



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

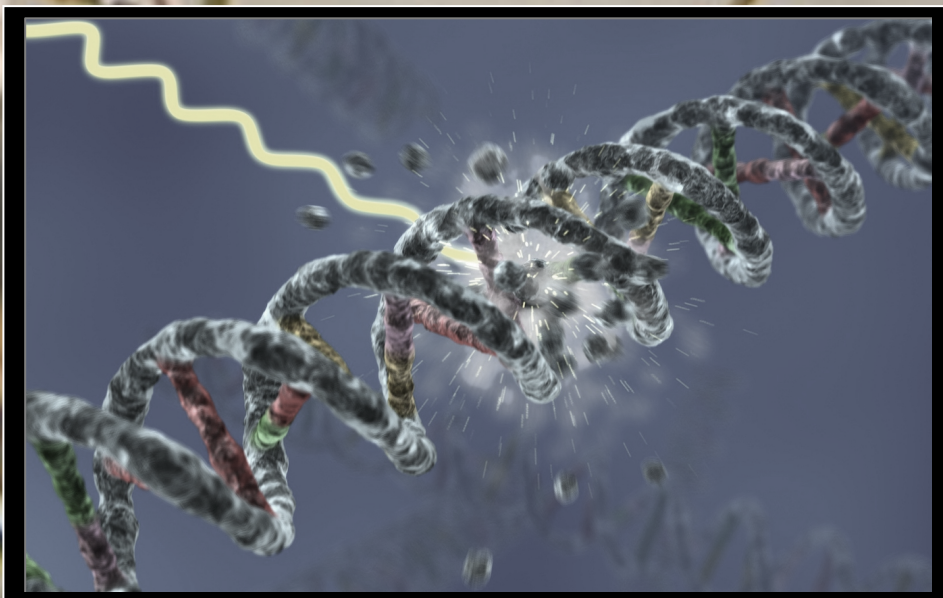
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήση του φυτού *Arabidopsis Thaliana* ως μοντέλο
για την απόκριση σε βλάβες DNA, έπειτα από έκθεση
σε Ιοντίζουσα Ακτινοβολία, με χρήση δεδομένων
μικροσυστοιχιών



ΙΩΑΝΝΗΣ ΣΤΕΦΑΝΙΔΗΣ

ΑΘΗΝΑ (2020)



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

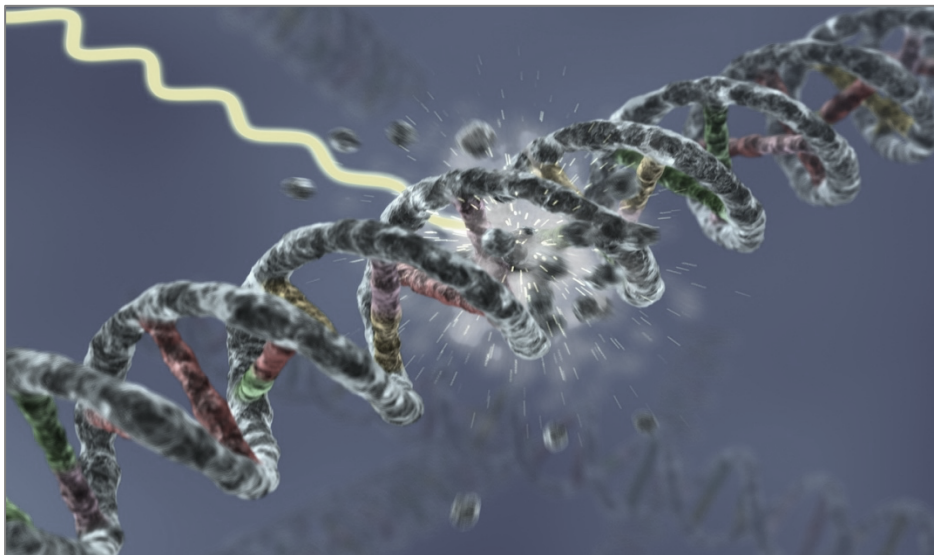
HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY

MASTER IN «BIOINFORMATICS»

Master Diploma Thesis

***Arabidopsis Thaliana* as a model organism for
studying DNA Damage Response, after exposure
to Ionizing Radiation, using microarray data**



ΙΩΑΝΝΗΣ ΣΤΕΦΑΝΙΔΗΣ

Computer Science & Telecommunications,
Technological Institute of Larissa

ATHENS (2020)



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

**Χρήση του φυτού *Arabidopsis Thaliana* ως μοντέλο
για την απόκριση σε βλάβες DNA, έπειτα από έκθεση
σε Ιοντίζουσα Ακτινοβολία, με χρήση δεδομένων
μικροσυστοιχιών.**

Τριμελής εξεταστική επιτροπή:

Αλέξανδρος Γεωργακίλας, Αναπλ. Καθηγητής (Επιβλέπων)

Τομέας Φυσικής, Ε.Μ.Π

Ιωάννης Τρουγκάκος, Καθηγητής

Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Ε.Κ.Π.Α

Αριστοτέλης Χατζηϊωάννου, Ερευνητής Α

Τομέας Βιολογικών Ερευνών και Βιοτεχνολογίας, Ε.Ι.Ε

ΕΥΧΑΡΙΣΤΙΕΣ

Το ευχαριστώ είναι δείγμα ευγνωμοσύνης σε ανθρώπους που συνέβαλλαν στην προσπάθεια επίτευξης των στόχων σου. Το μεγαλύτερο ευχαριστώ μου στη συμφοιτήτρια, φίλη και επιστήμονα Θεοδώρα – Δάφνη Μιχαλέττου που η συμμετοχή της στην εκπόνηση της διπλωματικής ήταν πραγματικά ουσιαστική. Την ευχαριστώ από τα βάθη της ψυχής μου. Ένα μεγάλο ευχαριστώ για την εμπιστοσύνη που έδειξε στο πρόσωπο μου, αναλαμβάνοντας με, ο επιβλέπων της διπλωματικής κύριος Αλέξανδρος Γεωργακίλας. Το εγχείρημα αυτό ήταν πρόκληση για όλους μας, καθώς άνθρωποι από διαφορετικά πεδία συμπράξαν για την υλοποίηση της. Άπειρες ώρες συζητήσεων, σε ένα νεόδμητο πεδίο, καθένας να προσφέρει με τις γνώσεις του. Ευχαριστώ επίσης τους κυρίους Ιωάννη Τρουγκάκο και Αριστοτέλη Χατζηιώαννου που δέχτηκαν τη πρόσκληση συμμετοχής ως μέλη της τριμελούς επιτροπής. Οι ευχαριστίες δε σταματούν εδώ και αποτελούν ένδειξη ότι ένα ταξίδι μόλις τελείωσε και ότι ένα άλλο ξεκινά με καινούργιες γνώσεις και εμπειρίες. Οι μόνοι σταθεροί άνθρωποι σε αυτή τη διαδρομή οι γονείς, ίσως η πιο σπουδαία λέξη που θα συναντήσει κανείς στην παρούσα διπλωματική. Τους ευχαριστώ που στέκονται δίπλα μου στα όμορφα και στα άσχημα, για τις θυσίες και τους κόπους χρόνων.

Αφιερωμένη στους γονείς μου,
Ελένη και Χρήστο.

ΠΕΡΙΛΗΨΗ

Είναι γνωστό ότι η έκθεση βιολογικών ιστών σε ιοντίζουσες ακτινοβολίες (IR) έχει ως αποτέλεσμα την επαγωγή μιας σειράς μηχανισμών ως απόκριση στις βλάβες που δημιουργούνται στο DNA, γνωστοί και ως DNA Damage Response (DDR). Σκοπός της παρούσας διπλωματικής είναι η διερεύνηση των μηχανισμών αυτών στην περίπτωση του φυτού μοντέλου *Arabidopsis Thaliana*. Η προσέγγιση που επιλέχθηκε ήταν η ανάλυση διαφορικής έκφρασης (Differential Gene Expression Analysis - DGEA) βασισμένη σε δεδομένα μικροσυστοιχιών από δημόσια αποθετήρια.

Η αναζήτηση των πειραματικών μας δεδομένων έγινε στη βάση ArrayExpress, καταλήγοντας σε ένα μόνο dataset με κωδικό E-GEOD-61484. Τα δείγματα του νεαρού βλαστού της *A. Thaliana* είχαν εκτεθεί σε 100Gy ακτινοβολίας-γ ή 30Gy ακτινοβολίας ΗΖΕ. Οι χρόνοι συλλογής των δειγμάτων έπειτα από ακτινοβόληση ήταν στις 1.5, 3, 6, 12, 24 ώρες. Τα στατιστικώς σημαντικά διαφορικά εκφρασμένα γονίδια αναδείχθηκαν με χρήση της γλώσσας R. Το σύνολο των διαφορικά εκφραζόμενων γονιδίων που προκύπτει, γνωστά ως DEGs, χρησιμοποιήθηκε για τη δημιουργία δικτύων πρωτεϊνικών αλληλεπιδράσεων μέσω της STRING για τα δύο είδη ακτινοβολίας ξεχωριστά καθώς επίσης για τη διεξαγωγή ανάλυσης εμπλουτισμού (enrichment analysis) σχετικά με βιολογικές διεργασίες της GO (Gene Ontology). Τέλος, έπειτα από επιλογή των σημαντικών DEGs χρησιμοποιήθηκε το εργαλείο DIOPT για την ανάδειξη των αντιστοίχων ορθολόγων στον άνθρωπο.

Τα αποτελέσματα μας υποδεικνύουν ενεργοποίηση γνωστών μηχανισμών DDR και για τα δύο είδη ακτινοβολίας, όπως σημεία ελέγχου του κυτταρικού κύκλου, επιδιορθωτικών μηχανισμών DNA και μηχανισμών γήρανσης. Παρόλα αυτά, μηχανισμοί απόπτωσης εμφανίζονται μόνο στην περίπτωση έκθεσης σε ΗΖΕ. Επιπλέον βασικές διαφορές στην περίπτωση των ΗΖΕ είναι η έντονη ενεργοποίηση του έμφυτου ανοσοποιητικού και της αυτοφαγίας. Κάτι ακόμη που παρατηρείται είναι η τάση του κυττάρου να επανέρχεται στη φυσιολογική του κατάσταση 24 ώρες μετά την ακτινοβόληση. Τέλος, από την αναζήτηση ζευγών ορθολόγων μεταξύ φυτού και ανθρώπου προκύπτει η έντονη συσχέτιση με μηχανισμούς DDR στον άνθρωπο, γεγονός που καθιστά την *A. thaliana* ως ικανοποιητικό μοντέλο για την μελέτη των επιδράσεων IR.

ABSTRACT

It is known that the exposure of biological tissues to ionizing radiation (IR) results in the induction of a series of mechanisms as a response to DNA damage, known as DNA Damage Response DDR. The aim of this master thesis is the investigation of these mechanisms in the model organism *Arabidopsis Thaliana*. Our selected approach was Differential Gene Expression Analysis (DGEA) based on microarray data from public repositories.

The search for our experimental data was conducted in the ArrayExpress database, resulting in only one dataset with accession code E-GEOD-61484. The samples of *Arabidopsis Thaliana* seedlings were exposed to 100Gy gamma or 30Gy HZE radiation. They were collected across a range of time points 1.5, 3, 6, 12, 24 hours respectively. The statistically significantly differentially expressed genes emerged through the use of R programming language. The total number of differentially expressed genes, known as DEGs, were used for the creation of protein-protein interaction networks. To this end, we utilized STRING database for both ionizing radiation types. The same DEG lists were used for performing a functional enrichment analysis concerning biological processes of GO (Gene Ontology). In the final step of our analysis, we opted for DEGs involved in main processes to predict pairs of orthologs between plant and human through the DIOPT tool.

Our results indicate activation of common DDR mechanisms for both radiation types, such as cell cycle check points, DNA repair and aging mechanisms. Nonetheless, the phenomenon of apoptosis is appeared in the case of HZE exposure. Additionally in the same case, the main differences observed in our enrichment analysis results were the intense activation of innate immune system as well as the induction of autophagy. We also observed was the tendency of cell to obtain its homeostasis 24 hours after exposure to HZE. Finally, a strong correlation concerning DDR mechanisms emerged from the predicted pairs of orthologs, verify *A. thaliana* as acceptable model organism for studying the effects of IR.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	1
1.1 Ιοντίζουσες Ακτινοβολίες	1
1.1.1 Τύποι Ιοντίζουσας Ακτινοβολίας	2
1.1.2 Πηγές Έκθεσης Ιοντίζουσας Ακτινοβολίας.....	5
1.1.3 Μονάδες Μέτρησης Ακτινοβολίας	8
1.2 Βιολογικές Επιπτώσεις Ιοντίζουσας Ακτινοβολίας	9
1.2.1 Βλάβες Γενετικού Υλικού	10
1.2.2 Απόκριση στις Βλάβες του DNA.....	11
1.2.3 Επιδιορθωτικοί Μηχανισμοί του DNA	14
1.2.4 Συσχέτιση με την Απόκριση του Ανοσοποιητικού	18
1.3 Το <i>Arabidopsis Thaliana</i> ως Οργανισμός Μοντέλο	19
1.3.1 Το Γονιδίωμα του Φυτού <i>Arabidopsis Thaliana</i>	20
1.3.2 Σύγκριση με τον Άνθρωπο	22
1.4 Μικροσυστοιχίες DNA	24
1.4.1 Πλατφόρμες Μικροσυστοιχιών.....	25
1.4.2 Ανάλυση Δεδομένων Μικροσυστοιχιών	26
2. ΜΕΘΟΔΟΣ & ΔΙΑΔΙΚΑΣΙΑ	34
2.1 Αναζήτηση Πρωτογενών Δεδομένων	34
2.2 Πειραματικός Σχεδιασμός	36
2.3 Χρήση της R για την ανάλυση των δεδομένων	37
2.3.1 Εισαγωγή δεδομένων στην κλάση ExpressionSet.....	39
2.3.2 Έλεγχος Ποιότητας Δεδομένων.....	41
2.3.3 Απεικόνιση RLE για περαιτέρω έλεγχο ποιότητας δεδομένων	43
2.3.4 Κανονικοποίηση	45
2.3.5 Φιλτράρισμα βασισμένο στην ένταση	49
2.3.5 Σχολιασμός των γονιδίων (transcript clusters/probesets)	52
2.3.6 Στατιστική Ανάλυση Διαφορικής Εκφρασμένων Γονιδίων -Γραμμικά Μοντέλα	55
2.4 Διαγράμματα VENN	59
2.5 Ανάλυση Λειτουργικού Εμπλουτισμού	61
2.6 Πρόβλεψη Ορθολόγων μέσω DIOPT	62
3. ΑΠΟΤΕΛΕΣΜΑΤΑ	64

3.1 Διαφορικώς Εκφρασμένα Γονίδια	65
3.2 Βιολογικά Δίκτυα – HUB Genes.....	66
3.3 Αποτελέσματα Ανάλυσης Εμπλουτισμού.....	72
3.4 Ορθόλογα Γονίδια.....	74
4. ΣΥΖΗΤΗΣΗ.....	79
5. ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	87

ΕΙΚΟΝΕΣ

- Εικόνα 1.** Τύποι ακτινοβόλησης και βαθμός διαπερατότητας στην ύλη (Radiation -- ANS / About Nuclear). 2
- Εικόνα 2.** Άμεση και έμμεση βλάβη βιομορίων από ιοντίζουσα ακτινοβολία (Wang et al.). 9
- Εικόνα 3.** Σηματοδοτικό μονοπάτι απόκρισης στις βλάβες του DNA (Ray and Fry).12
- Εικόνα 4.** Γενική δομή ενός MEGA-TAB αρχείου (Rocca-Serra et al.).36
- Εικόνα 5.** Δομή της κλάσης ExpressionSet της Bioconductor (Klaus and Reisenauer).39
- Εικόνα 6.** Έλεγχος ποιότητας δεδομένων μέσω θηκογραμμάτων των λογαριθμησμένων πρωτογενών τιμών έκφρασης.42
- Εικόνα 7.** Έλεγχος ποιότητας των δεδομένων μέσω θηκογραμμάτων με χρήση του γραφήματος οπτικοποίησης RLE εκτελώντας τον αλγόριθμο RMA χωρίς όμως να προηγηθεί κανονικοποίηση.45
- Εικόνα 8.** Απεικόνιση PCA με βάση την ακτινοβολία και το πειραματικό αντίγραφο. Παρατηρούμε ότι το είδος της ακτινοβόλησης είναι υπεύθυνο για τη διασπορά των δειγμάτων σε μεγαλύτερο ποσοστό από τις πειραματικές επαναλήψεις, το οποίο είναι βιολογικώς αναμενόμενο. Επιπλέον, παρατηρούμε ότι στα cluster των δειγμάτων εμφανίζονται δείγματα και από τις δύο πειραματικές επαναλήψεις, γεγονός που υποδηλώνει ότι δεν υπάρχει batch effect.47
- Εικόνα 9.** Απεικόνιση PCA με βάση την ακτινοβολία και το χρόνο συλλογής του δείγματος. Παρατηρούμε ότι το είδος της ακτινοβόλησης είναι υπεύθυνο για τη διασπορά των δειγμάτων σε μεγαλύτερο ποσοστό από το χρόνο συλλογής. Επιπλέον, παρατηρούμε το σχηματισμό cluster ανά ώρα συλλογής, το οποίο είναι βιολογικώς αναμενόμενο.48
- Εικόνα 10.** Απεικόνιση θηκογραμμάτων των κανονικοποιημένων δεδομένων μας μέσω της συνάρτησης boxplot(), έπειτα από την εφαρμογή του αλγορίθμου RMA. Τα δεδομένα μας φαίνεται να ακολουθούν μια προσεγγιστικά κανονική κατανομή.49
- Εικόνα 11.** Χρήση της συνάρτησης hist() για τη δημιουργία ιστογράμματος που απεικονίζει τις διαμέσους των εντάσεων ανά μετάγραφο. Από το γράφημα βλέπουμε μια υψηλή συγκέντρωση γονιδίων με χαμηλές διαμέσους και ενδεικτική τοποθέτηση γραμμής κατωφλίου για την αφαίρεση αυτών που αποτελούν πιθανά θόρυβο.51
- Εικόνα 12.** Στο διάγραμμα Venn απεικονίζονται με διαφορετικό χρώμα τα διαφορικώς εκφραζόμενα γονίδια στις υπό σύγκριση συνθήκες, δηλαδή στην ώρα συλλογής των δειγμάτων έπειτα από ακτινοβόληση με 100Gy γ-ακτινοβολίας στον άγριο τύπο της Arabidopsis Thaliana.60

Εικόνα 13. Στο διάγραμμα Venn απεικονίζονται με διαφορετικό χρώμα τα διαφορετικώς εκφραζόμενα γονίδια στις υπό σύγκριση συνθήκες, δηλαδή στην ώρα συλλογής των δειγμάτων έπειτα από ακτινοβολήση με 30Gy ΗΖΕ στον άγριο τύπο της *Arabidopsis thaliana*.....61

Εικόνα 14. Αναπαράσταση δικτύου πρωτεϊνικών αλληλεπιδράσεων των DEGs έπειτα από έκθεση σε ακτινοβολία-γ του φυτού *A. thaliana*. Οι πρωτεΐνες αναπαρίστανται ως κόμβοι και οι μεταξύ τους σχέσεις με ακμές. Το πάχος της γραμμής υποδεικνύει την ισχύ των δεδομένων σύνδεσης. Τα διαφορετικά χρώματα μέσα στους κόμβους αντιστοιχούν στις διάφορες βιολογικές διεργασίες που αυτοί συμμετέχουν.....67

Εικόνα 15. Αναπαράσταση δικτύου πρωτεϊνικών αλληλεπιδράσεων των DEGs έπειτα από έκθεση σε ΗΖΕ ακτινοβολία του φυτού *A. thaliana*. Οι πρωτεΐνες αναπαρίστανται ως κόμβοι και οι μεταξύ τους σχέσεις με ακμές. Το πάχος της γραμμής υποδεικνύει την ισχύ των δεδομένων σύνδεσης. Τα διαφορετικά χρώματα μέσα στους κόμβους αντιστοιχούν στις διάφορες βιολογικές διεργασίες που αυτοί συμμετέχουν.69

ΠΙΝΑΚΕΣ

- Πίνακας 1:** Συνολικός αριθμός DEGs έπειτα από ακτινοβολήση-γ του άγριου τύπου της *A. thaliana* ανά time point (Εικόνα 12).65
- Πίνακας 2:** Συνολικός αριθμός DEGs έπειτα από HZE ακτινοβολήση του άγριου τύπου της *A. thaliana* ανά time point (Εικόνα 13).65
- Πίνακας 3:** Μοναδικά διαφορικώς εκφραζόμενα γονίδια ανά τύπο ακτινοβολίας & ανά time point του άγριου τύπου της *A. thaliana*, που δε βρέθηκαν σε καμία άλλη ομάδα (Εικόνα 12, Εικόνα 13).65
- Πίνακας 4:** Κοινά διαφορικώς εκφραζόμενα γονίδια ανά τύπο ακτινοβολίας & ανά time point του άγριου τύπου της *A. thaliana*, σε όλες τις ομάδες (Εικόνα 12, Εικόνα 13).66
- Πίνακας 5 :** Genes & Hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση της *A. thaliana* σε ακτινοβολία-γ και οι βιολογικές διεργασίες στις οποίες συμμετέχουν που σχετίζονται άμεσα με το DDR (Εικόνα 14).68
- Πίνακας 6:** Υπολειπόμενα hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση ακτινοβολίας-γ και οι βιολογικές διεργασίες (GO) που συμμετέχουν (Εικόνα 14).68
- Πίνακας 7:** Genes & Hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση της *A. thaliana* σε HZE ακτινοβολία και οι βιολογικές διεργασίες (GO) στις οποίες συμμετέχουν που σχετίζονται άμεσα με το DDR (Εικόνα 15).70
- Πίνακας 8:** Υπολειπόμενα hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση HZE ακτινοβολίας και οι βιολογικές διεργασίες (GO) που συμμετέχουν (Εικόνα 15).71
- Πίνακας 9:** Τα βασικότερα αποτελέσματα που προέκυψαν από τη λειτουργική ανάλυση της GO, εισάγοντας τα DEGs που προέκυψαν από το νεαρό βλαστό της *A.thaliana* έπειτα από ακτινοβολήση-γ των 100Gy σε 5 time points.72
- Πίνακας 10:** Τα βασικότερα αποτελέσματα που προέκυψαν από τη λειτουργική ανάλυση της GO, εισάγοντας τα DEGs που προέκυψαν από το νεαρό βλαστό της *A.thaliana* έπειτα από HZE ακτινοβολήση των 30Gy σε 5 time points.73
- Πίνακας 11:** Αποτελέσματα ομολογίας γονιδίων μεταξύ της *A.thaliana* και του ανθρώπου. Από τα κοινά DEGs που προέρχονται από ακτινοβολήση-γ, επιλέχθηκαν ως input στο DIOPT εκείνα που εμπλέκονται στις βασικότερες βιολογικές διεργασίες έτσι όπως προέκυψαν από τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων της STRING.75
- Πίνακας 12:** Αποτελέσματα ομολογίας γονιδίων μεταξύ της *A.thaliana* και του ανθρώπου. Από τα κοινά DEGs που προέρχονται από HZE ακτινοβολήση, επιλέχθηκαν ως input στο DIOPT εκείνα που εμπλέκονται στις βασικότερες βιολογικές λειτουργίες έτσι όπως προέκυψαν από τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων της STRING.76

1. ΕΙΣΑΓΩΓΗ

1.1 Ιοντίζουσες Ακτινοβολίες

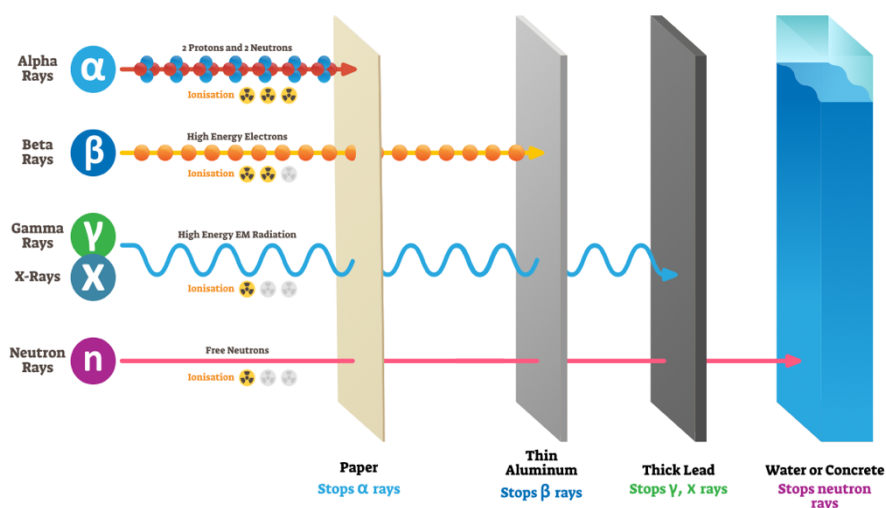
Για χρόνια οι επιστήμονες προσπαθούσαν να κατανοήσουν τη δομή του ατόμου. Σήμερα είναι γνωστό ότι το άτομο έχει έναν εξαιρετικά μικρό σε μέγεθος θετικά φορτισμένο πυρήνα που περιβάλλεται από ένα νέφος αρνητικά φορτισμένων ηλεκτρονίων. Αν και εκατό χιλιάδες φορές μικρότερος από το μέγεθος ολόκληρου του ατόμου, ο πυρήνας είναι τόσο πυκνός που αντιπροσωπεύει σχεδόν ολόκληρη τη μάζα του. Το εσωτερικό του αποτελείται από μία συστάδα προσκολλημένων μεταξύ τους σωματιδίων. Τα σωματίδια αυτά αποτελούνται από τα θετικά φορτισμένα πρωτόνια και τα ηλεκτρικά ουδέτερα νετρόνια. Ο αριθμός των πρωτονίων που περιέχει ο πυρήνας καθορίζει και το είδος του ατόμου. Άτομα που φέρουν τον ίδιο αριθμό πρωτονίων αλλά διαφορετικό αριθμό νετρονίων στον πυρήνα τους, καλούνται ισότοπα. Τα άτομα ως σύνολο είναι ηλεκτρικά ουδέτερα, καθώς τα πρωτόνια και τα ηλεκτρόνια έχουν αντίθετο στοιχειώδες ηλεκτρικό φορτίο και ο αριθμός των πρωτονίων είναι ίσος με τον αριθμό των ηλεκτρονίων. Αν και τα περισσότερα άτομα στη φύση είναι σταθερά, πολλά στοιχεία έχουν ισότοπα τα οποία είναι ασταθή. Οι πυρήνες τους διασπώνται σε άλλα είδη πυρήνων, με ταυτόχρονη ελευθέρωση ενέργειας, φαινόμενο το οποίο καλείται ραδιενέργεια. Τα ισότοπα με ασταθείς πυρήνες ονομάζονται ραδιονουκλίδια. Η ενέργεια που ελευθερώνεται είναι ικανή να αλληλεπιδράσει με άλλα άτομα και να προκαλέσει τον ιοντισμό τους. Ιοντισμός του ατόμου είναι η βίαιη απομάκρυνση ενός ή περισσότερων ηλεκτρονίων από τα τροχιακά του, εξαιτίας της δράσης κάποιου εξωτερικού αιτίου, με αποτέλεσμα την παραγωγή αντίθετα φορτισμένων ιόντων (ζεύγη φορτίων). Μεταξύ των αιτιών που μπορούν να προκαλέσουν ιοντισμό είναι οι ακτινοβολίες υψηλής ενέργειας, οι οποίες για την ικανότητα τους αυτή αποκαλούνται ιοντίζουσες (United

Nations. Scientific Committee On The Effects Of Atomic Radiation and United Nations Environment Programme).

1.1.1 Τύποι Ιοντίζουσας Ακτινοβολίας

Υπάρχουν αρκετοί τύποι ιοντίζουσας ακτινοβολίας (Εικόνα 1). Στις ιοντίζουσες ακτινοβολίες κατατάσσονται οι ηλεκτρομαγνητικές ακτινοβολίες όπως οι ακτίνες-γ και οι ακτίνες-Χ και οι ακτινοβολίες σωματιδιακής φύσεως όπως η άλφα, η βήτα, τα πρωτόνια, τα νετρόνια και κάποια υποατομικά σωματίδια που παράγονται τεχνητά ή αποτελούν συστατικές της κοσμικής ακτινοβολίας.

Η ενέργεια που εναποθέτει η ακτινοβολία στην ύλη ανά μονάδα μήκους διαδρομής καλείται γραμμική μεταβίβαση ενέργειας LET (Linear Energy Transfer). Μεγάλες τιμές LET έχουν τα βαριά φορτισμένα σωματίδια όπως τα νετρόνια, προξενώντας περισσότερες βλάβες στους ιστούς, ενώ τα ηλεκτρόνια και τα φωτόνια αποτελούν ακτινοβολίες μικρών τιμών LET (Joonhong Ahn; Humans).



Εικόνα 1. Τύποι ακτινοβολίας και βαθμός διαπερατότητας στην ύλη (Radiation -- ANS / About Nuclear).

Σωματίδια άλφα (α)

Η ακτινοβολία α είναι σωματιδιακή ακτινοβολία, η οποία εκπέμπεται από μεγάλους ραδιενεργούς πυρήνες με ατομικό αριθμό μεγαλύτερο του 82. Τα σωματίδια α είναι σχετικά βαριά σωματίδια αποτελούμενα από δύο πρωτόνια και δύο νετρόνια είναι δηλαδή πυρήνες του στοιχείου του ηλίου 4He^{+2} . Είναι τέσσερις φορές βαρύτερα του πυρήνα του υδρογόνου και μεταφέρουν σχετικά μεγάλο ηλεκτρικό φορτίο. Η εμβέλεια αυτών των σωματίων είναι μικρή και όταν προσβάλλουν την ύλη, λόγω των φυσικών ιδιοτήτων τους, επιβραδύνονται έντονα. Η ακτινοβολία α χαρακτηρίζεται από υψηλό LET και είναι δυνατόν να απορροφηθεί πλήρως από ένα και μόνο λεπτό φύλλο χαρτιού ή το δέρμα. Μπορεί να γίνει επιβλαβής εάν εισχωρήσει στον οργανισμό με την εισπνοή ή την κατάποση. (L'Annunziata; RoC Profile: Ionizing Radiation; 14th RoC 2016).

Σωματίδια βήτα (β)

Τα σωματίδια β είναι σωματίδια υψηλής ενέργειας, έχουν την ίδια μάζα και το ίδιο φορτίο όπως αυτή του ηλεκτρονίου. Εκπέμπονται από τους ραδιενεργούς πυρήνες των ατόμων όταν αυτός μεταπίπτει από μία ασταθή σε μια πιο σταθερή κατάσταση και μπορεί να είναι είτε αρνητικά (β^-) φορτισμένα ηλεκτρόνια, είτε θετικά (β^+) φορτισμένα ποζιτρόνια. Ο βραδύτερος ρυθμός απώλειας ενέργειας των σωματιδίων β τους επιτρέπει να διανύουν μεγαλύτερες αποστάσεις έναντι των σωματιδίων α και να διεισδύουν με περισσότερη ευκολία στην ύλη ή τους ιστούς, συνεπώς χαρακτηρίζεται από χαμηλό LET. Παρόλο που είναι πιο διεισδυτική, μπορεί να ανακοπεί από μερικά χιλιοστά πλαστικού ή λεπτά μεταλλικά φύλλα. Θεωρούνται επιβλαβής μόνο κατά την εισπνοή ή την κατάποση τους από τον ανθρώπινο οργανισμό (Peirce et al; Iaea, International Atomic Energy Agency).

Ακτινοβολία γ

Η ακτινοβολία γ είναι ηλεκτρομαγνητική ακτινοβολία με εξαιρετικά μικρό μήκος κύματος εκπέμποντας υψηλής ενέργειας φωτόνια. Δημιουργούνται από αλλαγές που υφίστανται στο εσωτερικό του πυρήνα του ατόμου. Εάν ο αριθμός των νουκλεονίων βρίσκεται σε τέτοια αναλογία που εφιστούν το άτομο σταθερό αλλά η ενέργεια του πυρήνα είναι πολύ μεγάλη, τότε η πλεονάζουσα ενέργεια αποβάλλεται με την μετάβαση ισομερούς. Η μετάβαση ισομερούς είναι μια διαδικασία που προκύπτει σε ένα άτομο όταν έχει προηγηθεί μια δραστηριότητα π.χ. εκπομπή α ή β σωματιδίων και η περίσσεια ενέργειας των νουκλεονίων αποβάλλεται υπό τη μορφή ακτινοβολίας. Η ενέργεια η οποία ακτινοβολείται ονομάζεται ακτινοβολία γάμμα και είναι ουσιαστικά φωτόνιο το οποίο εκπέμπεται από τον πυρήνα που βρίσκεται σε διέγερση. Το φωτόνιο δεν έχει μάζα και δεν μεταφέρει ηλεκτρικό φορτίο. Έτσι η διέλευσή της ακτινοβολίας γ μέσα από τα πλέγματα των ατόμων της ύλης, δεν εμποδίζεται, ενώ η πιθανότητά της να αλληλεπιδράσει με τα ηλεκτρόνια ή τους πυρήνες των ατόμων είναι σχετικά μικρή. Συνεπώς είναι διεισδυτική ακτινοβολία χαμηλού LET και αποκόπτεται δύσκολα. Πυκνά υλικά όπως ο χάλυβας και ο μόλυβδος παρέχουν ασπίδα προστασίας έναντι της ακτινοβολίας αυτής (Alburger; Humans).

Ακτινοβολία νετρονίων

Τα νετρόνια είναι ηλεκτρικώς ουδέτερα σωματίδια. Επομένως δεν αλληλεπιδρούν με τα ηλεκτρόνια της ύλης αλλά με τους πυρήνες των ατόμων, απελευθερώνοντας φορτισμένα σωματίδια που εφόσον έχουν αρκετή κινητική ενέργεια προκαλούν διέγερση και ιοντισμό της ύλης. Αλληλεπιδρούν κυρίως με πυρήνες ατόμων μικρού ατομικού αριθμού όπως του υδρογόνου, οι οποίοι συνήθως αποκτούν αρκετή κινητική ενέργεια για να διεγείρουν και να ιοντίσουν γειτονικά άτομα. Σε

βιολογικό ιστό, τα νετρόνια αλληλεπιδρούν πρωτίστως με το υδρογόνο του ύδατος, παράγοντας πρωτόνια με σημαντική κινητική ενέργεια. Η ακτινοβολία νετρονίων είναι πολύ διεισδυτική, όπως και η ακτινοβολία γάμμα (γ), και ιοντίζει την ύλη έμμεσα, μεταφέροντας την ενέργειά της σε συνήθως βαριά φορτισμένα σωματίδια. Μπορούν επίσης να συλληφθούν από πυρήνες ατόμων και να επανεκπεμφθούν ή να δημιουργήσουν διαφορετικό πυρήνα, οπότε η περίσσεια της ενέργειας θα εκπεμφθεί με τη μορφή ακτινοβολίας γ . Η ακτινοβολία νετρονίου χαρακτηρίζεται από υψηλό LET (Nikolopoulos et al.).

Ακτίνες Χ

Οι ακτίνες Χ είναι υψηλής ενέργειας φωτόνια όπως και η ακτινοβολία γάμμα (γ). Παράγονται τεχνητά από την ταχεία επιβράδυνση δεσμών ηλεκτρονίου από διεργασίες που συντελούνται εκτός του πυρήνα των ατόμων. Οι ακτίνες Χ είναι και αυτές αρκετά διεισδυτικές και στην απουσία θωράκισης από πυκνά υλικά μπορεί να εναποθέσει μεγάλα ποσά ενέργειας στο εσωτερικό των οργάνων. Θεωρούνται ακτινοβολίες χαμηλού LET. Στους ιστούς μεταφέρουν την ενέργεια τους κυρίως στα ηλεκτρόνια (International Atomic Energy Agency).

1.1.2 Πηγές Έκθεσης Ιοντίζουσας Ακτινοβολίας

Ο άνθρωπος εκτίθεται σε διάφορες φυσικές και τεχνητές πηγές ιοντιζουσών ακτινοβολιών. Οι φυσικές πηγές υπάρχουν ανεξάρτητα από την ανθρώπινη δραστηριότητα ενώ οι τεχνητές έχουν προκύψει ως αποτέλεσμα των τεχνολογικών δραστηριοτήτων του ανθρώπου. Οι φυσικές πηγές βρίσκονται στο διάστημα και στο γήινο περιβάλλον, ενώ οι τεχνητές πηγές προκύπτουν από την ιατρική χρήση ιοντιζουσών ακτινοβολιών στην ακτινοδιαγνωστική, στην ακτινοθεραπεία και στην πυρηνική ιατρική. Η έκθεση του ανθρώπου σε ιοντίζουσες ακτινοβολίες μπορεί να είναι είτε εσωτερική είτε εξωτερική. Η εσωτερική

ακτινοβολήση οφείλεται σε εισπνοή και κατάποση ραδιενεργών ισοτόπων ενώ η εξωτερική σε ραδιοστοιχεία του εδάφους και την κοσμική ακτινοβολία (Métivier).

1.1.2.1 Φυσικές Πηγές

Η μεγαλύτερη φυσική πηγή έκθεσης ακτινοβολίας στον άνθρωπο είναι το ραδόνιο, το οποίο βρίσκεται στο περιβάλλον ως φυσικό ραδιενεργό υλικό. Το ευγενές αυτό ραδιενεργό αέριο εισέρχεται στον ατμοσφαιρικό αέρα όταν αυτό διαφύγει από το έδαφος και τα πετρώματα και εισχωρεί στο εσωτερικό των κτιρίων μέσω κενών ή ρωγμών. Εφόσον οι συγκεντρώσεις ραδονίου εντός των κτιρίων είναι μη επιτρεπτές, τότε ο κίνδυνος έκθεσης από τα προϊόντα διάσπασης του είναι αρκετά μεγάλος (Vogiannis and Nikolopoulos; Charles). Άλλες φυσικές πηγές που προέρχονται από τα προϊόντα διάσπασης ραδιενεργών στοιχείων του εδάφους και των πετρωμάτων είναι το Ουράνιο και το Θόριο.

Η πρόσληψη ραδιενεργών στοιχείων μέσω της τροφικής αλυσίδας αποτελεί μία δεύτερη φυσική πηγή έκθεσης ακτινοβολίας στον ανθρώπινο οργανισμό. Τα σημαντικότερα ραδιοϊσότοπα που εμπíπτουν σε αυτή την κατηγορία είναι ^{40}K , το οποίο βρίσκεται σε σταθερή ισοτοπική αναλογία σε σχέση με το φυσικό κάλιο και ο άνθρακας ^{14}C , που παράγεται στα ανώτερα στρώματα της ατμόσφαιρας και αποτελεί το σημαντικότερο ισότοπο κοσμικής προέλευσης.

Μια τρίτη φυσική πηγή εξωτερική έκθεσης είναι η κοσμική ακτινοβολία. Οι περισσότερες από τις ακτινοβολίες που τη συνιστούν, πηγάζουν από το βαθύ διαστρικό χώρο. Απελευθερώνονται επίσης από τον ήλιο κατά τη διάρκεια των ηλιακών εκλάμψεων. Παρόλο που αποτελούν την κύρια πηγή ακτινοβολίας στο διάστημα, ακτινοβολούν άμεσα τη Γη

αλληλεπιδρώντας με την ατμόσφαιρα, παράγοντας έτσι διαφορετικούς τύπους ακτινοβολίας και ραδιενεργού υλικού (Exposure Due to Natural Sources).

1.1.2.2 Τεχνητές Πηγές

Οι τεχνητές πηγές έκθεσης ακτινοβολίας στον άνθρωπο περιλαμβάνουν τη διαγνωστική ραδιολογία, που χρησιμοποιεί πολύ σύνθετα απεικονιστικά συστήματα παραγωγής ακτίνων Χ. Η δέσμη των ακτίνων αυτών επικεντρώνεται στην περιοχή του σώματος που πρέπει να απεικονισθεί, με πιο σύνηθες τις ακτινογραφικές εξετάσεις του θώρακος και της λεκάνης-ισχίων.

Άλλη μία βασική τεχνητή πηγή έκθεσης σε ιοντίζουσα ακτινοβολία αποτελεί η πυρηνική ιατρική, που χρησιμοποιεί ραδιοφάρμακα, τα οποία χορηγούνται σε εξεταζόμενους για την απεικόνιση της μορφολογίας και της λειτουργίας των οργάνων. Θεραπευτικά ραδιοφάρμακα χορηγούνται κυρίως για την αντιμετώπιση νεοπλασιών όπως ο καρκίνος του θυροειδούς και για παρηγορικές θεραπείες. Το ραδιοφάρμακο, αφού εισέλθει στο σώμα, κατευθύνεται και συγκεντρώνεται στην υπό εξέταση περιοχή εκπέμποντας ακτινοβολία, η οποία και ανιχνεύεται από κατάλληλα συστήματα, όπως γ camera, SPECT/CT, PET/CT.

Τέλος, στις τεχνητές πηγές συνεισφέρουν η επεμβατική ακτινολογία και η ακτινοθεραπεία, η οποία χρησιμοποιείται για την καταστροφή νεοπλασματικών κύτταρων. Η θεραπεία σχεδιάζεται προσεκτικά για να εξασφαλιστεί ότι σωστή δόση ακτινοβολίας προσδίδεται στη σωστή περιοχή του όγκου, ενώ ταυτόχρονα η δόση στους φυσιολογικούς ιστούς ελαχιστοποιείται, ώστε η βλάβη στα υγιή όργανα να είναι ελάχιστη. Η πιο συνήθης ακτινοθεραπεία καρκινικών όγκων γίνεται με

εξωτερική ακτινοβολήση. Δέσμες ακτίνων γάμμα ή Χ ή δέσμες ηλεκτρονίων κατευθύνονται υπό διάφορες γωνίες και θέσεις εξωτερικά του ασθενούς και εστιάζουν στην περιοχή του όγκου (“Ελληνική Επιτροπή Ατομικής Ενέργειας - Αρχική”).

1.1.3 Μονάδες Μέτρησης Ακτινοβολίας

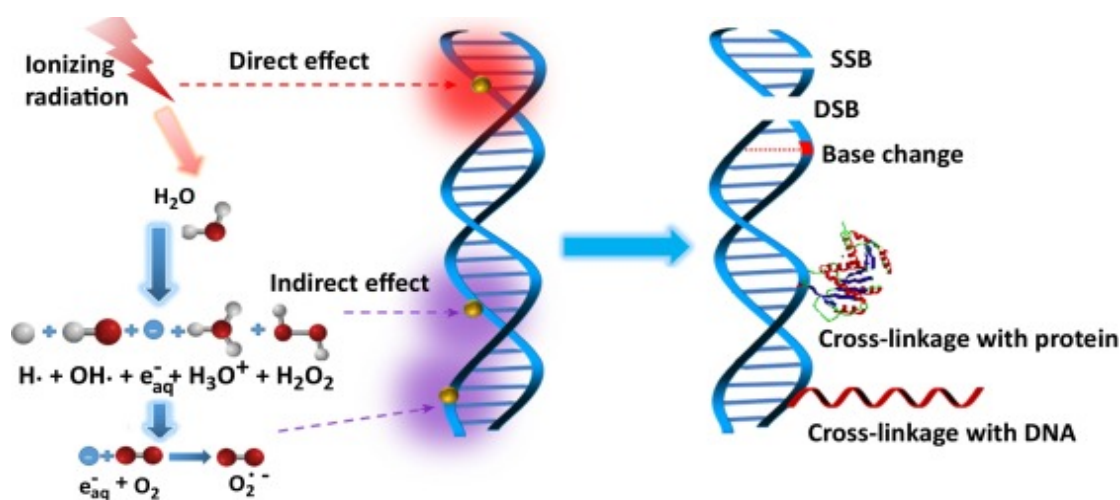
Η ραδιενέργεια μετράται με διαφορετικούς τρόπους. Όταν ένας ιστός λαμβάνει ακτινοβολία από ραδιενεργή πηγή μετράται η δόση ακτινοβολήσης, μετράται όμως και το απορροφούμενο ποσό της ραδιενέργειας στο νερό, στο έδαφος και στον αέρα. Οι μονάδες μέτρησης είναι διαφορετικές σε κάθε μία από τις παραπάνω περιπτώσεις. Υπάρχουν διαφορετικές αλλά συσχετισμένες μεταξύ τους μονάδες μέτρησης ραδιενέργειας και των αποτελεσμάτων που αυτές προκαλούν (US EPA, “Radiation Terms and Units”).

Η ραδιενέργεια αναφέρεται στην αυθόρμητη μεταστοιχείωση ενός ασταθούς ισοτόπου που ακολουθείται από εκπομπή ιοντίζουσας ακτινοβολίας. Ανεξάρτητα από τον τύπο ακτινοβολίας, η ποσότητα ενός ραδιενεργού υλικού σε δεδομένη χρονική στιγμή εκφράζεται με τον αριθμό των ραδιενεργών διασπάσεων στη μονάδα του χρόνου. Σαν μονάδα μέτρησης της ενεργότητας έχουμε τη μία διάσπαση ανά δευτερόλεπτο, που ονομάζεται Becquerel (Bq). Η δόση έκθεσης ορίζεται ως το ηλεκτρικό φορτίο που παράγεται από την ιοντίζουσα ηλεκτρομαγνητική ακτινοβολία στη μονάδα της μάζας που διαχέεται στον αέρα. Μονάδα μέτρησης της δόσης έκθεσης είναι το 1 Coulomb/kg. Το μέγεθος που εκφράζει την ενέργεια που εναποτίθεται από την προσπίπτουσα ακτινοβολία κατά τη διέλευση της στη μονάδα μάζας της προσβαλλόμενης ύλης ονομάζεται απορροφούμενη δόση (D) με μονάδα μέτρησης το Gray (Gy) που αντιστοιχεί σε απορροφούμενη ενέργεια 1 Joule/kg προσβαλλόμενης ύλης. Ισοδύναμη δόση καλείται το

δοσιμετρικό μέγεθος που σχετίζεται με τη βιολογικές επιδράσεις που ενδέχεται να προκαλέσει σε έναν ιστό η απορροφούμενη δόση ενός συγκεκριμένου τύπου ακτινοβολίας. Μονάδα ισοδύναμης δόσης είναι το Sievert (Sv) (“NRC: Measuring Radiation”).

1.2 Βιολογικές Επιπτώσεις Ιοντίζουσας Ακτινοβολίας

Όταν ακτινοβολία απορροφηθεί από το κύτταρο ενδεχομένως να επηρεάσει μια σειρά κρίσιμων κυτταρικών στόχων με σημαντικότερο αυτό του γενετικού υλικού. Ενδείξεις υποδεικνύουν ότι συγκεκριμένα οι βλάβες του DNA επιφέρουν κυτταρικό θάνατο, γενετικές μεταλλάξεις καθώς και καρκινογένεση. Οι επιπτώσεις της ιοντίζουσας ακτινοβολίας σε ένα βιολογικό υλικό, χωρίζονται σε δύο βασικές κατηγορίες, την άμεση και την έμμεση (Εικόνα 2) (Elgazzar and Kazem).



Trends in Pharmacological Sciences

Εικόνα 2. Άμεση και έμμεση βλάβη βιομορίων από ιοντίζουσα ακτινοβολία (Wang et al.).

Ο μηχανισμός της άμεσης δράσης της ακτινοβολίας αφορά κυρίως στο γενετικό υλικό, καθώς σχετίζεται με την εναπόθεση ενέργειας απευθείας στο DNA, προκαλώντας ιονισμό των ατόμων του και

διασπώντας τους χημικούς δεσμούς του, διαταράσσοντας έτσι τη δομή και τη λειτουργία του (Flowers et al.). Εξαιτίας του σχετικά μικρού όγκου που καταλαμβάνει το DNA στο εσωτερικό του κυττάρου, η πιθανότητα της άμεσης αλληλεπίδρασης είναι εξαιρετικά χαμηλή και εξαρτάται από τον τύπο της ακτινοβολίας και την πυκνότητα των ιοντισμών που προκαλεί.

Αντιθέτως οι βλάβες που προκαλούνται από το μηχανισμό της έμμεσης δράσης είναι εκείνες κατά τις οποίες η ακτινοβολία δεν απορροφάται απευθείας από το γενετικό υλικό αλλά παραλαμβάνεται έμμεσα από άλλα ιονισμένα ή διεγερμένα μόρια. Στην περίπτωση αυτή, η αλληλεπίδραση της ακτινοβολίας με μόρια νερού, τα οποία βρίσκονται σε αφθονία εντός του κυττάρου, έχουν ως αποτέλεσμα την παραγωγή ελευθέρων ριζών. Οι ελεύθερες ρίζες έχουν ένα αζευγάρωτο ηλεκτρόνιο που τις καθιστά εξαιρετικά αντιδραστικά μόρια, δίνοντας έτσι την δυνατότητα να πυροδοτούν μια σειρά τροποποιήσεων σε μόρια του κυττάρου που κατέχουν ζωτικής σημασίας ρόλους για την εύρυθμη λειτουργία του.

1.2.1 Βλάβες Γενετικού Υλικού

Είναι γνωστό ότι το γενετικό υλικό, που αποτελεί τη βασική μονάδα της κληρονομικότητας, είναι ένα εγγενώς αντιδραστικό μόριο επιρρεπές σε χημικές τροποποιήσεις από ενδογενείς και εξωγενείς παράγοντες. Επιπροσθέτως, η DNA πολυμεράση που εμπλέκεται στους μηχανισμούς αντιγραφής και επιδιόρθωσης του DNA δημιουργεί τυπογραφικά λάθη, επιβαρύνοντας τα κύτταρα με ενδεχόμενες δυσμενείς για τον οργανισμό μεταλλάξεις.

Όπως αναφέρθηκε παραπάνω, οι βλάβες του γενετικού υλικού μπορούν να ταξινομηθούν σε δύο βασικές κατηγορίες: τις ενδογενείς και τις

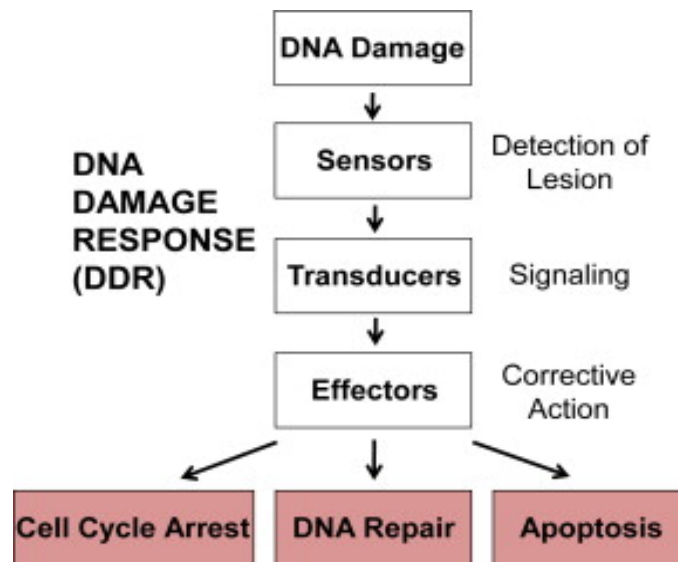
εξωγενείς. Η πλειοψηφία των ενδογενών βλαβών προκύπτει από το χημικά ενεργό DNA το οποίο εμπλέκεται σε υδρολυτικές και οξειδωτικές αντιδράσεις και τις δραστικές μορφές οξυγόνου (ROS). Οι εξωγενείς βλάβες από την άλλη μεριά, πραγματοποιούνται όταν περιβαλλοντικοί, φυσικοί και χημικοί παράγοντες βλάπτουν το DNA. Παράδειγμα αυτών είναι η ιοντίζουσα και η υπεριώδης ακτινοβολία που μπορούν να οδηγήσουν στην παραγωγή διμερών πυριμιδινών και χημικών μεταλλαξογόνων (Chatterjee and Walker).

Από τις πιο σημαντικές άμεσες βλάβες που η ιοντίζουσα ακτινοβολία προκαλεί είναι η θραύση της διπλής αλυσίδας του DNA (Double Strand Breaks - DSBs). Αντίθετα, βλάβες που συντελούνται σε επίπεδο βάσεων επάγονται συνήθως εξαιτίας της έμμεσης επίδρασης της ακτινοβολίας, μέσω του σχηματισμού αντιδραστικών μορφών οξυγόνου (ROS). Γενικότερα, στις βλάβες DNA που προκύπτουν από τις έμμεσες επιδράσεις περιλαμβάνονται οι αλλοιώσεις βάσεων (οξειδώσεις, μεθυλιώσεις, απαμινώσεις), η απώλεια μίας βάσης, η δημιουργία αβασικών σημείων (AP sites) και η θραύση του ενός κλώνου της αλυσίδας DNA (Single Strand Breaks - SSBs) (Borrego-Soto et al.).

1.2.2 Απόκριση στις Βλάβες του DNA

Τα κύτταρα υπόκεινται συνεχώς σε βλάβες του γενετικού τους υλικού από ενδογενείς και εξωγενείς παράγοντες. Με σκοπό να διατηρήσουν τη γενωμική τους ακεραιότητα, τα κύτταρα έχουν αναπτύξει ένα σύνολο μηχανισμών απόκρισης στις βλάβες DNA (DDR – DNA damage response), ένα πολυεπίπεδο σηματοδοτικό μονοπάτι που συμπεριλαμβάνει διαφορετικές λειτουργικές πρωτεΐνες. Αυτές περιλαμβάνουν πρωτεΐνες ανίχνευσης, μεταγωγής σήματος και ενεργοποίησης της εκκίνησης απόκρισης σε διαδικασίες όπως η καταστολή του κυτταρικού κύκλου, η επιδιόρθωση του DNA και η απόπτωση (Εικόνα 3). Οι διαδικασίες αυτές

πρέπει να είναι άμεσες και αποτελεσματικές για την επιβίωση του κυττάρου (Ray and Fry). Οι βλάβες που προκύπτουν πρέπει ορθά να επιδιορθωθούν, ώστε να αποφευχθεί οποιαδήποτε απώλεια ή διαίωσιση εσφαλμένης πληροφορίας κατά την κυτταρική διαίρεση. Τέτοιου είδους λάθη μπορεί να προκαλέσουν αναπτυξιακές ανωμαλίες και καρκινογένεση. Ωστόσο, η επιλογή του κατάλληλου μηχανισμού επιδιόρθωσης από το κύτταρο, βασίζεται σε δύο πράγματα, στον τύπο της βλάβης DNA και στην εκάστοτε φάση του κυτταρικού κύκλου που προκαλείται η βλάβη (Branzei and Foiani).



Εικόνα 3. Σηματοδοτικό μονοπάτι απόκρισης στις βλάβες του DNA (Ray and Fry).

Ο κυτταρικός κύκλος αποτελείται από μία σειρά προκαθορισμένων γεγονότων που οδηγούν στο διπλασιασμό και τη διαίρεση του κυττάρου. Η διαδικασία περιλαμβάνει την αντιγραφή του γενετικού υλικού, το διαχωρισμό των χρωμοσωμάτων, τη μίτωση και τη διαίρεση. Εξωκυττάρια σήματα ελέγχουν την είσοδο, την έξοδο και την πρόοδο του κυτταρικού κύκλου. Τα σημεία ελέγχου, που αποτελούν μονοπάτια σηματοδότησης, επιτηρούν την πορεία του και είναι υπεύθυνα για την

καταστολή του στην περίπτωση ανίχνευσης βλαβών στο DNA, έως ότου αυτές επιδιορθωθούν. Στην περίπτωση που οι βλάβες δεν είναι δυνατόν να επιδιορθωθούν, τα σημεία ελέγχου μπορούν να οδηγήσουν στην ενεργοποίηση του μηχανισμού απόπτωσης. Τα στάδια του κυτταρικού κύκλου διαιρούνται επιγραμματικά στις εξής φάσεις (Trent):

- G₀ - φάση μηδέν ή φάση ηρεμίας (quiescent stage)
- G₁ - ανάπτυξη του κυττάρου
- S - σύνθεση του DNA
- G₂ – ανάπτυξη του κυττάρου
- M - μιτωτική φάση

Οι θέσεις των σημείων ελέγχου για την επιδιόρθωση των βλαβών είναι κομβικής σημασίας αφενός γιατί στο τέλος της φάσης G₁ το κύτταρο προετοιμάζεται για χρωμοσωμική αντιγραφή και αφετέρου γιατί στο τέλος της φάσης G₂ το κύτταρο προετοιμάζεται για τη μιτωτική του διαίρεση. Αν το κύτταρο υποστεί βλάβη από εξωγενή πηγή και ένα τρίτο σημείο ελέγχου ενεργοποιείται κατά τη μετάφαση που καλείται (metaphase ή spindle assembly checkpoint) όπου ελέγχει για το σωστό διαχωρισμό των ζευγών χρωμοσωμάτων. Οι εξιδεικευμένοι μηχανισμοί που διέπουν αυτά τα συστήματα ελέγχου έχουν εκτενώς μελετηθεί. Κεντρικό ρόλο στα μονοπάτια απόκρισης στις βλάβες του DNA (DDR) διαδραματίζουν οι πρωτεϊνικές κινάσες θρεονίνης/σερίνης ATM, ATR και DNA-PK, μέλη της οικογένειας των κινασών PI3. Η ATM ενεργοποιείται έπειτα από ανίχνευση double strand breaks στο γενετικό υλικό ενώ η ATR ενεργοποιείται από την παρουσία single-strand breaks (SSBs) λόγω αντιγραφικού στρες. Ως αντιγραφικό στρες ορίζεται η καταστολή της διχάλας αντιγραφής έπειτα από αδυναμία σύζευξης του συμπλόκου ελικάσης και DNA πολυμεράσης για τη σύνθεση του DNA. Η αρχή της σηματοδότησης του μονοπατιού απόκρισης σε τέτοιου είδους βλάβες επιτυγχάνεται με τις πρωτεϊνικές

κινάσες Chk1 και Chk2, οι οποίες ενεργοποιούνται και φωσφορυλιώνονται με ATR και ATM εξαρτώμενο τρόπο και οι οποίες με τη σειρά τους φωσφορυλιώνουν την CDC25. Η φωσφορυλιωμένη φωσφατάση CDC25 μεταφέρεται από τον πυρήνα στο κυτταρόπλασμα όπου διακόπτεται ο κυτταρικός κύκλος. Διακοπή του κυτταρικού κύκλου μπορεί να προκληθεί και από τη δράση της πρωτεΐνης p53 αφού αρχικά φωσφορυλιωθεί από την Chk2 (Santivasi and Xia; Zeman and Cimprich). Η πρωτεΐνη p53 φαίνεται να αποκτά ρόλο μεσολαβητή και να εμπλέκεται σε όλους τους μηχανισμούς επιδιόρθωσης του DNA καθώς επίσης και σε ένα ευρύ φάσμα κυτταρικών διαδικασιών όπως η απόπτωση. Ο ρόλος της p53 στην απόπτωση έχει συνδεθεί με τη δραστηριότητα καταστολής του όγκου. Σε απόκριση του κυτταρικού στρες, η p53 ρυθμίζει επίσης μια σειρά γονιδίων που πυροδοτούν αποπτωτικές αντιδράσεις (Xu and El-Gewely; Aubrey et al.).

Παρά τους αποτελεσματικούς μηχανισμούς επιδιόρθωσης που έχουν αναπτυχθεί κατά την πορεία εξέλιξης της ζωής, το γενετικό υλικό του κυττάρου ενδέχεται να παραμείνει τελείως ανεπιδιόρθωτο ή μερικώς επιδιορθωμένο, συνεχίζοντας τον κυτταρικό του κύκλο διαιώνοντας τις μεταλλάξεις (Goodarzi et al.). Μία πιθανή αιτία αδυναμίας των μηχανισμών επιδιόρθωσης να αποκαταστήσουν τη βλάβη είναι η επαγωγή σύνθετων βλαβών DNA (complex/clustered DNA damage). Τα λεγόμενα και multiply damaged sites (MDS) αποτελούν μία πληθώρα συσσωρευμένων, διαφορετικών τύπων βλαβών του DNA σε απόσταση 15-20 bp που φαίνεται να οδηγούν το κύτταρο σε γενωμική αστάθεια (Mavragani et al.).

1.2.3 Επιδιορθωτικοί Μηχανισμοί του DNA

Εκτιμάται πως το κύτταρο υφίσταται καθημερινά περί τις 70.000 αλλοιώσεις στο σύνολο του (LINDAHL and BARNES). Η πλειοψηφία

αυτών, σχετιζόμενων με αλλοιώσεις στο γενετικό υλικό, προέρχεται από single-strand breaks (SSBs) που προκύπτουν από οξειδωτικές βλάβες, κατά τη διάρκεια του μεταβολισμού ή την υδρόλυση των βάσεων. Παρόλα αυτά, η επαγωγή των double-strand breaks (DSBs) αν και δεν είναι συχνή, αποδεικνύεται επικινδυνότερη. Κατά τη διάρκεια της εξέλιξης τους τα κύτταρα για να διατηρήσουν την ακεραιότητα τους, δεδομένου του καθημερινού καταγισμού ενδογενών και εξωγενών βλαβών του γενετικού τους υλικού, έπρεπε να αποκτήσουν ένζυμα για την επιδιόρθωση των βλαβών που υφίστανται στο DNA. (Friedberg; Tubbs and Nussenzweig).

Ως πρώτη γραμμή άμυνας έναντι όλων αυτών των επιβλαβών παραγόντων, οι οργανισμοί ανέπτυξαν μονοπάτια επιδιόρθωσης στις βλάβες του DNA (Kusakabe et al.), τα πέντε βασικότερα εκ των οποίων παρατίθενται παρακάτω (Hakem) :

Επιδιόρθωση Λανθασμένων Ζευγών Βάσεων (Mismatch Repair - MMR)

Αποτελεί σημαντικό μονοπάτι καταστολής σφαλμάτων, που εισάγονται κατά την αντιγραφή του γενετικού υλικού, απομακρύνοντας παράταιρα ζεύγη βάσεων (G-T αντί για G-C) και διορθώνοντας επίσης και μικρούς βρόχους ένθεσης ή απαλοιφής εξαιτίας της ολίσθησης της DNA πολυμεράσης. Αρχικά ο μηχανισμός αναγνωρίζει τη βλάβη. Έπειτα αφαιρεί το τμήμα του κλώνου όπου βρίσκεται η λανθασμένη βάση με πολυμερισμό του κενού και συγκόλληση των φωσφοδιεστερικών δεσμών από την DNA λιγάση. Αν η επιδιόρθωση επιτευχθεί στον πατρικό κλώνο ή αποτύχει πριν ξεκινήσει ο επόμενος κύκλος αντιγραφής του DNA, η αταίριαστη βάση θα οδηγήσει σε μετάλλαξη. Ελλαττωματική επιδιόρθωση του (MMR) καταλήγει σε αυξημένα ποσοστά κυτταρικών μεταλλάξεων που μπορεί να προκαλέσουν πολυποδικό καρκίνο του παχέος εντέρου (σύνδρομο Lynch) (Graham V et al.).

Επιδιόρθωση με Εκτομή Βάσεων (Base Excision Repair - BER)

Η εκτομή της αυθόρμητης απαμίνωσης της κυτοσίνης σε ουρακίλη από το ένζυμο DNA γλυκοζυλάση είναι ένα παράδειγμα της επιδιόρθωσης εκτομής βάσεων (BER). Τέτοιου είδους τροποποιήσεις προκαλούν μικρές αλλαγές στη δομή του DNA και οφείλονται συνήθως στην εγγενή αστάθεια του γενετικού υλικού. Άλλα χημικά μεταλλαξογόνα που τροποποιούν το DNA αλλάζοντας τη συμπληρωματικότητα των βάσεων περιλαμβάνουν τις αλκαλιωμένες πουρίνες όπως η 3'- μεθυλαδενίνη (3Me-A). Οι τροποποιήσεις αυτές προκαλούνται επίσης από περιβαλλοντικούς παράγοντες όπως την ιοντίζουσα ακτινοβολία που δημιουργεί αντιδραστικά είδη οξυγόνου. Οι ενζυμικά καταλυόμενες απαλοιφές από τις γλυκοζυλάσες διασπών τον N-γλυκοζιτικό δεσμό αφήνοντας μια οπή που ονομάζεται AP ή αβασική περιοχή. Αν η απομάκρυνση της βάσης που έχει υποστεί βλάβη αφήνει ένα κενό μήκους ενός νουκλεοτιδίου, ο μηχανισμός ακολουθεί το μονοπάτι SP-BER (Short Patch - BER). Αντίθετα, αν το κενό που προκύπτει έχει μήκος 2-9 νουκλεοτίδια, ο μηχανισμός ακολουθεί το μονοπάτι LP-BER (Long Patch - BER) (Krokan et al. ; Parikh et al.).

Επιδιόρθωση Εκτομής Νουκλεοτιδίων (Nucleotide Excision Repair-NER)

Ο μηχανισμός επιδιόρθωσης με εκτομή νουκλεοτιδίων απομακρύνει μία πληθώρα ογκωδών αλλοιώσεων, που παραμορφώνουν τη δομή της διπλής έλικας. Οι πιο σημαντικές από τις βλάβες αυτές είναι τα διμερή πυριμιδινών που προκαλούνται από την υπεριώδη ακτινοβολία. Ο μηχανισμός αυτός αναγνωρίζει τη μη φυσιολογική δομή που αποκτούν περιοχές του DNA και τις εκτέμνει. Αρχικά αναγνωρίζει τη βλάβη, και αμέσως δεσμεύει ένα πρωτεϊνικό σύμπλοκο στην περιοχή της παραμόρφωσης. Έπειτα απομακρύνει την αλλοιωμένη αλυσίδα που απαρτίζεται από 12 νουκλεοτίδια και τέλος αποκαθιστά και επανασυνθέτει το ελλιπών τμήμα από την DNA πολυμεράση και λιγάση.

Αδυναμία επιδιόρθωσης του DNA έχει ως αποτέλεσμα την εμφάνιση διαφόρων αυτοσωμικών νοσημάτων, όπως η μελαγρωματική ξηροδερμία (Xeroderma Pigmentosum, XP) που οδηγεί σε υπερευαισθησία στο ηλιακό φως (Dexheimer).

Ομόλογος Ανασυνδυασμός (Homologous Recombination - HR)

Ο ομόλογος ανασυνδυασμός (HR) συμπεριλαμβάνει το ξετύλιγμα της κατεστραμμένης έλικας και της διείσδυσης των κατεστραμμένων κλώνων σε ένα ομόλογο DNA μόριο που μπορεί να είναι είτε η αδερφή χρωματίδα είτε το ομόλογο χρωμόσωμα, δηλαδή ομόλογες αλληλουχίες μέσα στο γονιδίωμα. Η επιδιόρθωση από τη σπάσιμο της διπλής έλικας του DNA μέσω του ομόλογου ανασυνδυασμού επιτελείται μέσω της αντιγραφής, χρησιμοποιώντας τον ομόλογο κλώνο σαν μήτρα από το σημείο του ρήγματος και μετά. Ο ομόλογος ανασυνδυασμός συνεπώς απαιτεί εκτεταμένες περιοχές ομολογίας σε ένα άλλο δίκλωνο μόριο DNA για να μη χαθεί η γενετική πληροφορία.

Μη Ομόλογη Ένωση Άκρων (Non Homologous End Joining - NHEJ)

Σε αντίθεση με τον ομόλογο ανασυνδυασμό (HR), η μη ομόλογη ένωση άκρων (NHEJ) δεν απαιτεί ομολογία με ένα δεύτερο δίκλωνο μόριο DNA και καθόλου ή ελάχιστα ζεύγη βάσεων ομόλογα μεταξύ των δύο σπασμένων άκρων του DNA. Η πρώτη πρωτεΐνη που αναγνωρίζει τη ρήξη της διπλής έλικας του DNA φαίνεται πως είναι η Ku, ένα ετεροδιμερές που αποτελείται από δύο πολυπεπτιδικές αλυσίδες των 70kDa και 80kDa. Το ετεροδιμερές προσδένεται και στα δύο άκρα του σπασμένου DNA και εν συνεχεία σχηματίζει σύμπλοκο με την καταλυτική υπομονάδα της DNA-PKcs εξαρτώμενης πρωτεϊνικής κινάσης, μεταναστεύοντας προς τα δίκλινα ρήγματα. Η DNA-PKcs είναι αναγκαία για την προσέλευση και φωσφορυλίωση πρωτεϊνών που επεξεργάζονται τα σπασμένα άκρα του DNA. Τις περισσότερες φορές τα άκρα από τη ρήξη του δίκλωνου DNA δεν καταλήγουν σε ένα ζεύγος

βάσεων (blunt ended DNA) αλλά αποτελούνται από ένα μονόκλωνο τμήμα στο 3' ή 5' που προεξέχει. Ένα πρωτεϊνικό σύμπλοκο που περιέχει την ενδονουκλεάση Artemis φαίνεται να εμπλέκεται στο μηχανισμό επιδιόρθωσης του δίκλωνου DNA. Το σύμπλοκο αυτό αποτελείται από τις πρωτεΐνες Mre11p, Rad50p και Nbs1. Η τελευταία φαίνεται να σηματοδοτεί την παρουσία βλάβης στα σημεία ελέγχου του κυτταρικού κύκλου και να ρυθμίζει την ενεργότητα του συμπλόκου Mre11p-Rad50p. Το τελευταίο στάδιο στην επιδιόρθωση του DNA είναι η συγκόλληση των επεξεργασμένων άκρων η οποία επιτελείται από το σύμπλοκο DNA λιγάση IV και XRCC4 (Featherstone and Jackson).

1.2.4 Συσχέτιση με την Απόκριση του Ανοσοποιητικού

Σημαντική εξέλιξη έχει πραγματοποιηθεί στην κατανόηση του ρόλου που διαδραματίζει το DNA στην ενεργοποίηση του ανοσοποιητικού συστήματος. Όταν ξένο DNA προσβάλλει το κύτταρο ξενιστή επάγεται η έκφραση κάποιων γονιδίων γνωστών και ως ιντερφερόνες (IFN) που ενεργοποιούν το έμφυτο ανοσοποιητικό σύστημα και διεγείρουν φλεγμονώδης αποκρίσεις. Τέτοιες αποκρίσεις έχει διαπιστωθεί ότι πυροδοτούνται και από βλάβες του ενδογενούς DNA, όχι μόνο στον άνθρωπο αλλά και σε άλλα είδη. Η ενεργοποίηση του ανοσοποιητικού ως απόκριση στις βλάβες του DNA περιλαμβάνει μια σειρά διαδοχικών διεργασιών, χωρίς όμως αυτές να έχουν αποσαφηνιστεί πλήρως. Στις διεργασίες αυτές εμπλέκονται πρωτεΐνες ανίχνευσης της βλάβης, μεταγωγής σήματος και τελεστών που καθορίζουν την τελική κυτταρική απόκριση.

Όσο αφορά τις πρωτεΐνες εντοπισμού της βλάβης, όπως η Ku70 και η DNA-PK που λειτουργούν ως υποδοχείς ανίχνευσης παθογόνων (PRRs), δεν περιορίζουν τη λειτουργία τους μόνο στο πυρηνικό DNA αλλά ανιχνεύουν την εισβολή ξένου DNA στο κυτταρόπλασμα επάγοντας την

παραγωγή ιντερφερονών, κυτοκινών και χημειοκινών ενεργοποιώντας μονοπάτια σηματοδότησης του ανοσοποιητικού συστήματος. Επιπλέον ενδείξεις προκύπτουν από σηματοδοτικά πολυπρωτεϊνικά σύμπλοκα μεταξύ της Rad50 πρωτεΐνης ανίχνευσης βλαβών με τον προσαρμογέα του έμφυτου ανοσοποιητικού συστήματος CARD9 έπειτα από επιμόλυνση βακτηριακού ή dsDNA. Ο σχηματισμός του συμπλόκου dsDNA-Rad50-CARD9 επάγει την ενεργοποίηση του NF-κB μεταγραφικού παράγοντα και την παραγωγή της προφλεγμονώδης κυτοσίνης pro-IL-1β. Ο NF-κB είναι ένας ισχυρός μεταγραφικός παράγοντας, η δραστηριότητα του οποίου παρεμποδίζεται από μια μη φωσφορυλιωμένη μορφή της κυτταροπλασματικής πρωτεΐνης IκB προκαλώντας την αποδέσμευση του NF-κB ο οποίος μεταναστεύει στον πυρήνα. Η παρουσία του NF-κB στον πυρήνα επάγει τη μεταγραφή πολλών γονιδίων που είναι απαραίτητα για την αποτελεσματική λειτουργία της έμφυτης ανοσίας. Ο μηχανισμός DDR δεν επάγει μόνο την παραγωγή προφλεγμονωδών σημάτων όπως οι ιντερφερόνες (IFNs) αλλά και την παραγωγή πρωτεΐνων (ligands) που έχουν την ικανότητα να προσδένουν στους ανοσολογικούς υποδοχείς. Από όλα αυτά τα ευρήματα καταδεικνύεται ότι ενεργοποιείται μια σειρά αντιδράσεων του ανοσοποιητικού που σχετίζεται με τους μηχανισμούς DDR (Nakad and Schumacher).

1.3 Το *Arabidopsis Thaliana* ως Οργανισμός Μοντέλο

Το φυτό *Arabidopsis Thaliana* ανακαλύφθηκε από τον Johannes Thal στα βουνά της Γερμανίας το 16° αιώνα, ο οποίος το ονόμασε *Pilosella siliquosa*. Μεσολάβησαν αρκετά ονόματα ώσπου τελικά να επικρατήσει το όνομα *Arabidopsis Thaliana*. Ένα ετήσιο, μικρό σε μέγεθος φυτό που ευδοκιμεί ως ζιζάνιο στους αγρούς, η *A. thaliana* ανήκει στην οικογένεια Brassicaceae και είναι συγγενικό με το γογγύλι και το λάχανο (Μπαρδούτσος N. et al, 2005). Εξαιτίας μιας σειράς φυσικών

χαρακτηριστικών η *A. thaliana* αποτελεί ένα εξαιρετικό μοντέλο στη βιολογία φυτών. Αρχικά, είναι πάρα πολύ μικρό σε μέγεθος ενώ και ο κύκλος ζωής του είναι πολύ σύντομος. Επιπλέον, αυτογονιμοποιείται και δίνει μεγάλο αριθμό απογόνων ενώ το μικρό διπλοειδές γονιδίωμα του μειώνει σημαντικά το χρόνο και την εργασία που απαιτούνται για την κλωνοποίηση και τη μεταχείριση των γονιδίων του. Τέλος, ο μικρός αριθμός χρωμοσωμάτων του απλουστεύει τη γενετική χαρτογράφηση (Meinke).

1.3.1 Το Γονιδίωμα του Φυτού *Arabidopsis Thaliana*

Η *A. thaliana* ήταν το πρώτο πλήρες πυρηνικό γονιδίωμα ανώτερου φυτού που αλληλουχήθηκε. Περίπου 146Mb εκατομμύρια ζεύγη βάσεων απαρτίζουν το σχετικά μικρό γονιδίωμα της, το οποίο κατανέμεται σε πέντε χρωμοσώματα. Έχει βρεθεί ότι τα γονίδια της *A. thaliana* που κωδικοποιούν πρωτεΐνες περιέχουν κατά μέσο όρο 5.4 εξόνια, μέσου μήκος περίπου 276bp, που διαχωρίζονται από σχετικά μικρά σε μήκος ιντρόνια. Μικρή επίσης είναι και η διαγονιδιακή τους απόσταση. Ένα ιδιαίτερο χαρακτηριστικό γνώρισμα των γονιδίων του φυτού αυτού είναι η περιεκτικότητα των εξονίων τους σε GC (44%) η οποία είναι υψηλότερη από αυτή των ιντρονίων (32%).

Η δομή του γονιδιώματος της *A. thaliana* αποκαλύπτει διπλασιασμούς τόσο σε τοπικό όσο και σε επίπεδο ολόκληρου γονιδιώματος. Πιθανώς να υπήρξαν τρεις πολυπλοειδισμοί, οι εκτιμώμενοι χρόνοι των οποίων ποικίλλουν σημαντικά. Διαθέτει επίσης μια μιτοχονδριακή και μια χλωροπλαστική αλληλουχία υπό τη μορφή κυκλικού DNA. Τα κυτταροπλασματικά αυτά οργανίδια των φυτικών κυττάρων φέρουν μεγαλύτερο μεταβολικό φορτίο, σε σύγκριση με τα ζωικά κύτταρα, εξαιτίας των δραστηριοτήτων των χλωροπλαστών. Η γονιδιωματική ανάλυση των χλωροπλαστών αποκάλυψε μια σχετικά μεγάλη πυκνότητα

συντηρημένων γονιδίων ενώ μια πιο αραιή παρουσία γονιδίων στα μιτοχόνδρια. Και στις δύο περιπτώσεις η ανάλυση υπέδειξε λιγότερα ιντρόνια σε σχέση με το πυρηνικό γονιδίωμα του φυτού.

Η πλειοψηφία του πρωτεώματος της *A. thaliana* αντιστοιχίζεται σε γονίδια τα οποία σχετίζονται με βασικές διεργασίες των φυτών, συμπεριλαμβανομένων όσων εμπλέκονται στη διαδικασία της φωτοσύνθεσης και στο μεταβολισμό των συστατικών των κυτταρικών τοιχωμάτων. Η *A. thaliana* παρουσιάζει αφθονία σε γονίδια που κωδικοποιούν κανάλια μεταφοράς νερού, μεταφορείς πεπτιδίων-ορμονών, μεταβολικά και βιοσυνθετικά ένζυμα και πρωτεΐνες που εμπλέκονται στην άμυνα, την αποτοξίνωση και την περιβαλλοντική ανίχνευση. Τα φυτά διαθέτουν ειδικές μεταβολικές οδούς, για τη φωτοσύνθεση και για το μεταβολισμό των συστατικών του κυτταρικού τοιχώματος, των αλκαλοειδών και των ρυθμιστικών παραγόντων όπως οι αυξίνες και οι γιββερελλίνες. Η σύνθετη μεταβολική τους δραστηριότητα, προϋποθέτει ένα μεγάλο και ποικίλο σύνολο ενζύμων. Επιπλέον, εξαιτίας του ουσιαστικού ρόλου του φωτός στα φυτά, η *A. thaliana* διαθέτει πολλούς αισθητήρες που ρυθμίζουν την ανάπτυξη και τις αποκρίσεις του κερκάδιου ρυθμού.

Τα φυτά απειλούνται επίσης από παθογόνους μικροοργανισμούς και έχουν αναπτύξει αμυντικούς μηχανισμούς οι οποίοι διαφέρουν από το ανοσοποιητικό μας σύστημα. Ένα όπλο που τα φυτά παρατάσσουν ενάντια στα παθογόνα περιλαμβάνει την παραγωγή αντιδραστικών μορφών οξυγόνου. Τα φυτά συνθέτουν πρόσθετα αμυντικά μόρια σε σχέση με αυτά των ζώων αλλά και μόρια που προσελκύουν τους επικονιαστές. Αυτές οι προσελκυστικές ουσίες αποτελούν χρήσιμες πηγές γεύσεων, αρωμάτων και φαρμάκων ενοποιώντας την

παραδοσιακή φυτοθεραπεία με τη σύγχρονη φαρμακολογία (Feldmann and Goff; Lesk).

1.3.2 Σύγκριση με τον Άνθρωπο

Συγκρίνοντας τις πρωτεΐνες που κωδικοποιούνται από το πυρηνικό γονιδίωμα της *A. thaliana* με τις ανθρώπινες πρωτεΐνες, το ποσοστό της ομολογίας που παρατηρείται ποικίλλει ανάλογα με τη λειτουργική κατηγορία στην οποία εμπίπτει. Όσο αφορά την πρωτεϊνοσύνθεση, το 60% των πυρηνικά κωδικοποιημένων γονιδίων της *A. thaliana* εμφανίζει ομολογία με τα ανθρώπινα γονίδια. Για τη ρύθμιση της μεταγραφής, το ποσοστό αυτό μειώνεται στο 30%. Το γεγονός αυτό δεν οφείλεται στο ότι η μεταγραφική διαδικασία δεν εκπροσωπείται ισχυρά στα φυτικά γονιδιώματα, αλλά στο ότι στα φυτά διεξάγεται με διαφορετικό τρόπο. Στην πραγματικότητα, τα φυτά διαθέτουν πολύ περισσότερους μεταγραφικούς παράγοντες συγκριτικά με άλλους οργανισμούς μοντέλα. Τέλος, σχετικά με τις συνιστώσες των μονοπατιών μεταγωγής σήματος, τα φυτά έχουν αναπτύξει οικογένειες ειδικών μεταγραφικών παραγόντων, που δεν εμφανίζουν ομολογία με τις γνωστές συνιστώσες μεταγωγής σήματος σε άλλους οργανισμούς.

Πολλά γονίδια της *A. thaliana* είναι ομόλογα ανθρωπίνων γονιδίων τα οποία εμπλέκονται σε διάφορες ασθένειες. Τα φυτά και τα θηλαστικά εμφανίζουν παρόμοια συστήματα επιδιόρθωσης του DNA, κάτι που συνεπάγεται και ομολογία μεταξύ των γονιδίων. Ένα χαρακτηριστικό παράδειγμα ομολογίας μεταξύ της *A. thaliana* και του ανθρώπου είναι το γονίδιο BRCA2 το οποίο συμμετέχει στο μονοπάτι του ομόλογου ανασυνδυασμού. Για κάποια γονίδια που σχετίζονται με την εμφάνιση ασθενειών στον άνθρωπο, η ομολογία που παρουσιάζουν οι πρωτεΐνες των φυτών σε σχέση με τις πρωτεΐνες των υπολοίπων οργανισμών όπως της *D.melanogaster* ή του *C.elegans* είναι μεγαλύτερη (Lesk).

Ομολογία ανάμεσα σε φυτά και άνθρωπο φαίνεται πως παρουσιάζουν και οι μηχανισμοί απόκρισης στις βλάβες του DNA (DDR) έπειτα από έκθεση σε ακτινοβολία. Οι γενοτοξικές βλάβες που επάγονται από ακτινοβολία οδηγούν στην ενεργοποίηση των μονοπατιών DDR, που χαρακτηρίζονται από ένα σύνολο συνεργατικών λειτουργιών που είτε θα επιδιορθώσουν τις βλάβες του DNA είτε θα κατευθύνουν το κύτταρο σε απόπτωση. Αυτοί οι εξελικτικά συντηρημένοι μηχανισμοί ανάμεσα σε φυτά και άνθρωπο αποτελούν εξέχων παράδειγμα της αυτογνωσίας του DNA να διατηρεί τη γονιδιωματική του ακεραιότητα, ενεργοποιώντας ένζυμα και μονοπάτια επιδιόρθωσης (Elledge). Κατά τη διάρκεια της εξέλιξης, τα φυτά έχουν αποκτήσει ένα σύνολο γονιδίων που είτε συμμετέχουν στα μονοπάτια επιδιόρθωσης του DNA με σαφή προφίλ έκφρασης είτε συνεισφέρουν ως σύμπλοκα στα μονοπάτια σηματοδότησης που μεσολαβούν από την ανίχνευση βλαβών του DNA. Αυτά τα γονίδια παρουσιάζουν μεγάλες ομοιότητες με τα αντίστοιχα ανθρώπινα ως απόκριση στην ακτινοβολία, μολονότι τα επίπεδα ραδιοευαισθησίας διαφέρουν σημαντικά. Στον άνθρωπο τα επίπεδα ανοχής στην ακτινοβολία κυμαίνονται από 0,001-1Gy ενώ στα φυτά από 1-100Gy (Nikitaki et al.)

Πιο συγκεκριμένα, ορθόλογα των γονιδίων ATM και ATR που είναι γνωστά για την ενεργοποίησή τους κατά την ανίχνευση βλαβών του DNA στον άνθρωπο, έχουν βρεθεί στο γονιδίωμα της *A. Thaliana*. Στα φυτά, η πλήρης απώλεια της φυσιολογικής λειτουργίας των γονιδιακών προϊόντων τους δεν επηρεάζει την επιβίωσή τους ενώ μεταλλάξεις των ATM/ATR παρουσιάζουν υπερευαισθησία στην ακτινοβολία-γ, τόσο στην *A. thaliana* όσο και στον άνθρωπο (Culligan et al.). Ευαισθησία στην ακτινοβολία σχετίζεται με μειωμένη ικανότητα αποτελεσματικής επιδιόρθωσης του DNA άρα και ενεργοποίησης των μονοπατιών DDR (Borràs-Fresneda et al.). Σε κάθε περίπτωση, αυτό που αναμένουμε

έπειτα από έκθεση σε ιοντίζουσα ακτινοβολία είναι η δημιουργία complex DNA damage. Οι συσσωρευμένες αυτές βλάβες θα πυροδοτήσουν διάφορα μονοπάτια επιδιόρθωσης, με τον ομόλογο ανασυνδυασμό να αποτελεί το κύριο μονοπάτι επιδιόρθωσης για δίκλωνες θραύσεις του DNA. Τα γονίδια που φέρεται να συμμετέχουν στο μονοπάτι του ομόλογου ανασυνδυασμού στα φυτά, όπως τα BRCA1 και PARP1 παρουσιάζουν ομολογία με τον άνθρωπο.

Για αυτό το λόγο επιλέξαμε να μελετήσουμε τη διαφορική έκφραση των γονιδίων του φυτού *A. thaliana* ύστερα από έκθεση σε ιοντίζουσα ακτινοβολία. Τα δημοσίως διαθέσιμα δεδομένα προέρχονται από πειράματα μικροσυστοιχιών.

1.4 Μικροσυστοιχίες DNA

Οι μικροσυστοιχίες DNA ή DNA microarrays ήταν μέχρι πρότινος η πιο διαδεδομένη μέθοδος ανάλυσης της γονιδιακής έκφρασης. Η ευκολία στο χειρισμό, η μεγάλη διάχυση της τεχνολογίας και το σχετικά χαμηλό κόστος της είναι οι κύριοι λόγοι για τη μεγάλη τους εξάπλωση. Παρά την ολοένα αυξανόμενη χρήση μεθόδων αλληλούχισης επόμενης γενιάς, η μεγάλη πλειοψηφία των πειραμάτων έκφρασης που είναι διαθέσιμα έχουν διενεργηθεί με τη χρήση μικροσυστοιχιών (Duggan et al.). Σε ένα πείραμα γονιδιακής έκφρασης σε μικροσυστοιχία DNA, το πλακίδιο σχεδιάζεται ώστε να περιέχει χαρακτηριστικά τμήματα DNA από όσο το δυνατόν περισσότερα γονίδια του υπό μελέτη οργανισμού. Τα τμήματα αυτά ονομάζονται ανιχνευτές (probes) και μπορούν έτσι να υβριδοποιηθούν με συμπληρωματικά τμήματα cDNA των αντίστοιχων γονιδίων. Τόσο η ανίχνευση όσο και η ποσοτικοποίηση του mRNA γίνεται μέσω της μέτρησης φθορισμού που εκπέμπεται από τη στιγμή που τα δύο συμπληρωματικά μόρια υβριδοποιηθούν. Ο φθορισμός προκύπτει καθώς το δείγμα, πριν περάσει πάνω από το πλακίδιο, έχει

σημανθεί με μια συγκεκριμένη φθορίζουσα χρωστική, η οποία ενεργοποιείται και εκπέμπει μόνο σε δίκλωνη διαμόρφωση.

1.4.1 Πλατφόρμες Μικροσυστοιχιών

Υπάρχουν δύο τύποι DNA μικροσυστοιχιών, οι cDNA μικροσυστοιχίες όπου προκατασκευασμένα τμήματα cDNA τυπώνονται σε γυάλινο πλακίδιο και οι ολιγονουκλεοτιδικές στις οποίες πραγματοποιείται *in situ* σύνθεση των ολιγονουκλεοτιδίων. Και οι δύο είναι βασισμένοι στην αρχή της συμπληρωματικότητας, επιδεικνύοντας με αυτό τον τρόπο τη σχετική αφθονία των mRNA μορίων στο δείγμα (Herrera and Gancino).

Στην παρούσα διπλωματική τα πειράματα διενεργήθηκαν με την *in situ* σύνθεση ολιγονουκλεοτιδίων σε μικροσυστοιχίες υψηλής πυκνότητας όπως οι GeneChips της Affymetrix. Οι ολιγονουκλεοτιδικοί ανιχνευτές που αποτελούνται από μόλις 20-25 ζεύγη βάσεων (bp), συντίθενται απευθείας σε επιφάνεια πυριτίου μόλις 1.2cm² τετραγωνικών εκατοστών. Κάθε ολιγονουκλεοτίδιο είναι τοποθετημένο σε μία γνωστή θέση της συστοιχίας γνωρίζοντας έτσι την ταυτότητα κάθε γονιδίου. Για τον οργανισμό μοντέλο *Arabidopsis Thaliana* που μελετάμε, περισσότερες από 24.000 αλληλουχίες γονιδίων είναι ακινητοποιημένες πάνω στο γυάλινο υπόστρωμα της GeneChip *Arabidopsis* ATH1 Genome Array. Συνήθως 11 πανομοιότυπα ολιγονουκλεοτίδια αντιστοιχούν σε κάθε γονίδιο βελτιώνοντας την ευαισθησία, την εξειδίκευση και τη στατιστική ακρίβεια της έντασης υβριδισμού για την εξαγωγή αποτελεσμάτων σχετικά με την έκφραση κάθε γονιδίου (Dalma-Weiszhausz et al.). Η χρήση των probe sets αυξάνει επιπλέον την ακρίβεια. Κάθε ανιχνευτής (probe) αποτελείται από ένα ζεύγος ολιγονουκλεοτιδίων στο οποίο η μία αλληλουχία υβριδοποιείται επακριβώς με το γονίδιο στόχο PM (perfect match) ενώ

η άλλη αλληλουχία του ζεύγους MM (mismatch) διαφέρει από την PM αλληλουχία κατά μία βάση στο κέντρο του 25μερούς (Miller and Tang).

1.4.2 Ανάλυση Δεδομένων Μικροσυστοιχιών

Οι μετρήσεις φθορισμού που προκύπτουν από ένα πείραμα μικροσυστοιχιών διατηρούν κάποια χαρακτηριστικά που δυσχεραίνουν την περαιτέρω ανάλυση. Πιο συγκεκριμένα, οι ανεπεξέργαστες τιμές φθορισμού εμφανίζουν μεγάλη διασπορά σε ό,τι αφορά το εύρος της έντασης, με πολλές τιμές να είναι εξαιρετικά μικρές και λίγες να είναι πολύ μεγάλες. Κάτι τέτοιο είναι προβληματικό καθώς ιδανικά και προκειμένου τα πειράματα να είναι συγκρίσιμα, θα πρέπει η διασπορά των τιμών να είναι ανεξάρτητη της έντασης φθορισμού. Οι τιμές δηλαδή θα πρέπει να κατανέμονται με όσο το δυνατόν πιο ομοιόμορφο τρόπο. Προκειμένου να μειωθεί η επίδραση αυτής της τάσης για διασπορά που σχετίζεται με την ένταση, οι αρχικές τιμές ενός πειράματος μικροσυστοιχιών μετασχηματίζονται με τη λήψη λογάριθμου. Η λογαρίθμηση των τιμών φθορισμού οδηγεί σε μια προσεγγιστικά κανονική κατανομή των τιμών που κάνει τον περαιτέρω χειρισμό τους πιο εύκολο.

1.4.2.1 Αλγόριθμος RMA

Μια πληθώρα πηγών μεταβλητότητας και θορύβου είναι δυνατόν να υπεισέλθει στα πειράματα μικροσυστοιχιών. Τα τρία βασικά βήματα για την προεπεξεργασία των δεδομένων από πειράματα μικροσυστοιχιών είναι τα εξής:

- Διόρθωση Υποβάθρου
- Κανονικοποίηση

- Σύνοψη Πολλαπλών Χαρτογραφήσεων

Ο αλγόριθμος RMA της Affymetrix GeneChip πραγματοποιεί διαδοχικά τις παραπάνω διαδικασίες. Ο συγκεκριμένος αλγόριθμος σχεδιάστηκε για να χρησιμοποιεί τη νέα τεχνολογία των chip που πραγματοποιείται μόνο PM (perfect match) υβριδισμό. Με αυτό τον τρόπο μειώνεται η διασπορά. Παρόλο που οι μισοί ανιχνευτές δε χρησιμοποιούνται ο αλγόριθμος δίνει καλά αποτελέσματα μειώνοντας την ακρίβεια αλλά αυξάνοντας την πιστότητα.

Διόρθωση Υποβάθρου

Η διόρθωση υποβάθρου (background correction) αποτελεί μία μέθοδο αφαίρεσης του θορύβου στο υπόβαθρο από τις εντάσεις σήματος προερχόμενες από κάθε spot εντός της μικροσυστοιχίας. Ο θόρυβος υποβάθρου μπορεί να προκύψει από διάφορες πηγές μεταξύ αυτών η μη ειδική πρόσδεση του επισημασμένου εκχυλίσματος προς υβριδοποίηση ή η ανομοιόμορφη κατανομή του στην επιφάνεια της μικροσυστοιχίας. Άλλες πηγές μεταβλητότητας αποτελούν η ατελής έκπλυση μετά την υβριδοποίηση κατά την πειραματική διαδικασία καθώς και ο θόρυβος που υπεισέρχεται από το σύστημα σάρωσης (Irizarry).

Κανονικοποίηση Ποσοστημορίου

Πολλοί μέθοδοι έχουν προταθεί για κανονικοποίηση των GeneChip μικροσυστοιχιών της Affymetrix, με τη μέθοδο της κανονικοποίησης ποσοστημορίου (quantile normalization) να αποδίδει καλύτερα. Σκοπός της κανονικοποίησης ποσοστημορίου είναι να δημιουργήσει πανομοιότυπες κατανομές των τιμών των εντάσεων έκφρασης σε όλες τις μικροσυστοιχίες ως προς τις στατιστικές τους ιδιότητες (Bolstad et al.). Στην περίπτωση του RMA για να κανονικοποιηθούν δύο η περισσότερες κατανομές μεταξύ τους, αρχικά πραγματοποιείται μια

ταξινόμηση κατά αύξουσα σειρά και έπειτα ορίζεται ο αριθμητικός μέσος όρος των κατανομών. Στη συνέχεια, ο υψηλότερος μέσος όρος που προέκυψε αντικαθιστά τις υψηλότερες τιμές κάθε κατανομής, ο δεύτερος υψηλότερος μέσος όρος τις αμέσως επόμενες στη διάταξη τιμές και ούτω καθεξής.

Σύνοψη Πολλαπλών Χαρτογραφήσεων

Σειρά έχει η διαδικασία της σύνοψης (summarization) κατά την οποία πραγματοποιείται ταυτοποίηση των τιμών έντασης των γονιδίων που έχουν υβριδοποιηθεί με τα αντίστοιχα ονόματα τους. Τα γονίδια αυτά βρίσκονται σε προκαθορισμένες θέσεις πάνω στη μικροσυστοιχία, οι οποίες βασίζονται σε ειδικά annotation files. Στα ειδικά αυτά αρχεία καταγράφεται η ακριβή τοποθεσία των γονιδίων πάνω στο chip. Σε κάθε spot τοποθετούνται πολλαπλοί ιχνηθέτες για να ποσοτικοποιήσουν το γονιδιακό στόχο. Για τα GeneChip της Affymetrix χρησιμοποιείται ένα σύνολο 11 έως 20 ιχνηθετών (probes) σε κάθε spot αποτελούμενα από 25μερή ολιγονουκλεοτιδίων γνωστών γονιδίων. Κάθε γονίδιο όμως αποτελείται από πολλαπλά μετάγραφα που βρίσκονται σε διαφορετικές θέσεις πάνω στο πλακίδιο. Το σύνολο των probes που αφορούν μετάγραφα του ίδιου γονιδίου συνοψίζονται σε ένα probeset για να δώσουν την τελική τιμή έντασης του γονιδίου (Chen et al.).

1.4.2.2 Principal Component Analysis

Για την ανάλυση δεδομένων γονιδιακής έκφρασης που προέρχονται από περισσότερες από δύο συνθήκες, με σκοπό να εξάγουμε ομάδες γονιδίων με κοινά χαρακτηριστικά σε ότι αφορά τα επίπεδα έκφρασης χρησιμοποιούμε τη μέθοδο της Ανάλυσης Κυρίων Συνιστωσών (Principal Component Analysis - PCA). Η μέθοδος αυτή αποσκοπεί στη μείωση διαστασιμότητας ή πιο απλά στη μείωση της πολυπλοκότητας στο χώρο των διαστάσεων. Με τον όρο διαστάσεις αναφερόμαστε στη δομή των δεδομένων που περιλαμβάνουν πολύ συχνά τιμές έκφρασης για

χιλιάδες ή για εκατοντάδες χιλιάδες γονιδίων και για μεγάλο αριθμό διαφορετικών συνθηκών. Αυτό ουσιαστικά συνιστά ένα αντικείμενο με τόσες διαστάσεις για κάθε γονίδιο όσος και ο αριθμός των συνθηκών. Στόχος της PCA είναι να εντοπίσει την οπτική γωνία παρατήρησης του αντικειμένου που προσφέρει τη μεγαλύτερη δυνατή πληροφορία. Η αρχή της μεθόδου βασίζεται σε μια διαδοχική επανάληψη της προβολής των δεδομένων στη διάσταση που μεγιστοποιεί την διασπορά των τιμών. Ο μετασχηματισμός των διαστάσεων σε νέες διαστάσεις δίνει πάντα μία κύρια συνιστώσα και αντιστοιχεί ουσιαστικά σε μία ορθογώνια προβολή των σημείων σε έναν νέο άξονα. Η πρώτη κύρια συνιστώσα θα είναι πάντα αυτή με τη μεγαλύτερη διακριτική ικανότητα ενώ ο αριθμός των κυρίων συνιστωσών που μπορούν να υπολογιστούν είναι ο αριθμός των αρχικών διαστάσεων μείον ένα. Συνήθως οι δύο πρώτες αρκούν για να προκύψουν ικανοποιητικά συμπεράσματα. Ως μέθοδος, η PCA αποτελεί μία από τις πρώτες επιλογές μας για μια διερευνητική ανάλυση των δεδομένων μας. Βασικός της στόχος να αναδείξει τις σημαντικότερες διαφορές μεταξύ των στοιχείων μιας δομής δεδομένων κρατώντας και παρουσιάζοντας εκείνα τα χαρακτηριστικά τους που ευθύνονται για τη μεγαλύτερη ποικιλομορφία στα δείγματά μας (Abdi and Williams).

1.4.2.3 Στατιστική Ανάλυση για Διαφορική Έκφραση Γονιδίων

Ο υπολογισμός του βαθμού της μεταβολής της έκφρασης του ίδιου γονιδίου μεταξύ δύο διαφορετικών συνθηκών γίνεται με τη χρήση του λογάριθμου του λόγου των τιμών έκφρασης στη συνθήκη μελέτης (test) προς τη συνθήκη ελέγχου (control). Υπολογίζουμε το λογάριθμο του λόγου τους ως εξής() :

$$\log_2 FC(g) = \log_2 \frac{E(g)_{test}}{E(g)_{control}}$$

Τιμές του λόγου > 1 θα είναι ενδεικτικές μεγαλύτερης έκφρασης στη συνθήκη μελέτης και επομένως υπερέκφρασης του γονιδίου ενώ τιμές του λόγου < 1 θα είναι ενδεικτικές υποέκφρασης του. Η εφαρμογή του λογάριθμου γίνεται για δύο λόγους. Αρχικά, για να μειώσουμε τη διασπορά των τιμών λόγων έκφρασης και κατά δεύτερον να μετατρέψει το ουδέτερο σημείο της ποσότητας από το 1 στο 0. Οι τιμές \log_2FC είναι θετικές στην περίπτωση της υπερέκφρασης του γονιδίου g και αρνητικές στην περίπτωση υποέκφρασης του. Επιπλέον, η χρήση του δυαδικού λογάριθμου επιτρέπει μια απευθείας ανάγνωση του βαθμού της διαφορικής έκφρασης. Μια τιμή $\log_2FC = 1$ σημαίνει διπλάσια έκφραση σε σχέση με τη συνθήκη ελέγχου, ενώ μια τιμή $\log_2FC = -1$ υποδηλώνει μείωση στο μισό. Στην περίπτωση που η έκφραση του κάθε γονιδίου έχει μετρηθεί περισσότερες από μία φορές σε επαναλήψεις του ίδιου πειράματος η εξίσωση δεν αλλάζει αλλά στη θέση των $E(g)_{test}$ και $E(g)_{control}$ χρησιμοποιούνται οι αντίστοιχες μέσες τιμές των επαναλήψεων. Η σημασία της πραγματοποίησης επαναλήψεων του ίδιου πειράματος είναι πολύ μεγάλη για λόγους στατιστικής αξιολόγησης. Χρειαζόμαστε ένα στατιστικό έλεγχο, ώστε να αξιολογήσουμε σε ποιο βαθμό η διπλάσια έκφραση αντανακλά ένα υπαρκτό βιολογικό φαινόμενο ή μια τυχαία διακύμανση στα επίπεδα του mRNA. Οποσδήποτε ακραίες τιμές είναι ενδεικτικές πραγματικών διαφορών, ωστόσο ένα απλό ερώτημα σχετίζεται με το κατά πόσο μια τιμή είναι επαναλήψιμη. Για αυτό είναι σημαντικό κάθε πείραμα έκφρασης να διενεργείται τουλάχιστον σε τρεις επαναλήψεις, που σημαίνει ότι για κάθε συνθήκη και για κάθε γονίδιο θα πρέπει να έχουμε τουλάχιστον τρεις μετρήσεις έκφρασης (Lee et al.). Ο λόγος για αυτόν τον περιορισμό είναι ότι για να αξιολογήσουμε τη διαφορική έκφραση στατιστικά, θα πρέπει να υπολογίσουμε μια πιθανότητα που να εκτιμά τη διαφορά μεταξύ των μέσων τιμών των μετρήσεων για κάθε συνθήκη. Σε πειράματα έκφρασης, όπου εξετάζουμε μια συνθήκη μελέτης με μια συνθήκη ελέγχου, ο στατιστικός έλεγχος γίνεται με τη

χρήση του ελέγχου t-test που αποτελεί ένα στατιστικό έλεγχο υποθέσεων. Η υπόθεση αυτή που ονομάζεται null hypothesis, στην περίπτωση του t-test σημαίνει ότι οι μέσες τιμές δύο συνόλων τιμών είναι ταυτόσημες. Το t-test υπολογίζει ένα μέγεθος t που όσο μεγαλύτερο είναι τόσο μικρότερη η πιθανότητα οι δύο μέσες τιμές να ταυτίζονται. Σε κάθε περίπτωση η αποδιδόμενη τιμή είναι η πιθανότητα ταύτισης των δύο μέσων τιμών, που αντιστοιχίζεται σε μια τιμή (p-value). Όσο μικρότερη είναι η τιμή p-value τόσο μικρότερη είναι η πιθανότητα οι δύο μέσες τιμές να είναι ίδιες και συνεπώς τα δύο δείγματα να προέρχονται από την ίδια κατανομή. Με αυτό τον τρόπο, μικρές τιμές p-value σε πειράματα έκφρασης είναι ισχυρή ένδειξη ότι οι διαφορές που παρατηρούνται στα επίπεδα έκφρασης ενός γονιδίου είναι στατιστικά σημαντικές. Ο προσδιορισμός της στατιστικά σημαντικής διαφορικής έκφρασης γονιδίων γίνεται ορίζοντας κάποια όρια τιμών \log_2FC και p-value. Ως όριο σημαντικότητας βιβλιογραφικά τίθεται το $p\text{-value} \leq 0.05$ και ως όριο διαφορικής έκφρασης η απόλυτη τιμή $|\log_2FC| > 1.5$. Γονίδια που πληρούν τα συγκεκριμένα όρια ονομάζονται διαφορικά εκφραζόμενα (differentially expressed genes, DEGs). Τα όρια αυτά είναι αυθαίρετα και συχνά χρησιμοποιούνται λιγότερο αυστηρές τιμές κατωφλίων. Ακόμη όμως και χαμηλές τιμές p-value θα πρέπει να αντιμετωπίζονται με προσοχή. Στην περίπτωση που υπολογίζουμε μεγάλο αριθμό από p-values αυτό που κάνουμε είναι να διενεργήσουμε έναν έλεγχο πολλαπλών υποθέσεων. Αυτό που συμβαίνει σε αυτές τις περιπτώσεις είναι ότι για καθαρά στατιστικούς λόγους κάποιες τιμές p-values, που είναι αρκετά μικρές ώστε να χαρακτηρίσουν στατιστικά σημαντικές αλλαγές στα επίπεδα έκφρασης, μπορούν να έχουν προκύψει τυχαία (ψευδώς θετικά αποτελέσματα). Αυτές θα είναι περισσότερες όσο μεγαλύτερος είναι ο αριθμός των πολλαπλών υποθέσεων. Η διόρθωση των αποτελεσμάτων που προκύπτουν από τον έλεγχο πολλαπλών υποθέσεων χρησιμοποιούν μεθόδους που όλες συγκλίνουν στη διαδικασία απόρριψης ενός μέρους

των αρχικά στατιστικά σημαντικών p-values. Για πειράματα γονιδιακής έκφρασης η πιο αποτελεσματική μέθοδος για τη διόρθωση πολλαπλών ερωτημάτων είναι αυτή των Benjamini και Hochberg γνωστή και ως False Discovery Rate FDR (Cui and Churchill).

1.4.2.4 Gene Ontology – Functional Enrichment

Οι σχέσεις γονιδίων-λειτουργιών αποτελούν τη βάση της λειτουργικής ανάλυσης της γονιδιακής έκφρασης. Έπειτα από τη διενέργεια πειραμάτων έκφρασης σε γονιδιωματικό επίπεδο προκύπτει μία λίστα από διαφορετικώς εκφραζόμενα γονίδια. Η ανάγκη για κατηγοριοποίηση αυτών των γονιδίων ανάλογα με τις διαπιστωμένες ή προβλεπόμενες λειτουργίες τους οδήγησε σε αυτό που αποκαλούμε γονιδιακές οντολογίες. Η Γονιδιακή Οντολογία (Gene Ontology) αποτέλεσε ένα από τα πιο σημαντικά πρώτα βήματα προς την επίλυση ενός σημαντικού προβλήματος, την αποτύπωση του συνόλου της βιολογικής γνώσης σχετικά με τη γονιδιακή λειτουργία με ένα σαφή και οργανωμένο τρόπο. Αποτελεί το αποτέλεσμα της δουλειάς του αντίστοιχου Consortium που συστήθηκε το 1988 με σκοπό τη δημιουργία μιας συνεκτικής, παγκόσμιας ονοματολογίας γονιδίων για όλους τους οργανισμούς (Ashburner et al.). Στο πλαίσιο αυτής της προσπάθειας, η βάση δεδομένων της Γονιδιακής Οντολογίας έχει δημιουργήσει τρεις βασικές, ανεξάρτητες μεταξύ τους, κατηγορίες που περιγράφουν τα γονίδια όλων των μελετημένων οργανισμών στη βάση: της κυτταρικής διεργασίας στην οποία συμμετέχουν οι πρωτεΐνες τους, της μοριακής λειτουργίας των πρωτεϊνικών προϊόντων τους και του κυτταρικού εντοπισμού των πρωτεϊνών τους.

Το αμέσως επόμενο βήμα που αποτελεί και πυρήνα της λειτουργικής ανάλυσης είναι η ποσοτικοποίηση του βαθμού εμπλουτισμού μιας λίστας γονιδίων σε διάφορες βιολογικές διεργασίες. Πρόκειται

ουσιαστικά για μία εκτίμηση του βαθμού, βάση του οποίου μπορούμε να αξιολογήσουμε εάν ένα σύνολο γονιδίων που σχετίζονται με συγκεκριμένες βιολογικές διεργασίες έχει προκύψει στην τύχη ή όχι. Για πειράματα γονιδιακής έκφρασης χρησιμοποιούμε ως δεδομένα τη λίστα των διαφορικά εκφραζόμενων γονιδίων και γνώση από βάσεις δεδομένων όπως η Gene Ontology. Από τη λίστα αυτή, που αποτελεί ένα δείγμα γονιδίων έκφρασης εκτιμούμε το βαθμό εμπλουτισμού σε συγκεκριμένες λειτουργίες καθώς και το επίπεδο στατιστικής σημαντικότητας του κάνοντας χρήση της υπεργεωμετρικής κατανομής. Η αθροιστική πιθανότητα της υπεργεωμετρικής κατανομής μπορεί να λειτουργήσει ως εκτιμητής της τυχαιότητας μιας διαδικασίας δειγματοληψίας χωρίς εναπόθεση. Η συγκεκριμένη τιμή χρησιμοποιείται ως τιμή p-value που μπορεί να χρησιμεύσει αργότερα στον έλεγχο υποθέσεων.

2. ΜΕΘΟΔΟΣ & ΔΙΑΔΙΚΑΣΙΑ

2.1 Αναζήτηση Πρωτογενών Δεδομένων

Στην παρούσα διπλωματική, αναζητήσαμε να βρούμε πειράματα γονιδιακής έκφρασης κυρίως με τη μέθοδο των μικροσυστοιχιών. Η αναζήτηση έγινε με βασικό κριτήριο, τα πειράματα να προέρχονται από το φυτό *Arabidopsis Thaliana* και να έχουν εκτεθεί σε ιοντίζουσα ακτινοβολία. Τα αποτελέσματα που προέκυψαν, για το συγκεκριμένο οργανισμό μοντέλο, δε μας έδωσαν πληθώρα τέτοιων υψηλής απόδοσης γενωμικών πειραμάτων σε δημόσια αποθετήρια, όπως η GEO (Gene Expression Omnibus) (Barrett et al.) και η ArrayExpress (Athar et al.). Έτσι, καταλήξαμε σε ένα dataset, κατατεθειμένο στην ArrayExpress.

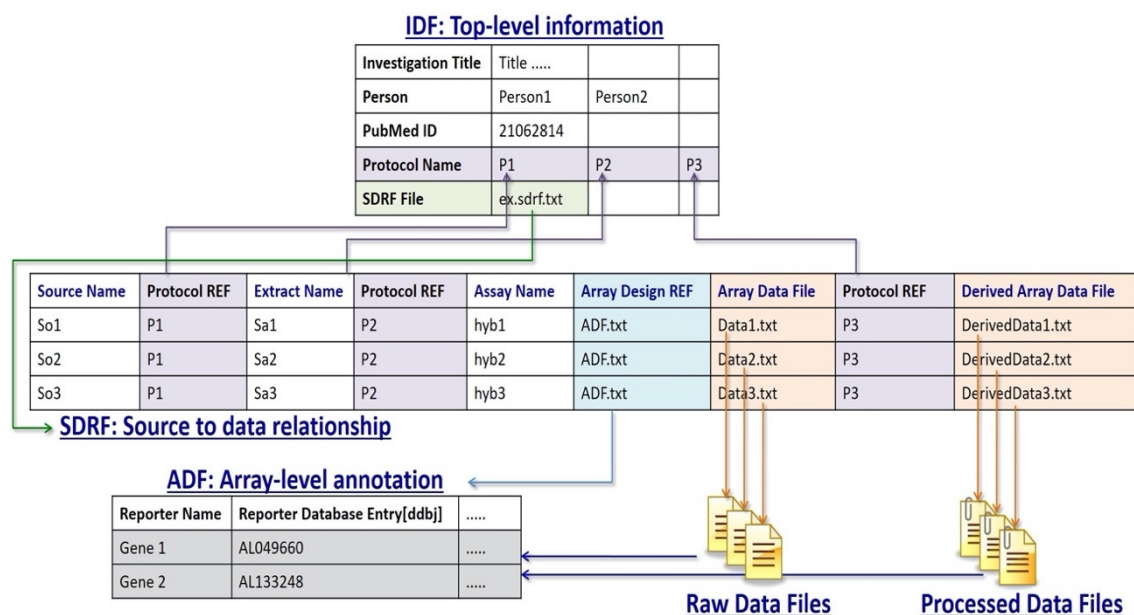
Όπως ήδη αναφέρθηκε, η ArrayExpress αποτελεί ένα δημόσιο αποθετήριο δεδομένων υψηλής απόδοσης λειτουργικής γονιδιωματικής (Parkinson et al.), από όπου έγινε και η αναζήτηση των πρωταρχικών μας δεδομένων. Συμμορφωμένο προς τις απαιτήσεις του MIAME (Brazma et al.), περιέχει τα αρχικά δεδομένα δημοσιευμένων πειραμάτων μικροσυστοιχιών, με ένα δομημένο και τυποποιημένο τρόπο. Για κάθε σύνολο δεδομένων (dataset), αποθηκεύεται στη βάση μία MEGA-TAB δομή που περιέχει πέντε διαφορετικούς τύπους αρχείων (Εικόνα 4):

- Investigation Description Format (IDF)
- Array Design Format (ADF)

- Sample and Data Relationship Format (SDRF)
- The raw data files &
- The processed data files

Όλα τα αρχεία είναι tab-delimited. Το αρχείο IDF περιλαμβάνει βασικές πληροφορίες για το πείραμα, συμπεριλαμβανομένου του τίτλου, της περιγραφής, του πρωτοκόλλου που ακολουθεί και τέλος λεπτομέρειες που αφορούν την υποβολή του πειράματος στη βάση. Το αρχείο ADF περιγράφει το σχεδιασμό μίας διάταξης. Πιο συγκεκριμένα, τη θέση της κάθε ακολουθίας στη διάταξη καθώς και τα στοιχεία σχολιασμού της. Το SDRF αρχείο περιέχει το σχολιασμό των δειγμάτων και τη σχέση μεταξύ των διατάξεων όπως αυτά παρέχονται από τον υποβάλλοντα. Η πληροφορία σε ένα SDRF αρχείο οργανώνεται με τέτοιο τρόπο ώστε να ακολουθεί τη φυσική ροή ενός πειράματος λειτουργικής γονιδιωματικής. Ξεκινά με την περιγραφή των δειγμάτων και ολοκληρώνεται με τα ονόματα των αρχείων δεδομένων που παράγονται από την ανάλυση των πειραματικών αποτελεσμάτων. Για δεδομένα μονοκάναλων μικροσυστοιχιών, μία γραμμή στο SDRF είναι ισοδύναμη με μία υβριδοποίηση. Το raw zip file, περιέχει τα πρωτογενή δεδομένα (όπως αυτά παράγονται από τις μικροσυστοιχίες της Affymetrix) ενώ το processed zip file περιλαμβάνει όλα τα επεξεργασμένα δεδομένα.

Ως λέξη κλειδί στη βασική μας αναζήτηση χρησιμοποιήθηκε ο όρος ionizing radiation και τα αποτελέσματα που προέκυψαν φιλτραρίστηκαν με βάση τον οργανισμό μελέτης μας Arabidopsis Thaliana. Η αναζήτηση έδωσε τρία αποτελέσματα, εκ των οποίων κρίθηκαν κατάλληλα προς ανάλυση μόνο τα δεδομένα του πειράματος με κωδικό E-GEOD-61484. Τα οποία διαθέτουν πληροφορίες σχετικές με την περιγραφή του πειράματος, της διάταξης και των πρωτοκόλλων που χρησιμοποιήθηκαν, τον αριθμό των δειγμάτων και τον οργανισμό μελέτης στον οποίον διενεργήθηκε το πείραμα.



Εικόνα 4. Γενική δομή ενός MEGA-TAB αρχείου (Rocca-Serra et al.).

2.2 Πειραματικός Σχεδιασμός

Σκοπός της μελέτης του πειράματος με κωδικό E-GEOD-61484 (Missirian et al.), ήταν να εξεταστεί η επίδραση δυο διαφορετικών τύπων ιοντίζουσας ακτινοβολίας (IR), HZE και ακτινοβολία-γ στο μεταγράφημα των νεαρών βλαστών της *Arabidopsis Thaliana*. Μετά την ακτινοβολήση των δειγμάτων, το υλικό συλλέχθηκε σε ένα εύρος διαφορετικών χρονικών σημείων που κυμαίνεται από 1.5 έως και 24 ώρες. Ο άγριος τύπος ενός νεαρού βλαστού 5 μόλις ημερών ακτινοβολήθηκε με 100 Gy γάμμα ακτινοβολίας για περίπου 15 λεπτά και 30 Gy HZE ακτινοβολήσης για περίπου 12 λεπτά. Κατέληξαν έτσι, σε ένα σύνολο 22 πειραματικών δειγμάτων, συνθήκης και ελέγχου, με δύο επαναλήψεις (replicates) το καθένα, δίνοντας συνολικά 44 δείγματα. Η πλατφόρμα που χρησιμοποιήθηκε για το μεταγραφικό προφίλ του πειράματος ήταν της Affymetrix, η Arabidopsis Genome ATH1 Array, ειδικά σχεδιασμένη να περιέχει όλα τα γνωστά γονίδια της *Arabidopsis Thaliana*.

2.3 Χρήση της R για την ανάλυση των δεδομένων

Θα ξεκινήσουμε την ανάλυση μας κατεβάζοντας τα ανεπεξέργαστα δεδομένα (raw data) των CEL αρχείων και θα προχωρήσουμε στην εισαγωγή αυτών στην κλάση ExpressionSet της Bioconductor, μίας τυποποιημένης δομής δεδομένων για την αναπαράσταση των γενωμικών μας δεδομένων. Έπειτα θα ελέγξουμε την ποιότητα των δεδομένων μας, θα διενεργήσουμε κανονικοποίηση, για να καταλήξουμε στην ανάλυση της διαφορικής έκφρασης γονιδίων. Η παρούσα ανάλυση πραγματοποιήθηκε βάση προτεινόμενης ροής εργασίας (workflow) για ανάλυση μικροσυστοιχιών (Klaus and Reisenauer).

- R version : R version 4.0.2 (2020-06-22)
- Bioconductor Version : 3.11

Για την εγκατάσταση των πακέτων μέσω της BiocManager απαιτούνται οι ακόλουθες γραμμές κώδικα :

```
>if(!requireNamespace("BiocManager", quietly =TRUE))
  install.packages("BiocManager")
>BiocManager::install("library")
```

Αρχικά εγκαθιστούμε το πακέτο της ArrayExpress . Όπως αναφέρθηκε παραπάνω, το πρώτο βήμα για την ανάλυση των δεδομένων είναι να κατεβάσουμε τα raw data των CEL αρχείων που είναι αποθηκευμένα στο αποθετήριο της ArrayExpress με κωδικό πρόσβασης E-GEOD-61484. Τα αρχεία αυτά δημιουργούνται από ένα λογισμικό σάρωσης μικροσυστοιχιών και περιέχει τις μετρούμενες τιμές έντασης φθορισμού. Μέσω της συνάρτησης getAE() του πακέτου της ArrayExpress (Kauffmann et al.) αποκτούμε τα ftp links για να έχουμε πρόσβαση και να κατεβάσουμε τα raw data μας και μαζί με αυτά όλα

τα αρχεία σχολιασμού σε έναν ειδικό κατάλογο (directory) που ορίσαμε ως *raw_data_dir*. Τα ονόματα των αρχείων θα επιστραφούν ως λίστα.

```
>raw_data_dir<-getwd()
> library(ArrayExpress)
> annoAE<-getAE("E-GEOD-61484", path=raw_data_dir, type=
"raw")
```

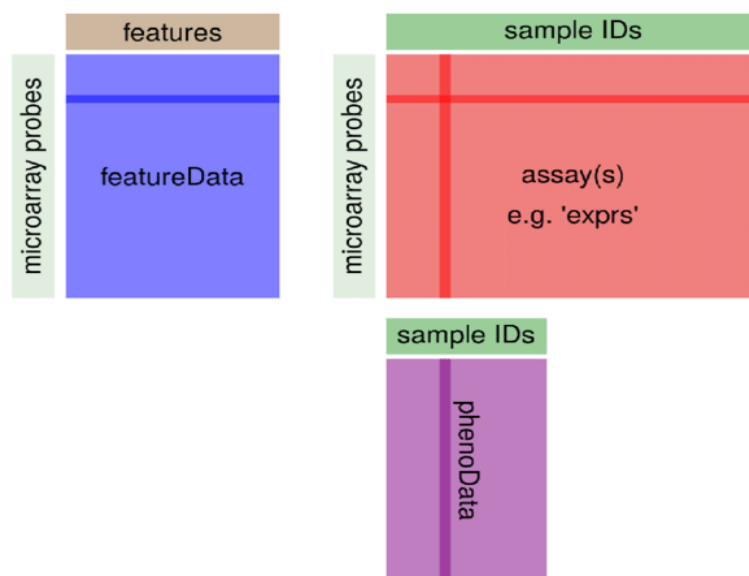
Η Biobase (Huber et al.) μέρος της Bioconductor, είναι ένα πακέτο που περιέχει τυποποιημένες δομές δεδομένων για να αναπαριστά τα γενωμικά δεδομένα. Εξαιτίας του ότι αυτά είναι αρκετά περίπλοκα και αποτελούνται από διαφορετικές συνιστώσες όπως σχολιασμός των γενωμικών δεδομένων καθώς και πληροφορίες για τα πειραματικά δείγματα γεννήθηκε η ανάγκη για σχεδιασμό μιας κλάσης που να συνδυάζει όλες αυτές τις διαφορετικές πηγές πληροφοριών σε μια πιο απλή και πρακτική δομή, την ExpressionSet.

Τα δεδομένα σε ένα ExpressionSet αποτελούνται :

- *assayData* : περιέχει δεδομένα έκφρασης από πειράματα μικροσυστοιχιών, με τα γονίδια να αποτελούν τις γραμμές και τα δείγματα τις στήλες του.
- *metadata*
 - *phenoData* : Περιγραφή των δειγμάτων στο πείραμα με τα δείγματα να αποτελούν τις γραμμές και περιγραφή των στοιχείων του πειράματος στις στήλες. Περιέχει τα περιεχόμενα ενός SDRF αρχείου.
 - *featureData* : μεταδεδομένα για τα γονίδια πάνω στη μικροσυστοιχία ή πληροφορίες σχετικές με την τεχνολογία που χρησιμοποιήθηκε για τα πειράματα με τις γραμμές να αποτελούνται από τα γονίδια (probes) όπως στο *assayData* και ελεύθερα εκχωρηθείς στήλες.

- `experimentData` : Μια ευπροσάρμοστη δομή για να περιγράψει το πείραμα.

Η κλάση `ExpressionSet` της `Bioconductor` συγχρονίζει όλα αυτά τα δεδομένα με τέτοιο αποτελεσματικό τρόπο, υπενθυμίζοντας μας όμως ότι οι γραμμές του `phenoData` θα πρέπει να ταιριάζουν ακριβώς με τις στήλες του `assayData`, καθώς και οι γραμμές του `assayData` με τις γραμμές του `featureData` (Εικόνα 5) (O'Brien).



Εικόνα 5. Δομή της κλάσης `ExpressionSet` της `Bioconductor` (Klaus and Reisenauer).

2.3.1 Εισαγωγή δεδομένων στην κλάση `ExpressionSet`

Τα φαινοτυπικά δεδομένα συνοψίζουν πληροφορίες για τα δείγματα που αναφέρονται σαν `covariates`. Για να αποκτήσουμε το αρχείο με το σχολιασμό των δειγμάτων μας, θα εισάγουμε το αρχείο SDRF με την συνάρτηση `read.delim()`, η οποία διαβάζει `tab-delimited` αρχεία. Η στήλη `Array.Data.File` του αρχείου SDRF, μας δίνει την ακριβή ονομασία των δειγμάτων, την οποία και θα χρησιμοποιήσουμε σαν γραμμές του πίνακα για να κατασκευάσουμε με ορθότητα τη δομή της κλάσης

ExpressionSet. Οι μεταβλητές “covariates” ενός φαινοτυπικού πίνακα ενσωματώνονται στο ExpressionSet ως metadata αφού μετατραπούν σε data frame. Η κλάση AnnotatedDataFrame σχεδιάστηκε από το πακέτο της Biobase για να αποθηκεύει και να διαχειρίζεται όλες αυτές τις πληροφορίες με αυτό τον τρόπο.

```
>pDataFile<-file.path(raw_data_dir,"E-GEOD-61484.sdrf.txt")
>SDRF<-read.delim(pDataFile)
>rownames(SDRF)<-SDRF$Array.Data.File
>SDRF<-AnnotatedDataFrame(SDRF)
```

Εν συνεχεία, δημιουργούμε το αντικείμενο της κλάσης ExpressionSet, το οποίο θα ονομάσουμε raw_data και περιέχει πληροφορίες σχετικά με τις ανεπεξέργαστες τιμές έκφρασης, τα φαινοτυπικά δεδομένα από το αρχείο SDRF καθώς και πληροφορίες του πακέτου που χρησιμοποιήθηκε για το σχολιασμό της πειραματικής διάταξης. Η ανάλυση των μονοκάναλων μικροσυστοιχιών της Affymetrix, ξεκινούν με τα CEL αρχεία. Αυτά είναι τα αποτέλεσμα της επεξεργασίας των πρωτογενών αρχείων εικόνας χρησιμοποιώντας κατάλληλο λογισμικό και περιέχουν τις εντάσεις των τιμών έκφρασης. Από το πακέτο της oligo (Carvalho and Irizarry) θα εισάγουμε τα CEL αρχεία και με τη βοήθεια της συνάρτησης read.celfiles() θα τα διαβάσουμε. Με την αποθήκευση των CEL αρχείων στο αντικείμενο raw_data, συμπληρώσαμε τον πίνακα assayData της κλάσης ExpressionSet, τα οποία διαθέτουν και το σωστό πακέτο σχολιασμού της διάταξης “pd.ath1.121501”. Για να μπορέσουμε να εξασφαλίσουμε ότι τα CEL αρχεία θα διαβασθούν και θα εισαχθούν με την ίδια σειρά που εμφανίζονται στην κλάση AnnotatedDataFrame, χρησιμοποιήσαμε τη στήλη Array.Data.File του πίνακα SDRF σαν όρισμα της συνάρτησης filenames().

```
>library(oligo)
```

```
>raw_data<-read.celfiles(filenamees = file.path(raw_data_dir,
SDRF$Array.Data.File), verbose =FALSE, phenoData = SDRF)
>stopifnot(validObject(raw_data))
>all(rownames(pData(SDRF)) == colnames(exprs(raw_data)))
```

Τέλος, ελέγξαμε εάν το αντικείμενο που δημιουργήσαμε είναι έγκυρο και το κατά πόσο τα δείγματα ταιριάζουν απόλυτα μεταξύ των διαφορετικών πινάκων. Μέσω της συνάρτησης pData() του πακέτου της Biobase έχουμε πρόσβαση στα φαινοτυπικά δεδομένα του αντικειμένου raw_data της κλάσης ExpressionSet. Από τις στήλες του πίνακα που προκύπτουν κρατάμε εκείνες με τους πειραματικούς παράγοντες του ενδιαφέροντος μας. Στήλες του ενδιαφέροντος μας, από το αρχείο SDRF, είναι οι ακόλουθες :

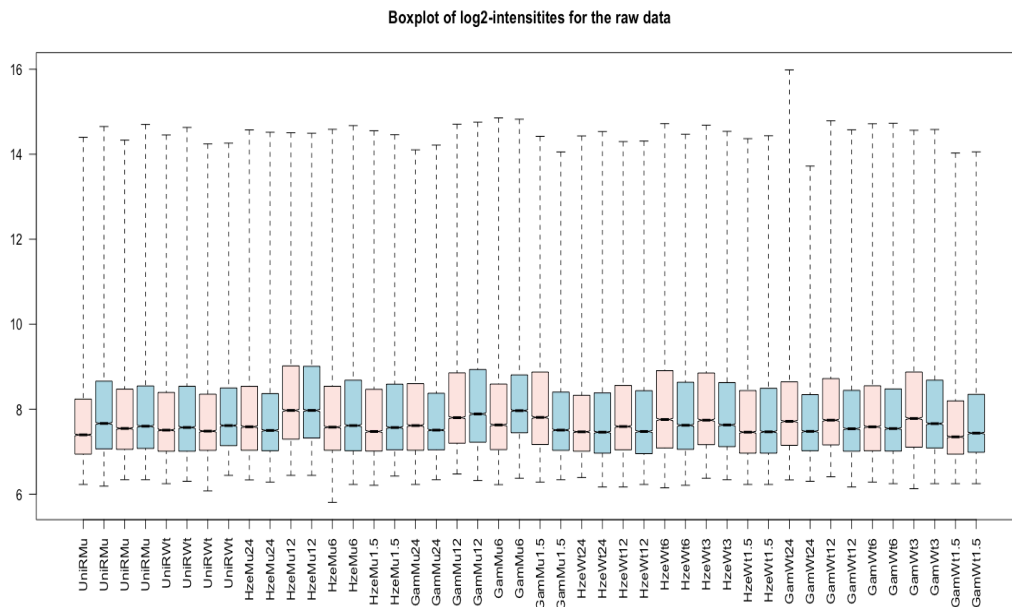
- FactorValue [REPLICATE]
- FactorValue [GENOTYPE]
- FactorValue [TIME AFTER TREATMENT]
- FactorValue [TREATED WITH]

```
>Biobase::pData(raw_data)<-Biobase::pData(raw_data)[, c(42,
43, 45, 46)]
>head(Biobase::pData(raw_data))
```

2.3.2 Έλεγχος Ποιότητας Δεδομένων

Η αξιολόγηση της ποιότητας των δεδομένων αποτελεί ουσιαστική διαδικασία της ανάλυσης μικροσυστοιχιών (Kauffmann and Huber). Μία πρώτη προσέγγιση για την αξιολόγηση αυτών, αποτελεί η χρήση θηκογραμμάτων, πριν και μετά την υλοποίηση του αλγορίθμου RMA. Τα θηκογράμματα απεικονίζουν γραφικά την κατανομή των τιμών των εντάσεων φθορισμού ανά μικροσυστοιχία. Αυτό επιτυγχάνεται με τη βοήθεια της συνάρτησης boxplot() του πακέτου της oligo που δέχεται ως όρισμα τα σύνολα των τιμών έκφρασης φθορισμού. Κάθε θηκόγραμμα αντιστοιχεί σε ένα δείγμα, υλοποιώντας εξορισμού

λογαριθμικό μετασχηματισμό με βάση το 2, για λόγους που έχουν να κάνουν με την εύκολη ερμηνεία των αποδιδόμενων τιμών σε περιπτώσεις σύγκρισης δύο ή περισσότερων δειγμάτων. Η λογαρίθμηση των τιμών φθορισμού έχει ως σκοπό να οδηγήσει σε μια προσεγγιστικά κανονική κατανομή που θα κάνει τον περαιτέρω χειρισμό των δεδομένων μας πιο εύκολη. Από τον παρακάτω κώδικα απεικονίζονται σχηματικά μέσω θηκογραμμάτων οι τιμές έκφρασης φθορισμού των λογαριθμησμένων πλέον δεδομένων μας ανά μικροσυστοιχία (Εικόνα 6). Είναι φανερό η ανάγκη κανονικοποίησης μιας που τα θηκογράμματα ανά μικροσυστοιχία της παρακάτω απεικόνισης δεν ακολουθούν μια προσεγγιστικά κανονική κατανομή, όπως καταλαβαίνουμε από τις θέσεις των διαμέσων των δειγμάτων.



Εικόνα 6. Έλεγχος ποιότητας δεδομένων μέσω θηκογραμμάτων των λογαριθμησμένων πρωτογενών τιμών έκφρασης.

```
>oligo::boxplot(raw_data, main = "Boxplot of log2
intensities for the raw data", names = c("UniRMu", "UniRMu",
"UniRMu", "UniRMu", "UniRwt", "UniRwt", ..., "Gamwt1.5"), las
= 2, col = c("mistyrose", "lightblue"), notch = TRUE)
```

2.3.3 Απεικόνιση RLE για περαιτέρω έλεγχο ποιότητας δεδομένων

Η ανεπιθύμητη διασπορά των τιμών φθορισμού σε ό,τι αφορά το εύρος της έντασης μπορεί να αποδειχθεί εξαιρετικά προβληματική στην ανάλυση δεδομένων μικροσυστοιχιών και η ανίχνευση της καθίσταται καίριας σημασίας. Το φαινόμενο αυτό προέρχεται συνήθως από τεχνικούς παράγοντες που δεν είναι βιολογικού ενδιαφέροντος, όπως η διεξαγωγή των πειραμάτων σε διαφορετικές συνθήκες περιβάλλοντος, οδηγώντας σε παρεκκλίσεις μεταξύ των δειγμάτων (batch effects). Για αυτό το λόγο θα καταφύγουμε σε μία ακόμη διαδικασία για την αξιολόγηση της ποιότητας των δεδομένων μας. Η ανάλυση αυτή θα πραγματοποιηθεί μέσω γραφημάτων οπτικοποίησης Relative Log Expression (RLE) που υλοποιεί τον αλγόριθμο RMA χωρίς να εκτελεί κανονικοποίηση των δεδομένων. Η απεικόνιση γραφημάτων (RLE) βασίζεται στον υπολογισμό της διάμεσου των τιμών έντασης κάθε μεταγράφου μεταξύ των δειγμάτων. Δεν εφαρμόζουμε επιπλέον log2 λογαριθμικό μετασχηματισμό καθώς τα δεδομένα λογαριθμίζονται μέσω της συνάρτησης RMA(). Εν συνεχεία, υπολογίζουμε τις αποκλίσεις από τη διάμεσο μέσω της συνάρτησης sweep() και αποθηκεύουμε τα αποτελέσματα στο αντικείμενο RLE το οποίο αναδιαμορφώνουμε σε data frame με σκοπό τη δημιουργία θηκογραμμάτων (Gandolfo and Speed).

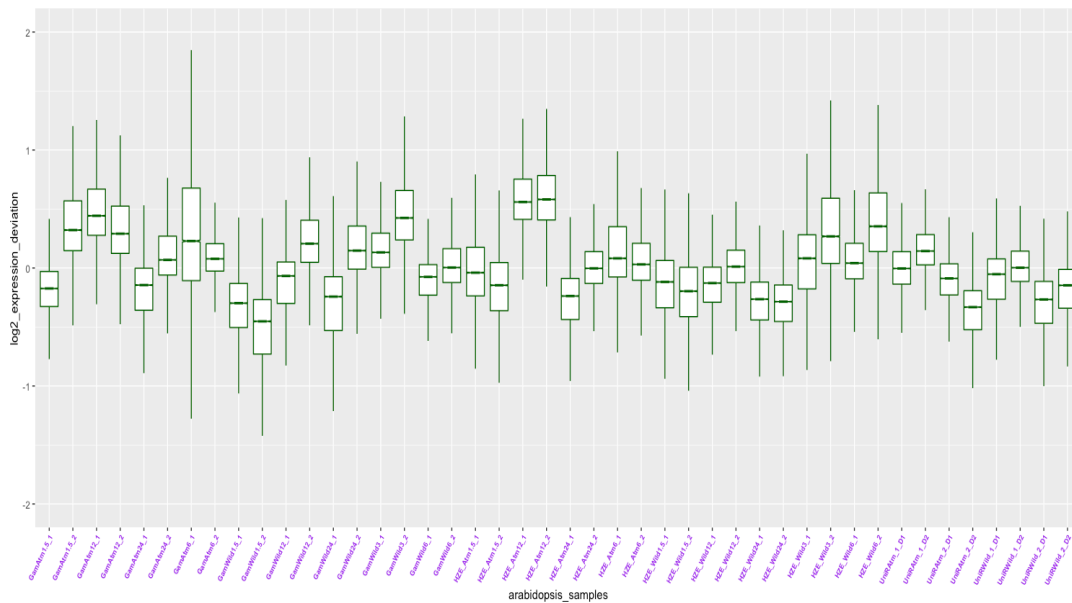
```
>ioan.eSet.NoNorm <- oligo::rma(raw_data, normalize = FALSE)
>transcript_medians<-Biobase::rowMedians(as.matrix(Biobase ::
exprs(ioan.eSet.NoNorm)))
>subtract_median <- sweep(Biobase::exprs(ioan.eSet.NoNorm),
1, transcript_medians, "-")
>RLE <- as.data.frame(subtract_median)
```

Θα μετονομάσουμε τις στήλες του data frame κάνοντας χρήση του πακέτου dplyr ενώ με το πακέτο tidyr (Wickham) θα αναδιαμορφώσουμε τα δεδομένα κάνοντας χρήση της συνάρτησης

gather() και τέλος θα κάνουμε χρήση του πακέτου οπτικοποίησης δεδομένων ggplot2 για την αναπαράσταση με θηκογράμματα (Wilkinson).

```
>library(dplyr)
>colnames(RLE)
>RLE<-RLE%>%
  rename("UnirAtm_2_D2"="GSM1506057_0706A10_L2.CEL","Unir
wild_2_D2"="GSM1506053_0706A10_K2.CEL","Unirwild_1_D2"
="GSM1506052_0706A10_K1.CEL","HZE_At6_2"="GSM1506045_
1006A-03_HA6-20_2.CEL","HZE_At6_1"="GSM1506044_1006A-
03_HA6-19_2.CEL", "GamAt6_2"="GSM1506037_1006A03_GA6-
8_2.CEL",...,"HZE_wild24_2"="GSM1506033_0806A10_I2.CEL",
"HZE_wild3_1"="GSM1506026_1006A03_HW13.CEL","Gamwild12
_2"="GSM1506021_1006A03_GW129.CEL","Gamwild3_2"="GSM15
06017_1006A03_GW32_2.CEL","Gamwild1.5_1"="GSM1506014_0
806A-10_C1.CEL").
>library(tidyr)
>RLE_gather<tidyr::gather(RLE,arabidopsis_samples,log2_expr
ession_deviation)
>library(ggplot2)
>ggplot2::ggplot(RLE_gather,aes(arabidopsis_samples,log2_ex
pression_deviation))+geom_boxplot(outlier.shape = NA, color
= "darkgreen", notch = TRUE) + ylim(c(-2,
2))+theme(axis.text.x = element_text(colour = "purple",
angle = 60, size = 6.5,hjust = 1 , face = "bold"))
```

Μετά την εκτέλεση του κώδικα, στον άξονα των x παρουσιάζεται ο συνολικός αριθμός των δειγμάτων ενώ στον άξονα των y παρατίθενται οι αποκλίσεις των τιμών έκφρασης κάθε μεταγράφου όλων των δειγμάτων από τη διάμεσο (Εικόνα 7). Τα boxplots που εκτείνονται αρκετά είναι και εκείνα που υποδεικνύουν τη μεγαλύτερη διακύμανση των μεταγράφων από τη διάμεσο τιμή. Αυτό μπορεί να σχετίζεται με την ποιότητα των δεδομένων λόγω κάποιου τεχνικού παράγοντα (batch effect). Αν το σχήμα και η διάμεσος κάποιου θηκογράμματος διαφέρει σημαντικά από τα υπόλοιπα, θα πρέπει να ελεγχθεί και ενδεχομένως να απομακρυνθεί. Στην δική μας περίπτωση δε θα αφαιρεθεί κάποιο δείγμα για την μετέπειτα ανάλυση.



Εικόνα 7. Έλεγχος ποιότητας των δεδομένων μέσω θηκογραμμάτων με χρήση του γραφήματος οπτικοποίησης RLE εκτελώντας τον αλγόριθμο RMA χωρίς όμως να προηγηθεί κανονικοποίηση.

2.3.4 Κανονικοποίηση

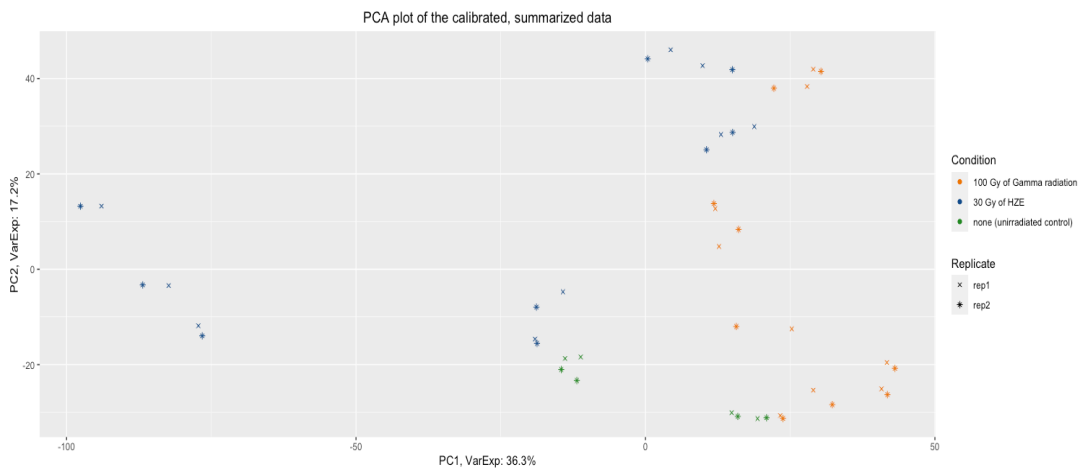
Ακολουθεί η εφαρμογή του αλγορίθμου RMA (Robust Multi-array Average) στα δεδομένα μας με σκοπό να πραγματοποιηθούν οι διαδικασίες της διόρθωσης θορύβου υποβάθρου, κανονικοποίησης και σύνοψης των probes σε probesets.

```
> eset_norm <- oligo::rma(raw_data)
> exp_eSet <- Biobase::exprs(ioan.eSet)
```

Για την αξιολόγηση της ποιότητας των κανονικοποιημένων πλέον δεδομένων μας διενεργούμε ανάλυση κυρίων συνιστωσών PCA (Principal Component Analysis) (Εικόνα 8, (Εικόνα 9) (Abdi and Williams). Πρόκειται για μια τεχνική που αποσκοπεί στο να μειώσει την πολυπλοκότητα στο χώρο των διαστάσεων, σε μια προσπάθεια να διακρίνουμε ομάδες που συμπεριφέρονται με παρόμοιο τρόπο. Πιο συγκεκριμένα, για να απεικονίσουμε μια PCA χρησιμοποιούμε το

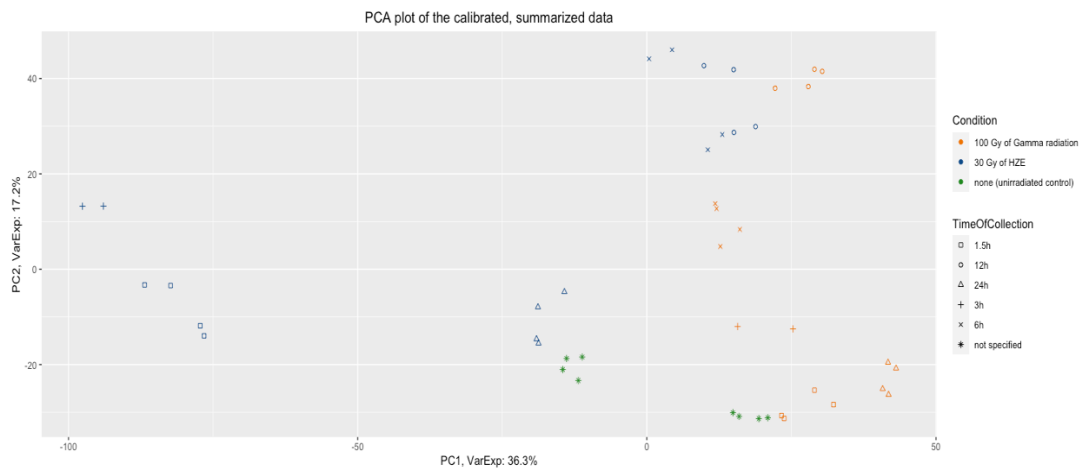
πακέτο βασικών γραφημάτων ggplot2 και τη συνάρτηση prcomp(). Εξ'ορισμού, η συνάρτηση prcomp() αναμένει να δει τα δείγματα ως γραμμές και τα γονίδια ως στήλες οπότε αντιμετωπίζουμε τον πίνακα χρησιμοποιώντας τη συνάρτηση transpose - αντιμετάθεση t(). Εάν δεν αντιμετωπίσουμε τον πίνακα θα καταλήξουμε με ένα γράφημα που δείχνει ότι τα γονίδια σχετίζονται μεταξύ τους. Αυτό που επιστρέφει η συνάρτηση prcomp() είναι τρία πράγματα, το x, sdev και το rotation. Το x περιέχει τις κύριες συνιστώσες PCs για το σχεδιασμό του γραφήματος. Χρησιμοποιούμε τις πρώτες δύο στήλες του x για να σχεδιάσουμε ένα δισδιάστατο γράφημα που χρησιμοποιεί τις δύο πρώτες PCs. Η πρώτη κύρια συνιστώσα είναι και εκείνη με τη μεγαλύτερη διακύμανση των αρχικών δεδομένων μας. Για να απεικονίσουμε ένα δισδιάστατο γράφημα, συνήθως χρησιμοποιούμε τις πρώτες δύο. Η PC1 είναι στον άξονα των x και η PC2 είναι στον άξονα των y. Η PC1 είναι η πρώτη στήλη στο x και η PC2 είναι η δεύτερη. Καθώς το ποσοστό της μεταβλητότητας για κάθε κύρια συνιστώσα είναι πιο ενδιαφέρον από ότι η πραγματική τιμή, υπολογίζουμε τα ποσοστά.

```
>PCA <- prcomp(t(exp_eSet), scale = FALSE)
>percentVar <- round(100*PCA$sdev^2/sum(PCA$sdev^2),1)
>sd_ratio <- sqrt(percentVar[2] / percentVar[1])
>data1 <- data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],
  replicate=pData(ioan.eSet)$FactorValue..REPLICAT
  E.,Condition=pData(ioan.eSet)$FactorValue..TREAT
  ED.WITH.)
>ggplot(data1, aes(PC1, PC2)) +
  geom_point(aes(shape = replicate, colour = condition))
+ ggtitle("PCA plot of the calibrated, summarized 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(4,8)) +
  scale_color_manual(values = c("darkorange2",
  "dodgerblue4","forestgreen"))
```



Εικόνα 8. Απεικόνιση PCA με βάση την ακτινοβολία και το πειραματικό αντίγραφο. Παρατηρούμε ότι το είδος της ακτινοβολήσης είναι υπεύθυνο για τη διασπορά των δειγμάτων σε μεγαλύτερο ποσοστό από τις πειραματικές επαναλήψεις, το οποίο είναι βιολογικώς αναμενόμενο. Επιπλέον, παρατηρούμε ότι στα cluster των δειγμάτων εμφανίζονται δείγματα και από τις δύο πειραματικές επαναλήψεις, γεγονός που υποδηλώνει ότι δεν υπάρχει batch effect.

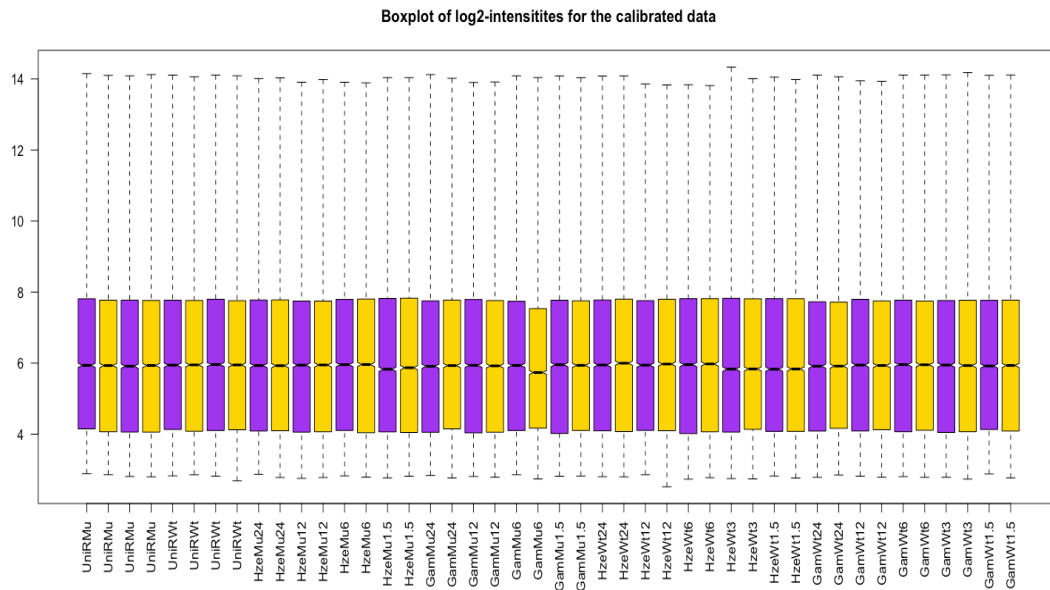
```
>data2 <- data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2],
  TimeOfCollection=pData(ioan.eSet)$FactorValue..T
  IME.AFTER.TREATMENT.,Condition=pData(ioan.eSet)$
  FactorValue..TREATED.WITH.)
>ggplot(data2, aes(PC1, PC2)) +
  geom_point(aes(shape = TimeOfCollection, colour =
  Condition)) + ggtitle("PCA plot of the calibrated,
  summarized 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(4,8)) +
  scale_color_manual(values = c("darkorange2",
  "dodgerblue4","forestgreen"))
```



(Εικόνα 9. Απεικόνιση PCA με βάση την ακτινοβολία και το χρόνο συλλογής του δείγματος. Παρατηρούμε ότι το είδος της ακτινοβολίας είναι υπεύθυνο για τη διασπορά των δειγμάτων σε μεγαλύτερο ποσοστό από το χρόνο συλλογής. Επιπλέον, παρατηρούμε το σχηματισμό cluster ανά ώρα συλλογής, το οποίο είναι βιολογικώς αναμενόμενο.

```
> oligo::boxplot(ioan.eSet, main = "Boxplot of log2-intensities for the calibrated data", names = c("UnirMu", "UnirWt", "HzeMu24", "HzeMu1.5", "GamMu24", "GamMu6", "Hzewt24", "Hzewt24", ... , "Hzewt12", "Hzewt3", "Hzewt1.5", "GamWt6", "GamWt1.5"), las = 2, col = c("purple", "gold"), notch = TRUE)
```

Με τη χρήση της συνάρτησης `boxplot()` του πακέτου της `oligo` θα απεικονίσουμε πλέον μέσω θηκογραμμάτων τα κανονικοποιημένα δεδομένα μας (Εικόνα 10). Κάθε θηκόγραμμα αναπαριστά τις τιμές εντάσεων φθορισμού ανά μικροσυστοιχία. Με βάση τον παραπάνω κώδικα που παρατέθηκε, η γραφική απεικόνιση των θηκογραμμάτων ανά μικροσυστοιχία ακολουθεί πλέον μια προσεγγιστικά κανονική κατανομή.



Εικόνα 10. Απεικόνιση θηκογραμμάτων των κανονικοποιημένων δεδομένων μας μέσω της συνάρτησης `boxplot()`, έπειτα από την εφαρμογή του αλγορίθμου RMA. Τα δεδομένα μας φαίνεται να ακολουθούν μια προσεγγιστικά κανονική κατανομή.

2.3.5 Φιλτράρισμα βασισμένο στην ένταση

Σε αυτό το στάδιο της ανάλυσης θα φιλτράρουμε όλα τα γονίδια των οποίων οι τιμές έκφρασης είναι εξαιρετικά χαμηλές. Οι ανιχνευτές (probes) από δεδομένα μικροσυστοιχιών δείχνουν συχνά ένα μεγάλο εύρος έντασης υποβάθρου. Αυτός ο αριθμός ανιχνευτών, συνήθως είναι παρόμοιος σε κάθε διάταξη μικροσυστοιχιών συνδυάζοντας μικρή απόκλιση με χαμηλή ένταση. Για τους παραπάνω λόγους καταλήγουμε να τα διαχειριζόμαστε σαν διαφορετικώς εκφραζόμενα αν και είναι λίγο πιο πάνω από το όριο ανίχνευσης και δεν προσφέρουν καμία ουσιαστική πληροφορία. Έτσι λοιπόν εφαρμόζουμε ένα φιλτράρισμα των εντάσεων υπολογίζοντας την τιμή της διαμέσου κάθε μεταγράφου από τα συνολικά 44 δείγματα που διαθέτουμε με τη βοήθεια της συνάρτησης `rowMedians()` και εξάγουμε τα αποτελέσματα στο αντικείμενο `mediansAssayData`.

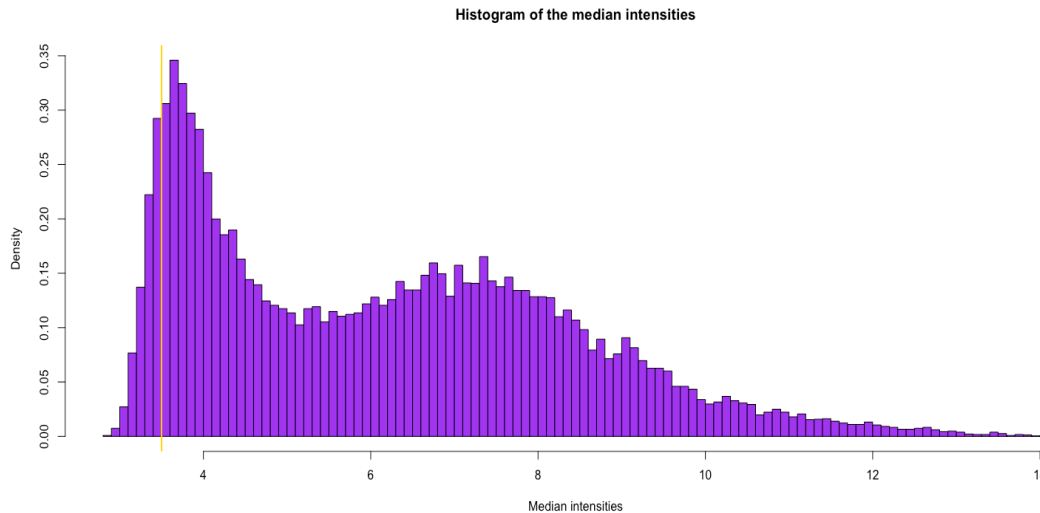
```
>mediansAssayData <-rowMedians(Biobase::exprs(ioan.eSet))
```

Με δεδομένες τις τιμές της διαμέσου κάθε μετάγραφου προχωράμε στη δημιουργία ιστογραμμάτων με χρήση της συνάρτησης `hist()`. Αυτό που παρατηρούμε είναι μία υψηλή συγκέντρωση γονιδίων με χαμηλές διαμέσους στα αριστερά του γραφήματος. Τα γονίδια αυτά θα φιλτραριστούν με ένα ενδεδειγμένο κατώφλι. Εξετάζουμε πάλι το ιστόγραμμα θέτοντας τη γραμμή κατωφλίου στο 3.5 αφαιρώντας όλα τα γονίδια που βρίσκονται πριν από αυτήν. Για να εισάγουμε τη γραμμή χρησιμοποιούμε τη συνάρτηση `abline()`. Η γραφική απεικόνιση του οποίου φαίνεται παρακάτω (Εικόνα 11).

```
>hist_medians <- hist(mediansAssayData, 100, col = "purple", freq = FALSE, main = "Histogram of the median intensities", border = "black", xlab = "Median intensities")
>threshold <- 3.500000
>hist_medians <- hist(mediansAssayData, 100, col = "purple", freq = FALSE, main = "Histogram of the median intensities", border = "black", xlab = "Median intensities")
> abline(v = threshold, col = "gold", lwd = 2)
```

Μετάγραφα που δεν έχουν εντάσεις φθορισμού μεγαλύτερες από το κατώφλι που θέσαμε σε τόσα δείγματα όσο η μικρότερη πειραματική ομάδα αφαιρούνται. Για να συμβεί αυτό ανακτούμε τη λίστα με τον αριθμό των δειγμάτων ανά πειραματική ομάδα `no_of_samples`, το αποτέλεσμα του οποίου είναι 2 το οποίο εκχωρούμε στο αντικείμενο `samples_threshold`.

```
>no_of_samples<-table(paste(pData(ioan.eSet)$FactorValue..TREATED.WITH., "_", pData(ioan.eSet)$FactorValue..GENOTYPE.,"_", pData(ioan.eSet)$FactorValue..TIME.AFTER.TREATMENT.))
>samples_threshold <- min(no_of_samples)
```



Εικόνα 11. Χρήση της συνάρτησης `hist()` για τη δημιουργία ιστογράμματος που απεικονίζει τις διαμέσους των εντάσεων ανά μετάγραφο. Από το γράφημα βλέπουμε μια υψηλή συγκέντρωση γονιδίων με χαμηλές διαμέσους και ενδεικτική τοποθέτηση γραμμής κατωφλίου για την αφαίρεση αυτών που αποτελούν πιθανά θόρυβο.

Κάνοντας χρήση της συνάρτησης `apply()` εφαρμόζεται σε κάθε μετάγραφο κατά μήκος όλων των δειγμάτων η συνάρτηση που θα αξιολογήσει αν ο συνολικός αριθμός των δειγμάτων όπου η διάμεσος ξεπερνά το κατώφλι είναι μεγαλύτερη από τη μικρότερη πειραματική ομάδα και επιστρέφει `TRUE` ή `FALSE` για κάθε μετάγραφο. Με τη συνάρτηση `subset()` θα αφαιρεθούν όλες οι `FALSE` τιμές από το αντικείμενο `idx_threshold` που αποτελεί υποσύνολο του `expressionSet ioan.eSet` και θα εκχωρηθούν στο αντικείμενο `filteredTranscripts`.

```
> idx_threshod <- apply(Biobase::exprs(ioan.eSet), 1, function(x){
  sum(x > threshold) >= samples_threshold})
> table(idx_threshod)
> filteredTranscripts <- subset(ioan.eSet, idx_threshod)
> nrow(filteredTranscripts)
```

2.3.5 Σχολιασμός των γονιδίων (transcript clusters/probesets)

Σε αυτό το σημείο θα προσθέσουμε πληροφορίες σχολιασμού για τα αναγνωριστικά των μεταγράφων που ομαδοποιούνται σχηματίζοντας clusters. Η συγχώνευση αυτή των μεταγράφων στο αντίστοιχο γονίδιο αποθηκεύεται εν συνεχεία στο data frame featureData που αποτελεί μέρος της δομής δεδομένων ExpressionSet. Χρησιμοποιούμε τη μέθοδο select() από το πακέτο σχολιασμού AnnotationDbi για να αιτηθούμε γονιδιακών συμβόλων (SYMBOL), μιας σύντομης περιγραφής των γονιδίων (GENENAME) καθώς και των αναγνωριστικών όπως έχουν εκχωρηθεί στη βάση TAIR για το φυτό *Arabidopsis Thaliana*. Αμέσως μετά αφαιρούμε όσα probes δεν αντιστοιχούν σε κάποιο σύμβολο γονιδίου. Με την συνάρτηση subset() θέλουμε να ικανοποιηθεί η συνθήκη !is.na(TAIR). Αυτό που επιστρέφεται είναι ένα data frame που θα πληρεί τη συνθήκη που θέσαμε. Η συνάρτηση is.na() δημιουργεί έναν πίνακα από λογικές τιμές, μέσω του οποίου TRUE υποδεικνύει ελλιπή τιμή NA. Με τον τελεστή ! μπροστά από τη συνάρτηση αντιστρέφεται η σημασία των λογικών τιμών. Σε αυτή την περίπτωση το TRUE υποδηλώνει ότι η τιμή δεν είναι ελλιπής επιστρέφοντας έτσι ένα data frame με αναγνωριστικά γονιδίων της *Arabidopsis Thaliana*.

```
>annot.feataData<-AnnotationDbi::select(ath1121501.db, keys =
(featureNames(filteredTranscripts)), columns = c("SYMBOL",
"GENENAME", "ENTREZID", "TAIR"), keytype = "PROBEID")
> annot.feataData <- subset(annot.feataData, !(is.na(TAIR)))
```

Με τη βοήθεια του πακέτου dplyr θα χρησιμοποιήσουμε τη συνάρτηση group_by() που θα δημιουργήσει ένα ομαδοποιημένο αντίγραφο του αρχικού μας data frame – annot.feataData με βάση τη στήλη PROBEID. Η συνάρτηση n_distinct() μετράει αποτελεσματικά τον αριθμό των μοναδικών τιμών στο σύνολο ενός διανύσματος και στη συγκεκριμένη περίπτωση το διάνυσμα μας είναι η στήλη TAIR. Η συνάρτηση

`summarise()` χρησιμοποιείται τα δεδομένα που δημιουργήθηκαν από την `group_by()` και το αποτέλεσμα της εντολής επιστρέφει μία γραμμή για κάθε ομάδα. Με τη συνάρτηση `filter()` αποκτούμε όλα εκείνα τα probes που υπάρχουν περισσότερες από μία φορές στη στήλη TAIR. Ο τελεστής `%in%` ελέγχει εάν ένα στοιχείο, στην περίπτωσή μας εάν το όνομα του γονιδίου μέσω της συνάρτησης `featureNames(filteredTranscripts)` εμπεριέχεται, στα probes εκείνα που εμφανίζονται πάνω από μια φορά. Το αποτέλεσμα αποθηκεύεται στο αντικείμενο `excludeIDs` και μέσω της εντολής `table(excludeIDs)` εμφανίζεται ένας πίνακας συχνοτήτων λογικών τιμών, όπου FALSE οι τιμές που δεν ταιριάζουν. Με την εντολή `subset()` αφαιρούμε εκείνες που είναι ταυτόσημες από το αντικείμενο `filteredTranscripts` και το αποτέλεσμα εκχωρείται στη δομή δεδομένων `ioan.Eset`. Η συνάρτηση `fData()` μας δίνει τη δυνατότητα να έχουμε πρόσβαση στα feature data του expressionSet `ioan.Eset`. Μέχρι τώρα πληροφορίες για τα γονίδια δεν έχουν αποθηκευτεί στο `fData(ioan.Eset)`. Μόνο τα ονόματα PROBEIDs των μεταγράφων/γονιδίων έχουν εκχωρηθεί στις γραμμές του. Δημιουργούμε μια στήλη PROBEID στα `fData(ioan.Eset)` και εκχωρούμε τις γραμμές με τα ονόματα σε αυτήν.

```
>groupByID<-group_by(annot.feaData, PROBEID)
>nrow(groupByID)
>summariseByID<-dplyr::summarize(groupByID, no_of_matches =
n_distinct(TAIR))
>nrow(summariseByID)
>filterByMatch <- filter(summariseByID, no_of_matches > 1)
>nrow(filterByMatch)
>excludeIDs<-(featureNames(filteredTranscripts)%in%
filterByMatch$PROBEID)
>table(excludeIDs)
>ioan.Eset<-subset(filteredTranscripts, !excludeIDs)
> validObject(ioan.Eset)
>fData(ioan.Eset)$PROBEID<-rownames(fData(ioan.Eset))
```


Για να δούμε τι είδους δεδομένα μπορούμε να ανακτήσουμε από το πακέτο σχολιασμού `ath1121501.db` χρησιμοποιούμε τη μέθοδο `columns`. Έπειτα εκχωρούμε τις γραμμές με τα ονόματα των PROBEIDs του `fData(ioan.Eset)` στο αντικείμενο `probes`. Μέχρι τώρα έχουμε αναφέρει μεθόδους που μπορούν να εξάγουν πληροφορίες από αρκετές στήλες, πρόσφατα προστέθηκε η μέθοδος `mapIds` που επιτρέπει την εξαγωγή μίας στήλης τη φορά. Για να μην επιστρέφονται πολλαπλές τιμές του ίδιου αναγνωριστικού όταν εξάγουμε πληροφορίες για μια συγκεκριμένη στήλη χρησιμοποιούμε τη μέθοδο `multiVals` και ορίζουμε να παίρνει την τιμή "first".

```
>columns(ath1121501.db)
>probes<-rownames(fData(ioan.Eset))
>fData(ioan.Eset)$ENTREZ <- mapIds(ath1121501.db,
                                   keys=probes,
                                   column="ENTREZID",
                                   keytype="PROBEID",
                                   multiVals="first")

>fData(ioan.Eset)$SYMBOL <- mapIds(ath1121501.db,
                                   keys=probes,
                                   column="SYMBOL",
                                   keytype="PROBEID",
                                   multiVals="first")

>fData(ioan.Eset)$GENENAME <- mapIds(ath1121501.db,
                                       keys=probes,
                                       column="GENENAME",
                                       keytype="PROBEID",
                                       multiVals="first")

>fData(ioan.Eset)$TAIR <- mapIds(ath1121501.db,
                                  keys=probes,
                                  column="TAIR",
                                  keytype="PROBEID",
                                  multiVals="first")
```

Εξορισμού γνωρίζουμε ότι η κλάση ExpressionSet συγχρονίζει έτσι τα δεδομένα της ώστε οι γραμμές του assayData να αντιστοιχούν στις γραμμές του featureData.

```
> all(rownames(exprs(ioan.ESet)) == rownames(fData(ioan.ESet)))  
> validObject(ioan.ESet)
```

2.3.6 Στατιστική Ανάλυση Διαφορικών Εκφρασμένων Γονιδίων - Γραμμικά Μοντέλα

Για την ανάλυση της γονιδιακής έκφρασης θα ανατρέξουμε στο πακέτο της limma (Ritchie et al.) που χρησιμοποιεί γραμμικά μοντέλα για την ανάλυση πειραμάτων μικροσυστοιχιών. Αρχικά θα δημιουργήσουμε τον πίνακα σχεδιασμού (design matrix), ο οποίος υποδεικνύει ποιο εκχύλισμα RNA έχει εφαρμοσθεί σε κάθε μικροσυστοιχία. Η συνάρτηση paste() θα μας βοηθήσει να συνδέσουμε τους τρεις πειραματικούς μας παράγοντες (factors) που είναι ο γονότυπος, η ώρα συλλογής του δείγματος RNA μετά την ακτινοβολία και το είδος της ακτινοβολίας. Το αντικείμενο *GTT* που δημιουργείται αποτελεί μεταβλητή χαρακτήρα και τη μετατρέπουμε σε παράγοντα με τη συνάρτηση factor(). Με αυτό τον τρόπο παρατηρούνται όλοι οι δυνατοί συνδυασμοί αποκαλώντας τον ως παραγοντικό σχεδιασμό (factorial design). Ο πίνακας σχεδιασμού δημιουργήθηκε μέσω της συνάρτησης model.matrix() εφαρμόζοντας ένα μοντέλο με ένα συντελεστή για κάθε ένα από τους συνδυασμούς παραγόντων.

```
>GTT<-paste(ioan.ESet$FactorValue..GENOTYPE.,ioan.ESet$Factor  
Value..TIME.AFTER.TREATMENT.,ioan.ESet$FactorValue..TREATED.WITH.  
,sep = ".")  
>GTT <- factor(GTT, levels = c("atm-1.not specified.none (unirradiated  
control)","wild type.not specified.none (unirradiated control)","atm-  
1.24h.30 Gy of HZE","atm-1.12h.30 Gy of HZE","atm-1.6h.30 Gy of
```

```
HZE","atm-1.1.5h.30 Gy of HZE","atm-1.24h.100 Gy of Gamma
radiation","atm-1.12h.100 Gy of Gamma radiation","atm-1.6h.100 Gy of
Gamma radiation","atm-1.1.5h.100 Gy of Gamma radiation","wild
type.24h.30 Gy of HZE","wild type.12h.30 Gy of HZE","wild type.6h.30
Gy of HZE","wild type.3h.30 Gy of HZE","wild type.1.5h.30 Gy of
HZE","wild type.24h.100 Gy of Gamma radiation","wild type.12h.100 Gy
of Gamma radiation","wild type.6h.100 Gy of Gamma radiation","wild
type.3h.100 Gy of Gamma radiation","wild type.1.5h.100 Gy of Gamma
radiation"))
```

```
>design <- model.matrix(~0+GTT)
>colnames(design)<c("mu.non.control","wt.non.control","mu.24hr.HZE",
"mu.12hr.HZE","mu.6hr.HZE","mu.1.5hr.HZE","mu.24hr.Gam","mu.12hr.
Gam","mu.6hr.Gam","mu.1.5hr.Gam","wt.24hr.HZE","wt.12hr.HZE","wt.6
hr.HZE","wt.3hr.HZE","wt.1.5hr.HZE","wt.24hr.Gam","wt.12hr.Gam","wt.
6hr.Gam","wt.3hr.Gam","wt.1.5hr.Gam")
```

Για να βρεθούν τα διαφορικώς εκφραζόμενα γονίδια χρησιμοποιούμε τη συνάρτηση `lmFit()` η οποία παράγει ένα μοντέλο γραμμικής προσαρμογής χρησιμοποιώντας το `design matrix` που σχεδιάσαμε παραπάνω. Καθώς πλέον το γραμμικό μοντέλο έχει σχηματισθεί ένας δεύτερος πίνακας (`contrast matrix`) απαιτείται κάνοντας χρήση της συνάρτησης `makeContrasts()` για τη δημιουργία συγκρίσεων ενδιαφέροντος μεταξύ των RNA δειγμάτων. Το αντικείμενο `fit` (του πόσο καλά εφαρμόζονται τα δεδομένα μας) και το `contrast_matrix` (ο πίνακας που περιέχει τις συγκρίσεις) χρησιμοποιούνται ως παράμετροι της συνάρτησης `contrasts.fit()` για να υπολογίσει τα `fold changes` και `t-statistics` ώστε να μπορέσουμε να διενεργήσουμε τις συγκρίσεις. Τα τυπικά σφάλματα περιορίζονται χρησιμοποιώντας το απλό `empirical Bayes` μοντέλο (`eBayes`).

```
>fit <-lmFit(ioan.ESet, design)
>contrast_matrix <- makeContrasts(wt.1.5hr.Gam-
wt.non.control, wt.3hr.Gam-
wt.non.control,wt.6hr.Gamwt.non.control,wt.12hr.Ga m-
wt.non.control,wt.24hr.Gam-wt.non.control,wt.1.5hr.HZE-
wt.no n.control,wt.3hr.HZE-wt.non.control,wt.6hr.HZE-
```

```

wt.non.control,wt        .12hr.HZE-wt.non.control,wt.24hr.HZE-
wt.non.control, mu.1.5hr.G
am-mu.non.control,mu.6hr.Gam-mu.non.control,mu.12hr.Gam-
mu        .non.control,mu.24hr.Gam-mu.non.control,mu.1.5hr.HZE-
mu.non.
control,mu.6hr.HZE-mu.non.control,mu.12hr.HZE-
mu.non.control,                mu.24hr.HZE-
mu.non.control,mu.non.controlwt.non.control,levels = design)
>fit <- contrasts.fit(fit, contrast_matrix)
>fit.eb <- eBayes(fit)

```

Η συνάρτηση topTable() εξάγει τα πολύ υψηλά στην κατάταξη διαφορικά εκφραζόμενα γονίδια μετά την προσαρμογή του γραμμικού μοντέλου. Τα κριτήρια που επιλέξαμε είναι $\log_2FC = 1.0$ και $p\text{-value} = 0.05$, ενώ η διόρθωση πολλαπλών ερωτημάτων έγινε με τη μέθοδο των Benjamini - Hochberg. Για να τυπώσουμε σε αρχείο τα αποτελέσματα που επιστρέφονται χρησιμοποιούμε τη συνάρτηση write.table(). Απομονώνουμε τη στήλη του ενδιαφέροντος μας που είναι τα αναγνωριστικά της TAIR και με τη χρήση της συνάρτησης na.omit() απομακρύνουμε όλες τις ελλιπείς τιμές του και ξανά τυπώνουμε πάλι με την write.table() τα αποτελέσματα σε αρχείο.

```

>DEG1.b <- topTable(fit.eb, coef = 1, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG1.b, file = "wt.1.5.Gamma.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_g1.b <- data.frame(DEG1.b[,5], row.names =
DEG1.b$PROBEID)
>de_g1.b <- na.omit(de_g1.b)
>write.table(de_g1.b, file="wt.1.5.GammaTAIR.txt",
row.names = FALSE, col.names = FALSE, sep = "\t", quote =
FALSE)

```

```

>DEG2.b <- topTable(fit.eb, coef = 2, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG2.b, file = "wt.3.Gamma.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_g2.b <- data.frame(DEG2.b[,5], row.names =
DEG2.b$PROBEID)
>de_g2.b <- na.omit(de_g2.b)

```

```

>write.table(de_G2.b, file="wt.3.GammaTAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG3.b <- topTable(fit.eb, coef = 3, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG3.b, file = "wt.6.Gamma.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G3.b <- data.frame(DEG3.b[,5], row.names =
DEG3.b$PROBEID)
>de_G3.b <- na.omit(de_G3.b)
>write.table(de_G3.b, file="wt.6.GammaTAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG4.b <- topTable(fit.eb, coef = 4, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG4.b, file = "wt.12.Gamma.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G4.b <- data.frame(DEG4.b[,5], row.names =
DEG4.b$PROBEID)
>de_G4.b <- na.omit(de_G4.b)
>write.table(de_G4.b, file="wt.12.GammaTAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG5.b <- topTable(fit.eb, coef = 5, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG5.b, file = "wt.24.Gamma.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G5.b <- data.frame(DEG5.b[,5], row.names =
DEG5.b$PROBEID)
>de_G5.b <- na.omit(de_G5.b)
>write.table(de_G5.b, file="wt.24.GammaTAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG6.b <- topTable(fit.eb, coef = 6, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG6.b, file = "wt.1.5.HZE.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G6.b <- data.frame(DEG6.b[,5], row.names =
DEG6.b$PROBEID)
>de_G6.b <- na.omit(de_G6.b)
>write.table(de_G6.b, file="wt.1.5.HZETAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG7.b <- topTable(fit.eb, coef = 7, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)

```

```

>write.table(DEG7.b, file = "wt.3.HZE.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G7.b <- data.frame(DEG7.b[,5], row.names =
DEG7.b$PROBEID)
>de_G7.b <- na.omit(de_G7.b)
>write.table(de_G7.b, file="wt.3.HZETAIR.txt", row.names =
FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG8.b <- topTable(fit.eb, coef = 8, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG8.b, file = "wt.6.HZE.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G8.b <- data.frame(DEG8.b[,5], row.names =
DEG8.b$PROBEID)
>de_G8.b <- na.omit(de_G8.b)
>write.table(de_G8.b, file="wt.6.HZETAIR.txt", row.names =
FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

>DEG9.b <- topTable(fit.eb, coef = 9, n = Inf, adjust.method
= "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG9.b, file = "wt.12.HZE.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G9.b <- data.frame(DEG9.b[,5], row.names =
DEG9.b$PROBEID)
>de_G9.b <- na.omit(de_G9.b)
>write.table(de_G9.b, file="wt.12.HZETAIR.txt", row.names =
FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

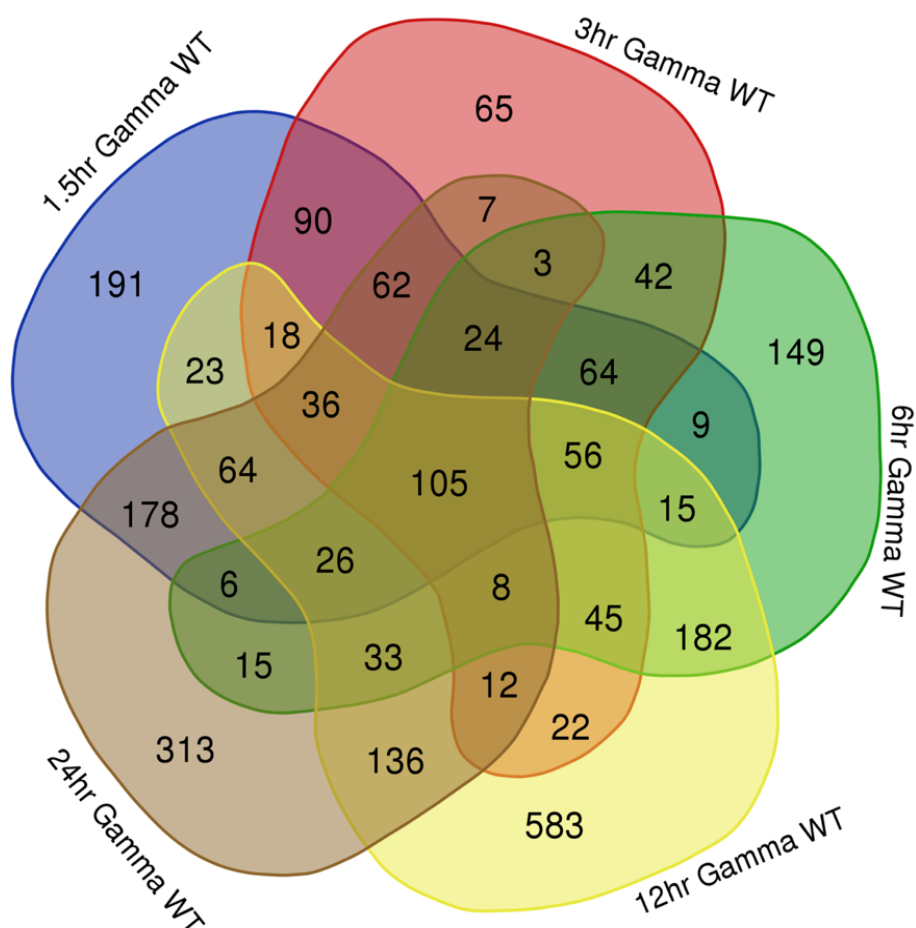
>DEG10.b <- topTable(fit.eb, coef = 10, n = Inf,
adjust.method = "BH", lfc = 1.0, p.value = 0.05)
>write.table(DEG10.b, file = "wt.24.HZE.txt", sep = "\t",
row.names = FALSE, quote = FALSE)
>de_G10.b <- data.frame(DEG10.b[,5], row.names =
DEG10.b$PROBEID)
>de_G10.b <- na.omit(de_G10.b)
>write.table(de_G10.b, file="wt.24.HZETAIR.txt", row.names
= FALSE, col.names = FALSE, sep = "\t", quote = FALSE)

```

2.4 Διαγράμματα VENN

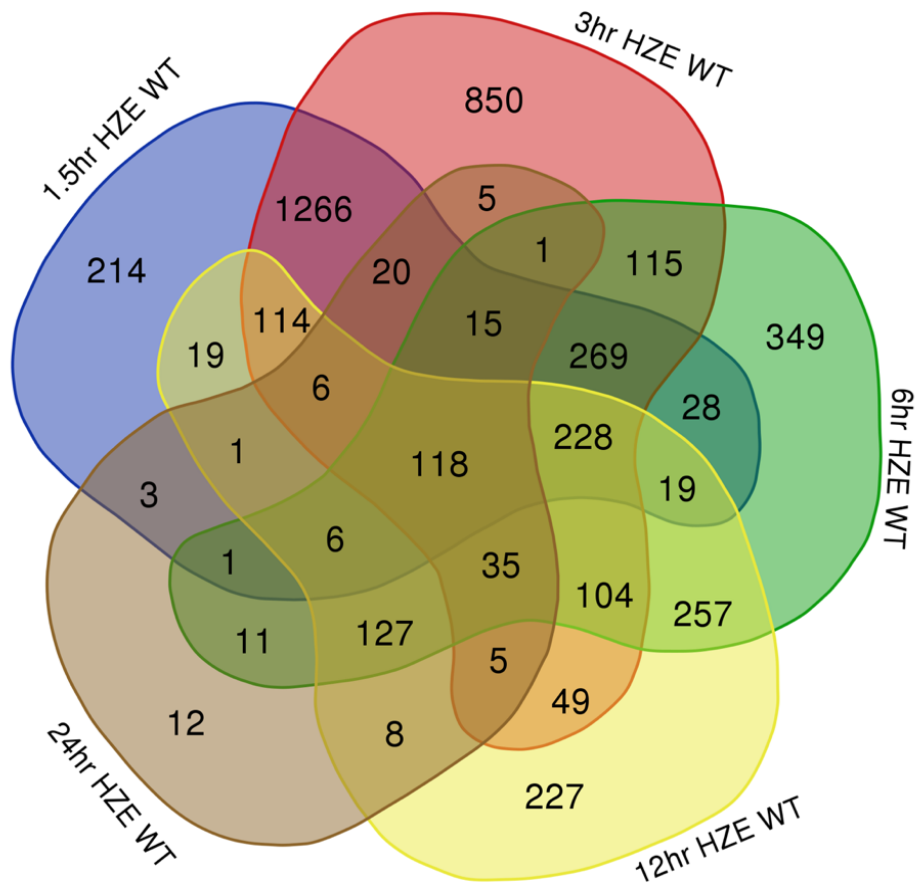
Σκοπός της χρήσης διαγραμμάτων Venn (Εικόνα 12, Εικόνα 13) είναι η σύγκριση των στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων (differentially expressed genes DEGs), με επιλεγμένα κριτήρια

p -value = 0.05 και log fold change = 1.0 για τον άγριο τύπο της *Arabidopsis Thaliana*. Επιδιώκουμε να αναδείξουμε κοινά DEGs μεταξύ δειγμάτων που έχουν υποστεί ακτινοβολήση- γ και ΗΖΕ ιόντων σιδήρου ξεχωριστά, λαμβάνοντας την ποσοτικοποίηση των mRNA επιπέδων σε διακριτές χρονικές περιόδους των 1.5, 3, 6, 12, 24 ωρών.¹



Εικόνα 12. Στο διάγραμμα Venn απεικονίζονται με διαφορετικό χρώμα τα διαφορετικώς εκφραζόμενα γονίδια στις υπό σύγκριση συνθήκες, δηλαδή στην ώρα συλλογής των δειγμάτων έπειτα από ακτινοβολήση με 100Gy γ -ακτινοβολίας στον άγριο τύπο της *Arabidopsis Thaliana*.

¹ <http://bioinformatics.psb.ugent.be/webtools/Venn/>



Εικόνα 13. Στο διάγραμμα Venn απεικονίζονται με διαφορετικό χρώμα τα διαφορετικώς εκφραζόμενα γονίδια στις υπό σύγκριση συνθήκες, δηλαδή στην ώρα συλλογής των δειγμάτων έπειτα από ακτινοβολήση με 30Gy HZE στον άγριο τύπο της *Arabidopsis Thaliana*.

2.5 Ανάλυση Λειτουργικού Εμπλουτισμού

Σε αυτό το στάδιο θα εστιάσουμε την προσοχή μας στη βιολογική σημασία των DEGs που προέκυψαν από την ανάλυση μας. Από τα διαγράμματα Venn παίρνουμε τα κοινά και τα γειτονικά γονίδια που βρέθηκαν να είναι διαφορετικώς εκφρασμένα σε όλα τα time points ανά ακτινοβολία στον άγριο τύπο της *Arabidopsis Thaliana*. Σκοπός μας είναι να βρούμε σε ποιες βασικές βιολογικές διεργασίες εμπλέκονται κάνοντας ανάλυση εμπλουτισμού μέσω της γονιδιακής οντολογίας GO εισάγοντας τα αναγνωριστικά των γονιδίων όπως καταχωρούνται στη

βάση TAIR. Τα αποτελέσματα που προέκυψαν αξιολογήθηκαν με βάση δύο κριτήρια: τον εμπλουτισμό (fold enrichment), που αξιολογεί το βαθμό εμπλουτισμού του συνόλου διαφορετικώς εκφραζόμενων γονιδίων μας σε συγκεκριμένες βιολογικές λειτουργίες και το Ποσοστό Εσφαλμένων Ανακαλύψεων (FDR) που προκύπτουν από τον έλεγχο πολλαπλών υποθέσεων. Από την ανάλυση FDR οροθετήθηκε ως κατώφλι, η τιμή $p\text{-value} < 0.05$, που αντιστοιχεί στο ποσοστό του συνόλου των ανακαλύψεων που αναμένετε να είναι εσφαλμένες.

Έπειτα μέσω της βιολογικής βάσης δεδομένων STRING v11 (Szklarczyk et al.) κατασκευάσαμε δίκτυα για την απεικόνιση των αλληλεπιδράσεων μεταξύ των διαφορετικώς εκφρασμένων γονιδίων. Ταυτοχρόνως αναδείχθηκαν τα hub genes των δικτύων βάση το αριθμό συνδεσιμότητας τους με άλλους κόμβους. Ως πηγές εξαγωγής των αλληλεπιδράσεων χρησιμοποιήθηκαν τα εργαστηριακά πειράματα και η αυτοματοποιημένη εξόρυξη κειμένου (text mining). Ο βαθμός εμπιστοσύνης τέθηκε στο 0.400 για να έχουμε όσο το δυνατόν περισσότερες αλληλεπιδράσεις μεταξύ των γονιδίων μετριάζοντας τη στατιστική σημαντικότητα. Τέλος, οι κόμβοι που δε συνδέονται στο δίκτυο αγνοούνται ενώ κόμβοι με περισσότερες από πέντε ακμές θεωρήθηκαν ως hub genes.

2.6 Πρόβλεψη Ορθολόγων μέσω DIOPT

Πολλά εργαλεία έχουν προκύψει στην ανάγκη ταυτοποίησης ορθολόγων. Ωστόσο η χαμηλή κάλυψη και η ανομοιογένεια τους αποτελεί εμπόδιο ώστε να διακρίνουμε υψηλής εμπιστοσύνης ορθόλογα ανάμεσα στα διάφορα είδη. Το εργαλείο DIOPT (Hu et al.) ενσωματώνει μια σειρά από προβλεπόμενα ορθόλογα γονίδια όπως αυτά προέκυψαν από 9 διαφορετικές βάσεις δεδομένων όπως η Ensembl Compara και η HomoloGene. Παρέχει μια βελτιστοποιημένη μέθοδο ενσωμάτωσης,

σύγκρισης και πρόσβασης σε ορθολογικές προβλέψεις που προέρχονται από αλγόριθμους βασισμένους στην ομολογία ακολουθιών, τα φυλογενετικά δέντρα και τη λειτουργική ομοιότητα. Υπολογίζει ένα απλό σκορ που υποδεικνύει τον αριθμό των εργαλείων που υποστηρίζει τη σχέση ενός ορθολογικού ζεύγους γονιδίων καθώς και ένα σταθμισμένο σκορ (weighted score) που βασίζεται στην ομοιότητα της λειτουργικής εκτίμησης ενός ζεύγους ορθόλογων γονιδίων που προβλέπεται από συσχέτιση ορών στη GO. Στην πρόσφατη έκδοση της DIOPT 5.2.1 κατατάσσει τα ζεύγη ορθόλογων γονιδίων σε high/moderate/low. Με high κατατάσσει τα ορθόλογα των οποίων το DIOPT score > 2, δηλαδή το σκορ του ζεύγους ορθόλογων γονιδίων είναι το υψηλότερο σε τουλάχιστον 2 εργαλεία και προς τις δύο κατευθύνσεις αναζήτησης. Με moderate κατατάσσει τα ορθόλογα των οποίων η μία εκ των δύο κατευθύνσεων δίνει το καλύτερο σκορ. Στην περίπτωση αυτή το DIOPT score θα πρέπει να είναι μεγαλύτερο του 2, δηλαδή να έχει βρεθεί σε τουλάχιστον δύο εργαλεία ως προς την αναζήτηση μιας κατεύθυνσης. Σε όλες τις υπόλοιπες περιπτώσεις οι σχέσεις των ορθόλογων κατατάσσεται ως low.

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Έχοντας στα χέρια μας ένα μεγάλο αριθμό διαφορεικά εκφραζόμενων γονιδίων όπως προέκυψαν από τις υπό σύγκριση συνθήκες μελέτης μέσω ανάλυσης στην R, θα περάσουμε στην αποτύπωση των στοιχείων αυτών σε πίνακες. Σε πίνακες, θα καταγράψουμε επίσης την πληροφορία για τα κοινά αλλά και τα μοναδικά DEGs όπως προέκυψαν από τα διαγράμματα Venn, που αντανακλούν την επίδραση που έχει η ακτινοβολία στο πρότυπο γονιδιακής έκφρασης του φυτού *A. thaliana* σε διακριτά χρονικά σημεία. Στη συνέχεια, θα απεικονίσουμε μέσω της STRING, δίκτυα πρωτεϊνικών αλληλεπιδράσεων για τις πιθανές βιολογικές διεργασίες στις οποίες εμπλέκονται τα προϊόντα των κοινών διαφορεικώς εκφραζόμενων γονιδίων ανά τύπο ακτινοβολίας. Έπειτα θα χρησιμοποιήσουμε την πλατφόρμα της Γονιδιακής Οντολογίας (Gene Ontology GO) για να κάνουμε ανάλυση εμπλουτισμού. Η GO συνδέεται πλέον με την ιστοσελίδα της PANTHER που οπτικοποιεί με οργανωμένο τρόπο τα αποτελέσματα της ανάλυσης εμπλουτισμού χρησιμοποιώντας τις σχέσεις της GO. Οι πίνακες με τα αποτελέσματα εμπλουτισμού έχουν ταξινομηθεί βάση του FDR (False Discovery Rate). Τέλος, μέσω του εργαλείου DIOPT εισάγουμε τα γονίδια που βρήκαμε να εμπλέκονται στις βασικότερες βιολογικές διεργασίες από τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων, για κάθε τύπο ακτινοβολίας ξεχωριστά και καταγράφουμε σε πίνακες την ομολογία των γονιδίων μεταξύ της *A.thaliana* και του ανθρώπου. Για την αξιολόγηση του βαθμού εμπιστοσύνης των ορθόλογων γονιδίων εννέα εργαλεία ενσωματώνονται στο DIOPT.

3.1 Διαφορικώς Εκφρασμένα Γονίδια

Πίνακας 1: Συνολικός αριθμός DEGs έπειτα από ακτινοβολήση-γ του άγριου τύπου της *A. thaliana* ανά time point (Εικόνα 12).

Συνθήκη	Αριθμός DEGs	Non-Redundant DEGs
1.5hr Gamma WT	972	967
3hr Gamma WT	662	659
6hr Gamma WT	782	782
12hr Gamma WT	1368	1364
24hr Gamma WT	1033	1028
Τελικός αριθμός DEGs		2582

Πίνακας 2: Συνολικός αριθμός DEGs έπειτα από HZE ακτινοβολήση του άγριου τύπου της *A. thaliana* ανά time point (Εικόνα 13).

Συνθήκη	Αριθμός DEGs	Non-Redundant DEGs
1.5hr HZE WT	2339	2327
3hr HZE WT	3214	3200
6hr HZE WT	1687	1683
12hr HZE WT	1326	1323
24hr HZE WT	374	374
Τελικός αριθμός DEGs		4482

Πίνακας 3: Μοναδικά διαφορικώς εκφραζόμενα γονίδια ανά τύπο ακτινοβολίας & ανά time point του άγριου τύπου της *A. thaliana*, που δε βρέθηκαν σε καμία άλλη ομάδα (Εικόνα 12, Εικόνα 13).

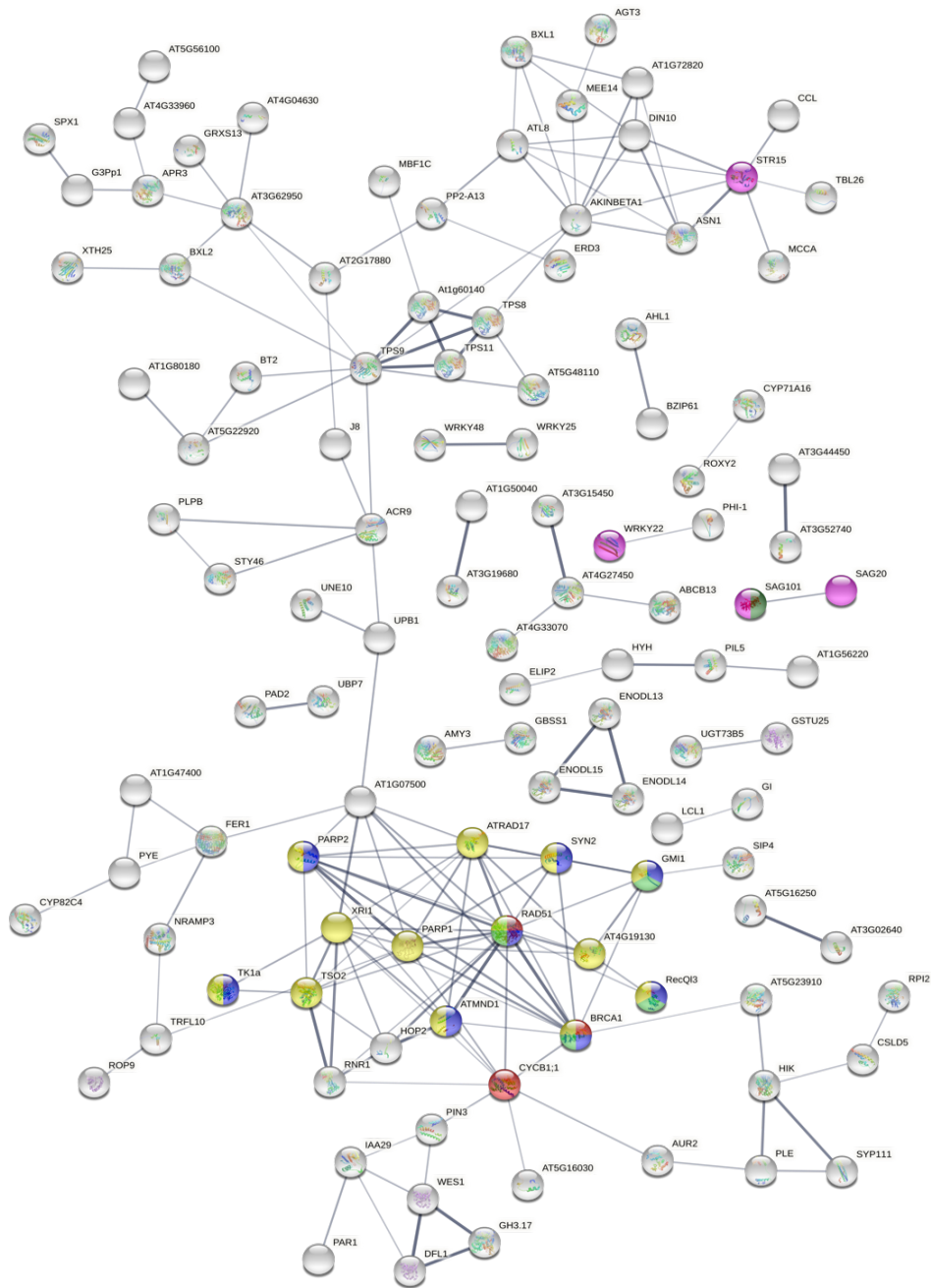
DEGs				
1.5hr γ	3hr γ	6hr γ	12hr γ	24hr γ
191	65	149	583	313
1.5hr HZE	3 HZE	6 HZE	12 HZE	24 HZE
214	850	349	227	12

Πίνακας 4: Κοινά διαφορετικώς εκφραζόμενα γονίδια ανά τύπο ακτινοβολίας & ανά time point του άγριου τύπου της *A. thaliana*, σε όλες τις ομάδες (Εικόνα 12, Εικόνα 13).

DEGs	
γ-rays 100Gy	HZE 30Gy
105	118

3.2 Βιολογικά Δίκτυα – HUB Genes

Για τη δημιουργία δικτύων πρωτεϊνικών αλληλεπιδράσεων χρησιμοποιούμε τη βάση δεδομένων STRING v11. Ως input εισάγουμε τη λίστα με τα αναγνωριστικά της TAIR που βρέθηκαν να είναι κοινά σε όλα τα time points και κοινά σε τέσσερα από τα πέντε time points, για τους δύο τύπους ακτινοβολίας ξεχωριστά (**Error! Reference source not found.**, Εικόνα 15). Η κατασκευή του δικτύου υλοποιήθηκε με βάση κάποιων επιλεγμένων κριτηρίων: το σκορ της ελάχιστης απαιτούμενης αλληλεπίδρασης και τις πηγές από όπου προέρχεται η πληροφορία αλληλεπίδρασης. Το σκορ τέθηκε στο 0.4, που είναι ένα μεσαίας εμπιστοσύνης κατώφλι. Η χρήση κατωφλίων με χαμηλότερα σκορ δίνουν περισσότερες αλληλεπιδράσεις αλλά και περισσότερα ψευδώς θετικά αποτελέσματα. Στην περίπτωση μας, δοκιμές με υψηλότερα σκορ συρρίκνωναν κατά πολύ την πληροφορία του δικτύου οπότε αναπόφευκτα οδηγηθήκαμε στη χρήση του medium confidence = 0.400. Οι δύο πηγές δεδομένων που επιλέχθηκαν για την αναπαράσταση των αλληλεπιδράσεων, ήταν τα εργαστηριακά πειράματα και η αυτοματοποιημένη εξόρυξη κειμένου (text mining). Τέλος, ως hub genes θεωρήσαμε όλα τα γονίδια που έχουν βαθμό συνδεσιμότητας 5 κόμβων και άνω.



Εικόνα 14. Αναπαράσταση δικτύου πρωτεϊνικών αλληλεπιδράσεων των DEGs έπειτα από έκθεση σε ακτινοβολία-γ του φυτού *A. thaliana*. Οι πρωτεΐνες αναπαρίστανται ως κόμβοι και οι μεταξύ τους σχέσεις με ακμές. Το πάχος της γραμμής υποδεικνύει την ισχύ των δεδομένων σύνδεσης. Τα διαφορετικά χρώματα μέσα στους κόμβους αντιστοιχούν στις διάφορες βιολογικές διεργασίες που αυτοί συμμετέχουν.

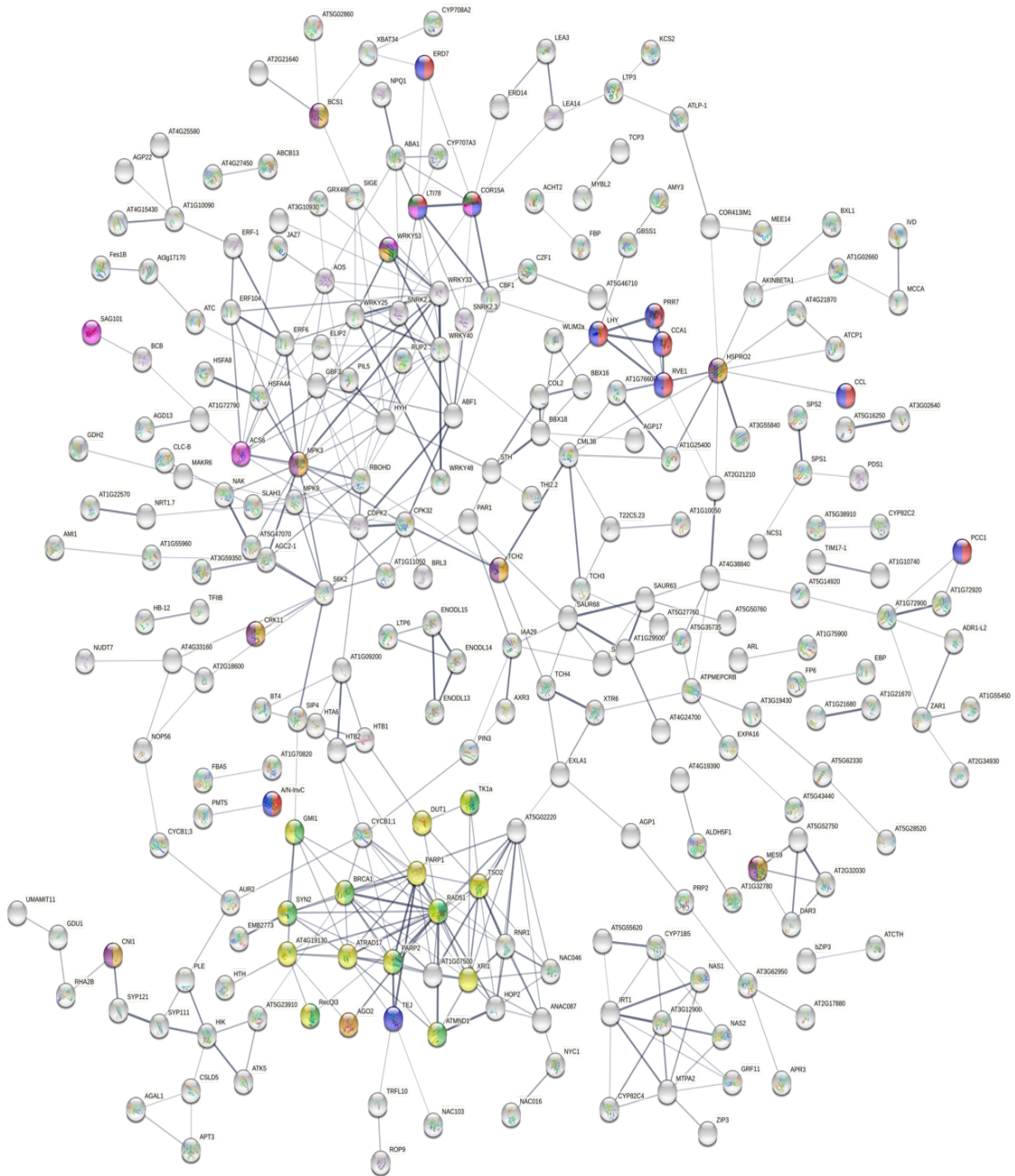
Πίνακας 5 : Genes & Hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση της *A. thaliana* σε ακτινοβολία-γ και οι βιολογικές διεργασίες στις οποίες συμμετέχουν που σχετίζονται άμεσα με το DDR (*Error! Reference source not found.*).

HUB GENES	GENES	DNA repair		Senescence		Gamma Radiation
		DSB repair	DSB via HR	Aging	Leaf Senescence	
AT4G19130	AT4G19130					
ATMND1	ATMND1	√				
ATRAD17	ATRAD17					
BRCA1	BRCA1	√	√			√
GMI1	GMI1	√	√			
PARP1	PARP1					
PARP2	PARP2	√				
RAD51	RAD51	√	√			√
	RecQI3	√	√			
SYN2	SYN2	√				
	TK1a	√				
TSO2	TSO2					
XRI1	XRI1					
	SAG101			√	√	
	SAG20			√		
STR15	STR15			√		
	WRKY22			√		
CYCB1;1	CYCB1;1					√

Πίνακας 6: Υπολειπόμενα hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση ακτινοβολίας-γ και οι βιολογικές διεργασίες (GO) που συμμετέχουν (*Error! Reference source not found.*).

GO Biological process		Number of genes	Hub Genes
GO:0007049	cell cycle	13	HOP2, AUR2, AT1G07500
GO:0042592	homeostatic process	11	AT3G62950, APR3, G3Pp1, PIN3, PYE
GO:0009266	response to temperature stimulus	10	MBF1C
GO:0009314	response to radiation	17	ASN1
GO:0010035	response to inorganic substance	12	MBF1C, RNR1
GO:0005975	carbohydrate metabolic process	12	AKINBETA1, DIN10, TPS9
GO:1901700	response to oxygen-containing compound	18	ASN1, MBF1C
GO:0005992	trehalose biosynthetic process	3	TPS9

GO:0044260	cellular macromolecule metabolic process	48	ATL8, MBF1C, HOP2, RNR1
------------	--	----	-------------------------



Εικόνα 15. Αναπαράσταση δικτύου πρωτεϊνικών αλληλεπιδράσεων των DEGs έπειτα από έκθεση σε ΗΖΕ ακτινοβολία του φυτού *A. thaliana*. Οι πρωτεΐνες αναπαρίστανται ως κόμβοι και οι μεταξύ τους σχέσεις με ακμές. Το πάχος της γραμμής υποδεικνύει την ισχύ των δεδομένων

σύνδεσης. Τα διαφορετικά χρώματα μέσα στους κόμβους αντιστοιχούν στις διάφορες βιολογικές διεργασίες που αυτοί συμμετέχουν.

Πίνακας 7: Genes & Hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση της *A. thaliana* σε ΗΖΕ ακτινοβολία και οι βιολογικές διεργασίες (GO) στις οποίες συμμετέχουν που σχετίζονται άμεσα με το DDR (Εικόνα 15).

HUB GENES	GENES	Circadian Rhythm	DNA repair		Senescence		Immune System
			DSB repair	DSB via HR	Aging	Leaf	
	A/N-InvC	✓					
	CCA1	✓					
	CCL	✓					
COR15A	COR15A	✓					
	ERD7	✓					
LHY	LHY	✓					
LT178	LT178	✓			✓	✓	
	PCC1	✓					
	PRR7	✓					
RVE1	RVE1	✓					
TEJ	TEJ	✓					
AT4G19130	AT4G19130						
ATMND1	ATMND1		✓				
ATRAD17	ATRAD17						
BRCA1	BRCA1		✓	✓			
DUT1	DUT1						
GMI1	GMI1		✓	✓			
PARP1	PARP1						
PARP2	PARP2		✓				
RAD51	RAD51		✓	✓			
	RecQ13		✓	✓			
SYN2	SYN2		✓				
	TK1a		✓				
TSO2	TSO2						
XRI1	XRI1						
ACS6	ACS6				✓		
COR15A	COR15A				✓	✓	
	SAG101				✓		
	AGO2						✓
	BCS1						✓
	CNI1						✓
	CRK11						✓
HSPRO2	HSPRO2						✓

	MES9						√
MPK3	MPK3						√
TCH2	TCH2						√
	WRKY53				√	√	√

Πίνακας 8: Υπολειπόμενα hub genes του δικτύου πρωτεϊνών έπειτα από έκθεση ΗΖΕ ακτινοβολίας και οι βιολογικές οι βιολογικές διεργασίες (GO) που συμμετέχουν (Εικόνα 15).

GO Biological process		Number of genes	Hub Genes
GO:0007049	cell cycle	18	AUR2, HOP2, CYCB1;1, AT1G07500, AT5G02220
GO:0042592	homeostatic process	16	MTPA2, GRX480, APR3, RBOHD, SLAH3
GO:0055114	oxidation-reduction process	26	ALDH5F1, ABA1, RBHOHD, APR3, RNR1, AOS, GRX480
GO:0050793	regulation of developmental process	21	NAC046, CDPK2
GO:0032870	cellular response to hormone stimulus	30	SAUR63, IAA29, GRX480, EBP, SNRK2.3, ABF1
GO:0051707	response to other organism	27	ATPMEPCRB, WRKY33, AOS, RBHOHD, EBP, LEA14
GO:0009620	response to fungus	12	ERF104, RBHOHD, AOS, WRKY33, WRKY40, LEA14
GO:0009266	response to temperature stimulus	27	ATPMEPCRB, ALDH5F1, RBHOHD, ABA1, CBF1, EBP, WRKY33
GO:0009408	response to heat	15	ABA1, ALDH5F1, EBP, RBHOHD, WRKY33
GO:0009409	response to cold	16	ATPMEPCRB, CBF1, WRKY33
GO:0009414	response to water deprivation	17	ABA1, WRKY33, LEA14, SLAH3
GO:0009611	response to wounding	11	CML38, LEA14, WRKY40, AOS, RBHOHD
GO:0009639	response to red or far red light	11	HYH, SIGE, IAA29, ABA1
GO:0010200	response to chitin	12	WRKY40, WRKY33
GO:0009733	response to auxin	10	IAA29, SAUR63
GO:0009737	response to abscisic acid	21	ABF1, GBF3, SNRK2.3
GO:0010212	response to ionizing radiation	5	CYCB1;1
GO:0009411	response to UV	5	HYH

3.3 Αποτελέσματα Ανάλυσης Εμπλουτισμού

Μέσω της πλατφόρμας Gene Ontology πραγματοποιήθηκε ανάλυση εμπλουτισμού στο σύνολο των διαφορικά εκφραζόμενων γονιδίων ως προς τις βιολογικές διεργασίες. Αρχικά, αυτό που υπολογίζεται είναι ο λόγος εμπλουτισμού ή εμπλουτισμός (fold enrichment) για κάθε λειτουργική κατηγορία. Για να αποφασιστεί αν για κάθε κατηγορία ο εμπλουτισμός είναι σημαντικός, εφαρμόζεται η υπεργεωμετρική κατανομή. Η τιμή που προκύπτει για κάθε διεργασία χρησιμοποιείται ως τιμή p-value που είναι απαραίτητη για τον έλεγχο υποθέσεων. Για να ξεπεραστεί το πρόβλημα πολλαπλών ερωτημάτων, που οδηγούν σε ποσοστό εσφαλμένων ανακαλύψεων, χρησιμοποιείται η μέθοδος των Benjamini και Hochberg, που είναι γνωστή και ως False Discovery Rate (FDR), βάση του οποίου επιλέχθηκαν τα αποτελέσματα του εμπλουτισμού. Με αυτή τη μέθοδο απορρίπτεται μέρος των αρχικά στατιστικά σημαντικών p-values περιορίζοντας κατά πολύ τα τελικά αποτελέσματα σε πειράματα έκφρασης.

Πίνακας 9: Τα βασικότερα αποτελέσματα που προέκυψαν από τη λειτουργική ανάλυση της GO, εισάγοντας τα DEGs που προέκυψαν από το νεαρό βλαστό της A.thaliana έπειτα από ακτινοβολήση-γ των 100Gy σε 5 time points.

GO Biological process	1.5 γ	3 γ	6 γ	12 γ	24 γ
	False Discovery Rate				
response to ionizing radiation (GO:0010212)	3.84E-02	1.40E-02	1.71E-02	2,49E-02	
response to oxidative stress (GO:0006979)	8.00E-06	3.53E-02		3.06E-05	1.88E-05
regulation of reactive oxygen species metabolic process (GO:2000377)				7.33E-03	
cell cycle (GO:0007049)			5.51E-05		
mitotic cell cycle process (GO:1903047)		5.96E-05	3.48E-04		
mitotic cell cycle (GO:0000278)		2.80E-04	5.41E-04		

regulation of cell cycle process (GO:0010564)		6.04E-04	2.01E-03		
cell division (GO:0051301)		3.96E-03	1.18E-02		
negative regulation of mitotic cell cycle (GO:0045930)		2.47E-02			
mitotic cell cycle phase transition (GO:0044772)			7.21E-03		
regulation of mitotic cell cycle (GO:0007346)		3.74E-02			
DNA repair (GO:0006281)	1.21E-04	8.97E-05			
double-strand break repair (GO:0006302)	1.92E-03	2.50E-04	2.36E-02		
homologous recombination (GO:0035825)		3.80E-02			
circadian rhythm (GO:0007623)	5.44E-03		2.02E-05	3.70E-16	1.90E-02
regulation of circadian rhythm (GO:0042752)				7.93E-04	
circadian regulation of gene expression (GO:0032922)				2.87E-02	
rhythmic process (GO:0048511)	2.96E-04		8.99E-06	5.66E-17	3.98E-03
aging (GO:0007568)			4.38E-02	1.67E-05	
systemic acquired resistance (GO:0009627)			8.73E-03	3.64E-02	
homeostatic process (GO:0042592)	1.02E-06	5.37E-04	8.36E-05	4.50E-04	1.29E-02
cellular homeostasis (GO:0019725)	8.80E-04	1.98E-02	3.55E-04	4.84E-03	
cell killing (GO:0001906)	9.76E-03		3.34E-02		
regulation of post-embryonic development (GO:0048580)				2.89E-06	6.63E-03
positive regulation of auxin mediated signaling pathway (GO:0010929)	4.91E-02	3.14E-02			

Πίνακας 10: Τα βασικότερα αποτελέσματα που προέκυψαν από τη λειτουργική ανάλυση της GO, εισάγοντας τα DEGs που προέκυψαν από το νεαρό βλαστό της *A.thaliana* έπειτα από HZE ακτινοβολήση των 30Gy σε 5 time points.

GO Biological process	1.5 HZE	3 HZE	6 HZE	12 HZE	24 HZE
	False Discovery Rate				
response to ionizing radiation (GO:0010212)			6.45E-03		3.07E-03
response to oxidative stress (GO:0006979)	8.42E-08	3.09E-10	4.83E-10	3.19E-08	6.77E-06
regulation of cell cycle process (GO:0010564)			2.86E-02		
DNA repair (GO:0006281)					9.24E-03
double-strand break repair (GO:0006302)					7.80E-04
circadian rhythm (GO:0007623)	9.16E-03	8.85E-05	1.28E-10	5.83E-12	

regulation of circadian rhythm (GO:0042752)			4.71E-03	2.22E-03	
rhythmic process (GO:0048511)	1.41E-02	1.50E-05	1.83E-12	2.82E-14	
aging (GO:0007568)	1.84E-02	7.91E-03	6.99E-03		1.66E-03
leaf senescence (GO:0010150)		4.69E-02			2.26E-02
regulation of leaf senescence (GO:1900055)				2.30E-02	
cell death (GO:0008219)		1.03E-03	1.72E-03	2.51E-02	1.55E-02
programmed cell death (GO:0012501)		1.28E-03	6.20E-04	3.46E-02	2.59E-02
immune response (GO:0006955)		2.33E-04	1.45E-08	6.63E-06	1.97E-07
regulation of immune response (GO:0050776)			4.61E-02		5.13E-04
immune system process (GO:0002376)		3.61E-04	5.16E-09	1.38E-06	6.37E-09
regulation of immune system process (GO:0002682)			5.61E-03		7.23E-07
innate immune response (GO:0045087)		2.80E-04	1.83E-08	1.98E-05	6.74E-07
regulation of innate immune response (GO:0045088)			3.65E-02		3.79E-04
immune effector process (GO:0002252)					1.18E-02
regulation of immune effector process (GO:0002697)					2.86E-03
systemic acquired resistance (GO:0009627)		1.69E-02	8.45E-03	1.35E-02	
autophagy (GO:0006914)		3.34E-02			
process utilizing autophagic mechanism (GO:0061919)		3.35E-02			
regulation of neurotransmitter levels (GO:0001505)	4.24E-02		1.28E-02		
positive regulation of auxin mediated signaling pathway (GO:0010929)					3.37E-03
homeostatic process (GO:0042592)	4.33E-02		4.75E-06	9.47E-03	

3.4 Ορθόλογα Γονίδια

Για την εύρεση ορθόλογων γονιδίων μεταξύ της *A. thaliana* και του ανθρώπου κάναμε χρήση του εργαλείου DIOPT. Τα αποτελέσματα που προέκυψαν καταγράφηκαν στους παρακάτω πίνακες. Η πρώτη και η τρίτη στήλη αναφέρονται στην αντιστοιχία των ορθόλογων γονιδίων μεταξύ των οργανισμών προς σύγκριση. Η τέταρτη στήλη υπολογίζει ένα σκορ που υποδεικνύει τον αριθμό των εργαλείων που ταυτοποιεί τη σχέση αυτή, με μέγιστο σκορ το εννέα όσα δηλαδή και τα ενσωματωμένα εργαλεία. Το DIOPT score αποτελεί έναν πρώτο εκτιμητή του βαθμού εμπιστοσύνης μεταξύ των ορθόλογων ζευγών. Στην πέμπτη στήλη υπολογίζεται ένα σταθμισμένο σκορ (weighted score) που βασίζεται στη λειτουργική αξιολόγηση. Ουσιαστικά το weighted score πρόκειται για ένα μέσο όρο που προκύπτει χρησιμοποιώντας τον υψηλής ποιότητας σχολιασμό σχετικά με μοριακές λειτουργίες της GO για όλα τα ζεύγη ορθόλογων που προβλέφθηκαν από κάθε εργαλείο. Για κάθε ζεύγος υπάρχουν βαθμίδες: με high ταξινομούνται τα ζεύγη που αμφίδρομα δίνουν το υψηλότερο σκορ (συνήθως δύο εργαλεία και άνω), με moderate τα ζεύγη εκείνα που δίνουν το υψηλότερο σκορ προς τη μία κατεύθυνση αναζήτησης και με low εκείνα τα ζεύγη που μόνο ένα εργαλείο δίνει πρόγνωση ομολογίας.

Πίνακας 11: Αποτελέσματα ορθολόγων γονιδίων μεταξύ της A.thaliana και του ανθρώπου. Από τα κοινά DEGs που προέρχονται από ακτινοβολήση-γ, επιλέχθηκαν ως input στο DIOPT εκείνα που εμπλέκονται στις βασικότερες βιολογικές διεργασίες έτσι όπως προέκυψαν από τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων της STRING.

Thale cress GeneID	TAIR ID	Human GeneID	DIOPT Score	Weighted Score	Rank	Best Score	Best Score Reverse
AT4G19130	AT4G19130	RPA1	8	8.07	high	Yes	Yes
ATMND1	AT4G29170	MND1	9	8.98	high	Yes	Yes
ATRAD17	AT5G66130	RAD17	9	8.98	high	Yes	Yes
BRCA1	AT4G21070	BARD1	4	4.06	moderate	Yes	No
BRCA1	AT4G21070	BRCA1	2	2.01	moderate	No	Yes
BRCA1	AT4G21070	ANKRD12	2	2.01	moderate	No	Yes

BRCA1	AT4G21070	ANKRD11	2	2.01	moderate	No	Yes
PARP1	AT2G31320	PARP1	8	7.98	high	Yes	Yes
PARP1	AT2G31320	PARP2	3	2.91	low	No	No
PARP2	AT4G02390	PARP2	9	8.98	high	Yes	Yes
PARP2	AT4G02390	PARP3	2	2.11	moderate	No	Yes
PARP2	AT4G02390	PARP1	2	1.91	low	No	No
RAD51	AT5G20850	RAD51	9	8.98	high	Yes	Yes
RAD51	AT5G20850	DMC1	2	1.91	low	No	No
RecQ13	AT4G35740	RECQL5	5	4.87	high	Yes	Yes
RecQ13	AT4G35740	RECQL	2	1.91	low	No	No
RecQ13	AT4G35740	WRN	2	1.91	moderate	No	Yes
RecQ13	AT4G35740	BLM	2	1.91	low	No	No
SYN2	AT5G40840	RAD21L1	2	2.01	moderate	Yes	No
SYN2	AT5G40840	RAD21	2	2.01	moderate	Yes	No
TK1a	AT3G07800	TK1	8	7.98	high	Yes	Yes
TSO2	AT3G27060	RRM2	9	8.98	high	Yes	Yes
TSO2	AT3G27060	RRM2B	8	7.98	moderate	No	No
CYCB1;1	AT4G37490	CCNB2	7	7.07	high	Yes	Yes
CYCB1;1	AT4G37490	CCNB1	5	5	moderate	No	Yes
CYCB1;1	AT4G37490	CCNE2	2	2.11	moderate	No	Yes
CYCB1;1	AT4G37490	CCNB3	2	2.11	low	No	No
CYCB1;1	AT4G37490	CCNE1	2	2.11	moderate	No	Yes
CYCB1;1	AT4G37490	CCNA1	2	1.91	low	No	No

Πίνακας 12: Αποτελέσματα ορθολόγων γονιδίων μεταξύ της A.thaliana και του ανθρώπου. Από τα κοινά DEGs που προέρχονται από HZE ακτινοβολία, επιλέχθηκαν ως input στο DIOPT εκείνα που εμπλέκονται στις βασικότερες βιολογικές λειτουργίες έτσι όπως προέκυψαν από τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων της STRING.

Thale cress GeneID	TAIR ID	Human Symbol	DIOPT Score	Weighted Score	Rank	Best Score	Best Score Reverse
CCA1	AT2G46830	MYSM1	2	2.1	high	Yes	Yes
ERD7	AT2G17840	SPART	2	2.15	moderate	Yes	No
LHY	AT1G01060	MYSM1	2	2.1	high	Yes	Yes
RVE1	AT5G17300	MYSM1	2	2.1	high	Yes	Yes
TEJ	AT2G31870	BPHL	4	3.92	high	Yes	Yes
TEJ	AT2G31870	PARG	2	2	moderate	No	Yes
AT4G19130	AT4G19130	RPA1	8	8.07	high	Yes	Yes

ATMND1	AT4G29170	MND1	9	8.98	high	Yes	Yes
ATRAD17	AT5G66130	RAD17	9	8.98	high	Yes	Yes
BRCA1	AT4G21070	BAR1	4	4.06	moderate	Yes	No
BRCA1	AT4G21070	BRCA1	2	2.01	moderate	No	Yes
BRCA1	AT4G21070	ANKRD12	2	2.01	moderate	No	Yes
BRCA1	AT4G21070	ANKRD11	2	2.01	moderate	No	Yes
DUT1	AT3G46940	DUT	4	3.91	high	Yes	Yes
PARP1	AT2G31320	PARP1	8	7.98	high	Yes	Yes
PARP1	AT2G31320	PARP2	3	2.91	low	No	No
PARP2	AT4G02390	PARP2	9	8.98	high	Yes	Yes
PARP2	AT4G02390	PARP3	2	2.11	moderate	No	Yes
PARP2	AT4G02390	PARP1	2	1.91	low	No	No
RAD51	AT5G20850	RAD51	9	8.98	high	Yes	Yes
RAD51	AT5G20850	DMC1	2	1.91	low	No	No
RecQ13	AT4G35740	RECQL5	5	4.87	high	Yes	Yes
RecQ13	AT4G35740	RECQL	2	1.91	low	No	No
RecQ13	AT4G35740	WRN	2	1.91	moderate	No	Yes
RecQ13	AT4G35740	BLM	2	1.91	low	No	No
SYN2	AT5G40840	RAD21L1	2	2.01	moderate	Yes	No
SYN2	AT5G40840	RAD21	2	2.01	moderate	Yes	No
TK1a	AT3G07800	TK1	8	7.98	high	Yes	Yes
TSO2	AT3G27060	RRM2	9	8.98	high	Yes	Yes
TSO2	AT3G27060	RRM2B	8	7.98	moderate	No	No
ACS6	AT4G11280	ACCSL	9	8.98	high	Yes	Yes
ACS6	AT4G11280	ACCS	8	7.98	moderate	No	No
AGO2	AT1G31280	PIWIL3	1	1.1	low	Yes	No
AGO2	AT1G31280	PIWIL4	1	1.1	low	Yes	No
AGO2	AT1G31280	PIWIL2	1	1.1	low	Yes	No
AGO2	AT1G31280	PIWIL1	1	1.1	low	Yes	No
AGO2	AT1G31280	AGO4	1	1.01	low	Yes	No
AGO2	AT1G31280	AGO1	1	1.01	low	Yes	No
AGO2	AT1G31280	AGO2	1	1.01	low	Yes	No
AGO2	AT1G31280	AGO3	1	1.01	low	Yes	No
BCS1	AT3G50930	BCS1L	6	6.07	high	Yes	Yes
CNI1	AT5G27420	RNF133	1	0.91	low	Yes	No
CRK11	AT4G23190	ROR1	2	1.96	high	Yes	Yes
MPK3	AT3G45640	MAPK1	6	6.07	moderate	Yes	No
MPK3	AT3G45640	MAPK3	6	6.07	high	Yes	Yes
MPK3	AT3G45640	MAPK8	3	3.12	moderate	No	Yes
MPK3	AT3G45640	MAPK9	3	3.12	moderate	No	Yes
MPK3	AT3G45640	MAPK10	3	3.12	moderate	No	Yes

MPK3	AT3G45640	MAPK14	3	3.11	moderate	No	Yes
MPK3	AT3G45640	MAPK7	3	2.91	low	No	No
MPK3	AT3G45640	MAPK6	2	2.11	moderate	No	Yes
MPK3	AT3G45640	MAPK4	2	2.11	moderate	No	Yes
MPK3	AT3G45640	MAPK13	2	2.11	moderate	No	Yes
MPK3	AT3G45640	MAPK12	2	2.11	moderate	No	Yes
TCH2	AT5G37770	KCNIP3	1	1.01	low	Yes	No
TCH2	AT5G37770	CIB2	1	1.01	low	Yes	Yes
TCH2	AT5G37770	CIB1	1	1.01	low	Yes	Yes
TCH2	AT5G37770	GUCA1B	1	1.01	low	Yes	No
TCH2	AT5G37770	PPP3R2	1	1.01	low	Yes	No
TCH2	AT5G37770	PPP3R1	1	1.01	low	Yes	No
TCH2	AT5G37770	CIB4	1	1.01	low	Yes	Yes
TCH2	AT5G37770	KCNIP1	1	1.01	low	Yes	No
TCH2	AT5G37770	CIB3	1	1.01	low	Yes	Yes
TCH2	AT5G37770	KCNIP2	1	1.01	low	Yes	No
TCH2	AT5G37770	GUCA1A	1	1.01	low	Yes	No
TCH2	AT5G37770	TESC	1	1.01	low	Yes	Yes
TCH2	AT5G37770	CHP2	1	1.01	low	Yes	No
TCH2	AT5G37770	GUCA1C	1	1.01	low	Yes	No
TCH2	AT5G37770	CHP1	1	1.01	low	Yes	No
TCH2	AT5G37770	RCVRN	1	1.01	low	Yes	No

4. ΣΥΖΗΤΗΣΗ

Από την απεικόνιση της PCA ((Εικόνα 9), που με χρώμα παρουσιάζονται οι διαφορετικές συνθήκες ακτινοβολήσης και με σύμβολα οι ώρες συλλογής των δειγμάτων παρατηρούμε ότι η διασπορά των τιμών είναι τέτοια που καταδεικνύει τις σημαντικότερες διαφορές τους. Από τις πρώτες κιόλας ώρες συλλογής των δειγμάτων (1.5 και 3 ωρών) διακρίνουμε άμεση και έντονη μεταγραφική δραστηριότητα στα ακτινοβολημένα με 30Gy HZE δείγματα, κάτι που δε συμβαίνει με τα δείγματα που έχουν υποστεί 100Gy ακτινοβολία-γ. Η σταδιακή αύξηση των DEGs έπειτα από έκθεση σε ακτινοβολία-γ κορυφώνεται μετά τη συλλογή του δείγματος στις 12 ώρες ενώ από τη συλλογή του δείγματος μετά τις 12 ώρες και εντός 24ώρου τα επίπεδα έκφρασης των γονιδίων των κυττάρων που έχουν εκτεθεί σε HZE ακτινοβολήση αρχίζουν να επανέρχονται σε φυσιολογικά επίπεδα φανερώνοντας τις διαφορές μεταξύ του τύπου και της δόσης ακτινοβολίας.

Τα ίδια συμπεράσματα προκύπτουν και από τα διαγράμματα Venn (Εικόνα 12, Εικόνα 13). Ο συνολικός αριθμός των διαφορετικώς εκφραζόμενων γονιδίων που προκύπτει μετά από HZE ακτινοβολήση στις 3 πρώτες ώρες ανέρχεται στα 850, διπλάσιος από κάθε άλλη ώρα συλλογής δείγματος. Αξιοσημείωτο είναι ότι όπως υπέδειξε η απεικόνιση PCA, τα κύτταρα φαίνεται σχεδόν να επανέρχονται στην ομοιόστασή τους μετά το πέρας του 24ώρου με μόλις 82 γονίδια να εκφράζονται διαφορετικώς μόνο στην περίπτωση έκθεσης HZE. Αντιθέτως, έντονη δραστηριότητα των επιπέδων έκφρασης των γονιδίων του

79

νεαρού βλαστού της *A. Thaliana* που έχει εκτεθεί σε 100Gy ακτινοβολίας-γ φαίνεται να συντελείται μετά τη συλλογή των δειγμάτων στις 12 και 24 ώρες.

Στην προσπάθεια μας να αναδείξουμε διαφορές στις βιολογικές διεργασίες που επάγονται έπειτα από ακτινοβολία θα πρέπει να λαμβάνεται υπόψιν τόσο ο τύπος όσο και η δόση της ακτινοβολίας. Μία πρώτη προσέγγιση ως προς την κατανόηση αυτής της σύνθετης απόκρισης είναι η δημιουργία βιολογικών δικτύων με τη βοήθεια της βάσης STRING εισάγοντας τα DEGs που βρέθηκαν σε όλα αλλά και στα 4 από τα 5 time points ανά τύπο ακτινοβολίας ξεχωριστά (Εικόνα 14, Εικόνα 15). Από την αναπαράσταση των δικτύου πρωτεϊνικών αλληλεπιδράσεων είναι φανερό ότι και οι δύο τύποι ιοντίζουσας ακτινοβολίας επάγουν μια ποικιλία βλαβών στο DNA με ταυτόχρονη ενεργοποίηση μηχανισμών απόκρισης στις βλάβες του DNA καθώς και μηχανισμών επιδιόρθωσης με κυριότερο τον ομόλογο ανασυνδυασμό (Homologous Recombination - HR). Το μονοπάτι του ομόλογου ανασυνδυασμού είναι ο κύριος μηχανισμός επιδιόρθωσης DSBs, το οποίο ενεργοποιείται κατά τη φάση G2 του κυτταρικού κύκλου, δηλαδή μετά το διπλασιασμό του DNA και λίγο πριν την έναρξη της μίτωσης (Zhao et al.). Τα γονιδιακά προϊόντα που βρέθηκαν να εμπλέκονται στο μηχανισμό επιδιόρθωσης του ομόλογου ανασυνδυασμού και στα δύο δίκτυα πρωτεϊνικών αλληλεπιδράσεων είναι τα BRCA1, RAD51, GMI1, και RecQ13. Μία ακόμη κοινή βιολογική διεργασία ανάμεσα στους δύο τύπους ακτινοβολίας είναι η κυτταρική γήρανση (senescence). Πρόκειται για μια μόνιμη κατάσταση παύσης του κυτταρικού κύκλου που πυροδοτείται από ποικίλους μηχανισμούς όπως οι βλάβες του DNA. Από την αναπαράσταση του δικτύου τα σχετιζόμενα με την κυτταρική γήρανση γονίδια έπειτα από έκθεση σε ακτινοβολία-γ είναι τα SAG101, SAG20, STR15 και WRKY22 ενώ τα σχετιζόμενα γονίδια έπειτα από

έκθεση σε HZE ακτινοβολία είναι τα ACS6, COR15A, SAG101, LTI78 και WRKY53.

Οι βασικές διαφορές που προκύπτουν ανάμεσα στα δύο δίκτυα είναι η εμφάνιση διεργασιών όπως το έμφυτο ανοσοποιητικό σύστημα καθώς και ο κιρκάδιος ρυθμός όταν το φυτό εκτεθεί σε ακτινοβολία HZE. Τα DEGs που συμμετέχουν στο έμφυτο ανοσοποιητικό σύστημα είναι τα MPK3, TCH2, HSPRO2, BCS1 και MES9. Η ισχυρή δόση αλλά και ο τύπος ακτινοβολήσης φαίνεται πως πυροδοτεί την ενεργοποίηση του έμφυτου ανοσοποιητικού συστήματος, του μόνου συστήματος ανοσίας που διαθέτουν τα φυτά. Τα φυτικά κύτταρα αδυνατώντας να επιδιορθώσουν τις βλάβες από τέτοιας εμβέλειας ακτινοβολήση οδηγούνται τελικά σε απόπτωση. Τα ενδογενή βλαπτικά μόρια που απελευθερώνονται από κατεστραμμένα, στρεσογόνα ή αποπτωτικά κύτταρα, γνωστά και ως DAMPs αλληλεπιδρούν με υποδοχείς αναγνώρισης προτύπων (PRRs) (Roh and Sohn). Ένας τέτοιος υποδοχέας φαίνεται πως αποτελεί το προϊόν της έκφρασης του γονιδίου MPK3. Μοναδικό εύρημα στο ίδιο δίκτυο αποτελεί επίσης και ο κιρκάδιος ρυθμός με τη συμμετοχή αρκετών γονιδίων όπως τα ERD7, CCL, COR15A, LTI78, LHY, CCA1, PRR7, RVE1 και PCC1. Τα γονίδια που σχετίζονται με τον κιρκάδιο ρυθμό είναι γνωστό ότι ρυθμίζουν μια πληθώρα κυτταρικών διεργασιών όπως ο κυτταρικός κύκλος, η απόπτωση και η επιδιόρθωση βλαβών του DNA. Τόσο οι οξειδωτικοί αμυντικοί μηχανισμοί όσο και η επισκευή των DSBs από HZE συγχρονίζονται από κιρκάδιους ρυθμούς (Kanakoglou et al.).

Από την ανάλυση εμπλουτισμού των DEGs ανά time point, που προήλθαν από έκθεση σε ακτινοβολία-γ (Πίνακας 9), προέκυψαν στατιστικά σημαντικά μεγάλες τιμές εμπλουτισμού σε λειτουργικές κατηγορίες που δεν αναδείχθηκαν από το δίκτυο πρωτεϊνικών αλληλεπιδράσεων. Βασικότερη από αυτές είναι η εμφάνιση του κιρκάδιου ρυθμού (circadian rhythm GO:0007623) σε όλα σχεδόν τα

time points, με πολύ υψηλό FDR μετά τη συλλογή του δείγματος στις 12 ώρες. Μία δεύτερη λειτουργική κατηγορία είναι αυτή της κυτταρικής θανάτωσης cell killing GO:0001906), μιας διαδικασίας που είναι περισσότερο ταυτόσημη με αυτή της νέκρωσης και ενεργοποιείται κατά τις πρώτες ώρες ακτινοβολήσης του φυτού. Στον πίνακα των αποτελεσμάτων παρατηρούμε υψηλές τιμές FDR για διαδικασίες που ρυθμίζουν τη μετά-εμβρυϊκή ανάπτυξη (regulation of post-embryonic development GO:0048580) καθώς επίσης θετική ρύθμιση του μονοπατιού απόκρισης στην αυξίνη (positive regulation of auxin mediated signaling pathway GO:0010929), μιας σημαντικής ορμόνης που συμμετέχει στην ανάπτυξη και μορφογένεση των φυτών.

Στην περίπτωση των αποτελεσμάτων της ανάλυσης εμπλουτισμού για τα DEGs που προέκυψαν από έκθεση του φυτού σε HZE ακτινοβολία (Πίνακας 10), τα αποτελέσματα συγκλίνουν σημαντικά. Σημαντικότερο εύρημα αποτελεί η ανοσολογική απόκριση του φυτού στην ακτινοβολία HZE, με την ενεργοποίηση του έμφυτου ανοσοποιητικού συστήματος (innate immune response GO:0045087) που επάγεται αδιαλείπτως μετά τη συλλογή του δείγματος την τρίτη ώρα. Οι τιμές FDR είναι αρκετά υψηλές μέχρι και το πέρας του 24ώρου. Η ενεργοποίηση του μηχανισμού της γήρανσης (aging GO:0007568) είναι επίσης εμφανής σε όλα σχεδόν τα time points. Πιθανές αιτίες που φαίνεται πως προάγουν το φαινότυπο της γήρανσης είναι η συσσώρευση βλαβών του DNA λόγω ακτινοβολίας που δεν μπορούν να επιδιορθωθούν ορθά, είτε επάγοντας την υπερέκφραση ογκοκατασταλτικών γονιδίων είτε αναστέλλοντας μη αναστρέψιμα τον κυτταρικό κύκλο (Chen). Η απόκριση στο οξειδωτικό στρες (response to oxidative stress GO:0006979) με την παραγωγή ελεύθερων ριζών ROS έχει συνδεθεί επίσης με το φαινόμενο της γήρανσης. Ο πίνακας αποτελεσμάτων έρχεται να επαληθεύσει τις υποθέσεις μας καθώς η απόκριση στο οξειδωτικό στρες είναι παρούσα σε όλα τα time point με πολύ υψηλές τιμές FDR ενώ απουσιάζει η

δραστηριότητα των μηχανισμών επιδιόρθωσης βλαβών του DNA οδηγώντας τα κύτταρα σε κυτταρική γήρανση ή σε απόπτωση από τις πρώτες κιόλας ώρες ακτινοβολήσης. Η απόπτωση που είναι μια διεργασία προγραμματισμένου κυτταρικού θανάτου (programmed cell death GO:0012501) αποτελεί έναν θεμελιώδη μηχανισμό ομοιόστασης, μια επαγόμενη και εντεταλμένη διαδικασία κατά την οποία το ίδιο το κύτταρο προκαλεί το θάνατο του χωρίς να προκαλεί τοπικές φλεγμονώδεις αποκρίσεις. Τέλος, η εμφάνιση της αυτοφαγίας (autophagy GO:0006914), μια διεργασίας γνωστή τόσο για την εμπλοκή της σε πρωτεΐνες που συμμετέχουν στους μηχανισμούς επιδιόρθωσης DSBs όσο και στην αποικοδόμηση των πρωτεϊνών αυτών όταν υπάρξει ανεπανόρθωτη βλάβη αποτελεί ισχυρή ένδειξη ότι μπορεί να οδηγήσει στην έναρξη προγραμματισμένου κυτταρικού θανάτου (Kindt et al.). Σε αυτή την υπόθεση συνηγορεί και η χρονική περίοδος ενεργοποίησης της αυτοφαγίας που συμπίπτει με την έναρξη της διεργασίας του προγραμματισμένου κυτταρικού θανάτου.

Σχετικά με τους πίνακες ορθολόγων γονιδίων (Πίνακας 11, Πίνακας 12) για κάθε είδους ακτινοβολία, θα παραθέσουμε εκείνα τα γονίδια ανάμεσα σε φυτό και άνθρωπο που βρέθηκαν να έχουν την υψηλότερη ομολογία βάση σκορ (DIOPT score) και ταξινόμησης (rank). Στα αποτελέσματα που προέκυψαν θα πρέπει να λάβουμε ισχυρά υπόψιν μας την αντίσταση που προβάλλουν τα φυτά στους διάφορους τύπους αλλά και τις δόσεις ακτινοβολήσης (Kim et al.).

Στην περίπτωση ακτινοβολήσης-γ ξεχώρισαν τα εξής ζεύγη ορθολόγων γονιδίων : AT4G19130-PRA1, ATMND1-MDN1, ATRAD17-RAD17, BRCA1-BARD1, BRCA1-BRCA1, PARP1-PARP1, PARP2-PARP2, RAD51-RAD51, RecQ13-RECQL5, TK1a-TK1, TSO2-RRM2, TSO2-RRM2B, CYCB1;1-CCNB2, CYCB1;1-CCNB1. Το γονίδιο RAD51 είναι γνωστό για την εμπλοκή του στον ομόλογο ανασυνδυασμό και την επιδιόρθωση του DNA. Επίσης

βρέθηκε να αλληλεπιδρά με τα γονίδια BRAC1 και BRAC2 τα οποία είναι σημαντικά για την κυτταρική απόκριση στις βλάβες του DNA. Η πρωτεΐνη που κωδικοποιείται από το γονίδιο RAD17 απαιτείται για την παύση του κυτταρικού κύκλου και την επιδιόρθωση βλαβών του DNA, προσδέεται σε σημείο του DNA πριν την εντοπισμένη βλάβη και φωσφορυλιώνεται από την ATR πρωτεϊνική κινάση. Οι πολυμεράσες της πόλυ (ADP-ριβόζης) PARP1, PARP2 αποτελούν βασικά μέλη οικογένειας πρωτεϊνικών ενζύμων που συμμετέχουν στην επιδιόρθωση του DNA. Τέλος το γονίδιο RECQL5 που κωδικοποιεί μία ελικάση διαδραματίζει σημαντικό ρόλο στη σταθερότητα του γονιδιώματος. Μία επίσης σημαντική λειτουργία του γονιδίου αυτού είναι ότι αποτρέπει παρεκκλίσεις του ομόλογου ανασυνδυασμού εκτοπίζοντας το γονίδιο RAD51 που έχει προσδεθεί σε ssDNA (Stelzer et al.). Στα φυτά, τα αντίστοιχα γονίδια επιτελούν παρόμοιες διεργασίες. Οι αλλαγές που συντελούνται στο μεταγράφημα ως απόκριση στις βλάβες του DNA από 100Gy ακτινοβολίας-γ καταδεικνύει μια έντονη δραστηριότητα υπερέκφρασης εκατοντάδων γονιδίων. Κύριος μεσολαβητής για αυτές τις αλλαγές είναι η πρωτεϊνική κινάση ATM που ενεργοποιείται από γενοτοξικές βλάβες που προκαλούν DSBs. Πολλά από τα επαγόμενα γονίδια κωδικοποιούν πρωτεΐνες που εμπλέκονται σε διεργασίες όπως η επιδιόρθωση του DNA, πιο συγκεκριμένα με τον μεταβολισμό του DNA και τον κυτταρικό κύκλο. Επίσης, παράγοντες που συμμετέχουν στο μονοπάτι του ομόλογου ανασυνδυασμού (HR) όπως το γονίδιο RAD51 καθώς και παραγόντων σηματοδότησης όπως τα RAD17, BRAC1 και PARP (Waterworth et al.).

Τα πλέον σημαντικά ορθόλογα γονίδια μεταξύ φυτού και ανθρώπου όπως προέκυψαν βάση του DIOPT σκορ έπειτα από HZE ακτινοβολία είναι τα εξής: CCA1-MYSM1, LHY-MYSM1, RVE1-MYSM1, TEJ-BPHL, AT4G19130-RPA1, ATMND1-MND1, ATRAD17-RAD17, BRCA1-BARD1, DUT1-DUT, PARP1-PARP1, PARP2-PARP2, RAD51-RAD51, RecQI3-RECQL5,

TK1a-TK1, TSO2-RRM2, ACS6-RRM2B, ACS6-ACCSL, ACS6-ACCS, BCS1-BCS1L, CRK11-ROR1, MPK3-MAPK1, MPK3-MAPK3. Τα ζεύγη ορθολόγων που σχετίζονται με μηχανισμούς επιδιόρθωσης του DNA παραμένουν και εδώ ίδια.

Ξεκινώντας με τη σειρά εμφάνισης των ορθολόγων, τα πρώτα γονίδια σχετίζονται με τον κirkάδιο ρυθμό. Το κirkάδιο ρολόι όπως και στους υπόλοιπους οργανισμούς έτσι και στα φυτά είναι ένας ενδογενής, μηχανισμός που συγχρονίζει βιολογικές διαδικασίες όταν συντελούνται καθημερινές και εποχικές αλλαγές στο περιβάλλον (Inoue et al.). Τα προϊόντα των γονιδίων CCA1 και LHY λειτουργούν συνεργατικά στη ρύθμιση του κirkάδιου ρυθμού της *A. Thaliana*, σχηματίζοντας ένα ετεροδιμερές. Πρόκειται ουσιαστικά για δύο στενά συνδεδεμένους μεταγραφικούς παράγοντες που περιέχουν τμήμα του MYB πρωτεϊνικού domain και προσδέονται στο DNA (Z. Y. Wang et al.). Η πρόσδεση τους στον υποκινητή του γονιδίου TOC1 ρυθμίζει αρνητικά την έκφραση του (Alabadi). Πέραν του ότι αποτελούν μέλη μιας θηλιάς ανατροφοδότησης για την κirkάδια ρυθμικότητα, πιθανόν να ρυθμίζουν την έκφραση γονιδίων που σχετίζονται με τις βιοχημικές και φυσιολογικές διεργασίες ως συνάρτηση του 24ώρου (Green and Tobin). Το αντίστοιχο ορθόλογο γονίδιο στον άνθρωπο είναι το MYSM1, το οποίο πιστεύεται ότι εμπλέκεται με στην καρκινογένεση και στο μελάνωμα. Μία μη λειτουργική μορφή του γονιδίου έχει αρκετές αναπτυξιακές ανωμαλίες που εξαρτώνται από το γονίδιο TP53, και συμπεριλαμβάνουν την αλλαγμένη μορφή και χρώση του δέρματος (Wilms et al.).

Έχει αποδειχθεί ότι η οικογένεια των MPKs γονιδίων που κωδικοποιούν πρωτεϊνικές κινάσες ενεργοποιούμενες από μιτογόνα αποτελούν βασικά στοιχεία του έμφυτου ανοσοποιητικού συστήματος του φυτού *A. Thaliana*. Πιο συγκεκριμένα, το γονίδιο MPK3 ενεργοποιείται από παθογόνα ή από ενδογενείς διεγέρτες των αμυντικών μηχανισμών όπως

τα ολιγογαλακτουρονίδια (OGs), τα οποία λειτουργούν ως σχετιζόμενα με βλάβες μοριακά πρότυπα (DAMPs) (Galletti et al.). Οι ενδογενείς διεγέρτες γνωστοί και ως damage-associated molecular patterns (DAMPs) που απελευθερώνονται από κατεστραμμένα, στρεσογόνα ή αποπτωτικά κύτταρα του φυτού ως απόκριση στο στρες που προκαλείται από δυσμενείς αβιοτικούς παράγοντες όπως η ακτινοβολία έχει ως αποτέλεσμα την ενεργοποίηση του έμφυτου ανοσοποιητικού συστήματος (Pavlioroulou et al.). Το αντίστοιχο ορθόλογο γονίδιο στον άνθρωπο κωδικοποιεί την πρωτεΐνη MAPK3 μέλος της οικογένειας MAPKs κινασών γνωστή ως εξωκυττάρια κινάση ρυθμιζόμενη από σήματα (ERKs). Λειτουργεί σαν σημείο ολοκλήρωσης για πολλαπλά βιοχημικά σήματα και εμπλέκεται σε μια ποικιλία κυτταρικών διεργασιών όπως ο πολλαπλασιασμός, η διαφοροποίηση και η πορεία του κυτταρικού κύκλου ως απόκριση σε μια ποικιλία εξωκυττάρων σημάτων. Η συμβολή του επίσης στην ενεργοποίηση του έμφυτου ανοσοποιητικού συστήματος είναι καταλυτική (Huang et al.).

Εν κατακλείδι, και στις δύο περιπτώσεις ακτινοβολίας η παρουσία επιδιορθωτικών μηχανισμών ως απόκριση στις βλάβες που επάγονται είναι εμφανή. Παρόλα αυτά, το φυτό με 30Gy ακτινοβολίας HZE αν και επάγει μηχανισμούς επιδιόρθωσης φαίνεται ότι ταυτόχρονα προσπαθεί να οδηγήσει το κύτταρο σε απόπτωση. Επιπλέον στις σοβαρές επιπτώσεις που φέρει αυτού του είδους η ακτινοβολία στο φυτό ανήκουν η ενεργοποίηση διεργασιών όπως το έμφυτο ανοσοποιητικό σύστημα και η απορρύθμιση του κερκάδιου ταλαντωτή. Παρόλα αυτά για να αυξηθεί η πιστότητα των αποτελεσμάτων της ανάλυσης μας, θα ήταν θεμιτό να είχαμε στη διάθεσή μας περισσότερα datasets μικροσυστοιχιών. Παράλληλα, ενώ έχει αλληλουχηθεί πλήρως το γενετικό υλικό της *A.Thaliana* και τα γονίδια της είναι καλά σχολιασμένα, δυστυχώς δεν υπάρχει πληθώρα εργαλείων για ανάλυση λειτουργικού εμπλουτισμού. Η βιβλιογραφία όσον αφορά στην

απόκριση του φυτού σε ακτινοβολία ΗΖΕ είναι ελλιπής. Καθώς όμως το σύνολο των αποτελεσμάτων για έκθεση σε ακτινοβολία-γ εμφανίζει μεγάλο ποσοστό συσχέτισης με τους μηχανισμούς DDR στον άνθρωπο, εξάγεται το συμπέρασμα ότι το φυτό μπορεί να χρησιμοποιηθεί ως μοντέλο για την απόκριση ΙR στον άνθρωπο.

5. ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

1. Abdi, Hervé, and Lynne J. Williams. "Principal Component Analysis." *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, 30 June 2010, pp. 433–459, 10.1002/wics.101.
2. Alabadi, D. "Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock." *Science*, vol. 293, no. 5531, 3 Aug. 2001, pp. 880–883, 10.1126/science.1061320. Accessed 12 Mar. 2019.
3. Alburger, D. E. "Nuclear Isomerism." *Kernreaktionen III / Nuclear Reactions III*, 1957, pp. 1–108, 10.1007/978-3-642-45878-1_1. Accessed 29 Sept. 2020.
4. Ashburner, Michael, et al. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics*, vol. 25, no. 1, May 2000, pp. 25–29, www.nature.com/articles/ng0500_25, 10.1038/75556.
5. Athar, Awais, et al. "ArrayExpress Update – from Bulk to Single-Cell Expression Data." *Nucleic Acids Research*, vol. 47, no. D1, 24 Oct. 2018, pp. D711–D715, 10.1093/nar/gky964. Accessed 28 Apr. 2020.
6. Aubrey, Brandon J, et al. "How Does P53 Induce Apoptosis and How Does This Relate to P53-Mediated Tumour Suppression?" *Cell Death & Differentiation*, vol. 25, no. 1, 17 Nov. 2017, pp. 104–113, www.ncbi.nlm.nih.gov/pmc/articles/PMC5729529/, 10.1038/cdd.2017.169.
7. Barrett, Tanya, et al. "NCBI GEO: Archive for Functional Genomics Data Sets—Update." *Nucleic Acids Research*, vol. 41, no. D1, 26 Nov. 2012, pp. D991–D995, 10.1093/nar/gks1193. Accessed 17 June 2019.
8. Bolstad, B.M., et al. "A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias." *Bioinformatics*, vol. 19, no. 2, 22 Jan. 2003, pp. 185–193, 10.1093/bioinformatics/19.2.185. Accessed 24 Feb. 2020.
9. Borràs-Fresneda, Mireia, et al. "Differences in DNA Repair Capacity, Cell Death and Transcriptional Response after Irradiation between a Radiosensitive and a Radioresistant Cell

- Line.” *Scientific Reports*, vol. 6, no. 1, 1 June 2016, 10.1038/srep27043. Accessed 18 Oct. 2020.
10. Borrego-Soto, Gissela, et al. “Ionizing Radiation-Induced DNA Injury and Damage Detection in Patients with Breast Cancer.” *Genetics and Molecular Biology*, vol. 38, no. 4, 24 Nov. 2015, pp. 420–432, 10.1590/s1415-475738420150019. Accessed 21 Nov. 2019.
 11. Branzei, Dana, and Marco Foiani. “Regulation of DNA Repair throughout the Cell Cycle.” *Nature Reviews Molecular Cell Biology*, vol. 9, no. 4, 20 Feb. 2008, pp. 297–308, 10.1038/nrm2351. Accessed 14 Mar. 2020.
 12. Brazma, Alvis, et al. “Minimum Information about a Microarray Experiment (MIAME)—toward Standards for Microarray Data.” *Nature Genetics*, vol. 29, no. 4, Dec. 2001, pp. 365–371, 10.1038/ng1201-365. Accessed 29 Apr. 2020.
 13. Carvalho, Benilton S., and Rafael A. Irizarry. “A Framework for Oligonucleotide Microarray Preprocessing.” *Bioinformatics*, vol. 26, no. 19, 5 Aug. 2010, pp. 2363–2367, 10.1093/bioinformatics/btq431. Accessed 7 May 2020.
 14. CDC. “Radiation Studies: Ionizing Radiation.” Centers for Disease Control and Prevention, 7 Dec. 2015, www.cdc.gov/nceh/radiation/ionizing_radiation.html.
 15. Charles, M. “Effects of Ionizing Radiation: United Nations Scientific Committee on the Effects of Atomic Radiation: UNSCEAR 2006 Report, Volume 1--Report to the General Assembly, with Scientific Annexes A and B.” *Radiation Protection Dosimetry*, vol. 138, no. 2, 26 Nov. 2009, pp. 187–189, 10.1093/rpd/ncp262. Accessed 2 Nov. 2019.
 16. Chatterjee, Nimrat, and Graham C Walker. “Mechanisms of DNA Damage, Repair, and Mutagenesis.” *Environmental and Molecular Mutagenesis*, vol. 58, no. 5, 2017, pp. 235–263, www.ncbi.nlm.nih.gov/pubmed/28485537, 10.1002/em.22087.
 17. Chen, Jiandong. “The Cell-Cycle Arrest and Apoptotic Functions of P53 in Tumor Initiation and Progression.” *Cold Spring Harbor Perspectives in Medicine*, vol. 6, no. 3, Mar. 2016, p. a026104, www.ncbi.nlm.nih.gov/pmc/articles/PMC4772082/, 10.1101/cshperspect.a026104.
 18. Chen, Z., et al. “A Distribution Free Summarization Method for Affymetrix GeneChip(R) Arrays.” *Bioinformatics*, vol. 23, no. 3, 5 Dec. 2006, pp. 321–327, 10.1093/bioinformatics/btl609. Accessed 17 Nov. 2020.
 19. Cui, Xiangqin, and Gary A Churchill. “Statistical Tests for Differential Expression in CDNA Microarray Experiments.” *Genome Biology*, vol. 4, no. 4, 2003, p. 210, www.ncbi.nlm.nih.gov/pmc/articles/PMC154570/, 10.1186/gb-2003-4-4-210. Accessed 28 Oct. 2020.

20. Culligan, Kevin M., et al. "ATR and ATM Play Both Distinct and Additive Roles in Response to Ionizing Radiation." *The Plant Journal*, vol. 48, no. 6, Dec. 2006, pp. 947–961, 10.1111/j.1365-313x.2006.02931.x. Accessed 5 May 2020.
21. Dalma-Weiszhausz, Dennise D., et al. "The Affymetrix GeneChip Platform: An Overview." *Methods in Enzymology*, vol. 410, 2006, pp. 3–28, pubmed.ncbi.nlm.nih.gov/16938544/, 10.1016/S0076-6879(06)10001-4. Accessed 3 Oct. 2020.
22. Dexheimer, Thomas S. "DNA Repair Pathways and Mechanisms." *DNA Repair of Cancer Stem Cells*, 27 July 2012, pp. 19–32, 10.1007/978-94-007-4590-2_2. Accessed 24 Sept. 2020.
23. "Draw Venn Diagram." *Bioinformatics.Psb.Ugent.Be*, bioinformatics.psb.ugent.be/webtools/Venn/.
24. Duggan, D. J., et al. "Expression Profiling Using CDNA Microarrays." *Nature Genetics*, vol. 21, no. 1 Suppl, 1 Jan. 1999, pp. 10–14, pubmed.ncbi.nlm.nih.gov/9915494/, 10.1038/4434. Accessed 3 Oct. 2020.
25. Elgazzar, Abdelhamid H, and Nafisah Kazem. *The Pathophysiologic Basis of Nuclear Medicine*. Berlin ; New York, Springer, 2006.
26. Elledge, Stephen J. "The DNA Damage Response—Self-Awareness for DNA." *JAMA*, vol. 314, no. 11, 15 Sept. 2015, p. 1111, 10.1001/jama.2015.10387. Accessed 10 Apr. 2020.
27. Exposure Due to Natural Sources.
28. Featherstone, Carol, and Stephen P Jackson. "DNA Double-Strand Break Repair." *Current Biology*, vol. 9, no. 20, Oct. 1999, pp. R759–R761, 10.1016/s0960-9822(00)80005-6.
28. Feldmann, Kenneth A., and Stephen A. Goff. "Chapter Four - The First Plant Genome Sequence—*Arabidopsis Thaliana*." *ScienceDirect, Academic Press*, 1 Jan. 2014, www.sciencedirect.com/science/article/pii/B978012417163300044. Accessed 16 Oct. 2020.
29. Flowers, Paul, et al. "21.6 Biological Effects of Radiation – Chemistry." *Opentextbc.Ca*, 2019, opentextbc.ca/chemistry/chapter/21-6-biological-effects-of-radiation/.
30. Friedberg, Errol C. "A Brief History of the DNA Repair Field." *Cell Research*, vol. 18, no. 1, 24 Dec. 2007, pp. 3–7, 10.1038/cr.2007.113. Accessed 25 Sept. 2020.
31. Galletti, Roberta, et al. "Arabidopsis MPK3 and MPK6 Play Different Roles in Basal and Oligogalacturonide- or Flagellin-Induced Resistance against *Botrytis Cinerea*." *Plant Physiology*, vol. 157, no. 2, 1 Oct. 2011, pp. 804–814.
32. Gandolfo, Luke C., and Terence P. Speed. "RLE Plots: Visualizing Unwanted Variation in High Dimensional Data." *PLOS ONE*, vol.

- 13, no. 2, 5 Feb. 2018, p. e0191629, 10.1371/journal.pone.0191629.
33. Gao, Xiaolian, et al. "In Situ Synthesis of Oligonucleotide Microarrays." *Biopolymers*, vol. 73, no. 5, 2004, pp. 579–596, 10.1002/bip.20005. Accessed 3 Oct. 2020.
 34. "GeneChip™ Arabidopsis Genome ATH1 Array." www.thermofisher.com, www.thermofisher.com/order/catalog/product/900385?SID=src-h-srp-900385#/900385?SID=src-h-srp-900385. Accessed 4 Oct. 2020.
 35. Goodarzi, A.A., et al. "Environmental Sources of Ionizing Radiation and Their Health Consequences." *Genome Stability*, 2016, pp. 569–581, 10.1016/b978-0-12-803309-8.00033-1. Accessed 30 Sept. 2020.
 36. Graham V, William J., et al. "DNA Mismatch Repair: Mechanisms and Cancer Genetics." Reference Module in Biomedical Sciences, 2018, 10.1016/b978-0-12-801238-3.96130-0. Accessed 24 Sept. 2020.
 37. Green, Rachel M., and Elaine M. Tobin. "The Role of CCA1 and LHY in the Plant Circadian Clock." *Developmental Cell*, vol. 2, no. 5, May 2002, pp.516–518, 10.1016/s1534-5807(02)00184-3. Accessed 30 Oct. 2019.
 38. Hakem, Razqallah. "DNA-Damage Repair; the Good, the Bad, and the Ugly." *The EMBO Journal*, vol. 27, no. 4, 20 Feb. 2008, pp. 589–605, 10.1038/emboj.2008.15. Accessed 6 Oct. 2019.
 39. Harper, J. Wade, and Stephen J. Elledge. "The DNA Damage Response: Ten Years After." *Molecular Cell*, vol. 28, no. 5, Dec. 2007, pp. 739–745, 10.1016/j.molcel.2007.11.015. Accessed 23 June 2020.
 40. Herrera, Henry J., and Marlon Gancino. "DNA Microarrays: Recent Advances." *Bionatura*, vol. 2, no. 3, 15 Aug. 2017, pp. 404–406, revistabionatura.com/files/2017.02.03.13.pdf, 10.21931/rb/2017.02.03.13. Accessed 17 Nov. 2019.
 41. Hu, Pingzhao, et al. "Tests for Differential Gene Expression Using Weights in Oligonucleotide Microarray Experiments." *BMC Genomics*, vol. 7, no. 1, 22 Feb. 2006, 10.1186/1471-2164-7-33.
 42. Hu, Yanhui, et al. "An Integrative Approach to Ortholog Prediction for Disease-Focused and Other Functional Studies." *BMC Bioinformatics*, vol. 12, no. 1, 2011, p. 357, 10.1186/1471-2105-12-357. Accessed 12 Oct. 2020.
 43. Huang, Gonghua, et al. "Regulation of JNK and P38 MAPK in the Immune System: Signal Integration, Propagation and Termination." *Cytokine*, vol. 48, no. 3, Dec. 2009, pp. 161–169, 10.1016/j.cyto.2009.08.002. Accessed 16 Oct. 2019.
 44. Huber, Wolfgang, et al. "Orchestrating High-Throughput Genomic Analysis with Bioconductor." *Nature Methods*, vol. 12,

- no. 2, 29 Jan. 2015, pp. 115–121, 10.1038/nmeth.3252. Accessed 6 May 2020.
45. Humans, IARC Working Group on the Evaluation of Carcinogenic Risks to. Ionizing Radiation, Part 1: X- and Gamma (γ)-Radiation, and Neutrons. PubMed, International Agency for Research on Cancer, 2000, www.ncbi.nlm.nih.gov/books/NBK401325/. Accessed 29 Sept. 2020.
 46. Iaea, International Atomic Energy Agency. Radiation, People and the Environment : ... A Broad Overview of Ionizing Radiation, Its Effects and Uses, as Well as the Measures in Place to Use It Safely. Vienna, Iaea International Atomic Energy Agency, 2004.
 47. Iliakis, George, et al. "DNA Damage Checkpoint Control in Cells Exposed to Ionizing Radiation." *Oncogene*, vol. 22, no. 37, 28 Aug. 2003, pp. 5834–5847, 10.1038/sj.onc.1206682. Accessed 19 Mar. 2020.
 48. Inoue, Keisuke, et al. "Circadian Clock during Plant Development." *Journal of Plant Research*, vol. 131, no. 1, 13 Nov. 2017, pp. 59–66, 10.1007/s10265-017-0991-8. Accessed 21 Nov. 2020.
 49. International Atomic Energy Agency. Radiation, People and the Environment : A Broad Overview of Ionizing Radiation, Its Effects and Uses, as Well as the Measures in Place to Use It Safely. Austria, International Atomic Energy Agency, 2007.
 50. Irizarry, R. A. "Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data." *Biostatistics*, vol. 4, no. 2, 1 Apr. 2003, pp. 249–264, 10.1093/biostatistics/4.2.249.
 51. ---. "Summaries of Affymetrix GeneChip Probe Level Data." *Nucleic Acids Research*, vol. 31, no. 4, 15 Feb. 2003, pp. 15e–115, 10.1093/nar/gng015. Accessed 5 Apr. 2020.
 52. Joonhong Ahn. *Reflections on the Fukushima Daiichi Nuclear Accident : Toward Social-Scientific Literacy and Engineering Resilience*. Cham, Springeropen, 2015.
 53. Kanakoglou, Dimitrios S., et al. "Effects of High-Dose Ionizing Radiation in Human Gene Expression: A Meta-Analysis." *International Journal of Molecular Sciences*, vol. 21, no. 6, 12 Mar. 2020, p. 1938, 10.3390/ijms21061938. Accessed 24 Nov. 2020.
 54. Kauffmann, Audrey, et al. "Importing ArrayExpress Datasets into R/Bioconductor." *Bioinformatics*, vol. 25, no. 16, 8 June 2009, pp. 2092–2094, 10.1093/bioinformatics/btp354. Accessed 7 Dec. 2019.
 55. Kauffmann, Audrey, and Wolfgang Huber. "Microarray Data Quality Control Improves the Detection of Differentially Expressed Genes." *Genomics*, vol. 95, no. 3, 1 Mar. 2010, pp. 138–142,

- www.sciencedirect.com/science/article/pii/S0888754310000042, 10.1016/j.ygeno.2010.01.003. Accessed 5 June 2020.
56. Kesari, Santosh, et al. "DNA Damage Response and Repair: Insights into Strategies for Radiation Sensitization of Gliomas." *Future Oncology*, vol. 7, no. 11, Nov. 2011, pp. 1335–1346, 10.2217/fon.11.111. Accessed 21 Sept. 2020.
 57. Kim, Jin-Baek, et al. "Differentially Expressed Genes in Response to Gamma-Irradiation during the Vegetative Stage in *Arabidopsis Thaliana*." *Molecular Biology Reports*, vol. 41, no. 4, 19 Jan. 2014, pp. 2229–2241, 10.1007/s11033-014-3074-0. Accessed 24 Nov. 2020.
 58. Kindt, Thomas J, et al. *Kuby Immunology*. New York, W.H. Freeman And Company, 2007.
 59. Klaus, Bernd, and Stefanie Reisenauer. "An End to End Workflow for Differential Gene Expression Using Affymetrix Microarrays." *F1000Research*, vol. 5, 3 July 2018, p. 1384, 10.12688/f1000research.8967.2.
 60. Krokan, Hans E, et al. "Base Excision Repair of DNA in Mammalian Cells." *FEBS Letters*, vol. 476, no. 1–2, 26 June 2000, pp. 73–77, 10.1016/s0014-5793(00)01674-4. Accessed 29 May 2020.
 61. Kusakabe, Masayuki, et al. "Mechanism and Regulation of DNA Damage Recognition in Nucleotide Excision Repair." *Genes and Environment*, vol. 41, no. 1, 25 Jan. 2019, 10.1186/s41021-019-0119-6. Accessed 25 Sept. 2020.
 62. L'Annunziata, Michael F. "1 - Alpha Radiation." *ScienceDirect, Elsevier Science B.V.*, 1 Jan. 2007, www.sciencedirect.com/science/article/pii/B9780444527158500049. Accessed 28 Sept. 2020.
 63. Lee, M.-L. T., et al. "Importance of Replication in Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive CDNA Hybridizations." *Proceedings of the National Academy of Sciences*, vol. 97, no. 18, 29 Aug. 2000, pp. 9834–9839, 10.1073/pnas.97.18.9834. Accessed 12 July 2020.
 64. Lesk, Arthur M. *Introduction to Genomics*. Oxford U.A., Oxford University Press, 2017.
 65. LINDAHL, T., and D.E. BARNES. "Repair of Endogenous DNA Damage." *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 65, no. 0, 1 Jan. 2000, pp. 127–134, 10.1101/sqb.2000.65.127. Accessed 25 Sept. 2020.
 66. Mavragani, Ifigeneia, et al. "Complex DNA Damage: A Route to Radiation-Induced Genomic Instability and Carcinogenesis." *Cancers*, vol. 9, no. 12, 18 July 2017, p. 91, 10.3390/cancers9070091. Accessed 30 Sept. 2020.
 67. Meinke, D. W. "*Arabidopsis Thaliana*: A Model Plant for Genome Analysis." *Science*, vol. 282, no. 5389, 23 Oct. 1998, pp. 662–

- 682, science.sciencemag.org/content/282/5389/662, 10.1126/science.282.5389.662. Accessed 13 Apr. 2019.
68. Métivier, Henri. "Natural Radiation Sources, Including Some Lessons for Nuclear Waste Management." *Comptes Rendus Physique*, vol. 3, no. 7–8, Sept. 2002, pp. 1035–1048, 10.1016/s1631-0705(02)01360-9. Accessed 24 Apr. 2020.
69. Mi, Huaiyu, et al. "Large-Scale Gene Function Analysis with the PANTHER Classification System." *Nature Protocols*, vol. 8, no. 8, 18 July 2013, pp. 1551–1566, www.nature.com/articles/nprot.2013.092, 10.1038/nprot.2013.092. Accessed 8 May 2019.
70. Miller, M. B., and Y.-W. Tang. "Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology." *Clinical Microbiology Reviews*, vol. 22, no. 4, 1 Oct. 2009, pp. 611–633, 10.1128/cmr.00019-09. Accessed 23 July 2019.
71. Missirian, Victor, et al. "Corrigendum: High Atomic Weight, High-Energy Radiation (HZE) Induces Transcriptional Responses Shared with Conventional Stresses in Addition to a Core 'DSB' Response Specific to Clastogenic Treatments." *Frontiers in Plant Science*, vol. 7, 21 Nov. 2016, 10.3389/fpls.2016.01753.
72. "Mutation, DNA Repair, and DNA Integrity | Learn Science at Scitable." *Nature.com*, 2014,
73. Nakad, Rania, and Björn Schumacher. "DNA Damage Response and Immune Defense: Links and Mechanisms." *Frontiers in Genetics*, vol. 7, 9 Aug. 2016, 10.3389/fgene.2016.00147. Accessed 1 Oct. 2020.
74. Nikitaki, Zacharenia, et al. "Bridging Plant and Human Radiation Response and DNA Repair through an In Silico Approach." *Cancers*, vol. 9, no. 6, 6 June 2017, 10.3390/cancers9060065. Accessed 18 Oct. 2020.
75. Nikolopoulos, Dimitrios, et al. *Ακτινοβολίες Περιβάλλοντος Και Άνθρωπος*. Repository.Kallipos.Gr, 2015,.
76. "NRC: Measuring Radiation." *Nrc.Gov*, 2017, www.nrc.gov/about-nrc/radiation/health-effects/measuring-radiation.html.
77. O'Brien, C.M. "Bioconductor Case Studies by Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon." *International Statistical Review*, vol. 77, no. 1, Apr. 2009, pp. 158–159, 10.1111/j.1751-5823.2009.00074_14.x.
78. O'Connor, Mark J. "Targeting the DNA Damage Response in Cancer." *Molecular Cell*, vol. 60, no. 4, Nov. 2015, pp. 547–560, , 10.1016/j.molcel.2015.10.040. Accessed 10 Nov. 2019.
79. Parikh, Sudip S, et al. "Base Excision Repair Enzyme Family Portrait: Integrating the Structure and Chemistry of an Entire DNA Repair Pathway." *Structure*, vol. 5, no. 12, Dec. 1997, pp.

- 1543–1550, 10.1016/s0969-2126(97)00303-1. Accessed 25 Sept. 2020.
80. Parkinson, H., et al. "ArrayExpress Update--from an Archive of Functional Genomics Experiments to the Atlas of Gene Expression." *Nucleic Acids Research*, vol. 37, no. Database, 1 Jan. 2009, pp. D868–D872, 10.1093/nar/gkn889. Accessed 29 Apr. 2020.
 81. Pavlopoulou, Athanasia, et al. "In Silico Phylogenetic and Structural Analyses of Plant Endogenous Danger Signaling Molecules upon Stress." *Oxidative Medicine and Cellular Longevity*, vol. 2019, 15 July 2019, pp. 1–14, 10.1155/2019/8683054. Accessed 22 Nov. 2020.
 82. Peirce, J. Jeffrey, et al. "Radioactive Waste." *Environmental Pollution and Control*, 1998, pp. 211–231, 10.1016/b978-075069899-3/50017-1. Accessed 28 Sept. 2020.
 83. "Radiation -- ANS / About Nuclear." *Www.Ans.org*, www.ans.org/nuclear/radiation/. Accessed 29 Sept. 2020.
 84. Ray, Paul D., and Rebecca C. Fry. "Chapter 2 - The Cell: The Fundamental Unit in Systems Biology." *ScienceDirect*, Academic Press, 1 Jan. 2015, www.sciencedirect.com/science/article/pii/B97801280156430002X. Accessed 21 Sept. 2020.
 85. Ritchie, Matthew E., et al. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research*, vol. 43, no. 7, 20 Jan. 2015, pp. e47–e47, academic.oup.com/nar/article/43/7/e47/2414268, 10.1093/nar/gkv007. Accessed 30 May 2019.
 86. RoC Profile: Ionizing Radiation; 14th RoC 2016. 2016.
 87. Rocca-Serra, Philippe, et al. "ArrayExpress: A Public Database of Gene Expression Data at EBI." *Comptes Rendus Biologies*, vol. 326, no. 10–11, Oct. 2003, pp. 1075–1078, 10.1016/j.crv.2003.09.026. Accessed 6 May 2020.
 88. Roh, Jong Seong, and Dong Hyun Sohn. "Damage-Associated Molecular Patterns in Inflammatory Diseases." *Immune Network*, vol. 18, no. 4, 2018, 10.4110/in.2018.18.e27.
 89. "RSSC BIOLOGICAL EFFECTS OF IONIZING RADIATION 08/11 5-1 CHAPTER 5 BIOLOGICAL EFFECTS OF IONIZING RADIATION." *Semantic Scholar*, 2012, www.semanticscholar.org/paper/RSSC-BIOLOGICAL-EFFECTS-OF-IONIZING-RADIATION-08%2F11/d99fd4b62ccf6e351c449be1d5ef9022a1370ffc.
 90. Safety Commission, Canadian Nuclear. "Introduction to Radiation." *Nuclearsafety.Gc.Ca*, 3 Feb. 2014, nuclearsafety.gc.ca/eng/resources/radiation/index.cfm. Accessed 28 Sept. 2020.

91. Santivasi, Wil L., and Fen Xia. "Ionizing Radiation-Induced DNA Damage, Response, and Repair." *Antioxidants & Redox Signaling*, vol. 21, no. 2, 10 July 2014, pp. 251–259, 10.1089/ars.2013.5668. Accessed 7 Apr. 2019.
92. Stelzer, Gil, et al. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." *Current Protocols in Bioinformatics*, 20 June 2016, pp. 1.30.1-1.30.33, 10.1002/cpbi.5. Accessed 4 Apr. 2019.
93. Szklarczyk, Damian, et al. "STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets." *Nucleic Acids Research*, vol. 47, no. Database issue, 8 Jan. 2019, pp. D607–D613, www.ncbi.nlm.nih.gov/pmc/articles/PMC6323986/, 10.1093/nar/gky1131. Accessed 19 Feb. 2020.
94. Trent, Ronald J. "Development, Aging and Cancer." *Molecular Medicine*, vol. 346, no. 978-0-12-381451–7, 2012, pp. 203–243, 10.1016/b978-0-12-381451-7.00007-4.
95. Tubbs, Anthony, and André Nussenzweig. "Endogenous DNA Damage as a Source of Genomic Instability in Cancer." *Cell*, vol. 168, no. 4, 2017, pp. 644–656, www.ncbi.nlm.nih.gov/pubmed/28187286, 10.1016/j.cell.2017.01.002. Accessed 16 Oct. 2019.
96. United Nations. Scientific Committee On The Effects Of Atomic Radiation, and United Nations Environment Programme. *Radiation Effects and Sources : What Is Radiation? What Does Radiation Do to Us? Where Does Radiation Come From?. Nairobi, Kenya?*, United Nations Environment Programme, 2016.
97. US EPA, OAR. "Radiation Basics." US EPA, 12 Nov. 2014, www.epa.gov/radiation/radiation-basics.
98. ---. "Radiation Terms and Units." US EPA, 20 June 2017, www.epa.gov/radiation/radiation-terms-and-units. Accessed 17 Sept. 2020.
99. Vogiannis, Efstratios G., and Dimitrios Nikolopoulos. "Radon Sources and Associated Risk in Terms of Exposure and Dose." *Frontiers in Public Health*, vol. 2, 5 Jan. 2015, 10.3389/fpubh.2014.00207. Accessed 29 June 2020.
100. Wang, Hao, et al. "Cancer Radiosensitizers." *Trends in Pharmacological Sciences*, vol. 39, no. 1, Jan. 2018, pp. 24–48, 10.1016/j.tips.2017.11.003. Accessed 25 Oct. 2020.
101. Wang, Z. Y., et al. "A Myb-Related Transcription Factor Is Involved in the Phytochrome Regulation of an Arabidopsis Lhcb Gene." *The Plant Cell*, vol. 9, no. 4, 1 Apr. 1997, pp. 491–507, 10.1105/tpc.9.4.491. Accessed 21 Nov. 2020.
102. Waterworth, Wanda M., et al. "Repairing Breaks in the Plant Genome: The Importance of Keeping It Together." *New*

- Phytologist, vol. 192, no. 4, 11 Oct. 2011, pp. 805–822, 10.1111/j.1469-8137.2011.03926.x. Accessed 18 Nov. 2020.
103. Wickham, Hadley. “Tidy Data.” *Journal of Statistical Software*, vol. 59, no. 10, 2014, 10.18637/jss.v059.i10. Accessed 19 Dec. 2019.
 104. Wilkinson, Leland. “Ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H.” *Biometrics*, vol. 67, no. 2, June 2011, pp. 678–679, 10.1111/j.1541-0420.2011.01616.x.
 105. Williamson, Elizabeth A., et al. “Chapter 8 - Overview for the Histone Codes for DNA Repair.” *ScienceDirect*, Academic Press, 1 Jan. 2012, www.sciencedirect.com/science/article/pii/B9780123876652000080. Accessed 20 Sept. 2020.
 106. Wilms, Christina, et al. “MYSM1/2A-DUB Is an Epigenetic Regulator in Human Melanoma and Contributes to Tumor Cell Growth.” *Oncotarget*, vol. 8, no. 40, 27 June 2017, pp. 67287–67299, www.ncbi.nlm.nih.gov/pmc/articles/PMC5620173/, 10.18632/oncotarget.18617. Accessed 23 Nov. 2020.
 107. World. “Ionizing Radiation, Health Effects and Protective Measures.” *Who.Int*, World Health Organization: WHO, 29 Apr. 2016, www.who.int/news-room/fact-sheets/detail/ionizing-radiation-health-effects-and-protective-measures.
 108. Xu, H., and M. R. El-Gewely. “P53 Network — Its Downstream Regulated Genes.” *Biochemical Society Transactions*, vol. 28, no. 5, 1 Oct. 2000, pp. A227–A227, 10.1042/bst028a227a. Accessed 21 Sept. 2020.
 109. Zeman, Michelle K., and Karlene A. Cimprich. “Causes and Consequences of Replication Stress.” *Nature Cell Biology*, vol. 16, no. 1, 24 Dec. 2013, pp. 2–9, 10.1038/ncb2897. Accessed 25 Sept. 2019.
 110. Zhao, Xin, et al. “Cell Cycle-Dependent Control of Homologous Recombination.” *Acta Biochimica et Biophysica Sinica*, vol. 49, no. 8, 25 May 2017, pp. 655–668, 10.1093/abbs/gmx055. Accessed 23 Nov. 2020.
 111. “Ελληνική Επιτροπή Ατομικής Ενέργειας - Αρχική.” *Www.Eeae.Gr*, www.eeae.gr.
 112. Abdi, Hervé, and Lynne J. Williams. “Principal Component Analysis.” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, 30 June 2010, pp. 433–459, 10.1002/wics.101.
 113. Alabadi, D. “Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock.” *Science*, vol. 293, no. 5531, 3 Aug. 2001, pp. 880–883, 10.1126/science.1061320. Accessed 12 Mar. 2019.

114. Alburger, D. E. "Nuclear Isomerism." *Kernreaktionen III / Nuclear Reactions III*, 1957, pp. 1–108, 10.1007/978-3-642-45878-1_1. Accessed 29 Sept. 2020.
115. Ashburner, Michael, et al. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics*, vol. 25, no. 1, May 2000, pp. 25–29, www.nature.com/articles/ng0500_25, 10.1038/75556.
116. Athar, Awais, et al. "ArrayExpress Update – from Bulk to Single-Cell Expression Data." *Nucleic Acids Research*, vol. 47, no. D1, 24 Oct. 2018, pp. D711–D715, 10.1093/nar/gky964. Accessed 28 Apr. 2020.