



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

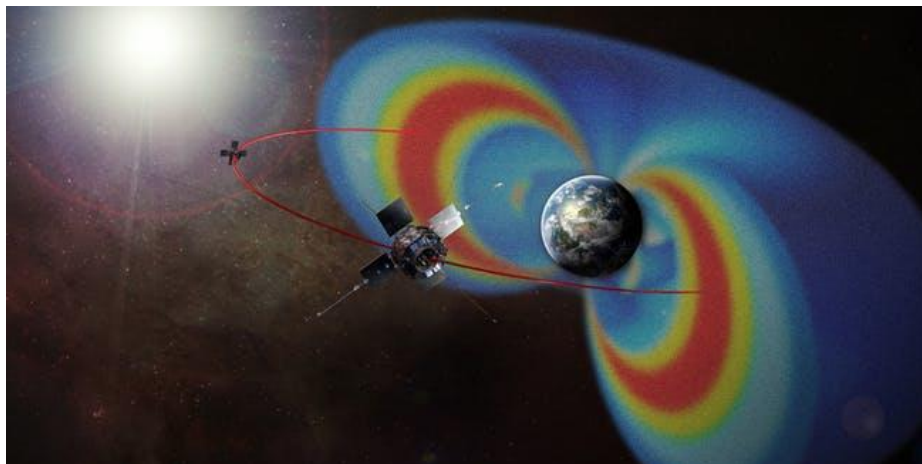
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Ανάλυση δεδομένων μικροσυστοιχιών για την
επίδραση ακτινοβολιών υψηλού LET με
εφαρμογές στη θεραπεία του καρκίνου ή το
διάστημα»**



Θεοδώρα-Δάφνη Μιχαλέττου

Πτυχιούχος Σχολής Εφαρμοσμένων Μαθηματικών & Φυσικών
Επιστημών, ΕΜΠ

ΑΘΗΝΑ (2019-20)



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

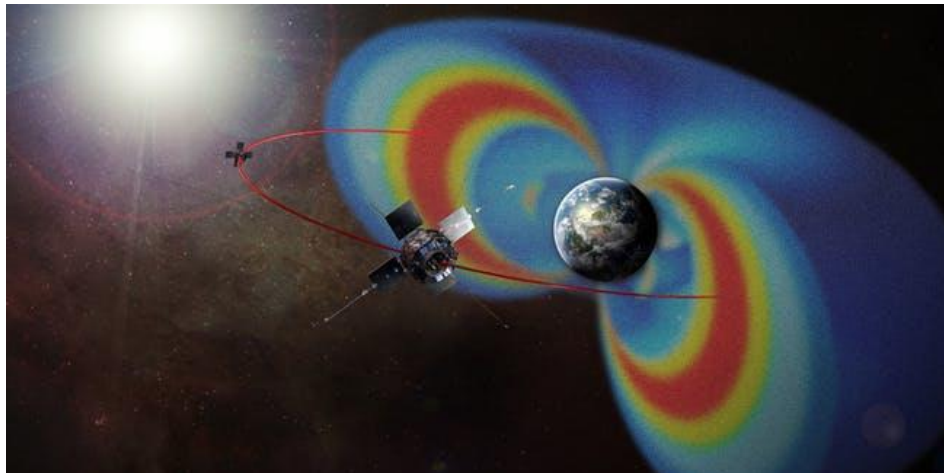
HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY

MASTER IN «BIOINFORMATICS»

MASTER DIPLOMA THESIS

«Microarray data analysis for studying the effects of high-LET ionizing radiation and emerging applications in cancer treatment or space »



Theodora-Dafni Michalettu

School of Applied Mathematical & Physical Sciences Graduate,
N.T.U.A.

ATHENS (2019-20)



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

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Ανάλυση δεδομένων μικροσυστοιχιών για
την επίδραση ακτινοβολιών υψηλού LET με
εφαρμογές στη θεραπεία του καρκίνου ή το
διάστημα»**

Τριμελής εξεταστική επιτροπή
Αναπληρωτής Καθηγητής Αλέξανδρος Γεωργακίλας (Επιβλέπων)
Τομέα Φυσικής,
Σ.Ε.Μ.Φ.Ε., Ε.Μ.Π.

Καθηγητής Παντελής Μπάγκος
Τμήμα Πληροφορικής με Εφαρμογές στην Βιοϊατρική,
Πανεπιστήμιο Θεσσαλίας

Καθηγητής Βιολογίας Ιωάννης Τρουγκάκος
Τομέας Βιολογίας Κυττάρου & Βιοφυσικής,
Τμήμα Βιολογίας, Ε.Κ.Π.Α.

Ευχαριστίες

Η εργασία αυτή εκπονήθηκε στο Εργαστήριο Βλαβών DNA του τομέα Φυσικής στο Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ). Ευχαριστώ τον Αναπ. Καθ. κ. Α. Γεωργακίλα για την ανάθεση του θέματος, την καθοδήγηση του κατά την διεξαγωγή της εργασίας αυτής και τις εποικοδομητικές συζητήσεις στα θέματα ραδιοβιολογίας. Ευχαριστώ θερμά τον συνάδελφο και υποψήφιο διδάκτορα Έκτωρ - Ξαβιέ Δελαστίκ για την συμβολή του σε θέματα ανάλυσης δεδομένων μικροσυστοιχιών. Στη συνέχεια ευχαριστώ τους Καθ. κ. Ι. Τρουγκάκο και Καθ. κ. Π. Μπάγκο που δέχθηκαν να συμμετέχουν στην τριμελή εξεταστική επιτροπή μου. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου και τον αδερφό μου για την αμέριστη συμπαράσταση τους κατά την διάρκεια των σπουδών μου και την συναισθηματική υποστήριξη.

Περίληψη

Εισαγωγή

Η χρήση ιοντίζουσας ακτινοβολίας (IR) υψηλής γραμμικής μεταφοράς ενέργειας (LET), προοδευτικά ενσωματώνεται στην ακτινοθεραπεία (RT) λόγω της στοχευμένης εναπόθεσης ενέργειας και της υψηλής σχετικής βιολογικής αποτελεσματικότητας (RBE), η οποία έχει συσχετιστεί και με την επαγωγή σύνθετων βλαβών DNA. Αυτά τα πλεονεκτήματα της σωματιδιακής ακτινοβολίας καθίστανται παράλληλα παράγοντας κινδύνου για τους αστροναύτες, εξαιτίας της αναπόφευκτης έκθεσης στην κοσμική ακτινοβολία. Κάνοντας λόγο λοιπόν για υγιείς ιστούς και οργανισμούς, η απόκριση σε βλάβες DNA (DDR) ως αποτέλεσμα της επαγωγής σύνθετων βλαβών DNA, οι οποίες προκύπτουν από τέτοιους τύπους ακτινοβολίας, μπορεί να συμβάλει στις διάφορες χρόνιες και όψιμες επιδράσεις της IR. Η μέθοδος προσέγγισης που προτείνουμε προς μελέτη των πιθανών υποβοσκοτώντων πολύπλοκων μηχανισμών, είναι η μελέτη των μεταβολών στην έκφραση γονιδίων ύστερα από έκθεση σε IR υψηλού LET, μέσω βιοπληροφορικής.

Μέθοδος

Επιλέχθηκαν ανθρώπινα δείγματα ακτινοβολημένα με υψηλής LET IR και μη-ακτινοβολημένα (δείγματα ελέγχου) στα οποία καθορίστηκε η γονιδιακή έκφραση μέσω τεχνολογίας μικροσυστοιχιών cDNA. Χρησιμοποιήθηκαν μη επεξεργασμένα δεδομένα από κατατεθειμένες μελέτες στα δημόσια αποθετήρια του GEO και ArrayExpress. Η προ-επεξεργασία και η ανάλυση διαφορικής έκφρασης πραγματοποιήθηκε σε γλώσσα προγραμματισμού R. Τα διαφορικά εκφρασμένα γονίδια που προέκυψαν, υποβλήθηκαν σε ανάλυση λειτουργικού εμπλουτισμού ώστε να ταυτοποιηθούν όροι γονιδιακής οντολογίας (GO) και βιολογικών μονοπατιών που σχετίζονται με τις επιδράσεις της υψηλής LET IR, καθώς και για σχετιζόμενες παθήσεις με τα εργαλεία WebGestalt και Enrichr, αντίστοιχα. Εκτελέστηκε βιβλιογραφική αναζήτηση βασισμένη στο GO και στην κατασκευή δικτύων πρωτεϊνικών αλληλεπιδράσεων (PPI networks) με το εργαλείο STRING, ώστε να αναδειχθούν γονίδια ενδιαφέροντος σχετιζόμενα με την DDR. Τέλος, διευρύνουμε την μελέτη, προς αναζήτηση συγγενών παθήσεων που έχουν συνδεθεί με έκθεση σε βαρέα ιόντα (HZE), συνδυάζοντας συσχετίσεις γονιδίων-ασθενειών από το DisGeNET και δίκτυα ασθενειών από το DISNOR.

Αποτελέσματα-Συζήτηση

Τα ευρήματά μας υποδεικνύουν την επαγωγή μιας διαφορικής βιολογικής απόκρισης σε ακτινοβολία υψηλής LET (σε σχέση με χαμηλού LET όπως ακτίνων X), κυρίως μέσω ενισχυμένης φλεγμονώδους απόκρισης που συμπεριλαμβάνει μηχανισμούς που σχετίζονται με το επίκτητο ανοσοποιητικό σύστημα. Η ενεργοποίηση της σηματοδότησης JAK-STAT υποστηρίζει περαιτέρω αυτή την υπόθεση. Επιπλέον, οι αποπτωτικοί μηχανισμοί φαίνεται να αναστέλλονται, όπως υποστηρίζεται και από την διέγερση της εστιακής προσκόλλησης (focal adhesion), γεγονός που μπορεί να οδηγήσει στην επαγωγή κυτταρικής γήρανσης. Επεκτείνοντας την μελέτη αυτή προς αναζήτηση πιθανών συγγενών παθήσεων της έκθεσης σε HZE, η ανάλυση λειτουργικού εμπλουτισμού ανέδειξε τη συμμετοχή μηχανισμών που συσχετίζονται με νευροεμφυλιστικές παθήσεις, μέσω των μονοπατιών σηματοδότησης Wnt και ασβεστίου, καθώς και του σχηματισμού αμυλοειδών ινιδίων. Η ενεργοποίηση του μονοπατιού σηματοδότησης NOTCH, η οποία υποστηρίζεται και από την υπερ-έκφραση του SIRT1 σε επίπεδο βιολογίας συστημάτων, είναι ενδεικτική της σύνδεσης με καρδιαγγειακές παθήσεις. Τέλος, μέσω της απορρύθμισης στην έκφραση του PLA2G4A, εξάγεται η σύνδεση με φλεγμονώδεις μεταβολικές ασθένειες.

Λέξεις-κλειδιά: ιοντίζουσες ακτινοβολίες, κοσμική ακτινοβολία, υψηλή LET, απόκριση σύνθετων βλαβών DNA, βιολογία συστημάτων, βιοπληροφορική, R-Bioconductor, διαφορική έκφραση γονιδίων, λειτουργικός εμπλουτισμός, βιολογικά δίκτυα, συγγενείς παθήσεις, καρκίνος

Abstract

Introduction

The use of high Linear Energy Transfer (LET) Ionizing Radiation (IR) is progressively being incorporated in Radiation Therapy (RT) due to its precise dose localization and high relative biological effectiveness, also associated with the induction of highly complex DNA damage. At the same time, these benefits of particle radiation become a high risk in the case of inevitable, cosmic radiation exposure for astronauts. Nonetheless, DNA Damage Response (DDR) activated via complex DNA damage on healthy tissue, occurring from such types of radiation, may be instrumental in the induction of various chronic and late effects. A method of approach for better understanding the possible complicated underlying response mechanisms, is studying alterations in gene expression after exposure to high-LET IR with the use of bioinformatics.

Methods

To this end, human samples irradiated with high LET IR and non-irradiated (control) were selected and gene expression was determined using cDNA microarray data analysis. Raw data from various studies were downloaded from GEO and ArrayExpress public databases. Pre-processing and differential expression analysis was performed using R programming language. Differentially expressed genes were subjected to functional enrichment analysis to identify Gene Ontology (GO) terms and biological pathways related to high LET IR effects as well as associated diseases, using WebGestalt and Enrichr, respectively. A bibliographical research was performed using GO and Protein to Protein Interaction (PPI) network construction via STRING, to select a group of genes of interest, regarding DDR. Finally, we expanded our study in search of possible comorbidities for heavy ion (HZE) exposure, combining gene-disease associations from DisGeNET platform and disease networks from DISNOR.

Results-Discussion

Our findings suggest the induction of a differential biological response for high-LET radiations, in comparison to low-LET IR, mainly through an enhanced inflammatory response, including several mechanisms associated with the adaptive immune system. Activation of JAK-STAT signaling pathway further upholds this premise. In addition, apoptotic mechanisms seem to be inhibited, further supported by stimulation of focal adhesion, which may imply the induction of cellular senescence. By expanding our study in search of possible comorbidities for HZE particle exposure, functional enrichment analysis suggests the involvement of mechanisms, tightly correlated with neurodegenerative disorders through Wnt and Calcium signaling pathways as well as amyloid fibrils formation. NOTCH signaling pathway activation, also supported by the up-regulation of SIRT1 at a systems biology level, is indicative of a connection to cardiovascular disease. Finally, through dysregulation in the expression of PLA2G4A, inflammatory metabolic diseases are suggested.

Keywords: ionizing radiation, cosmic radiation, high LET, complex DNA damage response, systems biology, bioinformatics, microarrays, R-Bioconductor, differential gene expression, functional enrichment, biological networks, comorbidities, cancer

ΠΕΡΙΕΧΟΜΕΝΑ

1.	ΕΙΣΑΓΩΓΗ.....	15
1.1.	Ιοντίζουσες Ακτινοβολίες στην Βιολογία.....	15
1.1.1.	Γραμμική Εναπόθεση Ενέργειας.....	16
1.1.2.	Επίδραση στον Βιολογικό Ιστό.....	19
1.1.3.	Σύνθετες Βλάβες DNA	20
1.1.4.	Απόκριση στις Βλάβες DNA.....	21
1.1.5.	Διαστημικές Αποστολές	24
1.1.6.	Συγγενείς Παθήσεις.....	25
1.2.	Μικροσυστοιχίες.....	28
1.2.1.	Κατασκευή & Λειτουργία Μικροσυστοιχιών	29
1.2.2.	Ανάλυση Δεδομένων Μικροσυστοιχιών.....	32
2.	ΜΕΘΟΔΟΣ & ΔΙΑΔΙΚΑΣΙΑ.....	36
2.1.	Εισαγωγή	36
2.2.	Αναζήτηση Πρωτογενών Δεδομένων	37
2.2.1.	Αποθετήρια Δεδομένων Γονιδιωματικής	37
2.2.2.	Πειραματικός Σχεδιασμός Επιλεγμένων Σειρών	39
2.3.	Ανάλυση Δεδομένων Μικροσυστοιχιών	44
2.3.1.	Εισαγωγή Πρωταρχικών Δεδομένων	45
2.3.2.	Προ-επεξεργασία Δεδομένων	51
2.3.3.	Διαφορική Έκφραση Γονιδίων	84
2.4.	Μελέτη των επιδράσεων ακτινοβολίας	90
2.4.1.	Λειτουργική Ανάλυση Εμπλουτισμού.....	90
2.4.2.	Ομαδοποίηση Δεδομένων	91
2.4.3.	Ανάλυση Εμπλουτισμού για Ασθένειες & Παθήσεις	93
2.4.4.	Αναζήτηση Σημαντικών Γονιδίων της Απόκρισης σε IR.....	94
2.4.5.	Συγγενείς Παθήσεις.....	101
3.	ΑΠΟΤΕΛΕΣΜΑΤΑ	102

3.1.	Μελέτη Εντός των Σειρών Δεδομένων	102
3.1.1.	Λειτουργική Ανάλυση Εμπλουτισμού.....	102
3.1.2.	Ποσοτική Σύγκριση Έκφρασης Σημαντικών Γονιδίων.....	109
3.2.	Μελέτη Ομαδοποιημένων Δεδομένων	111
3.2.1.	Ομαδοποίηση Δεδομένων – Διαγράμματα Venn.....	111
3.2.2.	Λειτουργική Ανάλυση Εμπλουτισμού.....	117
3.2.3.	Ανάλυση Εμπλουτισμού για Όρους Ασθενειών	124
3.2.4.	Συγγενείς Παθήσεις & Δίκτυα Ασθενειών	132
4.	ΣΥΖΗΤΗΣΗ.....	136
5.	ΠΑΡΑΡΤΗΜΑ	143
6.	ΒΙΒΛΙΟΓΡΑΦΙΑ	150

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 1: Στοιχεία για τον πειραματικό σχεδιασμό των μελετών που επιλέχθηκαν από το GEO και το ArrayExpress και αναλύθηκαν για το σκοπό της εργασίας.....	42
Πίνακας 2: Κωδικός και πλήρες όνομα της τεχνολογίας της μικροσυστοιχίας (πλατφόρμας) που χρησιμοποιήθηκε σε κάθε σειρά δεδομένων που επιλέχθηκε και αναλύθηκε. Οι σειρές με αστερίσκο βρέθηκαν και στα δύο αποθετήρια.	43
Πίνακας 3: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με την μέθοδο GSEA στο WebGestalt.	91
Πίνακας 4: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με την μέθοδο ORA στο WebGestalt.	92
Πίνακας 5: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με το Enrichr ως προς όρους ασθενειών και παθήσεων.	94
Πίνακας 6: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά κάθε συστάδας κατά την ανάλυση της μεγαλύτερης συστάδας του δικτύου πρωτεϊνών επιδιόρθωσης DNA.....	97
Πίνακας 7: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά των συστάδων 2-4 του δικτύου πρωτεϊνών επιδιόρθωσης DNA.....	97
Πίνακας 8: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά των συστάδων 1-7 του δικτύου πρωτεϊνών του ανοσοποιητικού συστήματος.	100
Πίνακας 9: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικώς εκφρασμένων γονιδίων της σειράς GSE44282, που αποτελείται από δείγματα ανθρωπίνων βρογχικών επιθηλιακών κυττάρων, ακτινοβολημένα με 1Gy ακτίνων-X και ιόντων σιδήρου αντίστοιχα που συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.....	102
Πίνακας 10: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικώς εκφρασμένων γονιδίων της σειράς E-MTAB-5754, που αποτελείται από δείγματα ανθρωπίνων ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα με 2Gy ακτίνων-X και ιόντων σιδήρου αντίστοιχα που συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.	103
Πίνακας 11: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικώς εκφρασμένων γονιδίων των σειρών E-MTAB-5761/3463, που αποτελείται από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος, ακτινοβολημένα με 1Gy ιόντων σιδήρου, ιόντων άνθρακα και ακτίνων-X αντίστοιχα που συλλέχθηκαν 8 ώρες μετά την ακτινοβολήση.	105
Πίνακας 12: Κρίσιμα διαφορικώς εκφρασμένα γονίδια, που χρησιμοποιήθηκαν για τον σχεδιασμό των διαγραμμάτων στηλών έκφρασης, για κάθε μία από τις σειρές δεδομένων που περιλάμβαναν δείγματα εκτεθειμένα σε ακτινοβολία υψηλής και χαμηλής τιμής LET.....	111
Πίνακας 13: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5754, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X, τα οποία συλλέχθηκαν 8-24 ώρες μετά την ακτινοβολήση.	117
Πίνακας 14: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από την σειρά GSE81067, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων και ινοβλαστών πνεύμονα, ακτινοβολημένα με υψηλές δόσεις σωματιδίων άλφα, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.	118

Πίνακας 15: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές GSE63952/64375, που προήλθαν από δείγματα ανθρωπίνων λευκοκυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 6 ώρες μετά την ακτινοβόληση.	118
Πίνακας 16: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα ομαδοποιημένα διαφορικώς εκφρασμένα γονίδια από την σειρά GSE44282, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις ιόντων σιδήρου, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβόληση.	120
Πίνακας 17: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές GSE44282/63952/64375 και E-MTAB-5761/5754, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, απομονωμένων λευκοκυττάρων, μονοπύρηνων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 1-24 ώρες μετά την ακτινοβόληση.	121
Πίνακας 18: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5754, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X, τα οποία συλλέχθηκαν 8-24 ώρες μετά την ακτινοβόληση.	124
Πίνακας 19: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από την σειρά GSE81067, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων και ινοβλαστών πνεύμονα, ακτινοβολημένα υψηλές δόσεις σωματιδίων άλφα, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβόληση.	125
Πίνακας 20: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές GSE6395/64375, που προήλθαν από δείγματα ανθρωπίνων λευκοκυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 6 ώρες μετά την ακτινοβόληση.	126
Πίνακας 21: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα ομαδοποιημένα διαφορικώς εκφρασμένα γονίδια από την σειρά GSE44282, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, ακτινοβολημένα με υψηλές δόσεις ακτίνων-X και βαρέων ιόντων, τα οποία συλλέχθηκαν 12-24 ώρες μετά την ακτινοβόληση.	127
Πίνακας 22: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα διαφορικώς εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5761, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος, ακτινοβολημένα υψηλές δόσεις ακτίνων-X και βαρέων ιόντων, τα οποία συλλέχθηκαν 8 ώρες μετά την ακτινοβόληση.	129
Πίνακας 23: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα διαφορικώς εκφρασμένα γονίδια από την σειρά E-MTAB-5754, που προήλθαν από δείγματα ανθρωπίνων ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X και ιόντων σιδήρου, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβόληση.	131
Πίνακας 24: Πειραματικά διαφορικώς εκφρασμένα γονίδια κοινά με την βιβλιογραφία για τις 4 κατηγορίες συγγενών παθήσεων.	133
Πίνακας 25: Συντομογραφίες που αναφέρθηκαν σε αυτή την εργασία.	143
Πίνακας 26: Παράθεση των ενδεικτικών γονιδίων που επιλέχθηκαν προς ποσοτική μελέτη στις σειρές δεδομένων μικροσυστοιχιών, ώστε να αντιπροσωπευθούν οι διεργασίες επιδιόρθωσης DNA και ανοσοποιητικού συστήματος.	144

Πίνακας 27: Όροι και κωδικοί των ασθενειών από το DisGeNET για τις τέσσερις βασικές κατηγορίες συγγενών παθήσεων που έχουν συσχετιστεί με έκθεση σε σωματιδιακή ακτινοβολία.	145
Πίνακας 28: Γονίδια που αντιστοιχούν σε κάθε κατηγορία συγγενών παθήσεων της έκθεσης σε σωματιδιακή ακτινοβολία από το DisGeNET.	147

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 1: (a) Τροχιές ιοντισμών ηλεκτρομαγνητικής ακτινοβολίας χαμηλής LET και (b)τροχιά ιοντισμών σωματιδιακής ακτινοβολίας υψηλής LET σε ένα κύτταρο.....	18
Εικόνα 2: Εξιδακτυκωμένη αναπαράσταση διάφορων τύπων ομαδοποιημένων βλαβών, συμπεριλαμβανομένων ομαδοποιημένων μονόκλωνων θραύσεων (SSB) και δίκλωνων θραύσεων (DSBs) που μπορούν να επέλθουν έπειτα από έκθεση σε ιοντίζουσα ακτινοβολία [22].	21
Εικόνα 3: Απεικόνιση των μηχανισμών απόκρισης βλαβών DNA [28].....	22
Εικόνα 4: Συνοπτική παρουσίαση των ρών εργασίας για την ανάλυση δεδομένων.	45
Εικόνα 5: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα πριν και μετά την εφαρμογή της συνάρτησης $\log_2()$, η οποία εκτελεί διόρθωση υποβάθρου και κανονικοποίηση, για την σειρά GSE80547.	52
Εικόνα 6: Αναπαράσταση με θηκογράμματα (boxplots) των μέσων εντάσεων έκφρασης ανά κατηγορία ετικέτας ποιότητας ανιχνευτή για την σειρά GSE80547	54
Εικόνα 7: Διάγραμμα της PCA για την σειρά GSE80547. Τα δείγματα φαίνεται να ομαδοποιούνται με βάση τα σετ πειραματικών επαναλήψεων και όχι με βάση την δόση της ακτινοβολίας.....	55
Εικόνα 8: Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE80547. Κάθε σημείο αναπαριστά ένα δείγμα όπου ο αριθμός στην αρχή του ονόματός του υποδηλώνει το σετ δειγμάτων και τα Ctrl, A, B, C την δόση της ακτινοβολίας.	57
Εικόνα 9: Διάγραμμα MDS για τα διορθωμένα δεδομένα της σειράς GSE80547. Κάθε σημείο αναπαριστά ένα δείγμα όπου ο αριθμός στην αρχή του ονόματός του υποδηλώνει το σετ δειγμάτων και τα Ctrl, A, B, C την δόση της ακτινοβολίας.	58
Εικόνα 10: Διαγράμματα της PCA για την σειρά GSE80547 πριν και μετά την διαδικασία διόρθωσης του Batch Effect. Στο δεύτερο διάγραμμα οι αρχικές συστάδες δειγμάτων με βάση το πειραματικό αντίγραφο έχουν απαλειφθεί και διακρίνεται ομαδοποίηση με βάση την δόση της ακτινοβολίας.	58
Εικόνα 11: Αναπαράσταση δεδομένων έκφρασης με θηκόγραμμα (boxplot) ανά δείγμα, για τα επεξεργασμένα δεδομένα της σειράς GSE81067, όπως λήφθηκαν από το αντίστοιχο αρχείο <code>series.matrix</code>	59
Εικόνα 12: Απεικόνιση των συστάδων στο διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE81067. Κάθε σημείο αναπαριστά ένα δείγμα ενώ οι 4 συστάδες που παρουσιάζονται, αντιστοιχούν στους τέσσερις τύπους ιστού που χρησιμοποιήθηκαν στο πείραμα (A549, HLF, HPEAC, THP-1). Error! Bookmark not defined.	
Εικόνα 13 Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE81067 που αντιστοιχούν σε φυσιολογικό ανθρώπινο ιστό.	61
Εικόνα 14: Αναπαράσταση δεδομένων έκφρασης με θηκόγραμμα (boxplot) ανά δείγμα, για τα επεξεργασμένα δεδομένα της σειράς GSE44282, όπως λήφθηκαν από το αρχείο <code>series.matrix</code>	62
Εικόνα 15: Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE44282. Κάθε σημείο αναπαριστά ένα δείγμα ενώ διαφαίνονται τρεις συστάδες που αντιστοιχούν στα τρία από τα πέντε βιολογικά αντίγραφα (επαναλήψεις).	63
Εικόνα 16: Διάγραμμα PCA για τα κανονικοποιημένα δεδομένα της σειράς GSE44282. Με χρώμα υποδηλώνεται η δόση σε συνδυασμό με τον τύπο της ακτινοβολίας ενώ τα σχήματα αντιστοιχούν στις διαφορετικές πειραματικές επαναλήψεις.....	65

Εικόνα 17: Διάγραμμα PCA για τα διορθωμένα δεδομένα της σειράς GSE44282, όπου οι συστάδες με βάση τις διαφορετικές πειραματικές επαναλήψεις έχουν εξαλειφθεί.	66
Εικόνα 18: Αναπαράσταση μη επεξεργασμένων δεδομένων έκφρασης με θηκογράμματα (boxplots) της σειράς GSE64375 με χρήση του πακέτου oligo.....	66
Εικόνα 19: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε κάθε δείγμα των πρωταρχικών δεδομένων της σειράς GSE64375.	68
Εικόνα 20: Αναπαράσταση επεξεργασμένων δεδομένων έκφρασης της σειράς GSE64375 με θηκογράμματα έπειτα από την εφαρμογή του αλγορίθμου RMA.....	69
Εικόνα 21: Ιστόγραμμα για τις μέσες επεξεργασμένες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς GSE64375.	70
Εικόνα 22: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE64375.	72
Εικόνα 23: Διάγραμμα MDS έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE64375. Κάθε σημείο αναπαριστά ένα δείγμα ενώ διαφαίνονται τέσσερις συστάδες που αντιστοιχούν στους τέσσερις διαφορετικούς εθελοντές (πειραματικά αντίγραφα).	73
Εικόνα 24: Διάγραμμα MDS των διορθωμένων δεδομένων της σειράς GSE64375. Κάθε σημείο αναπαριστά ένα δείγμα ενώ οι συστάδες που αντιστοιχούσαν στους τέσσερις διαφορετικούς εθελοντές (βιολογικά αντίγραφα) έχουν εξαλειφθεί.	74
Εικόνα 25: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς GSE63952.	75
Εικόνα 26: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς GSE63952.....	75
Εικόνα 27: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά GSE63952.	76
Εικόνα 28: Διαγράμματα PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE63952. Στο πρώτο διάγραμμα φαίνεται ότι οι διαφορές στην έκφραση διαμορφώνονται βάση των ωρών συλλογής των δειγμάτων έπειτα από ακτινοβολία. Από το δεύτερο διάγραμμα φαίνεται ότι οι συστάδες είναι ανεξάρτητες των πειραματικών αντιγράφων, με εξαίρεση την επανάληψη 9 (rep9).	77
Εικόνα 29: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς E-MTAB-5754.....	78
Εικόνα 30: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-5754.	78
Εικόνα 31: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-5754.	79
Εικόνα 32: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-5754.....	79
Εικόνα 33: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς E-MTAB-5761.....	80
Εικόνα 34: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-5761.	80
Εικόνα 35: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-5761.	81
Εικόνα 36: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-5761.....	81

Εικόνα 37: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μεταγράφο σε όλα τα δείγματα της σειράς E-MTAB-3463.	82
Εικόνα 38: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-3463.	82
Εικόνα 39: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-3463.	83
Εικόνα 40: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-3463. Πρώτη κύρια συνιστώσα αποτελεί ο τύπος και η δόση της ακτινοβολίας και δεύτερη τα πειραματικά αντίγραφα.	83
Εικόνα 41: Απεικόνιση των συστάδων που προέκυψαν με εφαρμογή της μεθόδου k-means clustering για k=4, στην μεγαλύτερη αρχική συστάδα του δικτύου πρωτεϊνών επιδιόρθωσης DNA, όπως προέκυψε από την πρώτη εφαρμογή της μεθόδου για k=5.	96
Εικόνα 42: Απεικόνιση των συστάδων 2-4 που προέκυψαν με εφαρμογή της μεθόδου k-means clustering για k=5 στο δίκτυο πρωτεϊνών επιδιόρθωσης DNA.	97
Εικόνα 43: Απεικόνιση της 1 ^{ης} συστάδας που προέκυψε με εφαρμογή της μεθόδου k-means clustering για k=7 στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.	98
Εικόνα 44: Απεικόνιση των συστάδων 2-4 που προέκυψαν με εφαρμογή της μεθόδου k-means clustering για k=7 στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.	99
Εικόνα 45: Απεικόνιση των συστάδων 5-7 που προέκυψαν με εφαρμογή της μεθόδου k-means clustering για k=7 στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.	99
Εικόνα 46: Διαγράμματα Venn των διαφορικώς εκφρασμένων γονιδίων από τις τρεις σειρές δεδομένων που περιείχαν δείγματα ακτινοβολημένα με ιοντίζουσα υψηλής και χαμηλής τιμής LET, σε συνδυασμό με τα «σημαντικά» γονίδια που επιλέχθηκαν από την βιβλιογραφική αναζήτηση. Οι γκρίζες περιοχές υποδεικνύουν τα γονίδια που αναπαραστάθηκαν με γραφικές παραστάσεις για κάθε σειρά δεδομένων (Πίνακας 12). ...	100
Εικόνα 47: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικώς εκφρασμένα γονίδια της σειράς E-MTAB-5754. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας ακτινοβολημένα με ακτίνες-X και ιόντα σιδήρου, δόσης των 2Gy και συλλεγμένα 24 ώρες μετά την ακτινοβολήση.	109
Εικόνα 48: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικώς εκφρασμένα γονίδια της σειράς E-MTAB-5761/3463. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα κύτταρα περιφερειακού αίματος ακτινοβολημένα με ακτίνες-X, ιόντα σιδήρου και ιόντα άνθρακα, δόσης του 1Gy και συλλεγμένα 8 ώρες μετά την ακτινοβολήση.	110
Εικόνα 49: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικώς εκφρασμένα γονίδια της σειράς GSE44282. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα βρογχικά επιθηλιακά κύτταρα ακτινοβολημένα με ακτίνες-X και ιόντα σιδήρου, δόσης του 1Gy και συλλεγμένα 12-24 ώρες μετά την ακτινοβολήση.	110
Εικόνα 50: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών E-MTAB-3464, E-MTAB-5754 και GSE44282, έπειτα από ακτινοβολήση με υψηλές δόσεις ακτίνων-X. Όλα τα δείγματα συλλέχθηκαν έως 8 ώρες έπειτα από την ακτινοβολήση και περιλαμβάνουν 3 τύπους ανθρωπίνου ιστού: μονοπύρηνια κύτταρα περιφερειακού αίματος, ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας και βρογχικά επιθηλιακά κύτταρα.	112
Εικόνα 51: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών GSE80547 και GSE81067 έπειτα από ακτινοβολήση με υψηλές δόσεις σωματιδίων άλφα. Όλα τα δείγματα συλλέχθηκαν 24 ώρες έπειτα από την ακτινοβολήση και περιλαμβάνουν 3 τύπους ανθρωπίνου ιστού: ανθρώπινα μονοπύρηνια κύτταρα περιφερειακού αίματος, ινοβλάστες πνευμόνων και πνευμονικά ενδοθηλιακά κύτταρα.	112
Εικόνα 52: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών GSE63952 και GSE64375, που περιλαμβάνουν δείγματα από ανθρώπινα απομονωμένα λευκοκύτταρα	

ακτινοβολημένα με βαρέα ιόντα υψηλών και χαμηλών δόσεων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-2 και 6 ωρών έπειτα από την ακτινοβόληση.....113

Εικόνα 53: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών E-MTAB-3464 και E-MTAB-5761, που περιλαμβάνουν δείγματα από ανθρώπινα μονοπύρρηνα κύτταρα περιφερειακού αίματος ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-X υψηλής δόσης. Τα δείγματα συλλέχθηκαν 8 ώρες έπειτα από την ακτινοβόληση.....113

Εικόνα 54: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια της σειράς E-MTAB-5754, που περιλαμβάνουν δείγματα από ανθρώπινα ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-X υψηλής δόσης. Τα δείγματα συλλέχθηκαν 24 ώρες έπειτα από την ακτινοβόληση.....114

Εικόνα 55: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια της σειράς GSE44282, που περιλαμβάνουν δείγματα από ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-X υψηλής δόσης. Τα δείγματα συλλέχθηκαν 12-24 ώρες έπειτα από την ακτινοβόληση.114

Εικόνα 56: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια της σειράς GSE44282, που περιλαμβάνουν δείγματα από ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα ακτινοβολημένα με ιόντα σιδήρου υψηλών και χαμηλών δόσεων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-4 και 24 ωρών έπειτα από την ακτινοβόληση.115

Εικόνα 57: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών GSE44282, GSE64375, GSE63952, E-MTAB-5761 και E-MTAB-5754 έπειτα από ακτινοβόληση με υψηλές και χαμηλές δόσεις βαρέων ιόντων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-4 και 24 ωρών έπειτα από την ακτινοβόληση και περιλαμβάνουν: ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα, απομονωμένα λευκοκύτταρα, μονοπύρρηνα κύτταρα περιφερειακού αίματος και ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας.116

Εικόνα 58: Διάγραμμα Venn των πειραματικών γονιδίων της ανάλυσης προς σύγκριση με τα γονίδια που έχουν συσχετιστεί με τις 4 προς μελέτη κατηγορίες συγγενών παθήσεων: νευροεκφυλιστικές παθήσεις, καρδιαγγειακές παθήσεις, παθήσεις του μεταβολισμού και τέλος παθήσεις του ανοσοποιητικού & αυτοάνοσα νοσήματα. Στα κουτιά παρατίθενται τα ονόματα των γονιδίων που βρέθηκαν κοινά με τα πειραματικά διαφορικώς εκφρασμένα.....132

Εικόνα 59: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: SLC12A2, TRPV1, HSPA1B, GPX1, HMGB1, SOD2, SIRT1, CD36, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με καρδιαγγειακές παθήσεις.133

Εικόνα 60: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: SLC22A4, PTPN22, ELANE, COL1A1, OAS1, CCL20, HLA-DQA1, ITGA4, GLIS3, IL1B, ATG16L1, SIRT1, SOD2, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με παθήσεις του ανοσοποιητικού.....134

Εικόνα 61: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: PLA2G4A, ACADVL, HSD11B1, CTC1, SIRT1, CD36, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με παθήσεις του μεταβολισμού....134

Εικόνα 62: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: SERPINA1, AIMP1, GPX3, APP, SOD2, SIRT1, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με νευροεκφυλιστικές παθήσεις ..135

1. ΕΙΣΑΓΩΓΗ

1.1. Ιοντίζουσες Ακτινοβολίες στην Βιολογία

Τα άτομα που συνθέτουν την ύλη αποτελούνται από πυρήνες, οι οποίοι περιέχουν θετικά φορτισμένα πρωτόνια που φέρουν θετικό φορτίο ίσο με το αρνητικό φορτίο του ηλεκτρονίου, και νετρόνια, τα οποία φέρουν μηδενικό φορτίο. Ο αριθμός των ηλεκτρονίων στο άτομο και επομένως ο αριθμός των πρωτονίων στον πυρήνα, που ονομάζεται ατομικός αριθμός, δίνει ένα στοιχείο τα μοναδικά χαρακτηριστικά του. Έτσι, τα ατομικά είδη καθορίζονται από τους αριθμούς πρωτονίων και νετρονίων. Τα νουκλίδια ενός στοιχείου με τον ίδιο αριθμό πρωτονίων αλλά διαφορετικούς αριθμούς νετρονίων καλούνται ισότοπα αυτού του στοιχείου. Αν και πολλά νουκλίδια είναι σταθερά, τα περισσότερα δεν είναι. Φυσικώς ασταθή άτομα μετασχηματίζονται αυθόρμητα, απελευθερώνοντας ενέργεια υπό μορφή ακτινοβολίας. Αυτός ο αυθόρμητος μετασχηματισμός ενός πυρήνα ονομάζεται ραδιενέργεια και η περίσσεια εκπεμπόμενη ενέργεια παράγει τους διαφορετικούς τύπους ιοντίζουσας ακτινοβολίας (*Ionizing Radiation-IR*) [1]. Στους τύπους αυτούς περιλαμβάνονται:

1. Ηλεκτρομαγνητικές ακτινοβολίες: ακτίνες-Χ και ακτίνες-γ
2. Σωματιδιακές ακτινοβολίες:
 - Σωματίδια-α (πυρήνες ηλίου)
 - Σωματίδια-β (ηλεκτρόνια και ποζιτρόνια)
 - Νετρόνια (χαμηλής ή υψηλής ενέργειας ουδέτερα σωματίδια)
 - Πρωτόνια και άλλα βαρέα ιόντα (π.χ. ιόντα άνθρακα και σιδήρου)

Οι φυσικές ραδιενεργές σειρές Θορίου, Ραδίου-Ουρανίου και Ακτινίου, περιλαμβάνουν αυτά τα ισότοπα στοιχεία, που είναι από την φύση τους ασταθή και τα οποία συνιστούν την περιβαλλοντική ακτινοβολία [2]. Επιπλέον υπάρχει η κοσμική ακτινοβολία, που προέρχεται από το βαθύ διάστημα. Αποτελεί μείγμα πολλών διαφορετικών τύπων ακτινοβολίας όπως σωματίδια-α, ηλεκτρόνια και άλλα σωματίδια υψηλής ενέργειας που αλληλεπιδρούν έντονα με την ατμόσφαιρα. Ως αποτέλεσμα, η κοσμική ακτινοβολία, στο επίπεδο του εδάφους της γης, συνίσταται πρωτίστως από

νετρόνια, ηλεκτρόνια, ποζιτρόνια και φωτόνια [3]. Επιπλέον υπάρχουν και τεχνητές πηγές IR στις οποίες περιλαμβάνονται οι ακτινοβολίες από την διάγνωση, την θεραπεία και την χρήση ραδιοϊσοτόπων στην ιατρική [4].

Η έκθεση των ζωντανών οργανισμών σε ιοντίζουσα ακτινοβολία μπορεί να προκύψει λοιπόν κατά την αλληλεπίδραση με τις διάφορες πηγές ακτινοβολίας, όπως η περιβαλλοντική ακτινοβολία, η κοσμική ακτινοβολία και οι τεχνητές πηγές. Επίσης σε ευρύτερο πλαίσιο συνεισφέρουν τα ραδιενεργά απόβλητα και η ραδιενεργή σκόνη από πυρηνικές δοκιμές, ενώ οφείλει να σημειωθεί η επαγγελματική έκθεση από πυρηνικούς αντιδραστήρες και επιταχυντές.

Από την άποψη της αλληλεπίδρασης με την ύλη, ο όρος IR περιγράφει τους τύπους ακτινοβολίας που μεταφέρουν αρκετή ενέργεια ώστε να απελευθερώσουν ηλεκτρόνια από ένα άτομο [5]. Κάνουμε λόγο είτε για άμεσο ιοντισμό, όπως με τα σωματίδια-α, τα σωματίδια-β και άλλα βαρέα ιόντα (HZE) που απαντώνται κυρίως στο διάστημα, είτε έμμεσο, όπως με τις ακτίνες-γ, τις ακτίνες-X και τα νετρόνια [6]. Βασικό χαρακτηριστικό των σωματιδιακών ακτινοβολιών, με τις οποίες και θα ασχοληθούμε στην εργασία αυτή, είναι οι υψηλές τιμές γραμμικής μεταφοράς/εναπόθεσης ενέργειας (*Linear Energy Transfer-LET*).

1.1.1. Γραμμική Εναπόθεση Ενέργειας

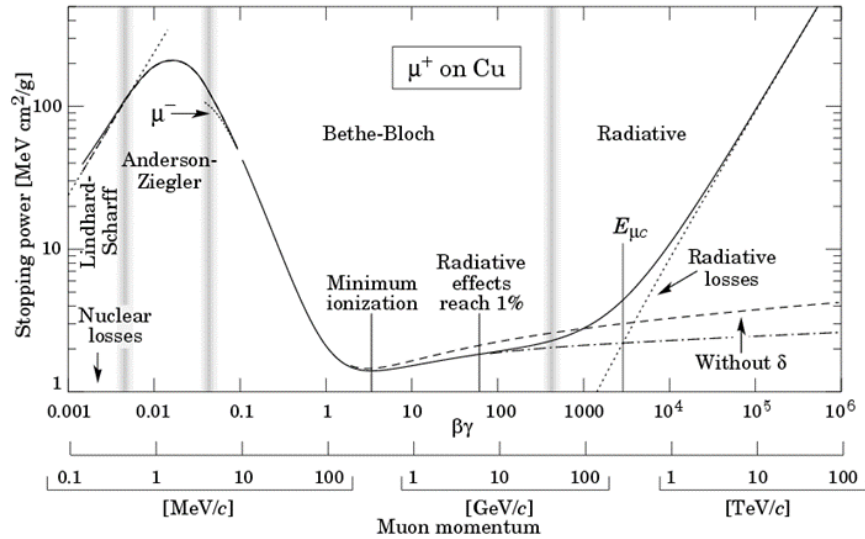
Ο ρυθμός απώλειας ενέργειας μιας δέσμης φορτισμένων σωματιδίων ως προς την απόσταση που διανύει σε ένα υλικό, δίνεται από την εξίσωση *Bethe-Bloch* (Εξίσωση 1), η οποία περιλαμβάνει παραμέτρους της ακτινοβολίας αλλά και του υλικού του στόχου.

$$-\frac{dE}{dx} = \left(\frac{1}{4\pi\epsilon_0}\right)^2 \frac{4\pi z^2 e^4}{m u^2} \frac{Z}{A} \ln\left(\frac{m u^2}{I}\right)$$

Εξίσωση 1: Γενική μορφή της εξίσωσης Bethe-Bloch, όπου $(1/4\pi\epsilon_0)$ σταθερή ποσότητα από τον Νόμο Coulomb, m , z και u καθορίζονται από την ακτινοβολία (μάζα, ατομικός αριθμός π.χ. $z=1$ για πρωτόνια, ταχύτητα), e το φορτίο του ηλεκτρονίου και I το δυναμικό ιοντισμού.

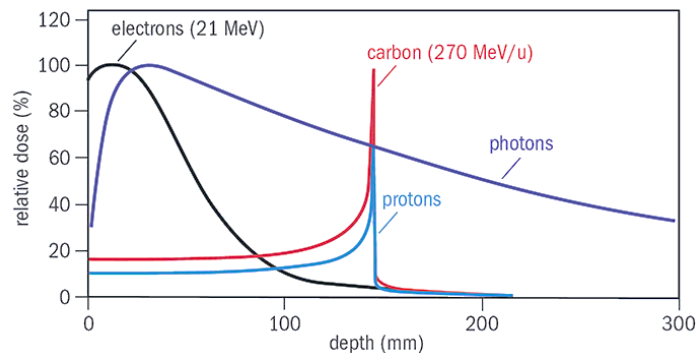
Ο ρυθμός απώλειας ενέργειας (ανασχετική ισχύς), εξαρτάται κυρίως από τον λόγο Z/A , που αποτελεί τον λόγο του ατομικού αριθμού προς τον μαζικό

αριθμό του στόχου-απορροφητή και περιγράφει την διαδρομή της ακτινοβολίας στην ύλη [7]. Για την απεικόνιση της εμβέλειας της ακτινοβολίας, η οποία καθορίζεται από τον ρυθμό με τον οποίο χάνει ενέργεια η δέσμη, λαμβάνονται υπόψιν κβαντικά και σχετικιστικά φαινόμενα (Γράφημα 1) [8].



Γράφημα 1: Ανασχετική ισχύς θετικά φορισμένων μιονίων σε χαλκό ως προς την σχετικιστική ορμή (http://meroli.web.cern.ch/Lecture_StragglngFunction.html).

Ειδικότερα στην περίπτωση μεγαλύτερων σωματιδίων (πρωτόνια, σωματίδια-α, βαρέα ιόντα) η απώλεια ενέργειας χαρακτηρίζεται από την καμπύλη Bragg (Γράφημα 2), η οποία σηματοδοτεί μέγιστη εναπόθεση ενέργειας στο τέλος της διαδρομής, η οποία ακολουθείται από την ολική απορρόφηση της δέσμης [9, 10].



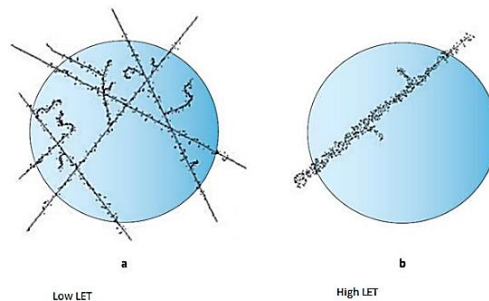
Γράφημα 2: Καμπύλη εναπόθεσης ενέργειας ως προς βάθος εισχώρησης διαφόρων ακτινοβολιών στο ίδιο υλικό. Ξεχωρίζει η περίπτωση πρωτονίων και ιόντων άνθρακα όπου σχηματίζεται η χαρακτηριστική καμπύλη Bragg (<https://cerncourier.com/a/the-changing-landscape-of-cancer-therapy/>).

Για τον προσδιορισμό της κατανομής της ενέργειας σε ένα υλικό άρα και σε έναν βιολογικό ιστό χρησιμοποιείται η LET και υποδηλώνει την απώλεια ενέργειας της ακτινοβολίας dE λόγω συγκρούσεων με τα ηλεκτρόνια του στόχου για συγκεκριμένο μήκος διαδρομής dx στον στόχο (Εξίσωση 2). Αν και στο Διεθνές Σύστημα Μονάδων (S.I.) μετριέται σε (N) *newton*, στην περίπτωση της ιατρικής φυσικής και της ακτινοθεραπείας, συνήθως χρησιμοποιείται το keV/μm.

$$LET \equiv \frac{dE}{dx}$$

Εξίσωση 2: Ορισμός γραμμικής εναπόθεσης ενέργειας (Linear Energy Transfer).

Κατά την έκθεση σε ηλεκτρομαγνητικές ακτινοβολίες δημιουργούνται πολλαπλές τροχιές με μικρό ποσοστό ιονισμών ανά τροχιά και μικρότερη εναπόθεση ενέργειας συνολικά. Για το λόγο αυτό χαρακτηρίζονται από χαμηλή τιμή LET. Αντίθετα στις σωματιδιακές ακτινοβολίες, παρατηρείται εμφάνιση λίγων τροχιών με πυκνούς ιοντισμούς κατά μήκος της τροχιάς τους στον απορροφητή και αρκετά μεγαλύτερη εναπόθεση ενέργειας [11] (Εικόνα 1). Έτσι χαρακτηρίζονται από υψηλή τιμή LET και συνολικά πιο στοχευμένη αλλά και καταστροφική δράση [12].



Εικόνα 1: (a) Τροχιές ιοντισμών ηλεκτρομαγνητικής ακτινοβολίας χαμηλής LET και (b) τροχιά ιοντισμών σωματιδιακής ακτινοβολίας υψηλής LET σε ένα κύτταρο.

Η ραδιενέργεια μιας πηγής μετριέται ως διασπάσεις ανά δευτερόλεπτο όπου 1Bq (*becquerel*) = 1 διάσπαση/sec ή εναλλακτικά με Ci (*curie*) όπου $1\text{Ci}=3.7 \cdot 10^{10}\text{Bq}$. Ως μέτρο της ραδιενέργειας που προσλαμβάνει ένας απορροφητής και κατ' επέκταση ένα βιολογικός ιστός, χρησιμοποιούμε την απορροφούμενη δόση, η οποία αποτελεί το ποσό της ενέργειας (dE) που

απορροφάται ανά μονάδα μάζας (dm) και ως μονάδα αυτής χρησιμοποιείται κυρίως το Gy (*gray*) όπου $1\text{Gy}=1\text{J/kg}=100\text{rad}$.

Διακρίνουμε έτσι την ακτινοβολία σε υψηλές και χαμηλές δόσεις θεωρώντας χαμηλές όσες είναι μικρότερες από 0.1Gy. Παρόλα αυτά σε πολλαπλά πειράματα θεωρούνται χαμηλές δόσεις έως και 0.5Gy [13]. Για την περιγραφή δόσης σε βιολογικό ιστό χρησιμοποιείται επίσης η ισοδύναμη δόση (*equivalent dose*), μονάδα της οποίας είναι το Sv (*sievert*), όπου $1\text{Sv}=100\text{rad}=100\text{Rem}$ και χρησιμοποιείται για την σύγκριση της σχετικής βιολογικής αποτελεσματικότητας (*Relative Biological Effectiveness-RBE*), η οποία αποτελεί μια εμπειρική τιμή που προκύπτει από πειραματικές μετρήσεις και εξαρτάται άμεσα από την τιμή LET [14]. Αντίστοιχα η ενεργός δόση (*effective dose*) χρησιμοποιείται για την εκτίμηση των κινδύνων έκθεσης του ανθρώπου στην ακτινοβολία και μετριέται επίσης σε Sv. Το ετήσιο ανεκτό όριο πρόσληψης ακτινοβολίας για τον οργανισμό που συσσωρεύεται και λειτουργεί αθροιστικά είναι 50 mSv [3, 6].

Επίδραση στον Βιολογικό Ιστό

Οι συνέπειες της ιοντίζουσας ακτινοβολίας στον βιολογικό ιστό εξαρτώνται συνολικά από διάφορες παραμέτρους, όπως ο τύπος της ακτινοβολίας, η δόση που θα απορροφηθεί, η χωρική και χρονική κατανομή της ενέργειας που μεταφέρθηκε αλλά και ο ίδιος ο βιολογικός στόχος. Αρχικά επέρχεται το φυσικό στάδιο αυτής της αλληλεπίδρασης, που σχετίζεται κυρίως με την μεταφορά ενέργειας μέσω ιοντισμών και πραγματοποιείται σε χρόνο μικρότερο των 10^{-18} sec [5].

Ακολουθεί το χημικό στάδιο, όπου τα συστατικά των κυττάρων αντιδρούν με ταχείες χημικές αντιδράσεις, το οποίο έχει ως αποτέλεσμα την μεταβολή της δομής των βιομορίων του ιστού, με έμφαση στις πρωτεΐνες και το DNA. Οι βασικοί μηχανισμοί στους οποίους οφείλονται οι βλάβες στο στάδιο αυτό, σχετίζονται είτε με άμεσες βλάβες στο βιομακρομόριο του κυττάρου, ή με έμμεσες βλάβες που προκύπτουν από τον σχηματισμό ελευθέρων ριζών, στο οποίο αναφερόμαστε και ως οξειδωτικό στρες [15]. Στην περίπτωση των άμεσων αποτελεσμάτων, η βιολογική επίδραση της IR σχετίζεται κυρίως με τον σχηματισμό δίκλωνων θραύσεων στο DNA (*Double Strand Breaks-DSBs*), που αποτελούν τους πιο θανατηφόρους

τύπους βλάβης του DNA και οδηγούν στον θάνατο των κυττάρων εάν δεν επιδιορθωθούν [16]. Το οξειδωτικό στρες, από την άλλη πλευρά, είναι στην ουσία μία διαδικασία ραδιόλυσης του νερού, που περιλαμβάνει τον σχηματισμό ενδιάμεσων, μερικώς μειωμένων ειδών οξυγόνου, τα οποία συλλογικά λαμβάνουν τον όρο των ενεργών ειδών οξυγόνου (*Reactive Oxygen Species-ROS*). Στη συνέχεια, τα ROS προκαλούν το σχηματισμό ριζών υδροξυλίου που παράγουν μία ποικιλία ανεπιθύμητων βιολογικών αντιδράσεων καθώς επιτίθενται στα δομικά και λειτουργικά μόρια [17], με αποτέλεσμα το γενικευμένο στρες σε επίπεδο κυττάρου. Οι ρίζες υδροξυλίου μπορεί να προκαλέσουν μονόκλωνες θραύσεις στο DNA (*Single-Strand Breaks-SSBs*) και μία πληθώρα αλλοιώσεων στις αζωτούχες βάσεις και τα μόρια σακχάρων που το συνιστούν, οι οποίες έχουν κυτταροτοξική ή μεταλλαξογόνο δράση [18]. Ταυτόχρονα μπορεί να δημιουργηθούν και διάδεσμοι μεταξύ των δύο συμπληρωματικών κλώνων του DNA [19].

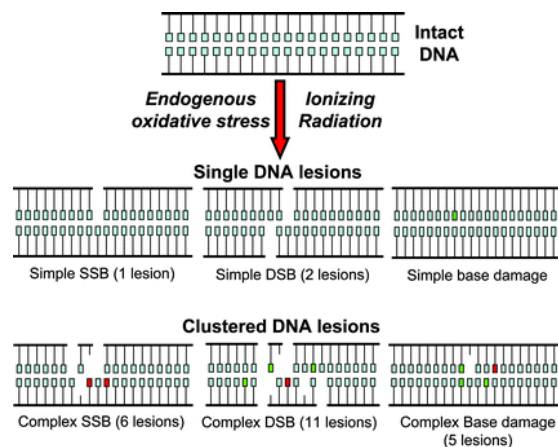
Το βιολογικό στάδιο της αλληλεπίδρασης λοιπόν περιγράφεται από μία ποικιλία αποκρίσεων και ένα τοπικό ή γενικευμένο οξειδωτικό στρες εξαιτίας της παραγωγής ελεύθερων ριζών στην ακτινοβολημένη περιοχή ή γενικότερα στον οργανισμό, μέσω συστημικών επιδράσεων. Αυτή η συνεχόμενη πρόκληση που αντιμετωπίζει το κύτταρο, μπορεί να οδηγήσει σε γενομική αστάθεια και σχηματισμό καρκίνου [20], ενώ ταυτόχρονα πυροδοτεί την απόκριση στις βλάβες DNA (*DNA Damage Response-DDR*), που αποτελεί την κύρια συνιστώσα των επιδράσεων της IR σε κυτταρικό επίπεδο [21].

1.1.2. Σύνθετες Βλάβες DNA

Οι βλάβες DNA που βρίσκονται σε κοντινή απόσταση (10-20 ζεύγη βάσεων), αναφέρονται ως σύνθετες ή ομαδοποιημένες βλάβες DNA (*complex DNA damage | DNA damage clusters*). Θεωρούνται εξαιρετικά σημαντικές βιολογικές βλάβες, καθώς είναι εξαιρετικά ανθεκτικές στην επιδιόρθωση ενώ σε πολλές περιπτώσεις δεν μπορούν να επιδιορθωθούν με καταστροφικά αποτελέσματα για το κύτταρο ή τον οργανισμό [22]. Οι

δύο βασικές κατηγορίες σύνθετων βλαβών DNA είναι οι δίκλωνες θραύσεις (DSBs) και οι μη-δίκλωνες οξειδωτικές [16].

Τέτοιου είδους βλάβες DNA αποτελούν την υπογραφή της high LET ακτινοβολίας ενώ η πολυπλοκότητα και η σοβαρότητα τους αυξάνεται με την αύξηση της τιμής LET. Σε αντίθεση με την ακτινοβολία χαμηλής LET, που δημιουργεί συστάδες (*clusters*) με έως και 10 αλλοιώσεις ανά σύμπλεγμα, η ακτινοβολία υψηλής LET δύναται να προκαλέσει ζημιά ακόμη μεγαλύτερης πολυπλοκότητας μέχρι και 25 αλλοιώσεις ανά σύμπλεγμα (Εικόνα 2). Τέλος, οι ομαδοποιημένες βλάβες DNA σε αντίθετους κλώνους παρουσιάζουν μία αξιοσημείωτη πολυπλοκότητα. Η επιδιόρθωσή τους καθίσταται δύσκολη και έτσι αυξάνεται η πιθανότητα πρόκλησης μεταλλαξιογόνων ή ακόμα και θανατηφόρων επιπτώσεων στο κύτταρο σε σχέση με τυχόν απομονωμένες αλλοιώσεις [23].



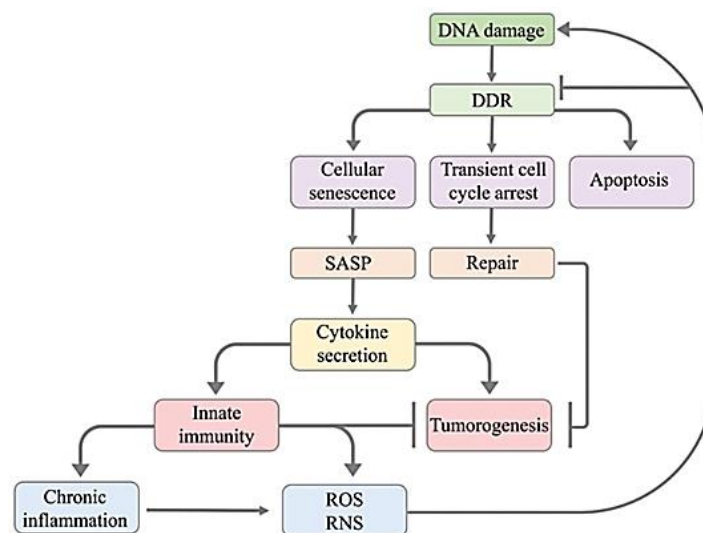
Εικόνα 2: Εξειδικευμένη αναπαράσταση διάφορων τύπων ομαδοποιημένων βλαβών, συμπεριλαμβανομένων ομαδοποιημένων μονόκλωνων θραύσεων (SSB) και δίκλωνων θραύσεων (DSBs) που μπορούν να επέλθουν έπειτα από έκθεση σε ιοντίζουσα ακτινοβολία [22].

1.1.3. Απόκριση στις Βλάβες DNA

Γενικά, η απόκριση στις βλάβες DNA (DDR) μπορεί να οριστεί ως το άθροισμα των λειτουργιών που ενορχηστρώνουν την ανίχνευση βλαβών DNA και τη μεταγωγή της ανάλογης σηματοδότησης (Εικόνα 3). Πιο συγκεκριμένα, η ογκο-κατασταλτική πρωτεΐνη TP53 αποτελεί βασικό παράγοντα των μηχανισμών DDR μέσω των αποκρίσεων που σηματοδοτεί, οι οποίες περιλαμβάνουν τη διακοπή του κυτταρικού κύκλου, την

επιδιόρθωση του DNA και την ενεργοποίηση της απόπτωσης [24]. Η ακριβής μετάβαση από τη φάση G1 του κυτταρικού κύκλου στη φάση S είναι ζωτικής σημασίας για τον ελεγχόμενο πολλαπλασιασμό των κυττάρων και η εσφαλμένη ρύθμισή αυτού προάγει την ογκογένεση [25]. Η διακοπή στη φάση G1 παρέχει στο κύτταρο τον απαιτούμενο χρόνο ώστε να αποκαταστήσει τις βλάβες DNA. Σε περίπτωση αποτυχίας της επιδιόρθωσης, τα επίπεδα της p53 μειώνονται και η δραστηριότητα της κινάσης CDK-cyclin επανεκκινείται, οδηγώντας σε είσοδο στη φάση S και σε πιθανή ενεργοποίηση της απόπτωσης. [26]. Οι λανθασμένα επιδιορθωμένες βλάβες DNA μπορούν να οδηγήσουν σε μεταλλάξεις ενώ μη επιδιορθωμένες βλάβες μπορεί να έχουν ως αποτέλεσμα την απόπτωση ή την κυτταρική γήρανση [27].

Η πρωτεΐνη p21 που κωδικοποιείται από το γονίδιο CDKN1A, ως βασικό γονίδιο που ενεργοποιείται από την p53, είναι υπεύθυνη για την αναστολή του κυτταρικού πολλαπλασιασμού σε απόκριση βλαβών DNA [28]. Επιπλέον, συσχετίζεται σε μεγάλο βαθμό με την επισκευή του DNA, ενώ παράλληλα συμβάλλει στην εκτέλεση της απόπτωσης. Αν και θεωρείται ότι παίζει βασικό ρόλο ως «φύλακας του γονιδιώματος», μπορεί εναλλακτικά να ενεργήσει ως μεσολαβητής γονιδιωματικής αστάθειας, κυτταρικής γήρανσης και καρκινογένεσης κάτω από ορισμένες συνθήκες, όπως η έκθεση σε IR [29].



Εικόνα 3: Απεικόνιση των μηχανισμών απόκρισης βλαβών DNA [30].

Η επιδιόρθωση του DNA αναφέρεται σε ένα σύνολο διαδικασιών μέσω των οποίων ένα κύτταρο αντιλαμβάνεται, αναγνωρίζει και επιδιορθώνει βλάβες στα μόρια DNA που κωδικοποιούν το γονιδίωμα του και διαδραματίζει κυρίαρχο ρόλο στην κανονική λειτουργία των κυττάρων και κατ' επέκταση όλου του οργανισμού [31]. Τα βασικά μονοπάτια επιδιόρθωσης του DNA περιλαμβάνουν την επισκευή μέσω εκτομής βάσης (*Base Excision Repair-BER*), την επισκευή μέσω εκτομής νουκλεοτιδίου (*Nucleotide Excision Repair-NER*) και την αποκατάσταση αναντιστοιχιών νουκλεοτιδίων (*Mismatch Repair-MMR*), το οποίο διορθώνει τις εσφαλμένα τοποθετημένες βάσεις και τους μικρούς βρόχους, οι οποίοι συναντώνται συνήθως στις επαναλαμβανόμενες αλληλουχίες DNA. Επιπλέον υπάρχουν ο ομόλογος ανασυνδυασμός (*Homologous Recombination-HR*) και η μη-ομόλογη σύνδεση άκρων (*Non-Homologous End Joining-NHEJ*) που δρουν ξεχωριστά ή συνδυαστικά για την επιδιόρθωση DSBs και σύνθετων συμβάντων όπως οι διάδεσμοι [32, 33].

Η έκθεση των κυττάρων σε IR προκαλεί ορισμένα επιπλέον συστημικά αποτελεσμάτων μέσω υποκίνησης των φλεγμονωδών σημάτων «κινδύνου» και της έμφυτης ανοσοαπόκρισης [15]. Οι χρόνιες φλεγμονώδεις αποκρίσεις που μπορεί να προκύψουν, προάγουν τον καρκίνο μέσω αυξημένης καταστροφής του DNA και αναστολής των μονοπατιών επιδιόρθωσης του [30]. Επιπλέον, η λανθασμένη ενεργοποίηση του μεταγραφικού παράγοντα NF-κB έχει συσχετιστεί με αριθμό φλεγμονωδών ασθενειών, ενώ η επίμονη αναστολή του NF-κB οδηγεί σε λανθασμένη ανάπτυξη ανοσοκυττάρων ή καθυστερημένη κυτταρική ανάπτυξη [34]. Η απορρύθμιση των μηχανισμών DDR γενικότερα μπορεί να προκαλέσει αρκετές ανθρώπινες ασθένειες και παθήσεις που σχετίζονται με την προδιάθεση σχηματισμού καρκίνου, την επιταχυνόμενη γήρανση και τις αναπτυξιακές ανωμαλίες [35].

Επομένως, για να συνοψίσουμε και να περιγράψουμε τα αποτελέσματα της έκθεσης σε IR σε οποιοδήποτε ζωντανό οργανισμό, υπάρχει μια ποικιλία αποκρίσεων που περιλαμβάνουν: Α) την παραγωγή ελεύθερων ριζών στην ακτινοβολημένη περιοχή ή γενικότερα σε ολόκληρο τον οργανισμό μέσω συστηματικών (μη στοχευμένων) επιδράσεων, Β) την ενεργοποίηση

μηχανισμών DDR και επιδιόρθωσης DNA Γ) την ενεργοποίηση της προφλεγμονώδους οδού [36].

Στην περίπτωση ακτινοβολίας υψηλής LET συγκεκριμένα, εμφανίζονται επιδράσεις στρες που επιμένουν, κυρίως λόγω της επαγωγής σύνθετων βλαβών DNA, όπου το οξειδωτικό και θερμικό στρες δεν περιορίζονται στον ιστό-στόχο, αλλά διαδίδονται και σε γειτονικά κύτταρα. Επίσης εμφανίζονται καινοφανείς αλληλουχίες στις ακτινοβολημένες περιοχές, που συμμετέχουν σε σύνθετες αναδιατάξεις και καταλήγουν σε ορατές χρωμοσωμικές αλλοιώσεις, οι οποίες συνήθως συνιστούν στη δημιουργία καρκινογένεσεων εάν δεν επέλθει κυτταρικός θάνατος [37, 38]. Κάνοντας λόγο λοιπόν για σύνθετες βλάβες DNA και συνεπώς τις κακώς επισκευασμένες βλάβες, έχουμε αύξηση των κυτταροτοξικών και μεταλλαξιογόνων επιδράσεων. Παρόλο που πολλαπλά μονοπάτια για την επαγωγή και διαιώνιση της γονιδιωματικής αστάθειας είναι πιθανά, θεωρείται ότι κυρίως οι DSBs και άλλες μορφές σύνθετων βλαβών εμπλέκονται σε αυτή, με αποτέλεσμα την γέννηση χρωμοσωμικών ανωμαλιών [19]. Συμπερασματικά, η έρευνα περί επαγωγής και επιδιόρθωσης σύνθετων βλαβών είναι απαραίτητη για τον υπολογισμό των παραγόντων κινδύνου της υψηλής LET ακτινοβολίας σχετικά με την πρόκληση καρκίνου αλλά και άλλων συγγενών παθήσεων, ιδιαίτερα στις περιπτώσεις που χρησιμοποιείται για θεραπευτικό σκοπό [39].

1.1.4. Διαστημικές Αποστολές

Εκτός από την χρήση της ηλεκτρομαγνητικής IR (ακτίνες-X, ακτίνες-γ) για τη θεραπεία διαφόρων τύπου καρκίνου, η ακτινοθεραπεία με χρήση υψηλής LET σωματιδιακής IR ενδείκνυται για την αντιμετώπιση συγκεκριμένων όγκων λόγω των βελτιωμένων φυσικών ιδιοτήτων και της βιολογικής αποτελεσματικότητας που επιδεικνύει. Σε αντίθεση όμως με την κλινική, τοπική εφαρμογή τέτοιου τύπου ακτινοβολίας, οι αστροναύτες εκτίθενται σε χρόνια κοσμική ακτινοβολία, στο σύνολο του οργανισμού τους.

Η έκθεση στη γαλαξιακή κοσμική ακτινοβολία λοιπόν, η οποία αποτελείται από πρωτόνια, σωματίδια-α και βαρέα ιόντα, αποτελεί μείζονα παράγοντα κινδύνου κατά τη διάρκεια μακροχρόνιων διαστημικών αποστολών. Για

παράδειγμα, κατά τη διάρκεια μιας αποστολής στον Άρη, ένας αστροναύτης θα συσσωρεύσει σημαντικές δόσεις γαλαξιακής κοσμικής ακτινοβολίας της τάξης του 1 Sv [40]. Οι επιπτώσεις αυτού του ακραίου περιβάλλοντος για μεγάλες χρονικές περιόδους έχουν αντίκτυπο πρώτα από όλα στο ανοσοποιητικό σύστημα, γεγονός που θα μπορούσε να θέσει σε κίνδυνο την υγεία των αστροναυτών και κατ' επέκταση την επιτυχία της αποστολής [41]. Άλλες πιθανές επιπτώσεις είναι τα καρκινικά και μη καρκινικά φαινόμενα όπως ο σχηματισμός καταρράκτη και οι εκφυλιστικές παρενέργειες στο κεντρικό νευρικό και καρδιαγγειακό σύστημα [42].

Η NASA (*National Aeronautics & Space Administration*) στη πρώτη της αποστολή που κράτησε έναν χρόνο, μέσω επιλογής ζεύγους μονοζυγωτικών διδύμων αστροναυτών, σύγκρινε τις επιπτώσεις του περιβάλλοντος μιας διαστημικής πτήσης στον έναν από τους διδύμους με τις επιπτώσεις του περιβάλλοντος της Γης στον έτερο, γενετικά πανομοιότυπο του. Η ανάλυση διαφορικής έκφρασης γονιδίων έδειξε εμπλουτισμένα μονοπάτια που σχετίζονται με αποκρίσεις βλαβών DNA. Επιπλέον, πολλά βιολογικά μονοπάτια που σχετίζονται με την ανοσοαπόκριση μεταβλήθηκαν σημαντικά, σε όλους τους κυτταρικούς τύπους, συμπεριλαμβανομένης τόσο της επίκτητης, όσο και της έμφυτης ανοσοαπόκρισης. Παρατηρήθηκε επίσης αύξηση του καρδιακού ρυθμού κατά 10% και μέτρια μείωση της συστολικής και μέσης αρτηριακής πίεσης. Η διαστολή της καρωτιδικής αρτηρίας που αναπτύχθηκε νωρίς κατά την πτήση, διατηρήθηκε σε όλη την διάρκεια της αποστολής. Συγχρόνως αυξήθηκε το πάχος της καρωτιδικής αρτηρίας και παρέμεινε έτσι έως 4 ημέρες μετά την προσγείωση. Γενικότερα, οι σημαντικές αλλαγές, στο σύνολό τους, φαίνονται ανάλογες με την χρονική διάρκεια της διαστημικής αποστολής και για αυτό το λόγο απαιτείται περαιτέρω μελέτη για αποστολές που ξεπερνούν το ένα έτος [43].

1.1.5. Συγγενείς Παθήσεις

Η υψηλής LET IR, όπως αναφέρθηκε παραπάνω, προκαλεί σύνθετες βλάβες DNA, με αποτέλεσμα την άμεση ενεργοποίηση αισθητήρων βλαβών DNA και πρωτεϊνών επιδιόρθωσης. Επειδή η αποκατάσταση των σύνθετων

βλαβών είναι χρονοβόρα και ελλιπής, μία έντονη και διατηρούμενη DDR μπορεί να διαμορφώσει την κυτταρική απόκριση ως προς πιθανές οξειές και όψιμες επιδράσεις. Έτσι το αντικείμενο μελέτης της εργασίας αυτής, διαμορφώνεται γύρω από την μελέτη των επιδράσεων της υψηλής LET IR. Η συνολική κυτταρική απόκριση πρέπει να μελετηθεί ώστε να διασαφηνιστούν οι μηχανισμοί ανάπτυξης των διαφόρων παρενεργειών και να αναπτυχθούν μέθοδοι πρόληψης σχετικά με τους κίνδυνους έκθεσης σε αυτή, τόσο για τις περιπτώσεις που χρησιμοποιείται για θεραπευτικό σκοπό, όσο για την έκθεση στην οποία υποβάλλονται οι αστροναύτες, καθώς η πλήρης θωράκιση τους στα διαστημικά σκάφη είναι ανέφικτη.

Αρχικά, η αρνητική επίδραση της κοσμικής ακτινοβολίας (υψηλής LET) στο ανοσοποιητικό σύστημα είναι ζήτημα προς μελέτη, δεδομένου ότι ορισμένες συνιστώσες του ανοσοποιητικού συστήματος συμπεριλαμβάνονται στους πιο ευαίσθητους στην ακτινοβολία ιστούς του οργανισμού [44]. Επιπρόσθετα, ο κατάλληλος συνδυασμός ακτινοθεραπείας και ανοσοθεραπείας έχει ανοίξει νέες προοπτικές για τη θεραπεία μεταστατικών και προχωρημένων περιπτώσεων καρκίνου [41].

Η κρανιακή ακτινοθεραπεία, που χρησιμοποιείται ώστε να αποτρέψει την πρόοδο κακοήθειας στον εγκέφαλο, μπορεί να προκαλέσει προοδευτικές και ίσως και ανεπανόρθωτες επιδράσεις στη γνωστική λειτουργία συμπεριλαμβανομένης της μάθησης, της μνήμης, της ταχύτητας επεξεργασίας πληροφορίας, της προσοχής, της γνωστικής ευελιξίας και της εκτελεστικής λειτουργίας [45]. Τέτοιου είδους θεραπείες έχουν ως αποτέλεσμα και άλλες συμπεριφορικές διαταραχές, που επηρεάζουν αρνητικά το άγχος, τη διάθεση και την κατάθλιψη [46, 47]. Για τους παραπάνω λόγους, οι μηχανισμοί μέσω των οποίων η έκθεση σε κοσμική ακτινοβολία μπορεί να διαταράξει το κεντρικό νευρικό σύστημα (*Central Nervous System-CNS*) προβληματίζουν την NASA, καθώς οι νευρογνωστικές (*neurocognitive*) επιπλοκές που μπορεί να προκύψουν βάζουν σε κίνδυνο την επιτυχία της αποστολής, την ασφάλεια των αστροναυτών, και την ποιότητα ζωής αυτών μετά την ολοκλήρωση της αποστολής [48].

Επιδημιολογικά δεδομένα από υπό θεραπεία ασθενείς με καρκίνο του μαστού, υποδηλώνουν ότι η σχετιζόμενη με IR καρδιαγγειακή νόσος (*Radiation Induced Cardio-Vascular Disease-RICVD*) μπορεί να προκύψει ακόμα και με δόσεις <2 Gy στοχευμένα στην περιοχή της καρδιάς [49]. Παρόλα αυτά, τα δεδομένα που προέρχονται από ασθενείς με καρκίνο οφείλουν να ερμηνεύονται προσεκτικά πριν από την παρέκταση τους στους αστροναύτες που αποτελούν μια υγιή ομάδα ατόμων, ενώ παράλληλα ο τρόπος και τα είδη της ακτινοβολίας που δέχονται ενδέχεται να διαφέρουν. Από την κοσμική ακτινοβολία, ειδικά τα βαρέα ιόντα (HZE) μπορούν να αυξήσουν σημαντικά τον κίνδυνο καρδιαγγειακής νόσου. Τα HZE μεταβάλλουν τη μεθυλίωση του DNA και την έκφραση γονιδίων που σχετίζονται με την καρδιαγγειακή λειτουργία και παθολογία, με αποτέλεσμα τις εκφυλιστικές μεταβολές ιστών και την επιταχυνόμενη έναρξη της αθηροσκλήρωσης. Παράλληλα αυξάνουν την προ-φλεγμονώδη απόκριση και εμποδίζουν την αγγειογένεση [50]. Μέχρι στιγμής λοιπόν είναι γνωστό ότι η RICVD επηρεάζει τόσο τον καρδιακό ιστό, όσο και το σύστημα των αγγείων. Ωστόσο, απαιτείται βελτίωση της εκτίμησης του κινδύνου για τις επανδρωμένες διαστημικές αποστολές.

Τέλος οφείλει να ερευνηθεί η συσχέτιση μεταξύ έκθεσης σε υψηλής LET IR και ανάπτυξης παθήσεων του μεταβολισμού. Η έκθεση σε IR γενικά, έχει συσχετιστεί με αυξημένο κίνδυνο ανάπτυξης χρόνιων μεταβολικών διαταραχών όπως η αντίσταση στην ινσουλίνη και ο διαβήτης τύπου 2 [51]. Παρόλα αυτά, διαταραχές στον οξειδωτικό μεταβολισμό έπειτα από έκθεση σε σωματιδιακή ακτινοβολία, έχουν συσχετιστεί με μειωμένη μεταφορά μιτοχονδριακών πρωτεϊνών και φαίνεται να παραμένουν για μεγάλο χρονικό διάστημα μετά την αποσύνθεση των ενεργών ειδών οξυγόνου (ROS). Η συνθήκη αυτή, ενδέχεται να διαταράζει τη σταθερότητα και τη δραστικότητα των πρωτεϊνών επιδιόρθωσης DNA [52]. Επιπλέον, ενώ οι οξειδωτικοί μηχανισμοί ρυθμίζονται αρνητικά, παρατηρείται ενεργοποίηση της ανοσολογικής απόκρισης [53]. Συμπερασματικά, οι τοξικές επιδράσεις της υψηλής LET IR και τα αποτελέσματα αυτών οφείλουν να μελετηθούν περαιτέρω διότι πιθανά αποτελούν έναν έμμεσο μονοπάτι για την εμφάνιση συγγενών παθήσεων.

1.2. Μικροσυστοιχίες

Οι μικροσυστοιχίες (*microarrays*) αποτελούν δισδιάστατα πλακίδια στα οποία δοκιμάζονται μεγάλες ποσότητες βιολογικού υλικού. Πιο συγκεκριμένα ως ανιχνευτές σε μία μικροσυστοιχία μπορεί να χρησιμοποιηθούν πρωτεΐνες, τμήματα ιστών, ανιχνευτές μεταβολιτών κ.α., που αντιπροσωπεύουν γνωστά ή άγνωστα γονίδια, και στις οποίες μπορεί να υβριδοποιηθεί το DNA ή το RNA προς μελέτη, αφού έχει πρώτα σημανθεί με φθορίζουσες χρωστικές [54]. Οι μικροσυστοιχίες επιτρέπουν την ανάλυση της γονιδιακής έκφρασης, της ποικιλότητας, της αλληλουχίας του DNA, των επιπέδων και τροποποιήσεων των πρωτεϊνών κ.λπ. ενώ δίνουν τη δυνατότητα ανάλυσης ολόκληρου του μεταγραφώματος ενός οργανισμού σε ένα μόνο πείραμα [55]. Υπάρχουν τρεις βασικές κατηγορίες μικροσυστοιχιών:

- Μικροσυστοιχίες πρωτεϊνών
- Μικροσυστοιχίες ιστών
- Μικροσυστοιχίες cDNA

Στην παρούσα εργασία θα ασχοληθούμε με την τρίτη κατηγορία, τις μικροσυστοιχίες cDNA, που αποτελούν μια μέθοδο υψηλής απόδοσης μέσω της οποίας μπορούμε να συλλέξουμε πληροφορίες για την διαφορική έκφραση των γονιδίων σε διάφορους οργανισμούς. Τα προτερήματα αυτού του τύπου μικροσυστοιχιών προκύπτουν από την εξαιρετική εξειδίκευση και δυνατότητα συσχέτισης που επιτυγχάνουν λόγω συμπληρωματικότητας των αζωτούχων βάσεων. Ένα αντίγραφο DNA ενός απομονωμένου γονιδίου αποτελεί σχεδόν τέλειο αντιδραστήριο για ειδική και ποσοτική ανίχνευση [56]. Οι μικροσυστοιχίες cDNA λοιπόν είναι μια διάταξη μεγάλου αριθμού ανιχνευτών DNA που αντιπροσωπεύουν συγκεκριμένους γενετικούς τόπους. Σε κάθε θέση ανιχνευτή (*probe*) συνήθως υπάρχει μια «ομάδα» ουσιών που σχετίζονται με μια συγκεκριμένη διεργασία [57]. Έτσι επιτρέπεται η ανάλυση της γονιδιακής έκφρασης και της ποικιλότητας με μαζική και παράλληλη επεξεργασία. Μπορεί κανείς να εξάγει χρήσιμες πληροφορίες για τη βιολογική λειτουργία ενός οργανισμού, βρίσκοντας ποια γονίδια επάγονται ή καταστέλλονται σε κάποια φάση του κυτταρικού

κύκλου, σε κάποια αναπτυξιακή στιγμή ή σε απόκριση σε ερεθίσματα του περιβάλλοντος, όπως η απόκριση σε ορμόνες, σε υψηλή θερμοκρασία ή έπειτα από έκθεση σε ιοντίζουσα ακτινοβολία [58].

Για την ετοιμασία του βιολογικού υλικού, αρχικά απομονώνεται το mRNA από τα υπό μελέτη δείγματα. Στη συνέχεια, μέσω αντίστροφης μεταγραφής, λαμβάνεται το αντίστοιχο συμπληρωματικό DNA (cDNA), το οποίο σημαίνεται με φθορίζουσα χρωστική [56]. Κατόπιν ακολουθεί η υβριδοποίηση των σημασμένων στόχων με τους ανιχνευτές της μικροσυστοιχίας. Σε αυτό το στάδιο, λόγω της συμπληρωματικότητας των βάσεων τμήματα του γενετικού υλικού των δειγμάτων «κολλάνε» στο πλακίδιο σε διαφορετικές αναλογίες. Το επόμενο βήμα είναι η σάρωση της επιφάνειας της μικροσυστοιχίας, απ' όπου προκύπτει μια ψηφιακή εικόνα, που προέρχεται από τη διέγερση των μορίων σήμανσης που βρίσκονται στους στόχους και φθορίζουν σε συγκεκριμένα μήκη κύματος, φανερώνοντας τις συγκεντρώσεις σε κάθε ανιχνευτή. Κατ' αυτό τον τρόπο, αυξημένη φωτεινότητα σημαίνει ποσοτικά αυξημένη έκφραση γονιδίου στο δείγμα [59].

Στη συνέχεια προχωράμε σε ανάλυση των δεδομένων αυτών, από την οποία προκύπτουν ποσοτικοποιημένα δεδομένα για τις εκφράσεις [60, 61]. Τα δεδομένα αυτά επεξεργάζονται στον ηλεκτρονικό υπολογιστή με πληθώρα αλγορίθμων, ώστε να απαλειφθούν τα σφάλματα και να προκύψουν συμπεράσματα τα οποία ο άνθρωπος δε θα μπορούσε να εξάγει, λόγω του μεγάλου όγκου πληροφορίας.

1.2.1. Κατασκευή & Λειτουργία Μικροσυστοιχιών

Ανάλογα με τον τρόπο κατασκευής της μικροσυστοιχίας, μια γενική κατηγοριοποίηση που μπορούμε να κάνουμε είναι: Α) οι τυπωμένες μικροσυστοιχίες Β) οι μικροσυστοιχίες με χρήση σφαιριδίων Γ) οι μικροσυστοιχίες «επιτόπιας σύνθεσης» [62]. Σε αυτό το κεφάλαιο θα επικεντρωθούμε στους δύο τύπους που χρησιμοποιήθηκαν σε αυτή την εργασία.

1.2.1.1. *Illumina BeadArrays*

Οι μικροσυστοιχίες με χρήση σφαιριδίων κατασκευάζονται μέσω τοποθέτησης ιχνηθετών σε δισδιάστατο υπόστρωμα γνωρίζοντας εκ των προτέρων τη θέση του καθενός. Η τεχνολογία της Illumina ωστόσο βασίζεται στην τυχαία εναπόθεση σφαιριδίων (BeadArray®) σε υπόστρωμα, συνήθως γυαλί. Η τεχνολογία Illumina BeadArray λοιπόν, αφορά σε μικροσυστοιχίες ολιγονουκλεοτιδίων, ακινητοποιημένων σε μαγνητικά σφαιρίδια πυριτίου, τα οποία είναι τοποθετημένα τυχαία σε «πηγάδια» διαμέτρου της τάξης μερικών μm στην επιφάνεια ενός υποστρώματος με δεδομένη διάταξη. Τα πηγάδια είναι διαμορφωμένα ώστε να χωράνε ένα σφαιρίδιο το καθένα.

Τα ολιγονουκλεοτίδια που χρησιμοποιούνται για μικροσυστοιχίες γονιδιακής έκφρασης είναι μεγέθους 79 βάσεων και αποτελούνται από 2 μέρη. Το πρώτο μέρος είναι ο ανιχνευτής (*probe*) μεγέθους 50 βάσεων. Βρίσκεται στην 5' θέση και έχει σχεδιαστεί για να υβριδίζει ακολουθίες γονιδίων με γνωστές αλληλουχίες κατατεθειμένες σε δημόσια αποθετήρια. Οι ανιχνευτές αυτοί εναποτίθενται τυχαία κατά μήκος της συστοιχίας. Μια μικροσυστοιχία 12-θέσεων για τον άνθρωπο περιέχει περίπου 40.000 ανιχνευτές. Το δεύτερο μέρος που βρίσκεται στο 3' άκρο αποτελείται από τις υπόλοιπες 29 βάσεις. Ονομάζεται τμήμα διεύθυνσης (*address*) και είναι μια μοναδική ακολουθία που δημιουργείται από την Illumina ώστε να επιτρέπει την μοναδική ταυτοποίηση χάριν κωδικοποίησης της εκάστοτε θέσης. Οι αλληλουχίες του *address* σχεδιάζονται με μεθόδους Βιοπληροφορικής ώστε να μην εμφανίζουν ομολογία με γενωμικές ακολουθίες γονιδίων για τον υπό μελέτη οργανισμό. Ονομάζονται *IllumiCodes* και αποτελούν το βασικό εργαλείο για την αποκωδικοποίηση της συστοιχίας. Κάθε σφαιρίδιο καλύπτεται από εκατοντάδες χιλιάδες αντίγραφα ενός συγκεκριμένου ολιγονουκλεοτιδίου, το οποίο εμφανίζεται περίπου 30 φορές σε διάφορες θέσεις της συστοιχίας [63].

1.2.1.2. Affymetrix GeneChips

Η δημιουργία αυτού του είδους μικροσυστοιχιών πραγματοποιείται με την επιτόπια σύνθεσή τους πάνω σε ένα στερεό φορέα. Η σύνθεση αυτή δημιουργείται μέσω ομοιοπολικού δεσμού μεταξύ του υδροξυλίου του 5' άνθρακα του σακχάρου του τελευταίου συνδεδεμένου νουκλεοτιδίου και της φωσφορικής ομάδας του επόμενου νουκλεοτιδίου. Σε κάθε νουκλεοτίδιο που προστίθεται προσδένεται μία προστατευτική ομάδα για αποφυγή σύνδεσης περισσότερων της μίας βάσης σε κάθε κύκλο σύνθεσης. Στη συνέχεια, πριν ξεκινήσει ο επόμενος κύκλος σύνθεσης, η προστατευτική ομάδα μετατρέπεται σε ομάδα υδροξυλίου, είτε μέσω έκθεσης σε οξύ (χημική αποπροστασία) ή μέσω έκθεσης σε φως (φωτολιθογραφία). Η τεχνολογία που εφαρμόζει η εταιρεία Affymetrix για την κατασκευή μικροσυστοιχιών DNA είναι η φωτολιθογραφία με χρήση масκών [64].

Οι μικροσυστοιχίες GeneChip® της εταιρείας Affymetrix αποτελούνται από μονόκλωνα 25μερή ολιγονουκλεοτίδια-ανιχνευτές. Η διαδικασία ξεκινάει με το γυάλινο πλακίδιο, το οποίο εμβαπτίζεται σε σιλάνιο (SiH_4) και τα μόρια του σιλανίου συνδυάζονται με το γυαλί. Ένα μόριο συνδέτης μαζί με ένα φωτοευαίσθητο μόριο προστίθενται σε κάθε μόριο σιλανίου. Ο συνδέτης είναι το σημείο έναρξης πρόσδεσης του πρώτου δεσοξυριβονουκλεοτιδίου. Κάθε νουκλεοτίδιο είναι φωτοχημικά τροποποιημένο αφού φέρει μια προστατευτική ομάδα η οποία απομακρύνεται μετά από επίδραση υπεριώδους ακτινοβολίας και αποτελεί το υπόστρωμα, όπου θα προσδεθεί το επόμενο νουκλεοτίδιο. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να σχηματιστούν συγκεκριμένα 25μερή ολιγονουκλεοτίδια σε κάθε θέση ανιχνευτική [65].

Κάθε νουκλεοτιδική αλληλουχία (γονίδιο) προς εξέταση, αντιπροσωπεύεται από 11 έως 20 μοναδικούς ανιχνευτές που δημιουργούνται μέσω υπολογιστή και είναι διάσπαρτοι στη μικροσυστοιχία, ώστε να αποφευχθεί η λανθασμένη εκτίμηση ποσοτικοποίησης της έκφρασης. Συνήθως, οι ανιχνευτές υβριδοποιούνται σε ανεξάρτητες περιοχές της αλληλουχίας, όμως ορισμένες φορές μπορεί να έχουν μικρή αλληλοεπικάλυψη (εφόσον αυτό κρίνεται απαραίτητο). Η ομάδα των ανιχνευτών που αφορά

συγκεκριμένο γονίδιο ή ομάδα παρόμοιων γονιδίων, είναι γνωστή σαν σύνολο ανιχνευτών (*probeset*), το οποίο παρέχει με υψηλή ακρίβεια τον υπολογισμό της έκφραση του γονιδίου στόχου [64].

1.2.2. Ανάλυση Δεδομένων Μικροσυστοιχιών

Σε αυτή την εργασία επιδιώκουμε να συνδυάσουμε δεδομένα από διαφορετικές μελέτες που χρησιμοποιούν μικροσυστοιχίες. Χάριν εγκυρότητας της ανάλυσης και στατιστικής αξιοπιστίας των αποτελεσμάτων, θα αναλυθούν τα πρωταρχικά δεδομένα, τα οποία βρίσκονται κατατεθειμένα σε δημόσια αποθετήρια και όχι εκείνα που αναλύθηκαν από τους ερευνητές. Στόχος είναι να εντοπιστούν τα συνολικά στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια σε κάθε μελέτη.

Βασική διαδικασία κατά την επεξεργασία πρωταρχικών δεδομένων μικροσυστοιχιών είναι η κανονικοποίηση των τιμών των εντάσεων έκφρασης. Η ανάγκη αυτή προκύπτει λόγω πολλαπλών ανεπιθύμητων παρεκκλίσεων που προκύπτουν από τεχνικά «λάθη». Αυτές οι παρεκκλίσεις οφείλονται σε μεταβολές των συνθηκών του εκάστοτε πειράματος που δεν μπορούν να αποφευχθούν και που σε συνδυασμό με τα βιολογικά αποτελέσματα μπορεί να οδηγήσουν σε ψευδείς ανακαλύψεις. Η κανονικοποίηση ποσοστημορίου (*quantile normalization*), την οποία και θα χρησιμοποιήσουμε, αναπτύχθηκε εξ αρχής και στοχευμένα για μικροσυστοιχίες γονιδιακής έκφρασης [66]. Η μέθοδος αυτή θεωρεί ότι η στατιστική κατανομή κάθε δείγματος είναι η ίδια. Η μέση κατανομή, η οποία λαμβάνεται από τον μέσο όρο έκφρασης των γονιδίων σε κάθε δείγμα, χρησιμοποιείται ως αναφορά έτσι ώστε όλες οι κατανομές να είναι συγκρίσιμες μεταξύ τους. Έτσι δημιουργούνται πανομοιότυπες κατανομές ως προς τις στατιστικές τους ιδιότητες. Παράλληλα απαιτείται οι κανονικοποιημένες τιμές να λογαριθμίζονται με βάση το 2 καθώς υπακούουν στον νόμο του Benford ή νόμο του πρώτου ψηφίου [67]. Ο νόμος αυτός προκύπτει από την φαινομενολογική παρατήρηση της κατανομής της συχνότητας των αρχικών ψηφίων σε πολλά φυσικά σύνολα αριθμητικών δεδομένων και υποδηλώνει ότι σε αυτά τα σύνολα τα μη σημαντικά ψηφία δυσανάλογα συχνά απαντώνται ως σημαντικά ψηφία. Η

λογαρίθμιση λοιπόν είναι απαραίτητη για την εξάλειψη αυτού του φαινομένου [68].

Η διαδικασία της κανονικοποίησης ανήκει στο ευρύτερο στάδιο της προ-επεξεργασίας των δεδομένων μικροσυστοιχιών έκφρασης. Κατά την προ-επεξεργασία θα πρέπει να ληφθούν υπόψη ο συνολικός πειραματικός σχεδιασμός, τα τεχνικά χαρακτηριστικά της κάθε μικροσυστοιχίας, η βιολογική πληροφορία που συνθέτει τους ανιχνευτές, καθώς και οι πρακτικές σωστού χειρισμού μεγάλου όγκου δεδομένων. Πριν από την διαδικασία αυτή οφείλει να γίνει διόρθωση των εκφράσεων βάση του σήματος υποβάθρου, διαδικασία που εξαρτάται από την τεχνολογία της εκάστοτε πλατφόρμας. Στη συνέχεια τα δεδομένα έκφρασης υπόκεινται σε διαδικασίες φιλτραρίσματος βάση τεχνικών κατασκευαστικών χαρακτηριστικών και βιβλιοθηκών σχολιασμού σχετικά με την αντιστοιχία ανιχνευτή-γονιδίου. Για την επιθεώρηση των δεδομένων κατά τη διάρκεια της προ-επεξεργασίας υπάρχει μια ποικιλία διαγραμμάτων όπως τα θηκογράμματα (*boxplots*), τα ιστογράμματα, τα διαγράμματα ανάλυσης κυρίων συνιστωσών (*Principal Component Analysis-PCA*) και τα διαγράμματα πολυδιάστατης κλιμάκωσης (*Multi-Dimensional Scaling-MDS*).

Το *boxplot* αποτελεί μια μέθοδο γραφικής απεικόνισης αριθμητικών δεδομένων στην περιγραφική στατιστική με χρήση τεταρτημορίων. Οι αποστάσεις μεταξύ των διαφορετικών τμημάτων των κουτιών απεικονίζουν την διασπορά και την ασυμμετρία των δεδομένων [69] ενώ οι ακραίες τιμές αναπαρίστανται ως τελείες (*outliers*). Η κεντρική γραμμή υποδηλώνει την θέση της διαμέσου (*median*), που περιλαμβάνει το 50% των παρατηρήσεων. Το μέγεθος του κουτιού στο άνω και κάτω τεταρτημόριο, πάνω και κάτω από την διάμεσο αντίστοιχα, δίνει εικόνα για την συμμετρικότητα της κατανομής της μεταβλητής. Οι μύστακες καθορίζουν την μέγιστη και ελάχιστη τιμή εξαιρώντας τις ακραίες τιμές [70].

Η μέθοδος PCA, αποτελεί μία γραμμική μέθοδο συμπίεσης δεδομένων η οποία βασίζεται στον επαναπροσδιορισμό των συντεταγμένων ενός συνόλου δεδομένων (στην περίπτωση μας τις εκφράσεις ανά δείγμα) σε ένα άλλο σύστημα συντεταγμένων. Αυτές οι νέες συντεταγμένες είναι το

αποτέλεσμα ενός γραμμικού συνδυασμού προερχόμενου από τις αρχικές μεταβλητές που εκπροσωπούνται σε ορθογώνιο άξονα. Για το λόγο αυτό, η πρώτη κύρια συνιστώσα διατηρεί τις περισσότερες πληροφορίες για τα αρχικά δεδομένα και αναπαριστά την κατεύθυνση που μεγιστοποιεί τη μεταβλητότητα τους όταν προβάλλονται σε αυτόν τον άξονα. Η δεύτερη κύρια συνιστώσα είναι ανεξάρτητη από την πρώτη και αναπαριστά την κατεύθυνση μεταβλητότητας ως προς τον ορθογώνιο άξονα, κάθετα στην πρώτη. Η συνολική ποσότητα των συνιστωσών είναι ίση με τη ποσότητα των αρχικών μεταβλητών (αριθμός δειγμάτων) και περιέχει τις ίδιες στατιστικές πληροφορίες. Με εφαρμογή αυτής της μεθόδου επιτυγχάνεται μείωση του συνόλου των μεταβλητών, καθώς οι πρώτες κύριες συνιστώσες διατηρούν περισσότερο από το 90% της στατιστικής πληροφορίας από τα αρχικά δεδομένα [71]. Σχεδιάζοντας το διάγραμμα των συμπιεσμένων δεδομένων, όπου κάθε σημείο αναπαριστά ένα δείγμα, αναζητούμε τις παραμέτρους που καθορίζουν εντονότερα τις αλλαγές στην έκφραση.

Εναλλακτικά μπορεί να χρησιμοποιηθεί η μέθοδος MDS που χρησιμοποιείται για την συσχέτιση μεταβλητών βάση αποστάσεων. Από τον συνολικό πίνακα των εκφράσεων ανά δείγμα, προκύπτει ένας πίνακας αποστάσεων για κάθε ανιχνευτή-γονίδιο με τις σχετικές αποστάσεις έκφρασης για κάθε δυάδα δειγμάτων. Από το σύνολο των ανιχνευτών-γονιδίων υπολογίζεται ο μέσος όρος απόστασης για κάθε δυάδα δειγμάτων. Έτσι προκύπτει ο συνολικός πίνακας αποστάσεων μεταξύ των δειγμάτων. Με την προσέγγιση αυτή παίρνουμε υπόψη μας τη σχετική θέση που τοποθετείται κάθε δείγμα με βάση τις εκφράσεις, αφαιρώντας από την ανάλυση την υποκειμενικότητα που προκύπτει βάση διακυμάνσεων ανά ανιχνευτή-γονίδιο. Κατά τον σχεδιασμό του διαγράμματος οι θέσεις των παρατηρήσεων (δειγμάτων) αναπαρίστανται ως σημεία σε δισδιάστατη επιφάνεια προβολής ώστε να διαφαίνονται οι σχετικές αποστάσεις μεταξύ αυτών. Ονοματίζοντας κάθε σημείο-δείγμα με βάση τα πειραματικά ή βιολογικά χαρακτηριστικά του, λαμβάνουμε μια εικόνα για τις παραμέτρους που υποκινούν περισσότερο τις αλλαγές στην έκφραση [72].

Τελικό στάδιο αποτελεί η ανάλυση της διαφορικής έκφρασης για την ανάδειξη στατιστικά σημαντικών διαφορικά εκφρασμένων γονιδίων. Για το

σκοπό αυτό επιλέγουμε την εμπειρική Μπεύζιανή που βασίζεται στην ανάλυση γραμμικής παλινδρόμησης (Linear Regression) και χρησιμοποιεί την κοινή φύση των συμπερασμάτων που εξάγονται από τα δεδομένα μικροσυστοιχιών ώστε να αντλεί ξεχωριστά συμπεράσματα για κάθε γονίδιο. Στη ουσία υποθέτει ότι οι μέσες τιμές και οι διακυμάνσεις έχουν κατανομές που εκτιμώνται από τα δεδομένα. Αποτελεί λοιπόν μια διαδικασία κατά την οποία βελτιώνεται η εκ των προτέρων γνώση που έχουμε για μια παράμετρο (επεξεργασμένα δεδομένα έκφρασης), μέσω της πιθανοφάνειας στην τελική πληροφορία που θα έχουμε για αυτή [73, 74]. Η έκφραση κάθε γονιδίου μοντελοποιείται ως γραμμική συνάρτηση των επεξηγηματικών μεταβλητών (χαρακτηριστικά ακτινοβολήσης, βιολογικός ιστός, πειραματικά αντίγραφα κ.λπ.) ενώ γίνεται η εκτίμηση των γραμμικών παραμέτρων (συντελεστών) του μοντέλου. Η στατιστική σημαντικότητα των συντελεστών εξετάζεται με δοκιμή T-test [75]. Καθώς ο όγκος των δεδομένων για κάθε πείραμα μικροσυστοιχιών είναι αρκετά μεγάλος, επιλέγουμε να χρησιμοποιούμε την προσαρμοσμένη τιμή p-value (adjusted p-value) όπως προκύπτει από την μέθοδο διόρθωσης πολλαπλών ερωτημάτων *False Discovery Rate-FDR* [76]. Μέσω αυτής μειώνεται η πιθανότητα αποδοχής ενός αποτελέσματος ως σημαντικό λόγω μεγάλης στατιστικής και το οποίο στην πραγματικότητα είναι ψευδώς θετικό. Αυτή η συνάρτηση υπολογίζει και το λογαριθμημένο με βάση το 2, κλάσμα συγκρινόμενων εκφράσεων ($\log_2 \text{fold change} - \log_2 \text{FC}$), για κάθε γονίδιο σε κάθε σύγκριση, ο οποίος είναι χρήσιμος χάριν ποσοτικοποίησης και πρόσημου παρέχοντας μια αριθμητική τιμή για υπέρ ή υπό-έκφραση [77, 78].

2. ΜΕΘΟΔΟΣ & ΔΙΑΔΙΚΑΣΙΑ

Για την παρούσα εργασία χρησιμοποιήθηκε υπολογιστής Dell Inspiron 3576 (Intel Core i7 8550U CPU 1.80GHz) της Dell Inc., με περιβάλλον Microsoft Windows 10 Home. Σε αυτόν εγκαταστάθηκαν και χρησιμοποιήθηκαν: η γλώσσα προγραμματισμού της R v.3.6.1 [79] και η σουίτα του Bioconductor v.3.9 [80] στο περιβάλλον RStudio [81]. Για την διεξαγωγή ανάλυσης λειτουργικού εμπλουτισμού των στατιστικώς σημαντικών διαφορικώς εκφρασμένων γονιδίων χρησιμοποιήθηκαν χάριν σύγκρισης δύο διαδικτυακά εργαλεία: το WebGestalt [82] και το Enrichr [83]. Για την ανάδειξη κρίσιμων γονιδίων ενδιαφέροντος χρησιμοποιήθηκαν η βάση δεδομένων γονιδιακής οντολογίας (Gene Ontology-GO) [84], και το εργαλείο κατασκευής πρωτεϊνικών δικτύων STRING v 11.0 [85]. Τέλος για την αναζήτηση των πιθανών τρόπων με τους οποίους αναπτύσσονται οι συγγενείς παθήσεις που έχουν συσχετιστεί με την έκθεση σε υψηλής LET IR από την βιβλιογραφία, συνδυάστηκαν η πλατφόρμα του DisGeNET [86, 87], η οποία περιέχει συλλογές γονιδίων που σχετίζονται με τις ασθένειες του ανθρώπου και το διαδικτυακό εργαλείο DisNor [86] για τον σχεδιασμό των αντιστοίχων δικτύων.

2.1. Εισαγωγή

Το πρώτο βήμα για την έναρξη μιας μελέτης γονιδιωματικής είναι η διατύπωση ενός συγκεκριμένου βιολογικού ερωτήματος, προκειμένου να καθοριστεί ο συνολικός πειραματικός σχεδιασμός με τον καλύτερο δυνατό τρόπο. Εάν τα πρωταρχικά δεδομένα προς ανάλυση, προέρχονται από δημόσιο αποθετήριο, το ερώτημα αυτό θα καθορίσει τα πειράματα που θα επιλεγούν μέσω των τελικών χαρακτηριστικών των δειγμάτων τους. Εφόσον στοχεύουμε στη μελέτη της επίδρασης IR υψηλού LET, συνιστάται η επιλογή ενός μόνο οργανισμού καθώς τα διαθέσιμα πειράματα ήδη εμπεριέχουν πολλαπλές παραμέτρους όπως η δόση και ο τύπος της ακτινοβολίας, ο χρόνος συλλογής των δειγμάτων έπειτα από την ακτινοβολία και ο τύπος του ιστού. Για το σκοπό αυτό αποφασίσαμε να μελετήσουμε την διαφορική έκφραση γονιδίων στον ανθρώπινο ιστό μέσω

μεταγραφικού προφίλ, χρησιμοποιώντας cDNA μικροσυστοιχίες. Στα επιλεγμένα πειράματα αυτής της ανάλυσης λοιπόν χρησιμοποιούνται υγιή δείγματα από ανθρώπινο ιστό. Σε κάθε μελέτη υπάρχουν μη ακτινοβολημένα (*control samples*) και ακτινοβολημένα με υψηλής LET IR δείγματα όπως και με χαμηλής LET IR εάν περιλαμβάνονταν στην μελέτη. Απώτερος σκοπός στην μελέτη αυτή είναι η διαφορική έκφραση των γονιδίων μεταξύ ακτινοβολημένων και μη δειγμάτων με σκοπό την μελέτη και σύγκριση των επιδράσεων υψηλής και χαμηλής LET IR. Κατά αυτόν τον τρόπο το βιολογικό προφίλ της έκθεσης στον συγκεκριμένο τύπο ακτινοβολίας θα προκύψει μέσω λειτουργικής ανάλυσης εμπλουτισμού των διαφορικά εκφρασμένων γονιδίων σε συνδυασμό με άλλες μεθόδους Βιοπληροφορικής και Βιολογίας Συστημάτων.

2.2. Αναζήτηση Πρωτογενών Δεδομένων

2.2.1. Αποθετήρια Δεδομένων Γονιδιωματικής

Η αναζήτηση για τα πρωταρχικά μας δεδομένα έγινε στο Gene Expression Omnibus¹ (GEO) [88, 89] του Εθνικού Κέντρου Βιοτεχνολογικών Πληροφοριών (National Center for Biotechnology Information - NCBI), που αποτελεί δημόσιο αποθετήριο δεδομένων γονιδιακής έκφρασης, συμμορφωμένων προς τις απαιτήσεις του MIAME (Minimal Information About Microarray Experiments) [90, 91] και στο ArrayExpress² [92] ως βασικό αρχείο δεδομένων λειτουργικής γονιδιωματικής του Ευρωπαϊκού Εργαστηρίου Μοριακής Βιολογίας (European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI). Στο GEO τα δεδομένα αρχειοθετούνται ως εξής: κάθε πειραματική πλατφόρμα που αντιστοιχεί σε συγκεκριμένη τεχνολογία μικροσυστοιχίας, έχει ένα μοναδικό αριθμό GPL. Κάθε δείγμα έχει ένα μοναδικό αριθμό GSM και ανήκει σε μία πλατφόρμα. Κάθε σειρά δειγμάτων έχει ένα μοναδικό αριθμό GSE και αποτελεί σύνολο από ένα ή περισσότερα δείγματα (GSM) που ανήκουν σε μία ή

¹ <https://www.ncbi.nlm.nih.gov/geo/>

² <https://www.ebi.ac.uk/arrayexpress/>

περισσότερες πλατφόρμες. Το ίδιο GSM μπορεί να ανήκει σε περισσότερες από μία σειρές (GSE). Τέλος κάθε σύνολο δεδομένων έχει ένα μοναδικό αριθμό GDS και αποτελεί μια επιμελημένη συλλογή περισσότερων του ενός δειγμάτων που ανήκουν σε μία πλατφόρμα και μία σειρά. Στο ArrayExpress αντίστοιχα κάθε μελέτη διαθέτει ένα μοναδικό αριθμό πρόσβασης (E-MTAB). Ενώ στο παρελθόν είχε γίνει μία απόπειρα εισαγωγής μελετών από το GEO με αντίστοιχους κωδικούς πρόσβασης E-GEOD, καθώς πλέον το Gene Expression Atlas³ [93] επιλέγει πειράματα για επανεπεξεργασία και επιμέλεια απευθείας από το GEO και παράλληλα το ArrayExpress πρόκειται να ενσωματωθεί στη βάση δεδομένων BioStudies⁴ [94, 95], έχουν σταματήσει οι εκ νέου εισαγωγές δεδομένων από το GEO στο ArrayExpress.

Αναζητήθηκαν πειράματα μονοκαναλικών cDNA μικροσυστοιχιών που χρησιμοποιούσαν ανθρώπινα δείγματα οποιουδήποτε είδους ιστού, ακτινοβολημένα με IR υψηλής τιμής LET, οποιασδήποτε δόσης και περιείχαν τα αντίστοιχα μη ακτινοβολημένα δείγματα για να υπηρετήσουν ως δείγματα αναφοράς (*controls*). Εφόσον κάποιο από τα πειράματα περιείχε και δείγματα ακτινοβολημένα με IR χαμηλού LET, αυτά χρησιμοποιήθηκαν και αναλύθηκαν χάριν σύγκρισης. Σε αυτό το στάδιο είναι σημαντικό να σημειωθεί ότι το πειραματικό πρωτόκολλο θα πρέπει να υποδηλώνει ότι η έκθεση σε IR αποτελεί τη μοναδική συνθήκη για τα τελικά δείγματα. Επιπλέον τα πρωτογενή πειραματικά δεδομένα θα πρέπει να είναι διαθέσιμα προς λήψη, καθώς συνιστάται να εκτελείται ομοιογενώς η διαδικασία προ-επεξεργασίας και στατιστικής ανάλυσης για το σύνολο των συλλεχθέντων δεδομένων, προκειμένου να καταλήξουμε στις βέλτιστες λίστες διαφορικώς εκφρασμένων γονιδίων

Η αναζήτησή στο GEO έγινε ως προς τα σύνολα δεδομένων με την επιλογή της προχωρημένης αναζήτησης. Αρχικά περιορίσαμε τα σύνολα δεδομένων επιλέγοντας εκείνα που περιέχουν προφίλ γονιδιακής έκφρασης μέσω μικροσυστοιχίας και χρησιμοποιούν ως οργανισμό τον άνθρωπο. Επιπλέον

³ <https://www.ebi.ac.uk/gxa/home>

⁴ <https://www.ebi.ac.uk/biostudies/>

λόγω της περιπλοκότητας του όρου IR επιλέξαμε να χρησιμοποιήσουμε το αντίστοιχο MeSH Term [96] για την ιονίζουσα ακτινοβολία και τους όρους «high LET», «HZE particles» και «alpha particles» που υποδεικνύουν ιονίζουσα ακτινοβολία υψηλού LET:

```
((("expression profiling by array"[DataSet Type]) AND "homo sapiens"[Organism]) AND "radiation, ionizing"[Mesh Terms]) AND ("high LET" OR "HZE particles" OR "alpha particles")
```

Έτσι λάβαμε 20 σειρές δεδομένων, όπου ελέγχθηκαν για την καταλληλότητά τους βάση πειραματικού σχεδιασμού, πρωτοκόλλων εργαστηριακής επεξεργασίας, σχολιασμού δειγμάτων και τύπου μικροσυστοιχίας. Κατά αυτόν τον τρόπο καταλήξαμε σε 5 τελικές σειρές δεδομένων προς ανάλυση.

Κατά την αναζήτηση στο ArrayExpress εντοπίστηκαν 2 εκ των 5 παραπάνω σειρών. Επιπλέον χρησιμοποιώντας τις σύνθετες αναζητήσεις:

```
"ionizing radiation" AND "high-LET"
```

```
"ionizing radiation" AND "high LET"
```

και φιλτράροντας τα αποτελέσματα ως προς τον οργανισμό και τον τύπο του πειράματος μπορέσαμε να επιλέξουμε 2 ακόμα κατάλληλες σειρές δεδομένων. Έτσι καταλήξαμε σε σύνολο 7 σειρών προς μελέτη.

2.2.2. Πειραματικός Σχεδιασμός Επιλεγμένων Σειρών

GSE44282: Χρησιμοποιήθηκαν ανθρώπινα βρογχικά επιθηλιακά κύτταρα (HBEC3KT), που ακτινοβολήθηκαν με τρεις διαφορετικούς τύπους ακτινοβολίας: ιόντα σιδήρου (Fe) και ιόντα πυριτίου (Si) σε δόσεις 0.5 και 1 Gy καθώς και ακτίνες-X δόσεων 1 και 1 Gy. Το ολικό RNA λήφθηκε μετά από 1, 4, 12 και 24 ώρες έπειτα από την ακτινοβολήση, δίνοντας έτσι σύνολο 134 δειγμάτων εκ των οποίων τα 80 αποτελούσαν δείγματα αναφοράς. Τα δείγματα κατηγοριοποιούνται σε πέντε βιολογικά αντίγραφα σύμφωνα με τον σχεδιασμό του πειράματος και αναλύθηκαν με μικροσυστοιχίες Illumina HumanWG-6 v2.0 expression beadchip [97].

GSE80547: Σκοπός αυτής της μελέτης ήταν η ταυτοποίηση γονιδίων των οποίων η έκφραση επηρεάζεται από την έκθεση σε ακτινοβολία

σωματιδίων-α. Ανθρώπινα μονοπύρρηνα κύτταρα περιφερειακού αίματος εκτέθηκαν σε ακτινοβολία σωματιδίων άλφα, που προήλθε από πηγή Αμερικού (^{241}Am) σε δόσεις 0.5, 1 και 1.5 Gy. Τα κύτταρα συλλέχθηκαν 24 ώρες μετά την έκθεση, δίνοντας σύνολο 48 δειγμάτων που διαχωρίστηκαν σε 12 σετ με 1 δείγμα αναφοράς στο κάθε ένα. Το RNA εκχυλίστηκε και υποβλήθηκε σε ανάλυση μικροσυστοιχιών τεχνολογίας Illumina HumanHT-12 V4.0 expression beadchip [98].

GSE81067: Χρησιμοποιήθηκαν τέσσερις ανθρώπινες κυτταρικές σειρές: μετασχηματισμένα μονοκύτταρα (THP-1), μετασχηματισμένα πνευμονικά επιθηλιακά κύτταρα (A549), φυσιολογικά πνευμονικά ενδοθηλιακά κύτταρα (HPEAC) και φυσιολογικοί ινοβλάστες πνευμόνων (HFL-1). Οι κυτταρικοί τύποι εκτέθηκαν σε ακτινοβολία σωματιδίων άλφα που προήλθε από πηγή Αμερικού (^{241}Am) σε δόσεις 0.5, 1 και 1.5 Gy. Τα κύτταρα συλλέχθηκαν 24 ώρες μετά την έκθεση δίνοντας σύνολο 110 δειγμάτων, εκ των οποίων τα 40 αντιστοιχούν σε φυσιολογικό ιστό (HPEAC και HFL-1). Το RNA εκχυλίστηκε και υποβλήθηκε σε ανάλυση μικροσυστοιχιών τεχνολογίας Illumina HumanHT-12 V3.0 expression beadchip [99].

GSE63952: Πραγματοποιήθηκε μεταγραφική ανάλυση βάση δόσολογίας και χρονικής πορείας για έκθεση σε ακτινοβολία δόσεων 0.15, 0.30 και 1.5 Gy με χρονικά σημεία συλλογής των δειγμάτων 1, 2 και 6 ώρες μετά την έκθεση, χρησιμοποιώντας μικροσυστοιχίες Affymetrix® Human PrimeArrayTM. Χρησιμοποιήθηκαν δείγματα ανθρώπινου αίματος, τα οποία λήφθηκαν από δέκα εθελοντές και ακτινοβολήθηκαν με σωματίδια υψηλής ενέργειας (HZE ions). Τα λευκοκύτταρα απομονώθηκαν και μετρήθηκε η έκφραση γονιδίων. Δείγματα και από τους δέκα εθελοντές μετρήθηκαν σε 1 και 2 ώρες έπειτα από την ακτινοβολία, για τα τέσσερα επίπεδα δόσης συμπεριλαμβάνοντας την δόση αναφοράς των 0 Gy, ενώ δείγματα από τους εθελοντές 1,2,3,4 και 7 μετρήθηκαν στις 6 ώρες. Κατά αυτό τον τρόπο σύνολο 25 δειγμάτων αντιστοιχείται σε κάθε ένα από τα τέσσερα επίπεδα ακτινοβολίας δημιουργώντας συνολικά 100 δείγματα [100].

GSE64375: Πραγματοποιήθηκε μεταγραφική ανάλυση βάση δόσολογίας για έκθεση σε ακτινοβολία 0, 0.3, 1.5 και 3 Gy με χρονικό σημείο συλλογής των δειγμάτων τη 1 ώρα μετά την έκθεση, χρησιμοποιώντας μικροσυστοιχίες

Affymetrix® GeneChip® Human Gene 1.0 ST v1. Χρησιμοποιήθηκαν δείγματα ανθρωπίνου αίματος από τέσσερις εθελοντές και ακτινοβολήθηκαν με σωματίδια υψηλής ενέργειας (HZE ions). Τα λευκοκύτταρα απομονώθηκαν και μετρήθηκε η έκφραση γονιδίων. Τα δείγματα και για τους τέσσερις εθελοντές μετρήθηκαν στη 1 ώρα μετά την ακτινοβολήση και για τα τέσσερα επίπεδα δόσεων, με αποτέλεσμα την ύπαρξη τεσσάρων επαναλήψεων σε κάθε επίπεδο δόσης. Κατά αυτό τον τρόπο σύνολο 4 δειγμάτων αντιστοιχείται σε κάθε ένα από τα τέσσερα επίπεδα ακτινοβολίας δημιουργώντας συνολικά 16 δείγματα [101].

E-MTAB-5761|3463: Η μελέτη αυτή χωρίζεται σε δύο σύνολα δεδομένων με τα δείγματα να αποτελούνται από ανθρώπινα μονοπύρρηνα κύτταρα περιφερειακού αίματος, που συλλέχθηκαν 8 ώρες μετά την ακτινοβολήση. Το πρώτο σύνολο απαρτίζεται από 4 σετ των 4 δειγμάτων ακτινοβολημένα με ιόντα σιδήρου (Fe) και άνθρακα (C) με τα αντίστοιχα δείγματα αναφοράς από 8 εθελοντές δίνοντας τελικά 16 δείγματα. Στο δεύτερο σύνολο, τα δείγματα αντιστοιχούν σε 3 σετ των 20 δειγμάτων ακτινοβολημένα με ακτίνες-X των 0.1 και 1 Gy και τα αντίστοιχα δείγματα αναφοράς (σύνολο 60), που λήφθηκαν από 10 εθελοντές. Το RNA εκχυλίστηκε και υποβλήθηκε σε ανάλυση μικροσυστοιχιών τεχνολογίας Affymetrix GeneChip Human Gene 1.0 ST Array [102].

E-MTAB-5754: Προκειμένου να προσδιοριστούν οι μοριακοί μηχανισμοί επίδρασης ακτινοβολίας υψηλού LET (ιόντα σιδήρου) και χαμηλού LET (ακτίνες-X) σε ενδοθηλιακά κύτταρα της στεφανιαίας αρτηρίας, πραγματοποιήθηκε ανάλυση με χρήση μικροσυστοιχιών (Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]). Τα ανθρώπινα δείγματα ακτινοβολήθηκαν με 2 Gy σε σετ των 3 δειγμάτων, ενώ χρησιμοποιήθηκαν και τα αντίστοιχα μη ακτινοβολημένα. Το ολικό RNA εκχυλίστηκε σε δύο διαφορετικά χρονικά σημεία μετά την ακτινοβολήση (1 ημέρα και 7 ημέρες) δίνοντας σύνολο 24 δειγμάτων [103].

Τα συνοπτικά χαρακτηριστικά για τον πειραματικό σχεδιασμό και την τεχνολογία μικροσυστοιχιών που χρησιμοποιήθηκε σε κάθε σειρά δεδομένων συνοψίζονται στους παρακάτω πίνακες (Πίνακας 1, Πίνακας 2).

Πίνακας 1: Στοιχεία για τον πειραματικό σχεδιασμό των μελετών που επιλέχθηκαν από το GEO και το ArrayExpress και αναλύθηκαν για το σκοπό της εργασίας.

Series Accession	Human Tissue	IR Type	Dose [Gy]	Time Points [hrs]	Samples
GSE44282 ⁵ E-GEOD-44282	Bronchial Epithelial Cells (HBEC3KT)	X-rays Fe Si Ions	1,3 0.5,1	1, 4, 12, 24	134
GSE80547 ⁶	Peripheral blood mononuclear cells	alpha particles	0.5, 1, 1.5	24	48
GSE81067 ⁷	Pulmonary Endothelial Cells (HPEAC) Lung Fibroblasts (HFL-1)	alpha particles	0.5, 1, 1.5	24	110
GSE63952 ⁸	Isolated Leucocytes	HZE particles	0.15, 0.3, 1.5	1, 2, 6	100
GSE64375 ⁹ E-GEOD-64375	Isolated Leucocytes	HZE particles	0.3, 1.5, 3	1	16
E-MTAB-3463 ¹⁰	Peripheral blood mononuclear cells	X-rays Fe C Ions	0.1, 1 1	8 8	16 60
E-MTAB-5754 ¹²	Coronary Artery Endothelial Cells	X-rays Fe Ions	2	24, 168	24

⁵ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44282>

⁶ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80547>

⁷ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81067>

⁸ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63952>

⁹ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64375>

¹⁰ <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3463/>

¹¹ <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5761/>

¹² <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5754/>

Πίνακας 2: Κωδικός και πλήρες όνομα της τεχνολογίας της μικροσυστοιχίας (πλατφόρμας) που χρησιμοποιήθηκε σε κάθε σειρά δεδομένων που επιλέχθηκε και αναλύθηκε. Οι σειρές με αστερίσκο βρέθηκαν και στα δύο αποθετήρια.

Series Accession	Technology – Platform Name	Platform Accession
GSE442828* E-GEOD-44282	Illumina HumanWG-6 v2.0 expression beadchip	GPL13376 ¹³
GSE80547	Illumina HumanHT-12 V4.0 expression beadchip	GPL10558 ¹⁴
GSE81067	Illumina HumanHT-12 V3.0 expression beadchip	GPL6947 ¹⁵
GSE63952	[PrimeView] Affymetrix Human Gene Expression Array	GPL15207 ¹⁶
GSE64375* E-GEOD-64375	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244 ¹⁷
E-MTAB-5761	[HuGene-1_0-st-v1] Affymetrix GeneChip Human Gene 1.0 ST Array	A-AFFY-141 ¹⁸
E-MTAB-3463	[HuGene-1_0-st-v1] Affymetrix GeneChip Human Gene 1.0 ST Array NetAffx Probeset annotation release 31	A-AFFY-187 ¹⁹
E-MTAB-5754	[HuGene-2_0-st] Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]	GPL16686 ²⁰

¹³ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13376>

¹⁴ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10558>

¹⁵ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6947>

¹⁶ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL15207>

¹⁷ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6244>

¹⁸ <https://www.ebi.ac.uk/arrayexpress/arrays/A-AFFY-141/>

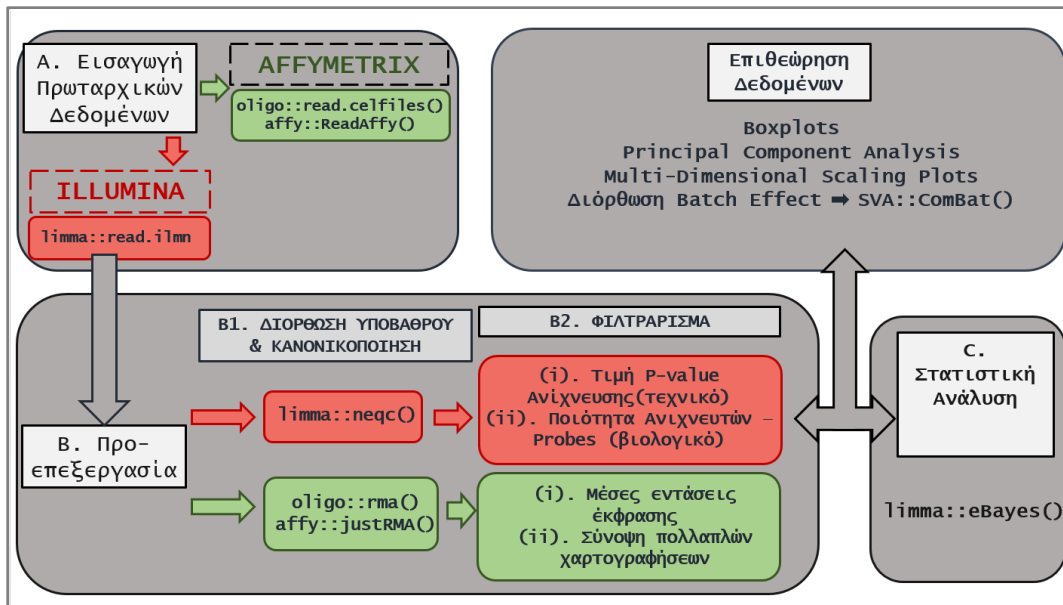
¹⁹ <https://www.ebi.ac.uk/arrayexpress/arrays/A-AFFY-187/>

²⁰ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16686>

2.3. Ανάλυση Δεδομένων Μικροσυστοιχιών

Το άνοιγμα και η επεξεργασία των δεδομένων των μικροσυστοιχιών στην παρούσα εργασία έγινε με χρήση της προγραμματιστικής γλώσσας R [79] στο ολοκληρωμένο περιβάλλον ανάπτυξης RStudio [81]. Η R αποτελεί ένα περιβάλλον λογισμικού για στατιστικούς υπολογισμούς και γραφικά. Μία μεγάλη συλλογή πακέτων της R που χρησιμοποιείται ευρέως στην Βιοπληροφορική είναι η σουίτα Bioconductor [80]. Κάθε πλατφόρμα εμφανίζει διαφορές ως προς την διαχείριση της μέσω της R, το βασικό πακέτο που χρησιμοποιείται όμως σε κάθε περίπτωση είναι το *limma* (*Linear Models for Microarray and RNA-Seq data*) [104]. Το πακέτο *limma* αποτελεί πακέτο λογισμικού για την ανάλυση πειραμάτων γονιδιακής έκφρασης, κυρίως με χρήση γραμμικών μοντέλων, με βασικό στόχο την αξιολόγηση της διαφορικής έκφρασης [78].

Αφού καταλήξαμε στις συγκεκριμένες σειρές δεδομένων, προχωρούμε στην ανάλυση των πρωταρχικών δεδομένων έκφρασης. Τα τρία βασικά βήματα που ακολουθήθηκαν είναι: η εισαγωγή δεδομένων, η προ-επεξεργασία και η στατιστική ανάλυση. Τα πρώτα δύο βήματα εξαρτώνται κυρίως από την τεχνολογία της εκάστοτε πλατφόρμας, ενώ υπάρχουν πολλοί τρόποι επιθεώρησης των δεδομένων κατά τη διάρκεια της διαδικασίας. Παρακάτω θα περιγραφούν αναλυτικά οι ροές εργασίας (Εικόνα 4), οι οποίες αντιστοιχούν σε κοινώς χρησιμοποιούμενες τεχνολογίες από την Illumina και την Affymetrix. Επιπλέον, παρατίθεται ο κώδικας για την εγκατάσταση των βασικών πακέτων που χρησιμοποιήθηκαν κατά την ανάλυση, σύμφωνα με την έκδοση 3.6.1 της R. Τα πακέτα εφόσον εγκατασταθούν πρέπει να κληθούν με την εντολή *library()* ώστε να χρησιμοποιηθούν.



Εικόνα 4: Συνοπτική παρουσίαση των ροών εργασίας για την ανάλυση δεδομένων.

```

BiocManager::install(c("Biobase", "limma", "GEOquery", "oligo",
/"affy", "affxparser", "sva"))
BiocManager::install(c("illuminaHumanv2.db",illuminaHumanv3.db,
/illuminaHumanv4.db))
BiocManager::install(c("pd.hugene.1.0.st.v1", "pd.hugene.2.0.st",
/"primeview.db", "hugene10sttranscriptcluster.db",
/"hugene20sttranscriptcluster.db"))
BiocManager::install(c("gplots", "ggplot2"))

```

2.3.1. Εισαγωγή Πρωταρχικών Δεδομένων

2.3.1.1 Illumina

GSE80547: Για να αναλύσουμε τις σειρές δεδομένων Illumina, ξεκινούμε από το αρχείο κειμένου «*GSE80547_non-normalized.txt*» που περιέχει τις μη επεξεργασμένες εκφράσεις. Σε αυτό το αρχείο οι γραμμές αντιπροσωπεύουν τους ανιχνευτές στη μικροσυστοιχία (γονίδια) ενώ οι στήλες αντιστοιχούν στα δείγματα της μελέτης. Συνήθως για κάθε δείγμα δίπλα από τη στήλη των εντάσεων υπάρχει μια επιπλέον στήλη με τιμές p-value ανίχνευσης (*detection p-values*). Η έννοια αυτής της τιμής χρησιμοποιείται από την Illumina καθώς στις BeadArray μικροσυστοιχίες υπάρχει ένα σύνολο ανιχνευτών αρνητικού ελέγχου (*negative controls*),

που χρησιμοποιούνται για τον υπολογισμό της έντασης υποβάθρου. Οι στήλες αυτές απαιτούνται αργότερα στην ανάλυση κατά την διαδικασία διόρθωσης υποβάθρου και κανονικοποίησης. Μία μικρή τιμή *p-value* ανίχνευσης υποδεικνύει ότι η μετρούμενη ένταση είναι πολύ πιθανό να είναι ένα στατιστικά σημαντικό σήμα έκφρασης και όχι θόρυβος υποβάθρου [105]. Για να διαβάσουμε αυτό το αρχείο στην R, χρησιμοποιούμε τη συνάρτηση *read.ilmn()* από το πακέτο *limma*, προσδιορίζοντας τις στήλες που πρέπει να διαβαστούν ως ανιχνευτές στο όρισμα «*probeid*», ως τιμές έκφρασης στο «*expr*» και ως τιμές *p-value* ανίχνευσης στο «*other.columns*»:

```
rawData80547=read.ilmn(files="GSE80547_non-normalized.txt",  
/probeid="ID_REF", expr="GSM", other.columns="Detection")
```

Έτσι με αυτή τη συνάρτηση δημιουργείται ένα αντικείμενο κλάσης *EListRaw* με όνομα *rawData80547* που μπορεί να περιέχει επιπλέον πληροφορίες σχολιασμού των ανιχνευτών (*probe annotation*), οι οποίες εισάγονται ως αντικείμενο *data.frame* στο όρισμα «*genes*» και πληροφορίες για τον φαινότυπο των δειγμάτων στο όρισμα «*targets*». Συνεπώς σε αυτό το στάδιο είναι χρήσιμο να εισαχθούν και οι φαινοτυπικές-πειραματικές πληροφορίες για τα δείγματα του πειράματος. Εάν η μελέτη ενδιαφέροντος βρίσκεται και στο *ArrayExpress*, υπάρχει διαθέσιμο αρχείο κειμένου (*Sample and Data Relationship Format-SDRF*), που περιέχει την πληροφορία αυτή για τα δείγματα. Εφόσον η μελέτη βρίσκεται μόνο στο *GEO*, αυτό το αρχείο μπορεί να δημιουργηθεί χειροκίνητα με τις σειρές να αντιστοιχούν σε αναγνωριστικά δείγματος και τις στήλες να περιέχουν τις φαινοτυπικές-πειραματικές πληροφορίες που θα χρησιμοποιηθούν, όπως η δόση ακτινοβολίας, ο χρόνος μετά την ακτινοβολήση ή το πειραματικό αντίγραφο στο οποίο αντιστοιχεί το εκάστοτε δείγμα. Κάποιος μπορεί εύκολα να ανακτήσει αυτή την πληροφορία από το αρχείο κειμένου «*GSE80547_series_matrix.txt*» της σειράς στο *GEO*, από τις γραμμές αυτού που αρχίζουν με «*! Sample_*». Οι φαινοτυπικές-πειραματικές πληροφορίες μπορούν εναλλακτικά να φορτωθούν μεμονωμένα ως απλός πίνακας *pheData80547*, χρησιμοποιώντας την συνάρτηση *read.delim()* στο αντίστοιχο αρχείο «*80547.sdrf.txt*»:

```
pheData80547=read.delim("80547.sdrf.txt", header=TRUE, sep = "\t",
/quote /= "\"", dec = ".", fill = TRUE)
```

GSE81067: Σε αυτή την μελέτη εκτελώντας την ανάλυση για τα μη επεξεργασμένα δεδομένα εκφράσεων η διαφορική στατιστική ανάλυση δεν κατέληγε σε στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια. Έτσι δοκιμάσαμε να χρησιμοποιήσουμε τα επεξεργασμένα από τους ερευνητές δεδομένα από το αρχείο «*GSE81067_series_matrix.txt*» της μελέτης που περιέχει, εκτός της φαινοτυπικές-πειραματικές πληροφορίες όπως αναφέρθηκε παραπάνω, και τις κανονικοποιημένες τιμές έκφρασης, χωρίς όμως τις τιμές p-value ανίχνευσης. Για το σκοπό αυτό, μπορούμε να χρησιμοποιήσουμε τα πακέτα GEOquery (συνάρτηση *getGEO()*) [106] ή ArrayExpress (συνάρτηση *getAE()*) [107] ώστε να δημιουργήσουμε ένα αντικείμενο *data.matrix.81067* που περιέχει τα κανονικοποιημένα δεδομένα έκφρασης με τους ανιχνευτές ως γραμμές και τα δείγματα ως στήλες. Για να συμπεριλάβουμε φαινοτυπικές-πειραματικές πληροφορίες, αφού διαβάσουμε το αρχείο «*81067.sdrf.txt*» σε πίνακα *pheData81067*, το μετατρέψαμε σε αντικείμενο *AnnotatedDataFrame* και το ενσωματώσαμε στο πεδίο «*phenoData*» του τελικού αντικειμένου *ExpressionSet* με όνομα *normalized81067*:

```
rawData81067=getGEO(GEO="GSE81067",
/filename="GSE81067_family.soft.gz", GSEMatrix = TRUE,
/AnnotGPL = TRUE, getGPL = FALSE)
gsmlist=Filter(function(gsm){Meta(gsm)$platform_id=='GPL6947'},
/GSMList(rawData81067))
probesets = Table(GPLList(raw)[[1]])$ID
data.matrix.81067 = do.call('cbind',lapply(gsmlist,function(x)
{tab <- Table(x)
mymatch <- match(probesets,tab$ID_REF)
return(tab$VALUE[mymatch])}))
data.matrix.81067=apply(data.matrix.81067,2,function(x)
/{as.numeric(as.character(x))})
rownames(data.matrix.81067) = probesets
colnames(data.matrix.81067) = names(gsmlist)
pheData81967=read.delim("81067.sdrf.txt", header = TRUE, sep = "\t",
/quote = "\"", dec = ".", fill = TRUE)
pheno = as(pheData81067,"AnnotatedDataFrame")
rownames(pheno) = names(gsmlist)
normalized81067=new('ExpressionSet',exprs=data.matrix.81067,
/phenoData=pheno)
```

GSE44282: Στην σειρά αυτή το αρχείο των μη επεξεργασμένων εκφράσεων δεν περιλαμβάνει στήλες τιμής p-value ανίχνευσης. Συνεπώς προτείνεται να διαβαστεί απευθείας το αρχείο «GSE44282_series_matrix.txt» της σειράς από το GEO σε αντικείμενο *data.matrix.44282* με την ίδια διαδικασία που αναφέρθηκε παραπάνω. Χωρίς αυτές τις τιμές ανίχνευσης, δεν είναι δυνατή η διόρθωση υποβάθρου και η κανονικοποίηση. Επιπλέον οι φαινοτυπικές-πειραματικές πληροφορίες ενσωματώθηκαν στο τελικό ExpressionSet με όνομα *normalized44282*, από το αντίστοιχο αρχείο «44282.sdrf.txt» όπως λήφθηκε από το ArrayExpress, ως πίνακας *pheData44282* που μετατράπηκε σε αντικείμενο AnnotatedDataFrame και εισήχθη στο όρισμα «*phenoData*»:

```
rawData44282=getGEO(GEO="GSE44282",
/filename="GSE44282_family.soft.gz", GSEMatrix = TRUE,
/AnnotGPL = TRUE, getGPL = FALSE)
gsmlist=Filter(function(gsm){Meta(gsm)$platform_id=='GPL13376'},
/GSMList(rawData44282))
probesets = Table(GPLList(raw)[[1]])$ID
data.matrix.44282 = do.call('cbind',lapply(gsmlist,function(x)
{tab = Table(x)
mymatch = match(probesets,tab$ID_REF)
return(tab$VALUE[mymatch])}))
data.matrix.44282=apply(data.matrix.44282,2,function(x)
/{as.numeric(as.character(x))})
rownames(data.matrix.44282) = probesets
colnames(data.matrix.44282) = names(gsmlist)
pheData44282=read.delim("44282.sdrf.txt", header = TRUE, sep = "\t",
/quote = "\"", dec = ".", fill = TRUE)
pheno = as(pheData44282,"AnnotatedDataFrame")
rownames(pheno) = names(gsmlist)
normalized44282=new('ExpressionSet',exprs=data.matrix.44282,
/phenoData=pheno)
```

2.3.1.2 Affymetrix

GSE64375: Τα σύνολα δεδομένων Affymetrix περιλαμβάνουν τα μη επεξεργασμένα δεδομένα έκφρασης σε αρχεία μορφοποίησης τύπου CEL, αλλά απαιτούν και τον αντίστοιχο σχολιασμό που εμπεριέχει τις πληροφορίες χαρτογράφησης για τη μικροσυστοιχία μέσω συσχέτισης των ανιχνευτών (*probes*) με τα *probesets* [64]. Σε ορισμένες εκδόσεις, όπως η GeneChip, οι πληροφορίες σχεδίασης της πλατφόρμας αποθηκεύονται στα

CEL αρχεία και μπορούν εύκολα να προσπελαστούν από το αντίστοιχο πακέτο, στην περίπτωση αυτή το `rd.hugene.1.0.st.v1` [108]). Η φόρτωση τους απαιτεί μόνο την εισαγωγή των ονομάτων των αρχείων προς ανάγνωση με την κατάληξη `.CEL`, καθώς και του αντίστοιχου αρχείου «`GSE64375.sdrf.txt`» που βρέθηκε στο ArrayExpress. Χρησιμοποιείται η συνάρτηση `read.celfiles()` του πακέτου `oligo` [109]. Η συνάρτηση δημιουργεί αντικείμενο τύπου `ExpressionFeatureSet`, το οποίο ονομάζουμε `rawData64375` και απαρτίζεται μεταξύ άλλων από τα τμήματα «`assayData`» (εντάσεις έκφρασης), «`featureData`» (σχολιασμός ανιχνευτών) και «`phenoData`» (φαινοτυπικά-πειραματικά δεδομένα):

```
SDRF = read.delim("GSE64375.sdrf.txt")
rownames(SDRF) = SDRF$Array.Data.File
SDRF = AnnotatedDataFrame(SDRF)
celfiles = list.celfiles(full.names=TRUE)
rawData64375 = oligo::read.celfiles(celfiles, phenoData = SDRF)
```

GSE63952: Στην περίπτωση αυτής της μελέτης, όπου συναντούμε την έκδοση `PrimeView`, απαιτείται μια συγκεκριμένη βιβλιοθήκη (`Common Data Format-CDF`), που περιέχει τις πληροφορίες σχεδιασμού της μικροσυστοιχίας. Ένας τρόπος διαχείρισης τέτοιων μικροσυστοιχιών απαιτεί τη λήψη του αρχείου της βιβλιοθήκης «`PrimeView.cdf`» μέσα από την επίσημη ιστοσελίδα της `Affymetrix`²¹. Έπειτα χρησιμοποιώντας το αρχείο αυτό και τη συνάρτηση `convertCdf()` του πακέτου `affxparser` [110], μετασχηματίζουμε τη βιβλιοθήκη σε δυαδική μορφοποίηση (`version="4"` στα ορίσματα) προκειμένου να χρησιμοποιηθεί στη συνάρτηση `ReadAffy()` του πακέτου `affy` [111, 112] ώστε να εισαχθούν τα μη επεξεργασμένα δεδομένα στην R. Αυτή η συνάρτηση διαβάζει όλα τα αρχεία με κατάληξη `.CEL` στον τρέχοντα φάκελο εργασίας και τα ενσωματώνει σε ένα αντικείμενο `AffyBatch` που ονομάζουμε `rawData63952`. Παράλληλα ενσωματώνουμε τις πληροφορίες πειραματικού σχεδιασμού (`phenoData`) από το αρχείο «`GSE63952.sdrf.txt`» που κατασκευάσαμε:

21

<http://www.affymetrix.com/support/technical/byproduct.affx?product=primeview>

```
convertCdf("PrimeView.cdf","PrimeView2.cdf",version="4",
/force=FALSE, validate=TRUE)
SDRF = read.delim("GSE63952.sdrf.txt")
SDRF = AnnotatedDataFrame(SDRF)
rawData63952 = ReadAffy(phenoData=SDRF, cdfname="primeview2")
```

Και στις δύο περιπτώσεις που αφορούν δεδομένα μικροσυστοιχιών της Affymetrix, ανεξαρτήτως της συνάρτησης μέσω της οποίας φορτώσαμε τα δεδομένα στην R, τα δεδομένα έκφρασης, φαινοτύπου και ανιχνευτών μπορούν να προσπελαστούν αντίστοιχα με τις εντολές από το πακέτο Biobase [113]:

```
Biobase::exprs(rawData63952)
Biobase::pData(rawData63952)
Biobase::fData(rawData63952)
```

E-MTAB-5754|5761|3463: Ο κώδικας για την ανάγνωση των πρωταρχικών δεδομένων για τις τρεις μελέτες που βρέθηκαν μόνο στο ArrayExpress είναι ανάλογος με την πρώτη περίπτωση πλατφόρμας Affymetrix (GSE64375) καθώς αντιστοιχούν στην έκδοση GeneChip:

```
SDRF <- read.delim("E-MTAB-5754.sdrf.txt")
rownames(SDRF) <- SDRF$Array.Data.File
SDRF <- AnnotatedDataFrame(SDRF)
celfiles = list.celfiles(full.names=TRUE)
rawData5754 = oligo::read.celfiles(celfiles, phenoData = SDRF)
```

```
SDRF <- read.delim("E-MTAB-5761.sdrf.txt")
rownames(SDRF) <- SDRF$Array.Data.File
SDRF <- AnnotatedDataFrame(SDRF)
celfiles = list.celfiles(full.names=TRUE)
rawData5761 = oligo::read.celfiles(celfiles, phenoData = SDRF)
```

```
SDRF <- read.delim("E-MTAB-3463.sdrf.txt")
rownames(SDRF) <- SDRF$Array.Data.File
SDRF <- AnnotatedDataFrame(SDRF)
celfiles = list.celfiles(full.names=TRUE)
rawData3463 = oligo::read.celfiles(celfiles, phenoData = SDRF)
```

Μετά τη φόρτωση των πρωτογενών δεδομένων και των απαραίτητων σχολιασμών, είμαστε έτοιμοι να προχωρήσουμε στο στάδιο της προεπεξεργασίας. Αυτό το βήμα είναι κρίσιμο για να διασφαλιστεί το ότι τα τελικά αποτελέσματα διαφορικής στατιστικής ανάλυσης, θα περιέχουν τον

μέγιστο αριθμό αληθώς θετικών, στατιστικώς σημαντικών διαφορικώς εκφραζόμενων γονιδίων.

2.3.2. Προ-επεξεργασία Δεδομένων

Το στάδιο αυτό περιλαμβάνει κατά βάση τη διόρθωση σήματος υποβάθρου, την κανονικοποίηση των και λογαριθμικό μετασχηματισμό με βάση το 2 των δεδομένων έκφρασης [66, 68, 114]. Επιπλέον συμπεριλαμβάνονται μέθοδοι για φιλτράρισμα δειγμάτων και ανιχνευτών βάση των εντάσεων έκφρασης, που αφορά σε τεχνικά χαρακτηριστικά. Σε περιπτώσεις που υπάρχουν οι αντίστοιχες βιβλιοθήκες σχολιασμού των ανιχνευτών συνίσταται να πραγματοποιείται περεταίρω φιλτράρισμα βάση της ποιότητας των ανιχνευτών. Η σύνοψη των ανιχνευτών λόγω πολλαπλής χαρτογράφησης σε μικροσυστοιχίες τεχνολογίας Affymetrix είναι απαραίτητη. Υπάρχουν επίσης πολλά χρήσιμα γραφήματα που μπορούν να κατασκευαστούν για την επιθεώρηση των δεδομένων, ως προς τον έλεγχο της επεξεργασίας και την ύπαρξη *Batch Effect* [115]. Εάν για οποιονδήποτε λόγο κάποιος επιλέξει να ξεκινήσει την ανάλυση με τα επεξεργασμένα από τους ερευνητές δεδομένα, τότε το στάδιο της προ-επεξεργασίας ξεκινά με το φιλτράρισμα των δεδομένων βάση τεχνικών χαρακτηριστικών.

2.3.2.1 Illumina

Η συνάρτηση *neqc()* του πακέτου *limma*, ιδιαίτερα χρήσιμη για τις μικροσυστοιχίες τύπου *BeadArray* [105], εκτελεί διόρθωση υποβάθρου με την μέθοδο «*normexp*» [116] χρησιμοποιώντας ανιχνευτές αρνητικού ελέγχου. Επιπρόσθετα εκτελεί κανονικοποίηση ποσοστημορίου, η οποία χρησιμοποιεί ανιχνευτές αρνητικού και θετικού ελέγχου, οι οποίοι προκύπτουν από τις στήλες τιμών *p-value* ανίχνευσης.

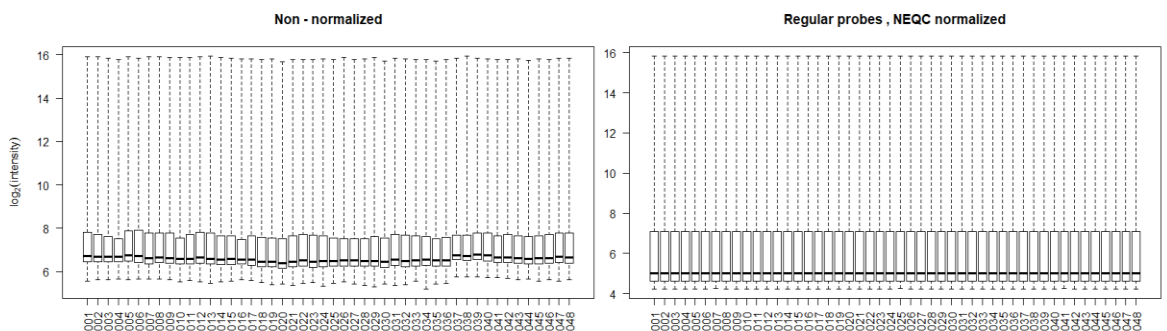
GSE80547: Με χρήση της παραπάνω συνάρτησης δημιουργείται ένα αντικείμενο *EList-class* που ονομάσαμε *normalized80547* και περιέχει τις κανονικοποιημένες και λογαριθμημένες με βάση το 2 εντάσεις έκφρασης:

```
normalized80547=neqc(rawData80547, status=NULL, negctrl="negative",  
/regular="regular",offset=15,robust=FALSE, detection.p="Detection")
```

Οι γραμμές του αντικειμένου αντιστοιχούν σε ανιχνευτές από τις οποίες έχουν αφαιρεθεί εκείνες που αντιστοιχούν σε ανιχνευτές ελέγχου. Οι στήλες αντιστοιχούν στα δείγματα της σειράς και τις αντίστοιχες στήλες p-value ανίχνευσης. Εφόσον δεν διαχειριζόμαστε αντικείμενο τύπου ExpressionSet, τα φαινοτυπικά-πειραματικά δεδομένα εισάγονται σε ξεχωριστό αντικείμενο data.matrix όπως αναφέρθηκε παραπάνω, με όνομα *pheData80547*:

Για να ελέγξουμε ότι η διαδικασία της κανονικοποίησης εκτελέστηκε ορθά, σχεδιάζουμε τα θηκογράμματα [69, 70] πριν και μετά την εφαρμογή της συνάρτησης *neqc()* (Εικόνα 5):

```
boxplot(log2(rawData80547$E),range=0,las=2,xlab="",ylab=expression
/(log [2]( intensity )),main=" Non - normalized ")
boxplot(normalized80547$E,range=0,las=2,xlab="",ylab=expression(log
/[2](intensity)),main=" Regular probes , NEQC normalized ")
```



Εικόνα 5: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα πριν και μετά την εφαρμογή της συνάρτησης *neqc()*, η οποία εκτελεί διόρθωση υποβάθρου και κανονικοποίηση, για την σειρά GSE80547.

Έτσι λαμβάνουμε την γραφική αναπαράσταση των αριθμητικών δεδομένων έκφρασης σε παραστατική στατιστική με χρήση θηκογραμμάτων. Εφόσον οι διάμεσοι κάθε δείγματος, οι οποίοι αναπαρίστανται με τη μαύρη γραμμή σε κάθε κουτί είναι ευθυγραμμισμένοι στο δεύτερο διάγραμμα, αυτομάτως συνάγεται το συμπέρασμα ότι η διαδικασία ήταν επιτυχής. Συνεχίζουμε με την διαδικασία φιλτραρίσματος βάση τεχνικών χαρακτηριστικών της μικροσυστοιχίας, βασιζόμενοι στις τιμές p-value ανίχνευσης. Θα αφαιρέσουμε τους ανιχνευτές που δεν παρουσιάζουν αποδεκτή τιμή p-value για αριθμό δειγμάτων ίσο ή μεγαλύτερο του μικρότερου αριθμού δειγμάτων που υπάγονται σε μία κατηγορία. Για παράδειγμα:

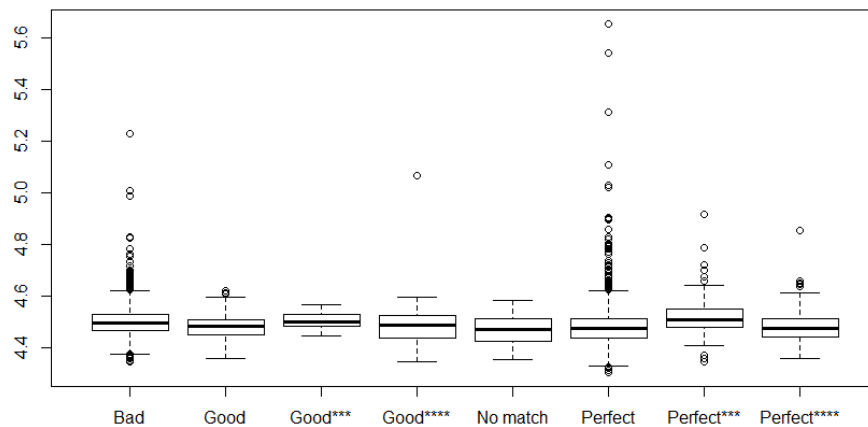
```
expressed = rowSums(normalized80547$other$Detection <= 0.1) >= 3
normalized80547_2 = normalized80547[expressed,]
dim(normalized80547_2)
```

Το αποδεκτό κατώφλι για την τιμή p-value (0.1) και ο αριθμός των δειγμάτων (3) που καθορίζουν ποιοι ανιχνευτές γίνονται αποδεκτοί στο αντικείμενο *expressed* είναι ενδεικτικά. Σε κάθε περίπτωση προτείνεται ο πειραματισμός με διάφορες τιμές έως να παραμένουν τουλάχιστον 7,000 ανιχνευτές, αριθμός που μπορεί να ελεγχθεί από τις διαστάσεις και ειδικότερα τον αριθμό των γραμμών του *normalized80547_2*. Αν σε μία μελέτη οι τιμές ανίχνευσης δεν διατίθενται, το βήμα αυτό προφανώς δεν είναι εφικτό.

Για τις περισσότερες μικροσυστοιχίες της Illumina, υπάρχουν βάσεις δεδομένων που περιέχουν πληροφορίες για την αντιστοίχιση των κωδικών ταυτοποίησης ανιχνευτών του κατασκευαστή με πολλαπλές ονοματολογίες γονιδίων και πληροφορίες για τη θέση στο χρωμόσωμα και τις αλληλουχίες των γονιδίων. Επιπλέον, περιέχουν την πληροφορία ανάθεσης χαρακτηρισμού ποιότητας σε κάθε ανιχνευτή βασιζόμενη στο πόσο κατάλληλα και μοναδικά ταιριάζει στο μετάγραφο/στόχο την οποία και θα χρησιμοποιήσουμε στο στάδιο αυτό. Για την μελέτη αυτή έχουμε την βιβλιοθήκη *illuminaHumanv4.db* [117, 118] και ειδικότερα την συνάρτηση *illuminaHumanv4PROBEQUALITY*, που δίνει την δυνατότητα για περαιτέρω φιλτράρισμα των ανιχνευτών με βάση τον σχολιασμό ποιότητας (Εικόνα 6):

```
ids = as.character(rownames(normalized80547_2))
qual=unlist(mget(ids,illuminaHumanv4PROBEQUALITY,ifnotfound=NA))
AveSignal = rowMeans(normalized80547_2$E)
boxplot( AveSignal ~ qual )
rem = qual == "No match " | qual == "Bad"
normalized80547.filt = normalized80547_2 [!rem , ]
dim(normalized80547.filt)
```

Κατά αυτόν τον τρόπο αφαιρούμε τους ανιχνευτές που έχουν σημειωθεί ως «*Bad*» (κακή ταύτιση) ή «*No Match*» (μηδενική ταύτιση) από τους δημιουργούς της βάσης, σύμφωνα με δεδομένα από δημόσια αποθετήρια. Και σε αυτή την περίπτωση συνίσταται ο έλεγχος του αριθμού των ανιχνευτών που παραμένουν, πάνω στους οποίους θα βασιστεί η ανάλυση στη συνέχεια.



Εικόνα 6: Αναπαράσταση με θηκογράμματα (boxplots) των μέσων εντάσεων έκφρασης ανά κατηγορία ετικέτας ποιότητας ανιχνευτή για την σειρά GSE80547

Σε αυτό το σημείο θα επιθεωρήσουμε τα δεδομένα χρησιμοποιώντας την μέθοδο ανάλυσης κυρίων συνιστωσών (PCA) [71], που χρησιμοποιεί ορθογώνιο μετασχηματισμό για να μετατρέψει ένα σύνολο παρατηρήσεων από πιθανώς συσχετισμένες μεταβλητές σε ένα σύνολο από γραμμικώς ανεξάρτητες μεταβλητές ονομαζόμενες «κύριες συνιστώσες». Κατά αυτόν τον τρόπο μπορούμε να συμπεράνουμε ποιο φαινοτυπικό-πειραματικό χαρακτηριστικό ορίζει κατά κύριο λόγο τις διαφορές στην γονιδιακή έκφραση. Σε αυτή την περίπτωση έχουμε μία μελέτη κατά την οποία εφαρμόζονται τρεις διαφορετικές δόσεις ακτινοβολίας σωματιδίων-α. Όλα τα δείγματα συλλέχθηκαν 24 ώρες μετά την ακτινοβολία ενώ έγιναν 12 επαναλήψεις του πειράματος από τα 12 σετ των δειγμάτων που δημιουργήθηκαν. Η ανάλυση πραγματοποιείται με χρήση της συνάρτησης `prcomp()` του πακέτου `stats` [119] ενώ το διάγραμμα της PCA σχεδιάζεται με χρήση της συνάρτησης `ggplot()` του πακέτου `ggplot2` [120] ως εξής:

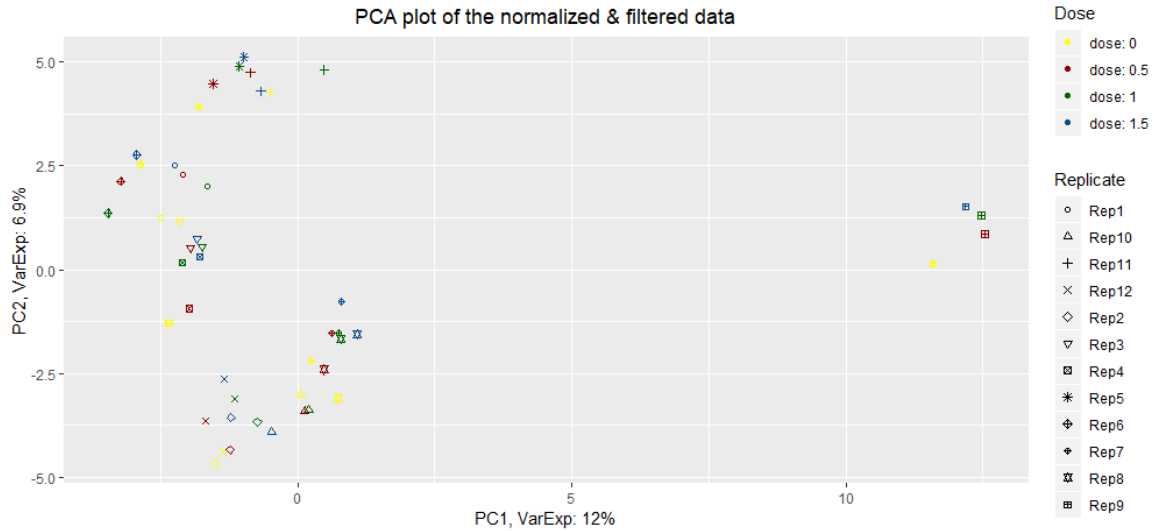
```

exprs_norm = normalized80547.filt$t$E
PCA = prcomp(t(exprs_norm), scale = FALSE)
percentVar = round(100*PCA$sdev^2/sum(PCA$sdev^2),1)
sd_ratio = sqrt(percentVar[2] / percentVar[1])

dataGG = data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],
Dose = pheData$Dose,
Replicate = c("Rep1","Rep1","Rep1","Rep1","Rep2","Rep2","Rep2",
/"Rep2","Rep3","Rep3","Rep3","Rep3","Rep4","Rep4","Rep4","Rep4",
/"Rep5","Rep5","Rep5","Rep5","Rep6","Rep6","Rep6","Rep6","Rep7",
/"Rep7","Rep7","Rep7","Rep8","Rep8","Rep8","Rep8","Rep9","Rep9",
/"Rep9","Rep9","Rep10","Rep10","Rep10","Rep10","Rep11","Rep11",
/"Rep11","Rep11","Rep12","Rep12","Rep12","Rep12"))

ggplot(dataGG, aes(PC1, PC2)) +
geom_point(aes(shape = Replicate, colour = Dose)) +
ggtitle("PCA plot of the normalized & filtered data") +
xlab(paste0("PC1, VarExp: ", percentVar[1], "%")) +
ylab(paste0("PC2, VarExp: ", percentVar[2], "%")) +
theme(plot.title = element_text(hjust = 0.5)) +
coord_fixed(ratio = sd_ratio) +
scale_shape_manual(values = c(1,2,3,4,5,6,7,8,9,10,11,12)) +
scale_color_manual(values = c("yellow", "darkred", "darkgreen",
/"dodgerblue4"))

```

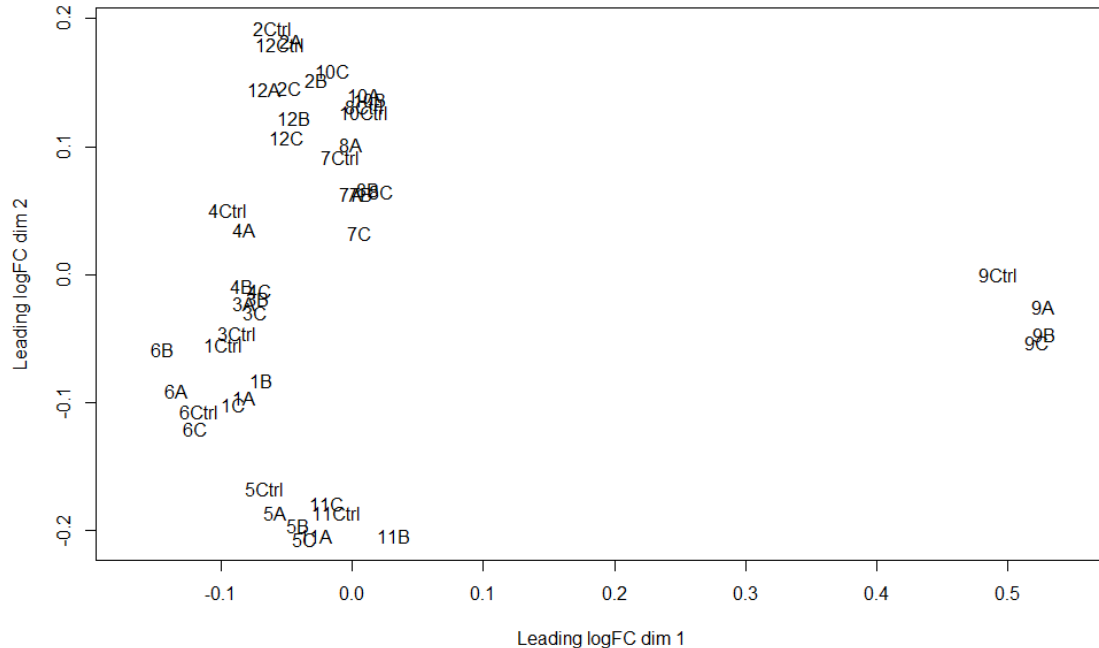


Εικόνα 7: Διάγραμμα της PCA για την σειρά GSE80547. Τα δείγματα φαίνεται να ομαδοποιούνται με βάση τα σετ πειραματικών επαναλήψεων και όχι με βάση την δόση της ακτινοβολίας.

Στο διάγραμμα (Εικόνα 7) βλέπουμε ότι τα δείγματα παρουσιάζουν συστάδες βάση των επαναλήψεων. Το γεγονός ότι η διαφοροποίηση στην γονιδιακή έκφραση οφείλεται σε θέματα πειραματικού σχεδιασμού και όχι

σε βιολογικές παραμέτρους, υποδηλώνει την ύπαρξη *Batch Effect* [115]. Θα πρέπει συνεπώς να εφαρμοστεί μια διαδικασία διόρθωσης των δεδομένων πριν προχωρήσουμε σε διαφορική ανάλυση. Παρόλα αυτά για να βεβαιωθούμε ότι πράγματι συμβαίνει κάτι τέτοιο, θα σχεδιάσουμε και ένα διάγραμμα πολυδιάστατης κλιμάκωσης (MDS) [72] χρησιμοποιώντας την συνάρτηση *plotMDS()* του πακέτου *limma*. Διαθέτουμε σύνολο 48 δειγμάτων που προκύπτουν από 12 επαναλήψεις, όπου σε κάθε μία από αυτές έχουμε τρεις διαφορετικές δόσεις ακτινοβολίας (A,B,C) και τα δείγματα αναφοράς (Ctrl), όλα συλλεγμένα στις 24 ώρες μετά την ακτινοβολία. Το διάγραμμα MDS είναι ιδιαίτερα χρήσιμο σε περιπτώσεις που το πείραμα διαθέτει περισσότερες πολλαπλές παραμέτρους όπως θα δούμε στη συνέχεια. Ονοματίζοντας κατάλληλα τα δείγματα μπορεί κανείς να δει με βάση ποια παράμετρο σχηματίζονται οι συστάδες των δειγμάτων. Στο συγκεκριμένο διάγραμμα (Εικόνα 8) φαίνεται καθαρά ο σχηματισμός συστάδων βάση πειραματικής επανάληψης στα δείγματα. Προχωρούμε στη διόρθωση του *Batch Effect* με χρήση της συνάρτησης *ComBat()* του πακέτου *SVA* [121]. Η συνάρτηση απαιτεί μία ετικέτα περιγραφής της παραμέτρου που ορίζεται ως «παρτίδα» (στην περίπτωση αυτή οι πειραματικές επαναλήψεις), την οποία ονομάζουμε *batch.lab* και έναν πίνακα μοντελοποίησης των υπολοίπων παραμέτρων, που ονομάζουμε *mod*.

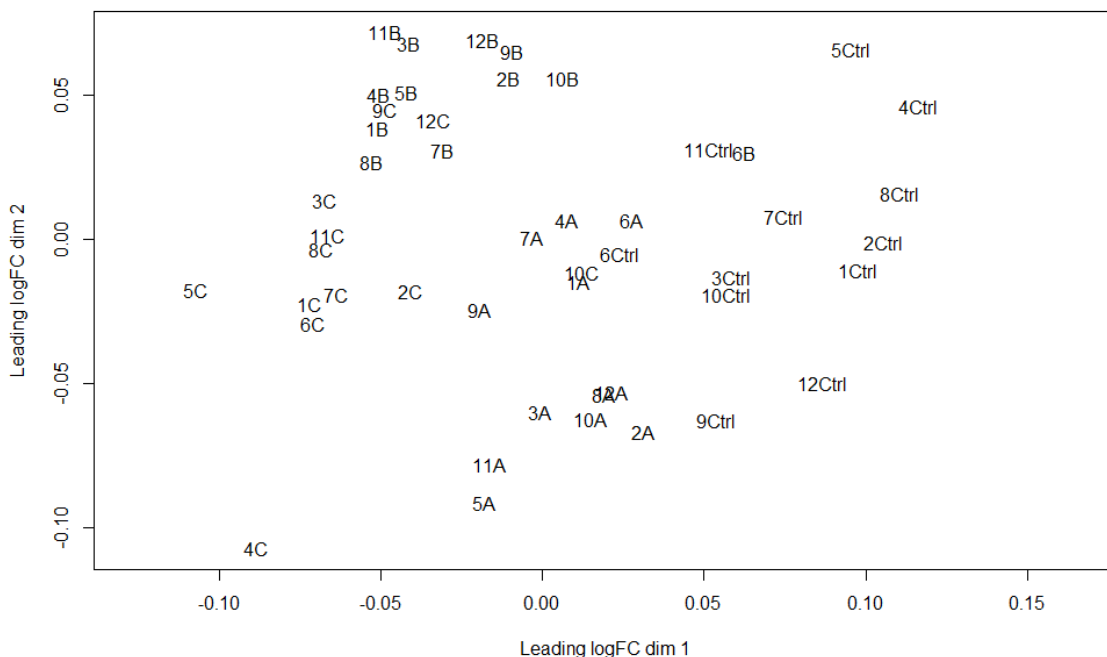
```
labels= c("1Ctrl","1A","1B","1C","2Ctrl","2A","2B","2C","3Ctrl",  
/"3A","3B","3C","4Ctrl","4A","4B","4C","5Ctrl","5A","5B","5C",  
/"6Ctrl","6A","6B","6C","7Ctrl","7A","7B","7C","8Ctrl","8A","8B",  
/"8C","9Ctrl","9A","9B","9C","10Ctrl","10A","10B","10C","11Ctrl",  
/"11A","11B","11C","12Ctrl","12A","12B","12C")  
limma::plotMDS(normalized80547.filt$E, labels=labels)
```

Εικόνα 8: Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE80547. Κάθε σημείο αναπαριστά ένα δείγμα όπου ο αριθμός στην αρχή του ονόματός του υποδηλώνει το σετ δειγμάτων και τα Ctrl, A, B, C την δόση της ακτινοβολίας.

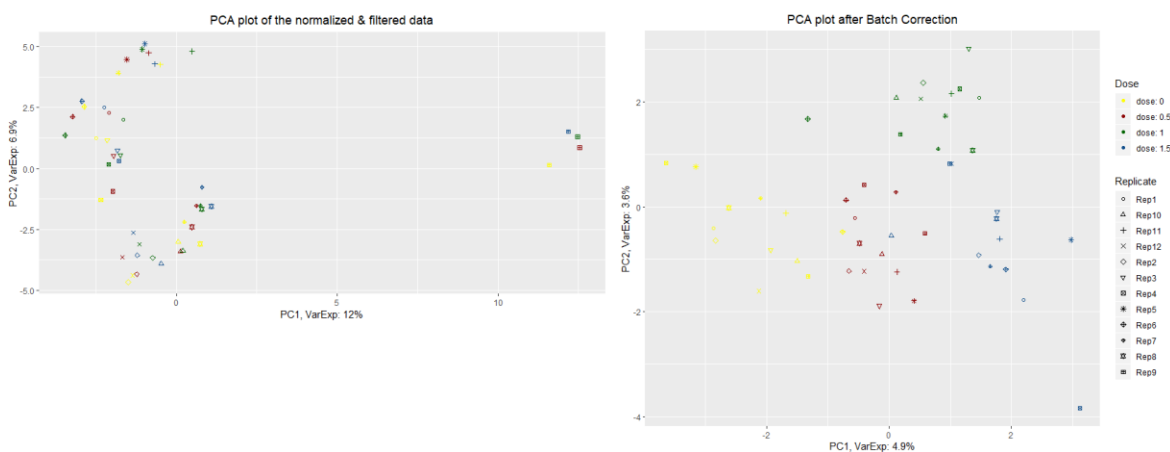
Για να βεβαιωθούμε ότι τα δεδομένα μετασχηματίστηκαν επιτυχώς, σχεδιάζουμε ξανά το διάγραμμα MDS (Εικόνα 9). Βλέπουμε λοιπόν ότι οι συστάδες των δειγμάτων έχουν διασκορπιστεί και τα δείγματα ίδιας δόσης εμφανίζονται σε γειτονικές θέσεις, γεγονός που συνάδει με το βιολογικώς αναμενόμενο προφίλ.

```
batch.lab = c (1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5,
/5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10,
/10, 11, 11, 11, 11, 12, 12, 12, 12)
mod = model.matrix(~ 0+factor(rep(c(1,2,3,4),12)))
80547.combat_data = Combat(dat = normalizedData2.filt$E, batch =
/batch.lab, mod = mod[,1:3], par.prior = TRUE, prior.plots = TRUE)
plotMDS(80547.combat_data, labels=labels)
```



Εικόνα 9: Διάγραμμα MDS για τα διορθωμένα δεδομένα της σειράς GSE80547. Κάθε σημείο αναπαριστά ένα δείγμα όπου ο αριθμός στην αρχή του ονόματός του υποδηλώνει το σετ δειγμάτων και τα Ctrl, A, B, C την δόση της ακτινοβολίας.

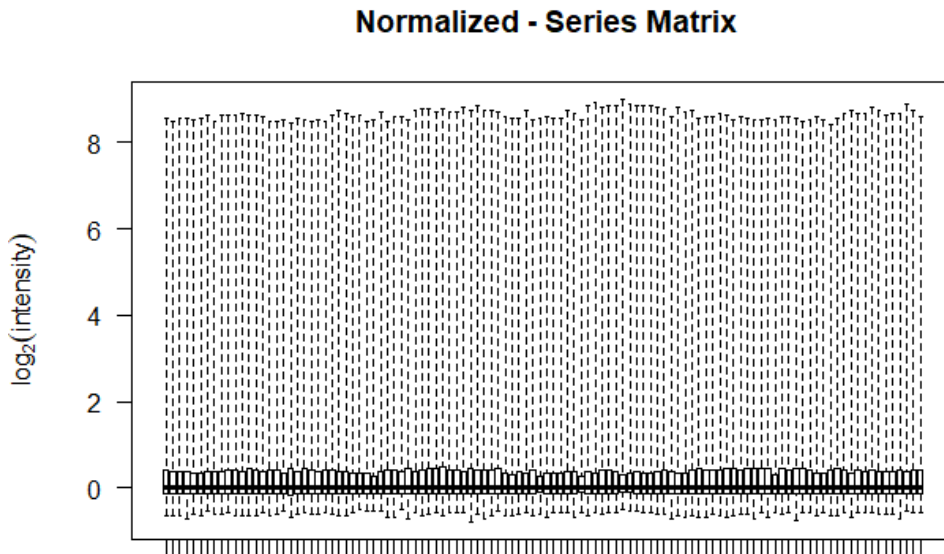
Σημειώνεται ότι η διόρθωση είναι ορατή και στην περίπτωση του διαγράμματος της PCA (Εικόνα 10). Συνεπώς τα δεδομένα έχουν υποστεί την κατάλληλη προ-επεξεργασία και μπορούν να αναλυθούν στατιστικά ως προς τη διαφορική έκφραση γονιδίων.



Εικόνα 10: Διαγράμματα της PCA για την σειρά GSE80547 πριν και μετά την διαδικασία διόρθωσης του Batch Effect. Στο δεύτερο διάγραμμα οι αρχικές συστάδες δειγμάτων με βάση το πειραματικό αντίγραφο έχουν απαλειφθεί και διακρίνεται ομαδοποίηση με βάση την δόση της ακτινοβολίας.

GSE81067: Σε αυτή τη σειρά, εφόσον διαβάσαμε κατευθείαν τις επεξεργασμένες από τους ερευνητές τιμές έκφρασης από το αρχείο «GSE81067_series_matrix.txt», συνίσταται ο έλεγχος των δεδομένων αυτών (Εικόνα 11).

```
boxplot(exprs(normalized81067), range=0, las=2, xlab="GSMS",  
/names=FALSE, ylab=expression(log [2]( intensity )), main="Normalized  
/- Series Matrix")
```



Εικόνα 11: Αναπαράσταση δεδομένων έκφρασης με θηκόγραμμα (boxplot) ανά δείγμα, για τα επεξεργασμένα δεδομένα της σειράς GSE81067, όπως λήφθηκαν από το αντίστοιχο αρχείο series.matrix.

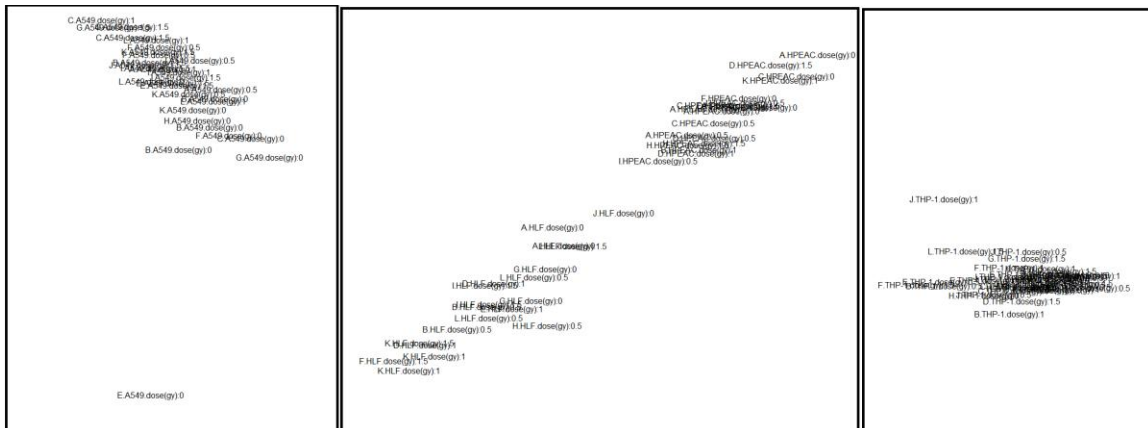
Εφόσον οι διάμεσοι των δειγμάτων είναι ευθυγραμμισμένοι, μπορούμε να προχωρήσουμε στη διαδικασία του φιλτραρίσματος με βάση την ποιότητα με χρήση της illuminaHumanv3.db [117]:

```
ids = as.character(rownames(normalized81067))  
qual = unlist(mget(ids, illuminaHumanv3PROBEQUALITY, ifnotfound = NA))  
rem = qual == "No match " | qual == "Bad"  
normalized81067.filt = normalized81067 [!rem , ]
```

Σχεδιάζοντας το διάγραμμα MDS (Εικόνα 12) φαίνεται ότι τα δείγματα δημιουργούν συστάδες με βάση τους τέσσερις διαφορετικούς ιστούς που χρησιμοποιούνται και όχι τις πειραματικές επαναλήψεις, το οποίο είναι και βιολογικώς αναμενόμενο. Οφείλει να σημειωθεί σε αυτό το σημείο πως ακολουθώντας τις ίδιες διαδικασίες για τις μη επεξεργασμένες εκφράσεις,

τα δείγματα πάλι δημιουργούσαν συστάδες όχι όμως βάση του ιστού. Καθώς δεν ήταν δυνατό να εντοπιστεί ο λόγος για τον οποίο συνέβαινε αυτό επιλέχθηκε η ανάλυση των επεξεργασμένων από τους ερευνητές δεδομένων.

```
dose_labels = normalized81067$Dose
type_labels = normalized81067$Celltype
rep_labels = normalized81067$REP.
plotMDS(normalized81067.filt, labels=paste(rep_labels,type_labels,
/dose_labels, sep="."))
```



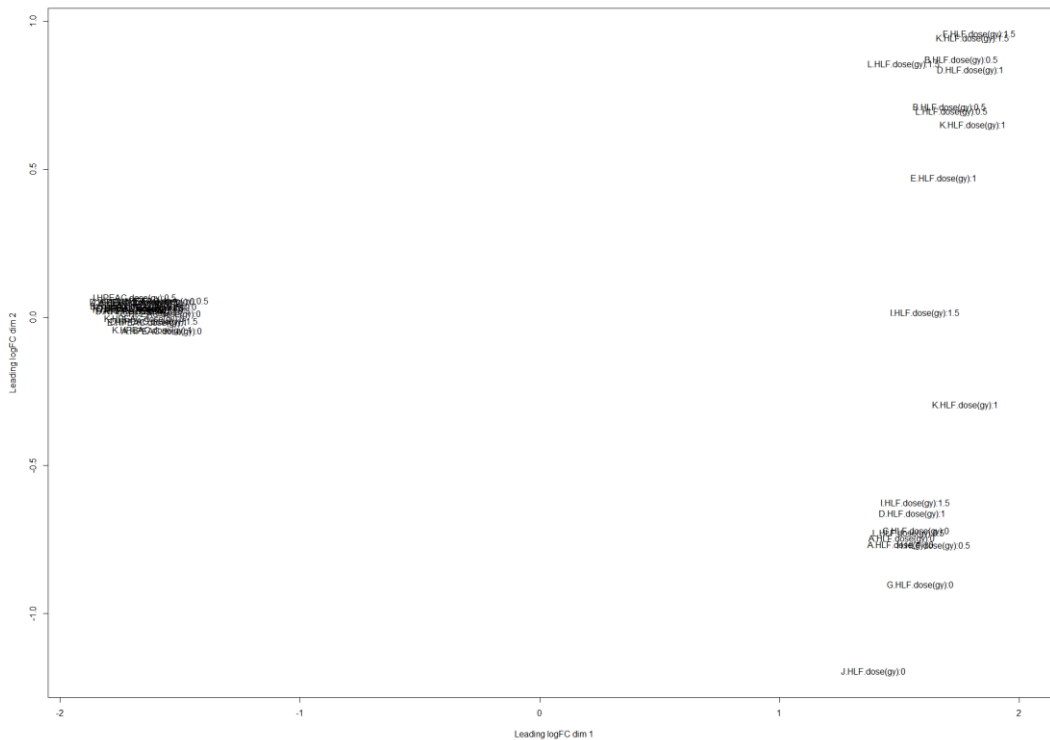
Εικόνα 12: Απεικόνιση των συστάδων στο διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE81067. Κάθε σημείο αναπαριστά ένα δείγμα ενώ οι 4 συστάδες που παρουσιάζονται, αντιστοιχούν στους τέσσερις τύπους ιστού που χρησιμοποιήθηκαν στο πείραμα (A549, HLF, HPEAC, THP-1).

Σε αυτό το σημείο επιθυμούμε να κρατήσουμε μόνο τα δείγματα που αντιστοιχούν σε φυσιολογικό ιστό: πνευμονικά ενδοθηλιακά κύτταρα (HPEAC) και ινοβλάστες πνεύμονα (HLF-1). Συνεπώς τα αποθηκεύουμε σε αντικείμενο ExpressionSet με όνομα *normalized81067.filt.normal*. Για το σκοπό αυτό, τα ονόματα των δειγμάτων που αντιστοιχούν σε φυσιολογικό ιστό αποθηκεύονται στην λίστα *normal.GSMs*. Σχεδιάζουμε ξανά το διάγραμμα MDS (Εικόνα 13) με ετικέτες που υποδηλώνουν κατά σειρά το πειραματικό αντίγραφο, τον ιστό και την δόση ακτινοβολίας για δείγματα φυσιολογικού ιστού και μόνο.

```

normalTissue=subset(pData(normalized81067.filt),
/normalized81067.filt$CellType=="HPEAC"|
/normalized81067.filt$CellType=="HLF",
select=c(Dose, CellType, REP., TimePoint))
normal.GSMs = rownames(normalTissue)
normalized81067.filt.normal = normalized81067.filt[,normal.GSMs]
dose_labelsN = normalized81067.filt.normal$Dose
type_labelsN = normalized81067.filt.normal$CellType
rep_labelsN = normalized81067.filt.normal$REP.
plotMDS(normalized81067.filt.normal,
/labels=paste(rep_labelsN,type_labelsN,dose_labelsN,sep="."))

```

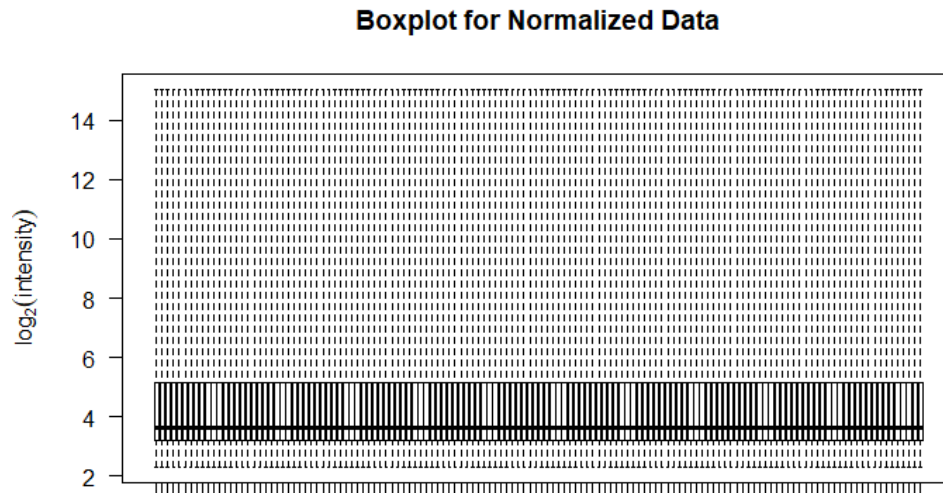


Εικόνα 13 Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE81067 που αντιστοιχούν σε φυσιολογικό ανθρώπινο ιστό.

Στην περίπτωση των δειγμάτων από ινοβλάστες πνεύμονα (HLF-1) παρατηρούμε ξεκάθαρα την ύπαρξη διαφοροποιήσεων με βάση την δόση ακτινοβολίας. Αντίθετα οι διαφορές στην έκφραση μεταξύ δειγμάτων από πνευμονικά ενδοθηλιακά κύτταρα (HPEAC) φαίνονται αρκετά μικρές (Εικόνα 13). Παρόλα αυτά τα δεδομένα έχουν προετοιμαστεί με τον βέλτιστο δυνατό τρόπο για διαφορική ανάλυση.

GSE44282: Ομοίως σε αυτή τη σειρά, εφόσον διαβάσαμε τις επεξεργασμένες από τους ερευνητές τιμές έκφρασης, προχωρούμε στην κατασκευή θηκογράμματος για έλεγχο των δεδομένων έκφρασης του αντικειμένου *normalized44282*:

```
boxplot(exprs(normalized44282), range=0, las=2, xlab="GSMS",
        /names=FALSE, ylab=expression(log [2]( intensity )), main="Boxplot
        /for Normalized Data")
```



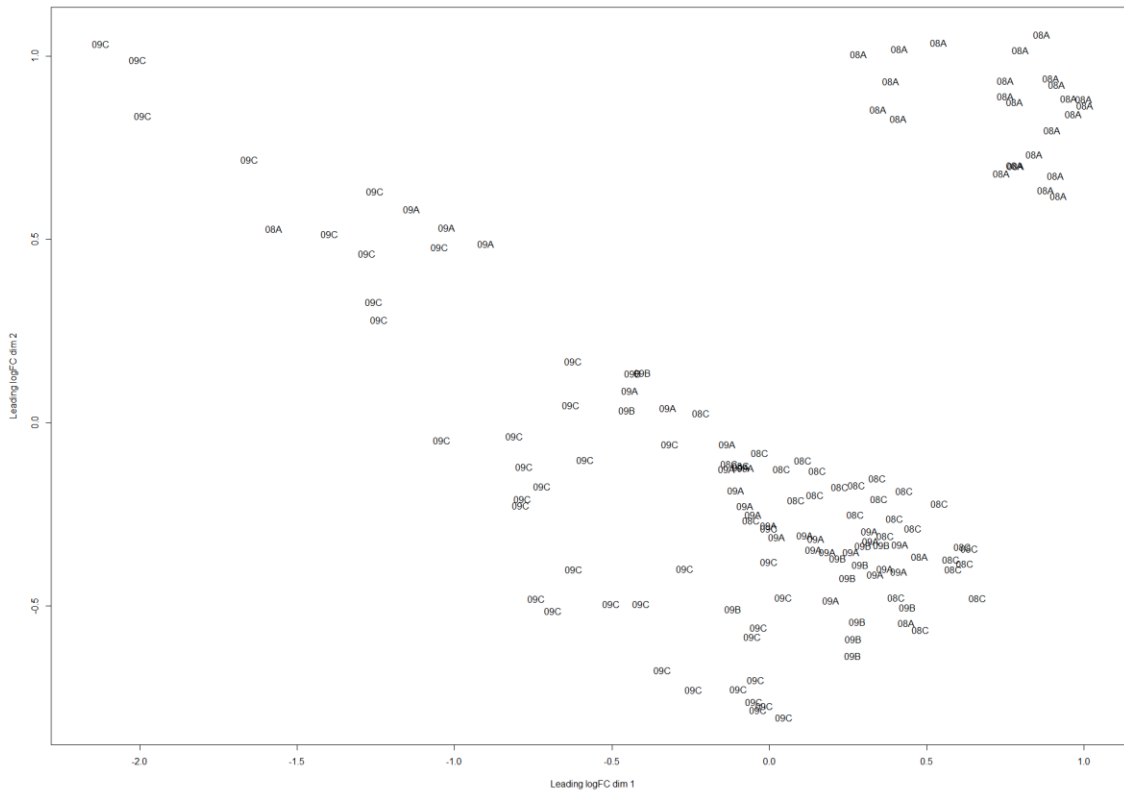
Εικόνα 14: Αναπαράσταση δεδομένων έκφρασης με θηκόγραμμα (boxplot) ανά δείγμα, για τα επεξεργασμένα δεδομένα της σειράς GSE44282, όπως λήφθηκαν από το αρχείο *series.matrix*.

Παρατηρούμε στο διάγραμμα (Εικόνα 14) ότι τα δεδομένα έχουν την απαραίτητη μορφή και άρα μπορούμε να συνεχίσουμε με την διαδικασία φιλτραρίσματος με βάση την ποιότητα των ανιχνευτών (*illuminaHumanv2.db* [122]), καθώς το φιλτράρισμα με βάση τα τεχνικά χαρακτηριστικά είναι ανέφικτο λόγω έλλειψης των τιμών *p-value* ανίχνευσης:

```
ids = as.character(rownames(exprs(normalized44282)))
qual = unlist(mget(ids, illuminaHumanv2PROBEQUALITY, ifnotfound = NA))
rem <- qual == "No match" | qual == "Bad"
normdata44282.filt = data.matrix[!rem, ]
```

Σχεδιάζοντας πολλαπλά διαγράμματα MDS καταλήξαμε στο συμπέρασμα ότι τα δείγματα σχηματίζουν συστάδες με βάση τις πειραματικές

επαναλήψεις (Εικόνα 15), υποδεικνύοντας έτσι την ύπαρξη Batch Effect, το οποίο και θα πρέπει να διορθωθεί.



Εικόνα 15: Διάγραμμα MDS για τα φιλτραρισμένα δεδομένα της σειράς GSE44282. Κάθε σημείο αναπαριστά ένα δείγμα ενώ διαφαίνονται τρεις συστάδες που αντιστοιχούν στα τρία από τα πέντε βιολογικά αντίγραφα (επαναλήψεις).

Για να είναι πιο ξεκάθαρο το συμπέρασμα αυτό, σχεδιάσαμε και το αντίστοιχο διάγραμμα PCA υποδεικνύοντας με διαφορετικό χρώμα τον συνδυασμό τύπου και δόσης ακτινοβολίας και με διαφορετικά σχήματα τις πειραματικές επαναλήψεις (Εικόνα 16).

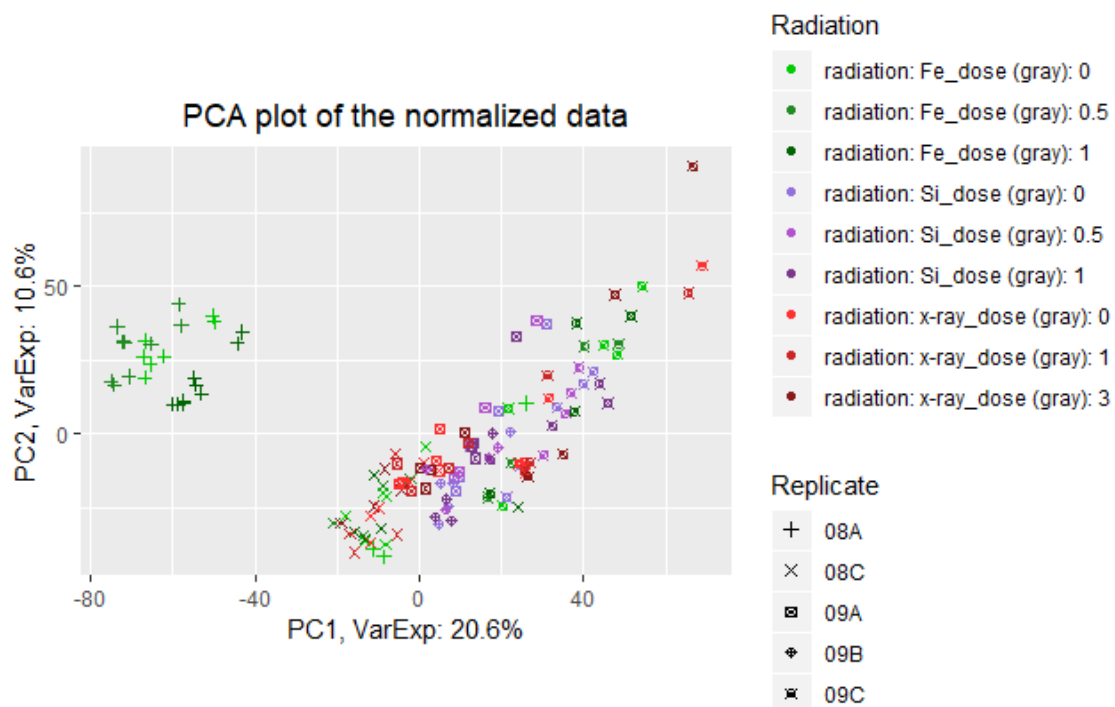
```

labels = normalized44282$CONDITION
rep_labels = normalized44282$Replicate
plotMDS(normalized44282.filt, labels=rep_labels)
exprs_norm = exprs(normalized44282.filt)
PCA = prcomp(t(exprs_norm), scale = FALSE)
percentVar = round(100*PCA$sdev^2/sum(PCA$sdev^2),1)
sd_ratio = sqrt(percentVar[2] / percentVar[1])

dataGG = data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],
/Radiation=paste(normalized44282$RadiationType,normalized44282$Dose
/,sep="_"),
/Replicate = normalized44282$Replicate)

ggplot(dataGG, aes(PC1, PC2)) +
geom_point(aes(shape = Replicate, colour = Radiation)) +
ggtitle("PCA plot of the normalized data") +
xlab(paste0("PC1, VarExp: ", percentVar[1], "%")) +
ylab(paste0("PC2, VarExp: ", percentVar[2], "%")) +
theme(plot.title = element_text(hjust = 0.5)) +
coord_fixed(ratio = sd_ratio) +
scale_shape_manual(values = c(3,4,7,10,13)) +
scale_color_manual(values = c("green3", "forestgreen", "darkgreen",
/"mediumpurple", "mediumorchid3", "mediumorchid4", "firebrick1",
/"firebrick3", "firebrick4"))

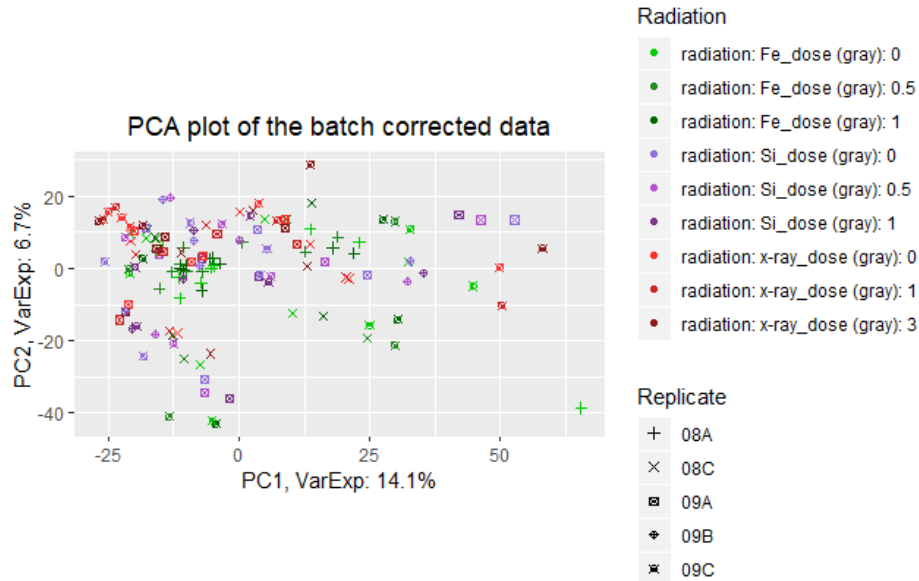
```



Εικόνα 16: Διάγραμμα PCA για τα κανονικοποιημένα δεδομένα της σειράς GSE44282. Με χρώμα υποδηλώνεται η δόση σε συνδυασμό με τον τύπο της ακτινοβολίας ενώ τα σχήματα αντιστοιχούν στις διαφορετικές πειραματικές επαναλήψεις.

Προχωρούμε στη διόρθωση του Batch Effect εισάγοντας τις πειραματικές επαναλήψεις στο *batch.lab* και κατασκευάζοντας τον απαιτούμενο πίνακα μοντελοποίησης *mod* με βάση τα φαινοτυπικά-πειραματικά χαρακτηριστικά τύπου και δόσης ακτινοβολίας καθώς και του χρόνου συλλογής των δειγμάτων έπειτα από την ακτινοβολήση. Σχεδιάζοντας το νέο διάγραμμα MDS είδαμε τις συστάδες των δειγμάτων να εξαλείφονται. Η επιτυχημένη διαδικασία διόρθωσης των δεδομένων διαφαίνεται και από το αντίστοιχο διάγραμμα PCA (Εικόνα 17). Υπάρχει ένα δείγμα (GSM1082166) σε μεγάλη απόσταση από τα υπόλοιπα, το οποίο αποτελεί δείγμα αναφοράς για τα ακτινοβολημένα με ιόντα σιδήρου (Fe) δείγματα. Καθώς για κάθε τύπο ακτινοβολίας υπάρχουν πολλά δείγματα αναφοράς, το παραπάνω δεν θα αποτελέσει πρόβλημα για την μετέπειτα διαφορική ανάλυση λόγω στατιστικής.

```
batch.lab = normalized44282.filt$Replicate
mod=model.matrix(~as.factor(CONDITION),
data=normdata44282.filt$CONDITION)
colnames(mod)=c("Fe.05_1", "Fe.05_12", "Fe.05_24", "Fe.05_4", "Fe.1_1",
/"Fe.1_12", "Fe.1_24", "Fe.1_4", "Fe.C_0", "Fe.C_1", "Fe.C_12", "Fe.C_24"
/"Fe.C_4", "Si.05_1", "Si.05_12", "Si.05_24", "Si.05_4", "Si.1_1",
/"Si.1_12", "Si.1_24", "Si.1_4", "Si.C_0", "Si.C_1", "Si.C_12",
/"Si.C_24", "Si.C_4", "Xray.1_1", "Xray.1_12", "Xray.1_24", "Xray.1_4",
/"Xray.3_1", "Xray.3_12", "Xray.3_24", "Xray.3_4", "Xray.C_0",
/"Xray.C_1", "Xray.C_12", "Xray.C_24", "Xray.C_4")
rownames(mod) = normdata44282.filt$GEO_Accession
44282.combat_data = Combat(dat = exprs(normdata44282.filt),
/ batch = batch.lab, mod = mod[1:38], par.prior = TRUE,
/ prior.plots = TRUE)
plotMDS(44282.combat_data, labels=paste(rep_labels, labels,
/sep="."))
```

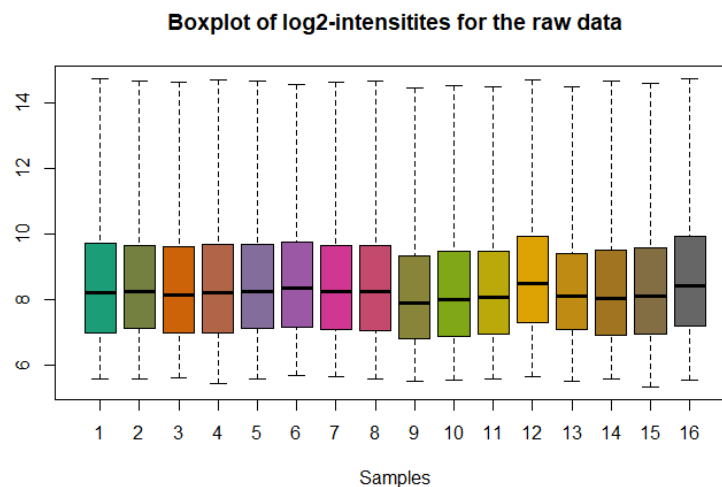


Εικόνα 17: Διάγραμμα PCA για τα διορθωμένα δεδομένα της σειράς GSE44282, όπου οι συστάδες με βάση τις διαφορετικές πειραματικές επαναλήψεις έχουν εξαλειφθεί.

2.3.2.2 Affymetrix

GSE64375: Αρχικά σχεδιάζουμε το θηκόγραμμα για τις μη επεξεργασμένες τιμές έκφρασης στο αντικείμενο `rawData64375` (Εικόνα 18) εφαρμόζοντας όμως σε αυτές μόνο λογάριθμο με βάση το 2 μέσω της συνάρτησης `log2()`:

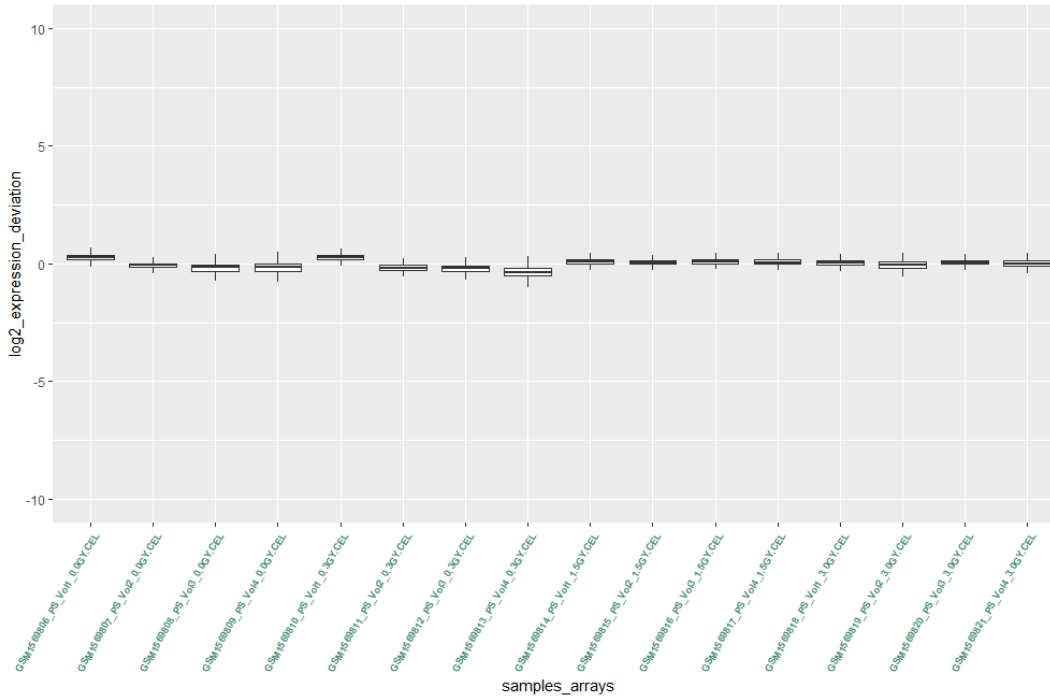
```
oligo::boxplot(log2(rawData64375), target="core", main="Boxplot of log2-intensities for the raw data", names=TRUE, xlab="Samples")
```



Εικόνα 18: Αναπαράσταση μη επεξεργασμένων δεδομένων έκφρασης με θηκογράμματα (boxplots) της σειράς GSE64375 με χρήση του πακέτου `oligo`.

Κατά την διαχείριση δεδομένων Affymetrix συνίσταται να εκτελείται ένας έλεγχος των μη επεξεργασμένων δεδομένων μέσω υπολογισμού των μέσων λογαριθμημένων εντάσεων όλων των μεταγράφων (ανιχνευτών) για όλες τις μικροσυστοιχίες (δείγματα) [123]. Για το σκοπό αυτό υπολογίζουμε τις διαμέσους των γραμμών του πίνακα `exprs(rawData64375)`, που υποδεικνύουν την διάμεσο έντασης έκφρασης από όλα τα δείγματα για κάθε ανιχνευτή. Σημειώνουμε ότι δεν πρέπει να λογαριθμήσουμε με βάση το δύο, καθώς η έξοδος της συνάρτησης `rma()` του πακέτου `oligo` [109] χρησιμοποιεί τον αλγόριθμο *Robust Multi-array Average* (RMA) [124] και η έξοδός της είναι προκαθορισμένα λογαριθμημένη. Η συνάρτηση παράλληλα μετατρέπει ένα αντικείμενο `AffyBatch` ή `ExpressionFeatureSet` σε `ExpressionSet`. Στη συνέχεια αφαιρούμε αυτή την μέση ένταση από κάθε ένταση για κάθε μετάγραφο μέσω της συνάρτησης `sweep()` και μετασχηματίζουμε τα δεδομένα ώστε να σχεδιάσουμε θηκογράμματα για κάθε δείγμα (Εικόνα 19):

```
RMA_notNorm=oligo::rma(rawData64375,target="core",normalize=FALSE)
row_medians=Biobase::rowMedians(as.matrix
/(Biobase::exprs(RMA_notNorm)))
RLE_data = sweep(Biobase:: (RMA_notNorm), 1, row_medians)
RLE_data = as.data.frame(RLE_data)
RLE_data_gathered = tidyr::gather(RLE_data, samples_arrays,
/log2_expression_deviation)
ggplot2::ggplot(RLE_data_gathered,
aes(samples_arrays,log2_expression_deviation))+
geom_boxplot(outlier.shape=NA)+ylim(c(-10,10))+
theme(axis.text.x = element_text(colour = "aquamarine4", angle = 60,
/size = 6.5, hjust = 1 , face = "bold"))
```



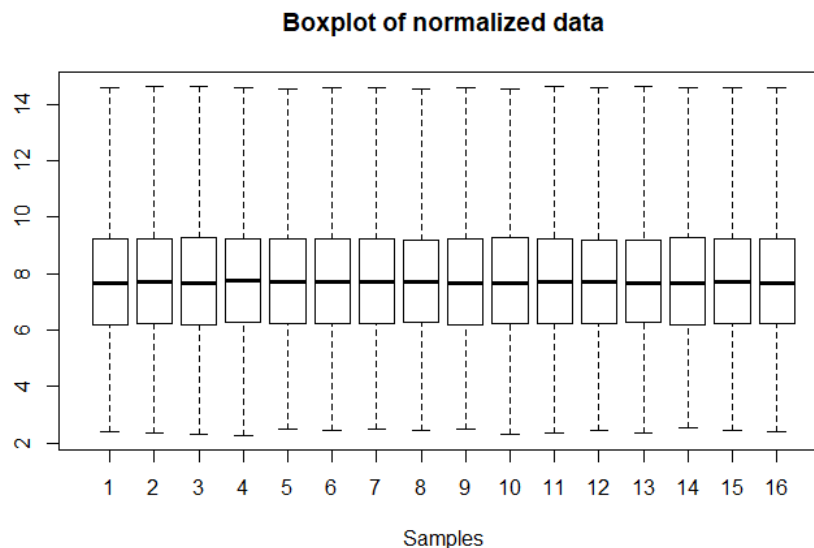
Εικόνα 19: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε κάθε δείγμα των πρωταρχικών δεδομένων της σειράς GSE64375.

Ο γ-άξονας στο διάγραμμα αναπαριστά για κάθε μικροσυστοιχία την απόκλιση έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα τσιπ. Συνεπώς κουτιά μεγαλύτερου ύψους υποδεικνύουν ασυνήθιστα μεγάλη απόκλιση από την διάμεσο σε πολλά μετάγραφα, υποδηλώνοντας ότι αυτά τα τσιπ διαφέρουν με κάποιο τρόπο από την πλειοψηφία των υπολοίπων. Κουτιά μετατοπισμένα κατά τον γ-άξονα υποδεικνύουν μια συστηματικά υψηλότερη ή χαμηλότερη έκφραση της πλειοψηφίας των μεταγραφών σε σύγκριση με την πλειοψηφία των υπολοίπων τσιπ. Κάτι τέτοιο θα μπορούσε να προκληθεί από ζητήματα ποιότητας ή Batch Effect. Επομένως εάν το σχήμα και η διάμεσος ενός συγκεκριμένου κουτιού διαφέρουν πάρα πολύ από το σύνολο, θα πρέπει να επιθεωρούνται και ενδεχομένως να αφαιρούνται [123]. Ωστόσο στην συγκεκριμένη μελέτη δεν εντοπίζεται κάτι τέτοιο.

Συνεπώς, προχωράμε στην εφαρμογή του πλήρους αλγορίθμου RMA προκειμένου να διορθώσουμε το υπόβαθρο, να κανονικοποιήσουμε και να συνοψίσουμε τα δεδομένα σε ένα ExpressionSet με όνομα

normalized64375. Επιπλέον σχεδιάζουμε τα αντίστοιχα θηκογράμματα των επεξεργασμένων δεδομένων (Εικόνα 20) για την επισκόπηση της διαδικασίας.

```
normalized64375 = oligo::rma(rawData64375, target = "core")
oligo::boxplot(normalized, main = "Boxplot of normalized data",
/Names=TRUE, xlab="Samples")
```



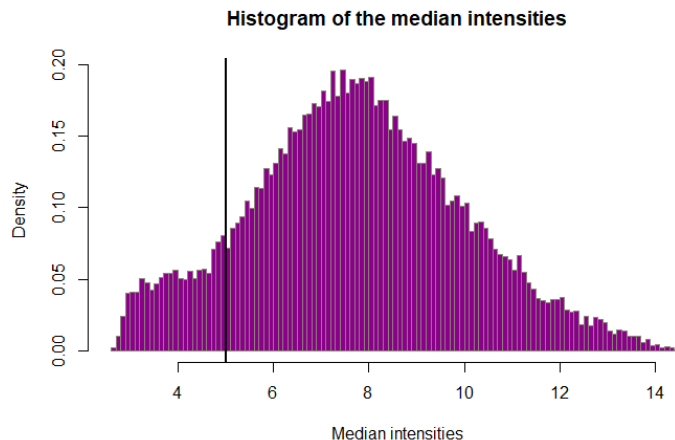
Εικόνα 20: Αναπαράσταση επεξεργασμένων δεδομένων έκφρασης της σειράς GSE64375 με θηκογράμματα έπειτα από την εφαρμογή του αλγορίθμου RMA.

Το επόμενο βήμα είναι να εκτελέσουμε ένα φιλτράρισμα με βάση τις εντάσεις. Πιο συγκεκριμένα στοχεύουμε στο να αφαιρέσουμε χαμηλώς υβριδοποιημένους ανιχνευτές, καθώς τα δεδομένα μικροσυστοιχιών συνήθως παρουσιάζουν μεγάλο αριθμό ανιχνευτών στην περιοχή εντάσεων υποβάθρου. Για το σκοπό αυτό υπολογίζουμε τις διαμέσους των εντάσεων ανά γραμμή ως διαμέσους των μεταγράφων και τις αναθέτουμε στο *gene_medians*. Έτσι κατασκευάζουμε ένα ιστόγραμμα των μέσων εντάσεων ανά γονίδιο (Εικόνα 21), όπου αναζητούμε την συσσώρευση χαμηλών μέσων στην αριστερή πλευρά που θα αντιπροσωπεύει τα γονίδια που θέλουμε να φιλτράρουμε. Προκειμένου να ορίσουμε το κατώφλι αποκοπής με όνομα *threshold*, πρέπει να επιθεωρήσουμε το ιστόγραμμα οπτικά.

```

gene_medians = rowMedians(Biobase::exprs(normalized64375))
threshold = 4
hist_res = hist(gene_medians, 100, col = "darkmagenta", freq = FALSE,
main = "Histogram of the median intensities", border =
"antiquewhite4", xlab = "Median intensities")
abline(v = threshold, col = "black", lwd = 2)

```



Εικόνα 21: Ιστόγραμμα για τις μέσες επεξεργασμένες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς GSE64375.

Μετάγραφα με εντάσεις μικρότερες του κατωφλίου σε τουλάχιστον τόσα τσιπ όσα η μικρότερη πειραματική ομάδα (*samples_cutoff*) θα αφαιρεθούν. Για το σκοπό αυτό εφαρμόζεται η συνάρτηση *idx_threshold()* σε κάθε γραμμή ώστε να αξιολογηθεί το αν ο αριθμός των τσιπ όπου η διάμεσος ξεπερνά το κατώφλι, είναι μεγαλύτερος από το *samples_cutoff* και επιστρέφει TRUE ή FALSE για κάθε γραμμή. Τέλος μετασχηματίσουμε το ExpressionSet ως *normalized64375.filt*, κρατώντας τις γραμμές που έλαβαν τιμή TRUE. Έγινε αφαίρεση 3461 γραμμών ανιχνευτών.

```

samples_cutoff = min(no_of_samples)
idx_threshold=apply(Biobase::exprs(normalized64375),1,function(x)
/{sum(x > threshold) >= samples_cutoff}c)
normalized64375.filt = subset(normalized64375, idx_threshold)

```

Το επόμενο βήμα είναι να σχολιάσουμε τα μετάγραφα ώστε να αφαιρέσουμε τις πολλαπλές χαρτογραφήσεις. Αρχικά, προσθέτουμε τα Feature Data, που αποτελούνται από το Hugo Gene Nomenclature Committee (HGNC) [125] Gene Symbol και το όνομα του κάθε γονιδίου σύμφωνα με τους κωδικούς ταυτοποίησης του κατασκευαστή (probesIDs) που είναι

αποθηκευμένα στο ExpressionSet. Για το σκοπό αυτό χρησιμοποιείται η βιβλιοθήκη `hugene10sttranscriptcluster.db` [126]. Στη συνέχεια αφαιρούμε τους ανιχνευτές στους οποίους δεν έχει ανατεθεί όνομα γονιδίου. Επιπλέον οι ανιχνευτές που αντιστοιχήθηκαν σε πολλαπλά γονίδια αφαιρούνται [123].

```
anno_exprs = AnnotationDbi::select(hugene10sttranscriptcluster.db,
                                   keys = (featureNames(normalizedData.filt)),
                                   columns = c("SYMBOL", "GENENAME"), keytype = "PROBEID")
anno_exprs = subset(anno_exprs, !is.na(SYMBOL))
anno_grouped = group_by(anno_exprs, PROBEID)
anno_summarized = dplyr::summarize(anno_grouped,
                                   /no_of_matches=n_distinct(SYMBOL))
head(anno_summarized)
# A tibble: 6 x 2
  PROBEID no_of_matches
  <chr>      <int>
1 7896740         3
2 7896742         3
3 7896744         4
4 7896754         1
5 7896756         2
6 7896759         1
anno_filtered = filter(anno_summarized, no_of_matches > 1)
probe_stats = anno_filtered
ids_ex=(featureNames(normalized64375.filt)%in% probe_stats$PROBEID)
table(ids_ex)
  ids_ex
  FALSE TRUE
  27723 2113
```

Αυτό σημαίνει ότι πρέπει να αφαιρεθούν 2113 ανιχνευτές (μετάγραφα). Έτσι δημιουργούμε το τελικό ExpressionSet `exprs_final` στο οποίο προσαρτώνται τα αντίστοιχα HGNC Gene Symbols με τα ονόματα των γονιδίων που αντιστοιχούν στους τελικώς επιλεγμένους ανιχνευτές:

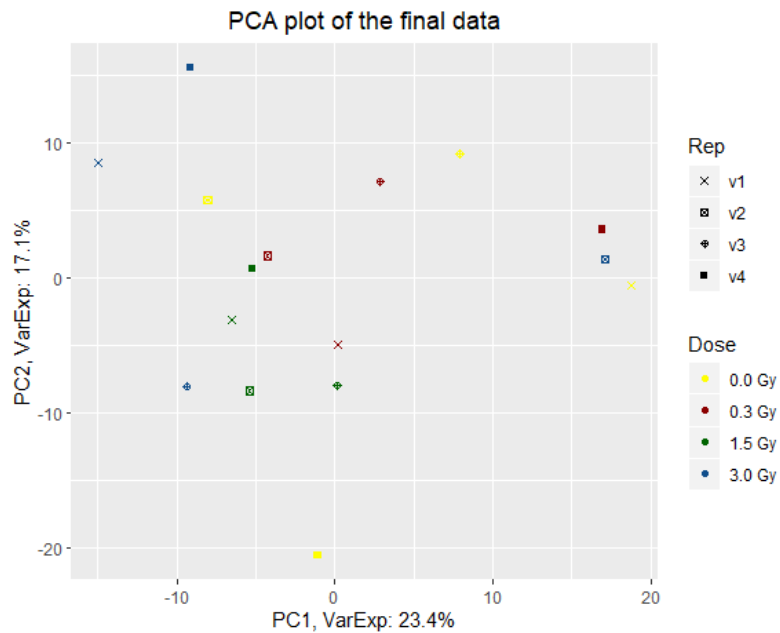
```
exprs_final = subset(normalized64375.filt, !ids_ex)
validObject(exprs_final)
[1] TRUE
fData(exprs_final)$PROBEID = rownames(fData(exprs_final))
fData(exprs_final) = left_join(fData(exprs_final), anno_exprs)
rownames(fData(exprs_final)) = fData(exprs_final)$PROBEID
validObject(exprs_final)
[1] TRUE
```

Τώρα που έχουμε κανονικοποιήσει, φιλτράρει και συνοψίσει τα δεδομένα μπορούμε να προχωρήσουμε στην επιθεώρηση ξεκινώντας με ένα διάγραμμα PCA (Εικόνα 22). Στην περίπτωση αυτή κάθε σημείο στην γραφική απεικόνιση αντιπροσωπεύει ένα δείγμα με το χρώμα να υποδεικνύει τη δόση ακτινοβολίας και το σχήμα τον εθελοντή (πειραματικό αντίγραφο).

```

exprs_final = biobase::exprs(exprs_final)
PCA = prcomp(t(exprs_final), scale = FALSE)
percentVar = round(100*PCA$sdev^2/sum(PCA$sdev^2),1)
sd_ratio = sqrt(percentVar[2] / percentVar[1])
dataGG = data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],
Dose = pData(exprs_final)$FactorValue..radiation.dose.,
Rep = pData(exprs_final)$FactorValue..volunteer.)
ggplot(dataGG, aes(PC1, PC2)) +
geom_point(aes(shape = Rep, colour = Dose)) +
ggtitle("PCA plot of the final data") +
xlab(paste0("PC1, VarExp: ", percentVar[1], "%")) +
ylab(paste0("PC2, VarExp: ", percentVar[2], "%")) +
theme(plot.title = element_text(hjust = 0.5)) +
coord_fixed(ratio = sd_ratio) +
scale_shape_manual(values = c(4,7,10,15)) +
scale_color_manual(values = c("yellow", "darkred", "darkgreen",
"dodgerblue4"))

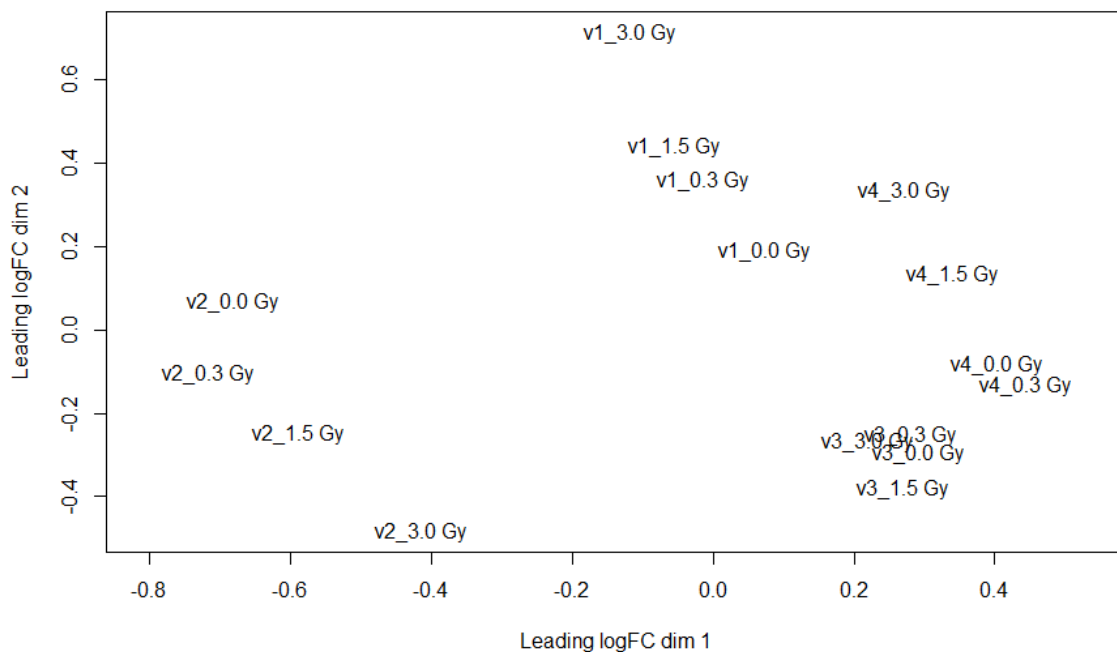
```



Εικόνα 22: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE64375.

Στο διάγραμμα μπορεί κανείς να παρατηρήσει ότι τα διαφορετικά βιολογικά αντίγραφα επηρεάζουν σημαντικότερα τις διαφορές στην γονιδιακή έκφραση (PC1) σε σχέση με τη δόση ακτινοβολίας, γεγονός που μπορεί να υποδηλώνουν την ύπαρξη Batch Effect. Παρόλα αυτά, σχεδιάστηκε το αντίστοιχο διάγραμμα MDS (Εικόνα 23), για το οποίο δημιουργήσαμε ετικέτες που συνδυάζουν την πληροφορία εθελοντή και δόσης:

```
labDose=as.factor(pData(exprs_final)$FactorValue..radiation.dose/.)
labRep=as.factor(pData(normalizedData)$FactorValue..volunteer.)
plotMDS(exprs(exprs_final), labels=paste(labRep,labDose,sep="_"))
```



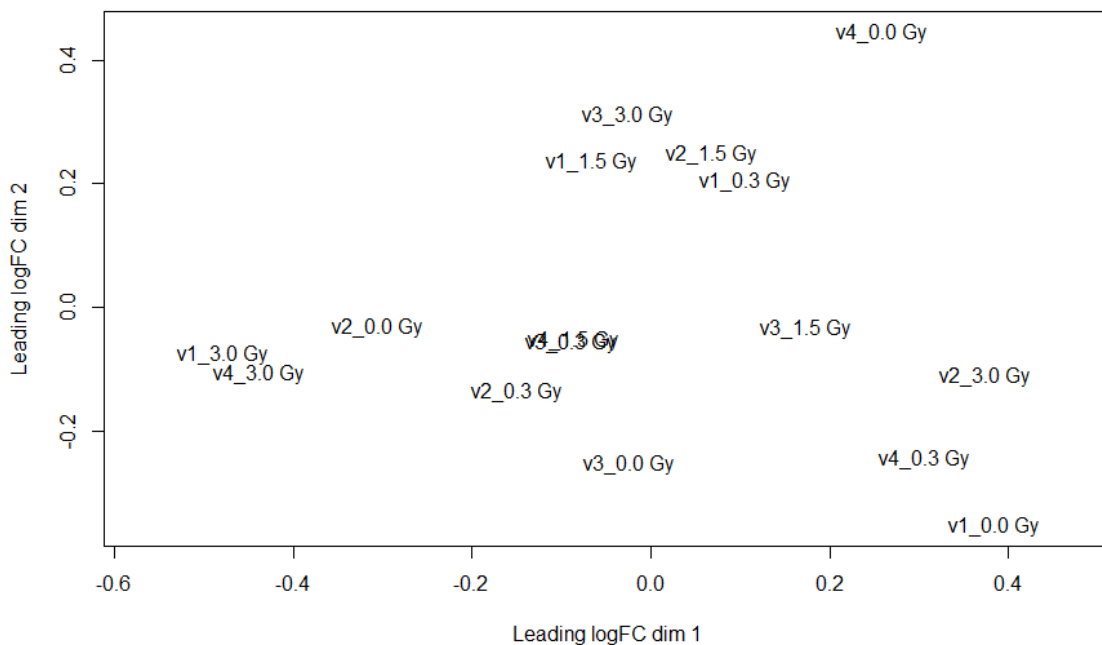
Εικόνα 23: Διάγραμμα MDS έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE64375. Κάθε σημείο αναπαριστά ένα δείγμα ενώ διαφαίνονται τέσσερις συστάδες που αντιστοιχούν στους τέσσερις διαφορετικούς εθελοντές (πειραματικά αντίγραφα).

Καθώς τα δείγματα όντως δημιουργούν συστάδες βάση των πειραματικών αντιγράφων, εκτελούμε την απαιτούμενη διόρθωση στα δεδομένα και σχεδιάζουμε το νέο διάγραμμα MDS (Εικόνα 24) που υποδεικνύει ότι τα δεδομένα έχουν προετοιμαστεί κατάλληλα για διαφορική ανάλυση:

```

batch.lab = c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4)
mod = matrix(0,ncol=4,nrow=16)
rownames(mod) = colnames(exprs_final)
colnames(mod) = c("Ctrl", "D03", "D15", "D30")
mod[label_dose=="0.0 Gy",2] = 1
mod[label_dose=="0.3 Gy",3] = 1
mod[label_dose=="1.5 Gy",4] = 1
mod[label_dose=="3.0 Gy",5] = 1
64375.combat_data = Combat(dat = exprs(exprs_final), batch =
batch.lab, mod = mod[,1:3], par.prior = TRUE, prior.plots = TRUE)
plotMDS(64375.combat_data, labels=paste(label_rep, label_dose,
/sep="_"))

```



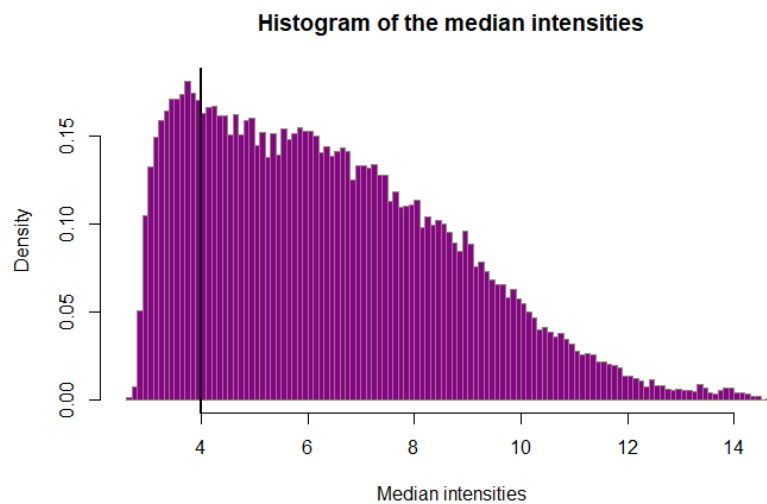
Εικόνα 24: Διάγραμμα MDS των διορθωμένων δεδομένων της σειράς GSE64375. Κάθε σημείο αναπαριστά ένα δείγμα ενώ οι συστάδες που αντιστοιχούσαν στους τέσσερις διαφορετικούς εθελοντές (βιολογικά αντίγραφα) έχουν εξαλειφθεί.

GSE63952: Διαχειριστήκαμε αυτή την σειρά δεδομένων με παρόμοιο τρόπο με την προηγούμενη, χρησιμοποιώντας όμως για τον σχολιασμό των ανιχνευτών την βιβλιοθήκη primeview.db [127]. Οφείλει να σημειωθεί το συγκεκριμένο πακέτο δεν διατίθεται για την έκδοση της R 3.6.1 αλλά μεταφέρθηκε από προηγούμενη έκδοση κατά την αναβάθμιση. Από τον έλεγχο των μη επεξεργασμένων δεδομένων με θηκογράμματα (Εικόνα 25) παρατηρήθηκε ότι δεν υπάρχει κάποιο δείγμα που πρέπει να αφαιρεθεί από την ανάλυση.



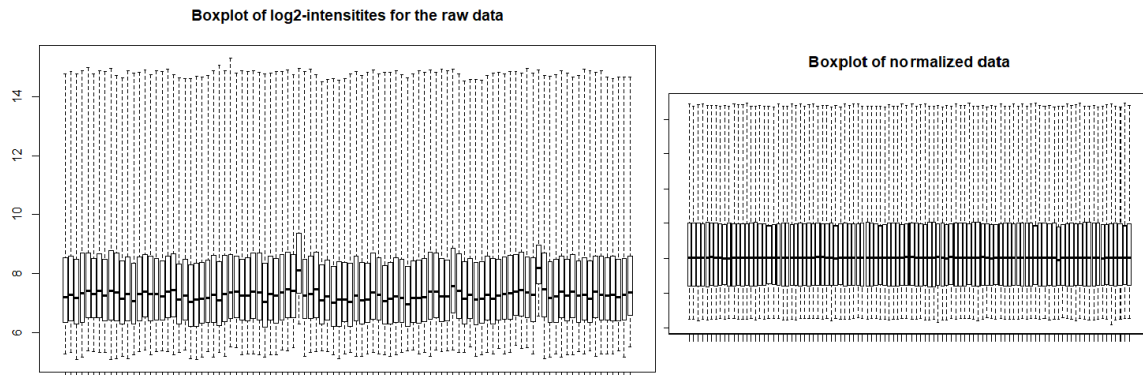
Εικόνα 25: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς GSE63952.

Στη συνέχεια σχεδιάζοντας το ιστόγραμμα των μέσων εντάσεων (Εικόνα 26) καταλήξαμε να αφαιρέσουμε 6224 μεταγράφα.



Εικόνα 26: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς GSE63952.

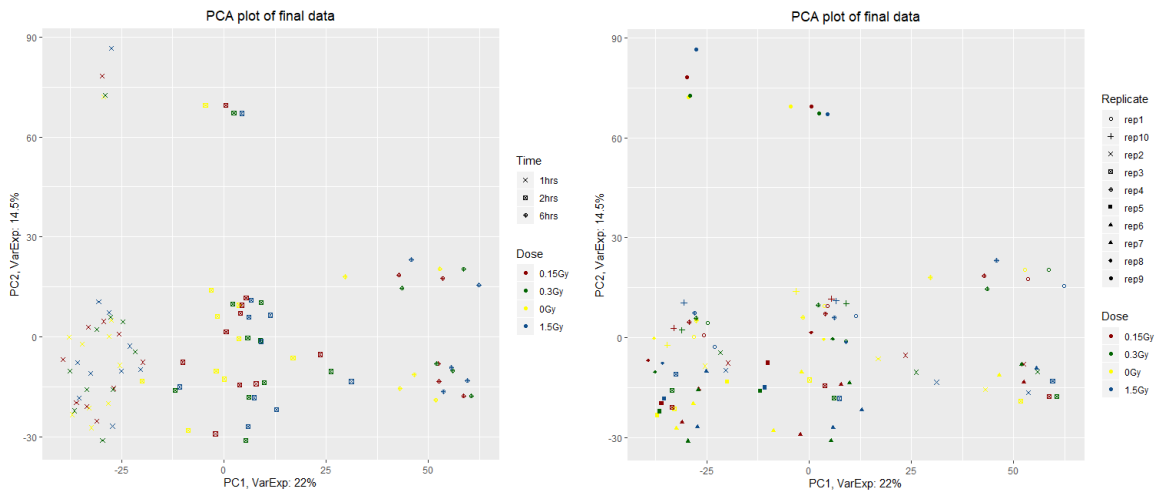
Εφαρμόσαμε τον πλήρες αλγόριθμο RMA και ελέγξαμε ότι εφαρμόστηκε επιτυχώς με τα αντίστοιχα θηκογράμματα (Εικόνα 27).



Εικόνα 27: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά GSE63952.

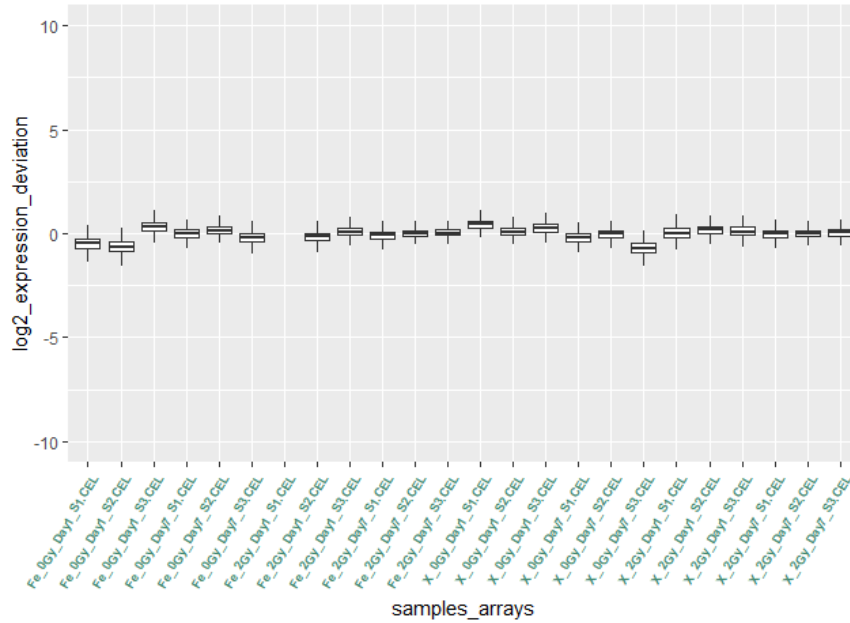
Κατά την διαδικασία απαλοιφής των πολλαπλών χαρτογραφήσεων ο κώδικας ήταν ο ίδιος με τον παραπάνω με τη μόνη διαφορά ότι εφόσον έχουμε την τεχνολογία Affymetrix Primeview χρησιμοποιούμε την αντίστοιχη βιβλιοθήκη. Αφαιρέθηκαν άλλα 2276 μετάγραφα. Τέλος σχεδιάστηκαν δύο διαγράμματα PCA, όπου υποδεικνύεται με χρώμα η δόση ακτινοβολίας και με σχήματα ο χρόνος και τα πειραματικά αντίγραφα αντίστοιχα (Εικόνα 28). Διαπιστώθηκε ότι δεν υπάρχει Batch Effect και άρα μπορούμε να προχωρήσουμε σε διαφορική ανάλυση.

```
dataGGTIME <- data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],  
Dose = Biobase::pData(exprs_final)$Dose,  
Time = Biobase::pData(exprs_final)$Time)  
dataGGREP <- data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],  
Dose = Biobase::pData(exprs_final)$Dose,  
Replicate = Biobase::pData(exprs_final)$Replicate)
```



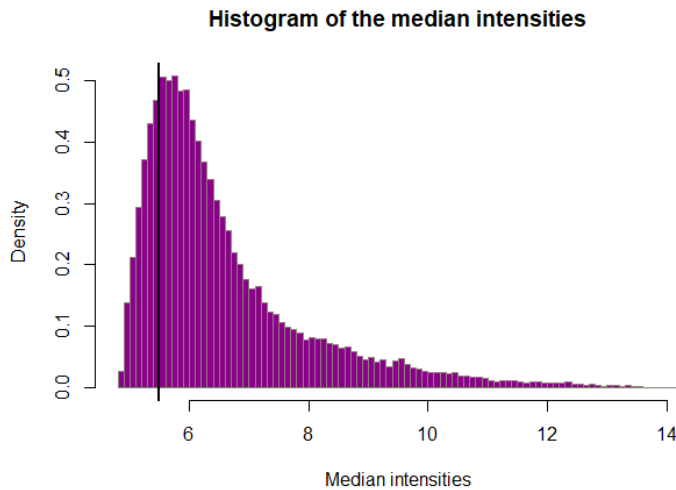
Εικόνα 28: Διαγράμματα PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς GSE63952. Στο πρώτο διάγραμμα φαίνεται ότι οι διαφορές στην έκφραση διαμορφώνονται βάση των ωρών συλλογής των δειγμάτων έπειτα από ακτινοβολήση. Από το δεύτερο διάγραμμα φαίνεται ότι οι συστάδες είναι ανεξάρτητες των πειραματικών αντιγράφων, με εξαίρεση την επανάληψη 9 (rep9).

E-MTAB-5754: Και οι τρεις παρακάτω μελέτες που βρέθηκαν στο ArrayExpress ανήκουν σε εκδόσεις GeneChip της Affymetrix. Από τον έλεγχο των μη επεξεργασμένων δεδομένων με θηκογράμματα (Εικόνα 29) παρατηρήθηκε ότι δεν υπάρχει κάποιο δείγμα που πρέπει να αφαιρεθεί από την ανάλυση. Παρόλα αυτά ένα δείγμα ακτινοβολημένο με ιόντα σιδήρου που συλλέχθηκε μία ημέρα μετά την ακτινοβολήση εμφάνιζε κενές τιμές κατά την διόρθωση υποβάθρου. Έτσι ο αλγόριθμος RMA εφαρμόστηκε χωρίς διόρθωση υποβάθρου καθώς κατά πάσα πιθανότητα η πληροφορία για την ένταση υποβάθρου στα .CEL αρχεία ήταν λανθασμένη και η κανονικοποίηση ποσοστημορίου απέδιδε κενές τιμές. Στη συνέχεια σχεδιάζοντας το ιστόγραμμα των μέσων εντάσεων (Εικόνα 30) καταλήξαμε να αφαιρέσουμε 8250 μετάγραφα.



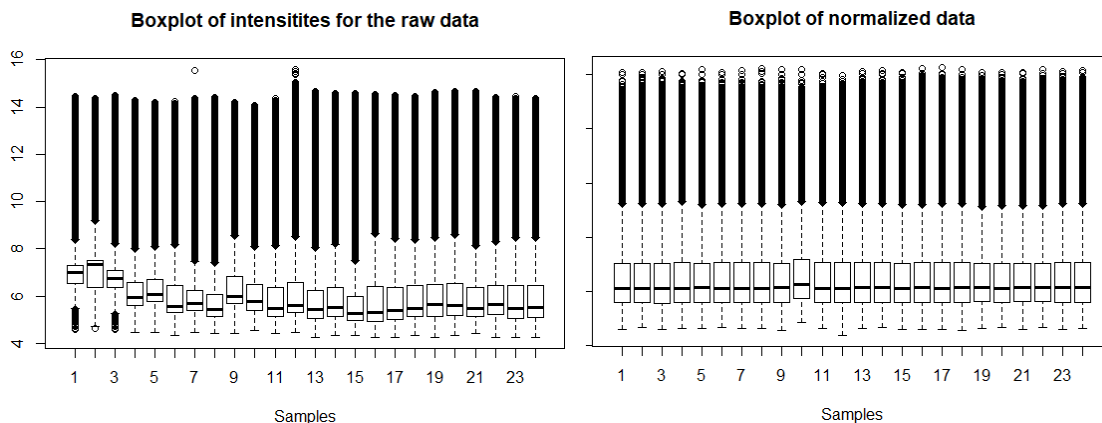
Εικόνα 29: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς E-MTAB-5754.

```
normalized5754 = oligo::rma(raw, target="core", background=FALSE)
```



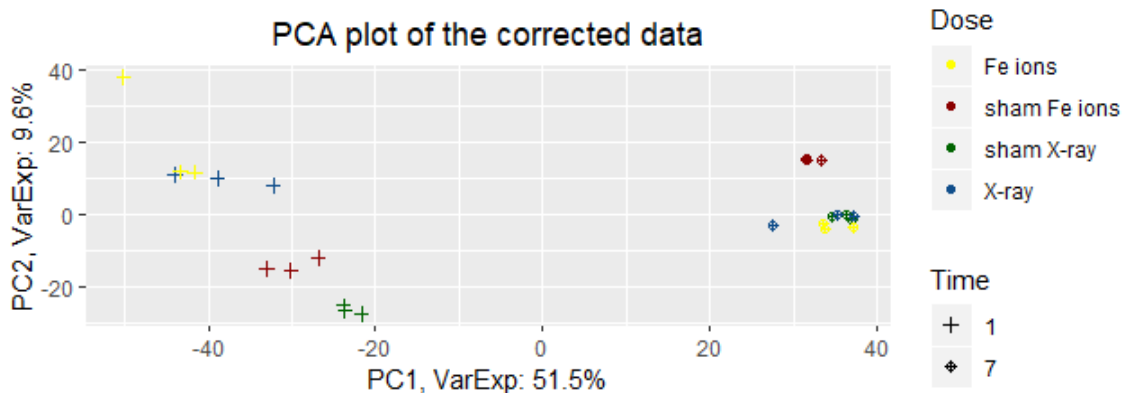
Εικόνα 30: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-5754.

Μετά από την εφαρμογή της κανονικοποίησης ελέγχοντας τα δεδομένα έκφρασης με θηκογράμματα (Εικόνα 31), βεβαιωθήκαμε ότι η ανάλυση μπορεί να συνεχιστεί παρά την παράλειψη της διαδικασίας διόρθωσης υποβάθρου.



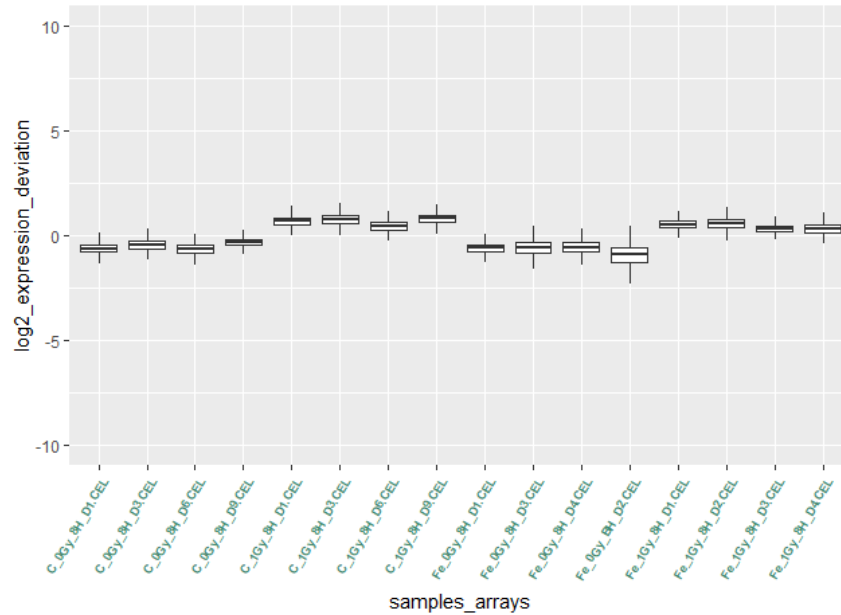
Εικόνα 31: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-5754.

Κατά την διαδικασία απαλοιφής των πολλαπλών χαρτογραφήσεων με την χρήση της βιβλιοθήκης `hugene20sttranscriptcluster.db` [128] αφαιρέθηκαν ακόμα 2163 ανιχνευτές. Στη συνέχεια σχεδιάστηκε το διάγραμμα PCA (Εικόνα 32) με πρώτη κύρια συνιστώσα τις ώρες συλλογής των δειγμάτων έπειτα από ακτινοβολία και δεύτερη τον τύπο σε συνδυασμό με την δόση της ακτινοβολίας και διαπιστώθηκε πως δεν υπάρχει Batch Effect.



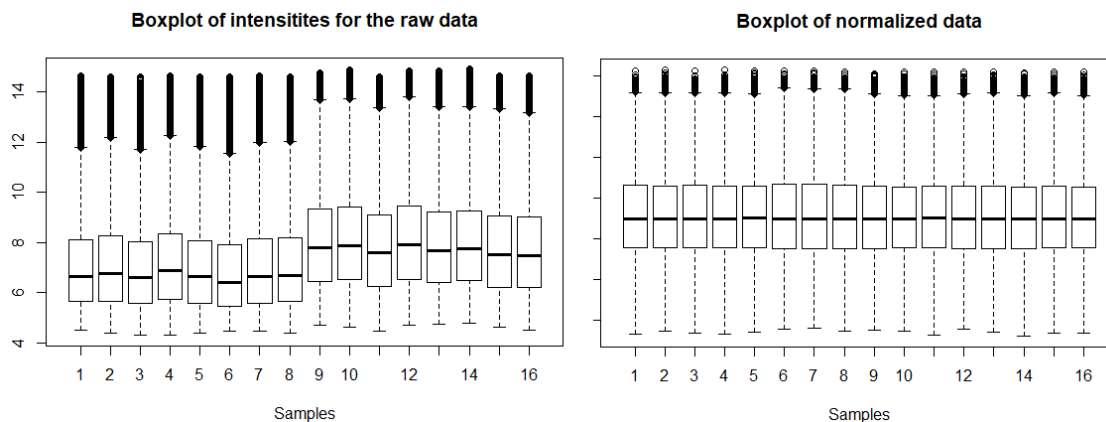
Εικόνα 32: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-5754.

E-MTAB-5761: Σε αυτή τη μελέτη από τον έλεγχο των μη επεξεργασμένων δεδομένων με θηκογράμματα (Εικόνα 33) παρατηρήθηκε ότι δεν υπάρχει κάποιο δείγμα που πρέπει να αφαιρεθεί από την ανάλυση.

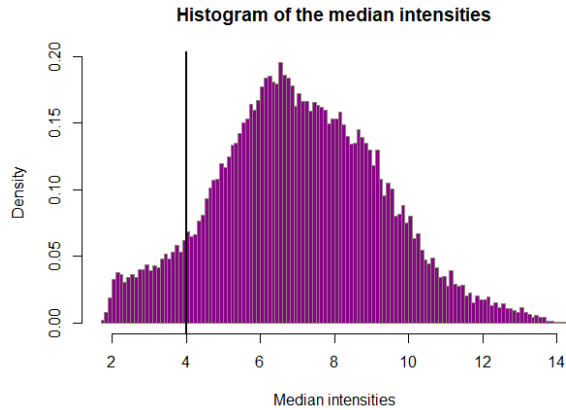


Εικόνα 33: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς E-MTAB-5761.

Ο αλγόριθμος RMA εφαρμόστηκε και ελέγχθηκε (Εικόνα 34) για την επεξεργασία των δεδομένων έκφρασης. Στη συνέχεια, μέσω της διαδικασίας φιλτραρίσματος με βάση την ένταση (Εικόνα 35), αφαιρέθηκαν 2517 μεταγράφα.

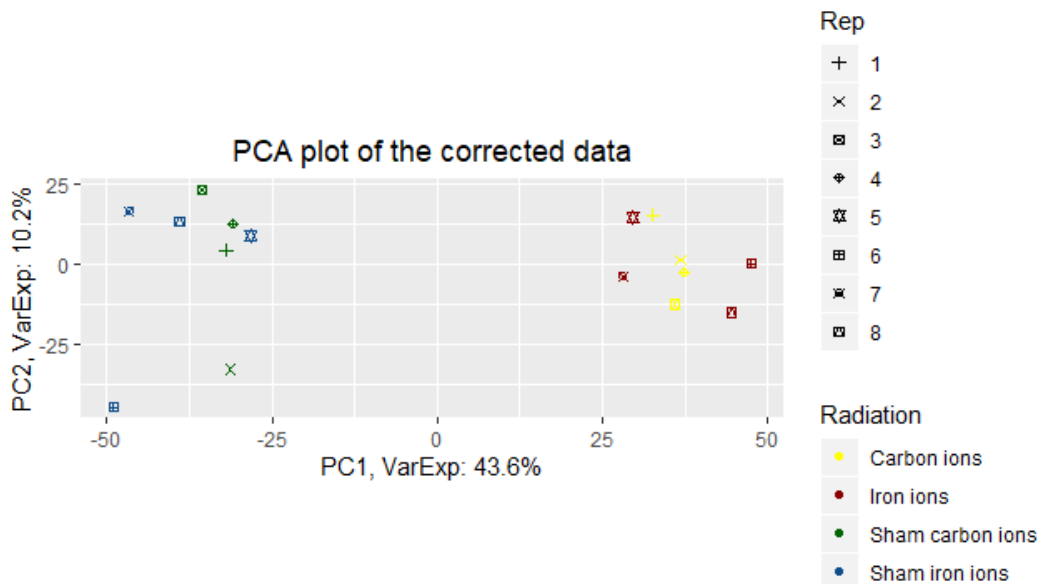


Εικόνα 34: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-5761.



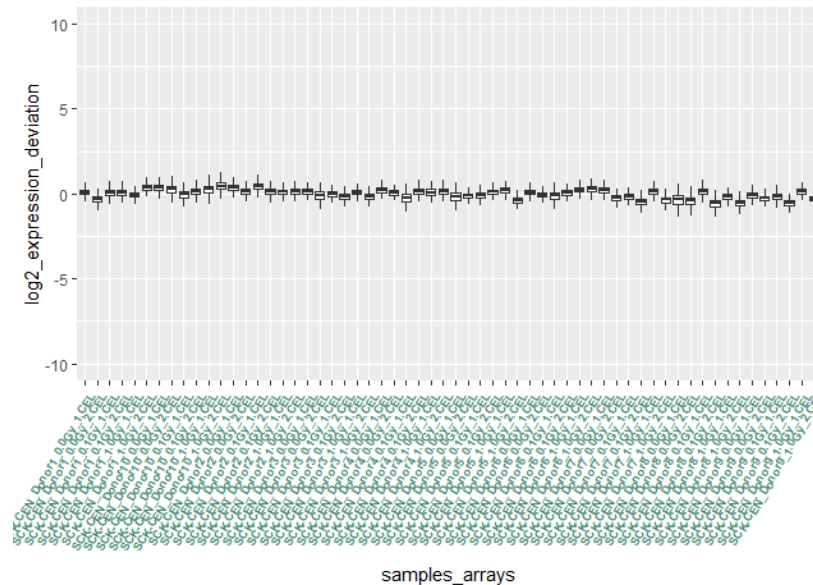
Εικόνα 35: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-5761.

Κατά την διαδικασία απαλοιφής των πολλαπλών χαρτογραφήσεων με την χρήση της βιβλιοθήκης hugene10sttranscriptcluster.db [126] αφαιρέθηκαν ακόμα 2144 μεταγράφα. Τέλος σχεδιάστηκε το διάγραμμα PCA (Εικόνα 36). Στο διάγραμμα γίνεται αντιληπτό ότι ο βασικός διαχωρισμός στο προφίλ έκφρασης εμφανίζεται μεταξύ ακτινοβολημένων δειγμάτων και δειγμάτων αναφοράς, γεγονός που είναι βιολογικώς αναμενόμενο. Έτσι συμπεραίνουμε ότι δεν υπάρχει Batch Effect.



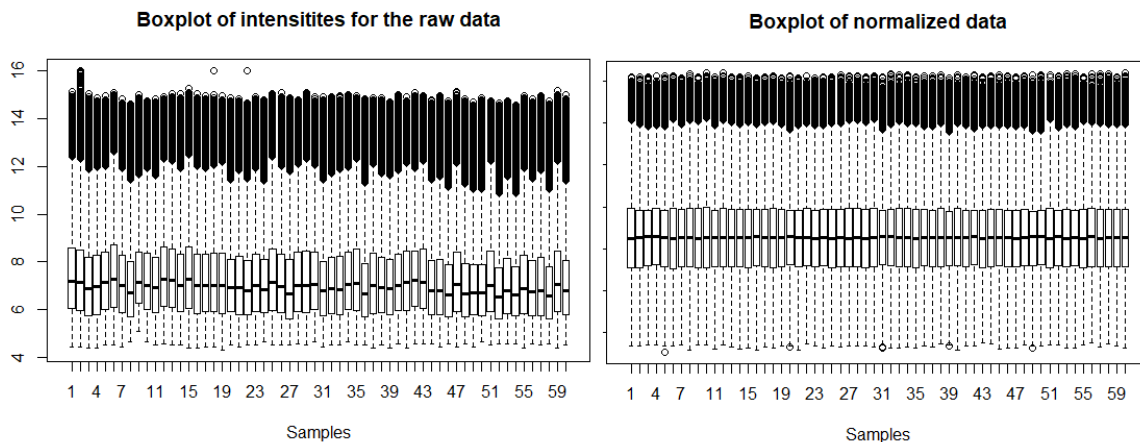
Εικόνα 36: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-5761.

E-MTAB-3463: Και σε αυτή τη μελέτη από τον έλεγχο των μη επεξεργασμένων δεδομένων με θηκογράμματα (Εικόνα 37) παρατηρήθηκε ότι δεν υπάρχει κάποιο δείγμα που πρέπει να αφαιρεθεί από την ανάλυση.

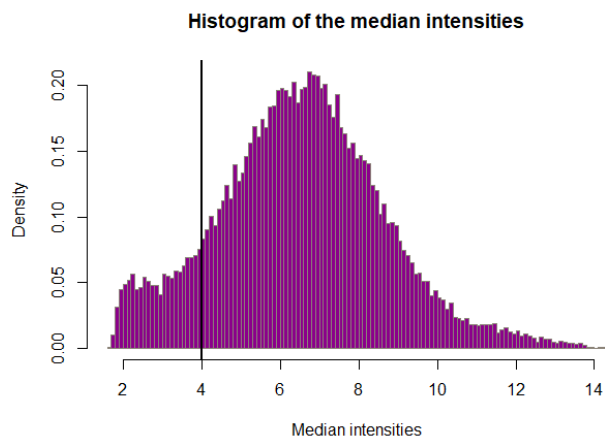


Εικόνα 37: Θηκογράμματα (boxplots) της απόκλισης της έντασης έκφρασης από την μέση ένταση για κάθε μετάγραφο σε όλα τα δείγματα της σειράς E-MTAB-3463.

Ο αλγόριθμος RMA εφαρμόστηκε και ελέγχθηκε (Εικόνα 38) κανονικά για την επεξεργασία των δεδομένων. Η διαδικασίας φιλτραρίσματος με βάση την ένταση (Εικόνα 39 Εικόνα 35) οδήγησε στην αφαίρεση 3702 μεταγράφων.

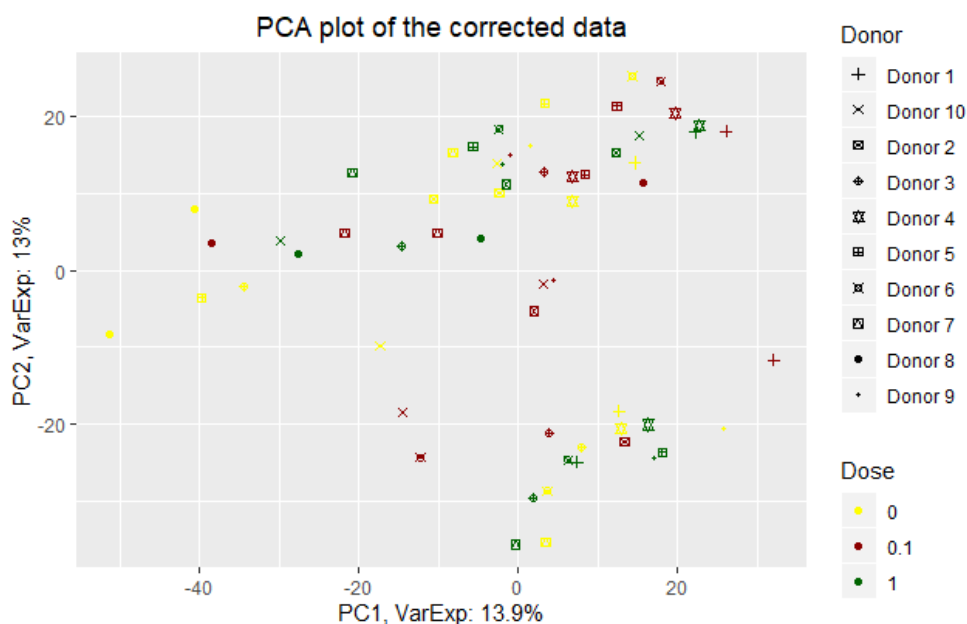


Εικόνα 38: Αναπαράσταση δεδομένων έκφρασης με θηκογράμματα (boxplots) πριν και μετά την εφαρμογή του αλγορίθμου RMA για την σειρά E-MTAB-3463.



Εικόνα 39: Ιστόγραμμα για τις μέσες εντάσεις έκφρασης των μεταγράφων (ανιχνευτών) της σειράς E-MTAB-3463.

Κατά την διαδικασία απαλοιφής των πολλαπλών χαρτογραφήσεων όπως παραπάνω, αφαιρέθηκαν ακόμα 2079 μεταγράφα. Τέλος σχεδιάστηκε το διάγραμμα PCA (Εικόνα 40). Καθώς τα δείγματα δεν σχηματίζουν συστάδες, συμπεραίνουμε ότι δεν υπάρχει Batch Effect.



Εικόνα 40: Διάγραμμα της PCA έπειτα από κανονικοποίηση, φιλτράρισμα και σύνοψη των δεδομένων της σειράς E-MTAB-3463. Πρώτη κύρια συνιστώσα αποτελεί ο τύπος και η δόση της ακτινοβολίας και δεύτερη τα πειραματικά αντίγραφα.

2.3.3. Διαφορική Έκφραση Γονιδίων

Υπό την προϋπόθεση ότι έχουμε επεξεργαστεί και ελέγξει τα δεδομένα μας, βρισκόμαστε στο στάδιο παραγωγής των λιστών διαφορικά εκφρασμένων γονιδίων για κάθε σύνολο δεδομένων. Δημιουργούμε συγκρίσεις μεταξύ των εξεταζόμενων ή υπό συνθήκη δειγμάτων και των δειγμάτων αναφοράς (*control vs condition*) σε κάθε μελέτη. Η «συνθήκη» εξαρτάται από τη δόση και τον τύπο της ακτινοβολίας, το χρόνο μετά την έκθεση, καθώς και από τον τύπο των κυττάρων σε περιπτώσεις όπου χρησιμοποιούνται διαφορετικοί τύποι σε ένα πείραμα. Χρησιμοποιώντας τη συνάρτηση *eBayes()* του πακέτου *limma* παίρνουμε τιμές \log_2FC , *p-value* και προσαρμοσμένου *p-value* για κάθε γονίδιο σε κάθε σύγκριση, όπως υπολογίζεται μέσω Μπεϋζιανής εκτιμητικής [77, 78].

2.3.3.1 *Illumina*

Η διαδικασία αντιστοίχισης των ανιχνευτών σε γονίδια ως προς το HGNC Gene Symbol συστήνεται να εκτελείται σε αυτό το στάδιο ακριβώς πριν την διαφορική ανάλυση, ώστε να μην χάνεται σημαντική στατιστική πληροφορία. Για το σκοπό αυτό στην περίπτωση της *Illumina* χρησιμοποιείται η εκάστοτε βάση δεδομένων, που αναφέρθηκε παραπάνω κατά την διαδικασία φιλτραρίσματος των ανιχνευτών με βάση την ποιότητα. Στη συνέχεια εκτελείται διαφορική ανάλυση με χρήση της συνάρτησης *eBayes()* η οποία απαιτεί ως ορίσματα έναν πίνακα μοντελοποίησης, που περιέχει τον χαρακτηρισμό των δειγμάτων με βάση τα φαινοτυπικά-πειραματικά χαρακτηριστικά και ένα μοντέλο γραμμικής προσαρμογής (*linear model fit*) για κάθε γονίδιο δεδομένου ενός συνόλου μικροσυστοιχιών, το οποίο λαμβάνεται από την συνάρτηση *lfit()*. Έτσι κατασκευάζεται πίνακας *result_Bayes*, που περιέχει τις τροποποιημένες υπολογιζόμενες τιμές *t* και *F* στατιστικής καθώς και τα \log_2FC μέσω εμπειρικής Μπεϋζιανής τροποποίησης των τυπικών σφαλμάτων προς μια κοινή τιμή.

GSE80547: Για να αντιστοιχήσουμε τους ανιχνευτές στα HGNC Gene Symbols χρησιμοποιούμε την συνάρτηση *illuminaHumanv4SYMBOL* της

βάσης *illuminaHumanv4* [118] στα διορθωμένα δεδομένα του πίνακα *combat_data*. Αντιστοιχούμε τα ονόματα των ανιχνευτών από την πρώτη στήλη των δεδομένων αυτών με τα HGNC Gene Symbols από την βάση δεδομένων και τα αντικαθιστούμε με αυτά:

```
ids = as.character(rownames(80547.combat_data))
symbol = mget(ids, illuminaHumanv4SYMBOL , ifnotfound = NA)
anno = data.frame(Ill_ID = ids, Symbol = as.character(symbol))
rownames(80547.combat_data) = anno[,2]
```

Στη συνέχεια χρησιμοποιούμε τον πίνακα μοντελοποίησης *mod* που κατασκευάστηκε παραπάνω για την διόρθωση του Batch Effect ονοματίζοντας τις στήλες του ώστε να μπορούμε να ορίσουμε στην συνέχεια τις στήλες που θα συγκριθούν. Στο αντικείμενο *fit* περιέχεται το μοντέλο γραμμικής προσαρμογής, το οποίο χρησιμοποιεί τα διορθωμένα δεδομένα και τα αντίστοιχα υπολογιζόμενα βάρη από την συνάρτηση *arrayWeights()*. Τέλος μέσω της συνάρτησης *makeContrasts()*, χρησιμοποιώντας τα ονόματα στηλών του *mod*, καθορίζονται οι συγκρίσεις που επιθυμούμε να διεξαχθούν. Ο πίνακας *result_Bayes* περιέχει τα τελικά στατιστικά δεδομένα τα οποία μπορούμε να προσπελάσουμε με την συνάρτηση *topTable()*, όπου το όρισμα «*coef*» αναπαριστά τον αύξοντα αριθμό σύγκρισης που διεξήχθη, ενώ στο όρισμα «*p_value*» επιλέγεται η τιμή κατωφλίου για το υπολογιζόμενο adjusted *p_value* για κάθε γονίδιο σε κάθε σύγκριση:

```
colnames(mod) = c("Ctrl", "Dose1", "Dose2", "Dose3")
aw = arrayweights (80547.combat_data , mod )
fit = lmFit(80547.combat_data , mod, weights = aw)
contrasts = makeContrasts( Ctrl - Dose1 , Ctrl - Dose2, Ctrl - Dose3,
/levels = mod )
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
C_Dose1 = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
C_Dose2 = topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
C_Dose3 = topTable(results.Bayes, n=10000, coef = 3, p.value = 0.05)
```

GSE81067: Για να αντιστοιχήσουμε τους ανιχνευτές στα HGNC Gene Symbols σε αυτή την περίπτωση χρησιμοποιούμε την συνάρτηση *illuminaHumanv3SYMBOL* της βάσης *illuminaHumanv3* [117] στα διορθωμένα δεδομένα έκφρασης του πίνακα *normalized81067.filt.normal*,

όπου περιέχονται μόνο οι στήλες που αντιστοιχούν σε δείγματα από φυσιολογικό ανθρώπινο ιστό κατά ανάλογο τρόπο:

```
f.exprs = exprs(normalized81067.filt.normal)
ids = as.character(rownames(f.exprs))
symbol = mget(ids, illuminaHumanv3SYMBOL , ifnotfound = NA)
anno = data.frame(ill_ID = ids, symbol = as.character(symbol))
rownames(f.exprs) = anno[,2]
```

Κατασκευάζουμε στην συνέχεια πίνακα μοντελοποίησης *mod* με βάση τις οχτώ κατηγορίες δειγμάτων όπως προκύπτουν από τις τέσσερις τιμές δόσης στους δύο τύπους ιστού, ονοματίζουμε ανάλογα τις στήλες αυτού και χρησιμοποιούμε την λίστα *normal_GSMs* για τα ονόματα των γραμμών. Η διαφορική ανάλυση εκτελείται στη συνέχεια όπως και παραπάνω:

```
mod=model.matrix(~0+factor(c(5,8,2,6,4,5,1,1,7,3,4,5,3,4,2,6,2,3,1,
/7,7,3,6,8,1,2,8,5,4,8,8,7,3,5,4,1,6,7,2,6)))
rownames(mod) = normal.GSMs
colnames(mod)=c("HLF.0", "HLF.05", "HLF.1", "HLF.15", "HPEAC.0",
/"HPEAC.05", "HPEAC.1", "HPEAC.15")
aw = arrayweights (f.exprs, mod[1:8])
fit = lmFit(f.exprs, mod[1:8], weights = aw)
contrasts = makeContrasts(HLF.0 - HLF.05, HLF.0 - HLF.1,
/HLF.0 - HLF.15, HPEAC.0 - HPEAC.05, HPEAC.0 - HPEAC.1,
/HPEAC.0 - HPEAC.15, levels = mod[1:8])
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
PE_Dose1 = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
PE_Dose2 = topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
PE_Dose3 = topTable(results.Bayes, n=10000, coef = 3, p.value = 0.05)
LF_Dose1 = topTable(results.Bayes, n=10000, coef = 4, p.value = 0.05)
LF_Dose2 = topTable(results.Bayes, n=10000, coef = 5, p.value = 0.05)
LF_Dose3 = topTable(results.Bayes, n=10000, coef = 6, p.value = 0.05)
```

GSE44282: Στην περίπτωση αυτής της σειράς δεδομένων χρησιμοποιούμε την συνάρτηση *illuminaHumanv2SYMBOL* της βάσης *illuminaHumanv2* [122] στα διορθωμένα δεδομένα του πίνακα *44282.combat_data* για την αντιστοίχιση ανιχνευτών σε HGNC Gene Symbol:

```
ids = as.character(rownames(44282.combat_data))
symbol = mget(ids, illuminaHumanv2SYMBOL , ifnotfound = NA)
anno = data.frame(ill_ID = ids, symbol = as.character(symbol))
rownames(44282.combat_data) = anno[,2]
```

Ομοίως κατασκευάζουμε πίνακα μοντελοποίησης *mod2* για τις κατηγορίες των δειγμάτων με τα ανάλογα ονόματα στηλών και διεξάγουμε την διαφορική ανάλυση:

```
mod2=model.matrix(~0+factor(c(1,1,1,1,1,1,1,1,1,1,1,2,2,5,2,3,5,2,3,
/5,6,8,9,6,8,6,7,8,9,1,1,1,1,1,1,2,3,4,5,6,7,8,9,8,19,19,19,19,19,
/20,21,22,23,24,25,26,27,10,10,10,10,10,11,12,13,14,15,16,17,18,19,
/19,19,19,19,20,21,22,23,24,25,26,27,10,10,10,10,10,11,12,13,14,15,
/16,17,18,1,1,1,1,1,2,3,4,5,6,7,8,9,10,10,10,10,10,11,12,13,14,15,
/16,17,18,19,19,19,19,19,20,21,22,23,24,25,26,27)))
colnames(mod2)=c("FeCtrl", "Fe05_1", "Fe05_4", "Fe05_12", "Fe05_24",
/"Fe1_1", "Fe1_4", "Fe1_12", "Fe1_24", "SiCtrl", "Si05_1", "Si05_4",
/"Si05_12", "Si05_24", "Si1_1", "Si1_4", "Si1_12", "Si1_24", "XrayCtrl",
/"Xray1_1", "Xray1_4", "Xray1_12", "Xray1_24", "Xray3_1", "Xray3_4",
/"Xray3_12", "Xray3_24")
aw = arrayweights(44282.combat_data , mod2)
fit = lmFit(44282.combat_data, mod2, weights = aw)
contrasts = makeContrasts(FeCtrl - Fe05_1, FeCtrl - Fe05_4,
/FeCtrl - Fe05_12, FeCtrl - Fe05_24, FeCtrl - Fe1_1,
/FeCtrl - Fe1_4, FeCtrl - Fe1_12, FeCtrl - Fe1_24, SiCtrl - Si05_1,
/SiCtrl - Si05_4, SiCtrl - Si05_12, SiCtrl - Si05_24,
/SiCtrl - Si1_1, SiCtrl - Si1_4, SiCtrl - Si1_12,
/SiCtrl - Si1_24, XrayCtrl - Xray1_1, XrayCtrl - Xray1_4,
/XrayCtrl - Xray1_12, XrayCtrl - Xray1_24, XrayCtrl - Xray3_1,
/XrayCtrl - Xray3_4, XrayCtrl - Xray3_12, XrayCtrl - Xray3_24,
/levels = mod2)
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
C_Fe05.1 = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
```

...

2.3.3.2 Affymetrix

Για τα δεδομένα της Affymetrix η αντιστοίχιση των ανιχνευτών σε HGNC Gene Symbols έχει ήδη λάβει χώρα κατά το στάδιο αφαίρεσης των πολλαπλών χαρτογραφήσεων. Έτσι πριν την διεξαγωγή της Μπεϋζιανής διαφορικής ανάλυσης θα πρέπει να προστεθεί στους πίνακες διορθωμένων δεδομένων έκφρασης, οι στήλες με τα αντίστοιχα HGNC Gene Symbols. Εφόσον προστεθεί η στήλη με τα ονόματα των γονιδίων, ο κώδικας για διεξαγωγή της διαφορικής ανάλυσης είναι ανάλογος με παραπάνω με τις διαφορές να εμφανίζονται στον πίνακα μοντελοποίησης κάθε συνόλου δεδομένων.

GSE64375

```
probes = fData(64375.exprs_final)
rownames(64375.combat_data)=paste(probes$PROBEID,probes$SYMBOL,
/sep="_")
aw = arrayWeights(64375.combat_data, mod[1:3])
fit = lmFit(64375.combat_data, mod[1:3], weights = aw)
contrasts = makeContrasts(Ctrl-Dose0.3,Ctrl-Dose1.5,Ctrl-Dose3.0,
/levels = mod[1:3])
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
C_Dose03 = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
C_Dose15 = topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
C_Dose30 = topTable(results.Bayes, n=10000, coef = 3, p.value = 0.05)
```

GSE63952

```
rownames(63952.exprs_final)=paste(probes$PROBEID,probes$SYMBOL,
/sep="_")
mod=model.matrix(~0+factor(c(1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,
/2,2,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,4,5,5,5,5,5,5,5,5,5,5,5,6,6,6,6,6,6,7,
/7,7,7,7,7,7,7,7,7,8,8,8,8,8,8,8,8,8,8,8,9,9,9,9,9,10,10,10,10,10,10,
/10,10,10,10,11,11,11,11,11,11,11,11,11,11,11,11,11,11,12,12,12,12,12)))
aw = arrayWeights(63952.exprs_final , mod)
fit = lmFit(63952.exprs_final, mod, weights = aw)
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
Dose015.1=topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
Dose030.1=topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
Dose150.1=topTable(results.Bayes, n=10000, coef = 3, p.value = 0.05)
Dose015.2=topTable(results.Bayes, n=10000, coef = 4, p.value = 0.05)
Dose030.2=topTable(results.Bayes, n=10000, coef = 5, p.value = 0.05)
Dose150.2=topTable(results.Bayes, n=10000, coef = 6, p.value = 0.05)
Dose015.6=topTable(results.Bayes, n=10000, coef = 7, p.value = 0.05)
Dose030.6=topTable(results.Bayes, n=10000, coef = 8, p.value = 0.05)
Dose150.6=topTable(results.Bayes, n=10000, coef = 9, p.value = 0.05)
```

E-MTAB-5754

```
labelDose = as.factor(pData(normalized)$Factor.Value..irradiate.)
labelTime=as.factor(pData(normalized)$Factor.Value..time.)
mod=model.matrix(~0+as.factor(paste(labelDose,labelTime,sep="_")))
colnames(mod)=c("Fe1","Fe7","CFe1","CFe7","X1","X7","CX1","CX7")
rownames(mod) = pData(5754.exprs_final)$Assay.Name
aw = arrayWeights(5754.exprs_final, mod)
fit =- lmFit(5754.exprs_final, mod, weights = aw)
contrasts = makeContrasts(CFe1-Fe1,CX1-X1, levels = mod)
results.Bayes = eBayes(contrasts.fit(fit , contrasts))
C_Fe = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
C_Xray = topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
```


E-MTAB-5761

```
mod=model.matrix(~0+as.factor(Factor.value.irradiate.),
/data=pData(5761norm.filt))
colnames(mod)=c("Carbon","Fe", "C.Carbon", "C.Fe")
aw = arrayweights(5761.exprs_final, mod)
fit = lmFit(5761.exprs_final, mod, weights = aw)
contrasts = makeContrasts(C.Carbon-Carbon,C.Fe-Fe, levels = mod)
results.Bayes = eBayes(contrasts.fit(fit ,contrasts))
C_Fe = topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
C_Carbon = topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
```

E-MTAB-3463

```
mod=model.matrix(~0+as.factor(Factor.value.dose.),
/data=pData(3463norm.filt))
colnames(mod)=c("Ctrl","Dose0.1","Dose1")
aw = arrayweights(3463.exprs_final, mod)
fit = lmFit(exprs_final, mod, weights = aw)
contrasts = makeContrasts(Ctrl-Dose0.1,Ctrl-Dose1, levels = mod)
results.Bayes = eBayes(contrasts.fit(fit ,contrasts))
C_Xrays01=topTable(results.Bayes, n=10000, coef = 1, p.value = 0.05)
C_Xrays10=topTable(results.Bayes, n=10000, coef = 2, p.value = 0.05)
```

Σε αυτό το στάδιο έχουμε μια πληθώρα από λίστες διαφορετικώς εκφρασμένων γονιδίων, που προέκυψαν από τις σειρές δεδομένων μικροσυστοιχιών από το GEO και το ArrayExpress. Οι λίστες αυτές περιέχουν τα ονόματα των γονιδίων υπό την μορφή των HGNC Gene Symbols, το \log_2FC και τα υπολογιζόμενα adjusted p-values. Το όριο στατιστικής σημαντικότητας τέθηκε ως $\text{adjusted } p\text{-value} < 0.05$. Οι λίστες προέκυψαν μέσω της σύγκρισης ακτινοβολημένων και μη ακτινοβολημένων ανθρώπινων δειγμάτων, ομαδοποιημένων βάση παραμέτρων όπως η δόση και ο τύπος ακτινοβολίας, ο χρόνος μετά την ακτινοβολήση και ο τύπος κυττάρων του δείγματος.

2.4. Μελέτη των επιδράσεων ακτινοβολίας

Στο στάδιο αυτό επιδιώκουμε να προσδώσουμε βιολογικό νόημα στις λίστες που προέκυψαν κατά την ανάλυση των δεδομένων μικροσυστοιχιών με χρήση μεθόδων λειτουργικής ανάλυσης εμπλουτισμού. Σκοπός είναι η ανάδειξη των διαφορών στην απόκριση του κυττάρου έπειτα από έκθεση σε ακτινοβολία υψηλής και χαμηλής τιμής LET, μέσω των όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών που θα προκύψουν, καθώς και η συνολική κατανόηση της απόκρισης σε κάθε περίπτωση. Επιπλέον θα μελετηθούν τέσσερις βασικές κατηγορίες συγγενών παθήσεων, που έχουν συσχετιστεί με την έκθεση σε σωματιδιακές ακτινοβολίες και ο τρόπος με τον οποίο αυτές μπορεί να προκύπτουν ως αποτέλεσμα της έκθεσης σε ακτινοβολία υψηλού LET.

2.4.1. Λειτουργική Ανάλυση Εμπλουτισμού

Για την λειτουργική ανάλυση εμπλουτισμού επιλέχθηκε και χρησιμοποιήθηκε το διαδικτυακό εργαλείο WebGestalt²² [82]. Χρησιμοποιήθηκε η μέθοδος ανάλυσης εμπλουτισμού συνόλου γονιδίων (Gene Set Enrichment Analysis-GSEA) [129] που επιτρέπει την ενσωμάτωση ποσοτικοποιημένης μεταβλητής για υπερ- και υπο-έκφραση. Η μέθοδος εφαρμόστηκε μόνο για τις σειρές δεδομένων που περιλαμβάνουν δείγματα εκτεθειμένα σε σωματιδιακή αλλά και ηλεκτρομαγνητική ακτινοβολία (Πίνακας 3). Εκτελέστηκαν 7 αναλύσεις που αντιπροσώπευαν τις 7 λίστες διαφορικών εκφρασμένων γονιδίων των σειρών αυτών (Πίνακας 9, Πίνακας 10, Πίνακας 11). Μέσω ενσωμάτωσης της μετρικής του \log_2FC για κάθε γονίδιο σε κάθε λίστα, αυξάνεται η πιστότητα των αποτελεσμάτων της ανάλυσης εμπλουτισμού. Σε αυτό το στάδιο επιλέχθηκαν ως αναφορά όροι γονιδιακής οντολογίας [84] στην κατηγορία των βιολογικών διεργασιών και όροι βιολογικών μονοπατιών από τις βάσεις δεδομένων KEGG [130, 131], Reactome [132], Panther [133] και WikiPathways [134].

²² <http://www.webgestalt.org/option.php>

Πίνακας 3: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με την μέθοδο GSEA στο WebGestalt.

Series Accession	Cell Type	Radiation Type	Dose [Gy]	Time [hrs]
GSE44282	Bronchial Epithelial Cells	X-rays Fe Ions	1	24
E-MTAB-5754	Coronary Artery Endothelial Cells	X-rays Fe Ions	2	24
E-MTAB-5761 3463	Peripheral Blood Mononuclear Cells	X-rays Fe Ions Carbon Ions	1	8

2.4.2. Ομαδοποίηση Δεδομένων

Προκειμένου να λάβουμε μία συνολικότερη εικόνα της απόκρισης στην ακτινοβολία και να εκμεταλλευτούμε όλα τα δεδομένα μικροσυστοιχιών που αναλύθηκαν, χρησιμοποιήσαμε την μέθοδο ανάλυσης υπερεκπροσώπησης (Over-Representation Analysis-ORA) [135], που παρέχει το WebGestalt. Η μέθοδος ORA εκτελεί μία στατιστική αξιολόγηση του κλάσματος των γονιδίων σε συγκεκριμένο όρο (βιολογική διαδικασία, μονοπάτι κ.λπ.) από το σύνολο των γονιδίων που παρουσιάζουν αλλαγές στην έκφραση. Η βασική διαφορά με την μέθοδο GSEA είναι ότι αξιολογεί όλα τα γονίδια με τον ίδιο τρόπο, καθώς δεν υπάρχει δυνατότητα ταξινόμησης μέσω κατάλληλης μετρικής και συνεπώς είναι πιθανό να χαθεί σημαντικό ποσοστό πληροφορίας [136].

Χρησιμοποιήθηκαν οι λίστες κοινών γονιδίων που προέκυψαν κατά την ομαδοποίηση των αρχικών λιστών από την Μπεϋζιανή ανάλυση, μέσω σχεδιασμού διαγραμμάτων Venn²³ (Εικόνα 50, Εικόνα 51, Εικόνα 52, Εικόνα 53, Εικόνα 54, Εικόνα 55, Εικόνα 56, Εικόνα 57). Σκοπός ήταν οι ομάδες αυτές να επιδεικνύουν όσο το δυνατό πιο ομοιόμορφα χαρακτηριστικά και να διαφοροποιούνται σε μία από τις παραπάνω παραμέτρους, προς την κατεύθυνση της απάντησης ενός συγκεκριμένου βιολογικού ερωτήματος.

²³ <http://bioinformatics.psb.ugent.be/webtools/Venn/>

Για την παράμετρο του τύπου ακτινοβολίας, το κύριο χαρακτηριστικό που λαμβάνεται υπόψιν είναι η τιμή της Γραμμικής Εναπόθεσης Ενέργειας (LET) [39], γεγονός που υποδηλώνει ότι η ηλεκτρομαγνητική ακτινοβολία χαμηλής LET και η σωματιδιακή ακτινοβολία υψηλής LET μπορούν να θεωρηθούν ως οι δύο κύριες συνιστώσες που προκαλούν τις μέγιστες διαφοροποιήσεις στη βιολογική απόκριση. Στη συνέχεια, όσον αφορά στο χρόνο μετά την ακτινοβολήση, είναι αποδεκτό να μελετάται το προφίλ έκφρασης σε ομάδες ωρών, καθώς έχουν παρατηρηθεί διαφορές μεταξύ πρώιμων (περίπου 1-6) και μεταγενέστερων (24) ωρών μετά την έκθεση [22, 137, 138]. Τέλος, σχετικά με τη δόση της ακτινοβολίας, με την ίδια λογική μπορούν να σχηματιστούν μεμονωμένες συγκρίσεις βασισμένες σε μία μόνο τιμή ή σε ομάδες δόσεων που βασίζονται στο κατώφλι των 0.5Gy (τυπικά 0.01Gy) για χαμηλές και υψηλές δόσεις ακτινοβολίας αντίστοιχα [15, 139, 140]. Σημειώνουμε ότι ο τύπος του ιστού πιθανά να καθορίζει το προφίλ διαφορικής έκφρασης λόγω της προϋπάρχουσας κυτταρικής διαφοροποίησης, συνεπώς συγκρίσεις μεταξύ ιστών οφείλουν να γίνονται με προσοχή.

Πίνακας 4: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με την μέθοδο ORA στο WebGestalt.

Series Accession	Cell Type	Radiation Type	Dose	Time [hrs]
E-MTAB-5754 3463	Coronary Artery Endothelial Cells Peripheral Blood Mononuclear Cells	X-rays	High	8-24
GSE81067	Pulmonary Endothelial Cells (HPEAC) Lung Fibroblasts (HFL-1)	Alpha	High	24
GSE63952 64375	Isolated Leucocytes	HZE	Low High	6
GSE44282	Bronchial Epithelial Cells (HBEC3KT)	Fe Ions	Low High	24
E-MTAB-5754 5761	Coronary Artery Endothelial Cells Peripheral Blood Mononuclear Cells		Low	1-4 6-24
GSE63952 64375	Isolated Leucocytes	HZE		1-4
GSE44282	Bronchial Epithelial Cells (HBEC3KT)		High	6-24

Με την εφαρμογή των παραπάνω καταλήξαμε σε 10 αναλύσεις (Πίνακας 4) με την μέθοδο ORA, χρησιμοποιώντας ως αναφορά τους όρους που χρησιμοποιήθηκαν και παραπάνω ενώ προστέθηκαν οι όροι γονιδιακής οντολογίας στις κατηγορίες μοριακής λειτουργίας και κυτταρικών τμημάτων (Πίνακας 13, Πίνακας 14, Πίνακας 15, Πίνακας 16, Πίνακας 17). Σημειώνεται ότι το επίπεδο στατιστικής σημαντικότητας για τους εμπλουτισμένους όρους, και τις δύο περιπτώσεις που χρησιμοποιήθηκε το WebGestalt, ορίστηκε βάση του υπολογιζόμενου προσαρμοσμένου p-value με την μέθοδο FDR [76].

2.4.3. Ανάλυση Εμπλουτισμού για Ασθένειες & Παθήσεις

Αρχικά εκτελέστηκε μια ανάλυση εμπλουτισμού για όρους ασθενειών και παθήσεων με το εργαλείο Enrichr²⁴ [83, 141]. Ο λόγος για τον οποίο επιλέχθηκε το εργαλείο αυτό ήταν αρχικά γιατί εμφανίζει μεγαλύτερη ποικιλία ως προς τις βάσεις δεδομένων με τις οποίες συνδέεται γενικότερα αλλά και συγκεκριμένα ως προς εκείνες που αφορούν σε ασθένειες και παθήσεις, συγκριτικά με το WebGestalt. Το Enrichr χρησιμοποιεί την μέθοδο του Fisher's exact test [142], η οποία ενσωματώνεται στα περισσότερα εργαλεία εμπλουτισμού, εκτελώντας επιπλέον διορθώσεις βασιζόμενες στην τυπική απόκλιση προς τον υπολογισμό ενός τελικού σκορ [141]. Καθώς καθένα εργαλείο εμφανίζει διαφορές ως προς τις μεθόδους εμπλουτισμού, επιδιώκουμε να συγκρίνουμε αυτά τα δύο εργαλεία ως προς την πιστότητα των μεθόδων ανάλυσης που χρησιμοποιούν. Επιπλέον, θεμιτή είναι και μια συνολικότερη ανασκόπηση για την παρουσίαση των αποτελεσμάτων, το πόσο φιλικά είναι προς τον χρήστη και σε ποιες περιπτώσεις ενδείκνυται η χρήση του καθενός.

Εκτελέστηκαν 10 αναλύσεις (Πίνακας 5) με βάση τις ομαδοποιημένες λίστες γονιδίων όπως προέκυψαν παραπάνω, καθώς το εργαλείο δεν δίνει δυνατότητα ενσωμάτωσης πληροφορίας ποσοτικοποίησης. Ως αναφορά για την ανάλυση εμπλουτισμού επιλέχθηκαν όροι ασθενειών από το δημόσιο

²⁴ <https://amp.pharm.mssm.edu/Enrichr/#>

αρχείο της ClinVar [143-145] του NCBI και από την πλατφόρμα του DisGeNET [87, 146, 147]. Οι πίνακες των αποτελεσμάτων της ανάλυσης εμπλουτισμού με χρήση του Enrichr παρατίθενται στο κεφάλαιο των αποτελεσμάτων (Πίνακας 18, Πίνακας 19, Πίνακας 20, Πίνακας 21, Πίνακας 22, Πίνακας 23).

Πίνακας 5: Πειραματικές πληροφορίες για τις λίστες γονιδίων που αναλύθηκαν με το Enrichr ως προς όρους ασθενειών και παθήσεων.

Series Accession	Cell Type	Radiation Type	Dose	Time [hrs]
E-MTAB-5754 3463	Coronary Artery Endothelial Cells Peripheral Blood Mononuclear Cells	X-rays	High	8-24
GSE81067	Pulmonary Endothelial Cells (HPEAC) Lung Fibroblasts (HFL-1)	Alpha	High	24
GSE63952 64375	Isolated Leucocytes	HZE	Low High	6
GSE44282	Bronchial Epithelial Cells (HBEC3KT)	X-rays Fe, Si Ions	High	12-24
E-MTAB-5761 3463	Peripheral Blood Mononuclear Cells	X-rays Fe Ions	High	8
E-MTAB-5754	Coronary Artery Endothelial Cells	X-rays Fe Ions	High	24

2.4.4. Αναζήτηση Σημαντικών Γονιδίων της Απόκρισης σε IR

Στη συνέχεια θεωρήθηκε χρήσιμο να διενεργηθεί μία βιβλιογραφική αναζήτηση ώστε να προσδιοριστεί μία ομάδα κρίσιμων γονιδίων, η οποία θα αντιπροσωπεύει τις βασικές βιολογικές διεργασίες του κυττάρου που σχετίζονται με την επίδραση της ιοντίζουσας ακτινοβολίας ώστε να μελετηθεί η συμπεριφορά των γονιδίων αυτών στα δεδομένα της παρούσας εργασίας.

Αρχικά έγινε αναζήτηση για όρους γονιδιακής οντολογίας μέσω της εφαρμογής AmiGO [148] στην επίσημη ιστοσελίδα²⁵, σχετικά με την επιδιόρθωση βλαβών DNA και την ανοσολογική απόκριση. Κατά αυτό τον τρόπο καταλήξαμε στους δύο αντιπροσωπευτικούς κωδικούς, που στο δέντρο της γονιδιακής οντολογίας για την επιδιόρθωση βλαβών DNA και την απόκριση του ανοσοποιητικού συστήματος αντίστοιχα τοποθετούνται ως:

biological process > cellular process > cellular response to stimulus > cellular response to stress > cellular response to DNA damage stimulus > DNA repair (GO:0006281) – 2352 Genes

biological process > immune system process > immune response (GO:0006955) – 581 Genes

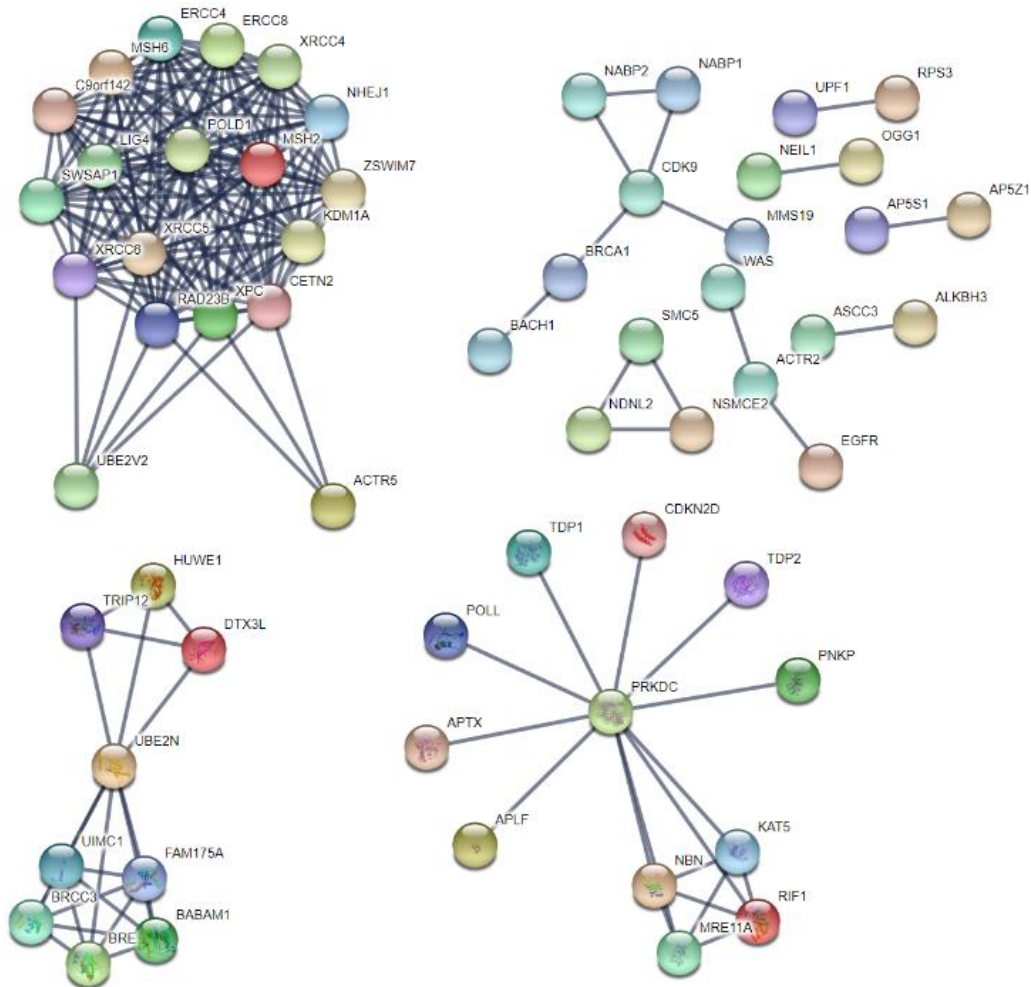
Στη συνέχεια χρησιμοποιήθηκαν αυτές οι δύο λίστες ανθρώπινων γονιδίων, ώστε να σχεδιαστούν δύο δίκτυα πρωτεϊνών με το εργαλείο STRING v11²⁶[85]. Από την επιλογή αναζήτησης για πολλαπλές πρωτεΐνες, με οργανισμό αναφοράς τον άνθρωπο κατασκευάστηκαν τα δύο δίκτυα. Οι ακμές των δικτύων σχεδιάστηκαν βάση του μεγίστου σκορ αλληλεπίδρασης ενώ η πληροφορία για τις αλληλεπιδράσεις αυτές προήλθε από βάσεις δεδομένων αναφοράς και πειράματα. Καθώς και τα δύο δίκτυα ήταν αρκετά μεγάλα για να καθοριστούν τα κρίσιμα γονίδια με τον βέλτιστο τρόπο εφαρμόστηκε η μέθοδος k-means clustering, η οποία παρέχεται από το STRING. Πρόκειται για έναν αλγόριθμο μηχανικής μάθησης που εκπαιδεύεται σε ένα σύνολο δεδομένων και στη συνέχεια το ταξινομεί σε ένα νέο σύνολο δεδομένων χρησιμοποιώντας αριθμό k προκαθορισμένων συστάδων (*clusters*) [149, 150]. Ακολουθούν οι συστάδες που δημιουργήθηκαν και σχεδιάστηκαν ξεχωριστά καθώς και τα αντιπροσωπευτικά γονίδια που επιλέχθηκαν σε κάθε περίπτωση.

²⁵ <http://geneontology.org/>

²⁶ <https://string-db.org/cgi/input.pl>

Επιδιόρθωση Βλαβών DNA

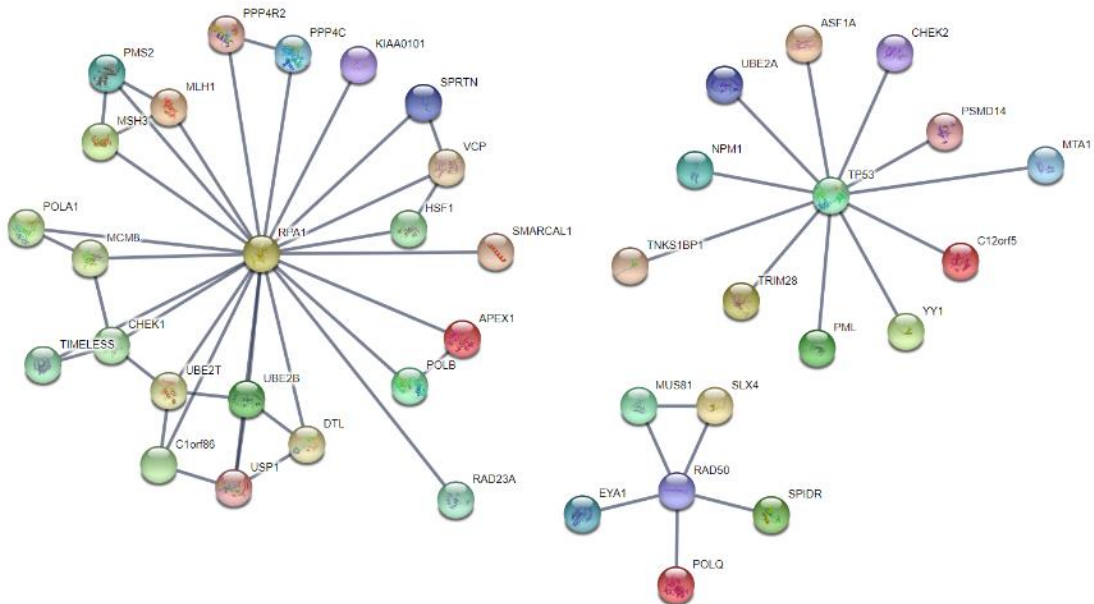
Στα πλαίσια της διεργασίας επιδιόρθωσης βλαβών DNA, ο βέλτιστος αριθμός συστάδων κατά την εφαρμογή της μεθόδου ήταν πέντε ($k=5$). Η πρώτη συστάδα όντας η μεγαλύτερη αναλύθηκε ξεχωριστά για $k=4$ (Εικόνα 41), όπου η πρώτη συστάδα περιλαμβάνει γονίδια χαρακτηριστικών μονοπατιών επιδιόρθωσης από την οποία επιλέχθηκαν και τα περισσότερα γονίδια (Πίνακας 6). Από τις υπόλοιπες συστάδες (Εικόνα 42) οι τρεις εκ των τεσσάρων αναδεικνύουν με ξεκάθαρο τρόπο ακόμα 3 αντιπροσωπευτικά γονίδια (Πίνακας 7) του δικτύου πρωτεϊνών επιδιόρθωσης DNA.



Εικόνα 41: Απεικόνιση των συστάδων που προέκυψαν με εφαρμογή της μεθόδου k -means clustering για $k=4$, στην μεγαλύτερη αρχική συστάδα του δικτύου πρωτεϊνών επιδιόρθωσης DNA, όπως προέκυψε από την πρώτη εφαρμογή της μεθόδου για $k=5$.

Πίνακας 6: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά κάθε συστάδας κατά την ανάλυση της μεγαλύτερης συστάδας του δικτύου πρωτεϊνών επιδιόρθωσης DNA.

Cluster	Genes
1	ERCC1, ERCC3, ERCC8, FAN1, LIG4, MSH2, MSH6, NHEJ1, PCNA, POLD3, POLE, POLK, RPA2, RPA3, SWSAP1, XPC
2	CDK9, BRCA1
3	UBE2N
4	PRKDC



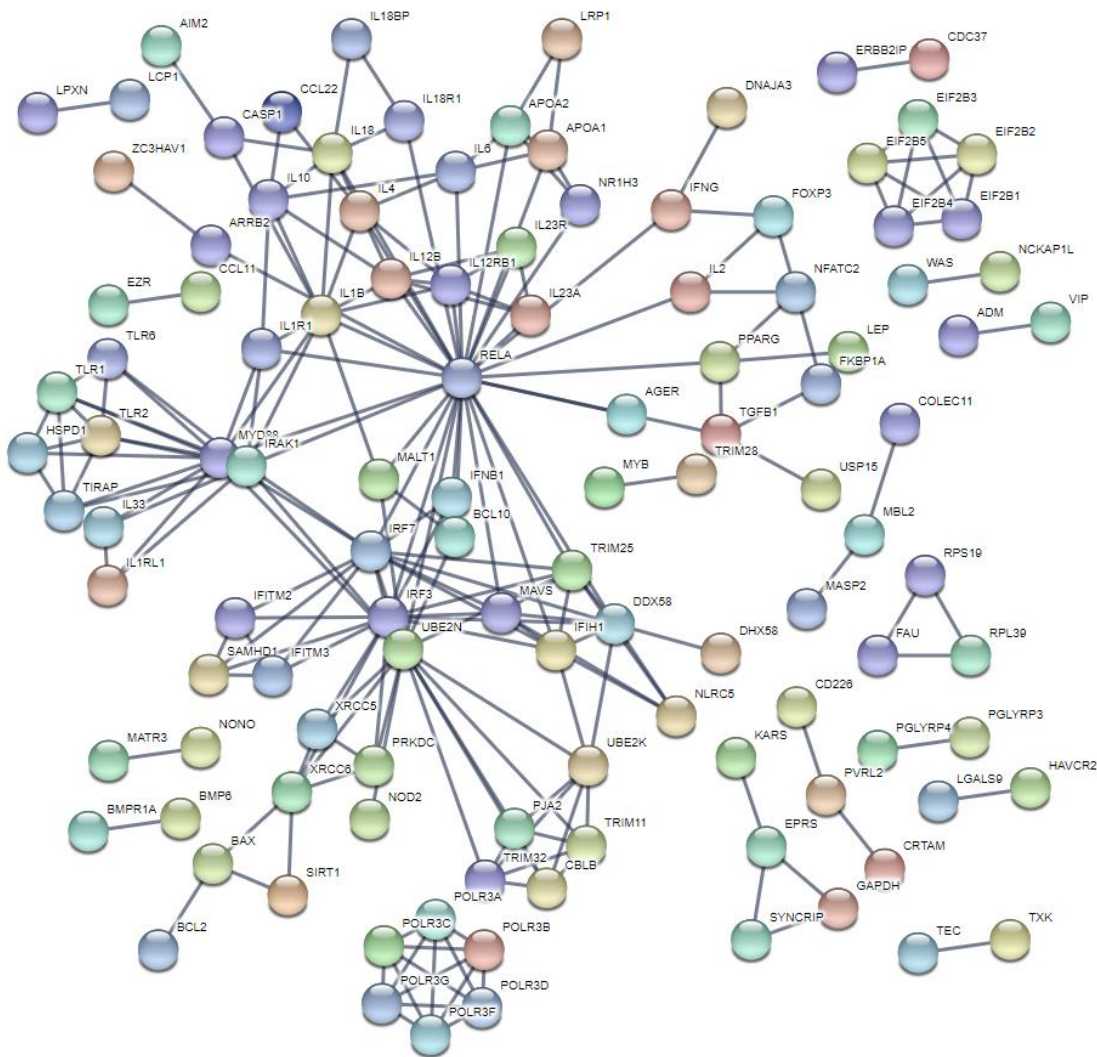
Εικόνα 42: Απεικόνιση των συστάδων 2-4 που προέκυψαν με εφαρμογή της μεθόδου *k-means clustering* για $k=5$ στο δίκτυο πρωτεϊνών επιδιόρθωσης DNA.

Πίνακας 7: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά των συστάδων 2-4 του δικτύου πρωτεϊνών επιδιόρθωσης DNA.

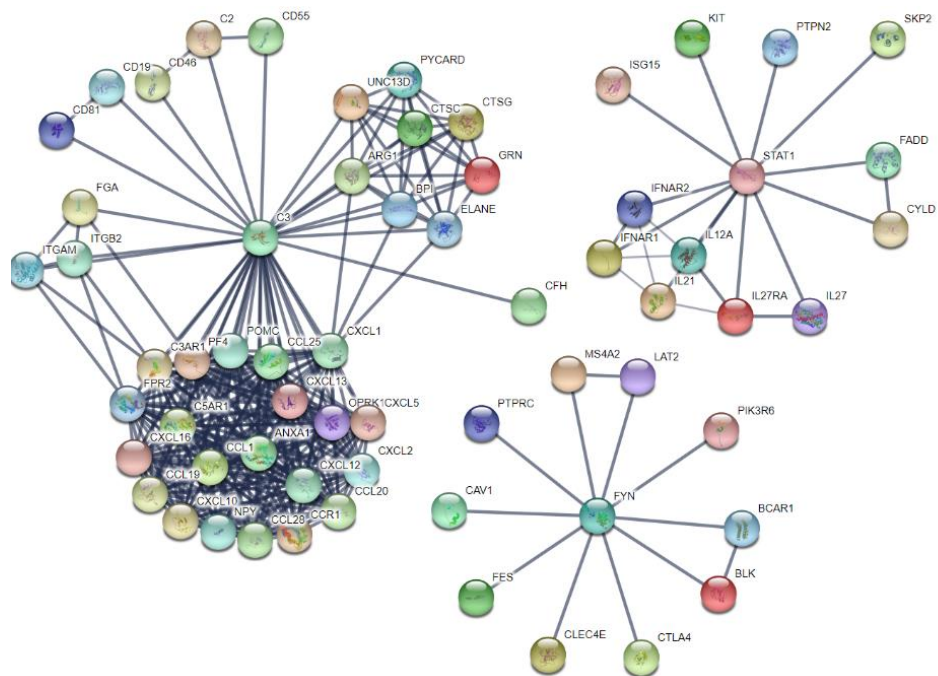
Cluster	Genes
1	RPA1
2	TP53
3	RAD50

Απόκριση Ανοσοποιητικού

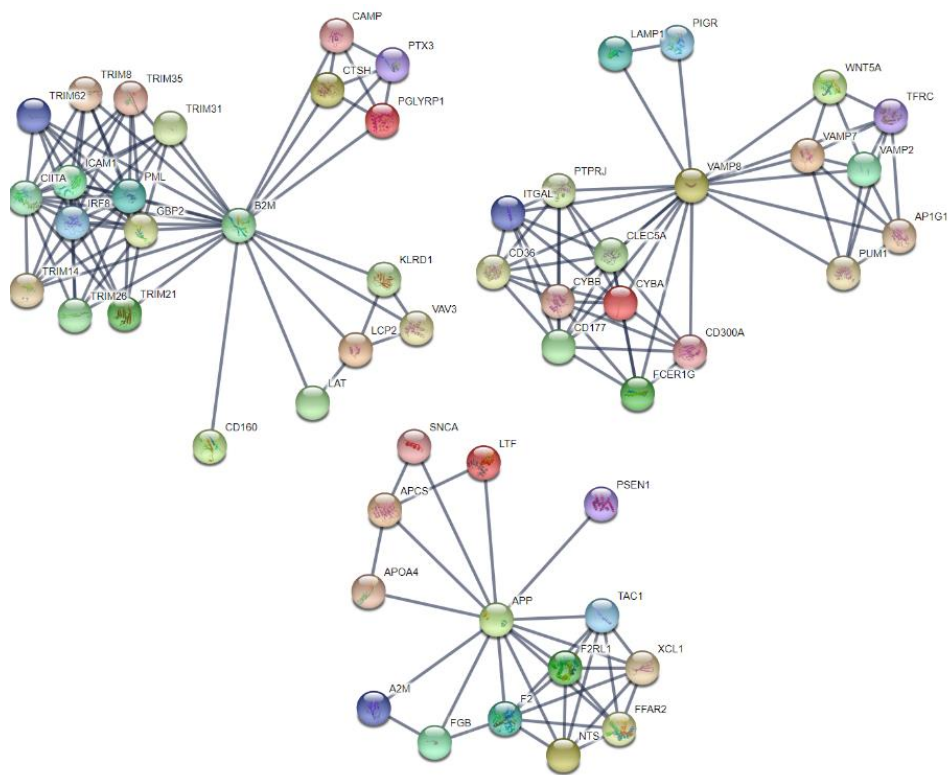
Στα πλαίσια της διεργασίας επιδιόρθωσης βλαβών DNA ο βέλτιστος αριθμός συστάδων ήταν επτά ($k=7$) κατά την εφαρμογή της μεθόδου. Από την πρώτη συστάδα (Εικόνα 43) επιλέχθηκαν 4 ενδεικτικά γονίδια, ενώ από τις υπόλοιπες συστάδες (Εικόνα 44, Εικόνα 45) αναδεικνύονται με ξεκάθαρο τρόπο ακόμα 6 αντιπροσωπευτικά γονίδια (Πίνακας 8).



Εικόνα 43: Απεικόνιση της 1^{ης} συστάδας που προέκυψε με εφαρμογή της μεθόδου k -means clustering για $k=7$ στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.



Εικόνα 44: Απεικόνιση των συστάδων 2-4 που προέκυψαν με εφαρμογή της μεθόδου *k*-means clustering για $k=7$ στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.

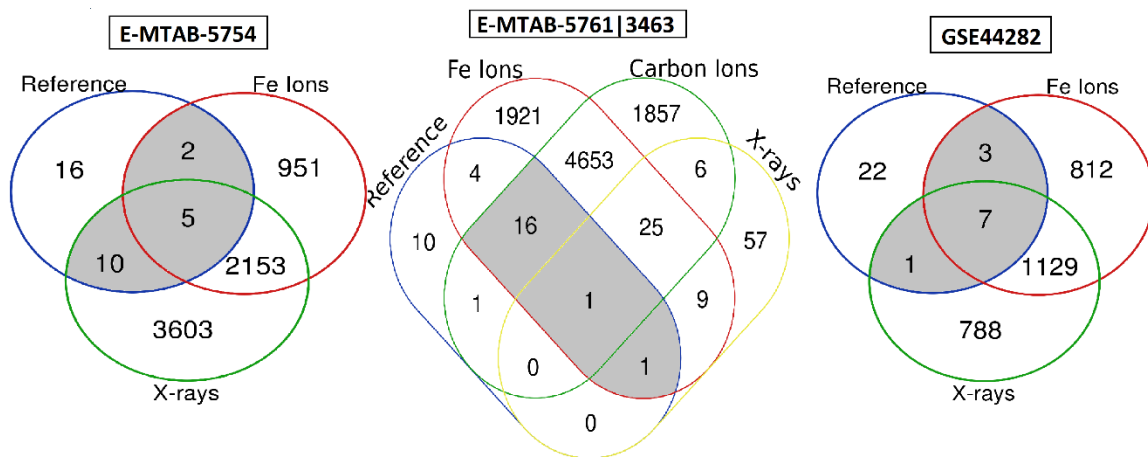


Εικόνα 45: Απεικόνιση των συστάδων 5-7 που προέκυψαν με εφαρμογή της μεθόδου *k*-means clustering για $k=7$ στο δίκτυο πρωτεϊνών του ανοσοποιητικού συστήματος.

Πίνακας 8: Πίνακας γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά των συστάδων 1-7 του δικτύου πρωτεϊνών του ανοσοποιητικού συστήματος.

Cluster	Genes
1	RELA, IRF3, MUD88, IRAK1
2	C3
3	STAT1
4	FYN
5	B2M
6	VAMP8
7	APP

Το σύνολο των γονιδίων που επιλέχθηκαν ως αντιπροσωπευτικά των βιολογικών διεργασιών της επιδιόρθωσης DNA και του ανοσοποιητικού συστήματος που φαίνονται στο Παράρτημα (Πίνακας 26), μελετήθηκαν ως προς την έκφραση. Επικεντρωθήκαμε στα αποτελέσματα των σειρών που περιείχαν δείγματα ακτινοβολημένα με ιοντίζουσα υψηλής και χαμηλής τιμής LET. Τα κοινά με την βιβλιογραφική αναζήτηση γονίδια κάθε σειράς (Εικόνα 46) αναπαραστάθηκαν με διαγράμματα στηλών έκφρασης. Η έκφραση ποσοτικοποιήθηκε βάση της τιμής logFC.



Εικόνα 46: Διαγράμματα Venn των διαφορικά εκφρασμένων γονιδίων από τις τρεις σειρές δεδομένων που περιείχαν δείγματα ακτινοβολημένα με ιοντίζουσα υψηλής και χαμηλής τιμής LET, σε συνδυασμό με τα «σημαντικά» γονίδια που επιλέχθηκαν από την βιβλιογραφική αναζήτηση. Οι γκριζές περιοχές υποδεικνύουν τα γονίδια που αναπαραστάθηκαν με γραφικές παραστάσεις για κάθε σειρά δεδομένων (Πίνακας 12).

Για τον σχεδιασμό των διαγραμμάτων (Εικόνα 47, Εικόνα 48, Εικόνα 49) χρησιμοποιήθηκε το πρόγραμμα OriginPro8 [151, 152]. τα διαγράμματα παρατίθενται στο κεφάλαιο των αποτελεσμάτων.

2.4.5. Συγγενείς Παθήσεις

Στα πλαίσια των συγγενών παθήσεων που έχουν συσχετιστεί με την έκθεση σε ιοντίζουσα ακτινοβολία και ειδικότερα με ακτινοβολίες υψηλής τιμής LET επιλέχθηκαν 4 βασικές κατηγορίες παθήσεων όπως αναφέρθηκαν στην εισαγωγή:

1. Παθήσεις του ανοσοποιητικού & αυτοάνοσα νοσήματα
2. Παθήσεις του μεταβολισμού
3. Καρδιαγγειακές παθήσεις
4. Νευροεκφυλιστικές παθήσεις

Για αυτές τις κατηγορίες εκτελέστηκε μια εκτενής αναζήτηση στην πλατφόρμα του DisGeNET, η οποία περιέχει συσχετίσεις γονιδίων-ασθενειών, ώστε να συγκεντρωθούν οι όροι ασθενειών και παθήσεων για κάθε κατηγορία. Οι όροι και οι κωδικοί των ασθενειών καθώς και τα γονίδια που αντιστοιχούν σε κάθε κατηγορία παρατίθενται στο Παράρτημα (Πίνακας 27, Πίνακας 28). Στη συνέχεια σχεδιάστηκε ένα συγκεντρωτικό διάγραμμα Venn (Εικόνα 58) των παραπάνω γονιδίων σε συνδυασμό με το σύνολο των διαφορετικώς εκφρασμένων γονιδίων έπειτα από έκθεση σε ακτινοβολία βαρέων ιόντων (HZE). Τέλος χρησιμοποιήθηκε το εργαλείο DISNOR [86] για την κατασκευή δικτύων αλληλεπιδράσεων (Εικόνα 59, Εικόνα 60, Εικόνα 61, Εικόνα 62), τα οποία ορίζονται ως το σύνολο των αιτιωδών αλληλεπιδράσεων μεταξύ γονιδίων που έχουν παρατηρηθεί να μεταλλάσσονται στην ίδια παθολογία. Χρησιμοποιήθηκαν τα γονίδια που προέκυψαν από την ανάλυση των δεδομένων μικροσυστοιχιών έπειτα από έκθεση σε HZE, τα οποία βρέθηκαν κοινά με κάθε μία κατηγορία συγγενών παθήσεων (Πίνακας 24).

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1. Μελέτη Εντός των Σειρών Δεδομένων

3.1.1. Λειτουργική Ανάλυση Εμπλουτισμού

Παρατίθενται τα αποτελέσματα των λειτουργικών αναλύσεων εμπλουτισμού από το WebGestalt με την μέθοδο GSEA. Η μέθοδος αυτή χρησιμοποιήθηκε μόνο για τις σειρές που περιλάμβαναν δείγματα ακτινοβολημένα με σωματιδιακή και ηλεκτρομαγνητική ακτινοβολία (υψηλής και χαμηλής LET αντίστοιχα) χάριν σύγκρισης των αποκρίσεων σε κάθε περίπτωση:

Πίνακας 9: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικώς εκφρασμένων γονιδίων της σειράς GSE44282, που αποτελείται από δείγματα ανθρωπίνων βρογχικών επιθηλιακών κυττάρων, ακτινοβολημένα με 1Gy ακτίνων-X και ιόντων σιδήρου αντίστοιχα που συλλέχθηκαν 24 ώρες μετά την ακτινοβόληση.

WebGestalt - GSEA			
GSE44282			
Bronchial Epithelial Cells X-rays 1Gy 24hrs			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0002526	acute inflammatory response
		GO:1990868	response to chemokine
		GO:0071496	cellular response to external stimulus
		GO:0070555	response to interleukin-1
		GO:0060326	cell chemotaxis
Pathway	KEGG	hsa04064	NF-kappa B signaling pathway
		hsa04657	IL-17 signaling pathway
		hsa04668	TNF signaling pathway
		hsa04060	Cytokine-cytokine receptor interaction
	Panther	P06959	CCKR signaling map
P00006		Apoptosis signaling pathway	
Bronchial Epithelial Cells Fe Ions 1Gy 24hrs			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0032103	positive regulation of response to external stimulus
		GO:0071216	cellular response to biotic stimulus
		GO:0060326	cell chemotaxis
		GO:0097696	STAT cascade
		GO:0002526	acute inflammatory response
		GO:0006260	DNA replication

		GO:0070371	ERK1 and ERK2 cascade
		GO:0006959	humoral immune response
		GO:0006284	base-excision repair
Pathway	KEGG	hsa04630	JAK-STAT signaling pathway
		hsa04657	IL-17 signaling pathway
		hsa04668	TNF signaling pathway
		hsa04060	Cytokine-cytokine receptor interaction
		hsa04621	NOD-like receptor signaling pathway
		hsa04064	NF-kappa B signaling pathway
		hsa04110	Cell cycle
		hsa03030	DNA replication
		hsa00240	Pyrimidine metabolism
		hsa03410	Base excision repair
	Panther	P06959	CCKR signaling map
		P00031	Inflammation mediated by chemokine and cytokine signaling pathway
		P00006	Apoptosis signaling pathway
		P00059	p53 pathway
		P00052	TGF-beta signaling pathway
		P00053	T cell activation
	Wiki Pathways	R-HSA-1280215	Cytokine Signaling in Immune system
		R-HSA-6783783	Interleukin-10 signaling
		R-HSA-449147	Signaling by Interleukins
		R-HSA-453279	Mitotic G1-G1/S phases
		R-HSA-69206	G1/S Transition
		R-HSA-168256	Immune System
		R-HSA-2555396	Mitotic Metaphase and Anaphase

Πίνακας 10: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικών εκφρασμένων γονιδίων της σειράς E-MTAB-5754, που αποτελείται από δείγματα ανθρωπίνων ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα με 2Gy ακτίνων-X και ιόντων σιδήρου αντίστοιχα που συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.

WebGestalt - GSEA

E-MTAB-5754

Coronary Artery Endothelial Cells | X-rays | 2Gy | 24 hrs

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0006310	DNA recombination
		GO:0006302	double-strand break repair
		GO:0006333	chromatin assembly or disassembly
		GO:0000075	cell cycle checkpoint
		GO:0006260	DNA replication
		GO:0007059	chromosome segregation
Pathway	KEGG	hsa03440	Homologous recombination
		hsa04115	p53 signaling pathway

	hsa04218	Cellular senescence
	hsa03030	DNA replication
	hsa05322	Systemic lupus erythematosus
	hsa04110	Cell cycle
Panther	P00011	Blood coagulation
	P04398	p53 pathway feedback loops 2
	P00034	Integrin signalling pathway
	P00017	DNA replication
Reactome	R-HSA-69242	S Phase
	R-HSA-69206	G1/S Transition
	R-HSA-453279	Mitotic G1-G1/S phases
	R-HSA-194315	Signaling by Rho GTPases
	R-HSA-195258	RHO GTPase Effectors
	R-HSA-1640170	Cell Cycle
	R-HSA-69278	Cell Cycle, Mitotic
Wiki Pathways	WP3959	DNA IR-Double Strand Breaks (DSBs) and cellular response via ATM
	WP2516	ATM Signaling Pathway
	WP707	DNA Damage Response
	WP1530	miRNA Regulation of DNA Damage Response
	WP45	G1 to S cell cycle control
	WP4016	DNA IR-damage and cellular response via ATR
	WP2361	Gastric Cancer Network 1
	WP466	DNA Replication
	WP179	Cell Cycle

Coronary Artery Endothelial Cells | Fe Ions | 2Gy | 24 hrs

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0007059	chromosome segregation
		GO:0010948	negative regulation of cell cycle process
		GO:0000075	cell cycle checkpoint
		GO:0034341	response to interferon-gamma
		GO:0032612	interleukin-1 production
		GO:0008380	RNA splicing
Pathway	KEGG	hsa04110	Cell cycle
		hsa04115	p53 signaling pathway
		hsa04218	Cellular senescence
		hsa04510	Focal adhesion
		hsa04020	Calcium signaling pathway
		hsa05414	Dilated cardiomyopathy (DCM)
		hsa04620	Toll-like receptor signaling pathway
		hsa04621	NOD-like receptor signaling pathway
Pathway	Panther	P00034	Integrin signalling pathway
		P00059	p53 pathway
		P00057	Wnt signaling pathway

	P00045	Notch signaling pathway
	P00011	Blood coagulation
	P00048	PI3 kinase pathway
	P00006	Apoptosis signaling pathway
	P00053	T cell activation
	P00046	Oxidative stress response
Reactome	R-HSA-69278	Cell Cycle, Mitotic
	R-HSA-1640170	Cell Cycle
	R-HSA-68886	M Phase
	R-HSA-195258	RHO GTPase Effectors
	R-HSA-194315	Signaling by Rho GTPases
	R-HSA-453279	Mitotic G1-G1/S phases
	R-HSA-72163	mRNA Splicing - Major Pathway
Wiki Pathways	WP2446	Retinoblastoma Gene in Cancer
	WP179	Cell Cycle
	WP4016	DNA IR-damage and cellular response via ATR
	WP1530	miRNA Regulation of DNA Damage Response
	WP2361	Gastric Cancer Network 1
	WP707	DNA Damage Response
	WP4240	Regulation of sister chromatid separation at the metaphase-anaphase transition
	WP466	DNA Replication
	WP306	Focal Adhesion
	WP2516	ATM Signaling Pathway
	WP254	Apoptosis
	WP75	Toll-like Receptor Signaling Pathway

Πίνακας 11: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο GSEA. Χρησιμοποιήθηκαν οι λίστες διαφορικών εκφρασμένων γονιδίων των σειρών E-MTAB-5761/3463, που αποτελείται από δείγματα ανθρώπινων κυττάρων περιφερειακού αίματος, ακτινοβολημένα με 1Gy ιόντων σιδήρου, ιόντων άνθρακα και ακτίνων-X αντίστοιχα που συλλέχθηκαν 8 ώρες μετά την ακτινοβόληση.

WebGestalt - GSEA

E-MTAB-5761|3463

Peripheral Blood Mononuclear Cells | Fe Ions | 1Gy | 8hrs

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0071216	cellular response to biotic stimulus
		GO:0009593	detection of chemical stimulus
		GO:0010573	vascular endothelial growth factor production
		GO:0070555	response to interleukin-1
		GO:0006968	cellular defense response
		GO:0002532	production of molecular mediator involved in inflammatory response
		GO:0002576	platelet degranulation
		GO:0002446	neutrophil mediated immunity

Pathway	KEGG	hsa04060 hsa04668 hsa05321 hsa04657 hsa05322 hsa04611 hsa04650 hsa04612	Cytokine-cytokine receptor interaction TNF signaling pathway Inflammatory bowel disease (IBD) IL-17 signaling pathway Systemic lupus erythematosus Platelet activation Natural killer cell mediated cytotoxicity Antigen processing and presentation
	Panther	P00029 P04391 P04377 P04378 P04380	Huntington disease Oxytocin receptor mediated signaling pathway Beta1 adrenergic receptor signaling pathway Beta2 adrenergic receptor signaling pathway Corticotropin releasing factor receptor signaling pathway
	Reactome	R-HSA-420499 R-HSA-6783783 R-HSA-6785807 R-HSA-5693579 R-HSA-5693616 R-HSA-380108 R-HSA-114608 R-HSA-5625740 R-HSA-76005 R-HSA-5334118 R-HSA-427359 R-HSA-168249 R-HSA-977225 R-HSA-6798695	Class C/3 (Metabotropic glutamate/pheromone receptors) Interleukin-10 signaling Interleukin-4 and Interleukin-13 signaling Homologous DNA Pairing and Strand Exchange Presynaptic phase of homologous DNA pairing and strand exchange Chemokine receptors bind chemokines Platelet degranulation RHO GTPases activate PKNs Response to elevated platelet cytosolic Ca ²⁺ DNA methylation SIRT1 negatively regulates rRNA expression Innate Immune System Amyloid fiber formation Neutrophil degranulation

Peripheral Blood Mononuclear Cells | Carbon Ions | 1Gy | 8hrs

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0071216	cellular response to biotic stimulus
		GO:0010573	vascular endothelial growth factor production
		GO:0009593	detection of chemical stimulus
		GO:0002576	platelet degranulation
		GO:0032418	lysosome localization
		GO:0002446	neutrophil mediated immunity
Pathway	KEGG	hsa04060	Cytokine-cytokine receptor interaction
		hsa04668	TNF signaling pathway
		hsa04657	IL-17 signaling pathway
		hsa05323	Rheumatoid arthritis
		hsa04612	Antigen processing and presentation
	Panther	P00016	Cytoskeletal regulation by Rho GTPase
		P04379	Beta3 adrenergic receptor signaling pathway

	P04376	5HT4 type receptor mediated signaling pathway
	P04378	Beta2 adrenergic receptor signaling pathway
	P04380	Corticotropin releasing factor receptor signaling pathway
Reactome	R-HSA-6783783	Interleukin-10 signaling
	R-HSA-420499	Class C/3 (Metabotropic glutamate/pheromone receptors)
	R-HSA-6785807	Interleukin-4 and Interleukin-13 signaling
	R-HSA-159236	Transport of Mature mRNA derived from an Intron-Containing Transcript
	R-HSA-1650814	Collagen biosynthesis and modifying enzymes
	R-HSA-76002	Platelet activation, signaling and aggregation
	R-HSA-977225	Amyloid fiber formation
	R-HSA-168249	Innate Immune System
	R-HSA-6798695	Neutrophil degranulation
Wiki Pathways	WP530	Cytokines and Inflammatory Response
	WP3617	Photodynamic therapy-induced NF-kB survival signaling
	WP3624	Lung fibrosis
	WP2849	Hematopoietic Stem Cell Differentiation
	WP615	Senescence and Autophagy in Cancer
	WP4329	miRNAs involvement in the immune response in sepsis
	WP1533	Vitamin B12 Metabolism
	WP2855	Dopaminergic Neurogenesis
	WP408	Oxidative Stress
	WP4211	Transcriptional cascade regulating adipogenesis
	WP3937	Microglia Pathogen Phagocytosis Pathway

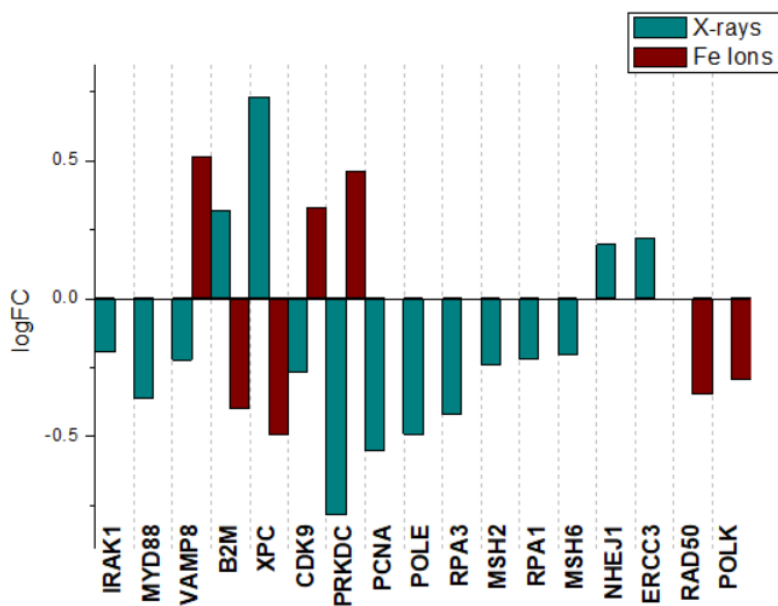
Peripheral Blood Mononuclear Cells | X-rays | 1Gy | 8hrs

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0072331	signal transduction by p53 class mediator
		GO:0044772	mitotic cell cycle phase transition
		GO:2001020	regulation of response to DNA damage stimulus
		GO:0010948	negative regulation of cell cycle process
		GO:0000075	cell cycle checkpoint
		GO:0045930	negative regulation of mitotic cell cycle
		GO:1901987	regulation of cell cycle phase transition
		GO:0097193	intrinsic apoptotic signaling pathway
		GO:0009314	response to radiation
		GO:0104004	cellular response to environmental stimulus
Pathway	KEGG	hsa05220	Chronic myeloid leukemia
		hsa04110	Cell cycle
		hsa05214	Glioma
		hsa05218	Melanoma
		hsa04210	Apoptosis

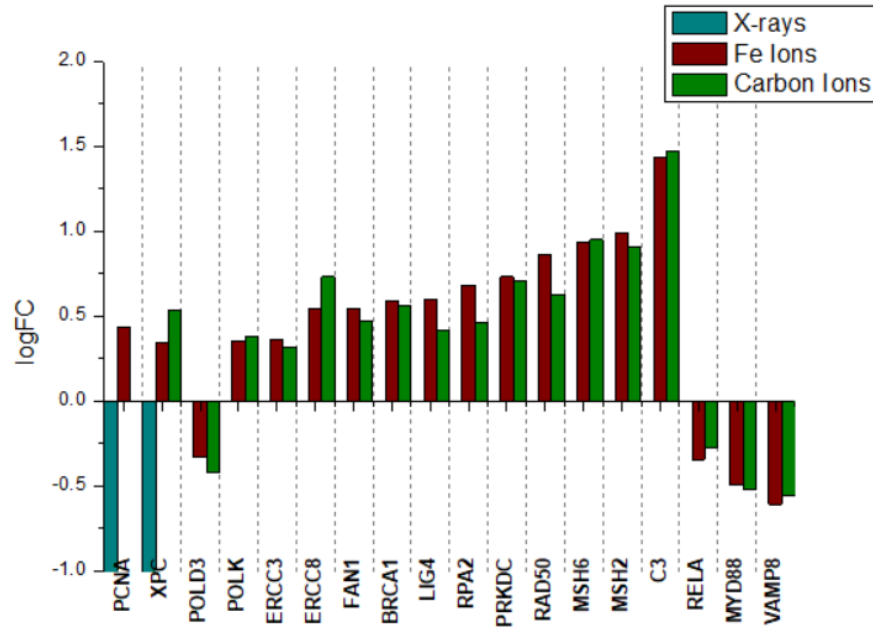
	hsa05200	Pathways in cancer
	hsa04060	Cytokine-cytokine receptor interaction
	hsa04115	p53 signaling pathway
Panther	P04398	p53 pathway feedback loops 2
	P00006	Apoptosis signaling pathway
	P00059	p53 pathway
Reactome	R-HSA-5633008	TP53 Regulates Transcription of Cell Death Genes
	R-HSA-212436	Generic Transcription Pathway
	R-HSA-73857	RNA Polymerase II Transcription
	R-HSA-74160	Gene expression (Transcription)
	R-HSA-597592	Post-translational protein modification
	R-HSA-5696398	Nucleotide Excision Repair
	R-HSA-392499	Metabolism of proteins
	R-HSA-3700989	Transcriptional Regulation by TP53
	R-HSA-73894	DNA Repair
Wiki Pathways	WP4155	Endometrial cancer
	WP4262	Breast cancer pathway
	WP45	G1 to S cell cycle control
	WP179	Cell Cycle
	WP254	Apoptosis
	WP3982	miRNA regulation of p53 pathway in prostate cancer
	WP707	DNA Damage Response
	WP1530	miRNA Regulation of DNA Damage Response

3.1.2. Ποσοτική Σύγκριση Έκφρασης Σημαντικών Γονιδίων

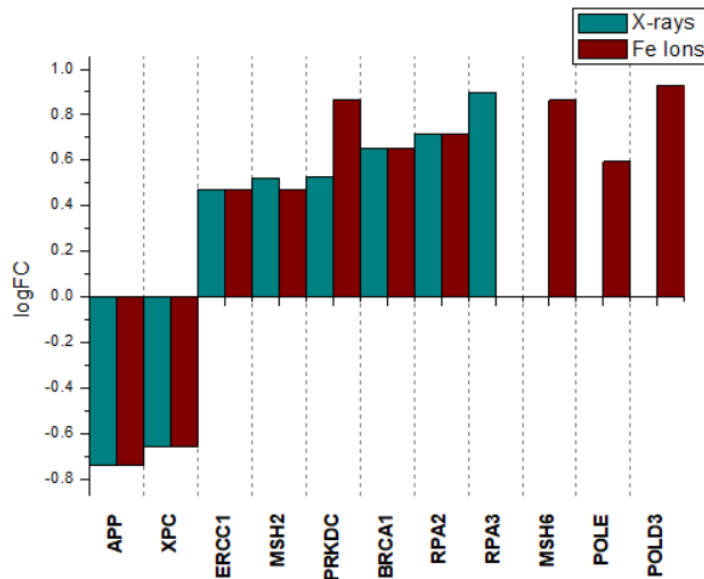
Με χρήση του λογισμικού OriginPro8 σχεδιάστηκαν τα διαγράμματα στηλών έκφρασης (Εικόνα 47, Εικόνα 48, Εικόνα 49) των διαφορικά εκφρασμένων γονιδίων. Η ανάλυση αφορά στα γονίδια από τις 3 σειρές δεδομένων με δείγματα εκτεθειμένα σε ακτινοβολία υψηλής και χαμηλής τιμής LET, τα οποία βρέθηκαν κοινά με εκείνα που συμμετέχουν σε διαδικασίες επιδιόρθωσης DNA και ανοσοαπόκρισης, από την βιβλιογραφική αναζήτηση που εκτελέστηκε βάση όρων γονιδιακής οντολογίας (Πίνακας 12). Σημειώνεται ότι ορισμένα από τα κρίσιμα γονίδια που επιλέχθηκαν και σχεδιάστηκαν σε κάθε μελέτη δεν εμφανίστηκαν ως στατιστικώς σημαντικά διαφορικά εκφρασμένα για όλους τους τύπους ακτινοβολίας.



Εικόνα 47: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικά εκφρασμένα γονίδια της σειράς E-MTAB-5754. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας ακτινοβολημένα με ακτίνες-X και ιόντα σιδήρου, δόσης των 2Gy και συλλεγμένα 24 ώρες μετά την ακτινοβόληση.



Εικόνα 48: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικώς εκφρασμένα γονίδια της σειράς E-MTAB-5761/3463. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα κύτταρα περιφερειακού αίματος ακτινοβολημένα με ακτίνες-X, ιόντα σιδήρου και ιόντα άνθρακα, δόσης του 1Gy και συλλεγμένα 8 ώρες μετά την ακτινοβόληση.



Εικόνα 49: Διάγραμμα στηλών έκφρασης για τα κρίσιμα διαφορικώς εκφρασμένα γονίδια της σειράς GSE44282. Σε αυτήν χρησιμοποιήθηκαν ανθρώπινα βρογχικά επιθηλιακά κύτταρα ακτινοβολημένα με ακτίνες-X και ιόντα σιδήρου, δόσης του 1Gy και συλλεγμένα 12-24 ώρες μετά την ακτινοβόληση.

Πίνακας 12: Κρίσιμα διαφορικώς εκφρασμένα γονίδια, που χρησιμοποιήθηκαν για τον σχεδιασμό των διαγραμμάτων στηλών έκφρασης, για κάθε μία από τις σειρές δεδομένων που περιλάμβαναν δείγματα εκτεθειμένα σε ακτινοβολία υψηλής και χαμηλής τιμής LET

Series Radiation	E-MTAB-5754		E-MTAB-3463 5761			GSE44282	
	X-rays	Fe Ions	X-rays	Fe Ions	C Ions	X-rays	Fe Ions
Genes	XPC	XPC	XPC	XPC	XPC	BRCA1	BRCA1
	B2M	B2M	PCNA	PCNA	-	XPC	XPC
	CDK9	CDK9		BRCA1	BRCA1	ERCC1	ERCC1
	PRKDC	PRKDC		ERCC8	ERCC8	PRKDC	PRKDC
	VAMP8	VAMP8		RAD50	RAD50	RPA2	RPA2
	IRAK1	RAD50		POLK	POLK	MSH2	MSH2
	POLE	POLK		C3	C3	APP	APP
	MSH6			POLD3	POLD3	RPA3	MSH6
	RPA3			FAN1	FAN1		POLE
	PCNA			RPA2	RPA2		POLD3
	NHEJ1			ERCC3	ERCC3		
	ERCC3			MYD88	MYD88		
	MYD88			MSH2	MSH2		
	RPA1			RELA	RELA		
	MSH2			MSH6	MSH6		
				PRKDC	PRKDC		
				VAMP8	VAMP8		
			LIG4	LIG4			
Gene Count	17		18			11	

3.2. Μελέτη Ομαδοποιημένων Δεδομένων

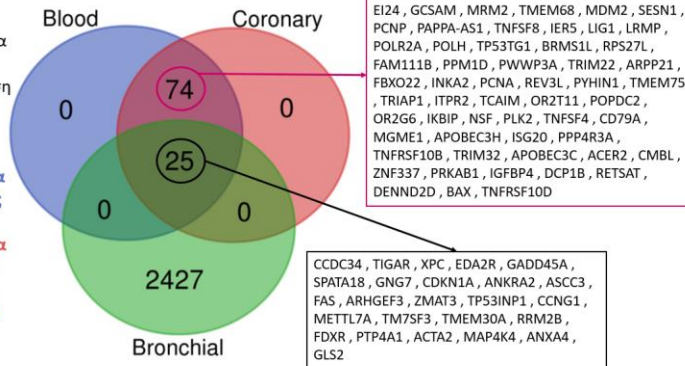
3.2.1. Ομαδοποίηση Δεδομένων – Διαγράμματα Venn

Επιλέγοντας τις 3 σειρές δεδομένων οι οποίες περιλάμβαναν δείγματα ακτινοβολημένα με υψηλές δόσεις ακτίνων-Χ και επιλέγοντας τα δείγματα που συλλέχθηκαν έως 8 ώρες έπειτα από την ακτινοβολήση σχεδιάστηκε το αντίστοιχο διάγραμμα Venn (Εικόνα 50).

Διάγραμμα Venn

Διαφορικά εκφρασμένα γονίδια έως 8 ώρες έπειτα από ακτινοβόληση με υψηλές δόσεις ακτίνων-Χ σε 3 τύπους ιστού:

1. Μονοπύρηννα κύτταρα περιφερειακού αίματος (E-MTAB-3463)
2. Ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας (E-MTAB-5754)
3. Βρογχικά επιθηλιακά κύτταρα (GSE44282)



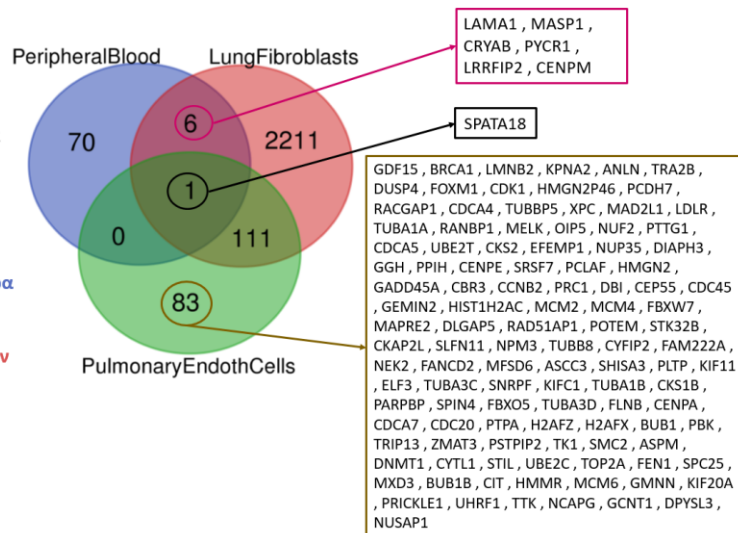
Εικόνα 50: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικά εκφρασμένα γονίδια των σειρών E-MTAB-3464, E-MTAB-5754 και GSE44282, έπειτα από ακτινοβόληση με υψηλές δόσεις ακτίνων-Χ. Όλα τα δείγματα συλλέχθηκαν έως 8 ώρες έπειτα από την ακτινοβόληση και περιλαμβάνουν 3 τύπους ανθρωπίνου ιστού: μονοπύρηννα κύτταρα περιφερειακού αίματος, ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας και βρογχικά επιθηλιακά κύτταρα.

Κατά τον ίδιο τρόπο, επιλέχθηκαν τα δείγματα από τις 2 σειρές δεδομένων που χρησιμοποίησαν ακτινοβόληση με σωματίδια-α υψηλής δόσης και περιλάμβαναν 3 διαφορετικούς τύπους ανθρωπίνου ιστού ώστε να σχεδιαστεί το διάγραμμα Venn (Εικόνα 51) και να αναδειχθούν τα κοινά διαφορικά εκφρασμένα γονίδια μεταξύ ιστών.

Διάγραμμα Venn

Διαφορικά εκφρασμένα γονίδια 24 ώρες έπειτα από ακτινοβόληση με υψηλές δόσεις σωματιδίων άλφα σε 3 τύπους ιστού:

1. Μονοπυρηνικά κύτταρα περιφερειακού αίματος (GSE80547)
2. Ινοβλάστες πνευμόνων (GSE81067)
3. Πνευμονικά ενδοθηλιακά κύτταρα (GSE81067)



Εικόνα 51: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικά εκφρασμένα γονίδια των σειρών GSE80547 και GSE81067 έπειτα από ακτινοβόληση με υψηλές δόσεις σωματιδίων άλφα. Όλα τα δείγματα συλλέχθηκαν 24 ώρες έπειτα από την ακτινοβόληση και περιλαμβάνουν 3 τύπους ανθρωπίνου ιστού: ανθρώπινα μονοπύρηννα κύτταρα περιφερειακού αίματος, ινοβλάστες πνευμόνων και πνευμονικά ενδοθηλιακά κύτταρα.

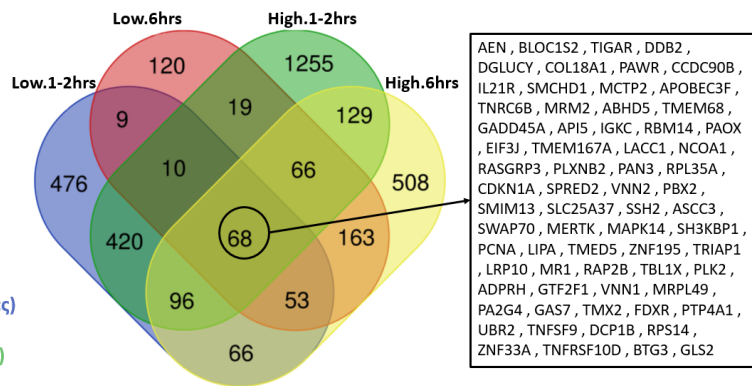
Έχοντας στην διάθεσή μας αρκετά δείγματα απομονωμένων λευκοκυττάρων από 2 σειρές δεδομένων, ομαδοποιήσαμε το διαφορικά εκφρασμένα γονίδια με βάση την δόση σε χαμηλή και υψηλή με όριο τα 0.05Gy αλλά και με βάση την ώρα συλλογής, δημιουργώντας 4 ομάδες για τον σχεδιασμό του διαγράμματος Venn (Εικόνα 52).

Διάγραμμα Venn

Διαφορικά εκφρασμένα γονίδια σε **απομονωμένα λευκοκύτταρα** έπειτα από ακτινοβολία με **βαρέα ιόντα** σε 4 κατηγορίες ανάλογα με την δόση της ακτινοβολίας και τον χρόνο συλλογής μετά την ακτινοβολία

(GSE63952|GSE64375):

1. Χαμηλές δόσεις (1-2 ώρες)
2. Χαμηλές δόσεις (6 ώρες)
3. Υψηλές δόσεις (1-2 ώρες)
4. Υψηλές δόσεις (6 ώρες)



AEN, BLOC1S2, TIGAR, DDB2, DGLUCY, COL18A1, PAWR, CCDC90B, IL21R, SMCHD1, MCTP2, APOBEC3F, TNRC6B, MRM2, ABHD5, TMEM68, GADD45A, API5, IGC, RBM14, PAOX, EIF3J, TMEM167A, LACC1, NCOA1, RASGRP3, PLXNB2, PAN3, RPL35A, CDKN1A, SPRED2, VNN2, PBX2, SMIM13, SLC25A37, SSH2, ASCC3, SWAP70, MERTK, MAPK14, SH3KBP1, PCNA, LIPA, TMED5, ZNF195, TRIAP1, LRP10, MR1, RAP2B, TBL1X, PLK2, ADPRH, GTF2F1, VNN1, MRPL49, PA2G4, GAS7, TMX2, FDXR, PTP4A1, UBR2, TNFSF9, DCP1B, RPS14, ZNF33A, TNFRSF10D, BTG3, GLS2

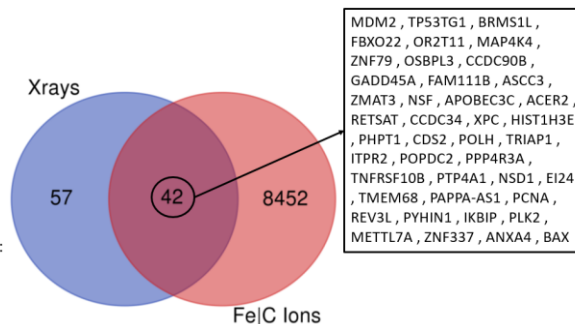
Εικόνα 52: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικά εκφρασμένα γονίδια των σειρών GSE63952 και GSE64375, που περιλαμβάνουν δείγματα από ανθρώπινα απομονωμένα λευκοκύτταρα ακτινοβολημένα με βαρέα ιόντα υψηλών και χαμηλών δόσεων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-2 και 6 ωρών έπειτα από την ακτινοβολία.

Από τις δύο σειρές δεδομένων που ανήκουν στην ίδια μελέτη, με δείγματα ανθρώπινων μονοκύριων κυττάρων περιφερειακού αίματος ακτινοβολημένα με υψηλές δόσεις βαρέων ιόντων και ακτίνων-X, σχεδιάστηκε το αντίστοιχο διάγραμμα Venn (Εικόνα 53). Όλα τα δείγματα συλλέχθηκαν 8 ώρες μετά την ακτινοβολία.

Διάγραμμα Venn

Κοινά διαφορικά εκφρασμένα γονίδια των σειρών E-MTAB-5761|3463 σε **μονοκύρινα κύτταρα περιφερειακού αίματος**, 8 ώρες έπειτα από ακτινοβολία με **υψηλές δόσεις**:

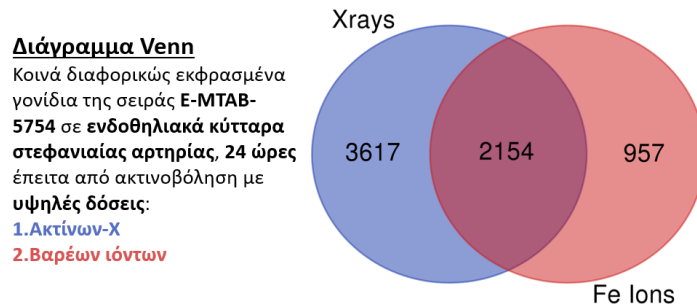
1. Ακτίνων-X
2. Βαρέων ιόντων



MDM2, TP53TG1, BRMS1L, FBXO22, OR2T11, MAP4K4, ZNF79, OSBPL3, CCDC90B, GADD45A, FAM111B, ASCC3, ZMAT3, NSF, APOBEC3C, ACER2, RETSAT, CCDC34, XPC, HIST1H3E, PHPT1, CDS2, POLH, TRIAP1, ITPR2, POPDC2, PPP4R3A, TNFRSF10B, PTP4A1, NSD1, EI24, TMEM68, PAPPAS1, PCNA, REV3L, PYHIN1, IKBIP, PLK2, METTL7A, ZNF337, ANXA4, BAX

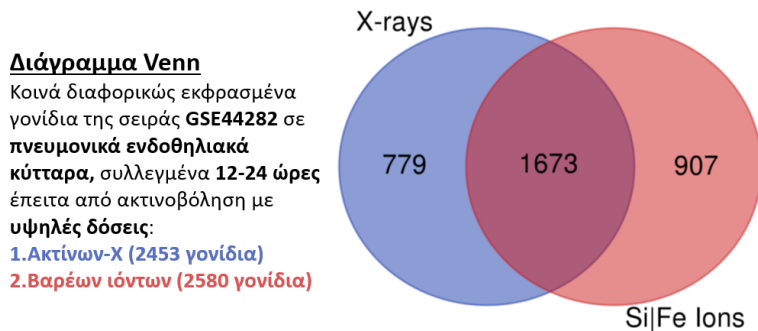
Εικόνα 53: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικά εκφρασμένα γονίδια των σειρών E-MTAB-3464 και E-MTAB-5761, που περιλαμβάνουν δείγματα από ανθρώπινα μονοκύρινα κύτταρα περιφερειακού αίματος ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-X υψηλής δόσης. Τα δείγματα συλλέχθηκαν 8 ώρες έπειτα από την ακτινοβολία.

Στη συνέχεια κατά τον ίδιο τρόπο σχεδιάστηκε το διάγραμμα Venn (Εικόνα 54) για την σειρά με δείγματα ανθρώπινων ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας ακτινοβολημένα με υψηλές δόσεις βαρέων ιόντων και ακτίνων-Χ. Όλα τα δείγματα συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.



Εικόνα 54: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορετικώς εκφρασμένα γονίδια της σειράς E-MTAB-5754, που περιλαμβάνουν δείγματα από ανθρώπινα ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-Χ υψηλής δόσης. Τα δείγματα συλλέχθηκαν 24 ώρες έπειτα από την ακτινοβολήση.

Ομοίως και για την σειρά που χρησιμοποιεί δείγματα πνευμονικών ενδοθηλιακών κυττάρων ακτινοβολημένα με υψηλές δόσεις βαρέων ιόντων και ακτίνων-Χ σχεδιάστηκε το διάγραμμα Venn (Εικόνα 55) για δείγματα που συλλέχθηκαν 12-24 ώρες μετά την ακτινοβολήση.



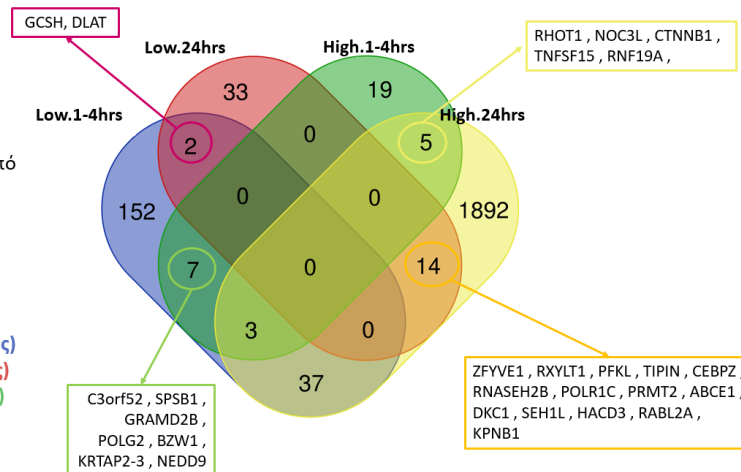
Εικόνα 55: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορετικώς εκφρασμένα γονίδια της σειράς GSE44282, που περιλαμβάνουν δείγματα από ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα ακτινοβολημένα με βαρέα ιόντα ή ακτίνες-Χ υψηλής δόσης. Τα δείγματα συλλέχθηκαν 12-24 ώρες έπειτα από την ακτινοβολήση.

Στην ίδια σειρά δεδομένων αποφασίσαμε να σχεδιάσουμε ένα ακόμα διάγραμμα Venn (Εικόνα 56) ώστε να μελετήσουμε και ξεχωριστά τα δείγματα που ακτινοβολήθηκαν με ιόντα σιδήρου, ομαδοποιώντας με βάση την δόση και τον χρόνο έπειτα από την ακτινοβολήση.

Διάγραμμα Venn

Διαφορικώς εκφρασμένα γονίδια σε πνευμονικά ενδοθηλιακά κύτταρα της σειράς GSE44282 έπειτα από ακτινοβολήση με ιόντα σιδήρου σε 4 κατηγορίες ανάλογα με την δόση της ακτινοβολίας και τον χρόνο συλλογής μετά την ακτινοβολήση:

1. Χαμηλές δόσεις (1-4 ώρες)
2. Χαμηλές δόσεις (24 ώρες)
3. Υψηλές δόσεις (1-4 ώρες)
4. Υψηλές δόσεις (24 ώρες)



Εικόνα 56: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια της σειράς GSE44282, που περιλαμβάνουν δείγματα από ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα ακτινοβολημένα με ιόντα σιδήρου υψηλών και χαμηλών δόσεων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-4 και 24 ωρών έπειτα από την ακτινοβολήση.

Τέλος ομαδοποιήσαμε τα διαφορικώς εκφρασμένα γονίδια από όλες τις σειρές που περιλάμβαναν δείγματα ακτινοβολημένα με βαρέα ιόντα, ανεξαρτήτως ιστού, ώστε να εντοπιστούν τυχόν κοινά γονίδια μέσω του διαγράμματος Venn (Εικόνα 57). Δημιουργήθηκαν 4 ομάδες με βάση την δόση (χαμηλή και υψηλή) αλλά και με βάση τις ώρες συλλογής μετά την ακτινοβολήση διαχωρίζοντας την πρώιμη απόκριση (1-4 ώρες) και την μεταγενέστερη (24 ώρες).

Διάγραμμα Venn

Διαφορικώς εκφρασμένα γονίδια σε πνευμονικά ενδοθηλιακά κύτταρα, απομονωμένα λευκοκύτταρα, μονοπύρρηνα περιφερειακού αίματος και ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας των σειρών:

GSE44282/64375/63952

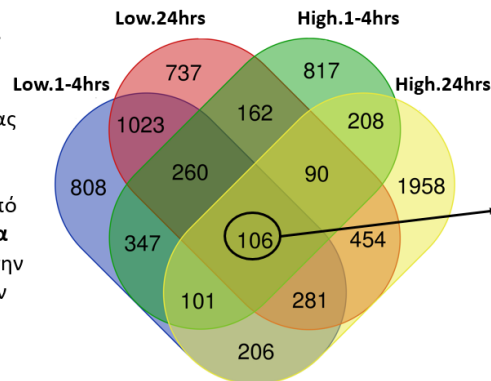
E-MTAB-5761/5754 έπειτα από ακτινοβολήση με **βαρέα ιόντα** σε 4 κατηγορίες ανάλογα με την δόση της ακτινοβολίας και τον χρόνο συλλογής μετά την ακτινοβολήση:

1. Χαμηλές δόσεις (1-4 ώρες)

2. Χαμηλές δόσεις (24 ώρες)

3. Υψηλές δόσεις (1-4 ώρες)

4. Υψηλές δόσεις (24 ώρες)



FARP1 , BRIX1 , AP1S2 , TRAF5 , UFM1 , COL18A1 , TMEM59 , DDX18 , NOP58 , SVIL , TOMM22 , KIF23 , URB2 , YBX1 , IPPK , MRPL14 , EIF4ENIF1 , NCL , SYT11 , CDKN1A , UPF2 , N4BP2L1 , MCPH1 , MAFF , CCT6A , ERGIC1 , RAE1 , SLC25A37 , SPATA5L1 , SEPTIN6 , DOCK8 , ZMYND8 , MERTK , MEPE , PHF19 , CDC42EP2 , TMEM169 , TMED5 , FTSJ3 , ZNF195 , TRIAP1 , PABPC4 , NCBP2 , GADD45GIP1 , SF3B1 , IPO7 , PPP4R3A , WSB2 , HNRNPLL , ST13 , UBR2 , PRICKLE1 , PEG10 , SSBP1 , SESN2 , RHOT1 , AVEN , CCT8 , ZNF467 , NOC3L , ZNF561 , CCND1 , SMCHD1 , NCOA5 , ADARB1 , FAM207A , EIF5B , NAMPT , FAM57A , ARL4D , FEZ1 , CTPS1 , DNAJC21 , PCMTD1 , RIOX1 , IMPDH1 , HSDL2 , SMIM13 , C1orf43 , ANKRA2 , AFF1 , UTP4 , MPPE1 , SWAP70 , WDR43 , PDPR , DRAM1 , MRPL17 , TWNK , EMC7 , SNTB2 , PSMD14 , KIAA0040 , NSUN2 , UGDH , ATP6V1G1 , AAGAB , PA2G4 , ARHGDI1 , NOLC1 , SART3 , KLF10 , ZNF33A , NCOA4 , GGA2 , RPS6KA3

Εικόνα 57: Διάγραμμα Venn για τα στατιστικώς σημαντικά διαφορικώς εκφρασμένα γονίδια των σειρών GSE44282, GSE64375, GSE63952, E-MTAB-5761 και E-MTAB-5754 έπειτα από ακτινοβολήση με υψηλές και χαμηλές δόσεις βαρέων ιόντων. Τα δείγματα διαχωρίστηκαν με βάση την ώρα συλλογής σε ομάδες 1-4 και 24 ωρών έπειτα από την ακτινοβολήση και περιλαμβάνουν: ανθρώπινα πνευμονικά ενδοθηλιακά κύτταρα, απομονωμένα λευκοκύτταρα, μονοπύρρηνα κύτταρα περιφερειακού αίματος και ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας.

3.2.2. Λειτουργική Ανάλυση Εμπλουτισμού

Παρατίθενται τα αποτελέσματα των λειτουργικών αναλύσεων εμπλουτισμού από το WebGestalt με την μέθοδο ORA. Σε αυτό το στάδιο χρησιμοποιήθηκαν οι λίστες γονιδίων που προέκυψαν από την ομαδοποίηση των δειγμάτων όλων των σειρών σε συνδυασμό με τα παραπάνω διαγράμματα Venn:

Πίνακας 13: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορετικώς εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5754, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X, τα οποία συλλέχθηκαν 8-24 ώρες μετά την ακτινοβόληση.

WebGestalt - ORA			
Peripheral Blood Mononuclear Cells (E-MTAB-3464) Coronary Artery Endothelial Cells (E-MTAB-5754)			
X-rays High Dose 8-24hrs 99 GENES			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0042770	signal transduction in response to DNA damage
		GO:0072331	signal transduction by p53 class mediator
		GO:2001020	regulation of response to DNA damage stimulus
		GO:0007050	cell cycle arrest
		GO:0000075	cell cycle checkpoint
		GO:0044843	cell cycle G1/S phase transition
		GO:0045862	positive regulation of proteolysis
		GO:0090305	nucleic acid phosphodiester bond hydrolysis
		GO:0009314	response to radiation
		GO:0009896	positive regulation of catabolic process
Pathway	KEGG	hsa04115	p53 signaling pathway
	Panther	P00059	p53 pathway
	Reactome	R-HSA-3700989	Transcriptional Regulation by TP53
		R-HSA-212436	Generic Transcription Pathway
		R-HSA-73857	RNA Polymerase II Transcription
		R-HSA-74160	Gene expression (Transcription)
	Wiki Pathways	WP4286	Genotoxicity pathway
		WP3982	miRNA regulation of p53 pathway in prostate cancer
		WP707	DNA Damage Response
		WP1530	miRNA Regulation of DNA Damage Response

Πίνακας 14: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από την σειρά GSE81067, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων και ινοβλαστών πνεύμονα, ακτινοβολημένα με υψηλές δόσεις σωματιδίων άλφα, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.

WebGestalt - ORA			
Pulmonary Endothelial Cells (HPEAC) Lung Fibroblasts (HLF-1) (GSE81067)			
Alpha Particles High Dose 24hrs 112 GENES			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:1902850	microtubule cytoskeleton organization involved in mitosis
		GO:0007059	chromosome segregation
		GO:0051321	meiotic cell cycle
		GO:0044772	mitotic cell cycle phase transition
		GO:1901987	regulation of cell cycle phase transition
		GO:0033044	regulation of chromosome organization
	GO Cellular Component	GO:0000793	condensed chromosome
		GO:0098687	chromosomal region
Pathway	KEGG	hsa04110	Cell cycle
	Reactome	R-HSA-2500257	Resolution of Sister Chromatid Cohesion
		R-HSA-68877	Mitotic Prometaphase
		R-HSA-2555396	Mitotic Metaphase and Anaphase
		R-HSA-69278	Cell Cycle, Mitotic
		R-HSA-68886	M Phase
		R-HSA-1640170	Cell Cycle
		R-HSA-69620	Cell Cycle Checkpoints
R-HSA-195258	RHO GTPase Effectors		

Πίνακας 15: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές GSE63952/64375, που προήλθαν από δείγματα ανθρωπίνων λευκοκυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 6 ώρες μετά την ακτινοβολήση.

WebGestalt - ORA			
Isolated Leucocytes (GSE63952 GSE64375)			
HZE Particles Low Dose 6hrs 509 GENES			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0042770	signal transduction in response to DNA damage
		GO:0071826	ribonucleoprotein complex subunit organization
		GO:0022613	ribonucleoprotein complex biogenesis
		GO:0072331	signal transduction by p53 class mediator
		GO:0045930	negative regulation of mitotic cell cycle
		GO:0009314	response to radiation

		GO:0006310	DNA recombination
Pathway	KEGG	hsa04115	p53 signaling pathway
	Reactome	R-HSA-3700989	Transcriptional Regulation by TP53
		R-HSA-74160	Gene expression (Transcription)
		R-HSA-73857	RNA Polymerase II Transcription
		R-HSA-8953854	Metabolism of RNA
		R-HSA-72165	mRNA Splicing - Minor Pathway
		R-HSA-212436	Generic Transcription Pathway
	Wiki Pathways	WP4286	Genotoxicity pathway
		WP707	DNA Damage Response
WP1530		miRNA Regulation of DNA Damage Response	

HZE Particles | High Dose | 6hrs | 1149 GENES

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0042770	signal transduction in response to DNA damage
		GO:0045930	negative regulation of mitotic cell cycle
		GO:0036230	granulocyte activation
		GO:0002446	neutrophil mediated immunity
		GO:0043087	regulation of GTPase activity
	GO Cellular Component	GO:0030667	secretory granule membrane
Pathway	KEGG	hsa04115	p53 signaling pathway
		hsa05220	Chronic myeloid leukemia
	Panther	P00059	p53 pathway
	Reactome	R-HSA-5628897	TP53 Regulates Metabolic Genes
		R-HSA-3700989	Transcriptional Regulation by TP53
		R-HSA-6798695	Neutrophil degranulation
		R-HSA-74160	Gene expression (Transcription)
		R-HSA-73857	RNA Polymerase II Transcription
		R-HSA-212436	Generic Transcription Pathway
		R-HSA-168249	Innate Immune System
R-HSA-168256	Immune System		
Wiki Pathways	WP4286	Genotoxicity pathway	
	WP4255	Non-small cell lung cancer	

Πίνακας 16: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα ομαδοποιημένα διαφορικώς εκφρασμένα γονίδια από την σειρά GSE44282, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις ιόντων σιδήρου, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβόληση.

WebGestalt - ORA			
Bronchial Epithelial Cells – HBEC3KT (GSE44282)			
Fe Ions Low Dose 24hrs 35 GENES			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0006403	RNA localization
		GO:0006399	tRNA metabolic process
		GO:0034470	ncRNA processing
	GO Cellular Component	GO:0005732	small nucleolar ribonucleoprotein complex
		GO:0044452	nucleolar part
		GO:0101002	ficolin-1-rich granule
		GO:0005635	nuclear envelope
GO:0005759	mitochondrial matrix		
GO Molecular Function	GO:0140098	catalytic activity, acting on RNA	
Pathway	KEGG	hsa03008	Ribosome biogenesis in eukaryotes
	Reactome	R-HSA-6787450	tRNA modification in the mitochondrion
		R-HSA-168276	NS1 Mediated Effects on Host Pathways
		R-HSA-168253	Host Interactions with Influenza Factors
		R-HSA-1169410	Antiviral mechanism by IFN-stimulated genes
		R-HSA-6790901	rRNA modification in the nucleus and cytosol
		R-HSA-1169408	ISG15 antiviral mechanism
		R-HSA-72306	tRNA processing
		R-HSA-913531	Interferon Signaling
R-HSA-8953854	Metabolism of RNA		
Fe Ions High Dose 24hrs 1950 GENES			
Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0022613	ribonucleoprotein complex biogenesis
		GO:0034470	ncRNA processing
		GO:0016072	rRNA metabolic process
		GO:0006397	mRNA processing
		GO:0071826	ribonucleoprotein complex subunit organization
		GO:0034248	regulation of cellular amide metabolic process
		GO:0010608	posttranscriptional regulation of gene expression
		GO:0008380	RNA splicing
		GO:0032200	telomere organization
	GO Cellular Component	GO:0030684	preribosome
	GO:0044452	nucleolar part	

	GO Molecular Function	GO:0140098	catalytic activity, acting on RNA
Pathway	Reactome	R-HSA-8953854	Metabolism of RNA
		R-HSA-72312	rRNA processing
		R-HSA-8868773	rRNA processing in the nucleus and cytosol
		R-HSA-6791226	Major pathway of rRNA processing in the nucleolus and cytosol
		R-HSA-73894	DNA Repair
		R-HSA-6790901	rRNA modification in the nucleus and cytosol
		R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA
		R-HSA-3371556	Cellular response to heat stress

Πίνακας 17: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων γονιδιακής οντολογίας και βιολογικών μονοπατιών από το WebGestalt με την μέθοδο ORA. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές GSE44282/63952/64375 και E-MTAB-5761/5754, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, απομονωμένων λευκοκυττάρων, μονοπύρηνων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 1-24 ώρες μετά την ακτινοβολήση.

WebGestalt - ORA

HZE Particles (GSE44282|63952|64375, E-MTAB-5761|5754)

Low Dose | 1-4hrs | 3133 GENES

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0016072	rRNA metabolic process
		GO:0097193	intrinsic apoptotic signaling pathway
		GO:0022613	ribonucleoprotein complex biogenesis
		GO:0034470	ncRNA processing
		GO:0006401	RNA catabolic process
		GO:0034248	regulation of cellular amide metabolic process
		GO:0010608	posttranscriptional regulation of gene expression
	GO Cellular Component	GO:0005840	ribosome
Pathway	Reactome	R-HSA-72312	rRNA processing
		R-HSA-8868773	rRNA processing in the nucleus and cytosol
		R-HSA-6791226	Major pathway of rRNA processing in the nucleolus and cytosol
		R-HSA-72766	Translation
		R-HSA-3700989	Transcriptional Regulation by TP53
		R-HSA-8953854	Metabolism of RNA
		R-HSA-5663205	Infectious disease
		R-HSA-74160	Gene expression (Transcription)
		R-HSA-73857	RNA Polymerase II Transcription
		R-HSA-168256	Immune System

Low Dose | 6-24hrs | 3496 GENES

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0006403	RNA localization
		GO:0006399	tRNA metabolic process
		GO:0016072	rRNA metabolic process
		GO:0022613	ribonucleoprotein complex biogenesis
		GO:0034470	ncRNA processing
		GO:0071826	ribonucleoprotein complex subunit organization
		GO:0034248	regulation of cellular amide metabolic process
		GO:0002446	neutrophil mediated immunity
		GO:0010608	posttranscriptional regulation of gene expression
		GO:0036230	granulocyte activation
	GO Molecular Function	GO:0140098	catalytic activity, acting on RNA
Pathway	Reactome	R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA
		R-HSA-8953854	Metabolism of RNA
		R-HSA-3700989	Transcriptional Regulation by TP53
		R-HSA-6798695	Neutrophil degranulation
		R-HSA-168256	Immune System

High Dose | 1-4hrs | 2090 GENES

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0016072	rRNA metabolic process
		GO:0006403	RNA localization
		GO:0022613	ribonucleoprotein complex biogenesis
		GO:0034470	ncRNA processing
	GO Cellular Component	GO:0030684	preribosome
Pathway	KEGG	hsa05169	Epstein-Barr virus infection
	Reactome	R-HSA-6790901	rRNA modification in the nucleus and cytosol
		R-HSA-5607761	Dectin-1 mediated noncanonical NF-kB signaling
		R-HSA-1168372	Downstream signaling events of B Cell Receptor (BCR)
		R-HSA-8878159	Transcriptional regulation by RUNX3
		R-HSA-202424	Downstream TCR signaling
		R-HSA-983705	Signaling by the B Cell Receptor (BCR)
		R-HSA-8953854	Metabolism of RNA
		R-HSA-983168	Antigen processing: Ubiquitination & Proteasome degradation
		R-HSA-74160	Gene expression (Transcription)
		R-HSA-1280218	Adaptive Immune System
		R-HSA-73857	RNA Polymerase II Transcription
		R-HSA-212436	Generic Transcription Pathway

High Dose | 6-24hrs | 3498 GENES

Category	Database	GeneSet	Description
Gene Ontology	GO Biological Process	GO:0016072	rRNA metabolic process
		GO:0006260	DNA replication
		GO:0007059	chromosome segregation
		GO:0034470	ncRNA processing
		GO:0022613	ribonucleoprotein complex biogenesis
		GO:0008380	RNA splicing
		GO:0006397	mRNA processing
		GO:0048285	organelle fission
	GO:0044772	mitotic cell cycle phase transition	
	GO Cellular Component	GO:0000793	condensed chromosome
GO:0098687		chromosomal region	
GO:0005819		spindle	
GO:0030055		cell-substrate junction	
GO Molecular Function	GO:0140098	catalytic activity, acting on RNA	
Pathway	KEGG	hsa04142	Lysosome
	Reactome	R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA
		R-HSA-69620	Cell Cycle Checkpoints
		R-HSA-69278	Cell Cycle, Mitotic
		R-HSA-1640170	Cell Cycle
		R-HSA-8953854	Metabolism of RNA

3.2.3. Ανάλυση Εμπλουτισμού για Όρους Ασθενειών

Παρατίθενται τα αποτελέσματα των λειτουργικών αναλύσεων εμπλουτισμού από το Enrichr. Σε αυτό το στάδιο χρησιμοποιήθηκαν οι λίστες γονιδίων που προέκυψαν από την ομαδοποίηση των δειγμάτων όλων των σειρών σε συνδυασμό με τα παραπάνω διαγράμματα Venn:

Πίνακας 18: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορικώς εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5754, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος και ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X, τα οποία συλλέχθηκαν 8-24 ώρες μετά την ακτινοβόληση.

Enrichr				
Peripheral Blood Mononuclear Cells (E-MTAB-3464) Coronary Artery Endothelial Cells (E-MTAB-5754)				
X-rays High Dose 8-24hrs 99 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	xeroderma pigmentosum	2/9	60.06	459.20
	mitochondrial dna deletion syndromes	1/8	33.78	119.35
	myocardial infarction 1	1/13	20.79	63.54
	neoplasm of the breast	1/27	10.01	23.53
DisGeNET	BARDET-BIEDL SYNDROME 11	2/5	108.11	964.01
	Mycobacterium avium infection	2/6	90.09	767.03
	Grade III Chondrosarcoma	2/7	77.22	631.65
	Common wart	2/7	77.22	631.65
	Poikiloderma	3/19	42.67	426.98
	Leukemia, Plasma Cell	4/44	24.57	264.70
	Varicose Ulcer	3/34	23.85	196.14
	Adenovirus Infections	4/89	12.15	97.29
	Creatine phosphokinase serum increased	5/158	8.55	69.38
	Infection	9/491	4.95	46.69
	B-Cell Lymphomas	10/729	3.71	29.61
	Primary malignant neoplasm of lung	22/2267	2.62	29.19
	Carcinoma of lung	22/2475	2.40	23.45
	Malignant neoplasm of lung	21/2448	2.32	20.41

Πίνακας 19: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορετικώς εκφρασμένα γονίδια από την σειρά GSE81067, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων και ινοβλαστών πνεύμονα, ακτινοβολημένα υψηλές δόσεις σωματιδίων άλφα, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.

Enrichr				
Pulmonary Endothelial Cells (HPEAC) Lung Fibroblasts (HLF-1) (GSE81067)				
Alpha Particles High Dose 24hrs 112 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	primary autosomal recessive microcephaly	3/15	36.04	343.69
	familial hypercholesterolemia	1/5	36.04	129.57
	malformation of cortical development	1/6	30.03	102.58
	familial hypercholesterolemias	1/8	22.52	70.58
	neoplasm of the breast	2/27	13.35	61.75
	carcinoma of colon	2/30	12.01	53.13
DisGeNET	Warburton Anyane Yeboa syndrome	3/5	108.11	1439.41
	MOSAIC VARIEGATED ANEUPLOIDY SYNDROME	3/6	90.09	1137.43
	Increased nuchal translucency	3/8	67.57	784.05
	Seckel syndrome	5/35	25.74	347.57
	Cortical gyral simplification	4/25	28.83	330.80
	Sclerocystic Ovaries	10/165	10.92	189.92
	Neoplasms, Intracranial	7/96	13.14	180.70
	Precursor cell lymphoblastic lymphoma	5/61	14.77	158.16
	Triple Negative Breast Neoplasms	17/631	4.85	79.91
	Colorectal Neoplasms	24/1072	4.03	78.26
	Malignant neoplasm of lung	40/2448	2.94	68.20
	Non-Small Cell Lung Carcinoma	37/2242	2.97	63.96
	Liver carcinoma	50/3592	2.51	60.10
	Squamous cell carcinoma of the head and neck	20/933	3.86	59.87
	Malignant tumor of cervix	19/883	3.88	57.41
	Malignant neoplasm of urinary bladder	22/1143	3.47	52.51
	Primary malignant neoplasm of lung	35/2267	2.78	51.43
	Pancreatic Neoplasm	15/664	4.07	50.61
	Epithelial ovarian cancer	24/1328	3.26	49.94
	Glioblastoma	31/1936	2.89	49.26
	Hematologic Neoplasms	13/556	4.21	47.45
	Colon Carcinoma	32/2090	2.76	45.79
	Malignant neoplasm of pancreas	29/1845	2.83	43.99
Malignant neoplasm of breast	58/5053	2.07	42.91	
leukemia	29/1940	2.69	38.99	
Breast Carcinoma	56/4962	2.03	38.89	
Secondary malignant neoplasm of lymph node	21/1270	2.98	35.92	

Πίνακας 20: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα κοινά διαφορικά εκφρασμένα γονίδια από τις σειρές GSE6395/64375, που προήλθαν από δείγματα ανθρωπίνων λευκοκυττάρων, ακτινοβολημένα με χαμηλές και υψηλές δόσεις βαρέων ιόντων, τα οποία συλλέχθηκαν 6 ώρες μετά την ακτινοβολήση.

Enrichr				
Isolated Leucocytes (GSE63952 GSE64375)				
HZE Particles Low Dose 6hrs 509 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	xeroderma pigmentosum	4/9	17.50	174.44
	congenital disorder of glycosylation	4/37	4.26	18.17
DisGeNET	Bone marrow toxicity	4/8	19.69	207.41
	Mycobacterium avium infection	3/6	19.69	159.18
	Abnormal upper motor neuron morphology	4/14	11.25	89.96
	Basal cell nevus	5/26	7.57	58.54
	Leukemia, Myelocytic, Acute	80/1702	1.85	31.04
	Hematologic Neoplasms	32/556	2.27	24.99
	Breast Carcinoma	178/4962	1.41	22.52
	Prostatic Neoplasms	69/1553	1.75	21.92
	Malignant neoplasm of breast	178/5053	1.39	20.31
	Colonic Neoplasms	39/777	1.98	19.82
	Lymphoma	58/1306	1.75	18.62
	Chronic Lymphocytic Leukemia	51/1119	1.79	18.21
	Malignant neoplasm of prostate	121/3238	1.47	18.14
	Diffuse Large B-Cell Lymphoma	30/574	2.06	17.88
	Solid Neoplasm	40/839	1.88	17.15
	Colon Carcinoma	82/2090	1.54	15.54
	Neuroblastoma	69/1697	1.60	15.45
	Nasopharyngeal carcinoma	36/781	1.81	14.00
Autoimmune Diseases	46/1059	1.71	13.98	
HZE Particles High Dose 6hrs 1149 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	mitochondrial dna deletion syndromes	3/8	6.53	31.11
	xeroderma pigmentosum	3/9	5.80	25.55
	cutis laxa, autosomal recessive	2/5	6.96	24.57
	muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type a, 1	2/5	6.96	24.57
	alpha-dystroglycan related dystrophy	2/6	5.80	18.34
	disorder of the urea cycle metabolism	2/6	5.80	18.34
	leigh syndrome	3/13	4.02	13.46
	neoplasm of the breast	5/27	3.22	13.06
	familial cancer of breast	5/27	3.22	13.06

	arrhythmogenic right ventricular cardiomyopathy	2/8	4.35	11.37
DisGeNET	Mucosal telangiectasiae	8/16	8.70	120.37
	Acute lymphocytic leukemia	103/906	1.98	48.92
	Chronic Lymphocytic Leukemia	120/1119	1.87	46.27
	Lymphoma	128/1306	1.71	34.91
	Hematologic Neoplasms	65/556	2.03	34.72
	Leukemia, Myelocytic, Acute	155/1702	1.59	30.75
	Multiple Myeloma	125/1311	1.66	30.45
	Acute leukemia	57/491	2.02	30.10
	Lymphoma, Non-Hodgkin	60/538	1.94	27.63
	Adenocarcinoma of lung (disorder)	112/1186	1.64	26.32
	Breast Carcinoma	368/4962	1.29	24.08
	Myeloid Leukemia	54/500	1.88	22.55
	Stomach Neoplasms	81/834	1.69	21.89
	Myeloid Leukemia, Chronic	84/875	1.67	21.59
	B-Cell Lymphomas	71/729	1.70	19.67
	Central neuroblastoma	140/1654	1.47	19.58
	Colon Carcinoma	170/2090	1.42	19.07
Glioblastoma	156/1936	1.40	16.65	
Malignant neoplasm of lung	190/2448	1.35	16.11	

Πίνακας 21: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα ομαδοποιημένα διαφορικώς εκφρασμένα γονίδια από την σειρά GSE44282, που προήλθαν από δείγματα ανθρωπίνων πνευμονικών ενδοθηλιακών κυττάρων, ακτινοβολημένα με υψηλές δόσεις ακτίνων-Χ και βαρέων ιόντων, τα οποία συλλέχθηκαν 12-24 ώρες μετά την ακτινοβόληση.

Enrichr				
Bronchial Epithelial Cells – HBEC3KT (GSE44282)				
X-rays High Dose 12-24hrs 2452 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	diamond-blackfan anemia	11/19	4.72	60.54
	hepatic failure, early-onset, and neurologic disorder due to cytochrome c oxidase deficiency	5/12	3.40	15.53
	hepatocellular carcinoma	4/9	3.63	14.76
	congenital myopathy with fiber type disproportion	3/6	4.08	14.64
	thoracic aortic aneurysm and aortic dissection	4/10	3.26	11.95
	DisGeNET	Mitochondrial Diseases	101/383	2.15
Neoplasm Invasiveness		50/182	2.24	39.58
Breast Carcinoma		761/4962	1.25	38.28
Non-Small Cell Lung Carcinoma		369/2242	1.34	29.24
Malignant neoplasm of lung		398/2448	1.33	29.05
Primary malignant neoplasm of lung		372/2267	1.34	28.89

Fanconi Anemia	66/289	1.86	27.76
Colorectal Carcinoma	463/2930	1.29	27.63
Uterine Fibroids	76/348	1.78	26.71
Carcinoma of lung	396/2475	1.31	25.76
Malignant neoplasm of ovary	328/2025	1.32	23.25
Colorectal Cancer	503/3297	1.24	22.64
Liver carcinoma	542/3592	1.23	22.25
Neuroblastoma	279/1697	1.34	21.97
Malignant neoplasm of stomach	377/2397	1.28	21.53
Epithelial ovarian cancer	224/1328	1.38	21.01
Ovarian Carcinoma	348/2202	1.29	20.62
Squamous cell carcinoma	302/1875	1.31	20.61
Colorectal Neoplasms	185/1072	1.41	20.15
Stomach Carcinoma	371/2377	1.27	19.96
Central neuroblastoma	269/1654	1.33	19.72
Adenocarcinoma of lung (disorder)	201/1186	1.38	19.55
Colon Carcinoma	330/2090	1.29	19.43

Fe, Si Ions | High Dose | 12-24hrs | 2580 GENES

Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	galloway-mowat syndrome	4/5	6.20	41.51
	xeroderma pigmentosum	5/9	4.31	25.23
	familial adenomatous polyposis 1	3/5	4.65	18.81
DisGeNET	Mitochondrial Diseases	94/383	1.90	41.67
	Breast Carcinoma	774/4962	1.21	28.17
	Fibrosarcoma	69/297	1.80	25.71
	Cerebral atrophy	74/337	1.70	22.00
	Malignant neoplasm of urinary bladder	206/1143	1.40	21.40
	Undifferentiated carcinoma	54/236	1.77	19.60
	Bladder Neoplasm	214/1216	1.36	19.17
	Epithelial ovarian cancer	230/1328	1.34	18.50
	Degenerative brain disorder	40/166	1.87	18.26
	Neoplasm Invasiveness	43/182	1.83	18.12
	Glioma	358/2210	1.26	17.27
	Renal Cell Carcinoma	230/1347	1.32	16.83
	Liver carcinoma	553/3592	1.19	16.81
	Malignant neoplasm of prostate	503/3238	1.20	16.66
	Squamous cell carcinoma	307/1875	1.27	16.23
Hepatocarcinogenesis	101/526	1.49	15.93	
Fanconi Anemia	61/289	1.64	15.86	

Πίνακας 22: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα διαφορικά εκφρασμένα γονίδια από τις σειρές E-MTAB-3463/5761, που προήλθαν από δείγματα ανθρωπίνων κυττάρων περιφερειακού αίματος, ακτινοβολημένα υψηλές δόσεις ακτίνων-X και βαρέων ιόντων, τα οποία συλλέχθηκαν 8 ώρες μετά την ακτινοβόληση.

Enrichr				
Peripheral Blood Mononuclear Cells (E-MTAB-3464 5761)				
X-rays High Dose 8hrs 100 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	xeroderma pigmentosum	3/9	67.34	777.53
	mitochondrial dna deletion syndromes	2/8	50.51	369.40
	autoimmune lymphoproliferative syndrome	1/6	33.67	118.82
	thoracic aortic aneurysm and aortic dissection	1/10	20.20	61.17
	myocardial infarction 1	1/13	15.54	43.09
	limb-girdle muscular dystrophy, autosomal recessive	1/14	14.43	38.98
	acute myeloid leukemia	1/20	10.10	23.83
	mitochondrial diseases	2/60	6.73	22.45
	neoplasm of the breast	1/27	7.48	15.53
DisGeNET	Common wart	3/7	86.58	1074.86
	Varicose Ulcer	4/34	23.77	253.45
	Gastric lymphoma	3/26	23.31	190.56
	Basal cell nevus	3/26	23.31	190.56
	Leukemia, Plasma Cell	4/44	18.37	176.82
	Cardiac Lymphoma	3/28	21.65	172.10
	keratoacanthoma	4/47	17.19	161.04
	Intestinal Cancer	4/57	14.18	122.05
	Precursor cell lymphoblastic lymphoma	4/61	13.25	110.56
	stage, non-small cell lung cancer	4/65	12.43	100.71
	Squamous cell carcinoma of skin	8/204	7.92	93.08
	Malignant Testicular Germ Cell Tumor	5/105	9.62	83.21
	Ki-1+ Anaplastic Large Cell Lymphoma	7/227	6.23	55.47
	Hepatoblastoma	7/234	6.04	52.68
	Malignant neoplasm of skin	8/306	5.28	46.85
	Burkitt Lymphoma	9/406	4.48	38.49
	Colonic Neoplasms	14/777	3.64	38.13
	Chronic Lymphocytic Leukemia	18/1119	3.25	37.95
	Primary malignant neoplasm of lung	29/2267	2.58	35.80
	Malignant neoplasm of liver	14/804	3.52	35.54
B-Cell Lymphomas	13/729	3.60	34.90	
Adult T-Cell Lymphoma/Leukemia	10/499	4.05	34.85	
Diffuse Large B-Cell Lymphoma	11/574	3.87	34.73	
Malignant neoplasm of lung	30/2448	2.48	33.29	

Liver and Intrahepatic Biliary Tract Carcinoma	11/606	3.67	31.17
Retinoblastoma	10/538	3.76	30.07
Lymphoma, Non-Hodgkin	10/538	3.76	30.07
Malignant tumor of cervix	14/883	3.20	29.15
Conventional (Clear Cell) Renal Cell Carcinoma	11/630	3.53	28.81
Carcinoma of lung	29/2475	2.37	28.59
Esophageal Neoplasms	11/636	3.49	28.25
Lymphoma	18/1306	2.78	26.80
Renal Cell Carcinoma	18/1347	2.70	24.92
cervical cancer	14/964	2.93	24.04

Fe, C Ions | High Dose | 8hrs | 8495 GENES

Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	renal dysplasia and retinal aplasia	6/7	2.59	12.97
	nephronophthisis	7/9	2.35	11.33
DisGeNET	Degenerative brain disorder	85/166	1.54	21.21
	Muscle degeneration	119/258	1.39	16.16
	Neurogenic muscle atrophy, especially in the lower limbs	100/214	1.41	15.03
	Skeletal muscle atrophy	100/214	1.41	15.03
	Curvature of spine	175/407	1.30	14.17
	Neurogenic Muscular Atrophy	104/227	1.38	13.83
	Epileptic encephalopathy	197/467	1.27	13.59
Cerebral atrophy	147/337	1.32	13.52	

Πίνακας 23: Αποτελέσματα της ανάλυσης εμπλουτισμού όρων ασθενειών και παθήσεων από το Enrichr. Χρησιμοποιήθηκαν τα διαφορετικώς εκφρασμένα γονίδια από την σειρά E-MTAB-5754, που προήλθαν από δείγματα ανθρωπίνων ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας, ακτινοβολημένα υψηλές δόσεις ακτίνων-X και ιόντων σιδήρου, τα οποία συλλέχθηκαν 24 ώρες μετά την ακτινοβολήση.

Enrichr				
Coronary Artery Endothelial Cells (E-MTAB-5754)				
X-rays High Dose 24hrs 5769 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	fanconi anemia	10/14	2.48	16.67
	glycogen storage disease	10/17	2.04	9.50
	neuronal ceroid lipofuscinosis	6/9	2.31	8.95
	neoplasm of the breast	12/27	1.54	4.30
	aplastic anemia	3/5	2.08	3.97
	thoracic aortic aneurysm and aortic dissection	5/10	1.73	3.52
DisGeNET	stage, bladder cancer	7/7	3.47	30.16
	Pyridoxine-responsive sideroblastic anemia	16/22	2.52	26.64
	Malformations of Cortical Development	49/98	1.73	20.34
	Chromosome Breakage	25/43	2.01	19.68
	Irregular hyperpigmentation	22/40	1.91	14.61
	Pancreatic Neoplasm	240/664	1.25	13.47
	Decreased platelet count	65/151	1.49	13.32
	Mental and motor retardation	353/1020	1.20	12.75
	Breast Cancer, Familial	65/153	1.47	12.47
Thrombocytopenia	130/340	1.33	12.10	
Fe Ions High Dose 24hrs 3110 GENES				
Database	Term	Overlap	Odds Ratio	Combined Score
ClinVar	cutis laxa, autosomal recessive	4/5	5.14	30.69
	fanconi anemia	7/14	3.21	18.98
	adult junctional epidermolysis bullosa	3/5	3.86	13.61
DisGeNET	Cutis Laxa	24/57	2.71	36.41
	Malignant mesothelioma	91/334	1.75	30.57
	Clouding of corneal stroma	26/74	2.26	23.75
	Fibrosarcoma	76/297	1.65	20.16
	Breast Carcinoma	889/4962	1.15	18.60
	Malignant neoplasm of pancreas	361/1845	1.26	17.82
	Malignant neoplasm of breast	899/5053	1.14	17.19
	Squamous cell carcinoma of the head and neck	196/933	1.35	16.93
	Thrombocytopenia	82/340	1.55	16.60
	Glioblastoma	374/1936	1.24	16.56
	Pancreatic carcinoma	362/1868	1.25	16.49
Pancreatic Neoplasm	145/664	1.40	16.38	

3.2.4. Συγγενείς Παθήσεις & Δίκτυα Ασθενειών

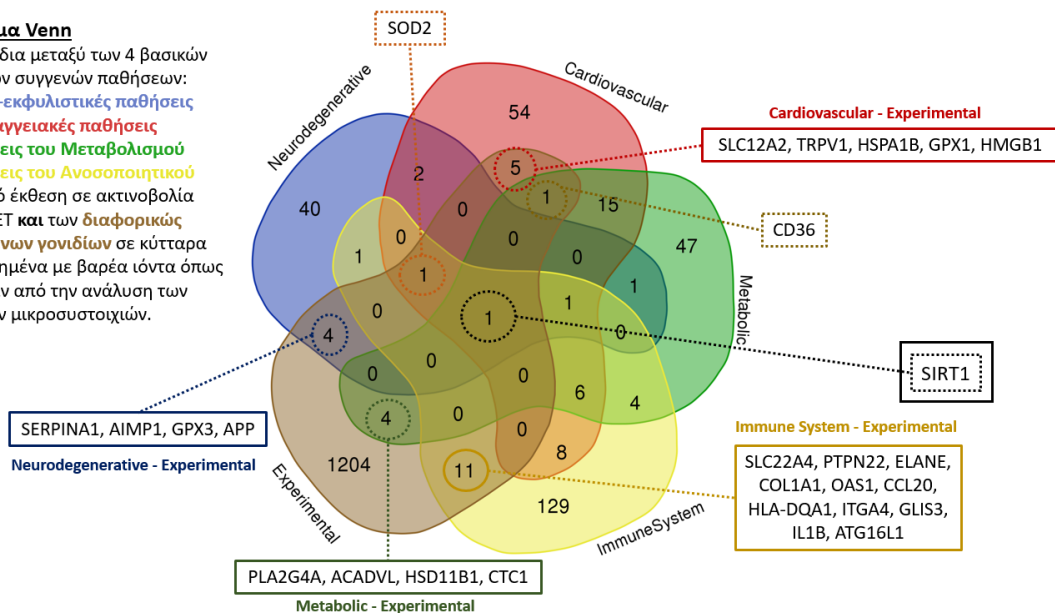
Αρχικά σχεδιάστηκε το διάγραμμα Venn (Εικόνα 58) για τα γονίδια των 4 επιλεγμένων κατηγοριών συγγενών παθήσεων της έκθεσης σε ακτινοβολία υψηλού LET από την αναζήτηση στην πλατφόρμα του DisGeNET, σε συνδυασμό με τα γονίδια που προέκυψαν από την ανάλυση δεδομένων μικροσυστοιχιών από δείγματα εκτεθειμένα σε βαρέα ιόντα.

Διάγραμμα Venn

Κοινά γονίδια μεταξύ των 4 βασικών κατηγοριών συγγενών παθήσεων:

- **Νευρο-εκφυλιστικές παθήσεις**
- **Καρδιαγγειακές παθήσεις**
- **Παθήσεις του Μεταβολισμού**
- **Παθήσεις του Ανοσοποιητικού**

έπειτα από έκθεση σε ακτινοβολία υψηλού LET και των **διαφορικώς εκφρασμένων γονιδίων** σε κύτταρα ακτινοβολημένα με βαρέα ιόντα όπως προέκυψαν από την ανάλυση των δεδομένων μικροσυστοιχιών.



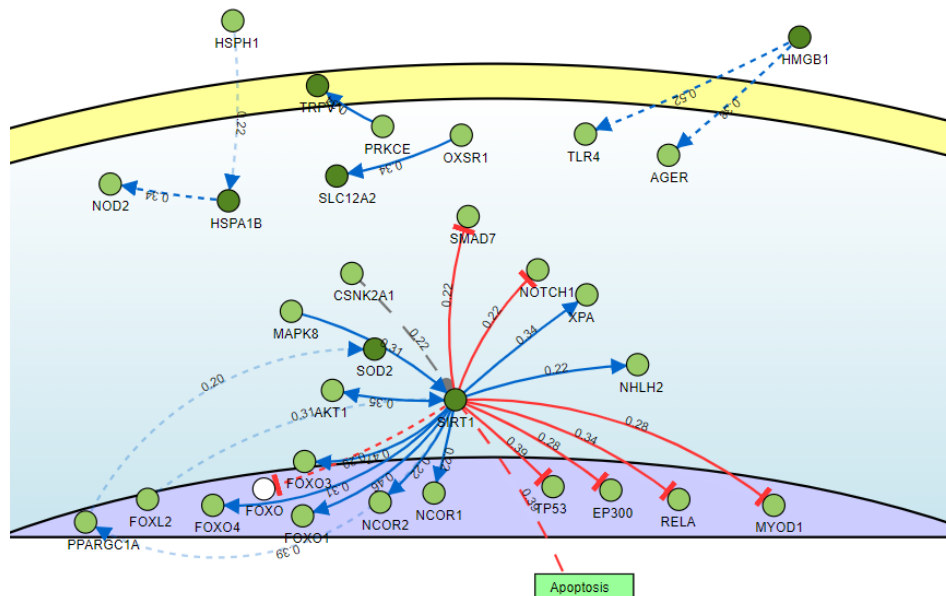
Εικόνα 58: Διάγραμμα Venn των πειραματικών γονιδίων της ανάλυσης προς σύγκριση με τα γονίδια που έχουν συσχετιστεί με τις 4 προς μελέτη κατηγορίες συγγενών παθήσεων: νευροεκφυλιστικές παθήσεις, καρδιαγγειακές παθήσεις, παθήσεις του μεταβολισμού και τέλος παθήσεις του ανοσοποιητικού & αυτοάνοσα νοσήματα. Στα κουτιά παρατίθενται τα ονόματα των γονιδίων που βρέθηκαν κοινά με τα πειραματικά διαφορικώς εκφρασμένα.

Στη συνέχεια χρησιμοποιήθηκαν τα πειραματικά γονίδια, τα οποία βρέθηκαν κοινά με κάθε κατηγορία συγγενών παθήσεων (Πίνακας 24), ως είσοδος στο εργαλείο DISNOR. Σκοπός ήταν να απεικονιστούν οι αιτιώδεις αλληλεπιδράσεις μεταξύ αυτών (Εικόνα 59, Εικόνα 60, Εικόνα 61, Εικόνα 62) και να αναδειχθούν πιθανοί μηχανισμοί που ενεργοποιούνται ή απορρυθμίζονται ως αποτέλεσμα της έκθεσης σε IR υψηλής τιμής LET και που μπορεί να οδηγήσουν στην εκδήλωση παθολογικού φαινοτύπου. Κατά τον σχεδιασμό των δικτύων επιλέχθηκε το μέγιστο επίπεδο πολυπλοκότητας αναζήτησης (level 3).

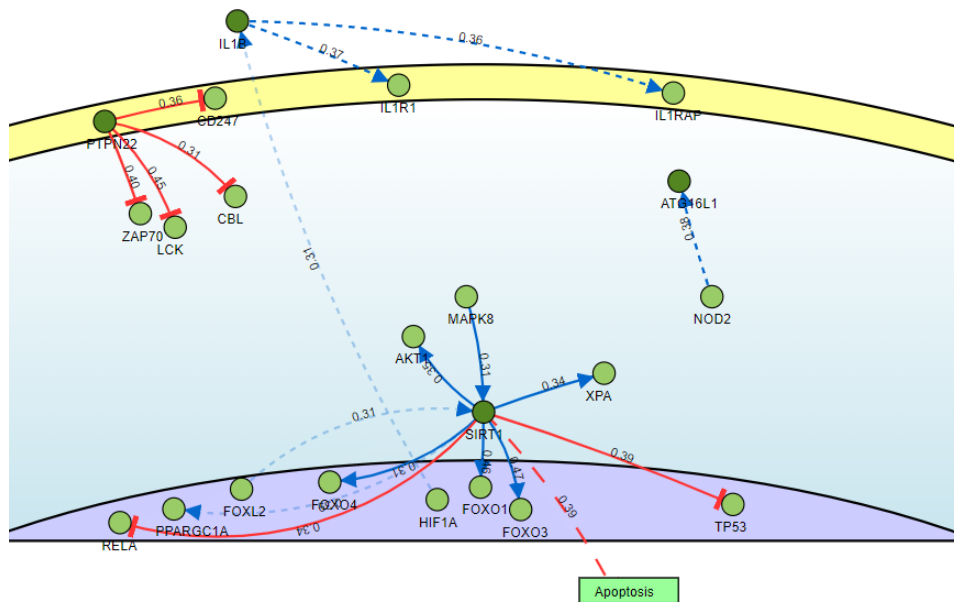
Πίνακας 24: Πειραματικά διαφορικώς εκφρασμένα γονίδια κοινά με την βιβλιογραφία για τις 4 κατηγορίες συγγενών παθήσεων.

Comorbidity	Genes
Immune Disorder	SLC22A4, PTPN22, ELANE, COL1A1, OAS1, CCL20, HLA-DQA1, ITGA4, GLIS3, IL1B, ATG16L1, SIRT1, SOD2
Cardiovascular Disorder	SLC12A2, TRPV1, HSPA1B, GPX1, HMGB1, SOD2, SIRT1, CD36
Neurodegenerative Disorder	SERPINA1, AIMP1, GPX3, APP, SOD2, SIRT1
Metabolic Disorder	PLA2G4A, ACADVL, HSD11B1, CTC1, SIRT1, CD36

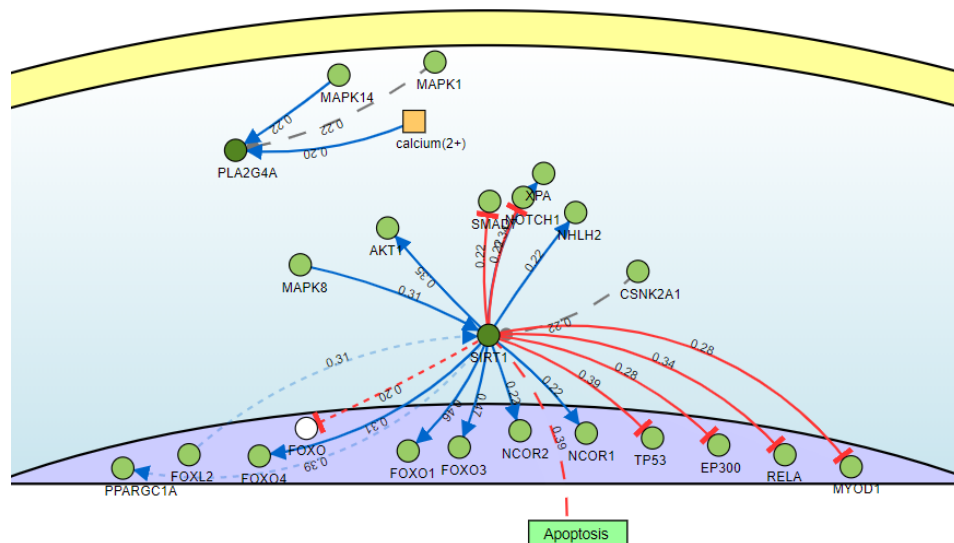
Με το επιλεγόμενο επίπεδο πολυπλοκότητας αναζήτησης, επιτρέπεται η εισαγωγή αλληλεπιδράσεων σηματοδότησης που αφορούν σε οποιαδήποτε από τις πρωτεΐνες που αντιστοιχούν στα γονίδια εισόδου αλλά και σε όλες τις υπόλοιπες πρωτεΐνες με τις οποίες μπορεί να υπάρξει αλληλεπίδραση, σύμφωνα με την βάση δεδομένων SIGNOR για την οργάνωση και απεικόνιση σηματοδοτικών αλληλεπιδράσεων μεταξύ βιολογικών οντοτήτων [153, 154]. Για την βέλτιστη απεικόνιση των δικτύων, χρησιμοποιήθηκε ως μέθοδος φιλτραρίσματος μια αυστηρότερη επιλογή στο σκορ αλληλεπίδρασης (0.2-0.3), που εξαρτάται από την πιθανότητα να αναφερθούν δύο πρωτεΐνες στο ίδιο άρθρο.



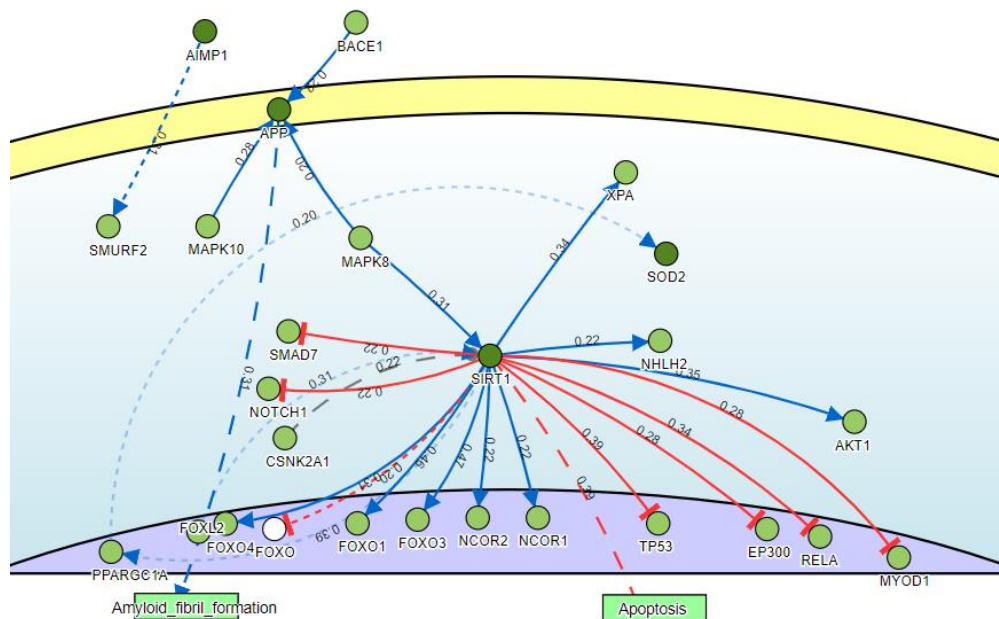
Εικόνα 59: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: SLC12A2, TRPV1, HSPA1B, GPX1, HMGB1, SOD2, SIRT1, CD36, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με καρδιαγγειακές παθήσεις.



Εικόνα 60: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: SLC22A4, PTPN22, ELANE, COL1A1, OAS1, CCL20, HLA-DQA1, ITGA4, GLIS3, IL1B, ATG16L1, SIRT1, SOD2, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με παθήσεις του ανοσοποιητικού.



Εικόνα 61: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: PLA2G4A, ACADVL, HSD11B1, CTC1, SIRT1, CD36, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με παθήσεις του μεταβολισμού.



Εικόνα 62: Δίκτυο αιτιωδών αλληλεπιδράσεων με είσοδο τις εγγραφές: *SERPINA1*, *AIMP1*, *GPX3*, *APP*, *SOD2*, *SIRT1*, οι οποίες αντιστοιχούν στα γονίδια που συσχετίστηκαν με νευροεκφυλιστικές παθήσεις

4. ΣΥΖΗΤΗΣΗ

Κατά την διαδικασία ανάλυσης δεδομένων μικροσυστοιχιών αναπτύχθηκε μία προσεκτική και λεπτομερής ροή εργασίας που μπορεί να εφαρμοστεί σε πληθώρα δεδομένων έκφρασης από μελέτες που χρησιμοποιούν μικροσυστοιχίες τεχνολογιών Illumina και Affymetrix. Λαμβάνονται υπόψιν οι ιδιαιτερότητες κάθε πλατφόρμας και οι διαδικασίες σχολιασμού των ανιχνευτών σε κάθε περίπτωση. Επιπλέον, προτείνονται μέθοδοι φιλτραρίσματος των δεδομένων βασισμένοι τόσο στα τεχνικά χαρακτηριστικά, όσο και στις βιολογικές παραμέτρους του εκάστοτε πειράματος. Τέλος παρουσιάζονται διαδικασίες επισκόπησης της ανάλυσης, με τις αντίστοιχες ερμηνείες, προς την βέλτιστη ανάλυση διαφορικής έκφρασης γονιδίων.

Συγκρίνοντας το δύο εργαλεία που χρησιμοποιήθηκαν για ανάλυση λειτουργικού εμπλουτισμού, το WebGestalt, παρότι διασυνδέεται με λιγότερες βάσεις δεδομένων, έχει πιο φιλική προς το χρήστη διεπαφή. Επιπλέον επιτρέπει την μεταβολή ορισμένων παραμέτρων της ανάλυσης εμπλουτισμού (π.χ. μέγιστος και ελάχιστος αριθμός γονιδίων ανά κατηγορία, μέθοδος διόρθωσης πολλαπλών ερωτημάτων, επίπεδο στατιστικής σημαντικότητας), ώστε ο χρήστης να μπορεί να εξαγάγει τα δικά του συμπεράσματα. Ένα αρνητικό στοιχείο της αναβαθμισμένης έκδοσης του εργαλείου [82, 136] είναι ότι δεν επιτρέπει την εξαγωγή των δέντρων γονιδιακής οντολογίας σε μορφή εικόνας. Στις 3 μεθόδους ανάλυσης που παρέχει το εργαλείο, εμπεριέχεται η GSEA που δίνει την δυνατότητα κατάταξης των γονιδίων εισαγωγής μέσω ποσοτικής μεταβλητής (\log_2FC). Αντίθετα το Enrichr επιδιώκει μια διαδικασία κατάταξης ενσωματωμένη στον αλγόριθμο που χρησιμοποιεί, καταλήγοντας σε ένα *combined score*. Παρατηρήθηκε ότι σε περιπτώσεις που η λίστα γονιδίων εισαγωγής ήταν μεγάλη, ο αριθμός των εμπλουτισμένων όρων ήταν σημαντικά μικρότερος. Το γεγονός αυτό πιθανά υποδηλώνει κάποια αδυναμία του αλγορίθμου, η οποία πρέπει να ληφθεί υπόψιν κατά την βιολογική ερμηνεία των αποτελεσμάτων.

Επιδιώκοντας να προσδώσουμε μία βιολογική ερμηνεία στα αποτελέσματα ανάλυσης λειτουργικού εμπλουτισμού εντός των σειρών δεδομένων,

διαπιστώθηκαν σημαντικές διαφορές στην απόκριση έπειτα από έκθεση σε υψηλής και χαμηλής τιμής LET IR. Βασικό εύρημα αποτελεί η ενισχυμένη απόκριση του ανοσοποιητικού συστήματος στην περίπτωση της σωματιδιακής ακτινοβολίας, κάτι που πιθανότατα σχετίζεται με την αρχική επαγωγή πολύ σύνθετων βλαβών σε κυτταρικό επίπεδο [155]. Επιπλέον, υπάρχουν ενδείξεις απορρύθμισης βιολογικών μονοπατιών, οι οποίες μπορεί να οδηγήσουν στον σχηματισμό νευροεκφυλιστικών και καρδιαγγειακών παθήσεων.

Πιο συγκεκριμένα, στην περίπτωση βρογχικών επιθηλιακών κυττάρων εκτεθειμένων σε 1Gy IR και συλλεγμένων 24 ώρες έπειτα από την ακτινοβόληση (Πίνακας 9) διαφαίνεται η ενεργοποίηση αποπτωτικών μηχανισμών και μηχανισμών φλεγμονής, μέσω ενεργοποίησης σηματοδοτικών μονοπατιών NF-κB, IL-17 και TNF [131]. Παρόλα αυτά, στην περίπτωση έκθεσης σε ιόντα σιδήρου παρατηρούνται επιπλέον ευρήματα σε σύγκριση με την έκθεση σε ακτίνες-X, ενδεικτικά ανοσοαπόκρισης, όπως η ενεργοποίηση των T-λεμφοκυττάρων. Επίσης υποδηλώνεται η παύση του κυτταρικού κύκλου στη μετάβαση G1/S, με ταυτόχρονη ενεργοποίηση της ογκοκατασταλτικής πρωτεΐνης TP53 και διαφόρων μηχανισμών επιδιόρθωσης βλαβών DNA (π.χ. BER). Τα σηματοδοτικά μονοπάτια CCKR και ERK2-ERK2 [133, 156] υποδεικνύουν διαδικασίες απόκρισης σε στρες και μηχανισμούς ελέγχου αποφάσεων κυτταρικού πολλαπλασιασμού ή θανάτου. Τέλος, το μονοπάτι JAK-STAT που φαίνεται να ενεργοποιείται και σχετίζεται με τα παραπάνω, λαμβάνει επιπρόσθετα ρόλο σε ανοσολογικές διαδικασίες και στον σχηματισμό καρκίνου [157].

Στην περίπτωση ενδοθηλιακών κυττάρων στεφανιαίας αρτηρίας εκτεθειμένων σε 2Gy IR και συλλεγμένων 24 ώρες έπειτα από την ακτινοβόληση (Πίνακας 10), οι αποκρίσεις και στις δύο περιπτώσεις (ακτίνες-X και ιόντα σιδήρου) υποδεικνύουν σημεία ελέγχου του κυτταρικού κύκλου, απόκριση σε βλάβες DNA, επιδιορθωτικούς μηχανισμούς και ενεργοποίηση της TP53. Μόνο στην περίπτωση έκθεσης σε ιόντα σιδήρου φαίνεται να ενεργοποιείται η ανοσοαπόκριση, μέσω ενεργοποίησης T-λεμφοκυττάρων και παραγωγής ιντερλευκίνης-1. Παράλληλα, η ενεργοποίηση των Toll-like και NOD-like υποδοχέων στηρίζει περαιτέρω

αυτή την πρόταση [131]. Εύρημα προς συζήτηση αποτελεί το «*focal adhesion*». Πρόκειται για μία βασική διαδικασία μέσω της οποίας τα κύτταρα αλληλεπιδρούν και επικοινωνούν με το περιβάλλον και σχετίζεται με τα αποτελέσματα παραγωγής ιντεγκρίνης και «RHO GTPases Effectors» [158]. Έχει παρατηρηθεί ότι η υπερ-έκφραση της σχετιζόμενης κινάσης FAK αποδίδει καταστάσεις που εμποδίζουν την απόπτωση [159]. Η συνθήκη αυτή αποτελεί παράγοντα κινδύνου στην περίπτωση κακώς ή/και μη επιδιορθωμένων βλαβών, αυξάνοντας την πιθανότητα διαιώνισης της λανθασμένης αλληλουχίας. Τα αποτελέσματα επίσης υποδεικνύουν την απορρύθμιση των σηματοδοτικών μονοπατιών ασβεστίου και Wnt πρωτεϊνών που έχουν συσχετιστεί με νευροεκφυλιστικές παθήσεις [160]. Τέλος, το αποτέλεσμα «Dilated Cardiomyopathy» (διατατική μυοκαρδιοπάθεια) σε συνδυασμό με την ενεργοποίηση του «NOTCH signaling pathway» μπορεί να αποτελέσει ένδειξη για τον αυξημένο κίνδυνο εμφάνισης καρδιαγγειακών νοσημάτων και ειδικότερα ανάπτυξης της CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy) [161].

Η μελέτη της έκθεσης κυττάρων περιφερειακού αίματος σε 1Gy ακτίνων-X και συλλεγμένων 8 ώρες έπειτα από την ακτινοβολήση (Πίνακας 11) απέδωσε αναμενόμενα αποτελέσματα σχετικά με την απόκριση σε IR όπως σημεία ελέγχου του κυτταρικού κύκλου, ενεργοποίηση της TP53 και επαγωγή επιδιορθωτικών μηχανισμών (π.χ. NER). Επιπλέον, προέκυψαν και αποτελέσματα ενδεικτικά του σχηματισμού καρκίνου. Αντίθετα, στην περίπτωση έκθεσης σε ιόντα σιδήρου και άνθρακα διαφαίνεται έντονα η παρουσία ανοσοαπόκρισης μέσω ενεργοποίησης των μονοπατιών ιντερλευκινών, TNF, αντιγονοπαρουσίασης, ανοσίας προκαλούμενης από ουδετερόφιλα κ.α. (Πίνακας 11). Επιπλέον, το αποτέλεσμα «Senescence and Autophagy in Cancer» υποδηλώνει την παρεμπόδιση αποπτωτικών μηχανισμών που τελικά οδηγεί στην κυτταρική γήρανση [162]. Η εμφάνιση ενεργοποίησης των β -1, β -2 και β -3 αδρενεργικών υποδοχέων (beta-1,2,3 adrenergic receptors) σχετίζεται με την λειτουργία του μυοκαρδίου [133], οι οποίοι μελετώνται εκτενώς για τον ρόλο τους στην καρδιακή ανεπάρκεια [163]. Τέλος, το αποτέλεσμα «Amyloid Fiber Formation» σε συνδυασμό με το «Corticotropin releasing factor receptor signaling pathway»

υποδεικνύουν πιθανότητα ανάπτυξης νευροεκφυλιστικών παθήσεων και έχουν συσχετιστεί με τις νόσους του Αλτσχάιμερ (Alzheimer's), του Πάρκινσον (Parkinson's) και του Χάντιγκτον (Huntington's) [164, 165].

Τα διαγράμματα ποσοτικοποιημένης έκφρασης συνάδουν με τα αποτελέσματα της ανάλυσης εμπλουτισμού για την σύγκριση των επιδράσεων χαμηλής και υψηλής τιμής LET IR. Στην περίπτωση των βρογχικών επιθηλιακών κυττάρων (Εικόνα 49), μόνο έπειτα από έκθεση σε ιόντα σιδήρου εντοπίζεται υπερ-έκφραση των γονιδίων MSH6, POLE και POLD3, τα οποία σχετίζονται με τους μηχανισμούς επιδιόρθωσης DNA MMR και NER αντίστοιχα [84]. Επιπλέον, το PRKDC που σχετίζεται με την σηματοδότηση των κινασών ERK [166], εμφανίζεται διπλασίως υπερ-εκφρασμένο έπειτα από έκθεση σε HZE συγκριτικά με την έκθεση σε ακτίνες-Χ. Στα ενδοθηλιακά κύτταρα στεφανιαίας αρτηρίας (Εικόνα 47) εμφανίζεται υπο-έκφραση της πλειοψηφίας των γονιδίων που σχετίζονται με επιδιορθωτικούς μηχανισμούς έπειτα από έκθεση σε ακτίνες-Χ. Αξιοσημείωτη είναι και η υπερ-έκφραση του VAMP8, το οποίο σχετίζεται με την έμφυτη ανοσοαπόκριση [166] και βρίσκεται υπο-εκφρασμένο στην περίπτωση των ακτίνων-Χ. Τέλος, στα κύτταρα περιφερειακού αίματος (Εικόνα 48) οι διαφορές στην έκφραση έπειτα από έκθεση σε ιόντα σιδήρου και άνθρακα είναι παρόμοιες. Σημαντική είναι και η έντονη υπερ-έκφραση ($\log_2FC \approx 1.5$) του C3 στις περιπτώσεις έκθεσης σε σωματιδιακή ακτινοβολία, που σχετίζεται με διαδικασίες του ανοσοποιητικού συστήματος [34].

Ο σχεδιασμός των διαγραμμάτων Venn έγινε με σκοπό την ομαδοποίηση των διαφορετικώς εκφρασμένων γονιδίων προς την κατεύθυνση μίας συνολικότερης εικόνας της επίδρασης της ακτινοβολίας στον ανθρώπινο ιστό. Παρόλα αυτά, μέσω των διαγραμμάτων αυτών εξήχθησαν ορισμένα ενδιαφέροντα συμπεράσματα. Αρχικά, το προφίλ διαφορετικής έκφρασης μεταξύ κυττάρων αίματος και στεφανιαίας αρτηρίας, έπειτα από έκθεση σε ακτίνες-Χ, φαίνεται κοινό και κατά πολύ διαφορετικό από εκείνο των βρογχικών επιθηλιακών κυττάρων (Εικόνα 50). Κάτι ανάλογο συμβαίνει και έπειτα από έκθεση σε σωματίδια-α, όπου εμφανίζεται μεγαλύτερη ομοιότητα μεταξύ ινοβλαστών πνεύμονα και πνευμονικών ενδοθηλιακών

κυττάρων και ταυτόχρονη απόκλιση στην περίπτωση κυττάρων περιφερειακού αίματος (Εικόνα 51). Συγκρίνοντας τα διαφορετικώς εκφρασμένα γονίδια εντός των μελετών που περιέχουν δείγματα ακτινοβολημένα με IR υψηλής και χαμηλής LET διαπιστώθηκε μεγάλος αριθμός κοινών γονιδίων στις περιπτώσεις συλλογής 12-24 ωρών έπειτα από την ακτινοβολήση (Εικόνα 54, Εικόνα 55). Αντίθετα, στην περίπτωση συλλογής των δειγμάτων 8 ώρες μετά την ακτινοβολήση (Εικόνα 53), τα διαφορετικώς εκφρασμένα γονίδια εξαιτίας της έκθεσης σε ακτίνες-Χ είναι πολύ λιγότερα στο σύνολο, με 42 μόνο κοινά με εκείνα της έκθεσης σε βαρέα ιόντα. Κάτι τέτοιο πιθανά σχετίζεται με τις διαφορές τις όψιμες επιδράσεις των ακτινοβολιών χαμηλής και υψηλής LET αντίστοιχα.

Από τα αποτελέσματα ανάλυσης εμπλουτισμού με την μέθοδο ORA, εφαρμοσμένη σε ομάδες γονιδίων που προέκυψαν από τα διαγράμματα Venn, αναμένουμε μια συνολική και συνοπτική εικόνα των επιδράσεων της IR. Στην περίπτωση έκθεσης σε υψηλές δόσεις ακτίνων-Χ (Πίνακας 13), παρατηρούμε αναμενόμενα αποτελέσματα απόκρισης βλαβών DNA, σημείων ελέγχου του κυτταρικού κύκλου και ενεργοποίηση της TP53. Αξιοσημείωτο είναι το αποτέλεσμα «RHO GTPases Effectors» έπειτα από έκθεση σε υψηλές δόσεις σωματιδίων-α, 24 ώρες μετά την ακτινοβολήση (Πίνακας 14), που σχετίζεται με διαδικασίες αντι-απόπτωσης, όπως προαναφέρθηκε [159]. Τα αποτελέσματα της έκθεσης απομονωμένων λευκοκυττάρων σε βαρέα ιόντα (Πίνακας 15) υποδηλώνουν επίσης ενεργοποίηση μηχανισμών απόκρισης βλαβών DNA αλλά και ύπαρξη γενεοτοξικού στρες. Στη περίπτωση υψηλών δόσεων εμφανίζεται επιπρόσθετα εμπλουτισμός όρων σχετιζόμενων με την ανοσοαπόκριση και τον σχηματισμό καρκίνου μέσω των όρων «Chronic Myeloid Leukemia» και «Non-small Cell Lung Carcinoma». Στα βρογχικά επιθηλιακά κύτταρα (Πίνακας 16) παρατηρείται επαγωγή επιδιόρθωσης DNA, μόνο στην περίπτωση υψηλών δόσεων. Τέλος, η ανάλυση εμπλουτισμού στις συγκεντρωτικές λίστες διαφορετικώς εκφρασμένων γονιδίων από έκθεση σε βαρέα ιόντα (Εικόνα 57) επιβεβαιώνει την αυξημένη ενεργοποίηση ανοσολογικών μηχανισμών, ειδικά στην περίπτωση των υψηλών δόσεων (Πίνακας 17). Πέρα από τις ενδείξεις επαγωγής φλεγμονώδους περιβάλλοντος, διακρίνεται η ενεργοποίηση διαδικασιών επίκτητης

ανοσίας, η ενεργοποίηση της δράσης ουδετερόφιλων και του υποδοχέα TCR (T-Cell antigen Receptor), ο οποίος αποτελεί τον μηχανισμό ανίχνευσης αντιγόνων από τα Τα-λεμφοκύτταρα [167] και σχετίζεται με μεταβολές στην έκφραση του μεταγραφικού παράγοντα NF-κΒ [42, 132].

Τα αποτελέσματα της ανάλυσης εμπλουτισμού για ασθένειες και παθήσεις καταδεικνύουν πολλαπλώς τον σχηματισμό καρκίνου στους πνεύμονες, το ήπαρ, το πάγκρεας και τον μαστό. Παράλληλα στις περιπτώσεις έκθεσης σε ακτινοβολία υψηλής LET, υπερισχύουν τα αποτελέσματα σχετιζόμενα με λευχαιμία, λεμφώματα και παθήσεις του εγκεφάλου, όπως το νευροβλάστωμα και το γλοιοβλάστωμα (Πίνακας 18, Πίνακας 19, Πίνακας 20). Για την άμεση σύγκριση παθήσεων που μπορεί να προκύψουν από υψηλής και χαμηλής LET IR (Πίνακας 21, Πίνακας 22, Πίνακας 23) τα αποτελέσματα δεν είναι αξιολογήσιμα, καθώς προκύπτει ότι το εργαλείο Enrichr δεν αποδίδει ικανοποιητικά για μεγάλες λίστες γονιδίων.

Τα γονίδια από την ανάλυση δεδομένων μικροσυστοιχιών, που βρέθηκαν κοινά με εκείνα από την βιβλιογραφική αναζήτηση για συγγενείς παθήσεις στο DisGeNET (Εικόνα 58), πιθανά να μπορούν να χρησιμοποιηθούν ως βιοδείκτες για την εμφάνιση ασθενειών και παθήσεων. Το γονίδιο SIRT1 που εντοπίστηκε σε όλες τις κατηγορίες νοσημάτων, σχετίζεται με διαδικασίες γήρανσης [166] ενώ παράλληλα έχει βρεθεί να εμπλέκεται σε παθήσεις του μεταβολισμού όπως η παχυσαρκία, σε καρδιαγγειακές και νευροεκφυλιστικές παθήσεις καθώς και στον σχηματισμό καρκίνου [168]. Στα δίκτυα πρωτεϊνικών αλληλεπιδράσεων από το DISNOR, για κάθε κατηγορία συγγενών παθήσεων της ακτινοβολίας, εντοπίζεται ένα τμήμα με κεντρικό κόμβο το SIRT1. Φαίνεται ότι το γονίδιο αυτό προκαλεί υπο-έκφραση των TP53 και NOTCH1 καταλήγοντας στην απενεργοποίηση μηχανισμών απόπτωσης. Στο δίκτυο που συνθέτουν τα γονίδια σχετιζόμενα με καρδιαγγειακές παθήσεις (Εικόνα 59), το PRKCE που προκαλεί την υπερ-έκφραση του TRPV1, σχετίζεται με την ισχαιμία [169]. Στην περίπτωση των παθήσεων του ανοσοποιητικού (Εικόνα 60), σημαντική είναι η παρουσία του PTPN22, υπεύθυνο για την υπο-έκφραση των CD247, CBL, LCK και ZAP70. Το σύμπλεγμα αυτών των γονιδίων πιθανά υποδηλώνει την ανάπτυξη αυτοάνοσων νοσημάτων όπως η

ανοσοανεπάρκεια, ο συστηματικός ερυθρηματώδης λύκος, η ρευματοειδής αρθρίτιδα και η αυτοάνοση μυοκαρδιοπάθεια [166]. Από τα γονίδια που σχετίζονται με παθήσεις του μεταβολισμού, το PLA2G4A, ως μέλος της οικογένειας ενζύμων A2-φωσφολιπάσης, εξυπηρετεί σημαντικό ρόλο στη δημιουργία δραστικών μεταβολιτών λιπιδίων που προάγουν φλεγμονώδεις μεταβολικές παθήσεις όπως η αθηροσκλήρωση, η υπερλιπιδαιμία, η παχυσαρκία και ο διαβήτης [170]. Τέλος, στο δίκτυο των σχετιζόμενων με νευροεκφυλιστικές παθήσεις γονιδίων, παρατηρούμε την ενεργοποίηση σχηματισμού αμυλοειδών ινιδίων μέσω του APP που σχετίζεται με την εγκεφαλική αμυλοειδή αγγειοπάθεια και τη νόσο του Αλτσχάιμερ [166].

Τα αποτελέσματα μας συνοψίζονται στο ότι οι βασικές διαφορές στις επιδράσεις των σωματιδιακών ακτινοβολιών συνδέονται άμεσα με την επαγωγή της επίκτητης ανοσοαπόκρισης. Το εύρημα αυτό μπορεί να αποτελεί μέσο για τον σχηματισμό επίμονης φλεγμονής, γονιδιωματικής αστάθειας και ανάπτυξης αυτοάνοσων νοσημάτων [171]. Επιπλέον, διαφαίνεται μέσω όλων των προσεγγίσεων ότι υπάρχει διασύνδεση με τις τέσσερις κατηγορίες συγγενών παθήσεων που στηρίζονται από την βιβλιογραφία [172]. Γενικότερα πρέπει να τονιστεί η σημαντικότητα της περαιτέρω μελέτης των αποτελεσμάτων έκθεσης σε υψηλής LET IR σε υγιείς ιστούς, τόσο για σκοπούς βελτιστοποίησης των μεθόδων ακτινοθεραπείας, όσο για την προστασία των αστροναυτών σε διαστημικές αποστολές μεγάλης διάρκειας.

5. ΠΑΡΑΡΤΗΜΑ

Πίνακας 25: Συντομογραφίες που αναφέρθηκαν σε αυτή την εργασία.

IR	Ionizing Radiation
HZE	High atomic number (Z) Energy (E)
LET	Linear Energy Transfer
RBE	Relative Biological Effectiveness
ROS	Reactive Oxygen Species
SSB	Single-Strand Break
DSB	Double Strand Break
DDR	DNA Damage Response
BER	Base Excision Repair
NER	Nucleotide Excision Repair
MMR	MisMatch Repair
HR	Homologous Recombination
NHEJ	Non-Homologous End Joining
NASA	National Aeronautics & Space Administration
CNS	Central Nervous System
RICVD	Radiation-Induced CardioVascular Disease
PCA	Principal Component Analysis
MDS	Multi-Dimensional Scaling
FDR	False Discovery Rate
log₂FC	log ₂ fold-change
GO	Gene Ontology
GEO	Gene Expression Omnibus
MIAME	Minimal Information About Microarray Experiments
SDRF	Sample and Data Relationship Format
CDF	Common Data Format
RMA	Robust Multi-array Average
HGNC	HUGO Gene Nomenclature Committee
GSEA	Gene Set Enrichment Analysis
ORA	Over-Representation Analysis
CADASIL	Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
TCR	T-Cell antigen Receptor

Πίνακας 26: Παράθεση των ενδεικτικών γονιδίων που επιλέχθηκαν προς ποσοτική μελέτη στις σειρές δεδομένων μικροσυστοιχιών, ώστε να αντιπροσωπευθούν οι διεργασίες επιδιόρθωσης DNA και ανοσοποιητικού συστήματος.

HGNC ID	Gene Name
Immune Response (GO:0006955)	
APP	amyloid beta precursor protein
B2M	beta-2-microglobulin
C3	complement C3
FYN	FYN proto-oncogene, Src family tyrosine kinase
IRAK1	interleukin 1 receptor associated kinase 1
IRF3	interferon regulatory factor 3
MYD88	MYD88 innate immune signal transduction adaptor
RELA	RELA proto-oncogene, NF-kB subunit
STAT1	signal transducer and activator of transcription 1
VAMP8	vesicle associated membrane protein 8
DNA Repair (GO:0006281)	
UBE2N	ubiquitin conjugating enzyme E2 N
TP53	tumor protein p53
BRCA1	BRCA1 DNA repair associated
CDK9	cyclin dependent kinase 9
ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit
ERCC3	ERCC excision repair 3, TFIIH core complex helicase subunit
ERCC8	ERCC excision repair 8, CSA ubiquitin ligase complex subunit
FAN1	FANCD2 and FANCI associated nuclease 1
LIG4	DNA ligase 4
MSH2	mutS homolog 2
MSH6	mutS homolog 6
NHEJ1	non-homologous end joining factor 1
PCNA	proliferating cell nuclear antigen
POLD3	DNA polymerase delta 3, accessory subunit
POLE	DNA polymerase epsilon, catalytic subunit
POLK	DNA polymerase kappa
PRKDC	protein kinase, DNA-activated, catalytic subunit
RAD50	RAD50 double strand break repair protein
RPA1	replication protein A1
RPA2	replication protein A2
RPA3	replication protein A3
SWSAP1	SWIM-type zinc finger 7 associated protein 1
XPC	XPC complex subunit, DNA damage recognition and repair factor

Πίνακας 27: Όροι και κωδικοί των ασθενειών από το DisGeNET για τις τέσσερις βασικές κατηγορίες συγγενών παθήσεων που έχουν συσχετιστεί με έκθεση σε σωματιδιακή ακτινοβολία.

DisGeNET Accession	Disease Term
Immune-Inflammatory	
C0004364	Autoimmune Diseases
C0021053	Immune System Diseases
C4021753	Abnormality of Immunhe System
C2976853	Immune Reconstitution Inflammatory Syndrome
C0020951	Immune Complex Diseases
C1844666	Immune Dysregulation
C1290884	Inflammatory Disorder
C1290886	Chronic Inflammatory Disorder
C3860213	Autoinflammatory Disorder
C3267073	Autoinflammatory Disease
C3890737	Autoinflammatory Syndrome
(Auto)Immune-Disease	
C0398650	Immune Thrombocytopenic Purpura
C0178468	Autoimmune Thyroid Disease
C0205734	Diabetes, Autoimmune
C0242584	Autoimmune Thrombocytopenia
C0920350	Autoimmune Thyroiditis
C3160909	Autoimmune Arthritis
C0002880	Autoimmune Hemolytic Anemia
C1328840	Autoimmune Lymphoproliferative Syndrome
C3887639	Autoimmune Gastritis
C2609129	Autoimmune Pancreatitis
C0400836	Autoimmune Liver Disease
C3494489	Autoimmune Polyendocrinopathy Syndrome
C0340971	Autoimmune Neutropenia
C0341305	Autoimmune Enteropathy
C0406632	Autoimmune Skin Disease
C1328843	Autoimmune Vasculitis
C0021390	Inflammatory Bowel Disease
C0751356	Idiopathic Inflammatory Myopathies
C0683381	Inflammatory Joint Disease
C0333355	Inflammatory Disease of Mucous Membranes
C0270933	Inflammatory Neuropathy
C3532222	Inflammatory Cardiomyopathy
Metabolic	
C0025517	Metabolic Diseases
C0005944	Metabolic Bone Disorder

C0006112	Brain Disorder, Metabolic
C0750968	CNS Metabolic Disorder
C0751743	Metabolic Disorder, CNS, Acquired
C0751744	Brain Disease, Metabolic, Acquired
C1135773	Acquired Metabolic Disease, Nervous System
C0524620	Metabolic Syndrome X

Cardiovascular

C0007222	Cardiovascular Diseases
C0741949	Cardiovascular Pathology
C3665365	Arteriosclerotic Cardiovascular Disease
C0243050	Cardiovascular Abnormalities
C0497243	Neoplasm of Cardiovascular System
C4072686	Cardiovascular & Pulmonary Disease
C0004153	Atherosclerotic Cardiovascular Disease

Neurodegenerative

C0524851	Neurodegenerative Disorders
C0270715	CNS Neurodegenerative Disorder

Πίνακας 28: Γονίδια που αντιστοιχούν σε κάθε κατηγορία συγγενών παθήσεων της έκθεσης σε σωματιδιακή ακτινοβολία από το DisGeNET.

Neuro-degenerative	Cardiovascular	Metabolic	Immune-Inflammatory	(Auto)Immune Disease
PANK2	LDLR	LEP	IL6	FAS
TTC19	AGER	NOS3	STAT3	FASLG
APP	AGT	SIRT1	PTPN22	IL10
MAPT	HP	IL18	FAS	TNF
ATXN1	SIRT1	PPARG	CD28	AIRE
SNCA	APOE	CRP	CD40LG	INS
SIRT1	CYP27A1	IL6	FCGR2B	FCGR2C
PSEN1	ABCA1	CCL2	IFNG	TNFSF15
GDNF	ALOX5	SHBG	IL4	ATG16L1
HMOX1	ALOX15	ADIPOQ	IL17A	NOD2
IL6	CETP	RETN	TG	IL23R
NGF	CRP	PON1	TNFAIP3	IRGM
CRYAB	CYBA	MTTP	TNFRSF1A	IL1B
FTL	ICAM1	HTR2C	TPO	CARD9
HCRT	IFNG	CRP	SIRT1	IL18RAP
EPO	IL6	INPPL1	C4B	IL6
GSTM1	LPA	INS	AICDA	PLCG2
SOD2	NOS3	ESRRA	CXCR3	PTGS2
GSTO1	SERPINE1	ETFDH	SIAE	STAT3
APOD	PLA2G2A	LBP	PECAM1	MUC19
APP	PON1	NR1I2	HARS	TG
CRYAB	PPARG	ALPL	HP	ICAM1
EPO	PTGS2	INS	TIMP1	APC
FTH1	CCL2	HMGA1	TNFRSF4	PTPN22
GDNF	TLR4	TRIB3	B2M	SLC11A1
GPX3	TNF	NEIL1	EGR1	IL2RA
GSR	VCAM1	CYP27A1	FBL	HLA-DRB1
GSTM1	VEGFA	DHCR7	F8	CASP10
GSTM2	ADIPOQ	ESR1	RBP3	TGFB1
GSTM4	AGT	GC	C1S	IL10
GSTM5	ALB	IGF1	COL1A1	PTPN22
HCRT	APOB	KCNMA1	COPA	TSHR
HMOX1	APOC3	LRP2	CRYAA	DEFA5
IL6	APOE	MMP2	CRYBB2	HLA-DQB1
IDO1	CETP	SLC20A1	HMGCR	NRAS
MAPT	CRP	ANKRD11	HPX	ITGA4
MGST1	ACE	CTC1	HSPA9	SFRP1
NGF	HP	CYP2R1	ITGB6	SFRP2
NGFR	ICAM1	PHOSPHO1	KRT19	VNN1
SERPINA1	LPA	CPT2	PF4	APC2
PKD2	LPL	ALDH5A1	RAPSN	INAVA
SEPT5	MTHFR	AHR	SOD2	HP

Neuro-degenerative	Cardiovascular	Metabolic	Immune-Inflammatory	(Auto)Immune Disease
PSEN1	NOS3	DSG1	TGFB2	IFNG
ATXN1	PON1	HTR2A	C1QA	IGF1
SELENOP	PTGS2	PON1	IFNG	IL2RA
SNCA	CCL2	UPB1	MASP2	KCNJ11
SOD2	VWF	APOA1	ARG1	HNF1A
SPTAN1	APOA4	CD36	CNR2	TNF
VIM	MMP1	HSD11B1	POMC	CLEC16A
PDE8B	CBS	SERPINE1	EIF2AK4	IL1RN
KYNU	MPO	PPARG	AIRE	PRKCD
INA	NPPB	TNF	CD19	SLC14A1
GSTO1	VCAM1	NR1I3	CTLA4	ITGAL
SIRT1	ESR1	CPT2	HLA-DPB1	ITGB8
TTC19	CYP2C19	ALDH5A1	HLA-DQB1	CUL2
AGPAT3	EDN1	CPT2	HLA-DRB1	RASSF1
PANK2	ALOX5AP	ALDH5A1	IL2RA	SLAMF8
APOD	GSTM1	CPT2	IRF5	CAT
FTH1	PLAT	ALDH5A1	TNFRSF13B	CD69
GPX3	ADRB1	CPT2	TSSK1B	CHRM2
GSR	CST3	ALDH5A1	IL23R	CP
GSTM2	SOD1	APOC3		CRP
GSTM4	STAT3	IRS1		CTSH
GSTM5	ABCG5	AGT		DDIT3
IDO1	PON2	IGF2		HLA-DQA1
MGST1	SELE	PPARGC1A		HSD11B2
NGFR	CNR2	HMOX1		ITPR3
SERPINA1	NOS2	PLA2G7		NOS1
PKD2	GPX1	KL		NOS3
SELENOP	NPY	EPHX2		OAS1
SPTAN1	AHR	HMGCR		PAX4
VIM	IGF2	PTGS2		PRKCQ
PDE8B	GH1	HSPA8		CCL20
KYNU	SORT1	PTPN11		SLC11A1
INA	LTA4H	ACADVL		STAT3
AGPAT3	CCL4	CYBA		ABCC8
LRRK2	SOD2	G6PD		AIFM1
AIMP1	SOCS3	HP		FGF21
CAT	PARP1	LPL		IL19
FTL	FASLG	PLA2G4A		IL20
ABCD1	HRH1	REG1A		FOXP3
NQO1	HSPA1B	CCL5		SLC29A3
FUS	CCL3	TGFB1		BACH2
ATXN2	GRK5	VCAM1		C1QTNF6
SOD1	AGT	PLA2G6		GLIS3
ATG7	NOS2	FADD		IL27

Neuro-degenerative	Cardiovascular	Metabolic	Immune-Inflammatory	(Auto)Immune Disease
TARDBP	CD36	HDAC3		CASP12
TREM2	CD40	FGF19		CSF2
C9orf72	CD40LG	NAMPT		CSF3
	IL18	CHEK2		IL1RN
	NPC1	FTO		IL10
	SELP	TP53INP1		MTHFR
	TTPA			TFF3
	TRPV1			TGFA
	GRK2			ADAMTS13
	SLC12A2			IGHG1
	TNNT2			IGHG2
	EDN1			IGKC
	FGF8			TARS
	FOXC2			ZFAT
	GATA2			FCRL3
	STRA6			CASP8
	ANKS6			IL2
	AGER			XBP1
	CAV1			IFNG
	HMGB1			RUNX3
	F2			ESR2
	CLU			CD40LG
	CASR			MUC2
	MMP2			NFIL3
				IL4
				TSLP
				IGF2
				KLKB1
				NOS2
				SLC22A4
				SLC6A14
				ELANE
				MKI67
				MUC2
				MUC4
				NOS2
				PCNA
				ULK1
				GJA1
				GJB1
				GJB2

6. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Nenoi, M., *Current Topics in Ionizing Radiation Research*. 2012, Croatia: InTech.
2. UNEP, U.N.E.P., *Radiation Effects and Sources*. 2016, Austria: United Nations Environment Programme.
3. IAEA, I.A.E.A., *Radiation, People and the Environment*. 2004, Austria: Division of Radiation and Waste Safety.
4. Little, M.P., *Risks associated with ionizing radiation*. Br Med Bull, 2003. **68**: p. 259-75.
5. Leroy, C. and P.-G. Rancoita, *Principles of radiation interaction in matter and detection*. 2012, New Jersey: World Scientific.
6. Del Guerra, A., *Ionizing Radiation Detectors for Medical Imaging*. 2004, Singapore: World Scientific Publishing Co. Re. Ltd.
7. Ziegler, J.F., M.D. Ziegler, and B.J. P., *SRIM – The stopping and range of ions in matter*. Science Direct, 2010(Nuclear Instruments and Methods in Physics Research B).
8. Porter, L.E., *Bethe-bloch stopping power parameters for light projectiles at energies near the stopping power maximum*. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 1985. **12**(1): p. 50-55.
9. Kjellberg , R.N., et al., *Bragg-Peak Proton-Beam Therapy for Arteriovenous Malformations of the Brain*. New England Journal of Medicine, 1983. **309**(5): p. 269-274.
10. Bragg, W.H. and R. Kleeman, *LXXIV. On the ionization curves of radium*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1904. **8**(48): p. 726-738.
11. Goodhead, D.T. and H. Nikjoo, *Track Structure Analysis of Ultrasoft X-rays Compared to High- and Low-LET Radiations*. International Journal of Radiation Biology, 1989. **55**(4): p. 513-529.
12. Held, K.D., et al., *Effects of Charged Particles on Human Tumor Cells*. Front Oncol, 2016. **6**: p. 23.
13. Mehta, S.R., et al., *Radiotherapy: Basic Concepts and Recent Advances*. Med J Armed Forces India, 2010. **66**(2): p. 158-62.
14. Keith, S., et al., *Toxicological Profile for Radon*, in *Toxicological Profile for Radon*. 2012: Atlanta (GA).
15. Georgakilas, A.G., *Bystander and non-targeted effects: a unifying model from ionizing radiation to cancer*. Cancer Lett, 2015. **356**(1): p. 3-4.
16. Georgakilas, A.G., P. O'Neill, and R.D. Stewart, *Induction and repair of clustered DNA lesions: what do we know so far?* Radiat Res, 2013. **180**(1): p. 100-9.

17. Riley, P.A., *Free radicals in biology: oxidative stress and the effects of ionizing radiation*. Int J Radiat Biol, 1994. **65**(1): p. 27-33.
18. Wallace, S.S., *Enzymatic processing of radiation-induced free radical damage in DNA*. Radiat Res, 1998. **150**(5 Suppl): p. S60-79.
19. Mavragani, I.V., et al., *Complex DNA Damage: A Route to Radiation-Induced Genomic Instability and Carcinogenesis*. Cancers (Basel), 2017. **9**(7).
20. Nikitaki, Z., et al., *Stress-induced DNA damage biomarkers: applications and limitations*. Front Chem, 2015. **3**: p. 35.
21. Ogrunc, M., et al., *Oncogene-induced reactive oxygen species fuel hyperproliferation and DNA damage response activation*. Cell Death Differ, 2014. **21**(6): p. 998-1012.
22. Georgakilas, A.G., *Processing of DNA damage clusters in human cells: current status of knowledge*. Mol Biosyst, 2008. **4**(1): p. 30-5.
23. Georgakilas, A.G., et al., *Processing of bistranded abasic DNA clusters in gamma-irradiated human hematopoietic cells*. Nucleic Acids Res, 2004. **32**(18): p. 5609-20.
24. Saini, D., et al., *Transcription profile of DNA damage response genes at G(0) lymphocytes exposed to gamma radiation*. Mol Cell Biochem, 2012. **364**(1-2): p. 271-81.
25. Bertoli, C., J.M. Skotheim, and R.A. de Bruin, *Control of cell cycle transcription during G1 and S phases*. Nat Rev Mol Cell Biol, 2013. **14**(8): p. 518-28.
26. Shu, K.X., B. Li, and L.X. Wu, *The p53 network: p53 and its downstream genes*. Colloids Surf B Biointerfaces, 2007. **55**(1): p. 10-8.
27. Mendoza, J., et al., *Association between ERCC1 and XPA expression and polymorphisms and the response to cisplatin in testicular germ cell tumours*. Br J Cancer, 2013. **109**(1): p. 68-75.
28. Harris, S.L. and A.J. Levine, *The p53 pathway: positive and negative feedback loops*. Oncogene, 2005. **24**(17): p. 2899-908.
29. Georgakilas, A.G., O.A. Martin, and W.M. Bonner, *p21: A Two-Faced Genome Guardian*. Trends Mol Med, 2017. **23**(4): p. 310-319.
30. Nakad, R. and B. Schumacher, *DNA Damage Response and Immune Defense: Links and Mechanisms*. Front Genet, 2016. **7**: p. 147.
31. Toprani, S.M. and B. Das, *Role of base excision repair genes and proteins in gamma-irradiated resting human peripheral blood mononuclear cells*. Mutagenesis, 2015. **30**(2): p. 247-61.
32. Knijnenburg, T.A., et al., *Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas*. Cell Rep, 2018. **23**(1): p. 239-254 e6.

33. Friedberg, E.C., *A history of the DNA repair and mutagenesis field: The discovery of base excision repair*. DNA Repair (Amst), 2016. **37**: p. A35-9.
34. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation*. Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
35. Pan, M.R., et al., *Connecting the Dots: From DNA Damage and Repair to Aging*. Int J Mol Sci, 2016. **17**(5).
36. Hatzi, V.I., et al., *Non-targeted radiation effects in vivo: a critical glance of the future in radiobiology*. Cancer Lett, 2015. **356**(1): p. 34-42.
37. Singleton, B.K., C.S. Griffin, and J. Thacker, *Clustered DNA damage leads to complex genetic changes in irradiated human cells*. Cancer Res, 2002. **62**(21): p. 6263-9.
38. Stewart, R.D., et al., *Effects of radiation quality and oxygen on clustered DNA lesions and cell death*. Radiat Res, 2011. **176**(5): p. 587-602.
39. Hada, M. and A.G. Georgakilas, *Formation of clustered DNA damage after high-LET irradiation: a review*. J Radiat Res, 2008. **49**(3): p. 203-10.
40. Hassler, D.M., et al., *Mars' surface radiation environment measured with the Mars Science Laboratory's Curiosity rover*. Science, 2014. **343**(6169): p. 1244797.
41. Fernandez-Gonzalo, R., S. Baatout, and M. Moreels, *Impact of Particle Irradiation on the Immune System: From the Clinic to Mars*. Front Immunol, 2017. **8**: p. 177.
42. Hellweg, C.E., et al., *The Role of the Nuclear Factor kappaB Pathway in the Cellular Response to Low and High Linear Energy Transfer Radiation*. Int J Mol Sci, 2018. **19**(8).
43. Garrett-Bakelman, F.E., et al., *The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight*. Science, 2019. **364**(6436).
44. Choukèr, A., *Stress challenges and immunity in space: From mechanisms to monitoring and preventive strategies*. 2012. 1-469.
45. Makale, M.T., et al., *Mechanisms of radiotherapy-associated cognitive disability in patients with brain tumours*. Nat Rev Neurol, 2017. **13**(1): p. 52-64.
46. Andersen, B.L. and H.H. Tewfik, *Psychological reactions to radiation therapy: reconsideration of the adaptive aspects of anxiety*. J Pers Soc Psychol, 1985. **48**(4): p. 1024-32.
47. Wellisch, D.K., et al., *Predicting major depression in brain tumor patients*. Psychooncology, 2002. **11**(3): p. 230-8.
48. Cucinotta, F.A., *Space radiation risks for astronauts on multiple International Space Station missions*. PLoS One, 2014. **9**(4): p. e96099.

49. Darby, S.C., et al., *Risk of ischemic heart disease in women after radiotherapy for breast cancer*. N Engl J Med, 2013. **368**(11): p. 987-98.
50. Hughson, R.L., A. Helm, and M. Durante, *Heart in space: effect of the extraterrestrial environment on the cardiovascular system*. Nat Rev Cardiol, 2018. **15**(3): p. 167-180.
51. Nylander, V., et al., *Ionizing Radiation Potentiates High-Fat Diet-Induced Insulin Resistance and Reprograms Skeletal Muscle and Adipose Progenitor Cells*. Diabetes, 2016. **65**(12): p. 3573-3584.
52. Azzam, E.I., J.P. Jay-Gerin, and D. Pain, *Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury*. Cancer Lett, 2012. **327**(1-2): p. 48-60.
53. Ina, Y. and K. Sakai, *Activation of immunological network by chronic low-dose-rate irradiation in wild-type mouse strains: analysis of immune cell populations and surface molecules*. Int J Radiat Biol, 2005. **81**(10): p. 721-9.
54. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
55. Schena, M., *Genome analysis with gene expression microarrays*. Bioessays, 1996. **18**(5): p. 427-31.
56. Duggan, D.J., et al., *Expression profiling using cDNA microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 10-4.
57. Lou, X.J., et al., *Expression monitoring using cDNA microarrays. A general protocol*. Methods Mol Biol, 2001. **175**: p. 323-40.
58. Brown, P.O. and D. Botstein, *Exploring the new world of the genome with DNA microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 33-7.
59. Schena, F.P., *[Research and patents in the biomedical field]*. G Ital Nefrol, 2002. **19**(1): p. 2-3.
60. Butte, A., *The use and analysis of microarray data*. Nat Rev Drug Discov, 2002. **1**(12): p. 951-60.
61. Lyng, H., et al., *Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction*. BMC Genomics, 2004. **5**(1): p. 10.
62. Cheung, V.G., et al., *Making and reading microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 15-9.
63. Heller, M.J., *DNA microarray technology: devices, systems, and applications*. Annu Rev Biomed Eng, 2002. **4**: p. 129-53.
64. Yap, G., *Affymetrix, Inc*. Pharmacogenomics, 2002. **3**(5): p. 709-11.
65. Lipshutz, R.J., et al., *High density synthetic oligonucleotide arrays*. Nat Genet, 1999. **21**(1 Suppl): p. 20-4.

66. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. *Bioinformatics*, 2003. **19**(2): p. 185-93.
67. Formann, A.K., *The Newcomb-Benford law in its relation to some common distributions*. *PLoS One*, 2010. **5**(5): p. e10541.
68. Qiu, X., H. Wu, and R. Hu, *The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis*. *BMC Bioinformatics*, 2013. **14**: p. 124.
69. Pagano, M., *The histogram and boxplot: a picture is worth a thousand words*. *Nutrition*, 1992. **8**(5): p. 374-5.
70. Royeen, C.B., *The boxplot: a screening test for research data*. *Am J Occup Ther*, 1986. **40**(8): p. 569-71.
71. Raychaudhuri, S., J.M. Stuart, and R.B. Altman, *Principal components analysis to summarize microarray experiments: application to sporulation time series*. *Pac Symp Biocomput*, 2000: p. 455-66.
72. Chen, Y. and P.S. Meltzer, *Gene expression analysis via multidimensional scaling*. *Curr Protoc Bioinformatics*, 2005. **Chapter 7**: p. Unit 7 11.
73. Townsend, J.P., *Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays*. *BMC Bioinformatics*, 2004. **5**: p. 54.
74. Dudoit, S., et al., *Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments*. *Statistica Sinica*, 2002. **12**: p. 111–140.
75. Huber, P.J. and E.M. Ronchetti, *Robust Statistics*. 2nd ed. Vol. Hoboken, New Jersey
Canada. 2009: John Wiley & Sons, Inc.
76. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995. **57**(1): p. 289-300.
77. Smyth, G.K., *Linear models and empirical Bayes methods for assessing differential expression in microarray experiments*. *Stat Appl Genet Mol Biol*, 2004. **3**(1): p. 3.
78. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. *Nucleic Acids Res*, 2015. **43**(7): p. e47.
79. RCoreTeam, *R: A Language and Environment for Statistical Computing*. 2019, R Foundation for Statistical Computing.
80. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. *Genome Biol*, 2004. **5**(10): p. R80.

81. RStudio Team, *RStudio: Integrated Development for R*. RStudio, Inc., 2019.
82. Liao, Y., et al., *WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs*. *Nucleic Acids Res*, 2019. **47**(W1): p. W199-W205.
83. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update*. *Nucleic Acids Res*, 2016. **44**(W1): p. W90-7.
84. Gene Ontology, C., *Gene Ontology Consortium: going forward*. *Nucleic Acids Res*, 2015. **43**(Database issue): p. D1049-56.
85. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. *Nucleic Acids Res*, 2019. **47**(D1): p. D607-D613.
86. Lo Surdo, P., et al., *DISNOR: a disease network open resource*. *Nucleic Acids Res*, 2018. **46**(D1): p. D527-D534.
87. Pinero, J., et al., *The DisGeNET knowledge platform for disease genomics: 2019 update*. *Nucleic Acids Res*, 2019.
88. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update*. *Nucleic Acids Res*, 2013. **41**(Database issue): p. D991-5.
89. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. *Nucleic Acids Res*, 2002. **30**(1): p. 207-10.
90. Brazma, A., *Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges*. *ScientificWorldJournal*, 2009. **9**: p. 420-3.
91. Oliver, S., *On the MIAME Standards and Central Repositories of Microarray Data*. *Comp Funct Genomics*, 2003. **4**(1): p. 1.
92. Athar, A., et al., *ArrayExpress update - from bulk to single-cell expression data*. *Nucleic Acids Res*, 2019. **47**(D1): p. D711-D715.
93. Papatheodorou, I., et al., *Expression Atlas: gene and protein expression across multiple studies and organisms*. *Nucleic Acids Res*, 2018. **46**(D1): p. D246-D251.
94. Sarkans, U., et al., *The BioStudies database-one stop shop for all data supporting a life sciences study*. *Nucleic Acids Res*, 2018. **46**(D1): p. D1266-D1270.
95. McEntyre, J., U. Sarkans, and A. Brazma, *The BioStudies database*. *Mol Syst Biol*, 2015. **11**(12): p. 847.
96. Dhammi, I.K. and S. Kumar, *Medical subject headings (MeSH) terms*. *Indian J Orthop*, 2014. **48**(5): p. 443-4.
97. Ding, L.H., et al., *Distinct transcriptome profiles identified in normal human bronchial epithelial cells after exposure to gamma-rays and*

- different elemental particles of high Z and energy*. BMC Genomics, 2013. **14**: p. 372.
98. Chauhan, V. and M. Howland, *Microarray analysis of lymphocytes isolated from healthy donors ex-vivo exposed to alpha particle radiation*. 2016.
 99. Chauhan, V. and M. Howland, *Microarray analysis of cell types exposed to alpha particle radiation*. 2016.
 100. Rouchka, E.C., et al., *Early transcriptional changes in response to low dose radiation exposure*. 2017.
 101. Rouchka, E.C., et al., *Transcriptional profile of immediate response to ionizing radiation exposure*. Genom Data, 2016. **7**: p. 82-5.
 102. Baselet, B., et al., *Differential Impact of Single-Dose Fe Ion and X-Ray Irradiation on Endothelial Cell Transcriptomic and Proteomic Responses*. Front Pharmacol, 2017. **8**: p. 570.
 103. Macaeva, E., et al., *Radiation-induced alternative transcription and splicing events and their applicability to practical biodosimetry*. Sci Rep, 2016. **6**: p. 19251.
 104. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. 2005: New York. p. 397–420.
 105. Ritchie, M.E., et al., *BeadArray expression analysis using bioconductor*. PLoS Comput Biol, 2011. **7**(12): p. e1002276.
 106. Davis, S. and P.S. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor*. Bioinformatics, 2007. **23**(14): p. 1846-7.
 107. Kauffmann, A., et al., *Importing ArrayExpress datasets into R/Bioconductor*. Bioinformatics, 2009. **25**(16): p. 2092-4.
 108. Carvalho, B., *pd.hugene.1.0.st.v1: Platform Design Info for Affymetrix HuGene-1_0-st-v1*. 2015.
 109. Carvalho, B.S. and R.A. Irizarry, *A Framework for Oligonucleotide Microarray Preprocessing*. Bioinformatics, 2010. **26**(19): p. 2363-7.
 110. Bengtsson, H., J. Bullard, and K.D. Hansen, *affxparser: Affymetrix File Parsing SDK*. 2019.
 111. Gautier, L., et al., *affy--analysis of Affymetrix GeneChip data at the probe level*. Bioinformatics, 2004. **20**(3): p. 307-15.
 112. Gautier, L., et al., *{affy---analysis of Affymetrix GeneChip data at the probe level*. Bioinformatics. Vol. 20. 2004, Oxford, UK: Oxford University Press. 307--315.
 113. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor*. Nature Methods, 2015. **12**(2): p. 115--121.

114. Smyth, G.K. and T. Speed, *Normalization of cDNA microarray data*. *Methods*, 2003. **31**(4): p. 265-73.
115. Muller, C., et al., *Removing Batch Effects from Longitudinal Gene Expression - Quantile Normalization Plus ComBat as Best Approach for Microarray Transcriptome Data*. *PLoS One*, 2016. **11**(6): p. e0156594.
116. Ritchie, M.E., et al., *A comparison of background correction methods for two-colour microarrays*. *Bioinformatics*, 2007. **23**(20): p. 2700-7.
117. Dunning, M., A. Lynch, and M. Eldridge, *illuminaHumanv3.db: Illumina HumanHT12v3 annotation data (chip illuminaHumanv3)*. 2015.
118. Dunning, M., A. Lynch, and M. Eldridge, *illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4)*. 2015.
119. RCoreTeam, *R: A Language and Environment for Statistical Computing*. 2019.
120. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York.
121. Leek, J.T., et al., *sva: Surrogate Variable Analysis*. 2019.
122. Dunning, M., A. Lynch, and M. Eldridge, *illuminaHumanv2.db: Illumina HumanWG6v2 annotation data (chip illuminaHumanv2)*. 2015.
123. Klaus, B. and S. Reisenauer, *An end to end workflow for differential gene expression using Affymetrix microarrays*. *F1000Res*, 2016. **5**: p. 1384.
124. Harr, B. and C. Schlotterer, *Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons*. *Nucleic Acids Res*, 2006. **34**(2): p. e8.
125. Yates, B., et al., *Genenames.org: the HGNC and VGNC resources in 2017*. *Nucleic Acids Res*, 2017. **45**(D1): p. D619-D625.
126. MacDonald, J.W., *hugene10sttranscriptcluster.db: Affymetrix hugene10 annotation data (chip hugene10sttranscriptcluster)*. 2017.
127. Carlson, M., *primeview.db: Affymetrix ChipName not specified annotation data (chip primeview)*.
128. MacDonald, J.W., *hugene20sttranscriptcluster.db: Affymetrix hugene20 annotation data (chip hugene20sttranscriptcluster)*. 2017.
129. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. *Proc Natl Acad Sci U S A*, 2005. **102**(43): p. 15545-50.
130. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. *Nucleic Acids Res*, 2000. **28**(1): p. 27-30.
131. Kanehisa, M., et al., *New approach for understanding genome variations in KEGG*. *Nucleic Acids Res*, 2019. **47**(D1): p. D590-D595.
132. Fabregat, A., et al., *The Reactome Pathway Knowledgebase*. *Nucleic Acids Res*, 2018. **46**(D1): p. D649-D655.

133. Mi, H., et al., *Large-scale gene function analysis with the PANTHER classification system*. Nat Protoc, 2013. **8**(8): p. 1551-66.
134. Pico, A.R., et al., *WikiPathways: pathway editing for the people*. PLoS Biol, 2008. **6**(7): p. e184.
135. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges*. PLoS Comput Biol, 2012. **8**(2): p. e1002375.
136. Wang, J., et al., *WEB-based GENE SeT Analysis Toolkit (WebGestalt): update 2013*. Nucleic Acids Res, 2013. **41**(Web Server issue): p. W77-83.
137. Castelletti, N., et al., *Risk of lung adenocarcinoma from smoking and radiation arises in distinct molecular pathways*. Carcinogenesis, 2019.
138. Nikitaki, Z., et al., *Systemic mechanisms and effects of ionizing radiation: A new 'old' paradigm of how the bystanders and distant can become the players*. Semin Cancer Biol, 2016. **37-38**: p. 77-95.
139. Betlazar, C., et al., *The impact of high and low dose ionising radiation on the central nervous system*. Redox Biol, 2016. **9**: p. 144-156.
140. Ding, X., et al., *Integrative analysis of gene expression in response to low-dose ionizing radiation in a human skin model*. Med Oncol, 2015. **32**(6): p. 621.
141. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. BMC Bioinformatics, 2013. **14**: p. 128.
142. Fisher, R.A., *On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P*. Journal of the Royal Statistical Society, 1922. **85**(1): p. 87-94.
143. Landrum, M.J., et al., *ClinVar: improvements to accessing data*. Nucleic Acids Res, 2020. **48**(D1): p. D835-D844.
144. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants*. Nucleic Acids Res, 2016. **44**(D1): p. D862-8.
145. Landrum, M.J., et al., *ClinVar: public archive of relationships among sequence variation and human phenotype*. Nucleic Acids Res, 2014. **42**(Database issue): p. D980-5.
146. Pinero, J., et al., *DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants*. Nucleic Acids Res, 2017. **45**(D1): p. D833-D839.
147. Pinero, J., et al., *DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes*. Database (Oxford), 2015. **2015**: p. bav028.
148. Carbon, S., et al., *AmiGO: online access to ontology and annotation data*. Bioinformatics, 2009. **25**(2): p. 288-9.

149. Brohee, S. and J. van Helden, *Evaluation of clustering algorithms for protein-protein interaction networks*. BMC Bioinformatics, 2006. **7**: p. 488.
150. Wang, J., et al., *Recent advances in clustering methods for protein interaction networks*. BMC Genomics, 2010. **11 Suppl 3**: p. S10.
151. Seifert, E., *OriginPro 9.1: scientific data analysis and graphing software-software review*. J Chem Inf Model, 2014. **54**(5): p. 1552.
152. OriginLabCorporation, *OriginPro*. Northampton, MA, USA.
153. Lo Surdo, P., et al., *SIGNOR: A Database of Causal Relationships Between Biological Entities-A Short Guide to Searching and Browsing*. Curr Protoc Bioinformatics, 2017. **58**: p. 8 23 1-8 23 16.
154. Perfetto, L., et al., *SIGNOR: a database of causal relationships between biological entities*. Nucleic Acids Res, 2016. **44**(D1): p. D548-54.
155. Mavragani, I.V., et al., *Ionizing Radiation and Complex DNA Damage: From Prediction to Detection Challenges and Biological Significance*. Cancers (Basel), 2019. **11**(11).
156. Wortzel, I. and R. Seger, *The ERK Cascade: Distinct Functions within Various Subcellular Organelles*. Genes Cancer, 2011. **2**(3): p. 195-209.
157. Harrison, D.A., *The Jak/STAT pathway*. Cold Spring Harb Perspect Biol, 2012. **4**(3).
158. Wu, C., *Focal adhesion: a focal point in current cell biology and molecular medicine*. Cell Adh Migr, 2007. **1**(1): p. 13-8.
159. Kasahara, T., et al., *Antiapoptotic action of focal adhesion kinase (FAK) against ionizing radiation*. Antioxid Redox Signal, 2002. **4**(3): p. 491-9.
160. Bezprozvanny, I., *Calcium signaling and neurodegenerative diseases*. Trends Mol Med, 2009. **15**(3): p. 89-100.
161. Penton, A.L., L.D. Leonard, and N.B. Spinner, *Notch signaling in human development and disease*. Semin Cell Dev Biol, 2012. **23**(4): p. 450-7.
162. Alessio, N., et al., *Low dose radiation induced senescence of human mesenchymal stromal cells and impaired the autophagy process*. Oncotarget, 2015. **6**(10): p. 8155-66.
163. Feldman, D.S., et al., *Mechanisms of disease: beta-adrenergic receptors-alterations in signal transduction and pharmacogenomics in heart failure*. Nat Clin Pract Cardiovasc Med, 2005. **2**(9): p. 475-83.
164. De Souza, E.B., *Corticotropin-releasing factor receptors: physiology, pharmacology, biochemistry and role in central nervous system and immune disorders*. Psychoneuroendocrinology, 1995. **20**(8): p. 789-819.
165. Stroo, E., et al., *Cellular Regulation of Amyloid Formation in Aging and Disease*. Front Neurosci, 2017. **11**: p. 64.

166. Belinky, F., et al., *PathCards: multi-source consolidation of human biological pathways*. Database (Oxford), 2015. **2015**.
167. Courtney, A.H., W.L. Lo, and A. Weiss, *TCR Signaling: Mechanisms of Initiation and Propagation*. Trends Biochem Sci, 2018. **43**(2): p. 108-123.
168. Elibol, B. and U. Kilic, *High Levels of SIRT1 Expression as a Protective Mechanism Against Disease-Related Conditions*. Front Endocrinol (Lausanne), 2018. **9**: p. 614.
169. Rappaport, N., et al., *MalaCards: an integrated compendium for diseases and their annotation*. Database (Oxford), 2013. **2013**: p. bat018.
170. Hui, D.Y., *Phospholipase A(2) enzymes in metabolic and cardiovascular diseases*. Curr Opin Lipidol, 2012. **23**(3): p. 235-40.
171. Georgakilas, A.G., et al., *Emerging molecular networks common in ionizing radiation, immune and inflammatory responses by employing bioinformatics approaches*. Cancer Lett, 2015.
172. Hellweg, C.E., et al., *Molecular Signaling in Response to Charged Particle Exposures and its Importance in Particle Therapy*. Int J Part Ther, 2018. **5**(1): p. 60-73.