



ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

“Μεταγραφικό προφίλ της μη-αλκοολικής λιπώδους νόσου του ήπατος και ο ρόλος των μακρύ μη κωδικών RNAs στην εξέλιξή της”

Κοζωνάκης Άγγελος
AM: 2372

Επιβλέποντες:

Ευγενία Ντίνη
Χαράλαμπος Σπηλιανάκης

Ηράκλειο Κρήτης, 2022

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	4
ΣΥΝΤΜΗΣΕΙΣ	5
ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ: ΕΙΣΑΓΩΓΗ	
Υποκεφάλαιο 1.1: Τα κύτταρα του ήπατος	7
Υποκεφάλαιο 1.1.1: Ηπατοκύτταρα	8
Υποκεφάλαιο 1.1.2: Ενδοθηλιακά κύτταρα των ηπατικών κολποειδών αγγείων	8
Υποκεφάλαιο 1.1.3: Ηπατικά αστεροειδή κύτταρα	8
Υποκεφάλαιο 1.2: Νόσοι & ασθένειες του ήπατος	9
Υποκεφάλαιο 1.3: Η μη-αλκοολική λιπώδης νόσος του ήπατος	10
Υποκεφάλαιο 1.3.1: Χαρακτηριστικά της NAFLD	10
Υποκεφάλαιο 1.3.2: Ιστοπαθολογικές αλλαγές της NAFLD	11
Υποκεφάλαιο 1.4: Κατηγοριοποίηση ιστολογικών αλλαγών της NAFLD	13
Υποκεφάλαιο 1.5: Στάδια και βαθμοί κατάταξης της μη-αλκοολικής στεατοηπατίδας	14
Υποκεφάλαιο 1.6: Διάγνωση της NAFLD	15
Υποκεφάλαιο 1.6.1: Βιοψία	17
Υποκεφάλαιο 1.6.2: Μη-επεμβατικές μέθοδοι	17
Υποκεφάλαιο 1.7: Κωδικά και μη-κωδικά γονίδια	18
Υποκεφάλαιο 1.8: Μη-κωδικό RNA	18
Υποκεφάλαιο 1.9: Μακρύ μη-κωδικό RNA & Βασικά χαρακτηριστικά	18
Υποκεφάλαιο 1.9.1: Μακρύ μη-κωδικό RNA – Ρόλος και Λειτουργίες	19
Υποκεφάλαιο 1.10: Μακρύ μη-κωδικό RNA – eRNAs	20
Υποκεφάλαιο 1.10.1: Μακρύ μη-κωδικό RNA – eRNAs	21
Υποκεφάλαιο 1.11: Πρωτεΐνες που προσδένουν RNA	22
ΣΚΟΠΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ	24
ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ: ΜΕΘΟΔΟΛΟΓΙΑ	
Υποκεφάλαιο 2.1: Συλλογή δεδομένων	25
Υποκεφάλαιο 2.2: Εύρεση PCGs & lncRNAs	25
Υποκεφάλαιο 2.3: Ανάλυση κύριων συνιστώσων	26
Υποκεφάλαιο 2.4: Ανάλυση διαφορικής έκφρασης	26
Υποκεφάλαιο 2.4.1: Απεικόνιση αποτελεσμάτων διαφορικής έκφρασης	27
Υποκεφάλαιο 2.5: Αυτοματοποιημένη μηχανική μάθηση	27
Υποκεφάλαιο 2.6: Οντολογία γονιδίων & Εγκυκλοπαίδεια του Κιότο για τα γονίδια και τα γονιδιώματα	29
ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ: ΑΠΟΤΕΛΕΣΜΑΤΑ	
Υποκεφάλαιο 3.1: Ανάλυση διαφορικής έκφρασης	30
Υποκεφάλαιο 3.1.1: Ανάλυση διαφορικής έκφρασης για PCGs	31
Υποκεφάλαιο 3.1.2: Ανάλυση διαφορικής έκφρασης για lncRNAs	35
Υποκεφάλαιο 3.1.3: Γονίδια ενδιαφέροντος για lncRNAs	35
Υποκεφάλαιο 3.2: Ανάλυση διαφορικής έκφρασης σε γράφημα τύπου ηφαιστείου	39

Υποκεφάλαιο 3.3: Αυτοματοποιημένη μηχανική μάθηση	41
Υποκεφάλαιο 3.4: Αποτελέσματα ανάλυσης GO & KEGG	43
ΣΥΖΗΤΗΣΗ	46
ΒΙΒΛΙΟΓΡΑΦΙΑ	48
ΠΑΡΑΡΤΗΜΑ	56

ΠΕΡΙΛΗΨΗ

Η καθημερινότητα πολλών πολιτισμών προωθεί την κατανάλωση ανθυγιεινών τροφίμων υψηλά σε θερμίδες, με αποτέλεσμα να αυξάνονται φαινόμενα παχυσαρκίας που οδηγούν στην μη-αλκοολική λιπώδη νόσο. Ιδιαίτερη προσοχή έχει λάβει η μελέτη των ιστοπαθολογικών αλλαγών της νόσου, αλλά και η αναγνώριση πιθανών βιοδεικτών. Ταυτόχρονα, νέες πληροφορίες για τα μακρά μη-κωδικά RNA (lncRNAs), τονίζουν την σημαντικότητά τους σε διαφορετικά επίπεδα, όπως η αναδιαμόρφωση χρωματίνης και η ρύθμιση μεταγραφής γονιδίων, συνεργατικά με πρωτεΐνες που προσδένονται RNA (RBPs). Δεδομένα από αλληλούχιση συνολικού RNA (RNA-Seq) ατόμων που νοσούν, δημοσιεύτηκαν και η ανάλυσή τους εστίασε στον ρόλο των γονιδίων που κωδικοποιούν πρωτεΐνες (PCGs). Με σκοπό να φανεί το μεταγραφικό προφίλ από μια άλλη όψη, στα ίδια δεδομένα έγινε ανάλυση πάνω στα lncRNAs. Ανάλυση διαφορικής έκφρασης και αυτοματοποιημένη μηχανική μάθηση, έδειξαν αρκετά γονίδια των lncRNAs να είναι στατιστικά σημαντικά διαφορικά εκφρασμένα, με δύο γονίδια να κυριαρχούν, το GAS6-AS1 και το MIR4435-1HG. Τα γονίδια αυτά επιλέχθηκαν ως γονίδια ενδιαφέροντος για μελλοντικές πειραματικές προσεγγίσεις. Νέα δεδομένα που χαρακτηρίζουν RBPs βιόθησαν στον εντοπισμό 5 σημαντικών γονιδίων που η έκφρασή τους διαφοροποιείται σημαντικά στην εξέλιξη της νόσου: ENO1, ZFP36, NOB1, PCBP2, SRSF10. Τα αποτελέσματα τις πτυχιακής περιγράφουν και απεικονίζουν λεπτομερώς τις αλλαγές που συμβαίνουν στα γονίδια καθώς η νόσος εξελίσσεται και παρέχει τον κώδικα με τον οποίο διεξάχθηκαν οι αναλύσεις.

ΣΥΝΤΜΗΣΕΙΣ

LSECs: Ενδοθηλιακά κύτταρα των ηπατικών κολποειδών αγγείων (Liver Sinusoidal Endothelial Cells)

HTCs: Ηπατικά αστεροειδή κύτταρα (Hepatic Stellate Cells)

NAFLD: Μη-αλκοολική λιπώδης νόσος του ήπατος (Non-Alcoholic Fatty Liver Disease)

NASH: Μη-αλκοολική στεατοηπατίτιδα (Non-Alcoholic Steatohepatitis)

NASH F0-4: Μη-αλκοολική στεατοηπατίτιδα με ίνωση (fibrosis). Η ίνωση έχει 4 στάδια τα οποία χαρακτηρίζονται με F0 (όχι ίνωση) και F1-F4 (στάδιο ίνωσης)

HCC: Ηπατοκυτταρικός καρκίνος (Hepatocellular Carcinoma)

NASH-CRN (ή NAS): Σκορ κατηγοριοποίησης των ιστοπαθολογικών αλλαγών της μη-αλκοολικής λιπώδης νόσου του ήπατος (NAFLD Activity Score). Παίρνει τιμές 0-8.

SAF: Σκορ κατηγοριοποίησης των ιστοπαθολογικών αλλαγών της μη-αλκοολικής λιπώδης νόσου του ήπατος (Steatosis, Activity, Fibrosis). Παίρνει τιμές 0-4. Διαφέρει από το NAS

PCG: Γονίδιο που κωδικοποιεί πρωτεΐνη (Protein Coding Gene)

lncRNA: Μη-κωδικό RNA

lncRNA: Μακρύ μη-κωδικό RNA (long non-coding RNA). Η συντομία lncRNA χρησιμοποιείται στη μεθοδολογία και στα αποτελέσματα για να περιγράψει τα γονίδια που κωδικοποιούν lncRNAs

TFs: Μεταγραφικοί παράγοντες (Transcription Factors)

RBP: Πρωτεΐνη που προσδένει RNA (RNA Binding Protein)

RBD: Τμήμα πρόσδεσης του RNA (RNA Binding Domain)

eRNA: Enhancer RNA

elncRNA: enhancer-associated μακρύ μη-κωδικό RNA

TSS: Σημείο έναρξης μεταγραφής (Transcription Start Site)

DEG: Διαφορικά εκφραζόμενο γονίδιο (Differentially Expressed Gene)

FC: Αλλαγή διπλώματος (Fold Change)

RNA-Seq: Πείραμα αλληλούχισης RNA (RNA Sequencing)

PCA: Ανάλυση κύριων συνιστωσών (Principal Component Analysis)

GO: Οντολογία γονιδίων (Gene Ontology)

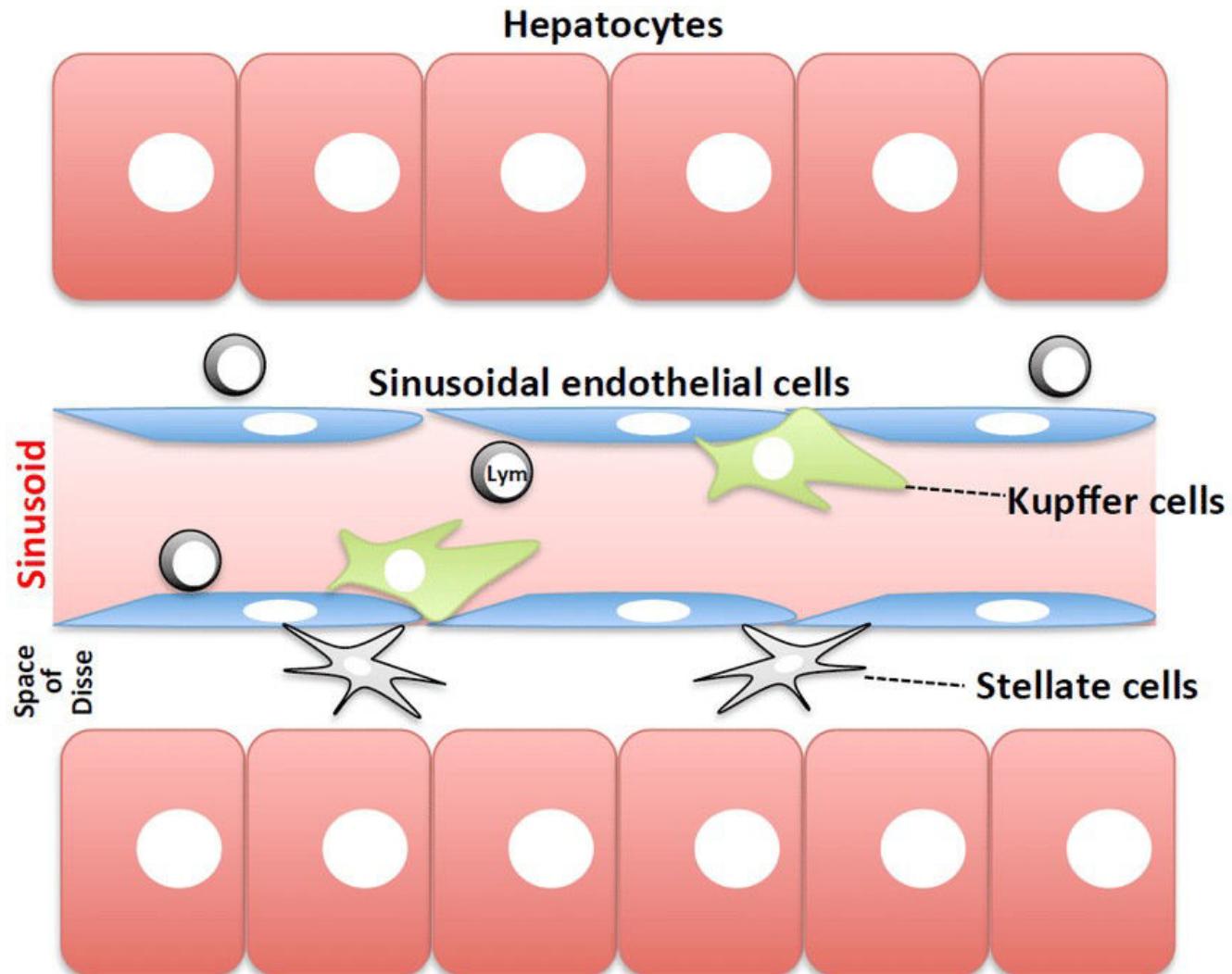
KEGG: Εγκυκλοπαίδεια του Κιότο για τα γονίδια και τα γονιδιώματα (Kyoto Encyclopedia of Genes and Genomes)

AML: Αυτοματοποιημένη μηχανική μάθηση (Automated Machine Learning)

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ: ΕΙΣΑΓΩΓΗ

Υποκεφάλαιο 1.1: Τα κύτταρα του ήπατος

Το ήπαρ συμμετέχει σε μια σειρά από μεταβολικά και ομοιοστατικά μονοπάτια, απολύτως αναγκαία για την σωστή λειτουργία και συνεπώς την υγεία του οργανισμού[1]. Στις λειτουργίες αυτές συμβάλλουν μια σειρά διαφορετικών κυττάρων τα οποία θα αναλυθούν παρακάτω(Εικόνα 1).



Εικόνα 1. Κύτταρα του ήπατος. Ηπατοκύτταρα (hepatocytes), ενδοθηλιακά κύτταρα των ηπατικών κολποειδών αγγείων (Sinusoidal endothelial cells) τα ηπατικά αστεροειδή κύτταρα (stellate cells) και τα μακροφάγα του ήπατος (Kupffer cells). DOI εικόνας: <https://doi.org/10.3390/ijms15057711>

Υποκεφάλαιο 1.1.1: Ηπατοκύτταρα

Η πολυπληθέστερη και σημαντικότερη ομάδα κυττάρων που συντελούν το συκώτι, είναι τα ηπατικά κύτταρα (ηπατοκύτταρα), φτάνοντας το 60% των συνολικών κυττάρων και το 80% της μάζας του οργάνου[2]. Ο ρόλος τους περιλαμβάνει την επεξεργασία και αποθήκευση συστατικών της τροφής, αποτοξίνωση από αμμωνία και αλκοόλη, νέο-γλυκογένεση και σύνθεση πρωτεΐνών[3]. Συνοπτικά, τα ηπατοκύτταρα διαχωρίζονται με βάση τη λειτουργία τους σε 3 κύριες ζώνες. Πρώτη είναι η περιπυλαία ζώνη 1 που αποτελείται από 6-8 στρώματα ηπατοκυττάρων που δέχονται αίμα εμπλουτισμένο με οξυγόνο και θρεπτικές ουσίες. Ελέγχουν το μεταβολισμό του γλυκογόνου, αποτοξίνωση αμμωνίας και συνθέτουν πρωτεΐνες του πλάσματος. Δεύτερη είναι η ενδιάμεση ή περιφλεβική ζώνη 2 η οποία έχει 6-10 στρώματα ηπατοκυττάρων με βασικό ρόλο στο ξενοβιοτικό μεταβολισμό. Τρίτη και τελευταία είναι η κεντρολοβιακή ζώνη 3 με 2-3 στρώματα ηπατικών κυττάρων γύρω από τη κεντρική φλέβα. Επιτελούν αντιδράσεις βιομετασχηματισμού, σύνθεση γλουταμίνης, γλυκόλυση και συμβάλλουν στην αποτοξίνωση[4,5](Εικόνα 2). Προβλήματα στη λειτουργία τους, που αφορούν στο μεταβολισμό και στην αποθήκευση λιπιδίων οδηγούν σε συσσώρευση λίπους και τελικά σε ηπατική στεάτωση (steatosis, Εικόνα 3).

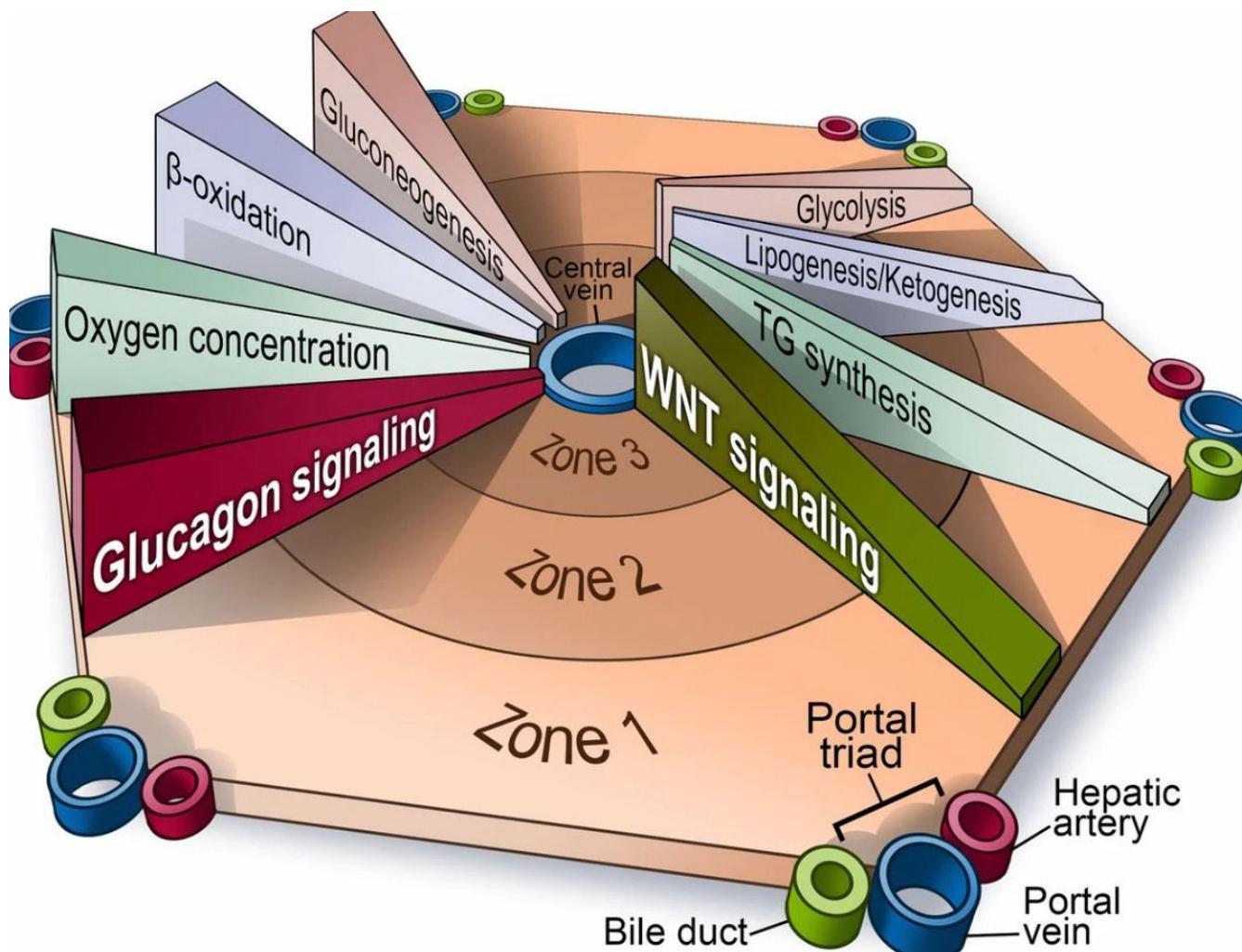
Υποκεφάλαιο 1.1.2: Ενδοθηλιακά κύτταρα των ηπατικών κολποειδών αγγείων

Εξίσου βασική ομάδα κυττάρων που απαντώνται στο συκώτι, είναι τα ενδοθηλιακά κύτταρα των ηπατικών κολποειδών αγγείων (Liver Sinusoidal Endothelial Cells, LSECs). Είναι απαραίτητα για την διατήρηση της ομοιόστασης του μεταβολισμού και του ανοσοποιητικού. Σχετίζονται με την παθοφυσιολογία πολλών ασθενειών όπως η ογκογένεση (ηπατικό καρκίνωμα)[6] και οι χρόνιες ιογενείς λοιμώξεις (ηπατίτιδα) [7] ενώ ταυτόχρονα συμβάλλουν στην εξέλιξη της κίρρωσης[8] (cirrhosis, Εικόνα 3).

Υποκεφάλαιο 1.1.3: Ηπατικά αστεροειδή κύτταρα

Η τρίτη ομάδα κυττάρων που απαντάται στο συκώτι στο συκώτι είναι τα ηπατικά αστεροειδή κύτταρα (Hepatic Stellate Cells, HSCs) με βασική λειτουργία την αποθήκευση βιταμινών (όπως η ρετινόλη) σε λιπίδια. Υπό φυσιολογικές συνθήκες είναι μεταβολικά αδρανή[3,9], όμως η συνεχής έκθεση σε

φλεγμονές και εξωγενές στρες - όπως χρόνιες ηπατικές νόσοι - οδηγεί σε δυσλειτουργία των κυττάρων αυτών. Στην περίπτωση αυτή τα αστεροειδή κύτταρα παίρνουν ένα μυοϊνοβλαστικό φαινότυπο και παράγουν κολλαγόνο, το οποίο σε μεγάλες ποσότητες οδηγεί σε συσσώρευση ουλώδη ιστού, δηλαδή σε ίνωση (fibrosis, Εικόνα 3). Η ινογένεση ενισχύεται στην περίοδο νόσησης, από την δράση των μακριφάγων του ήπατος (Kupffer κύτταρα) τα οποία πολλαπλασιάζονται ως απόκριση στις φλεγμονές[9].



Εικόνα 2. Ζώνες ηπατοκυττάρων. Περιπυλαία ζώνη 1 (zone 1), ενδιάμεση ή περιφερειακή ζώνη 2 (zone 2) και κεντρολοβιακή ζώνη 3 (zone 3). DOI εικόνας: <https://doi.org/10.1073/pnas.1804203115>

Υποκεφάλαιο 1.2: Νόσοι & ασθένειες του ήπατος

Τα κύτταρα του ήπατος χαρακτηρίζονται από διάφορες δυσλειτουργίες που οδηγούν σε μια πληθώρα νόσων επηρεάζοντας πολλαπλά όργανα. Πιο κοινές από τις νόσους, είναι η αλκοολική ηπατική νόσος, η ηπατίτιδα, η αιμοχρωματωση, η πρωτοπαθής χολική κίρρωση και η μη-αλκοολική λιπώδης νόσος

τους ήπατος. Ως αίτια πρόκλησης των νοσημάτων αυτών θεωρούνται η κατανάλωση αλκοόλ, προσβολή από ιό, κληρονομικότητα γονιδίου, προβλήματα στο ανοσοποιητικό και η παχυσαρκία αντίστοιχα.

Υποκεφάλαιο 1.3: Η μη-αλκοολική λιπώδης νόσος του ήπατος

Το λιπώδες ήπαρ, είναι η πιο κοινή νόσος που συναντάται στο συκώτι. Χαρακτηρίζεται ιστολογικά από συσσώρευση λιπιδίων και πιο συγκεκριμένα όταν τα τριγλυκερίδια είναι περισσότερα από το 5% του συνολικού βάρους του ήπατος[10]. Στην περίπτωση που αυτό συμβαίνει χωρίς την επίδραση του αλκοόλ, τότε ονομάζεται μη-αλκοολική λιπώδης νόσος του ήπατος(Non-Alcoholic Fatty Liver Disease, NAFLD) και έχουμε σημάδια απλής στεάτωσης. Όταν στο λιπώδης ήπαρ παρατηρηθεί φλεγμονή , ηπατικά “μπαλόνια” (διόγκωση κυττάρων) και προχωρημένη στεάτωση, τότε η νόσος ονομάζεται μη-αλκοολική στεατοηπατίτιδα (Non-Alcoholic Steatohepatitis, NASH) και είναι το επόμενο στάδιο της νόσου.

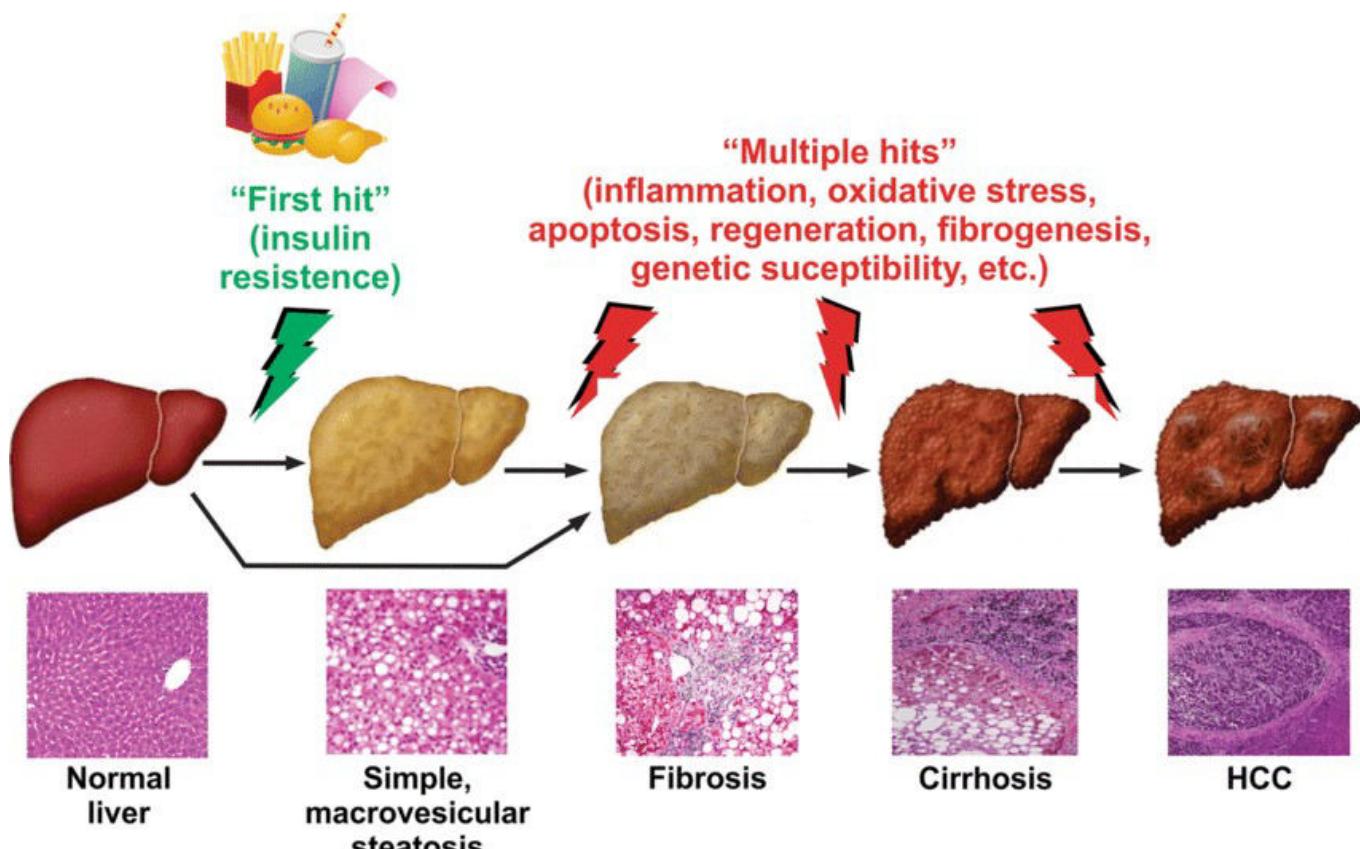
Υποκεφάλαιο 1.3.1: Χαρακτηριστικά της NAFLD

Ο τρόπος ζωής και η καθημερινότητα ειδικά στους δυτικούς πολιτισμούς, προωθεί την κατανάλωση τροφίμων υψηλά σε θερμίδες, με αποτέλεσμα να υπάρχει ραγδαία αύξηση της παχυσαρκίας και συνεπώς περιπτώσεων μη-αλκοολικής λιπώδης νόσου[11,12]. Περίπου το 15-30% όλου του πληθυσμού πάσχει από τη νόσο, και πλησιάζει το 50% με 90% σε περιπτώσεις παχυσαρκίας με το ποσοστό να αυξάνεται όσο αυξάνονται κλινικά τα στάδια της παχυσαρκίας[13,14]. Πιο συγκεκριμένα,, όταν η νόσος εξελίσσεται σε στεατοηπατίτιδα, φαίνεται να φτάνει το 65% σε πληθυσμούς με παχυσαρκία πρώτου και δεύτερου βαθμού (Δείκτης μάζας σώματος = 30-39.9 kg/m²) ενώ κορυφώνεται στο 85% σε πληθυσμούς με παχυσαρκία τρίτου βαθμού (Δείκτης μάζας σώματος = 40-59 kg/m²). Άλλοι παράγοντες που επηρεάζουν σε μεγάλο βαθμό την εξέλιξη της νόσου προς μη-αλκοολική στεατοηπατίτιδα, είναι η ηλικία (μεγαλύτερη επίδραση σε μεγαλύτερες ηλικίες) η εθνικότητα, περιβαλλοντικοί παράγοντες, ο διαβήτης (μεγαλύτερες επιπλοκές σε άτομα με διαβήτη) και το φύλο[15,16]. Παρόλο που φαίνεται να υπάρχουν περισσότερες περιπτώσεις της ασθένειας σε αρσενικά άτομα, αυτό αλλάζει μετά την εμμηνόπαυση, με τις γυναίκες να νοσούν πιο συχνά και με πιο έντονα συμπτώματα [16]. Η νόσος επίσης φαίνεται να επηρεάζεται αλλά και να επηρεάζει μεταβολικά

σύνδρομα πέρα από την παχυσαρκία, όπως την αντίσταση στην ινσουλίνη, τον τύπου 2 διαβήτη και την δυσλιπιδαιμία [17].

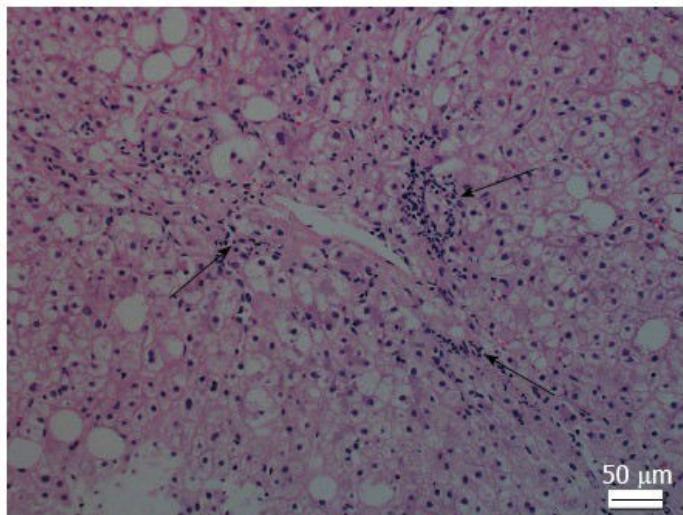
Υποκεφάλαιο 1.3.2: Ιστοπαθολογικές αλλαγές της NAFLD

Η βαρύτητα της μη αλκοολικής λιπώδης νόσου του ήπατος μπορεί να φανεί μελετώντας τις ιστολογικές αλλαγές του ήπατος. Όπως αναφέρθηκε ήδη, κύριο χαρακτηριστικό της νόσου είναι η συσσώρευση λιπιδίων στα ηπατικά κύτταρα με μοναδικό σύμπτωμα στα πρώτα στάδια, την απλή στεάτωση (Non-Alcoholic Fatty Liver, NAFL, Εικόνα 3) [18]. Όταν ξεκινήσουν να εμφανίζονται εστίες φλεγμονής(Εικόνα 4A), η νόσος μεταπίπτει σε στεατοηπατίτιδα όπου έχουμε κυτταρικό θάνατο και απόπτωση. Πέρα από τη φλεγμονή συχνά παρατηρείται ηπατικό “μπαλόνι”, δηλαδή διόγκωση κυττάρων κατά 1.5-2 φορές τη διάμετρο των φυσιολογικών και αραιό κυτταρόπλασμα [19](Εικόνα 4B). Επιπλέον σύμπτωμα στο οποίο συνεισφέρουν τα “μπαλόνια”, είναι η ίνωση (fibrosis), δηλαδή αυξημένο πάχος του συνδετικού ιστού που οδηγεί σε ουλώδη ιστό. Αυτό συμβαίνει όταν τα κύτταρα προσπαθούν να επουλωθούν και γίνεται εναπόθεση εξωκυττάριων πρωτεΐνων, με κυριότερη και πολυπληθέστερη το κολλαγόνο [20,21]. Στην προσπάθειά τους τα κύτταρα να επουλωθούν, δημιουργείται συνεχώς ουλώδης ιστός και το ήπαρ οδηγείται σε κίρρωση δηλαδή το τελικό στάδιο της ίνωσης όπου το συκώτι δεν μπορεί να λειτουργήσει σωστά. Στα πρώτα στάδια όπου δεν υπάρχουν μεγάλες επιπλοκές ονομάζεται αντιρροπούμενη κίρρωση, ενώ στο τελικό της στάδιο ονομάζεται μη αντιρροπούμενη ή ηπατική νόσος τελικού σταδίου όπου στις περισσότερες περιπτώσεις, δεν είναι αντιστρέψιμη η βλάβη στο ήπαρ και μοναδική λύση είναι η μεταμόσχευση ήπατος[21,22](Εικόνα 5). Η κίρρωση επίσης αποτελεί ένα ευνοϊκό περιβάλλον για την μετάπτωση σε ηπατοκυτταρικό καρκίνο (Hepatocellular Carcinoma, HCC, Εικόνα 3) [23,24].

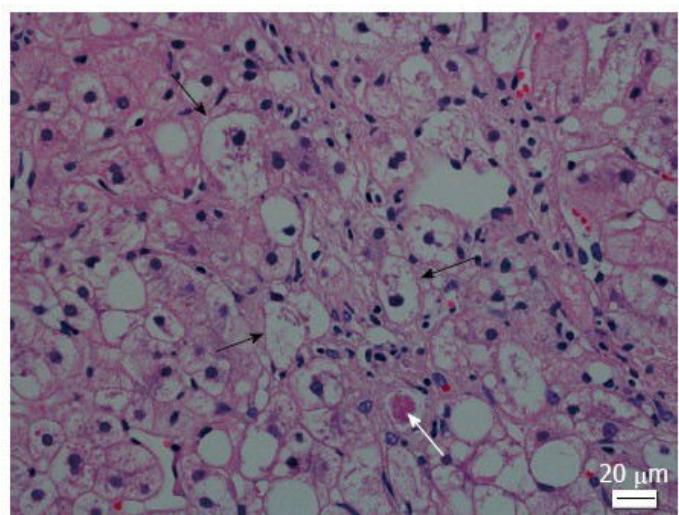


Εικόνα 3. Η εξέλιξη της μη-αλκοολικής λιπώδης νόσου του ήπατος από υγιές ήπαρ μέχρι τον ηπατοκυτταρικό καρκίνο, με έμφαση στις ιστολογικές αλλαγές. Φαίνεται το υγιές ήπαρ (normal liver), η απλή στεάτωση (simple steatosis), η ίνωση (fibrosis), η κίρρωση (cirrhosis) και η πατοκυτταρικός καρκίνος (HCC). DOI εικόνας: 10.1007/s00018-018-2947-0

A



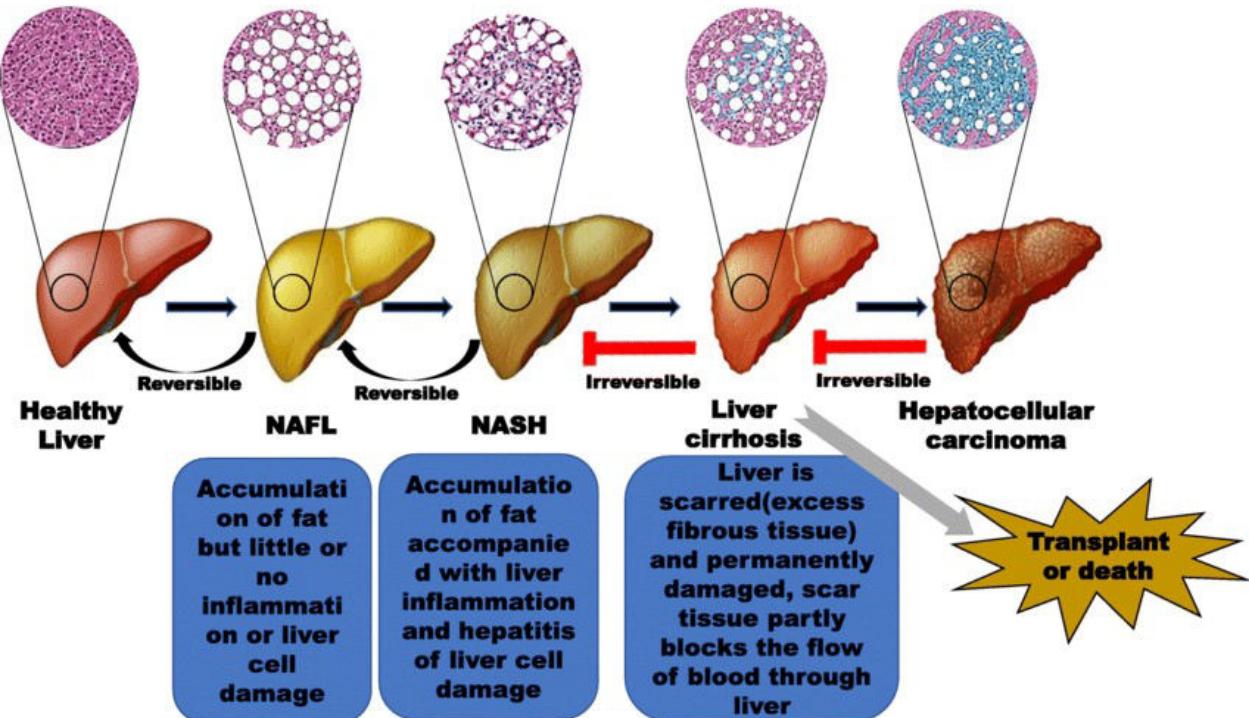
B



Εικόνα 4. Ιστολογικές αλλαγές κατά την εξέλιξη της μη-αλκοολικής λιπώδης νόσου του ήπατος. (A) Λοβιακή φλεγμονή στη μη αλκοολική στεατοηπατίτιδα. Με μαύρα βελάκια φαίνονται νεκροφλεγμονώδης εστίες (necroinflammatory foci). (B) Διογκωμένα ηπατοκύτταρα (μαύρο βελάκι) και σωμάτια Mallory-Denk (άσπρο βελάκι) κατά τη μη αλκοολική στεατοηπατίτιδα. Τα σωμάτια Mallory-Denk είναι ηωσινόφιλα συσσωματόματα στο κυτταρόπλασμα ηπατοκυττάρων.

DOI εικόνων: 10.3748/wjg.v20.i42.15539

The spectrum of NAFLD



Εικόνα 5. Η εξέλιξη της μη-αλκοολικής λιπώδης νόσου του ήπατος από υγιές ήπαρ μέχρι τον ηπατοκυτταρικό καρκίνο. Φαίνονται τα ιστολογικά χαρακτηριστικά των κυττάρων στις εικόνες πάνω από το ήπαρ. DOI εικόνας: 10.1186/s12944-020-01210-0

Υποκεφάλαιο 1.4: Κατηγοριοποίηση ιστολογικών αλλαγών της NAFLD

Με βάση τις ιστολογικές αλλαγές που συμβαίνουν στο ήπαρ κατά τη νόσο, έχει δημιουργηθεί ένας τρόπος να κατηγοριοποιηθεί η βαρύτητα των αλλαγών αυτών και ονομάζεται NAS (NAFLD Activity Score)[25]. Τα τρία κύρια χαρακτηριστικά που λαμβάνονται υπόψη είναι η βαρύτητα της στεάτωσης, τα ηπατικά “μπαλόνια” και η λοβιακή φλεγμονή στα οποία γίνεται ημιποσοτική μέτρηση. Για την στεάτωση γίνεται μέτρηση της επιφάνειας του ήπατος που έχει υποστεί αλλαγές και παίρνει τιμές από 0 έως 3 (0 όταν η επιφάνεια είναι μικρότερη από 5% ενώ 3 όταν η επιφάνεια είναι πάνω από 66%).

Για την λοβιδιακή φλεγμονή γίνεται μέτρηση των εστιών και παίρνει σκορ από 0 έως 3. Τέλος γίνεται μέτρηση των διογκωμένων κυττάρων (“μπαλόνια”) και αποδίδεται ένα σκορ από 0 έως 2. Έτσι το τελικό σκορ (NAS) μπορεί να διαμορφωθεί βάση των ιστολογικών αλλαγών από 0 έως 8 [26,27] (Πίνακας 1). Παρόλο που ο δείκτης αυτός δεν αποτελεί διαγνωστικό μέσο, οι ιστολογικές αλλαγές συσχετίζονται με την ύπαρξη ή όχι στεατοηπατίτιδας. Ασθενείς όπου το σκορ είναι μεγαλύτερο ή ίσο του 5 ($NAS \geq 5$) στις περισσότερες περιπτώσεις παρατηρείται στεατοηπατίτιδα, ενώ όταν το σκορ είναι

μικρότερο ή ίσο του 2 ($NAS \leq 2$) συνήθως δεν χαρακτηρίζονται από στεατοηπατίτιδα [25]. Αυτό διαφέρει ανάλογα τον ασθενή και δεν αποτελεί απόλυτη κατηγοριοποίηση [28].

Αντίστοιχα με το NAS, έχει προταθεί το SAF (Steatosis, Activity and Fibrosis) σκορ, το οποίο υπολογίζεται με τον αλγόριθμο FLIP (Fatty Liver Inhibition of Progression). Βασική διαφορά είναι πως σε αυτή τη περίπτωση διαχωρίζεται η εκτίμηση της στεάτωσης (S), το άθροισμα των διογκωμένων κυττάρων (“μπαλόνια”) και της λοβιδιακής φλεγμονής (A) και τέλος της ίνωσης (F), με ημιποσοτικές τακτικές κλίμακες [29,30](Πίνακας 1). Αυτό δίνει τη δυνατότητα μεγαλύτερου βαθμού λεπτομέρειας όσον αφορά την δραστηριότητα της νόσου και πιθανώς καλύτερη κατάταξη ασθενών με μη-αλκοολική στεατοηπατίτιδα.

Πίνακας 1. Τα δύο σκορ κατάταξης της μη-αλκοολικής λιπώδης νόσου του ήπατος, NAS (NASH-CRN) και SAF. Μέγιστο σκορ για το NAS είναι το 8 ενώ για το SAF το μέγιστο σκορ είναι το 4. Αναγράφονται τα χαρακτηριστικά υπολογισμού κάθε ιστοπαθολογικής αλλαγής για τα σκορ (αριστερά για το NAS και δεξιά για το SAF) καθώς και χαρακτηριστικά της ίνωσης (fibrosis). DOI πίνακα: <https://doi.org/10.3389/fmed.2021.615978>

	Score	NASH-CRN	SAF	Score
Steatosis	0	<5%	<5%	0
	1	5–33%	5–33%	1
	2	>33–67%	>33–67%	2
	3	>67%	>67%	3
Lobular inflammation	0	No foci	No foci	0
	1	<2 foci/20X	<2 foci/20X	1
	2	2–4 foci/20X	>2 foci/20X	2
	3	>4 foci/20X		
Ballooning	0	No ballooning	Normal hepatocytes	0
	1	Few ballooned cells	Clusters of rounded, pale hepatocytes	1
	2	Many ballooned cells	Many enlarged (2X normal size) hepatocytes	2
Fibrosis	0	No fibrosis	No fibrosis	0
	1	1a Mild, zone 3 perisinusoidal/pericellular fibrosis 1b Moderate, zone 3 perisinusoidal/pericellular fibrosis 1c Portal/periportal fibrosis	Mild fibrosis perisinusoidal/pericellular	1
	2	Perisinusoidal/pericellular and portal/periportal fibrosis	Perisinusoidal/pericellular and portal/periportal fibrosis	2
	3	Bridging fibrosis	Bridging fibrosis	3
	4	Cirrhosis	Cirrhosis	4
Composite score for activity	0–8	NAS = NAFLD Activity Score = steatosis + ballooning + lobular inflammation	A = ballooning + lobular inflammation	0–4

Υποκεφάλαιο 1.5: Στάδια και βαθμοί κατάταξης της μη-αλκοολικής στεατοηπατίτιδας

Η μη-αλκοολική στεατοηπατίτιδα (NASH) κατηγοριοποιείται με βάση τη στεάτωση, τα διογκωμένα και αποδιοργανωμένα κύτταρα και την φλεγμονή (ενδολοβιακή ή πυλαία), σε 3 βαθμούς(Πίνακας 2) .

Η πρώτου βαθμού ή αλλιώς ήπιας έντασης, δεύτερου βαθμού ή μέτριας έντασης και η τρίτου βαθμού ή βαριάς έντασης[31]. Επιπρόσθετα έχει γίνει κατηγοριοποίηση της μη-αλκοολικής στεατοηπατίτιδας ανάλογα με την βαρύτητα και την ιστολογική τοποθεσία της ίνωσης. Υπάρχουν τέσσερα βασικά στάδια, με το πρώτο να χωρίζεται περαιτέρω. Συγκεκριμένα, το πρώτο στάδιο (NASH F1, Non-Alcoholic Steatohepatitis Fibrosis Stage 1) χωρίζεται στα υποστάδια 1A (ήπια ίνωση στα περικολποειδή αγγεία της ζώνης 3), 1B (μέτρια ίνωση στα περικολποειδή αγγεία της ζώνης 3) και 1C (μόνο πυλαία και περιπυλαία ίνωση)[25]. Το δεύτερο στάδιο (NASH F2) χαρακτηρίζεται ταυτόχρονα από ίνωση στα περικολποειδή τριχοειδή της ζώνης 3 και από πυλαία ή περιπυλαία ίνωση. Στο τρίτο στάδιο (NASH F3) πέρα από τα συμπτώματα του δεύτερου σταδίου έχουμε εμφάνιση γεφυροποιούς ίνωσης, δηλαδή λωρίδες ινώδη ιστού διαφόρου πάχους, με το πάχος να προσδιορίζει και το πόσο έχει προχωρήσει η ασθένεια (μεγαλύτερος πάχος σημαίνει πιο προχωρημένη). Συμβαίνει συνήθως μετά από κεντρολοβιακή νέκρωση και παράγει νέες αγγειακές συνδέσεις μεταξύ πυλαία περιοχής και κεντρικών φλεβών. Τα συμπτώματα αυτά θεωρούνται στάδια αρχικής κίρρωσης. Τέταρτο και τελευταίο στάδιο (NASH F4) σημαίνει πως η ίνωση έχει προχωρήσει σε κίρρωση(Πίνακας 1, Εικόνα 6). Πέρα από τα τέσσερα αυτά στάδια μπορεί να υπάρξει και στεατοηπατίτιδα χωρίς ίνωση (NASH F0) και καθώς δεν διαφέρει πάρα πολύ από το πρώτο στάδιο συνήθως κατατάσσονται μαζί (NASH F0-F1).

Υποκεφάλαιο 1.6: Διάγνωση της NAFLD

Η νόσος ακόμα και σήμερα δημιουργεί πολλά ερωτήματα που δεν μπορούν να απαντηθούν επαρκώς και η ακριβής διάγνωσή της αποτελεί μια πρόκληση. Χωρίς κατάλληλα μέτρα για έλεγχο ασυμπτωματικών ατόμων (screening) και λόγω έλλειψης καλά προσαρμοσμένων κλινικών κατευθυντήριων οδηγιών, η εξακρίβωση πιθανών ασθενών βασίζεται σε πληθυσμιακές μελέτες οι οποίες όμως δίνουν ανομοιογενή αποτελέσματα λόγω εθνικών διαφορών [32] και συνεπώς συστηματικές υποδιαγνώσεις [33]. Παρόλο που η βιοψία αποτελεί την κύρια μέθοδο για εξακρίβωση των ιστοπαθολογικών αλλαγών και άρα του σταδίου [10], μεγάλη βάση έχει δοθεί σε μη-επεμβατικές μεθόδους, δηλαδή αναγνώριση βιοδεικτών που δίνουν τις ίδιες αξιόπιστες πληροφορίες χωρίς την ανάγκη για βιοψία.

Πίνακα 2. Οι 3 βαθμοί κατηγοριοποίησης της μη-αλκοολικής στεατοηπατίτιδας και τα χαρακτηριστικά τους. Η πρώτον βαθμού του βαθμού ή αλλιώς ήπιας έντασης (mild), δεύτερου βαθμού ή μέτριας έντασης (moderate) και η τρίτου βαθμού ή βαριάς έντασης (severe). DOI πίνακα: 10.3748/wjg.v20.i42.15539

Grade 1 (mild)

Steatosis: up to 66%

Ballooning: occasional in zone 3

Intralobular inflammation: scattered polymorphs ± lymphocytes

Portal inflammation: no or mild

Grade 2 (moderate)

Steatosis: any degree

Ballooning: obvious, predominantly zone 3

Intralobular inflammation: polymorphs and chronic inflammation noted

Portal inflammation: mild to moderate

Grade 3 (severe)

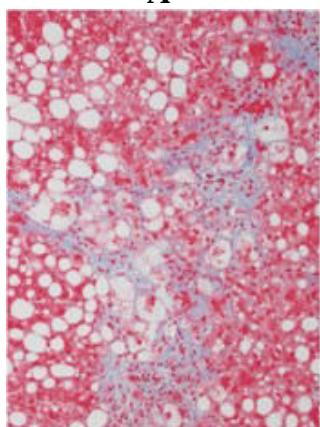
Steatosis: panacinar

Ballooning: ballooning and disarray obvious, predominantly in zone 3

Intralobular inflammation: scattered polymorphs ± mild chronic inflammation

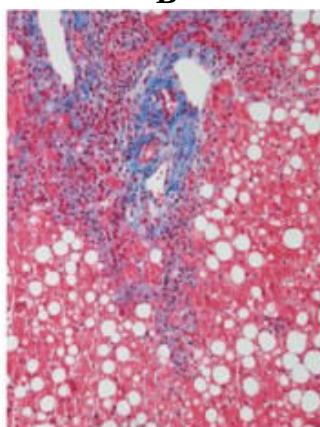
Portal inflammation: mild or moderate

A



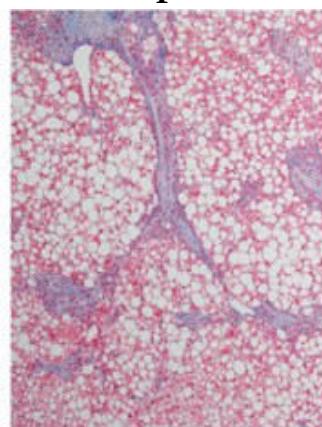
Perisinusoidal
fibrosis

B



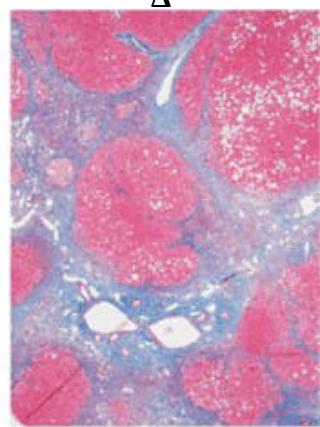
Periportal
fibrosis

Γ



Bridging
Fibrosis
(Stage 3)

Δ



Cirrhosis
(Stage 4)

Εικόνα 6. Τα 4 στάδια της μη-αλκοολικής στεατοηπατίτιδας βάση της ίνωσης. Με μπλε χρώμα φαίνεται το κολλαγόνο. (Α) Πρώτο στάδιο (NASH F1). Ίνωση στα περικολποειδή αγγεία της ζώνης 3. Το κολλαγόνο βρίσκεται γύρω από τα περικολποειδή που απαντώνται γύρω από την κεντρική φλέβα. (Β) Δεύτερο στάδιο (NASH F2). Ίνωση στα περικολποειδή τριχοειδή της ζώνης 3 με λιγότερη εναπόθεση κολλαγόνου και περιπλαίσια ίνωση όπου έχουμε μεγαλύτερη συσσώρευση κολλαγόνου. (Γ) Τρίτο στάδιο (NASH F3), γεφυροποιούσα ίνωση. Ένα αγγειωμένο διάφραγμα ινώδους ιστού τέμνει το ηπατικό παρέγχυμα. (Δ) Τέταρτη στάδιο (NASH F4), κίρρωση. Διάφραγμα ινώδους ιστού βρίσκεται γύρω από οζίδια ηπατοκυττάρων. DOI εικόνας: 10.1007/s40290-019-00305-z

Υποκεφάλαιο 1.6.1: Βιοψία

Οι τεχνικές δυσκολίες των μη-επεμβατικών μεθόδων και αρκετοί συγχυτές όπως τα τα λιπίδια και γλυκαυμικοί παράγοντες[34] (γλυκόζη, γλυκοζυλιωμένη αιμοσφαιρίνη - HbA1c), καθιστούν την βιοψία το πιο αξιόπιστο εργαλείο για την αξιολόγηση της νόσου αλλά επέτρεψαν και την δημιουργία της κατηγοριοποίησης με βάση το NAS (NAFLD Activity Score) που βασίζεται σε ιστολογικές αλλαγές του ήπατος [25]. Πρέπει να σημειωθεί πως η βιοψία συνήθως αποφεύγεται σε περιπτώσεις όπου ο ασθενής είναι επιρρεπής στη νόσο και προτιμάται σε περιπτώσεις όπου υπάρχει υποψία για στεατοηπατίτιδα. Ο λόγος είναι γιατί η βιοψία όπως και όλες οι επεμβατικές διαδικασίες συνοδεύονται από επικινδυνότητα. Η αιμορραγία και ο πόνος στο σημείο που γίνεται η δειγματοληψία (72 ώρες) είναι το πιο κοινό, με περαιτέρω επιπλοκές να σπανίζουν. Άλλοι περιοριστικοί παράγοντες αποτελούν η ασυνέπεια στην δειγματοληψία και η ανάλυση του δείγματος με αποτέλεσμα λάθος διάγνωση [35,36]. Φυσικά αυτά τα προβλήματα μπορούν εύκολα να παρακαμφθούν με σωστή χρήση των κλινικών οδηγιών.

Υποκεφάλαιο 1.6.2: Μη-επεμβατικές μέθοδοι

Κλινικά, για προκαταρκτικό έλεγχο, προτιμάται η χρήση τεχνικών εικονογράφησης (μαγνητική τομογραφία, υπέρηχος, υπολογιστική τομογραφία, CAP [37], H-MRS [38]) με κυριότερη τον υπέρηχο λόγω της προσβασιμότητας και απλότητας στη χρήση του [10]. Σε αντίθεση όμως με τα ακριβότερα μηχανήματα (όπως ο μαγνητικός τομογράφος), ο υπέρηχος είναι λιγότερο ευαίσθητος και υπάρχει μεγαλύτερη πιθανότητα σφάλματος [39]. Λόγω της ποιοτικής φύσεως των τεχνικών εικονογράφησης, γίνεται χρήση κυκλοφορούντων βιοδεικτών (πολλές φορές και σε συνδυασμό με προ-κλινικές παραμέτρους) για την συμπλήρωση και ποσοτικοποίηση της διάγνωσης. Ο συνδυασμός βιοχημικών και ανθρωπομετρικών παραμέτρων παρόλα αυτά δεν μπορεί να ποσοτικοποιήσει επακριβώς την στεάτωση [40,41] ή να ξεχωρίσει την απλή στεάτωση (NAFL) από τη στεατοηπατίτιδα (NASH) δημιουργώντας την ανάγκη για βιοψίες [10]. Επιπλέον πρόβλημα των μη-επεμβατικών μεθόδων είναι η επιρρέπεια σε παράγοντες όπως η σπλαχνική παχυσαρκία [42] και η γήρανση [43], χωρίς αυτό να αποκλείει όμως την χρησιμότητα τους ως μεσολαβητές της νόσου. Πρέπει επίσης να σημειωθεί πως η δραστηριότητά των βιοδεικτών *in vivo* δεν αντιπροσωπεύει πάντα και την κυκλοφορική τους αφθονία.

Επειδή στην παρούσα μελέτη εξετάστηκε ο ρόλος των μη κωδικών RNA στη ηπατική νόσο NAFLD, στη συνέχεια παρουσιάζονται δεδομένα που αφορούν στα μη κωδικά RNAs.

Υποκεφάλαιο 1.7: Κωδικά και μη-κωδικά γονίδια

Στο γονιδίωμα οι περιοχές που μεταγράφονται, μπορούν να χωριστούν σε κωδικές, δηλαδή περιοχές του DNA που μπορούν να μεταφραστούν σε πρωτεΐνες και σε μη κωδικές, δηλαδή περιοχές που δεν θα μεταφραστούν. Στον άνθρωπο περίπου το 41% του γονιδιώματος αποτελείται από γονίδια που κωδικοποιούν πρωτεΐνες (Protein Coding Genes, PCGs) και άρα μπορεί να θεωρηθεί λειτουργικό. Παρόλα αυτά η αλληλουχία των γονιδίων αυτών δεν ξεπερνά το 1-2% του γονιδιώματος (σύμφωνα με το NCBI), με το υπόλοιπο γονιδίωμα να θεωρείται μη-κωδικό. Με τη βοήθεια μεγάλων πρότζεκτ όπως το ENCODE και το FANTOM, έχουμε αρχίσει τα τελευταία χρόνια να μαθαίνουμε και να κατανοούμε περισσότερα πράγματα για τη πολυπλοκότητα και τη λειτουργία των RNA. Συγκεκριμένα φαίνεται πως το 80% του γενωμικού DNA των θηλαστικών μεταγράφεται και ελέγχεται αποτελεσματικά, με το μεγαλύτερο ποσοστό να ανήκει σε γονίδια μη-κωδικών RNA (non coding RNA, ncRNA) [44].

Υποκεφάλαιο 1.8: Μη-κωδικό RNA

Τα μη-κωδικά RNA χωρίζονται σε 2 βασικές ομάδες με βάση το μέγεθός τους. Η πρώτη είναι τα μικρά μη-κωδικά RNA που έχουν μήκος λιγότερο από 200 νουκλεοτίδια και περιλαμβάνει κατηγορίες όπως τα microRNAs (miRNAs), τα μικρά πυρηνικά RNAs (snRNAs), small interfering RNAs (siRNAs), μεταφορικά RNAs (tRNAs) καθώς και θραύσματα αυτών (transfer RNA-derived fragments, tRFs). Η δεύτερη και μεγαλύτερη ομάδα των μη-κωδικών RNAs, είναι τα μακρά μη κωδικά RNAs (long non-coding RNAs, lncRNAs) με μήκος 200 ή περισσότερα νουκλεοτίδια και συχνά ταξινομούνται ανάλογα με την θέση του γενετικού τους τόπου σε σχέση με τα κοντινά τους γονίδια σε διάφορες κατηγορίες (π.χ. intergenic, intronic, antisense)[45].

Υποκεφάλαιο 1.9: Μακρύ μη-κωδικό RNA & Βασικά χαρακτηριστικά

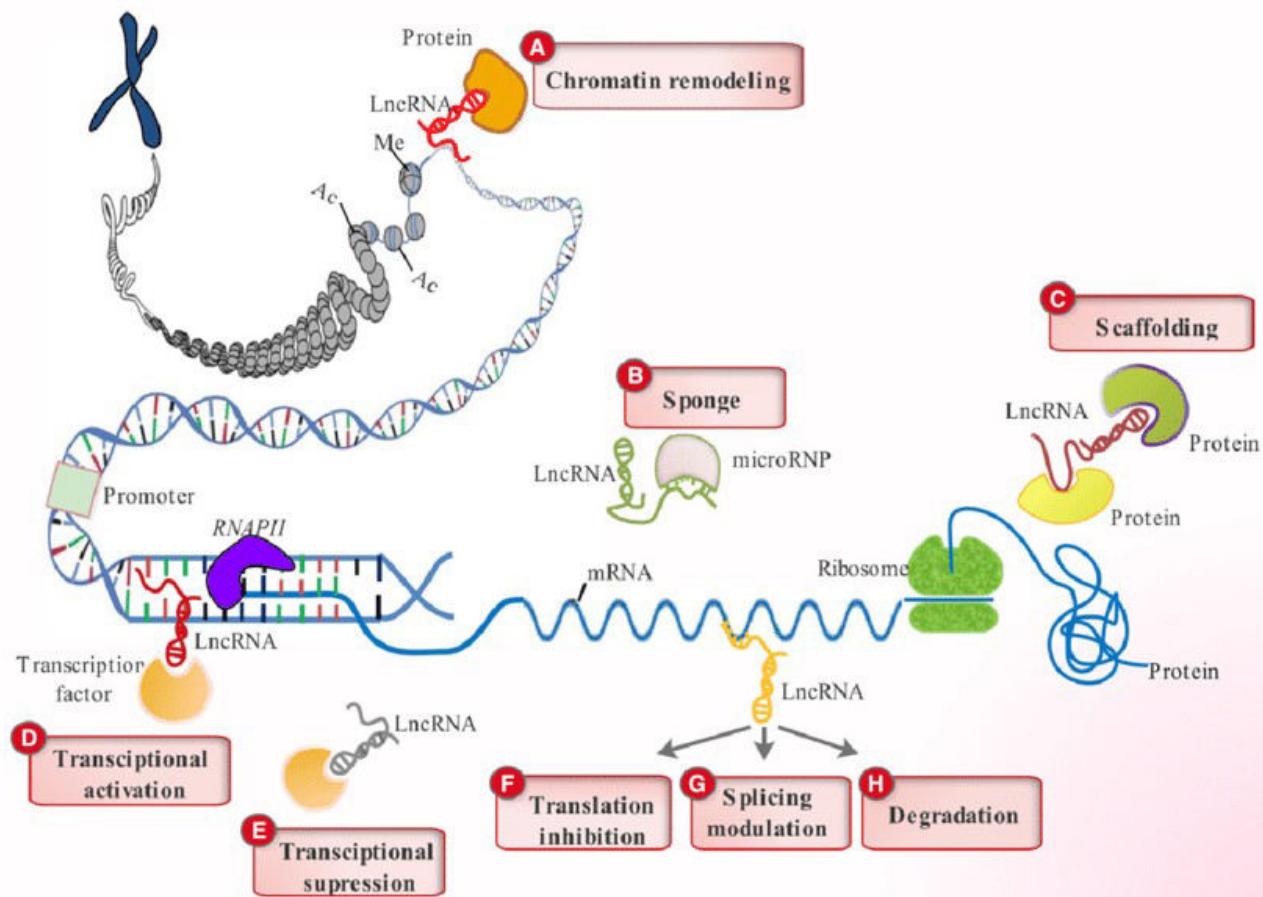
Στη μεταγραφή των μακρών μη-κωδικών RNA συμμετέχουν διάφορες RNA πολυμεράσες, αλλά βασικότερο ρόλο έχει η RNA πολυμεράση II. Παρατηρούνται ομοιότητες με το αγγελιαφόρο RNA

(mRNA) ως προς τις τροποποιήσεις και την ωρίμανσή τους, καθώς ο μηχανισμός επεξεργασίας και ωρίμανσης τους είναι ίδιος. Συγκεκριμένα, παρατηρείται συχνά προσθήκη καλύπτρας (cap) στο 5' άκρο με 7-μέθυλ-γουανοσίνη (m7G) και προσθήκη πολύ-A ουράς (poly-A tail) στο 3' άκρο. Ωστόσο, τα lncRNAs παρουσιάζουν χαμηλότερα επίπεδα ματίσματος (splicing) από τα mRNAs, και υψηλότερα επίπεδα εναλλακτικού ματίσματος[46]. Η διαφορική έκφρασή των μακρών μη-κωδικών RNA σε διαφορετικά στάδια της ανάπτυξης, καθώς και η ταυτόχρονη ρύθμισή τους με άλλα γονίδια που κωδικοποιούν πρωτεΐνες, όπως μεταγραφικούς παράγοντες (Transcription Factors, TFs) και παράγοντες αναδιαμόρφωσης της χρωματίνης (Chromatin Remodelers), έχουν ως αποτέλεσμα την οργανωμένη αναπροσαρμογή του νεοσυντιθέμενου μεταγραφώματος κατά τη διαδικασία της διαφοροποίησης[47]. Επίσης, η κυτταροειδική και ιστοειδική έκφρασή των lncRNAs τα καθιστά σημαντικά στη μελέτη ασθενειών ενώ ταυτόχρονα είναι πιθανοί στόχοι για θεραπεία[48].

Υποκεφάλαιο 1.9.1: Μακρύ μη-κωδικό RNA – Ρόλος και Λειτουργίες

Τα lncRNAs αλληλεπιδρούν με DNA, mRNA, πρωτεΐνες που δεσμεύουν RNA (RNA-Binding Proteins, RBPs), miRNA και επιτελούν μια σειρά από λειτουργίες(Εικόνα 7). Ελέγχουν την έκφραση γονιδίων σε πολλά επίπεδα (επιγενετικά, μεταγραφικά, μετα-μεταγραφικά, μεταφραστικά και μετα-μεταφραστικά)[49] καθώς και ενισχύουν ή περιορίζουν τον σχηματισμό μεταγραφικής λούπας, ενώ μπορούν να στρατολογήσουν ή να μπλοκάρουν άλλους ρυθμιστές. Έχουν απαραίτητο και αναγκαίο ρόλο στη μεταγραφική ρύθμιση, τον οποίο τον επιτελούν αλληλεπιδρώντας με άλλα μόρια κοντά στους υποκινητές ή τους ενισχυτές των γονιδίων στόχων. Ταυτόχρονα λειτουργούν σαν πρόδρομα άλλων μη-κωδικών RNA (όπως τα miRNAs), και επηρεάζουν το μάτισμα των mRNAs [50]. Ο ρόλος τους όμως δεν σταματάει εκεί καθώς συμμετέχουν σε πολλά άλλα μονοπάτια, όπως την επεξεργασία μορίων RNA, την αναδιαμόρφωση της χρωματίνης (αναδιάπλαση χρωματίνης, τροποποιήσεις ιστόνης, μεθυλίωση DNA), τη μετάφραση mRNA και δρουν ως ογκογονίδια ή καταστολείς όγκων, ρυθμίζοντας μια σειρά σηματοδοτικών μονοπατιών [51]. Με βάση τη λειτουργικότητά τους, τα lncRNA διαχωρίζονται στα μη-λειτουργικά, που αποτελούν πιθανά παραπροϊόντα της μεταγραφής, σε αυτά που η διαδικασία της μεταγραφής τους είναι επαρκής ώστε να δράσουν αλλά το μετάγραφο δεν έχει κάποια περεταίρω λειτουργική σημασία, και τέλος αυτά που παρουσιάζουν διάφορες λειτουργίες. Αυτές μπορεί να είναι *in cis*, είτε παραμένοντας προσδεμένα πάνω ή κοντά στην περιοχή μεταγραφής τους,

είτε αλληλεπιδρώντας με πρωτεΐνες όπως παράγοντες της μεταγραφής και της αναδιαμόρφωσης της χρωματίνης. [52,53,54,55].

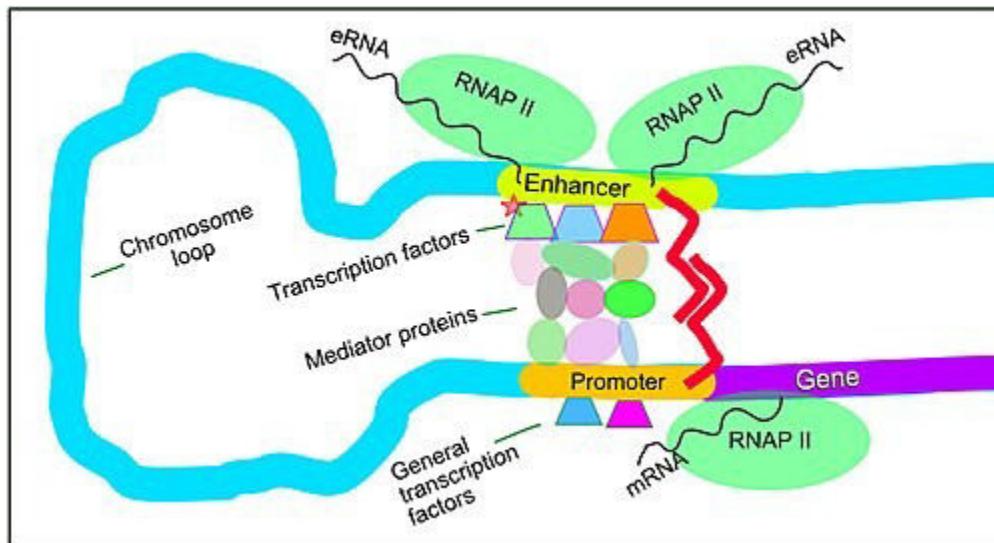


Εικόνα 7. Λειτουργίες των lncRNAs. (A) Τα lncRNAs στρατολογούν δομικά στοιχεία πρωτεϊνών που συμμετέχουν στο σύμπλοκο αναδιαμόρφωσης της χρωματίνης και αλλάζουν το μοτίβο οργάνωσης. (B) Προσδένονται με το συμπληρωματικό τους miRNA και μειώνουν την δράση τους. (C) Παρέχουν θέσεις πρόσδεσης για πρωτεΐνες (RBPs) που που λειτουργούν στο ίδιο μονοπάτι με αυτές, έτσι αποκτούν ρόλο ικριωμάτων (scaffolding). (D,E) Ελέγχουν τη μεταγραφή. (D) Ενεργοποιούν τη μεταγραφή κάποιων γονιδίων με το να προσδένουν και να καθοδηγούν μεταγραφικούς παράγοντες (TFs) στους υποκινητές τους. (E) Καταστέλλουν τη μεταγραφή απομακρύνοντας μεταγραφικούς παράγοντες από τους υποκινητές. (F,G,H) Ρυθμίζουν τις λειτουργίες του mRNA με πρόσδεση σε αυτό. (F: Καταστολή μετάφρασης, G: Αλλαγή μοτίβου ματίσματος, H: Τα οδηγούν σε αποδόμηση). DOI εικόνας: 10.1111/jcmm.13238

Υποκεφάλαιο 1.10: Μακρύ μη-κωδικό RNA – eRNAs

Τα eRNAs (enhancer RNAs) είναι ένα είδος lncRNA που παράγεται από αμφίδρομη (bidirectional) μεταγραφή περιοχής όπου βρίσκονται ενεργοί ενισχυτές γονιδίων και δρουν μαζί με το DNA για να ενισχύσουν τη μεταγραφή δημιουργώντας τη μεταγραφική λούπα[56,57](Εικόνα 8). Πρόκειται για μετάγραφα μικρού μήκους (200-600 νουκλεοτίδια) που αποδομούνται πολύ γρήγορα από το εξώσωμα

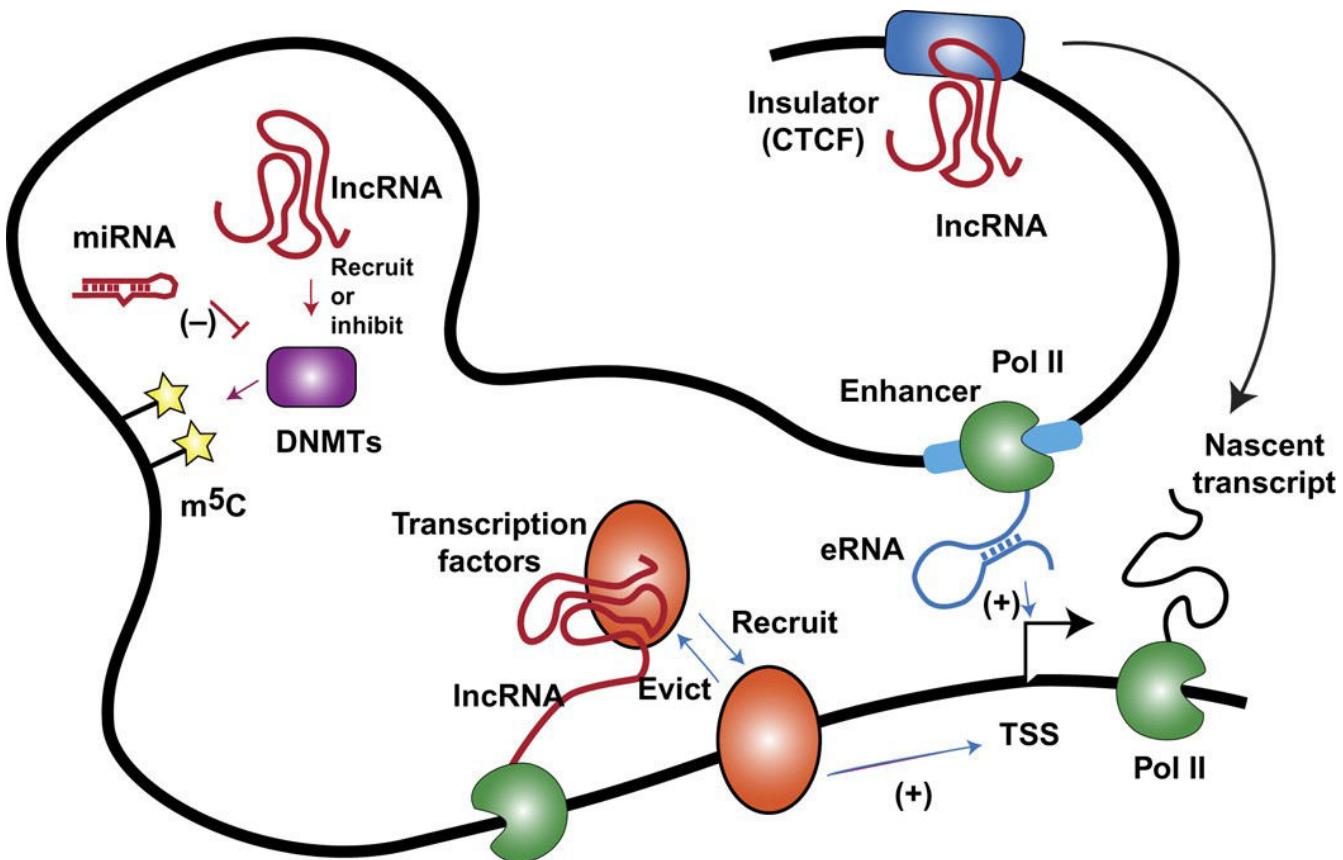
(exosome depletion) και είναι δύσκολο να εντοπιστούν σε σταθερή κατάσταση (steady state). Για τον λόγο αυτό είναι πολύ σημαντική η μελέτη του νεοσυντιθέμενου RNA μεταγραφώματος (nascent RNA transcriptome), όπου μπορούν να εντοπιστούν περισσότερα lncRNAs προτού αποδομηθούν[58] (Εικόνα 9).



Εικόνα 8. Μεταγραφική λούπα όπου φαίνεται η αμφίδρομη παραγωγή του eRNA. Η περιοχή όπου βρίσκονται ενεργοί ενισχυτές γονιδίων αλληλεπιδρά με τους υποκινητές της περιοχής του γονιδίου που πρόκειται να μεταγραφεί και σχηματίζει τη μεταγραφική λούπα. Έτσι μπορεί να ξεκινήσει η σύνθεση mRNA από την RNA πολυμεράση II, προσδεμένη στον υποκινητή του σημείου έναρξης μεταγραφής. Η λούπα σταθεροποιείται από την αλληλεπίδραση 2 πρωτεϊνών όπου η μία βρίσκεται προσδεμένη στον στον ενισχυτή και η άλλη στον υποκινητή (κόκκινες γραμμές). DOI εικόνας: https://en.wikipedia.org/wiki/Enhancer_RNA

Υποκεφάλαιο 1.10.1: Μακρύ μη-κωδικό RNA – elncRNAs

Μία άλλη κατηγορία των lncRNAs που παράγονται από αμφίδρομα μεταγραφόμενες περιοχές ενισχυτών, είναι τα enhancer-associated μακρά μη-κωδικά RNAs (elncRNAs). Τα elncRNAs μεταγράφονται από μια εκ των δύο κατευθύνσεων και κοντά στο κέντρο ενός μικρού ποσοστού ενεργών ενισχυτών (3-5%) και παρουσιάζουν μεγαλύτερη σταθερότητα από τα eRNAs. Επίσης, παρουσιάζουν υψηλότερα επίπεδα ματίσματος, συγκρίσιμα με τα mRNA, καθώς επίσης τίθενται υπό επεξεργασία από τα σωμάτια ματίσματος όσο ακόμα μεταγράφονται (co-transcriptional splicing) [59,60].

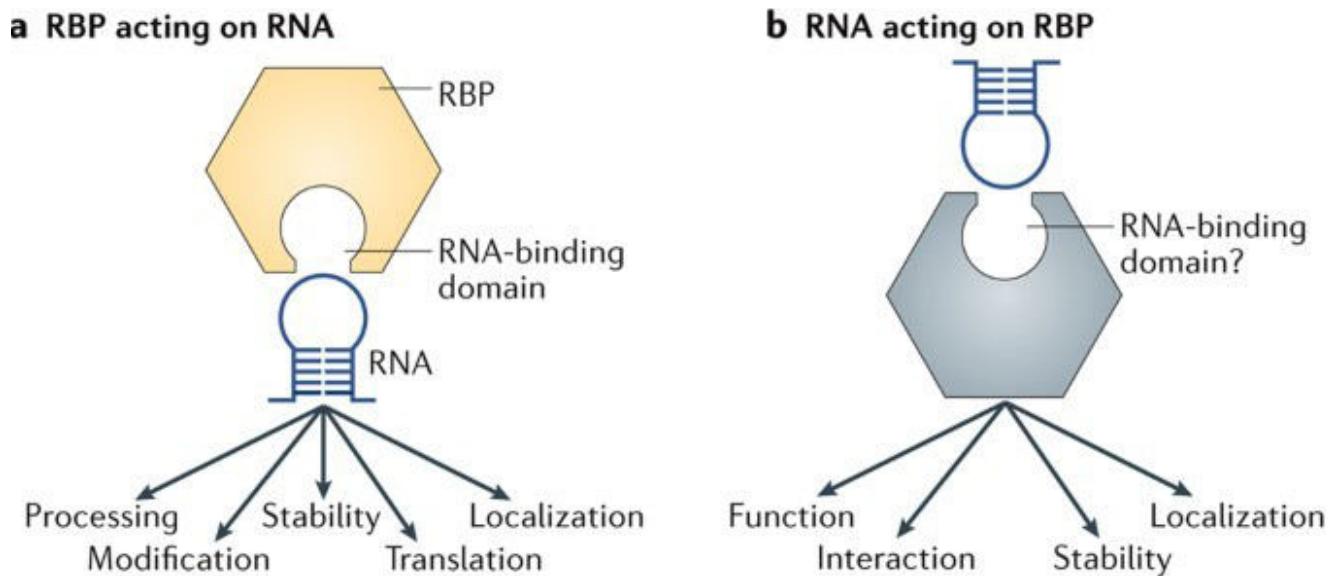


Εικόνα 9. Μεταγραφική λούπα. Η αλληλεπίδραση lncRNAs με μεταγραφικούς παράγοντες (TFs) in cis. Φαίνεται η αλληλεπίδραση lncRNA με τον CTCF (chromatin insulator CCCTC-binding factor) για τη ρύθμιση μεταγραφής. Αναπαριστάται το νεοσυντιθέμενο μεταγράφωμα (nascent transcript). (TSS: Transcription start site – Σημείο έναρξης μεταγραφής) Η μεταγραφή ενισχυτών οδηγεί στη παραγωγή eRNAs και συνεισφέρουν στην ενεργοποίηση γονιδίων ή στην αναδιαμόρφωση χρωματίνης. DOI εικόνας: 10.1126/sciadv.aoa2110

Υποκεφάλαιο 1.11: Πρωτεΐνες που προσδένουν RNA

Οι πρωτεΐνες που προσδένουν RNA (RNA Binding Proteins, RBPs) είναι ζωτικής σημασίας για τους οργανισμούς διότι κάποιες από τις βασικές λειτουργίες τους είναι να ελέγχουν ρυθμιστικά μονοπάτια (Εικόνα 10) . Μεγάλης σημασίας είναι η ικανότητα τους να ρυθμίζουν την έκφραση γονιδίων, καθώς και οι μετα-μεταγραφικές λειτουργίες όπως το μάτισμα, η μεταφορά των μεταγράφων και η ρύθμιση της μετάφρασης και της αποδόμησης (turnover) του mRNA[61]. Η έκφραση και ενεργότητα τους προσαρμόζεται στις κυτταρικές ανάγκες των RNA μεταγράφων και φαίνεται πως η μη-έκφρασή τους οδηγεί σε πολλαπλά προβλήματα στη φυσιολογία του κυττάρου ενώ ταυτόχρονα συμβάλλουν σε φαινοτυπικές ανωμαλίες άτυπων και καρκινικών κυττάρων[62]. Παρόλο που δεν έχουν αναγνωριστεί πολλά RBPs στον άνθρωπο, έχουν βρεθεί πέρα από τα κανονικά RBPs (canonical RBPs), τα μη-

κανονικά RBPs (non-canonical RBPs), τα οποία είναι πρωτεΐνες που η βασική τους λειτουργία δεν είναι η πρόσδεση RNA αλλά την επιτελούν σε συγκεκριμένες συνθήκες[63].



Εικόνα 10. (a) Οι πρωτεΐνες που προσδένουν RNA (RBPs) προσδένουν το RNA σε ειδικές περιοχές που ονομάζονται RBD (RNA-binding domain) ώστε να ρυθμίσουν το μεταβολισμό και τη λειτουργία των RNAs. (b) Αντίθετα η πρόσδεση του RNA στην RBP αλλάζει την λειτουργία και την τύχη της. DOI εικόνας: <https://doi.org/10.1038/nrm.2017.130>

ΣΚΟΠΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Το μεταγραφικό προφίλ με δείγματα ηπατοκυττάρων από ασθενείς με μη αλκοολική λιπώδης νόσο του ήπατος χαρακτηρίστηκε πρόσφατα μέσω πειραμάτων αλληλούχισης RNA (RNA-Seq). Η χρήση αλληλούχισης συνολικού RNA αντί μεθόδων υβριδοποίησης σε μικροσυστοιχίες, προσεγγίζει καλύτερα το μεταγραφικό προφίλ. Η ανάλυση των δεδομένων αυτών εστίασε στην έκφραση γονιδίων που κωδικοποιούν πρωτεΐνες (PCGs)[64], όμως ο ρόλος μη κωδικών μεταγραφών στην εξέλιξη της νόσου παραμένει ασαφής. Σκοπός της πτυχιακής εργασίας ήταν η μελέτη του μεταγραφικού προφίλ της μη-αλκοολικής λιπώδης νόσου του ήπατος με μεθόδους βιοπληροφορικής και έμφαση στο μακρύ μη κωδικό RNA (lncRNA) ώστε να καταλάβουμε το ρόλο του καθώς η ασθένεια εξελίσσεται. Η αναγνώριση στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων (Differentially Expressed Genes, DEGs), πιθανώς οδηγήσει στην καλύτερη κατανόηση της νόσου και ίσως εμφανίσει νέους υποψήφιους βιοδείκτες και θεραπευτικούς στόχους. Για τη μελέτη αυτή, έγινε χρήση δημόσια διαθέσιμων δεδομένων (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135251>) που περιέχουν τις εκφράσεις γονιδίων (gene specific read counts) μετά από πείραμα αλληλούχισης (RNA-Seq -216 άτομα -10 control, 206 ασθενείς).

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ: ΜΕΘΟΔΟΛΟΓΙΑ

Υποκεφάλαιο 2.1: Συλλογή δεδομένων

Για την ανάλυση του μεταγραφικού προφίλ έγινε χρήση αρχείων που προήλθαν από πειράματα αλληλούχισης RNA (RNA-Seq) και περιέχουν τις εκφράσεις γονιδίων (gene specific read counts) από 216 άτομα. Από τα 216 άτομα, 10 δεν ήταν ασθενείς και χρησιμοποιήθηκαν ως control, 51 ήταν ασθενής μη-αλκοολικής λιπώδης νόσου του ήπατος με απλή στεάτωση (NAFL) και 155 ασθενής με μη-αλκοολική στεατοηπατίτιδα (NASH). Συγκεκριμένα υπήρχαν 34 ασθενής με μη-αλκοολική στεατοηπατίτιδα χωρίς ίνωση ή με ίνωση πρώτου σταδίου (NASH F0-F1), 52 με ίνωση δευτέρου σταδίου (NASH F2), 55 με ίνωση τρίτου σταδίου (NASH F3) και 14 με ίνωση τετάρτου σταδίου (NASH F4). Στη συνέχεια έγινε σύμπτυξη των 216 αρχείων σε 1 κοινό αρχείο (dataframe).

Υποκεφάλαιο 2.2: Εύρεση PCGs & lncRNAs

Για τον διαχωρισμό των γονιδίων και σύγκριση αποτελεσμάτων έγινε χρήση του γονιδιώματος GRCh38, Ensemble release 76, το οποίο είχε χρησιμοποιηθεί ως γονιδιώματα αναφοράς σε προηγούμενη έρευνα[64]. Κάθε αρχείο περιέχει 64,258 γονίδια στα οποία περιέχονται τα κωδικά και τα μη-κωδικά γονίδια. Για να μπορέσουμε να κάνουμε την ανάλυση του μεταγραφικού προφίλ, βρήκαμε όλα τα γονίδια στο γονιδιώμα αναφοράς και χωρίσαμε τα κωδικά και τα μακρά μη-κωδικά. Κρατήσαμε το Ensemble Gene ID, το όνομα του γονιδίου και το είδος του γονιδίου στη περίπτωση των lncRNAs (π.χ antisense, intronic κλπ). Αυτό έγινε σε γλώσσα AWK με τη χρήση των παρακάτω εντολών:

```
awk '$3 ~/gene/' Homo_sapiens.GRCh38.76.gtf > out
```

Για τα κωδικά γονίδια

```
awk '$18 ~/protein_coding/' out > prot
awk '{if ( $3 ~/gene/ && $18 ~/protein_coding/) print $10 "\t" $14 "\t" $18}'
Homo_sapiens.GRCh38.76.gtf > outprot1
awk '{gsub ('/','"', $0); print $0}' outprot1 > outprot2
awk '{gsub ('/','"', $0); print $0}' outprot2 > outprot3
awk '{print $1 "\t" $2 "\t" $3}' outprot3 > outprot4
```

Για τα μακρά μη-κωδικά γονίδια

```
awk '{if ($18 !~/protein_coding/) print $18}' out | sort | uniq
awk '$18 ~3prime_overlapping_ncrna/ || $18 ~antisense/ || $18 ~lincRNA/ || $18
~/sense_intronic/ || $18 ~sense_overlapping/' out > lnc
awk '{print $10 "\t" $14 "\t" $18}' lnc > outlnc1
awk '{gsub ('/','"', $0); print $0}' outlnc1 > outlnc2
awk '{gsub ('/','"', $0); print $0}' outlnc2 > outlnc3
awk '{print $1 "\t" $2 "\t" $3}' outlnc3 > outlnc4
```

Από τις εντολές αυτές πήραμε 21,983 γονίδια που κωδικοποιούν πρωτεΐνες (PCGs) και 14,407 μακρά μη-κωδικά γονίδια (lncRNAs).

Τα επόμενα βήματα εκτός του AML (Automated Machine Learning) έγιναν στη γλώσσα R μέσω Emacs Speaks Statistics (ESS) σε Debian 11 (Bullseye). Τα γονίδια που βρέθηκαν από το γονιδίωμα αναφοράς συγχωνεύτηκαν με τις εκφράσεις γονιδίων των control και ασθενών, ξεχωριστά για τα PCGs και για τα lncRNAs. Στη συνέχεια έγινε αφαίρεση των γονιδίων που δεν εκφραζόντουσαν σε κανένα από τα άτομα και τελικά είχαμε 18,401 PCGs και 10,902 lncRNAs.

Υποκεφάλαιο 2.3: Ανάλυση κύριων συνιστωσών

Ανάλυση κύριων συνιστωσών (Principal Component Analysis, PCA) διεξήχθη μεταξύ των δειγμάτων control και των ασθενών NASH F2, καθώς 2 από τα control δείγματα είχαν χαρακτηριστεί με ίνωση δεύτερου σταδίου (F2). Για την PCA έγινε χρήση των πακέτων “pls”, “ROCR” και η απεικόνιση με το πακέτο “ggbiplot”. Τα δεδομένα που εισάγθηκαν ήταν τα 18,401 PCGs μετά από κανονικοποίηση με την λειτουργία (function) VST (17,940 μετά την κανονικοποίηση).

Υποκεφάλαιο 2.4: Ανάλυση διαφορικής έκφρασης

Για να δούμε τη διαφορική έκφραση των γονιδίων, πήραμε 3 επίπεδα αναφοράς (Control, NAFL, NASH F0-F1) και τα συγκρίναμε κάθε φορά με τα επόμενα στάδια. Για παράδειγμα συγκρίναμε τα control δείγματα σε μια ξεχωριστή ανάλυση με τους NAFL ασθενείς, σε μία άλλη με τους NASH F0-F1 ασθενείς κ.ο.κ. Αυτό έγινε τόσο για τα PCGs όσο για τα lncRNAs. Συνολικά έγιναν 10 αναλύσεις με επίπεδο αναφοράς τα control δείγματα (5 για τα PCGs και 5 για τα lncRNAs), 8 αναλύσεις με επίπεδο αναφοράς τα NAFL δείγματα (4 για τα PCGs και 4 για τα lncRNAs) και 6 αναλύσεις με επίπεδο αναφοράς τα NASH F0-F1 δείγματα (3 για τα PCGs και 3 για τα lncRNAs). Η ανάλυση διαφορικής έκφρασης έγινε με το πακέτο “DESeq2”. Το πακέτο αυτό υπολογίζει το p-value με Wald test, το οποίο στη συνέχεια προσαρμόζεται με τη μέθοδο Benjamini και Hochberg και προκύπτει το padj, το οποίο χρησιμοποιούμε. Στη συνέχεια, σε κάθε μια από τις 24 αναλύσεις, έγινε αφαίρεση των N/A με βάση το p value adjusted (padj). Από το dataframe που προέκυψε, για να βρούμε τα στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια (Differentially Expressed Genes, DEGs) χρησιμοποιήσαμε

2 κατώφλια (cutoff), ένα για $\text{padj} < 0,005$ και ένα για $\text{padj} < 0,001$ με την λειτουργία (function) subset. Έγινε επιλογή πιο αυστηρών στατιστικών κριτηρίων με σκοπό να βρεθούν τα πιο σημαντικά από τα διαφορικά εκφραζόμενα γονίδια. Όλες οι αναλύσεις έγιναν και με $\text{padj} < 0,01$ για επαλήθευση πως δεν υπάρχει πολύ μεγάλη απόκλιση όταν δίνουμε μικρότερο περιθώριο p-value.

Μία επιπλέον ανάλυση διαφορικής έκφρασης έγινε μόνο για τα γονίδια των RBPs, ακολουθώντας τα ίδια βήματα, με διαφορά πως εστίασε μόνο στα αρχικά στάδια της εξέλιξης της νόσου, δηλαδή έγινε σύγκριση των control δειγμάτων με τα NAFL και τα NASH F0-F1 δείγματα ξεχωριστά. Τα RBPs που χρησιμοποιήθηκαν εδώ ήταν ~2,300, προήλθαν από αντίστοιχη έρευνα στο ήπαρ[65] και έγινε σύμπτυξη τους βάση Ensembl ID με τα αρχικά αρχεία των ασθενών[64].

Υποκεφάλαιο 2.4.1: Απεικόνιση αποτελεσμάτων διαφορικής έκφρασης

Η απεικόνιση αποτελεσμάτων έγινε με γράφημα τύπου ηφαιστείου (Volcano Plot) μέσω του πακέτου “ggplot2” και με διάγραμμα Venn μέσω του πακέτου “ggVennDiagram” και “ggplot2” συνδυαστικά με “ggrepel”. Για το γράφημα τύπου ηφαιστείου έγινε χρήση των αρχικών αποτελεσμάτων της διαφορικής έκφρασης (DESeq2) μετά από την αφαίρεση των N/A τιμών και χρήση του αρνητικού λογαρίθμου του $\text{padj}(-\log_{10}(\text{padj}))$ και του λογαρίθμου της αλλαγής διπλώματος ($\log_2(\text{FC})$) στην απεικόνιση. Το διάγραμμα Venn επιτρέπει να φανούν τα κοινά διαφορικά εκφραζόμενα γονίδια καθώς η νόσος εξελίσσεται. Στη προκειμένη περίπτωση τα δεδομένα που χρησιμοποιήθηκαν ήταν τα αποτελέσματα της διαφορικής έκφρασης μετά τον διαχωρισμό βάση του padj cutoff.

Υποκεφάλαιο 2.5: Αυτοματοποιημένη μηχανική μάθηση

Ένας πιο εξειδικευμένος τρόπος να δούμε πόσο σημαντικά ή επαναλήψιμα είναι τα αποτελέσματα της ανάλυσης διαφορικής έκφρασης αλλά και να λάβουμε περισσότερες γνώσεις πάνω σε αυτά, είναι η χρήση αυτοματοποιημένης μηχανικής μάθησης (Automated Machine Learning, AML). Η AML προβλέπει την καλύτερη προσέγγιση (π.χ Random Forest, Ridge Logistic Regression, Support Vector Machine) της ανάλυσης που θέλουμε να κάνουμε.

Για το σκοπό αυτό έγιναν οι παρακάτω δυαδικές συγκρίσεις (δηλαδή σύγκριση δύο σταδίων της νόσου):

- Control vs NAFL
- Control vs NASH F0-F1
- NAFL vs NASH F4
- NASH F0-F1 vs NASH F4

Οι παραπάνω συγκρίσεις έγιναν ξεχωριστά για τα PCGs, lncRNAs, RBPs αλλά έγιναν και με σύμπτυξη PCGs & lncRNAs με σκοπό να δούμε εάν υπάρχει κάποια συσχέτιση μεταξύ τους αλλά και αν προβλέπουν καλύτερα την εξέλιξη της νόσου.

Επιπλέον έγιναν οι εξής αναλύσεις μόνο για τα lncRNAs:

- Παλινδρόμηση βάση του σκορ NAS
- Παλινδρόμηση βάση τα στάδια ίνωσης
- Πολυμεταβλητή (Multiclassification) ανάλυση (όλα τα στάδια της νόσου λήφθηκαν υπόψη)

Τα δεδομένα που χρησιμοποιήθηκαν ήταν οι εκφράσεις γονιδίων (gene specific read counts) προτού γίνει κάποια ανάλυση. Η προ-επεξεργασία (preprocessing) έγινε με αφαίρεση γονιδίων χωρίς έκφραση και γονιδίων με μέγιστη σχετική έκφραση 1 και χρήση του συντελεστή μεταβλητότητας (Coefficient of Variation) για έλεγχο πιθανών έκτοπων τιμών με τις εντολές:

```
df1 ← sapply(df, function(x) sd(x, na.rm=T) / mean(x, na.rm=T) * 100)
boxplot(df1)
```

Όταν τα αποτελέσματα της AML ήταν από 1 έως 5 γονίδια, ίσως η έκφραση των γονιδίων αυτών υποσκίαζε τα υπόλοιπα. Συνεπώς, γινόταν αφαίρεση των γονιδίων αυτών και επανάληψη της ανάλυσης. Οι αναλύσεις αυτές διεξήχθη στη πλατφόρμα JADBio (<https://jadbio.com/>, Ιωάννης Τσαμαρδίνος).

Υποκεφάλαιο 2.6: Οντολογία γονιδίων & Εγκυκλοπαίδεια του Κιότο για τα γονίδια και τα γονιδιώματα

Η ανάλυση οντολογίας γονιδίων (Gene Ontology, GO) μας παρέχει πληροφορίες και μπορούμε να αναγνωρίσουμε τις βιολογικές διαδικασίες (Biological Processes, BP), το κυτταρικό διαμέρισμα (Cellular Compartment, CC) και τις μοριακές λειτουργίες (Molecular Functions, MF) στις οποίες εμπλέκονται τα γονίδια που μας ενδιαφέρουν. Συγκεκριμένα, η βιολογική διεργασία, περιλαμβάνει όρους που περιγράφουν μια ακολουθία διακριτών μοριακών λειτουργιών με καθορισμένη αρχή και τέλος, χωρίς όμως τη δυναμική και τις εξαρτήσεις μεταξύ τους. Το κυτταρικό διαμέρισμα, περιλαμβάνει όρους που περιγράφουν τις υποκυττάριες δομές του κυττάρου όπου εντοπίζονται τα προϊόντα του γονιδίου. Τέλος, η μοριακή λειτουργία, περιλαμβάνει όρους που περιγράφουν λειτουργίες μεμονωμένων ή συμπλόκων γονιδιακών προϊόντων οι οποίες εκτελούνται σε μοριακό επίπεδο, χωρίς λεπτομέρειες για το πλαίσιο που συμβαίνουν.

Η ανάλυση έγινε ξεχωριστά για τα στατιστικά σημαντικά ($p<0,01$) PCGs με αυξημένη έκφραση, βάση της αλλαγής διπλώματος (Fold Change, $FC \geq 1,5$). Για επαλήθευση των αποτελεσμάτων, έγινε ξεχωριστή ανάλυση με προσθήκη γονιδίων πολύ μικρής αλλαγής διπλώματος ή σταθερής έκφρασης ($FC = 0$). Αντίστοιχα, η εγκυκλοπαίδεια του Κιότο για τα γονίδια και τα γονιδιώματα (Kyoto Encyclopedia of Genes and Genomes, KEGG), είναι μια ελεύθερη βάση δεδομένων για τη συστηματική ανάλυση των γονιδιακών λειτουργιών και συνδέει γονιδιωματικές πληροφορίες με ανώτερης τάξης λειτουργικές πληροφορίες. Η ανάλυση KEGG έγινε με τα ίδια βήματα και παραμέτρους που αναφέρθηκαν στην ανάλυση οντολογίας γονιδίων. Το πακέτο με το οποίο έγιναν οι 2 αναλύσεις είναι το “DOSE” και η βάση δεδομένων αναρτήθηκε με την εντολή library(org.Hs.eg.db). Οι αναλύσεις έγιναν ξεχωριστά για BP, CC, MF και KEGG. Η απεικόνιση αποτελεσμάτων έγινε με το πακέτο “clusterProfiler”.

Τα αρχεία και οι εντολές που χρησιμοποιήθηκαν είναι διαθέσιμα στο Παράρτημα 13.

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ: ΑΠΟΤΕΛΕΣΜΑΤΑ

Υποκεφάλαιο 3.1: Ανάλυση διαφορικής έκφρασης

Μετά την αφαίρεση των τιμών N/A, τα αποτελέσματα της διαφορικής ανάλυσης φιλτραρίστηκαν με βάση το padj ώστε να κρατήσουμε μόνο τα στατιστικά σημαντικά γονίδια. Στους πίνακες 3-5 φαίνεται ο αριθμός των γονιδίων που βρέθηκαν μετά το κατώφλι αποκοπής(cutoff) $p<0,01$, $p<0,005$ & $p<0,001$. Να αναφερθεί πως η σύγκριση NAFL με NASH F0-F1 και NASH F0-F1 vs NASH F2 για τα lncRNAs δεν έδειξε κανένα στατιστικά σημαντικό διαφορικά εκφραζόμενο γονίδιο οπότε δεν αναπαριστάται στις εικόνες.

- Για τα PCGs τα αρχικά γονίδια ήταν 18,401 και μετά την αφαίρεση N/A ~16,000.
- Για τα lncRNAs τα αρχικά γονίδια ήταν 10,902 και μετά την αφαίρεση N/A ~4,500.
- Για τα RBPs τα αρχικά γονίδια ήταν 2,180 και μετά την αφαίρεση N/A ~2,160.

Πίνακας 3. Αποτελέσματα ανάλυσης διαφορικής έκφρασης για $p<0,01$ (DESeq2). Αναγράφονται οι αριθμοί των γονιδίων.

padj < 0,01	PCGs	lncRNAs
Control vs NAFL	5135	500
Control vs NASH F0-F1	4771	538
Control vs NASH F2	5417	601
Control vs NASH F3	4854	517
Control vs NASH F4	4259	322
NAFL vs NASH F0-F1	0	0
NAFL vs NASH F2	95	13
NAFL vs NASH F3	1791	136
NAFL vs NASH F4	1886	166
NASH F0-F1 vs NASH F2	25	0
NASH F0-F1 vs NASH F3	924	38
NASH F0-F1 vs NASH F4	2167	196

Πίνακας 4. Αποτελέσματα ανάλυσης διαφορικής έκφρασης για $p < 0,005$ (DESeq2). Αναγράφονται οι αριθμοί των γονιδίων.

padj < 0,005	PCGs	lncRNAs	RBPs
Control vs NAFL	4589	434	821
Control vs NASH F0-F1	4238	428	763
Control vs NASH F2	4853	493	-
Control vs NASH F3	4275	422	-
Control vs NASH F4	3661	249	-
NAFL vs NASH F0-F1	0	0	-
NAFL vs NASH F2	58	9	-
NAFL vs NASH F3	1389	88	-
NAFL vs NASH F4	1551	132	-
NASH F0-F1 vs NASH F2	11	0	-
NASH F0-F1 vs NASH F3	756	25	-
NASH F0-F1 vs NASH F4	1769	149	-

Πίνακας 5. Αποτελέσματα ανάλυσης διαφορικής έκφρασης για $p < 0,001$ (DESeq2). Αναγράφονται οι αριθμοί των γονιδίων. Σε παρένθεση φαίνονται τα αποτελέσματα της διαφορικής ανάλυσης από το άρθρο αναφοράς (Limma-Voom)[64].

padj < 0,001	PCGs	lncRNAs	RBPs
Control vs NAFL	3672	295	676
Control vs NASH F0-F1	3241	288	617
Control vs NASH F2	3815	327	-
Control vs NASH F3	3360	278	-
Control vs NASH F4	2555	147	-
NAFL vs NASH F0-F1	0 (vs 0)	0	-
NAFL vs NASH F2	21 (vs 50)	2	-
NAFL vs NASH F3	915 (vs 907)	53	-
NAFL vs NASH F4	1049 (vs 1369)	71	-
NASH F0-F1 vs NASH F2	7 (vs 0)	0	-
NASH F0-F1 vs NASH F3	482 (vs 434)	16	-
NASH F0-F1 vs NASH F4	1175 (vs 1194)	75	-

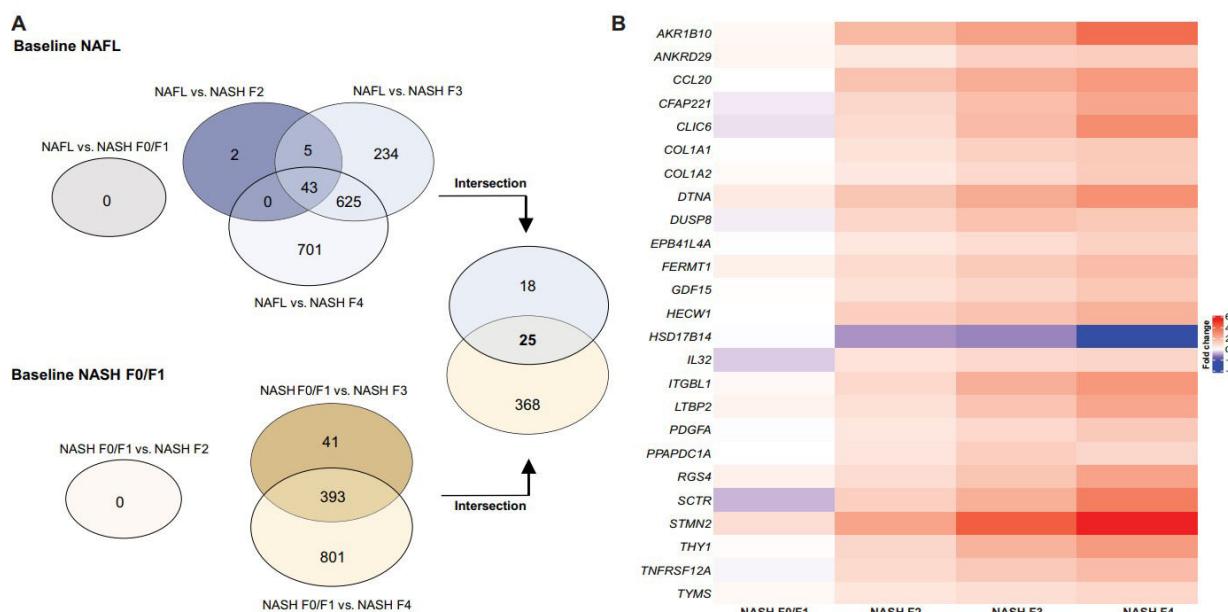
Παρατηρούμε πως υπάρχει μια σταθερή αύξηση των διαφορικά εκφραζόμενων γονιδίων καθώς τα στάδια της νόσου εξελίσσονται, με εξαίρεση όταν παίρνουμε επίπεδο αναφοράς τα control δείγματα, όπου υπάρχει μια διακύμανση, με κορύφωση στη σύγκριση Control vs NASH F2 και μετά πτώση (Πίνακες 3-5).

Υποκεφάλαιο 3.1.1: Ανάλυση διαφορικής έκφρασης για PCGs

Τα αποτελέσματα για τα PCGs συγκρίθηκαν τα αποτελέσματα της έρευνας από την οποία χρησιμοποιήσαμε τα αρχεία, ώστε δούμε πως υπάρχει συνοχή. Σε αντίθεση με την προσέγγισή μας με

Voomto πακέτο “DESeq2”, στο άρθρο αναφοράς η ανάλυση διαφορικής έκφρασης έγινε με το πακέτο “limma” (Limma-Voom). Δεν φαίνεται να υπάρχουν μεγάλες διαφορές στα στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια (Differentialy Expressed Genes, DEGs), με εξαίρεση όταν το επίπεδο αναφοράς ήταν τα δείγματα NASH F0-F1 όπως αναγράφεται στους πίνακες 3-5.

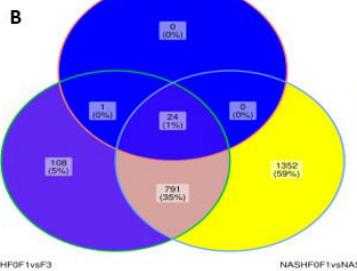
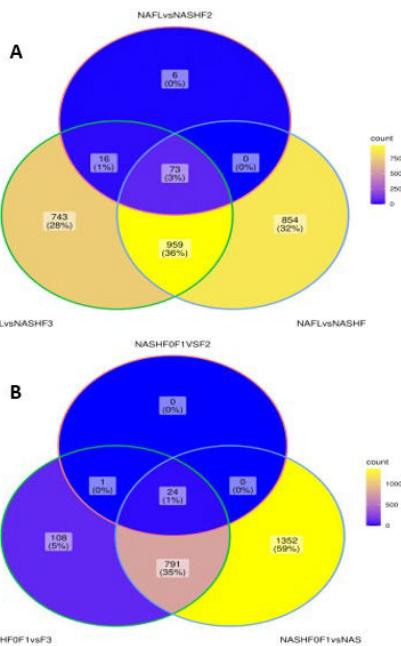
Αυτό όπως φαίνεται παρακάτω (Εικόνα 12) επηρεάζει τα διαγράμματα Venn συγκριτικά με αυτά του άρθρου αναφοράς. Δεν έγινε ανάλυση με επίπεδο αναφοράς τα control δείγματα για τα PCGs, ούτε των RBPs ή των lncRNAs στο άρθρο αναφοράς[64], συνεπώς δεν υπάρχει κάποια σύγκριση.



Εικόνα 11. Αποτελέσματα ανάλυσης διαφορικής έκφρασης (limma) του άρθρου αναφοράς[64]. (Α) Διάγραμμα Venn με τους αριθμούς των διαφορικά εκφραζόμενων γονιδίων με τα 2 επίπεδα αναφοράς (πάνω είναι επίπεδο αναφοράς τα δείγματα NAFL και κάτω τα δείγματα NASH F0-F1). (Β) Χάρτης θερμότητας με τα 25 γονίδια που είναι κοινώς διαφορικά εκφραζόμενα σε όλα τα στάδια της νόσου.

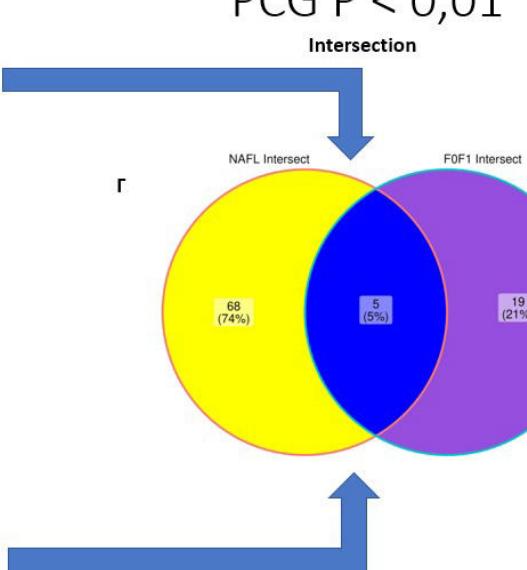
Στο άρθρο βρέθηκαν 25 PCGs που είναι κοινώς διαφορικά εκφραζόμενα σε όλα στάδια της νόσου (Εικόνα 11B). Η ανάλυση διαφορικής έκφρασης που κάναμε, εμφάνισε DEGs για NASH F0-F1 vs NASH F2 (Εικόνα 12), με 2/25 κοινά με εκείνα του άρθρου. Έγινε λοιπόν επιπλέον σύγκριση της ανάλυσης του άρθρου αναφοράς με τη δική μας χωρίς το NASH F0-F1 vs NASH F2 και βρήκαμε 44 κοινώς DEGs (Intersection), από αυτά, 19/25 ήταν κοινά με το άρθρο αναφοράς, για $p < 0,01$ (Εικόνα 13) ενώ 13/25 για $p < 0,005$ (Παράρτημα 2).

PCG P < 0,01



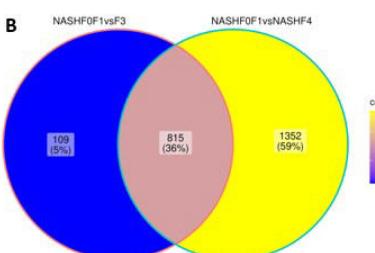
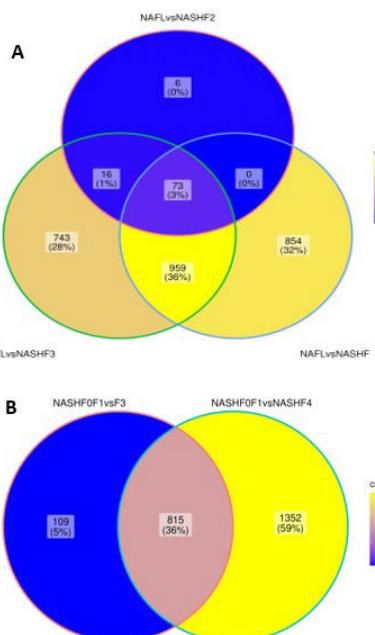
- **NASH F0-F1-baseline**
- 1) **CCL20**
 - 2) CDH24
 - 3) **CLIC6**
 - 4) CXCL1
 - 5) CXCL6
 - 6) CXCL8
 - 7) EEF1A2
 - 8) FAP
 - 9) GABRE
 - 10) **ITGBL1**
 - 11) KRT7
 - 12) L3MBTL3
 - 13) LAMA1
 - 14) MOXD1
 - 15) MVP
 - 16) PDE9A
 - 17) PDZK1IP1
 - 18) **SCTR**
 - 19) SLC35F6
 - 20) SOX9
 - 21) SPATA21
 - 22) SPINT1
 - 23) TACSTD2
 - 24) TMEM125

- **Intersection**
- 1) **CCL20**
 - 2) FAP
 - 3) GABRE
 - 4) **ITGBL1**
 - 5) SPATA21



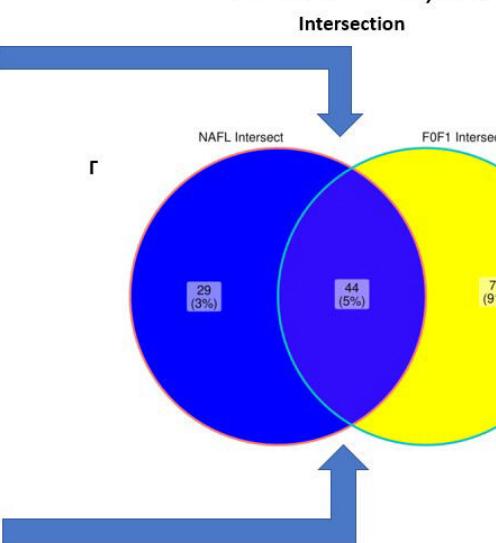
Εικόνα 12. Διάγραμμα Venn με τους αριθμούς των DEGs για τα PCGs. Τα 2 επίπεδα αναφοράς NAFL (Α), NASH F0-F1 (Β) και (Γ) τα κοινά DEGs μεταξύ των δύο. (Δ) Τα κοινά DEGs για το επίπεδο αναφοράς NASH F0-F1 (αριστερά) και τα κοινά μεταξύ των 2 επιπέδων αναφοράς (Intersection). Με κόκκινο φαίνονται τα κοινά DEGs μεταξύ της ανάλυσης μας και του άρθρου αναφοράς. Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.

PCG P < 0,01



- **Δ**
- 1) **AKR1B10**
 - 2) AMICA1
 - 3) ANXA2
 - 4) C6orf223
 - 5) CABYR
 - 7) **CCL20**
 - 8) CFAP221
 - 6) CLIC6
 - 9) COL1A2
 - 10) COL3A1
 - 11) COL5A1
 - 12) CYP2C19
 - 13) DCAF11
 - 14) DTNA
 - 15) EPB41L1

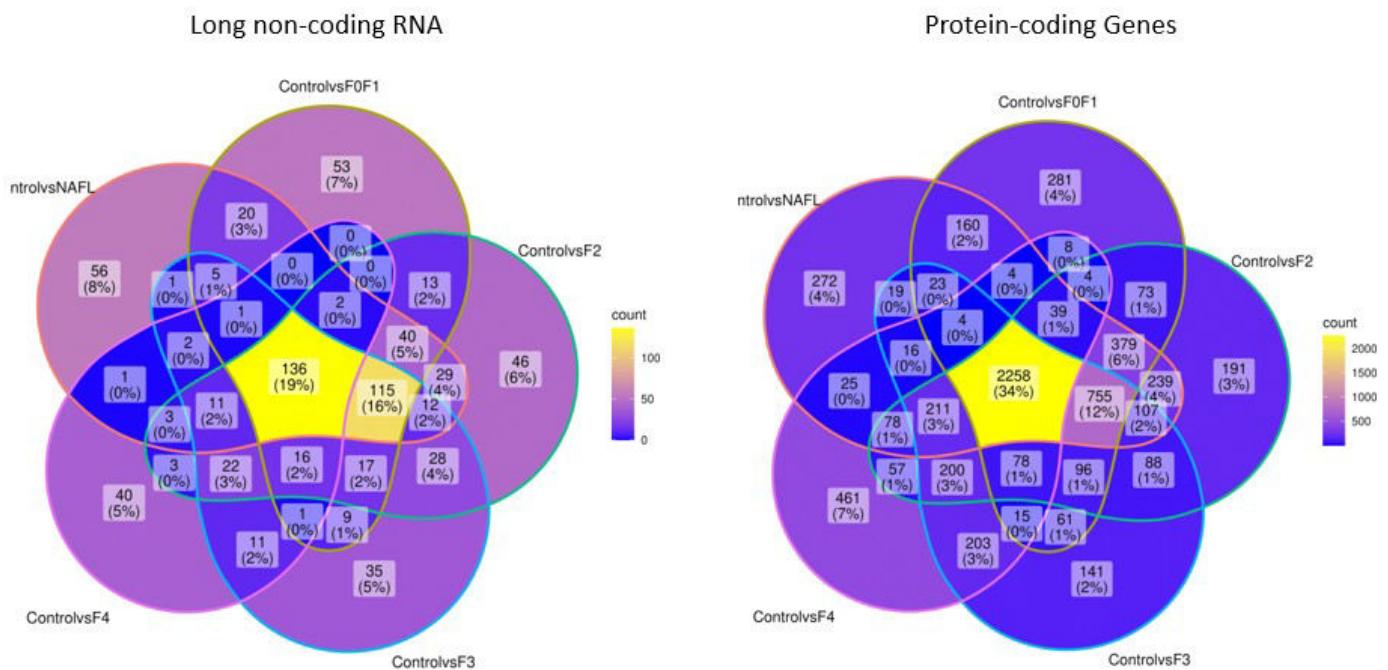
- 16) **FERMT1**
- 17) FAP
- 18) GABRE
- 19) **GDF15**
- 20) **HECW1**
- 21) ICAM3
- 22) **IL32**
- 23) **ITGBL1**
- 24) LOXL4
- 25) **LTPB2**
- 26) MTCL1
- 27) NQO1
- 28) PCK2
- 29) **PDGFA**
- 30) PLAU
- 31) PLP2
- 32) **PPAPDC1A**
- 33) **RGS4**
- 34) **SCTR**
- 35) SLC6A8
- 36) SLC7A6
- 37) SPATA21
- 38) **STMN2**
- 39) SULF2
- 40) THBS2
- 41) **THY1**
- 42) **TNFRSF12A**
- 43) TTL
- 44) WNT4



Εικόνα 13. Διάγραμμα Venn με τους αριθμούς των DEGs για τα PCGs. Τα 2 επίπεδα αναφοράς (Α) NAFL, (Β) NASH F0-F1 και (Γ) τα κοινά DEGs μεταξύ των δύο. Η σύγκριση NASH F0-F2 vs NASH F2 δεν λήφθηκε υπόψιν ώστε να γίνει παρόμοια σύγκριση με το άρθρο αναφοράς. (Δ) Αναγράφονται τα κοινά DEGs μεταξύ των 2 επιπέδων αναφοράς. Με κόκκινο φαίνονται τα κοινά DEGs της ανάλυσης μας και του άρθρου αναφοράς (19/25). Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.

Για να καταλάβουμε την εξέλιξη της νόσου και να απαντήσουμε το ερώτημα - πρόβλεψη της νόσου στα αρχικά στάδια - χρησιμοποιήσαμε ως επίπεδο αναφοράς τα control δείγματα και παρατηρήσαμε μια πληθώρα γονιδίων, τόσο PCGs και lncRNAs να είναι κοινώς διαφορικά εκφραζόμενα μεταξύ όλων των σταδίων της νόσου (Πίνακες 3-5). Για την ακρίβεια φαίνεται 2258 PCGs και 136 lncRNAs να είναι κοινώς διαφορικά εκφραζόμενα για $p < 0,005$ (Εικόνα 14). Για $p < 0,01$ & $p < 0,001$ ανατρέξτε το Παράρτημα 3.

Control-baseline $P < 0,005$



Εικόνα 14. Διαγράμματα Venn με τους αριθμούς των DEGs (αριστερά - lncRNAs, δεξιά - PCGs) για cutoff με $p-value < 0,005$. Επίπεδο αναφοράς είναι τα δείγματα control ($n=10$) και γίνεται σύγκριση κάθε φορά με ένα στάδιο της νόσου (Control vs NAFL, Control vs NASH F0-F1 κ.ο.κ). Με κίτρινο φαίνονται τα κοινώς διαφορικά εκφραζόμενα γονίδια σε όλα τα στάδια της νόσου. Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.

Υποκεφάλαιο 3.1.2: Ανάλυση διαφορικής έκφρασης για lncRNAs

Για την ανάλυση διαφορικής έκφρασης των lncRNAs δεν υπήρχε μέτρο σύγκρισης με το άρθρο αναφοράς. Η ανάλυση έγινε πάλι για 3 διαφορετικά cutoff ($p<0,01$, $p<0,005$, $p<0,001$). Τα αποτελέσματα μας έδειξαν για $p<0,01$, τρία κοινώς διαφορικά εκφραζόμενα lncRNAs μεταξύ των 2 επιπέδων αναφοράς: το GAS6-AS1, το MIR4435-1HG και το LINC00665 (Εικόνα 15). Αυτό δηλαδή σημαίνει πως έχουν στατιστικά σημαντική διαφορική έκφραση των τριών αυτών γονιδίων καθώς η νόσος εξελίσσεται σε μετέπειτα στάδια. Να σημειωθεί πως η ανάλυση διαφορικής έκφρασης για NASH F0-F1 vs NASH F2 δεν μας έδωσε στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια και δεν απεικονίζονται στα διαγράμματα Venn για τα lncRNAs.

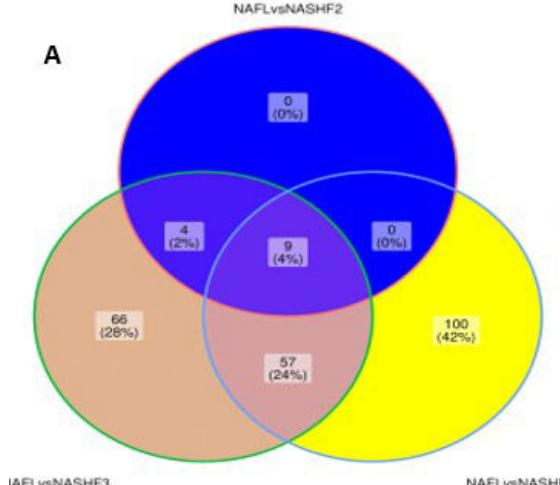
Ανάλυση με πιο αυστηρά cutoff, έδειξε πως το GAS6-AS1 είναι το μοναδικό στατιστικά σημαντικό διαφορικά εκφραζόμενο γονίδιο κοινό μεταξύ των δύο επιπέδων αναφοράς (Εικόνα 16, Παράρτημα 4). Το MIR4435-1HG φαίνεται να είναι στατιστικά σημαντικό διαφορικά εκφραζόμενο μόνο στις συγκρίσεις με επίπεδο αναφοράς NASH F0-F1 (Εικόνα 16, Παράστημα 4). Το LINC00665 δεν εμφανίστηκε ως στατιστικά σημαντικό με τα αυστηρότερα cutoffs και δεν επιλέχθηκε για μελέτη.

Υποκεφάλαιο 3.1.3: Γονίδια ενδιαφέροντος για lncRNAs

Για να δούμε σε περισσότερη λεπτομέρεια τις αλλαγές των γονιδίων GAS6-AS1 και MIR4435-1HG, ανακαλέσαμε το padj και την αλλαγή διπλώματος (Fold Change, FC) από τα αρχικά αποτελέσματα της ανάλυσης διαφορικής έκφρασης, δηλαδή χωρίς να έχουμε κάποιο φιλτράρισμα (Πίνακας 6, Πίνακας 7) με τη χρήση της εντολής “grep”. Να σημειωθεί πως η αλλαγή διπλώματος υπολογίζεται από το πακέτο DESeq2 ως $\log_2(FC)$, ενώ οι αριθμοί που αναγράφονται στους πίνακες είναι στρογγυλοποιημένα αποτελέσματα του $2^{\log_2(FC)}$.

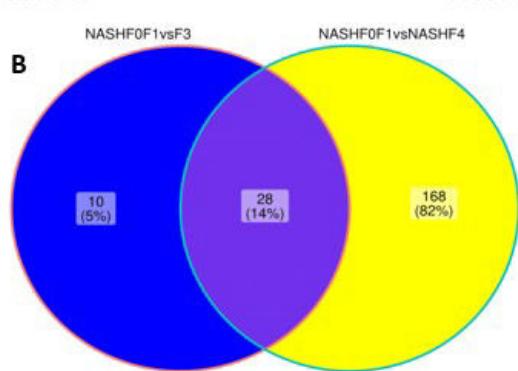
LNC RNA P < 0,01

A



count
100
75
50
25
0

B



count
160
120
80
40
0

Intersection

Γ

NAFL Intersect F0F1 Intersect



count
25
20
15
10
5
0

Intersection DEGs

Δ

- **NAFL-baseline**

- 1) **MIR4435-1HG**
- 2) FAM201A
- 3) **GAS6-AS1**
- 4) HORMAD2-AS1
- 5) LINC00665
- 6) LINC01123
- 7) RP11-118B22.4
- 8) RP11-119D9.1
- 9) RP1-140K8.5

- **NASH F0-F1-baseline**

- 1) B3GALT5-AS1
- 2) FAM222A-AS1
- 3) **GAS6-AS1**
- 4) GDNF-AS1
- 5) VIM-AS1
- 6) DLGAP1-AS3
- 7) LINC00152
- 8) LINC00511
- 9) LINC00665
- 10) LINC00702
- 11) LINC01279
- 12) EGOT
- 13) CTC-273B12.8
- 14) CTD-2008P7.9

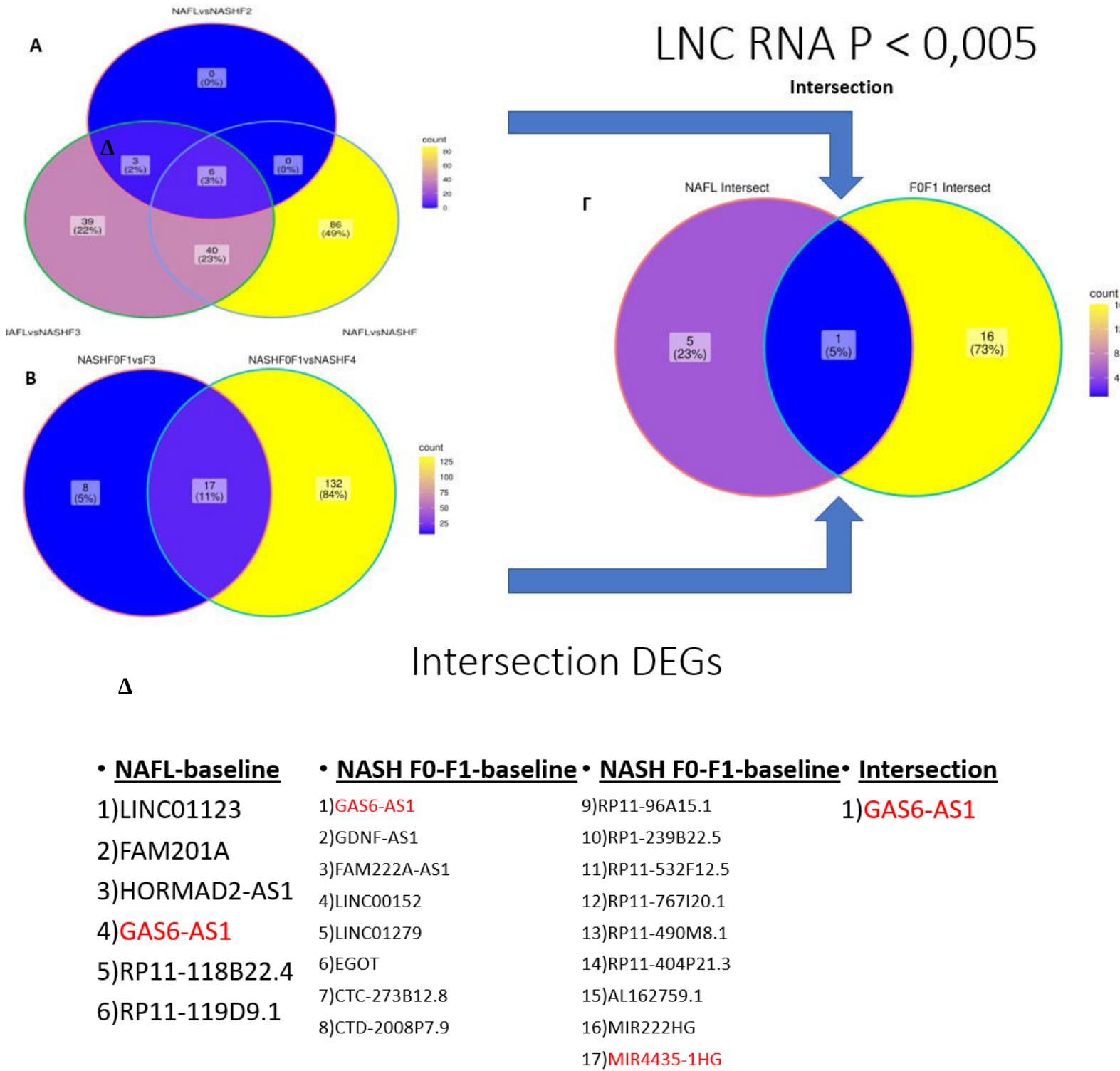
- **NASH F0-F1-baseline**

- 15) RP11-96A15.1
- 16) RP1-239B22.5
- 17) RP11-465B22.8
- 18) RP11-163F15.1
- 19) RP11-532F12.5
- 20) RP11-757F18.5
- 21) RP11-273G15.2
- 22) RP5-1033H22.2
- 23) RP11-767I20.1
- 24) RP11-490M8.1
- 25) RP11-404P21.3
- 26) AL162759.1
- 27) MIR222HG
- 28) **MIR4435-1HG**

- **Intersection**

- 1) **MIR4435-1HG**
- 2) LINC00665
- 3) **GAS6-AS1**

Εικόνα 15. Διάγραμμα Venn με τους αριθμούς των DEGs για τα lncRNAs σε cutoff p<0,01. Τα 2 επίπεδα αναφοράς NAFL (Α), NASH F0-F1 (Β) και (Γ) τα κοινά DEGs μεταξύ των δύο. (Δ) Αναγράφονται τα κοινά DEGs για το επίπεδο αναφοράς NAFL (αριστερά), για το επίπεδο αναφοράς NASH F0-F1 (δύο κεντρικές στήλες) και τα κοινά μεταξύ των 2 επιπέδων αναφοράς (Intersection). Με κόκκινο φαίνονται τα 2 υπογήφια lncRNAs με ενδιαφέρον για περιταίρω μελέτη. Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.



Εικόνα 16. Διάγραμμα Venn με τους αριθμούς των DEGs για τα lncRNAs σε cutoff p<0,005. Τα 2 επίπεδα αναφοράς NAFL (Α), NASH F0-F1 (Β) και (Γ) τα κοινά DEGs μεταξύ των δύο. (Δ) Αναγράφονται τα κοινά DEGs για το επίπεδο αναφοράς NAFL (αριστερά), για το επίπεδο αναφοράς NASH F0-F1 (δύο κεντρικές στήλες) και τα κοινά μεταξύ των 2 επιπέδων αναφοράς (Intersection). Με κόκκινο φαίνονται τα 2 υποψήφια lncRNAs με ενδιαφέρον για περεταίρω μελέτη. Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.

To GAS6-AS1 όπως φαίνεται και στον Πίνακα 6, Εικόνα 17 (εκτός από την σύγκριση NASH F0-F1 vs NASH F2 και NAFL vs NASH F0-F1 που σε καμία ανάλυση δεν μας έδωσε στατιστικά σημαντικά αποτελέσματα) έχει μια τάση να εκφράζεται σε μεγαλύτερο βαθμό (upregulated) όσο η νόσος μεταπίπτει σε επόμενα στάδια. Συγκεκριμένα όταν το επίπεδο αναφοράς είναι τα δείγματα Control παρατηρούμε μία πολύ μεγάλη αλλαγή διπλώματος (έως και 20 φορές περισσότερο στο τέταρτο στάδιο ίνωσης F4) και πολύ μικρό padj που καθιστά το γονίδιο ιδιαίτερα στατιστικά σημαντικό. Παρατηρούμε το ίδιο μοτίβο υπερέκφρασης (θετική αλλαγή διπλώματος) στα υπόλοιπα επίπεδα αναφοράς (NAFL, NASH F0-F1), με όλες τις τιμές padj να είναι ιδιαίτερα μικρές (άρα στατιστικά σημαντικές).

Πίνακας 6. Αποτελέσματα ανάλυσης διαφορικής έκφρασης (DESeq2), μόνο για το lncRNA γονίδιο GAS6-AS1. Η πρώτη στήλη αναφέρει την σύγκριση από την οποία εξάχθηκαν τα στοιχεία. Στη δεύτερη φαίνεται το padj που προκύπτει εφαρμόζοντας τη μέθοδο Benjamini και Hochberg στο αρχικό p-value. Στην τρίτη βλέπουμε την αλλαγή διπλώματος και στην τέταρτη αναγράφεται το γονίδιο και το Ensembl gene ID (ENSG). Το e-N (N = αριθμός στα δεξιά του e) αναδεικνύει ότι αριθμός στα αριστερά του ε πολλαπλασιάζεται με 10^e . Το πρόσημο λαμβάνεται υπόψιν και στη περίπτωση αυτή είναι αρνητικό (αν ήταν θετικό θα είχαμε 10^{+N} ή 10^N).

DESeq2 ανάλυση	padj (e-N = 10^e -N)	Αλλαγή διπλώματος (Fold Change, FC)	Γονίδιο GAS6-AS1 ENSG00000233695
Control vs NAFL	3.755284e-12	5,96	GAS6-AS1
Control vs NASH F0-F1	3.908204e-16	6,6	GAS6-AS1
Control vs NASH F2	1.376609e-20	11,58	GAS6-AS1
Control vs NASH F3	2.503909e-30	12,47	GAS6-AS1
Control vs NASH F4	1.041937e-36	19,15	GAS6-AS1
NAFL vs NASH F0-F1	-	-	GAS6-AS1
NAFL vs NASH F2	7.142160e-04	1,97	GAS6-AS1
NAFL vs NASH F3	3.647324e-07	2,1	GAS6-AS1
NAFL vs NASH F4	4.402547e-07	3,18	GAS6-AS1
NASH F0-F1 vs NASH F2	-	-	GAS6-AS1
NASH F0-F1 vs NASH F3	1.553668e-04	1,89	GAS6-AS1
NASH F0-F1 vs NASH F4	8.090035e-07	2,88	GAS6-AS1

Αντίστοιχα, το MIR4435-1HG (Πίνακας 7, Εικόνα 18-19) έχει μια παρόμοια τάση να αυξάνεται καθώς η νόσος εξελίσσεται. Δεν φαίνεται να είναι στατιστικά σημαντικά διαφορικά εκφρασμένο για τις πρώτες 3 συγκρίσεις όπου επίπεδο αναφοράς είναι τα δείγματα control. Παρατηρούμε επίσης ότι όσο πιο προχωρημένο επίπεδο αναφοράς χρησιμοποιούμε, η στατιστική του σημαντικότητά και η αλλαγή διπλώματος αυξάνεται περισσότερο με την εξέλιξη της νόσου συγκριτικά αυτής του GAS6-AS1.

Από τα αποτελέσματα (Πίνακας 7-8) μπορεί να γίνει μια αρχική υπόθεση πως το GAS6-AS1 ίσως παίζει αντισταθμιστικό ρόλο για τα βασικά συμπτώματα της νόσου και για τον λόγο αυτό παρατηρούμε τόσο μεγάλη αλλαγή διπλώματος από μετάβαση υγιούς ατόμου (control) σε ασθενές.

Αντίθετα, το MIR4435-1HG ίσως να έχει παρόμοια αντισταθμιστικό ρόλο, αλλά για τις πιο προχωρημένες ιστοπαθολογικές αλλαγές που παρατηρούνται στην εξέλιξη της νόσου.

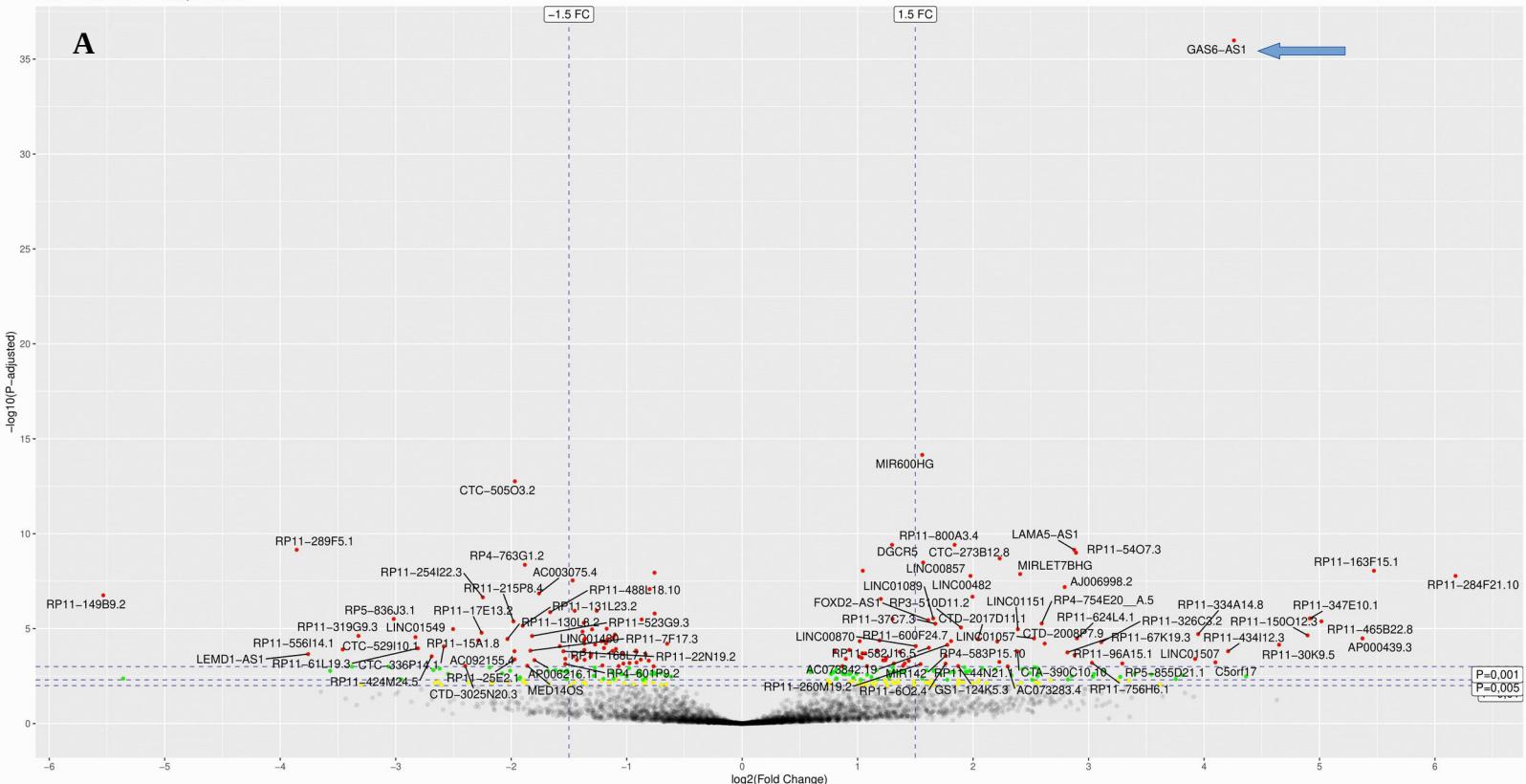
Πίνακας 7. Αποτελέσματα ανάλυσης διαφορικής έκφρασης (DESeq2), μόνο για το lncRNA γονίδιο MIR4435-1HG. Η πρώτη στήλη αναφέρει την σύγκριση από την οποία εξάχθηκαν τα στοιχεία. Στη δεύτερη φαίνεται το padj που προκύπτει εφαρμόζοντας τη μέθοδο Benjamini και Hochberg στο αρχικό p-value. Στην τρίτη βλέπουμε την αλλαγή διπλώματος και στην τέταρτη αναγράφεται το γονίδιο και το Ensembl gene ID (ENSG). Το e-N (N = αριθμός στα δεξιά του e) αναδεικνύει ότι αριθμός στα αριστερά του ε πολλαπλασιάζεται με 10^e -N. Το πρόσημο λαμβάνεται υπόψιν και στη περίπτωση αυτή είναι αρνητικό (αν ήταν θετικό θα είχαμε 10^{+N} ή 10^N).

DESeq2 ανάλυση	padj (e-N = 10^e -N)	Αλλαγή διπλώματος (Fold Change, FC)	Γονίδιο MIR4435-1HG ENSG00000172965
Control vs NAFL	4.568646e-01	0,8	MIR4435-1HG
Control vs NASH F0-F1	2.984088e-01	0,75	MIR4435-1HG
Control vs NASH F2	4.260882e-01	1,28	MIR4435-1HG
Control vs NASH F3	4.593759e-02	1,78	MIR4435-1HG
Control vs NASH F4	6.843181e-04	2,65	MIR4435-1HG
NAFL vs NASH F0-F1	-	-	MIR4435-1HG
NAFL vs NASH F2	7.279800e-03	1,62	MIR4435-1HG
NAFL vs NASH F3	2.021223e-08	2,24	MIR4435-1HG
NAFL vs NASH F4	3.255382e-10	3,32	MIR4435-1HG
NASH F0-F1 vs NASH F2	-	-	MIR4435-1HG
NASH F0-F1 vs NASH F3	9.970357e-08	2,37	MIR4435-1HG
NASH F0-F1 vs NASH F4	8.639334e-12	3,54	MIR4435-1HG

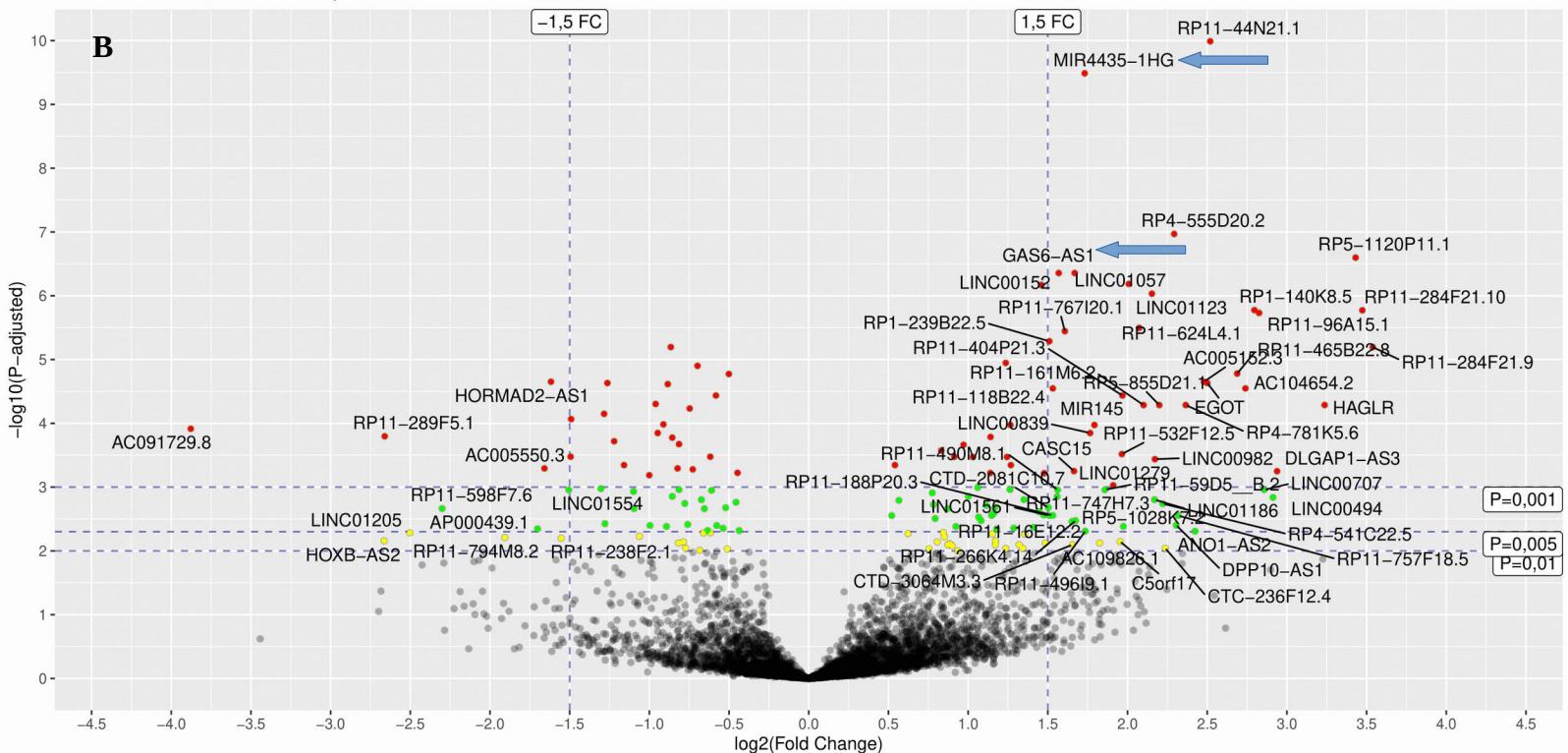
Υποκεφάλαιο 3.2: Ανάλυση διαφορικής έκφρασης σε γράφημα τύπου ηφαιστείου

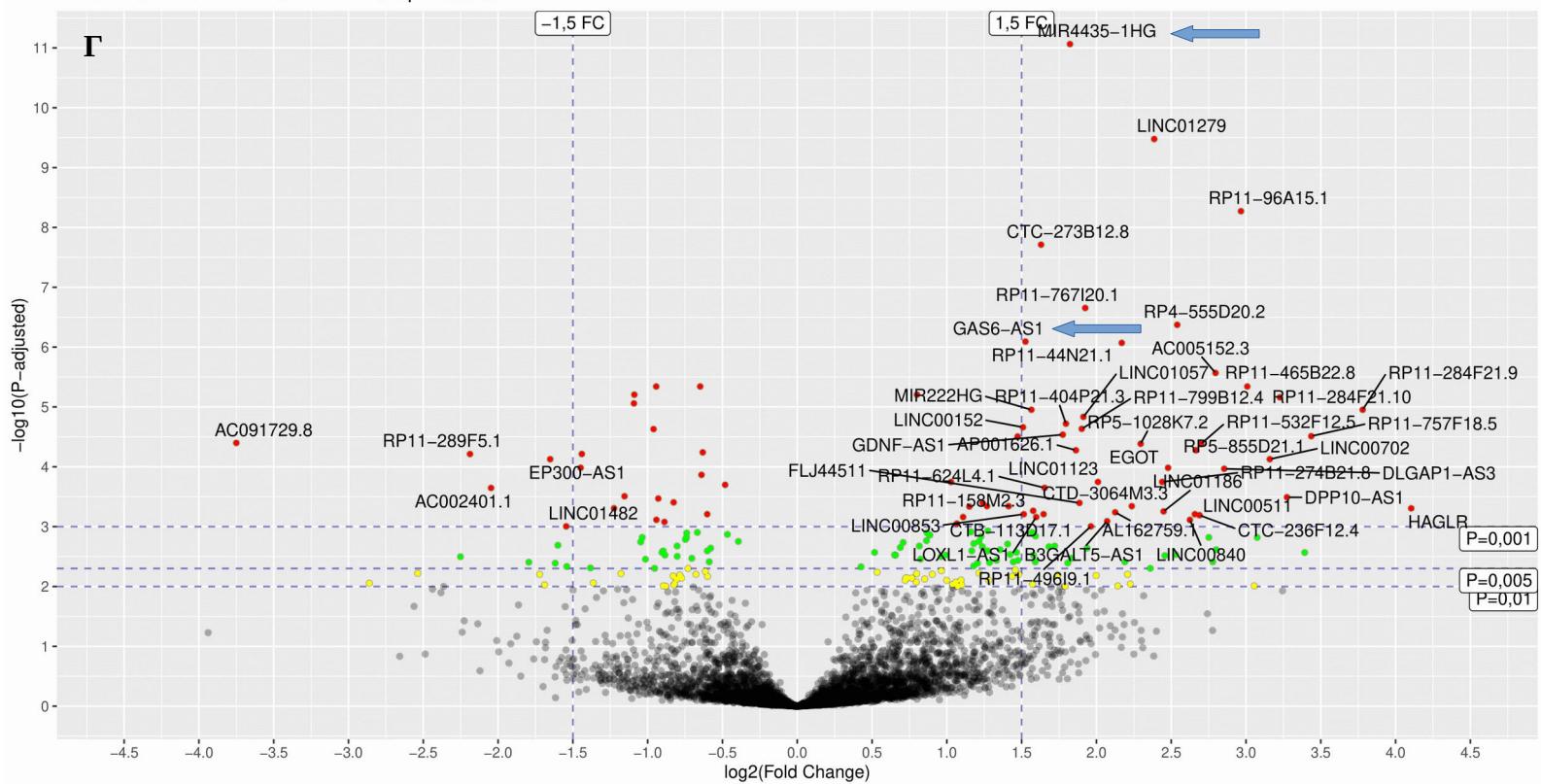
Περεταίρω απεικόνιση των αποτελεσμάτων της ανάλυσης διαφορικής έκφρασης έγινε με γράφημα τύπου ηφαιστείου (ggplot2 + ggrepel). Με σκοπό να φανούν τα κοινά διαφορικά εκφραζόμενα γονίδια, έγινε επιλογή των κοινών μεταξύ των 2 επιπέδων αναφοράς NAFL και NASH F0-F1, δηλαδή το Intersection (όπως αναφέρετε και προηγουμένως) και έγινε έμφασή τους στο γράφημα τύπου ηφαιστείου. Καθώς εστιάσαμε στα γονίδια των lncRNAs, στην Εικόνα 17 απεικονίζονται τα γραφήματα για την ανάλυση επιπέδων αναφοράς και το τελευταίο στάδιο της νόσου, το NASH F4. Επιπλέον γραφήματα για όλες τις αναλύσεις lncRNAs και PCGs μπορούν να βρεθούν στο Παράρτημα 5-9. Με μπλε βελάκι μπορούμε να δούμε τα αποτελέσματα του Πίνακα 6-7 για τις συγκρίσεις με το τελευταίο στάδιο της νόσου (NASH F4), των δύο γονιδίων ενδιαφέροντος GAS6-AS1 και MIR4435-1HG.

LNC ControlsF4 DESeq2 Results

A

LNC NAFLvsNASHF4 DESeq2 Results

B



Εικόνα 17. Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των lncRNAs μετά από την αφαίρεση των N/A τιμών. Έγινε χρήση του αρνητικού λογαρίθμου του padj (-log10(padj)) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος (log2(FC)) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με p-value < 0,01 (padj), με πράσινο τα γονίδια μεταξύ 0,01 < p-value < 0,005 και με κόκκινο τα γονίδια με p-value < 0,001. Αναγράφονται τα ονόματα (Gene ID) των γονιδίων με στατιστικά σημαντικό p-value και μεγάλη αλλαγή διπλώματος (FC). Με βελάκι τονίζουμε τα γονίδια ενδιαφέροντος GAS6-AS1 & MIR4435-1HG. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel. (A) Σύγκριση control δειγμάτων με NASH F4 (B) Σύγκριση NAFL δειγμάτων με NASH F4 (Γ) Σύγκριση NASH F0-F1 δειγμάτων με NASH F4.

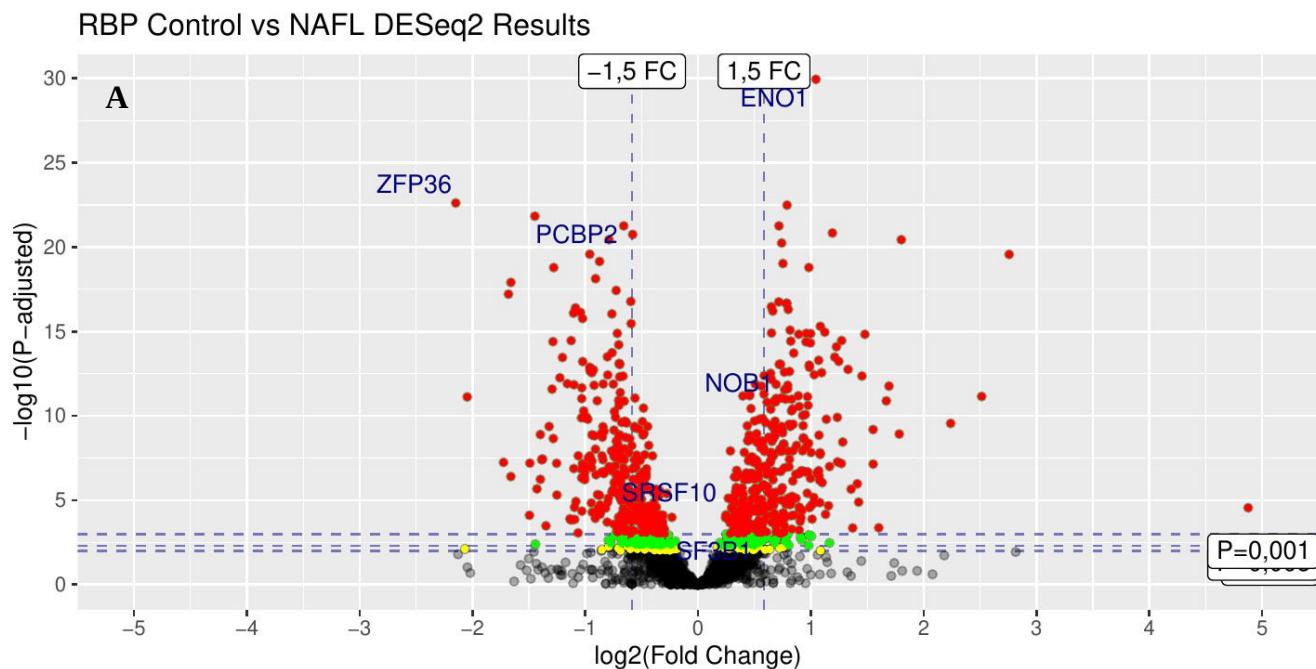
Υποκεφάλαιο 3.3: Αυτοματοποιημένη μηχανική μάθηση

Με τη βοήθεια της πλατφόρμας JADBio (<https://jadbio.com/>), μπορέσαμε μέσω αυτοματοποιημένης μηχανικής μάθησης (Automated Machine Learning, AML) να τονίσουμε την σημαντικότητα των γονιδίων GAS6-AS1 και MIR4435-1HG. Στην προσπάθεια να βρούμε υποψήφια γονίδια που θα μπορούν να χρησιμοποιηθούν ως βιοδείκτες για την διάγνωση και πρόβλεψη της ασθένεια, συγκρίναμε τα 2 επίπεδα αναφοράς (NAFL, NASH F0-F1) με το τελευταίο στάδιο της νόσου, την μη-αλκοολική στεατοηπατίτιδα με ίνωση τέταρτου σταδίου (NASH F4). Οι συγκρίσεις έγιναν για PCGs και lncRNAs. Για τα lncRNAs, παρατηρήσαμε για την ανάλυση NASH F0-F1 vs NASH F4 πως το γονίδιο MIR4435-1HG φάνηκε ως μοναδικό στοιχείο της AML, δηλαδή ως γονίδιο που προβλέπει καλύτερα την εξέλιξη της νόσου στο τελικό στάδιο. Στη συνέχεια αφαιρέσαμε το γονίδιο και εκτελέσαμε την ίδια ανάλυση, και βρήκαμε 26 στοιχεία, στα οποία βρισκόταν το GAS6-AS1. Για την ανάλυση NAFL vs NASH F4 και τα 2 γονίδια ενδιαφέροντος φάνηκαν ως στοιχεία της AML. Τα αποτελέσματα Control vs NAFL & Control vs NASH F0-F1 δεν εμφάνισαν τα γονίδια ενδιαφέροντος ως στοιχεία του μοτίβου

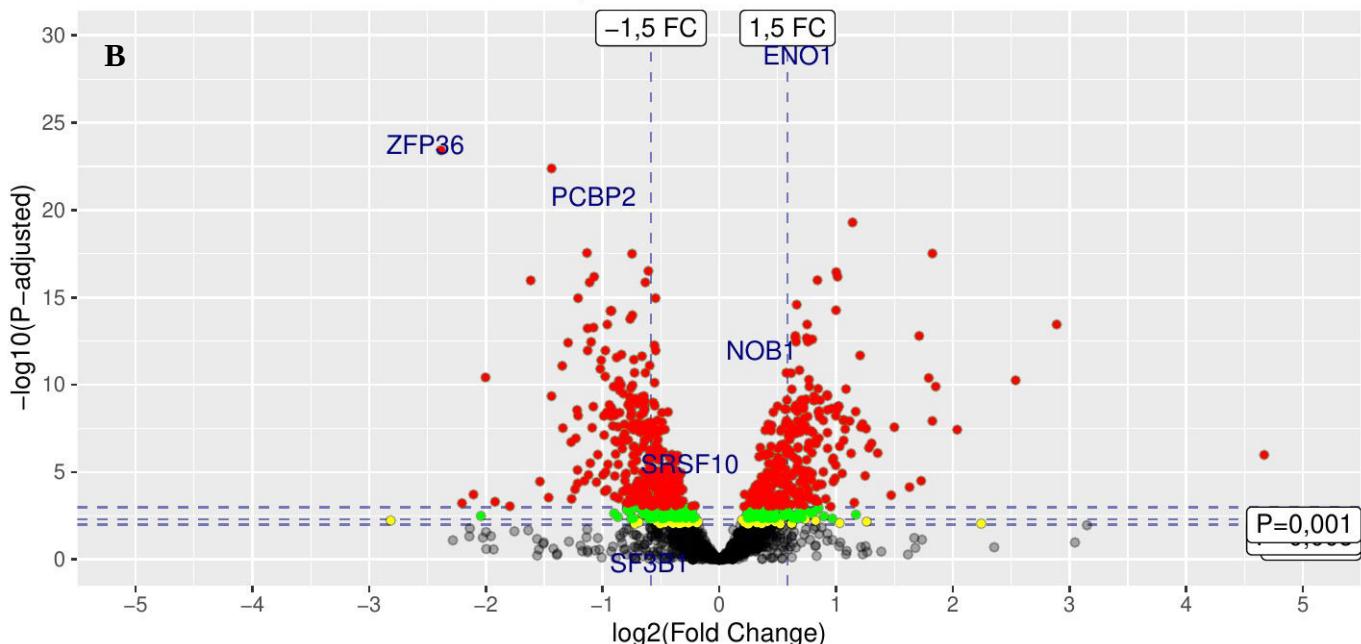
που πρόβλεψε. Οι παλινδρομήσεις βάση του σκορ NAS και βάση τα στάδια ίνωσης, δεν μας έδωσε κάποιο στατιστικά σημαντικό αποτέλεσμα (R-Squared: 0,009 & -0,004 αντίστοιχα).

Στη συνέχεια συνδυάσαμε τα PCGs και τα γονίδια των lncRNAs σε ένα dataframe και παρατηρήσαμε πως το μοτίβο που πρόβλεψε η AML ήταν 1(AUC Mean:0,958), με 5 στοιχεία (γονίδια), 4 PCGs (SEMA3G, GABRE, GSN και DTNA το οποίο ήταν στα 25 του άρθρου αναφοράς[64]) και 1 lncRNA, το MIR4435-1HG. Με την χρήση λίστας με ~2300 RBPs[65] και μετά από ανάλυση διαφορικής έκφρασης(Εικόνα 18), επαναλάβαμε τις αναλύσεις στην AML μόνο για τα RBPs, και βρήκαμε μεταξύ των μοτίβων των διαφορετικών αναλύσεων πιο συχνά να εμφανίζονται 6 RBPs: NSRP1, HMGN5, UBBP4, ZFP36, NOB1, PCBp2. Στη συνέχεια προσπαθήσαμε από τα αποτελέσματα της AML για τα PCGs και PCGs μαζί με lncRNAs, να αναγνωρίσουμε RBPs που εμφανίζονται ως στοιχεία του μοτίβου πρόβλεψης. Τα RBPs ENO1 (non-canonical RBP), ZFP36 (canonical RBP), NOB1 (canonical RBP) και PCBp2 (canonical RBP) βρέθηκαν ως στοιχεία ανάμεσα στα μοτίβα. Ένα επιπλέον RBP που χρησιμοποιήθηκε ως γονίδιο ενδιαφέροντος είναι το SRSF10 (Serine and Arginine Rich Splicing Factor 10, canonical RBP). Το SRSF10 ενώ κανονικά βρίσκεται σε μεγάλη έκφραση κοντά σε μοτίβα πολυαδενιλίωσης, σε άτομα με NAFLD η έκφρασή του μειώνεται[66].

Περισσότερα στοιχεία για τη διαδικασία και τα αποτελέσματα της AML υπάρχουν στο Παράρτημα 10.



RBP Control vs NASH F0–F1 DESeq2 Results



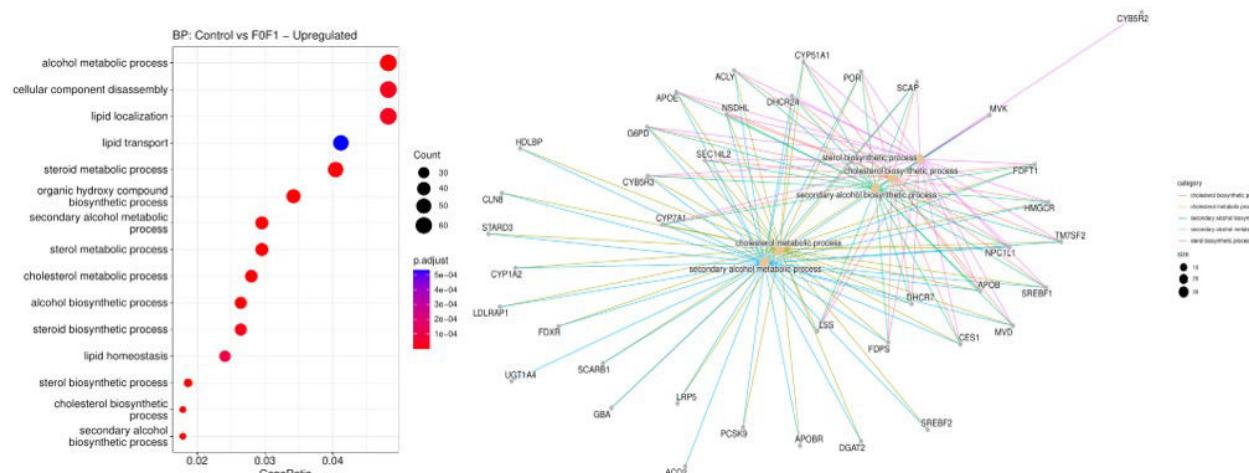
Εικόνα 18. Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των RBPs[65] για (A) Control vs NAFL και (B) Control vs NASH F0-F1. Έγινε χρήση του αρνητικού λογαρίθμου του παρδί ($-\log_{10}(\text{padj})$) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος ($\log_2(\text{FC})$) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με $p\text{-value} < 0,01$ (padj), με πράσινο τα γονίδια μεταξύ $0,01 < p\text{-value} < 0,005$ και με κόκκινο τα γονίδια με $p\text{-value} < 0,001$. Η ταμπέλα “-1,5 FC” και “1,5 FC” εδώ δεν πρόκειται για $\log_2(\text{FC})$ σε αντίθεση με τα άλλα γραφήματα. Αναγράφονται με μπλε τα 5 ονόματα (Gene ID) των RBP ενδιαφέροντος. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.

Υποκεφάλαιο 3.4: Αποτελέσματα ανάλυσης GO & KEGG

Τα αποτελέσματα της ανάλυσης διαφορικής έκφρασης γονιδίων για τα PCGs με $\text{padj} < 0,01$ & $\text{FC} \geq 1,5$ (Πίνακας 3) αναλύθηκαν με GO και KEGG μέσω του πακέτου “DOSE”. Οι αναλύσεις έγιναν με σκοπό να κατανοήσουμε πως εμπλέκονται τα γονίδια που είναι διαφορικά εκφρασμένα στην έναρξη της νόσου και στην εξέλιξη στο τελικό στάδιο (βιολογικές διαδικασίες – BP, κυτταρικό διαμέρισμα – CC, μοριακές λειτουργίες – MF). Συνεπώς η ανάλυση οντολογίας γονιδίων GO (Παράρτημα 11) & KEGG (Παράστημα 12) έγινε στα επίπεδα Control vs NAFL, Control vs NASH F0-F1, NAFL vs NASH F4, NASH F0-F1 vs NASH F4. Όπως είναι αναμενόμενο, η μετάβαση ενός υγιούς ατόμου στη νόσο συνοδεύεται από πολλαπλές αλλαγές στην έκφραση γονιδίων. Όσον αφορά τις βιολογικές διαδικασίες (Εικόνα 19, Παράρτημα 11; 1-4A), παρατηρούμε γονίδια να σχετίζονται με αποσυναρμολόγηση κυτταρικού διαμερίσματος, πρωτογενή και δευτερογενή μεταβολισμό αλκοόλης, μεταβολισμό χοληστερόλης και φυσικά μεταβολισμό λιπιδίων, τον κύριο λόγο πρόκλησης της νόσου. Όσον αφορά τις μοριακές λειτουργίες (MF), παρατηρούμε γονίδια που εμπλέκονται στην πρόσδεση καδερίνης, ένα μόριο που συμμετέχει στην προσκόλληση κυττάρων (cell adhesion) και έχει βρεθεί ως βιοδείκτης της

νόσου κυρίως σε μικρής ηλικίας άτομα[73]. Προϊόντα των διαφορικά εκφραζόμενων γονιδίων (κυτταρικό διαμέρισμα, CC) παρατηρούνται σε σταγονίδια λιπιδίων (συσσώρευση) και στο μονοπάτι TORC1, το οποίο έχει βρεθεί όταν υπερεκφράζεται συνεισφέρει στη συσσώρευση λιπιδίων μέσω φρουκτόζης[73]. Η KEGG ανάλυση για το Control vs NAFL αναδεικνύει γονίδια να εμπλέκονται με την έκκριση χολής η οποία σε άτομα με NAFLD εκκρίνεται σε μεγαλύτερες ποσότητες από το ήπαρ και δεν μετατρέπεται σε δευτερεύον προϊόντα (όπως δεοξυχολικό οξύ, DCA)[74]. Με μέτρο αναφοράς ένα ήδη ασθενή άτομο (NAFL vs NASH F4), πέρα από τις προαναφερθέντες λειτουργίες παρατηρούμε γονίδια επιπλέον να συμμετέχουν στην μεταφορά και localization λιπιδίων (Παράρτημα 11-3A).

GO BP

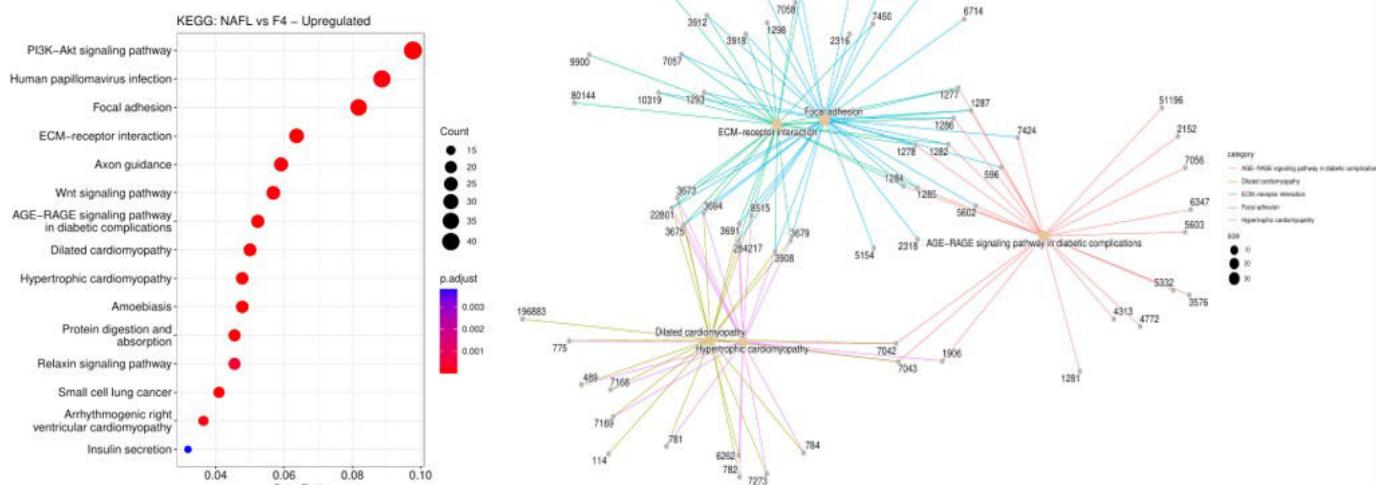


Εικόνα 19. Αποτελέσματα ανάλυσης οντολογίας γονιδίων για τις βιολογικές διαδικασίες (GO:BP) της Control vs NASH F0-F1 διαφορικής ανάλυσης. Έγινε χρήση στατιστικά σημαντικών διαφορικά εκφρασμένων γονιδίων ($p<0,01$) με αλλαγή διπλώματος (Fold Change) $FC \geq 1,5$ (upregulated). Αριστερά βρίσκεται το διάγραμμα κουκκίδας και δεξιά το cnetplot της αντίστοιχης ανάλυσης όπου φαίνονται το ονόματα των γονιδίων που εμπλέκονται στις βιολογικές διαδικασίες. Για την ανάλυση έγινε χρήση $p<0,05$ στις παραμέτρους του πακέτου (DOSE). Το padj στα γραφήματα κουκκίδων αναφέρεται στη στατιστική σημαντικότητα που προβλέπει η ανάλυση. Στα cnetplot απεικονίζονται τα 5 πιο στατιστικά σημαντικά αποτελέσματα. Counts = Αριθμός γονιδίων που εμπλέκονται στις βιολογικές διαδικασίες. Οι εικόνες για τα διαγράμματα κουκκίδων & cnetplot (enrichplot) παράχθηκαν από το πακέτο “clusterProfiler”.

Ιδιαίτερο ενδιαφέρον βέβαια έχουν τα αποτελέσματα της KEGG ανάλυσης για NAFL vs F4 (Εικόνα 20, Παράρτημα 12) και ο λόγος είναι διότι ενώ στις αναλύσεις με χρήση των διαφορικά εκφραζόμενων γονιδίων για Control vs NAFL και Control vs NASH F0-F1 κυρίως παρατηρείται τα γονίδια να εμπλέκονται στην έκκριση χολής δεν ισχύει το ίδιο όταν συγκρίνουμε ένα άτομο που νοσεί με τα

επόμενα στάδια. Πέρα από έκκριση ινσουλίνης, παρατηρούμε γονίδια να εμπλέκονται σε μία σειρά από μονοπάτια, όπως Wnt/β-κατενίνη, AGE-RAGE και PI3K-Akt.

KEGG



Εικόνα 20. Αποτελέσματα ανάλυσης KEGG για NAFL vs NASH F4. Έγινε χρήση στατιστικά σημαντικών διαφορικά εκφρασμένων γονιδίων ($p<0,01$) με αλλαγή διπλώματος (Fold Change) $FC \geq 1,5$ (upregulated). Αριστερά βρίσκεται το διάγραμμα κουκκίδας και δεξιά το cnetplot της αντίστοιχης ανάλυσης. Οι αριθμοί που αναγράφονται στο cnetplot (δεξιά) είναι τα ονόματα γονιδίων σε ENTREZID (ειδικό βάσης δεδομένων). Στα cnetplot απεικονίζονται τα 5 πιο στατιστικά σημαντικά αποτελέσματα. Counts = Αριθμός γονιδίων που εμπλέκονται. Οι εικόνες για τα διαγράμματα κουκκίδων & cnetplot (enrichplot) παράχθηκαν από το πακέτο “clusterProfiler”.

Το σηματοδοτικό μονοπάτι Wnt/β-κατενίνης έχει σημαντικό ρόλο στην παθοφυσιολογία του ήπατος και ιδιαίτερα με την ομοιόσταση της χολής, ενώ έχει συσχετιστεί με το ηπατοκυτταρικό καρκίνωμα (HCC), όπου υπάρχει αυξημένη ενεργοποίηση του μονοπατιού[75]. Επιπλέον, το AGE-RAGE (Advanced Glycation End products-AGE Receptor) μονοπάτι φαίνεται να συνεισφέρει στη συσσώρευση λιπιδίων και άρα την εμφάνιση ιστοπαθολογικών αλλαγών της νόσου (ίνωση, φλεγμονή κλπ)[76]. Τέλος, το PI3K-Akt σηματοδοτικό μονοπάτι υπό φυσιολογικές συνθήκες προωθεί τον μεταβολισμό, κυτταρικό κύκλο και αγγειογένεση ως ανταπόκριση σε εξωκυττάριες μεταβολές. Απορρύθμιση του μονοπατιού έχει επίσης συσχετιστεί με το ηπατοκυτταρικό καρκίνωμα, δηλαδή μια πιθανή μετάβαση ενός ασθενούς ατόμου από μη-αλκοολική στεατοηπατίτιδα με ίνωση τέταρτου σταδίου[77].

ΣΥΖΗΤΗΣΗ

Το ενδιαφέρον προς μελέτη και κατανόηση του ρόλου των lncRNAs σε συνδυασμό με την εξέλιξη μεθόδων προσέγγισης, όπως τα πειράματα αλληλούχισης συνολικού RNA (RNA-seq) έχουν οδηγήσει σε πολλές μελέτες να στρέφονται γύρω από αυτά. Ιδιαίτερο ενδιαφέρον έχει εμφανίσει και ο ρόλος των lncRNAs στην μη-αλκοολική λιπώδη νόσο[67,68,69,70]. Τα αποτελέσματα για την ανάλυση διαφορικής έκφρασης εστιασμένη στα lncRNAs δείχνουν αρκετά γονίδια με αλλαγμένη έκφραση, αυξημένη ή μειωμένη. Ανάμεσα στα γονίδια αυτά, το GAS6-AS1 και το MIR4435-1HG φαίνεται να έχουν ένα σταθερό μοτίβο αυξημένης έκφρασης με την εξέλιξη των σταδίων της νόσου. Η αυτοματοποιημένη μηχανική μάθηση ενίσχυσε τα ευρήματα των δύο γονιδίων ενδιαφέροντος με διαφορετικούς τρόπους προσέγγισης. Περεταίρω μελέτη με μηχανική μάθηση (όχι αυτοματοποιημένη) όπως με γενικευμένο γραμμικό μοντέλο ή τυχαία δάση (Random Forest) και ίσως δώσει καλύτερα μοτίβα πρόβλεψης τόσο στοιχείων (γονιδίων) που θα χαρακτηρίζουν προδιάθεση προς αυτήν όσο και στοιχείων που θα επιτρέψουν πιο εύκολο χαρακτηρισμό σταδίων (με-επεμβατικές μέθοδοι). Ενδιαφέρον έχει η μελέτη των δύο γονιδίων πειραματικά (σε κυτταρικό τύπο), με στοχευμένη αλλαγή της έκφρασης τους (υπερέκφραση ή σίγηση) και παρατήρηση των αποτελεσμάτων.

Επιπλέον, πολλές μελέτες έχουν στραφεί στην παρατήρηση RBPs συνδυαστικά με τα lncRNAs, δεδομένου ότι τα RBPs προσδένονται σε lncRNAs για να επιτελέσουν τις λειτουργίες τους[61,62,63, 65,66,71]. Η ανάλυση διαφορικής έκφρασης και η αυτοματοποιημένη μηχανική μάθηση παρουσίασαν 4 γονίδια κανονικών (canonical) RBPs με στατιστικά σημαντικό ενδιαφέρον (ZFP36, NOB1, PCBP2, SRSF10) και 1 μη-κανονικό (non-canonical) RBP. Για την προσέγγιση των αποτελεσμάτων, μπορεί να γίνει χρήση eCLIP δεδομένων (enhanced crosslinking and immunoprecipitation) όπου παρέχουν πληροφορίες σχετικά με το σημείο πρόσδεσης ενός RBP πάνω στο μετάγραφο. Τα δεδομένα αυτά υπάρχουν και για HepG2 κύτταρα, (δημόσια διαθέσιμα) αν και ο χαρακτηρισμός των RBPs είναι ακόμα ένα μη-ολοκληρωμένο έργο. Χρήση -των ως τώρα- χαρακτηρισμένων RBPs για HepG2 κύτταρα (~310) θα εμφανίσει σημεία πρόσδεσης πάνω στα μετάγραφα διαφορικά εκφραζόμενων γονιδίων υπογραμμίζοντας πιθανές αλληλεπιδράσεις με lncRNAs ή/και mRNAs και στην αρχή αποκρυπτογράφησης των δικτύων που ρυθμίζουν τη μεταγραφή στα ηπατικά κύτταρα.

Όπως αναφέρθηκε, τα lncRNAs επεξεργάζονται με μάτισμα (splicing) και τα μοτίβα τους μπορούν να μελετηθούν για να εξάγουν επιπρόσθετες πληροφορίες (από τα δεδομένα που εφαρμόστηκαν στις παραπάνω αναλύσεις) για την αλλαγή του ματίσματος στην πρόοδο της νόσου. Οι πληροφορίες μπορούν στην συνέχεια να συσχετιστούν με RBPs που προσδένονται σε παράγοντες ματίσματος (splicing factors) μέσω eCLIP δεδομένων και να ενισχύσουν ακόμα περισσότερο τα δίκτυα που ρυθμίζουν τη μεταγραφή. Στη περίπτωση όπου τα δίκτυα εμφανίσουν τα lncRNAs να προηγούνται της μεταγραφής των mRNAs, πιθανώς να αποτελέσουν βιοδείκτες και θεραπευτικούς στόχους. Μονοπάτια όπως το wnt/β-κατενίνη[75], AGE-RAGE[76], PI3K-Akt[77], ίσως διευκολύνουν την αναζήτηση γονιδίων στόχων, ανατρέχοντας βιβλιογραφίες για τον προσδιορισμό προϊόντων που συμμετέχουν σε αυτές και στη συνέχεια ανάλυσή τους (χτίσιμο δικτύου).

Παρόλο που η νόσος του μη-αλκοολικού λιπώδης ήπατος (NAFLD) χαρακτηρίζεται μέχρι το τέταρτο στάδιο ίνωσης της μη αλκοολικής στεατοηπατίτιδας (NASH F4), μπορεί να εξελιχθεί σε ηπατοκυτταρικό καρκίνωμα (HCC). Αυτό θέτει περισσότερα ερωτήματα για το πως μεταπίπτει και τους παράγοντες/γονίδια που το επάγουν. Συνεπώς μπορεί να χτιστεί ένα δίκτυο (π.χ με μηχανική μάθηση) από OMICs δεδομένα του HCC με σκοπό να κατανοήσουμε τις αλλαγές που λαμβάνουν χώρα.

БІБЛІОГРАФІА

- 1)Gillian Pocock, Christopher D. Richards, and David A. Richards (2013). The Gastrointestinal System. Human Physiology 5th edition.
- 2)Katrien Vekemans and Filip Braet (2005). Structural and functional aspects of the liver and liver sinusoidal cells in relation to colon carcinoma metastasis. *World J Gastroenterol* Sep 7; 11(33): 5095–5102.
- 3)Zbigniew Kmiec (2001). Cooperation of liver cells in health and disease. *Advances in Anatomy, Embryology and Cell Biology (ADVSANAT)*, volume 161).
- 4)Rolf Gebhardt and Madlen Matz-Soja (2014). Liver zonation: Novel aspects of its regulation and its impact on homeostasis. *World J Gastroenterol*. Jul 14; 20(26): 8491-8504.
- 5)Kurt Jungermann and Nick Katz (1989). Functional specialization of different hepatocyte populations. *Physiol Rev*. Jul;69(3):708-64.
- 6)Michiie Sakamoto (2009). Early HCC: diagnosis and molecular markers. *Journal of Gastroenterology* volume 44, pages108–111.
- 7)Andrea Baiocchini, Franca Del Nonno, Chiara Taibi, Ubaldo Visco-Comandini, Gianpiero D'Offizi, Mauro Piacentini and Laura Falasca (2019). Liver sinusoidal endothelial cells (LSECs) modifications in patients with chronic hepatitis C. *Scientific Reports* volume 9, Article number: 8760.
- 8)Daniel A Patten, Emma L Shepherd, Christopher J Weston and Shishir Shetty. (2019). Novel Targets in the Immune Microenvironment of the Hepatic Sinusoids for Treating LiverDiseases. *Semin. Liver Dis.*39, 111–123.
- 9)Ramón Bataller and David A. Brenner (2005). Liver fibrosis. *J Clin Invest*. Feb 1; 115(2):209–218.
- 10) European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO) (2016). EASL-EASD-EASO Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease. *Obes Facts*: 9:65-90
- 11)Rocío Aller, Conrado Fernández-Rodríguez, Oreste Lo Iacono, Rafael Bañares, Javier Abad, José Antonio Carrión, Carmelo García-Monzón, Joan Caballería, Marina Berenguer, Manuel Rodríguez-Péralvarez, José López Miranda, Eduardo Vilar-Gómez, Javier Crespo, Miren García-Cortés, María Reig, José María Navarro, Rocío Gallego, Joan Genescà, María Teresa Arias-Loste, María Jesús Pareja, Agustín Albillos, Jordi Muntané, Francisco Jorquera, Elsa Solà, Manuel Hernández-Guerra, Miguel Ángel Rojo, Javier Salmerón, Llorenç Caballería, Moisés Diago, Esther Molina, Ramón Bataller and Manuel Romero-Gómez (2018). Management of non-alcoholic fatty liver disease (NAFLD). Clinical practice guideline. *Gastroenterol Hepatol*. May;41(5):328–349.
- 12)Min-Sun Kwak and Donghee Kim (2018). Non-alcoholic fatty liver disease and lifestyle modifications, focusing on physical activity. *Korean J Intern Med*. Jan;33(1):64–74

- 13)Giorgio Bedogni, Lucia Miglioli, Flora Masutti, Claudio Tiribelli, Giulio Marchesini and Stefano Bellentani (2005). Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* Jyl; 42(1):44-52.
- 14)Mariana Machado, Pedro Marques-Vidal and Helena Cortez-Pinto (2006). Hepatic histology in obese patients undergoing bariatric surgery. *J Hepatol* Oct; 45(4):600-6.
- 15)Michelle T Long, Alison Pedley, Joseph M Massaro, Udo Hoffmann, Jiantao Ma, Rohit Loomba, Raymond T Chung and Emelia J Benjamin (2017). A simple clinical model predicts incident hepatic steatosis in a community-based cohort: The Framingham Heart Study. *Liver Int.* 2018 Aug; 38(8):1495-1503.
- 16)Zhongli Wang, Ming Xu, Zhengguo Hu, Michael Hultström and Enyin Lai (2014). Sex-specific prevalence of fatty liver disease and associated metabolic factors in Wuhan, south central China. *Eur J Gastroenterol Hepatol.* Sep; 26(9):1015-21.
- 17)Giulio Marchesini, Mara Brizi, Giampaolo Bianchi, Sara Tomassetti, Elisabetta Bugianesi, Marco Lenzi, Arthur J. McCullough, Stefania Natale, Gabriele Forlani and Nazario Melchionda (2001). Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. *Diabetes.* Aug;50(8):1844-50.
- 18)Christiane Stern, Laurent Castera (2016). Non-invasive diagnosis of hepatic steatosis. *Hepatol Int.* Jan; 11(1):70-78.
- 19)Stephen Caldwell, Yoshihiro Ikura, Daniela Dias, Kosuke Isomoto, Akito Yabu, Christopher Moskaluk, Patcharin Pramoonjago, Winsor Simmons, Harriet Scruggs, Nicholas Rosenbaum, Timothy Wilkinson, Patsy Toms, Curtis K Argo, Abdullah MS Al-Osaimi, and Jan A Redick (2010). Hepatocellular Ballooning in NASH. *J Hepatol.* Oct;53(4):719-23.
- 20)Robert F. Schwabe, Ira Tabas and Utpal B. Pajvani (2020). Mechanisms of Fibrosis Development in NASH. *Gastroenterology.* May; 158(7): 1913–1928.
- 21)Leen J M Heyens, Dana Busschots, Ger H Koek, Geert Robaeys, Sven Francque (2021). Liver Fibrosis in Non-alcoholic Fatty Liver Disease: From Liver Biopsy to Non-invasive Biomarkers in Diagnosis and Treatment. *Front Med (Lausanne).* Apr 14; 8:615978.
- 22)Jen-Kuei Peng, Nilay Hepgul, Irene J Higginson and Wei Gao (2018). Symptom prevalence and quality of life of patients with end-stage liver disease: A systematic review and meta-analysis. *Palliat Med.* Jan; 33(1):24-36.
- 23)Kazuo Tarao,corresponding author Akito Nozaki,Takaaki Ikeda,Akira Sato, Hirokazu Komatsu, Tatsujii Komatsu, Masataka Taguri, and Katsuaki Tanaka (2019). Real impact of liver cirrhosis on the development of hepatocellular carcinoma in various liver diseases—meta-analytic assessment. *Cancer Med.* Mar; 8(3): 1054–1065.
- 24)Prashanth Rawla, Tagore Sunkara, Pradhyumna Muralidharan, Jeffrey Pradeep Raj (2018). Update in global trends and aetiology of hepatocellular carcinoma. *Contemp Oncol (Pozn).* ;22(3):141-150.

- 25)David E Kleiner, Elizabeth M Brunt, Mark Van Natta, Cynthia Behling, Melissa J Contos, Oscar W Cummings, Linda D Ferrell, Yao-Chang Liu, Michael S Torbenson, Aynur Unalp-Arida, Matthew Yeh, Arthur J McCullough, Arun J Sanyal, Nonalcoholic Steatohepatitis Clinical Research Network (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. Jun; 41(6):1313-21.
- 26)Zobair M. Younossi, Rohit Loomba, Quentin M. Anstee, Mary E. Rinella, Elisabetta Bugianesi, Giulio Marchesini, Brent A. Neuschwander-Tetri, Lawrence Serfaty, Francesco Negro, Stephen H. Caldwell, Vlad Ratziu, Kathleen E. Corey, Scott L. Friedman, Manal F. Abdelmalek, Stephen A. Harrison, Arun J. Sanyal, Joel E. Lavine, Philippe Mathurin, Michael R. Charlton, Zachary D. Goodman, Naga P. Chalasani, Kris V. Kowdley, Jacob George, and Keith Lindor (2019). Diagnostic Modalities for Nonalcoholic Fatty Liver Disease, Nonalcoholic Steatohepatitis, and Associated Fibrosis. *Hepatology*. Jul; 68(1):349-360.
- 27)Yoshihisa Takahashi and Toshio Fukusato (2014). Histopathology of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol*. Nov 14; 20(42): 15539–15548.
- 28)Elizabeth M Brunt, David E Kleiner, Laura A Wilson, Patricia Belt, Brent A Neuschwander-Tetri, NASH Clinical Research Network (CRN). Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology*. Mar; 53:810–820.
- 29)Pierre Bedossa, Christine Poitou, Nicolas Veyrie, Jean-Luc Bouillot, Arnaud Basdevant, Valerie Paradis, Joan Tordjman and Karine Clement. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* Nov; 56:1751-1759.
- 30)Pierre Bedossa and the FLIP Pathology Consortium (2014). Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology*. Aug; 60(2):565-75.
- 31)Brunt, Elizabeth M MD; Janney, Christine G MD; Di Bisceglie, Adrian M MD; Neuschwander-Tetri, Brent A MD; Bacon, Bruce R MD (1999). Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *American Journal of Gastroenterology*: September - Volume 94 - Issue 9 – p.2467-2474.
- 32)Alonso Romo-Romo, Carlos A Aguilar-Salinas, Griselda X Brito-Córdova, Rita A Gómez Díaz, David Vilchis Valentín, Paloma Almeda-Valdes (2016). Effects of the Non-Nutritive Sweeteners on Glucose Metabolism and Appetite Regulating Hormones: Systematic Review of Observational Prospective Studies and Clinical Trials. *PLoS One*. Aug 18;11(8):e0161264.
- 33)Myriam Alexander, A. Katrina Loomis, Jolyon Fairburn-Beech, Johan van der Lei, Talita Duarte-Salles, Daniel Prieto-Alhambra, David Ansell, Alessandro Pasqua, Francesco Lapi, Peter Rijnbeek, Mees Mosseveld, Paul Avillach, Peter Egger, Stuart Kendrick, Dawn M. Waterworth, Naveed Sattar and William Alazawi (2018). Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Medicine* volume 16, Article number: 130.

- 34)Janae Wentong Wai, Charmaine Fu, and Vincent Wai-Sun Wong (2020). Confounding factors of non-invasive tests for nonalcoholic fatty liver disease. *J Gastroenterol.* Aug; 55(8):731-741.
- 35)Vlad Ratziu, Frédéric Charlotte, Agnès Heurtier, Sophie Gombert, Philippe Giral, Eric Bruckert, André Grimaldi, Frédérique Capron, Thierry Poinard, LIDO Study Group. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Clinical Trial Gastroenterology.* Jun;128(7):1898-906.
- 36)Raj Vuppalanchi, Aynur Unalp, Mark L Van Natta, Oscar W Cummings, Kumar E Sandrasegaran, Tariq Hameed, James Tonascia and Naga Chalasani (2009). Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol.* Apr; 7(4):481-6.
- 37)Ke Pu, Yuping Wang, Suyang Bai, Hui Wei, Yongning Zhou, Jiangao Fan and Liang Qiao (2019). Diagnostic accuracy of controlled attenuation parameter (CAP) as a non-invasive test for steatosis in suspected non-alcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Gastroenterology* volume 19, Article number: 51.
- 38)Stefan Traussnigg, Christian Kienbacher, Martin Gajdošík, Ladislav Valkovič, Emina Halilbasic, Judith Stift, Christian Rechling, Harald Hofer, Petra Steindl-Munda, Peter Ferenci, Fritz Wrba, Siegfried Trattning, Martin Krššák and Michael Trauner (2017). Ultra-high-field magnetic resonance spectroscopy in non-alcoholic fatty liver disease: Novel mechanistic and diagnostic insights of energy metabolism in non-alcoholic steatohepatitis and advanced fibrosis. *Liver Int.* Oct; 37(10):1544-1553.
- 39)Scott B Reeder 1, Irene Cruite, Gavin Hamilton and Claude B Sirlin (2011). Quantitative Assessment of Liver Fat with Magnetic Resonance Imaging and Spectroscopy. *Review J Magn Reson Imaging.* Oct; 34(4):729-749.
- 40)Larissa Fedchuk, Fabio Nascimbeni, R Pais, Frédéric Charlotte, C Housset, Vlad Ratziu, LIDO Study Group (2014). Performance and limitations of steatosis biomarkers in patients with nonalcoholic fatty liver disease. *Alimentary Pharmacology & Therapeutics,* 29 Sep, 40(10):1209-1222.
- 41)Naim Alkhouri, MD and Arthur J. McCullough, MD (2012). Noninvasive Diagnosis of NASH and Liver Fibrosis Within the Spectrum of NAFLD. *Gastroenterol Hepatol (N Y).* Oct; 8(10): 661–668.
- 42)Laurent Castéra, Juliette Foucher, Pierre-Henri Bernard, Françoise Carvalho, Daniele Allaix, Wassil Merrouche, Patrice Couzigou and Victor de Lédinghen (2010). Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology.* Mar; 51(3):828-35.
- 43)Stuart McPherson, Tim Hardy, Jean-Francois Dufour, Salvatore Petta, Manuel Romero-Gomez, Mike Allison, Claudia P Oliveira, Sven Francque, Luc Van Gaal, Jörn M Schattenberg, Dina Tiniakos, Alastair Burt, Elisabetta Bugianesi, Vlad Ratziu, Christopher P Day and Quentin M Anstee (2017). Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. *Am J Gastroenterol.* May; 112(5):740-751.
- 44)Chung-Chau Hon, Jordan A. Ramilowski, Jayson Harshbarger, Nicolas Bertin, Owen J. L. Rackham, Julian Gough, Elena Denisenko, Sebastian Schmeier, Thomas M. Poulsen, Jessica Severin, Marina Lizio, Hideya Kawaji, Takeya Kasukawa, Masayoshi Itoh, A. Maxwell Burroughs, Shohei

Noma, Sarah Djebali, Tanvir Alam, Yulia A. Medvedeva, Alison C. Testa, Leonard Lipovich, Chi-Wai Yip, Imad Abugessaisa, Mickaël Mendez and Alistair R. R (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*. Mar 9; 543(7644):199-204.

45)Juliane C R Fernandes, Stephanie M Acuña, Juliana I Aoki, Lucile M Floeter-Winter and Sandra M Muxel (2019). Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA*. Feb 17; 5(1):17.

46)Melé M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res*. Jan; 27(1):27-37.

47)Viiri LE, Rantapero T, Kiamehr M, Alexanova A, Oittinen M, Viiri K, Niskanen H, Nykter M, Kaikkonen MU, Aalto-Setälä K (2019). Extensive reprogramming of the nascent transcriptome during iPSC to hepatocyte differentiation. *Sci Rep*. Mar 5; 9(1):3562.

48)Lorna W. Harries (2019). RNA Biology Provides New Therapeutic Targets for Human Disease. *Front Genet*. Mar 8; 10:205.

49)Quinn JJ and Chang HY. Unique features of long non-coding RNA biogenesis and function (2016). *Nature Reviews Genetics*; (1):47-62.

50)Jarroux J, Morillon A and Pinskaya M. History, discovery, and classification of lncRNAs. *Long Non Coding RNA Biology*; 1008:1-46.

51)Maite Huarte (2015). The emerging role of lncRNAs in cancer. *Nature Medicine*; (11):1253-1261.

52)Aleksandra E Kornienko, Philipp M Guenzl, Denise P Barlow and Florian M Pauler (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*; (1):1-14.

53)Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K and Guttman M (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. Aug 16; 341(6147):1237973.

54)Simão Teixeira da Rocha, Valentina Boeva, Martin Escamilla-Del-Arenal, Katia Ancelin, Camille Granier, Neuza Reis Matias, Serena Sanulli, Jen Chow, Edda Schulz, Christel Picard, Syuzo Kaneko, Kristian Helin, Danny Reinberg, A Francis Stewart, Anton Wutz, Raphaël Margueron and Edith Heard (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell*. Jan 23; 53(2):301-16.

55)Yuntao S Mao, Hongjae Sunwoo, Bin Zhang and David L Spector (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature Cell Biology* ;(1):95-101.

56)Robert S. Young, Yatendra Kumar, Wendy A. Bickmore and Martin S. Taylor (2017). Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biology* volume 18, Article number: 242.

- 57) Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Müller F, Forrest ARR, Carninci P, Rehli M and Sandelin A (2014). An atlas of active enhancers across human cell types and tissues. *Nature*. Mar 27; 507(7493):455-461.
- 58) Erin M Wissink, Anniina Vihervaara, Nathaniel D Tippens and John T Lis (2019). Nascent RNA Analyses: Tracking Transcription and Its Regulation. *Nat Rev Genet*. Dec; 20(12):705-723.
- 59) Evgenia Ntini, Annita Louloupi, Julia Liz, Jose M Muino, Annalisa Marsico, Ulf Andersson Vang Ørom (2018). Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nature Communications* volume 9, Apr 24;9(1):1636.
- 60) Jennifer Y Tan, Adriano Biasini, Robert S Young, and Ana C Marques (2020). Splicing of enhancer-associated lincRNAs contributes to enhancer activity. *Life Sci Alliance*. Feb 21; 3(4):e202000663.
- 61) Shivendra Kishore, Sandra Luber and Mihaela Zavolan (2010). Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief Funct Genomics*. Dec; 9:391–404.
- 62) Bruno Pereira, Marc Billaud and Raquel Almeida (2017). RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer* Jul; 3:506–528.
- 63) Alfredo Castello, Bernd Fischer, Katrin Eichelbaum, Rastislav Horos, Benedikt M Beckmann, Claudia Strein, Norman E Davey, David T Humphreys, Thomas Preiss, Lars M Steinmetz, Jeroen Krijgsveld and Matthias W Hentze (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. Jun 8; 149:1393–1406.
- 64) Olivier Govaere, Simon Cockell, Dina Tiniakos, Rachel Queen, Ramy Younes, Michele Vacca, Leigh Alexander, Federico Ravaioli, Jeremy Palmer, Salvatore Petta, Jerome Boursier, Chiara Rosso, Katherine Johnson, Kristy Wonders, Christopher P. Day, Mattias Ekstedt, Matej Orešič, Rebecca Darlay, Heather J. Cordell, Fabio Marra, Antonio Vidal-Puig, Pierre Bedossa, Jörn M. Schattenberg, Karine Clément, Michael Allison, Elisabetta Bugianesi, Vlad Ratziu, Ann K. Daly and Quentin M. Anstee (2020) Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med*. Dec 2; 12(572):eaba4448.
- 65) Jonas Nørskov Søndergaard, Christian Sommerauer, Ionut Atanasoai, Laura C Hinte, Keyi Geng, Giulia Guiducci, Lars Bräutigam, Myriam Aouadi, Lovorka Stojic, Isabel Barragan and Claudia Kutter (2022) CCT3-LINC00326 axis regulates hepatocarcinogenic lipid metabolism. *Gut*. Jan 12; 71(10):2081-2092.
- 66) Andrew M. Jobbins, Nejc Haberman, Natalia Artigas, Christopher Amourda, Helen A.B. Paterson, Sijia Yu, Samuel J.I. Blackford , Alex Montoya, Marian Dore, Yi-Fang Wang, Alessandro Sardini, Inês

Cebola, Johannes Zuber, Sheikh Tamir Rashid, Boris Lenhard and Santiago Vernia (2022). Dysregulated RNA polyadenylation contributes to metabolic impairment in non-alcoholic fatty liver disease. Nucleic Acids Research Apr 8; 50(6):3379-3393.

67) Timothy Hardy, Kristy Wonders, Ramy Younes, Guruprasad P. Aithal, Rocio Aller, Michael Allison, Pierre Bedossa, Fay Betsou, Jerome Bourquier, M. Julia Brosnan, Alastair Burt, Jeremy Cobbold, Helena Cortez-Pinto, Chris P. Daya,b, Jean-Francois Dufour, Mattias Ekstedt, Sven Francquen, Stephen Harrison, Luca Miele, Patrik Nasr, George Papathodoridis, Salvatore Petta, Dina Tiniakos, Richard Torstenson, Luca Valenti, Adriaan G. Holleboomw, Hannele Yki-Jarvinen, Andreas Geier, Manuel Romero-Gomez, Vlad Ratiua, Elisabetta Bugianesi, Jörn M. Schattenberg and Quentin M. Ansteea on behalf of the LITMUS Consortium (2020). The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemporary Clinical Trials 98. Oct 9; 106175.

68) Mirolyuba Ilieva, James Dao, Henry E. Miller, Jens Hedelund Madsen, Alexander J. R. Bishop, Sakari Kauppinen and Shizuka Uchida (2022). Systematic Analysis of Long Non-Coding RNA Genes in Nonalcoholic Fatty Liver Disease. Non-Coding RNA; 8, 56. <https://doi.org/10.3390/ncrna8040056>

69) Biljana Atanasovska, Sander S. Rensen, Glenn Marsman, Ronit Shiri-Sverdlov, Sebo Withoff, Folkert Kuipers, Cisca Wijmenga, Bart van de Sluis and Jingyuan Fu (2021). Long Non-Coding RNAs Involved in Progression of Non-Alcoholic Fatty Liver Disease to Steatohepatitis. Cells 2021, 10, 1883. <https://doi.org/10.3390/cells10081883>

70) Arezoo Gowhari Shabgah, Fatemeh Norouzi, Mahdiyeh Hedayati-Moghadam, Davood Soleimani⁴, Naseh Pahlavani and Jamshid Gholizadeh Navashenaq (2021). A comprehensive review of long non-coding RNAs in the pathogenesis and development of non-alcoholic fatty liver disease. Nutrition & Metabolism (Lond) 2021; 18:22. <https://doi.org/10.1186/s12986-021-00552-5>

71) Christian Sommerauer and Claudia Kutter (2022). Noncoding RNAs and RNA-binding proteins: emerging governors of liver physiology and metabolic diseases. Am J Physiol Cell Physiol 323: C1003–C1017. 10.1152/ajpcell.00232.2022

72) Xuan Kan, Geli Liu, Yong Yang, Qingyan Yang, Yapu Li, Feng Wang (2016). Serum vascular endothelial cadherin and thrombomodulin are markers of non-alcoholic fatty liver disease in children. J Pediatr Endocrinol Metab. Dec 1;29(12):1331-1335.

73) Valerie Sapp, Leah Gaffney, Steven F EauClaire, Randolph P Matthews (2014). Fructose leads to hepatic steatosis in zebrafish that is reversed by mechanistic target of rapamycin (mTOR) inhibition. Nov;60(5):1581-92. doi: 10.1002/hep.27284

74) Na Jiao, Susan S Baker, Adrian Chapa-Rodriguez, Wensheng Liu, Colleen A Nugent, Maria Tsompana, Lucy Mastrandrea, Michael J Buck, Robert D Baker, Robert J Genco, Ruixin Zhu and Lixin Zhu (2018). Suppressed hepatic bile acid signalling despite elevated production of primary and secondary bile acids in NAFLD. Gut. Oct;67(10):1881-1891.

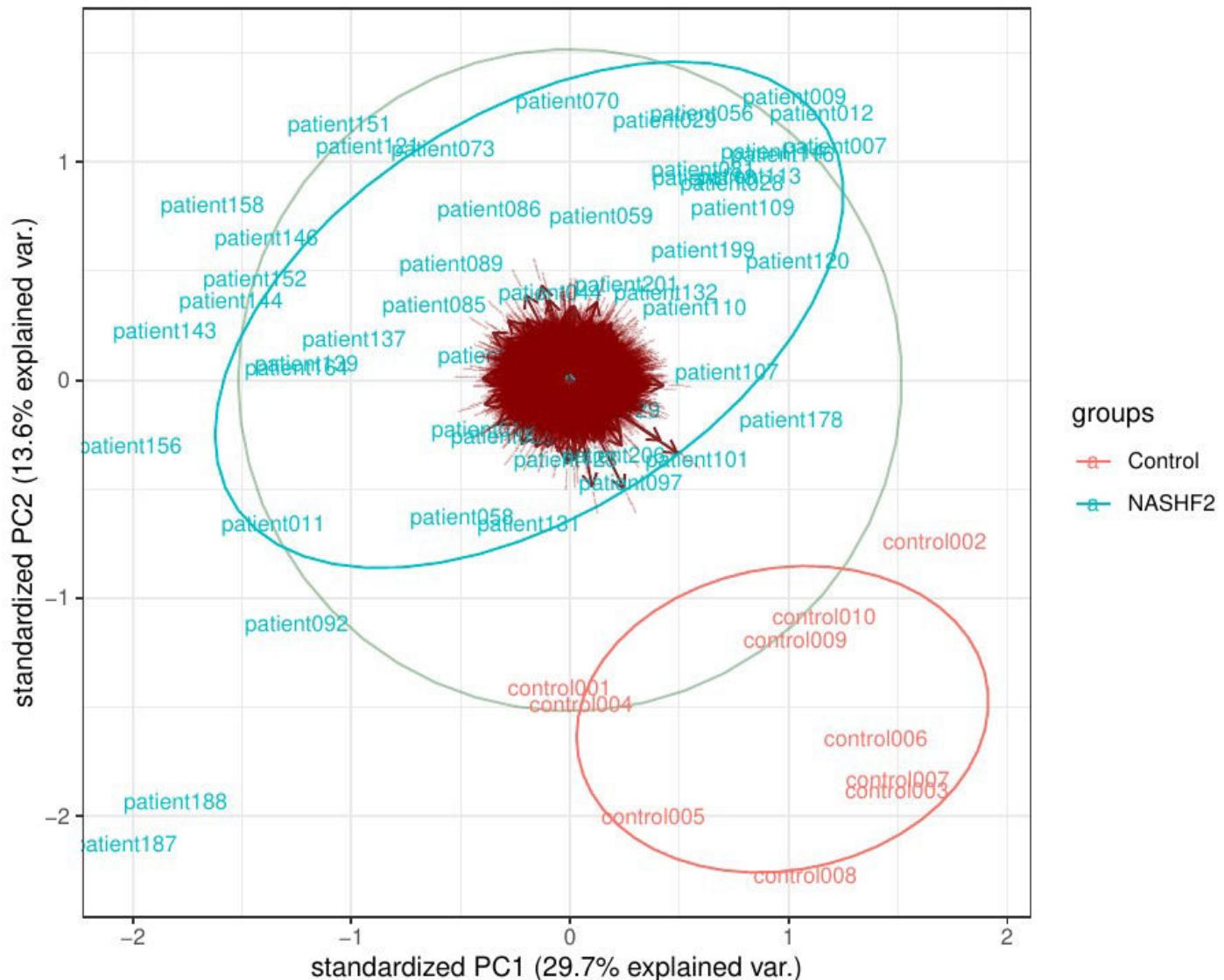
75)Yuan Tian, Myth T.S. Mok, Pengyuan Yang, and Alfred S.L. Cheng (2016). Samy L. Habib, Academic Editor. Epigenetic Activation of Wnt/β-Catenin Signaling in NAFLD-Associated Hepatocarcinogenesis. *Cancers* (Basel). Aug; 8(8): 76. P. doi: 10.3390/cancers8080076

76)Kamyar Asadipooya, Kamran B. Lankarani, Rishi Raj and Mohammadreza Kalantarhormozi (2019). RAGE is a Potential Cause of Onset and Progression of Nonalcoholic Fatty Liver Disease. *Int J Endocrinol.* Sep 18; 2151302. doi: 10.1155/2019/2151302

77)Eun Jin Sun, Miriam Wankell, Pranavan Palamuthusingam, Craig McFarlane, and Lionel Hebbard (2021). Targeting the PI3K/Akt/mTOR Pathway in Hepatocellular Carcinoma. *Biomedicines*. Nov 8; 9(11): 1639. doi: 10.3390/biomedicines9111639

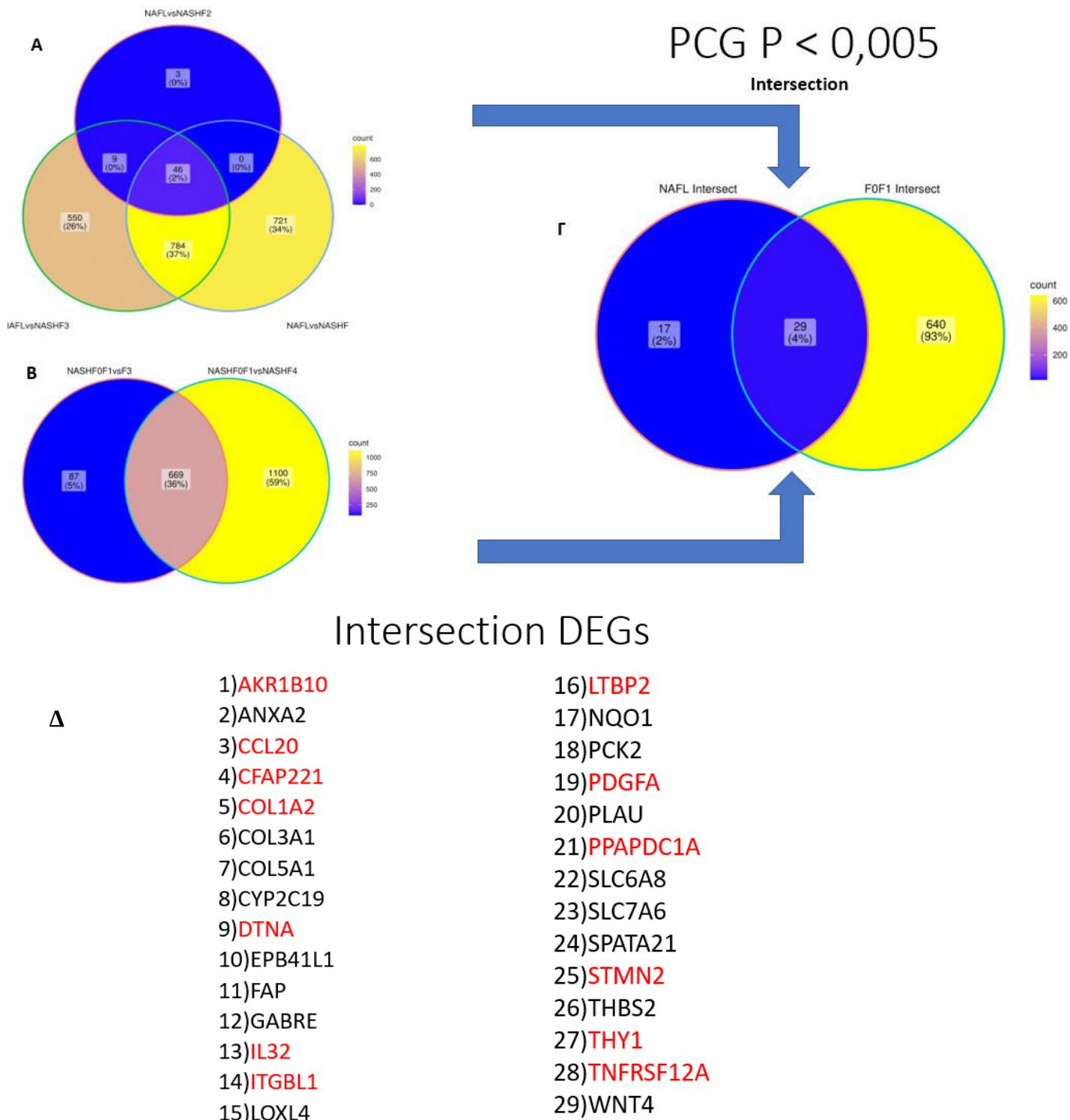
ΠΑΡΑΡΤΗΜΑ

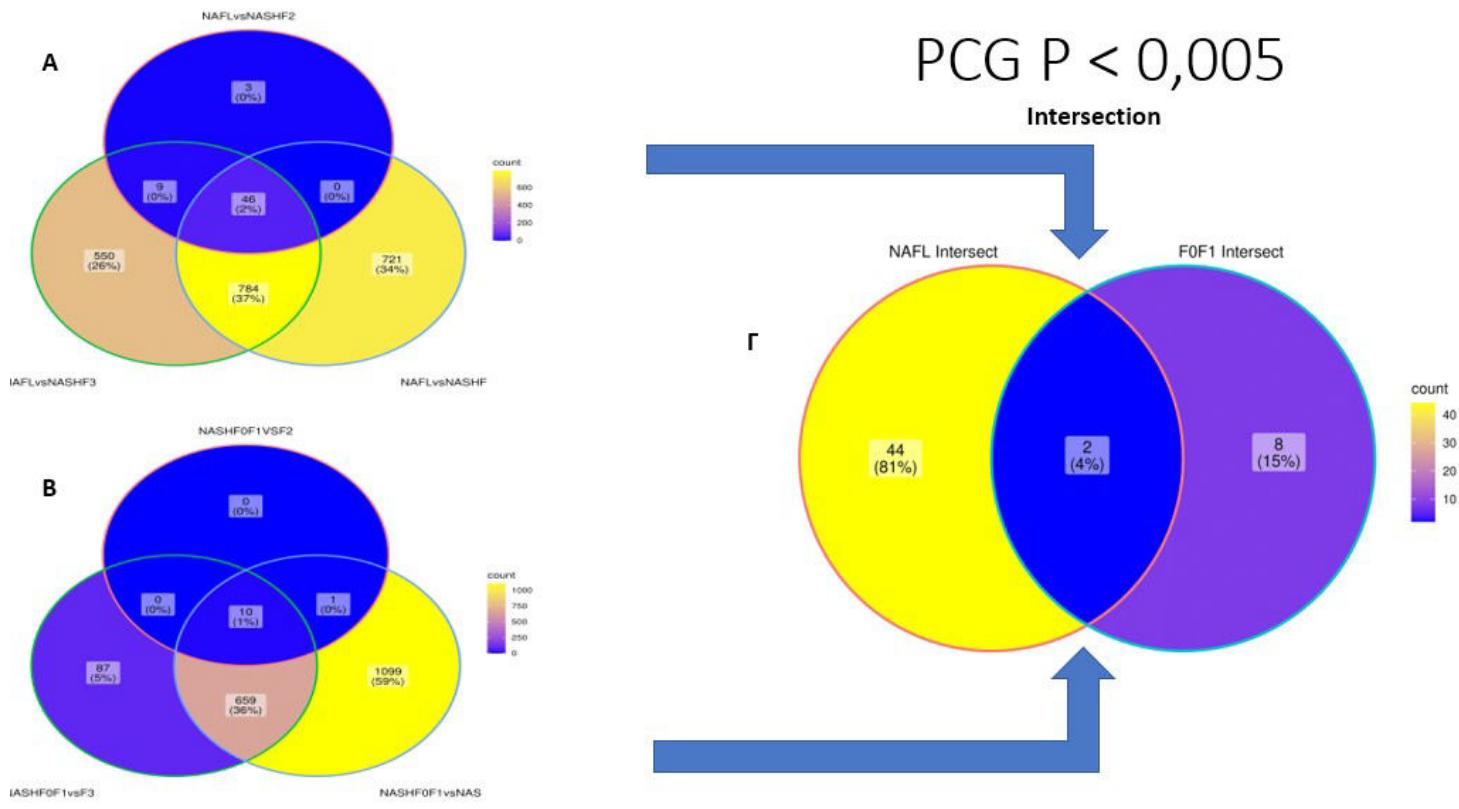
Παράρτημα 1. Ανάλυση κυρίων συνιστωσών για επαλήθευση πως τα δείγματα control δεν κατηγοριοποιούνται μαζί με τα δείγματα μη-αλκοολικής στεατοηπατίδας με ίνωση δεύτερου σταδίου (NASH F2). Με κόκκινο είναι τα δείγματα control (n=10) και μπλε τα δείγματα NASH F2(n= 52). Τα δείγματα control 1 και control 10 είχαν χαρακτηριστεί με ίνωση δεύτερου σταδίου, αλλά δεν φαίνεται να ομαδοποιούνται μαζί με τα δείγματα των NASH F2. Η εικόνα παράχθηκαν από το πακέτο ggbiplot.



Παράτημα 2.

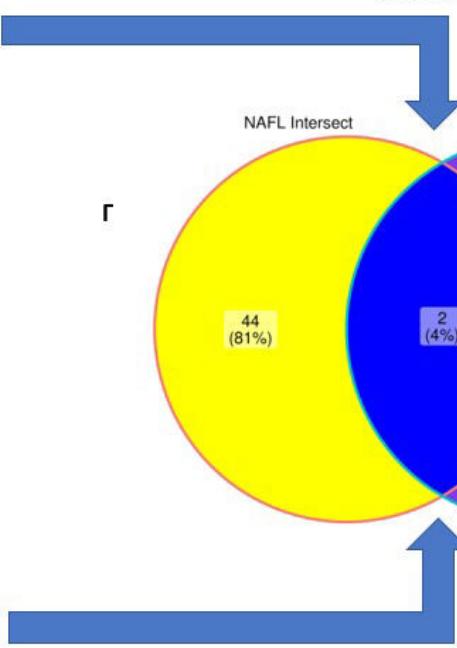
Διαγράμματα Venn με τους αριθμούς των διαφορικά εκφραζόμενων γονιδίων σε διαφορετικά p-value cutoffs. Τα 2 επίπεδα αναφοράς (Α) NAFL, (Β) NASH F0-F1 και (Γ)τα κοινά DEGs μεταξύ των δύο. Όπου παραλείπεται η σύγκριση για NASH F0-F1 vs NASH F2 έχει γίνει ώστε να υπάρχει ανάλογη σύγκριση με το άρθρο αναφοράς. (Δ) Αναγράφονται τα κοινά DEGs μεταξύ των 2 επιπέδων αναφοράς (Intersection DEGs). Με κόκκινο φαίνονται τα κοινά DEGs μεταξύ της ανάλυσης μας και του άρθρου αναφοράς. Οι εικόνες παράχθηκαν από το πακέτο ggVennDiagram.





PCG P < 0,005

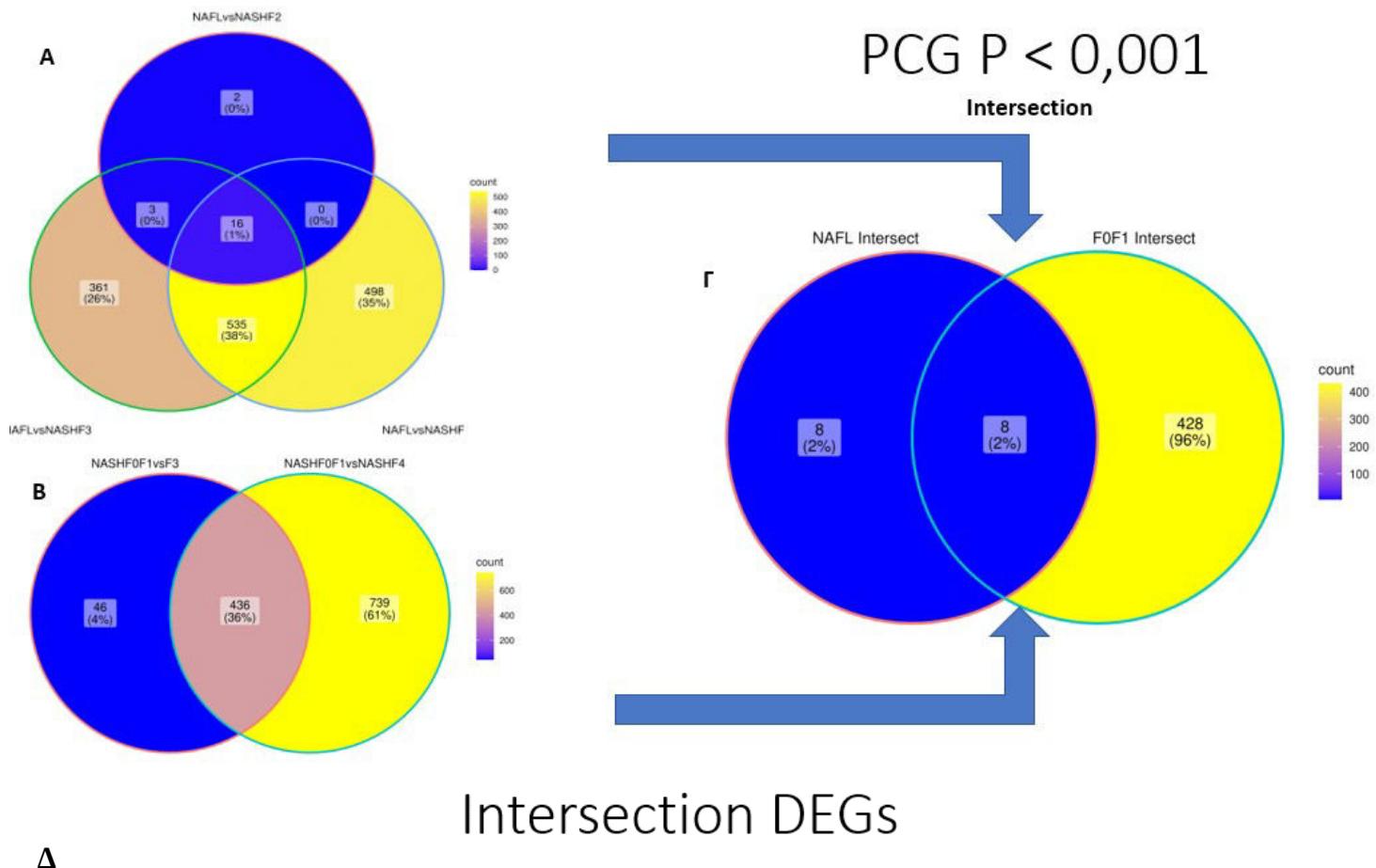
↑



Intersection DEGs

Δ

- NASH F0-F1-baseline
 - 1) **CCL20**
 - 2) **CLIC6**
 - 3) CXCL1
 - 4) CXCL8
 - 5) EEF1A2
 - 6) GABRE
 - 7) KRT7
 - 8) **SCTR**
 - 9) SPINT1
 - 10) TACSTD2
- Intersection
 - 1) **CCL20**
 - 2) GABRE



Intersection DEGs

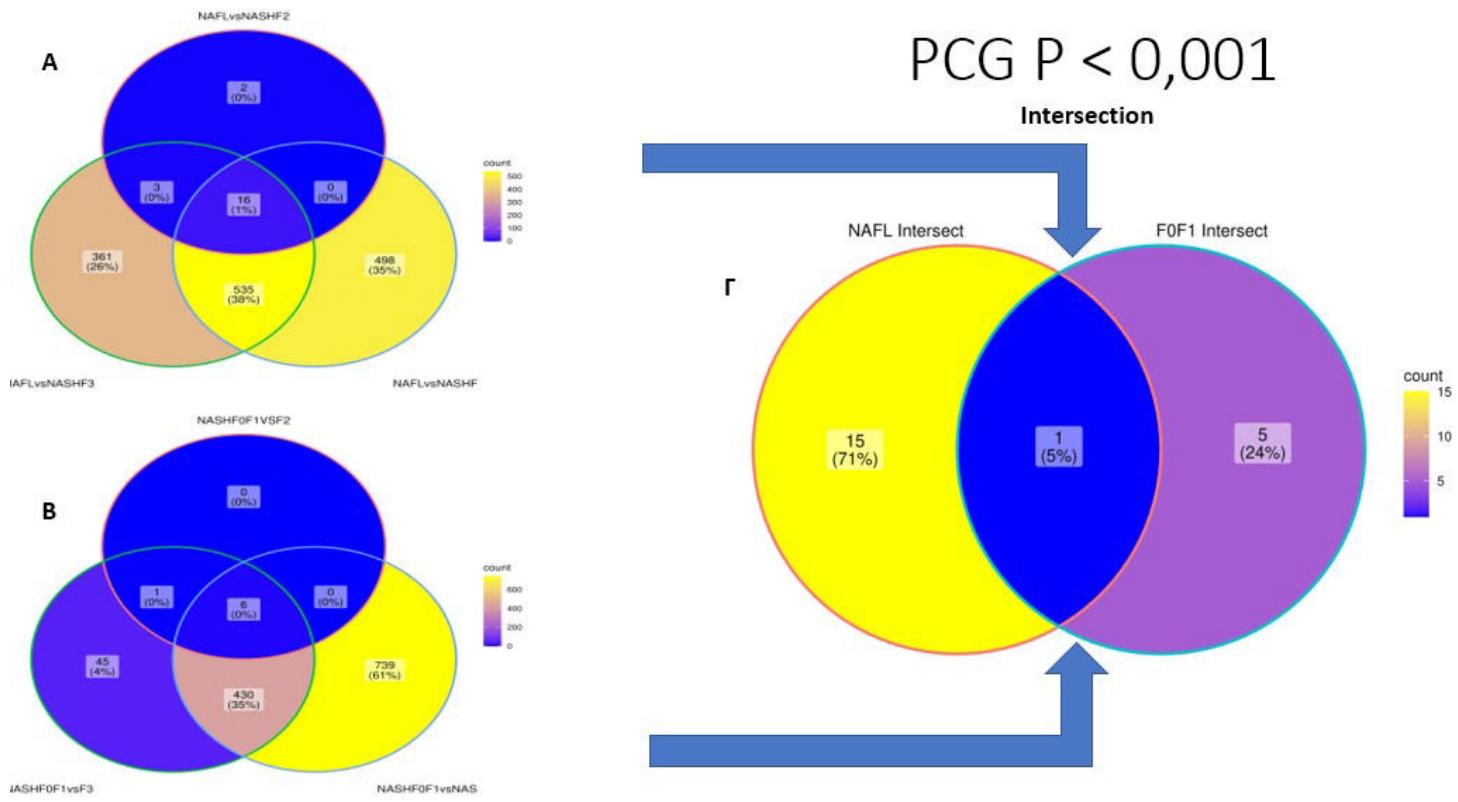
Δ

• NAFL-baseline

- | | |
|------------------|----------------------|
| 1)AHCY | 9)LOXL4 |
| 2)AMICA1 | 10) LTBP2 |
| 3)CCDC64 | 11)NAAA |
| 4) CCL20 | 12) PDGFA |
| 5)COL1A1 | 13)RAB7B |
| 6) COL1A2 | 14)SLC7A6 |
| 7)COL5A1 | 15) STMN2 |
| 8)CYP3A7 | 16) TNFRSF12A |

• Intersection

- | |
|---------------------|
| 1) CCL20 |
| 2) COL1A2 |
| 3)LOXL4 |
| 4) LTBP2 |
| 5)PDGFA |
| 6)SLC7A6 |
| 7) STMN2 |
| 8) TNFRSF12A |



Intersection DEGs

Δ

- **NAFL-baseline**

- 1)AHCY 9)LOXL4
- 2)AMICA1 10)**LTBP2**
- 3)CCDC64 11)NAAA
- 4)**CCL20** 12)**PDGFA**
- 5)COL1A1 13)RAB7B
- 6)**COL1A2** 14)SLC7A6
- 7)COL5A1 15)**STMN2**
- 8)CYP3A7 16)**TNFRSF12A**

- **NASH F0-F1-baseline**

- 1)**CCL20**
- 2)CXCL1
- 3)GABRE
- 4)KRT7
- 5)SCTR
- 6)TACSTD2

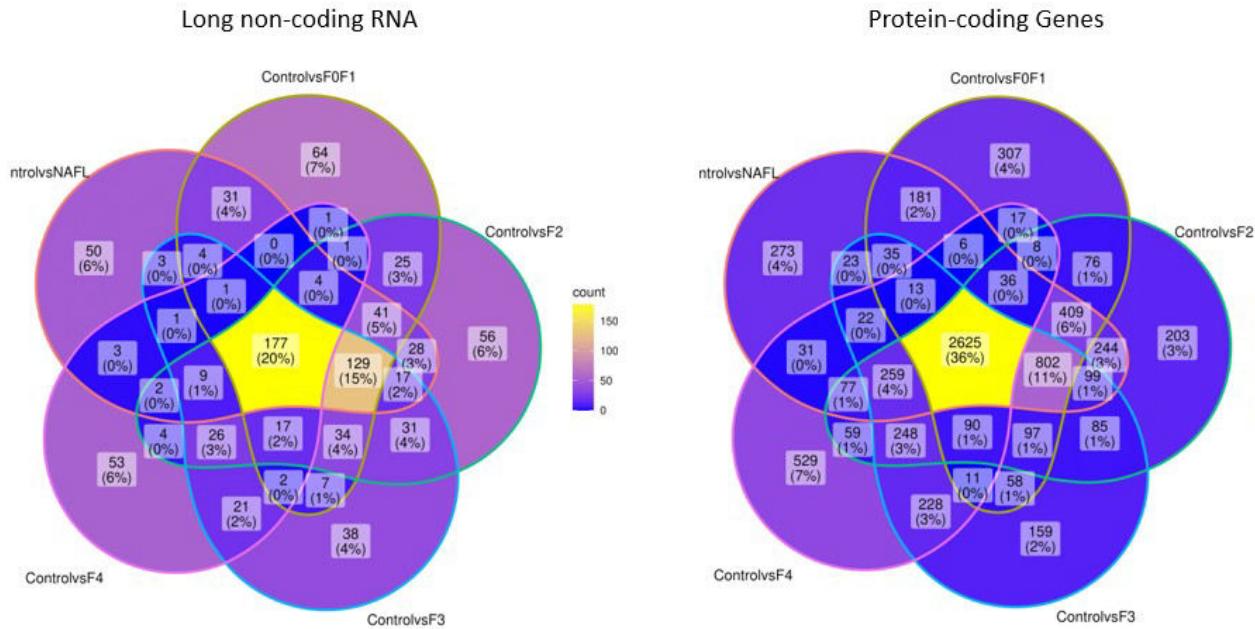
- **Intersection**

- 1)**CCL20**

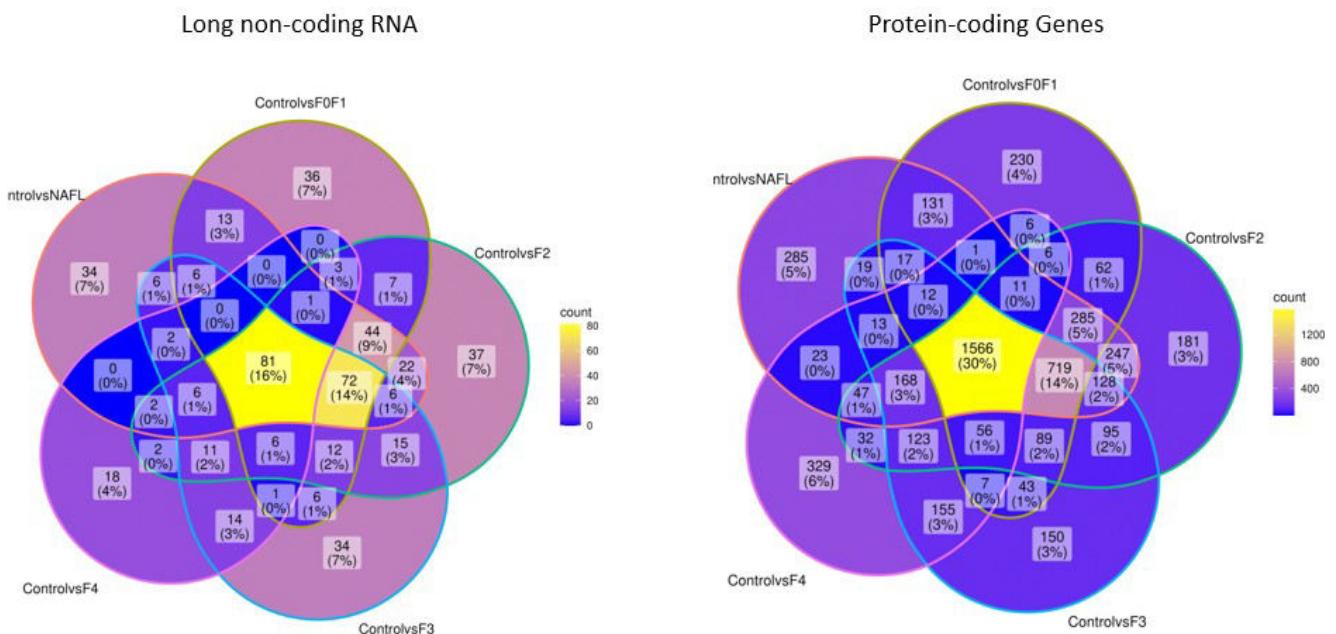
Παράτημα 3.

Διαγράμματα Venn με τους αριθμούς των διαφορικά εκφραζόμενων γονιδίων για cutoff με p-value<0,005. Επίπεδο αναφοράς είναι τα δείγματα control (n=10) και γίνεται σύγκριση κάθε φορά με ένα στάδιο της νόσου (Control vs NAFL, Control vs NASH F0-F1 κ.ο.κ). Με κίτρινο φαίνονται τα κοινώς διαφορικά εκφραζόμενα γονίδια σε όλα τα στάδια της νόσου. Οι εικόνες παράχθηκαν από το πακέτο ggVennDiagram.

Control-baseline P < 0,01

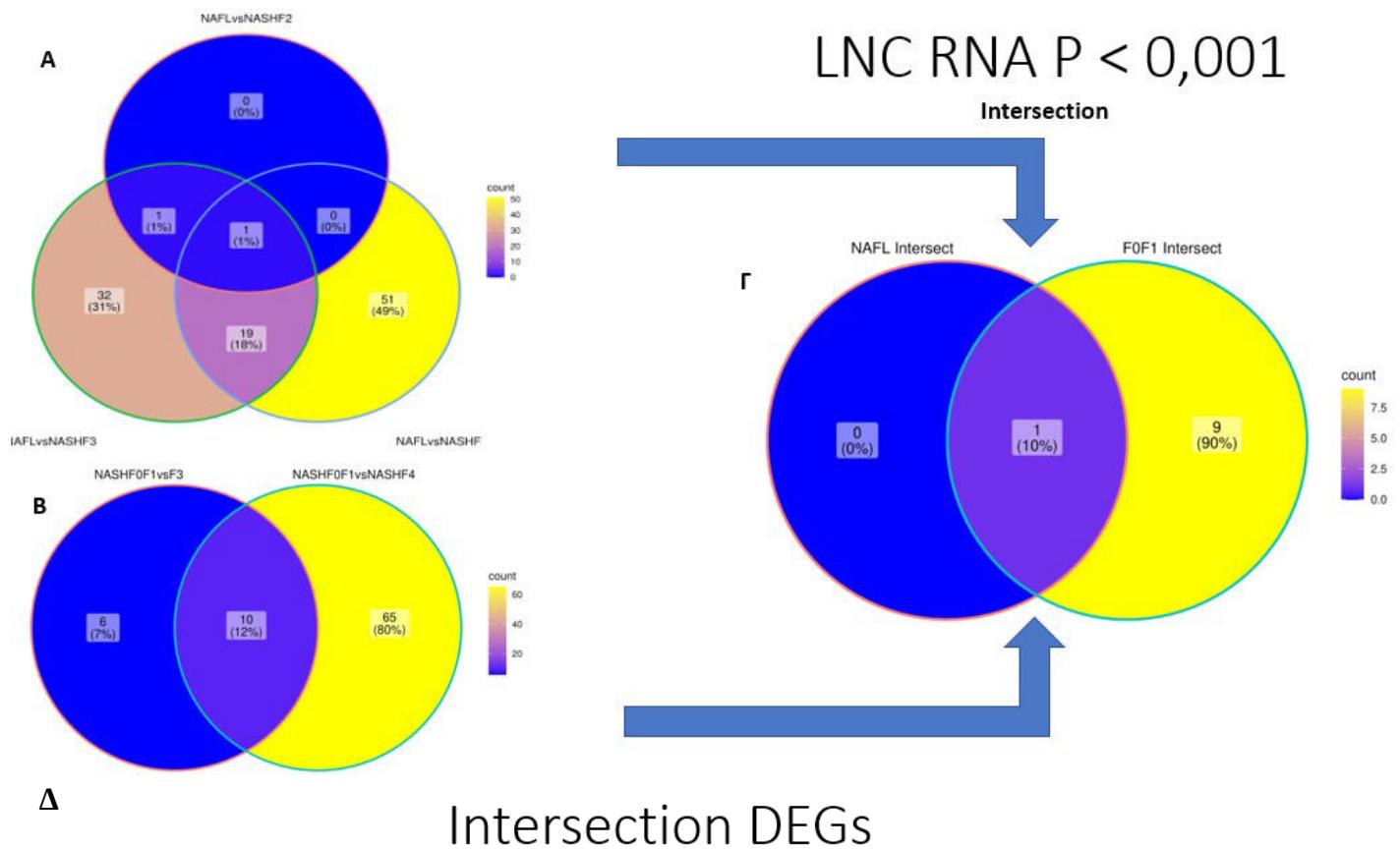


Control-baseline P < 0,001



Παράτημα 4.

Εικόνα 16. Διάγραμμα Venn με τους αριθμούς των DEGs για τα lncRNAs σε cutoff p<0,001. Τα 2 επίπεδα αναφοράς NAFL (Α), NASH F0-F1 (Β) και (Γ) τα κοινά DEGs μεταξύ των δύο. (Δ) Αναγράφονται τα κοινά DEGs για το επίπεδο αναφοράς NAFL (αριστερά), για το επίπεδο αναφοράς NASH F0-F1 (δύο κεντρικές στήλες) και τα κοινά μεταξύ των 2 επιπέδων αναφοράς (Intersection). Με κόκκινο φαίνονται τα 2 υποψήφια lncRNAs με ενδιαφέρον για περεταίρω μελέτη. Η εικόνα παράχθηκε από το πακέτο ggVennDiagram.



- **NAFL-baseline**

- 1)**GAS6-AS1**

- **NASH F0-F1-baseline**

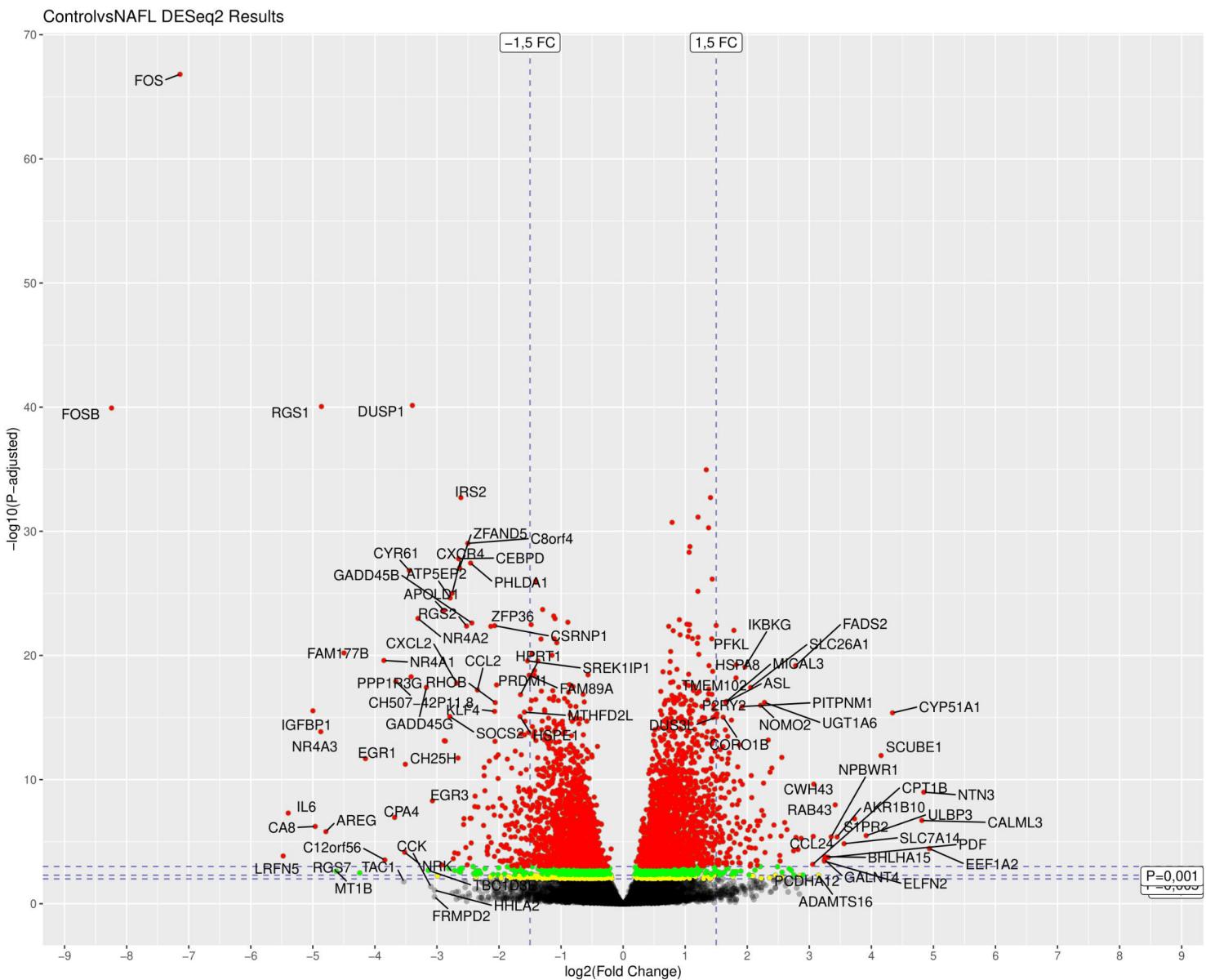
- 1)**GAS6-AS1**
- 2)GDNF-AS1
- 3)LINC01279
- 4)EGOT
- 5)CTC-273B12.8
- 6)RP11-96A15.1
- 7)RP11-532F12.5
- 8)RP11-767I20.1
- 9)RP11-404P21.3
- 10)**MIR4435-1HG**

- **Intersection**

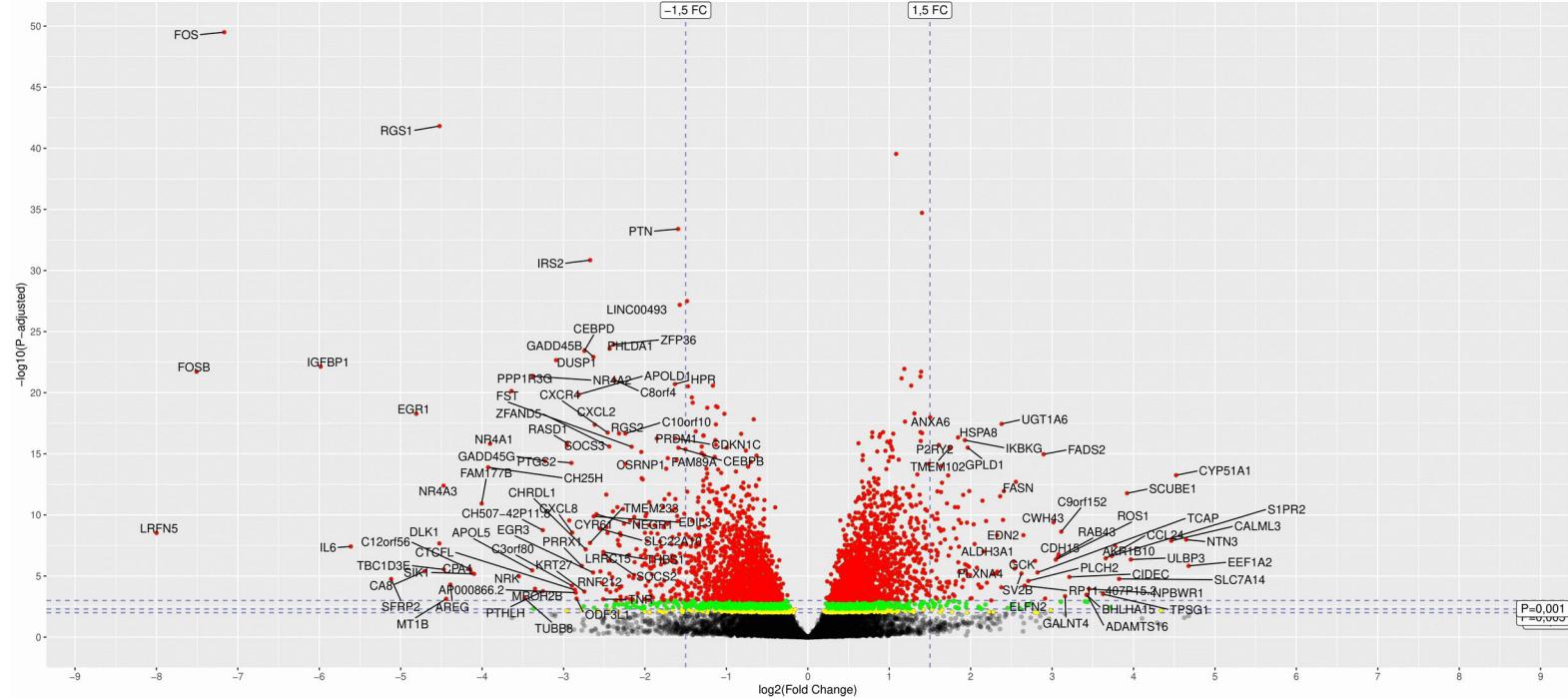
- 1)**GAS6-AS1**

Παράρτημα 5.

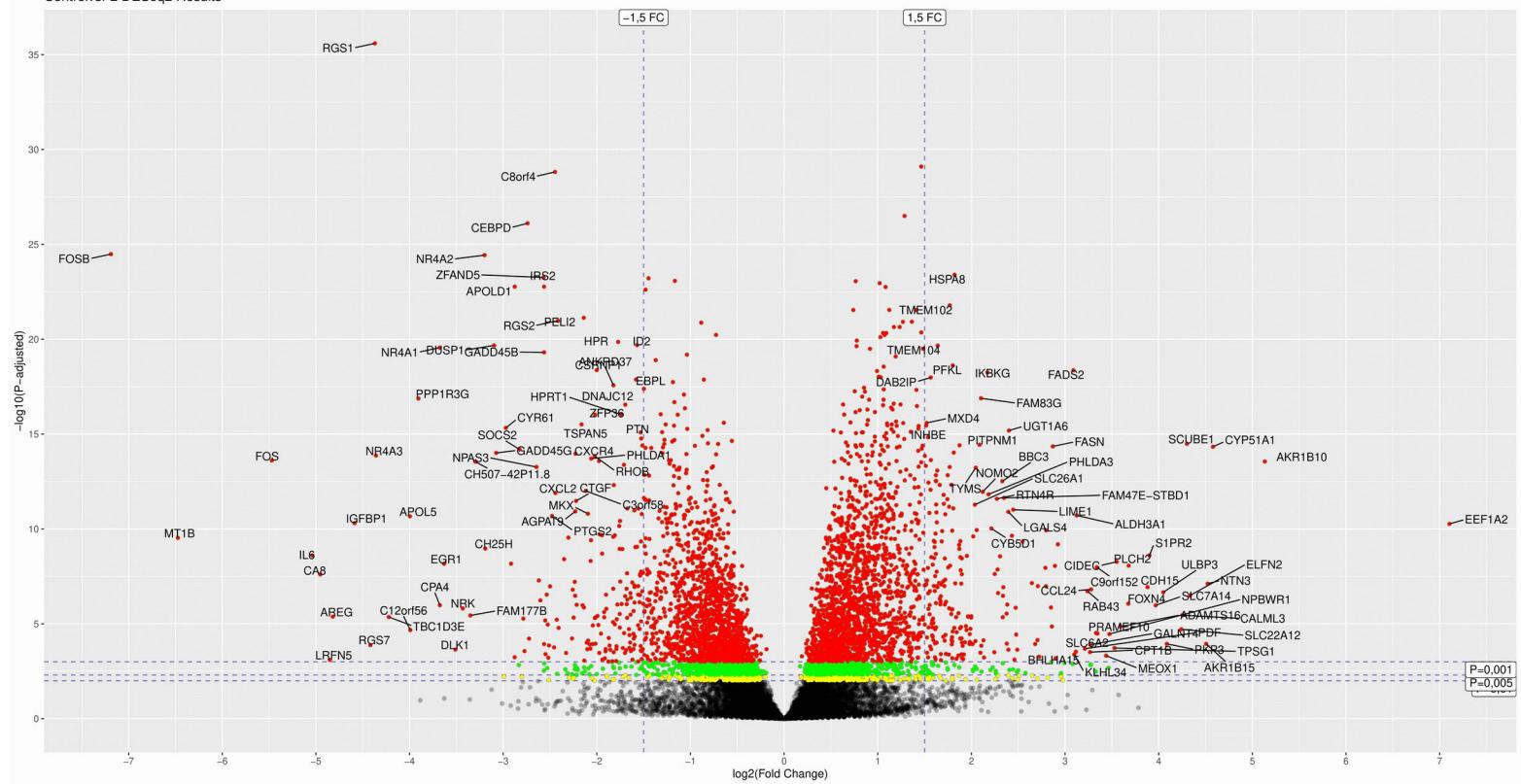
Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των PCGs για το επίπεδο αναφοράς control, μετά από την αφαίρεση των N/A τιμών.. Έγινε χρήση του αρνητικού λογαρίθμου του padj ($-\log_{10}(\text{padj})$) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος ($\log_2(\text{FC})$) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με $p\text{-value} < 0,01$ (padj), με πράσινο τα γονίδια μεταξύ $0,01 < p\text{-value} < 0,005$ και με κόκκινο τα γονίδια με $p\text{-value} < 0,001$. Αναγράφονται τα ονόματα (Gene ID) των γονιδίων με $p < 0,001$ και μεγάλη αλλαγή διπλώματος (πολύ αυξημένη ή πολύ μειωμένη έκφραση). Πάνω αριστερά αναφέρεται η σύγκριση η οποία απεικονίζεται. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.



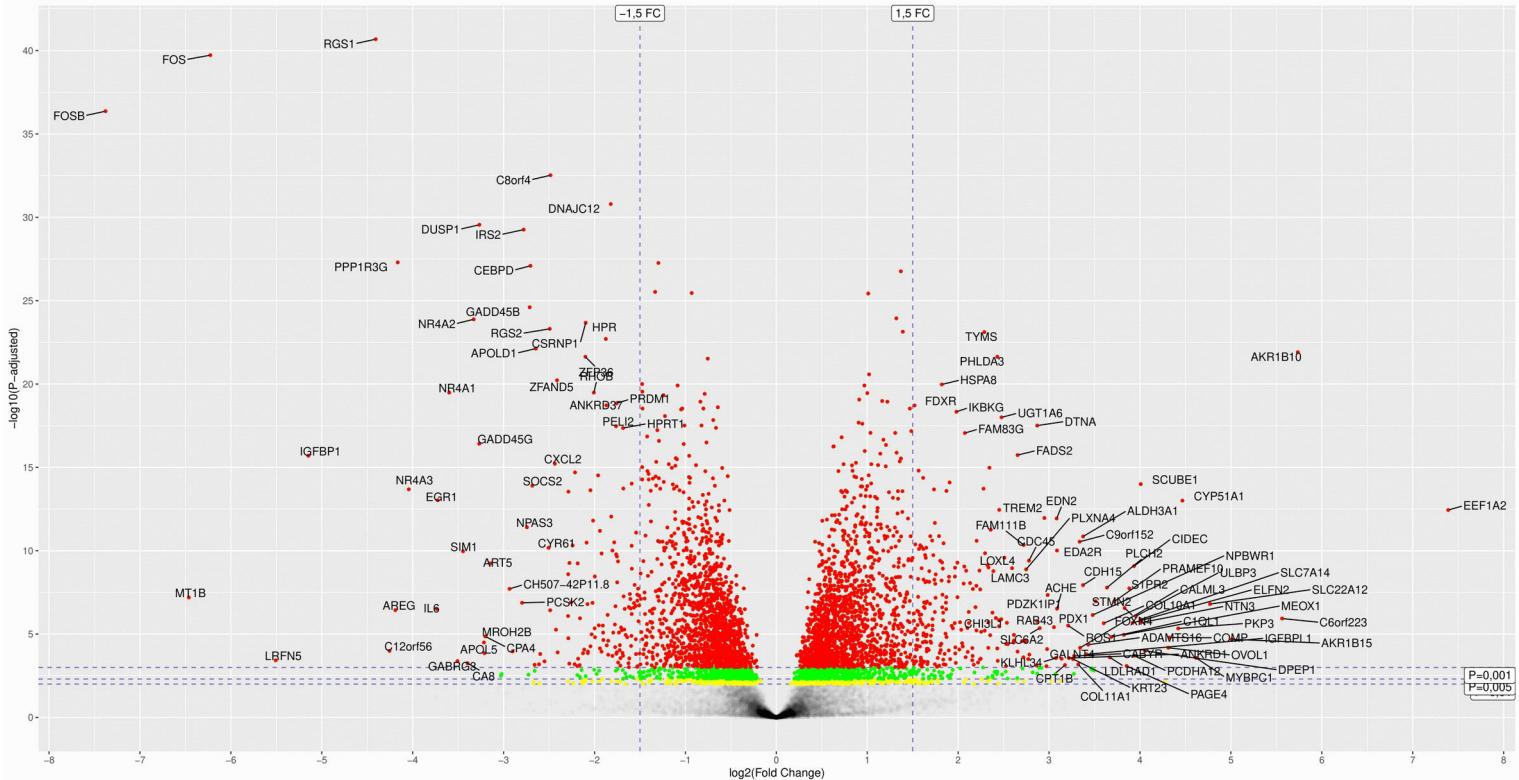
Controls vs F0F1 DESeq2 Results



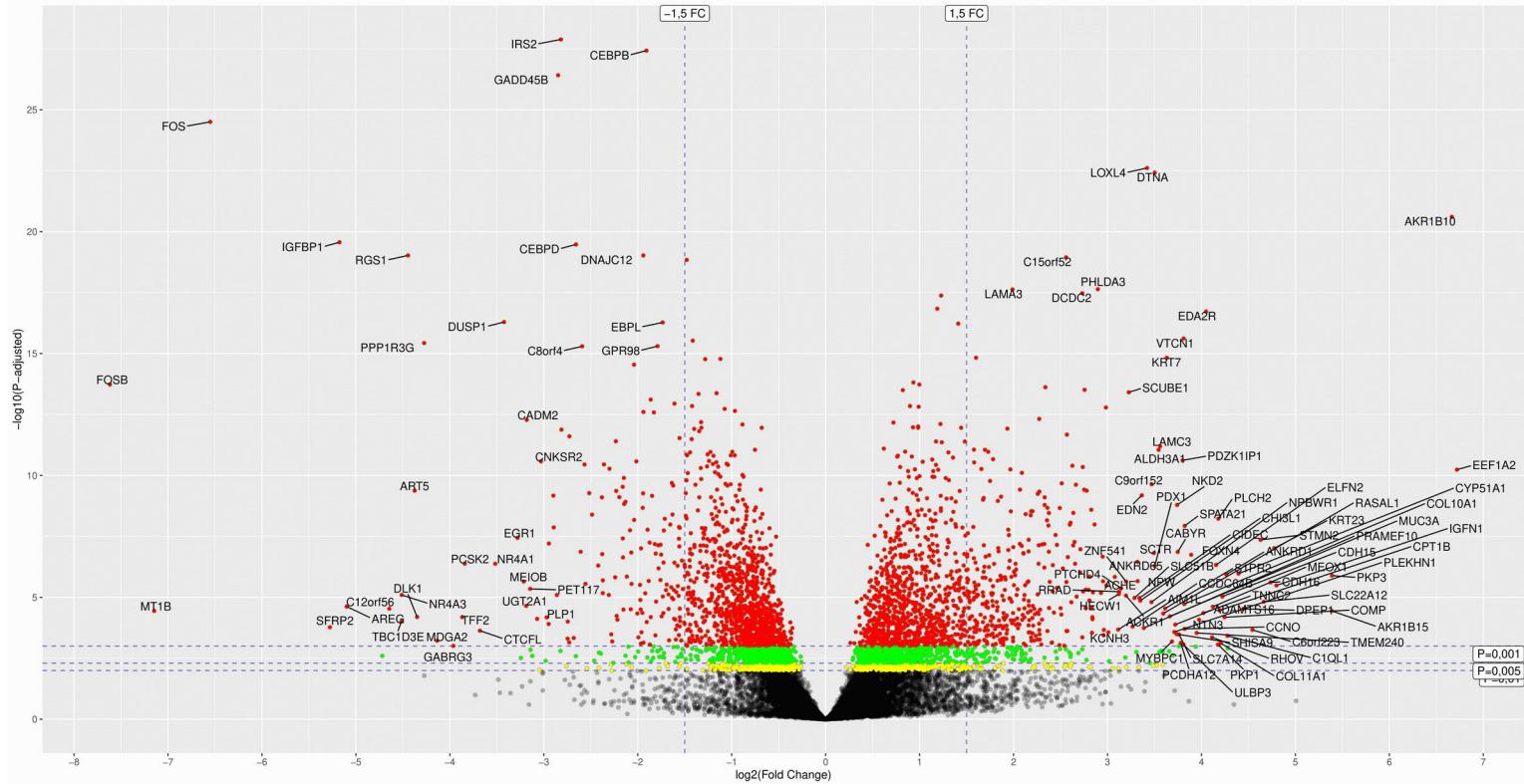
Controls vs F2 DESeq2 Results



ControlsF3 DESeq2 Results

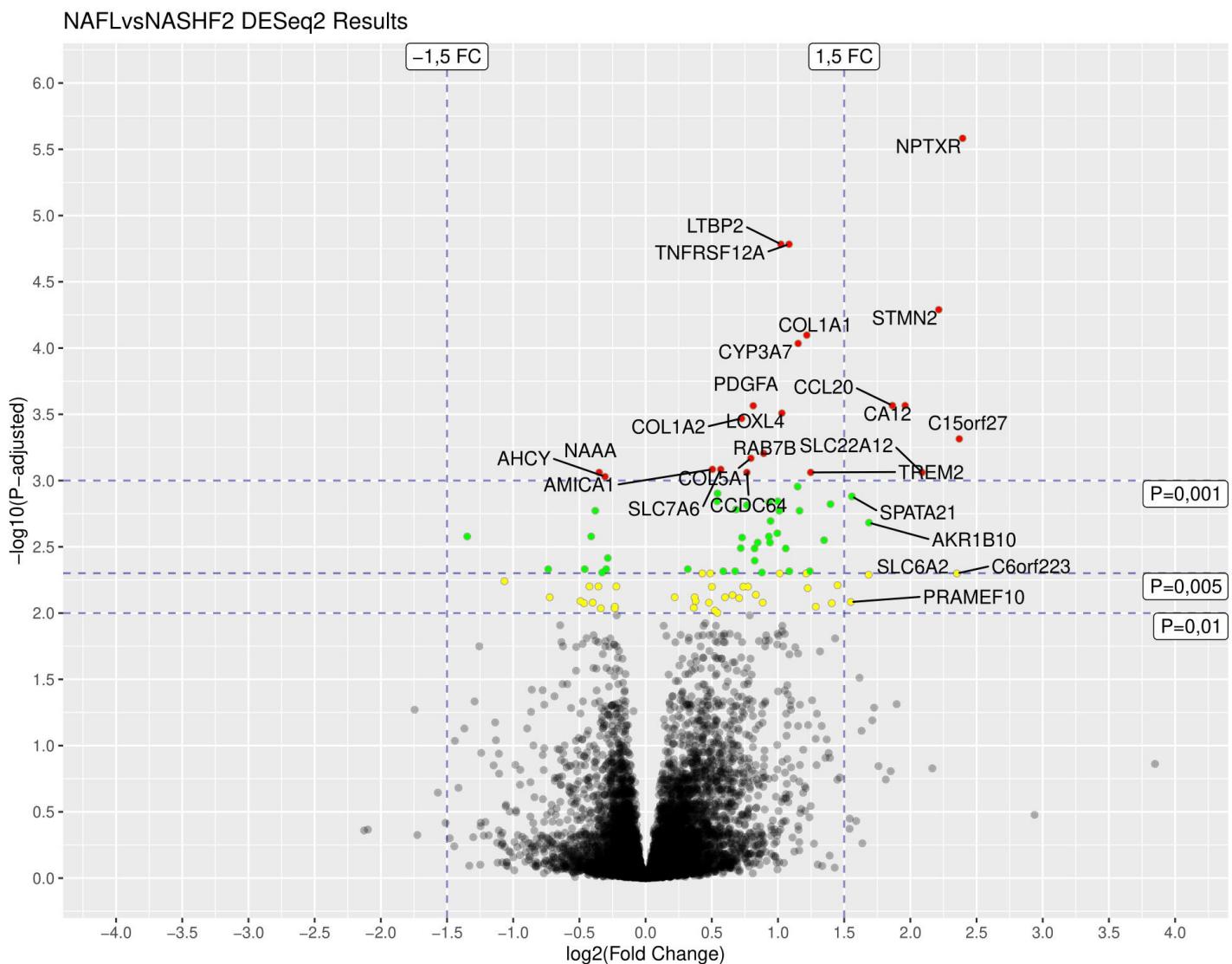


ControlsF4 DESeq2 Results

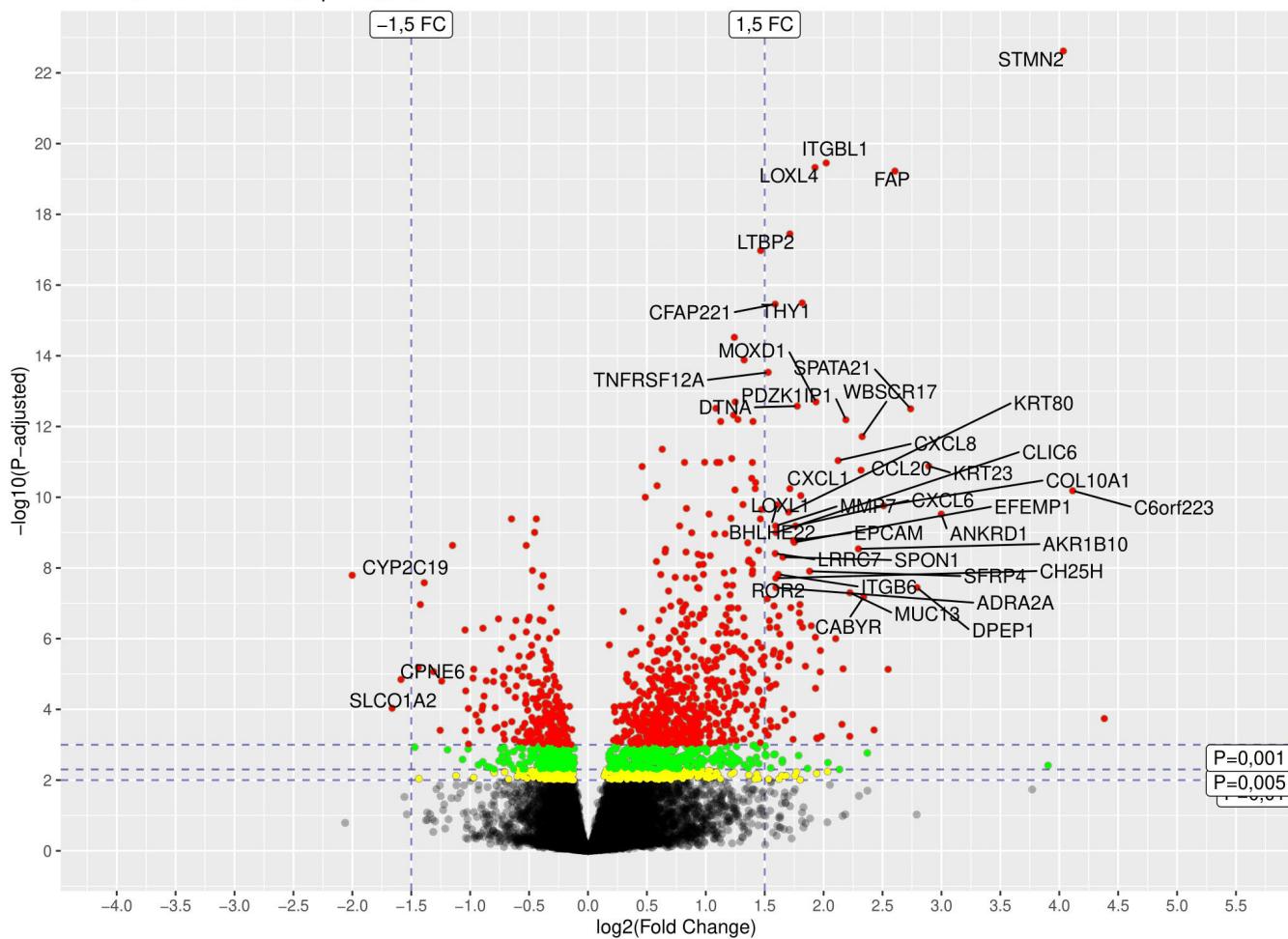


Παράτημα 6.

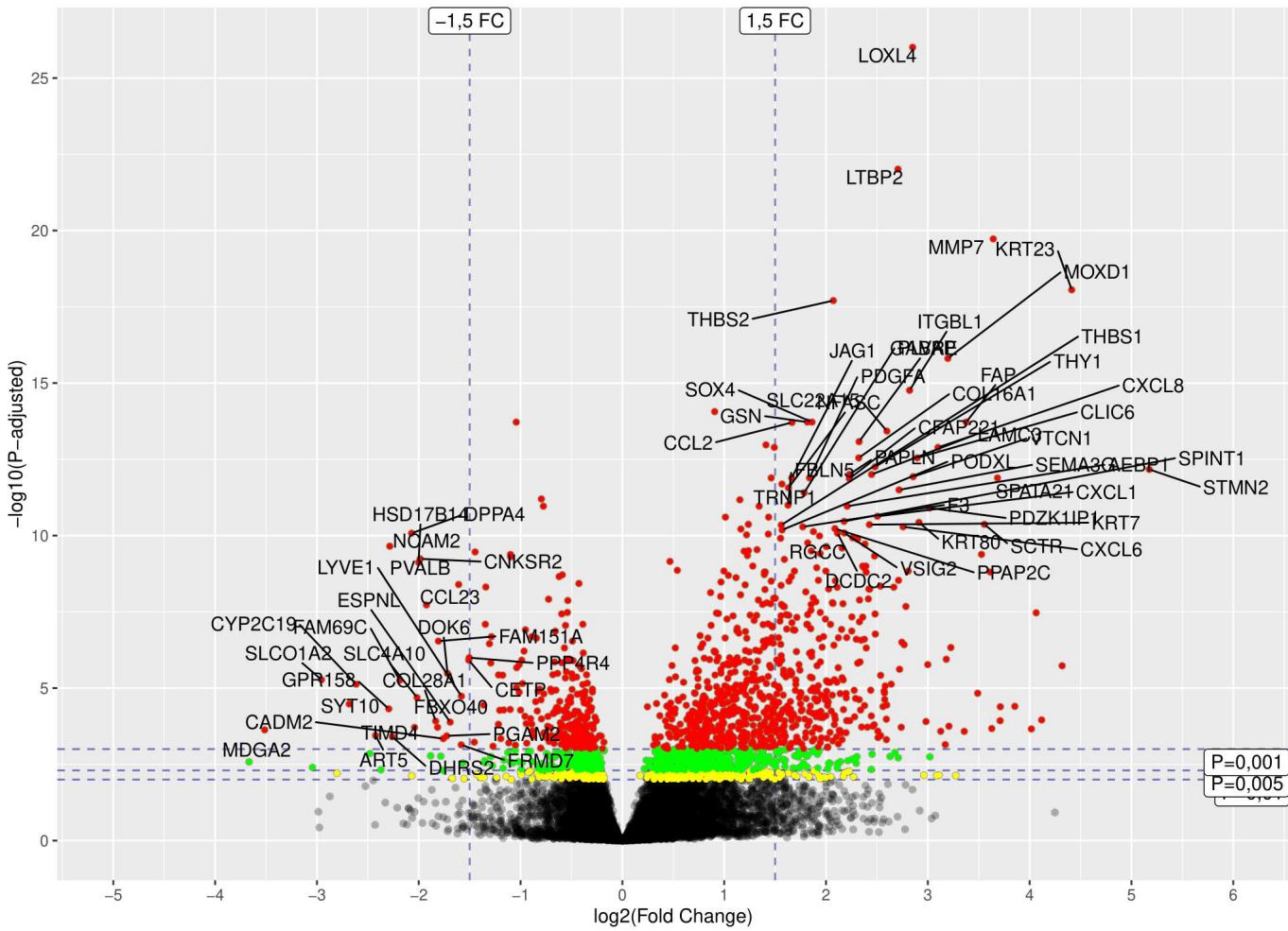
Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των PCGs για τα 2 επίπεδα αναφοράς (NAFL, NASH F0-F1), μετά από την αφαίρεση των N/A τιμών.. Έγινε χρήση του αρνητικού λογαρίθμου του padj (-log10(padj)) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος (log2(FC)) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με p-value<0,01 (padj), με πράσινο τα γονίδια μεταξύ 0,01 < p-value < 0,005 και με κόκκινο τα γονίδια με p-value < 0,001. Αναγράφονται τα ονόματα (Gene ID) των γονιδίων με στατιστικά σημαντικό p-value και μεγάλη αλλαγή διπλώματος (FC). Πάνω αριστερά αναφέρεται η σύγκριση η οποία απεικονίζεται. Δεν έγινε γράφημα τύπου για την ανάλυση NAFL vs NASH F0-F1 καθώς δεν βρέθηκαν στατιστικά σημαντικά γονίδια. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.



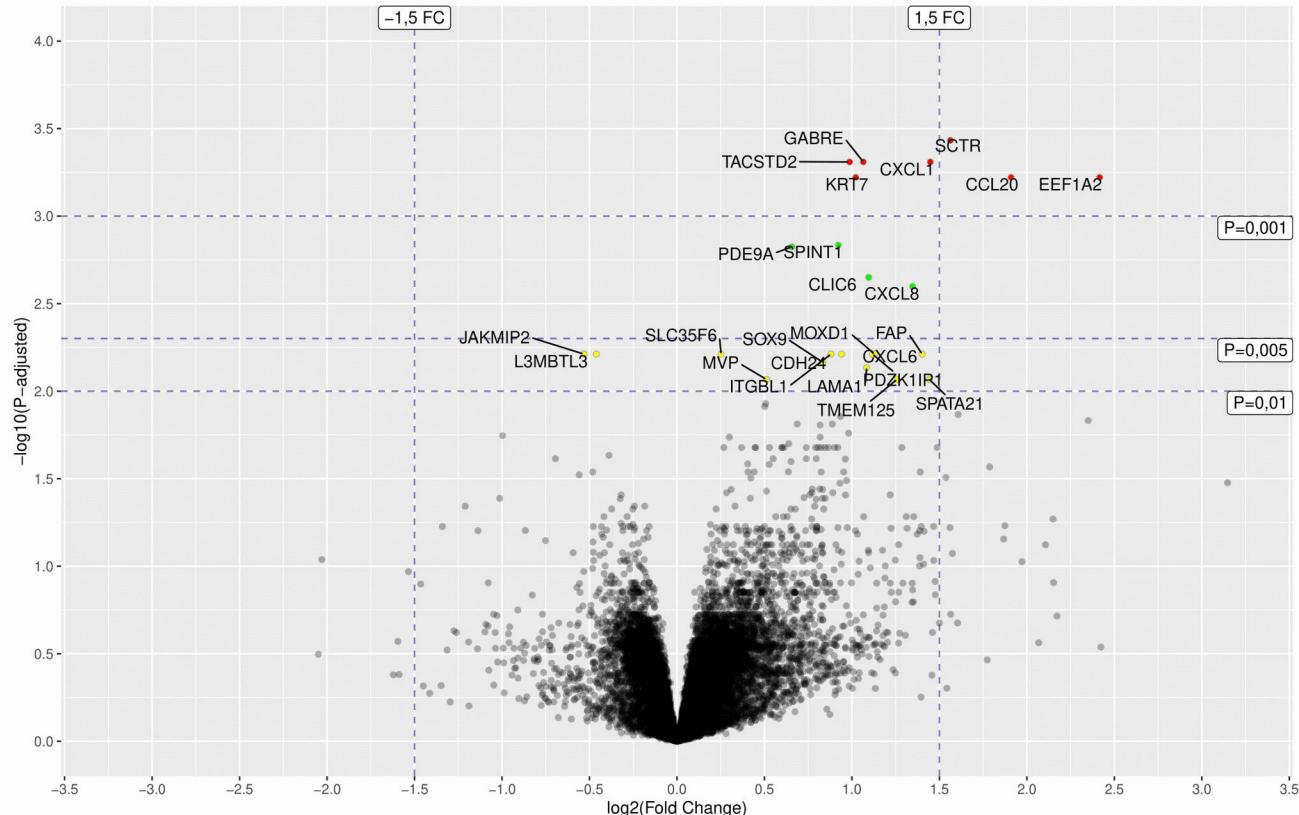
NAFLvsNASH-F3 DESeq2 Results



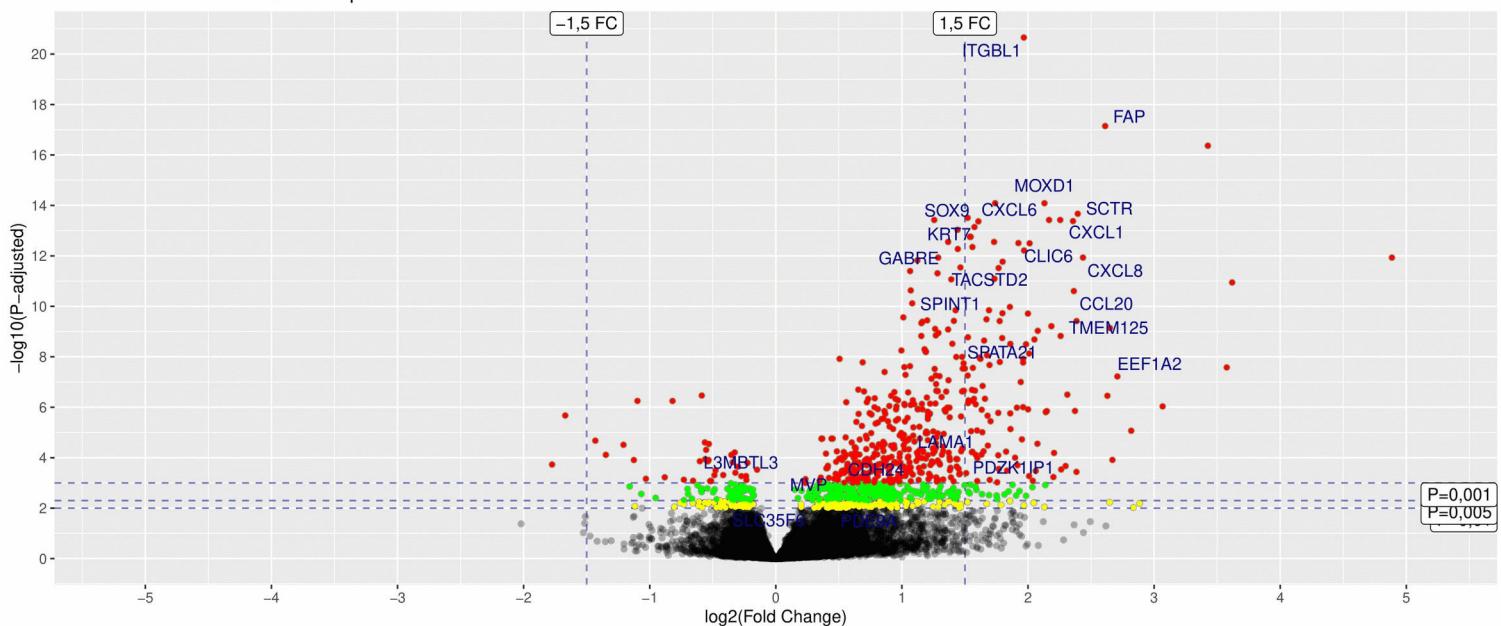
NAFLvsNASH-F4 DESeq2 Results



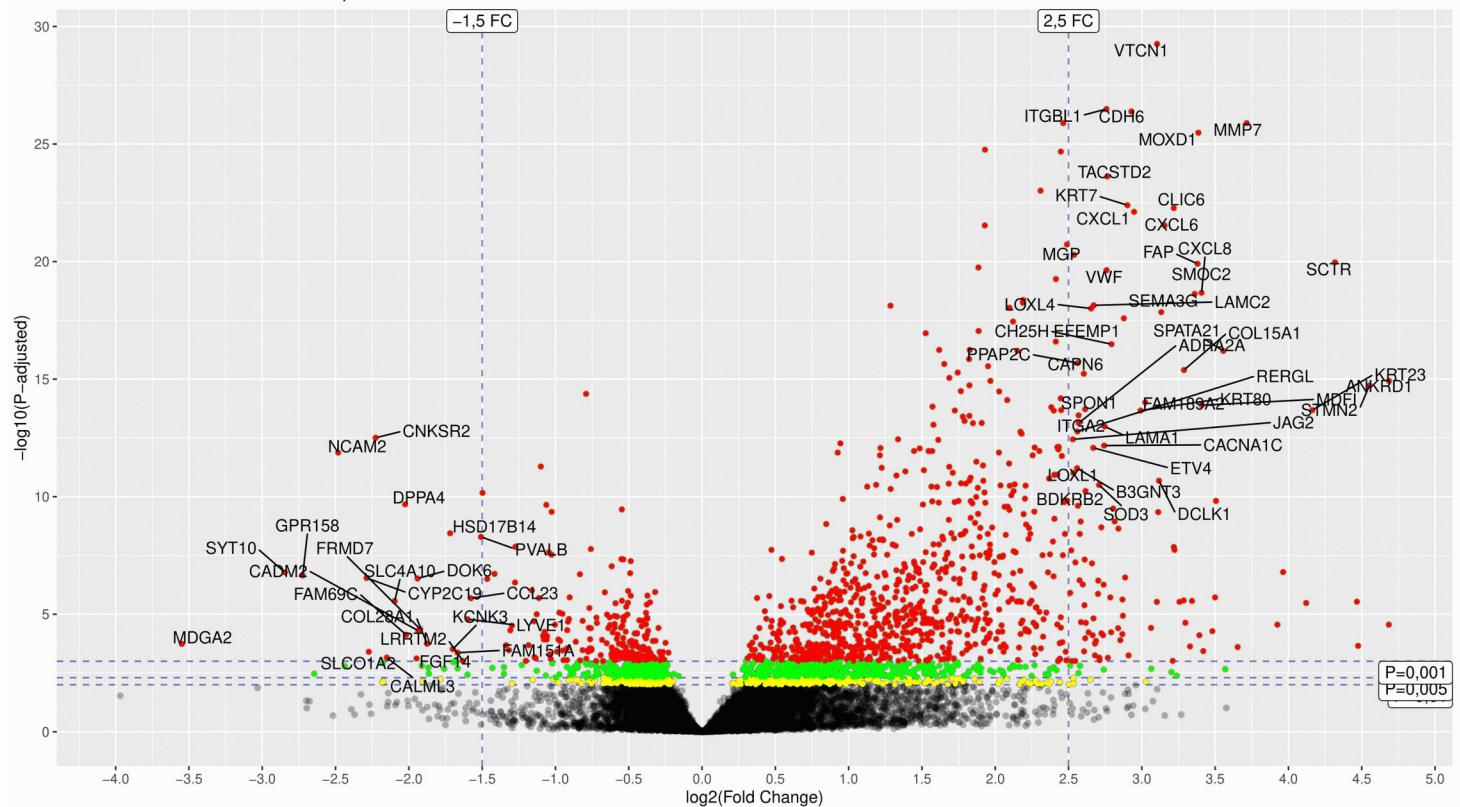
NASH-F0F1vsNASH-F2 DESeq2 Results



NASH-F0F1vsNASH-F3 DESeq2 Results



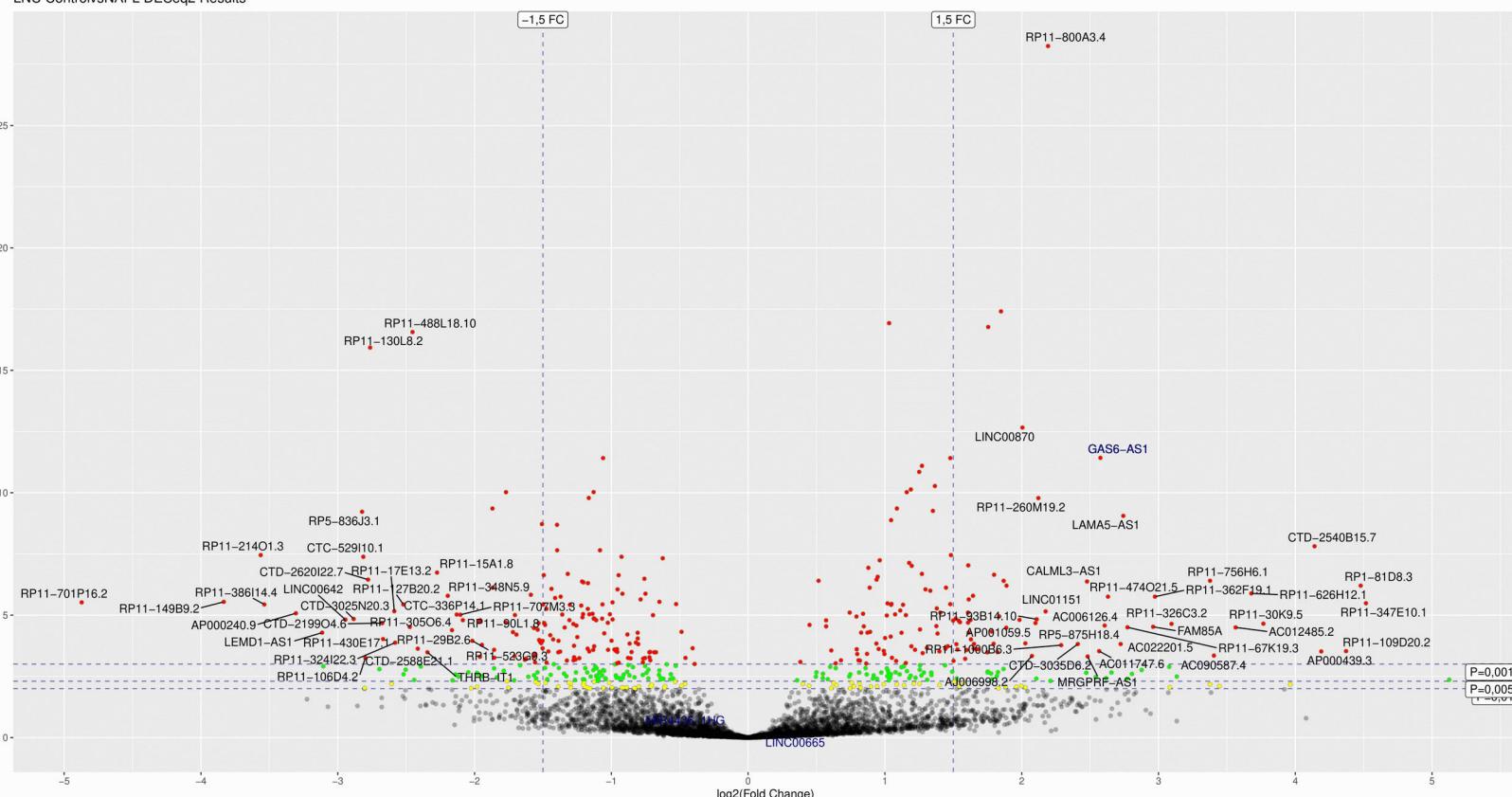
NASH-F0F1vsNASH-F4 DESeq2 Results



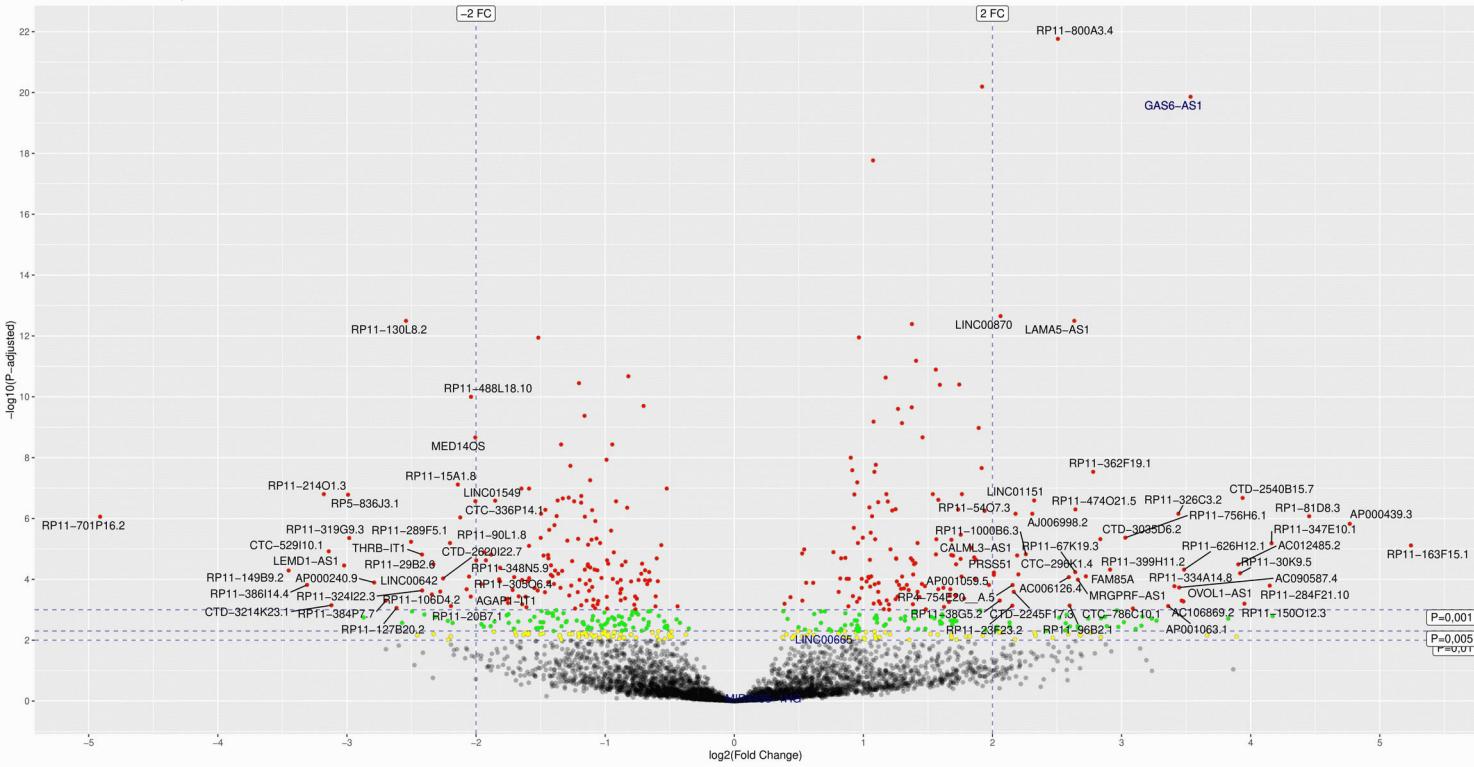
Παράρτημα 7.

Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των lncRNAs για το επίπεδο αναφοράς control, μετά από την αφαίρεση των N/A τιμών. Έγινε χρήση του αρνητικού λογαρίθμου του padj (-log10(padj)) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος (log2(FC)) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με p-value<0,01 (padj), με πράσινο τα γονίδια μεταξύ 0,01 < p-value < 0,005 και με κόκκινο τα γονίδια με p-value < 0,001. Αναγράφονται τα ονόματα (Gene ID) των γονιδίων με p<0,001 και με μεγάλη αλλαγή διπλώματος (πολύ αυξημένη ή πολύ μειωμένη έκφραση). Με μπλε χρώμα έχει τονιστεί το GAS6-AS1. Πάνω αριστερά αναφέρεται η σύγκριση η οποία απεικονίζεται. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.

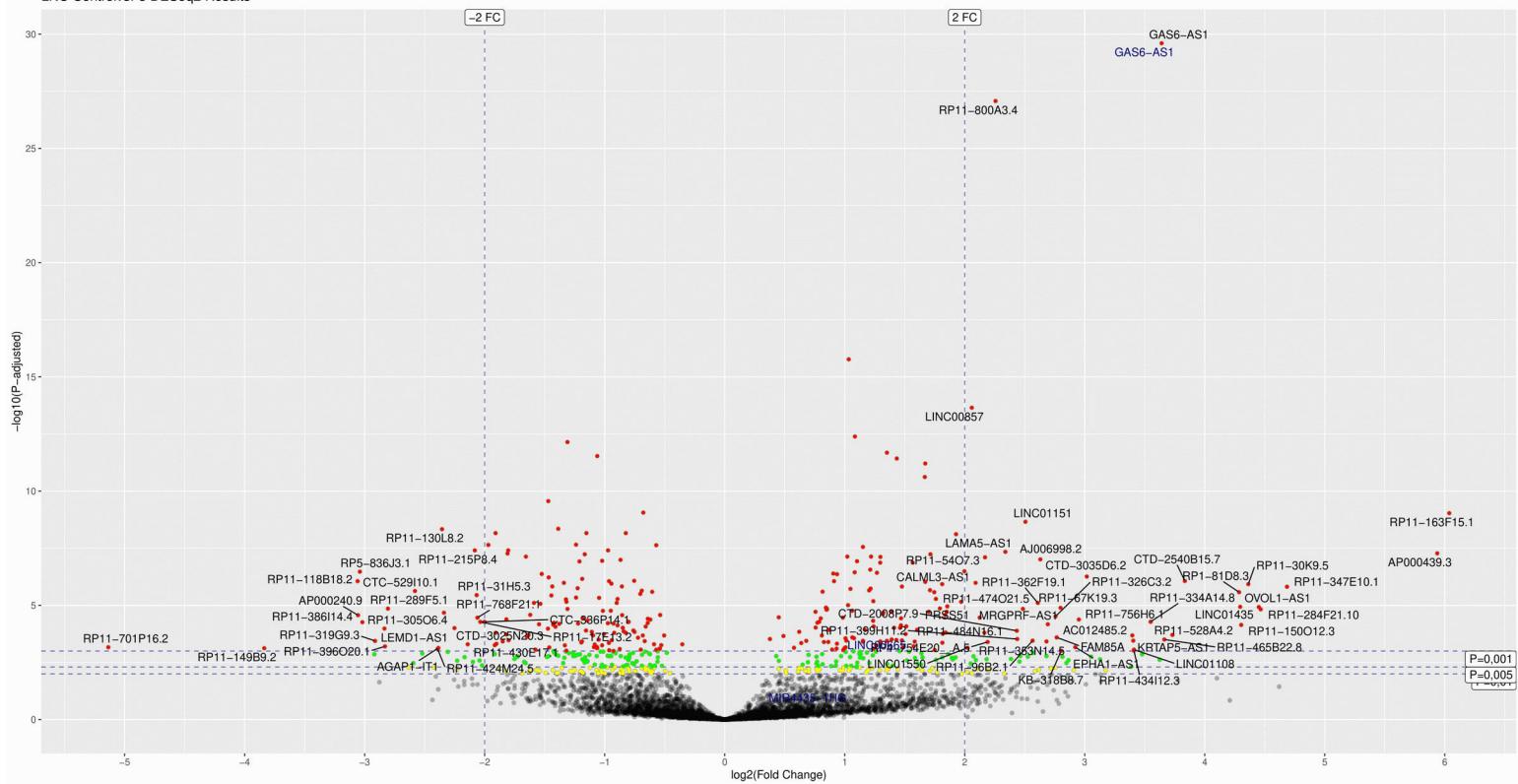
LNC ControlvsNAFL DESeq2 Results



LNC ControlsF2 DESeq2 Results



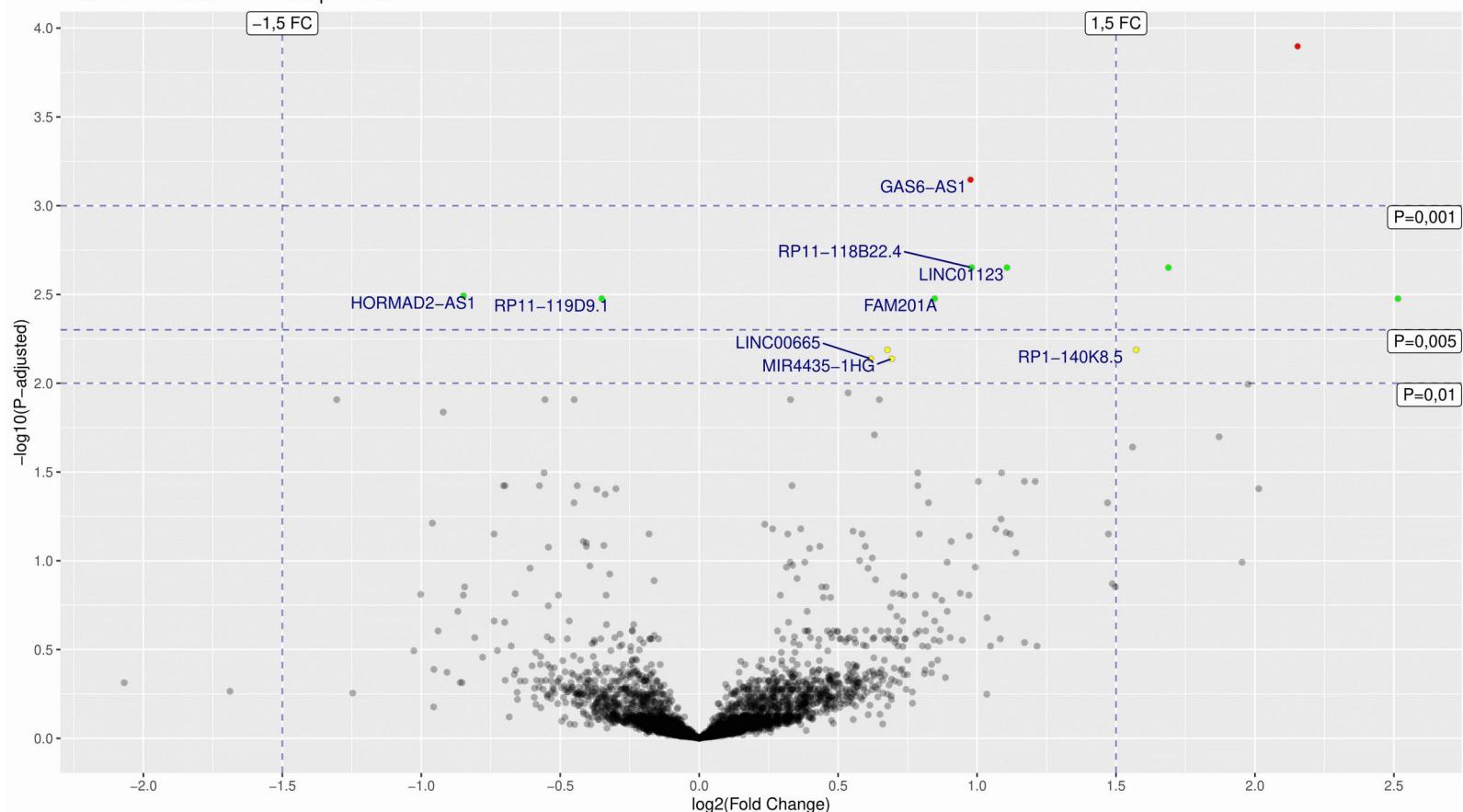
LNC ControlsF3 DESeq2 Results



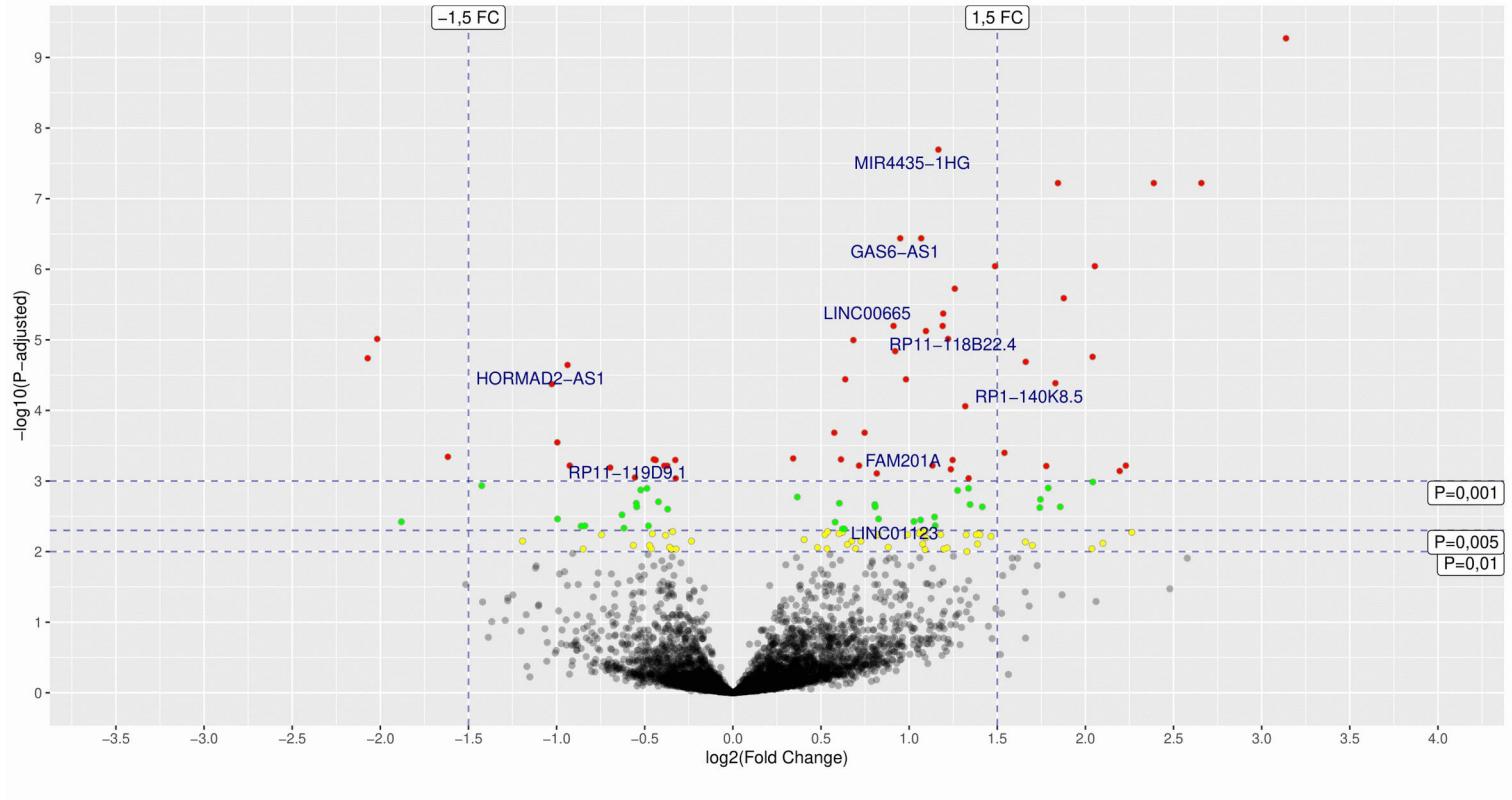
Παράτημα 8.

Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των lncRNAs για το επίπεδο αναφοράς NAFL, μετά από την αφαίρεση των N/A τιμών. Έγινε χρήση του αρνητικού λογαρίθμου του padj ($-\log_{10}(\text{padj})$) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος ($\log_2(\text{FC})$) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με $p\text{-value} < 0,01$ (padj), με πράσινο τα γονίδια μεταξύ $0,01 < p\text{-value} < 0,005$ και με κόκκινο τα γονίδια με $p\text{-value} < 0,001$. Αναγράφονται με μπλε τα 9 ονόματα (Gene ID) γονιδίων που είναι κοινά μεταξύ των 3 συγκρίσεων με επίπεδο αναφοράς NAFL (Εικόνα 15Α,Δ - NAFL-Baseline). Πάνω αριστερά αναφέρεται η σύγκριση η οποία απεικονίζεται. Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.

LNC NAFLvsNASHF2 DESeq2 Results



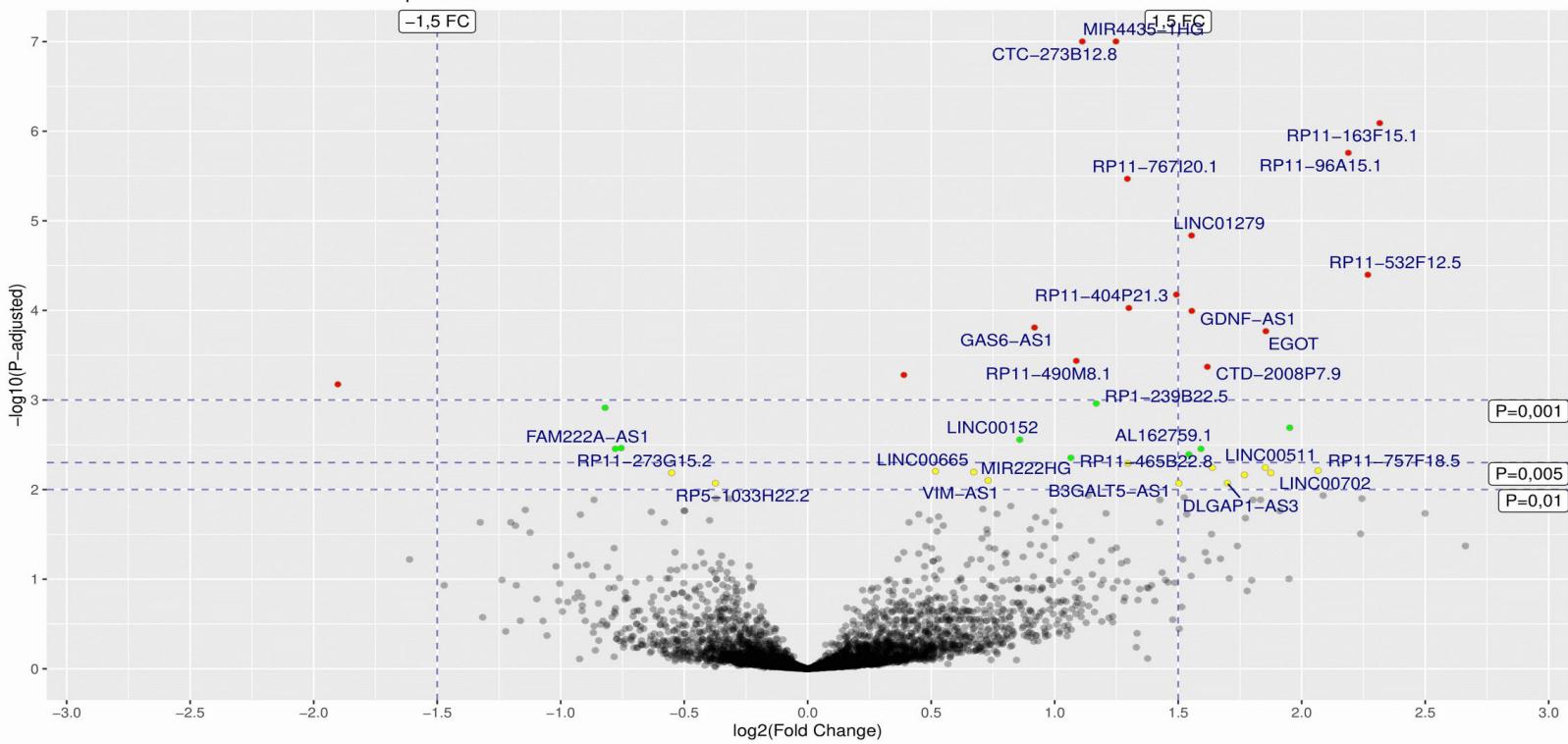
LNC NAFLvsNASHF3 DESeq2 Results



Παράρτημα 9.

Γράφημα τύπου ηφαιστείου της ανάλυσης διαφορικής έκφρασης των lncRNAs για NASH F0-F1 vs NASH F3, μετά από την αφαίρεση των N/A τιμών. Έγινε χρήση του αρνητικού λογαρίθμου του padj ($-\log_{10}(\text{padj})$) για τον άξονα x και του λογαρίθμου της αλλαγής διπλώματος ($\log_2(\text{FC})$) για τον άξονα y. Με μαύρο/γκρι απεικονίζονται τα μη-στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια. Με κίτρινο τα γονίδια με $p\text{-value} < 0,01$ (padj), με πράσινο τα γονίδια με $0,01 < p\text{-value} < 0,005$ και με κόκκινο τα γονίδια με $p\text{-value} < 0,001$. Αναγράφονται με μπλε τα 28 ονόματα (Gene ID) γονιδίων που είναι κοινά μεταξύ των 2 συγκρίσεων με επίπεδο αναφοράς NASH F0-F1 (Εικόνα 15Β,Δ – NASH F0-F1-Baseline). Οι εικόνες παράχθηκαν από τα πακέτα ggplot2 + ggrepel.

LNC NASH-F0F1vsNASH-F3 DESeq2 Results



Παράτημα 10.

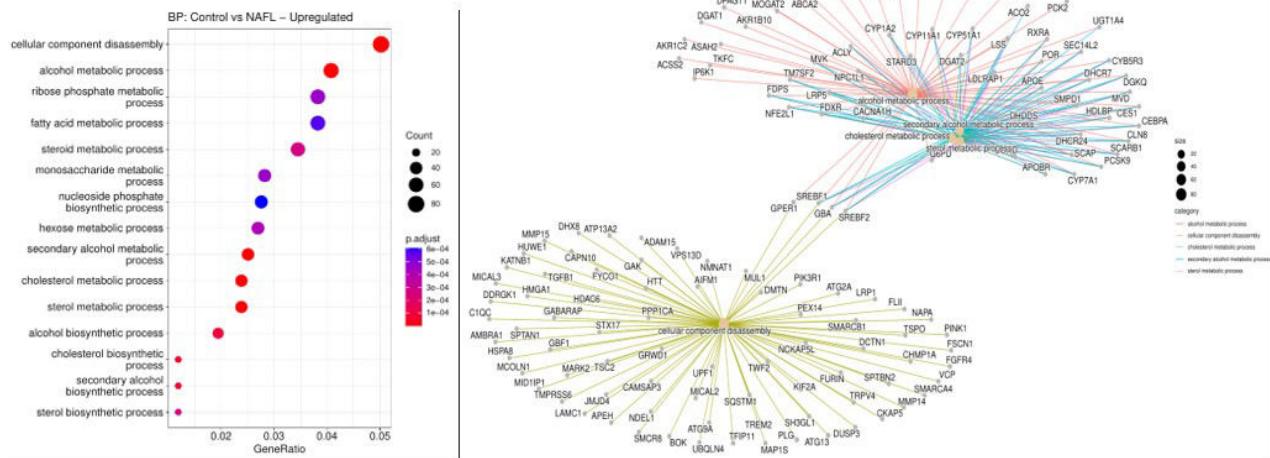
Αποτελέσματα της αυτοματοποιημένης μηχανικής μάθησης μέσω της πλατφόρμας JADBio. Επισυνάπτονται στο αρχείο “Παράτημα 10”.

Παράτημα 11.

Αποτελέσματα ανάλυσης οντολογίας γονιδίων (Gene Ontology, GO). Έγινε χρήση στατιστικά σημαντικών διαφορικά εκφρασμένων γονιδίων ($p<0,01$) με αλλαγή διπλώματος (Fold Change) $FC \geq 1,5$ (upregulated) από τις συγκρίσεις (1) Control vs NAFL, (2) Control vs NASH F0-F1, (3) NAFL vs NASH F4 και (4) NASH F0-F1 vs NASH F4. Οι αναλύσεις για (Α) τις βιολογικές διαδικασίες – BP, (Β) το κυτταρικό διαμέρισμα – CC και (Γ) τις μοριακές λειτουργίες – MF, έγιναν ξεχωριστά. Αριστερά βρίσκεται το διάγραμμα κουκκίδας και δεξιά το cnetplot της αντίστοιχης ανάλυσης όπου φαίνονται το ονόματα των γονιδίων που εμπλέκονται με το εκάστοτε επίπεδο ανάλυσης (BP, MF, CC). Για την ανάλυση έγινε χρήση $p<0,05$ στις παραμέτρους του πακέτου (DOSE). Το παρόν στα γραφήματα κουκκίδων αναφέρεται στη στατιστική σημαντικότητα που προβλέπει η ανάλυση. Στα cnetplot απεικονίζονται τα 5 πιο στατιστικά σημαντικά αποτελέσματα. Counts = Αριθμός γονιδίων που εμπλέκονται στο εκάστοτε επίπεδο ανάλυσης (BP, MF, CC). Οι εικόνες για τα διαγράμματα κουκκίδων & cnetplot (enrichplot) παράχθηκαν από το πακέτο “clusterProfiler”.

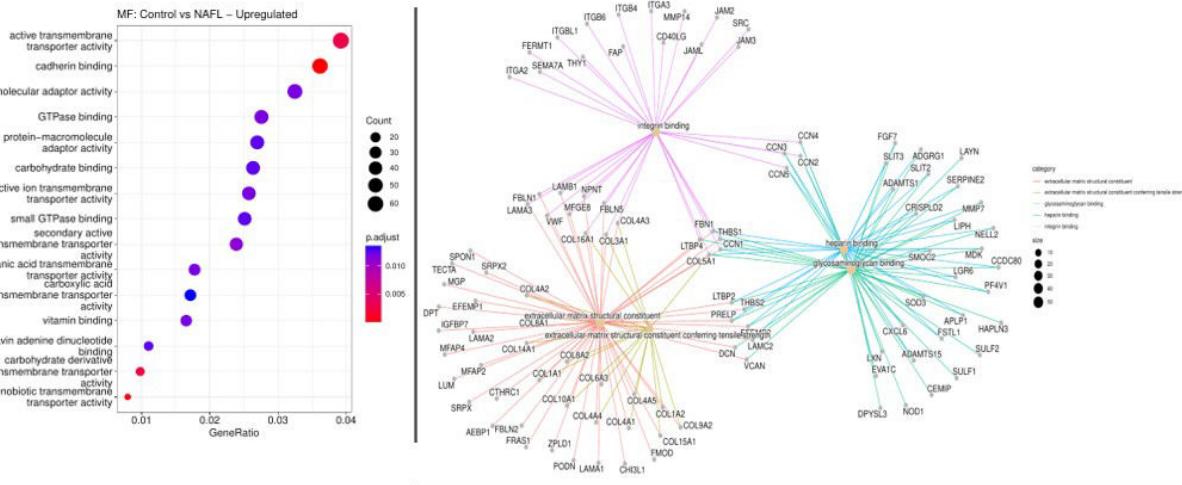
1A

GO BP

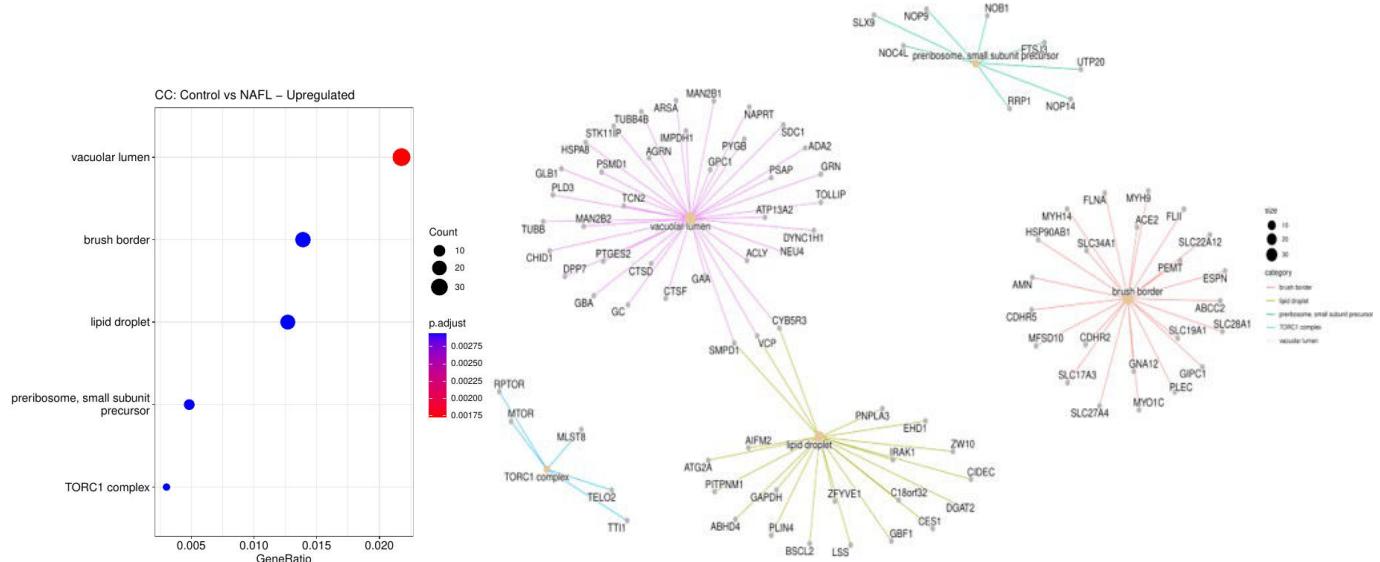


1B

GO MF

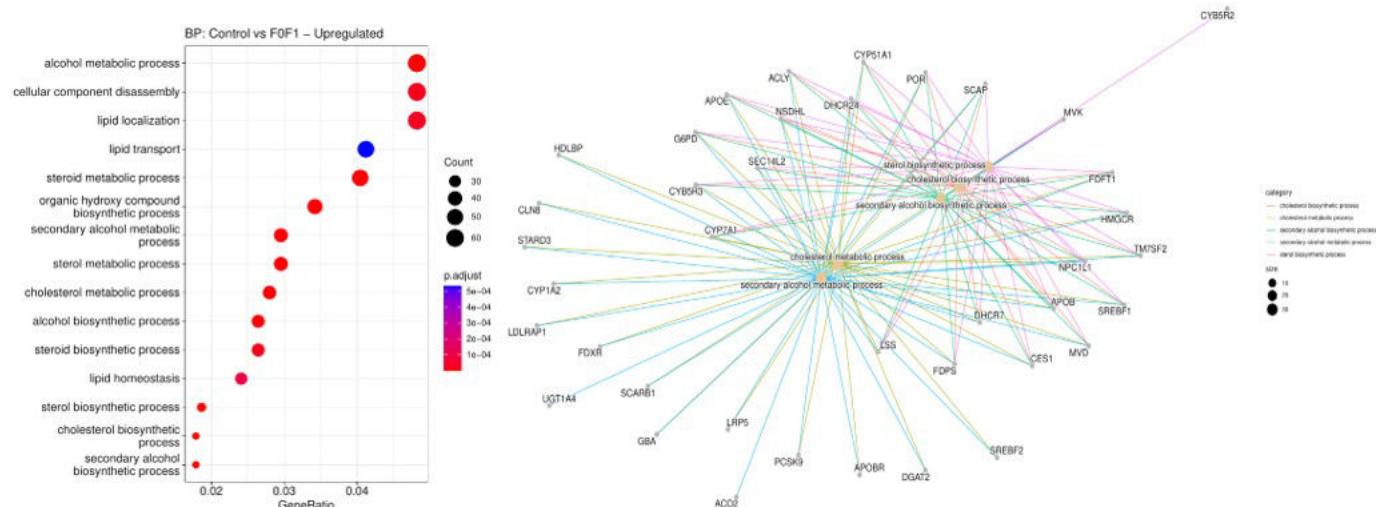


GO CC



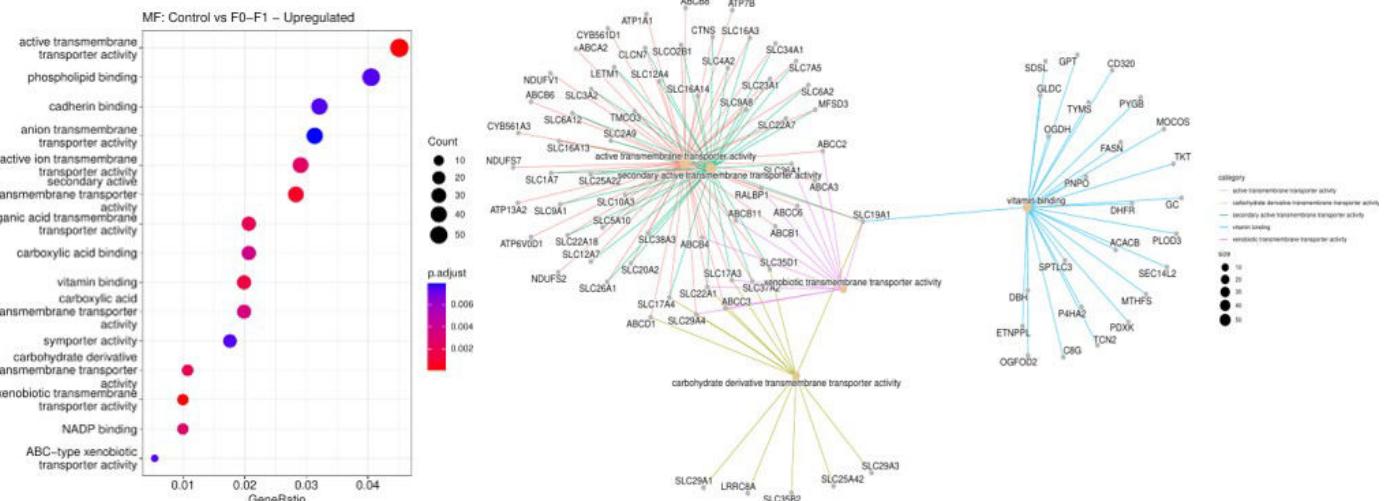
2A

GO BP



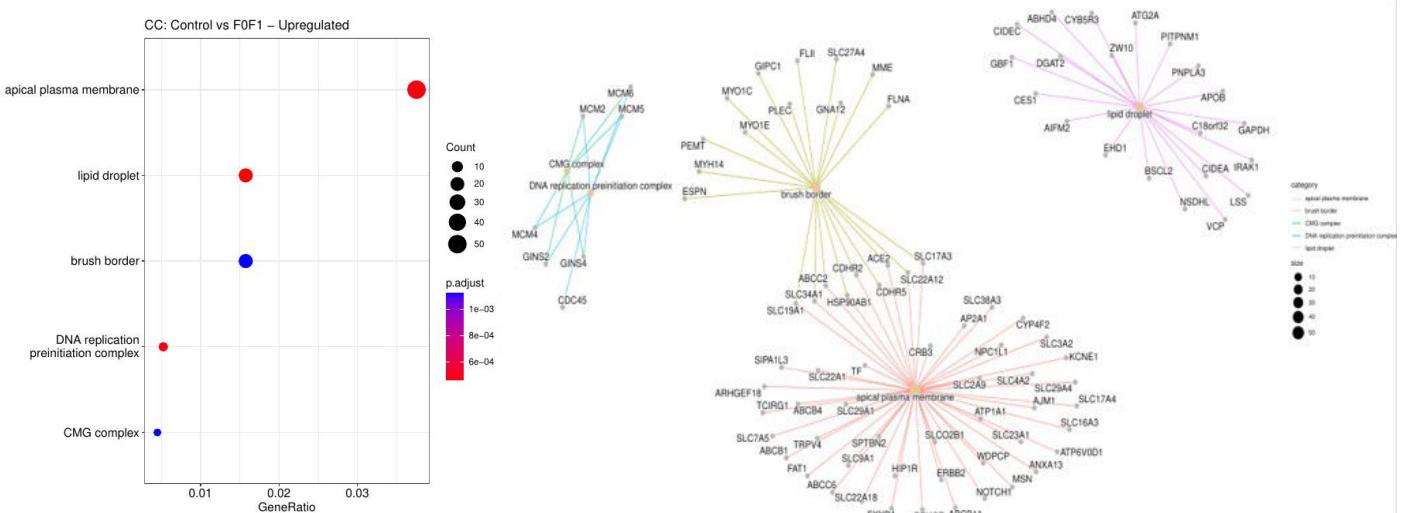
2B

GO MF



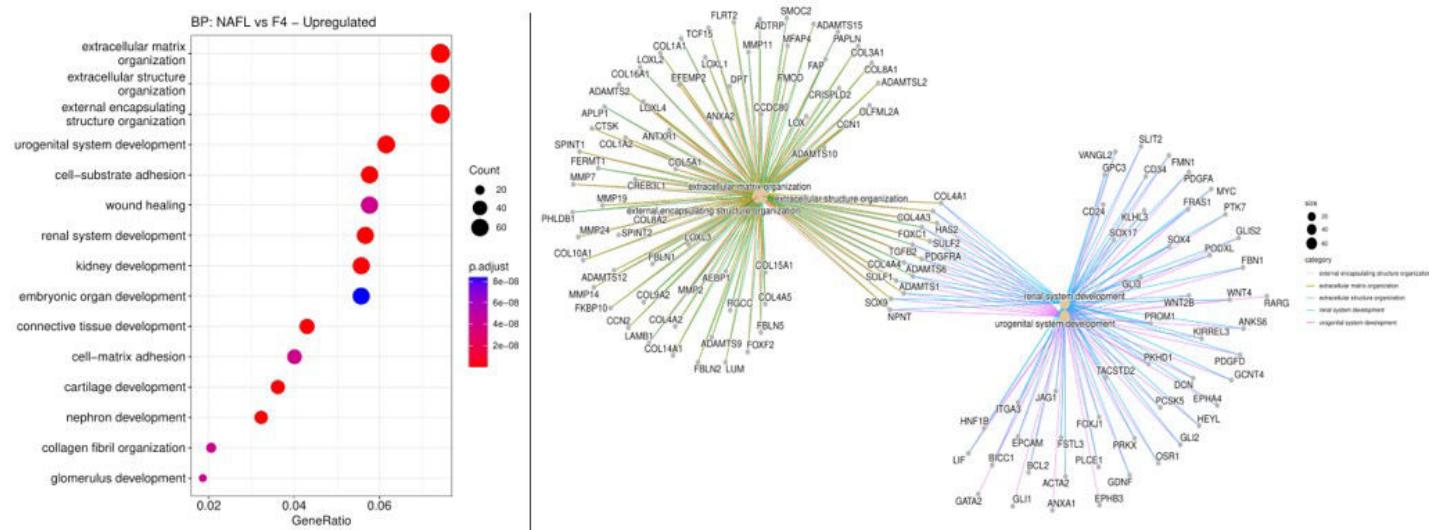
GO CC

2Γ

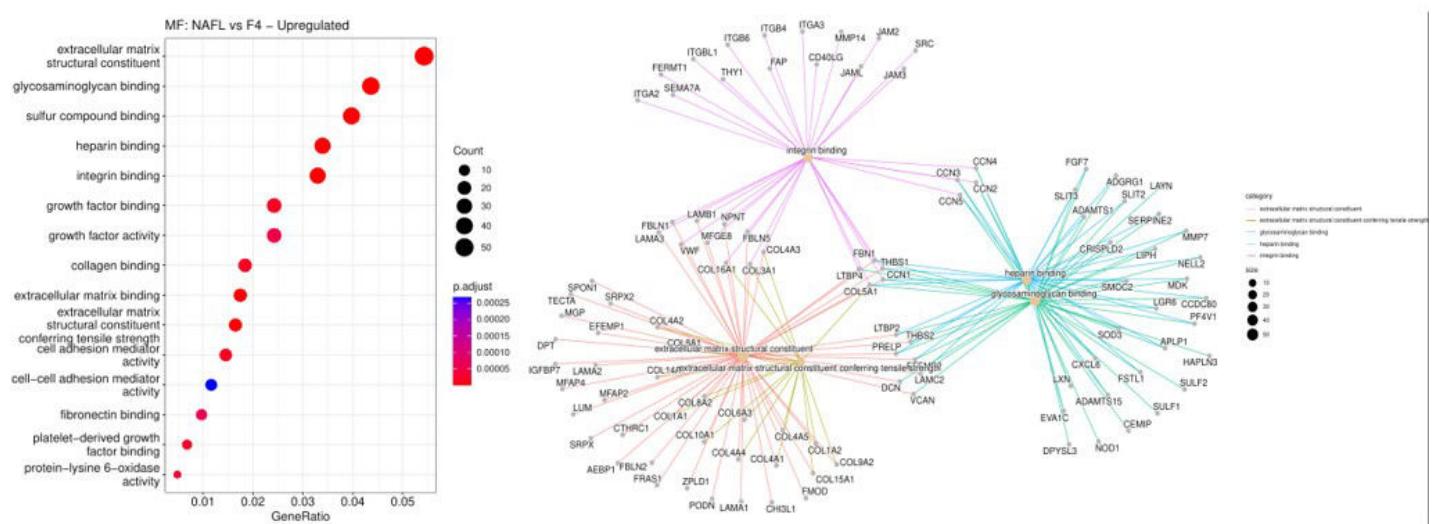


3A

GO BP

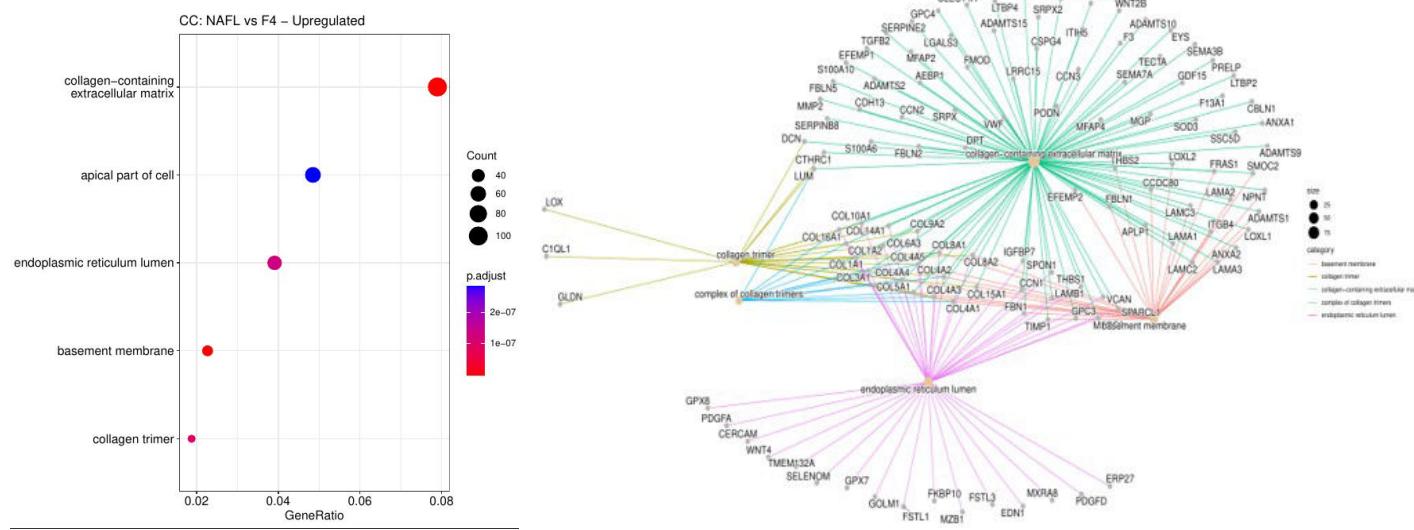


3B



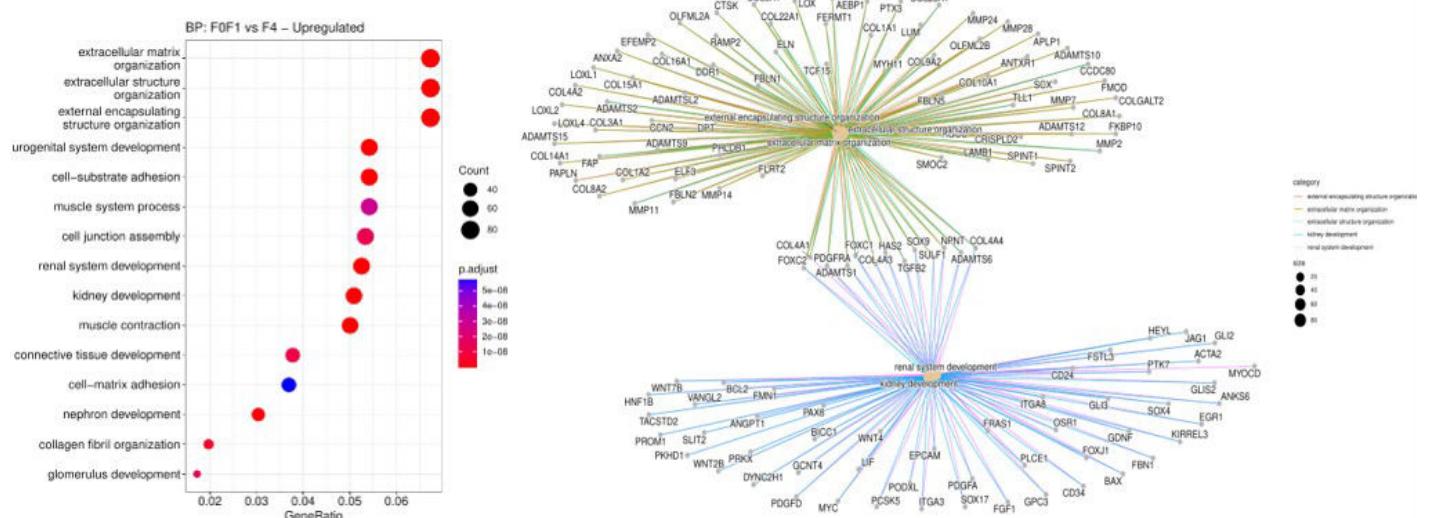
GO CC

3Γ



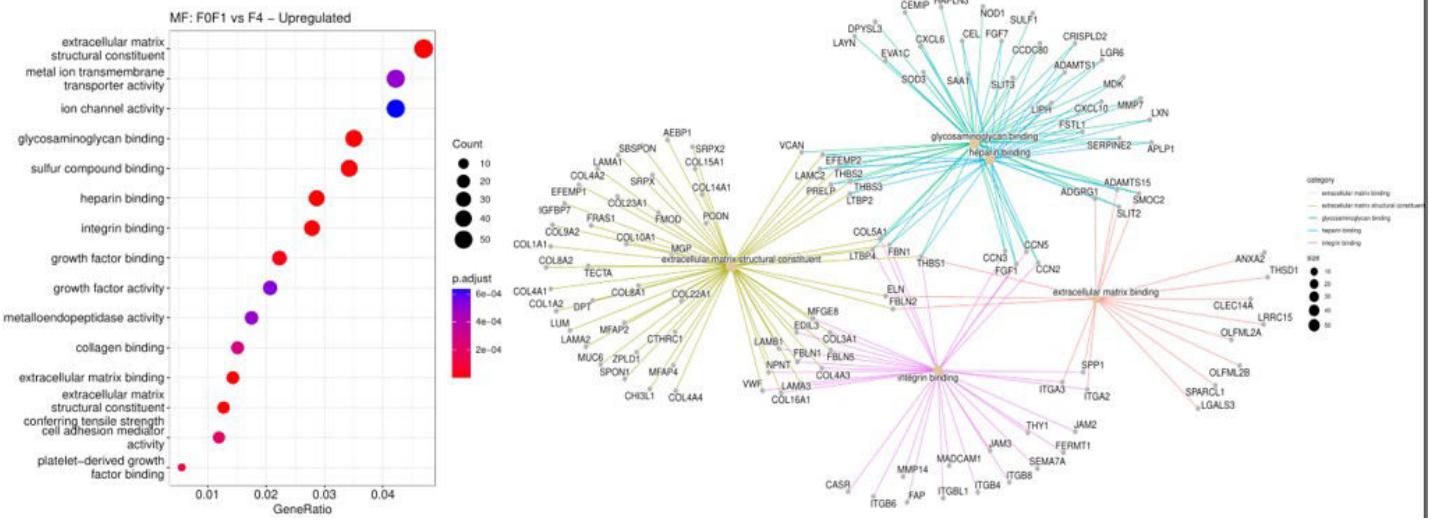
GO BP

4A



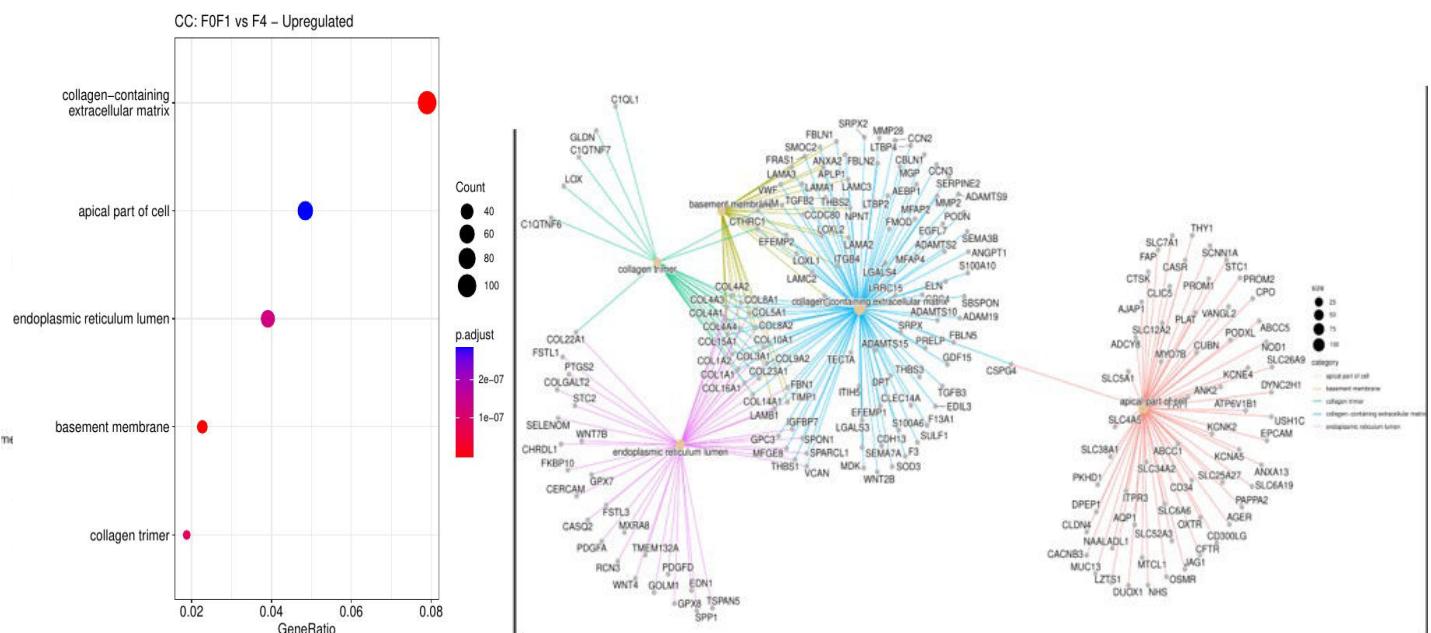
4B

GO MF



4C

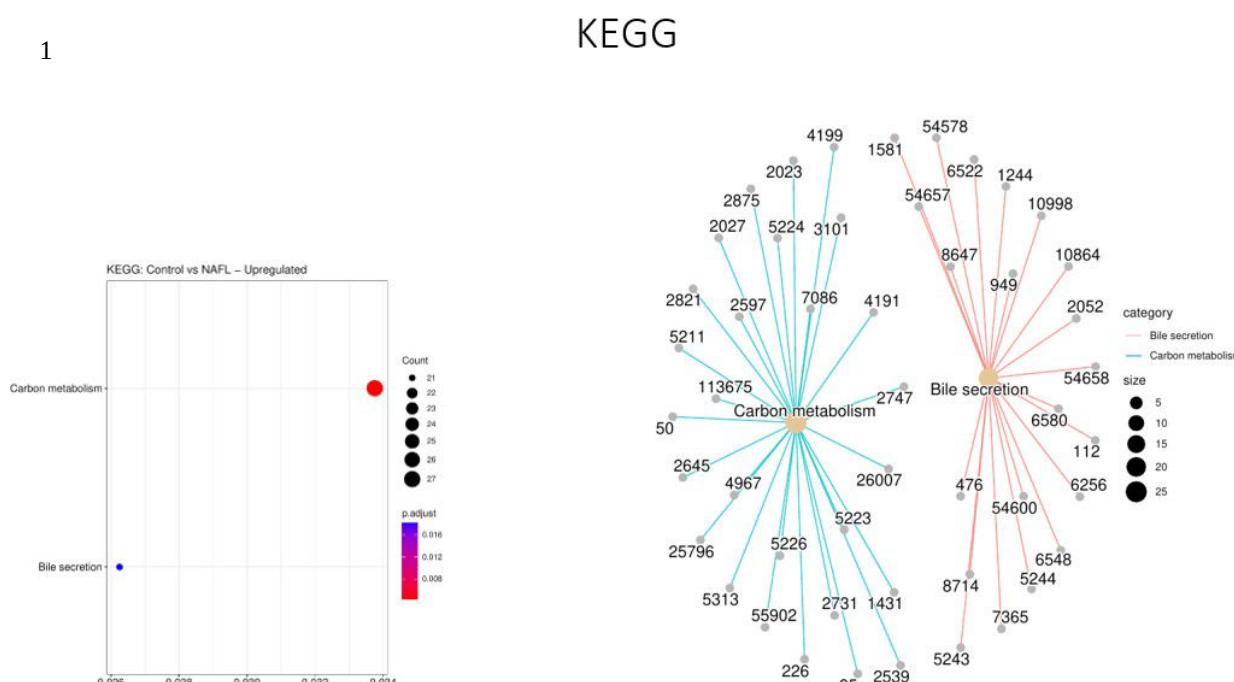
GO CC



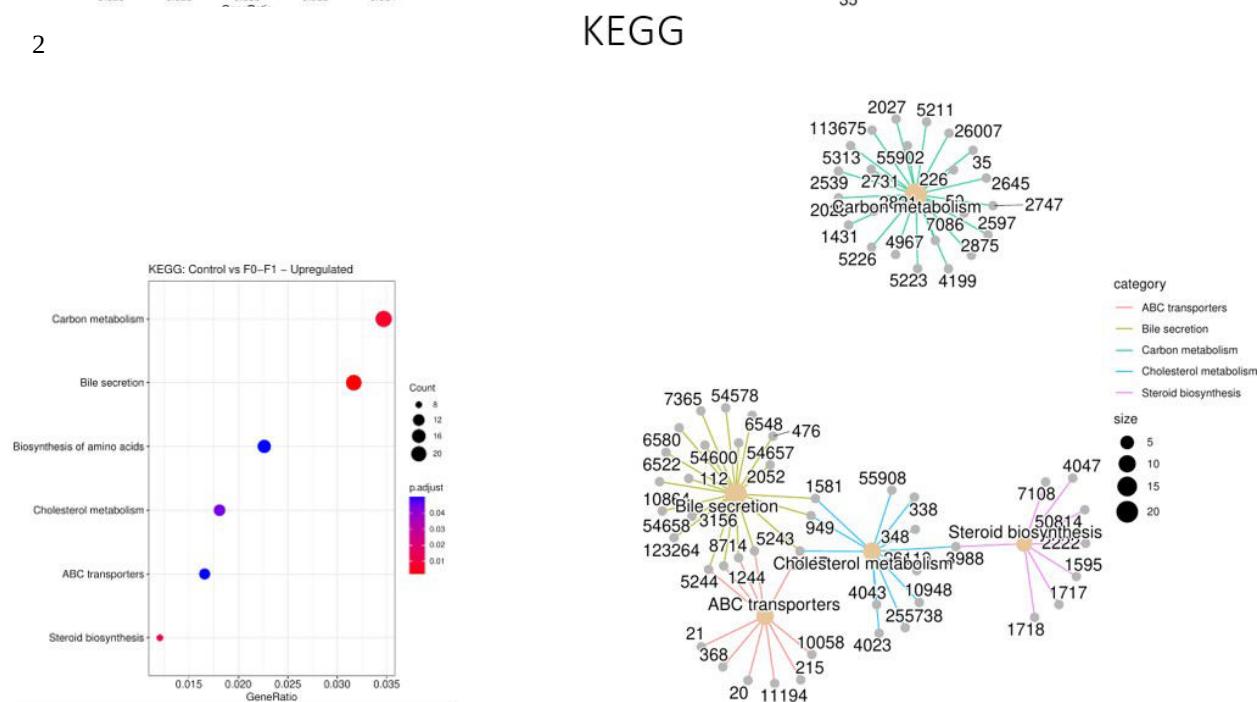
Παράτημα 12.

Αποτελέσματα ανάλυσης KEGG. Έγινε χρήση στατιστικά σημαντικών διαφορικά εκφρασμένων γονιδίων ($p<0,01$) με αλλαγή διπλώματος (Fold Change) $FC \geq 1,5$ (upregulated), από τις συγκρίσεις (1) Control vs NAFL, (2) Control vs NASH F0-F1, (3) NAFL vs NASH F4 και (4) NASH F0-F1 vs NASH F4. Αριστερά βρίσκεται το διάγραμμα κουκκίδας και δεξιά το cnetplot της αντίστοιχης ανάλυσης. Οι αριθμοί που αναγράφονται στο cnetplot (δεξιά) είναι τα ονόματα γονιδίων σε ENTREZID (ειδικό βάσης δεδομένων). Στα cnetplot απεικονίζονται τα 5 πιο στατιστικά σημαντικά αποτελέσματα. Counts = Αριθμός γονιδίων που εμπλέκονται. Οι εικόνες για τα διαγράμματα κουκκίδων & cnetplot (enrichplot) παράχθηκαν από το πακέτο “clusterProfiler”.

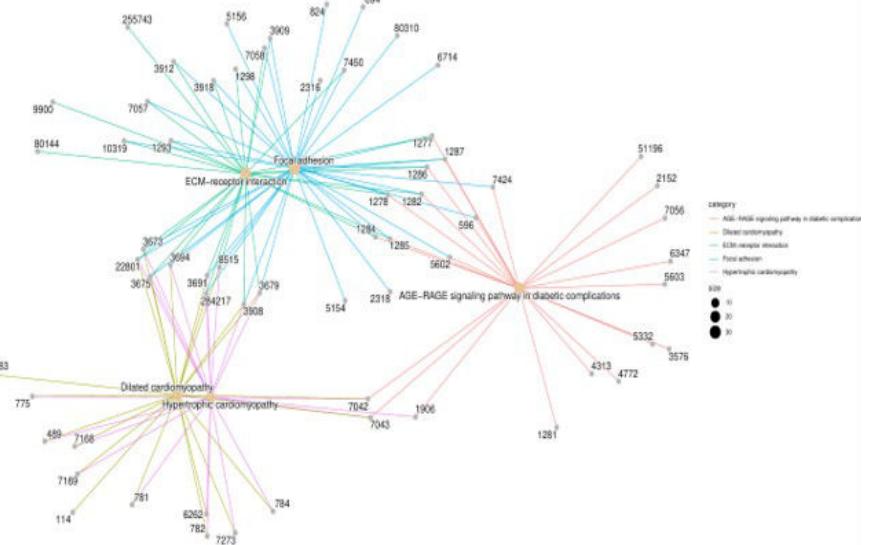
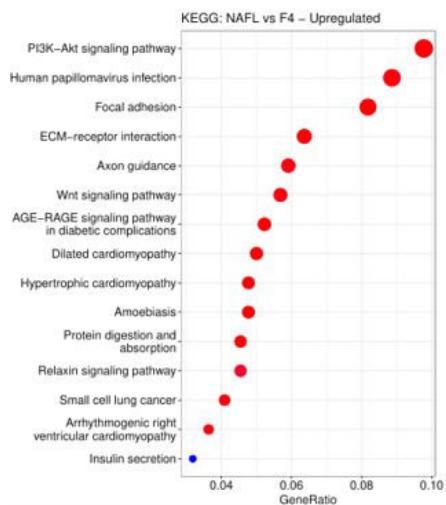
1



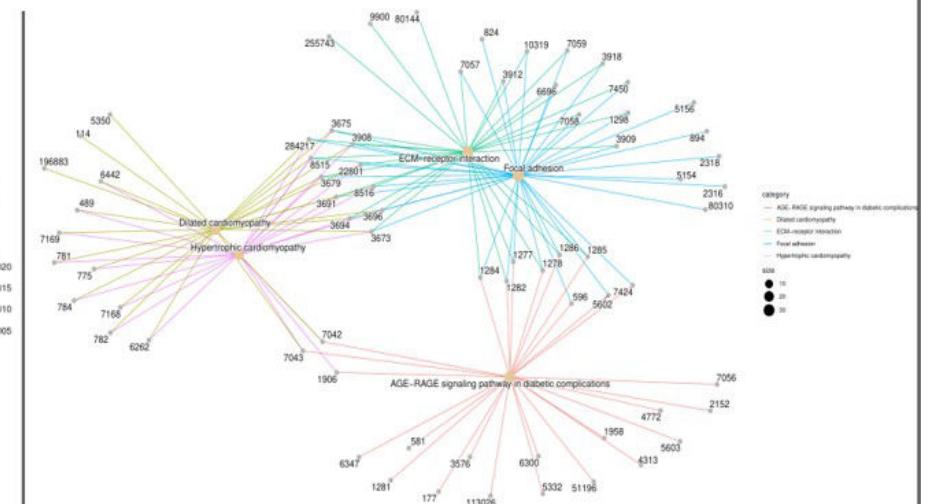
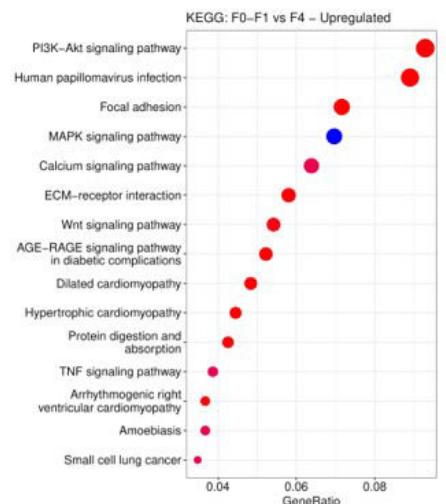
2



KEGG



KEGG



Παράρτημα 13.

Στο φάκελο Παράρτημα 13 βρίσκονται τα αρχεία με τις εντολές που χρησιμοποιήθηκαν για την πτυχιακή εργασία (PCA, DESeq2, Volcano plots-Scatterplots, Venn diagrams, AML).

Επιπλέον τα αρχεία/δεδομένα για την αναπαραγωγή της ανάλυσης βρίσκονται στο URL:
<https://www.dropbox.com/sh/6jia8hl4tcapufc/AACOs0KSkmaY9qxFjhZaZV0a?dl=0>