



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Μητρώο Ασθενών με Δερματικό Μελάνωμα

Παζινού Μαρία

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνοι
Καρανίκας Χαράλαμπος
Λέκτορας

Παπαδόδημα Όλγα
Research Scientist στο Εθνικό Ίδρυμα Ερευνών

Λαμία, 2021



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

Μητρώο Ασθενών με Δερματικό Μελάνωμα

Παξινού Μαρία

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Επιβλέποντες
Καρανίκας Χαράλαμπος
Λέκτορας**

**Παπαδόδημα Όλγα
Research Scientist στο Εθνικό Ίδρυμα Ερευνών**

Λαμία, έτος 2021

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, 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 παρουσιάσή 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: 29/09/2021

Ο – Η Δηλ.

Παξινού Μαρία

(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.

Μητρώο Ασθενών με Δερματικό Μελάνωμα

Παζινού Μαρία

Τριμελής Επιτροπή:

Καρανίκας Χαράλαμπος, Λέκτορας (επιβλέπων)

Κακαρούνας Αθανάσιος, Αναπληρωτής Καθηγητής

Τασουλής Σωτήριος, Επίκουρος Καθηγητής

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κύριο Καρανίκα Χαράλαμπο και την κυρία Παπαδόδημα Όλγα για την καθοδήγηση, την εμπιστοσύνη που μου έδειξαν και τον χρόνο που διέθεσαν ώστε να φέρω σε πέρας αυτήν την εργασία. Επίσης ευχαριστώ πολύ την οικογένειά μου που με στηρίζει όλα αυτά τα χρόνια και πιστεύει σε εμένα.

ΠΕΡΙΛΗΨΗ

Το μελάνωμα είναι μια μορφή κακοήθους όγκου που εμφανίζεται κυρίως στο δέρμα μιας και προκαλείται από τα μελανοκύτταρα. Στόχος της παρούσας εργασίας είναι η δημιουργία ενός ενισχυμένου μητρώου ασθενών με μελάνωμα το οποίο θα περιέχει κλινικά αλλά και μοριακά δεδομένα με σκοπό την απόκτηση περαιτέρω γνώσεων για την νόσο. Για την πραγμάτωση έγινε βιβλιογραφική έρευνα στα ήδη υπάρχοντα μητρώα ασθενών με μελάνωμα. Έπειτα έγινε γονιδιακή ανάλυση σε δεδομένα γονιδιακής έκφρασης από τα οποία εξήχθησαν λίστες γονιδίων με διαφορά στα επίπεδα έκφρασής τους ανάμεσα σε υγιείς και νοσούντες ιστούς, μέσω του λογισμικού Bioconductor χρησιμοποιώντας το RStudio. Αυτές οι λίστες γονιδίων χρησιμοποιήθηκαν τέλος στην υλοποίηση του ενισχυμένου μητρώου ασθενών για μελάνωμα με τη βοήθεια της `mysql`.

ABSTRACT

Melanoma is a form of malignant tumor that occurs mainly on the skin due to the melanocytes. The aim of this work is to create an enhanced melanoma registry which will contain clinical and molecular data in order to gain further knowledge around the disease. A bibliographic research was carried out about the already existing melanoma registries. Afterward a genomic analysis on gene expression was performed, from which gene lists were exported using the Bioconductor software in the RStudio environment. These gene lists, whose expression is different between healthy and diseased tissues, were used in the implementation of melanoma registry using MySQL.

Πίνακας Περιεχομένων

Περιεχόμενα

Κεφάλαιο 1 ^ο : Εισαγωγή.....	11
1.1 Γενικά για το Μελάνωμα.....	11
1.2 Είδη Μελανώματος.....	13
1.3 Διάγνωση και Θεραπεία.....	14
1.4 Ρόλος Πληροφορικής Υγείας.....	15
1.5 Ρόλος Βιοπληροφορικής.....	16
Κεφάλαιο 2 ^ο : Μητρώα ασθενών με μελάνωμα.....	18
2.1 Υπάρχοντα μητρώα ασθενών.....	18
2.1.1 Melanoma UK.....	18
2.1.2 War on Melanoma.....	19
2.1.3 CURE OM Patient Registry.....	20
2.1.4 Yale Melanoma Registry.....	21
2.1.5 Dutch Melanoma Treatment Registry.....	22
2.1.6 EUMelaReg.....	23
2.1.7 The Spanish National Cutaneous Melanoma Registry.....	24
2.1.8 Danish Melanoma Database.....	25
2.1.9 Swedish Melanoma Registry (SweMR).....	26
2.2 Συγκριτική μελέτη Μητρώων Ασθενών για Μελάνωμα.....	26
Κεφάλαιο 3 ^ο : Δεδομένα Γονιδιώματος - Ανάλυση.....	28
3.1 Μικροσυστοιχίες γονιδίων.....	28
3.2 Υλικό και μέθοδοι.....	29
3.3 Προέλευση Δείγματος.....	31
3.4 Λογισμικό.....	32
3.5 Προεπεξεργασία Δεδομένων.....	33
3.5.1 Διόρθωση Θορύβου υποβάθρου εικόνας (Background Correction).....	35
3.5.2 Κανονικοποίηση (Normalization).....	35
3.5.3 Λογαριθμικός Μετασχηματισμός Δεδομένων (log ₂ transformation).....	36
3.5.4 Αλγόριθμος Robust Multiarray Average (RMA).....	38
3.6 Έλεγχος επεξεργασίας (Quality control).....	40
3.6.1 Διάγραμμα πυκνότητας (Density Plot).....	40
3.6.2 Θηκόγραμμα (Box Plot).....	42

3.6.3 Διάγραμμα σχετικής λογαριθμικής έκφρασης (Relative Log Expression – RLE Plot)	43
3.6.4 MA Plot	44
3.7 Φιλτράρισμα (Filtering panel).....	45
3.8 Στατιστική Ανάλυση	47
3.8.1 False Discovery Rate.....	47
3.8.2 Στατιστικός έλεγχος	48
3.8.3 Limma (Linear Models for Microarray Data)	50
3.9 Αντιστοίχιση των διαφορικά εκφρασμένων μεταγράφων σε γονιδιακά προϊόντα.....	50
3.10 Αποτελέσματα	51
3.10.1 Volcano Plots για την απεικόνιση των αποτελεσμάτων.....	52
3.10.2 Πίνακες Γονιδίων	53
Κεφάλαιο 4 ^ο : Υλοποίηση Μητρώου Ασθενών για Μελάνωμα	54
4.1 Γενικά οι Βάσεις Δεδομένων	54
4.2 Εννοιολογικό μοντέλο Βάσεων Δεδομένων	56
4.2.1 Το μοντέλο Οντοτήτων-Συσχετίσεων	56
4.2.2 ER Diagram Μητρώου Ασθενών με Μελάνωμα.....	58
4.3 Data Dictionary.....	61
4.4 Υλοποίηση με MySQL	75
Βιβλιογραφία	80
Παράρτημα	82

Κεφάλαιο 1^ο : Εισαγωγή

1.1 Γενικά για το Μελάνωμα

Το μελάνωμα ορίζεται ως μια μορφή κακοήθους όγκου, όπου εμφανίζεται κυρίως στο δέρμα, μιας και προκαλείται από τα μελανοκύτταρα. Σε σπάνιες περιπτώσεις το μελάνωμα μπορεί να εμφανιστεί στον οφθαλμό (στον επιπεφυκότα και στην ίριδα), στις μήνιγγες καθώς και σε άλλους βλεννογόνους, στους οποίους υπάρχουν μελανοκύτταρα. Στην πράξη το δέρμα αποτελείται από 3 στοιβάδες, την επιδερμίδα που είναι η εξωτερική στοιβάδα η οποία δηλαδή έρχεται σε επαφή με το εξωτερικό περιβάλλον, τη δερμίδα που βρίσκεται στη μέση και την υποδερμίδα η οποία αποτελείται κυρίως από λιπώδη ιστό. Η επιδερμίδα από μόνη της χωρίζεται και αυτή σε κάποιες στοιβάδες με την πιο βαθιά από αυτές να ονομάζεται βασική στοιβάδα και στην οποία βρίσκονται τα μελανοκύτταρα. Στην φυσιολογική μορφή τους τα μελανοκύτταρα ευθύνονται για τον σχηματισμό των σπύλων για αυτό και παράγουν μια χρωστική ουσία τη μελανίνη που είναι υπεύθυνη για το χρώμα στο δέρμα, στα μαλλιά και στα μάτια. Ρόλος της μελανίνης είναι να προστατεύει τα κύτταρα του δέρματος από την επίδραση της ηλιακής ακτινοβολίας. Το μελάνωμα διακρίνεται σε δύο φάσεις ανάπτυξης, την κάθετη (vertical growth phase), την οριζόντια ή ακτινωτή (radial growth phase) ή κάθετη και οριζόντια / ακτινωτή. Η κάθετη φάση ανάπτυξης είναι πάντα διηθητική όπως συμβαίνει στο οζώδες μελάνωμα και μπορεί να αναπτυχθεί σε οριζόντια φάση ανάπτυξης προϋπάρχοντος μελανώματος άλλου τύπου. Η βιολογική συμπεριφορά αυτής της φάσης είναι επιθετικότερη και περιλαμβάνει μεγαλύτερου μεγέθους κύτταρα ή αθροίσεις σε σύγκριση με το ενδοεπιδερμικό (in situ) ή και διηθητικό μελάνωμα όταν αυτό βρίσκεται σε οριζόντια φάση ανάπτυξης. Συνεπώς η οριζόντια φάση ανάπτυξης ή αλλιώς ακτινωτή φάση μπορεί να είναι ενδοεπιδερμική (in situ) ή διηθητική με το μέγεθος των κυττάρων και των αθροίσεων να είναι παρόμοιο σε αυτές τις δύο περιπτώσεις. Γνωρίζουμε ότι το μελάνωμα είναι ένας σπάνιος τύπος καρκίνου που όμως είναι ο πιο επικίνδυνος τύπος καρκίνου του δέρματος διότι υπάρχει αυξημένη πιθανότητα μετάστασης. Για να γίνει μια μετάσταση χρησιμοποιείται το λεμφικό σύστημα το οποίο αποτελείται από λεπτά αγγεία, όπως τα αιμοφόρα αγγεία που σχηματίζουν ένα δίκτυο σε όλους τους ιστούς του σώματος. Μέσω του υγρού που περιέχουν τα λεμφικά αγγεία (λέμφος) είναι δυνατό να αποσπαστούν καρκινικά κύτταρα από τον πρωτοπαθή όγκο και να

οδηγηθούν στη συνέχεια στους λεμφαδένες. Λεμφαδένες συναντώνται κυρίως στις μασχάλες, στο λαιμό, τους βουβώνες και τη κοιλιά. Επιπλέον μεταστάσεις είναι δυνατό να συμβούν και μέσω του αίματος, η λεγόμενη αιματογενής διασπορά. Γενικά το μελάνωμα αποτελεί πλέον το 3% του συνόλου των καρκίνων και είναι ένα από τα πιο γρήγορα εξαπλούμενα κακοήθη νοσήματα λόγω της αυξανόμενης συχνότητάς του παγκοσμίως. Πρόσφατες στατιστικές μελέτες έχουν δείξει πως το μελάνωμα είναι ο έκτος συχνότερος τύπος καρκίνου στους άνδρες και ο έβδομος στις γυναίκες.

Σύμφωνα με μελέτες για την κατανομή και την εξέλιξη της νόσου γνωρίζουμε πλέον πως το μελάνωμα ποικίλει ανάλογα με τη φυλή, τη γεωγραφική περιοχή, το φύλο, την ηλικία, τον ιστολογικό τύπο και την ανατομική εντόπιση της νόσου. Με βάση αυτά τα στοιχεία είμαστε σε θέση να γνωρίζουμε πλέον πως η νόσος τα τελευταία δέκα χρόνια ευδοκμεί στην Ευρώπη όπου η επίπτωση εμφανίζεται μεγαλύτερη στον Βορρά από ότι στον Νότο εν αντιθέσει της υπερϊώδους ακτινοβολίας, ίσως λόγω των φαινοτυπικών χαρακτηριστικών του πληθυσμού, τη Βόρεια Αμερική και την Αυστραλία, ενώ παράλληλα σπανίζει στους Ασιατικούς λαούς και σε άτομα της μαύρης φυλής, στους τελευταίους μάλιστα συνηθίζει να κάνει την εμφάνισή της στα άκρα. Συγκεκριμένα για την Ελλάδα η επίπτωση θεωρείται ότι είναι 2,2 ανά 100000. Όσον αφορά την ηλικία φαίνεται πως η νόσος δεν κάνει «διακρίσεις» αν και εμφανίζεται σπανιότερα στα παιδιά και συνεπώς η μέση ηλικία διάγνωσης είναι τα 55 έτη. Σημαντικό είναι να αναφέρουμε πως τα παραπάνω χαρακτηριστικά ενδέχεται να διαφέρουν στην περίπτωση που συνδυάζονται με κάποιο οικογενειακό ιστορικό μελανώματος. Η κυριότερη περιβαλλοντική αιτία της νόσου είναι η υπερϊώδης ακτινοβολία, είτε φυσική είτε τεχνητή (με τεχνητές πηγές μαυρίσματος) και μπορεί να αποφευχθεί με την τακτική χρήση αντιηλιακών. Αναφερόμενοι στην θνησιμότητα λόγω του μελανώματος θα μπορούσαμε να πούμε πως δεν είναι αντίστοιχη με την επίπτωσή του, πιθανότατα εξ' αιτίας της πρώιμης διάγνωσης λόγω της καλύτερης και επαρκέστερης ενημέρωσης του κοινού από τους οργανισμούς υγείας ή λόγω της καλύτερης παροχής ιατρικών υπηρεσιών που υποστηρίζουν κι αυτές την έγκαιρη και ακριβέστερη διάγνωση. Μεγαλύτερο κίνδυνο για την εκδήλωση μελανώματος έχουν άτομα με ανοιχτόχρωμο δέρμα δηλαδή εκείνοι που «καίγονται» εύκολα και «μαυρίζουν» δύσκολα ή με ιστορικό υπέρμετρης έκθεσης στον ήλιο ή σε τεχνητές πηγές «μαυρίσματος» οι οποίοι μάλιστα παρουσιάζουν 60% μεγαλύτερη πιθανότητα εμφάνισης της νόσου σε σύγκριση με άτομα που δεν έχουν εκτεθεί στον ήλιο. Επιπλέον άτομα που φέρουν πολλαπλούς κοινούς ή άτυπους μελανοκυτταρικούς

σπίλους θεωρείται επικίνδυνο να νοσήσουν, όπως και άτομα με οικογενειακό ιστορικό μελανώματος στα οποία η πιθανότητα εμφάνισης γίνεται διπλάσια. Επίσης αυξάνεται η πιθανότητα εμφάνισης σε ασθενείς που λαμβάνουν ανοσοκατασταλτική θεραπεία έπειτα από μεταμόσχευση οργάνου, όπως και σε ασθενείς που έχουν ιστορικό κακοήθους νεοπλασματος .

1.2 Είδη Μελανώματος

Μπορούμε να διακρίνουμε τέσσερις κλινικούς τύπους μελανώματος οι οποίοι διαφέρουν ως προς την επιδημιολογία τους, τα ιστολογικά τους χαρακτηριστικά, το μοριακό τους υπόβαθρο και τον συσχετισμό τους με την έκθεση στην ηλιακή ακτινοβολία. Σύμφωνα με την συχνότητα εμφάνισης τους έχουμε κατά σειρά:

Το Επιφανειακώς Επεκτεινόμενο Μελάνωμα, το οποίο αποτελεί το 70% του συνόλου των μελανωμάτων και συναντάται ευκολότερα στη λευκή φυλή με τη διάγνωση του να είναι συνηθέστερη σε άτομα ηλικίας 30-50 ετών. Κατά βάση εντοπίζεται στον κορμό στους άνδρες και στις κνήμες στις γυναίκες. Μορφολογικά θα μπορούσαμε να πούμε ότι χαρακτηρίζεται από παζετοειδή διασπορά κακοήθων μελανοκυττάρων σε όλο το πάχος της επιδερμίδας, γεγονός που το κατατάσσει στην οριζόντια φάση ανάπτυξης του όγκου.

Το Οζώδες Μελάνωμα που αντιστοιχεί στο 10-15% των μελανωμάτων, διαγιγνώσκεται κατά κύριο λόγο σε άτομα ηλικίας 50-60 ετών και μάλιστα κυρίως σε άνδρες. Εντοπίζεται συχνότερα στον κορμό, την κεφαλή και τον τράχηλο. Όσον αφορά την μορφολογία του, αυτή ποικίλει ανάμεσα σε καστανό, κυανό ή μελανό οζίδιο με σχετική συμμετρία, ομοιογένεια και μονοχρωμία που όμως μεγαλώνει γρήγορα και συχνά εξελκώνεται και αιμορραγεί. Ωστόσο υπάρχει ένα 5% αυτών των μελανωμάτων που είναι αμελανωτικά και εμφανίζονται με ερυθρό ή ρόδινο χρώμα. Επομένως τα παραπάνω καθιστούν αυτήν την κατηγορία μελανώματος εξαιρετικά επιθετική μιας και η διάγνωση γίνεται σε πιο προχωρημένο στάδιο και συνεπώς έχει χειρότερη πρόγνωση. Για αυτό και αντιστοιχεί στην κάθετη φάση ανάπτυξης του όγκου, έπειτα από σύντομη ή απύσα οριζόντια φάση.

Το Μελάνωμα Αναπτυσσόμενο Επί Κακοήθους Φακίδας το οποίο ποσοστιαία αντιστοιχεί στο 15% των μελανωμάτων. Αναπτύσσεται αργά μέσα στα χρόνια πάνω σε κάποια κακοήθη φακίδα σε περιοχές που εκτίθενται στον ήλιο όπως στην κεφαλή,

κυρίως στην ρίνα και στις παρειές και τον τράχηλο και έτσι η ηλικία διάγνωσης είναι συνήθως τα 60-70 χρόνια. Μορφολογικά είναι σαν ευμεγέθους καστανόμαυρη κηλίδα με υπόχρωμες περιοχές και ακανόνιστη παρυφή στην οποία αναπτύσσεται οζίδιο.

Περίπου το 5-10% αυτών εξελίσσεται σε διηθητικό μελάνωμα.

Το Μελάνωμα Των Άκρων Τύπου Εφηλίδας φαίνεται να είναι σπάνιο στη λευκή φυλή με ποσοστό μόλις 5-10% αλλά αποτελεί τη συνηθέστερη μορφή μελανώματος στη μαύρη φυλή και την κίτρινη με ποσοστά 70% και 45% αντίστοιχα. Εμφανίζεται σαν μία ακανόνιστη, ευμεγέθους, καστανόμαυρη (ποικιλόχρωμη) κηλίδα κατά την οριζόντια φάση ανάπτυξης του όγκου, πάνω στην οποία όταν επέλθει η κάθετη φάση ανάπτυξης μπορεί να εμφανιστεί οζίδιο. Συναντάται κυρίως στις παλάμες και στα πέλματα ή στην ονυχία μονάδα και συνεπώς λόγω της θέσης του υπάρχει δυσκολία στην αναγνώρισή του και η διάγνωση δεν είναι έγκαιρη.

Σπανιότεροι τύποι μελανώματος είναι το αμελανωτικό, το δεσμοπλαστικό, το σπιτσοειδές, το μικροκυτταρικό μελάνωμα, το εκ διαυγών κυττάρων μελάνωμα, ο κακοήθους κυανός σπίλος, το ζωικού τύπου μελάνωμα και το μελάνωμα των βλεννογόνων.

1.3 Διάγνωση και Θεραπεία

Η διάγνωση επέρχεται έπειτα από την κλινική εξέταση του ασθενούς και την επισκόπηση της δερματικής βλάβης. Συνεπώς χρειάζεται η λήψη του ιατρικού ιστορικού του ασθενούς, του οικογενειακού ιστορικού όσον αφορά το μελάνωμα αλλά και του ιστορικού εξέλιξης της βλάβης έτσι ώστε οι γιατροί να είναι σε θέση να γνωρίζουν το χρονικό διάστημα από την παρατήρηση της βλάβης όπως και το εάν παρουσιάζει κνησμό, πόνο ή αιμορραγία. Στη συνέχεια ο γιατρός θα εξετάζει κλινικά την οποιαδήποτε μελαγχρωματική βλάβη έτσι ώστε να αποτυπώσει το χρώμα, το σχήμα, την παρουσία εξέλκωσης, αιμορραγίας, εφελκιδοποίησης ή επιφανειακού οζιδίου. Ένας τρόπος για να γίνει εύκολα η εξέταση είναι ο κανόνας ABCDE, ο οποίος είναι ένα ακρωνύμιο όπου το «Α» αναφέρεται στην Ασυμμετρία, το «Β» στα όρια (Border), το «C» στο χρώμα (Color), το «D» στη διάμετρο (Diametre) και το «E» στην εξέλιξη (Evolution). Μία άλλη διαγνωστική in vivo μέθοδος είναι η δερματοσκόπηση κατά την οποία το δέρμα εξετάζεται υπό ειδικό φωτισμό (πολωμένο ή μη πολωμένο φως) και μεγέθυνση για να γίνουν ορατές δομές και χρώματα που δεν μπορεί να τα δει το γυμνό μάτι, έτσι ώστε η διαγνωστική ακρίβεια στη διάγνωση του

μελανώματος να αυξηθεί. Ειδικότερα όταν γίνεται από εξιδεικευμένους και έμπειρους γιατρούς η διαγνωστική ακρίβεια αυξάνεται κατά 30%. Πολύ σημαντικό είναι επίσης πως με τη χρήση αυτής της τεχνικής μειώνεται σε μεγάλο ποσοστό ο αριθμός των περιττών επεμβάσεων για την αφαίρεση σπίλων. Η δερματοσκόπηση αφορά όλες τις βλάβες και όχι μόνο τις αναφερόμενες ως «ύποπτες» για αυτό και η εξέταση είναι πάντα ολόσωμη και σε αυτήν συνεκτιμώνται επιδημιολογικά στοιχεία όπως ηλικία, φύλο, φυλή, ιστορικό μελανώματος και κλινικά δεδομένα που αφορούν τα χαρακτηριστικά της βλάβης (εντόπιση και εξέλιξη βλάβης). Μία άλλη τεχνική είναι η ψηφιακή δερματοσκόπηση γνωστή και σαν χαρτογράφηση σπίλων η οποία έχει σκοπό την καταμέτρηση και την καταγραφή των σπίλων έτσι ώστε να παρακολουθούνται τα άτομα με υψηλό κίνδυνο για μελάνωμα. Αυτή κυρίως αφορά τα άτομα που συγκαταλέγονται στις ομάδες υψηλού κινδύνου για μελάνωμα, δηλαδή αυτών που είτε έχουν πολλούς άτυπους σπίλους, είτε έχουν ατομικό ή οικογενειακό ιστορικό μελανώματος και μεγάλο αριθμό σπίλων, είτε ατόμων με σύνδρομο δυσπλαστικών σπίλων (FAMMM). Επιπλέον προορίζεται και για άτομα που χρειάζονται την παρακολούθηση μιας μεμονωμένης επίπεδης μελαγχρωματικής βλάβης όταν αυτή εμφανίζει άτυπες αλλαγές. Τέλος μια ακόμα εξέταση που κάνει δυνατό τον εντοπισμό και την δερματοσκοπική αξιολόγηση μιας ενδεχόμενης βλάβης είναι η ολόσωμη φωτογράφιση και αφορά κι αυτή τις ομάδες υψηλού κινδύνου για μελάνωμα.

Μία μελαγχρωματική βλάβη εάν θεωρηθεί ύποπτη έπειτα από την εξέταση με κάποια από τις παραπάνω μεθόδους από τον επιβλέποντα γιατρό θα πρέπει να αφαιρεθεί άμεσα με χειρουργική επέμβαση και στη συνέχεια να εξεταστεί ιστολογικά μέσω μίας βιοψίας. Υπάρχουν διάφορες μέθοδοι λήψης βιοψίας οι οποίες εφαρμόζονται κατά περίπτωση και με βάση πληροφορίες που αφορούν χαρακτηριστικά και τη μορφολογία της βλάβης όπως ο χρόνος εμφάνισης και οι μεταβολές ή διάφορες θεραπευτικές παρεμβάσεις που έχουν προηγηθεί και το κλινικό ιστορικό του ασθενή. Εάν το αποτέλεσμα της βιοψίας δείξει κακοήθεια τότε ο ασθενής θα πρέπει να εξεταστεί για τυχόν μεταστάσεις και στην περίπτωση αυτή να του χορηγηθεί κάποια συμπληρωματική θεραπεία.

1.4 Ρόλος Πληροφορικής Υγείας

Η πληροφορική της υγείας είναι ένας διαρκώς αναπτυσσόμενος κλάδος της πληροφορικής μιας και οι ανάγκες για μηχανογράφηση πληθαίνουν όσο πληθαίνουν και τα δεδομένα στον τομέα της υγείας. Συνεπώς έχοντας πλέον τεχνολογική δύναμη και γνώση η επιστήμη της πληροφορικής είναι σε θέση να αποθηκεύει όλα τα δεδομένα και τις πληροφορίες και να τα επεξεργάζεται έτσι ώστε να αντλεί από αυτά τις χρήσιμες γνώσεις και να διευκολύνεται η λήψη αποφάσεων από τους ειδικούς υγείας.

Πιο συγκεκριμένα ένας από τους τρόπους με τον οποίο βοηθά η πληροφορική της υγείας και θα μελετηθεί στην παρούσα εργασία είναι τα μητρώα ασθενών τα οποία αποτελούν ένα οργανωμένο σύστημα που συλλέγει, αναλύει και διαδίδει τα δεδομένα και τις πληροφορίες για μια ομάδα ανθρώπων που ορίζονται από μια συγκεκριμένη νόσο, κατάσταση, έκθεση σε ιατρικό προϊόν ή μια υπηρεσία υγείας και εξυπηρετεί έναν προκαθορισμένο επιστημονικό, κλινικό, ή/και δημόσιας υγείας (πολιτικό) σκοπό. Μερικοί από τους σημαντικότερους σκοπούς για την δημιουργία και τη λειτουργία μητρώου ασθενών είναι: η περιγραφή της φυσικής πορείας μιας νόσου (natural history of disease), ο προσδιορισμός της κλινικής ή/και της κοστολογικής-αποτελεσματικότητας (determine clinical and/or cost-effectiveness), η αξιολόγηση της ασφάλειας ή της βλάβης (assess safety or harm), η εκτίμηση της ποιότητας περίθαλψης, καθώς και για να εξυπηρετήσει την επιτήρηση της δημόσιας υγείας και τον έλεγχο ασθενειών (public health surveillance and disease control).

Κατ' αυτόν τον τρόπο έχουν δημιουργηθεί και μητρώα ασθενών με μελάνωμα μερικά από τα οποία θα εξετάσουμε στο επόμενο κεφάλαιο. Στόχος τους είναι η γενικότερη βελτίωση της φροντίδας των ασθενών με μελάνωμα (η ενημέρωση, η πρόληψη, η έγκαιρη διάγνωση και θεραπεία κ.α.) και ο καλύτερος τρόπος για να επιτευχθεί αυτό είναι ακολουθώντας τους ασθενείς στο χρόνο και τον τόπο, έτσι ώστε να μπορούν να παρατηρηθούν μεσοπρόθεσμα και μακροπρόθεσμα αποτελέσματα. Τέλος τα μητρώα αυτά εξυπηρετούν και σε σκοπούς αξιολόγησης των τεχνολογιών υγείας των φαρμακευτικών και ιατροτεχνολογικών προϊόντων που έχουν όλο και περισσότερο την ανάγκη δεδομένων πραγματικής κλινικής πρακτικής ('real-world' data), προκειμένου να εξετάζεται η κλινική αποτελεσματικότητα (clinical effectiveness) και να πραγματοποιείται οικονομική αξιολόγηση (economic evaluation).

1.5 Ρόλος Βιοπληροφορικής

Βιοπληροφορική είναι ο επιστημονικός χώρος όπου η σύμπραξη της Βιολογίας με την Πληροφορική, την Στατιστική και τα Μαθηματικά εξερευνά νέους τρόπους για την προσέγγιση των βιολογικών προβλημάτων, καθώς και την αντίληψη βασικών αρχών της Βιολογίας. Οι επιστήμονες που ασχολούνται με την βιοπληροφορική έχουν αναπτύξει διάφορα εργαλεία και τεχνολογίες ώστε να γίνονται αναλύσεις στο ανθρώπινο γονιδίωμα και να λαμβάνονται στατιστικά σημαντικά αποτελέσματα για την μελέτη και τη λήψη αποφάσεων που αφορούν τις λειτουργίες των διάφορων γονιδίων. Επιπλέον αυτές οι τεχνολογίες αποτελούν αντικείμενο της μοριακής βιολογίας καθώς αφορούν λειτουργικές μελέτες γονιδιωματικής που έχουν ως πρωταρχικό στόχο την κατανόηση της δυναμικής και σύνθετης σχέσης μεταξύ του γονιδιώματος ενός οργανισμού και του φαινοτύπου του. Έτσι γίνονται μελέτες που διερευνούν μια σειρά βιολογικών διαδικασιών όπως η μεταγραφή και η μετάφραση, με κύριο στόχο την απόκτηση γνώσης και εμπειρίας σχετικά με το πότε και πού εκφράζονται τα γονίδια, πώς διαφέρουν τα επίπεδα γονιδιακής έκφρασης σε διάφορους τύπους κυττάρων και καταστάσεις, ποιοι είναι οι λειτουργικοί ρόλοι διαφορετικών γονιδίων και σε ποιες κυτταρικές διαδικασίες συμμετέχουν, πώς ρυθμίζονται τα γονίδια, πώς μεταβάλλεται η γονιδιακή έκφραση σε διάφορες ασθένειες ή έπειτα από κάποια θεραπεία. Η λειτουργική γονιδιωματική ενσωματώνει πληροφορίες από διάφορες μοριακές μεθοδολογίες για να κατανοήσει πώς η αλληλουχία του DNA μεταφράζεται σε πολύπλοκες πληροφορίες σε ένα κύτταρο. Η τεχνολογία των μικροσυστοιχιών (microarray technology) αποτελεί ένα ισχυρό εργαλείο που χρησιμοποιείται για να προσδιοριστούν τα επίπεδα της έκφρασης των γονιδίων σε βιολογικά δείγματα και παράλληλα να συγκριθεί η γονιδιακή έκφραση σε διάφορες βιολογικές καταστάσεις. Συνεπώς τα τελευταία χρόνια αυτές οι νέες τεχνολογίες που επιτρέπουν τη μαζική αλληλούχιση του ανθρώπινου γονιδιώματος και την ανάλυσή του, τόσο στο επίπεδο της γονιδιακής έκφρασης όσο και στο πεδίο των μεταλλάξεων, έχουν συμβάλει αδιαμφισβήτητα στην έρευνα και την αντιμετώπιση ασθενειών όπως ο καρκίνος. Παράλληλα όμως αυτές οι τεχνολογίες έχουν αυξήσει τον όγκο και την πολυπλοκότητα των δεδομένων που παράγουν, γεγονός που έχει δημιουργήσει την ανάγκη για την παραγωγή λογισμικού και εξειδικευμένων εργαλείων που βοηθούν στην βιοπληροφορική ανάλυση. Στα πλαίσια αυτής της εργασίας θα ασχοληθούμε με την ανάλυση γονιδιωματικών δεδομένων μικροσυστοιχιών στον καρκίνο και συγκεκριμένα στο μελάνωμα. Στόχος είναι η ενδεχόμενη εύρεση στατιστικά σημαντικών γονιδίων με διαφορά στα επίπεδα

έκφρασής τους ανάμεσα σε υγιείς και νοσούντες ιστούς. Τα δεδομένα που χρησιμοποιήθηκαν αντλήθηκαν από την GEO (Gene Expression Omnibus), η οποία είναι μια βάση γονιδιωματικών δεδομένων που υποστηρίζεται από την NCBI (National Center for Biotechnology Information) και αναλύθηκαν στο RStudio με γλώσσα προγραμματισμού την R και εργαλείο το Bioconductor.

Κεφάλαιο 2^ο : Μητρώα ασθενών με μελάνωμα

2.1 Υπάρχοντα μητρώα ασθενών

Τα μητρώα ασθενών με μελάνωμα βοηθούν στην καταγραφή, τη συλλογή και την επεξεργασία των δεδομένων των ασθενών με σκοπό να δημιουργηθεί μια πλατφόρμα για έρευνα έτσι ώστε να αποκτηθούν περισσότερες γνώσεις για την νόσο. Τα μητρώα διαφοροποιούνται μεταξύ τους ως προς την εμβέλειά τους (πχ εθνικά μητρώα, ευρωπαϊκά κα), τους ασθενείς στους οποίους αναφέρονται (πχ ανάλογα το στάδιο της νόσου ή τον τύπο κα) και τον σκοπό για τον οποίο αναπτύχθηκαν (πχ περιγραφή της πορείας της νόσου, επιδημιολογικοί σκοποί, κλινικοί, εκτίμηση της ποιότητας περίθαλψης κα). Παρακάτω δίνονται ενδεικτικά κάποια υπάρχοντα μητρώα και γίνεται ανάλυση του τρόπου λειτουργίας του καθενός και του σκοπού του.

2.1.1 Melanoma UK

Το Melanoma UK είναι ένα μητρώο ασθενών για μελάνωμα το οποίο είναι σε λειτουργία από το 2017 μέσω μιας δωρεάν εφαρμογής (Melanoma UK app) για Android και iPhone και απευθύνεται σε ανθρώπους που ζουν με μελάνωμα σε οποιοδήποτε στάδιο. Η εφαρμογή αναπτύχθηκε από το Melanoma UK και το Vitaccess σε συνεργασία με το Royal Marsden NHS Foundation και χρηματοδοτείται από το Vitaccess με πρωταρχικό σκοπό την βελτίωση της φροντίδας των ασθενών. Τα δεδομένα που συλλέγονται τα εισάγουν οι ίδιοι οι χρήστες της εφαρμογής μέσω του κινητού τους τηλεφώνου, συμπληρώνοντας σχετικά ερωτηματολόγια και έρευνες από την άνεση του σπιτιού τους. Στη συνέχεια αυτά τα δεδομένα αποθηκεύονται στην βάση δεδομένων και όντας ανώνυμα είναι διαθέσιμα οποιαδήποτε στιγμή στους χρήστες της πλατφόρμας για να τα δουν ή να τα χρησιμοποιήσουν για κάποια δευτερογενή ανάλυση. Για αυτό και πρόσβαση σε αυτά έχουν ακαδημαϊκοί ερευνητές

σε νοσοκομεία ή πανεπιστήμια και φαρμακευτικές εταιρείες που αναπτύσσουν φάρμακα και θεραπείες για το μελάνωμα. Τα δεδομένα είναι δημογραφικά και επιδημιολογικά, επίσης αφορούν την ποιότητα ζωής του ασθενή και τα πρότυπα θεραπείας που λαμβάνει. Οι χρήστες είναι σε θέση να καταγράψουν την κατάσταση της υγείας τους και να συνδεθούν με άλλους ασθενείς γρήγορα και εύκολα μέσω των παρακάτω πεδίων που θα βρουν στην εφαρμογή. Στο πεδίο symptom tracker μπορούν να αποθηκεύσουν το πώς νιώθουν και τα συμπτώματά τους. Στο πεδίο Knowledge Feature, το οποίο ανανεώνεται τακτικά έχει δημιουργηθεί από τις γνώσεις ειδικών υγείας, και σε αυτό οι χρήστες μπορούν να βρουν γενικές πληροφορίες για το μελάνωμα, τις διαγνώσεις και τα στάδια, διάφορα χρήσιμα στατιστικά, πιθανές αιτίες και παράγοντες της νόσου κ.α. Το πεδίο Community επιτρέπει στους ασθενείς να βρουν ο ένας τον άλλον και να έρθουν σε επαφή μεταξύ τους λειτουργώντας με τρόπο τέτοιοι ώστε να αντιστοιχίζονται ασθενείς με παρόμοια διάγνωση ή ασθενείς που ζουν σε κοντινές περιοχές. Τέλος το πεδίο Medical Profile όπου ο κάθε χρήστης προσθέτει και ανανεώνει τις προσωπικές του πληροφορίες για διαγνώσεις, στάδια, θεραπείες, νοσηλείες ή χειρουργεία και τον καθημερινό του τρόπο ζωής. Στόχος είναι η καλύτερη κατανόηση του πώς αισθάνονται οι ασθενείς κατά την διάρκεια που λαμβάνουν κάποια θεραπεία αλλά και έπειτα από αυτήν. Προς το παρόν το μοντέλο της εφαρμογής είναι πιλοτικό και εφαρμόζεται μόνο στο Ηνωμένο Βασίλειο αλλά το σχέδιο είναι να επεκταθεί στις Ηνωμένες Πολιτείες, στην Ευρώπη και την Αυστραλία. Τα πλεονεκτήματα της εφαρμογής μεταξύ άλλων περιλαμβάνουν το ότι οι έρευνες γίνονται μέσω του κινητού τηλεφώνου κι έτσι οι χρήστες μπορούν να συμπληρώσουν τα διάφορα ερωτηματολόγια οποιαδήποτε στιγμή το επιθυμούν. Επιπλέον μειώνονται τυχόν λάθη όπως και ο διοικητικός φόρτος για τους συμμετέχοντες και τους διαχειριστές.

2.1.2 War on Melanoma

Το Melanoma Community Registry ξεκίνησε τον Μάιο του 2014, χρηματοδοτείται από το OHSU Department of Dermatology and Knight Cancer Institute, το Colorado Foundation for Public Health & the Environment, το N.L. Tartar Trust αλλά και από δωρεές και είναι μια σημαντική πηγή από εθελοντές που απασχολούνται στο War on Melanoma. Το μητρώο αυτό περιλαμβάνει ανθρώπους που έχουν επιβιώσει από το μελάνωμα, τις οικογένειές τους ή και φίλους τους. Οι εγγεγραμμένοι λαμβάνουν

πληροφορίες για μελλοντικά project και εκδηλώσεις που αφορούν το μελάνωμα. Επίσης καλούνται να απαντήσουν σε έρευνες και ερωτηματολόγια δίνοντας έτσι ζωτικής σημασίας πληροφορίες για τους ερευνητές που αναζητούν λύσεις για την πρόληψη και την θεραπεία της ασθένειας. Ένας τρόπος με τον οποίο μπορούν οι εθελοντές να συμμετέχουν είναι μέσω του Mole Mapper μιας εφαρμογής για συσκευές iPhone που επιτρέπει την ερευνητική συμμετοχή των χρηστών καθώς ανιχνεύουν ύποπτες ελιές (moles) στο σώμα τους. Διαφορετικά οι ενδιαφερόμενοι μπορούν να εγγραφούν στο μητρώο από την σελίδα του War on Melanoma συμπληρώνοντας μια φόρμα με στοιχεία όπως την διεύθυνση του email, την ημερομηνία γέννησης, την διεύθυνση κατοικίας, προσωπικό και οικογενειακό ιατρικό ιστορικό σχετικό με το μελάνωμα, χαρακτηριστικά όπως χρώμα μαλλιών, ματιών και δέρματος, διάφορες καθημερινές συμπεριφορές που ενδεχομένως αποτελούν παράγοντα κινδύνου για την ασθένεια όπως για παράδειγμα η έκθεση στην ηλιακή ακτινοβολία, συνθήκες που επηρεάζουν την ικανότητα πρόσβασης σε ιατρική περίθαλψη και άλλα κοινωνικά δεδομένα. Επιπλέον ζητείται φωτογραφικό υλικό από την εξέλιξη της δερματικής βλάβης και συναίνεση του εγγεγραμμένου για τον διαμοιρασμό των δεδομένων που υποβάλλει με άλλους φορείς έρευνας όπως το BioLibrary, ιατρικά αρχεία και παθολογικές αναφορές. Τέλος οι χρήστες δηλώνουν το ενδιαφέρον τους για εθελοντικές δράσεις και την διαθεσιμότητά τους για αποκτήσουν οι ίδιοι σχετικές γνώσεις και εκπαιδευτικές δεξιότητες με σκοπό να προωθήσουν αυτές τις γνώσεις στην κοινότητα και κατ' αυτόν τον τρόπο να μεταφέρουν ευρύτερα με διάφορες προσεγγιστικές δραστηριότητες τη «φωνή του ασθενούς». Τα δεδομένα που συλλέγονται βοηθούν τους ερευνητές να αποκτήσουν μια περισσότερο ολοκληρωμένη εικόνα των λειτουργιών του ανθρώπινου σώματος, να αναπτύξουν καινούργιους και ακριβέστερους τρόπους για να γίνεται η διάγνωση και να βρίσκουν καινούριες θεραπείες και φάρμακα για την αντιμετώπιση του μελανώματος. Στόχος του μητρώου είναι η καλύτερη ενημέρωση του απλού κόσμου με σκοπό την πρόληψη και την έγκαιρη διάγνωση.

2.1.3 CURE OM Patient Registry

Το CURE OM (Ocular Melanoma) Patient Registry είναι ένα ερευνητικό πρόγραμμα που βασίζεται σε ασθενείς με οφθαλμικό μελάνωμα, έγινε γνωστό το 2016 στο CURE OM Patient and Caregiver Symposium και από τότε ειδικοί υγείας για σπάνιες

νόσους, άνθρωποι με ενδιαφέρον από την κοινότητα και προσωπικό υγείας αφιέρωσαν αρκετό χρόνο για έρευνα στο μητρώο έτσι ώστε να βελτιωθεί παίρνοντας την τελική μορφή του και δημιουργώντας μια εκτενή βάση δεδομένων. Η χρηματοδότηση γίνεται από το Melanoma Foundation's CURE OM και από δωρεές. Στόχος ήταν το μητρώο να τεθεί σε λειτουργία στις αρχές του 2021 ξεκινώντας από τις Ηνωμένες Πολιτείες με πρόθεση να επεκταθεί και να αποκτήσει συμμετέχοντες από όλον τον κόσμο και για αυτό συνεργάζονται επιστήμονες από πολλούς και διαφορετικούς κλάδους της υγείας και της πληροφορικής. Τα δεδομένα που θα συλλέγονται πρόκειται να είναι σχετικά με τη γενετική επίπτωση και τα χαρακτηριστικά της νόσου, το περιβάλλον στο οποίο ζει ο ασθενής, του τύπους των θεραπειών, τα φυσικά αποτελέσματα και την ποιότητα ζωής του ασθενούς. Επίσης θα ενημερώνει για τις ερευνητικές πρωτοβουλίες που αφορούν την πολιτική και τις προτιμήσεις των ασθενών στα πρότυπα φροντίδας. Σκοπός είναι μέσω αναλύσεων στα δεδομένα του μητρώου να παραχθεί χρήσιμη γνώση σχετικά με την ασθένεια, διάφορες ανωμαλίες, και ευρήματα που δεν είναι ακόμα γνωστά έτσι ώστε οι ειδικοί υγείας να οδηγηθούν σε ακριβέστερες θεραπείες. Προς το παρόν μέσα στα χρόνια προετοιμασίας του μητρώου η ομάδα του CURE OM έχει καταφέρει να βρει χρηματοδότηση, να έρθει σε επαφή με την κοινότητα των ασθενών με οφθαλμικό μελάνωμα, να προσλάβει αξιόλογους ειδικούς όπως η κυρία Jacqueline Kraska επιβλέπουσα της προόδου του μητρώου με εμπειρία 50 χρόνων σε σχετικές κλινικές έρευνες και μητρώα ασθενών αλλά και πολλούς ακόμα ειδικούς σε κλάδους όπως η επιδημιολογία, η βιοστατιστική και ερευνητές από το National Eye Institute. Επιπροσθέτως η ομάδα έχει παρουσιάσει και κατά συνέπεια προωθήσει το μητρώο σε πολλά συνέδρια στο Ηνωμένο Βασίλειο αλλά και στο εξωτερικό. Μερικά από αυτά είναι το 7th Annual Eyes on a Cure: Patient & Caregiver Symposium, το 8th Annual Eyes on a Cure: Patient & Caregiver Symposium, το Society for Melanoma Research Congress in Manchester, England αλλά και το International Society Ocular Oncology in Los Angeles CA.

2.1.4 Yale Melanoma Registry

Το Yale University και το Yale New Haven Hospital μέσω της πρωτοβουλίας του δόκτωρ Stephan Ariyan και του συμβουλίου όγκου μελανώματος με την υποστήριξη του YCCI έχουν δημιουργήσει μια ολοκληρωμένη βάση δεδομένων που περιέχει

λεπτομερείς πληροφορίες για περίπου 3800 ασθενείς διαγνωσμένους με μελάνωμα και προσβλέπουν στην άμεση ένταξη στην πλατφόρμα ανθρώπων με μη επιβεβαιωμένη διάγνωση. Το μητρώο δημιουργήθηκε με σκοπό την έρευνα και την διευκόλυνση στην λήψη αποφάσεων για την κλινική θεραπεία του μελανώματος. Η βάση δεδομένων τεχνικά είναι διασυνδεδεμένη με το YSPORE, το οποίο είναι μια ερευνητική ομάδα του Yale που μελετά τον καρκίνο του δέρματος, και συμπεριλαμβάνει ένα project της ομάδας το Specimen Resource Core. Αυτό έχει ως σκοπό να ενσωματώσει στο μητρώο την ιατρική περίθαλψη του ασθενούς με μια ανάλυση μετάλλαξης όγκου έτσι ώστε να μπορεί να υποστηρίξει μελλοντικές παρατηρητικές και μηχανιστικές μελέτες που αφορούν το μελάνωμα. Σύμφωνα με τους δημιουργούς η έρευνα για το μελάνωμα και τη θεραπεία των ασθενών εξελίσσονται ταχέως καθοδηγούμενες σε μεγάλο βαθμό από την ανάπτυξη νέων μεθόδων θεραπείας και συνεπώς το Yale Melanoma Registry επιδιώκει να διαδραματίσει δυναμικό ρόλο στην υποβοήθηση της προσαρμογής χειρουργικών, ακτινολογικών και θεραπευτικών παρεμβάσεων για το μελάνωμα, για τη βελτίωση της ποιότητας ζωής των ασθενών και την πληρέστερη ενημέρωση τόσο των ίδιων όσο και των ειδικών επιστημόνων υγείας.

2.1.5 Dutch Melanoma Treatment Registry

Το Dutch Melanoma Treatment Registry (DMTR) ιδρύθηκε τον Ιούλιο του 2013 για να διαφυλάξει την ασφάλεια και την ποιότητα της παροχής φροντίδας των ασθενών που έχουν διαγνωστεί με μελάνωμα στην Ολλανδία. Η βάση δεδομένων περιλαμβάνει λεπτομερείς πληροφορίες για όλους τους Ολλανδούς ασθενείς με μη αναστρέψιμο στάδιο 3 μελανώματος και στάδιο 4. Οι πληροφορίες αυτές είναι σχετικές με τα εξωτερικά χαρακτηριστικά των ασθενών αλλά και της ίδιας της δερματικής βλάβης, διάφορα πρότυπα θεραπειών, κλινικά αποτελέσματα, αναφορές στην ποιότητα ζωής των ασθενών όπως και της υγειονομικής περίθαλψης στην οποία υποβλήθηκαν και παρατηρήσεις για τυχόν απώλεια της παραγωγικότητας του ατόμου. Πρωταρχικός σκοπός αυτής της συλλογής δεδομένων είναι ο κλινικός έλεγχος έτσι ώστε να γίνει όσο το δυνατόν πιο διαφανής η φροντίδα των ασθενών και να παραχθούν χρήσιμες πληροφορίες για την πραγματική σχέση κόστους και αποτελεσματικότητας. Επιμέρους σκοπό αποτελεί η δημιουργία μιας πλατφόρμας για έρευνα μέσω της οποίας με δεδομένα μόλις δύο ετών οι ειδικοί έχουν ήδη οδηγηθεί σε συμπεράσματα.

Ένα από αυτά αφορά τους δείκτες απόδοσης ποιότητας οι οποίοι έδειξαν ότι οι αναστολείς BRAF και το ipilimumab, τα οποία αποτελούν στοχευμένες θεραπείες, έχουν ενσωματωθεί με ασφάλεια στο σύστημα φροντίδας της Ολλανδίας με ποσοστά τοξικότητας που ήταν συνεπείς με τις κλινικές δοκιμές της φάσης 3 που πραγματοποιήθηκαν. Σημειώνεται επίσης ότι η μέση συνολική επιβίωση των ασθενών που έλαβαν συστηματική θεραπεία ήταν 10,1 μήνες (με διάστημα εμπιστοσύνης 95% [CI] 9,1-11,1) κατά το πρώτο έτος καταχώρησης στο μητρώο και 12,7 μήνες (με διάστημα εμπιστοσύνης 95% [CI] 11,6-13,7) κατά το δεύτερο έτος. Η ομάδα του DMTR προσβλέπει στο μέλλον μέσω του μητρώου να παρέχονται πληροφορίες για δύσκολες ερωτήσεις σχετικά με τον ορισμό των πιθανών υποομάδων ασθενών που φαίνεται να επωφελούνται περισσότερο από τα νέα φάρμακα.

2.1.6 EUMelaReg

Το Ευρωπαϊκό μητρώο για ασθενείς με μελάνωμα συλλέγει δεδομένα από ασθενείς της Ευρώπης και τα ενσωματώνει σε μια ενιαία βάση δεδομένων δημιουργώντας κατ' αυτόν τον τρόπο μια μεγάλη πλατφόρμα για έρευνα στο μελάνωμα και στους άλλους τύπους καρκίνου του δέρματος. Προς το παρόν το μητρώο συνεργάζεται με 12 ευρωπαϊκές χώρες και τις αντίστοιχες ομάδες τους που αναφέρονται ονομαστικά παρακάτω: ADOReg – German Skin Cancer Registry (Germany), GEM Registry – Spanish Multidisciplinary Melanoma Group Registry (Spanish), Dutch Melanoma Database (Danish), Central South Eastern European Registry (Croatia, Serbia, Bosnia and Herzegovina), RoMelaReg – Romanian Melanoma Registry (Romania), SwiMelaReg – Swiss Melanoma Registry (Switzerland), MelaExpert (Poland), HESMO – Hellenic Society of Medical Oncology (Greece), CNMR – Clinic National Melanoma Registry (Italy). Οι πληροφορίες που συλλέγονται σε αυτό περιλαμβάνουν ένα ευρύ φάσμα από επιδημιολογικά πρωτογενή δεδομένα, πληροφορίες κλινικής θεραπείας και πτυχές οικονομικών σημείων της υγείας. Κινητήριες δυνάμεις για την επιτυχία του μητρώου αποτελούν η ευημερία των ασθενών με κακοήθες μελάνωμα και άλλους τύπους καρκίνου του δέρματος αλλά και η εύρεση ακριβέστερης και αποτελεσματικότερης θεραπείας. Παράλληλα κύριος στόχος είναι η δημιουργία και η διατήρηση ενός κεντρικού και ενωτικού μητρώου που θα περιέχει κλινικά δεδομένα διάγνωσης, θεραπείας και αποτελεσμάτων από επιβεβαιωμένα περιστατικά μελανώματος από διάφορες ευρωπαϊκές χώρες. Επιπλέον οι δημιουργοί αποσκοπούν

στην εξασφάλιση υψηλής ποιότητας και καλά καθορισμένων συνόλων δεδομένων που θα αναφέρονται στην κλινική εικόνα των ασθενών προκειμένου να διατηρηθεί η περιεκτικότητα για να μπορεί το μητρώο να καλύπτει ένα μεγάλο σύνολο ασθενών που θα προσθέτουν λεπτομερείς πληροφορίες για την κλινική τους κατάσταση. Τα δεδομένα διατίθενται σε όλες τις ενδιαφερόμενες επιστημονικές και ιατρικές εταιρείες αλλά και σε μεμονωμένα κέντρα θεραπείας του μελανώματος.

2.1.7 The Spanish National Cutaneous Melanoma Registry

Το The Spanish National Cutaneous Melanoma Registry δημιουργήθηκε το 1997 για να καταγράψει τα χαρακτηριστικά των διαγνώσεων του μελανώματος. Το μητρώο περιέχει πληροφορίες για 14.039 ασθενείς και εξυπηρετεί στην έρευνα για τη νόσο. Οι ασθενείς που περιλαμβάνονται έχουν διαγνωστεί με μελάνωμα έπειτα από διαγνωστική βιοψία και δίνουν πληροφορίες σχετικές με τα αποτελέσματά της, την ημερομηνία γέννησης τους, το φύλλο τους, την ημερομηνία της διάγνωσης, την τοποθεσία τη δερματικής βλάβης, την πρόγνωση και την ύπαρξη ή μη κάποιας ενδεχόμενης μετάστασης. Επιπλέον παρέχονται πληροφορίες για επιδημιολογικά δεδομένα που σχετίζονται με τον όγκο, την ιστοπαθολογία, το μέγεθος του όγκου, την παρουσία ή μη εξέλκωσης, τα επίπεδα γαλακτικής αφυδρογονάσης (LDH) και ταξινόμηση σύμφωνα με την Αμερικανική επιτροπή για τον καρκίνο (AJCC). Έπειτα από έρευνες βασισμένες στα δεδομένα του μητρώου βρέθηκε πως για τα χρόνια μεταξύ 1997 και 2011, το 56.5% των εγγεγραμμένων ασθενών είναι γυναίκες, ενώ το 43.5% άντρες και οι διαγνώσεις στο 37.1% αφορούν μελάνωμα στον κορμό που είναι και το πιο σύνηθες. Επιπλέον ο πιο συνήθης κλινικός-παθολογικός τύπος μελανώματος φαίνεται να ήταν το επιφανειακώς εκτεινόμενο μελάνωμα, ακολουθούμενο από το οζώδες μελάνωμα με ποσοστά 62.6% και 16.8% αντίστοιχα. Οι περιπτώσεις στις οποίες δεν παρατηρήθηκε μετάσταση αφορούσαν το 86.2% των ασθενών, ενώ το 9.9% διακρίνονταν από μεταστάσεις σε κοντινούς ιστούς και το 3.9% σε μεταστάσεις σε πιο μακρινούς ιστούς. Παράλληλα βρέθηκε πως ανεξαρτήτως ηλικίας οι άντρες είχαν παχύτερους όγκους, πολύ συχνά δημιουργία εξελκώματος, υψηλότερα επίπεδα γαλακτικής αφυγοδρονάσης και υψηλότερο ποσοστό μετάστασης από τις γυναίκες. Η μελέτη αυτή κατέληγε στο συμπέρασμα πως οι εκστρατείες πρόληψης για το μελάνωμα θα πρέπει να εστιάζουν περισσότερο

στους άντρες άνω των 50 ετών διότι έχουν την τάση να αναπτύσσουν παχύτερους όγκους και επομένως έχουν χειρότερη πρόγνωση.

2.1.8 Danish Melanoma Database

Το Danish Melanoma Database (DMD) αποτελεί μέρος του Danish Multidisciplinary Cancer Groups (οργανισμός που εστιάζει στην βελτίωση της φροντίδας των Δανών ασθενών με μελάνωμα), χρηματοδοτείται από το Danish Regions, διοικείται από το Danish Clinical Registries (RKKP) και από το 1985 σε εθνικό επίπεδο συλλέγει λεπτομερή δεδομένα με σκοπό την παρακολούθηση και τη βελτίωση της θεραπείας και της επιβίωσης των ασθενών με μελάνωμα. Αυτή η πλατφόρμα απευθύνεται σε όλους τους Δανούς ασθενείς με δερματικό μελάνωμα και μελάνωμα σταδίου 0 (situ melanoma). Το DMD είναι μια παλιά βάση δεδομένων αλλά καινούριο εργαλείο ως μητρώο κλινικής ποιότητας. Το 2011 η βάση δεδομένων θεωρήθηκε ότι περιέχει υψηλής ποιότητας δεδομένα υπό την υποστήριξη του Danish Regions' Clinical Quality Program, μέχρι τότε ήταν μια ιδιωτική πρωτοβουλία που καθοδηγούνταν από ειδικούς για το μελάνωμα και η χρηματοδότηση ήταν ασταθής. Τα δεδομένα τα οποία συλλέγονται είναι δημογραφικά, κλινικά και παθολογικά και περιλαμβάνουν χαρακτηριστικά όπως το πάχος του όγκου (Breslow Depth), τη δημιουργία εξελκώματος, τις μιτώσεις και το στάδιο μετάστασης του όγκου. Επιπλέον περιέχονται πληροφορίες σχετικές με την ημερομηνία διάγνωσης, την θεραπεία, τον τύπο της χειρουργικής επέμβασης στην οποία ίσως έχουν υποβληθεί, τη λεμφοσκόπηση (ημερομηνία εξέτασης, τεχνική, αποτελέσματα) και τη βιοψία (ημερομηνία, μέθοδο, δραστηριότητα των αφαιρούμενων δειγμάτων, εναπομένουσα δραστηριότητα στην περιοχή) σε ασθενείς για τους οποίους αυτό υποδείχθηκε, διάφορες πληροφορίες παρακολούθησης συμπεριλαμβανομένων τυχόν υποτροπή και θεραπεία. Συγκεκριμένα για την παρακολούθηση γίνεται ενδελεχής εξέταση της χειρουργημένης περιοχής και του υπόλοιπου δέρματος του ασθενή και διεξοδική ψηλάφηση των περιοχών των λεμφαδένων, αναζήτηση μεταστάσεων και καινούργιες εξετάσεις μέσω τομογραφιών σε περιπτώσεις υποψίας κάποιας μετάστασης. Το πρόγραμμα παρακολούθησης εφαρμόζεται συχνότερα σε ασθενείς με «φτωχότερη» πρόγνωση και λιγότερο συχνά σε ασθενείς με καλύτερη πρόγνωση. Σε περίπτωση θανάτου περιλαμβάνονται η αιτία και η ημερομηνία και ο εγγεγραμμένος καθίσταται

εκτός μελέτης. Στόχος είναι να επεκταθεί και να αποκτήσει ακόμα περισσότερες κλινικές παραμέτρους για να ωφελήσει τόσο τη θεραπεία του ίδιου του ασθενούς, όσο και την έρευνα. Η λίστα των δεικτών που παρακολουθούνται επεκτείνεται συνεχώς και εκδίδονται ετήσιες εκθέσεις ποιότητας ανά γεωγραφική περιοχή. Αρκετές επιστημονικές μελέτες που αφορούν επιδημιολογικές, κλινικές, γενετικές, παθολογικές και κυτταρικές έρευνες βασίζονται σε δεδομένα του DMD. Για παράδειγμα μια από τις επιδημιολογικές μελέτες με βάση τα δεδομένα του μητρώου παρακολουθεί την εξέλιξη της επίπτωσης και της θνησιμότητας με την πάροδο του χρόνου. Παρατηρείται πως τα περιστατικά μελανώματος αυξάνονται κατά 4% - 5% κάθε χρόνο.

2.1.9 Swedish Melanoma Registry (SweMR)

Το Swedish Melanoma Registry δημιουργήθηκε το έτος 1990 και περιέχει δεδομένα για ανθρώπους όλων των ηλικιών που αφορούν την διάγνωση, την θεραπεία, την πρωτοβάθμια φροντίδα υγείας και περιπτώσεις ανθρώπων που έχουν επιζήσει από το μελάνωμα. Αξιοσημείωτο είναι πως έχει κατασκευαστεί ένα προγνωστικό μοντέλο μελανώματος το οποίο είναι βασισμένο σε όλα τα δεδομένα του SweMR και επιπλέον ενημερώνεται από το Σουηδικό μητρώο θανάτων. Κάποιοι από τους σκοπούς διατήρησης του μητρώου περιλαμβάνουν την καταγραφή ειδικών παραγόντων κινδύνου που σχετίζονται με την ύπαρξη καρκίνου, τις θεραπείες που υφίστανται οι ασθενείς όπως και την επιβίωση αυτών. Ακόμα σκοπός είναι η δημιουργία μιας εθνικής βάσης για έρευνα, η οποία μεταξύ άλλων θα καθιστά δυνατή την παρακολούθηση των εθνικών εξελίξεων για να γίνουν παγκόσμιες συγκρίσεις επικράτησης της νόσου, επίπτωσης της νόσου και επιβίωσης από αυτήν. Επίσης παρέχονται σε μεμονωμένες κλινικές ή νοσοκομεία συγκρίσεις με περιφερειακά και εθνικά δεδομένα σε μια Ηλεκτρονική Έκθεση η οποία ενημερώνεται τρεις φορές τον χρόνο. Παράλληλα προσφέρει και μια ετήσια αναφορά στους παρόχους φροντίδας υγείας. Τέλος βασικό σκοπό επίσης αποτελεί η περιγραφή της φροντίδας των ασθενών με μελάνωμα μέσα στα χρόνια με την βοήθεια κάποιων δεικτών ποιότητας.

2.2 Συγκριτική μελέτη Μητρώων Ασθενών για Μελάνωμα

Τα μητρώα που αναλύθηκαν στο προηγούμενο κεφάλαιο έχουν κατ' ουσίαν τον ίδιο σκοπό, δηλαδή την βοήθεια εξαγωγής περαιτέρω γνώσεων σχετικά με το μελάνωμα, αλλά όπως φαίνεται διαφέρουν στον τρόπο που συλλέγουν τα δεδομένα τους, στους

τύπους δεδομένων που συλλέγουν, στους ασθενείς που απευθύνονται, στη διασύνδεσή τους με άλλες βάσεις δεδομένων και στην εξυπηρέτηση περαιτέρω σκοπών. Αναλυτικότερα όσον αφορά των τρόπο συλλογής των δεδομένων, είδαμε πως υπάρχουν μητρώα όπως το Melanoma UK και το War on Melanoma τα οποία διαθέτουν εξειδικευμένη εφαρμογή που βοηθάει στην συλλογή των δεδομένων, στην καλύτερη ενημέρωση των ασθενών, μέχρι και την ανίχνευση ύποπτων ελιών. Επιπλέον παρατηρήθηκαν διαφορές στα δεδομένα που συλλέγουν τα μητρώα, μιας και εκτός από δημογραφικά και κλινικά δεδομένα, βρέθηκαν μητρώα που εξέταζαν ένα σύνολο πληροφοριών που αφορούν μοριακά δεδομένα (όπως τα μητρώα War on Melanoma, Cure OM Registry και Yale Melanoma Registry) αλλά και δεδομένα που αφορούν κοινωνικές συνήθειες των ασθενών (όπως τα μητρώα EUMelaReg, War on Melanoma) μέσω των οποίων μπορούν να εξαχθούν χρήσιμα επιδημιολογικά δεδομένα. Αξιοσημείωτο είναι επίσης πως κάθε μητρώο απευθύνονταν σε διαφορετικού τύπου ασθενής, δηλαδή είτε ασθενείς με καρκίνο του δέρματος σε διαφορετικό στάδιο είτε πρώην ασθενείς που πλέον έχουν ξεπεράσει τη νόσο και τον κίνδυνο. Όμως υπήρχαν και κάποια μητρώα όπως το War on Melanoma το οποίο περιλαμβάνει την οικογένεια ή και τους φίλους των ασθενών, αλλά και το Yale Melanoma Registry που σκοπεύει να εντάξει στο μητρώο του ανθρώπους που έχουν την υποψία μελανώματος. Ακόμα μια καίριας σημασίας παρατήρηση είναι η διασύνδεση των μητρώων με άλλες βάσεις δεδομένων με τις οποίες μοιράζονται δεδομένα, παράδειγμα αποτελεί το μητρώο Yale Melanoma Registry. Τέλος βρέθηκαν διαφορές στον σκοπό των μητρώων αφού υπάρχουν μητρώα με μοναδικό σκοπό την ίαση, την θεραπεία των ασθενών και την έρευνα όπως τα περισσότερα, αλλά υπάρχουν και άλλα που εστιάζουν στην ενημέρωση όπως το War on Melanoma ή και κοινωνικούς σκοπούς που αφορούν τα οικονομικά της υγείας όπως το μητρώο EUMelaReg. Στον πίνακα που ακολουθεί συνοψίζονται αυτά που συζητήθηκαν στο παρόν κεφάλαιο.

Πίνακας 1: Συγκριτική μελέτη Μητρώων Ασθενών με Μελάνωμα

Μητρώα	Ενεργά/ Σε λειτουργία	Διασύνδεση με μοριακές βάσεις δεδομένων	Αφορούν μόνο ασθενείς	Περιέχουν μοριακά δεδομένα	Σκοπός: Κοινωνικός	Σκοπός: Ίαση/ Θεραπείες	Διαθέτουν εφαρμογή
--------	-----------------------------	--	-----------------------------	----------------------------------	-----------------------	-------------------------------	-----------------------

Melanoma Uk	NAI	OXI	NAI	OXI	OXI	NAI	NAI
War On Melanoma	NAI	NAI	OXI	NAI	OXI	NAI	NAI
CURE OM Patient Registry	OXI	OXI	NAI	NAI	OXI	NAI	OXI
Yale Melanoma Registry	NAI	NAI	NAI	NAI	OXI	NAI	OXI
Dutch Melanoma Treatment Registry	NAI	OXI	NAI	OXI	NAI	OXI	OXI
EUMelaReg	NAI	OXI	NAI	OXI	NAI	NAI	OXI
The Spanish Cutaneous Melanoma Registry	NAI	OXI	NAI	OXI	OXI	NAI	OXI
Danish Melanoma Database	NAI	OXI	NAI	OXI	OXI	NAI	OXI
SweMR	NAI	OXI	NAI	OXI	OXI	NAI	OXI

Κεφάλαιο 3^ο : Δεδομένα Γονιδιακά - Ανάλυση

3.1 Μικροσυστοιχίες γονιδίων

Οι μικροσυστοιχίες DNA είναι μια από τις τεχνολογίες που αξιοποιούν τις πληροφορίες και τα αντιδραστήρια που προέκυψαν από τα προγράμματα αλληλούχησης γονιδιώματος και έχουν επιτρέψει τη διερεύνηση των προτύπων γονιδιακής έκφρασης σε καθολική κλίμακα. Καθώς γίνονται γνωστές οι μικροσυστοιχίες και άλλες τεχνολογίες, οι δυσκολίες που συνδέονται με τη συλλογή και τη διαχείριση των δεδομένων από τα διάφορα πειράματα αυξάνονται σημαντικά.

Η βελτίωση της αξιοπιστίας των εργαστηριακών πρωτοκόλλων η μείωση της τιμής των μικροσυστοιχιών του εμπορίου και η βελτιωμένη κατανόηση των περιπλοκών του πειραματικού σχεδιασμού έχουν οδηγήσει σε περισσότερο περίπλοκα πειράματα, τα οποία παράγουν τεράστιες ποσότητες δεδομένων. Χαρακτηριστικά αναφέρεται ότι λίγα χρόνια πριν οι μελέτες με τη χρήση μικροσυστοιχιών περιλάμβαναν κατά κανόνα δέκα περίπου πειραματικές δοκιμασίες, ενώ σήμερα οι μελέτες συνηθίζεται να περιλαμβάνουν 100 ή και περισσότερες τέτοιες δοκιμασίες. Ένας αυξανόμενος αριθμός προσεγγίσεων χρησιμοποιείται για την κατασκευή μικροσυστοιχιών DNA. Σε αυτούς περιλαμβάνεται και η μέθοδος παραγωγής των συστοιχιών GeneChip της Affymetrix, που χρησιμοποιήθηκε στα δείγματα της παρούσας εργασίας, μέσω της οποίας παράχθηκε μια συστοιχία με μοναδικές «διευθυνσιοδοτήσιμες» θέσεις (addressable positions) που περιέχουν διαφορετικούς, διακριτούς ανιχνευτές οι οποίοι αντιπροσωπεύουν συνήθως μεμονωμένα γονίδια, με σκοπό τη μελέτη της έκφρασής τους. Συνεπώς οι μικροσυστοιχίες DNA είναι μια διάταξη συστηματικά τοποθετημένων αλληλουχιών γονιδίων τα οποία εμφανίζονται ως μικροσκοπικές κουκίδες (spots) πάνω στην επιφάνεια τοποθέτησης και ονομάζονται ανιχνευτές (probes). Στην πράξη οι μικροσυστοιχίες δεν μετρούν άμεσα τη γονιδιακή έκφραση αλλά μέσω της έντασης φωτός (που μετράται αντιστοιχούμενη σε απόλυτους αριθμούς). Έτσι, για ένα συγκεκριμένο εξεταζόμενο γονίδιο, περισσότερα αντίγραφα οδηγούν σε μεγαλύτερη ένταση στον αντίστοιχο ανιχνευτή. Πειράματα DNA μικροσυστοιχιών ενσωματώνουν μεγάλο αριθμό βημάτων από την απομόνωση του βιολογικού υλικού στα εργαστήρια μέχρι την τελική ανάλυση και ερμηνεία, κάθε ένα από τα οποία εισάγει το δικό του θόρυβο στα τελικά αποτελέσματα που τα κάνει να χρήζουν επεξεργασίας. Κατά συνέπεια είναι πολύ σημαντικός ο σωστός πειραματικός σχεδιασμός και η προσεκτική και ακριβής βιοπληροφορική και στατιστική ανάλυση, έτσι ώστε να είναι επιτυχής και αξιόπιστη η τελική ερμηνεία των πρωτογενών δεδομένων.

3.2 Υλικό και μέθοδοι

Όσον αφορά το υλικό που απαιτείται για την διαδικασία, χρησιμοποιούνται ολιγονουκλεοτιδικές DNA μικροσυστοιχίες υψηλής πυκνότητας. Αυτός ο τύπος μικροσυστοιχιών αποτελείται από συγκεκριμένες ολιγονουκλεοτιδικές αλληλουχίες

που συντίθενται πάνω σε σταθερή επιφάνεια πυριτίου με χρήση φωτοευαίσθητων υποστρωμάτων και φωτολιθογραφικών πλακών. Εφαρμόζεται μονοχρωματικός υβριδισμός (single channel) καθώς μόνο ένα δείγμα σημαίνεται με κατάλληλη χρωστική και υβριδοποιείται στο υπόστρωμα της μικροσυστοιχίας κάθε φορά. Ο όρος "συστοιχία ολιγονουκλεοτιδίων" αναφέρεται στην ειδική τεχνική κατασκευής, η οποία περιλαμβάνει την άμεση σύνθεση των επιθυμητών ολιγονουκλεοτιδικών αλληλουχιών στην επιφάνεια συστοιχίας (σε αντίθεση με άλλες τεχνολογίες εναπόθεσης ακέραιων αλληλουχιών νουκλεϊκού οξέος). Στη συγκεκριμένη εργασία θα χρησιμοποιηθούν δεδομένα που προκύπτουν από τις συνηθέστερα χρησιμοποιούμενες συστοιχίες ολιγονουκλεοτιδίων, τις συστοιχίες ανιχνευτών Affymetrix GeneChip. Πρόκειται για μικροσυστοιχίες υψηλής πυκνότητας ολιγονουκλεοτιδίων που καλύπτουν πλήρως το ανθρώπινο γονιδίωμα (περισσότερες από 1.00.000 θέσεις καταγραφής φθορισμού νουκλεοτιδίων, περισσότεροι από 70.000 κλώνοι, περισσότερα από 33.000 γονίδια). Κάθε εξεταζόμενο γονίδιο διερευνάται από ένα σύνολο από 14 ζεύγη ολιγονουκλεοτιδικών ανιχνευτών, το καθένα με ένα μήκος 25 βάσεων. Συγκεκριμένα, κάθε γονίδιο πάνω στη μικροσυστοιχία αντιπροσωπεύεται από διαφορετικά probes και ενσωματώνει τη στρατηγική ανιχνευτών πλήρους/ελλιπούς ταύτισης (Perfect Match (PM) / Mismatch (MM) probes). Κάθε ζεύγος ανιχνευτών αποτελείται από έναν ανιχνευτή τέλειαν αντιστοίχισης (PM) (πλήρως συμπληρωματικός) που έχει σχεδιαστεί για να υβριδοποιείται με το μεταγραφικό στόχο πλήρως, και από έναν ανιχνευτή ελλιπούς ταύτισης (MM) που σχεδιάστηκε για τη μέτρηση του θορύβου του περιβάλλοντος και των μη ειδικών σημάτων. Οι ανιχνευτές MM κατασκευάστηκαν πανομοιότυπα με τα αντίστοιχα ζεύγη PM, με τη μόνη τους διαφορά να βρίσκεται στην 13η βάση της αλληλουχίας (ακριβώς στο κέντρο των 25 βάσεων της αλληλουχίας), η οποία αντικαταστάθηκε από τη συμπληρωματική της βάση (Irizarry et al., 2003) (Lipshutz et al., 1999). Έτσι γίνεται ποσοτικοποίηση και αφαίρεση ψευδών σημάτων που οφείλονται σε μη ειδική ή διασταυρούμενη υβριδοποίηση, δίνοντας μια εκτίμηση του θορύβου που προκύπτει. Μέχρι να φτάσουμε στο σημείο της στατιστικής ανάλυσης για τον εντοπισμό των διαφορικά εκφρασμένων γονιδίων και των λειτουργιών όπου συμμετέχουν, απαιτείται να προηγηθούν ορισμένα βήματα. Το βιολογικό πείραμα που πραγματοποιείται πριν την ανάλυση, έχει εν συντομία ως εξής: Αρχικά, ολικό mRNA (αγγελιαφόρο RNA που μεταφέρει τη γενετική πληροφορία από το DNA στα ριβοσώματα για την πρωτεϊνοσύνθεση των κυττάρων) απομονώνεται από τα κύτταρα στόχους ή ιστούς

βάσει προκαθορισμένων πρωτοκόλλων και αξιολογείται η ποιότητά του. Στη συνέχεια μέσα από εξειδικευμένες τεχνικές και διαδικασίες σημαίνεται με φθορίζουσες ή ραδιενεργές ουσίες και τα μόρια που προκύπτουν υβριδοποιούνται με τα ακινητοποιημένα τμήματα DNA (probes) στο πλακίδιο. Ακολουθούν εκπλύσεις με κατάλληλα διαλύματα ώστε να ελαχιστοποιηθούν οι πιθανότητες διασταυρούμενης υβριδοποίησης μεταξύ παρόμοιων γονιδίων μέσω διάχυσης περίσσειας υλικού. Ακολουθεί σάρωση του πλακιδίου της μικροσυστοιχίας με υψηλής ταχύτητας σαρωτές laser, που διεγείρουν τις φθορίζουσες χρωστικές δημιουργώντας φωτεινά σημεία (spots) και παράγουν ψηφιακά αρχεία εικόνων. Η ανίχνευση του σήματος γίνεται με τη χρήση φωτοπολλαπλασιαστικών αγωγών. Κάθε μια από τις χρησιμοποιούμενες χρωστικές διεγείρεται σε συγκεκριμένο μήκος κύματος και κατά την διέγερσή της εκπέμπει ακτινοβολία διαφορετικού μήκους κύματος η οποία ενισχύεται από τους φωτοπολλαπλασιαστικούς αγωγούς που τη μετατρέπουν τελικά σε ηλεκτρικό σήμα. Οι εικόνες που προκύπτουν υφίστανται περαιτέρω επεξεργασία από ειδικό λογισμικό ώστε να ποσοτικοποιηθούν οι εντάσεις φθορισμού για την κάθε θέση της μικροσυστοιχίας. Η ένταση φθορισμού λειτουργεί ως μέτρο του προσδιορισμού της σχετικής αφθονίας του εκάστοτε μεταγράφου, καθώς η ποσότητα φθορισμού που μετράται είναι ανάλογη της ποσότητας mRNA του γονιδίου το οποίο έχει μεταγραφεί και αντιπροσωπεύεται από δεδομένο σημείο της μικροσυστοιχίας.

3.3 Προέλευση Δείγματος

Αυτού του τύπου οι τεχνολογίες έχουν οδηγήσει στη δημιουργία μεγάλου όγκου μοριακών δεδομένων και γονιδιακής πληροφορίας για αυτό και κρίθηκε σημαντική η ανάπτυξη γονιδιακών βάσεων δεδομένων για να είναι δυνατή η αποθήκευση, η διατήρηση, η διάδοση της πληροφορίας και η χρήση της σε δευτερογενής έρευνες. Μια τέτοια γονιδιωματική βάση δεδομένων είναι και η Gene Expression Omnibus (GEO) του Εθνικού Κέντρου Βιοτεχνολογικών Πληροφοριών των ΗΠΑ (National Center for Biotechnology Information – NCBI). Η GEO περιέχει μεγάλο αριθμό πειραμάτων και δειγμάτων που συνοδεύονται από πληθώρα πληροφοριών σχετικών με τον τύπο των δεδομένων που περιέχουν, τον τύπο του βιολογικού υλικού, τον τύπο της μικροσυστοιχίας που χρησιμοποιήθηκε, τις πειραματικές συνθήκες κ.α. Τα δεδομένα στις βάσεις αφορούν επεξεργασμένα δεδομένα, δηλαδή δεδομένα που έχουν υποστεί προ-επεξεργασία μετά την σάρωση του πλακιδίου (chip) και την

απόκτηση των τιμών φθορισμού για την κάθε θέση, αλλά και τα πρωταρχικά μη επεξεργασμένα δεδομένα (raw data), δίνοντας κατ' αυτόν τον τρόπο την ελευθερία στο χρήστη να επαναλάβει την ανάλυση με όποιο αλγόριθμο επιθυμεί και να εξάγει τα δικά του αποτελέσματα και συμπεράσματα. Στην συγκεκριμένη εργασία επιλέχθηκαν τα μη επεξεργασμένα δεδομένα και προχωρήσαμε στην εξ αρχής επεξεργασία τους. Το δείγμα το οποίο αναλύθηκε στην παρούσα εργασία περιλαμβάνει 34 αρχεία από μελέτη που έχει πραγματοποιηθεί για το μελάνωμα και έχουν προκύψει έπειτα από την χειρουργική αφαίρεση ύποπτων σπίλων και την βιοψία αυτών των ιστών. Σύμφωνα με πληροφορίες που αντλήθηκαν από την GEO στα 34 αυτά αρχεία συμπεριλαμβάνονται 5 αρχεία για κοινούς σπίλους, 7 για δυσπλαστικούς σπίλους, 6 για κανονικό δέρμα και 16 που αφορούν πρωτοπαθές μελάνωμα (χωρίς να παρέχεται περαιτέρω πληροφορία για τα χαρακτηριστικά των ασθενών). Για τον προσδιορισμό της γονιδιακής έκφρασης χρησιμοποιήθηκε η πλατφόρμα μικροσυστοιχίας Affymetrix Human Genome U133 Plus 2.0 Array (hgu133plus2) και το αρχικό μας δείγμα με τα πρωταρχικά μη επεξεργασμένα δεδομένα αντιστοιχούσε σε 1.354.896 probe ids.

3.4 Λογισμικό

Η επεξεργασία των πρωταρχικών δεδομένων και η βιοπληροφορική ανάλυση που έγινε για να εξάγουμε τα αποτελέσματά μας, πραγματοποιήθηκαν με τη βοήθεια του RStudio και της γλώσσας προγραμματισμού R. Η R είναι μια γλώσσα προγραμματισμού ελεύθερου λογισμικού για στατιστική πληροφορική και γραφικά και συνεπώς χρησιμοποιείται ευρέως από στατιστικούς και επιστήμονες που ασχολούνται με την εξόρυξη γνώσης για την ανάπτυξη στατιστικού λογισμικού και την ανάλυση δεδομένων. Επιπλέον χρησιμοποιήθηκε το λογισμικό Bioconductor που παρέχει πακέτα βιοπληροφορικής και στατιστικής ανάλυσης. Το Bioconductor είναι ένα δωρεάν λογισμικό ανοιχτού κώδικα και ανοιχτού λογισμικού για την ανάλυση και την κατανόηση των γονιδιωματικών δεδομένων που δημιουργούνται από εργαστηριακά βιολογικά πειράματα. Στην πράξη είναι ένα αποθετήριο λογισμικού ανοιχτού κώδικα που περιλαμβάνει ένα ευρύ φάσμα στατιστικών εργαλείων. Επιπλέον αυτό βασίζεται κυρίως στη στατιστική γλώσσα προγραμματισμού R και παρέχει πολλά πακέτα σχολιασμού γονιδιώματος που είναι κυρίως προσανατολισμένα στους διαφορετικούς τύπους μικροσυστοιχιών.

3.5 Προεπεξεργασία Δεδομένων

Το τελικό σήμα που παράγεται από το σαρωτή laser στην επιφάνεια της μικροσυστοιχίας του πειράματος είναι ένα αρχείο εικόνας και αντιπροσωπεύει τις τιμές εντάσεων φθορισμού. Για να γίνει η ανάλυση των εικόνων που έχουν ληφθεί από τον σαρωτή αποκόπονται τα επιθυμητά τμήματα κάθε εικόνας και εφαρμόζεται πλέγμα (gridding) για την οριοθέτηση των σημείων. Κατά τη διαδικασία εφαρμογής πλέγματος αντιστοιχίζονται συντεταγμένες σε κάθε μια από τις κουκίδες της εικόνας έτσι ώστε να καθοριστεί με ακρίβεια η θέση τους. Έτσι δημιουργείται ένα αρχείο πλέγματος το οποίο εκτός από τις συντεταγμένες κάθε κουκίδας περιέχει και πληροφορίες για το τί αντιπροσωπεύει. Στη συνέχεια πραγματοποιείται κατάτμηση της εικόνας (segmentation) για το διαχωρισμό των κουκίδων και τη διάκριση των pixels που αντιστοιχούν σε σημεία ενδιαφέροντος (foreground) και αυτών που αντιστοιχούν στο υπόβαθρο (background) (Yang et al., 2001) (Carvalho, Irizarry, 2010). Τέλος προσδιορίζεται η ένταση φθορισμού (intensity extraction) η οποία είναι χαρακτηριστική και θα χρησιμοποιηθεί στη συνέχεια για τον τελικό μας στόχο που είναι η εκτίμηση της γονιδιακής έκφρασης. Συνεπώς στην πλατφόρμα που εμείς εξετάζουμε της Affymetrix δημιουργούνται για τα δείγματά μας αρχεία αποτελεσμάτων, τα λεγόμενα CEL files που αφορούν τα μη επεξεργασμένα δεδομένα (raw data) και περιέχουν τις συντεταγμένες κάθε σημείου, τη μέση τιμή της εντάσεως φθορισμού του υποβάθρου για κάθε κουκίδα, τις διαμέσους τιμές, τις τυπικές αποκλίσεις αλλά και μερικά χαρακτηριστικά που αφορούν την ποιότητα της μικροσυστοιχίας όπως ενδεχόμενες γρατζουνιές ή μη αξιόπιστες μετρήσεις. Έτσι για να προχωρήσουμε στην ανάλυσή μας το πρώτο βήμα που χρειάζεται να κάνουμε είναι να φορτώσουμε αυτά τα δεδομένα στο περιβάλλον του RStudio όπου δουλεύουμε. Αυτό θα γίνει με τη βοήθεια της βιβλιοθήκης GEOquery και της συνάρτησης `getGEOSuppFiles("GSE114445")` η οποία θα κατεβάσει και θα φορτώσει στο RStudio όλα τα CEL files για το δείγμα που εξετάζουμε. Στη συνέχεια μέσω της βιβλιοθήκης `affy` και της συνάρτησης `ReadAffy(verbose = TRUE, filenames = cels)` θα δημιουργηθεί ένα αντικείμενο τύπου `AffyBatch` το οποίο περιέχει τις τιμές έντασης φθορισμού από όλες τις θέσεις ανιχνευτών όλων των δειγμάτων. Έπειτα γίνεται η αντιστοίχιση των χαρακτηριστικών των δειγμάτων με το αρχείο που δημιουργήθηκε και το αποτέλεσμα είναι να έχουμε καταφέρει να μετατρέψουμε την

πληροφορία από τα αρχεία εικόνων σε ποσοτικοποιημένα και εύκολα μετρήσιμα δεδομένα τα οποία αντιστοιχούν μέσω του αντικειμένου AffyBatch στα χαρακτηριστικά των ατόμων του δείγματος.

Επομένως αφού μετατρέψαμε τα δεδομένα μας σε μορφή διαχειρίσιμη θα προχωρήσουμε στη διαδικασία της επεξεργασίας τους που απαιτείται πριν την ανάλυση των ολιγονουκλεοτιδικών DNA μικροσυστοιχιών της Affymetrix που μελετάμε. Γνωρίζουμε ότι τα μη επεξεργασμένα δεδομένα (raw data) που έχουν προκύψει από πειράματα μικροσυστοιχιών περιλαμβάνουν μεγάλα ποσοστά θορύβου και μεταβλητότητας. Οπότε για να έχουμε περισσότερο ακριβή αποτελέσματα που δεν θα παραποιούνται δηλαδή από κανενός είδους μεταβλητότητα, πριν γίνει η στατιστική ανάλυση θα πρέπει να γίνει περαιτέρω επεξεργασία των πρωταρχικών δεδομένων μας (raw data) για την απαλοιφή του τεχνικού θορύβου (συστημικά σφάλματα) και κατ' επέκταση τη λήψη περισσότερο αξιόπιστων τιμών γονιδιακής έκφρασης. Ο θόρυβος και η μεταβλητότητα που αναφέρθηκαν παραπάνω είναι δυνατό να προέρχονται από κάθε στάδιο της πειραματικής διαδικασίας και υπάρχει περίπτωση να είναι είτε βιολογικές είτε να προκύπτουν από τις διαδικασίες του πειράματος μέχρι το τελικό αποτέλεσμά του. Δηλαδή μια σειρά παραγόντων όπως το πειραματικό πρωτόκολλο που εκτελείται, ο εξοπλισμός του εργαστηρίου, τα υλικά και οι συνθήκες του πειράματος, η εμπειρία και ικανότητα του ερευνητή σε τέτοια είδους πειράματα, εισάγουν μεταβλητότητα και θόρυβο που επιθυμούμε να εξαλείψουμε για να έχουμε όσο το δυνατόν περισσότερο «καθαρά» αποτελέσματα. Για να γίνει αυτό θα ξεκινήσουμε με τη διαδικασία Διόρθωσης του θορύβου υποβάθρου (background noise), η οποία πραγματοποιεί την αφαίρεση που εγγενούς φθορισμού που παράγεται από το ίδιο το πλακίδιο της μικροσυστοιχίας. Αυτό στην πράξη αποσκοπεί στην απαλοιφή φαινομένων όπως η μη ειδική υβριδοποίηση, η παρουσία θορύβου από ξένα στοιχεία στο πλακίδιο όπως σκόνη, αλλά και διακυμάνσεις που έχουν δημιουργηθεί από τη διαδικασία σάρωσης της μικροσυστοιχίας. Το επόμενο βήμα περιλαμβάνει την κανονικοποίηση των δεδομένων εντός κάθε μικροσυστοιχίας (within array normalization) έτσι ώστε να διώξουμε τα συστημικά λάθη, αλλά και κανονικοποίηση μεταξύ μικροσυστοιχιών (between array normalization) έτσι ώστε να ισοσταθμιστούν οι διασπορές μεταξύ διαφορετικών συνθηκών. Στη συνέχεια τα αριθμητικά δεδομένα έκφρασης θα μετασχηματιστούν σε λογαριθμική κλίμακα με βάση το 2, για να υπάρχει συμμετρική μεταχείριση των λόγων έκφρασης. Έπειτα υπολογίζεται η συνοπτική τιμή έντασης

που αντιστοιχεί σε κάθε γονίδιο αφού συνδυαστούν οι τιμές έντασης όλων των ανιχνευτών που αντιστοιχούν σε κάθε γονίδιο. Ο κάθε ανιχνευτής χαρακτηρίζεται ως probe και το σύνολο αυτών ως probe set. Το κάθε probe set αντιστοιχεί σε ένα μετάγραφο, το οποίο αντιπροσωπεύει ένα γονίδιο. Τελικό στάδιο της επεξεργασίας είναι το φιλτράρισμα το οποίο γίνεται μέσω των τιμών p-value για την εξάλειψη γονιδίων ελλιπούς πληροφορίας ή γονιδίων στα οποία τα επίπεδα θορύβου είναι υψηλά σε σχέση με τα επίπεδα έκφρασης.

3.5.1 Διόρθωση Θορύβου υποβάθρου εικόνας (Background Correction)

Ξεκινώντας την επεξεργασία το πρώτο στάδιο είναι η διόρθωση του υποβάθρου της έντασης του κάθε ανιχνευτή. Αυτό το κάνουμε για να διαχωρίσουμε το αποτέλεσμα μας από τον πρόσθετο θόρυβο και να κρατήσουμε μόνο το σήμα που λαμβάνουμε από τον φθορισμό των χρωστικών ουσιών εξ αιτίας της υβριδοποίησης που έχει γίνει κατά την πειραματική διαδικασία. Ο θόρυβος που αναφέραμε παραπάνω θα μπορούσε να διακριθεί σε δύο κατηγορίες. Η πρώτη είναι ο θόρυβος πηγής όπως είναι ο θόρυβος των φωτονίων, η σκόνη στα πλακίδια, η επεξεργασία που έχει υποστεί το πλακίδιο για την διεξαγωγή του πειράματος. Η δεύτερη κατηγορία αφορά τον θόρυβο ανίχνευσης δηλαδή αυτόν που έχει προκύψει στη διαδικασία της ενίσχυσης και της ψηφιοποίησης του σήματος. Πρόβλημα επίσης αποτελούν οι ακραίες εντάσεις φθορισμού, δηλαδή οι πολύ χαμηλές εντάσεις και οι πολύ υψηλές. Στην πρώτη περίπτωση το πρόβλημα έγκειται στο ότι τις πολύ χαμηλές εντάσεις είναι δύσκολο να τις ξεχωρίσουμε από τον τοπικό θόρυβο και συνεπώς δεν μπορούμε να προσδιορίσουμε αν τελικά παρατηρείται έκφραση του γονιδίου ή όχι. Στη δεύτερη περίπτωση υψηλά επίπεδα γονιδιακής έκφρασης παράγουν και υψηλής εντάσεως φθορισμό ο οποίος ενδέχεται να φτάνει στο επίπεδο κορεσμού του σαρωτή με αποτέλεσμα να μην είναι δυνατή η διάκριση διαφορών στην έκφραση αυτών των γονιδίων, μιας και όταν η ένταση των pixels ξεπερνάει το όριο ανίχνευσης τότε της αποδίδεται η τιμή του ανώτατου ορίου, γεγονός που εισάγει κι αυτό θόρυβο στον υπολογισμό της γονιδιακής έκφρασης.

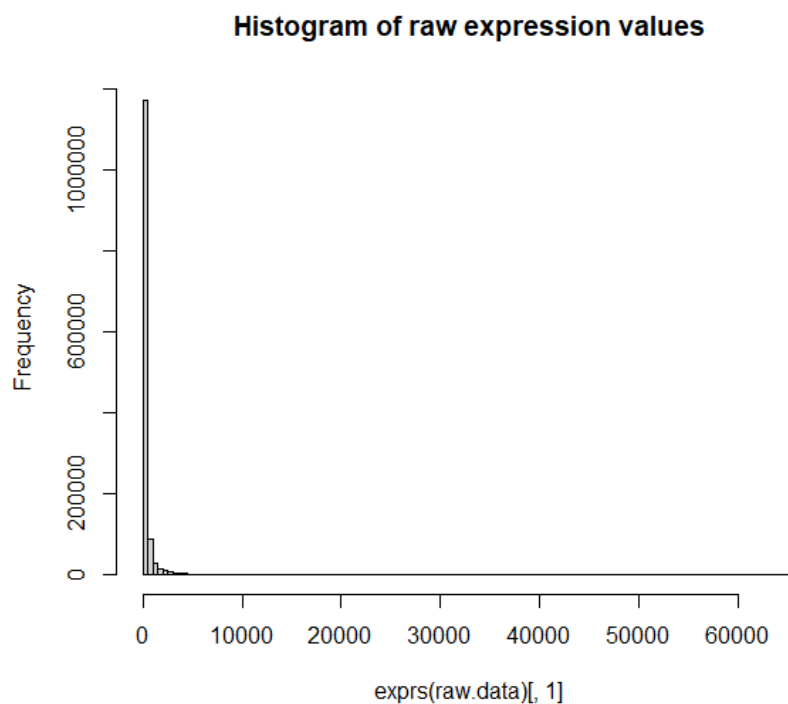
3.5.2 Κανονικοποίηση (Normalization)

Το δεύτερο βήμα στην επεξεργασία των δεδομένων είναι η κανονικοποίηση που αποσκοπεί στην απαλοιφή των συστηματικών σφαλμάτων και του τεχνικού θορύβου. Στόχος της κανονικοποίησης είναι να φέρει τα δεδομένα των επιπέδων έκφρασης που προέρχονται από κάθε δείγμα – πλακίδιο (chip) σε μορφή συγκρίσιμη για να γίνει η όσο το δυνατόν ακριβέστερη σύγκριση ομοειδών. Ο λόγος που χρειάζεται αυτό είναι επειδή υπάρχουν μικρές αποκλίσεις μεταξύ των διαδικασιών υβριδισμού για κάθε μικροσυστοιχία, οι οποίες αργότερα θα οδηγήσουν σε μεγάλες αποκλίσεις ανάμεσα στις συνολικές εντάσεις για διαφορετικές μικροσυστοιχίες. Αυτές οι αποκλίσεις μπορούν να έχουν ισχυρή επίδραση στα αποτελέσματα και ενδεχομένως να οφείλονται στον όγκο ενός δείγματος, στην ποσότητα του RNA σε κάποιο δείγμα αλλά ακόμα και στον χρόνο που χρειάζεται για να γίνει η υβριδοποίηση του δείγματος. Επιπλέον οι διακυμάνσεις πιθανώς να αφορούν τα υλικά του πειράματος, δηλαδή να υπάρχουν διαφορές μεταξύ των πλακιδίων ή μεταξύ των σαρωτών που χρησιμοποιούνται για να μετρηθούν οι τιμές φθορισμού και έτσι να οδηγούμαστε σε μη καθαρά αποτελέσματα.

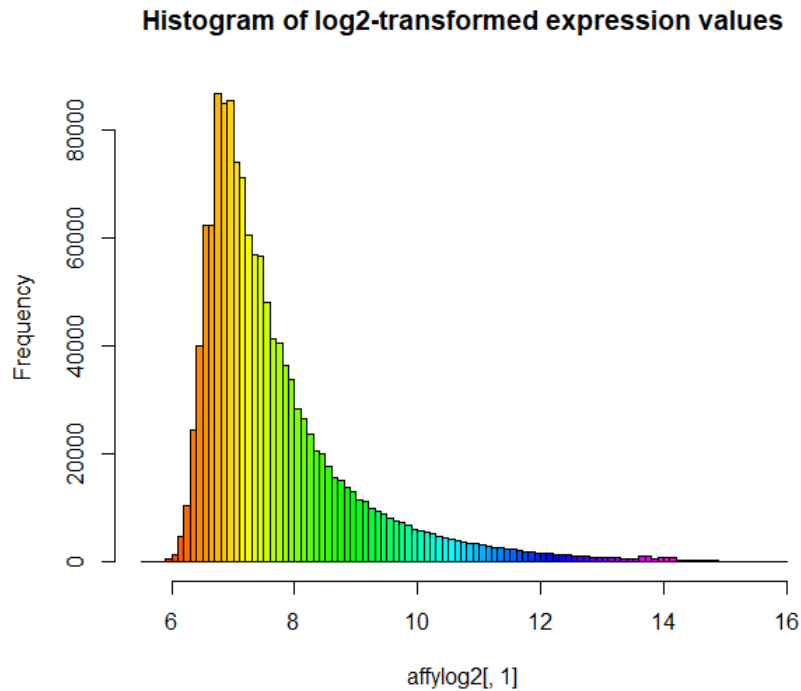
3.5.3 Λογαριθμικός Μετασχηματισμός Δεδομένων (\log_2 transformation)

Στη συνέχεια της επεξεργασίας των δεδομένων είναι πολύ βασικό να μετατρέψουμε τις αριθμητικές τιμές που έχουμε σε λογαριθμική κλίμακα. Αυτή η μετατροπή αποσκοπεί στο να γίνει η κατανομή της έντασης περισσότερο συμμετρική σε σχήμα «καμπάνας» στο διάγραμμα και συνεπώς να καταστήσει δυνατή την εφαρμογή στατιστικών μεθόδων. Διαφορετικά θα είχαμε τις εντάσεις φθορισμού να είχαν ασύμμετρη κατανομή, μιας και μαζί με τα γονίδια χαμηλής έκφρασης, υπάρχουν και πολλά γονίδια που βρίσκονται κάτω από το όριο ανίχνευσης του σαρωτή που σημαίνει ότι υπάρχουν περισσότερες εντάσεις φθορισμού στο επίπεδο έντασης φθορισμού υποβάθρου ή λίγο πάνω από το επίπεδο του υποβάθρου και πολύ λίγα γονίδια με πολύ υψηλά επίπεδα έκφρασης, γεγονός που καθιστά δύσκολο τον υπολογισμό διάφορων στατιστικών δεδομένων όπως η μέση τιμή των δεδομένων έντασης και άρα των δεδομένων έκφρασης. Επιπλέον γνωρίζουμε ότι η διακύμανση στις εντάσεις συνήθως αυξάνεται καθώς η μέση ένταση αυξάνεται. Οπότε αφού η διακύμανση εξαρτάται από τη μέση τιμή σε εκείνα τα πλακίδια όπου ο μέσος όρος θα είναι υψηλότερος, θα είναι υψηλότερη και η διακύμανση τους. Το παραπάνω αποτελεί παραβίαση της γενικής παραδοχής των παραμετρικών μοντέλων ότι όλες οι

ομάδες, δηλαδή τα πλακίδια, θα πρέπει να έχουν παρόμοιες διακυμάνσεις. Επομένως αφού με τον λογαριθμικό μετασχηματισμό διορθώνεται η εξάρτηση της διακύμανσης από τον μέσο όρο, αυξάνεται η ακρίβεια των στατιστικών αναλύσεων και κατ' επέκταση αυξάνεται η πιθανότητα ανίχνευσης της πραγματικής διαφορικής έκφρασης. Στις παρακάτω εικόνες ιστογράμματος είναι εύκολο να διακριθεί η διαφορά των τιμών εντάσεων φθορισμού ενός δείγματος πριν την εφαρμογή του λογαριθμικού μετασχηματισμού και μετά από αυτόν. Στον Y άξονα σημειώνονται οι συχνότητες εμφάνισης κάθε τιμής έντασης φθορισμού του X άξονα, με την εικόνα πριν τον μετασχηματισμό να μας δείχνει πως η συντριπτική πλειοψηφία των εντάσεων ήταν κοντά στο 0, γεγονός που αλλάζει όπως βλέπουμε στην εικόνα μετά τον μετασχηματισμό.



Εικόνα 1: Histogram of raw expression values



Εικόνα 2: Histogram of log2-transformed expression values

3.5.4 Αλγόριθμος Robust Multiarray Average (RMA)

Για την προεπεξεργασία των πρωταρχικών δεδομένων (raw data) που κάναμε δηλαδή της διόρθωσης θορύβου υποβάθρου (background correction), της κανονικοποίησης (normalization), του λογαριθμικού μετασχηματισμού (\log_2 transformation) και του υπολογισμού της συνοπτικής τιμής έντασης που αντιστοιχεί στο κάθε γονίδιο που περιγράψαμε (expression calculation) χρησιμοποιήσαμε τον αλγόριθμο RMA. Η επιλογή αυτού του αλγορίθμου έγινε διότι υπολογίζει με πολύ μεγάλη ακρίβεια την πραγματική ένταση φθορισμού και συνεπώς δίνει και ακριβέστερη εκτίμηση για την διαφοροποίηση στην έκφραση των γονιδίων. Το πρώτο βήμα που εκτελεί ο συγκεκριμένος αλγόριθμος είναι η διόρθωση υποβάθρου μέσω της αφαίρεσης του θορύβου που εξηγήσαμε παραπάνω. Η μέθοδος με την οποία το πραγματοποιεί αυτό στηρίζεται μόνο στο σήμα των ανιχνευτών πλήρους ταύτισης (PM) οι οποίοι θα μας δώσουν πληροφορία που χρησιμεύει στην ορθή εκτίμηση της γονιδιακής έκφρασης και δεν περιλαμβάνει και το σήμα ανιχνευτών ελλιπούς ταύτισης (MM) όπως συμβαίνει με άλλους αλγόριθμους. Η διόρθωση του υποβάθρου εφαρμόζεται ξεχωριστά σε κάθε μικροσυστοιχία (πλακίδιο) που συμμετέχει στη μελέτη. Για να

αποφευχθεί λανθασμένη εκτίμηση της ποσοτικοποίησης λόγω φαινομένων θέσης οι PM ανιχνευτές του κάθε γονιδίου είναι διάσπαρτοι στη μικροσυστοιχία. Η ύπαρξη πολλών διαφορετικών ανιχνευτών - κλώνων για διαφορετικές περιοχές του ίδιου γονιδίου βελτιώνει σημαντικά το λόγο της έντασης σήματος ως προς το θόρυβο, καθώς συνοψίζονται για να υπολογιστεί μια μέση τιμή των εντάσεων των πολλαπλών θέσεων, που παρέχει ακρίβεια στην ποσοτικοποίηση του RNA, ενώ μειώνει δραστικά το εύρος των ψευδώς θετικών σημάτων (Lipshutz et al., 1999) . Ο αλγόριθμος RMA εφαρμόζει μια αμερόληπτη μέθοδο παλινδρόμησης για να προσδιορίσει την έκφραση των μεταγράφων-στόχων των μικροσυστοιχιών και το τυπικό σφάλμα , βασιζόμενος αποκλειστικά στις «διορθωμένες» τιμές έντασης των θέσεων ανιχνευτή πλήρους ταύτισης PM και αγνοώντας τις MM τιμές (Bolstad et al., 2003) (Irizarry et al., 2003). Έπειτα ο αλγόριθμος πραγματοποιεί την διαδικασία της κανονικοποίησης η οποία είναι μια μη παραμετρική μέθοδος κανονικοποίησης που ονομάζεται «Κανονικοποίηση Ποσοστημορίου» (Quantile Normalization) (Bolstad et al., 2003). Αυτή η μέθοδος καταφέρνει να κανονικοποιεί οριζόντια όλες τις μικροσυστοιχίες (πλακίδια) έτσι ώστε να έχουν την ίδια κατανομή διορθώνοντας ενδεχόμενες αποκλίσεις ανάμεσά τους. Συνεπώς θεωρούμε ότι η κατανομή της αφθονίας των μεταγράφων – στόχων είναι σχεδόν ίδια σε όλα τα δείγματα. Αναλυτικότερα αυτό γίνεται συγκρίνοντας τα επίπεδα έκφρασης των ακολουθιών – στόχων ανάμεσα στις μικροσυστοιχίες για τα διάφορα ποσοστημόρια. Στο πρώτο στάδιο θα δημιουργηθεί μια συνολική κατανομή της έντασης των ανιχνευτών (probes) από όλες τις μικροσυστοιχίες η οποία αντιπροσωπεύει την κατανομή έκφρασης μιας θεωρητικής «μικροσυστοιχίας αναφοράς» (reference chip). Στη συνέχεια, για κάθε πραγματική μικροσυστοιχία υπολογίζεται για κάθε τιμή έντασης ανιχνευτή, το ποσοστημόριο της πάνω στην συνολική κατανομή της έντασης των ανιχνευτών με σκοπό η αρχική τιμή να μετασχηματιστεί στην τιμή του ποσοστημορίου που προέκυψε πάνω στην μικροσυστοιχία αναφοράς. Η μέθοδος είναι αρκετά αποτελεσματική ως προς τη μείωση της μεταβλητότητας των μεταγράφων χαμηλής συγκέντρωσης και διακρίνει καλύτερα τα μετάγραφα που παρουσιάζουν διαφορεική έκφραση από αυτά που παραμένουν σταθερά (Choe et al., 2005).

Τη διαδικασία της κανονικοποίησης ακολουθεί ο λογαριθμικός μετασχηματισμός με βάση το 2 των τιμών των εντάσεων φθορισμού που πραγματώνεται μέσω μια μεθόδου μη παραμετρικής παλινδρόμησης του αθροιστικού γραμμικού μοντέλου λογαριθμικής κλίμακας. Αυτό το στάδιο όπως αναφέραμε και παραπάνω εξυπηρετεί

στο να αποκτήσουν οι τιμές των εντάσεων κατανομή που να προσεγγίζει την κανονική έτσι ώστε να μπορεί στη συνέχεια να γίνει η στατιστική ανάλυση. Τέλος ο αλγόριθμος RMA χρειάζεται να συνοψίσει τους ανιχνευτές (probes) σε μια τιμή έκφρασης.

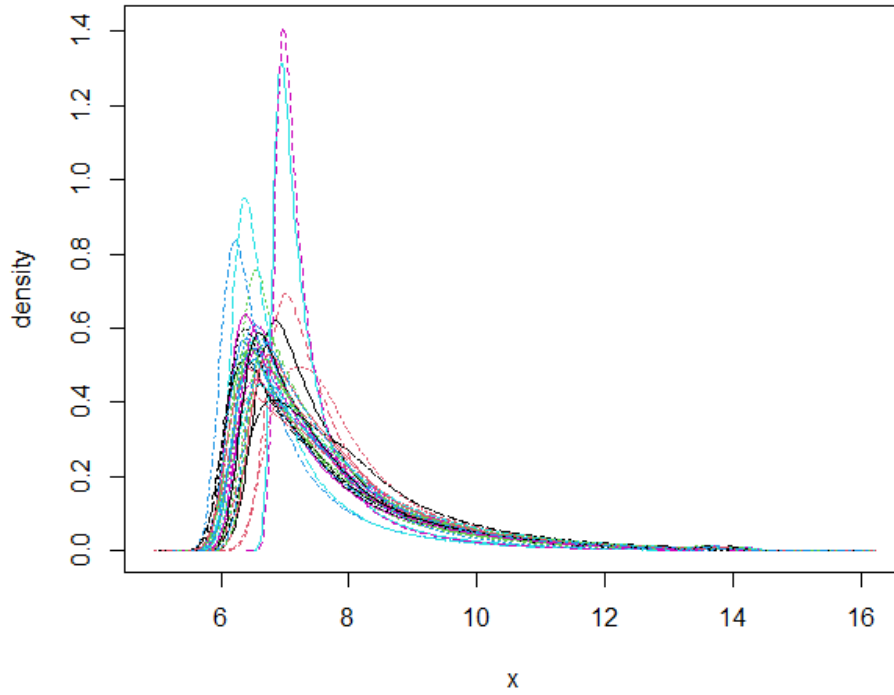
3.6 Έλεγχος επεξεργασίας (Quality control)

Ένα αδιαμφισβήτητο σημαντικό στάδιο της μελέτης που πραγματοποιήθηκε μετά τον αλγόριθμο RMA είναι ο ποιοτικός έλεγχος των δεδομένων για να είμαστε σε θέση να γνωρίζουμε τις αλλαγές που έγιναν στην διαδικασία της προεπεξεργασίας, να συγκρίνουμε το πριν και το μετά και να μπορούμε να δούμε γραφικές απεικονίσεις των δεδομένων μας. Στα πλαίσια αυτής της εργασίας ο έλεγχος έγινε με τέσσερα διαφορετικά διαγράμματα: το διάγραμμα πυκνότητας (density plot), το θηκόγραμμα (box plot), το διάγραμμα σχετικής λογαριθμικής έκφρασης (Relative Log Expression - RLE plot) , και το διάγραμμα MA (M (log ratio) – A (mean average) scales plot) με τη βοήθεια των αντίστοιχων εντολών και συναρτήσεων που παρατίθενται παρακάτω.

3.6.1 Διάγραμμα πυκνότητας (Density Plot)

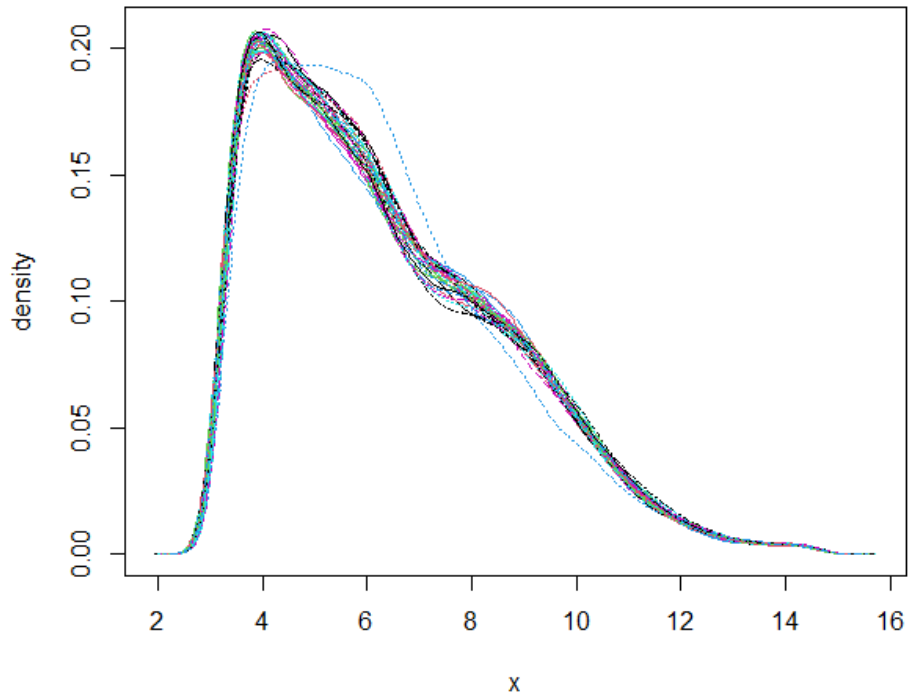
Τα διαγράμματα πυκνότητας αποτελούν παραλλαγή ιστογραμμάτων που έχουν υποστεί εξομάλυνση. Η ομαλότητα της τελικής καμπύλης ελέγχεται από μια παράμετρο, το εύρος της ζώνης που είναι ανάλογη με το εύρος της κάθε μπάρας στο ιστόγραμμα, με τις κορυφές των καμπύλων να συμβάλλουν στο να γίνει πιο εμφανές το διάστημα στο οποίο εμφανίζονται οι περισσότερες τιμές. Στην πράξη κάθε καμπύλη απεικονίζει τα δεδομένα ενός δείγματος και υποδεικνύει πόσο συχνά εμφανίζεται μια συγκεκριμένη ένταση. Στις παρακάτω εικόνες παρατηρούμε τη διαφορά των διαγραμμάτων πυκνότητας πριν και μετά την εφαρμογή του αλγορίθμου RMA που είδαμε αναλυτικά προωτέρω.

Before normalization (log2-transformed) - Probe level



Εικόνα 3: Density plot before normalization

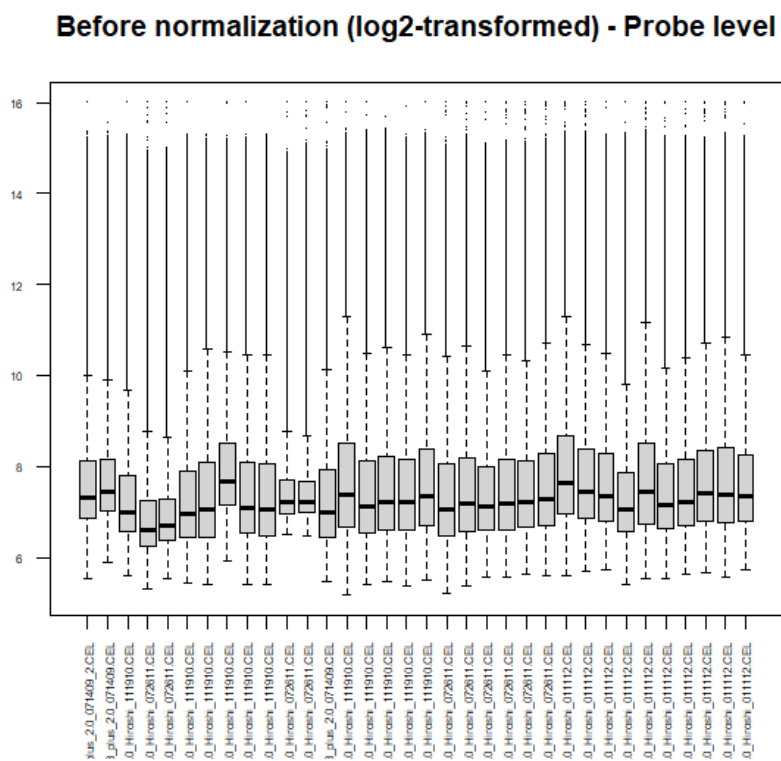
RMA-normalized - Probeset level



Εικόνα 4 : Density plot RMA Normalized

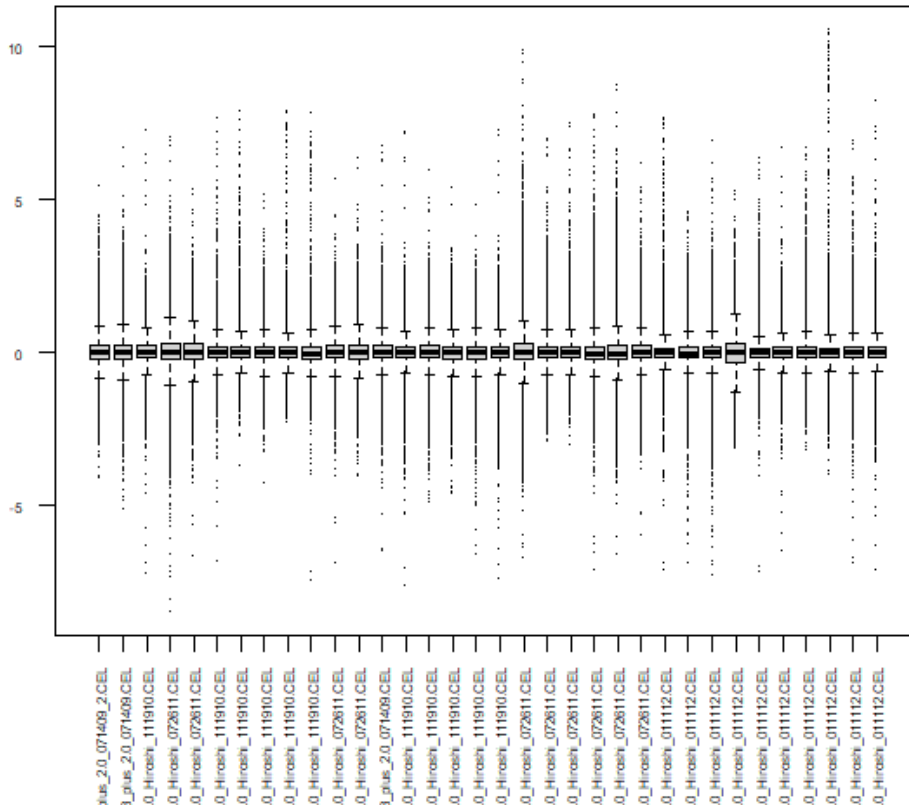
3.6.2 Θηκόγραμμα (Box Plot)

Το θηκόγραμμα προσφέρει μια κατανοητή εικόνα για την κατανομή των τιμών έκφρασης των γονιδίων σε μια μικροσυστοιχία και χρησιμοποιείται κυρίως για να μπορούμε με μια γρήγορη ματιά να συγκρίνουμε τις κατανομές των εντάσεων μεταξύ των δειγμάτων. Για αυτό και με τα θηκογράμματα είναι δυνατό να ελεγχθεί εάν έχει γίνει σωστά η διόρθωση θορύβου υποβάθρου και η κανονικοποίηση. Το κάθε ορθογώνιο παραλληλόγραμμα «κουτάκι» του θηκογράμματος αντιπροσωπεύει την κατανομή των τιμών έντασης φθορισμού της κάθε μικροσυστοιχίας, δηλαδή του κάθε πλακιδίου, το οποίο περιέχει το 50% των τιμών που βρίσκονται γύρω από τη διάμεσο (2ο κα 3ο τεταρτημόριο) και απεικονίζεται στο διάγραμμα μια μία οριζόντια γραμμή μέσα στο κάθε «κουτάκι». Το άνω άκρο του ορθογωνίου και πάνω αντιπροσωπεύει το 75% των τιμών και πάνω (4ο τεταρτημόριο) μέχρι την οριζόντια γραμμή που αντιπροσωπεύει τη μεγαλύτερη τιμή έντασης της κατανομής. Αντιστοίχως, το κάτω άκρο του ορθογωνίου οριοθετεί το 25% των τιμών και κάτω από αυτό (1ο τεταρτημόριο) μέχρι την οριζόντια γραμμή που αντιπροσωπεύει τη μικρότερη τιμή έντασης. Στις παρακάτω εικόνες βλέπουμε τη διαφορά των θηκογραμμάτων πριν και μετά την εφαρμογή του αλγορίθμου RMA.



Εικόνα 5 Box plot before Normalization

RLE plot

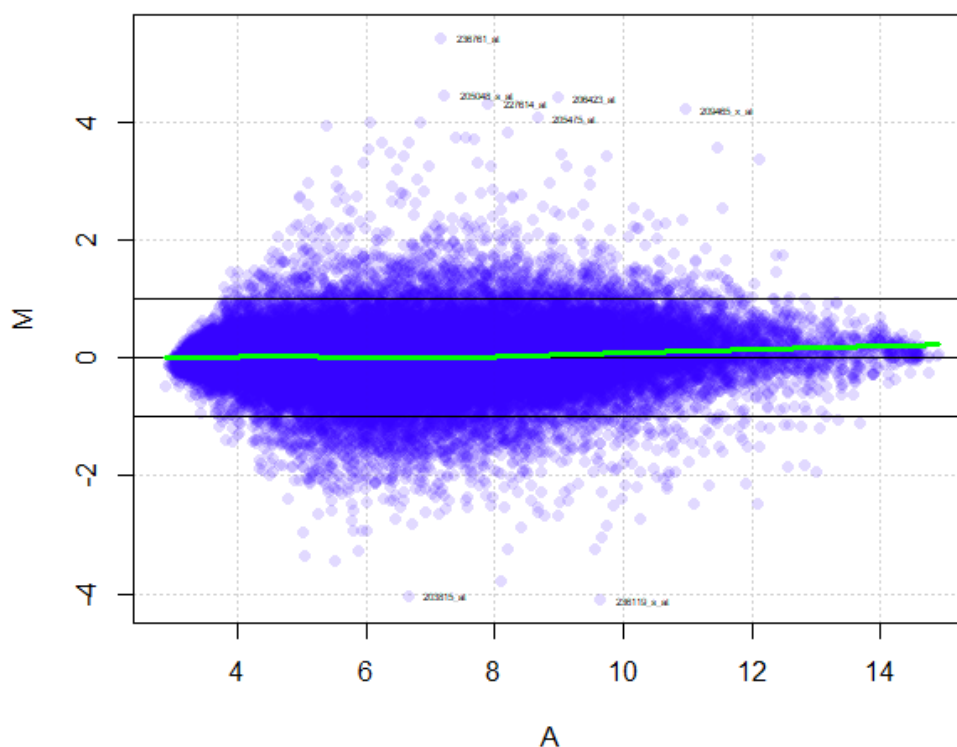


Εικόνα 7: RLE plot

3.6.4 MA Plot

Το MA plot περιέχει πληροφορία αντίστοιχη με το RLE plot. Ξεκινώντας χρειάστηκε να υπολογιστούν δύο ποσότητες, η A και η M. Η ποσότητα A αντιστοιχεί στον μέσο όρο \log_2 μετασηματισμένης έντασης του κάθε probe set και της αντίστοιχης διάμεσης έντασης του probe set της μικροσυστοιχίας αναφοράς, ενώ η ποσότητα M αφορά τη διαφορά τους. Στόχος είναι οι δύο αυτές ποσότητες στο γράφημα να είναι ανεξάρτητες διότι έτσι θα ξέρουμε πως η διαδικασία του RMA ήταν επιτυχής. Επιπλέον θα πρέπει το νέφος των σημειακών δεδομένων (data points) που απεικονίζεται στο γράφημα να είναι κεντραρισμένο γύρω από το 0 ($M=0$), κάνοντας την υπόθεση ότι η πλειοψηφία των γονιδίων δεν έχει διαφορές στα επίπεδα έκφρασης και πως ο αριθμός των υπερεκφρασμένων και υποεκφρασμένων γονιδίων είναι περίπου ίσος. Τέλος στο γράφημα έχει προστεθεί η πράσινη γραμμή τοπικής παλινδρόμησης lowess (locally weighted scatter plot smoothing) για επιπλέον οπτική επιβεβαίωση της «κατανομής» του νέφους γύρω από το 0 και επιπροσθέτως έχουν

σημανθεί ονομαστικά τα probe sets με ιδιαίτερα μεγάλη διαφορά εντάσεων σε σχέση με την μικροσυστοιχία αναφοράς ($M > 4$).



Εικόνα 8: MA plot

3.7 Φιλτράρισμα (Filtering panp)

Τελευταίο και πολύ σημαντικό βήμα της προεπεξεργασίας των δεδομένων μικροσυστοιχιών είναι το φιλτράρισμα που αποσκοπεί στην αφαίρεση αυτών των σημάτων υβριδοποίησης που θεωρούνται λιγότερο αξιόπιστα. Οι μικροσυστοιχίες μπορούν να μετρήσουν ταυτόχρονα την έκφραση των γονιδίων ολόκληρου του γονιδιώματος. Σε ένα δεδομένο πείραμα όμως, είναι βέβαιο ότι δεν αλλάζει η έκφραση των γονιδίων ολόκληρου του γονιδιώματος. Αρχικά, δεν αναμένεται να εκφράζονται όλα τα γονίδια. Επίσης, μεταξύ των εκφρασμένων γονιδίων, ένα μικρό μέρος αναμένεται να εκφραστεί διαφορετικά από τις διαφορετικές συνθήκες που χρησιμοποιούνται στο πείραμα (Calza et al., 2007). Το πρόβλημα έγκειται στο ότι είναι πιθανό να υπάρχουν γονίδια των οποίων τα επίπεδα έκφρασης παρόλο που διατηρούνται σταθερά από πείραμα σε πείραμα μπορεί να εμφανίζουν διακύμανση

λόγω θορύβου. Στην περίπτωση που παραβλέψουμε το βήμα του φιλτραρίσματος και αυτά τα γονίδια δεν απομακρυνθούν από την ανάλυσή μας ο θόρυβος αυτός μπορεί να πολλαπλασιαστεί στη συνέχεια της επεξεργασίας και λανθασμένα να ερμηνευτεί ως πραγματικό σήμα. Υπάρχουν διαφορετικές μέθοδοι φιλτραρίσματος που είτε έχουν σχεδιαστεί για να εκτιμούν δεδομένα χαμηλής ποιότητας, είτε η χρήση τους είναι να προσδιορίσουν διαφορετικά εκφρασμένα γονίδια ή ακόμα και για να ενισχύσουν ένα συγκεκριμένο χαρακτηριστικό των δεδομένων. Για να καταλήξει κανείς στην κατάλληλη μέθοδο που θα χρειαστεί για την ανάλυσή του θα πρέπει να δίνεται προσοχή στην ισορροπία ανάμεσα στην ποιότητα και την ποσότητα των δεδομένων του έτσι ώστε να μην διαταράσσεται. Με τη διαδικασία του φιλτραρίσματος αυξάνεται σε μεγάλο βαθμό η ευαισθησία της ανάλυσης μας, καθώς παράλληλα μειώνεται ο μεγάλος όγκος των δεδομένων γεγονός που βοηθάει στην πραγματοποίηση των στατιστικών ελέγχων σε πολλά επίπεδα. Για παράδειγμα με μικρότερο όγκο υπολογιστικών δεδομένων βελτιώνεται το πρόβλημα των πολλαπλών ελέγχων κι επίσης μειώνεται η ανάγκη για μεγάλη υπολογιστική δύναμη που σε άλλη περίπτωση θα ήταν απαραίτητη. Ωστόσο είναι καίριας σημασίας το φιλτράρισμα να γίνεται με ιδιαίτερη προσοχή καθώς δεν θέλουμε να αφαιρεθούν από την ανάλυσή μας πληροφοριακά στοιχεία από τα δεδομένα.

Στην συγκεκριμένη εργασία επιλέξαμε να χρησιμοποιήσουμε το φιλτράρισμα με βάση την ένταση. Η επιλογή αυτή έγινε για να εντοπιστούν και στη συνέχεια να αφαιρεθούν probes ή probe sets τα οποία έχουν χαμηλή τιμή έντασης φθορισμού. Σε αυτό το στάδιο χρησιμοποιήσαμε το στατιστικό πακέτο της R, το PANP (Presence – Absence calls with Negative Probe sets). Η μέθοδος αυτή λειτουργεί χάρη στην χρήση κάποιων probe sets της Affymetrix τα οποία είναι γνωστό ότι δεν συμμετέχουν σε κάποιο πρότυπο (pattern) υβριδοποίησης για αυτό και πήραν το όνομα probe sets κλώνων αρνητικής αντιστοίχισης (Negative Strand Matching Probe Sets - NSMPs). Έτσι έπειτα από τα στάδια της προεπεξεργασίας εφαρμόζοντας την εντολή `pa.calls()` στα δεδομένα μας υπολογίζεται μια «συνάρτηση επιβίωσης» των NSMPs και με βάση αυτήν τα δεδομένα μας διαιρούνται σε 3 κατηγορίες, ενεργά, οριακά και ανενεργά, με τη βοήθεια των αντίστοιχων p-values που δηλώνουν την βεβαιότητα αυτής της ταξινόμησης. Αναλυτικότερα ως ενεργά ονομάζουμε εκείνα τα probe sets τα οποία βρίσκονται κάτω από ένα συγκεκριμένο p-value το οποίο ορίζεται από τον κάθε ερευνητή. Στην δική μας περίπτωση επιλέξαμε αυτό το όριο να είναι στο 0,01. Επιπλέον ως τελικό κριτήριο - φίλτρο θέσαμε την επιλογή όσων probe sets

προέκυψαν μη ανενεργά σε ποσοστό ίσο ή μεγαλύτερο του 50% σε κάποια κατηγορία (ασθενών και υγιών) των δειγμάτων μας. Για παράδειγμα, για να περάσει το στάδιο φιλτραρίσματος ένα probe set του δείγματός μας θα πρέπει να έχει χαρακτηριστεί ως μη ανενεργό όχι απλά σε 18 από τα 34 δείγματα, αλλά σε 10 από τους 18 υγιείς ή 9 από τους 16 ασθενείς.

3.8 Στατιστική Ανάλυση

Βασικός στόχος στις μελέτες μικροσυστοιχιών είναι η εύρεση γονιδίων που η έκφρασή τους διαφοροποιείται υπό συγκεκριμένες βιολογικές καταστάσεις. Για αυτό και χρειάζεται να υπολογιστεί η διαφορά στα επίπεδα έκφρασης και να εκτιμηθεί εάν αυτή η παρατηρούμενη διαφορά είναι στατιστικά σημαντική. Οπότε για να επιλεγεί η κατάλληλη στατιστική μέθοδος είναι σημαντικό να λαμβάνονται υπόψιν οι εξαρτήσεις που μπορεί να υπάρχουν μεταξύ των γονιδίων, ο διαφορετικός ρυθμός αλλαγής κάθε γονιδίου ως απόκριση σε ένα ερέθισμα αλλά και το γεγονός ότι σε χαμηλές εντάσεις φθορισμού ενδέχεται να παρατηρείται σημαντική μεταβλητότητα. Το πρόβλημα που μπορεί να προκύψει σε αυτό το στάδιο της ανάλυσης αφορά το λεγόμενο σφάλμα τύπου I, δηλαδή την ύπαρξη μεγάλου αριθμού ψευδώς θετικών αποτελεσμάτων που ενδέχεται να προκύψουν κατά τη διαδικασία ανίχνευσης διαφορικά εκφρασμένων γονιδίων εξ' αιτίας των πολλαπλών δοκιμών (σε ένα σχετικά μικρό αριθμό δειγμάτων, διεξάγεται αριθμός δοκιμών ίσος με τον αριθμό των γονιδίων). Συνεπώς για να αποφευχθεί το παραπάνω θα πρέπει να γίνει διόρθωση του στατιστικού αποτελέσματος και να επιλεγεί η κατάλληλη μέθοδος τέτοια ώστε να αυξάνει την αξιοπιστία της ανάλυσής μας και παράλληλα να μειώνει με προσοχή τα ψευδώς θετικά αποτελέσματα γιατί διαφορετικά είναι πιθανό να οδηγηθούμε σε αύξηση του σφάλματος τύπου II, δηλαδή των ψευδώς αρνητικών αποτελεσμάτων, με τραγική συνέπεια την απώλεια χρήσιμης πληροφορίας από τα δεδομένα του δείγματός μας.

3.8.1 False Discovery Rate

Όπως περιεγράφηκε και παραπάνω το βασικό πρόβλημα στην στατιστική ανάλυση των μικροσυστοιχιών είναι οι πολλαπλοί έλεγχοι και αυτό κυρίως οφείλεται στην επιλογή του επιπέδου σημαντικότητας το οποίο συνηθίζεται να είναι με όριο την τιμή

$p\text{-value} \leq 0.05$ για να είμαστε αρκετά σίγουροι ότι τα εξαγόμενα συμπεράσματα είναι και τα σωστά. Στην πράξη αυτό σημαίνει πως υπάρχει μια πιθανότητα της τάξεως του 5% ψευδών θετικών αποτελεσμάτων κατά τον έλεγχο διαφορικής έκφρασης ενός γονιδίου το οποίο είναι πολύ καλό μιας και μας δίνει και 95% διάστημα εμπιστοσύνης. Όμως στα πειράματα μικροσυστοιχιών όπου ο αριθμός των δοκιμών είναι ίσος με τον αριθμό των γονιδίων, πράγμα που σημαίνει ότι είναι πάρα πολύ μεγάλος, το διάστημα εμπιστοσύνης που δίνεται δεν είναι αρκετά ικανοποιητικό. Μιλώντας με αριθμούς εάν είχαμε ένα δείγμα με 10.000 γονίδια το 5% των ψευδώς θετικών αποτελεσμάτων μεταφράζεται σε 500 γονίδια τα οποία θα είχαν βρεθεί διαφορετικά εκφρασμένα πιθανότατα από λάθος. Αυτό ανέπτυξε την ανάγκη για την εύρεση μεθόδων που λειτουργούν επεμβατικά διορθώνοντας την τιμή $p\text{-value}$. Για τον σκοπό αυτόν στα βιολογικά πειράματα χρησιμοποιούνται κυρίως μέθοδοι γνωστοί ως FDR (False Discovery Rate) οι οποίες μάλιστα θεωρούνται λιγότερο αυστηρές και σκοπός τους είναι η ελαχιστοποίηση του σφάλματος τύπου I ως προς το σύνολο των απορρίψεων της μηδενικής απόρριψης.

Είναι πολύ σημαντικό να λαμβάνουμε υπόψιν μας ότι η διαδικασία που περιεγράφηκε είναι μια καθαρά στατιστική διαδικασία και δεν λαμβάνει υπόψιν βιολογικές παραμέτρους που αφορούν την ανεξαρτησία των γονιδίων αλλά και την παραδοχή ότι τα γονίδια είναι πιθανό να επηρεάσουν το ένα το άλλο με τρόπους που προς το παρόν δεν γνωρίζουμε. . Το ζήτημα της μη ανεξαρτησίας των γονιδίων αντιμετωπίστηκε μέσω του συγκεκριμένου εργαλείου ανάλυσης που επιλέξαμε στο περιβάλλον του RStudio, του πακέτου limma.

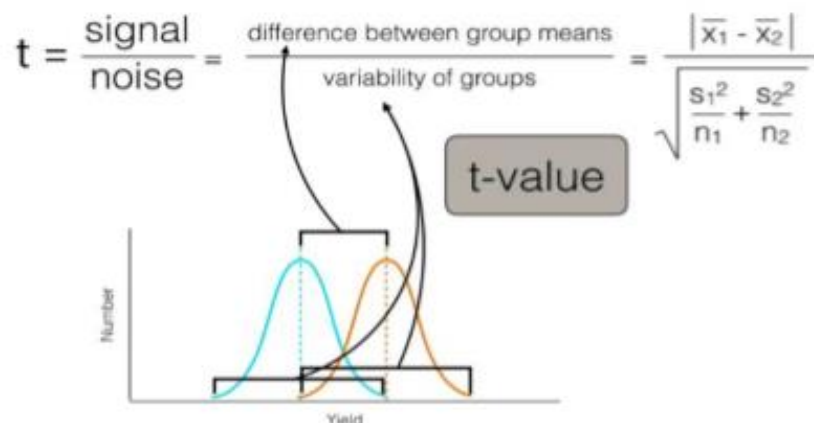
3.8.2 Στατιστικός έλεγχος

Ο στατιστικός έλεγχος που χρησιμοποιήθηκε στην παρούσα εργασία είναι ο Student (Student t-test). Για να εξεταστεί η διαφορική έκφραση των γονιδίων αρχικά ορίστηκε μια μηδενική υπόθεση (null hypothesis) για το κάθε γονίδιο της μικροσυστοιχίας. Αυτή η μηδενική υπόθεση μας λέει πως το εκάστοτε γονίδιο δεν παρουσιάζει διαφορική έκφραση και δοκιμάστηκε έναντι της υπόθεσης ελέγχου η οποία μας λέει πως υπάρχει διαφορά στην έκφραση του γονιδίου ανάμεσα στις διαφορετικές καταστάσεις. Σύμφωνα με τον στατιστικό έλεγχο Student χρησιμοποιήσαμε ένα στατιστικό μέτρο t το οποίο αποτέλεσε κριτήριο ελέγχου για τον λόγο σήματος – θορύβου για κάθε γονίδιο, συγκρίνοντας τα επίπεδα έκφρασής του για τις δύο

διαφορετικές συνθήκες που μελετώνται. Γενικά βασική προϋπόθεση για την εφαρμογή του t-test είναι να υπάρχει κανονική κατανομή των τιμών των εντάσεων, να μη διαφέρει πολύ η διακύμανση, και παράλληλα να υπάρχει ο ίδιος αριθμός μετρήσεων σε κάθε κατάσταση.

Αναλυτικά θεωρήσαμε δύο καταστάσεις 1 και 2 με μέση τιμή \bar{x}_1 και \bar{x}_2 αντίστοιχα. Επιπλέον βρέθηκαν οι τιμές της τυπικής απόκλισης s_1 και s_2 για κάθε δείγμα, οι οποίες υψωμένες στο τετράγωνο μας δίνουν τη διακύμανση για την κάθε κατάσταση και υπολογίστηκε ο τύπος της παρακάτω φωτογραφίας με τις τιμές n_1 και n_2 να αναφέρονται στον αριθμό των παρατηρήσεων.

Ωστόσο σε πειράματα μικροσυστοιχιών ο αριθμός των δειγμάτων είναι μικρός, τα αποτελέσματα του απλού t-test δεν είναι τόσο αξιόπιστα (Jeanmougin et al., 2010). Συνεπώς κρίνεται απαραίτητο να χρησιμοποιηθούν παραλλαγές του κλασσικού t-test όπως το «ελεγχόμενο» t-test (moderated t-statistics) που βελτιώνουν τα αποτελέσματα των διαφορικά εκφρασμένων γονιδίων και ενσωματώνουν εμπειρικά-μπεϋζιανά μοντέλα (empirical Bayes models). Με αυτό τον τρόπο το τυπικό σφάλμα προσαρμόζεται για το κάθε γονίδιο συνδυάζοντας την πληροφορία μετρήσεων από το εκάστοτε γονίδιο και το σύνολο όλων των υπόλοιπων γονιδίων. Επιτυγχάνεται έτσι αύξηση των βαθμών ελευθερίας. Οι βαθμοί ελευθερίας που προστίθενται αντανακλούν την επιπλέον πληροφορία που υπάρχει ανάμεσα στα γονίδια, η οποία χρησιμοποιείται για να εξαχθεί το αποτέλεσμα για το κάθε ένα από αυτά (Baldi, Long, 2001), (Fox, Dimmic, 2006). Στη συνέχεια θα γίνει ο υπολογισμός του p-value που θα υποδείξει εάν οι διαφορές που παρατηρούμε είναι τελικά στατιστικά σημαντικές ή όχι. Η τιμή του p-value που συνήθως ορίζεται στο 0.05 ($p \leq 0.05$) είναι η πιθανότητα να παρατηρούμε ένα τέτοιο αποτέλεσμα στο t-test από τύχη, δηλαδή να ισχύει η μηδενική υπόθεση και άρα να μην υπάρχει στατιστικά σημαντική διαφορά. Όσο μικρότερη η τιμή του p-value τόσο μικρότερη αυτή πιθανότητα και άρα πιο σημαντική η διαφορά που παρατηρούμε.



Εικόνα 9: Student t-test

3.8.3 Limma (Linear Models for Microarray Data)

Με βάση τα συμπεράσματα της προηγούμενης ενότητας χρειάστηκε τελικώς για την πραγματοποίηση του στατιστικού ελέγχου η χρήση του πακέτου limma στο περιβάλλον του RStudio. Το πακέτο limma είναι ένα ισχυρό εργαλείο ανάλυσης γονιδιακής έκφρασης μικροσυστοιχιών το οποίο χρησιμοποιείται επιπλέον και για την ανάλυση και άλλων τύπων ή άλλης τεχνολογίας δεδομένων, σε μεγάλη ποικιλία πειραμάτων και σχεδιασμών. Το εργαλείο αυτό προσδιορίζει τη διαφοροποίηση στην γονιδιακή έκφραση με τη χρήση γραμμικών μοντέλων και τη χρήση εμπειρικής μπεϋζιανής στατιστικής για εξαγωγή ισχυρών συμπερασμάτων ακόμη και για μικρά δείγματα.

3.9 Αντιστοίχιση των διαφορικά εκφρασμένων μεταγράφων σε γονιδιακά προϊόντα

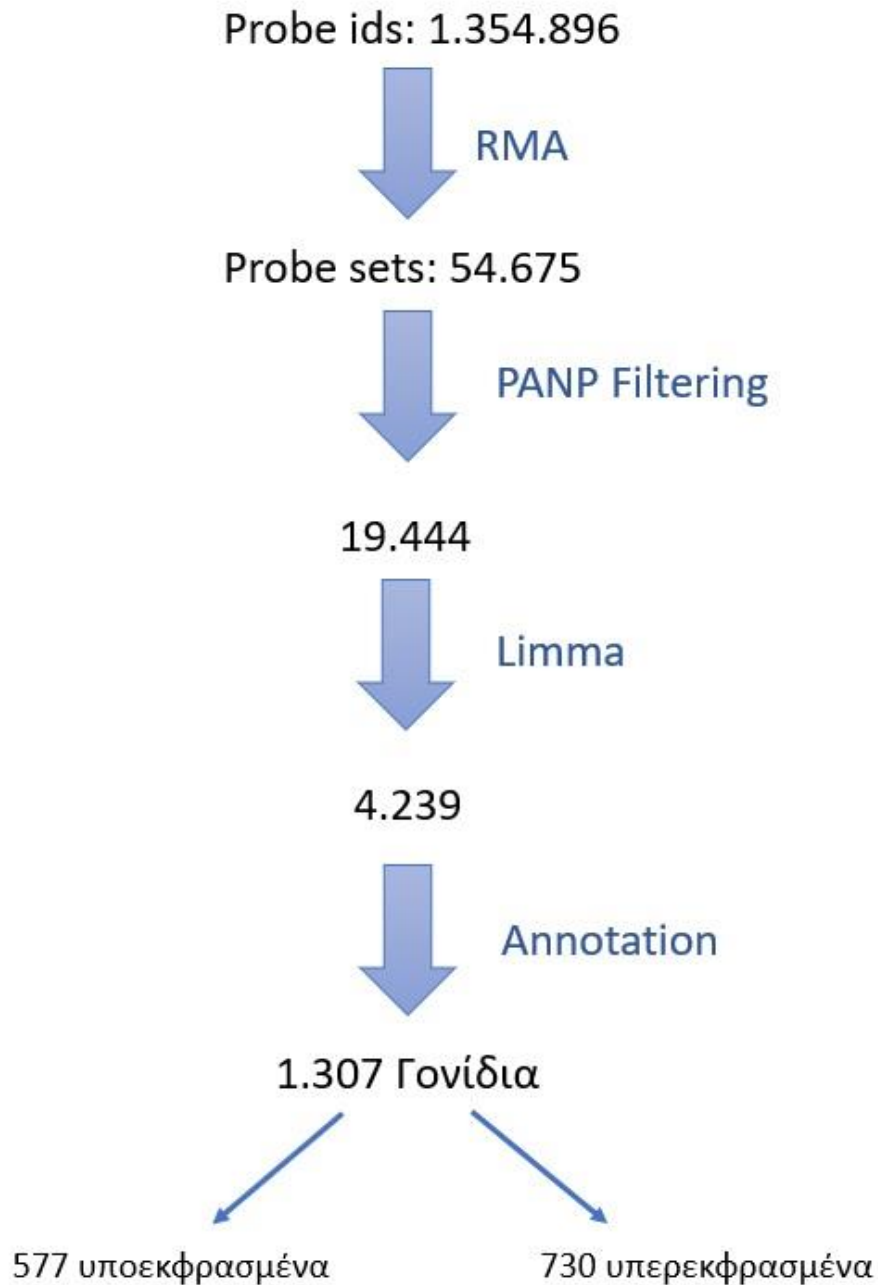
Οι προαναφερθείσες διαδικασίες πραγματοποιήθηκαν με τη μέγιστη δυνατή ακρίβεια για να φτάσουμε σε αυτό το τελευταίο βήμα της ανάλυσης μας, στο οποίο θα καταλήξουμε σε μια πλήρη λίστα των διαφορικά εκφρασμένων γονιδίων. Δηλαδή στην αντιστοίχιση των κωδικών του κάθε μετάγραφου (probe set) που προέκυψε στατιστικά σημαντικά διαφοροποιημένο, με το αντίστοιχο γονιδιακό του προϊόν. Στην πράξη σε αυτό το στάδιο έγινε ο σχολιασμός (annotation) των μεταγράφων που είναι διαφορικά εκφρασμένα με τα ονόματα των γονιδίων στα οποία αντιστοιχούν και τους αναγνωριστικούς κωδικούς τους σε άλλες βάσεις. Επομένως στο RStudio με τη

χρήση αρχείων που παρέχονται από την ίδια την εταιρεία Affymetrix και περιέχουν όλες τις απαραίτητες πληροφορίες για όλα τα μετάγραφα της πλατφόρμας, μέσω κατάλληλων βιβλιοθηκών στην R (hgu133plus2.db , hgu133a.db) και με τη βοήθεια του προγράμματος Excel, εξαγάγαμε μια πλήρη λίστα (με ονόματα, σύμβολα και κωδικούς σε άλλες βάσεις) των διαφορεικά εκφρασμένων γονιδίων.

Γνωρίζουμε ότι στα πειράματα μικροσυστοιχιών εξετάζεται η έκφραση των μεταγράφων ολόκληρου του γονιδιώματος, πράγμα που σημαίνει ότι κάποια από τα μετάγραφα μπορεί να μην έχουν αντιστοιχίση σε συγκεκριμένο ή γνωστό γονίδιο, για αυτό και τα μετάγραφα αυτά δεν μπορούν να προχωρήσουν στο επόμενο στάδιο λειτουργικής ανάλυσης οπότε αφαιρούνται από τη μελέτη. Επιπλέον υπάρχουν περισσότερα από ένα μετάγραφα τα οποία αντιστοιχίζονται στο ίδιο γονίδιο αλλά και μετάγραφα που αντιστοιχίζονται ταυτόχρονα σε δύο διαφορετικά γονίδια. Ο τρόπος που λειτουργούμε σε τέτοιες περιπτώσεις είναι επιλέγοντας ποιο μετάγραφο θέλουμε να κρατήσουμε στην έρευνά μας. Συνεπώς καταλήγουμε στο ότι η λίστα γονιδίων στην οποία οδηγηθήκαμε είναι φανερά μικρότερη από την αρχική λίστα των διαφορεικά εκφρασμένων probe sets που προήλθε από την στατιστική ανάλυση.

3.10 Αποτελέσματα

Στην εικόνα παρουσιάζεται η επίδραση που είχε το κάθε βήμα που αναλύθηκε στο κεφάλαιο αυτό και φαίνεται αναλυτικά ο αριθμός των γονιδίων όπως τροποποιήθηκε έπειτα από κάθε διαδικασία επεξεργασίας. Στο τέλος καταλήγουμε στο πλήθος των διαφορεικά εκφρασμένων γονιδίων για το δείγμα που μελετήθηκε στην παρούσα εργασία. Σημειώνεται πως τα κριτήρια επιλογής ήταν $|\log_{2}fc| > 0.75$ και $p\text{-value} < 0.05$.

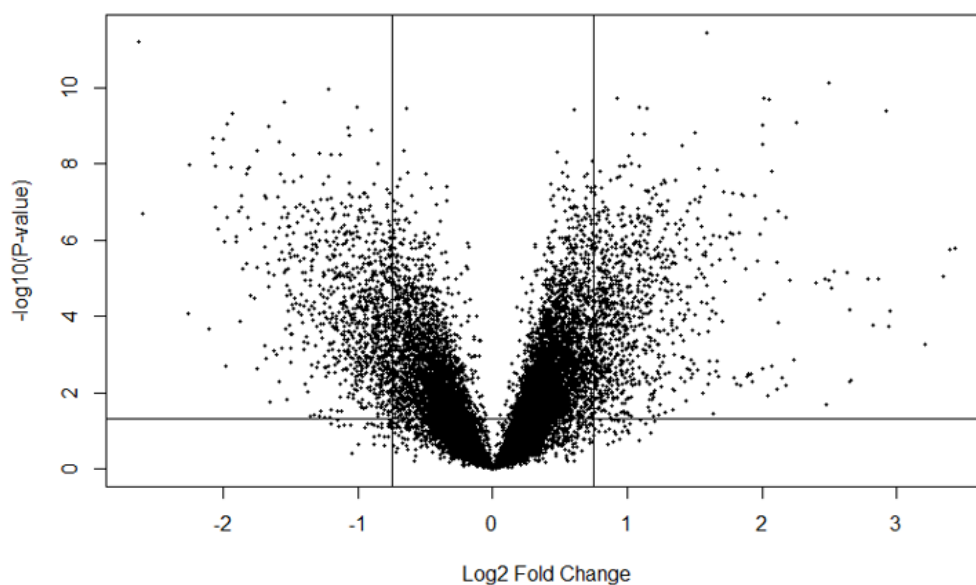


Εικόνα 10: Βήματα Ανάλυσης

3.10.1 Volcano Plots για την απεικόνιση των αποτελεσμάτων

Τα Volcano plots, διαγράμματα ηφαιστείου στα ελληνικά, είναι διαγράμματα τα οποία απεικονίζουν σημεία που στην περίπτωση μας κάθε ένα τέτοιο σημείο αντιστοιχεί σε ένα γονίδιο. Ο άξονας Y του διαγράμματος αντιστοιχεί στην αρνητική τιμή του λογαριθμικά μετασχηματισμένου (με βάση το δεκαδικό λογάριθμο) p-value και στον άξονα X αντιπροσωπεύονται οι τιμές logFC. Συνεπώς παρατηρούμε πως όσο μεγαλύτερη είναι η τιμή στον άξονα Y του γονιδίου τόσο περισσότερο στατιστικά

σημαντική είναι η μεταβολή της έκφρασής του. Ενώ παράλληλα στον άξονα X όσο μεγαλύτερη η απόλυτη τιμή logFC δηλαδή όσο περισσότερο απέχει από το 0 τόσο μεγαλύτερη η διαφορά στην έκφραση. Στην πράξη ο συνδυασμός αυτών σημαίνει πως οι κουκίδες οι οποίες στο γράφημα είναι ψηλά αριστερά και δεξιά αντιπροσωπεύουν γονίδια που είναι περισσότερο στατιστικά σημαντικά και έχουν μεγαλύτερη διαφορά στην έκφραση.



Εικόνα 11: Volcano plot

3.10.2 Πίνακες Γονιδίων

Τελικώς στους πίνακες παρακάτω παρουσιάζονται τα 20 πιο έντονα υπερεκφρασμένα και υποεκφρασμένα γονίδια από τα 730 που προέκυψαν με στατιστικά σημαντική διαφοροποίηση στην έκφραση, έπειτα από την ανάλυση που έγινε με το πακέτο limma και με βάση τα κριτήρια επιλογής που θέσαμε. Συνεπώς βλέπουμε τα 20 πρώτα με βάση το logfc και το αντίστοιχο διορθωμένο p-value.

Πίνακας 2: Top20 υπερεκφρασμένα γονίδια

Σύμβολο γονιδίου	Logfc	Adj p-value
PRAME	3.434385469	0.0000658
S100A9	3.391545625	0.0000691
LINC00518	3.346700188	0.000206396
SERPINB4	3.211808258	0.003798586
SPP1	2.955545994	0.00091625
MMP1	2.937319186	0.001786636

CCL8	2.919698698	0.000000557
CXCL9	2.866500142	0.000226928
S100A8	2.829846607	0.001635935
ADAMDEC1	2.785992111	0.000232345
S100A7	2.665982147	0.019170114
IGHG1	2.651378597	0.020049196
GDF15	2.649092762	0.000856431
CXCL10	2.633029102	0.000173735
PI3	2.531926945	0.000165207
RGS1	2.500439142	0.000248568
CCL3	2.498254768	0.000000475
CHIT1	2.468184492	0.000228019
SAMSN1	2.258512682	0.000001
EGFL6	2.233728627	0.00759874

Πίνακας 3: Top20 υποεκφρασμένα γονίδια

Σύμβολο γονιδίου	Logfc	Adj p-value
LAMB4	-2.626401487	6.18E-08
CCL27	-2.599391681	0.0000207
WIF1	-2.261143465	0.00098796
BTC	-2.111232018	0.001992924
RORA	-2.08130672	0.00000155
CDHR1	-2.07614454	0.00000292
CHP2	-2.06061789	0.0000168
POU2F3	-2.039121955	0.000035
EXPH5	-1.999032245	0.00000162
SCGB1D2	-1.982329753	0.010110312
CNTN1	-1.974667777	0.0000228
GATA3	-1.94617155	0.0000047
ZNF273	-1.934955968	0.000000627
CLDN8	-1.900923738	0.0000441
DSC3	-1.884513743	0.0000184
PATJ	-1.825574633	0.00000578
ADGRF4	-1.825317137	0.000122909
CHL1	-1.770747424	0.000513202
KRT31	-1.749594033	0.011402465
SMPDL3A	-1.746276627	0.00000275

Κεφάλαιο 4^ο : Υλοποίηση Μητρώου Ασθενών για Μελάνωμα

4.1 Γενικά οι Βάσεις Δεδομένων

Με τον όρο βάση δεδομένων εννοείται μια συλλογή από συστηματικά μορφοποιημένα συσχετιζόμενα δεδομένα στα οποία είναι δυνατή η ανάκτηση δεδομένων μέσω αναζήτησης κατ' απαίτηση, σύμφωνα με τον ορισμό της Wikipedia. Τα δεδομένα αυτά μέσω επεξεργασίας από κάποιους προκαθορισμένους κανόνες γίνονται πληροφορία και η πληροφορία είναι πολύτιμη. Για αυτό η σωστή συλλογή, αποθήκευση, ταξινόμηση και διαχείριση αυτών σε βάσεις δεδομένων κρίνεται απαραίτητη ειδικά στην σημερινή εποχή όπου ο όγκος των δεδομένων και κατ' επέκταση της πληροφορίας είναι τεράστιος. Ωστόσο η ανάγκη για ηλεκτρονική καταχώρηση και συστηματική οργάνωση των πληροφοριών αυξήθηκε κατακόρυφα ήδη από τη δεκαετία του 1970 στις μεγάλες επιχειρήσεις και οδήγησε πολύ γρήγορα στην ανάπτυξη Τραπεζών Πληροφοριών ή Βάσεων Δεδομένων (Data Bases). Στην σημερινή εξέλιξή τους οι βάσεις δεδομένων αποτελούνται από σύνολα αρχείων που έχουν υψηλό βαθμό οργάνωσης και συνδέονται μεταξύ τους με λογικές σχέσεις. Επιπλέον αυτά χρησιμοποιούνται ταυτόχρονα από πολλές εφαρμογές και πολλούς και διαφορετικούς χρήστες και για να γίνει αυτό απαιτείται η μεσολάβηση ειδικού λογισμικού που αποκαλείται Σύστημα Διαχείρισης Βάσης Δεδομένων (ΣΔΒΔ) ή DBMS (Data Base Management System).

Ένα σύστημα Διαχείρισης Βάσεων Δεδομένων αποτελείται στην ουσία από προγράμματα και υπορουτίνες που όπως λέει και το όνομά του αφορούν τη διαχείριση της βάσης δεδομένων, δηλαδή την προσθήκη, τον μετασχηματισμό, τη διαγραφή των πληροφοριών με ελέγχους ασφαλείας κ.α. Αναλυτικότερα ένα ΣΔΒΔ απαρτίζεται από το υλικό (hardware), το λογισμικό (software), τη βάση δεδομένων και τους διαφορετικούς χρήστες ανάλογα το επίπεδο διαχείρισης. Συγκεκριμένα στο υλικό δηλαδή σε σκληρούς δίσκους, ηλεκτρονικούς υπολογιστές κ.α. είναι αποθηκευμένα τα αρχεία της βάσης δεδομένων όπως και τα προγράμματα που χρειάζονται για τη διαχείρισή τους. Το λογισμικό αφορά τα προγράμματα επεξεργασίας των δεδομένων της βάσης. Η βάση δεδομένων αφορά τη συλλογή των αρχείων που είναι αποθηκευμένα τα δεδομένα του συστήματος, τα οποία είναι πιθανό να βρίσκονται σε πολλούς διαφορετικούς υπολογιστές, ωστόσο στον χρήστη παρέχεται η εντύπωση πως αυτά τα δεδομένα είναι ενοποιημένα (data integration) και πως είναι μερισμένα (data sharing). Τέλος οι χρήστες μιας βάσης δεδομένων χωρίζονται στις παρακάτω τέσσερις κατηγορίες: Τελικοί χρήστες, Προγραμματιστές εφαρμογών, Διαχειριστές δεδομένων και Διαχειριστές βάσης δεδομένων. Τα ΣΔΒΔ είναι σημαντικό να οργανωμένα με τρόπο τέτοιο ώστε να διατηρούν την

ανεξαρτησία των δεδομένων τους για να μην έχουμε απώλεια χρήσιμης πληροφορίας κατά την τροποποίηση ή τον μετασχηματισμό του λογισμικού. Για αυτόν τον λόγο δημιουργήθηκε η λεγόμενη αρχιτεκτονική τριών επιπέδων. Το πρώτο επίπεδο ονομάζεται «Εσωτερικό επίπεδο» (internal level) και αφορά ξεκάθαρα την αποθήκευση των αρχείων στο υλικό (με άλλα λόγια την πραγματική ή φυσική κατάσταση τους. Το δεύτερο επίπεδο ονομάζεται «Εξωτερικό επίπεδο» (external level) και αφορά όλες τις κατηγορίες χρηστών που αναφέρθηκαν παραπάνω. Το τρίτο επίπεδο ονομάζεται «Εννοιολογικό επίπεδο» (conceptual level), το οποίο θα μας απασχολήσει στα πλαίσια αυτής της εργασίας και είναι ένα ενδιάμεσο επίπεδο των προηγούμενων δύο που αφορά τη λογική σχεδίαση της βάσης δεδομένων.

4.2 Εννοιολογικό μοντέλο Βάσεων Δεδομένων


Το πρώτο βήμα για τη δημιουργία ενός συστήματος βάσεων δεδομένων είναι η σχεδίαση του, η οποία αποτελείται από κάποια επιμέρους στάδια. Αρχικά χρειάζεται η συλλογή και η ανάλυση των απαιτήσεων αλλά και των λειτουργιών, δεδομένα που θα προκύψουν από τις ανάγκες των τελικών χρηστών. Στο επόμενο στάδιο εμπεριέχεται ο εννοιολογικός σχεδιασμός της βάσης δεδομένων, με τον οποίο και θα ασχοληθούμε στο παρόν κεφάλαιο.

Κατά τον εννοιολογικό σχεδιασμό της βάσης δεδομένων γίνεται η δημιουργία του εννοιολογικού σχήματος χρησιμοποιώντας ένα υψηλού επιπέδου εννοιολογικό μοντέλο. Με αυτόν τον τρόπο καθορίζονται οι δομές που θα χρησιμοποιηθούν για την καταχώρηση των δεδομένων όπως και οι περιορισμοί που έχουν δοθεί από τους τελικούς χρήστες. Για να περιγραφούν αυτές οι δομές και οι περιορισμοί στην παρούσα εργασία χρησιμοποιήθηκε το μοντέλο Οντοτήτων-Συσχετίσεων μέσω του οποίου στοχεύουμε το σχήμα που θα δημιουργηθεί να είναι απλό, κατανοητό αλλά και ακριβές έτσι ώστε να είναι δυνατή η άμεση μετατροπή του σε ένα μοντέλο υλοποίησης.

4.2.1 Το μοντέλο Οντοτήτων-Συσχετίσεων


Το μοντέλο Οντοτήτων-Συσχετίσεων είναι ένα γραφικό μοντέλο και το σχήμα μιας βάσης δεδομένων με τη χρήση αυτού του μοντέλου ονομάζεται και διάγραμμα Οντοτήτων-Συσχετίσεων (ER Diagram). Αναλυτικότερα για τον σχεδιασμό

χρησιμοποιούμε (σημειώνονται μόνο οι συμβολισμοί που χρησιμοποιήθηκαν στο ER Diagram που δημιουργήθηκε στα πλαίσια αυτής της εργασίας):

1. Οντότητες: 

Μια οντότητα αντιστοιχεί σε ένα αντικείμενο του πραγματικού κόσμου και περιγράφεται από ένα όνομα και μια λίστα γνωρισμάτων (ιδιοτήτων).

Επιπλέον υπάρχουν οι ασθενείς ή μη ισχυροί ή οι αδύναμοι (weak) τύποι οντοτήτων οι οποίοι ξεχωρίζουν διότι δεν έχουν αρκετά γνωρίσματα για να σχηματίσουν δικό τους πρωτεύον κλειδί (η έννοια του κλειδιού επεξηγείται παρακάτω).

2. Γνωρίσματα: 

Τα γνωρίσματα είναι ιδιότητες που χαρακτηρίζουν μια οντότητα.


Διακρίνονται σε απλά και σύνθετα (συνένωση απλών γνωρισμάτων).

Επιπλέον υπάρχουν τα μονότιμα (single value) που παίρνουν δηλαδή μία μόνο τιμή και τα πλειότιμα (multi value) που δέχονται ένα σύνολο από τιμές.

Επίσης έχουμε τα παραγόμενα (derived) γνωρίσματα τα οποία είναι δυνατό να υπολογιστούν από άλλα και τα αποθηκευμένα.


Τα γνωρίσματα χαρακτηρίζονται από μία έννοια που ονομάζεται Κλειδί (key).

Το κλειδί είναι ένα σύνολο από γνωρίσματα τέτοια ώστε δεν μπορεί να υπάρχουν δυο οντότητες με την ίδια τιμή σε αυτά, δηλαδή την προσδιορίζουν μοναδικά. Το υποψήφιο κλειδί (candidate key) είναι το ελάχιστο κλειδί, δηλαδή ένα κλειδί από το οποίο εάν αφαιρέσουμε ένα γνώρισμά του, παύει να είναι κλειδί. Τέλος το πρωτεύον κλειδί (primary key) είναι εκείνο το υποψήφιο κλειδί που επιλέγουμε.

3. Συσχετίσεις: 

Αποτελούν τη διασύνδεση μεταξύ δύο ή περισσότερων οντοτήτων. Με τον όρο βαθμός μιας συσχέτισης εννοούμε το πλήθος των οντοτήτων που συμμετέχουν σε αυτήν. Οι συσχέτισεις δύναται να έχουν κι αυτές δικά τους γνωρίσματα. Επιπλέον χαρακτηρίζονται με τον λόγο πληθικότητας, για παράδειγμα στις δυαδικές συσχέτισεις οι λόγοι πληθικότητας που θα συναντήσουμε είναι: «ένα προς ένα 1-1», «ένας προς πολλά 1:N», «πολλά προς ένα N-1», «πολλά προς πολλά N-M». Τέλος ένας ακόμα χαρακτηρισμός που αφορά την συμμετοχή των οντοτήτων προσδιορίζει εάν η συμμετοχή ενός συνόλου οντοτήτων E σε ένα σύνολο συσχετίσεων R είναι ολική στην περίπτωση δηλαδή που κάθε οντότητα του E συμμετέχει τουλάχιστον σε μια

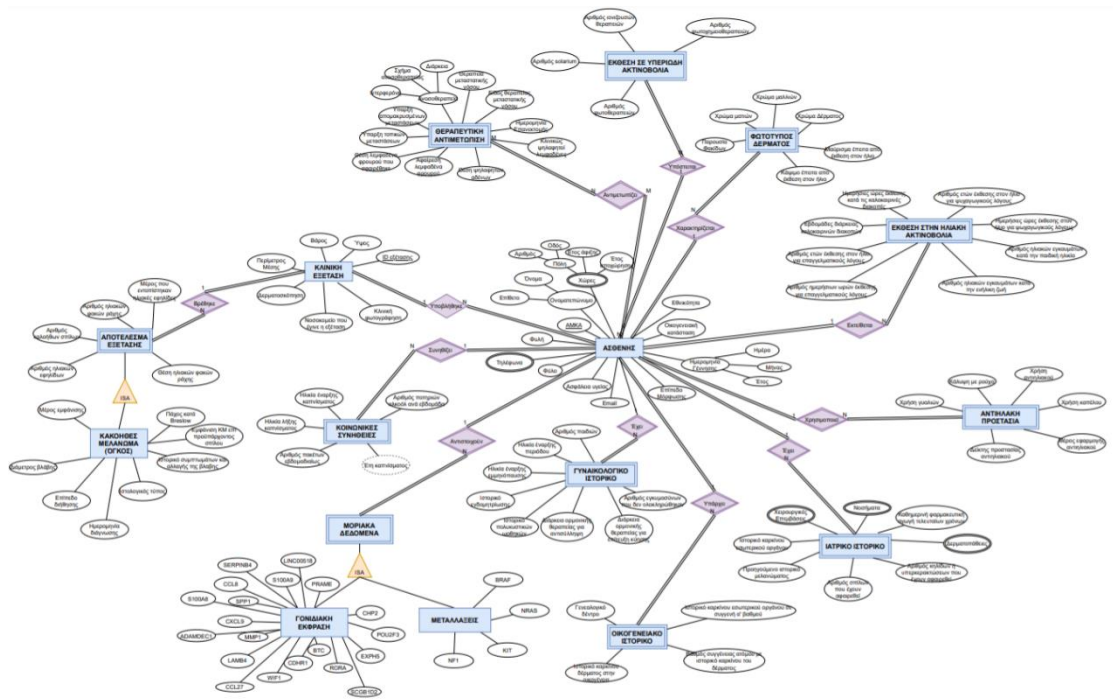
συσχέτιση στο R. Αν κάποιες οντότητες του E δεν συμμετέχουν στο R τότε η συμμετοχή χαρακτηρίζεται ως μερική.

4. Εξειδίκευση ISA: 

Μια οντότητα δύναται να περιλαμβάνει υπό-ομάδες οντοτήτων οι οποίες διακρίνονται από επιπρόσθετα γνωρίσματα. Οι υπό-ομάδες μιας οντότητας κληρονομούν τα γνωρίσματα και τις συσχετίσεις της.

4.2.2 ER Diagram Μητρώου Ασθενών με Μελάνωμα

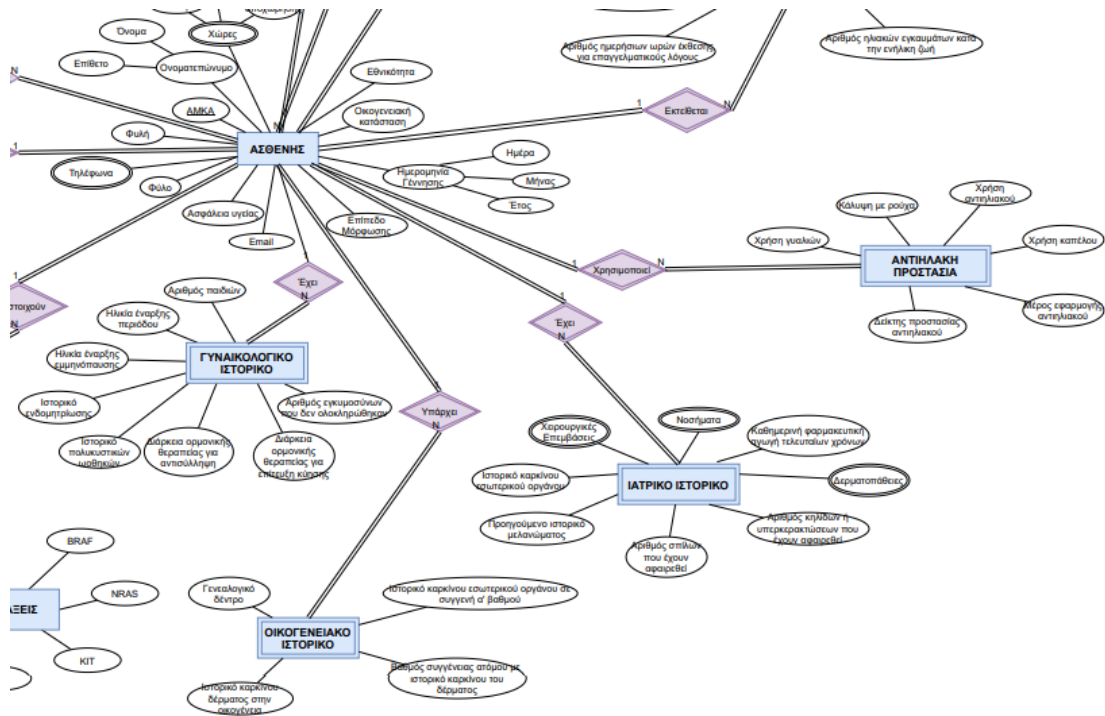
Χρησιμοποιώντας τις γνώσεις και τα εργαλεία που αναλύθηκαν πρωτύτερα δημιουργήθηκε το ER Diagram του μητρώου ασθενών με μελάνωμα που αποτυπώνεται στην παρακάτω φωτογραφία.



Εικόνα 12: ER Diagram Μητρώου Ασθενών με Μελάνωμα

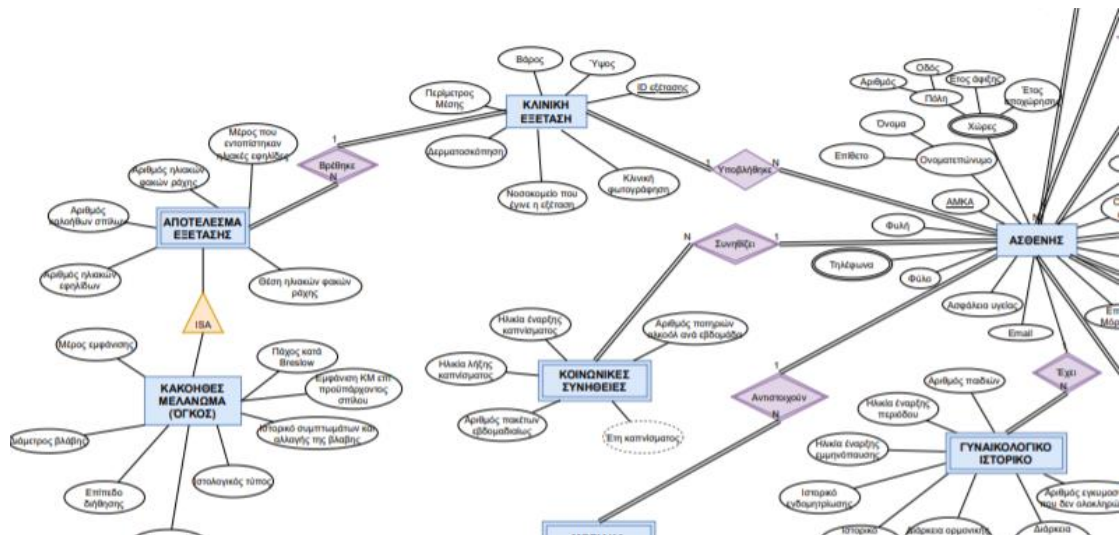
Χωρίζοντας την παραπάνω φωτογραφία σε τέσσερα τμήματα μπορούμε να δούμε καλύτερα τις οντότητες με τα γνωρίσματά τους και τις συσχετίσεις μεταξύ τους.

1^ο τμήμα: Παρουσιάζονται οι συσχετίσεις και τα γνωρίσματα της οντότητας «Ασθενής» με τις οντότητες «Θεραπευτική αντιμετώπιση», «Έκθεση σε υπεριώδη



Εικόνα 14: 2ο τμήμα ER Diagram

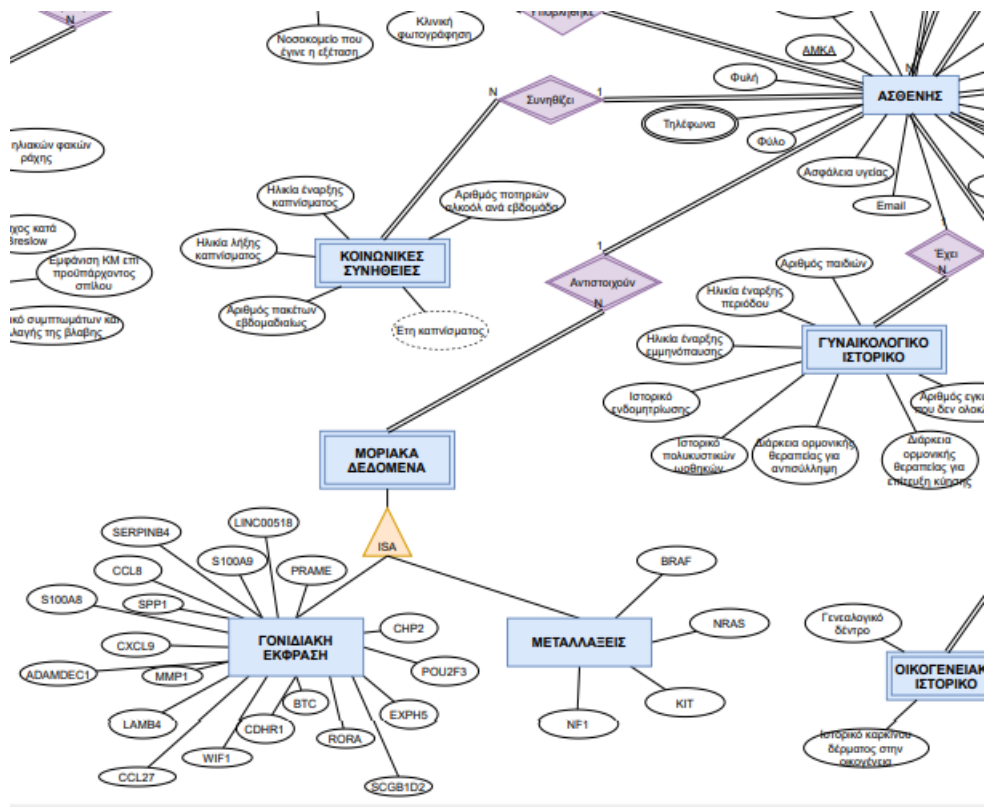
3^ο τμήμα: Παρουσιάζονται οι συσχετίσεις και τα γνωρίσματα της οντότητας «Ασθενής» με τις οντότητες «Κλινική Εξέταση», «Κοινωνικές Συνήθειες», «Αποτελέσματα Εξέτασης», «Κακοήθες Μελάνωμα».



Εικόνα 15: 3ο τμήμα ER Diagram

4^ο τμήμα: Παρουσιάζονται οι συσχετίσεις και τα γνωρίσματα της οντότητας «Ασθενής» με τις οντότητες «Μοριακά Δεδομένα», «Γονιδιακή Έκφραση», «Μεταλλάξεις». Τα γνωρίσματα στην οντότητα Μεταλλάξεις επιλέχθηκαν με βάση την βιβλιογραφία και την συχνότητα εμφάνισης στο μελάνωμα. Ενώ τα γνωρίσματα

στην οντότητα «Γονιδιακή Έκφραση» προέκυψαν από την ανάλυση που έγινε στο 3^ο κεφάλαιο.



Εικόνα 16: 4ο τμήμα ER Diagram

4.3 Data Dictionary

Παρακάτω αναπαρίστανται σε πίνακες αναλυτικά τα γνωρίσματα της κάθε οντότητας του ER Diagram με επιπλέον πληροφορίες που αφορούν τον τύπο δεδομένων και μια σύντομη περιγραφή.

- ΑΣΘΕΝΗΣ

Ο «Ασθενής» είναι μια οντότητα με πρωτεύον κλειδί το ΑΜΚΑ.

Πίνακας 4: Ασθενής

Γνώρισμα	Data Type	Description
A.M.K.A.	Int	Primary Key
Όνοματεπώνυμο	Varchar(200)	Έχει επιμέρους γνωρίσματα το όνομα και το επίθετο
Email	Varchar(200)	Η Ηλεκτρονική Διεύθυνση

Εθνικότητα	Varchar(200)	Εθνικότητα
Φυλή	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Καυκάσια», «Κίτρινη φυλή», «Λατινο-αμερικανική», «Μαύρη φυλή»
Χώρες	Varchar(200)	Πλειότιμο γνώρισμα που χωρίζεται σε επιμέρους γνωρίσματα για να έχουμε επιδημιολογική πληροφορία για τα μέρη που έζησε ο ασθενής
Οικογενειακή Κατάσταση	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Έγγαμος\η», «Άγαμος\η», «Διαζευγμένος\η», «Χήρος\α»
Ημερομηνία Γέννησης	Date	Ημέρα, Μήνας, Έτος γέννησης
Επίπεδο Μόρφωσης	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Απόφοιτος Γυμνασίου», «Απόφοιτος Λυκείου», «Απόφοιτος ΤΕΙ», «Απόφοιτος Πανεπιστημίου», «Απόφοιτος Μεταπτυχιακού», «Άλλο»
Ασφάλεια Υγείας	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Κρατική ασφάλιση υγείας (ΕΟΠΠΥ ή άλλο ταμείο)», «Ιδιωτική ασφάλιση υγείας», «Κρατική και ιδιωτική ασφάλιση υγείας», «Πρόνοια», «Όχι»
Τηλέφωνα	Int	Πλειότιμο γνώρισμα
Φύλο	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Γυναίκα», «Άντρας», «Άλλο»

- ΓΥΝΑΙΚΟΛΟΓΙΚΟ ΙΣΤΟΡΙΚΟ

Το «Γυναικολογικό Ιστορικό» είναι μια αδύναμη οντότητα που συνδέεται με τον «Ασθενή» με τη συσχέτιση «Έχει».

Πίνακας 5: Γυναικολογικό Ιστορικό

Γνώρισμα	Data Type	Description
Ηλικία Έναρξης Περιόδου	Int	Πρώτη ημέρα έμμηνου ρήσεως
Ηλικία Έναρξης Εμμηνόπαυσης	Int	Ηλικία διακοπής έμμηνου ρήσεως
Αριθμός παιδιών	Int	Αριθμός παιδιών
Ιστορικό ενδομητρίωσης	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Ιστορικό Πολυκυστικών Ωοθηκών	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Διάρκεια ορμονικής θεραπείας για αντισύλληψη	Int	Αριθμός σε μήνες
Διάρκεια ορμονικής θεραπείας για επίτευξη κύησης	Int	Αριθμός σε μήνες
Αριθμός εγκυμοσύνων που δεν ολοκληρώθηκαν	Int	Αριθμός εγκυμοσύνων που δεν ολοκληρώθηκαν

- ΚΟΙΝΩΝΙΚΕΣ ΣΥΝΗΘΕΙΕΣ

Η οντότητα «Κοινωνικές Συνήθειες» είναι μια αδύναμη οντότητα που συνδέεται με τον «Ασθενή» με την συσχέτιση «Συνηθίζει».

Πίνακας 6: Κοινωνικές Συνήθειες

Γνώρισμα	Data Type	Description
Ηλικία έναρξης καπνίσματος	Int	Ηλικία έναρξης καπνίσματος
Ηλικία λήξης καπνίσματος	Int	Ηλικία λήξης καπνίσματος
Έτη καπνίσματος	Int	Γνώρισμα που συμβολίζεται

		με διακεκομμένη γραμμή γιατί μπορεί να υπολογιστεί από άλλα
Αριθμός πακέτων τσιγάρων εβδομαδιαίως	Int	Αριθμός πακέτων τσιγάρων
Αριθμός ποτηριών αλκοόλ εβδομαδιαίως	Int	Αριθμός ποτηριών αλκοόλ

- ΚΛΙΝΙΚΗ ΕΞΕΤΑΣΗ

Η «Κλινική Εξέταση» είναι μια οντότητα που συνδέεται με τον «Ασθενή» με την συσχέτιση «Υποβλήθηκε» και έχει ως πρωτεύον κλειδί το ID Εξέτασης.

Πίνακας 7: Κλινική Εξέταση

Γνώρισμα	Data Type	Description
ID Εξέτασης	Int	Primary key
Ύψος	Int	Ύψος σε εκατοστά πχ 173
Βάρος	Float	Βάρος σε κιλά πχ 60,5
Περίμετρος Μέσης	Int	Περίμετρος μέσης σε εκατοστά πχ 54
Δερματοσκόπηση	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Κλινική φωτογράφιση	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Νοσοκομείο που έγινε η εξέταση	Varchar(200)	Το όνομα του Νοσοκομείου

- ΑΠΟΤΕΛΕΣΜΑ ΕΞΕΤΑΣΗΣ

Η οντότητα «Αποτέλεσμα Εξέτασης» είναι μια αδύναμη οντότητα που συνδέεται με την συσχέτιση «Βρέθηκε» με την «Κλινική Εξέταση».

Πίνακας 8: Αποτελέσματα Εξέτασης

Γνώρισμα	Data Type	Description
Αριθμός καλοήθων σπύλων	Int	Κοινοί, καλοήθεις σπύλοι
Αριθμός ηλιακών φακών	Int	Οι ηλιακές φακές είναι

ράχης		γνωστές και ως πανάδες
Θέση ηλιακών φακών ράχης	Varchar(200)	Μέρος που εντοπίστηκαν οι ηλιακές φακές ράχης
Αριθμός ηλιακών εφηλίδων	Int	Καλοήθεις επιδερμидικές καφεοειδείς βλάβες, που οφείλονται σε αύξηση της μελανίνης των μελανοκυττάρων
Θέση ηλιακών εφηλίδων	Varchar(200)	Μέρος που εντοπίστηκαν οι ηλιακές εφηλίδες

- ΚΑΚΟΗΘΕΣ ΜΕΛΑΝΩΜΑ

Η οντότητα «Κακοήθες Μελάνωμα» είναι εξειδίκευση της οντότητας «Αποτέλεσμα Εξέτασης» και αφορά τις εξετάσεις στις οποίες έχει γίνει διάγνωση για κακοήθες μελάνωμα έπειτα από βιοψία.

Πίνακας 9: Κακοήθες Μελάνωμα

Γνώρισμα	Data Type	Description
ID τομής βιοψίας	Int	Primary Key
Διάμετρος Βλάβης	Int	Διάμετρος βλάβης σε εκατοστά
Μέρος εμφάνισης	Varchar(200)	Θέση που εμφανίστηκε η βλάβη
Ημερομηνία Διάγνωσης	Date	Ημερομηνία διάγνωσης
Επίπεδο διήθησης	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «I Ενδοεπιδερμικό», «II Αρχόμενη διήθηση του θηλώδους χορίου», «III Πλήρης διήθηση και πλήρωση του χορίου», «IV Επέκταση του δικτυωτού χορίου», «V Διήθηση του υποδόριου λίπους»
Ιστολογικός Τύπος	Enum(val1, val2, val3,...)	Παίρνει τις τιμές:

		«Κακοήθες μελάνωμα επί εδάφους κακοήθους φακής(LMM)», «Οζώδες κακοήθες μελάνωμα(SSM)», «Οζώδες κακοήθες μελάνωμα(NM)», «Μεταστατικό κακοήθες μελάνωμα», «Μη-ταξινομούμενο», «Μελάνωμα αγνώστου πρωτοπαθούς»
Πάχος κατά Breslow	Float	Πάχος τους όγκου σε χιλιοστά σύμφωνα με τον δείκτη Breslow.
Ιστορικό συμπτωμάτων αλλαγής της βλάβης	Varchar(300)	Αναλυτική περιγραφή αλλαγής της βλάβης στον χρόνο
Εμφάνιση ΚΜ επί προϋπάρχοντος σπίλου	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Όχι (de novo βλάβη)», «Ναι, επί επίκτητου σπίλου», «Ναι, επί συγγενούς σπίλου»

- **ΕΚΘΕΣΗ ΣΕ ΥΠΕΡΙΩΔΗ ΑΚΤΙΝΟΒΟΛΙΑ**

Η αδύναμη οντότητα «Εκθεση σε υπεριώδη ακτινοβολία» συνδέεται με τον «Ασθενή» με την συσχέτιση «Υπόκειται».

Πίνακας 10: Έκθεση σε υπεριώδη ακτινοβολία

Γνώρισμα	Date Type	Description
Αριθμός Solarium	Int	Αριθμός συνεδριών για solarium στη ζωή του ατόμου
Αριθμός φωτοθεραπειών	Int	Αριθμός φωτοθεραπειών στη ζωή του ατόμου
Αριθμός ιονίζουσων	Int	Αριθμός ιονίζουσων

Θεραπειών		θεραπειών στη ζωή του ατόμου
Αριθμός φωτοχημειοθεραπειών	Int	Αριθμός φωτοχημειοθεραπειών στη ζωή του ατόμου

- **ΦΩΤΟΤΥΠΟΣ ΔΕΡΜΑΤΟΣ**

Η αδύναμη οντότητα «Φωτότυπος Δέρματος», αφορά την περιγραφή των φυσικών χρωμάτων του ασθενή και την αντίδραση του δέρματός του στον ήλιο, συνδέεται με τον «Ασθενή» με την συσχέτιση «Χαρακτηρίζεται».

Πίνακας 11: Φωτότυπος Δέρματος

Γνώρισμα	Data Type	Description
Χρώμα Ματιών	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Γκρι», «Μπλε», «Πράσινα», «Ανοιχτά καστανά», «Σκούρα καστανά», «Μαύρα»
Χρώμα Μαλλιών	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Ξανθά», «Κόκκινα», «Ανοιχτά καστανά», «Σκούρα καστανά», «Μαύρα» (αφορά το φυσικό χρώμα μαλλιών)
Χρώμα Δέρματος	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Ανοιχτόχρωμο», «Ελαφρά μελαχρινό», «Σκούρο μελαχρινό», «Μαύρο» (αφορά το έσω μέρος του μπράτσου ή των γλουτών)
Μαύρισμα έπειτα από έκθεση στον ήλιο	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Δεν μαυρίζω καθόλου», «Μαυρίζω λίγο», «Μαυρίζω μέτρια», «Μαυρίζω έντονα»
Κάψιμο έπειτα από έκθεση	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Καίγομαι

στον ήλιο		άσχημα», «Καίγομαι μέτρια», «Καίγομαι λίγο», «Δεν καίγομαι», «Είμαι μελαχρινό άτομα», «Ανήκω στη μαύρη φυλή»
Παρουσία φακίδων	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Όχι», «Λίγες», «Μερικές», «Πολλές» (αφορά την ηλικία των 21 ετών)

- ΕΚΘΕΣΗ ΣΤΗΝ ΗΛΙΑΚΗ ΑΚΤΙΝΟΒΟΛΙΑ

Η αδύναμη οντότητα «Έκθεση στην Ηλιακή Ακτινοβολία» συνδέεται με τον «Ασθενή» με την συσχέτιση «Εκτίθεται».

Πίνακας 12: Έκθεση στην ηλιακή ακτινοβολία

Γνώρισμα	Date Type	Description
Αριθμός ηλιακών εγκαυμάτων κατά την ενήλικη ζωή	Int	Πόσες φορές ο ασθενής «κάηκε» από την έκθεση του στον ήλιο στην ενήλικη ζωή του
Αριθμός ηλιακών εγκαυμάτων κατά την παιδική ηλικία	Int	Πόσες φορές ο ασθενής «κάηκε» από την έκθεση του στον ήλιο όσο ήταν παιδί
Αριθμός ετών έκθεσης στον ήλιο για ψυχαγωγικούς λόγους	Int	Ψυχαγωγικοί λόγοι μπορούν να θεωρηθούν οι ακόλουθοι: αθλητισμός, κολύμβηση, ψάρεμα, ιστιοπλοΐα, κηπουρική κλπ
Ημερήσιες ώρες έκθεσης στον ήλιο για ψυχαγωγικούς λόγους	Int	Ψυχαγωγικοί λόγοι μπορούν να θεωρηθούν οι ακόλουθοι: αθλητισμός, κολύμβηση, ψάρεμα, ιστιοπλοΐα, κηπουρική κλπ

Εβδομάδες διάρκειας καλοκαιρινών διακοπών	Int	Εβδομάδες διάρκειας καλοκαιρινών διακοπών
Ημερήσιες ώρες έκθεσης κατά τις καλοκαιρινές διακοπές	Int	Ημερήσιες ώρες έκθεσης στην ηλιακή ακτινοβολία κατά τις καλοκαιρινές διακοπές
Αριθμός ετών έκθεσης στον ήλιο για επαγγελματικούς λόγους	Int	Τα έτη για τα οποία οι ασθενείς εκτίθενται σε ηλιακή ακτινοβολία κατά την εργασία τους
Αριθμός ημερήσιων ωρών έκθεσης για επαγγελματικούς λόγους	Int	Οι ημερήσιες ώρες στις οποίες οι ασθενείς εκτίθενται σε ηλιακή ακτινοβολία κατά την εργασία τους

- ΑΝΤΙΗΛΙΑΚΗ ΠΡΟΣΤΑΣΙΑ

Η αδύναμη οντότητα «Αντιηλιακή Προστασία» συνδέεται με τον «Ασθενή» με την συσχέτιση «Χρησιμοποιεί» (όλα τα γνωρίσματα αφορούν τους καλοκαιρινούς μήνες).

Πίνακας 13: Αντιηλιακή Προστασία

Γνώρισμα	Data Type	Description
Χρήση αντιηλιακού	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Ποτέ», «<50% του χρόνου», «>50% του χρόνου», «Πάντα»
Χρήση γυαλιών	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Ποτέ», «<50% του χρόνου», «>50% του χρόνου», «Πάντα»
Χρήση καπέλου	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Ποτέ», «<50% του χρόνου», «>50% του χρόνου», «Πάντα»
Κάλυψη με ρούχα	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Ποτέ», «<50% του χρόνου», «>50% του χρόνου», «Πάντα»
Δείκτης προστασίας	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Συνήθως»

αντηλιακού		μικρότερο του 15», «Συνήθως μεταξύ 10-30», «Συνήθως μεγαλύτερο του 30»
Μέρος εφαρμογής αντηλιακού	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Μόνο στο πρόσωπο», «Πρόσωπο και βραχίονες», «Σε όλα τα ακάλυπτα σημεία του σώματος»

- ΙΑΤΡΙΚΟ ΙΣΤΟΡΙΚΟ

Η αδύναμη οντότητα «Ιατρικό Ιστορικό» συνδέεται με τον «Ασθενή» με την συσχέτιση «Έχει».

Πίνακας 14: Ιατρικό Ιστορικό

Γνώρισμα	Data Type	Description
Νοσήματα	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Σακχαρώδης Διαβήτης», «Στεφανιαία Νόσος», «Νόσος Parkinson», «Χρόνια αυτοάνοσα νοσήματα ή αρθρίτιδες», «Κακοήθειες», «Άλλο»
Δερματοπάθειες	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Χρόνια φλεγμονώδη νοσήματα δέρματος», «Λεύκη», «Φωτοδερματοπάθειες», «Άλλο»
Χειρουργικές Επεμβάσεις	Varchar(400)	Αναλυτική περιγραφή των χειρουργικών επεμβάσεων
Καθημερινή Φαρμακευτική αγωγή τελευταίων χρόνων	Varchar(400)	Συγκεκριμένα φάρμακα που παίρνει καθημερινά ο ασθενής τα τελευταία χρόνια

Προηγούμενο ιστορικό μελανώματος	Varchar(400)	Αναλυτική περιγραφή ύπαρξης ή μη ιστορικού μελανώματος
Αριθμός σπύλων που έχουν αφαιρεθεί	Int	Αριθμός σπύλων που έχουν αφαιρεθεί στο παρελθόν
Αριθμός κηλίδων ή υπερκερακτώσεων που έχουν αφαιρεθεί	Int	Αριθμός κηλίδων ή υπερκερακτώσεων που έχουν αφαιρεθεί στο παρελθόν
Ιστορικό καρκίνου εσωτερικού οργάνου	Varchar(400)	Αναλυτική περιγραφή ύπαρξης ιστορικού εσωτερικού οργάνου

- ΟΙΚΟΓΕΝΕΙΑΚΟ ΙΣΤΟΡΙΚΟ

Η αδύναμη οντότητα «Οικογενειακό Ιστορικό» συνδέεται με τον «Ασθενή» με την συσχέτιση «Υπάρχει».

Πίνακας 15: Οικογενειακό Ιστορικό

Γνώρισμα	Data Type	Description
Ιστορικό καρκίνου δέρματος στην οικογένεια	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Όχι», «Βασικοκυτταρικός», «Ακανθοκυτταρικός», «Μελάνωμα», «Δε γνωρίζω»
Βαθμός συγγένεια ατόμου με ιστορικό καρκίνου του δέρματος	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Μητέρα», «Πατέρας», «Παππούς/Γιαγιά», «Αδελφός/Αδελφή», «Τέκνο», «Άλλο»
Ιστορικό καρκίνου εσωτερικού οργάνου σε συγγενή πρώτου βαθμού	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Γενεαλογικό δέντρο	Image	Φωτογραφία Γενεαλογικού δέντρου με περιγραφή κακοηθειών σε πρώτου βαθμού συγγενείς

- ΘΕΡΑΠΕΥΤΙΚΗ ΑΝΤΙΜΕΤΩΠΙΣΗ

Η αδύναμη οντότητα «Θεραπευτική Αντιμετώπιση» συνδέεται με τον «Ασθενή» με την συσχέτιση «Αντιμετωπίζει».

Πίνακας 16: Θεραπευτική αντιμετώπιση

Γνώρισμα	Data Type	Description
Ημερομηνία επανεκτομής	Date	Ημερομηνία επανεκτομής (στην περίπτωση που έγινε)
Κλινικός ψηλαφητοί λεμφαδένες	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Θέση ψηλαφητών λεμφαδένων	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Προσθιωτιαίοι», «Οπισθοωτιαίοι», «Τραχηλικοί», «Αυχενικοί», «Μασχαλιαίοι», «Βουβωνικοί», «Άλλοι»
Αφαίρεση λεμφαδένα φρουρού	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Θέση λεμφαδένα φρουρού που αφαιρέθηκε	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Προσθιωτιαίοι», «Οπισθοωτιαίοι», «Τραχηλικοί», «Αυχενικοί», «Μασχαλιαίοι», «Βουβωνικοί», «Άλλοι»
Υπαρξη τοπικών μεταστάσεων	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
Υπαρξη απομακρυσμένων μεταστάσεων	Enum(val1, val,2, val3,...)	Παίρνει τις τιμές: «Οχι», «Πνεύμονες», «Ήπαρ», «Εγκέφαλος», «Δέρμα», «Υποδόριο», «Λεμφαδενικές απομακρυσμένες μεταστάσεις»
Ανοσοθεραπεία	Set(val1, val,2, val3,...)	Περιλαμβάνει επιμέρους γνωρίσματα τα οποία παίρνουν τις ακόλουθες τιμές: Ιντερφερόνη: «ΝΑΙ(Intron)», «ΝΑΙPEG-

		Intron», «OXI», Σχήμα ανοσοθεραπείας: «High dose», «Low dose», «Και τα 2», Διάρκεια: «1 μήνας», «12 μήνες»
Θεραπεία μεταστατικής νόσου	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «Θεραπεία», «Παρακολούθηση», «Χειρουργική», «Χημειοθεραπεία», «Ακτινοθεραπεία», «Ανοσοθεραπεία», «Στοχευμένη θεραπεία»
Είδος θεραπείας μεταστατικής νόσου	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «DTIC», «Temozolamide», «Ιντερφερόν», «Ipilimumab», «Vemourafenib»

- ΜΕΤΑΛΛΑΞΕΙΣ

Η οντότητα «Μεταλλάξεις» είναι εξειδίκευση της οντότητας «Μοριακά Δεδομένα». Τα γνωρίσματα αυτής της οντότητας είναι γονίδια που φαίνεται να είναι τα πιο αντιπροσωπευτικά στο μελάνωμα. Ένας ασθενής κάνοντας μια εξέταση μπορεί να διαπιστωθεί εάν φέρει μετάλλαξη σε κάποιο από αυτά τα γονίδια, για αυτό και θα παίρνουν τις τιμές «Ναι», «Όχι».

Πίνακας 17: Μεταλλάξεις

Γνώρισμα	Data Type	Description
BRAF	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
NRAS	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
KIT	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ», «ΟΧΙ»
NF1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «ΝΑΙ»,

		«ΟΧΙ»
--	--	-------

- ΓΟΝΙΔΙΑΚΗ ΕΚΦΡΑΣΗ

Η οντότητα «Γονιδιακή Έκφραση» είναι εξειδίκευση της οντότητας «Μοριακά Δεδομένα». Σε αυτήν τα γνωρίσματα είναι γονίδια που βρέθηκαν στην έρευνα που έγινε στο 3^ο κεφάλαιο ότι υπερεκφράζονται και υποεκφράζονται. Θα παίρνουν τις τιμές -1, 0, 1, οι οποίες είναι ποιοτικές και συμβολίζουν υποεκφρασμένο, χωρίς ενδιαφέρον, υπερεκφρασμένο γονίδιο αντίστοιχα.

Πίνακας 18: Γονιδιακή έκφραση

Γνώρισμα	Data Type	Description
PRAME	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
S100A9	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
LINC00518	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
SERPINB4	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
SPP1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
MMP1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
CCL8	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
CXCL9	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
S100A7	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
ADAMDEC1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
LAMB4	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
CCL27	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
WIF1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»

		«1»
BTC	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
RORA	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
CDHR1	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
CHP2	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
POU2F3	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
EXPH5	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»
SCGB1D2	Enum(val1, val2, val3,...)	Παίρνει τις τιμές: «-1», «0», «1»

4.4 Υλοποίηση με **mySQL**

Στα πλαίσια αυτού του κεφαλαίου γίνεται η υλοποίηση των πινάκων με την **mySQL**.

```
CREATE TABLE PATIENT(
```

```
    amka INT NOT NULL PRIMARY KEY,
```

```
    firstName VARCHAR(200) NOT NULL,
```

```
    lastName VARCHAR(200) NOT NULL,
```

```
    email VARCHAR(200) NOT NULL,
```

```
    nationality VARCHAR(200),
```

```
    tribe ENUM('Καυκάσια', 'Κίτρινη φυλή', 'Λατινο-Αμερικάνικη', 'Μαύρη  
φυλή'),
```

```
    city VARCHAR(200),
```

```
    street VARCHAR(200),
```

```
    streetNum INT,
```

```
    yearofa INT,
```

```
    yearofl INT,
```

```
    familyStatus ENUM('Έγγαμος\η', 'Άγαμος\η', 'Διαζευγμένος\η', 'Χήρος\α'),
```

```

    dateOfBirth DATE,
    education ENUM('Απόφοιτος Γυμνασίου', 'Απόφοιτος Λυκείου', 'Απόφοιτος
ΤΕΙ', 'Απόφοιτος Πανεπιστημίου', 'Απόφοιτος Μεταπτυχιακού', 'Άλλο'),
    healthInsurance ENUM('Κρατική ασφάλιση υγείας (ΕΟΠΠΥ ή άλλο ταμείο)',
'Ιδιωτική ασφάλιση υγείας', 'Κρατική και ιδιωτική ασφάλιση υγείας', 'Πρόνοια', 'Όχι'),
    phoneNum INT NOT NULL,
    gender ENUM('Γυναίκα', 'Αντρας', 'Άλλο')
);

```

```

CREATE TABLE CLINICAL_EXAMINATION(
    id INT NOT NULL PRIMARY KEY,
    hight INT,
    weight INT,
    waistCircumference FLOAT,
    dermatoscopy ENUM('NAI', 'OXI'),
    clinicalPhotos ENUM('NAI', 'OXI'),
    hospital VARCHAR(200)
);

```

```

CREATE TABLE RESULTS_OF_EXAMINATION(
    nevusNum INT,
    solbackNum INT,
    solefilNum INT,
    solefilPosition VARCHAR(200),
    solbackPosition VARCHAR(200),
    id INT,

    FOREIGN KEY (id) REFERENCES CLINICAL_EXAMINATION(id)
);

```

```

CREATE TABLE MALIGNANT_MELANOMA(
    idOfBiopsy INT NOT NULL PRIMARY KEY,
    damageCircumference INT,

```

```

damagePosition VARCHAR(200),
diagnosisDate DATE,
infiltrationLevel ENUM('I Ενδοεπιδερμικό', 'II Αρχόμενη διήθηση του
θηλώδους χορίου', 'III Πλήρης διήθηση και πλήρωση του χορίου', 'IV Επέκταση του
δικτυωτού χορίου', 'V Διήθηση του υποδόριου λίπους'),
histologicalType ENUM('Κακοήθες μελάνωμα επί εδάφους κακοήθους
φακής(LMM)', 'Οζώδες κακοήθες μελάνωμα(SSM)', 'Οζώδες κακοήθες
μελάνωμα(NM)', 'Μεταστατικό κακοήθες μελάνωμα', 'Μη-ταξινομούμενο',
'Μελάνωμα αγνώστου πρωτοπαθούς'),
breslowThickness FLOAT,
damageChangingRec VARCHAR(300),
melanOnPreNevus ENUM('Όχι (de novo βλάβη)', 'Ναι, επί επίκτητου σπίλου',
'Ναι, επί συγγενούς σπίλου')
);

```

```

CREATE TABLE SKIN_PHOTOTYPE(
eyeColor ENUM('Γκρι', 'Μπλε', 'Πράσινα', 'Ανοιχτά καστανά', 'Σκούρα
καστανά', 'Μαύρα'),
hairColor ENUM('Ξανθά', 'Κόκκινα', 'Ανοιχτά καστανά', 'Σκούρα καστανά',
'Μαύρα'),
skinColor ENUM('Ανοιχτόχρωμο', 'Ελαφρά μελαχρινό', 'Σκούρο μελαχρινό',
'Μαύρο'),
tanningAfterSun ENUM('Δεν μαυρίζω καθόλου', 'Μαυρίζω λίγο', 'Μαυρίζω
μέτρια', 'Μαυρίζω έντονα'),
sunburn ENUM('Καίγομαι άσχημα', 'Καίγομαι μέτρια', 'Καίγομαι λίγο', 'Δεν
καίγομαι', 'Είμαι μελαχρινό άτομα', 'Ανήκω στη μαύρη φυλή'),
freckles ENUM('Όχι', 'Λίγες', 'Μερικές', 'Πολλές'),
amka INT,

FOREIGN KEY (amka) REFERENCES PATIENT(amka)
);

```

```

CREATE TABLE MEDICAL_RECORD(

```

```

diseases ENUM('Σακχαρώδης Διαβήτης', 'Στεφανιαία Νόσος', 'Νόσος
Parkinson', 'Χρόνια αυτοάνοσα νοσήματα ή αρθρίτιδες', 'Κακοήθειες', 'Άλλο'),
skinDiseases ENUM('Χρόνια φλεγμονώδη νοσήματα δέρματος', 'Λεύκη',
'Φωτοδερματοπάθειες', 'Άλλο'),
surgicalProcedures VARCHAR(400),
dailyMedication VARCHAR(400),
melanomaRec VARCHAR(400),
nevusRemoveNum INT,
kilidesRemoveNum INT,
intenalCancerRec VARCHAR(400),
amka INT,

FOREIGN KEY (amka) REFERENCES PATIENT(amka)
);

```

```

CREATE TABLE FAMILY_RECORD(
skinCancerRecFam ENUM('Όχι', 'Βασικοκυτταρικός', 'Ακανθοκυτταρικός',
'Μελάνωμα', 'Δε γνωρίζω'),
degreeOfRelWithSkinCancer ENUM('Μητέρα', 'Πατέρας', 'Παππούς/Γιαγιά',
'Αδελφός/Αδελφή', 'Τέκνο', 'Άλλο'),
intenalCancerRecFam ENUM('ΝΑΙ', 'ΟΧΙ'),
pedigree BLOB,
amka INT,
FOREIGN KEY (amka) REFERENCES PATIENT(amka)
);

```

```

CREATE TABLE TREATMENT(
resectomy DATE,
lymphnodes ENUM('ΝΑΙ', 'ΟΧΙ'),
lymphnodesPosition ENUM('Προσθωτιαίοι', 'Οπισθοωτιαίοι', 'Τραχηλικοί',
'Αυχενικοί', 'Μασχαλιαίοι', 'Βουβωνικοί', 'Άλλοι'),
lymphnodeRemoval ENUM('ΝΑΙ', 'ΟΧΙ'),
lymphnodeRemovalPosition ENUM('Προσθωτιαίοι', 'Οπισθοωτιαίοι',
'Τραχηλικοί', 'Αυχενικοί', 'Μασχαλιαίοι', 'Βουβωνικοί', 'Άλλοι'),

```

```
metastasis ENUM('Όχι', 'Πνεύμονες', 'Ήπαρ', 'Εγκέφαλος', 'Δέρμα',
'Υποδόριο', 'Λεμφαδενικές απομακρυσμένες μεταστάσεις'),
metastasisTreatment ENUM('Θεραπεία', 'Παρακολούθηση', 'Χειρουργική',
'Χημειοθεραπεία', 'Ακτινοθεραπεία', 'Ανοσοθεραπεία', 'Στοχευμένη θεραπεία'),
kindOfMetastasisTreatment ENUM('DTIC', 'Temozolamide', 'Ιντερφερόν',
'Ipilimumab', 'Vemourafenib'),
amka INT,
```

```
FOREIGN KEY (amka) REFERENCES PATIENT(amka)
```

```
);
```

```
CREATE TABLE MUTATIONS(  
  BRAF ENUM('NAI', 'OXI'),  
  NRAS ENUM('NAI', 'OXI'),  
  KIT ENUM('NAI', 'OXI'),  
  NF1 ENUM('NAI', 'OXI')  
  amka INT,
```

```
FOREIGN KEY (amka) REFERENCES PATIENT(amka)
```

```
);
```

```
CREATE TABLE GENE EXPRESSION(  
  PRAME ENUM('-1', '0', '1'),  
  S100A9 ENUM('-1', '0', '1'),  
  LINC00518 ENUM('-1', '0', '1'),  
  SERPINB4 ENUM('-1', '0', '1'),  
  SPP1 ENUM('-1', '0', '1'),  
  MMP1 ENUM('-1', '0', '1'),  
  CCL8 ENUM('-1', '0', '1'),  
  CXCL9 ENUM('-1', '0', '1'),  
  S100A8 ENUM('-1', '0', '1'),  
  ADAMDEC1 ENUM('-1', '0', '1'),  
  LAMB4 ENUM('-1', '0', '1'),  
  CCL27 ENUM('-1', '0', '1'),  
  WIF1 ENUM('-1', '0', '1'),
```

BTC ENUM('-1', '0', '1'),
RORA ENUM('-1', '0', '1'),
CDHR1 ENUM('-1', '0', '1'),
CHP2 ENUM('-1', '0', '1'),
POU2F3 ENUM('-1', '0', '1'),
EXPH5 ENUM('-1', '0', '1'),
SCGB1D2 ENUM('-1', '0', '1'),
amka INT,

FOREIGN KEY (amka) REFERENCES PATIENT(amka)

);

Βιβλιογραφία

- [1] A. Ι. Στρατηγός, Κατευθυντήριες Οδηγίες Διάγνωσης και Αντιμετώπισης Μελανώματος, Αθήνα, 2016.
- [2] Z. Metka and K. Marcel, Methodological guidelines and recommendations for efficient and rational governance of patient registries, Ljubljana: National Institute of Public Health, Trubarjeva 2, 1000 Ljubljana, Slovenia , 2015.
- [3] "Melanoma Uk," 2016. [Online]. Available: <https://www.melanomauk.org.uk/>.
- [4] "War on Melanoma," [Online]. Available: <https://www.ohsu.edu/war-on-melanoma/melanoma-community-registry>.
- [5] «Cure OM Patient Registry,» [Ηλεκτρονικό]. Available: <https://melanoma.org/cure-om-patient-registry/>.
- [6] «YCCI Research Services,» [Ηλεκτρονικό]. Available: <https://medicine.yale.edu/ycci/researchservices/supportservices/general/melanoma/>.
- [7] A. Jochems, M. G. Schouwenburg, B. Leeneman, G. Vreugdenhil, M. W. Wouters and J. Van der Hoeven, "Dutch Melanoma Treatment Registry: Quality assurance in the care of patients with metastatic melanoma in the Netherlands," 25 December 2016.
- [8] M. Weichenthal, A. van Akkooi, P. Mohr, E. Ellebaek, I. Svane, S. Ugurel-Becker, U. Leiter, L. Hojberg, J. Haanen, H. Schmidt, T. Oellegaard, M. Wouters, D. Schadendorf and L. Bastholt, "EUMelaReg - a European platform for outcome research on real world treatment data of patients with advanced melanoma.," 21 October 2018.
- [9] L Ríos, E. Nagore, J. López, P Redondo, R. M. Martí, R. Fernandez-de-Misa and B. Soler, "The Spanish National Cutaneous Melanoma Registry. Tumour characteristics at Diagnosis: 15 Years of Experience.," April 2013.
- [10] L. Rosenkrantz Hölmich, S. Klausen, E. Spaun, G. Schmidt, D. Gad, I. M. Svane, H. Schmidt, H. F. Lorentzen και E. H. Ibfelt, «The Danish Melanoma

- Database,» 25 October 2016.
- [11] «Swedish Melanoma Registry (SweMR),» [Ηλεκτρονικό]. Available: <https://skr.se/en/kvalitetsregister/hittaregister/registerarkiv/hudmelanom.44558.html>.
- [12] R. A. Irizarry, B. Bolstad, F. Collin, L. M. Cope, B. Hobbs και T. P. Speed, Summaries of Affymetrix GeneChip probe level data., 2003.
- [13] R. Irizarry, B. Hobbs, F. Collin, L. Cope, Y. Beazer-Barclay, K. Antonellis, U. Scherf και T. Speed, Exploration, normalization, and summaries of high density oligonucleotide array probe level data, 2003.
- [14] R. Irizarry, C. Wang, Y. Zhou και T. Speed, Gene set enrichment analysis made simple, 2009.
- [15] R. Lipshutz, S. Fodor, T. Gingeras και D. Lockhart, High density synthetic oligonucleotide arrays, 1999.
- [16] B. Carvalho και R. Irizarry, «A framework for oligonucleotide microarray preprocessing,» 2010.
- [17] Q. Yang, J. Cui, I. Chazaro, L. Cupples και S. Demissie, «Power and type I error rate of false discovery rate approaches in genome-wide association studies,» 2005.
- [18] B. Bolstad, R. Irizarry, M. Astrand και T. Speed, «A comparison of normalization methods for high density oligonucleotide array data based on variance and bias,» 2003.
- [19] S. Choe, M. Boutros, A. Michelson, G. Church και M. Halfon, «Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset,» 2005.
- [20] S. Calza, W. Raffelsberger, A. Ploner, J. Sahel, T. Leveillard και Y. Pawitan, «Filtering genes to improve sensitivity in oligonucleotide microarray data analysis,» 2007.
- [21] M. Jeanmougin, A. de Reynies, L. Marisa, C. Paccard, G. Nuel και M. Guedj, «Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies,» 2010.
- [22] P. Baldi και A. Long, «A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes,» 2001.
- [23] R. Fox και M. Dimmic, «A two-sample Bayesian t-test for microarray data,» 2006.
- [24] «Wikipedia,» 2019. [Ηλεκτρονικό]. Available: https://el.wikipedia.org/wiki/%CE%92%CE%AC%CF%83%CE%B7_%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CF%89%CE%BD.
- [25] Π. Ευαγγελία, «Εννοιολογική Μοντελοποίηση Δεδομένων».

Παράρτημα

Κώδικας σε R

```
install.packages("installr")
library(installr)
updateR()
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install()
BiocManager::install("GEOquery")
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("affy")
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("hgu133acdf")

library(GEOquery)
library(affy)
library(limma)
library(hgu133acdf)

#Set working directory
setwd("C:/users/Maraki/Documents/RProjects1/Analysis_GSE114445")

#Download Files
getGEOSuppFiles("GSE114445")

setwd("C:/users/Maraki/Documents/RProjects1/Analysis_GSE114445/GSE114445")

#unzip CELL
untar("GSE114445_RAW.tar", exdir = "data")
cels = list.files("data/", pattern = "CEL")
```

```

sapply(paste("data", cels, sep = "/"), gunzip)
cels = list.files("data/", pattern = "CEL")

#AffyBatch Objects
setwd("C:/users/Maraki/Documents/RProjects1/Analysis_GSE114445/GSE114445/data")
raw.data <- ReadAffy(verbose = TRUE, filenames = cels)

#number of probe ids and samples
dim(exprs(raw.data))

#normalization
eset <- rma(raw.data)
save(raw.data, eset, file =
"C:/users/Maraki/Documents/RProjects1/Analysis_GSE114445/new")

#number of probe sets after normalization
dim(eset)

#Quality Control
#CEL Image
image(raw.data[,1])

#histogram of raw expression values for the first sample
hist(exprs(raw.data)[,1], breaks = 200, main = "Histogram of raw expression values")

#log2 transformation of expression values
affylog2 <- log2(exprs(raw.data))

# histogram of log2-transformed expression values
hist(affylog2[,1], br=100, col=rainbow(100), main="Histogram of log2-transformed
expression values")

# densities before-after normalization

```

```
plotDensity(log2(exprs(raw.data)), main="Before normalization (log2-transformed) -  
Probe level")
```

```
plotDensity(exprs(eset), main="RMA-normalized - Probeset level")
```

```
# boxplots before-after normalization
```

```
boxplot(log2(exprs(raw.data)), pch=".", las=2, cex.axis=0.5, main="Before  
normalization (log2-transformed) - Probe level")
```

```
boxplot(exprs(eset), pch=".", las=2, cex.axis=0.5, main="RMA-normalized - Probeset  
level")
```

```
# RLE plot
```

```
#Compute a median for each row (probe)
```

```
m <- rowMedians(exprs(eset))
```

```
# Centring subtract the median value of each probe
```

```
rle <- sweep(exprs(eset),1, m, "-")
```

```
# Plot a box of the probe-wise centred values
```

```
boxplot(rle, pch=".", las=2, cex.axis=0.5, main="RLE plot")
```

```
# MA plot
```

```
ref<-rowMedians(exprs(eset))
```

```
A <- (exprs(eset)[,1] + ref)/2
```

```
M <- exprs(eset)[,1] - ref
```

```
plot(A,M,pch=16,col=rainbow(1,alpha=0.15,start=0.7), panel.first=grid())
```

```
# Label the IDs of up- and down-regulated genes
```

```
text(A[abs(M) > 4], M[abs(M) > 4], featureNames(eset)[abs(M) > 4], cex=0.4,pos=4)
```

```
abline(h=0)
```

```
abline(h=1)
```

```
abline(h=-1)
```

```

# Draw a local regression line
lines(lowess(A,M),col="green",lwd=3)

#Filtering
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("panp")
#Set new levels
DATA<-eset
DATA$Sample.source.name = factor(DATA$Sample.source.name)
new.levels<-c("Non-Melanoma","Non-Melanoma","Non-Melanoma","Primary
Melanoma")
DATA$Sample.source.name <- factor(new.levels[DATA$Sample.source.name])
DATA$Sample.source.name
library(panp)
pa.calls()
PA <- pa.calls(DATA, looseCutoff=0.02, tightCutoff=0.01, verbose = FALSE)
Pacalls <- PA$Pcalls
Pvalues <- PA$Pvals
write.table(Pacalls, file="Pacalls_gcrma.csv", sep="," , col.names=NA)
write.table(Pvalues, file="Pvalues_gcrma.csv", sep="," , col.names=NA)
P_list <-rownames(Pacalls)[Pacalls[,1]=="P"]
M_list <-rownames(Pacalls)[Pacalls[,1]=="M"]
A_list <-rownames(Pacalls)[Pacalls[,1]=="A"]
datadraft = pData(DATA)
nonmelan = datadraft$Row.names[datadraft$Sample.source.name=="Non-
Melanoma"]
melan = datadraft$Row.names[datadraft$Sample.source.name=="Primary
Melanoma"]
xxnonmelan = Pacalls[,nonmelan]
rrmelan = Pacalls[,melan]
qq=Pacalls[which((rowMeans(xxnonmelan == "A") < 0.50 | rowMeans(rrmelan
=="A") < 0.50)),]

```

```

#Number of probe sets after filtering
dim(qq)
eset2 <- DATA[rownames(qq),]

#Limma
library("limma")
table(eset2$Sample.source.name)
design <- model.matrix(~factor(eset2$Sample.source.name))
design
fit <- lmFit(eset2, design)
fit
ebayes <- eBayes(fit, robust = T, trend = T)
#TopTable
tableTop <- topTable(ebayes, coef=2, number = Inf, lfc=0.5)
#Number of probe sets after limma analysis
dim(tableTop)
#Volcano plot
volcanoplot(ebayes, coef=2)
abline(v=-0.5)
abline(v=0.5)
abline(h=1.3)

#Annotation
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("hgu133plus2.db")
library(hgu133plus2.db)
ann <- select(hgu133plus2.db, keys=row.names(tableTop),
  columns=c("SYMBOL", "GENENAME"), keytype="PROBEID")
write.csv(ann, file = "MelannoVSall.csv")
write.csv(tableTop, file = "toptablemelanomVSALL.csv")

```

