



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Μηχανολόγων Μηχανικών
Τομέας Μηχανολογικών Κατασκευών
και Αυτομάτου Ελέγχου
Εργαστήριο Εμβιομηχανικής και
Συστημικής Βιολογίας

Διπλωματική Εργασία

Υπολογιστική ανάλυση βιολογικών δικτύων για επαναστόχευση φαρμάκων στη Μη-Αλκοολική Λιπώδη Νόσο του Ήπατος

Ναυσικά Χαλά

Επιβλέπων:
Επ. Καθηγητής ΕΜΠ
Λεωνίδας Αλεξόπουλος

Οκτώβριος 2017

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κύριο Λεωνίδα Αλεξόπουλο, Επίκουρο Καθηγητή του Ε.Μ.Π., για την καθοδήγηση και την υποστήριξή του, καθώς και για την ευκαιρία που μου έδωσε να γίνω μέλος του εργαστηρίου «Εμβιομηχανικής και Συστημικής Βιολογίας». Επίσης θα ήθελα να ευχαριστήσω ιδιαιτέρως τη Δανάη Σ. Ζαρέιφη, μεταπτυχιακή φοιτήτρια για τη βοήθεια και την υπομονή της καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου. Τέλος, θα ήθελα να ευχαριστήσω τον Asier Antoranz, υποψήφιο διδάκτωρ Ε.Μ.Π. για τη βοήθειά του στο υπολογιστικό κομμάτι.

Περίληψη

Τα τελευταία χρόνια αναπτύσσονται υπολογιστικά μοντέλα τα οποία περιγράφουν διάφορες βιολογικές διεργασίες. Με αυτά τα μοντέλα ασχολείται ο κλάδος της συστημικής βιολογίας και δίνει νέες λύσεις στην διάγνωση, πρόγνωση και θεραπεία ασθενειών. Πεδίο εφαρμογής αποτελεί η μελέτη των μηχανισμών δράσης των φαρμάκων και ο πιθανός επαναπροσδιορισμός της χρήσης τους σε νέες ασθένειες (drug repositioning). Στην παρούσα εργασία γίνεται επαναστόχευση υπάρχοντων φαρμάκων για την αντιμετώπιση της μη αλκοολικής λιπώδης νόσου του ήπατος (Non-Alcoholic Fatty Liver Disease; NAFLD), η οποία είναι η πιο συχνή παθολογική κατάσταση του ήπατος, με συχνότητα εμφάνισης στο γενικό πληθυσμό έως 30% στις ανεπτυγμένες χώρες. Πολλές φαρμακολογικές ουσίες έχουν δοκιμαστεί για τη διαχείριση της νόσου, αλλά καμία θεραπεία δεν έχει εγκριθεί από τον US Food and Drug Administration (FDA) συγκεκριμένα για την αντιμετώπιση της νόσου.

Στην εργασία αυτή αναπτύσσονται μοντέλα ανάλυσης των βιολογικών διεργασιών που σχετίζονται με τη νόσο NAFLD. Αρχικά προσδιορίστηκαν οι μηχανισμοί της νόσου μέσω ανάλυσης δικτύων από γονιδιακά δεδομένα δειγμάτων βιοψίας ασθενών με NAFLD. Έπειτα, με τη βοήθεια του βιοπληροφορικού εργαλείου Connectivity Map ελέγχονται οι εμπορικά διαθέσιμες βιοδραστικές ουσίες και επιλέγονται αυτές οι οποίες επηρεάζουν τα ίδια μονοπάτια με τη νόσο. Τέλος, η αποτελεσματικότητα ορισμένων ουσιών δοκιμάζεται σε πειραματικά μοντέλα της νόσου *in vitro*.

Λέξεις κλειδιά: Συστημική βιολογία, βιολογικά δίκτυα, επαναστόχευση φαρμάκων, Μη-Αλκοολική Λιπώδης Νόσος του Ήπατος, R Bioconductor, ανάλυση μονοπατιών, Connectivity Map, *in vitro* πειράματα.

Abstract

There is an increased development of computational models over the last years in order to describe various biological processes. Systems biology uses these models as powerful tools in diagnosis, prognosis and treatment of diseases. These methods are also employed for the study of pharmacological agents mechanisms and drug repositioning. The aim of this project is to develop a new platform for drug repositioning in Non-Alcoholic Fatty Liver Disease (NAFLD), which is the most common liver disease with a prevalence of more than 30% in developed countries. Many pharmacological agents have been tested for the management of the disease, but none is approved from the USA Food and Drug Administration (FDA) so far.

In this study, models for the analysis of biological networks related with NAFLD are developed. NAFLD mechanisms were identified through pathway analysis from gene expression data sets of biopsy proven patients. Next, the commercially available pharmacological compounds are screened through the bioinformatics tool Connectivity Map in order to choose those that affect the same pathways as the disease. Finally, the efficacy of these compounds is being tested on *in vitro* models of the disease.

Key words: Systems biology, biological networks, drug repositioning, Non-Alcoholic Fatty Liver Disease, R Bioconductor, pathway analysis, Connectivity Map, *in vitro* experiments.

Περιεχόμενα

Κεφάλαιο 1 : Εισαγωγή	1
1.1. Συστημική βιολογία.....	1
1.2. Βιολογικά δίκτυα	2
1.2.1. Ρυθμιστικά δίκτυα.....	2
1.2.2. Μεταβολικά δίκτυα.....	3
1.2.3. Σηματοδοτικά δίκτυα.....	4
1.2.4. Δίκτυα πρωτεϊνικών αλληλεπιδράσεων	5
1.3. Pathway Analysis	5
1.4. Επαναστόχευση φαρμάκων.....	7
1.5. Εφαρμογή : Μη-Αλκοολική Λιπώδης Νόσος του Ήπατος.....	9
1.6. Σκοπός	11
Κεφάλαιο 2 : Μέθοδοι και Υλικά	13
2.1. Υπολογιστική Ανάλυση.....	13
2.1.1. Gene Expression Omnibus.....	13
2.1.2. Διάγραμμα κρατήρα ηφαιστείου (Volcano plot).....	13
2.1.3. Γονιδιακή ανάλυση (Gene Level Statistics)	15
2.1.4. Ανάλυση μονοπατιών (Gene Set Analysis)	16
2.1.5. Drugbank.....	18
2.1.6. Connectivity Map.....	18
2.1.7. ToxDB	20
2.1.8. LiverTox.....	20
2.1.9. Μεθοδολογία	20
2.2. Πειραματική Επαλήθευση.....	24
2.2.1. Cell Culture.....	24
2.2.2. Επαγωγή NAFLD	24
2.2.3. Πειράματα προκαταρκτικής αξιολόγησης βιοδραστικών ουσιών για τη θεραπεία της νόσου	25
2.2.4. Μέθοδος Resazurin για τη μέτρηση κυτταρικής βιωσιμότητας..	25
2.2.5. High Content Screening των σταγονιδίων λιπιδίων	26
Κεφάλαιο 3 : Αποτελέσματα	29
3.1. Υπολογιστική Ανάλυση.....	29
3.1.1. Μικροσυστοιχίες από τη GEO	29
3.1.2. Επιλογή σετ δεδομένων με τη χρήση Volcano plot.....	29
3.1.3. Ανάλυση μικροσυστοιχιών σε επίπεδο γονιδίων (GLS).....	31
3.1.4. Ανάλυση μικροσυστοιχιών σε επίπεδο μονοπατιών (GSA).....	34
3.1.5. Προσδιορισμός των μονοπατιών που επηρεάζουν οι στεατογενικές φαρμακολογικές ουσίες.....	39

3.1.6. Εύρεση βιοδραστικών ουσιών για αντιστροφή της νόσου με τη χρήση του cMap	40
3.1.7. Τελική επιλογή βιοδραστικών ουσιών.....	43
3.2. Πειραματική Επαλήθευση.....	44
3.2.1. Καμπύλες δόσης-επιβίωσης.....	44
3.2.2. High Content Screening	46
Κεφάλαιο 4 : Συμπεράσματα	73
Μελλοντική εργασία.....	77
Βιβλιογραφία.....	79
Παράρτημα Α.....	85
Παράρτημα Β.....	Error! Bookmark not defined.
Κύριο κομμάτι κώδικα.....	Error! Bookmark not defined.
Συνάρτηση pianoPathWays.....	Error! Bookmark not defined.
Συνάρτηση pianoAnalysis.....	Error! Bookmark not defined.
Συνάρτηση plotGenes	Error! Bookmark not defined.

Κεφάλαιο 1

Εισαγωγή

1.1. Συστημική βιολογία

Η συστημική βιολογία είναι κλάδος της βιολογίας που αναπτύσσεται συνεχώς. Η γενική ιδέα είναι ότι τα διάφορα βιολογικά φαινόμενα ερμηνεύονται με βάση τη θεωρία των συστημάτων. Η θεωρία των συστημάτων χρησιμοποιείται σε πολλές επιστήμες και έχει ισχυρό μαθηματικό υπόβαθρο. Η συνεργασία επιστημόνων από πολλούς και διαφορετικούς τομείς, όπως η βιολογία, η φυσική, τα μαθηματικά και η πληροφορική, δίνει τη δυνατότητα για μαθηματική μοντελοποίηση και ανάλυση οικοσυστημάτων, οργανισμών και κυττάρων. Με αυτόν τον τρόπο μπορούν να συνδυαστούν οι διαφορετικές προσεγγίσεις κάθε επιστημονικού κλάδου για τη δημιουργία ενός νέου τρόπου αντίληψης της λειτουργίας των βιολογικών συστημάτων. Διαφορετικά, η κατανόηση αυτών είναι πολλές φορές δύσκολη και αβέβαιη όταν παρατηρείται το πρόβλημα μόνο από μια σκοπιά.

Τα τελευταία χρόνια έχουν αναπτυχθεί οι τεχνολογίες υψηλής απόδοσης (high-throughput), οι οποίες, σε αντίθεση με τις προηγούμενες, επιτρέπουν τη μέτρηση πολλαπλών στόχων σε πολλά δείγματα με υψηλή ταχύτητα, απόδοση, ακρίβεια και ευαισθησία. Τα αποτελέσματα που προκύπτουν από αυτές τις μετρήσεις, η συστημική βιολογία μπορεί να τα αξιοποιήσει χρησιμοποιώντας μαθηματικές θεωρίες και μοντέλα. Με τον τρόπο αυτό ανακαλύπτονται νέες θεωρίες, οι οποίες μπορούν να εξηγήσουν τον τρόπο λειτουργίας και τη δομή του εκάστοτε συστήματος που μελετάται [1].

Αυτό που έχει σημασία είναι ότι ένα σύστημα αποτελείται όχι μόνο από τα μέρη του, αλλά και από τον τρόπο που αυτά συνδέονται. Χαρακτηριστικό παράδειγμα αποτελεί η μελέτη της γονιδιακής έκφρασης. Η ανεξάρτητη μελέτη ενός ή μερικών γονιδίων δεν παρέχει αρκετά δεδομένα για την κατάσταση ενός οργανισμού, ενώ η συνδυαστική μελέτη της έκφρασης μεγάλου πλήθους γονιδίων με τη φυσιολογία του οργανισμού οδηγεί σε πιο ασφαλή συμπεράσματα. Για παράδειγμα, το ανοσοποιητικό σύστημα ενός οργανισμού δεν μπορεί να περιγράψει ικανοποιητικά μόνο με τη μέτρηση της συγκέντρωσης ή της δραστηριότητας ορισμένων ανοσοσφαιρινών, καθώς η λειτουργία του δεν είναι αποτέλεσμα απλών μηχανισμών ή γονιδίων. Σύμφωνα με τη συστημική βιολογία, το ανοσοποιητικό σύστημα είναι το σύνολο των αλληλεπιδράσεων πολλών διαφορετικών γονιδίων, πρωτεϊνών, μηχανισμών και εξωτερικών

παραγόντων που έχουν ως στόχο να αντιμετωπίσουν ασθένειες και μολύνσεις. Επομένως, χρειάζονται οι πληροφορίες που είναι κρίσιμες για τη σωστή πρόβλεψη της αντίδρασης του οργανισμού στην εισαγωγή κάποιας ξένης ουσίας ή κάποιου ξένου οργανισμού μπορεί να είναι είτε στο γονιδίωμα, είτε στο σύνολο των πρωτεϊνών, είτε στη σύνθεση των παραπάνω [2].

Η συστημική βιολογία λοιπόν, μοντελοποιεί τη λειτουργία των βιολογικών οντοτήτων σε δίκτυα και σκοπός της είναι να καθορίσει τη δομή και τη δυναμική τους για να τα περιγράψει με το βέλτιστο τρόπο. Επομένως με τις κατάλληλες εισόδους στα κατάλληλα μοντέλα μπορεί να προβλεφθεί η συμπεριφορά του συστήματος με την πάροδο του χρόνου. Οι εφαρμογές της ποικίλουν, με ορισμένα παραδείγματα την ανάπτυξη βιοδεικτών για τη διάγνωση ασθενειών και την πρόβλεψη της ανταπόκρισης ή όχι ασθενών σε συγκεκριμένα φάρμακα.

1.2. Βιολογικά δίκτυα

Ως βιολογικά δίκτυα χαρακτηρίζονται όλα αυτά που σχετίζονται και σχηματίζουν δικτυακές κοινότητες είτε μεταξύ διαφορετικών έμβιων οργανισμών είτε μεταξύ λειτουργιών των οργανισμών αυτών [4]. Βιολογικά δίκτυα μπορούν να συναντηθούν σε όλα τα επίπεδα της ζωής, από το πιο μακροσκοπικό, όπως τα οικοσυστήματα, μέχρι το πιο μικροσκοπικό. Τα συστήματα στο μοριακό επίπεδο είναι ιδιαίτερα πολύπλοκα λόγω του πλήθους των αλληλεπιδράσεων. Ακόμα και ο πιο απλός οργανισμός χρειάζεται να ρυθμίσει χιλιάδες βιοχημικές διεργασίες για να εκφραστούν τα κατάλληλα γονίδια την κατάλληλη χρονική στιγμή, για να συντεθούν τα απαραίτητα μεταβολικά προϊόντα και να εξαχθεί από αυτά η ενέργεια που χρειάζονται ώστε να ανταποκριθούν σε ερεθίσματα. Αυτό το τεράστιο πλήθος διεργασιών που αλληλοδιαπλέκονται δημιουργεί πυκνά και σύνθετα δίκτυα.

Στη συνέχεια θα γίνει μια περιληπτική αναφορά στα σημαντικότερα δίκτυα μοριακής βιολογίας. Το χαρακτηριστικό των δικτύων αυτών είναι ότι αποτελούνται από μόρια, γονίδια, πρωτεΐνες και μεταβολίτες και από την αλληλεπίδραση μεταξύ αυτών.

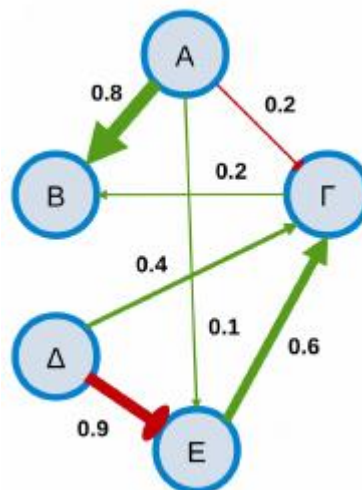
1.2.1. Ρυθμιστικά δίκτυα

Ρυθμιστικά δίκτυα (regulatory networks) ονομάζονται εκείνα που ορίζουν το πλέγμα των λειτουργικών αλληλεπιδράσεων της ρύθμισης της έκφρασης μεταξύ γονιδίων. Τα δίκτυα αυτά αποτελούνται από γονίδια ή εναλλακτικά από τις πρωτεΐνες που αυτά κωδικοποιούν, ενώ οι αλληλεπιδράσεις τους αντιπροσωπεύουν τις σχέσεις ρύθμισης μεταξύ τους [5]. Είναι γνωστό ότι για να ρυθμιστεί το επίπεδο της γονιδιακής έκφρασης, ρυθμίζεται η μεταγραφή των πρωτεϊνών, οι οποίες ονομάζονται και μεταγραφικοί παράγοντες (transcription factors). Κάθε μεταγραφικός παράγοντας έχει την ικανότητα να εντοπίσει και να προσδεθεί σε συγκεκριμένες σύντομες αλληλουχίες DNA με σκοπό να ρυθμίσει θετικά ή αρνητικά τα επίπεδα μεταγραφής των γονιδίων. Αξίζει να αναφερθεί ότι ένας μεταγραφικός παράγοντας μπορεί να ρυθμίζει πάνω από ένα γονίδιο,

ενώ ένα γονίδιο μπορεί να ρυθμίζεται από έναν ή περισσότερους μεταγραφικούς παράγοντες (έκφραση γονιδίων).

Τα δίκτυα αυτά έχουν ως κόμβους γονίδια ή μεταγραφικούς παράγοντες, ενώ οι ακμές αναπαριστούν τη μεταξύ τους αλληλεπίδραση. Είναι κατευθυνόμενα δίκτυα και μπορεί να δείχνουν το είδος της ρύθμισης, δηλαδή αν πρόκειται για ενεργοποίηση ή για καταστολή. Επίσης μπορεί να λαμβάνεται υπόψιν και ο βαθμός στον οποίο η κάθε ρύθμιση διαμορφώνει την έκφραση του κάθε στοιχείου. Στην Εικόνα 1.1. παρουσιάζεται ένα παράδειγμα.

Τα ρυθμιστικά δίκτυα είναι πολύ σημαντικά καθώς παρουσιάζουν το επίπεδο της κυτταρικής λειτουργίας που αποτελεί τη βάση για την κυτταρική εξειδίκευση, την ανάπτυξη και την απόκριση των οργανισμών σε περιβαλλοντικά ερεθίσματα. Εξαιτίας αυτού, τα δίκτυα αυτά είναι ιδιαίτερα εύπλαστα και μετασχηματίζονται δυναμικά, μεταβάλλοντας τα στοιχεία τους καθώς και τις συνδέσεις μεταξύ τους και το είδος τους. Τέλος συνδέονται άμεσα με τα υπόλοιπα είδη βιολογικών δικτύων διότι αυτά καθορίζουν τα στοιχεία που υπάρχουν σε αυτά.

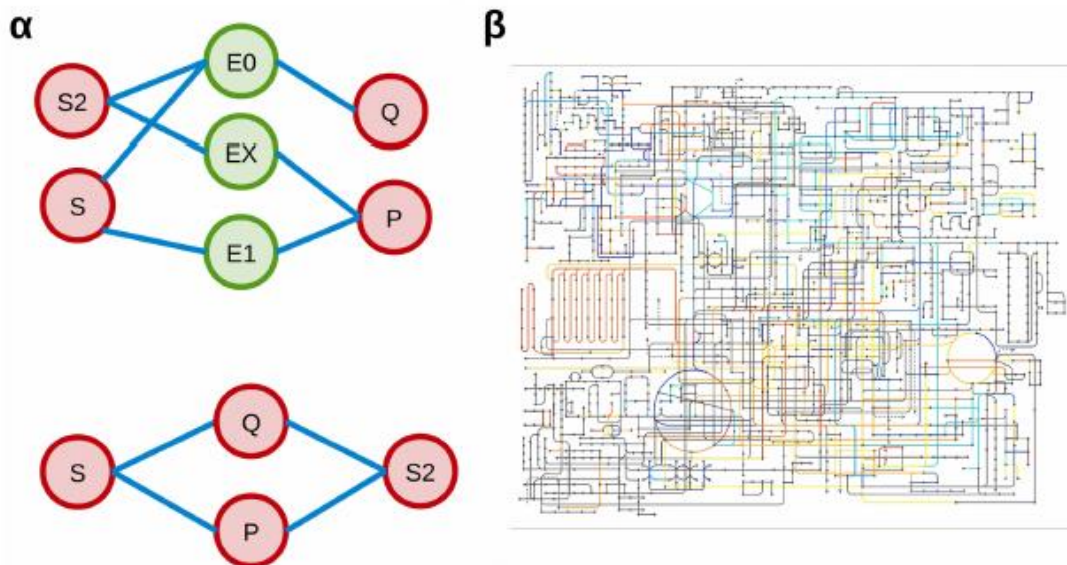


Εικόνα 1.1. Ρυθμιστικό δίκτυο (Nikolaou et al., 2015)

1.2.2. Μεταβολικά δίκτυα

Μεταβολικά δίκτυα ονομάζονται τα δίκτυα που αποτελούνται από μεταβολικά μονοπάτια, τα οποία είναι ακολουθίες βιοχημικών αντιδράσεων και σχετίζονται με την παραγωγή και την κατανάλωση ενέργειας. Τα στοιχεία τους ανήκουν σε δύο διαφορετικές κατηγορίες και ονομάζονται διμερή (bipartite). Αποτελούνται από τους μεταβολίτες, δηλαδή τα υποστρώματα και τα προϊόντα των αντιδράσεων αυτών και τα ένζυμα που τις καταλύουν. Οι κόμβοι είναι μεταβολίτες και ένζυμα και οι ακμές να δείχνουν τη σύνδεσή τους [6]. Στην Εικόνα 1.2. οι κόκκινοι κόμβοι είναι οι μεταβολίτες και οι πράσινοι κόμβοι τα ένζυμα. Το ένζυμο E1 δρα πάνω στο υπόστρωμα S και παράγει το μεταβολίτη P. Για τον διαχωρισμό των υποστρωμάτων από τα προϊόντα χρησιμοποιούνται κατευθυνόμενα δίκτυα. Ορισμένες φορές τα δίκτυα αυτά αναπαριστώνται έτσι

ώστε οι κόμβοι να είναι μόνο οι μεταβολίτες, οπότε οι συνδέσεις γίνονται μεταξύ μεταβολιτών που μοιράζονται ένα ένζυμο. Αυτό συμβαίνει για να απλοποιηθούν και να μειωθεί ο όγκος των κόμβων.



Εικόνα 1.2. α) Παράδειγμα μεταβολικού δικτύου β) Μεταβολικό δίκτυο που αντιστοιχεί σε μέρος των αντιδράσεων μεταβολισμού ενός ανθρώπινου κυττάρου (Εικόνα από J3D3, CC BY-SA 4.0 από Wikimedia Commons) (Nikolaou et al., 2015)

1.2.3. Σηματοδοτικά δίκτυα

Σηματοδοτικά δίκτυα ονομάζονται εκείνα που αποτελούνται από σηματοδοτικά μονοπάτια (signaling pathways), τα οποία περιγράφουν διεργασίες κυτταρικής σηματοδότησης (cell signaling) μέσω της ενεργοποίησης και καταστολής πρωτεϊνών και ενζύμων. Τα στοιχεία τους είναι σχεδόν αποκλειστικά πρωτεΐνες, ενώ οι μεταξύ τους σχέσεις αντιστοιχούν σε αντιδράσεις ενεργοποίησης ή καταστολής που αποτελούν στάδια στη μετάδοση ενός σήματος. Τα στοιχεία μπορούν να συμμετέχουν και να μεταδίδουν πληροφορίες σε πολλά μονοπάτια ταυτόχρονα, επομένως δημιουργούνται πολύπλοκα δίκτυα. Κάθε μονοπάτι έχει συνήθως για είσοδο ένα περιβαλλοντικό ερέθισμα από τον εξωκυττάριο χώρο στις συνδέσεις των υποδοχέων της κυτταρικής μεμβράνης (ligands), που ενεργοποιούν την πρωτεΐνη υποδοχέα (receptor) της κυτταρικής μεμβράνης. Ακολουθεί μια αλυσίδα μορίων και αντιδράσεων σηματοδότησης μέσα από το κυτταρόπλασμα με τελικούς αποδέκτες μεταγραφικούς παράγοντες εντός του πυρήνα [7]. Έτσι προκαλούνται αλλαγές στις ενδοκυτταρικές συγκεντρώσεις σηματοδοτικών μορίων, όπως ασβέστιο ή κυκλικά νουκλεοτίδια [8].

Τα σηματοδοτικά και τα μεταβολικά δίκτυα έχουν σημαντικές ομοιότητες, αλλά διαφέρουν ως προς το είδος της πληροφορίας που μεταδίδουν και τον στόχο που επιτελούν. Τα πρώτα σχετίζονται με τη μετάδοση πληροφορίας μέσω μοριακών σημάτων, ενώ είναι εξαιρετικά δυναμικά και οι χρόνοι

απόκρισης μπορεί να διαρκούν από χιλιοστά του δευτερολέπτου μέχρι ολόκληρα λεπτά ή και περισσότερο. Επίσης, ενώ είναι δύσκολο να μελετηθούν, είναι πολύ σημαντικά ιδιαίτερα για τομείς όπως η μοριακή ανοσολογία [9]. Τα δεύτερα είναι στατικά και σχετίζονται με την παραγωγή και κατανάλωση ενέργειας στο κύτταρο [10].

1.2.4. Δίκτυα πρωτεϊνικών αλληλεπιδράσεων

Με την αυστηρή έννοια του όρου όλα τα βιολογικά δίκτυα που αναφέραμε προηγουμένως είναι δίκτυα πρωτεϊνικών αλληλεπιδράσεων, καθώς περιέχουν πρωτεΐνες στα στοιχεία τους. Έτσι ονομάζουμε τα δίκτυα για το οποία δεν μπορούμε να εξακριβώσουμε τη φύση των αλληλεπιδράσεων ή όταν οι αλληλεπιδράσεις είναι ποικίλων τύπων. Συχνά πρόκειται για 'μικτά' δίκτυα με πολλές και διαφορετικές λειτουργικές σχέσεις μεταξύ των πρωτεϊνών [11].

Ο όρος αναφέρεται επίσης και στα δίκτυα που η φύση των αλληλεπιδράσεων είναι πολύ συγκεκριμένη και αφορά πρωτεΐνες που συνδέονται μοριακά μεταξύ τους κατά τη δημιουργία πρωτεϊνικών συμπλοκών, τα οποία αποτελούν αυτόνομες λειτουργικές μονάδες. Η ταυτοποίηση αυτού του είδους των συνδέσεων είναι πολύ δύσκολο να γίνει πειραματικά, οπότε συχνά αυτά τα δίκτυα περιλαμβάνουν ετερόκλητες πληροφορίες. Οι ακμές τους είναι μη-κατευθυνόμενες, ενώ η αλληλεπίδραση μπορεί να είναι μεταβατική, δηλαδή η δημιουργία ενός συμπλόκου να τροποποιήσει μια πρωτεΐνη η οποία με τη σειρά της επηρεάζει άλλες αλληλεπιδράσεις. Αυτού του είδους οι αλληλεπιδράσεις αποτελούν το δυναμικό τμήμα του δικτύου [8].

1.3. Pathway Analysis

Συνήθως η γονιδιωματική ανάλυση περιλαμβάνει τη μελέτη της έκφρασης των γονιδίων, δηλαδή τον προσδιορισμό των διαφορετικά εκφρασμένων γονιδίων. Διαφορικά εκφρασμένα γονίδια καλούνται τα γονίδια που είναι είτε υπερεκφρασμένα είτε υποεκφρασμένα στη συνθήκη μελέτης – παθολογικά δείγματα – σε σχέση με τη συνθήκη ελέγχου – φυσιολογικά δείγματα. Ωστόσο προσδιορίζοντας τα μεμονωμένα γονίδια που εκφράζονται διαφορετικά δεν παρέχονται αρκετές πληροφορίες σχετικά με τον μηχανισμό που διαφοροποιείται στη συνθήκη μελέτης σε σχέση με τη συνθήκη ελέγχου.

Σύμφωνα με τη θεωρία συστημάτων, τα γονίδια δεν δρουν μόνα τους, αλλά αλληλεπιδρούν με άλλα, καθώς ανήκουν σε δίκτυα. Τα μονοπάτια είναι βιολογικά δίκτυα που σχετίζονται με μια συγκεκριμένη διαδικασία μέσα σε ένα κύτταρο ή έναν οργανισμό. Για παράδειγμα υπάρχουν μονοπάτια που αφορούν τη μείωση και την ενδοκύττωση των κυττάρων. Περνώντας από τη διαφορετική έκφραση των γονιδίων στη διαφορετική έκφραση μονοπατιών, είναι δυνατόν να κατανοηθούν ποια φαινόμενα διαφοροποιούν τη συνθήκη μελέτης από τη συνθήκη ελέγχου. Συνήθως η συνθήκη μελέτης είναι μια ασθένεια, επομένως μέσα από αυτή την ανάλυση προσφέρεται μια σε βάθος κατανόηση της

ασθένειας, καθώς αναγνωρίζονται ποιες βιολογικές διαδικασίες σχετίζονται με αυτή σε κυτταρικό επίπεδο [12].

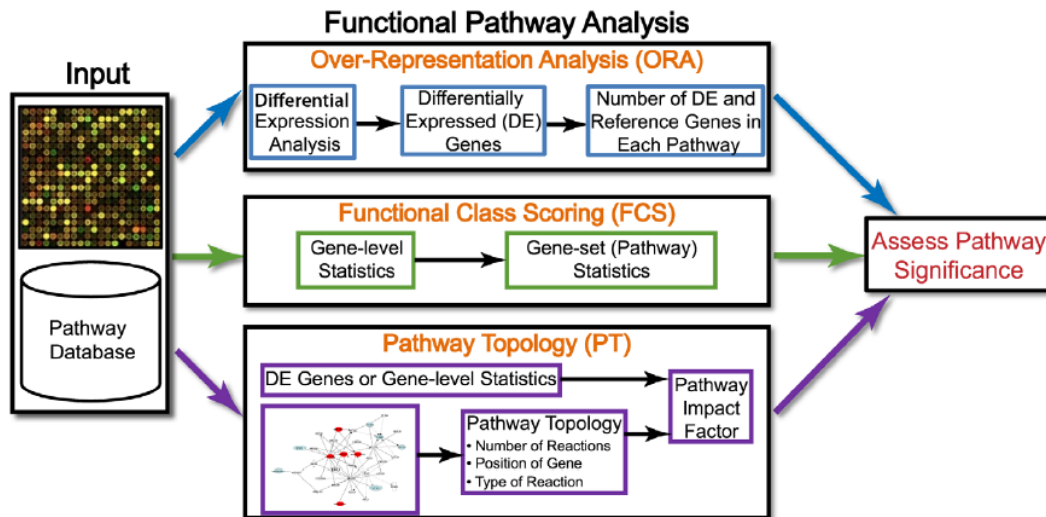
Επιπρόσθετα, υπάρχει η αντίληψη ότι σε πολλές ασθένειες οι διαφορές στην έκφραση των γονιδίων είναι μικρή και δύσκολο να εντοπιστεί για μεμονωμένα γονίδια. Αν όμως μελετιέται η έκφραση των μονοπατιών και υπάρχουν πολλά γονίδια που έχουν λίγο πιο αυξημένη έκφραση από ότι συνήθως, τότε το μονοπάτι αυτό μπορεί να εντοπιστεί ως διαφορεικά εκφρασμένο. Ένα παράδειγμα αποτελεί η περίπτωση του διαβήτη τύπου II. Ενώ δεν φαίνεται να υπάρχουν μεμονωμένα διαφορετικά εκφρασμένα γονίδια, ένα σετ γονιδίων που συμμετέχει στην οξειδωτική φωσφορυλίωση εμφάνισε συντονισμένη μείωση σε διαβητικούς ανθρώπους [13].

Υπάρχουν τρεις διαφορετικοί τρόποι ανάλυσης σε επίπεδο μονοπατιών και παρουσιάζονται στην Εικόνα 1.3.. Ο πρώτος ονομάζεται Over-Representation Analysis; ORA. Αυτή η μέθοδος έχει ως είσοδο τα διαφορετικά εκφρασμένα γονίδια και τα μονοπάτια στα οποία αυτά ανήκουν. Μετά από υπολογισμό της ποσότητας των υπερεκφρασμένων και υποεκφρασμένων γονιδίων που υπάρχουν σε κάθε μονοπάτι, δίνονται ως έξοδο τα μονοπάτια με το μεγαλύτερο πλήθος τέτοιων γονιδίων. Οι κύριοι περιορισμοί αυτής της μεθόδου είναι ότι αξιοποιεί μόνο ένα μέρος των γονιδίων, αυτά που είναι διαφορετικά εκφρασμένα, και μάλιστα δεν λαμβάνει υπόψιν του τα επίπεδα διαφορετικής έκφρασης. Επίσης θεωρεί ότι τόσο τα γονίδια, όσο και τα μονοπάτια είναι ανεξάρτητα μεταξύ τους.

Η δεύτερη μέθοδος ονομάζεται Functional Class Scoring; FCS. Αυτή η μέθοδος έχει ως είσοδο την ανάλυση σε επίπεδο γονιδίων, δηλαδή χρησιμοποιεί όλα τα γονίδια και την τιμή της έκφρασης τους και κάνει στατιστική ανάλυση των μονοπατιών λαμβάνοντας υπόψιν την έκταση των μονοπατιών και την συσχέτιση μεταξύ των γονιδίων. Ως έξοδο δίνει στα μονοπάτια μια τιμή στατιστικής σημαντικότητας. Ωστόσο, η μέθοδος αυτή εξαιρεί το γεγονός πως κάποια γονίδια συμμετέχουν σε πάνω από ένα μονοπάτια, επομένως αν ένα μονοπάτι είναι στατιστικά σημαντικό και περιέχει γονίδια που συμμετέχουν και σε άλλα μονοπάτια, τότε υπάρχει πιθανότητα να φαίνεται ότι και αυτά είναι σημαντικά, ενώ δεν είναι. Ένας άλλος περιορισμός, ο οποίος όμως μπορεί να αντιμετωπιστεί χρησιμοποιώντας τις κατάλληλες στατιστικές μεθόδους, είναι ότι ενώ αρχικά γίνεται μια ταξινόμηση των γονιδίων με βάση την έκφρασή τους, στη συνέχεια αγνοείται η τιμή αυτής της έκφρασης, με αποτέλεσμα γονίδια που υπερεκφράζονται ή υποεκφράζονται πολύ να μην λαμβάνουν την απαιτούμενη βαρύτητα.

Ο τρίτος τρόπος ονομάζεται Pathway Topology Based Analysis (PT). Η μέθοδος αυτή ακολουθεί τα βήματα της FCS, αλλά χρησιμοποιεί και την τοπολογία των μονοπατιών, δηλαδή τον τρόπο που αλληλοεπιδρούν τα γονίδια σε ένα μονοπάτι όταν αλλάζουν οι συνθήκες. Οι κύριοι περιορισμοί αυτής της μεθόδου είναι ότι η πραγματική τοπολογία των μονοπατιών εξαρτάται από το είδος του κυττάρου που μελετάται και η πληροφορία αυτή δεν είναι επαρκώς μελετημένη για την πλειονότητα των περιπτώσεων. Επομένως, είναι πολύ

δύσκολο να γενικευθεί και να ενσωματωθεί σε βάσεις δεδομένων με τα σημερινά βιοπληροφορικά μέσα. Τέλος, αδυνατεί να μοντελοποιήσει τις δυναμικές καταστάσεις του συστήματος και τις αλληλεπιδράσεις μεταξύ των μονοπατιών.



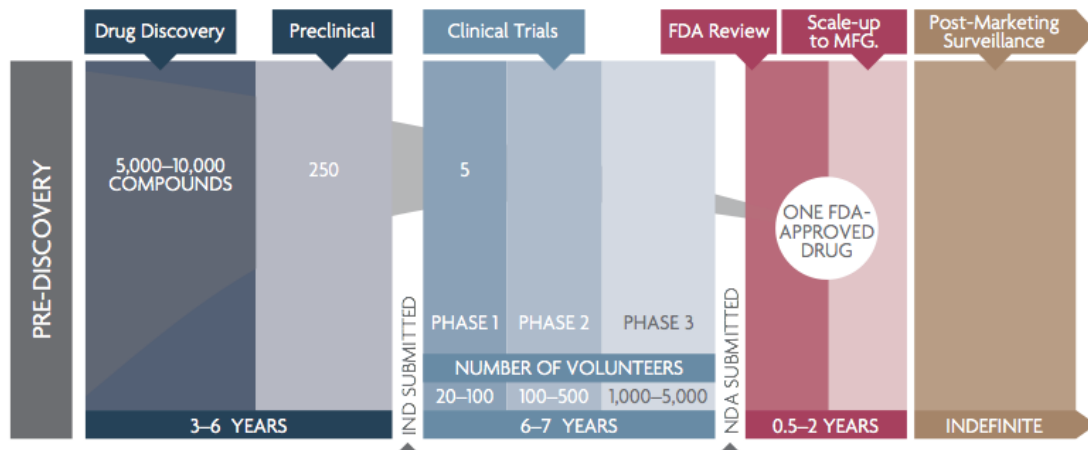
Εικόνα 1.3. Μέθοδοι ανάλυσης μονοπατιών (Khatri et al., 2012)

1.4. Επαναστόχευση φαρμάκων

Η επαναστόχευση/ επανατοποθέτηση φαρμάκων (drug repositioning/ geruprosing) είναι μια μέθοδος που έχει κερδίσει έδαφος τα τελευταία χρόνια. Αναφέρεται στη διαδικασία όπου σε ένα υπάρχον φάρμακο δίνεται ένας νέος στόχος, διαφορετικός από αυτόν που είχε μέχρι τώρα. Τέτοια φάρμακα μπορούν να είναι όσα βρίσκονται στην αγορά, όσα βρίσκονται σε κλινικές μελέτες ή ακόμα και όσα αποσύρθηκαν πριν βγουν στην αγορά.

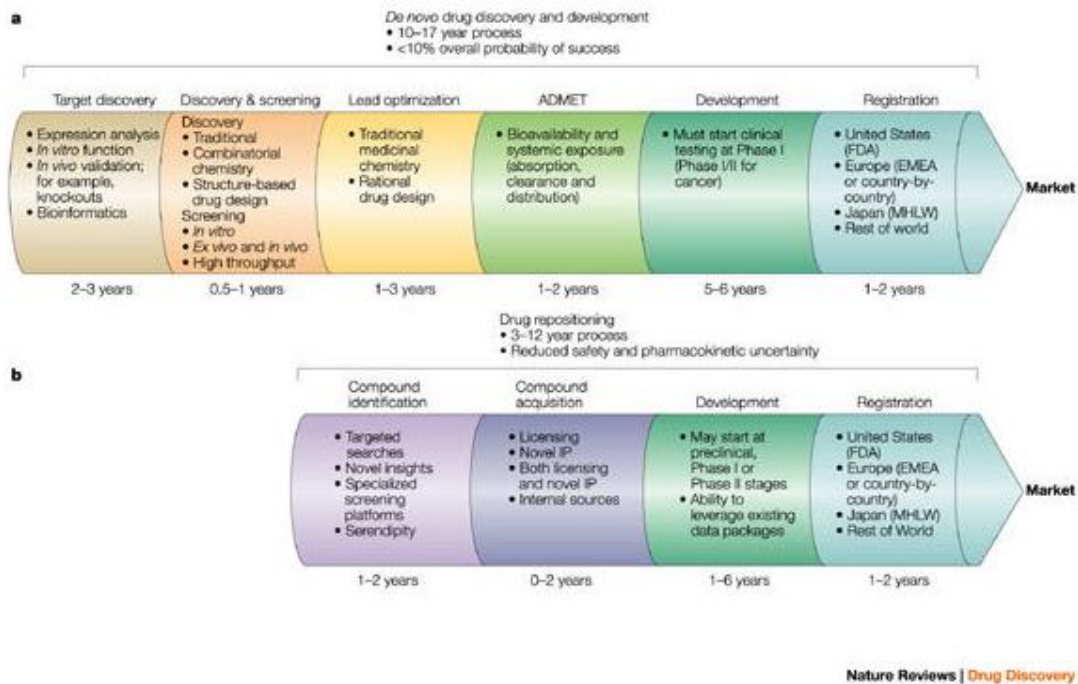
Κάποιο φάρμακο είναι δυνατόν να έχει αποσυρθεί για τους εξής λόγους: 1) απέτυχε στον αρχικό του στόχο, 2) προκαλούσε ανεπιθύμητες παρενέργειες σε συγκεκριμένη ομάδα πληθυσμού, ενώ σε άλλους ασθενείς είχε ευεργετικό ρόλο, 3) η φαρμακευτική εταιρία εγκατέλειψε την προώθηση του φαρμάκου λόγω αναθεώρησης στρατηγικής.

Η διαδικασία ανάπτυξης ενός καινούργιου φαρμάκου είναι μια διαδικασία ιδιαίτερα χρονοβόρα και δαπανηρή, καθώς διαρκεί κατά μέσο όρο δέκα ή περισσότερα χρόνια και κοστίζει περίπου 2.6 δισεκατομμύρια δολάρια. Οι κλινικές δοκιμές μπορούν να διαρκέσουν έξι με επτά χρόνια, ενώ μόνο το 12% των φαρμάκων φτάνουν τη φάση Ι των κλινικών δοκιμών. Τελικά, μόνο το 1-2% των φαρμάκων καταφέρνουν να κυκλοφορήσουν στην αγορά [16]. Στην Εικόνα 1.4. φαίνεται το χρονοδιάγραμμα της διαδικασίας αυτής.

Source: PhRMA⁶

Εικόνα 1.4. R&D φαρμάκων στην PhRMA (2007)

Αντίθετα, η επαναστόχευση φαρμάκων έχει πολλά οφέλη. Κατ' αρχάς, επιλέγοντας φάρμακα που είναι ήδη στην αγορά ή έχουν περάσει από ελέγχους τοξικότητας και έχουν κριθεί ασφαλή στη χρήση, είναι πιο πιθανό ότι δεν θα προκαλέσουν ανεπιθύμητες παρενέργειες. Εφόσον το 30% των αποτυχιών των φαρμάκων στις κλινικές δοκιμές οφείλονται στην ασφάλεια, αυτό αποτελεί μεγάλο πλεονέκτημα της επαναστόχευσης φαρμάκων. Το παραπάνω σε συνδυασμό με την ήδη ανεπτυγμένη φόρμουλα για το φάρμακο συνεπάγεται και σημαντική μείωση του χρόνου και του κόστους μέχρι να κυκλοφορήσει το φάρμακο στην αγορά. Μια πρόσφατη μελέτη έδειξε ότι το κόστος της κυκλοφορίας ενός επαναστοχευμένου φαρμάκου είναι περίπου 8.4 εκατομμύρια δολάρια, ενώ η κυκλοφορία ενός επαναστοχευμένου φαρμάκου, η φόρμουλα του οποίου έχει τροποποιηθεί, κοστίζει 41.3 εκατομμύρια δολάρια, ποσά πολύ μικρότερα από το αντίστοιχο για την δημιουργία ενός νέου φαρμάκου από την αρχή [17]. Γενικά η μέθοδος αυτή συμφέρει τις βιομηχανίες, καθώς μειώνει τον απαιτούμενο χρόνο και πόρους για την ανάπτυξη νέων φαρμάκων [18].

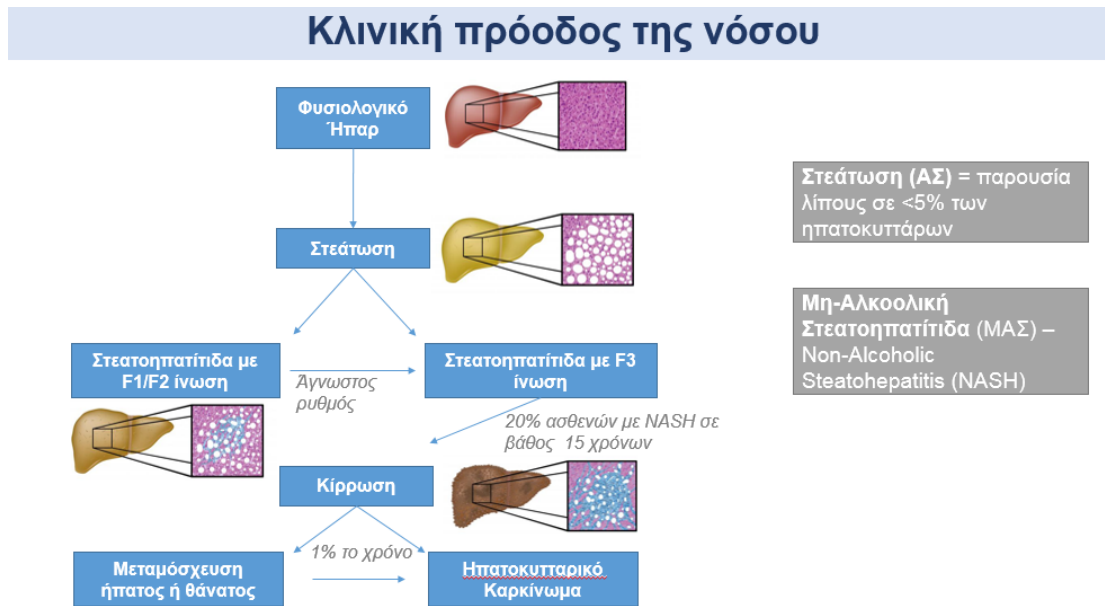


Εικόνα 1.5. Σύγκριση μεταξύ ανάπτυξης νέου φαρμάκου και επανατοποθέτησης φαρμάκου (Ashburn et al., 2004)

1.5. Εφαρμογή : Μη-Αλκοολική Λιπώδης Νόσος του Ήπατος

Η Μη-Αλκοολική Λιπώδης Νόσος του Ήπατος (Non-Alcoholic Fatty Liver Disease; NAFLD) είναι η ηπατική συνιστώσα του μεταβολικού συνδρόμου και ορίζεται ως η συσσώρευση λίπους στο ηπατικό παρέγχυμα χωρίς φλεγμονή, απουσία αυξημένης κατανάλωσης αλκοόλ (λιγότερο από 21 μονάδες στους άνδρες και 14 μονάδες στις γυναίκες) [20]. Τις τελευταίες δεκαετίες παρατηρείται αύξηση των ποσοστών παχυσαρκίας σε παγκόσμιο επίπεδο, γεγονός που έχει συμβάλει στην ανάδειξη της NAFLD σε μείζων θέμα παγκόσμιας υγείας.

Η NAFLD χαρακτηρίζεται από ποικίλες παθολογικές καταστάσεις, οι οποίες μπορεί να αφορούν από απλή στεάτωση έως μη αλκοολική στεατοηπατίτιδα (Non-alcoholic Steatohepatitis; NASH), ίνωση, κύρωση και ηπατοκυτταρικό καρκίνωμα (Hepatocellular Carcinoma; HCC), με όλες τις κλινικές προεκτάσεις τους. Η απλή στεάτωση ορίζεται ως η παρουσία λίπους σε <5% των ηπατοκυττάρων, ενώ το 20-25% των περιπτώσεων εξελίσσεται σε NASH. Το 20% των ασθενών με NASH θα αναπτύξει ίνωση και τελικά κίρρωση.



Εικόνα 1.6. Κλινική πρόοδος της νόσου NAFLD (Alkhoury et al., 2010)

Σχετίζεται στενά με ανοχή στην ινσουλίνη και άλλους παράγοντες παχυσαρκίας, όπως ο διαβήτης, η κεντρική κοιλιακή παχυσαρκία και η δυσλιπιδαιμία. Αποτελεί επίσης ανεξάρτητο παράγοντα ρίσκου για καρδιοαγγειακές παθήσεις ανεξαρτήτως ηλικίας, φύλου και χοληστερίνης. Σχετίζεται με αυξημένο ρίσκο θνητότητας από κάθε αίτιο, στο οποίο συμβάλλουν τόσο παράγοντες που σχετίζονται με το ήπαρ όσο και δευτερογενείς παράγοντες, όπως κακοήθειες, διαβήτης και στεφανιαία νόσος [22].

Η επικράτηση της μη-αλκοολική λιπώδης νόσου του ήπατος σε άτομα φυσιολογικού βάρους, χωρίς παρουσία μεταβολικών παραγόντων κινδύνου εμφανίζεται να είναι 16%, και ανέρχεται σε 43-60% σε ασθενείς με διαβήτη, και έως 91% σε ασθενείς με υπερλιπιδαιμία. Η επικράτηση της αυξάνεται επίσης με την ηλικία, από λιγότερο του 20% σε άτομα κάτω των 20 ετών, σε περισσότερο από 60% σε άτομα άνω των 60.

Η Μη-Αλκοολική Λιπώδης Νόσος του Ήπατος είναι κατά βάση μη-συμπτωματική ασθένεια και η διάγνωσή της γίνεται συχνά τυχαία. Οι κλινικοί ιατροί θα πρέπει να εξετάζουν τη πιθανότητα ύπαρξης της νόσου σε ασθενείς με μη κανονικούς ηπατικούς δείκτες και με ταυτόχρονη παρουσία ενός ή παραπάνω μεταβολικών παραγόντων επικινδυνότητας: η πιθανότητα εμφάνισης της νόσου αυξάνεται αναλογικά με τον αριθμό παραγόντων του μεταβολικού συνδρόμου που υπάρχουν.

Το υπερηχογράφημα αποτελεί καλό διαγνωστικό εργαλείο για περιπτώσεις με τουλάχιστον 30% ηπατική στεάτωση με 64% ευαισθησία και 85% ειδικότητα, παράγοντες που βελτιώνονται συνεχώς. Άλλα μέσα για τη διάγνωση ηπατικής στεάτωσης αποτελούν η παράμετρος ελεγχόμενης εξασθένισης (Control Attenuation Parameter; CAP, λογισμικό του συστήματος Fibroscan) και η σπεκτροσκοπία μαγνητικού συντονισμού (Magnetic Resonance Spectroscopy;

MRS), τα οποία είναι πιο ευαίσθητα και προσφέρουν ποσοτική ανάλυση της ηπατικής κατάστασης.

Μέχρι σήμερα ωστόσο, η βιοψία ήπατος αποτελεί το πιο αξιόπιστο και διαδεδομένο μέσο για τη διάγνωση και εκτίμηση του βαθμού NAFLD/NASH και είναι το πρωτεύον τελικό σημείο για τα τελικά στάδια των κλινικών δοκιμών [23].

Οι σύγχρονοι μέθοδοι αντιμετώπισης της NAFLD είναι η αλλαγή τρόπου ζωής, δηλαδή η εισαγωγή γυμναστικής και η παρέμβαση στις διαιτητικές συνήθειες. Κανένα φάρμακο δεν έχει μέχρι στιγμής εγκριθεί που να χορηγείται για την αντιμετώπιση της νόσου. Πολλές μελέτες ασχολούνται με την ανεύρεση φαρμάκου και κάποιες έχουν βγάλει αποτελέσματα από *in vitro* ή *in vivo* πειράματα, ενώ άλλες είναι στο στάδιο των κλινικών δοκιμών.

1.6. Σκοπός

Λόγω της μεγάλης εξάπλωσης της νόσου στον πληθυσμό είναι απαραίτητο να βρεθούν βιοδείκτες και τρόποι φαρμακευτικής αντιμετώπισης αυτής. Πολλοί επιστήμονες έχουν στραφεί προς τις τεχνολογίες υψηλής απόδοσης (high-throughput). Σε αντιδιαστολή με τη χρονοβόρα διαδικασία δημιουργίας νέου φαρμάκου του FDA, οι *in silico* μέθοδοι, όπως η μελέτη για επαναστόχευση φαρμάκων μπορούν να συμβάλουν στην γρηγορότερη παραγωγή φαρμακευτικών ουσιών για την αντιμετώπιση της νόσου.

Σκοπός της παρούσας διπλωματικής είναι η δημιουργία μιας πλατφόρμας επανατοποθέτησης φαρμάκων για τη NAFLD, συνδυάζοντας *in vitro* μοντέλα ηπατοκυτταρικών σειρών με ανάλυση δικτύων από δεδομένα γονιδιακής έκφρασης ασθενών με NAFLD.

Κεφάλαιο 2

Μέθοδοι και Υλικά

2.1. Υπολογιστική Ανάλυση

2.1.1. *Gene Expression Omnibus*

Η Gene Expression Omnibus (GEO) του National Center for Biotechnology Information (NCBI) είναι μια δημόσια διαδικτυακή αποθήκη γονιδιωματικών (genomics) δεδομένων. Η πλειοψηφία των δεδομένων που περιέχει προέρχονται από μελέτες γονιδιακής ανάλυσης που μελετούν ένα μεγάλο εύρος βιολογικών θεμάτων, όπως είναι διάφορες ασθένειες, η εξέλιξη, η ανοσολογία, η τοξικολογία, ο μεταβολισμός και άλλα. Τα δεδομένα κάθε μελέτης παρέχονται στη μορφή μικροσυστοιχιών [25].

Το πρώτο βήμα για να ξεκινήσει η ανάλυση για να γίνει η επιλογή των δεδομένων που θα επεξεργαστούν. Η αναζήτηση πραγματοποιήθηκε με όρους «NAFLD» και λήφθηκαν 628 αποτελέσματα στην GEO DataSets Database.

Για κάθε σετ δεδομένων γίνεται εξαγωγή ορισμένων πληροφοριών, όπως τίτλος, περίληψη, οργανισμός, αναφορά στο άρθρο και στον συγγραφέα, κωδικός αναφοράς και αριθμός δειγμάτων. Με βάση τον οργανισμό και την περίληψη, αρχικά επιλέγονται σετ δεδομένων που αφορούν άνθρωπο (*Homo sapiens sp.*) και περιλαμβάνουν και υγιείς ανθρώπους και ανθρώπους που πάσχουν από κάποια μορφή NAFLD.

2.1.2. *Διάγραμμα κρατήρα ηφαιστείου (Volcano plot)*

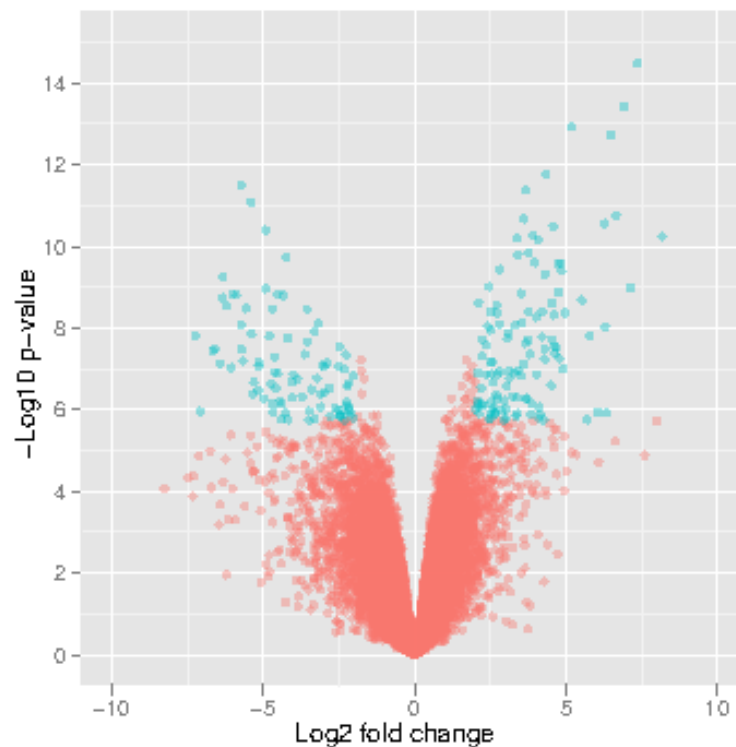
Για κάθε δείγμα παρέχονται μετρήσεις συγκεντρώσεων οι οποίες ποσοτικοποιούν την έκφραση των γονιδίων στα κύτταρα. Σκοπός είναι να υπολογιστεί ο αριθμός των διαφορεικά εκφρασμένων γονιδίων. Ο βαθμός μεταβολής της έκφρασης ενός γονιδίου υπολογίζεται με τη χρήση του λογαρίθμου του λόγου των τιμών έκφρασης στη συνθήκη μελέτης προς τη συνθήκη ελέγχου:

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

Έτσι το ουδέτερο σημείο της ποσότητας είναι στο μηδέν και λόγοι θετικοί αντιστοιχούν σε υπερέκφραση του γονιδίου, ενώ λόγοι αρνητικοί αντιστοιχούν σε υποέκφραση του γονιδίου. Επίσης η τιμή $\log_2 FC(g) = 1$ σημαίνει διπλάσια έκφραση, ενώ η τιμή $\log_2 FC(g) = -1$ σημαίνει υποδιπλάσια έκφραση.

Ωστόσο, για να αποφευχθεί η χρήση μιας τιμής που φαινομενικά μοιάζει σημαντική αλλά οφείλεται σε τυχαία διακύμανση, χρειάζεται στατιστικός έλεγχος των δεδομένων. Απαιτείται λοιπόν να γίνει σύγκριση ενός ζεύγους κατανομών για κάθε γονίδιο και όχι ενός ζεύγους τιμών. Έτσι κάθε πείραμα έκφρασης πρέπει να εκτελείται τουλάχιστον τρεις φορές [26]. Στην περίπτωση αυτή, ο στατιστικός έλεγχος γίνεται με τη χρήση του Student's t-test. Ορίζεται ως «μηδενική υπόθεση» ότι οι μέσες τιμές των κατανομών των δύο συνθηκών είναι ταυτόσημες και υπολογίζουμε το μέγεθος t. Το μέγεθος t αντιστοιχίζεται στο μέγεθος p-value, το οποίο όσο πιο μικρό είναι, τόσο ισχυρότερη ένδειξη είναι ότι η μηδενική υπόθεση δεν ισχύει και οι μέσες τιμές αντιστοιχούν σε διαφορετικές κατανομές, που σημαίνει ότι η διαφορά που παρατηρείται είναι στατιστικά σημαντική και υπάρχει διαφορική έκφραση.

Τα δύο παραπάνω μεγέθη αναπαριστώνται στο διάγραμμα «κρατήρας ηφαιστείου» (Volcano plot). Κάθε σημείο στο διάγραμμα αντιστοιχεί σε ένα γονίδιο, ενώ στον y – άξονα είναι ο αρνητικός δεκαδικός λογάριθμος του $p - value$ και στον x – άξονα είναι η τιμή $\log_2 FC$. Όσο ψηλότερα είναι στον κάθετο άξονα ένα γονίδιο, τόσο πιο σημαντικά στατιστικά είναι η διαφορική του έκφραση, ενώ όσο πιο μακριά είναι από το σημείο μηδέν στον οριζόντιο άξονα, τόσο πιο μεγάλη είναι η έντασή της (Εικόνα 2.1.). Ως όρια για τη διαφορική έκφραση ενός γονιδίου συνήθως χρησιμοποιούνται οι τιμές $|\log_2 FC| > 1$ και $p - value \leq 0.05$.



Εικόνα 2.1. Διάγραμμα κρατήρα ηφαιστείου – Volcano plot (Gillespie, 2011)

2.1.3. Γονιδιακή ανάλυση (Gene Level Statistics)

Η στατιστική μελέτη στο επίπεδο των γονιδίων (Gene Level Statistics; GLS) γίνεται με το πακέτο της R Bioconductor «Linear Models for Microarray and RNA-Seq Data; limma» [28].

Για κάθε γονίδιο δημιουργείται ένα γραμμικό μοντέλο της μορφής:

$$y_g = \beta_{g0} + \beta_{g1} * x_{g1} + \dots + \beta_{gn} * x_{gn} + \varepsilon_g, g \in 1 \dots G \quad (1)$$

Όπου :

y_g : η έκφραση του γονιδίου g

β_{g0} : η έκφραση της τιμής του γονιδίου g στο δείγμα αναφοράς (control)

$\beta_{g1}, \dots, \beta_{gn}$: οι διαφορές των τιμών έκφρασης του γονιδίου g των δειγμάτων 1-n από το δείγμα αναφοράς

x_{g1}, \dots, x_{gn} : οι μεταβλητές με τιμές 1 ή 0, ανάλογα με το αν η σύγκριση αφορά ή όχι το n-οστό δείγμα αντίστοιχα

ε_g : το σφάλμα υπολογισμού της γονιδιακής έκφρασης

Απαλείφοντας το σφάλμα ε_g , η τιμή y_g εκτιμάται με την τιμή \widehat{y}_g και οι υπόλοιπες παράμετροι μετασχηματίζονται σε $\widehat{\beta}_{g0}, \dots, \widehat{\beta}_{gn}$. Ορίζεται ως μηδενική υπόθεση H_0 ότι δεν υπάρχει διαφορά μεταξύ της έκφρασης του γονιδίου στο δείγμα αναφοράς και στο εκάστοτε δείγμα, δηλαδή $\widehat{\beta}_{gn} = 0$. Από τη γραμμική άλγεβρα το διάνυσμα αυτό υπολογίζεται ως:

$$\widehat{\beta}_{gn} = (X^T X)^{-1} X^T Y$$

Όπου,

X : ο πίνακας των μεταβλητών x_{g0}, \dots, x_{gn} , όπου οι σειρές αντιστοιχούν στο διαφορετικά γονίδια και οι στήλες στα διαφορετικά δείγματα

Y : το διάνυσμα των τιμών της γονιδιακής έκφρασης

Το τυπικό σφάλμα υπολογίζεται ως: $se(\widehat{\beta}_{gn}) = \sqrt{s_g^2 (X^T X)^{-1}}$

Όπου s_g^2 η διακύμανση του γονιδίου g και η δοκιμασία σημαντικότητας (significance test), από την οποία υπολογίζεται το στατιστικό t για κάθε γονίδιο

λαμβάνει τη μορφή: $t - test = \frac{\widehat{\beta}_{gn}}{se(\widehat{\beta}_{gn})}$

Η παραπάνω μεθοδολογία έχει ένα σημαντικό πρόβλημα, όταν το τυπικό σφάλμα είναι αρκετά μικρό δίνει πολλά ψευδώς στατιστικά σημαντικά (false positive; FP) εκφρασμένα γονίδια. Το πρόβλημα αυτό αντιμετωπίζεται με τη χρήση ιεραρχικών μοντέλων, τα οποία περιγράφουν τη μεταβολή των συντελεστών $\widehat{\beta}_{gn}$ και της διακύμανσης (s_g^2) συναρτήσει των γονιδίων.

Πιο συγκεκριμένα, πρώτα διαχωρίζουμε τους ασθενείς σε τρεις ομάδες, τους υγιείς (Healthy), αυτούς που πάσχουν από NASH και αυτούς που έχουν στεάτωση. Έπειτα, με τον τρόπο που αναλύθηκε προηγουμένως, δημιουργείται ένα γραμμικό μοντέλο για κάθε γονίδιο. Ωστόσο, από την εξίσωση (1) αξιοποιούνται μόνο οι όροι x_{g1} και x_{g2} , καθώς αντί για μεμονομένα δείγματα χρησιμοποιούνται οι τρεις ομάδες δειγμάτων. Στη συνέχεια γίνεται σύγκριση μεταξύ των υγιών και αυτών που πάσχουν από NASH και μεταξύ των υγιών ασθενών και αυτών έχουν στεάτωση (Εικόνα 2.2.).

Τέλος, μέσω της εμπειρικής Μπεϋζιανής μεθόδου (empirical Bayes) υπολογίζονται οι απαραίτητες υπερπαράμετροι, η τιμή B-value καθώς εφαρμόζεται και ένα μετριοπαθές t – στατιστικό (moderated t-statistic). Έτσι προκύπτουν οι παράμετροι που φαίνονται στην Εικόνα 2.3. για κάθε σύγκριση και για κάθε γονίδιο.

Levels	Contrasts	
	NASHvHealthy	stvHealthy
Healthy	-1	-1
NASH	1	0
Steatosis	0	1

Εικόνα 2.2. Πίνακας σύγκρισης

	logFC	AveExpr	t	P.value	adj.P.val	B	TF
1	-3.979716	11.02524	-12.821248	3.202044e-19	2.222112e-15	33.23330	FOSB
2	-2.857281	10.13552	-8.021568	3.259439e-11	4.712379e-09	15.33726	IL6
3	-2.768875	12.93391	-10.355020	3.134180e-15	2.104854e-12	24.34196	FOS
4	2.706911	12.30086	8.700111	2.134833e-12	5.108631e-10	17.99350	CYP7A1
5	-2.687602	11.97376	-13.629670	1.856159e-20	1.932169e-16	35.97085	MYC
6	-2.632605	11.47382	-12.073915	4.796385e-18	1.426513e-14	30.62267	JUNB

Εικόνα 2.3. Μορφή αποτελεσμάτων της μελέτης στο επίπεδο γονιδιακής έκφρασης (GLS)

2.1.4. Ανάλυση μονοπατιών (Gene Set Analysis)

Η ανάλυση στο επίπεδο των μονοπατιών γίνεται με το πακέτο της R Bioconductor «Platform for integrative analysis of omics data; riano» και είναι μια μέθοδος FCS [29]. Χρησιμοποιώντας τα σετ γονιδίων που δίνει η βάση δεδομένων Molecular Signatures Database (MSigDB [30]) και τα αποτελέσματα της προηγούμενης ανάλυσης πραγματοποιείται η μέθοδος της GSA με εννέα διαφορετικές στατιστικές μεθόδους. Αυτές είναι οι: “Fisher”, “Stouffer”, “Reporter”, “Tail strength”, “page”, “maxmean”, “sum”, “mean” και “median”.

- “Fisher”: Η βασική μορφή του ακριβές τεστ του Fisher ελέγχει την ανεξαρτησία δύο μεταβλητών για πίνακες συνάφειας 2x2 υπολογίζοντας τιμές p-value [31].
- “Stouffer”: Η μέθοδος του Stouffer βασίζεται στα p-value, δημιουργεί τα Z-score και επιτρέπει τη χρήση συντελεστών βαρύτητας [32].
- “Reporter”: Ο αλγόριθμος Reporter Features χαρτογραφεί τη γονιδιακή ανάλυση σε μεταβολικά δίκτυα γονιδίων αναγνωρίζοντας

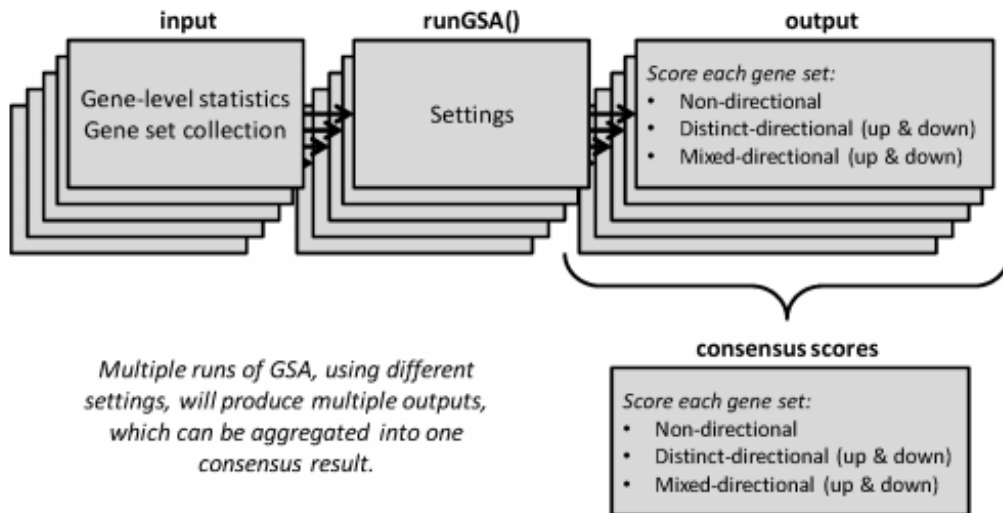
ομάδες γειτονικών γονιδίων τα οποία είναι διαφορετικά εκφρασμένα. Ο αλγόριθμος αυτός βασίζεται στον «Reporter Metabolites Algorithm» των Patil & Nielsen [33].

- “Tail strength”: Έστω ότι έχουμε μηδενική υπόθεση και συσχετισμένες τιμές p-value p_i $i=1,2,\dots,m$. Θεωρώντας ως μηδενική υπόθεση ότι οι τιμές p_i είναι ανεξάρτητες και πανομοιότυπα κατανομημένα τυχαίες μεταβλητές στο διάστημα $[0,1]$ και ότι ισχύει $p_1 \leq p_2 \leq \dots \leq p_m$. Το μέγεθος tail strength (TS) υπολογίζει την απόκλιση μιας τιμής p-value από την αναμενόμενη. Θετικές τιμές συνεπάγονται απόρριψη της μηδενικής υπόθεσης, που σημαίνει ότι υπάρχουν περισσότερες μικρές τιμές p-value από ότι θα περιμέναμε αν ήταν τυχαίες. Τέλος το μέγεθος αυτό δίνει περισσότερη βαρύτητα στις χαμηλότερες τιμές p-value που σημαίνει ότι είναι πιο ευαίσθητο σε αποκλίσεις στην ουρά (tail) [34].
- “page”: Η μέθοδος αυτή με τις τιμές των fold change υπολογίζει Z score για κάθε σετ γονιδίων και χρησιμοποιώντας την κανονική κατανομή καταλήγει σε στατιστικά σημαντικά σετ γονιδίων. (Parametric Analysis of Gene set Enrichment; page) [35].
- “maxmean”: Στη μέθοδο αυτή υπολογίζεται η μέση τιμή των θετικών και των αρνητικών τιμών έκφρασης για κάθε γονίδιο σε κάθε σετ γονιδίων και επιλέγεται το σετ γονιδίων με τη μεγαλύτερη απόλυτη τιμή [36].
- “sum”, “mean”, “median”: Στις μεθόδους αυτές υπολογίζονται το άθροισμα, η μέση τιμή και η διάμεσος για κάθε σετ γονιδίων. Στην πρώτη χρησιμοποιείται το μέγεθος t , ενώ στις άλλες δύο το μέγεθος fold change.

Επειδή καμία από τις παραπάνω μεθόδους δεν μπορεί να θεωρηθεί πολύ καλύτερη των υπολοίπων, επιλέγεται να συνδυαστούν τα αποτελέσματα από όλες τις μεθόδους ώστε να κυριαρχήσει η επικρατούσα τάση (Εικόνα 2.4.). Η στατιστική ανάλυση των γονιδίων ανάγεται λοιπόν σε στατιστική ανάλυση των μονοπατιών και αυτά κατατάσσονται σε πέντε κατηγορίες ανάλογα με το p – value που τα χαρακτηρίζει. Οι κατηγορίες είναι οι εξής: Distinct up, mixed up, non directional, mixed down και distinct down.

- Distinct up: Τα γονίδια που είναι υπερεκφρασμένα και τα γονίδια που είναι υποεκφρασμένα αλληλοαναιρούνται. Υπάρχει συνολικά υψηλότερη υπερέκφραση στο μονοπάτι.
- Mixed up: Υπάρχει μια υψηλή συνιστώσα υπερεκφρασμένων γονιδίων. Η συνιστώσα των υποεκφρασμένων γονιδίων δεν λαμβάνεται υπόψιν.
- Non directional: Χρησιμοποιείται η απόλυτη τιμή της διαφορικής έκφρασης των γονιδίων. Υπάρχει υψηλή διαφορική έκφραση γονιδίων, χωρίς να λαμβάνεται υπόψιν η κατεύθυνση αυτής.

- Mixed down: Υπάρχει μια υψηλή συνιστώσα υποεκφρασμένων γονιδίων. Η συνιστώσα των υπερεκφρασμένων γονιδίων δεν λαμβάνεται υπόψιν.
- Distinct down: Τα γονίδια που είναι υπερεκφρασμένα και τα γονίδια που είναι υποεκφρασμένα αλληλοαναιρούνται. Υπάρχει συνολικά υψηλότερη υποέκφραση στο μονοπάτι.



Εικόνα 2.4. Συνδυασμός των διαφορετικών μεθόδων GSA (Ritchie et al., 2015)

2.1.5. Drugbank

Η Drugbank είναι μια διαδικτυακή βάση δεδομένων βιοπληροφορικής και χημειοπληροφορικής που συνδυάζει λεπτομερή δεδομένα για κάθε φαρμακολογικό παράγοντα καθώς και για τις μοριακές και γονιδιακές του αλληλεπιδράσεις. Η τελευταία έκδοση της Drugbank περιέχει 10.500 φάρμακα. Πιο αναλυτικά, περιέχει 1.737 εγκεκριμένα μικρά μόρια φαρμάκων, 870 εγκεκριμένα βιοτεχνολογικά φάρμακα, 103 συμπληρώματα διατροφής και πάνω από 5.023 πειραματικά φάρμακα. Επιπλέον 4.772 πρωτεΐνες (Drug target/ enzyme/ transporter/ carrier) είναι συνδεδεμένες σε αυτά τα φάρμακα. Για κάθε φάρμακο οι μισές πληροφορίες είναι αφιερωμένες σε χημικά δεδομένα και οι άλλες μισές στους στόχους (drug target) [37].







2.1.6. Connectivity Map

Το Connectivity Map (cMap) είναι ένα διαδικτυακό εργαλείο βιοπληροφορικής, προϊόν συνεργασίας των πανεπιστημίων Harvard και MIT. Περιέχει μια βάση δεδομένων με γονιδιακές υπογραφές φαρμάκων και ένα εργαλείο για να εντοπίζονται ομοιότητες μεταξύ των υπογραφών αυτών.

Η βάση δεδομένων περιέχει στοιχεία γονιδιακής έκφρασης που λαμβάνονται με την τεχνολογία *in situ* ολιγονουκλεοτιδικών μικροσυστοιχιών. Σε πέντε σειρές καρκινικών κυττάρων εφαρμόζονται 1.309 βιοδραστικά μόρια/φάρμακα σε διαφορετικές συγκεντρώσεις και για διαφορετικό χρόνο. Οι καρκινικές σειρές που χρησιμοποιήθηκαν είναι: λευχαιμικά κύτταρα (HL60), καρκινικά κύτταρα μαστού (MCF7, ssMCF7), καρκινικά κύτταρα προστάτη (PC3) και κύτταρα

μελανώματος (SKMEL5). Οι συγκεντρώσεις είναι ανάλογες αυτών που χρησιμοποιούνται επιτυχημένα σε πειραματικές *in vitro* εφαρμογές, ενώ σε περιπτώσεις που έλειπε αυτή η γνώση χρησιμοποιούνται 10μΜ. Η διάρκεια κάθε πειράματος είναι 6 ή/και 12 ώρες με σκοπό να καταγραφεί ο άμεσος μηχανισμός δράσης του φαρμάκου. Τα βιοδραστικά μόρια αποκαλούνται διαταράκτες (*perturbagens*), καθώς η εισαγωγή τους διαταράσσει τη φυσιολογική λειτουργία των κυττάρων και οδηγεί σε διαφορετική γονιδιακή έκφραση. Αυτά τα μόρια διατρέχουν ένα ευρύ φάσμα ουσιών από αντιβιώσεις μέχρι φάρμακα χημειοθεραπειών και φυτικά εκχυλίσματα. Σε κάθε πείραμα σε ορισμένες μικροσυστοιχίες κυττάρων εφαρμόζεται το φάρμακο (*treatment*), ενώ σε άλλες εφαρμόζεται ψευδοφάρμακο (*vehicle*) ως συνθήκη ελέγχου (*control*). Έτσι δημιουργείται ένα ζεύγος *treatment* – *control*. Χρησιμοποιώντας μια λίστα ανιχνευτών (*probe set*) καταγράφονται τα γονίδια που είναι διαφορετικά εκφρασμένα στη συνθήκη μελέτης *-treatment-* προς τη συνθήκη ελέγχου *-control*. Δημιουργείται μια λίστα, η οποία στην αρχή της έχει τα υπερεκφρασμένα γονίδια και στο τέλος της τα υποεκφρασμένα γονίδια. Κάθε πείραμα αποδίδει παραδείγματα (*treatment instances*) που στο σύνολο τους αριθμούν τα 6.100. Σε κάθε παράδειγμα προσδίδεται ένας αναγνωριστικός αριθμός και χαρακτηρίζεται από το πείραμα από το οποίο προήλθε, το είδος της κυτταρικής σειράς, το όνομα και τη συγκέντρωση του φαρμάκου και τον αριθμό σάρωσης των μικροσυστοιχιών θεραπείας και ελέγχου. Σαν αποτέλεσμα δημιουργείται ένα γονιδιακό προφίλ (γονιδιακή υπογραφή; *gene signature*) για κάθε φάρμακο.

Το εργαλείο που εντοπίζει ομοιότητες μεταξύ γονιδιακών υπογραφών λειτουργεί θέτοντας μια υπογραφή-ερώτηση. Αυτή μπορεί να είναι η γονιδιακή υπογραφή κάποιας ασθένειας που παρέχεται από τον χρήστη ή η γονιδιακή υπογραφή κάποιου άλλου φαρμάκου. Η σύγκριση γίνεται με το στατιστικό τεστ Kolmogorov-Smirnov (K-S), μια μη παραμετρική μέθοδο που προσδιορίζει εάν οι κατανομές δύο δειγμάτων διαφέρουν [38]. Πρακτικά, ελέγχεται αν η υπογραφή ερώτηση έχει παρόμοια ή αντίθετη γονιδιακή υπογραφή με τα φάρμακα που είναι καταχωρημένα στη βάση δεδομένων. Κάθε γονίδιο στην υπογραφή ερώτησης φέρει ένα σήμα που υποδεικνύει εάν είναι υπερεκφρασμένο ή υποεκφρασμένο. Όταν η υπογραφή-ερώτηση συγκρίνεται με τα φάρμακα της βάσης, εξετάζεται αν τα υπερεκφρασμένα γονίδια της είναι στην κορυφή της λίστας του εκάστοτε φαρμάκου και τα υποεκφρασμένα γονίδια της στον πάτο της λίστας ή αντίστροφα. Στην πρώτη περίπτωση έχουμε παρόμοια γονιδιακή υπογραφή, ενώ στην δεύτερη έχουμε αντίστροφη γονιδιακή υπογραφή. Κατά την παραπάνω διαδικασία υπολογίζεται μια τιμή που ονομάζεται τιμή εμπλουτισμού (*Enrichment Score*; ES) και παίρνει τιμές στο διάστημα [-1,1]. Οι θετικές τιμές αντιστοιχούν σε παρόμοια γονιδιακή υπογραφή, ενώ οι αρνητικές σε αντίθετη. Για κάθε ερώτηση που τίθεται παράγεται ένας πίνακας με βιοδραστικές ουσίες, όπως φαίνεται στην παρακάτω εικόνα.

rank	cmap name	mean	n	enrichment	p	specificity	% non-null	
1	cefamandole	-0.633	4	-0.930	0.00002	0.0000	100	
2	glycopyrronium bromide	-0.608	5	-0.883	0.00006	0.0000	100	
3	cefoperazone	-0.589	3	-0.943	0.00028	0.0000	100	
4	meclocycline	-0.623	4	-0.872	0.00056	0.0000	100	
5	amiodarone	0.753	5	0.800	0.00078	0.0069	80	
6	sulfamethizole	-0.560	4	-0.833	0.00145	0.0000	100	

Εικόνα 2.5. Μορφή αποτελεσμάτων του cMap

2.1.7. ToxDB

Το ToxDB είναι ένα διαδικτυακό εργαλείο βιοπληροφορικής που επιχειρεί να περιγράψει την επίδραση φαρμάκων στην έκφραση γονιδίων και σε επίπεδο μονοπατιών. Η αναζήτηση μπορεί να γίνει είτε με φάρμακα είτε με μονοπάτια. Επίσης παρέχει πληροφορίες για το επίπεδο τοξικότητας των φαρμάκων [42].

2.1.8. LiverTox

Το LiverTox είναι μια ανοιχτή ιστοσελίδα που προσφέρει ενημερωμένες, κατανοητές και αμερόληπτες πληροφορίες για φάρμακα που προκαλούν ζημιά στο ήπαρ. Τα φάρμακα αυτά μπορεί να είναι φάρμακα που θέλουν συνταγογράφηση ή όχι καθώς και φυτικά εκχυλίσματα ή πρόσθετα διατροφής. Προσφέρει μια εισαγωγή σχετικά με τη ζημιά που μπορούν να προκαλέσουν τα διάφορα φάρμακα στο ήπαρ, συγκεκριμένα στοιχεία φαρμάκων με λεπτομερείς πληροφορίες σχετικά με την ηπατοτοξικότητα τους, καθώς είναι ανοιχτό και για την κατάθεση κλινικών μελετών και σχολίων επί των αναρτημένων δεδομένων [43].

2.1.9. Μεθοδολογία

Όπως αναφέρθηκε προηγουμένως, από τη GEO επιλέγονται ορισμένα σετ δεδομένων που αφορούν στον άνθρωπο (*Homo sapiens sp.*) και περιλαμβάνουν και υγιείς ανθρώπους και ανθρώπους που πάσχουν από κάποια μορφή NAFLD. Από αυτά, αφού σχεδιαστούν τα volcano plots, γίνεται μια δεύτερη διαλογή που βασίζεται στο πλήθος των διαφορετικά εκφρασμένων γονιδίων, καθώς όσο πιο πολλά είναι αυτά, τόσο περισσότερη πληροφορία δίνει το συγκεκριμένο σετ για τη νόσο.

Το επόμενο βήμα είναι η πλήρης ανάλυση των δεδομένων που έχουν επιλεγεί. Ενώ προηγουμένως έγινε ένας γρήγορος υπολογισμός των διαφορετικά εκφρασμένων γονιδίων, τώρα ο στόχος είναι να βρεθούν ποια μονοπάτια είναι στατιστικά σημαντικά για τη νόσο. Πρώτα γίνεται στατιστική μελέτη στο επίπεδο των γονιδίων (GLS) και έπειτα ακολουθεί η ανάλυση στο επίπεδο των μονοπατιών (GSA). Να σημειωθεί ότι για κάθε σετ δεδομένων η ανάλυση διεξάγεται ξεχωριστά.

Αρχικά χρησιμοποιούνται τα ανεπεξέργαστα δεδομένα από τα σετ δεδομένων που επιλέχθηκαν και γίνεται RMA κανονικοποίηση ποσοστημορίων. «Η συγκεκριμένη μέθοδος οδηγεί σε συγκρίσιμες κατανομές

τιμών ακόμα και στην περίπτωση που οι αρχικές κατανομές δεν είναι κανονικές. Η διαδικασία βασίζεται ουσιαστικά στη σύγκριση της κατάταξης των τιμών και για το λόγο αυτό είναι ανεξάρτητη των ροπών (μέση τιμή, διασπορά κλπ).» [3] Τα ανεπεξέργαστα δεδομένα είναι μια μέτρηση ανά δείγμα, για κάθε μεταγραφικό παράγοντα που αντιστοιχεί σε κάποιο γονίδιο. Πολλές φορές ο ίδιος μεταγραφικός παράγοντας αντιστοιχεί σε πολλά γονίδια ή διαφορετικοί μεταγραφικοί παράγοντες αντιστοιχούν στο ίδιο γονίδιο. Οι πρώτοι εξαιρέθηκαν από τα δεδομένα, ενώ οι δεύτεροι ομαδοποιήθηκαν με χρήση της μέσης τιμής. Η στατιστική μελέτη στο επίπεδο των γονιδίων εκτελείται όπως περιεγράφηκε παραπάνω και δίνει δεδομένα όπως φαίνονται στην Εικόνα 2.3. για κάθε σύγκριση, για κάθε γονίδιο.

Στη συνέχεια για κάθε σετ δεδομένων και για κάθε σύγκριση εκτελείται η ανάλυση στο επίπεδο των μονοπατιών και τα σετ γονιδίων κατηγοριοποιούνται στις πέντε κατηγορίες (distinct up, mixed up, non directional, mixed down, distinct down) και επιλέγονται τα δώδεκα (12) μονοπάτια που είναι υψηλότερα βαθμολογημένα σε κάθε κατηγορία. Αυτά είναι τα μονοπάτια που έχουν τη σημαντικότερη διαφορική έκφραση για τη νόσο NAFLD και αποτελούν την Ομάδα 1.

Στο εργαστήριο έχει προηγηθεί η μελέτη και η δημιουργία μοντέλων επαγωγής NAFLD σε πρωτογενή ανθρώπινα ηπατοκύτταρα και ηπατοκυτταρικές σειρές με πέντε διαφορετικές χημικές/φαρμακολογικές ουσίες. Οι ουσίες αυτές επιλέχθηκαν ώστε να σχετίζονται με διαφορετικούς τρόπους εκδήλωσης της νόσου *in vivo*. Εφόσον η πειραματική επιβεβαίωση των δεδομένων που θα προκύψουν από την ανάλυση θα γίνει σε αυτά τα μοντέλα, πρέπει να ληφθεί υπόψιν το πως επηρεάζουν οι ουσίες αυτές το κύτταρο σε μοριακό επίπεδο (περισσότερα για την επαγωγή της νόσου στην παράγραφο 2.2.2.). Αυτό που είναι σημαντικό σε αυτό το στάδιο είναι να βρεθούν τα μονοπάτια που επηρεάζουν τα στεατογενικά φάρμακα τα οποία επάγουν τη νόσο.

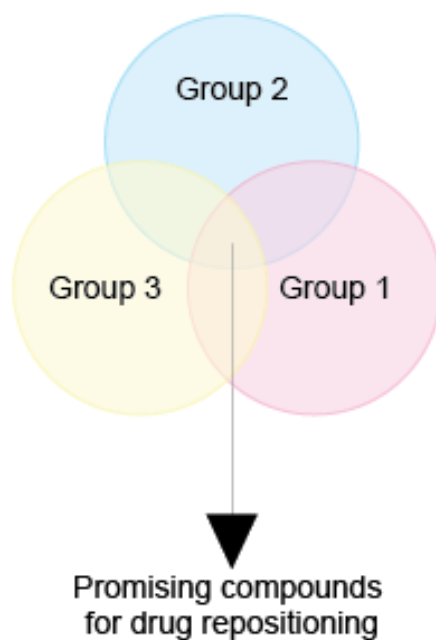
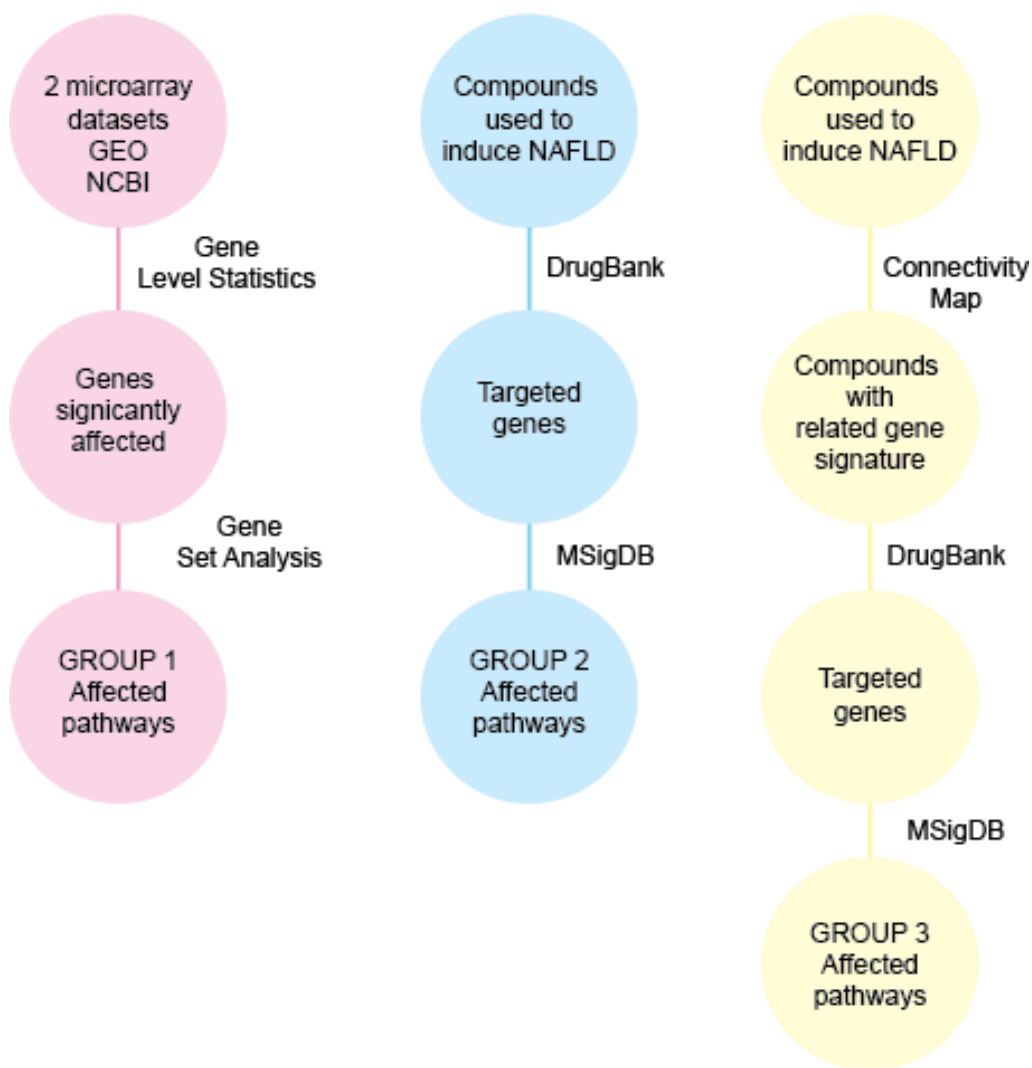
Αναζητώντας τα στεατογενικά φάρμακα στην Drugbank αντλούνται τα γονίδια στόχοι (drug target) καθώς και τα γονίδια των πρωτεϊνών που φέρουν και μεταφέρουν το φάρμακο εντός του κυττάρου (enzyme/ transporter/ carrier). Στη συνέχεια, μέσω της MSigDB καταγράφονται τα μονοπάτια στα οποία ανήκει το κάθε γονίδιο. Τα μονοπάτια αυτά αποτελούν την Ομάδα 2. Τέλος, συγκρίνονται με αυτά της Ομάδας 1 και επιλέγεται η τομή τους.

Στη συνέχεια, χρησιμοποιείται το εργαλείο του cMap που εντοπίζει ομοιότητες μεταξύ γονιδιακών υπογραφών. Ως υπογραφή-ερώτηση θέτονται τα στεατογενικά φάρμακα, το καθένα ξεχωριστά και βάζοντας τρεις διαφορετικές κυτταρικές σειρές (HL60, MCF7, PC3) και συγκέντρωση όσο πιο κοντινή γίνεται σε αυτή που χρησιμοποιείται πειραματικά για την επαγωγή της νόσου. Επιλέχθηκαν οι βιοδραστικές ουσίες που έχουν αντίθετη ($ES < 0$) ή παρόμοια ($ES > 0$) γονιδιακή έκφραση και με $p\text{-value} < 0.05$. Κάθε βιοδραστική ουσία

αναζητήθηκε στην Drugbank και μέσω της MSigDB βρέθηκαν τα μονοπάτια που αυτή επηρεάζει. Τα μονοπάτια αυτά αποτελούν την Ομάδα 3.

Τα στεατογενικά φάρμακα έχουν μια διαφορετική κύρια λειτουργία και έχουν ως παρενέργεια την NAFLD. Επομένως για να μην επιλεγούν βιοδραστικές ουσίες που αντιτίθενται στην κύρια λειτουργία της καθεμίας, το ενδιαφέρον επικεντρώνεται στις ουσίες που αφορούν σε μονοπάτια που σχετίζονται με τη νόσο. Επομένως επιλέγονται οι βιοδραστικές ουσίες που επηρεάζουν μονοπάτια που ανήκουν στην τομή των Ομάδων 1, 2 και 3 (Εικόνα 2.6.).

Από τις ουσίες που επιλέχθηκαν, ορισμένες από αυτές μπορεί να είναι ηπατοτοξικές ή μπορεί να προκαλούν NAFLD. Με τη βοήθεια των ToxDB, LiverTox και διάφορων άλλων βιβλιογραφικών αναφορών αποκλείονται οι ουσίες αυτές.



Εικόνα 2.6. Μεθοδολογία

2.2. Πειραματική Επαλήθευση

2.2.1. Cell Culture

Η ηπατοκυτταρική σειρά HuH7 καλλιεργήθηκε σε θρεπτικό υλικό Dubelccos Modified Eagles Medium (DMEM)/High glucose με προσθήκη 10% Fetal Bovine Serum (FBS) και 1% μείγμα πενικιλίνης/στρεπτομυκίνης (Penicillin/Streptomycin) (BIOSERA, Γαλλία), σε επωαστήρα 37°C, 5.0% CO₂, 90% υγρασία.

Τα κύτταρα επιστρώθηκαν σε πλάκες καλλιέργειας 96 θέσεων (96-well plates) με πυκνότητα 15.000 κύτταρα/θέση για τη διεξαγωγή των πειραμάτων.

2.2.2. Επαγωγή NAFLD

Η επαγωγή της Μη Αλκοολικής Λιπώδης Νόσου του ήπατος έγινε με κατεργασία των κυττάρων με μείγμα ολεικού και παλμιτικού οξέος (Cayman Chemicals, 90260, 10006627), Tamoxifen citrate (Cayman Chemicals, 11629), Amiodarone hydrochloride (Cayman Chemicals, 15213), Tetracycline hydrochloride (Cayman Chemicals, 14328) και Valproic Acid sodium salt (Cayman Chemicals, 13033).

Τα λιπαρά οξέα διαλύθηκαν σε 100% αιθανόλη σε θερμοκρασία 90°C, σύμφωνα με τις οδηγίες του παρασκευαστή. Παρασκευάστηκε μείγμα ολεικού και παλμιτικού οξέος σε μοριακή αναλογία 1:2 παλμιτικό οξύ:ολεικό οξύ. Τα κύτταρα εκτέθηκαν για 24 ώρες σε συγκέντρωση 500 μM του μείγματος σε 1% v/v στο θρεπτικό υλικό.

Οι φαρμακολογικές ουσίες Tamoxifen citrate, Amiodarone hydrochloride, Tetracycline hydrochloride διαλύθηκαν σε DMSO σύμφωνα με τις οδηγίες του παρασκευαστή. Το Valproic Acid sodium salt διαλύθηκε σε 100% αιθανόλη σε θερμοκρασία 90°C, σύμφωνα με τις οδηγίες του παρασκευαστή. Τα κύτταρα εκτέθηκαν στις συγκεντρώσεις του Πίνακα 2.1. για 24 ώρες 0,1% v/v DMSO και 1% αιθανόλη στο θρεπτικό υλικό.

Δραστική Ουσία	Συγκέντρωση (μM)
Tamoxifen citrate	8
Amiodarone hydrochloride	20
Tetracycline hydrochloride	150
Valproic acid sodium salt	6000

Πίνακας 2.1. Συγκεντρώσεις των στεατογενικών ουσιών

2.2.3. Πειράματα προκαταρκτικής αξιολόγησης βιοδραστικών ουσιών για τη θεραπεία της νόσου

Για την προκαταρκτική αξιολόγηση βιοδραστικών ουσιών για τη θεραπεία της νόσου επιλέχθηκαν ουσίες οι οποίες βρίσκονται στη βιβλιοθήκη φαρμακολογικών ουσιών LOPAC (LOPAC1280 – The Library of Pharmacologically Active Compounds, SigmaAldrich). Στη βιβλιοθήκη αυτή, οι φαρμακολογικές ουσίες βρίσκονται διαλυμένες σε συγκέντρωση 10mM σε DMSO.

Για την κατασκευή των καμπυλών δόσης-επιβίωσης οι ουσίες διαλύθηκαν περαιτέρω σε DMSO και προστέθηκαν στο θρεπτικό υλικό των κυττάρων 0,1% v/v. Μετά την επιλογή των κατάλληλων συγκεντρώσεων, που δεν μειώνουν τη βιωσιμότητα των κυττάρων κάτω από 90%, οι ουσίες αυτές προστέθηκαν σε κύτταρα στα οποία είχε προηγηθεί επαγωγή της NAFLD για 24 ώρες.

2.2.4. Μέθοδος Resazurin για τη μέτρηση κυτταρικής βιωσιμότητας

Το resazurin είναι ένας κυτταρικά διαπερατός οξειδοαναγωγικός δείκτης των κυττάρων που χρησιμοποιείται για την μέτρηση του αριθμού βιώσιμων κυττάρων. Μπορεί να διαλυθεί σε φυσιολογικούς διαλύτες (π.χ. PBS) σχηματίζοντας ένα σκούρο μπλε διάλυμα, και να προστεθεί απευθείας σε κυτταροκαλλιέργειες με ομογενή τρόπο. Ζωντανά κύτταρα με ενεργό μεταβολισμό μπορούν να ανάγουν το resazurin στο ροζ και φθορίζον resofurin.

Δεν απαιτείται προσθήκη ενός ενδιάμεσου φορέα ηλεκτρονίων για να πραγματοποιηθεί η αντίδραση, αλλά η προσθήκη θα μπορούσε να επιταχύνει την αντίδραση. Η ποσότητα resofurin που παράγεται είναι ανάλογη του αριθμού βιώσιμων κυττάρων και μπορεί να ποσοτικοποιηθεί με μέτρηση φθορισμού σε μήκος κύματος διέγερσης 560nm και μήκος κύματος εκπομπής 590nm. Ποσοτικοποίηση μπορεί να γίνει και με μέτρηση οπτικής απορρόφησης, αλλά παρουσιάζει πολύ μικρότερη ευαισθησία.

Ο χρόνος επώασης που απαιτείται για την παραγωγή επαρκούς φθορίζον σήματος είναι συνήθως 1-4 ώρες και εξαρτάται από την μεταβολική ενεργότητα του κάθε κυτταρικού τύπου, της πυκνότητας των κυττάρων και των συνθηκών καλλιέργειας, όπως το θρεπτικό υλικό. Ο χρόνος επώασης πρέπει συνεπώς να προσδιορίζεται πειραματικά και να είναι αρκετά μικρός ώστε να αποφευχθεί η τοξικότητα του αντιδραστήριου αλλά αρκετά μεγάλος ώστε να είναι αρκετή η ευαισθησία της μεθόδου. Περαιτέρω αναγωγή του resofurin οδηγεί σε σχηματισμό hydroresofurin, το οποίο είναι άχρωμο και μη-φθορίζον, και συνεπώς μπορεί να οδηγήσει σε εσφαλμένα αποτελέσματα.

Στην παρούσα μελέτη χρησιμοποιήθηκε Resazurin sodium salt (SIGMA, R7017) διαλυμένο αρχικά σε Dulbecco's Phosphate Buffered Saline pH 7.4 (Biosera, PM-B2092), και περαιτέρω αραιωμένο σε θρεπτικό υλικό πριν από κάθε πείραμα. Μετά από την 24ωρη έκθεση των κυττάρων στις επαγωγικές ουσίες, αντικαταστάθηκε το θρεπτικό υλικό και προστέθηκε Resazurin 60μg/mL. Έγινε επώαση για 2 ώρες σε επωαστήρα 37°C, 5.0% CO₂, 90%

υγρασία και μετρήθηκε ο φθορισμός σε μήκη κύματος Ex560nm/Em590nm στο Varioskan™ LUX multimode microplate reader (Thermo Scientific™). Η βιωσιμότητα των κυττάρων εκφράστηκε ως ποσοστό των εκτεθειμένων στις επαγωγικές ουσίες κυττάρων προς τα μη εκτεθειμένα [44].

Η μέθοδος αυτή χρησιμοποιήθηκε για την κατασκευή καμπύλης δόσης-επιβίωσης για τις ουσίες που χρησιμοποιήθηκαν για την θεραπεία της NAFLD, με σκοπό να προσδιοριστούν οι συγκεντρώσεις που δεν μειώνουν τη βιωσιμότητα κάτω από 90% (EC₉₀). Για το σκοπό αυτό χρησιμοποιήθηκε το λογισμικό GraphPad Prism 6.0 [45]. Στο λογισμικό αυτό χρησιμοποιείται λογισμικό μοντέλο παλινδρόμησης τεσσάρων παραμέτρων. Πρώτα υπολογίζεται η τιμή EC₅₀ και από αυτή, η τιμή EC₉₀ σύμφωνα με την παρακάτω σχέση:

$$EC_F = EC_{50} * \left(\frac{F}{100 - F} \right)^{\frac{1}{HillSlope}}$$

2.2.5. High Content Screening των σταγονιδίων λιπιδίων

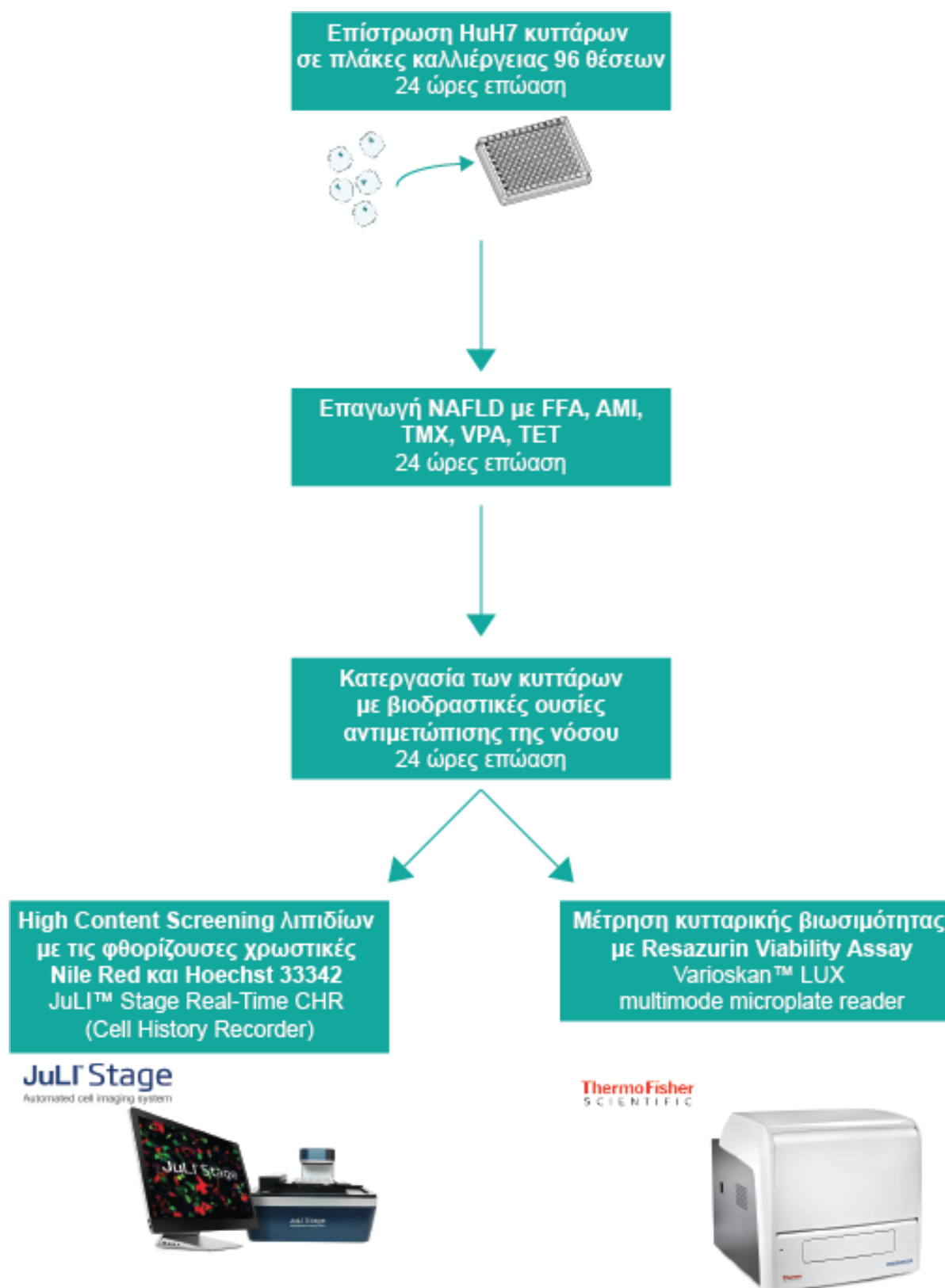
Το ενδοκυτταρικό φορτίου λίπους προσδιορίστηκε με μικροσκοπία φθορισμού μετά από χρώση των ενδοκυτταρικών σταγονιδίων λιπιδίων με την κόκκινη φθορίζουσα χρωστική Nile Red (Thermo Fisher Scientific, N1142). Ως αντιχρωματισμός χρησιμοποιήθηκε η μπλε φθορίζουσα χρωστική Hoechst 33342 (Thermo Fisher Scientific, H3570).

Το Nile Red χρησιμοποιείται για τον εντοπισμό και την ποσοτικοποίηση λιπιδίων, ειδικά ουδέτερων σταγονιδίων λιπιδίων μέσα στο κύτταρο. Είναι σχεδόν μη-φθορίζον στο νερό και σε άλλους πολικούς διαλύτες αλλά φθορίζει σε μη-πολικά περιβάλλοντα σε μήκος κύματος διέγερσης 552nm και μήκος κύματος εκπομπής 636nm.

Το Hoechst 33342 είναι κυτταρικά διαπερατή χρωστική που εκπέμπει μπλε φθορισμό όταν προσδεθεί στο dsDNA, ενώ έχει μήκος κύματος διέγερσης 392nm και μήκος κύματος εκπομπής 440nm.

Μετά την έκθεση των κυττάρων στις επαγωγικές ουσίες για 24 ώρες και στις βριοδραστικές ουσίες αντιστροφής της νόσου για άλλες 24 ώρες, έγινε αφαίρεση του θρεπτικού υλικού και πλύση των κυττάρων με PBS. Στη συνέχεια τα κύτταρα εκτέθηκαν για 45 λεπτά σε διάλυμα χρώσης αποτελούμενο από Nile Red 2μg/mL και Hoechst 33342 5μg/mL διαλυμένα σε θρεπτικό υλικό. Μετά την επώαση έγινε αφαίρεση του θρεπτικού υλικού, πλύση των κυττάρων με PBS και προσθήκη φρέσκου θρεπτικού υλικού.

Η λήψη των εικόνων έγινε με χρήση του JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek), με χρήση των καναλιών DAPI (Excitation 390/40, Emission 452/45) και RFP (Excitation 525/50, Emission 580) και 20x οπτικού φακού.



Εικόνα 2.7. Πειραματική διαδικασία

Κεφάλαιο 3

Αποτελέσματα

3.1. Υπολογιστική Ανάλυση

3.1.1. Μικροσυστοιχίες από τη GEO

Από την αναζήτηση στη GEO επιλέχθηκαν τέσσερις μελέτες θέτοντας ως κριτήρια να αφορούν στον άνθρωπο (*Homo sapiens sp.*) και να περιλαμβάνουν δεδομένα από υγιείς και ασθενείς. Τα βασικά στοιχεία για τα σετ δεδομένων που επιλέχθηκαν παρουσιάζονται στον Πίνακα 3.1..

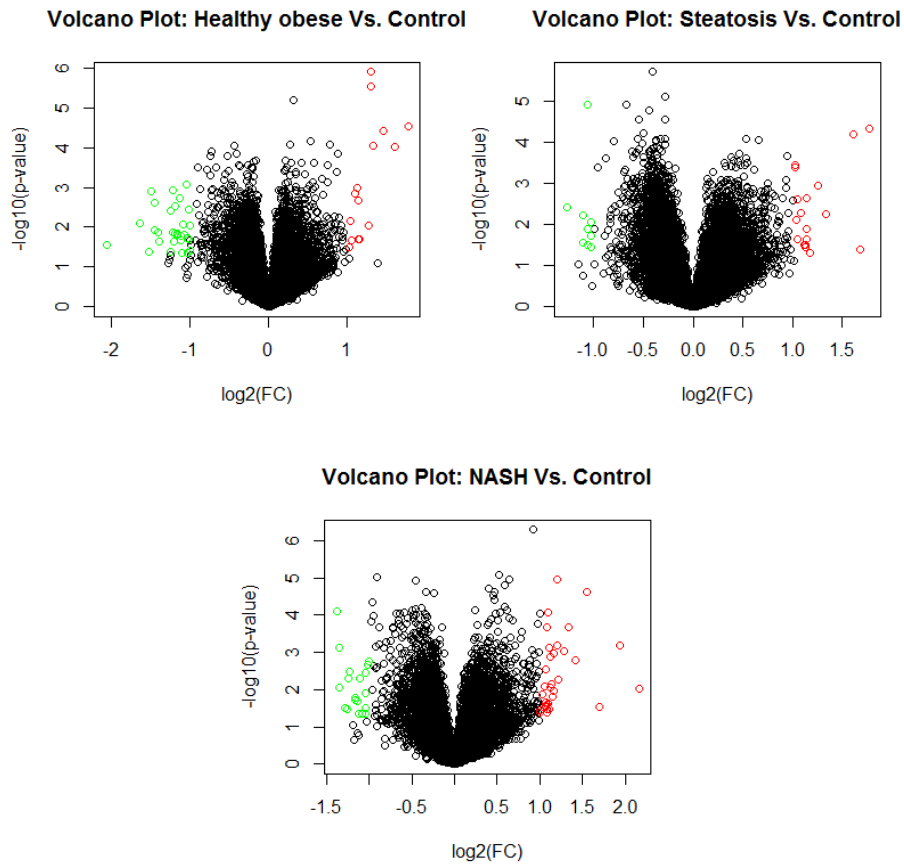
Αύξων αριθμός GEO	Συγγραφείς	Αριθμός υγιών δειγμάτων	Αριθμός ασθενών δειγμάτων	Είδος δείγματος	Παθολογικό στάδιο της νόσου NAFLD
GSE48452	Jochen H et al.	14	59	Βιοψία ήπατος	Healthy obese, NASH, Steatosis
GSE63067	Frades I	8	11	Βιοψία ήπατος	NASH, Steatosis
GSE72756	Chuanzheng S et al.	5	5	Βιοψία ήπατος	NAFLD
GSE89632	Allard JP et al.	24	39	Βιοψία ήπατος	NASH, Steatosis

Πίνακας 3.1. Σετ δεδομένων από την GEO

3.1.2. Επιλογή σετ δεδομένων με τη χρήση Volcano plot

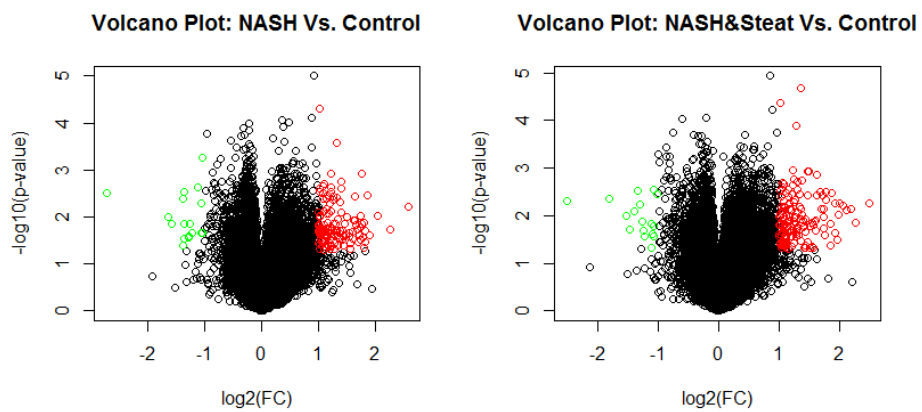
Τα σετ δεδομένων που παρουσιάστηκαν παραπάνω αντιστοιχούν στα εξής Volcano plots.

GSE48452



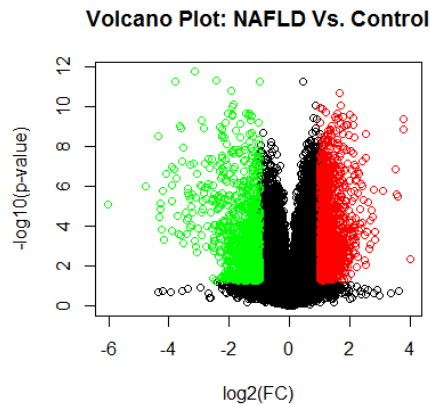
Εικόνα 3.1. Volcano plot για τη GSE48452. Πάνω αριστερά: σύγκριση Healthy Obese Vs. Control, πάνω δεξιά: σύγκριση Steatosis Vs. Control, κάτω: σύγκριση NASH Vs. Control

GSE63067



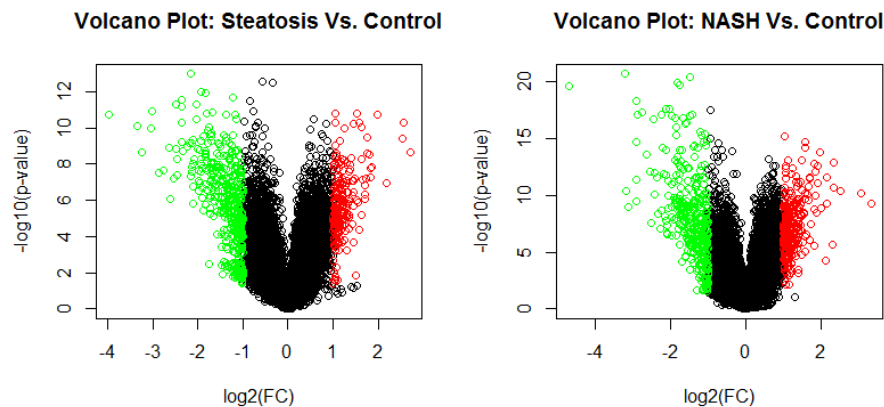
Εικόνα 3.2. Volcano plot για τη GSE63067. Αριστερά: σύγκριση NASH Vs. Control, δεξιά: σύγκριση NASH & Steatosis Vs. Control

GSE72756



Εικόνα 3.3. Volcano plot για τη GSE72756, σύγκριση NAFLD Vs. Control

GSE89632



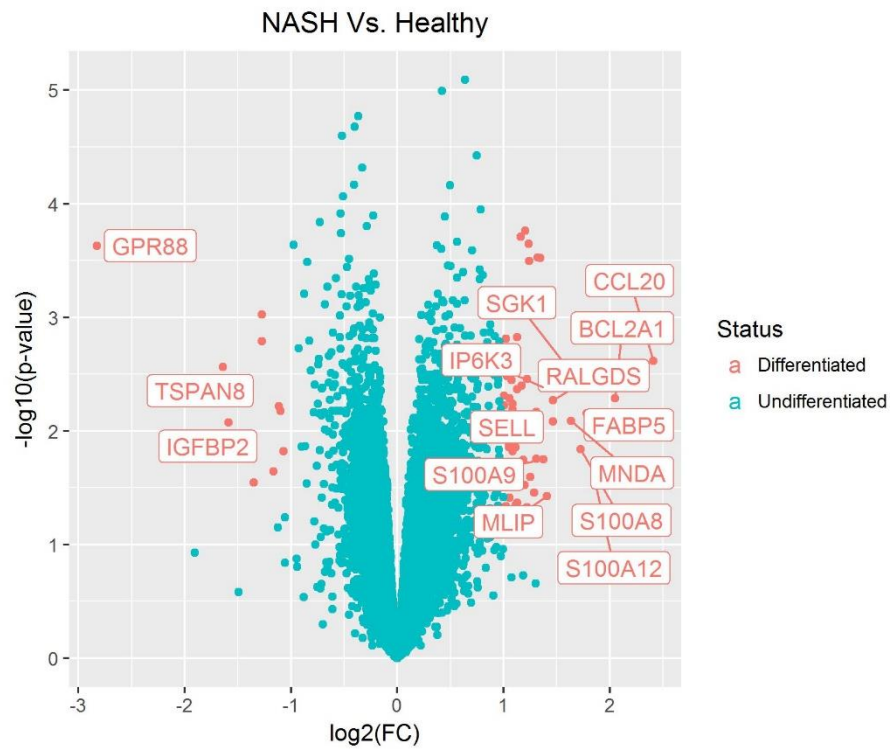
Εικόνα 3.4. Volcano plot για τη GSE89632, Αριστερά: σύγκριση Steatosis Vs. Control, δεξιά: σύγκριση NASH Vs. Control

Από αυτές, επιλέγονται οι GSE63067 και GSE89632 καθώς έχουν ικανοποιητικό πλήθος δειγμάτων και διαφορεικά εκφρασμένων γονιδίων, σε αντίθεση με τις GSE48542 και GSE72756.

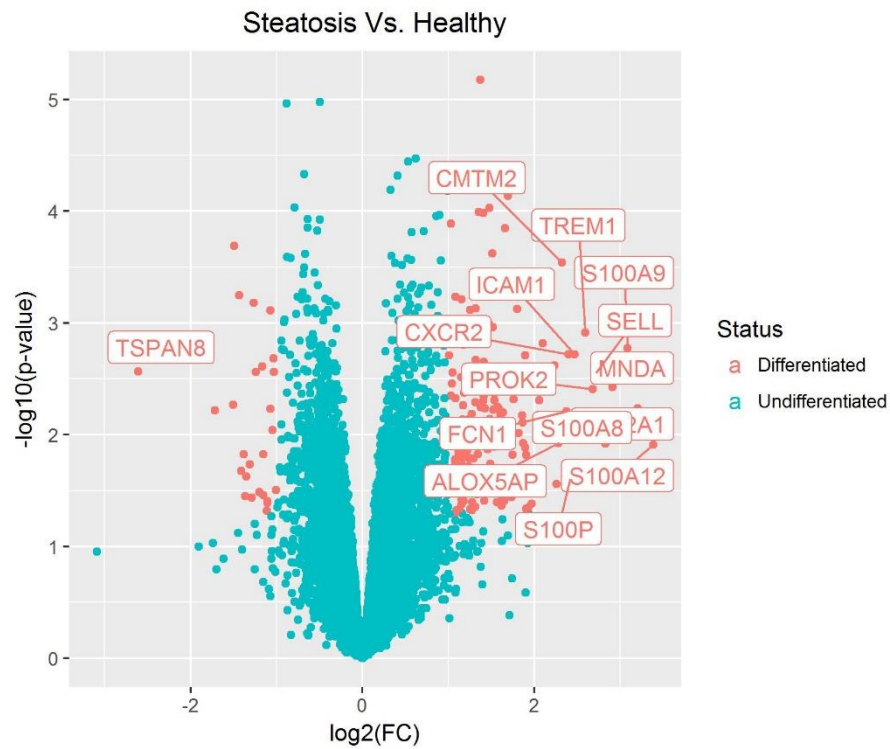
3.1.3. Ανάλυση μικροσυστοιχιών σε επίπεδο γονιδίων (GLS)

Κάνοντας την ανάλυση σε επίπεδο γονιδίων, για κάθε σετ δεδομένων και για κάθε σύγκριση, υπολογίζονται τα διαφορετικά εκφρασμένα γονίδια. Τα αποτελέσματα παρουσιάζονται στη μορφή Volcano plot, σημειώνοντας τα 15 πιο στατιστικά σημαντικά διαφορετικά εκφρασμένα γονίδια σε κάθε περίπτωση.

GSE63067

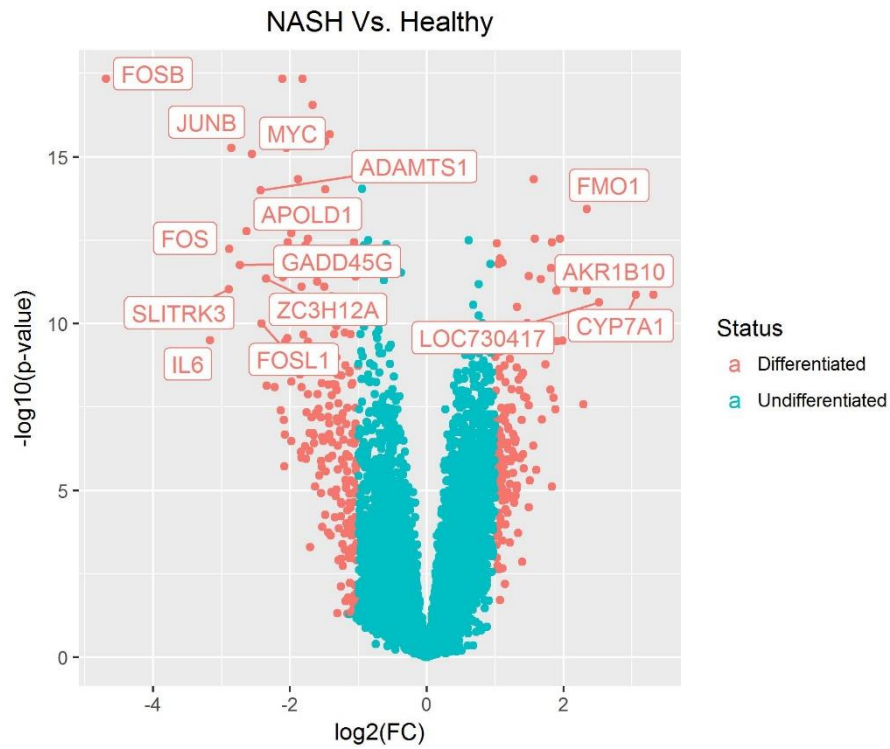


Εικόνα 3.5. Volcano plot για τη GSE63067 και τη σύγκριση NASH Vs Healthy

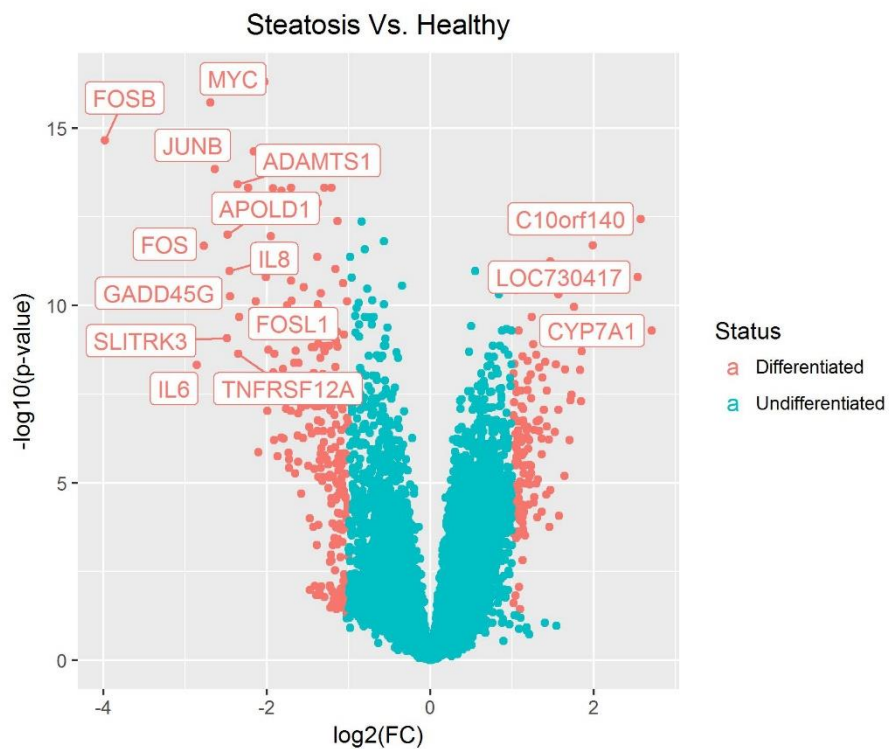


Εικόνα 3.6. Volcano plot για τη GSE63067 και τη σύγκριση Steatosis Vs Healthy

GSE89632



Εικόνα 3.7. Volcano plot για τη GSE89632 και τη σύγκριση NASH Vs Healthy

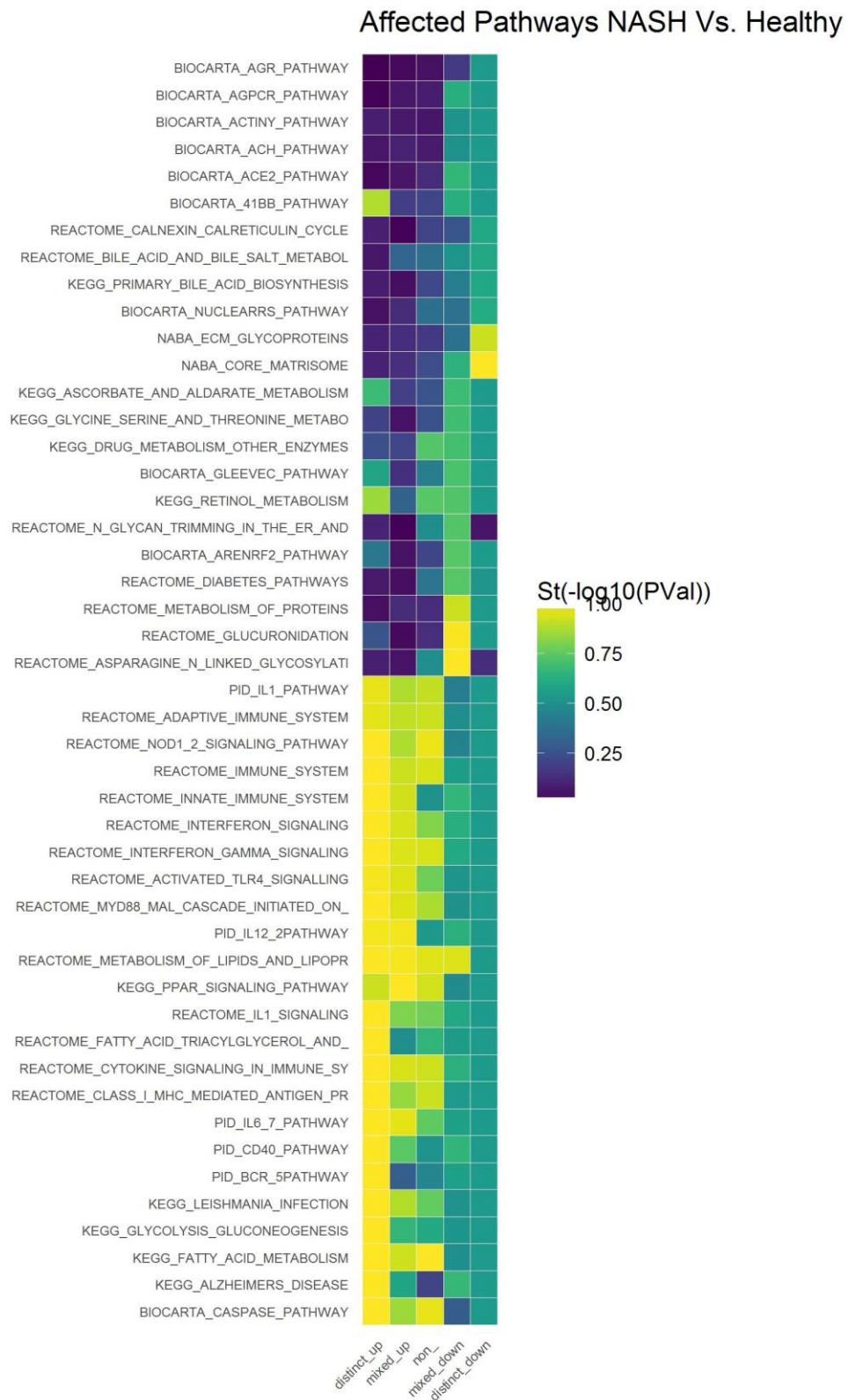


Εικόνα 3.8. Volcano plot για τη GSE89632 και τη σύγκριση Steatosis Vs Healthy

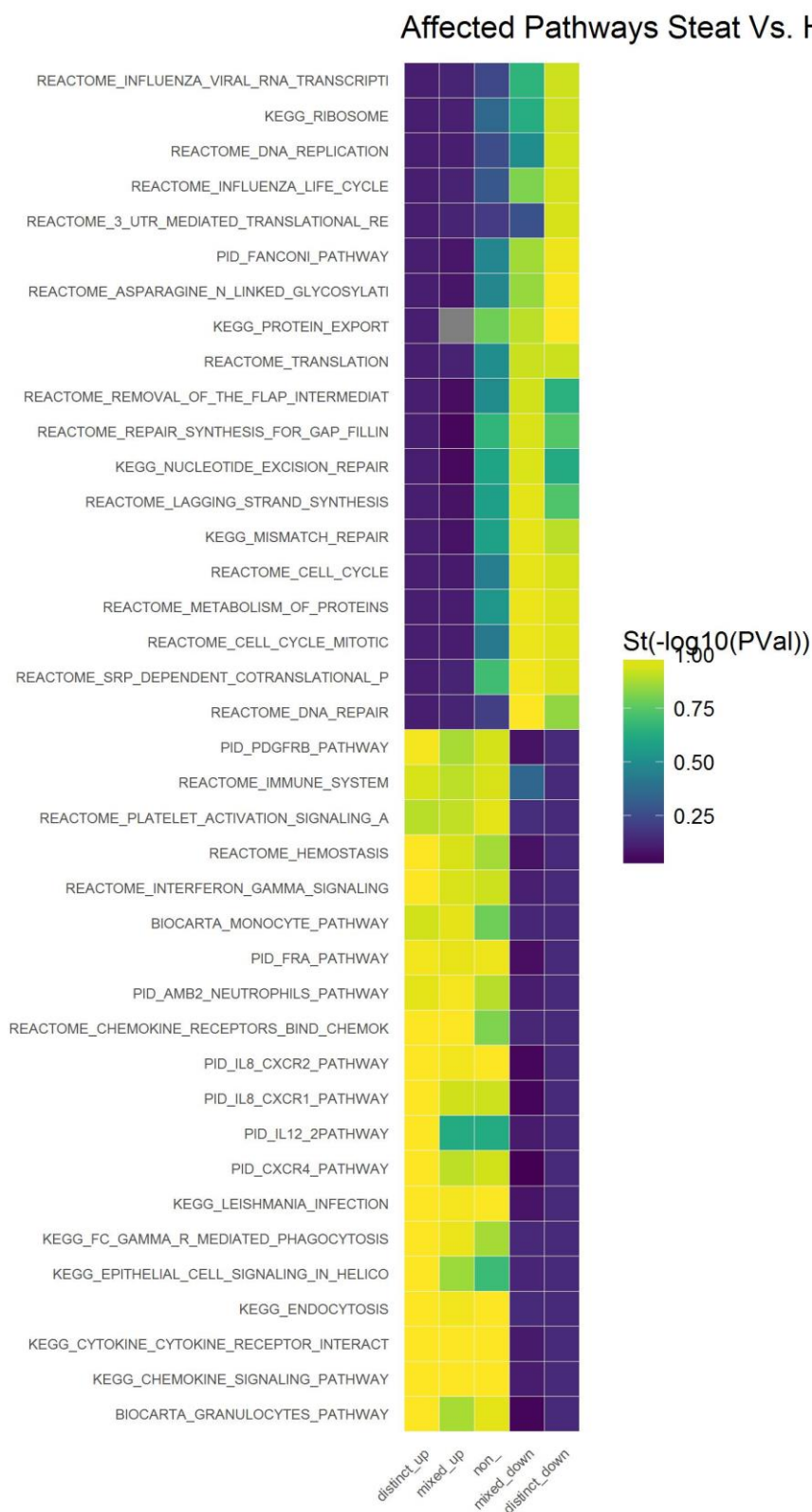
3.1.4. Ανάλυση μικροσυστοιχιών σε επίπεδο μονοπατιών (GSA)

Εκτελώντας την ανάλυση σε επίπεδο μονοπατιών, υπολογίζονται τα πιο σημαντικά στατιστικά μονοπάτια χωρισμένα στις πέντε κατηγορίες που αναφέρθηκαν παραπάνω. Τα αποτελέσματα παρουσιάζονται σε μορφή θερμικού χάρτη (heatmap). Στο χάρτη αυτό υπάρχουν πέντε στήλες, όπου η κάθε μια αντιστοιχεί σε μια από τις πέντε κατηγορίες (distinct up, mixed up, non directional, mixed down, distinct down). Στον οριζόντιο άξονα παρουσιάζονται τα ονόματα των στατιστικά σημαντικότερων μονοπατιών για τη νόσο. Το χρώμα που έχει κάθε μονοπάτι για κάθε κατηγορία αντιστοιχεί στη στατιστική του σημαντικότητα, όπου όσο πιο ανοιχτόχρωμα είναι, τόσο πιο υψηλή η στατιστική σημαντικότητα.

GSE63067

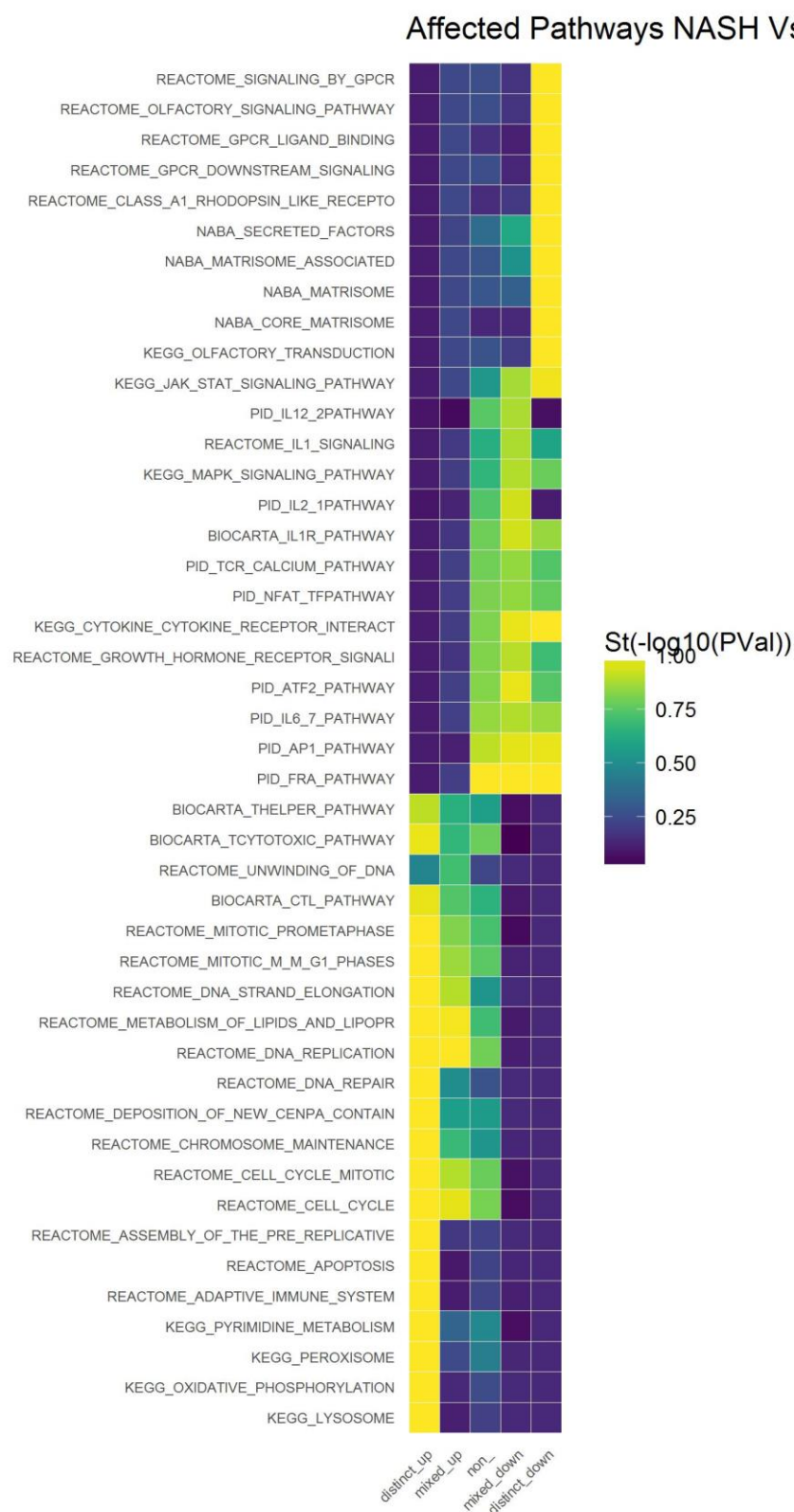


Εικόνα 3.9. Θερμικός χάρτης στατιστικά σημαντικών μονοπατιών για τη GSE63067 και τη σύγκριση NASH Vs Healthy

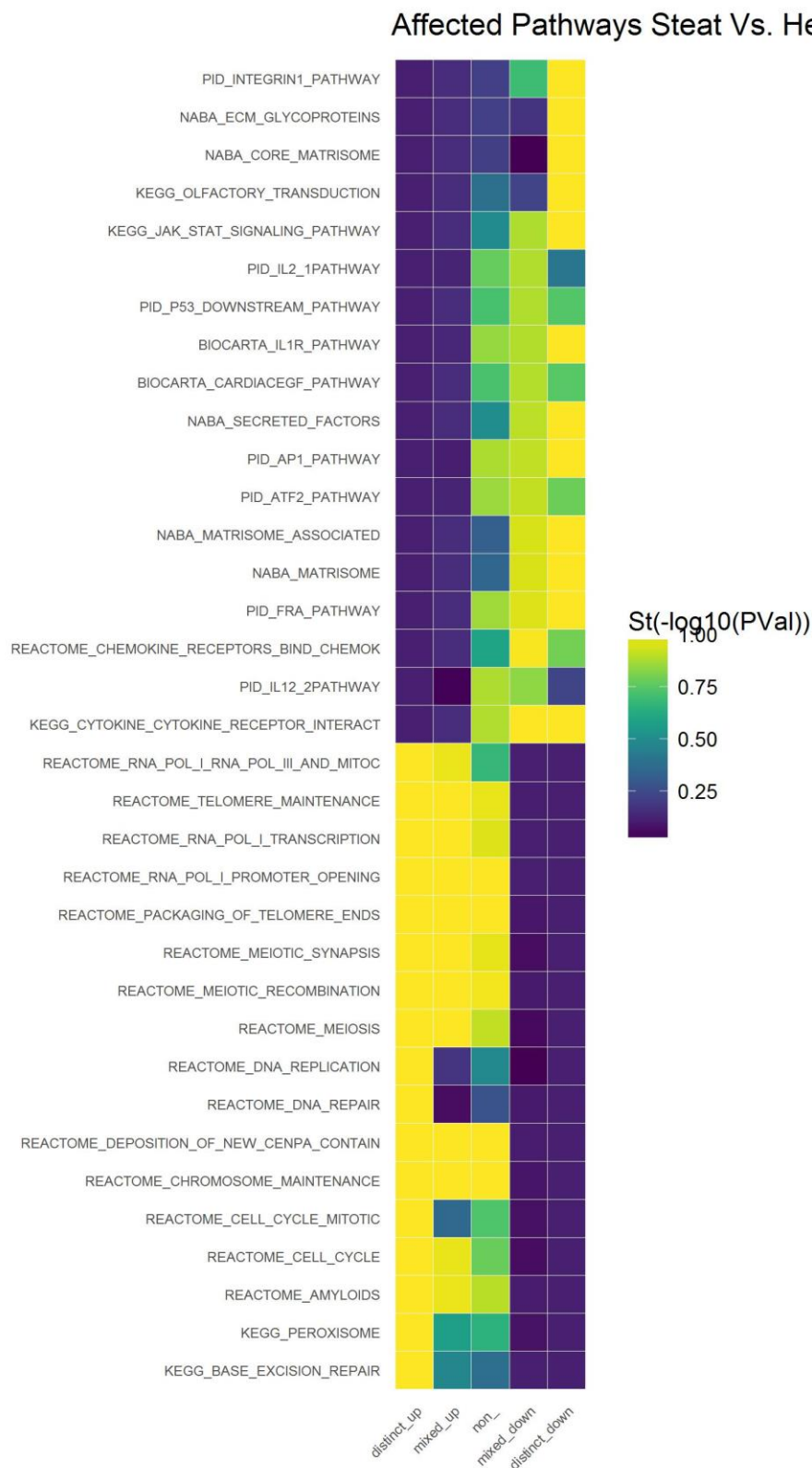


Εικόνα 3.10. Θερμικός χάρτης στατιστικά σημαντικών μονοπατιών για τη GSE63067 και τη σύγκριση Steatosis Vs Healthy

GSE89632



Εικόνα 3.11. Θερμικός χάρτης στατιστικά σημαντικών μονοπατιών για τη GSE89632 και τη σύγκριση NASH Vs Healthy



Εικόνα 3.12. Θερμικός χάρτης στατιστικά σημαντικών μονοπατιών για τη GSE89632 και τη σύγκριση Steatosis Vs Healthy

3.1.5. Προσδιορισμός των μονοπατιών που επηρεάζουν οι στεατογενικές φαρμακολογικές ουσίες

Χρησιμοποιώντας τη Drugbank και την MSigDB προσδιορίζονται τα μονοπάτια που επηρεάζουν οι στεατογενικές φαρμακολογικές ουσίες. Στον παρακάτω πίνακα παρουσιάζονται τα μονοπάτια που ανήκουν στην τομή των Ομάδων 1 και 2 (βλ. Παράγραφο 2.1.9.).

Φάρμακο	Γονίδιο	Μονοπάτια
Valproic acid	ACADSB	KEGG_FATTY_ACID_METABOLISM
		REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES
	OGDH	REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES
	PPARA	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
		REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
	PPARD	REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM
		KEGG_PPAR_SIGNALING_PATHWAY
	PPARG	BIOCARTA_NUCLEARRS_PATHWAY
		KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
		PID_NFAT_TFPATHWAY
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS		
	REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM	
Amiodarone	ADRB1	KEGG_ENDOCYTOSIS
	PPARG	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
		PID_NFAT_TFPATHWAY
		REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM		
Tetracycline	-	-
Tamoxifen	ESR1	PID_ATF2_PATHWAY
		PID_AP1_PATHWAY
	EBP	REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
Oleic acid	PPARA	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY

Oleic acid		REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
		REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM
	PPARD	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
	PPARG	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
		PID_NFAT_TFPATHWAY
		REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
	RXRA	REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM
		KEGG_PPAR_SIGNALING_PATHWAY
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS		
Palmitic acid	PPT1	KEGG_LYSOSOME
	PPARA	KEGG_PPAR_SIGNALING_PATHWAY
		BIOCARTA_NUCLEARRS_PATHWAY
		REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
		REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM

Πίνακας 3.2. Τα γονίδια που επηρεάζουν τα στεατογενικά φάρμακα και τα μονοπάτια στα οποία ανήκουν

3.1.6. Εύρεση βιοδραστικών ουσιών για αντιστροφή της νόσου με τη χρήση του cMap

Χρησιμοποιώντας ως είσοδο στο cMap τα στεατογενικά φάρμακα και με τη βοήθεια των Drugbank και MSigDB βρίσκονται οι βιοδραστικές ουσίες που σχετίζονται με μονοπάτια που ανήκουν στην τομή των Ομάδων 1 και 2 (βλ. Παράγραφο 2.1.9.). Παρακάτω παρουσιάζονται για τα πέντε στεατογενικά φάρμακα, οι υπογραφές-ερωτήσεις και τα ουσίες που επηρεάζουν μονοπάτια που ανήκουν στην τομή των Ομάδων 1 και 2, όπως εμφανίζονται στο εργαλείο cMap.

Στους πίνακες με τις υπογραφές-ερωτήσεις για κάθε φάρμακο φαίνονται τα τρία παραδείγματα που αποτελούν την υπογραφή-ερώτηση. Πιο συγκεκριμένα, παρουσιάζονται η δραστική ουσία, η δόση, η κυτταρική σειρά και ο κωδικός του παραδείγματος.

Στους πίνακες με τις βιοδραστικές ουσίες, η είναι ο αριθμός των παραδειγμάτων για την ουσία αυτή και rank η σειρά κατάταξης τους με βάση το

p value. Το μέγεθος enrichment παίρνει τιμές στο διάστημα [-1,1] και δείχνει πόσο μοιάζει η γονιδιακή έκφραση της συγκεκριμένης ουσίας με αυτή της υπογραφής-ερώτησης. Θετικές τιμές αντιστοιχούν σε παρόμοια γονιδιακή υπογραφή, ενώ αρνητικές σε αντίθετη. Η τιμή p value είναι μια εκτίμηση της πιθανότητας η τιμή του μεγέθους enrichment να είναι τυχαία. Το μέγεθος specificity είναι μια εκτίμηση της μοναδικότητας της σύνδεσης μεταξύ της υπογραφής-ερώτησης και της συγκεκριμένης ουσίας. Υπολογίζεται ξεχωριστά για θετικά και αρνητικά enrichment, ενώ υψηλές τιμές δείχνουν ότι η σύνδεση αυτή είναι συνηθισμένη. Τέλος, το μέγεθος % Non-null δείχνει ποιο ποσοστό των παραδειγμάτων που υπάρχουν για τη συγκεκριμένη ουσία συμφωνούν με την κατεύθυνση του enrichment.

Cmap name	Dose	Cell line	Instance id
Amiodarone	6mM	HL60	2434
Amiodarone	6mM	PC3	4657
Amiodarone	6mM	MCF7	3296

Πίνακας 3.3. Υπογραφή-ερώτηση για AMI

Rank	Cmap name	Mean	n	Enrichment	P	Specificity	% non null
39	Ketoconazole	-0.408	4	-0.689	0.0213	0.0000	75
41	Aminoglutethimide	-0.280	3	-0.781	0.02157	0.0179	66
48	Fusidic acid	-0.406	4	-0.666	0.02801	0.0063	75

Πίνακας 3.4. Βιοδραστικές ουσίες με αντίθετη γονιδιακή έκφραση από το AMI

Rank	Cmap name	Mean	n	Enrichment	P	Specificity	% non null
8	Ethisterone	0.432	6	0.698	0.00205	0.0000	8
14	Meclizine	0.412	5	0.709	0.00491	0.0131	80
24	Omeprazole	0.379	4	0.739	0.00889	0.0642	100
28	Trimetazidine	0.338	4	0.720	0.01243	0.0078	100
51	Prednisone	0.282	5	0.607	0.02842	0.0526	80
55	Amiloride	0.340	5	0.597	0.03276	0.0167	80
65	Ursodeoxycholic acid	0.390	3	0.725	0.04054	0.0517	100

Πίνακας 3.5. Βιοδραστικές ουσίες με παρόμοια γονιδιακή έκφραση με το AMI

Cmap name	Dose	Cell line	Instance id
Tamoxifen	7μM	HL60	1366
Tamoxifen	7μM	PC3	2050
Tamoxifen	7μM	MCF7	2212

Πίνακας 3.6. Υπογραφή-ερώτηση για TMX

Rank	Cmap name	Mean	N	Enrichment	P	Specificity	% non null
50	Carbimazole	-0.396	3	-0.792	0.01827	0.0480	66
60	Sulindac	-0.342	7	-0.524	0.02499	0.0388	57

Πίνακας 3.7. Βιοδραστικές ουσίες με αντίθετη γονιδιακή έκφραση από το TMX

Rank	Cmap name	Mean	N	Enrichment	P	Specificity	% non null
5	Sirolimus	0.215	4	0.330	0.00012	0.2590	65
23	Disulfiram	0.357	5	0.709	0.00495	0.0694	80
44	Ivermectin	0.321	5	0.651	0.01442	0.1075	80
46	Clomipramine	0.410	4	0.705	0.01578	0.1105	100
49	Albendazole	0.390	3	0.793	0.01815	0.0116	100
69	Clotrimazole	0.263	5	0.597	0.03258	0.1333	80
85	Mifepristone	0.241	4	0.629	0.04768	0.0584	75

Πίνακας 3.8. Βιοδραστικές ουσίες με παρόμοια γονιδιακή έκφραση με το TMX

Cmap name	Dose	Cell line	Instance id
Tetracycline	8μM	HL60	1397
Tetracycline	8μM	PC3	2080
Tetracycline	8μM	MCF7	2243

Πίνακας 3.9. Υπογραφή-ερώτηση για TET

Rank	Cmap name	Mean	N	Enrichment	P	Specificity	% non null
31	Raloxifen	-0.380	7	-0.519	0.02697	0.0520	57
54	Metoprolol	-0.417	4	-0.624	0.04953	0.0534	75

Πίνακας 3.10. Βιοδραστικές ουσίες με αντίθετη γονιδιακή έκφραση από το TET

Rank	Cmap name	Mean	N	Enrichment	P	Specificity	% non null
16	Aceclofenac	0.720	4	0.727	0.01108	0.0137	75
39	Cefotaxime	0.675	5	0.592	0.03475	0.0136	80

Πίνακας 3.11. Βιοδραστικές ουσίες με παρόμοια γονιδιακή έκφραση με το TET

Cmap name	Dose	Cell line	Instance id
Valproic acid	1mM	HL60	2669
Valproic acid	1mM	PC3	458
Valproic acid	1mM	MCF7	989

Πίνακας 3.12. Υπογραφή-ερώτηση για VPA

Rank	Cmap name	Mean	n	Enrichment	P	Specificity	% non null
63	Ketoprofen	-0.370	6	-0.525	0.04660	0.0385	50

Πίνακας 3.13. Βιοδραστικές ουσίες με αντίθετη γονιδιακή έκφραση από το VPA

Rank	Cmap name	Mean	n	Enrichment	P	Specificity	% non null
8	Resveratrol	0.294	9	0.655	0.00026	0.1226	77
36	Nabumetone	0.267	4	0.722	0.01182	0.0130	100
60	Mepacrine	0.330	2	0.860	0.03941	0.0677	100

Πίνακας 3.14. Βιοδραστικές ουσίες με παρόμοια γονιδιακή έκφραση με το VPA

Τελικά, τα μονοπάτια που βρίσκονται στην τομή των τριών Ομάδων είναι τα δέκα που εμφανίζονται παρακάτω. Στο Παράρτημα Α παρουσιάζεται ένα Volcano Plot για το καθένα, όπου φαίνεται η έκφραση των γονιδίων και ποια είναι τα διαφορικά εκφρασμένα.

Μονοπάτι πυρηνικών υποδοχέων (BIOCARTA_NUCLEARRS_PATHWAY), μονοπάτι της ενδοκύτωσης (KEGG_ENDOCYTOSIS), μονοπάτι μεταβολισμού λιπαρών οξέων (KEGG_FATTY_ACID_METABOLISM), σηματοδοτικό μονοπάτι των PPAR (KEGG_PPAR_SIGNALING_PATHWAY), μονοπάτι της Activator Protein 1 (PID_AP1_PATHWAY), μονοπάτι του Activating Transcription Factor 2 (PID_ATF2_PATHWAY), μονοπάτι των πυρηνικών παραγόντων ενεργοποιημένων T- κυττάρων (PID_NFAT_TFPATHWAY), μονοπάτι μεταβολισμού λιπαρών οξέων, τριακυλογλυκερόλης και κετονοσωμάτων (REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM), μονοπάτι μεταβολισμού των αμινοξέων και των παραγώγων τους (REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES), μονοπάτι μεταβολισμού λιπιδίων και λιποπρωτεϊνών (REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS)

3.1.7. Τελική επιλογή βιοδραστικών ουσιών

Από όλες τις βιοδραστικές ουσίες που έχουν επιλεγθεί στο προηγούμενο βήμα, χρησιμοποιώντας τις διαδικτυακές βάσεις ToxDB, LiverTox, και άλλες βιβλιογραφικές αναφορές απορρίπτονται οι ουσίες οι οποίες είτε είναι ηπατοτοξικές είτε έχει αποδειχθεί πειραματικά ότι προκαλούν NAFLD. Οι υπόλοιπες ουσίες είναι αυτές που σύμφωνα με την ανάλυση που προηγήθηκε είναι πιθανόν να μπορούν να αντιστρέψουν τη νόσο.

3.2. Πειραματική Επαλήθευση

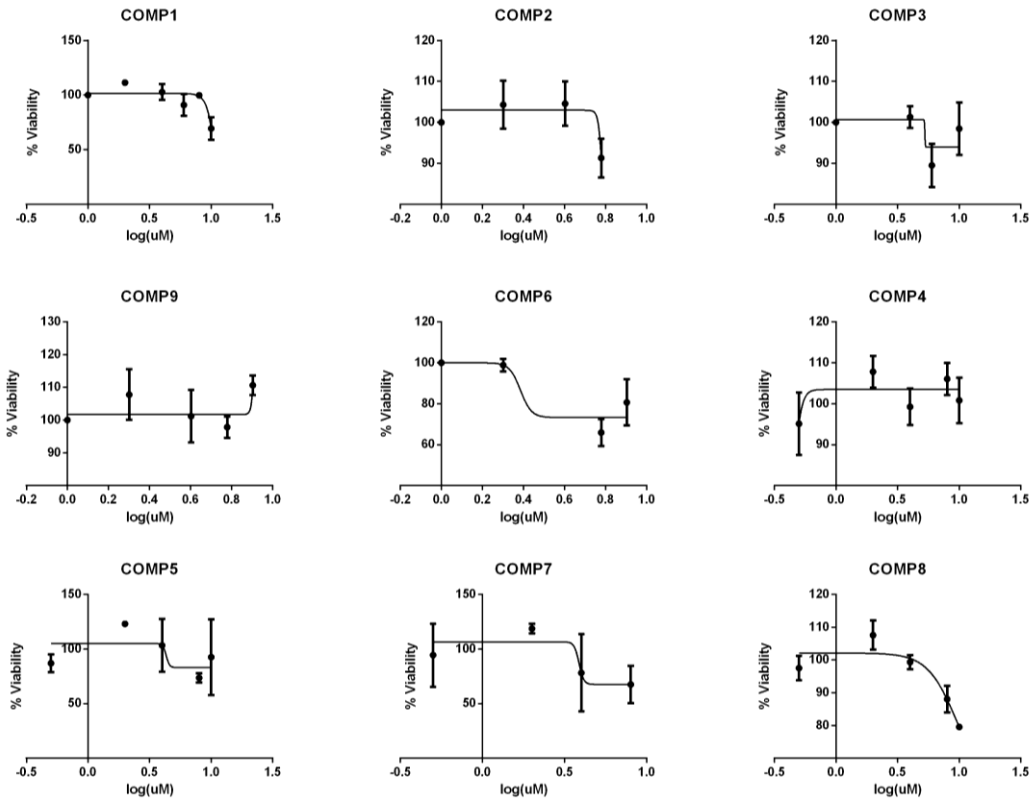
Από τις βιοδραστικές ουσίες της προηγούμενης ενότητας, επιλέχθηκαν εννέα, για να γίνει προκαταρκτική αξιολόγηση. Τα ουσίες αυτές παρουσιάζονται στον Πίνακα 3.15..

Βιοδραστικές ουσίες
COMP1
COMP2
COMP3
COMP4
COMP5
COMP6
COMP7
COMP8
COMP9

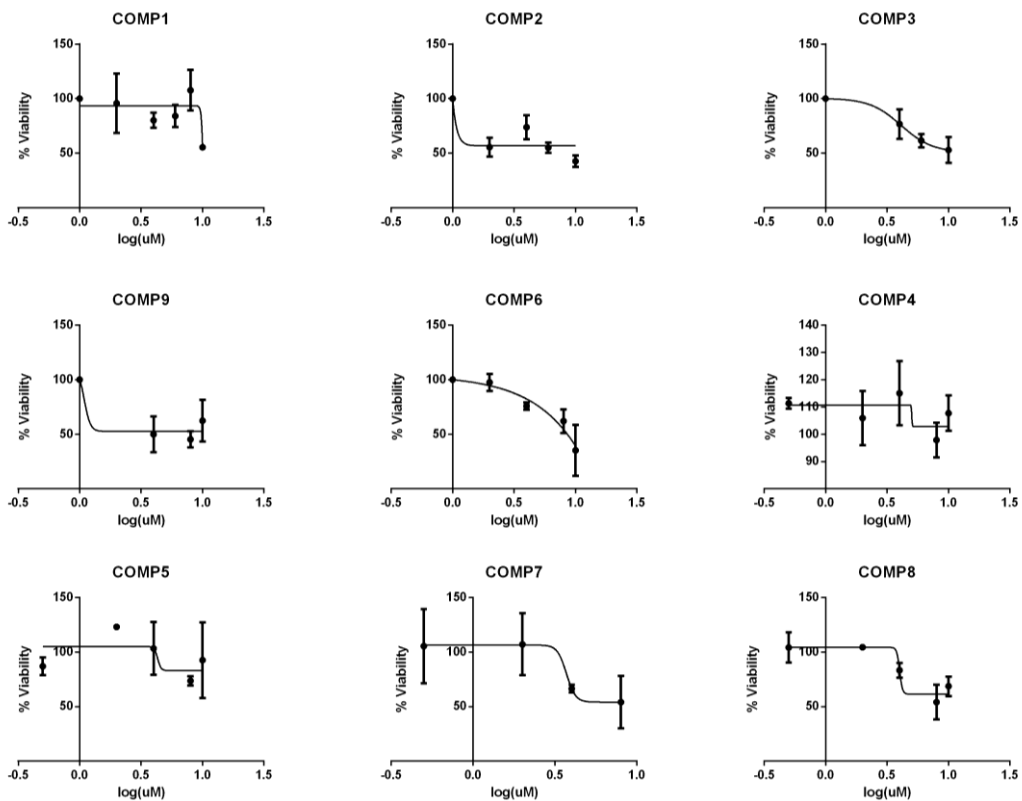
Πίνακας 3.15. Βιοδραστικές ουσίες που δοκιμάστηκαν πειραματικά

3.2.1. Καμπύλες δόσης-επιβίωσης

Χρησιμοποιώντας τα αποτελέσματα από τη μέθοδο Resazurin για τη μέτρηση της κυτταρικής βιωσιμότητας κατασκευάζονται οι καμπύλες δόσης επιβίωσης για τις ουσίες του Πίνακα 3.15.. Η πρώτη μέτρηση έγινε μετά από 24 ώρες και η δεύτερη μετά από 48 ώρες. Χρησιμοποιώντας το λογισμικό GraphPad Prism 6.0 προκύπτουν οι Εικόνες 3.13. και 3.14.. Στον οριζόντιο άξονα είναι ο δεκαδικός λογάριθμος της συγκέντρωσης σε μM και στον κάθετο άξονα είναι το ποσοστό της βιωσιμότητας. Στον Πίνακα 3.16. φαίνονται οι τιμές EC_{90} και στον Πίνακα 3.17. φαίνονται οι συγκεντρώσεις που χρησιμοποιήθηκαν στη συνέχεια των πειραμάτων.



Εικόνα 3.13. Καμπύλες δόσης-επιβίωσης για τις βιοδραστικές ουσίες του Πίνακα 3.16. μετά από 24 ώρες



Εικόνα 3.14. Καμπύλες δόσης-επιβίωσης για τις βιοδραστικές ουσίες του Πίνακα 3.16. μετά από 48 ώρες

Drug	Treatment	EC50	LogEC50	HillSlope	R square	logEC90	EC90
COMP1	24h	14.66	1.17	-11.12	0.71	1.08	12.03
	48h	11.69	1.07	-51.48	0.52	1.05	11.20
COMP2	24h	6.11	0.79	-45.85	0.60	0.76	5.82
	48h	0.59	-0.23	-13.09	0.70	-0.30	0.50
COMP3	24h	5.27	0.72	-267.7	0.33	0.72	5.23
	48h	4.12	0.62	-3.393	0.84	0.33	2.16
COMP4	24h	0.46	-0.34	14.52	0.28	-0.27	0.54
	48h	5.03	0.70	-186.9	0.20	0.70	4.97
COMP5	24h	4.32	0.64	-32.37	0.21	0.61	4.033
	48h	4.32	0.64	-32.37	0.21	0.61	4.03
COMP6	24h	2.42	0.38	-16.17	0.75	0.33	2.12
	48h	1.19	3.05	-1.91	0.82	2.55	353.87
COMP7	24h	3.87	0.59	-29.95	0.39	0.56	3.60
	48h	3.71	0.57	-15.77	0.60	0.51	3.23
COMP8	24h	9.60	0.98	-3.634	0.84	0.72	5.24
	48h	4.00	0.60	-37.5	0.84	0.58	3.78
COMP9	24h	9.77	0.99	58.14	0.34	1.01	10.14
	48h	1.09	0.04	-17.53	0.77	-0.02	0.97

Πίνακας 3.16. Υπολογισμός EC_{90}

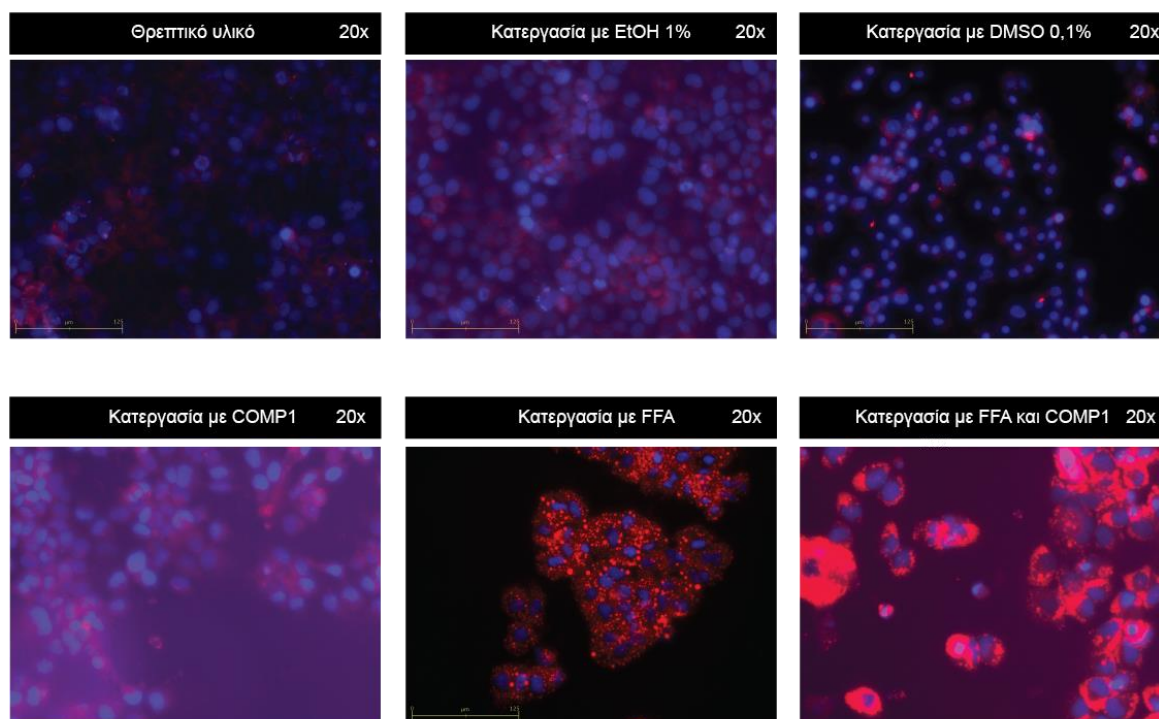
Βιοδραστικές ουσίες	Συγκέντρωση (μM)
COMP1	8
COMP2	6
COMP3	5
COMP4	4
COMP5	0.5
COMP6	3
COMP7	0.05
COMP8	5
COMP9	8

Πίνακας 3.17. Συγκεντρώσεις βιοδραστικών ουσιών

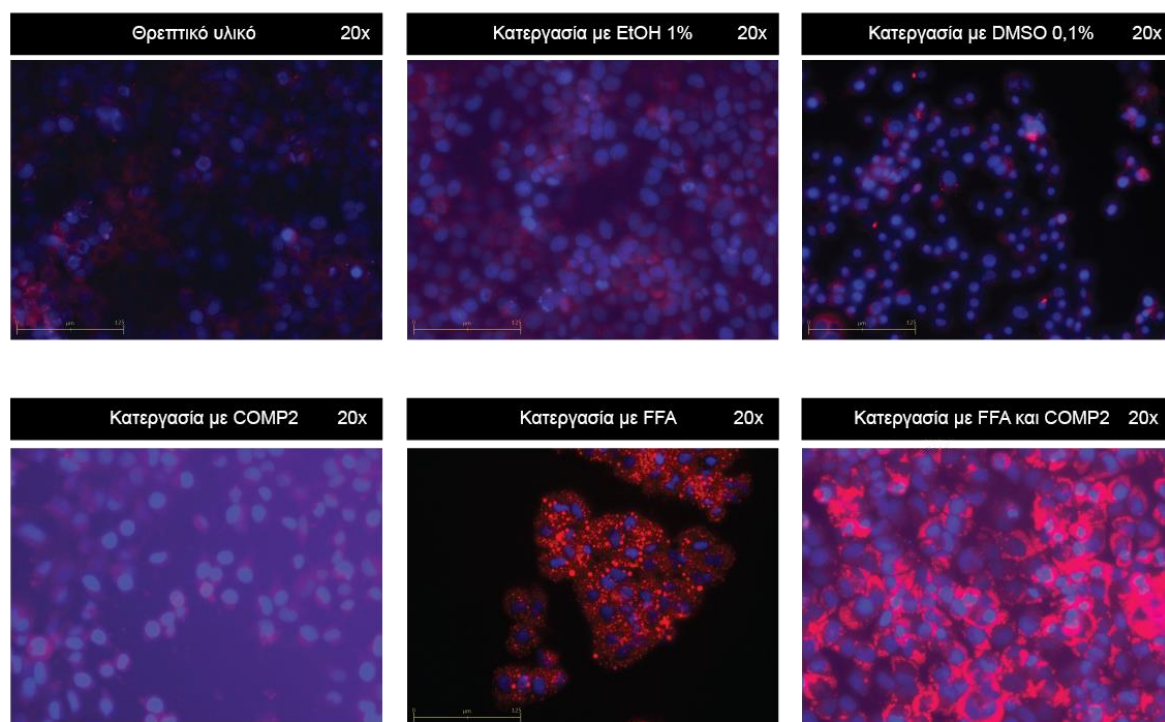
3.2.2. High Content Screening

Σκοπός της παρούσας πειραματικής διαδικασίας είναι ο έλεγχος της μείωσης του ενδοκυτταρικού λιπιδικού φορτίου από της βιοδραστικές ουσίες αντιμετώπισης της νόσου, όπως αυτές προέκυψαν από την *in silico* ανάλυση. Πραγματοποιήθηκε απεικόνιση των κυτάρων μετά από κατεργασία αυτών τόσο με τις στεατογενικές ουσίες που επάγουν NAFLD και τις βιοδραστικές ουσίες αντιμετώπισης, όσο και με τις χημικές ενώσεις στις οποίες διαλύθηκαν οι ουσίες αυτές (control group). Για τη χρώση των σταγονιδίων λιπιδίων

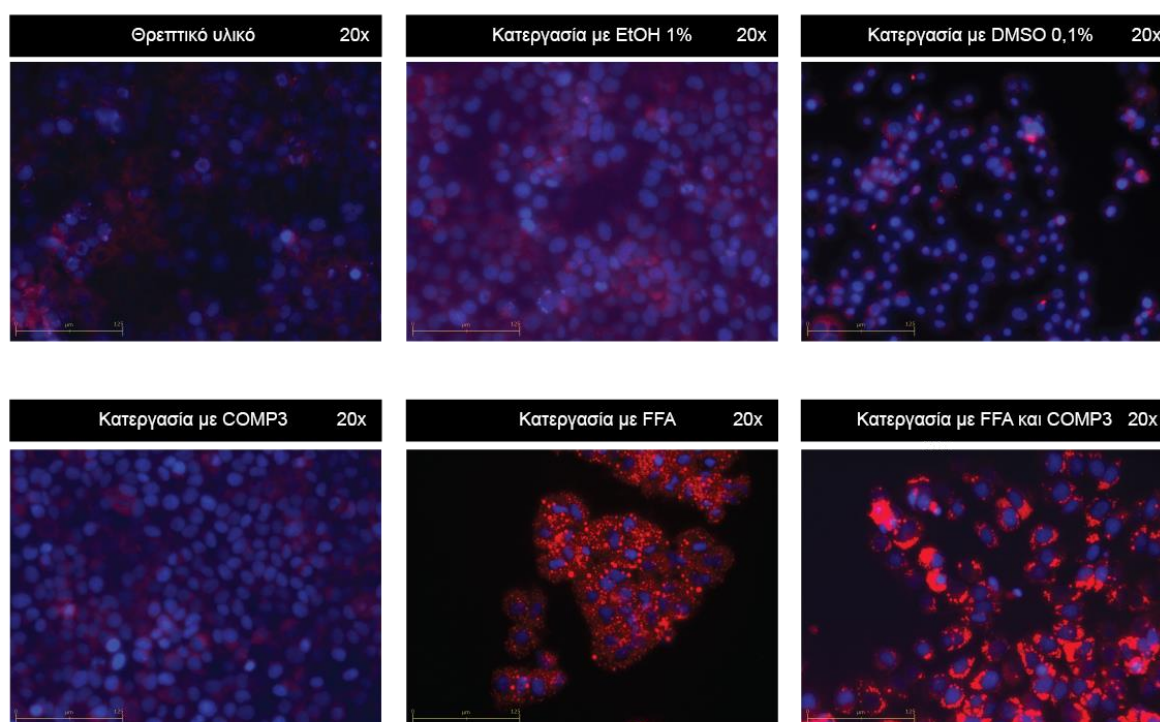
χρησιμοποιήθηκε η χρωστική Nile Red (κόκκινο), ενώ για τη χρώση των πυρήνων η Hoechst 33420 (μπλε). Η απεικόνιση των κυττάρων έγινε με μέθοδο high content screening στο μικροσκόπιο φθορισμού JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek).



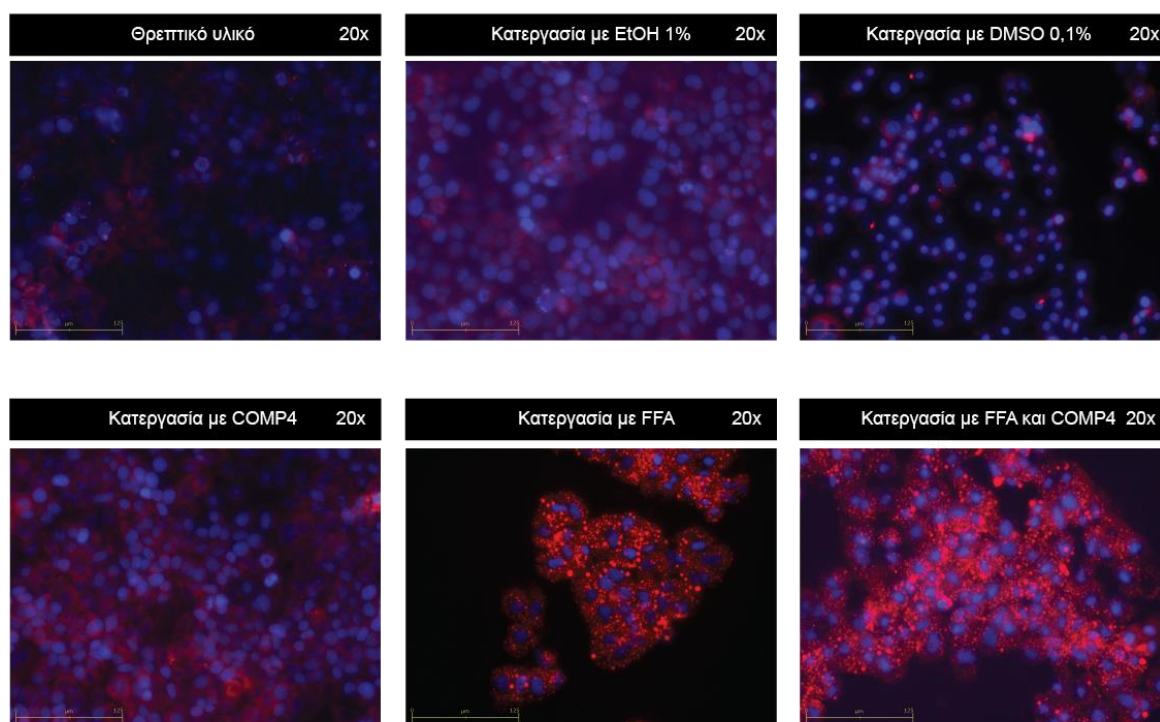
Εικόνα 3.15. Κατεργασία με FFA και COMP1



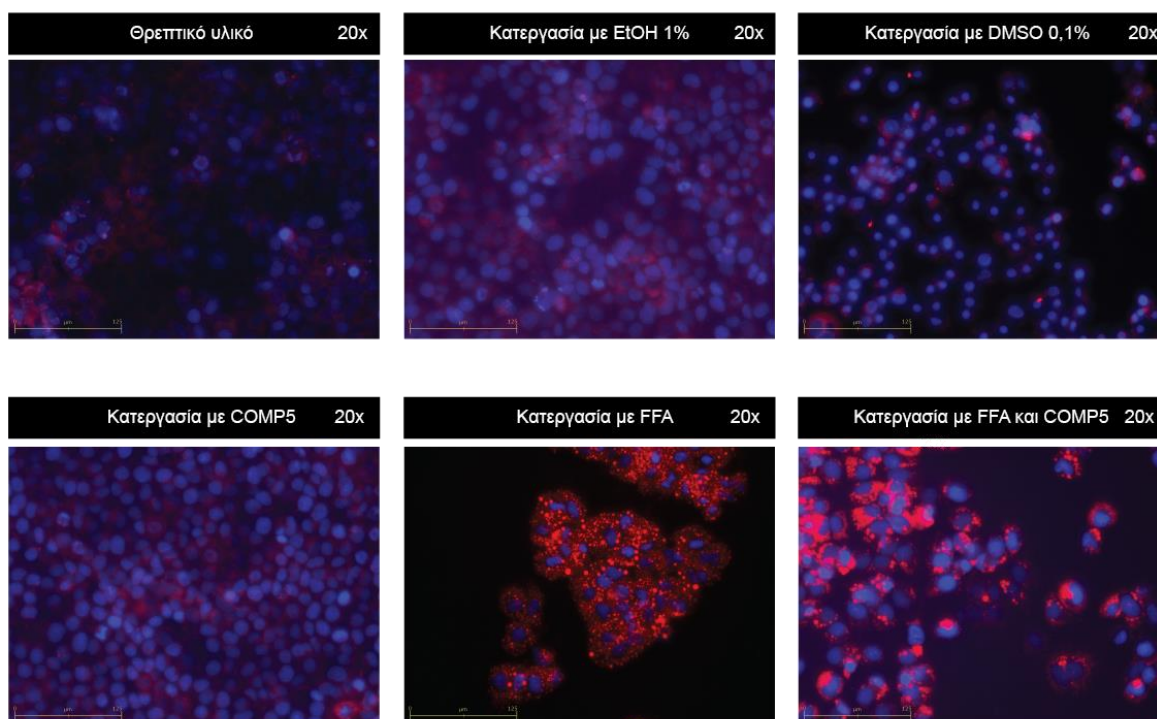
Εικόνα 3.16. Κατεργασία με FFA και COMP2



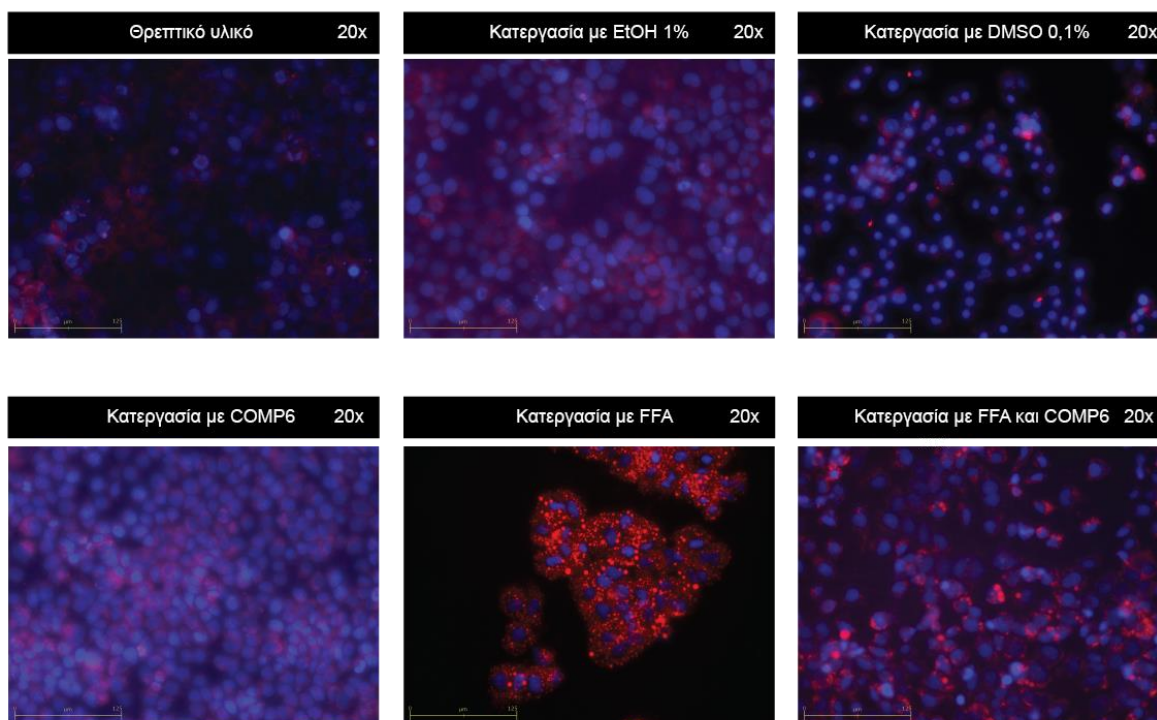
Εικόνα 3.17. Κατεργασία με FFA και COMP3



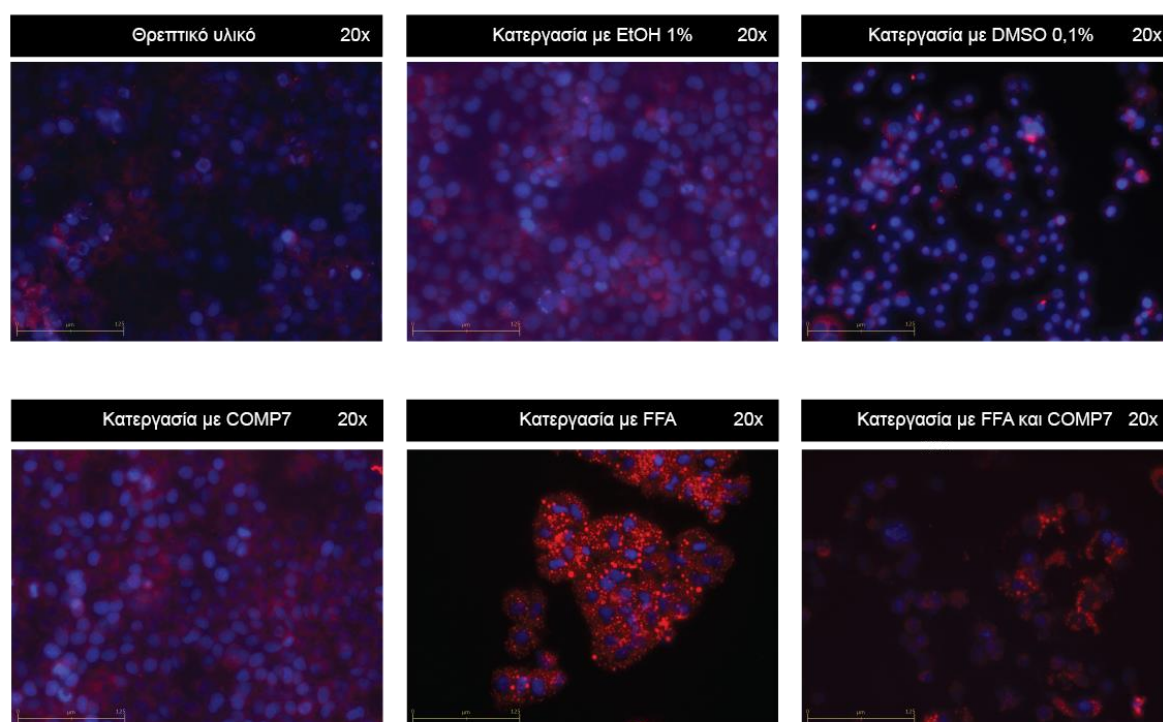
Εικόνα 3.18. Κατεργασία με FFA και COMP4



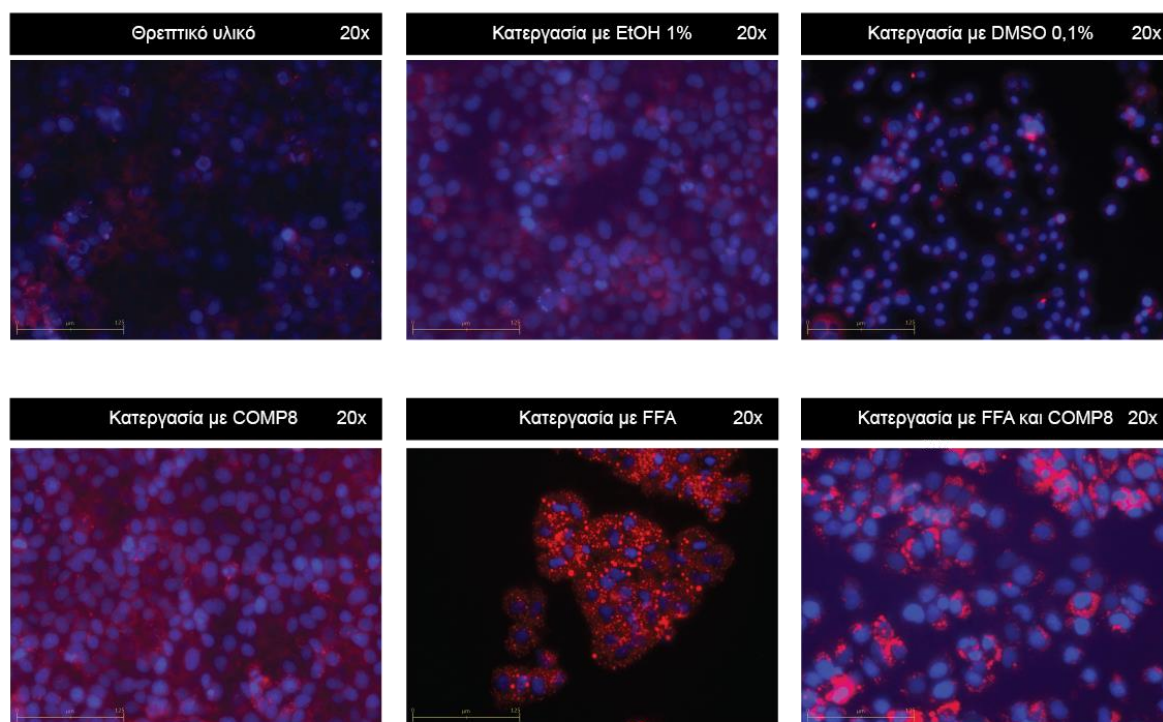
Εικόνα 3.19. Κατεργασία με FFA και COMP5



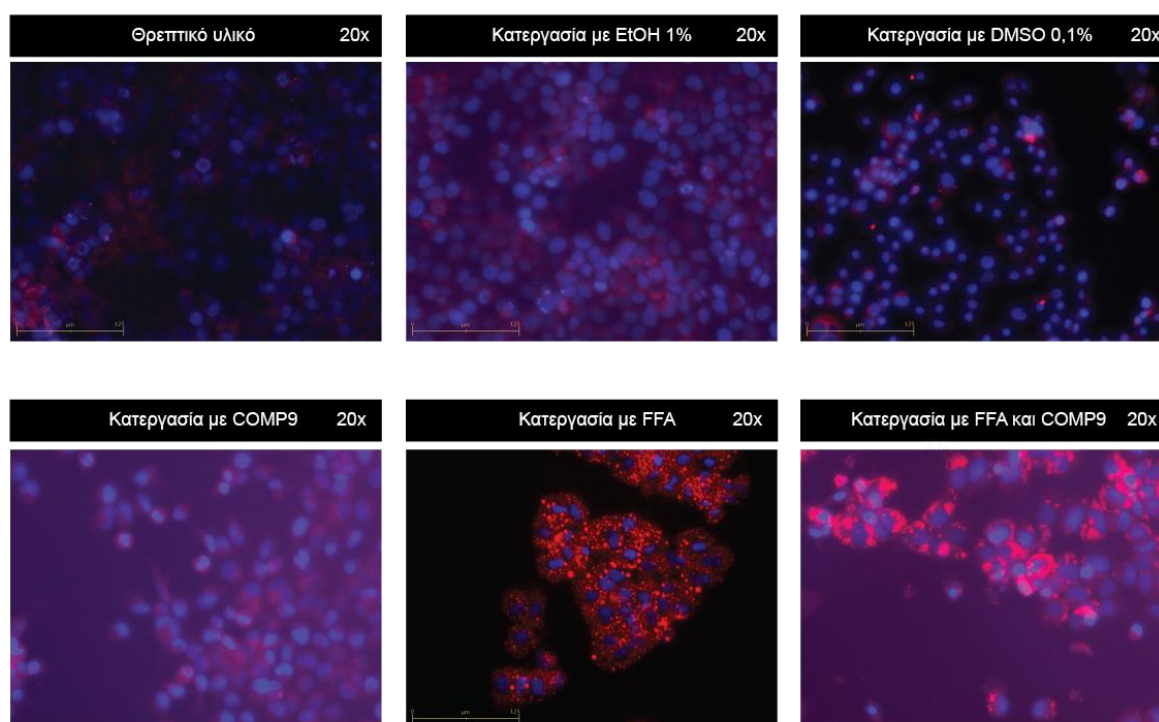
Εικόνα 3.20. Κατεργασία με FFA και COMP6



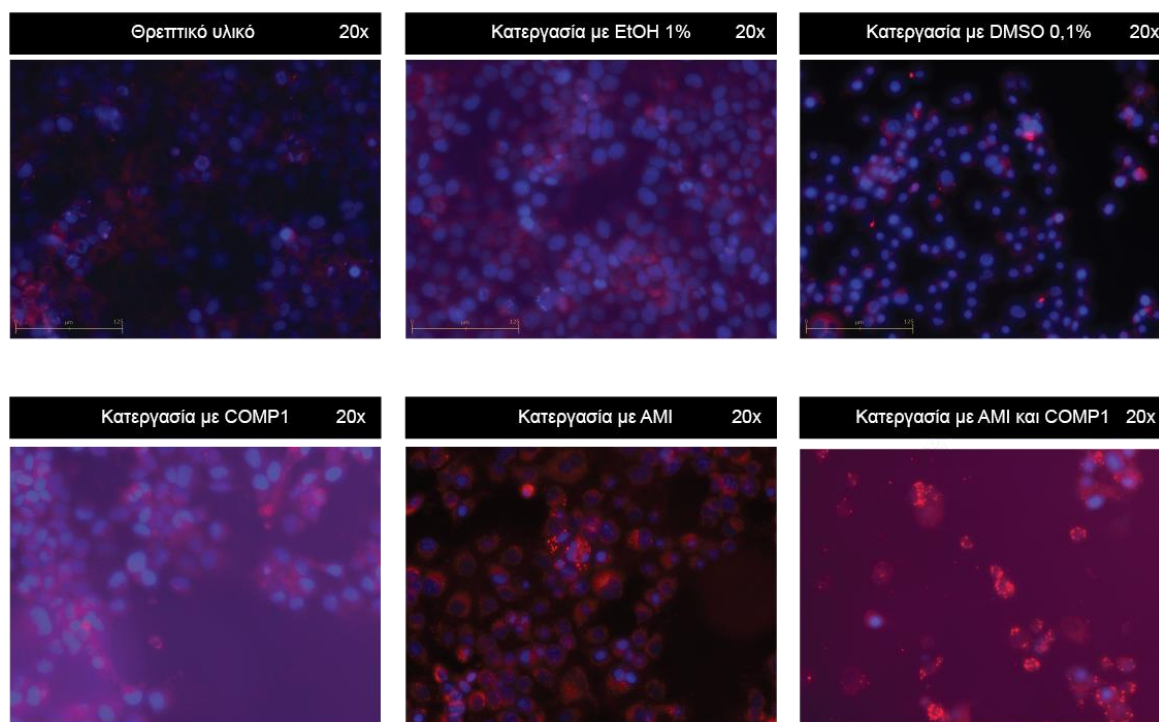
Εικόνα 3.21. Κατεργασία με FFA και COMP7



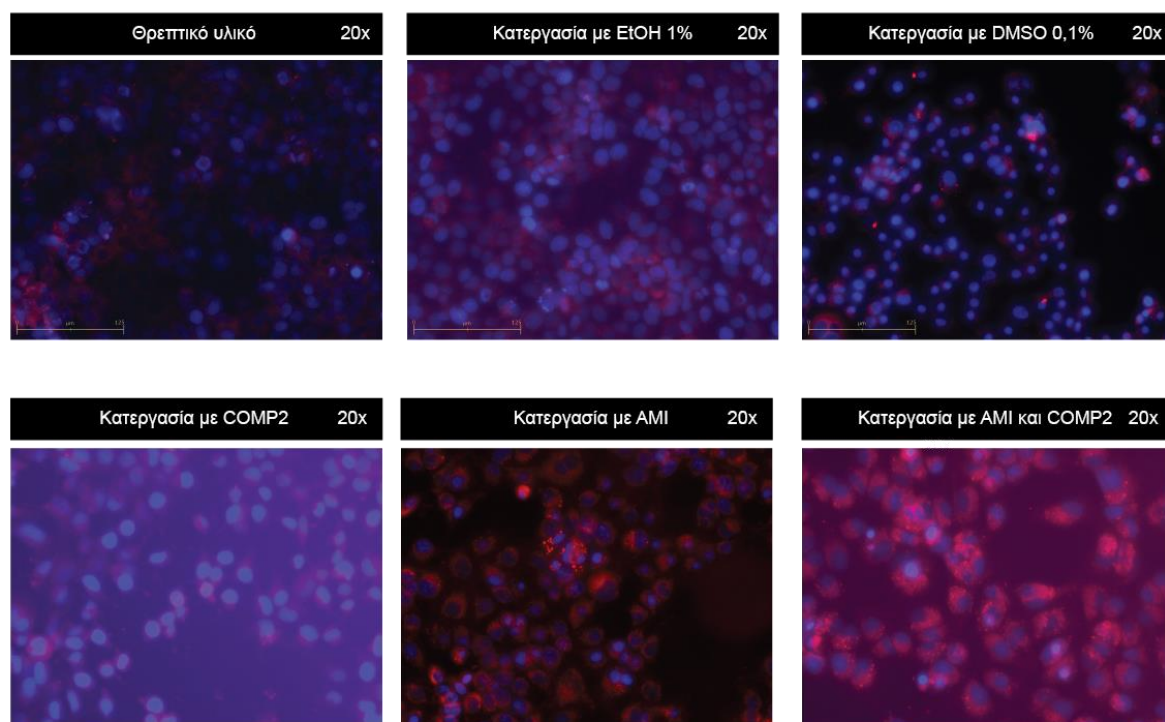
Εικόνα 3.22. Κατεργασία με FFA και COMP8



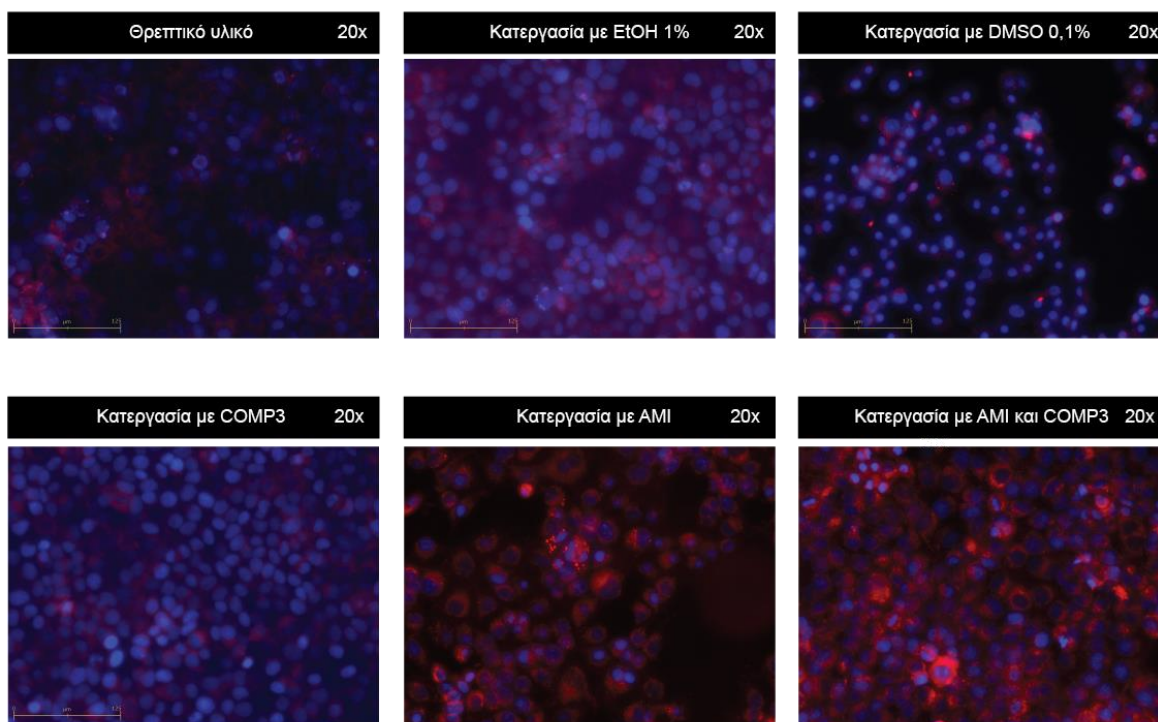
Εικόνα 3.23. Κατεργασία με FFA και COMP9



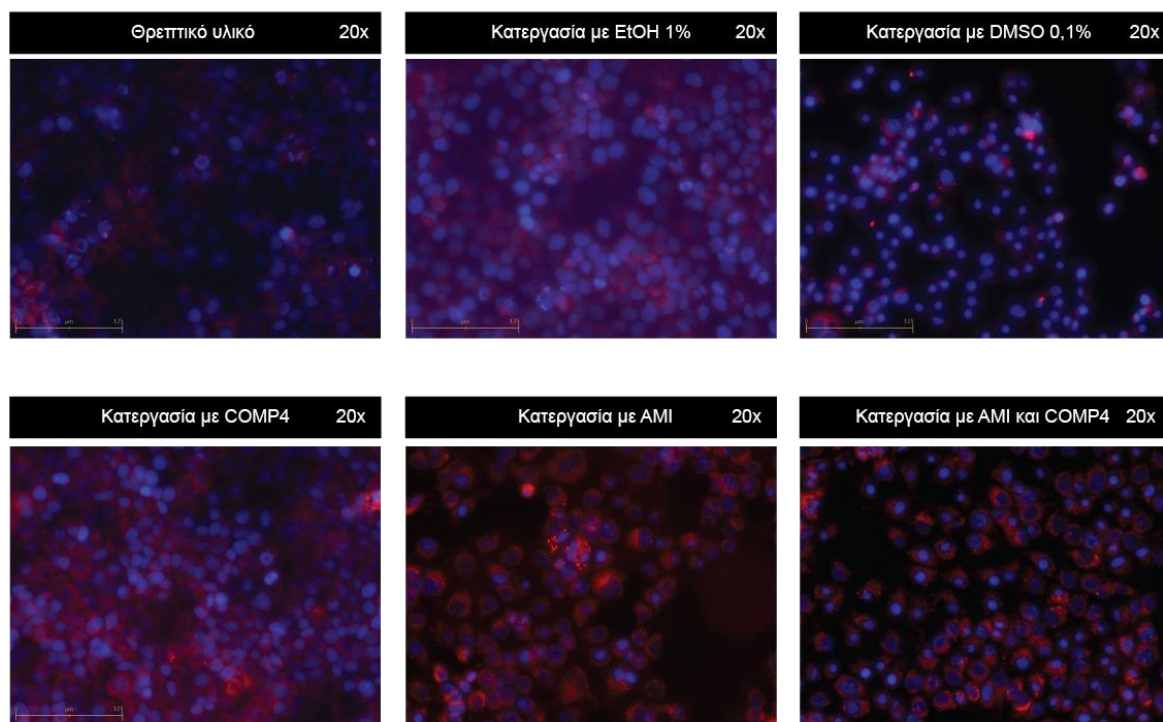
Εικόνα 3.24. Κατεργασία με AMI και COMP1



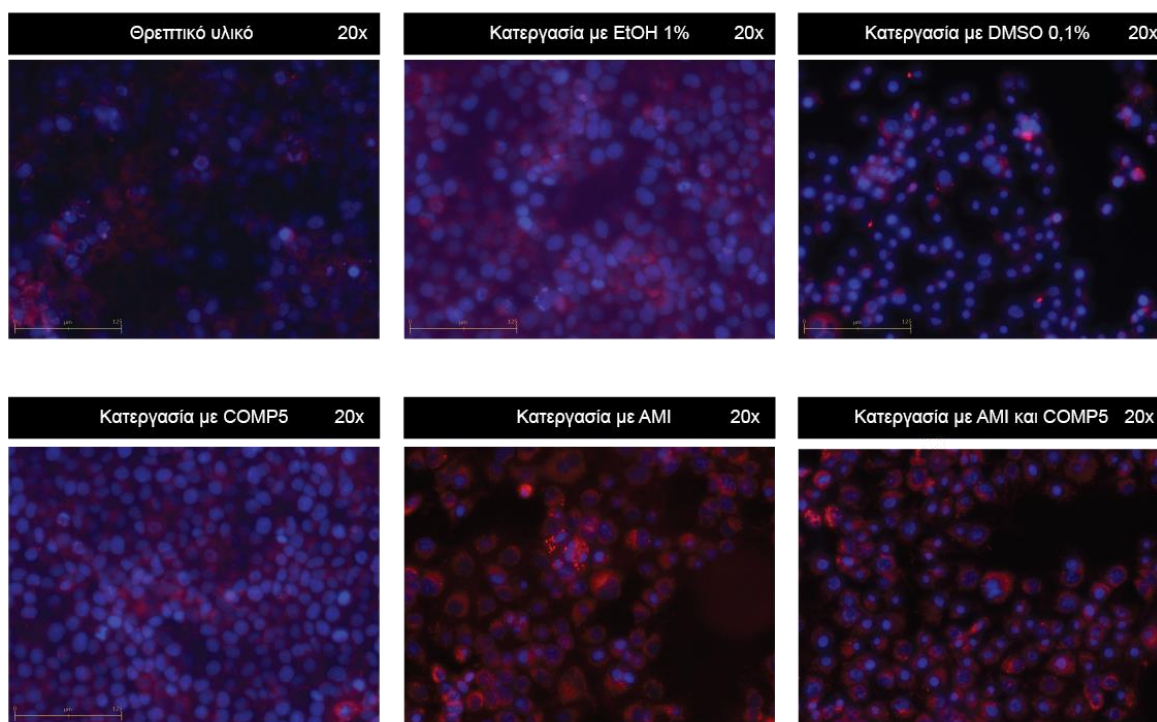
Εικόνα 3.25. Κατεργασία με AMI και COMP2



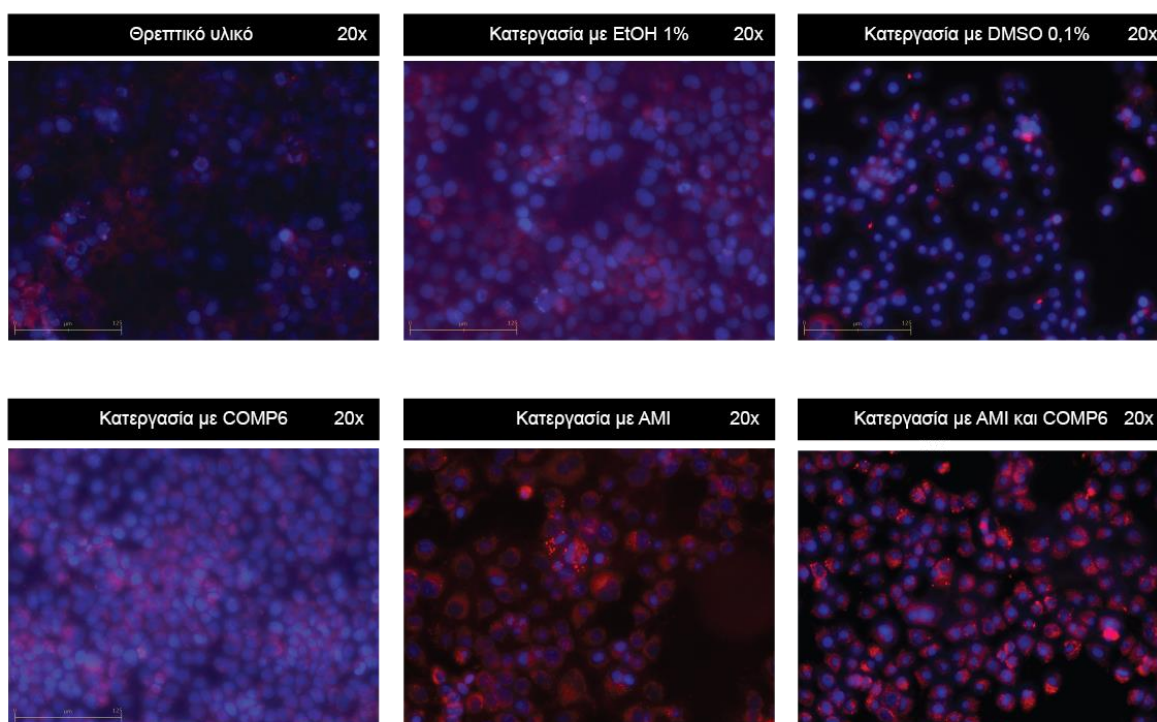
Εικόνα 3.26. Κατεργασία με AMI και COMP3



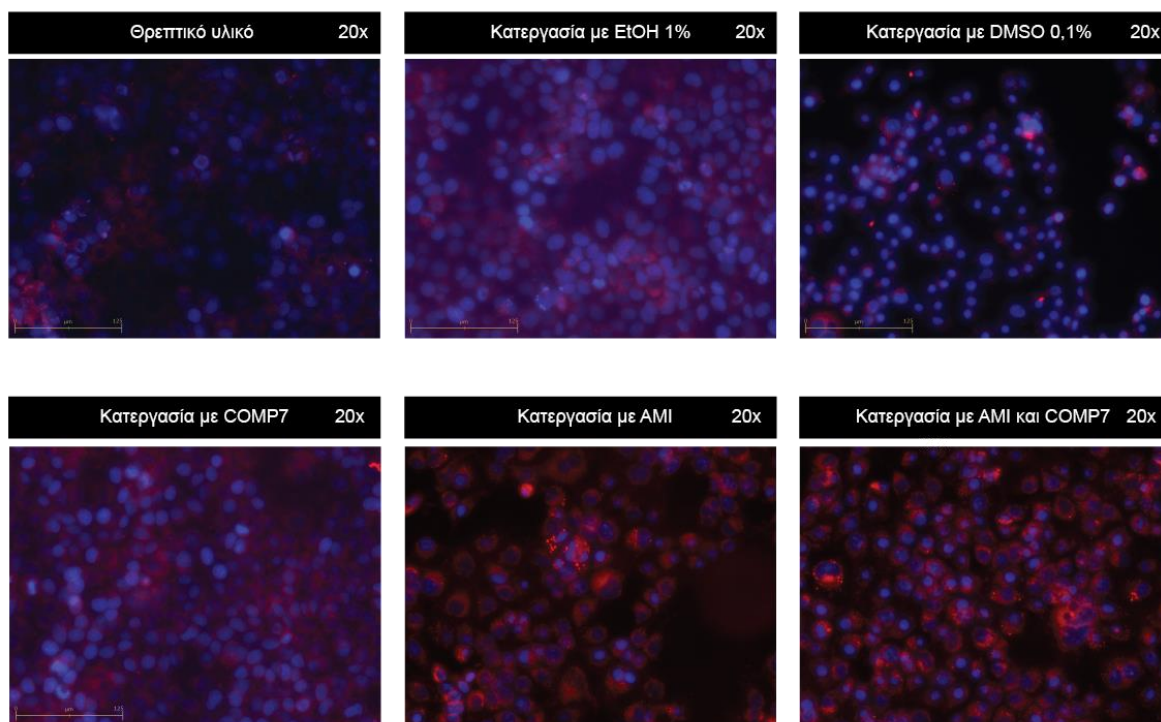
Εικόνα 3.27. Κατεργασία με AMI και COMP4



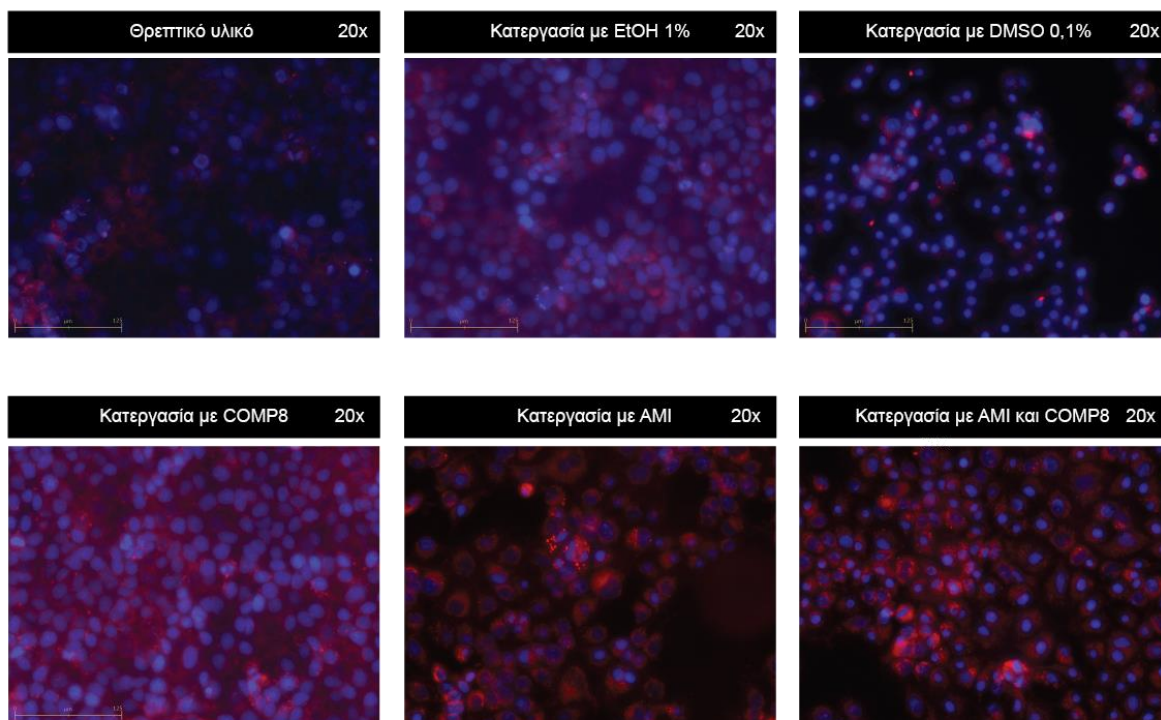
Εικόνα 3.27. Κατεργασία με AMI και COMP5



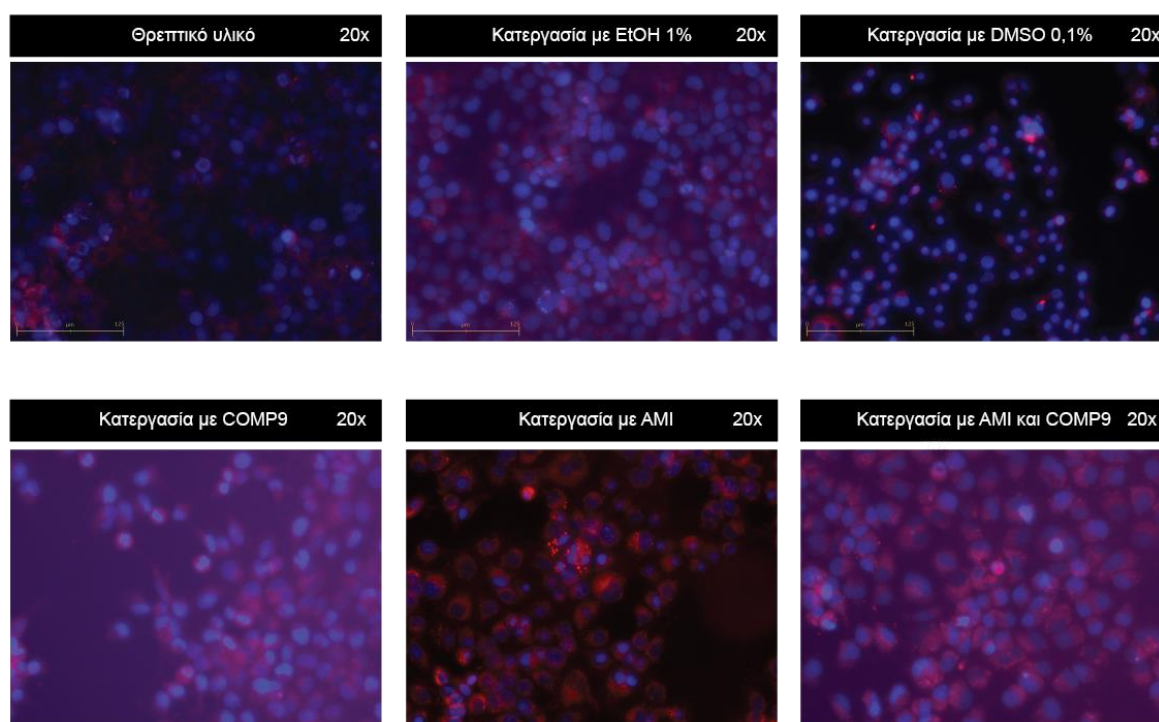
Εικόνα 3.29. Κατεργασία με AMI και COMP6



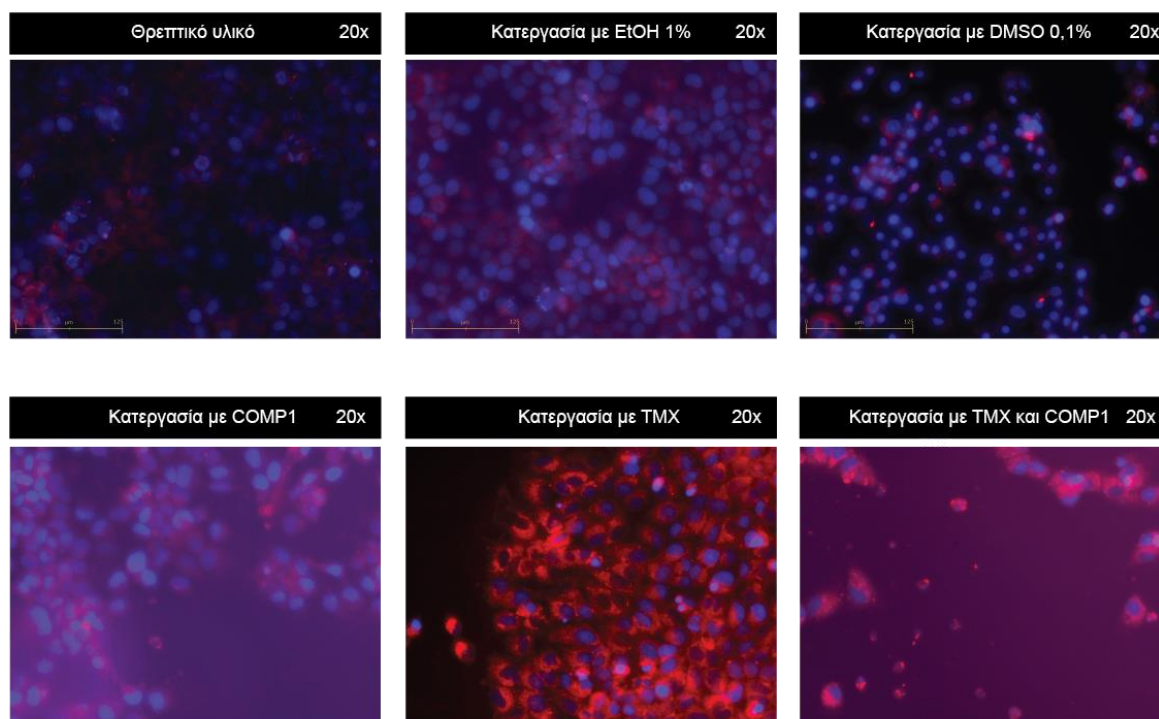
Εικόνα 3.30. Κατεργασία με AMI και COMP7



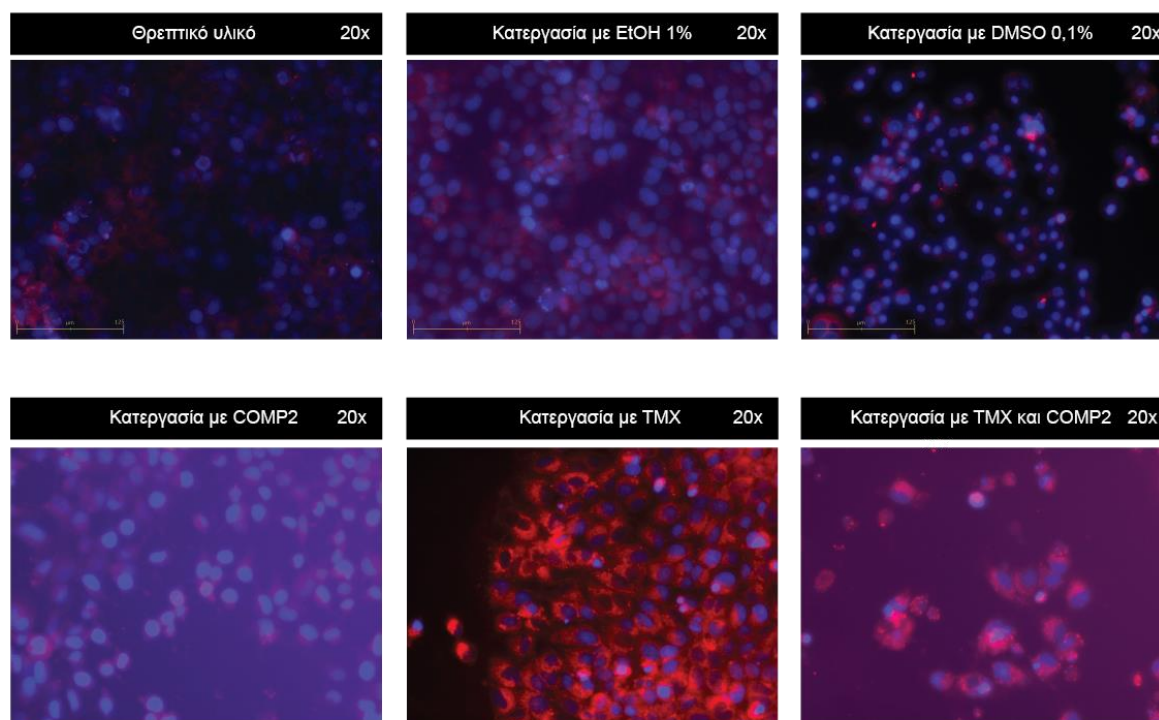
Εικόνα 3.31. Κατεργασία με AMI και COMP8



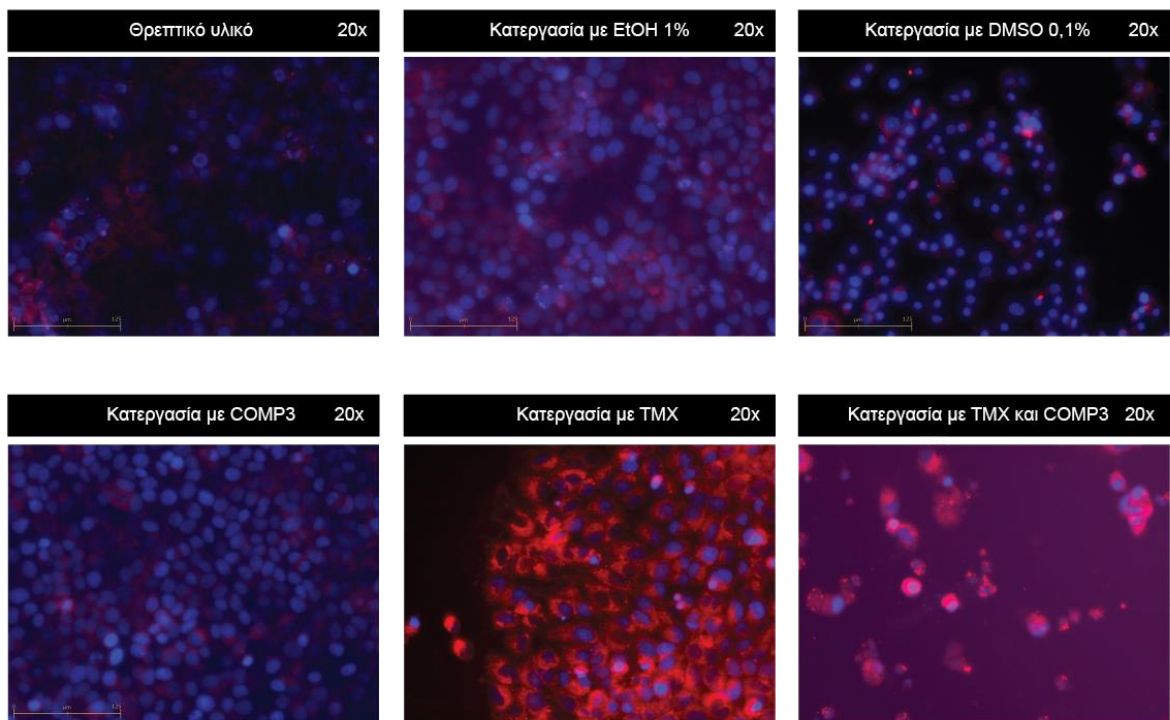
Εικόνα 3.32. Κατεργασία με AMI και COMP9



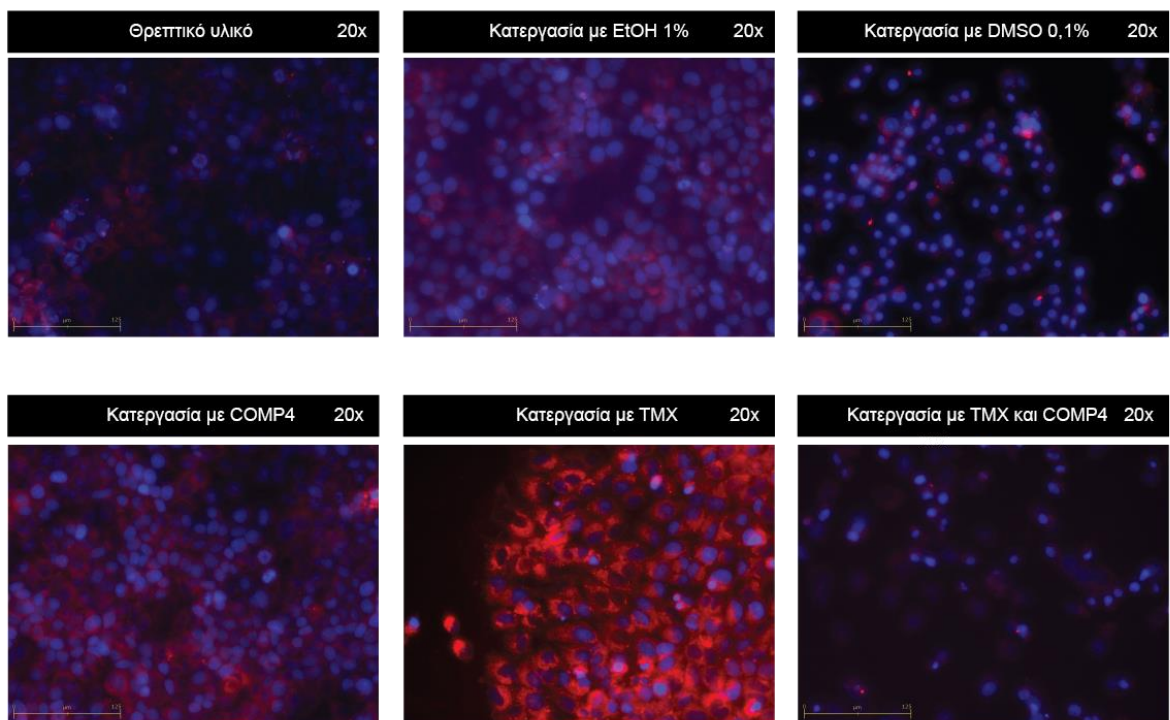
Εικόνα 3.33. Κατεργασία με TMX και COMP1



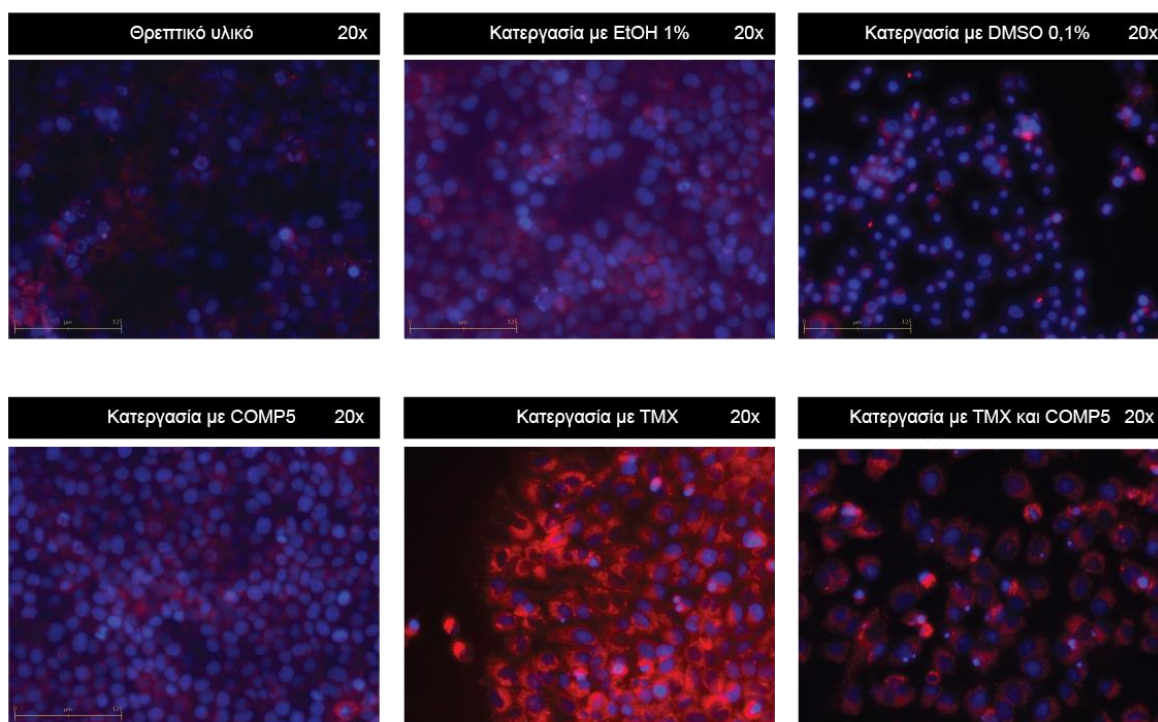
Εικόνα 3.34. Κατεργασία με TMX και COMP2



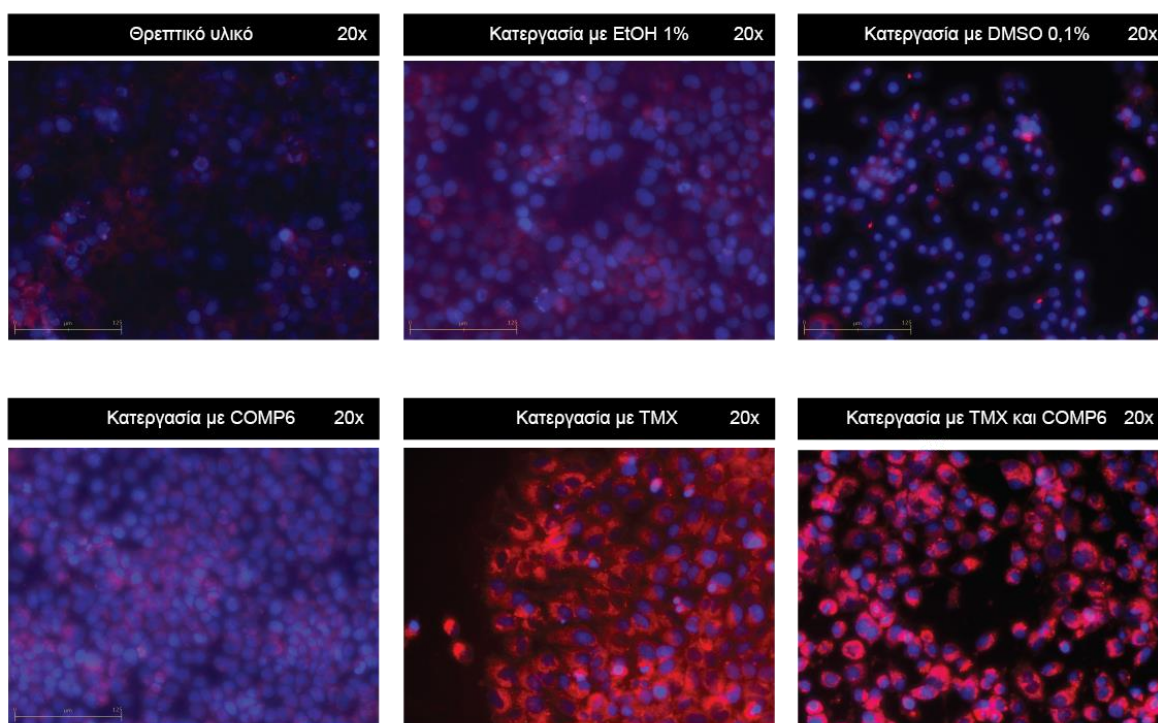
Εικόνα 3.35. Κατεργασία με TMX και COMP4



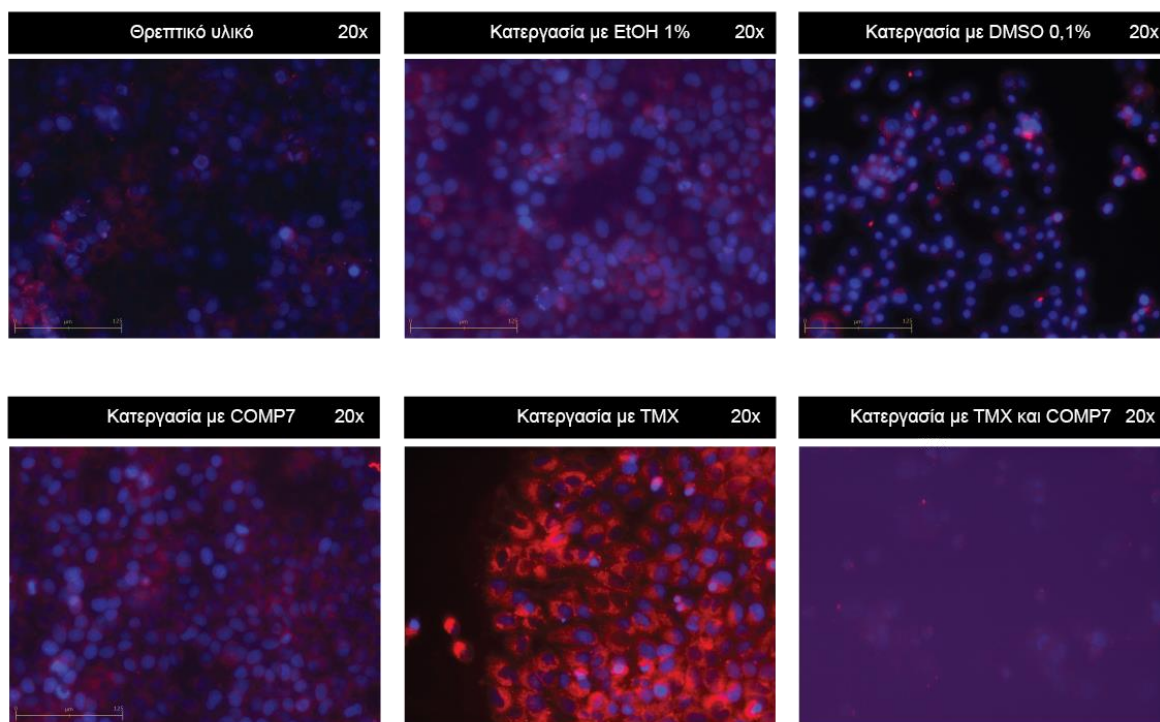
Εικόνα 3.35. Κατεργασία με TMX και COMP4



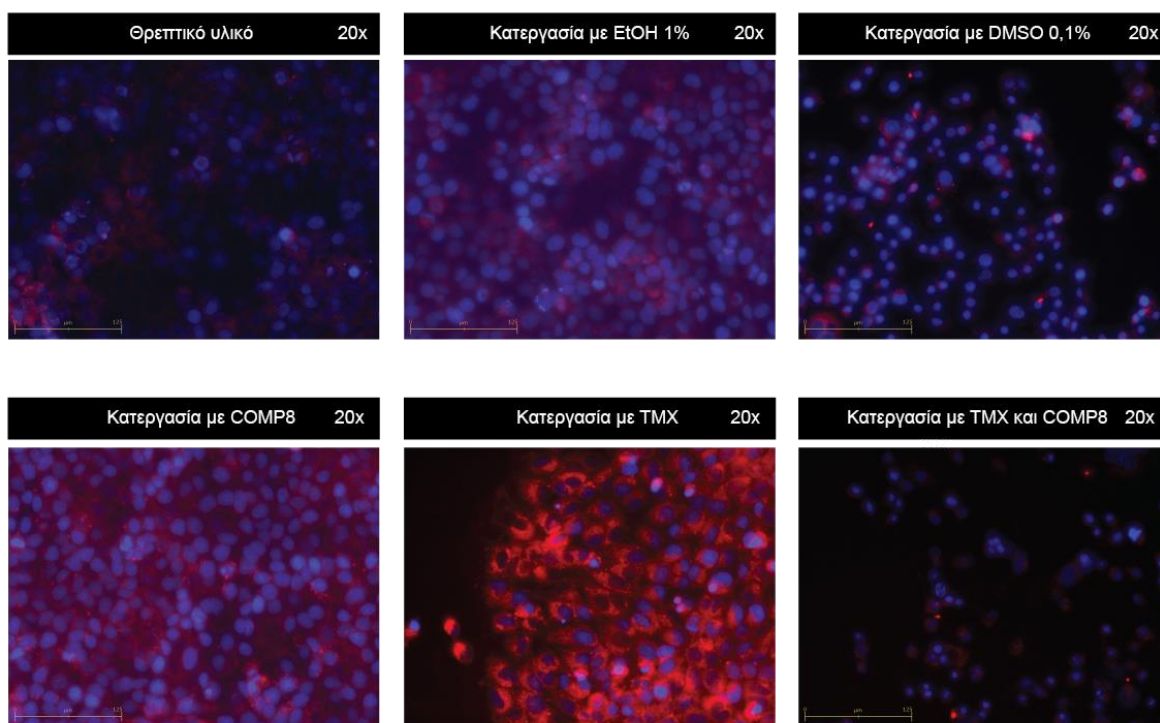
Εικόνα 3.36. Κατεργασία με TMX και COMP5



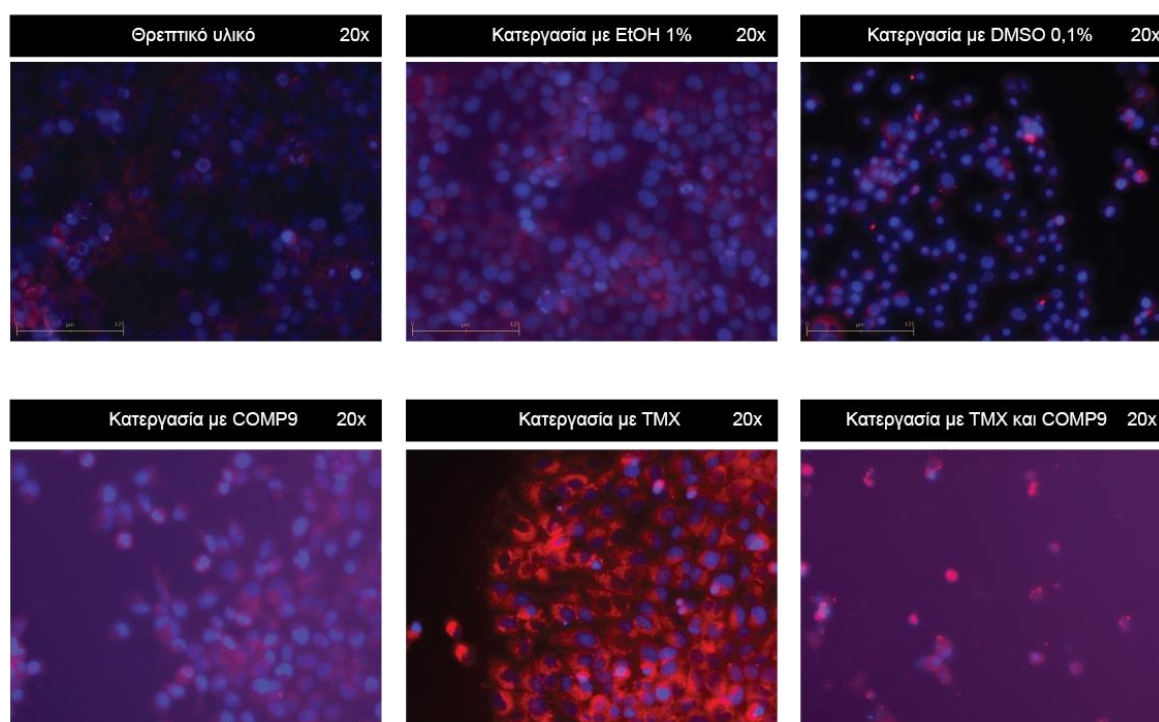
Εικόνα 3.37. Κατεργασία με TMX και COMP6



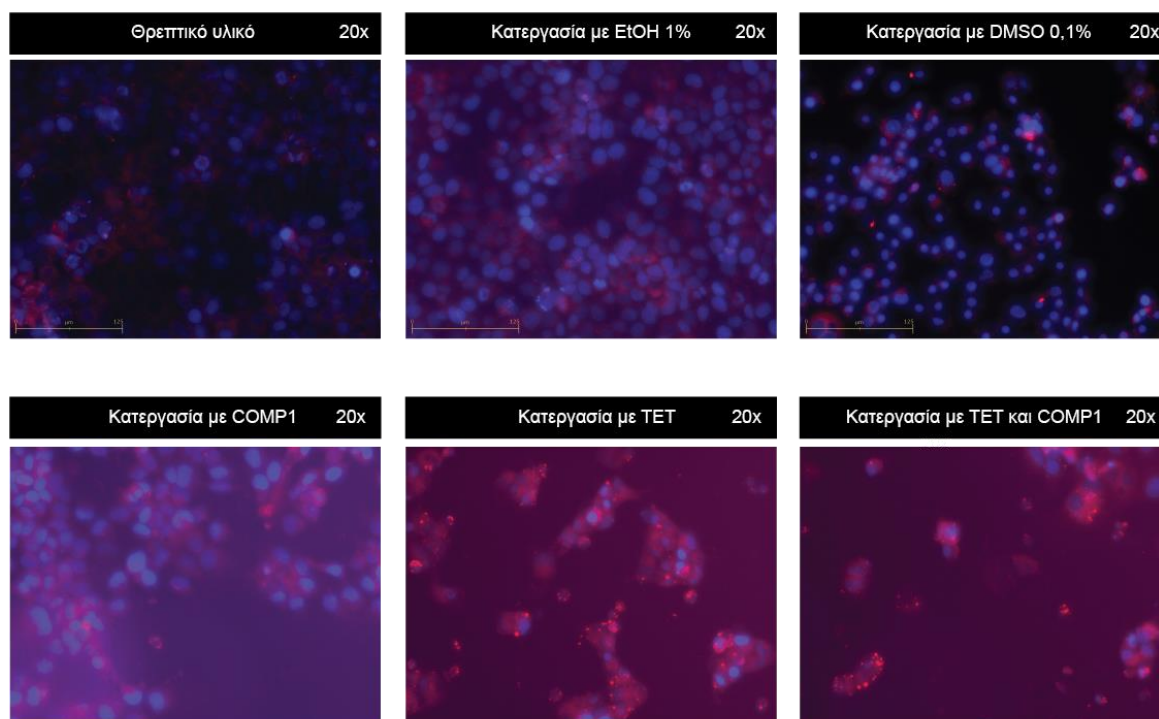
Εικόνα 3.38. Κατεργασία με TMX και COMP7



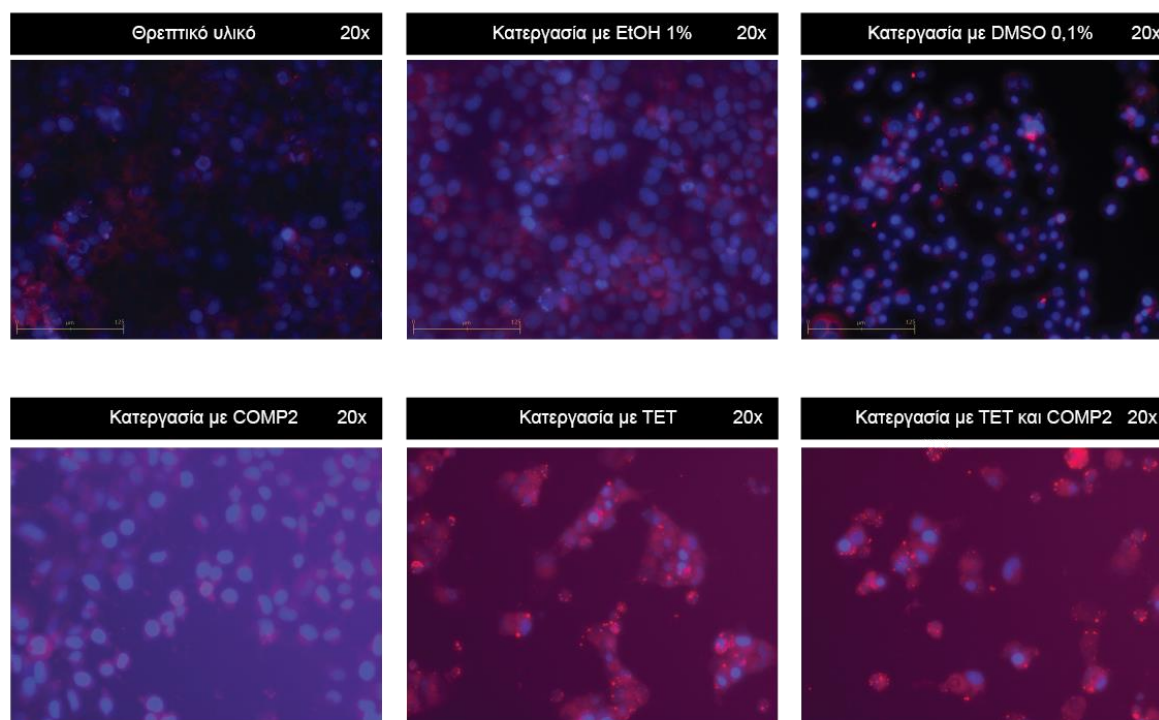
Εικόνα 3.39. Κατεργασία με TMX και COMP8



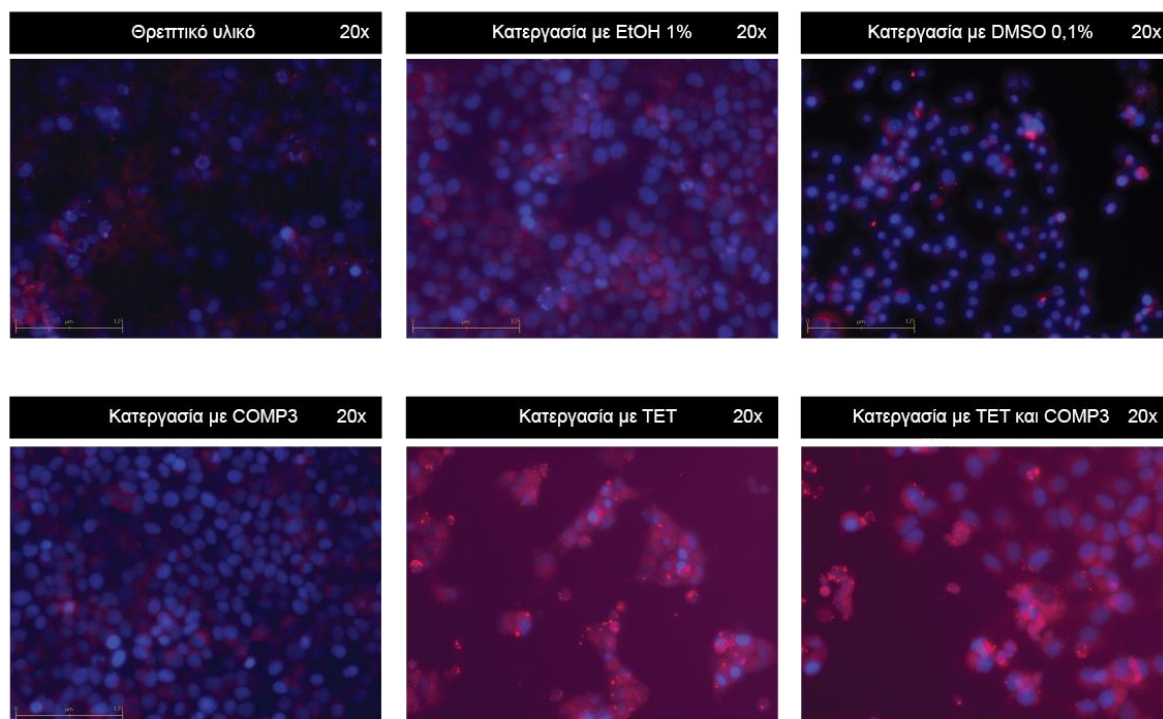
Εικόνα 3.40. Κατεργασία με TMX και COMP9



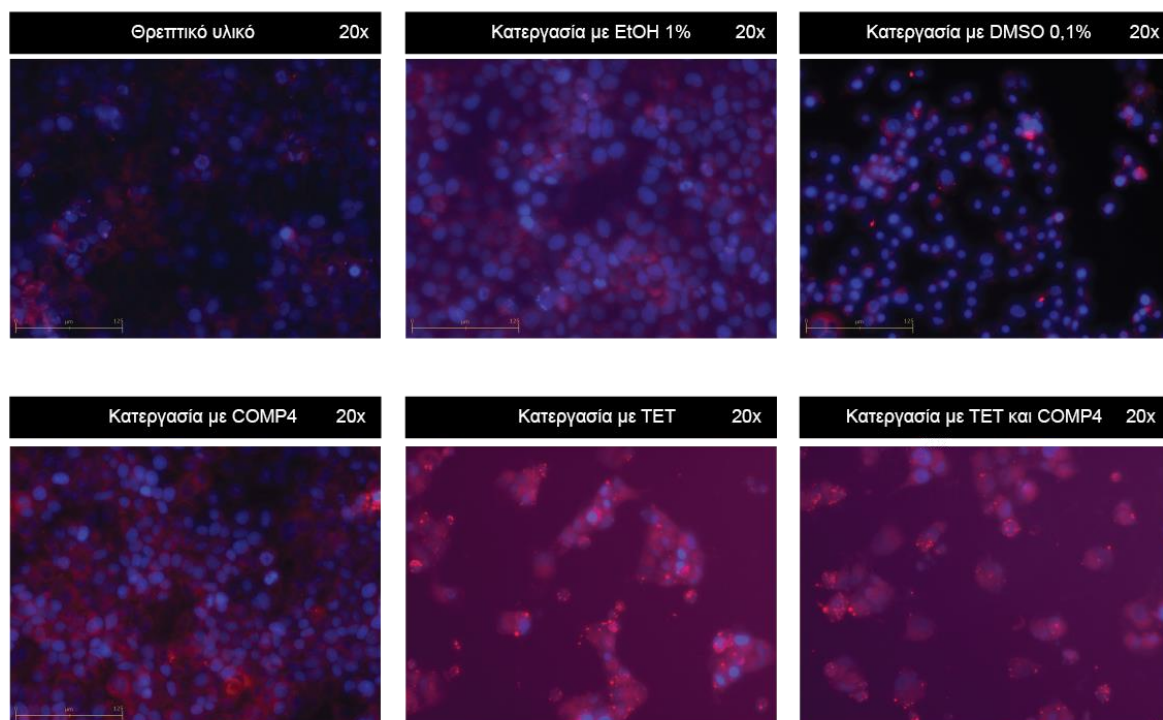
Εικόνα 3.41. Κατεργασία με TET και COMP1



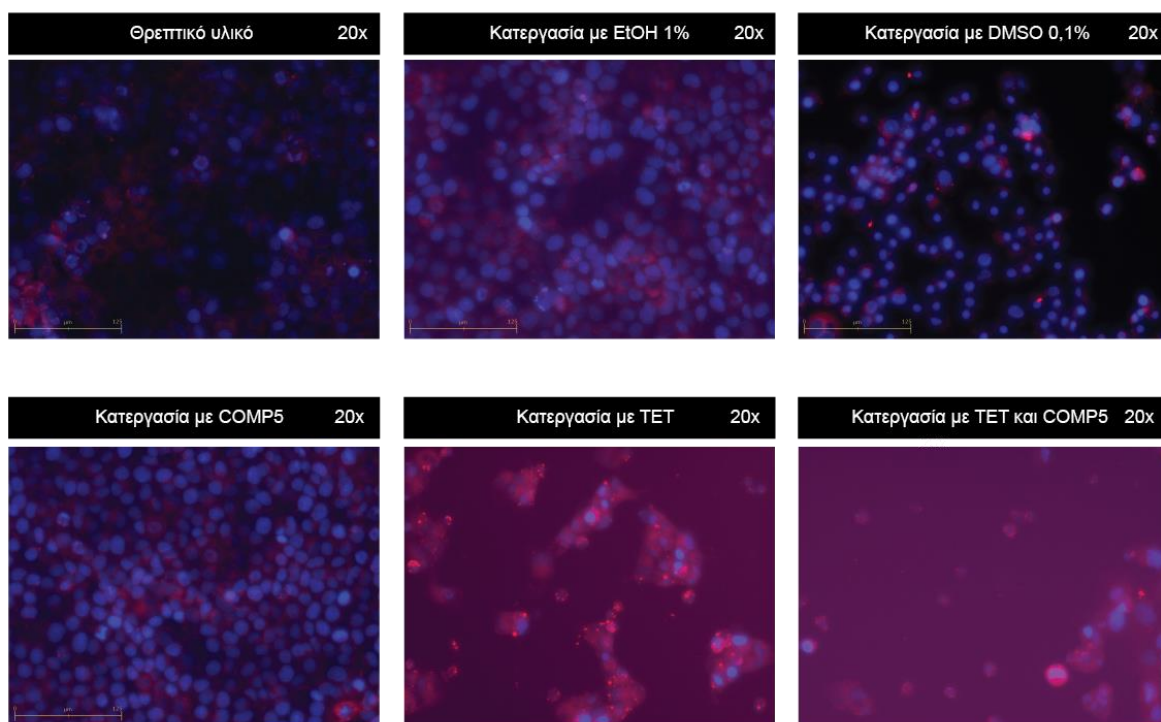
Εικόνα 3.42. Κατεργασία με TET και COMP2



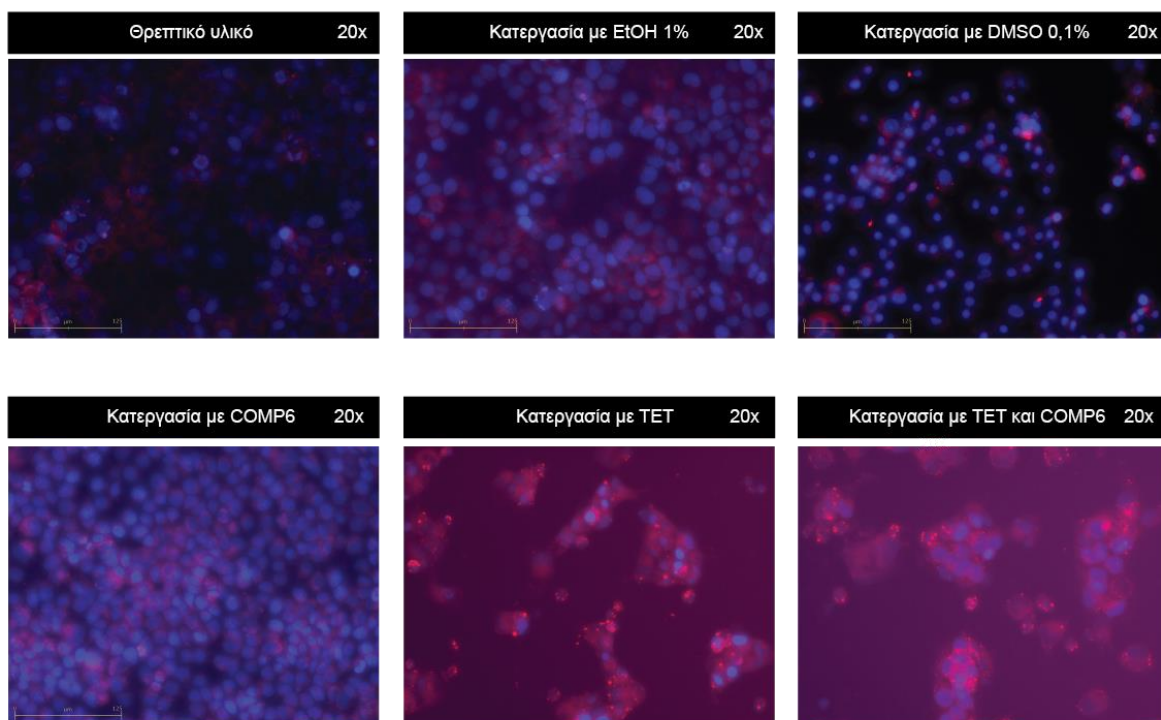
Εικόνα 3.43. Κατεργασία με TET και COMP3



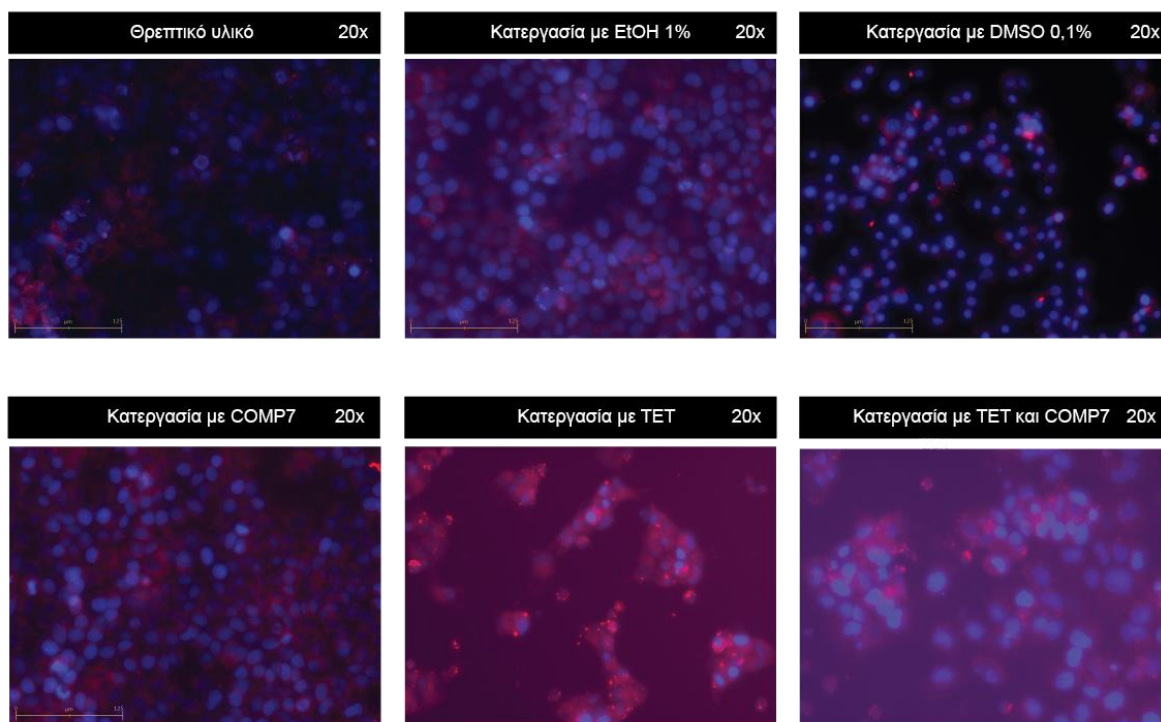
Εικόνα 3.44. Κατεργασία με TET και COMP4



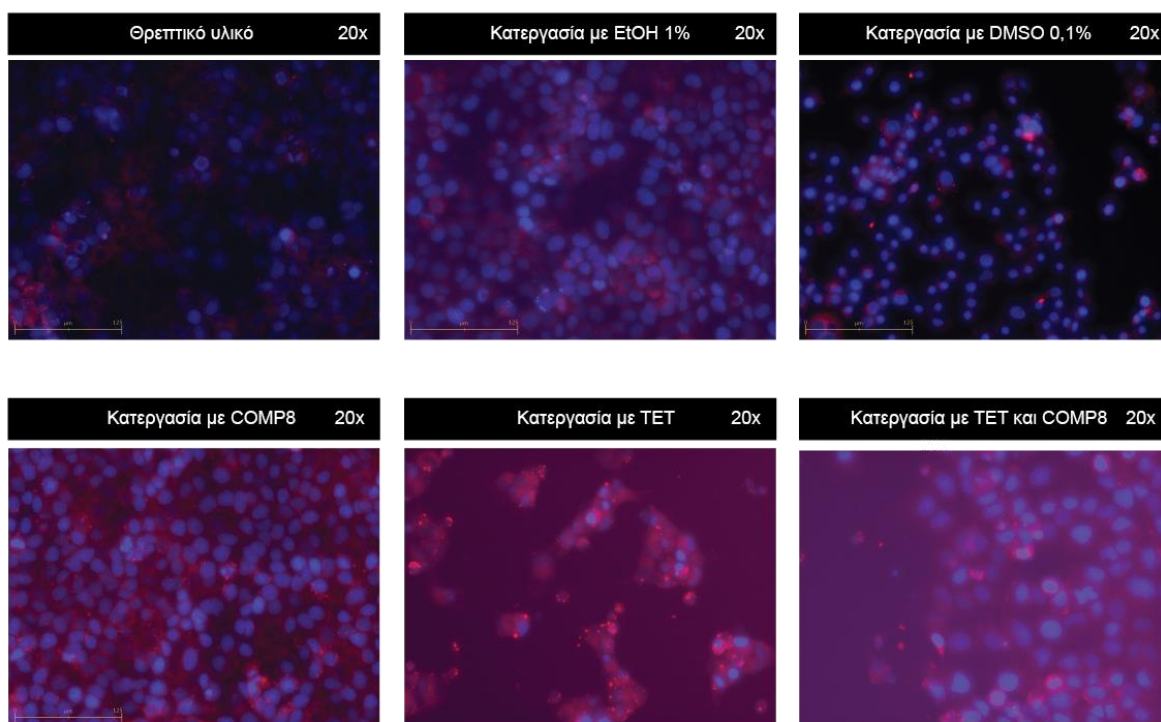
Εικόνα 3.45. Κατεργασία με TET και COMP5



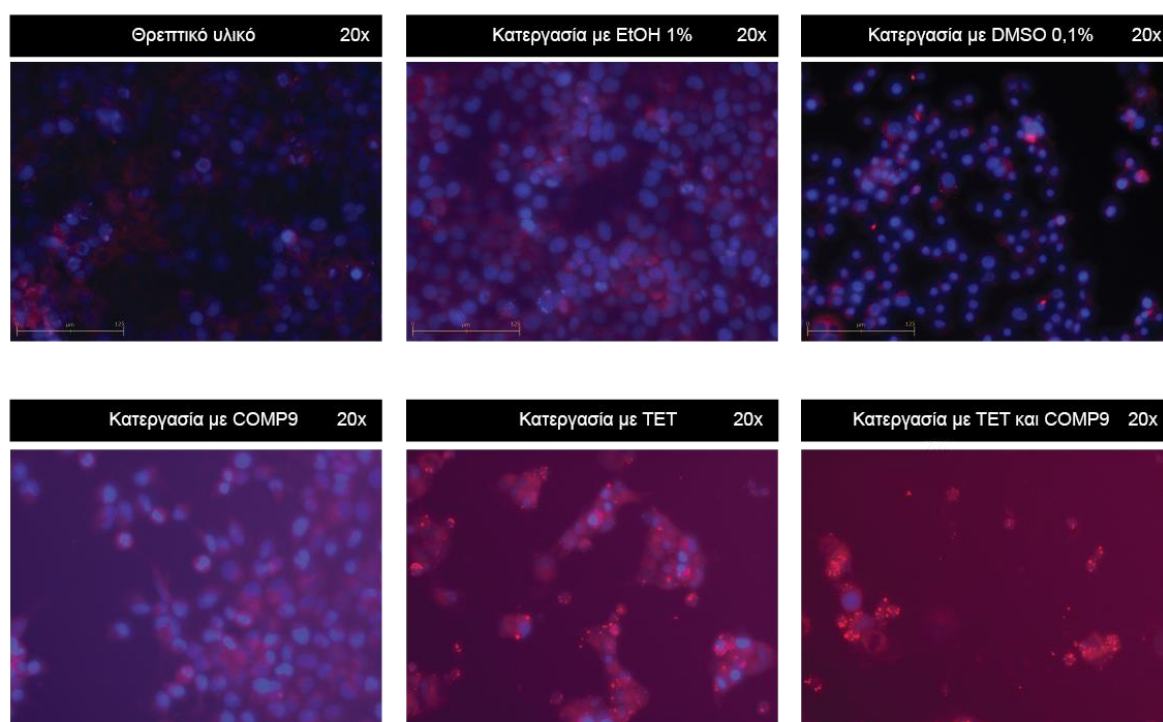
Εικόνα 3.46. Κατεργασία με TET και COMP6



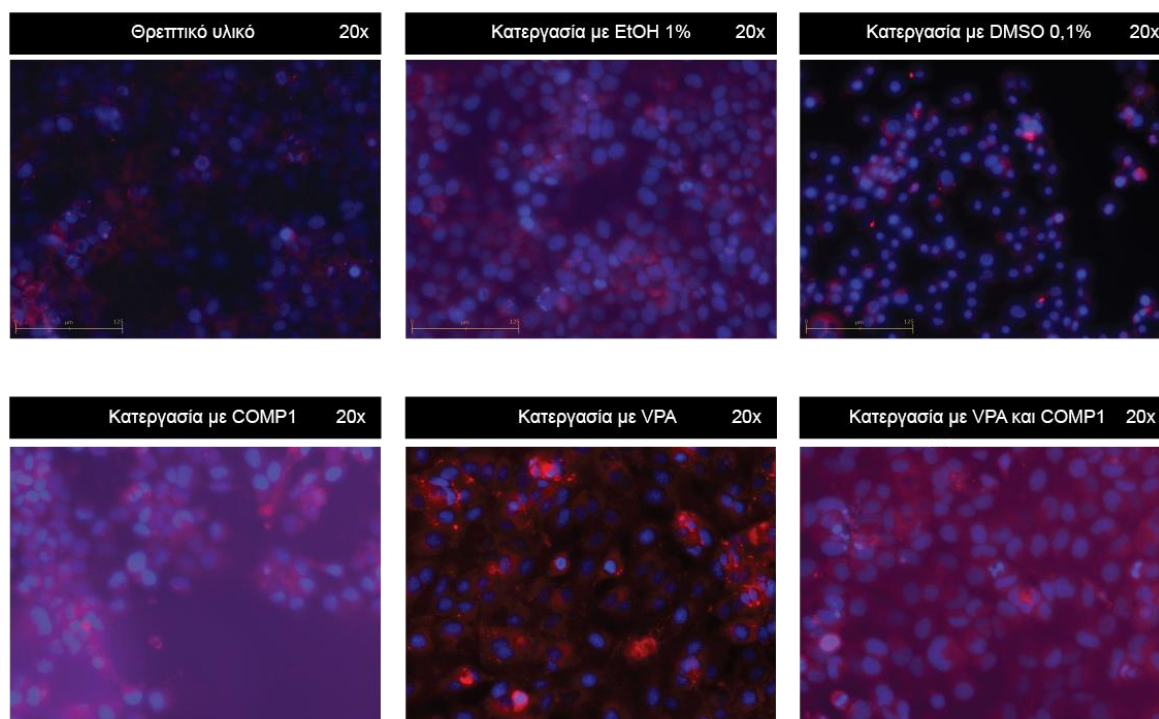
Εικόνα 3.47. Κατεργασία με TET και COMP7



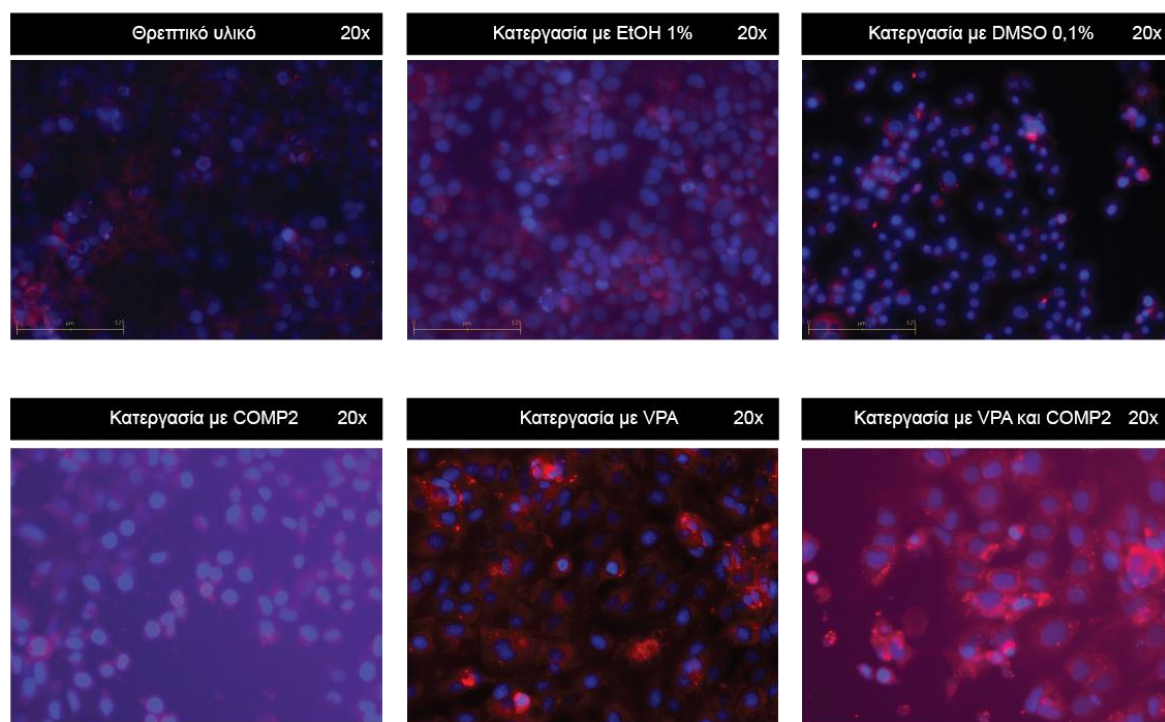
Εικόνα 3.48. Κατεργασία με TET και COMP8



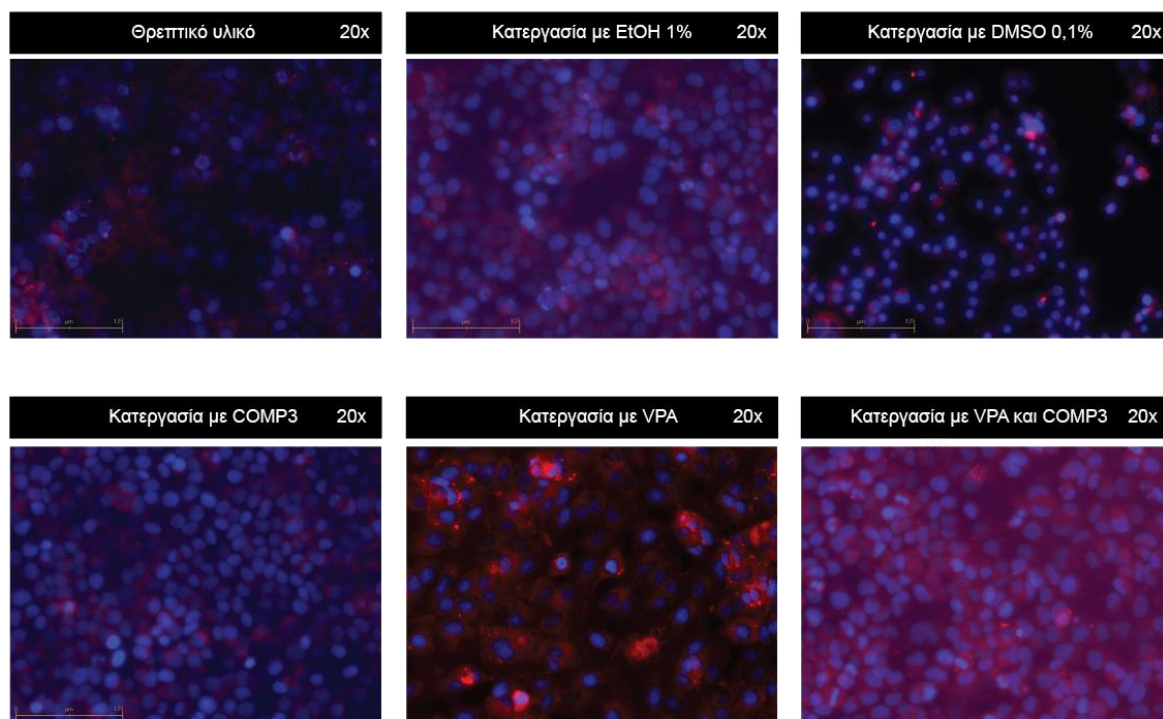
Εικόνα 3.49. Κατεργασία με TET και COMP9



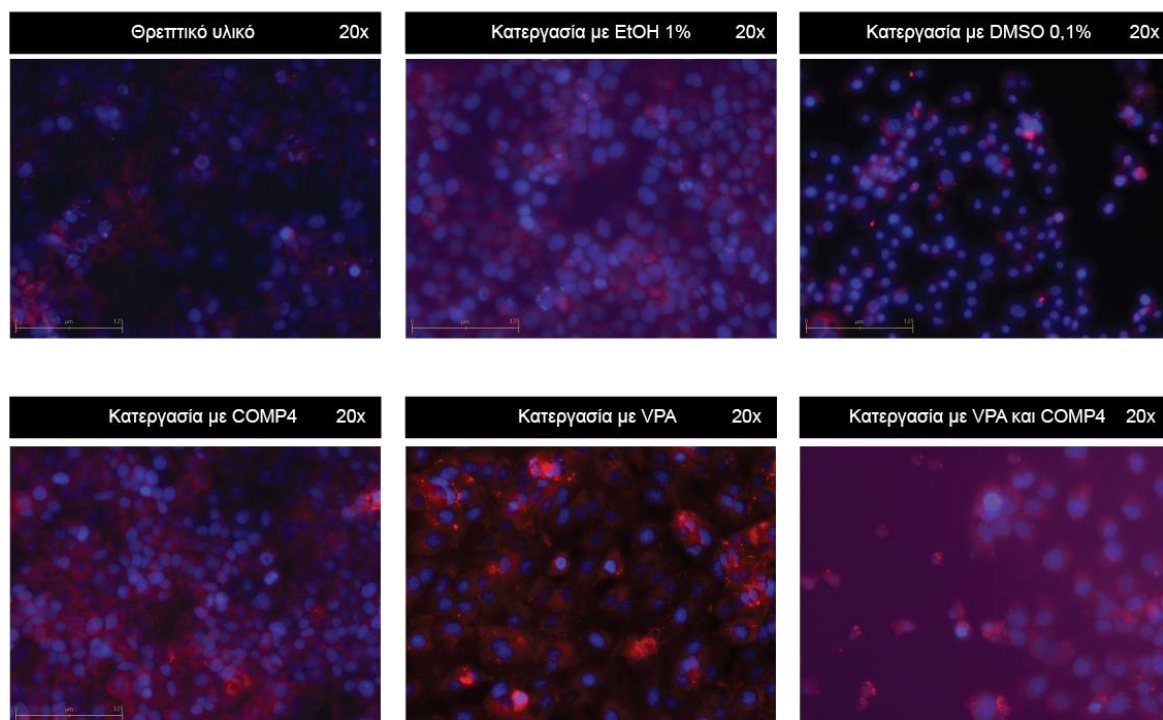
Εικόνα 3.50. Κατεργασία με VPA και COMP1



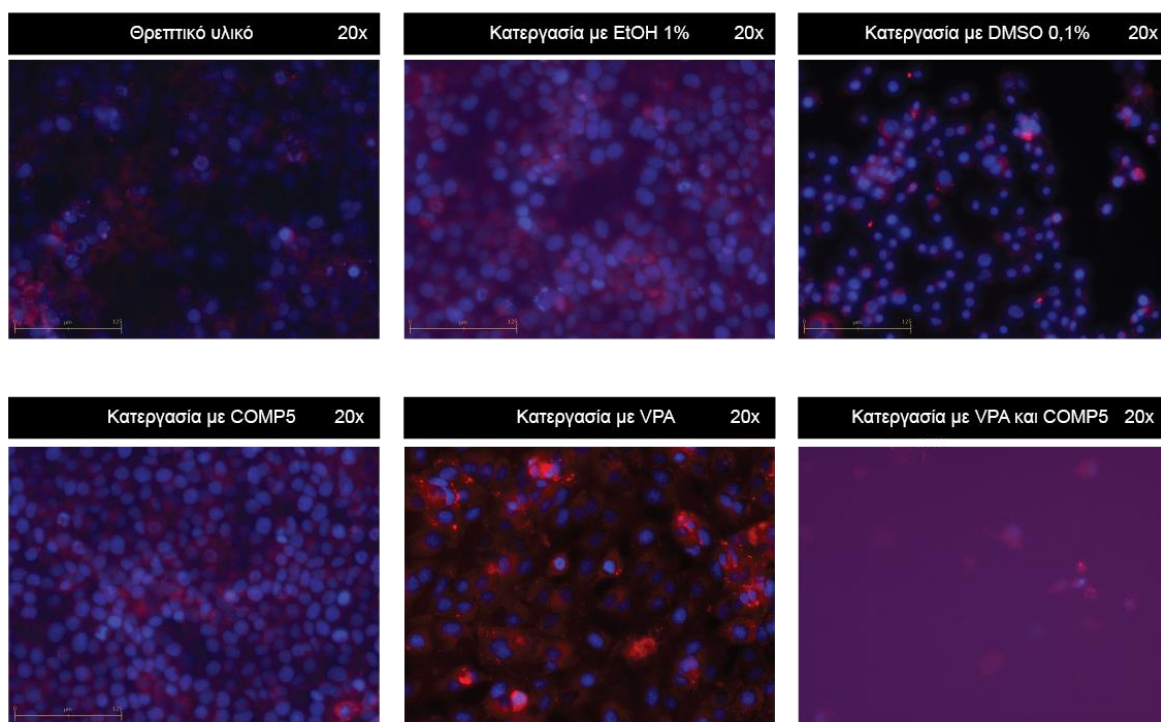
Εικόνα 3.51. Κατεργασία με VPA και COMP2



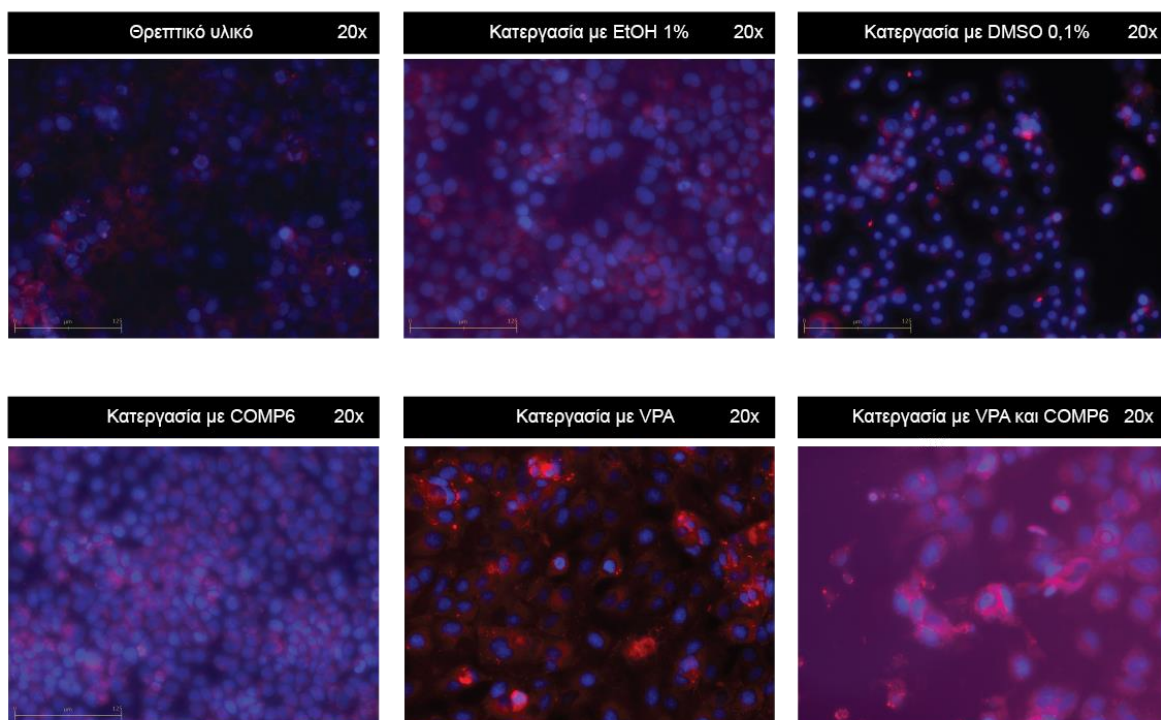
Εικόνα 3.52. Κατεργασία με VPA και COMP3



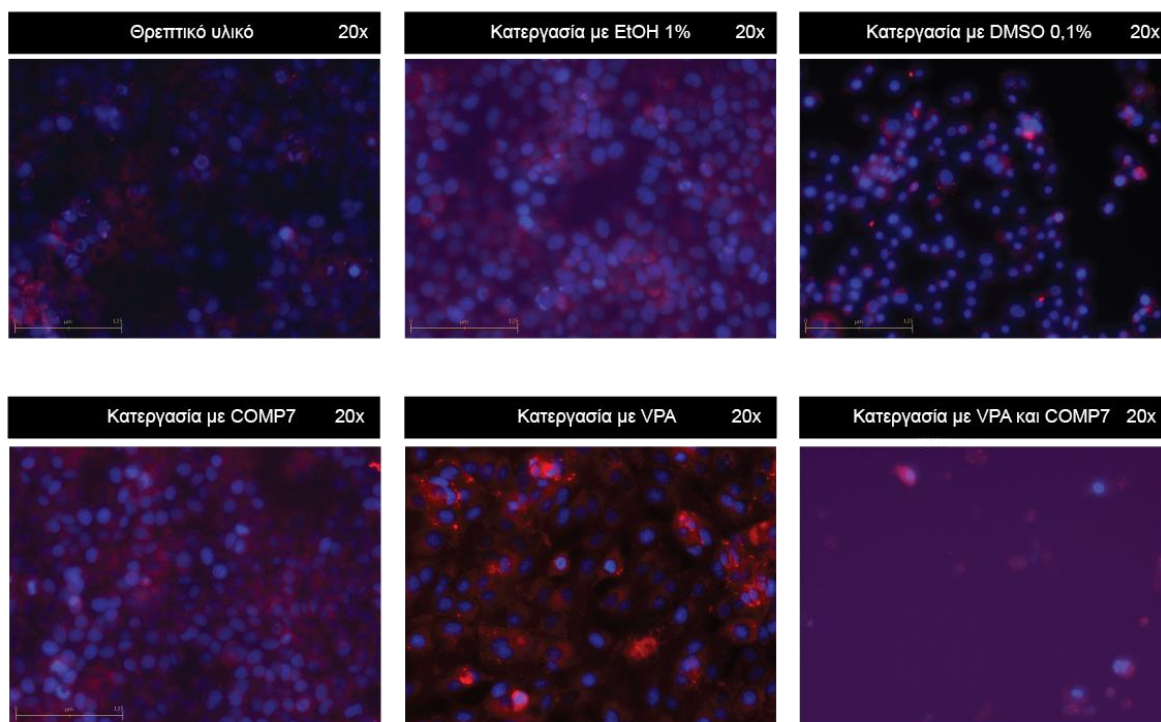
Εικόνα 3.53. Κατεργασία με VPA και COMP4



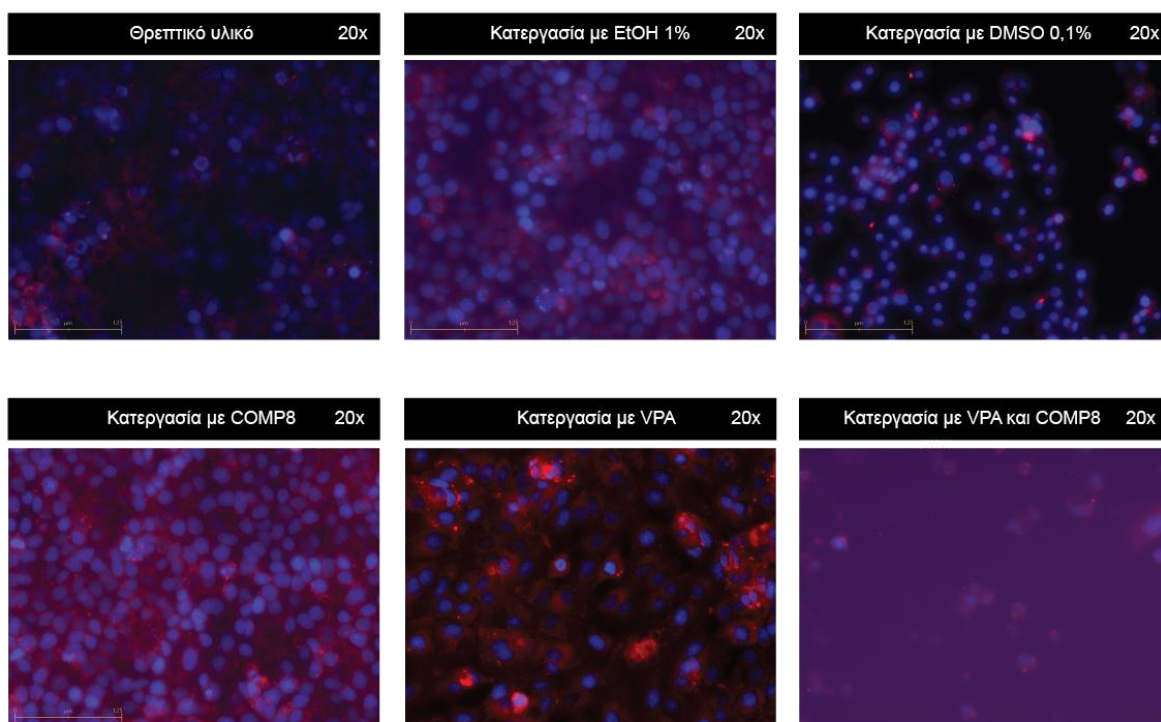
Εικόνα 3.54. Κατεργασία με VPA και COMP5



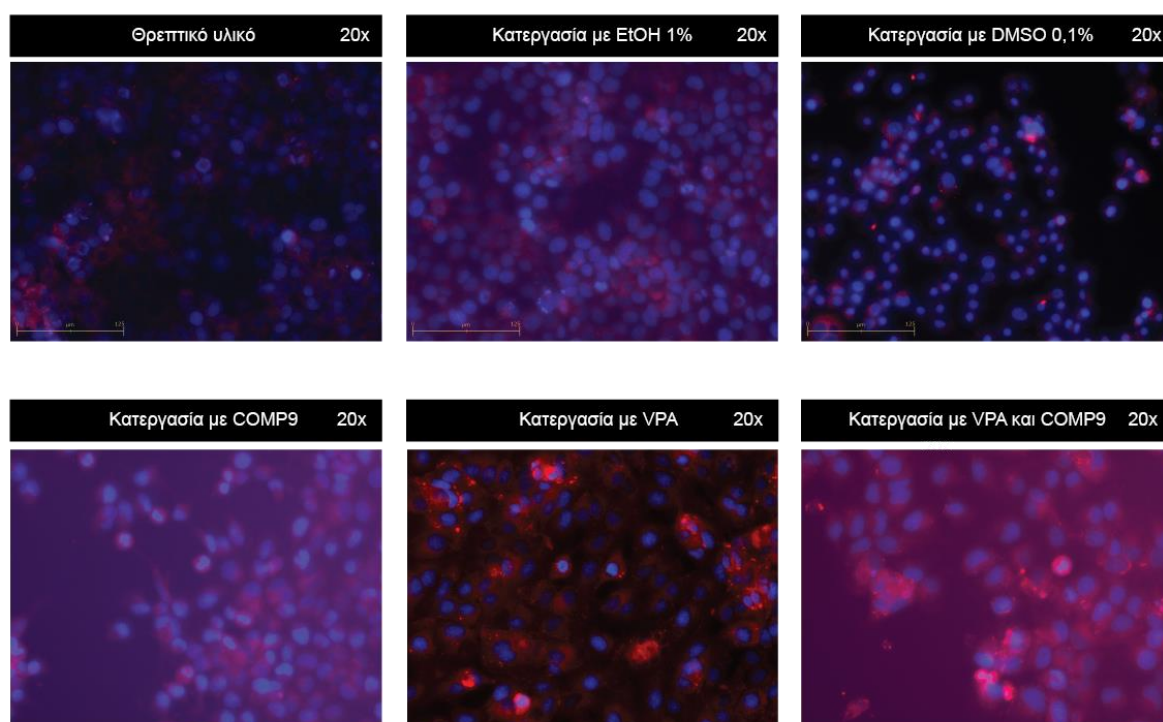
Εικόνα 3.55. Κατεργασία με VPA και COMP6



Εικόνα 3.56. Κατεργασία με VPA και COMP7



Εικόνα 3.56. Κατεργασία με VPA και COMP8



Εικόνα 3.57. Κατεργασία με VPA και COMP9

Κεφάλαιο 4

Συμπεράσματα

Στην παρούσα μελέτη κατασκευάστηκε επιτυχώς μια πλατφόρμα επαναστόχευσης φαρμάκων για τη NAFLD, βασισμένη σε ανάλυση βιολογικών δικτύων. Η πλατφόρμα αυτή βασίζεται στην ανάλυση μικροσυστοιχιών γονιδιακής έκφρασης για τον προσδιορισμό των σημαντικά επηρεασμένων βιολογικών δικτύων της νόσου, και σε συνδυασμό με πληροφορίες από προ-υπάρχοντα *in vitro* μοντέλα χρησιμοποιεί φαρμακολογικές βάσεις δεδομένων για να προσδιορίσει βιοδραστικές ουσίες που επιδρούν σε αυτά τα μονοπάτια. Η πλατφόρμα αυτή μπορεί να τροποποιηθεί σε μικρό βαθμό και να χρησιμοποιηθεί για την επαναστόχευση φαρμάκων για άλλες νόσους που πληρούν αυτές τις προϋποθέσεις.

Ένα από τα αποτελέσματα της μελέτης είναι ο προσδιορισμός των μονοπατιών που σχετίζονται με τη νόσο NAFLD. Παρακάτω παρατίθενται ορισμένες πληροφορίες για τα μονοπάτια αυτά.

- Μονοπάτι πυρηνικών υποδοχέων (BIOCARTA_NUCLEARRS_PATHWAY): Οι πυρηνικοί υποδοχείς είναι μεταγραφικοί παράγοντες που ενεργοποιούνται κατά την πρόσδεση στους συνδέτες τους. Αρχικά, κατηγοριοποιήθηκαν σε κλασικούς ενδοκρινικούς πυρηνικούς ορμονικούς υποδοχείς και ορφανούς υποδοχείς. Ωστόσο, περεταίρω μελέτες έχουν οδηγήσει στην ταυτοποίηση λιπιδικών συνδετών για κάποιους από αυτούς τους ορφανούς υποδοχείς, που είναι υπεύθυνοι για τον μεταβολισμό, αποθήκευση και μείωση των λιπιδίων. Αυτοί οι υποδοχείς συμπεριλαμβάνουν τους peroxisome proliferators-Activated receptors (PPARs) για λιπαρά οξέα, τους liver X receptors (LCR) για οξυστερόλες, Farnesoid X receptors (FXR) για χολικά οξέα και steroid xenobiotic receptor/X receptor (SXR/PXR or Nsil2) για ξενοβιοτικά [46].
- Μονοπάτι της ενδοκύτωσης (KEGG_ENDOCYTOSIS): Περιλαμβάνει γονίδια που συμμετέχουν στο μηχανισμό της ενδοκύτωσης. Κατά τον μηχανισμό αυτό, μόρια που βρίσκονται στο εξωτερικό του κυττάρου και δεν μπορούν να διαπεράσουν την κυτταρική μεμβράνη, μπορούν να μεταφερθούν στο εσωτερικό του κυττάρου μέσω εγκόλπωσης [47].
- Μονοπάτι μεταβολισμού λιπαρών οξέων (KEGG_FATTY_ACID_METABOLISM): Περιλαμβάνει γονίδια που σχετίζονται με το μεταβολισμό των λιπαρών οξέων [48].

- Σηματοδοτικό μονοπάτι των PPAR (KEGG_PPAR_SIGNALING_PATHWAY): Οι peroxisome proliferator-activated receptors (PPARs) είναι πυρηνικοί ορμονικοί υποδοχείς που ενεργοποιούνται από λιπαρά οξέα και παράγωγά τους. Οι PPAR έχουν τρεις υποτύπους, την PPARA, την PPARD και την PPARG, που εκφράζονται διαφορετικά στα σπονδυλωτά. Κάθε υποδοχέας κωδικοποιείται από ξεχωριστό γονίδιο και προσδένεται με λιπαρά οξέα και εικοσανοειδή. Η PPARA συμμετέχει στον καθαρισμό των κυκλοφορούντων ή κυτταρικών λιπιδίων, μέσω της ρύθμισης της έκφρασης γονιδίων που εμπλέκονται στο μεταβολισμό των λιπιδίων στο ήπαρ και στους σκελετικούς μύες. Η PPARD εμπλέκεται στην οξειδωση των λιπιδίων και στον κυτταρικό πολλαπλασιασμό. Τέλος, η PPARG προωθεί την διαφοροποίηση των λιποκυττάρων για την αύξηση της πρόσληψης γλυκόζης στο αίμα [49].
- Μονοπάτι της Activator Protein 1 (PID_AP1_PATHWAY): Η Activator Protein 1 είναι ένας μεταγραφικός παράγοντας, ο οποίος ρυθμίζει τη γονιδιακή έκφραση ως απόκριση σε διάφορα ερεθίσματα, όπως κυτοκίνες, αυξητικοί παράγοντες, στρες, βακτηριακές και ιικές λοιμώξεις. Η AP-1 ελέγχει ορισμένες κυτταρικές διεργασίες όπως η διαφοροποίηση, ο πολλαπλασιασμός και η απόπτωση. Η λειτουργία της AP-1 συχνά ρυθμίζεται μέσω μετά-μεταφραστικών τροποποιήσεων, δημιουργίας διμερών DNA και αλληλεπίδρασης με διάφορους παράγοντες πρόσδεσης. Ο μεταφραστικός παράγοντας AP-1 σχετίζεται με πολλές φυσιολογικές λειτουργίες, κυρίως με τον καθορισμό της διάρκειας ζωής ενός οργανισμού και την αναγέννηση των ιστών [51].
- Μονοπάτι του Activating Transcription Factor 2 (PID_ATF2_PATHWAY): Η Activating transcription factor 2 (ATF2), είναι μια πρωτεΐνη που στους ανθρώπους κωδικοποιείται από το γονίδιο ATF2. Αυτό το γονίδιο κωδικοποιεί έναν μεταγραφικό παράγοντα που είναι μέρος της οικογένειας των φερμουάρ λευκίνης των πρωτεϊνών πρόσδεσης στο DNA. Επίσης συνδέεται με το cAMP-responsive element (CRE) και ενεργοποιεί τη μεταγραφή, ενώ επιδρά και ως ακετυλοτρανσφεράση των ιστών [52].
- Μονοπάτι των πυρηνικών παραγόντων ενεργοποιημένων T- κυττάρων (PID_NFAT_TFPATHWAY): Οι πυρηνικοί παράγοντες ενεργοποιημένων T-κυττάρων (Nuclear factor of activated T-cells; NFAT) είναι μια οικογένεια μεταγραφικών παραγόντων που είναι σημαντικοί στην ανοσολογική απόκριση. Ένα ή περισσότερα μέλη της NFAT οικογένειας εκφράζονται στα περισσότερα κύτταρα του ανοσοποιητικού συστήματος. Οι NFAT συμμετέχουν επίσης στην ανάπτυξη του καρδιακού, σκελετικού μυ και νευρικού συστήματος [50].
- Μονοπάτι μεταβολισμού λιπαρών οξέων, τριακυλογλυκερόλης και κετονοσωμάτων (REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM): Περιλαμβάνει γονίδια που

σχετίζονται με το μεταβολισμό των λιπαρών οξέων, της τριακυλογλυκερόλης και κενοσωμάτων [53].

- Μονοπάτι μεταβολισμού των αμινοξέων και των παραγώγων τους (REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES): Αυτή η ομάδα των αντιδράσεων είναι υπεύθυνη για: 1) τη διάσπαση των αμινοξέων, 2) τη σύνθεση ουρίας από αμμωνία και αμινο-ομάδων που προκύπτουν από τη διάσπαση των αμινοξέων, 3) τη σύνθεση των δέκα αμινοξέων που δεν είναι απαραίτητα στοιχεία της ανθρώπινης διατροφής και 4) τη σύνθεση μορίων που περιέχουν καρτινίνη και κρεατίνη. Η μεταφορά αυτών των μορίων διαμέσου μεμβρανών που αποτελούνται από διπλοστοιβάδες λιπιδίων αναφέρεται ξεχωριστά ως μέρος της διαμεμβρανικής μεταφοράς μικρών μορίων [54].
- Μονοπάτι μεταβολισμού λιπιδίων και λιποπρωτεϊνών (REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS): Περιλαμβάνει γονίδια που σχετίζονται με τον μεταβολισμό λιπιδίων και λιποπρωτεϊνών [55].

Όπως φαίνεται από τις παραπάνω περιγραφές, τα μονοπάτια που έδειξε η ανάλυση ότι είναι σημαντικά φαίνεται να σχετίζονται με τη νόσο, όπως έχουν δείξει προηγούμενες μελέτες. Τα περισσότερα αφορούν στο μεταβολισμό λιπιδίων και λιποπρωτεϊνών ή παραγώγων τους, ή σε μονοπάτια στα οποία ανήκουν μεταγραφικοί παράγοντες οι οποίοι είναι γνωστό από τη βιβλιογραφία ότι εμπλέκονται στη NAFLD [56].

Στην παράγραφο 3.1.6. παρουσιάστηκαν ορισμένα φάρμακα που είναι υποσχόμενα για να αντιστρέψουν τη νόσο NAFLD. Από αυτά, σύμφωνα με τη βιβλιογραφία, η πλειοψηφία δεν έχουν χρησιμοποιηθεί για το σκοπό αυτό. Τα Estradiol [57], Ivermectin [58], Raloxifene [59] και Sirolimus [60] έχουν δοκιμαστεί *in vivo* σε αρουραίους (*Rattus norvegicus* sp), ενώ το Resveratrol [61] βρίσκεται σε κλινικές δοκιμές. Το γεγονός ότι αρκετά από τα φάρμακα έχουν ήδη δοκιμαστεί για την αντιμετώπιση της νόσου είναι πολύ θετικό καθώς δείχνει ότι η ανάλυση που ακολουθήθηκε έχει σωστή κατεύθυνση.

Από τα αποτελέσματα των πειραματικών διαδικασιών αξίζει να σχολιαστούν οι καμπύλες δόσης επιβίωσης. Όπως φαίνεται από τις χαμηλές τιμές R^2 , δεν προσεγγίζουν αρκετά καλά την θεωρητική μορφή τους και ο υπολογισμός της συγκέντρωσης EC_{90} δεν είναι ακριβής. Η μορφή της καμπύλης πιθανότατα να οφείλεται στην όχι καλή διάλυση των ουσιών στις κατάλληλες συγκεντρώσεις. Επίσης θα ήταν θεμιτό να υπάρχει μεγαλύτερη συγκέντρωση από 10μM ή μεγαλύτερος χρόνος επώασης, καθώς στις περισσότερες περιπτώσεις δεν παρουσιάζεται μεγάλη μεταβολή της βιωσιμότητας μετά το πρώτο 24ωρο.

Τέλος, οι εικόνες της παραγράφου 3.2.2. απεικονίζουν την επίδραση κάθε βιοδραστικής ουσίας σε κάθε μέθοδο επαγωγής NAFLD και τα συμπεράσματα παρουσιάζονται στον Πίνακα 4.1.. Με (✓) σημειώνονται οι περιπτώσεις στις οποίες οι βιοδραστικές ουσίες προκαλούν μείωση του ενδοκυτταρικού φορτίου λίπους και με (X) οι περιπτώσεις στις οποίες δεν παρατηρήθηκε κάποια

μεταβολή. Με (?) σημειώνονται οι συνδυασμοί στους οποίους παρατηρείται κάποια μείωση αλλά απαιτείται επιπλέον χρόνος επώασης ή υψηλότερη συγκέντρωση της βιοδραστικής ουσίας για εμφανή αποτελέσματα. Στις ίδιες εικόνες φαίνεται ότι καμία από τις βιοδραστικές ουσίες που δοκιμάστηκαν δεν προκαλεί NAFLD.

Μέθοδος επαγωγής NAFLD	Βιοδραστικές ουσίες								
	COM P1	COM P2	COM P3	COM P4	COM P5	COM P6	COM P7	COM P8	COM P9
Free Fatty Acid	X	X	√	X	√	√	√	√	√
Amiodarone	X	?	X	?	?	√	X	X	?
Tamoxifen	X	X	X	√	√	?	√	√	?
Tetracycline	X	X	X	X	?	X	√	√	?
Valproic acid	?	X	?	√	?	?	√	?	?

Πίνακας 4.1. Ποιες βιοδραστικές ουσίες επιδρούν σε κάθε μοντέλο

Μελλοντική εργασία

Κατά τη μελέτη, η μέθοδος με την οποία έγινε η ανάλυση των μονοπατιών είναι η Functional Class Scoring (FCS). Χρησιμοποιώντας την μέθοδο Pathway Topology Based (PT) μπορεί να ληφθεί υπόψιν σε μεγαλύτερο βαθμό η αλληλεπίδραση των γονιδίων και να ενισχυθούν τα παραπάνω αποτελέσματα.

Ένας άλλος τρόπος για να βελτιωθεί η ανάλυση, είναι να προσδιοριστεί πιο λεπτομερώς η επίδραση των ουσιών που επάγουν τη νόσο σε κάθε μονοπάτι. Αυτό μπορεί να πραγματοποιηθεί μέσω στατιστικής ανάλυσης του βάρους κάθε γονιδίου στόχου μέσα στο αντίστοιχο μονοπάτι. Αντίστοιχα μπορεί να προσδιοριστεί η επίδραση των βιοδραστικών ουσιών που αντιστρέφουν τη νόσο στα μονοπάτια αυτά.

Παραπάνω αναφέρθηκε ότι η πειραματική μεθοδολογία που ακολουθήθηκε μπορεί να εξάγει αξιόπιστα αποτελέσματα καθώς κάποιες ουσίες που είναι γνωστές και από τη βιβλιογραφία έδειξαν σημαντικό περιορισμό της νόσου. Στόχος μελλοντικής εργασίας είναι η *in vitro* μελέτη όλων των βιοδραστικών ουσιών που προήλθαν από την παρούσα ανάλυση, τόσο σε επίπεδο αντιστροφής της νόσου όσο και σε επίπεδο τοξικότητας και σε επίπεδο μεταγωγής σημάτων μέσω πρωτεωμικών αναλύσεων. Τέλος, θα ήταν καλό να αναπτυχθεί ένας αλγόριθμος για την ανάλυση των εικόνων που προκύπτουν από το High Content Screening για να ποσοτικοποιηθεί το ενδοκυτταρικό λιπιδικό φορτίο και να είναι δυνατό να συγκριθούν και μαθηματικά οι διαφορετικές περιπτώσεις.

Βιβλιογραφία

- [1] Alexopoulos, L, Αλεξόπουλος, Λ. Συστημική Βιολογία. Εμβιομηχανική – Βιοιατρική Τεχνολογία. Διαθέσιμο στο: <https://ocw.aoc.ntua.gr/modules/document/file.php/MECH114/%CE%A3%CE%B7%CE%BC%CE%B5%CE%B9%CF%8E%CF%83%CE%B5%CE%B9%CF%82/6o.pdf>
- [2] Theodoridis, G., Girousi, S., Zachariadis, G., Zotou, A. S., Samanidou, V., Θεοδωρίδης, Γ., ... & Σαμανίδου, Β. (2015). *Βιοαναλυτική χημεία*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/3667>
- [3] Νικολάου, C., Chouvardas, P., Νικολάου, Χ., & Χουβαρδάς, Π. (2015). Εισαγωγή. Τι είναι η Υπολογιστική Βιολογία. Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/1578>
- [4] Λεφέβρ, Ν. (2016). *Σύνθετα δίκτυα: μελέτη και ανάλυση βιολογικών δικτύων-εφαρμογή στην διάδοση ασθενειών*. (Πτυχιακή εργασία, ΤΕΙ Ηπείρου).
- [5] Davidson, E. H. (2010). *The regulatory genome: gene regulatory networks in development and evolution*. Academic press.
- [6] Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., & Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912), 190.
- [7] Eissmann, P., Evans, J. H., Mehrabi, M., Rose, E. L., Nedvetzki, S., & Davis, D. M. (2010). Multiple mechanisms downstream of TLR-4 stimulation allow expression of NKG2D ligands to facilitate macrophage/NK cell crosstalk. *The Journal of Immunology*, 184(12), 6901-6909.
- [8] Meyers, R. A. (2009). *Encyclopedia of complexity and systems science*. Springer.(pp.719-741).
- [9] Wilson, N. S., Dixit, V., & Ashkenazi, A. (2009). Death receptor signal transducers: nodes of coordination in immune signaling networks. *Nature immunology*, 10(4), 348-355.
- [10] Bhalla, U. S., & Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science*, 283(5400), 381-387.
- [11] Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316(5830), 1497-1502.
- [12] Emmert-Streib, F., & Glazko, G. V. (2011). Pathway analysis of expression data: deciphering functional building blocks of complex diseases. *PLoS computational biology*, 7(5), e1002053.
- [13] Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., ... & Houstis, N. (2003). PGC-1 [alpha]-responsive genes

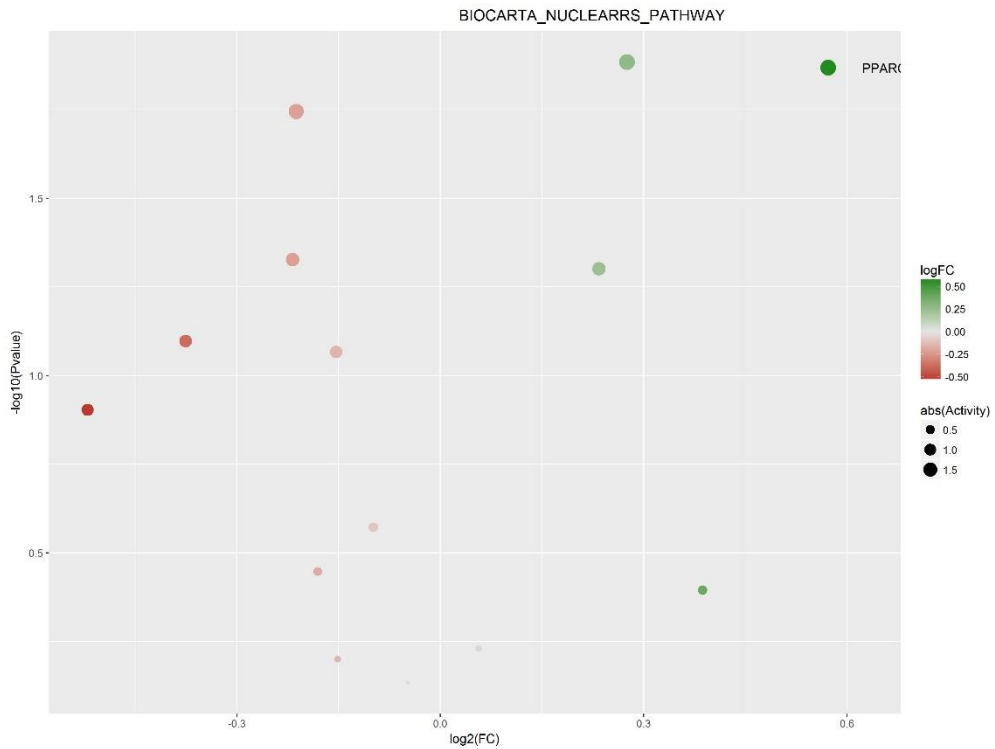
- involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*, 34(3), 267.
- [14] Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS computational biology*, 8(2), e1002375.
- [15] PhRMA. (n.d.). [Brochure]. Author. Retrieved October 10, 2017, from http://cmidd.northwestern.edu/files/2015/10/Drug_RD_Brochure-12e7vs6.pdf
- [16] Modernizing Drug Discovery, Development & Approval. (n.d.). Retrieved October 4, 2017, from <http://www.phrma.org/report/modernizing-drug-discovery-development-and-approval1>
- [17] Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., ... & Butte, A. J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine*, 3(96), 96ra77-96ra77.
- [18] Deotarse, P. P., Jain, A. S., Baile, M. B., Kolhe, N. S., & Kulkarni, A. A. (2015). Drug repositioning: a review. *International Journal of Pharma Research & Review*, 4(8), 51-58.
- [19] Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, 3(8)
- [20] Angulo, P. (2002). Nonalcoholic fatty liver disease. *New England Journal of Medicine*, 346(16), 1221-1231.
- [21] Alkhoury, N., Dixon, L. J., & Feldstein, A. E. (2009). Lipotoxicity in nonalcoholic fatty liver disease: not all lipids are created equal. *Expert review of gastroenterology & hepatology*, 3(4), 445-451.
- [22] Dowman, J. K., Tomlinson, J. W., & Newsome, P. N. (2009). Pathogenesis of non-alcoholic fatty liver disease. *QJM: An International Journal of Medicine*, 103(2), 71-83.
- [23] Dowman, J. K., Tomlinson, J. W., & Newsome, P. N. (2011). Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Alimentary pharmacology & therapeutics*, 33(5), 525-540.
- [24] Ζαρείφη, Δ. Σ. (2017). Πρωτεομική ανάλυση της μη-αλκοολικής λιπώδους νόσου του ήπατος σε *in vitro* μοντέλα πρωτογενών ανθρώπινων ηπατοκυττάρων. (Μεταπτυχιακή Μελέτη, Γεωπονικό Πανεπιστήμιο Αθηνών). Αθήνα.
- [25] Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, 30(1), 207-210.
- [26] Lee, M. L. T., Kuo, F. C., Whitmore, G. A., & Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of the National Academy of Sciences*, 97(18), 9834-9839.
- [27] Gillespie, C. (2011). Volcano plots of microarray data. *The Bioinformatics Knowledgeblog*.

- [28] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7), e47-e47.
- [29] Våremo, L., Nielsen, J., & Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic acids research*, 41(8), 4378-4391.
- [30] Molecular Signatures Database v6.1. (n.d.). Retrieved May 26, 2017, from <http://software.broadinstitute.org/gsea/msigdb>
- [31] Fisher, R. A. (1932). Statistical methods for research workers. Edinburgh: Oliver and Boyd, 1925. *Google Scholar*.
- [32] Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams Jr, R. M. (1949). The American soldier: Adjustment during army life. (Studies in social psychology in World War II), Vol. 1.
- [33] Oliveira, A. P., Patil, K. R., & Nielsen, J. (2008). Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC systems biology*, 2(1), 17.
- [34] Taylor, J., & Tibshirani, R. (2005). A tail strength measure for assessing the overall univariate significance in a dataset. *Biostatistics*, 7(2), 167-181.
- [35] Kim, S. Y., & Volsky, D. J. (2005). PAGE: parametric analysis of gene set enrichment. *BMC bioinformatics*, 6(1), 144.
- [36] [Efron, B., & Tibshirani, R. (2007). On testing the significance of sets of genes. *The annals of applied statistics*, 107-129.
- [37] Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., ... & Tang, A. (2013). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research*, 42(D1), D1091-D1097.
- [38] Toutenburg, H. (1975). Hollander, M., DA Wolfe: Nonparametric statistical methods. John Wiley & Sons, New York-Sydney-Tokyo-Mexico City 1973. 503 S., \$9.50. *Biometrical Journal*, 17(8), 526-526.
- [39] Qu, X. A., & Rajpal, D. K. (2012). Applications of Connectivity Map in drug discovery and development. *Drug discovery today*, 17(23), 1289-1298.
- [40] Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., ... & Reich, M. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *science*, 313(5795), 1929-1935.
- [41] Σιαβέλης, Ι. (2015). Μέθοδοι βιοπληροφορικής για τον επαναπροσδιορισμό φαρμάκων στη νόσο Αλτσχάιμερ (Μεταπτυχιακή διπλωματική εργασία, Πανεπιστήμιο Πατρών, ΕΜΠ). Αθήνα.
- [42] Hardt, C., Beber, M. E., Rasche, A., Kamburov, A., Hebels, D. G., Kleinjans, J. C., & Herwig, R. (2016). ToxDB: pathway-level interpretation of drug-treatment data. *Database*, 2016.
- [43] Search Livertox Database. (n.d.). Retrieved October 10, 2017, from <http://www.livertox.nih.gov/>

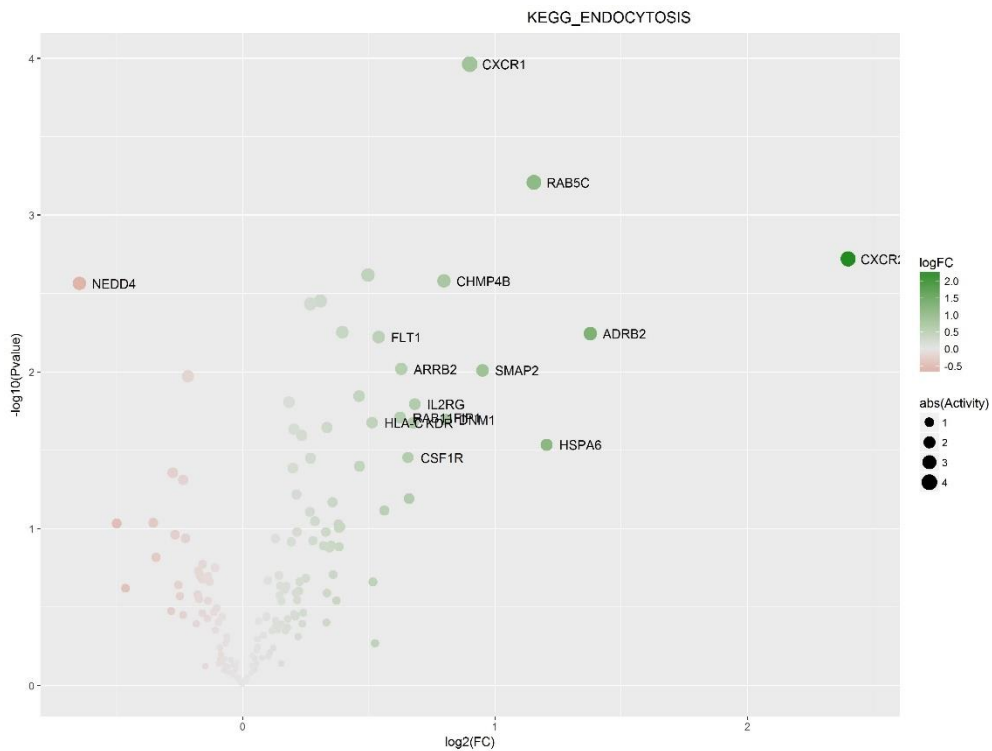
- [44] Riss, T. L., Moravec, R. A., Niles, A. L., Duellman, S., Benink, H. A., Worzella, T. J., & Minor, L. (2016). Cell viability assays.
- [45] Motulsky, H. (1999). *Analyzing data with GraphPad prism*. GraphPad Software Incorporated.
- [46] Gene Set: BIOCARTA_NUCLEARRS_PATHWAY. (n.d.). Retrieved October 15, 2017, from http://software.broadinstitute.org/gsea/msigdb/cards/BIOCARTA_NUCLEARRS_PATHWAY
- [47] Gene Set: KEGG_ENDOCYTOSIS. (n.d.). Retrieved October 15, 2017, from http://software.broadinstitute.org/gsea/msigdb/cards/KEGG_ENDOCYTOSIS
- [48] Gene Set: KEGG_FATTY_ACID_METABOLISM. (n.d.). Retrieved October 15, 2017, from http://software.broadinstitute.org/gsea/msigdb/cards/KEGG_FATTY_ACID_METABOLISM
- [49] (n.d.). Retrieved October 15, 2017, from http://www.genome.jp/dbget-bin/www_bget?pathway%3Amap03320
- [50] NFAT. (2017, July 25). Retrieved October 15, 2017, from <https://en.wikipedia.org/wiki/NFAT>
- [51] AP-1 transcription factor. (2017, October 11). Retrieved October 15, 2017, from https://en.wikipedia.org/wiki/AP-1_transcription_factor
- [52] Activating transcription factor 2. (2017, October 18). Retrieved October 18, 2017, from https://en.wikipedia.org/wiki/Activating_transcription_factor_2
- [53] Gene Set: REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM. (n.d.). Retrieved October 19, 2017, from http://software.broadinstitute.org/gsea/msigdb/cards/REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_METABOLISM
- [54] P., D., NYU School of Medicine, Department of Biochemistry, & P. (n.d.). Retrieved October 15, 2017, from http://www.reactome.org/content/detail/REACT_13
- [55] Gene Set: REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS. (n.d.). Retrieved October 15, 2017, from http://software.broadinstitute.org/gsea/msigdb/cards/REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
- [56] Berlanga, A., Guiu-Jurado, E., Porrás, J. A., & Auguet, T. (2014). Molecular pathways in non-alcoholic fatty liver disease. *Clinical and experimental gastroenterology*, 7, 221.
- [57] Zhang, H., Liu, Y., Wang, L., Li, Z., Zhang, H., Wu, J., ... & Huhtaniemi, I. (2013). Differential effects of estrogen/androgen on the prevention of nonalcoholic fatty liver disease in the male rat. *Journal of lipid research*, 54(2), 345-357.

- [58] Jin, L., Wang, R., Zhu, Y., Zheng, W., Han, Y., Guo, F., ... & Li, Y. (2015). Selective targeting of nuclear receptor FXR by avermectin analogues with therapeutic effects on nonalcoholic fatty liver disease. *Scientific reports*, 5.
- [59] Luo, F., Ishigami, M., Achiwa, K., Ishizu, Y., Kuzuya, T., Honda, T., ... & Goto, H. (2015). Raloxifene ameliorates liver fibrosis of nonalcoholic steatohepatitis induced by choline-deficient high-fat diet in ovariectomized mice. *Digestive diseases and sciences*, 60(9), 2730-2739.
- [60] Wang, Y., Shi, M., Fu, H., Xu, H., Wei, J., Wang, T., & Wang, X. (2010). Mammalian target of the rapamycin pathway is involved in non-alcoholic fatty liver disease. *Molecular medicine reports*, 3(6), 909-915.
- [61] Zhang, C., Yuan, W., Fang, J., Wang, W., He, P., Lei, J., & Wang, C. (2016). Efficacy of Resveratrol Supplementation against Non-Alcoholic Fatty Liver Disease: A Meta-Analysis of Placebo-Controlled Clinical Trials. *PloS one*, 11(8), e0161792.

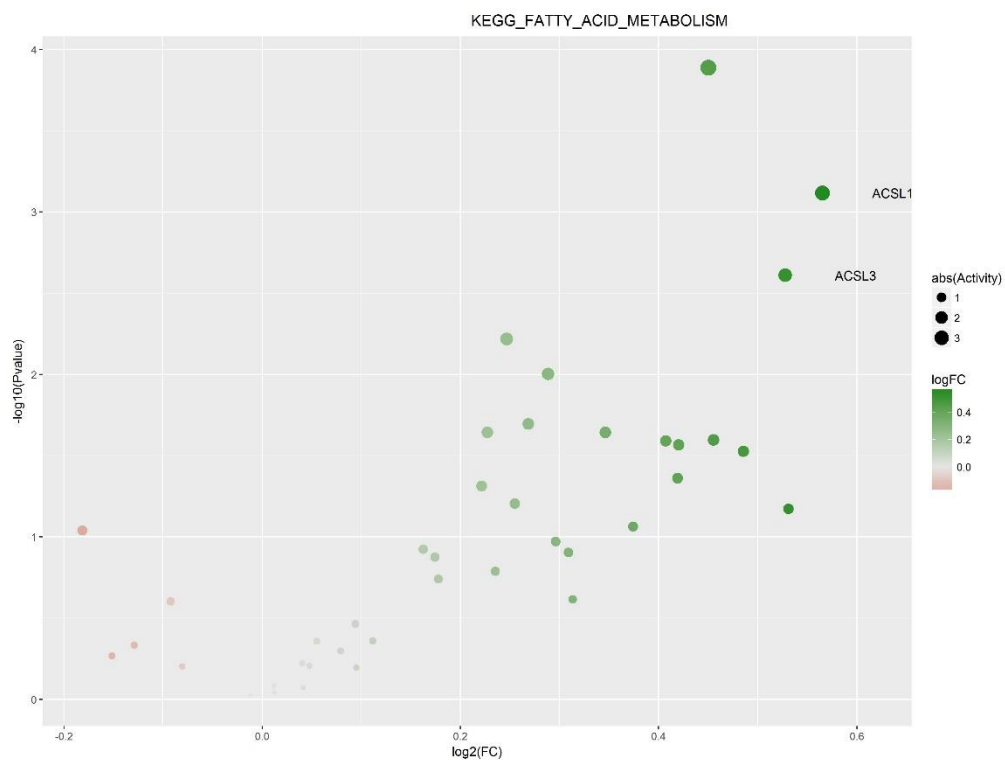
Παράρτημα Α



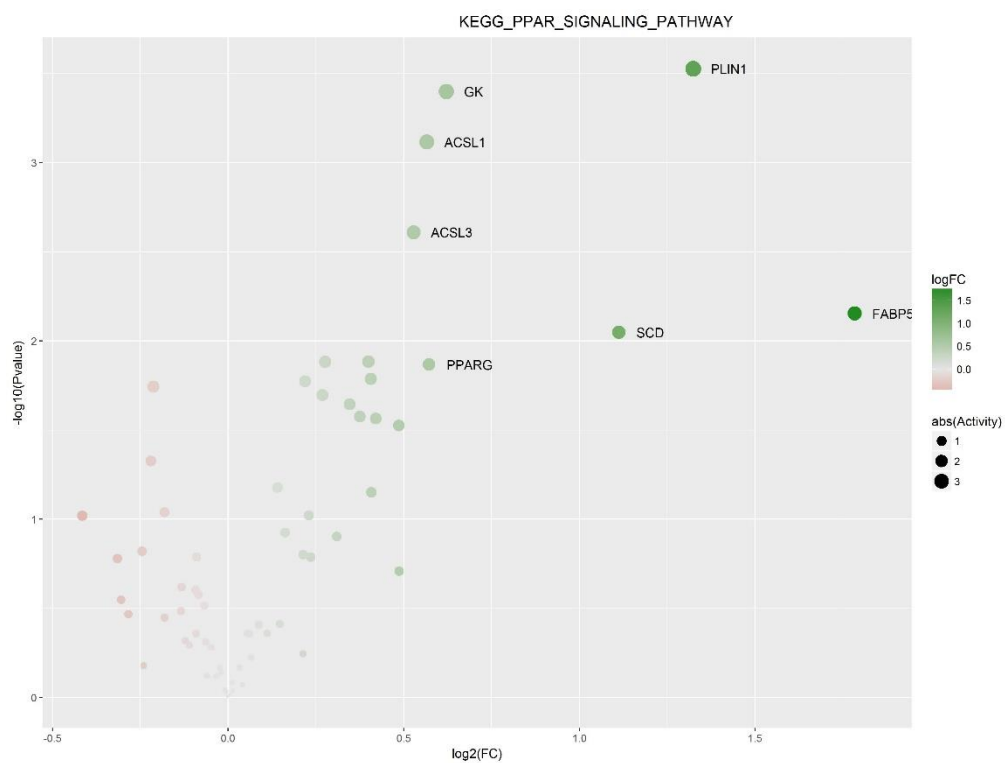
Εικόνα Α1. Volcano plot - Μονοπάτι πυρηνικών υποδοχέων (GSE63067 NASH VS. Healthy)



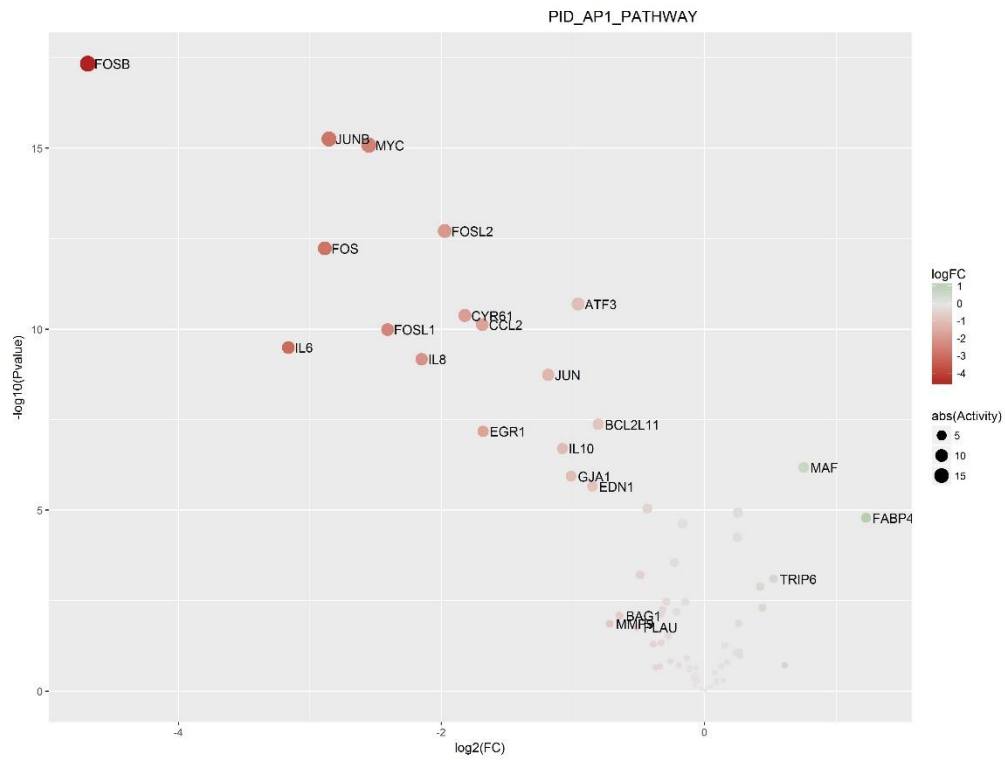
Εικόνα Α2. Volcano plot - Μονοπάτι της ενδοκύτωσης (GSE63067 Steatosis Vs. Healthy)



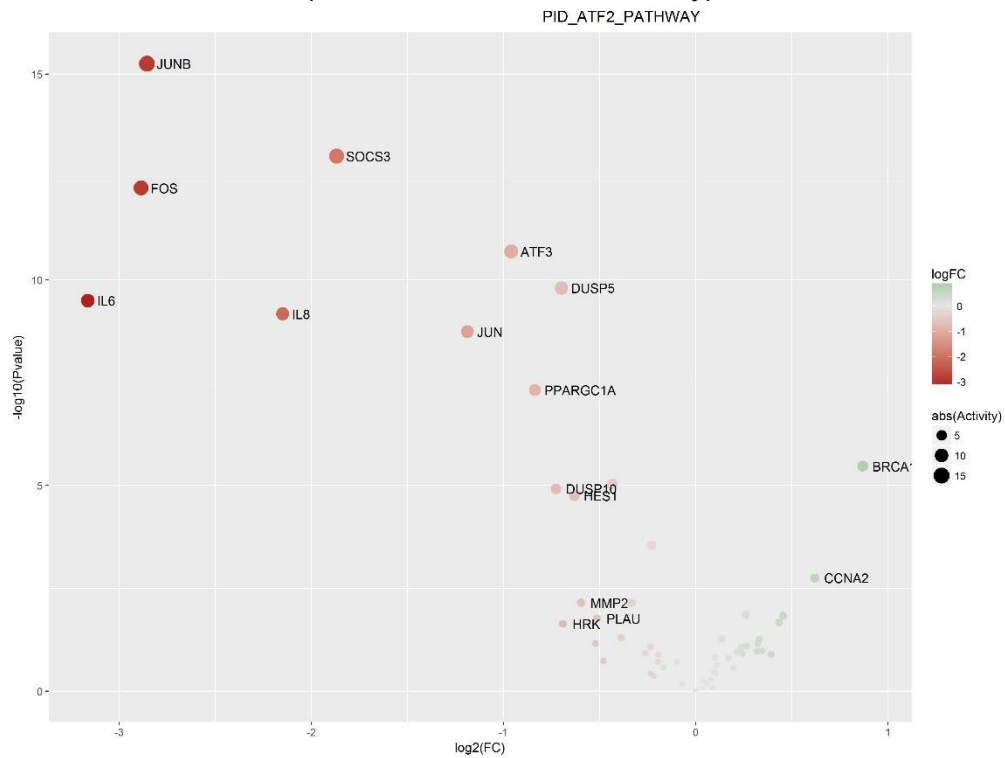
Εικόνα A3. Volcano plot - Μονοπάτι μεταβολισμού λιπαρών οξέων (GSE63067 NASH Vs. Healthy)



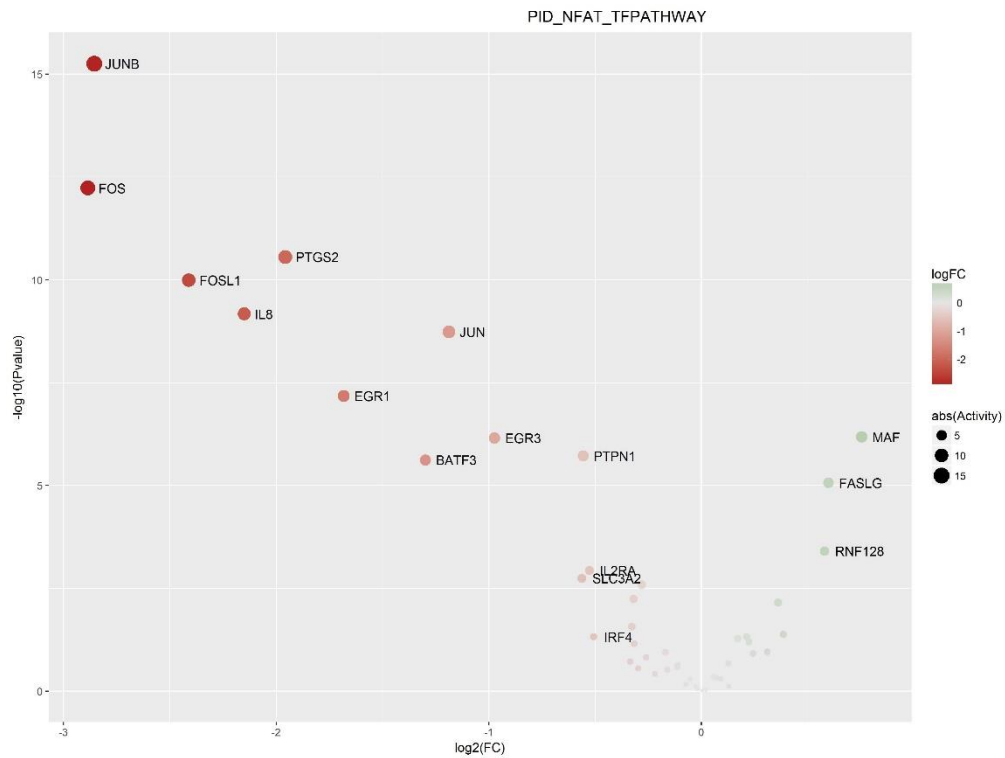
Εικόνα A4. Volcano plot – Σηματοδοτικό μονοπάτι των PPAR (GSE63067 NASH Vs. Healthy)



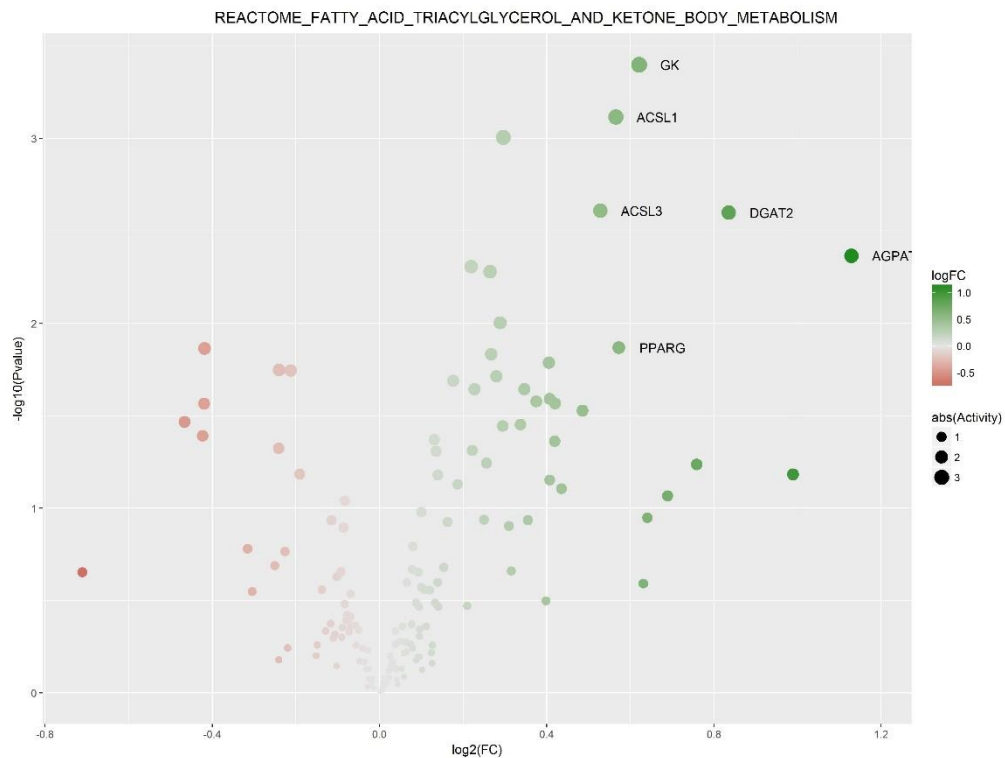
Εικόνα Α5. Volcano plot – Μονοπάτι της Activator Protein 1 (GSE89632 NASH Vs. Healthy)



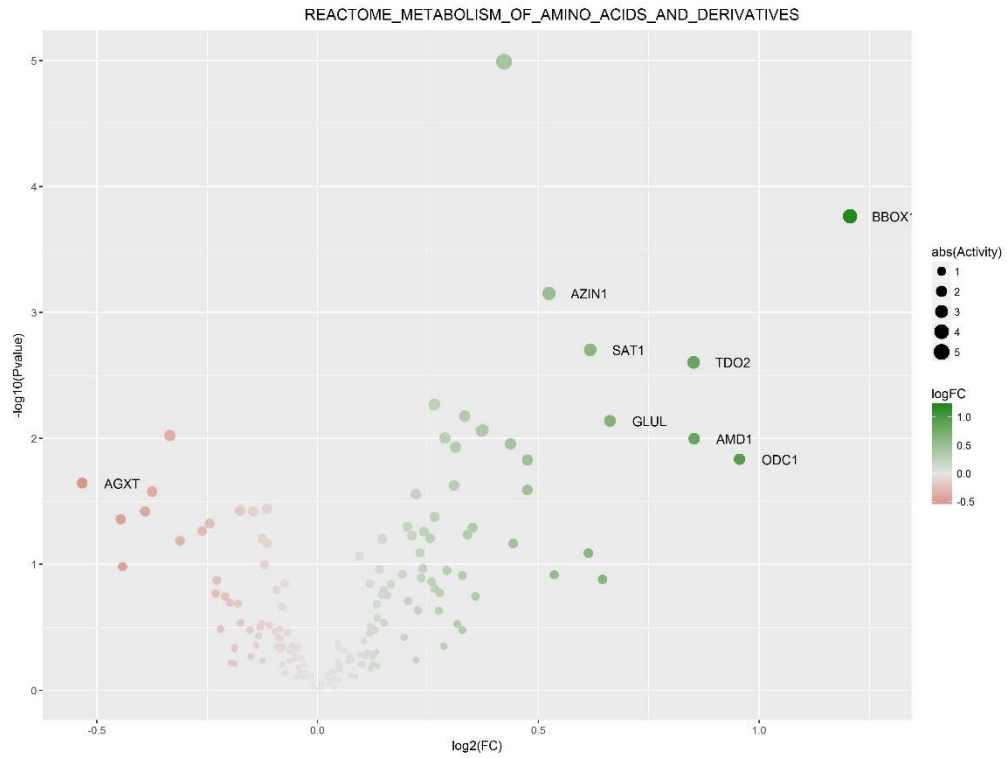
Εικόνα Α6. Volcano plot – Μονοπάτι του Activating Transcription Factor 2 (GSE89632 NASH Vs. Healthy)



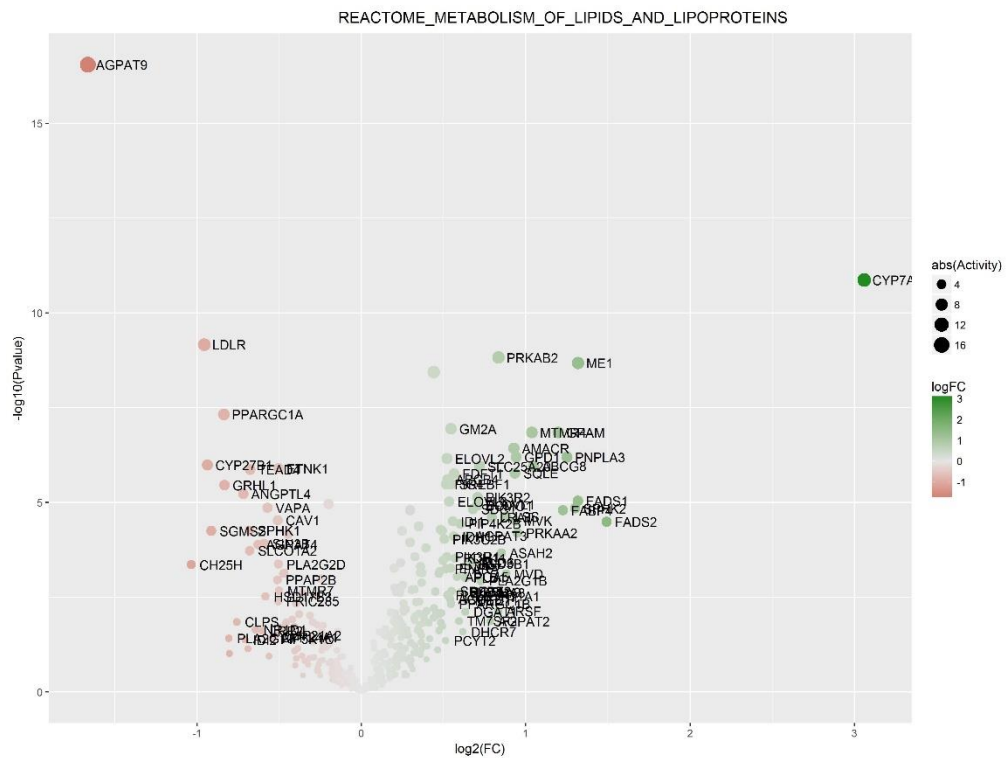
Εικόνα Α7. Volcano plot – Μονοπάτι των πυρηνικών παραγόντων ενεργοποιημένων T-κυττάρων (GSE89632 NASH Vs. Healthy)



Εικόνα Α8. Volcano plot – Μονοπάτι μεταβολισμού λιπαρών οξέων, τριακυλογλυκερόλης και κενοσωμάτων (GSE63067 NASH Vs. Healthy)



Εικόνα A9. Volcano plot – Μονοπάτι μεταβολισμού των αμινοξέων και των παραγώγων τους (GSE63067 NASH Vs. Healthy)



Εικόνα A10. Volcano plot – Μονοπάτι μεταβολισμού λιπιδίων και λιποπρωτεϊνών (GSE89632 NASH Vs. Healthy)