



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Χημικών Μηχανικών

Τομέας IV: Σύνθεσης και Ανάπτυξης Βιομηχανικών Διαδικασιών

Διπλωματική Εργασία

**Βιοπληροφορική ανάλυση μεταγραφωμικών δεδομένων για τη
μελέτη του μικροβιώματος στο ανθρώπινο ηπατοκυτταρικό
καρκίνωμα**

Διονύσιος Δημητρέας

Επιβλέπουσα:
Επίκουρη Καθηγήτρια ΕΜΠ Διομή Μαμμά

Αθήνα 2023

Διονύσιος Δημητριάς

**Βιοπληροφορική ανάλυση μεταγραφωμικών δεδομένων για τη μελέτη του
μικροβιώματος στο ανθρώπινο ηπατοκυτταρικό καρκίνωμα**

Διπλωματική Εργασία, Σχολή Χημικών Μηχανικών

Τομέας IV: Σύνθεσης και Ανάπτυξης Βιομηχανικών Διαδικασιών

Εθνικό Μετσόβιο Πολυτεχνείο

Επιβλέπουσα: Επίκουρη Καθηγήτρια ΕΜΠ Διομή Μαμμά

Περίληψη

Το ηπατοκυτταρικό καρκίνωμα (HCC) είναι η πιο κοινή μορφή καρκίνου του ήπατος που κατατάσσεται τρίτη σε συχνότητα αιτία θανάτου από καρκίνο. Ωστόσο μέχρι σήμερα δεν υπάρχουν διαθέσιμες επιλογές πρόγνωσης, έγκαιρης διάγνωσης και αποτελεσματικής θεραπείας. Τα τελευταία χρόνια έχει αποδειχτεί η στενή συσχέτιση μεταξύ του μικροβιώματος, ιδιαίτερα του εντέρου, και της εμφάνισης καρκίνου και έχει αναδειχτεί η παρουσία διακριτού ενδοογκικού μικροβιώματος σε πολλούς τύπους όγκων, υποστηρίζοντας ότι ο προσδιορισμός μικροβιακών βιοδεικτών μπορεί να συμβάλλει στην εξατομικευμένη αντιμετώπιση του HCC. Στην παρούσα Διπλωματική Εργασία χρησιμοποιήθηκαν διαθέσιμα δεδομένα αλληλούχισης RNA από ηπατικούς ιστούς 25 ασθενών με HCC και κατάλληλες μέθοδοι Βιοπληροφορικής και Βιοστατιστικής, με σκοπό τη μελέτη του ενδοογκικού μικροβιώματος και της σχέσης του με τον καρκίνο και τους υποτύπους υψηλής και χαμηλής διαφοροποίησης. Οι αναλύσεις πραγματοποιήθηκαν σε καρκινικά δείγματα και δείγματα ελέγχου, από παρακαρκινικούς ιστούς των ασθενών. Μία αυτοματοποιημένη ροή διεργασιών αναπτύχθηκε σε περιβάλλον γλώσσας προγραμματισμού R, με τη χρήση προγραμματιστικών πακέτων του αποθετηρίου Bioconductor.

Αρχικά, τα μη επεξεργασμένα αναγνώσματα των δειγμάτων αναλύθηκαν και επεξεργάστηκαν ως προς την ποιότητά τους, και ευθυγραμμίστηκαν στο ανθρώπινο γονιδίωμα με το εργαλείο HISAT2. Η Ανάλυση σε Κύριες Συνιστώσες για τα ευθυγραμμισμένα αναγνώσματα των καρκινικών δειγμάτων, σε συνδυασμό με την Ανάλυση Εμπλουτισμού σε Σύνολα Γονιδίων κατηγοριοποίησε τα καρκινικά δείγματα σε 2 φαινοτύπους, υψηλής και χαμηλής διαφοροποίησης. Τόσο τα καρκινικά με τα παρακαρκινικά, όσο και οι 2 υπο-ομάδες καρκινικών δειγμάτων, υποβλήθηκαν σε ανάλυση διαφορικής γονιδιακής έκφρασης με το εργαλείο edgeR, και στη συνέχεια, τα διαφορικά εκφρασμένα γονίδια χαρακτηρίστηκαν ως προς τις λειτουργίες του σε όρους Γονιδιακής Οντολογίας και μονοπάτια Reactome από το σετ εργαλείων BioInfoMiner. Τα αναγνώσματα που δεν ευθυγραμμίστηκαν στο ανθρώπινο γονιδίωμα εισήχθησαν στο εργαλείο μετα-μεταγραφωμικής SAMSA2, προκειμένου να εντοπιστούν τα μικροβιακά αναγνώσματα, δίνοντας πληροφορίες αναφορικά με τα βακτήρια και τις βιολογικές λειτουργίες που εκφράζονται. Οι αναλύσεις που πραγματοποιήθηκαν στο μικροβίωμα περιλαμβάνουν τη σχετική αφθονία, άλφα- και βήτα-ποικιλομορφία, και

διαφορική αφθονία των ειδών και των ενζύμων, τόσο για τα καρκινικά και παρακαρκινικά δείγματα, όσο και για τις υπο-ομάδες των καρκινικών δειγμάτων.

Κατά την ανάλυση διαφορικής έκφρασης των ανθρώπινων γονιδίων, εντοπίστηκαν 959 διαφορεικά εκφρασμένα γονίδια, 773 υπο- και 186 υπερ-εκφρασμένα γονίδια στα καρκινικά δείγματα, με τις κύριες βιολογικές λειτουργίες που εκφράζουν να συσχετίζονται με τον κυτταρικό κύκλο και την καταστολή του ανοσοποιητικού συστήματος. Για τις υπο-ομάδες των καρκινικών δειγμάτων, ο λειτουργικός χαρακτηρισμός με το BioInfoMiner έδειξε πως τα διαφορετικά εκφρασμένα γονίδια εκφράζουν λειτουργίες σχετικές με τον καταβολισμό και μεταβολισμό ουσιών, σχετικών με την παραγωγή ενέργειας στο κύτταρο. Η ανάλυση των βακτηριακών οργανισμών στα δείγματα, σε επίπεδο σχετικής αφθονίας εντόπισε στατιστικά σημαντική μείωση των *Actinobacteria* και αύξηση των *Proteobacteria* στα καρκινικά δείγματα. Η άλφα-ποικιλομορφία των καρκινικών δειγμάτων δεν βρέθηκε να έχει στατιστικά σημαντικές διαφορές σε σχέση με τα δείγματα ελέγχου, ενώ η ανάλυση της βήτα-ποικιλομορφίας έδειξε πως τα καρκινικά δείγματα εμφανίζουν υψηλότερη μεταβλητότητα μεταξύ τους. Επιπλέον, κατά την ανάλυση της διαφορικής αφθονίας, το είδος *Lactococcus lactis* βρέθηκε υπο-εκφρασμένο στα καρκινικά δείγματα, ενώ είδη των γενών *Aeromonas* και *Mycoplasma*, υπερ-εκφρασμένα. Σε επίπεδο βιολογικών λειτουργιών, ένζυμα που σχετίζονται με την παραγωγή ενέργειας σε φυσιολογικά κύτταρα, όπως η Phosphoenolpyruvate carboxykinase, εντοπίστηκαν υπο-εκφρασμένα στα καρκινικά δείγματα ενώ οι ATP synthase και Chaperone DnaK, που σχετίζονται με την ογκογένεση, αγγειογένεση και μετάσταση, βρέθηκαν υπερ-εκφρασμένες. Τέλος, η ανάλυση της διαφορικής αφθονίας, τόσο για τους μικροοργανισμούς, όσο και για τα ένζυμα, ανάμεσα στα δείγματα των υπο-ομάδων και ελέγχου, αλλά και μεταξύ των υπο-ομάδων, αναδεικνύει τον ενδιαμέσο χαρακτήρα της υπο-ομάδας καρκινικών δειγμάτων υψηλής διαφοροποίησης.

Λέξεις-κλειδιά: ηπατοκυτταρικό καρκίνωμα, ενδοογκικό μικροβίωμα, αλληλούχιση RNA, μεταγραφωμική

Abstract

Bioinformatics analysis of transcriptomic data for the study of microbiota in human hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is the most common form of liver cancer, ranking third most common cause of cancer death. However, to date there are no options available for prognosis, early diagnosis, and effective treatment. In recent years, the close association between the microbiome, particularly the gut microbiome, and cancer incidence has been demonstrated and the presence of a distinct intratumoral microbiome in many cancer types has been highlighted, suggesting that the identification of microbial biomarkers may contribute to the personalised treatment of HCC. In the present Diploma Thesis, available RNA sequencing data from liver tissues of 25 HCC patients and appropriate bioinformatics and biostatistics methods were used to study the intratumoral microbiome and its relationship with cancer, and high- and low-differentiation subtypes. The analyses were performed on cancer samples and peritumoral controls. An automated pipeline was developed in an R programming language environment, using programming packages from the Bioconductor repository.

Initially, raw sample reads were analyzed and processed for quality, and aligned to the human genome, using the HISAT2 tool. Principal Component Analysis for the aligned reads of tumor samples, combined with Gene Set Enrichment Analysis, categorized them into 2 phenotypes of well- and poorly-differentiated tumors. Differential gene expression analysis was conducted for the comparison between tumor and peritumoral samples, as well as between 2 tumor subgroups, with the edgeR tool, and the differentially expressed genes were annotated to their functions, in terms of Gene Ontology and Reactome pathways, by the BioInfoMiner tool set. Unmapped reads to the human genome were imported into the SAMSA2 metatranscriptomics tool, in order to identify microbial reads, providing information regarding the taxonomy and the biological functions expressed. Analyses performed on the microbiome included relative abundance, alpha- and beta-diversity, and differential abundance of species and enzymes for both tumor and peritumoral samples, as well as for the subgroups of tumor samples.

In the differential expression analysis of human genes, 959 differentially expressed genes, 773 down- and 186 up-regulated genes were identified in the tumor samples, with the

main biological functions they express correlating with the cell cycle and immune suppression. For the subgroups of tumor samples, functional characterization with BioInfoMiner indicated that differentially expressed genes express functions related to catabolism and metabolism of substances relevant to energy production in the cell. Analysis of bacterial organisms in the samples at the level of relative abundance identified a statistically significant decrease in *Actinobacteria*, and an increase in *Proteobacteria* in the tumor samples. Tumor samples' alpha-diversity was found to have no statistically significant differences compared to the control group, while the beta-diversity analysis showed that tumor samples exhibited higher intra-variability. Furthermore, in the differential abundance analysis, *Lactococcus lactis* was found to be under-expressed in tumor samples, whereas species of the genera *Aeromonas* and *Mycoplasma* were over-expressed. At the level of biological functions, enzymes related to energy production in normal cells, such as Phosphoenolpyruvate carboxykinase, were found under-expressed in tumor samples, while ATP synthase and Chaperone DnaK were found overexpressed, which are associated with tumorigenesis, angiogenesis, and metastasis. Finally, the analysis of differential abundance, for both microorganisms and enzymes, between the subgroups and control samples, and between the 2 subgroups highlights the intermediate characteristics of the well-differentiated tumor subgroup.

Keywords: hepatocellular carcinoma, intratumoral microbiome, RNA sequencing, transcriptomics

Κατάλογος Εικόνων

Εικόνα 1. Οι μοριακοί μηχανισμοί που συνδέονται με την ηπατοκαρκινογένεση.....	13
Εικόνα 2. Διαγραμματική απεικόνιση του μηχανισμού παθολογικών μεταβολών καρκινογένεσης: από τη φλεγμονή υγιούς ήπατος στο HCC.	15
Εικόνα 3. Η επίδραση του περιβάλλοντος αποτελεί ένα σύνθετο δίκτυο αλληλεπιδράσεων ανάμεσα σε ποικίλους παράγοντες	21
Εικόνα 4. Η επίδραση του αλκοόλ στο ήπαρ.....	22
Εικόνα 5. Μετατόπιση μεταβολιτών και μοριακών μοτίβων σχετιζόμενων με παθογόνα (PAMPs) που βιοσυντίθενται από μικρόβια στο έντερο μέσω της πυλαίας κυκλοφορίας στο ήπαρ, όπου ασκούν πολλαπλές επιδράσεις στην κατάσταση της υγείας του	24
Εικόνα 6. Αλλοιωμένος άξονας εντέρου-ήπατος σε ασθενείς με HCC.....	26
Εικόνα 7. Μοντέλο της βακτηριακής εισβολής στον καρκινικό ιστό του ξενιστή.....	30
Εικόνα 8. Διάγραμμα διαδικασίας RNA-seq.....	32
Εικόνα 9. Διάγραμμα απεικόνισης της αύξησης των μεταγραφωμικών ερευνών.....	37
Εικόνα 10. Η δομή της FASTA μορφής.....	39
Εικόνα 11. Η δομή της FASTQ μορφής.....	40
Εικόνα 12. Παραδείγματα της αναφοράς FastQC και των σημαντικότερων χαρακτηριστικών που λαμβάνονται υπόψη.	41
Εικόνα 13. Παράδειγμα του λογισμικού BioInfoMiner και των δεδομένων εισόδου για την ανάλυση.	47
Εικόνα 14. Μετα-μεταγραφωμική ανάλυση με το εργαλείο SAMSA2.....	49
Εικόνα 15. Κύριο διάγραμμα ροής περιγραφής της υπολογιστικής διαδικασίας.	56
Εικόνα 16. GSVA ανάλυση για τα καρκινικά δείγματα.....	60
Εικόνα 17. Ομαδοποίηση δειγμάτων με τη χρήση της μεθόδου Ck-means 1 διάστασης.	61
Εικόνα 18. Αναφορά αποτελεσμάτων FastQC για τον έλεγχο ποιότητας αλληλουχίας ανά βάση για το δείγμα SRR6188338, πριν (A) και μετά (B) τη χρήση του εργαλείου αποκοπής Trimmomatic.....	65
Εικόνα 19. Αρχείο κειμένου του αποτελέσματος ευθυγράμμισης από το εργαλείο HISAT2 για το δείγμα SRR6188338.....	66
Εικόνα 20. Τα φυσιολογικά (κόκκινο) και τα καρκινικά (γαλάζιο) δείγματα διαχωρίζονται ποιοτικά στην ανάλυση κυρίων συνιστωσών (PCA) βάσει των καταμετρήσεων των αναγνωσμάτων τους.....	67
Εικόνα 21. Θερμικός χάρτης απεικόνισης των καταμετρήσεων κάθε γονιδίου (γραμμές) στα δείγματα (στήλες).	68
Εικόνα 22. Volcano plot ανάλυσης διαφορικής έκφρασης γονιδίων της σύγκρισης Tumor vs Normal δειγμάτων.....	68
Εικόνα 23. PCA στα καρκινικά δείγματα, για τα 1000 γονίδια με την υψηλότερη μεταβλητότητα.	70
Εικόνα 24. Ομαδοποίηση καρκινικών δειγμάτων βάσει του βαθμού εμπλουτισμού του συνόλου γονιδίων HSIAO Liver Specific Genes στην PC1.....	71
Εικόνα 25. Ανάλυση διαφορικής έκφρασης γονιδίων μεταξύ Poorly και Well δειγμάτων.	72
Εικόνα 26. Ανάλυση εμπλουτισμού λειτουργιών GO συσχετίζοντας τα διαφορικά εκφρασμένα γονίδια με στατιστικά σημαντικές βιολογικές λειτουργίες.....	73
Εικόνα 27. Περιγραφική απεικόνιση της σχετικής αφθονίας του μικροβιώματος των δειγμάτων των δύο κατηγοριών, Normal (αριστερά) και Tumor (δεξιά), σε επίπεδο φύλων.....	74
Εικόνα 28. Στατιστικός έλεγχος Wilcoxon ($p < 0.05$) της σχετικής αφθονίας του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal και Tumor, σε επίπεδο φύλου.....	75
Εικόνα 29. Ανάλυση άλφα-ποικιλομορφίας των δειγμάτων του μικροβιώματος για τα παρατηρούμενα (Observed) OTUs, το δείκτη Shannon και Simpson.	76

Εικόνα 30. Heatmap απεικόνισης της ανομοιοτητας των δειγμάτων κατά την ανάλυση της βήτα ποικιλομορφίας με τη χρήση του μέτρου ανομοιοτητας Bray-Curtis.....	77
Εικόνα 31. Διάγραμμα διασποράς PCoA για την ανάλυση της βήτα-ποικιλομορφίας μεταξύ των δειγμάτων, με τη μέθοδο ανομοιοτητας Bray-Curtis.....	77
Εικόνα 32. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Tumor vs Normal. Κάθε κουκίδα αντιστοιχεί σε ένα είδος.....	78
Εικόνα 33. Περιγραφική απεικόνιση της σχετικής αφθονίας των Level 1 βιολογικών λειτουργιών του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal (αριστερά) και Tumor (δεξιά).	79
Εικόνα 34. Στατιστικός έλεγχος Wilcoxon ($p < 0.05$) της σχετικής αφθονίας των βιολογικών λειτουργιών Level 1 του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal και Tumor.	79
Εικόνα 35. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Tumor vs Normal.....	80
Εικόνα 36. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Well vs Normal.	81
Εικόνα 37. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Poorly vs Normal	82
Εικόνα 38. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Well vs Normal.....	83
Εικόνα 39. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Normal.....	84
Εικόνα 40. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Well	85
Εικόνα 41. Εύρεση βέλτιστου αριθμού υπο-ομάδων με το κριτήριο BIC.....	90
Εικόνα 42. Αποτελέσματα ομαδοποίησης βάσει του βαθμού εμπλουτισμού των δειγμάτων στα σύνολα γονιδίων.....	91
Εικόνα 43. Αποτελέσματα ομαδοποίησης των καρκινικών δειγμάτων βάσει του βαθμού εμπλουτισμού με των συνόλων γονιδίων στην PC1	92
Εικόνα 44. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης Tumor-Normal με βιολογικές λειτουργίες σε όρους GO από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner	93
Εικόνα 45. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης Tumor-Normal με βιολογικές λειτουργίες σε όρους Reactome Pathways από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner	94
Εικόνα 46. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης των υπο-ομάδων καρκινικών δειγμάτων, Poorly-Well, με βιολογικές λειτουργίες σε όρους Γονιδιακής Οντολογίας (Gene Ontology) από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner	94
Εικόνα 47. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης των υπο-ομάδων καρκινικών δειγμάτων, Poorly-Well, με βιολογικές λειτουργίες σε όρους Reactome Pathways από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner... ..	94

Κατάλογος Πινάκων

Πίνακας 1. Κατηγορίες HCC κατά Hoshida και τα χαρακτηριστικά τους.	17
Πίνακας 2. Διαφοροποιήσεις στο εντερικό μικροβίωμα ανθρώπων από έρευνες σε φυσιολογικούς και HCC ασθενείς.	27
Πίνακας 3. Συντελεστές συσχέτισης της PC1 με τους βαθμούς εμπλουτισμού των συνόλων γονιδίων.	71
Πίνακας 4. Οι υπο-ομάδες που σχηματίστηκαν διαχωρίζουν τα καρκινικά δείγματα βάσει της διαφοροποίησης των κυττάρων.	71
Πίνακας 5. Βαθμοί εμπλουτισμού (Enrichment Scores) των συνόλων γονιδίων στην PC1. ...	95
Πίνακας 6. Βαθμοί εμπλουτισμού (Enrichment Scores) της έκφρασης των Tumor δειγμάτων στα σύνολα γονιδίων.	95
Πίνακας 7. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τις εμπλεκόμενες βιολογικές διεργασίες GO από την ανάλυση διαφορικής έκφρασης μεταξύ Tumor και Normal δειγμάτων.	95
Πίνακας 8. Τα 30 κύρια διαφορικά εκφρασμένα γονίδια από τη σύγκριση Tumor-Normal δειγμάτων που εμπλέκονται με διεργασίες GO.	97
Πίνακας 9. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τα εμπλεκόμενα μονοπάτια Reactome από την ανάλυση διαφορικής έκφρασης μεταξύ Tumor και Normal δειγμάτων.	98
Πίνακας 10. Τα 30 κύρια διαφορικά εκφρασμένα γονίδια από τη σύγκριση Tumor-Normal δειγμάτων που εμπλέκονται με μονοπάτια Reactome.	98
Πίνακας 11. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τις εμπλεκόμενες βιολογικές διεργασίες GO από την ανάλυση διαφορικής έκφρασης μεταξύ Poorly και Well δειγμάτων.	99
Πίνακας 12. Τα 30 κύρια διαφορικά εκφρασμένα γονίδια από τη σύγκριση Poorly-Well δειγμάτων που εμπλέκονται με διεργασίες GO.	100
Πίνακας 13. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τα εμπλεκόμενα μονοπάτια Reactome από την ανάλυση διαφορικής έκφρασης μεταξύ Poorly και Well δειγμάτων.	101
Πίνακας 14. Τα διαφορικά εκφρασμένα γονίδια από τη σύγκριση Poorly-Well δειγμάτων που εμπλέκονται με μονοπάτια Reactome.	101
Πίνακας 15. Ανάλυση της άλφα ποικιλομορφίας των Normal και Tumor δειγμάτων, για τους 3 δείκτες: τα παρατηρούμενα OTUs (Observed), Shannon και Simpson.	101
Πίνακας 16. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Tumor vs Normal, από το edgeR.	102
Πίνακας 17. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Tumor vs Normal, από το edgeR.	103
Πίνακας 18. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Well vs Normal, από το edgeR.	104
Πίνακας 19. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Poorly vs Normal, από το edgeR.	105
Πίνακας 20. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Poorly vs Well, από το edgeR.	106
Πίνακας 21. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Well vs Normal, από το edgeR.	107
Πίνακας 22. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Normal, από το edgeR.	108
Πίνακας 23. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Well, από το edgeR.	108

Περιεχόμενα

Περίληψη	1
Abstract	3
Κατάλογος Εικόνων	5
Κατάλογος Πινάκων	7
Πρόλογος και ευχαριστίες	10
1 Θεωρητικό Μέρος	11
1.1 Ηπατοκυτταρικό Καρκίνωμα	11
1.1.1 Κύρια χαρακτηριστικά και αιτίες πρόκλησης	11
1.1.2 Μοριακή Κατηγοριοποίηση HCC κατά Hoshida	16
1.2 Μικροβίωμα	17
1.2.1 Εντερικό Μικροβίωμα	17
1.2.2 Σύσταση εντερικού μικροβιώματος	18
1.2.3 Άξονας Εντέρου-Ήπατος (Gut-Liver Axis)	22
1.2.4 Σχέση εντερικού μικροβιώματος με το HCC	25
1.2.5 Ενδοογκικό μικροβίωμα (Intratumoral Microbiome)	28
1.2.6 Ενδοογκικό μικροβίωμα και HCC	30
1.3 Μεταγραφωμική Ανάλυση RNA-seq Δεδομένων	32
1.3.1 Αλληλούχιση RNA (RNA Sequencing)	32
1.3.2 Μεταγραφωμική (Transcriptomics)	33
1.3.3 Μετα-μεταγραφωμική (Metatranscriptomics)	35
1.4 Θεωρητικό Υπόβαθρο της Υπολογιστικής Ανάλυσης	38
1.4.1 Συλλογή Αρχικών Δεδομένων	38
1.4.2 Ποιοτικός Έλεγχος (Quality Control)	40
1.4.3 Αποκοπή αλληλουχιών χαμηλής ποιότητας	41
1.4.4 Λειτουργικός Χαρακτηρισμός (Functional Annotation)	42
1.4.5 Ανάλυση διαφορικής γονιδιακής έκφρασης (Differential Gene Expression Analysis)	45
1.4.6 Λειτουργικός Χαρακτηρισμός Γονιδιακής Έκφρασης	47
1.4.7 Μετα-μεταγραφωμική ανάλυση – SAMSA2	48
1.4.8 Ανάλυση Μικροβιώματος	50
2 Μεθοδολογία	53
2.1 Λήψη RNA-Seq Δεδομένων	54
2.2 Προεπεξεργασία Αρχικών Δεδομένων	54
2.3 Ευθυγράμμιση στο Ανθρώπινο Γονιδίωμα (Human Genome Alignment)	55
2.4 Μεταγραφωμική Ανάλυση	57
2.4.1 Ποσοτικοποίηση (Quantification)	57
2.4.2 Ανάλυση σε Κύριες Συνιστώσες (PCA)	58
2.4.3 Ανάλυση Διαφορικής Έκφρασης Ανθρώπινων Γονιδίων	59
2.4.4 Ομαδοποίηση Ασθενών βάσει Μοριακών Χαρακτηριστικών	59
2.4.5 Ανάλυση Διαφορικής Έκφρασης Ανθρώπινων Γονιδίων μεταξύ των υπο-ομάδων	62
2.4.6 Λειτουργικός Χαρακτηρισμός Διαφορικά Εκφρασμένων Γονιδίων	62
2.5 Μετα-μεταγραφωμική Ανάλυση	62
2.5.1 Εισαγωγή δεδομένων στο SAMSA2	62
2.5.2 Ανάλυση Μικροβιώματος	63
3 Αποτελέσματα και Συζήτηση Αποτελεσμάτων	65

3.1	Προεπεξεργασία Αρχικών Δεδομένων	65
3.2	Ευθυγράμμιση με το Ανθρώπινο Γονιδίωμα	66
3.3	Ανάλυση Γονιδιακής Έκφρασης στο ανθρώπινο μεταγράφομα	67
3.4	Μοριακή ταξινόμηση καρκινικών δειγμάτων ασθενών με βάση τις υπογραφές γονιδιακών βιοδεικτών.	69
3.5	Διαφορική έκφραση γονιδίων μεταξύ των καρκινικών υπο-ομάδων	72
3.6	Λειτουργικός Χαρακτηρισμός Διαφορικά Εκφρασμένων Γονιδίων	72
3.7	Ανάλυση Μικροβιώματος	74
3.7.1	<i>Χαρακτηρισμός του ενδοσωγικού μικροβιώματος</i>	74
3.7.2	<i>Ανάλυση βιολογικών λειτουργιών του ενδοσωγικού μικροβιώματος</i>	78
3.7.3	<i>Σύγκριση του μικροβιώματος των καρκινικών υπο-ομάδων και εντοπισμός σχετικών μικροβιακών βιοδεικτών των καρκινικών υπο-ομάδων</i>	81
3.7.4	<i>Σύγκριση Βιολογικών Λειτουργιών Μικροβιώματος μεταξύ καρκινικών υπο-ομάδων και ως προς τους φυσιολογικούς ιστούς</i>	83
4	Περιορισμοί της μελέτης	86
5	Συμπεράσματα	88
	Παράρτημα	90
	Συντομογραφίες	109
	Βιβλιογραφία	110

Πρόλογος και ευχαριστίες

Η παρούσα Διπλωματική Εργασία εκπονήθηκε στα πλαίσια της ολοκλήρωσης του κύκλου σπουδών μου στη Σχολή Χημικών Μηχανικών ΕΜΠ, κατά το Ακαδημαϊκό έτος 2022-2023, υπό την επίβλεψη της επίκουρης καθηγήτριας ΕΜΠ Διομής Μαμμά, σε συνεργασία με την Δρ. Ελένη Λουτράρη στην Α' Κλινική Εντατικής Θεραπείας του Νοσοκομείου «Ευαγγελισμός» της Ιατρικής Σχολής ΕΚΠΑ.

Αρχικά, θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτριά μου κ. Μαμμά για την καθοδήγηση και την εμπιστοσύνη που μου έδειξε καθ' όλη τη διάρκεια εκπόνησης της Διπλωματικής Εργασίας. Είμαι ευγνώμων για το πρότυπο που αποτέλεσε σε όλη τη διάρκεια των σπουδών μου, ως καθηγήτρια και ως άνθρωπος.

Επιπλέον, η εργασία αυτή δε θα είχε πραγματοποιηθεί χωρίς τη συμβολή της Δρ. Ελένης Λουτράρη, η οποία με τις γνώσεις, την εμπειρία, και τις συμβουλές της αποτέλεσε τον ακρογωνιαίο λίθο της τελειοποίησης της εργασίας, καθώς και τον ομότιμο καθηγητή της Σχολής Χημικών Μηχανικών, κ. Κολίση, για την διαρκή υποστήριξη και βοήθεια.

Ακόμη, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στους συνεργάτες, Θωμά Γκέκα, Παναγιώτη Αγιουτάντη και Μαριάνθη Λογοθέτη, για την ενθάρρυνση, καθοδήγηση και βοήθεια κατά το έτος συνεργασίας μας.

Μία ιδιαίτερη ευχαριστία αποδίδεται στον καθηγητή μου για το μάθημα Βιοπληροφορικής στο πανεπιστήμιο ΚΤΗ της Στοκχόλμης, Olof Emanuelsson, ο οποίος με ενέπνευσε αρχικά να ασχοληθώ με τον τομέα της Βιοπληροφορικής, μεταδίδοντας το πάθος του για την επιστήμη, και παρέχοντάς μου τις απαραίτητες γνωστικές βάσεις για την εκπόνηση της Διπλωματικής Εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου, τους γονείς και τον αδερφό μου, για την απεριόριστη υποστήριξη, ηθική και υλική, και για την πίστη τους στις επιδιώξεις και τα όνειρά μου, αλλά και τους φίλους και φίλες, οι οποίοι συνέβαλαν στην απόκτηση πολύτιμων αναμνήσεων στα χρόνια των σπουδών μου.

1 Θεωρητικό Μέρος

1.1 Ηπατοκυτταρικό Καρκίνωμα

1.1.1 Κύρια χαρακτηριστικά και αιτίες πρόκλησης

Ο καρκίνος του ήπατος αποτελεί τον πέμπτο πιο συνήθη τύπο καρκίνου παγκοσμίως και κατέχει την τρίτη θέση ως συχνότερη αιτία θνησιμότητας από καρκίνο, ενώ οι περιπτώσεις του αυξάνονται παγκοσμίως (El-Serag & Rudolph, 2007). Το ηπατοκυτταρικό καρκίνωμα (HCC) είναι η πιο κοινή μορφή ηπατικού καρκίνου, καλύπτοντας, κατά προσέγγιση, το 90% των περιπτώσεων, ενώ οι υπόλοιπες περιπτώσεις αποδίδονται στο χολαγγειοκαρκίνωμα και το ηπατικό αγγειοσάρκωμα (Zhu et al., 2016). Το 50% των περιπτώσεων HCC σχετίζονται με τον ιό της ηπατίτιδας Β (HBV), ενώ οι περιπτώσεις που συνδέονται με τον ιό της ηπατίτιδας C (HCV) προσφάτως έχουν μειωθεί (Llovet et al., 2016). Αποτελεί ένα επιθετικό είδος καρκίνου με σταθερό ρυθμό θνησιμότητας, δίχως να απαντώνται αποτελεσματικές θεραπευτικές επιλογές (Suresh et al., 2020).

Το HCC χαρακτηρίζεται από ποικίλα ενδιαφέροντα επιδημιολογικά στοιχεία, όπως διαφοροποιήσεις ανάλογα τη γεωγραφική περιοχή, την εθνικότητα, το βιολογικό φύλο και την παρουσία πολλών περιβαλλοντικών, δυνητικά αποτρέψιμων κινδύνων. Επίσης, η ταχέως αναπτυσσόμενη κατανόηση των μοριακών μηχανισμών του καρκίνου, υποδεικνύουν πως η πρόκληση της καρκινογένεσης δεν απαντάται σχεδόν ποτέ σε υγιή ήπατα, αλλά ο κίνδυνος αυξάνεται ραγδαία σε συνθήκες χρόνιας ηπατικής φλεγμονής, κυρίως παρουσία κίρρωσης (El-Serag & Rudolph, 2007). Η εμφάνιση του HCC είναι υψηλότερη στις μέσες ηλικίες, μεταξύ 55 και 59 ετών στην Κίνα, και 63-65 ετών στη Βόρεια Αμερική και την Ευρώπη. Υψηλότερα ποσοστά HCC παρατηρούνται στην Ανατολική και Νοτιοανατολική Ασία, και την Αφρική (El-Serag & Rudolph, 2007; Jemal et al., 2011). Τα τελευταία στοιχεία που εκτιμήθηκαν από τη GLOBOCAN έδειξαν περίπου 900.000 νέες περιπτώσεις και 830.000 θανάτους από καρκίνο του ήπατος, το 2020, παγκοσμίως (+ref).

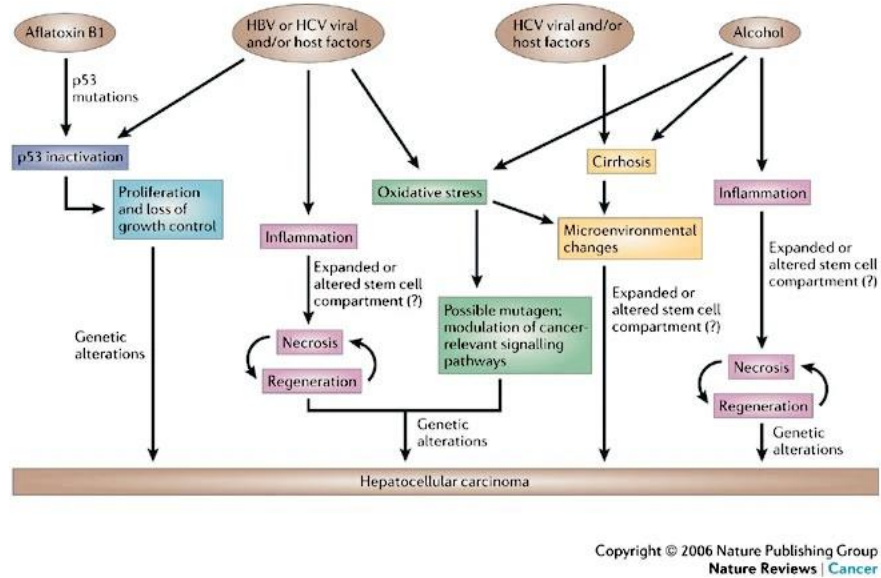
Πάνω από το 90% των περιπτώσεων HCC συνδέεται με χρόνια ηπατική φλεγμονή, καθώς και κίρρωση του ήπατος, από οποιαδήποτε αιτιολογία (Balogh et al., 2016). Οι κύριοι αιτιολογικοί παράγοντες του HCC είναι η χρόνια κατανάλωση αλκοόλ, ο διαβήτης, η μη αλκοολική στεατοηπατίτιδα (NASH) και η μόλυνση από HBV ή HCV (Llovet et al., 2016).

Τα τελευταία χρόνια, οι λιπώδεις νόσοι του ήπατος παρουσιάζουν άνοδο ως αιτίες ανάπτυξης καρκίνου στο ήπαρ. Η άνοδος αυτή οφείλεται στην εξέλιξη των αντικών θεραπειών, που παρεμποδίζουν την ανάπτυξη ιογενούς αιτιολογίας HCC. Η μη αλκοολική λιπώδης νόσος του ήπατος (NAFLD) χαρακτηρίζεται από την υπερβολική συσσώρευση ηπατικών λιπιδίων, που οδηγεί σε στεατοηπατίτιδα, και, ύστερα, σε κίρρωτικό στάδιο HCC (Piroola et al., 2022), και συσχετίζεται με μεταβολικές νόσους, όπως παχυσαρκία, υπέρταση, ινσουλινοαντίσταση και διαβήτη τύπου 2 (Suresh et al., 2020). Η NASH, φλεγμονή του ήπατος από συσσώρευση λίπους στα ηπατοκύτταρα, αποτελεί μία από τις βασικότερες αιτιολογίες του HCC στη Δύση, δίχως κατανάλωση αλκοόλ (Llovet et al., 2016). Η παθογένεση της NASH μπορεί να περιγραφεί με το συνδυασμό δύο παραγόντων. Ο κύριος παράγοντας είναι η συσσώρευση λιπιδίων με αλλαγές στην ομοιόστασή τους, που σχετίζονται με την παχυσαρκία, όπως την αντίσταση στην ινσουλίνη και τις ανωμαλίες της αδιποκίνης, ενώ ο δεύτερος παράγοντας είναι ένας συνδυασμός αυξημένου φορτίου του οξειδωτικού στρες, μιτοχονδριακής δυσλειτουργίας, τοξικότητας των χολικών οξέων και λοιπών λειτουργιών (Milosevic et al., 2019).

Η αλκοολική λιπώδης νόσος του ήπατος (AFLD) αποδίδεται αποκλειστικά στην υπερβολική κατανάλωση αλκοόλ, επιφέροντας συσσώρευση λίπους, φλεγμονή και τραυματισμό του ηπατικού ιστού, οδηγώντας, τελικά, σε κακοήθεια (Suresh et al., 2020). Ανεξάρτητα από τις διαφορές μεταξύ των δύο ειδών λιπώδους ηπατικής νόσου ο μηχανισμός καρκινογένεσης είναι ο ίδιος. Η ακεταλδεΰδη, προϊόν οξείδωσης της αιθανόλης, είναι πιθανός καρκινογόνος συντελεστής. Παρόλο που η μείζων μεταβολική οδός της αιθανόλης πραγματοποιείται από το ένζυμο CYP2E1 στα υπεροξυσώματα, η ακεταλδεΰδη και δραστικές μορφές οξυγόνου (ROS) εξακολουθούν να παράγονται, προκαλώντας μη αναστρέψιμες βλάβες στο DNA (Ganne-Carrié & Nahon, 2019).

Οι πιο διαδεδομένες περιπτώσεις χρόνιας ιογενούς ηπατίτιδας είναι οι HBV και HCV. Ο ιός της HBV είναι ένα δίκλωνο κυκλικό μόριο DNA με 8 διαφορετικούς γενοτύπους (από A ως H). Η μετάδοσή του προκύπτει, κυρίως, από μεταγγίσεις μολυσμένου αίματος, ενδοφλέβιες ενέσεις, σεξουαλική επαφή, καθώς και από την κάθετη μετάδοση, από τη μητέρα στο έμβρυο (Ott et al., 2012). Η ηπατοκαρκινογένεση από τον HBV μπορεί να ελαττωθεί με τη χορήγηση αντικής θεραπείας, σε ποσοστό ως και 10% για κίρρωτικές περιπτώσεις ήπατος. Ακόμη, με

χορήγηση εμβολίου κατά του HBV έχει παρουσιαστεί μείωση περιπτώσεων HCC προερχόμενων από HBV (Hosaka et al., 2013).



Εικόνα 1. Οι μοριακοί μηχανισμοί που συνδέονται με την ηπατοκαρκινογένεση.

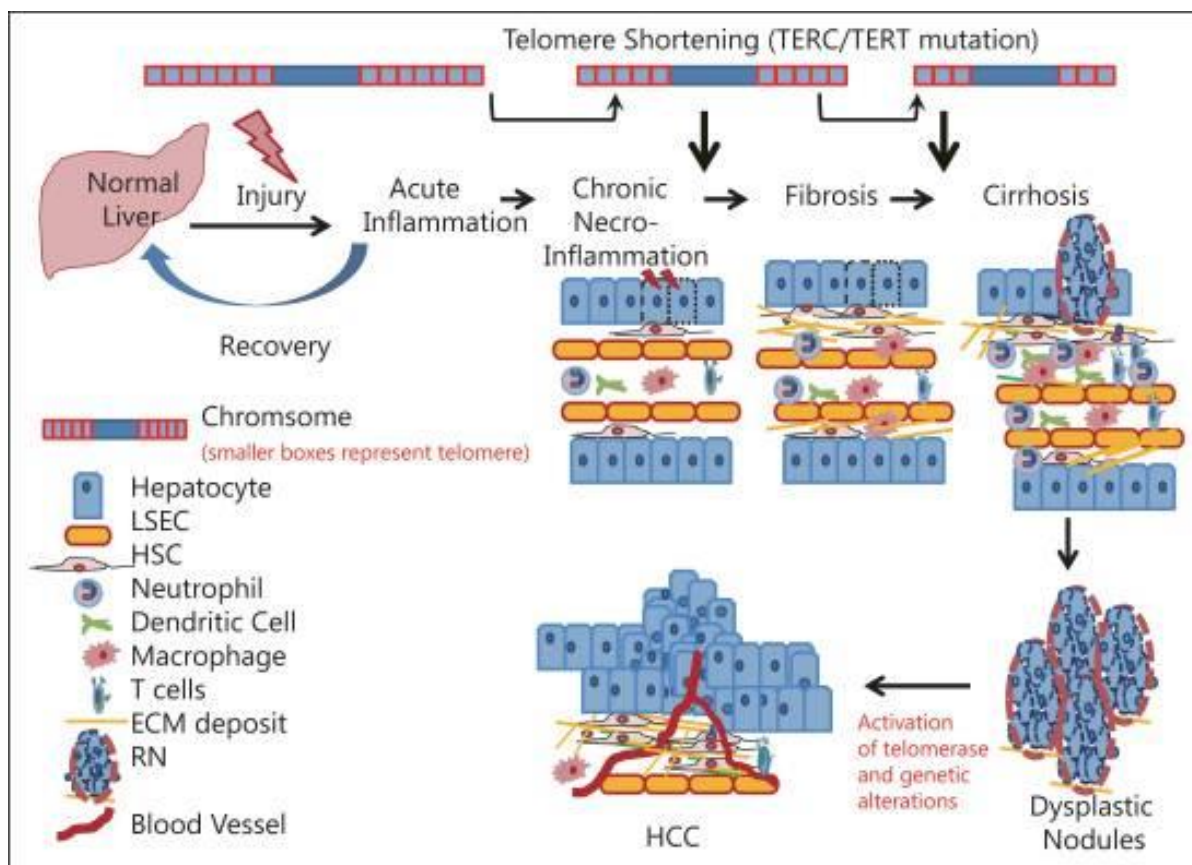
Η αλληλεπίδραση της μοριακής δομής του HBV με το ανθρώπινο γονιδίωμα, όπως και η εκτεταμένη έκθεση στην αφλατοξίνη Β, προκαλεί την απενεργοποίηση της πρωτεΐνης p53, η οποία αποτελεί ένα μεταγραφικό παράγοντα, καταστολέα της καρκινογένεσης (Herrero et al., 2016). Ακόμη, η ενεργοποίηση του μονοπατιού Wnt/β-catenin, κατά την ηπατοκαρκινογένεση, επηρεάζει πολλές σχετιζόμενες λειτουργίες, όπως την ομοιόσταση, τον κυτταρικό πολλαπλασιασμό και την απόπτωση. Το μονοπάτι αυτό ενεργοποιείται συνήθως από μεταλλάξεις της β-catenin, ακόμα και σε φυσιολογικούς ηπατικούς ιστούς (Khalaf et al., 2018). Επομένως, η απορρύθμιση της p53, η ενεργοποίηση του μονοπατιού Wnt/β-catenin, αλλά και η αναστολή του παράγοντα μετασχηματισμού των μεσεγχυματικών κυττάρων, TGF-β (Neuveut et al., 2010), οδηγούν στον πολλαπλασιασμό των κυττάρων, και, επομένως, στην ανάπτυξη κακοήθειας.

Ένας όγκος που χαρακτηρίζεται καλώς διαφοροποιημένος διαθέτει κύτταρα τα οποία μοιράζονται περισσότερα κοινά χαρακτηριστικά με τα αντίστοιχα φυσιολογικά ηπατικά κύτταρα. Αντίθετα, ένας όγκος χαμηλής διαφοροποίησης, διαθέτει καρκινικά κύτταρα μεγαλύτερα σε μέγεθος και σημαντικά διαφορετικά ως προς τις μορφολογικές τους ιδιότητες (K. Zhang et al., 2020).

Η διαδοχική εξέλιξη της χρόνιας ηπατικής νόσου σε ηπατική ίνωση (fibrosis) και τελικά σε κίρρωση, οδηγεί στη νεοπλασία, μέσω της δημιουργίας δυσπλαστικών οζιδίων. Η ενεργοποίηση της τελομεράσης σε συνδυασμό με ποικίλες γενετικές τροποποιήσεις προξενούν την καρκινογένεση, όπως παρουσιάζεται στην Εικόνα 2 (Ramakrishna et al., 2013).

Ο πολλαπλασιασμός των κυττάρων στο ήπαρ αυξάνεται υπό συνθήκες χρόνιας ηπατίτιδας. Ωστόσο, η κίρρωση χαρακτηρίζεται από την καταστολή του πολλαπλασιασμού των ηπατικών κυττάρων, υποδηλώνοντας πως η αναγεννητική ικανότητα του ήπατος εξασθενεί κατά το στάδιο της κίρρωσης (Neuveut et al., 2010). Επίσης, η βράχυνση των τελομερών περιορίζει την αναγεννητική ικανότητα των πρωτογενών ανθρώπινων κυττάρων, συμπεριλαμβανομένων των ηπατικών, οδηγώντας, τελικά, σε φθορές του DNA, διακοπή του κυτταρικού κύκλου, γήρανση ή απόπτωση (Fagagna et al., 2003). Η βράχυνση των τελομερών, σε συνδυασμό με τη γήρανση των ηπατοκυττάρων, συντελούν στην απώλεια αντιγραφικού αποθέματος, σε κίρρωτικό στάδιο (El-Serag & Rudolph, 2007).

Η πρωτεΐνη p21, κύριος ρυθμιστής του κυτταρικού πολλαπλασιασμού, λειτουργεί ως αναστολέας της κινάσης κυκλίνης και, κατά συνέπεια, του κυτταρικού κύκλου (Abbas & Dutta, 2009). Η πρωτεΐνη p53 έχει ογκοκατασταλτική δράση, δηλαδή ρυθμίζει την κυτταρική διαίρεση εμποδίζοντας τα κύτταρα να αναπτυχθούν και να διαιρεθούν πολύ γρήγορα ή με ανεξέλεγκτο τρόπο (L. Zhao & Sanyal, 2022). Η p53 συμμετέχει στην κύρια οδός καταστολής όγκου που περιορίζει την επιβίωση και τον πολλαπλασιασμό των κυττάρων, προκαλεί τη διακοπή του κυτταρικού κύκλου, ως απόκριση στην ενεργοποίηση ογκογονιδίων, και προστατεύει την ακεραιότητα του γονιδιώματος (di Micco et al., 2006).



Εικόνα 2. Διαγραμματική απεικόνιση του μηχανισμού παθολογικών μεταβολών καρκινογένεσης: από τη φλεγμονή υγιούς ήπατος στο HCC. Κατά τη διάρκεια της εξέλιξης ενός καλοήθους αναγεννητικού οζιδίου (benign regenerating nodule, BRN) σε δυσπλαστικό οζίδιο (dysplastic nodule, DN) και τελικά σε HCC, η παθολογική αιμάτωση του DN και του HCC γίνεται από την ηπατική αρτηρία (Ramakrishna et al., 2013).

Η πρωτεΐνη p21 καταστέλλει την λειτουργία της p53, και κατά συνέπεια την ανάπτυξη καρκίνου. Ωστόσο, σε ορισμένες κυτταρικές συνθήκες, η συγκεκριμένη πρωτεΐνη μπορεί να δράσει ως καρκινογόνος. Συνεπώς, σύμφωνα με την υπόθεση ότι ο εξασθενημένος πολλαπλασιασμός των ηπατικών κυττάρων που προκαλείται από τη βράχυνση των τελομερών θα μπορούσε να επιταχύνει την ανάπτυξη καρκίνου, έχει παρατηρηθεί συσχέτιση μεταξύ της αυξημένης έκφρασης της p21 και του κινδύνου HCC στην ανθρώπινη κίρρωση (C. Zhang et al., 2010).

Το HCC είναι ένας όγκος που εξαρτάται σε μεγάλο βαθμό από την αγγειογένεση, καθώς οι στέρεοι όγκοι προϋποθέτουν παροχή αίματος, προκειμένου να μπορούν να αναπτυχθούν για περισσότερο από ορισμένα χιλιοστά. Κατά την πρόοδο από καλώς έως μέτρια διαφοροποιημένο στάδιο, η αγγειογένεση επιτρέπει την εισβολή των κακοηθών κυττάρων στα αγγεία και τη μετάστασή τους. Τόσο η αγγειογένεση όσο και ο κυτταρικός πολλαπλασιασμός εμπλέκονται στην ανάπτυξη του HCC. Οι όγκοι μπορούν να προκαλέσουν τη δημιουργία αγγείων, εκκρίνοντας χημικούς σηματοδότες ενεργοποίησης της αγγειογένεσης (Nishida et al.,

2006). Οι σηματοδότες αυτοί, σε νεοπλασματικούς ιστούς, ενεργοποιούν πρωτεΐνες, όπως τον αγγειακό ενδοθηλιακό αυξητικό παράγοντα (VEGF), την αγγειοποιητίνη, το TGF- α και πολλές ακόμη, οι οποίες ονομάζονται αγγειογενετικοί ενεργοποιητές (Rajabi & Mousa, 2017). Ο πιο ισχυρός και κρίσιμος παράγοντας για την προώθηση της ανάπτυξης των αγγείων και της εξέλιξης του όγκου είναι ο VEGF. Τα υψηλά επίπεδα του VEGF και των υποδοχέων του ανιχνεύονται κυρίως σε περιπτώσεις χρόνιας ηπατικής φλεγμονής και σε κίρρωτικό ήπαρ, και συνδέονται με υψηλή επιθετικότητα (Alqahtani et al., 2019). Έτσι, η αυξημένη αγγειογένεση, σε συνδυασμό με την αναπτυσσόμενη υποξία και τη συμβολή της κίρρωσης του ήπατος, διαμορφώνουν ένα περιβάλλον που πληροί τις προϋποθέσεις ηπατοκαρκινογένεσης και μετάστασης (Li et al., 2022).

1.1.2 Μοριακή Κατηγοριοποίηση HCC κατά Hoshida

Ένα σύστημα ταξινόμησης HCC με βάση τα προφίλ γονιδιακής έκφρασης που είναι εξαιρετικά αναπαραγωγίσιμο μεταξύ συνόλων κλινικών δειγμάτων, διαιρεί το HCC σε τρεις κύριες υποκατηγορίες που ονομάζονται S1, S2 και S3. Οι όγκοι υποκατηγορίας S1 αποτελούν το 28-31% και οι όγκοι S2 αποτελούν το 23-24% του HCC στα σύνολα των κλινικών δεδομένων που εξετάστηκαν (Hoshida et al., 2009). Και οι δύο υποκατηγορίες όγκων, S1 και S2, σχετίζονται με χειρότερη πρόγνωση επιβίωσης σε σχέση με την υποκατηγορία S3. Οι όγκοι S1 είναι πιο διεισδυτικοί τοπικά, ενώ οι όγκοι S2 εκφράζουν τις E-cadherin, Glypican-3 και c-myc και εκκρίνουν άλφα φετοπρωτεΐνη (AFP). Αυτό το σύστημα ταξινόμησης έχει ταυτοποιήσει κυτταρικές σειρές ως S1 και S2 HCC, αλλά ποτέ δεν έχει ταυτοποιήσει μια κυτταρική σειρά S3, υποδηλώνοντας ότι κύτταρα αυτού του τύπου δεν είναι σε θέση να επιβιώσουν *in vitro*. Χαρακτηριστικά αυτών των συστημάτων μοριακής κατηγοριοποίησης έχουν προβλέψει την ανταπόκριση του HCC σε αναδυόμενες στοχευμένες θεραπείες σε προκλινικές μελέτες, συμπεριλαμβανομένων των αναστολέων του υποδοχέα του επιδερμικού αυξητικού παράγοντα (EGFR), του υποδοχέα του αυξητικού παράγοντα 1 που μοιάζει με ινσουλίνη (IGF-1R) και των Src/Abl (Schmidt et al., 2016).

Πίνακας 1. Κατηγορίες HCC κατά Hoshida και τα χαρακτηριστικά τους.

Υποκατηγορία	S1	S2	S3
Μεταλλάξεις DNA	TP53		CTNNB1
Μοριακά χαρακτηριστικά	TGF-β, ILs, IFN ↑	SHH ↑, IFN ↓	Μεταβολισμός ξενοβιοτικών, χολικού οξέος και λιπιδίων οξέων, λιπογένεση
Ιστολογικός φαινότυπος	Λιγότερο διαφοροποιημένα κύτταρα		Περισσότερο διαφοροποιημένα κύτταρα
Δείκτες		AFP, GPC3	
Κλινική επίδραση		Υψηλή επιθετικότητα	

Ανάμεσα στους 3 υποτύπους HCC, παρατηρήθηκε πως, μορφολογικά, οι όγκοι S2 είναι μεγαλύτεροι συγκριτικά με τους υπόλοιπους, ενώ οι S3 μικρότεροι, προσομοιάζοντας φυσιολογικούς ιστούς. Όπως επιβεβαιώνεται και στον Πίνακα 1, οι όγκοι S3 περιλαμβάνουν τους καλώς διαφοροποιημένους (Well differentiated) καρκινικούς ιστούς, ενώ βρέθηκε επίσης πως δεν υπάρχουν σημαντικές ιστολογικές διαφορές ανάμεσα στους υποτύπους S1 και S2, χαρακτηρίζοντας κακώς διαφοροποιημένους (Poorly differentiated) καρκινικούς ιστούς. Ακόμη, ανιχνεύθηκε σημαντικά υψηλότερη πρώιμη υποτροπή στους όγκους S1 (Hoshida et al., 2009), συμπεριφορά που σχετίζεται με τη διάδοση των πρωτοπαθών καρκινικών κυττάρων εντός του ήπατος, μέσω της εισβολής αυτών στα αγγεία (El-Domiaty et al., 2021).

1.2 Μικροβίωμα

1.2.1 Εντερικό Μικροβίωμα

Το μικροβίωμα αποτελεί ένα οικοσύστημα συμβιωτικών και παθογόνων μικροοργανισμών στο ανθρώπινο σώμα. Αυτό το σύνολο μικροοργανισμών περιλαμβάνει πρωτόζωα, μύκητες, βακτήρια και ιούς, οι οποίοι διαμορφώνουν ειδικές ανά περιοχή μικροβιακές κοινότητες (Maddi et al., 2019). Απαντάται σε περιοχές του σώματος οι οποίες αλληλοεπιδρούν με το περιβάλλον, συμπεριλαμβάνοντας το δέρμα, την αεροπεπτική οδό, το γαστρεντερικό σωλήνα και την κοιλική κοιλότητα. Η ύπαρξη διακριτού μικροβιώματος σε κάθε οργανισμό αποτελεί καταλυτικό παράγοντα στην ομοίωση και το ανοσοποιητικό σύστημα (Rattan et al., 2020). Αποτελεί ένα σύστημα με συμβιωτική και δυναμική σχέση με τον ξενιστή, με πολλούς από τους μικροοργανισμούς να ευθύνονται για θεμελιώδεις λειτουργίες που, κυρίως, αφορούν τη ρύθμιση μεταβολικών διεργασιών και αποκρίσεις του ανοσοποιητικού συστήματος. Οι αλλαγές στο μικροβίωμα σε ένα όργανο μπορούν να

επιφέρουν αλλοιώσεις της τοπικής ομοιόστασης και φλεγμονές, προκαλώντας καταστροφή ιστών και απορρύθμιση των ανοσοποιητικών αποκρίσεων (Sepich-Poore et al., 2021).

Η εντερική μικροχλωρίδα περιέχει δεκάδες τρισεκατομμύρια μικροοργανισμούς, που περιλαμβάνουν τουλάχιστον 1.000 διαφορετικά είδη (species) γνωστών βακτηρίων, η συντριπτική πλειοψηφία των οποίων ανήκει στα φύλα *Firmicutes* και *Bacteroidetes* (Ley et al., 2008). Η μικροχλωρίδα συμβάλλει στην εξαγωγή ενέργειας από τα τρόφιμα και στη σύνθεση βιταμινών και αμινοξέων και βοηθά στη δημιουργία φραγμών κατά των παθογόνων. Η διαταραχή της ομοιόστασης του μικροβιώματος - που ονομάζεται δυσβίωση - έχει συσχετιστεί με τη φλεγμονώδη νόσο του εντέρου, το σύνδρομο ευερέθιστου εντέρου, την κοιλιοκάκη νόσο, τροφικές αλλεργίες, διαβήτη τύπου 1 και 2, καρκίνο, παχυσαρκία και καρδιαγγειακή νόσο (Engen et al., 2015).

Η σχέση του μικροβιώματος με τον ξενιστή είναι σημαντική για την ορθή λειτουργία του ανοσοποιητικού συστήματος. Η ανάπτυξη των ανοσοκυττάρων, η παραγωγή αντισωμάτων, αντιμικροβιακών πεπτιδίων και η λειτουργία αρκετών συντελεστών της άμυνας του ξενιστή ρυθμίζονται από τη μικροχλωρίδα (Kamada et al., 2013). Μια κατηγορία μορίων που δρα ως σύνδεσμος μεταξύ της μικροχλωρίδας και του ανοσοποιητικού συστήματος του ξενιστή είναι τα λιπαρά οξέα βραχείας αλυσίδας (SCFAs). Πρόκειται για προϊόντα του μεταβολισμού των βακτηρίων, τα οποία αποτελούνται χημικά από ένα τμήμα καρβοξυλικού οξέος και μια μικρή αλυσίδα υδρογονάνθρακα. Τα πιο κοινά και πιο μελετημένα από αυτά είναι το οξικό, το προπιονικό και το βουτυρικό οξύ, τα οποία έχουν δύο, τρεις και τέσσερις άνθρακες στη χημική τους δομή, αντίστοιχα. Τα SCFAs χρησιμοποιούνται εν μέρει ως πηγή ATP από αυτά τα κύτταρα. Επιπλέον, δρουν ως ρυθμιστές διαφορετικών πτυχών της ανάπτυξης, της επιβίωσης και της λειτουργίας των επιθηλιακών κυττάρων του εντέρου και των λευκοκυττάρων, συντηρώντας, κατά συνέπεια, την εντερική ομοιόσταση (Corrêa-Oliveira et al., 2016).

1.2.2 Σύσταση εντερικού μικροβιώματος

Το μικροβίωμα στο έντερο αποτελείται από ορισμένους κυρίαρχους μικροοργανισμούς, που υπερσχύουν με μικρές διακυμάνσεις σε κάθε υγιή ξενιστή. Τα βακτηριακά φύλα *Firmicutes* και *Bacteroidetes* αντιπροσωπεύουν το 90% του εντερικού μικροβιώματος. Συγκεκριμένα, το πρώτο φύλο συγκροτείται από περισσότερα από 200 διαφορετικά γένη, εκ των οποίων το γένος *Clostridium* αντιπροσωπεύει το 95% αυτού στο μικροβίωμα. Το φύλο

Bacteroidetes αντιπροσωπεύεται από τα κυρίαρχα γένη *Bacteroides* και *Prevotella* (Rinninella et al., 2019). Ειδικότερα, το γένος *Prevotella* επηρεάζεται από διατροφή πλούσια σε φυτικές ίνες, πλούσιες σε υδατάνθρακες, με αποτέλεσμα να χρησιμοποιείται ως βιοδείκτης για την ύπαρξη ασθενειών, όπως παχυσαρκία, καρκίνος παχέος εντέρου, μεταβολικό σύνδρομο και διαβήτη τύπου 1 (Precup & Vodnar, 2019).

Η μικροβιακή ποικιλομορφία εντός ενός περιβάλλοντος στον ανθρώπινο οργανισμό, δηλαδή ο αριθμός των διακριτών τύπων μικροοργανισμών, μπορεί να συνδεθεί με σοβαρές νόσους: π.χ. η ελαττωμένη ποικιλομορφία στο έντερο σχετίζεται με την παχυσαρκία και με φλεγμονώδους νόσους του εντέρου (Turnbaugh et al., 2009), ενώ η αυξημένη στην περιοχή του κόλπου, με βακτηριακή κολπίτιδα (Fredricks et al., 2005).

Συνεχείς μελέτες του ανθρώπινου μικροβιώματος επέδειξαν ότι ακόμη και τα υγιή άτομα διαφοροποιούνται αξιολογούμενα ως προς το μικροβίωμά τους σε διάφορες θέσεις του οργανισμού έντερο, δέρμα και κόλπος. Μεγάλο μέρος της εν λόγω διαφοροποίησης παραμένει ανερμήνευτο, ωστόσο είναι γνωστό πως η διατροφή, το περιβάλλον, γενετικά χαρακτηριστικά και η πρώιμη έκθεση σε μικρόβια αποτελούν παράγοντες που διαμορφώνουν την ατομική μικροβιακή κοινότητα.

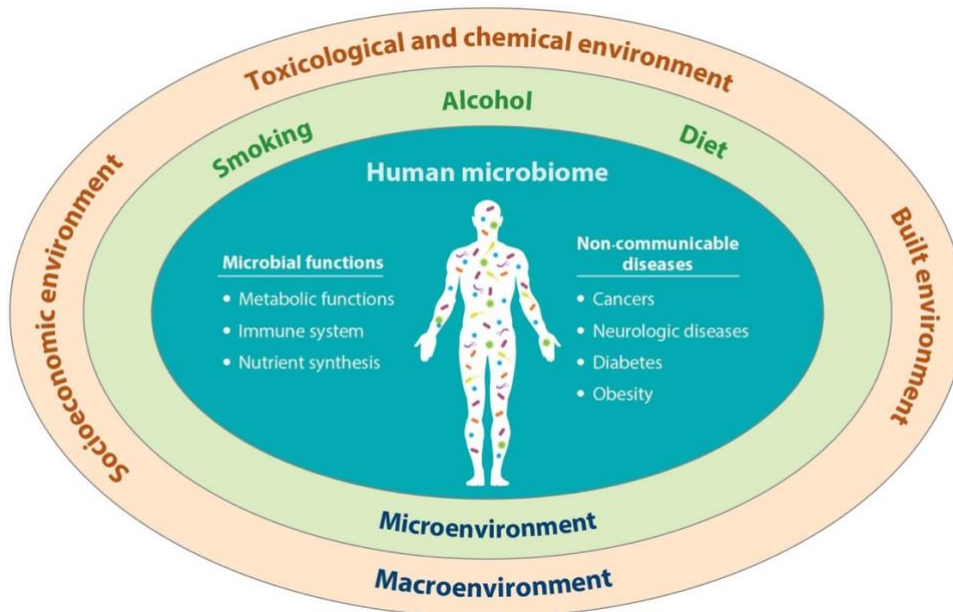
Οι διαφορετικές διατροφικές συνήθειες ανά γεωγραφική περιοχή έδειξαν πως επηρεάζουν άμεσα το μικροβίωμα του εντέρου. Παραδείγματος χάρη, η κατανάλωση υδατάνθρακα μπορεί να προσφέρει πληροφορίες για τη διαφορετικότητα του μικροβιώματος ανάμεσα σε δείγματα της Μπουρκίνα Φάσο και της Ιταλίας, παρουσιάζοντας στον πληθυσμό της πρώτης υπεροχή των *Bacteroidetes* και συγκεκριμένα του γένους *Prevotella*, σε σχέση με τον πληθυσμό της τελευταίας (de Filippo et al., 2010). Γενικότερα, ο τρόπος ζωής στο δυτικό κόσμο – από την κατανάλωση νερού και αντιβιοτικών μέχρι τις σύγχρονες συνθήκες εργασίας και το άγχος - φαίνεται να οδηγεί στην ελάττωση του μικροβιακού πλούτου και της βιοποικιλομορφίας, και, κατά συνέπεια, στην εμφάνιση ασθενειών που συνδέονται με αυτή (Mosca et al., 2016).

Η δυσβίωση μπορεί να προκληθεί από περιβαλλοντικούς παράγοντες που συναντώνται συνήθως στις δυτικές κοινωνίες, όπως η διατροφή, το κάπνισμα και η κατανάλωση αλκοολούχων ποτών. Είναι καλά τεκμηριωμένο ότι η διατροφή επηρεάζει τη σύνθεση και την ποικιλομορφία της εντερικής μικροχλωρίδας. Οι δίαιτες πλούσιες σε λίπος όπως κ οι «δυτικού

τύπου» δίαιτες, που περιλαμβάνουν υψηλή περιεκτικότητα σε λιπαρά και ζάχαρη αλλάζουν τη μικροχλωρίδα του εντέρου, (Engen et al., 2015) και συμβάλλουν στην ανάπτυξη παχυσαρκίας και ηπατικής βλάβης, καθώς και σύνδρομο ευερέθιστου εντέρου, κοιλιοκάκη νόσο, διαβήτη τύπου 1 και τύπου 2, τροφικές αλλεργίες και καρδιαγγειακές παθήσεις, τουλάχιστον σε γενετικά ευαίσθητα άτομα (Brown et al., 2012).

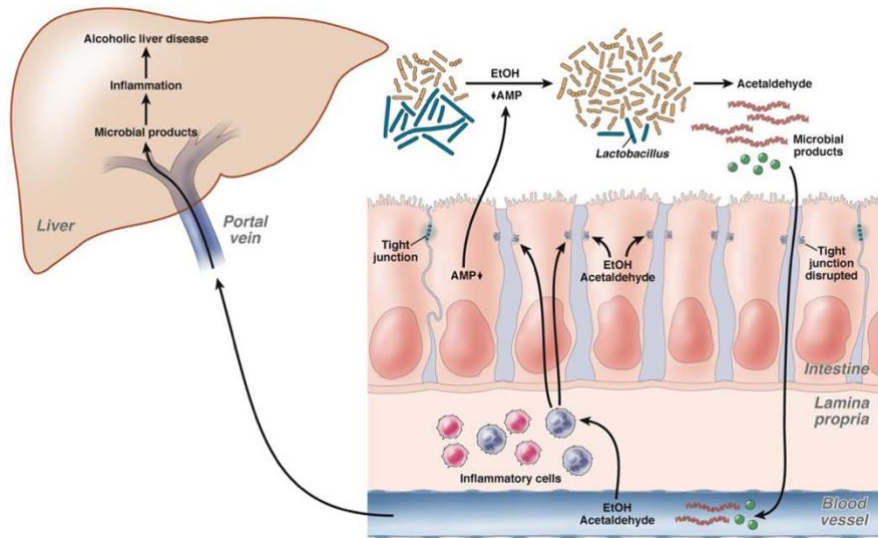
Έχειδειχτεί ότι δίαιτες υψηλής περιεκτικότητας σε λιπαρά οδηγούν σε δυσβίωση και υπερανάπτυξη βακτηρίων του εντέρου (Εικόνα 3). Οι αλλαγές στην εντερική μικροχλωρίδα αυξάνουν την εξαγωγή ενέργειας και το μεταβολισμό των διαιτητικών ινών σε ολιγοσακχαρίτες, μονοσακχαρίτες και λιπαρά οξέα βραχείας αλυσίδας (SCFA), αντίστοιχα. Η διατροφική χολίνη μεταβολίζεται από την εντερική μικροχλωρίδα, με αποτέλεσμα την ανεπάρκεια χολίνης. Η ανεπάρκεια ηπατικής χολίνης προκαλεί ηπατική στεάτωση, και οι αλλαγές στη μικροχλωρίδα παράγουν επίσης αιθανόλη (EtOH), η οποία απορροφάται και μεταβολίζεται στο ήπαρ. Η δυσβίωση, με τη σειρά της, προκαλεί διαταραχή των στενών συνδέσεων στα εντεροκύτταρα. Η αυξημένη εντερική διαπερατότητα οδηγεί σε μετατόπιση μικροβιακών προϊόντων στο ήπαρ και προκαλεί φλεγμονή (Schnabl & Brenner, 2014).

Η παχυσαρκία σχετίζεται με δυσβιοτική μικροχλωρίδα του εντέρου, με μειωμένη ποικιλομορφία και αυξημένη αναλογία *Firmicutes/Bacteroidetes* (Chakraborti, 2015). Το περιβάλλον και οι συνήθειες του ατόμου αποτελούν παράγοντες που επιδρούν κατά 20% στη διακύμανση της μικροβιακής ποικιλομορφίας. Η ατμοσφαιρική ρύπανση μειώνει την σχετική αφθονία των φύλων *Firmicutes* και *Bacteroidetes*, προκαλώντας δυσβίωση, ρήξη του εντερικού φραγμού, με κυριότερες συνέπειες την παχυσαρκία και το διαβήτη τύπου 2 (Bailey et al., 2020).



Εικόνα 3. Η επίδραση του περιβάλλοντος αποτελεί ένα σύνθετο δίκτυο αλληλεπιδράσεων ανάμεσα σε ποικίλους παράγοντες. Οι περιβαλλοντικές εκθέσεις έχουν άμεσο αντίκτυπο στο μικροβίωμα και στην εμφάνιση ασθενειών (Ahn & Hayes, 2021).

Το κάπνισμα εκθέτει τον οργανισμό σε ένα σύνθετο χημικό μείγμα ουσιών, συμπεριλαμβανομένων της νικοτίνης, αλδεϋδών, πολυκυκλικών αρωματικών υδρογονανθράκων, νιτροζαμινών κ.ο.κ., που διέρχονται στους πνεύμονες ως αερολύματα ή σε ελεύθερη μορφή (Ghio et al., 2008). Τα παραπάνω τοξικά συστατικά ελαττώνουν ενδογενής αντιοξειδωτικούς παράγοντες του οργανισμού, καθώς αυξάνουν την υπεροξειδωση των λιπιδίων, τα επίπεδα του οξειδωτικού στρες και τις συγκεντρώσεις προφλεγμονωδών παραγόντων στο αίμα (van der Vaart et al., 2004). Επιπλέον, οι τοξικές ουσίες του καπνού του τσιγάρου που διέρχονται στη γαστρεντερική οδό προκαλούν δυσβίωση του εντερικού μικροβιώματος μέσω διαφορετικών μηχανισμών, όπως η αντιμικροβιακή δράση (Berkowitz et al., 2019; Gui et al., 2021). Δεδομένου ότι το μικροβίωμα έχει πρόσφατα μελετηθεί ότι σχετίζεται με τη λειτουργία του ανοσοποιητικού συστήματος του ξενιστή, το κάπνισμα, το οποίο μπορεί να επηρεάσει τη λειτουργία του ανοσοποιητικού, μπορεί να επηρεάσει έμμεσα τη σύνθεση του μικροβιώματος του εντέρου (Corrêa-Oliveira et al., 2016; Soporì, 2002).



Εικόνα 4. Η επίδραση του αλκοόλ στο ήπαρ. Η δημιουργία διαρρέοντος εντέρου λόγω της επιβάρυνσης του οξειδωτικού στρες, οδηγεί σε ηπατική φλεγμονή και, κατά συνέπεια, σε αλκοολική νόσο του ήπατος. Συντμήσεις: EtOH: αιθυλική αλκοόλη, AMP: μονοφωσφορική αδενοσίνη

Το αλκοόλ είναι ένας ακόμη διατροφικός διαταράκτης του εντερικού μικροβιώματος, καθώς έχει διαπιστωθεί πως με την κατανάλωση αλκοόλ μειώνεται η αφθονία του φύλου *Bacteroidetes* και αυξάνεται εκείνη του φύλου *Proteobacteria*. Άλλες μελέτες δείχνουν ότι η δυσβιοτική μικροχλωρίδα στους αλκοολικούς συσχετίζεται, επίσης, με υψηλό επίπεδο ενδοτοξίνης στο αίμα, υποδεικνύοντας ότι η δυσβίωση μπορεί να συμβάλει στην εντερική διαπερατότητα ή στην αυξημένη μετατόπιση Gram-αρνητικών μικροβιακών προϊόντων από τον εντερικό αυλό στο κυκλοφορικό σύστημα (Engen et al., 2015).

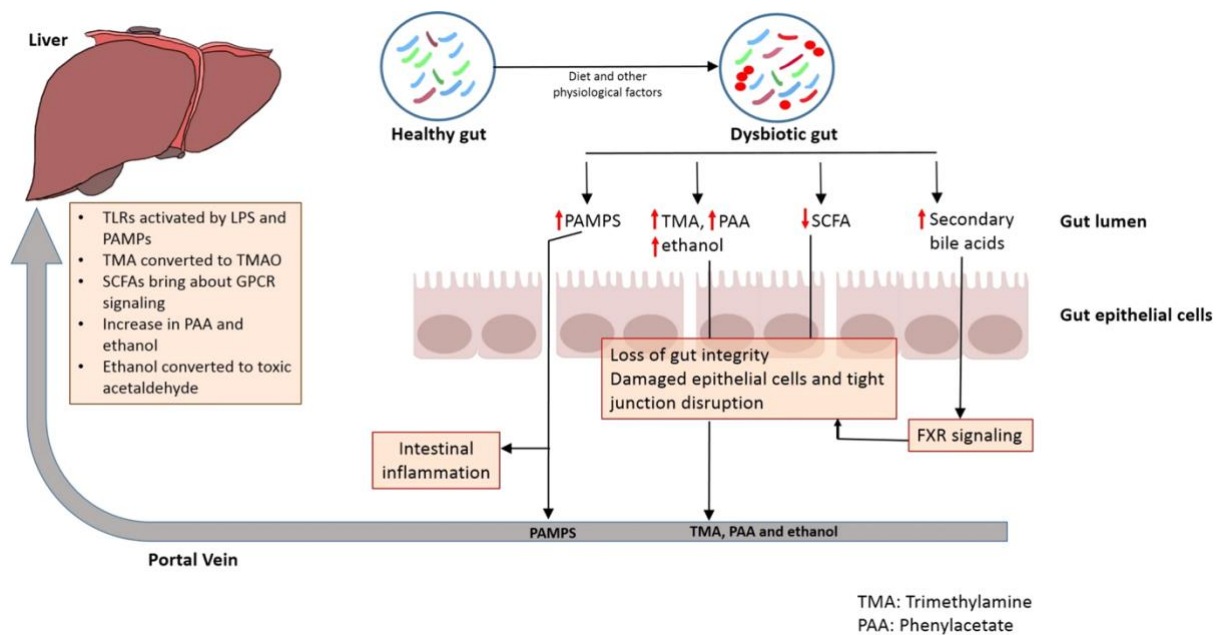
Η πλειονότητα των μελετών που εξετάστηκαν υποδεικνύουν μια συσχέτιση μεταξύ της βακτηριακής υπερανάπτυξης και δυσβίωσης στο έντερο και της ανάπτυξης αλκοολικής νόσου του ήπατος (ALD) και της κίρρωσης. Πράγματι, η διαταραγμένη λειτουργία του εντερικού φραγμού, η οποία σχετίζεται με την κατανάλωση αλκοόλ, σε συνδυασμό με τη βακτηριακή υπερανάπτυξη και δυσβίωση, θα μπορούσε να είναι πολύ σχετική με την ανάπτυξη ηπατικής παθολογίας που προκαλείται από το αλκοόλ, όπως της NAFLD, NASH ή ALD (Engen et al., 2015).

1.2.3 Άξονας Εντέρου-Ήπατος (Gut-Liver Axis)

Η κατανόηση της σύνδεσης μεταξύ του μικροβιώματος του εντέρου και του ήπατος έχει παράσχει σημαντικές γνώσεις για την παθοφυσιολογία των ηπατικών παθήσεων. Δεδομένου

ότι η μικροβιακή δυσβίωση του εντέρου αυξάνει τη διαπερατότητα του εντέρου, οι μεταβολίτες που συντίθενται από αυτά μπορούν να φτάσουν στο ήπαρ μέσω της πυλαίας κυκλοφορίας και να επηρεάσουν την ηπατική ανοσία και τη φλεγμονή. Τα κύτταρα του ανοσοποιητικού συστήματος που ενεργοποιούνται από αυτούς τους μεταβολίτες μπορούν, επίσης, να φτάσουν στο ήπαρ μέσω της λεμφικής κυκλοφορίας. Το ήπαρ επηρεάζει την ανοσία και το μεταβολισμό σε πολλά όργανα του σώματος, συμπεριλαμβανομένου του εντέρου. Απελευθερώνει χολικά οξέα (bile acids, BA) και άλλους μεταβολίτες στη χοληφόρο οδό από όπου εισέρχονται στη συστηματική κυκλοφορία (Anand & Mande, 2022).

Η NAFLD και η NASH είναι παθήσεις του ήπατος που έχουν συσχετιστεί με δυσβίωση στο μικροβίωμα του εντέρου. Παρόμοια συμπτώματα έχουν επίσης παρατηρηθεί σε άτομα που πάσχουν από αλκοολική ηπατική νόσο (ALD) λόγω υπερβολικής κατάχρησης αλκοόλ. Η μη προοδευτική μορφή αυτών των ασθενειών (π.χ. NAFLD) συχνά περιλαμβάνει συσσώρευση λίπους στο ήπαρ ή στεάτωση, ενώ η προοδευτική μορφή (π.χ. NASH) διαγιγνώσκεται από ηπατική βλάβη και φλεγμονή (Tripathi et al., 2018). Η ανάλυση δειγμάτων κοπράνων έδειξε να εμφανίζονται χαμηλότερα επίπεδα *Prevotella* και υψηλότερα *Bacteroides* και *Ruminococcus* στο έντερο ασθενών με NASH στο στάδιο 2 ή υψηλότερο σε σύγκριση με τα άτομα ελέγχου με ίνωση σταδίου 1 (Boursier et al., 2016). Αυτό δείχνει ότι αλλαγές στην εντερική μικροχλωρίδα συνδέονται με τη σοβαρότητα της νόσου. Έρευνες δείχνουν υψηλότερη αφθονία *Escherichia coli* και *Bacteroides vulgatus* στο έντερο ασθενών με NAFLD σε πρώιμα καθώς και προχωρημένα στάδια ίνωσης (Loomba et al., 2017).



Εικόνα 5. Μετατόπιση μεταβολιτών και μοριακών μοτίβων σχετιζόμενων με παθογόνα (PAMPs) που βιοσυντίθενται από μικρόβια στο έντερο μέσω της πυλαίας κυκλοφορίας στο ήπαρ, όπου ασκούν πολλαπλές επιδράσεις στην κατάσταση της υγείας του. Συντμήσεις: PAMPs: Μόρια που σχετίζονται με παθογόνα, PAA: πολυακρυλικό οξύ, TMA: τριμεθυλαμίνη, SCFA: λιπαρά οξέα βραχείας αλυσίδας, FXR: Υποδοχέας Farnesoid X (Anand et al., 2022).

Ορισμένοι μηχανισμοί διέπουν τη στενή αμφίδρομη επικοινωνία μεταξύ του εντέρου και του ήπατος (Εικόνα 5). Για παράδειγμα, μεταβολικοί μηχανισμοί του ξενιστή και του εντερικού μικροβιώματος, μεταβολίζουν αρκετά εξωγενή διατροφικά και περιβαλλοντικά συστατικά, καθώς και ενδογενή υποστρώματα, όπως αμινοξέα και BAs (Wahlström et al., 2016). Τα προϊόντα που παράγονται κατά τη διάρκεια αυτής της διαδικασίας μεταφέρονται στο ήπαρ μέσω της πυλαίας φλέβας, επηρεάζοντας έτσι την ηπατική φυσιολογία. Ομοίως, τα κύτταρα του ανοσοποιητικού που ενεργοποιούνται από διάφορα θρεπτικά συστατικά καθώς και μεταβολίτες από το μικροβίωμα του εντέρου μπορούν να εισέλθουν στο λεμφικό σύστημα και να ρυθμίσουν τις ανοσολογικές αποκρίσεις σε άπω όργανο όπως το ήπαρ. Από την άλλη πλευρά, το ήπαρ επικοινωνεί με το έντερο μέσω της απελευθέρωσης BA και άλλων μεταβολιτών στη χοληφόρο οδό της συστηματικής κυκλοφορίας (Chiang, 2013; Tripathi et al., 2018).

Ο εντερικός φραγμός, που αποτελείται από ισχυρά συνδεδεμένα κύτταρα, εξασφαλίζει επιλεκτική μεταφορά θρεπτικών ουσιών και περιορίζει την κίνηση παθογόνων οργανισμών από τον αυλό του εντέρου στο σύστημα του ξενιστή. Το μικροβίωμα του εντέρου επηρεάζει την ακεραιότητα του φραγμού του εντέρου είτε επηρεάζοντας τους μηχανισμούς σηματοδότησης του ανοσοποιητικού, είτε παράγοντας μεταβολίτες όπως τα λιπαρά οξέα βραχείας αλυσίδας

(SCFAs). Έτσι, διαταραχές σε οποιονδήποτε από αυτούς τους παράγοντες μπορεί να οδηγήσουν σε αύξηση της διαπερατότητας του εντέρου (Suzuki, 2020).

Μια υποβαθμισμένη ακεραιότητα του φραγμού του εντέρου είναι πιθανό να οδηγήσει σε μετατόπιση μικροοργανισμών και μορίων που προέρχονται από μικρόβια. Κάτω από αυτές τις συνθήκες, αυτά τα μικρόβια, καθώς και οι συντιθέμενοι μεταβολίτες τους μπορούν να μεταφερθούν μέσω του διαρρέντος εντέρου (leaky gut) στο ήπαρ, και κατά συνέπεια να μεταφερθούν μέσω του πυλαίου συστήματος στα απομακρυσμένα όργανα, προκαλώντας έτσι τη φλεγμονή και τη βλάβη τους (Εικόνα 6) (Anand & Mande, 2022).

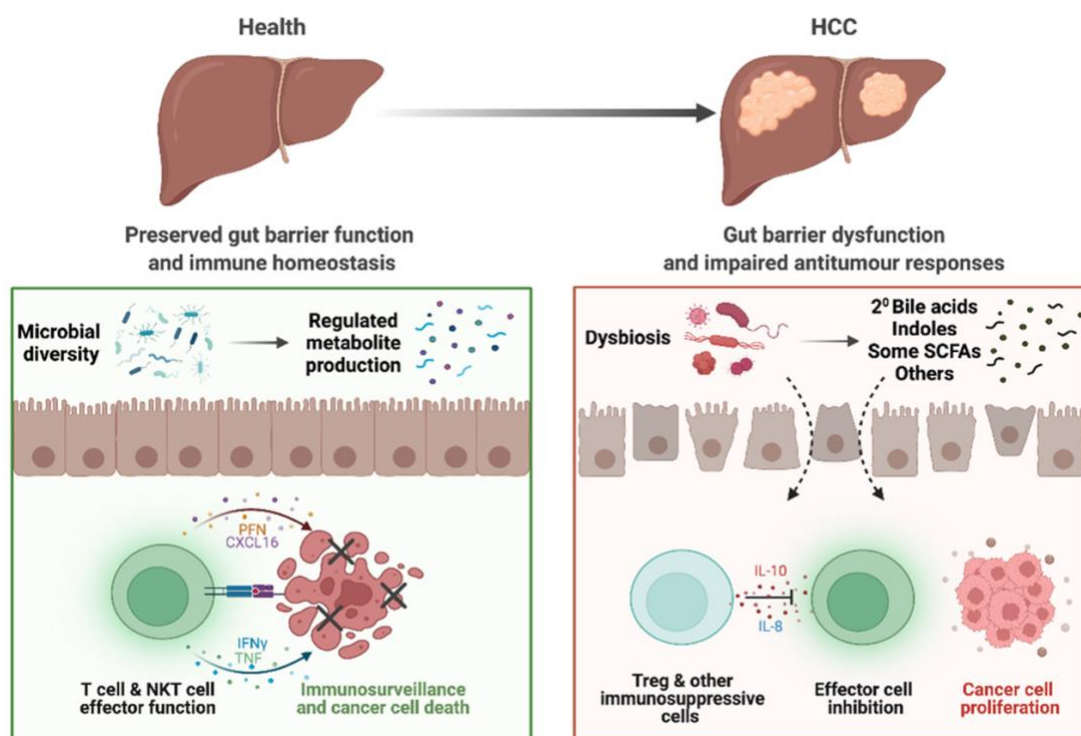
Η πλειονότητα των μελετών που εξετάστηκαν υποδεικνύουν μια συσχέτιση μεταξύ της βακτηριακής δυσβίωσης στο έντερο και της ανάπτυξης αλκοολικής νόσου του ήπατος (ALD) και κίρρωσης. Πράγματι, η διαταραγμένη λειτουργία του εντερικού φραγμού, η οποία σχετίζεται με την κατανάλωση αλκοόλ, σε συνδυασμό με τη βακτηριακή δυσβίωση, θα μπορούσε να είναι πολύ σχετική με την ανάπτυξη ηπατικής παθολογίας που προκαλείται από το αλκοόλ, όπως της NAFLD, NASH ή ALD (Engen et al., 2015). Η κατανάλωση αλκοόλ διαταράσσει τον εντερικό φραγμό, μέσω της επιβάρυνσης του οξειδωτικού στρες στο έντερο, το οποίο με τη σειρά του διαταράσσει τους στενούς συνδέσμους και προάγει την εντερική διαπερατότητα (Rao et al., 2004). Η αυξημένη εντερική διαπερατότητα επιτρέπει σε προφλεγμονώδη/παθογόνα μικροβιακά προϊόντα, όπως την ενδοτοξίνη (π.χ. λιποπολυσακχαρίτη [LPS] και πεπτιδογλυκάνη), να μετατοπίζονται από τον εντερικό αυλό στο ήπαρ μέσω της πυλαίας φλέβας. Η έκθεση σε αυτά τα βακτηριακά προϊόντα προκαλεί φλεγμονή στο ήπαρ, η οποία μπορεί να λειτουργήσει σε συνδυασμό με τις άμεσες επιδράσεις του αλκοόλ για να προκαλέσει ALD, καθώς και κίρρωση του ήπατος, όπως παρουσιάζεται και στην Εικόνα 4 (Schnabl & Brenner, 2014).

1.2.4 Σχέση εντερικού μικροβιώματος με το HCC

Η σύνδεση του μικροβιώματος με τον καρκίνο δεν είναι νέα. Περίπου το 20% των τύπων καρκίνου συσχετίζονται την τελευταία δεκαετία με λοιμώξεις από βακτήρια, ιούς ή παράσιτα (zur Hausen, 2009). Η καρκινογένεση σε κάθε σημείο του σώματος επηρεάζεται από εσωτερικούς και εξωτερικούς παράγοντες, που τελικά οδηγούν στον ανεξέλεγκτο πολλαπλασιασμό των κυττάρων και σε αντιδράσεις του ανοσοποιητικού, είτε υπέρ του ξενιστή, είτε υπέρ του καρκίνου. Τα μικρόβια δύνανται να επηρεάζουν την καρκινογένεση σε διάφορους

τύπους καρκίνου, επιδρώντας στη φλεγμονή, τη γενωμική σταθερότητα των κυττάρων του ξενιστή και παράγοντας μεταβολίτες που ρυθμίζουν τη γονιδιακή έκφραση των κυττάρων (Cogdill et al., 2018).

Σε έρευνα που εκπονήθηκε από τους Xie et al., βρέθηκε μια σημαντική μείωση στα επίπεδα του γένους *Akkermansia* σε ποντικά μοντέλου NASH-HCC (Xie et al., 2016). Το *Akkermansia* μπορεί να μειώσει τα επίπεδα LPS, και η αύξηση της αφθονίας του βελτιώνει το μεταβολικό προφίλ των ατόμων με παχυσαρκία που προκαλείται από τη διατροφή (Shin et al., 2014). Το *Akkermansia* παίζει επίσης ρόλο στη μείωση της συσσώρευσης λίπους, καθώς η ποσότητα του στο λεπτό έντερο συσχετίζεται αρνητικά με τη συνολική περιεκτικότητα σωματικού λίπους (Axling et al., 2012). Στα ίδια μοντέλα ποντικών, παρατηρήθηκε μείωση στα επίπεδα του γένους *Parasutterella*.



Εικόνα 6. Αλλοιωμένος άξονας εντέρου-ήπατος σε ασθενείς με HCC. Στους υγιείς οργανισμούς, το μικροβίωμα και οι βακτηριακοί μεταβολίτες που παράγονται διατηρούν την ομοιότητα του εντέρου. Οι διατηρημένες λειτουργίες του φραγμού του εντέρου αποτρέπουν τη βακτηριακή μετατόπιση και την επακόλουθη φλεγμονή, και επιτρέπουν μόνο επιλεγμένους μεταβολίτες να εισέλθουν στην κυκλοφορία. Τελικά, διατηρείται και η ομοιότητα στο ήπαρ, με ανοσολογική επιτήρηση που αποτρέπει την ηπατοκαρκινογένεση. Ωστόσο, σε ασθενείς με HCC, η δυσβίωση έχει ως αποτέλεσμα την απορυθμισμένη παραγωγή μεταβολιτών, συμπεριλαμβανομένων αυξημένων χολικών οξέων (BAs) και απορυθμισμένης σηματοδότησης των BA και λιπαρών οξέων βραχείας αλυσίδας (SCFA). Υπάρχει επίσης μειωμένη λειτουργία του εντερικού φραγμού, επιτρέποντας τη μετατόπιση βακτηρίων και μεταβολιτών προκαλώντας φλεγμονή.

Μελέτες σε ζώα έχουν αυξήσει την κατανόηση μας για τους μηχανισμούς ηπατοκαρκινογένεσης που σχετίζονται με τη δυσβίωση. Ωστόσο, δεδομένα από δείγματα

ανθρώπων είναι ακόμη σχετικά περιορισμένα (Πίνακας 2). Οι Grąt et al. έδειξαν, σε ένα μικρό δείγμα ασθενών με κίρρωση, ότι οι υψηλότεροι αριθμοί *E. coli* στα κόπρανα συσχετίζονται με HCC (Grąt et al., 2016). Σε ύστερη έρευνα αναλύθηκαν τα προφίλ βιοδεικτών και μικροχλωρίδας του εντέρου 20 ασθενών με κίρρωση, NAFLD ή πρώιμο HCC (Ponziani et al., 2019). Παρατηρήθηκε ότι όλοι οι ασθενείς με κίρρωση εμφάνισαν αύξηση στους δείκτες εντερικής διαπερατότητας και φλεγμονής, όπως LPS, και καλπροτεκτίνη κοπράνων, αντίστοιχα, σε σύγκριση με υγιείς ιστούς ελέγχου. Όλοι οι ασθενείς με κίρρωση εμφάνισαν, επίσης, μειωμένα επίπεδα του γένους *Akkermansia*, με αυξημένα *Enterobacteriaceae*, ενώ η υποομάδα με HCC παρουσίασε περαιτέρω αύξηση στην αφθονία των *Bacteroides* και *Ruminococcus*, με μειωμένο *Bifidobacterium*. Ακόμη, εντοπίστηκε μια αντίστροφη σχέση μεταξύ του *Akkermansia* και της καλπροτεκτίνης των κοπράνων, ενώ το *Bacteroides* συσχετίστηκε με αυξημένες τις προφλεγμονώδεις κυτοκίνες IL8 και IL13 (Ponziani et al., 2019).

Πίνακας 2. Διαφοροποιήσεις στο εντερικό μικροβίωμα ανθρώπων από έρευνες σε φυσιολογικούς και HCC ασθενείς.

Μελέτη	Σχετιζόμενα εντερικά μικρόβια	Βιβλιογραφία
HCC vs healthy controls	<i>Neisseria</i> ↑, <i>Enterobacteriaceae</i> ↑, <i>Veillonella</i> ↑, <i>Limnobacter</i> ↑, <i>Enterococcus</i> ↓, <i>Phyllobacterium</i> ↓, <i>Clostridium</i> ↓, <i>Ruminococcus</i> ↓, <i>Coprococcus</i> ↓	(Zheng et al., 2020)
HCC vs healthy controls	άλφα-ποικιλομορφία εντερικού μικροβιώματος ↓, <i>Proteobacteria</i> ↑, <i>Enterobacteriaceae</i> ↑, <i>Bacteroides xyloxyloformans</i> ↑, <i>B. caecimuris</i> ↑, <i>Ruminococcus gnavus</i> ↑, <i>Clostridium bolteae</i> ↑, <i>Veillonellaparvula</i> ↑, <i>Oscillospiraceae</i> ↓, <i>Erysipelotrichaceae</i> ↓	(Behary et al., 2021)
HCC vs healthy controls	<i>Proteobacteria</i> ↑, <i>Desulfococcus</i> ↑, <i>Enterobacter</i> ↑, <i>Prevotella</i> ↑, <i>Veillonella</i> ↑, <i>Cetobacterium</i> ↓	(Ni et al., 2019)
HCC vs healthy controls	<i>Bacteroides</i> ↑, <i>Akkermansia</i> ↓, <i>Bifidobacterium</i> ↓	(Ponziani et al., 2019)
HCC vs healthy controls	<i>Escherichia coli</i> ↑	(Grąt et al., 2016)
HCC vs healthy controls	<i>Klebsiella</i> ↑, <i>Haemophilus</i> ↑, <i>Alistipes</i> ↓, <i>Phascolarctobacterium</i> ↓, <i>Ruminococcus</i> ↓	(Ren et al., 2019)

Ο βαθμός δυσβίωσης που σχετίζεται με το HCC αυξάνεται καθώς αυξάνεται το στάδιο ανάπτυξης καρκίνου. Με την εισαγωγή ενός δείκτη, που ονομάζεται βαθμός δυσβίωσης (D_{dys}), οι Ni et al. διαπίστωσαν ότι οι περιπτώσεις HCC είχαν υψηλότερα μικροβιακά επίπεδα

βακτηρίων, που τα χαρακτηρίζαν ως προφλεγμονώδη, στα κόπρανα σε σύγκριση με υγιή άτομα. Σημειωτέον, ο D_{dys} αυξήθηκε σημαντικά μεταξύ των περιπτώσεων HCC, σε σχέση με τους φυσιολογικούς ιστούς. Επιπλέον, ο D_{dys} έτεινε να αυξάνεται, καθώς το στάδιο HCC αυξανόταν, παρόλο που η διαφορά δεν ήταν στατιστικά σημαντική μεταξύ των σταδίων (Ni et al., 2019).

Πρόσφατα, οι Ren et al. συγκρίνοντας το μικροβίωμα του εντέρου, σε ασθενείς με HCC και υγιείς χαρακτήρισαν το μικροβίωμα του εντέρου μεταξύ περιπτώσεων HCC και πρότειναν να αξιολογηθεί η πιθανότητα χρήσης του ως μη επεμβατικού βιοδείκτη για τη διάγνωση της νόσου (Ren et al., 2019). Διαπιστώθηκε ότι η μικροβιακή ποικιλομορφία των κοπράνων σε πρώιμο HCC με κίρρωση παρουσίασε αύξηση σε σύγκριση με αυτήν στους ασθενείς με κίρρωση. Επιπλέον, σε σύγκριση με την κίρρωση, η αφθονία του φύλου *Actinobacteria* αυξήθηκε στο πρώιμο HCC. Αντίστοιχα, οι αφθονίες 13 γενών όπως τα *Parabacteroides* και *Gemmiger* αυξήθηκαν στους ασθενείς με πρώιμο HCC σε σύγκριση με αυτούς με κίρρωση. Αντίθετα, οι αφθονίες των γενών που παράγουν βουτυρικό οξύ παρουσίασαν μείωση, ενώ αυτές των γενών που παράγουν LPS αυξήθηκαν σε ασθενείς με πρώιμο HCC σε σύγκριση με φυσιολογικά άτομα. Επιπλέον, οι συγγραφείς εντόπισαν τους βέλτιστους 30 μικροβιακούς δείκτες μεταξύ μη HCC και πρώιμων περιπτώσεων HCC. Συγκεκριμένα, επαληθεύτηκε η ισχυρή δυνατότητα των μικροβιακών δεικτών του εντέρου στη διάγνωση πρώιμη ή ακόμη και προχωρημένης HCC (Kang et al., 2022; Ren et al., 2019).

1.2.5 Ενδοογκικό μικροβίωμα (*Intratumoral Microbiome*)

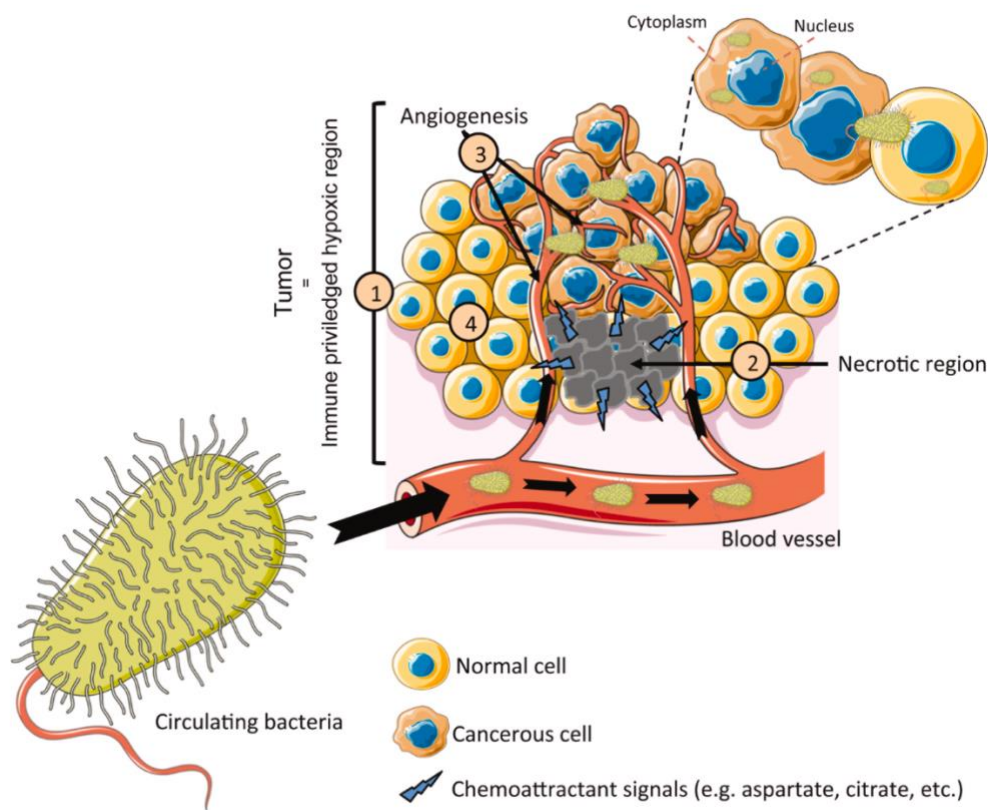
Πρόσφατα, η στενή συσχέτιση μεταξύ του μικροπεριβάλλοντος του όγκου και της εμφάνισης καρκίνου πλαισιώθηκε με την παρουσία συγκεκριμένων βακτηρίων μέσα σε καρκινικούς ιστούς. Η πρόοδος στην Αλληλούχιση Επόμενης Γενιάς (Next Generation Sequencing, NGS) έχει διευκολύνει σημαντικά τον εντοπισμό μικροοργανισμών σε ιστούς σε όλο το σώμα και τη σχέση τους με την υγεία και τις ασθένειες, συμπεριλαμβανομένου του καρκίνου (Gosiewski et al., 2017). Σε όλους τους τύπους όγκων, ο χαρακτηρισμός της μικροβιακής ποικιλομορφίας είχε ως αποτέλεσμα την ανίχνευση ενός ξεχωριστού ενδοογκικού μικροβιώματος για κάθε τύπο καρκίνου (Nejman et al., 2020).

Οι μικροοργανισμοί που βρίσκονται στον ιστό μπορεί να προάγουν την ογκογένεση, αλλάζοντας διάφορες πτυχές του γονιδιώματος του ξενιστή. Για παράδειγμα, τα βακτήρια που

κατοικούν στον ιστό μπορούν να προκαλέσουν άμεσα βλάβη στο DNA, μέσω των τοξινών που παράγουν. Συγκεκριμένα, το *Escherichia coli* που παράγει το μεταβολίτη colibactin (Cuevas-Ramos et al., 2010), και τα *Proteobacteria* και *Campylobacter jejuni* που παράγουν τοξίνες θανατηφόρες για τα κύτταρα (He et al., 2019), προκαλούν φθορές στο δίκλωνο DNA, ενώ βακτηριακά είδη όπως το *Bilophila wadsworthia*, το *Fusobacterium nucleatum* και το *Desulfonivibrio desulfurican* αυξάνουν τα επίπεδα του οξειδωτικού στρες, μέσω της παραγωγής ROS. Εκτός από την άμεση πρόκληση βλάβης στο DNA, μικρόβια που βρίσκονται στους ιστούς, όπως το *E. coli* και το *H. pylori* μπορούν να διαταράξουν τους μηχανισμούς αποκατάστασης της βλάβης του DNA, για να επιδεινώσουν περαιτέρω την αστάθεια του γονιδιώματος και να οδηγήσουν στην ογκογένεση (Park et al., 2022). Ο κυτταρικός κύκλος μπορεί, επίσης, να διαταραχθεί άμεσα από ιούς που βρίσκονται σε ιστούς, προάγοντας την ογκογένεση, όπως αποδεικνύεται από τους μηχανισμούς που κρύβονται πίσω από κακοήθειες που προκαλούνται από τον ιό των ανθρώπινων θηλωμάτων (HPV) (Suh et al., 2014) και τον ιό Epstein-Barr (EBV) (Yin et al., 2019). Οι μικροοργανισμοί και τα υπολείμματά τους, όπως στην περίπτωση των ενδογενών ρετροϊών, έχει επίσης αποδειχθεί ότι προξενούν την ογκογένεση, μέσω άλλων μηχανισμών που σχετίζονται με το γονιδίωμα του ξενιστή, όπως της αλλαγής του τοπικού επιγενετικού τοπίου, ή της πειρατείας της μεταγραφής του ξενιστή (Gonzalez-Cao et al., 2016).

Μετά από μια μεγάλη μεταγονιδιωματική μελέτη διαφόρων στερεών όγκων, αποκαλύφθηκε η ύπαρξη ενδοκυτταρικών βακτηρίων, τόσο στα κύτταρα του ανοσοποιητικού, όσο και στα καρκινικά του ξενιστή (Nejman et al., 2020). Προκειμένου να εξηγηθεί η παρουσία βακτηρίων σε καρκινικούς ιστούς, έχει προταθεί ένα μοντέλο (Εικόνα 7), στο οποίο φαίνεται πως οι όγκοι πληρούν όλες τις προαπαιτούμενες συνθήκες για την υποστήριξη υψηλής βακτηριακής ποικιλομορφίας (Bermudes et al., 2002). Με βάση τις εγγενείς ιδιότητες των καρκινικών ιστών, αποτελούν ένα επιτρεπτό περιβάλλον για την υποστήριξη της βακτηριακής εισβολής, επιβίωσης και ανάπτυξης. Οι όγκοι αποτελούν συχνά ένα υποξικό περιβάλλον, στο οποίο τα περιορισμένα επίπεδα οξυγόνου ευνοούν την ανάπτυξη συγκεκριμένων βακτηριακών ειδών, όπως προαιρετικά και υποχρεωτικά αναερόβιων βακτηρίων. Ο νεκρωτικός ιστός είναι πλούσιος σε θρεπτικά συστατικά (π.χ. πουρίνες), ωφελώντας τόσο στην επιβίωση των βακτηρίων, όσο και στον γρήγορο πολλαπλασιασμό των κυττάρων. Ακόμη, κατά την ανάπτυξη του καρκίνου, αναδύονται νέα αιμοφόρα αγγεία, τα οποία περιβάλλουν τους αναπτυσσόμενους ιστούς. Αυτή η αγγειογενετική διαδικασία οδηγεί στο σχηματισμό ενός δικτύου ανώμαλων αγγείων υψηλής διαπερατότητας, γύρω και μέσα στον όγκο, το οποίο στη συνέχεια

χρησιμοποιείται από τα βακτήρια για να εισέλθουν στους ιστούς. Τα βακτήρια μπορούν να πολλαπλασιαστούν εξωκυτταρικά ή ενδοκυτταρικά. Τα καρκινικά κύτταρα προστατεύονται από την ανοσολογική επιτήρηση του ξενιστή, καθώς περιβάλλονται από κύτταρα του ανοσοποιητικού συστήματος που εμποδίζουν την επέκταση των νεκρωτικών περιοχών του όγκου. Έτσι, τα βακτήρια που εισβάλλουν επωφελούνται από αυτή την έλλειψη επιτήρησης από το ανοσοποιητικό σύστημα και πολλαπλασιάζονται ευκολότερα, προστατευμένα από τους αμυντικούς μηχανισμούς του ξενιστή (Heymann et al., 2021).



Εικόνα 7. Μοντέλο της βακτηριακής εισβολής στον καρκινικό ιστό του ξενιστή. Λόγω της αγγειογένεσης και της προνομιούχας θέσης των καρκινικών κυττάρων ανάμεσα σε κύτταρα του ανοσοποιητικού συστήματος, παρέχεται προστασία και τα κατάλληλα θρεπτικά συστατικά στα βακτήρια, ώστε να επιβιώσουν και να αναπτυχθούν (Heymann et al., 2021).

1.2.6 Ενδοογκικό μικροβίωμα και HCC

Εκτεταμένα δεδομένα δείχνουν ότι η χρόνια μόλυνση μπορεί να οδηγήσει σε καρκίνο σε διάφορα όργανα. Τα παράσιτα, όπως τα σχιστοσώματα, και ένα βακτήριο, το *Helicobacter pylori*, ταξινομήθηκαν ως καρκινογόνα τύπου I από τον Διεθνή Οργανισμό Έρευνας για τον Καρκίνο το 1994 (Møller et al., 1995). Η ανακάλυψη της παρουσίας DNA του γένους *Helicobacter* στο υλικό του ήπατος από ασθενείς με ηπατική νόσο οδήγησε στην υπόθεση ότι αυτά τα βακτήρια μπορεί να παίζουν ρόλο στην εξέλιξη ηπατικών βλαβών από χρόνια ιογενή

ηπατίτιδα σε κίρρωση και HCC. Οι καθοριστικοί παράγοντες αυτής της εξέλιξης δεν είναι ακόμη πλήρως κατανοητοί, συμπεριλαμβανομένων εκείνων που εμφανίζονται σε HCV-θετικούς ασθενείς (Rocha et al., 2005).

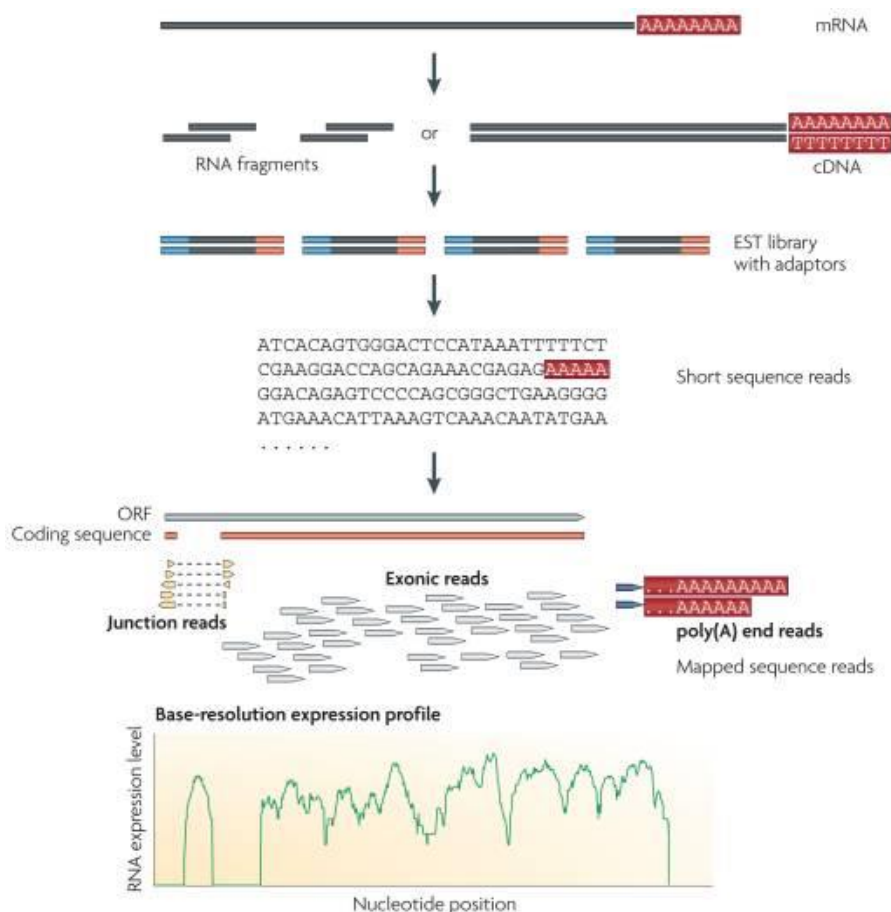
Σε έρευνες που συσχετίζουν το μικροβίωμα καρκινικών ιστών HCC με φυσιολογικούς παρακαρκινικούς ιστούς (Qu et al., 2022), παρατηρούνται διαφοροποιήσεις, όχι τόσο στην ποικιλομορφία μεταξύ των δύο ομάδων, αλλά όσον αφορά συγκεκριμένες οικογένειες ή γένη βακτηρίων, που αποτελούν βιοδείκτες για χαρακτηριστικά καρκίνου. Τα αποτελέσματα έδειξαν ότι τα βακτηριακά φύλα των *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* και *Firmicutes*, που μαζί αντιστοιχούσαν στο 90% των αλληλουχιών κατά μέσο όρο, ήταν οι τέσσερις κυρίαρχοι πληθυσμοί, τόσο στον όγκο, όσο και στους αντίστοιχους παρακείμενους φυσιολογικούς ιστούς. Επίσης, το γένος *Pseudomonas* μειώθηκε σημαντικά στους καρκινικούς ιστούς, κάτι που ενδεχομένως οφείλεται στις αντικαρκινικές του βιολογικές λειτουργίες (Pang et al., 2022).

Σε άλλη πρόσφατη έρευνα, η σύγκριση πραγματοποιήθηκε μεταξύ τριών ομάδων ιστών, υγιών, παρακαρκινικών (υγιή κύτταρα πλησίον του όγκου), και HCC, όπου ελέγχθηκαν στατιστικά οι ποικιλομορφίες στα 3 διακριτά περιβάλλοντα (Huang et al., 2022). Τα μικροβιακά προφίλ των HCC και παρακαρκινικών ιστών αποδείχθηκαν πως δε διαφέρουν σημαντικά σε επίπεδο ποικιλομορφίας, ενώ αυτές οι ομάδες συγκριτικά με τους υγιείς ιστούς είχαν αυξημένη ποικιλομορφία. Τα φύλα (Phyla) *Proteobacteria*, *Firmicutes* και *Actinobacteriota*, και οι βακτηριακές τάξεις (Class) *Bacilli* και *Actinobacteria*, εμπλουτίζονταν σταθερά σε παρακαρκινικούς ιστούς και HCC, ενώ τα *Gammaproteobacteria* υπήρχαν σε ιδιαίτερη αφθονία στους ιστούς HCC. Επιπλέον, τα φύλα *Firmicutes* και *Actinobacteriota* και η τάξη *Saccharimonadia* βρέθηκαν να σχετίζονται με τα κλινικοπαθολογικά χαρακτηριστικά των ασθενών με HCC, συμπεριλαμβανομένου του φύλου, της διαβάθμισης κίρρωσης ή του μεγέθους του όγκου, υποδηλώνοντας ότι αυτά τα είδη μπορεί να συνδέονται με την καρκινογένεση ή την εξέλιξη του HCC. Η αυξημένη επιβάρυνση στο HCC και η σημαντική συσχέτιση με τη σταδιοποίηση της κίρρωσης μπορεί να υποδηλώνουν σημαντικό ρόλο για τα *Proteobacteria* και τα *Gammaproteobacteria* στην καρκινογένεση ή την εξέλιξη του HCC, επιβεβαιώνοντας την παρατήρηση πολλών πρόσφατων ερευνών (Komiya et al., 2021; Loo et al., 2017).

1.3 Μεταγραφωμική Ανάλυση RNA-seq Δεδομένων

1.3.1 Αλληλούχιση RNA (RNA Sequencing)

Η αλληλούχιση RNA (RNA-Seq) χρησιμοποιεί πρόσφατα αναπτυγμένες τεχνολογίες βαθιάς αλληλούχισης. Γενικά, ένας πληθυσμός RNA, ολικό ή κλασματοποιημένο, μετατρέπεται σε μια βιβλιοθήκη θραυσμάτων cDNA με προσαρμογείς συνδεδεμένους στο ένα ή και στα δύο άκρα (Εικόνα 8). Κάθε μόριο αναλύεται με αλληλούχιση υψηλής απόδοσης (high throughput sequencing) για να ληφθούν σύντομες αλληλουχίες από το ένα άκρο (αλληλουχία ενός άκρου, SE) ή και τα δύο άκρα (αλληλουχία ζεύγους άκρων, PE). Τα αναγνώσματα είναι συνήθως μήκους 30-400 bp, ανάλογα με την τεχνολογία προσδιορισμού αλληλουχίας DNA που χρησιμοποιείται. Μετά τον προσδιορισμό της αλληλουχίας, τα προκύπτοντα αναγνώσματα είτε ευθυγραμμίζονται με ένα γονιδίωμα ή μεταγράφο αναφοράς, είτε συναρμολογούνται «de novo». (Z. Wang et al., 2009).



Εικόνα 8. Διάγραμμα διαδικασίας RNA-seq: Οι μακρές RNA αλληλουχίες μετατρέπονται αρχικά σε μία βιβλιοθήκη από τμήματα συμπληρωματικού DNA (cDNA) και στη συνέχεια οι προσαρμογείς (αντάπτορες) αλληλούχισης προστίθενται σε κάθε τμήμα cDNA, όπου λαμβάνεται τελικά μία βραχεία αλληλουχία. Το ανάγνωσμα αυτό ευθυγραμμίζεται με ένα γονιδίωμα ή μεταγράφο αναφοράς, και κατηγοριοποιείται σε 3 τύπους: αναγνώσματα εξονίων (exonic reads), αναγνώσματα χωρίς κενά (junction reads) και αναγνώσματα με πολυαδενυλιωμένο άκρο (polyA reads). Αυτοί οι τρεις τύποι χρησιμοποιούνται για τη δημιουργία ενός προφίλ έκφρασης ανάλογης βάσης για κάθε γονίδιο, όπως φαίνεται στο κάτω μέρος.

Αν και το RNA-Seq εξακολουθεί να είναι μια τεχνολογία υπό ενεργό ανάπτυξη, προσφέρει πολλά βασικά πλεονεκτήματα σε σχέση με τις υπάρχουσες τεχνολογίες. Πρώτον, σε αντίθεση με τις προσεγγίσεις που βασίζονται στον υβριδισμό, το RNA-Seq δεν περιορίζεται στην ανίχνευση μεταγραφών που αντιστοιχούν στην υπάρχουσα γονιδιωματική αλληλουχία. Για παράδειγμα, RNA-Seq με βάση το Roche 454 Life Science έχει χρησιμοποιηθεί για τον προσδιορισμό της αλληλουχίας του μεταγραφώματος της πεταλούδας *Glanville frillary* (Vera et al., 2008). Αυτό καθιστά το RNA-Seq ιδιαίτερα ελκυστικό για οργανισμούς που δεν αποτελούν πρότυπο, με γονιδιωματικές αλληλουχίες που δεν έχουν ακόμη προσδιοριστεί. Επιπλέον, τα σύντομα αναγνώσματα 30 ζευγών βάσεων (bp) από το RNA-Seq δίνουν πληροφορίες για το πώς συνδέονται δύο εξόνια, ενώ τα μεγαλύτερα αναγνώσματα ή τα σύντομα αναγνώσματα ζεύγους αποκαλύπτουν τη συνδεσιμότητα μεταξύ πολλαπλών εξονίων. Αυτοί οι παράγοντες καθιστούν το RNA-Seq χρήσιμο για τη μελέτη πολύπλοκων μεταγραφωμάτων. Ακόμη, το RNA-Seq μπορεί να αποκαλύψει παραλλαγές αλληλουχίας στις μεταγραφόμενες περιοχές (Morin et al., 2008).

Ένα δεύτερο πλεονέκτημα του RNA-Seq έναντι άλλων τεχνικών, όπως των μικροσυστοιχιών DNA, είναι ότι δεν έχει ανώτατο όριο για ποσοτικοποίηση, το οποίο συσχετίζεται με τον αριθμό των αλληλουχιών που λαμβάνονται. Κατά συνέπεια, έχει ένα μεγάλο δυναμικό εύρος επιπέδων έκφρασης στα οποία μπορούν να ανιχνευθούν οι μεταγραφές: εκτιμήθηκε ένα εύρος μεγαλύτερο από 9.000 φορές σε μια μελέτη που ανέλυσε 16 εκατομμύρια χαρτογραφημένες αναγνώσεις στο *Saccharomyces cerevisiae* (Nagalakshmi et al., 2008), και εκτιμήθηκε ένα εύρος πέντε τάξεων μεγέθους για 40 εκατομμύρια αναγνώσματα αλληλουχίας ποντικών (Mortazavi et al., 2008). Αντίθετα, οι μικροσυστοιχίες DNA στερούνται ευαισθησίας για γονίδια που εκφράζονται είτε σε χαμηλά, είτε σε πολύ υψηλά επίπεδα, και ως εκ τούτου έχουν μικρότερο δυναμικό εύρος. Το RNA-Seq έχει επίσης αποδειχθεί ότι είναι εξαιρετικά ακριβές για τον ποσοτικό προσδιορισμό των επιπέδων έκφρασης, όπως προσδιορίζεται χρησιμοποιώντας ποσοτική PCR (qPCR) (Z. Wang et al., 2009).

1.3.2 Μεταγραφωμική (Transcriptomics)

Τα πλεονεκτήματα του RNA-Seq επιτρέπουν τη δημιουργία μιας άνευ προηγουμένου συνολικής εικόνας του μεταγραφώματος και της οργάνωσής του για έναν αριθμό ειδών και τύπων κυττάρων. Πριν από την εκτεταμένη χρήση του RNA-Seq, ήταν γνωστό ότι ένα πολύ

μεγαλύτερο από το αναμενόμενο κλάσμα των γονιδιωμάτων της ζύμης και του ανθρώπου μεταγράφονται, και είχε βρεθεί ένας αριθμός διακριτών ισομορφών για πολλά γονίδια (Bertone et al., 2004; David et al., 2006). Ωστόσο, τα άκρα των περισσότερων μετάγραφων και εξονίων δεν είχαν αναλυθεί με ακρίβεια, και η έκταση της ετερογένειας παρέμενε ελάχιστα κατανοητή. Το RNA-Seq, με την υψηλή ανάλυση και ευαισθησία του έχει αποκαλύψει πολλές νέες μεταγραφόμενες περιοχές και ισομορφές γνωστών γονιδίων και έχει χαρτογραφήσει τα άκρα 5' και 3' για πολλά γονίδια.

Κατά την τελευταία δεκαετία, οι τεχνολογίες προσδιορισμού αλληλούχισης DNA και RNA έχουν σημειώσει τεράστια πρόοδο, όσον αφορά την απόδοση, την ταχύτητα και τη μείωση του κόστους αλληλούχισης. Ομοίως, η πρόσβαση σε γονιδιώματα και μεταγραφώματα έχει ωφελήσει σε μεγάλο βαθμό τη μελέτη των ζώων και των φυτών (Telenti et al., 2016). Οι τεχνολογίες αλληλούχισης δεύτερης ή επόμενης γενιάς (NGS) και τρίτης γενιάς που χρησιμοποιούνται σήμερα διαθέτουν τεράστια βελτίωση σε σχέση με τις τεχνολογίες αλληλούχισης πρώτης γενιάς, ειδικά όσον αφορά την αναλογία απόδοσης/κόστους και την ταχύτητα επεξεργασίας. Η παραδοσιακή αλληλούχιση Sanger, η οποία χρησιμοποιήθηκε ευρέως για σχεδόν τρεις δεκαετίες από τη δημοσίευσή της το 1977, μπορούσε να επιτύχει μόνο περιορισμένη ή πολύ χαμηλή απόδοση (Naranpranawa et al., 2020). Η μελέτη του ανθρώπινου γονιδιώματος χρησιμοποίησε μεθόδους προσδιορισμού αλληλούχισης Sanger και χρειάστηκε πάνω από 10 χρόνια και σχεδόν 3 δισεκατομμύρια \$ ΗΠΑ για την ολοκλήρωσή του. Αντίθετα, το σύστημα Illumina HiSeq μπορεί τώρα να αναλύσει τις αλληλουχίες πάνω από 45 ανθρώπινων γονιδιωμάτων για 1.000 \$ ΗΠΑ, το καθένα σε μια μέρα. Η Illumina είναι μια από τις τεχνολογίες NGS, της οποίας το όργανο HiSeq X μπορεί να επιτύχει συνολική απόδοση έως και 900 Gb με κόστος ανά Gb μόλις 714 \$ ΗΠΑ (Lander et al., 2001).

Λόγω αυτής της δραστηρικής προόδου, η ποσότητα των ακατέργαστων αναγνωσμάτων που παράγονται από τους προσδιοριστές αλληλουχίας είναι τεράστια και η υψηλή κάλυψη προσθέτει μια τεράστια ποσότητα επικαλυπτόμενων τμημάτων DNA ή RNA, ειδικά σε μεγάλα γονιδιώματα. Επειδή ο όγκος των δεδομένων προς επεξεργασία είναι πολύ μεγάλος, η συναρμολόγηση των σύντομων αναγνωσμάτων για την κατασκευή του πλήρους γονιδιώματος ή του μεταγραφώματος γίνεται δύσκολη, απαιτώντας υψηλή υπολογιστική ισχύ και χρόνο εκτέλεσης. Αυτό οδηγεί σε σημαντική συμφόρηση στην υπολογιστική βιολογία και τη βιοπληροφορική (Scholz et al., 2012).

Η συναρμολόγηση των πρωτογενών δεδομένων ακολουθίας ακολουθεί μία από τις δύο προσεγγίσεις: (1) βάσει αναφοράς και (2) *de novo assembly*. Η συναρμολόγηση βάσει αναφοράς, που ονομάζεται επίσης συγκριτική συναρμολόγηση, είναι η διαδικασία αναδημιουργίας του γονιδιώματος ή του μεταγραφώματος, χρησιμοποιώντας προηγούμενη γνώση. Σε αυτή τη μέθοδο, ένα προηγούμενος συναρμολογημένο γονιδίωμα ενός στενά συγγενούς οργανισμού χρησιμοποιείται ως πρότυπο για την ευθυγράμμιση των εν λόγω αναγνωσμάτων. Κάθε ανάγνωσμα τοποθετείται στην πιο πιθανή θέση απέναντι στο πρότυπο αναφοράς. Η προκύπτουσα διάταξη θα μπορούσε να είναι παρόμοια με την αλληλουχία αναφοράς, αλλά όχι εντελώς πανομοιότυπη, καθώς θα μπορούσαν να υπάρχουν περιοχές που είναι σημαντικά διαφορετικές (Nock et al., 2011; Zeevi et al., 2019).

Η συναρμολόγηση αλληλουχίας που διαβάζεται χωρίς προηγούμενη γνώση του μεταγραφώματος ή χωρίς γονιδίωμα αναφοράς ονομάζεται συναρμολόγηση εκ νέου, ή “*de novo assembly*”. Ενώ η συναρμολόγηση εκ νέου παρέχει την ευκαιρία να δημιουργηθεί οποιαδήποτε αλληλουχία νέου οργανισμού, η διαδικασία παρουσιάζει πολλές προκλήσεις, συμπεριλαμβανομένων τμηματικών διπλότυπων, επαναλήψεων αλληλουχιών, γονιδίων που λείπουν, και του τεράστιου αριθμού ακατέργαστων αναγνώσεων που πρέπει να χειριστούν (Alkan et al., 2011).

1.3.3 Μετα-μεταγραφωμική (Metatranscriptomics)

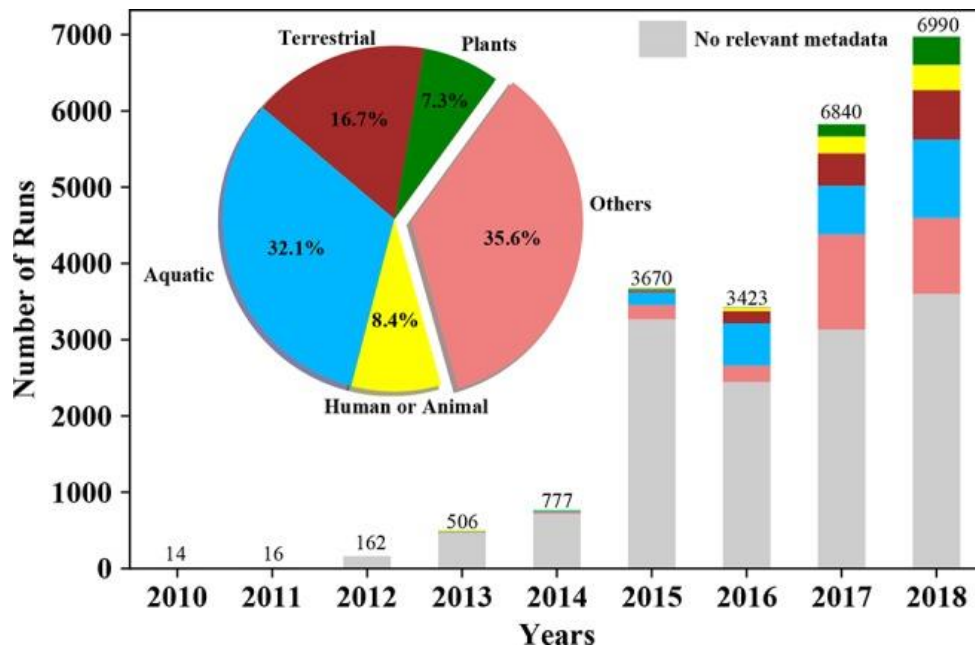
Τις τελευταίες δεκαετίες έχουν σημειωθεί σημαντικές πρόοδοι στις τεχνολογίες αλληλούχισης που έχουν μεταμορφώσει τον τρόπο με τον οποίο διεξάγονται βιολογικά πειράματα, ιδιαίτερα όταν πρόκειται για τη μελέτη πολύπλοκων συστημάτων μικροβιώματος. Ωστόσο, το μεγαλύτερο μέρος της αλληλούχισης υψηλής απόδοσης έχει επικεντρωθεί στον προσδιορισμό αλληλουχίας DNA ολόκληρων κοινοτήτων, χρησιμοποιώντας είτε στοχευμένες προσεγγίσεις, όπως η αλληλούχιση PCR-amplicon του γονιδίου 16S rRNA, είτε η “shotgun” αλληλούχιση όλου του διαθέσιμου DNA του δείγματος (μεταγονιδιωμική).

Αυτές οι μέθοδοι συνέβαλαν σε πολλές ανακαλύψεις την περασμένη δεκαετία, βοηθώντας στον καλύτερο χαρακτηρισμό συστημάτων μικροβιώματος από περιβάλλοντα που κυμαίνονται από το ανθρώπινο έντερο (Wang et al., 2015) έως το έδαφος (Feng et al., 2018) και τους ωκεανούς (Acinas et al., 2021). Αν και οι μελέτες που βασίζονται στην ανάλυση του

16S rRNA γονιδίου χαρακτηρίζουν μόνο άμεσα το ταξινομικό προφίλ ενός μικροβιώματος, αποτελούν μια οικονομικά αποδοτική επιλογή για την μελέτη της ποικιλομορφίας του μικροβιώματος (μετρώντας το μέγιστο δυναμικό εύρος σχετικής αφθονίας) πολλών δειγμάτων, χρησιμοποιώντας ελάχιστη αλληλουχία. Ωστόσο, όλο και περισσότερες μελέτες χρησιμοποιούν τη “shotgun” μεταγενωμική, καθώς οι εξελίξεις στις τεχνολογίες αλληλούχισης επιτρέπουν την ολοκληρωμένη σύλληψη των περισσότερων μελών του μικροβιώματος, ενώ ταυτόχρονα διευκρινίζουν πιθανά γονίδια και λειτουργικές οδούς. Ένας από τους κύριους περιορισμούς της μεταγενωμικής προσέγγισης είναι ότι δεν διακρίνει τα ενεργά από τα ανενεργά μέλη ενός μικροβιώματος και, επομένως, δεν μπορεί να βοηθήσει στη διάκριση αυτών που συμβάλλουν στην παρατηρούμενη συμπεριφορά του οικοσυστήματος, από εκείνα που είναι απλώς παρόντα.

Η χρήση της αλληλούχισης RNA (RNA-Seq) για την καταγραφή των εκφρασμένων μεταγράφων μέσα σε ένα μικροβίωμα, σε μια δεδομένη χρονική στιγμή, κάτω από ένα σύνολο περιβαλλοντικών συνθηκών, εστιάζει στα βιολογικά ενεργά μέλη του μικροβιώματος. Με το RNA-Seq, μπορούν να ανιχνευθούν και να συσχετιστούν με μεταβολικές οδούς, γονίδια σχετικά χαμηλής έκφρασης, συμπεριλαμβανομένου ολόκληρου του μεταγραφώματος που περιλαμβάνει μη κωδικοποιήσιμα RNA.

Πριν από την εμφάνιση της αλληλουχίας υψηλής απόδοσης, οι τεχνολογίες μικροσυστοιχιών (microarrays) χρησιμοποιούνταν, επίσης, ευρέως για τη μέτρηση των επιπέδων έκφρασης γνωστών μεταγράφων από οργανισμούς ή ακόμα και μικροβιακές κοινότητες (Shakya et al., 2019). Με την εφαρμογή τεχνολογιών NGS στο RNA, είναι πλέον δυνατό όχι μόνο να μετρηθούν γνωστοί στόχοι μεταγραφής, αλλά και να ανακαλυφθούν προηγουμένως άγνωστα μεταγραφώματα, και παραλλαγές μεταγράφων, απευθείας από τα δεδομένα της αλληλούχισης.



Εικόνα 9. Διάγραμμα απεικόνισης της αύξησης των μεταγραφωμικών ερευνών από το 2010 έως το 2018, και της απόδοσης της χρήσης των ερευνών σε διάφορους τομείς (Shakya et al., 2019).

Στο σύντομο χρονικό διάστημα, από τότε που εισήχθη για πρώτη φορά η τεχνολογία στις αρχές της δεκαετίας του 2000, ο αριθμός των μεταγραφωμικών ερευνών ή η αλληλούχιση RNA από μικροβιακές κοινότητες έχει αυξηθεί σημαντικά (Εικόνα 9). Όσον αφορά τις εφαρμογές, η τεχνική έχει χρησιμοποιηθεί για τον χαρακτηρισμό ενεργών μικροβίων σε μια κοινότητα (Bashiardes et al., 2016), την ανακάλυψη νέων μικροβιακών αλληλεπιδράσεων (Bikel et al., 2015), την ανίχνευση ρυθμιστικού RNA (Naroli et al., 2017), παρακολουθώντας την έκφραση των γονιδίων και προσδιορίζοντας τη σχέση μεταξύ των ιών και του ξενιστή (Moniruzzaman et al., 2017). Ωστόσο, αυτή η επαναστατική μέθοδος έχει τα δικά της μειονεκτήματα. Όπως συμβαίνει με τις περισσότερες μεταγραφικές μεθόδους, ο πειραματικός σχεδιασμός και ο τρόπος συλλογής των δειγμάτων είναι κρίσιμοι και απαιτείται επαρκές υλικό για αλληλούχιση. Επιπλέον, η μετα-μεταγραφωμική δεν είναι πάντα σε θέση να συλλάβει ολόκληρο το μεταγράφομα, λόγω εν μέρει της πολυπλοκότητας (υψηλή ποικιλομορφία και σχετικές αναλογίες μελών) ορισμένων μικροβιακών κοινοτήτων, του μεγάλου δυναμικού εύρους έκφρασης μεταγραφής, του μικρού χρόνου ημιζωής του RNA και λοιπών τεχνικών περιορισμών.

Παρόλο που οι προηγούμενες μελέτες είχαν επικεντρωθεί, κυρίως, στην περιγραφή της ταξινομικής σύνθεσης των μικροβιακών κοινοτήτων και του λειτουργικού τους δυναμικού, πολλές μελέτες χρησιμοποιούν τώρα μεταγραφωμικά εργαλεία για να κατανοήσουν καλύτερα τις αλληλεπιδράσεις μεταξύ των μικροβίων και του ξενιστή τους (Pérez-Losada et al., 2015),

για τον εντοπισμό σημαντικών ενεργών μεταβολικών μονοπατιών (Franzosa et al., 2014) και την κατανόηση του ρόλου της γονιδιακής έκφρασης στην εξέλιξη μίας νόσου (Nowicki et al., 2018). Με περαιτέρω προόδους στις τεχνολογίες αλληλούχισης και προσαρμοσμένες μεθόδους βιοπληροφορικής ανάλυσης, η μετα-μεταγραφωμική υπόσχεται να γίνει ένα αναπόσπαστο εργαλείο για τη διερεύνηση των μικροβιακών κοινοτήτων στον άνθρωπο, αλλά και σε άλλους ξενιστές.

1.4 Θεωρητικό Υπόβαθρο της Υπολογιστικής Ανάλυσης

1.4.1 Συλλογή Αρχικών Δεδομένων

Βασική προϋπόθεση για την ανάλυση αλληλουχιών RNA είναι η λήψη αξιόπιστων και χρήσιμων αρχικών δεδομένων. Οι παράμετροι που κατά κύριο λόγο λαμβάνονται υπόψη βασίζονται στις ανάγκες του εξεταζόμενου βιολογικού εγχειρήματος, με τις 2 βασικότερες να είναι το βάθος αλληλούχισης και ο αριθμός των αντιγράφων (replicates). Ωστόσο, πληροφορίες για τον αριθμό των δειγμάτων, τη χρονική στιγμή λήψης τους και λοιπές πειραματικές παράμετροι παίζουν καθοριστικό ρόλο στην επιλογή κατάλληλων δεδομένων (Koch et al., 2018).

Καταρχάς, το βάθος αλληλούχισης, δηλαδή ο αριθμός των RNA αναγνωσμάτων εντός ενός δείγματος, δηλώνει την ποσότητα μίας συγκεκριμένης αλληλουχίας κατά τη διάρκεια της αλληλούχισης. Μεγαλύτερο βάθος αλληλούχισης καθιστά τον αριθμό αναγνωσμάτων που θα καταγραφεί μεγαλύτερο, και την ποσοτικοποίησή τους πιο ακριβή. Για αναλύσεις RNA προτείνεται βάθος αλληλούχισης των τουλάχιστον 40 εκατομμυρίων αναγνωσμάτων. Ακόμη το μήκος της αλληλουχίας στην μεταγραφωμική και μετα-μεταγραφωμική παίζουν μεγάλο ρόλο, καθώς ιδιαίτερα στην τελευταία προτείνεται τα αναγνώσματα να είναι τουλάχιστον 100 ζεύγη αζωτούχων βάσεων (bp) (Westreich et al., 2018).

Επιπρόσθετα, για την ανάλυση RNA αλληλουχιών, καταλυτικός παράγοντας είναι αν οι προκυπτόμενες αλληλουχίες είναι SE ή PE. Στην SE αλληλούχιση, η αλληλουχία διαβάζεται από το ένα άκρο, ενώ κατά την PE αλληλούχιση, η αλληλουχία διαβάζεται και από τα δύο της άκρα. Η πρώτη υπερτερεί της δεύτερης αναφορικά με το κόστος και της ταχύτητας αλληλούχισης. Ωστόσο, η δεύτερη έχει αποδειχθεί πως είναι πιο ακριβής, και προσφέρει τον καλύτερο εντοπισμό επαναλαμβανόμενων τμημάτων της αλληλουχίας. Εάν η εξειδίκευση της

PE αλληλούχισης δεν είναι απαραίτητη στο πλαίσιο ενός πειράματος, είθισται να χρησιμοποιούνται SE αλληλουχίες (Corley et al., 2017).

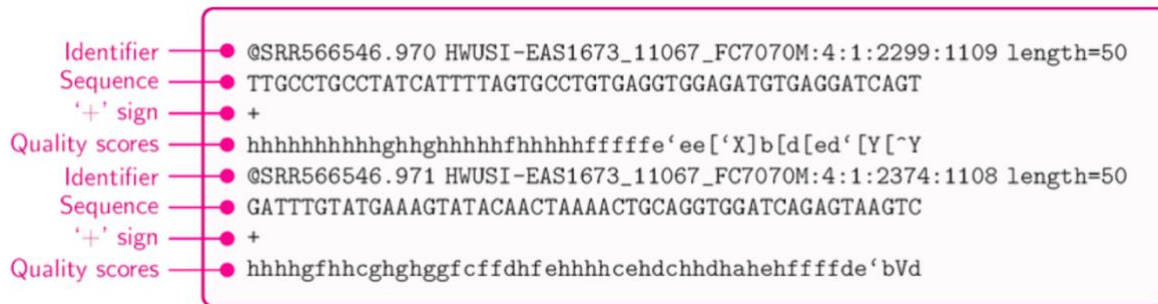
Τα δεδομένα που παραλαμβάνονται από τις τεχνικές αλληλούχισης διαθέτουν τη μορφή FASTA ή FASTQ. Τα αρχεία αυτά διαθέτουν κείμενο με την αλληλουχία και περαιτέρω αναγνωριστικές πληροφορίες.

Η μορφή FASTA αποτελεί την απλούστερη μορφή αλληλουχίας, καταγράφοντας μόνο τις αζωτούχες βάσεις της αλληλουχίας. Αποτελείται από δύο σειρές, με την πρώτη να αποτελεί την αναγνωριστική σειρά της αλληλουχίας (όνομα και σχόλια), και τη δεύτερη την αλληλουχία (Εικόνα 10) (Hosseini et al., 2016).



Εικόνα 10. Η δομή της FASTA μορφής. Η πρώτη σειρά αποτελεί τον τίτλο της αλληλουχίας στο ανάγνωσμα (Header), ενώ η επόμενη ή οι επόμενες αποτελούν την αλληλουχία. Οι αλληλουχίες μεταξύ τους διαχωρίζονται από τον τίτλο, ο οποίος έχει ως αναγνωριστικό στοιχείο το χαρακτήρα '>' στην αρχή του (Hosseini et al., 2016).

Η μορφή FASTQ είναι η πλέον διαδεδομένη στη χρήση της για την βιοπληροφορική ανάλυση δειγμάτων. Πέρα από τις πληροφορίες που παρέχει η FASTA, η συγκεκριμένη μορφή παρέχει πληροφορίες όσον αφορά την ποιότητα αλληλούχισης του κάθε αναγνώσματος. Η δομή της FASTQ μορφής αποτελείται από 4 γραμμές. Η πρώτη γραμμή αποτελείται από τον τίτλο της αλληλουχίας, περιλαμβάνοντας το όνομα, και ενδεχόμενα σχόλια όσον αφορά την αλληλουχία. Η δεύτερη γραμμή, όπως και στη FASTA μορφή, περιλαμβάνει τις αζωτούχες βάσεις της αλληλουχίας. Η τρίτη γραμμή περιέχει μόνο το σύμβολο '+' προκειμένου να σημάνει το τέλος της αλληλουχίας, συνεχίζοντας με την τέταρτη γραμμή και την περιγραφή της ποιότητας κάθε βάσης της αλληλουχίας (Εικόνα 11).



Εικόνα 11. Η δομή της FASTQ μορφής. Η πρώτη σειρά αποτελεί τον τίτλο της αλληλουχίας στο ανάγνωσμα (Identifier), ενώ η επόμενη ή οι επόμενες αποτελούν την αλληλουχία. Οι αλληλουχίες μεταξύ τους διαχωρίζονται από τον τίτλο, ο οποίος έχει ως αναγνωριστικό στοιχείο το χαρακτήρα '@' στην αρχή του. Το σύμβολο '+' διαχωρίζει την αλληλουχία από την περιγραφή της ποιότητας των βάσεων (Quality scores), η οποία πραγματοποιείται με κωδικοποιημένους χαρακτήρες που βαθμολογούν την ποιότητα (Hosseini et al., 2016).

1.4.2 Ποιοτικός Έλεγχος (Quality Control)

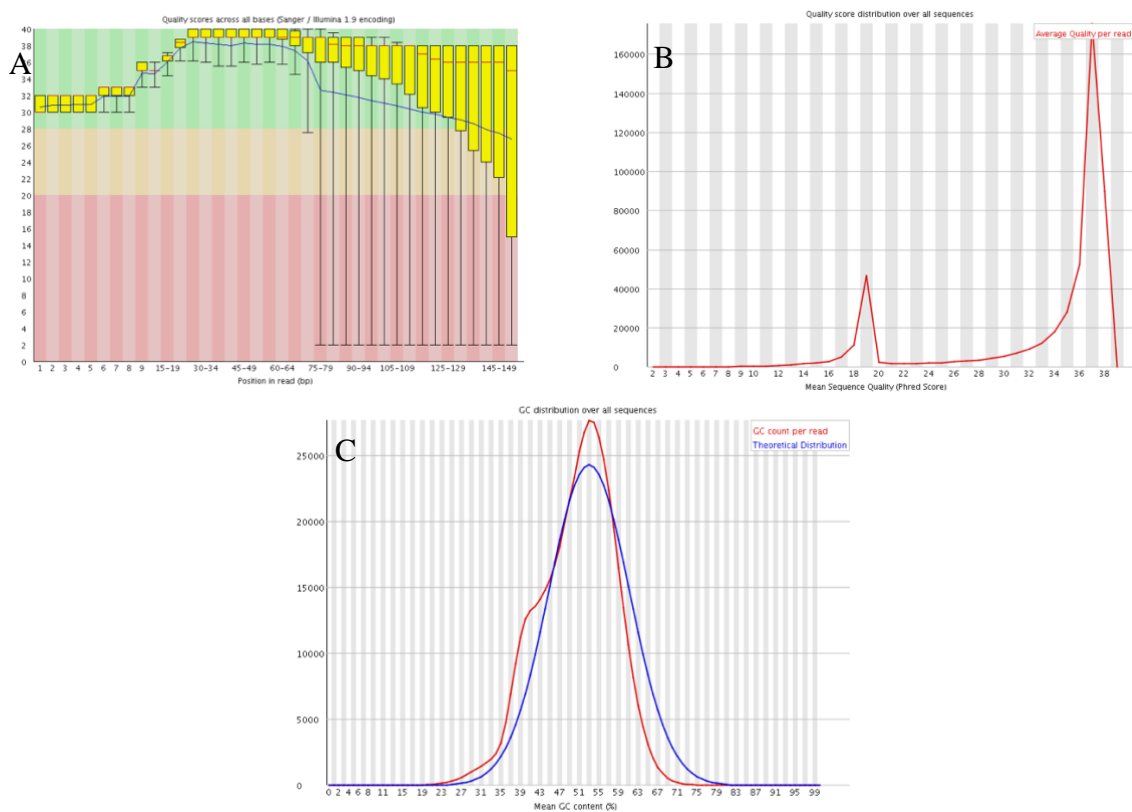
Το πιο σύνθητες πρώτο στάδιο κατά τη διαδικασία επεξεργασίας RNA-Seq δεδομένων είναι η ανάλυση ποιότητας των αλληλουχιών, που ονομάζεται ποιοτικός έλεγχος (Quality Control, QC). Το εργαλείο FASTQC προσφέρει ποικίλα περιγραφικά στατιστικά, όσον αφορά την ποιότητα και τη σύσταση των αλληλουχιών (Εικόνα 12). Ενδεικτικά, ο ποιοτικός έλεγχος περιλαμβάνει την ανάλυση ποιότητας της αλληλουχίας, το ποσοστό γουανίνης και κυτοσίνης (%GC content), την παρουσία προσαρμογέων (adaptors) και αλληλουχιών που εκπροσωπούνται σε μεγάλο βαθμό, καθώς και τον έλεγχο πολλών αντιγράφων k-mers (Leggett et al., 2013). Το εργαλείο αυτό χρησιμοποιείται για την επεξεργασία αλληλουχιών από την τεχνολογία Illumina.

Η αναφορά του FastQC περιλαμβάνει 10 σχηματικές αξιολογήσεις, με πρώτη, και μία από τις σημαντικότερες, την ποιότητα των αλληλουχιών ανά βάση (per base sequence quality) (Εικόνα 12A). Όσο υψηλότερη είναι η βαθμολογία, τόσο καλύτερη είναι η ποιότητα της βάσης. Η χρωματική διαφορά διαχωρίζει την ανάλυση σε βάσεις καλής (πράσινο), μέτριας (πορτοκαλί) και χαμηλής ποιότητας (κόκκινο), κατά τον κατακόρυφο άξονα.

Μία ακόμη αξιολόγηση είναι η ποιότητα των αλληλουχιών ανά δείγμα (per base sequence scores), μια αναφορά που επιτρέπει να αξιολογηθεί εάν ένα υποσύνολο των αλληλουχιών έχει συνολικά χαμηλές τιμές ποιότητας (Εικόνα 12B). Συχνά ισχύει ότι ένα υποσύνολο αλληλουχιών θα έχει καθολικά κακή ποιότητα, επειδή είναι κακώς απεικονισμένο. Ωστόσο αυτές οι αλληλουχίες αντιπροσωπεύουν μόνο ένα μικρό ποσοστό των συνολικών αναγνωσμάτων. Εάν ένα σημαντικό ποσοστό των αναγνωσμάτων σε ένα πείραμα έχει

συνολικά χαμηλή ποιότητα, τότε αυτό υποδηλώνει κάποιο είδος συστηματικού προβλήματος - πιθανώς σε ένα μόνο μέρος της πειραματικής διαδικασίας.

Η σύσταση γουανίνης και κυτοσίνης ανά αλληλουχία (per sequence GC content) μετράει το περιεχόμενο GC σε όλο το μήκος κάθε αλληλουχίας σε ένα αρχείο και το συγκρίνει με μια διαμορφωμένη κανονική κατανομή του GC περιεχομένου (Εικόνα 12C). Σε μια τυχαία βιβλιοθήκη αναμένεται μια σχεδόν κανονική κατανομή του GC περιεχομένου, όπου η κεντρική κορυφή αντιστοιχεί στο συνολικό περιεχόμενο GC του υποκείμενου γονιδιώματος. Δεδομένου ότι δεν είναι γνωστό το περιεχόμενο GC του γονιδιώματος, αυτό υπολογίζεται από τα παρατηρούμενα δεδομένα και χρησιμοποιείται για τη δημιουργία μιας κατανομής αναφοράς.



Εικόνα 12. Παραδείγματα της αναφοράς FastQC και των σημαντικότερων χαρακτηριστικών που λαμβάνονται υπόψη.

1.4.3 Αποκοπή αλληλουχιών χαμηλής ποιότητας

Τα αποτελέσματα που παραλαμβάνονται από τον ποιοτικό έλεγχο μπορούν να δώσουν μία συνολική εικόνα για την μετέπειτα επεξεργασία τους. Σε περιπτώσεις χαμηλής ποιότητας των δειγμάτων, χρησιμοποιούνται εργαλεία τα οποία θα εντοπίσουν και θα αφαιρέσουν τις χαμηλής ποιότητας αλληλουχίες, ή περιοχές αλληλουχιών. Για αλληλουχίες της τεχνολογίας Illumina, ένα εργαλείο που χρησιμοποιείται για την ανωτέρω διαδικασία είναι το Trimmomatic.

Ακόμη, αφαιρούνται οι χαμηλής ποιότητας αλληλουχίες και οι αντάπτορες. Παρόλο που πολλά εργαλεία είχαν αναπτυχθεί για τις ανωτέρω διαδικασίες, κανένα, πέρα από το Trimmomatic, δεν είχε την ευελιξία και την δυνατότητα υποδοχής PE αναγνωσμάτων (Bolger et al., 2014).

1.4.4 Λειτουργικός Χαρακτηρισμός (*Functional Annotation*)

Η ευθυγράμμιση ομοιότητας είναι η διαδικασία σύγκρισης μίας αλληλουχίας με μία αλληλουχία αναφοράς, με σκοπό την εύρεση της βέλτιστης θέσης του αναγνώσματος. Τα εργαλεία ευθυγράμμισης (aligners) αξιολογούν την ποιότητα της ευθυγράμμισης στη σχετική θέση, με το κάθε εργαλείο να χρησιμοποιεί το δικό του σύστημα. Συγκεκριμένα, για αναγνώσματα από ανθρώπινους οργανισμούς, η αλληλουχία αναφοράς που χρησιμοποιείται για τη σύγκριση είναι το ανθρώπινο γονιδίωμα, GRCh38.p13 (Nurk et al., 2022).

Η πρώτη μεγάλη κατηγορία αναφέρεται ως «εργαλεία ευθυγράμμισης αναγνωσμάτων χωρίς κενά» (unspliced aligners), οι οποίοι ευθυγραμμίζουν τα αναγνώσματα σε μια αναφορά, χωρίς να επιτρέπουν μεγάλα κενά, όπως αυτά που προκύπτουν από αναγνώσματα που εκτείνονται σε όρια εξονίων ή συνδέσεις κενών. Μπορούν να χρησιμοποιηθούν εργαλεία ευθυγράμμισης χωρίς κενά για την ευθυγράμμιση των αναγνωσμάτων με ένα μεταγράφομα ή γονιδίωμα αναφοράς. Η ευθυγράμμιση με ένα μεταγράφομα αναφοράς μειώνει την ανάγκη χειρισμού συνδέσεων των κενών, αλλά περιορίζεται στην ανάλυση γνωστών μεταγράφων. Τα εργαλεία ευθυγράμμισης χωρίς κενά χωρίζονται γενικά σε δύο υποκατηγορίες με βάση τη μεθοδολογία τους – οι «seed» μέθοδοι και οι «μέθοδοι μετασχηματισμού Burrows-Wheeler (BWT)» (Garber et al., 2011).

Οι μέθοδοι seed ευθυγραμμίζουν σύντομες υποαλληλουχίες, (seeds), από κάθε ανάγνωσμα σε μια αναφορά, απαιτώντας τέλεια ευθυγράμμιση στην υποαλληλουχία. Οι μέθοδοι BWT δημιουργούν ένα ευρετήριο Burrows-Wheeler της αλληλουχίας αναφοράς και αναζητούν αποτελεσματικά τέλειες ευθυγραμμίσεις. Αναντιστοιχίες μπορεί να επιτρέπονται με μια εκθετική αύξηση της υπολογιστικής πολυπλοκότητας. Γενικά, οι μέθοδοι μετασχηματισμού Burrows-Wheeler είναι ταχύτερες από τις μεθόδους seed, αλλά οι δεύτερες παρέχουν αυξημένη ευαισθησία (Martin & Wang, 2011). Η πιο διαδεδομένη και αποτελεσματική μέθοδος BWT είναι το Bowtie2. Επίσης, το εργαλείο HISAT2 χρησιμοποιεί στη βάση του το ίδιο μοντέλο με το Bowtie2, με ορισμένες επιπρόσθετες βελτιστοποιήσεις που επιτρέπουν την πιο επιτυχή ευθυγράμμιση αλληλουχιών μικρού μήκους. Έτσι, το HISAT2

αποτελεί πλέον το πιο εκτεταμένο σε χρήση εργαλείο για την ανάλυση RNA-Seq δεδομένων σε ευκαρυώτες, όπου οι αλληλουχίες είναι σχετικά μικρές (D. Kim et al., 2019).

Ωστόσο, οι μέθοδοι στην κατηγορία ευθυγράμμισης αναγνωσμάτων χωρίς κενά περιορίζονται από την αδυναμία τους να χειριστούν αναγνώσματα που εκτείνονται σε συνδέσεις των κενών κατά την ευθυγράμμιση σε ένα γονιδίωμα αναφοράς. Η δεύτερη μεγάλη κατηγορία μεθόδων ευθυγράμμισης, τα εργαλεία ευθυγράμμισης με κενά, ή «spliced aligners», ευθυγραμμίζουν τα αναγνώσματα σε ολόκληρο το γονιδίωμα, με τα αναγνώσματα που εκτείνονται σε ιντρόνια να απαιτούν μεγάλα κενά (Au et al., 2010). Τα εργαλεία αυτά χωρίζονται επίσης σε δύο μεγάλες κατηγορίες με βάση τη μεθοδολογία τους - τις μεθόδους «exon-first» και τις μεθόδους «seed-and-extend».

Οι μέθοδοι exon-first ξεκινούν αντιστοιχίζοντας ολόκληρες αναγνώσεις στο γονιδίωμα χρησιμοποιώντας εργαλεία ευθυγράμμισης αναγνωσμάτων χωρίς κενά και, στη συνέχεια, αναζητούν ευθυγραμμίσεις με κενά για τα υπόλοιπα αναγνώσματα. Οι προσεγγίσεις exon-first είναι αποτελεσματικές, αλλά μπορεί να χάνουν τις πραγματικές ευθυγραμμίσεις όταν είναι διαθέσιμη μια ευθυγράμμιση σε ψευδογονίδιο (Garber et al., 2011). Οι μέθοδοι seed-and-extend διαχωρίζουν τα αναγνώσματα σε υποαλληλουχίες που ευθυγραμμίζονται στο γονιδίωμα, και, οι υποψήφια θέσεις ευθυγράμμισης εξετάζονται με πιο ευαίσθητες μεθόδους. Όπως και οι ευθυγραμμιστές χωρίς κενά, οι μέθοδοι seed-and-extend είναι πιο αργές, αλλά πιο ευαίσθητες, και παρουσιάζουν βελτιωμένη απόδοση κατά την ευθυγράμμιση αναγνωσμάτων από πολυμορφικά δείγματα (Garber et al., 2011). Γνωστές μέθοδοι exon-first και seed-and-extend είναι το TopHat2 και το GSNAP, αντίστοιχα.

Μετά την ευθυγράμμιση, κατασκευάζεται ένας πίνακας που ονομάζεται μήτρα καταμέτρησης (count matrix), που παρουσιάζει την καταμέτρηση των αναγνωσμάτων (read counts) στα δείγματα. Έτσι, ο πίνακας αυτός αποτελείται από γραμμές για κάθε γονίδιο, και στήλες για τα δείγματα. Τα κελιά, λοιπόν, δείχνουν την ποσότητα του κάθε γονιδίου (γραμμή) στο δείγμα (στήλη). Οι μηδενικές τιμές των κελιών υποδηλώνουν την απουσία ενός γονιδίου στο εν προκειμένω δείγμα.

Ο στόχος της κανονικοποίησης είναι οι διαφορές στα κανονικοποιημένα πλήθη ανάγνωσης να αντιπροσωπεύουν διαφορές στην πραγματική έκφραση. Η κανονικοποίηση είναι σωστή όταν η σχέση μεταξύ των κανονικοποιημένων αριθμών ανάγνωσης είναι σωστή.

Δεδομένου ότι το πραγματικό προϊόν της γονιδιακής έκφρασης δεν μετρείται ποτέ, θεωρείται ότι η αληθινή έκφραση ενός γονιδίου είναι η ποσότητα mRNA/κυττάρου που παράγει. Αυτός φαίνεται να είναι ο ορισμός που χρησιμοποιείται συνήθως, καθώς προηγούμενες έρευνες θεωρούν ότι ένα γονίδιο είναι διαφορετικά εκφρασμένο (differentially expressed - DE) σε διαφορετικές βιολογικές συνθήκες εάν υπάρχει διαφορά στην ποσότητα mRNA/κυττάρου που παράγει υπό αυτές τις συνθήκες (Evans et al., 2018).

Οι βασικότερες παράμετροι που συνυπολογίζονται για την κανονικοποίηση των δεδομένων ενός δείγματος ή μίας ομάδας δειγμάτων είναι το βάθος αλληλούχισης, το μήκος των γονιδίων και η σύσταση του RNA.

Το βάθος αλληλούχισης είναι απαραίτητο να λαμβάνεται υπόψη στην κανονικοποίηση για σύγκριση δεδομένων μεταξύ διαφορετικών δειγμάτων. Τα αναγνώσματα ενός γονιδίου σε ένα δείγμα μπορούν να συγκριθούν με τα αναγνώσματα του ίδιου γονιδίου σε άλλο δείγμα μόνο αν τα συνολικά αναγνώσματα όλων των γονιδίων είναι ισόποσα σε κάθε δείγμα.

Η κανονικοποίηση ως προς το μήκος των γονιδίων είναι απαραίτητη για την ποσοτική σύγκριση της γονιδιακής έκφρασης μεταξύ διαφορετικών γονιδίων εντός ενός δείγματος. Η σύγκριση μεταξύ του αριθμού των αναγνωσμάτων δύο γονιδίων μπορεί να γίνει μόνο όταν τα δύο γονίδια έχουν ίδιο μήκος.

Όσον αφορά τη σύσταση του RNA, η ύπαρξη λίγων γονιδίων με μεγάλη ποσότητα αντιγράφων, οι διαφορές στον αριθμό των γονιδίων μεταξύ δειγμάτων και η ύπαρξη ξένων αλληλουχιών, μπορούν να παραποιήσουν τα αποτελέσματα της ποσοτικής σύγκρισης. Αυτές οι παράμετροι είναι, συνεπώς, απαραίτητο να συνυπολογιστούν στην κανονικοποίηση για ασφαλέστερα συμπεράσματα σε συγκρίσεις μεταξύ δειγμάτων. Είναι, επιπλέον, ιδιαίτερα σημαντικές στον έλεγχο διαφορικής έκφρασης γονιδίων.

Οι πιο συχνά χρησιμοποιούμενες μέθοδοι κανονικοποίησης είναι οι CPM, RPKM/FKPM, καθώς και μέθοδοι από εργαλεία μελέτης της διαφορικής έκφρασης, όπως ο διάμεσος λόγων (median of ratios) του DESeq2 (Love et al., 2014) και ο συντετμημένος μέσος των M τιμών (Trimmed Mean of M-values, TMM) του edgeR (Maza, 2016).

Οι μετρήσεις ανά γονίδιο κανονικοποιούνται σε CPM (counts per million) διαιρώντας τις με τον συνολικό αριθμό των χαρτογραφημένων αναγνωσμάτων ανά δείγμα και

πολλαπλασιάζοντας με 10^6 . Τα κανονικοποιημένα δεδομένα CPM στη συνέχεια μετασχηματίστηκαν με \log_2 , χρησιμοποιώντας μετατόπιση κατά 1, προκειμένου να αποφεύγονται μηδενικές τιμές στο λογάριθμο (Bushel et al., 2020).

Οι μετρήσεις ανά γονίδιο κανονικοποιούνται με τη χρήση της προσέγγισης «σταθμισμένου» συντημημένου μέσου M τιμών (TMM) στο πακέτο edgeR, του Bioconductor. Μετά την περικοπή των δεδομένων, δημιουργούνται συντελεστές κανονικοποίησης για κάθε δείγμα χρησιμοποιώντας τη συνάρτηση calcNormFactors. Στη συνέχεια χρησιμοποιούνται συντελεστές κανονικοποίησης για την προσαρμογή του συνολικού αντιστοιχισμένου αριθμού αναγνωσμάτων από κάθε δείγμα (Robinson & Oshlack, 2010).

Οι μέθοδοι RPKM/FPKM (Reads/Fragments per kilobase of exon per million reads/fragments) συνυπολογίζουν το βάθος αλληλούχισης και το μήκος γονιδίων και χρησιμοποιούνται κυρίως για συγκρίσεις γονιδίων του ίδιου δείγματος. Έτσι, οι μέθοδοι αυτές δεν εφαρμόζονται ικανοποιητικά για έλεγχο διαφορικής έκφρασης γονιδίων (Zhao et al., 2020).

1.4.5 Ανάλυση διαφορικής γονιδιακής έκφρασης (Differential Gene Expression Analysis)

Καθώς η ποσοτικοποίηση RNA-Seq βασίζεται σε μετρήσεις αναγνωσμάτων που αποδίδονται σε μετάγραφα, οι αρχικές προσεγγίσεις για την ανάλυση της διαφορικής έκφρασης χρησιμοποιούσαν διακριτές κατανομές πιθανοτήτων, όπως η κατανομή Poisson και η αρνητική διωνυμική κατανομή (Robinson & Smyth, 2007). Η κατανομή Poisson και τα εργαλεία που την αξιοποιούν, όπως το Myrna (Langmead et al., 2010) και το DEGseq (L. Wang et al., 2010), λαμβάνουν υπόψη τη διακύμανση μεταξύ των αντιγράφων σε ένα δείγμα. Η αρνητική διωνυμική κατανομή (γνωστή και ως κατανομή γάμμα-Poisson) είναι μια γενικευμένη κατανομή Poisson, επιτρέποντας πρόσθετη διακύμανση, που ονομάζεται υπερδιασπορά. Ορισμένες, πιο σύγχρονες, μέθοδοι, όπως το δημοφιλές edgeR, λαμβάνουν ακατέργαστες μετρήσεις αλληλουχιών ανάγνωσης και εισάγουν πιθανές πηγές πόλωσης στο στατιστικό μοντέλο για να πραγματοποιήσουν μια ολοκληρωμένη κανονικοποίηση καθώς και μια ανάλυση διαφορικής έκφρασης (Robinson et al., 2010). Το DESeq2, όπως το edgeR, χρησιμοποιεί την αρνητική διωνυμική ως κατανομή αναφοράς και παρέχει τη δική του προσέγγιση κανονικοποίησης (Love et al., 2014). Ορισμένες ανεξάρτητες μελέτες σύγκρισης έχουν αποδείξει ότι η επιλογή της μεθόδου μπορεί να επηρεάσει σημαντικά το αποτέλεσμα της

ανάλυσης και ότι καμία μεμονωμένη μέθοδος δεν είναι πιθανό να αποδώσει ευνοϊκά για όλα τα σύνολα δεδομένων (Seyednasrollah et al., 2015).

Τα βασικότερα βήματα της ανάλυσης διαφορικής έκφρασης γονιδίων παρουσιάζεται παρακάτω (Garber et al., 2011; Z. Wang et al., 2009):

- [1]. Κανονικοποίηση των μετρήσεων των αναγνωσμάτων (εφόσον δεν έχει προηγηθεί). Ποικίλα εργαλεία, όπως το edgeR και το DESeq2 πραγματοποιούν την κανονικοποίηση με δικές τους παραμέτρους.
- [2]. Υπολογισμός της διασποράς των μετρήσεων των αναγνωσμάτων για κάθε γονίδιο σε κάθε μετάγραφο.
- [3]. Υπολογισμός του p-value. Η τιμή αυτή δίνει την πιθανότητα που όταν η μηδενική υπόθεση είναι αληθής, η διαφορά των μετρήσεων θα είναι η ίδια ή μεγαλύτερη από τα παρατηρούμενα αποτελέσματα (ψευδώς θετικά). Ένα επίπεδο σημαντικότητας, α , αποφασίζεται. Εάν $p\text{-value} < \alpha$, τότε η διαφορά μεταξύ των μετρήσεων είναι στατιστικά σημαντική. Συχνά, το α ορίζεται από τις τιμές 0.05, 0.01 ή 0.001.
- [4]. Εκτίμηση του μεγέθους επίδρασης (effect size), γνωστό ως \log_2 Fold Change (\log_2FC):

$$\log_2FC = \log_2 \frac{\text{μετρήσεις στη συνθήκη A (π.χ. δείγμα A)}}{\text{μετρήσεις στη συνθήκη B (π.χ. δείγμα B)}} \quad (1)$$

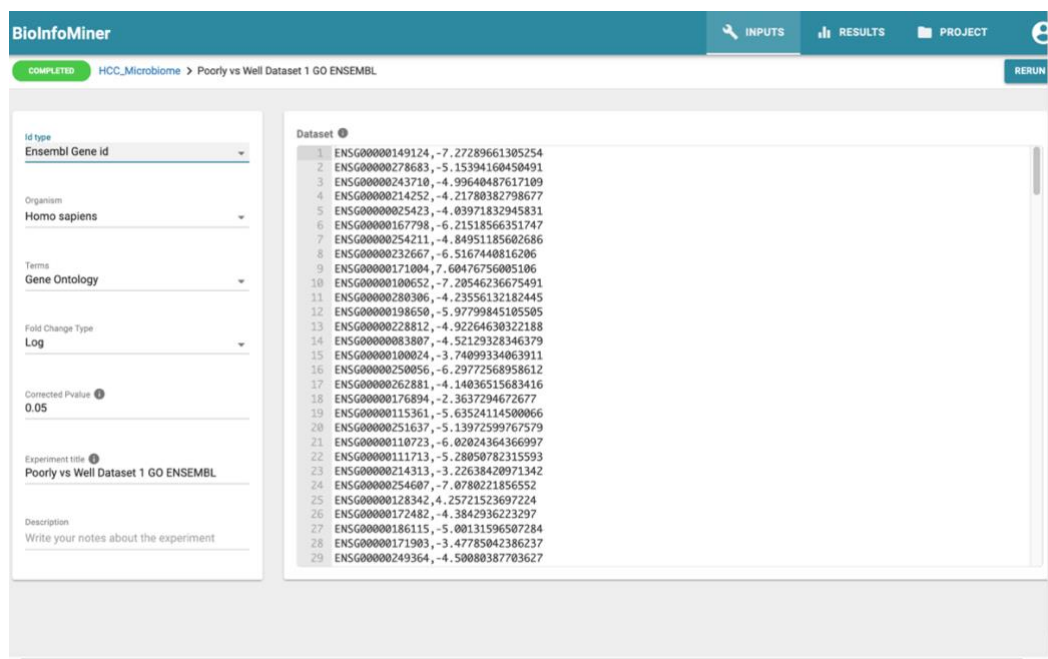
Συνήθως, το $\log_2FC = \pm 0.5$ αποτελεί το όριο της στατιστικής σημαντικότητας.

- [5]. Διόρθωση των p-value λόγω των πολλαπλών στατιστικών ελέγχων. Λόγω της πραγματοποίησης πολλαπλών στατιστικών ελέγχων ανεξάρτητων δειγμάτων στα ίδια δεδομένα, πρέπει να διορθωθεί το επίπεδο σημαντικότητας, α . Η απλούστερη μέθοδος που χρησιμοποιείται είναι η διόρθωση Bonferroni, που προσαρμόζει το α ανάλογα με τον αριθμό των ελέγχων που έλαβαν χώρα (δηλαδή των αριθμό των γονιδίων στην περίπτωση διαφορικής έκφρασης γονιδίων). Έτσι, το διορθωμένο p-value (adjusted p-value) αποτελεί το επιλεγμένο, διαιρεμένο με τον συνολικό αριθμό των γονιδίων.

Η απεικόνιση των αποτελεσμάτων συνήθως πραγματοποιείται με διαγράμματα-ηφαίστεια (volcano plots). Στον άξονα y που περιγράφει την τιμή p-value, συνήθως χρησιμοποιείται η αρνητικά τιμή της, λογαριθμημένη με βάση 10. Ο άξονας x αφορά την τιμή του \log_2FC για κάθε γονίδιο.

1.4.6 Λειτουργικός Χαρακτηρισμός Γονιδιακής Έκφρασης

Ο λειτουργικός χαρακτηρισμός των διαφορικά εκφρασμένων ανθρώπινων γονιδίων πραγματοποιήθηκε με το σετ εργαλείων BioInfoMiner (Koutsandreas et al., 2016; Valavanis et al., 2015). Το BioInfoMiner περιλαμβάνει εργαλεία για το λειτουργικό χαρακτηρισμό γονιδίων ή πρωτεϊνών σε γνωστά μεταβολικά μονοπάτια (Reactome pathways) και βιολογικές διεργασίες με όρους Γονιδιακής Οντολογίας (GO), καθώς και εργαλεία για την πρόβλεψη των λειτουργιών τους με βάση τις αλληλουχίες ή άλλα χαρακτηριστικά τους.



The screenshot displays the BioInfoMiner web application interface. The top navigation bar includes 'INPUTS', 'RESULTS', and 'PROJECT' tabs. The main content area is divided into a left sidebar and a central panel. The sidebar contains several dropdown menus: 'Id type' (set to 'Ensembl Gene id'), 'Organism' (set to 'Homo sapiens'), 'Terms' (set to 'Gene Ontology'), 'Fold Change Type' (set to 'Log'), 'Corrected Pvalue' (set to '0.05'), and 'Experiment title' (set to 'Poorly vs Well Dataset 1 GO ENSEMBL'). The central panel, titled 'Dataset', displays a list of 29 genes with their corresponding Ensembl IDs and log2 fold change values. The list is as follows:

Gene ID	log2FC
ENSG00000149124	-7.27289661385254
ENSG00000278683	-5.15394160450491
ENSG00000243710	-4.99640487617109
ENSG00000214252	-4.21780382798677
ENSG0000025423	-4.03971832945831
ENSG00000167798	-6.21518566351747
ENSG00000254211	-4.84951185602686
ENSG00000232667	-6.5167440816206
ENSG00000171004	7.60476756005106
ENSG00000180652	-7.20546236675491
ENSG00000280306	-4.23556132182445
ENSG00000198650	-5.977998451805505
ENSG00000228812	-4.92264630322188
ENSG00000083807	-4.52129328346379
ENSG00000180024	-3.74099334063911
ENSG00000250056	-6.29772568958612
ENSG00000262881	-4.14036515683416
ENSG00000176894	-2.3637294672677
ENSG00000115361	-5.63524114580066
ENSG00000251637	-5.13972599767579
ENSG00000110723	-6.02024364366997
ENSG00000111713	-5.28050782315593
ENSG00000214313	-3.22638420971342
ENSG00000254607	-7.0780221856552
ENSG00000128342	4.25721523697224
ENSG00000172482	-4.3842936223297
ENSG00000186115	-5.00131596507284
ENSG00000171903	-3.47785042386237
ENSG00000249364	-4.50080387703627

Εικόνα 13. Παράδειγμα του λογισμικού BioInfoMiner και των δεδομένων εισόδου για την ανάλυση.

Για τη λειτουργία του BioInfoMiner, εισάγονται ορισμένα δεδομένα εισόδου, όπως φαίνεται και στην Εικόνα 13:

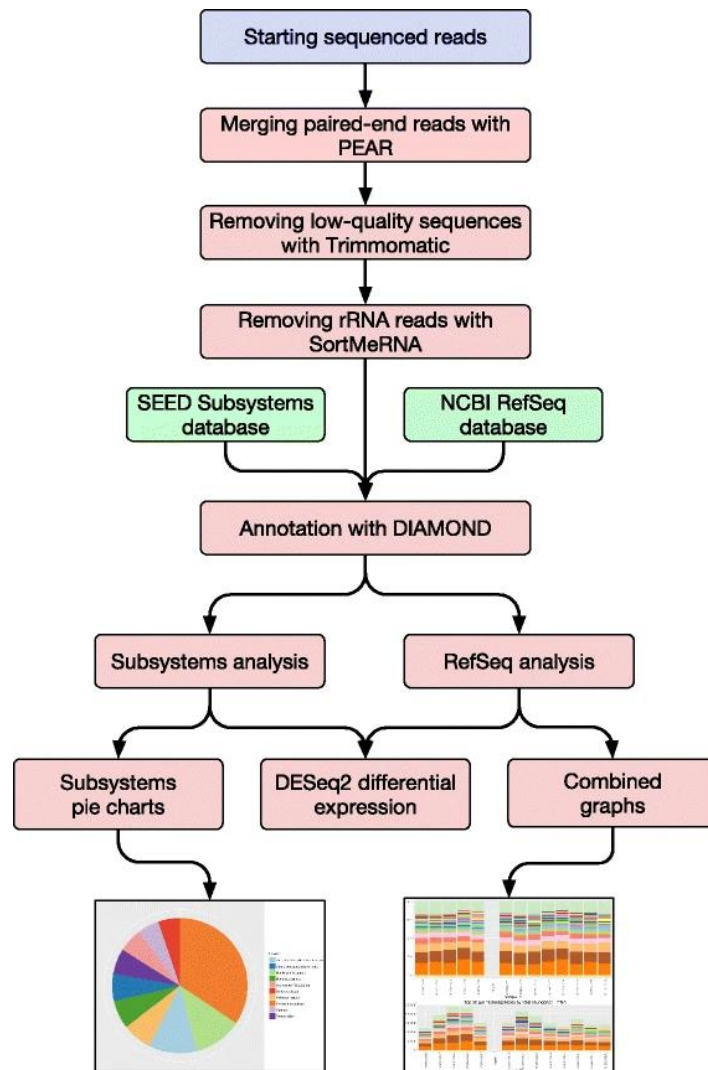
- Τα δεδομένα της διαφορικής γονιδιακής έκφρασης (Dataset): Παρέχεται η λίστα των γονιδίων από την ανάλυση γονιδιακής έκφρασης που πραγματοποιήθηκε με το εργαλείο edgeR, καθώς και οι αντίστοιχες τιμές τους σε \log_2FC ,
- Ο τύπος αναγνωριστικού γονιδίου (ID type),
- Ο οργανισμός που προέκυψαν τα δεδομένα (Organism),
- Οι όροι (Terms) των βιολογικών λειτουργιών ενδιαφέροντος,
- Η μορφή του βαθμού μεταβολής της έκφρασης (Fold Change),
- Το όριο της διορθωμένης p-value σύμφωνα με το οποίο ορίστηκε η διαφορική έκφραση,
- Σχόλια του χειριστή για την περιγραφή της σύγκρισης που πραγματοποιείται.

1.4.7 Μετα-μεταγραφωμική ανάλυση – SAMSA2

Πέρα από την ανάλυση του λειτουργικού χαρακτηρισμού των αναγνωσμάτων σε ανθρώπινα γονίδια, η μελέτη του ρόλου του μικροβιώματος σε RNA-Seq δεδομένα έχει δημιουργήσει την ανάγκη για την κατασκευή εργαλείων που να έχουν τη δυνατότητα ευθυγράμμισης αρχικών δεδομένων μορφής FASTQ σε βακτηριακά γονιδιώματα. Η μελέτη αυτή είναι απλή, όταν η μελέτη επικεντρώνεται σε ένα συγκεκριμένο βακτηριακό είδος, και ακολουθεί τον ίδιο σχεδιασμό διαδικασίας με αυτόν που περιεγράφηκε για το ανθρώπινο γονιδίωμα, δηλαδή την ευθυγράμμιση των αναγνωσμάτων σε ένα γονιδίωμα αναφοράς. Ωστόσο, η ανίχνευση και ο χαρακτηρισμός ολόκληρων μικροβιακών πληθυσμών εντός δείγματος είναι μία λιγότερο απλή διαδικασία και περιλαμβάνει τη χρήση μίας βάσης δεδομένων με τα γονιδιώματα, ή αποσπασματικά γονίδια, των πιο αντιπροσωπευτικών βακτηρίων. Έτσι, με τη χρήση μετα-μεταγραφωμικών εργαλείων δίνονται πληροφορίες για την αφθονία βακτηρίων στο δείγμα (όπως και στη μεταγενωμική), αλλά και ποια από αυτά είναι ενεργά και ποιες οι βιολογικές τους λειτουργίες (Westreich et al., 2018).

Το SAMSA2 χρησιμοποιεί πολλά αναγνωρισμένα εργαλεία, συμπεριλαμβανομένων των PEAR (J. Zhang et al., 2014) και Trimmomatic για προεπεξεργασία και περικοπή ανάγνωσης (Bolger et al., 2014), SortMeRNA για φιλτράρισμα ριβοσωμικών RNA (Kopylova et al., 2012) και DIAMOND για την ευθυγράμμιση (Buchfink et al., 2015).

Μια επισκόπηση της ροής δεδομένων μέσω της διοχέτευσης SAMSA2 βρίσκεται στην Εικόνα 14. Το πρώτο βήμα στη διεργασία του SAMSA2 είναι η συγχώνευση αρχείων PE, εάν χρησιμοποιήθηκε αυτός ο τύπος αλληλουχίας, ο οποίος και προτείνεται. Το PEAR είναι μια γρήγορη και αποτελεσματική συγχώνευση PE αναγνωσμάτων (J. Zhang et al., 2014). Όταν χρησιμοποιείται σε δύο αρχεία PE, δημιουργείται ένα συγχωνευμένο αρχείο εξόδου που περιέχει όλα τα αναγνώσματα, καθώς και δύο αρχεία "notCombined" που περιέχουν τα μη συναρμολογημένα αναγνώσματα, των δύο άκρων. Γενικά, μόνο τα συγχωνευμένα αναγνώσματα χρησιμοποιούνται στην υπόλοιπη διεργασία, παρόλο που εάν δεν μπορεί να συγχωνευθεί μεγάλος αριθμός αναγνωσμάτων, είναι σκόπιμο να χρησιμοποιηθούν τόσο το συγχωνευμένο, όσο και το προς τα εμπρός ανάγνωσμα "notCombined" για να διασφαλιστεί ότι λαμβάνεται ένας επαρκής αριθμός ευθυγραμμίσεων.



Εικόνα 14. Μετα-μεταγραφωμική ανάλυση με το εργαλείο SAMSA2.

Μετά τη συγχώνευση των αναγνωσμάτων PE (εάν έχει επιλεγεί), οι αλληλουχίες χαμηλής ποιότητας ή/και οι αλληλουχίες των προσαρμογέων αφαιρούνται. Το Trimmomatic (Bolger et al., 2014) παράγει ένα καθαρισμένο αρχείο εξόδου που είναι έτοιμο για το επόμενο βήμα στην προεπεξεργασία.

Τέλος, αν και η πλειονότητα των ριβοσωμικών αλληλουχιών αφαιρούνται από το μετα-μεταγράφομα στο προηγούμενο βήμα, μέσω ριβοεξάλειψης (ribodepletion), το SortMeRNA παρέχει μια ακόμη ψηφιακή σάρωση, διασφαλίζοντας ότι τα ριβοσωμικά αναγνώσματα αφαιρούνται, πριν από το λειτουργικό χαρακτηρισμό (Korylova et al., 2012). Η αφαίρεση των ριβοσωμικών αναγνωσμάτων σε αυτό το βήμα μειώνει τον συνολικό αριθμό των αλληλουχιών που θα ευθυγραμμιστούν, αυξάνοντας την ταχύτητα της διεργασίας και μειώνοντας τα μεγέθη αρχείων. Το SortMeRNA χρησιμοποιεί πολλές βάσεις δεδομένων rRNA αναφοράς, τόσο για βακτηριακές, όσο και για ευκαρυωτικές αλληλουχίες ριβοσώματος. Το SortMeRNA εξάγει

ταυτοποιημένες ριβοσωμικές αλληλουχίες και αλληλουχίες που δεν ταιριάζουν με γνωστά ριβοσώματα. Αυτές οι ακολουθίες, που δεν ταυτοποιήθηκαν ως ριβοσωμικές, μεταφέρονται στο βήμα ευθυγράμμισης στο SAMSA2.

Το SAMSA2 χρησιμοποιεί το DIAMOND (Buchfink et al., 2015), έναν ταχύ αλγόριθμο εκτέλεσης ευθυγραμμίσεων σε μία ή περισσότερες επιλεγμένες βάσεις δεδομένων αναφοράς. Το DIAMOND έχει σχεδιαστεί ειδικά για το λειτουργικό χαρακτηρισμό μεγάλου αριθμού αναγνωσμάτων, σε μια βάση δεδομένων αναφοράς ταυτόχρονα, με ταχύτητες έως και 10.000 φορές μεγαλύτερες από παραδοσιακές μεθόδους, όπως το BLAST.

Για να μπορέσουν οι αλληλουχίες εισόδου να ευθυγραμμιστούν από το DIAMOND, πρέπει να ενταχθούν στη διεργασία βάσεις δεδομένων αναφοράς, όπως η RefSeq του NCBI (Tatusova et al., 2014) και η ιεραρχική βάση δεδομένων SEED Subsystems (Overbeek et al., 2014) για την ευθυγράμμιση των αναγνωσμάτων σε λειτουργικές δραστηριότητες. Ακόμη, είναι δυνατή η προσθήκη άλλων βάσεων δεδομένων, όπως η βάση δεδομένων Carbohydrate Active Enzyme (CAZy) (Cantarel et al., 2009). Μετά τη μετατροπή μιας βάσης δεδομένων σε ένα δυαδικό αρχείο με δυνατότητα αναζήτησης με DIAMOND, τα αρχεία πειραματικής ακολουθίας σχολιάζονται έναντι αυτής της αναφοράς. Όταν η βάση δεδομένων μετατραπεί σε δυαδικό αρχείο με συμβατότητα αναζήτησης στο DIAMOND, οι πειραματικές αλληλουχίες εισόδου είναι έτοιμες προς ευθυγράμμιση.

Το SAMSA2 δίνει πληροφορίες ξεκινώντας από την έξοδο του DIAMOND, δηλαδή ευθυγραμμισμένες καταμετρήσεις αναγνωσμάτων, καταλήγοντας σε στοιχεία για τη σχετική αφθονία ενός βακτηριακού είδους ή ενός μεταβολικού μονοπατιού (relative abundance), την ποικιλομορφία των ειδών μεταξύ και μέσα στα δείγματα (άλφα και βήτα ποικιλομορφία, αντίστοιχα), καθώς και τη διαφορική αφθονία (differential abundance) των ειδών στα δείγματα (Westreich et al., 2018).

1.4.8 Ανάλυση Μικροβιώματος

Η αξιολόγηση της ταξινομικής ποικιλομορφίας του μικροβιώματος πραγματοποιείται χρησιμοποιώντας διάφορα μέτρα. Η άλφα ποικιλομορφία είναι ένας όρος που περιγράφει την ποικιλομορφία εντός του κάθε δείγματος, και συνήθως λαμβάνει υπόψη τα διαφορετικά παρατηρούμενα είδη, με τη χρήση ποικίλων δεικτών. Αρχικά, ο πιο απλός δείκτης της άλφα

ποικιλομορφίας είναι τα παρατηρούμενα είδη, όπου καταμετρούνται τα διαφορετικά είδη σε κάθε δείγμα. Ο δείκτης Shannon εκτιμά την ομοιομορφία (evenness) και την αφθονία (richness) των τιμών του συνόλου των ειδών μέσα στο δείγμα. Δηλαδή, στην περίπτωση που απαντάται μόνο ένα είδος στο δείγμα, η τιμή του δείκτη Shannon είναι μηδενική. Ο δείκτης Simpson εκτιμάει την πιθανότητα δύο στοιχείων (μικροβιακών ειδών ή αναγνωσμάτων), επιλεγμένων τυχαία, από ένα δείγμα να ταυτίζονται ως προς το είδος. Επομένως, αυτό το μέτρο λαμβάνει τιμές από 0 έως 1.

Η ποικιλομορφία βήτα είναι ένας όρος που εκφράζει τις διαφορές μεταξύ δειγμάτων. Στη μελέτη του μικροβιώματος χρησιμοποιείται για να εντοπίσει τις διαφορές μεταξύ δύο ομάδων, όπως τις ομάδες ασθένειας και ελέγχου (control). Η βήτα ποικιλομορφία μετράται ως η μέση απόσταση (ή ανομοιότητα) από μια μεμονωμένη μονάδα στο κέντρο της ομάδας, χρησιμοποιώντας ένα κατάλληλο μέτρο ανομοιότητας. Το κέντρο πρέπει να οριστεί στον κύριο χώρο συντεταγμένων του επιλεγμένου μέτρου ανομοιότητας. Αυτή η έννοια της ποικιλομορφίας βήτα είναι αρκετά ευέλικτη, επειδή μπορεί να βασίζεται σε οποιοδήποτε επιλεγμένο μέτρο ανομοιότητας με βιολογική ή οικολογική σημασία. Έχει επίσης το πρόσθετο πλεονέκτημα σε σχέση με το αρχικό μέτρο του Whittaker (Whittaker, 1952), καθώς μπορεί να χρησιμοποιηθεί για τον έλεγχο διαφορών στην ποικιλομορφία βήτα μεταξύ περιοχών ή ομάδων, μέσω μιας πολυπαραγοντικής δοκιμής για ομοιογένεια στις διασπορές (Anderson et al., 2006).

Το τεστ για την ομοιογένεια των πολυμεταβλητών διασπορών είναι ένα πολυμεταβλητό ανάλογο του τεστ του Levene και μπορεί να βασίζεται σε οποιοδήποτε μέτρο ανομοιότητας επιλογής. Ουσιαστικά, υπολογίζεται μια στατιστική τιμή F για να συγκρίνει τη μέση απόσταση των μονάδων παρατήρησης με το κέντρο της ομάδας τους, που ορίζεται στο χώρο που προσδιορίζεται από το επιλεγμένο μέτρο ανομοιότητας. Στη συνέχεια λαμβάνεται μια τιμή p (p-value) μεταθέτοντας τα υπολείμματα ελαχίστων τετραγώνων (Anderson et al., 2006).

Μια σημαντική πολυπλοκότητα αφορά τον υπολογισμό των κεντροειδών για μέτρα που δεν είναι ευκλείδεια. Δύο από τις μεθόδους που μπορούν να εξετάσουν τη μεταβλητότητα αυτή είναι η UniFrac, η Bray – Curtis και η Jaccard. Η πρώτη λαμβάνει υπόψη το φυλογενετικό δέντρο που δομείται κατόπιν ανάλυσης στις αλληλουχίες γονιδίων 16S rRNA των δειγμάτων. Η δεύτερη βασίζεται στη μέθοδο Ανάλυσης Κύριων Συνιστωσών (Principal Component Analysis, PCA), η οποία δε λαμβάνει υπόψη τη δομή του φυλογενετικού δέντρου. Η PCA είναι

μία ανάλυση που ορίζεται ως ο ορθογώνιος γραμμικός μετασχηματισμός μετατροπής δεδομένων σε ένα νέο σύστημα συντεταγμένων, όπου η μεγαλύτερη διακύμανση βρίσκεται στην πρώτη συντεταγμένη (PC1). Αποτελεί μια μέθοδο ελαχιστοποίησης συνιστωσών σε δύο συντεταγμένες, οι οποίες προκύπτουν από το αποτέλεσμα ενός γραμμικού συνδυασμού της σχετικής αφθονίας των δειγμάτων (Walters & Martiny, 2020).

Επακόλουθα, η ποικιλομορφία των μικροβιακών ειδών εξετάζεται περαιτέρω, στο επίπεδο της διαφορικής αφθονίας. Όπως και στην ανάλυση της διαφορικής έκφρασης γονιδίων, έτσι και στην ανάλυση διαφορικής αφθονίας των βακτηριακών ειδών, τα εργαλεία edgeR και DESeq2, που είναι διαθέσιμα σε ορισμένα πακέτα του Bioconductor της R, δίνουν αποδοτικά αποτελέσματα μέσω στατιστικών ελέγχων.

Το τελευταίο βήμα στις αναλύσεις μεταγραφωμικών και μετα-μεταγραφωμικών δεδομένων είναι ο λειτουργικός χαρακτηρισμός των διαφορικά εκφρασμένων γονιδίων σε βιολογικές λειτουργίες. Ο μοριακός χαρακτηρισμός αυτός εφαρμόζεται και στα ανθρώπινα γονίδια και στα βακτηριακά, στην περίπτωση που η ανάλυση τα εμπλέκει και τα δύο. Η κυρίαρχη προσέγγιση του προβλήματος είναι η σύγκριση της λίστας των διαφορικά εκφρασμένων γονιδίων και η ανάλυση ομάδων γονιδίων (gene set enrichment), ομαδοποιώντας τα γονίδια ανάλογα με την κυριότερη βιολογική λειτουργία που εμπíπτουν (Zito et al., 2021). Για τη χρησιμοποίηση τέτοιων μεθόδων είναι απαραίτητος ο λειτουργικός χαρακτηρισμός των δεδομένων για το προς μελέτη μεταγράφομα. Πολλά εργαλεία που χρησιμοποιούνται για την gene set enrichment ανάλυση σε μεταγραφωμικά και μεταμεταγραφωμικά δεδομένα είναι το Enrichr (E. Y. Chen et al., 2013), το GSEA (Hänzelmann et al., 2013), και το AllEnricher (D. Zhang et al., 2020).

2 Σκοπός της μελέτης

Οι στόχοι της ανάλυσης της παρούσας Διπλωματικής Εργασίας συνοψίζονται στα παρακάτω:

- Η αξιολόγηση της προσέγγισης βιοπληροφορικής ανάλυσης διαθέσιμων μεταγραφωμικών δεδομένων από παρακαρκινικά και καρκινικά δείγματα ανθρώπινου HCC, για τη μελέτη του ενδο-ογκικού μικροβιώματος
- Η διερεύνηση των ταξινομικών και λειτουργικών χαρακτηριστικών του ενδο-ογκικού μικροβιώματος σε σύγκριση με τους φυσιολογικούς, παρακαρκινικούς ιστούς ήπατος
- Η εξέταση της πιθανής συσχέτισης του ενδο-ογκικού μικροβιώματος με την ετερογένεια του HCC και η ανάδειξη εμπλεκόμενων μικροβιακών βιοδεικτών και λειτουργιών

3 Μεθοδολογία

Η αρχική υπολογιστική διαδικασία της προ-επεξεργασίας των αρχικών δεδομένων πραγματοποιήθηκε σε περιβάλλον Linux/Unix, ενώ η μετέπειτα ανάλυση, στη γλώσσα προγραμματισμού R, με τη χρήση προγραμματιστικών πακέτων του αποθετηρίου Bioconductor (<https://bioconductor.org/install/#install-bioconductor-packages>).

3.1 Λήψη RNA-Seq Δεδομένων

Τα αρχικά, μη επεξεργασμένα δεδομένα που χρησιμοποιήθηκαν, ελήφθησαν από τη βάση δεδομένων Gene Expression Omnibus (GEO) με χαρακτηριστικό αριθμό GSE105130, της εργασίας PRJNA414787 (Jin et al., 2019). Τα δεδομένα αυτά αφορούν 25 ασθενείς HCC, και συγκεκριμένα, τα RNA-Seq αποτελέσματα από 25 καρκινικούς (Tumor) και 27 φυσιολογικούς παρακαρκινικούς (Normal), ηπατικούς ιστούς. Τα δείγματα υποβλήθηκαν σε αλληλούχιση RNA με το όργανο Illumina HiSeq 2000 και είναι αναγνώσματα PE με μέσο μήκος 180bp, της μορφής FASTQ.

3.2 Προεπεξεργασία Αρχικών Δεδομένων

Το πρώτο στάδιο της επεξεργασίας των μη επεξεργασμένων δεδομένων RNA-Seq ήταν ο ποιοτικός έλεγχος, το οποίο πραγματοποιήθηκε με το εργαλείο FastQC. Η είσοδος του εργαλείου FastQC ήταν αρχεία των αρχικά μη επεξεργασμένων δεδομένων RNA-Seq, σε μορφή FASTQ. Η ανάλυση στο FastQC εκτελείται από μια σειρά αναλυτικών εφαρμογών.

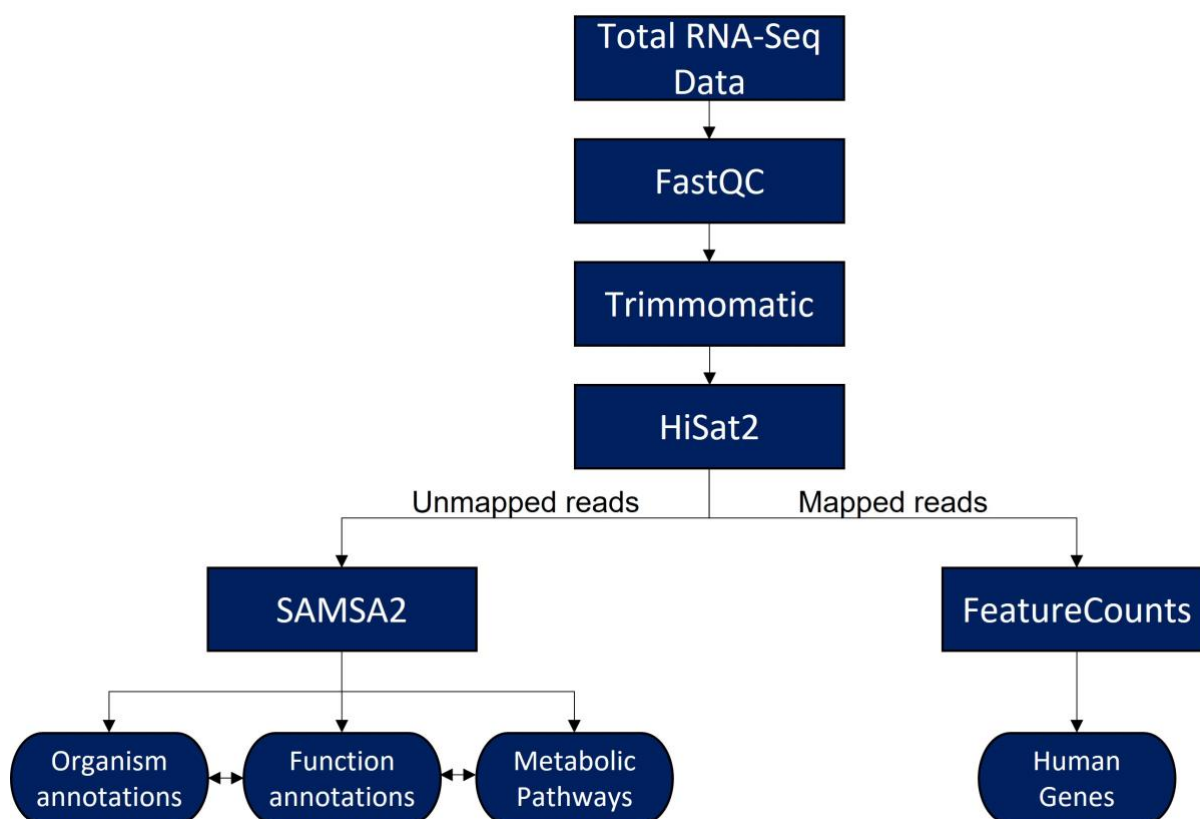
Έπειτα, εισήλθαν στο εργαλείο Trimmomatic, με σκοπό την αποκοπή αλληλουχιών χαμηλής ποιότητας και την αφαίρεση αλληλουχιών των προσαρμογών. Το εργαλείο μπορεί να λειτουργήσει για SE και PE αλληλουχίες. Η PE λειτουργία διατηρεί την αντιστοιχία των ζευγών των αναγνωσμάτων, και επίσης χρησιμοποιεί τις πρόσθετες πληροφορίες που περιέχονται στις κατά ζεύγη αλληλουχίες για την καλύτερη εύρεση θραυσμάτων προσαρμογών ή εκκινητών PCR που εισάγονται κατά τη διαδικασία προετοιμασίας της βιβλιοθήκης αναγνωσμάτων. Αφού τα αρχικά δεδομένα που επιλέχθηκαν είναι PE αναγνώσματα, η λειτουργία του Trimmomatic ήταν 'PE'. Οι υπόλοιπες παράμετροι που ορίστηκαν κατά τη λειτουργία του εργαλείου ήταν οι εξής (Bolger et al., 2014):

- ILLUMINACLIP: καταχωρούνται οι αλληλουχίες προσαρμογέων, σε FASTA μορφή, προκειμένου να αποκοπούν,
- LEADING: κόβει τις βάσεις από την αρχή μιας αλληλουχίας, αν είναι κάτω από το όριο ποιότητας, το οποίο ορίστηκε 3,
- TRAILING: κόβει τις βάσεις από το τέλος μιας αλληλουχίας, αν είναι κάτω από το όριο ποιότητας, το οποίο ορίστηκε 3,
- SLIDINGWINDOW: εκτελεί προσέγγιση οριζόντιας ολίσθησης με οριοθετημένο τμήμα. Η σάρωση ξεκινάει από το 5' άκρο και αποκόβει την αλληλουχία όταν η μέση ποιότητα είναι μικρότερη από ένα ορισμένο κατώτατο όριο. Η παράμετρος αυτή ορίστηκε ως 4:15, δηλαδή για τις 15 πρώτες βάσεις της αλληλουχίας, ξεκινώντας από το 5' άκρο, η μέση ποιότητα να υπερβαίνει την ποιότητα 4,
- MINLEN: το ελάχιστο μήκος της αλληλουχίας προκειμένου να μην απορριφθεί. Ορίστηκε η τιμή των 36bp.

Το σύστημα αξιολόγησης ποιότητας των αλληλουχιών που επιλέχθηκε είναι το 'PHRED33'.

3.3 Ευθυγράμμιση στο Ανθρώπινο Γονιδίωμα (Human Genome Alignment)

Κατόπιν της αξιολόγησης ποιότητας και αποκοπής των αλληλουχιών, τα προεπεξεργασμένα δεδομένα ευθυγραμμίστηκαν με το ανθρώπινο γονιδίωμα GRCh38.p13. Οι αλληλουχίες που δε ευθυγραμμίστηκαν με αυτό, θεωρήθηκαν μη ανθρώπινες, και υποβλήθηκαν μετέπειτα σε μετα-μεταγραφωμική ανάλυση, προκειμένου μέρος αυτών να ευθυγραμμιστεί με βακτηριακές αλληλουχίες. Η ροή της επεξεργασίας των δεδομένων φαίνεται στην Εικόνα 15.



Εικόνα 15. Κύριο διάγραμμα ροής περιγραφής της υπολογιστικής διαδικασίας. Αρχικά, η προεπεξεργασία των δειγμάτων περιλαμβάνει τον ποιοτικό έλεγχο και την αποκοπή των αναγνωσμάτων. Τα αναγνώσματα ευθυγραμμίζονται με το ανθρώπινο γονιδίωμα και διαχωρίζονται σε ανθρώπινα και μη ανθρώπινα αναγνώσματα. Τα ανθρώπινα υφίστανται λειτουργικό χαρακτηρισμό αφού καταμετρηθούν, ενώ τα μη ανθρώπινα αναλύονται μετα-μεταγραφωμικά για την εύρεση διάφορων ποσοτικών χαρακτηριστικών του μικροβιώματος.

Το εργαλείο που επιλέχθηκε να χρησιμοποιηθεί είναι το HISAT2, το οποίο προτείνεται για ευθυγραμμίσεις σε ένα γονιδίωμα αναφοράς (D. Kim et al., 2019). Το εργαλείο δέχθηκε ως είσοδο τα δύο PE αρχεία FASTQ για κάθε δείγμα, με δύο εξόδους: τα ευθυγραμμισμένα ανθρώπινα γονίδια (mapped reads) σε μορφή BAM, και τα υπόλοιπα που δεν ευθυγραμμίστηκαν με το ανθρώπινο γονιδίωμα (unmapped reads), σε μορφή FASTQ.

Η μορφή BAM αποτελεί μία συμπιεσμένη δυαδική έκδοση των αρχείων SAM, που χρησιμεύει στην καταγραφή ευθυγραμμισμένων αλληλουχιών. Η δομή ενός αρχείου BAM είναι μία επικεφαλίδα, που περιέχει πληροφορίες για ολόκληρο το αρχείο, όπως όνομα δείγματος, μήκος δείγματος και μέθοδο ευθυγράμμισης, και την ενότητα «ευθυγραμμίσεις» (alignments), που περιέχει το όνομα του αναγνώσματος, την αλληλουχία, την ποιότητά της, πληροφορίες για την ευθυγράμμιση και προσαρμοσμένες ετικέτες. Το όνομα αναγνώσματος περιλαμβάνει το χρωμόσωμα, τη συντεταγμένη έναρξης, την ποιότητα ευθυγράμμισης και τη συμβολοσειρά του περιγραφέα ευθυγράμμισης.

3.4 Μεταγραφωμική Ανάλυση

3.4.1 Ποσοτικοποίηση (*Quantification*)

Η ποσοτικοποίηση των ανθρώπινων αναγνωσμάτων πραγματοποιήθηκε με το εργαλείο featureCounts. Το featureCounts (Liao et al., 2014) είναι ένα εργαλείο για την καταμέτρηση των αναγνωσμάτων που ευθυγραμμίζονται με ένα σύνολο γονιδίων ή γονιδιωματικών χαρακτηριστικών. Χρησιμοποιείται συνήθως στην ανάλυση RNA-seq για να ποσοτικοποιήσει τα επίπεδα έκφρασης των γονιδίων σε ένα δείγμα μετρώντας τον αριθμό των αναγνωσμάτων που αντιστοιχούν σε κάθε γονίδιο. Αυτές οι πληροφορίες χρησιμοποιούνται, στη συνέχεια, για τη σύγκριση της γονιδιακής έκφρασης μεταξύ διαφορετικών δειγμάτων ή συνθηκών, ή για τον εντοπισμό διαφορετικά εκφρασμένων γονιδίων.

Το featureCounts χρησιμοποιείται ευρέως επειδή είναι γρήγορο, ακριβές και εύκολο στη χρήση. Μπορεί να χειριστεί πολλές μορφές αρχείων, συμπεριλαμβανομένων των BAM και SAM, και μπορεί να λειτουργήσει με αναγνώσματα αλληλούχισης PE και SE. Διαθέτει, επίσης, επιλογές για φιλτράρισμα και σύνοψη των καταμετρήσεων, και μπορεί να παράξει τα αποτελέσματα σε διάφορες μορφές για περαιτέρω ανάλυση.

Το αρχείο GTF (Gene Transfer Format) χρησιμοποιείται στο featureCounts, για να παρέχει το χαρακτηρισμό για το γονιδίωμα που αναλύεται. Καθορίζει τη θέση και άλλες πληροφορίες σχετικά με τα γονίδια και άλλα γονιδιωματικά χαρακτηριστικά στο γονιδίωμα, όπως εξόνια, ιντρόνια και διαγονιδιακές περιοχές.

Όταν εκτελείται το featureCounts, το αρχείο GTF εισάγεται μαζί με τις ευθυγραμμισμένες αναγνώσεις σε ένα αρχείο BAM ή SAM. Το featureCounts χρησιμοποιεί, στη συνέχεια, τις πληροφορίες στο αρχείο GTF, για να προσδιορίσει ποια αναγνώσματα ευθυγραμμίζονται με ποια ανθρώπινα γονίδια, και να μετρήσει τον αριθμό των αναγνωσμάτων που αντιστοιχούν σε κάθε γονίδιο.

Δύο αλληλουχίες PE που έχουν ευθυγραμμιστεί σε διαφορετικό γονίδιο ή χρωμόσωμα, και αναγνώσματα με βαθμό ευθυγράμμισης (mapping quality score) χαμηλότερο του 10 δεν λήφθηκαν υπόψη στην ποσοτικοποίηση.

Το αποτέλεσμα εξόδου του εργαλείου featureCounts ήταν ένας πίνακας που παραθέτει τον αριθμό των αναγνώσεων που καταμετρήθηκαν για κάθε γονίδιο. Ο πίνακας περιλαμβάνει τις ακόλουθες στήλες:

- Αναγνωριστικό γονιδίου (Gene ID): Το μοναδικό αναγνωριστικό για το γονίδιο ή το γονιδιωματικό χαρακτηριστικό,
- Χρωμόσωμα (Chromosome): Το χρωμόσωμα ή το ικρίωμα στο οποίο βρίσκεται το γονίδιο ή το χαρακτηριστικό,
- Έναρξη (Start): Η αρχική θέση του γονιδίου ή του χαρακτηριστικού,
- Τέλος (End): Η τελική θέση του γονιδίου ή του χαρακτηριστικού,
- Κλώνος (Strand): Ο κλώνος στον οποίο βρίσκεται το γονίδιο ή το χαρακτηριστικό (+ για θετικό κλώνο, - για αρνητικό κλώνο),
- Μήκος (Length): Το μήκος του γονιδίου ή του χαρακτηριστικού σε ζεύγη βάσεων,
- Καταμέτρηση (Count): Ο αριθμός των αναγνώσεων που μετρήθηκαν για το γονίδιο ή το χαρακτηριστικό,

Προαιρετικά, ο πίνακας μπορεί επίσης να περιλαμβάνει στήλες για το όνομα του δείγματος, την ομάδα και άλλες πληροφορίες (Liao et al., 2014). Η έξοδος των featureCounts υπέστη περαιτέρω επεξεργασία και αναλύθηκε χρησιμοποιώντας εργαλεία, όπως το edgeR, για τον εντοπισμό γονιδίων που εκφράζονται διαφορετικά μεταξύ των διαφορετικών δειγμάτων.

3.4.2 Ανάλυση σε Κύριες Συνιστώσες (PCA)

Από τον πίνακα καταμετρήσεων που δίνει ως έξοδο το εργαλείο featureCounts, λαμβάνονται υπόψη τα 1000 γονίδια με τη μεγαλύτερη διάμεση απόλυτη απόκλιση (median absolute deviation). Το διάγραμμα διασποράς PCA κατασκευάστηκε με σκοπό την ποιοτική ανάδειξη του γονιδιακού προφίλ των δειγμάτων, με είσοδο τον κανονικοποιημένο πίνακα καταμετρήσεων, με τη μέθοδο TMM. Οι υπολογιστικές διαδικασίες στη γλώσσα προγραμματισμού R, με τη χρήση πακέτων όπως το 'stats' και το 'vegan' (Oksanen, 2010), εμπεριέχουν τον υπολογισμό ιδιοτιμών και ιδιοδιανυσμάτων των πινάκων των καταμετρήσεων των αναγνωσμένων γονιδίων, και, τελικά, την προβολή των δεδομένων σε ένα χώρο 2 διαστάσεων.

3.4.3 Ανάλυση Διαφορικής Έκφρασης Ανθρώπινων Γονιδίων

Η ανάλυση της διαφορικής γονιδιακής έκφρασης εκτελέστηκε με το εργαλείο edgeR. Εισάγοντας τον πίνακα καταμετρήσεων των αναγνωσμάτων για κάθε γονίδιο των δειγμάτων, χρησιμοποιήθηκε η συνάρτηση ‘DGEList’. Η συνάρτηση ‘calcNormFactors’ χρησιμοποιήθηκε για να κανονικοποιηθούν τα δεδομένα (με τη μέθοδο κανονικοποίησης TMM), και η συνάρτηση εκτίμησης ‘estimateDisp’ προκειμένου να εκτιμηθεί η διασπορά. Επιπλέον, η ανάλυση της διαφορικής έκφρασης πραγματοποιήθηκε για γονίδια που απαντώνται τουλάχιστον στο 40% των δειγμάτων, με τουλάχιστον 10 καταμετρήσεις στο κάθε δείγμα, με τη χρήση της συνάρτησης ‘filterByExpr’. Χρησιμοποιήθηκε η συνάρτηση ‘treat’ (McCarthy & Smyth, 2009), ώστε να θεωρούνται στατιστικά σημαντικά διαφορικά εκφρασμένα γονίδια των οποίων οι διαφορές σε όλα τα δείγματα να έχουν $|\log_2FC| > 1.0$. Τα όρια που τέθηκαν για την ανάλυση της διαφορικής έκφρασης στο edgeR είναι η διορθωμένη (adjusted) p-value < 0.05 .

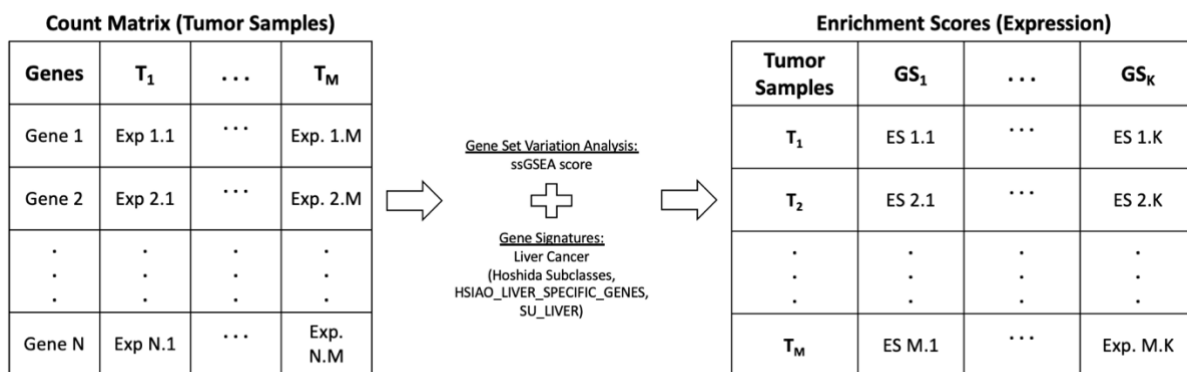
Η απεικόνιση της διαφορικής γονιδιακής έκφρασης των καρκινικών έναντι των φυσιολογικών δειγμάτων (Tumor vs Normal) πραγματοποιήθηκε με volcano plot, με τη χρήση του πακέτου ‘ggplot2’ στην R, όπου η κατακόρυφη γραμμή $x=0$ διαχωρίζει τα υπερ-εκφρασμένα (up-regulated) από τα υπο-εκφρασμένα (down-regulated) γονίδια. Δηλαδή, τα γονίδια που απαντώνται δεξιά είναι υπερ-εκφρασμένα στα Tumor δείγματα, ενώ στα αριστερά, υπο-εκφρασμένα.

3.4.4 Ομαδοποίηση Ασθενών βάσει Μοριακών Χαρακτηριστικών

Προκειμένου να κατηγοριοποιηθούν τα καρκινικά δείγματα σε υποομάδες, βάσει των μοριακών τους χαρακτηριστικών, δημιουργήθηκε ένας πίνακας μετρήσεων των 1000 πιο μεταβλητών γονιδίων, μόνο για τα 25 Tumor δείγματα. Όπως περιγράφεται και στην Εικόνα 16, στην ανάλυση μεταβλητότητας συνόλου γονιδίων (Gene Set Variation Analysis, GSVA), εντοπίστηκαν κοινά γονίδια ανάμεσα στα καρκινικά δείγματα και σε σύνολα γονιδίων (gene sets), προκειμένου να κατηγοριοποιηθούν τα δείγματα βάσει των μοριακών τους χαρακτηριστικών, χρησιμοποιώντας το πακέτο ‘GSVA’ στο περιβάλλον της R. Έτσι, κάθε δείγμα χαρακτηρίστηκε από ένα βαθμό εμπλουτισμού (ssGSEA enrichment score), ο οποίος δηλώνει το βαθμό στον οποίο τα γονίδια του δείγματος υπερ-εκφράζονται ή υπο-εκφράζονται σε ένα συγκεκριμένο σύνολο.

Τα σύνολα γονιδίων που χρησιμοποιήθηκαν αποτελούν ορισμένες γονιδιακές υπογραφές που χαρακτηρίζουν το ήπαρ ή το HCC. Συγκεκριμένα, αυτά είναι το HSIAO_LIVER_SPECIFIC_GENES (Hsiao et al., 2001) και το SU_LIVER (Su et al., 2002), γονιδιακές υπογραφές που χαρακτηρίζουν ένα υγιές ήπαρ, καθώς και οι γονιδιακές υπογραφές των 3 υποκατηγοριών του Hoshida για το HCC (Hoshida et al., 2009), S1, S2 και S3, οι οποίες περιεγράφηκαν στην παράγραφο 1.1.2. Τα παραπάνω σύνολα γονιδίων ελήφθησαν από τη βάση δεδομένων MSigDB (Liberzon et al., 2015; Subramanian et al., 2005).

Ο σκοπός της ανάλυσης ήταν η διερεύνηση της πιθανής κατηγοριοποίησης των δειγμάτων σε 2 ομάδες: στα δείγματα όγκων με υψηλή διαφοροποίηση (Well-differentiated tumors) ή χαμηλή διαφοροποίηση (Poorly-differentiated tumors), αντιστοίχως. Αν για παράδειγμα, ένα καρκινικό δείγμα παρουσιάζει υψηλό βαθμό εμπλουτισμού για τη γονιδιακή υπογραφή “HSIAO_LIVER_SPECIFIC_GENES” έχει περισσότερα μοριακά χαρακτηριστικά ενός υγιούς ήπατος, και, συμπερασματικά, χαρακτηρίζεται ως “well-differentiated”. Αντίστοιχα, ένα καρκινικό δείγμα με χαμηλό βαθμό εμπλουτισμού για τη γονιδιακή υπογραφή Hoshida Liver Cancer Subclass S3 -που αντιστοιχεί σε καλά διαφοροποιημένα καρκινικά κύτταρα- θα χαρακτηριστεί ως “poorly differentiated”.



Εικόνα 16. GSVA ανάλυση για τα καρκινικά δείγματα (T_i). Για τις μετρήσεις των γονιδίων (Gene j) από τον πίνακα για τα καρκινικά δείγματα (Count Matrix), σε συνδυασμό με τις γονιδιακές υπογραφές (Gene Signatures) για υγιή ήπατα και HCC, η ανάλυση κατασκευάζει έναν πίνακα με τις αξιολογήσεις των βαθμών εμπλουτισμού για κάθε γονιδιακή υπογραφή.

Η κατασκευή του πίνακα των βαθμών εμπλουτισμού των δειγμάτων για κάθε γονιδιακή υπογραφή αποτελεί την είσοδο για την ομαδοποίηση των δειγμάτων (clustering), βάσει των μοριακών χαρακτηριστικών (Εικόνα 17), η οποία πραγματοποιείται με τη μέθοδο Ckmeans

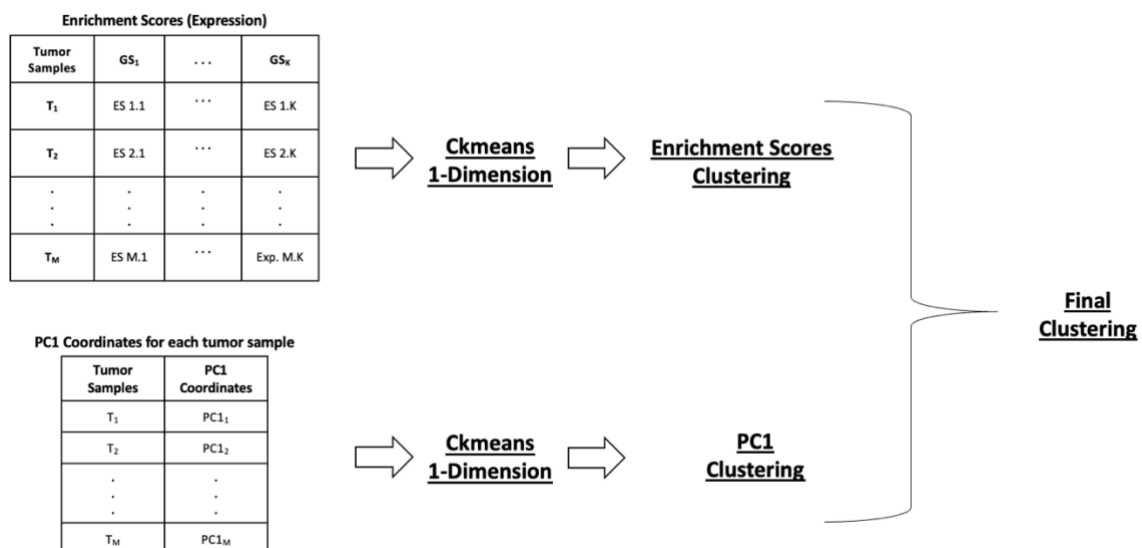
μίας διάστασης, χρησιμοποιώντας το πακέτο ‘Ckmeans.1d.dp’ στην R (H. Wang & Song, 2011).

Ο αριθμός των υπο-ομάδων (clusters) που διακρίθηκαν τα δείγματα υπολογίστηκε από το βέλτιστο αποτέλεσμα του κριτηρίου Bayesian Information Criterion (BIC), που αποτελεί μέτρο προσαρμογής λαμβάνοντας υπόψη την πολυπλοκότητα του μοντέλου και τον αριθμό των σημείων-δεδομένων. Στο πλαίσιο της μεθόδου ομαδοποίησης Ckmeans 1 διάστασης, το BIC υπολογίζει την προσαρμογή διάφορων αριθμών ομάδων στα δεδομένα και εντοπίζει το βέλτιστο. Ο υπολογισμός αυτός γίνεται βάσει της εξίσωσης:

$$BIC = n \cdot \log(SSE) + k \cdot \log(n) \quad (2)$$

όπου n ο αριθμός των σημείων-δεδομένων, SSE το άθροισμα των τετραγωνικών σφαλμάτων (το άθροισμα των τετραγωνικών διαφορών μεταξύ των σημείων-δεδομένων και των κέντρων των υπο-ομάδων τους), και k ο αριθμός των υπο-ομάδων (Webster, 2020).

Έτσι, προκειμένου να εξασφαλιστεί ένα εγκυρότερο αποτέλεσμα, η μέθοδος ομαδοποίησης Ckmeans 1 διάστασης εφαρμόζεται και για την ομαδοποίηση των δειγμάτων μέσω ανάλυσης PCA βάσει της έκφρασης των 1000 πιο μεταβλητών γονιδίων.



Εικόνα 17. Ομαδοποίηση δειγμάτων με τη χρήση της μεθόδου Ck-means 1 διάστασης. Το άνω σκέλος περιγράφει την ομαδοποίηση των καρκινικών δειγμάτων βάσει της ανάλυσης GSVA, ενώ το κάτω σκέλος την ομαδοποίηση των δειγμάτων βάσει της κύριας συνιστώσας τους κατά την PCA. Το τελικό αποτέλεσμα προκύπτει από τον συνυπολογισμό των 2 μεθόδων.

Προκειμένου να προκύψει η τελική ομαδοποίηση, υπολογίστηκε η συσχέτιση Pearson του βαθμού εμπλουτισμού των δειγμάτων στο κάθε σύνολο γονιδίων με την τιμή των δειγμάτων στην PC1. Η ομαδοποίηση προκύπτει από το σύνολο γονιδίων το οποίο έχει τον κατ' απόλυτο υψηλότερο συντελεστή συσχέτισης με την PC1. Για τα δείγματα που κατηγοριοποιούνται σε διαφορετική υπο-ομάδα κατά PC1 σε σχέση με αυτή του βαθμού εμπλουτισμού στο επιλεγμένο σύνολο γονιδίων, δε μπορεί να γίνει κατάταξη οπότε δεν συμπεριλαμβάνονται στην περαιτέρω ανάλυση.

3.4.5 Ανάλυση Διαφορικής Έκφρασης Ανθρώπινων Γονιδίων μεταξύ των υπο-ομάδων

Οι υπο-ομάδες των καρκινικών δειγμάτων ελέγχθηκαν, έπειτα, ως προς τη διαφορική γονιδιακή έκφραση, με το εργαλείο edgeR. Κατά τον έλεγχο αυτό, χρησιμοποιήθηκαν οι ίδιες παράμετροι και συναρτήσεις στην R με τη σύγκριση Tumor vs Normal. Ωστόσο, επειδή η σύγκριση αυτή πραγματοποιήθηκε μεταξύ υπο-ομάδων των καρκινικών δειγμάτων, εφαρμόστηκε ένα λιγότερο αυστηρό όριο στη συνάρτηση 'treat', $|\log_2FC| > 0.50$.

3.4.6 Λειτουργικός Χαρακτηρισμός Διαφορικά Εκφρασμένων Γονιδίων

Για το λειτουργικό χαρακτηρισμό των διαφορικά εκφρασμένων γονιδίων σε βιολογικές λειτουργίες με όρους GO και μονοπάτια Reactome, χρησιμοποιήθηκε το εργαλείο BioInfoMiner. Για την ανάλυση αυτή, εισήχθησαν οι λίστες των διαφορικά εκφρασμένων γονιδίων, με τις αντίστοιχες τιμές \log_2FC , τόσο μεταξύ των Tumor και Normal δειγμάτων, όσο και μεταξύ των υπο-ομάδων των Tumor δειγμάτων που κατηγοριοποιήθηκαν κατά το στάδιο της ομαδοποίησης με τη μέθοδο Ckmeans μίας διάστασης. Το όριο της στατιστικής σημαντικότητας των βιολογικών λειτουργιών και μονοπατιών τέθηκε ως η διορθωμένη p-value 0.05.

3.5 Μετα-μεταγραφωμική Ανάλυση

3.5.1 Εισαγωγή δεδομένων στο SAMSA2

Η είσοδος στο εργαλείο SAMSA2 είναι τα μη ευθυγραμμισμένα αναγνώσματα με το ανθρώπινο γονιδίωμα, της μορφής FASTQ, τα οποία μετά την εσωτερική προ-επεξεργασία του εργαλείου (Trimmomatic, PEAR και SortMeRNA), ευθυγραμμίστηκαν με το εργαλείο

ευθυγράμμισης πρωτεϊνών DIAMOND, χρησιμοποιώντας τις βάσεις δεδομένων NCBI RefSeq και SEED Subsystems Database (Εικόνα 14). Η βάση δεδομένων SEED τροποποιήθηκε, έτσι ώστε οι πρωτεΐνες που εμπεριέχονται να είναι κοινές με αυτές της βάσης δεδομένων RefSeq. Ακόμη, τροποποιήθηκαν δύο παράμετροι του εργαλείου SAMSA2, προκειμένου να αυστηροποιηθεί η ευθυγράμμιση των πρωτεϊνών. Συγκεκριμένα, το στατιστικό E-value (ο αριθμός των προβλεπόμενων τυχαίων ευθυγραμμίσεων) μειώθηκε από 10^{-1} σε 10^{-4} , και το identity match (το ποσοστό των ορθών ευθυγραμμίσεων) αυξήθηκε σε 90%. Έτσι, τα τελικά αποτελέσματα του SAMSA2 περιλαμβάνουν τη λίστα των ευθυγραμμισμένων πρωτεϊνών, καθώς και τους βακτηριακούς οργανισμούς, σε κάθε δυνατή ταξινομική βαθμίδα, και μεταβολικά μονοπάτια, που εκφράζουν οι εντοπισμένες πρωτεΐνες.

3.5.2 Ανάλυση Μικροβιώματος

Η ανάλυση του μικροβιώματος πραγματοποιήθηκε για τις συγκρίσεις μεταξύ των Normal δειγμάτων και των καρκινικών υπο-ομάδων, που προέκυψαν από την ομαδοποίηση των δειγμάτων βάσει της γονιδιακής τους έκφρασης. Ένα από τα σημαντικότερα περιγραφικά στατιστικά για την ανάλυση του μικροβιώματος είναι η σχετική αφθονία των βακτηρίων, η οποία πραγματοποιείται σε επίπεδο φύλου, και περιγράφει το λόγο της ποσότητας ενός φύλου σε όλο το δείγμα. Τα αποτελέσματα της σχετικής αφθονίας οπτικοποιούνται χρησιμοποιώντας ραβδογράμματα (barplots).

Για τον υπολογισμό της άλφα ποικιλομορφίας εντός των δειγμάτων, λήφθηκαν υπόψη οι δείκτες Shannon και Simpson, και υπολογίστηκε η p-value από ένα στατιστικό έλεγχο Wilcoxon, ανάμεσα στις ομάδες δειγμάτων. Η απεικόνιση γίνεται με θηκογράμματα (boxplots), όπου δίνεται μία εικόνα της κατανομής της αφθονίας των βακτηρίων στα καρκινικά και υγιή δείγματα, απεικονίζοντας τις διαφορές μεταξύ τους.

Για τον υπολογισμό της βήτα-ποικιλομορφίας μεταξύ των δειγμάτων, η μέθοδος ανομοιότητας που εφαρμόστηκε είναι η Bray-Curtis (Ricotta & Podani, 2017), η οποία εξετάζει τις ποσότητες των κοινών μικροβίων ανάμεσα σε ένα ζεύγος δειγμάτων, με ένα εύρος τιμών από 0 έως 1. Αν τα δύο δείγματα του εξεταζόμενου ζεύγους μοιράζονται τα ίδια μικροβιακά είδη, στην ίδια ποσότητα, τότε το μέτρο ανομοιότητας ισούται με 0. Αντίθετα, αν δεν έχουν κανένα κοινό βακτήριο, κάποιας ταξινομικής βαθμίδας, τότε το μέτρο ανομοιότητας λαμβάνει

τη μέγιστη τιμή του, 1. Η Ανάλυση Κύριων Συντεταγμένων (PCoA) αποτελεί μία μέθοδο PCA η οποία χρησιμοποιήθηκε για τον υπολογισμό και απεικόνιση της βήτα ποικιλομορφίας στα δείγματα, με διάγραμμα διασποράς.

Η ανάλυση της διαφορικής αφθονίας ειδών αναφέρεται στη διαδικασία αναγνώρισης βακτηριακών ειδών που απαντώνται διαφορεικά μεταξύ δύο βιολογικών καταστάσεων, χρησιμοποιώντας, όπως και στην ανάλυση της διαφορικής γονιδιακής έκφρασης, το εργαλείο edgeR, δίχως, ωστόσο, τη χρήση της συνάρτησης 'treat'. Η ανάλυση αυτή θα μπορούσε να πραγματοποιηθεί για κάθε ταξινομικό επίπεδο. Ωστόσο, επιλέχθηκε η ανάλυση στο ειδικότερο επίπεδο για την εξαγωγή περισσότερων πληροφοριών, δηλαδή τα είδη. Κατά τις συγκρίσεις μεταξύ των ομάδων δειγμάτων, για να χαρακτηριστεί ένα είδος σε διαφορεική αφθονία, τα όρια ορίστηκαν ως διορθωμένη p -value < 0.05 και $|\log_2FC| > 0.5$.

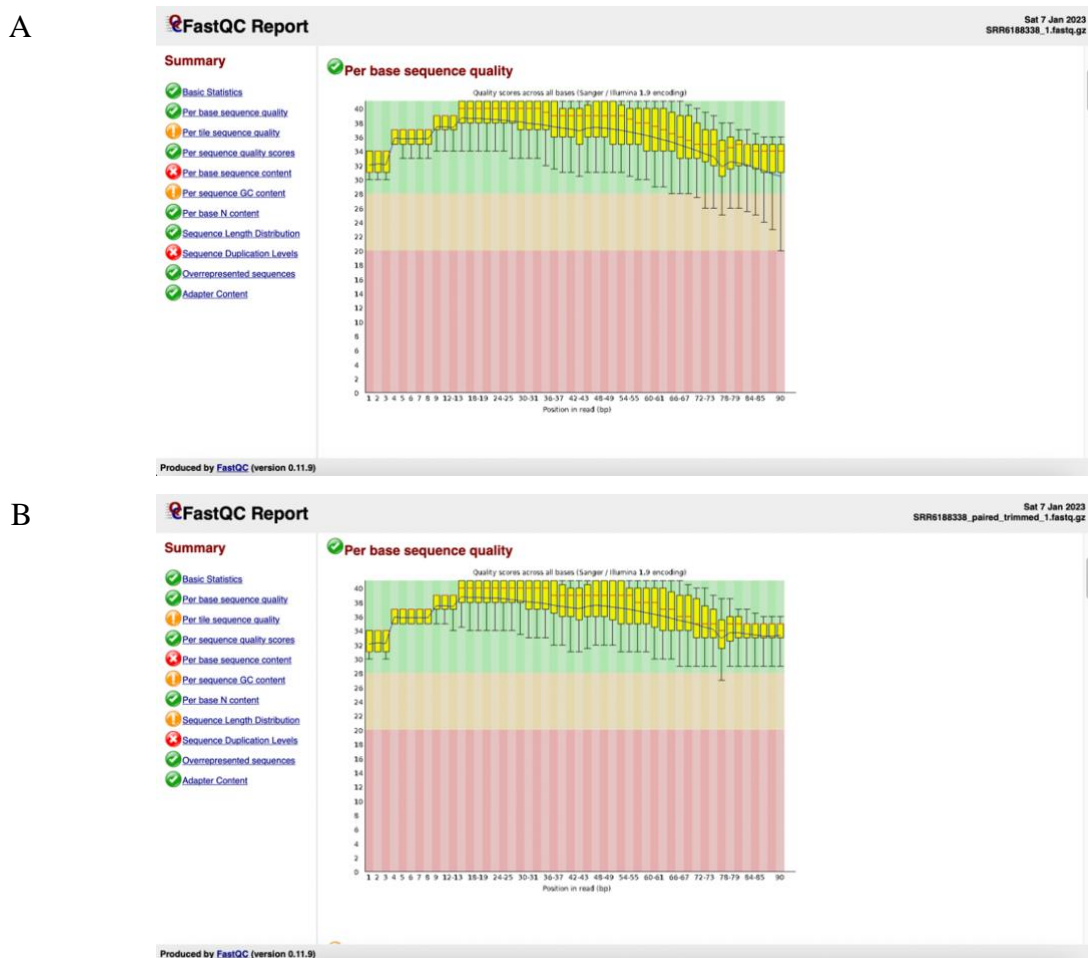
Το SAMSA2 έδωσε ως έξοδο και τις βιολογικές λειτουργίες που έχουν αντιστοιχιστεί στις ευθυγραμμισμένες πρωτεΐνες, οι οποίες εντοπίζονται με τη βοήθεια της βάσης δεδομένων SEED Subsystems Database, και διακρίνονται σε 4 διαφορετικά επίπεδα. Το επίπεδο 1 (Level 1) είναι το γενικότερο και αφορά κατηγορίες βιολογικών λειτουργιών, το επίπεδο 2 (Level 2) αναφέρεται σε μία γενική ομαδοποίηση μεταβολικών μονοπατιών, και το επίπεδο 3 (Level 3) σε επιμέρους μεταβολικά μονοπάτια, ενώ το επίπεδο 4 (Level 4) αφορά ένζυμα.

Έτσι, οι αναλύσεις σχετικής και διαφορικής αφθονίας εκτελούνται και στις βιολογικές λειτουργίες. Συγκεκριμένα, για την σχετική αφθονία, χρησιμοποιήθηκαν οι μετρήσεις των βιολογικών λειτουργιών Level 1, ενώ για τη διαφορική αφθονία, οι μετρήσεις των βιολογικών λειτουργιών Level 4, χρησιμοποιώντας τις ίδιες παραμέτρους που τέθηκαν στην ανάλυση διαφορικής αφθονίας των ειδών.

4 Αποτελέσματα και Συζήτηση Αποτελεσμάτων

4.1 Προεπεξεργασία Αρχικών Δεδομένων

Ο ποιοτικός έλεγχος πραγματοποιήθηκε με το εργαλείο FastQC, εις διπλούν, πριν και μετά την αποκοπή με το Trimmomatic, προκειμένου να γίνει αντιληπτή η βελτίωση της ποιότητας των δεδομένων. Στην Εικόνα 18 παρατηρείται ο ποιοτικός έλεγχος του αρχείου FASTQ ενός εκ των 2 αναγνώσμάτων του δείγματος SRR6188338, και φαίνεται πως η ποιότητα βελτιώθηκε σημαντικά, ιδιαίτερα προς το τέλος της αλληλουχίας, στις τελευταίες 30 βάσεις από τις 90 συνολικά. Επιπλέον, τα αναγνώσματα μετά την αποκοπή είναι κατά 3% λιγότερα, κατά προσέγγιση, στα δείγματα.



Εικόνα 18. Αναφορά αποτελεσμάτων FastQC για τον έλεγχο ποιότητας αλληλουχίας ανά βάση για το δείγμα SRR6188338, πριν (A) και μετά (B) τη χρήση του εργαλείου αποκοπής Trimmomatic.

4.2 Ευθυγράμμιση με το Ανθρώπινο Γονιδίωμα

Η ευθυγράμμιση με το ανθρώπινο γονιδίωμα αναφοράς έδωσε συνεπή αποτελέσματα σε όλα τα δείγματα. Το εργαλείο HISAT2 δίνει ένα αρχείο κειμένου για κάθε δείγμα που παρουσιάζει την περίληψη της ευθυγράμμισης που πραγματοποιείται. Για παράδειγμα, για το δείγμα SRR6188338 (Εικόνα 19), η ευθυγράμμιση με το ανθρώπινο γονιδίωμα ανήλθε στο ποσοστό 98.77%.

```
27932768 reads; of these:
 27932768 (100.00%) were paired; of these:
   765801 (2.74%) aligned concordantly 0 times
  24667339 (88.31%) aligned concordantly exactly 1 time
  2499628 (8.95%) aligned concordantly >1 times
-----
  765801 pairs aligned concordantly 0 times; of these:
   115713 (15.11%) aligned discordantly 1 time
-----
  650088 pairs aligned 0 times concordantly or discordantly; of these:
  1300176 mates make up the pairs; of these:
   685221 (52.70%) aligned 0 times
   521632 (40.12%) aligned exactly 1 time
   93323 (7.18%) aligned >1 times
98.77% overall alignment rate
```

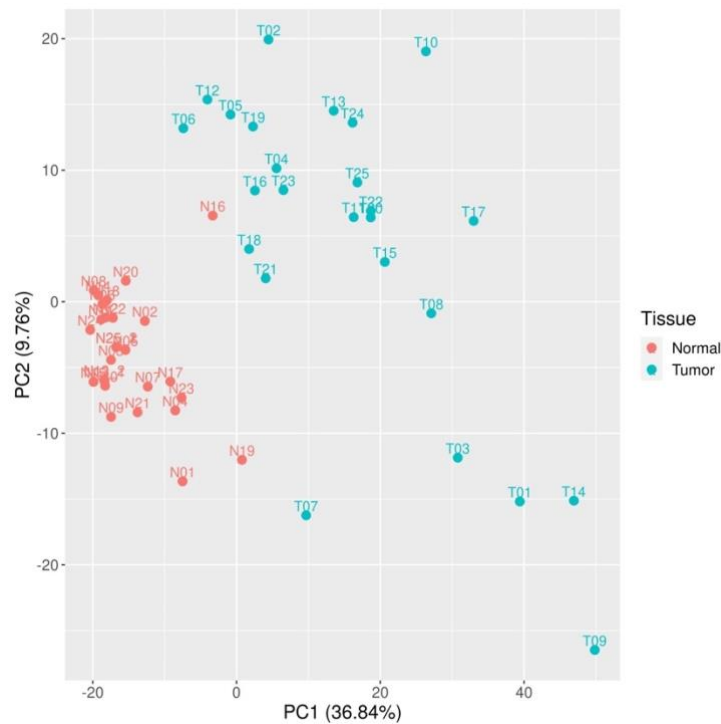
Εικόνα 19. Αρχείο κειμένου του αποτελέσματος ευθυγράμμισης από το εργαλείο HISAT2 για το δείγμα SRR6188338.

Συνολικά, το μέσο ποσοστό ευθυγράμμισης (alignment rate) των δειγμάτων στο ανθρώπινο γονιδίωμα υπολογίστηκε 98.73%. Επομένως, το 1.27% των αλληλουχιών, κατά μέσο όρο, στα δείγματα, δεν ευθυγραμμίζεται με το ανθρώπινο γονιδίωμα.

Μετά την εισαγωγή των δεδομένων της ποσοτικοποίησης των ευθυγραμμισμένων γονιδίων από το featureCounts, ακολούθησε η απλοποίηση των ονομασιών των δειγμάτων. Οι ασθενείς αριθμήθηκαν από το 1 ως το 25, και, ανάλογα με τη φύση του δείγματος, φυσιολογικό ή καρκινικό, επισημάνθηκε ένας χαρακτήρας N (Normal) ή T (Tumor), αντίστοιχα. Για τους ασθενείς 13 και 25, που διαθέτουν από 2 φυσιολογικά δείγματα ο καθένας, η ετικέτα δόθηκε ως N13_1 και N13_2, και N25_1 και N25_2.

4.3 Ανάλυση Γονιδιακής Έκφρασης στο ανθρώπινο μεταγράφομα

Τα αποτελέσματα της ανάλυσης PCA βάσει των μετρήσεων των αναγνωσμάτων σε κάθε δείγμα, για τα 1000 πιο μεταβλητά γονίδια, διαχωρίζουν τα δείγματα διακριτικά σε δύο ομάδες που αντιστοιχούν σε φυσιολογικούς και καρκινικούς ιστούς ασθενών (Εικόνα 20).



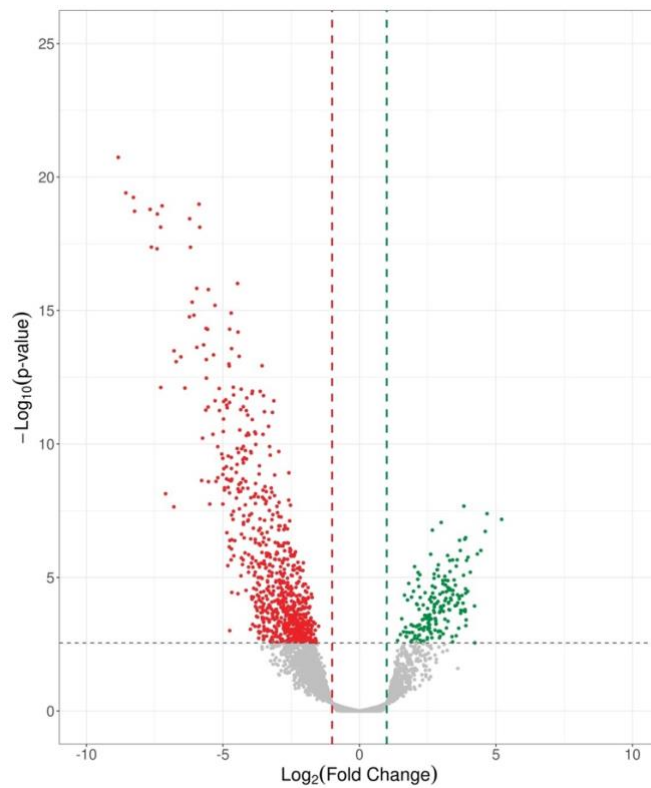
Εικόνα 20. Τα φυσιολογικά (κόκκινο) και τα καρκινικά (γαλάζιο) δείγματα διαχωρίζονται ποιοτικά στην ανάλυση κυρίων συνιστωσών (PCA) βάσει των καταμετρήσεων των αναγνωσμάτων τους.

Παρόμοια εικόνα δίνεται και από το θερμικό χάρτη (heatmap) που κατασκευάστηκε για τα 1000 πιο μεταβλητά γονίδια στα δείγματα (Εικόνα 21). Παρατηρείται πως τα δείγματα έχουν διαχωριστεί και πάλι βάσει του γονιδιακού τους προφίλ σε καρκινικά και φυσιολογικά, με εξαίρεση τον υγιή ιστό N16, και τον καρκινικό T07. Τα αποτελέσματα αυτά επιβεβαιώνουν την ορθότητα της βιοπληροφορικής επεξεργασίας και ανάλυσης των δεδομένων που πραγματοποιήθηκε για το ανθρώπινο μεταγράφομα.



Εικόνα 21. Θερμικός χάρτης απεικόνισης των καταμετρήσεων κάθε γονιδίου (γραμμές) στα δείγματα (στήλες).

Στη συνέχεια πραγματοποιήθηκε ανάλυση της διαφορικής έκφρασης των γονιδίων (Tumor vs Normal), όπως φαίνεται και στο volcano plot της Εικόνας 22.



Εικόνα 22. Volcano plot ανάλυσης διαφορικής έκφρασης γονιδίων της σύγκρισης Tumor vs Normal δειγμάτων. Κάθε κουκίδα αντιστοιχεί σε ένα γονίδιο. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται τα στατιστικά σημαντικά υπο-εκφρασμένα γονίδια στα καρκινικά δείγματα έναντι των φυσιολογικών. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται τα στατιστικά σημαντικά σημαίνει υπερ-εκφρασμένα γονίδια στα καρκινικά δείγματα έναντι των φυσιολογικών. Με γκρι χρώμα είναι τα γονίδια χωρίς στατιστικά σημαντική διαφορική έκφραση.

Κατά τη σύγκριση Tumor vs Normal, βρέθηκαν 959 διαφορεικά εκφρασμένα γονίδια, εκ των οποίων τα 773 είναι υπο-εκφρασμένα στα καρκινικά δείγματα σε σχέση με τα φυσιολογικά, ενώ τα 186 είναι υπερ-εκφρασμένα.

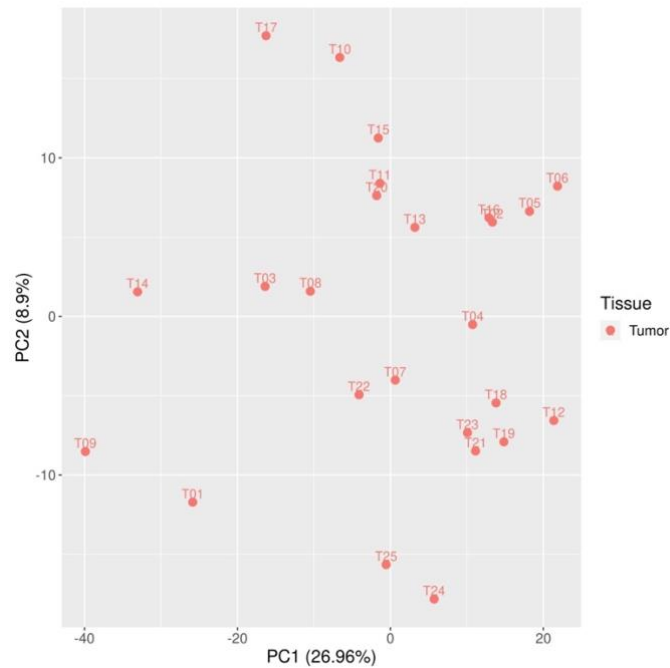
Τα αποτελέσματα που προέκυψαν από την ανωτέρω ανάλυση σε σχέση με τη διαφορετική έκφραση αρκετών γονιδίων και τη βιολογική σημασία της επιβεβαιώνονται βιβλιογραφικά, γεγονός που επαληθεύει την εγκυρότητα της ανάλυσης, αλλά και την ποιότητα της ευθυγράμμισης με το HISAT2. Για παράδειγμα, 3 υπο-εκφρασμένα γονίδια της σύγκρισης Tumor vs Normal είναι τα CLEC4M ($\log_2FC = -8.83$), MARCO ($\log_2FC = -7.67$) και CYP1A2 ($\log_2FC = -6.71$). Η CLEC4M είναι μία πρωτεΐνη που εκφράζεται στο ήπαρ και παίζει ρόλο στη λειτουργία του ανοσοποιητικού, και η MARCO είναι μια πρωτεΐνη που εκφράζεται στην επιφάνεια ορισμένων κυττάρων του ανοσοποιητικού που ονομάζονται μακροφάγα, και έχει αποδειχθεί ότι παίζει ρόλο στην ανοσολογική απόκριση στον καρκίνο. Ενώ παλαιότερα δεν ήταν ξεκάθαρη η έκφραση των πρωτεϊνών αυτών στο HCC, πρόσφατα έχει εντοπιστεί η υπο-έκφρασή τους στα καρκινικά κύτταρα (Luo et al., 2020; Xiao et al., 2019). Τέλος, η πρωτεΐνη CYP1A2 είναι ένα ένζυμο, του οποίου η απορρύθμιση έχει βρεθεί ότι υποδηλώνει καρκινογένεση (Yu et al., 2021).

Ακόμη, 2 γονίδια που υπερ-εκφράζονται στον καρκίνο βρέθηκαν να είναι τα TERT ($\log_2FC = +5.21$) και GPC3 ($\log_2FC = +4.45$), τα οποία και βιβλιογραφικά επαληθεύτηκε πως υπερ-εκφράζονται στα καρκινικά κύτταρα. Συγκεκριμένα, η πρωτεΐνη TERT είναι ένα ένζυμο που εμπλέκεται στη διατήρηση των τελομερών, τα οποία είναι προστατευτικές δομές στα άκρα των χρωμοσωμάτων (Jang et al., 2021), και η GPC3 έχει αποδειχθεί πως υπερ-εκφράζεται στο HCC, αποτελώντας σημαντικό βιοδείκτη, καθώς και πως η έκφρασή του είναι ανάλογη του βαθμού διαφοροποίησης των κυττάρων (M. Suzuki et al., 2010).

4.4 Μοριακή ταξινόμηση καρκινικών δειγμάτων ασθενών με βάση τις υπογραφές γονιδιακών βιοδεικτών.

Η ανάλυση PCA στα καρκινικά δείγματα σε σχέση με την έκφραση των 1000 πιο μεταβλητών γονιδίων φανερώνει από την αρχή μία μεταξύ τους διαφοροποίηση (Εικόνα 23). Για παράδειγμα, παρατηρείται πως βάσει της PC1, της κύριας συνιστώσας που καταλαμβάνει

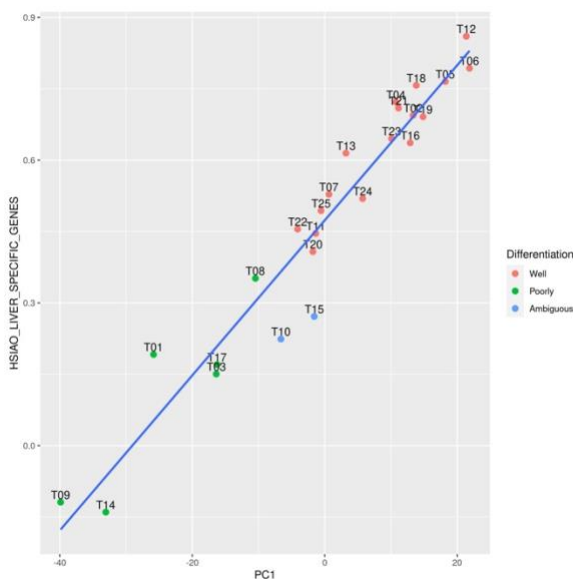
το 26.96% της πληροφορίας, τα δείγματα T09 και T12, διαφέρουν σημαντικά ως προς τη γονιδιακή τους έκφραση



Εικόνα 23. PCA στα καρκινικά δείγματα, για τα 1000 γονίδια με την υψηλότερη μεταβλητότητα.

Για την μοριακή ταξινόμηση των δειγμάτων, η ανάλυση BIC προτείνει 2 υπο-ομάδες ως το βέλτιστο αριθμό, για κάθε γονιδιακή υπογραφή και για την ομαδοποίηση βάσει της θέσης των δειγμάτων στην PC1. Η μόνη εξαίρεση αποτελεί η ομαδοποίηση βάσει του συνόλου γονιδίων HOSHIDA_LIVER_CANCER_SUCLASS_S3, όπου το κριτήριο BIC υπολόγισε των βέλτιστο αριθμό στις 4 υπο-ομάδες (βλ. Εικόνα 41).

Κατά τη ομαδοποίηση με βάση το βαθμό εμπλουτισμού της PC1, τα καρκινικά δείγματα ομαδοποιήθηκαν όπως φαίνεται στην Εικόνα 42, με τη χρήση της μεθόδου Ck-means 1 διάστασης. Καθώς το αποτέλεσμα αυτό θα συμψηφιστεί με τις ομαδοποιήσεις που προέκυψαν από τους βαθμούς εμπλουτισμού των 5 συνόλων γονιδίων (Εικόνα 43), υπολογίστηκαν οι συντελεστές συσχέτισης Pearson των συνόλων γονιδίων με την PC1 (Πίνακας 3). Τα δύο σύνολα γονιδίων με την υψηλότερη συσχέτιση είναι οι HOSHIDA_LIVER_CANCER_SUCLASS_S3 και HSIAO_LIVER_SPECIFIC_GENES. Επιλέχθηκε το δεύτερο για την εύρεση της τελικής ομαδοποίησης, καθώς οι προκυπτόμενες υπο-ομάδες ταυτίζονται για όλα τα δείγματα (εκτός των T10 και T15), με αυτές από την ομαδοποίηση βάσει της PC1 (Εικόνα 24).



Εικόνα 24. Ομαδοποίηση καρκινικών δειγμάτων βάσει του βαθμού εμπλουτισμού του συνόλου γονιδίων Hsiao Liver Specific Genes στην PC1.

Πίνακας 3. Συντελεστές συσχέτισης της PC1 με τους βαθμούς εμπλουτισμού των συνόλων γονιδίων.

Σύνολο Γονιδίων	Συντελεστής συσχέτισης, ρ	p-value
Hoshida Subclass S1	-0.58	2.59E-03
Hoshida Subclass S2	-0.82	4.94E-07
Hoshida Subclass S3	0.97	6.21E-16
Hsiao Liver Specific Genes	0.97	4.20E-15
SU Liver	0.95	9.33E-13

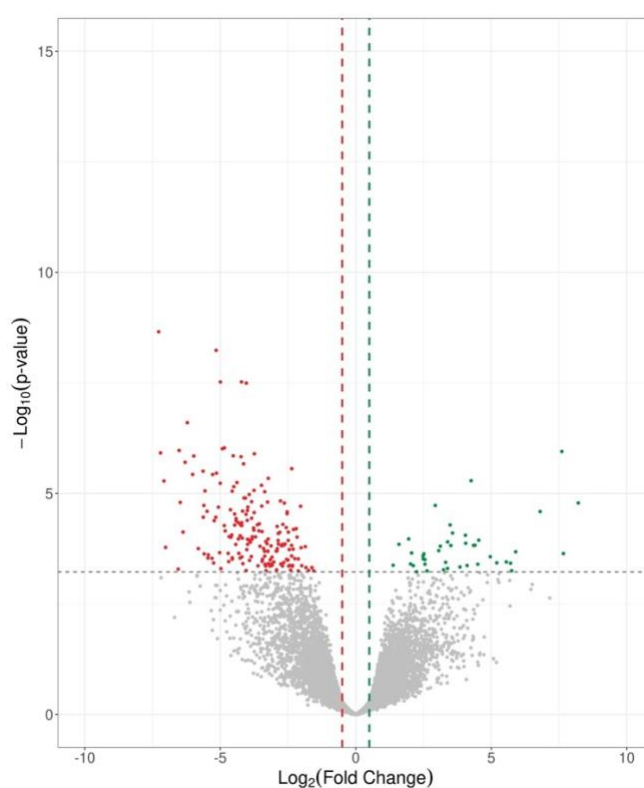
Αναλυτικότερα, τα καρκινικά δείγματα τα οποία έχουν υψηλό βαθμό εμπλουτισμού σε αυτό το σύνολο γονιδίων σημαίνει πως διαθέτουν μοριακά χαρακτηριστικά που προσομοιάζουν τα υγιή διαφοροποιημένα ηπατικά κύτταρα. Αντίθετα, τα καρκινικά δείγματα με χαμηλό βαθμό εμπλουτισμού σε αυτό το σύνολο γονιδίων τείνουν να παρουσιάζουν χαρακτηριστικά χαμηλής διαφοροποίησης (Hsiao et al., 2001). Επομένως, οι 2 υπο-ομάδες που προέκυψαν και παρουσιάζονται στον Πίνακα 4 αφορούν τη διαφοροποίηση των καρκινικών κυττάρων και κατατάσσουν κάθε δείγμα όγκου, σε χαμηλής (Poorly) και υψηλής (Well) διαφοροποίησης. Τα δύο καρκινικά δείγματα που δεν ομαδοποιήθηκαν, χαρακτηρίστηκαν ως ασαφή (“Ambiguous”) και αγνοήθηκαν στη συνέχεια της ανάλυσης.

Πίνακας 4. Οι υπο-ομάδες που σχηματίστηκαν διαχωρίζουν τα καρκινικά δείγματα βάσει της διαφοροποίησης των κυττάρων.

Well	Poorly	Ambiguous
T02, T04, T05, T06, T07, T11, T12, T13, T16, T18, T19, T20, T21, T22, T23, T24, T25	T01, T03, T08, T09, T14, T17	T10, T15

4.5 Διαφορική έκφραση γονιδίων μεταξύ των καρκινικών υπο-ομάδων

Στη συνέχεια εξετάστηκε η διαφορική έκφραση γονιδίων μεταξύ μεταξύ Poorly και Well καρκινικών δειγμάτων (Poorly vs Well), όπως ορίστηκαν παραπάνω κατά την ομαδοποίηση (Εικόνα 25). Στην ανάλυση αυτή εντοπίστηκαν 222 συνολικά διαφορικά εκφρασμένα γονίδια, εκ των οποίων τα 180 είναι υπο-εκφρασμένα και τα 42 είναι υπερ-εκφρασμένα στα Poorly δείγματα έναντι των Well. Το αποτέλεσμα αυτό ενισχύει την παραπάνω ταξινόμηση καθώς δείχνει πως σε επίπεδο γονιδιακής έκφρασης και επομένως σχετιζόμενων μοριακών χαρακτηριστικών οι δύο υπο-ομάδες διαφέρουν σημαντικά μεταξύ τους.



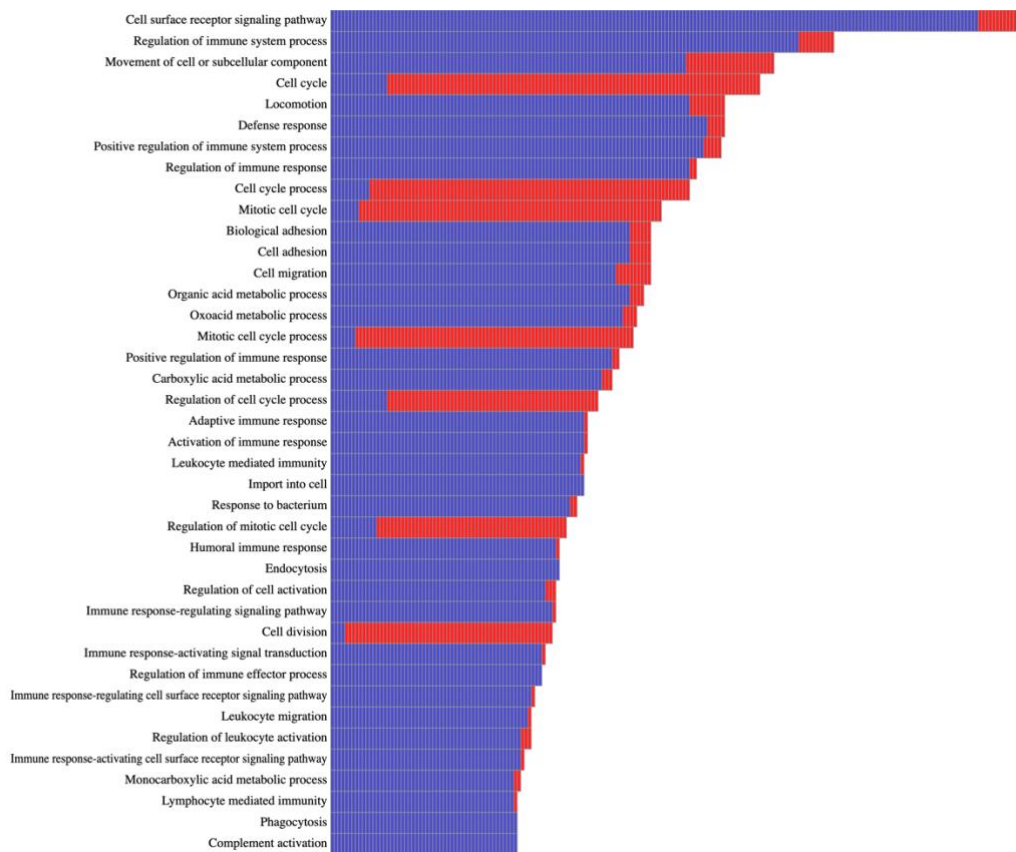
Εικόνα 25. Ανάλυση διαφορικής έκφρασης γονιδίων μεταξύ Poorly και Well δειγμάτων. Κάθε κουκίδα αντιστοιχεί σε ένα γονίδιο. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται τα στατιστικά σημαντικά υπο-εκφρασμένα γονίδια στα Poorly δείγματα έναντι των Well. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται τα στατιστικά σημαντικά σημαίνει υπερ-εκφρασμένα γονίδια στα Poorly δείγματα έναντι των Well. Με γκρι χρώμα είναι τα γονίδια χωρίς στατιστικά σημαντική διαφορική έκφραση.

4.6 Λειτουργικός Χαρακτηρισμός Διαφορικά Εκφρασμένων Γονιδίων

Για τη λειτουργική ανάλυση των διαφορικά εκφρασμένων γονιδίων από τις συγκρίσεις Tumor vs Normal και Poorly vs Well δειγμάτων χρησιμοποιήθηκε το εργαλείο BioInfoMiner και εξετάστηκαν συσχετίσεις με γονιδιακές οντολογίες, καθώς και βιολογικά μονοπάτια της βάσης δεδομένων Reactome. Οι λειτουργίες οι οποίες έχουν διορθωμένο p-value (corrected p-

value) < 0.05 θεωρούνται στατιστικά σημαντικές όσον αφορά τον εμπλουτισμό τους στα διαφορικά εκφρασμένα γονίδια.

Βάσει αυτής της ανάλυσης προέκυψαν 112 διεργασίες GO (Πίνακας 7) και 27 βιολογικά μονοπάτια Reactome (Πίνακας 9) που σχετίζονται σημαντικά με τα διαφορικά εκφρασμένα γονίδια Tumor vs Normal. Επίσης, σημαντική σύνδεση προέκυψε για 74 διεργασίες GO (Πίνακας 11) και 12 βιολογικά μονοπάτια Reactome (Πίνακας 13) σε σχέση με τα διαφορικά εκφρασμένα γονίδια Poorly vs Well. Βρέθηκαν όροι που σχετίζονται με την επούλωση τραυμάτων, την πήξη του αίματος και την αιμόσταση, τη μετανάστευση και κινητικότητα των κυττάρων. Επιπλέον, λειτουργίες που σχετίζονται με αλλαγές στον μεταβολισμό και καταβολισμό καθώς και τη βιοσύνθεση λιπιδίων και οξέων ήταν εκφρασμένες σε υψηλό βαθμό.



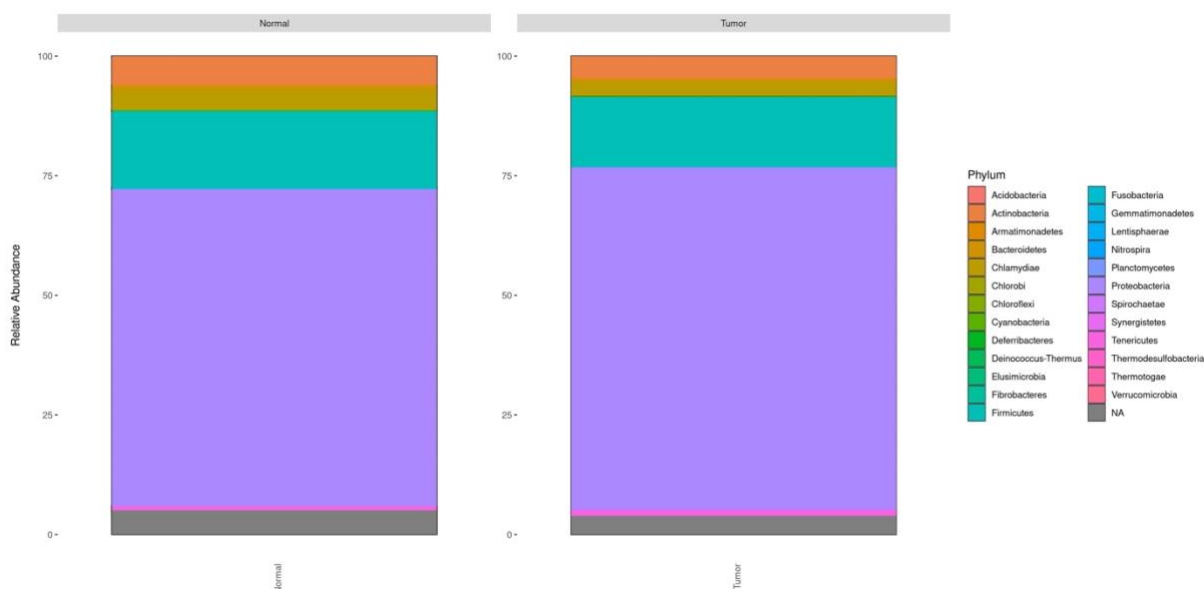
Εικόνα 26. Ανάλυση εμπλουτισμού λειτουργιών GO συσχετίζοντας τα διαφορικά εκφρασμένα γονίδια με στατιστικά σημαντικές βιολογικές λειτουργίες (διορθωμένη p-value < 0.05). Με κόκκινο χρώμα παρουσιάζεται η υπερ-έκφραση, ενώ με μπλε η υπο-έκφραση των γονιδίων. Άξονας y: βιολογικές λειτουργίες σε όρους GO, Άξονας x: η ποσότητα των συσχετιζόμενων γονιδίων με κάθε βιολογική λειτουργία.

4.7 Ανάλυση Μικροβιώματος

Τα αναγνώσματα τα οποία δεν ευθυγραμμίστηκαν με το ανθρώπινο γονιδίωμα αναλύθηκαν από το εργαλείο SAMSA2, το οποίο ύστερα από την ευθυγράμμιση και το χαρακτηρισμό των αναγνωσμάτων στις βάσεις δεδομένων NCBI RefSeq και SEED Subsystems Database, έδωσε αποτελέσματα όσον αφορά τους αντιστοιχισμένους οργανισμούς και βιολογικές λειτουργίες. Συγκεκριμένα, μετά το λειτουργικό χαρακτηρισμό αντιστοιχίστηκαν 3403 μοναδικά βακτηριακά είδη σε όλα τα δείγματα - 2580 στα Normal και 2526 στα Tumor - και 1982 μοναδικές βιολογικές λειτουργίες Επιπέδου 4 (Level 4), 1439 στα Normal και 1374 στα Tumor).

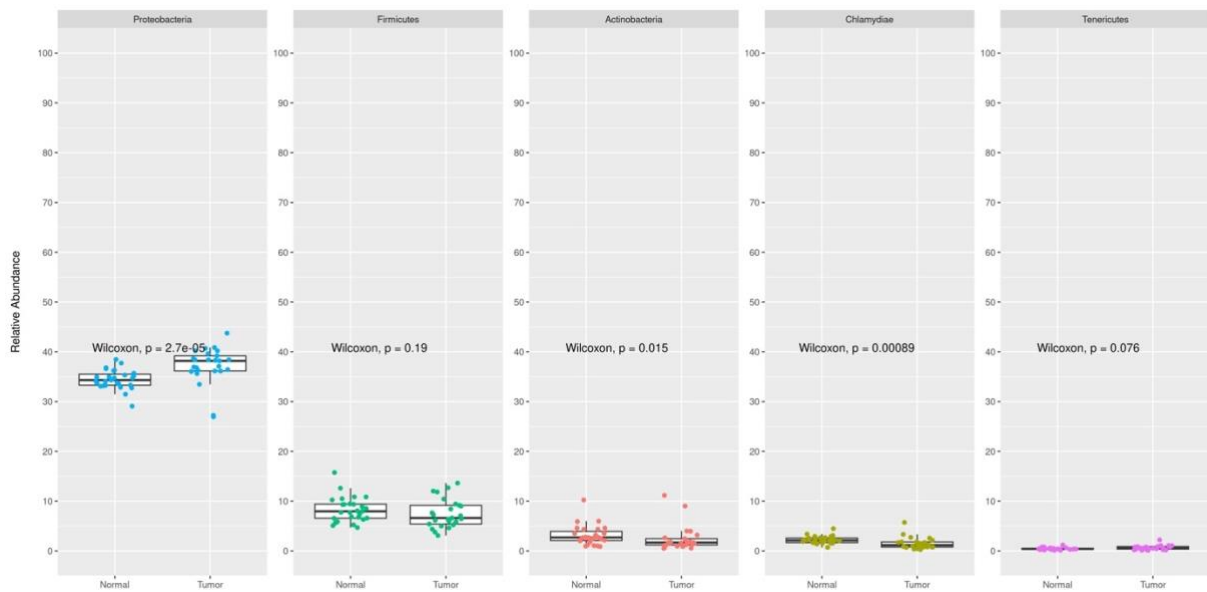
4.7.1 Χαρακτηρισμός του ενδοογκικού μικροβιώματος

Κατά την ανάλυση της σχετικής αφθονίας στα δείγματα Tumor και Normal (Εικόνα 27), παρατηρείται πως τα 5 κυριότερα φύλα στα δείγματα είναι τα *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Chlamydiae* και *Tenericutes*. Συγκεκριμένα, παρουσιάζεται μείωση της σχετικής αφθονίας των φύλων *Actinobacteria* και *Firmicutes*, και αύξηση του φύλου *Proteobacteria*, στα Tumor δείγματα.



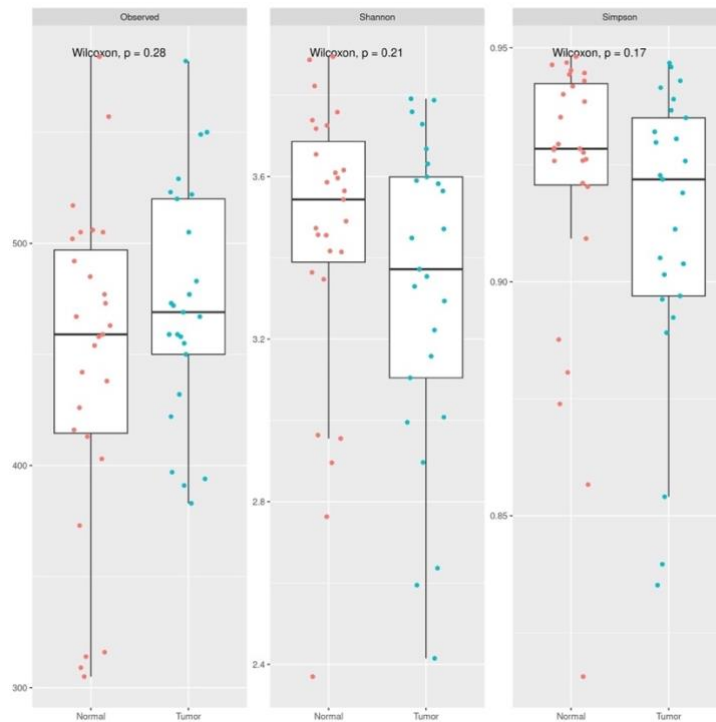
Εικόνα 27. Περιγραφική απεικόνιση της σχετικής αφθονίας του μικροβιώματος των δειγμάτων των δύο κατηγοριών, Normal (αριστερά) και Tumor (δεξιά), σε επίπεδο φύλων.

Με την εφαρμογή στατιστικού ελέγχου Wilcoxon ($p < 0.05$) στην Εικόνα 28, φαίνεται πως η διαφορά της σχετικής αφθονίας των φύλων *Proteobacteria* και *Actinobacteria* μεταξύ Tumor και Normal δειγμάτων είναι στατιστικά σημαντική. Αύξηση της σχετικής αφθονίας των *Proteobacteria* στα καρκινικά δείγματα έχει βρεθεί σε πρόσφατες μελέτες για το μικροβίωμα στο HCC, σε συγκρίσεις μεταξύ παρακαρκινικών και καρκινικών δειγμάτων ασθενών (Huang et al., 2022). Ακόμη, σε πρόσφατη μελέτη της σχετικής αφθονίας μεταξύ ασθενών HCC και δειγμάτων ελέγχου εντοπίστηκε μείωση των *Actinobacteria* στα καρκινικά δείγματα (Xue et al., 2022).



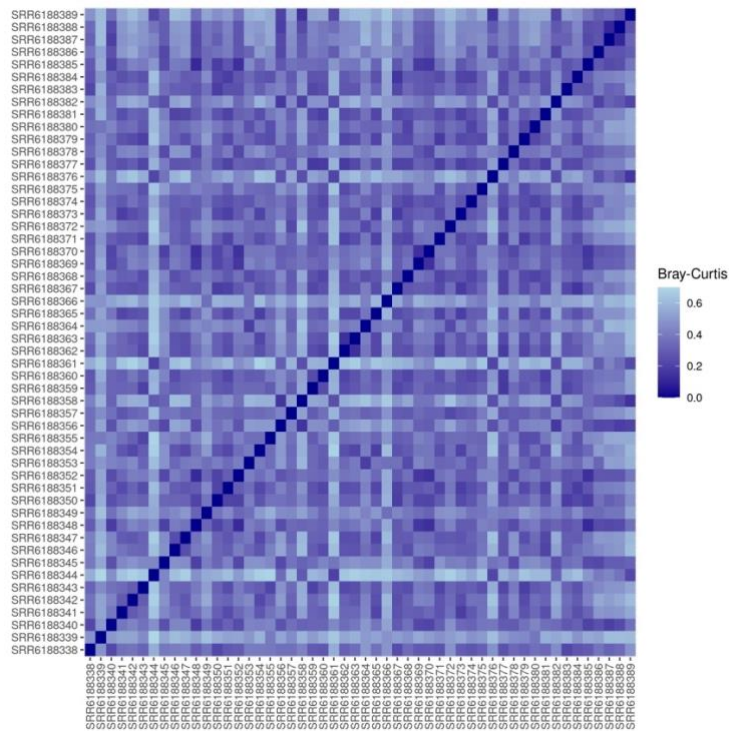
Εικόνα 28. Στατιστικός έλεγχος Wilcoxon ($p < 0.05$) της σχετικής αφθονίας του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal και Tumor, σε επίπεδο φύλου.

Στην ανάλυση της άλφα ποικιλομορφίας (Εικόνα 29), δεν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των ομάδων δειγμάτων, τόσο στα παρατηρούμενα OTU (Observed), όσο και στους δείκτες Shannon και Simpson. Βιβλιογραφικά επιβεβαιώνεται πως σε συγκρίσεις HCC με παρακαρκινικά δείγματα, δεν υπάρχουν στατιστικά σημαντικές διαφορές σε επίπεδο άλφα ποικιλομορφίας (Qu et al., 2022).

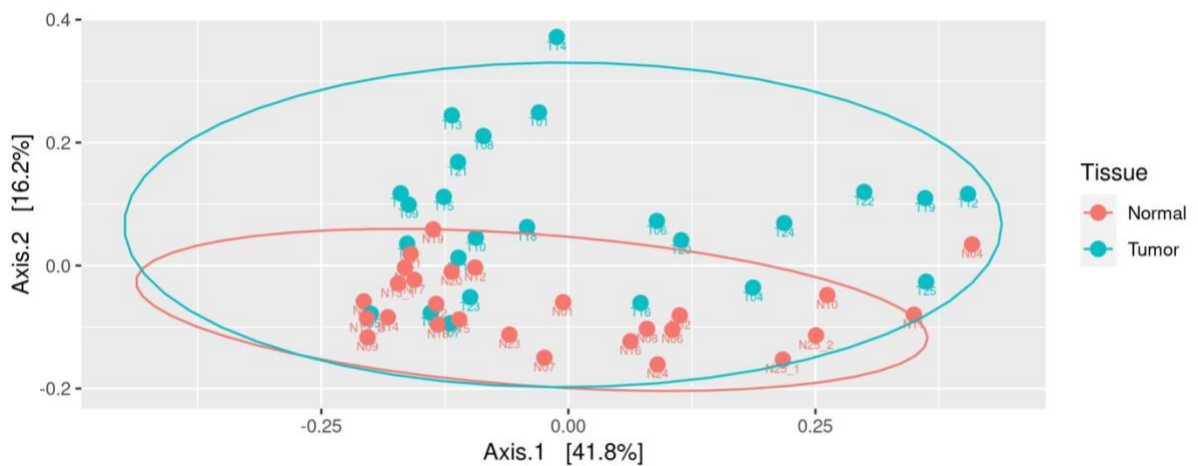


Εικόνα 29. Ανάλυση άλφα-ποικιλιομορφίας των δειγμάτων του μικροβιώματος για τα παρατηρούμενα (*Observed*) OTUs, το δείκτη *Shannon* και *Simpson*.

Στην ανάλυση της βήτα ποικιλιομορφίας, αρχικά, υπολογίστηκαν οι συντελεστές ανομοιότητας Bray-Curtis για κάθε ζεύγος δειγμάτων, που απεικονίζονται στο heatmap της Εικόνας 30. Από την ανάλυση PCoA της Εικόνας 31, παρατηρείται πως, σε επίπεδο βήτα ποικιλιομορφίας, υπάρχουν δύο διακριτά μικροβιώματα στα Tumor και Normal δείγματα, αντίστοιχα. Συγκεκριμένα, η μεταβλητότητα μεταξύ των Tumor δειγμάτων φαίνεται αυξημένη τόσο κατά την πρώτη κύρια συνιστώσα (Axis.1), όσο και κατά τη δεύτερη (Axis.2). Φαίνεται, δηλαδή, πως το μικροβίωμα στα φυσιολογικά δείγματα είναι πιο σταθερό ως προς η σύστασή του, σε σχέση με τα καρκινικά.



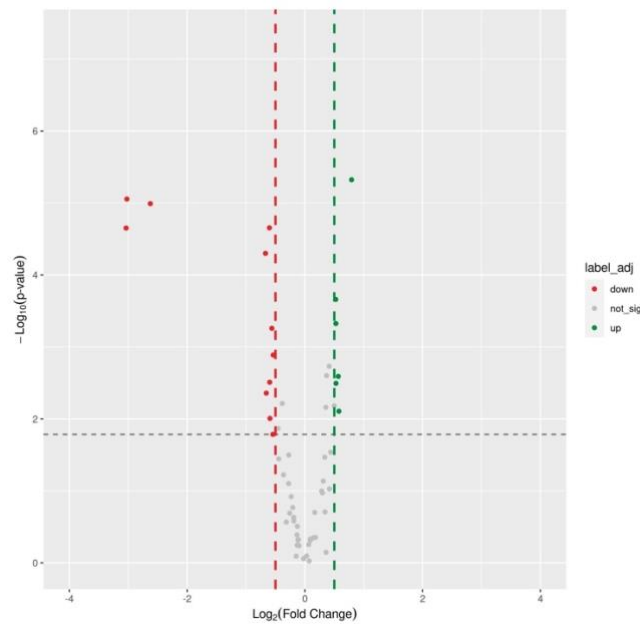
Εικόνα 30. Heatmap απεικόνισης της ανομοιότητας των δειγμάτων κατά την ανάλυση της βήτα ποικιλομορφίας με τη χρήση του μέτρου ανομοιότητας Bray-Curtis. Η τιμή 0 δηλώνει την πλήρη ομοιότητα μεταξύ δειγμάτων, ενώ η τιμή 1 την πλήρη ανομοιότητα.



Εικόνα 31. Διάγραμμα διασποράς PCoA για την ανάλυση της βήτα-ποικιλομορφίας μεταξύ των δειγμάτων, με τη μέθοδο ανομοιότητας Bray-Curtis.

Στην ανάλυση διαφορικής αφθονίας των Tumor vs Normal εντοπίστηκαν 17 είδη σε διαφορετική αφθονία, και συγκεκριμένα, 11 υπο-εκφρασμένα και 6 υπερ-εκφρασμένα (Εικόνα 32). Το *Lactococcus lactis* ($\log_2FC = -3.04$), που βρέθηκε υπο-εκφρασμένο στα καρκινικά δείγματα έναντι των φυσιολογικών, είναι γνωστό για τις αντικαρκινικές του ιδιότητες, καταστέλλοντας την ανάπτυξη του καρκίνου, ρυθμίζοντας την αγγειογένεση και προκαλώντας τον άμεσο θάνατο των καρκινικών κυττάρων (S. Kim et al., 2022). Επιπρόσθετα, δύο είδη του

γένους *Aeromonas*, *A. piscicola* ($\log_2FC = +0.79$) και *A. hydrophila* ($\log_2FC = +0.53$), βρέθηκαν υπερ-εκφρασμένα στα Tumor δείγματα. Έχει αναφερθεί, πως τα παθογόνα βακτήρια του εν λόγω γένους παρουσιάζουν αύξηση σε περιπτώσεις HCC, λόγω της συσσώρευσης χολής που προκαλείται από τον καρκίνο και, ταυτόχρονα, της αντοχής αυτών των βακτηρίων σε αυτό το περιβάλλον (Y.-W. Chen et al., 2021).

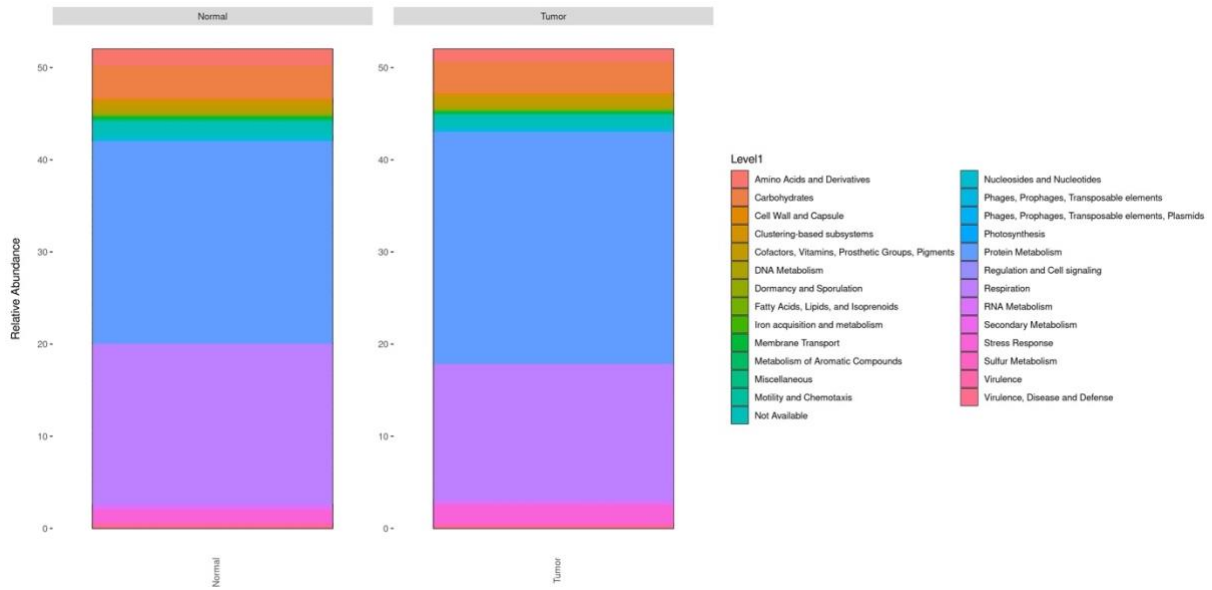


Εικόνα 32. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Tumor vs Normal. Κάθε κουκίδα αντιστοιχεί σε ένα είδος. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται τα στατιστικά σημαντικά υπο-εκφρασμένα είδη στα Tumor δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται τα στατιστικά σημαντικά υπερ-εκφρασμένα είδη στα Tumor δείγματα, έναντι των Normal. Με γκρι χρώμα είναι τα είδη χωρίς στατιστικά σημαντική διαφορική αφθονία

4.7.2 Ανάλυση βιολογικών λειτουργιών του ενδοογκικού μικροβιώματος

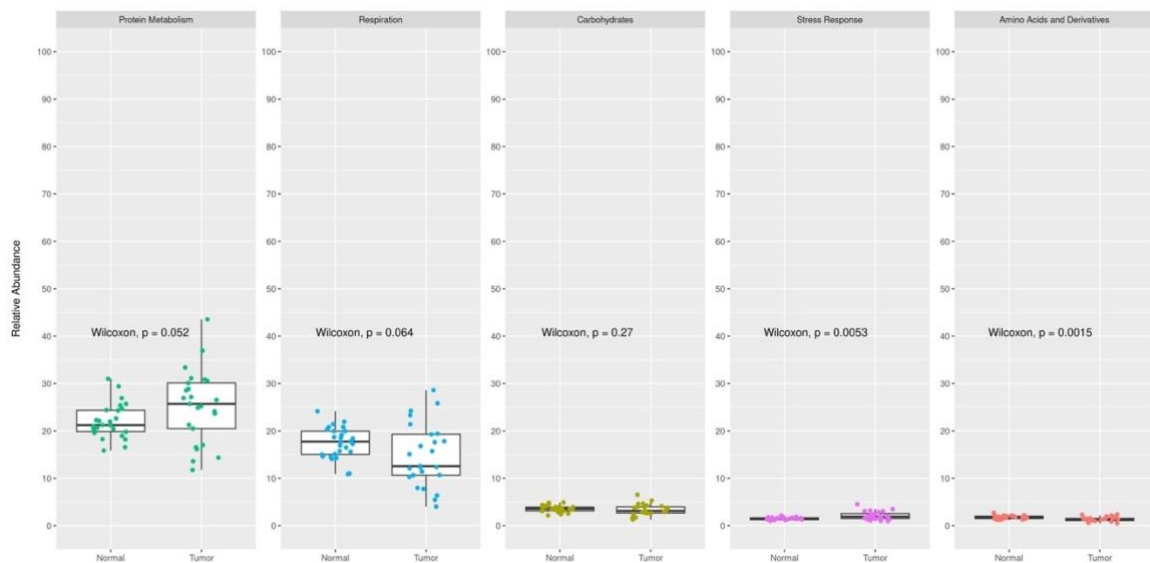
Ο λειτουργικός χαρακτηρισμός των μικροβιακών γονιδίων από το SAMSA2, που πραγματοποιήθηκε με τη βάση δεδομένων SEED Subsystems Database, δίνει αποτελέσματα για τις βιολογικές λειτουργίες σε 4 διαφορετικά επίπεδα (Levels), από το γενικότερο (Level 1) που αφορά γενικές κατηγορίες μεταβολικών μονοπατιών, ως το ειδικότερο (Level 4) που αφορά ένζυμα. Η ανάλυση της σχετικής αφθονίας των λειτουργιών πραγματοποιήθηκε στο γενικότερο επίπεδο, Level 1. Αρχικά, όπως φαίνεται και στο bar plot της Εικόνας 33, οι κυριότερες λειτουργίες των οποίων η σχετική αφθονία μεταβλήθηκε στα Tumor δείγματα, συγκριτικά με τα Normal, είναι ο μεταβολισμός πρωτεϊνών (Protein Metabolism), η κυτταρική αναπνοή (Respiration), λειτουργίες

υδατανθράκων (Carbohydrates), η απόκριση οξειδωτικού στρες (Stress Response) και λειτουργίες αμινοξέων (Amino Acids and Derivatives).



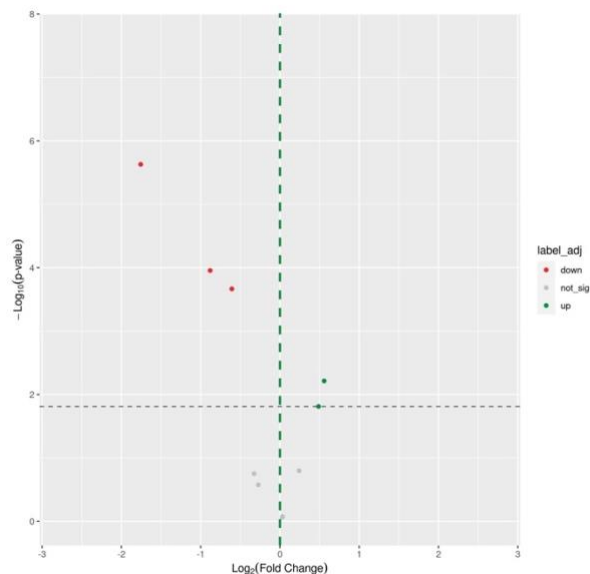
Εικόνα 33. Περιγραφική απεικόνιση της σχετικής αφθονίας των Level 1 βιολογικών λειτουργιών του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal (αριστερά) και Tumor (δεξιά).

Με την εφαρμογή στατιστικού ελέγχου Wilcoxon (Εικόνα 34), παρατηρείται πως, για τις 5 κυριότερες βιολογικές λειτουργίες Level 1, ανάμεσα στα Normal και τα Tumor δείγματα υπάρχει στατιστικά σημαντική διαφορά ($p < 0.05$) για τις λειτουργίες που σχετίζονται με το οξειδωτικό στρες και τα αμινοξέα. Ειδικότερα, λειτουργίες που αφορούν το μεταβολισμό πρωτεϊνών είναι αυξημένες στα Tumor δείγματα έναντι των Normal, ενώ για τις λειτουργίες που αφορούν την κυτταρική αναπνοή παρατηρείται το αντίθετο.



Εικόνα 34. Στατιστικός έλεγχος Wilcoxon ($p < 0.05$) της σχετικής αφθονίας των βιολογικών λειτουργιών Level 1 του μικροβιώματος στις δύο κατηγορίες δειγμάτων, Normal και Tumor.

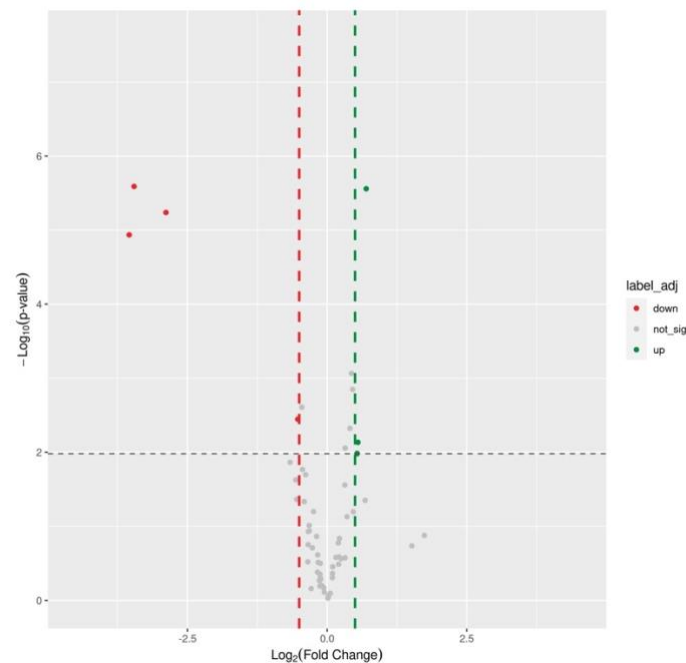
Η ανάλυση της διαφορικής αφθονίας των βιολογικών λειτουργιών, πραγματοποιήθηκε με το εργαλείο edgeR, για το ειδικότερο επίπεδο, Level 4, που αφορά ένζυμα που εκφράζονται από τα μικροβιακά γονίδια. Αρχικά, για τη σύγκριση Tumor vs Normal εντοπίστηκαν 5 λειτουργίες σε διαφορεική αφθονία, εκ των οποίων 2 είναι υπερ-εκφρασμένες και 3 υπο-εκφρασμένες στα Tumor δείγματα (Εικόνα 35). Συγκεκριμένα, η υπερ-εκφρασμένη πρωτεΐνη Chaperone protein DnaK ($\log_2FC = +0.56$) είναι μία πρωτεΐνη των βακτηρίων του γένους *Mycoplasma*, παρεμβαίνει στην ικανότητα των κυττάρων να εντοπίζουν και να επιδιορθώνουν βλάβες στο DNA και έχει συσχετιστεί με την ογκογένεση και εξέλιξη του καρκίνου (Benedetti et al., 2021). Ακόμη, παρατηρείται υπο-έκφραση του ενζύμου Phosphoenolpyruvate carboxykinase (EC 4.1.1.32) με $\log_2FC = -1.76$, του οποίου ο ρόλος είναι ο μεταβολισμός πυροσταφυλικού στον κύκλο κιτρικού οξέος, με σκοπό την παραγωγή γλυκόζης. Επομένως, η παρατηρούμενη υπο-έκφρασή του από το μικροβίωμα στα Tumor δείγματα συνάδει με το γεγονός ότι τα καρκινικά κύτταρα τείνουν να μετατρέπουν το πυροσταφυλικό σε γαλακτικό, παρακάμπτοντας τον κύκλο κιτρικού οξέος (Faubert et al., 2017).



Εικόνα 35. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Tumor vs Normal. Κάθε κουκίδα αντιστοιχεί σε μια λειτουργία. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται οι στατιστικά σημαντικά υπο-εκφρασμένες λειτουργίες στα Tumor δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται οι στατιστικά σημαντικά υπερ-εκφρασμένες λειτουργίες στα Tumor δείγματα, έναντι των Normal. Με γκρι χρώμα είναι οι λειτουργίες χωρίς στατιστικά σημαντική διαφορική αφθονία.

4.7.3 Σύγκριση του μικροβιώματος των καρκινικών υπο-ομάδων και εντοπισμός σχετικών μικροβιακών βιοδεικτών των καρκινικών υπο-ομάδων

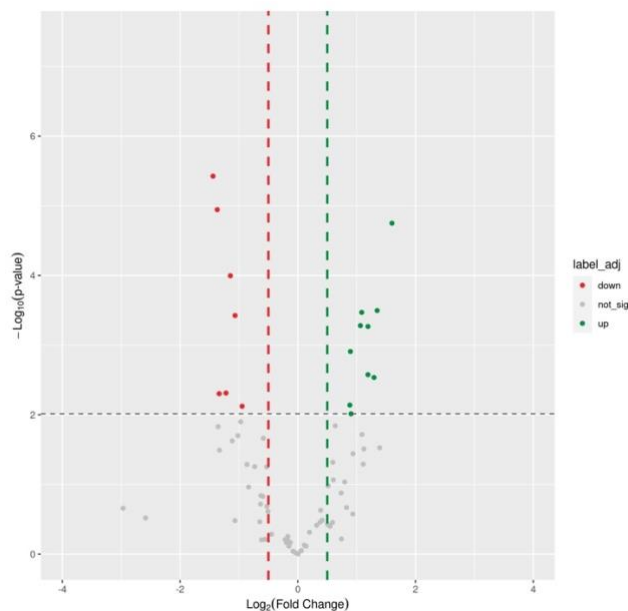
Η διαφορική αφθονία των βακτηριακών ειδών μελετήθηκε, για τις συγκρίσεις μεταξύ των καρκινικών υπο-ομάδων με τα δείγματα ελέγχου (Well vs Normal και Poorly vs Normal) και των καρκινικών υπο-ομάδων μεταξύ τους (Poorly vs Well). Αρχικά, από τη σύγκριση Well vs Normal, βρέθηκαν 8 βακτηριακά είδη σε διαφορική αφθονία, εκ των οποίων τα 4 υπο-εκφρασμένα, και 4 υπερ-εκφρασμένα (Εικόνα 36). Κατά τη σύγκριση αυτή, το είδος *Lactococcus lactis* είναι υπο-εκφρασμένο στα Well δείγματα ($\log_2FC = -3.71$). Όπως προαναφέρθηκε, το συγκεκριμένο είδος του φύλου *Proteobacteria* θεωρείται σημαντικό στα φυσιολογικά κύτταρα, λόγω της ικανότητας καταστολής ογκογενετικών δραστηριοτήτων.



Εικόνα 36. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Well vs Normal. Κάθε κουκίδα αντιστοιχεί σε ένα είδος. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται τα στατιστικά σημαντικά υπο-εκφρασμένα είδη στα Well δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται τα στατιστικά σημαντικά υπερ-εκφρασμένα είδη στα Well δείγματα, έναντι των Normal. Με γκρι χρώμα είναι τα είδη χωρίς στατιστικά σημαντική διαφορική αφθονία.

Στη συνέχεια, για τη σύγκριση δειγμάτων Poorly vs Normal, βρέθηκαν 17 στατιστικά σημαντικά είδη σε διαφορική αφθονία (Εικόνα 37). Από αυτή τη σύγκριση εντοπίζεται ο μεγαλύτερος αριθμός ειδών σε διαφορική αφθονία, συγκριτικά με τις υπόλοιπες συγκρίσεις (ακολουθεί η σύγκριση Well vs Normal και, τέλος, Poorly vs Well). Τα 7 από αυτά βρέθηκαν υπο-εκφρασμένα στα Poorly δείγματα, ενώ τα υπόλοιπα 10, υπερ-εκφρασμένα. Το βακτήριο *Aeromonas piscicola*, είναι υπερ-εκφρασμένο ($\log_2FC = +1.60$), και έχει συσχετιστεί με

ποικίλες ασθένειες του ήπατος, όπως κίρρωση ήπατος αλλά και καρκίνο (P.-L. Chen et al., 2016). Ακόμη, υπερ-εκφρασμένα στα Poorly δείγματα βρέθηκαν τα είδη *Mycoplasma ovipneumoniae* ($\log_2FC = +1.19$), και *Pseudomonas aeruginosa* ($\log_2FC = +1.35$). Πράγματι, έχει βρεθεί πως είδη του γένους *Mycoplasma* αναστέλλουν τη δραστηριότητα της πρωτεΐνης p53 και ενεργοποιούν ογκογενετικά μεταβολικά μονοπάτια, όπως το Wnt/ β -catenin (Yacoub et al., 2021). Όπως αναμενόταν, το *P. aeruginosa* είναι υπερ-εκφρασμένο στα Poorly δείγματα, καθώς αποτελεί έναν από τους πιο γνωστούς οργανισμούς ενδοογκικού μικροβιώματος και έχει συσχετιστεί με συμπτωματικές μεταστάσεις από πρωτοπαθείς κακοήθειες (Paprocka et al., 2022).



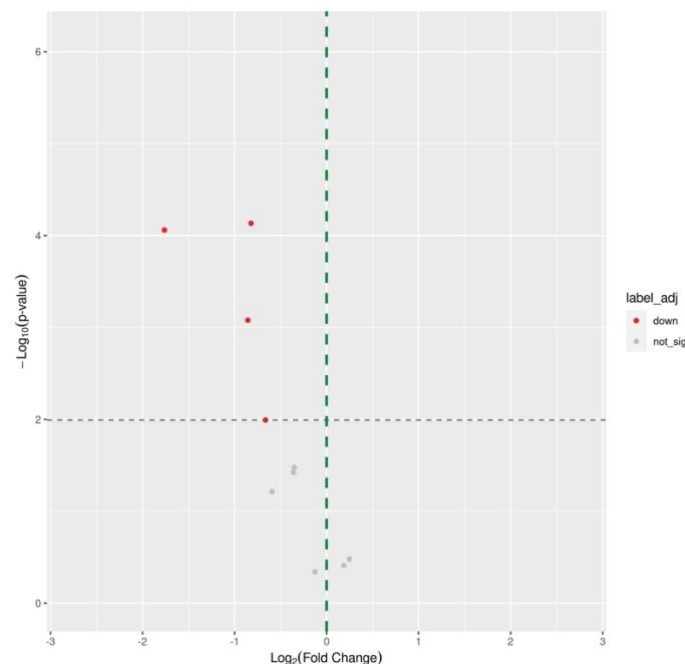
Εικόνα 37. Volcano plot ανάλυσης διαφορικής αφθονίας ειδών για τη σύγκριση Poorly vs Normal. Κάθε κουκίδα αντιστοιχεί σε ένα είδος. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται τα στατιστικά σημαντικά υπο-εκφρασμένα είδη στα Poorly δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται τα στατιστικά σημαντικά υπερ-εκφρασμένα είδη στα Poorly δείγματα, έναντι των Normal. Με γκρι χρώμα είναι τα είδη χωρίς στατιστικά σημαντική διαφορική αφθονία.

Η τελευταία σύγκριση με το edgeR, Poorly vs Well, δεν εντόπισε στατιστικά σημαντικά είδη σε διαφορική αφθονία. Το αποτέλεσμα αυτό ήταν αναμενόμενο, καθώς η ανάλυση πραγματοποιήθηκε ανάμεσα στις υποκατηγορίες των Tumor δειγμάτων. Ωστόσο, τα αποτελέσματα, παρόλο που δεν είναι στατιστικά σημαντικά, δίνουν ποικίλα πορίσματα όσον αφορά τις διαφοροποιήσεις του μικροβιώματος μεταξύ των βαθμών διαφοροποίησης του HCC. Συγκεκριμένα, για τα υπερ-εκφρασμένα είδη *P. aeruginosa* ($\log_2FC = +1.08$) και *M. ovipneumoniae* ($\log_2FC = +0.92$) παρατηρείται μικρότερη θετική τιμή \log_2FC , σε σχέση με τη σύγκριση Poorly vs Normal, καθιστώντας αυτούς τους μικροοργανισμούς εν δυνάμει

βιοδείκτες για την κατηγοριοποίηση καρκινικών δειγμάτων βάσει της κυτταρικής διαφοροποίησης. Ακόμη, έχει ενδιαφέρον η υπερ-έκφραση του *Mycobacterium tuberculosis* ($\log_2FC = +0.57$), το οποίο έχει χαρακτηριστεί ως 1^{ης} τάξης καρκινογόνο βακτήριο, και υπεύθυνο, συγκεκριμένα, για την ανάπτυξη ηπατικού καρκίνου (Roy et al., 2021).

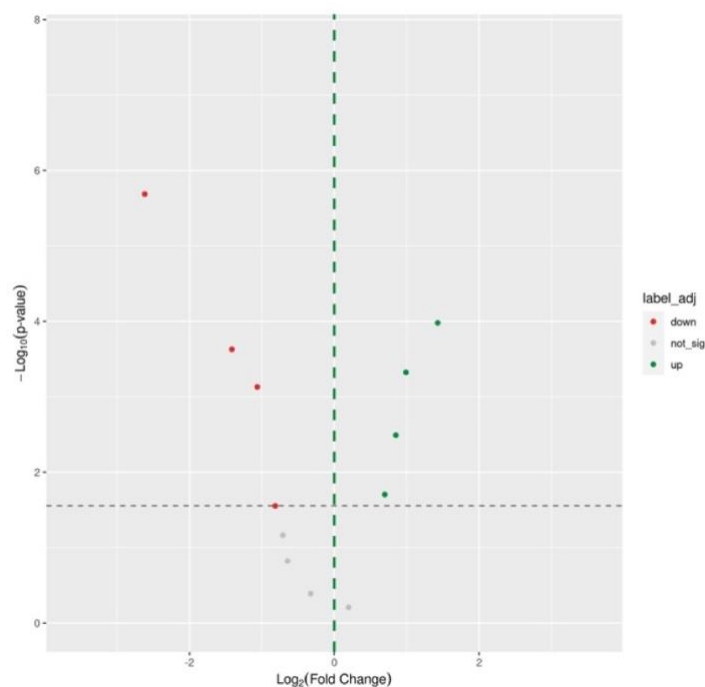
4.7.4 Σύγκριση Βιολογικών Λειτουργιών Μικροβιώματος μεταξύ καρκινικών υπο-ομάδων και ως προς τους φυσιολογικούς ιστούς

Στη σύγκριση Well vs Normal, βρέθηκαν 4 ένζυμα σε διαφορεική αφθονία, υπο-εκφρασμένα στα Well δείγματα, όπως παρουσιάζεται στο volcano plot της Εικόνας 38. Τα ένζυμα Phosphoenolpyruvate carboxykinase ($\log_2FC = -1.76$) και Aldehyde dehydrogenase (EC 1.2.1.3) ($\log_2FC = -0.86$) εντοπίστηκαν και κατά την προηγούμενη σύγκριση, επιβεβαιώνοντας την ελάττωση της διαφορεικής αφθονίας τους ακόμη και στα καλώς διαφοροποιημένα καρκινικά κύτταρα.



Εικόνα 38. Volcano plot ανάλυσης διαφορεικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Well vs Normal. Κάθε κουκίδα αντιστοιχεί σε μια λειτουργία. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται οι στατιστικά σημαντικά υπο-εκφρασμένες λειτουργίες στα Well δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται οι στατιστικά σημαντικά υπερ-εκφρασμένες λειτουργίες στα Well δείγματα, έναντι των Normal. Με γκρι χρώμα είναι οι λειτουργίες χωρίς στατιστικά σημαντική διαφορεική αφθονία.

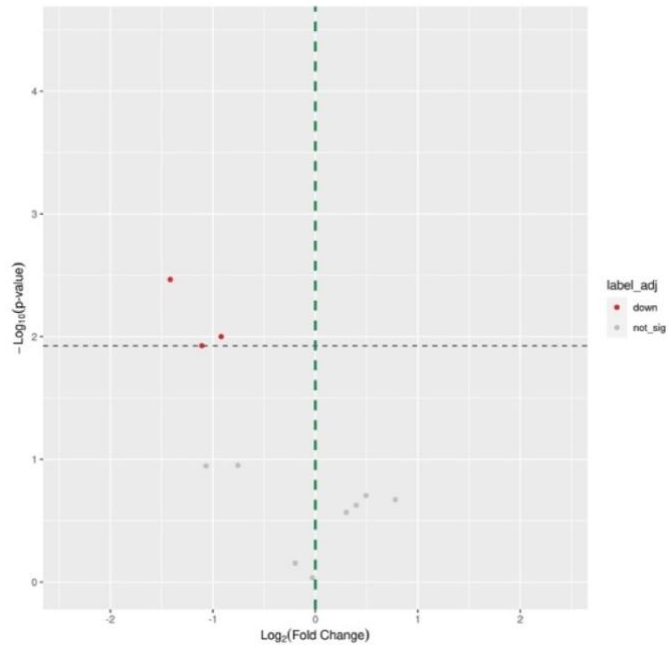
Κατά την ανάλυση των δειγμάτων Poorly vs Normal, βρέθηκαν 8 ένζυμα σε διαφορετική αφθονία, με 4 από αυτά να είναι υπο-εκφρασμένα και 4 υπερ-εκφρασμένα στα Poorly δείγματα έναντι των Normal (Εικόνα 39). Πέραν από την υπερ-έκφραση της Chaperone protein DnaK, που αναλύθηκε και κατά τη σύγκριση Tumor vs Normal, ένα ακόμη υπερ-εκφρασμένο ένζυμο είναι η ATP συνθάση (ATP synthase, EC 3.6.3.14), με $\log_2FC = 0.85$. Πράγματι, το ένζυμο αυτό παράγει μεγάλες ποσότητες ATP, που αποτελεί την κινητήριο δύναμη του κυττάρου. Η έκτοπη δραστηριότητά της στην κυτταρική μεμβράνη, όμως, συνδέεται με λειτουργίες όπως η αγγειογένεση, η ογκογένεση, και η μετάσταση του καρκίνου (T. Wang et al., 2021). Επιπλέον, ένα υπο-εκφρασμένο ένζυμο στα Poorly δείγματα είναι η Cytochrome C oxidase (EC 1.9.3.1) με $\log_2FC = -1.06$, η οποία παράγει ενέργεια καταλύοντας αντιδράσεις οξείδωσης. Είναι γνωστό πως η καταστολή της Cytochrome C oxidase οδηγεί στην τάση του μεταβολισμού προς γλυκόλυση, και κατά συνέπεια στην καρκινογένεση (Sharma et al., 2017; Srinivasan et al., 2016).



Εικόνα 39. Volcano plot ανάλυσης διαφορετικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Normal. Κάθε κουκίδα αντιστοιχεί σε μια λειτουργία. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται οι στατιστικά σημαντικά υπο-εκφρασμένες λειτουργίες στα Poorly δείγματα, έναντι των Normal. Δεξιά του άξονα $y=0$ (πράσινο) εντοπίζονται οι στατιστικά σημαντικά υπερ-εκφρασμένες λειτουργίες στα Poorly δείγματα, έναντι των Normal. Με γκρι χρώμα είναι οι λειτουργίες χωρίς στατιστικά σημαντική διαφορετική αφθονία.

Τέλος, κατά τη σύγκριση Poorly vs Well βρέθηκαν 3 υπο-εκφρασμένα ένζυμα σε διαφορετική αφθονία, όπως απεικονίζονται και στην Εικόνα 40. Ένα από αυτά είναι η Cytochrome C oxidase ($\log_2FC = -1.11$), υποδεικνύοντας ότι η καταστολή της λειτουργίας του

φαίνεται να επηρεάζει και το βαθμό διαφοροποίησης των καρκινικών κυττάρων. Ακόμη, το μέγεθος επίδρασης \log_2FC , της Aldehyde dehydrogenase (-1.41) είναι μεγαλύτερο, κατ' απόλυτη τιμή, από την αντίστοιχη τιμή της στη σύγκριση Well vs Normal (-0.86), υποδηλώνοντας την ενδιάμεση λειτουργικότητα των Well δειγμάτων.



Εικόνα 40. Volcano plot ανάλυσης διαφορικής αφθονίας των βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Well. Κάθε κουκίδα αντιστοιχεί σε μια λειτουργία. Αριστερά του άξονα $y=0$ (κόκκινο) εντοπίζονται οι στατιστικά σημαντικά υπο-εκφρασμένες λειτουργίες στα Poorly δείγματα, έναντι των Normal. Με γκρι χρώμα είναι οι λειτουργίες χωρίς στατιστικά σημαντική διαφορική αφθονία.

4 Περιορισμοί της μελέτης

- Κατά την αρχική ανάλυση των δειγμάτων είχε πραγματοποιηθεί εμπλουτισμός σε πολυαδενυλιωμένα μετάγραφα (Jin et al., 2019). Η μέθοδος αυτή εφαρμόζεται κυρίως σε περιπτώσεις όπου επιθυμείται η ανάλυση συνολικού ευκαρυωτικού mRNA που φέρει αυτή την μετα-μεταγραφική τροποποίηση της πολυαδενυλίωσης. Καθώς το mRNA των βακτηρίων δε διαθέτει αυτό το χαρακτηριστικό, είναι πιθανό από τις διαθέσιμες αλληλουχίες να έχουν περιοριστεί τα βακτηριακά μετάγραφα μετά το στάδιο του εμπλουτισμού. Αντίθετα, η εξάλειψη του ριβοσωμικού RNA, κατά την οποία αφαιρούνται αλληλουχίες rRNA, οι οποίες βρίσκονται σε μεγάλο ποσοστό στο ανθρώπινο και βακτηριακό μεταγράφομα και επηρεάζουν την ανάλυση, είναι επιθυμητή. Τα δεδομένα της τελευταίας μεθόδου καθιστούν ευκολότερη την ανάλυση κυτταρικών μεταγραφικών λειτουργιών (S. Zhao et al., 2018). Η προσέγγιση της χρησιμοποίησης των μη ευθυγραμμισμένων αναγνωσμάτων που προκύπτουν μετά την ευθυγράμμιση στο ανθρώπινο γονιδίωμα με το εργαλείο HISAT2, εμπεριέχει την πιθανότητα να μην έχουν διαχωριστεί πλήρως ανθρώπινες αλληλουχίες, όπως επαναλαμβανόμενες ή υπερ-επεξεργασμένο RNA, καθιστώντας τα μη επεξεργασμένα αναγνώσματα μη αμιγώς μικροβιακά. Προκειμένου να περιοριστούν αυτά τα υπολείμματα ανθρώπινων αλληλουχιών στα μη ευθυγραμμισμένα αναγνώσματα, αυστηροποιήθηκαν οι παράμετροι ευθυγράμμισης με το εργαλείο DIAMOND, όπως το ποσοστό ορθών ευθυγραμμίσεων (από 50% σε 90%) και το στατιστικό E-value (από 10^{-1} σε 10^{-4}).
- Ο περιορισμένος αριθμός των ασθενών από όπου προήλθαν τα δεδομένα περιορίζει την εξαγωγή στατιστικά σημαντικών και άρα έγκυρων αποτελεσμάτων. Η έλλειψη κλινικών δεδομένων για τα δείγματα αυτά περιόρισε επίσης την ανάλυση όσον αφορά τις συσχετίσεις που θα μπορούσαν να πραγματοποιηθούν ανάμεσα στο μικροβίωμα και στοιχεία όπως η ηλικία, το στάδιο/βαθμός διαφοροποίησης του καρκίνου, σοβαρότητα συμπτωμάτων, δεδομένα επιβίωσης κ.ο.κ. Ωστόσο, με την Ανάλυση Εμπλουτισμού Συνόλου Γονιδίων επιτεύχθηκε ομαδοποίηση των ασθενών ως προς το βαθμό διαφοροποίησης των Tumor δειγμάτων, βάσει της συσχέτισης έκφρασης των ανθρώπινων γονιδίων στα καρκινικά δείγματα με γνωστές γονιδιακές υπογραφές του ήπατος.

- Κατά την ανάλυση της διαφορικής αφθονίας στη σύγκριση Poorly vs Well δειγμάτων παρατηρήθηκε περιορισμένος αριθμός στατιστικά σημαντικών βακτηριακών ειδών και βιολογικών λειτουργιών, ενδεχομένως και λόγω του μικρού αριθμού Poorly δειγμάτων ($n = 6$), σε σχέση με τα Well ($n = 17$) που δεν επέτρεψε τη διεξαγωγή ενός μεγάλου και έγκυρου εύρους αποτελεσμάτων.

5 Συμπεράσματα

Η ανάλυση της έκφρασης των ανθρώπινων γονιδίων επέτρεψε τη διάκριση των καρκινικών δειγμάτων σε δύο φαινοτύπους, υψηλής και χαμηλής διαφοροποίησης. Η ομαδοποίηση αυτή προκύπτει από την Ανάλυση Εμπλουτισμού Συνόλου Γονιδίων.

Η χρήση της βάσης δεδομένων NCBI RefSeq στο εργαλείο SAMSA2 διευκολύνει την ταξινόμηση των ευθυγραμμισμένων γονιδίων του μικροβιώματος σε οργανισμούς, καθιστώντας την ανάλυση γρήγορη και εύκολη. Ακόμη, η αντίστοιχη των γονιδίων με τη χρήση της βάσης δεδομένων SEED Subsystems σε βιολογικές λειτουργίες δίνει ένα μεγάλο εύρος πληροφορίας αναφορικά με τον πιθανό ρόλο του ενδο-ογκικού μικροβιώματος στο HCC.

Τα αποτελέσματα της διαφορικής γονιδιακής έκφρασης που λαμβάνονται από τη σύγκριση Tumor vs Normal δειγμάτων, μέσω των αντίστοιχων διαγραμμάτων και πινάκων, επιβεβαιώνουν προηγούμενες συγκριτικές μελέτες γονιδιακής έκφρασης μεταξύ φυσιολογικών και καρκινικών ιστών. Ακόμη, όσον αφορά τη σύγκριση Poorly vs Well δειγμάτων, τα αποτελέσματα προβάλλουν μία ειδικότερη διαφορά στη γονιδιακή έκφραση, ανάλογη με τη διαφοροποίηση των καρκινικών κυττάρων. Πολλά από τα γονίδια που εντοπίστηκαν ως διαφορετικά εκφρασμένα, και στις 2 συγκρίσεις, επιβεβαιώνεται από τη βιβλιογραφία πως σχετίζονται με την ογκογένεση και την ανάπτυξη καρκίνου.

Στο πλαίσιο της λειτουργικής ανάλυσης των ανθρώπινων γονιδίων, το εργαλείο BioInfoMiner δίνει πληροφορίες τόσο σε όρους Γονιδιακής Οντολογίας (GO), όσο και σε όρους βιολογικών μονοπατιών Reactome, εντοπίζοντας και τα διαφορετικά εκφρασμένα γονίδια με το σημαντικότερο λειτουργικό ρόλο. Η υπερ-έκφραση των ανοσοσφαιρινών στα Tumor δείγματα υποδηλώνει την καταστολή του ανοσοποιητικού μηχανισμού στους ασθενείς.

Κατά την ανάλυση της σχετικής αφθονίας των ειδών του μικροβιώματος στα δείγματα, τα κυριότερα φύλα βακτηρίων που εμφανίζονται είναι τα *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Chlamydiae* και *Tenericutes*. Η άλφα ποικιλομορφία στα Tumor δείγματα, είναι μειωμένη σε σχέση με τα Normal, με τη χρήση των δεικτών Shannon και Simpson, δίχως όμως στατιστικά σημαντική διαφορά. Όσον αφορά την ανάλυση της βήτα ποικιλομορφίας, παρατηρήθηκε αυξημένη μεταβλητότητα μεταξύ των Tumor δειγμάτων, σε σχέση με τα Normal.

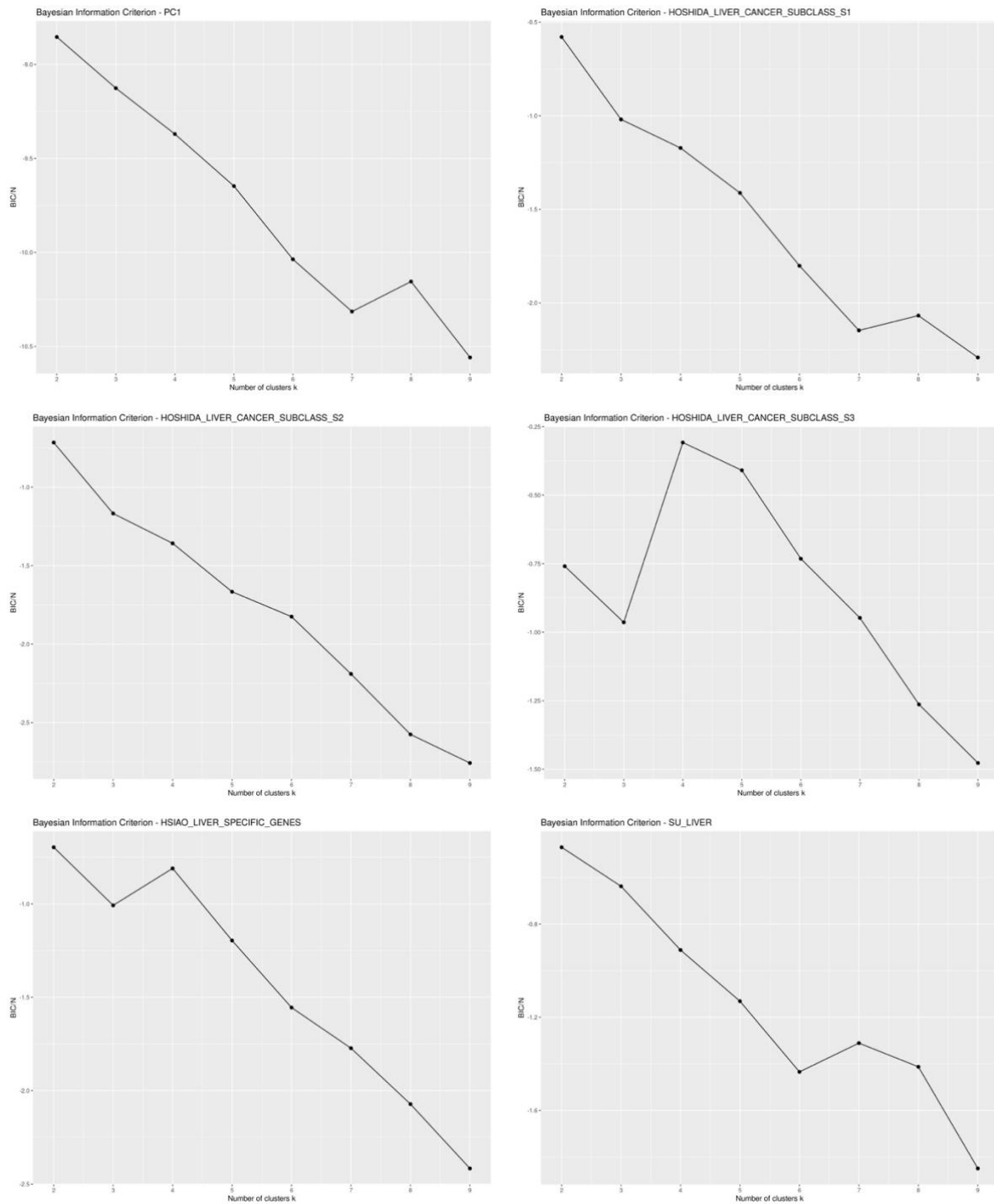
Κατά την ανάλυση της διαφορικής αφθονίας στη σύγκριση Tumor vs Normal, βρέθηκαν υπερ-εκφρασμένα είδη του γένους *Aeromonas*, τα οποία έχει επιβεβαιωθεί πως είναι παρόντα σε καρκινικούς ιστούς, προωθώντας την ανάπτυξη της κακοήθειας. Επίσης, βρέθηκε υπο-εκφρασμένο το *Lactococcus lactis*, του οποίου οι ιδιότητες έχουν αποδειχθεί αντικαρκινικές.

Συμπληρωματικά, βακτήρια από τα γένη *Aeromonas* και *Mycoplasma* βρέθηκαν εμπλουτισμένα στα δείγματα Poorly, έναντι των Normal. Για αυτά τα βακτηριακά είδη, έχει αποδειχθεί πως η παρουσία τους σχετίζεται με την ανάπτυξη καρκίνου. Ακόμη, το βακτήριο *Lactococcus lactis* βρέθηκε σε διαφορεική αφθονία στα Normal δείγματα, τόσο στη σύγκριση Well vs Normal, όσο και στη Poorly vs Normal, επιβεβαιώνοντας τις αντικαρκινικές του ιδιότητες που αναφέρονται στη βιβλιογραφία.

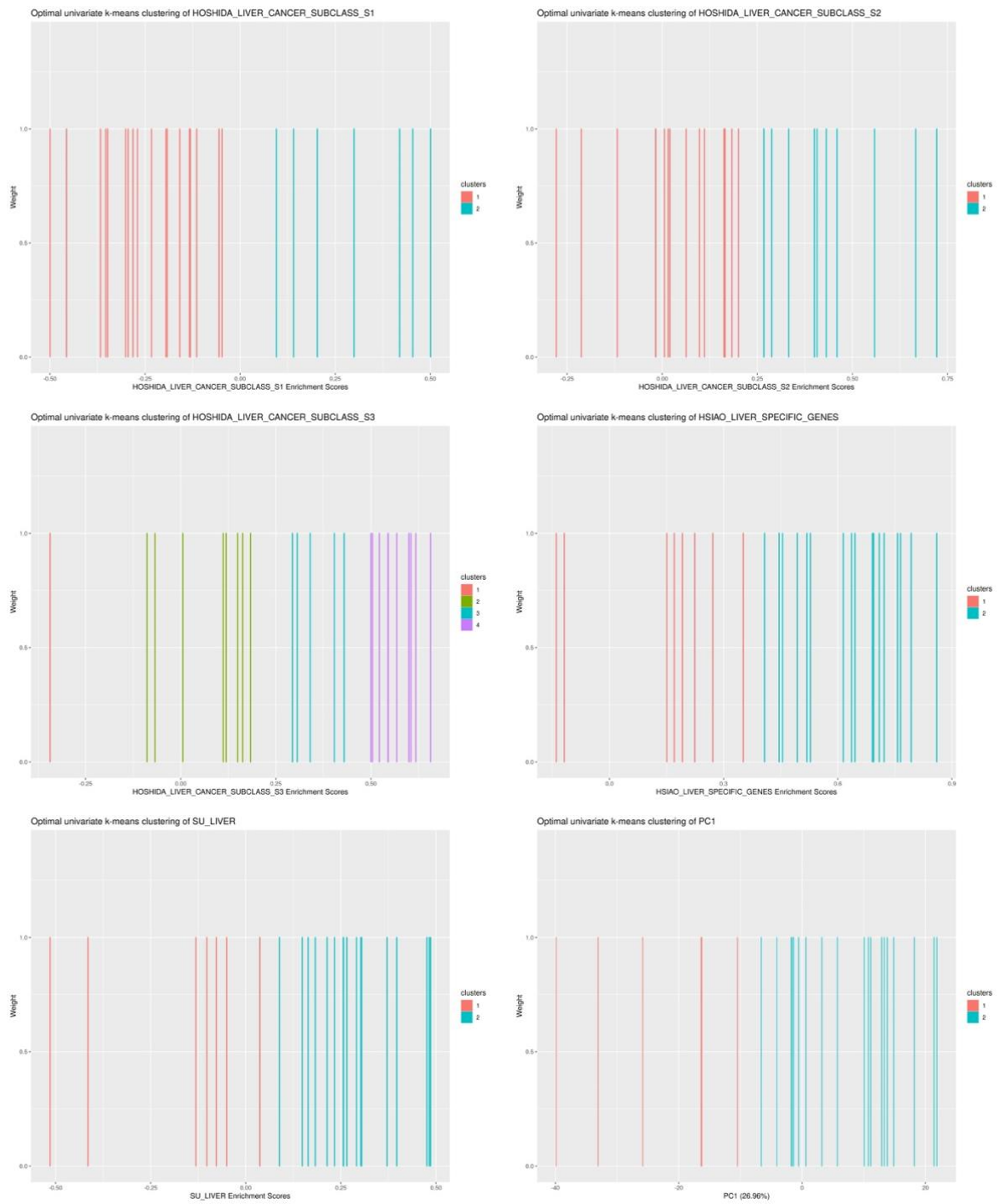
Η ανάλυση της διαφορικής αφθονίας των βιολογικών λειτουργιών με το edgeR έδωσε πληροφορίες τόσο για 3την έκφραση ενζύμων στα Tumor δείγματα έναντι των Normal, όσο και για την έκφραση αυτών στους διαφορετικούς βαθμούς διαφοροποίησης. Ένζυμα όπως τα Phosphopyruvate carboxykinase, Aldehyde dehydrogenase και Cytochrome c oxidase, τα οποία σχετίζονται με μονοπάτια παραγωγής ενέργειας στα υγιή κύτταρα, βρέθηκαν υπο-εκφρασμένα, με στατιστική σημαντικότητα, στα καρκινικά δείγματα. Το ένζυμο ATP synthase εντοπίστηκε υπερ-εκφρασμένο στα Tumor δείγματα, σε συμφωνία με ευρήματα που το συσχετίζουν με δραστηριότητες όπως η ογκογένεση και η αγγειογένεση.

Η ανάλυση της διαφορικής αφθονίας ανάμεσα στις καρκινικές υπο-ομάδες και τα Normal δείγματα, αλλά και μεταξύ των υπο-ομάδων, επισημαίνει τα σημαντικά βακτηριακά είδη και ένζυμα, με τις τιμές των \log_2FC να υποδηλώνουν τον ενδιάμεσο χαρακτήρα των Well δειγμάτων.

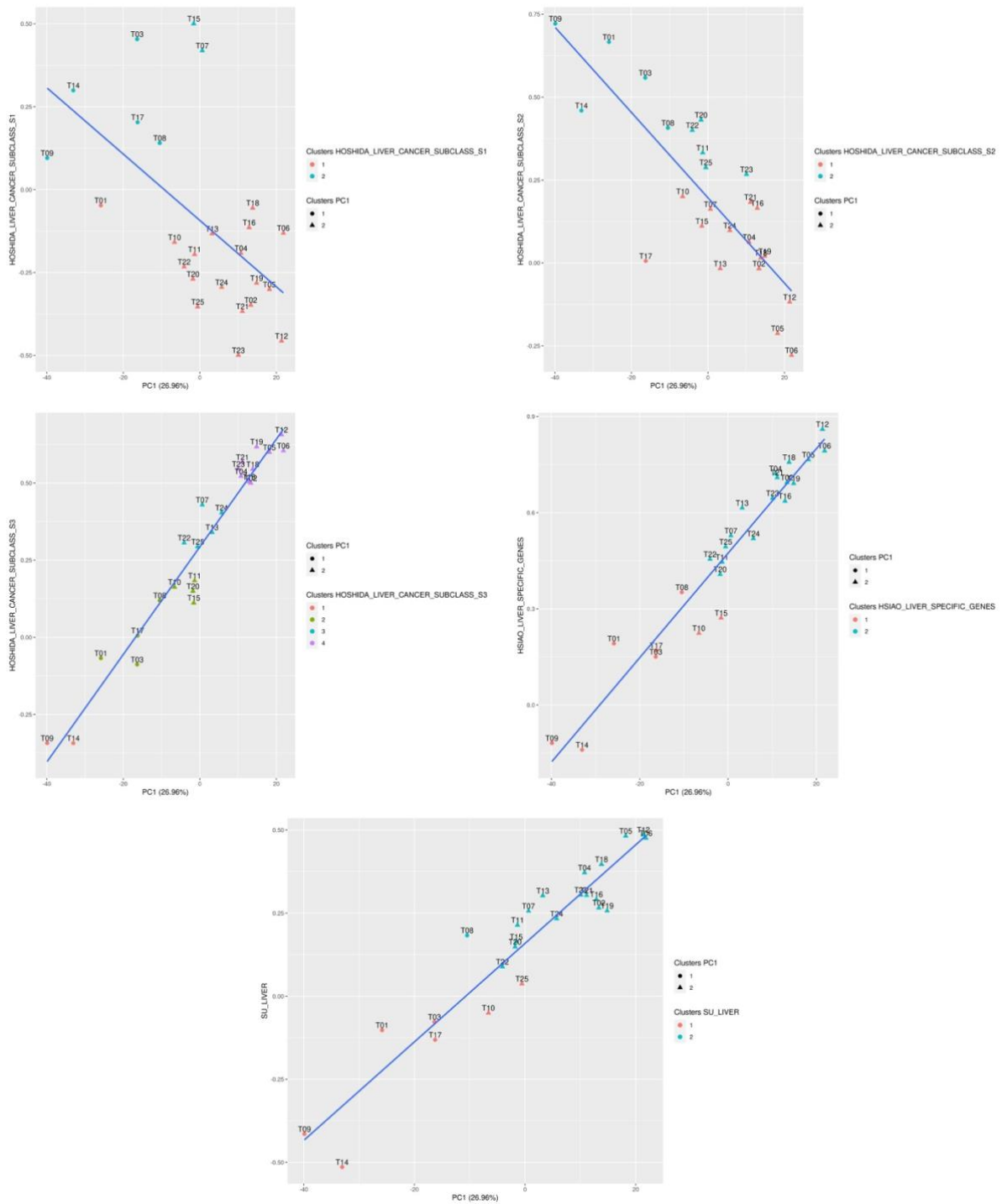
Παράρτημα



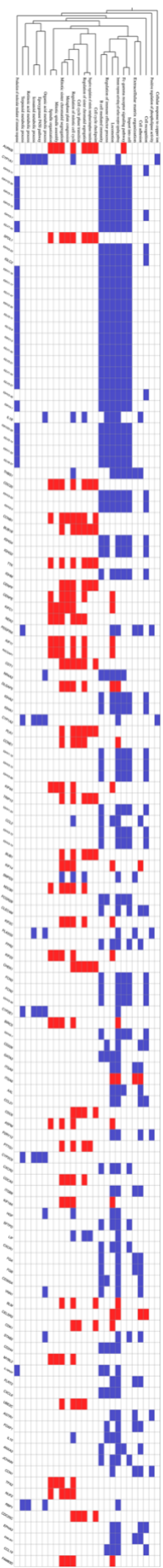
Εικόνα 41. Εύρεση βέλτιστου αριθμού υπο-ομάδων με το κριτήριο BIC. Από τα αριστερά προς τα δεξιά, και από τα άνω προς τα κάτω: PCI, Hoshida Subclass S1, S2, S3, HSHAO Liver Specific Genes, SU Liver.



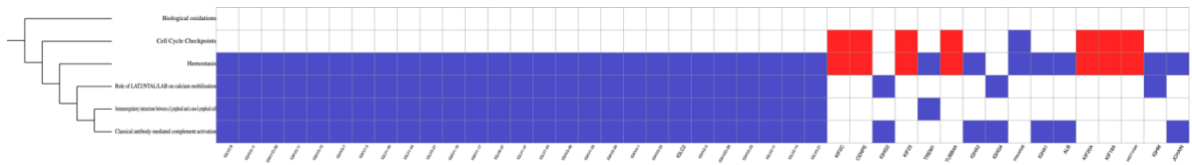
Εικόνα 42. Αποτελέσματα ομαδοποίησης βάσει του βαθμού εμπλουτισμού των δειγμάτων στα σύνολα γονιδίων. Από τα αριστερά προς τα δεξιά, και από τα άνω προς τα κάτω: Hoshida Subclass S1, S2, S3, HSIAO Liver Specific Genes, SU Liver, PC1.



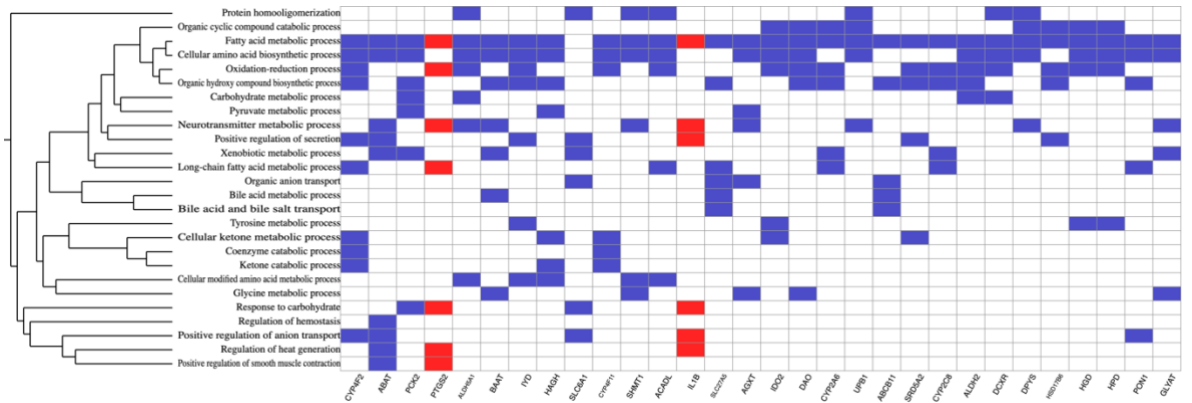
Εικόνα 43. Αποτελέσματα ομαδοποίησης των καρκινικών δειγμάτων βάσει του βαθμού εμπλουτισμού με των συνόλων γονιδίων στην PC1. Από τα αριστερά προς τα δεξιά, και από τα άνω προς τα κάτω: Hoshida Subclass S1, S2, S3, HSIAO Liver Specific Genes, SU Liver.



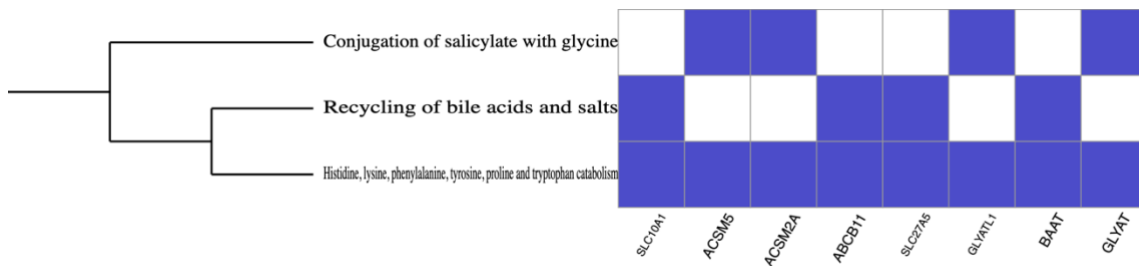
Εικόνα 44. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης Tumor-Normal με βιολογικές λειτουργίες σε όρους GO από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner. Με μπλε χρώμα παρουσιάζονται τα γονίδια που είναι υπο-εκφρασμένα στα καρκινικά δείγματα ενώ με κόκκινο χρώμα τα γονίδια που είναι υπερ-εκφρασμένα στα καρκινικά δείγματα.



Εικόνα 45. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης Tumor-Normal με βιολογικές λειτουργίες σε όρους Reactome Pathways από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner. Με μπλε χρώμα παρουσιάζονται τα γονίδια που είναι υπο-εκφρασμένα στα καρκινικά δείγματα ενώ με κόκκινο χρώμα τα γονίδια που είναι υπερ- εκφρασμένα στα καρκινικά δείγματα.



Εικόνα 46. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης των υπο-ομάδων καρκινικών δειγμάτων, Poorly-Well, με βιολογικές λειτουργίες σε όρους Γονιδιακής Οντολογίας (Gene Ontology) από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner. Με μπλε χρώμα παρουσιάζονται τα γονίδια που είναι υπο-εκφρασμένα στα Poorly δείγματα, ενώ με κόκκινο χρώμα τα γονίδια που είναι υπερ-εκφρασμένα στα Poorly δείγματα.



Εικόνα 47. Heatmap συσχέτισης των κυριότερων διαφορικά εκφρασμένων γονιδίων της σύγκρισης των υπο-ομάδων καρκινικών δειγμάτων, Poorly-Well, με βιολογικές λειτουργίες σε όρους Reactome Pathways από το εργαλείο λειτουργικού χαρακτηρισμού BioInfoMiner. Με μπλε χρώμα παρουσιάζονται τα γονίδια και οι αντίστοιχες λειτουργίες που είναι υπο-εκφρασμένες στα Poorly δείγματα.

Πίνακας 5. Βαθμοί εμπλουτισμού (Enrichment Scores) των συνόλων γονιδίων στην PC1.

Gene Signature	PC1 Enrichment Loading Scores
HOSHIDA_LIVER CANCER SUBCLASS S1	-0,3562021
HOSHIDA_LIVER CANCER SUBCLASS S2	-0,6439312
HOSHIDA_LIVER CANCER SUBCLASS S3	0,8901471
HSIAO LIVER SPECIFIC GENES	0,8892632
SU LIVER	0,7763024

Πίνακας 6. Βαθμοί εμπλουτισμού (Enrichment Scores) της έκφρασης των Tumor δειγμάτων στα σύνολα γονιδίων.

Samples	HOSHIDA SUBCLASS S1	HOSHIDA SUBCLASS S2	HOSHIDA SUBCLASS S3	HSIAO LIVER SPECIFIC GENES	SU LIVER
T01	-0,047601886	0,666419266	-0,067834098	0,191481302	-0,102338348
T02	-0,348172933	-0,017103971	0,499711813	0,694178627	0,26600753
T03	0,453884308	0,55852176	-0,088379383	0,150417708	-0,077502493
T04	-0,191543419	0,0629716	0,522012861	0,721938531	0,371600028
T05	-0,300972194	-0,212481758	0,599530639	0,765206036	0,482192376
T06	-0,131014238	-0,278352201	0,604835148	0,79301725	0,475780904
T07	0,419319049	0,162229465	0,429372026	0,52816047	0,256393535
T08	0,140545676	0,407216093	0,119410278	0,351511002	0,182935818
T09	0,095395115	0,721647799	-0,343048957	-0,118934863	-0,414606391
T10	-0,158691019	0,200582708	0,162811298	0,223869628	-0,050227751
T11	-0,19514053	0,332439796	0,183518797	0,445844882	0,213962872
T12	-0,456444694	-0,118045842	0,656593347	0,860142052	0,485766058
T13	-0,133254601	-0,016939936	0,340378874	0,614629503	0,302597741
T14	0,299306164	0,459535413	-0,343406653	-0,139857948	-0,514233942
T15	0,500482463	0,111147609	0,111608258	0,271345618	0,164088269
T16	-0,114306711	0,165300678	0,504041986	0,636278739	0,291504719
T17	0,203047302	0,006057255	0,005402078	0,170211931	-0,131127862
T18	-0,055341906	0,015742208	0,54469268	0,757075212	0,396886092
T19	-0,28169264	0,020295752	0,617706245	0,691072692	0,257099611
T20	-0,269782617	0,431261906	0,149441265	0,407492227	0,148707893
T21	-0,367022852	0,183354839	0,567721295	0,709326983	0,303336582
T22	-0,232811582	0,400170632	0,306253144	0,454953968	0,088677973
T23	-0,499517537	0,267198573	0,545491545	0,645192813	0,304651554
T24	-0,294505907	0,097960143	0,403706127	0,519098749	0,23340633
T25	-0,353342492	0,287944439	0,293745322	0,493690242	0,037321291

Πίνακας 7. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τις εμπλεκόμενες βιολογικές διεργασίες GO από την ανάλυση διαφορικής έκφρασης μεταξύ Tumor και Normal δειγμάτων.

Term id	Term Definition	Enrichment	Corrected pvalue
GO:0002455	humoral immune response mediated by circulating immunoglobulin	48/154	0,0007
GO:0006956	complement activation	53/165	0,0008
GO:0006958	complement activation, classical pathway	47/148	0,0012
GO:0030449	regulation of complement activation	40/112	0,0014
GO:0002920	regulation of humoral immune response	42/131	0,0019
GO:0016064	immunoglobulin mediated immune response	50/190	0,0023
GO:0019724	B cell mediated immunity	50/192	0,0023
GO:0006959	humoral immune response	65/341	0,0034
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	53/259	0,0039
GO:0000278	mitotic cell cycle	94/684	0,0042
GO:0002449	lymphocyte mediated immunity	53/248	0,0045
GO:1903047	mitotic cell cycle process	86/587	0,0047
GO:0002431	Fc receptor mediated stimulatory signaling pathway	33/130	0,005
GO:0006909	phagocytosis	53/308	0,0058
GO:0006898	receptor-mediated endocytosis	49/273	0,006
GO:0007059	chromosome segregation	47/260	0,0065

GO:0038094	Fc-gamma receptor signaling pathway	32/132	0,0079
GO:0050900	leukocyte migration	57/367	0,0082
GO:0000280	nuclear division	47/277	0,0084
GO:0098813	nuclear chromosome segregation	40/209	0,0089
GO:0140014	mitotic nuclear division	31/140	0,0089
GO:0002250	adaptive immune response	73/541	0,0093
GO:0048285	organelle fission	47/302	0,0097
GO:0007052	mitotic spindle organization	21/72	0,0106
GO:0022402	cell cycle process	102/958	0,0111
GO:0000819	sister chromatid segregation	29/133	0,0113
GO:0051783	regulation of nuclear division	37/210	0,0115
GO:0002253	activation of immune response	73/603	0,0123
GO:0000070	mitotic sister chromatid segregation	25/106	0,0127
GO:0002697	regulation of immune effector process	60/453	0,013
GO:0002377	immunoglobulin production	30/147	0,0144
GO:0038095	Fc-epsilon receptor signaling pathway	31/167	0,0145
GO:1902850	microtubule cytoskeleton organization involved in mitosis	23/95	0,0151
GO:0051301	cell division	63/496	0,0154
GO:0007088	regulation of mitotic nuclear division	34/185	0,0158
GO:0006897	endocytosis	65/533	0,0162
GO:0006910	phagocytosis, recognition	22/93	0,0171
GO:0002684	positive regulation of immune system process	111/1118	0,0172
GO:0098657	import into cell	72/648	0,0174
GO:0002768	immune response-regulating cell surface receptor signaling pathway	58/460	0,0176
GO:0038093	Fc receptor signaling pathway	36/232	0,0176
GO:0007049	cell cycle	122/1316	0,0185
GO:0002440	production of molecular mediator of immune response	31/169	0,0195
GO:0006911	phagocytosis, engulfment	23/114	0,0198
GO:0002429	immune response-activating cell surface receptor signaling pathway	55/426	0,0202
GO:0010965	regulation of mitotic sister chromatid separation	16/58	0,0212
GO:0007166	cell surface receptor signaling pathway	198/2439	0,0219
GO:0051983	regulation of chromosome segregation	22/103	0,022
GO:0030071	regulation of mitotic metaphase/anaphase transition	15/53	0,0223
GO:0099024	plasma membrane invagination	24/123	0,0225
GO:0002764	immune response-regulating signaling pathway	64/570	0,0232
GO:0022610	biological adhesion	91/925	0,0233
GO:0002682	regulation of immune system process	143/1659	0,0233
GO:0002757	immune response-activating signal transduction	61/534	0,0235
GO:0007155	cell adhesion	91/919	0,0238
GO:0010324	membrane invagination	24/130	0,0244
GO:1902099	regulation of metaphase/anaphase transition of cell cycle	15/55	0,0248
GO:0042573	retinoic acid metabolic process	10/25	0,0262
GO:1905818	regulation of chromosome separation	16/63	0,0265
GO:0008037	cell recognition	33/224	0,0269
GO:0034754	cellular hormone metabolic process	22/112	0,0278
GO:0019373	epoxygenase P450 pathway	9/20	0,028
GO:0016477	cell migration	91/935	0,0286
GO:0050778	positive regulation of immune response	82/826	0,0287
GO:0050776	regulation of immune response	104/1113	0,0291
GO:0051303	establishment of chromosome localization	16/71	0,0294
GO:0007093	mitotic cell cycle checkpoint	26/162	0,0299
GO:0051310	metaphase plate congression	14/53	0,0303
GO:0033047	regulation of mitotic sister chromatid segregation	16/69	0,0313
GO:0009617	response to bacterium	70/692	0,0318
GO:2000816	negative regulation of mitotic sister chromatid separation	10/28	0,0325
GO:0033045	regulation of sister chromatid segregation	17/81	0,0331
GO:0090307	mitotic spindle assembly	11/34	0,0332
GO:0044772	mitotic cell cycle phase transition	35/260	0,0337
GO:0042742	defense response to bacterium	40/323	0,0345
GO:0050000	chromosome localization	16/72	0,0348
GO:0040011	locomotion	112/1278	0,035
GO:0019369	arachidonic acid metabolic process	13/52	0,0354

GO:1905819	negative regulation of chromosome separation	10/29	0,0357
GO:0007051	spindle organization	23/138	0,0359
GO:0032787	monocarboxylic acid metabolic process	54/505	0,0367
GO:0051784	negative regulation of nuclear division	13/50	0,0372
GO:0044770	cell cycle phase transition	35/268	0,0372
GO:0016053	organic acid biosynthetic process	37/295	0,0381
GO:0007080	mitotic metaphase plate congression	12/43	0,0385
GO:0046394	carboxylic acid biosynthetic process	37/294	0,0397
GO:0050865	regulation of cell activation	64/634	0,0399
GO:0043062	extracellular structure organization	44/380	0,0399
GO:0007346	regulation of mitotic cell cycle	67/677	0,0405
GO:0033048	negative regulation of mitotic sister chromatid segregation	10/31	0,0408
GO:0010564	regulation of cell cycle process	76/803	0,0411
GO:0050853	B cell receptor signaling pathway	20/115	0,0417
GO:0000075	cell cycle checkpoint	28/201	0,0421
GO:0045132	meiotic chromosome segregation	17/89	0,0422
GO:0006928	movement of cell or subcellular component	126/1506	0,0424
GO:0002443	leukocyte mediated immunity	72/747	0,0426
GO:0045839	negative regulation of mitotic nuclear division	11/41	0,0444
GO:0051985	negative regulation of chromosome segregation	10/34	0,0445
GO:0033046	negative regulation of sister chromatid segregation	10/33	0,0455
GO:0071280	cellular response to copper ion	9/28	0,0461
GO:0045841	negative regulation of mitotic metaphase/anaphase transition	9/26	0,0462
GO:0043436	oxoacid metabolic process	87/983	0,0467
GO:0050867	positive regulation of cell activation	45/412	0,047
GO:0006082	organic acid metabolic process	89/1004	0,048
GO:0030198	extracellular matrix organization	38/334	0,0483
GO:0002694	regulation of leukocyte activation	57/591	0,0485
GO:0010518	positive regulation of phospholipase activity	13/58	0,0487
GO:0006690	icosanoid metabolic process	17/102	0,0487
GO:0006952	defense response	112/1319	0,0488
GO:0019752	carboxylic acid metabolic process	80/893	0,049
GO:1902100	negative regulation of metaphase/anaphase transition of cell cycle	9/27	0,0496
GO:0006721	terpenoid metabolic process	18/111	0,0497

Πίνακας 8. Τα 30 κύρια διαφορετικά εκφρασμένα γονίδια από τη σύγκριση Tumor-Normal δειγμάτων που εμπλέκονται με διεργασίες GO.

Gene Symbol	Definition	Systemic Processes	Interactions	Fold Change
AURKB	aurora kinase B	8	16	2,78
CYP1A1	cytochrome P450 family 1 subfamily A member 1	8	1	-3,54
IGHV3-11	immunoglobulin heavy variable 3-11 (gene/pseudogene)	7	0	-3,51
IGKV1D-39	immunoglobulin kappa variable 1D-39	7	0	-3,68
IGKV3-11	immunoglobulin kappa variable 3-11	7	0	-3,37
IGKV3-15	immunoglobulin kappa variable 3-15	7	0	-3,29
IGHV3-7	immunoglobulin heavy variable 3-7	7	0	-3,3
IGLV7-43	immunoglobulin lambda variable 7-43	7	0	-3,61
SPDL1	spindle apparatus coiled-coil protein 1	7	0	1,81
IGHV3-23	immunoglobulin heavy variable 3-23	7	0	-3,29
IGLC2	immunoglobulin lambda constant 2	7	0	-3,09
IGKV1-16	immunoglobulin kappa variable 1-16	7	0	-3,63
IGKV1-17	immunoglobulin kappa variable 1-17	7	0	-3,63
IGLV1-47	immunoglobulin lambda variable 1-47	7	0	-3,27
IGKV2-30	immunoglobulin kappa variable 2-30	7	0	-3,78
IGLV2-11	immunoglobulin lambda variable 2-11	7	0	-3,55
IGLV2-8	immunoglobulin lambda variable 2-8	7	0	-3,43
IGKV1-5	immunoglobulin kappa variable 1-5	7	0	-3,21
IGLV1-40	immunoglobulin lambda variable 1-40	7	0	-3,54
IGLV3-21	immunoglobulin lambda variable 3-21	7	0	-3,52
IGLV1-44	immunoglobulin lambda variable 1-44	7	0	-3,37

IGLV3-27	immunoglobulin lambda variable 3-27	7	0	-3,4
IGHV3-48	immunoglobulin heavy variable 3-48	7	0	-3,74
IGKV4-1	immunoglobulin kappa variable 4-1	7	0	-3,77
IL1B	interleukin 1 beta	7	3	-2,79
IGKV2D-28	immunoglobulin kappa variable 2D-28	7	0	-3,8
IGLV2-14	immunoglobulin lambda variable 2-14	7	0	-3,52
IGKV1-33	immunoglobulin kappa variable 1-33	7	0	-3,5
IGLV6-57	immunoglobulin lambda variable 6-57	7	0	-3,06
THBS1	thrombospondin 1	7	1	-2,96

Πίνακας 9. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τα εμπλεκόμενα μονοπάτια Reactome από την ανάλυση διαφορικής έκφρασης μεταξύ Tumor και Normal δειγμάτων.

Term Definition	Enrichment	Corrected pvalue
Binding and Uptake of Ligands by Scavenger Receptors	40/98	0,0015
Scavenging of heme from plasma	34/69	0,0052
Creation of C4 and C2 activators	34/71	0,0059
Initial triggering of complement	35/79	0,0062
Complement cascade	41/114	0,0092
Regulation of Complement cascade	37/103	0,0116
Classical antibody-mediated complement activation	30/63	0,012
CD22 mediated BCR regulation	28/61	0,0151
FCGR activation	29/69	0,0186
Role of phospholipids in phagocytosis	31/82	0,0198
Role of LAT2/NTAL/LAB on calcium mobilization	28/71	0,0198
FCERI mediated MAPK activation	30/87	0,0216
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	41/186	0,0246
Resolution of Sister Chromatid Cohesion	30/125	0,0258
Cell surface interactions at the vascular wall	38/194	0,0275
Regulation of actin dynamics for phagocytic cup formation	31/118	0,0284
Fcgamma receptor (FCGR) dependent phagocytosis	33/143	0,0302
Cell Cycle Checkpoints	46/291	0,032
Hemostasis	81/676	0,0355
Mitotic Spindle Checkpoint	26/110	0,0366
FCERI mediated NF-kB activation	28/136	0,038
Cell Cycle, Mitotic	66/538	0,0394
Mitotic Prometaphase	34/198	0,0418
Cell Cycle	74/644	0,0426
Biological oxidations	34/221	0,0448
RHO GTPases Activate Formins	27/139	0,0458
Signaling by the B Cell Receptor (BCR)	28/166	0,0481

Πίνακας 10. Τα 30 κύρια διαφορικά εκφρασμένα γονίδια από τη σύγκριση Tumor-Normal δειγμάτων που εμπλέκονται με μονοπάτια Reactome.

Gene Symbol	Definition	Systemic Processes	Interactions	Fold Change
IGLV2-8	immunoglobulin lambda variable 2-8	4	0	-3,43
IGHV3-11	immunoglobulin heavy variable 3-11 (gene/pseudogene)	4	0	-3,51
IGKV1D-39	immunoglobulin kappa variable 1D-39	4	0	-3,68
IGKV3-11	immunoglobulin kappa variable 3-11	4	0	-3,37
IGKV3-15	immunoglobulin kappa variable 3-15	4	0	-3,29
IGHV3-7	immunoglobulin heavy variable 3-7	4	0	-3,3
IGKV1-5	immunoglobulin kappa variable 1-5	4	0	-3,21
IGLV1-40	immunoglobulin lambda variable 1-40	4	0	-3,54
IGLV1-44	immunoglobulin lambda variable 1-44	4	0	-3,37
IGLV3-27	immunoglobulin lambda variable 3-27	4	0	-3,4
IGKV1-16	immunoglobulin kappa variable 1-16	4	0	-3,63
IGKV1-17	immunoglobulin kappa variable 1-17	4	0	-3,63
IGLV6-57	immunoglobulin lambda variable 6-57	4	0	-3,06

IGLV1-47	immunoglobulin lambda variable 1-47	4	0	-3,27
IGLV7-43	immunoglobulin lambda variable 7-43	4	0	-3,61
IGHV3-48	immunoglobulin heavy variable 3-48	4	0	-3,74
IGKV1-33	immunoglobulin kappa variable 1-33	4	0	-3,5
IGKV2-30	immunoglobulin kappa variable 2-30	4	0	-3,78
IGKV4-1	immunoglobulin kappa variable 4-1	4	0	-3,77
IGHV3-53	immunoglobulin heavy variable 3-53	4	0	-3,15
IGLC2	immunoglobulin lambda constant 2	4	0	-3,09
IGHV2-5	immunoglobulin heavy variable 2-5	4	0	-3,36
IGKV2D-28	immunoglobulin kappa variable 2D-28	4	0	-3,8
IGHV3-23	immunoglobulin heavy variable 3-23	4	0	-3,29
IGLV2-11	immunoglobulin lambda variable 2-11	4	0	-3,55
IGLV2-14	immunoglobulin lambda variable 2-14	4	0	-3,52
IGLV3-21	immunoglobulin lambda variable 3-21	4	0	-3,52
KIF2C	kinesin family member 2C	2	5	3,13
CENPE	centromere protein E	2	3	2,95
IGHG2	immunoglobulin heavy constant gamma 2 (G2m marker)	2	1	-3,87

Πίνακας 11. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τις εμπλεκόμενες βιολογικές διεργασίες GO από την ανάλυση διαφορικής έκφρασης μεταξύ Poorly και Well δειγμάτων.

Term id	Term Definition	Enrichment	Corrected pvalue
GO:0019752	carboxylic acid metabolic process	48/893	0,0008
GO:0006082	organic acid metabolic process	49/1004	0,0013
GO:0043436	oxoacid metabolic process	48/983	0,0025
GO:0044281	small molecule metabolic process	60/1712	0,0026
GO:0044282	small molecule catabolic process	27/427	0,0033
GO:0006520	cellular amino acid metabolic process	23/328	0,0045
GO:1901605	alpha-amino acid metabolic process	20/227	0,0054
GO:0009063	cellular amino acid catabolic process	15/127	0,0058
GO:0055114	oxidation-reduction process	36/960	0,0068
GO:0032787	monocarboxylic acid metabolic process	27/505	0,007
GO:1901606	alpha-amino acid catabolic process	13/107	0,0071
GO:0017144	drug metabolic process	24/487	0,008
GO:0044283	small molecule biosynthetic process	24/574	0,0085
GO:0042737	drug catabolic process	14/139	0,0089
GO:0006629	lipid metabolic process	36/1199	0,0102
GO:0008202	steroid metabolic process	14/257	0,0116
GO:0046394	carboxylic acid biosynthetic process	16/294	0,0124
GO:0016053	organic acid biosynthetic process	16/295	0,0124
GO:0042133	neurotransmitter metabolic process	9/93	0,0131
GO:1901615	organic hydroxy compound metabolic process	19/444	0,0142
GO:0006732	coenzyme metabolic process	13/248	0,0142
GO:0051186	cofactor metabolic process	18/417	0,0144
GO:0009064	glutamine family amino acid metabolic process	8/76	0,0145
GO:0006570	tyrosine metabolic process	4/10	0,0163
GO:0009410	response to xenobiotic stimulus	14/295	0,0163
GO:0006805	xenobiotic metabolic process	9/122	0,0185
GO:0006572	tyrosine catabolic process	3/5	0,0187
GO:1901575	organic substance catabolic process	40/1748	0,0188
GO:0009056	catabolic process	43/2030	0,0195
GO:0006631	fatty acid metabolic process	14/308	0,0203
GO:0006090	pyruvate metabolic process	25750	0,0212
GO:0005975	carbohydrate metabolic process	17/473	0,0221
GO:1900046	regulation of hemostasis	7/79	0,0221
GO:0030193	regulation of blood coagulation	7/78	0,0231
GO:0072330	monocarboxylic acid biosynthetic process	10/180	0,0238
GO:0071466	cellular response to xenobiotic stimulus	10/181	0,0242
GO:0001505	regulation of neurotransmitter levels	14/355	0,0247
GO:0050818	regulation of coagulation	7/84	0,0249
GO:0006544	glycine metabolic process	4/21	0,0249

GO:0044248	cellular catabolic process	37/1759	0,0261
GO:0009072	aromatic amino acid family metabolic process	5/39	0,0274
GO:0072329	monocarboxylic acid catabolic process	8/116	0,0293
GO:0008610	lipid biosynthetic process	17/580	0,0294
GO:0055086	nucleobase-containing small molecule metabolic process	16/526	0,0302
GO:0008652	cellular amino acid biosynthetic process	7/86	0,0312
GO:0042180	cellular ketone metabolic process	6/76	0,0323
GO:1900047	negative regulation of hemostasis	5/50	0,0331
GO:0042182	ketone catabolic process	3/10	0,0334
GO:0009743	response to carbohydrate	9/184	0,0337
GO:0009069	serine family amino acid metabolic process	5/46	0,0338
GO:0030195	negative regulation of blood coagulation	5/49	0,0346
GO:1903793	positive regulation of anion transport	5/51	0,036
GO:1901565	organonitrogen compound catabolic process	24/1042	0,0367
GO:0044255	cellular lipid metabolic process	23/950	0,0373
GO:1901617	organic hydroxy compound biosynthetic process	9/185	0,038
GO:0015721	bile acid and bile salt transport	4/29	0,0385
GO:0050819	negative regulation of coagulation	5/54	0,0398
GO:1901568	fatty acid derivative metabolic process	8/155	0,0399
GO:0006575	cellular modified amino acid metabolic process	9/196	0,0401
GO:0006694	steroid biosynthetic process	7/127	0,0415
GO:0051289	protein homotetramerization	6/83	0,0421
GO:0031650	regulation of heat generation	3/13	0,0423
GO:0051047	positive regulation of secretion	13/426	0,0435
GO:0032370	positive regulation of lipid transport	5/60	0,0436
GO:1901361	organic cyclic compound catabolic process	14/475	0,0441
GO:0045987	positive regulation of smooth muscle contraction	4/33	0,045
GO:0010817	regulation of hormone levels	15/528	0,045
GO:0009109	coenzyme catabolic process	3/15	0,0468
GO:0043648	dicarboxylic acid metabolic process	6/103	0,0482
GO:0009117	nucleotide metabolic process	13/435	0,0484
GO:0015711	organic anion transport	13/445	0,0486
GO:0051260	protein homooligomerization	11/328	0,0487
GO:0001676	long-chain fatty acid metabolic process	6/108	0,0499
GO:0008206	bile acid metabolic process	4/38	0,05

Πίνακας 12. Τα 30 κύρια διαφορετικά εκφρασμένα γονίδια από τη σύγκριση Poorly-Well δειγμάτων που εμπλέκονται με διεργασίες GO.

Gene Symbol	Definition	Systemic Processes	Interactions	Fold Change
CYP4F2	cytochrome P450 family 4 subfamily F member 2	10	0	-5,00
ABAT	4-aminobutyrate aminotransferase	9	0	-2,78
PCK2	phosphoenolpyruvate carboxykinase 2, mitochondrial	7	0	-2,45
PTGS2	prostaglandin-endoperoxide synthase 2	7	0	3,24
ALDH5A1	aldehyde dehydrogenase 5 family member A1	7	0	-1,86
BAAT	bile acid-CoA:amino acid N-acyltransferase	7	0	-2,9
IYD	iodotyrosine deiodinase	7	0	-3,99
HAGH	hydroxyacylglutathione hydrolase	7	0	-1,99
SLC6A1	solute carrier family 6 member 1	6	0	-3,09
CYP4F11	cytochrome P450 family 4 subfamily F member 11	6	0	-3,48
SHMT1	serine hydroxymethyltransferase 1	6	0	-2,85
ACADL	acyl-CoA dehydrogenase long chain	6	1	-5,64
IL1B	interleukin 1 beta	6	0	2,51
SLC27A5	solute carrier family 27 member 5	6	0	-4,52
AGXT	alanine--glyoxylate and serine--pyruvate aminotransferase	6	0	-4,38
IDO2	indoleamine 2,3-dioxygenase 2	6	0	-5,59
DAO	D-amino acid oxidase	6	0	-3,95
CYP2A6	cytochrome P450 family 2 subfamily A member 6	6	0	-7,02
UPB1	beta-ureidopropionase 1	5	0	-3,74
ABCB11	ATP binding cassette subfamily B member 11	5	0	-4,05
SRD5A2	steroid 5 alpha-reductase 2	5	0	-5,22

CYP2C8	cytochrome P450 family 2 subfamily C member 8	5	1	-5,56
ALDH2	aldehyde dehydrogenase 2 family member	5	1	-2,17
DCXR	dicarbonyl and L-xylulose reductase	5	1	-2,69
DPYS	dihydropyrimidinase	5	0	-3,28
HSD17B6	hydroxysteroid 17-beta dehydrogenase 6	5	0	-4,04
HGD	homogentisate 1,2-dioxygenase	5	0	-2,49
HPD	4-hydroxyphenylpyruvate dioxygenase	5	0	-5,45
PON1	paraoxonase 1	5	0	-3,89
GLYAT	glycine-N-acyltransferase	5	0	-7,27

Πίνακας 13. Αναλυτικά αποτελέσματα του λειτουργικού χαρακτηρισμού για τα εμπλεκόμενα μονοπάτια Reactome από την ανάλυση διαφορικής έκφρασης μεταξύ Poorly και Well δειγμάτων.

Term Definition	Enrichment	Corrected pvalue
Metabolism	64/2103	0,0039
Biological oxidations	15/221	0,0087
Metabolism of amino acids and derivatives	18/369	0,0096
Conjugation of salicylate with glycine	4/8	0,0163
Metabolism of lipids	22/737	0,0191
Arachidonic acid metabolism	7/59	0,0239
Eicosanoids	4/12	0,0267
Histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolism	6/46	0,0312
Phase I - Functionalization of compounds	8/105	0,0352
Glyoxylate metabolism and glycine degradation	5/31	0,0377
Recycling of bile acids and salts	4/16	0,0394
Cytochrome P450 - arranged by substrate type	6/65	0,047

Πίνακας 14. Τα διαφορικά εκφρασμένα γονίδια από τη σύγκριση Poorly-Well δειγμάτων που εμπλέκονται με μονοπάτια Reactome.

Gene Symbol	Definition	Systemic Processes	Interactions	Fold Change
SLC10A1	solute carrier family 10 member 1	2	0	-7,21
ACSM5	acyl-CoA synthetase medium chain family member 5	2	0	-3,44
ACSM2A	acyl-CoA synthetase medium chain family member 2A	2	0	-3,7
ABCB11	ATP binding cassette subfamily B member 11	2	0	-4,05
SLC27A5	solute carrier family 27 member 5	2	0	-4,52
GLYATL1	glycine-N-acyltransferase like 1	2	0	-3,4
BAAT	bile acid-CoA:amino acid N-acyltransferase	2	0	-2,9
GLYAT	glycine-N-acyltransferase	2	0	-7,27

Πίνακας 15. Ανάλυση της άλφα ποικιλομορφίας των Normal και Tumor δειγμάτων, για τους 3 δείκτες: τα παρατηρούμενα OTUs (Observed), Shannon και Simpson.

Patient	Samples	Tissue	Observed	Shannon	Simpson
P01	N01	Normal	473	3,543399573	0,941777723
P01	T01	Tumor	422	3,104659372	0,892357103
P02	N02	Normal	506	3,364081699	0,920294117
P02	T02	Tumor	397	3,589972533	0,941468246
P03	N03	Normal	403	3,887099439	0,944291846
P03	T03	Tumor	483	3,667915057	0,935026702
P04	N04	Normal	467	2,369702432	0,815623419
P04	T04	Tumor	550	3,157851195	0,901518474
P05	N05	Normal	557	3,609198153	0,909231062
P05	T05	Tumor	582	3,79133701	0,932037623
P06	N06	Normal	454	3,347151097	0,921048712
P06	T06	Tumor	450	3,222156557	0,91899011

P07	N07	Normal	463	3,596254667	0,938500659
P07	T07	Tumor	520	3,728832661	0,942943365
P08	N08	Normal	458	3,416612549	0,926177909
P08	T08	Tumor	459	3,293577368	0,905089153
P09	N09	Normal	438	3,894800524	0,946354857
P09	T09	Tumor	549	3,582140447	0,925813558
P10	N10	Normal	426	2,963756458	0,887649916
P10	T10	Tumor	505	3,759170464	0,94588565
P11	N11	Normal	485	2,763171917	0,856663792
P11	T11	Tumor	432	3,631087418	0,939023965
P12	N12	Normal	505	3,585881543	0,928566757
P12	T12	Tumor	383	2,414610476	0,839640501
P13	N13_1	Normal	314	3,456679025	0,927616364
P13	N13_2	Normal	305	3,490090501	0,928423909
P13	T13	Tumor	469	3,329668247	0,896240591
P14	N14	Normal	502	3,822697303	0,944609562
P14	T14	Tumor	391	2,995297667	0,903831142
P15	N15	Normal	373	3,717695239	0,948104003
P15	T15	Tumor	473	3,599147111	0,929780868
P16	N16	Normal	584	3,472898258	0,92939826
P16	T16	Tumor	459	3,354130254	0,921877987
P17	N17	Normal	459	3,738448186	0,945173203
P17	T17	Tumor	477	3,564301116	0,930548729
P18	N18	Normal	413	3,758294028	0,946838494
P18	T18	Tumor	529	3,470992337	0,922734554
P19	N19	Normal	492	3,455576676	0,928200035
P19	T19	Tumor	472	2,636498308	0,854051948
P20	N20	Normal	416	3,615581336	0,935160563
P20	T20	Tumor	394	3,448634562	0,936644009
P21	N21	Normal	442	3,564649946	0,925887346
P21	T21	Tumor	467	3,372036997	0,911241259
P22	N22	Normal	505	3,725510542	0,940052418
P22	T22	Tumor	455	2,896703458	0,889114966
P23	N23	Normal	517	3,654514864	0,942925417
P23	T23	Tumor	522	3,787682858	0,946763203
P24	N24	Normal	477	3,414398824	0,92582244
P24	T24	Tumor	523	3,007879507	0,896978685
P25	N25_1	Normal	316	2,955498886	0,880632453
P25	N25_2	Normal	309	2,895759173	0,873887375
P25	T25	Tumor	458	2,594876365	0,835170936

Πίνακας 16. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Tumor vs Normal, από το edgeR.

Bacteria	log ₂ FC	adj.P.Val	Up/Down Regulation
Lactococcus lactis	-3,04	2,63E-04	down
Microbacterium sp.	-3,02	2,01E-04	down
Streptomyces sparsogenes	-2,63	2,01E-04	down
Chlamydia sp.	-0,67	4,93E-04	down
Bathymodiolus thermophilus	-0,66	1,62E-02	down
Bacillus thuringiensis	-0,60	2,63E-04	down
Chlamydia trachomatis	-0,60	1,26E-02	down
Marteella sp.	-0,59	2,77E-02	down
Chlamydia psittaci	-0,56	3,61E-03	down
Clostridioides difficile	-0,54	4,21E-02	down
alpha proteobacterium	-0,54	7,65E-03	down
Pacificimonas flava	-0,46	3,65E-02	not_sig
Flavobacterium sp.	-0,44	7,85E-02	not_sig
Fusobacterium nucleatum	-0,39	2,12E-02	not_sig
Sphingomonas melonis	-0,36	1,26E-01	not_sig
Paenibacillus sp.	-0,32	3,90E-01	not_sig

Chlamydia abortus	-0,28	1,55E-01	not_sig
Bacteria	-0,27	7,48E-02	not_sig
Gordonia polyisoprenivorans	-0,26	3,17E-01	not_sig
Sphingomonas sp.	-0,23	2,08E-01	not_sig
Terrabacter sp.	-0,21	2,87E-01	not_sig
Curvibacter lanceolatus	-0,19	3,53E-01	not_sig
Curvibacter gracilis	-0,19	3,87E-01	not_sig
Clostridium citroniae	-0,15	8,36E-01	not_sig
Rhizobium sp.	-0,13	5,61E-01	not_sig
Paenibacillus odorifer	-0,13	6,30E-01	not_sig
Collimonas pratensis	-0,13	4,37E-01	not_sig
Pelomonas sp.	-0,12	6,30E-01	not_sig
Candidatus Regiella	-0,12	5,72E-01	not_sig
Streptococcus dysgalactiae	-0,11	5,72E-01	not_sig
Vibrio parahaemolyticus	-0,10	6,30E-01	not_sig
Acinetobacter baumannii	-0,02	8,90E-01	not_sig
Ruminococcus albus	0,03	8,36E-01	not_sig
Pseudomonas syringae	0,07	6,30E-01	not_sig
Klebsiella pneumoniae	0,07	9,41E-01	not_sig
Enterococcus faecium	0,09	5,72E-01	not_sig
Pseudomonas sp.	0,09	5,72E-01	not_sig
Mycobacterium sp.	0,11	5,72E-01	not_sig
Mycoplasma hyopneumoniae	0,15	5,72E-01	not_sig
Gammaproteobacteria bacterium	0,17	3,17E-01	not_sig
Bacillus wiedmannii	0,18	5,72E-01	not_sig
Mycoplasma ovipneumoniae	0,28	1,85E-01	not_sig
Helicobacter pylori	0,30	1,91E-01	not_sig
Streptococcus pneumoniae	0,31	1,49E-01	not_sig
Pseudomonas aeruginosa	0,34	7,74E-02	not_sig
Escherichia coli	0,34	3,17E-01	not_sig
Mycobacterium tuberculosis	0,36	2,14E-02	not_sig
Bacillus cereus	0,36	7,68E-01	not_sig
Salmonella enterica	0,37	1,17E-02	not_sig
Herbaspirillum rubrisubalbicans	0,41	9,97E-03	not_sig
Alcanivorax hongdengensis	0,42	1,79E-01	not_sig
Klebsiella aerogenes	0,44	7,14E-02	not_sig
Acinetobacter idrijaensis	0,50	2,14E-02	not_sig
Gammaproteobacteria	0,52	1,84E-03	up
Aeromonas hydrophila	0,53	3,49E-03	up
Loktanella sp.	0,53	1,26E-02	up
Lactobacillus harbinensis	0,57	1,17E-02	up
Streptococcus agalactiae	0,58	2,31E-02	up
Aeromonas piscicola	0,79	2,01E-04	up

Πίνακας 17. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Tumor vs Normal, από το edgeR.

Pathways Level 4	log ₂ FC	adj.P.Val	Up/Down Regulation
Phosphoenolpyruvate carboxykinase [GTP] (EC 4.1.1.32)_739	-1,76	2,34E-05	down
Aldehyde dehydrogenase (EC 1.2.1.3)_729	-0,88	5,54E-04	down
Cytochrome c oxidase polypeptide I (EC 1.9.3.1)_1837	-0,61	7,20E-04	down
Cytochrome c oxidase polypeptide III (EC 1.9.3.1)_1839	-0,33	2,53E-01	not_sig
Fumarate hydratase class II (EC 4.2.1.2)_769	-0,27	3,32E-01	not_sig
Translation elongation factor 1 alpha subunit_1700	0,03	8,55E-01	not_sig
ATP synthase alpha chain (EC 3.6.3.14)_1816	0,03	8,55E-01	not_sig
ATP synthase beta chain (EC 3.6.3.14)_1818	0,24	2,53E-01	not_sig
Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)_1136	0,49	3,09E-02	up
Chaperone protein DnaK_1973	0,56	1,53E-02	up

Πίνακας 18. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Well vs Normal, από το edgeR.

Bacteria	log ₂ FC	adj.P.Val	Up/Down Regulation
Lactococcus lactis	-3,71	1,39E-04	down
Microbacterium sp.	-3,63	6,11E-05	down
Chlamydia	-0,66	9,89E-03	down
Bacillus thuringiensis	-0,60	6,90E-03	down
Aeromonas piscicola	0,64	6,11E-05	up
Herbaspirillum rubrisubalbicans	0,32	4,98E-02	not_sig
Aeromonas hydrophila	0,31	6,42E-02	not_sig
Gammaproteobacteria bacterium	0,46	6,42E-02	not_sig
Marteella sp.	-0,65	6,42E-02	not_sig
Terrabacter sp.	-0,34	6,42E-02	not_sig
Gammaproteobacteria	0,37	8,79E-02	not_sig
alpha proteobacterium	-0,47	9,89E-02	not_sig
Sphingomonas melonis	-0,48	9,89E-02	not_sig
Fusobacterium nucleatum	-0,31	1,04E-01	not_sig
Gordonia polyisoprenivorans	-0,46	1,04E-01	not_sig
Pelomonas sp.	-0,56	1,04E-01	not_sig
Klebsiella aerogenes	0,48	1,51E-01	not_sig
Salmonella enterica	0,22	1,57E-01	not_sig
Clostridioides difficile	-0,47	1,63E-01	not_sig
Sphingomonas sp.	-0,24	1,70E-01	not_sig
Bradyrhizobium sp.	-0,31	1,76E-01	not_sig
Loktanella sp.	0,35	1,76E-01	not_sig
Curvibacter lanceolatus	-0,27	1,88E-01	not_sig
Bacteria	-0,21	2,47E-01	not_sig
Lactobacillus harbinensis	0,30	2,47E-01	not_sig
Collimonas pratensis	-0,20	2,62E-01	not_sig
Curvibacter gracilis	-0,23	2,87E-01	not_sig
Flavobacterium sp.	-0,39	2,87E-01	not_sig
Acinetobacter idrijaensis	0,29	2,97E-01	not_sig
Escherichia coli	0,34	3,24E-01	not_sig
Mycobacterium tuberculosis	0,20	3,24E-01	not_sig
Paenibacillus sp.	-0,41	3,24E-01	not_sig
Streptococcus agalactiae	0,31	3,24E-01	not_sig
Streptococcus pneumoniae	0,23	3,24E-01	not_sig
Klebsiella pneumoniae	1,05	4,70E-01	not_sig
Mycoplasma ovipneumoniae	0,17	4,70E-01	not_sig
Ruminococcus albus	-0,16	4,83E-01	not_sig
Helicobacter pylori	0,16	5,69E-01	not_sig
Acinetobacter baumannii	-0,10	5,91E-01	not_sig
Pseudomonas aeruginosa	0,10	5,91E-01	not_sig
Bacillus cereus	0,63	6,12E-01	not_sig
Pseudomonas sp.	-0,06	6,20E-01	not_sig
Mycoplasma hyopneumoniae	-0,10	6,61E-01	not_sig
Mycobacterium sp.	0,00	9,94E-01	not_sig
Pseudomonas syringae	0,00	9,94E-01	not_sig

Πίνακας 19. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Poorly vs Normal, από το edgeR.

Bacteria	log₂FC	adj.P.Val	Up/Down Regulation
Chlamydia	-1,44	2,89E-04	down
Chlamydia psittaci	-1,37	4,37E-04	down
Aeromonas piscicola	1,60	4,56E-04	up
Bacillus thuringiensis	-1,15	1,94E-03	down
alpha proteobacterium	-1,07	4,13E-03	down
Salmonella enterica	1,09	4,13E-03	up
Pseudomonas aeruginosa	1,35	4,13E-03	up
Herbaspirillum rubrisubalbicans	1,06	4,62E-03	up
Aeromonas hydrophila	1,19	4,62E-03	up
Pseudomonas sp.	0,89	9,51E-03	up
Mycoplasma ovipneumoniae	1,19	1,86E-02	up
Lactobacillus harbinensis	1,30	1,87E-02	up
Marteella sp.	-1,34	2,74E-02	down
Chlamydia trachomatis	-1,22	2,74E-02	down
Chlamydia abortus	-0,95	3,63E-02	down
Gammaproteobacteria	0,88	3,63E-02	up
Mycobacterium tuberculosis	0,91	4,39E-02	up
Flavobacterium sp.	-0,97	5,39E-02	not_sig
Beijerinckia indica	-1,35	5,71E-02	not_sig
Pseudomonas syringae	0,64	5,71E-02	not_sig
Pacificimonas flava	-1,02	6,97E-02	not_sig
Loktanella sp.	1,09	6,97E-02	not_sig
Fusobacterium nucleatum	-0,59	7,26E-02	not_sig
Aurantimonas coralicida	-1,12	7,62E-02	not_sig
Leptolyngbya sp.	1,39	9,16E-02	not_sig
Klebsiella aerogenes	1,12	9,18E-02	not_sig
Clostridioides difficile	-1,33	9,18E-02	not_sig
Mycoplasma hyopneumoniae	0,94	9,97E-02	not_sig
Gammaproteobacteria bacterium	0,60	1,28E-01	not_sig
Rhizobium pusense	-0,86	1,28E-01	not_sig
Streptococcus agalactiae	1,11	1,28E-01	not_sig
Rhizobium sullae	-0,73	1,31E-01	not_sig
Bacteria	-0,53	1,31E-01	not_sig
Mycobacterium sp.	0,60	1,95E-01	not_sig
Helicobacter pylori	0,80	2,03E-01	not_sig
Ruminococcus albus	0,52	2,24E-01	not_sig
Methylobacterium sp.	-0,84	2,27E-01	not_sig
Streptococcus pneumoniae	0,74	2,69E-01	not_sig
Candidatus Regiella	-0,63	2,86E-01	not_sig
Rhizobium sp.	-0,59	2,86E-01	not_sig
Bathymodiolus thermophilus	-0,63	3,60E-01	not_sig
Sphingomonas melonis	-0,52	3,79E-01	not_sig
Alcanivorax hongdengensis	0,83	3,84E-01	not_sig
Klebsiella pneumoniae	-2,97	3,85E-01	not_sig
Terrabacter sp.	0,39	4,02E-01	not_sig
Sphingomonas taxi	-0,51	4,08E-01	not_sig
Paeniglutamicibacter antarcticus	0,93	4,35E-01	not_sig
Bacillus cereus	-2,59	4,85E-01	not_sig
Microbacterium sp.	-1,07	5,09E-01	not_sig
Methylopila sp.	-0,65	5,09E-01	not_sig
Streptococcus dysgalactiae	0,38	5,09E-01	not_sig
Acinetobacter idrijaensis	0,42	5,09E-01	not_sig
Sanguibacteroides justesenii	0,59	5,09E-01	not_sig
Cyanothece sp.	0,51	5,35E-01	not_sig
Acinetobacter baumannii	0,32	5,39E-01	not_sig

Bacillus wiedmannii	0,55	5,47E-01	not_sig
Enterococcus faecium	0,20	6,56E-01	not_sig
Bradyrhizobium canariense	-0,45	6,88E-01	not_sig
Sphingomonas sp.	-0,17	7,29E-01	not_sig
Lactococcus lactis	-0,62	7,51E-01	not_sig
Streptomyces sparsogenes	-0,56	7,51E-01	not_sig
Gordonia polyisoprenivorans	-0,22	7,51E-01	not_sig
Curvibacter lanceolatus	-0,18	7,51E-01	not_sig
Clostridium citroniae	0,74	7,51E-01	not_sig
Pelomonas puraquae	-0,19	7,85E-01	not_sig
Sphingomonas	-0,15	7,85E-01	not_sig
Curvibacter gracilis	-0,13	7,85E-01	not_sig
Meiothermus silvanus	-0,15	8,35E-01	not_sig
Collimonas pratensis	0,11	8,35E-01	not_sig
Pelomonas sp.	0,12	8,35E-01	not_sig
Bradyrhizobium sp.	0,14	8,35E-01	not_sig
Curvibacter	-0,08	9,48E-01	not_sig
Vibrio parahaemolyticus	0,05	9,48E-01	not_sig
Asinibacterium sp.	0,06	9,48E-01	not_sig
Paenibacillus sp.	-0,04	9,84E-01	not_sig
Paenibacillus odorifer	-0,02	9,92E-01	not_sig
Escherichia coli	0,00	9,95E-01	not_sig

Πίνακας 20. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βακτηριακών ειδών για τη σύγκριση Poorly vs Well, από το edgeR.

Bacteria	log₂FC	adj.P.Val	Up/Down Regulation
Klebsiella pneumoniae	-4,53	1,59E-01	not_sig
Bacillus cereus	-4,42	1,59E-01	not_sig
Chlamydia psittaci	-1,01	1,86E-01	not_sig
Clostridioides difficile	-1,01	1,59E-01	not_sig
Chlamydia	-0,96	1,86E-01	not_sig
Chlamydia trachomatis	-0,91	2,22E-01	not_sig
Chlamydia abortus	-0,87	1,86E-01	not_sig
Marteella sp.	-0,85	3,51E-01	not_sig
Bacillus thuringiensis	-0,71	2,21E-01	not_sig
Escherichia coli	-0,69	3,51E-01	not_sig
Candidatus Regiella	-0,68	1,86E-01	not_sig
Flavobacterium sp.	-0,67	3,88E-01	not_sig
alpha proteobacterium	-0,67	3,51E-01	not_sig
Pacificimonas flava	-0,66	3,51E-01	not_sig
Aurantimonas coralicida	-0,56	5,36E-01	not_sig
Rhizobium sp.	-0,54	3,73E-01	not_sig
Rhizobium pusense	-0,48	5,79E-01	not_sig
Bacteria	-0,42	3,51E-01	not_sig
Fusobacterium nucleatum	-0,40	3,51E-01	not_sig
Methylopila sp.	-0,39	7,00E-01	not_sig
Acinetobacter idrijaensis	-0,24	7,28E-01	not_sig
Sphingomonas echinoides	-0,18	7,26E-01	not_sig
Sphingomonas melonis	-0,18	7,57E-01	not_sig
Bathymodiolus thermophilus	-0,13	9,11E-01	not_sig
Curvibacter lanceolatus	-0,02	9,88E-01	not_sig
Paenibacillus odorifer	0,01	9,88E-01	not_sig
Curvibacter gracilis	0,01	9,88E-01	not_sig
Sphingomonas sp.	0,05	9,11E-01	not_sig
Enterococcus faecium	0,07	8,49E-01	not_sig
Gordonia polyisoprenivorans	0,07	9,11E-01	not_sig
Pelomonas puraquae	0,08	9,11E-01	not_sig

Vibrio parahaemolyticus	0,08	9,11E-01	not_sig
Collimonas pratensis	0,13	7,52E-01	not_sig
Meiothermus silvanus	0,14	7,93E-01	not_sig
Sphingomonas	0,20	7,52E-01	not_sig
Sanguibacteroides justesenii	0,24	6,62E-01	not_sig
Paenibacillus sp.	0,25	7,57E-01	not_sig
Acinetobacter baumannii	0,28	4,78E-01	not_sig
Pelomonas sp.	0,30	6,62E-01	not_sig
Bradyrhizobium sp.	0,31	6,53E-01	not_sig
Ruminococcus albus	0,33	2,89E-01	not_sig
Streptococcus pneumoniae	0,33	3,37E-01	not_sig
Gammaproteobacteria bacterium	0,36	3,64E-01	not_sig
Bacillus wiedmannii	0,40	3,87E-01	not_sig
Streptococcus dysgalactiae	0,40	3,29E-01	not_sig
Alcanivorax hongdengensis	0,45	3,51E-01	not_sig
Gammaproteobacteria	0,46	3,37E-01	not_sig
Helicobacter pylori	0,49	3,37E-01	not_sig
Loktanella sp.	0,52	3,43E-01	not_sig
Streptococcus agalactiae	0,53	3,51E-01	not_sig
Pseudomonas syringae	0,53	2,55E-01	not_sig
Terrabacter sp.	0,55	3,37E-01	not_sig
Mycobacterium tuberculosis	0,57	1,59E-01	not_sig
Mycobacterium sp.	0,58	2,55E-01	not_sig
Paeniglutamicibacter antarcticus	0,59	4,63E-01	not_sig
Herbaspirillum rubrisubalbicans	0,62	1,86E-01	not_sig
Lactobacillus harbinensis	0,73	1,86E-01	not_sig
Aeromonas hydrophila	0,73	1,86E-01	not_sig
Salmonella enterica	0,75	1,59E-01	not_sig
Klebsiella aerogenes	0,76	1,86E-01	not_sig
Pseudomonas sp.	0,80	1,86E-01	not_sig
Mycoplasma hyopneumoniae	0,81	3,08E-01	not_sig
Mycoplasma flocculare	0,89	2,46E-01	not_sig
Aeromonas piscicola	0,89	1,86E-01	not_sig
Leptolyngbya sp.	0,91	1,86E-01	not_sig
Mycoplasma ovipneumoniae	0,92	1,86E-01	not_sig
Clostridium citroniae	1,00	6,53E-01	not_sig
Pseudomonas aeruginosa	1,08	1,59E-01	not_sig
Microbacterium sp.	2,11	3,85E-01	not_sig
Streptomyces sparsogenes	2,12	3,51E-01	not_sig
Lactococcus lactis	2,61	3,51E-01	not_sig

Πίνακας 21. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Well vs Normal, από το edgeR.

Pathways Level 4	log ₂ FC	adj.P.Val	Up/Down Regulation
Phosphoenolpyruvate carboxykinase [GTP] (EC 4.1.1.32)_739	-1,76	4,35E-04	down
Aldehyde dehydrogenase (EC 1.2.1.3)_729	-0,86	2,78E-03	down
Cytochrome c oxidase polypeptide I (EC 1.9.3.1)_1837	-0,82	4,35E-04	down
Fumarate hydratase class II (EC 4.2.1.2)_769	-0,66	2,54E-02	down
Cytochrome c oxidase polypeptide III (EC 1.9.3.1)_1839	-0,59	8,74E-02	not_sig
ATP synthase alpha chain (EC 3.6.3.14)_1816	-0,36	6,28E-02	not_sig
Translation elongation factor 1 alpha subunit_1700	-0,35	6,28E-02	not_sig
ATP synthase beta chain (EC 3.6.3.14)_1818	-0,13	4,56E-01	not_sig
Chaperone protein DnaK_1973	0,19	4,32E-01	not_sig
Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)_1136	0,25	4,14E-01	not_sig

Πίνακας 22. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Normal, από το edgeR.

Pathways Level 4	log ₂ FC	adj.P.Val	Up/Down Regulation
Phosphoenolpyruvate carboxykinase [GTP] (EC 4.1.1.32)_739	-2,62	2,47E-05	down
Aldehyde dehydrogenase (EC 1.2.1.3)_729	-1,41	9,40E-04	down
Cytochrome c oxidase polypeptide I (EC 1.9.3.1)_1837	-1,06	1,78E-03	down
Cytochrome c oxidase polypeptide III (EC 1.9.3.1)_1839	-0,81	4,18E-02	down
ATP synthase beta chain (EC 3.6.3.14)_1818	0,70	3,39E-02	up
ATP synthase alpha chain (EC 3.6.3.14)_1816	0,85	6,46E-03	up
Translation elongation factor 1 alpha subunit_1700	0,99	1,42E-03	up
Chaperone protein DnaK_1973	1,43	6,29E-04	up
Methylmalonyl-CoA mutase (EC 5.4.99.2)_920	-0,71	9,15E-02	not_sig
Fumarate hydratase class II (EC 4.2.1.2)_769	-0,65	1,81E-01	not_sig
Electron transfer flavoprotein, alpha subunit_825	-0,32	4,44E-01	not_sig
Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)_1136	0,20	6,20E-01	not_sig

Πίνακας 23. Αποτελέσματα ανάλυσης διαφορικής αφθονίας βιολογικών λειτουργιών (Level 4) για τη σύγκριση Poorly vs Well, από το edgeR.

Pathways Level 4	log ₂ FC	adj.P.Val	Up/Down Regulation
Aldehyde dehydrogenase (EC 1.2.1.3)_729	-1,41	3,76E-02	down
Cytochrome c oxidase polypeptide I (EC 1.9.3.1)_1837	-1,11	4,35E-02	down
Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)_1136	-0,92	4,35E-02	down
Fumarate hydratase class II (EC 4.2.1.2)_769	-0,76	2,49E-01	not_sig
Cytochrome c oxidase polypeptide III (EC 1.9.3.1)_1839	-1,07	2,49E-01	not_sig
Translation elongation factor 1 alpha subunit_1700	0,49	3,26E-01	not_sig
Chaperone protein DnaK_1973	0,40	3,26E-01	not_sig
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)_2029	0,78	3,26E-01	not_sig
ATP synthase alpha chain (EC 3.6.3.14)_1816	0,30	3,30E-01	not_sig
Glucose-6-phosphate isomerase (EC 5.3.1.9)_718	-0,20	7,70E-01	not_sig
ATP synthase beta chain (EC 3.6.3.14)_1818	-0,03	9,23E-01	not_sig

Συντομογραφίες

Συντομογραφία	Εξήγηση
AFLD	Αλκοολική Λιπώδης Νόσος Ήπατος
AFP	Άλφα-Φετοπρωτεΐνη
ALD	Αλκοολική Νόσος του Ήπατος
ATP	Τριφωσφορική αδενοσίνη
BA	Χολικά Οξέα
BIC	Κριτήριο Πληροφορίας κατά Bayes
DNA	Δεοξυριβονουκλεϊκό οξύ
EBV	Ιός Epstein-Barr
GO	Γονιδιακή Οντολογία
GTF	Μορφή μεταφοράς γονιδίου
HBV	Ιός Ηπατίτιδας Β
HCC	Ηπατοκυτταρικό καρκίνωμα
HCV	Ιός Ηπατίτιδας C
HPV	Ιός Ανθρώπινων Θηλωμάτων
LPS	Λιποπολυσακχαρίτης
NAFLD	Μη Αλκοολική Λιπώδης Νόσος Ήπατος
NASH	Μη Αλκοολική Στεατοηπατίτιδα
NGS	Αλληλούχιση Επόμενης Γενιάς
OTU	Λειτουργική Ταξονομική Μονάδα
PCA	Ανάλυση Κυρίων Συνιστωσών
PCoA	Ανάλυση Κυρίων Συντεταγμένων
PE	(Αναγνώσματα) Ζεύγους Άκρων
RNA	Ριβονουκλεϊκό οξύ
mRNA	Αγγελιαφόρο Ριβονουκλεϊκό οξύ
rRNA	Ριβοσωμικό Ριβονουκλεϊκό οξύ
ROS	Αντιδραστικά Είδη Οξυγόνου
SCFA	Λιπαρά Οξέα Βραχείας Αλυσίδας
SE	(Αναγνώσματα) Μονού Άκρου
VEGF	Αγγειακός Ενδοθηλιακός Αυξητικός Παράγοντας

Βιβλιογραφία

- Abbas, T., & Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. *Nature Reviews Cancer*, 9(6), 400–414. <https://doi.org/10.1038/nrc2657>
- Acinas, S. G., Sánchez, P., Salazar, G., Cornejo-Castillo, F. M., Sebastián, M., Logares, R., Royo-Llonch, M., Paoli, L., Sunagawa, S., Hingamp, P., Ogata, H., Lima-Mendez, G., Roux, S., González, J. M., Arrieta, J. M., Alam, I. S., Kamau, A., Bowler, C., Raes, J., ... Gasol, J. M. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, 4(1), 604. <https://doi.org/10.1038/s42003-021-02112-2>
- Ahn, J., & Hayes, R. B. (2021). Environmental Influences on the Human Microbiome and Implications for Noncommunicable Disease. *Annual Review of Public Health*, 42(1), 277–292. <https://doi.org/10.1146/annurev-publhealth-012420-105020>
- Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), 61–65. <https://doi.org/10.1038/nmeth.1527>
- Alqahtani, A., Khan, Z., Alloghbi, A., S. Said Ahmed, T., Ashraf, M., & M. Hammouda, D. (2019). Hepatocellular Carcinoma: Molecular Mechanisms and Targeted Therapies. *Medicina*, 55(9), 526. <https://doi.org/10.3390/medicina55090526>
- Anand, S., & Mande, S. S. (2022). Host-microbiome interactions: Gut-Liver axis and its connection with other organs. *Npj Biofilms and Microbiomes*, 8(1), 89. <https://doi.org/10.1038/s41522-022-00352-6>
- Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, 9(6), 683–693. <https://doi.org/10.1111/j.1461-0248.2006.00926.x>
- Au, K. F., Jiang, H., Lin, L., Xing, Y., & Wong, W. H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Research*, 38(14), 4570–4578. <https://doi.org/10.1093/nar/gkq211>
- Axling, U., Olsson, C., Xu, J., Fernandez, C., Larsson, S., Ström, K., Ahrné, S., Holm, C., Molin, G., & Berger, K. (2012). Green tea powder and *Lactobacillus plantarum* affect gut microbiota, lipid metabolism and inflammation in high-fat fed C57BL/6J mice. *Nutrition & Metabolism*, 9(1), 105. <https://doi.org/10.1186/1743-7075-9-105>
- Bailey, M. J., Naik, N. N., Wild, L. E., Patterson, W. B., & Alderete, T. L. (2020). Exposure to air pollutants and the gut microbiota: a potential link between exposure, obesity, and type 2 diabetes. *Gut Microbes*, 11(5), 1188–1202. <https://doi.org/10.1080/19490976.2020.1749754>
- Balogh, J., Victor, D., Asham, E. H., Burroughs, S. G., Boktour, M., Saharia, A., Li, X., Ghobrial, M., & Monsour, H. (2016). Hepatocellular carcinoma: a review. *Journal of Hepatocellular Carcinoma*, Volume 3, 41–53. <https://doi.org/10.2147/JHC.S61146>
- Bashiardes, S., Zilberman-Schapira, G., & Elinav, E. (2016). Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and Biology Insights*, 10, BBI.S34610. <https://doi.org/10.4137/BBI.S34610>
- Behary, J., Amorim, N., Jiang, X.-T., Raposo, A., Gong, L., McGovern, E., Ibrahim, R., Chu, F., Stephens, C., Jebeili, H., Fragomeli, V., Koay, Y. C., Jackson, M., O’Sullivan, J., Weltman, M., McCaughan, G., El-Omar, E., & Zekry, A. (2021). Gut microbiota impact on the peripheral immune response in non-alcoholic fatty liver disease related hepatocellular carcinoma. *Nature Communications*, 12(1), 187. <https://doi.org/10.1038/s41467-020-20422-7>

- Benedetti, F., Curreli, S., Gallo, R. C., & Zella, D. (2021). Exogenous bacterial DnaK increases protein kinases activity in human cancer cell lines. *Journal of Translational Medicine*, *19*(1), 60. <https://doi.org/10.1186/s12967-021-02734-4>
- Berkowitz, L., Pardo-Roa, C., Salazar, G. A., Salazar-Echegarai, F., Miranda, J. P., Ramírez, G., Chávez, J. L., Kalergis, A. M., Bueno, S. M., & Álvarez-Lobos, M. (2019). Mucosal Exposure to Cigarette Components Induces Intestinal Inflammation and Alters Antimicrobial Response in Mice. *Frontiers in Immunology*, *10*. <https://doi.org/10.3389/fimmu.2019.02289>
- Bermudes, D., Low, B., & Pawelek, J. (2002). *Tumor-Targeted Salmonella* (pp. 57–63). https://doi.org/10.1007/0-306-46817-4_6
- Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M., & Snyder, M. (2004). Global Identification of Human Transcribed Sequences with Genome Tiling Arrays. *Science*, *306*(5705), 2242–2246. <https://doi.org/10.1126/science.1103388>
- Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., del Pozo-Yauner, L., & Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, *13*, 390–401. <https://doi.org/10.1016/j.csbj.2015.06.001>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Boursier, J., Mueller, O., Barret, M., Machado, M., Fizanne, L., Araujo-Perez, F., Guy, C. D., Seed, P. C., Rawls, J. F., David, L. A., Hunault, G., Oberti, F., Calès, P., & Diehl, A. M. (2016). The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. *Hepatology (Baltimore, Md.)*, *63*(3), 764–775. <https://doi.org/10.1002/hep.28356>
- Brown, K., DeCoffe, D., Molcan, E., & Gibson, D. L. (2012). Diet-Induced Dysbiosis of the Intestinal Microbiota and the Effects on Immunity and Disease. *Nutrients*, *4*(8), 1095–1119. <https://doi.org/10.3390/nu4081095>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Bushel, P. R., Ferguson, S. S., Ramaiahgari, S. C., Paules, R. S., & Auerbach, S. S. (2020). Comparison of Normalization Methods for Analysis of TempO-Seq Targeted RNA Sequencing Data. *Frontiers in Genetics*, *11*. <https://doi.org/10.3389/fgene.2020.00594>
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database issue), D233–8. <https://doi.org/10.1093/nar/gkn663>
- Chakraborti, C. K. (2015). New-found link between microbiota and obesity. *World Journal of Gastrointestinal Pathophysiology*, *6*(4), 110. <https://doi.org/10.4291/wjgp.v6.i4.110>
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. <https://doi.org/10.1186/1471-2105-14-128>
- Chen, P.-L., Lamy, B., & Ko, W.-C. (2016). *Aeromonas dhakensis*, an Increasingly Recognized Human Pathogen. *Frontiers in Microbiology*, *7*. <https://doi.org/10.3389/fmicb.2016.00793>
- Chen, Y.-W., Su, S.-L., Li, C.-W., Tsai, C.-S., Lo, C.-L., Syue, L.-S., Li, M.-C., Lee, C.-C., Lee, N.-Y., Ko, W.-C., & Chen, P.-L. (2021). Pancreaticobiliary Cancers and *Aeromonas*

- Isolates Carrying Type III Secretion System Genes *ascF-ascG* Are Associated With Increased Mortality: An Analysis of 164 *Aeromonas* Infection Episodes in Southern Taiwan. *Frontiers in Cellular and Infection Microbiology*, 11. <https://doi.org/10.3389/fcimb.2021.749269>
- Chiang, J. Y. L. (2013). Bile Acid Metabolism and Signaling. In *Comprehensive Physiology* (pp. 1191–1212). Wiley. <https://doi.org/10.1002/cphy.c120023>
- Cogdill, A. P., Gaudreau, P. O., Arora, R., Gopalakrishnan, V., & Wargo, J. A. (2018). The Impact of Intratumoral and Gastrointestinal Microbiota on Systemic Cancer Therapy. *Trends in Immunology*, 39(11), 900–920. <https://doi.org/10.1016/j.it.2018.09.007>
- Corley, S. M., MacKenzie, K. L., Beverdam, A., Roddam, L. F., & Wilkins, M. R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, 18(1), 399. <https://doi.org/10.1186/s12864-017-3797-0>
- Corrêa-Oliveira, R., Fachi, J. L., Vieira, A., Sato, F. T., & Vinolo, M. A. R. (2016). Regulation of immune cell function by short-chain fatty acids. *Clinical & Translational Immunology*, 5(4), e73. <https://doi.org/10.1038/cti.2016.17>
- Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., & Nougayrède, J.-P. (2010). *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences*, 107(25), 11537–11542. <https://doi.org/10.1073/pnas.1001261107>
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., & Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences*, 103(14), 5320–5325. <https://doi.org/10.1073/pnas.0601091103>
- de Filippo, C., Cavalieri, D., di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences*, 107(33), 14691–14696. <https://doi.org/10.1073/pnas.1005963107>
- di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre', M., Giovanni Nuciforo, P., Bensimon, A., Maestro, R., Giuseppe Pelicci, P., & d'Adda di Fagagna, F. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*, 444(7119), 638–642. <https://doi.org/10.1038/nature05327>
- El-Domiaty, N., Saliba, F., Vibert, E., Karam, V., Sobesky, R., Ibrahim, W., Pittau, G., Ciaccio, O., Salloum, C., Amer, K., Saeed, M. A., Shawky, J. A., Sa Cunha, A., Rosmorduc, O., Cherqui, D., Adam, R., & Samuel, D. (2021). Early Versus Late Hepatocellular Carcinoma Recurrence After Transplantation: Predictive Factors, Patterns, and Long-term Outcome. *Transplantation*, 105(8), 1778–1790. <https://doi.org/10.1097/TP.0000000000003434>
- El-Serag, H. B., & Rudolph, K. L. (2007). Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis. *Gastroenterology*, 132(7), 2557–2576. <https://doi.org/10.1053/j.gastro.2007.04.061>
- Engen, P. A., Green, S. J., Voigt, R. M., Forsyth, C. B., & Keshavarzian, A. (2015). The Gastrointestinal Microbiome: Alcohol Effects on the Composition of Intestinal Microbiota. *Alcohol Research : Current Reviews*, 37(2), 223–236.
- Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, 19(5), 776–792. <https://doi.org/10.1093/bib/bbx008>

- Fagagna, F. d'Adda di, Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., von Zglinicki, T., Saretzki, G., Carter, N. P., & Jackson, S. P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, *426*(6963), 194–198. <https://doi.org/10.1038/nature02118>
- Faubert, B., Li, K. Y., Cai, L., Hensley, C. T., Kim, J., Zacharias, L. G., Yang, C., Do, Q. N., Doucette, S., Burguete, D., Li, H., Huet, G., Yuan, Q., Wigal, T., Butt, Y., Ni, M., Torrealba, J., Oliver, D., Lenkinski, R. E., ... DeBerardinis, R. J. (2017). Lactate Metabolism in Human Lung Tumors. *Cell*, *171*(2), 358-371.e9. <https://doi.org/10.1016/j.cell.2017.09.019>
- Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., Wei, W., Wang, M., & Zhao, Y. (2018). Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiology*, *18*(1), 11. <https://doi.org/10.1186/s12866-018-1152-5>
- Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., Izard, J., Garrett, W. S., Chan, A. T., & Huttenhower, C. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences*, *111*(22). <https://doi.org/10.1073/pnas.1319284111>
- Fredricks, D. N., Fiedler, T. L., & Marrazzo, J. M. (2005). Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *New England Journal of Medicine*, *353*(18), 1899–1911. <https://doi.org/10.1056/NEJMoa043802>
- Ganne-Carrié, N., & Nahon, P. (2019). Hepatocellular carcinoma in the setting of alcohol-related liver disease. *Journal of Hepatology*, *70*(2), 284–293. <https://doi.org/10.1016/j.jhep.2018.10.008>
- Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, *8*(6), 469–477. <https://doi.org/10.1038/nmeth.1613>
- Ghio, A. J., Hilborn, E. D., Stonehuerner, J. G., Dailey, L. A., Carter, J. D., Richards, J. H., Crissman, K. M., Foronjy, R. F., Uyeminami, D. L., & Pinkerton, K. E. (2008). Particulate Matter in Cigarette Smoke Alters Iron Homeostasis to Produce a Biological Effect. *American Journal of Respiratory and Critical Care Medicine*, *178*(11), 1130–1138. <https://doi.org/10.1164/rccm.200802-334OC>
- Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., & Rosell, R. (2016). Human endogenous retroviruses and cancer. *Cancer Biology & Medicine*, *13*(4), 483. <https://doi.org/10.20892/j.issn.2095-3941.2016.0080>
- Gosiewski, T., Ludwig-Galezowska, A. H., Huminska, K., Sroka-Oleksiak, A., Radkowski, P., Salamon, D., Wojciechowicz, J., Kus-Slowinska, M., Bulanda, M., & Wolkow, P. P. (2017). Comprehensive detection and identification of bacterial DNA in the blood of patients with sepsis and healthy volunteers using next-generation sequencing method - the observation of DNAemia. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *36*(2), 329–336. <https://doi.org/10.1007/s10096-016-2805-7>
- Grąt, M., Wronka, K. M., Krasnodębski, M., Masior, Ł., Lewandowski, Z., Kosińska, I., Grąt, K., Stypułkowski, J., Rejowski, S., Wasilewicz, M., Gałęcka, M., Szachta, P., & Krawczyk, M. (2016). Profile of Gut Microbiota Associated With the Presence of Hepatocellular Cancer in Patients With Liver Cirrhosis. *Transplantation Proceedings*, *48*(5), 1687–1691. <https://doi.org/10.1016/j.transproceed.2016.01.077>
- Gui, X., Yang, Z., & Li, M. D. (2021). Effect of Cigarette Smoke on Gut Microbiota: State of Knowledge. *Frontiers in Physiology*, *12*. <https://doi.org/10.3389/fphys.2021.673341>

- Hänzelmann, S., Castelo, R., & Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, *14*(1), 7. <https://doi.org/10.1186/1471-2105-14-7>
- He, Z., Gharaibeh, R. Z., Newsome, R. C., Pope, J. L., Dougherty, M. W., Tomkovich, S., Pons, B., Mirey, G., Vignard, J., Hendrixson, D. R., & Jobin, C. (2019). *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut*, *68*(2), 289–300. <https://doi.org/10.1136/gutjnl-2018-317200>
- Herrero, A., Rojas, E., Misiewicz-Krzeminska, I., Krzeminski, P., & Gutiérrez, N. (2016). Molecular Mechanisms of p53 Deregulation in Cancer: An Overview in Multiple Myeloma. *International Journal of Molecular Sciences*, *17*(12), 2003. <https://doi.org/10.3390/ijms17122003>
- Heymann, C. J. F., Bard, J.-M., Heymann, M.-F., Heymann, D., & Bobin-Dubigeon, C. (2021). The intratumoral microbiome: Characterization methods and functional impact. *Cancer Letters*, *522*, 63–79. <https://doi.org/10.1016/j.canlet.2021.09.009>
- Hosaka, T., Suzuki, F., Kobayashi, M., Seko, Y., Kawamura, Y., Sezaki, H., Akuta, N., Suzuki, Y., Saitoh, S., Arase, Y., Ikeda, K., Kobayashi, M., & Kumada, H. (2013). Long-term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology*, *58*(1), 98–107. <https://doi.org/10.1002/hep.26180>
- Hoshida, Y., Nijman, S. M. B., Kobayashi, M., Chan, J. A., Brunet, J.-P., Chiang, D. Y., Villanueva, A., Newell, P., Ikeda, K., Hashimoto, M., Watanabe, G., Gabriel, S., Friedman, S. L., Kumada, H., Llovet, J. M., & Golub, T. R. (2009). Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Research*, *69*(18), 7385–7392. <https://doi.org/10.1158/0008-5472.CAN-09-1089>
- Hosseini, M., Pratas, D., & Pinho, A. (2016). A Survey on Data Compression Methods for Biological Sequences. *Information*, *7*(4), 56. <https://doi.org/10.3390/info7040056>
- Hsiao, L. L., Dangond, F., Yoshida, T., Hong, R., Jensen, R. v, Misra, J., Dillon, W., Lee, K. F., Clark, K. E., Haverty, P., Weng, Z., Mutter, G. L., Frosch, M. P., MacDonald, M. E., Milford, E. L., Crum, C. P., Bueno, R., Pratt, R. E., Mahadevappa, M., ... Gullans, S. R. (2001). A compendium of gene expression in normal human tissues. *Physiological Genomics*, *7*(2), 97–104. <https://doi.org/10.1152/physiolgenomics.00040.2001>
- Huang, J.-H., Wang, J., Chai, X.-Q., Li, Z.-C., Jiang, Y.-H., Li, J., Liu, X., Fan, J., Cai, J.-B., & Liu, F. (2022). The Intratumoral Bacterial Metataxonomic Signature of Hepatocellular Carcinoma. *Microbiology Spectrum*, *10*(5). <https://doi.org/10.1128/spectrum.00983-22>
- Jang, J.-W., Kim, J.-S., Kim, H.-S., Tak, K.-Y., Lee, S.-K., Nam, H.-C., Sung, P.-S., Kim, C.-M., Park, J.-Y., Bae, S.-H., Choi, J.-Y., & Yoon, S.-K. (2021). Significance of TERT Genetic Alterations and Telomere Length in Hepatocellular Carcinoma. *Cancers*, *13*(9). <https://doi.org/10.3390/cancers13092160>
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: A Cancer Journal for Clinicians*, *61*(2), 69–90. <https://doi.org/10.3322/caac.20107>
- Jin, Y., Lee, W. Y., Toh, S. T., Tennakoon, C., Toh, H. C., Chow, P. K.-H., Chung, A. Y.-F., Chong, S. S., Ooi, L. L.-P.-J., Sung, W.-K., & Lee, C. G.-L. (2019). Comprehensive analysis of transcriptome profiles in hepatocellular carcinoma. *Journal of Translational Medicine*, *17*(1), 273. <https://doi.org/10.1186/s12967-019-2025-x>
- Kamada, N., Seo, S.-U., Chen, G. Y., & Núñez, G. (2013). Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews Immunology*, *13*(5), 321–335. <https://doi.org/10.1038/nri3430>

- Kang, Y., Cai, Y., & Yang, Y. (2022). The Gut Microbiome and Hepatocellular Carcinoma: Implications for Early Diagnostic Biomarkers and Novel Therapies. *Liver Cancer, 11*(2), 113–125. <https://doi.org/10.1159/000521358>
- Khalaf, A. M., Fuentes, D., Morshid, A. I., Burke, M. R., Kaseb, A. O., Hassan, M., Hazle, J. D., & Elsayes, K. M. (2018). Role of Wnt/ β -catenin signaling in hepatocellular carcinoma, pathogenesis, and clinical significance. *Journal of Hepatocellular Carcinoma, Volume 5*, 61–73. <https://doi.org/10.2147/JHC.S156701>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology, 37*(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kim, S., Kim, Y., Lee, S., Kim, Y., Jeon, B., Kim, H., & Park, H. (2022). Live Biotherapeutic *Lactococcus lactis* GEN3013 Enhances Antitumor Efficacy of Cancer Treatment via Modulation of Cancer Progression and Immune System. *Cancers, 14*(17), 4083. <https://doi.org/10.3390/cancers14174083>
- Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A Beginner's Guide to Analysis of RNA Sequencing Data. *American Journal of Respiratory Cell and Molecular Biology, 59*(2), 145–157. <https://doi.org/10.1165/rcmb.2017-0430TR>
- Komiyama, S., Yamada, T., Takemura, N., Kokudo, N., Hase, K., & Kawamura, Y. I. (2021). Profiling of tumour-associated microbiota in human hepatocellular carcinoma. *Scientific Reports, 11*(1), 10589. <https://doi.org/10.1038/s41598-021-89963-1>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England), 28*(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Koutsandreas, T., Binenbaum, I., Pilalis, E., Valavanis, I., Papadodima, O., & Chatziioannou, A. (2016). Analyzing and Visualizing Genomic Complexity for the Derivation of the Emergent Molecular Networks. *International Journal of Monitoring and Surveillance Technologies Research, 4*(2), 30–49. <https://doi.org/10.4018/IJMSTR.2016040103>
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860–921. <https://doi.org/10.1038/35057062>
- Langmead, B., Hansen, K. D., & Leek, J. T. (2010). Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology, 11*(8), R83. <https://doi.org/10.1186/gb-2010-11-8-r83>
- Leggett, R. M., Ramirez-Gonzalez, R. H., Clavijo, B. J., Waite, D., & Davey, R. P. (2013). Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics, 4*. <https://doi.org/10.3389/fgene.2013.00288>
- Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., & Gordon, J. I. (2008). Evolution of Mammals and Their Gut Microbes. *Science, 320*(5883), 1647–1651. <https://doi.org/10.1126/science.1155725>
- Li, X.-Y., Shen, Y., Zhang, L., Guo, X., & Wu, J. (2022). Understanding initiation and progression of hepatocellular carcinoma through single cell sequencing. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1877*(3), 188720. <https://doi.org/10.1016/j.bbcan.2022.188720>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- Llovet, J. M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., & Gores, G. (2016). Hepatocellular carcinoma. *Nature Reviews Disease Primers*, 2(1), 16018. <https://doi.org/10.1038/nrdp.2016.18>
- Loo, T. M., Kamachi, F., Watanabe, Y., Yoshimoto, S., Kanda, H., Arai, Y., Nakajima-Takagi, Y., Iwama, A., Koga, T., Sugimoto, Y., Ozawa, T., Nakamura, M., Kumagai, M., Watashi, K., Taketo, M. M., Aoki, T., Narumiya, S., Oshima, M., Arita, M., ... Ohtani, N. (2017). Gut Microbiota Promotes Obesity-Associated Liver Cancer through PGE2-Mediated Suppression of Antitumor Immunity. *Cancer Discovery*, 7(5), 522–538. <https://doi.org/10.1158/2159-8290.CD-16-0932>
- Looma, R., Seguritan, V., Li, W., Long, T., Klitgord, N., Bhatt, A., Dulai, P. S., Caussy, C., Bettencourt, R., Highlander, S. K., Jones, M. B., Sirlin, C. B., Schnabl, B., Brinkac, L., Schork, N., Chen, C.-H., Brenner, D. A., Biggs, W., Yooseph, S., ... Nelson, K. E. (2017). Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism*, 25(5), 1054-1062.e5. <https://doi.org/10.1016/j.cmet.2017.04.001>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Luo, L., Chen, L., Ke, K., Zhao, B., Wang, L., Zhang, C., Wang, F., Liao, N., Zheng, X., Liu, X., Wang, Y., & Liu, J. (2020). High expression levels of CLEC4M indicate poor prognosis in patients with hepatocellular carcinoma. *Oncology Letters*. <https://doi.org/10.3892/ol.2020.11294>
- Maddi, A., Sabharwal, A., Violante, T., Manuballa, S., Genco, R., Patnaik, S., & Yendamuri, S. (2019). The microbiome and lung cancer. *Journal of Thoracic Disease*, 11(1), 280–291. <https://doi.org/10.21037/jtd.2018.12.88>
- Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews. Genetics*, 12(10), 671–682. <https://doi.org/10.1038/nrg3068>
- Maza, E. (2016). In Papyro Comparison of TMM (edgeR), RLE (DESeq2), and MRN Normalization Methods for a Simple Two-Conditions-Without-Replicates RNA-Seq Experimental Design. *Frontiers in Genetics*, 7. <https://doi.org/10.3389/fgene.2016.00164>
- McCarthy, D. J., & Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 25(6), 765–771. <https://doi.org/10.1093/bioinformatics/btp053>
- Milosevic, I., Vujovic, A., Barac, A., Djelic, M., Korac, M., Radovanovic Spurnic, A., Gmizic, I., Stevanovic, O., Djordjevic, V., Lekic, N., Russo, E., & Amedei, A. (2019). Gut-Liver Axis, Gut Microbiota, and Its Modulation in the Management of Liver Diseases: A Review of the Literature. *International Journal of Molecular Sciences*, 20(2), 395. <https://doi.org/10.3390/ijms20020395>
- Møller, H., Heseltine, E., & Vainio, H. (1995). Working group report on schistosomes, liver flukes and *Helicobacter pylori*. Meeting held at IARC, LYON, 7–14 june 1994. *International Journal of Cancer*, 60(5), 587–589. <https://doi.org/10.1002/ijc.2910600502>
- Moniruzzaman, M., Wurch, L. L., Alexander, H., Dyhrman, S. T., Gobler, C. J., & Wilhelm, S. W. (2017). Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nature Communications*, 8(1), 16054. <https://doi.org/10.1038/ncomms16054>
- Morin, R. D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T. J., McDonald, H., Varhol, R., Jones, S. J. M., & Marra, M. A. (2008). Profiling the HeLa S3

- transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, 45(1), 81–94. <https://doi.org/10.2144/000112900>
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621–628. <https://doi.org/10.1038/nmeth.1226>
- Mosca, A., Leclerc, M., & Hugot, J. P. (2016). Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.00455>
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, 320(5881), 1344–1349. <https://doi.org/10.1126/science.1158441>
- Napoli, S., Piccinelli, V., Mapelli, S. N., Pisignano, G., & Catapano, C. v. (2017). Natural antisense transcripts drive a regulatory cascade controlling c-MYC transcription. *RNA Biology*, 14(12), 1742–1755. <https://doi.org/10.1080/15476286.2017.1356564>
- Naranpanawa, D. N. U., Chandrasekara, C. H. W. M. R. B., Bandaranayake, P. C. G., & Bandaranayake, A. U. (2020). Raw transcriptomics data to gene specific SSRs: a validated free bioinformatics workflow for biologists. *Scientific Reports*, 10(1), 18236. <https://doi.org/10.1038/s41598-020-75270-8>
- Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., ... Straussman, R. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*, 368(6494), 973–980. <https://doi.org/10.1126/science.aay9189>
- Neuveut, C., Wei, Y., & Buendia, M. A. (2010). Mechanisms of HBV-related hepatocarcinogenesis. *Journal of Hepatology*, 52(4), 594–604. <https://doi.org/10.1016/j.jhep.2009.10.033>
- Ni, J., Huang, R., Zhou, H., Xu, X., Li, Y., Cao, P., Zhong, K., Ge, M., Chen, X., Hou, B., Yu, M., Peng, B., Li, Q., Zhang, P., & Gao, Y. (2019). Analysis of the Relationship Between the Degree of Dysbiosis in Gut Microbiota and Prognosis at Different Stages of Primary Hepatocellular Carcinoma. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.01458>
- Nishida, N., Yano, H., Nishida, T., Kamura, T., & Kojiro, M. (2006). Angiogenesis in cancer. *Vascular Health and Risk Management*, 2(3), 213–219. <https://doi.org/10.2147/vhrm.2006.2.3.213>
- Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., & Henry, R. J. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, 9(3), 328–333. <https://doi.org/10.1111/j.1467-7652.2010.00558.x>
- Nowicki, E. M., Shroff, R., Singleton, J. A., Renaud, D. E., Wallace, D., Drury, J., Zirnheld, J., Colleti, B., Ellington, A. D., Lamont, R. J., Scott, D. A., & Whiteley, M. (2018). Microbiota and Metatranscriptome Changes Accompanying the Onset of Gingivitis. *MBio*, 9(2). <https://doi.org/10.1128/mBio.00575-18>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizkadze, A. v., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44–53. <https://doi.org/10.1126/science.abj6987>
- Oksanen, J. (2010). Vegan: community ecology package. [Http://CRAN.R-Project.Org/Package= Vegan](http://CRAN.R-Project.Org/Package=Vegan).

- Ott, J. J., Stevens, G. A., Groeger, J., & Wiersma, S. T. (2012). Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine*, *30*(12), 2212–2219. <https://doi.org/10.1016/j.vaccine.2011.12.116>
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(Database issue), D206–14. <https://doi.org/10.1093/nar/gkt1226>
- Pang, Z., Gu, M.-D., & Tang, T. (2022). *Pseudomonas aeruginosa* in Cancer Therapy: Current Knowledge, Challenges and Future Perspectives. *Frontiers in Oncology*, *12*, 891187. <https://doi.org/10.3389/fonc.2022.891187>
- Paprocka, P., Durnaś, B., Mańkowska, A., Król, G., Wollny, T., & Bucki, R. (2022). *Pseudomonas aeruginosa* Infections in Cancer Patients. *Pathogens*, *11*(6), 679. <https://doi.org/10.3390/pathogens11060679>
- Park, E. M., Chelvanambi, M., Bhutiani, N., Kroemer, G., Zitvogel, L., & Wargo, J. A. (2022). Targeting the gut and tumor microbiota in cancer. *Nature Medicine*, *28*(4), 690–703. <https://doi.org/10.1038/s41591-022-01779-2>
- Pérez-Losada, M., Castro-Nallar, E., Bendall, M. L., Freishtat, R. J., & Crandall, K. A. (2015). Dual Transcriptomic Profiling of Host and Microbiota during Health and Disease in Pediatric Asthma. *PLOS ONE*, *10*(6), e0131819. <https://doi.org/10.1371/journal.pone.0131819>
- Pirola, C. J., Salatino, A., Quintanilla, M. F., Castaño, G. O., Garaycochea, M., & Sookoian, S. (2022). The influence of host genetics on liver microbiome composition in patients with NAFLD. *EBioMedicine*, *76*, 103858. <https://doi.org/10.1016/j.ebiom.2022.103858>
- Ponziani, F. R., Bhoori, S., Castelli, C., Putignani, L., Rivoltini, L., del Chierico, F., Sanguinetti, M., Morelli, D., Paroni Sterbini, F., Petito, V., Reddel, S., Calvani, R., Camisaschi, C., Picca, A., Tuccitto, A., Gasbarrini, A., Pompili, M., & Mazzaferro, V. (2019). Hepatocellular Carcinoma Is Associated With Gut Microbiota Profile and Inflammation in Nonalcoholic Fatty Liver Disease. *Hepatology*, *69*(1), 107–120. <https://doi.org/10.1002/hep.30036>
- Precup, G., & Vodnar, D.-C. (2019). Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. *British Journal of Nutrition*, *122*(2), 131–140. <https://doi.org/10.1017/S0007114519000680>
- Qu, D., Wang, Y., Xia, Q., Chang, J., Jiang, X., & Zhang, H. (2022). Intratumoral Microbiome of Human Primary Liver Cancer. *Hepatology Communications*, *6*(7), 1741–1752. <https://doi.org/10.1002/hep4.1908>
- Rajabi, M., & Mousa, S. (2017). The Role of Angiogenesis in Cancer Treatment. *Biomedicines*, *5*(4), 34. <https://doi.org/10.3390/biomedicines5020034>
- Ramakrishna, G., Rastogi, A., Trehanpati, N., Sen, B., Khosla, R., & Sarin, S. K. (2013). From Cirrhosis to Hepatocellular Carcinoma: New Molecular Insights on Inflammation and Cellular Senescence. *Liver Cancer*, *2*(3–4), 367–383. <https://doi.org/10.1159/000343852>
- Rao, R. K., Seth, A., & Sheth, P. (2004). Recent Advances in Alcoholic Liver Disease I. Role of intestinal permeability and endotoxemia in alcoholic liver disease. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *286*(6), G881–G884. <https://doi.org/10.1152/ajpgi.00006.2004>
- Rattan, P., Minacapelli, C. D., & Rustgi, V. (2020). The Microbiome and Hepatocellular Carcinoma. *Liver Transplantation*, *26*(10), 1316–1327. <https://doi.org/10.1002/lt.25828>
- Ren, Z., Li, A., Jiang, J., Zhou, L., Yu, Z., Lu, H., Xie, H., Chen, X., Shao, L., Zhang, R., Xu, S., Zhang, H., Cui, G., Chen, X., Sun, R., Wen, H., Lerut, J. P., Kan, Q., Li, L., & Zheng,

- S. (2019). Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. *Gut*, *68*(6), 1014–1023. <https://doi.org/10.1136/gutjnl-2017-315084>
- Ricotta, C., & Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity*, *31*, 201–205. <https://doi.org/10.1016/j.ecocom.2017.07.003>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Robinson, M. D., & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, *23*(21), 2881–2887. <https://doi.org/10.1093/bioinformatics/btm453>
- Rocha, M., Avenaud, P., Ménard, A., le Bail, B., Balabaud, C., Bioulac-Sage, P., de Magalhães Queiroz, D. M., & Mégraud, F. (2005). Association of Helicobacter species with hepatitis C cirrhosis with or without hepatocellular carcinoma. *Gut*, *54*(3), 396–401. <https://doi.org/10.1136/gut.2004.042168>
- Roy, D., Ehtesham, N. Z., & Hasnain, S. E. (2021). Is Mycobacterium tuberculosis carcinogenic to humans? *The FASEB Journal*, *35*(9). <https://doi.org/10.1096/fj.202001581RR>
- Schmidt, B., Wei, L., DePeralta, D. K., Hoshida, Y., Tan, P. S., Sun, X., Sventek, J. P., Lanuti, M., Tanabe, K. K., & Fuchs, B. C. (2016). Molecular subclasses of hepatocellular carcinoma predict sensitivity to fibroblast growth factor receptor inhibition. *International Journal of Cancer*, *138*(6), 1494–1505. <https://doi.org/10.1002/ijc.29893>
- Schnabl, B., & Brenner, D. A. (2014). Interactions Between the Intestinal Microbiome and Liver Diseases. *Gastroenterology*, *146*(6), 1513–1524. <https://doi.org/10.1053/j.gastro.2014.01.020>
- Scholz, M. B., Lo, C.-C., & Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology*, *23*(1), 9–15. <https://doi.org/10.1016/j.copbio.2011.11.013>
- Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science*, *371*(6536). <https://doi.org/10.1126/science.abc4552>
- Seyednasrollah, F., Laiho, A., & Elo, L. L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*, *16*(1), 59–70. <https://doi.org/10.1093/bib/bbt086>
- Shakya, M., Lo, C.-C., & Chain, P. S. G. (2019). Advances and Challenges in Metatranscriptomic Analysis. *Frontiers in Genetics*, *10*. <https://doi.org/10.3389/fgene.2019.00904>
- Sharma, V., Jambrina, P. G., Kaukonen, M., Rosta, E., & Rich, P. R. (2017). Insights into functions of the H channel of cytochrome *c* oxidase from atomistic molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, *114*(48). <https://doi.org/10.1073/pnas.1708628114>
- Shin, N.-R., Lee, J.-C., Lee, H.-Y., Kim, M.-S., Whon, T. W., Lee, M.-S., & Bae, J.-W. (2014). An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut*, *63*(5), 727–735. <https://doi.org/10.1136/gutjnl-2012-303839>

- Srinivasan, S., Guha, M., Dong, D. W., Whelan, K. A., Ruthel, G., Uchikado, Y., Natsugoe, S., Nakagawa, H., & Avadhani, N. G. (2016). Disruption of cytochrome c oxidase function induces the Warburg effect and metabolic reprogramming. *Oncogene*, *35*(12), 1585–1595. <https://doi.org/10.1038/onc.2015.227>
- Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., Patapoutian, A., Hampton, G. M., Schultz, P. G., & Hogenesch, J. B. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(7), 4465–4470. <https://doi.org/10.1073/pnas.012025199>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- Suh, Y., Amelio, I., Guerrero Urbano, T., & Tavassoli, M. (2014). Clinical update on cancer: molecular oncology of head and neck cancer. *Cell Death & Disease*, *5*(1), e1018–e1018. <https://doi.org/10.1038/cddis.2013.548>
- Suresh, D., Srinivas, A. N., & Kumar, D. P. (2020). Etiology of Hepatocellular Carcinoma: Special Focus on Fatty Liver Disease. *Frontiers in Oncology*, *10*. <https://doi.org/10.3389/fonc.2020.601710>
- Suzuki, M., Sugimoto, K., Tanaka, J., Tameda, M., Inagaki, Y., Kusagawa, S., Nojiri, K., Beppu, T., Yoneda, K., Yamamoto, N., Ito, M., Yoneda, M., Uchida, K., Takase, K., & Shiraki, K. (2010). Up-regulation of glypican-3 in human hepatocellular carcinoma. *Anticancer Research*, *30*(12), 5055–5061.
- Suzuki, T. (2020). Regulation of the intestinal barrier by nutrients: The role of tight junctions. *Animal Science Journal*, *91*(1). <https://doi.org/10.1111/asj.13357>
- Tatusova, T., Ciufu, S., Fedorov, B., O'Neill, K., & Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Research*, *42*(Database issue), D553-9. <https://doi.org/10.1093/nar/gkt1274>
- Telenti, A., Pierce, L. C. T., Biggs, W. H., di Iulio, J., Wong, E. H. M., Fabani, M. M., Kirkness, E. F., Moustafa, A., Shah, N., Xie, C., Brewerton, S. C., Bultara, N., Garner, C., Metzker, G., Sandoval, E., Perkins, B. A., Och, F. J., Turpaz, Y., & Venter, J. C. (2016). Deep sequencing of 10,000 human genomes. *Proceedings of the National Academy of Sciences*, *113*(42), 11901–11906. <https://doi.org/10.1073/pnas.1613365113>
- Tripathi, A., Debelius, J., Brenner, D. A., Karin, M., Loomba, R., Schnabl, B., & Knight, R. (2018). The gut–liver axis and the intersection with the microbiome. *Nature Reviews Gastroenterology & Hepatology*, *15*(7), 397–411. <https://doi.org/10.1038/s41575-018-0011-z>
- Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. <https://doi.org/10.1038/nature07540>
- Valavanis, I., Pilalis, E., Georgiadis, P., Kyrtopoulos, S., & Chatziioannou, A. (2015). Cancer Biomarkers from Genome-Scale DNA Methylation: Comparison of Evolutionary and Semantic Analysis Methods. *Microarrays*, *4*(4), 647–670. <https://doi.org/10.3390/microarrays4040647>
- van der Vaart, H., Postma, D. S., Timens, W., & ten Hacken, N. H. (2004). Acute effects of cigarette smoke on inflammation and oxidative stress: a review. *Thorax*, *59*(8), 713–721. <https://doi.org/10.1136/thx.2003.012468>

- Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I., & Marden, J. H. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, *17*(7), 1636–1647. <https://doi.org/10.1111/j.1365-294X.2008.03666.x>
- Wahlström, A., Sayin, S. I., Marschall, H.-U., & Bäckhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, *24*(1), 41–50. <https://doi.org/10.1016/j.cmet.2016.05.005>
- Walters, K. E., & Martiny, J. B. H. (2020). Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. *PloS One*, *15*(9), e0233872. <https://doi.org/10.1371/journal.pone.0233872>
- Wang, H., & Song, M. (2011). Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming. *The R Journal*, *3*(2), 29. <https://doi.org/10.32614/RJ-2011-015>
- Wang, L., Feng, Z., Wang, X., Wang, X., & Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics (Oxford, England)*, *26*(1), 136–138. <https://doi.org/10.1093/bioinformatics/btp612>
- Wang, T., Ma, F., & Qian, H. (2021). Defueling the cancer: ATP synthase as an emerging target in cancer therapy. *Molecular Therapy - Oncolytics*, *23*, 82–95. <https://doi.org/10.1016/j.omto.2021.08.015>
- Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., & Zheng, S.-S. (2015). Application of metagenomics in the human gut microbiome. *World Journal of Gastroenterology*, *21*(3), 803. <https://doi.org/10.3748/wjg.v21.i3.803>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Webster, A. J. (2020). *Bayesian information criteria for clustering normally distributed data*.
- Westreich, S. T., Treiber, M. L., Mills, D. A., Korf, I., & Lemay, D. G. (2018). SAMSA2: a standalone metatranscriptome analysis pipeline. *BMC Bioinformatics*, *19*(1), 175. <https://doi.org/10.1186/s12859-018-2189-z>
- Whittaker, R. H. (1952). A Study of Summer Foliage Insect Communities in the Great Smoky Mountains. *Ecological Monographs*, *22*(1), 1–44. <https://doi.org/10.2307/1948527>
- Xiao, Y., Chen, B., Yang, K., Wang, Q., Liu, P., Gu, Y., Zhong, Q., Liu, Z., He, Y., & Liu, Q. (2019). Down-regulation of MARCO associates with tumor progression in hepatocellular carcinoma. *Experimental Cell Research*, *383*(2), 111542. <https://doi.org/10.1016/j.yexcr.2019.111542>
- Xie, G., Wang, X., Liu, P., Wei, R., Chen, W., Rajani, C., Hernandez, B. Y., Alegado, R., Dong, B., Li, D., & Jia, W. (2016). Distinctly altered gut microbiota in the progression of liver disease. *Oncotarget*, *7*(15), 19355–19366. <https://doi.org/10.18632/oncotarget.8466>
- Xue, C., Jia, J., Gu, X., Zhou, L., Lu, J., Zheng, Q., Su, Y., Zheng, S., & Li, L. (2022). Intratumoral bacteria interact with metabolites and genetic alterations in hepatocellular carcinoma. *Signal Transduction and Targeted Therapy*, *7*(1), 335. <https://doi.org/10.1038/s41392-022-01159-9>
- Yacoub, E., Saed Abdul-Wahab, O. M., Al-Shyarba, M. H., & ben Abdelmoumen Mardassi, B. (2021). The Relationship between Mycoplasmas and Cancer: Is It Fact or Fiction? Narrative Review and Update on the Situation. *Journal of Oncology*, *2021*, 1–13. <https://doi.org/10.1155/2021/9986550>
- Yin, H., Qu, J., Peng, Q., & Gan, R. (2019). Molecular mechanisms of EBV-driven cell cycle progression and oncogenesis. *Medical Microbiology and Immunology*, *208*(5), 573–583. <https://doi.org/10.1007/s00430-018-0570-1>

- Yu, J., Xia, X., Dong, Y., Gong, Z., Li, G., Chen, G. G., & Lai, P. B. S. (2021). CYP1A2 suppresses hepatocellular carcinoma through antagonizing HGF/MET signaling. *Theranostics*, *11*(5), 2123–2136. <https://doi.org/10.7150/thno.49368>
- Zeevi, D., Korem, T., Godneva, A., Bar, N., Kurilshikov, A., Lotan-Pompan, M., Weinberger, A., Fu, J., Wijmenga, C., Zhernakova, A., & Segal, E. (2019). Structural variation in the gut microbiome associates with host health. *Nature*, *568*(7750), 43–48. <https://doi.org/10.1038/s41586-019-1065-y>
- Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., Mao, Y., Zhang, X., Pang, X., Wei, C., Zhao, G., Chen, Y., & Zhao, L. (2010). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *The ISME Journal*, *4*(2), 232–241. <https://doi.org/10.1038/ismej.2009.112>
- Zhang, D., Hu, Q., Liu, X., Zou, K., Sarkodie, E. K., Liu, X., & Gao, F. (2020). AllEnricher: a comprehensive gene set function enrichment tool for both model and non-model species. *BMC Bioinformatics*, *21*(1), 106. <https://doi.org/10.1186/s12859-020-3408-y>
- Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. <https://doi.org/10.1093/bioinformatics/btt593>
- Zhang, K., Yang, X., Zhao, Q., Li, Z., Fu, F., Zhang, H., Zheng, M., & Zhang, S. (2020). Molecular Mechanism of Stem Cell Differentiation into Adipocytes and Adipocyte Differentiation of Malignant Tumor. *Stem Cells International*, *2020*, 1–16. <https://doi.org/10.1155/2020/8892300>
- Zhao, L., & Sanyal, S. (2022). p53 Isoforms as Cancer Biomarkers and Therapeutic Targets. *Cancers*, *14*(13), 3145. <https://doi.org/10.3390/cancers14133145>
- Zhao, S., Ye, Z., & Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA (New York, N.Y.)*, *26*(8), 903–909. <https://doi.org/10.1261/rna.074922.120>
- Zhao, S., Zhang, Y., Gamini, R., Zhang, B., & von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific Reports*, *8*(1), 4781. <https://doi.org/10.1038/s41598-018-23226-4>
- Zheng, R., Wang, G., Pang, Z., Ran, N., Gu, Y., Guan, X., Yuan, Y., Zuo, X., Pan, H., Zheng, J., & Wang, F. (2020). Liver cirrhosis contributes to the disorder of gut microbiota in patients with hepatocellular carcinoma. *Cancer Medicine*, *9*(12), 4232–4250. <https://doi.org/10.1002/cam4.3045>
- Zhu, R. X., Seto, W.-K., Lai, C.-L., & Yuen, M.-F. (2016). Epidemiology of Hepatocellular Carcinoma in the Asia-Pacific Region. *Gut and Liver*, *10*(3). <https://doi.org/10.5009/gnl15257>
- Zito, A., Lualdi, M., Granata, P., Cocciadiferro, D., Novelli, A., Alberio, T., Casalone, R., & Fasano, M. (2021). Gene Set Enrichment Analysis of Interaction Networks Weighted by Node Centrality. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.577623>
- zur Hausen, H. (2009). The Search for Infectious Causes of Human Cancers: Where and Why (Nobel Lecture). *Angewandte Chemie International Edition*, *48*(32), 5798–5808. <https://doi.org/10.1002/anie.200901917>