant Effect Predictor

Variant Effect Predictor (γνωστό ως VEP), ένα ιδιαίτερα χρήσιμο σύνολο εργαλείων που αποσκοπεί στην ανάλυση, τον σχολιασμό και την ιεράρχηση των γονιδιωματικών μεταλλάξεων σε κωδικοποιημένες και μη περιοχές ενός γονιδίου. Προσφέρει πρόσβαση σε μία εκτενή συλλογή γενετικών μεταλλάξεων και δυνατότητα επιλογής ανάμεσα σε πολλαπλές διεπαφές, σύμφωνα με τις διαφορετικές απαιτήσεις της εκάστοτε ανάλυσης. Ο σχολιασμός αφορά SNPs, εισαγωγές, διαγραφές, CNVs ή δομικές παραλλαγές σε γονίδια, μεταγραφές και πρωτεϊνικές αλληλουχίες, καθώς και ρυθμιστικές περιοχές. Για γνωστές ή αλληλεπικαλυπτόμενες μεταλλάξεις στα αποτελέσματα συμπεριλαμβάνονται συχνότητες αλληλομόρφων και

πληροφορίες για ασθένειες ή και φαινοτύπους. Το λογισμικό αυτό πιο αναλυτικά χρησιμοποιείται με στόχο τον εντοπισμό:

1. Γονιδίων και μεταγραφών που επηρεάζονται από τις μεταλλάξεις.
2. Της τοποθεσίας των μεταλλάξεων στο γονιδίωμα (π.χ. αν βρίσκεται έναντι κάποιας μεταγραφής, στην αλληλουχία κωδικοποίησης, σε τμήμα RNA που δεν κωδικοποιείται, σε ρυθμιστικές περιοχές κ.ο.κ.).
3. Ενδεχόμενων συνεπειών των μεταλλάξεων στις πρωτεϊνικές αλληλουχίες (π.χ. διακοπή απόκτησης, απώλεια ή διακοπή απώλειας, αλλαγή του πλαισίου πρωτεΐνης και άλλα).
4. Γνωστών μεταλλάξεων που μπορεί να ταιριάζουν στο γονιδιακό προφίλ άλλου ασθενούς και ενδεχομένως να συσχετιζόμενες μικρές συχνότητες αλληλομόρφων από το 1000 Genomes Project.
5. Βαθμολογιών SIFT και PolyPhen-2, που αφορούν αλλαγές στις πρωτεϊνικές ακολουθίες [35].

Κάθε έκδοση VEP συνδέεται με μια συγκεκριμένη έκδοση ensembl. Με τον τρόπο αυτόν εξασφαλίζεται ότι όλα τα αποτελέσματα είναι σταθερά σε μια κυκλοφορία, γεγονός κρίσιμο για την προέλευση και την αναπαραγωγιμότητα. Για να αποφευχθεί η παρερμηνεία μίας μετάλλαξης που βασίζεται σε προηγούμενη έκδοση μεταγραφής ή πρωτεΐνης, το αποτέλεσμα (output) περιλαμβάνει το αναγνωριστικό και την έκδοση στις HGVS coding περιγραφές. Οι σχολιασμοί που προκύπτουν από την VEP χρησιμοποιούνται ως inputs (εισαγωγές) σε εργαλεία για την ενδελεχή διερεύνηση των εναλλακτικών σχολιασμών, όπως το GEMINI. Μπορεί να απλοποιηθεί σε μεγάλο βαθμό την ερμηνεία των μεταλλάξεων και είναι ένα ευέλικτο εργαλείο αξίας για κάθε έργο που απαιτεί λεπτομερή σχολιασμό μεταλλάξεων ακολουθίας. Είναι δωρεάν στην χρήση και διαθέσιμο προς όλους [19].

## 2.2 Διαδικασία επεξεργασίας δεδομένων πενήντα έξι (56) φακέλων ασθενών.

Η διαδικασία επεξεργασίας των δεδομένων πραγματοποιήθηκε συνολικά σε τρία στάδια. Στο πρώτο στάδιο της διαδικασίας πραγματοποιήθηκε η επεξεργασία των δεδομένων των 56 (πενήντα έξι) ασθενών στην βάση δεδομένων της ensembl. Ακολούθως στο δεύτερο στάδιο, η επεξεργασία γίνεται στην σελίδα της PharmKGB με σκοπό την επαλήθευση των αποτελεσμάτων που προέκυψαν στην πρώτη φάση της μελέτης. Η διαδικασία ολοκληρώνεται κατά το τρίτο στάδιο της επεξεργασίας, όπου για τον κάθε rsID γίνεται εύρεση της θέσης του χρωμοσώματος του.

### 2.2.1 Πρώτο Στάδιο Επεξεργασίας:

Για την εκτέλεση του πρώτου σταδίου της επεξεργασίας πραγματοποιήθηκε η παραλαβή των 56 vcf αρχείων, όπως αυτά προκύπτουν από τον όμιλο διαγνωστικών

βιοπαθολογικών εργαστηρίων «Ανάλυση Ιατρική Α.Ε.». Τα αρχεία αυτά αφορούν Έλληνες ασθενείς με ως επί το πλείστον κοινά χαρακτηριστικά (π.χ. ηλικία, φύλο). Στην συνέχεια έγινε εξαγωγή των αρχείων αυτών σε υπολογιστικά φύλλα excel (.txt) και πραγματοποιήθηκε εισαγωγή των vcf αρχείων στην βάση δεδομένων «ensembl» (<https://www.ensembl.org/index.html>) με σκοπό την επεξεργασία τους (Εικόνα 1, Παράρτημα).

Απαραίτητη προϋπόθεση αποτέλεσε η επιλογή κατάλληλης έκδοσης του γονιδιώματος, ώστε να είναι συμβατό με αυτόν που εντοπίζεται στα προς επεξεργασία αρχεία, εν προκειμένω η GRCh37 (Εικόνα 2, Παράρτημα) και μεταφορά στην αντίστοιχη σελίδα (Εικόνα 3, Παράρτημα).

Στην συνέχεια πραγματοποιείται η εκκίνηση της VEP (Εικόνες 4,5 του Παραρτήματος) καθώς και η μεταφόρτωση του εκάστοτε αρχείου στην VEP και επιλογή των απαραίτητων συνθηκών με σκοπό την εξαγωγή επιθυμητών αποτελεσμάτων (Εικόνα 6, Παράρτημα). Τα κριτήρια επιλογής των συνθηκών είναι απολύτως συνυφασμένα με τον σκοπό της αναζήτησης της κάθε μελέτης. Στην προκειμένη περίπτωση ήταν γνωστές οι θέσεις των γονιδίων και η αναζήτηση αφορούσε το ποιες από αυτές τις θέσεις αποτελούν βιοδείκτες.

Σύμφωνα με την σειρά εμφάνισης στο πρόγραμμα, ακολουθήθηκαν τα εξής βήματα:

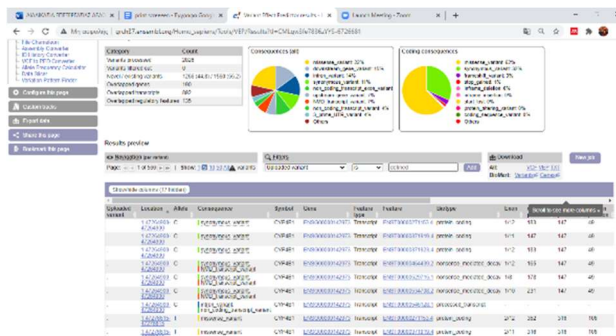
1. Από το μενού *Transcript database to use* επιλογή των «Ensembl/GENCODE transcripts». (Εικόνα 7, Παράρτημα)
2. Από το μενού *Additional configurations*, επιλογή: «Gene symbol, Transcript version, CCDS, Protein, UniProt, HGVS». (Εικόνα 8, Παράρτημα)
3. Από το μενού *Variants and frequency data* επιλέχθηκαν τα «1000 Genomes global minor allele frequency, 1000 Genomes continental allele frequencies, ESP allele frequencies, gnomAD (exomes) allele frequencies, PubMed IDs for citations of co-located variants, Include flagged variants». (Εικόνα 9, Παράρτημα)
4. Από το μενού *Additional annotation* τα «Transcript biotype, Exon and intron numbers, Transcript support level, APPRIS, MANE, Identify canonical transcripts, Protein domains, Phenotypes». (Εικόνα 10, Παράρτημα)
5. Από το μενού: Predictions, επιλέχθηκαν τα : SIFT, PolyPhen, dbNSFP, CADD, Condel, LoFtool, MPC, BLOSUM62. (Εικόνες 11, 12 στο Παράρτημα)

Πιο συγκεκριμένα, για την dbNSFP, αφού πραγματοποιηθεί η ενεργοποίηση της, επιλέχθηκαν τα εξής χωρία: SIFT (score, pred), PolyPhen2\_HVAR (score, pred), mutationtaster (score, pred), meta\_R (score, pred), PROVEAN (score, pred), DANN (score), CADD\_raw, CADD\_phred, GERP\_NR, GERP++\_R.

Σημειώνεται ότι η επιλογή αυτών των in silico scores, γίνεται με χρήση του πλήκτρου ctrl. (Εικόνα 13, Παράρτημα)

6. Τέλος γίνεται επιλογή του πλήκτρου RUN και πραγματοποιείται η εκκίνηση του προγράμματος. (Εικόνα 14, Παράρτημα)

Μετά το τέλος της αναζήτησης αποθηκεύονται και οι τρεις τύποι αρχείων, όπως αυτοί προκύπτουν. Τα αποτελέσματα της αναζήτησης φαίνονται στην εικόνα 15.



Εικόνα 15. Αποτελέσματα αναζήτησης στην ensembl.

Η διαδικασία αυτή όπως παρουσιάστηκε επαναλαμβάνεται και για τα 56 (πενήντα έξι) αρχεία ασθενών, με τον ίδιο ακριβώς τρόπο. Τα αρχεία διαγράφονται από την βάση δεδομένων της ensembl, δηλαδή δεν διατηρούνται μετά το πέρας της αναζήτησης.

Στην πορεία πραγματοποιήθηκε επιπλέον μετατροπή των vcf/ ver αρχείων σε αρχεία excel, ενώ τα διπλότυπα (επαναλαμβανόμενες καταχωρήσεις) που αντιστοιχούν σε πολλαπλές εγγραφές μιας παραλλαγής σε πολλά μετάγραφα, αφαιρέθηκαν εισάγοντας στην γραμμή εντολών τα παρακάτω:

```
for i in path/to/*.txt;do awk '!seen[$2]H' ${i}>${i}_cut.txt; done
```

Τέλος στα αρχεία excel που προκύπτουν γίνεται επιλογή των στηλών : #Uploaded\_variation, Location, Allele, Consequence, IMPACT, SYMBOL, Gene, Feature\_type, Feature, Existing\_variation, HGNC\_ID, SIFT, PolyPhen, DOMAINS, HGVS\_OFFSET, AF, AFR\_AF, AMR\_AF, EAS\_AF, EUR\_AF, SAS\_AF, AA\_AF, EA\_AF, gnomAD\_AF, gnomAD\_AFR\_AF, gnomAD\_AMR\_AF, gnomAD\_ASJ\_AF, gnomAD\_EAS\_AF, gnomAD\_FIN\_AF, gnomAD\_NFE\_AF, gnomAD\_OTH\_AF, gnomAD\_SAS\_AF, CLIN\_SIG, SOMATIC, PHENO, PUBMED, MOTIF\_NAME, MOTIF\_POS, HIGH\_INF\_POS, MOTIF\_SCORE\_CHANGE, CADD\_phred, CADD\_raw, DANN\_score, GERP++\_NR, GERP++\_RS, MetaLR\_pred, MetaLR\_score, MutationTaster\_pred, MutationTaster\_score, PROVEAN\_pred, PROVEAN\_score, Polyphen2\_HVAR\_pred, Polyphen2\_HVAR\_score, SIFT\_pred, SIFT\_score2, MPC, ada\_score, rf\_score, BLOSUM62, LoFtool, Condol, CADD\_PHRED, CADD\_RAW, PHENOTYPES.

Στην «Εικόνα 15» που ακολουθεί, συνοψίζονται τα αποτελέσματα της αναζήτησης, όπως αυτά έχουν προκύψει σύμφωνα με τα χαρακτηριστικά (scores) που επιλέχθηκαν σε κάθε στάδιο κατά την εκτέλεση του προγράμματος. Στην πρώτη στήλη των αποτελεσμάτων όπως αυτά εμφανίστηκαν από το πρόγραμμα παρουσιάζεται η χρωμοσωμική θέση, ενώ στην δεύτερη στήλη το αλληλόμορφο. Η τέταρτη στήλη περιλαμβάνει τον συμβολισμό του κάθε γονιδίου, για την διευκόλυνση του χρήστη.

Αναφορικά με τα διαγράμματα της εικόνας, είναι ενδεικτικά για την συχνότητα εμφάνισης των χαρακτηριστικών (scores) που επιλέχθηκαν κατά την εκτέλεση του



προγράμματος και ανάλυση αυτών. Με μπλε χρώμα απεικονίζεται η μεταφόρτωση από την ensembl 38 στην ensembl 37, ενώ ιδιαίτερα ενδιαφέρον είναι το κόκκινο τμήμα του διαγράμματος, καθώς σε αυτό περιέχονται γονίδια τα οποία είτε δεν περιέχουν μετάλλαξη που αφορά το πλήθος των ασθενών της μελέτης είτε γονίδια που δεν υπάρχουν γενικά στο δείγμα. Συμπερασματικά αυτό ερμηνεύεται με δύο τρόπους: τα γονίδια αυτά είτε δεν περιέχονται καθόλου στον συγκεκριμένο πληθυσμό είτε τυχαία δεν συμπεριλαμβάνονται στο δείγμα.

Εδώ ολοκληρώνεται το πρώτο στάδιο επεξεργασίας των δεδομένων.

### 2.2.2 Δεύτερο Στάδιο Επεξεργασίας

Για την εκτέλεση του δεύτερου σταδίου επεξεργασίας χρειάστηκε μετάβαση στην ιστοσελίδα της PharmKGB (<https://www.pharmgkb.org/>).

Αρχικά αποθηκεύτηκαν τα gene-specific-information-tables και στην συνέχεια έγινε αναζήτηση των rsIDs (μοναδική ταυτότητα της κάθε συντεταγμένης/ μετάλλαξης στο γονίδιο) που εντοπίζονται στους 18 (δέκα – οχτώ) πίνακες της PharmKGB στην βάση δεδομένων dbSNP.

Στην συνέχεια συλλέχθηκε η προκύπτουσα για τον κάθε rsIDs πληροφορία και έγινε ενσωμάτωσή της στους αντίστοιχους allele definition tables. Ακολούθως πραγματοποιείται εντοπισμός των rsIDs από τους 18 αυτούς πίνακες σε κάθε ένα από τα 56 αρχεία και δημιουργείται φύλλο excel, το οποίο περιέχει τα rsIDs που εντοπίστηκαν και την συχνότητα εμφάνισής τους.

Τέλος δημιουργείται φύλλο excel που παρουσιάζει τα rsIDs, την αντίστοιχη πληροφορία από τα vcf/ ver αρχεία και τους allele definition tables/ dbSNP.

### 2.2.3 Τρίτο Στάδιο Επεξεργασίας

Κατά το τρίτο στάδιο επεξεργασίας και με άξονα τον εκάστοτε rsID, πραγματοποιείται έρευνα για την εύρεση της θέσης χρωμοσώματος του.

Η θέση αυτή αναζητείται και στα 56 vcf (αρχικά) αρχεία, ώστε να γίνει συνολική καταγραφή γονοτύπου των ασθενών που φέρουν τα rsID.

Ολοκληρώνοντας την διαδικασία, πραγματοποιείται δημιουργία νέου φύλλου excel, στο οποίο καταγράφεται η συχνότητα με την οποία εμφανίζεται κάθε rsID.

### **3 Κεφάλαιο. Αποτελέσματα**

Τα αρχικά δεδομένα κωδικοποιήθηκαν και οργανώθηκαν με άξονα τον εκάστοτε πολυμορφισμό (rsID) που υπάρχει καταχωρημένος στην PharmGKB και εντοπίστηκε στον φάκελο του κάθε ασθενούς, που παρουσίαζε την ανάλυση του γονιδιώματος του και συγκεκριμένα των εξωνίων του. Στόχος είναι, αρχικά, ο εντοπισμός των ασθενών που εμφανίζουν τους πιο συνήθεις πολυμορφισμούς, αυτούς δηλαδή που βρίσκονται καταχωρημένοι στην PharmGKB και στην συνέχεια σύμφωνα με τους πολυμορφισμούς που αναγνωρίστηκαν στον καθένα από αυτούς τους ασθενείς, η καταγραφή των γονοτύπων (ομόζυγος για το μεταλλαγμένο ή το φυσιολογικό αλληλόμορφο, ή ετερόζυγος για το αλληλόμορφο αναφοράς) στο σύνολο των ασθενών του δείγματος, ώστε τελικά να εξαχθεί κάποιο συμπέρασμα για τις συνέπειες που ενδέχεται να προκληθούν από τους πολυμορφισμούς στο δείγμα. Στην κλινική αξιοποίηση αυτού, μία ομαδοποίηση και οπτικοποίηση των δεδομένων θα καταστήσει δυνατή όχι μόνο την ευκολότερη κατανόηση στη συσχέτισης των γονοτύπων και των οδηγιών για την προσαρμογή της θεραπείας, αλλά και το ποσοστό εμφάνισης μεταλλάξεων στον υποπληθυσμό τον οποίο αφορά η μελέτη. Για τον σκοπό αυτό η πληροφορία οργανώθηκε σε πίνακα, στον οποίο περιέχονται όλοι οι πολυμορφισμοί, οι πληροφορίες που αφορούν αυτούς τους πολυμορφισμούς και τα γονίδια στα οποία αυτοί ανήκουν και που υπάρχουν στην PharmGKB και εντοπίστηκαν στους ασθενείς. Κατόπιν προστέθηκαν οδηγίες φαρμακογονιδιωμιακού χαρακτήρα για την πιθανή οδηγία ή/και προσαρμογή θεραπειών. Παράλληλα τα δεδομένα οργανώθηκαν και σε δεύτερο πίνακα, σύμφωνα με τον γονότυπο των ασθενών σε σχέση με το εκάστοτε γονίδιο.

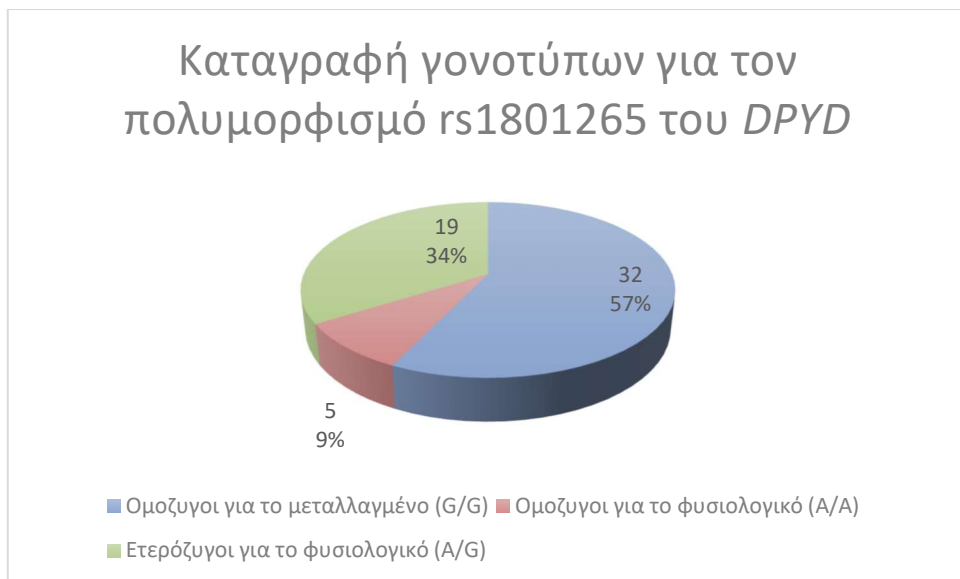
Ενδεικτικά καταγράφηκε και οπτικοποιήθηκε ο γονότυπος της μετάλλαξης rs1801265 του γονιδίου *DPYD* στο δείγμα των 56 ασθενών. Το γονίδιο *DPYD* κωδικοποιεί την αφυδρογονάση διυδροπυριμιδίνης (DPD), ένα ένζυμο που καταλύει το στάδιο του ρυθμού του περιορισμού στον μεταβολισμό της φθοροουρακίλης.

Όπως προκύπτει από την dbSNP για τον πολυμορφισμό rs1801265 το φυσιολογικό αλληλόμορφο συμβολίζεται ως A και το μεταλλαγμένο ως G. Έτσι προκύπτουν τρία είδη γονοτύπων:

- A/A που αντιπροσωπεύει το ομόζυγο φυσιολογικό
- A/G που αντιπροσωπεύει το ετερόζυγο άτομο
- G/G για την ομόζυγη μεταλλαγμένη έκφραση

Συγκεκριμένα στο δείγμα των ασθενών που μελετήθηκαν στην παρούσα εργασία, ο πολυμορφισμός rs1801265 εντοπίστηκε σε 51 (πενήντα έναν) από αυτούς, στην αναλογία που παρουσιάζεται στο «Διάγραμμα 1» που ακολουθεί. Αυτό πρακτικά υποδηλώνει ότι ένα πολύ μεγάλο ποσοστό εμφάνισης στο δείγμα, της τάξης του 91% , αναμένεται να παρουσιάσει παρόμοια χαρακτηριστικά που αφορούν αυτόν τον πολυμορφισμό, κάτι το οποίο μπορεί να αποτελεί ενδιαφέρουσα μελέτη για το εγγύς μέλλον.

## Καταγραφή γονοτύπων για τον πολυμορφισμό rs1801265 του *DPYD*



Διάγραμμα 1.

Καταγραφή της γονοτυπικής έκφρασης του πολυμορφισμού για το γονίδιο *DPYD*.

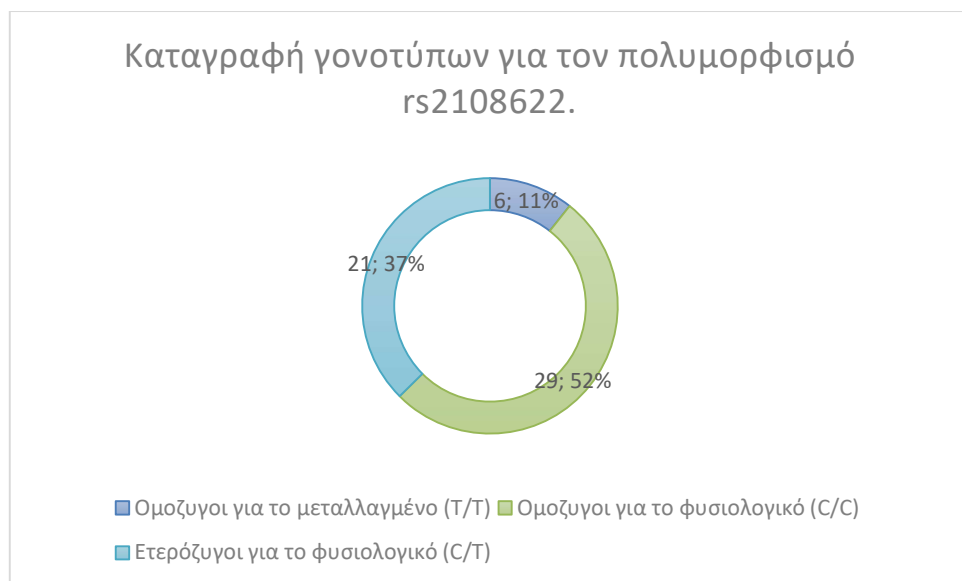
Ιδιαίτερο ενδιαφέρον αποτελεί το γεγονός ότι ο συγκεκριμένος πολυμορφισμός δεν εμφανίζεται (πρόκειται για τα ομόζυγα για το φυσιολογικό αλληλόμορφο αναφοράς άτομα) μόλις στο 9% του δείγματος. Παρόλο που στην συγκεκριμένη εργασία το δείγμα είναι περιορισμένο, δεν αποκλείεται η παρατήρηση αυτή να φανερώνει την ενδεχομένως έντονη συχνότητα εμφάνισης του πολυμορφισμού και σε κάποιο μεγαλύτερο δείγμα.

Η διαδικασία επαναλήφθηκε για τον πολυμορφισμό rs2108622 του γονιδίου *CYP4F2*, που κωδικοποιεί την πρωτεΐνη CYP4F2. Η πρωτεΐνη αυτή καταλύει την υδροξυλίωση της βιταμίνης K (VK1) στην υδροξυλιωμένη μορφή της. Η μετάλλαξη rs2108622 του γονιδίου *CYP4F2*, η οποία αφορά μία λανθασμένη έκφραση V433M με μειωμένη δραστηριότητα της CYP4F2 και μειωμένο μεταβολισμό της VK1, έχει συσχετιστεί με αυξημένη απαίτηση δόσης βαρφαρίνης στους πάσχοντες [7].

Για τον πολυμορφισμό rs2108622 το αλληλόμορφο που χαρακτηρίζεται ως αλληλόμορφο αναφοράς είναι το C, ενώ το μεταλλαγμένο είναι το T. Οι τρεις πιθανοί γονότυποι που εμφανίζονται είναι:

- C/C που αντιπροσωπεύει το ομόζυγο φυσιολογικό
- C/T για τον ετερόζυγο γονότυπο
- T/T για το ομόζυγο μεταλλαγμένο άτομο

Στην παρούσα εργασία ο πολυμορφισμός rs2108622 εντοπίστηκε σε 27 (είκοσι επτά) ασθενείς. Οι 21 (είκοσι ένα) από αυτούς ήταν ετερόζυγοι για το αλληλόμορφο αναφοράς, εμφανίζουν δηλαδή (C/T), ενώ 6 (έξι) ήταν ομόζυγοι για την έκφραση της συγκεκριμένης μετάλλαξης (T/T). Αυτό πρακτικά σημαίνει ότι ο πολυμορφισμός εμφανίζεται σε ένα συνολικό ποσοστό της τάξης του 48%, ενώ οι υπόλοιποι 29 (είκοσι εννέα) ασθενείς που φέρουν φυσιολογικό γονότυπο (C/C) δεν εμφανίζουν τον συγκεκριμένο πολυμορφισμό και αποτελούν το 52%. Αναλυτικά τα δεδομένα παρουσιάζονται στο «Διάγραμμα 2».



Διάγραμμα 2. Απεικόνιση για την μετάλλαξη rs2108622 του γονιδίου CYP4F2.

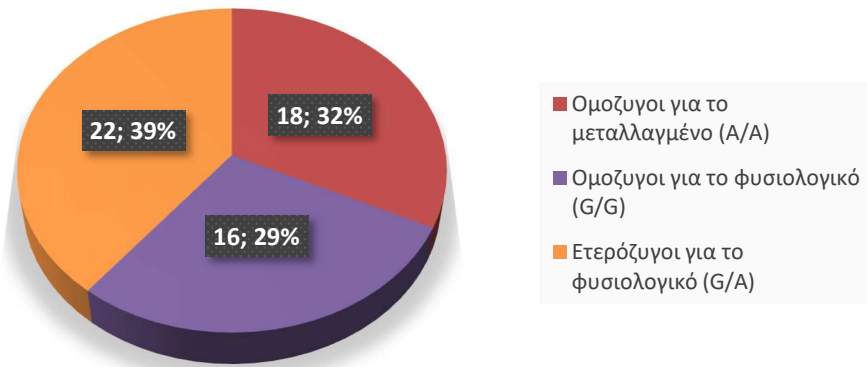
Τέλος η διαδικασία πραγματοποιήθηκε και για την γονοτυπική καταγραφή και απεικόνιση του πολυμορφισμού rs16947 του γονιδίου *CYP2D6*. Το *CYP2D6* καταλύει τον μεταβολισμό ενός μεγάλου αριθμού κλινικά σημαντικών φαρμάκων, συμπεριλαμβανομένων αντικαταθλιπτικών, νευροληπτικών, μερικών αντιαρρυθμικών, λιπόφιλων αναστολέων β-αδρενεργικών υποδοχέων και οπιοειδών [3].

Στην περίπτωση αυτή το αλληλόμορφο αναφοράς είναι το G ενώ ως μετάλλαξη χαρακτηρίζεται η εμφάνιση του αλληλομόρφου A. Οι τρεις γονότυποι για τα άτομα είναι:

- G/G για το ομόζυγο φυσιολογικό
- G/A για το ετερόζυγο
- A/A για το ομόζυγο μεταλλαγμένο

Όπως περιγράφεται και στο «Διάγραμμα 3» που ακολουθεί, η μετάλλαξη αυτή παρατηρήθηκε στους 40 (σαράντα) από τους 56 (πενήντα έξι) ασθενείς του δείγματος. Από τους 40 (σαράντα) αυτούς ασθενείς, οι 18 (δέκα οχτώ) είναι ομόζυγοι για το μεταλλαγμένο γονίδιο (A/A), ενώ οι υπόλοιποι 22 (είκοσι δύο) ετερόζυγοι για το αλληλόμορφο αναφοράς (G/A). Το σύνολο αυτό αντιπροσωπεύει το 71% του δείγματος, κάτι που επίσης υποδηλώνει ότι όπως και στην περίπτωση του *DPYD* πολλοί περισσότεροι από τους μισούς ασθενείς εμφανίζουν την μεταλλαγμένη αυτή εκδοχή του γονιδίου. Ωστόσο και σε αυτήν την περίπτωση μέρος των ασθενών, της τάξης του 29%, φαίνεται να εξαιρείται από τα συμπεράσματα, καθώς τα άτομα είναι ομόζυγα για το αλληλόμορφο αναφοράς (G/G) και συνεπώς δεν εμφανίζουν τον πολυμορφισμό rs16947 που μελετάται για το *CYP2D6*.

### Καταγραφή γονοτύπων για τον πολυμορφισμό rs16947 του CYP2D6



Διάγραμμα 3. Γονοτυπική καταγραφή του πολυμορφισμού rs16947 του CYP2D6.

## **4 Κεφάλαιο. Συμπεράσματα**

Οι βιοδείκτες αποτελούν αναμφίβολα ένα σημαντικό εργαλείο, τόσο στον τομέα της πρόληψης όσο και στην εξατομικευμένη και επιτυχημένη θεραπεία των ασθενών. Στην εποχή μας έχουν γίνει σημαντικά βήματα προόδου στο κομμάτι της αξιοποίησης των πληροφοριών που εξάγονται από τις μελέτες των βιοδεικτών, για την εφαρμογή τους στην παροχή εξειδικευμένων εξετάσεων αλλά και για την εξέλιξη της ιατρικής περίθαλψης.

Σε ένα δείγμα τόσο πληθυσμιακά περιορισμένο όπως αυτό που μελετήθηκε, είναι ιδιαίτερα ενδιαφέρον αυτό που παρουσιάζει ο πολυμορφισμός για το γονίδιο *DPYD*. Η φθοριοουρακίλη ή 5-φλουουρακίλη (5-FU), είναι ένας χημειοθεραπευτικός παράγοντας που ανήκει στην κατηγορία φαρμάκων των φθοροπυριμιδινών. Γενετικές μεταλλάξεις στο γονίδιο *DPYD* μπορεί να οδηγήσουν σε ένζυμα με μειωμένη ή και μηδενική δραστηριότητα. Σε άτομα με τουλάχιστον ένα αντίγραφο μιας μη λειτουργικής μετάλλαξης *DPYD* (όπως για παράδειγμα, *DPYD\*2A* (c.1905+1G> A) ή *DPYD\*13*(c.1679T> G)) ενδέχεται ο μεταβολισμός της φθοριοουρακίλης σε μη φυσιολογικούς ρυθμούς. Κατά συνέπεια, διατρέχουν σοβαρό κίνδυνο λόγω της τοξικότητας της φθοριοουρακίλης, σε βαθμό απειλητικό για την ζωή. Ο κακός μεταβολισμός της φθοριοουρακίλης μπορεί να οδηγήσει σε καταστολή του μυελού των οστών, προβλήματα ευκοιλίας καθώς και νευροτοξικότητα [8].

Η συχνότητα εμφάνισης του πολυμορφισμού στο δείγμα υποδηλώνει την ανεπάρκεια σε μεγάλο ποσοστό στον μεταβολισμό της αφυδρογονάση της διυδροπυριμιδίνης. Η ανεπάρκεια αυτή μπορεί είτε να είναι υπεύθυνη εξ ολοκλήρου είτε απλώς να συμμετέχει (να είναι δηλαδή συνυπεύθυνη) στην δημιουργία κάποια ασθένειας.

Αντίστοιχα αποτελέσματα παρατηρήθηκαν και κατά την μελέτη του πολυμορφισμού rs2108622 του *CYP4F2*. Η δράση του γονιδίου *CYP4F2* σχετίζεται άμεσα με το ένζυμο που κωδικοποιεί το γονίδιο της βαρφαρίνης και την αυξημένη ανάγκη του οργανισμού για την χορήγηση αυτής. Πρέπει να αναφερθεί, ότι η συσχέτιση που περιγράφεται δεν έχει επαληθευθεί με ακρίβεια και απαιτούνται περαιτέρω μελέτες για την εξακρίβωση της. [6] Η βαρφαρίνη (warfarin) αναστέλλει την αναγωγή του εποξειδίου της βιταμίνης K, με αποτέλεσμα την μείωση της έκφρασης της βιταμίνης K (βιταμίνη KH<sub>2</sub>). Κατ' επέκταση αναστέλλεται η πήξη των παραγόντων II, VII, IX και X και των αντιπηκτικών πρωτεϊνών C και S που εξαρτώνται από την βιταμίνη K. Η ελάττωση των τριών εκ των τεσσάρων παραγόντων πήξης που εξαρτώνται από την βιταμίνη K (παράγοντες II, VII και X), οδηγεί σε μειωμένα επίπεδα προθρομβίνης και μία μείωση στο ποσοστό της θρομβίνης που παράγεται και δεσμεύεται σε ινώδες. Το γεγονός αυτό μειώνει την θρομβογεννητικότητα των θρόμβων [7].

Σε ασθενείς οι οποίοι φέρουν την μετάλλαξη rs2108622 για το γονίδιο *CYP4F2* παρατηρείται αδυναμία συντήρησης των επιπέδων της βαρφαρίνης στα φυσιολογικά όρια και αδυναμία στον μεταβολισμό του VK1, με συνέπεια να απαιτείται ως επί το πλείστον άμεση χορήγηση αυξημένης δόσης βαρφαρίνης. Οι ασθενείς που φέρουν γονότυπο C/T, T/T ή φορείς του T χρειάζονται μεγαλύτερη δόση βαρφαρίνης συγκριτικά με τους ομοζυγωτικούς *CYP4F2*. Η παραλλαγή *CYP4F2* rs2108622 που εντοπίστηκε και στο δείγμα των 56 (πενήντα έξι) που μελετήθηκε, οδηγεί σε αύξηση της δόσης της βαρφαρίνης για κάθε T αλληλόμορφο. Ασθενείς με 2 T/T αλληλόμορφα χρειάζονται αύξηση της βαρφαρίνης κατά 1mg/day σε σχέση με τους ασθενείς που φέρουν 2 C/C αλληλόμορφα. Συνεπώς, πρέπει η θεραπεία να είναι εξατομικευμένη και σύμφωνη με τις ανάγκες του κάθε ασθενούς ξεχωριστά.



Τέλος το γονίδιο *CYP2D6*, που μελετήθηκε στο «Διάγραμμα 3» των αποτελεσμάτων, είναι το πιο εκτενώς χαρακτηριζόμενο πολυμορφικό ένζυμο μεταβολισμού φαρμάκων. Είναι πολυμορφικό, με περισσότερες από 70 (εβδομήντα) αλληλικές παραλλαγές να έχουν καταγραφεί μέχρι και σήμερα. Από αυτές, περισσότερες από 15 (δεκαπέντε) κωδικοποιούν ένα ανενεργό ή ακόμα και κανένα ένζυμο, ενώ άλλες κωδικοποιούν ένζυμα με μειωμένη, φυσιολογική ή αυξημένη ενζυμική δράση. Η ανεπάρκεια του ενζύμου *CYP2D6* κληρονομείται στα άτομα ως αυτοσωμικό υπολειπόμενο χαρακτηριστικό. Η έκφραση του *CYP2D6* παρατηρείται έντονα στο ήπαρ αλλά και σε περιοχές του κεντρικού νευρικού συστήματος και έχει μελετηθεί σε πολύ μεγάλο βαθμό με αποτέλεσμα να προκύψουν τέσσερις κατηγορίες μεταβολιστών.

Τα άτομα τα οποία φέρουν αυτή τον πολυμορφισμό rs16947 (7% των Καυκάσιων, περίπου 1% των Ανατολικών) ταξινομούνται ως φτωχοί μεταβολιστές (poor metabolizers). Μεταξύ των υπολοίπων (εκτεταμένοι μεταβολιστές), η ενζυμική δραστηριότητα είναι μεταβλητή, από εξαιρετικά υψηλή (υπερταχείς μεταβολιστές), έως σημαντικά μειωμένη (ενδιάμεσοι μεταβολιστές). Για τα φάρμακα που μεταβολίζονται από το *CYP2D6* σε ανενεργούς μεταβολιστές, οι αναστολές του *CYP2D6* μπορεί να οδηγήσουν σε τοξικότητα, ενώ για φάρμακα που μεταβολίζονται από ενεργούς μεταβολιστές του *CYP2D6*, η προσθήκη ενός αναστολέα του *CYP2D6* θα τείνει να αναστείλει την αποτελεσματικότητα του φαρμάκου. Η γενετική μεταβλητότητα στη δραστηριότητα του *CYP2D6* μπορεί επίσης να επηρεάσει το αποτέλεσμα των φαρμακευτικών αλληλεπιδράσεων του *CYP2D6* [14].

Το *CYP2D6* εμφανίζει έντονη μεταβλητότητα σύμφωνα με τα διαφορετικά εθνικά χαρακτηριστικά του πληθυσμού που μελετάται κάθε φορά, με διαφορές που εντοπίζονται τόσο στη συχνότητα εμφάνισης των αλληλομόρφων όσο και στην ύπαρξη «ειδικών για τον πληθυσμό» αλληλικών παραλλαγών. Στο συγκεκριμένο πληθυσμιακό δείγμα ο πολυμορφισμός rs16947 για το γονίδιο εντοπίστηκε σε ένα αρκετά μεγάλο ποσοστό, καθώς αφορούσε το 71%. Εύκολα μπορεί να εξαχθεί το συμπέρασμα ότι τόσο η έντονη εμφάνιση της μετάλλαξης στο δείγμα, όσο και τα αναμενόμενα προς εμφάνιση χαρακτηριστικά της μετάλλαξης του γονιδίου στους ασθενείς, σχετίζονται ενδεχομένως με την πληθυσμιακή ομάδα του δείγματος (στην προκειμένη περίπτωση καυκάσιοι ασθενείς).

Στο πλαίσιο που αφορά την μελέτη των φαρμακογονιδιωματικών βιοδεικτών και την βιοπληροφορική, οι βάσεις δεδομένων μπορούν σε τέτοιες περιπτώσεις να επέμβουν και με τις πληροφορίες που προσφέρουν να καθοδηγήσουν τον ασθενή σωστά προς την αποφυγή παρενεργειών ή περεταίρω προβλημάτων που μπορούν να δημιουργηθούν από ανεπαρκή ή λανθασμένη φαρμακευτική αγωγή.

Στο το παράδειγμα της αφυδρογονάσης της διυδροπυριμιδίνης που μελετήθηκε παραπάνω, είναι σημαντικό οι επιστήμονες υγείας ή ακόμα και ο ίδιος ο ασθενής που πάσχει να μπορεί να γνωρίζει ότι η αλληλεπίδραση με την φθοριοουρακίλη (π.χ. αγωγή με fluouracil) είναι τοξική και μπορεί να οδηγήσει σε πολυμορφισμό, προκειμένου να αποφεύγεται. Για το παράδειγμα που αφορά το *CYP2D6*, παρά τα κοινά χαρακτηριστικά που μπορεί να εμφανίζονται λόγω εθνικής ομοιογένειας, η διαφορετική μεταβολική ικανότητα του εκάστοτε ασθενούς διαμορφώνει με ξεχωριστό τρόπο σε κάθε περίπτωση την φαρμακευτική αγωγή που θα πρέπει να ακολουθήσει, προκειμένου να αποφεύγονται φαινόμενα κακού μεταβολισμού. Κατ' αυτόν τον τρόπο αντίστοιχα συμπεράσματα μπορούν να διεξαχθούν και για άλλους βιοδείκτες.

Παρά τις συνθήκες κάτω από τις οποίες πραγματοποιήθηκε η συγκεκριμένη μελέτη, οι πληροφορίες που εξάγονται αποτελούν ένδειξη για τα σημαντικά οφέλη

της γνώσης των ιδιαίτερων χαρακτηριστικών που μπορεί να προκύψουν από τις διαφορετικές παραλλαγές των γονιδίων στα άτομα και της σωστής εφαρμογής της εξατομικευμένης θεραπείας. Η μελέτη αυτή μπορεί να επαναληφθεί και σε μεγαλύτερο πληθυσμιακό δείγμα, με μεγαλύτερη ποικιλομορφία και χαρακτηριστικά από όπου θα προκύψουν περισσότερες πληροφορίες σχετικά με τους εκάστοτε πολυμορφισμούς. Πρόκειται για ένα σημαντικό έργο παρά τις μικρές αδυναμίες του, το οποίο προς το παρόν μπορεί να υλοποιηθεί με την βοήθεια της βιοπληροφορικής και τις εξειδικευμένες γνώσεις του βιοπληροφορικού. Στο μέλλον, ωστόσο, θα μπορούσε να γίνει αυτοματοποίηση της διαδικασίας, παραδείγματος χάριν, μέσω χρήσης εφαρμογής στην οποία ο χρήστης θα μπορεί να αναζητά τους πολυμορφισμούς που υπάρχουν αναφορικά με κάποια πάθηση που μπορεί να φέρει και τις συνέπειες που μπορεί να προκύψουν σε συνδυασμό με τα μοναδικά χαρακτηριστικά που ο ίδιος εμφανίζει, προκειμένου οι μελέτες και οι πληροφορίες που αφορούν τους βιοδείκτες να γίνουν πιο προσιτές και ευνόητες για το ευρύ κοινό και κυρίως για τον πάσχοντα ασθενή και άμεσα ενδιαφερόμενο.

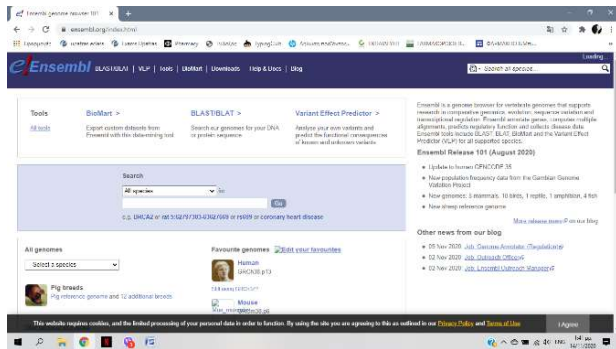
## **5 Κεφάλαιο. Βιβλιογραφία**

- [1] J.K. Aronson, R.E. Ferner, Biomarkers—A General Review, *Curr. Protoc. Pharmacol.* 76 (2017) 9.23.1-9.23.17.
- [2] F. Balloux, O. Brønstad Brynildsrud, L. van Dorp, L.P. Shaw, H. Chen, K.A. Harris, H. Wang, V. Eldholm, From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic, *Trends Microbiol.* 26 (2018) 1035–1048.
- [3] L. Bertilsson, M.-L. Dahl, P. Dalén, A. Al-Shurbaji, Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs, *Br. J. Clin. Pharmacol.* 53 (2002) 111–122.
- [4] S. Best, K. Wou, N. Vora, I.B. Van der Veyver, R. Wapner, L.S. Chitty, Promises, pitfalls and practicalities of prenatal whole exome sequencing, *Prenat. Diagn.* 38 (2018) 10–19.
- [5] J. Chung, K.-W. Lee, C. Lee, S.-H. Shin, S. Kyung, H.-J. Jeon, S.-Y. Kim, E. Cho, C.E. Yoo, D.-S. Son, W.-Y. Park, D. Park, Performance evaluation of commercial library construction kits for PCR-based targeted sequencing using a unique molecular identifier, *BMC Genomics.* 20 (2019) 216.
- [6] J.Y. Chyou, J.L. Mega, M.S. Sabatine, *Pharmacogenetics*, Fourth Ed, Elsevier Inc., 2012.
- [7] A.K. Daly, Chapter 24 - Pharmacogenomics of Warfarin, in: S.B.T.-H. of P. and S.M. Padmanabhan (Ed.), *Academic Press*, San Diego, 2014: pp. 497–507.
- [8] L. Dean, Fluorouracil Therapy and DPYD Genotype, *Med. Genet. Summ.* (2012) 1–18.
- [9] C.E. French, I. Delon, H. Dolling, A. Sanchis-Juan, O. Shamardina, K. Mégy, S. Abbs, T. Austin, S. Bowdin, R.G. Branco, H. Firth, S. Tuna, T.J. Aitman, S. Ashford, W.J. Astle, D.L. Bennet, M. Bleda, K.J. Carss, P.F. Chinnery, S.V. V Deevi, D. Fletcher, D.P. Gale, S.F. Gräf, F. Hu, R. James, M.A. Kasanicki, N. Kingston, A.B. Koziell, H.L. Allen, E.R. Maher, H.S. Markus, S. Meacham, N.W. Morrell, C.J. Penkett, I. Roberts, A. Sanchis-Juan, K.G.C. Smith, H. Stark, K.E. Stirrups, E. Turro, H. Watkins, C. Williamson, T. Young, J.R. Bradley, W.H. Ouwehand, F.L. Raymond, S. Agrawal, R. Armstrong, K. Beardsall, G. Belteki, M. Bohatschek, S. Broster, R. Campbell, R. Chaudhary, C. Costa, A. D'Amore, A. Fitzsimmons, J. Hague, J. Harley, S. Hoodbhoy, R. Kayani, W. Kelsall, S.G. Mehta, R. O'Donnell, S. O'Hare, A. Ogilvy-Stuart, S. Papakostas, S.-M. Park, A. Parker, N. Pathan, M. Prapa, A. Sammut, R. Sandford, K. Schon, Y. Singh, K. Spike, A.L.T. Tavares, D. Wari-Pepple, H.S. Wong, C.G. Woods, D.H. Rowitch, F.L. Raymond, N.B. Disease, N.G.C. Project, Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children, *Intensive Care Med.* 45 (2019) 627–636.
- [10] D.N.C. and G.P.P. George Potamias, Kleanthi Lakiotaki, Theodora Katsila, Ming Ta Michael Lee, Stavros Topouzis, Deciphering next-generation pharmacogenomics: an information technology perspective, *R. Soc.* (2014).
- [11] J.Y. Hehir-Kwa, R. Pfundt, J.A. Veltman, Exome sequencing and whole genome sequencing for the detection of copy number variation, *Expert Rev. Mol. Diagn.* 15 (2015) 1023–1032.
- [12] W. Huber, S. Davis, F. Pepin, M. Smith, Package ‘biomaRt’, (2021).
- [13] A.C. Jelin, N. Vora, Whole Exome Sequencing: Applications in Prenatal Genetics, *Obstet. Gynecol. Clin. North Am.* 45 (2018) 69–81.
- [14] P. John R. Horn, PharmD, FCCP, Philip D. Hansten, Get to Know an Enzyme: CYP2D6, *Pharm. Timew.* (2008).

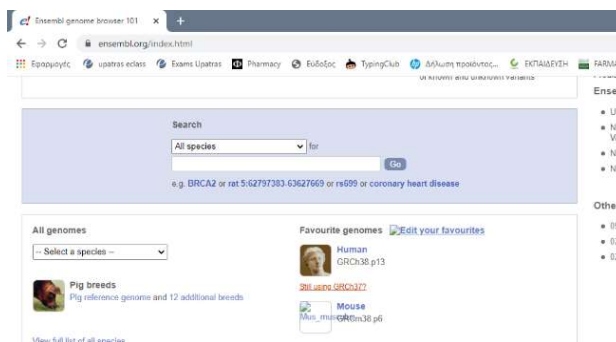
- [15] K.H. Joseph P Kitzmiller, Philip F. Binkley, Saurabh R. Pandey, Adam M. Suhy, Damiano Baldassarre, *Statin Pharmacogenomics: Pursuing Biomarkers for Predicting Clinical Outcomes*, DiscovMed. (2013).
- [16] T. Katsila, G.P. Patrinos, Whole genome sequencing in pharmacogenomics, *Front. Pharmacol.* 6 (2015) 61.
- [17] X. Liu, H. Xu, H. Xu, Q. Geng, W.-H. Mak, F. Ling, Z. Su, F. Yang, T. Zhang, J. Chen, H. Yang, J. Wang, X. Zhang, X. Xu, H. Jia, Z. Zhang, X. Liu, S. Zhong, New genetic variants associated with major adverse cardiovascular events in patients with acute coronary syndromes and treated with clopidogrel and aspirin, *Pharmacogenomics J.* (2021).
- [18] D.J. McCarthy, P. Humburg, A. Kanapin, M.A. Rivas, K. Gaulton, J.-B. Cazier, P. Donnelly, T.W. Consortium, Choice of transcripts and software has a large effect on variant annotation, *Genome Med.* 6 (2014) 26.
- [19] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor, *Genome Biol.* 17 (2016) 1–14.
- [20] C. Mizzi, E. Dalabira, J. Kumuthini, N. Dzimiri, I. Balogh, N. Başak, R. Böhm, J. Borg, P. Borgiani, N. Bozina, H. Bruckmueller, B. Burzynska, A. Carracedo, I. Cascorbi, C. Deltas, V. Dolzan, A. Fenech, G. Grech, V. Kasiulevicius, L. Kádaši, V. Kučinskis, E. Khusnutdinova, Y.L. Loukas, M. Macek, H. Makukh, R. Mathijssen, K. Mitropoulos, C. Mitropoulou, G. Novelli, I. Papantoni, S. Pavlovic, G. Saglio, J. Setric, M. Stojiljkovic, A.P. Stubbs, A. Squassina, M. Torres, M. Turnovec, R.H. van Schaik, K. Voskarides, S.M. Wakil, A. Werk, M. del Zompo, B. Zukic, T. Katsila, M.T.M. Lee, A. Motsinger-Rief, H.L. McLeod, P.J. van der Spek, G.P. Patrinos, A European Spectrum of Pharmacogenomic Biomarkers: Implications for Clinical Pharmacogenomics, *PLoS One.* 11 (2016) e0162866.
- [21] W.H. Organization, I.P. on C. Safety, Biomarkers and risk assessment : concepts and principles / published under the joint sponsorship of the United Nations environment Programme, the International Labour Organisation, and the World Health Organization, (1993).
- [22] K. Pruitt, G. Brown, T. Tatusova, D. Maglott, *The NCBI Handbook. The Reference Sequence (RefSeq) Database*, <https://www.ncbi.nlm.nih.gov/books/NBK21101/>. (2012).
- [23] B. Rabbani, H. Nakaoka, S. Akhondzadeh, M. Tekin, N. Mahdieh, Next generation sequencing: implications in personalized medicine and pharmacogenomics, *Mol. Biosyst.* 12 (2016) 1818–1830.
- [24] M. Simmaco, M. Borro, S. Missori, P. Martelletti, Pharmacogenomics in migraine: catching biomarkers for a predictable disease control, *Expert Rev. Neurother.* 9 (2009) 1267–1269.
- [25] S. Uchida, 2 - Databases and software to make your research life easier, in: S.B.T.-A.N.G. Uchida (Ed.), *Woodhead Publ. Ser. Biomed.*, Woodhead Publishing, 2012: pp. 7–47.
- [26] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, *Nucleic Acids Res.* 38 (2010) e164–e164.
- [27] Z.W. Yang Lin, Lu Yulan, Wang Huijun, No Title, *Pediatr. Precis. Med.* (2016).
- [28] Π.Χ. Μ. Μαρσέλος, Αικ.Αντωνίου, Μ.Κωνσταντή, Γ. Λεονταρίτης,Ευ. Μανωλόπουλος,Π.Παππάς, Βιοχημικηφαρμακολογια Μηχανισμοί Δράσης Των

- Φαρμάκων, (2015) 278.
- [29] Pharmacogenomics and Clinical Biomarkers in Drug Discovery and Development, *Pathol. Patterns Rev.* 124 (n.d.) S29–S41.
  - [30] dbNSFP, (n.d.).
  - [31] BLAST, (n.d.).
  - [32] BLAT, (n.d.).
  - [33] encode, (n.d.).
  - [34] GENCODE, (n.d.).
  - [35] e!Ensembl, (n.d.).

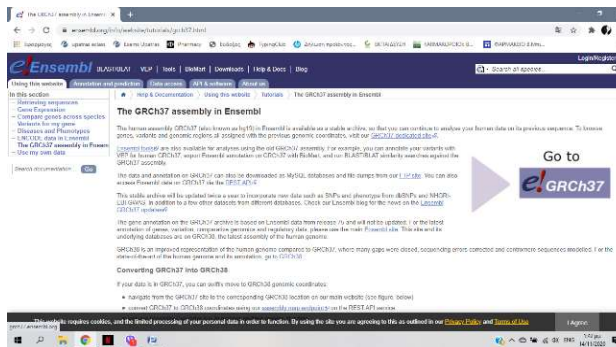
## **6 Κεφάλαιο. Παράρτημα**



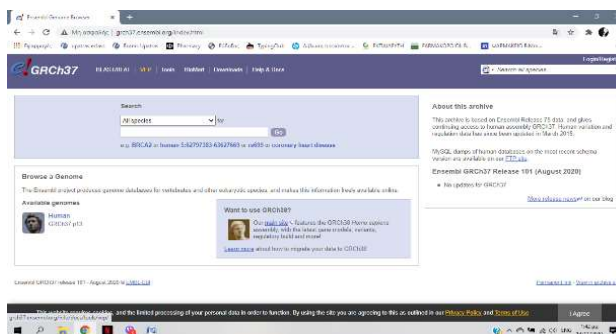
Εικόνα 1. Εισαγωγή των νcf αρχείων των 56 ασθενών στην βάση δεδομένων.



Εικόνα 2. Επιλογή κατάλληλης έκδοσης- Εν προκειμένη GRCh37.

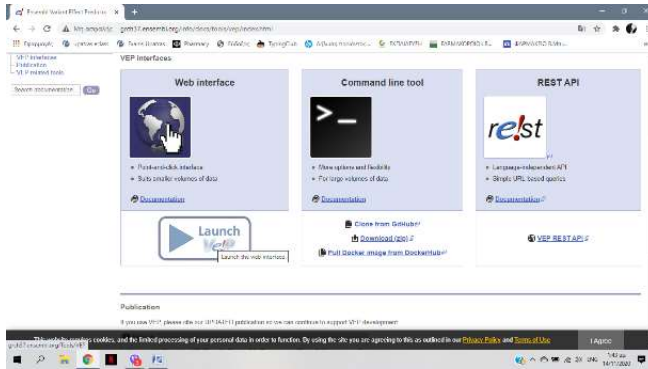


Εικόνα 3. Μεταφορά στην σελίδα της GRCh37.

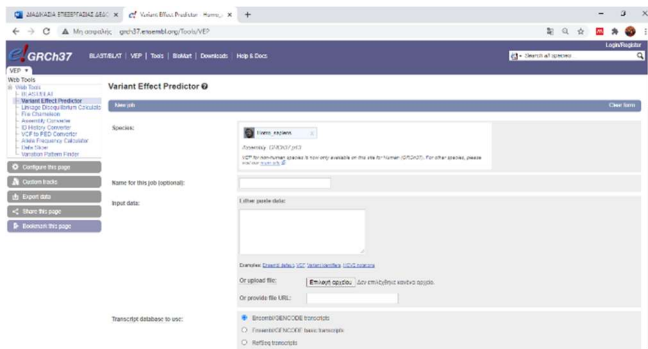


Εικόνα 4. Εκκίνηση της VEP.

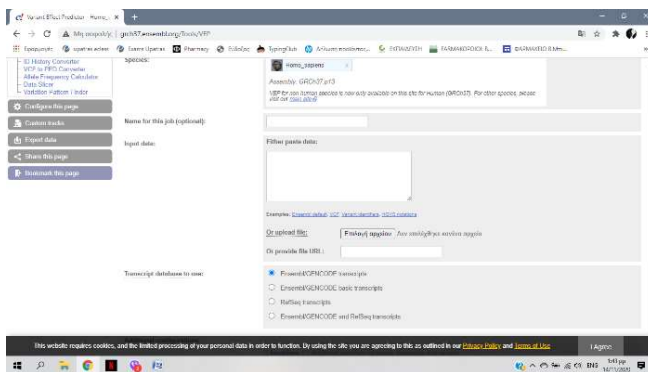




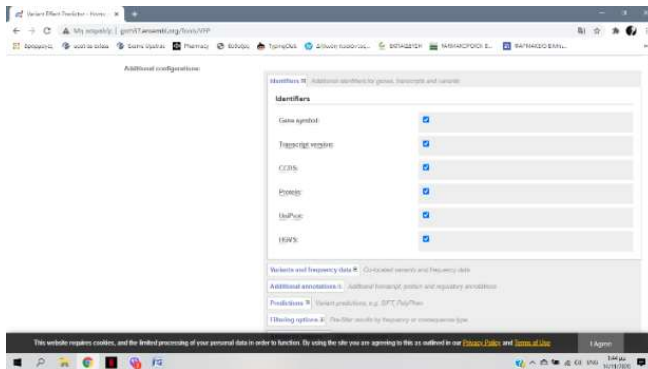
Εικόνα 5. Εκκίνηση της VEP.



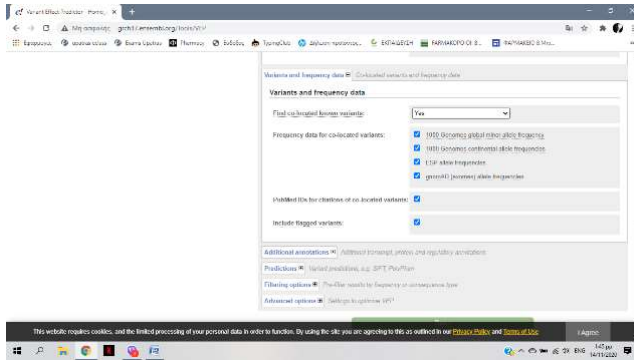
Εικόνα 6. Επιλογή των απαραίτητων συνθηκών για την εξαγωγή αποτελεσμάτων.



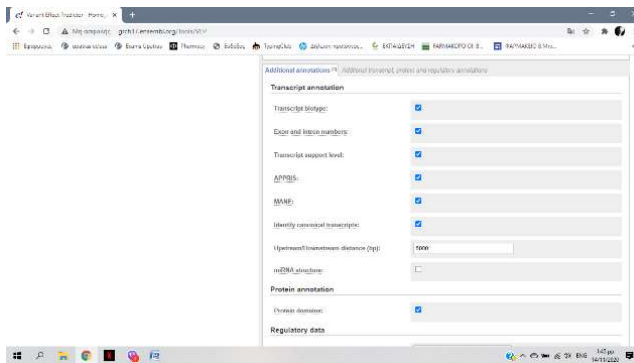
Εικόνα 7. Επιλογή των «Ensembl/GENCODE transcripts»



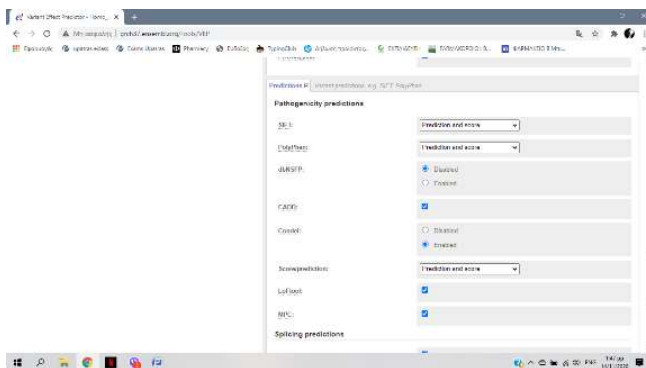
Εικόνα 8. Επιλογή χαρακτηριστικών από το μενού Additional configurations.



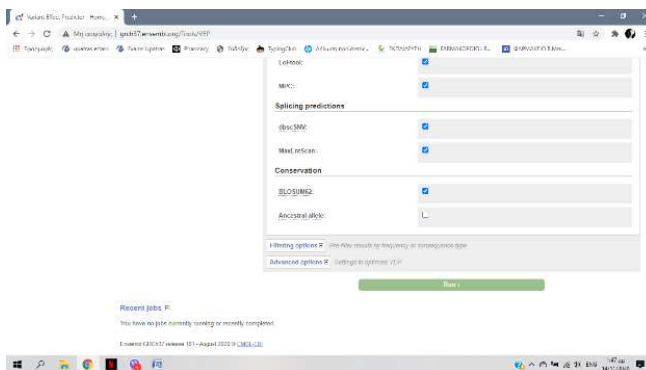
Εικόνα 9. Επιλογή χαρακτηριστικών από Variants and frequency data.



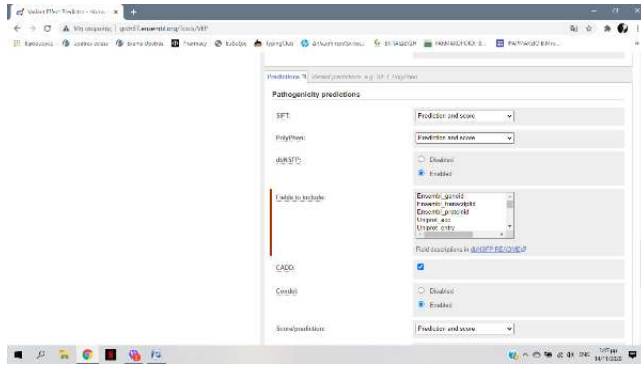
Εικόνα 10. Επιλογή χαρακτηριστικών από Additional annotation.



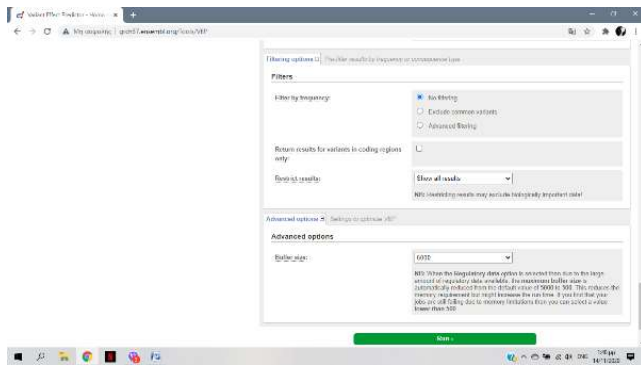
Εικόνα 11. Επιλογές χαρακτηριστικών από μενού Predictions.



Εικόνα 12. Επιλογές χαρακτηριστικών από μενού Predictions.



Εικόνα 13. Επιλογή *in silico* scores.



Εικόνα 14. Εκκίνηση του προγράμματος.